Probabilités discrètes [1]

Olivier Hénard

11 janvier 2022

1. Notes du cours « Probabilités » (MEU254) de la deuxième année de licence de Mathématiques à l'Université Paris-Saclay

Table des matières

Chapitre 1

Espaces de probabilité et événements, cas finis et dénombrables

1.1 Dénombrement

On commence par quelques rappels de première année, menés tambour battant.

1.1.1 Application

Soit E et F deux ensembles et $f : E \to F$ une application [1]. Si $x \in E$ et $y \in F$ vérifient $f(x) = y$, on dit que y est l'image de x par f, et que x est un antécédant de y par f. À la différence de l'image (unique par définition d'une application), un antécédant peut ne pas exister ou au contraire, il peut exister plusieurs antécédants.

Ceci donne lieu aux définitions suivantes. Une fonction $f : E \to F$ est dite

- *surjective* si tout élément de F admet au moins un antécédant par f : pour tout $y \in F$, il existe $x \in E$ tel que $y = f(x)$.
- *injective* si tout élément de F admet au plus un antécédant par f : pour tout $x_1, x_2 \in E$, $f(x_1) = f(x_2)$ implique $x_1 = x_2$.
- *bijective* si elle est à la fois injective et surjective : pour tout $y \in F$, il existe un unique $x \in E$ tel que $y = f(x)$.

Noter qu'une fonction peut n'être ni injective ni surjective, et que ces notions dépendent étroitement du choix de E et de F, comme le montre l'exemple suivant.

Exemple 1.1. Considérons l'application $x \mapsto x^2$ par exemple :

- Considérée de $\mathbb{R}$ dans $\mathbb{R}^+$ elle est surjective ($y \geq 0$ est le carré de $\sqrt{y}$) mais non injective ($y \geq 0$ est aussi le carré de $-\sqrt{y}$, distinct de $\sqrt{y}$ dès lors que $y \neq 0$).
- Considérée de $\mathbb{R}^+$ dans $\mathbb{R}^+$ elle est bijective.
- Considérée de $\mathbb{R}$ dans $\mathbb{R}$, elle n'est ni injective ni surjective.

1. une fonction dont le domaine de définition est égal à l'ensemble E entier, c'est-à-dire que chaque élément de E admet une image

1.1.2 Image directe et image réciproque

On peut étendre f en une fonction de l'ensemble $\mathcal{P}(E)$ des parties (ou sous-ensembles, nous emploierons les deux termes indifféremment) de E vers l'ensemble $\mathcal{P}(F)$ des parties de F :

$$f : \mathcal{P}(E) \to \mathcal{P}(F), A \mapsto f(A) = \{f(x) : x \in A\}$$

$f(A)$ est appelée *image directe* de l'ensemble A, ses éléments sont les images des éléments de A. (Il s'agit à proprement parler d'un abus de notation, puisque f désigne déjà une fonction de E dans F ; comme la nature des arguments de ces deux applications, un élément ou un ensemble, diffère, il n'y a toutefois pas de risque d'ambiguité ; par exemple, $f(\{x\}) = \{f(x)\}$). On a par définition $f(E) \subset F$, avec égalité ssi si f est surjective.

L'image directe est compatible avec l'opération d'union : si A_1 et A_2 sont deux parties de E,

$$f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$$

(le vérifier à l'aide des définitions). Attention en revanche : en général, on a seulement

$$f(A_1 \cap A_2) \subset f(A_1) \cap f(A_2)$$

et l'inclusion peut être stricte, si par exemple il existe un élément $y \in F$ qui admet exactement deux antécédants $x \in A \cap B^c$ et $y \in B \cap A^c$. En effet $z \in f(A) \cap f(B)$ mais $z \notin f(A \cap B)$

On peut *toujours* associer à f une fonction f^{-1} de l'ensemble $\mathcal{P}(F)$ des parties de F vers l'ensemble $\mathcal{P}(E)$ des parties de E :

$$f^{-1} : \mathcal{P}(F) \to \mathcal{P}(E), B \mapsto f^{-1}(B) = \{x : f(x) \in B\}$$

$f^{-1}(B)$ est appelée l'image réciproque de la l'ensemble B, c'est l'ensemble des antécédants des éléments de B.

L'image réciproque est encore compatible avec les opérations d'union et d'intersection : $f^{-1}(F) = E$, et si B_1 et B_2 sont deux parties de F,

$$f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2) \text{ et } f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2).$$

(le vérifier à l'aide des définitions). On peut reformuler injectivité et surjectivité à l'aide de la notion d'image réciproque :

— f est injective si $\text{Card}\, f^{-1}(\{y\}) \leq 1$ pour tout $y \in F$,
— f est surjective si $\text{Card}\, f^{-1}(\{y\}) \geq 1$ pour tout $y \in F$,
— f est donc bijective si $\text{Card}\, f^{-1}(\{y\}) = 1$ pour tout $y \in F$.

Dans le cas où f est bijective, *et dans ce cas seulement*, f^{-1} fait sens en tant que fonction de F dans E, on note simplement $f^{-1}(y)$ l'unique antécédant de y par f, c'est-à-dire qu'on *définit* $\{f^{-1}(y)\} := f^{-1}(\{y\})$. On appelle alors f^{-1} l'application réciproque de f.

Noter que si l'on restreint l'ensemble d'arrivée F de f à $f(E)$, alors l'application $f : E \to f(E)$ est toujours surjective.

Rappel de théorie des ensembles

Soit Ω un ensemble, et A, B, C trois parties (ou sous-ensembles) de Ω. Rappelons la symétrie et distributivité des symboles d'intersection $\cap$ et d'union $\cup$:

$$A \cap B = B \cap A, \quad A \cup B = B \cup A$$

$$(A \cap B) \cup C = (A \cap C) \cup (B \cup C), \quad (A \cup B) \cap C = (A \cap C) \cup (B \cup C).$$

On rappelle la notation $A^c := \Omega \setminus A$ pour le complémentaire de l'ensemble A, l'ensemble des éléments de Ω qui n'appartiennent pas à A On rappelle alors les lois de Morgan :

$$(A \cup B)^c = A^c \cap B^c \text{ et } (A \cap B)^c = A^c \cup B^c$$

Ces identités se montrent par double inclusion.

On dit que les ensembles $(A_i)_{1 \leq i \leq n}$ forment une partition de Ω s'ils vérifient :
— pour tout $i \in \{1, \ldots\}$, $A_i \neq \emptyset$.
— si $i \neq j$, $A_i \cap A_j = \emptyset$ (*ensembles deux à deux disjoints*)
— $\cup_{i \in \{1,\ldots,n\}} A_i = \Omega$.

En toutes lettres, les $(A_i)_{i \in I}$ sont non vides, deux à deux disjoints, et leur réunion donne l'ensemble Ω entier.

1.1.3 Cardinal d'un ensemble fini.

Définition 1.2. Soit Ω un ensemble. Ω est dit *fini* s'il existe un entier $n \in \mathbb{N}$ et une bijection $f : \{1, \ldots, n\} \to \Omega$. Dans ce cas, l'entier n est appelé *cardinal* de l'ensemble Ω.

Remarque 1.3.
— *Puisque les ensembles* $\{1, \ldots, n\}$ *et* $\{1, \ldots, p\}$ *sont en bijection ssi* $n = p$*, le cardinal est bien défini de manière unique.*
— *Si un ensemble* Ω *est de cardinal* n*, la bijection* f *permet d'écrire* $\Omega = \{\omega_1, \ldots, \omega_n\}$ *si l'on pose* $\omega_i := f(i)$*. Le cardinal correspond donc bien à l'intuition qu'on s'en fait : c'est le nombre d'éléments dans l'ensemble.*
— *On adopte la convention que l'ensemble* $\{1, \ldots, n\}$ *pour* $n = 0$ *correspond à l'ensemble vide, et donc l'ensemble vide est de cardinal* 0*.*

En particulier, on a la remarque simple mais cruciale suivante :

Proposition 1.4. *Deux ensembles finis sont de même cardinal ssi ils sont en bijection. Formellement* $\mathrm{Card}(A) = \mathrm{Card}(B)$ *ssi il existe* $f : A \to B$ *bijection.*

Démonstration. Si A et B sont de même cardinal fini, disons n, notant f_A et f_B les bijections associées de A et B vers $\{1, \ldots, n\}$ respectivement, on a que la composée $f_A \circ f_B^{-1}$ est une bijection de A sur B. Réciproquement, si A est fini, de cardinal n disons, notant f_A la bijection associée de A vers $\{1, \ldots, n\}$, et g la bijection de A vers B, la composée $f_A \circ g^{-1}$ donne une bijection de B vers $\{1, \ldots, n\}$ donc A et B sont de même cardinal. □

Il en découle le point méthodologique suivant important, à la base du dénombrement et de la combinatoire en général : il suffit pour dénombrer un ensemble de cardinal inconnu de le mettre en bijection avec un ensemble mieux compris.

1.1.4 Fonction indicatrice

Définition 1.5. Soit Ω un ensemble, et $A \in \mathcal{P}(\Omega)$. Alors la fonction indicatrice de A (relativement à Omega) est la fonction

$$\mathbb{1}_A : \Omega \to \{0, 1\}, x \mapsto \mathbb{1}_A(x) = \begin{cases} 0 \text{ si } x \notin A \\ 1 \text{ si } x \in A. \end{cases}$$

Cette fonction pernet d'éviter de faire en toutes lettres des disjonctions de cas ; l'avantage est qu'on faire du calcul avec ces fonctions, les ajouter, les soustraire, les multiplier ... ; cela permet donc de systématiser des disjonctions de cas pénibles, comme nous allons essayer de le montrer.

D'abord, le plus naturel est d'interpréter les opérations de réunion et d'intersection en terme de max et de mins :

Lemme 1.6. *On a :*

$$\mathbb{1}_{A\cup B} = \max\{\mathbb{1}_A, \mathbb{1}_B\}, \tag{1.1}$$
$$\mathbb{1}_{A\cap B} = \min\{\mathbb{1}_A, \mathbb{1}_B\}. \tag{1.2}$$

Remarque : ces égalités sont fonctionnelles, la première signifie par exemple "pour tout $x \in \Omega$, $\mathbb{1}_{A\cup B}(x) = \max\{\mathbb{1}_A(x), \mathbb{1}_B(x)\}$".

Démonstration. On note que, pour tout $x \in \Omega$,

$$\begin{array}{ll} & \max\{\mathbb{1}_A(x), \mathbb{1}_B(x)\} = 1 \\ \text{ssi} & (\mathbb{1}_A(x) = 1 \text{ ou } \mathbb{1}_B(x) = 1) \\ \text{ssi} & x \in A \text{ ou } x \in B \\ \text{ssi} & x \in A \cup B \\ \text{ssi} & \mathbb{1}_{A\cup B}(x) = 1. \end{array}$$

□

Démonstration alternative. On peut aussi considérer une partition adaptée au problème. Par exemple, ici :

$$\Omega = (A \cup A^c) \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B) \cup (A^c \cap B^c)$$

Maintenant, on compare les deux membres de
- sur $A \cap B$, les deux membres de (**??**) valent 1 et $\max\{1, 1\} = 1$.
- sur $A \cap B^c$, les deux membres valent 1 et $\max\{1, 0\} = 1$.
- sur $A^c \cap B$, les deux membres valent 1 et $\max\{0, 1\} = 1$.
- sur $A^c \cap B^c$, les deux membres valent 0 et $\max\{0, 0\} = 0$.

□

On laisse au lecteur le soin de prouver (**??**) en utilisant les deux mêmes méthodes.

Lemme 1.7. *Si A et B sont disjoints,*

$$\mathbb{1}_{A\cup B} = \mathbb{1}_A(x) + \mathbb{1}_B(x)$$

En particulier,

$$1 = \mathbb{1}_A(x) + \mathbb{1}_{A^c}(x)$$

Démonstration. On peut déduire ces résultats du lemme précédent en utilisant que pour $x, y \in \{0, 1\}$ tels que $xy = 0$,

$$\max\{x, y\} = x + y.$$

Le cas particulier suivant correspond au choix de $B = A^c = \Omega \setminus A$. □

Néanmoins, on préfère écrire ces formules sous forme de produit et somme de $\mathbb{1}_A$ et $\mathbb{1}_B$, car sommes et produits sont des opérations plus pratiques que max et min dans les calculs. Les deux formules à retenir sont donc les suivantes :

Proposition 1.8. *On a :*

$$\mathbb{1}_{A\cap B} = \mathbb{1}_A \cdot \mathbb{1}_B, \tag{1.3}$$
$$\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A \cdot \mathbb{1}_B = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A\cap B} \tag{1.4}$$

Attention : max et min de fonctions ne s'écrivent pas toujours aussi simplement, on utilise ici la particularité des fonctions indicatrices d'être à valeurs dans $\{0, 1\}$. L'intérêt de la formule (**??**) est qu'elle connecte la réunion et l'intersection et admet une traduction simple en terme de cardinal (paragraphe suivant).

Démonstration. On peut déduire ces résultats du lemme précédent en utilisant que pour $x, y \in \{0, 1\}$,

$$\min\{x, y\} = xy$$
$$\max\{x, y\} = x + y - xy$$

Le vérifier simplement sur les 4 couples $(0, 0)$, $(0, 1)$, $(1, 0)$, et $(1, 1)$. □

Démonstration alternative. Néanmoins, on peut aussi procéder de façon plus algébrique, en passant au complémentaire, en utilisant les lois de Morgan et le résultat tout juste démontré (**??**) :

$$\mathbb{1}_{A\cup B} = 1 - \mathbb{1}_{(A\cup B)^c} = 1 - \mathbb{1}_{A^c\cap B^c} = 1 - \mathbb{1}_{A^c} \cdot \mathbb{1}_{B^c} = 1 - (1 - \mathbb{1}_A)(1 - \mathbb{1}_B) = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A\cap B},$$

ce qui fournit une preuve alternative de (**??**). □

Remarque 1.9. *Plus généralement, on va être capable d'exprimer l'indicatrice de la réunion d'ensembles comme une combinaison linéaire, à coefficients dans $\{-1, +1\}$, des indicatrices portant sur des intersections (arbitraires) de ces mêmes ensembles. On va ici se contenter d'écrire la formule sous forme factorisée : si $A_1, \ldots, A_n$ sont des sous-ensembles de Ω, alors :*

$$\mathbb{1}_{\cup A_i} = 1 - \mathbb{1}_{(\cup A_i)^c} = 1 - \mathbb{1}_{\cap(A_i^c)} = 1 - \prod_i \mathbb{1}_{A_i^c} = 1 - \prod_i (1 - \mathbb{1}_{A_i}).$$

Cette généralisation (ou plutôt sa forme développée, que nous avons ici omise) porte le nom de formule du crible de Poincaré*, elle est un peu compliquée à écrire formellement, elle permet d'exprimer l'indicatrice d'une réunion en fonction de l'indicatrice d'une intersection.*

Corollaire 1.10. *On a :*

$$\mathbb{1}_{A\cup B} \leq \mathbb{1}_A + \mathbb{1}_B$$

avec égalité ssi A et B sont disjoints, ie $A \cap B = \emptyset$.

Démonstration. L'inégalité découle de (**??**) en notant que le terme supplémentaire $-\mathbb{1}_{A\cap B} \leq 0$; ensuite, on a égalite ssi $\mathbb{1}_{A\cap B}$ est nulle (c'est-à-dire la fonction identiquement nulle sur Ω), soit $A \cap B = \emptyset$. □

Rappelons que le produit cartésien $A \times B := \{(x, y) \in \Omega^2 : x \in A \text{ et } y \in B\}$ de A et de B est constitué des couples dont la première coordonnée est dans A et la seconde dans B.

Proposition 1.11. *On a :*

$$\mathbb{1}_{A\times B}(x, y) = \mathbb{1}_A(x)\mathbb{1}_B(y).$$

Démonstration. $\mathbb{1}_{A\times B}(x, y)$ vaut 1 ssi $x \in A$ et $y \in B$ ssi $\mathbb{1}_A(x) = 1$ et $\mathbb{1}_B(y) = 1$ ssi $\mathbb{1}_A(x)\mathbb{1}_B(y) = 1$ □

1.1.5 Propriétés du cardinal

On considère dans toute la suite du chapitre un ensemble Ω **fini**. On va considérer deux types de démonstrations : avec les indicatrices (les résultats sont alors des corollaires des propositions démontrées dans la sous-section précédent), en utilisant que pour Ω fini,

$$\mathrm{Card}(A) = \sum_{x\in\Omega} \mathbb{1}_A(x).$$

Proposition 1.12. *Soit A, B deux parties de Ω disjointes, c'est-à-dire qui vérifient $A \cap B = \emptyset$:*

$$\mathrm{Card}(A \cup B) = \mathrm{Card}(A) + \mathrm{Card}(B). \tag{1.5}$$

Démonstration. Notons n_A et n_B les deux quantités $\mathrm{Card}(A)$ et $\mathrm{Card}(B)$. On dipose de deux bijections $f_A : \{1, \dots, n_A\} \to A$, et $f_B : \{1, \dots, n_B\} \to B$. Alors $f : \{1, \dots, n_A + n_B\} \to A \cup B$ définie par $f(i) = f_A(i)$ si $1 \le i \le n_A$ et $f(n_A + i) = f_B(i)$ si $1 \le i \le n_B$ est encore une bijection. C'est clairement une surjection. De plus deux éléments distincts de $\{1, \dots, n_A + n_B\}$ ont des images distinctes, puisque A et B sont disjoints. □

Avec les indicatrices. On somme sur les $x \in \Omega$ la relation

$$\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B.$$

□

Plus généralement, par récurrence finie, si $A_1, \dots, A_p \subset \Omega$ sont des ensembles deux à deux disjoints, on a la propriété d'additivité finie :

$$\mathrm{Card}(\bigcup_{i=1}^{p} A_i) = \sum_{i=1}^{p} \mathrm{Card}(A_i).$$

En particulier, si les $(A_i)_{1\le i\le n}$ réalisent une partition de Ω, puisque $\Omega = \bigcup_{i=1}^{p} A_i$, on obtient que

$$\mathrm{Card}(\Omega) = \sum_{i=1}^{p} \mathrm{Card}(A_i).$$

Corollaire 1.13. *Soit A, B deux parties de Ω arbitraires,*

$$\mathrm{Card}(A \cup B) = \mathrm{Card}(A) + \mathrm{Card}(B) - \mathrm{Card}(A \cap B). \tag{1.6}$$

En particulier, on a la propriété de sous-additivité :

$$\mathrm{Card}(A \cup B) \le \mathrm{Card}(A) + \mathrm{Card}(B).$$

De même par réccurence finie, on peut étendre la propriété d'additivité en propriété de sous-additivité :

$$\mathrm{Card}(\bigcup_{1\leq i\leq p} A_i) \leq \sum_{i=1}^{p} \mathrm{Card}(A_i).$$

Démonstration. On rappelle que $A\cup B = (A\cap B^c)\cup(A\cap B)\cup(A^c\cap B)$ et les trois ensembles de droite sont disjoints, donc, de **??** :

$$\mathrm{Card}(A\cup B) = \mathrm{Card}(A\cap B^c) + \mathrm{Card}(A\cap B^c) + \mathrm{Card}(A^c\cap B)$$

Ensuite on a les deux décompositions : $A = A\cap\Omega = A\cap(B\cup B^c) = (A\cap B)\cup(A\cap B^c)$ et $B = (B\cap A)\cup(B\cap A^c)$ qui impliquent :

$$\mathrm{Card}(A) = \mathrm{Card}(A\cap B^c) + \mathrm{Card}(A\cap B)$$

$$\mathrm{Card}(B) = \mathrm{Card}(A^c\cap B) + \mathrm{Card}(A\cap B)$$

Finalement :

$$\begin{aligned}\mathrm{Card}(A\cup B) &= (\mathrm{Card}(A\cap B^c) + \mathrm{Card}(A\cap B)) + (\mathrm{Card}(A^c\cap B) + \mathrm{Card}(A\cap B)) - \mathrm{Card}(A\cap B)\\ &= \mathrm{Card}(A) + \mathrm{Card}(B) - \mathrm{Card}(A\cap B)\end{aligned}$$

□

Démonstration alternative avec les indicatrices. On somme sur les $x\in\Omega$ la relation (**??**). □

On peut considérer ensuite le cas de trois ensembles :

Corollaire 1.14. *Soit A, B, C trois parties de Ω,*

$$\begin{aligned}\mathrm{Card}(A\cup B\cup C) = &\ \mathrm{Card}(A) + \mathrm{Card}(B) + \mathrm{Card}(C)\\ &- \mathrm{Card}(A\cap B) - \mathrm{Card}(A\cap C) - \mathrm{Card}(B\cap C)\\ &+ \mathrm{Card}(A\cap B\cap C).\end{aligned}$$

Démonstration. Bien sûr, il est possible de refaire la preuve en utilisant une partition adaptée. Nous donnons seulement le début de la preuve : on développe par distributivité :

$$\Omega = (A\cup A^c)\cap(B\cup B^c)\cap(C\cup C^c) = (A^c\cap B^c\cap C^c)\cup(8\text{ termes})$$

