Probabilités et Statistiques S4
Université de Strasbourg
L2 M/ MPA/I/ CMI

Davide Giraudo

21 février 2023

Table des matières

1. 10 heures de cours magistral, vendredi de 8h à 10h sur 5 séances.
2. 17 heures de travaux dirigés.
3. 2 contrôles continus de 1 heure 30 chacun

   .
4. Note :
   - assiduité $0,1$ (présence à tous les TD),
   - $0,45$ pour chaque contrôle continu.

Table des matières

1 Espaces probabilisés

1.1 Terminologie

1.1.1 Réalisation, événements

On considère le lancer d'un dé à six faces.

- Un résultat possible est appelé une réalisation ; il est noté ω.
- L'ensemble de tous les résultats possibles est appelé espace d'états et est notée Ω. Ici : $\Omega = \{1,2,3,4,5,6\}$.
- Un événement est une collection de résultats. Par exemple, si A est l'événement "le résultat du lancer est pair", $A = \{2,4,6\} \subset \Omega$.
- L'événement contraire de A est noté A^c ; dans l'exemple précédent,

$$A^c = \Omega \setminus A = \{1,2,3,4,5,6\} \setminus \{2,4,6\} = \{1,3,5\}.$$

Événements particuliers

- L'événement certain : Ω.
- L'événement impossible : $\emptyset$.
- L'événement A et B : $A \cap B$.
- L'événement A ou B ("ou" non exclusif) : $A \cup B$.

Définition 1.1. *Les événements A et B sont incompatibles si $A \cap B = \emptyset$.*

1.2 Ensembles et tribus

1.2.1 Ensembles finis/ ensembles dénombrables

Définition 1.2 (Ensemble fini)**.** *Un ensemble Ω est fini s'il existe un entier $n \in \mathbb{N}^*$ et une bijection de Ω dans $\{1, \dots, n\}$. On dit alors que* $\operatorname{card}(\Omega) = n$.

Définition 1.3 (Ensemble dénombrable)**.** *On dit que Ω est dénombrable s'il existe une bijection de Ω dans $\mathbb{N}$.*

Définition 1.4 (Ensemble au plus dénombrable)**.** *Un ensemble est dit au plus dénombrable s'il est fini ou dénombrable.*

Exemples 1.1.

- $\mathbb{Z}$, $\mathbb{N} \times \mathbb{N}$ et $\mathbb{Q}$ sont dénombrables.
- L'intervalle $[0, 1]$ n'est pas dénombrable.

1.2.2 Rappel sur les opérations ensemblistes

Définition 1.5 (Partition)**.** *On dit que $(A_i)_{i\in I}$ forme une partition de A si pour tous $i, i' \in I$ tels que $i \neq i'$, $A_i \cap A_{i'} = \emptyset$ et $\bigcup_{i\in I} A_i = A$.*

Proposition 1.6. *Soient Ω un ensemble, I un ensemble d'indices, $(A_i)_{i\in I}$ une collection de sous-ensembles de Ω et $B \subset \Omega$. Alors*

$$\left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c, \quad \left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c \tag{1.2.1}$$

$$B \cap \bigcap_{i\in I} A_i = \bigcap_{i\in I} (B \cap A_i), \quad B \cap \bigcup_{i\in I} A_i = \bigcup_{i\in I} (B \cap A_i). \tag{1.2.2}$$

1.2.3 Tribus

Définition 1.7 (Tribu)**.** *Soit $\mathcal{F}$ une collection de parties de Ω. On dit que $\mathcal{F}$ est une tribu si*

1. *$\Omega \in \mathcal{F}$,*
2. *si $A \in \mathcal{F}$, alors $A^c = \Omega \setminus A \in \mathcal{F}$ et*
3. *si $(A_n)_{n\in\mathbb{N}}$ est une suite d'éléments de $\mathcal{F}$, alors $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.*

Exemples 1.2.

- $\mathcal{F} = \mathcal{P}(\Omega) = \{A : A \subset \Omega\}$ est une tribu,
- de même que $\{\emptyset, \Omega\}$ (dite tribu triviale)
- ou encore $\{\emptyset, A_0, A_0^c, \Omega\}$.
- Mais $\{\emptyset, A_0, \Omega\}$ n'est pas une tribu si $A_0 \neq \emptyset$ et $A_0 \neq \Omega$.

1.3 Probabilités

Définition 1.8 (Probabilité)**.** *Une probabilité* $\mathbb{P}$ *est une fonction définie sur une tribu* $\mathcal{F}$ *à valeurs dans* $[0,1]$ *telle que*

1. $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$ *et*
2. *si* $(A_i)_{i \in I}$ *est une collection au plus dénombrable d'éléments de* $\mathcal{F}$ *deux à deux disjoints* ($A_i \cap A_{i'} = \emptyset$ *si* $i \neq i'$), *alors*

$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mathbb{P}(A_i). \tag{1.3.1}$$

[cette propriété est appelée σ*-additivité]*

On peut interpréter $\mathbb{P}(A)$ de la façon suivante : si on répète l'expérience aléatoire considérée N fois et on note $N(A)$ le nombre de fois où A se réalise, alors $\mathbb{P}(A) \approx \frac{N(A)}{N}$.

Définition 1.9 (Espaces probabilisés)**.** *Soient* Ω *un ensemble,* $\mathcal{F}$ *une tribu sur* Ω *et* $\mathbb{P} \colon \mathcal{F} \to [0,1]$ *une probabilité. Le triplet* $(\Omega, \mathcal{F}, \mathbb{P})$ *est appelé espace probabilisé.*

Exemple 1.10. Lancer d'un dé équilibré : $\Omega = \{1,2,3,4,5,6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ et $\mathbb{P}(\{\omega\}) = 1/6$ pour tout $\omega \in \Omega$.

Un événement $A \in \mathcal{F}$ est :

- presque sûr si $\mathbb{P}(A) = 1$;
- négligeable si $\mathbb{P}(A) = 0$.

Proposition 1.11 (Propriétés des mesures de probabilité)**.** *1.* $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ *;*

2. *Si* $A, B \in \mathcal{F}$ *et* $A \subset B$, *alors* $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$ *;*
3. *monotonie : si* $A, B \in \mathcal{F}$ *et* $A \subset B$, *alors* $\mathbb{P}(A) \leqslant \mathbb{P}(B)$ *;*
4. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ *;*
5. *soit* $(A_i)_{i \in I}$ *une collection au plus dénombrable d'ensembles deux à deux disjoints et telle que* $\sum_{i \in I} \mathbb{P}(A_i) = 1$. *Alors pour* $B \in \mathcal{F}$,

$$\mathbb{P}(B) = \sum_{i \in I} \mathbb{P}(A_i \cap B). \tag{1.3.2}$$

6. *Soit* $(A_n)_{n \in \mathbb{N}}$ *une suite croissante d'événements* ($A_n \subset A_{n+1}$ *pour tout* n). *Alors*

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \to \infty} \mathbb{P}(A_n) \qquad \textit{convergence monotone.} \tag{1.3.3}$$

Démonstration. 1. On applique (1.3.1) à $I = \{1,2\}$, $A_1 = A$, $A_2 = A^c$: $\mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$ et on conclut par le fait que $A \cup A^c = \Omega$ et $\mathbb{P}(\Omega) = 1$.

2. On applique (1.3.1) à $I = \{1,2\}$, $A_1 = A$, $A_2 = B \setminus A$.

3. C'est une conséquence directe de la propriété précédente car $\mathbb{P}(B \setminus A) \geqslant 0$.
4. On applique (1.3.1) à $I = \{1, 2, 3\}$, $A_1 = A \setminus (A \cap B)$, $A_2 = B \setminus (A \cap B)$ et $A_3 = A \cap B$.
5. On applique (1.3.1) avec A_i remplacé par $A_i \cap B$.
6. On applique (1.3.1) avec $I = \mathbb{N}$, et A_i remplacé par $A_i \setminus A_{i-1}$ pour $i \geqslant 1$.

□

1.4 Probabilité sur un ensemble fini ou dénombrable

1.4.1 Expression de $\mathbb{P}(A)$ en fonction de $\mathbb{P}(\{\omega\})$

Dans cette sous-section, on suppose Ω au plus dénombrable. Dans ce cas, $\mathbb{P}$ est déterminée uniquement par la collection $(\mathbb{P}(\{\omega\}))_{\omega\in\Omega}$. Plus précisément :

Lemme 1.12. *Pour tout $A \in \mathcal{F}$,*

$$\mathbb{P}(A) = \sum_{\omega\in A} \mathbb{P}(\{\omega\}). \tag{1.4.1}$$

Démonstration. La collection $(A \cap \{\omega\})_{\omega\in\Omega}$ est constituée d'événement deux à deux disjoints. Par σ-additivité,

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{\omega\in\Omega} (A \cap \{\omega\})\right) = \sum_{\omega\in\Omega} \mathbb{P}(A \cap \{\omega\}). \tag{1.4.2}$$

De plus,

$$\mathbb{P}(A) = \sum_{\omega\in\Omega} \mathbb{P}(A \cap \{\omega\}) = \sum_{\omega\in A} \mathbb{P}(A \cap \{\omega\}) + \sum_{\omega\in A^c} \mathbb{P}(A \cap \{\omega\}) \tag{1.4.3}$$

et

$$\mathbb{P}(A \cap \{\omega\}) = \begin{cases} \mathbb{P}(\{\omega\}) & \text{si } \omega \in A, \\ 0 & \text{si } \omega \in A^c, \end{cases} \tag{1.4.4}$$

d'où le résultat.

