Chapitre 21

Dénombrement, probabilités et variables aléatoires

21.1 Introduction historique

On voit, par cet Essai, que la théorie des probabilités n'est, au fond, que le bon sens réduit au calcul ; elle fait apprécier avec exactitude ce que les esprits justes sentent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte.

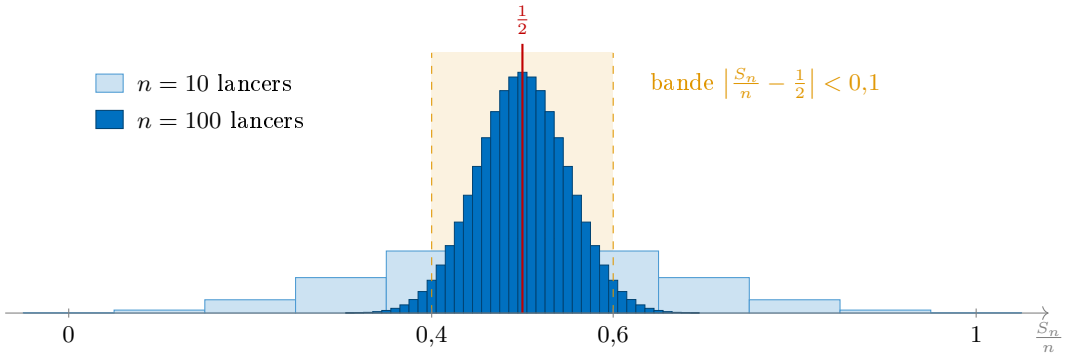
Pierre-Simon Laplace - *Essai philosophique sur les probabilités*, Paris, 1814

21.1.1 Partager une mise interrompue

En 1654, le chevalier de Méré soumet à **Blaise Pascal** le problème des partis : deux joueurs ont misé la même somme, le premier qui gagne trois manches emporte le tout, et la partie est interrompue alors que l'un mène deux manches à une. Comment partager équitablement la mise ? Dans sa correspondance avec **Pierre de Fermat**, la réponse se dégage : deux manches au plus suffiraient à conclure, et des quatre suites également possibles de deux manches, trois donnent la victoire au joueur qui mène. Il doit donc recevoir les trois quarts de la mise. Tout le chapitre est en germe dans ce calcul : on *dénombre* des issues également possibles (section 21.2), on en tire une *probabilité* (21.3), on *conditionne* par ce que l'on sait déjà (21.4), et l'on calcule la *valeur moyenne* d'un gain aléatoire (21.5 et 21.6).

21.1.2 De la fréquence à la probabilité

En 1657, **Christiaan Huygens** publie le premier traité systématique du calcul des chances, *De ratiociniis in ludo aleae*. L'*Ars conjectandi* de **Jacques Bernoulli**, publié en 1713, après sa mort survenue en 1705, contient son « théorème d'or » : la fréquence d'un événement dans des épreuves répétées se rapproche de sa probabilité. L'essai de **Thomas Bayes**, publié en 1763, renverse la question et remonte des effets aux causes ; l'*Essai* de Laplace, en 1814, fait le bilan d'un siècle et demi. **Bienaymé** en 1853, puis **Tchebychev** en 1867, démontrent l'inégalité qui rend le théorème de Bernoulli quantitatif (section 21.8) ; **Markov** étudie à partir de 1906 des suites de variables dépendantes, les chaînes qui portent son nom, et **Kolmogorov** donne en 1933 les axiomes sur lesquels repose aujourd'hui la théorie. La figure illustre le théorème de Bernoulli pour une pièce équilibrée : quand on passe de 10 à 100 lancers, l'histogramme de la fréquence des piles se resserre autour de $\frac{1}{2}$, et la probabilité d'un écart au moins égal à 0,1, hors de la bande orange, tombe de 0,754 à 0,057.



21.2 Dénombrement

Dénombrer, c'est compter les éléments d'un ensemble fini sans les énumérer un à un. Le chapitre 1 a utilisé les propriétés intuitives du nombre d'éléments et annonçait leur justification ; c'est le chapitre 2 qui l'a donnée, en définissant le cardinal par les numérotations et en démontrant ses propriétés (lemmes 1 à 6 et théorème 2). Le chapitre 3 a ensuite défini les coefficients binomiaux par une formule et montré qu'ils comptent les parties d'un ensemble. Cette section rassemble ces résultats, complète la liste des formules utiles, et donne aux formules de Pascal et du binôme leur démonstration naturelle, celle qui compte.

21.2.1 Cardinal d'un ensemble fini

Définition : *(Rappel du chapitre 2.)* Un ensemble E est **fini** s'il existe un entier $n \in \mathbb{N}$ et une bijection de $[\![1\,,n]\!]$ sur E ; cet entier est unique, c'est le **cardinal** de E, noté $\mathrm{card}(E)$, $|E|$ ou $\mathrm{Card}(E)$.

Proposition 1 : *(Rappels.)* Soit E et F deux ensembles finis.

(i) S'il existe une bijection de E sur F, alors $|E| = |F|$, et réciproquement.

(ii) Toute partie A de E est finie, et $|A| \leqslant |E|$, avec égalité si et seulement si $A = E$.

(iii) Si $|E| = |F|$, une application $f : E \to F$ est injective si et seulement si elle est surjective, si et seulement si elle est bijective.

∘ Démonstration : Le point (i) est le lemme 2 du chapitre 2, complété par la remarque qui suit la définition de l'équipotence ; le point (ii) est le lemme 3, et le point (iii) le théorème 2 du même chapitre. □

Remarque : Le point (iii) est le plus utile en pratique : pour montrer qu'une application entre deux ensembles finis de même cardinal est bijective, il suffit de montrer qu'elle est injective, ce qui est souvent la moitié facile du travail. Le théorème 10 du chapitre 15 est l'analogue pour les applications linéaires entre espaces de même dimension finie, où le cardinal est remplacé par la dimension.

21.2.2 Opérations sur les cardinaux

Pour une partie A d'un ensemble E, on note $\overline{A} = E \setminus A$ son complémentaire dans E.

Proposition 2 : Soit E un ensemble fini, A et B deux parties de E.

(i) *Réunion disjointe.* Si A_1, …, A_r sont des parties de E deux à deux disjointes, alors

$$\left|\bigcup_{i=1}^{r} A_i\right| = \sum_{i=1}^{r} |A_i|.$$

(ii) *Différence et complémentaire.* $|B \setminus A| = |B| - |A \cap B|$, et en particulier $|\overline{A}| = |E| - |A|$.

(iii) *Réunion quelconque.* $|A \cup B| = |A| + |B| - |A \cap B|$.

(iv) *Produit cartésien.* Si E_1, …, E_r sont des ensembles finis, alors

$$|E_1 \times \cdots \times E_r| = |E_1| \times \cdots \times |E_r|.$$

(v) *Lemme des bergers.* Si $f : E \to F$ est une application vers un ensemble fini F, alors

$$|E| = \sum_{y \in F} |f^{-1}(\{y\})|.$$

∘ Démonstration :

▷ *Réunion disjointe.* Pour $r = 2$, c'est le lemme 4 du chapitre 2. Si le résultat est vrai pour $r-1$ parties, la réunion $A_1 \cup \cdots \cup A_{r-1}$ est disjointe de A_r, et le cas $r = 2$ donne le rang r.

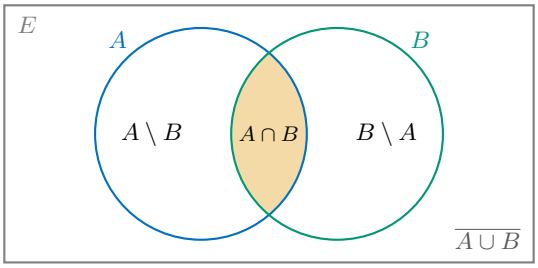
▷ *Différence.* La partie B est la réunion disjointe de $B \cap A$ et de $B \setminus A$, donc $|B| = |A \cap B| + |B \setminus A|$. Pour $B = E$, on obtient $|E| = |A| + |\overline{A}|$.

▷ *Réunion quelconque.* La réunion $A \cup B$ est la réunion disjointe de A et de $B \setminus A$, donc, par ce qui précède,

$$|A \cup B| = |A| + |B \setminus A| = |A| + |B| - |A \cap B|.$$

C'est l'exercice 25 de la feuille du chapitre 2.

▷ *Produit et bergers.* Le point (iv) est le lemme 6 du chapitre 2, étendu par récurrence sur r. Le point (v) est la forme générale du lemme des bergers que donne la remarque qui suit la proposition 7 du chapitre 3 : les fibres sont deux à deux disjointes et recouvrent E, et l'on applique (i). □



Dans la somme $|A| + |B|$, les éléments de la lentille orange $A \cap B$ sont comptés deux fois, une fois avec A, en bleu, et une fois avec B, en vert : la soustraction de $|A \cap B|$ corrige ce double comptage. Les trois morceaux $A \setminus B$, $A \cap B$ et $B \setminus A$ sont disjoints, et c'est sur eux que repose toute la démonstration.

Exemples :

- Parmi les entiers de $[\![1\,,100]\!]$, il y a 50 multiples de 2, 33 multiples de 3, et 16 multiples de 6, c'est-à-dire de 2 et de 3. Il y a donc $50+33-16 = 67$ entiers divisibles par 2 ou par 3, et $100 - 67 = 33$ entiers premiers avec 6.
- Le chapitre 2 annonçait que l'on dénombrerait une réunion en sommant des indicatrices. Pour une partie A de E, on a $|A| = \sum_{x\in E} \mathbf{1}_A(x)$, et la proposition 1 du chapitre 2 donne $\mathbf{1}_{A\cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A\mathbf{1}_B$, avec $\mathbf{1}_A\mathbf{1}_B = \mathbf{1}_{A\cap B}$. En sommant sur $x \in E$, on retrouve le point (iii) : une égalité de fonctions est devenue une égalité de nombres.

Remarque : Pour trois parties, le même calcul donne

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

La formule générale, appelée **formule du crible** ou d'inclusion-exclusion, est hors programme ; on la démontrera néanmoins en quelques lignes à la fin de cette section, car elle découle du développement d'un produit (proposition 5), ce qui tient la promesse de l'exercice 25 de la feuille du chapitre 2. L'erreur la plus répandue est de l'appliquer à des parties disjointes, pour lesquelles le point (i) suffit, ou d'oublier le dernier terme.

21.2.3 Applications et parties

Pour deux ensembles E et F, on note F^E l'ensemble des applications de E dans F ; c'est l'ensemble $\mathcal{F}(E, F)$ du chapitre 2.

Proposition 3 : Soit E et F deux ensembles finis, avec $|E| = p$ et $|F| = n$.

$$|F^E| = n^p \qquad \text{et} \qquad |\mathcal{P}(E)| = 2^p.$$

∘ Démonstration :

▷ *Applications.* Soit $\varphi : [\![1\,,p]\!] \to E$ une numérotation de E. L'application $f \mapsto \big(f(\varphi(1)),\dots,f(\varphi(p))\big)$ est une bijection de F^E sur F^p : une application est déterminée par ses valeurs, que l'on peut choisir librement. Les propositions 1 (i) et 2 (iv) donnent $|F^E| = |F^p| = n^p$.

▷ *Parties.* L'application $A \mapsto \mathbf{1}_A$ est une bijection de $\mathcal{P}(E)$ sur $\{0,1\}^E$, de réciproque $g \mapsto g^{-1}(\{1\})$ (exercice 25 de la feuille du chapitre 2). Le premier point, avec $|\{0,1\}| = 2$, donne $|\mathcal{P}(E)| = 2^p$. □

La proposition 6 du chapitre 1 avait établi $|\mathcal{P}(E)| = 2^p$ par récurrence, en séparant les parties qui contiennent un élément fixé de celles qui ne le contiennent pas ; la bijection avec les indicatrices remplace la récurrence par un seul argument. Cinq lancers d'un dé donnent ainsi $6^5 = 7\,776$ résultats, et il y a $2^8 = 256$ octets, c'est-à-dire de mots de huit chiffres binaires.

21.2.4 Listes, arrangements et permutations

Définition : Soit E un ensemble et $p \in \mathbb{N}$. Une **p-liste** d'éléments de E est un élément $(x_1,\dots,x_p)$ de E^p ; elle est **d'éléments distincts** si les x_i sont deux à deux distincts. On appelle aussi **arrangement** de p éléments de E une p-liste d'éléments distincts, et **permutation** de E une bijection de E sur lui-même.

Une p-liste d'éléments de E n'est autre qu'une application $[\![1\,,p]\!] \to E$, et elle est d'éléments distincts si et seulement si cette application est injective. Dans une liste, l'ordre compte : $(1,2)$ et $(2,1)$ sont deux listes différentes.

Théorème 1 : Listes d'éléments distincts, injections, permutations :
Soit E et F deux ensembles finis, avec $|E| = p$ et $|F| = n$.

- Si $p \leqslant n$, le nombre de p-listes d'éléments distincts de F, et le nombre d'injections de E dans F, valent
$$A_n^p = n(n-1)\cdots(n-p+1) = \frac{n!}{(n-p)!}.$$
Si $p > n$, il n'y en a aucune.
- Le nombre de permutations de F est $n!$.

∘ Démonstration :

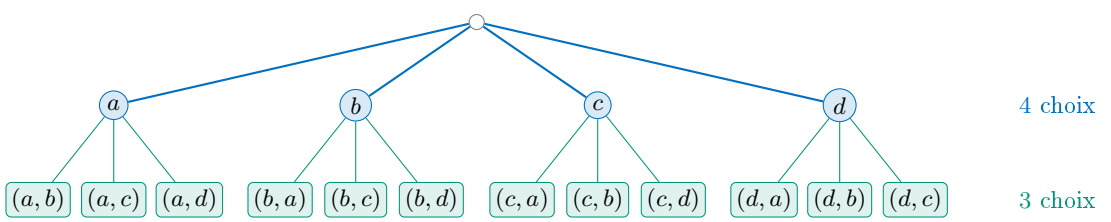
▷ *Listes.* Notons L_p l'ensemble des p-listes d'éléments distincts de F, et montrons par récurrence sur p que $|L_p| = n(n-1)\cdots(n-p+1)$, produit qui est nul dès que $p > n$, puisqu'il contient alors le facteur 0. Pour $p = 0$, il y a une seule liste, la liste vide, et le produit vide vaut 1. Supposons le résultat vrai au rang $p-1$, et considérons l'application

$$\pi : \left| \begin{array}{ccc} L_p & \longrightarrow & L_{p-1} \\ (x_1,\dots,x_p) & \longmapsto & (x_1,\dots,x_{p-1}) \end{array} \right.$$

qui oublie le dernier terme. La fibre de π au-dessus d'une liste $(x_1,\dots,x_{p-1})$ est formée des listes obtenues en lui ajoutant un élément x_p de $F \setminus \{x_1,\dots,x_{p-1}\}$, qui a $n-(p-1)$ éléments par la proposition 2 (ii). Le lemme des bergers (proposition 2 (v)) et l'hypothèse de récurrence donnent

$$|L_p| = \big(n-p+1\big)\,|L_{p-1}| = n(n-1)\cdots(n-p+2)(n-p+1).$$

▷ *Injections et permutations.* Une numérotation φ de E fournit une bijection $f \mapsto \big(f(\varphi(1)),\dots,f(\varphi(p))\big)$ entre les injections de E dans F et L_p. Enfin, une permutation de F est une injection de F dans F (proposition 1 (iii)), et il y en a $A_n^n = n!$; pour $F = [\![1\,,n]\!]$, c'est la proposition 1 du chapitre 17, $\operatorname{card}\mathfrak{S}_n = n!$. □



L'arbre représente les 2-listes d'éléments distincts de $\{a,b,c,d\}$: le premier terme se choisit de 4 façons, en bleu, et, quel que soit ce premier choix, le second de 3 façons, en vert. Le nombre de feuilles est $4\times 3 = 12 = A_4^2$. C'est le lemme des bergers qui justifie la multiplication : chaque nœud du premier niveau porte le même nombre de feuilles.

