

Mathématiques Appliquées 2

Filières SYND et ETE

4ème semestre
printemps 2025

Table des matières

I	Algèbre linéaire	3
1	L'espace à n dimensions \mathbb{R}^n	4
1.1	Généralités	4
1.2	Produit scalaire et projection orthogonale	7
1.3	Droite, plan et hyperplan	8
2	Matrices	9
2.1	Définitions et opérations	9
2.2	Déterminant	12
2.3	Matrice inverse	14
3	Applications linéaires	17
3.1	Définitions et propriétés	17
3.2	Applications linéaires orthogonales	20
II	Statistiques appliquées	22
4	Statistique descriptive	23
4.1	Données univariées	24
4.1.1	Représentations graphiques	24
4.1.2	Résumés numériques	25
4.1.3	Données standardisées	29
4.2	Données bivariées	29
4.3	Régression linéaire	31
5	Probabilités	33
5.1	Définitions et notations	33
5.2	Probabilité conditionnelle, indépendance et formule de Bayes . .	35
5.3	Arbre de probabilités	36

6	Variables aléatoires	37
6.1	Variables discrètes	37
6.1.1	Loi de Bernoulli	40
6.1.2	Loi binomiale	40
6.1.3	Loi de Poisson	41
6.2	Variable continue	43
6.2.1	Loi normale	46
6.2.2	Théorème limite-central	48
7	Introduction aux tests statistiques	50
7.1	Test de proportion	54

Première partie

Algèbre linéaire

Chapitre 1

L'espace à n dimensions \mathbb{R}^n

1.1 Généralités

De façon analogue au plan \mathbb{R}^2 et à l'espace à 3 dimensions \mathbb{R}^3 , on définit (mathématiquement) l'espace à n dimensions \mathbb{R}^n .

- Un **point** $A \in \mathbb{R}^n$ a la forme

$$A = (a_1, a_2, \dots, a_n).$$

Les n nombres réels a_1, a_2, \dots, a_n sont les **coordonnées** du point A .

- Le point $O = (0, 0, \dots, 0) \in \mathbb{R}^n$ est l'**origine** de \mathbb{R}^n .
- La **distance** entre les points $A = (a_1, \dots, a_n)$ et $B = (b_1, \dots, b_n)$ est définie par

$$\text{dist}(A, B) = \sqrt{(b_1 - a_1)^2 + \dots + (b_n - a_n)^2}.$$

La **sphère** centrée au point $A = (a_1, a_2, \dots, a_n) \in \mathbb{R}^n$ et de rayon $r > 0$ est l'ensemble des points $P = (x_1, x_2, \dots, x_n)$ tels que

$$\begin{aligned} \text{dist}(P, A) &= r \\ \Leftrightarrow \\ (x_1 - a_1)^2 + (x_2 - a_2)^2 + \dots + (x_n - a_n)^2 &= r^2. \end{aligned}$$

À partir de points, on peut définir des vecteurs.

- Le **vecteur-lieu** du point $A = (a_1, \dots, a_n) \in \mathbb{R}^n$ est défini par :

$$\overrightarrow{OA} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

- Pour $A = (a_1, \dots, a_n)$ et $B = (b_1, \dots, b_n)$, le vecteur \overrightarrow{AB} est défini par :

$$\overrightarrow{AB} = \begin{pmatrix} b_1 - a_1 \\ b_2 - a_2 \\ \vdots \\ b_n - a_n \end{pmatrix}.$$

- Le **vecteur nul** de \mathbb{R}^n , noté $\vec{0}$ est le vecteur

$$\vec{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Pour \vec{a} et $\vec{b} \in \mathbb{R}^n$ et $\lambda \in \mathbb{R}$, l'**addition** et la **multiplication par un scalaire** sont définies de manière analogue au cas de vecteurs de \mathbb{R}^2 et \mathbb{R}^3 :

$$\vec{a} + \vec{b} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix}, \quad \lambda \vec{a} = \begin{pmatrix} \lambda a_1 \\ \lambda a_2 \\ \vdots \\ \lambda a_n \end{pmatrix}.$$

Pour $\vec{a} \in \mathbb{R}^n$, la **norme** de \vec{a} est aussi analogue au cas de vecteurs de \mathbb{R}^2 et \mathbb{R}^3 :

$$\|\vec{a}\| = \sqrt{a_1^2 + a_2^2 + \dots + a_n^2}.$$

- Un vecteur \vec{v} est une **combinaison linéaire** des vecteurs

$$\vec{v}_1, \dots, \vec{v}_m,$$

s'il peut s'écrire sous la forme :

$$\vec{v} = \lambda_1 \vec{v}_1 + \lambda_2 \vec{v}_2 + \dots + \lambda_m \vec{v}_m,$$

avec

$$\lambda_1, \dots, \lambda_m \in \mathbb{R}.$$

- Des vecteurs

$$\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m$$

sont **linéairement indépendants** si l'équation

$$\lambda_1 \vec{a}_1 + \lambda_2 \vec{a}_2 + \dots + \lambda_m \vec{a}_m = \vec{0}$$

n'a pas d'autre solution que

$$\lambda_1 = \lambda_2 = \dots = \lambda_m = 0.$$

(Cela signifie qu'aucun des vecteurs ne peut s'écrire comme combinaison linéaire des autres. Dans ce cas, les m vecteurs engendrent un (sous-)espace à m dimensions de \mathbb{R}^n).

On peut montrer :

Théorème : Une famille de n vecteurs $\vec{v}_1, \dots, \vec{v}_n \in \mathbb{R}^n$ linéairement indépendants forme une **base** de \mathbb{R}^n . Dans une base, tout vecteur $\vec{v} \in \mathbb{R}^n$ s'écrit de façon unique comme combinaison linéaire des vecteurs de base :

$$\vec{v} = \lambda_1 \vec{v}_1 + \lambda_2 \vec{v}_2 + \dots + \lambda_n \vec{v}_n.$$

Les nombres $\lambda_1, \lambda_2, \dots, \lambda_n$ sont les **composantes** de \vec{v} dans la base $(\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n)$.

1.2 Produit scalaire et projection orthogonale

On définit le **produit scalaire** des vecteurs

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad \text{et} \quad \vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

de \mathbb{R}^n par

$$\vec{a} \cdot \vec{b} = a_1 b_1 + a_2 b_2 + \dots + a_n b_n = \sum_{i=1}^n a_i b_i.$$

On dit que \vec{a} et \vec{b} sont **perpendiculaires** (ou **orthogonaux**) si

$$\vec{a} \cdot \vec{b} = 0.$$

L'**angle** entre les vecteurs \vec{a} et $\vec{b} \in \mathbb{R}^n$ est donné par

$$\varphi = \arccos \left(\frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|} \right)$$

Pour 2 vecteurs \vec{a} et \vec{b} quelconques (avec $\vec{b} \neq 0$), la **projection orthogonale** (vectorielle) de \vec{a} sur \vec{b} est définie par

$$\vec{a}_{\vec{b}} = \frac{\vec{a} \cdot \vec{b}}{\|\vec{b}\|^2} \vec{b}.$$

Théorème : Si $\vec{v}_1, \dots, \vec{v}_n$ forment une **base orthonormée** de \mathbb{R}^n .

C'est-à-dire :

- Tous les vecteurs sont perpendiculaires entre eux.
- Tous les vecteurs ont une norme égal à 1.

Alors tout vecteur \vec{v} est la somme de ses projections orthogonales sur les vecteurs de base :

$$\vec{v} = (\vec{v} \cdot \vec{v}_1) \vec{v}_1 + \dots + (\vec{v} \cdot \vec{v}_n) \vec{v}_n = \sum_{i=1}^n (\vec{v} \cdot \vec{v}_i) \vec{v}_i.$$

1.3 Droite, plan et hyperplan

- Soient A un point et \vec{d} un vecteur de \mathbb{R}^n . La **droite** d passant par A et de vecteur directeur \vec{d} est l'ensemble des points $P \in \mathbb{R}^n$ tels que

$$d : \overrightarrow{OP} = \overrightarrow{OA} + t\vec{d}, \quad t \in \mathbb{R}.$$

- Soient A un point et \vec{u} et \vec{w} deux vecteurs linéairement indépendants de \mathbb{R}^n , l'ensemble des points P tels que

$$\overrightarrow{OP} = \overrightarrow{OA} + t\vec{u} + s\vec{w}, \quad \text{avec } t, s \in \mathbb{R}$$

forme un **plan** à 2 dimensions dans \mathbb{R}^n passant par A .

- Soient A un point et $\vec{u}_1, \dots, \vec{u}_{n-1}$, $(n - 1)$ vecteurs linéairement indépendants de \mathbb{R}^n . L'ensemble des points P tels que

$$\overrightarrow{OP} = \overrightarrow{OA} + t_1\vec{u}_1 + \dots + t_{n-1}\vec{u}_{n-1}, \quad \text{avec } t_1, \dots, t_{n-1} \in \mathbb{R}$$

forme un (sous-)espace de dimension $(n - 1)$ de \mathbb{R}^n . Un tel (sous-)espace est appelé un **hyperplan**.

On peut également le décrire par son **équation cartésienne** : Il s'agit de l'ensemble des points $X = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ tels que

$$n_1x_1 + n_2x_2 + \dots + n_nx_n = b,$$

où le vecteur

$$\vec{n} = \begin{pmatrix} n_1 \\ n_2 \\ \vdots \\ n_n \end{pmatrix}$$

est perpendiculaire à l'hyperplan et

$$b = \overrightarrow{OA} \cdot \vec{n}.$$

La **distance** entre le point $P = (p_1, p_2, \dots, p_n) \in \mathbb{R}^n$ et l'hyperplan $H : n_1x_1 + n_2x_2 + \dots + n_nx_n = b$ est donnée par

$$\text{dist}(P, H) = \frac{|n_1p_1 + n_2p_2 + \dots + n_np_n - b|}{\sqrt{n_1^2 + n_2^2 + \dots + n_n^2}}.$$

Chapitre 2

Matrices

2.1 Définitions et opérations

Une **matrice** $m \times n$ est un tableau à m lignes et n colonnes. On l'écrit généralement entre parenthèses et on numérote les éléments selon leur position dans la matrice :

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

a_{ij} désigne l'élément qui se trouve à la $i^{\text{ème}}$ ligne et à la $j^{\text{ème}}$ colonne de la matrice A .

On note aussi

$$A = (a_{ij})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}.$$

Types de matrices et matrices particulières :

- Une matrice $n \times n$ est une **matrice carrée** (d'ordre n). La **diagonale** d'une matrice carrée est l'ensemble des éléments a_{ii} (commence en haut à gauche et se termine en bas à droite).
- Une **matrice triangulaire supérieure** (resp. **inférieure**) est une matrice carrée dont tous les éléments situés en-dessous (resp. au-dessus) de la diagonale sont nuls.
- Une matrice carrée $A = (a_{ij})$ est **symétrique**, si $a_{ij} = a_{ji}$ pour tout i et j , c'est-à-dire si la matrice est symétrique par rapport à la diagonale.

- Une matrice carrée dont tous les éléments en dehors de la diagonale sont nuls est une **matrice diagonale**.
- La **matrice identité** de dimension n est la matrice $n \times n$ dont tous les éléments diagonaux sont égaux à 1 et tous les autres sont nuls. On la note I (ou I_n) :

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & 1 \end{pmatrix}$$

- La **matrice nulle** $m \times n$ est la matrice O qui ne contient que des zéros.

Opérations matricielles :

- Soit $A = (a_{ij})$ une matrice $m \times n$. La **transposée** de A est la matrice $n \times m$ définie par

$$A^T = (a_{ji}).$$

Les lignes de A deviennent les colonnes de A^T et réciproquement.

- La **somme** de deux matrices $m \times n$ A et B (de mêmes dimensions!) est définie par

$$A + B = (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij}).$$

On additionne les éléments correspondants.

- La **multiplication** d'une matrice A par un **nombre** $\lambda \in \mathbb{R}$ est définie par

$$\lambda A = \lambda(a_{ij}) = (\lambda a_{ij}).$$

On multiplie chaque élément par λ .

