Probabilités et Statistiques S4
Université de Strasbourg
L2 Cursus master ingénierie, Informatique, Mathématiques, Mathématiques et Physique Approfondies

Davide Giraudo

23 février 2024

Table des matières

- 10 heures de cours magistral, vendredi de 8h à 10h sur 5 séances.
- 17 heures de travaux dirigés.
- 2 contrôles continus de 1 heure 30 chacun.
- Note :
  — assiduité $0,1$ (présence à tous les TD),
  — $0,45$ pour chaque contrôle continu.

Introduction

Ce cours constitue une première introduction à la probabilité et aux statistiques descriptives. On verra comment modéliser une expérience aléatoire ayant un nombre fini de résultats.

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 $\{1, \ldots, n\}$. 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

Puisque l'on est amené à s'intéresser à des événements s'exprimant comme des opérations sur des événements, il est important d'être familier avec celles-ci.

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

La collection d'événements qui nous intéresse doit être stable par rapport à certaines opérations ensemblistes.

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* $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$*.*

Bien que l'on ait pas encore introduit les probabilités, voici une explication intuive de cette définition. On essaie d'associer à un événement A une quantité $\mathbb{P}(A) \in [0,1]$ qui est d'autant plus grande que A a de chances de se produire. On doit être capable de déterminer la probabilité que rien ne se passe (l'ensemble vide), que quelque chose se passe (Ω tout entier). De plus, si on sait déterminer $\mathbb{P}(A)$, la probabilité que A se produise, alors on doit être également capable de déterminer la probabilité que A ne se produise pas. Enfin, la probabilité que certains événements se produisent doit aussi avoir un sens.

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 $A_0 \neq \Omega$.

Dans le cas où il y a un nombre fini de résultats possibles, on prendra $\mathcal{F} = \mathcal{P}(\Omega)$.

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 $i \geqslant 1$.

□

1.3.1 Exercices

Exercice 1. Donner une condition nécessaire et suffisante sur A et B pour que $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Exercice 2. Soient A et B deux événements tels que $\mathbb{P}(A) = \mathbb{P}(B) = 1$. Que vaut $\mathbb{P}(A \cap B)$?

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 $A \subset \Omega$,

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

1.4.2 Exercices

Exercice 3. Soit $\Omega = \{1, \ldots, n\}$ avec $\mathbb{P}(\{k\}) = 1/n$ pour tout $k \in \{1, \ldots, n\}$. Soient $A = \{1, \ldots, i\}$ et $B = \{1, \ldots, j\}$. Déterminer $\mathbb{P}(A)$, $\mathbb{P}(B)$, $\mathbb{P}(A \cap B)$ et $\mathbb{P}(A \cup B)$.

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

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

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 $\sum_{k=0}^{n} C_n^k = 2^n$.

1.6.1 Exercices

Exercice 4. On doit asseoir sur un banc de 10 places, 4 Allemands, 3 Français, et 3 Anglais ; les gens de même nationalité devant rester groupés. Combien de dispositions sont possibles ?

Exercice 5. On veut former un comité de 7 personnes, constitué de 2 démocrates, 2 républicains, et 3 indépendants. On a le choix parmi 6 démocrates, 5 républicains, et 4 indépendants. Combien de choix sont possibles ?

Exercice 6. Huit nouveaux professeurs vont être envoyés dans 4 écoles.

1. Combien d'affectations sont possibles ?
2. Qu'en est-il si l'on impose que chaque école recevra 2 professeurs ?

Ici, on entend par effectation l'attribution à un professeur d'une école, en particulier, il est possible qu'une ou plusieurs écoles ne soient pas affectée(s).

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

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

1.7.1 Exercices

Exercice 7. Soient A et B deux événements tels que $\mathbb{P}(B) > 0$. Calculer $\mathbb{P}(A \mid B) + \mathbb{P}(A^c \mid B)$.

Exercice 8. Soient A et B deux événements tels que $B \subset A$. Exprimer $\mathbb{P}(A \mid B)$ et $\mathbb{P}(B \mid A)$ en fonction de $\mathbb{P}(A)$ et $\mathbb{P}(B)$?

Exercice 9. On suppose que l'on a autant de chance d'avoir une fille ou un garçon à la naissance. Votre voisin de palier vous dit qu'il a deux enfants.

1. Quelle est la probabilité qu'il ait au moins un garçon ?
2. Quelle est la probabilité qu'il ait un garçon, sachant que l'ainée est une fille ?
3. Quelle est la probabilité qu'il ait un garçon, sachant qu'il a au moins une fille ?
4. Vous téléphonez à votre voisin. Une fille décroche le téléphone. Vous savez que dans les familles à un garçon et une fille, la fille décroche le téléphone avec probabilité p, quelle est la probabilité que votre voisin ait un garçon ?
5. Vous sonnez à la porte de votre voisin. Une fille ouvre la porte. Sachant que l'ainé(e) ouvre la porte avec probabilité p, et ce indépendamment de la répartition de la famille, quelle est la probabilité que votre voisin ait un garçon ?

