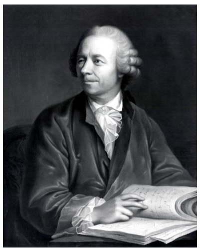


Cours Euler

Module 1

Combinatoire, probabilités et statistiques

C. Mermoud

2024-2025

I. Combinatoire

Nous commençons avec ce sujet le premier module de cette année consacré à la combinatoire, aux probabilités et aux statistiques. Pour le moment, nous voulons développer des méthodes et des formules de comptage d'événements.
Combien y a-t-il de tirages possibles au loto ?
Combien y a-t-il de manières différentes de ranger les livres d'une bibliothèque ?
Combien de possibilités différentes y a-t-il au tournoi de Roland-Garros d'arriver à la finale ?
C'est à ces questions, et à bien d'autres encore, que nous voulons répondre !

1 Le principe fondamental de la combinatoire

Comment dénombrer les résultats possibles d'un certain nombre d'expériences indépendantes ?

Exemple 1.1. L'entraîneur de l'équipe mixte de patinage artistique doit choisir un homme parmi les 7 que compte l'équipe suisse et une femme parmi les 5 que compte son équipe.
Combien de couples différents peut-il former ?

Chacun des 7 patineurs peut former un couple avec chacune de 5 patineuses $\Rightarrow$ $7 \cdot 5 = 35$ couples possibles.

Soit X un ensemble fini. On note $|X|$ le nombre d'éléments de X, aussi appelé sa *cardinalité*. Souvent on aime ordonner les éléments d'un ensemble de cardinalité k et pour nommer chaque élément, on écrira parfois $X = \{x_1, \ldots, x_k\}$.

Exemple 1.2. Si X est l'ensemble des mois de l'année, alors $|X| = 12$. On peut écrire $X =$\{janvier, février, mars, avril, mai, juin, juillet, août, septembre, octobre, novembre, décembre\}, ou encore si l'on ne s'intéresse qu'au comptage $X = \{m_1, m_2, \ldots, m_{12}\}$, où m_j est le j^{e} mois de l'année.

Lorsque X et Y sont deux sous-ensembles d'un même ensemble Z, on peut former leur union $X \cup Y$ et leur intersection $X \cap Y$ par

$$X \cup Y = \{z \in Z \mid z \in X \text{ ou } z \in Y\} \quad \text{et} \quad X \cap Y = \{z \in Z \mid z \in X \text{ et } z \in Y\}$$

Lemme 1.3. *On a* $|X \cup Y| = |X| + |Y| - |X \cap Y|$.

Démonstration. Notons $z_1, \dots z_k$ les k éléments de $X \cap Y$.
Alors X est formé de ces éléments et de quelques autres : $X = \{z_1, \dots z_k, x_1, \dots x_n\}$
De même, $Y = \{z_1, \dots, z_k, y_1, \dots y_m\}$
Par suite, $X \cup Y = \{z_1, \dots, z_k, x_1, \dots, x_n, y_1, \dots, y_m\}$.
Ainsi, $|X| = k+n$, $|Y| = k+m$, $|X \cup Y| = k+m+n$ et $|X \cap Y| = k$ □

Si l'intersection de X et Y est vide, on dit que l'union $X \cup Y$ est *disjointe* et on note $X \cup Y = X \coprod Y$. Ainsi la cardinalité d'une union disjointe est la somme des cardinalités :

$$|X \coprod Y| = |X| + |Y|.$$

Théorème 1.4. *Soient X et Y des ensembles finis. Alors* $|X \times Y| = |X| \cdot |Y|$.

Démonstration. L'ensemble produit $X \times Y$ est l'ensemble de tous les couples (x, y) avec $x \in X$ et $y \in Y$. Ainsi, si on fait varier y, on peut décomposer le produit en une union disjointe

$$X \times Y = \coprod_{y \in Y} X \times \{y\}.$$

Chaque ensemble $X \times \{y\}$ est de cardinalité $|X|$ puisqu'il existe une bijection entre $X \times \{y\}$ et X pour tout y. Le résultat précédent montre que $|X \times Y| = \sum_{y \in Y} |X \times \{y\}| = |X| \cdot |Y|$. □

C'est ce résultat assez anodin qui mérite le nom de principe fondamental de la combinatoire ! En voici la formulation combinatoire.

Théorème 1.5. *Soit k un nombre entier ≥ 2. Supposons que l'on réalise k expériences indépendantes l'une de l'autre. Si la j-ème a n_j résultats possibles pour $1 \leq j \leq k$, alors il existe $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ résultats différents pour les k expériences prises ensemble.*

Démonstration. On effectue la preuve par récurrence sur k. Les cas $k = 1$ et $k = 2$ sont connus. Supposons donc que $k \geq 3$ et que le cas de $k - 1$ expériences est vrai.

On décompose les k expériences en deux "blocs", le premier étant formé des $k-1$ premières expériences et le deuxième de la dernière expérience.
Par hypothèse de récurrence, le premier bloc admet $n_1 \cdot n_2 \cdot \ldots \cdot n_{k-1}$ résultats.
Les deux blocs constituent un cas où $k = 2$. Ainsi, il y a finalement $(n_1 \cdot n_2 \cdot \ldots \cdot n_{k-1}) \cdot n_k$ résultats différents. □

Exemple 1.6. On lance en même temps un dé traditionnel à six faces numérotées de 1 à 6 et un dé octogonal dont les faces sont gravées d'une lettre majuscule A, B, C, D, E, F, G, H.
Combien y a-t-il de lancers différents ?

On peut écrire dans un tableau à double entrée les résultats possibles de tous les lancers :

$1A$	$2A$	$3A$	$4A$	$5A$	$6A$
$1B$	$2B$	$3B$	$4B$	$5B$	$6B$
$1C$	$2C$	$3C$	$4C$	$5C$	$6C$
$1D$	$2D$	$3D$	$4D$	$5D$	$6D$
$1E$	$2E$	$3E$	$4E$	$5E$	$6E$
$1F$	$2F$	$3F$	$4F$	$5F$	$6F$
$1G$	$2G$	$3G$	$4G$	$5G$	$6G$
$1H$	$2H$	$3H$	$4H$	$5H$	$6H$

Il y a donc $6 \cdot 8 = 48$ lancers différents.

Exemple 1.7. Combien y a-t-il de nombres à deux chiffres ?

Un nombre à deux chiffres s'écrit "xy" avec $x \in \{1, \dots 9\}$ et $y \in \{0; \dots 9\}$

Il y a donc $9 \cdot 10 = 90$ nombres à deux chiffres.

Exemple 1.8.
Combien y a-t-il de nombres à trois chiffres <u>tous distincts</u> qui se terminent par un 3 ou par un 7 ?

centaine dizaine unité
$\neq 0$ ____ 3 ou 7

On choisit d'abord les éléments contraints :

$$\Rightarrow \underset{\text{unité}}{2} \cdot \underset{\text{centaine}}{8} \cdot \underset{\text{dizaine}}{8} = 128 \text{ possibilités.}$$

Les notions de permutation, arrangement et combinaison permettent de dénombrer une grande classe de situations différentes. Nous en établissons maintenant les méthodes générales.

2 Permutations

Avant de définir ce qu'est une permutation, commençons avec un exemple.
Combien de mots différents de trois lettres peut-on former avec les trois lettres A, D et N.

$\square \ \square \ \square$

$3 \cdot 2 \cdot 1 = 3! = 6$ possibilités, à savoir :

ADN, AND, DAN, DNA, NAD, NDA.

Définition 2.1. Soit X un ensemble fini de cardinalité n.
Une *permutation* de X est une disposition ordonnée $(x_1, x_2, \dots, x_n)$ de tous les éléments de X.

Théorème 2.2. *Le nombre de permutations d'un ensemble X fini de cardinalité n vaut*

$$P_n = n!$$

Démonstration. On démontre le résultat par induction sur n.
Si $n = 1$, il y a un seul élément x dans X. Ainsi la seule permutation est (x).
Si $n > 1$, notre hypothèse d'induction est qu'il y a $(n-1)!$ permutations d'un ensemble de cardinalité $n-1$. Une permutation de X est déterminée par le choix du premier élément, appelons-le x, puis d'une permutation de $X - \{x\}$.
Le principe fondamental de la combinatoire nous dit qu'il y a $n \cdot (n-1)! = n!$ résultats possibles. □

Exemple 2.3. On joue au loto en tirant dans un sac des jetons numérotés de 1 à 45.
Combien de tirages différents y a-t-il ?

Il y en a $45! \cong 10^{56}$.