- Soit A une matrice $m \times n$ et B une matrice $n \times p$ (c'est-à-dire : le nombre de colonnes de A est égal au nombre de lignes de B). Dans ce cas, on définit le **produit** $A \cdot B$ par la matrice $C = (c_{ij})$ de dimensions $m \times p$ telle que

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}.$$

On fait le produit scalaire de la $i^{\text{ème}}$ ligne de A et de la $j^{\text{ème}}$ colonne de B .

Propriétés :

1. $A + B = B + A$
2. $(A + B) + C = A + (B + C)$
3. $(\lambda A)B = A(\lambda B) = \lambda(AB)$.
4. $AI = A$ et $IA = A$, où I sont des matrices identité (de dimensions adéquates)
5. $(AB)C = A(BC)$
6. $A(B + C) = AB + AC$ et $(B + C)A = BA + CA$.
7. $(A^T)^T = A$;
8. $A = A^T$ si et seulement si A est symétrique.
9. $(A + B)^T = A^T + B^T$
10. $(AB)^T = B^T A^T$.

Remarque : Le produit de 2 matrices n'est **PAS** commutatif. En général, on a

$$AB \neq BA.$$

Le système d'équations linéaires

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{cases}$$

peut s'écrire sous la **forme matricielle** :

$$A\vec{x} = \vec{b},$$

c'est-à-dire

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix},$$

A : matrice $m \times n$, $\vec{x} \in \mathbb{R}^n$ et $\vec{b} \in \mathbb{R}^m$ vecteurs.

2.2 Déterminant

Dans ce paragraphe, on considère seulement des **matrices carrées**.

Soit A une matrice $n \times n$. Le **déterminant** de A est un nombre noté $\det(A)$ ou $|A|$.

Calcul du déterminant :

On définit le déterminant de manière récursive :

- Si

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

est une matrice 2×2 , son déterminant est défini par :

$$\det(A) = ad - bc.$$

- Si $A = (a_{ij})$ est une matrice $n \times n$ ($n \geq 3$), on note A_{ij} la matrice $(n-1) \times (n-1)$ qu'on obtient en supprimant la $i^{\text{ème}}$ ligne et la $j^{\text{ème}}$ colonne de A .

Le déterminant de A est alors défini de manière récursive de l'une des 2 manières suivantes :

a) (développement suivant une colonne)

On choisit une colonne de A , disons la $j^{\text{ème}}$, et on pose

$$\det(A) = (-1)^{1+j}a_{1j}\det(A_{1j}) + (-1)^{2+j}a_{2j}\det(A_{2j}) \\ + \dots + (-1)^{n+j}a_{nj}\det(A_{nj}).$$

Autrement dit, on multiplie chaque élément a_{ij} de la colonne j par le déterminant de la matrice A_{ij} et on munit ces termes du signe $(-1)^{i+j}$ qui suit le schéma

$$\begin{pmatrix} + & - & + & - & \dots \\ - & + & - & + & \dots \\ + & - & + & - & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

et on additionne tous ces termes.

b) (développement suivant une ligne)

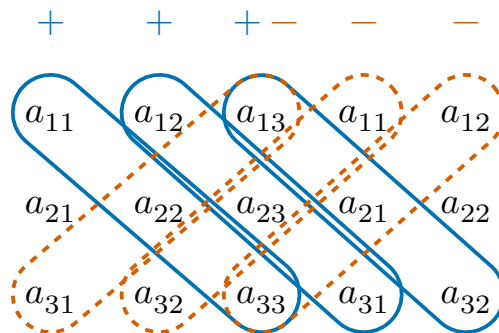
On choisit une ligne de A , disons la $i^{\text{ème}}$, et on pose

$$\det(A) = (-1)^{i+1}a_{i1} \det(A_{i1}) + (-1)^{i+2}a_{i2} \det(A_{i2}) \\ + \dots + (-1)^{i+n}a_{in} \det(A_{in}).$$

En d'autres termes, on fait avec une ligne ce qu'on a fait dans a) avec une colonne.

On peut montrer que le résultat ne dépend pas de la colonne ou de la ligne choisie pour développer $\det(A)$.

Remarque : Pour une matrice 3×3 , on peut également calculer $\det(A)$ à l'aide de la **règle de Sarrus** :



$$\Rightarrow \det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12}$$

Quelques propriétés :

1. $\det(A^T) = \det(A)$.
2. $\det(AB) = \det(A) \det(B)$.
3. Si A est une matrice triangulaire, alors $\det(A)$ est égal au produit des éléments de la diagonale. En particulier : $\det(I) = 1$.
4. $\det(\lambda A) = \lambda^n \det(A)$, si A est une matrice $n \times n$.

Interprétation géométrique en dimensions 2 et 3 :

- Si A est une matrice 2×2 , $|\det(A)|$ = aire du parallélogramme dont les côtés sont formés par les colonnes de A .
- Si A est une matrice 3×3 , $|\det(A)|$ = volume du parallélépipède dont les arêtes sont formées par les colonnes de A .

Conséquences :

- $\det(A) = 0$ si et seulement si les colonnes de A sont des vecteurs linéairement dépendants.
- $\det(A) = 0$ si et seulement si les lignes de A sont des vecteurs linéairement dépendants.

Application : Valeurs propres

Soit A une matrice carrée $n \times n$ et I_n la matrice identité de dimensions $n \times n$. Le **polynôme caractéristique** de A est défini par

$$\chi_A(\lambda) = \det(A - \lambda I_n).$$

Les **valeurs propres** de A sont les zéros du polynôme caractéristique $\chi_A(\lambda)$. Autrement dit, les valeurs propres de A sont les solutions de l'équation

$$\chi_A(\lambda) = 0.$$

Théorème : Soit A une matrice $n \times n$.

1. $\chi_A(\lambda)$ est un polynôme de degré n dans la variable λ . Il y a donc n valeurs propres (complexes) de A , en comptant leur multiplicité.
2. Le déterminant de A est égal au produit des valeurs propres, c-à-d. :

$$\det(A) = \lambda_1 \lambda_2 \dots \lambda_n,$$

où $\lambda_1, \dots, \lambda_n$ sont les valeurs propres de A .

2.3 Matrice inverse

Dans ce paragraphe, on considère seulement des **matrices carrées**.

Théorème : Soit A et B deux matrices $(n \times n)$, alors

$$AB = I \implies BA = I.$$

On dit que B est la **matrice inverse** de A et que A est l'**inverse** de B .

On note $B = A^{-1}$ et $A = B^{-1}$.

Si A possède une matrice inverse, on dit que A est **inversible** (ou **régulière**).
Sinon, on dit que A est **singulière**.

Propriétés et notation :

- Si A possède une matrice inverse, elle est **unique**.
- Si A est inversible, alors A^{-1} est **inversible** et

$$(A^{-1})^{-1} = A.$$

- Si A et B sont inversibles, alors AB est **inversible** et

$$(AB)^{-1} = B^{-1}A^{-1}.$$

- Si A est inversible, alors A^T est **inversible** et

$$(A^T)^{-1} = (A^{-1})^T.$$

- Si A est inversible, alors

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

Critères pour savoir si une matrice est inversible

- A est inversible si et seulement si

$$\det(A) \neq 0.$$

- A est inversible si et seulement si les colonnes de A sont des vecteurs **linéairement indépendants**.
- A est inversible si et seulement si les lignes de A sont des vecteurs **linéairement indépendants**.
- A est inversible si et seulement si pour tout \vec{b} le système $A\vec{x} = \vec{b}$ a une **unique solution** \vec{x} .

Remarque : Si A est inversible, l'unique solution du système $A\vec{x} = \vec{b}$ est donnée par

$$\vec{x} = A^{-1}\vec{b}.$$

Calcul de A^{-1} :

- Si A est une matrice 2×2

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

et si $\det(A) \neq 0$, alors l'inverse de A est donnée par

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

(Remarque : Une formule similaire, basée sur les déterminants, existe dans le cas d'une matrice inversible $n \times n$, $n \geq 3$.)

- Algorithme de Gauss-Jordan

Pour inverser une matrice A de taille $n \times n$, on forme une matrice $n \times (2n)$ en juxtaposant à A la matrice I (de dimension n), ce qui donne $(A \mid I)$ et on essaie d'amener cette matrice à la forme $(I \mid B)$ en répétant les opérations suivantes (basées sur l'algorithme d'élimination de Gauss pour les systèmes d'équations) :

- a) multiplier une ligne par un nombre $\neq 0$;
- b) ajouter à une ligne un multiple d'une autre ligne.
- c) intervertir 2 lignes.

$$\left(\begin{array}{ccc|ccc} a_{11} & \dots & a_{1n} & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} & 0 & \dots & 1 \end{array} \right) \xrightarrow[\text{Gauss}]{\text{Algorithme}} \left(\begin{array}{ccc|ccc} 1 & \dots & 0 & b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & b_{n1} & \dots & b_{nn} \end{array} \right).$$

Remarque : Avec cette méthode on peut (sans calculer le déterminant) savoir si une matrice est inversible ou non : Si on réussit à atteindre la forme $(I \mid B)$, alors A est **inversible** et

$$A^{-1} = B.$$

Si, en partant de $(A \mid I)$, on obtient uniquement des zéros dans la moitié gauche d'une ligne, alors A n'est **PAS** inversible.

Chapitre 3

Applications linéaires

3.1 Définitions et propriétés

Une application $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ est dite **linéaire**, s'il existe une matrice A de dimension $(m \times n)$ telle que

$$f(\vec{x}) = A\vec{x}.$$

Théorème : Une application f est linéaire si et seulement si

$$\begin{aligned} f(\vec{x} + \vec{y}) &= f(\vec{x}) + f(\vec{y}) \\ f(\lambda\vec{x}) &= \lambda f(\vec{x}) \end{aligned}$$

pour tous les \vec{x}, \vec{y} et tous les $\lambda \in \mathbb{R}$.

(c'est-à-dire : l'image d'une somme est égal à la somme des images, et l'image de $\lambda\vec{x}$ est égal à $\lambda \cdot$ image de \vec{x} .)

Remarques :

- Si f est une application linéaire, alors

$$f(\vec{0}) = \vec{0}.$$

- Une application de la forme $f(\vec{x}) = A\vec{x} + \vec{b}$ n'est pas linéaire si

$$\vec{b} \neq \vec{0}.$$

On dit qu'il s'agit d'une application **affine**.

Dans la suite du chapitre, on considère seulement des applications linéaires de \mathbb{R}^n dans \mathbb{R}^n .

Une telle application, f , peut être vue comme une **transformation** de l'espace \mathbb{R}^n : Tout vecteur $\vec{x} \in \mathbb{R}^n$ est transformé par f en un vecteur $f(\vec{x})$ de \mathbb{R}^n (le vecteur-image de \vec{x}).

Exemples :

- Dans \mathbb{R}^n , la projection orthogonale sur l'axe x_1 est l'application linéaire qui transforme tout vecteur

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{en} \quad f(\vec{x}) = \begin{pmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

La matrice de f est

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}.$$

- Dans \mathbb{R}^3 , la projection orthogonale sur le plan xy transforme tout vecteur

$$\vec{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \text{en} \quad f(\vec{x}) = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}.$$

Sa matrice est

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Théorème : Soit f une application linéaire de $\mathbb{R}^n : f(\vec{x}) = A\vec{x}$. Alors les colonnes de A sont les images des vecteurs de la base canonique de \mathbb{R}^n , c'est-à-dire :

$$A = \left(f(\vec{e}_1) \dots f(\vec{e}_n) \right)$$

où

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \vec{e}_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Propriétés :

- Interprétation géométrique :
 - En dimension 2, c'est-à-dire, si f est une application linéaire de \mathbb{R}^2 , le déterminant de A représente l'aire (orientée) du parallélogramme engendré par $f(\vec{e}_1)$ et $f(\vec{e}_2)$.
 - En dimension 3, $\det(A)$ représente le volume (orienté) du parallélépipède qui est l'image par f du cube unité (dont les arêtes sont $\vec{e}_1, \vec{e}_2, \vec{e}_3$).
 - En dimension $n \geq 4$, $\det(A)$ représente le volume orienté de l'image par f de l'hypercube unité.
- Si f et g sont des applications linéaires, alors la composée $g \circ f$ est aussi linéaire. Si A_f et A_g sont les matrices de f et g , alors la matrice de $g \circ f$ est donnée par le produit :

$$A_{g \circ f} = A_g A_f.$$

- Une application linéaire f est bijective, si pour tout $\vec{y} \in \mathbb{R}^n$ il existe un unique $\vec{x} \in \mathbb{R}^n$ tel que $\vec{y} = f(\vec{x})$. Dans ce cas, l'application f^{-1} existe et est définie sur tout \mathbb{R}^n par

$$f^{-1}(\vec{x}) = A^{-1}\vec{x},$$

où A^{-1} est l'inverse de la matrice de f .