On pourra considérer les événements :

$$\begin{array}{rc} G: & \text{"il y a au moins un garçon"}, \\ A: & \text{"l'aîné est une fille"}, \\ F: & \text{"il y a au moins une fille"}, \\ D: & \text{"une fille décroche le téléphone"}, \\ P: & \text{"une fille ouvre la porte"}. \end{array}$$

Exercice 10. Les laboratoires pharmaceutiques indiquent pour chaque test sa sensibilité α, qui est la probabilité que le test soit positif si le sujet est malade, et sa spécificité β, qui est la probabilité que le test soit négatif si le sujet est sain.

1. Sachant que la probabilité d'être malade est de 1 pour 1000 personnes, calculer la probabilité pour que vous soyez un sujet sain alors que votre test est positif, avec $\alpha = 98\%$ et $\beta = 97\%$.
2. Calculer la probabilité d'être malade alors que le test est négatif.

Exercice 11. Le jeu télévisé "la porte de la fortune" comporte trois portes : A, B et C. Derrière l'une d'entre elles se trouve un cadeau et rien derrière les deux autres. Vous choisissez au hasard une des trois portes sans l'ouvrir, par exemple la porte A. À ce moment-là, le présentateur, qui sait derrière quelle porte se trouve le cadeau, ouvre une porte parmi les deux B et C, derrière laquelle il n'y a évidement rien. On vous propose alors de changer ou non de porte, le but étant d'ouvrir la porte qui cache le cadeau afin de gagner. L'objectif de cet exercice est de déterminer votre meilleure stratégie.

1. On suppose que si le cadeau est derrière la porte A, alors le présentateur choisit au hasard entre les deux autres portes. Calculer la probabilité pour que le cadeau soit derrière la porte B sachant que le présentateur ouvre la porte C. Que faites-vous ?

2. On suppose que si le cadeau est derrière la porte A, alors le présentateur choisit systématiquement la porte B. Que faites-vous si le présentateur ouvre la porte B ?, la porte C ?

3. Montrer que quelle que soit la valeur de la probabilité pour que le présentateur ouvre la porte B (respectivement C) sachant que le cadeau est en A, vous avez intérêt à changer de porte. En déduire que la meilleure stratégie consiste à changer systématiquement de porte.

On pourra considérer les événements :

$$\begin{aligned} G_A &: \text{"le cadeau se trouve derrière la porte A"}, \\ G_B &: \text{"le cadeau se trouve derrière la porte B"}, \\ G_C &: \text{"le cadeau se trouve derrière la porte C"}, \\ O_A &: \quad \text{"le présentateur ouvre la porte A"}, \\ O_B &: \quad \text{"le présentateur ouvre la porte B"}, \\ O_C &: \quad \text{"le présentateur ouvre la porte C"}. \end{aligned}$$

1.8 Indépendance

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

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

Remarque 1.20. Si A et B sont indépendants et $\mathbb{P}(B) > 0$, alors $\mathbb{P}(A \mid B) = \mathbb{P}(A)$. Intuitivement, cela veut dire que la probabilité que A se réalise sachant que B se réalise est la même que celle que la probabilité de A. Notons que si $(B) = 0$, alors A est automatiquement indépendant de B.

On aimerait à présent définir le concept d'indépendance de plusieurs événements. Commençons par le cas de trois événements A, B et C. Intuitivement, C devrait être indépendant de tout événement que l'on peut fabriquer à partir de A et de B. Donc C devrait être indépendant de $A \cap B$, $A \cap B^c$, $A^c \cap B$ et de $A^c \cap B^c$, ce qui impose

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap B)\,\mathbb{P}(C) \tag{1.8.2}$$

et l'indépendance entre C et A, ainsi qu'entre C et B. En appliquant le même raisonnement en échangeant les rôles de B et C, on trouve que A doit être indépendant de B et que B doit être indépendant de C. Toutes les contraintes se traduisent par les égalités suivantes :

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C), \tag{1.8.3}$$

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B), \tag{1.8.4}$$
$$\mathbb{P}(A \cap C) = \mathbb{P}(A)\,\mathbb{P}(C), \tag{1.8.5}$$
$$\mathbb{P}(B \cap C) = \mathbb{P}(B)\,\mathbb{P}(C). \tag{1.8.6}$$

Ceci nous amène à la définition suivante.

Définition 1.21. *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,*

$$\mathbb{P}\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} \mathbb{P}(A_j). \tag{1.8.7}$$

Remarque 1.22. Si $(A_i)_{i\in I}$ est indépendante et $I' \subset I$, alors $(A_i)_{i\in I'}$ est indépendante.

