Probabilités

1 Equiprobabilité sur un ensemble fini, dénombrement **4**
1.1 Opérations sur les parties d'un ensemble 5
1.2 Quelques dénombrements classiques 7
1.2.1 Produit cartésien, listes ordonnées et règle du produit 7
1.2.2 Permutations et listes ordonnées sans répétition 9
1.2.3 Arrangements et tirages ordonnés sans remise 9
1.2.4 Coefficients binomiaux, parties d'un ensemble et tirages non ordonnés 10

2 Espace de probabilité **12**
2.1 Tribu et événements 13
2.2 Probabilité 13
2.3 Probabilité sur un ensemble discret 15
2.3.1 Caractérisation 15
2.3.2 Rappels sur les séries à termes positifs 16
2.3.3 Un exemple d'ensemble infini non dénombrable 18
2.4 Probabilité à densité sur $\mathbb{R}$ 18

3 Indépendance d'événements et probabilité conditionnelle **20**
3.1 Indépendance d'événements 21
3.1.1 Indépendance de deux événements 21
3.1.2 Indépendance d'une famille d'événements 21
3.2 Application : le jeu de pile ou face, ou schéma de Bernoulli fini 22
3.2.1 Probabilité d'un événement élémentaire. 22
3.2.2 Probabilité d'obtenir k piles 23
3.3 Probabilité conditionnelle 23
3.3.1 Définition 24
3.3.2 Propriétés 25
3.4 Complément 26
3.4.1 Le jeu infini de pile ou face, ou schéma de Bernoulli 26
3.4.2 Notion de presque sûr 27

4 Premier TP. Prise en main de Scilab, premières simulations **28**
4.1 Prise en main 28
4.2 Premières simulations 30
4.2.1 Histogramme pour le générateur `rand()` de loi uniforme sur $[0, 1]$ 30
4.2.2 Simulation de lancers d'un dé équilibré. 31
4.2.3 Evolution de la proportion de pile dans une suite de lancers 31

5 Variables aléatoires réelles **33**
5.1 Variable aléatoire réelle . . . 34
5.1.1 Définitions . . . 34
5.1.2 Variables aléatoires discrètes . . . 34
5.1.3 Variable aléatoire à densité . . . 35
5.1.4 Autres variables aléatoires réelles . . . 36
5.2 Simulation de variables aléatoires réelles . . . 36
5.2.1 Exemples de simulation de variables aléatoires discrètes . . . 36
5.2.2 Exemples de simulation de variables aléatoires à densité . . . 37

6 Espérance et variance des variables aléatoires réelles **38**
6.1 Espérance des variables aléatoires réelles . . . 39
6.1.1 Définitions . . . 39
6.1.2 Propriétés de l'espérance . . . 40
6.1.3 Espérance d'une fonction d'une variable aléatoire . . . 41
6.2 Variance d'une variable aléatoire réelle . . . 41
6.3 Calculs pour les v.a. discrètes classiques . . . 43
6.3.1 Loi de Bernoulli de paramètre $p \in [0,1]$ . . . 43
6.3.2 Loi binomiale de paramètres $n \in \mathbb{N}^*$, $p \in [0,1]$ . . . 43
6.3.3 Loi uniforme sur $\{0,\dots,n\}$ . . . 44
6.3.4 Loi de Poisson de paramètre $\lambda \in]0,\infty[$ . . . 44
6.3.5 Loi géométrique de paramètre $p \in]0,1]$ . . . 45
6.4 Calculs pour les v.a. à densité classiques . . . 46
6.4.1 Lois uniformes . . . 46
6.4.2 Lois exponentielles . . . 46
6.4.3 Lois normales . . . 47

7 Second TP. Autour des variables aléatoires **49**

8 Vecteur aléatoire discret et indépendance entre variables aléatoires discrètes **52**
8.1 Vecteur aléatoire de $\mathbb{R}^2$ . . . 53
8.2 Calcul des lois marginales . . . 54
8.3 Théorème de transfert pour un vecteur aléatoire discret . . . 54
8.4 Espérance et covariance . . . 55
8.5 Indépendance de deux variables aléatoires : définitions et critères . . . 57
8.6 Indépendance et covariance . . . 58
8.7 Indépendance de plusieurs variables aléatoires discrètes . . . 58
8.8 Loi d'une somme de variables aléatoire indépendantes . . . 59
8.8.1 Sommes de variables de Bernoulli indépendantes . . . 59
8.8.2 Sommes de variables de Poisson indépendantes . . . 60

9 Théorèmes limite **61**
9.1 Lois faible et forte des grands nombres . . . 62
9.1.1 Loi faible des grands nombres : énoncé et commentaires . . . 62
9.1.2 Démonstration de la loi faible des grands nombres . . . 63
9.1.3 Applications aux intervalles de fluctuations . . . 63
9.1.4 Loi forte des grands nombres . . . 64
9.2 Théorème de la limite centrale . . . 64
9.2.1 Enoncé et commentaires . . . 64

9.2.2 Applications aux intervalles de fluctuations 66
9.2.3 Approximation de la loi binomiale 67
9.3 Initiation à l'estimation statistique 67
9.3.1 Modélisation probabiliste. 67
9.3.2 Estimation à l'aide de la loi des grands nombres. 68
9.3.3 Contrôle de l'erreur d'approximation à l'aide de la loi faible des grands nombres. 68
9.3.4 Contrôle de l'erreur d'approximation à l'aide du théorème central limite. 69

Chapitre 1

Equiprobabilité sur un ensemble fini, dénombrement

Sommaire

1.1 Opérations sur les parties d'un ensemble 5
1.2 Quelques dénombrements classiques 7
1.2.1 Produit cartésien, listes ordonnées et règle du produit 7
1.2.2 Permutations et listes ordonnées sans répétition 9
1.2.3 Arrangements et tirages ordonnés sans remise 9
1.2.4 Coefficients binomiaux, parties d'un ensemble et tirages non ordonnés 10

Objectifs :

• Savoir manipuler des ensembles, des parties et les opérations simples sur les ensembles.

• Savoir passer de la description non mathématisée d'une expérience aléatoire équiprobable à une formalisation mathématique.

• Savoir faire des dénombrements simples.

Mots-clés :

• Ensemble fini, cardinal.

• Probabilité uniforme sur un ensemble fini.

• Produit cartésien d'ensembles, n-uplet.

• Nombres de combinaisons, nombre d'arrangements, nombre de permutations.

Pour décrire une expérience aléatoire, on commence par décrire l'ensemble Ω des résultats possibles. Un événement est alors un sous-ensemble de Ω :

▷ **Exemple.** Lancer d'un dé à six faces. On peut modéliser cette expérience par l'univers $\Omega = \{1, 2, 3, 4, 5, 6\}$, dans lequel on liste les résultats possibles de l'expérience. L'événement "le résultat est divisble par 3" correspond à la partie $\{3, 6\}$.

Pour décrire complètement l'expérience aléatoire, il faut dire aussi quelle est la probabilité de chacun des résultats de l'expérience. Un cas particulèrement simple est celui de l'équiprobabilité, dans lequel toutes les issues possibles ont la même probabilité de se réaliser :

Définition 1.1 *Soit Ω un univers fini. L'*équiprobabilité*, ou* probabilité uniforme*, sur Ω, est l'application $\mathbb{P} : \mathcal{P}(\Omega) \to [0, 1]$ qui à une partie A de Ω associe*

$$\mathbb{P}(A) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)}.$$

On remarque que tous les singletons ont la même probabilité $1/\,\mathrm{Card}(\Omega)$, ce qui justifie le terme d'*équi*probabilité ou de probabilité *uniforme*. Dans ce cadre, le calcul de la probabilité d'un événement se ramène au dénombrement d'ensembles, c'est-à-dire au calcul de cardinal d'ensembles finis.

▷ **Exemple.** Si on veut modéliser un lancer de dé équilibré, on va prendre pour $\mathbb{P}$ l'équiprobabilité :

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \mathbb{P}(\{3\}) = \mathbb{P}(\{4\}) = \mathbb{P}(\{5\}) = \mathbb{P}(\{6\}) = \frac{1}{6}.$$

Par exemple, on peut calculer

$$\mathbb{P}(\text{« résultat divisible par trois »}) = \mathbb{P}(\{3, 6\}) = \frac{2}{6} = \frac{1}{3}.$$

Dans ce chapitre, nous rappelons un certain nombres d'opérations sur les parties d'un ensemble et nous introduisons certains dénombrements élémentaires. Mais nous verrons plus tard la définition générale d'un espace de probabilité, qui permettra de modéliser des expériences aléatoires plus générales (avec un nombre d'issues possibles infini, ou non-équiprobables)

1.1 Opérations sur les parties d'un ensemble

Pour décrire un *ensemble*, on peut donner la liste de ses éléments entre accolades : c'est ce qu'on appelle la *définition en extension*. Un ensemble n'est pas ordonné, ainsi $\{1, 3, 2\}$ et $\{1, 2, 3\}$ décrivent le même ensemble. Dans cet exemple, 2 est un *élément* de $\{1, 2, 3\}$, ce qui se note

$$2 \in \{1, 2, 3\} \qquad \text{et se lit « 2 } \textit{appartient à } \{1, 2, 3\}.$$

$\{1, 3\}$ est *une partie* ou *un sous-ensemble* de $\{1, 2, 3\}$; on dit que $\{1, 3\}$ est *inclus* dans $\{1, 2, 3\}$, ce qui se note

$$\{1, 3\} \subset \{1, 2, 3\}.$$

On peut aussi définir un ensemble *en compréhension*, c'est-à-dire en donnant la ou les propriétés qui caractérisent ses éléments :

$$\{2, 4, 6\} = \{x \in \{1, \dots, 7\} : \ 2 \text{ divise } x\}.$$

Quelques ensembles usuels en mathématiques :
– l'ensemble vide qui ne contient aucun éléments : $\varnothing = \{\,\}$,
– l'ensemble des entiers naturels $\mathbb{N} = \{0, 1, 2, ...\}$.

On note $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω.

$\triangleright$ **Exemple.** Soit $\Omega = \{1, 2\}$: listons ses parties :
– il y a des parties qui ne comportent qu'un seul élément : $\{1\}$ et $\{2\}$;
– il y a une partie à deux éléments $\{1, 2\}$;
– ne pas oublier la partie vide : $\varnothing$.
Au final, $\mathcal{P}(\Omega) = \{\varnothing, \{1\}, \{2\}, \{1, 2\}\}$.

♣ **Exercice 1.1.** Soit $\Omega = \{1, 2, 3\}$. Écrire l'ensemble $\mathcal{P}(\Omega)$ (on pourra classer ces parties suivant leur nombre d'éléments ; ne pas oublier l'ensemble vide !). Combien Ω a-t-il de parties distinctes ? de parties distinctes à deux éléments ?

$\triangleright$ **Exemple.** Je prends simultanément deux cartes au hasard dans un jeu de cartes. Pour modéliser cette expérience, je numérote mes cartes de 1 à 5, et je prends pour Ω l'ensemble des parties à 2 éléments de l'ensemble $\{1, 2, 3, 4, 5\}$:

$$\Omega = \{\{1, 2\},\ \{1, 3\},\ \{1, 4\},\ \{1, 5\},\ \{2, 3\},\ \{2, 4\},\ \{2, 5\},\ \{3, 4\},\ \{3, 5\},\ \{4, 5\}\}.$$

Sans précisions supplémentaires, je peux supposer que toutes ces issues ont la même probabilité. Comme il y en a 10, je leur affecte donc chacune la probabilité $1/10$.
Quelle est la probabilité que j'ai, dans mes deux cartes, la carte numéro 1 ?
Je note A cet événement : $A = \{\{1, 2\},\ \{1, 3\},\ \{1, 4\},\ \{1, 5\}\}$, donc $\mathbb{P}(A) = 4/10 = 2/5$.

Soit Ω un ensemble, et soient A et B deux parties de Ω.

- la *réunion* de A et B, notée $A \cup B$, est la partie de Ω composée des éléments de A et des éléments de B :

  $$\{1, 2, 5, 6\} \cup \{2, 3, 5, 7\} = \{1, 2, 3, 5, 6, 7\}.$$

  La réunion de deux ensembles correspond au lien logique « *ou* » : $x \in A \cup B \Leftrightarrow x \in A$ **ou** $x \in B$.

  ♠ **Attention !** Noter que si x est dans $A \cup B$, alors x peut être à la fois dans A **et** dans B.
  La réunion d'une famille d'ensembles correspond au quantificateur $\exists$ (« il existe ») : soit $(A_i)_{i \in I}$ une famille de parties d'un ensemble Ω. Alors
  $$x \in \bigcup_{i \in I} A_i \Leftrightarrow \exists i \in I, x \in A_i.$$

- l'*intersection* de A et B, notée $A \cap B$, est la partie de Ω composée des éléments qui sont à la fois dans A et dans B :
  $$\{1, 2, 5, 6\} \cap \{2, 3, 5, 7\} = \{2, 5\}.$$
  L'intersection de deux ensembles correspond au lien logique « *et* » : $x \in A \cap B \Leftrightarrow x \in A$ **et** $x \in B$.
  Deux événements sont *disjoints* si leur intersection est vide : $A \cap B = \varnothing$.
  L'intersection d'une famille d'ensembles correspond au quantificateur $\forall$ (« pour tout » ou « quel que soit ») : soit $(A_i)_{i \in I}$ une famille de parties d'un ensemble Ω. Alors

  $$x \in \bigcap_{i \in I} A_i \Leftrightarrow \forall i \in I, x \in A_i.$$

- le *complémentaire* de A dans Ω, noté $\bar{A}$ (ou parfois A^c ou $\Omega \backslash A$), est la partie de Ω composée des éléments de Ω qui ne sont pas dans A :

  $$\Omega = \{1, 2, ..., 10\},\ A = \{1, 3, 5, 7, 8, 9, 10\},\ \bar{A} = \{2, 4, 6\}.$$

  Le passage au complémentaire correspond à la négation logique : $x \in \bar{A} \Leftrightarrow x \notin A$.
  Notons que
  $$A \cup \bar{A} = \Omega \ \text{ et } \ A \cap \bar{A} = \varnothing.$$

Propriétés 1.2 *Soient A, B et C trois parties d'un ensemble* Ω. *Alors*

- $\overline{A \cup B} = \bar{A} \cap \bar{B}$
- $\overline{A \cap B} = \bar{A} \cup \bar{B}$
- $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$
- $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

On peut visualiser ces propriétés à l'aide de *diagrammes de Venn*, ou diagramme en patate.

♣ **Exercice 1.2.** Soit A et B deux parties d'un ensemble Ω. Montrer et illustrer sur un diagramme de Venn que $A \cup B$ est l'union disjointe des trois parties $A \cap B^c$, $B \cap A^c$ et $A \cap B$.

Propriétés 1.3 *Pour toute suite de parties* $(A_i)_{i \geq 1}$ *d'un ensemble* E, *pour tout* $n \in \{1, 2, \ldots\} \cup \{\infty\}$

- $\overline{\left(\bigcap_{i=1}^{n} A_i\right)} = \bigcup_{i=1}^{n} \overline{A_i}$
- $\left(\bigcup_{i=1}^{n} A_i\right) \cap A = \bigcup_{i=1}^{n} (A_i \cap A)$
- $\overline{\left(\bigcup_{i=1}^{n} A_i\right)} = \bigcap_{i=1}^{n} \overline{A_i}$
- $\left(\bigcap_{i=1}^{n} A_i\right) \cup A = \bigcap_{i=1}^{n} (A_i \cup A)$

1.2 Quelques dénombrements classiques

1.2.1 Produit cartésien, listes ordonnées et règle du produit

Définition 1.4 *(produit cartésien d'ensembles)*
Soit E_1, E_2 *des ensembles. On définit le produit cartésien des ensembles* E_1 *et* E_2 *de la façon suivante :*

$$E_1 \times E_2 = \{(x_1, x_2) \text{ tels que } x_1 \in E_1 \text{ et } x_2 \in E_2\}.$$

Un élément (x_1, x_2) *de* $E_1 \times E_2$ *est appelé un* 2-uplet *ou* couple.

Remarque. $E \times F$ se lit « E croix F ». La définition se généralise à un produit de n ensembles.

▷ **Exemple.** Les ensembles E et F peuvent être les mêmes, et on considère alors $E \times E$? Lorsque l'on fait le produit cartésien n fois du même ensemble E, on note le résultat E^n. Par exemple, $E^3 = E \times E \times E$.
Pour décrire le résultat de quatre lancers successifs d'une pièce, on va prendre

$$\Omega = \{P, F\}^4 = \left\{ \begin{array}{l} (P,P,P,P), (P,P,P,F), (P,P,F,P), (P,P,F,F), \\ (P,F,P,P), (P,F,P,F), (P,F,F,P), (P,F,F,F), \\ (F,P,P,P), (F,P,P,F), (F,P,F,P), (F,P,F,F), \\ (F,F,P,P), (F,F,P,F), (F,F,F,P), (F,F,F,F) \end{array} \right\}.$$

Ω est composé de 24 quadruplets de $\{P, F\}$.

♠ **Attention !** Un n-uplet est ordonné et se note entre parenthèses, contrairement à une partie à n éléments qui est sans ordre et se note entre accolades. Par exemple, $(1, 2, 3) \neq (1, 3, 2)$ mais $\{1, 2, 3\} = \{1, 3, 2\}$.

Proposition 1.5 *(règle du produit)*
Soit $E_1, ..., E_n$ *des ensembles finis. Alors*

$$\mathrm{Card}(E_1 \times E_2 \times ... \times E_n) = \prod_{1 \leq i \leq n} \mathrm{Card}(E_i).$$

Démonstration. $(x_1, x_2, ..., x_n) \in E_1 \times E_2 \times ... \times E_n$ si et seulement si pour tout $i \in \{1, \dots, n\}$, $x_i \in E_i$. Il y a donc $\mathrm{Card}(E_i)$ choix pour chaque x_i, et donc $\prod_{1 \leq i \leq n} \mathrm{Card}(E_i)$ éléments dans $E_1 \times E_2 ... \times E_n$. □

En particulier, on a

$$\mathrm{Card}(E^n) = (\mathrm{Card}(E))^n.$$

◊ **En pratique.** En pratique, la règle du produit se voit bien quand on fait un dénombrement par arbre.

▷ **Exemple.** On lance deux dés standard non truqués, un rouge et un bleu. Quelle est la probabilité que le résultat du dé rouge soit pair et le résultat du dé bleu soit divisible par 3 ?
Solution : L'énoncé nous dit que les dés ne sont pas truqués, il s'agit donc d'une expérience aléatoire décrite par une équiprobabilité.
Décrivons l'univers de cette expérience aléatoire. Les deux dés étant de couleurs différentes, je peux reconnaître le résultat bleu du résultat rouge. Je dois donc considérer des couples ; je choisis de les ordonner dans l'ordre (résultat rouge, résultat bleu). Bien sûr, on aurait également pu choisir l'ordre inverse. Chaque dé a pour résultat un élément de $E = \{1, 2, 3, 4, 5, 6\}$. L'univers est donc le suivant :

$$\Omega = E \times E = E^2 = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} = \{1, 2, 3, 4, 5, 6\}^2.$$

Par la proposition précédente, $\mathrm{Card}(\Omega) = \mathrm{Card}(E) \times \mathrm{Card}(E) = 6 \times 6 = 36$.
Déterminons maintenant la partie A de Ω qui correspond à l'événement « le résultat du dé rouge est pair et le résultat du dé bleu est divisible par 3 » :

$$A = \{2, 4, 6\} \times \{3, 6\}.$$

Par la proposition précédente, $\mathrm{Card}(A) = \mathrm{Card}(\{2, 4, 6\}) \times \mathrm{Card}(\{3, 6\}) = 3 \times 2 = 6$. Donc

$$\mathbb{P}(A) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)} = \frac{6}{36} = \frac{1}{6}.$$

♣♣ **Exercice 1.3.** Quelle est la probabilité que la somme de deux lancers de dés équilibrés soit 5 ?

♣ **Exercice 1.4.** On lance trois dés, un à six faces (numérotées de 1 à 6), un à neuf faces (numérotées de 1 à 9) et un à treize faces (numérotées de 1 à 13), non truqués. Donner l'univers Ω adapté à cette expérience et calculer son cardinal. Calculer la probabilité que la somme des trois faces obtenues soit égale à 26.

▷ **Exemple.** Je tire successivement trois cartes au hasard dans un jeu de 32, en remettant à chaque fois la carte piochée dans le paquet avant de piocher la suivante. Donner un univers associé à cette expérience aléatoire et calculer la probabilité d'obtenir trois as.
Solution : Décrivons l'univers de cette expérience aléatoire : je peux décider, pour simplifier, de numéroter les cartes de 1 à 32, et je suppose que les quatre as sont les cartes numérotées de 1 à 4. Ceci sert à simplifier les choses et n'est pas obligatoire, on peut aussi nommer les cartes par leurs noms habituels ($2\clubsuit$, $R\heartsuit \dots$) ou choisir une autre convention. Une issue de mon expérience est un triplet d'éléments de $\{1, 2, ..., 32\}$ puisque le tirage est avec remise, on choisit donc

$$\Omega = \{1, 2, ..., 32\} \times \{1, 2, ..., 32\} \times \{1, 2, ..., 32\} = \{1, 2, ..., 32\}^3.$$

Ainsi, $\mathrm{Card}(\Omega) = 32^3$. Je suppose que je pioche les cartes sans tricher, et je fais donc l'hypothèse d'équiprobabilité. Il me reste à déterminer la partie A de Ω correspondant à l'événement « obtenir trois as » :

$$A = \{1, 2, 3, 4\} \times \{1, 2, 3, 4\} \times \{1, 2, 3, 4\} = \{1, 2, 3, 4\}^3.$$

Ainsi, $\mathrm{Card}(A) = 4^3$ et donc,

$$\mathbb{P}(A) = \frac{\mathrm{Card}\, A}{\mathrm{Card}(\Omega)} = \frac{4^3}{32^3} = \frac{1}{2^9}.$$

♣ **Exercice 1.5.** J'ai dans une chaussette 26 jetons, chacun avec une lettre de l'alphabet. Je pioche successivement avec remise 4 jetons. Donner un univers associé à cette expérience aléatoire et calculer la probabilité de ne piocher que des « A » ou des « B ».

1.2.2 Permutations et listes ordonnées sans répétition

Définition 1.6 *(permutation d'un ensemble fini)*
Soit E un ensemble fini de cardinal $n \in \mathbb{N}^*$. Une permutation des éléments de E est une bijection de $\{1, 2, ...n\}$ dans E ; c'est aussi une liste ordonnée sans répétition de tous les éléments de E ; c'est encore un tirage ordonné sans remise de tous les éléments de E.

▷ **Exemple.** Le 5-uplet $(3, 7, 2, 8, 4)$ est une liste ordonnée sans répétition de tous les éléments de l'ensemble $E = \{2, 3, 4, 7, 8\}$.

♣ **Exercice 1.6.** Donner la liste de toutes les liste ordonnées sans répétition des éléments de l'ensemble $\{1, 2, 3\}$, puis de l'ensemble $\{1, 2, 3, 4\}$. (Vérification : il y en a respectivement 6 et 24).

Proposition 1.7 *(permutations d'un ensemble fini)*
Le nombre de permutations d'un ensemble E à n éléments est $n! = n \times (n - 1) \times \cdots \times 1$.
Par convention, $0! = 1$.

Démonstration. Quitte à numéroter les éléments de E, il suffit de travailler avec les ensembles de la forme $E = \{1, 2, \dots, n\}$. Une permutation de $\{1, \dots, n\}$ est une bijection de $\{1, \dots, n\}$ dans $\{1, \dots, n\}$. En particulier, les images des points $\{1, \dots, n\}$ sont deux à deux distinctes. On a donc n choix pour l'image de 1, puis une fois l'image de 1 fixée, on a $n - 1$ choix pour l'image de 2...puis $n - (n - 2) = 2$ choix pour l'image de $n - 1$, puis $n - (n - 1) = 1$ choix pour l'image de n. Il y a donc $n!$ permutations d'un ensemble à n éléments. □

▷ **Exemple.** Lucie a 5 cartons devant elle, chacun portant une des lettres de son prénom. Elle les aligne au hasard devant elle. Quelle est la probabilité qu'elle ait écrit son prénom ?
Solution : Bien que ce ne soit pas explicitement demandé dans l'énoncé, je commence par décrire l'univers correspondant à cette expérience aléatoire : l'alignement réalisé par Lucie constitue une liste ordonnée des cinq cartons, donc Ω est l'ensemble des permutations de F où $F = \{C, E, I, L, U\}$; par conséquent $\text{Card}(\Omega) = 5! = 120$. Je note A l'événement « Lucie a écrit son prénom ». C'est donc le singleton

$$A = \{(L, U, C, I, E)\}.$$

Finalement $\mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)} = \frac{1}{120}$.

♣ **Exercice 1.7.** pour faire passer un examen oral dans sa classe, le professeur décide de tirer l'ordre de passage des 26 étudiants au hasard. Quelle est la probabilité que Sophie passe en dernier ? Quelle est la probabilité que Paul passe avant Marie ?

1.2.3 Arrangements et tirages ordonnés sans remise

Définition 1.8 *(Arrangements)*
Soit E un ensemble fini de cardinal $n \in \mathbb{N}^*$ et $k \in \mathbb{N}$. Un arrangement *de k éléments de E est une k-liste d'éléments de E* deux à deux distincts.

▷ **Exemple.** $(5, 1, 8)$ est un arrangement de 3 éléments de $\{1, 2, 3, 5, 7, 8, 9\}$.

Proposition 1.9 *(Nombre d'arrangements)*
Pour tous entiers naturels n et k, on pose $A_n^k = \frac{n!}{(n-k)!}$ si $k \leq n$ et $A_n^k = 0$ sinon.
Le nombre d'arrangements à k éléments d'un ensemble à n éléments est A_n^k.

Démonstration. Soit E un ensemble de cardinal k et soit F un ensemble de cardinal n. Quitte à numéroter les éléments de E et F, on peut se ramener au cas $E = \{1, \dots, k\}$ et $F = \{1, \dots, n\}$.
Si $k > n$, alors il n'existe pas d'injection de E dans F et comme, par définition, $A_n^k = 0$, il y a bien A_n^k injections.
Si $k \leq n$, une injection de E dans F est une application de E dans F telle que les éléments de E ont des images deux à deux distinctes dans F. Pour l'élément 1 de E, on a n choix possibles ; puis une fois l'image de 1 fixée, on a $n-1$ choix pour l'image de 2...puis $n-(k-2) = n-k+2$ choix pour l'image de $k-1$, puis $n-(k-1) = n-k+1$ choix pour l'image de k. Il y a donc $n \times (n-1) \times ... \times (n-k+2) \times (n-k+1)$ injections de E dans F, et c'est exactement A_n^k. □

▷ **Exemple.** 27 chevaux sont au départ du tiercé. Combien de résultats possibles existe-t-il ?
Solution : On doit exactement compter le nombre de 3-listes sans répétition de 27 chevaux, donc le nombre cherché vaut :

$$A_{27}^3 = 27 \times 26 \times 25.$$

♣ **Exercice 1.8.** Le bureau de l'association des étudiants est composé d'un président, d'un trésorier, et d'un secrétaire. L'association compte 73 membres. Combien de bureaux différents peut-on former ?