Théorème : Si f est une application linéaire bijective, alors f transforme des vecteurs linéairement indépendants en vecteurs linéairement indépendants. En particulier, si $\vec{u}_1, \dots, \vec{u}_n$ est une base de \mathbb{R}^n , alors $f(\vec{u}_1), \dots, f(\vec{u}_n)$ est à nouveau une base de \mathbb{R}^n .

3.2 Applications linéaires orthogonales

Une application linéaire $f(\vec{x}) = A\vec{x}$ est dite **orthogonale** si

$$A^T = A^{-1}.$$

Dans ce cas, on dit que A est une **matrice orthogonale**.

Propriétés :

- Une matrice A est orthogonale si et seulement si ses colonnes (et ses lignes) forment une base orthonormée de \mathbb{R}^n .
- Le déterminant d'une matrice orthogonale est 1 ou -1 .
- Une application linéaire orthogonale préserve les distances entre les points. C'est-à-dire

$$\|f(\vec{x}) - f(\vec{y})\| = \|\vec{x} - \vec{y}\|,$$

et donc les angles.

- La composée de 2 applications linéaires orthogonales est aussi orthogonale.

Classification des applications linéaires orthogonales en dimensions 2 et 3 :

- Une application linéaire orthogonale de \mathbb{R}^2 est de l'un des 2 types suivants :

1. Une **rotation** autour de l'origine.

La matrice de la rotation d'angle α (dans le sens positif) est

$$\begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}.$$

2. Une **symétrie** par rapport à une droite passant par l'origine.

La matrice est alors

$$\begin{pmatrix} \cos(2\varphi) & \sin(2\varphi) \\ \sin(2\varphi) & -\cos(2\varphi) \end{pmatrix},$$

où φ est l'angle (orienté) entre cette droite et l'axe des x .

- Une application linéaire orthogonale de \mathbb{R}^3 est de l'un des 3 types suivants :
 1. une **rotation** autour d'une droite passant par l'origine ;
 2. une **symétrie** par rapport à un plan contenant l'origine ;
 3. une **rotation** autour d'une droite passant par l'origine, **suivie** de la **symétrie** par rapport au plan passant par l'origine et perpendiculaire à cette droite.

Exemples :

- La rotation d'angle α autour de l'axe des z a pour matrice

$$\begin{pmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

- La symétrie par rapport au plan xy a pour matrice

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

- La rotation d'angle α autour de l'axe des z , suivie de la symétrie par rapport au plan xy a pour matrice

$$\begin{pmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Deuxième partie

Statistiques appliquées

Chapitre 4

Statistique descriptive

But : Organiser et décrire (graphiquement et numériquement) des données.

Les données proviennent de l'observation d'une ou plusieurs caractéristiques d'un processus. Ces caractéristiques sont appelées des **variables**.

Type de variables :

- Une variable est dite **quantitative**, si elle prend des valeurs numériques que l'on peut additionner et multiplier entre elles. Sinon, on parle de variable **qualitative**.
- Une variable quantitative est dite **discrète**, si ses valeurs sont dénombrables (on peut donc les énumérer). Sinon on dit qu'elle est **continue**.
- Une variable qualitative est dite **ordinale**, si l'on peut ordonner ses valeurs (de manière croissante ou décroissante), sinon on dit qu'elle est **nominale**.

Exemple :

Nom	Age	Revenu mensuel (CHF)	Niveau d'études
Alice	21	3214.60	Maturité
Jean	32	6124.75	Bachelor
Jeanne	28	7891.10	Doctorat

La variable **Nom** est qualitative et nominale, la variable **Age** est quantitative et discrète, la variable **Revenu mensuel** est quantitative et continue, enfin la variable **Niveau d'étude** est qualitative et ordinale.

4.1 Données univariées

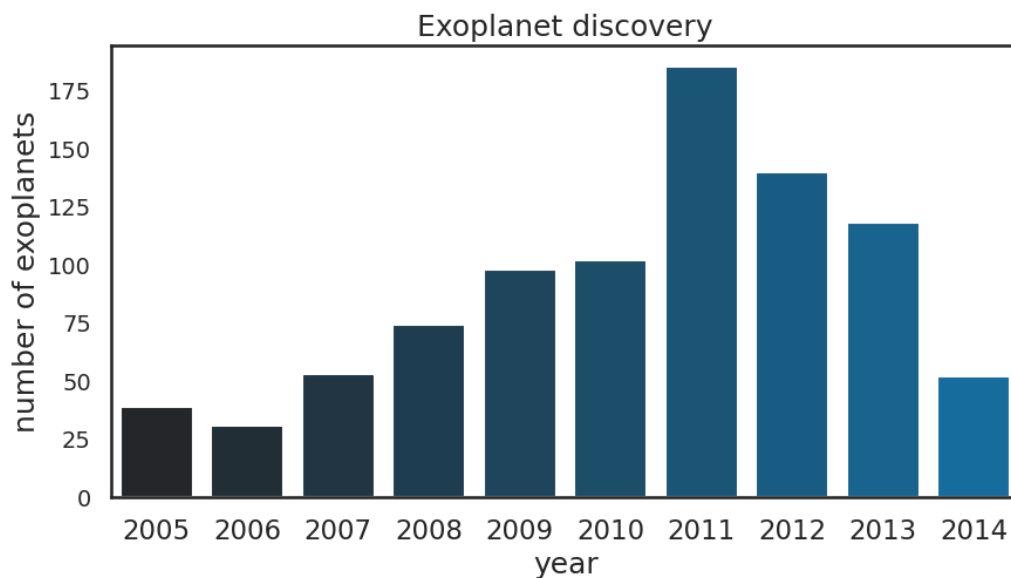
Dans cette partie, les exemples seront faits à partir de la base de données "planets". Cette base de données est disponible dans le package seaborn de Python et se présente comme suit¹.

method	number	orbital_period	mass	distance	year
Radial Velocity	1	269.300	7.10	77.40	2006
Radial Velocity	1	874.774	2.21	56.95	2008
Radial Velocity	1	763.000	2.60	19.84	2011

4.1.1 Représentations graphiques

Diagramme en bâtons

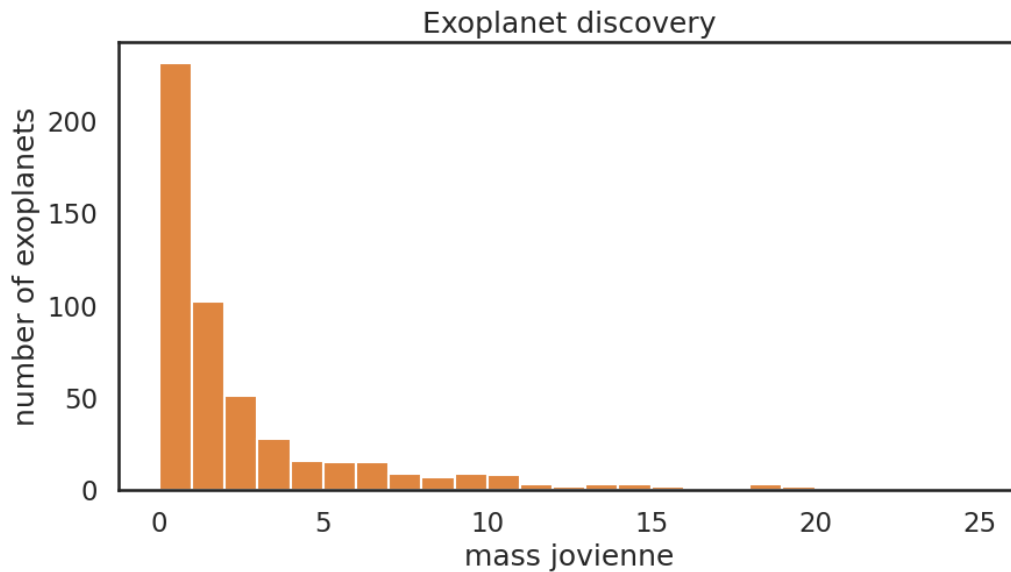
Pour des données x_1, x_2, \dots, x_n provenant d'une variable discrète x et si les données ne sont pas trop nombreuses, on peut représenter toute l'information. Habituellement, on représente, pour chaque valeur de x , son effectif (c'est-à-dire le nombre de fois que cette valeur apparaît). Pour cela on peut utiliser un diagramme en bâtons.



Histogramme

Pour une **variable continue**, on peut utiliser un histogramme. Il reprend le principe du diagramme en bâtons, mais en regroupant les valeurs dans des intervalles, appelés **classes**.

1. **method** : méthode de détection, **number** : nombre de planètes orbitant autour de la même étoile, **orbital_period** : période orbitale en jours, **mass** : en masse jovienne, **distance** : en parsec, **year** : année de découverte.



Remarques :

- En règle générale, chaque classe à la même largeur (sur l'axe des x).
- La somme de toutes les hauteurs des rectangles est égal aux nombre de données.
- On peut diviser la hauteur de chaque rectangle par le nombre total de données pour que la somme des hauteurs soit 1. Dans ce cas, la forme de l'histogramme ne change pas.
- On peut utiliser un histogramme pour une variable discrète, si les valeurs sont trop nombreuses pour être représentées individuellement.

4.1.2 Résumés numériques

On cherche à décrire un grand nombre de données x_1, x_2, \dots, x_n , provenant d'une variable quantitative x , à l'aide de quelques valeur numériques (mesures de position, de dispersion, ...)

Médiane, quartiles, boxplot

On considère une série de données **ordonnées**

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}.$$

La valeur $x_{(i)}$ représente la valeur à la i^{eme} position des données ordonnées. On parle de **rang** i .

Définition :

- La **médiane**, notée Q_2 (ou \tilde{x}) est définie comme suit :

$$Q_2 = \tilde{x} = \begin{cases} x_{(\frac{n+1}{2})} & \text{si } n \text{ est impair} \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} & \text{si } n \text{ est pair.} \end{cases}$$

C'est une valeur telle que la moitié des x_i est plus grande que \tilde{x} et la moitié des x_i est plus petite que \tilde{x} . Elle correspond au milieu de la liste ordonnée (mesure de position).

- Le **premier quartile**, noté Q_1 est défini par

$$Q_1 = x_{(\frac{n+3}{4})}$$

- Le **troisième quartile**, noté Q_3 est défini par

$$Q_3 = x_{(\frac{3n+1}{4})}$$

Lorsque le rang de Q_1 ou Q_3 n'est pas une valeur entière, on appelle :

- R_{inf} la valeur juste en dessous du rang obtenu.
- R_{sup} la valeur juste au dessus du rang obtenu.

On applique ensuite la règle suivante :

- Si le rang se termine par 0.25, alors

$$Q_1 = \frac{3R_{\text{inf}} + R_{\text{sup}}}{4}$$

- Si le rang se termine par 0.5, alors

$$Q_1 = \frac{R_{\text{inf}} + R_{\text{sup}}}{2}$$

- Si le rang se termine par 0.75, alors

$$Q_1 = \frac{R_{\text{inf}} + 3R_{\text{sup}}}{4}$$

On procède de même pour Q_3 .