Exemple 1.23 (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 1.24*.* 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 1.25 (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

$$\Omega = \Omega_1 \times \Omega_2 = \{\omega = (\omega_1, \omega_2), \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}. \tag{1.8.8}$$

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,

$$\mathbb{P}(\{(\omega_1, \omega_2)\}) = \mathbb{P}_1(\{\omega_1\})\,\mathbb{P}_2(\{\omega_2\}). \tag{1.8.9}$$

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 l'exemple précédent, 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

$$\mathbb{P}(A) = \frac{2}{36} = \mathbb{P}(B) \tag{1.8.10}$$

et

$$\mathbb{P}(A \cap B) = \mathbb{P}(\{(1,1)\}) = \frac{1}{36} \neq \mathbb{P}(A)\,\mathbb{P}(B). \tag{1.8.11}$$

Par conséquent, A et B ne sont pas indépendants.

Exercice 12. Soient A_1, A_2, A_3 et A_4 quatre événements. Écrire les conditions pour que ces évènements soient mutuellement indépendants.

Exercice 13. Soient A, B et C trois événements mutuellement indépendants. Démontrer que A, B et C^c le sont aussi.

Exercice 14. Soient A, B et C trois événements mutuellement indépendants. Démontrer que C est indépendant de $A \cup B$.

Exercice 15. On lance deux pièces de monnaie et on considère les événements :
A "la première pièce donne face"
B "la deuxième pièce donne pile"
C "les deux pièces donnent le même résultat"

1. Les événements A, B, C sont-ils indépendants deux à deux ?
2. Les événements A, B, C sont-ils globalement indépendants ?

Exercice 16. Un atelier comporte 3 machines indépendantes entre elles A, B, C. Les probabilités de défaillance sont respectivement $\mathbb{P}\left(\text{défaillance de } A\right) = 0.1$; $\mathbb{P}\left(\text{défaillance de } B\right) = 0.2$; $\mathbb{P}\left(\text{défaillance de } C\right) = 0.3$. Quelle est la probabilité d'avoir exactement une machine en panne ?

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*

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

L'application $\mathbb{P}_X \colon \mathcal{P}\left(E\right) \to [0, 1]$ *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 $p_k = \mathbb{P}\left(X = x_k\right) = \mathbb{P}\left(\{\omega \in \Omega; X\left(\omega\right) = x_k\}\right)$.
- Si E est dénombrable, soit $\tau \colon E \to \mathbb{N}$ une bijection, $x_k = \tau^{-1}\left(k\right)$ et $p_k = \mathbb{P}\left(X = x_k\right)$ (si $E = \mathbb{N}$, on a simplement $p_k = \mathbb{P}\left(X = k\right)$).

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.

Exercice 17. Soit X une variable aléatoire correspondant au résulat du lancer d'un dé équilibré à six faces. Déterminer $\mathbb{E}[X]$.

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*

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

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

Proposition 2.7 (Inégalités sur l'espérances)**.**

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

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

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

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

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

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

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

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

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

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

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

Définition 2.9 (Écart-type)**.** *L'écart-type de X, noté σ_X, est défini par*

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

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

Exercice 18. Soit N un entier naturel fixé et X une variable aléatoire prenant les valeurs $-N$, 0 et N avec probabilité 1/3. Calculer l'espérance, la variance et l'écart-type de X.

Exercice 19. Neuf chevaux, quatre blancs et cinq noirs pénètrent un par un sur la piste d'un cirque. Ils sont distincts les uns par rapport aux autres par leur harnachement.

1. Combien existe-t-il de façons de les faire arriver sur la piste ?
2. On appelle X la variable aléatoire : nombre de chevaux blancs précédent le premier cheval noir. Donner la loi de probabilité de X.
3. Calculer l'espérance mathématique de X et sa variance.

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.

Exercice 20. Soit X une variable aléatoire de loi de Poisson de paramètre $\theta > 0$ (i.e. $\mathbb{P}(X = k) = e^{-\theta}\frac{\theta^k}{k!}, k \in \mathbb{N}$).

1. Vérifier que $\frac{1}{1+X}$ est une variable intégrable. Calculer $\mathbb{E}\left[\frac{1}{1+X}\right]$.
2. Calculer $\mathbb{E}\left[\frac{1}{(1+X)(2+X)}\right]$ et en déduire $\mathbb{E}\left[\frac{1}{2+X}\right]$.

2.4 Exemple de lois discrètes

2.4.1 Variable constante

Définition 2.10. *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 $\mathrm{Var}\,(X) = 0$.