1.2.4 Coefficients binomiaux, parties d'un ensemble et tirages non ordonnés

Proposition 1.10 *(Coefficients binomiaux)*
Pour tous entiers naturels n et k, on pose $C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$ si $k \leq n$ et $\binom{n}{k} = 0$ sinon.
i) Pour tout entier naturel non nul n et pour tout entier naturel k tel que $0 \leq k \leq n-1$:

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}.$$

ii) Pour tout entier naturel n et pour tous nombres réels x et y, on a la formule du binôme *:*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

iii) Le nombre de parties à k éléments d'un ensemble à n éléments est $\binom{n}{k}$.
iv) Un ensemble à n éléments admet 2^n parties différentes.

Démonstration.
i) On fait le calcul explicite.
ii) et iii) se font par récurrence.
iv) On peut appliquer ii) avec $x = y = 1$ et utiliser iii).

□

♣ **Exercice 1.9.** Combien y a-t-il de façons de choisir un groupe de trois personnes pour faire la vaisselle dans une famille de 5 personnes ?

◊ **En pratique.** i) permet une construction par récurrence des nombres $\binom{n}{p}$, appelée le *triangle de Pascal* :

	0	1	2	3	4	5	6
0	1						
1	1	1					
2	1	2	1				
3	1	3	3	1			
4	1	4	6	4	1		
5	1	5	10	10	5	1	
6	1	6	15	20	15	6	1

▷ **Exemple.** Je pioche trois cartes en même temps dans un jeu de 32. Quelle est la probabilité que ce soit trois as ?

Solution : Il s'agit d'un tirage non ordonné sans remise (les trois cartes sont piochées en même temps). Je vais donc prendre pour univers Ω l'ensemble des parties à 3 éléments de l'ensemble des 32 cartes. Donc $\mathrm{Card}(\Omega) = \binom{32}{3}$.
Comme toujours dans ce genre d'exercice, on suppose que le tirage se fait sans pouvoir favoriser de cartes et on modélise donc l'expérience à l'aide de l'équiprobabilité.
Soit A l'événement « piocher trois as ». L'ensemble A est constitué des parties à trois éléments qui ne comportent que des as. Donc A est l'ensemble des parties à trois éléments de l'ensemble des 4 as, donc $\mathrm{Card}(A) = \binom{4}{3}$.
Finalement $\mathbb{P}(A) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)} = ... = \frac{1}{1240}$.

♣ **Exercice 1.10.** Combien de mots de 10 lettres comportant exactement 3 "A" et 7 "B" peut-on écrire ?

Chapitre 2

Espace de probabilité

Sommaire

2.1 Tribu et événements 13
2.2 Probabilité 13
2.3 Probabilité sur un ensemble discret 15
    2.3.1 Caractérisation 15
    2.3.2 Rappels sur les séries à termes positifs 16
    2.3.3 Un exemple d'ensemble infini non dénombrable 18
2.4 Probabilité à densité sur $\mathbb{R}$ 18

Objectifs :

• Modéliser une expérience aléatoire, c'est-à-dire construire un ensemble Ω des issues possibles (un univers) de l'expérience aléatoire, et le munir d'un outil (une probabilité) permettant de mesurer la chance d'obtenir par cette expérience un résultat donné, ou un ensemble de résultats donnés.
• Donner les propriétés d'une probabilité permettant de faire les calculs.

Mots-clés :

• Univers, probabilité.
• Événement, événement élémentaire, événements disjoints.
• Ensemble fini, cardinal, ensemble dénombrable.
• Probabilité uniforme sur un ensemble fini et probabilité uniforme sur $[0, 1]$.

Outils :

• Formule d'inclusion-exclusion, formule des probabilités totales.
• Axiomes et propriétés d'une probabilité.
• Caractérisations des probabilités sur les ensembles finis et dénombrables.

Pour modéliser une expérience aléatoire, on introduit un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ composé de trois objets : un ensemble Ω, appelé *univers*, une famille $\mathcal{F}$ de parties de Ω qui doit vérifier un certain nombre de propriétés, appelée *tribu*, et dont on parlera le moins possible dans ce cours, et enfin une *probabilité*, qui va attribuer à chaque élément de la tribu un nombre compris entre 0 et 1.
On voudrait généraliser le cas fini équiprobable pour pouvoir considérer des expériences aléatoires plus générales : modélisation d'un lancer de dé non équilibré, temps d'apparition d'un pile dans une suite de lancer, générateur de nombres aléatoires d'un ordinateur, modélisation de la durée de vie d'un composant électronique...

2.1 Tribu et événements

Définition 2.1 *Soit Ω un ensemble. Une famille $\mathcal{F}$ de parties de Ω est appelée une* tribu *si elle vérifie les propriétés suivantes :*
i) Ω est un élément de $\mathcal{F}$;
ii) (stabilité par complémentaire) Si A est un élément de $\mathcal{F}$, alors $\bar{A}$ est un élément de $\mathcal{F}$;
iii) (stabilité par union dénombrable) Si les $(A_i)_{i\in\mathbb{N}}$ sont des éléments de $\mathcal{F}$, alors $\bigcup_{i\in\mathbb{N}} A_i$ est un élément de $\mathcal{F}$.
L'ensemble Ω est appelé univers*. Les éléments de $\mathcal{F}$ sont appelés des* événements.

Remarque. Dans iii), on considère une famille *dénombrable* de parties de Ω : on en a besoin pour des raisons techniques. On retiendra dans un premier temps :
– Si $A \in \mathcal{F}$, alors $\bar{A} \in \mathcal{F}$.
– Si $A \in \mathcal{F}$ et $B \in \mathcal{F}$, alors $A \cup B \in \mathcal{F}$.
L'ensemble $\mathcal{P}(\Omega)$ de toutes les parties de Ω est une tribu. Si Ω est un ensemble fini, on prend (quasiment) toujours $\mathcal{F} = \mathcal{P}(\Omega)$.

♣ **Exercice 2.1.** Soit Ω un ensemble, muni d'une tribu $\mathcal{F}$.
• Montrer que $\varnothing \in \mathcal{F}$.
• Montrer que si $A \in \mathcal{F}$ et $B \in \mathcal{F}$, alors $A \cap B \in \mathcal{F}$.

2.2 Probabilité

Définition 2.2 *Une probabilité $\mathbb{P}$ sur un espace mesurable $(\Omega, \mathcal{F})$ est une application de $\mathcal{F}$ dans $[0, 1]$:*

$$\begin{array}{rcl} \mathbb{P} : \mathcal{F} & \to & [0,1] \\ A & \mapsto & \mathbb{P}(A), \end{array}$$

vérifiant les deux propriétés suivantes :
i) $\mathbb{P}(\Omega) = 1$,
ii) (σ-additivité) Si les $(A_i)_{i\in\mathbb{N}}$ sont des éléments de $\mathcal{F}$ deux à deux disjoints*, alors*

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

Le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ est alors appelé un espace de probabilité.

♠ **Attention !** Une probabilité est une *application*, qui "mange" un *événement* A qui donne en résultat un *nombre* $\mathbb{P}(A)$ dans $[0, 1]$. Le nombre $\mathbb{P}(A)$ est appelé la probabilité de A.

Remarque. L'équiprobabilité sur un ensemble fini est un cas particulier de probabilité.

Remarque. Pourquoi a-t-on choisi ces deux propriétés pour la définition d'une probabilité ? Tout d'abord, remarquons qu'elles sont assez intuitives :

– $\mathbb{P}(\Omega) = 1$ signifie intuitivement que l'ensemble Ω décrit bien tous les résultats possibles de l'expérience aléatoire, autrement dit, « il est impossible de tomber à l'extérieur de l'univers Ω ».
– Quant à la deuxième propriété, pensons au cas de deux événements A et B disjoints : il semble naturel que la probabilité ait une propriété du type $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
– Ces deux propriétés sont à la fois suffisantes pour pouvoir faire des calculs, et suffisamment peu contraignantes pour être satisfaites dans des cas très différents.

Pour faire des calculs de probabilités dévénements, on va avoir besoin de règles :

Proposition 2.3 *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité.*
i) $\mathbb{P}(\varnothing) = 0$.
ii) Si A et B sont deux événements disjoints (c'est-à-dire $A \cap B = \varnothing$), alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
iii) Si A_1, A_2, ..., A_n sont des événements deux à deux disjoints (pour tous $i \neq j$, on a $A_i \cap A_j = \varnothing$) alors

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

iv) Si $A \in \mathcal{F}$, alors $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$.
v) Si A et B sont deux éléments de $\mathcal{F}$ tels que $A \subset B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$.
vi) (Formule d'inclusion-exclusion) Si A et B sont deux éléments de $\mathcal{F}$, alors

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Démonstration. vi) On écrit $A \cup B$ comme la réunion disjointe de $A \cap \bar{B}$, $A \cap B$ et $B \cap \bar{A}$ (vérifier et faire un dessin), et on remarque que A est la réunion disjointe de $A \cap \bar{B}$ et $A \cap B$, tandis que B est la réunion disjointe de $A \cap B$ et $B \cap \bar{A}$. On obtient donc

$$\begin{aligned}
\mathbb{P}(A \cup B) &= \mathbb{P}(A \cap \bar{B}) + \mathbb{P}(A \cap B) + \mathbb{P}(B \cap \bar{A}) \\
&= \big(\mathbb{P}(A \cap \bar{B}) + \mathbb{P}(A \cap B)\big) + \big(\mathbb{P}(A \cap B) + \mathbb{P}(B \cap \bar{A})\big) - \mathbb{P}(A \cap B) \\
&= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B),
\end{aligned}$$

ce qui est l'égalité recherchée. □

♣ **Exercice 2.2.** On suppose qu'on tire un nombre au hasard dans l'ensemble $\{1, 2, \ldots, 20\}$, pas nécessairement uniformément. On note X le résultat de cette expérience aléatoire.

1. On suppose que la probabilité d'obtenir un nombre pair est $0,4$. Quelle est la probabilité d'obtenir un nombre impair ?
2. On note A l'événement "X est pair" et B l'événement "X est impair et plus petit que 10". On suppose $\mathbb{P}(A) = 0,6$ et $\mathbb{P}(B) = 0,25$. Que vaut $\mathbb{P}(A \cup B)$?
3. On note maintenant respectivement A, B, C les événements "X est divisible par 2, 3, 6". On suppose $\mathbb{P}(A) = 0,6$, $\mathbb{P}(B) = 0,3$ et $\mathbb{P}(C) = 0,2$. Que vaut $\mathbb{P}(A \cup B)$?

Une des idées de base du calcul des probabilités est de découper les événements en événements plus petits dont on connaît la probabilité. Ceci est formalisé par la *formule des probabilités totales*, qui généralise le principe de partition :

Définition 2.4 *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité. Un* système complet d'événements $(A_i)_{i \in I}$ *est une famille* finie *d'éléments de* $\mathcal{F}$ *telle que*
– $\displaystyle\bigcup_{i \in I} A_i = \Omega$,
– *si $i, j \in I$ et $i \neq j$, alors $A_i \cap A_j = \varnothing$.*

Proposition 2.5 *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité et soit* $(A_i)_{i\in I}$ *un système complet d'événements. Alors pour tout événement* $B \in \mathcal{F}$*, on a la* formule des probabilités totales *:*

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

Démonstration. Comme $\bigcup_{i\in I} A_i = \Omega$,

$$B = B \cap \Omega = B \cap \bigcup_{i\in I} A_i = \bigcup_{i\in I} (B \cap A_i).$$

Comme les $(A_i)_{i\in I}$ sont deux à deux disjoints, les $(B \cap A_i)_{i\in I}$ sont a fortiori deux à deux disjoints, donc

$$\mathbb{P}(B) = \mathbb{P}\left(\bigsqcup_{i\in I} (B \cap A_i)\right) = \sum_{i\in I} \mathbb{P}(B \cap A_i).$$

□

La proposition suivante regroupe des propriétés plus avancées, qu'on peut omettre en première lecture :

Proposition 2.6 *i) Si les* $(A_i)_{i\in\mathbb{N}}$ *sont des éléments de* $\mathcal{F}$*, alors*

$$\mathbb{P}\left(\bigcup_{i\in\mathbb{N}} A_i\right) \leq \sum_{i\in\mathbb{N}} \mathbb{P}(A_i).$$

ii) (Continuité monotone croissante) Si les $(A_i)_{i\in\mathbb{N}}$ *forment une suite croissante d'éléments de* $\mathcal{F}$*, c'est-à-dire s'ils vérifient* $\forall i \in \mathbb{N},\ A_i \subset A_{i+1}$*, alors*

$$\mathbb{P}\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \lim_{i\to+\infty} \mathbb{P}(A_i).$$

iii) (Continuité monotone décroissante) Si les $(A_i)_{i\in\mathbb{N}}$ *forment une suite décroissante d'éléments de* $\mathcal{F}$*, c'est-à-dire s'ils vérifient* $\forall i \in \mathbb{N}, A_{i+1} \subset A_i$*, alors*

$$\mathbb{P}\left(\bigcap_{i\in\mathbb{N}} A_i\right) = \lim_{i\to+\infty} \mathbb{P}(A_i).$$

2.3 Probabilité sur un ensemble discret

2.3.1 Caractérisation

Définition 2.7 • *Un ensemble* E *non vide est dit fini si on peut le mettre en bijection avec un ensemble de la forme* $\{1, 2, \ldots, n\}$*. Le nombre* n *est alors appelé le* cardinal *de* E *: c'est le nombre d'éléments de* E*. Si* E *n'est pas fini, on dit qu'il est... infini !*
• *Un ensemble* E *est dit dénombrable si on peut le mettre en bijection avec* $\mathbb{N}$*.*
• *Un ensemble* E *est dit discret s'il est fini ou dénombrable.*

▷ **Exemple.** On pourrait avoir envie de dire que $\mathbb{N}^*$ "est plus petit" que $\mathbb{N}$, et cependant, l'application

$$\begin{array}{rcl} \varphi : \mathbb{N} & \to & \mathbb{N}^* \\ x & \mapsto & x+1 \end{array}$$

est une bijection. De ce point de vue, $\mathbb{N}$ et $\mathbb{N}^*$ ont la même taille.

▷ **Exemple.** $\mathbb{N}$, $2\mathbb{N}$, $\mathbb{Z}$, $\mathbb{N}^d$, $\mathbb{Z}^d$, $\mathbb{Q}$ sont des ensembles dénombrables.

Proposition 2.8 *Soit Ω un ensemble fini ou dénombrable, muni de la tribu $\mathcal{P}(\Omega)$. Une probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{P}(\Omega))$ est entièrement déterminée par une famille $(p_\omega)_{\omega\in\Omega}$ telle que*

1. $\forall \omega \in \Omega, \quad p_\omega \geq 0$,
2. $\sum_{\omega\in\Omega} p_\omega = 1$;

en posant, pour tout $A \in \mathcal{P}(\Omega)$,

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

Cela montre que pour définir une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ lorsque Ω est fini, il suffit de donner la famille des valeurs $\mathbb{P}(\{\omega\})$, $\omega \in \Omega$.

♣ **Exercice 2.3.** Un dé non équilibré. On résume les probabilités des événements élémentaires dans le tableau suivant :

ω	1	2	3	4	5	6
p_i	1/3	1/6	1/12	1/12	1/4	p_6

Déterminer p_6 pour que les p_ω définissent une probabilité. Calculer la probabilité que le résultat du dé soit pair.

♣♣ **Exercice 2.4.** On lance deux dés équilibrés, un bleu et un rouge. Donner un espace de probabilité pour décrire le résultat de cette expérience aléatoire.
Maintenant, je lance deux dés équilibrés, un bleu et un rouge et je relève la sommes des deux dés. Donner un espace de probabilité pour décrire le résultat de cette expérience aléatoire.

▷ **Exemple.** J'ai un composant qui, à chaque utilisation, a une probabilité $p \in]0, 1[$ de tomber en panne, et je voudrais construire un espace de probabilité pour modéliser son nombre de fonctionnements corrects. A priori, je ne peux pas borner ce nombre, et je dois donc considérer l'univers $\Omega = \mathbb{N}^* = \{1, 2, \ldots\}$. D'après la proposition, une probabilité sur $\mathbb{N}^*$ est caractérisée par une famille $(p_n)_{n\in\mathbb{N}^*}$ telle que

1. $\forall n \in \mathbb{N}^*, \quad p_n \geq 0$,
2. $\sum_{n\in\mathbb{N}} p_n = 1$.

C'est donc une série à termes positifs, de somme 1.
Ici, on a

$$p_1 = \mathbb{P}(\{1\}) = p,\ p_2 = \mathbb{P}(\{2\}) = (1-p)p,\ \ldots,\ p_k = \mathbb{P}(\{k\}) = (1-p)^{k-1}p,\ \ldots$$

Cette suite vérifie-elle des deux propriétés ?

♣♣♣ **Exercice 2.5.** Peut-on définir une probabilité uniforme sur l'ensemble $\mathbb{N}$ des entiers naturels ?

2.3.2 Rappels sur les séries à termes positifs

On sait très bien ce qu'est la somme d'un nombre fini de termes. Les séries tentent de donner un sens à des sommes avec un nombre *infini* de termes.

Définition 2.9 *Soit $(u_n)_{n\in\mathbb{N}}$ une suite de nombres réels positifs.*
- *la somme finie $S_n = \sum_{k=0}^{n} u_k$ est appelée la* somme partielle *de rang n (c'est donc une somme "normale") ;*
- *on dit que la série $\sum_{n\in N} u_n$* converge *si et seulement si la suite des sommes partielles $(S_n)_{n\in N}$ converge. Dans ce cas, $S = \lim_{n\to+\infty} S_n$ est appelée* la somme *de la série. On note*

$$\sum_{n=0}^{+\infty} u_n = S = \lim_{n\to+\infty} S_n.$$

▷ **Exemple.** Soit $p \in]0,1[$ fixé. Pour tout $k \in \mathbb{N}^*$, on pose $p_k = p\,(1-p)^{k-1}$. Vérifions que la famille de réels $(p_k)_{k\in\mathbb{N}^*}$ définit une probabilité sur $\mathbb{N}^*$.

1. $\forall k \in \mathbb{N}^*, \quad p_k \geq 0.$
2. Rappel sur la série géométrique :

Proposition 2.10 *(Série géométrique) Soit $s \in \mathbb{R}$ et $s \neq 1$. Alors*

$$\sum_{k=0}^{n} s^k = \frac{s^{n+1}-1}{s-1}.$$

La série de terme général $(s^n)_{n\in\mathbb{N}}$ converge si et seulement si $|s|<1$ et dans ce cas $\sum_{k=0}^{+\infty} s^k = \frac{1}{1-s}$.

Reprenons notre calcul. Comme $0 < 1-p < 1$,

$$\sum_{k=1}^{+\infty} p_k = \sum_{k=1}^{+\infty} p\,(1-p)^{k-1} = p\sum_{l=0}^{+\infty}(1-p)^l = \frac{p}{1-(1-p)} = 1.$$

Donc la famille $(p_k)_{k\in\mathbb{N}^*}$ définit bien une probabilité sur $\mathbb{N}^*$. On peut alors, par exemple,

– calculer la probabilité que le composant fonctionne moins de 10 fois :

$$\mathbb{P}(\{1,2,\ldots,10\}) = \sum_{k=1}^{10} p_k = \sum_{k=1}^{10} p\,(1-p)^{k-1} = p\sum_{k=0}^{9} p^k = p\frac{1-(1-p)^{10}}{1-(1-p)} = 1-(1-p)^{10}.$$

– calculer la probabilité que le composant fonctionne au moins de 11 fois :

$$\mathbb{P}(\{11,12,\ldots\}) = \mathbb{P}(\{1,2,\ldots,10\}^c) = 1-\mathbb{P}(\{1,2,\ldots,10\}) = (1-p)^{10}.$$

– calculer la probabilité de l'ensemble des nombres pairs non nuls, noté $2\mathbb{N}^*$:

$$\mathbb{P}(2\mathbb{N}^*) = \sum_{k=1}^{+\infty} p_{2k} = \sum_{k=1}^{+\infty} p\,(1-p)^{2k-1} = p\,(1-p)\sum_{k=1}^{+\infty}[(1-p)^2]^{k-1} = p\,(1-p)\sum_{l=0}^{+\infty}[(1-p)^2]^l = \frac{p\,(1-p)}{1-(1-p)^2} = \frac{1-p}{2-p}.$$

– calculer la probabilité de l'ensemble des nombres impairs, noté $1+2\mathbb{N}$:

$$\mathbb{P}(1+2\mathbb{N}) = \mathbb{P}((2\mathbb{N}^*)^c) = 1-\mathbb{P}(2\mathbb{N}^*) = 1-\frac{1-p}{2-p} = \frac{1}{2-p}.$$

Proposition 2.11 *(Critères de convergence des séries à termes positifs) Soit $(u_n)_{n\in\mathbb{N}}$ une suite de nombres réels* positifs.

1. *Si la suite des sommes partielles $(S_n)_{n\in\mathbb{N}}$ est majorée, alors la série $\sum_{n\in\mathbb{N}} u_n$ converge.*
2. *S'il existe une suite $(v_n)_{n\in\mathbb{N}}$ de nombres réels positifs telle que*

$$\forall n \in \mathbb{N},\ u_n \leq v_n \quad \text{et} \quad \sum_{n\in\mathbb{N}} v_n \text{ converge},$$

   alors la série $\sum_{n\in\mathbb{N}} u_n$ converge.
3. *(Critère de d'Alembert) Si $\frac{u_{n+1}}{u_n}$ converge vers l et que $l<1$, alors la série $\sum_{n\in\mathbb{N}} u_n$ converge.*

♣ **Exercice 2.6.** On prend $\Omega = \mathbb{N}$ et pour tribu l'ensemble $\mathcal{P}(\Omega)$ de toutes les parties de Ω. On fixe $\lambda > 0$ et $C > 0$ et on pose, pour $k \in \mathbb{N}$,

$$\mathbb{P}(\{k\}) = C\frac{\lambda^k}{k!}.$$

Déterminer C pour que $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ soit un espace de probabilité.

2.3.3 Un exemple d'ensemble infini non dénombrable

Proposition 2.12 $\{0,1\}^{\mathbb{N}}$ *n'est pas dénombrable.*

Démonstration. On utilise le procédé de la diagonale de Cantor. Raisonnons par l'absurde et supposons que $\{0,1\}^{\mathbb{N}}$ soit dénombrable. On peut donc numéroter ses éléments $\{0,1\}^{\mathbb{N}} = \{x_0, x_1, \dots\}$. Un point de $\{0,1\}^{\mathbb{N}}$ est une suite de 0 et de 1 : on note ainsi, pour tout $i \in \mathbb{N}$,

$$x_i = (x_{i,0}, x_{i,1}, \dots) = (x_{i,j})_{j\in\mathbb{N}}.$$

On va maintenant construire un élément $y = (y_j)_{j\in\mathbb{N}}$ de $\{0,1\}^{\mathbb{N}}$ différent de tous les x_i : on pose

$$\forall j \in \mathbb{N},\ y_j = \begin{cases} 0 \text{ si } x_{j,j} = 1 \\ 1 \text{ si } x_{j,j} = 0 \end{cases}$$

Pour un entier naturel i quelconque, par construction, $y_i \neq x_{i,i}$, donc $y \neq x_i$; ainsi, pour tout entier $i \geq 0$, $y \neq x_i$ et ainsi $y \notin \{x_0, x_1, \dots\}$; cependant, $y \in \{0,1\}^{\mathbb{N}}$. Ceci contredit donc $\{0,1\}^{\mathbb{N}} = \{x_0, x_1, \dots\}$. □

Remarque. Si à un nombre $x \in [0,1]$, on associe la suite des chiffres de son développement en base 2, on construit une injection de $[0,1[$ dans $\{0,1\}^{\mathbb{N}}$: donc $[0,1]$ n'est pas dénombrable, donc $\mathbb{R}$ non plus...

2.4 Probabilité à densité sur $\mathbb{R}$

Tous les phénomènes aléatoires ne sont pas discrets : par exemple, je peux souhaiter modéliser la durée de vie d'un composant (par un nombre de $\mathbb{R}_+$), ou le générateur de nombre aléatoire de mon ordination (par un nombre dans $[0,1]$)...

Définition 2.13 • *Soit $f : \mathbb{R} \to \mathbb{R}$ une application. On dit que f est une densité sur $\mathbb{R}$ si et seulement si elle satisfait les conditions suivantes :*

1. *f est continue par morceaux,*
2. *pour tout $x \in \mathbb{R}$, $f(x) \geq 0$,*
3. *L'intégrale, éventuellement impropre, de f sur $\mathbb{R}$ existe et $\int_{\mathbb{R}} f(x)dx = 1$.*

• *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. On dit que $\mathbb{P}$ est* de densité *f si pour tous $a, b \in \mathbb{R}$ avec $a < b$,*

$$\mathbb{P}([a,b]) = \int_a^b f(x)dx.$$

▷ **Exemple.** Pour modéliser modéliser le générateur de nombres aléatoire de la calculatrice ou du tableur, on aimerait comprendre ce que signifie "tirer un nombre au hasard dans $[0,1]$ de façon uniforme" : on prend pour tout $x \in \mathbb{R}$, $f(x) = \mathbf{1}_{0\leq x\leq 1}$. Alors f est une densité sur $\mathbb{R}$. En effet,

1. f est continue sur les trois morceaux $]-\infty, 0[$, $]0,1[$ et $]1, +\infty[$.
2. pour tout $x \in \mathbb{R}$, $f(x) \geq 0$,
3. Etudions l'intégrale de f sur $\mathbb{R}$:

$$\int_0^{+\infty} f(x)dx = \int_0^1 dx = 1.$$

Soit $\mathbb{P}$ la probabilité sur $\mathbb{R}$ de densité f. On peut alors calculer la probabilité d'un intervalle $[a,b]$, avec $0 \leq a < b \leq 1$:

$$\mathbb{P}([a,b]) = \int_a^b f(x)dx = int_a^b dx = (b-a).$$

C'est donc bien une loi uniforme, au sens où la probabilité de tomber dans un intervalle dépend de sa longueur, mais pas de sa position.

Proposition 2.14 *Sur* $\Omega = [0,1]$ *muni de la tribu* $\mathcal{B}([0,1])$*, on considère la probabilité* $\mathbb{P}$ *uniforme sur* $[0,1]$.
i) Soit $a, b \in [0,1]$ *tels que* $a < b$ *:*

$$\mathbb{P}([a,b]) = \mathbb{P}(]a,b[) = \mathbb{P}(]a,b]) = \mathbb{P}([a,b[) = b - a.$$

ii) Pour tout $a \in [0,1]$, $\mathbb{P}(\{a\}) = 0$.