Puis, de la loi de Morgan,

$$\Omega = (A\cup B\cup C) = (A\cup B\cup C)^c\cup(A\cup B\cup C) = (A^c\cap B^c\cup C^c)\cup(A\cup B\cup C)$$

ce qui permet d'identifier les 8 termes comme une partition de $A\cup B\cup C$. On en déduit une expression de $\mathrm{Card}(A\cup B\cup C)$ comme une somme de 8 termes. De même on exprime A, B et C à l'aide des mêmes ensembles en utilisant la distributivité :

$$A = A\cap(B\cup B^c)\cap(C\cap C^c) = (4\text{ termes})$$

et aussi les ensembles $A\cap B$, $A\cap C$ et $B\cap C$ comme suit :

$$A\cap B = (A\cap B)\cap(C\cup C^c) = (2\text{ termes}).$$

□

Démonstration avec les indicatrices. On comprend néanmoins qu'il devient bien plus simple et systématique d'utiliser les indicatrices. D'une part

$$\begin{aligned}
\mathbb{1}_{A\cup B\cup C}(x) &= 1 - \mathbb{1}_{(A\cup B\cup C)^c}(x) \\
&= 1 - \mathbb{1}_{A^c\cap B^c\cap C^c}(x) \\
&= 1 - \mathbb{1}_{A^c}\mathbb{1}_{B^c}\mathbb{1}_{C^c}(x) \\
&= 1 - (1-\mathbb{1}_A(x))(1-\mathbb{1}_B(x))(1-\mathbb{1}_C(x)) \\
&= \mathbb{1}_A(x) + \mathbb{1}_B(x) + \mathbb{1}_C(x) \\
&- \mathbb{1}_A\mathbb{1}_B(x) - \mathbb{1}_A(x)\mathbb{1}_C(x) - \mathbb{1}_B(x)\mathbb{1}_C(x) \\
&+ \mathbb{1}_A(x)\mathbb{1}_B(x)\mathbb{1}_C(x) \\
&= \mathbb{1}_A(x) + \mathbb{1}_B(x) + \mathbb{1}_C(x) \\
&- \mathbb{1}_{A\cap B}(x) - \mathbb{1}_{A\cap C}(x) - \mathbb{1}_{B\cap C}(x) \\
&+ \mathbb{1}_{A\cap B\cap C}(x)
\end{aligned}$$

d'où le résultat en sommant sur $x \in \Omega$ cette relation. □

Corollaire 1.15. *Soit A, B deux parties de Ω. Alors le produit cartésien $A \times B$ satisfait :*

$$\mathrm{Card}(A \times B) = \mathrm{Card}(A)\,\mathrm{Card}(B)$$

En particulier, si l'on note A^p le produit cartésien de p copies de A avec lui-même :

$$\mathrm{Card}(A^p) = \mathrm{Card}(A)^p$$

Démonstration. L'idée est toujours de se ramener à une collection d'ensembles disjoints. Dans le cas du produit cartésien, $A \times B$ peut s'écrire comme la réunion disjointe suivante : $A \times B = \bigcup_{x\in A}(\{x\} \times B)$, donc

$$\begin{aligned}
\mathrm{Card}(A \times B) &= \mathrm{Card}\Big(\bigcup_{x\in A} \{x\} \times B\Big) \\
&= \sum_{x\in A} \mathrm{Card}(\{x\} \times B) \\
&= \sum_{x\in A} \mathrm{Card}(B) \\
&= \mathrm{Card}(A) \cdot \mathrm{Card}(B)
\end{aligned}$$

□

Démonstration à l'aide d'indicatrices. Il est encore possible d'écrire la preuve à l'aide d'indicatrices ; toutefois, comme nous le montrons ci-dessous, cela ne simplifie pas beaucoup la

preuve :

$$\begin{aligned}\mathrm{Card}(A\times B) &= \sum_{(x,y)\in\Omega^2} \mathbb{1}_{A\times B}(x,y)\\ &= \sum_{(x,y)\in\Omega^2} \mathbb{1}_A(x)\cdot\mathbb{1}_B(y)\\ &= \sum_{x\in\Omega}\sum_{y\in\Omega} \mathbb{1}_A(x)\cdot\mathbb{1}_B(y)\\ &= \sum_{x\in\Omega} \mathbb{1}_A(x)(\sum_{y\in\Omega}\mathbb{1}_B(y))\\ &= (\sum_{y\in\Omega}\mathbb{1}_B(y))\cdot(\sum_{x\in\Omega}\mathbb{1}_A(x))\\ &= \mathrm{Card}(A)\cdot\mathrm{Card}(B)\end{aligned}$$

Cette preuve utilise implicitement la décomposition du produit cartésien $\Omega\times\Omega = \bigcup_x \{x\}\times\Omega$ qui constitue essentiellement la preuve précédente - avec $\Omega\times\Omega$ remplacé par $A\times B$. □

On termine par un lemme très utile. Comme dit précédemment, on s'en remet souvent à des bijections explicites pour dénombrer des ensembles : une bijection est une application telle que l'image réciproque de tout élément est un singleton. D'autres types de fonctions peuvent aussi nous être utiles.

Lemme 1.16 (Lemme des bergers)**.** *Si $f : E \to F$ est une fonction telle que pour tout $y \in F$,* $\mathrm{Card}(f^{-1}(\{y\}))$ *soit constant égal à n, alors*

$$\mathrm{Card}\, E = n\,\mathrm{Card}\, F.$$

Le cas $n = 0$ est impossible si E est non vide (pourquoi ?), le cas $n = 1$ correspond au cas d'une fonction f bijective ; enfin une fonction f qui satisfait à l'hypothèse est dite "n-to-1" dans la terminologie anglo-saxonne. L'appellation lemme des bergers fait référence à la situation suivante : l'application qui à une patte de mouton associe son mouton est une application "4-to-1" donc on peut compter les moutons en comptant les pattes puis en divisant le résultat par 4.

Démonstration. Tout d'abord sans indicatrice : les ensembles $(f^{-1}(\{y\}), y \in F)$ forment une partition de E et donc :

$$\mathrm{Card}\, E = \sum_{y\in F} \mathrm{Card}\, f^{-1}(\{y\}) = \sum_{y\in F} n = n\,\mathrm{Card}\, F$$

□

Démonstration avec des indicatrices. On peut aussi faire avec des indicatrices : ça n'est (vraiment) pas plus simple en l'occurence, mais le but est de s'habituer à ces manipulations. Le point clef est que pour tout x, par définition d'une application :

$$1 = \sum_{y\in F} \mathbb{1}_{y=f(x)}$$

De cette relation découle :

$$\operatorname{Card} E = \sum_{x\in E} 1 = \sum_{x\in E}(\sum_{y\in F} \mathbb{1}_{y=f(x)}) = \sum_{y\in F}\sum_{x} \mathbb{1}_{y=f(x)} = \sum_{y\in F}\sum_{x} \mathbb{1}_{f^{-1}(\{y\})}(x) = \sum_{y\in F} \operatorname{Card} f^{-1}(\{y\})$$

et puisque $\operatorname{Card} f^{-1}(\{y\})$ est constant égal à n, on obtient bien $\operatorname{Card} E = n \operatorname{Card} F$. □

1.1.6 Ensembles de référence

Dans cette section, on détaille quelques ensembles de référence en combinatoire, qu'on associe ensuite aux tirages de boules distinguables avec ou sans remise, en tenant compte ou non de l'ordre dans lequel les boules sont tirées (ce qui donnera lieu a $2 \times 2 = 4$ types de tirage).

p-uplets d'éléments de $\{1, \dots, n\}$

Pour $0 \leq p \leq n$, on appelle p-uplet une suite (ordonnée) de p éléments de $\{1, \dots, n\}$. C'est encore le produit cartésien de l'ensemble $\{1, \dots, n\}$ avec lui-même, p fois :

$$\{1, \dots, n\}^p = \{(x_1, \dots, x_p) : x_1, \dots, x_p \in \{1, \dots, n\}\}.$$

Son cardinal vaut donc :

$$\operatorname{Card}(\{1, \dots, n\}^p) = \operatorname{Card}(\{1, \dots, n\})^p = n^p.$$

C'est aussi le cardinal de l'ensemble $\mathcal{F}(\{1, \dots, p\}, \{1, \dots, n\})$ des fonctions de $\{1, \dots, p\}$ dans $\{1, \dots, n\}$, en interprétant la i-ème coordonnée du p-uplet comme une l'image de i par une fonction. Formellement, l'application

$$\varphi : \begin{array}{ll} \{1, \dots, n\}^p & \to \mathcal{F}(\{1, \dots, p\}, \{1, \dots, n\}), \\ x & \mapsto f : \{1, \dots, p\} \to \{1, \dots, n\}, i \mapsto x_i \end{array}$$

est une bijection.

p-uplets d'éléments distincts de $\{1, \dots, n\}$ (arrangements)

Pour $0 \leq p \leq n$, l'ensemble $\mathcal{A}_n^p$ des arrangements de p éléments de $\{1, \dots, n\}$ est le sous-ensemble des p-uplets de $\{1, \dots, n\}$ dont les p coordonnées sont deux à deux distinctes.

$$\mathcal{A}_n^p := \{(x_1, \dots, x_p) \in \{1, \dots, n\}^p : i \neq j \Rightarrow x_i \neq x_j\}.$$

Rappelons que pour $n \in \mathbb{N}$, $n!$ (lire : factorielle n) est l'entier égal au produit des entiers de 1 à n, soit $n! = \prod_{i=1}^{n} i$, avec la convention que le produit sur l'ensemble vide vaut $1 : 0! = 1$.

Proposition 1.17. *Le cardinal A_n^p de $\mathcal{A}_n^p$ vaut :*

$$A_n^p = \operatorname{Card} \mathcal{A}_n^p = n \times (n-1) \times \dots (n-p+1) = \frac{n!}{(n-p)!}$$

(Attention, dans cette notation A_n^p, p est un indice supérieur, pas une puissance !).

Démonstration intuitive. L'explication intuitive est la suivante : on a n choix pour le premier élément, puis, cet élément étant interdit, $n-1$ choix seulement pour le second élément, $n-2$ pour le troisième, et ainsi de suite jusqu'au p-ième élément pour lequel il reste $n-(p-1) = n-p+1$ choix, puisque $p-1$ éléments ont déjà été choisis. □

Démonstration. Une preuve plus formelle de la formule pour A_n^p est basée sur la relation de récurrence suivante :

$$A_n^p = nA_{n-1}^{p-1},$$

qui elle-même découle de l'observation suivante : l'application

$$\begin{array}{ll} \mathcal{A}_n^p & \to \{1,\dots,n\} \times \mathcal{A}_{n-1}^{p-1} \\ (x_1, x_2, \dots, x_n) & \to (x_1, (y_2, \dots, y_n)), \text{ avec } y_i = x_{i-\mathbb{1}_{x_1 \leq x_i}} \end{array}$$

est une bijection. Exhibons la fonction inverse pour s'en convaincre. Si x_1 et $(y_2, \dots, y_n)$ sont donnés : on pose pour $2 \leq i \leq n$, $x_i = y_i + \mathbb{1}_{y_i \geq x_1}$. De la relation de récurrence et de l'initialisation $A_n^1 = n$ on tire :

$$A_n^p = nA_{n-1}^{p-1} = n(n-1)A_{n-2}^{p-2} = n(n-1)\dots(n-p+2)A_{n-p+1}^1 = n(n-1)\dots(n-p+1)$$

□

C'est aussi le cardinal de l'ensemble des fonctions injectives de $\{1,\dots,p\}$ dans $\{1,\dots,n\}$, puisque l'application φ précédente de domaine d'arrivée restreint à ces injections définit encore une bijection.

Permutations de $\{1,\dots,n\}$

C'est l'ensemble des bijections de $\{1,\dots,n\}$, traditionnellement noté :

$$\Sigma_n = \{\sigma : \{1,\dots,n\} \to \{1,\dots,n\}, \sigma \text{ bijection}\}$$

Bien sûr, on peut encore voir $(\sigma(1),\dots,\sigma(n))$ comme un n-uplet de n éléments distincts de $\{1,\dots,n\}$, c'est-à-dire que Σ_n est en bijection avec $\mathcal{A}_n^n$, et donc

Corollaire 1.18. *L'ensemble Σ_n des bijections de l'ensemble $\{1,\dots,n\}$a pour cardinal $n!$.*

C'est le nombre de façons distinctes d'ordonner n éléments : si on a n objets à ordonner, que l'on numérote de 1 à n, alors $\sigma(1)$ est le premier objet, suivi de $\sigma(2)$, $\sigma(3)\dots$

Exemple 1.19. De combien de façons peut on disposer 6 livres distincts sur une étagère ?

Parties de $\{1,\dots,n\}$

L'ensemble des parties, ou sous-ensembles, de $\{1,\dots,n\}$, est traditionnellement noté :

$$\mathcal{P}(\{1,\dots,n\}) = \{A : A \subset \{1,\dots,n\}\}$$

L'application

$$\varphi : \begin{array}{ll} \mathcal{P}(\{1,\dots,n\}) & \to \{0,1\}^n \\ A & \mapsto (\mathbb{1}_A(1),\dots,\mathbb{1}_A(n)) \end{array}$$

qui à une partie associe le n-uplet de $\{0,1\}^n$, avec 1 en position $i \in \{1,\dots,n\}$ ssi i appartient à la partie A, est une bijection. Par exemple, avec $n=6$, $\varphi(\{2,4,5\}) = (0,1,0,1,1,0)$. Comme $\text{Card}(\{0,1\}^n) = \text{Card}(\{0,1\})^n = 2^n$, on a donc :

Proposition 1.20. *L'ensemble des parties de $\{1, \ldots, n\}$ a pour cardinal :*

$$\operatorname{Card} \mathcal{P}(\{1, \ldots, n\}) = 2^n.$$

Notons que l'ensemble vide, qui correspond au choix du n-uplet $(0, \ldots, 0)$, est une partie de $\{1, \ldots, n\}$. Les parties non-vide sont donc en nombre $2^n - 1$.

Si l'on distingue les parties de $\{1, \ldots, n\}$ selon leur nombre d'éléments, on obtient une partition de l'ensemble $\mathcal{P}(\{1, \ldots, n\})$ en les parties à p éléments, lorsque p parcourt $\{0, \ldots, n\}$, et on peut essayer de comprendre comment dénombrer les éléments de cette partition. C'est l'objet du paragraphe suivant.

Parties à p éléments de $\{1, \ldots, n\}$

Définition 1.21. Pour $0 \leq p \leq n$, c'est l'ensemble noté :

$$\mathcal{C}_n^p := \{\{x_1, \ldots, x_p\} \subset \{1, \ldots, n\} | (x_1, \ldots, x_p) \in \mathcal{A}_n^p\}.$$

Il y a une différence clef par rapport aux arrangements cependant : on considère cette fois-ci l'*ensemble* $\{x_1, \ldots, x_p\}$ et non la *liste ordonnée* $(x_1, \ldots, x_p)$. Le point essentiel est que deux arrangements peuvent correspondent à la même partie dès lors que leurs éléments différent par une permutation, par exemple :

$$(4, 8, 2, 3) \neq (8, 2, 4, 3) \text{ mais } \{4, 8, 2, 3\} = \{8, 2, 4, 3\}.$$

Plus généralement, l'application :

$$\begin{array}{ll} \mathcal{A}_n^p & \to \mathcal{C}_n^p \\ (x_1, \ldots, x_p) & \mapsto \{x_1, \ldots, x_p\} \end{array}$$

est "p!-to-1". En effet, on peut décrire explicitement l'ensemble des antécédants du p-uplet strictement croissant : $\{x_1, \ldots, x_p\}$: il s'agit de l'ensemble

$$\{x_\sigma = (x_{\sigma(1)}, \ldots, x_{\sigma(p)}) | \sigma \in \Sigma_p\}$$

indicé par σ l'ensemble des permutations de $\{1, \ldots, p\}$ Ainsi, du lemme des bergers, on conclut que

Proposition 1.22. *Le cardinal de $\mathcal{C}_n^p$ l'ensemble des parties de $\{1, \ldots, n\}$ à p éléments vaut*

$$\operatorname{Card}(\mathcal{C}_n^p) = \frac{A_n^p}{p!} = \frac{n!}{p!(n-p)!} =: \binom{n}{p} \in \mathbb{N}^\star$$

Le quotient par $p!$ dans la formule précédente correspond au nombre d'ordres possibles pour un p-uplet d'éléments distincts.

Noter en particulier que la dernière expression nous dit que $p!(n-p)!$ divise $n!$ ce qui n'est pas a priori évident : que chacun des deux nombres $p!$ et $(n-p)!$ divise $n!$ est clair, que leur produit divise encore $n!$ est moins évident. Cette formule vaut encore pour $p = 0$ ou $p = n$ avec la convention que $0! = 1$.

L'application φ ci-dessus met cet ensemble en bijection avec le sous-ensemble

$$\{j = (j_1, \ldots, j_n) \in \{0, 1\}^n : \sum_{1 \leq k \leq n} j_k = p\}. \tag{1.7}$$

dont la cardinal vaut donc encore $\binom{n}{p}$

Enfin, l'application qui à une partie à p éléments associe le p-uplet de ses éléments classés par ordre croissant est un bijection, de sorte que l'ensemble

$$\{(i_1, \ldots, i_n) \in \{1, \ldots, n\}^p : 1 \leq i_1 < \ldots < i_p \leq n\}. \tag{1.8}$$

est encore de cardinal $\binom{n}{p}$.

Autour du coefficient binomial

Triangle de Pascal Supposons $1 \leq p \leq n$. Observons que l'application

$$A \in \mathcal{C}_n^p \to (A \cap \{1, \ldots, n-1\}, \mathbb{1}_A(n))$$

induit une bijection de $\mathcal{C}_n^p$ sur la réunion (disjointe) des deux ensembles suivants

$$(\mathcal{C}_{n-1}^p \times \{0\}) \bigcup \left(\mathcal{C}_{n-1}^{p-1} \times \{1\}\right) :$$

Le premier ensemble correspond au cas où n n'est pas dans A, et alors la restriction de A à $\{1, \ldots, n-1\}$ comprend p éléments parmi $n-1$; le second ensemble correspond au cas où n est pas dans A, et alors la restriction de A à $\{1, \ldots, n-1\}$ comprend encore $p-1$ éléments parmi $n-1$. Si l'on interprète ceci en terme de cardinaux, on obtient donc la relation dite du triangle de Pascal :

$$\binom{n}{p} = \binom{n-1}{p} + \binom{n-1}{p-1},$$

que l'on peut aussi vérifier directement à l'aide du calcul et qui permet de calculer récursivement les coefficients binomiaux de proche en proche.

Binôme de Newton Puisque l'ensemble des parties de $\{1, \ldots, n\}$ de cardinal p réalise une partition de l'ensemble des parties de $\{1, \ldots, n\}$ lorsque p varie entre 0 et n, nous avons donc prouvé que :

$$\begin{aligned} 2^n &= \mathrm{Card}(\{A \subset \{1, \ldots, n\}\}) \\ &= \sum_{k=0}^{n} \mathrm{Card}(\{A \subset \{1, \ldots, n\} \mid \mathrm{Card}(A) = k\}) \\ &= \sum_{k=0}^{n} \binom{n}{k} \end{aligned} \tag{1.9}$$

Rappelons que, pour un ensemble fini I et des nombres x_i,

$$(\sum_{i \in I} x_i)^2 = \sum_{(i_1, i_2) \in I \times I} x_{i_1} x_{i_2}$$

et plus généralement,

$$(\sum_{i \in I} x_i)^p = \sum_{(i_1, i_2, \ldots, i_p) \in I^p} x_{i_1} x_{i_2} \ldots x_{i_p}$$

avec I^p le produit cartésien de I avec lui-même p fois. Nous montrons maintenant la formule du binôme de Newton. Nous prenons soin de formaliser un maximum sa démonstration. D'abord on note que $a+b$ peut sécrire comme suit :

$$a+b = a^1b^0 + a^0b^1 = \sum_{\epsilon\in\{0,1\}} a^\epsilon b^{1-\epsilon}.$$

Ensuite, on élève cette quantité à la puissance n :

$$\begin{aligned}
(a+b)^n &= (\sum_{\epsilon\in\{0,1\}} a^\epsilon b^{1-\epsilon})^n \\
&= \sum_{(\epsilon_1,\ldots,\epsilon_n)\in\{0,1\}^n} a^{\epsilon_1+\ldots+\epsilon_n} b^{n-(\epsilon_1+\ldots+\epsilon_n)} \\
&= \sum_{(\epsilon_1,\ldots,\epsilon_n)\in\{0,1\}^n} a^{\epsilon_1+\ldots+\epsilon_n} b^{n-(\epsilon_1+\ldots+\epsilon_n)} \cdot (\sum_{j=0}^{n} \mathbb{1}_{\epsilon_1+\ldots+\epsilon_n=j}) \\
&= \sum_{j=0}^{n} \sum_{(\epsilon_1,\ldots,\epsilon_n)\in\{0,1\}^n} a^j b^{n-j} \mathbb{1}_{\epsilon_1+\ldots+\epsilon_n=j} \\
&= \sum_{j=0}^{n} a^j b^{n-j} \sum_{(\epsilon_1,\ldots,\epsilon_n)\in\{0,1\}^n} \mathbb{1}_{\epsilon_1+\ldots+\epsilon_n=j} \\
&= \sum_{j=0}^{n} a^j b^{n-j} \,\mathrm{Card}\{(\epsilon\in\{0,1\}^n : \epsilon_1+\ldots+\epsilon_n = j\}) \\
&= \sum_{j=0}^{n} \binom{n}{j} a^j b^{n-j}
\end{aligned}$$

La formule du binôme de Newton généralise (**??**) qui correspond au cas particulier $a=b=1$.

Symétrie des coefficients binomiaux La bijection $\mathcal{P}(\{1,\ldots,n) \to \mathcal{P}(\{1,\ldots,n), A \mapsto A^c$ envoie les parties à p éléments de $\{1,\ldots,n\}$ sur les parties à $n-p$ éléments de $\{1,\ldots,n\}$. Ceci montre que :

$$\binom{n}{p} = \binom{n}{n-p},$$

ce qu'on peut encore vérifier directement par le calcul puisque la formule $\binom{n}{p} = \frac{n!}{p!(n-p)!}$ est symétrique sous l'effet de l'échange de p et $n-p$ au dénominateur.