Exemple :

On considère les données ordonnées suivantes

1, 11, 15, 19, 20, 24, 28, 34, 37, 47, 50, 61

- Pour la médiane, comme $n = 12$ est pair, on obtient :

$$Q_2 = \frac{x_{(6)} + x_{(7)}}{2} = \frac{24 + 28}{2} = 26.$$

- Pour le premier quartile, $\frac{n+3}{4} = 3.75$, $R_{\text{inf}} = 15$, $R_{\text{sup}} = 19$:

$$Q_1 = \frac{1 \cdot 15 + 3 \cdot 19}{4} = \frac{72}{4} = 18.$$

- Pour le troisième quartile, $\frac{3n+1}{4} = 9.25$, $R_{\text{inf}} = 37$, $R_{\text{sup}} = 47$:

$$Q_3 = \frac{3 \cdot 37 + 1 \cdot 47}{4} = \frac{158}{4} = 39.5.$$

On définit l'écart interquartile par $IQR = Q_3 - Q_1$.

Le **boxplot** est une représentation graphique des informations fournies par la médiane et les quartiles. Il est construit selon les règles suivantes :

- On dessine un rectangle (=boîte) allant du quartile Q_1 au quartile Q_3 .
- Dans le rectangle, un trait de séparation désigne la médiane Q_2
- On dessine, en partant de Q_1 un trait (=moustache) jusqu'à la première valeur $x_{(i)}$ telle que

$$x_{(i)} \geq Q_1 - 1.5 IQR.$$

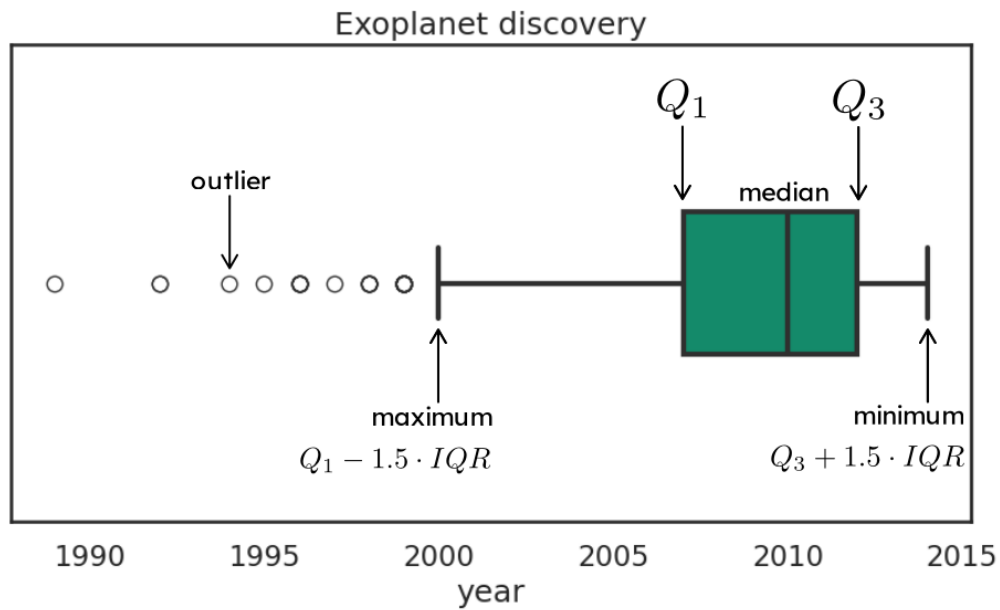
- On dessine, en partant de Q_3 un trait (=moustache) jusqu'à la dernière valeur $x_{(i)}$ telle que

$$x_{(i)} \leq Q_3 + 1.5 IQR.$$

- On représente par un point (ou un petit cercle) chaque valeur $x_{(i)}$ telle que

$$x_{(i)} < Q_1 - 1.5 IQR \quad \text{ou} \quad x_{(i)} > Q_3 + 1.5 IQR.$$

On appelle ces valeurs des **valeurs aberrantes** (en anglais : **outliers**).



Moyenne, variance, écart-type

- La **moyenne** des données est

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

(mesure de position).

- La **variance** (empirique) des données est

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right).$$

- L'**écart-type** (empirique) des données est

$$s = \sqrt{s^2}$$

(mesure de dispersion).

- Le **coefficient de variation**

$$CV(x) = \frac{s}{\bar{x}}$$

est une mesure relative (sans unités) de la dispersion.

Remarques :

- La somme des écarts à la moyenne vaut zéro : $\sum_{i=1}^n (x_i - \bar{x}) = 0$.
- Bien que la variance mesure aussi la dispersion des données, on lui préfère souvent l'écart-type, car il a les mêmes unités que les données (et que la moyenne).

4.1.3 Données standardisées

Si x_1, x_2, \dots, x_n sont des données de moyenne \bar{x} et d'écart-type s , alors la substitution

$$z_i = \frac{x_i - \bar{x}}{s}$$

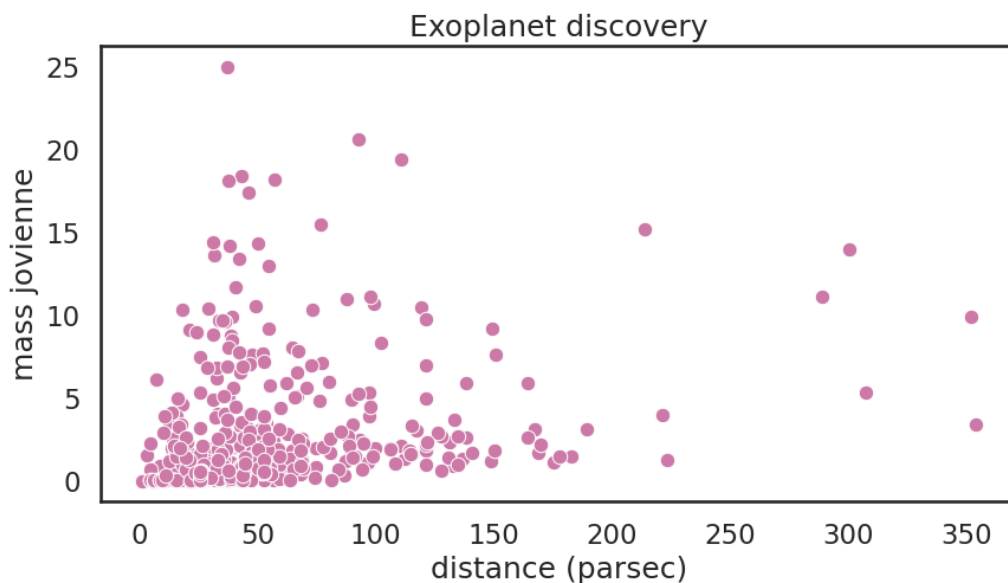
transforme les données x_i en des données z_i de moyenne 0 et d'écart-type 1, dites **standardisées** (ou **normalisées**).

4.2 Données bivariées

Pour visualiser une série de données du type

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

provenant de deux variables x et y , on peut représenter chacune des paires (x_i, y_i) comme un point du plan à la position indiquée par les coordonnées $x = x_i$ et $y = y_i$. On représente ainsi les données comme un **nuage de points** (ou **scatterplot**).



La position du **centre de gravité** du nuage est donnée par le point

$$C = (\bar{x}, \bar{y}) \quad \text{avec} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{et} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

La dispersion dans les directions x et y est décrite par les variances :

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{et} \quad s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

et les écarts-type

$$s_x = \sqrt{s_x^2} \quad \text{et} \quad s_y = \sqrt{s_y^2}.$$

La **covariance** (empirique), définie par

$$\text{Cov}(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n\bar{x} \cdot \bar{y} \right)$$

mesure la variation conjointe des variables x et y .

$\text{Cov}(x, y)$ est d'autant plus grande (en valeur absolue) que les points (x_i, y_i) sont proches d'une droite.

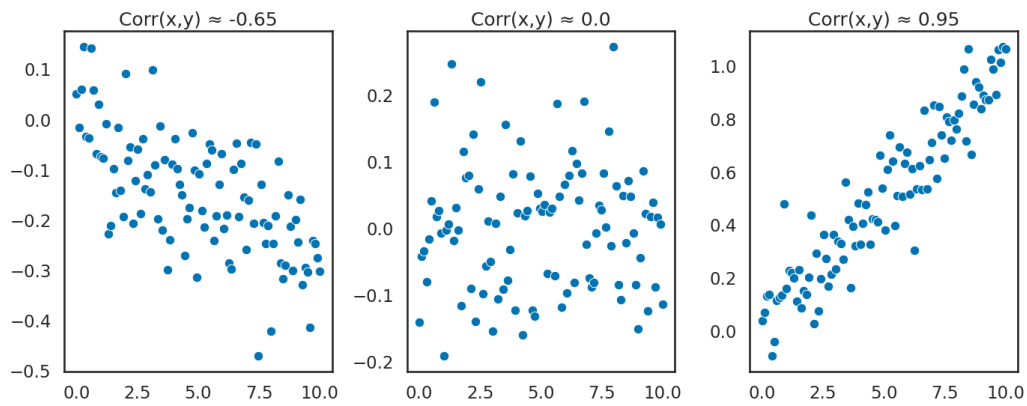
Le **coefficient de corrélation** (empirique) :

$$\text{Corr}(x, y) = \frac{\text{Cov}(x, y)}{s_x s_y}$$

est une mesure relative de la variation conjointe de x et y .

Remarque :

- $\text{Corr}(x, y) \in [-1, 1]$ a le même signe que $\text{Cov}(x, y)$
- $\text{Corr}(x, y) = 1 \Leftrightarrow$ les points sont alignés sur une droite de pente positive
- $\text{Corr}(x, y) = -1 \Leftrightarrow$ les points sont alignés sur une droite de pente négative.
- $\text{Corr}(x, y)$ proche de 0 signifie que les variations de x et de y ne sont pas dues à une relation linéaire.



4.3 Régression linéaire

Idée : Trouver "la meilleure droite" qui passe par une série de données bivariées : la **droite de régression**.

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n).$$

Dans l'idéal, on aimerait bien résoudre le système suivant :

$$\begin{cases} m \cdot x_1 + q = y_1 \\ m \cdot x_2 + q = y_2 \\ \vdots \\ m \cdot x_n + q = y_n. \end{cases}$$

Malheureusement, ce système n'admet aucune solution si les données ne sont pas parfaitement alignées.

Algorithme :

1. On définit les matrices A , B et X comme suit :

$$A = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}, \quad B = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \text{et} \quad X = \begin{pmatrix} m \\ q \end{pmatrix}.$$

L'équation

$$AX = B$$

correspond alors à notre système insoluble.

2. On résout, à la place, l'équation modifiée

$$A^T A X = A^T B \Rightarrow X = (A^T A)^{-1} A^T B$$

Comme $A^T A$ est inversible (si les x_i ne sont pas tous égaux), il y a une unique solution pour X : les paramètres de la droite que l'on cherche. C'est ce qu'on appelle la **solution du problème des moindres carrés**.

Exemple :

On considère les données

$$(1, 2), (2, 3), (3, 5), (4, 4)$$

et on y applique l'algorithme.

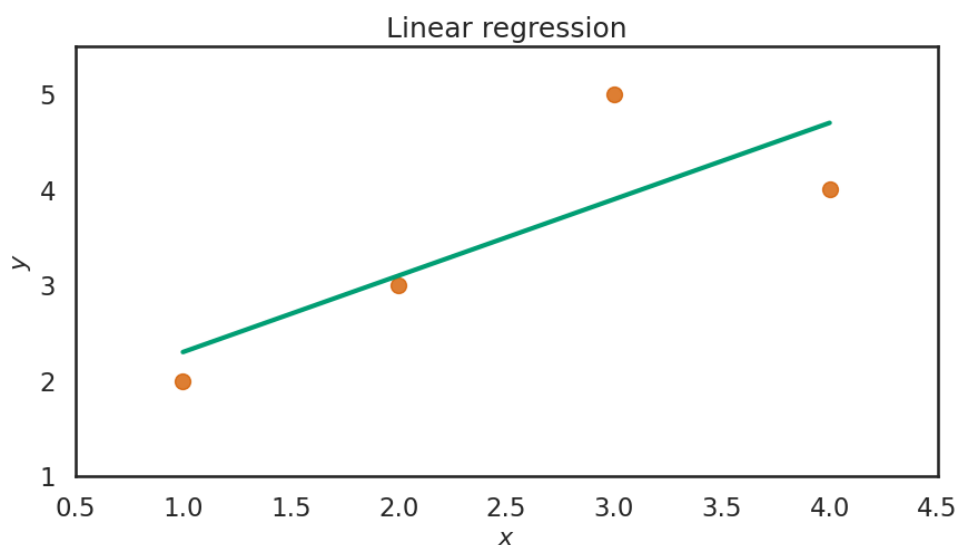
On commence par définir les matrices :

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 2 \\ 3 \\ 5 \\ 4 \end{pmatrix} \quad \text{et} \quad X = \begin{pmatrix} m \\ q \end{pmatrix}.$$

On résout

$$X = (A^T A)^{-1} A^T B = \frac{1}{20} \begin{pmatrix} 4 & -10 \\ -10 & 30 \end{pmatrix} \cdot \begin{pmatrix} 39 \\ 14 \end{pmatrix} = \begin{pmatrix} \frac{4}{5} \\ \frac{3}{2} \end{pmatrix}.$$

La droite de régression est donc $y = \frac{4}{5}x + \frac{3}{2}$.



