

Mathématiques pour littéraires

Polycopié de cours

Vadim Lebovici¹

2021/2022 - S2

1. vadim.lebovici@ens.fr, <https://vadimlebovici.github.io>

Table des matières

1	Ensembles et applications	3
1.1	Introduction	4
1.1.1	Un peu d'Histoire	4
1.1.2	Comment comparer la taille d'ensembles infinis ?	5
1.1.3	Objectif du chapitre	6
1.2	Kit de manipulation des ensembles	7
1.2.1	Encore un peu d'Histoire	7
1.2.2	Enfin, des mathématiques	9
1.3	Fonctions	11
1.3.1	Motivations	11
1.3.2	Définitions	11
1.3.3	Comment se servir des applications pour notre problème ?	13
1.3.4	Injection	14
1.3.5	Surjection	14
1.3.6	Bijection	15
1.3.7	Reformulation de nos questions initiales	15
1.3.8	Définition des ensembles finis et infinis	16
1.4	Il y a autant de nombres pairs que de nombres entiers	16
1.5	Il y a plus de points sur une droite que de nombre entiers	16
1.5.1	Court rappel sur les nombres réels	16
1.5.2	$\mathbb{N} \preccurlyeq \mathbb{R}$	17
1.5.3	Est-ce que $\mathbb{N} \sim \mathbb{R}$?	17
1.6	Appendice : Paradoxe de Russel	20
1.7	Références	21
2	Suites et limites	22
2.1	Introduction	23
2.1.1	Paradoxe de la dichotomie	23
2.1.2	Quelle valeur ont les sommes infinies ?	24
2.1.3	Deux sommes divergentes	25
2.2	Plan du chapitre	28
2.3	Suite numérique	29
2.3.1	Peu d'Histoire	29

2.3.2	Définition et premiers exemples	29
2.3.3	Termes de la série géométrique (calcul de Zénon)	30
2.3.4	Propagation d'une épidémie	30
2.4	Limites	31
2.4.1	Reformulation de nos problèmes initiaux	36
2.4.2	Suites géométriques	37
2.5	Une bonne définition mathématique	38
2.5.1	Encadrement	38
2.5.2	Opérations	39
2.6	Résolution de nos problèmes initiaux	39
2.6.1	Paradoxe de Zénon	39
2.6.2	Propagation d'une épidémie	43
2.6.3	Série de Grandi	44
2.6.4	Série harmonique	44
3	Probabilités et statistiques	46
3.1	Introduction	47
3.1.1	Problème jouet	47
3.1.2	Le hasard	48
3.2	Probabilités sur un univers fini	50
3.2.1	Définition	50
3.2.2	Propriétés	53
3.2.3	Exemples	54
3.3	Variables aléatoires	56
3.3.1	Vers les statistiques	56
3.3.2	Définition	58
3.3.3	Indépendance	59
3.3.4	Variables aléatoires suivant une loi uniforme	60
3.3.5	Variables aléatoires suivant une loi de Bernoulli	61
3.4	Espérance	62
3.4.1	Introduction	62
3.4.2	Définition	64
3.4.3	Propriétés	65
3.4.4	Variance	66
3.4.5	Mesure de l'écart à la moyenne	68
3.5	Statistiques et loi des grands nombres	69

Chapitre 1

Ensembles et applications

1.1 Introduction

«L'éternité, c'est long... surtout vers la fin. »

Woody Allen

«L'infini, c'est grand... mais ça dépend surtout duquel. »

Woody Allen (s'il avait été mathématicien)

Dans ce premier chapitre, nous nous intéresserons à un concept qui peut sembler mystérieux à première vue : l'infini. Ce concept, qui signifie *ce qui n'est pas limité*, n'a cessé de fasciner les intellectuelles à travers les époques et sera le fil rouge de ce cours. La raison en est que son étude motiva les mathématiciennes à développer des outils et des théories d'une grande richesse. Le premier de ces fruits s'est révélé utile pour fonder tout l'édifice mathématique au XX^e siècle : la théorie des ensembles.

Note bibliographique. *Toute cette section d'introduction, notamment les points historiques, a été écrite très largement grâce au très bon livre Histoire de la Théorie des Ensembles de Jean-Pierre Belna. Elle s'inspire également de l'article de Patrick Dehornoy, Georg Cantor, et les infinis furent, que je vous recommande vivement.*

1.1.1 Un peu d'Histoire

Si j'osais simplifier l'Histoire, voici ce que je dirais¹ : l'infini pose déjà problème aux Grecques. Zénon d'Elée expose au V^eme siècle avant J.C. une série de paradoxes destinée à montrer que le mouvement est impossible. Cette thèse n'étant pas des plus intuitives, on est en droit de se demander si la démonstration tient la route. Or, ces paradoxes reposent tous sur des divisions *infinies* (et indéfinies) de l'espace et/ou du temps, et nous verrons au Chapitre 2 qu'une définition mathématique de ces opérations fait disparaître les paradoxes.

Un siècle plus tard, c'est au tour d'Aristote d'étudier l'infini. Selon lui, tout être est à la fois *en puissance* et *en acte*. En puissance, au sens *qui est possible mais n'est pas encore réalisé, qui n'a d'existence que virtuelle*. En acte, au sens *qui est réalisé ou achevé* et a donc une existence concrète. L'infini ne peut alors pas être en acte, puisqu'il ne s'achève jamais et reste ainsi irrémédiablement en puissance. Il n'a donc qu'une existence virtuelle et n'est pas concret.

Cette vision de l'infini imprègnera longtemps la pensée mathématique. Ce qui peut paraître tout aussi étonnant est que par ailleurs, note Aristote lui-même, le caractère virtuel de l'infini n'est pas gênant pour faire des mathématiques : pour faire des raisonnements sur des droites, il suffit de réfléchir à des segments que l'on pourrait prolonger

1. Les parties de ce cours qui ne sont pas *stricto sensu* des mathématiques, comme la tentative d'histoire de la pensée qui va suivre, doivent être prises avec beaucoup de précautions, car si je suis qualifié pour parler de mathématiques, je suis (très) loin d'être expert du reste, et serait ravi de me voir corriger par un retour de mail, pour améliorer ce cours.

aussi longuement que l'on veut, il n'est pas nécessaire de considérer la droite dans son intégralité, dans toute son étendue, qui elle est infinie.

Un des points qui fut l'objet de nombre de réflexions fut l'axiome du tout et la partie d'Euclide (III^{ème} siècle avant J.C.) figurant dans *Les Eléments*, et s'exprimant ainsi :

Axiome 1.1. *Le tout est plus grand que la partie.*

Une manière naturelle d'interpréter cet axiome est la suivante : si l'on considère un certain ensemble d'éléments, n'importe lequel de ses sous-ensembles² est plus petit que l'ensemble de départ. Cela semble pour le moins évident... si l'on ne considère que des ensemble finis.

Si l'on considère des ensembles infinis en revanche, comme le fait remarquer Duns Scot (XIII^{ème} siècle), l'axiome pose question. Considérons par exemple l'ensemble des nombres entiers naturels 0, 1, 2, 3, ..., et le sous-ensemble des entiers pairs 0, 2, 4, ... D'une part, l'un contient strictement l'autre, puisque 3 par exemple est un entier qui n'est pas pair. D'autre part, les deux sont infinis et pourraient donc être considérés comme ayant la même taille.

Remarque 1.2. Avant que ne naisse la théorie axiomatique des ensembles, les mathématiciens ont historiquement toujours résolu cette contradiction en s'interdisant de comparer la taille d'ensembles infinis, afin de ne pas remettre en cause l'axiome d'Euclide.

1.1.2 Comment comparer la taille d'ensembles infinis ?

Ce que le paradoxe de Duns Scot fait ressortir, c'est qu'outre le fait que l'infini est un concept à manier avec précaution, il se pose un problème lorsque l'on cherche à comparer deux ensembles infinis : la notion de comparaison de taille. Qu'entend-on exactement « par tel ensemble est *plus grand* que tel autre » ? Duns Scot énonce essentiellement deux manières de faire.

Première approche. On pourrait définir qu'un ensemble est plus petit qu'un autre si le premier est inclus dans le second. Le problème c'est que cette méthode ne permet de comparer la taille que d'ensembles dont les éléments sont de même nature : un ensemble fini de pommes et un ensemble fini de poires sont incomparables...

Deuxième approche. On pourrait suivre notre manière de procéder pour les ensembles finis :

1. on compte le nombre d'éléments du premier ensemble, appelé *cardinal* en mathématiques,
2. on compte le nombre d'éléments du second ensemble,

2. i.e n'importe quel ensemble d'éléments présents dans l'ensemble initial.

3. on compare les cardinaux obtenus.

Dans le cas infini, on pourrait très bien se dire : les deux premières étapes renvoient le même résultat, infini, les cardinaux sont donc égaux, si tant est que le terme de *cardinaux* ait un sens dans ce cas. Cependant, c'est une grave erreur : les deux premières étapes ne renvoient pas le même résultat, elles ne se terminent jamais ! Il n'y a donc jamais de moment où l'on pourra comparer les résultats et il n'y a non plus de *résultat* qui ait un sens.

Résolution par Cantor. En 1873, le mathématicien allemand Georg Cantor (1845 – 1918) souhaite comparer la taille de deux ensembles que vous connaissez. Vous avez peut-être déjà entendu dire qu'il y a une infinité de points sur une droite.³ Vous savez également qu'il y a une infinité de nombres entiers naturels : 0, 1, 2, ... Cantor se pose alors la question suivante :

Question 1.3. *Est-ce qu'il y a autant de points sur une droite que de nombres entiers naturels ?*

C'est en cherchant une réponse à cette question que Cantor introduira un vocabulaire précis pour parler des ensembles ainsi qu'une approche fructueuse pour comparer leur taille. Il parviendra à la résoudre en 1874 : *non*, il n'y en a pas autant, il y a plus de points sur une droite que de nombres entiers, en un sens précis et satisfaisant.

1.1.3 Objectif du chapitre

La question 1.3 est une des questions que nous allons chercher à résoudre dans ce chapitre et qui va nous servir de motivation pour introduire des notions fondamentales en mathématiques : les ensembles et les applications. Ces notions sont équivalentes aux lettres et aux mots d'une langue, et sont aussi intéressantes qu'elles vous seront utiles si vous souhaitez lire des mathématiques, quelle qu'en soit la forme, un jour. Le chapitre est donc structuré ainsi :

1. Nous commencerons par introduire le vocabulaire utilisé en mathématiques pour parler d'ensembles.
2. Nous définirons ensuite une bonne notion de comparaison de taille via le concept mathématique d'*application*.
3. Nous répondrons à la question de Duns Scot : y a-t-il plus de nombres entiers que de nombres pairs ?
4. Pour enfin répondre à la question de Cantor : y a-t-il plus de points sur une droite que de nombres entiers naturels ?

3. Notez que ce fait, aujourd'hui enseigné en primaire, fut historiquement loin d'être une évidence : il était considéré comme faux par Aristote et plus tard par Leibniz.

1.2 Kit de manipulation des ensembles

1.2.1 Encore un peu d'Histoire

Note bibliographique. Cette partie tire grandement son contenu et même ses formulations du brillant article de Patrick Dehornoy *Théorie axiomatique des ensembles pour Encyclopedia Universalis*, que vous pouvez trouver sur son site internet. Je dresse ici un court résumé des idées essentielles que je veux vous présenter, mais je vous incite vivement à lire au moins l'intégralité de la première section de cet article. Pour un point de vue plus détaillé et contenant l'expression explicite des axiomes, vous pouvez consulter la seconde partie du polycopié de cours de Jean Feydy, Culture mathématique, disponible sur son site.

Les recherches de Cantor sur l'infini ont été le point de départ de la théorie des ensembles modernes, avec les travaux de Richard Dedekind (1831 – 1916), mathématicien allemand. La théorie des ensembles, au sens de la théorie *axiomatique* des ensembles, née à la fin du XIX^e siècle, a pour but de fournir un cadre conceptuel dans lequel la notion d'ensemble est bien définie et peut être étudiée. Elle a occupé un rôle prépondérant lors de son développement tout au long du XX^e siècle, jusqu'à servir un temps de base à l'édifice mathématique tout entier. Aujourd'hui, la théorie des ensembles développée au XX^e siècle ne peut plus servir de base à la totalité de l'édifice mathématique, mais cela dépasse le cadre de ce cours.

Pourquoi cette théorie est-elle dite axiomatique ? La raison en est que définir les ensembles est très délicat : si l'on choisit une définition trop vague, elle mène à des contradictions, comme le paradoxe de Russel (voir Section 1.6), du nom du mathématicien gallois du même nom Bertrand Russel (1872 – 1970). Or, il est raisonnable de penser, et c'est l'hypothèse que nous ferons, que le monde n'est pas contradictoire, donc que si notre définition l'est, c'est qu'elle ne retranscrit pas la réalité que nous voulions décrire, elle est donc à changer.

Face à une difficulté si grande pour expliciter la *nature* du concept d'ensemble, les mathématiciens ont recours à un procédé très intéressant, dont l'idée est la suivante :

Idée fondamentale. Il n'est pas nécessaire, pour travailler avec le concept d'ensemble d'un point de vue mathématique, d'en donner la *nature*. Il suffit de décrire de manière non-ambiguë⁴ les propriétés minimales (appelées *axiomes*) que doivent satisfaire les choses que l'on nommera ensuite *ensembles*. Nous pourrions alors étudier le comportement des choses satisfaisant ces propriétés en établissant les conséquences logiques⁵, appelées *théorèmes*, qu'ont ces propriétés.

Dans le cas des ensembles, ce sont principalement Ernst Zermelo, Abraham Frænkel et Thoralf Skolem qui ont élaboré au XX^e siècle une théorie axiomatique des ensembles, que l'on appelle la théorie ZF (Zermelo-Frænkel).