□

Définition 1.13. *Si Ω fini et pour tout $\omega \in \Omega$, $\mathbb{P}(\{\omega\}) = 1/\operatorname{card}(\Omega)$, alors on dit que $\mathbb{P}$ est la probabilité uniforme.*

Par le Lemme 1.12, si $\mathbb{P}$ est la probabilité uniforme, alors pour tout $A \subset \Omega$,

$$\mathbb{P}(A) = \frac{\operatorname{card}(A)}{\operatorname{card}(\Omega)} = \frac{\text{nombre de cas favorables à } A}{\text{nombre de cas possibles}}. \tag{1.4.5}$$

1.5 Modélisation

On souhaite donner un modèle probabiliste pour la somme des faces d'un lancer de deux dés équilibrés. L'espace des résultats possibles est $\Omega_1 = \{2, 3, \ldots, 12\}$. Soit $\mathbb{P}_1$ la probabilité uniforme sur Ω. Alors $\mathbb{P}_1(\{7\}) = \mathbb{P}_1(\{12\}) = 1/11$, ce qui ne correspond pas à ce que l'on peut observer. Ceci s'explique par le fait que ce modèle ne tienne pas compte du fait que l'on somme les faces des deux dés.

On peut considérer alors un espace d'état prenant en compte le résultat de chaque dé : $\Omega_2 = \{(\omega_1, \omega_2), \omega_1, \omega_2 \in \{1, \ldots, 6\}\}$. Autrement dit, pour chaque réalisation $\omega = (\omega_1, \omega_2)$, ω_1 est le résultat du lancer du premier dé et ω_2 celui du second. On a

$$\mathbb{P}_2\,(\text{la somme est } 7) = \frac{6}{36} = \frac{1}{6} \tag{1.5.1}$$

$$\mathbb{P}_2\,(\text{la somme est } 12) = \frac{1}{36} \tag{1.5.2}$$

On peut également déterminer $\mathbb{P}_2\,(\text{la somme est } x)$, $x \in \{2, 3, \ldots, 12\}$.

Dans le choix précédent de Ω_2, on a distingué les dés. Lorsqu'ils sont indicernables, on peut prendre

$$\Omega_3 = \{(\omega_1, \omega_2), 1 \leqslant \omega_1 \leqslant \omega_2 \leqslant 6\} \tag{1.5.3}$$

muni de la probabilité uniforme. On a

$$\operatorname{card}(\Omega_3) = 1 + 2 + 3 + 4 + 5 + 6 = 21 \tag{1.5.4}$$

et

$$\mathbb{P}_3\,(\text{la somme est } 7) = \frac{3}{21} = \frac{1}{7} \tag{1.5.5}$$

$$\mathbb{P}_3\,(\text{la somme est } 12) = \frac{1}{21} \tag{1.5.6}$$

1.6 Dénombrement

Lorsque l'on considère la probabilité uniforme, on est amené à déterminer le cardinal de certains ensembles, et donc à un problème de dénombrement.

Définition 1.14 (Permutations)**.** *Le nombre de permutations de $\{1, \ldots, n\}$ dans $\{1, \ldots, n\}$ (nombre de bijections d'un ensemble de n éléments dans un ensemble de n éléments) est*

$$n! = \prod_{k=1}^{n} k = n \times (n-1) \times \cdots \times 2 \times 1, \tag{1.6.1}$$

avec la convention $0! = 1$.

C'est le nombre de façons de placer n objets dans n cases, avec exactement un objet par case.

Définition 1.15 (Arrangements)**.** *Le nombre d'arrangements de k objets parmi n avec $k \leqslant n$ est noté A_n^k. Il correspond au nombre d'injections de $\{1, \ldots, k\}$ dans $\{1, \ldots, n\}$.*

C'est le nombre de façons de tirer k boules, sans remise et en tenant compte de l'ordre, dans une urne contenant n boules. On a

$$A_n^k = \frac{n!}{(n-k)!}. \tag{1.6.2}$$

Définition 1.16 (Coefficient binomial)**.** *Le nombre de sous-ensembles à k éléments d'un ensemble à n éléments, $0 \leqslant k \leqslant n$, est le coefficient binomial*

$$C_n^k = \frac{n!}{k!\,(n-k)!}. \tag{1.6.3}$$

C'est le nombre de façons de choisir k boules, sans remise et sans tenir compte de l'ordre, dans une urne contenant n boules.

Les coefficients binomiaux interviennent également dans la formule du binôme de Newton :

$$(a+b)^n = \sum_{k=0}^{n} C_n^k a^k b^{n-k}. \tag{1.6.4}$$

En particulier, avec $a = b = 1$, on obtient $\sum_{k=0}^{n} C_n^k = 2^n$.

1.7 Probabilités conditionnelles

On considère le lancer de deux dés équilibré distingables. On cherche à déterminer la probabilité de $A =$ "la somme des deux dés fait 10 ou plus" sachant B : "la face du deuxième dé est 5".

Si l'on effectue un grand nombre de lancers N, on note $N(A \cap B)$ le nombre de fois où $A \cap B$ se réalise et $N(B)$ le nombre de fois où B se réalise. En ne considérant que les réalisations où le deuxième dé indique 5, on a

$$\frac{N(A \cap B)}{N(B)} = \frac{N(A \cap B)}{N} \frac{N}{N(B)} \tag{1.7.1}$$

qui se comporte comme $\mathbb{P}(A \cap B) / \mathbb{P}(B)$.

Définition 1.17 (Probabilités conditionnelles)**.** *Soient A et B deux événements tels que $\mathbb{P}(B) > 0$. La probabilité conditionnelle de A sachant B, notée $\mathbb{P}(A \mid B)$, est définie par*

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{1.7.2}$$

Si A et B sont deux événements tels que $\mathbb{P}(B) > 0$ alors

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\,\mathbb{P}(B) \tag{1.7.3}$$

Proposition 1.18. *Si A et B sont deux événements tels que $0 < \mathbb{P}(B) < 1$ alors on a*

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\,\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\,\mathbb{P}(B^c) \tag{1.7.4}$$

et la formule de Bayes

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\,\mathbb{P}(B)}{\mathbb{P}(A \mid B)\,\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\,\mathbb{P}(B^c)}. \tag{1.7.5}$$

1.8 Indépendance

Définition 1.19 (Indépendance)**.** *On dit que deux événements A et B sont indépendants si*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B). \tag{1.8.1}$$

On dit qu'une collection au plus dénombrable d'événements $(A_i)_{i\in I}$ est indépendante si pour tout sous-ensemble fini J de I,

$$\mathbb{P}\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} \mathbb{P}(A_j). \tag{1.8.2}$$

Remarque 1.20. Si A et B sont indépendants et $\mathbb{P}(B) > 0$, alors $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.

Remarque 1.21. Si $(A_i)_{i\in I}$ est indépendante et $I' \subset I$, alors $(A_i)_{i\in I'}$ est indépendante.

Exemple : jeu de cartes

On prend un jeu de 32 cartes numérotées de 1 à 8, chaque numéro étant présent en quatre exemplaires, chacun d'une couleur : $\spadesuit$, $\clubsuit$, $\diamondsuit$, $\heartsuit$. On tire une carte au hasard dans ce jeu de carte. Les événements A : "la carte tirée est un 3" et B : "la carte tirée est un $\spadesuit$" sont indépendants.

Question 1.22. On considère le même jeu mais sans le 8$\clubsuit$ et 4$\diamondsuit$. Proposer un espace probabilisé pour modéliser cette expérience ainsi que la précédente et déterminer si les événements A et B définis précédemment sont indépendants.

Exemple : lancer de dés

On lance deux dés distingables. L'espace d'états pour le premier dé est $\Omega_1 = \{1, \ldots, 6\}$ avec la probabilité uniforme $\mathbb{P}_1$ et pour le second $\Omega_2 = \{1, \ldots, 6\}$ avec la probabilité uniforme $\mathbb{P}_2$. L'espace d'états associé au lancer des deux dés est

$$\Omega = \Omega_1 \times \Omega_2 = \{\omega = (\omega_1, \omega_2), \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}. \tag{1.8.3}$$

Une probabilité $\mathbb{P}$ sur Ω est déterminée par $\mathbb{P}(\{(\omega_1, \omega_2)\})$; $\omega_1 \in \Omega_1, \omega_2 \in \Omega_2$. Comme les résultats des deux lancers sont indépendants,

$$\mathbb{P}(\{(\omega_1, \omega_2)\}) = \mathbb{P}_1(\{\omega_1\})\,\mathbb{P}_2(\{\omega_2\}). \tag{1.8.4}$$

On dit que $\mathbb{P}$ est la probabilité produit, et on note $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$.

Plus généralement, les espaces produit permettent de modéliser des expériences indépendantes.