Chapitre 5

Probabilités

5.1 Définitions et notations

Définitions :

- Une **expérience aléatoire** est un processus dont le résultat est influencé par le hasard.
- L'**ensemble fondamental** d'une expérience aléatoire, noté Ω , est l'ensemble de tous les résultats possibles de l'expérience.
- Un **événement** est un sous-ensemble de Ω (c'est-à-dire un ensemble de résultats possibles).

Rappel : Si A et B sont deux sous-ensembles de Ω , alors

- L'**intersection**, notée

$$A \cap B$$

est l'ensemble des éléments qui sont communs à A et à B .

- L'**union**, notée

$$A \cup B$$

est l'ensemble de tous les éléments qui se trouvent dans A ou dans B (ou dans les deux).

- La **différence**, notée

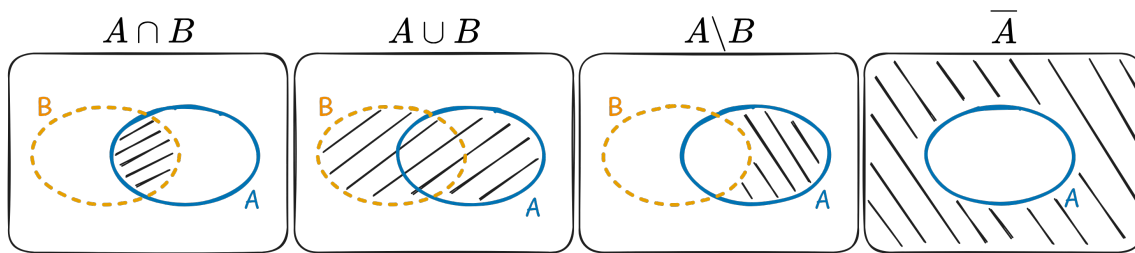
$$A \setminus B$$

est l'ensemble des éléments qui sont dans A mais pas dans B .

- Le **complément**, noté

$$\overline{A}$$

est l'ensemble des éléments qui ne sont pas dans A . (On le note aussi $\Omega \setminus A$).



On appelle **probabilité** une fonction P qui, à tout événement $A \subset \Omega$, fait correspondre un nombre réel $P(A) \in [0, 1]$ et qui satisfait

1. $P(\Omega) = 1$;
2. Si $A \cap B = \emptyset$, alors $P(A \cup B) = P(A) + P(B)$.

Remarque :

- $P(A)$ représente les "chances" que l'événement A se produise.
- Lorsque $A \cap B = \emptyset$, on dit que les événements sont **incompatibles**.

Propriétés :

À partir de ces axiomes, on peut déduire les propriétés suivantes :

1. $P(\emptyset) = 0$;
2. $P(\bar{A}) = 1 - P(A)$;
3. Pour tout $A, B \subset \Omega$: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Exemple : (Equiprobabilité sur un ensemble discret)

Pour un ensemble fondamental discret $\Omega = \{\omega_1, \dots, \omega_n\}$, la fonction P définie par :

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{nombre d'éléments de } A}{\text{nombre d'éléments de } \Omega}, \quad \text{pour tout } A \subset \Omega,$$

est une probabilité.

Dans ce cas, les résultats individuels ω_i ont tous la même probabilité :

$$P(\{\omega_i\}) = \frac{1}{n}, \quad \text{pour tout } i.$$

Cet exemple s'applique au lancer d'un dé (à 6 faces) équilibré :

- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- $P(\{1\}) = \dots = P(\{6\}) = \frac{1}{6}$.

5.2 Probabilité conditionnelle, indépendance et formule de Bayes

Définitions :

- Si A et B sont deux événements, la **probabilité conditionnelle** de A sachant B est définie par

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{si } P(B) \neq 0 \quad \text{et} \quad = 0 \quad \text{si } P(B) = 0.$$

- Deux événements A et B sont **indépendants** si

$$P(A \cap B) = P(A) \cdot P(B).$$

Remarque :

A et B sont indépendants si et seulement si

$$P(A|B) = P(A) \quad \text{et} \quad P(B|A) = P(B).$$

Formule de la probabilité totale :

Pour tous les événements A et B on a :

$$P(A) = P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B})$$

Formule de Bayes :¹

Pour tous les événements A et B on a :

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

1. Plus généralement, si B_1, \dots, B_n sont des événements tels que $B_i \cap B_j = \emptyset$ pour tout $i \neq j$ et $B_1 \cup \dots \cup B_n = \Omega$, alors

- la formule de la probabilité totale généralisée s'écrit :

$$P(A) = P(A|B_1) \cdot P(B_1) + \dots + P(A|B_n) \cdot P(B_n) \quad \text{pour tout événement } A.$$

- La formule de Bayes généralisée s'écrit :

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{P(A|B_1) \cdot P(B_1) + \dots + P(A|B_n) \cdot P(B_n)}, \quad \text{pour tout événement } A.$$

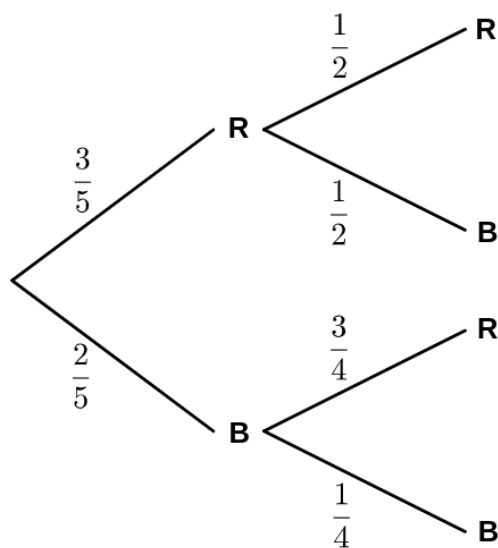
5.3 Arbre de probabilités

Souvent, il est possible de représenter une épreuve aléatoire sous la forme d'un **arbre de probabilités** afin de faciliter la représentation d'une épreuve aléatoire.

- Chaque étage de l'arbre représente une étape différente de l'épreuve aléatoire.
- On écrit les différentes probabilités de l'épreuve aléatoire sur les branches de l'arbre.
- À chaque étage, la somme des probabilités des branches partant du même sommet doit être égale à 1.
- Lorsqu'on suit un chemin sur l'arbre, on **multiplie** les probabilités.
- Lorsque on change de chemin, on **additionne** les probabilités.

Exemple :

On considère une urne contenant 3 jetons rouges et 2 jetons bleus, tous identiques. On tire successivement et sans remise deux jetons. Cette épreuve aléatoire peut être représentée par l'arbre de probabilité suivant.



- A : obtenir deux jetons rouges

$$P(A) = \frac{3}{5} \cdot \frac{1}{2} = \frac{3}{10}.$$

- B : obtenir exactement un jeton bleu

$$P(B) = \frac{3}{5} \cdot \frac{1}{2} + \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{10} + \frac{6}{20} = \frac{4}{5}.$$

Chapitre 6

Variables aléatoires

Une **variable aléatoire** (réelle) est une variable dont la valeur dépend du résultat d'une expérience aléatoire. Plus précisément, une variable aléatoire X est une fonction qui, à chaque résultat $\omega \in \Omega$, fait correspondre un nombre réel $x = X(\omega)$. On dit que x est une **réalisation** de la variable X .

6.1 Variables discrètes

X est une variable aléatoire **discrète**, si elle prend un nombre fini ou infini, mais dénombrable de valeurs x_1, x_2, \dots . Pour décrire la variable X , il suffit de donner pour chaque x_i la probabilité p_i que la variable prenne cette valeur, c'est-à-dire la probabilité de l'événement $\{X = x_i\}$. On l'abrège par :

$$p_i = P(\{X = x_i\}) = P(X = x_i).$$

Remarque : Les p_i doivent satisfaire à la condition :

$$\sum_i p_i = 1.$$

En effet, comme les événements $\{X = x_i\}$ sont deux à deux incompatibles :

$$\begin{aligned} \Omega = \bigcup_i \{X = x_i\} &\implies P(\Omega) = P\left(\bigcup_i \{X = x_i\}\right) \\ &= \sum_i P(\{X = x_i\}) = \sum_i p_i = 1. \end{aligned}$$

Les valeurs x_i et les probabilités p_i définissent la **loi de probabilité**¹ de X

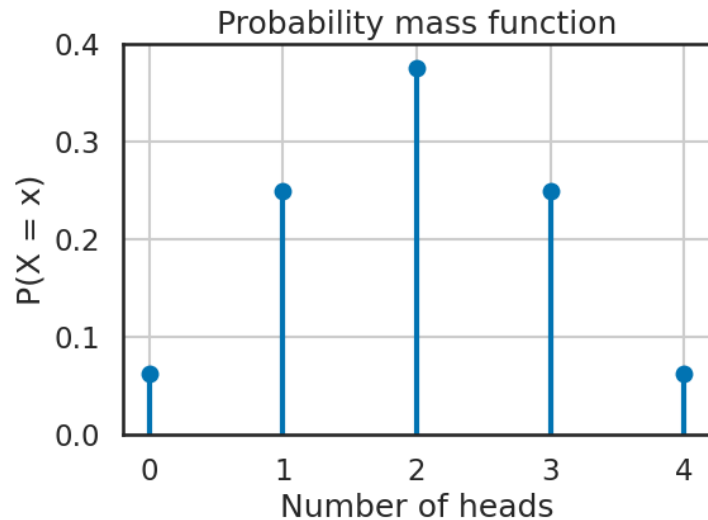
1. En anglais : **probability mass function (pmf)**

Exemple :

On considère la variable aléatoire X définie par

X = nombre de faces obtenus lors de 4 lancers d'une pièce non truquée.

x_i	p_i
0	$\frac{1}{16} = 0.0625$
1	$\frac{4}{16} = 0.25$
2	$\frac{6}{16} = 0.375$
3	$\frac{4}{16} = 0.25$
4	$\frac{1}{16} = 0.0625$



Une autre manière de définir une variable aléatoire X consiste à donner sa **fonction de répartition**²

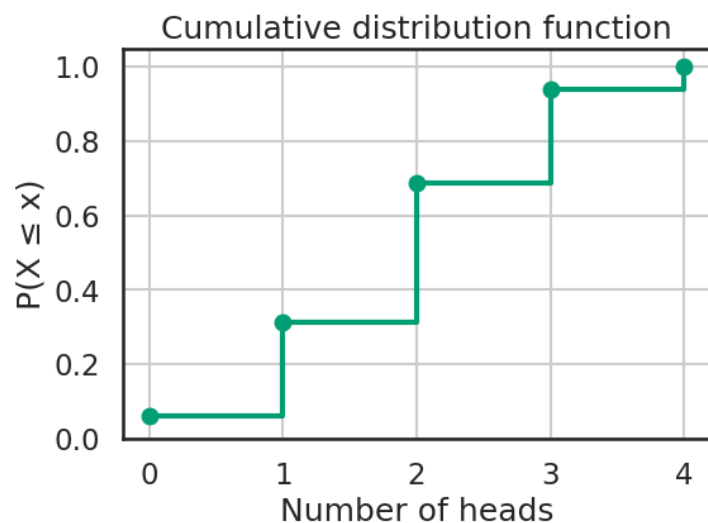
$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p_i, \quad x \in \mathbb{R}.$$

Elle donne la somme cumulée des probabilités. Il s'agit d'une fonction en escaliers qui est croissante de 0 à 1.