4. Ceci peut être réalisé grâce au langage symbolique de la logique.

5. Pour des règles de logique que l'on aura pris soin de spécifier de manière non-ambiguë.

Cette idée fondamentale n'est pas utilisée que dans le cas de la théorie des ensembles. Pour étudier les nombres, il suffit d'étudier les conséquences logiques de leurs propriétés fondamentales prises comme axiomes : chaque nombre a un successeur, il existe un nombre plus petit que tous les autres que l'on note 0, etc. C'est l'arithmétique du mathématicien italien Peano (1858 – 1932). De même, pour faire de la géométrie, il n'y a nul besoin de définir la nature d'un point ou d'une droite, il suffit de prendre leurs propriétés essentielles pour axiomes (par deux points passe une unique droite, etc), et d'en établir les conséquences logiques. C'est ce qu'a fait le mathématicien antique Euclide (~ 300 av. J.C.).

Cette approche pour appréhender un concept peut sembler décevante car la nature des ensembles nous reste inaccessible. Les mathématiques apparaissent comme une simple *modélisation* d'objets concrets, au sens où l'on a encodé les propriétés des objets à étudier dans un langage symbolique, pour ne plus étudier que cette formalisation.

Pourquoi n'est-ce pas si terrible ?

1. Il est bon de se rappeler que c'est l'exigence de rigueur qui impose un tel choix faute de mieux.
2. De plus, les fruits de cette approche sont d'une remarquable généralité : n'importe quel concept satisfaisant les axiomes de l'arithmétique de Peano se comporte comme un nombre, et satisfait donc *tous les théorèmes* valides sur les nombres. En mathématiques, il est fréquent et très fructueux de remarquer que certains objets se comportent comme d'autres déjà étudiés. On peut alors gratuitement appliquer aux premiers tous les théorèmes valides sur les seconds. Il est par exemple possible de définir une addition entre courbes fermées, qui permet de faire de l'algèbre avec des objets géométriques...
3. Enfin, de telles approches sont *redoutablement efficaces*. Certes, les mathématiques sont fondées sur des axiomes logiques exprimés en des termes symboliques, et peuvent donc sembler ne pas nous renseigner sur la nature des objets en question. Cependant, ce sont ces mêmes mathématiques qui permettent de décrire le réel avec tant de précision dans les autres sciences, et de faire ainsi des prédictions très précises de phénomènes naturels.

Une dernière remarque sur ce point : ces questions d'axiomatisation nous rappellent à quel point les mathématiques sont soumises à une part de subjectivité, le choix des axiomes, inhérente à toute forme de pensée. Cependant, il est à noter que les mathématiques ont la particularité de chercher à réduire le plus possible cette part de subjectivité.

Pourquoi la théorie des ensembles peut-elle servir de base à l'édifice mathématique ? Il se trouve que la théorie des ensembles permet d'encoder les ensembles, mais pas seulement. On peut construire une *représentation*, un encodage, de quasiment tous les objets mathématiques à l'aide d'ensembles. Les ensembles ont le rôle de briques élémentaires, des *atomes*, qui permettent de construire les autres concepts mathématiques. L'avantage de ce procédé est qu'il n'est pas nécessaire d'ajouter des axiomes à

la théorie pour construire de nouveaux objets. Par exemple, les nombres sont construits comme des ensembles satisfaisant certaines propriétés (celles énoncées par Peano). Ce paragraphe mériterait un approfondissement, mais ceci sort du cadre de ce cours.

Conclusion. Nous sommes désormais convaincus qu'il est nécessaire de définir proprement le concept d'ensemble, une définition naïve n'étant pas satisfaisante. De plus, nous avons vu que *définir proprement* signifie : établir – à l'aide du vocabulaire de la logique – un nombre minimal d'axiomes qui décrivent les caractéristiques essentielles des ensembles. Dans ce cours, nous n'allons utiliser qu'une définition *intuitive* des ensembles ainsi que leurs propriétés les plus utiles en nous appuyant sur le fait que la définition et ces propriétés ont été bien définies par des mathématiciennes avant nous. Tout ce que l'on fera sera donc bien fondé, même si nous n'avons pas le temps de voir comment. Vous saurez ainsi manipuler des ensembles, en définir comme bon vous semble et travailler avec. C'est tout ce qui est nécessaire pour faire des mathématiques, même à un niveau Master.

1.2.2 Enfin, des mathématiques

Comme promis, voici quelques bases pour bien travailler avec les ensembles en mathématiques.

Définition 1.4 (Informelle). Un ensemble est une collection d'éléments munie d'un critère qui permet de déterminer si un élément est ou non dans l'ensemble.

Pour noter leurs concepts simplement et sans ambiguïté, les mathématiciens utilisent des symboles.

Notations 1.5.

- Pour spécifier les éléments d'un ensemble, on les écrira entre accolades. Ainsi, l'ensemble constitué des entiers 0 et 1, et du mot "zoo" sera noté $\{0, 1, \text{"zoo"}\}$. Pour lui donner un nom, on écrira $X = \{0, 1, \text{"zoo"}\}$.
- **Appartenance.** Pour indiquer qu'un élément e appartient à un ensemble E , on notera $e \in E$. Le symbole \notin indique au contraire que e n'appartient pas à E . Ainsi, on a $0 \in X$ et $3 \notin X$.
- **Inclusion.** Si tous les éléments d'un ensemble A appartiennent également à E , on dit que A est inclus dans E , que E contient A , ou que A est un sous-ensemble de E et on note $A \subseteq E$. Ainsi, $\{0, \text{"zoo"}\} \subseteq X$.

Voici un exemple fondamental d'ensemble, le seul dont l'existence est spécifiquement stipulée dans les axiomes de la théorie ZF⁶.

Exemple 1.6 (Ensemble vide). L'ensemble qui ne contient aucun élément est appelé *ensemble vide* et noté \emptyset .

6. La théorie des ensembles de Zermelo et Fraenkel.

Un des axiomes de ZF détermine quand on a le droit de dire que deux ensembles sont égaux :

Axiome 1.7 (Extensionnalité). *Deux ensembles sont égaux s'ils ont les mêmes éléments.*

Notation 1.8. Pour écrire le contenu d'un ensemble, l'ordre dans lequel sont notés les éléments ne compte pas et on ne s'autorise pas de répétition. Ainsi on a $X = \{0, 1, \text{"zoo"}\} = \{\text{"zoo"}, 1, 0\} = \{\text{"zoo"}, 1, 0, 0\}$. Il n'y a qu'un seul nombre 0, ce n'est pas la peine de le noter deux fois.

Quels ensembles peut-on définir ? Pour l'instant, nous ne pouvons définir les ensembles qu'en énumérant leurs éléments entre parenthèses, on dit qu'on les définit par *extension*. Pourtant, nous avons envie de pouvoir parler de l'ensemble des entiers naturels ou de l'ensemble des entiers pairs, qui sont infinis et que l'on ne va donc pas pouvoir énumérer. C'est pour cela que l'on s'autorise à définir des ensembles par *compréhension*, au sens suivant :

Définition 1.9 (Informelle). Si P est une propriété (raisonnable) qui peut être vérifiée par les éléments d'un ensemble E , alors on peut définir l'ensemble des éléments de E vérifiant P , et on le note

$$\{e \in E \mid e \text{ vérifie } P\},$$

où le symbole \mid se lit "tel que".

Remarque 1.10. Notez comme la condition que les éléments e sélectionnés fassent déjà partie d'un ensemble E empêche la présence du paradoxe de Russel.

Exemple 1.11.

- Considérons l'ensemble F des fruits et considérons la propriété $P = \text{"être un fruit rouge"}$, qui est bien une propriété qui peut être vérifiée ou non par les éléments de l'ensemble F . On peut alors définir l'ensemble $\{f \in F \mid f \text{ vérifie } P\}$ qui n'est autre que l'ensemble des fruits rouges.
- Considérons l'ensemble \mathcal{P} des nombres pairs. Cet ensemble n'est autre que $\{n \in \mathbb{N} \mid n \text{ est pair}\}$, où \mathbb{N} désigne l'ensemble des entiers naturels.

Voici d'autres façons de construire de nouveaux ensembles à partir de ceux déjà à notre disposition.

Définition 1.12. Soient deux ensembles E et F .

- **Union.** On appelle *union* de E et F , et l'on note $E \cup F$, l'ensemble contenant tous les éléments de E et tous les éléments de F .
- **Intersection.** On appelle *intersection* de E et F , et l'on note $E \cap F$, l'ensemble contenant les éléments qui sont à la fois dans E et dans F .

- **Complémentaire.** Soit A un ensemble tel que $A \subseteq E$. On appelle *complémentaire de A dans E* , et l'on note $E \setminus A$ l'ensemble

$$E \setminus A = \{e \in E \mid e \notin A\}.$$

- **Ensemble des parties.** On appelle *ensemble des parties de E* , et l'on note 2^E ou $\mathcal{P}(E)$, l'ensemble des sous-ensembles de E .

Remarque 1.13. Le fait que l'union, l'intersection, le complémentaire et les parties précédemment définis soient bel et bien des ensembles découle des axiomes de la théorie ZF.

Exemple 1.14. Prenons pour exemple $E = \{0, 1, 2, 3\}$ et $F = \{2, 3, 4\}$.

- **Union.** $E \cup F = \{0, 1, 2, 3, 4\}$.
- **Intersection.** $E \cap F = \{2, 3\}$.
- **Complémentaire.** Soit $A = \{1, 3\}$. On a bien $A \subseteq E$. De plus, $E \setminus A = \{0, 2\}$.
- **Ensemble des parties.**

$$\begin{aligned} \mathcal{P}(E) = & \{\emptyset, \{0\}, \{1\}, \{2\}, \{3\}, \{0, 1\}, \{0, 2\}, \{0, 3\}, \{1, 2\}, \\ & \{1, 3\}, \{2, 3\}, \{0, 1, 2\}, \{0, 1, 3\}, \{0, 2, 3\}, \{1, 2, 3\}, \{0, 1, 2, 3\}\} \end{aligned}$$

1.3 Fonctions

1.3.1 Motivations

Ce qui nous intéresse est de trouver un moyen de comparer la taille (en un sens à définir) de deux ensembles. Le problème, c'est qu'on ne peut s'autoriser à utiliser la méthode qui consiste à compter le nombre d'éléments puis comparer les résultats puisqu'elle n'a pas de sens dans le cas d'ensembles infinis. De plus, souvenez-vous, on ne peut pas prendre la relation d'inclusion de l'un dans l'autre car elle ne permet pas de comparer des ensembles dont les éléments ne sont pas de même nature. Il faut que nous trouvions un moyen de définir une relation d'ordre entre deux ensembles, un lien d'infériorité de taille, pour pouvoir dire ensuite : le premier ensemble est lié par notre lien d'infériorité au deuxième ensemble, il est donc plus petit.

Dans la partie qui va suivre, nous allons définir ces liens entre ensembles : les applications.

1.3.2 Définitions

Nous arrivons à la partie où nous allons définir un des objets les plus fondamentaux des mathématiques : les fonctions. Si la théorie des ensembles est la grammaire des mathématiques, la fonction en est le verbe. C'est un concept fondamental, universel et extrêmement puissant.

Vous avez normalement, au cours de votre scolarité, déjà rencontré la définition d'une fonction. Vous leurs avez donné des formes algébriques, vous en avez dessiné des graphes... En voici une définition plus générale.

Définition 1.15 (Application). Soient X et Y deux ensembles. Une *application*⁷ f de X dans Y , noté $f : X \rightarrow Y$, fait correspondre à chaque élément x de X un unique élément de Y , noté $f(x)$.

Pour une application $f : X \rightarrow Y$ et un élément $x \in X$, si $y \in Y$ est tel que $y = f(x)$, on dit que y est l'*image de x par f* et que x est l'*antécédent de y par f* . Une application est donc un objet mathématique qui prend en entrée un élément x d'un ensemble X (appelé *ensemble de définition* de l'application), et renvoie un élément y de Y (appelé *ensemble but* de l'application), de sorte que pour un x fixé, la fonction renvoie toujours le même y , que l'on note $f(x)$.

Exemple 1.16. Un distributeur de boissons se modélise très bien par une application. Entrez un numéro x , et la boisson sélectionnée $f(x)$ vous tombe dans les mains. Pour un numéro donné, vous aurez toujours la même boisson⁸. L'ensemble X de définition du distributeur est l'ensemble des numéros correspondant aux boissons, et l'ensemble but Y est l'ensemble des boissons disponibles au distributeur.

Remarque 1.17. Ici, on ne fait que décrire le comportement d'une fonction, il faudrait ajouter un axiome à notre théorie stipulant qu'il existe de tels objets si nous voulons travailler avec. En fait, il est plus malin de construire un objet mathématique *application* à partir d'ensembles, des objets bien axiomatisés, qui se comporte comme on voudrait qu'une application se comporte. On aura alors défini une application sans ajouter d'axiome ! Comme pour les ensembles, la définition n'est pas le plus important pour nous ici, tout ce dont il faut se souvenir pour travailler est la propriété essentielle d'une application, et c'est justement la définition que je viens de vous donner.

Comment définir une application soi-même ? Pour définir une application, il suffit donc, pour tout élément $x \in X$ de définir un unique élément $f(x)$ de Y , celui qui est associé à x . On peut noter également une application f de la manière suivante :

$$f : \begin{array}{l} X \rightarrow Y \\ x \mapsto f(x) \end{array}$$

Exemple 1.18. La fonction *carré* est l'application c qui a tout entier naturel associé son carré. On la note

$$c : \begin{array}{l} \mathbb{N} \rightarrow \mathbb{N} \\ x \mapsto x^2 \end{array} .$$

7. souvent appelé *fonction*. En toute rigueur, il y a une nuance entre les deux termes mais même les mathématiciens font cet abus de langage.

8. Si vous entrez un numéro différent, vous pouvez quand même avoir la même boisson (si plusieurs rangées contiennent la même boisson par exemple).

Exercice 1.19. Prendre deux ensembles finis et définir une application entre eux.

Quelques définitions supplémentaires :

Définition 1.20 (Image directe). Soit $f : X \rightarrow Y$ une application et soit $A \subseteq X$. On appelle *image directe de A par f*, et on note $f(A)$, l'ensemble des éléments de Y qui ont un antécédent appartenant à A par f . Autrement dit, on a

$$f(A) = \{y \in Y \mid \text{il existe } x \in A \text{ tel que } y = f(x)\}.$$

On appelle *image de f*, et l'on note $\text{Im}(f)$, l'ensemble $f(X)$.