Démonstration. i) Soit $a, b \in [0,1]$ tels que $a < b$. Pour tout n (suffisamment grand)

$$[a - \frac{1}{n}, b + \frac{1}{n}] \subset]a,b[\subset [a,b], \text{ donc } \mathbb{P}([a - \frac{1}{n}, b + \frac{1}{n}]) \leq \mathbb{P}(]a,b[) \leq \mathbb{P}([a,b]).$$

Or $\mathbb{P}([a,b]) = b - a$ et $\mathbb{P}\left([a + \frac{1}{n}, b - \frac{1}{n}]\right) = b - a - \frac{2}{n}$, donc pour tout n (suffisamment grand),

$$b - a - \frac{2}{n} \leq \mathbb{P}(]a,b[) \leq b - a,$$

et on obtient le résultat en faisant tendre n vers $+\infty$. □

♣♣♣ **Exercice 2.7.** On reprend le même cadre.

1. Pourquoi ne peut-on pas écrire $\mathbb{P}([0,1]) = \sum_{a\in[0,1]} \mathbb{P}(\{a\})$?
2. Quelle est la probabilité de $[0,1] \cap \mathbb{Q}$?

♣♣ **Exercice 2.8.** Comment faire pour définir une probabilité uniforme sur $[2,5]$?

▷ **Exemple.** On prend, pour tout $x \in \mathbb{R}$, $f(x) = \frac{1}{2}\exp(-2x)\mathbf{1}_{x\geq 0}$. Alors f est une densité sur $\mathbb{R}$. En effet

1. f est continue sur les deux morceaux $]-\infty, 0[$ et $]0, +\infty[$.
2. pour tout $x \in \mathbb{R}$, $f(x) \geq 0$,
3. Etudions l'intégrale de f sur $\mathbb{R}$: c'est une intégrale impropre : $\int_R f = \int_0^{+\infty} \frac{1}{2}\exp(-2x)dx$. On commence par étudier l'intégrale normale sur $[0, M]$:

   $$\int_0^M \frac{1}{2}\exp(-2x)dx = [-\exp(-2x)]_0^M = 1 - \exp(-2M).$$

   Maintenant, on fait tendre M vers $+\infty$:

   $$\lim_{M\to+\infty} \int_0^M \frac{1}{2}\exp(-2x)dx = \lim_{M\to+\infty}(1 - \exp(-2M)) = 1,$$

   donc l'intégrale impropre de f sur $\mathbb{R}$ existe et $\int_{\mathbb{R}} f(x)dx = 1$.

Soit $\mathbb{P}$ la probabilité sur $\mathbb{R}$ de densité f. On peut alors calculer la probabilité d'un intervalle $[a,b]$, avec $0 < a < b < +\infty$:

$$\mathbb{P}([a,b]) = \int_a^b f(x)dx = int_a^b \frac{1}{2}\exp(-2x)dx = [-\exp(-2x)]_a^b = \exp(-2a) - \exp(-2b).$$

♣ **Exercice 2.9.** Soit $\lambda > 0$. On pose pour tout $x \in \mathbb{R}$, $f_\lambda(x) = \frac{1}{\lambda}\exp(-\lambda x)\mathbf{1}_{x\geq 0}$.

1. Montrer que f_λ est une densité.
2. Tracer l'allure de f_λ pour différentes valeurs de λ.
3. On note $\mathbb{P}_\lambda$ la probabilité sur $\mathbb{R}$ de densité f_λ. Calculer, pour tout $t > 0$, $\mathbb{P}_\lambda([0,t])$.

♣ **Exercice 2.10.** Soit $C > 0$. On pose, pour tout $x \in \mathbb{R}$, $f(x) = \frac{C}{1+x^2}$.
Déterminer C pour que f soit une densité sur $\mathbb{R}$.

Chapitre 3

Indépendance d'événements et probabilité conditionnelle

Sommaire

3.1 Indépendance d'événements . . . 21
3.1.1 Indépendance de deux événements . . . 21
3.1.2 Indépendance d'une famille d'événements . . . 21
3.2 Application : le jeu de pile ou face, ou schéma de Bernoulli fini . . . 22
3.2.1 Probabilité d'un événement élémentaire. . . . 22
3.2.2 Probabilité d'obtenir k piles . . . 23
3.3 Probabilité conditionnelle . . . 23
3.3.1 Définition . . . 24
3.3.2 Propriétés . . . 25
3.4 Complément . . . 26
3.4.1 Le jeu infini de pile ou face, ou schéma de Bernoulli . . . 26
3.4.2 Notion de presque sûr . . . 27

Objectifs :
Formaliser la notion d'indépendance entre deux événements, et introduire la notion de probabilité conditionnelle.

Mots-clés :
• Indépendance, probabilité conditionnelle.
• Schéma de Bernoulli.

Outils :
• Coefficients du binôme, formule du triangle de Pascal, formule du binôme de Newton, nombre de parties d'un ensembles.
• Formule des probabilités totales, formule de Bayes.

3.1 Indépendance d'événements

3.1.1 Indépendance de deux événements

▷ **Exemple.** Je tire une carte au hasard dans un jeu de 32 cartes. Je considère les 3 événements suivants : A : « la carte tirée est rouge », B : « la carte tirée est un cœur » et C : « la carte tirée est un roi ». Savoir que C est réalisé ne me donne a priori aucune indication quant au fait que cette carte soit un cœur. En revanche, si je sais que A est réalisé, je sais que j'ai « plus de chance » d'avoir tiré un cœur.

Nous allons formaliser ces réponses intuitives.

Définition 3.1 *Soient A et B deux événements d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que A et B sont indépendants si*

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

▷ **Exemple.** Reprenons notre exemple du jeu de 32 cartes.

$$\begin{aligned} \mathbb{P}(B \cap C) &= \mathbb{P}(\text{« la carte tirée est le roi de cœur »}) = 1/32 \\ \mathbb{P}(B)\mathbb{P}(C) &= (8/32) \times (4/32) = 1/32. \end{aligned}$$

Donc B et C sont indépendants.

$$\begin{aligned} \mathbb{P}(A \cap B) &= \mathbb{P}(\text{« la carte tirée est un cœur »}) = 8/32 = 1/4 \\ \mathbb{P}(A)\,\mathbb{P}(B) &= (16/32) \times (8/32) = 1/8. \end{aligned}$$

Donc A et B ne sont pas indépendants.

♠ **Attention !** Ne pas confondre des événements indépendants et des événements disjoints !
• si A et B sont indépendants, alors $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$;
• si A et B sont disjoints, alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Proposition 3.2 *Soient A et B deux événements d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.*
A et B sont indépendants ssi A et B^c sont indépendants A^c et B^c sont indépendants.

Démonstration. a faire □

3.1.2 Indépendance d'une famille d'événements

La notion d'indépendance s'étend aux familles de plusieurs événements de la manière suivante :

Définition 3.3 *On dit que les événements A_1, A_2,..., A_n sont* indépendants *si pour tout sous-ensemble non vide $\{j_1, \ldots, j_p\}$ de $\{1, \ldots, n\}$, on a*

$$\mathbb{P}\big(A_{j_1} \cap A_{j_2} \cap \ldots \cap A_{j_p}\big) = \mathbb{P}\big(A_{j_1}\big)\,\mathbb{P}\big(A_{j_2}\big) \ldots \mathbb{P}\big(A_{j_p}\big).$$

♣ **Exercice 3.1.** Un circuit électrique est formé de 3 composants, qui ont respectivement des probabilités d'être en panne p_1, p_2 et p_3. On suppose que les pannes des différents composants sont indépendantes. Calculer la probabilité que le circuit soit en panne dans chacun des cas suivants :

1. les 3 composants sont montés en série,
2. les trois composants sont montés en parallèle,
3. deux sont en parallèle, et le troisième en série avec ce groupe de deux.

♣ **Exercice 3.2.** Considérons l'espace correspondant à deux lancers de pile ou face (pièce non truquée) et prenons

A : « pile au premier lancer »

B : « pile au second lancer »

C : « même résultat aux deux lancers »

Montrez que les événements A, B, C sont indépendants deux à deux mais non indépendants.

La notion d'indépendance s'étend aux familles infinies de la façon suivante :

Définition 3.4 *(peut être omis en première lecture) Soit I un ensemble quelconque d'indices et $(A_i)_{i\in I}$ une famille d'événements. On dit que les A_i, $i \in I$ sont indépendants si toute sous-famille finie d'événements de la famille $(A_i)_{i\in I}$ est constituée d'événements indépendants.*

3.2 Application : le jeu de pile ou face, ou schéma de Bernoulli fini

On souhaite modéliser l'expérience aléatoire consistant à lancer n avec fois une pièce et à noter le nombre de pile obtenu ($n \geq 1$ est un entier nature non nul). On suppose que cette pièce tombe sur pile (codé par 1) avec probabilité $p \in]0,1[$ et sur face (codé par 0) avec probabilité $1-p$.
L'univers naturel pour cette expérience est donc l'ensemble

$$\Omega = \{0,1\}^n$$

des n-uplets à valeurs dans $\{0,1\}$. On prend comme tribu l'ensemble des parties de Ω. Par exemple, si $n = 4$, l'issue $\omega = (1,0,0,1) \in \Omega$ code l'événement "j'ai obtenu pile au premier tirage, puis face au second, puis face au troisième, puis pile au dernier tirage". Nous aimerions définir une probabilité sur Ω modélisant n lancers successifs de la pièce.

3.2.1 Probabilité d'un événement élémentaire.

Notons, pour tout $k \in \{1,\ldots,n\}$, A_k l'événement "obtenir 1 au k-ème lancer". Alors, pour $k \in \{1,\ldots,n\}$, $\mathbb{P}(A_k) = p$. De plus, il est naturel de supposer que les $(A_k)_{1\leq k\leq n}$ sont indépendants.
Ceci nous permet de déterminer la probabilité de chaque événement élémentaire : soit $\omega = (\omega_i)_{1\leq i\leq n} \in \Omega$ fixé. Cet ω code donc le résultat de la suite des n lancers de pile ou face. On note $I(\omega)$ l'ensemble des numéros des lancers ayant donné pile pour ce résultat ω :

$$I(\omega) = \{i \in \{1,\ldots,n\} :\ \omega_i = 1\} \quad \text{et} \quad \sum_{i=1}^n \omega_i = \mathrm{Card}(I(w)).$$

Alors $\{\omega\} = \left(\bigcap_{i\in I(w)} A_i\right) \cap \left(\bigcap_{j\notin I(w)} \overline{A_j}\right)$; comme les événements $(A_k)_{1\leq k\leq n}$ sont indépendants,

$$\mathbb{P}(\{\omega\}) = \mathbb{P}\left(\left(\bigcap_{i\in I(w)} A_i\right) \cap \left(\bigcap_{j\notin I(w)} \overline{A_j}\right)\right) = \left(\prod_{i\in I(w)} \mathbb{P}(A_i)\right)\left(\prod_{j\notin I(w)} \mathbb{P}(\overline{A_j})\right) = p^{\mathrm{Card}(I(w))}(1-p)^{n-\mathrm{Card}(I(w))}.$$

On remarque en particulier que $\mathbb{P}(\{\omega\})$ ne dépend que du nombre de 1 dans ω (c'est-à-dire du nombre de pile obtenus en faisant les n lancers), et non pas des positions de ces 1 (c'est-à-dire des numéros des lancers ayant donné pile).

Schéma de Bernoulli fini. Nous venons de définir un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ modélisant le résultat de n lancers indépendants d'une pièce tombant sur pile avec probabilité p :

$$\Omega = \{0,1\}^n,\ \mathcal{F} = \mathcal{P}(\Omega) \quad \text{et} \quad \forall \omega \in \Omega,\ \mathbb{P}(\omega) = p^{\sum_{i=1}^n \omega_i}(1-p)^{n-\sum_{i=1}^n \omega_i}.$$

Comme Ω est fini, connaître $\mathbb{P}$ sur chacun des événements élémentaires suffit à déterminer $\mathbb{P}$.

3.2.2 Probabilité d'obtenir k piles

Soit $k \in \{0, \dots, n\}$. Notons B_k l'événement "obtenir exactement k fois pile". Quelle est la probabilité de B_k ?

$$\mathbb{P}(B_k) = \sum_{\omega \in B_k} \mathbb{P}(\omega).$$

Or $\omega \in B_k \Leftrightarrow \sum_{i=1}^n \omega_i = k$, donc tous les événements élémentaires $\omega \in B_k$ ont la même probabilité $\mathbb{P}(\omega) = p^k(1-p)^{n-k}$. Ainsi,

$$\mathbb{P}(B_k) = \sum_{\omega \in B_k} \mathbb{P}(\omega) = \sum_{\omega \in B_k} p^k(1-p)^{n-k} = p^k(1-p)^{n-k}\mathrm{Card}(B_k).$$

Il nous reste donc à déterminer le cardinal de B_k, ou autrement dit, à compter combien il existe d'éléments de $\{0,1\}^n$ écrits avec exactement k fois le chiffre 1 et $n-k$ fois le chiffre 0. On voit qu'il suffit de déterminer la place des k chiffres 1 parmi les n places possibles, et de compléter ensuite avec des 0 : on doit donc compter le nombre de parties à k éléments d'un ensemble à n éléments, qui est $\binom{n}{k}$. Ainsi,

Schéma de Bernoulli fini. Quand on réalise n lancers indépendants d'une pièce tombant sur pile avec probabilité p, la probabilité d'obtenir exctement k piles est

$$\mathbb{P}(B_k) = \binom{n}{k} p^k(1-p)^{n-k}.$$

Remarque. Remarquons que les $(B_k)_{0 \le k \le n}$ forment un système complet. La formule du binôme assure que

$$\sum_{k=0}^n \mathbb{P}(B_k) = \sum_{k=0}^n \binom{n}{k} p^k(1-p)^{n-k} = (p + (1-p))^n = 1;$$

ceci est toujours vrai pour un système complet

♣ **Exercice 3.3.** Une machine fabrique des boulons, et la probabilité qu'un boulon soit défectueux est $0,02$. Je prélève 10 boulons. Quelle est la probabilité qu'au moins 8 de ces boulons soient fonctionnels ?

3.3 Probabilité conditionnelle

Sur une population de vingt-sept malades, on donne à dix malades un nouveau médicament et aux dix-sept autres un placebo. Parmi les dix malades ayant reçu le médicament, sept déclarent avoir senti une amélioration, et huit, parmi ceux ayant reçu le placebo, ressentent aussi un mieux.

Un univers adapté à cette expérience peut être le suivant : à chacun des vingt-sept malades, j'associe une lettre, M ou P, suivant s'il a reçu le médicament ou le placébo, et un signe, + ou − suivant s'il a ressenti une amélioration ou non. L'ensemble $\Omega = \{(M,+), (M,-), (P,+), (P,-)\}$ rassemble toutes les configurations possibles, et on va attribuer à chaque événement élémentaire de Ω une probabilité traduisant le résultat du test :

$$\mathbb{P}(\{(M,-)\}) = \frac{10-7}{27} = \frac{3}{27},\ \mathbb{P}(\{(M,+)\}) = \frac{7}{27},\ \mathbb{P}(\{(P,-)\}) = \frac{17-8}{27} = \frac{9}{27} \text{ et } \mathbb{P}(\{(P,+)\}) = \frac{8}{27}.$$

On peut par exemple calculer la proportion de personnes ayant constaté une amélioration, c'est-à-dire calculer la probabilité

$$\mathbb{P}(\{(M,+), (P,+)\}) = \frac{7+8}{27} = \frac{15}{27}.$$

Mais dans cette proportion, on ne prend pas en compte le fait que certains ont reçu un placebo et d'autres le médicament. On a plutôt envie de calculer, parmi les gens qui ont reçu le médicament, la proportion de ceux qui ont senti une amélioration, $7/10 = 0,7$, et de la comparer avec la proportion correspondante chez ceux ayant reçu le placebo, $8/17 \approx 0,47$. On a ainsi envie de conclure que le médicament est plus efficace que le placebo.

On a donc envie d'introduire une « condition » (ici le fait d'avoir pris le médicament) qui modifie notre application probabilité. Nous allons formaliser cette idée à l'aide de la notion de *probabilité conditionnelle*.

3.3.1 Définition

Définition 3.5 *Soient A et B deux événements d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ tels que $\mathbb{P}(B) > 0$. On appelle* probabilité conditionnelle *de A sachant B le nombre suivant, noté $\mathbb{P}_B(A)$ ou $\mathbb{P}(A \,|\, B)$:*

$$\mathbb{P}_B(A) = \mathbb{P}(A \,|\, B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

▷ **Exemple.** On pioche au hasard une carte dans un jeu non truqué de 32 cartes. On a alors par exemple

$$\mathbb{P}(\text{« cœur »} \,|\, \text{« dame »}) = \frac{\mathbb{P}(\text{« dame de cœur »})}{\mathbb{P}(\text{« dame »})} = \frac{1/32}{4/32} = \frac{1}{4}$$

et

$$\mathbb{P}(\text{« cœur »} \,|\, \text{« rouge »}) = \frac{\mathbb{P}(\text{« cœur rouge »})}{\mathbb{P}(\text{« rouge »})} = \frac{8/32}{16/32} = \frac{1}{2}.$$

Dans cet exemple, on a choisi de noté les événements avec du texte. Avec ces notations,

$$\text{« dame »} \cap \text{« cœur »} = \text{« dame de cœur »} \quad \text{et} \quad \text{« cœur »} \cap \text{« rouge »} = \text{« cœur rouge »}.$$

En fait, comme « cœur » $\subset$ « rouge », on a même « cœur rouge » = « cœur ».
Notez que cela correspond bien à notre intuition : si l'on sait qu'on a pioché une dame, la probabilité d'avoir pioché un cœur est 1/4 car l'information « on a pioché une dame » n'apporte rien. En revanche, si l'on sait qu'on a pioché une carte rouge, on a une chance sur deux que ce soit un cœur (car il y a autant de cœur que de carreaux).

Proposition 3.6 *Soit B un événement d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ tel que $\mathbb{P}(B) > 0$. Alors $(\Omega, \mathcal{F}, \mathbb{P}_B)$ est un espace de probabilité.*

On peut considérer les choses de la façon suivante : la probabilité $\mathbb{P}$ contient toute l'information disponible sur l'expérience aléatoire décrite. Une fois que B est réalisé, on a davantage d'information : il est donc naturel de changer de probabilité en prenant $\mathbb{P}_B$ afin de prendre en compte l'information supplémentaire.

Remarque. Les deux notations $\mathbb{P}_B(A)$ et $\mathbb{P}(A \,|\, B)$ ont toutes deux des avantages et des inconvénients. La seconde est plus simple à lire (la barre | se lisant « sachant », on lit « probabilité de A sachant B ») mais la première a l'avantage qu'on se souvient mieux que l'application $\mathbb{P}_B$ est une probabilité. Par exemple,

$$\mathbb{P}_B(\bar{A}) = 1 - \mathbb{P}_B(A).$$

♠ **Attention !** Même si l'on écrit $\mathbb{P}(A \,|\, B)$, ce qui est entre parenthèses, $A \,|\, B$, n'est ni un événement ni rien du tout, ce n'est qu'une écriture qui ne veut rien dire toute seule. C'est un peu comme si l'on écrivait ${}_B(A)$, cela n'a aucun sens tout seul !

On peut caractériser l'indépendance de deux événements à l'aide des probabilités conditionnelles.

Proposition 3.7 *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Soient A et B deux événements tels que $\mathbb{P}(B) > 0$. Alors*

$$A \text{ et } B \text{ sont indépendants} \Leftrightarrow \mathbb{P}(A \,|\, B) = \mathbb{P}(A).$$

Démonstration. En effet, comme $\mathbb{P}(B) > 0$, on peut diviser :

$$A \text{ et } B \text{ sont indépendants} \Leftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B) \Leftrightarrow \mathbb{P}(A \,|\, B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A).$$ □

3.3.2 Propriétés

Proposition 3.8 *(Formule de Bayes)*
Soit $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité. Soient* A *et* B *deux événements tels que* $\mathbb{P}(A) > 0$ *et* $\mathbb{P}(B) > 0$*. Alors*

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\, \mathbb{P}(A)}{\mathbb{P}(B)}.$$

♣ **Exercice 3.4.** Écrire une démonstration.
Cette proposition sert à « retourner » le conditionnement : on l'utilise pour calculer $\mathbb{P}(A \mid B)$ lorsqu'on connaît $\mathbb{P}(B \mid A)$. Par exemple on connaît la probabilité d'avoir ressenti une amélioration sachant qu'on est malade et on cherche la probabilité d'être malade sachant qu'on a ressenti une amélioration.

Proposition 3.9 *(Formule des probabilités totales)*
Soit $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité et soit* $(A_i)_{i\in I}$ *un système complet d'événements tel que* $\forall i \in I$, $\mathbb{P}(A_i) > 0$*. Alors pour tout événement* $B \in \mathcal{F}$*, on a*

$$\mathbb{P}(B) = \sum_{i\in I} \mathbb{P}(B \mid A_i)\, \mathbb{P}(A_i).$$

♣ **Exercice 3.5.** Écrire une démonstration. Préciser quelles égalités nécessitent que la famille d'événements soit finie ou dénombrable.
On peut combiner les deux dernières propositions pour obtenir l'énoncé suivant :

Proposition 3.10 *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité,* $(A_i)_{i\in I}$ *un système complet d'événements tel que* $\forall i \in I$, $\mathbb{P}(A_i) > 0$ *et* B *un événement de probabilité non nulle. Alors pour tout* $j \in I$*, on a*

$$\mathbb{P}(A_j \mid B) = \frac{\mathbb{P}(B \mid A_j)\, \mathbb{P}(A_j)}{\sum_{i\in I} \mathbb{P}(B \mid A_i)\, \mathbb{P}(A_i)}.$$

Un cas très utile est donné par le système complet $(A, \bar{A})$. Si $0 < \mathbb{P}(A) < 1$ et $\mathbb{P}(B) > 0$, alors

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\, \mathbb{P}(A)}{\mathbb{P}(B \mid A)\, \mathbb{P}(A) + \mathbb{P}(B \mid \bar{A})\, \mathbb{P}(\bar{A})}.$$

▷ **Exemple.** On m'a donné deux pièces, en me disant que l'une est équilibrée, l'autre est truquée et donne pile dans 75% des cas. Je choisis une des deux piéces au hasard, et en la lançant 6 fois, j'obtiens 4 fois pile. Quelle est la probabilité que la piéce que j'ai lancée soit la piéce truquée ?
Solution : Nous n'allons pas expliciter complètement l'espace de probabilité qui régit cette expérience, mais calculer seulement la probabilité demandée. Soit A l'événement « j'ai obtenu 4 fois pile et 2 fois face parmi les 6 lancers », T l'événement « j'ai lancé la pièce truquée » et donc $\bar{T}$ l'événement « j'ai lancé la pièce équilibrée ». Avec ces notations, on cherche donc $\mathbb{P}(T \mid R)$.
Traduisons dans un premier temps les données de l'énoncé. Puisqu'on choisit une des deux pièces au hasard, il est raisonnable de supposer que $\mathbb{P}(T) = \mathbb{P}(\bar{T}) = 1/2$. Avec la pièce équilibrée, on obtient pile et face à chaque lancer avec probabilité 1/2. Comme les différents lancers sont indépendants, on en déduit que

$$\mathbb{P}(R \mid \bar{T}) = \binom{6}{4} (1/2)^6 .$$

Un raisonnement similaire donne

$$\mathbb{P}(R \mid T) = \binom{6}{4} (3/4)^4 \, (1/4)^2 .$$

La formule précédente donne alors

$$\mathbb{P}(T \mid R) = \frac{\mathbb{P}(R \mid T)\, \mathbb{P}(T)}{\mathbb{P}(R \mid T)\, \mathbb{P}(T) + \mathbb{P}(R \mid \bar{T})\, \mathbb{P}(\bar{T})} = \frac{\binom{6}{4}(3/4)^4\,(1/4)^2\,(1/2)}{\binom{6}{4}(3/4)^4\,(1/4)^2\,(1/2) + \binom{6}{4}(1/2)^6(1/2)} = \frac{3^4}{2^6+3^4} \approx 0,56.$$

♣ **Exercice 3.6.** Deux usines A et B fabriquent des trottinettes. L'usine A fabrique deux fois plus de trottinettes que l'usine B. Les trottinettes provenant de l'usine A ont dans 5% des cas un défaut, alors que celles provenant de l'usine B ont dans 2% des cas un défaut. Je viens d'acheter une trottinette pour Théodule, laquelle se révéle être défectueuse. Quelle est la probabilité qu'elle provienne de l'usine A ?

♣♣ **Exercice 3.7.** Reprendre l'exercice avec le circuit électrique. Calculer, dans chacun des trois cas, la probabilité que le composant 1 soit en panne sachant que le circuit ne fonctionne pas.

3.4 Complément

3.4.1 Le jeu infini de pile ou face, ou schéma de Bernoulli

Nous avons déjà vu la modélisation de l'expérience consistant à lancer n fois une pièce. Mais cela peut être insuffisant. Par exemple, si l'on veut étudier l'instant où l'on obtient le premier « pile », on ne peut pas se restreindre à n lancers, il faut considérer une suite infinie de lancers. On doit donc modéliser une suite infinie de lancers indépendants, tombant avec probabilité $p \in]0,1[$ sur pile et avec probabilité $1-p$ sur face. L'univers correspondant à cette expérience est l'ensemble $\Omega = \{0,1\}^{\mathbb{N}}$ des suites à valeurs dans $\{0,1\}$. Nous avons vu que cet ensemble était infini non dénombrable.

Quelle est la probabilité d'un événement élémentaire ? Notons, pour tout $k \geq 1$, A_k l'événement "obtenir pile au k-ème lancer". Alors, pour $k \geq 1$, $\mathbb{P}(A_k) = p$. De plus, il est naturel de supposer que les $(A_k)_{k\in\mathbb{N}^*}$ sont indépendants.
Prenons par exemple la suite $\omega_0 = (P, F, P, F, \dots)$. Alors pour tout $n \geq 1$, $\{\omega\} \subset \{$j'ai obtenu n fois le motif "PF" à la suite$\}$.
Or

$$\begin{aligned}\{\text{j'ai obtenu } n \text{ fois le motif “PF” à la suite}\} &= \left(\bigcap_{0\leq k\leq n-1} A_{2k+1}\right) \cap \left(\bigcap_{1\leq k\leq n} \overline{A_{2k}}\right) \\ \mathbb{P}(\{\omega_0\}) &= \left(\prod_{0\leq k\leq n-1} \mathbb{P}(A_{2k+1})\right) \cap \left(\prod_{1\leq k\leq n} \mathbb{P}(\overline{A_{2k}})\right) \\ &= p^n(1-p)^n.\end{aligned}$$

Donc, en passant à la limite quand n tend vers $+\infty$, il vient $\mathbb{P}(\omega_0) = 0$.
On est donc, comme dans le cas des probabilités continues, dans un cas où on ne peut pas, en général, écrire une formule du type $\mathbb{P}(A) = \sum_{\omega\in A} \mathbb{P}(\{\omega\})$. Ceci est dû au fait que $\Omega = \{P, F\}^{\mathbb{N}}$ n'est pas dénombrable (tout comme $[0,1]$). Dans l'écriture $A = \bigsqcup_{a\in A}\{a\}$, l'union n'est pas dénombrable dès lors que A n'est pas dénombrable ; on ne peut donc pas appliquer l'axiome ii) de σ-additivité de la définition.
Contrairement au schéma de Bernoulli fini, ici, il ne suffit pas de connaître la probabilité des événements élémentaires pour décrire la probabilité.