Exemple : lancer de dés Avec les mêmes notations que dans la page précédente, si $A = E_1 \times \Omega_2$ et $B = \Omega_1 \times E_2$, avec $E_1 \subset \Omega_1$, $E_2 \subset \Omega_2$, alors A et B sont indépendants (intuitivement, la réalisation de A ne dépend que du résultat du lancer du premier dé et celle de B seulement de celle du second).

Mais si $A = \{(1,1), (1,2)\}$ et $B = \{(1,1), (1,6)\}$, alors

$$\mathbb{P}(A) = \frac{2}{36} = \mathbb{P}(B) \tag{1.8.5}$$

et

$$\mathbb{P}(A \cap B) = \mathbb{P}(\{(1,1)\}) = \frac{1}{36} \neq \mathbb{P}(A)\,\mathbb{P}(B). \tag{1.8.6}$$

Par conséquent, A et B ne sont pas indépendants.

2 Variables aléatoires discrètes

2.1 Loi d'une variable aléatoire discrète

Définition 2.1 (Variable aléatoire)**.** *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace probabilisé où* Ω *est au plus dénombrable. Une variable aléatoire discrète* X *est une application* $X\colon \Omega \to E$ *où* E *est au plus dénombrable.*

La plupart du temps, E sera fini ou égal à $\mathbb{N}$.

Interprétation : à chaque résultat d'une expérience aléatoire, on attribue un nombre (de "piles" après un certain nombre de lancers d'une pièce, nombre de clients dans un magasin, gain ou surtout perte lors d'un jeu de hasard).

Définition 2.2 (Loi d'une variable aléatoire)**.** *Pour tout* $A \subset E$*, on note*

$$\mathbb{P}_X(A) = \mathbb{P}(\{\omega \in \Omega, X(\omega) \in A\}) \tag{2.1.1}$$

L'application $\mathbb{P}_X\colon \mathcal{P}(E) \to [0,1]$ *est une probabilité sur* E *et est appelée loi de la variable aléatoire* X*.*

Détermination de la loi de X

Soit X une variable aléatoire discrète, de loi $\mathbb{P}_X$. On suppose que X prend ses valeurs dans un ensemble E au plus dénombrable.

- Si $E = \{x_1, \ldots, x_N\}$ est fini, soit $p_k = \mathbb{P}(X = x_k) = \mathbb{P}(\{\omega \in \Omega; X(\omega) = x_k\})$.
- Si E est dénombrable, soit $\tau\colon E \to \mathbb{N}$ une bijection, $x_k = \tau^{-1}(k)$ et $p_k = \mathbb{P}(X = x_k)$ (si $E = \mathbb{N}$, on a simplement $p_k = \mathbb{P}(X = k)$).

La loi $\mathbb{P}_X$ est entièrement déterminée par les valeurs de p_k. En effet, par le Lemme 1.12,

$$\mathbb{P}_X(A) = \sum_{k : k \in A} p_k. \tag{2.1.2}$$

2.2 Moments d'une variable aléatoire discrète

2.2.1 Espérance d'une variable aléatoire discrète

On dit que X est une variable aléatoire réelle si $X\colon \Omega \to \mathbb{R}$, autrement dit, les valeurs prises par X sont réelles.

Définition 2.3 (Espérance)**.** *Soient* $E \subset \mathbb{R}$ *un ensemble au plus dénombrable et* $X\colon \Omega \to E$ *une variable aléatoire discrète. Si*

$$\sum_{x \in E} |x| \, \mathbb{P}(X = x) < \infty, \tag{2.2.1}$$

on dit que la variable aléatoire X *admet un moment d'ordre un et on pose*

$$\mathbb{E}[X] = \sum_{x \in E} x \mathbb{P}(X = x). \tag{2.2.2}$$

La quantité $\mathbb{E}[X]$ *est appelée espérance de* X*.*

L'espérance de X peut être interprétée comme la moyenne pondérée des valeurs prises par X.

2.2.2 Moments d'ordre supérieur

Définition 2.4. *On dit que la variable aléatoire réelle discrète X admet un moment d'ordre $n \in \mathbb{N}^*$ si X^n admet un moment d'ordre* 1.

Remarque 2.5. Si X est bornée, au sens où il existe une constante C telle que pour tout $\omega \in \Omega$, $|X(\omega)| \leqslant C$, alors X admet un moment d'ordre n pour chaque n.

Si X prend les valeurs $x_1, \ldots, x_N$, alors

$$\mathbb{E}[X] = \sum_{k=1}^{N} x_k \mathbb{P}(X = x_k) \tag{2.2.3}$$

et plus généralement

$$\mathbb{E}[X^n] = \sum_{k=1}^{N} x_k^n \mathbb{P}(X = x_k) \tag{2.2.4}$$

2.2.3 Propriétés de l'espérance

Proposition 2.6. *Soient X et Y deux variables aléatoires discrètes réelles ayant un moment d'ordre un fini. L'espérance vérifie les propriétés suivantes :*

- *linéarité : si $\alpha, \beta \in \mathbb{R}$, alors $\alpha X + \beta Y$ admet un moment d'ordre un et*

$$\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y]. \tag{2.2.5}$$

- *Positivité : si X est positive presque sûrement, c'est-à-dire $\mathbb{P}(X \geqslant 0) = 1$, alors $\mathbb{E}[X] \geqslant 0$.*
- *Croissance : si $X \leqslant Y$ presque sûrement, c'est-à-dire $\mathbb{P}(X \leqslant Y) = 1$, alors $\mathbb{E}[X] \leqslant \mathbb{E}[Y]$.*

Inégalités sur l'espérances

- Inégalité de Markov : pour tout $a > 0$,

$$\mathbb{P}(|X| > a) \leqslant \frac{1}{a} \mathbb{E}[|X|]. \tag{2.2.6}$$

- Inégalité de Tchebychev. Soit $a > 0$. On a

$$\mathbb{P}(|X| > a) \leqslant \frac{1}{a^2} \mathbb{E}\left[X^2\right]. \tag{2.2.7}$$

- Inégalité de Jensen : soit $\phi \colon \mathbb{R} \to \mathbb{R}$ une fonction convexe ($\phi(\alpha x + (1-\alpha) y) \leqslant \alpha \phi(x) + (1-\alpha)\phi(y)$ pour tous $x, y \in \mathbb{R}$ et tout $\alpha \in [0,1]$) et X une variable aléatoire telle que X et $\phi(X)$ admettent un moment d'ordre un. Alors

$$\phi(\mathbb{E}[X]) \leqslant \mathbb{E}[\phi(X)]. \tag{2.2.8}$$

- Inégalité de Cauchy-Schwarz : soient X et Y deux variables aléatoires réelle admettant un moment d'ordre deux fini. Alors

$$\mathbb{E}[|XY|] \leqslant \sqrt{\mathbb{E}[X^2]\,\mathbb{E}[Y^2]}. \tag{2.2.9}$$

2.3 Variance d'une variable aléatoire

Soit X une variable aléatoire admettant un moment d'ordre deux fini. Par l'inégalité de Jensen (2.2.8) ou bien par l'inégalité de Cauchy-Schwarz (2.2.9), X admet un moment fini d'ordre un.

Définition 2.7 (Variance)**.** *Soit X une variable aléatoire admettant un moment d'ordre deux fini. La variance de X, notée $\operatorname{Var}(X)$, est définie par*

$$\operatorname{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]. \tag{2.3.1}$$

Intuitivement : plus $\operatorname{Var}(X)$ est proche de 0 et plus les valeurs de X sont concentrées autour de la moyenne, et plus $\operatorname{Var}(X)$ est grande, plus X aura tendance à prendre des valeurs éloignées de la moyenne.

Définition 2.8 (Écart-type)**.** *L'écart-type de X, noté σ_X, est défini par*

$$\sigma_X = \sqrt{\operatorname{Var}(X)}. \tag{2.3.2}$$

Remarquons que $\sigma_{cX} = |c|\,\sigma_X$ pour tout $c \in \mathbb{R}$ car $\operatorname{Var}(cX) = c^2 \operatorname{Var}(X)$.

2.3.1 Espérance d'une fonction d'une variable aléatoire

Soit X une variable aléatoire discrète réelle, prenant ses valeurs dans un ensemble E au plus dénombrable. Soit $g\colon \mathbb{R} \to \mathbb{R}$ une fonction telle que

$$\sum_{x \in E} |g(x)|\, \mathbb{P}(X = x) < \infty. \tag{2.3.3}$$

La variable aléatoire $g \circ X =: g(X)$, définie par $g \circ X(\omega) = g(X(\omega))$, admet un moment d'ordre 1 et

$$\mathbb{E}[g(X)] = \sum_{x \in E} g(x)\, \mathbb{P}(X = x). \tag{2.3.4}$$

On se servira en particulier du cas $g\colon x \mapsto x^2$ pour calculer la variance.

2.4 Exemple de lois discrètes

2.4.1 Variable constante

Définition 2.9. *On dit qu'une variable aléatoire X est constante s'il existe un réel c tel que $\mathbb{P}(X = c) = 1$.*

On remarque que le réel c est unique (car pour $c \neq c'$, les événements $\{X = c\}$ et $\{X = c'\}$ sont incompatibles).