Exemples :

- Dix coureurs disputent une course : il y a $A_{10}^3 = 10\times 9\times 8 = 720$ podiums possibles, puisqu'un podium est une 3-liste d'éléments distincts, le premier, le deuxième et le troisième.
- Les dates d'anniversaire de n personnes forment une n-liste d'éléments de $[\![1\,,365]\!]$; il y a 365^n telles listes, dont A_{365}^n sont d'éléments distincts. La section 21.3 en tirera le paradoxe des anniversaires.

21.2.5 Parties à p éléments, formules de Pascal et du binôme

Théorème 2 : Nombre de parties à p éléments :
Soit E un ensemble fini à n éléments et $p\in[\![0\,,n]\!]$. Le nombre de parties de E à p éléments est

$$\binom{n}{p} = \frac{A_n^p}{p!} = \frac{n!}{p!\,(n-p)!}.$$

∘ Démonstration : Le théorème 1 du chapitre 3 l'a établi par récurrence sur n, à l'aide de la formule de Pascal démontrée par le calcul. Voici une démonstration qui compte. Soit L_p l'ensemble des p-listes d'éléments distincts de E et $\mathcal{P}_p$ l'ensemble des parties de E à p éléments. L'application

$$\rho : \left|\begin{array}{ccc} L_p & \longrightarrow & \mathcal{P}_p \\ (x_1,\dots,x_p) & \longmapsto & \{x_1,\dots,x_p\}\end{array}\right.$$

qui oublie l'ordre est bien définie, puisque les x_i sont distincts. La fibre au-dessus d'une partie A est l'ensemble des p-listes d'éléments distincts de A, c'est-à-dire des façons d'ordonner A : elle a $A_p^p = p!$ éléments par le théorème 1. Le lemme des bergers donne $|L_p| = p!\,|\mathcal{P}_p|$, d'où $|\mathcal{P}_p| = \frac{A_n^p}{p!}$. □

Choisir une partie à p éléments, c'est choisir p éléments *sans ordre* ; choisir une p-liste d'éléments distincts, c'est les choisir *avec ordre*, et chaque partie correspond à $p!$ listes. Il y a ainsi $\binom{32}{5} = 201\,376$ mains de cinq cartes dans un jeu de 32, comme l'annonçait le chapitre 3.

Proposition 4 : *(Démonstrations combinatoires.)* Pour $n\in\mathbb{N}$ et $p\in[\![0\,,n]\!]$,

$$\binom{n}{p} = \binom{n}{n-p}, \qquad \text{et, si } p<n, \qquad \binom{n}{p} + \binom{n}{p+1} = \binom{n+1}{p+1}.$$

∘ Démonstration :

▷ *Symétrie.* Dans un ensemble E à n éléments, l'application $A\mapsto \overline{A}$ est une bijection, égale à sa propre réciproque, de l'ensemble des parties à p éléments sur l'ensemble des parties à $n-p$ éléments (proposition 2 (ii)).

▷ *Pascal.* Soit E un ensemble à $n+1$ éléments et $a\in E$. Les parties de E à $p+1$ éléments qui ne contiennent pas a sont les parties à $p+1$ éléments de $E\setminus\{a\}$, au nombre de $\binom{n}{p+1}$; celles qui contiennent a s'écrivent $B\cup\{a\}$, où B est une partie à p éléments de $E\setminus\{a\}$ déterminée de façon unique, et il y en a $\binom{n}{p}$. La proposition 2 (i) donne la formule. □

La démonstration de la formule de Pascal est celle de l'hérédité du théorème 1 du chapitre 3 lue à l'envers : là-bas, la formule calculée servait à compter ; ici, le comptage démontre la formule. La même idée, compter un ensemble de deux façons, démontre la formule de Vandermonde de l'exercice 36 de la feuille du chapitre 3 : pour a, $b\in\mathbb{N}$ et $q\in[\![0\,,a+b]\!]$, choisir q éléments dans la réunion disjointe d'un ensemble à a éléments et d'un ensemble à b éléments, c'est en choisir k dans le premier et $q-k$ dans le second, d'où

$$\binom{a+b}{q} = \sum_{k=0}^{q}\binom{a}{k}\binom{b}{q-k}.$$

Pour démontrer la formule du binôme de la même façon, on développe un produit de sommes.

Proposition 5 : *(Développement d'un produit.)* Soit I un ensemble fini, $(a_i)_{i\in I}$ et $(b_i)_{i\in I}$ deux familles de réels ou de complexes. Alors

$$\prod_{i\in I}(a_i+b_i)=\sum_{A\in\mathcal{P}(I)}\Big(\prod_{i\in A}a_i\Big)\Big(\prod_{i\in I\setminus A}b_i\Big).$$

∘ Démonstration : Par récurrence sur $|I|$. Si I est vide, les deux membres valent 1, le seul terme de la somme étant associé à $A=\varnothing$. Supposons le résultat vrai pour les ensembles à m éléments, et soit I à $m+1$ éléments, $j\in I$ et $I'=I\setminus\{j\}$. Par hypothèse de récurrence et distributivité,

$$\prod_{i\in I}(a_i+b_i)=\sum_{A'\in\mathcal{P}(I')}a_j\Big(\prod_{i\in A'}a_i\Big)\Big(\prod_{i\in I'\setminus A'}b_i\Big)+\sum_{A'\in\mathcal{P}(I')}\Big(\prod_{i\in A'}a_i\Big)\Big(\prod_{i\in I'\setminus A'}b_i\Big)b_j.$$

Les termes de la première somme sont ceux de la formule associés aux parties $A=A'\cup\{j\}$ qui contiennent j, ceux de la seconde aux parties $A=A'$ qui ne le contiennent pas ; comme dans la démonstration de la formule de Pascal, les parties de I se répartissent exactement en ces deux catégories, et la relation de Chasles conclut. □

Théorème 3 : Formule du binôme, démonstration combinatoire :
Pour tous réels ou complexes a et b et tout $n\in\mathbb{N}$,

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

∘ Démonstration : On applique la proposition 5 avec $I=[\![1\,,n]\!]$, $a_i=a$ et $b_i=b$: le terme associé à une partie A vaut $a^{|A|}b^{n-|A|}$. On regroupe les parties selon leur cardinal $k\in[\![0\,,n]\!]$; il y en a $\binom{n}{k}$ de cardinal k (théorème 2), et chacune fournit le même terme a^kb^{n-k}. □

Remarques :

- Le chapitre 3 avait démontré la formule par récurrence, et le chapitre 11 l'avait étendue à deux éléments qui commutent dans un anneau (théorème 3). La démonstration ci-dessus explique d'où viennent les coefficients : $\binom{n}{k}$ compte les façons de choisir les k facteurs qui fournissent a.
- Avec $a=b=1$, la formule du binôme donne $\sum_{k=0}^{n}\binom{n}{k}=2^n$: c'est le dénombrement de $\mathcal{P}(E)$ selon le cardinal des parties, qui retrouve la proposition 3.

La proposition 5 permet aussi de compter des objets qui évitent certaines propriétés. Le corrigé de l'exercice 53 de la feuille du chapitre 10 annonçait l'interprétation suivante de la somme S_p qu'il calculait.

Proposition 6 : *(Nombre de surjections.)* Soit E et F deux ensembles finis, avec $|E|=p$ et $|F|=n\geqslant 1$. Le nombre de surjections de E sur F est

$$\sum_{k=0}^{n}(-1)^{n-k}\binom{n}{k}k^p.$$

∘ Démonstration : Pour $f\in F^E$, on a f surjective si et seulement si, pour tout $y\in F$, l'élément y est atteint. Notons $\delta_y(f)=1$ si $y\notin f(E)$, et 0 sinon. Le produit $\prod_{y\in F}\big(1-\delta_y(f)\big)$ vaut 1 si f est surjective et 0 sinon. La proposition 5, avec $a_y=-\delta_y(f)$ et $b_y=1$, donne

$$\prod_{y\in F}\big(1-\delta_y(f)\big)=\sum_{T\in\mathcal{P}(F)}(-1)^{|T|}\prod_{y\in T}\delta_y(f),$$

et le produit $\prod_{y\in T}\delta_y(f)$ vaut 1 si f ne prend aucune valeur dans T, c'est-à-dire si $f\in(F\setminus T)^E$, et 0 sinon. En sommant sur $f\in F^E$ et en intervertissant les sommes (chapitre 3, proposition 6), le nombre de surjections vaut

$$\sum_{T\in\mathcal{P}(F)}(-1)^{|T|}\big|(F\setminus T)^E\big|=\sum_{T\in\mathcal{P}(F)}(-1)^{|T|}(n-|T|)^p=\sum_{j=0}^{n}(-1)^j\binom{n}{j}(n-j)^p,$$

en regroupant les parties T selon leur cardinal j. Le changement d'indice $k = n - j$ et la symétrie $\binom{n}{n-k} = \binom{n}{k}$ donnent la formule. □

Remarques :

- Pour $p = 4$ et $n = 2$, on trouve $2^4 - 2 \times 1^4 = 14$, c'est-à-dire les 16 applications moins les 2 constantes, comme dans l'exercice 25 de la feuille du chapitre 2. Il n'y a aucune surjection si $p < n$, et il y en a $n!$ si $p = n$, ce sont les bijections (proposition 1 (iii)) : on retrouve sans calcul les valeurs $S_p = 0$ pour $p < n$ et $S_n = n!$ des exercices 49 de la feuille du chapitre 9 et 53 de la feuille du chapitre 10. Pour $p = 5$ et $n = 3$, la formule donne $243 - 96 + 3 = 150$, valeur confirmée par une énumération exhaustive.
- Le même calcul, appliqué aux indicatrices de parties A_1, …, A_r d'un ensemble E, donne la formule du crible, hors programme : on développe $\mathbf{1}_{\overline{A_1 \cup \cdots \cup A_r}} = \prod_{i=1}^{r}(1 - \mathbf{1}_{A_i})$ par la proposition 5, puis on somme sur $x \in E$,

$$\left|\bigcup_{i=1}^{r} A_i\right| = \sum_{\varnothing \neq J \subset [\![1,r]\!]} (-1)^{|J|+1} \left|\bigcap_{i \in J} A_i\right|.$$

21.2.6 Méthode

♡ **Méthode : Dénombrer :**

- ▶ On décrit précisément les objets à compter, et l'on se demande si l'ordre compte et si les répétitions sont permises.
- ▶ On reconnaît l'un des quatre modèles du tableau ci-dessous, ou bien on découpe l'ensemble en morceaux disjoints que l'on compte séparément (proposition 2 (i)).
- ▶ Si l'on construit les objets par choix successifs, on vérifie que le nombre de possibilités à chaque étape ne dépend pas des choix précédents, avant de multiplier (lemme des bergers).
- ▶ Pour une condition « au moins un », on compte le complémentaire, « aucun ».

p **éléments parmi** n	**avec répétitions**	**sans répétition**
ordre compté	p-listes, applications : n^p	arrangements : $\dfrac{n!}{(n-p)!}$
ordre non compté	hors programme : $\dbinom{n+p-1}{p}$	parties à p éléments : $\dbinom{n}{p}$

Exemples :

- Une main de cinq cartes d'un jeu de 32 contient un *carré* si elle contient les quatre cartes d'une même hauteur : on choisit la hauteur, de 8 façons, puis la cinquième carte parmi les 28 autres, d'où $8 \times 28 = 224$ mains. Pour un *full*, trois cartes d'une hauteur et deux d'une autre, on choisit la hauteur du brelan, puis ses trois cartes, puis la hauteur de la paire, puis ses deux cartes : $8 \times \binom{4}{3} \times 7 \times \binom{4}{2} = 1\,344$ mains. L'ordre des choix intermédiaires ne crée pas de double comptage, car le brelan et la paire jouent des rôles différents.
- Le nombre de mains contenant au moins un as se compte par le complémentaire :

$$\binom{32}{5} - \binom{28}{5} = 201\,376 - 98\,280 = 103\,096.$$

  L'erreur la plus répandue est de choisir « d'abord un as, puis quatre cartes quelconques », ce qui donne $4 \times \binom{31}{4} = 125\,860$: une main qui contient deux as est alors comptée deux fois, une fois pour chacun de ses as. Une énumération exhaustive confirme la valeur 103 096.

Remarque : Le nombre $\binom{n+p-1}{p}$ du tableau compte, par exemple, les façons de répartir p boules identiques dans n boîtes. Il ne sert pas en probabilités, car les répartitions qu'il compte ne sont pas également probables lorsque l'on lance les boules au hasard : c'est un piège classique de modélisation, que la section suivante permettra de préciser.

21.3 Espaces probabilisés finis

Une expérience aléatoire est une expérience dont on ne peut pas prévoir le résultat, comme le lancer d'un dé ou le tirage d'une main de cartes. Pour en parler mathématiquement, on se donne un ensemble fini Ω dont chaque élément décrit une issue possible, avec assez de détails pour que tout ce qui nous intéresse se lise sur elle. On verra à la section 21.5 qu'il est souvent préférable de ne pas décrire Ω explicitement, et de raisonner sur les quantités observées : l'univers n'est alors qu'une source de hasard, suffisante pour porter ces quantités.

21.3.1 Univers et événements

Définition : Un **univers fini** est un ensemble fini non vide Ω, dont les éléments ω sont appelés **issues**. Un **événement** est une partie de Ω ; l'événement A est *réalisé* lorsque l'issue ω de l'expérience appartient à A. Un **événement élémentaire** est un singleton $\{\omega\}$. Deux événements A et B sont **incompatibles** si $A \cap B = \varnothing$. Un **système complet d'événements** est une famille $(A_1, \dots, A_r)$ d'événements deux à deux incompatibles dont la réunion est Ω.

Le vocabulaire probabiliste traduit celui des ensembles du chapitre 1 :

Notation	Vocabulaire ensembliste	Vocabulaire probabiliste
$\omega \in \Omega$	élément de Ω	issue, résultat de l'expérience
$A \subset \Omega$	partie de Ω	événement
$\varnothing$, Ω	partie vide, ensemble entier	événement impossible, certain
$\overline{A}$	complémentaire de A	événement contraire, « non A »
$A \cap B$, $A \cup B$	intersection, réunion	« A et B », « A ou B »
$A \cap B = \varnothing$	A et B disjointes	A et B incompatibles
$A \subset B$	A incluse dans B	A implique B

Remarques :

- Un système complet d'événements est une partition de Ω au sens du chapitre 1, à ceci près que certains des A_i peuvent être vides. Le chapitre 1 annonçait que les partitions seraient à la base de la formule des probabilités totales : c'est le théorème 5 de la section suivante.
- Le « ou » des probabilités est inclusif, comme celui de la logique du chapitre 1 : l'événement $A \cup B$ est réalisé lorsque l'un au moins des deux événements l'est.

21.3.2 Probabilité sur un univers fini

Définition : Une **probabilité** sur un univers fini Ω est une application $\mathrm{P} : \mathcal{P}(\Omega) \to [0\,,1]$ telle que

- $\mathrm{P}(\Omega) = 1$;
- pour tous événements A et B incompatibles, $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$ *(additivité)*.

Le couple (Ω, P) est appelé un **espace probabilisé fini**.

Proposition 7 : Soit (Ω, P) un espace probabilisé fini, A, B, A_1, …, A_r des événements.

(i) $\mathrm{P}(\varnothing) = 0$, et si les A_i sont deux à deux incompatibles, $\mathrm{P}(A_1 \cup \cdots \cup A_r) = \mathrm{P}(A_1) + \cdots + \mathrm{P}(A_r)$.

(ii) $\mathrm{P}(\overline{A}) = 1 - \mathrm{P}(A)$, et $\mathrm{P}(B \setminus A) = \mathrm{P}(B) - \mathrm{P}(A \cap B)$.

(iii) *Croissance.* Si $A \subset B$, alors $\mathrm{P}(A) \leqslant \mathrm{P}(B)$.

(iv) $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$, et $\mathrm{P}(A_1 \cup \cdots \cup A_r) \leqslant \mathrm{P}(A_1) + \cdots + \mathrm{P}(A_r)$.