Quelle est la probabilité de n'obtenir que des faces ? L'argument précédent montre que la probabilité d'obtenir la suite $(F, F, F, \dots)$ est nulle. Donc, avec probabilité 1, on finira par voir apparaître un pile.

Quelle est la probabilité que le premier pile sorte au k^e lancer ? Pour $k \geq 1$, on note B_k l'événement "pile sort pour la première fois au k-ème tirage". Alors, en utilisant l'indépendance des événements $(A_k)_{k\geq 1}$,

$$\forall k \geq 1 \quad B_k = \left(\bigcap_{1\leq i\leq k-1} \overline{A}_i\right) \cap A_k \quad \text{et donc} \quad \mathbb{P}(B_k) = \left(\prod_{1\leq i\leq k-1} \mathbb{P}(\overline{A}_i)\right) \mathbb{P}(A_k) = (1-p)^{k-1}p,$$

retenons ce résultat :

Schéma de Bernoulli infini. Quand on réalise une suite de lancers indépendants d'une pièce tombant sur pile avec probabilité p, la probabilité d'obtenir le premier pile au lancer numéro k est

$$\mathbb{P}(B_k) = (1-p)^{k-1}p.$$

♣♣ **Exercice 3.8.** (le singe dactylographe.) Mon singe Théodule s'amuse à taper au hasard des lettres sur mon clavier d'ordinateur (qui ne comporte que les 26 touches des 26 lettres de l'alphabet). On suppose qu'il n'a pas de préférence particulière pour certaines lettres, et que ses différents choix sont indépendants. Montrer que si je suis suffisamment patient, je verrai un jour Théodule taper le mot « abracadabra ».
Indication : « abracadabra » compte 11 lettres. Noter A_k l'événement : « abracadabra » apparaît entre les lettres $11k+1$ et $11k+11$ et imiter la preuve du jeu de pile ou face.

3.4.2 Notion de presque sûr

Nous avons vu dans l'exemple de la suite infinie de lancers de pile ou face que l'événement « il y a un pile dans la suite » est de probabilité 1, alors que cet événement n'est pas égal à l'ensemble Ω des tirages possibles (car le tirage $(F, F, F, \dots)$ n'est pas dans cet événement). Dans ce cas, on ne peut pas vraiment dire qu'il y aura *toujours* un pile dans la suite, on a besoin d'introduire la notion suivante.

Définition 3.11 *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité et soit* $A \in \mathcal{F}$.

1. *Si* $\mathbb{P}(A) = 1$, *on dit que* A *est un événement* presque sûr *ou que* A *se produit* presque sûrement.
2. *Si* $\mathbb{P}(A) = 0$, *on dit que* A *est un événement* négligeable.

Remarque. Ces notions dépendent de la probabilité $\mathbb{P}$: un événement peut être presque sûr pour une probabilité sans l'être pour une autre. Notez également que ce cas n'apparait naturellement que lorsque l'univers est non-dénombrable.

▷ **Exemple.**
1) Considérons une suite infini de lancers indépendants d'une pièce tombant sur pile avec probabilité $p \in]0,1[$. Nous avons vu que presque sûrement, il y aura au moins un pile dans la suite des lancers.
2) Considérons la probabilité uniforme sur $[0,1]$; nous avons vu que presque sûrement, le nombre obtenu est irrationnel.

♣ **Exercice 3.9.** Qu'est-ce que le complémentaire d'un événement négligeable ? D'un événement presque sûr ?

♣ **Exercice 3.10.** Montrer qu'une union dénombrable d'événements négligeables est un événement négligeable. Montrer qu'une intersection dénombrable d'événements presque sûrs est un événement presque sûr.

Chapitre 4

Premier TP. Prise en main de Scilab, premières simulations

4.1 Prise en main

Lancer Scilab :

```
scilab-5.3.3
Consortium Scilab (DIGITEO)
Copyright (c) 1989-2011 (INRIA)
Copyright (c) 1989-2007 (ENPC)
Initialisation : Chargement de l'environnement de travail
->
```

Le symbole $--\ >$ est le prompteur : on tape en mode calculatrice sur la ligne, on appuie sur "entrée" pour terminer la ligne, le résultat apparait sur la ligne suivante.
On peut utiliser les flèches haut/bas pour réutiliser les lignes de commande précédentes. Je n'écris pas les réponses de Scilab, seulement les lignes de commandes... N'hésitez pas à expérimenter par vous-même. Après chaque groupe de commendes, vous trouverez éventuellement des commentaires.
Introduction. Les commandes de base sont $*$, $+$, $-$, $/$.

```
-> 2+3
-> 2*3
-> 2/3
-> 2∧3
-> 2*(3+1)∧2
```

Utilisation des variables. Les résultats peuvent être stockés dans des variables :

```
-> x = 2+3
-> x
-> 2+3 // blabla
-> y = 1+2;
-> x*y
-> ans
-> mavariable = x∧y
-> mavariable
-> clear
-> mavariable
```

Le symbole // indique qu'on fait un commentaire : Scilab ne lit pas la suite de la ligne. Le `clear` efface le contenu de toutes les variables. Le ; à la fin d'une ligne empêche l'affichage, mais l'affectation de la variable est faite. La variable `ans` contient le résultat de la dernière opération.

Vecteurs. Un vecteur est une liste ; les délimiteurs d'un vecteur sont les crochets [et].

```
-> [1 2 3]
-> V=[1,2,3,1.4]
-> [2;5;1.7]
```

On peut faire des listes d'entiers et des subdivisions régulières :

```
-> 1:5
-> 0:2:10
-> 0:0.1:1
```

On peut faire des opérations sur les vecteurs :

```
-> x = 1:3
-> x+0.1
-> x*7
-> x/2
-> x ∧ 3
-> x
-> y = [2 4 6 8]
-> x+y
```

Bon, évidemment, il faut que les vecteurs aient des taille compatibles

```
-> y = [4 5 6]
-> x+y
-> x-y
-> x*y
-> z=x'
-> z*y
-> x.*y
```

Le .∗ fait la multiplication coefficient par coefficient. Essayer aussi ./. On peut récupérer les composantes d'un vecteur :

```
-> x = 2:2:10
-> x(1)
-> x(2)
-> help length
```

La commande `help` ouvre la fenêtre d'aide à la commande désirée. A la fin de la page se trouvent des exemples.

```
-> length(x)
-> x([1 2])
-> x([3 4])
```

On peut concaténer des vecteurs :

```
-> x=1:2:7
-> [ x 1 ]
-> [x x ]
```

Fonctions de vecteurs. On a déjà vu `length`.

```
-> sin(1)
-> %pi
-> x= [0: 1/12 : 1]*2*%pi
```

```
-> sin(x)
-> y = 1:3
-> z = exp(y)
-> log(z)
-> e
-> %e
```

Que font les lignes de commande suivantes ? Ne pas oublier l'aide si besoin...

```
-> x=[0:0.1:2*%pi];
-> y=sin(x);
-> plot2d(x,y)
-> clf
```

☐ **TP 4.1.** Tracer le graphe de la fonction $x \mapsto x^2$ sur $[-2, 2]$.
Que font les lignes de commande suivantes ? Ne pas oublier l'aide si besoin...

```
-> x=[1 2 3 4 5 6]
-> sum(x)
-> cumsum(x)
-> mean(x)
```

☐ **TP 4.2.** Calculer la somme des 30 premiers entiers et la somme des carrés des 30 premiers entiers.
Matrices. C'est essentiellement comme des vecteurs :

```
-> A = [1, 2, 3; 4, 5, 6]
-> A'
-> A(2,1)
-> A(1, :)
-> A(2, :)
-> A(:,2)
-> size(A)
-> size(A')
-> A∧2
-> A.∧2
-> B = [A , [7;8]]
-> C = [A ; [11 12 13]]
-> ones(2,3)
-> zeros(2,5)
-> A
-> sum(A)
-> sum(A,'r')
-> sum(A,'c')
```

4.2 Premières simulations

4.2.1 Histogramme pour le générateur `rand()` de loi uniforme sur $[0, 1]$

. Scilab possède une fonction `rand()` dont les appels successifs fournissent une suite de nombres parfaitement déterministes, mais ayant pour un certain nombre de tests statistiques les propriétés d'une suite de variables aléatoires indépendantes et identiquement distribuées, de loi uniforme sur $[0, 1]$.

```
-> rand
-> rand
-> rand(10,10)
-> x=rand(1,10)
```

Dans x, nous avons un 10 échantillon de loi uniforme sur $[0,1]$, c'est-à-dire la réalisation de 10 variables aléatoires iid de loi uniforme sur $[0,1]$. Scilab nous permet de résumer un échantillon dans un histogramme :

```
-> x=rand(1,20);
-> c=[0:0.1:1]
-> histplot(c,x);
```

Le vecteur c détermine les extrémités des classes : ainsi, notre histogramme a 10 classes : $[0,0.1]$, $[0.1,0.2]$, ..., $[0.9,1]$. La fonction `histplot` calcule la proportion d'éléments de x qui tombent dans chaque classe, et trace au dessus de chaque classe un bâton de hauteur égale à la proportion correspondante.

□ **TP 4.3.** Tracer l'histogramme d'un 500-échantillon puis d'un 5000-échantillon de loi uniforme sur $[0,1]$. Qu'observez-vous ?

4.2.2 Simulation de lancers d'un dé équilibré.

Scilab possède un générateur de nombres aléatoires plus sophistiqué : `grand`. Consulter son aide. Par exemple, que font les commandes suivantes ?

```
-> x = grand(1,20,'uin',1,6)
-> x == 6
-> bool2s(ans)
-> x = grand(1,20,'uin',1,6)
-> sum(bool2s(x == 6))
```

□ **TP 4.4.** On voudrait estimer la probabilité d'obtenir au moins un 6 en lancant 3 fois un dé équilibré.

1. fabriquer une matrice de taille 3×10, chaque composante représentant un lancer de dé. Une colonne va représenter une tentative de trois lancer de dé : on a donc 10 tentatives.
2. fabriquer la matrice qui indique si chaque lancer a donné un 6 ou non, puis le vecteur qui indique si chaque tentative a eu au moins un 6 ou non.
3. Quelle proportion des tentatives a permis d'obtenir un 6 ? Comparer avec la probabilité théorique ?
4. Reprendre avec un plus grand nombre (100, 1000...) de tentatives, et proposer une conjecture.

□ **TP 4.5.** Le problème du chevalier de Méré posé à Blaise Pascal. Qu'est ce qui est le plus probable : obtenir un six en lançant quatre fois un dé, ou obtenir au moins un double six en lançant 24 fois deux dés ?
Essayer de résoudre ce problème par simulation et par un calcul théorique.

□ **TP 4.6.** Le problème de Galilée. Le prince de Toscane demande un jour à Galilée : "Pourquoi lorsqu'on jette trois dés obtient-on plus souvent la somme 10 que la somme 9, bien que ces deux sommes soient obtenues de six façons différentes ?"
Essayer de résoudre ce problème par simulation et par un calcul théorique.

4.2.3 Evolution de la proportion de pile dans une suite de lancers

□ **TP 4.7.** Simuler 100 tirages d'une pièce tombant sur pile avec probabilité 0.43. Calculer la proportion de pile obtenue. Expérimentez et faites une conjecture.
Que font les lignes suivantes ?

```
-> p=0.7, n=10
-> tirage=bool2s(rand(1,n)<p)
-> somme=cumsum(tirage)
-> temps=[1:n]
-> moyenne=somme./temps
-> plot2d(temps,somme)
```

Que font les lignes suivantes ?

```
-> p=0.7, n=10
-> temps=[1:n];
-> moyenne=cumsum(bool2s(rand(1,n)<p))./temps;
-> plot2d(temps,somme)
```

Chapitre 5

Variables aléatoires réelles

Sommaire

5.1 Variable aléatoire réelle 34
- 5.1.1 Définitions 34
- 5.1.2 Variables aléatoires discrètes 34
- 5.1.3 Variable aléatoire à densité 35
- 5.1.4 Autres variables aléatoires réelles 36

5.2 Simulation de variables aléatoires réelles 36
- 5.2.1 Exemples de simulation de variables aléatoires discrètes 36
- 5.2.2 Exemples de simulation de variables aléatoires à densité 37

Objectifs :

- Introduire des fonctions du résultat d'une expérience aléatoire, ou variables aléatoires.
- Comprendre la notion de loi et savoir donner la loi d'une variable aléatoire réelle.
- Savoir calculer la loi de la variable aléatoire à partir de sa fonction de répartition, et réciproquement.
- Savoir simuler des variables aléatoires simples.
- Connaître les lois classiques, discrètes et à densité.

Mots-clés :

- Variable aléatoire, variable aléatoire discrète, variable aléatoire à densité.
- Loi d'une v.a., fonction de répartition.

Je lance trois dés équilibrés. Quel résultat vais-je obtenir en moyenne ? On pourrait construire un espace de probabilité pour modéliser cette expérience. Mais on peut aussi définir X_1 le résultat du premier dé, X_2 le résultat du deuxième dé et X_3 le résultat du troisième dé, et s'intésser étudier la quantité $S = X_1 + X_2 + X_2$. Ces quantités sont des *fonctions du hasard*, on va les appeler des variables aléatoires.

5.1 Variable aléatoire réelle

5.1.1 Définitions

Définition 5.1 *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité. Une* variable aléatoire *(abrégée v.a.) est une application* $X : \Omega \to \mathbb{R}$. *On note* $X(\Omega)$ *l'ensemble des valeurs prises par la variable aléatoire* X.

La *loi (de probabilité)* de X est la probabilité $\mathbb{P}_X$ définie sur $X(\Omega)$ qui caractérise le comportement statistique de X. On n'en donne pas de définition en général, seulement dans le cas des v.a. discrètes et des v.a. à densité. Dans ce cadre, l'espace de probabilité du départ, $(\Omega, \mathcal{F}, \mathbb{P})$ n'a pas d'importance et on ne le précise pas : ce qui nous intéresse, c'est le comportement statistique de X.

♠ **Attention !** Ne pas confondre la *variable aléatoire* (la fonction du hasard) avec sa *loi* (la probabilité qui décrit son comportement « statistique ») !

Définition 5.2 *Soit* X *une variable aléatoire définie sur un espace de probabilité* $(\Omega, \mathcal{F}, \mathbb{P})$. *On appelle* fonction de répartition *de* X *l'application*

$$F_X : \begin{array}{l} \mathbb{R} \to [0,1] \\ t \mapsto \mathbb{P}_X(\,]-\infty, t]) = \mathbb{P}(X \leq t) \end{array}.$$

Proposition 5.3 *Soit* X *une variable aléatoire réelle définie sur un espace de probabilité* $(\Omega, \mathcal{F}, \mathbb{P})$, *et soit* F_X *sa fonction de répartition. Alors*

1. F_X *est une fonction croissante, dont la limite en* $-\infty$ *est* 0 *et la limite en* $+\infty$ *est* 1.
2. *La fonction de répartition caractérise la loi : deux v.a.* X *et* Y *ont la même loi si et seulement si elles ont la même fonction de répartition.*

5.1.2 Variables aléatoires discrètes

Définition 5.4 *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité. Une* variable aléatoire discrète *est une application* $X : \Omega \to \mathbb{R}$, *telle que l'ensemble* $X(\Omega)$ *des valeurs prises* X *est fini ou dénombrable. La* loi (de probabilité) *de* X *est la probabilité* $\mathbb{P}_X$ *définie sur l'ensemble discret* $X(\Omega)$ *et caractérisée par :*

$$\forall A \subset X(\Omega) \quad \mathbb{P}_X(A) = \mathbb{P}(X \in A).$$

♢ **En pratique.** Pour donner la loi d'une variable aléatoire discrète X, on donne l'ensemble $X(\Omega)$ des valeurs prises par X et, pour tout $x \in X(\Omega)$, la probabilité $\mathbb{P}_X(\{x\}) = \mathbb{P}(X = x)$.
On n'oublie pas de vérifier que $\displaystyle\sum_{x \in X(\Omega)} \mathbb{P}_X(\{x\}) = 1$.

♣ **Exercice 5.1.** On lance un dé truqué : il tombe deux fois plus souvent sur les nombres pairs que les nombres impairs ; par contre, il tombe sur chacun des nombres pairs avec la même probabilité, et sur chacun des nombres impairs avec la même probabilités. On note X le résultat. Donner la loi de X.

♣ **Exercice 5.2.** On lance deux fois un dé équilibré, et on note S la somme des résultats. Déterminer la loi de S.

Proposition 5.5 *La fonction de répartition F_X d'une variable aléatoire* discrète *X est une fonction en escalier croissante et continue à droite.*
L'ensemble $X(\Omega)$ des valeurs prises par X est l'ensemble des points de discontinuité de F_X ; la probabilité de l'événement $\{X = a\}$ est la hauteur du saut de F_X en a.

♣ **Exercice 5.3.** Déterminer la fonction de répartition d'une variable aléatoire de loi uniforme sur $\{1, \ldots, 6\}$.

♣ **Exercice 5.4.** On donne la fonction de répartition

$$F_X(t) = \begin{cases} 0 & \text{si} \quad t < 2/3 \\ 0,3 & \text{si} \quad 2/3 \leq t < 1 \\ 0,6 & \text{si} \quad 1 \leq t < 7/5 \\ 1 & \text{si} \quad t \geq 7/5. \end{cases}$$

Tracer F_X et déterminer la loi de X.

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $X : \Omega \to \mathbb{R}$ une variable aléatoire discrète. On donne dans le tableau suivant les caractéristiques de quelques lois classiques.

Nom	*Paramètre(s)*	*Support*	$\mathbb{P}(X = k)$	*Exemple*
Loi de Bernoulli	$p \in]0,1[$	$\{0,1\}$	p si $k=1$, $1-p$ si $k=0$	Tirage dans une urne Pile ou face biaisé
Loi binômiale	$n \in \mathbb{N}^*$ $p \in]0,1[$	$\{0,\ldots,n\}$	$\binom{n}{k} p^k (1-p)^{n-k}$	Nombre de succès pour n tirages avec remise
Loi géométrique	$p \in]0,1[$	$\mathbb{N}^*$	$p\,(1-p)^{k-1}$	Temps du premier succès pour des tirages avec remise
Loi uniforme	E fini	E	$\dfrac{1}{\lvert E\rvert}$	Phénomènes équiprobables
Loi de Poisson	$\lambda > 0$	$\mathbb{N}$	$e^{-\lambda}\,\dfrac{\lambda^k}{k!}$	Nombre de bus arrivant en 1h si le temps d'attente moyen est $1/\lambda$ h

♣ **Exercice 5.5.** On pourra vérifier pour chacun de ces lois que la somme des $\mathbb{P}(X = k)$ pour k dans le support de la loi vaut 1.

5.1.3 Variable aléatoire à densité

Définition 5.6 *Soit f une densité sur $\mathbb{R}$. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, et $X : \Omega \to \mathbb{R}$ une variable aléatoire. On dit que X est une variable aléatoire réelle* de densité *f si pour tous $a, b \in \mathbb{R}$ avec $a < b$, sa loi $\mathbb{P}_X$ est caractérisée par*

$$\mathbb{P}_X([a,b]) = \mathbb{P}(X \in [a,b]) = \int_a^b f(x)dx.$$

◊ **En pratique.** Pour donner la loi d'une variable aléatoire réelle à densité, on donne simplement sa densité. On n'oublie pas de vérifier à la fin que pour tout $x \in \mathbb{R}$, $f(x) \geq 0$ et que $\int_{\mathbb{R}} f(x) = 1$. Ceci permet de détecter d'éventuelles erreurs de calcul.

Proposition 5.7 *La fonction de répartition F_X d'une variable aléatoire X* de densité *f est une fonction continue, C^1 par morceaux. C'est l'unique primitive de la densité f dont la limite en $-\infty$ est nulle.*
Réciproquement, si F_X est la fonction de répartition d'une variable aléatoire X, et si F_X est une fonction continue, C^1 par morceaux, alors X est une variable aléatoire à densité, et la densité f de X s'obtient de la façon suivante :
– si F_X est dérivable en x, alors on prend $f(x) = F'_X(x)$;
– sinon, on prend pour $f(x)$ une valeur arbitraire.

♣ **Exercice 5.6.** Soit X une variable aléatoire de loi uniforme sur $]0,1[$. Déterminer et tracer sa fonction de répartition.

♣ **Exercice 5.7.** Soit X une variable aléatoire définie sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, suivant la loi uniforme sur $[0,1]$. On pose $Y = X^2$. Montrer que Y admet une densité g que l'on déterminera.
Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, et $X : \Omega \to \mathbb{R}$ une variable aléatoire à densité. On donne dans le tableau suivant les caractéristiques de quelques lois classiques :

Nom	*Paramètre(s)*	*Support*	*Densité $f(x)$*	*Exemple*
Loi uniforme	$a, b \in \mathbb{R}$, $a < b$	$[a,b]$	$\frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$	Phénoménes équiprobables
Loi exponentielle	$\lambda \in \mathbb{R}_+^*$	$\mathbb{R}^+$	$\lambda \exp(-\lambda x)\mathbb{1}_{\mathbb{R}^+}(x)$	Durée de vie
Loi Gaussienne, **Loi normale**	$m \in \mathbb{R}$ $\sigma^2 \in \mathbb{R}_+^*$	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$	Théorème Central Limite, erreurs

5.1.4 Autres variables aléatoires réelles

♠ **Attention !** il existe des variables aléatoires réelles qui ne sont ni discrétes ni à densité. On n'insistera pas sur ces autres variables aléatoires, mais il faut garder en tête qu'elles existent.

5.2 Simulation de variables aléatoires réelles

La problématique dans cette section est la suivante : Scilab possède une fonction `rand()` dont les appels successifs fournissent une suite de nombres parfaitement déterministes, mais ayant pour un certain nombre de tests statistiques les propriétés d'une suite de variables aléatoires indépendantes et identiquement distribuées, de loi uniforme sur $[0,1]$. On travaillera dans la suite en supposant que les appels successifs de `rand()` forment une suite $(U_n)_{n\in\mathbb{N}}$ de vaiid de loi uniforme sur $[0,1]$.
Scilab possède aussi un générateur de nombres aléatoires `grand()` plus évolué, permettant de simuler les lois classiques.

<u>Problème</u> : Comment simuler une variable aléatoire suivant une loi donnée, à partir de la loi uniforme sur $[0,1]$?
Il s'agit d'écrire un programme qui donne en sortie un nombre aléatoire X, suivant la loi fixée.
Avant de se lancer, regarder le tutoriel Scilab...

5.2.1 Exemples de simulation de variables aléatoires discrètes

□ **TP 5.1.** On souhaite simuler le lancer d'une pièce qui tombre sur pile avec probabilité p. Ecrire une fonction `bernoulli1(p)` qui renvoit un tirage selon la loi de Bernoulli de paramètre p.

□ **TP 5.2.** Ecrire une fonction `deequilibre` qui simule le lancer d'un dé équilibré.

□ **TP 5.3.** Ecrire une fonction `binomiale(n,p)` qui renvoit un tirage selon la loi binomiale de paramètres (n,p).

Penser à l'apparition de cette loi dans le schéma de Bernoulli fini.

□ **TP 5.4.** En s'appuyant sur l'apparition de la loi géométrique dans le schéma de Bernoulli infini, écrire une fonction `geometrique1(p)` qui renvoit un tirage suivant la loi géométrique de paramètre p.

5.2.2 Exemples de simulation de variables aléatoires à densité

□ **TP 5.5.** Soit X une variable aléatoire définie sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, suivant la loi uniforme sur $[0, 1]$. Soit $\alpha > 0$ et $\beta \in \mathbb{R}$; on pose $Y = \alpha X + \beta$. En utilisant sa fonction de répartition, montrer que Y admet une densité g que l'on déterminera.
En déduire une façon de simuler une variable aléatoire de loi uniforme sur $[a, b]$.

□ **TP 5.6.** Soit X une variable aléatoire définie sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, suivant la loi uniforme sur $[0, 1]$. Soit $\lambda > 0$; on pose $Y = \frac{1}{\lambda} \ln(X)$. Montrer que Y admet une densité g que l'on déterminera.
En déduire une façon de simuler une variable aléatoire de loi exponentielle de paramètre λ.

Chapitre 6

Espérance et variance des variables aléatoires réelles

Sommaire

6.1 Espérance des variables aléatoires réelles 39
 6.1.1 Définitions 39
 6.1.2 Propriétés de l'espérance 40
 6.1.3 Espérance d'une fonction d'une variable aléatoire 41
6.2 Variance d'une variable aléatoire réelle 41
6.3 Calculs pour les v.a. discrètes classiques 43
 6.3.1 Loi de Bernoulli de paramètre $p \in [0,1]$ 43
 6.3.2 Loi binomiale de paramètres $n \in \mathbb{N}^*$, $p \in [0,1]$ 43
 6.3.3 Loi uniforme sur $\{0,\ldots,n\}$ 44
 6.3.4 Loi de Poisson de paramètre $\lambda \in]0,\infty[$ 44
 6.3.5 Loi géométrique de paramètre $p \in]0,1]$ 45
6.4 Calculs pour les v.a. à densité classiques 46
 6.4.1 Lois uniformes 46
 6.4.2 Lois exponentielles 46
 6.4.3 Lois normales 47

Objectifs :

• Définir l'espérance et la variance d'une variable aléatoire.

• Savoir calculer l'espérance et la variance d'une variable aléatoire, pour une variable aléatoire discrète et pour une variable aléatoire à densité.

Outils :

• Propriétés de l'espérance : linéarité, positivité, calcul par découpage selon une partition.

• Propriétés de la variance.

• Théorème de transfert pour calculer $\mathbb{E}(\varphi(X))$.

Mots-clés :

• Espérance, variance, intégrabilité, moment d'ordre 2.

La loi d'une variable aléatoire comporte toute l'information statistique sur cette variable aléatoire. Mais on a parfois besoin de quantités plus simples qui résument cette information statistique : l'espérance va résumer le comportement moyen d'une variable aléatoire, tandis que la variance va résumer sa dispersion.

6.1 Espérance des variables aléatoires réelles

L'espérance d'une variable aléatoire X donne son comportement statistique moyen : c'est une moyenne des valeurs prises par X pondérées par le poids que leur donne la loi de X. La définition mathématique prend des formes différente suivante que la loi de X est discrète ou à densité.

6.1.1 Définitions

Définition 6.1 • *Soit X une variable aléatoire discrète, définie sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On dit qu'elle est* intégrable *ou qu'elle* admet une espérance *si et seulement si*

$$\text{la somme ou la série} \sum_{x \in X(\Omega)} |x| \, \mathbb{P}(X = x) \text{ converge.}$$

Dans ce cas, son espérance *est le nombre* $\mathbb{E}[X] = \sum_{x \in X(\Omega)} x \, \mathbb{P}(X = x)$.