S'il existe un réel c tel que $\mathbb{P}(X = c) = 1$, alors $\mathbb{E}[X] = c$ et $\operatorname{Var}(X) = 0$.

2.4.2 Loi de Bernoulli

Définition 2.10. *On dit que X suit une loi de Bernoulli de paramètre $p \in [0,1]$ si*

$$\mathbb{P}(X=0) = 1-p, \quad \mathbb{P}(X=1) = p. \tag{2.4.1}$$

On a alors

$$\mathbb{E}[X] = p; \quad \mathrm{Var}(X) = p(1-p). \tag{2.4.2}$$

2.4.3 Loi Binomiale

Définition 2.11. *On dit que X suit une loi binomiale de paramètres (n,p) avec $n \in \mathbb{N}^*$ et $p \in [0,1]$ si*

$$\mathbb{P}(X=k) = C_n^k p^k (1-p)^{n-k}, \quad k \in \{0,1,\ldots,n\}. \tag{2.4.3}$$

Interprétation : X compte le nombre de succès dans une répétition de n expériences aléatoire indépendantes, chacune ayant une probabilité de succès p et d'échec $1-p$.

On a

$$\mathbb{E}[X] = np, \quad \mathrm{Var}(X) = np(1-p). \tag{2.4.4}$$

Ceci peut se déduire grâce aux identités

$$k \cdot C_n^k = nC_{n-1}^{k-1}, 1 \leqslant k \leqslant n, \tag{2.4.5}$$

$$k(k-1) \cdot C_n^k = n(n-1)C_{n-2}^{k-2}, 2 \leqslant k \leqslant n. \tag{2.4.6}$$

2.4.4 Loi Géométrique

Définition 2.12. *On dit que X suit une loi géométrique de paramètre $p \in]0,1[$ si pour chaque $n \in \mathbb{N}^*$,*

$$\mathbb{P}(X=n) = (1-p)^{n-1}p. \tag{2.4.7}$$

Interprétation : si on répète des expériences de Bernoulli de manière indépendante avec probabilité de succès p, X est le nombre d'expériences nécessaires à l'obtention d'un premier succès.

Grâce aux formules

$$\sum_{n \geqslant 1} r^n = \frac{r}{1-r} =: f(r), 0 \leqslant r < 1, \tag{2.4.8}$$

$$\sum_{n \geqslant 1} nr^{n-1} = f'(r), \sum_{n \geqslant 1} n(n-1)r^{n-2} = f''(r) \tag{2.4.9}$$

on déduit que

$$\mathbb{E}[X] = \frac{1}{p}, \quad \mathrm{Var}(X) = \frac{1-p}{p^2}. \tag{2.4.10}$$

2.4.5 Loi de Poisson

Définition 2.13. *On dit que X suit une loi de Poisson de paramètre $\lambda \in]0,\infty[$ si pour tout $k \in \mathbb{N}$,*

$$\mathbb{P}(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}. \tag{2.4.11}$$

Notons que pour de grandes valeurs de k, $\mathbb{P}(X=k)$ devient très petit (décroissance plus rapide que celle d'une suite géométrique).

Interprétation : X peut être interprétée comme le nombre de clients dans un magasin durant une certaine période de temps fixée.

Grâce à la formule

$$\sum_{k\geqslant 0}\frac{\lambda^k}{k!} = e^{\lambda}, \tag{2.4.12}$$

on peut déduire que

$$\mathbb{E}[X] = \mathrm{Var}(X) = \lambda. \tag{2.4.13}$$

2.5 Fonction génératrice d'une v.a. à valeurs entières positives

Définition de la fonction génératrice d'une v.a. Dans cette sous-section, on suppose que X est une variable aléatoire à valeurs dans $\mathbb{N}$.

Définition 2.14. *La fonction génératrice d'une variable aléatoire $X\colon \Omega \to \mathbb{N}$ est la fonction $G_X\colon [0,1] \to [0,1]$ définie par*

$$G_X(s) = \sum_{n=0}^{\infty} s^n\mathbb{P}(X=n), 0<s\leqslant 1, \quad G_X(0) = \mathbb{P}(X=0). \tag{2.5.1}$$

Remarque 2.15. La fonction G_X est bien définie car la série en question converge (grâce à $0 \leqslant s^n\mathbb{P}(X=n) \leqslant \mathbb{P}(X=n)$ et $\sum_{n=0}^{\infty}\mathbb{P}(X=n)=1$).

Proposition 2.16. *Soit X une variable aléatoire à valeurs entières et*

$$G_X(s) = \sum_{n=0}^{\infty} s^n\mathbb{P}(X=n), 0<s\leqslant 1, \quad G_X(0) = \mathbb{P}(X=0). \tag{2.5.2}$$

Alors G_X caractérise la loi de X ; plus précisément

$$\mathbb{P}(X=n) = \frac{G_X^{(n)}(0)}{n!}, \tag{2.5.3}$$

où $G_X^{(0)} = G_X$ et pour $n \geqslant 1$, $G_X^{(n)} = \left(G_X^{(n-1)}\right)'$.

Si on connaît G_X, alors on peut déterminer les moments de X en particulier $\mathbb{E}[X]$ et $\mathbb{E}[X^2]$. En effet,

$$G_X(s) = \sum_{n=0}^{\infty} s^n\mathbb{P}(X=n), 0<s\leqslant 1, \quad G_X(0) = \mathbb{P}(X=0), \tag{2.5.4}$$

$$G_X{}'(s) = \sum_{n=0}^{\infty} n s^{n-1} \mathbb{P}(X = n). \tag{2.5.5}$$

donc

$$G'_X(1) = \sum_{n=0}^{\infty} n \mathbb{P}(X = n) = \mathbb{E}[X]. \tag{2.5.6}$$

De même,

$$G_X{}''(s) = \sum_{n=0}^{\infty} n(n-1) s^{n-2} \mathbb{P}(X = n), \tag{2.5.7}$$

donc $G''_X(1) = \mathbb{E}[X(X-1)]$.

En résumé, on a les formules suivantes :

$$\mathbb{E}[X] = G'_X(1), \tag{2.5.8}$$

$$\mathbb{E}\left[X^2\right] = G''_X(1) + G'_X(1) \tag{2.5.9}$$

et

$$\mathrm{Var}(X) = G''_X(1) + G'_X(1) - \left(G'_X(1)\right)^2. \tag{2.5.10}$$

On peut aussi les retrouver grâce à

$$G_X(s) = \mathbb{E}\left[s^X\right] \tag{2.5.11}$$

et en dérivant sous l'espérance :

$$G'_X(s) = \mathbb{E}\left[X s^{X-1}\right], \quad G''_X(s) = \mathbb{E}\left[X(X-1) s^{X-2}\right]. \tag{2.5.12}$$

Exemples de fonction génératrice

- Loi de Bernoulli de paramètre p :

$$G_X(s) = \mathbb{P}(X = 0) + s^1 \mathbb{P}(X = 1) = 1 - p + ps. \tag{2.5.13}$$

- Loi binomiale de paramètres (n, p) : en utilisant (1.6.4),

$$G_X(s) = \sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} s^k = (1 - p + ps)^n. \tag{2.5.14}$$

- Loi géométrique de paramètre p : par (2.4.8),

$$G_X(s) = \sum_{k=0}^{\infty} (1-p)^{k-1} p s^k = \frac{ps}{1 - (1-p)s}. \tag{2.5.15}$$

- Loi de Poisson de paramètre λ : par (2.4.12),

$$G_X(s) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} s^k = e^{\lambda(s-1)}. \tag{2.5.16}$$

2.6 Loi conditionnelle et espérance conditionnelle

Définition 2.17 (Loi conditionnelle)**.** *Soient X et Y deux variables aléatoires discrètes définies sur un même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et prenant leurs valeurs respectives dans les ensembles E et F. On dit que X et Y sont indépendantes si*

$$\forall x \in E, y \in F, \quad \mathbb{P}\left(\{X = x\} \cap \{Y = y\}\right) = \mathbb{P}\left(X = x\right) \mathbb{P}\left(Y = y\right). \tag{2.6.1}$$

Définition 2.18. *Soient X et Y deux variables aléatoires discrètes définies sur un même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et prenant leurs valeurs respectives dans les ensembles E et F. La loi conditionnelle de Y sachant X est la donnée de*

$$\mathbb{P}\left(Y = y \mid X = x\right), x \in E, y \in F. \tag{2.6.2}$$

Remarque 2.19. Si X est indépendante de Y, alors la loi conditionnelle de Y sachant X est simplement la loi de Y.