Une formule de Van der Monde Soit $n, m \geq 0$, et $0 \leq k \leq n+m$. Alors :

$$\sum_p \binom{n}{p}\binom{m}{k-p} = \binom{n+m}{k}$$

où la somme dans le terme de gauche porte sur $\max\{0, k-m\} \leq p \leq \min\{k,n\}$: c'est une formule de Van der Monde. En effet,

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

tandis que

$$\begin{aligned}(a+b)^n \cdot (a+b)^m &= \left(\sum_{j_1=0}^{n} \binom{n}{j_1} a^{j_1} b^{n-j_1}\right) \cdot \left(\sum_{j_2=0}^{m} \binom{m}{j_2} a^{j_2} b^{n-j_2}\right) \\ &= \sum_{j_1=0}^{n} \sum_{j_2=0}^{m} \binom{n}{j_1}\binom{m}{j_2} a^{j_1+j_2} b^{n+m-(j_1+j_2)} \\ &= \sum_{k=0}^{n+m} \left(\sum_{p} \binom{n}{p}\binom{m}{k-p}\right) a^k b^{n+m-k}\end{aligned}$$

Et on peut ensuite identifier les termes facteurs de $a^k b^{n-k}$ (pourquoi ?)

Enumérer les quatre types de tirage

On dispose de n boules distinguables dans une urne. Combien de tirages différents de p boules existe-t-il :

— si l'on tire avec ou sans remise ?

— si l'on tient ou non compte de l'ordre des boules tirées ?

Cela fait quatre types de tirage distincts. Pour les dénombrer, numérotons les boules de 1 à n (cela correspond à l'hypothèse que les boules sont distinguables). Nous procédons dans l'ordre de difficulté :

— Avec ordre, avec remise : un tirage est un p-uplet d'éléments de $\{1,\dots,n\}$: le premier élément du p-uplet est la première boule tirée, le second la second boule tirée, etc... Il y donc $\mathrm{Card}(\{1,\dots,n\}^p) = n^p$ tels tirages possibles.

— Avec ordre, sans remise : un tirage est un p-uplet d'éléments distincts de $\{1,\dots,n\}$, il y donc $A_n^p = n!/(n-p)!$ tels tirages possibles.

— Sans ordre, sans remise : On part d'un tirage sans remise, c'est-à-dire un arrangement dans $\mathcal{A}_n^p$, et, puisqu'on ne considère pas l'ordre, deux arrangements $(x_1,\dots,x_p)$ et $(y_1,\dots,y_p)$ sont identifiés si il existe une permutation σ telle que $x_i = y_{\sigma(i)}, \quad 1 \le i \le n$. Autrement dit, deux arrangements sont donc identifiés si

$$\{x_i, 1 \le i \le n\} = \{y_i, 1 \le i \le n\}.$$

c'est-à-dire s'ils définissent la même partie à p éléments de $\{1,\dots,n\}$. Or nous avons vu que le nombre de parties à p éléments d'un ensemble à n éléments est

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

comme on l'a déjà vu. On rappelle à toutes fins utiles que cet ensemble est également en bijection avec

$$\{(i_1,\dots,i_p) \in \{1,\dots,n\}^p : 1 \le i_1 < i_2 < \dots < i_p \le n\}$$

et

$$\{(j_1,\dots,j_n) \in \{0,1\}^n : \sum_{k=1}^{n} j_k = p\}.$$

— Sans ordre, avec remise : c'est un ensemble encore non dénombré précédemment. On part d'un tirage avec remise, c'est-à-dire un p-uplet d'éléments de $\{1,\dots,n\}$, et, puisqu'on ne considère pas l' ordre, ǹouveau, deux p-uplets $(x_1,\dots,x_p)$ et $(y_1,\dots,y_p)$ sont identifiés si il existe une permutation σ telle que

$$x_i = y_{\sigma(i)}, \quad 1 \le i \le n.$$

[2] Notons que $(x_1,\dots,x_p)$ et $(y_1,\dots,y_p)$ définissent le même tirage ssi chaque boule apparaît le même nombre de fois, c'est-à-dire si

$$\sum_{i=1}^{p} \mathbb{1}_{\{x_i=k\}} = \sum_{i=1}^{p} \mathbb{1}_{\{y_i=k\}}, \qquad k = 1,\dots,n.$$

Ainsi, donc le nombre de tirages avec remise, sans remise correspond à l'énumération de l'ensemble :

$$\{(j_1,\dots,j_n) \in \mathbb{N}^n : \sum_{k=1}^{n} j_k = p\}$$

(noter que la différence avec (**??**) est cette fois-ci que l'on autorise les coordonnées j_k à prendre leurs valeurs dans $\mathbb{N}$ plutôt que $\{0,1\}$.) (la différence avec (**??**) est cette fois que l'on autorise les répétitions d'éléments.) La bijection entre des deux ensembles est comme suit : $j_k = \sum_\ell \mathbb{1}_{\{i_\ell=k\}}$. Maintenant on peut encoder un tel n-uplet $(j_1,\dots,j_n)$ par un chemin du plan du point de coordonnées $(1,0)$ au point de coordonnées (n,p) qui ne fait que des pas vers le haut ou vers la droite : le nombre de pas vers le haut vaut j_k lorsque l'abscisse vaut k. La longueur de ce chemin vaut $n+p-1$ (p pas vers le haut et $n-1$ vers la droite) et l'on doit choisir la position des p pas de type Nord, soit

$$\binom{n+p-1}{p}.$$

Une autre façon d'obtenir le même résultat est de noter que l'application qui à $(j_1,\dots,j_n) \in \mathbb{N}^n$ associe

$$(\underbrace{0,\dots,0}_{j_1 \text{ fois}},1,\underbrace{0,\dots,0}_{j_2 \text{ fois}},1,0,\dots,1,\underbrace{0,\dots,0}_{j_n \text{ fois}})$$

est une bijection entre les ensembles suivants (noter que le chiffre 1 apparaît $n-1$ fois) :

$$\{(j_1,\dots,j_n) \in \mathbb{N}^n : \sum_{k=1}^{n} j_k = p\} \text{ et } \{(i_1,\dots,i_{n+p-1}) \in \{0,1\}^{n+p-1} : \sum_{k=1}^{n} j_k = p\},$$

(voir TD2 pour la vérification de ce fait) et le cardinal de ce dernier ensemble vaut bien

$$\binom{n+p-1}{p}$$

Noter que deux p-uplets sont identifiés ssi leur réordonnancement par ordre croissant coïncide, c'est-à-dire que l'ensemble des tirages peut aussi être identifé avec :

$$\{(i_1,\dots,i_p) \in \{1,\dots,n\}^p : 1 \le i_1 \le i_2 \le \dots \le i_p \le n\}$$

2. À la différence du cas sans remise, un tirage ne s'identifie pas à une partie à p éléments de $\{1,\dots,n\}$, car il faut compter le nombre d'occurences de chaque élément

À la différence du cas sans remise, noter qu'on a des inégalités larges et pas des inégalités strictes. Un tel élément caractérise en fait un multi-ensemble (*multiset* en anglais), soit un ensemble dans lequel les 'eĺéments peuvent être répétes : ils apparaissent avec une certaine multiplicité, qui est un entier naturel.

1.2 Espaces de probabilités : définitions

1.2.1 L'univers Ω et les événements.

Définition 1.23. L'univers Ω est un ensemble non vide Ω qui rassemble au moins tous les résultats possibles de l'expérience considérée.

Attention : le choix de Ω n'est pas unique.

Exemple 1.24.
— Un tirage de pile ou face : $\Omega = \{0, 1\}$ ou toute paire avec deux symboles distincts de votre choix, $\{P, F\}$ par exemple.
— n tirages de pile ou face : $\Omega = \{0, 1\}^n$.

Dans ce cours, on s'intéressera aux espaces de probabilité finis, puis, dans un deuxième temps dénombrables.

Définition 1.25.
1. Une partie de Ω est appelée un événement, il décrit un ensemble d'états possibles de l'univers. L'ensemble des événements est l'ensemble des parties $\mathcal{P}(\Omega)$.
2. Le singleton $\{\omega\}$ est appelé événement élémentaire, il décrit un état possible de l'univers.
3. Ω est l'événement sûr ou certain, et $\emptyset$ l'événement impossible.

Exemple 1.26. Dans le cas du lancer d'un dé, si $\Omega = \{1, \ldots, 6\}$, l'événement $A = \{\text{le lancer est pair}\}$ s'écrit : $A = \{2, 4, 6\}$.

Définition 1.27. Deux événements A et B sont dits incompatibles si A et B sont des ensembles disjoints : $A \cap B = \emptyset$.

Cela signifie intuitivement que A et B ne peuvent être réalisés simultanément. Par exemple, toujours dans le cas du lancer de dé, les événements $A = \{\text{le lancer est pair}\}$ et $B = \{\text{le lancer est impair}\}$ sont incompatibles.

1.2.2 Mesure de probabilité, et germe de probabilité

La théorie des probabilités vise à mesurer les événements observables à l'issue d'une expérience, i.e. à décrire la chance de les observer concrètement en pratique. Ces probabilités peuvent venir d'une situation combinatoire impliquant un cardinal, mais pas toujours. Le cadre théorique adéquat est de permettre des pondérations arbitraires avec des axiomes minimaux.

Définition 1.28. On appelle mesure de probabilité sur l'univers Ω, toute fonction $\mathbb{P}$ de l'ensemble des événements $\mathcal{P}(\Omega)$ et à valeurs dans $[0, 1]$ vérifiant :

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

2. Pour $A, B \subset \Omega$ t.q. $A \cap B = \emptyset$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ *(additivité)*

Le couple $(\Omega, \mathbb{P})$ s'appelle "espace de probabilité (fini)". La propriété numéro 1 exprime le fait que l'événement certain est de probabilité 1. La propriété numéro 2, dit propriété d'additivité, traduit l'idée que la mesure d'un événement qui se décompose en événements incompatibles est la somme des mesures de probabilité de ces événements. Remarquons, en choisissant $A = \Omega$ et $B = \emptyset$, que $\mathbb{P}(\emptyset) = 0$. (On ne peut pas substituer l'axiome $\mathbb{P}(\Omega) = 1$ par $\mathbb{P}(\emptyset) = 0$ cependant, car on ne peut alors récupérer cette normalisation).

Définition 1.29. Un événement A tel que $\mathbb{P}(A) = 1$ est appelé un événement *presque sûr*, ou *presque certain*.

On notera qu'on peut avoir $A \neq \Omega$ mais A événement presque sûr, dès lors que A^c est de probabilité nulle. On *peut* itérer la relation 2 jusqu'à arriver aux événements élémentaires (ou singletons). Cela justifie de poser la définition suivante dans ce cadre.

Définition 1.30. On appelle germe de probabilité sur Ω une application $p : \Omega \to [0, 1]$ telle que $\sum_{\omega \in \Omega} p(\omega) = 1$.

Le lien entre les deux notions est exprimé dans la Proposition qui suit.

Proposition 1.31. *Il existe une correspondance entre mesures de probabilité sur Ω et germes de probabilité sur Ω :*

- *la restriction d'une mesure de probabilité $\mathbb{P}$ à la classe des événements élémentaires définit un germe $\Omega \to [0, 1], \omega \mapsto \mathbb{P}(\{\omega\})$.*
- *réciproquement, étant donné un germe de probabilité p, il existe une unique mesure de probabilité dont la restriction aux singletons coïncide avec le germe,*

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

Remarque 1.32. *Si l'approche par les germes peut sembler plus simple, c'est l'approche axiomatique des mesures de probabilités donnée en* **??** *qui se généralisera l'an prochain lorsque vous étudierez les probabilités continues.*

Démonstration. La restriction d'une mesure de probabilité $\mathbb{P}$ aux singletons définit un germe, puisque pour tout $\omega \in \Omega$, $\mathbb{P}(\{\omega\}) \geq 0$, tandis que

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(\bigcup_{\omega \in \Omega} \{\omega\}) = \sum_{\omega} \mathbb{P}(\{\omega\})$$

de la propriété d'additivité. Réciproquement, si $P(\{\omega\}) = p(\omega)$ pour un germe p, alors nécessairement $\mathbb{P}(A) = \sum_{\omega \in A} p(\omega)$ par additivité ; $\mathbb{P}$ ainsi définie est bien une probabilité. □

Définition 1.33. Si Ω est fini, on appelle mesure de probabilité uniforme $\mathbb{P}$ la mesure de probabilité définie par :

$$\mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}, \qquad A \subset \Omega$$

elle a pour germe $\omega \mapsto p(\omega) = \frac{1}{\text{Card}(\Omega)}$.

La formule d'additivité pour $\mathbb{P}$ découle alors de la propriété d'additivité du cardinal.
Une remarque élémentaire au sujet des germes est la suivante :

Lemme 1.34. *Deux germes de probabilité $p, q : \Omega \to [0,1]$proportionnels, c'est-à-dire tels qu'il existe $\alpha \in \mathbb{R}$ tel que pour tout $\omega \in \Omega$,*

$$p(\omega) = \alpha \cdot q(\omega)$$

sont égaux (et définissent donc la même mesure de probabilité.

Démonstration. En effet, en sommant sur $\omega \in \Omega$ la relation de définition, on obtient :

$$1 = \sum_{\omega} p(\omega) = \sum_{\omega} \alpha q(\omega) = \alpha \sum_{\omega} q(\omega) = \alpha.$$

□

En particulier,

Corollaire 1.35. *La mesure de probabilité uniforme est l'unique mesure de probabilité de germe constant, c'est-à-dire telle que pour tout $\omega, \omega' \in \Omega$, $p(\omega) = p(\omega')$.*

Démonstration. Seule la réciproque n'est pas évidente. Mais le germe de probabilité associé à la mesure de probabilité uniforme est un germe constant, il suffit alors d'appliquer le lemme **??** □

Aussi, on a des propriétés analogues à celles vérifiées par le cardinal. Les deux premières propriétés concernent des ensembles non nécessairement disjoints.

Proposition 1.36. *Soit $A, B \subset \Omega$. Alors*

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

En particulier, on a :

$$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B) \qquad \textit{(sous-additivité)}$$

Ensuite, si les $(B_i)_{1 \leq i \leq n}$ forment une partition de Ω,

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i) \qquad \textit{(probabilités totales)}$$

Bien entendu, par récurrence finie, ces propriétés restent valables pour un nombre fini d'ensembles. Par exemple,

— si les $(A_i)_{1 \leq i \leq n}$ sont deux à deux disjoints, on a la propriété d'additivité finie :

$$\mathbb{P}(\bigcup_{1 \leq i \leq n} A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i)$$

— et dans le cas général où on ne les suppose pas deux à deux disjoints, la propriété de *sous*-additivité finie :

$$\mathbb{P}(\bigcup_{1 \leq i \leq n} A_i) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)$$

1.2.3 Mesure de probabilité produit

Définition 1.37. Si $(\Omega_1, \mathbb{P}_1)$ et $(\Omega_2, \mathbb{P}_2)$ sont deux espaces de probabilité, de germes de probabilité respectifs p_1 et p_2, la mesure de probabilité produit, notée $\mathbb{P}_1 \times \mathbb{P}_2$, est la mesure de probabilité sur le produit cartésien $\Omega_1 \times \Omega_2$ de germe p donnée par

$$p((\omega_1, \omega_2)) = p_1(\omega_1)p_2(\omega_2), \qquad (\omega_1, \omega_2) \in (\Omega_1 \times \Omega_2).$$

Elle satisfait :

$$(\mathbb{P}_1 \otimes \mathbb{P}_2)\,(A \times B) = \mathbb{P}_1(A)\,\mathbb{P}_2(B), \qquad A \subset \Omega_1, B \subset \Omega_2$$

On appelle les sous-ensembles de $\Omega_1 \times \Omega_2$ de la forme $A \times B$ les pavés ; noter que tous les ensembles de $\Omega_1 \times \Omega_2$ ne sont pas des pavés ; si Ω_1 comporte deux éléments a et b, et Ω_2 comporte deux éléments c et d, alors l'ensemble $\{(a, c), (b, d)\}$ n'est pas un pavé.

Une mesure de probabilité sur l'espace produit est caractérisée par ses valeurs sur les pavés, puisque la classe des pavés comprend les singletons.

Exemple 1.38. Pour le jet d'un dé, on a déjà mentionné qu'un choix naturel de $(\Omega, \mathbb{P})$ était $\{1, \ldots, 6\}$ muni de la probabilité uniforme $\mathbb{P}$; pour le jet de deux dés, il est raisonnable de considérer l'espace produit $(\Omega^2, \mathbb{P} \times \mathbb{P})$; nous verrons dans la section suivante que ceci revient à postuler que les deux dés sont *indépendants*.

Démonstration. p donné ci-dessus est bien un germe car

$$\begin{aligned}
\sum_{(\omega_1,\omega_2)\in\Omega_1\times\Omega_2} p_1(\omega_1)p_2(\omega_2) &= \sum_{\omega_1\in\Omega_1}\sum_{\omega_2\in\Omega_2} p_1(\omega_1)p_2(\omega_2) \\
&= \sum_{\omega_1\in\Omega_1}\Big(p_1(\omega_1)\sum_{\omega_2\in\Omega_2} p_2(\omega_2)\Big) \\
&= (\sum_{\omega_1\in\Omega_1} p_1(\omega_1))(\sum_{\omega_2\in\Omega_2} p_2(\omega_2)) \\
&= 1 \times 1 = 1
\end{aligned}$$

Ensuite :

$$\begin{aligned}
(\mathbb{P}_1 \otimes \mathbb{P}_2)(A \times B) &= \sum_{(\omega_1,\omega_2)\in A\times B} p(\omega_1, \omega_2) \\
&= \sum_{(\omega_1,\omega_2)\in A\times B} p_1(\omega_1)p_2(\omega_2) \\
&= \sum_{\omega_1\in A}\sum_{\omega_2\in B} p_1(\omega_1)p_2(\omega_2) \\
&= (\sum_{\omega_1\in A} p_1(\omega_1))(\sum_{\omega_2\in B} p_2(\omega_2)) \\
&= \mathbb{P}_1(A)\,\mathbb{P}_2(B)
\end{aligned}$$

□

De façon plus générale, étant donnés n espaces de probabilité $(\Omega_1, \mathbb{P}_1), \ldots, (\Omega_n, \mathbb{P}_n)$, on peut définir sur le produit cartésien $\Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$ la mesure produit $\mathbb{P}_1 \otimes \ldots \otimes \mathbb{P}_n$: associée au germe $p((\omega_1, \ldots, \omega_1)) = p_1(\omega_1) \ldots p_n(\omega_n)$. Elle satisfait

$$(\mathbb{P}_1 \otimes \ldots \otimes \mathbb{P}_n)(A_1 \times \ldots \times A_n) = \mathbb{P}_1(A_1) \ldots \mathbb{P}_n(A_n)$$

Dans le cas où $(\Omega_i, \mathbb{P}_i) = (\Omega, \mathbb{P})$ pour $i = 1, \dots, n$, alors on note simplement $\mathbb{P}^{\otimes n}$ la mesure de probabilité produit sur l'espace produit Ω^n.

Aussi, on peut vérifier sans difficulté que

$$\mathbb{P}_1 \otimes \mathbb{P}_2 \otimes \mathbb{P}_3 = (\mathbb{P}_1 \otimes \mathbb{P}_2) \otimes \mathbb{P}_3 = \mathbb{P}_1 \otimes (\mathbb{P}_2 \otimes \mathbb{P}_3)$$

Proposition 1.39. *Si $\mathbb{P}_1$ et $\mathbb{P}_2$ sont les mesure de probabilité uniforme sur les espaces Ω_1 et Ω_2 respectivement, alors la mesure de probabilité produit $\mathbb{P}_1 \otimes \mathbb{P}_2$ sur $\Omega_1 \times \Omega_2$ est la mesure de probabilité uniforme sur $\Omega_1 \times \Omega_2$.*

Démonstration. Il suffit de noter que le germe produit $p((\omega_1, \omega_2)) = \frac{1}{\mathrm{Card}(\Omega_1)} \frac{1}{\mathrm{Card}(\Omega_2)}$ est constant, et d'utiliser le lemme **??** □

1.2.4 Indépendance

Définition 1.40. Soit $(\Omega, \mathbb{P})$ un espace de probabilité. Deux événements A et B sur cet espace sont dits indépendants lorsque

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

Intuitivement, deux événements sont indépendants si la réalisation de l'un ne donne pas d'indication sur la réalisation de l'autre.

Un événement A presque sûr ou de complémentaire presque sûr est indépendant de tout autre évenement B, puisque $\mathbb{P}(A) \in \{0, 1\}$.