Parfois, on effectue des expériences où les répétitions sont autorisées. Imaginons l'expérience suivante. On place dans un sac opaque 3 boules noires parfaitement identiques et 2 boules blanches pareilles. On sort du sac une boule après l'autre.
Combien y a-t-il de tirages, étant donné que les boules d'une même couleur sont indistinguables ?
Si l'on pouvait les distinguer, on aurait bien sûr 5! tirages différentes. Mais pensons par exemple au tirage "noir - blanc - noir- noir- blanc", que l'on notera $NBNNB$.
Si l'on numérote les boules noires N_1, N_2, N_3, on voit que les tirages

$$N_1BN_2N_3B,\ N_1BN_3N_2B,\ N_2BN_1N_3B,\ N_2BN_3N_1B,\ N_3BN_1N_2B,\ N_3BN_2N_1B$$

sont indistinguables. Il y en a $3! = 6$ correspondant aux permutations des boules noires.
De même, il y a dans chaque cas de figure 2 tirages indistinguables dû au fait que l'on tire deux boules blanches. Par conséquent, il y a

$$\frac{5!}{3!\ 2!} = \frac{5 \cdot 4 \cdot \cancel{3 \cdot 2 \cdot 1}}{\cancel{3 \cdot 2 \cdot 1} \cdot 2 \cdot 1} = \frac{20}{2} = 10 \text{ tirages différents.}$$

Il s'agit de

$$NNNBB,\ NNBNB,\ NNBBN,\ NBNNB,\ NBNBN,$$
$$NBBNN,\ BNNNB,\ BNNBN,\ BNBNN,\ BBNNN.$$

Définition 2.4. Soit X un ensemble de k éléments $x_1, \dots, x_k$. Une *permutation de n objets avec $r_1, r_2, \dots, r_k$ répétitions* de X est un n-uplet d'éléments de X où x_i apparaît exactement r_i fois.

En particulier, on a $r_1 + \dots + r_k = n$. On peut aussi penser à ces permutations avec répétitions de la manière suivante. Soit X un ensemble de n éléments dans lequel il y a des sous-ensembles de $r_1, \dots, r_k$ objets identiques. Une permutation avec répétitions de X est une permutation de X où l'ordre des objets indistingables n'est pas pris en compte.

Théorème 2.5. *Le nombre de permutations de n avec $r_1, r_2, \dots, r_k$ répétitions est*

$$\overline{P}_n(r_1; \dots; r_k) = \frac{n!}{r_1! \dots r_k!}.$$

Démonstration. Ce résultat reviendra dans un prochain cours, où nous le démontrerons. □

Exemple 2.6. On veut classer 6 livres de Jules Verne dans un rayon de librairie : 2 exemplaires de "20000 lieues sous les mers", 3 exemplaires de "5 semaines en ballon" et 1 seul du "Tour du monde en 80 jours". De combien de manières peut-on le faire ?

Il y a $\dfrac{6!}{3!\ 2!\ 1!} = 60$ possibilités.

3 Arrangements et combinaisons

Définition 3.1. Soit X un ensemble de n éléments distincts. Un *arrangement* de r éléments de X est une disposition ordonnée de r éléments de X. On note A_r^n le nombre de ces arrangements.

Exemple 3.2. Trente-deux équipes participent à la coupe du monde de football.
Combien de podiums possibles y a-t-il ?

Le même raisonnement permet de calculer le nombre d'arrangements.

Théorème 3.3. $$A_r^n = n \cdot (n-1) \cdot \dots \cdot (n-r+1) = \frac{n!}{(n-r)!}.$$

Exemple 3.4. Mélissa, 3 ans, possède 3 animaux en peluche : un ours, un kangourou et une souris. Elle aime les placer sur un canapé à 5 places. De combien de manières différentes peut-elle le faire ?

Définition 3.5. Soit X un ensemble de n éléments distincts et $r \leq n$.
Un *arrangement avec répétitions* de r éléments de X est une disposition ordonnée de r éléments non nécessairement distincts de X. On note $\overline{A_r^n}$ le nombre de ces arrangements avec répétitions.

Théorème 3.6. $\overline{A_r^n} = n \cdot n \cdot \dots \cdot n = n^r$.

Exemple 3.7. Combien de mots de trois lettres (de a à z) peut-on former ?

Définition 3.8. Soit X un ensemble de n éléments distincts. Une *combinaison* de r éléments de X est une disposition non ordonnée de r éléments de X. On note C_r^n le nombre de ces combinaisons.

Exemple 3.9. Au SwissLoto, on tire 5 nombres entre 1 et 45. Combien y a-t-il de tirages différents ?

Il y a A_5^{45} arrangements, mais puisqu'on ne tient pas compte de l'ordre des éléments (le tirage $5, 32, 40, 10, 3$ est le même que $40, 3, 32, 5, 10$), il faut encore . Ainsi, il y a

$$\frac{\qquad}{\qquad} = 1'221'759 \quad \text{tirages de loto différents.}$$

Chaque joueur devrait être conscient qu'il a moins d'une chance sur un million de gagner ! Cela fait moins de $0,0001\%$...

Théorème 3.10. $C_r^n = \begin{pmatrix} n \\ r \end{pmatrix} = \dfrac{n!}{(n-r)! \cdot r!}$.

Exemple 3.11.
Dans une classe de 10 élèves, combien a-t-on de choix de 3 personnes pour organiser une fête ?

Le nombre de combinaisons C_r^n est aussi appelé *coefficient binomial* pour des raisons que nous allons voir tout de suite. La proposition suivante donne une relation entre les combinaisons de n éléments et celles de $n-1$ éléments.

Proposition 3.12. *On a* $\begin{pmatrix} n \\ r \end{pmatrix} = \begin{pmatrix} n-1 \\ r \end{pmatrix} + \begin{pmatrix} n-1 \\ r-1 \end{pmatrix}$ *pour* $1 \leq r \leq n-1$.

Démonstration. Voici un argument combinatoire pour démontrer cette égalité.
Soit X un ensemble de n objets et x l'un d'entre eux. Les combinaisons de r objets peuvent être classées en deux groupes, celles qui contiennent x et celles qui ne le contiennent pas.
Ainsi, il y a $\begin{pmatrix} n-1 \\ r-1 \end{pmatrix}$ combinaisons qui contiennent x car il faut alors former une combinaison de $r-1$ éléments de $X - \{x\}$ et il y a $\begin{pmatrix} n-1 \\ r \end{pmatrix}$ combinaisons qui ne contiennent pas x car il s'agit des combinaisons de r éléments de $X - \{x\}$. □

Exemple 3.13. Un relais radiophonique est mis en place à l'aide de n antennes, dont m sont défectueuses. Le relais ne fonctionne que si deux antennes défectueuses ne sont jamais côte à côte. Combien peut-on trouver de configurations qui fonctionnent ?

Imaginons les $n-m$ antennes qui fonctionnent, alignées comme ceci :

$$- \; F \; - \; F \; - \; F \; \ldots \; - \; F \; - \; F \; -$$

où F indique une antenne qui fonctionne et $-$ une place pour une éventuelle antenne défectueuse.
Il y a donc positions pour installer une telle antenne et on doit en choisir .
La réponse est donc $\begin{pmatrix} \quad \\ \quad \end{pmatrix}$.

Le nombre de combinaisons C_k^n est aussi appelé *coefficient binomial* pour des raisons que nous allons voir maintenant.
La notion de combinaison permet en effet d'établir une formule particulièrement utile :

Théorème 3.14 (Binôme de Newton)**.** *On a* $$(x+y)^n = \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} x^k y^{n-k}.$$

Démonstration. Considérons le produit

$$(x_1 + y_1)(x_2 + y_2) \ldots (x_n + y_n).$$

Lorsque l'on développe cette expression, on obtient 2^n termes, chaque terme étant un produit de n facteurs x_i ou y_j. De plus, chacun de ces termes contient soit x_i, soit y_i pour tout $1 \leq i \leq n$. Il nous faut maintenant compter le nombre de termes qui contiennent exactement k facteurs x_i et $(n-k)$ facteurs y_j. Ces facteurs correspondent au choix de "k parmi n" (des combinaisons de k éléments dans un ensemble de n objets), c'est-à-dire $\begin{pmatrix} n \\ k \end{pmatrix}$.
On obtient donc la formule en posant $x_i = x$ et $y_i = y$ pour tout i. □

Remarque 3.15. Un peu d'histoire. Isaac Newton (1643-1727) est un philosophe, mathématicien, physicien et astronome anglais.

Figure emblématique des sciences, il est surtout reconnu pour sa théorie de la gravitation universelle et la création, en concurrence avec Leibniz, du calcul infinitésimal. Il est aussi connu pour la généralisation du théorème du binôme et l'invention dite de la méthode de Newton permettant de trouver des approximations d'un zéro (ou racine) d'une fonction d'une variable réelle à valeurs réelles.

II. Probabilités

Vous avez vu l'année passée le début de la combinatoire, avec les permutations, les arrangements et les combinaisons. Nous terminons aujourd'hui le sujet de combinatoire avec quelques formules de comptage et commençons le calcul des probabilités, intimement lié à la théorie des ensembles...

1 Les coefficients multinomiaux

Pour terminer le chapitre consacré à la combinatoire, nous généralisons la notion de coefficient binomial.