• *Soit X une variable aléatoire de densité f, définie sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On dit qu'elle est* intégrable *ou qu'elle* admet une espérance *si et seulement si*

$$\text{l'intégrale} \int_{\mathbb{R}} |x| f(x) dx \text{ est convergente.}$$

Dans ce cas, son espérance *est le nombre* $\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x) dx$.

Remarque. Si X est une variable aléatoire qui ne prend qu'un nombre fini de valeurs, son espérance, définie par une somme d'un nombre fini de termes, a toujours un sens. En revanche, si X est une variable aléatoire qui prend un nombre dénombrable de valeurs, l'espérance est définie par une série, dont il faut étudier la convergence. La condition d'intégrabilité assure que la série qui donne l'espérance est absolument convergente, donc convergente. Si X est une variable aléatoire de densité f, telle que le support de f soit borné, son espérance, définie par une intégrale sur un intervalle borné, a toujours un sens. En revanche, si le support de f n'est pas borné, l'espérance est définie par une intégrale impropre, dont il faut étudier la convergence.

▷ **Exemple.** Considérons une variable aléatoire X de loi uniforme sur $\{0, \ldots, n\}$ où n est un entier naturel fixé. C'est une variable aléatoire discrète, qui ne prend qu'un nombre fini de valeurs, à savoir $n + 1$, donc elle admet une espérance :

$$\mathbb{E}[X] = \sum_{k=0}^{n} k \, \mathbb{P}(X = k) = \sum_{k=0}^{n} k \, \frac{1}{n+1} = \frac{1}{n+1} \sum_{k=1}^{n} k = \frac{n(n+1)}{2(n+1)} = \frac{n}{2}.$$

▷ **Exemple.** On considère une variable aléatoire X qui suit la loi de Poisson de paramètre $\lambda > 0$, c'est-à-dire que

$$X(\Omega) = \mathbb{N} \quad \text{et} \quad \forall k \in \mathbb{N}, \; \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

On pose, pour tout $k \in \mathbb{N}$,

$$u_k = |k| \mathbb{P}(X = k) = k e^{-\lambda} \frac{\lambda^k}{k!},$$

et on regarde la série de terme général $(u_k)_{k \geq 0}$. Pour $k \geq 1$,

$$u_k = \lambda e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!}.$$

On reconnaît, à une constante multiplicative près, le terme général de la série exponentielle de paramètre λ, donc la série de terme général $(u_k)_{k\geq 0}$ converge, donc X est intégrable.
Maintenant,

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k\,\mathbb{P}(X=k) = \sum_{k=0}^{\infty} k\,e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty}\lambda\frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}\sum_{l=0}^{\infty}\frac{\lambda^l}{l!} = \lambda e^{-\lambda}e^{\lambda} = \lambda.$$

♣ **Exercice 6.1.** Soit $\alpha > 0$. Soit X une variable aléatoire à valeurs dans $\mathbb{N}^*$ telle que

$$\forall k \in \mathbb{N}^*, \quad \mathbb{P}(X=k) = \frac{C_\alpha}{k^\alpha}$$

pour un certain $C_\alpha > 0$.

1. À quelle condition sur α une telle probabilité existe-elle ?
2. Comment doit-on choisir C_α pour avoir une loi de probabilité ?
3. À quelle condition sur α la variable aléatoire X est-elle intégrable ?

▷ **Exemple.** Considérons une variable aléatoire X de loi uniforme sur $[a,b]$ où $a < b$ sont deux nombres réels fixés. C'est une variable aléatoire à densité, le support de sa densité est $[a,b]$, donc borné et donc X est intégrable :

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{+\infty} x\frac{1}{b-a}\mathbb{1}_{[a,b]}(x)dx = \frac{1}{b-a}\int_a^b xdx = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}.$$

▷ **Exemple.** On considère une variable aléatoire X qui suit la loi exponentielle de paramètre $\lambda > 0$. La densité f de X est donc

$$\forall x \in \mathbb{R} \quad f(x) = \lambda\exp(-\lambda x)\mathbb{1}_{\mathbb{R}_+}(x).$$

Pour savoir si X est intégrable, on doit étudier la convergence de l'intégrale suivante :

$$\int_{-\infty}^{+\infty} |x|f(x)dx = \int_{-\infty}^{+\infty} |x|\lambda\exp(-\lambda x)\mathbb{1}_{\mathbb{R}_+}(x)dx = \int_0^{+\infty}\lambda x\exp(-\lambda x)dx.$$

C'est une intégrale impropre en $+\infty$. Comme la fonction à intégrer est le produit d'un polynôme et d'une fonction exponentielle, on souhaite intégrer par parties. Soit donc $M > 0$:

$$\begin{aligned}\int_0^M \lambda x\exp(-\lambda x)dx &= [-x\exp(-\lambda x)]_0^M + \int_0^M \exp(-\lambda x)dx\\ &= -M\exp(-\lambda M) + [-\frac{1}{\lambda}\exp(-\lambda x)]_0^M\\ &= -M\exp(-\lambda M) + \frac{1}{\lambda} - \frac{1}{\lambda}\exp(-\lambda M).\end{aligned}$$

On voit alors que quand M tend vers $+\infty$, cette expression converge vers $\frac{1}{\lambda}$, donc l'intégrale impropre $\int_{-\infty}^{+\infty}|x|f(x)dx$ converge, donc la variable aléatoire X est intégrable et

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^{+\infty}\lambda x\exp(-\lambda x)dx = \frac{1}{\lambda}.$$

6.1.2 Propriétés de l'espérance

Pour ce paragraphe, on donne les propriétés de l'espérance utiles dans la pratique.

Proposition 6.2 (Critère d'intégrabilité par comparaison) *Soit $(\Omega,\mathcal{F},\mathbb{P})$ un espace de probabilité, et soit X et Y deux variables aléatoires réelles définies sur $(\Omega,\mathcal{F},\mathbb{P})$. On suppose que $|Y|$ est intégrable et que*

$$\forall\omega\in\Omega \quad |X(\omega)| \leq |Y(\omega)|.$$

Alors X est intégrable.
En particulier, une variable aléatoire bornée est toujours intégrable.

Proposition 6.3 (Calcul avec le principe de partition) *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité, et soit* X *une variable aléatoire réelle définie sur* $(\Omega, \mathcal{F}, \mathbb{P})$ *et intégrable, et soit* $(A_i)_{i \in I}$ *un système complet. Alors*

$$\mathbb{E}(X) = \sum_{i \in I} \mathbb{E}(X \mathbb{1}_{A_i}).$$

Proposition 6.4 (Linéarité et positivité) *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité.*

1. *L'espérance est linéaire : si* a *et* $b \in \mathbb{R}$ *et si* X *et* Y *sont deux v.a. définies sur* $(\Omega, \mathcal{F}, \mathbb{P})$ *et admettant une espérance, alors* $aX + bY$ *admet une espérance et*

$$\mathbb{E}[aX + bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y].$$

   Cas particulier : si a *est une constante réelle,* $\mathbb{E}[a] = a$.

2. *L'espérance est une application positive : si* $X \geq 0$ *(ce qui signifie que* $\forall \omega \in \Omega,\ X(\omega) \geq 0$*) et si* X *admet une espérance, alors* $\mathbb{E}[X] \geq 0$.

   Cas particulier : si $X \leq Y$ *(ce qui signifie que* $\forall \omega \in \Omega,\ X(\omega) \leq Y(\omega)$*), et si* X *et* Y *sont intégrables, alors* $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

6.1.3 Espérance d'une fonction d'une variable aléatoire

Théorème 6.5 (Théorème de transfert) • *Soit* X *une variable aléatoire discrète, définie sur un espace de probabilité* $(\Omega, \mathcal{F}, \mathbb{P})$ *et soit* φ *une application de* $\mathbb{R}$ *dans* $\mathbb{R}$. *Alors* $\varphi(X) : \Omega \to \mathbb{R}$ *est une variable aléatoire discrète. Elle est intégrable si et seulement si* $\sum_{x \in X(\Omega)} |\varphi(x)|\, \mathbb{P}(X = x) < +\infty$. *Dans le cas où elle est intégrable,*

$$\mathbb{E}[\varphi(X)] = \sum_{x \in X(\Omega)} \varphi(x)\, \mathbb{P}(X = x).$$

• *Soit* X *une variable aléatoire de densité* f, *définie sur un espace de probabilité* $(\Omega, \mathcal{F}, \mathbb{P})$ *et soit* φ *une application continue par morceaux de* $\mathbb{R}$ *dans* $\mathbb{R}$. *Alors* $\varphi(X) : \Omega \to \mathbb{R}$ *est une variable aléatoire. Elle est intégrable si et seulement si* $\int_{-\infty}^{+\infty} |\varphi(x)| f(x) dx$ *est convergente. Dans le cas où elle est intégrable,*

$$\mathbb{E}[\varphi(X)] = \int_{-\infty}^{+\infty} \varphi(x) f(x) dx.$$

▷ **Exemple.** Soit $\varphi : x \mapsto x^2$. Soit X une v.a. discrète. La v.a. X^2 est intégrable ssi $\displaystyle\sum_{x \in X(\Omega)} x^2\, \mathbb{P}(X = x) < +\infty$.
Si c'est le cas,

$$\mathbb{E}[X^2] = \sum_{x \in X(\Omega)} x^2\, \mathbb{P}(X = x).$$

Soit X une v.a. de densité f. La v.a. X^2 est intégrable ssi $\int_{-\infty}^{+\infty} x^2 f(x) dx < +\infty$. Si c'est le cas,

$$\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} x^2 f(x) dx.$$

6.2 Variance d'une variable aléatoire réelle

La variance d'une variable aléatoire est un nombre que traduit sa dispersion par rapport à son espérance.

Lemme 6.6 *Soit* X *une variable aléatoire définie sur un espace de probabilité* $(\Omega, \mathcal{F}, \mathbb{P})$ *telle que la variable aléatoire* X^2 *est intégrable. Alors* X *est intégrable (c'est-à-dire* $\mathbb{E}[|X|] < \infty$*).*

Démonstration. Pour tout $x \in \mathbb{R}$, $(1-|x|)^2 \geq 0 \Leftrightarrow |x| \leq \frac{1}{2}(1+x^2)$. Ainsi,

$$\forall \omega \in \Omega \quad |X(\omega)| \leq \frac{1}{2}(1+X(\omega)^2),$$

et donc $|X|$ est intégrable. On obtient aussi que $\mathbb{E}[|X|] \leq \dfrac{1}{2}(1+\mathbb{E}[X^2])$. □

Définition 6.7 *Soit X une variable aléatoire définie sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que X* admet un moment d'ordre 2*, ou que X est* de carré intégrable*, si X^2 est intégrable. Dans ce cas, on définit sa* variance $\mathrm{Var}(X)$ *et son* écart-type σ_X *par*

$$\mathrm{Var}(X) = \mathbb{E}\big[(X-\mathbb{E}[X])^2\big] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \qquad \textit{et} \qquad \sigma_X = \sqrt{\mathrm{Var}(X)}.$$

Démonstration. Soit X une variable aléatoire telle que $\mathbb{E}[X^2] < +\infty$. Montrons que $\mathrm{Var}(X)$ est bien définie : il faut montrer que la variable aléatoire Y définie comme suit est intégrable :

$$\forall \omega \in \Omega \quad Y(\omega) = (X(\omega) - \mathbb{E}[X])^2.$$

D'après le lemme précédent, X est intégrable : notons $m = \mathbb{E}(X)$. Alors

$$\forall \omega \in \Omega \quad Y(\omega) = (X(\omega)-m)^2 = X(\omega)^2 - 2mX(\omega) + m^2.$$

Comme X^2, X et les constantes sont intégrables, on en déduit par linéarité que Y est intégrable, et donc que $\mathrm{Var}(X)$ est bien définie. De plus,

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[(X-\mathbb{E}[X])^2] = \mathbb{E}[(X-m)^2] = \mathbb{E}[X^2 - 2mX + m^2] \\ &= \mathbb{E}[X^2] - 2m\,\mathbb{E}[X] + m^2 = \mathbb{E}[X^2] - m^2.\end{aligned}$$

□

Remarque. La variance est la moyenne du carré de l'écart entre X et sa moyenne. Elle mesure à quel point X varie autour de sa moyenne. Dans les applications pratiques, si X est exprimé dans une certaine unité (par exemple des mètres), $\mathrm{Var}(X)$ est exprimé dans cette unité au carré alors que σ_X est exprimé dans la même unité que X. Pour cette raison, on préférera généralement l'écart-type dans les applications numériques.

Proposition 6.8 *Soit X une v.a. discrète.*

1. $\mathrm{Var}(X) \geq 0$.

2. $\mathrm{Var}(X) = 0 \Leftrightarrow X$ *est constante presque sûrement.*

3. $\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X)$.

Démonstration. 1. $(X-\mathbb{E}[X])^2 \geq 0$ donc $\mathrm{Var}(X) \geq 0$ par positivité de l'espérance. Remarquons que cela implique, pour toute variable aléatoire telle que X^2 est intégrable, que

$$\mathbb{E}(X^2) \geq (\mathbb{E}(X))^2.$$

3. ♣ **Exercice 6.2.** □

♠ **Attention !** En général, $\mathrm{Var}(X+Y) \neq \mathrm{Var}(X) + \mathrm{Var}(Y)$. Regarder par exemple le cas $X = Y$.

6.3 Calculs pour les v.a. discrètes classiques

6.3.1 Loi de Bernoulli de paramètre $p \in [0,1]$

• $X(\Omega) = \{0,1\}$, $\mathbb{P}(X=1) = p$ et $\mathbb{P}(X=0) = 1-p$.
La variable aléatoire X est bornée, donc intégrable.

$$\begin{aligned}\mathbb{E}[X] &= 1.p + 0.(1-p) = p\ ;\\ \mathbb{E}[X^2] &= 1^2.p + 0^2.(1-p) = p\ ;\\ \mathrm{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p\,(1-p).\end{aligned}$$

Définition 6.9 *On définit la* fonction indicatrice *de la partie $A \subset \Omega$ par*

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & si\ \omega \in A,\\ 0 & si\ \omega \notin A.\end{cases}$$

Si A est un événement, $\mathbb{1}_A$ est une variable aléatoire à valeurs dans $\{0,1\}$ et $\mathbb{1}_A(\Omega) = \{0,1\}$.

$$\begin{aligned}\mathbb{P}(\mathbb{1}_A = 1) &= \mathbb{P}(\{\omega \in \Omega\ :\ \mathbb{1}_A(\omega) = 1\})\\ &= \mathbb{P}(\{\omega \in \Omega\ :\ \omega \in A\}) = \mathbb{P}(A).\end{aligned}$$

Un calcul similaire donne $\mathbb{P}(\{\mathbb{1}_A = 0\}) = \mathbb{P}(\bar{A})$. (On peut aussi remarquer que $\mathbb{P}(\{\mathbb{1}_A = 0\}) = 1 - \mathbb{P}(\{\mathbb{1}_A = 1\}) = 1 - \mathbb{P}(A) = \mathbb{P}(\bar{A})$.) Ainsi, la v.a. $\mathbb{1}_A$ suit une loi de Bernoulli de paramètre $\mathbb{P}(A) \in [0,1]$.

♢ **En pratique.** Ces variables aléatoires servent à modéliser des expériences aléatoires n'ayant que deux issues possibles, que l'on code avec 0 ou 1 : lancer d'un pile/face, expérience avec succès/échec, état d'un composant fonctionne/en panne, vote à un référendum oui/non, etc

6.3.2 Loi binomiale de paramètres $n \in \mathbb{N}^*$, $p \in [0,1]$

• $X(\Omega) = \{0,\dots,n\}$, $\forall k \in \{0,\dots,n\}$, $\mathbb{P}(X=k) = \binom{n}{k} p^k\,(1-p)^{n-k}$.
La variable aléatoire X est bornée, donc intégrable.
Notez que pour n, $k \geq 1$, on a $k\binom{n}{k} = n\binom{n-1}{k-1}$.

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=0}^{n} k\binom{n}{k} p^k\,(1-p)^{n-k} = \sum_{k=1}^{n} k\binom{n}{k} p^k\,(1-p)^{n-k}\\ &= \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k\,(1-p)^{(n-1)-(k-1)} = np\sum_{k=1}^{n}\binom{n-1}{k-1} p^{k-1}(1-p)^{(n-1)-(k-1)}\\ &= np\,\big(p + (1-p)\big)^{n-1} = np.\end{aligned}$$

En appliquant deux fois la formule précédente,

$$\begin{aligned}\mathbb{E}[X(X-1)] &= \sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k\,(1-p)^{n-k} = \sum_{k=2}^{n} k(k-1)\binom{n}{k} p^k\,(1-p)^{n-k}\\ &= \sum_{k=2}^{n} n(n-1)\binom{n-2}{k-2} p^k\,(1-p)^{(n-2)-(k-2)} = n(n-1)\,p^2\sum_{k=2}^{n}\binom{n-2}{k-2} p^{k-2}(1-p)^{(n-2)-(k-2)}\\ &= n(n-1)\,p^2\,\big(p + (1-p)\big)^{n-2} = n(n-1)\,p^2.\end{aligned}$$

Ainsi,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 = n(n-1)\,p^2 + np - n^2p^2 = np\,(1-p).$$

Remarque. On verra plus tard une méthode plus rapide pour retrouver ces résultats.
Nous avons déjà croisé cette loi. Si l'on considère n lancers successifs indépendants d'une pièce tombant sur "pile" avec probabilité $p \in]0,1[$, et que l'on note X le nombre de pile obtenu. Nous avons vu que X suit la loi binômialede paramètres (n,p).
Cette loi apparait en fait chaque fois que l'on somme n variables aléatoires indépendantes et identiquement distribuées de loi de Bernoulli de paramètre p. Voir le chapitre sur l'indépendance des variables aléatoires.

6.3.3 Loi uniforme sur $\{0,\dots,n\}$

• $X(\Omega) = \{0,\dots,n\}$, $\forall k \in \{0,\dots,n\}$, $\mathbb{P}(X=k) = 1/(n+1)$.
La variable aléatoire X est bornée, donc intégrable.

Rappel : $\displaystyle\sum_{k=0}^{n} k = \frac{n(n+1)}{2}$ et $\displaystyle\sum_{k=0}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$.

$$\mathbb{E}[X] = \sum_{k=0}^{n} \frac{k}{n+1} = \frac{n}{2} \ ;$$
$$\mathbb{E}[X^2] = \sum_{k=0}^{n} \frac{k^2}{n+1} = \frac{n(2n+1)}{6} \ ;$$
$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{n(n+2)}{12}.$$

Ces variables aléatoires servent à modéliser des expériences aléatoires équiprobables : lancer d'un dé équilibré, choix d'une carte au hasard dans un jeu de cartes, tirage dans une urne... On les retrouve dans les exercices de dénombrement.

6.3.4 Loi de Poisson de paramètre $\lambda \in]0,\infty[$

• $X(\Omega) = \mathbb{N}$ et $\forall k \in \mathbb{N}$, $\mathbb{P}(X=k) = e^{-\lambda}\,\dfrac{\lambda^k}{k!}$.
Montrons que X^2 est intégrable. On regarde

$$\sum_{x \in X(\Omega)} |x|^2 \mathbb{P}(X=x) = \sum_{k=0}^{+\infty} k^2\, e^{-\lambda}\,\frac{\lambda^k}{k!}.$$

C'est une série à terme positife, de terme général $u_k = k^2\, e^{-\lambda}\,\frac{\lambda^k}{k!}$.

$$\frac{u_{k+1}}{u_k} = \frac{(k+1)^2}{k^2}\frac{\lambda^{k+1}}{\lambda^k}\frac{k!}{(k+1)!} = \left(1+\frac{1}{k}\right)^2 \frac{\lambda}{k+1},$$

qui tend evrs 0 quand k tend vers $+\infty$. Donc (critère de d'Alembert) le série est convergente, donc X^2 est intégrable. Donc X et aussi $X(X-1)$ sont intégrables.

$$\mathbb{E}[X] = \sum_{k=0}^{+\infty} k\,e^{-\lambda}\,\frac{\lambda^k}{k!} = \sum_{k=1}^{+\infty} k\,e^{-\lambda}\,\frac{\lambda^k}{k!} = \lambda\,e^{-\lambda}\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda\,e^{-\lambda}e^{\lambda} = \lambda \ ;$$
$$\mathbb{E}[X(X-1)] = \sum_{k=0}^{+\infty} k(k-1)\,e^{-\lambda}\,\frac{\lambda^k}{k!} = \sum_{k=2}^{+\infty} k(k-1)\,e^{-\lambda}\,\frac{\lambda^k}{k!} = \lambda^2\,e^{-\lambda}\sum_{k=2}^{+\infty}\frac{\lambda^{k-2}}{(k-2)!} = \lambda^2\,e^{-\lambda}e^{\lambda} = \lambda^2 \ ;$$
$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

6.3.5 Loi géométrique de paramètre $p \in]0,1]$

• $X(\Omega) = \mathbb{N}^*$ et $\forall k \in \mathbb{N}^*,\ \mathbb{P}(X = k) = p\,(1-p)^{k-1}$.

Montrons que X^2 est intégrable. On regarde

$$\sum_{x \in X(\Omega)} |x|^2 \mathbb{P}(X = x) = \sum_{k=1}^{+\infty} k^2 p(1-p)^{k-1}.$$

C'est une série à terme positife, de terme général $u_k = k^2 p(1-p)^{k-1}$.

$$\frac{u_{k+1}}{u_k} = \frac{(k+1)^2}{k^2} \frac{p(1-p)^k}{p(1-p)^{k-1}} = \left(1 + \frac{1}{k}\right)^2 (1-p),$$

qui tend vers $1 - p < 1$ quand k tend vers $+\infty$. Donc (critère de d'Alembert) le série est convergente, donc X^2 est intégrable. Donc X et aussi $X(X-1)$ sont intégrables.
Nous utiliserons les formules suivantes :

$$\boxed{\forall x \in]-1,1[\qquad \sum_{k=1}^{\infty} k\, x^{k-1} = \frac{1}{(1-x)^2} \qquad \text{et} \qquad \sum_{k=2}^{\infty} k(k-1)\, x^{k-2} = \frac{2}{(1-x)^3}.}$$

(on obtient ces formules en dérivant $\displaystyle\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$).

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{+\infty} k\, p\, (1-p)^{k-1} = \sum_{k=1}^{+\infty} k\, p\, (1-p)^{k-1} = p \sum_{k=1}^{+\infty} k\, (1-p)^{k-1} = \frac{p}{\big(1-(1-p)\big)^2} = \frac{1}{p}\ ; \\
\mathbb{E}[X(X-1)] &= \sum_{k=0}^{+\infty} k(k-1)\, p\, (1-p)^{k-1} = \sum_{k=2}^{+\infty} k(k-1)\, p\, (1-p)^{k-1} = p\,(1-p) \sum_{k=2}^{+\infty} k(k-1)\, (1-p)^{k-2} \\
&= \frac{2p\,(1-p)}{\big(1-(1-p)\big)^3} = \frac{2(1-p)}{p^2}\ ; \\
\mathrm{Var}(X) &= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}.
\end{aligned}$$

La loi de Poisson est utilisée de façon fondamentale dans les problèmes de file d'attente : quand on veut modéliser par exemple le nombre de clients qui se présentent à un guichet dans un intervalle de temps donné, il est naturel d'utiliser une loi de Poisson.

Loi géométrique

Nous l'avons déjà rencontrée aussi. On considère une suite infinie de lancers indépendants d'une pièce qui tombe sur pile avec probabilité $p \in [0,1]$ à chaque lancer. On note X le numéro du premier lancer qui donne pile. Alors nous avons vu que X suit la loi géométrique de paramètre p.

♣ **Exercice 6.3.** Soit $X : \Omega \to \mathbb{N}^*$ une variable aléatoire. On dit que X est *sans mémoire* si

$$\forall k, l \in \mathbb{N}, \qquad \mathbb{P}(X > k + l \mid X > k) = \mathbb{P}(X > l).$$

1. Pourquoi dit-on d'une telle variable aléatoire qu'elle est sans mémoire ?
2. Vérifier qu'une variable aléatoire de loi géométrique est sans mémoire.
3. Soit $X : \Omega \to \mathbb{N}^*$ une variable aléatoire sans mémoire. Trouver une relation de récurrence pour la suite $\big(\mathbb{P}(X > n)\big)_{n \in \mathbb{N}^*}$, puis exprimer $\mathbb{P}(X > n)$ en fonction de n et de $\mathbb{P}(X > 1)$. En déduire que X suit une loi géométrique dont on donnera le paramètre.

6.4 Calculs pour les v.a. à densité classiques

6.4.1 Lois uniformes

• Soit X une variable aléatoire de densité $\forall x \in \mathbb{R} \quad f(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$.
X^2 est une variable aléatoire bornée, donc intégrable.

$$\mathbb{E}[X] = \int_{\mathbb{R}} xf(x)dx = \int_{\mathbb{R}} \frac{1}{b-a}x\mathbb{1}_{[a,b]}(x)dx = \frac{1}{b-a}\int_a^b xdx = \frac{1}{b-a}[x^2/2]_a^b = \frac{a+b}{2};$$
$$\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2f(x)dx = \int_{\mathbb{R}} \frac{1}{b-a}x^2\mathbb{1}_{[a,b]}(x)dx = \frac{1}{b-a}\int_a^b x^2dx = \frac{1}{b-a}[x^3/3]_a^b = \frac{a^2+ab+b^2}{3};$$
$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{a+b}{2} - \frac{a^2+ab+b^2}{3} = \frac{(b-a)^2}{12}.$$

Elle généralisent la loi uniforme sur $[0,1]$. Attention à ne pas les confondre avec les lois uniformes sur des ensembles finis.