2.4.2 Loi de Bernoulli

Définition 2.11. *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 2.12. *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 uniforme discrète

Définition 2.13. *On dit que X suit une loi uniform discrète de paramètre $N \in \mathbb{N}^*$ si*

$$\mathbb{P}(X=k) = \frac{1}{N}, \quad k \in \{1,\ldots,N\}. \tag{2.4.7}$$

En utilisant les formules $\sum_{k=1}^{N} k = N(N+1)/2$ et $\sum_{k=1}^{N} k^2 = N(N+1)(2N+1)/6$, on trouve que $\mathbb{E}[X] = (N+1)/2$ et $\mathrm{Var}(X) = (N^2-1)/12$.

2.4.5 Loi Géométrique

Définition 2.14. *On dit que X suit une loi géométrique de paramètre $p \in]0,1[$ si pour chaque $n \in \mathbb{N}^*$,*

$$\mathbb{P}(X=n) = (1-p)^{n-1} p. \tag{2.4.8}$$

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

$$\sum_{n=1}^{\infty} r^n = \frac{r}{1-r} =: f(r), 0 \leqslant r < 1, \tag{2.4.9}$$

$$\sum_{n=1}^{\infty} nr^{n-1} = f'(r), \sum_{n=1}^{\infty} n(n-1)r^{n-2} = f''(r) \tag{2.4.10}$$

on déduit que

$$\mathbb{E}[X] = \frac{1}{p}, \quad \mathrm{Var}(X) = \frac{1-p}{p^2}. \tag{2.4.11}$$

2.4.6 Loi de Poisson

Définition 2.15. *On dit que X suit une loi de Poisson de paramètre $\lambda \in]0, \infty[$ si pour tout $k \in \mathbb{N}$,*

$$\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}. \tag{2.4.12}$$

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

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}, \tag{2.4.13}$$

on peut déduire que

$$\mathbb{E}[X] = \mathrm{Var}(X) = \lambda. \tag{2.4.14}$$

Remarque 2.16. Malgré l'aspect « catalogue » de la présentation des lois usuelles, il faut avoir à l'esprit que toutes les lois ne rentrent pas dans une de ces catégories. Par exemple, si X prend les valeurs 0, 1, 2 et 3 avec probabilités respectives 1/4, 1/4, 0 et 1/2, cela ne correspond ni à la loi de Bernoulli (qui ne prend que deux valeurs), ni à la loi binomiale de paramètres $(3, p)$ pour un certain p, ni à la loi discrète uniforme, etc... Par conséquent, lorsque l'on demande la loi d'une variable aléatoire X donnée, on peut :

1. si c'est une loi usuelle, l'identifier et donner les paramètres ;
2. si ce n'est pas une loi usuelle, il faut donner les valeurs prises ainsi que les probabilités correspondantes.

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 2.17. *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 2.18. 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 2.19. *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

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

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 (2.4.9),

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

- Loi de Poisson de paramètre λ : par (2.4.13),

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

2.6 Loi conditionnelle et espérance conditionnelle

Définition 2.20 (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*

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

Définition 2.21. *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*

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

Remarque 2.22. Si X est indépendante de Y, alors la loi conditionnelle de Y sachant X est simplement la loi de Y.

Définition 2.23 (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 $x_1 \in E_1, \ldots, x_n \in E_n$,*

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

Proposition 2.24. *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 de loi conditionnelle 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 $k \in \{1, \ldots, n\}$,

$$\mathbb{P}(X_n = 1 \mid S_n = k) = \frac{\mathbb{P}(X_n = 1, S_n = k)}{\mathbb{P}(S_n = k)} \tag{2.6.4}$$
$$= \frac{\mathbb{P}(X_n = 1, S_{n-1} = k-1)}{\mathbb{P}(S_n = k)} \tag{2.6.5}$$

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

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

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 $\mathbb{P}(X_n = 1 \mid S_n = k) = k/n$, formule valide aussi pour $k = 0$. Similairement, on trouve $\mathbb{P}(X_n = 0 \mid S_n = k) = 1 - k/n$.

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 2.25. *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 2.26. 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 $x \in \{0, \ldots, n\}$,

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

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

Exercice 21. Soient X_1 de loi Poisson de paramètre $\lambda_1 > 0$ et X_2 de loi Poisson de paramètre $\lambda_2 > 0$. Ces deux variables sont supposées indépendantes entre elles.

1. Déterminer la loi de $X_1 + X_2$.
2. Pour $\ell \in \mathbb{N}$, calculer la loi de X_1 sachant l'événement $\{X_1 + X_2 = \ell\}$.

2.6.2 Loi conditionnelle de X sachant un événement E

Définition 2.27. *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 2.28 (Concept de couples de variables aléatoires)**.** *Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X\colon \Omega \to I$, $Y\colon \Omega \to J$ des variables aléatoires, où $I, J \subset \mathbb{R}$. 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 2.29 (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 :*

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

Exemple 2.30. On peut représenter la loi jointe dans un tableau.