Définition 1.1. Soit $n_1, \ldots, n_r$ des nombres entiers et $n = n_1 + \cdots + n_r$.
Le *coefficient multinomial* est

$$\binom{n}{n_1, \ldots, n_r} = \frac{n!}{n_1! \cdot \ldots \cdot n_r!}.$$

Lorsque $r = 2$, on voit que $\binom{n}{n_1, n_2}$ est simplement

Ce dernier compte le nombre de combinaisons possibles de n_1 objets choisis dans un ensemble de n objets. Ou encore le nombre de manières différentes de répartir n objets en

Mais que compte-t-on en général avec des coefficients multinomiaux ?

Théorème 1.2. *Le nombre de répartitions possibles de n objets distinguables en r groupes de $n_1, n_2, \ldots, n_r$ objets est égal au coefficient multinomial*

$$\binom{n}{n_1, \ldots, n_r} = \frac{n!}{n_1! \cdot \ldots \cdot n_r!}\,.$$

Démonstration. Procédons par récurrence. Lorsque $r = 2$, c'est la cas binomial connu.
Supposons que $r > 2$ et que le cas $r - 1$ est vrai.
Pour répartir n objets en r groupes de $n_1, \ldots, n_r$ éléments,

$\square$

Exemple 1.3. Dans un camp, 14 élèves doivent choisir l'une des activités suivantes : canoé, escalade et spéléologie. Sachant qu'il y a trois canoés de 2 places, dont une est occupée par le moniteur, 4 baudriers disponibles, et 5 lampes de poches, combien de répartitions y a-t-il ?

Pour terminer, voici le *Théorème multinomial*, qui généralise celui du binôme de Newton.
La preuve est élémentaire.

Théorème 1.4. *On a* $\quad (x_1 + x_2 + \cdots + x_r)^n \; = \sum_{(n_1,\ldots,n_r)} \begin{pmatrix} n \\ n_1, \ldots, n_r \end{pmatrix} x_1^{n_1} x_2^{n_2} \ldots x_r^{n_r}.$

2 Répartitions des boules dans des urnes

Voici une application des principes combinatoires que vous avez vus l'année passée.
Le problème est de répartir n boules identiques dans r urnes.

Exemple 2.1. On cherche à répartir 8 boules noires dans 3 urnes. On peut par exemple en placer 3 dans la première urne, puis une dans la deuxième, et 4 dans la dernière.
Pour compter le nombre de répartitions possibles, imaginons que nous alignions les huit boules comme ceci :

$$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet$$

Il faut former trois groupes.

Plus généralement, on a le résultat suivant.

Théorème 2.2. *Le nombre de répartition de n boules identiques en r urnes distinctes est*

$$\binom{n-1}{r-1}$$

si chaque urne doit contenir au moins une boule et que $r \geq n$.

Et si l'on autorise des urnes vides ? Dans la résolution précédente, on pourra donc placer plusieurs barres verticales au même endroit, y compris avant la première boule ou après la dernière.

Théorème 2.3. *Le nombre de répartition de n boules identiques dans r urnes distinctes est*

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

si l'on peut laisser certaines urnes vides.

Démonstration. Cette fois, il s'agit simplement d'aligner n boules et $r-1$ séparateurs :

Il y a donc $=$ $=$ possibilités. □

Exemple 2.4. Monsieur et Madame Aubert ont fait quelques économies et pensent que le moment est venu d'investir 20000 francs en bourse. Ils peuvent placer un certain nombre de milliers de francs en actions suisses, en actions de la zone Euro, en action américaines, ou encore en obligations. Combien de stratégies se présentent ?

3 Les événements

Une *expérience aléatoire* est une expérience dont on connait à l'avance les issues possibles, bien que l'on ne sache pas laquelle sera finalement réalisée.

On appelle *ensemble fondamental* l'ensemble de toutes les issues possibles d'une expérience aléatoire. On le note S.

Les éléments de S et tout sous-ensemble de S sont des *événements*.

Exemple 3.1. Si on lance une pièce de monnaie, l'ensemble fondamental $S = \{P, F\}$, où P est l'abréviation de pile et F celle de face.

Si l'expérience consiste à lancer deux dés distincts, alors il y a $6 \cdot 6 = 36$ issues possibles et

$$S = \{(i; j) \,|\, 1 \leq i, j \leq 6\}.$$

Les issues $\{(3; 6)\}$ ou $\{(i; i) \,|\, 1 \leq i \leq 6\}$ sont des événements ; le second est l'événement “les deux dés indiquent le même résultat”.

Dans ces deux exemples, l'ensemble fondamental est fini, mais il se peut aussi qu'il soit infini.

Dans le cas où l'expérience consiste à mesurer la durée de vie en heures d'une ampoule économique, on a $S = \mathbb{R}_+ = [0, \infty[$. Ici $E = [0; 24]$ est l'événement “l'ampoule dure moins d'un jour”.

On utilisera les notations usuelles de la théorie des ensembles pour effectuer des opérations sur des événements. Ainsi, si E et F sont deux événements, l'événement E^c est le *complémentaire* de E, c'est-à-dire l'événement où l'issue est n'importe quel résultat qui ne se trouve pas dans E. L'*intersection* $E \cap F$, aussi notée EF, est l'événement où l'issue est à la fois dans E et dans F. Les lois de De Morgan sont des principes élémentaires, mais fort utiles !

Proposition 3.2. Lois de De Morgan
Soient $E_1, \ldots, E_n$ des événements. Alors

$$\left(\cup_{i=1}^n E_i\right)^c = \cap_{i=1}^n E_i^c \text{ et } \left(\cap_{i=1}^n E_i\right)^c = \cup_{i=1}^n E_i^c.$$

Démonstration. Pour démontrer la première loi, nous allons voir que l'on a la "double inclusion" :

$$\left(\cup_{i=1}^n E_i\right)^c \subset \cap_{i=1}^n E_i^c \quad \text{et} \quad \cap_{i=1}^n E_i^c \subset \left(\cup_{i=1}^n E_i\right)^c.$$

Pour la première inclusion, considérons un élément $x \in \left(\cup_{i=1}^n E_i\right)^c$.

La deuxième inclusion se fait en exercice dans la série 2.
Pour démontrer la seconde loi de Morgan, définissons $F_i = E_i^c$.

$\square$

Remarque 3.3. Un peu d'histoire. Auguste De Morgan (1806 - 1871) est un mathématicien et logicien britannique, né en Inde. Il est le fondateur avec Boole de la logique moderne.

4 Les axiomes des probabilités

Vous avez certainement toutes et tous une idée intuitive de ce qu'est une probabilité.
Par exemple, la probabilité d'obtenir 6 lors d'un lancé de dé équilibré vaut .
Si on tire 13 cartes dans un jeu de 36 cartes, la probabilité qu'il y ait exactement trois rois dans le lot tiré vaut

Lorsque l'ensemble fondamental S est fini et constitué d'issues équiprobables et dénombrables, la probabilité $P(A)$ d'un événement A se calcule avec la formule

$$P(A) =$$

Si les issues ne sont pas équiprobables ou que l'ensemble fondamental est infini ou de nature continue, cette formule ne suffit plus. Nous allons donc adopter une approche plus conceptuelle.

Andreï Kolmogorov (1903-1983), mathématicien russe, a posé les axiomes qui portent son nom et qui définissent formellement ce qu'est une probabilité. Les voici.

Définition 4.1. Soit S l'ensemble fondamental d'une expérience.
Une *probabilité* est une application $P : \mathcal{P}(S) \to [0, 1]$ telle que :

1. $P(S) = 1$
2. Si les événements $E_1, E_2, E_3, \ldots$ sont disjoints, alors $P\left(\coprod_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$.

Le nombre réel $P(E)$ est appelé *probabilité* de l'événement E.

On attribue donc un nombre compris entre 0 et 1 à chaque événement. Plus ce nombre est grand, plus la probabilité de cet événement est grande. C'est pourquoi le premier axiome dit simplement que l'issue de l'expérience est à coup sûr (avec une probabilité égale à 1) l'un des éléments de S. Le deuxième axiome dit que la probabilité de l'union d'une infinité d'événements mutuellement exclusifs est la somme des probabilités de chacun d'eux.

Exemple 4.2. On lance un dé comme l'un de ceux que tu vois sur cette photo truquée :

Il se trouve que le coin commun aux faces 1, 2 et 3 est plombé si bien qu'on obtient deux fois plus souvent un grand nombre qu'un petit nombre.
Quelle est la probabilité d'obtenir un nombre pair ?

Déduisons à présent quelques propriétés élémentaires qui découlent des deux axiomes de la définition 4.1. La première de ces propriétés affirme que la probabilité de l'événement "vide" est nulle car il y a toujours une issue à toute expérience aléatoire.

Proposition 4.3. *On a* $P(\emptyset) = 0.$

Démonstration. Considérons les événements mutuellement exclusifs $E_1 = S$, $E_2 = \emptyset$, $E_3 = \emptyset$, etc. Alors nous savons que

□

Ceci nous permet dans un premier temps de montrer que l'axiome (2) est vrai aussi si on traite le cas d'une union disjointe *finie* d'événements. Il suffit d'appliquer le deuxième axiome en ajoutant un nombre infini d'événements vides.