Définition 1.21 (Image réciproque). Soit $f : X \rightarrow Y$ une application et $B \subseteq Y$. On appelle *image réciproque de B par f*, et l'on note $f^{-1}(B)$, l'ensemble des éléments x de X tels que $f(x) \in B$. Autrement dit, on a

$$f^{-1}(B) = \{x \in X \mid f(x) \in B\}.$$

Définition 1.22 (Restriction). Soit $f : X \rightarrow Y$ une application et soit $A \subseteq X$. On appelle *restriction de f à A*, et l'on note $f|_A$, l'application $f|_A : A \rightarrow Y$ définie pour tout $a \in A$ par $f|_A(a) = f(a)$.

Définition 1.23 (Composition d'applications). Soient X, Y , et Z trois ensembles et soient $f : X \rightarrow Y$ et $g : Y \rightarrow Z$ deux applications. On définit l'*application composée de f par g à gauche* ou *application composée de g par f à droite*, et l'on note $g \circ f$, l'application h définie pour tout élément $x \in X$ par $h(x) = g(f(x))$.

Notation 1.24. Si X est un ensemble, on appelle *application identité de X* l'application notée Id_X définie pour tout élément $x \in X$ par $\text{Id}_X(x) = x$.

1.3.3 Comment se servir des applications pour notre problème ?

Pour l'instant, on a pu définir des applications d'un ensemble vers un ensemble plus petit, ou vers un ensemble plus grand, et le fait que l'on puisse définir une application d'un ensemble dans un autre ne nous donne pas d'information comparative de taille comme nous aurions pu le souhaiter. Une application générique n'est pas le lien d'infériorité que nous aurions souhaité.

Notre stratégie va donc être de définir des types d'applications particuliers qui joueront ce rôle de lien d'infériorité. Nous définirons "être plus petit que", "être plus grand que" et "être de la même taille" comme l'existence de certains types d'applications entre les deux ensembles. Nous aurons alors *généralisé* la notion intuitive de comparaison de taille dans le cas des ensembles finis aux cas des ensembles infinis.

1.3.4 Injection

Dans quels cas est-ce qu'il y a moins d'éléments dans un ensemble que dans un autre ?

Définition 1.25 (Injectivité). On dit qu'une application $f : X \rightarrow Y$ est *injective* si f ne fait jamais correspondre le même élément de Y à deux éléments distincts de X . Autrement dit, f est injective si :

Pour tous $x \in X$ et $x' \in X$ tels que $x \neq x'$, on a $f(x) \neq f(x')$.

Remarques 1.26.

- Une autre manière (c'est la contraposée) très utile de formuler la définition est de dire que f est injective si :

Pour tous $x \in X$ et $x' \in X$, on a que si $f(x) = f(x')$, alors $x = x'$.

- Encore une autre manière de formuler l'injectivité serait de dire que si un élément $y \in Y$ est image d'au moins un élément de X , ce ne peut être que d'un seul.

Exercice 1.27.

- Donner un exemple d'une application injective entre deux ensembles finis.
- Donner un exemple d'une application injective de \mathbb{N} dans \mathbb{N} .

Définition 1.28 (Subpotence). Soient deux ensembles X et Y . On dit que X est *subpotent* à Y , et l'on note $X \preceq Y$, s'il existe une injection $f : X \rightarrow Y$.

1.3.5 Surjection

Définition 1.29 (Surjectivité). On dit qu'une application $f : X \rightarrow Y$ est *surjective* si pour tout élément de $y \in Y$, il existe un élément de $x \in X$ tel que $y = f(x)$.

Remarque 1.30. Par définition, une application est surjective si, et seulement si $\text{Im}(f) = Y$. Autrement dit, il est équivalent de dire qu'une application f est surjective ou de dire que $\text{Im}(f) = Y$. On pourrait reformuler informellement la définition en disant que f est surjective si elle atteint tous les éléments de Y au moins une fois.

Exercice 1.31.

- Donner un exemple d'une application surjective entre deux ensembles finis.
- Donner un exemple d'une application surjective de \mathbb{N} dans \mathbb{N} .

Définition 1.32 (Surpotence). Soient deux ensembles X et Y . On dit que X est *surpotent* à Y , et l'on note $X \succcurlyeq Y$, s'il existe une surjection $f : X \rightarrow Y$.

1.3.6 Bijection

Définition 1.33 (Bijektivité). On dit qu'une application $f : X \rightarrow Y$ est *bijective* si elle est injective et surjective.

Remarque 1.34. Voici des reformulations de la définition :

- f est bijective si tout élément de Y est l'image d'un (surjectivité) et un seul (injectivité) élément de X par f .
- Ou encore, f est bijective si :
 - f est une application : pour tout élément $x \in X$, il existe un unique élément $y \in Y$ tel que $y = f(x)$, et
 - f est bijective : pour tout élément $y \in Y$, il existe un unique élément $x \in X$ tel que $y = f(x)$.

Lorsqu'il existe une bijection entre X et Y , elle induit une correspondance entre les éléments de X et les éléments de Y de sorte que chaque élément de X est lié à un unique élément de Y et réciproquement.

Définition 1.35 (Equipotence). Soient deux ensembles X et Y . On dit que X est *équipotent* à Y , et l'on note $X \sim Y$, s'il existe une bijection $f : X \rightarrow Y$.

On utilise aussi souvent la terminologie : X est en bijection avec Y .

Proposition 1.36. La relation d'équipotence \sim est transitive, c'est-à-dire qu'elle vérifie : pour tous ensembles X , Y et Z , si $X \sim Y$ et $Y \sim Z$, alors $X \sim Z$.

Exercice 1.37. Le démontrer.

Conclusion. Nous avons finalement abouti à une notion de comparaison de taille entre ensembles infinis. Nous dirons que deux ensembles ont le même cardinal s'ils sont équipotents, i.e. s'il existe une bijection de l'un vers l'autre. Cette méthode permet de comparer des ensembles de même nature et, contrairement à la méthode naive consistant à compter le nombre d'éléments de chaque ensemble, elle a un sens dans le cas d'ensembles infinis. Le processus de comptage d'éléments ne termine jamais pour un ensemble infini, alors que l'on peut très bien définir en temps fini une bijection entre deux ensembles infinis, comme vous allez le voir dans la suite.

1.3.7 Reformulation de nos questions initiales

Donnons des noms aux ensembles dont nous parlions au début de ce cours. L'ensemble des entiers naturels est noté \mathbb{N} et l'ensemble des nombres pairs positifs $2\mathbb{N}$. Voici les questions que nous allons chercher à résoudre :

1. Est-ce que \mathbb{N} et $2\mathbb{N}$ sont équipotents ?
2. Est-ce que \mathbb{N} et l'ensemble des points sur une droite sont équipotents ?

1.3.8 Définition des ensembles finis et infinis

Avant de répondre à ces questions, faisons un petit aparté. Depuis le début de ce cours, nous parlons d'ensembles finis, d'ensembles infinis, sans jamais avoir donné une définition de ces termes. Avec les nouveaux outils que nous venons de définir, il est désormais possible d'en donner une qui soit satisfaisante.

Définition 1.38 (Ensemble fini). On dit qu'un ensemble E est *fini* s'il existe un entier $n \in \mathbb{N}$ tel que E soit en bijection avec l'ensemble $\{1, \dots, n\}$.

Proposition 1.39. *S'il existe,⁹ l'entier $n \in \mathbb{N}$ de la définition précédente est unique, et on l'appelle le cardinal de E .*

Définition 1.40 (Ensemble infini). On dit qu'un ensemble E est *infini* s'il n'est pas fini.

1.4 Il y a autant de nombres pairs que de nombres entiers

Théorème 1.41. *L'ensemble \mathbb{N} est équipotent à l'ensemble $2\mathbb{N}$.*

Démonstration. Considérer l'application

$$f : \begin{array}{l} \mathbb{N} \rightarrow 2\mathbb{N} \\ n \mapsto 2n \end{array}$$

et vérifier qu'elle définit bien une bijection entre \mathbb{N} et $2\mathbb{N}$. □

1.5 Il y a plus de points sur une droite que de nombre entiers

1.5.1 Court rappel sur les nombres réels

L'ensemble des nombres réels, noté \mathbb{R} , est le plus grand ensemble de nombres que vous avez pu rencontrer au cours de votre scolarité. C'est l'ensemble de tous les nombres que vous connaissez : les entiers positifs, les négatifs, les fractions, les *irrationnels*, π , ... tous !

Si l'on considère par exemple les deux entiers 0 et 1, il existe des nombres réels entre les deux comme 0.1 ou 0.43543. Mais il en existe aussi avec une infinité de nombres après la virgule comme $\frac{1}{3} \simeq 0.3333\dots$ ou encore $\frac{2}{7} \simeq 0.28571428571\dots$. Il existe également des nombres qui ne peuvent pas s'écrire sous forme d'une fraction de deux entiers comme $\sqrt{2}/2 \simeq 0.707106781\dots$ par exemple. En fait, tous les nombres compris en 0 et 1 forment

9. c'est-à-dire si l'ensemble E est fini

un continuum..., il y en a une infinité, et "il n'y a pas de trou" : les tracer tous revient à tracer *une segment de droite*. Ces termes sont flous, et n'ont pour but que de donner une vague intuition, mais ils ont un sens bien précis en mathématiques. Les définir dépasserait malheureusement le cadre de ce cours.

Il y a deux choses à retenir pour ce cours :

1. l'ensemble des nombres réels peut être représenté par une droite, *la droite réelle*, et il y a donc autant de point sur une droite que de nombres dans \mathbb{R} . En particulier, **résoudre la question 1.3 est équivalent à se demander si les ensembles \mathbb{N} et \mathbb{R} sont équipotents.**
2. Tout nombre réel admet un unique¹⁰ *développement décimal*, c'est-à-dire une unique écriture à virgule (potentiellement infinie).

Au Palais de la découverte, il y a une salle avec les décimales de π inscrites sur les murs.

1.5.2 $\mathbb{N} \preccurlyeq \mathbb{R}$

Tout d'abord, notons que puisque $\mathbb{N} \subseteq \mathbb{R}$, on peut définir une injection de \mathbb{N} dans \mathbb{R} par

$$\begin{aligned} \mathbb{N} &\rightarrow \mathbb{R} \\ n &\mapsto n \end{aligned}$$

En fait, on a même le théorème suivant :

Théorème 1.42. *Il existe une injection de l'ensemble \mathbb{N} dans l'ensemble $[0, 1] = \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$.*

Démonstration. Considérons l'application

$$f : \begin{aligned} \mathbb{N} &\rightarrow [0, 1] \\ n &\mapsto \begin{cases} 0 & \text{si } n = 0, \\ \frac{1}{n} & \text{si } n > 1, \end{cases} \end{aligned}$$

et montrons qu'elle est injective. Soient $n \in \mathbb{N}$ et $n' \in \mathbb{N}$ tels que $f(n) = f(n')$. Montrons que $n = n'$. Si $f(n) = f(n') = 0$, alors c'est que $n = n' = 0$ car pour tout $m \geq 1$, $\frac{1}{m} \neq 0$. Si maintenant $f(n)$ et $f(n')$ sont différents de 0, c'est que n et n' sont différents de 0 aussi, et on a que $\frac{1}{n} = \frac{1}{n'}$ par définition de f , ce qui équivaut à $n = n'$. \square

1.5.3 Est-ce que $\mathbb{N} \sim \mathbb{R}$?

On va utiliser l'argument diagonal de Cantor, une idée extrêmement riche, publié en 1891 par le mathématicien Georg Cantor, pour démontrer que \mathbb{R} et \mathbb{N} ne sont pas équipotents.

10. Ce n'est pas vrai, mais presque.

Théorème 1.43 (Cantor 1874). *Les ensembles \mathbb{N} et \mathbb{R} ne sont pas équipotents. Autrement dit, il n'existe pas de bijection entre \mathbb{N} et \mathbb{R} .*

Démonstration. On va montrer qu'il n'existe pas de bijection entre $]0, 1[$ ¹¹ et \mathbb{N} .

Ceci impliquera alors que \mathbb{N} n'est pas équipotent à \mathbb{R} . En effet, nous l'admettrons ici, mais nous avons une bijection $]0, 1[\sim \mathbb{R}$, donc si nous avons $\mathbb{N} \sim \mathbb{R}$ alors nous aurions $\mathbb{N} \sim \mathbb{R} \sim]0, 1[$ et donc $\mathbb{N} \sim]0, 1[$, ce dont nous allons précisément montrer le contraire.

Pour montrer qu'il n'existe pas de bijection entre \mathbb{N} et $]0, 1[$, nous allons montrer qu'il ne peut en fait déjà pas exister de surjection de \mathbb{N} sur $]0, 1[$.

Supposons par l'absurde qu'il existe une surjection $f : \mathbb{N} \rightarrow]0, 1[$. On cherche à aboutir à une contradiction logique (que quelque chose et son contraire soient vrais). Ceci montrera alors qu'il ne peut exister de telle surjection.

Comme f est une application, elle associe à chaque entier $n \in \mathbb{N}$ un nombre $f(n)$ entre 0 et 1. On ne sait pas quel nombre f associe à chaque entier, mais on sait qu'ils peuvent tous s'écrire de manière unique sous la forme décimale $0, \dots$ avec une certaine suite de chiffres après la virgule. De plus, à un nombre réel entre 0 et 1 est associé une unique suite, et réciproquement, à une suite de décimales, un unique nombre réel entre 0 et 1.

Par exemple, f pourrait être :

$$\begin{aligned} f(0) &= 0,5 \\ f(1) &= 0,7952838236253 \\ f(2) &= \pi/4 \simeq 0,78539816339\dots \\ f(3) &= \sqrt{2} - 1 \simeq 0,41421356\dots \\ f(4) &= \frac{2}{7} \simeq 0,28571428571\dots \\ f(5) &= \frac{1}{3} \simeq 0,3333\dots \\ f(6) &= 0,1 \\ f(7) &= 0,48 \\ f(8) &= 0,34 \\ &\vdots \end{aligned}$$

Comme f est surjective, chaque nombre $x \in]0, 1[$ a un antécédent par f . Autrement dit, il existe un $n \in \mathbb{N}$ tel que $f(n) = x$. Pour montrer une absurdité, on va en fait montrer que f n'atteint pas tous les éléments de $]0, 1[$.