Exemple 1.41. Sur $\{1, \dots, 6\}$ muni de la mesure de probabilité uniforme $\mathbb{P}$ (la modélisation usuelle d'un lancer de dé à six faces), les événements

$$A = \{\text{lancer pair}\} = \{2, 4, 6\} \text{ et } B = \{\text{lancer multiple de 3}\} = \{3, 6\}$$

sont indépendants, puisque

$$\mathbb{P}(A \cap B) = \frac{\mathrm{Card}(A \cap B)}{\mathrm{Card}(\Omega)} = \frac{1}{6}$$

tandis que

$$\mathbb{P}(A)\mathbb{P}(B) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)} \cdot \frac{\mathrm{Card}(B)}{\mathrm{Card}(\Omega)} = \frac{3}{6} \cdot \frac{2}{6} = \frac{1}{6}.$$

Exemple 1.42. Plus généralement, si p_1 et p_2 sont deux entiers *premiers entre eux* (=dont le plus grand diviseur commun vaut 1), alors sur $\Omega = \{1, \dots, p_1 p_2\}$ muni de la probabilité uniforme $\mathbb{P}$, les événements,

$$A = \{jp_1, 1 \leq j \leq p_2\} \text{ et } B = \{jp_2, 1 \leq j \leq p_1\}$$

sont indépendants. En effet $A \cap B$ est constitué des entiers multiples de p_1 et de p_2, donc de $p_1 p_2$, et consiste donc en le seul singleton $\{p_1 p_2\}$ (puisqu'on travaille sur Ω). Ainsi,

$$\mathbb{P}(A \cap B) = \frac{\mathrm{Card}(A \cap B)}{\mathrm{Card}(\Omega)} = \frac{1}{p_1 p_2}$$

tandis que

$$\mathbb{P}(A)\mathbb{P}(B) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)} \cdot \frac{\mathrm{Card}(B)}{\mathrm{Card}(\Omega)} = \frac{p_2}{p_1 p_2} \cdot \frac{p_1}{p_1 p_2} = \frac{1}{p_1 p_2}$$

Exercice 1.43. *p étant un entier premier, sur $\Omega = \{1, \ldots, p\}$ muni de la probabilité uniforme $\mathbb{P}$, peut-on définir deux évènements non triviaux de probabilité distinctes de 0 ou de 1 qui ne soient pas indépendants ?*

Exemple 1.44. Sur l'espace produit $(\Omega_1 \times \Omega_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$, pour $A \subset \Omega_1$ et $B \subset \Omega_2$ arbitraires, les évènements $A \times \Omega_2$ et $\Omega_1 \times B$ sont indépendants puisque :

$$\begin{aligned}(\mathbb{P}_1 \otimes \mathbb{P}_2)(A \times \Omega_2 \cap \Omega_1 \times B) &= (\mathbb{P}_1 \otimes \mathbb{P}_2)(A \times B) \\ &= \mathbb{P}_1(A)\, \mathbb{P}_2(B) \\ &= (\mathbb{P}_1 \otimes \mathbb{P}_2)(A \times \Omega_2)\, (\mathbb{P}_1 \otimes \mathbb{P}_2)(\Omega_1 \times B)\end{aligned}$$

Lemme 1.45. *Si deux événements A et B sont indépendants, alors les trois événements :*

1. *A^c et B*
2. *A et B^c*
3. *A^c et B^c*

sont indépendants.

Démonstration. Traitons par exemple le cas de A^c et B. On a $\mathbb{P}(A^c \cap B) + \mathbb{P}(A \cap B) = \mathbb{P}(B)$ donc $\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(B)(1 - \mathbb{P}(A)) = \mathbb{P}(B)(1 - \mathbb{P}(A))$. Et aussi celui de A^c et B^c.

$$\begin{aligned}\mathbb{P}(A^c \cap B^c) &= 1 - \mathbb{P}(A \cup B) \\ &= 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A \cap B) \\ &= 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(B) \\ &= (1 - \mathbb{P}(A))(1 - \mathbb{P}(B)) \\ &= \mathbb{P}(A^c)\mathbb{P}(B^c)\end{aligned}$$

□

Définition 1.46. Soit $(\Omega, \mathbb{P})$ un espace de probabilité. n événements $A_1, A_2, \ldots, A_n$ sur cet espace sont dits *indépendants* si pour tout $1 \leq p \leq n$ et $1 \leq i_1 < i_2 < \ldots < i_p \leq n$, on a :

$$\mathbb{P}(\bigcap_{j=1}^{p} A_{i_j}) = \prod_{1 \leq j \leq p} \mathbb{P}(A_{i_j})$$

En particulier, il ne suffit donc *pas* seulement de vérifier que :

$$\mathbb{P}(\bigcap_{i=1}^{n} A_i) = \prod_{1 \leq i \leq n} \mathbb{P}(A_i),$$

il faut aussi établir cette propriété pour toute sous-famille des n événements. Par exemple, vérifier que trois événements A, B et C sont indépendants requiert de vérifier les quatre identités suivantes :

$$\begin{aligned}\mathbb{P}(A\cap B\cap C)&=\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)\\ \mathbb{P}(A\cap B)&=\mathbb{P}(A)\mathbb{P}(B)\\ \mathbb{P}(A\cap C)&=\mathbb{P}(A)\mathbb{P}(C)\\ \mathbb{P}(B\cap C)&=\mathbb{P}(B)\mathbb{P}(C)\end{aligned}$$

Notons que de la définition, si n événements sont indépendants, toute sous-famille de ces n événements est encore indépendante.

Se pose tout d'abord la question de l'existence de n événements indépendants. Les mesures de probabilité de type produit vont nous aider à fournir des exemples faciles.

Exemple 1.47 (n événements indépendants)**.** On peut adapter comme suit l'exemple donné sur les pavés dans les espaces produit : si $(\Omega_1,\mathbb{P}_1),\ldots,(\Omega_n,\mathbb{P}_n)$ sont n espaces de probabilités, nous affirmons que les événements

$$\begin{aligned}B_1&:=A_1\times\Omega_2\times\ldots\times\Omega_n,\\ B_2&:=\Omega_1\times A_2\times\ldots\times\Omega_n,\\ &\ldots\\ B_n&:=\Omega_1\times\Omega_2\times\ldots\times A_n.\end{aligned}$$

sont indépendants sur l'espace produit $\Omega_1\times\ldots\times\Omega_n$ muni de la probabilité produit $\mathbb{P}_1\otimes\ldots\otimes\mathbb{P}_n$: ceci découle du fait que $\bigcap_{j=1}^{p}B_{i_j}$ est le produit cartésien $\prod C_i$ où C_i vaut A_i si $i\in\{i_1,\ldots,i_p\}$ et Ω_i sinon ; alors par définition de la probabilité produit :

$$\begin{aligned}(\mathbb{P}_1\otimes\ldots\otimes\mathbb{P}_n)(\bigcap_{j=1}^{p}B_{i_j})&=(\mathbb{P}_1\otimes\ldots\otimes\mathbb{P}_n)(\prod C_i)\\ &=\prod_{i=1}^{n}\mathbb{P}_i(C_i)\\ &=\prod_{i=1}^{p}\mathbb{P}_{i_j}(A_{i_j})\\ &=\prod_{i=1}^{p}(\mathbb{P}_1\otimes\ldots\otimes\mathbb{P}_n)(B_{i_j})\end{aligned}$$

Exemple 1.48 (n événements indépendants, bis)**.** Si l'exemple précédent peut sembler abstrait, voici le même exemple un peu déguisé. On considère $n=\prod_{k=1}^{j}p_k$ où les $p_k, k=1\ldots j$, sont des entiers deux à deux premiers entre eux (cela signifie que le plus grand multiple commun de deux de ces nombres est égal à 1). Sur l'espace $\{1,\ldots,n\}$, on considère alors $\mathbb{P}$ la mesure uniforme, et les événements

$$A_j=\{\ell p_j,\ell=1,\ldots,n/p_j\},$$

soit l'ensemble des multiples de p_j dans $\{1,\ldots,n\}$. On a alors

$$\mathbb{P}(A_j)=\frac{\mathrm{Card}(A_j)}{\mathrm{Card}(\{1,\ldots,n\})}=\frac{(n/p_j)}{n}=\frac{1}{p_j}.$$

et plus généralement, $1 \leq i_1 < \ldots < i_r \leq n$ étant donnés, alors un entier est multiple de $p_{i_1}, \ldots, p_{i_r}$ ssi il est multiple de $p := \prod_{j=1}^r p_{i_j}$ (on utilise que les p_{i_j} sont deux à deux premiers entre eux), et donc l'intersection des A_{i_j} est l'ensemble des multiples de p dans $\{1, \ldots, n\}$:

$$\bigcap_{i=1}^{p} A_{i_j} = \{\ell p, \ell = 1, \ldots, n/p\},$$

puis

$$\mathbb{P}(\bigcap_{i=1}^{p} A_{i_j}) = \frac{\mathrm{Card}(\bigcap_{i=1}^p A_{i_j})}{\mathrm{Card}(\{1,\ldots,n\})} = \frac{1}{p} = \frac{1}{\prod_{j=1}^r p_{i_j}} = \prod_{j=1}^{r} \frac{1}{p_{i_j}} = \prod_{j=1}^{r} \mathbb{P}(A_{i_j}).$$

Une notion plus faible est la suivante :

Définition 1.49. Soit $(\Omega, \mathbb{P})$ un espace de probabilité. n événements $A_1, A_2, \ldots, A_n$ sur cet espace sont dits *deux à deux indépendants* si pour tout $1 \leq i < j \leq n$,

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

Cette définition étant posée, existe-t-il trois événements deux à deux indépendants mais qui ne sont pas indépendants ?

Exemple 1.50 (3 événements deux à deux indépendants mais pas indépendants)**.** On pose Ω l'espace produit $\{0,1\}^2$, muni de la mesure de probabilité uniforme. Puis

$$A_1 = \{\omega = (\omega_1, \omega_2) : \omega_1 = 0\} = \{(0,0), (0,1)\}$$

$$A_2 = \{\omega = (\omega_1, \omega_2) : \omega_2 = 0\} = \{(0,0), (1,0)\}$$

$$A_3 = \{\omega = (\omega_1, \omega_2) : \omega_1 = \omega_2\} = \{(0,0), (1,1)\}$$

On a alors

$$\mathbb{P}(A_i) = \frac{\mathrm{Card}(A_i)}{\mathrm{Card}(\Omega)} = \frac{2}{2^2} = \frac{1}{2}$$

tandis que, si $i \neq j$, $A_1 \cap A_2 \cap A_3 = \{(0,0)\}$, et donc

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$$

En conclusion, A_1, A_2 et A_3 sont *deux à deux* indépendants, mais A_1, A_2, A_3 ne sont pas indépendants. Interprétation : on lance deux pièces de monnaie équilibrées indépendantes, si $\{0,1\}$ modélisent pile et face respectivement, alors

$$A_1 = \{\text{1er lancer pile}\}, A_2 = \{\text{2ième lancer pile}\}, A_3 = \{\text{1er lancer} = \text{2ième lancer}\}$$

Exemple 1.51 (3 événements deux à deux indépendants mais pas indépendants)**.** On pose Ω l'espace produit $\{0,1\}^3$, muni de la mesure de probabilité uniforme. Puis

$$A_1 = \{\omega = (\omega_1, \omega_2, \omega_3) : \omega_2 = \omega_3\}$$

$$A_2 = \{\omega = (\omega_1, \omega_2, \omega_3) : \omega_1 = \omega_3\}$$

$$A_3 = \{\omega = (\omega_1, \omega_2, \omega_3) : \omega_1 = \omega_2\}$$

On a alors

$$\mathbb{P}(A_i) = \frac{\mathrm{Card}(A_i)}{\mathrm{Card}(\Omega)} = \frac{2^2}{2^3} = \frac{1}{2}$$

tandis que, si $i \neq j$, $A_1 \cap A_2 \cap A_3 = A_i \cap A_j = \{\omega_1 = \omega_2 = \omega_3\}$, et donc

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_i \cap A_j) = \frac{\mathrm{Card}(A_i \cap A_j)}{\mathrm{Card}(\Omega)} = \frac{2}{2^3} = \frac{1}{4}$$

En conclusion, A_1, A_2 et A_3 sont *deux à deux* indépendants, mais A_1, A_2, A_3 ne sont pas indépendants. Interprétation : on lance trois pièces de monnaie équilibrées indépendantes, si $\{0, 1\}$ modélisent pile et face respectivement, alors

$$A_1 = \{\text{2ième lancer} = \text{3ième lancer}\}$$
$$A_2 = \{\text{1er lancer} = \text{3ième lancer}\}$$
$$A_3 = \{\text{2ième lancer} = \text{3ième lancer}\}.$$

L'indépendance de n événements entraine en cascade celles de nombreux autres événements qui lui sont liés :

Proposition 1.52. *$A_1, \ldots, A_n$ sont indépendants ssi*

$$\mathbb{P}(B_1 \cap \ldots \cap B_n) = \mathbb{P}(B_1) \ldots \mathbb{P}(B_n)$$

où pour tout $i \in \{1, \ldots, n\}$, $B_i \in \{\emptyset, A_i, A_i^c, \Omega\}$.

Démonstration. La condition donnée est plus forte que la définition de l'indépendance de n événements : si $1 \leq i_1 < \ldots < i_p \leq n$ est un p-uplet croissant donné, on pose $B_i = A_i$ si $i \in \{i_1, \ldots, i_p\}$ et $B_i = \Omega$ sinon, puis on calcule comme suit :

$$\mathbb{P}\Big(\bigcap_{1 \leq j \leq p} A_{i_j}\Big) = \mathbb{P}(\bigcap_{1 \leq i \leq n} B_i) = \prod_{i=1}^{n} \mathbb{P}(B_i) = \prod_{i=1}^{p} \mathbb{P}(A_{i_j}).$$

Réciproquement il va nous suffire de montrer que si $C_1, \ldots, C_p$ sont indépendants, $C_1^c, C_2, \ldots, C_p$ le sont aussi (pourquoi cela suffit ?). Mais ceci découle du calcul suivant ; on commence par noter que

$$C_2 \cap \ldots \cap C_p = (C_1 \cap C_2 \cap \ldots \cap C_p) \cup (C_1^c \cap C_2 \cap \ldots \cap C_p)$$

où la réunion est disjointe, ce qui implique :

$$\begin{aligned}
\mathbb{P}(C_1^c \cap \ldots \cap C_p) &= \mathbb{P}(C_2 \cap \ldots \cap C_p) - \mathbb{P}(C_1 \cap \ldots \cap C_p) \\
&= \mathbb{P}(C_2) \ldots \mathbb{P}(C_p) - \mathbb{P}(C_1)\mathbb{P}(C_2) \ldots \mathbb{P}(C_p) \\
&= (1 - \mathbb{P}(C_1))\mathbb{P}(C_2) \ldots \mathbb{P}(C_p) \\
&= \mathbb{P}(C_1^c)\mathbb{P}(C_2) \ldots \mathbb{P}(C_p)
\end{aligned}$$

□

Les compléments suivants sont non exigibles : Un corollaire simple est le suivant : on appelle *tribu engendrée* par $A_1, \ldots, A_n \subset \Omega$ un événement qui est réunion d'événements du type $\cap_{i=1}^n B_i$, où pour tout $i \in \{1, \ldots, n\}$, $B_i \in \{\emptyset, A_i, A_i^c, \Omega\}$ [3]

3. Nous ne détaillons pas dans ce cours la notion de tribu ni de tribu induite ; dans le cadre fini que nous considérons pour le moment, une tribu est simplement une famille de parties stable par réunion et passage au complémentaire (donc aussi par intersection), et contient l'univers Ω entier ; ni la notion de tribu induite qui est la plus petite tribu qui contient une famille de parties donnée.

Corollaire 1.53. *Si $A_1, \dots, A_n$ sont indépendants et I_1 et I_2 sont deux sous-ensembles disjoints de $\{1, \dots, n\}$, alors tout événement dans la tribu engendrée par $(A_i, i \in I_1)$ est indépendant de tout événement dans la tribu engendrée par $(A_i, i \in I_2)$.*

1.2.5 Probabilité conditionnelle

Définition 1.54. Soit $(\Omega, \mathbb{P})$ un espace de probabilité, A et B deux événements avec $\mathbb{P}(A) > 0$. On appelle probabilité conditionnelle de B sachant A la quantité

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} \in [0, 1]\cdot$$

Interprétation : c'est la probabilité de réalisation de l'événement B *sachant que* l'événement A est réalisé. Notons les trois cas particuliers suivants :

- si les deux événements A et B sont indépendants, $\mathbb{P}(B|A) = \mathbb{P}(B)$, c'est-à-dire que l'occurence de l'événement A n'impacte pas sur la probabilité de la réalisation de l'événement B.
- si les deux événements sont incompatibles, soit $A \cap B = \emptyset$, alors $\mathbb{P}(B|A) = 0$: la réalisation de l'événement A empêche celle de l'événement B.
- si B est inclus dans A, $B \subset A$, alors $\mathbb{P}(B|A) \geq \mathbb{P}(B)$, avec inégalité stricte dès lors que $\mathbb{P}(A) < 1$: la réalisation de l'événement A augmente la probabilité de réalisation de l'événement B.

Proposition 1.55.

- *(Passage au complémentaire)*

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

- *(Formule des probabilités totales, 2) Si $(A_i)_i$ partition de Ω, alors*

$$\mathbb{P}(B) = \sum_i \mathbb{P}(B|A_i)P(A_i).$$

Notons que les probabilités conditionnelles définissent en fait une mesure de probabilité.

Proposition 1.56. *Soit $(\Omega, \mathbb{P})$ un espace de probabilité, A un événement avec $\mathbb{P}(A) > 0$. Alors*

$$B \subset \Omega \mapsto \mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$$

définit une mesure de probabilité sur Ω, appelée mesure de probabilité conditionnelle sachant A, de germe :

$$\omega \mapsto \frac{p(\omega)}{\mathbb{P}(A)} \mathbb{1}_A(\omega)$$

On a une notion d'indépendance conditionnelle :

Définition 1.57. Soit $A, B, C \subset \Omega$ tel que $\mathbb{P}(C) > 0$. On dit que deux événements A et B sont indépendants conditionnellement à un événement C lorsque :

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

À nouveau, on peut vérifier que :

— A^c et B sont indépendants conditionnellement à C.
— A et B^c sont indépendants conditionnellement à C.
— A^c et B^c sont indépendants conditionnellement à C.

En revanche, attention,

— on peut avoir A et B indépendants conditionnellement à C, mais pas conditionnellement au complémentaire C^c : sur Ω le carré $\{0,1,2\}^2$ muni de la probabilité uniforme $\mathbb{P}$, considérons les événements

$$A = \{\omega = (\omega_1, \omega_2) \in \Omega : \omega_1 = 2\}$$
$$B = \{\omega = (\omega_1, \omega_2) \in \Omega : \omega_2 = 2\}$$
$$C = \{\omega = (\omega_1, \omega_2) \in \Omega : \omega_1 \geq 1, \omega_2 \geq 1\}$$

(On fera un dessin avec un carré de côté 3). Alors A et B sont indépendants (d'après les arguments généraux développés sur les espaces produits), A et B sont indépendants conditionnellement à C (d'après les arguments généraux développés sur les espaces produits) ; en effet,

$$\mathbb{P}(A \cap B|C) = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(C)} = \frac{\text{Card}(A \cap B \cap C)}{\text{Card}(C)} = \frac{1}{4}$$

tandis que

$$\mathbb{P}(A|C) = \mathbb{P}(B|C) = \frac{\mathbb{P}(B \cap C)}{\mathbb{P}(C)} = \frac{\text{Card}(B \cap C)}{\text{Card}(C)} = \frac{1}{2}$$

En revanche, A et B ne sont pas indépendants conditionnellement à C, puisque :

$$\mathbb{P}(A \cap B|C^c) = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(C)} = \frac{\text{Card}(A \cap B \cap C)}{\text{Card}(C)} = 0$$

tandis que

$$\mathbb{P}(A|C^c) = \mathbb{P}(B|C^c) = \frac{\mathbb{P}(B \cap C^c)}{\mathbb{P}(C^c)} = \frac{\text{Card}(B \cap C^c)}{\text{Card}(C^c)} = \frac{1}{5}$$

Applications des probabilités conditionnelles

Proposition 1.58 (Formule de Bayes)**.** *Soit $A, B \subset \Omega$ deux événements tels que $\mathbb{P}(A), \mathbb{P}(B) > 0$. Alors*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)}$$

L'intérêt de la fornule réside dans ses possibles applications. D'un point de vue théorique, il s'agit en effet d'un résultat élémentaire.

Démonstration. La preuve est élémentaire est consiste à developper le membre de droite. D'abord, les deux dénominateurs sont égaux et valent :

$$\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c),$$

tandis que le numérateur vaut simplement :

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

□

Exemple 1.59 (Paradoxe de la détection des événements rares)**.** On considère qu'une proportion $p = 1/000$ des conducteurs français s conduit sous l'emprise du cannabis. Un test disponible sur le marché présente une probabilité d'erreur de $q = 1/100$ (opn considère les deux probailités d'erreur, la probabilité de donner un faux positif, et la probabilité de donner un faux négatif, égales à ce nombre). Un conducteur est testé positif, quelle est la probabilité qu'il ait pris du cannabis ? Notons $+$ l'événement etre positif, $-$ l'événement être négatif et C l'événement avoir pris du cannabis. D'après les hypothèses, $\mathbb{P}(C|-) = \mathbb{P}(C^c|+) = q$, et $P(C) = p$ Maintenant,

$$\mathbb{P}(C|+) = \frac{\mathbb{P}(+|C)\mathbb{P}(C)}{\mathbb{P}(+|C)\mathbb{P}(C) + \mathbb{P}(+|C^c)\mathbb{P}(C^c)} = \frac{(1-q)p}{(1-q)p + q(1-p)} = 0,090164,$$

soit un peu plus de 9% seulement. Plus gńéralement lorsque p est très petit devant q, ceci est de l'ordre de p/q (ici égal à 10%, la correction est en effet petite). Le problème est que parmi les positifs, il existe beaucoup deplus de faux positifs quand la proba d'erreur du test excède la probabilité de l'événement rare à détecter.