Théorème 4.4. *Soit* $E_1, \ldots, E_n$ *des événements. Alors* $P\left(\coprod_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i).$

Le théorème suivant nous dit que la probabilité qu'un événement n'arrive pas vaut 1 moins la probabilité que cet événement arrive.

Théorème 4.5. *Soit* E *un événement. Alors* $P(E^c) = 1 - P(E).$

Démonstration.

□

Si F est un événement contenu dans E, alors E est plus probable que F.

Théorème 4.6. *Soient* $F \subset E$ *deux événements. Alors* $P(F) \leq P(E).$

Démonstration.
Puisque F est contenu dans E, on peut exprimer E comme union disjointe de F et de $F^c \cap E$. Par conséquent, $P(E) = P(F) + P(F^c \cap E)$ ce qui montre le théorème. □

Voici peut-être la formule élémentaire la plus utile en pratique :

Théorème 4.7. *Soient E et F deux événements. Alors*

$$P(E \cup F) = P(E) + P(F) - P(E \cap F).$$

Démonstration.

□

Exemple 4.8. On tire 13 cartes d'un jeu de 52.
Déterminer la probabilité qu'il y ait exactement 3 rois ou exactement 2 dames dans le lot.

Exemple 4.9. Quelle est la probabilité que deux élèves d'un cours Euler aient leur anniversaire le même jour s'il y a 25 élèves dans la classe ?

III. Probabilités conditionnelles

Le concept de probabilité conditionnelle est l'un des plus importants de cette théorie puisque l'on cherche souvent à savoir quelle est la probabilité d'un événement alors qu'on dispose d'une information partielle. La probabilité qu'il neige un matin n'est pas la même si l'on sait qu'il a fait 30 degrés la veille ou -5 degrés.

1 Rappels sur les probabilités

L'ensemble fondamental d'une expérience est l'ensemble de toutes les issues possibles d'une expérience.

Définition 1.1. Soit S l'ensemble fondamental d'une expérience.
Une *probabilité* est une application $P : \mathcal{P}(S) \to [0,1]$ telle que :

(1) $P(S) = 1$;

(2) Si les événements $E_1, E_2, E_3, \ldots$ sont disjoints, alors $P\left(\coprod_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$.

Le nombre réel $P(E)$ est appelé *probabilité* de l'événement E.

Nous avons vu la semaine passée que

a) $P(\emptyset) = 0$, c'est-à-dire que la probabilité que rien ne se passe est nulle, notre expérience a toujours une issue ;

b) $P\left(\coprod_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$, donc l'axiome (2) est vrai aussi si on traite le cas d'une union disjointe *finie* d'événements ;

c) $P(E^c) = 1 - P(E)$;

d) si $F \subset E$ des événement, alors $P(F) \leq P(E)$.

Exemple 1.2. On lance plusieurs fois de suite une paire de dés équilibrés et on s'intéresse à la somme des chiffres obtenus.
Quelle est la probabilité d'obtenir une somme égale à 5 avant une somme égale à 7 ?

Soit E_n l'événement "Pendant les $n-1$ premières épreuves ni 5 ni 7 ne sont obtenus, puis à la n-ème on obtient 5". La probabilité cherchée est la probabilité de la réunion disjointe $\coprod E_n$.

La probabilité d'obtenir un total de 5 vaut .
En effet,

De même, celle d'obtenir un total de 7 vaut .
Ainsi, la probabilité d'obtenir un total de 5 ou de 7 est .

Ainsi, puisque les lancers sont indépendants,

$$P(E_n) =$$

Par conséquent, on a

Nous avons terminé le cours de la semaine passée sur le résultat suivant :

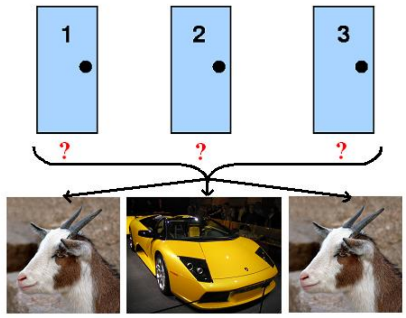
Théorème 1.3. *Soient E et F deux événements. Alors*

$$P(E \cup F) = P(E) + P(F) - P(E \cap F).$$

2 Probabilité conditionnelle

Commençons par un exemple élémentaire, que nous reprendrons par la suite pour expliquer le "paradoxe" du problème de Monty Hall.

Exemple 2.1. Supposez que vous êtes sur le plateau d'un jeu télévisé face à trois portes et que vous devez choisir d'en ouvrir une seule, en sachant que derrière l'une d'elles se trouve une voiture et derrière les deux autres des chèvres.



La probabilité de choisir la bonne porte vaut donc

Supposons maintenant que l'on sache, *avant de choisir*, qu'une chèvre se cache derrière la troisième porte. Dans ce cas, la probabilité que la voiture soit derrière la première porte vaut

On peut aussi raisonner comme suit : la probabilité que la voiture soit derrière la première porte et que simultanément une chèvre se trouve derrière la troisième porte vaut car si la voiture est derrière la première porte alors forcément, il y a une chèvre derrière la troisième porte.
Par ailleurs, la probabilité qu'une chèvre se trouve derrière la troisième porte vaut si bien que si l'on sait qu'une chèvre se trouve derrière la troisième porte, l'ensemble fondamental est réduit au de son état initial.
Ainsi, la probabilité que la voiture soit derrière la première porte sachant qu'une chèvre est derrière la troisième porte vaut

Définition 2.2. Soient E et F deux événements. On suppose que $P(F) > 0$. Alors, la *probabilité conditionnelle* $P(E|F)$ de E sachant que F est réalisé est

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Voici un exemple simple où l'on pourrait aussi raisonner directement.

Exemple 2.3. Une urne se trouve dans une pièce obscure et contient 10 billes rouges, 5 billes jaunes et 10 billes blanches lumineuses. On tire une boule et constate qu'elle n'est pas lumineuse. Quelle est la probabilité qu'elle soit jaune ?
Intuitivement :

Mathématiquement, avec la définition 2.2 :
Soit J l'événement produit par le tirage d'une bille jaune et B celui produit par le tirage d'une bille blanche.

Enfin, revenons au problème de Monty Hall.
Le candidat est placé devant trois portes fermées. Derrière l'une d'elles se trouve une voiture et derrière chacune des deux autres se trouve une chèvre. Il doit tout d'abord désigner une porte. Puis le présentateur, qui sait quelle est la bonne porte dès le début, doit ouvrir une porte qui n'est ni celle choisie par le candidat, ni celle cachant la voiture. Le candidat a alors le droit soit d'ouvrir la porte qu'il a choisie initialement, soit d'ouvrir la troisième porte.
Quelle stratégie doit-il adopter pour maximiser la probabilité de gagner la voiture ?

Exemple 2.4. Supposons que l'on choisisse la première porte.
Appelons A_i l'événement "la voiture se trouve derrière la i-ème porte", pour $1 \leq i \leq 3$.
Appelons X_j l'événement "le présentateur ouvre la j-ème porte" pour $2 \leq j \leq 3$.
Calculons toutes les probabilités $P(A_i \cap X_j)$.

D'abord, remarquons que pour tout i, $P(A_i) =$

- $P(A_1 \cap X_2) = P(A_1 \cap X_3) =$

- $P(A_2 \cap X_2) = P(A_3 \cap X_3) =$

- $P(A_2 \cap X_3) = P(A_3 \cap X_2) =$

Ceci nous permet aussi de calculer la probabilité que le présentateur ouvre la troisième porte.

Il faut changer de porte pour avoir le plus de chance de gagner la voiture !

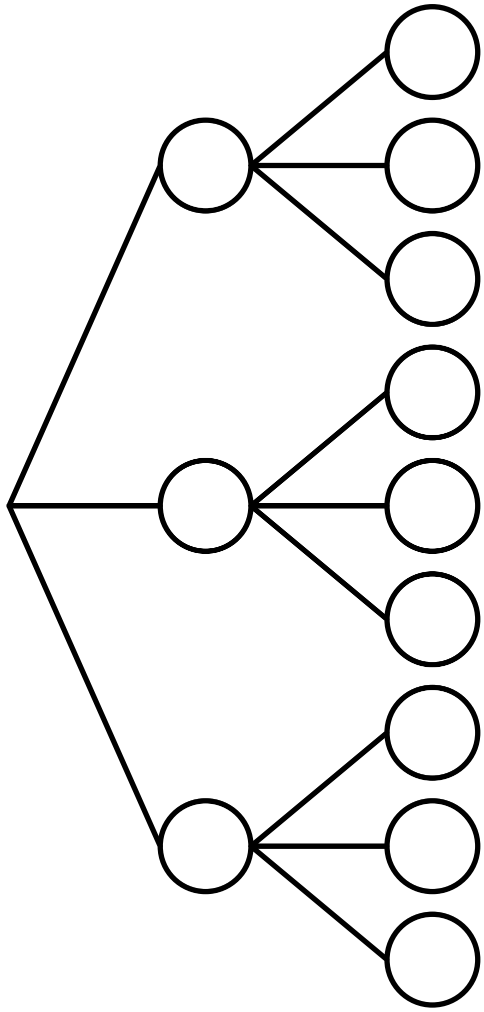
3 La formule des probabilités totales

Dans la section précédente, nous avons vu comment calculer une probabilité conditionnelle. Parfois il peut être utile de retourner cette formule et de l'appliquer pour calculer la probabilité d'un événement. Nous commençons par la formule de multiplication.