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

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

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

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

Remarque 2.31. 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 2.32. *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*

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

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

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

Exemple 2.33. 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 2.34. 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

x_i \ y_j	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

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

Si X est indépendante de Y, en prenant $g(x, y) = s^{x+y}$, $s \in [0, 1]$, on trouve que $G_{X+Y}(s) = G_X(s) G_Y(s)$.

2.7.4 Covariance

Définition 2.35 (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

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

Propriétés de la covariance

Proposition 2.36. *Si X et Y sont deux variables aléatoires indépendantes de carré intégrable, alors* $\operatorname{Cov}(X,Y) = 0$.

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

Proposition 2.37. *Soient X, Y et Z trois variables aléatoires discrètes de carré intégrable et $a, b \in \mathbb{R}$. On a :*

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

Proposition 2.38 (Covariance et variance)**.** *Si X et Y sont deux variables aléatoires de carré intégrable, alors*

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

Si de plus X et Y sont indépendantes, alors

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

2.7.5 Coefficient de corrélation linéaire

Définition 2.39 (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*

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

Proposition 2.40. *Si X et Y sont indépendantes, alors $\rho(X,Y) = 0$.*

Proposition 2.41. *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 2.42. *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 $X = aY + b$.*

2.8 Exercices

Exercice 22. Une variable aléatoire X prend les valeurs $k \in \{1, 2, 3, 6\}$, avec probabilité proportionnelle à k. Préciser la loi de X, son espérance, sa variance, et le coefficient de corrélation linéaire entre X et $(X-3)^2$.

Exercice 23. Soient X et Y deux variables aléatoires discrètes à valeurs dans $\{1, 2, \ldots, n\}$ telles que

$$\mathbb{P}\left(X = i, Y = j\right) = A\, C_{n-1}^{i-1} C_{n-1}^{j-1},\ \forall\,(i, j) \in \{1, 2, \ldots, n\}, \tag{2.8.1}$$

où A est une constante.

1. Trouver la constante A pour que (2.8.1) définisse une loi de probabilité.
2. Déterminer la loi marginale de X et celle de Y.
3. Les variables X et Y sont-elles indépendantes ?
4. Calculer la loi conditionnelle de X sachant $\{Y = j\}$.
5. Calculer $\mathbb{E}\left[X\right]$ et $\mathrm{Var}\left(X\right)$.

Exercice 24. On lance n fois une pièce. On désigne par $p \in]0, 1[$ la probabilité d'avoir pile. Soit X le nombre de piles obtenu après ces n lancers et Y celui de face. Calculer le coefficient de corrélation linéaire de X et Y.

Exercice 25. On considère un couple de variables aléatoires (X, Y) à valeurs dans $\mathbb{N}^* \times \mathbb{N}^*$ dont la loi jointe est donnée par

$$\mathbb{P}(X = k, Y = \ell) = \begin{cases} p^2(1-p)^{\ell-2} & \text{si } 1 \leqslant k < \ell, \\ 0 & \text{sinon,} \end{cases} \tag{2.8.2}$$

avec $p \in]0, 1[$.

1. Vérifier qu'on a bien ainsi défini une loi de probabilité.
2. Déterminer la loi marginale de X et de Y.
3. Les variables aléatoires X et Y sont-elles indépendantes ?
4. Pour les valeurs de ℓ pour lesquelles cela a un sens, déterminer la loi de X sachant $\{Y = \ell\}$.

Exercice 26. Un étudiant s'ennuie en cours et passe son temps à regarder par la fenêtre les feuilles tomber d'un arbre. On admet que le nombre de feuilles tombées à la fin du cours est une variable aléatoire X qui suit une loi de Poisson de paramètre $\lambda > 0$.

1. Donner $\mathbb{P}\left(X = k\right)$ et les valeurs possibles pour k.
2. Pour $\lambda > 0$, donner la valeur de
$$e^\lambda = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}.$$

3. Calculer l'espérance et la variance de X.
   À chaque fois qu'une feuille tombe par terre, l'étudiant lance une pièce qui donne pile avec une probabilité p et face avec une probabilité $q = 1 - p, p \in (0,1)$. On note F et P le nombre de faces et de piles obtenus respectivement.
4. Pour $k \in \mathbb{N}$ fixé, expliquer de manière simple pourquoi la loi de F sachant $X = k$ est une loi binomiale dont on précisera les paramètres. En déduire l'expression de $\mathbb{P}(F = a|X = k)$. On donnera les valeurs possibles pour a.
5. Pour $k \in \mathbb{N}$ et $a \in \mathbb{N}$, calculer la quantité $\mathbb{P}(\{X = k\} \cap \{F = a\})$.
6. Montrer que $\mathbb{P}(F = a) = e^{-\lambda q} \frac{(\lambda q)^a}{a!}$. Identifier la loi de F.
7. Donner, sans calculs, la loi de P.
8. Montrer que P et F sont indépendantes.
9. Calculer l'espérance de PF et la variance de $P + F$.