Définissons un nombre z , qui n'est pas atteint par f . On définit z par son écriture décimale. On choisit les décimales de z de sorte que z ait au moins une décimale de différence avec chacun des $f(n)$. Ainsi, z ne pourra pas être dans l'image de f ,

11. L'ensemble $]0, 1[$ désigne l'ensemble des nombres $x \in \mathbb{R}$ tels que $0 < x < 1$, c'est le segment ouvert des nombres réels compris strictement entre 0 et 1.

puisque sinon il serait égal à un certain $f(m)$ pour un $m \in \mathbb{N}$, et il aurait alors les mêmes décimales que lui.

Exemple Commençons par construire z dans le cas de l'exemple précédent de manière illustrée. Ecrivons pour cela les décimales de tous les $f(n)$ dans un grand tableau.

$$\begin{aligned} f(0) &= 0, \mathbf{5}000000000 \dots \\ f(1) &= 0, \mathbf{79}528382362 \dots \\ f(2) &= 0, \mathbf{785}39816339 \dots \\ f(3) &= 0, \mathbf{4142}1356237 \dots \\ f(4) &= 0, \mathbf{28571}428571 \dots \\ f(5) &= 0, \mathbf{333333}33333 \dots \\ f(6) &= 0, \mathbf{1000000}0000 \dots \\ f(7) &= 0, \mathbf{48000000}000 \dots \\ f(8) &= 0, \mathbf{340000000}00 \dots \\ &\vdots \end{aligned}$$

Pour être sûr que z a toujours une décimale de différence avec chacun des $f(n)$, on définit les décimales de z comme étant celles qui sont dans la diagonale de ce tableau (celles en rouge et en gras), mais auxquelles on fait subir d'abord une transformation, ce qui nous assure que les décimales de z seront différentes de celles de chacun des $f(n)$.

La transformation qu'on fait subir aux décimales de la diagonale avant de les placer dans l'écriture décimale de z est la suivante : si la décimale en rouge est différente de 9, on lui ajoute 1, si c'est 9, on la remplace par 1. Si l'on applique ce procédé pour définir z dans notre exemple, on obtient $z = 0,616324111\dots$

On peut alors vérifier que le procédé fonctionne : z peut-il être égal à $f(3)$? Non, puisque par construction, la 4^{ème} décimale de z est différente de la 4^{ème} décimale de $f(3)$. En effet, la 4^{ème} décimale de z est la 4^{ème} décimale de $f(3)$ à laquelle on a ajouté 1. Comme z et $f(3)$ ont une décimale de différence, ils sont différents. Par construction de z , ce sera vrai pour tous les $f(n)$ pour $n \in \mathbb{N}$, et z ne peut donc pas être dans l'image de f .

Généralisation Dans le cas général où les $f(n)$ ne sont pas donnés par les valeurs de l'exemple précédent, on peut tout de même définir le même procédé que dans l'exemple pour définir z : la n -ième décimale de z est définie de la manière suivante :

- ou bien comme la n -ième décimale¹² de $f(n-1)$ auquel on ajoute 1 si cette décimale n'est pas 9,
- ou bien comme 1 si la n -ième décimale de $f(n-1)$ est 9.

12. La n -ième décimale de $f(n-1)$ est la décimale de $f(n-1)$ se trouvant sur la diagonale, en rouge dans l'exemple.

Dans ce cas, z est bien un nombre réel entre 0 et 1. Cependant, il n'est pas dans l'image de f , car il a forcément au moins une décimale de différence avec $f(n)$ pour tout $n \in \mathbb{N}$. Il ne peut donc être aucun des $f(n)$. Ainsi, f n'est pas surjective. C'est une contradiction : nous avons supposé f surjective.

Ainsi, il ne peut exister de surjection de \mathbb{N} sur $]0, 1[$, donc encore moins de bijection entre eux, et donc, comme expliqué en début de cette preuve, \mathbb{N} et \mathbb{R} ne peuvent être équipotents. \square

1.6 Appendice : Paradoxe de Russel

Etablir une définition des ensembles qui soit assez rigoureuse pour ne pas donner lieu à des contradictions logiques est une tâche extrêmement difficile. Pour preuve, voici plus de détail sur le paradoxe de Russel. Dans ses *Fondements de l'arithmétique*, Gottlob Frege (mathématicien, logicien et philosophe allemand) souhaitait construire une théorie sur laquelle on pourrait fonder toutes les mathématiques. Cependant, Bertrand Russel lui adresse une lettre en 1902 dans laquelle il montre que l'une des bases de la théorie de Frege mène à une contradiction, rendant alors la théorie inutilisable. Frege publie tout de même son livre et écrit en annexe après avoir énoncé le paradoxe de Russel :

Pour un écrivain scientifique, il est peu d'infortunes pires que de voir l'une des fondations de son travail s'effondrer alors que celui-ci s'achève. C'est dans cette situation inconfortable que m'a mis une lettre de M. Bertrand Russell, alors que le présent volume allait paraître.

Quel est donc ce paradoxe ? Considérons une définition naïve des ensembles, comme par exemple : *un ensemble est une collection d'éléments munie d'un critère qui permet de déterminer si un élément est ou non dans l'ensemble*. Etudions les concepts que recouvre cette définition.

Considérons l'ensemble de tous les ensembles. Nous avons bien un ensemble au sens de notre définition : la propriété est : *"être un ensemble"*. Nous avons alors par exemple $\mathbb{N} \in E$, $\{\text{"zoo"}, 0, 1\} \in E$ et $\mathbb{R} \in E$. Oui mais... E est un ensemble n'est-ce pas ? On a donc aussi $E \in E$. Il s'appartient à lui-même. C'est un peu singulier pour un ensemble, non ? Cela n'est pas contradictoire a priori, mais ne correspond en tout cas pas à l'intuition que nous en avons car si notre intuition est celle d'un contenant, nous avons du mal à imaginer un objet étant à la fois le contenant et son contenu.

Considérons l'ensemble des ensembles qui ne s'appartiennent pas à eux-mêmes, et notons le A . On a alors qu'un ensemble X appartient à A si X n'appartient pas à X . En particulier, on a $\mathbb{N} \in A$, $\mathbb{R} \in A$, mais $E \notin A$. Mais voilà, la définition bancale que nous avons choisie va révéler sa véritable nature... En fait, A ne peut pas exister, car cela mènerait à une absurdité, c'est-à-dire à une proposition vraie et fausse à la fois.

Proposition 1.44. *L'existence de A est contradictoire.*

Démonstration. La preuve repose sur la question suivante : "Ce dernier ensemble s'appartient-il à lui-même?". Supposons par l'absurde qu'il existe un ensemble A qui contienne exactement les ensembles qui ne s'appartiennent pas à eux-mêmes. Nous allons étudier tous les cas qui peuvent alors se produire :

- Premier cas : $A \in A$, A s'appartient à lui-même, mais donc par la définition de A , il ne s'appartient pas : $A \notin A$.
- Second cas : $A \notin A$, mais alors par la définition de A , on a $A \in A$.

Les deux possibilités sont absurdes, A n'existe donc pas. □

1.7 Références

- Patrick Dehornoy, mathématicien ayant travaillé dans le domaine et bon vulgarisateur, est un auteur de référence sur le sujet de la théorie des ensembles. Il a fait de très bonnes [conférences grand public](#) et a écrit de nombreux [articles grand public](#). Je vous conseille en particulier la vidéo [Deux malentendus sur la théorie des ensembles](#) et les articles *Cantor et les infinis* et *L'infini est-il nécessaire ?*.
- Le livre sur lequel je me suis principalement appuyé pour la partie sur l'histoire de la théorie des ensembles est le suivant : Jean-Pierre Belna, *Histoire de la théorie des ensembles*, Volume 47 de Esprit des sciences, Éditeur Ellipses, 2009.
- Pour une version plus détaillée de la construction des la théorie axiomatique des ensembles, avec les formules et les axiomes écrits *in extenso*, voir la deuxième partie du très bon [polycopié du cours](#) *Culture mathématique* de Jean Feydy. Attention : ce polycopié, sans suivre le cours auquel il était rattaché, est plus difficile que le présent cours. Le polycopié de Jean Feydy contient également des références plus poussées pour approfondir votre connaissance de la théorie axiomatique des ensembles.
- Une [très bonne vidéo](#) sur le sujet de David Louapre, sur sa chaîne ScienceEtonnante.

Chapitre 2

Suites et limites

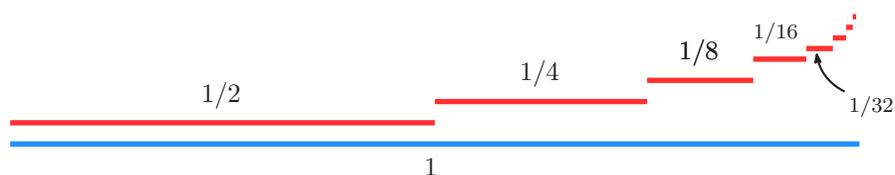


FIGURE 2.1 – Division du segment de longueur 1 en une infinité d'intervalles de longueur de plus en plus petite.

2.1 Introduction

Dans ce chapitre, nous continuons notre enquête sur les mystères de l'infini. Comme au chapitre précédent, notre but sera de donner un sens précis à une expression qui n'en a pas. Nous avons vu qu'avoir "la même taille" n'avait pas de sens pour des ensembles infinis, du moins pas avant d'en avoir donné une définition non-ambigüe. Désormais, nous nous attaquerons à l'expression "et ainsi de suite, indéfiniment", dont une utilisation connue est celle qu'en fait Zénon dans ses paradoxes.

2.1.1 Paradoxe de la dichotomie

Dans sa *Physique*, Aristote rapporte ainsi le *paradoxe de la dichotomie* dû à Zénon :

« [Zénon] prétend prouver que le mouvement n'existe pas, attendu que le mobile passe par la moitié avant d'arriver à la fin. »

Aucun usage de l'expression ambigüe en apparence ici, mais en réalité l'argument de Zénon pourrait se reformuler ainsi :

Supposons le mouvement possible. Pour me déplacer d'une distance de 1 mètre, je dois d'abord me déplacer d'une distance de $1/2$ mètre, puis à nouveau de $1/4$ mètre, puis de $1/8$ mètre, et ainsi de suite, *indéfiniment*. Le temps nécessaire au parcours de chacune des fractions de la distance étant non nul, le voyage nécessite une infinité de déplacements de durée non nulle, et ne terminera donc jamais. Le mouvement est donc impossible.

On a illustré ce raisonnement en Figure 2.2. Cependant, on peut argumenter contre Zénon que le temps nécessaire au parcours de fractions de plus en plus petites de la distance sera lui-même de plus en plus court, de sorte que la somme infinie de ces durées diminuant sans cesse puisse donner une valeur finie, précisément celle du temps qu'il me faut pour parcourir cette distance.

Le problème que nous nous posons alors ici est le suivant :

Comment calculer une somme infinie de nombres ?

Mathématiquement, la somme de deux nombres est bien définie : dans le cadre de la théorie des ensembles, on peut définir les nombres et donner un sens à leur addition. On peut même donner un sens à l'addition d'un nombre fini de nombres (addition de trois nombres à la fois, quatre nombres à la fois, etc...). En revanche, donner un sens

à l'addition d'un nombre infini de nombres est bien plus subtil, comme nous allons le voir dans les sections qui suivent.

2.1.2 Quelle valeur ont les sommes infinies ?

Prenons un exemple, pour étudier le paradoxe de Zénon sur des valeurs concrètes de distance et de temps. Considérons un déplacement à une vitesse de 1 km/h . L'expérience montre qu'il faut alors 1 heure pour parcourir 1 kilomètre, et non une infinité de temps comme le suggère Zénon. Cependant, suivons son raisonnement, et décomposons comme lui la durée de notre déplacement :

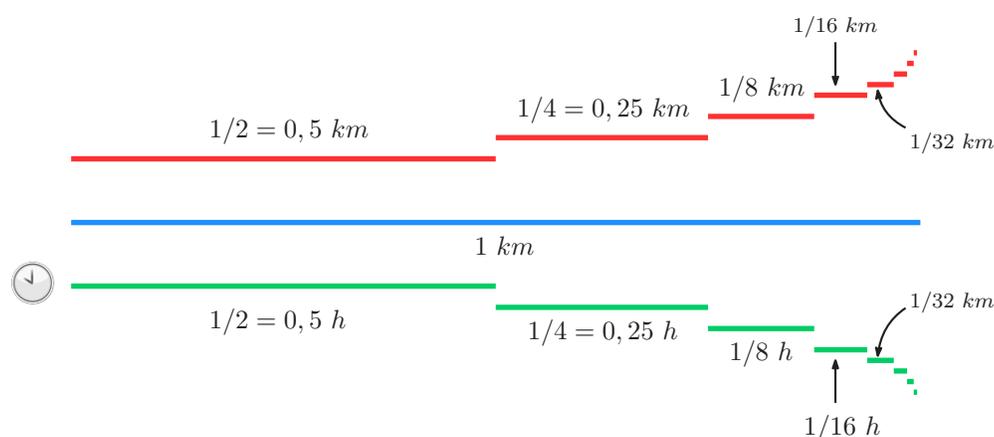


FIGURE 2.2 – Décomposition temporelle du mouvement comme le suggère Zénon.

De sorte que le temps que nous mettons à parcourir 1 km s'écrit selon cette décomposition :

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \dots$$

Deux diagnostics s'opposent : Zénon affirme que cette somme a une valeur infinie, alors que l'expérience (ou le bon sens) tendrait plutôt à nous montrer que cette somme vaut 1, la durée de notre parcours. En équation, cela donne :

$$1 = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \dots$$

Mais alors, qui a raison ?

Pour tenter de le découvrir, faisons le calcul nous même : ajoutons chacun des termes de la somme un à un, et calculons à chaque étape la valeur obtenue. Si Zénon a raison, la somme ne devrait jamais cesser d'augmenter, sans aucune borne, au fil des étapes. Si c'est le "bon sens", alors cette somme devrait se rapprocher de plus en plus de la valeur 1 au fil des étapes.