Proposition 3.1. Formule de multiplication
Soient A et B deux événements. Alors $P(A \cap B) = P(A) \cdot P(B|A)$.

Exemple 3.2. Un sac contient quatre billes rouges, deux billes bleues et trois billes vertes.
On cherche la probabilité des événements suivants : A : "les deux billes tirées sont rouges",
B : "la première est bleue et la seconde est verte" et C : "l'une des billes est rouge et l'autre bleue".

On désigne par R_1 l'événement "la première bille tirée est rouge", par R_2 "la seconde est rouge", par B_1 "la première est bleue", etc.
On dessine un diagramme en arbre où la *racine* indique le début de l'expérience, le premier niveau de branches indique le premier tirage, le second niveau le deuxième tirage.
On écrit la probabilité de chaque tirage sur la branche et la couleur tirée aux noeuds.



Le théorème suivant est une simple adaptation de cette formule.

Théorème 3.3. des probabilités totales
Soit $E_1, \dots, E_n$ *des événements tels que* $S = E_1 \coprod E_2 \coprod \cdots \coprod E_n$. *Alors pour tout événement* A,

$$P(A) = \sum_{i=1}^{n} P(A|E_i)P(E_i).$$

Démonstration. Montrons simplement le cas de deux événements : $S = E \coprod F$.

□

Exemple 3.4. Un inspecteur de police est convaincu à 60% que le suspect principal de son enquête sur le vol d'un tableau de Picasso est coupable. A ce stade de l'enquête on trouve un cheveu blond et court sur la scène du crime. Il se trouve que le suspect est blond ! Quelle est la probabilité qu'il ait volé le tableau sachant que 20% de la population a des cheveux blonds ?

Soit C l'événement "le suspect est coupable" et B "le coupable est blond".
Alors, avant d'avoir trouvé le cheveu blond, par le théorème des probabilités totales, on a

$P(B) = P(B|C)P(C) + P(B|C^c)P(C^c) =$

D'autre part, la probabilité conditionnelle que le suspect est coupable sachant que le coupable est blond se calcule directement avec la définition :

Dans cet exemple, nous avons utilisé le *Théorème de Bayes*, du nom de son auteur Thomas Bayes (1702–1761), mathématicien et pasteur britannique.

Théorème 3.5. Théorème de Bayes
Soit $E_1, \dots, E_n$ des événements tels que $S = E_1 \coprod E_2 \coprod \cdots \coprod E_n$.
Alors pour tout événement A de probabilité non nulle et tout k, on a

$$P(E_k|A) = \frac{P(A|E_k)P(E_k)}{\sum_{i=1}^n P(A|E_i)P(E_i)}.$$

Pour terminer ce cours, étudions un exemple de nature théorique.

Exemple 3.6. Problème de rencontre de Mont-mort (1708)
Lors d'une réunion de n hommes, chacun enlève son chapeau et le lance dans le vestibule. À la fin de la réunion, chacun prend un chapeau au hasard dans le tas. On dit qu'il y a *rencontre* si quelqu'un tire son propre chapeau. Quelle est la probabilité qu'il n'y ait aucune rencontre ?

Soit E_n l'événement "il n'y a aucune rencontre". L'idée est de conditionner $P(E_n)$ selon l'événement R : "le premier homme tire son propre chapeau".
Par la formule des probabilités totales, on a

Analysons maintenant $P(E_n|R^c)$. C'est la probabilité qu'il n'y ait pas de rencontre lorsque $n-1$ hommes tirent un chapeau dans un tas de $n-1$ chapeaux, tas dans lequel un chapeau n'appartient à personne, celui du premier homme, et l'un manque, celui que le premier homme a tiré.

Il y a deux cas de figure sans rencontre. Soit l'homme en trop ne tire pas le chapeau en trop, soit il le tire. S'il ne le tire pas, faisons comme si ce chapeau lui appartenait, si bien que cette situation est équivalente à E_{n-1}. Mais s'il le tire, et il y a une chance sur $n-1$ qu'il le fasse, il reste $n-2$ hommes qui doivent tirer un chapeau dans un tas constitué de leurs chapeaux :

Nous avons obtenu une formule de récurrence !
On peut commencer à calculer puisque $P(E_1) = 0$ et $P(E_2) = \dfrac{1}{2}$.
Il se trouve que la formule donne

une série qui tend vers $e^{-1} \cong 0,368$.

IV. Variables aléatoires

Aujourd'hui nous terminons la partie consacrée aux probabilités conditionnelles en parlant d'indépendance entre différents événements, puis nous étudions les variables aléatoires, ce qui nous permet de passer finalement aux statistiques dans le dernier chapitre du module.

1 Indépendance de deux événements

Nous avons vu que parfois la probabilité d'un événement est affectée par la réalisation d'un autre événement. Par exemple, le fait de tirer un as dans un paquet de cartes rend la probabilité d'en tirer un autre ensuite plus faible. Par contre, obtenir un six lors d'un lancer de dé n'affecte pas la probabilité d'en obtenir un autre au lancer suivant.

Définition 1.1. Les événements E et F sont *indépendants* si $P(E \cap F) = P(E)P(F)$.

En d'autres termes, on a

L'information partielle que F s'est réalisé n'influence pas la réalisation de E et vice-versa.

Exemple 1.2. Voici les résultats simplifiés d'une enquête sur la législation de la marijuana aux USA. On interroge des habitants sur tout le territoire des Etats-Unis.
On compte 30% de la population sur la côte Est.
La proportion de personnes sondées pour la dépénalisation et habitant sur la côte Est est de $7,8\%$.
Il y a $18,2\%$ de personnes en faveur de la dépénalisation et habitant dans les autres régions.

Les événements D : "être en faveur de la dépénalisation" et E : "vivre sur la côte Est" sont-ils dépendants ?

Que se passe-t-il lorsque l'on veut parler de l'indépendance de plus de deux événements ?
Il faut faire très attention avant de tirer des conclusions hâtives.
En effet, si A et B sont indépendants et que B et C le sont aussi, que dire de A et C ?

Par ailleurs, si A et B sont indépendants et que A et C le sont aussi, alors A et $B \cap C$ ... ne sont en général pas indépendants !

Exemple 1.3. On dispose d'un gros dé à six faces qu'on colorie de la manière suivante.
Les faces 1, 2, 3 et 4 sont blanches, les faces 5 et 6 sont bleues ;
les nombres 1, 2, 3 et 6 sont écrits à l'encre rouge, les nombres 4 et 5 à l'encre noire.
On considère les événements A : "obtenir un nombre pair", R : "obtenir un nombre rouge" et B : "obtenir une face blanche".

Un autre exemple surprenant (pour toi aussi je l'espère) sera analysé dans la série d'exercices. Ceci nous pousse à définir la notion d'indépendance totale.

Définition 1.4. Les événements E, F et G sont *totalement indépendants* s'ils sont indépendants deux à deux et de plus $P(E \cap F \cap G) = P(E)P(F)P(G)$.

Explicitement on demande donc que

$P(E \cap F) =$

$P(E \cap G) =$

$P(F \cap G) =$

et $P(E \cap F \cap G) =$

Exemple 1.5. *ou plutôt un contre-exemple :*
On lance deux pièces de monnaie et on considère les événements
A = "pile au 1er lancé", B = "face au 2ème lancé", C = "même résultat au deux lancés".

2 Variables aléatoires

Après avoir réalisé une expérience, il arrive qu'on s'intéresse plus à une fonction du résultat qu'au résultat lui-même. Par exemple, lors d'une élection au Conseil fédéral, peut-être que certains voudront seulement savoir combien de femmes ont été élues. S'il s'agit des élèves sélectionnés au cours Euler, ce sera le nombre d'élèves Genevois, ou alors le nombre d'élèves en 11ème année, etc.

Définition 2.1. Toute fonction réelle $X : S \to \mathbb{R}$ définie sur l'ensemble fondamental S d'une expérience est appelée *variable aléatoire*.