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 $\varepsilon > 0$,*

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

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

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

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*

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

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 $k \in \mathbb{N}$,*

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

Exemple 3.6. 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 de paramètre λ.

Exercice 27.

1. Soient $X_1, \dots, X_n$ n variables aléatoires indépendantes de lois respectives $\mathcal{P}(\lambda_1), \dots, \mathcal{P}(\lambda_n)$. Déterminer la loi de $\sum_{i=1}^n X_i$.
2. Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires de lois binomiales $\mathcal{B}(n, p_n)$ où la suite p_n vérifie $n \times p_n \to \lambda > 0$ (on remarquera que cette condition implique que $p_n \to 0$). Montrer que la suite $(X_n)_{n\geqslant 1}$ converge en loi vers une loi de Poisson $\mathcal{P}(\lambda)$.
3. Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires de lois binomiales $\mathcal{B}(n, p_n)$ où la suite p_n vérifie $n \times p_n \to 0$ Montrer que la suite $(X_n)_{n\geqslant 1}$ converge en loi vers une variable aléatoire constante (préciser la valeur de la constante).

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 3.7 (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 $\left(\overline{X_n} - \mathbb{E}[X_1]\right)_{n\geqslant 1}$ converge en moyenne d'ordre deux vers 0. En particulier,*

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

Démonstration. Il suffit de montrer que

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

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

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

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 limite central

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 3.8. *(Théorème limite central (TLC)] 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 TLC 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, $0,95$).

Exercice 28. Soient $X_1, ..., X_n, ...$ des variables aléatoires, indépendantes et de même loi de Poisson de paramètre 1.

1. Que vaut $\mathbb{P}(X_1 = k)$ et quelles sont les valeurs possibles prises par la variable aléatoire X_1 ?
2. Calculer la fonction génératrice de $\sum_{i=1}^n X_i$ et identifier la loi de $\sum_{i=1}^n X_i$.

3. Donner la définition de la convergence en probabilité.
4. Est-ce que $\sum_{i=1}^{n} X_i/n$ converge en probabilité ? Si oui vers quoi ?
5. Démontrer le résultat de la question 4. en utilisant une inégalité du cours.
6. Énoncer le théorème limite central.
7. En utilisant le théorème limite central donner la limite de $\mathbb{P}\left(\sum_{i=1}^{n} X_i \leqslant n\right)$ quand $n \to \infty$.
8. En déduire la valeur de
$$\lim_{n\to\infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!}.$$

4 Statistique

4.1 Introduction à la statistique

4.1.1 Objectifs

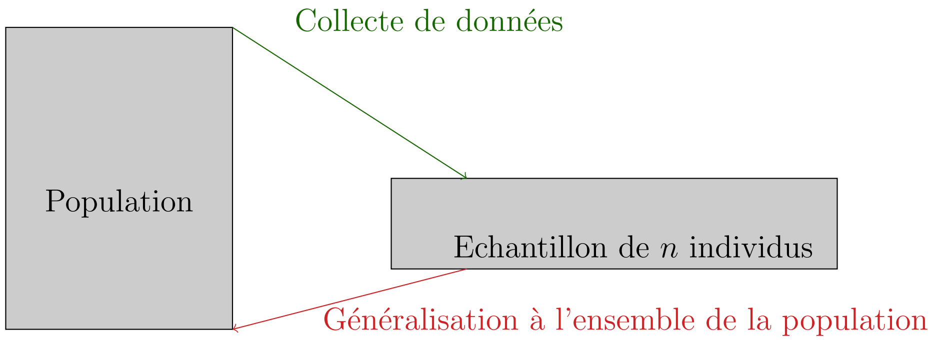
Définition 4.1. *La **statistique** est la science dont l'objet est de recueillir, de traiter et d'analyser des données issues de l'observation de phénomènes aléatoires, c'est-à-dire dans lesquels le hasard intervient.*

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 $x_1, \dots, x_n$.

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

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

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 $E = \{e_1, \dots, e_k\}$.

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%

FIGURE 1 – 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.