Définition 2.20 (Indépendance d'une collection de variables aléatoires discrètes)**.** *Soient $X_1, \ldots, X_n$ une suite de variables aléatoires discrètes où X_i prend ses valeurs dans l'ensemble E_i. La suite $(X_i, 1 \leqslant i \leqslant n)$ est indépendante si pour tous $x_1 \in E_1, \ldots, x_n \in E_n$,*

$$\mathbb{P}\left((X_1 = x_1) \cap \cdots \cap (X_n = x_n)\right) = \mathbb{P}\left(X_1 = x_1\right) \ldots \mathbb{P}\left(X_n = x_n\right). \tag{2.6.3}$$

Proposition 2.21. *Soit $(X_i, 1 \leqslant i \leqslant n)$ une suite indépendante de variables aléatoires de Bernoulli de paramètre p. Alors $\sum_{i=1}^n X_i$ suit une loi binomiale de paramètres (n, p).*

Exemple 2.22. Soit $(X_i, 1 \leqslant i \leqslant n)$ une suite indépendante de variables aléatoires de Bernoulli de paramètre p. On cherche la loi conditionnelle de X_n sachant $S_n = \sum_{i=1}^n X_i$. On calcule pour $k \in \{1, \ldots, n\}$,

$$\mathbb{P}\left(X_n = 1 \mid S_n = k\right) = \frac{\mathbb{P}\left(X_n = 1, S_n = k\right)}{\mathbb{P}\left(S_n = k\right)} \tag{2.6.4}$$

$$= \frac{\mathbb{P}\left(X_n = 1, S_{n-1} = k - 1\right)}{\mathbb{P}\left(S_n = k\right)} \tag{2.6.5}$$

et comme X_n est indépendante de S_{n-1}, on obtient

$$\mathbb{P}\left(X_n = 1 \mid S_n = k\right) = \frac{\mathbb{P}\left(X_n = 1\right) \mathbb{P}\left(S_{n-1} = k - 1\right)}{\mathbb{P}\left(S_n = k\right)} \tag{2.6.6}$$

En utilisant le fait que S_n suit une loi binomiale de paramètres (n, p) et S_{n-1} suit une loi binomiale de paramètres $(n - 1, p)$, on trouve que $\mathbb{P}\left(X_n = 1 \mid S_n = k\right) = k/n$, formule valide aussi pour $k = 0$. Similairement, on trouve $\mathbb{P}\left(X_n = 0 \mid S_n = k\right) = 1 - k/n$.

2.6.1 Espérance

Soient X et Y deux variables aléatoires discrètes. L'espérance conditionnelle $\mathbb{E}[Y \mid X]$ est la meilleure prévision possible de Y par une fonction de X.

Définition 2.23. *Soient X et Y deux variables aléatoires discrètes, où X prend ses valeurs dans un ensemble $E \subset \mathbb{R}$ au plus dénombrable. On définit l'espérance conditionnelle de Y sachant X, notée $\mathbb{E}[Y \mid X]$, par la variable aléatoire $\phi(X)$, où*

$$\phi(x) = \begin{cases} \frac{\mathbb{E}[Y\mathbf{1}\{X=x\}]}{\mathbb{P}(X=x)} & \text{si } \mathbb{P}(X=x) \neq 0 \\ 0 & \text{sinon,} \end{cases} \tag{2.6.7}$$

où la variable aléatoire $\mathbf{1}\{X = x\}$ vaut 1 si $X(\omega) = x$ et 0 sinon.

On note $\phi(x) = \mathbb{E}[Y \mid X = x]$.

Exemple 2.24. Soit $(X_i, 1 \leqslant i \leqslant n)$ une suite indépendante de variables aléatoires de Bernoulli de paramètre p et $S_n = \sum_{i=1}^n X_i$. La variable aléatoire $X_n\mathbf{1}\{S_n = x\}$ prend la valeur 1 si $X_n = 1$ et $S_n = x$, et 0 sinon donc par le même calcul que pour la loi conditionnelle de X_n sachant S_n, on a pour tout $x \in \{0, \dots, n\}$,

$$\frac{\mathbb{E}[X_n\mathbf{1}\{S_n = x\}]}{\mathbb{P}(S_n = x)} = \frac{x}{n}. \tag{2.6.8}$$

Par conséquent, $\mathbb{E}[X_n \mid S_n] = S_n/n$.

2.6.2 Loi conditionnelle de X sachant un événement E

Définition 2.25. *La loi conditionnelle de X sachant un événément E de probabilité non nulle est la donnée des couples $(x_i, \mathbb{P}(X = x_i \mid E))_{i\in I}$, où $(x_i)_{i\in I}$ est l'ensemble des valeurs prises par X.*

2.7 Couples de variables aléatoires discrètes

2.7.1 Loi jointe de (X, Y)

Définition 2.26 (Concept de couples de variables aléatoires)**.** *Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X : \Omega \to I$, $Y : \Omega \to J$ des variables aléatoires, où $I, J \subset \mathbb{N}$. Le couple de variables aléatoires (X, Y) est la variable aléatoire $(X, Y) : \Omega \to I \times J$ définie par*

$$(X, Y)(\omega) = (X(\omega), Y(\omega)). \tag{2.7.1}$$

Différents cas de figure :

- X et Y sont indépendantes ;
- Y est une fonction de X ;
- Y est une fonction de (X, X'), où X et X' sont indépendantes.

Définition 2.27 (Loi jointe d'un couple de variables aléatoires)**.** *Soit (X, Y) un couple de variables aléatoires discrètes. La loi jointe de (X, Y) est la donnée de*

$$\{(x_i, y_j, p_{i,j})\}_{i\in I, j\in J}\,, \tag{2.7.2}$$

où

- $\{(x_i, y_j)\}_{i\in I, j\in J}$ *sont les valeurs prises par le couple (X, Y),*
- $\{(p_{i,j})\}_{i\in I, j\in J}$ *sont les probabilités correspondantes :*

$$p_{i,j} = \mathbb{P}\left(X = x_i, Y = y_j\right), i \in I, j \in J. \tag{2.7.3}$$

Exemple 2.28. On peut représenter la loi jointe dans un tableau.

$y_j \backslash x_i$	0	2	4
1	1/8	0	1/8
2	1/16	1/16	1/16
8	1/16	1/4	1/4

Ici, $I = \{0, 2, 4\}$, $J = \{1, 2, 8\}$,

$$p_{0,1} = \mathbb{P}\left(X = 0, Y = 1\right) = 1/8, p_{0,2} = \mathbb{P}\left(X = 0, Y = 2\right) = 1/16, p_{0,8} = 1/16$$

$$p_{2,1} = 0, p_{2,2} = 1/16, p_{2,8} = 1/4$$

$$p_{4,1} = 1/8, p_{4,2} = 1/16, p_{4,8} = 1/4$$

Remarque 2.29. On doit avoir $\sum_{i\in I}\sum_{j\in J} p_{i,j} = 1$ et $p_{i,j} \geqslant 0$.

2.7.2 Lois marginales

Définition 2.30. *Soit (X, Y) un couple de variables aléatoires discrètes. Les lois marginales de X et Y sont les lois des variables aléatoires réelles X et Y. Si (X, Y) a pour loi jointe $\{(x_i, y_j, p_{i,j})\}_{i\in I, j\in J}$, alors*

- *la loi de X est donnée par $\{(x_i, p_{i,\bullet})\}_{i\in I}$, où*

$$p_{i,\bullet} = \mathbb{P}\left(X = x_i\right) = \sum_{j\in J} p_{i,j}. \tag{2.7.4}$$

- *la loi de Y est donnée par $\{(y_j, p_{\bullet,j})\}_{j\in J}$, où*

$$p_{\bullet,j} = \mathbb{P}\left(Y = y_j\right) = \sum_{i\in I} p_{i,j}. \tag{2.7.5}$$

Exemple 2.31. Revenons à l'exemple précédent.

$y_j \backslash x_i$	0	2	4	$\mathbb{P}(Y = y_j)$
1	1/8	0	1/8	1/4
2	1/16	1/16	1/16	3/16
8	1/16	1/4	1/4	9/16
$\mathbb{P}(X = x_i)$	1/4	5/16	7/16	1

Exemple 2.32. On a vu que la donnée de la loi jointe d'un couple permet de déterminer les lois marginales. Les couples (X, Y) et (X', Y'), dont les lois jointes sont données par

$y_j \backslash x_i$	0	1	$\mathbb{P}(Y = y_j)$
0	1/4	1/4	1/2
1	1/4	1/4	1/2
$\mathbb{P}(X = x_i)$	1/2	1/2	1

et

$y_j \backslash x_i$	0	1	$\mathbb{P}(Y' = y_j)$
0	1/2	0	1/2
1	0	1/2	1/2
$\mathbb{P}(X' = x_i)$	1/2	1/2	1

ont les mêmes lois marginales (lois de Bernoulli de paramètre 1/2) mais pas la même loi jointe.

2.7.3 Fonction de deux variables

Espérance d'une fonction d'un couple Soit (X, Y) un couple de variables aléatoires discrètes prenant ses valeurs dans $\{(x_i, y_j), i \in I, j \in J\}$, $g \colon \mathbb{R}^2 \to \mathbb{R}$ une fonction, et $Z = g(X, Y)$. Alors Z est une variable aléatoire discrète, prenant ses valeurs dans $\{g(x_i, y_j), i \in I, j \in J\}$. Par conséquent,

$$\mathbb{E}[Z] = \mathbb{E}[g(X, Y)] = \sum_{i \in I} \sum_{j \in J} g(x_i, y_j) \mathbb{P}(X = x_i, Y = y_j). \tag{2.7.6}$$

En particulier, si X est indépendante de Y et $g(x,y) = xy$, on a

$$\mathbb{E}[XY] = \sum_{i\in I}\sum_{j\in J} g(x_i, y_j)\,\mathbb{P}(X = x_i, Y = y_j) \tag{2.7.7}$$

$$= \sum_{i\in I}\sum_{j\in J} x_i, y_j \mathbb{P}(X = x_i)\,\mathbb{P}(Y = y_j) \tag{2.7.8}$$

d'où

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]. \tag{2.7.9}$$

2.7.4 Covariance

Définition 2.33 (Covariance)**.** *Soient X et Y deux variables aléatoire discrètes définies sur un même espace probabilisé $(\Omega\mathcal{F}, \mathbb{P})$. On suppose que X et Y admettent un moment fini d'ordre deux. La covariance de X et Y, notée* $\mathrm{Cov}(X,Y)$*, est définie par*

$$\mathrm{Cov}(X,Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]. \tag{2.7.10}$$

On a également

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]. \tag{2.7.11}$$

Voici quelques propriétés de la covariance.