Exemple :

En reprenant la variable aléatoire de l'exemple précédent.

x_i	$F(x_i)$
0	$\frac{1}{16} = 0.0625$
1	$\frac{5}{16} = 0.3125$
2	$\frac{11}{16} = 0.6875$
3	$\frac{15}{16} = 0.9375$
4	$\frac{16}{16} = 1$



2. En anglais : **cumulative distribution function (cdf)**

Définitions :

Soit X une variable discrète prenant les valeurs x_i avec les probabilités p_i .

- L'espérance de X est définie par

$$E(X) = \sum_i x_i p_i = x_1 p_1 + x_2 p_2 + \dots$$

L'espérance est une moyenne des valeurs de X , pondérées par leurs probabilités d'apparition. Elle se situe au centre de masse de la distribution de probabilité de X et correspond à la valeur de x attendue **en moyenne** sur une grande quantité d'observations de X .

- La variance de X est définie par

$$\begin{aligned} \text{Var}(X) &= \sum_i (x_i - E(X))^2 p_i \\ &= \sum_i x_i^2 p_i - (E(X))^2 = E(X^2) - (E(X))^2. \end{aligned}$$

La variance mesure la **dispersion** des valeurs de X autour de l'espérance $E(X)$.

- L'écart-type de X est défini par

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

Exemple :

En reprenant la variable aléatoire de l'exemple précédent.

- L'espérance de X est donnée par

$$E(X) = \sum_i x_i p_i = 0 \cdot \frac{1}{16} + 1 \cdot \frac{4}{16} + 2 \cdot \frac{6}{16} + 3 \cdot \frac{4}{16} + 4 \cdot \frac{1}{16} = \frac{32}{16} = 2.$$

- La variance de X est donnée par

$$\begin{aligned} \text{Var}(X) &= \sum_i x_i^2 p_i - (E(X))^2 \\ &= 0^2 \cdot \frac{1}{16} + 1^2 \cdot \frac{4}{16} + 2^2 \cdot \frac{6}{16} + 3^2 \cdot \frac{4}{16} + 4^2 \cdot \frac{1}{16} - 4 \\ &= \frac{80}{16} - 4 = 1. \end{aligned}$$

6.1.1 Loi de Bernoulli

La loi de Bernoulli modélise la situation suivante :

On effectue une expérience aléatoire qui ne peut avoir que deux résultats :

- un succès, avec probabilité p ,
- un échec, avec probabilité $1 - p$.

La variable aléatoire X qui décrit le succès ou l'échec de cette épreuve aléatoire et que prend la valeur 1 (succès) ou 0 (échec) suit une loi de Bernoulli de paramètre p .

On écrit

$$X \sim \text{Ber}(p)$$

On peut l'écrire formellement de la manière suivante :

$$P(X = k) = \begin{cases} p & \text{si } k = 1 \\ 1 - p & \text{si } k = 0. \end{cases}$$

On peut calculer l'espérance d'une variable aléatoire de Bernoulli :

$$E(X) = 1 \cdot p + 0 \cdot (1 - p) = p,$$

ainsi que sa variance :

$$\text{Var}(X) = E(X^2) - E(X)^2 = 1^2 \cdot p - p^2 = p - p^2 = p \cdot (1 - p).$$

6.1.2 Loi binomiale

La loi binomiale modélise la situation suivante :

On répète n fois une expérience aléatoire qui ne peut avoir que deux résultats :

- un succès, avec probabilité p ,
- un échec, avec probabilité $1 - p$.

Les n essais sont indépendants et la probabilité de succès, p , à chaque essai est fixe et ne dépend pas du nombre d'essais.

La variable aléatoire S_n qui compte le nombre de succès pour cette série de n essais peut être décrite comme une somme de variables aléatoires qui suivent une loi de Bernoulli :

$$S_n = \sum_{i=1}^n X_i \quad \text{avec} \quad X_i \sim \text{Ber}(p).$$

On dit que S_n suit une **loi binomiale de paramètres n et p** .

On écrit

$$S_n \sim \text{Bin}(n, p)$$

S_n prend les valeurs $0, 1, 2, \dots, n$ et la probabilité d'obtenir k succès parmi les n essais vaut :

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{pour } k = 0, 1, \dots, n.$$

Ici,

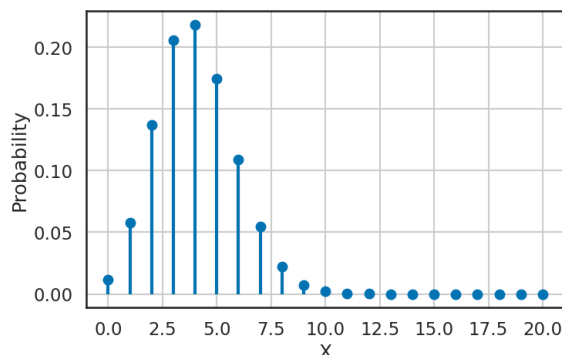
$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

est le **coefficient binomial** n sur k .

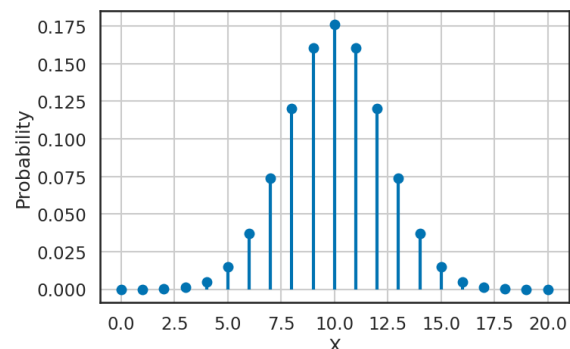
Un calcul montre que l'espérance et la variance de la loi binomiale de paramètres n et p sont :

$$E(X) = np \quad \text{et} \quad \text{Var}(X) = np(1-p).$$

Voici deux exemples de loi binomiales pour certaines valeurs de n et p .



$$n = 20, p = 0.2$$



$$n = 20, p = 0.5$$

6.1.3 Loi de Poisson

Une variable aléatoire X suit une **loi de Poisson** de paramètre $\lambda > 0$, si les valeurs de X sont tous les nombres naturels $k \in \mathbb{N}$ et si la probabilité que X prenne la valeur k vaut :

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{pour } k = 0, 1, 2, 3, \dots$$

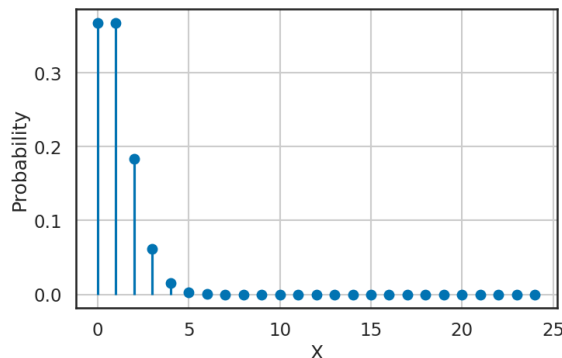
On écrit

$$X \sim \text{Poi}(\lambda)$$

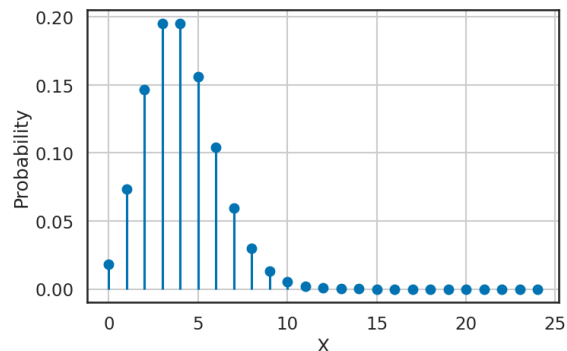
L'espérance et la variance de la loi de Poisson de paramètre λ sont :

$$E(X) = \lambda \quad \text{et} \quad \text{Var}(X) = \lambda.$$

Voici deux exemples de loi de Poisson pour certaines valeurs de λ .



$\lambda = 1$



$\lambda = 4$

Remarque :

On utilise la loi de Poisson pour modéliser le nombre d'événements se produisant dans un intervalle temporel (ou spatial) fixé, lorsque les événements ont peu de chance de se produire, mais qu'il y a de nombreuses opportunités. λ représente alors le nombre moyen d'événements qui se sont produits. La loi de Poisson est parfois appelée **loi des événements rares**.

Approximation d'une binomiale par une Poisson :

On considère

$$X \sim \text{Bin}(n, p)$$

avec

- $n > 10$
- $p < 0.05$.

Alors, on peut approximer X par une loi de Poisson avec $\lambda = n \cdot p$

$$X \sim \text{Poi}(n \cdot p)$$

6.2 Variable continue

X est une variable aléatoire **continue**, si elle peut prendre toutes les valeurs d'un intervalle.

Remarque :

Dans ce cas, comme la somme totale des probabilités doit valoir 1, on a

$$P(X = x) = 0$$

pour (presque) tout x .

Pour décrire la variable X , on ne peut donc pas se baser sur les probabilités

$$P(X = x).$$

On peut par contre utiliser la **fonction de répartition**

$$F(x) = P(X \leq x).$$

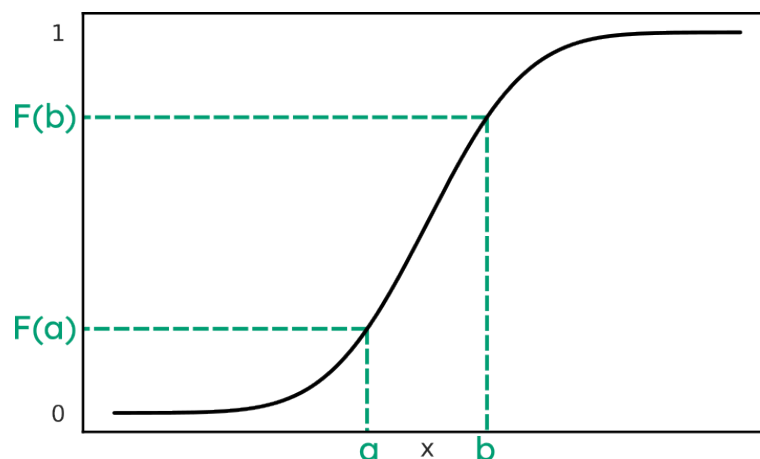
Remarque :

Lorsque X est continue, on a

$$P(X \leq x) = P(X < x).$$

Comme pour une variable discrète, la fonction $F(x)$ est croissante de 0 à 1. Dans le cas où X est continue, la fonction $F(x)$ est aussi continue. Connaissant $F(x)$, on peut alors calculer la probabilité que X prenne des valeurs dans un intervalle $[a, b]$:

$$P(a \leq X \leq b) = F(b) - F(a).$$



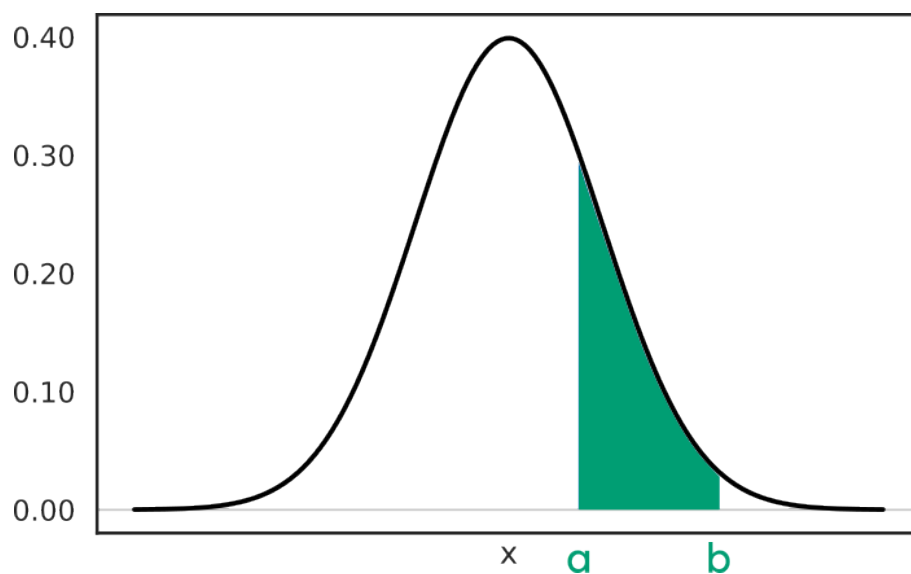
Dans la plupart des cas, la fonction de répartition $F(x)$ est aussi dérivable et sa dérivée

$$f(x) = F'(x)$$

s'appelle la **fonction de densité**³ de X .