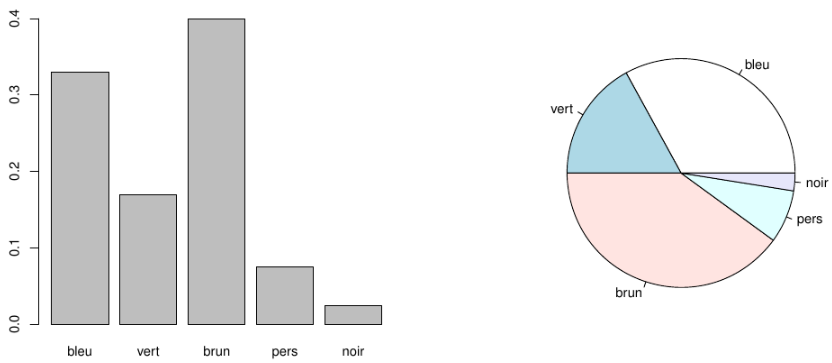
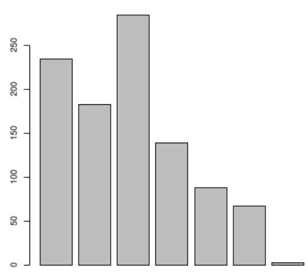


FIGURE 2 – Diagramme en bâtons (gauche), sectoriel (droite)

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.

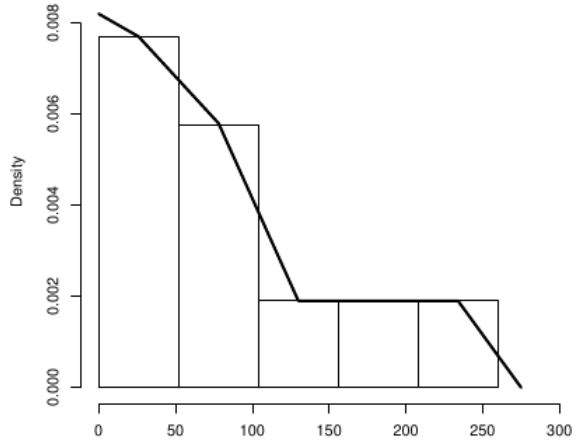


FIGURE 3 – Histogramme à classes de même largeur et polygone des fréquences

- 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).

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 :*

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

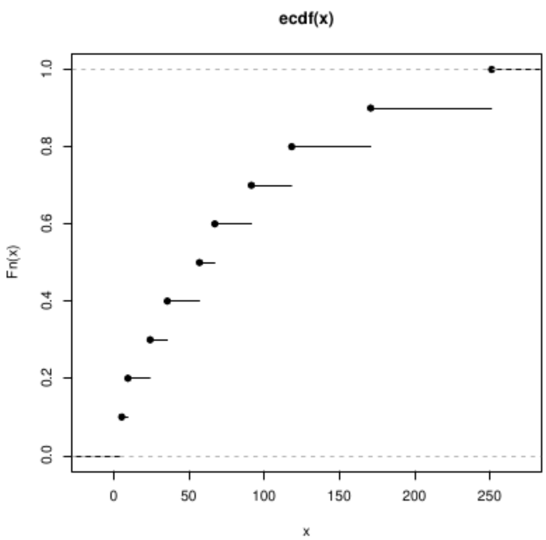


Figure 4 – Exemple de fonction de répartition empirique

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* $\frac{x^*_{\frac{n}{2}} + x^*_{\frac{n}{2}+1}}{2}$

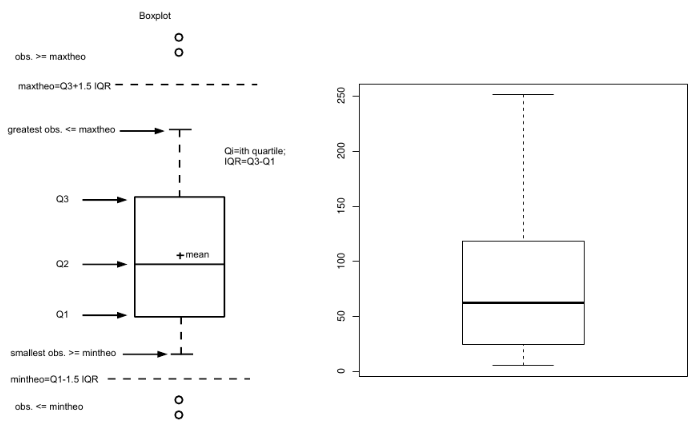


FIGURE 5 – Boxplot : explication (à gauche) et exemple (à droite)

La médiane n'est pas sensible aux valeurs aberrantes.

Définition 4.11 (Variance). *La variance empirique d'un échantillon $x_1, \dots, x_n$ 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 $s_n = \sqrt{s_n^2}$.*

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

Définition 4.14 (Quantile empirique). *Soit $x_1, \dots, x_n$ un échantillon et $x_1^* \leqslant x_2^* \leqslant \dots \leqslant x_n^*$ 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.*

4.3 Exercices

Exercice 29. Paul a noté pendant 12 jours la température en degré Celsius, au lever du jour :

$$-3;\ -4;\ 0;\ 1;\ 5;\ 5;\ 2;\ -1;\ -5;\ 2;\ 6;\ 7.$$

1. Calculer la moyenne de cette série.
2. Ranger cette série dans l'ordre croissant.
3. Déterminer la médiane de cette série.
4. Déterminer les premier, deuxième et troisième quartiles de cette série.
5. Calculer l'étendue de cette série.