Proposition 2.34. *Si X et Y sont deux variables aléatoires indépendantes de carré intégrable, alors* $\mathrm{Cov}(X,Y) = 0$.

La réciproque est fausse : soit X prenant les valeurs -1, 0 et 1 avec probabilité $1/3$ et $Y = X^2$.

Proposition 2.35. *Soient X, Y et Z trois variables aléatoires discrètes de carré intégrable et $a, b \in \mathbb{R}$. On a :*

- $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$;
- $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$;
- $\mathrm{Cov}(aX + bY, Z) = a\,\mathrm{Cov}(X,Z) + b\,\mathrm{Cov}(Y,Z)$;
- $\mathrm{Cov}(X,a) = 0$.

Proposition 2.36 (Covariance et variance)**.** *Si X et Y sont deux variables aléatoires de carré intégrable, alors*

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y). \tag{2.7.12}$$

Si de plus X et Y sont indépendantes, alors

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y). \tag{2.7.13}$$

2.7.5 Coefficient de corrélation linéaire

Définition 2.37 (Coefficient de corrélation linéaire)**.** *Soient X et Y deux variables aléatoires de carré intégrable. Le coefficient de corrélation linéaire, noté $\rho(X,Y)$, est défini par*

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}. \tag{2.7.14}$$

Proposition 2.38. *Si X et Y sont indépendantes, alors $\rho(X,Y) = 0$.*

Proposition 2.39. *Pour toutes variables aléatoires X et Y non constantes,*

$$-1 \leqslant \rho(X,Y) \leqslant 1. \tag{2.7.15}$$

L'inégalité (2.7.15) est une conséquence de l'inégalité de Cauchy-Schwarz (2.2.9).

Proposition 2.40. *Soient X et Y deux variables aléatoires de carré intégrable. On a*

1. *$\rho(X,Y) = 1$ si et seulement si il existe $a > 0$ et $b \in \mathbb{R}$ tels que $X = aY + b$.*
2. *$\rho(X,Y) = -1$ si et seulement si il existe $a < 0$ et $b \in \mathbb{R}$ tels que $X = aY + b$.*

3 Convergence et théorèmes en probabilité

3.1 Convergence en probabilité

Définition 3.1. *On dit qu'une suite de variables aléatoires $(X_n)_{n\geqslant 1}$ converge en probabilité vers une variable aléatoire X si pour tout $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0. \tag{3.1.1}$$

Exemple 3.2. Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires telles que $\mathbb{E}[|X_n|] \to 0$. Alors $X_n \to X = 0$ en probabilité. En effet, l'inégalité de Markov (2.2.6) donne

$$\mathbb{P}(|X_n - X| > \varepsilon) \leqslant \frac{1}{\varepsilon}\mathbb{E}[|X_n|]. \tag{3.1.2}$$

3.2 Convergence en moyenne d'ordre p

Définition 3.3. *On dit qu'une suite de variables aléatoires $(X_n)_{n\geqslant 1}$ converge en moyenne d'ordre $p \in \mathbb{N}^*$ vers une variable aléatoire X si*

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0. \tag{3.2.1}$$

Proposition 3.4. *Si $(X_n)_{n\geqslant 1}$ converge en moyenne d'ordre p pour un certain $p \in \mathbb{N}^*$ vers une variable aléatoire X, alors $(X_n)_{n\geqslant 1}$ converge en probabilité vers X.*

Démonstration. Soit $\varepsilon > 0$ fixé. L'inégalité

$$\mathbb{P}\left(|X_n - X| > \varepsilon\right) = \mathbb{P}\left(|X_n - X|^p > \varepsilon^p\right) \tag{3.2.2}$$

a lieu. En appliquant l'inégalité de Markov (2.2.6), on obtient que

$$\mathbb{P}\left(|X_n - X| > \varepsilon\right) \leqslant \frac{\mathbb{E}\left[|X_n - X|^p\right]}{\varepsilon^p}, \tag{3.2.3}$$

et le membre de droite tend vers 0 par hypothèse. □

3.3 Convergence en loi pour des variables aléatoires entières

Définition 3.5. *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires à valeurs dans $\mathbb{N}$. On dit que $(X_n)_{n\geqslant 1}$ converge en loi vers X si pour tout $k \in \mathbb{N}$,*

$$\lim_{n\to\infty} \mathbb{P}\left(X_n = k\right) = \mathbb{P}\left(X = k\right). \tag{3.3.1}$$

Exemples 3.1.

1. Si X_n suit une loi de Poisson de paramètre $\lambda_n > 0$ et $\lambda_n \to \lambda$, alors $(X_n)_{n\geqslant 1}$ converge en loi vers une variable aléatoire X de loi de Poisson.
2. Si X_n suit une loi binomiale de paramètres (n, p_n) et $np_n \to \lambda > 0$, alors $(X_n)_{n\geqslant 1}$ converge en loi vers une loi de Poisson de paramètre λ.

3.4 Loi faible des grands nombres

Supposons que l'on dispose d'une pièce non nécessairement équilibrée. On aimerait pouvoir déterminer la probabilité d'obtenir un pile, notée p. On lance donc la pièce un grand nombre de fois, et on compte le nombre de fois où la pièce tombe sur pile. Plus formellement, si on note X_i la variable aléatoire prenant la valeur 1 si le résultat du $i$ème lancer est pile et 0 sinon, on calcule $\sum_{i=1}^n X_i/n$. On s'attend à une convergence de $\sum_{i=1}^n X_i/n$ vers p, ce qui est formalisé par le résultat suivant.

Théorème 3.6 (Loi faible des grands nombres)**.** *Soit $(X_n)_{n\geqslant 1}$ une suite indépendante de variables aléatoires de même loi. Soit $\overline{X_n} := \frac{\sum_{i=1}^n X_i}{n}$ la moyenne empirique. On suppose que X_1 admet un moment fini d'ordre deux. Alors $\left(\overline{X_n} - \mathbb{E}\left[X_1\right]\right)_{n\geqslant 1}$ converge en moyenne d'ordre deux vers 0. En particulier,*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\overline{X_n} - \mathbb{E}\left[X_1\right]\right| > \varepsilon\right) = 0. \tag{3.4.1}$$

Démonstration. Il suffit de montrer que

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\overline{X_n} - \mathbb{E}\left[X_1\right]\right)^2\right] = 0. \tag{3.4.2}$$

Comme $\mathbb{E}\left[\overline{X_n}\right] = \mathbb{E}\left[X_1\right]$, on a

$$\mathbb{E}\left[\left(\overline{X_n} - \mathbb{E}\left[X_1\right]\right)^2\right] = \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^n X_i\right). \tag{3.4.3}$$

Par (2.7.13) avec $X = \sum_{i=1}^{n-1} X_i$ et $Y = X_n$, on a

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \operatorname{Var}\left(\sum_{i=1}^{n-1} X_i\right) + \operatorname{Var}(X_n) = \operatorname{Var}\left(\sum_{i=1}^{n-1} X_i\right) + \operatorname{Var}(X_1). \tag{3.4.4}$$

On obtient donc par récurrence que

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = n \operatorname{Var}(X_1) \tag{3.4.5}$$

et par (3.4.6), on déduit que

$$\mathbb{E}\left[\left(\overline{X_n} - \mathbb{E}[X_1]\right)^2\right] = \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\operatorname{Var}(X_1). \tag{3.4.6}$$

□

3.5 Théorème central limite

La loi faible des grands nombres dit en particulier que $\left(\overline{X_n} - \mathbb{E}[X_1]\right)_{n\geqslant 1}$ converge en probabilité vers 0. On peut caractériser l'erreur commise entre la moyenne empirique et la "vraie" moyenne de la manière suivante.