1.3 Extension au cas dénombrable

1.3.1 Ensembles dénombrables

Définition 1.60. Soit Ω un ensemble. On dit qu'un ensemble Ω est *dénombrable* s'il est en bijection avec une partie de $\mathbb{N}$.

Si la partie de $\mathbb{N}$ est finie, l'ensemble Ω est dit *fini*, et on peut choisir la partie de $\mathbb{N}$ sous la forme $\{1, \ldots, n\}$, avec n le cardinal de Ω. Sinon, l'ensemble Ω est dit *infini dénombrable* et on peut choisir comme partie de $\mathbb{N}$ l'ensemble $\mathbb{N}$ entier.

Notons en particulier qu'un ensemble fini est considéré comme dénombrable dans ce cours, alors que d'autres auteurs demandent à ce qu'un ensemble dénombrable soit en bijection avec $\mathbb{N}$ entier, ce qui exclut le cas des ensembles finis.

Une définition équivalente à la nôtre serait de demander qu'il existe une *injection* de Ω dans $\mathbb{N}$. Intuitivement, un ensemble "s'injecte" dans un autre ensemble (c'est-à-dire il existe une injection du premier ensemble vers le second) s'il est plus petit (au sens large). Un ensemble dénombrable est donc plus "petit" que l'ensemble des entiers naturels.

Proposition 1.61. *Si Ω_1 et Ω_2 sont deux ensembles dénombrables, le produit cartésien $\Omega_1 \times \Omega_2$ est encore dénombrable.*

Démonstration. Il suffit de montrer que $\mathbb{N} \times \mathbb{N}$ est dénombrable. Mais ceci découle d'un dessin ou l'on parcourt les diagonales $i + j = 0$, $i + j = 1$, $i + j = 2$ (de tailles finies respectivement égale à $1, 2, 3 \ldots$) successivement dans cet ordre. □

Proposition 1.62. *Si I est dénombrable est $(\Omega_i)_{i\in I}$ est une famille indicée par I un ensemble dénombrable, alors $\cup_{i\in I}\Omega_i$ est encore dénombrable.*

Les ensembles $\mathbb{N}$, $\mathbb{N}^2$, $\mathbb{Z}$ et $\mathbb{Q}$ sont dénombrables. Ni l'ensemble $\{0, 1\}^{\mathbb{N}}$ des suites binaires, ni l'ensemble $\mathbb{R}$ des nombres réels ne sont dénombrables en revanche. Nous proposons de vérifier certains de ces points dans les deux exercices suivants.

Exercice 1.63. *Montrer que $\mathbb{Z}$ et $\mathbb{Q}$ sont dénombrables. (On pourra écrire $\mathbb{Z}$ sous la forme $-\mathbb{N}\cup\mathbb{N}$ et noter que l'application qui à un nombre rationnel $q \in \mathbb{Q}$ associe son écriture sous la forme a/b, avec $a \in \mathbb{Z}$ et $b \in \mathbb{N}^\star$ premiers entre eux définit une injection de $\mathbb{Q}$ dans $\mathbb{Z} \times \mathbb{N}^\star$).*

Exercice 1.64 (Argument diagonal de Cantor)**.** Supposons *que $\{0,1\}^{\mathbb{N}}$ soit dénombrable.*

1. *Pourquoi peut-on écrire $\{0,1\}^{\mathbb{N}} = \{u_i, i \in \mathbb{N}\}$ pour $u_1, u_2, \ldots$ une suite d'éléments de $\{0,1\}^{\mathbb{N}}$?*
2. *Posons pour tout entier $n \in \mathbb{N}$, $v(n) = 1 - u_n(n)$. Observer que pour tout n, $v \neq u_n$ (au sens où les deux suites sont distinctes), et conclure à une absurdité.*

1.3.2 Famille sommable

La maitrise des outils de ce paragraphe est hors programme. Il s'agit de fournir un cadre mathématique précis aux sommes dénombrables qui apparaissent par la suite.

Cas positif

La notion de famille sommable a pour but de formaliser la somme d'un ensemble dont les termes ne sont pas naturellement *ordonnés* comme c'est le cas des termes d'une série.

Définition 1.65. Soit I un ensemble dénombrable, et $(a_i, i \in I)$ une famille de nombres réels positifs indicée par I. La famille est dite sommable si

$$\sup_{I_0 \text{ fini}} \{ \sum_{i\in I_0} a_i\} < \infty,$$

et dans ce cas, le supremum est appelé la somme de la famille, et simplement noté $\sum_{i\in I} a_i$.

Si S désigne la somme d'une famille $(a_i)_{i\in I}$ sommable, cela signifie encore que pour tout $\varepsilon > 0$, il existe un ensemble fini J_ε tel que

$$S - \varepsilon \leq \sum_{i\in J_\varepsilon} a_i \leq S.$$

On insiste sur le fait que cette définition de somme ne présuppose aucun ordre sur l'ensemble d'indice I (en tant qu'ensemble dénombrable, il n'existe pas nécessairement d'ordre naturel sur celui-ci). L'exemple le plus simple de famille sommable est une famille finie.

Soit $(u_n)_{n\in\mathbb{N}}$ une suite à termes positifs. On rappelle qu'une série numérique $\sum_{n\geq 0} u_n$ de terme général (u_n) converge lorsque la suite des sommes partielles $\sum_{k=1}^n u_k$ converge, ce qui signifie encore que cette suite est bornée. La limite est alors (par définition) la somme de la série. Que est le lien avec la notion de famille sommable ?

Proposition 1.66. *La série numérique de terme général $(u_n)_{n\in\mathbb{N}}$ converge ssi la famille $(u_n)_{n\in\mathbb{N}}$ est sommable, et la somme de la série coïncide avec la somme de la famille.*

Démonstration. Notons ℓ la somme de la série, et S la somme de la famille, et incluons la possibilité que ces deux quantités soient infinies pour simplifier la preuve. Soit I_0 ensemble fini, alors I_0 est inclus dans $\{1,\ldots,n\}$ pour $n := \max I_0$ et partant $\sum_{i\in I_0} u_i \leq \sum_{i=1}^n u_i \leq \ell$, d'où $S \leq \ell$. Maintenant, soit $\ell_0 < \ell$ donné, il existe $n_\varepsilon > 0$ tel que $\ell_0 \leq \sum_{i=1}^{n_\varepsilon} u_i \leq S$ (la dernière inégalité car $\{1,\ldots,n_\varepsilon\}$ est un ensemble fini) et donc, puisque ℓ_0 est arbitraire, cela permet d'établir l'inégalité réciproque $\ell \leq S$. □

Corollaire 1.67. *La somme d'une série numérique (qu'elle soit finie ou non) ne dépend pas de l'ordre des termes.*

Les propriétés élémentaires des sommes sont les suivantes.

Proposition 1.68. *Soit I un ensemble dénombrable, et $(a_i)_{i\in I}, (b_i)_{i\in I}$ deux famille de nombres réels positifs indicées par I et λ un réel positif. Alors :*

$$\sum_{i\in I}(\lambda a_i + b_i) = \lambda \sum_{i\in I} a_i + \sum_{i\in I} b_i$$

où l'égalité a lieu dans $[0, +\infty]$.

Proposition 1.69. *Soit I un ensemble dénombrable $(a_i)_{i\in I}$ une famille de nombres réels positifs et $I = I_0 \cup I_1$, la réunion étant disjointe. Alors :*

$$\sum_{i\in I_0\cup I_1} a_i = \sum_{i\in I_0} a_i + \sum_{i\in I_1} a_i$$

où l'égalité a lieu dans $[0, +\infty]$.

Plus généralement, il est possible de faire un nombre *dénombrable* de paquets.

Theorème 1.70 (Théorème de sommation par paquets)**.** *Soit I un ensemble dénombrable $(a_i)_{i\in I}$ une famille de nombres réels positifs, et $(I_j, j \in J)$ une partition de I. Alors*

$$\sum_{i\in I} a_i = \sum_{j\in J}\sum_{i\in I_j} a_i$$

où l'égalité a lieu dans $[0, +\infty]$.

Le théorène de Fubini positif est alors un simple cas particulier. Il dit qu'on peut calculer des sommes doubles en combinant deux sommes simples, et que l'ordre de sommation n'importe pas.

Corollaire 1.71 (Fubini positif, encore appelé Fubini Tonnelli)**.** *Soit $(a_{i,j})_{I\times J}$ une famille de nombres réels positifs indicée par le produit cartésien $I \times J$ de deux ensembles dénombrables I et J. Alors*

$$\sum_{(i,j)\in I\times J} a_{i,j} = \sum_{i\in I}(\sum_{j\in J} a_{i,j}) = \sum_{j\in J}(\sum_{i\in I} a_{i,j})$$

où l'égalité a lieu dans $[0, +\infty]$.

Cas général

On peut enfin définir la somme d'une famille $(a_i)_{i\in I}$ de nombres réels quelconques (pas forcements positifs). On note x_+ et x_- les parties positives et négative de x respectivement, à savoir $x_+ = \max\{x, 0\}$ et $x_- = \max\{-x, 0\}$.

Définition 1.72. Soit I un ensemble arbitraire, et $(a_i, i \in I)$ une famille de nombres réels indicée par I. La famille est dite sommable si les deux familles $((a_i)_+, i \in I)$ et $((a_i)_-, i \in I)$ le sont, et alors on pose

$$\sum_{i\in I} a_i = \sum_{i\in I}(a_i)_+ - \sum_{i\in I}(a_i)_-.$$

On aurait pu aussi formuler la définition de la sommabilité à l'aide de la valeur absolue.

Lemme 1.73. *La famille* $(a_i, i \in I)$ *est sommable ssi la famille de nombres* positifs $(|a_i|, i \in I)$ *soit sommable.*

Les résultats précédents (lien avec la somme d'une série, indépendance de la valeur de la somme quelque soit l'ordre choisi pour les termes, linéarité, sommation par paquets ....) restent tous valables *sous réserve* que les familles soient *sommables*, et la série *absolument convergente*.

Voici peut-être une définition plus pratique/concrète de la sommabilité dans le cas de nombres réels. Essentiellement la proposition dit que toute la "masse" de la somme infinie se trouve dans un ensemble fini.

Proposition 1.74. *Soit* I *un ensemble arbitraire, et* $(a_i, i \in I)$ *une famille de nombres réels indicée par* I*. La famille est dite sommable de somme* $S \in \mathbb{R}$ *si pour tout* $\varepsilon > 0$*, il existe* $J_\varepsilon \subset I$*,* J_ε *fini, tel que pour tout ensemble* $J \supset J_\varepsilon$*,* J *fini, on a :*

$$|S - \sum_{i \in J} a_i| \leq \varepsilon.$$

Démonstration. Si on a sommabilité de $((a_i)_+, i \in I)$ et de $((a_i)_-, i \in I)$, notant respectivement S_+ et S_- ces deux sommes, alors, $\varepsilon > 0$ étant fixé, on peut trouver deux sous-ensembles finis I_+ et I_- de I qui satisfont :

$$S_+ - \varepsilon \leq \sum_{i \in I_+} (a_i)_+ \leq S_+$$

et

$$S_- - \varepsilon \leq \sum_{i \in I_-} (a_i)_- \leq S_-.$$

Alors pour tout ensemble fini J tel que $J \supset I_+ \cup I_-$, on a bien que :

$$-\varepsilon \leq (S_+ - S_-) - \left(\sum_{i \in J} (a_i)_+ - \sum_{i \in J} (a_i)_- \right) \leq \varepsilon;$$

Réciproquement, si pour tout ε, il existe J_ε tel que pour tout I fini,

$$|S - \sum_{i \in J} a_i| \leq \varepsilon$$

alors en incluant dans I les seuls termes correspondants à des a_i positifs, on obtient que : $\sum_{i \in I \setminus J_\varepsilon} (a_i)_+ \leq \varepsilon$, et partant, $\sum_{i \in I} (a_i)_+ \leq \sum_{i \in J_\varepsilon} (a_i)_+ + \varepsilon < \infty$, ce qui montre que la famille $((a_i)_+)_{i \in I}$ est sommable. La famille $((a_i)_-)_{i \in I}$ se traite de façon similaire. □

Mentionnons le théorème de Fubini :

Corollaire 1.75 (Fubini)**.** *Soit* $(a_{i,j})_{I \times J}$ *une famille* sommable *de nombres réels indicée par le produit cartésien* $I \times J$ *de deux ensembles dénombrables* I *et* J*. Alors*

$$\sum_{(i,j) \in I \times J} a_{i,j} = \sum_{i \in I} (\sum_{j \in J} a_{i,j}) = \sum_{j \in J} (\sum_{i \in I} a_{i,j}).$$

En pratique pour vérifier que la famille est sommable, on applique Fubini positif à la famille $(|a_{i,j}|)_{i,j\in I\times J}$.

Noter que la condition de sommabilité est essentielle : par exemple, on peut considérer la famille :

$$a_{i,j} = \mathbb{1}_{i=j} - \mathbb{1}_{i=j+1}, \qquad (i,j) \in \mathbb{N}^2$$

(faire un dessin pour visualiser géométriquement où sont les $+1$ et les -1). Alors pour tout $i \in \mathbb{N}$, $\sum_{j\in\mathbb{N}} a_{i,j} = \sum_{j\in\mathbb{N}}(\mathbb{1}_{i=j} - \mathbb{1}_{i=j+1}) = \mathbb{1}_{i=0}$ et donc

$$\sum_{i\in\mathbb{N}}\sum_{j\in\mathbb{N}} a_{i,j} = \sum_{i\in\mathbb{N}} \mathbb{1}_{i=0} = 1$$

tandis que pour tout $j \in \mathbb{N}$, $\sum_{i\in\mathbb{N}} a_{i,j} = \sum_{i\in\mathbb{N}}(\mathbb{1}_{i=j} - \mathbb{1}_{i=j+1}) = 0$ et donc

$$\sum_{j\in\mathbb{N}}\sum_{i\in\mathbb{N}} a_{i,j} = \sum_{j\in\mathbb{N}} 0 = 0.$$

Le théorème de Fubini ne s'applique pas ici car la famille n'est pas sommable.

1.3.3 Définition d'une probabilité sur un univers dénombrable

On travaille désormais avec Ω un univers dénombrable. La définition de mesure de probabilité sur un tel univers doit être revue, avec la propriété d'additivité remplacée par la propriété d'*additivité dénombrable* encore appelée *sigma-additivité* :

Définition 1.76. On appelle mesure de probabilité sur l'univers Ω, toute fonction $\mathbb{P}$ de l'ensemble des événements $\mathcal{P}(\Omega)$ et à valeurs dans $[0,1]$ vérifiant :

1. $\mathbb{P}(\Omega) = 1$
2. Pour $(A_i)_{i\in\mathbb{N}}$ famille d'ensembles deux à deux disjoints, on a $\mathbb{P}(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} \mathbb{P}(A_i)$.

La proposition **??** qui lie probabilité et germe tient encore. Précisément, si $(p(\omega))_{\omega\in\Omega}$ est une famille sommable indicée par Ω dénombrable, alors

$$A \mapsto \mathbb{P}(A) := \sum_{\omega\in A} p(\omega)$$

définit bien une mesure de probabilité : la preuve de ce fait repose sur le théorème de sommation par paquets. Réciproquement, si $\mathbb{P}$ est une mesure de probabilité, $1 = \mathbb{P}(\Omega) = \mathbb{P}(\cup_{\omega\in\Omega}\{\omega\}) = \sum_{\omega\in\Omega}\mathbb{P}(\{\omega\})$ donc $p(\omega) = \mathbb{P}(\{\omega\})$ définit bien un germe.

Un point clef sur un espace infini dénombrable est qu'il *n'existe plus* de mesure de probabilité uniforme.

Chapitre 2

Variables aléatoires et moments

On travaille dans tout ce chapitre avec $(\Omega, \mathbb{P})$ un espace de probabilité pour lequel Ω est **dénombrable**.

2.1 Définition

Définition 2.1. Une *variable aléatoire* $X : \Omega \to \mathbb{R}$ est une application de Ω dans $\mathbb{R}$.

Nous pouvons comprendre une certaine déception chez le lecteur. Le vocable "variable aléatoire" rend simplement compte du fait que l'espace de probabilité $(\Omega, \mathbb{P})$ est généralement associé à une expérience aléatoire, dont X traduit un certain aspect ; mais du point de vue mathématique, une variable aléatoire n'est donc rien de plus qu'une application définie sur un espace de probabilité.

On aime à utiliser des lettres capitales pour désigner les variables aléatoires, c'est une convention très couramment retenue et donc importante. De la même façon un élément générique de l'univers Ω est souvent noté ω.

Dans le cadre simple dans lequel nous travaillerons cette année (espace de probabilité dénombrable), aucune condition supplémentaires n'est à exiger dans la définition de variable aléatoire (on aura tout le temps de se torturer l'an prochain avec la notion de mesurabilité...), et donc toute application est recevable.

Remarque 2.2. *On peut aussi considérer plusieurs aspects de l'expérience, et l'observation simultanée de ces différents aspects peut alors être consignée dans un vecteur aléatoire $X : \Omega \to \mathbb{R}^d$, avec $d \geq 1$ le nombre de caractéristiques que l'on observe. Pour ne pas alourdir ici les notations, on restera avec une variable aléatoire scalaire (i.e. à valeurs dans $\mathbb{R}$). Nous reviendrons plus tard sur le cas des vecteurs aléatoires, mais nous laissons de côté cet aspect pour le moment.*

Définition 2.3. On appelle *loi de X* la mesure *image* de $\mathbb{P}$ par l'application X.

$$P_X(A) := \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega : X(\omega) \in A\}) = \mathbb{P}(X \in A), \qquad A \subset \mathbb{R}$$

Il s'agit encore d'une mesure de probabilité, qui fait de $(X(\Omega), P_X)$ un espace de probabilité, et dont le germe est

$$p_X(x) := \mathbb{P}(X^{-1}(\{x\})) = \mathbb{P}(X = x), \qquad x \in X(\Omega).$$

Cette définition est en fait aussi une définition-proposition, et son contenu mathématique est prouvé ci-dessous :

Démonstration. Vérifions que P_X définit une mesure de probabilité. D'abord, $P_X(\mathbb{R}) = \mathbb{P}(X^{-1}(\mathbb{R})) = \mathbb{P}(X \in \mathbb{R}) = 1$. Ensuite, si les ensembles $(A_i)_{i\in\mathbb{N}}$ sont deux à deux disjoints, il en est de même des événements $X^{-1}(A_i) \subset \Omega$, et puisque $\mathbb{P}$ est une mesure de probabilité, il suit :

$$P_X(\bigcup_{i\in I} A_i) = \mathbb{P}(X^{-1}(\bigcup_i A_i)) = \mathbb{P}(\bigcup_i X^{-1}(A_i)) = \sum_i \mathbb{P}(X^{-1}(A_i)) = \sum_i \mathbb{P}(X \in A_i).$$

□

Exemple 2.4. Sur $\Omega = \{1, \dots, 6\}^2$ muni de $\mathbb{P}$ la mesure de probabilité uniforme sur cet espace (qui modélise le lancer de deux dés équilibrés indépendants), la variable aléatoire $X : (\omega_1, \omega_2) \in \Omega \to \omega_1 + \omega_2$ a pour loi (ou plutôt pour germe) :

$$p_X(k) = \mathbb{P}(\{(\omega_1, \omega) \in \Omega : \omega_1 + \omega_2 = k\}) = \frac{(k-1) \wedge (13-k)}{36},$$

et l'on notera que la fonction $k \mapsto (k-1) \wedge (13-k)$ se trouve maximisée en la valeur 7, qui est donc la valeur la plus probable pour la somme de deux dés équilibrés.

2.2 Moments

2.2.1 Espérance

Définition 2.5. Soit $X : \Omega \to \mathbb{R}$ variable aléatoire réelle définie sur $(\Omega, \mathbb{P})$, de germe p. On dit que X est intégrable si la famille $(|X(\omega)|p(\omega))_{\omega\in\Omega}$ est sommable et dans ce cas, on appelle espérance de X et on note $\mathbb{E}[X]$ la quantité :

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)p(\omega).$$

Par définition donc, X est intégrable lorsque les deux variables aléatoires *positives* $X_+ = \max\{X, 0\}$ et $X_- = \max\{-X, 0\}$ le sont, et, dans ce cas, on a l'égalité suivante (de définition) dans $\mathbb{R}$:

$$\mathbb{E}[X] = \mathbb{E}[X_+] - \mathbb{E}[X_-].$$

En pratique, on connaît souvent les variables aléatoires par leur loi, et la formule de changement de variables suivante, aussi appelée formule de transfert. est donc la formule que l'on utilisera dans la pratique :

Proposition 2.6 (Formule de transfert)**.** *Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire définie sur $(\Omega, \mathbb{P})$. La variables aléatoire $\varphi(X)$ est intégrable ssi la famille $(|\varphi(x)|p_X(x)_{x\in X(\Omega)})$ est sommable et dans ce cas,*

$$\mathbb{E}[\varphi(X)] = \sum_{x\in X(\Omega)} \varphi(x)p_X(x).$$

En particulier (prendre pour φ la fonction identité), X est intégrable ssi la famille $(|x|p_X(x)_{x\in X(\Omega)})$ est sommable et dans ce cas,

$$\mathbb{E}[X] = \sum_{x\in X(\Omega)} x p_X(x),$$

qui est la formule qu'on utilise en pratique.