Exemple 2.2. On jette trois pièces de monnaie équilibrées et on définit X par le nombre de piles obtenus. Ainsi X est une variable aléatoire pouvant prendre les valeurs 0, 1, 2 ou 3.
La probabilité de chacun de ces événements est :

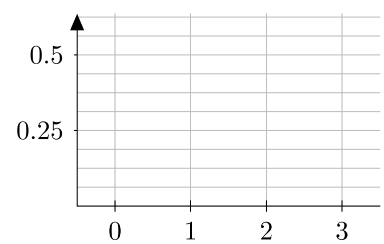
$P\{X = 0\} =$

$P\{X = 1\} =$

$P\{X = 2\} =$

$P\{X = 3\} =$

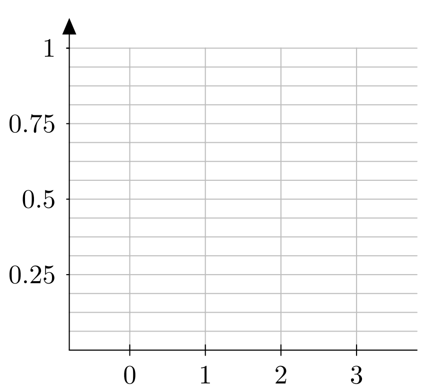
Représentation graphique de la distribution de X



Définition 2.3. La *fonction de répartition* d'une variable aléatoire X est la fonction réelle

$$F(b) = P\{X \leq b\}.$$

La fonction de répartition de l'exemple précédent se représente graphiquement comme suit :



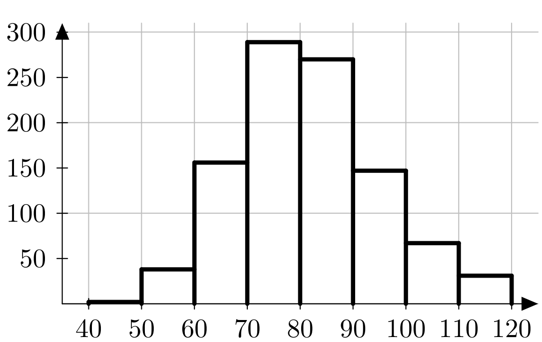
Nous étudierons principalement des variables aléatoires discrètes, qui ne prennent au plus qu'une quantité dénombrable de valeurs, mais il en existe d'autres, dites continues. Par exemple, les mesure, taille ou durée sont vues comme des grandeurs réelles plutôt qu'entières ou rationnelles.

Les concepts de fonction de distribution et de répartition se retrouvent aussi dans le domaine de la statistique comme le montre l'exemple ci-dessous.

Exemple 2.4. Le tableau suivant donne les résultats d'une étude du poids d'Australiens.
Le poids étant une mesure de type continu, les données ont été regroupés en classes

Poids X en kg	*Effectif*	*Pourcentage*	*Pourcentage cumulé*
$X < 50$	2	0.2%	
$50 \leqslant X < 60$	38	3.8%	
$60 \leqslant X < 70$	156	15.6%	
$70 \leqslant X < 80$	289	28.9%	
$80 \leqslant X < 90$	270	27.0%	
$90 \leqslant X < 100$	147	14.7%	
$100 \leqslant X < 110$	67	6.7%	
$110 \leqslant X$	31	3.1%	
Total	1000	100%	

Histogramme de distribution des poids.

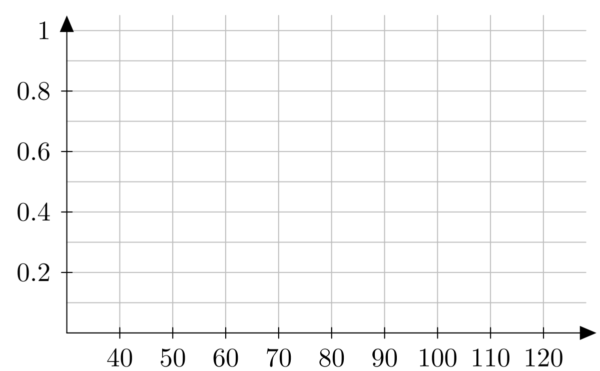


On peut alors considérer X comme une variable aléatoire continue.

Si on tire une personne au hasard dans l'échantillon, quelle est la probabilité que son poids soit compris entre 60 et 70 kg ?

En calculant les pourcentages cumulés des personnes ayant un poids inférieur à x kilos, on peut tracer la fonction de répartition de X.

Si on tire une personne au hasard dans l'échantillon, quelle est la probabilité son poids soit inférieur à 80 kg ?



Proposition 2.5. *Soit X une variable aléatoire et F sa fonction de répartition. Alors*

$$P\{a < X \leq b\} = F(b) - F(a) \text{ si } a < b.$$

Démonstration. Décrivons l'événement $\{X \leq b\}$ comme réunion disjointe des deux événements $\{X \leq a\}$ et $\{a < X \leq b\}$. Ainsi

□

Exemple 2.6. La roue de la fortune

On fait tourner une triple roue qui s'immobilise en laissant apparaître trois nombres compris entre 1 et 6. Le joueur mise un franc et choisit un nombre. Il gagne k fois sa mise si le nombre apparaît k fois. Il perd sa mise si le nombre n'apparaît pas.

Ce jeu est-il équitable ? Etudions la variable aléatoire $X =$ "gain net du joueur".

3 Espérance

L'espérance d'une variable aléatoire est la moyenne pondérée des valeurs possibles de cette variable. Elle mesure la valeur qu'on peut espérer obtenir en "moyenne" si on la répète un grand nombre de fois l'expérience aléatoire.

Définition 3.1. Soit X une variable aléatoire discrète. L'*espérance* de X, si elle existe, est le nombre réel

$$E[X] = \sum_{a \in \mathbb{R}} a\, P\{X = a\}.$$

Si le nombre de valeurs a pour lesquels $P\{X = a\} \neq 0$ est infini, l'espérance peut ne pas exister.

Exemple 3.2. Quelle est l'espérance du résultat du lancer d'un dé bien équilibré ?

En "moyenne" on obtient donc $3,5$, ce qui veut dire qu'on peut espérer obtenir 350 points si l'on additionne les résultats de 100 lancers.

Exemple 3.3. Quelle est l'espérance du gain net dans l'exemple 2.6 de la roue de la fortune ?

Il arrive aussi qu'on cherche à connaître l'espérance non pas d'une variable aléatoire X, mais d'une fonction de celle-ci.

Exemple 3.4. On tire au hasard un mot dans la phrase "TU VAS PAYER". Tu dois prédire en combien de lettres s'écrit le mot tiré et tu devras payer le carré de la différence entre la longueur prédite et la vraie longueur. Ainsi si tu prédis 5, tu ne devra rien payer si le mot "PAYER" est choisi, mais $(5-2)^2 = 9$ francs si le mot "TU" est choisi.
Quelle longueur faut-il prédire pour payer le moins possible en moyenne ?

Soit X la variable aléatoire donnant la différence entre la longueur prédite et la longueur du mot tiré. Si t est la longueur prédite, tu as donc une chance sur trois de devoir payer $(t-2)^2$, une sur trois de payer $(t-3)^2$ et enfin une sur trois de devoir payer $(t-5)^2$. L'espérance est donc

Il faut donc choisir $\dfrac{10}{3}$ pour espérer payer seulement $1,56$ francs en moyenne.

Théorème 3.5. *Soit X une variable aléatoire discrète et $g(x)$ une fonction réelle. Alors*

$$E[g(X)] = \sum_{a \in \mathbb{R}} g(a) P\{X = a\}.$$

Démonstration. Il s'agit en fait simplement de la définition de l'espérance appliquée à la variable aléatoire $g(X)$. Il faut remarquer néanmoins que $P\{g(X) = b\} = \sum_{a | g(a) = b} P\{X = a\}$. □

D'autre part l'espérance est linéaire :

Théorème 3.6. *Soient X, Y deux variables aléatoires discrètes. Alors $E[X + Y] = E[X] + E[Y]$.*

Démonstration. On vérifie cela en calculant

$$\begin{aligned} E[X+Y] &= \sum_c cP\{X+Y=c\} \\ &= \sum_c c \left(\sum_a P\{X=a, Y=c-a\} \right) \\ &= \sum_c \sum_a (a + (c-a)) P\{X=a, Y=c-a\} \\ &= \sum_c \sum_a aP\{X=a, Y=c-a\} + \sum_c \sum_a (c-a) P\{X=a, Y=c-a\} \end{aligned}$$

On calcule la première partie comme suit :

$$\begin{aligned} \sum_c \sum_a a \cdot P\{X=a, Y=c-a\} &= \sum_a a \cdot P\left\{ \coprod_c (\{X=a\} \cap \{Y=c-a\}) \right\} \\ &= \sum_a a \cdot P\{X=a\} = E[X] \end{aligned}$$

De même, la deuxième partie devient, avec le changement de variable $b = c - a$:

$$\begin{aligned} \sum_b \sum_a b \cdot P\{X=a, Y=b\} &= \sum_b b \cdot P\left\{ \coprod_a (\{X=a\} \cap \{Y=b\}) \right\} \\ &= \sum_b b \cdot P\{Y=b\} = E[Y] \end{aligned}$$

Ainsi, on obtient bien que $E[X+Y] = E[X] + E[Y]$. □

Exemple 3.7. Un problème de Daniel Bernoulli.
Daniel Bernoulli (1700–1782) est le fils de Jean et le neveu de Jacques Bernoulli, l'un des mathématiciens les plus illustres de cette famille bâloise. Ami d'Euler, il travailla avec lui à Saint-Petersbourg pendant plusieurs années, avant de s'installer à Bâle.