Théorème 3.7. *(Théorème central limite (TCL)] Soit $(X_n)_{n\geqslant 1}$ une suite indépendante de variables aléatoires de même loi. Soit $\overline{X_n} := \frac{\sum_{i=1}^n X_i}{n}$ la moyenne empirique et $\sigma^2 = \operatorname{Var}(X_1) > 0$. Alors pour tout réel t,*

$$\mathbb{P}\left(\sqrt{n}\left(\overline{X_n} - \mathbb{E}[X_1]\right) \leqslant t\sigma\right) \to \frac{1}{\sqrt{2\pi}}\int_{\infty}^{t} \exp\left(-\frac{s^2}{2}\right) \mathrm{d}s. \tag{3.5.1}$$

On peut utiliser le TCL pour fournir un intervalle aléatoire contenant la valeur moyenne avec la probabilité désirée. Soit l'intervalle aléatoire

$$I_{n,a} = \left[\overline{X_n} - \frac{a\sigma}{\sqrt{n}}, \overline{X_n} + \frac{a\sigma}{\sqrt{n}}\right]. \tag{3.5.2}$$

Alors

$$\lim_{n\to\infty} \mathbb{P}\left(\mathbb{E}[X_1] \in I_{n,a}\right) = \frac{1}{\sqrt{2\pi}}\int_{-a}^{a} \exp\left(-\frac{s^2}{2}\right) \mathrm{d}s \tag{3.5.3}$$

et on peut choisir a pour que le membre de droite aie la valeur comprise entre 0 et 1 voulue (souvent, $0,95$).

4 Statistique

4.1 Introduction à la statistique

4.1.1 Objectifs

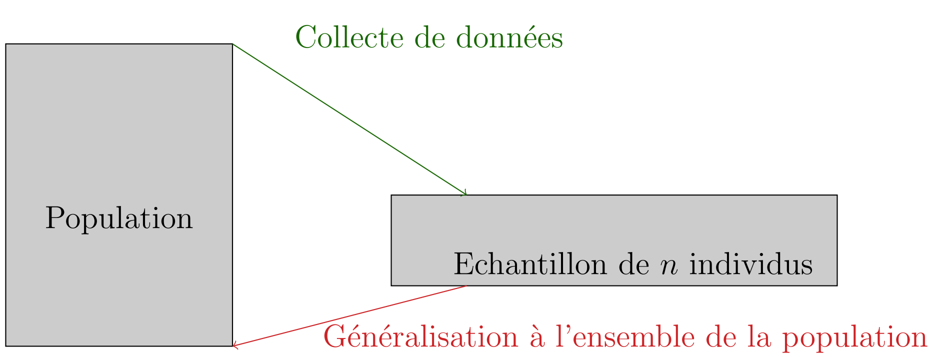
Définition 4.1. *La* ***statistique*** *est la science dont l'objet est de recueillir, de traiter et d'analyser des données issues de l'observation de phénomènes aléatoires, c'est-à-dire dans lesquels le hasard intervient.*

4.1.2 Différence entre statistique descriptive et statistique inférentielle

- La **statistique descriptive**
  - a pour but de résumer l'information contenue dans les données de façon synthétique et efficace ;
  - utilise la représentations de données sous forme de graphiques, de tableaux et d'indicateurs numériques (par exemple des moyennes) ;
  - permet de dégager les caractéristiques essentielles du phénomène étudié et de suggérer des hypothèses pour une étude ultérieure plus sophistiquée. Les probabilités n'ont ici qu'un rôle mineur.
- La **statistique inférentielle**
  - va au delà de la simple description des données.
  - Elle a pour but de faire des prévisions et de prendre des décisions au vu des observations.

  En général, il faut pour cela proposer des modèles probabilistes du phénomène aléatoire étudié et savoir gérer les risques d'erreurs.

4.1.3 Démarche statistique



4.2 Statistiques descriptives

4.2.1 Terminologie

- Les données dont nous disposons sont des mesures faites sur des **individus** (ou unités statistiques) issus d'une population.
- On s'intéresse à une ou plusieurs particularités des individus appelées **variables** ou caractères.
- L'ensemble des individus constitue l'échantillonétudié.

Exemple 4.2. Si l'échantillon est un groupe de TD de l'UFR math/info,

- un individu est un étudiant
- la population peut être l'ensemble des étudiants de l'UFR
- les variables étudiées peuvent être la taille, la filière choisie, la moyenne d'année, la couleur des yeux, ...
- Si l'échantillon est constitué de tous les individus de la population, on dit que l'on fait un **recensement**. Situation rare, essentiellement pour des raisons de coût.
- Quand l'échantillon n'est qu'une partie de la population, on parle de **sondage**. Le principe des sondages est d'étendre à l'ensemble de la population les enseignements tirés de l'étude de l'échantillon. Pour que cela ait un sens, il faut que l'échantillon soit représentatif de la population.

Dans ce chapitre, on ne s'intéresse qu'au cas où on ne mesure qu'une seule variable sur les individus. On dit alors que l'on fait de la statistique unidimensionnelle. Dans ce cas, les données sont sous la forme de la série des valeurs prises par la variable pour les n individus, notées $x_1, \ldots, x_n$.

On supposera que ces données sont les réalisations de n variables aléatoires $X_1, \ldots, X_n$ indépendantes et de même loi. On notera X une variable aléatoire de cette loi.

Le terme d'échantillon désignera $x_1, \ldots, x_n$.

4.2.2 Représentations graphiques, variables discrètes

Une variable discrète est une variable à valeurs dans un ensemble fini ou dénombrable.

- Si elles s'expriment par des nombres réels, elles sont appelées variables **quantitatives** ou numériques (ex : longueur, durée, coût, ...) ;
- si elles s'expriment par l'appartenance à une catégorie, elles sont appelées variables **qualitatives** ou catégorielles (ex : couleur, catégorie socio-professionnelle, ...).

Si la variable est qualitative, on appelle **modalités** les valeurs possibles de cette variable. L'ensemble des modalités est noté $E = \{e_1, \ldots, e_k\}$.

Par exemple, si la variable est la couleur des yeux d'un individu, l'ensemble des modalités est $E = \{\text{bleu}, \text{vert}, \text{brun}, \text{pers}, \text{noir}\}$. Si on interroge $n = 200$ personnes, les données sont sous la forme "noir, bleu, pers, vert, bleu..." ce qui n'est pas lisible.

Définition 4.3 (Fréquence absolue et relative)**.** *La **fréquence absolue** de la modalité e_j est le nombre total n_j d'individus de l'échantillon pour lesquels la variable a pris la modalité e_j , i.e. $n_j = \sum_{i=1}^n \mathbf{1}_{x_i=e_j}$.*

La fréquence relative de la modalité e_j est le pourcentage n_j/n d'individus de l'échantillon pour lesquels la variable a pris la modalité e_j.

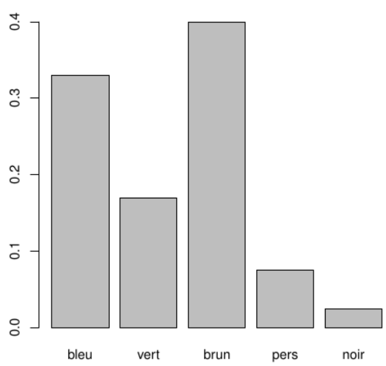
couleur des yeux	bleu	vert	brun	pers	noir
fréquences absolues	66	34	80	15	5
fréquences relatives	33%	17%	40%	7.5%	2.5%

Couleurs des yeux d'un échantillon de 200 personnes

Ici, $n = 200$, $k = 5$, $e_1 =$"bleu", $e_2 =$"vert", $e_3 =$"brun", $e_4 =$"pers" et $e_5 =$"noir". On a $n_1 = 66$, $n_2 = 34$, $n_3 = 80$, $n_4 = 15$ et $n_5 = 5$.

On peut utiliser deux types de représentations graphiques :

- **diagrammes en colonnes ou en bâtons** : à chaque modalité correspond un rectangle vertical dont la hauteur est proportionnelle à la fréquence relative de cette modalité ;
- **diagrammes sectoriels ou camemberts** : à chaque modalité correspond un secteur de disque dont l'aire (ou l'angle au centre) est proportionnelle à la fréquence relative de cette modalité.



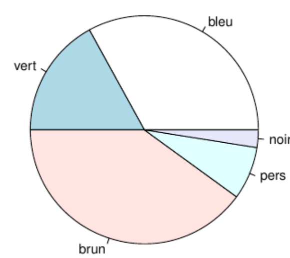
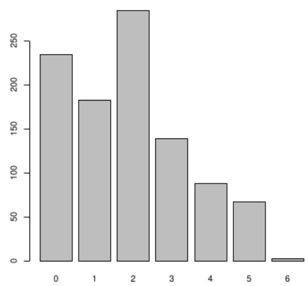


FIGURE 1 – Diagramme en bâtons (gauche), sectoriel (droite)

Pour les variables quantitatives, il y a un ordre naturel sur les modalités donc on n'utilise pas de diagramme sectoriel.

Nombre d'enfants	0	1	2	3	4	5	6	> 6
fréquences absolues	235	183	285	139	88	67	3	0
fréquences relatives	23.5%	18.3%	28.5%	13.9%	8.8%	6.7	0.3	0

TABLE 4.2 – Nombre d'enfants de 1000 couples



Lorsque la variable est continue (i.e., l'ensemble des modalités n'est pas dénombrable), on ne peut pas faire de représentation en diagramme en bâtons.

On peut cependant utiliser :

- l'histogramme et le polygone des fréquences qui lui est associé ;
- la fonction de répartition empirique, qui permet notamment de construire des graphes de probabilités.

On regroupe les observations proches en classes.