(v) Si $(A_1, \dots, A_r)$ est un système complet d'événements, alors

$$\sum_{i=1}^{r} \mathrm{P}(A_i) = 1 \qquad \text{et} \qquad \mathrm{P}(B) = \sum_{i=1}^{r} \mathrm{P}(B \cap A_i).$$

∘ Démonstration :

▷ *(i).* Les événements $\varnothing$ et $\varnothing$ sont incompatibles, donc $\mathrm{P}(\varnothing) = \mathrm{P}(\varnothing) + \mathrm{P}(\varnothing)$, d'où $\mathrm{P}(\varnothing) = 0$. L'additivité pour r événements s'obtient par récurrence sur r, la réunion $A_1 \cup \cdots \cup A_{r-1}$ étant incompatible avec A_r.

▷ *(ii) et (iii).* L'événement B est la réunion des événements incompatibles $B \cap A$ et $B \setminus A$, donc $\mathrm{P}(B) = \mathrm{P}(B \cap A) + \mathrm{P}(B \setminus A)$. Avec $B = \Omega$, on obtient $1 = \mathrm{P}(A) + \mathrm{P}(\overline{A})$. Si $A \subset B$, alors $B \cap A = A$, et $\mathrm{P}(B) = \mathrm{P}(A) + \mathrm{P}(B \setminus A) \geqslant \mathrm{P}(A)$, puisqu'une probabilité est positive.

▷ *(iv).* L'événement $A \cup B$ est la réunion des événements incompatibles A et $B \setminus A$, d'où la formule par (ii). Elle donne $\mathrm{P}(A \cup B) \leqslant \mathrm{P}(A) + \mathrm{P}(B)$, et l'inégalité pour r événements s'en déduit par récurrence.

▷ *(v).* Les A_i sont deux à deux incompatibles et de réunion Ω ; les $B \cap A_i$ sont deux à deux incompatibles et de réunion B. On applique (i). □

La formule (iv) est celle de la proposition 2 (iii), où le cardinal est remplacé par la probabilité : une probabilité « mesure » les événements comme le cardinal mesure les ensembles. L'inégalité de (iv), appelée **sous-additivité**, est grossière mais très utile, car elle ne demande aucune hypothèse sur les A_i : la probabilité que l'un au moins de plusieurs événements rares se produise est au plus la somme de leurs probabilités.

Définition : Une **distribution de probabilités** sur un ensemble fini E est une famille $(p_x)_{x \in E}$ de réels positifs tels que $\displaystyle\sum_{x \in E} p_x = 1$.

Théorème 4 : Une probabilité est déterminée par les événements élémentaires :
Soit Ω un univers fini. Pour toute distribution de probabilités $(p_\omega)_{\omega \in \Omega}$ sur Ω, il existe une unique probabilité P sur Ω telle que $\mathrm{P}(\{\omega\}) = p_\omega$ pour tout $\omega \in \Omega$. Elle est donnée par

$$\mathrm{P}(A) = \sum_{\omega \in A} p_\omega \qquad \text{pour tout } A \subset \Omega.$$

∘ Démonstration :

▷ *Unicité.* Soit P une telle probabilité et A un événement. Les singletons $\{\omega\}$, pour $\omega \in A$, sont deux à deux incompatibles et de réunion A ; la proposition 7 (i) donne $\mathrm{P}(A) = \sum_{\omega \in A} \mathrm{P}(\{\omega\}) = \sum_{\omega \in A} p_\omega$. La probabilité P est donc déterminée par les p_ω.

▷ *Existence.* On définit $\mathrm{P}(A)$ par la formule de l'énoncé. Les p_ω étant positifs, on a

$$0 \leqslant \mathrm{P}(A) \leqslant \sum_{\omega \in \Omega} p_\omega = 1,$$

et $\mathrm{P}(\Omega) = 1$. Si A et B sont incompatibles, la relation de Chasles (chapitre 3, proposition 1 (iii)) donne $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$. Enfin, $\mathrm{P}(\{\omega\}) = p_\omega$. □

Choisir un modèle probabiliste, c'est donc choisir un univers et une distribution sur cet univers ; les probabilités de tous les événements s'en déduisent par addition.

Définition : La **probabilité uniforme** sur un univers fini Ω est celle qui donne la même probabilité $\frac{1}{|\Omega|}$ à chaque événement élémentaire. Pour tout événement A,

$$\mathrm{P}(A) = \frac{|A|}{|\Omega|} = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}.$$

Sous la probabilité uniforme, calculer une probabilité revient à dénombrer : c'est pour cela que la section 21.2 précède celle-ci.

Exemples :

- Le problème des partis de l'introduction se modélise par $\Omega = \{G, P\}^2$, muni de la probabilité uniforme, où G désigne une manche gagnée par le joueur qui mène. Il perd la partie seulement si l'issue est (P, P) : il la gagne avec probabilité $1 - \frac{1}{4} = \frac{3}{4}$, ce qui justifie le partage de Pascal et Fermat.
- On lance deux dés discernables : $\Omega = [\![1, 6]\!]^2$, muni de la probabilité uniforme. L'événement « la somme vaut 7 » contient les six couples $(1, 6)$, $(2, 5)$, …, $(6, 1)$, et sa probabilité vaut $\frac{6}{36} = \frac{1}{6}$: c'est la somme la plus probable.

- Pour $n \geqslant 4$ pair, on choisit au hasard trois des n racines n-ièmes de l'unité, c'est-à-dire une partie à trois éléments de $\mathbb{U}_n$, sous la probabilité uniforme. D'après l'exercice 58 de la feuille du chapitre 4, il y a $\binom{n}{3}$ triangles, dont $\frac{n(n-2)}{2}$ sont rectangles : on choisit un diamètre, de $\frac{n}{2}$ façons, puis le troisième sommet, de $n-2$ façons. La probabilité d'obtenir un triangle rectangle vaut
$$\frac{n(n-2)/2}{n(n-1)(n-2)/6} = \frac{3}{n-1},$$
soit $\frac{3}{5}$ pour l'hexagone et $\frac{3}{11}$ pour le dodécagone. Le corrigé de cet exercice annonçait que cette proportion deviendrait une probabilité : elle tend vers 0, et un triangle choisi au hasard dans un polygone régulier à beaucoup de côtés n'est presque jamais rectangle.

Exemple : Le paradoxe des anniversaires. Les dates d'anniversaire de n personnes forment une n-liste de $\Omega = [\![1\,,365]\!]^n$, que l'on munit de la probabilité uniforme, en ignorant les années bissextiles. L'événement « les dates sont deux à deux distinctes » contient A_{365}^n issues (théorème 1), et la probabilité que deux personnes au moins partagent leur anniversaire vaut
$$p_n = 1 - \frac{A_{365}^n}{365^n} = 1 - \frac{365}{365} \times \frac{364}{365} \times \cdots \times \frac{365-n+1}{365}.$$
On trouve $p_{10} \approx 0{,}117$, $p_{23} \approx 0{,}507$, $p_{30} \approx 0{,}706$ et $p_{57} \approx 0{,}990$. Dès 23 personnes, une coïncidence est plus probable que son absence, ce qui surprend toujours : l'intuition compare 23 à 365, alors qu'il faut comparer à 365 le nombre de *paires* de personnes, $\binom{23}{2} = 253$, dont chacune peut produire une coïncidence.

Remarques :

- L'erreur la plus répandue est d'utiliser la probabilité uniforme sur un univers dont les issues ne sont pas également probables. Si l'on prend pour univers les 21 résultats non ordonnés d'un lancer de deux dés, le résultat « un 1 et un 2 » est deux fois plus probable que « deux 1 », car il correspond aux deux couples $(1,2)$ et $(2,1)$ des dés discernables. Pour appliquer la formule « cas favorables sur cas possibles », on choisit un univers où la symétrie de l'expérience rend les issues également probables : c'est le sens de la remarque qui clôt la section 21.2.
- Certaines expériences ne se décrivent pas par un univers fini : « lancer un dé jusqu'à obtenir un 6 » a une infinité d'issues. En seconde année, les probabilités sur un univers infini dénombrable s'écriront avec les familles sommables du chapitre 20, la somme des p_ω étant alors la somme d'une famille positive.

21.4 Probabilités conditionnelles et indépendance

Dans toute la suite du chapitre, (Ω, P) désigne un espace probabilisé fini.

21.4.1 Probabilité conditionnelle

Savoir qu'un événement B est réalisé modifie les chances des autres événements : on ne considère plus que les issues de B, et l'on renormalise pour que la probabilité totale reste égale à 1.

Définition : Soit B un événement tel que $\mathrm{P}(B) > 0$. Pour tout événement A, la **probabilité de A sachant B** est
$$\mathrm{P}_B(A) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}.$$
On la note aussi $\mathrm{P}(A \mid B)$; nous réservons cette seconde notation au cas où l'événement conditionnant est une intersection, comme $\mathrm{P}(A_3 \mid A_1 \cap A_2)$, dont l'écriture en indice serait illisible.

Proposition 8 : Si $\mathrm{P}(B) > 0$, l'application $\mathrm{P}_B : A \mapsto \mathrm{P}_B(A)$ est une probabilité sur Ω, et $\mathrm{P}_B(B) = 1$.

∘ Démonstration : Pour tout A, on a $0 \leqslant \mathrm{P}(A \cap B) \leqslant \mathrm{P}(B)$ par croissance, donc $\mathrm{P}_B(A) \in [0\,,1]$, et $\mathrm{P}_B(\Omega) = \mathrm{P}_B(B) = \frac{\mathrm{P}(B)}{\mathrm{P}(B)} = 1$. Si A et A' sont incompatibles, $A \cap B$ et $A' \cap B$ le sont aussi, et leur réunion est $(A \cup A') \cap B$; l'additivité de P donne celle de P_B après division par $\mathrm{P}(B)$. □

Exemple : Une famille a deux enfants ; on modélise leurs sexes, dans l'ordre des naissances, par $\Omega = \{G, F\}^2$ muni de la probabilité uniforme. Notons $U = \{(G,G),(G,F),(F,G)\}$ l'événement « la famille compte au moins un garçon » et $D = \{(G,G)\}$ l'événement « elle compte deux garçons ». Sachant U, la probabilité de D vaut

$$\mathrm{P}_U(D) = \frac{\mathrm{P}(D \cap U)}{\mathrm{P}(U)} = \frac{1/4}{3/4} = \frac{1}{3},$$

et non $\frac{1}{2}$. L'erreur la plus répandue est de raisonner sur « l'autre enfant », comme si l'on savait *lequel* est un garçon : sachant que l'aîné est un garçon, la probabilité vaudrait bien $\frac{1}{2}$, car l'information n'est pas la même.

Proposition 9 : *(Formule des probabilités composées.)* Soit $A_1, \ldots, A_n$ des événements tels que $\mathrm{P}(A_1 \cap \cdots \cap A_{n-1}) > 0$. Alors

$$\mathrm{P}(A_1 \cap \cdots \cap A_n) = \mathrm{P}(A_1)\,\mathrm{P}_{A_1}(A_2)\,\mathrm{P}(A_3 \mid A_1 \cap A_2) \cdots \mathrm{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}).$$

∘ Démonstration : Les intersections $A_1 \cap \cdots \cap A_k$, pour $k \leqslant n-1$, contiennent $A_1 \cap \cdots \cap A_{n-1}$ et sont donc de probabilité strictement positive, par croissance : toutes les probabilités conditionnelles sont définies. En les remplaçant par leurs définitions, le produit du second membre est télescopique (chapitre 3, proposition 3) :

$$\mathrm{P}(A_1) \times \frac{\mathrm{P}(A_1 \cap A_2)}{\mathrm{P}(A_1)} \times \cdots \times \frac{\mathrm{P}(A_1 \cap \cdots \cap A_n)}{\mathrm{P}(A_1 \cap \cdots \cap A_{n-1})} = \mathrm{P}(A_1 \cap \cdots \cap A_n).$$

□

Exemple : On tire successivement et sans remise trois cartes d'un jeu de 32. Notons A_k l'événement « la k-ième carte est un as ». Lorsque les $k-1$ premières cartes sont des as, il reste $5-k$ as parmi $33-k$ cartes, et il est naturel de poser $\mathrm{P}(A_k \mid A_1 \cap \cdots \cap A_{k-1}) = \frac{5-k}{33-k}$. La formule des probabilités composées donne

$$\mathrm{P}(A_1 \cap A_2 \cap A_3) = \frac{4}{32} \times \frac{3}{31} \times \frac{2}{30} = \frac{1}{1\,240},$$

qui est bien la proportion $\binom{4}{3}/\binom{32}{3} = \frac{4}{4\,960}$ des mains de trois cartes formées de trois as : les deux modèles, séquentiel et global, donnent le même résultat, et c'est ce qui légitime la modélisation par des probabilités conditionnelles.

Il est commode de donner un sens au produit $\mathrm{P}_B(A)\,\mathrm{P}(B)$ même lorsque $\mathrm{P}(B) = 0$: on convient alors qu'il vaut 0. Avec cette convention, l'égalité $\mathrm{P}(A \cap B) = \mathrm{P}_B(A)\,\mathrm{P}(B)$ est vraie dans tous les cas, puisque $\mathrm{P}(A \cap B) \leqslant \mathrm{P}(B) = 0$.

Théorème 5 : Formule des probabilités totales :
Soit $(B_1, \ldots, B_r)$ un système complet d'événements. Pour tout événement A,

$$\mathrm{P}(A) = \sum_{i=1}^{r} \mathrm{P}_{B_i}(A)\,\mathrm{P}(B_i).$$

∘ Démonstration : La proposition 7 (v) donne

$$\mathrm{P}(A) = \sum_{i=1}^{r} \mathrm{P}(A \cap B_i),$$

et chaque terme vaut $\mathrm{P}_{B_i}(A)\,\mathrm{P}(B_i)$, y compris lorsque $\mathrm{P}(B_i) = 0$, grâce à la convention. □

Théorème 6 : Formule de Bayes :
Soit A et B deux événements de probabilités strictement positives. Alors

$$\mathrm{P}_A(B) = \frac{\mathrm{P}_B(A)\,\mathrm{P}(B)}{\mathrm{P}(A)}.$$

Si $(B_1, \dots, B_r)$ est un système complet d'événements et $\mathrm{P}(A) > 0$, alors, pour tout $j \in [\![1\,, r]\!]$,

$$\mathrm{P}_A(B_j) = \frac{\mathrm{P}_{B_j}(A)\,\mathrm{P}(B_j)}{\sum_{i=1}^{r} \mathrm{P}_{B_i}(A)\,\mathrm{P}(B_i)}.$$

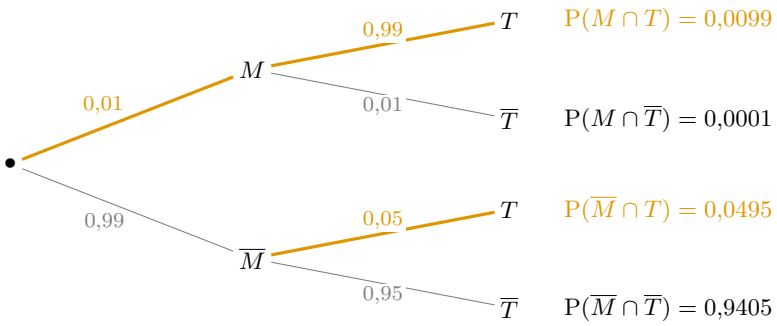
∘ Démonstration : Les deux membres de la première formule valent $\frac{\mathrm{P}(A\cap B)}{\mathrm{P}(A)}$. Pour la seconde, on écrit de même $\mathrm{P}_A(B_j) = \frac{\mathrm{P}(A\cap B_j)}{\mathrm{P}(A)}$, avec $\mathrm{P}(A \cap B_j) = \mathrm{P}_{B_j}(A)\,\mathrm{P}(B_j)$ grâce à la convention, et l'on développe $\mathrm{P}(A)$ par le théorème 5. □

Exemple : Un test de dépistage. Une maladie touche 1 % d'une population. Un test la détecte chez 99 % des malades, mais il est aussi positif chez 5 % des personnes saines. On choisit une personne au hasard ; on note M l'événement « elle est malade » et T l'événement « son test est positif ». Les données sont $\mathrm{P}(M) = 0{,}01$, $\mathrm{P}_M(T) = 0{,}99$ et $\mathrm{P}_{\overline{M}}(T) = 0{,}05$. Le couple $(M, \overline{M})$ est un système complet, et la formule des probabilités totales donne

$$\mathrm{P}(T) = 0{,}99 \times 0{,}01 + 0{,}05 \times 0{,}99 = 0{,}0099 + 0{,}0495 = 0{,}0594,$$

puis la formule de Bayes

$$\mathrm{P}_T(M) = \frac{\mathrm{P}_M(T)\,\mathrm{P}(M)}{\mathrm{P}(T)} = \frac{0{,}0099}{0{,}0594} = \frac{1}{6} \approx 0{,}167.$$



Les nombres portés par les branches issues d'un même nœud ont pour somme 1 ; la probabilité d'une feuille est le produit des nombres lus le long du chemin (formule des probabilités composées), et les quatre feuilles totalisent 1. Les deux chemins orange mènent à un test positif : la formule des probabilités totales additionne leurs probabilités, et la formule de Bayes mesure la part du premier dans cette somme, 0,0099 sur 0,0594. Un test positif ne donne qu'une chance sur six d'être malade, parce que les faux positifs, issus d'une population saine nombreuse, sont cinq fois plus nombreux que les vrais.