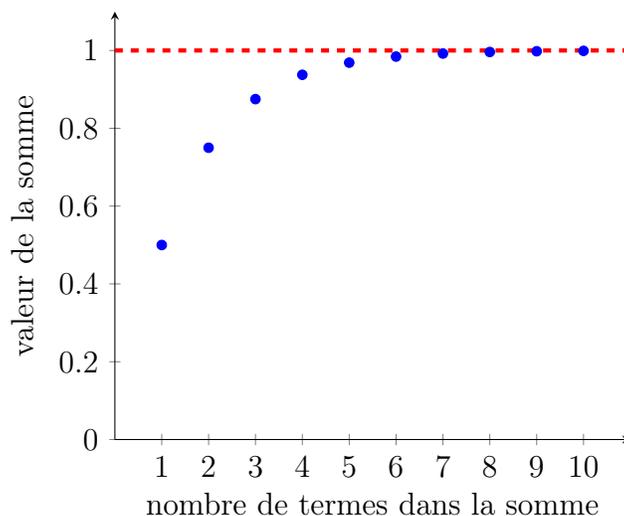


FIGURE 2.3 – Graphe de la somme des distances apparaissant dans le paradoxe de Zénon.

Somme	Total
$1/2 = 0,5$	0,5
$+1/4 = 0,25$	0,75
$+1/8 = 0,125$	0,875
$+1/16 = 0,0625$	$\sim 0,938$
$+1/32 = 0,03125$	$\sim 0,969$
$+1/64 = 0,015625$	$\sim 0,984$
$+1/128 = 0,0078125$	$\sim 0,992$
$+1/256 = 0,00390625$	$\sim 0,996$

Le calcul ci-dessus, ainsi que la figure 2.3 tendent à montrer que cette somme vaut 1, et que c'est donc notre bon sens qui avait raison.

Conclusion. Cet exemple pourrait nous laisser croire qu'une somme infinie de nombres de plus en plus petits a toujours une valeur finie. Cependant, cette impression est trompeuse, comme le montre la section suivante.

2.1.3 Deux sommes divergentes

Considérons la somme infinie suivante :¹

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \dots$$

1. On l'appelle la *série harmonique*.

Elle est très similaire à la somme considérée dans la partie précédente : c'est une somme infinie de nombres de plus en plus petits. Voyons ce que donne l'ajout de termes successifs :

Somme	Total
$1/2 = 0,5$	0,5
$+1/3 \sim 0,333$	$\sim 0,833$
$+1/4 = 0,25$	$\sim 1,083$
$+1/5 = 0,2$	$\sim 1,283$
$+1/6 \sim 0.167$	1,45
$+1/7 \sim 0.143$	$\sim 1,593$
$+1/8 = 0.125$	~ 1.718
$+1/9 \sim 0.111$	$\sim 1,829$

La somme semble augmenter, mais on ne peut pas vraiment se prononcer quant à sa divergence. Elle pourrait très bien converger vers 2. Calculons plus de termes (cf. figure 2.4).

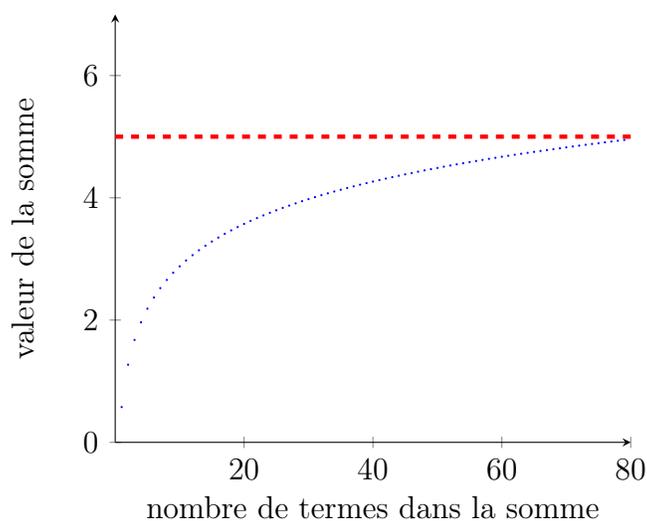


FIGURE 2.4 – Graphe des 80 premiers termes de la série harmonique.

Cette somme tenderait donc vers 5 ? Pourtant non, regardez la figure 2.5.

En fait, la somme diverge : plus l'on ajoute de termes, et plus la somme augmente, sans aucune borne. Il faut donc se méfier des sommes infinies de nombres de plus en plus petits, certaines semblent converger, d'autres diverger.

Puisque notre somme diverge, ça n'a pas vraiment de sens d'écrire :

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \dots$$

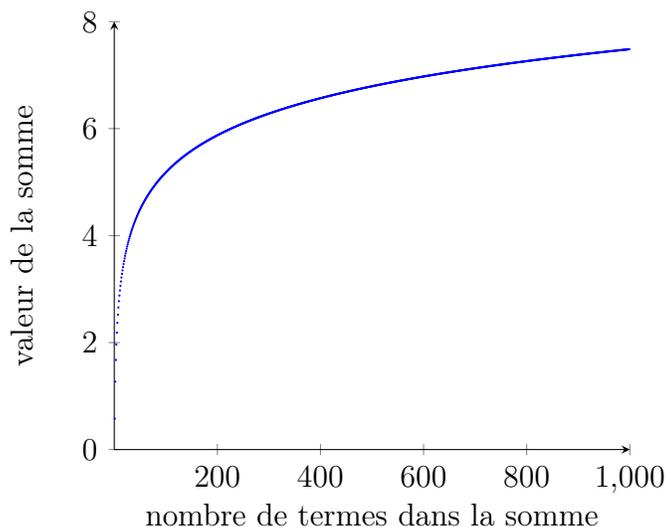


FIGURE 2.5 – Graphe des 1000 premiers termes de la série harmonique.

Sauf si on l’entend comme étant égale à l’infini peut-être ? Mais lequel ? Nous pourrions décréter que les sommes qui ne cessent d’augmenter sans aucune borne sont égale à une valeur, plus grande que tous les nombres et que l’on nommerait *infini*.

Très bien, mais considérez cette somme :²

$$1 - 1 + 1 - 1 + 1 - 1 + 1 - 1 + \dots$$

Elle vaut alternativement 1 et 0. Peut-on lui assigner une valeur ? Si oui, laquelle : 1, 0 ou 0.5 ? On pourrait choisir une des valeurs précédentes. Seulement, s’il est facile de se fixer des définitions, il est plus difficile de s’en fixer qui soient intéressantes. La question qu’il faut se poser lorsque l’on définit une notion en mathématiques est : est-ce que cette définition correspond bien à la réalité que je souhaitais décrire ? Est-ce que cette définition permet de montrer des théorèmes qui soient intéressants ? Par exemple, si je multiplie deux sommes convergentes entre elles, vont-elles converger vers le produit de leur valeurs ? Si oui, comment le montrer ? Il est raisonnable d’exiger, puisque c’est notre intuition, que la notion de somme que l’on va définir satisfasse par exemple cette propriété sur le produit, et des propriétés similaires pour toutes les opérations que l’on peut faire (addition, division, ...). Les calculs qui suivent montrent comme nos définitions hâtives sont bancales.

2. On l’appelle la *série de Grandi*, du nom du mathématicien, philosophe et prêtre italien Luigi Guido Grandi (1671 - 1742).

Notons $S = 1 - 1 + 1 - 1 + 1 - 1 + 1 - 1 + \dots$. On peut alors calculer naïvement :

$$\begin{aligned} 2S &= S + S \\ &= 1 - 1 + 1 - 1 + 1 - 1 + 1 - 1 + 1 + \dots \\ &\quad + 1 - 1 + 1 - 1 + 1 - 1 + 1 - 1 + \dots \\ &= 1 + 0 + 0 + 0 \dots \\ &= 1 \end{aligned}$$

Et donc $2S = 1$, soit $S = 0.5$. Mais l'on peut tout aussi calculer :

$$\begin{aligned} S &= (1 - 1) + (1 - 1) + (1 - 1) + (1 - 1) + \dots \\ &= 0 + 0 + 0 + 0 \dots \\ &= 0 \end{aligned}$$

ou bien

$$\begin{aligned} S &= 1 + (-1 + 1) + (-1 + 1) + (-1 + 1) + \dots \\ &= 1 + 0 + 0 + 0 \dots \\ &= 1 \end{aligned}$$

Horreur ! Trois calculs naïfs différents donnent trois valeurs différentes possibles pour la valeur de cette somme...

Conclusion. Nous avons vu trois exemples de sommes infinies : une (Zénon) qui semblait converger, une autre (harmonique) qui semblait augmenter sans borne et une dernière (Grandi) à laquelle il est difficile d'attribuer une valeur. Il nous faut une définition de somme infinie qui tienne la route — c'est-à-dire écrite dans un vocabulaire mathématique et correspondant à l'intuition que nous avons du concept de convergence — pour que nous puissions trouver des réponses satisfaisantes à nos questions initiales.

2.2 Plan du chapitre

L'objectif de ce chapitre est de donner un sens propre à la valeur d'une somme infinie. Pour ce faire, nous allons procéder en trois étapes :

1. Nous définirons d'abord la notion mathématique modélisant les processus qui se déroulent par étape,³ appelés *suites*,
2. Nous définirons ensuite une notion propre de *convergence* pour les suites,
3. Nous appliquerons (entre autres applications) cette théorie au cas du paradoxe de Zénon, pour démontrer mathématiquement l'équation :

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \dots = 1.$$

Remarque 2.1. Les notions de suites et de convergence seront également fondamentales pour faire des statistiques au Chapitre 3!

3. comme les sommes précédentes auxquelles nous ajoutons un nouveau terme à chaque étape.

2.3 Suite numérique

2.3.1 Peu d'Histoire

Les suites numériques font parties des objets mathématiques les plus anciens dont il nous reste des traces aujourd'hui. Elles modélisent les processus qui se déroulent par étape (comme nos sommes auxquelles on ajoute un nouveau terme à chaque étape). Il semble qu'elles aient été utilisées dès l'Égypte antique dans le but de calculer des approximations de nombres, de la même manière que notre série géométrique approchait de plus en plus le nombre 1. Pour calculer le nombre π par exemple, qui est égal au périmètre d'un cercle de diamètre 1, le mathématicien antique Archimède (287 av. J.-C. – 212 av. J.-C.) eu l'idée de calculer le périmètre de polygones réguliers dont les sommets sont sur le cercle. Plus les polygones ont de sommets, plus ils sont proches du cercle et donc plus leur périmètre est proche de celui du cercle. En ajoutant un sommet à chaque étape, on obtient une *suite* de nombres (les périmètres des polygones) qui se rapprochent de plus en plus du périmètre du cercle, c'est-à-dire du nombre π . Lorsque l'on s'arrête à une certaine étape, le périmètre du polygone à cette étape donne une approximation du nombre π .

Nous allons voir une autre utilisation théorique des suites lors du chapitre 3 où nous parlerons d'estimateurs statistiques. D'un point de vue plus pratique, nous verrons tout au long de ce chapitre un exemple concret et d'actualité : la modélisation de l'évolution d'une épidémie.

2.3.2 Définition et premiers exemples

Définition 2.2. On appelle *suite réelle* toute application de \mathbb{N} dans \mathbb{R} .

Lorsque l'on travaille avec les suites numériques, on utilise un nouveau vocabulaire pour parler des mêmes choses⁴.

Notation 2.3. Si $u : \mathbb{N} \rightarrow \mathbb{R}$ est une suite, on note u_n l'image de n par u plutôt que $u(n)$ comme il est d'usage pour les applications. De plus, on appelle u_n le *n -ième terme de la suite u* , plutôt que l'image de n par u comme il est d'usage pour les applications.

Exemples 2.4.

1. La suite constante égale à 1 est définie par :

$$u : \begin{array}{l} \mathbb{N} \rightarrow \mathbb{R} \\ n \mapsto 1 \end{array} .$$

2. La *suite identité* est définie par :

$$v : \begin{array}{l} \mathbb{N} \rightarrow \mathbb{R} \\ n \mapsto n \end{array} .$$

4. à mon avis pour des raisons historiques.

3. Pour un nombre réel $q \in \mathbb{R}$, on définit la *suite géométrique de raison q* par

$$x : \begin{array}{l} \mathbb{N} \rightarrow \mathbb{R} \\ n \mapsto q^n \end{array} .$$

4. On peut également définir la suite $y : \mathbb{N} \rightarrow \mathbb{R}$ définie par :

$$y : \begin{array}{l} \mathbb{N} \rightarrow \mathbb{R} \\ n \mapsto \begin{cases} 0 & \text{si } n = 0, \\ \frac{1}{n} & \text{si } n > 1. \end{cases} \end{array}$$

Exercice 2.5. Calculer les 4 premiers termes des suites précédentes.

2.3.3 Termes de la série géométrique (calcul de Zénon)

On peut décrire le chemin parcouru lors du déplacement décrit dans le paradoxe de Zénon par une suite numérique :

- après le premier déplacement, la distance parcourue est de $d_1 = 1/2$,
- après le deuxième déplacement, la distance parcourue est de $d_2 = 1/2 + 1/4$
- après le troisième déplacement, la distance parcourue est de $d_3 = 1/2 + 1/4 + 1/8$,
- ...

En fait, on peut réécrire ces termes ainsi :

$$\begin{aligned} d_1 &= \frac{1}{2^1}, \\ d_2 &= \frac{1}{2^1} + \frac{1}{2^2}, \\ d_3 &= \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^3}. \end{aligned}$$

La distance parcourue après le n -ième déplacement est alors de :

$$d_n = \frac{1}{2} + \cdots + \frac{1}{2^n}. \tag{2.1}$$

Ceci définit bien une suite $d : \mathbb{N} \rightarrow \mathbb{R}$.