Démonstration. On a le calcul suivant, valable dans $[0, +\infty]$,

$$\begin{aligned}\sum_{\omega\in\Omega} |\varphi(X(\omega))|p(\omega) &= \sum_{\omega\in\Omega} |\varphi(X(\omega))|p(\omega)\cdot(\sum_{x\in X(\Omega)} \mathbb{1}_{X(\omega))=x}) \\ &= \sum_{x\in X(\Omega)} |\varphi(x)| \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\})\mathbb{1}_{X(\omega))=x} \\ &= \sum_{x\in X(\Omega)} |\varphi(x)|\mathbb{P}(X=x) \\ &= \sum_{x\in X(\Omega)} |\varphi(x)|p_X(x)\end{aligned}$$

$\varphi(X)$ intégrable signifie $(|\varphi(X(\omega))|p(\omega)_{x\in X(\Omega)})$ sommable, qui est donc équivalent à $(|\varphi(x)|p_X(x))_{x\in X(\Omega)}$ sommable. Reprendre le calcul sans les valeurs absolues permet de conclure à la formule de l'énoncé. □

Notons que

Lemme 2.7 (Positivité de l'espérance)**.** *Soit $X : \Omega \to \mathbb{R}^+$ variable aléatoire réelle positive. Alors*

$$\mathbb{E}[X] \geq 0,$$

et

$$\mathbb{E}[X] = 0 \qquad ssi \qquad X = 0 \quad p.s.$$

Démonstration. Puisque X est à valeurs positives, on a $X(\omega)p(\omega) \geq 0$ quelque soit $\omega \in \Omega$, il suit par définition de la somme d'une famille que :

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)p(\omega) \geq 0$$

puis $\mathbb{E}[X] = 0$ ssi pour tout $\omega \in \Omega$, $X(\omega)p(\omega) = 0$ donc $X(\omega) > 0$ implique $p(\omega) = 0$. Ainsi $\mathbb{E}[X] = 0$ implique

$$\mathbb{P}(X > 0) = \sum_{\omega} \mathbb{1}_{X(\omega)>0}\, p(\omega) = \sum_{\omega} 0 = 0,$$

et partant $\mathbb{P}(X = 0) = \mathbb{P}(X \geq 0) - \mathbb{P}(X > 0) = 1 - 0 = 1$. □

Lemme 2.8 (Linéarité de l'espérance)**.** *Soit $X, Y : \Omega \to \mathbb{R}$ variables aléatoires réelles, et $\lambda \in \mathbb{R}$. Alors $\lambda X + Y$ est intégrable, et*

$$\mathbb{E}[\lambda X + Y] = \lambda\mathbb{E}[X] + \mathbb{E}[Y].$$

Démonstration. C'est un corollaire direct de la définition de l'espérance combinée avec la linéarité de la somme. □

On retiendra donc que pour développer une formule d'espérance par linéarité, il faut s'assurer que tous les termes sont intégrables. Noter que ce résultat ne requiert pas l'indépendance des variables aléatoires sous-jacentes X et Y (non encore définie certes, mais le lecteur a peut-être déjà été confronté à la défintion)

Approche alternative pour la définition de l'espérance

Notons le lemme simple suivant :

Lemme 2.9. *Soit X une variable aléatoire réelle, et $B \subset \mathbb{R}$. Alors*

$$\mathbb{E}[\mathbb{1}_B(X)] = \mathbb{P}(X \in B).$$

Démonstration. Par la définition de l'espérance :

$$\mathbb{E}[\mathbb{1}_B(X)] = \sum_{\omega\in\Omega} \mathbb{1}_B(X(\omega))p(\omega) = \sum_{\omega\in\Omega} \mathbb{1}_{\{X(\omega)\in B\}}p(\omega) = \mathbb{P}(\{\omega : X(\omega) \in B\})$$

□

Une autre possibilité eût donc été de prendre pour *définition* de l'espérance des variables aléatoires du type $\mathbb{1}_B(X)$ la quantité $\mathbb{P}(X \in B)$, puis d'étendre la définition en *postulant* la linéarité.

2.2.2 Variance et moments d'ordre supérieur.

Définition 2.10. Soit X variable aléatoire réelle. On dit que X admet un moment d'ordre $p \in \mathbb{N}$ si la variable aléatoire X^p est intégrable, et, dans ce cas, le moment d'ordre p est alors la quantité $\mathbb{E}[X^p]$

Notons que, si $1 \leq p < q$, alors $|x|^p \leq |x|^q + 1$ (en distinguant si $|x| \leq 1$ ou $|x| > 1$) et partant :

$$\mathbb{E}[|X|^p] \leq \mathbb{E}[|X|^q] + 1$$

donc l'existence d'un moment d'ordre q implique celle d'un moment d'ordre p. Les deux moments les plus importants sont les deux premiers moments (le premier moment est simplement l'espérance).

Les premier et second moments, $\mathbb{E}[X]$ et $\mathbb{E}[X^2]$, sont les plus importants. Le second moment $\mathbb{E}[X^2]$ est en lien avec la variance qui est une mesure de la dispersion de la variable aléatoire autour de son espérance.

Définition 2.11. Soit X variable aléatoire réelle de carré intégrable : $\mathbb{E}[X^2] < \infty$. On appelle variance de X la quantité :

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

On se rappellera du moyen mnémothechnique appris au lycée : "c'est la moyenne des carrés des écarts à la moyenne" (ici, moyenne doit être compris comme l'espérance). La variance mesure la *dispersion des valeurs autour de l'espérance*. Puisqu'un carré est positif, $\mathrm{Var}(X) \geq 0$ par positivité de l'espérance. En toute généralité, puisque $X - \mathbb{E}[X] \geq 0$, on peut définir la variance (égale à $+\infty$) même si X n'est pas de carré intégrable. Si on développe la relation de définition de la variance par linéarité, on obtient la formule suivante, quelquefois appelée :

Lemme 2.12 (Formule de Huyghens)**.** *Soit X variable aléatoire réelle de carré intégrable. On a :*

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

Au passage, on a que $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$, une formule qu'on peut voir comme un cas particulier de l'inégalité de Jensen **??** à venir.

Proposition 2.13. *Soit X une variable aléatoire réelle de carré intégrable.*

1. *Si $a, b \in \mathbb{R}$,*
$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$$
En particulier, $\mathrm{Var}(X + b) = \mathrm{Var}(X)$.
2. $\mathrm{Var}(X) = 0$ *si et seulement si X est p.s. constante, c'est-à-dire que* $\mathbb{P}(X = \mathbb{E}[X]) = 1$.

Démonstration. 1. Par linéarité de l'espérance, $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$, si bien que

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

2. On définit la variable aléatoire $Y = (X - \mathbb{E}[X])^2$. Alors $Y \geq 0$ et $Y = 0$ si et seulement si $X = \mathbb{E}[X]$. Le Lemme **??** ci-dessous appliqué à Y donne alors

$$X = \mathbb{E}[X] \text{ p.s. ssi } \mathbb{E}[Y] = 0 \text{ ssi } \mathrm{Var}(X) = 0.$$

□

Enfin, la proposition suivante connecte espérance et variance de façon naturelle.

Lemme 2.14. *Soit X une variable aléatoire réelle de carré intégrable. On a :*

$$\mathrm{Var}(X) = \min_{a \in \mathbb{R}} \mathbb{E}[(X - a)^2],$$

et le minimum est atteint en un unique élément a égal à $\mathbb{E}[X]$.

Cet énoncé dit que la *constante* qui approche le mieux la variable aléatoire X au sens d'une certaine distance (dite des "moindres carrés" ou encore du "risque quadratique", un terme plus particulièrement employé en économie) est la constante égale à $\mathbb{E}[X]$, et que la distance associée est alors la variance de la variable aléatoire.

Démonstration. Par linéarité de l'intégrale,

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

avec égalité si et seulement si $a = \mathbb{E}[X]$. □

2.3 Variables aléatoires usuelles

Il est temps de considérer les exemples les plus classiques de variables aléatoires et de calculer enfin quelques moments de façon concrète. Les lois usuelles sont présentées dans l'ordre de leur fréquence d'apparition, de la plus courante à la plus spécifique.

Loi uniforme

Définition 2.15. On dit que X suit la loi uniforme sur l'ensemble fini $S \subset \mathbb{R}$ (symboliquement : $X \sim \mathrm{Unif}(S)$), si

$$p_X(x) = \frac{1}{\mathrm{Card}(S)}, \qquad x \in S$$

Tous les éléments de l'ensemble fini S sont donc équiprobables, d'où le nom de la loi. Les formules de moments ne se simplifient pas spécialement, mais notons tout de même :

$$\mathbb{E}[X] = \frac{1}{\mathrm{Card}(S)} \sum_{x \in S} x \text{ et } \mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{\sum_{x\in S} x^2}{\mathrm{Card}(S)} - \left(\frac{\sum_{x\in S} x}{\mathrm{Card}(S)}\right)^2.$$

Loi de Bernoulli

Définition 2.16. On dit que X suit la loi de Bernoulli de paramètre $p \in [0,1]$ (symboliquement : $X \sim \mathrm{Ber}(p)$), si

$$p_X(1) = \mathbb{P}(X = 1) = p, \quad p_X(0) = \mathbb{P}(X = 0) = 1 - p.$$

Le nom de la loi de Bernoulli vient de *l'expérience de Bernoulli :* c'est une expérience aléatoire ayant deux issues possibles : *succès* ou *échec*. On pose alors $X = 1$ en cas de succès et $X = 0$ en cas d'échec, et si p est la probabilité de succès, alors X suit la loi de Bernoulli de paramètre p. Notons que dans le cas $p = 1/2$, $\mathrm{Ber}(p)$ coïncide avec $\mathrm{Unif}(\{0,1\})$. On peut calculer simplement ses moments comme suit : pour tout $k \in \mathbb{N}$,

$$\mathbb{E}[X^k] = 0^k \cdot (1-p) + 1^k \cdot p = p \text{ et } \mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p^2 - p = p(1-p).$$

Ainsi la variance est maximisée pour $p = 1/2$, et elle est bien sûr nulle pour les variables p.s. constantes associées aux valeurs $p = 0$ et $p = 1$.

Loi binomiale

La loi binomiale est une généralisation importante de la loi de Bernoulli.

Définition 2.17. On dit que X suit la loi binomiale de paramètres $n \in \mathbb{N}$ et $p \in [0,1]$ (symboliquement : $X \sim \mathrm{Bin}(n,p)$), si

$$p_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k \in \{0, \ldots, n\}.$$

Noter que $\mathrm{Bin}(1,p)$ coïncide avec $\mathrm{Ber}(p)$. Le fait que cette fonction de masse définit une probabilité correspond au théorème binomial de Newton :

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

En effet il suffit maintenant de poser $a = p$ et $b = 1 - p$ pour obtenir que

$$\sum_{k\in\mathbb{N}} p_X(k) = \sum_{0\leq k\leq n} \binom{n}{k} p^k (1-p)^{n-k} = 1$$

Interprétation : Il s'agit du nombre de nombre de succès lors de n répétitions indépendantes d'une expérience de Bernoulli de probabilité de succès p.

Pour le calcul des moments on procède comme suit : on commence par noter l'identité :

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

puis :

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} n\binom{n-1}{k-1} p^{k-1+1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{j=0}^{n} \binom{n-1}{j} p^j (1-p)^{(n-1)-j} \\
&= np
\end{aligned}$$

De même, on peut calculer ce qu'on appelle le *second moment factoriel* :

$$\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} n(n-1)\binom{n-2}{k-2} p^{k-2+2} (1-p)^{(n-2)-(k-2)} \\
&= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{(n-1)-j} \\
&= n(n-1)p^2
\end{aligned}$$

Ainsi

$$\begin{aligned}
\mathrm{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 \\
&= n(n-1)p^2 - np + (np)^2 \\
&= np(1-p)
\end{aligned}$$

c'est-à-dire n fois la variance de la loi de Bernoulli de paramètre p : ce lien se trouvera expliqué quand on aura justifié l'interprétation précédente.

Loi de Poisson

Définition 2.18. On dit que X suit la loi de Poisson de paramètre $\lambda \geq 0$ (symboliquement, $X \sim \mathrm{Po}(\lambda)$), si pour tout $k \in \mathbb{N}$,

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

Le fait que cette fonction de masse définit une probabilité correspond au développement en série entière de la fonction exponentielle : pour tout z complexe, $e^z = \sum_{k\geq 0} \frac{z^k}{k!}$, obtenu en calculant les dérivées successives de exp en 0.

On calcule comme suit le premier moment :

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

De même, on peut calculer le *second moment factoriel* :

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

Ainsi

$$\begin{aligned}
\mathrm{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 \\
&= \lambda^2 + \lambda - \lambda^2 \\
&= \lambda
\end{aligned}$$

Noter que ce sont les cas limites des valeurs pour les moments associés de $\mathrm{Bin}(n,p)$ lorsque $n \to \infty$ et $p = p(n)$ avec $np \to \lambda$. En effet, la loi de Poisson de paramètre $\lambda \geq 0$ peut-être vue comme une approximation de la loi binomiale de paramètres n et $p = \lambda/n$, *dès lors que p est petit*. En effet, on peut vérifier (exercice) que pour $\lambda \geq 0$ et $k \in \mathbb{N}$ fixés,

$$\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \to e^{-\lambda}\frac{\lambda^k}{k!}, \quad n \to \infty.$$

Loi géométrique

Définition 2.19. On dit que X suit la loi géométrique de paramètre $p \in]0,1]$ (symboliquement : $X \sim \text{Geom}(p)$), si

$$p_X(k) = \mathbb{P}(X = k) = p(1-p)^{k-1}, \quad k \in \mathbb{N}^* = \{1, 2, \ldots\}.$$

Interprétation : une v.a. X de loi géométrique de paramètre p représente le nombre de d'expériences de Bernoulli indépendantes de probabilité de succès p jusqu'à la première expérience fructueuse. En effet, la probabilité d'avoir eu des échecs pendant les $k-1$ premiers essais et un succès au k-ième essai est exactement égal à $(1-p)^{k-1} \times p$, si les expériences sont indépendantes. Nous reviendrons rigoureusement sur cette interprétation dans la section qui suit.

Le fait que cette fonction de masse définit une probabilité correspond au développement en série entière de la fonction $1/(1-z)$. Rappelons le raisonnement : pour tout complexe z,

$$\sum_{k=0}^{n} z^k \cdot (1-z) = \sum_{k=0}^{n} z^k - \sum_{k=0}^{n} z^{k+1} = 1 - z^{n+1}$$

donc pour $z \neq 1$, $\sum_{k=0}^{n} z^k = \frac{1-z^{n+1}}{1-z}$ puis par passage à la limite pour $|z| < 1$,

$$\sum_{k=0}^{\infty} z^k = \frac{1}{1-z}.$$

Ceci entraîne, pour $p \in]0,1]$, $\sum_{k=0}^{\infty}(1-p)^{k-1} = \frac{1}{1-(1-p)} = \frac{1}{p}$, c'est-à-dire $\sum_{k=0}^{\infty} p(1-p)^{k-1} = 1$. Aussi, par dérivation terme à terme de la série entière,

$$\sum_{k=0}^{\infty} k z^{k-1} = \frac{1}{(1-z)^2}.$$

donc

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} kp(1-p)^{k-1} = \frac{1}{p}.$$

De même, par une nouvelle dérivation :

$$\sum_{k=0}^{\infty} k(k-1) z^{k-2} = \frac{2}{(1-z)^3}.$$

donc

$$\mathbb{E}[X(X-1)] = \sum_{k=0}^{\infty} k(k-1)p(1-p)^{k-1} = \frac{2(1-p)}{p^2}.$$

donc

$$\text{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X^2] = \frac{2(1-p)}{p^2} + \frac{1}{p} - (\frac{1}{p})^2 = \frac{1-p}{p^2},$$

qui prend bien la valeur 0 en 1, et tend vers $+\infty$ quand $p \to 0$. Noter que la *queue* de la variable aléatoire prend une forme particulièrement simple :

$$\mathbb{P}(X > k) = \sum_{j>k} \mathbb{P}(X = j) = \sum_{j>k} p(1-p)^{j-1} = p(1-p)^k \sum_{\ell \geq 0} p^\ell = (1-p)^k, \quad k \in \mathbb{N}.$$

Remarque 2.20. *Il est également fréquent de rencontrer la loi géométrique sur* $\mathbb{N}$ *; il s'agit simplement de la loi géométrique décalée de 1 : précisément* $Y \sim \mathrm{Geom}_{\mathbb{N}}(p)$ *si* $Y + 1 \sim \mathrm{Geom}(p)$ *si*

$$\mathbb{P}(Y = k) = p(1-p)^k, \quad k \in \mathbb{N} = \{0, 1, 2 \ldots\}.$$

On fera bien attention à ajouter l'indice $\mathbb{N}$ *si on manipule cette loi ; les deux lois se rencontrent aussi fréquemment l'une que l'autre en pratique.*

Loi binomiale négative

On dit que X suit la loi binomiale négative de paramètres $r \in \mathbb{N}^\star$ et $p \in]0, 1]$ (symboliquement : $X \sim \mathrm{NegBin}(r, p)$, si

$$\mathbb{P}(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n \in \{r, r+1, \ldots\}$$

Interprétation : il s'agit du nombre de répétitions indépendantes dans une suite d'expériences de Bernoulli jusqu'à cumuler r succès. Noter que la dernière expérience (incluse dans le décompte) est nécessairement un succès ; ainsi les $r-1$ autres succès doivent être répartis parmi les $n-1$ expériences précédentes, ce qui explique le facteur combinatoire $\binom{n-1}{r-1}$ tandis que chaque configuration a pour probabilité $p^r(1-p)^{n-r}$. En particulier, $\mathrm{Geom}(p) = \mathrm{BN}(1, p)$. Le fait que cette fonction de masse définit une probabilité est un calcul intéressant :

$$(1-q)^{-r} = ((1-q)^{-1})^r = (\sum_{j \geq 0} q^j)^r = (\sum_{j_1 \geq 0} p^{j_1}) \times \ldots \times (\sum_{j_r \geq 0} q^{j_r}) = \sum_{j_1, \ldots, j_r \geq 0} q^{j_1 + \ldots + j_r}$$

Maintenant, pour trouver le coefficient de q^k dans cette expression, il faut se demander, pour $k \in \mathbb{N}$ fixé, combien de r-uplets d'entier $j_1, \ldots, j_r$ réalisent la contrainte $1_{\{j_1 + \ldots + j_r = k\}}$. Cet ensemble est en bijection avec les chemins Nord-Est du plan de $(0, 0)$ a $(k, r-1)$, dénombré par le coefficient binomial $\binom{r-1+k}{r-1}$, aussi :

$$\begin{aligned}(1-q)^{-r} &= \sum_{k \geq 0} q^k \Big(\sum_{j_1, \ldots, j_r \geq 0} 1_{\{j_1 + \ldots + j_r = k\}} \Big) \\ &= \sum_{k \geq 0} \binom{r-1+k}{r-1} q^k\end{aligned}$$

Ainsi, en réindiçant et en posant $q = 1 - p$, on tombe bien sur l'identité cherchée :

$$1 = \sum_{k \geq 0} \binom{r-1+k}{r-1} q^k (1-q)^r = \sum_{n \geq r} \binom{n-1}{r-1} (1-p)^{n-r} p^r$$

Remarque 2.21. *On rencontre quelquefois une autre convention pour la loi la loi Binomiale négative qui consiste à compter le nombre d'*échecs *avant de cumuler* r *succès : c'est donc simplement la loi de* $X - r$ *:*

$$\mathbb{P}(X - r = j) = \binom{r+j-1}{r-1} p^r (1-p)^j, \quad j \in \mathbb{N} = \{0, 1, \ldots\}$$

Loi hypergéométrique

On dit que X suit la loi hypergeométrique de paramètres $n, N, m \in \mathbb{N}$ avec $n \leq N$ et $m \leq N$ (symboliquement : $X \sim \text{Hypergeo}(n, N, m)$), si

$$\mathbb{P}(X = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}, \quad k = 0, \ldots, \min\{m, n\}.$$

Interprétation : on tire sans remise un échantillon de n boules d'une urne en contenant N, dont m blanches et $N - m$ noires. Si X désigne le nombre de boules blanches tirées, alors $X \sim \text{Hypergeo}(n, N, m)$.