♡ **Méthode : Utiliser un arbre pondéré :**

- ▶ On choisit un système complet adapté à la première étape de l'expérience, puis, pour chaque branche, un système complet adapté à l'étape suivante.
- ▶ On porte sur chaque branche la probabilité conditionnelle de son extrémité sachant le chemin qui y mène, et l'on vérifie que les branches issues d'un même nœud totalisent 1.
- ▶ La probabilité d'un chemin est le produit de ses branches ; celle d'un événement est la somme des probabilités des chemins qui le réalisent.

Remarque : L'erreur la plus répandue, en probabilités conditionnelles, est de confondre $\mathrm{P}_B(A)$ et $\mathrm{P}_A(B)$. Dans l'exemple, $\mathrm{P}_M(T) = 0{,}99$ alors que $\mathrm{P}_T(M) \approx 0{,}17$: un test qui détecte presque tous les malades n'est pas un test dont presque tous les positifs sont malades. La formule de Bayes est précisément le moyen de passer de l'une à l'autre, et elle fait intervenir la probabilité $\mathrm{P}(M)$, que l'intuition oublie.

21.4.2 Indépendance d'événements

Définition : Deux événements A et B sont **indépendants** si $\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B)$.
Des événements A_1, …, A_n sont **mutuellement indépendants** si, pour toute partie J de $[\![1\,,n]\!]$,

$$\mathrm{P}\Big(\bigcap_{i\in J} A_i\Big) = \prod_{i\in J} \mathrm{P}(A_i).$$

Ils sont **deux à deux indépendants** si A_i et A_j sont indépendants pour tous $i \neq j$.

Proposition 10 : Soit A et B deux événements.

(i) Si $\mathrm{P}(B) > 0$, alors A et B sont indépendants si et seulement si $\mathrm{P}_B(A) = \mathrm{P}(A)$.

(ii) Si A et B sont indépendants, alors A et $\overline{B}$ le sont, ainsi que $\overline{A}$ et $\overline{B}$.

(iii) Si A_1, …, A_n sont mutuellement indépendants, la famille obtenue en remplaçant certains des A_i par leurs contraires l'est aussi.

∘ Démonstration :

▷ *(i).* On divise l'égalité $\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B)$ par $\mathrm{P}(B) > 0$.

▷ *(iii), qui contient (ii).* Il suffit de remplacer un seul événement par son contraire, l'opération pouvant ensuite être répétée ; par symétrie des rôles, on remplace A_1. Soit $J \subset [\![1\,,n]\!]$. Si $1 \notin J$, l'égalité demandée est l'hypothèse. Sinon, on note $J' = J \setminus \{1\}$ et $C = \bigcap_{i\in J'} A_i$, avec la convention $C = \Omega$ si J' est vide. La proposition 7 (ii) donne

$$\mathrm{P}(\overline{A_1} \cap C) = \mathrm{P}(C) - \mathrm{P}(A_1 \cap C) = \prod_{i\in J'} \mathrm{P}(A_i) - \mathrm{P}(A_1) \prod_{i\in J'} \mathrm{P}(A_i) = \mathrm{P}(\overline{A_1}) \prod_{i\in J'} \mathrm{P}(A_i).$$

Le point (ii) est le cas $n = 2$, appliqué une fois puis deux fois. □

Le point (i) traduit l'idée intuitive : savoir que B est réalisé ne change pas la probabilité de A. Mais la définition par le produit est préférable, car elle est symétrique en A et B et ne suppose rien sur $\mathrm{P}(B)$.

Exemples :

- On choisit un entier X au hasard dans $[\![1\,,60]\!]$, sous la probabilité uniforme, et l'on note A_d l'événement « d divise X ». Comme d divise 60 pour $d \in \{2,3,5,6,10,15,30\}$, on a $\mathrm{P}(A_d) = \frac{1}{d}$ pour ces entiers, et un entier est divisible par 2 et 3 si et seulement s'il l'est par 6 (chapitre 12) : $\mathrm{P}(A_2 \cap A_3) = \frac{1}{6} = \mathrm{P}(A_2)\mathrm{P}(A_3)$. On vérifie de même les autres intersections, jusqu'à $\mathrm{P}(A_2 \cap A_3 \cap A_5) = \frac{1}{30}$: les trois événements sont mutuellement indépendants. Sur $[\![1\,,100]\!]$, en revanche, $\mathrm{P}(A_6) = \frac{16}{100}$ diffère de $\mathrm{P}(A_2)\,\mathrm{P}(A_3) = \frac{50}{100} \times \frac{33}{100}$: l'indépendance dépend de la probabilité, pas seulement des événements.
- On lance deux pièces équilibrées ; A est « la première donne pile », B « la seconde donne pile », C « les deux donnent le même côté ». Chacun a probabilité $\frac{1}{2}$, et chaque intersection de deux d'entre eux est réduite à l'issue « pile, pile », de probabilité $\frac{1}{4}$: ils sont deux à deux indépendants. Mais $\mathrm{P}(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8}$: ils ne sont pas mutuellement indépendants, puisque A et B réalisés ensemble imposent C.

Remarques :

- L'erreur la plus répandue est de confondre indépendants et incompatibles. Deux événements incompatibles de probabilités strictement positives ne sont jamais indépendants, puisque $\mathrm{P}(A \cap B) = 0 < \mathrm{P}(A)\,\mathrm{P}(B)$: savoir que l'un est réalisé apprend que l'autre ne l'est pas.
- Le premier exemple éclaire l'exercice 60 de la feuille du chapitre 20. Si deux entiers U et V sont choisis au hasard dans $[\![1\,,n]\!]$, l'événement « p divise U et V » a une probabilité proche de $\frac{1}{p^2}$; si ces événements étaient indépendants pour les différents nombres premiers p, la probabilité que U et V soient premiers entre eux vaudrait $\prod_p \left(1 - \frac{1}{p^2}\right) = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}$, d'après le produit eulérien de l'exercice 59. Ce raisonnement n'est qu'heuristique, mais l'exercice 60 démontre que la probabilité $\frac{q_n}{n^2}$ tend bien vers $\frac{6}{\pi^2} \approx 0{,}608$; pour $n = 1\,000$, elle vaut 0,608 383.

21.5 Variables aléatoires et lois

Dans la pratique, on ne s'intéresse pas à l'issue ω elle-même, mais à des quantités qui en dépendent : la somme de deux dés, le nombre de piles, le nombre de points fixes d'une permutation. Ce sont les variables aléatoires, et l'on a intérêt à formuler les événements à l'aide de ces variables le plus tôt possible.

21.5.1 Variables aléatoires

Définition : Une **variable aléatoire** sur Ω, à valeurs dans un ensemble E, est une application $X : \Omega \to E$. Elle est **réelle** si $E \subset \mathbb{R}$. Pour toute partie A de E et tout $x \in E$, on note

$$\{X \in A\} = (X \in A) = X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}, \qquad (X = x) = X^{-1}(\{x\}),$$

et, si X est réelle, $(X \leqslant x) = X^{-1}\big(]-\infty, x]\big)$. On écrit $\mathrm{P}(X \in A)$, $\mathrm{P}(X = x)$, $\mathrm{P}(X \leqslant x)$ pour les probabilités de ces événements.

Le nom est trompeur : une variable aléatoire n'est ni une variable ni aléatoire, c'est une fonction, parfaitement déterminée, de l'issue ω ; le hasard porte sur ω. Comme Ω est fini, l'image $X(\Omega)$ est finie. L'image réciproque respecte les opérations ensemblistes (chapitre 2, proposition 2) : par exemple $(X \in A) \cap (X \in B) = (X \in A \cap B)$ et $\overline{(X \in A)} = (X \in E \setminus A)$.

Proposition 11 : Soit X une variable aléatoire sur Ω. La famille des événements $(X = x)$, pour $x \in X(\Omega)$, est un système complet d'événements.

∘ Démonstration : Ces événements sont deux à deux incompatibles, puisqu'une issue ω n'a qu'une image, et leur réunion est Ω, puisque chaque ω appartient à $(X = X(\omega))$. Ce sont les fibres de X au sens du chapitre 2. □

Exemples :

- Pour un événement A, l'indicatrice $\mathbf{1}_A$ est une variable aléatoire à valeurs dans $\{0, 1\}$, et $(\mathbf{1}_A = 1) = A$. Tout événement est ainsi décrit par une variable aléatoire.
- Pour deux dés, avec $\Omega = [\![1, 6]\!]^2$, les variables $D_1 : (a, b) \mapsto a$ et $D_2 : (a, b) \mapsto b$ donnent les résultats de chaque dé, et $S = D_1 + D_2$ leur somme. L'événement « la somme vaut 7 » s'écrit $(S = 7)$, sans qu'il soit nécessaire d'énumérer les six couples qui le composent.

21.5.2 Loi d'une variable aléatoire

Définition : La **loi** d'une variable aléatoire $X : \Omega \to E$ est l'application

$$\mathrm{P}_X : A \longmapsto \mathrm{P}(X \in A),$$

définie sur les parties de $X(\Omega)$, ou de E. Deux variables aléatoires X et Y, éventuellement définies sur des espaces probabilisés différents, ont **même loi** si $\mathrm{P}(X \in A) = \mathrm{P}(Y \in A)$ pour toute partie A ; on note alors $X \sim Y$.

Proposition 12 : La loi P_X est une probabilité sur l'univers fini $X(\Omega)$. Elle est déterminée par la distribution de probabilités $\big(\mathrm{P}(X = x)\big)_{x \in X(\Omega)}$, et

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

∘ Démonstration : On a $\mathrm{P}_X\big(X(\Omega)\big) = \mathrm{P}(\Omega) = 1$, et si A et B sont disjointes, $(X \in A)$ et $(X \in B)$ sont incompatibles, de réunion $(X \in A \cup B)$: l'additivité de P donne celle de P_X. Le théorème 4, appliqué sur l'univers $X(\Omega)$, conclut. □

Donner la loi de X, c'est donc donner $X(\Omega)$ et les probabilités $\mathrm{P}(X = x)$; on les présente souvent dans un tableau. Réciproquement, toute distribution $(p_x)_{x \in E}$ sur un ensemble fini E est la loi d'une variable aléatoire : il suffit de munir $\Omega = E$ de la probabilité donnée par le théorème 4 et de prendre $X = \mathrm{id}_E$.

Proposition 13 : *(Loi de $f(X)$.)* Soit $X : \Omega \to E$ une variable aléatoire et $f : E \to F$ une application. La variable aléatoire $f(X) = f \circ X$ a pour loi

$$\mathrm{P}\big(f(X) = y\big) = \sum_{\substack{x \in X(\Omega) \\ f(x) = y}} \mathrm{P}(X = x) \qquad \text{pour } y \in F.$$

En particulier, si $X \sim Y$, alors $f(X) \sim f(Y)$.

∘ Démonstration : L'événement $\big(f(X) = y\big)$ est $\big(X \in f^{-1}(\{y\})\big)$, et la proposition 12 donne la formule. Le second membre ne dépend que de la loi de X, d'où la seconde affirmation. □

Remarque : Deux variables de même loi ne sont pas égales pour autant. Pour un dé, D_1 et D_2 ont la même loi, comme D_1 et $7 - D_1$, alors que $D_1 = 7 - D_1$ n'est réalisé pour aucune issue. L'erreur la plus répandue est d'en déduire des égalités de variables : de $X \sim Y$, on ne peut pas tirer $X + Y \sim 2X$, puisque $D_1 + (7 - D_1) = 7$ est constante alors que $2D_1$ ne l'est pas.

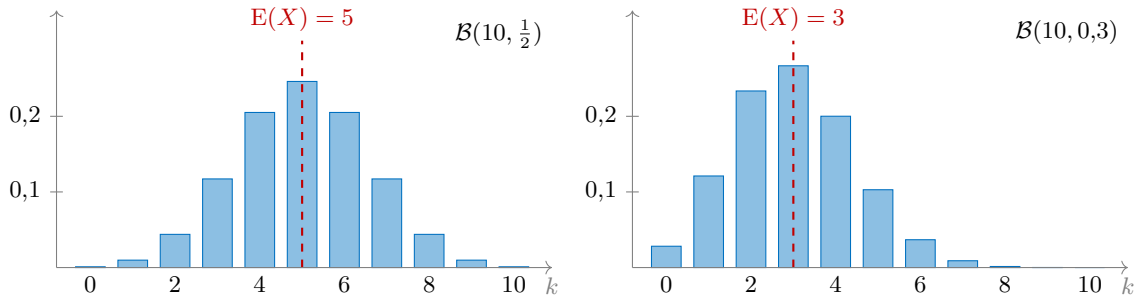
21.5.3 Lois usuelles

Définition : Soit E un ensemble fini non vide, $n \in \mathbb{N}^*$ et $p \in [0\,,1]$. Une variable aléatoire X suit

- la **loi uniforme** sur E, notée $\mathcal{U}(E)$, si $\mathrm{P}(X = x) = \frac{1}{|E|}$ pour tout $x \in E$;
- la **loi de Bernoulli** de paramètre p, notée $\mathcal{B}(p)$, si $X(\Omega) \subset \{0, 1\}$ et $\mathrm{P}(X = 1) = p$;
- la **loi binomiale** de paramètres n et p, notée $\mathcal{B}(n, p)$, si $X(\Omega) \subset [\![0\,, n]\!]$ et

$$\mathrm{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{pour tout } k \in [\![0\,, n]\!].$$

Ces trois familles sont bien des distributions de probabilités : pour la loi binomiale, la formule du binôme donne $\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$. L'indicatrice d'un événement A suit la loi $\mathcal{B}(\mathrm{P}(A))$, et une variable de Bernoulli est l'indicatrice de l'événement $(X = 1)$: on l'interprète comme le résultat, succès ou échec, d'une épreuve. Le théorème 8 montrera que la loi binomiale est celle du nombre de succès en n épreuves indépendantes.



Les barres bleues ont pour hauteurs les probabilités $\mathrm{P}(X = k)$. La loi $\mathcal{B}(10, \frac{1}{2})$ est symétrique autour de 5, puisque $\binom{10}{k} = \binom{10}{10-k}$; la loi $\mathcal{B}(10, 0{,}3)$ est décalée vers la gauche, et sa valeur la plus probable, 3, est aussi son espérance, en rouge, que la section 21.6 calculera. Les probabilités au-delà de 8 sont trop petites pour être visibles : $\mathrm{P}(X \geqslant 8) \approx 0{,}0016$.

Définition : Soit A un événement tel que $\mathrm{P}(A) > 0$. La **loi conditionnelle** de X sachant A est la loi de X dans l'espace probabilisé (Ω, P_A), c'est-à-dire la distribution $\big(\mathrm{P}_A(X = x)\big)_{x \in X(\Omega)}$.