6.4.2 Lois exponentielles

• Soit $\lambda > 0$, et soit X une variable aléatoire de loi exponentielle de paramètre λ : elle est de densité

$$\forall x \in \mathbb{R} \quad f(x) = \lambda\exp(-\lambda x)\mathbb{1}_{\mathbb{R}_+}(x).$$

Montrons que X^2 est intégrable. L'intégrale

$$\int_{\mathbb{R}} |x|^2f(x)dx = \int_{\mathbb{R}} x^2\lambda\exp(-\lambda x)\mathbb{1}_{\mathbb{R}_+}(x)dx = \int_0^{+\infty} x^2\lambda\exp(-\lambda x)dx$$

est l'intégrale, impropre en $+\infty$, d'une fonction positive.

$$\lim_{x\to+\infty} x^3\exp(-\lambda x) = 0$$

donc il existe $M > 0$ tel que $\forall x \geq M \quad 0 \leq x\exp(-\lambda x) \leq 1/x^2$.
Comme $\displaystyle\int_M^{+\infty} \frac{1}{x^2}dx < +\infty$, on en déduit que $\displaystyle\int_M^{+\infty} x\exp(-\lambda x)dx < +\infty$,
puis que $\displaystyle\int_0^{+\infty} x\lambda\exp(-\lambda x)dx < +\infty$, donc X^2 est intégrable.

$$\mathbb{E}[X] = \int_{\mathbb{R}} xf(x)dx = \int_{\mathbb{R}} x\lambda\exp(-\lambda x)\mathbb{1}_{\mathbb{R}_+}(x)dx = \int_0^{+\infty} x\lambda\exp(-\lambda x)dx;$$

soit M>0 $\displaystyle\int_0^M x\lambda\exp(-\lambda x)dx = [-x\exp(-\lambda x)]_0^M + \int_0^M \exp(-\lambda x)dx = -M\exp(-\lambda M) + [-\exp(-\lambda x)/\lambda]_0^M$

$$= -M\exp(-\lambda M) + \frac{1}{\lambda} - \frac{1}{\lambda}\exp(-\lambda M) \quad \longrightarrow \frac{1}{\lambda} \text{ quand } M \to +\infty,$$

donc $\mathbb{E}[X] = \dfrac{1}{\lambda}$;

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2f(x)dx = \int_{\mathbb{R}} x^2\lambda\exp(-\lambda x)\mathbb{1}_{\mathbb{R}_+}(x)dx = \int_0^{+\infty} x^2\lambda\exp(-\lambda x)dx;$$

soit M>0 $\displaystyle\int_0^M x^2\lambda\exp(-\lambda x)dx = [-x^2\exp(-\lambda x)]_0^M + \frac{2}{\lambda}\int_0^M x\lambda\exp(-\lambda x)dx \longrightarrow \frac{2}{\lambda}\mathbb{E}[X] = \frac{2}{\lambda^2}$ quand $M \to +\infty$,

donc $\mathbb{E}[X^2] = \dfrac{2}{\lambda^2}$;

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

♣ **Exercice 6.4.** (Propriété « sans mémoire ») Soit $X : \Omega \to \mathbb{R}_+$ une variable aléatoire. On dit que X est sans mémoire si et seulement si

$$\forall s, t \in \mathbb{R}_+,\ \mathbb{P}(X > s + t | X > s) = \mathbb{P}(X > t).$$

1. Pourquoi dit-on d'une telle variable aléatoire qu'elle est sans mémoire ?
2. Vérifier qu'une variable aléatoire de loi exponentielle est sans mémoire.

C'est à cause de cette propriété que la loi exponentielle est utilisée quand on veut modéliser la durée de vie d'un composant électronique, ou le temps entre les arrivées de deux clients successifs dans une file d'attente (il y a en particulier des liens trés forts entre les lois exponentielles et les lois de Poisson).

♣ **Exercice 6.5.** Soit $\lambda > 0$, et soit X une variable aléatoire de loi exponentielle de paramètre λ définie sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Calculer $G(t) = \mathbb{P}(X \geq t)$ pour $t \geq 0$.

6.4.3 Lois normales

• Soit X une variable aléatoire de loi normale de paramètres $(0, 1)$: elle est de densité

$$\forall x \in \mathbb{R} \quad f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Montrons que X^2 est intégrable. L'intégrale

$$\int_{\mathbb{R}} |x|^2 f(x) dx = \int_{\mathbb{R}} x^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 2 \int_0^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

est l'intégrale, impropre en $+\infty$, d'une fonction positive.

$$\lim_{x \to +\infty} x^4 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) = 0$$

donc il existe $M > 0$ tel que $\forall x \geq M \quad 0 \leq x^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \leq 1/x^2$.
Comme $\int_M^{+\infty} \frac{1}{x^2} dx < +\infty$, on en déduit que $\int_M^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx < +\infty$,
puis que $\int_0^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx < +\infty$, donc X^2 est intégrable.

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f(x) dx = \int_{\mathbb{R}} x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 0 \quad \text{car la fonction intégée est impaire;}$$

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2 f(x) dx = \int_{\mathbb{R}} x^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 2 \int_0^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx;$$

soit M>0 $$\int_0^M x^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = [-x \exp(-x^2/2)/\sqrt{2\pi}]_0^M + \int_0^M \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

$$\longrightarrow \frac{1}{2} \text{ quand } M \to +\infty,$$

$$\text{donc } \mathbb{E}[X^2] = 1;$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 1.$$

• Soit $m \in \mathbb{R}$, $\sigma^2 > 0$ et soit Y une variable aléatoire de loi normale de paramètres (m, σ^2) : elle est de densité

$$\forall x \in \mathbb{R} \quad g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

On peut reprendre les calculs précédents en faisant un changement de variable $y = (x - m)/\sigma$.

Mais on peut aussi procéder plus astucieusement : montrons que si X suit une loi gaussienne centrée réduite, alors $Y = m + \sigma X$ ($\sigma > 0$) suit une loi normale de paramètres (m, σ^2).
Notons Φ la primitive de $f : x \mapsto \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ dont la limite en $-\infty$ vaut 0 :

$$\forall t \in \mathbb{R} \quad \Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$

Comme f est de classe C^∞ sur $\mathbb{R}$, Φ est aussi de classe C^∞, et

$$\forall t \in \mathbb{R} \quad \Phi'(t) = f(t).$$

Remarquons que Φ est aussi la fonction de répartition de X. Calculons la fonction de répartition G de Y : soit $t \in \mathbb{R}$

$$G(t) = \mathbb{P}(Y \leq t) = \mathbb{P}(m + \sigma X \leq t) = \mathbb{P}\left(X \leq \frac{t-m}{\sigma}\right) = \Phi\left(\frac{t-m}{\sigma}\right).$$

On constate que $G : t \mapsto \mathbb{P}(Y \leq t)$ est de classe C^∞ sur $\mathbb{R}$ par composition de fonctions de classe C^∞ sur $\mathbb{R}$, et donc Y admet une densité g qu'on obtient en dérivant G :

$$\forall t \in \mathbb{R} \quad g(t) = G'(t) = \frac{1}{\sigma}\Phi'\left(\frac{t-m}{\sigma}\right) = \frac{1}{\sigma} f\left(\frac{t-m}{\sigma}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right).$$

Donc $Y = m + \sigma X$ suit une loi normale de paramètres (m, σ^2).
Il suffit alors d'utiliser les propriétés de l'espérance et de la variance :

$$\mathbb{E}[Y] = \mathbb{E}[m + \sigma X] = m + \sigma \mathbb{E}[X] = m;$$
$$\mathrm{Var}(Y) = \mathrm{Var}(m + \sigma X) = \sigma^2 \,\mathrm{Var}(X) = \sigma^2.$$

Ces lois apparaissent dans le théoréme central limite, que nous verrons plus tard. Elles sont souvent utilisées pour modéliser de petites erreurs ou variations aléatoires autour d'une quantité déterministe.
Il n'existe pas d'expression à l'aide de fonctions usuelles de la primitive de l'application $x \mapsto \exp(-x^2)$, mais on sait montrer que

$$I = \int_{\mathbb{R}} \exp(-x^2/2) dx = \sqrt{2\pi}.$$

♣ **Exercice 6.6.** Montrer que la densité de la loi $\mathcal{N}(m, \sigma^2)$ est bien d'intégrale 1 sur $\mathbb{R}$.
Indication. Faire un changement de variable dans l'intégrale pour se ramener au cas précédent.

Chapitre 7

Second TP. Autour des variables aléatoires

Rappels.

Scilab possède une fonction `rand()` dont les appels successifs fournissent une suite de nombres parfaitement déterministes, mais ayant pour un certain nombre de tests statistiques les propriétés d'une suite de variables aléatoires indépendantes et identiquement distribuées, de loi uniforme sur $[0, 1]$.

```
-> x=rand(1,10)
```

Dans x, nous avons un 10 échantillon de loi uniforme sur $[0, 1]$, c'est-à-dire la réalisation de 10 variables aléatoires iid de loi uniforme sur $[0, 1]$. Scilab nous permet de résumer un échantillon dans un histogramme :

```
-> x=rand(1,20);
-> c=[0:0.1:1]
-> histplot(c,x);
```

Le vecteur c détermine les extrémités des classes : ainsi, notre histogramme a 10 classes : $[0, 0.1]$, $[0.1, 0.2]$, ..., $[0.9, 1]$. La fonction `histplot` calcule la proportion d'éléments de x qui tombent dans chaque classe, et trace au dessus de chaque classe un bâton de hauteur égale à la proportion correspondante.

□ **TP 7.1. (lois exponentielles)**. On rappelle que la loi exponentielle de paramètre a est la loi de densité

$$\forall x \in \mathbb{R} \quad f(x) = a\exp(-ax)\mathbf{1}_{\mathbb{R}_+}(x).$$

Sa fonction de répartition est $\forall t \geq 0 \quad F(t) = \mathbb{P}(X \leq t) = \displaystyle\int_0^t a\exp(-ax)dx = 1 - \exp(-at)$.

Son espérance est $1/a$, sa variance $1/a^2$.

1. Tracer la densité de la loi exponentielle de paramètre 1 :

```
-> x = [0:0.1:10]';
-> y = exp(-x);
-> plot2d(x,y);
```

2. Tracer sur le même graphique les densités pour différentes valeurs du paramètre a.
3. Tracer sur le même graphique les fonctions de répartition pour différentes valeurs du paramètre a.
4. Utiliser `grand` pour simuler un 20-échantillon de loi exponentielle de paramètre 1 et le visualiser dans un histogramme :

```
-> x=grand(1,20,'exp',1);
-> c=[0:0.2:5]
-> histplot(c,x);
```

5. Tracer des histogrammes de n-échantillons, avec n grand, de lois exponentielles pour différents paramètres. Qu'observez-vous ?

6. Superposer aux histogrammes les densités correspondantes.
7. Calculer avec `mean` la moyenne de vos échantillons, ainsi que leur variance (Que fait `st deviation` ?) Comparer avec l'espérance et la variance de la loi correspondante.
8. Que font les lignes suivantes ?

```
-> x=rand(1,100);
-> c=[0:0.2:5];
-> histplot(c,-log(x));
```

9. Soit X une variable aléatoire de loi exponentielle de paramètre 1. Quelle est, pour $a > 0$, la loi de $Y = aX$? En calculant la fonction de répartition de Y, confirmez vos pronostics.
10. Soit U une variable aléatoire de loi uniforme sur $[0, 1]$. Je pose $X = -\frac{1}{a}\ln(U)$. Quelle est à votre avis la loi de X ? En calculant la fonction de répartition de X, confirmez vos pronostics.

□ **TP 7.2. (densité avec le sinus)**. On considère la densité (revoir l'exercice de TD)

$$\forall x \in \mathbb{R} \quad f(x) = \frac{1}{2}\sin(x)\mathbf{1}_{[0,\pi]}(x).$$

Sa fonction de répartition est $\forall t \in [0, \pi] \quad F(t) = \dfrac{1}{2}(1 - \cos(t))$.
Son espérance est $\pi/2$, sa variance ?.

1. Tracer sur un même graphique la densité et la fonction de répartition.
2. Vérifier que F est une bijection de $[0, \pi]$ dans $[0, 1]$, de réciproque $G(u) = \arccos(1 - 2u)$.
3. Soit U une variable aléatoire de loi uniforme sur $[0, 1]$. Je pose $X = \arccos(1-2U)$. En calculant la fonction de répartition de X, montrer que X a pour densité f.
4. Tracer l'histogramme d'un n-échantillons, avec n grand, de loi de densité f, et superposer sur le graphique la densité f.
5. Calculer avec `mean` la moyenne de vos échantillons, ainsi que leur variance. Comparer avec l'espérance et la variance de la loi correspondante.

□ **TP 7.3. (lois gaussiennes)**. On rappelle que la loi normale de paramètres (m, σ^2) est la loi de densité

$$\forall x \in \mathbb{R} \quad f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

1. Tracer sur le même graphique les densités pour différentes valeurs du paramètre m en fixant $\sigma^2 = 1$. Commenter.
2. Tracer sur le même graphique les densités pour différentes valeurs du paramètre σ^2 en fixant $m = 0$. Commenter.
3. Tracer des histogrammes de n-échantillons, avec n grand, de lois gaussiennes pour différents paramètres. Superposer aux histogrammes les densités correspondantes.
4. Calculer avec `mean` la moyenne de vos échantillons, ainsi que leur variance (Que fait `st deviation` ?) Comparer avec l'espérance et la variance de la loi correspondante.

□ **TP 7.4. (lois binomiales)**. On rappelle que la loi binomiale de paramètres (n, p) est la suivante

$$\forall k \in [0, \dots, n] \quad \mathbb{P}(X = k) = a_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

1. Que font les commandes suivantes ?

```
-> n=12;p=0.3; clf(); plot2d(0:n,binomial(p,n),-1);
```

2. Tracer sur un graphique, pour $n = 12$ et différentes valeurs de p, les points de coordonnées $(k, a_{n,p}(k)$.
3. Rappeler le lien entre le jeu de pile ou face et la loi binômiale.
4. Que font les commandes suivantes ?

```
-> n=12;p=0.3; y=sum(rand(1,12)<p);
```

5. Simuler 1000 fois le lancer de 12 pièces tombant sur pile avec probabilité p, puis compter, pour ces 1000 essais, le nombre de pile obtenus sur ces 12 pièces. Tracer un histrogramme de ces 1000 nombres. Commenter.

□ **TP 7.5. (lois géométriques)**. On rappelle que la loi géométrique de paramètre p est la suivante

$$\forall k \in \mathbb{N}^* \quad \mathbb{P}(X = k) = a_p(k) = (1 - p)^{k-1}p.$$

1. Que font les commandes suivantes ?

```
-> n=15;p=0.3; x=p*cumprod([1,(1-p)*ones(1,n)]);
```

2. Tracer sur un graphique, pour $n = 15$ et différentes valeurs de p, les points de coordonnées $(k, a_{n,p}(k)$ pour k variant de 1 à 16.
3. Rappeler le lien entre le jeu de pile ou face et la loi géométrique.
4. Que font les lignes suivantes :

```
-> p=0.3;i = 0;
-> while rand()<p, (i=i+1); end;
```

5. Simuler 1000 fois le nombre de lancers d'une pièce tombant sur pile avec probabilité p nécessaires à obtenir un premier pile. Tracer un histrogramme de ces 1000 nombres. Commenter.

Chapitre 8

Vecteur aléatoire discret et indépendance entre variables aléatoires discrètes

Sommaire

8.1 Vecteur aléatoire de $\mathbb{R}^2$ 53
8.2 Calcul des lois marginales 54
8.3 Théorème de transfert pour un vecteur aléatoire discret 54
8.4 Espérance et covariance 55
8.5 Indépendance de deux variables aléatoires : définitions et critères 57
8.6 Indépendance et covariance 58
8.7 Indépendance de plusieurs variables aléatoires discrètes 58
8.8 Loi d'une somme de variables aléatoire indépendantes 59
8.8.1 Sommes de variables de Bernoulli indépendantes 59
8.8.2 Sommes de variables de Poisson indépendantes 60

Objectifs :

- Étendre la notion de variable aléatoire à celle de vecteur aléatoire.
- Étendre la notion d'indépendance introduite pour les événements aux variables aléatoires.

Mots-clés :

- Vecteur aléatoire discret, loi d'un vecteur aléatoire discret, lois marginales, loi jointe.
- Covariance, coefficient de corrélation linéaire, matrice de covariance.
- Indépendance.

Outils :

- Théorème de transfert.
- Différents critères d'indépendance.
- Variance d'une somme de variables aléatoires indépendantes, covariance de variables aléatoires indépendantes.

Nous allons voir dans ce chapitre que presque tout ce que l'on a construit pour une variable aléatoire réelle se transpose aisément au cas d'un vecteur aléatoire. Les principales différences sont
– la notion de loi marginale
– la notion de covariance, qui essaie de rendre compte du lien entre les coordonnées d'un vecteur aléatoire.
On ne considérera cette année que des vecteurs aléatoires discrets, et à valeur dans $\mathbb{R}^2$.

8.1 Vecteur aléatoire de $\mathbb{R}^2$

Définition 8.1 *1. Un* vecteur aléatoire *à valeurs dans $\mathbb{R}^2$ sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ est une application Z de Ω dans $\mathbb{R}^2$:*

$$Z : \omega \in \Omega \mapsto Z(\omega) = (Z_1(\omega), Z_2(\omega)) \in \mathbb{R}^2.$$

2. Soit Z un vecteur aléatoire défini sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $\mathbb{R}^2$. La loi $\mathbb{P}_Z$ de $Z = (Z_1, Z_2)$ est une probabilité sur l'ensemble $Z(\Omega) \subset \mathbb{R}^2$ des valeurs prises par Z. La loi de la v.a. Z_i est appelée i-ème loi marginale *de Z. La loi $\mathbb{P}_Z$ est appelée* loi jointe *des v.a. (Z_1, Z_2).*

$\triangleright$ **Exemple.** On lance deux dés équilibrés et on note X et Y les résultats obtenus. Alors (X, Y), (X, X), ou encore $(X, X+Y)$ sont des vecteurs aléatoires discrets à valeurs dans $\mathbb{R}^2$.
3. Soit Z un vecteur aléatoire défini sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $\mathbb{R}^n$. On dit que Z est un vecteur aléatoire discret si l'ensemble $Z(\Omega)$ des valeurs prises par Z est fini ou dénombrable. Dans ce cas, pour la décrire la loi $\mathbb{P}_Z$ de Z, il suffit de donner :
– l'ensemble $Z(\Omega)$ des valeurs prises par Z,
– pour tout $z = (z_1, \ldots, z_n) \in Z(\Omega)$, le nombre

$$\mathbb{P}_Z\big(\{z\}\big) = \mathbb{P}(Z = z) = \mathbb{P}\big(\{Z_1 = z_1\} \cap \cdots \cap \{Z_n = z_n\}\big) = \mathbb{P}\big(Z_1 = z_1 \text{ et } \ldots \text{ et } Z_n = z_n\big).$$

$\triangleright$ **Exemple.** Présentation de la loi d'un vecteur aléatoire discret à valeurs dans $\mathbb{R}^2$ sous forme de tableau :

$Y \backslash X$	1	2	3	loi de Y
0	$0,1$	$0,05$	$0,4$	$0,55$
1	$0,15$	$0,3$	0	$0,45$
loi de X	$0,25$	$0,35$	$0,4$	1

La loi jointe de X et Y est $0,1\,\delta_{(1,0)} + 0,05\,\delta_{(2,0)} + 0,4\,\delta_{(3,0)} + 0,15\,\delta_{(1,1)} + 0,3\,\delta_{(2,1)}$.
La première loi marginale de (X, Y) est la loi de X, à savoir $0,25\,\delta_1 + 0,35\,\delta_2 + 0,4\,\delta_3$.
La seconde loi marginale de (X, Y) est la loi de Y, à savoir $\mathcal{B}(0,45)$.

♣ **Exercice 8.1.** On lance deux dés équilibrés et on note X et Y les résultats obtenus. On définit les vecteurs aléatoires $Z_1 = (X, Y)$, $Z_2 = (X, X)$ et $Z_3 = (X, X+Y)$. Déterminer les lois de ces trois vecteurs, et leurs lois marginales. On remarquera que les vecteurs aléatoires Z_1 et Z_2 ont les mêmes lois marginales sans qu'ils aient la même loi pour autant.

♠ **Attention !** Quand on connaît la loi d'un vecteur aléatoire, on peut en déduire ses lois marginales, mais la réciproque est fausse. La loi du vecteur comporte plus d'information, au sens où elle indique comment les coordonnées sont liées entre elles.

$\triangleright$ **Exemple.** On lance 5 fois de suite une pièce qui tombe sur pile avec probabilité $p \in [0, 1]$, on note X (resp. Y) le nombre de pile (resp. le nombre de face) et on définit $Z = (X, Y)$. Alors

$$Z(\Omega) = \big\{(k, 5-k) \ : \ k \in \{0, \ldots, 5\}\big\} \quad \text{et} \quad \forall k \in \{0, \ldots, 5\}, \quad \mathbb{P}\big(Z = (k, 5-k)\big) = \binom{5}{k} p^k (1-p)^{5-k}.$$

Calculons les lois marginales de Z :

$$X(\Omega) = \{0, \ldots, 5\} \quad \text{et} \quad \forall k \in \{0, \ldots, 5\}, \quad \mathbb{P}(X = k) = \binom{5}{k} p^k (1-p)^{5-k} \quad \text{donc } X \sim \mathcal{B}(5, p),$$
$$Y(\Omega) = \{0, \ldots, 5\} \quad \text{et} \quad \forall k \in \{0, \ldots, 5\}, \quad \mathbb{P}(Y = k) = \binom{5}{k} (1-p)^k p^{5-k} \quad \text{donc } Y \sim \mathcal{B}(5, 1-p).$$

Remarque. Si on prend, dans l'exemple précédent, $p = 1/2$, on voit que X et Y ont la même loi, et donc les vecteurs (X, Y) et (X, X) ont les mêmes lois marginales et pourtant les vecteurs n'ont pas la même loi car par exemple $\mathbb{P}\big((X, Y) = (1, 1)\big) = 0 \neq \mathbb{P}\big((X, X) = (1, 1)\big)$.

8.2 Calcul des lois marginales

Proposition 8.2 *Soit $Z = (X, Y)$ un vecteur aléatoire discret à valeurs dans $\mathbb{R}^2$. Alors*
- $X(\Omega) = \big\{x \in \mathbb{R} \ : \ \exists y \in \mathbb{R}$ *tel que* $(x, y) \in Z(\Omega)\big\}$ *;*
- $Y(\Omega) = \big\{y \in \mathbb{R} \ : \ \exists x \in \mathbb{R}$ *tel que* $(x, y) \in Z(\Omega)\big\}$ *;*
- *pour tout* $x \in X(\Omega)$, $\mathbb{P}(X = x) = \displaystyle\sum_{y \in Y(\Omega)} \mathbb{P}(X = x \text{ et } Y = y)$ *;*
- *pour tout* $y \in Y(\Omega)$, $\mathbb{P}(Y = y) = \displaystyle\sum_{x \in X(\Omega)} \mathbb{P}(X = x \text{ et } Y = y)$.

Démonstration. Il s'agit de la formule des probabilités totales : la famille $(\{Y = y\})_{y \in Y(\Omega)}$ est un système complet. Donc

$$\begin{aligned}\mathbb{P}(X = x) = \mathbb{P}\big(\{X = x\} \cap \Omega\big) &= \mathbb{P}\left(\{X = x\} \cap \bigsqcup_{y \in Y(\Omega)} \{Y = y\}\right) = \mathbb{P}\left(\bigsqcup_{y \in Y(\Omega)} (\{X = x\} \cap \{Y = y\})\right) \\ &= \sum_{y \in Y(\Omega)} \mathbb{P}\big(\{X = x\} \cap \{Y = y\}\big) = \sum_{y \in Y(\Omega)} \mathbb{P}(X = x \text{ et } Y = y).\end{aligned}$$

La seconde formule s'obtient de façon similaire. □

◊ **En pratique.** Plus généralement, soit $Z : \Omega \to \mathbb{R}^n$ un vecteur aléatoire discret. On note $Z_1, \ldots, Z_n$ ses coordonnées. On calcule la i-ème loi marginale de la façon suivante :

$$Z_i(\Omega) = \big\{z \in \mathbb{R} \ : \ \exists (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \in \mathbb{R}^{n-1} \text{ tel que } (x_1, \ldots, x_{i-1}, z, x_{i+1}, \ldots, x_n) \in Z(\Omega)\big\}$$
$$\text{et} \quad \forall z \in Z_i(\Omega), \quad \mathbb{P}(Z_i = z) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} \mathbb{P}\big(Z = (x_1, \ldots, x_{i-1}, z, x_{i+1}, \ldots, x_n)\big)$$

où la somme est sur l'ensemble des $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ tels que $(x_1, \ldots, x_{i-1}, z, x_{i+1}, \ldots, x_n) \in Z(\Omega)$.

8.3 Théorème de transfert pour un vecteur aléatoire discret

Comme dans le cas des variables aléatoires réelles, si $Z : \Omega \to \mathbb{R}^n$ est un vecteur aléatoire discret, le théorème de transfert permet de calculer l'espérance d'une fonction de Z.

Théorème 8.3 *(théorème de transfert)*
Soit $Z = (Z_1, Z_2, \ldots, Z_n) : \Omega \to \mathbb{R}^n$ un vecteur aléatoire discret et $h : \mathbb{R}^n \to \mathbb{R}$ une application telle que $h(Z)$ soit une v.a. Alors $h(Z)$ est intégrable si et seulement si $\displaystyle\sum_{z \in Z(\Omega)} \mathbb{P}(Z = z) \, |h(z)| < +\infty$.
Dans ce cas,

$$\mathbb{E}\big[h(Z)\big] = \sum_{z \in Z(\Omega)} \mathbb{P}(Z = z) \, h(z).$$

Démonstration. Même démonstration que dans le cas des variables aléatoires à valeurs réelles. $\square$

♣ **Exercice 8.2.** On lance deux dés équilibrés et on note X et Y les résultats obtenus. On définit les vecteurs aléatoires $Z_1 = (X, Y)$, $Z_2 = (X, X)$ et $Z_3 = (X, X + Y)$. Calculer $\mathbb{E}\big[X(X + Y)\big]$ de deux manières, avec la loi du vecteur Z_1 et avec celle du vecteur Z_3.

8.4 Espérance et covariance

Définition 8.4 *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $X = (X_1, X_2, \ldots, X_n) : \Omega \to \mathbb{R}^n$ un vecteur aléatoire discret. Lorsque tous les X_i sont intégrables, on définit l'*espérance *du vecteur aléatoire X par*

$$\mathbb{E}[X] = \big(\mathbb{E}[X_1], \mathbb{E}[X_2], \ldots, \mathbb{E}[X_n]\big) \in \mathbb{R}^n.$$

Remarque. L'espérance $\mathbb{E}[X]$ du vecteur aléatoire X dépend de la loi de X uniquement au travers des lois marginales.

Proposition 8.5 *(inégalité de Cauchy-Schwarz)*
Soient X et Y deux v.a. réelles définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Si X^2 et Y^2 sont intégrables, alors XY est intégrable et

$$\big|\mathbb{E}[XY]\big| \leq \sqrt{\mathbb{E}[X^2]\,\mathbb{E}[Y^2]}.$$

De plus, l'égalité $\big|\mathbb{E}[XY]\big| = \sqrt{\mathbb{E}[X^2]\,\mathbb{E}[Y^2]}$ est vérifiée si et seulement s'il existe $(\lambda, \mu) \in \mathbb{R}^2 \backslash \{(0, 0)\}$ tel que $\lambda X + \mu Y = 0$ presque sûrement.