- On se fixe une borne inférieure $a_0 < x_1^* = \min_{1\leqslant i\leqslant n} x_i$ et une borne supérieure $a_k > x_n^* = \max_{1\leqslant i\leqslant n} x_i$.
- On partitionne l'intervalle $]a_0, a_k]$ en k intervalles $]a_{j-1}, a_j]$ appelés classes. La largeur de la classe j est $h_j = a_j - a_{j-1}$.
- Si $h_j = h$ pour tout j, on dit que l'on fait un histogramme à pas fixe.
- Si les h_j ne sont pas tous égaux, l'histogramme est dit à pas variable.

Définition 4.4 (Effectif)**.** *L'**effectif** de la classe j est le nombre d'observations appartenant à cette classe :*

$$n_j = \sum_{i=1}^{n} \mathbf{1}_{]a_{j-1},a_j]}(x_i)\,. \tag{4.2.1}$$

La fréquence de la classe j est n_j/n.

Définition 4.5 (Histogramme)**.** *L'**histogramme** est la figure constituée des rectangles dont les bases sont des classes et dont les aires sont égales aux fréquences de ces classes. Autrement dit, la hauteur du j^{e} rectangle est $n_j/(nh_j)$.*

En pratique, on fera le choix de $a_0 = x_1^* - 0.025\,(x_n^* - x_1^*)$ et $a_k = x_n^* + 0.025\,(x_n^* - x_1^*)$.

On peut utiliser le polygone des fréquences, i.e. la ligne brisée reliant les milieux des sommets des rectangles, et prolongée de part et d'autre des bornes de l'histogramme de sorte que l'aire sous le polygone soit égale à 1 (comme une densité).

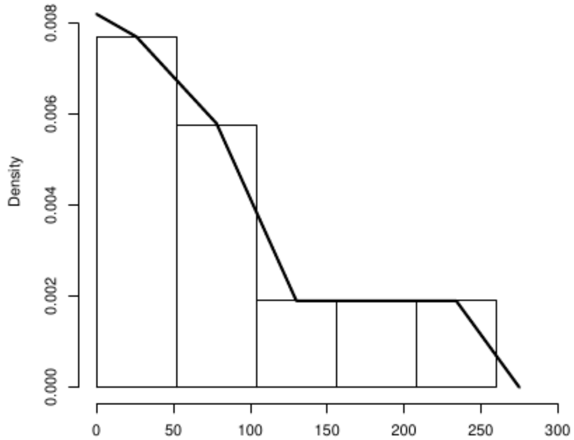


FIGURE 2 – Histogramme à classes de même largeur et polygone des fréquences

Si au lieu des effectifs n_j, on considère les effectifs cumulés $m_j := \sum_{i=1}^{j} n_i$, on construit un histogramme cumulé et un polygone des fréquences cumulées.

Si l'échantillon $x_1, \ldots, x_n$ provient de n réalisations indépendantes d'une variable aléatoire X, le polygone des fréquences cumulées donne une approximation de la fonction de répartition de X, c'est-à-dire la fonction $F \colon \mathbb{R} \to \mathbb{R}$ définie par

$$F(t) = \mathbb{P}(X \leqslant t). \tag{4.2.2}$$

Définition 4.6 (Échantillon ordonné)**.** *Soit $x_1, \ldots, x_n$ un échantillon. L'échantillon $x_1^*, \ldots, x_n^*$ est l'échantillon ordonné de $x_1, \ldots, x_n$ si*

1. $\{x_1, \ldots, x_n\} = \{x_1^*, \ldots, x_n^*\}$ *et*
2. $x_1^* \leqslant x_2^* \leqslant \ldots \leqslant x_n^*$.

Définition 4.7 (Fonction de répartition empirique)**.** *Soit $x_1, \ldots, x_n$ un échantillon dont toutes les valeurs sont distinctes et $x_1^*, \ldots, x_n^*$ l'échantillon ordonné. La fonction de répartition empirique F_n associée à l'échantillon $x_1, \ldots, x_n$ est la fonction en escalier suivante :*

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{x_i \leqslant t} = \begin{cases} 0 & si\ t < x_1^*, \\ i/n & si\ x_i^* \leqslant t < x_{i+1}^*, 1 \leqslant i \leqslant n-1, \\ 1 & si\ t \geqslant x_n^*. \end{cases} \tag{4.2.3}$$

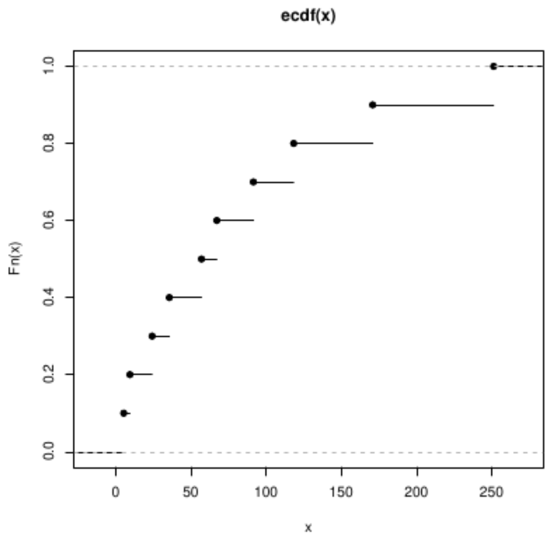


Figure 3 – Exemple de fonction de répartition empirique

4.2.3 Indicateurs statistiques

Définition 4.8 (Moyenne)**.** *La moyenne empirique d'un échantillon $x_1, \ldots, x_n$ est donné par*

$$\overline{x_n} = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{4.2.4}$$

Définition 4.9 (Valeurs extrêmes)**.** *Le minimum et le maximum d'un échantillon $x_1, \ldots, x_n$ sont donnés respectivement par*

$$x_1^* = \min_{1 \leqslant i \leqslant n} x_i, \quad x_n^* = \max_{1 \leqslant i \leqslant n} x_i. \tag{4.2.5}$$

Problème : très sensible en présence de valeurs aberrantes.

Définition 4.10 (Médiane)**.** *Soit $x_1, \ldots, x_n$ un échantillon et $x_1^* \leqslant x_2^* \leqslant \ldots \leqslant x_n^*$ l'échantillon ordonné.*

- *si n est impair, la médiane est définie par $x^*_{\frac{n+1}{2}}$.*
- *si n est pair, la médiane est définie par $\frac{x^*_{\frac{n}{2}} + x^*_{\frac{n}{2}+1}}{2}$*

La médiane n'est pas sensible aux valeurs aberrantes.

Définition 4.11 (Variance)**.** *La variance empirique d'un échantillon $x_1, \ldots, x_n$ est donnée par*

$$s_n^2 := \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x_n})^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{x_n}^2. \tag{4.2.6}$$

Définition 4.12 (Écart type)**.** *L'écart type est la racine de la variance empirique $s_n = \sqrt{s_n^2}$.*

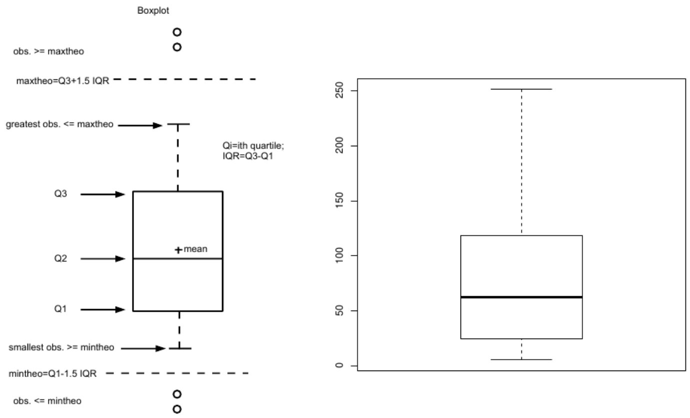


FIGURE 4 – Boxplot : explication (à gauche) et exemple (à droite)

Définition 4.13 (Étendue)**.** *Soit* $x_1, \ldots, x_n$ *un échantillon et* $x_1^* \leqslant x_2^* \leqslant \ldots \leqslant x_n^*$ *l'échantillon ordonné. Son étendue est donnée par* $e_n = x_n^* - x_1^*$.

Définition 4.14 (Quantile empirique)**.** *Soit* $x_1, \ldots, x_n$ *un échantillon et* $x_1^* \leqslant x_2^* \leqslant \ldots \leqslant x_n^*$ *l'échantillon ordonné. Le quantile empirique pour une probabilité* p *est donné par*

$$\forall p \in]0,1[, \quad q_{n,p} = \begin{cases} \frac{x_{np}^* + x_{np+1}^*}{2} & \text{si } np \in \mathbb{N} \\ x_{\lfloor np \rfloor + 1}^* & \text{sinon,} \end{cases} \tag{4.2.7}$$

où $\lfloor np \rfloor$ *est l'unique entier vérifiant* $\lfloor np \rfloor \leqslant np < \lfloor np \rfloor + 1$.

Les quartiles correspondent à $q_{n,1/4}, q_{n,2/4}$ (médiane) et $q_{n,3/4}$. La distance interquartile $q_{n,3/4} - q_{n,1/4}$ est un indicateur de volatilité.

Définition 4.15 (Boxplot)**.** *Le diagramme de boxplot affiche les premier et le troisième quartiles avec les bords du rectangle, la médiane avec le trait en gras.*