Exemple : Pour deux dés, la loi de D_1 sachant $(S = 7)$ est uniforme sur $[\![1\,,6]\!]$, puisque chaque valeur de D_1 correspond à un seul des six couples de somme 7. Sachant $(S = 4)$, elle est uniforme sur $\{1, 2, 3\}$: l'information sur la somme a restreint les valeurs possibles du premier dé.

21.5.4 Couples et n-uplets de variables aléatoires

Définition : Soit $X : \Omega \to E$ et $Y : \Omega \to F$ deux variables aléatoires. Le **couple** (X, Y) est la variable aléatoire $\omega \mapsto \big(X(\omega), Y(\omega)\big)$, à valeurs dans $E \times F$. Sa loi est appelée **loi conjointe** de X et Y ; elle est déterminée par les probabilités

$$\mathrm{P}(X = x, Y = y) = \mathrm{P}\big((X = x) \cap (Y = y)\big), \qquad (x, y) \in X(\Omega) \times Y(\Omega).$$

Les lois de X et de Y sont les **lois marginales** du couple. On définit de même un **n-uplet** $(X_1, \dots, X_n)$ de variables aléatoires, sa loi conjointe et ses n lois marginales.

Proposition 14 : Les lois marginales se déduisent de la loi conjointe : pour tout $x \in X(\Omega)$,

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

∘ Démonstration : C'est la proposition 7 (v), appliquée au système complet $\big((Y = y)\big)_{y \in Y(\Omega)}$ de la proposition 11. □

Exemple : Une urne contient trois boules numérotées 1, 2, 3 ; on en tire deux successivement, et l'on note X et Y les numéros obtenus. Selon que le tirage a lieu sans remise ou avec remise, les lois conjointes sont, en lisant x en ligne et y en colonne :

sans remise	$y = 1$	$y = 2$	$y = 3$	$\mathrm{P}(X = x)$
$x = 1$	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$
$x = 2$	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{3}$
$x = 3$	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{3}$
$\mathrm{P}(Y = y)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

avec remise	$y = 1$	$y = 2$	$y = 3$	$\mathrm{P}(X = x)$
$x = 1$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
$x = 2$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
$x = 3$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
$\mathrm{P}(Y = y)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

Les marges, obtenues en sommant les lignes et les colonnes, sont les mêmes dans les deux cas : X et Y suivent la loi uniforme sur $\{1, 2, 3\}$. Les lois conjointes sont pourtant différentes. L'erreur la plus répandue est de croire que les lois marginales déterminent la loi du couple : elles ne disent rien de la façon dont X et Y dépendent l'une de l'autre.

21.5.5 Variables aléatoires indépendantes

Définition : Deux variables aléatoires $X : \Omega \to E$ et $Y : \Omega \to F$ sont **indépendantes**, ce que l'on note $X \perp\!\!\!\perp Y$, si

$$\mathrm{P}(X \in A, Y \in B) = \mathrm{P}(X \in A)\,\mathrm{P}(Y \in B) \qquad \text{pour toutes parties } A \subset E \text{ et } B \subset F.$$

Des variables aléatoires $X_1, \dots, X_n$, à valeurs dans $E_1, \dots, E_n$, sont **mutuellement indépendantes** si, pour toutes parties $A_1 \subset E_1, \dots, A_n \subset E_n$,

$$\mathrm{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathrm{P}(X_1 \in A_1) \cdots \mathrm{P}(X_n \in A_n).$$

Proposition 15 : Les variables $X_1, \dots, X_n$ sont mutuellement indépendantes si et seulement si

$$\mathrm{P}(X_1 = x_1, \dots, X_n = x_n) = \mathrm{P}(X_1 = x_1) \cdots \mathrm{P}(X_n = x_n)$$

pour tout $(x_1, \dots, x_n) \in X_1(\Omega) \times \cdots \times X_n(\Omega)$, c'est-à-dire si la loi conjointe est le produit des lois marginales. Dans ce cas, toute sous-famille est formée de variables mutuellement indépendantes, et les événements $(X_1 \in A_1), \dots, (X_n \in A_n)$ sont mutuellement indépendants.

∘ Démonstration :

▷ *Condition nécessaire.* On prend $A_i = \{x_i\}$.

▷ *Condition suffisante.* Soit $A_i \subset E_i$, et $A'_i = A_i \cap X_i(\Omega)$. L'événement $(X_1 \in A_1, \dots, X_n \in A_n)$ est la réunion disjointe des événements $(X_1 = x_1, \dots, X_n = x_n)$, pour $(x_1, \dots, x_n) \in A'_1 \times \cdots \times A'_n$. Par l'hypothèse, puis par le développement d'un produit de sommes (chapitre 3, proposition 7, étendue à n facteurs par récurrence),

$$\mathrm{P}(X_1 \in A_1, \dots, X_n \in A_n) = \sum_{(x_1,\dots,x_n)} \prod_{i=1}^{n} \mathrm{P}(X_i = x_i) = \prod_{i=1}^{n} \Big(\sum_{x_i \in A'_i} \mathrm{P}(X_i = x_i) \Big) = \prod_{i=1}^{n} \mathrm{P}(X_i \in A_i).$$

▷ *Conséquences.* En prenant $A_i = E_i$, donc $(X_i \in A_i) = \Omega$, pour les indices qui n'appartiennent pas à une partie J de $[\![1\,,n]\!]$, on obtient l'égalité de la définition pour la sous-famille $(X_i)_{i\in J}$, ainsi que l'égalité qui définit l'indépendance mutuelle des événements $(X_i \in A_i)$. □

Proposition 16 : Si X_1, …, X_n sont mutuellement indépendantes, et si f_1, …, f_n sont des applications définies sur E_1, …, E_n, alors les variables $f_1(X_1)$, …, $f_n(X_n)$ sont mutuellement indépendantes. En particulier, si $X \perp\!\!\!\perp Y$, alors $f(X) \perp\!\!\!\perp g(Y)$.

∘ Démonstration : Pour des parties B_i des ensembles d'arrivée, $\big(f_i(X_i) \in B_i\big) = \big(X_i \in f_i^{-1}(B_i)\big)$, et l'on applique la définition aux parties $A_i = f_i^{-1}(B_i)$. □

Théorème 7 : Lemme des coalitions :
Soit X_1, …, X_n des variables aléatoires mutuellement indépendantes et $m \in [\![1\,,n-1]\!]$. Pour toutes applications f et g, les variables

$$f(X_1, \dots, X_m) \quad \text{et} \quad g(X_{m+1}, \dots, X_n)$$

sont indépendantes. Plus généralement, des fonctions de paquets disjoints de variables mutuellement indépendantes sont mutuellement indépendantes.

∘ Démonstration : Notons $U = (X_1, \dots, X_m)$ et $W = (X_{m+1}, \dots, X_n)$. Pour $u = (x_1, \dots, x_m)$ et $w = (x_{m+1}, \dots, x_n)$, la proposition 15, appliquée à la famille entière puis aux deux sous-familles, donne

$$\mathrm{P}(U = u, W = w) = \prod_{i=1}^{n} \mathrm{P}(X_i = x_i) = \Big(\prod_{i=1}^{m} \mathrm{P}(X_i = x_i) \Big) \Big(\prod_{i=m+1}^{n} \mathrm{P}(X_i = x_i) \Big) = \mathrm{P}(U = u)\,\mathrm{P}(W = w).$$

Les variables U et W sont donc indépendantes, par la proposition 15 avec $n = 2$, et la proposition 16 conclut. Pour plus de deux paquets, le même calcul montre que les variables U_1, …, U_r formées par les paquets sont mutuellement indépendantes, et la proposition 16 s'applique encore. □

Exemple : Si X_1, …, X_5 sont mutuellement indépendantes et réelles, alors $X_1 + X_2$, X_3X_4 et e^{X_5} sont mutuellement indépendantes. En revanche, $X_1 + X_2$ et $X_2 + X_3$ ne le sont pas en général, puisque les deux paquets partagent X_2 : c'est l'hypothèse de paquets disjoints que l'on oublie le plus souvent.

21.5.6 Modéliser des épreuves indépendantes

L'énoncé « on répète n fois une expérience, de façon indépendante » se traduit par la donnée de variables mutuellement indépendantes X_1, …, X_n. La proposition suivante garantit qu'un tel modèle existe toujours ; on n'a ensuite plus besoin de décrire l'univers.

Proposition 17 : Soit E_1, …, E_n des ensembles finis munis de distributions de probabilités $(p_x^{(1)})_{x\in E_1}$, …, $(p_x^{(n)})_{x\in E_n}$. Sur $\Omega = E_1 \times \cdots \times E_n$, la famille $p_\omega = p_{\omega_1}^{(1)} \cdots p_{\omega_n}^{(n)}$, pour $\omega = (\omega_1, \dots, \omega_n)$, est une distribution de probabilités. Pour la probabilité P qu'elle définit, les projections $X_i : \omega \mapsto \omega_i$ sont mutuellement indépendantes, et X_i a pour loi la distribution $p^{(i)}$.

∘ Démonstration : Les p_ω sont positifs, et le développement d'un produit de sommes donne

$$\sum_{\omega\in\Omega} p_\omega = \prod_{i=1}^{n}\Big(\sum_{x\in E_i} p_x^{(i)}\Big) = 1.$$

Le même calcul, où la i-ième somme est restreinte à $\{x\}$, donne $\mathrm{P}(X_i = x) = p_x^{(i)}$. Enfin,

$$\mathrm{P}(X_1 = x_1, \dots, X_n = x_n) = p_{(x_1,\dots,x_n)} = \prod_{i=1}^{n} \mathrm{P}(X_i = x_i),$$

et la proposition 15 conclut. □

Théorème 8 : Somme de variables de Bernoulli indépendantes :
Si X_1, …, X_n sont mutuellement indépendantes et suivent toutes la loi $\mathcal{B}(p)$, alors

$$X_1 + \cdots + X_n \sim \mathcal{B}(n,p).$$

∘ Démonstration : Soit $S = X_1 + \cdots + X_n$ et $k \in [\![0\,,n]\!]$. L'événement $(S = k)$ est la réunion disjointe des événements $(X_1 = x_1, \dots, X_n = x_n)$, pour les $(x_1, \dots, x_n) \in \{0,1\}^n$ comportant exactement k coordonnées égales à 1. Par indépendance (proposition 15), chacun a pour probabilité $p^k(1-p)^{n-k}$. Un tel n-uplet est déterminé par la partie de $[\![1\,,n]\!]$ formée des indices où il vaut 1 ; il y en a donc $\binom{n}{k}$ (théorème 2), et $\mathrm{P}(S = k) = \binom{n}{k} p^k (1-p)^{n-k}$. □

Le nombre de succès en n épreuves indépendantes de même probabilité de succès p suit donc la loi $\mathcal{B}(n,p)$: le nombre de piles en 10 lancers d'une pièce équilibrée suit la loi $\mathcal{B}(10, \frac{1}{2})$ de la figure, et le nombre de 6 en 10 lancers d'un dé la loi $\mathcal{B}(10, \frac{1}{6})$. L'hypothèse d'indépendance est essentielle : le nombre d'as parmi trois cartes tirées sans remise, dont les indicatrices ne sont pas indépendantes, ne suit pas une loi binomiale.

21.5.7 Trois modèles aléatoires

Exemple : Une marche aléatoire. Un marcheur part de 0 et fait n pas de longueur 1, vers la droite ou vers la gauche avec la même probabilité, indépendamment. On modélise les pas par des variables X_1, …, X_n mutuellement indépendantes de loi uniforme sur $\{-1, 1\}$, et sa position par $S_n = X_1 + \cdots + X_n$. Les variables $B_i = \frac{X_i+1}{2}$ sont mutuellement indépendantes de loi $\mathcal{B}(\frac{1}{2})$ (proposition 16), et $\frac{S_n+n}{2} = B_1 + \cdots + B_n$, nombre de pas vers la droite, suit la loi $\mathcal{B}(n, \frac{1}{2})$. En particulier, la probabilité d'être revenu à l'origine après $2n$ pas vaut

$$\mathrm{P}(S_{2n} = 0) = \mathrm{P}\Big(\sum_{i=1}^{2n} B_i = n\Big) = \binom{2n}{n}\frac{1}{4^n} \underset{n\to+\infty}{\sim} \frac{1}{\sqrt{\pi n}},$$

d'après l'équivalent que le chapitre 10 a tiré de la formule de Stirling et que le chapitre 18 a retrouvé par les intégrales de Wallis ; ces deux chapitres annonçaient qu'il servirait ici. Pour $n = 10$, on trouve 0,1762 contre 0,1784 pour l'équivalent, et pour $n = 100$, 0,056 35 contre 0,056 42. Le retour à l'origine devient rare, mais lentement : la série $\sum \frac{1}{\sqrt{n}}$ diverge (chapitre 20, théorème 5), et l'on peut montrer que le marcheur revient presque sûrement une infinité de fois à l'origine, ce qui dépasse le cadre fini de ce chapitre.

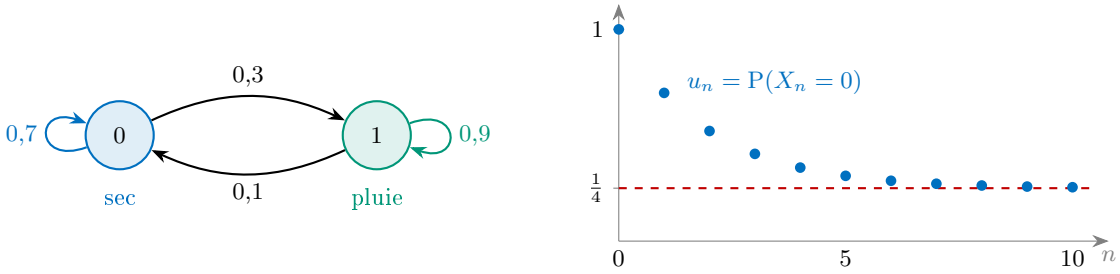
Exemple : Une chaîne de Markov à deux états. Chaque jour est sec, état 0, ou pluvieux, état 1. Un jour sec est suivi d'un jour pluvieux avec probabilité $a = 0{,}3$, un jour pluvieux d'un jour sec avec probabilité $b = 0{,}1$, quel que soit le passé. On modélise la météo des jours 0 à N par des variables X_0, …, X_N à valeurs dans $\{0,1\}$ telles que $\mathrm{P}_{(X_n=0)}(X_{n+1} = 1) = a$ et $\mathrm{P}_{(X_n=1)}(X_{n+1} = 0) = b$, le jour 0 étant sec. La formule des probabilités totales, appliquée au système complet $\big((X_n = 0), (X_n = 1)\big)$, donne pour $u_n = \mathrm{P}(X_n = 0)$

$$u_{n+1} = (1-a)\,u_n + b\,(1-u_n) = 0{,}6\,u_n + 0{,}1.$$

C'est une suite arithmético-géométrique (chapitre 7), de point fixe $\frac{1}{4}$, et $u_n = \frac{1}{4} + \frac{3}{4}\,(0{,}6)^n$. Matriciellement, la ligne $\mu_n = \big(\mathrm{P}(X_n = 0), \mathrm{P}(X_n = 1)\big)$ vérifie

$$\mu_{n+1} = \mu_n M, \qquad \text{où} \qquad M = \begin{pmatrix} 0{,}7 & 0{,}3 \\ 0{,}1 & 0{,}9 \end{pmatrix}$$

est la *matrice de transition*, et donc $\mu_n = \mu_0 M^n$ (chapitre 16).



À gauche, le graphe de la chaîne : chaque flèche porte la probabilité de passer d'un état à l'autre, et les flèches issues d'un même état totalisent 1, comme les branches d'un arbre pondéré. À droite, la probabilité u_n qu'il fasse sec le jour n, en bleu, décroît de 1 vers la valeur $\frac{1}{4}$, en rouge : quel que soit le temps du premier jour, la chaîne oublie son état initial, et passe à long terme un jour sur quatre au sec. La loi $\left(\frac{1}{4}, \frac{3}{4}\right)$, qui vérifie $\mu M = \mu$, est dite *stationnaire*.