Démonstration. • Supposons que X^2 et Y^2 sont intégrables. En écrivant que

$$0 \leq \big(|X| - |Y|\big)^2 = X^2 + Y^2 - 2|XY|, \quad \text{on voit que} \quad 2|XY| \leq X^2 + Y^2$$

et la v.a. XY est donc intégrable.
• Si $\mathbb{E}\big[Y^2\big] = 0$, alors Y^2 est une v.a. positive d'espérance nulle, donc $Y = 0$ presque sûrement. Ainsi $XY = 0$ presque sûrement et donc $\mathbb{E}[XY] = 0$. L'inégalité est vraie dans ce cas.
• Si $\mathbb{E}\big[Y^2\big] \neq 0$, on considère la fonction $f : u \in \mathbb{R} \mapsto \mathbb{E}\big[(X + uY)^2\big]$. En développant et en utilisant la linéarité de l'espérance, on obtient

$$f(u) = \mathbb{E}\big[(X + uY)^2\big] = \mathbb{E}\big[X^2\big] + 2u\,\mathbb{E}[XY] + u^2\,\mathbb{E}\big[Y^2\big].$$

f est un polynôme de degré 2 en u, qui est toujours positif ou nul. Son discriminant est donc négatif ou nul :

$$(2\mathbb{E}[XY])^2 - 4\,\mathbb{E}\big[X^2\big]\mathbb{E}\big[Y^2\big] \leq 0 \;\Leftrightarrow\; (\mathbb{E}[XY])^2 \leq \mathbb{E}\big[X^2\big]\mathbb{E}\big[Y^2\big] \;\Leftrightarrow\; \big|\mathbb{E}[XY]\big| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}.$$

• Cas d'égalité. Si $\mathbb{E}\big[Y^2\big] = 0$, alors $\mathbb{P}(Y = 0) = 1$ et donc $\lambda = 0$ et $\mu = 1$ conviennent. Si $\mathbb{E}\big[Y^2\big] \neq 0$ et $\big|\mathbb{E}[XY]\big| = \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$, alors le discriminant précédent est nul et f possède donc une racine double u_0. On a donc $f(u_0) = 0$, c'est-à-dire $\mathbb{E}\big[(X + u_0Y)^2\big] = 0$ et par suite $X + u_0Y = 0$ presque sûrement. $\square$

Cette proposition rend légitime la définition suivante.

Définition 8.6 *Soient X et Y deux v.a. réelles définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et de carré intégrable. On définit la* covariance *entre X et Y par*

$$\operatorname{Cov}(X,Y) = \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y].$$

On suppose de plus que X et Y ne sont pas constantes presque sûrement (ce qui équivaut à $\operatorname{Var}(X)\ \operatorname{Var}(Y) > 0$ d'après la proposition 6.8 page 42). Le coefficient de corrélation linéaire *entre X et Y est défini par*

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

♣ **Exercice 8.3.** Vérifiez que les deux définitions de la covariance sont équivalentes en utilisant la linéarité de l'espérance.

Remarque. Souvent, pour les applications théoriques (montrer des propositions), la première expression est plus intéressante alors que pour les applications pratiques (faire des calculs), la seconde est plus utile. Dans ce dernier cas, la seule quantité nouvelle à calculer est, en utilisant le théorème de transfert,

$$\mathbb{E}[XY] = \sum_{(x,y)\in(X,Y)(\Omega)} x\,y\,\mathbb{P}\big((X,Y) = (x,y)\big).$$

Proposition 8.7 *Soient X et Y deux v.a. réelles définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et de carré intégrable. Alors*

1. $\operatorname{Cov}(X,X) = \operatorname{Var}(X)$.
2. $|\operatorname{Cov}(X,Y)| \leq \sqrt{\operatorname{Var}(X)\ \operatorname{Var}(Y)}$.
3. $\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\ \operatorname{Cov}(X,Y)$.

On suppose de plus que X et Y ne sont pas constantes presque sûrement.

4. $|\rho_{X,Y}| \leq 1$. *De plus, $\rho_{(X,Y)} = 1$ si et seulement s'il existe $a > 0$ et $b \in \mathbb{R}$ tels que $Y = aX + b$ presque sûrement et $\rho_{(X,Y)} = -1$ si et seulement s'il existe $a < 0$ et $b \in \mathbb{R}$ tels que $Y = aX + b$ presque sûrement.*

Démonstration. Posons $U = X - \mathbb{E}[X]$ et $V = Y - \mathbb{E}[Y]$. On a alors $\operatorname{Cov}(X,Y) = \mathbb{E}[UV]$, $\operatorname{Var}(X) = \mathbb{E}\big[U^2\big]$ et $\operatorname{Var}(Y) = \mathbb{E}\big[V^2\big]$.

1. Si $Y = X$, $V = U$ et on a immédiatement l'égalité voulue.

2. Il s'agit de l'inégalité de Cauchy-Schwartz appliquée à U et V.

3. On développe $\operatorname{Var}(X+Y) = \mathbb{E}\big[(U+V)^2\big]$ (rappelez-vous la proposition 6.8 page 42).

4. La première partie découle immédiatement du point 2. Ensuite, $|\rho_{X,Y}| = 1$ est le cas d'égalité dans l'inégalité de Cauchy-Schwartz. Dans ce cas, il existe λ et μ tels que $\lambda U + \mu V = 0$ presque sûrement. Comme par ailleurs X et Y ne sont pas constantes presque sûrement, c'est-à-dire U et V ne sont pas nulles presque sûrement, on déduit que $\lambda \neq 0$ et $\mu \neq 0$. Ainsi

$$Y = -\frac{\lambda}{\mu}X + \frac{\lambda}{\mu}\,\mathbb{E}[X] + \mathbb{E}[Y] = aX + b \quad \text{avec } a = -\frac{\lambda}{\mu} \neq 0 \text{ et } b = \frac{\lambda}{\mu}\,\mathbb{E}[X] + \mathbb{E}[Y].$$

Ensuite, on observe que

$$\rho_{X,aX+b} = \frac{\operatorname{Cov}(X,aX+b)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(aX+b)}} = \frac{\mathbb{E}\big[(X - \mathbb{E}[X])(aX + b - a\mathbb{E}[X] - b)\big]}{\sqrt{\operatorname{Var}(X)\,a^2\operatorname{Var}(X)}} = \frac{a}{|a|} = \begin{cases} 1 & \text{si } a > 0, \\ -1 & \text{si } a < 0. \end{cases}$$

Ainsi, si $Y = aX + b$ presque sûrement, le calcul précédent montre que $\rho_{X,Y} = 1$ si $a > 0$ et $\rho_{X,Y} = -1$ si $a < 0$. Réciproquement, si $\rho_{X,Y} = 1$, on a vu que $Y = aX + b$ presque sûrement avec $a \neq 0$ et le calcul précédent montre que a ne peut pas être strictement négatif, donc $a > 0$. De même, si $\rho_{X,Y} = -1$, on conclut que $Y = aX + b$ presque sûrement avec $a < 0$. $\square$

Proposition 8.8 *La covariance est une forme bilinéaire symétrique positive sur l'espace vectoriel des v.a. de carré intégrable : si X, Y, Z sont des v.a. de carré intégrable, alors*

1. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$,
2. $\mathrm{Cov}(X+Y,Z) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z)$,
3. $\mathrm{Cov}(\lambda X,Y) = \lambda\,\mathrm{Cov}(X,Y)$ *pour tout réel* λ,
4. $\mathrm{Cov}(X,X) \geq 0$.

♣ **Exercice 8.4.** Montrez ces propriétés.

Remarque. En revanche, la covariance n'est pas un produit scalaire car elle n'est pas définie positive. En effet, $\mathrm{Cov}(X,a) = 0$ quel que soit $a \in \mathbb{R}$.

Définition 8.9 *Soit $X = (X_1, X_2, \ldots, X_n) : \Omega \to \mathbb{R}^n$ un vecteur aléatoire discret. On définit, lorsque cette quantité existe, la* matrice de covariance *de X par*

$$\mathbb{K}_X = \big(\mathrm{Cov}(X_i,X_j)\big)_{1\leq i,j\leq n}.$$

Remarque. Une condition suffisante d'existence de la matrice de covariance d'un vecteur aléatoire est l' intégrabilité des carrés de ses composantes ; cela découle de l'inégalité de Cauchy-Schwarz.

Proposition 8.10 *La matrice de covariance d'un vecteur aléatoire discret est symétrique et positive, c'est-à-dire que pour tout vecteur $(\lambda_1,\ldots,\lambda_n) \in \mathbb{R}^n$, le nombre $\displaystyle\sum_{i=1}^n\sum_{j=1}^n \lambda_i\,\mathrm{Cov}(X_i,X_j)\lambda_j \geq 0$.*

Démonstration. La matrice de covariance d'un vecteur aléatoire discret est symétrique d'après le point 1. de la proposition 8.8. Ensuite, notons que le nombre $\displaystyle\sum_{i=1}^n\sum_{j=1}^n \lambda_i\,\mathrm{Cov}(X_i,X_j)\lambda_j$ est la variance de la v.a. $\displaystyle\sum_{i=1}^n \lambda_i X_i$, il est donc positif. □

♣ **Exercice 8.5.** X et Y désignent les résultats des lancers de deux dés équilibrés. Donner les matrices de covariance des vecteurs aléatoires $Z_1 = (X,Y)$, $Z_2 = (X, X+Y, X-Y)$ et $Z_3 = (X, X+Y)$.

8.5 Indépendance de deux variables aléatoires : définitions et critères

Définition 8.11 *Deux v.a.* discrètes *X et Y définies sur un espace de probabilité $(\Omega,\mathcal{F},\mathbb{P})$ sont* indépendantes *si pour tout $x \in X(\Omega)$, pour tout $y \in Y(\Omega)$, les événements $\{X=x\}$ et $\{Y=y\}$ sont indépendants, c'est-à-dire*

$$\forall x \in X(\Omega), \quad \forall y \in Y(\Omega), \quad \mathbb{P}\big(\{X=x\}\cap\{Y=y\}\big) = \mathbb{P}(X=x)\,\mathbb{P}(Y=y). \tag{8.1}$$

L'idée est la suivante : quand X et Y sont indépendantes, connaître la valeur prise par Y ne modifie pas la loi de X. La loi de X sous $\mathbb{P}$ et la loi de X sous $\mathbb{P}_{\{Y=y\}}$ (probabilité sachant l'événement $Y=y$) coïncident quel que soit y du support de la loi de Y :

$$\mathbb{P}_{\{Y=y\}}(X=x) = \frac{\mathbb{P}\big(\{X=x\}\cap\{Y=y\}\big)}{\mathbb{P}(Y=y)} = \frac{\mathbb{P}(X=x)\,\mathbb{P}(Y=y)}{\mathbb{P}(Y=y)} = \mathbb{P}(X=x).$$

♣ **Exercice 8.6.** Montrer que deux événements A et B sont indépendants ssi les v.a. $\mathbb{1}_A$ et $\mathbb{1}_B$ sont indépendantes.

Proposition 8.12 *Soient X et Y deux v.a. discrètes. Les propositions suivantes sont équivalentes :*
- *X et Y sont indépendantes ;*
- $\forall A \subset X(\Omega), \quad \forall B \subset Y(\Omega), \quad \mathbb{P}\big(\{X \in A\} \cap \{Y \in B\}\big) = \mathbb{P}(X \in A)\, \mathbb{P}(Y \in B)$ *;*
- *$\forall f, g : \mathbb{R} \to \mathbb{R}$ telles que $f(X)$ et $g(Y)$ soient des v.a., $\quad f(X)$ et $g(Y)$ sont indépendantes ;*
- *$\forall f, g : \mathbb{R} \to \mathbb{R}$ telles que $f(X)$ et $g(Y)$ soient des v.a. intégrables,* $\quad \mathbb{E}[f(X)\, g(Y)] = \mathbb{E}[f(X)]\, \mathbb{E}[g(Y)]$ *;*

8.6 Indépendance et covariance

Proposition 8.13 *Soient X et Y deux v.a. réelles discrètes définies sur un espace de probabilité* $(\Omega, \mathcal{F}, \mathbb{P})$.

1. *Si X et Y sont indépendantes et intégrables, alors XY est intégrable et*
$$\mathbb{E}[XY] = \mathbb{E}[X]\, \mathbb{E}[Y].$$
2. *Si X et Y sont indépendantes et intégrables, alors* $\mathrm{Cov}(X, Y) = 0$.
3. *Si X et Y sont indépendantes et de carré intégrable, alors $X + Y$ est de carré intégrable et*
$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Démonstration. 1. On applique (**??**) aux fonctions $f : x \mapsto x$ et $g : y \mapsto y$.
2. $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\, \mathbb{E}[Y] = 0$ d'après 1.
3. On a vu que $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\, \mathrm{Cov}(X, Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ d'après 2. $\square$

♠ **Attention !** La réciproque de 2. est fausse ! Regardons le vecteur $Z = (X, Y)$ suivant : Z suit la loi uniforme sur $Z(\Omega) = \{(1, 0), (-1, 0), (0, 1), (0, -1)\}$, autrement dit, la probabilité que Z soit égal à l'un quelconque de ces points est $1/4$. Alors X et Y ont même loi, donnée par $X(\Omega) = \{-1, 0, 1\}$ et
$$\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = \frac{1}{4} \quad \text{et} \quad \mathbb{P}(X = 0) = \frac{1}{2}.$$
On voit donc facilement que $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. On remarque aussi que $XY = 0$ (il y a toujours une des deux coordonnées qui est nulle), donc $\mathrm{Cov}(X, Y) = 0$.
Cependant $\mathbb{P}(X = 0 \text{ et } Y = 0) = 0$ alors que $\mathbb{P}(X = 0)\, \mathbb{P}(Y = 0) = 1/4$, donc X et Y ne sont pas indépendantes.

8.7 Indépendance de plusieurs variables aléatoires discrètes

Définition 8.14 *Soit $(X_i)_{i \in I}$ une famille de variables aléatoires réelles définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que les $(X_i)_{i \in I}$ sont indépendantes si pour toute famille $(x_i)_{i \in I}$ de nombres réels, les événements $X_i = x_i$, $i \in I$, sont indépendants, c'est-à-dire si pour tout sous-ensemble non vide $\{j_1, \ldots, j_p\}$ de I et tous $x_1 \in X_{j_1}(\Omega)$, $\ldots$, $x_p \in X_{j_p}(\Omega)$, on a*
$$\mathbb{P}\big(X_{j_1} = x_1 \text{ et } X_{j_2} = x_2 \text{ et } \ldots \text{ et } X_{j_p} = x_p\big) = \mathbb{P}\big(X_{j_1} = x_1\big)\, \mathbb{P}\big(X_{j_2} = x_2\big) \ldots \mathbb{P}\big(X_{j_p} = x_p\big).$$

Proposition 8.15 *Soit $(X_1, \ldots, X_n)$ des variables aléatoires* indépendantes *admettant un moment d'ordre* 2. *Alors*
$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n).$$

Démonstration. Montrons un résultat plus général : si X_1, $\ldots$, X_n sont des v.a. de carrés intégrables, alors $\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) + 2 \displaystyle\sum_{1 \le i < j \le n} \mathrm{Cov}(X_i, X_j)$.

Cela entraînera le résultat voulu puisque, d'après la proposition 8.13, dans le cas d'un échantillon, dès que $i \neq j$, l'indépendance de X_i et X_j entraîne que $\mathrm{Cov}(X_i, X_j) = 0$.

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^n X_i\right) &= \mathbb{E}\left[\left(\sum_{i=1}^n X_i - \mathbb{E}\left[\sum_{i=1}^n X_i\right]\right)^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^n \big(X_i - \mathbb{E}[X_i]\big)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^n \sum_{j=1}^n \big(X_i - \mathbb{E}[X_i]\big)\big(X_j - \mathbb{E}[X_j]\big)\right] \\
&= \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}\big[\big(X_i - \mathbb{E}[X_i]\big)\big(X_j - \mathbb{E}[X_j]\big)\big] = \sum_{i=1}^n \sum_{j=1}^n \mathrm{Cov}(X_i, X_j).
\end{aligned}$$

En distinguant les termes diagonaux, cette somme peut s'écrire,

$$\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) + \sum_{1 \leq i \neq j \leq n} \mathrm{Cov}(X_i, X_j)$$

ou encore

$$\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j)$$

car pour tout $i \neq j$, la paire $\{i, j\}$ correspond au couples (i, j) et (j, i) et est donc comptée deux fois dans la somme du haut et une seule fois dans la somme du bas. □

8.8 Loi d'une somme de variables aléatoire indépendantes

8.8.1 Sommes de variables de Bernoulli indépendantes

Proposition 8.16 *Si X_1, X_2, ..., X_n sont n v.a. i.i.d. de loi $\mathcal{B}(p)$ alors $X_1 + X_2 + \cdots + X_n \sim \mathcal{B}(n, p)$.*

Démonstration. Soit $k \in \mathbb{N}$. Remarquons que $X_1 + X_2 + \cdots + X_n = k$ est équivalent à exactement k des X_i valent 1. Notons $\mathcal{C}_k = \big\{(x_1, \ldots, x_n) \in \{0,1\}^n \ : \ x_1 + \cdots + x_n = k\big\}$ l'ensemble des configurations où exactement k des x_i valent 1. Alors pour $(x_1, \ldots, x_n) \in \mathcal{C}_k$, on a par indépendance des X_i,

$$\mathbb{P}\big(X_1 = x_1 \text{ et } \ldots \text{ et } X_n = x_n\big) = \mathbb{P}\big(X_1 = x_1\big) \ldots \mathbb{P}\big(X_n = x_n\big) = p^k\,(1-p)^{n-k}$$

car k termes de ce produit valent p et $n - k$ termes valent $1 - p$. On a alors, en partitionnant

$$\begin{aligned}
\mathbb{P}(X_1 + X_2 + \cdots + X_n = k) &= \sum_{(x_1, \ldots, x_n) \in \mathcal{C}_k} \mathbb{P}\big(X_1 = x_1 \text{ et } \ldots \text{ et } X_n = x_n\big) \\
&= \sum_{(x_1, \ldots, x_n) \in \mathcal{C}_k} p^k\,(1-p)^{n-k} = \binom{n}{k} p^k\,(1-p)^{n-k}
\end{aligned}$$

car $|\mathcal{C}_k| = \dbinom{n}{k}$. Cela montre que $X_1 + X_2 + \cdots + X_n \sim \mathcal{B}(n, p)$. □

Proposition 8.17 *Si $X \sim \mathcal{B}(n, p)$ et $Y \sim \mathcal{B}(n', p)$ sont indépendantes, alors $X + Y \sim \mathcal{B}(n + n', p)$.*

Démonstration. Soit X_1, ..., X_n, Y_1, ..., $Y_{n'}$ des v.a. i.i.d. de loi $\mathcal{B}(p)$. D'après la proposition précédente, X a la même loi que $X_1 + X_2 + \cdots + X_n$ et Y a la même loi que $Y_1 + Y_2 + \cdots + Y_{n'}$. Ainsi, la v.a. $X + Y$ a la même loi que la somme de $n + n'$ v.a. i.i.d. de loi $\mathcal{B}(p)$, c'est donc une variable binomiale de paramètres $n + n'$ et p d'après la proposition précédente. □

8.8.2 Sommes de variables de Poisson indépendantes

Les lois de Poisson possèdent une propriété remarquable de stabilité par somme :

Proposition 8.18 *Soit $X_1, \dots, X_n$ des v.a.* indépendantes *de loi de Poisson de paramètres $\lambda_1, \dots, \lambda_n$. Alors*

$$X_1 + \dots + X_n \sim \mathcal{P}(\lambda_1 + \dots + \lambda_n).$$

Démonstration. Il suffit de montrer la propriété pour $n = 2$. En effet, si cela a été montré, le résultat général s'ensuit par récurrence. Supposons donc $n = 2$ et soit $k \in \mathbb{N}$. On décompose l'événement $\{X_1 + X_2 = k\}$ selon les valeurs de X_1 :

$$\begin{aligned}
\mathbb{P}(X_1 + X_2 = k) &= \sum_{l=0}^{k} \mathbb{P}(X_1 = l \text{ et } X_2 = k - l) = \sum_{l=0}^{k} \mathbb{P}(X_1 = l)\, \mathbb{P}(X_2 = k - l) \qquad \text{par indépendance} \\
&= \sum_{l=0}^{k} e^{-\lambda_1} \frac{\lambda_1^l}{l!}\, e^{-\lambda_2} \frac{\lambda_2^{k-l}}{(k-l)!} = \frac{e^{-\lambda_1-\lambda_2}}{k!} \sum_{l=0}^{k} \frac{k!}{l!(k-l)!} \lambda_1^l \lambda_2^{k-l} = e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1 + \lambda_2)^k}{k!}.
\end{aligned}$$

Si par ailleurs $k < 0$, on a $\mathbb{P}(X_1 + X_2 = k) = 0$, d'où le résultat. $\square$

Chapitre 9

Théorèmes limite

Sommaire

9.1 Lois faible et forte des grands nombres 62
9.1.1 Loi faible des grands nombres : énoncé et commentaires 62
9.1.2 Démonstration de la loi faible des grands nombres 63
9.1.3 Applications aux intervalles de fluctuations 63
9.1.4 Loi forte des grands nombres 64
9.2 Théorème de la limite centrale 64
9.2.1 Enoncé et commentaires 64
9.2.2 Applications aux intervalles de fluctuations 66
9.2.3 Approximation de la loi binomiale 67
9.3 Initiation à l'estimation statistique 67
9.3.1 Modélisation probabiliste. 67
9.3.2 Estimation à l'aide de la loi des grands nombres. 68
9.3.3 Contrôle de l'erreur d'approximation à l'aide de la loi faible des grands nombres. 68
9.3.4 Contrôle de l'erreur d'approximation à l'aide du théorème central limite. 69

Objectifs :

- Comprendre l'énoncé de la loi forte des grands nombres.
- Savoir déterminer des intervalles de fluctuations.
- Comprendre et savoir utiliser le théorème de la limite centrale.

Dans ce chapitre, on s'intéresse au problème qui consiste à observer un grand nombre fois le résultat d'une expérience aléatoire répétée. Typiquement, si je lance un très grand nombre de fois une pièce équilibée, que puis-je dire sur la proportion de pile obtenue ?
La loi des grands nombres donne une réponse mathématique à ce genre de questions.

9.1 Lois faible et forte des grands nombres

On considère une suite $(X_n)_{n\geq 1}$ de variables aléatoires indépendantes et identiquement distribuées, à valeurs dans $\mathbb{R}$. On suppose de plus ces variables de carré intégrable, c'est-à-dire que $\mathbb{E}[X_i^2] < +\infty$.

Définition 9.1 *Un n-échantillon est la donnée de n variables aléatoires indépendantes et identiquement distribuées $(X_i)_{1\leq i\leq n}$. La moyenne (au sens usuel) de ces n variables aléatoires*

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

est une variable aléatoire appelée moyenne empirique du n-échantillon.

Calculons son espérance et sa variance :

$$\mathbb{E}\left[\bar{X}_n\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \mathbb{E}[X_1]$$

car les v.a. étant i.i.d., elles ont toutes la même espérance $\mathbb{E}[X_1]$. Comme les v.a. sont indépendantes, la proposition 8.13 donne

$$\mathrm{Var}\left(\bar{X}_n\right) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n}\mathrm{Var}\left(X_1\right)$$

car toutes les v.a. ont la même variance. Le fait que $\mathrm{Var}\left(\bar{X}_n\right) \to 0$ lorsque $n \to \infty$ signifie que la v.a. $\bar{X}_n$ est de plus en plus concentrée autour de sa moyenne $\mathbb{E}[X_1]$. Le théorème suivant formalise ce résultat.

9.1.1 Loi faible des grands nombres : énoncé et commentaires

On considère une suite $(X_n)_{n\geq 1}$ de variables aléatoires indépendantes et identiquement distribuées, à valeurs dans $\mathbb{R}$. On suppose de plus ces variables de carré intégrable, c'est-à-dire que $\mathbb{E}[X_i^2] < +\infty$.

Théorème 9.2 *(loi faible des grands nombres)*
Soit $(X_n)_{n\geq 1}$ une suite de v.a.i.i.d. telles que $\mathbb{E}[X_i^2] < +\infty$. On pose $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Alors

$$\forall \varepsilon > 0, \quad \forall n \geq 1, \quad \mathbb{P}\left(\left|\bar{X}_n - \mathbb{E}[X_1]\right| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(X_1)}{n\,\varepsilon^2}.$$

En conséquence, pour tout $\varepsilon > 0$, $\mathbb{P}\left(\left|\bar{X}_n - \mathbb{E}[X_1]\right| \geq \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 0$.

Remarque. • L'hypothèse $\mathbb{E}[X_1^2] < +\infty$ implique l'existence de $\mathbb{E}[X_1]$ et de $\mathrm{Var}(X_1)$.
• La loi faible des grands nombres dit que, dans un certain sens, la moyenne empirique se rapproche, quand la taille n de l'échantillon tend vers l'infini, de l'espérance de la variable X_1.
• Il s'agit d'un résultat quantitatif : on a une majoration explicite de la probabilité que la moyenne empirique soit éloignée de plus de ε de $\mathbb{E}[X_1]$.

▷ **Exemple.** On souhaite modéliser des lancers successifs et indépendants d'une pièce qui tombe sur pile avec probabilité $p \in]0,1[$. Considérons la v.a X_i qui vaut 1 si pile sort au i-ème lancer et 0 si face sort au i-ème lancer. Les lancers étant indépendants, la suite $(X_i)_{i\geq 1}$ est une suite de v.a. i.i.d. de Bernoulli de paramètre $p = \mathbb{P}(X_i = 1)$ la probabilité d'obtenir pile à un lancer. Notons que $\mathbb{E}[X_1] = p$ et $\mathrm{Var}(X_1) = p(1-p)$.

Par ailleurs, $\displaystyle\sum_{i=1}^n X_i$ est le nombre de piles obtenus pendant les n premiers lancers, et donc $\bar{X}_n$ est la proportion de piles obtenus pendant les n premiers lancers.

La loi faible des grands nombres dit que, dans un certain sens, la proportion $\bar{X}_n$ de piles obtenus lors des n premiers lancers se rapproche, lorsque n tend vers $+\infty$, de p qui est la probabilité d'obtenir pile. Autrement dit, lorsque n est grand, on s'attend lors de n lancers à obtenir environ $p\,n$ piles. La loi faible des grands nombres donne un sens mathématique précis à cette phrase.

9.1.2 Démonstration de la loi faible des grands nombres

Commençons par deux résultats intermédiaires :

Proposition 9.3 (Inégalité de Markov) *Soit X une variable aléatoire positive et intégrable. Alors*

$$\forall a > 0, \quad \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Démonstration. On utilise le calcul avec les système complet $(\{X < a\}, \{X \geq a\})$, et la positivité de l'espérance :

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}\big[X(\mathbb{1}_{\{X<a\}} + \mathbb{1}_{\{X\geq a\}})\big] \\
&= \mathbb{E}\big[X\mathbb{1}_{\{X<a\}}\big] + \mathbb{E}\big[X\mathbb{1}_{\{X\geq a\}}\big] \\
&\geq 0 + \mathbb{E}\big[a\mathbb{1}_{\{X\geq a\}}\big] \\
&\geq a\,\mathbb{E}\big[\mathbb{1}_{\{X\geq a\}}\big] = a\,\mathbb{P}(X \geq a)
\end{aligned}$$

et on conclut en divisant par $a > 0$. □

Proposition 9.4 (Inégalité de Tchebitcheff) *Soit X une variable aléatoire de carré intégrable (c'est-à-dire $\mathbb{E}[X^2] < \infty$). Alors*

$$\forall \varepsilon > 0, \quad \mathbb{P}\big(|X - \mathbb{E}[X]| \geq \varepsilon\big) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

Démonstration. Notons que $\mathbb{P}\big(|X-\mathbb{E}[X]| \geq \varepsilon\big) = \mathbb{P}\big(|X-\mathbb{E}[X]|^2 \geq \varepsilon^2\big)$. On applique alors l'inégalité de Markov à la variable aléatoire positive et intégrable $(X - \mathbb{E}[X])^2$:

$$\mathbb{P}\big(|X - \mathbb{E}[X]|^2 \geq \varepsilon^2\big) \leq \frac{\mathbb{E}\big[(X - \mathbb{E}[X])^2\big]}{\varepsilon^2} = \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

□

Il nous reste à démontrer la loi faible : c'est maintenant très simple D'après l'inégalité de Tchebitcheff,

$$\mathbb{P}\left(\left|\bar{X}_n - \mathbb{E}[X_1]\right| \geq \varepsilon\right) = \mathbb{P}\left(\left|\bar{X}_n - \mathbb{E}[\bar{X}_n]\right| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\mathrm{Var}(X_1)}{n\,\varepsilon^2}$$

ce qui est l'inégalité voulue.