Imaginons pour comprendre la formule que les boules blanches et noires soient numérotées et considérons un tirage de n boules comme la liste ordonnée des numéros des boules ainsi que leur couleur. On a $(m \times \ldots \times (m - k + 1))((N - m) \times \ldots \times (N - m - (n - k) + 1)) \times \binom{n}{k}$ tirages qui conduisent à choisir m boules blanches : il faut choisir les numéros des boules puis l'ordre d'apparition des couleurs dans la liste. Le nombre total de tirages possibles est lui $N \times \ldots \times (N - n + 1)$ on a donc la formule.

$$\mathbb{P}(X = k) = \frac{(m \times \ldots \times (m - k + 1))((N - m) \times \ldots \times (N - m - (n - k) + 1)) \times \binom{n}{k}}{N \times \ldots \times (N - n + 1)}.$$

Il est facile de voir que cette formule coïncide avec celle proposée. Il existe un lien avec la loi binomiale dans la limite des grandes urnes :

Exercice 2.22. *Soit X_N de loi* Hypergeo(n, N, m). *On suppose que $m = m_N$ est une suite telle que $m_N/N \to p$ quand $N \to \infty$ (c'est-à-dire que la proportion de boules blanches converge dans la limite des grandes urnes). Montrer que, quand $N \to \infty$,*

$$\mathbb{P}(X_N = k) \longrightarrow \binom{n}{k} p^k (1-p)^{n-k}, \qquad k \in \{0, \ldots, n\}.$$

2.4 Indépendance de variables aléatoires

2.4.1 Définition

Définition 2.23. n variables aléatoires $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ sont dites indépendantes si pour tous événements $A_1, \ldots, A_n \subset \Omega$,

$$\mathbb{P}\Big(\bigcap_{1 \leq i \leq n} \{X_i \in A_i\} \Big) = \prod_{1 \leq i \leq n} \mathbb{P}(X_i \in A_i).$$

Remarque 2.24. *Un changement subtil de point de vue est le suivant : on peut aussi voir $X_1, \ldots, X_n$ comme les n coordonnées* d'un seul *vecteur aléatoire $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$, et l'indépendance décrit alors la loi du vecteur X en fonction de ses lois marginales, c'est-à-dire des lois des coordonnées X_i.*

C'est-à-dire que $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ sont indépendantes si pour tous $A_1, \ldots, A_n \subset \Omega$,

$$\text{les événements } \{X_1 \in A_1\}, \ldots, \{X_n \in A_n\} \text{ sont indépendants.}$$

Bien entendu, de la définition, si $X_1, \ldots, X_n$ sont indépendants, et $1 \leq i_1 < \ldots < i_p \leq n$, alors $X_{i_1}, \ldots, X_{i_p}$ sont indépendants : il suffit de prendre $A_i = \Omega$ pour $i \notin \{i_1, \ldots, i_p\}$. Aussi, on a la proposition simple suivante :

Proposition 2.25. *Soit $X_1, \dots, X_n : \Omega \to \mathbb{R}$ des variables aléatoires indépendantes et $f_1, \dots, f_n : \mathbb{R} \to \mathbb{R}$ des fonctions de la variables réelle. Alors $f_1(X_1), \dots, f_n(X_n)$ sont encore indépendantes.*

Démonstration. Soit $A_1, \dots, A_n \subset \mathbb{R}$ des parties de $\mathbb{R}$, on calcule :

$$\begin{aligned}\mathbb{P}(\bigcap_{1\le i\le n}\{f_i(X_i)\in A_i\}) &= \mathbb{P}(\bigcap_{1\le i\le n}\{X_i\in f_i^{-1}(A_i)\})\\ &= \prod_{1\le i\le n}\mathbb{P}(\{X_i\in f_i^{-1}(A_i)\}) \quad \text{par indépendance des } X_i\\ &= \prod_{1\le i\le n}\mathbb{P}(f_i(X_i)\in A_i)\end{aligned}$$

□

On a aussi le lemme suivant dit des coalitions (admis) :

Proposition 2.26. *Soit $n_1, n_2 \in \mathbb{N}^\star$, et $X_1, \dots, X_{n_1+n_2} : \Omega \to \mathbb{R}$ des variables aléatoires indépendantes, $f_1 : \mathbb{R}^{n_1} \to \mathbb{R}$, $f_2 : \mathbb{R}^{n_2} \to \mathbb{R}$ deux fonctions. Alors $f_1(X_1, \dots, X_{n_1})$ et $f_2(X_{n_1+1}, \dots, X_{n_1+n_2})$ sont indépendantes.*

Notons aussi que l'ensemble des atomes d'un couple de variables aléatoires indépendante prend nécessairement la forme d'un pavé.

Proposition 2.27. *Soit $X_1, X_2 : \Omega \to \mathbb{R}$ deux variables aléatoires. On note $S_{(X_1,X_2)} = \{(x_1, x_2) \in \mathbb{R}^2, \mathbb{P}((X_1, X_2) = (x_1, x_2)) > 0\}$ l'ensemble des atomes du couple X_1, X_2, S_{X_1} et S_{X_2} l'ensemble des atomes de X_1 et X_2 respectivement. Alors*

$$S_{X_1,X_2} \subset S_{X_1} \otimes S_{X_2},$$

et on a égalite si les deux variables sont indépendantes.

Notons qu'on peut avoir égalité sans pour autant que les deux variables ne soient pas indépendantes : il ne s'agit pas d'un "si et seulement si".

Démonstration. $(x_1, x_2) \in S_{X_1,X_2}$ signifie $\mathbb{P}((X_1, X_2) = (x_1, x_2)) > 0$ et donc $\mathbb{P}(X_1 = x_1) = \sum_x \mathbb{P}((X_1, X_2) = (x_1, x)) \ge \mathbb{P}((X_1, X_2) = (x_1, x_2)) > 0$ d'où $x_1 \in S_{X_1}$ (et par symétrie des rôles de X_1 et X_2, on a de même que $x_2 \in S_{X_2}$). Maintenant, si $x_1 \in S_{X_1}$ et $x_2 \in S_{X_2}$ et X_1, X_2 sont indépendantes, alors

$$\mathbb{P}((X_1, X_2) = (x_1, x_2)) = \mathbb{P}(\{X_1 = x_1\} \cap \{X_2 = x_2\}) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2) > 0$$

□

La définition de l'indépendance s'étend naturellement au cas d'une suite infinie de variables aléatoires.

Définition 2.28. Une collection infinie de variables aléatoires aléatoires $(X_i)_{i\in\mathbb{N}}$ est dite indépendante si pour tout $n \ge 1$ entier, les variables aléatoires sont $X_1, \dots, X_n$ sont indépendantes.

On utilisera l'acronyme *i.i.d.* pour une suite de variables aléatoires indépendantes et identiquement distribuées. L'existence d'une telle suite de loi donnée n'est pas évidente (c'est le théorème de Daniell-Kolmogorov) et sera ici admise.

2.4.2 Indépendance et moments

Proposition 2.29. *Soit $X, Y : \Omega \to \mathbb{R}$ des variables aléatoires, et $f, g : \mathbb{R} \to \mathbb{R}$ telles que $f(X)$ et $g(Y)$ soient intégrables. Si X et Y sont indépendantes, alors $f(X)g(Y)$ est une variable aléatoire intégrable et*

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].$$

Remarque 2.30. *Notons en particulier ($f = g = id$) que X, Y intégrables et indépendantes implique que le produit XY est intégrable (ce n'est pas forcément le cas si les variables X, Y ne sont pas indépendantes) avec*

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

Il peut être utile de clarifier la signification du membre de gauche $\mathbb{E}[f(X)g(Y)]$ dans l'égalité ci-dessus : $f(X)g(Y)$ est une fonction de ω, et donc une certaine variable aléatoire dont l'espérance vaut donc

$$\mathbb{E}[f(X)g(Y)] = \sum_{x,y} f(X(\omega))g(Y(\omega))p(\omega).$$

En pratique pour les calculs, on retiendra que l'on dispose de la formule de transfert suivante, dont l'extension au cas d'un vecteur aléatoire est sans difficulté.

Proposition 2.31 (Formule de transfert)**.** *Soit $X, Y : \Omega \to \mathbb{R}$ deux variables aléatoires définies sur $(\Omega, \mathbb{P})$, et $\varphi : \mathbb{R}^2 \to \mathbb{R}$ une fonction. $\varphi(X, Y)$ est intégrable ssi la famille $(|\varphi(x, y)|\mathbb{P}((X, Y) = (x, y)))_{x \in X(\Omega)}$ est sommable et dans ce cas,*

$$\mathbb{E}[\varphi(X, Y)] = \sum_{(x,y)} \varphi(x, y)\mathbb{P}((X, Y) = (x, y)) = \sum_{(x,y)} \varphi(x, y)\mathbb{P}(\{X = x\} \cap \{Y = y\}).$$

Dans le cas particulier $f = \mathbb{1}_A, g = \mathbb{1}_B$, on retombe sur la définition de l'indépendance :

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

On est maintenant en position de faire la démonstration de la Proposition **??**.

Démonstration de la proposition **??**. On utilise la définition de l'indépendance puis Fubini (pour vérifier que la collection $(f(x)g(y)\mathbb{P}(X = x)\mathbb{P}(Y = y))_{x,y}$ est sommable, on peut utiliser Fubini positif) :

$$\begin{aligned}
\mathbb{E}[f(X)g(Y)] = \sum_{(x,y)} f(x)g(y)\mathbb{P}((X, Y) = (x, y)) &= \sum_{(x,y)} f(x)g(y)\mathbb{P}(X = x \cap Y = y) \\
&= \sum_{(x,y)} f(x)g(y)\mathbb{P}(X = x)\mathbb{P}(Y = y) \\
&= \Big(\sum_x f(x)\mathbb{P}(X = x)\Big)\Big(\sum_y g(y)\mathbb{P}(Y = y)\Big) \\
&= \mathbb{E}[f(X)]\mathbb{E}[g(Y)]
\end{aligned}$$

□

2.4.3 Indépendance et variance

Une propriété essentielle de la variance est qu'elle est additive si les variables aléatoires sont *indépendantes*. (La linéarité de l'espérance ne nécessitait pas l'indépendance).

Proposition 2.32. *Soient $X, Y : \Omega \to \mathbb{R}$ deux v.a. indépendantes de carré intérgrable, c'est-à-dire telles que $\mathbb{E}[X^2] < \infty$ et $\mathbb{E}[Y^2] < \infty$. Alors*

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Démonstration. Il nous faut justifier que si X et Y sont de carré intégrables, c'est aussi le cas de XY. Mais ceci découle de l'inégalité : $2|XY| \leq X^2 + Y^2$ (qui vient elle-même en développant $(X+Y)^2 \geq 0$ et $(X-Y)^2 \geq 0$). On en donc en droit de développer dans les formules qui suivent :

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

par linéarité de l'espérance et indépendance. □

La variance est par définition une mesure de l'écart entre la variable aléatoire et son espérance. On peut aussi donner une nouvelle interprétation à l'aide d'un couple de variables aléatoires. Si X et Y sont deux variables aléatoires i.i.d. de carré intégrables, par linéarité de l'espérance puis par indépendance,

$$\mathbb{E}[(X-Y)^2] = \mathbb{E}[X^2] - 2\mathbb{E}[XY] + \mathbb{E}[Y^2] = 2(\mathbb{E}[X^2] - \mathbb{E}[X]^2) = 2\,\mathrm{Var}(X),$$

c'est à dire que la variance mesure aussi (à un facteur multiplicatif 2 près) l'espérance du carré de la distance entre deux réalisations *indépendantes*.

2.4.4 Construction à partir du schéma de Bernoulli.

On admettra l'existence de l'objet suivant, qui est le modèle canonique de suite binaire aléatoire :

Définition 2.33. On appelle *schéma de Bernoulli* $(B_i, i \geq 1)$ une suite de variables aléatoires de loi de Bernoulli indépendantes. On dit que le schéma est de paramètre $p \in [0,1]$ si p est la paramètre commun des lois de Bernoulli.

Il est essentiel de comprendre comment les variables aléatoires précédemment introduites peuvent être comprises à l'aide du schéma de Bernoulli, puisque ce dernier est en quelque sorte la *brique de base* du probabiliste.

Theorème 2.34. *Soit $(B_i, i \geq 1)$ un schéma de Bernoulli de paramètre $p \in]0,1]$, et $n, r \geq 1$ deux entiers. Alors :*
(i) $\sum_{i=1}^{n} B_i$ *suit une loi* $\mathrm{Bin}(n,p)$.
(ii) $\min\{i \geq 1 : B_i = 1\}$ *suit une loi* $\mathrm{Geom}(p)$.

(iii) $\min\{i \geq 1 : \sum_{1\leq j\leq i} B_j = r\}$ *suit une loi* $\mathrm{NegBin}(r,p)$.

Un corollaire immédiat est que $\mathrm{NegBin}(r,p)$ est la loi de la somme de r variables aléatoires indépendantes de loi $\mathrm{Geom}(p)$, résultat qui sera redémontré dans l'exercice **??**.

Démonstration. On prouve d'abord l'énoncé (i) au sujet de $\mathrm{Bin}(n,p)$. Soit $k \in \{0,1\}$. On observe que, si $k \in \{0,1\}$, on peut écrire

$$\mathbb{P}(B_1 = k) = p^k(1-p)^{1-k}.$$

Ainsi, par indépendance, pour $k_1, \ldots, k_n \in \{0,1\}$,

$$\mathbb{P}(\bigcap_{1\leq i\leq n} \{B_i = k_i\}) = p^{\sum k_i}(1-p)^{n-\sum_i k_i},$$

Et donc pour $0 \leq k \leq n$,

$$\begin{aligned}
\mathbb{P}(\sum_{i=1}^n B_i = k) &= \mathbb{P}\Big(\bigcup_{k_1+\ldots+k_n=k} \bigcap_{1\leq i\leq n} \{B_i = k_i\}\Big) \\
&= \sum_{k_1+\ldots+k_n=k} \mathbb{P}\Big(\bigcap_{1\leq i\leq n} \{B_i = k_i\}\Big) \\
&= \sum_{k_1+\ldots+k_n=k} p^k(1-p)^{n-k} \\
&= \binom{n}{k} p^k(1-p)^{n-k}.
\end{aligned}$$

On prouve maintenant l'énoncé (iii) au sujet de $\mathrm{NegBin}(r,p)$, dont le cas $r = 1$ implique l'énoncé (ii). On commence par remarquer que, pour $k \geq r$:

$$\Big\{\min\{i \geq 1 : \sum_{1\leq j\leq i} B_i = r\} = k\Big\} = \{\sum_{1\leq j\leq k-1} B_i = r-1\} \cap \{B_k = 1\}$$

Les deux événements de droite étant indépendants, on peut calculer comme suit en utilisant l'énoncé au sujet de la loi binomiale :

$$\mathbb{P}\Big(\min\{i \geq 1 : \sum_{1\leq j\leq i} B_i = r\} = k\Big) = \mathbb{P}\Big(\sum_{1\leq j\leq k-1} B_i = r-1\Big)\mathbb{P}(B_k = 1) = \binom{k-1}{r-1} p^r(1-p)^{k-r}$$

□

2.5 Inégalités et moments

L'inégalité de Markov est l'inégalité fondamentale permettant de borner des probabilités par des espérances. Cette inégalité repose sur une inégalité simple de fonctions élémentaires.

Lemme 2.35. *Soit* $y \geq 0$ *et* $x > 0$. *Alors*

$$\mathbb{1}_{y\geq x} \leq \frac{y}{x}.$$

Démonstration. On distingue les deux cas : si $y \geq x$, alors $\mathbb{1}_{y \geq x} = 1 \leq y/x$, et si $y < x$, alors $0 \leq y/x$, car y et x sont positifs. □

Proposition 2.36 (Inégalité de Markov)**.** *Soit X une v.a. positive. Alors pour tout $x > 0$,*

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

Démonstration. Soit $x > 0$. Puisque $X \geq 0$, le Lemme **??** montre que $\mathbb{1}_{X \geq x} \leq X/x$. Par le Lemme **??**,

$$\mathbb{P}(X \geq x) = \mathbb{E}[\mathbb{1}_{X \geq x}] \leq \mathbb{E}\left[\frac{X}{x}\right] = \frac{\mathbb{E}[X]}{x},$$

par linéarité de l'espérance. □

Inégalité de Chebychev

Theorème 2.37 (Inégalité de Chebychev)**.** *Soit X une v.a. d'espérance finie. Alors pour tout $x > 0$,*

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

Démonstration. On remarque que

$$\mathbb{P}(|X - \mathbb{E}[X]| > x) = \mathbb{P}((X - \mathbb{E}[X])^2 > x^2).$$

La variable $(X - \mathbb{E}[X])^2$ étant positive, on peut appliquer l'inégalité de Markov (Theorem **??**) pour obtenir

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

□

Inégalité de Jensen

Soit $I \subset \mathbb{R}$ un intervalle. On rappelle qu'une fonction $\varphi : I \to \mathbb{R}$ est *convexe* si pour tout $x, y \in I$, $t \in [0, 1]$,

$$\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y).$$

Remarque 2.38. *On note que si X est une variable aléatoire qui prend deux valeurs et telle que $\mathbb{P}(X = x) = p, \mathbb{P}(X = y) = 1 - p$, alors l'inégalité précédente s'écrit encore très simplement :*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]\cdot$$

L'inégalité de Jensen consiste en la généralisation de cette inégalité aux variables aléatoires réelles X intégrables.

Si $\varphi \in C^2(I)$, alors f est convexe si et seulement si $\varphi'' \geq 0$. Des exemples sont les fonctions

$$(x \mapsto e^x),\ (x \mapsto e^{-x}) \text{ et } (x \mapsto |x|^k) \text{ pour } k \geq 1.$$

Une fonction φ telle que $-\varphi$ est convexe est dite *concave*. Exemples :

$$(x > 0 \mapsto \log x) \text{ et } (x \geq 0 \mapsto x^k) \text{ pour } k \leq 1$$

On admet qu'une fonction convexe admet des *minorants affines* en tout point : si $x_0 \in I$, alors il existe $a \in \mathbb{R}$ tel que pour tout $x \in I$,

$$\varphi(x) \geq \varphi(x_0) + a(x - x_0).$$

(Dans le cas où f est dérivable en x_0, on peut simplement choisir $a = f'(x_0)$.)

Proposition 2.39 (Inégalité de Jensen)**.** *Soit X variable aléatoire réelle intégrable, à valeurs dans un intervalle $I \subset \mathbb{R}$ et $\varphi : I \to \mathbb{R}$ convexe. Alors*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)],$$

inégalité dans lequel le membre de droite peut être infini.

Démonstration. Posons $x_0 = \mathbb{E}[X]$. On peut alors montrer qu'il existe $a \in \mathbb{R}$, tel que pour tout $x \in \mathbb{R}$,

$$\varphi(x_0) + a(x - x_0) \leq \varphi(x).$$

Par conséquent,

$$\mathbb{E}[\varphi(X)] \geq \mathbb{E}[\varphi(x_0) + a(X - x_0)] = \varphi(x_0) + a(\mathbb{E}[X - x_0]) = \varphi(x_0) + a(\mathbb{E}[X] - x_0) = \varphi(x_0),$$

car $x_0 = \mathbb{E}[X]$. □

Exemple 2.40. Soit $t \in \mathbb{R}$. La fonction $x \mapsto e^{tx}$ est convexe sur $\mathbb{R}$ donc, pour X variable aléatoire réelle intégrable,

$$e^{t\mathbb{E}[X]} \leq \mathbb{E}[e^{tX}],$$

ce dernier terme pouvant être infini.

Remarque 2.41. *Les inégalités développées dans cette section valent dans un cadre plus général, pour des variables aléatoires réelles non nécessairement discrètes, avec une nouvelle définition étendue de l'espérance : tout cela sera donc revu l'an prochain.*

2.6 Somme de variables aléatoires

2.6.1 Fonction génératrice

Pour étudier les lois de somme en particulier, un outil très utile est la fonction génératrice de la variable aléatoire.

Définition 2.42. Soit $X : \Omega \to \mathbb{N}$ une variable aléatoire à valeurs entières. On appelle fonction génératrice de la variable aléatoire et on note φ_X la fonction :

$$\varphi_X : [-1, 1] \to \mathbb{R}, s \mapsto \varphi_X(s) = \mathbb{E}[s^X] = \sum_{k \in \mathbb{N}} s^k p_X(k).$$

Noter que si $|s| \leq 1$, $|s^X| = |s|^X \leq 1^X = 1$ donc la variable aléatoire s^X est bien intégrable. En d'autres termes, le rayon de convergence de la série entière $\varphi_X(s)$ est supérieur ou égal à 1 : le rayon de convergence peut prendre toutes les valeurs de $[1, +\infty]$, cette dernière valeur étant atteinte par exemple si l'ensemble $\{k \in \mathbb{N} : p_X(k) \neq 0\}$ est fini, c'est-à-dire si la variable aléatoire prend un nombre fini de valeurs.

Proposition 2.43. *φ_X est continue sur l'intervalle fermé $[-1,1]$, de classe $\mathcal{C}^\infty$ sur l'intervalle ouvert $]-1,1[$ et pour tout $\ell \in \mathbb{N}$,*

$$\begin{aligned}\varphi_X^{(\ell)}(s) &= \sum_{k\geq \ell} k(k-1)\dots(k-\ell+1)s^{k-\ell}p_X(k)\\ &= \mathbb{E}\left[X(X-1)\dots(X-\ell+1)s^X\right]\\ &= \mathbb{E}\left[\frac{X!}{(X-\ell)!}s^X\right]\end{aligned}$$

Démonstration. $\sum_{k\in\mathbb{N}} s^k\mathbb{P}(X=k)$ prend la valeur 1 en $s=1$, donc il s'agit d'une série entière de rayon de convergence $R\geq 1$: elle est donc de classe $\mathcal{C}^\infty$ sur $]-1,1[\subset]-R,R[$. La continuité sur $[-1,1]$ découle de la convergence normale : si $|s|\leq 1$, $|s^k p_X(k)| \leq p_X(k)$ et $\sum_k p_X(k) = 1$. Enfin, la formule concernant la dérivée ℓ-ième correspond au théorème de dérivation terme à terme des séries entières. □

Corollaire 2.44. *Pour tout $k\in\mathbb{N}$,*

$$p_X(k) = \frac{\varphi_X^{(k)}(0)}{k!}.$$

En particulier, il suit que la fonction génératrice caractérise la loi de X.

On pourrait même ajouter : la fonction génératrice *sur un voisinage de 0* caractérise la loi.

Démonstration. Il suffit d'évaluer la formule de la Proposition **??** en $s=0$. □

La fonction génératrice permet quelquefois de simplifier le calcul de moments.

Corollaire 2.45. *Soit $X:\Omega\to\mathbb{N}$ une variable aléatoire à valeurs entières. Alors on a les égalités*

$$\varphi_X^{(\ell)}(1-) = \mathbb{E}\left[\frac{X!}{(X-\ell)!}\right]$$

dans $[0,+\infty]$.