Le problème est le suivant. Une urne contient $2N$ cartes numérotées de 1 à N par paire. On tire m cartes au hasard. Quel est le nombre moyen de paires de cartes encore présentes dans l'urne ?

On définit X_i la variable aléatoire qui vaut 1 si la i-ème paire est intacte, et zéro sinon. Quelle est l'espérance de X_i ?

Le but est de trouver l'espérance de la somme de toutes les variables X_i.
Le théorème précédent affirme que cette espérance vaut

Ce modèle fut proposé pour déterminer combien il reste de couples après la mort de m personnes... Par exemple, s'il y a 100 couples au départ et que 50 personnes meurent, on peut espérer trouver encore $\cong 56$ couples "indemmes" après ce coup du sort.

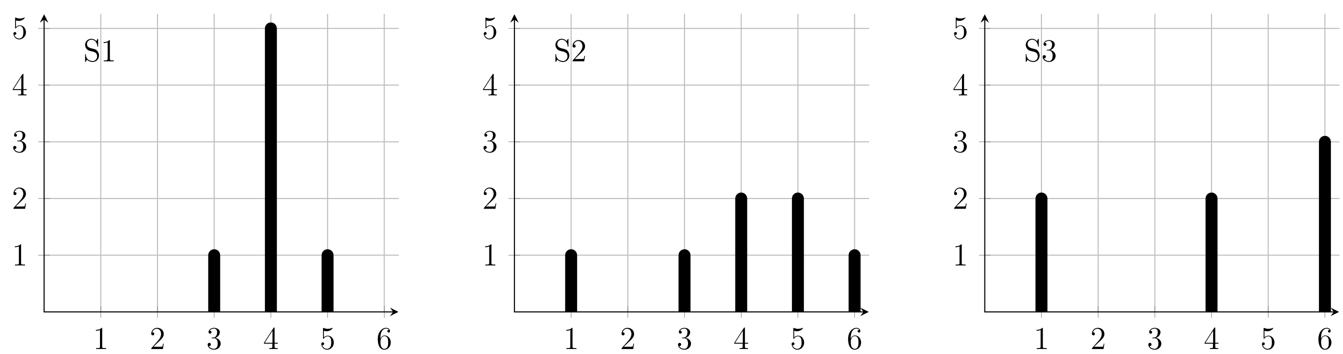
V. Variance et corrélation

La semaine passée, nous avons travaillé avec des variables aléatoires et défini l'espérance.

Pour terminer l'étude des variables aléatoires, nous définissons encore la variance, une mesure de dispersion des valeurs d'une variable autour de l'espérance, puis nous finissons en analysant la corrélation entre deux variables.

1 Variance

Exemple 1.1. Considérons les trois séries $S_1 = \{3, 4, 4, 4, 4, 4, 5\}$, $S_2 = \{1, 3, 4, 4, 5, 5, 6\}$ et $S_3 = \{1, 1, 4, 4, 6, 6, 6\}$. Elles possèdent la même moyenne et la même médiane, toutes deux égales à 4, mais il est clair que les données des séries 2 et 3 sont plus dispersées que celles de la série 1.



S'il s'agit des notes obtenues lors d'un test dans une petite classe, ses séries représentent des situations fort différentes. Il faut une **mesure de dispersion** pour compléter l'analyse statistique.

Pour cela, considérons dans chaque série les écarts entre les valeurs et la moyenne . Comme ces écarts peuvent être positifs ou négatifs, on les élève au carré.

Définition 1.2. Soit X une variable aléatoire et $\mu = E[X]$ son espérance. La *variance* de X vaut

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

L'unité de la variance diffère de l'unité de la variable aléatoire. On lui préfère l'écart-type.

Définition 1.3. Soit X une variable aléatoire. L'*écart-type* de X est le nombre

$$\sigma(X) = \sqrt{\operatorname{Var}(X)} = \sigma_X.$$

Remarque 1.4. En théorie des probabilités, l'inégalité de Bienaymé-Tchebychev montre que, quelle que soit la distribution d'une variable aléatoire ou statistique, au moins la moitié des valeurs se trouvent dans l'intervalle $[\bar{x} - \sqrt{2}\, s\ ; \bar{x} + \sqrt{2} s\ ]$.
Pratiquement, en statistique, l'intervalle $[\bar{x} - s\ ; \bar{x} + s]$ contient déjà souvent la majorité des données.

Exemple 1.5. Calculons les variances puis les écart-types des séries de l'exemple 1.1 et donnons-en une interprétation selon la remarque 1.4 ci-dessus.

A des fins calculatoires, il est souvent plus commode d'utiliser la formule donnée dans le théorème suivant. Il dit que la variance mesure la différence entre le carré de l'espérance et l'espérance au carré.

Proposition 1.6. Formule de König *Soit X une variable aléatoire. Alors*

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

Démonstration. Si $E[X] = \mu$, la définition de la variance utilise le carré $(X - \mu)^2 = X^2 - 2\mu X + \mu^2$.

□

Exemple 1.7. Calculons la variance des séries de l'exemple 1.1 avec la formule de König :

2 Covariance

Dans l'étude des variables aléatoires, il est parfois souhaité de décider à quel point deux variables sont corrélées. Lorsqu'elles sont indépendantes, elles ne le sont pas du tout, mais dans le cas contraire, il se peut qu'elles soient plus ou moins fortement liées.

Définition 2.1. Soient X et Y deux variables aléatoires définies sur le même ensemble fondamental. La *covariance* est définie par

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

Exemple 2.2. Soient X et Y deux variables aléatoires valant 1 ou zéro, selon que l'on tire un as ou non dans le cas de X, et une carte de trèfle ou non dans le cas de Y dans un jeu de 36 cartes.

Avant de continuer, observons deux choses. La première est que cette manière de calculer est quelque peu laborieuse, et nous allons développer d'autres méthodes de calcul. La seconde est que la covariance est nulle dans ce cas car les variables sont indépendantes, et nous verrons qu'effectivement la covariance permet de mesurer la corrélation entre X et Y.

Proposition 2.3. *On a* $\operatorname{Cov}(X,Y) = E[XY] - E[X]E[Y]$.

Démonstration. On développe le produit $(X - E[X])(Y - E[Y])$ et on applique la linéarité de l'espérance pour obtenir

□

Exemple 2.4. Calculons la covariance de l'exemple 2.2 à l'aide de la formule :

La notion d'indépendance que nous avons vue pour les événements se définit aussi pour variables aléatoires : deux variables sont indépendantes si tous les événements qu'elles décrivent le sont.

Définition 2.5. Deux variables aléatoires X et Y sont *indépendantes* si pour tout choix de sous-ensembles $A, B \subset \mathbb{R}$ on a

$$P\{X \in A, Y \in B\} = P\{X \in A\} \cdot P\{Y \in B\}$$

Proposition 2.6. *Si X et Y sont indépendantes, alors* $\operatorname{Cov}(X,Y) = 0$.

Démonstration. Avec la proposition 2.3, il suffit de montrer que $E[XY] = E[X]E[Y]$.

□

Vous verrez en exercice que la réciproque est fausse : la covariance peut être nulle sans pour autant que les variables soient indépendantes.

3 Régression linéaire des moindres carrés

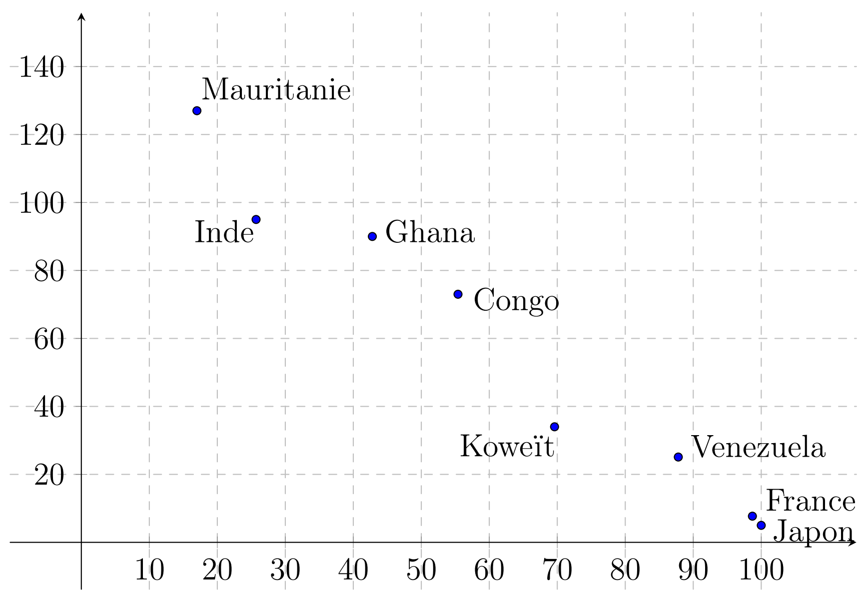
Exemple 3.1. Dans le tableau ci-dessous, X désigne le taux en % d'alphabétisation des femmes et Y le taux en ‰ de mortalité infantile dans les années 1980. On peut en effet supposer que le taux d'alphabétisation des femmes a de l'impact sur le taux de mortalité infantile.