Exemple : Une permutation aléatoire. Soit σ une variable aléatoire de loi uniforme sur $\mathfrak{S}_n$, le groupe symétrique du chapitre 17. Pour $i, j \in [\![1\,,n]\!]$, les permutations telles que $\sigma(i) = j$ sont en bijection avec les bijections de $[\![1\,,n]\!] \setminus \{i\}$ sur $[\![1\,,n]\!] \setminus \{j\}$; il y en a $(n-1)!$ par le théorème 1, donc

$$\mathrm{P}\big(\sigma(i) = j\big) = \frac{(n-1)!}{n!} = \frac{1}{n}, \qquad \text{et de même} \qquad \mathrm{P}\big(\sigma(i) = i, \sigma(j) = j\big) = \frac{(n-2)!}{n!} \quad \text{si } i \neq j.$$

Ainsi $\sigma(i)$ suit la loi uniforme sur $[\![1\,,n]\!]$. Les variables $\sigma(1)$ et $\sigma(2)$ ne sont pas indépendantes, puisqu'elles ne sont jamais égales. Enfin, pour $n \geqslant 2$, il y a autant de permutations paires que d'impaires (chapitre 17, remarques qui suivent la proposition 5), et une permutation aléatoire est paire avec probabilité $\frac{1}{2}$.

21.6 Espérance

Huygens évaluait un jeu de hasard par la moyenne des gains possibles, chacun pondéré par sa probabilité : c'est l'espérance. Dans toute la section, les variables aléatoires sont réelles.

21.6.1 Définition et premières propriétés

Définition : L'**espérance** d'une variable aléatoire réelle X est le réel

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

La variable X est dite **centrée** si $\mathrm{E}(X) = 0$.

L'espérance ne dépend que de la loi de X : deux variables de même loi ont même espérance. C'est la moyenne des valeurs de X pondérées par leurs probabilités ; pour un dé équilibré, $\mathrm{E}(X) = \frac{1+2+\cdots+6}{6} = \frac{7}{2}$, valeur que X ne prend jamais. L'espérance n'est donc pas « la valeur que l'on s'attend à observer », mais la moyenne autour de laquelle se répartissent les observations, ce que la section 21.8 précisera.

Proposition 18 : Pour toute variable aléatoire réelle X sur (Ω, P),

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

∘ Démonstration : On regroupe les issues selon la valeur de X : les événements $(X = x)$, pour $x \in X(\Omega)$, forment une partition de Ω (proposition 11), et la relation de Chasles étendue à cette partition, comme dans la démonstration de la proposition 6 du chapitre 3, donne

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

par le théorème 4 appliqué à l'événement $(X = x)$. □

La définition somme sur les valeurs, la proposition 18 sur les issues. La première sert à calculer, quand on connaît la loi ; la seconde sert à démontrer, car elle fait de l'espérance une combinaison linéaire des

valeurs $X(\omega)$, à coefficients fixés. Sous la probabilité uniforme, elle s'écrit $\mathrm{E}(X) = \frac{1}{|\Omega|}\sum_{\omega\in\Omega} X(\omega)$: c'est la moyenne arithmétique des valeurs de X. L'exercice 62 de la feuille du chapitre 16, qui faisait la moyenne des éléments d'un groupe fini de matrices, calculait ainsi l'espérance d'une matrice choisie au hasard dans ce groupe.

Théorème 9 : Propriétés de l'espérance :
Soit X et Y deux variables aléatoires réelles, a, b deux réels et A un événement.

(i) *Linéarité.* $\mathrm{E}(aX+bY) = a\,\mathrm{E}(X) + b\,\mathrm{E}(Y)$.

(ii) *Positivité, croissance.* Si $X \geqslant 0$, alors $\mathrm{E}(X) \geqslant 0$; si $X \leqslant Y$, alors $\mathrm{E}(X) \leqslant \mathrm{E}(Y)$.

(iii) *Inégalité triangulaire.* $|\mathrm{E}(X)| \leqslant \mathrm{E}(|X|)$.

(iv) *Valeurs particulières.* $\mathrm{E}(a) = a$ et $\mathrm{E}(\mathbf{1}_A) = \mathrm{P}(A)$.

(v) *Principe de la moyenne.* $\mathrm{P}\big(X \geqslant \mathrm{E}(X)\big) > 0$ et $\mathrm{P}\big(X \leqslant \mathrm{E}(X)\big) > 0$.

∘ Démonstration :

▷ *(i) à (iv).* Notons $p_\omega = \mathrm{P}(\{\omega\})$. Par la proposition 18 et la linéarité des sommes finies (chapitre 3, proposition 1),

$$\mathrm{E}(aX+bY) = \sum_{\omega\in\Omega} \big(aX(\omega)+bY(\omega)\big)p_\omega = a\,\mathrm{E}(X) + b\,\mathrm{E}(Y).$$

Si $X \geqslant 0$, tous les termes $X(\omega)p_\omega$ sont positifs ; la croissance s'en déduit en appliquant ce résultat à $Y-X$ et en utilisant (i). L'inégalité triangulaire des sommes finies donne (iii). Enfin, $\mathrm{E}(a) = a\sum_\omega p_\omega = a$, et $\mathrm{E}(\mathbf{1}_A) = 0\times\mathrm{P}(\overline{A}) + 1\times\mathrm{P}(A)$.

▷ *(v).* Soit $m = \mathrm{E}(X)$. Si l'on avait $\mathrm{P}(X \geqslant m) = 0$, toutes les issues ω telles que $p_\omega > 0$ vérifieraient $X(\omega) < m$, et comme il en existe au moins une, puisque $\sum_\omega p_\omega = 1$,

$$m = \sum_{\omega,\ p_\omega>0} X(\omega)\,p_\omega < \sum_{\omega,\ p_\omega>0} m\,p_\omega = m,$$

ce qui est absurde. On raisonne de même pour la seconde inégalité. □

Remarques :

- Pour toute variable X, la variable $X - \mathrm{E}(X)$ est centrée, par linéarité et (iv). L'espérance est une forme linéaire sur l'espace vectoriel $\mathbb{R}^\Omega$ des variables aléatoires réelles (chapitre 14), positive et qui vaut 1 sur la constante 1.
- Le point (v) dit qu'une variable prend, avec une probabilité non nulle, une valeur au moins égale à sa moyenne : on prouve ainsi l'*existence* d'un objet en calculant une moyenne, sans le construire. C'était l'argument de l'exercice 50 de la feuille du chapitre 19, que la section suivante relira en termes de probabilités.

Proposition 19 : *(Espérances usuelles.)* Soit $n \in \mathbb{N}^*$ et $p \in [0\,,1]$.

- Si $X \sim \mathcal{U}(\llbracket 1\,, n\rrbracket)$, alors $\mathrm{E}(X) = \dfrac{n+1}{2}$.
- Si $X \sim \mathcal{B}(p)$, alors $\mathrm{E}(X) = p$, et si $X \sim \mathcal{B}(n,p)$, alors $\mathrm{E}(X) = np$.

∘ Démonstration :

▷ *Uniforme et Bernoulli.* On a, dans le premier cas et dans le second,

$$\mathrm{E}(X) = \frac{1}{n}\sum_{k=1}^{n} k = \frac{n+1}{2} \qquad \text{et} \qquad \mathrm{E}(X) = 0\times(1-p) + 1\times p = p.$$

▷ *Binomiale, par le calcul.* Avec la formule du pion $k\binom{n}{k} = n\binom{n-1}{k-1}$ (chapitre 3, proposition 9) et la formule du binôme, en posant $q = 1-p$,

$$\mathrm{E}(X) = \sum_{k=1}^{n} k\binom{n}{k}p^k q^{n-k} = np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}q^{(n-1)-(k-1)} = np\,(p+q)^{n-1} = np.$$

▷ *Binomiale, sans calcul.* L'espérance ne dépend que de la loi. Par la proposition 17, il existe des variables $X_1, \dots, X_n$ mutuellement indépendantes de loi $\mathcal{B}(p)$, dont la somme suit la loi $\mathcal{B}(n,p)$ (théorème 8). Par linéarité, $\mathrm{E}(X) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n) = np$. □

La seconde démonstration est typique : on remplace une variable par une variable de même loi, plus commode, que l'on décompose en somme. Notons que la linéarité n'utilise pas l'indépendance des X_i ; seule l'identification de la loi de la somme en a besoin.

21.6.2 Formule de transfert

Théorème 10 : Formule de transfert :
Soit $X : \Omega \to E$ une variable aléatoire et $f : E \to \mathbb{R}$ une application. Alors

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

En particulier, pour un couple (X,Y) et $f : E \times F \to \mathbb{R}$,

$$\mathrm{E}\big(f(X,Y)\big) = \sum_{(x,y) \in X(\Omega)\times Y(\Omega)} f(x,y)\,\mathrm{P}(X = x, Y = y).$$

∘ Démonstration : Par la proposition 18, puis en regroupant les issues selon la valeur de X, comme dans sa démonstration,

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

Le cas d'un couple est le cas particulier où la variable est (X,Y), à valeurs dans $E \times F$; on somme sur $X(\Omega) \times Y(\Omega)$, qui contient l'image du couple, en ajoutant des termes nuls. Il s'étend de même aux n-uplets. □

La formule de transfert dispense de déterminer la loi de $f(X)$: seule celle de X intervient. Pour un dé, on obtient directement $\mathrm{E}(X^2) = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}$. L'erreur la plus répandue est d'écrire $\mathrm{E}\big(f(X)\big) = f\big(\mathrm{E}(X)\big)$, qui n'est vraie que pour f affine : ici $\mathrm{E}(X)^2 = \frac{49}{4} \neq \frac{91}{6}$.

Proposition 20 : Si X et Y sont deux variables aléatoires réelles indépendantes, alors

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

Plus généralement, si les variables $X_1, \dots, X_n$ sont mutuellement indépendantes, alors

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

∘ Démonstration : La formule de transfert, l'indépendance (proposition 15) et le produit de deux sommes (chapitre 3, proposition 7) donnent

$$\mathrm{E}(XY) = \sum_{(x,y)} xy\,\mathrm{P}(X=x)\,\mathrm{P}(Y=y) = \Big(\sum_{x\in X(\Omega)} x\,\mathrm{P}(X=x)\Big)\Big(\sum_{y\in Y(\Omega)} y\,\mathrm{P}(Y=y)\Big) = \mathrm{E}(X)\,\mathrm{E}(Y).$$

Pour n variables, on raisonne par récurrence : $X_1 \cdots X_{n-1}$ et X_n sont indépendantes par le lemme des coalitions (théorème 7), et la sous-famille $(X_1, \dots, X_{n-1})$ est formée de variables mutuellement indépendantes (proposition 15). □

Remarque : La réciproque est fausse. Si X suit la loi uniforme sur $\{-1,0,1\}$ et $Y = X^2$, alors $\mathrm{E}(XY) = \mathrm{E}(X^3) = 0 = \mathrm{E}(X)\,\mathrm{E}(Y)$, mais X et Y ne sont pas indépendantes, puisque $\mathrm{P}(X=0, Y=1) = 0$ alors que $\mathrm{P}(X=0)\,\mathrm{P}(Y=1) = \frac{1}{3} \times \frac{2}{3} = \frac{2}{9}$. La variable Y est même une fonction de X : la relation $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ ne mesure qu'une absence de dépendance *linéaire*, ce que la section 21.7 appellera décorrélation.

21.6.3 La méthode des indicatrices

♡ **Méthode : Calculer l'espérance d'un nombre d'objets :**

- On écrit la variable qui compte les objets comme une somme d'indicatrices, $N = \mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_m}$, où A_i est l'événement « le i-ième objet possible est présent ».
- On calcule chaque probabilité $\mathrm{P}(A_i)$, souvent toutes égales par symétrie.
- On conclut par linéarité : $\mathrm{E}(N) = \mathrm{P}(A_1) + \cdots + \mathrm{P}(A_m)$, sans avoir déterminé la loi de N et sans aucune hypothèse d'indépendance.

Exemples :

- *Points fixes d'une permutation aléatoire.* Soit σ de loi uniforme sur $\mathfrak{S}_n$ et N son nombre de points fixes. On a $N = \sum_{i=1}^{n} \mathbf{1}_{(\sigma(i)=i)}$, et chaque événement $(\sigma(i) = i)$ a pour probabilité $\frac{1}{n}$ (section 21.5), donc $\mathrm{E}(N) = n \times \frac{1}{n} = 1$. Quel que soit n, une permutation aléatoire a en moyenne un point fixe, alors que la loi de N est compliquée ; une énumération exhaustive le confirme pour $n \leqslant 7$.
- *Records.* On dit que i est un *record* de σ si $\sigma(i) > \sigma(j)$ pour tout $j < i$. Pour $j \leqslant i$, la composition à droite par la transposition $\tau = (j\ i)$ est une bijection $\sigma \mapsto \sigma \circ \tau$ de $\mathfrak{S}_n$, qui préserve donc la loi uniforme, et elle échange les événements « le plus grand des nombres $\sigma(1), \ldots, \sigma(i)$ est atteint en position i » et « …en position j ». Ces i événements forment un système complet de même probabilité : i est un record avec probabilité $\frac{1}{i}$, et le nombre moyen de records vaut
$$\sum_{i=1}^{n} \frac{1}{i} = H_n = \ln n + \gamma + o(1),$$
d'après le chapitre 20. Parmi 100 valeurs rangées dans un ordre aléatoire, on observe en moyenne $H_{100} \approx 5{,}19$ records seulement.
- *Triangles d'un graphe aléatoire.* Sur les sommets $1, \ldots, n$, on trace chacune des $\binom{n}{2}$ arêtes possibles avec probabilité p, indépendamment : on se donne des variables X_e mutuellement indépendantes de loi $\mathcal{B}(p)$, une par paire e de sommets. Le nombre T de triangles est la somme, sur les $\binom{n}{3}$ parties $\{i, j, k\}$, des produits $X_{ij}X_{jk}X_{ik}$, d'espérance p^3 par la proposition 20, puisque les trois arêtes sont distinctes. Ainsi $\mathrm{E}(T) = \binom{n}{3}p^3$; pour $n = 100$, on trouve $\mathrm{E}(T) \approx 0{,}162$ si $p = 0{,}01$, et $\mathrm{E}(T) \approx 20{,}2$ si $p = 0{,}05$.

Remarque : Le premier exemple se prolonge. La formule du crible, démontrée en remarque à la section 21.2 et appliquée aux événements $(\sigma(i) = i)$, donne la probabilité qu'une permutation aléatoire n'ait aucun point fixe : $\sum_{k=0}^{n} \frac{(-1)^k}{k!}$, qui tend vers $\frac{1}{\mathrm{e}} \approx 0{,}368$ d'après la série exponentielle du chapitre 20. Dès $n = 7$, la valeur exacte 0,3679 est indiscernable de la limite : environ une permutation sur trois est sans point fixe, quel que soit n.

21.7 Variance et covariance

L'espérance indique autour de quelle valeur se répartit une variable ; la variance mesure la dispersion de cette répartition. Les variables sont toujours réelles.

21.7.1 Variance et écart type

Définition : La **variance** et l'**écart type** d'une variable aléatoire réelle X sont
$$\mathrm{V}(X) = \mathrm{E}\Big(\big(X - \mathrm{E}(X)\big)^2\Big) \qquad \text{et} \qquad \sigma(X) = \sqrt{\mathrm{V}(X)}.$$
La variable X est dite **réduite** si $\sigma(X) = 1$.

La variance est l'espérance d'une variable positive, donc elle est positive, et l'écart type est bien défini. Il s'exprime dans la même unité que X, ce qui en fait la bonne mesure de l'écart typique à la moyenne. Par la formule de transfert, $\mathrm{V}(X) = \sum_x (x - m)^2\,\mathrm{P}(X = x)$, où $m = \mathrm{E}(X)$.

Théorème 11 : Propriétés de la variance :
Soit X une variable aléatoire réelle et a, b deux réels.

(i) *Formule de König-Huygens.* $\mathrm{V}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$.

(ii) $\mathrm{V}(aX+b) = a^2\,\mathrm{V}(X)$ et $\sigma(aX+b) = |a|\,\sigma(X)$.

(iii) $\mathrm{V}(X) = 0$ si et seulement si $\mathrm{P}\big(X = \mathrm{E}(X)\big) = 1$.