Démonstration. Il suffit de prendre la limite de la formule précédente en $s=1$. □

En particulier, $\mathbb{E}[X] = \varphi'_X(1-)$ et $\mathbb{E}[X(X-1)] = \varphi''_X(1-)$, et donc, pour X de carré intégrable, soit dès lors que $\varphi''_-(1) < \infty$,

$$\mathrm{Var}(X) = \varphi''_X(1-) + \varphi'_X(1-) - (\varphi'_X(1-))^2.$$

L'avantage de travailler avec la fonction génératrice est qu'on se ramène à l'étude d'une fonction de la variable réelle, et on peut faire sensiblement plus d'opérations avec une telle fonction qu'avec une suite (la donnée de la loi de X à valeurs entières correspond à la donnée de la suite $(\mathbb{P}(X=k))_{k\in\mathbb{N}}$). En outre, la fonction génératrice se comporte bien vis-à-vis de l'indépendance, comme en atteste la proposition ci-dessous.

Proposition 2.46. *Soit $X,Y:\Omega\to\mathbb{N}$ deux variables aléatoires à valeurs entières indépendantes. Alors, pour tout $s\in[-1,1]$,*

$$\varphi_{X+Y}(s) = \varphi_X(s)\varphi_Y(s)$$

Une application typique est donnée dans les deux exercices qui suivent :

Exercice 2.47. *Soit $n, m \in \mathbb{N}^\star$, et $p \in [0,1]$.*

1. *Calculer la fonction génératrice de la loi* $\mathrm{Bin}(n,p)$.
2. *En déduire espérance et variance de la loi* $\mathrm{Bin}(n,p)$.
3. *Soit X, Y deux variables aléatoires indépendantes de loi* $\mathrm{Bin}(n,p)$ *et* $\mathrm{Bin}(m,p)$. *Quelle est la loi de la somme $X+Y$?*

Corrigé 2.48. *Du théorème binomial, si $X \sim \mathrm{Bin}(n,p)$:*

$$\varphi_X(s) = \mathbb{E}[s^X] = \sum_{k=0}^{n} s^k \binom{n}{k} p^k (1-p)^{n-k} = (ps + (1-p))^n,$$

Cette fonction est définie et dérivable sur $\mathbb{R}$, de dérivée

$$\varphi'_X(s) = n(ps + (1-p))^{n-1} p$$

et

$$\varphi''_X(s) = n(n-1)(ps + (1-p))^{n-2} p^2$$

ainsi $\mathbb{E}[X] = \varphi'_X(1) = np$ et $\varphi''_X(1) = n(n-1)p^2$, d'où : $\mathrm{Var}(X) = n(n-1)p^2 + np - (np)^2 = np(1-p)$. *Donc si $Y \sim \mathrm{Bin}(m,p)$ est indépendante de X,*

$$\varphi_{X+Y}(s) = \varphi_X(s)\varphi_Y(s) = (ps + (1-p))^{n+m},$$

et donc, puisque la fonction caractéristique caractérise la loi, $X+Y$ suit une loi $\mathrm{Bin}(n+m,p)$

Exercice 2.49. *Soit $r, t \in \mathbb{N}^\star$, et $p \in]0,1]$. (Noter que $p = 0$ est cette fois-ci exclu).*

1. *Calculer la fonction génératrice de la loi* $\mathrm{NegBin}(r,p)$.
2. *En déduire espérance et variance de la loi* $\mathrm{NegBin}(r,p)$.
3. *Soit X, Y deux variables aléatoires indépendantes de loi* $\mathrm{NegBin}(r,p)$ *et* $\mathrm{NegBin}(t,p)$. *Quelle est la loi de la somme $X+Y$?*

Corrigé 2.50. *Rappelons que* $\mathrm{Card}\{(k_1,\dots,k_r) \in \mathbb{N}^r : k_1 + \dots + k_r = k\} = \binom{k+r-1}{r-1}$. *On repart de l'expression, pour $r \in \mathbb{N}^\star$,*

$$(1-z)^{-r} = (\sum_{k\geq 0} z^k)^r = \sum_{k_1,\dots,k_r \geq 0} z^{k_1+\dots+k_r} = \sum_{k\geq 0} \binom{r+k-1}{r-1} z^k$$

d'où

$$(1-z)^{-(r+1)} = \sum_{k\geq 0} \binom{r+k}{r} z^k = \sum_{k\geq r} \binom{k}{r} z^{k-r}$$

On en déduit le calcul de la fonction génératrice de $X \sim \mathrm{NegBin}(r,p)$, on a :

$$\begin{aligned}
\varphi_X(s) &= \sum_{k\geq r} s^k p_X(k) \\
&= \sum_{k\geq r} s^k \binom{k-1}{r-1} p^r (1-p)^{k-r} \\
&= (ps)^r \sum_{k-1\geq r-1} \binom{k-1}{r-1} ((1-p)s)^{(k-1)-(r-1)} \\
&= \left(\frac{ps}{(1-(1-p)s)}\right)^r
\end{aligned}$$

Maintenant, la fonction $\varphi_X(s)$ *est définie et dérivable sur* $]-1/(1-p), 1/(1-p)[$, *et*

$$\varphi'_X(s) = r\left(\frac{ps}{1-(1-p)s)}\right)^{r-1} \frac{p}{(1-(1-p)s)^2}$$

donc $\mathbb{E}[X] = \varphi(1) = r/p$. *Aussi,*

$$\varphi''_X(s) = r(r-1)\left(\frac{ps}{1-(1-p)s)}\right)^{r-2}\left(\frac{p}{(1-(1-p)s)^2}\right)^2 + r\left(\frac{ps}{1-(1-p)s)}\right)^{r-1}\frac{2p(1-p)}{(1-(1-p)s)^3}$$

donc $\varphi''_X(1) = \frac{r(r-1)+2r(1-p)}{p^2}$ *puis*

$$\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$$

Maintenant, si $Y \sim \mathrm{NegBin}(t,p)$ *est indépendante de* X, *pour* $|s| \leq 1$,

$$\varphi_{X+Y}(s) = \varphi_X(s)\varphi_Y(s) = \left(\frac{ps}{(1-(1-p)s)}\right)^{r+t},$$

et donc, puisque la fonction caractéristique caractérise la loi, $X+Y$ *suit une loi* $\mathrm{NegBin}(r+t,p)$. *En particulier, on retrouve que la somme de* r *loi géométriques indépendantes de paramètre de succès* p *suit une loi* $\mathrm{NegBin}(r,p)$.

2.6.2 Loi faible des grands nombres

Theorème 2.51 (Loi faible des grands nombres)**.** *Soient* $X_1, X_2, \ldots$ *indépendantes et de même loi, avec* $\mathbb{E}[|X_1|] < \infty$. *Alors, pour tout* $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{X_1+\cdots+X_n}{n} - \mathbb{E}[X_1]\right| > \varepsilon\right) \to 0 \text{ quand } n \to \infty.$$

La loi des grands nombres justifie l'interprétation de l'espérance comme « moyenne » de la v.a. : la *moyenne empirique* de n réalisations de cette v.a. approche son espérance quand n est grand. Du point de vue statistique, la loi des grands nombres permet l'estimation de l'espérance au travers de la donnée d'une suite de variables i.i.d.

Remarque 2.52. *On dit aussi qu'on a convergence de* $\frac{X_1+\cdots+X_n}{n}$ *vers* $\mathbb{E}[X_1]$ *en probabilité. La loi des grands nombres vaut encore avec un mode de convergence plus fort (elle est alors appelée loi* forte *des grands nombres), mais nous laissons cela pour l'année prochaine.*

Démonstration. Nous commençons par fournir la preuve sous l'hypothèse plus forte $\mathbb{E}[X_1^2] < \infty$. L'intérêt de cette hypothèse est qu'elle permet d'offrir une borne quantitative sur la quantité à contrôler :

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{X_1+\cdots+X_n}{n} - \mathbb{E}[X_1]\right| > \varepsilon\right) &= \mathbb{P}\left(\left|\frac{X_1+\cdots+X_n}{n} - \mathbb{E}\left[\frac{X_1+\cdots+X_n}{n}\right]\right| > \varepsilon\right) \\
&\leq \frac{\mathrm{Var}(\frac{X_1+\cdots+X_n}{n})}{\varepsilon^2} \quad \text{de Chebychev} \\
&= \frac{\mathrm{Var}(X_1)}{\varepsilon^2 n} \to 0.
\end{aligned}$$

quand $n \to \infty$, ayant noté que par indépendance, $\mathrm{Var}\left(\frac{X_1+\cdots+X_n}{n}\right) = \frac{1}{n^2}\left(\mathrm{Var}(X_1)+\cdots+\mathrm{Var}(X_n)\right) = \frac{\mathrm{Var}(X_1)}{n}$.

(Le reste de la preuve est très largement hors programme).

Nous présentons maintenant l'extension au cas intégrable, c'est-à-dire qu'on suppose seulement $\mathbb{E}[|X_1|] < \infty$. Posons $S_n = X_1 + \ldots + X_n$. Quitte à considérer $X_1 - \mathbb{E}[X_1]$ on peut supposer que les $(X_i)_{1\leq i\leq N}$ sont des variables aléatoires centrées, ce qu'on fera désormais. On suppose donc $\mathbb{E}[X_1] = 0$. Maintenant, pour N entier, on pose :

$$X_i^N = X_i \mathbb{1}_{|X_i|\leq N} \text{ et } S_n^N = \sum_{1\leq i\leq n} X_i^N,$$

c'est-à-dire qu'on considère les variables tronquées ainsi que leur somme. Soit $\varepsilon > 0$ fixé. Il nous suffit de montrer que pour tout n suffisamment grand, $\mathbb{P}(|S_n| > \varepsilon n) \leq \varepsilon$. Pour ce faire, on note que par convergence dominée (pas vu cette année) :

$$\mathbb{E}[X_1^N] \to \mathbb{E}[X_1] = 0 \text{ et } \mathbb{E}[|X_1 - X_1^N|] \to 0,$$

et donc on peut choisir N suffisamment grand tel que

$$|\mathbb{E}[X_1^N]| \leq \varepsilon/4 \text{ et } \mathbb{E}[|X_1 - X_1^N|] = \mathbb{E}[|X_1|\mathbb{1}_{|X_1^N|>N}] \leq \varepsilon/4.$$

(Le choix de ce dernier terme sera expliqué dans quelques lignes). Puisque les variables bornées sont de carré intégrable, la loi faible pour de telles variables implique :

$$\mathbb{P}(|S_n^N - n\mathbb{E}[|X_1^N|]| > \varepsilon n/4) \to 0,$$

et donc il existe un rang n_0 à partir duquel on a $\mathbb{P}(|S_n^N - n\mathbb{E}[|X_1^N]| > \varepsilon n/4) \leq \varepsilon/2$. Notons également que $S_n - S_n^N = \sum_{1\leq i\leq N} X_i \mathbb{1}_{|X_i|>N}$ donc de l'inégalité de Markov :

$$\mathbb{P}(|S_n - S_n^N| > \varepsilon n/2) \leq 2\frac{\mathbb{E}[|S_n - S_n^N|]}{\varepsilon n} \leq 2n\frac{\mathbb{E}[|X_1 - X_1^N|]}{\varepsilon n} \leq \frac{\varepsilon}{2}$$

grâce à la majoration précédente de $\mathbb{E}[|X_1|\mathbb{1}_{|X_1^N|>N}]$. Finalement pour $n \geq n_0$:

$$\begin{aligned}\mathbb{P}(|S_n| > \varepsilon n) &= \mathbb{P}(|S_n^N| > \varepsilon n/2) + \mathbb{P}(|S_n - S_n^N| > \varepsilon n/2)\\ &\leq \mathbb{P}(|S_n^N - n\mathbb{E}[X_1^N]| > \varepsilon n/4) + \mathbb{P}(|S_n - S_n^N| > \varepsilon n/2)\\ &\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \leq \varepsilon.\end{aligned}$$

□

Une application simple, dans le cas où les variables sont supposées de carré intégrable : $\mathbb{E}[X_1^2] < \infty$, est la construction d'intervalles de confiance pour $\mathbb{E}[X_1]$. Il s'agit d'un intervalle *aléatoire*, construit à partir des observations de $X_1, \ldots, X_n$ et qui contient la quantité supposée inconnue $\mathbb{E}[X_1]$ avec probabilité prescrite :

$$\mathbb{P}\left(\mathbb{E}[X_1] \in \left[\frac{X_1+\ldots+X_n}{n} - \alpha\sqrt{\frac{\mathrm{Var}(X_1)}{n}}, \frac{X_1+\ldots+X_n}{n} + \alpha\sqrt{\frac{\mathrm{Var}(X_1)}{n}}\right]\right) \geq 1 - \frac{1}{\alpha^2}$$

Le point de vue des statistiques est le suivant : la loi n'est pas connue, et on suppose estimer certaines de ses caractéristiques (comme les moments) à l'aide d'un échantillon de variables i.i.d. Dans cette optique, on voit mal pourquoi, si l'on cherche à estimer l'espérance $\mathbb{E}[X_1]$, la variance $\mathrm{Var}(X_1)$ serait supposée connue : en pratique, il faut déjà commencer par estimer $\mathrm{Var}(X_1)$ à l'aide des observations.

2.7 Variables dépendantes : covariance et corrélation

Lemme 2.53 (Inégalité de Cauchy–Schwarz)**.** *Soient $X, Y : \Omega \to \mathbb{R}$ des v.a. de carrés intégrables. Alors,*

$$\mathbb{E}[XY]^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2].$$

Démonstration. On peut supposer que $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 1$, sinon on remplace X et Y par $X' = X/\sqrt{\mathbb{E}[X^2]}$ et $Y' = Y/\sqrt{\mathbb{E}[Y^2]}$. On a alors :

$$0 \leq \mathbb{E}[(X-Y)^2] = \mathbb{E}[X^2 + Y^2 - 2XY] = \mathbb{E}[X^2] + \mathbb{E}[Y^2] - 2\mathbb{E}[XY] = 2(1 - \mathbb{E}[XY]).$$

et

$$0 \leq \mathbb{E}[(X+Y)^2] = \mathbb{E}[X^2 + Y^2 + 2XY] = \mathbb{E}[X^2] + \mathbb{E}[Y^2] + 2\mathbb{E}[XY] = 2(1 + \mathbb{E}[XY]).$$

Par conséquent, $|\mathbb{E}[XY]| \leq 1$. □

Définition 2.54. Soit $X, Y : \Omega \to \mathbb{R}$ variables aléatoires de carré intégrable, c'est-à-dire que $\mathbb{E}[X^2] < \infty$ et $\mathbb{E}[Y^2] < \infty$. On définit la *covariance* de X et Y par

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

En développant le produit et en utilisant la linéarité de l'espérance, on obtient

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY - X\mathbb{E}[Y] - \mathbb{E}[X]Y + \mathbb{E}[X]\mathbb{E}[Y]] = \mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y],$$

si bien que

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

une formule qui généralise la formule de Huyghens déjà vue : $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. D'ailleurs, par définition,

$$\mathrm{Var}(X) = \mathrm{Cov}(X, X),$$

donc la variance d'une v.a. est égale à sa covariance avec elle-même.

Proposition 2.55. *Soient X, Y, Z des v.a. de carrés intégrables et $a, b \in \mathbb{R}$,*

1. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
2. $\mathrm{Cov}(aX + bY, Z) = a\,\mathrm{Cov}(X, Z) + b\,\mathrm{Cov}(Y, Z)$.
3. $\mathrm{Cov}(a, X) = 0$.
4. $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$.

NB : Les deux premières propriétés de la proposition ci-dessus disent que la covariance est une *forme bilinéaire symétrique*.

Démonstration. 1. Evident par la définition.

2. Par linéarité de l'espérance,

$$\begin{aligned}\mathrm{Cov}(aX + bY, Z) &= \mathbb{E}[(aX + bY - \mathbb{E}[aX + bY])(Z - \mathbb{E}[Z])] \\ &= a\mathbb{E}[(X - \mathbb{E}[X])(Z - \mathbb{E}[Z])] + b\mathbb{E}[(Y - \mathbb{E}[Y])(Z - \mathbb{E}[Z])] \\ &= a\,\mathrm{Cov}(X, Z) + b\,\mathrm{Cov}(Y, Z).\end{aligned}$$

3. Puisque $\mathbb{E}[a] = a$,
$$\mathrm{Cov}(a, X) = \mathbb{E}[(a-a)X] = \mathbb{E}[0X] = 0.$$

4. On suppose que $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ quitte à remplacer les v.a. par leurs variables centrées. Alors,
$$\begin{aligned}\mathrm{Var}(X+Y) &= \mathbb{E}[(X+Y)^2] \\ &= \mathbb{E}[X^2 + Y^2 + 2XY] \\ &= \mathbb{E}[X^2] + \mathbb{E}[Y^2] + 2\mathbb{E}[XY] \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y).\end{aligned}$$

□

La covariance est une mesure de dépendance entre les variables X et Y. Pour illustrer cela, on considère un exemple :

Exemple 2.56. Soient A et B des événements et soient $\mathbb{1}_A$ et $\mathbb{1}_B$ leurs indicatrices. On calcule
$$\begin{aligned}&\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A), \quad \mathbb{E}[\mathbb{1}_B] = \mathbb{P}(B) \\ &\mathbb{E}[\mathbb{1}_A\mathbb{1}_B] = \mathbb{P}(A \cap B) \\ &\mathrm{Cov}(\mathbb{1}_A, \mathbb{1}_B) = \mathbb{E}[\mathbb{1}_A\mathbb{1}_B] - \mathbb{E}[\mathbb{1}_A]\mathbb{E}[\mathbb{1}_B] = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B).\end{aligned}$$

Par conséquent,
$$\mathrm{Cov}(\mathbb{1}_A, \mathbb{1}_B) = 0 \iff \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \iff A \text{ et } B \text{ sont indépendants.}$$

Ceci conforte l'idée de la covariance comme une mesure de dépendance entre les v.a. On peut même interpréter le signe de la covariance. Supposons que $\mathbb{P}(A) \neq 0$ et $\mathbb{P}(B) \neq 0$, alors on peut également écrire
$$\mathrm{Cov}(\mathbb{1}_A, \mathbb{1}_B) = \mathbb{P}(B)(\mathbb{P}(A \mid B) - \mathbb{P}(A)) = \mathbb{P}(A)(\mathbb{P}(B \mid A) - \mathbb{P}(B))$$
avec $\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$ la probabilité de A sachant B. Cette écriture montre que
$$\mathrm{Cov}(\mathbb{1}_A, \mathbb{1}_B) > 0 \iff \mathbb{P}(A \mid B) > \mathbb{P}(A) \iff \mathbb{P}(B \mid A) > \mathbb{P}(B).$$
Les deux dernières inégalités signifient que la réalisation d'un des deux événements augmente la chance que l'autre événement se réalise. On dit dans ce cas que les deux événements sont *positivement corrélés*. Dans le cas d'une inégalité dans l'autre sens, on dit qu'ils sont *négativement corrélés*.

Résumons : Soient A et B deux événements. Alors, A et B sont
— *positivement corrélés* si $\mathbb{P}(A \cap B) > \mathbb{P}(A)\mathbb{P}(B)$ ($\iff \mathrm{Cov}(\mathbb{1}_A, \mathbb{1}_B) > 0$)
— *négativement corrélés* si $\mathbb{P}(A \cap B) < \mathbb{P}(A)\mathbb{P}(B)$ ($\iff \mathrm{Cov}(\mathbb{1}_A, \mathbb{1}_B) < 0$)
— *indépendantes* si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ ($\iff \mathrm{Cov}(\mathbb{1}_A, \mathbb{1}_B) = 0$).

Au vu du dernier exemple, on dit que deux v.a. X et Y sont
— *positivement corrélées* si $\mathrm{Cov}(X, Y) > 0$

— *négativement corrélées* si $\mathrm{Cov}(X,Y) < 0$
— *non corrélées* si $\mathrm{Cov}(X,Y) = 0$

On remarque que l'absence de corrélation est équivalente à $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. On a alors l'implication

Lemme 2.57. *Si $X, Y : \Omega \to \mathbb{R}$ sont indépendantes, alors* $\mathrm{Cov}(X,Y) = 0$.

La réciproque est fausse en général, mais elle est vraie dans certains cas particuliers (par exemple quand X et Y sont des v.a. de Bernoulli, voir l'exemple ci-dessus).

Exemple 2.58. Soit X une v.a. de loi symétrique avec $\mathbb{E}[X^4] < \infty$. On pose $Y = X^2$ qui est de carré intégrable. Evidemment, X et Y ne sont en général pas indépendantes. En effet, on peut vérifier que X et Y sont indépendantes si et seulement si Y est une variable aléatoire constante. En revanche,

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[X^3] - \mathbb{E}[X]\mathbb{E}[X^2] = 0,$$

car $\mathbb{E}[X^3] = \mathbb{E}[X] = 0$ par symétrie de la loi de X. Ceci donne un exemple où les v.a. X sont non corrélées mais dépendantes.

Définition 2.59. Soit $X, Y : \Omega \to \mathbb{R}$ de carré intégrable, telle que $\mathrm{Var}(X) > 0$ et $\mathrm{Var}(Y) > 0$. On introduit le *coefficient de corrélation* $\rho(X,Y)$ comme suit :

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

Proposition 2.60. *Le coefficient de corrélation satisfait aux l'inégalités suivantes :*

$$-1 \leq \rho(X,Y) \leq 1.$$

Démonstration. Par l'inégalité de Cauchy–Schwarz,

$$|\,\mathrm{Cov}(X,Y)| = |\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]| \leq \sqrt{\mathbb{E}[(X-\mathbb{E}[X])^2]\mathbb{E}[(Y-\mathbb{E}[Y])^2]} = \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}.$$

Ces deux inégalités donnent $|\rho(X,Y)| \leq 1$. □