C'est cette fonction $f(x)$ qui est utilisée habituellement pour décrire la variable X . cette fonction, la probabilité que la valeur de X se trouve dans un intervalle $[a, b]$ est alors l'aire sous la courbe $y = f(x)$ entre $x = a$ et $x = b$:

$$P(a \leq X \leq b) = \int_a^b f(x) dx.$$



Justification :

Par le théorème fondamental du calcul différentiel et intégral,

$$P(a \leq X \leq b) = F(b) - F(a) = \int_a^b F'(x) dx = \int_a^b f(x) dx.$$

Normalisation :

La condition que la somme des probabilités vaut 1 se traduit par :

$$\int_{-\infty}^{+\infty} f(x) dx = 1.$$

L'aire totale sous la courbe de densité est égale à 1.

3. En anglais : **probability density function (pdf)**

Définitions :

Comme dans le cas discret, on définit pour une variable continue X avec densité $f(x)$,

- l'espérance de X par

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx,$$

- la variance de X par

$$\text{Var}(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x) dx$$

- et l'écart-type de X par

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{\int_{-\infty}^{+\infty} (x - E(X))^2 f(x) dx}.$$

Remarque :

Comme dans le cas discret : $\text{Var}(X) = E(X^2) - (E(X))^2$, c'est-à-dire

$$\text{Var}(X) = \int_{-\infty}^{+\infty} x^2 f(x) dx - \left(\int_{-\infty}^{+\infty} x f(x) dx \right)^2.$$

6.2.1 Loi normale

La loi normale est l'une des plus importantes lois de probabilité continues. On l'utilise par exemple pour modéliser des variables physiques telles que la longueur, la masse, la température, ... Elle est donc souvent un bon modèle pour l'erreur de mesure qui résulte de l'accumulation de nombreux facteurs dont aucun ne prédomine.

Définition :

Soient $\mu \in \mathbb{R}$ et $\sigma > 0$ deux nombres réels. Une variable aléatoire X suit une **loi normale** de paramètres μ et σ , si sa fonction de densité est la suivante :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}.$$

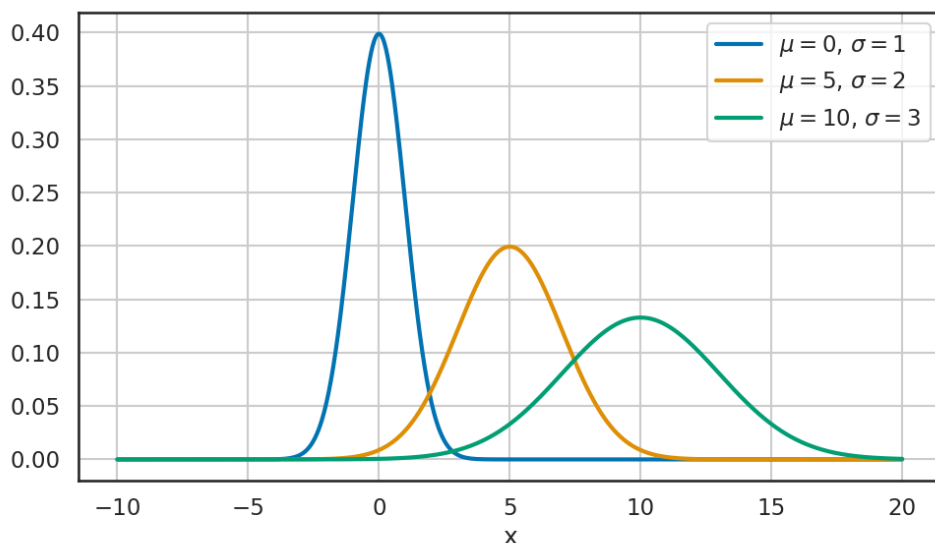
Sa **fonction de répartition** est alors :

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}} dt.$$

On écrit

$$X \sim \mathcal{N}(\mu, \sigma)$$

Le graphe de cette fonction est une courbe en forme de cloche, symétrique par rapport à son maximum $x = \mu$. Le paramètre σ donne la distance horizontale entre le maximum et chaque point d'inflexion. Si σ est petit, la courbe est étroite avec un maximum élevé, tandis que si σ est grand, la courbe est aplatie et étalée.



On peut montrer que si X suit une loi normale de paramètres μ et σ , l'espérance et la variance de X sont

$$E(X) = \mu \quad \text{et} \quad \text{Var}(X) = \sigma^2.$$

Loi normale centrée réduite :

La loi normale de paramètres $\mu = 0$ et $\sigma = 1$, notée

$$Z \sim \mathcal{N}(0, 1)$$

est appelée **loi normale centrée réduite** (ou **loi normale standard**).

La fonction de densité est

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

et sa fonction de répartition

$$\Phi(z) = \int_{-\infty}^z \varphi(t) dt.$$

Calcul de probabilité pour la loi normale :

La fonction de densité $f(x)$ de la loi normale ne possède pas de primitive élémentaire. Par conséquent, le calcul des probabilités pour une loi normale requiert l'utilisation d'une table ou d'un logiciel qui approxime ces valeurs par intégration numérique.

Standardisation :

Pour réduire la complexité de ce travail, on remarque que si X suit une loi normale $\mathcal{N}(\mu, \sigma)$, alors la variable standardisée

$$Z = \frac{X - \mu}{\sigma}$$

suit une loi normale $\mathcal{N}(0, 1)$.

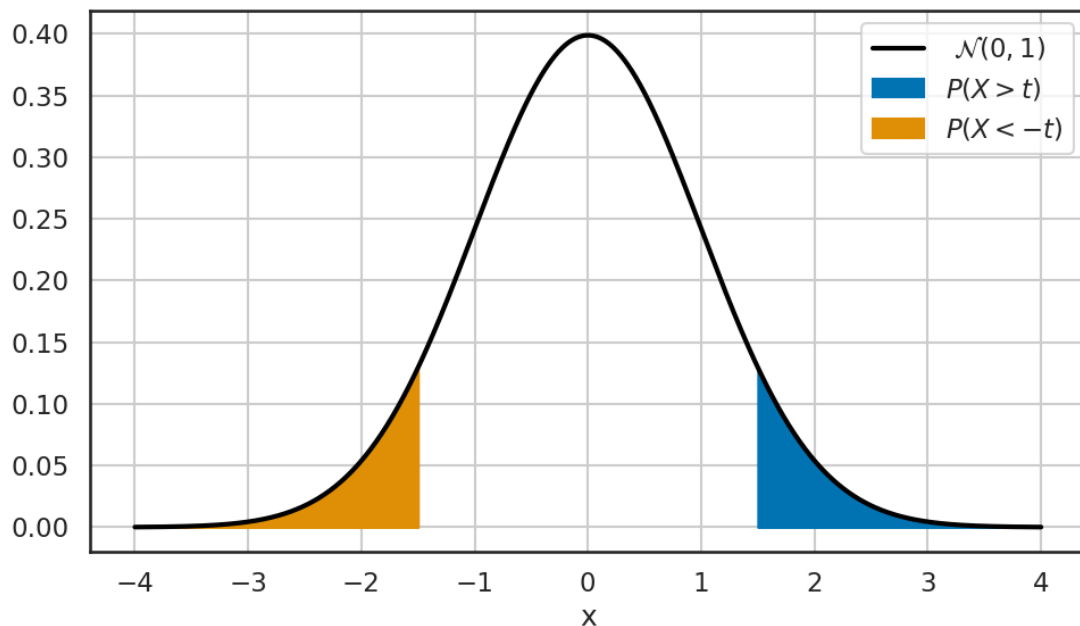
Le calcul d'une probabilité pour X se ramène donc à un calcul pour Z de la manière suivante

$$P(X \leq a) = P\left(Z \leq \frac{a - \mu}{\sigma}\right) = \Phi\left(\frac{a - \mu}{\sigma}\right), \quad \text{pour tout } a \in \mathbb{R}.$$

Symétrie :

Comme la fonction $\varphi(z)$ est paire, c'est-à-dire : $\varphi(-z) = \varphi(z)$ pour tout $z \in \mathbb{R}$, on a les symétries suivantes pour l'intégrale de φ :

$$\begin{aligned}\Phi(-a) &= P(Z \leq -a) = P(Z \geq a) \\ &= 1 - P(Z \leq a) = 1 - \Phi(a), \quad \text{pour tout } a \in \mathbb{R}.\end{aligned}$$



6.2.2 Théorème limite-central

Voici un théorème très important en probabilité, il montre comment la loi normale permet d'approximer des sommes et des moyennes de nombreuses lois de probabilité.

Théorème limite-central :

Soit X_1, X_2, \dots, X_n une suite de variables aléatoires de même loi que X , alors pour n grand :

$$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq t\right) \approx \Phi(t) = P(Z \leq t).$$

Avec

$$S_n = \sum_{i=1}^n X_i, \quad \sigma = \sqrt{\text{Var}(X)} \quad \text{et} \quad \mu = E(X).$$

Application : Approximation de la loi binomiale :

Soit

$$X \sim \text{Bin}(n, p)$$

Alors, pour n grand,

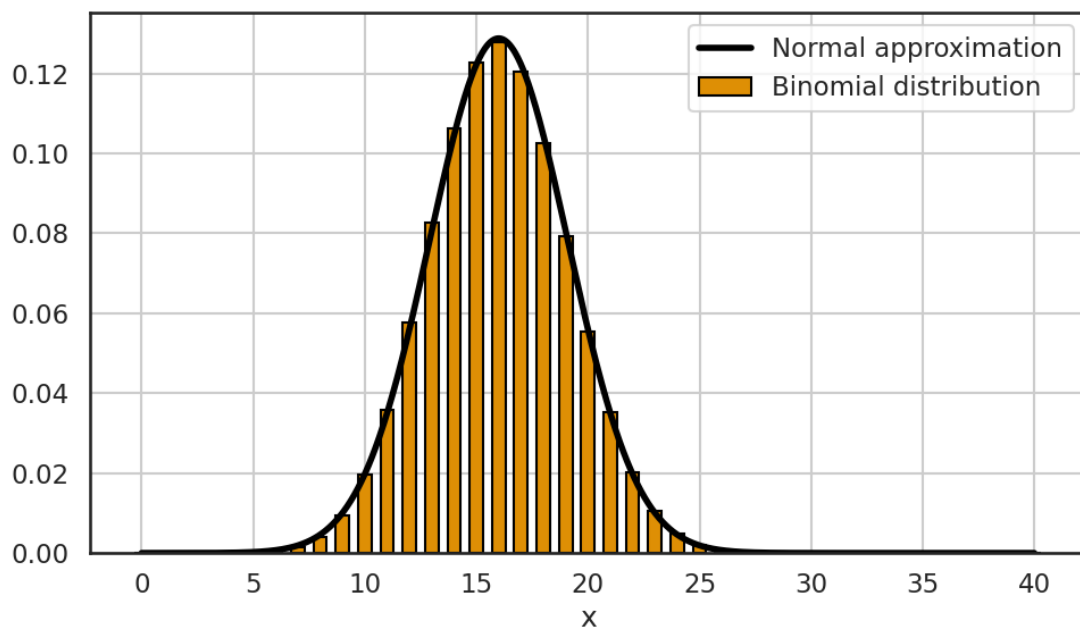
$$P(X \leq s) = P\left(\frac{X - np}{\sqrt{np(1-p)}} \leq \frac{s - np}{\sqrt{np(1-p)}}\right) \approx \Phi(t) = P(Z \leq t),$$

avec

$$t = \frac{s - np}{\sqrt{np(1-p)}}.$$

Remarque :

L'approximation est bonne si $np(1-p) > 9$.