Exercice 30. 1. On considère la série suivante :

5; 10; 5; 8; 4; 11; 6; 4; 8; 7; 9; 6; 13; 15; 8; 7; 5; 10

   (a) Quel est l'effectif total de cette série ?
   (b) Quelle est la médiane de cette série ?
   (c) Quel est le 1er quartile de cette série ?
   (d) Quel est l'écart interquartile de cette série ?

2. On considère la série suivante :

valeurs	22	23	24	25	26	27
fréquences cumulées	0.15	0.24	0.37	0.53	0.79	1

   (a) Quel est le pourcentage de la valeur 25 ?
   (b) Quelle est la médiane de cette série ?
   (c) Quel est le 1er quartile de cette série ?

Exercice 31. Suite aux mauvais résultats obtenus par les étudiants à un examen, un professeur envisage d'aug- menter toutes les notes des étudiants présents à l'examen de 10%. En d'autres termes, si un étudiant x a obtenu la note de x, celle-ci sera transformée en $y = x + 10\%x = x + x/10 = 11/10x$. Les affirmations suivantes sont-elles vraies ou fausses. Un justificatif précis (une ligne suffit pour chaque item) est demandé :

1. La moyenne empirique des résultats ne sera pas modifiée et leur écart-type empirique non plus.
2. La moyenne empirique des résultats sera multipliée par 10 et leur écart-type empirique restera inchangé.
3. La moyenne empirique des résultats augmentera de 10% et leur écart-type empirique restera inchangé.
4. La moyenne empirique des résultats et leur variance empirique augmenteront de 10%.
5. La moyenne empirique des résultats augmentera de 10% et leur variance empirique augmentera de 21%.
6. La moyenne empirique des résultats et leur variance empirique seront multipliées par 10.
7. La moyenne empirique des résultats augmentera de 10% et leur médiane restera inchangée.

8. La médiane et la variance empirique des résultats augmenteront de 10%.
9. L'écart interquartile et la variance empirique des résultats augmenteront de 21%.

Exercice 32. Voici différentes questions à choix unique (une seule réponse possible !). Justifier votre réponse en une ligne.

1. On dispose d'un échantillon de taille n noté $x_1, \ldots, x_n$. On ajoute une donnée à ce nouvel échantillon, qui contient donc maintenant $n+1$ données. Supposons que cette donnée est la médiane de l'échantillon initial $x_1, \ldots, x_n$. Quelle conséquence cela a sur la moyenne empirique de l'échantillon de taille $n+1$?
   (a) La moyenne empirique augmente
   (b) La moyenne empirique diminue
   (c) La moyenne empirique ne change pas
   (d) Aucune des réponses précédentes.
2. Plus la moyenne empirique de l'échantillon $x_1, \ldots, x_n$ est élevée et plus sa variance empirique est élevée.
   (a) Vrai.
   (b) Faux.
3. Soit un groupe de 10 sujets passant un test, noté sur 100, mais pouvant donner lieu à une note positive comme négative. La moyenne empirique obtenue au test par les 10 sujets est égale à 75. Si nous enlevons 75 points à chacun des sujets, la variance empirique
   (a) augmente.
   (b) diminue.
   (c) ne change pas.
   (d) devient égale à 0.
4. Soient $x_1, \ldots, x_n$, n distances mesurées en centimètres. Notons $\overline{x_n}$ la moyenne empirique de cet échantillon. Le fait de retraduire les mesures en mètres nous fournit une nouvelle série $y_1, \ldots, y_n$, de moyenne empirique $\overline{y_n}$. Les moyennes empiriques xn et y n sont liées entre elles par la relation suivante
   (a) $y_n = x_n$.
   (b) $\overline{y_n} = 100\overline{x_n}$.
   (c) $\overline{y_n} = \overline{x_n}/100$.
   (d) $\overline{y_n} = \overline{x_n} - 100$.
   (e) $\overline{y_n} = \overline{x_n}/10000$.
5. On se place dans le même contexte que dans 4) mais cette fois-ci on s'intéresse à la variance empirique. Les variances empiriques s_x^2 et s_y^2 sont liées entre elles par la relation suivante
   (a) $s_y^2 = 100 s_x^2$.

(b) $s_y^2 = 10000 s_x^2$.

(c) $s_y^2 = s_x^2/100$.

(d) $s_y^2 = s_x^2/1000$.

(e) $s_y^2 = s_x^2/10000$.