(iv) Si $\sigma(X) > 0$, la variable $\dfrac{X - \mathrm{E}(X)}{\sigma(X)}$ est centrée et réduite.

∘ Démonstration : Notons $m = \mathrm{E}(X)$. Par linéarité de l'espérance,

$$\mathrm{V}(X) = \mathrm{E}(X^2 - 2mX + m^2) = \mathrm{E}(X^2) - 2m^2 + m^2,$$

d'où (i). Pour (ii), $\mathrm{E}(aX+b) = am + b$, donc $\mathrm{V}(aX+b) = \mathrm{E}\big(a^2(X-m)^2\big) = a^2\mathrm{V}(X)$. Pour (iii), la proposition 18 écrit $\mathrm{V}(X) = \sum_\omega \big(X(\omega) - m\big)^2 \mathrm{P}(\{\omega\})$ comme une somme de termes positifs : elle est nulle si et seulement si $X(\omega) = m$ pour toute issue ω de probabilité non nulle, c'est-à-dire si et seulement si $\mathrm{P}(X \neq m) = 0$. Le point (iv) découle de (ii), avec $a = \frac{1}{\sigma(X)}$ et $b = -\frac{m}{\sigma(X)}$. □

Remarques :

- La formule de König-Huygens est la plus commode pour calculer, mais elle soustrait deux nombres souvent voisins. Elle montre aussi que $\mathrm{E}(X)^2 \leqslant \mathrm{E}(X^2)$, ce qui précise la mise en garde qui suit le théorème 10 pour $f(x) = x^2$.
- Le point (ii) dit qu'une translation ne change pas la dispersion, et qu'une homothétie de rapport a la multiplie par $|a|$. L'erreur la plus répandue est d'écrire $\mathrm{V}(aX) = a\mathrm{V}(X)$, ou $\mathrm{V}(-X) = -\mathrm{V}(X)$, qui serait négative.

Proposition 21 : *(Variances usuelles.)* Soit $n \in \mathbb{N}^*$ et $p \in [0\,,1]$.

- Si $X \sim \mathcal{U}(\llbracket 1\,, n \rrbracket)$, alors $\mathrm{V}(X) = \dfrac{n^2-1}{12}$.
- Si $X \sim \mathcal{B}(p)$, alors $\mathrm{V}(X) = p(1-p)$, et si $X \sim \mathcal{B}(n,p)$, alors $\mathrm{V}(X) = np(1-p)$.

∘ Démonstration :

▷ *Uniforme.* Par la formule de transfert et la somme des carrés du chapitre 3, $\mathrm{E}(X^2) = \frac{(n+1)(2n+1)}{6}$, et la formule de König-Huygens donne

$$\mathrm{V}(X) = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}.$$

▷ *Bernoulli.* On a $X^2 = X$, donc $\mathrm{V}(X) = p - p^2$.

▷ *Binomiale.* On calcule d'abord $\mathrm{E}\big(X(X-1)\big)$ par la formule de transfert, en appliquant deux fois la formule du pion, $k(k-1)\binom{n}{k} = n(n-1)\binom{n-2}{k-2}$ pour $k \geqslant 2$, puis la formule du binôme, avec $q = 1-p$:

$$\mathrm{E}\big(X(X-1)\big) = \sum_{k=2}^{n} k(k-1)\binom{n}{k}p^k q^{n-k} = n(n-1)p^2 \sum_{k=2}^{n}\binom{n-2}{k-2}p^{k-2}q^{n-k} = n(n-1)p^2.$$

Ainsi $\mathrm{E}(X^2) = n(n-1)p^2 + np$, et $\mathrm{V}(X) = n(n-1)p^2 + np - n^2p^2 = np(1-p)$. □

Pour un dé, $\mathrm{V}(X) = \frac{35}{12}$ et $\sigma(X) \approx 1{,}71$; pour la loi $\mathcal{B}(10\,;0{,}3)$ de la figure de la section 21.5, $\mathrm{V}(X) = 2{,}1$ et $\sigma(X) \approx 1{,}45$, ce qui donne l'ordre de grandeur de la largeur de l'histogramme autour de son espérance 3.

Proposition 22 : Si X est à valeurs dans $[0\,,1]$ et $m = \mathrm{E}(X)$, alors $\mathrm{V}(X) \leqslant m(1-m) \leqslant \dfrac{1}{4}$, avec égalité dans la première inégalité si et seulement si $\mathrm{P}\big(X \in \{0,1\}\big) = 1$.

∘ Démonstration : Comme $0 \leqslant X \leqslant 1$, on a $X^2 \leqslant X$, donc $\mathrm{E}(X^2) \leqslant m$ par croissance, et König-Huygens donne $\mathrm{V}(X) \leqslant m - m^2$. L'égalité a lieu si et seulement si $\mathrm{E}(X - X^2) = 0$, c'est-à-dire, la variable $X - X^2$ étant positive, si et seulement si elle est nulle pour toute issue de probabilité non nulle, comme dans la démonstration du théorème 11 (iii). Enfin, $m(1-m) = \frac{1}{4} - \left(m - \frac{1}{2}\right)^2 \leqslant \frac{1}{4}$. □

C'est l'analogue probabiliste de l'exercice 49 de la feuille du chapitre 6, qui majorait de même $\int_0^1 (g-m)^2$ pour une fonction continue g à valeurs dans $[0\,,1]$; son corrigé annonçait ce résultat. La variance d'une variable à valeurs dans $[0\,,1]$ est maximale pour une variable de Bernoulli de paramètre $\frac{1}{2}$: la dispersion est la plus grande quand la variable ne prend que les valeurs extrêmes, également probables.

21.7.2 Covariance

Définition : La **covariance** de deux variables aléatoires réelles X et Y est

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

Les variables X et Y sont dites **décorrélées** si $\mathrm{Cov}(X,Y) = 0$.

Proposition 23 : Soit X, Y et Z des variables aléatoires réelles, et a, b deux réels.

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

(ii) La covariance est symétrique et bilinéaire : $\mathrm{Cov}(aX + bY, Z) = a\,\mathrm{Cov}(X,Z) + b\,\mathrm{Cov}(Y,Z)$.

(iii) Si X et Y sont indépendantes, alors elles sont décorrélées.

∘ Démonstration : En développant, puis par linéarité de l'espérance,

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

et $\mathrm{Cov}(X,X) = \mathrm{V}(X)$ par définition. La formule (i) est symétrique, et linéaire en X par linéarité de l'espérance, d'où (ii). Le point (iii) découle de (i) et de la proposition 20. □

La réciproque de (iii) est fausse, comme le montre l'exemple de X uniforme sur $\{-1,0,1\}$ et $Y = X^2$, qui suit la proposition 20. Le signe de la covariance indique si X et Y ont tendance à s'écarter de leurs moyennes dans le même sens ou en sens contraires.

Exemple : Le corrigé de l'exercice 22 de la feuille du chapitre 3 annonçait que deux variables qui varient dans le même sens ont une covariance positive. Soit $a_1 \leqslant \cdots \leqslant a_n$ et $b_1 \leqslant \cdots \leqslant b_n$, et K, K' deux variables indépendantes de loi uniforme sur $[\![1\,,n]\!]$. On pose $X = a_K$, $Y = b_K$, $X' = a_{K'}$ et $Y' = b_{K'}$. Les suites étant croissantes, $(X - X')(Y - Y') \geqslant 0$, puisque $a_K - a_{K'}$ et $b_K - b_{K'}$ ont le signe de $K - K'$. Or (X', Y') a même loi que (X, Y), et X, Y' sont indépendantes, comme fonctions de K et de K' (proposition 16) ; en développant,

$$0 \leqslant \mathrm{E}\big((X - X')(Y - Y')\big) = 2\,\mathrm{E}(XY) - 2\,\mathrm{E}(X)\,\mathrm{E}(Y) = 2\,\mathrm{Cov}(X,Y).$$

C'est l'inégalité de Tchebychev de cet exercice :

$$\frac{1}{n}\sum_{k=1}^{n} a_k b_k \geqslant \Big(\frac{1}{n}\sum_{i=1}^{n} a_i\Big)\Big(\frac{1}{n}\sum_{j=1}^{n} b_j\Big).$$

Pour $a = (1,2,4)$ et $b = (0,3,5)$, on trouve $\mathrm{Cov}(X,Y) = \frac{22}{9}$, qui est bien la somme double de l'exercice divisée par $2n^2$.

Théorème 12 : Inégalité de Cauchy-Schwarz :
Pour toutes variables aléatoires réelles X et Y,

$$\big|\,\mathrm{Cov}(X,Y)\big| \leqslant \sigma(X)\,\sigma(Y).$$

∘ Démonstration :

▷ *Un trinôme positif.* Pour tout réel t, la bilinéarité et la symétrie de la covariance donnent

$$0 \leqslant \mathrm{V}(X+tY) = \mathrm{V}(Y)\,t^2 + 2\,\mathrm{Cov}(X,Y)\,t + \mathrm{V}(X).$$

▷ *Conclusion.* Si $\mathrm{V}(Y) > 0$, ce trinôme positif a un discriminant négatif ou nul :

$$4\,\mathrm{Cov}(X,Y)^2 - 4\mathrm{V}(X)\mathrm{V}(Y) \leqslant 0.$$

Si $\mathrm{V}(Y) = 0$, la fonction affine qui à t associe $2\,\mathrm{Cov}(X,Y)\,t + \mathrm{V}(X)$ est positive sur $\mathbb{R}$, donc constante : $\mathrm{Cov}(X,Y) = 0$, et l'inégalité est vraie. □

Remarques :

- C'est la démonstration du théorème 1 du chapitre 19, et le bilan de ce chapitre annonçait qu'on la retrouverait ici. La covariance n'est pourtant pas un produit scalaire sur l'espace des variables aléatoires : elle est positive, mais pas définie, puisque $\mathrm{V}(X) = 0$ pour toute constante. La démonstration n'utilise que la positivité, ce qui explique qu'elle passe sans changement ; seul le cas $\mathrm{V}(Y) = 0$ demande un argument propre.
- Lorsque $\sigma(X)\sigma(Y) > 0$, le *coefficient de corrélation* $\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$ appartient à $[-1\,,1]$. Si l'on munit un nuage de points (x_1, y_1), …, (x_m, y_m) de la probabilité uniforme, et que X et Y désignent l'abscisse et l'ordonnée d'un point tiré au hasard, la pente de la droite des moindres carrés de la proposition 14 du chapitre 19 s'écrit $a = \frac{\mathrm{Cov}(X,Y)}{\mathrm{V}(X)}$, et ρ mesure l'alignement du nuage.

Théorème 13 : Variance d'une somme :
Pour toutes variables aléatoires réelles X_1, …, X_n,

$$\mathrm{V}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{V}(X_i) + 2 \sum_{1 \leqslant i < j \leqslant n} \mathrm{Cov}(X_i, X_j).$$

Si les X_i sont deux à deux décorrélées, en particulier si elles sont mutuellement indépendantes,

$$\mathrm{V}(X_1 + \cdots + X_n) = \mathrm{V}(X_1) + \cdots + \mathrm{V}(X_n).$$

∘ Démonstration : Par bilinéarité de la covariance, puis en séparant les termes diagonaux de la somme double et en regroupant les termes d'indices (i,j) et (j,i), égaux par symétrie,

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

Les variables mutuellement indépendantes sont deux à deux indépendantes (proposition 15), donc deux à deux décorrélées (proposition 23). □

C'est l'identité qui développe $\|x_1+\cdots+x_n\|^2$ au chapitre 19, où la proposition 3 (ii) traitait le cas de deux vecteurs, et le cas décorrélé est le théorème de Pythagore (théorème 3 du même chapitre). On retrouve ainsi la variance binomiale : si X_1, …, X_n sont indépendantes de loi $\mathcal{B}(p)$, leur somme suit la loi $\mathcal{B}(n,p)$ et a pour variance $np(1-p)$, sans calcul de somme. L'erreur la plus répandue est d'oublier l'hypothèse : $\mathrm{V}(X+X) = 4\mathrm{V}(X)$, et non $2\mathrm{V}(X)$.

Exemples :

- *Points fixes, suite.* Pour σ uniforme sur $\mathfrak{S}_n$, avec $n \geqslant 2$, notons $I_i = \mathbf{1}_{(\sigma(i)=i)}$. Chaque I_i suit la loi $\mathcal{B}(\frac{1}{n})$, et, pour $i \neq j$, la section 21.5 donne $\mathrm{E}(I_iI_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$, d'où $\mathrm{Cov}(I_i, I_j) = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)} > 0$. Le théorème 13 donne

$$\mathrm{V}(N) = n \times \frac{1}{n}\Big(1 - \frac{1}{n}\Big) + n(n-1) \times \frac{1}{n^2(n-1)} = 1 - \frac{1}{n} + \frac{1}{n} = 1.$$

  Le nombre de points fixes a une espérance et une variance égales à 1 pour tout $n \geqslant 2$, ce que confirme l'énumération pour $n \leqslant 7$.

- *Des signes au hasard.* Le corrigé de l'exercice 50 de la feuille du chapitre 19 annonçait qu'on tirerait les signes au hasard. Soit u_1, …, u_n des vecteurs d'un espace préhilbertien et ε_1, …, ε_n des variables mutuellement indépendantes de loi uniforme sur $\{-1,1\}$. Pour $i \neq j$, $\mathrm{E}(\varepsilon_i\varepsilon_j) = \mathrm{E}(\varepsilon_i)\,\mathrm{E}(\varepsilon_j) = 0$, et $\varepsilon_i^2 = 1$; en développant le carré de la norme,

$$\mathrm{E}\Big(\Big\|\sum_{i=1}^n \varepsilon_i u_i\Big\|^2\Big) = \sum_{i,j}\mathrm{E}(\varepsilon_i\varepsilon_j)\langle u_i,u_j\rangle = \sum_{i=1}^n \|u_i\|^2.$$

Par le principe de la moyenne (théorème 9 (v)), il existe des signes pour lesquels la variable $\big\|\sum \varepsilon_i u_i\big\|^2$ dépasse sa moyenne $\sum\|u_i\|^2$, et d'autres pour lesquels elle ne la dépasse pas. Pour $u_1 = (1,0)$ et $u_2 = (1,1)$, les quatre valeurs 5, 1, 1, 5 ont bien pour moyenne $3 = 1+2$.

21.8 Inégalités probabilistes et loi faible des grands nombres

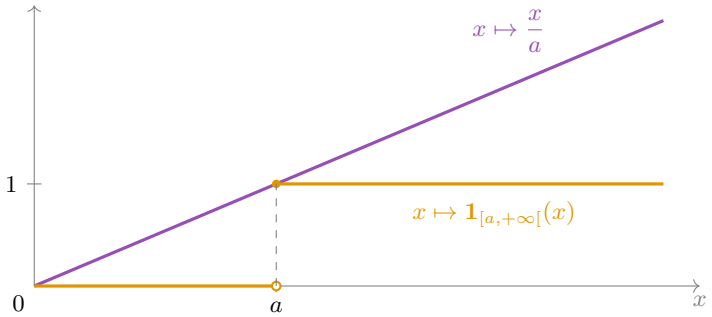
Connaître l'espérance et la variance d'une variable, sans connaître sa loi, suffit à majorer la probabilité qu'elle s'écarte beaucoup de sa moyenne. Ces majorations, grossières mais universelles, conduisent à la loi faible des grands nombres, qui justifie l'interprétation d'une probabilité comme une fréquence.

21.8.1 Inégalité de Markov

Théorème 14 : Inégalité de Markov :
Soit X une variable aléatoire réelle *positive* et $a > 0$. Alors

$$\mathrm{P}(X \geqslant a) \leqslant \frac{\mathrm{E}(X)}{a}.$$

∘ Démonstration : Pour toute issue ω, on a $a\,\mathbf{1}_{(X\geqslant a)}(\omega) \leqslant X(\omega)$: si $X(\omega) \geqslant a$, le premier membre vaut a, et sinon il vaut $0 \leqslant X(\omega)$. Par croissance de l'espérance et le théorème 9 (iv), $a\,\mathrm{P}(X \geqslant a) \leqslant \mathrm{E}(X)$. □



Sur $[0\,,+\infty[$, la fonction indicatrice de $[a\,,+\infty[$, en orange, reste sous la droite $x \mapsto \frac{x}{a}$, en violet. Composées avec X puis intégrées contre sa loi, c'est-à-dire passées à l'espérance, ces deux fonctions donnent $\mathrm{P}(X \geqslant a)$ et $\frac{\mathrm{E}(X)}{a}$. La figure montre aussi où l'inégalité est presque une égalité : lorsque X ne prend que les valeurs 0 et a, là où les deux graphes se touchent.