Pays	**Inde**	**Koweït**	**Mauritanie**	**France**	**Ghana**	**Congo**	**Venezuela**	**Japon**
X [%]	25.7	69.6	17	98.7	42.8	55.4	87.8	100
Y [‰]	95	34	127	7.7	90	73	25.1	5

Dans cet exemple, l'ensemble fondamental S est

Pour tout élément $s \in S$, nous disposons de deux valeurs $x_s = X(s)$ et $y_s = Y(s)$.
Si nous plaçons toutes les paires (x_s, y_s) dans le plan $\mathbb{R}^2$, nous obtenons un nuage de points.
Le but est de tracer une droite qui approche "le mieux possible" ce nuage.

Taux d'alphabétisation des femmes et taux de mortalité infantile.



La forme du nuage de points suggère que Y est liée à X par une relation affine. Mais laquelle ? Nous allons déterminer la droite de régression obtenue avec la méthode des moindres carrés.

Historiquement, le premier texte paru, faisant mention de la méthode des moindres carrés, est dû à Adrien-Marie Legendre (1752-1833), dans un article sur ses « nouvelles méthodes pour la détermination des orbites des comètes » publié en 1805. Un an plus tard, Carl Friedrich Gauss (1777-1855) fait aussi allusion à cette méthode.

On cherche donc une droite d'équation $y = ax + b$ telle que les valeurs $ax_s + b$ soient proches des valeurs y_s. Concrètement, on aimerait minimiser la distance entre ces deux valeurs, pour tous les s. Puisque la différence $\varepsilon_s = y_s - ax_s - b$ peut être positive ou négative, on va minimiser la somme des carrés de ces différence ε_s.
Calculons donc a et b pour minimiser l'expression

$$S = \sum_s (y_s - ax_s - b)^2$$

Développons d'abord cette somme de carrés comme une fonction de b :

$S(b) = \sum_s (\;\; y_s - ax_s \;\; - b)^2 \;\; =$

En d'autre terme, la droite de régression que nous cherchons passe par le point "moyen" (μ_X, μ_Y) du nuage de points. Il reste donc à déterminer a.
Pour déterminer a, nous remplaçons b par la valeur $\mu_y - a\mu_x$ que nous venons de trouver dans l'expression $\sum_s (y_s - ax_s - b)^2$ de départ. La somme à minimiser est maintenant

$$\begin{aligned} S(a) &= \sum_s (y_s - ax_s - (\mu_y - a\mu_x))^2 \;=\; \sum_s (y_s - \mu_y - a(x_s - \mu_x))^2 \\ &= \sum_s (y_s - \mu_y)^2 - 2\,a \sum_s (y_s - \mu_y)(x_s - \mu_x) + a^2 \sum_s (x_s - \mu_x)^2 \\ &= n\,\mathrm{Var}(Y) - 2\,n\,a\,\mathrm{Cov}(X,Y) + n\,a^2\,\mathrm{Var}(X) \\ &= n\left(a\sqrt{\mathrm{Var}(X)} - \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}\right)^2 + \;\ldots \end{aligned}$$

Comme précédemment, on conclut que $S(a)$ est minimale lorsque $a = \dfrac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}$

Nous avons démontré le résultat suivant :

Théorème 3.2. *Soit S l'ensemble fondamental de deux variables aléatoires X et Y. Alors l'approximation affine de Y en fonction de X qui minimise la somme des carrés $(y_s - x_s)^2$ pour autant que la variances de X soit non nulle est*

$$Z = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}(X - \mu_X) + \mu_Y.$$

Exemple 3.3. Dans l'exemple 3.1 des taux d'alphabétisation et de mortalité, nous calculons $E[X] = 62,125$, $E[Y] = 57,1$ puis $\mathrm{Var}(X) = E[X^2] - E[X]^2 = 4768,15 - (62,125)^2 = 908,63$.

Et pour la covariance, $\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = 2290,86 - 62,125 \cdot 57,1 = -1256,48$.

D'où le modèle affine $Z = \dfrac{-1256,48}{908,63}(X - 62,125) + 57,1$ c'est-à-dire $Z = -1,38\,X + 143,01$.

4 Coefficient de détermination et coefficient de corrélation

La droite des moindres carrés peut être déterminée quelle que soit la forme du nuage de points pour autant que $\mathrm{Var}(X) \neq 0$, mais ce n'est pas forcément un modèle judicieux de relation entre les variables X et Y. Il faut donc disposer d'outils de mesure de la qualité de la relation obtenue. Vous montrerez en exercice qu'avec la droite des moindres carrés, la moyenne des carrés des écarts vaut

$$\frac{1}{n}\sum_s \varepsilon_s^2 \; = \; \mathrm{Var}(Y)\left(1 - \frac{(\mathrm{Cov}(X,Y))^2}{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}\right)$$

Ce résultat motive l'introduction du coefficient de détermination et du coefficient de corrélation comme indicateurs de pertinence du modèle des moindres carrés.

Définition 4.1. Soient X et Y deux variables aléatoires. Le *coefficient de détermination* $\rho^2(X,Y)$ ou plus simplement ρ^2 est défini pour autant que les variances de X et Y soient non nulles par

$$\rho^2 = \frac{(\mathrm{Cov}(X,Y))^2}{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}.$$

Par définition, ρ^2 est compris entre 0 et 1 et peut être interprété comme un pourcentage.
Si $\rho^2 = 1$, alors $\sum \varepsilon_s^2 = 0$, ce qui signifie que la relation linéaire établie entre X et Y explique le 100% de la variance de Y.
Si $\rho^2 < 1$, alors $\sum \varepsilon_s^2 > 0$ et les points du nuage ne sont pas parfaitement alignés. Seule une proportion de ρ^2 de la variance de Y s'explique par la relation établie entre X et Y alors qu'une proportion de $(1 - \rho^2)$ de la variance de Y provient d'autres facteurs d'influence que X.

Pour une **interprétation correcte** du coefficient de détermination ρ^2, on peut utiliser la formulation suivante :

> Selon le modèle de régression, *nom de la variable* X explique $\rho^2 \cdot 100\%$ de la variance de *nom de la variable* Y ; $(1-\rho^2)\cdot 100\%$ de cette variance est imputable à d'autres facteurs.

Exemple 4.2. Pour le modèle de régression de l'exemple 3.3, le coefficient de détermination, après calcul de $\mathrm{Var}(Y) = E[Y^2] - E[Y]^2 = 5056,66 - (57,1)^2 = 1796,25$, est obtenu comme suit :

$$\rho^2 = \frac{(\mathrm{Cov}(X,Y))^2}{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)} = \frac{(-1256,48)^2}{908,63\cdot 1796,25} \cong 0,9678 \cong 97\%.$$

Interprétation : le taux d'analphabétisme des femmes explique de la variance du taux de mortalité infantile. En conséquence, seulement de cette variance s'explique par d'autres facteurs comme le taux de vaccination par exemple.

Définition 4.3. Soient X et Y deux variables aléatoires. Le *coefficient de corrélation* $\rho(X,Y)$ est défini pour autant que les variances de X et Y soient non nulles par

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

Exemple 4.4. Dans l'exemple 3.1 des taux d'alphabétisation et de mortalité, nous obtenons

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{-1256,48}{\sqrt{908,63\cdot 1796,25}} \cong -0,984.$$

Ce nombre est négatif car le taux de mortalité infantile décroît lorsque le taux d'alphabétisation des femmes croît. Par ailleurs, il est très proche de 1 en valeur absolue, ce qui indique une corrélation forte des deux variables.

Remarque 4.5.

a) Le coefficient de corrélation linéaire prend des valeurs comprises entre -1 et 1.

b) Si $\rho = 1$ ou -1, les points sont parfaitement alignés.

c) On considère généralement que la corrélation linéaire est
   - **forte** si $|\rho| \geq 0.9$; la régression linéaire exprime parfaitement le lien entre les données ;
   - **moyenne** si $0.6 \leq |\rho| < 0.9$; le modèle linéaire peut être considéré comme acceptable ;
   - **faible** si $0.2 \leq |\rho| < 0.6$; le modèle linéaire doit être remis en cause ;
   - **nulle** si $|\rho| \leq 0.2$; dans ce cas le modèle linéaire doit être rejeté. On dit alors que les variables X et Y sont **non-corrélées** linéairement.

d) Si $\rho > 0$, X et Y sont **corrélées positivement** ; la droite de régression a une pente positive.
   Si $\rho < 0$, X et Y sont **corrélées négativement** ; la droite de régression a une pente négative.

e) le coefficient de détermination est égal au carré du coefficient de corrélation.