Loi binomiale $n = 40$, $p = 0.4$ et loi normale $\mu = 16$, $\sigma = \sqrt{9.6}$

Chapitre 7

Introduction aux tests statistiques

Dans de nombreuses situations concrètes, on cherche à comparer deux alternatives sans pouvoir vérifier directement laquelle est meilleure. Par exemple :

- Un nouveau médicament est-il plus efficace que l'ancien ?
- Une nouvelle méthode d'enseignement est-elle supérieure à l'ancienne ?
- Est-ce qu'une tartine tombe réellement plus souvent du côté de la confiture ?

Pour répondre à ce type de question, on procède à un **test statistique**, à partir d'un **échantillon** issu d'une **variable aléatoire**.

Le plus souvent, on applique un test statistique lorsque l'on observe une différence/anomalie par rapport au comportement attendu/habituel.

Les hypothèses du test :

- L'**hypothèse nulle** H_0 représente le **comportement attendu** (habituel). Elle suppose que la différence constatée est due au hasard.
- L'**hypothèse alternative** (souvent notée H_1) correspond à l'existence d'un **véritable effet** ou d'une différence significative.

Comment décide-t-on d'accepter ou de rejeter H_0 ?

1. On calcule une **statistique de test** T , c'est-à-dire une fonction de l'échantillon.
2. On observe la valeur prise par T pour les données recueillies.
3. On évalue la compatibilité de cette valeur avec l'hypothèse H_0 :
 - si la valeur est compatible avec H_0 : on **accepte** H_0 ,
 - sinon : on **rejette** H_0 en faveur de l'hypothèse alternative.

Exemple :

On s'intéresse au problème capital de la "chute d'une tartine". Une affirmation fréquente prétend qu'une tartine a tendance à tomber du mauvais côté, c'est-à-dire du côté confiture.

- Désignons par p la probabilité pour que, lors d'une chute, une tartine donnée tombe du côté confiture.
- On choisit pour l'hypothèse nulle

$$H_0 : p = 0.5$$

ce qui signifie que la confiture n'a aucune incidence sur le côté où tombe la tartine.

- On choisit pour l'hypothèse alternative

$$H_1 : p > 0.5$$

ce qui signifie que la tartine a significativement plus de chance de tomber du côté de la confiture.

On verra plus tard comment prendre la décision entre accepter et rejeter H_0 .

Erreur de première et deuxième espèce :

En décidant d'accepter ou de rejeter H_0 , on peut commettre une erreur. On distingue 2 types d'erreurs :

- Rejeter l'hypothèse H_0 alors qu'elle est vraie. Cette erreur est appelée **erreur de type I** ou **erreur de première espèce**.
- Accepter l'hypothèse H_0 , alors qu'elle est fausse. Cette erreur est appelée **erreur de type II** ou **erreur de deuxième espèce**.

Définition :

- La probabilité de commettre une erreur de première espèce est notée α et on l'appelle le **niveau de signification** (ou **risque**) du test.
- La probabilité de commettre une erreur de deuxième espèce est notée β et la probabilité $(1 - \beta)$ est appelée la **puissance** du test.

Remarque :

La notion de puissance d'un test ne sera pas abordée dans ce cours.

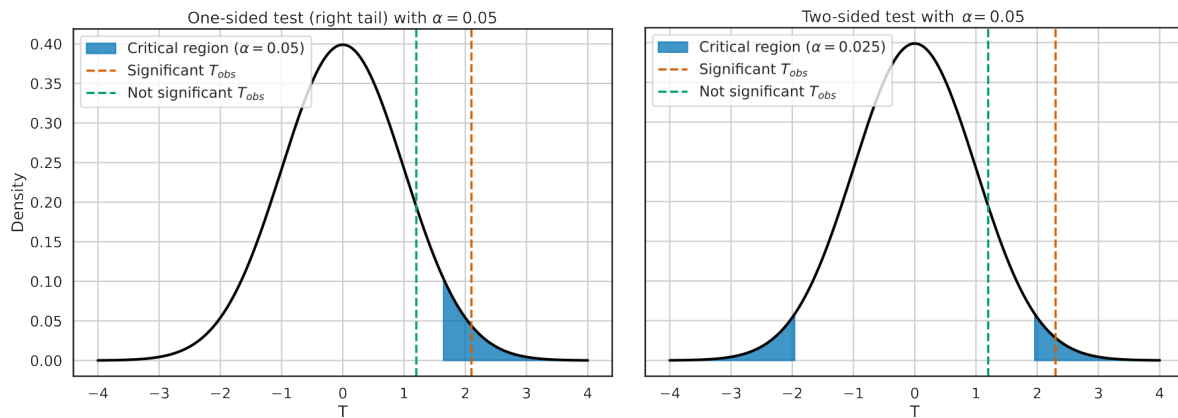
On peut résumer les erreurs à l'aide du tableau suivant :

	H_0 vraie	H_0 fausse
Rejeter H_0	Erreur de type I Proba. = α	Décision correcte Proba. = $1 - \alpha$
Accepter H_0	Décision correcte Proba. = $1 - \beta$	Erreur de type II Proba. = β

Procédure générale :

La procédure générale d'un test d'hypothèses est la suivante :

1. On détermine l'hypothèse H_0 et on fixe le niveau de signification α . (Habituellement : $\alpha = 0.05$).
2. On détermine, en fonction de α le **domaine critique** (ou **domaine de rejet**). Il est formé de l'ensemble des valeurs de la statistique de test T pour lesquelles on rejette H_0 .
 - Dans le cas d'un test **bilatéral**, le domaine critique est composé de deux parties : la partie formée par le $\frac{\alpha}{2} \cdot 100 \%$ de plus grandes valeurs de T et la partie formée par le $\frac{\alpha}{2} \cdot 100 \%$ des plus petites valeurs. On note $T_{1-\alpha/2}$ la valeur de T qui délimite ce domaine des deux côtés.
 - Dans le cas d'un test **unilatéral**, le domaine critique est composé d'une seule partie (le $\alpha \cdot 100 \%$ des valeurs les plus grandes ou les plus petites, selon ce que l'on cherche à tester). On note $T_{1-\alpha}$ la valeur de T qui délimite ce domaine.
3. En utilisant l'échantillon, on calcule la valeur que prend la statistique de test T pour cet échantillon qu'on appelle T_{obs} .
4. On prend la décision : Si la valeur T_{obs} se trouve dans le domaine critique, on rejette H_0 . Sinon, on accepte H_0 .



Exemple de test unilatéral à droite et bilatéral.

Voici une petite table pour la loi normale avec quelques valeurs souvent utilisées pour α .

α	5%	1%
$T_{1-\alpha}$	1.645	2.326
$T_{1-\alpha/2}$	1.96	2.576

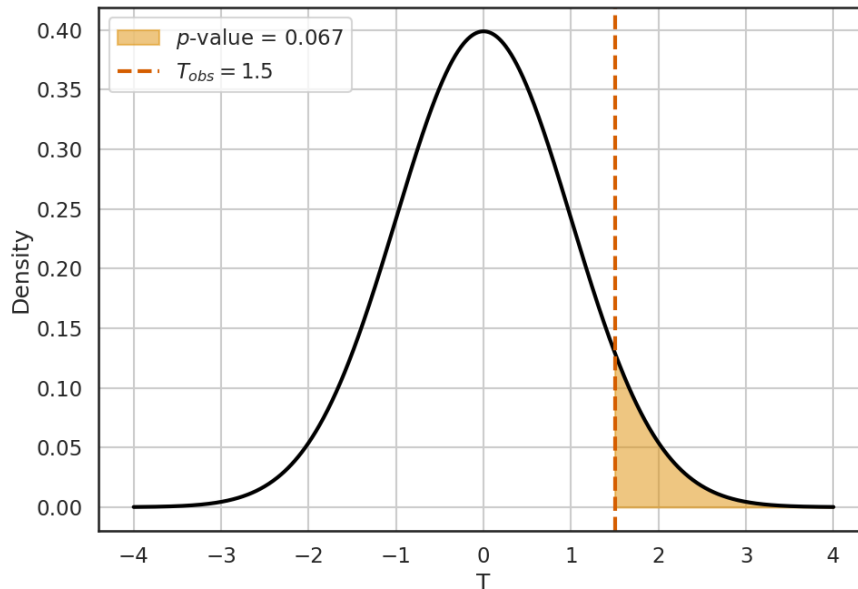
Définition :

Soit T une statistique de test dont la loi sous l'hypothèse nulle H_0 est connue. Soit T_{obs} la valeur observée de cette statistique. On définit la **p-valeur** comme la probabilité associée à T_{obs} .

Plus formellement :

Type de test	Définition de la p-valeur
Test bilatéral	$2 \cdot P(T \geq T_{\text{obs}} \mid H_0) = 2 \cdot (1 - F(T_{\text{obs}}))$
Test unilatéral à droite	$P(T \geq T_{\text{obs}} \mid H_0) = 1 - F(T_{\text{obs}})$
Test unilatéral à gauche	$P(T \leq T_{\text{obs}} \mid H_0) = F(T_{\text{obs}})$

où $F(t) = P(T \leq t)$ est la fonction de répartition cumulative de la statistique T sous H_0 .



7.1 Test de proportion

Les tests de proportions servent à comparer la proportion observée dans un échantillon à une proportion théorique.

Statistique de test :

Soit un échantillon de taille n dans lequel on observe S_n succès. La proportion observée est

$$\hat{p} = \frac{S_n}{n}.$$

On souhaite tester l'hypothèse nulle

$$H_0 : p = p_0,$$

où p_0 est une proportion théorique.

La statistique de test utilisée est

$$T = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{S_n - np_0}{\sqrt{np_0(1-p_0)}}.$$

Sous l'hypothèse nulle H_0 , et pour un échantillon suffisamment grand, T suit approximativement une loi normale centrée réduite,

$$T \sim \mathcal{N}(0,1).$$

Hypothèses et rejet :

Type de test	Hypothèses	Règle de rejet
Test bilatéral	$H_0 : p = p_0$	Rejeter H_0 si
	$H_1 : p \neq p_0$	$ T_{obs} > T_{1-\alpha/2}$
Test unilatéral à droite	$H_0 : p \leq p_0$	Rejeter H_0 si
	$H_1 : p > p_0$	$T_{obs} > T_{1-\alpha}$
Test unilatéral à gauche	$H_0 : p \geq p_0$	Rejeter H_0 si
	$H_1 : p < p_0$	$T_{obs} < T_{1-\alpha}$

Exemple :

On lance 2000 tartines, dont 1040 tombent du côté de la confiture. Soit p la probabilité que la tartine tombe du côté de la confiture. Vérifiez au niveau $\alpha = 0.05$ la légende urbaine qui affirme que les tartines tombent plus souvent du côté de la confiture.

$$H_0 : p = p_0 = 0.5$$

$$H_1 : p > p_0 = 0.5$$

La statistique correspondant aux observations est donnée par

$$T_{obs} = \frac{S_n^{obs} - np_0}{\sqrt{np_0(1-p_0)}} = \frac{1040 - 2000 \cdot 0.5}{\sqrt{2000 \cdot \frac{1}{4}}} = 1.79.$$

Comme le test est unilatéral et le niveau est fixé à $\alpha = 0.05$

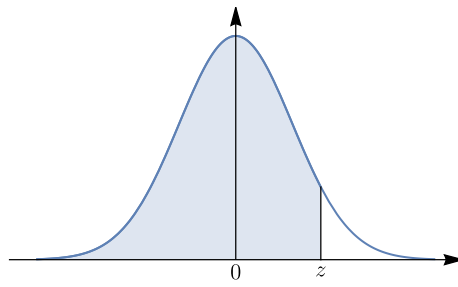
$$T_{0.95} = 1.645 \implies T_{obs} = 1.79 > 1.645$$

Donc on rejette H_0 .

La p-valeur est donnée par :

$$P(T \geq T_{obs} \mid H_0) = P(T \geq 1.79 \mid H_0) = 1 - \Phi(1.79) = 1 - 0.963 = 0.037.$$

Loi normale $\mathcal{N}(0, 1)$



$$\Phi(z) = P(Z \leq z)$$

[illegible]