Exemples :

- Pour un dé, $\mathrm{P}(X \geqslant 5) = \frac{1}{3}$, et l'inégalité de Markov donne seulement $\frac{7/2}{5} = 0{,}7$: elle ignore tout de la loi, hormis la positivité et l'espérance. Elle est pourtant optimale en général : pour $X = a\,\mathbf{1}_A$, les deux membres valent $\mathrm{P}(A)$.
- *La méthode du premier moment.* Le nombre T de triangles du graphe aléatoire de la section 21.6 est un entier positif, et l'inégalité de Markov avec $a = 1$ donne $\mathrm{P}(T \geqslant 1) \leqslant \mathrm{E}(T)$. Pour $n = 100$ sommets et $p = 0{,}01$, le graphe ne contient aucun triangle avec une probabilité au moins égale à $1 - 0{,}162 = 0{,}838$, sans que l'on ait déterminé la loi de T.

21.8.2 Inégalité de Bienaymé-Tchebychev

Théorème 15 : Inégalité de Bienaymé-Tchebychev :
Soit X une variable aléatoire réelle et $\varepsilon > 0$. Alors

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

∘ Démonstration : Les événements $\big(|X-\mathrm{E}(X)| \geqslant \varepsilon\big)$ et $\Big(\big(X-\mathrm{E}(X)\big)^2 \geqslant \varepsilon^2\Big)$ sont égaux, et l'on applique l'inégalité de Markov à la variable positive $\big(X-\mathrm{E}(X)\big)^2$, d'espérance $\mathrm{V}(X)$, avec $a = \varepsilon^2$. □

Avec $\varepsilon = k\,\sigma(X)$, l'inégalité s'écrit $\mathrm{P}\big(|X - \mathrm{E}(X)| \geqslant k\sigma(X)\big) \leqslant \frac{1}{k^2}$: quelle que soit sa loi, une variable s'écarte de sa moyenne de plus de deux écarts types avec une probabilité au plus $\frac{1}{4}$, et de plus de dix écarts types avec une probabilité au plus $\frac{1}{100}$. C'est en ce sens que l'écart type mesure la dispersion.

Exemple : Le nombre S de piles en 100 lancers d'une pièce équilibrée suit la loi $\mathcal{B}(100, \frac{1}{2})$, d'espérance 50 et de variance 25. L'inégalité de Bienaymé-Tchebychev donne

$$\mathrm{P}(|S - 50| \geqslant 10) \leqslant \frac{25}{10^2} = 0{,}25,$$

alors que la valeur exacte, calculée à partir de la loi binomiale, vaut 0,0569. La majoration est quatre fois trop grande, mais elle a été obtenue sans aucun calcul de loi.

21.8.3 Loi faible des grands nombres

Théorème 16 : Loi faible des grands nombres :
Soit $X_1, \ldots, X_n$ des variables aléatoires réelles mutuellement indépendantes, de même loi, d'espérance m et de variance σ^2. On note $\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$ leur moyenne. Pour tout $\varepsilon > 0$,

$$\mathrm{P}\big(|\overline{X}_n - m| \geqslant \varepsilon\big) \leqslant \frac{\sigma^2}{n\varepsilon^2}.$$

En particulier, si $S_n \sim \mathcal{B}(n,p)$, alors $\mathrm{P}\Big(\Big|\dfrac{S_n}{n} - p\Big| \geqslant \varepsilon\Big) \leqslant \dfrac{p(1-p)}{n\varepsilon^2} \leqslant \dfrac{1}{4n\varepsilon^2}$.

∘ Démonstration : Par linéarité, $\mathrm{E}(\overline{X}_n) = \frac{1}{n} \times nm = m$. Les X_i étant indépendantes, le théorème 13 et le théorème 11 (ii) donnent

$$\mathrm{V}(\overline{X}_n) = \frac{1}{n^2}\big(\mathrm{V}(X_1) + \cdots + \mathrm{V}(X_n)\big) = \frac{\sigma^2}{n},$$

et l'inégalité de Bienaymé-Tchebychev appliquée à $\overline{X}_n$ conclut. Pour le cas particulier, la variable $\frac{S_n}{n}$ a même loi que la moyenne de n variables indépendantes de loi $\mathcal{B}(p)$ (proposition 17 et théorème 8), de variance $p(1-p) \leqslant \frac{1}{4}$ (proposition 22), et la probabilité cherchée ne dépend que de cette loi. □

Remarques :

- L'énoncé est *non asymptotique* : il donne, pour chaque n, une majoration explicite. Il contient la forme asymptotique, puisque le majorant tend vers 0 : pour tout $\varepsilon > 0$,

  $$\mathrm{P}\big(|\overline{X}_n - m| \geqslant \varepsilon\big) \xrightarrow[n\to+\infty]{} 0.$$

  C'est le théorème d'or de Jacques Bernoulli, évoqué dans l'introduction, pour des variables de Bernoulli.

- *Interprétation fréquentiste.* Si l'on répète n fois, de façon indépendante, une expérience dans laquelle un événement A a la probabilité p, la fréquence d'apparition de A suit la loi de $\frac{S_n}{n}$, avec $S_n \sim \mathcal{B}(n,p)$. Le théorème dit qu'elle est proche de p avec une grande probabilité lorsque n est grand : la probabilité, définie ici comme une mesure abstraite, se lit sur les fréquences observées. C'est ce qui justifie l'estimation d'une probabilité inconnue par une fréquence, et la simulation.

- Pour garantir, avec une probabilité au moins 0,95, que la fréquence observée est à moins de 0,01 de p, il suffit que $\frac{1}{4n\varepsilon^2} \leqslant 0{,}05$ avec $\varepsilon = 0{,}01$, c'est-à-dire $n \geqslant 50\,000$, quel que soit p. Ce nombre est très pessimiste, et la sous-section suivante l'améliore.

21.8.4 Une inégalité de concentration

L'inégalité de Markov appliquée à une variable mieux choisie que $\big(X-\mathrm{E}(X)\big)^2$ donne des majorations bien plus fines. On le montre sur les lancers d'une pièce équilibrée.

Proposition 24 : Si $S_n \sim \mathcal{B}(n,\frac{1}{2})$, alors, pour tout $\varepsilon>0$,

$$\mathrm{P}\Big(\Big|\frac{S_n}{n}-\frac{1}{2}\Big|\geqslant\varepsilon\Big)\leqslant 2\,\mathrm{e}^{-2n\varepsilon^2}.$$

∘ Démonstration :

▷ *Réduction.* La probabilité ne dépend que de la loi de S_n. On se donne donc des variables X_1, …, X_n mutuellement indépendantes de loi uniforme sur $\{-1,1\}$ (proposition 17) ; comme pour la marche aléatoire de la section 21.5, $\frac{1}{2}(X_1+\cdots+X_n)+\frac{n}{2}$ suit la loi $\mathcal{B}(n,\frac{1}{2})$, et l'événement étudié s'écrit $\big(|Z|\geqslant n\varepsilon\big)$, avec $Z=\frac{1}{2}(X_1+\cdots+X_n)$.

▷ *Markov exponentiel.* Soit $t>0$. La fonction exponentielle étant strictement croissante, les événements $(Z\geqslant n\varepsilon)$ et $\big(\mathrm{e}^{tZ}\geqslant \mathrm{e}^{tn\varepsilon}\big)$ sont égaux, et l'inégalité de Markov appliquée à la variable positive e^{tZ} donne

$$\mathrm{P}(Z\geqslant n\varepsilon)\leqslant \mathrm{e}^{-tn\varepsilon}\,\mathrm{E}\big(\mathrm{e}^{tZ}\big).$$

Les variables $\mathrm{e}^{tX_i/2}$ sont mutuellement indépendantes (proposition 16), d'espérance $\frac{\mathrm{e}^{t/2}+\mathrm{e}^{-t/2}}{2}=\mathrm{ch}\,\frac{t}{2}$, et la proposition 20 donne

$$\mathrm{E}\big(\mathrm{e}^{tZ}\big)=\mathrm{E}\Big(\prod_{i=1}^{n}\mathrm{e}^{tX_i/2}\Big)=\Big(\mathrm{ch}\,\frac{t}{2}\Big)^n.$$

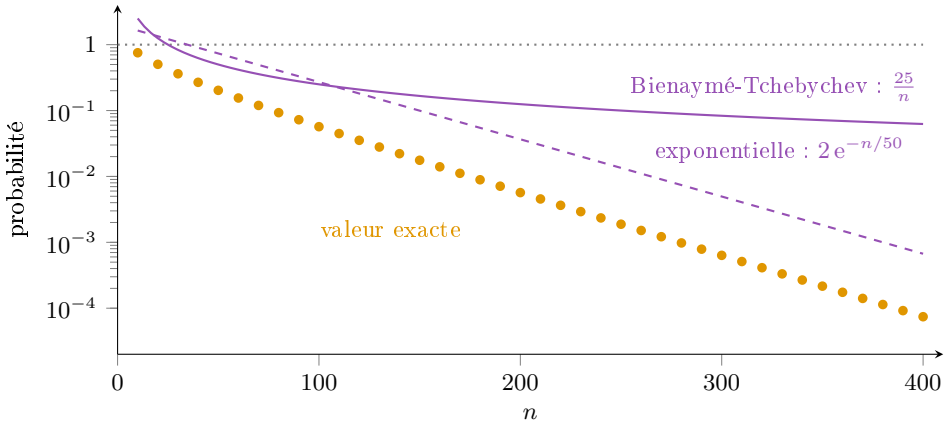
▷ *Majoration du cosinus hyperbolique.* Pour tout réel u, la série exponentielle (chapitre 20, théorème 2) donne, les termes impairs se compensant,

$$\mathrm{ch}\,u=\frac{\mathrm{e}^u+\mathrm{e}^{-u}}{2}=\sum_{k=0}^{+\infty}\frac{u^{2k}}{(2k)!}.$$

Comme $(2k)!=\prod_{j=1}^{k}(2j-1)(2j)\geqslant\prod_{j=1}^{k}2j=2^k k!$, on obtient, par comparaison terme à terme des sommes partielles,

$$\mathrm{ch}\,u\leqslant\sum_{k=0}^{+\infty}\frac{(u^2/2)^k}{k!}=\mathrm{e}^{u^2/2}.$$

▷ *Choix de t.* Ainsi $\mathrm{P}(Z\geqslant n\varepsilon)\leqslant\exp\big(-tn\varepsilon+\frac{nt^2}{8}\big)$ pour tout $t>0$. Le trinôme $\frac{nt^2}{8}-tn\varepsilon$ est minimal en $t=4\varepsilon$, où il vaut $-2n\varepsilon^2$: d'où $\mathrm{P}(Z\geqslant n\varepsilon)\leqslant\mathrm{e}^{-2n\varepsilon^2}$. Enfin, le n-uplet $(-X_1,\dots,-X_n)$ a même loi que $(X_1,\dots,X_n)$, puisque sa loi conjointe est le produit des mêmes lois marginales (proposition 15) ; donc $-Z\sim Z$ par la proposition 13, et $\mathrm{P}(Z\leqslant -n\varepsilon)=\mathrm{P}(-Z\geqslant n\varepsilon)\leqslant\mathrm{e}^{-2n\varepsilon^2}$. La sous-additivité (proposition 7 (iv)) conclut. □



La figure, à échelle logarithmique en ordonnée, compare pour $\varepsilon=0{,}1$ la probabilité exacte de l'événement $\big(|\frac{S_n}{n}-\frac{1}{2}|\geqslant 0{,}1\big)$, en orange, à ses deux majorants, en violet. La borne de Bienaymé-Tchebychev $\frac{25}{n}$ ne décroît que comme $\frac{1}{n}$; la borne exponentielle, en pointillés, décroît comme la valeur exacte, en ligne droite sur cette échelle, et devient la meilleure à partir de $n=108$. Pour $n=400$, on lit $7{,}4\times10^{-5}$, $6{,}7\times10^{-4}$ et 0,0625.

Remarque : Avec la proposition 24, il suffit que $2\,\mathrm{e}^{-2n\varepsilon^2} \leqslant 0{,}05$, soit $n \geqslant \frac{\ln 40}{2\times 10^{-4}} \approx 18\,444$ lancers, pour que la fréquence des piles soit à moins de 0,01 de $\frac{1}{2}$ avec probabilité au moins 0,95, au lieu des 50 000 lancers demandés par Bienaymé-Tchebychev. Appliquée à des variables indépendantes et bornées quelconques, la même méthode conduit à l'inégalité de Hoeffding, hors programme, première des *inégalités de concentration*.

21.9 Bilan

Notion	Énoncé	Référence
cardinaux	$\lvert A\cup B\rvert = \lvert A\rvert + \lvert B\rvert - \lvert A\cap B\rvert$; $\lvert F^E\rvert = n^p$; $\lvert\mathcal{P}(E)\rvert = 2^p$	propositions 1 à 3
listes et parties	$A_n^p = \frac{n!}{(n-p)!}$ listes, $n!$ permutations, $\binom{n}{p}$ parties ; Pascal et binôme par dénombrement	théorèmes 1 à 3, propositions 4 à 6
probabilité	déterminée par les $\mathrm{P}(\{\omega\})$; uniforme : $\mathrm{P}(A) = \frac{\lvert A\rvert}{\lvert\Omega\rvert}$	proposition 7, théorème 4
conditionnement	probabilités composées, totales ; formule de Bayes	propositions 8 et 9, théorèmes 5 et 6
indépendance	$\mathrm{P}(A\cap B) = \mathrm{P}(A)\mathrm{P}(B)$; loi conjointe produit des marginales ; coalitions	propositions 10, 15 à 17, théorème 7
lois	loi de $f(X)$; $\mathcal{U}(E)$, $\mathcal{B}(p)$, $\mathcal{B}(n,p)$; somme de Bernoulli indépendantes	propositions 11 à 14, théorème 8
espérance	linéarité, transfert ; $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ si $X \perp\!\!\!\perp Y$; $\mathrm{E}(\mathcal{B}(n,p)) = np$	propositions 18 à 20, théorèmes 9 et 10
variance	$\mathrm{V}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$; $\lvert\mathrm{Cov}(X,Y)\rvert \leqslant \sigma(X)\sigma(Y)$; variance d'une somme	propositions 21 à 23, théorèmes 11 à 13
inégalités	Markov, Bienaymé-Tchebychev ; $\mathrm{P}(\lvert\overline{X}_n - m\rvert \geqslant \varepsilon) \leqslant \frac{\sigma^2}{n\varepsilon^2}$	théorèmes 14 à 16, proposition 24

Trois idées sont à retenir. Dénombrer, c'est trouver la bonne bijection ou le bon découpage, en se demandant toujours si l'ordre compte et si les répétitions sont permises ; sous la probabilité uniforme, toute probabilité est un rapport de cardinaux. Ensuite, on raisonne sur des variables aléatoires et leurs lois plutôt que sur l'univers : l'indépendance se lit sur la loi conjointe, et la linéarité de l'espérance, sans aucune indépendance, calcule des moyennes dont la loi est inaccessible. Enfin, la variance, qui se comporte comme le carré d'une norme, contrôle les écarts à la moyenne par l'inégalité de Bienaymé-Tchebychev, et la loi faible des grands nombres en déduit que les fréquences observées approchent les probabilités. Le chapitre 22 étudiera les fonctions de deux variables réelles. En seconde année, les probabilités s'étendront aux univers infinis, où les sommes finies deviendront des sommes de familles sommables (chapitre 20) : lois géométriques et de Poisson, événement « une infinité des E_n se réalisent » de l'exercice 32 de la feuille du chapitre 1.