2.3.4 Propagation d'une épidémie

En guise d'exemple, nous allons modéliser la propagation d'une épidémie. Supposons qu'une toute nouvelle maladie ait été détectée en France, transformant lentement la population française en zombies affamés. Notre but est de prédire le nombre total de personnes qui deviendront des zombies d'ici la fin de l'épidémie. Avant que l'épidémie ne démarre (la semaine 0), on supposera n'avoir qu'une seule personne est infectée (le patient 0). La maladie étant contagieuse, ce patient contaminera un certain nombre de

personnes, et ces personnes de nouvelles, et ainsi de suite, avec de nouvelles personnes à chaque nouvelle *génération* de malades. Nous allons devoir compter le nombre de personnes de chaque génération, que l'on modélise par la suite numérique suivante :

$$\begin{aligned} \mathbb{N} &\longrightarrow \mathbb{R} \\ p : n &\longmapsto p_n := \text{nombre de personnes de la} \\ &\quad n\text{-ième génération de malades.} \end{aligned}$$

Comme on l'a dit précédemment, on a un seul patient à la génération 0, avant que l'épidémie ne démarre, i.e.

$$p(0) = 1.$$

Comment donner une expression plus explicite de la suite p ? Les modélisations les plus simples d'évolution d'une épidémie introduisent un paramètre réel, noté r , représentant le nombre moyen de personnes qu'une personne malade infectera à son tour. La n -ième génération compte un nombre p_n de personnes, et ces personnes vont infecter $r \cdot p_n$ nouvelles personnes. On aura donc un nombre de personnes à la génération $n + 1$:

$$p_{n+1} = r p_n$$

On peut alors calculer les premiers termes de la suite p :

$$\begin{aligned} p_1 &= r, \\ p_2 &= r^2, \\ p_3 &= r^3, \\ &\vdots \end{aligned}$$

et on peut en fait montrer le fait intuitif suivant : pour tout $n \in \mathbb{N}$,

$$p_n = r^n. \tag{2.2}$$

C'est-à-dire que la suite p est une suite géométrique de raison r . Selon que r est plus grand ou plus petit que 1, la suite p croît ou décroît (cf figure 2.6). Ainsi, le nombre de personnes qu'aura touché l'épidémie au bout de n générations d'infections vaudra :

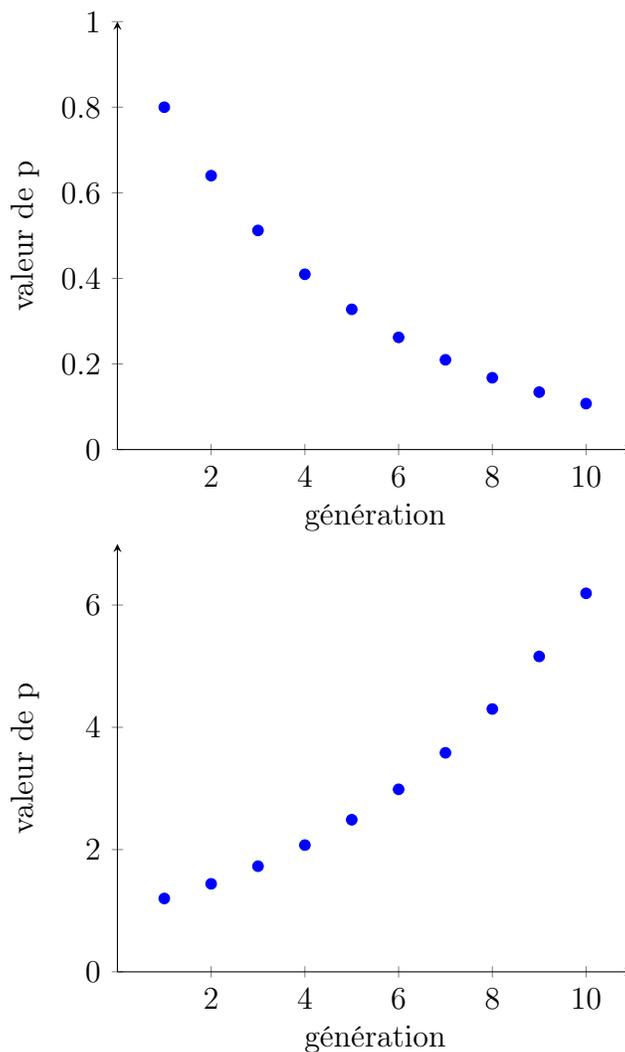
$$p_0 + p_1 + p_2 + \cdots + p_n = 1 + r + r^2 + \cdots + r^n.$$

La question que l'on se pose, et à laquelle on répondra grâce aux outils introduits dans ce chapitre est la suivante :

Question. *Combien de personnes seront touchées par l'épidémie au total?*

2.4 Limites

Nous avons vu en introduction que la suite d correspondant à la distance parcourue dans le déplacement du paradoxe de Zénon semblait converger vers la valeur 1, sans donner véritablement de sens à cette notion de convergence. C'est que ce nous allons faire maintenant.

FIGURE 2.6 – Graphe de p pour $r = 0,8$ (haut) et $r = 1,2$ (bas).

Définition 2.6. Soit $u : \mathbb{N} \rightarrow \mathbb{R}$ et $\ell \in \mathbb{R}$. On dit que u converge vers ℓ , ou que u tend vers ℓ quand n tend vers l'infini, et l'on note $u_n \xrightarrow[n \rightarrow +\infty]{} \ell$, si :

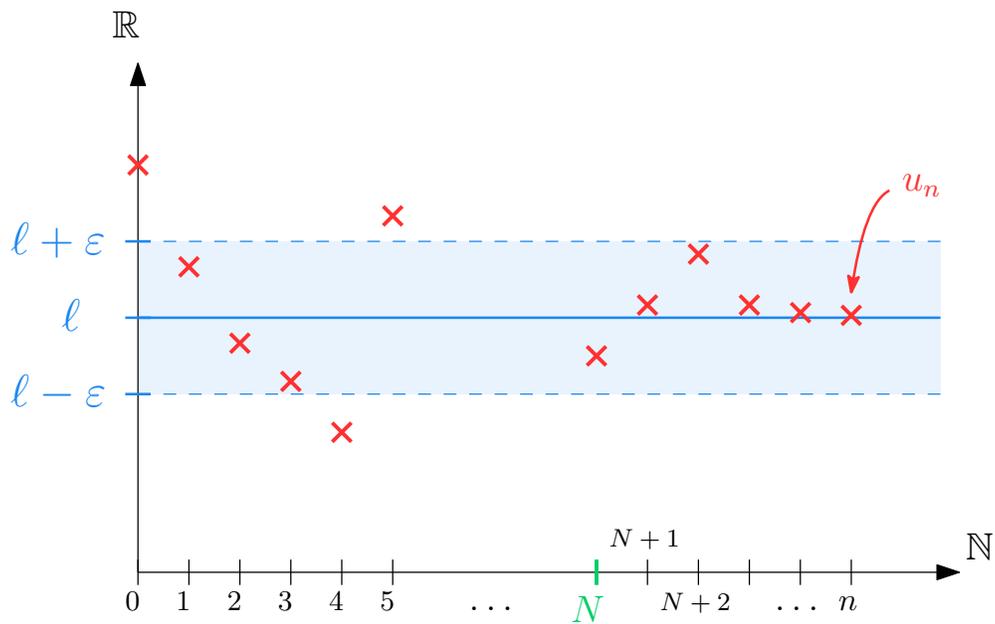
pour tout réel $\varepsilon > 0$, il existe $N \in \mathbb{N}$, tel que pour tout entier $n \geq N$,

$$\ell - \varepsilon < u_n < \ell + \varepsilon.$$

On a illustré cette définition en Figure 2.7.

Remarque 2.7. Il y a différentes façons informelles de reformuler la définition précédente :

- u finira par être à distance aussi proche que l'on veut de ℓ .
- peu importe la distance ε , il existe un rang à partir duquel u sera à distance plus petite que ε de ℓ .



× termes de la suite u

■ zones du graphe où les termes de u sont à distance plus petite que ε de l

N rang à partir duquel u vérifie :
pour tout $n \geq N$, $l - \varepsilon < u_n < l + \varepsilon$.

FIGURE 2.7 – Illustration de la définition 2.6.

mais seule la dernière formulation, i.e celle de la définition, est rigoureuse.

Remarque 2.8. Attention, l'entier N de la définition dépend du ε choisi ! Plus ε est petit, plus l'entier N a des chances de devoir être grand pour fonctionner.

Remarque 2.9. Pour un $\varepsilon > 0$ fixé, si $N \in \mathbb{N}$ fonctionne dans la définition, alors n'importe quel $N' \geq N$ fonctionne aussi : pour tout $n \geq N'$, on aura bien $\ell - \varepsilon < u_n < \ell + \varepsilon$, car $n \geq N' \geq N$ donc $n \geq N$, et on a supposé que c'était le cas pour tout $n \geq N$.

S'il existe un $\ell \in \mathbb{R}$ tel que $u : \mathbb{N} \rightarrow \mathbb{R}$ converge vers ℓ , on dit simplement que u converge, et on appellera ℓ sa *limite*⁵. Dans le cas contraire, on dit que u diverge.

Exemples 2.10. En prenant les mêmes notations que dans les exemples 2.4, on a :

1. La suite constante u converge : $u_n \xrightarrow[n \rightarrow +\infty]{} 1$.
2. La suite identité v diverge.
3. Nous étudierons la convergence de la suite géométrique plus tard.
4. La suite y converge : $y_n \xrightarrow[n \rightarrow +\infty]{} 0$.

Point pratique : comment prouver qu'une suite converge ? Pour montrer qu'une suite converge, on trouvera généralement ℓ à l'aide de notre intuition (comme nous l'avons fait dans les exemples précédents), puis on écrira la démonstration de "Pour tout $\varepsilon > 0$, ..." comme nous en avons désormais l'habitude :

1. On se donnera un réel $\varepsilon > 0$ générique : *Soit* $\varepsilon > 0$.
2. On posera un entier $N \in \mathbb{N}$ qui existe bel et bien : *On pose* $N = \dots$
3. On se donnera alors un entier $n \geq N$: *Soit* $n \geq N$.
4. Pour finalement vérifier que : $\ell - \varepsilon < u_n < \ell + \varepsilon$.

Point pratique : comment prouver qu'une suite diverge ? Pour montrer qu'une suite diverge, on procédera généralement par l'absurde en supposant que la suite converge vers une limite $\ell \in \mathbb{R}$ et en aboutissant à une contradiction. En supposant que u converge, la définition de la convergence indiquera que u satisfait une certaine propriété pour tout $\varepsilon > 0$, elle sera donc vraie pour celui que l'on veut et il s'agit en général de faire un choix judicieux pour aboutir à une contradiction.

Démonstrations des exemples. Avant de passer à la démonstration des faits énoncés dans les exemples 2.10, il faut que je vous énonce une propriété remarquable qu'ont les nombres réels et que nous n'allons cesser d'utiliser durant ce chapitre.

Théorème 2.11. Soient $a > 0$ et $b > 0$ deux nombres réels. Il existe un entier $n \geq 1$ tel que $b < na$.

5. On peut parler de sa limite car on peut montrer qu'il y a unicité de la limite lorsqu'elle existe, i.e. que si ℓ et ℓ' satisfont la définition 2.6 pour une suite u alors $\ell = \ell'$.

On dit que \mathbb{R} est *archimédien*. Je vous demande d'admettre ce fait, plutôt intuitif, dont la démonstration dépend de la définition de l'ensemble \mathbb{R} que l'on choisit. La définition de \mathbb{R} étant subtile et complexe elle dépasse le cadre de ce cours. On pourrait reformuler le théorème 2.11 ainsi : *peu importe la taille a de vos pas, et peu importe la distance b que vous devez parcourir, il existe un nombre de pas n (peut-être très grand) qui vous permettra d'arriver à destination.*

Démonstrations des affirmations des exemples 2.10.

1. On va donc montrer que u converge vers 1 en appliquant la Définition 2.6. On souhaite donc montrer que :

pour tout réel $\varepsilon > 0$, il existe $N \in \mathbb{N}$, tel que pour tout entier $n \geq N$,

$$1 - \varepsilon < u_n < 1 + \varepsilon.$$

Soit $\varepsilon > 0$. On cherche un $N \in \mathbb{N}$ tel que pour tout entier $n \geq N$, on a $1 - \varepsilon < u_n < 1 + \varepsilon$. Or, $1 - \varepsilon < 1 < 1 + \varepsilon$ est tout le temps vrai. De plus, pour tout $n \geq 0$, on a $u_n = 1$. Ainsi, pour tout $n \geq 0$, on a effectivement bien que $1 - \varepsilon < u_n < 1 + \varepsilon$, et l'entier $N = 0$ satisfait la propriété voulue.

4. Soit $\varepsilon > 0$. On cherche un $N \in \mathbb{N}$ tel que pour tout $n \geq N$, on a $0 - \varepsilon < y_n < 0 + \varepsilon$, i.e.

$$-\varepsilon < y_n < \varepsilon,$$

c'est-à-dire

$$-\varepsilon < \frac{1}{n} < \varepsilon.$$

Au vu de la définition de y , on ne peut remplacer y_n par $1/n$ que pour $n \geq 1$. Cependant, on peut toujours supposer que c'est le cas car les entiers n que l'on considérera satisferont toujours $n \geq N$ et on pourra toujours supposer $N \geq 1$ puisque si $N = 0$ fonctionne, alors $N = 1$ fonctionne aussi en vertu de la remarque 2.9.

Comme $1/n > 0$, on peut avoir toujours que $1/n > 0 > -\varepsilon$ et donc la partie gauche de l'inégalité est vérifiée pour tout $n \geq 1$.

La partie difficile est donc de trouver un $N \geq 1$ tel que pour tout $n \geq N$, on ait $\frac{1}{n} < \varepsilon$, i.e

$$1 < n \cdot \varepsilon.$$

Il suffit en fait d'appliquer le théorème 2.11 avec $a = \varepsilon$ et $b = 1$, pour obtenir qu'il existe un entier $N \geq 1$ tel que $b < Na$, i.e. tel que $1 < N\varepsilon$. De plus, pour tout $n \geq N$, on a que $N \cdot \varepsilon < n \cdot \varepsilon$ car $\varepsilon > 0$, donc aussi :

$$1 < n\varepsilon.$$

et finalement

$$\frac{1}{n} < \varepsilon.$$

Conclusion : On a bien montré : pour tout $\varepsilon > 0$, il existe un entier $N \in \mathbb{N}$ tel que pour tout entier $n \geq N$, $-\varepsilon < y_n < \varepsilon$, i.e y converge vers 0.