9.1.3 Applications aux intervalles de fluctuations

Soit $(X_n)_{n\geq 1}$ une suite de v.a.i.i.d. telles que $\mathbb{E}[X_i^2] < +\infty$. On pose $\bar{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i$.

Définition 9.5 *Un* intervalle de fluctuation *pour la moyenne empirique $\bar{X}_n$ au* niveau de confiance $1-\alpha$ *est un intervalle I tel que $\mathbb{P}(\bar{X}_n \in I) \geq 1 - \alpha$.*

On peut utiliser la loi faible des grands nombres pour déterminer des intervalles de fluctuations pour la moyenne empirique $\bar{X}_n$. Soit $n \geq 1$ fixé et $\varepsilon > 0$:

$$\begin{aligned}
\mathbb{P}\left(\left|\bar{X}_n - \mathbb{E}[X_1]\right| \geq \varepsilon\right) &\leq \frac{\mathrm{Var}(X_1)}{n\,\varepsilon^2}\\
\text{donc } \mathbb{P}\left(\left|\bar{X}_n - \mathbb{E}[X_1]\right| \leq \varepsilon\right) &\geq 1 - \frac{\mathrm{Var}(X_1)}{n\,\varepsilon^2}\\
\text{donc } \mathbb{P}\left(\bar{X}_n \in [\mathbb{E}[X_1]-\varepsilon, \mathbb{E}[X_1]+\varepsilon]\right) &\geq 1 - \frac{\mathrm{Var}(X_1)}{n\,\varepsilon^2}.
\end{aligned}$$

Ainsi, la loi faible des grands nombres assure que

l'intervalle $[\mathbb{E}[X_1]-\varepsilon, \mathbb{E}[X_1]+\varepsilon]$ est un intervalle de fluctuation pour $\bar{X}_n$ au niveau $1-\alpha$, avec $\alpha = \dfrac{\mathrm{Var}(X_1)}{n\,\varepsilon^2}$:

$$\mathbb{P}\left(\bar{X}_n \in [\mathbb{E}[X_1]-\varepsilon, \mathbb{E}[X_1]+\varepsilon]\right) \geq 1 - \frac{\mathrm{Var}(X_1)}{n\,\varepsilon^2}.$$

En pratique, on a trois paramètres : la taille n de l'échantillon, la demi-largeur ε de l'intervalle de confiance, et le niveau de confiance $1-\alpha$, liés par la relation

$$\alpha = \frac{\mathrm{Var}(X_1)}{n\,\varepsilon^2}.$$

Si on fixe deux des trois paramètres, cette relation permet de déterminer le troisième :

$$\alpha = \frac{\mathrm{Var}(X_1)}{n\,\varepsilon^2} \quad \Leftrightarrow \quad \varepsilon = \sqrt{\frac{\mathrm{Var}(X_1)}{n\alpha}} \quad \Leftrightarrow \quad n = \frac{\mathrm{Var}(X_1)}{\alpha\varepsilon^2}.$$

9.1.4 Loi forte des grands nombres

Théorème 9.6 *(Loi forte des grands nombres)*

Soit $(X_n)_{n\geq 1}$ une suite de v.a.i.i.d. telles que $\mathbb{E}[|X_1|] < +\infty$. On pose $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Alors

$$\mathbb{P}\big(\{\omega \in \Omega \ : \ \textit{la suite numérique } (\bar{X}_n(\omega))_{n\geq 1} \textit{ converge vers le nombre } \mathbb{E}[X_1]\}\big) = 1.$$

Autrement dit, presque sûrement, $\bar{X}_n(\omega) \to \mathbb{E}[X_1]$ lorsque $n \to \infty$.

Remarque. • C'est un résultat plus fort que la loi faible, puisque les hypothèses sont plus faibles...
• Par contre, c'est un résultat asymptotique, et non quantitatif. Il est donc moins efficace dans les utilisations pratiques.

9.2 Théorème de la limite centrale

9.2.1 Enoncé et commentaires

La loi des grands nombres dit que $\bar{X}_n$ tend en un certain sens vers $\mathbb{E}[X_1]$. L'énoncé suivant quantifie la vitesse de cette convergence : très grossièrement, il dit que que $\bar{X}_n - \mathbb{E}[X_1]$ tend vers 0 à la même vitesse que $1/\sqrt{n}$.

Théorème 9.7 *(théorème de la limite centrale* ou *théorème central limite)*

Soit $(X_n)_{n\geq 1}$ une suite de v.a. discrètes i.i.d. telles que $\mathbb{E}[X_1^2] < +\infty$. On pose $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

Alors, pour tous $-\infty \leq a < b \leq +\infty$,

$$\mathbb{P}\left(a \leq \sqrt{\frac{n}{\mathrm{Var}(X_1)}}\Big(\bar{X}_n - \mathbb{E}[X_1]\Big) \leq b\right) \underset{n\to\infty}{\longrightarrow} \int_a^b \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx.$$

Démonstration. Moins difficile que la démonstration de la loi forte des grands nombres, mais utilise des notions sortant du cadre de ce cours, résultat admis. □

Remarque. L'hypothèse $\mathbb{E}[X_1^2] < \infty$ assure que les deux grandeurs $\mathbb{E}[X_1]$ et $\mathrm{Var}(X_1)$ sont bien définies.

Remarquons que $\int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \mathbb{P}(a \leq Z \leq b)$, où la variable Z suit une loi *normale centre réduite*.
Avec cette interprétation, si l'on pose $Z_n = \sqrt{\frac{n}{\mathrm{Var}(X_1)}}\Big(\bar{X}_n - \mathbb{E}[X_1]\Big)$, le théorème peut être reformulé en, pour tous $-\infty \leq a < b \leq +\infty$, $\mathbb{P}(a \leq Z_n \leq b) \to \mathbb{P}(a \leq Z \leq b)$. Il s'agit là encore d'une sorte de convergence de la v.a. Z_n vers la v.a. Z : on dit que Z_n *converge en loi* vers Z.

Ce résultat est très profond : avec peu d'hypothèses, il assure un résultat de convergence général. Il met aussi en évidence le rôle fondamental de la loi normale, elle apparaît en effet comme limite quelles que soient les lois au départ : c'est pourquoi on la rencontre souvent dans l'étude statistique de nombreux phénomènes.

Vérifier que $\mathbb{E}[Z_n] = 0$ et $\mathrm{Var}(Z_n) = 1$. En fait, on a pris la v.a. $\bar{X}_n$, on l'a *centrée* en lui retranchant son espérance afin d'obtenir une v.a. d'espérance nulle, puis on l'a *réduite* en la divisant par sa variance de sorte à obtenir une v.a. de variance 1.

Remarque. Vitesse de convergence dans la loi des grands nombres. Le théorème central limite précise la convergence donnée par la loi des grands nombres.
Plaçons-nous dans le cas d'une suite réelle $(u_n)_{n\in\mathbb{N}}$ qui converge vers une limite l. On suppose qu'on ne connaît pas exactement la valeur de l, et qu'on sait facilement calculer les termes successifs de $(u_n)_{n\in\mathbb{N}}$. On aimerait donc donner une valeur approchée de l en disant que pour n suffisamment grand, u_n n'est pas loin de l. Toute la question est de savoir dire rigoureusement ce que veut dire « u_n n'est pas loin de l », autrement dit d'être capable d'estimer l'erreur commise entre u_n et l. Ceci revient à étudier la vitesse de convergence de la suite $(u_n)_{n\in\mathbb{N}}$ vers l, c'est-à-dire de trouver le terme suivant dans le développement asymptotique de u_n quand n tend vers l'infini : par exemple,

$$u_n = l + \frac{3}{n^2} + o\left(\frac{1}{n^2}\right) \qquad \Leftrightarrow \qquad \lim_{n\to\infty} n^2\,(u_n - l) = 3.$$

Dans ce cas, la vitesse est en $1/n^2$. Remarquons que

$$\begin{aligned} &\text{si } \alpha < 2, \text{ alors } \lim_{n\to\infty} n^\alpha\,(u_n - l) = 0 \\ &\text{si } \alpha > 2, \text{ alors } \lim_{n\to\infty} n^\alpha\,|u_n - l| = +\infty. \end{aligned}$$

Le cas $\alpha = 2$, celui qui donne une limite finie non triviale, est donc celui qui donne la bonne vitesse de convergence.
Revenons maintenant au cas de variables aléatoires i.i.d. La loi des grands nombres affirme que

$$\frac{1}{n}\sum_{i=1}^n X_i \text{ converge en un certain sens vers } \mathbb{E}[X_1].$$

Pour étudier la vitesse de convergence de cette suite, on cherche « le bon α » tel que

$$n^\alpha \left(\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}[X_1]\right) \text{ converge, en un certain sens, vers une limite ni nulle ni infinie.}$$

Le théorème central limite nous dit que la bonne valeur de α à considérer est $\alpha = 1/2$: la vitesse de convergence dans la loi des grands nombres est donc en $1/\sqrt{n}$.

9.2.2 Applications aux intervalles de fluctuations

Loi normale centrée réduite. La densité de la loi normale centrée réduite est

$$\forall x \in \mathbb{R} \quad f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2).$$

Si Z est une variable aléatoire définie sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et suivant la loi normale centrée réduite, alors sa fonction de répartition est :

$$\forall t \in \mathbb{R} \quad \Phi(t) = \mathbb{P}(Z \leq t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx.$$

On ne peut pas calculer explicitement cette fonction de répartition. On peut vérifier qu'elle est bijective de $\mathbb{R}$ dans $]0, 1[$, et on notera Φ^{-1} son inverse. Pour avoir des valeurs particulières de Φ ou de Φ^{-1}, on utilise des valeurs approchées données par des tables ou calculées numériquement par des programmes d'approximation.

♣ **Exercice 9.1.** Soit $a > 0$. Montrer que

$$\mathbb{P}(-a \leq Z \leq a) = \Phi(a) - \Phi(-a) = 2\Phi(a) - 1.$$

Penser à tracer la courbe de la densité f et utiliser les propriétés de symétrie de cette courbe.

Soit $(X_n)_{n \geq 1}$ une suite de v.a.i.i.d. telles que $\mathbb{E}[X_i^2] < +\infty$. On pose $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

On peut utiliser le théorème central limite pour déterminer des intervalles de fluctuations pour la moyenne empirique $\bar{X}_n$. Soit $n \geq 1$ grand, et $a > 0$: on fait l'approximation

$$\begin{aligned}
\mathbb{P}\left(-a \leq \sqrt{\frac{n}{\mathrm{Var}(X_1)}} \left(\bar{X}_n - \mathbb{E}[X_1]\right) \leq a\right) &\simeq \int_{-a}^{a} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 2\Phi(a) - 1, \\
\text{donc } \mathbb{P}\left(-a\sqrt{\frac{\mathrm{Var}(X_1)}{n}} \leq \bar{X}_n - \mathbb{E}[X_1] \leq a\sqrt{\frac{\mathrm{Var}(X_1)}{n}}\right) &\simeq 2\Phi(a) - 1, \\
\text{donc } \mathbb{P}\left(\bar{X}_n \in \left[\mathbb{E}[X_1] - a\sqrt{\frac{\mathrm{Var}(X_1)}{n}}, \mathbb{E}[X_1] + a\sqrt{\frac{\mathrm{Var}(X_1)}{n}}\right]\right) &\simeq 2\Phi(a) - 1,
\end{aligned}$$

On prend alors

$$\begin{aligned}
2\Phi(a) - 1 = 1 - \alpha \quad &\Leftrightarrow \quad \Phi(a) = 1 - \frac{\alpha}{2} \quad \Leftrightarrow \quad a = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \\
\text{et } \varepsilon = a\sqrt{\frac{\mathrm{Var}(X_1)}{n}} &= \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\frac{\mathrm{Var}(X_1)}{n}}.
\end{aligned}$$

Ainsi, le théorème central limite assure que

l'intervalle $[\mathbb{E}[X_1] - \varepsilon, \mathbb{E}[X_1] + \varepsilon]$ est un intervalle de fluctuation pour $\bar{X}_n$ au niveau $1 - \alpha$, avec

$$\varepsilon = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\frac{\mathrm{Var}(X_1)}{n}}.$$

En pratique, on a trois paramètres : la taille n de l'échantillon, la demi-largeur ε de l'intervalle de confiance, et le niveau de confiance $1 - \alpha$, liés ici par la nouvelle relation

$$\varepsilon = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\frac{\mathrm{Var}(X_1)}{n}}.$$

Si on fixe deux des trois paramètres, cette relation permet de déterminer le troisième.

9.2.3 Approximation de la loi binomiale

Prenons le cas particulier où les variables aléatoires X_n sont des v.a. i.i.d. de loi de Bernoulli de paramètre p. Dans ce cas, la v.a. $S_n = \sum_{i=1}^{n} X_i$ suit la loi binômiale de paramètres (n,p) ; le théorème central limite donne une façon d'approcher ces lois binomiales par des lois normales.

$\triangleright$ **Exemple.** Une usine fabrique des boulons, qui s'avèrent défectueux dans 3% des cas. On dispose de 100 boulons, quelle est la probabilité d'observer au moins 80 boulons corrects ?

On définit les v.a. X_i par $X_i = 1$ si le i-ème boulon est défectueux et $X_i = 0$ s'il ne l'est pas. La suite $(X_n)_{n\in\mathbb{N}}$ est donc une suite de v.a. i.i.d. suivant la loi de Bernoulli de paramètre $p = 0.03$. Par ailleurs, $S_n = \sum_{i=1}^{n} X_i$ compte le nombre de boulons défectueux si l'on en prend n, et suit la loi binômiale de paramètres $(n,p) = (100, 0.03)$.

– La loi des grands nombres assure que

$$\frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n} \simeq \mathbb{E}[X_1] = p \qquad \text{et donc} \qquad \frac{S_{100}}{100} \simeq 0.03 \qquad \text{et donc} \qquad S_{100} \simeq 3.$$

On s'attend donc, dans notre paquet de 100 boulons, à en trouver environ 3 défectueux. En particulier, on s'attend à ce que la probabilité d'avoir dans notre paquet de 100 boulons au moins 80 boulons corrects soit proche de 1.

– Revenons à la question posée. On a au moins 80 boulons corrects est l'événement $\{S_{100} \le 20\}$ et on veut estimer sa probabilité. On s'attend par le point précédent à ce que cette probabilité soit proche de 1 et pour être plus précis, on va utiliser le théorème central limite. Pour cela, on commence par transformer l'écriture de l'événement $\{S_{100} \le 20\}$ de sorte à faire apparaître $Z_{100} = \sqrt{\frac{100}{\mathrm{Var}(X_1)}}\Big(\bar{X}_{100} - \mathbb{E}[X_1]\Big) = \sqrt{\frac{100}{0,03\,(1-0,03)}}\Big(\bar{X}_{100} - 0,03\Big)$.

$$\begin{aligned}\mathbb{P}(S_{100} \le 20) &= \mathbb{P}\left(\bar{X}_{100} \le 0,2\right) = \mathbb{P}\left(\bar{X}_{100} - 0,03 \le 0,17\right)\\ &= \mathbb{P}\left(Z_{100} \le 0,17\sqrt{\frac{100}{0,03\,(1-0,03)}}\right) \simeq \int_{-\infty}^{9,97} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx.\end{aligned}$$

On ne peut pas calculer explicitement cette intégrale ; pour en connaître une valeur approchée, on utilise des tables de valeurs ; on trouve que ce nombre est très très proche de 1.

♣ **Exercice 9.2.** On prend cette fois-ci un échantillon de 1000 boulons. Donner une valeur approchée de la probabilité d'avoir moins de 20 boulons défectueux, puis de la probabilité d'avoir entre 20 et 40 boulons défectueux. On pourra utiliser l'approximation $\int_{1,85}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx \simeq 0,032$.

9.3 Initiation à l'estimation statistique

On souhaite estimer la proportion θ de gens dans la population française qui préfèrent le chocolat noir au chocolat blanc. Cette proportion, à une date fixée, est un paramètre non aléatoire θ, que l'on ne connaît pas.
On décide d'interroger n personnes, de compter le nombre N_n de personnes qui préfèrent le chocolat noir, et d'estimer θ par N_n/n :

$$\theta \simeq \frac{N_n}{n}.$$

Mais on aimerait aussi savoir quelle erreur on commet en faisant une telle approximation.

9.3.1 Modélisation probabiliste.

On décide de coder la réponse de la i-ème personne interrogée par $X_i = 1$ si elle préfère le chocolat noir, et $X_i = 0$ si elle préfère le chocolat blanc. Il est naturel de supposer que les personnes ne s'influencent pas entre elles et de modéliser cette hypothèse en disant que les X_i sont des variables indépendantes. Par ailleurs, comme une proportion θ de personnes préfère le chocolat noir, il est cohérent de supposer que $X_i = 1$ avec probabilité θ et $X_i = 0$ avec probabilité $1-\theta$. Ainsi, on modélise les résultats du sondage à l'aide d'une suite $(X_i)_{1\le i\le n}$ de v.a. i.i.d. de loi de Bernoulli de paramètre θ.

9.3.2 Estimation à l'aide de la loi des grands nombres.

Avec ce formalisme, $\mathbb{E}[X_i] = \theta$, $\mathrm{Var}(X_i) = \theta\,(1-\theta)$,

$$N_n = \sum_{i=1}^n X_i \qquad \text{et donc} \qquad \frac{N_n}{n} = \bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i.$$

La loi des grands nombres assure que $\bar{X}_n$ est proche, quand n est grand, de $\mathbb{E}(X_1) = \theta$: on dit que la moyenne empirique $\bar{X}_n$ est un estimateur de l'espérance $\mathbb{E}(X_1) = \theta$:

$$\theta \simeq \bar{X}_n = \frac{N_n}{n}.$$

9.3.3 Contrôle de l'erreur d'approximation à l'aide de la loi faible des grands nombres.

Quantifions maintenant à quel point cet estimateur est proche de la valeur réelle inconnue θ. Soit $\varepsilon > 0$ fixé. D'après la loi faible des grands nombres,

$$\mathbb{P}\left(\left|\bar{X}_n - \theta\right| \ge \varepsilon\right) = \mathbb{P}\left(\left|\bar{X}_n - \mathbb{E}(X-1)\right| \ge \varepsilon\right) \le \frac{\mathrm{Var}(X_1)}{n\,\varepsilon^2}.$$

Ici, comme X_1 suit la loi de Bernoulli de paramètre θ, on a $\mathrm{Var}(X_1) = \theta\,(1-\theta) \le 1/4$. Donc

$$\mathbb{P}\left(\left|\bar{X}_n - \theta\right| \ge \varepsilon\right) \le \frac{1}{4n\varepsilon^2} \text{ donc } \mathbb{P}\left(\left|\bar{X}_n - \theta\right| \le \varepsilon\right) \ge 1 - \frac{1}{4n\varepsilon^2},$$
$$\text{donc } \mathbb{P}\left(\theta \in \left[\bar{X}_n - \varepsilon, \bar{X}_n + \varepsilon\right]\right) \ge 1 - \frac{1}{4n\varepsilon^2}.$$

Définition 9.8 *Un* intervalle de confiance *pour un paramètre à estimer θ au* niveau de confiance $1-\alpha$ *est un intervalle aléatoire I tel que $\mathbb{P}(\theta \in I) \ge 1-\alpha$.*

Ici, la loi faible des grands nombres assure que

l'intervalle aléatoire $\left[\bar{X}_n - \varepsilon, \bar{X}_n + \varepsilon\right]$ est un intervalle de confiance pour θ au niveau de confiance $1-\alpha$, avec

$$\alpha = \frac{1}{4n\varepsilon^2}.$$

Rappelons que θ est un paramètre inconnu, mais fixé. Par contre, $\bar{X}_n$ est aléatoire, puisqu'elle dépend des réponses des n personnes que j'ai interrogées (un autre institut de sondage peut par exemple trouver un autre intervalle de confiance simplement parce qu'il n'a pas intérrogé les mêmes personnes).
Ici encore, la relation $\alpha = \dfrac{1}{4n\varepsilon^2}$ lie les trois paramètres ε, α et n. Deux de ces paramètres étant fixés, elle permet de déterminer le troisième, comme dans l'exemple ci-dessous.

▷ **Exemple.** On reprend la situation de l'exemple précédent.

1. On interroge $n = 1000$ personnes, et on obtient $\bar{X}_{1000} = 0,45$. Donner des intervalles de confiance pour θ aux niveaux $1-\alpha = 0,9$ et $1-\alpha = 0,95$.
   Solution : Comme $1-\alpha = 1 - \dfrac{1}{4n\varepsilon^2}$, on trouve $\varepsilon = \frac{1}{\sqrt{4\alpha n}}$.
   – Pour $1-\alpha = 0,9$, on trouve $\varepsilon = 0,05$. Donc $[0,4; 0,5]$ est un intervalle de confiance au niveau 90% pour θ.

– Pour $1-\alpha = 0,95$, on trouve $\varepsilon = 0,07$. Donc $[0,38; 0,52]$ est un intervalle de confiance au niveau 95% pour θ.

2. On interroge $n = 1000$ personnes, et on obtient $\bar{X}_{1000} = 0,45$. Quel est le niveau de confiance de l'intervalle $[0,43; 0,47]$?

   Solution : On veut que $\varepsilon = 0,02$. On trouve $1-\alpha = 0,375$. Donc $[0,43; 0,47]$ est un intervalle de confiance au niveau $37,5\%$.

3. Pour que $[0,445; 0,455]$ soit un intervalle de confiance de niveau 99%, combien faut-il interroger de personnes ?

   Solution : Comme $\alpha = \dfrac{1}{4n\varepsilon^2}$, on trouve $n = \dfrac{1}{4\alpha\varepsilon^2}$.
   On veut que $\varepsilon = 0,005$ et $1-\alpha = 0,99$, il faut donc choisir $n = 10^6$. On doit donc considérer un échantillon de taille au moins un million.

9.3.4 Contrôle de l'erreur d'approximation à l'aide du théorème central limite.

Quantifions maintenant à quel point cet estimateur est proche de la valeur réelle inconnue θ. Soit $n \geq 1$ grand, et $a > 0$: on fait l'approximation

$$\begin{aligned}
&\mathbb{P}\left(-a \leq \sqrt{\frac{n}{\mathrm{Var}(X_1)}}\Big(\bar{X}_n - \mathbb{E}[X_1]\Big) \leq a\right) \simeq \int_{-a}^{a} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx = 2\Phi(a) - 1, \\
\text{donc } &\mathbb{P}\left(-a\sqrt{\frac{\mathrm{Var}(X_1)}{n}} \leq \theta - \bar{X}_n \leq a\sqrt{\frac{\mathrm{Var}(X_1)}{n}}\right) \simeq 2\Phi(a) - 1, \\
\text{donc } &\mathbb{P}\left(\theta \in \left[\bar{X}_n - a\sqrt{\frac{\mathrm{Var}(X_1)}{n}}, \bar{X}_n + a\sqrt{\frac{\mathrm{Var}(X_1)}{n}}\right]\right) \simeq 2\Phi(a) - 1,
\end{aligned}$$

Mais $\mathrm{Var}(X_1) = \theta(1-\theta)$ dépend du paramètre inconnu θ. Comme précedemment, on majore la variance $\mathrm{Var}(X_1) = \theta(1-\theta) \leq 1/4$, et on obtient

$$\mathbb{P}\left(\theta \in \left[\bar{X}_n - a\sqrt{\frac{1}{4n}}, \bar{X}_n + a\sqrt{\frac{1}{4n}}\right]\right) = \mathbb{P}\left(\theta \in \left[\bar{X}_n - \frac{a}{2\sqrt{n}}, \bar{X}_n + \frac{a}{2\sqrt{n}}\right]\right) \geq 2\Phi(a) - 1.$$

On prend alors

$$\begin{aligned}
&2\Phi(a) - 1 = 1 - \alpha \quad \Leftrightarrow \quad \Phi(a) = 1 - \frac{\alpha}{2} \quad \Leftrightarrow \quad a = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \\
&\text{et } \varepsilon = \frac{a}{2\sqrt{n}} = \frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}{2\sqrt{n}}.
\end{aligned}$$

Ainsi, le théorème central limite assure que

> l'intervalle aléatoire $[\bar{X}_n - \varepsilon, \bar{X}_n + \varepsilon]$ est un intervalle de confiance pour θ au niveau $1-\alpha$, avec
> $$\varepsilon = \frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}{2\sqrt{n}}.$$

$\triangleright$ **Exemple.** On reprend la situation de l'exemple précédent.

1. On interroge $n = 1000$ personnes, et on obtient $\bar{X}_{1000} = 0,45$. Donner des intervalles de confiance pour θ aux niveaux $1-\alpha = 0,9$ et $1-\alpha = 0,95$.

   Solution :
   – Pour $1-\alpha = 0,9$, on trouve $\Phi^{-1}\left(1 - \dfrac{\alpha}{2}\right) = 1,64$, et $\varepsilon = 0,026$. Donc $[0,424; 0,476]$ est un intervalle de confiance au niveau 90% pour θ.
   – Pour $\alpha = 0,95$, on trouve on trouve $\Phi^{-1}\left(1 - \dfrac{\alpha}{2}\right) = 1,96$, et $\varepsilon = 0,031$. Donc $[0,419; 0,481]$ est un intervalle de confiance au niveau 95% pour θ.

2. On interroge $n = 1000$ personnes, et on obtient $\bar{X}_{1000} = 0,45$. Quel est le niveau de confiance de l'intervalle $[0,43; 0,47]$?

   Solution : On veut que $\varepsilon = 0,02$, donc $\Phi^{-1}\left(1 - \dfrac{\alpha}{2}\right) = 1,26$ donc $\alpha = 0,2$. Donc $[0,43; 0,47]$ est un intervalle de confiance au niveau 80%.

3. Pour que $[0,445; 0,455]$ soit un intervalle de confiance de niveau 99%, combien faut-il interroger de personnes ?

   Solution : Comme $\alpha = 0,01$, $\Phi^{-1}\left(1 - \dfrac{\alpha}{2}\right) = 2,76$. Avec $\varepsilon = 0,05$, on obtient $n = \left(\dfrac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}{2\varepsilon}\right)^2$, donc $n = 76000$.