2. Supposons par l'absurde qu'elle converge vers un nombre réel $\ell \in \mathbb{R}$, i.e. $v_n \xrightarrow[n \rightarrow +\infty]{} \ell$, et montrons que nous aboutissons à une contradiction. Choisissons par exemple $\varepsilon = 1$. Par définition de la convergence, il existe $N \in \mathbb{N}$ tel que pour tout $n \geq N$, on a $\ell - 1 < v_n < \ell + 1$. On va montrer qu'en fait :

il existe un $n \geq N$ tel que $v_n > \ell + 1$,

ce qui contredira bien l'affirmation précédente.

Comme $\ell + 1 > v_n = n \geq 0$, on peut appliquer le Théorème 2.11 avec $a = 1$ et $b = \ell + 1$, qui vérifient bien $a > 0$ et $b > 0$. Ce théorème affirme alors qu'il existe un $N' \in \mathbb{N}$ tel que $N'a > b$, c'est-à-dire tel que $N' > \ell + 1$. Poser $n = \max(N', N)$ donne bien un $n \geq N$ tel que $v_n = n \geq N' > \ell + 1$, i.e. une contradiction.

Conclusion : la convergence de v vers ℓ est contradictoire, i.e v diverge. □

2.4.1 Reformulation de nos problèmes initiaux

Nous pouvons enfin donner un sens à ce que nous entendons par *somme infinie* dans nos exemples initiaux. Puisque l'addition d'un nombre fini de nombres est bien définie⁶ et puisque nous avons défini proprement la notion de suite et de limite, nous allons définir la valeur d'une somme infinie comme la limite (si elle existe!) de la suite qui à $n \in \mathbb{N}$ associe la somme des n premiers termes (les suites de ce type sont appelées *séries*). Si la limite n'existe pas, nous nous refuserons à considérer que la somme infinie à une valeur.

1. Dans le cas du paradoxe de Zénon, la distance parcourue après le n -ième déplacement vaut $d_n = \frac{1}{2} + \dots + \frac{1}{2^n}$. Autrement dit, la distance d parcourue dans le paradoxe de Zénon se modélise par la suite :

$$d : \begin{array}{l} \mathbb{N} \rightarrow \mathbb{R} \\ n \mapsto \frac{1}{2} + \dots + \frac{1}{2^n} \end{array} .$$

2. En considérant pour chaque entier $n \in \mathbb{N}$ la somme des n premiers termes de la série harmonique, nous obtenons une suite :

$$h : \begin{array}{l} \mathbb{N} \rightarrow \mathbb{R} \\ n \mapsto 1 + \frac{1}{2} + \dots + \frac{1}{n} \end{array} .$$

3. Pour la série de Grandi, il faut tout d'abord remarquer que puisque $(-1)^n = 1$ si n est pair, et $(-1)^n = -1$ si n est impair, la série de Grandi peut s'écrire comme la somme

$$1 - 1 + 1 - 1 + \dots = (-1)^0 + (-1)^1 + (-1)^2 + (-1)^3 + \dots$$

6. Nous ne l'avons pas montré, je vous demande de l'admettre.

et en considérons les sommes des n premiers termes pour chaque $n \in \mathbb{N}$, nous obtenons donc la suite :

$$g : \begin{array}{l} \mathbb{N} \rightarrow \mathbb{R} \\ n \mapsto 1 - 1 + \dots + (-1)^n \end{array} .$$

4. Dans le cas de la propagation d'une épidémie de paramètre r , le nombre de personnes qu'aura touché l'épidémie au bout de n générations d'infections vaut $t_n = 1 + r + r^2 + \dots + r^n$. Autrement dit, le total de personnes touchées se modélise par la suite :

$$t : \begin{array}{l} \mathbb{N} \rightarrow \mathbb{R} \\ n \mapsto 1 + r + r^2 + \dots + r^n \end{array} .$$

On peut donc reformuler proprement nos problèmes de la manière suivante :

Questions.

1. Est-ce que la suite d converge ? Si oui, vers quelle limite ?
2. Même question pour h .
3. Même question pour g .
4. Même question pour t .

2.4.2 Suites géométriques

Vous aurez peut-être remarqué que les termes de deux⁷ des sommes qui nous intéressent sont les termes d'une suite géométrique : $(1/2)^n$ et r^n . Avant de pouvoir déterminer si d et t convergent, nous allons tout d'abord étudier le comportement des suites géométriques (sans somme donc) de la forme :

$$\begin{array}{l} \mathbb{N} \rightarrow \mathbb{R} \\ n \mapsto q^n \end{array} ,$$

pour un nombre réel $q > 0$.

Remarque 2.12. On a vu en Figure 2.6 le graphe de deux suites géométriques, l'une de raison $q = 0,8$ et une autre de raison $q = 1,2$. Ces deux graphes suggèrent que lorsque $q < 1$, la suite géométrique de raison q converge vers 0, et lorsque $q > 1$, elle diverge. De plus, on peut facilement se convaincre que lorsque $q = 1$, alors elle converge vers 1 (elle est en fait constante égale à 1, et on a montré qu'une telle suite convergeait vers 1). Ces intuitions sont vraies, et sont résumées dans la proposition suivante.

Proposition 2.13. *Soit $q \in \mathbb{R}$ tel que $q > 0$.*

(i) *Si $q < 1$, alors $q^n \xrightarrow[n \rightarrow +\infty]{} 0$.*

7. En fait trois, les termes de la série de Grandi sont de la forme q^n pour $q = -1$ mais nous allons procéder autrement pour étudier la convergence de g .

(ii) Si $q = 1$, alors $q^n \xrightarrow[n \rightarrow +\infty]{} 1$.

(iii) Si $q > 1$, la suite géométrique de raison q diverge.

Démonstration. Cette proposition est admise. Notez cependant que le (ii) revient à montrer que la suite constante égale à 1 converge vers 1, ce que nous avons déjà fait. \square

2.5 Une bonne définition mathématique

Comme nous l'avons fait remarquer en introduction lors de l'étude de la série de Grandi, pour qu'une définition mathématique soit intéressante, il faut qu'elle coïncide avec l'intuition que nous avons du concept défini. Lors de la résolution de problèmes mathématiques, nous sommes souvent amenés⁸ à multiplier ou additionner des suites entre elles. L'intuition que nous avons du concept de limite nous donne le sentiment que si deux suites u et v convergent respectivement vers deux limites ℓ et ℓ' , alors la somme $u + v$ de u et v doit converger vers la somme des limites $\ell + \ell'$. En utilisant la définition mathématique de la convergence (Définition 2.6), on peut montrer que cette intuition est fondée (Théorème 2.15), c'est donc une preuve de plus que la définition 2.6 est une bonne définition de convergence.

Dans cette section, nous allons prouver d'autres propriétés intuitives que satisfont les suites convergentes, et qui s'avèrent très utiles en pratique pour montrer la convergence de suites sans avoir à revenir à la définition de la convergence et faire une preuve "epsilonesque"⁹.

2.5.1 Encadrement

La première propriété, qui nous sera en particulier utile au chapitre 3, est le théorème d'encadrement des limites, aussi appelé théorème des gendarmes.

Théorème 2.14. Soient u , v et w trois suites réelles et soit $\ell \in \mathbb{R}$. Si

$$(i) \quad u_n \xrightarrow[n \rightarrow +\infty]{} \ell,$$

$$(ii) \quad w_n \xrightarrow[n \rightarrow +\infty]{} \ell,$$

(iii) pour tout $n \in \mathbb{N}$, $u_n \leq v_n \leq w_n$,

alors, $v_n \xrightarrow[n \rightarrow +\infty]{} \ell$.

Démonstration. Soit $\varepsilon > 0$. Comme u converge vers ℓ , il existe $N' \in \mathbb{N}$ tel que pour tout $n \geq N'$, on a $\ell - \varepsilon < u_n < \ell + \varepsilon$, et comme w converge vers ℓ , il existe $N'' \in \mathbb{N}$ tel que pour tout $n \geq N''$, on a $\ell - \varepsilon < w_n < \ell + \varepsilon$. Alors, on peut poser $N = \max(N', N'')$. Soit $n \geq N$. On a alors que $u_n > \ell - \varepsilon$ car $n \geq N'$ et $w_n < \ell + \varepsilon$ car $n \geq N''$. Ainsi, on a :

$$\ell - \varepsilon < u_n \leq v_n \leq w_n < \ell + \varepsilon,$$

8. Si vous ne me croyez pas encore, attendez le chapitre 3...

9. i.e. avec des $\varepsilon > 0$.

et donc $\ell - \varepsilon < v_n < \ell + \varepsilon$.

Conclusion : On a donc bien montré que pour tout $\varepsilon > 0$, il existe $N \in \mathbb{N}$ tel que pour tout $n \geq N$, on a $\ell - \varepsilon < v_n < \ell + \varepsilon$, i.e. v converge vers ℓ . \square

2.5.2 Opérations

Voici maintenant l'énoncé et la preuve de la propriété sur la somme de deux suites convergentes.

Théorème 2.15. Soient $\ell \in \mathbb{R}$ et $\ell' \in \mathbb{R}$ et soient u et v deux suites réelles telles que $u_n \xrightarrow[n \rightarrow +\infty]{} \ell$ et $v_n \xrightarrow[n \rightarrow +\infty]{} \ell'$. Alors,

$$(i) \quad u_n + v_n \xrightarrow[n \rightarrow +\infty]{} \ell + \ell',$$

$$(ii) \quad u_n \cdot v_n \xrightarrow[n \rightarrow +\infty]{} \ell \cdot \ell'.$$

Démonstration. On va montrer la propriété sur la somme et on admettra celle sur le produit. Soit $\varepsilon > 0$. Comme alors $\varepsilon/2 > 0$ et que u converge, il existe $N' \in \mathbb{N}$ tel que pour tout $n \geq N'$, on a $\ell - \varepsilon/2 < u_n < \ell + \varepsilon/2$. Comme également v converge, il existe $N'' \in \mathbb{N}$ tel que pour tout $n \geq N''$, on a $\ell' - \varepsilon/2 < v_n < \ell' + \varepsilon/2$. On pose $N = \max(N', N'')$. Soit $n \geq N$. On a, puisque $n \geq N'$ et $n \geq N''$,

$$\ell - \varepsilon/2 < u_n < \ell + \varepsilon/2,$$

et

$$\ell' - \varepsilon/2 < v_n < \ell' + \varepsilon/2.$$

En additionnant ces deux inégalités, on a :

$$\ell + \ell' - \varepsilon/2 - \varepsilon/2 < u_n + v_n < \ell + \ell' + \varepsilon/2 + \varepsilon/2,$$

i.e.

$$\ell + \ell' - \varepsilon < u_n + v_n < \ell + \ell' + \varepsilon.$$

Conclusion : on a bien montré que pour tout $\varepsilon > 0$, il existe $N \in \mathbb{N}$ tel que pour tout $n \geq N$, on a $\ell + \ell' - \varepsilon < u_n + v_n < \ell + \ell' + \varepsilon$, i.e. $u_n + v_n \xrightarrow[n \rightarrow +\infty]{} \ell + \ell'$. \square

Exercice 2.16. Avec les mêmes hypothèses que le théorème 2.15, montrer que pour tout $\lambda > 0$, on a $\lambda \cdot u_n \xrightarrow[n \rightarrow +\infty]{} \lambda \cdot \ell$.

2.6 Résolution de nos problèmes initiaux

2.6.1 Paradoxe de Zénon

Pour rappel, la distance d parcourue dans le paradoxe de Zénon se modélise par la suite :

$$d: \begin{array}{l} \mathbb{N} \rightarrow \mathbb{R} \\ n \mapsto \frac{1}{2} + \dots + \frac{1}{2^n} \end{array} .$$

La question que l'on se pose alors est la suivante : est-ce que la suite d converge, et si oui, vers quelle valeur ? Pour répondre à cette question, nous allons étudier plus en détail un objet plus général : les suites définies par des sommes de termes qui suivent une progression géométrique, i.e. les suites de la forme :

$$\begin{aligned} \mathbb{N} &\rightarrow \mathbb{R} \\ n &\mapsto 1 + q + \cdots + q^n, \end{aligned}$$

pour un paramètre réel $q > 0$. Ces suites sont appelées *séries géométriques de raison q* .

L'approche que nous adoptons ici est typique en mathématique : pour résoudre un problème précis (la convergence de d), nous allons étudier l'objet mis en jeu avec le plus de généralité possible (les séries géométriques). Nous récapitulerons les avantages de cette approche après en avoir présenté les fruits, qui consistent en les deux résultats mathématiques qui suivent.

Théorème 2.17. *Soit $q \in \mathbb{R}$ tel que $q \neq 1$. Alors, pour tout $n \in \mathbb{N}$,*

$$1 + q + \cdots + q^n = \frac{1 - q^{n+1}}{1 - q}.$$

Démonstration. On peut calculer :

$$\begin{aligned} (1 - q)(1 + q + \cdots + q^n) &= 1 + q + q^2 \cdots + q^n - q \cdot (1 + q + \cdots + q^{n-1} + q^n) \\ &= 1 + q + q^2 \cdots + q^n - q - q^2 - \cdots - q^n - q^{n+1} \\ &= 1 - q^{n+1}. \end{aligned}$$

Comme $q \neq 1$, on peut alors diviser par $(1 - q)$ de chaque côté pour obtenir :

$$1 + q + \cdots + q^n = \frac{1 - q^{n+1}}{1 - q}.$$

□

Remarque 2.18. Dans la preuve, les calculs effectués sur les sommes sont justes et bien définis car les sommes mises en jeu sont *finies*. Cela n'a rien à voir avec les calculs (faux) que nous avons fait sur la série de Grandi en introduction et qui avaient abouti à des absurdités : ils portaient sur des sommes infinies, pour lesquelles nous n'avons pas défini ce que voulait dire *additionner*, *soustraire*, etc.

Grâce à ce théorème, nous allons pouvoir déduire un résultat de convergence pour les séries géométriques.

Théorème 2.19. *Soit $q \in \mathbb{R}$ tel que $0 < q < 1$. Alors,*

$$1 + q + \cdots + q^n \xrightarrow{n \rightarrow +\infty} \frac{1}{1 - q}.$$

Démonstration. Nous savons par la proposition 2.13 que $q^n \xrightarrow[n \rightarrow +\infty]{} 0$. Par le théorème sur le produit des limites, on a également $q_{n+1} \xrightarrow[n \rightarrow +\infty]{} 0$ (on a multiplié par la suite constante égale à q , qui tend vers q). Par le théorème sur le produit des limites (Théorème 2.15) on a donc¹⁰ que $-q^{n+1} \xrightarrow[n \rightarrow +\infty]{} 0$, et par le même théorème sur la somme des limites¹¹ : $1 - q^{n+1} \xrightarrow[n \rightarrow +\infty]{} 1$. A nouveau par le théorème sur le produit¹², on a :

$$\frac{1 - q^{n+1}}{1 - q} \xrightarrow[n \rightarrow +\infty]{} \frac{1}{1 - q}.$$

D'où le résultat par le théorème 2.17 □

Corollaire 2.20.

$$\frac{1}{2} + \cdots + \left(\frac{1}{2}\right)^n \xrightarrow[n \rightarrow +\infty]{} 1.$$

Démonstration. On peut calculer :

$$\frac{1}{2} + \cdots + \left(\frac{1}{2}\right)^n = 1 + \frac{1}{2} + \cdots + \left(\frac{1}{2}\right)^n - 1.$$

Par le théorème 2.19, on a

$$1 + \frac{1}{2} + \cdots + \left(\frac{1}{2}\right)^n \xrightarrow[n \rightarrow +\infty]{} \frac{1}{1 - \frac{1}{2}} = \frac{1}{\frac{1}{2}} = 1 \cdot \frac{2}{1} = 2.$$

Ainsi, par le théorème sur les opérations sur les limites (Théorème 2.15), on a :

$$1 + \frac{1}{2} + \cdots + \left(\frac{1}{2}\right)^n - 1 \xrightarrow[n \rightarrow +\infty]{} 2 - 1 = 1,$$

d'où le résultat. □

Nous pouvons désormais conclure sur le paradoxe de Zénon : la distance totale parcourue, lorsque nous aurons divisé notre segment de longueur 1 km "indéfiniment", nous pouvons le définir proprement comme la limite de la somme des distances successivement parcourues, i.e. la limite de la suite d , et cette limite vaut 1 km, par le dernier corollaire.

10. On a multiplié par la suite constante égale à -1 .

11. On a additionné la suite constante égale à 1.

12. On a multiplié par la suite constante égale à $1/(1 - q)$.

Un mot sur l'approche. L'approche consistant à abstraire le problème en étudiant les séries géométriques dans leur généralité plutôt qu'uniquement celle du paradoxe de Zénon (i.e. pour $q = 1/2$) est une des étapes d'une démarche d'abstraction plus générale que nous suivons depuis le début de ce chapitre. Nous cherchions à résoudre le paradoxe de Zénon et nous avons pour cela largement abstrait le problème. Résumons notre démarche.

Dans un premier temps, nous avons dégagé l'objet au cœur de notre problème : les processus numériques évoluant par étape (comme la distance parcourue dans le paradoxe de Zénon). Plutôt que d'étudier celui donné par le paradoxe de Zénon, nous en avons donné une définition la plus générale possible, les *suites numériques*. Ainsi muni de la notion de suite, il nous fallait encore donner un sens à la notion de *convergence* pour résoudre notre problème de définition des sommes infinies. Ayant ainsi abstrait le problème, nous avons pu dégager des propriétés très générales de la convergence des suites (le théorème d'encadrement et celui sur les opérations) qui nous ont permis de prouver sans effort le théorème de convergence sur les séries géométriques. Le résultat sur la somme intervenant dans le paradoxe de Zénon en a découlé trivialement.

Tentons de dégager les avantages de cette approche :

1. elle permet d'utiliser les fruits de l'étude générale à un plus grand nombre de situations que si nous l'avions étudié dans un cas particulier, comme nous allons le voir tout de suite.
2. De plus, et surtout, elle permet d'obtenir une meilleure compréhension des objets étudiés. En effet, pour que notre étude englobe plus de situations, nous avons dû l'abstraire, c'est-à-dire extraire les caractéristiques essentielles de notre problème. Par exemple, nous comprenons désormais que si :

$$\frac{1}{2} + \dots + \left(\frac{1}{2}\right)^n \xrightarrow{n \rightarrow +\infty} 1,$$

ce n'est pas dû à une particularité du nombre $1/2$, mais qu'il y a derrière ce résultat quelque chose de plus général, le théorème 2.19, et que cette convergence est en fait due à la structure même des séries géométriques, dont les termes additionnés sont de nature, justement, géométrique. C'est en fait la convergence de la *suite géométrique* (Proposition 2.13) et la formule de la somme géométrique (Théorème 2.17) qui sont les éléments clés de cette convergence. Nous avons réussi à déterminer les caractéristiques essentielles de notre objet responsables des propriétés qui nous intéressent, ce qui est bien plus fort que d'avoir simplement résolu le problème.

Généraliser les problèmes mathématiques pour les résoudre est une approche qui était chère au mathématicien Alexandre Grothendieck (1928-2014).

«Prenons par exemple la tâche de démontrer un théorème qui reste hypothétique (à quoi, pour certains, semblerait se réduire le travail mathématique). Je vois deux approches extrêmes pour s'y prendre. L'une est celle du marteau et du burin, quand

le problème posé est vu comme une grosse noix, dure et lisse, dont il s'agit d'atteindre l'intérieur, la chair nourricière protégée par la coque. Le principe est simple : on pose le tranchant du burin contre la coque, et on tape fort. Au besoin, on recommence en plusieurs endroits différents, jusqu'à ce que la coque se casse — et on est content.

Je pourrais illustrer la deuxième approche, en gardant l'image de la noix qu'il s'agit d'ouvrir. La première parabole qui m'est venue à l'esprit tantôt, c'est qu'on plonge la noix dans un liquide émoullissant, de l'eau simplement pourquoi pas, de temps en temps on frotte pour qu'elle pénètre mieux, pour le reste on laisse faire le temps. La coque s'assouplit au fil des semaines et des mois — quand le temps est mûr, une pression de la main suffit, la coque s'ouvre comme celle d'un avocat mûr à point !

[...] C'est "l'approche de la mer", par submersion, absorption, dissolution — celle où, quand on n'est très attentif, rien ne semble se passer à aucun moment : chaque chose à chaque moment est si évidente, et surtout, si naturelle, qu'on se ferait presque scrupule souvent de la noter noir sur blanc, de peur d'avoir l'air de bombiner, au lieu de taper sur un burin comme tout le monde... »

Alexander Grothendieck, *Récoltes et Semailles*.

2.6.2 Propagation d'une épidémie

L'étude réalisée dans la section précédente permet également de résoudre notre problème d'épidémie. Le total de personnes infectées après chaque génération est modélisé par la suite suivante :

$$t : \begin{array}{l} \mathbb{N} \rightarrow \mathbb{R} \\ n \mapsto 1 + r + r^2 + \dots + r^n \end{array} .$$

On peut donc utiliser le théorème 2.19 pour montrer que si $0 < r < 1$, alors $t_n \xrightarrow[n \rightarrow +\infty]{} 1/(1-r)$. Par exemple, si chaque malade infecte en moyenne 0,5 personnes durant sa période de contagion, alors le nombre total de personnes infectées tend vers 2. Avec $r = 0,98$, le nombre total de personnes infectées tend vers 50.

En revanche, que se passe-t-il si $r > 1$? Chaque personne infectée en infecte en moyenne légèrement plus qu'une. On a le sentiment que le nombre total de personnes infectées ne va cesser de croître vers l'infini ou simplement vers le nombre total de personnes dans la population dans une situation réelle. La proposition suivante montre que du moins, la suite t ne converge pas.

Proposition 2.21. *Si $r \geq 1$, alors la suite t diverge.*

Démonstration. Si $r = 1$, alors pour tout $n \in \mathbb{N}$, $t_n = n$ et on a déjà montré que dans ce cas, t diverge (Exemples 2.10). Supposons désormais que $r > 1$ et montrons que dans ce cas, t diverge. Supposons par l'absurde que t converge. Dans ce cas, on peut calculer, puisque $r \neq 1$, par le théorème 2.17, on a pour tout $n \in \mathbb{N}$,

$$(1-r) \cdot t_n = 1 - r^{n+1}.$$

Ainsi, pour tout $n \in \mathbb{N}$, on a :

$$r^n = \frac{1 - (1-r) \cdot t_n}{r} \quad (2.3)$$

car $r \neq 0$. De plus, on peut montrer par le théorème sur les limites (Théorème 2.15) que puisque t converge, la suite définie par :

$$\begin{aligned} \mathbb{N} &\rightarrow \mathbb{R} \\ n &\mapsto \frac{1-(1-r)t_n}{r} \end{aligned}$$

converge également, et donc la suite géométrique de raison r également par (2.3). Ceci est une absurdité par la proposition 2.13 car $r > 1$. \square

Nous n'avons pas défini pour une suite la propriété de *tendre vers l'infini*, qui est une notion bien définie en mathématiques, et l'on pourrait montrer que lorsque le facteur r est supérieur ou égal à 1, la suite t tend en fait vers l'infini, ce qui validerait notre intuition sur la croissance sans limite du nombre de personnes infectées. Dans ce cas, l'épidémie s'envenime et la totalité de la population finit tôt ou tard par être contaminée.

2.6.3 Série de Grandi

On va utiliser le dernier exercice de la feuille d'exercices. Pour ce faire, on pourrait tout d'abord montrer, si nous avons l'outil adéquat¹³, que pour tout $n \in \mathbb{N}$,

$$1 + (-1) + \dots + (-1)^n = \begin{cases} 1 & \text{si } n \text{ est pair,} \\ 0 & \text{si } n \text{ est impair.} \end{cases}$$

Or la suite définie par

$$\begin{aligned} \mathbb{N} &\rightarrow \mathbb{R} \\ n &\mapsto \begin{cases} 1 & \text{si } n \text{ est pair,} \\ 0 & \text{si } n \text{ est impair,} \end{cases} \end{aligned}$$

est à valeurs entières et n'est pas stationnaire¹⁴ (cf le dit exercice). Par la conclusion du même exercice, elle ne peut donc pas être convergente.

2.6.4 Série harmonique

On va montrer que la série harmonique diverge. Ici encore, on pourrait en fait montrer que la série harmonique tend vers l'infini, mais cela dépasse (légèrement) le cadre de ce cours.

Proposition 2.22. *La suite h diverge.*

13. la démonstration par récurrence

14. montrez-le.

Démonstration. Supposons par l'absurde que la suite h converge vers $\ell \in \mathbb{R}$. On va admettre le lemme suivant ¹⁵ :

Lemme 2.23. *Si $u : \mathbb{N} \rightarrow \mathbb{R}$ converge vers $\ell \in \mathbb{R}$, alors $u_{2n} \xrightarrow[n \rightarrow +\infty]{} \ell$.*

Ainsi, $h_{2n} \xrightarrow[n \rightarrow +\infty]{} \ell$. Dans ce cas, la différence $h_{2n} - h_n \xrightarrow[n \rightarrow +\infty]{} 0$ par opérations sur les limites. On va montrer qu'en fait, cette différence est toujours plus grande que $1/2$, et conclure à une absurdité par l'exercice 5 de la feuille d'exercices.

Soit $n \geq 1$ un entier. On peut calculer :

$$\begin{aligned} h_{2n} - h_n &= 1 + \frac{1}{2} + \cdots + \frac{1}{n} + \frac{1}{n+1} + \cdots + \frac{1}{2n} - 1 - \frac{1}{2} - \cdots - \frac{1}{n} \\ &= \frac{1}{n+1} + \cdots + \frac{1}{2n}. \end{aligned}$$

Or, pour tout entier $1 \leq k \leq n$, on a $n+k \leq n+n = 2n$, donc :

$$\frac{n+k}{2n} \leq 1,$$

car $2n > 0$ et enfin :

$$\frac{1}{2n} \leq \frac{1}{n+k}.$$

car $n+k > 0$. En injectant cela dans la première équation, on obtient :

$$\begin{aligned} h_{2n} - h_n &= \frac{1}{n+1} + \cdots + \frac{1}{2n} \\ &\geq \frac{1}{2n} + \cdots + \frac{1}{2n} \text{ (avec } n \text{ termes)} \\ &\geq n \cdot \frac{1}{2n} \\ &\geq \frac{1}{2}. \end{aligned}$$

Comme la limite de la suite constante égale à $1/2$ est $1/2$, on peut appliquer l'exercice 5 de la feuille d'exercices pour obtenir par l'équation précédente :

$$0 \geq \frac{1}{2},$$

une absurdité. □

15. Vous pouvez parfaitement le montrer, appliquez la définition de la convergence.

Chapitre 3

Probabilités et statistiques

3.1 Introduction

Ce chapitre a pour but d'introduire la théorie des probabilités et les statistiques inférentielles par l'étude d'un problème jouet classique de statistiques. Ceci permettra de motiver l'introduction des concepts clés des probabilités, et de voir le lien entre les probabilités et les statistiques inférentielles, deux "sœurs" parmi les théories mathématiques.

Attention : encore une fois, toutes les considérations qui ne sont pas *stricto sensu* des mathématiques sont à lire pour ce qu'elles sont : des réflexions de quelqu'un qui n'y connaît (presque) rien et qui n'a que trop peu de culture sur le sujet.

3.1.1 Problème jouet

Je vous propose de jouer à un jeu de pile ou face dont la règle est la suivante :

Si la pièce tombe sur pile, je vous donne 100 €. Sinon, c'est vous qui me donnez 100 €.

Il y a tout de même beaucoup d'enjeu pour vous et j'ai sorti cette pièce de ma poche : elle pourrait bien être truquée, vous avez toutes les raisons de ne pas me faire confiance. Bon joueur, je vous autorise à étudier la pièce pour tester sa bonne et due forme, à condition que vous fondiez votre étude sur une théorie mathématique solidement construite. Voici donc le problème qu'il vous faut résoudre :

Question. *La pièce est-elle truquée ?*

L'énoncé de ce problème soulève immédiatement une suite de questions complexes : qu'est-ce qu'une pièce qui n'est pas truquée ? Intuitivement : une pièce où l'on aurait autant de chances de faire pile que de faire face. Que veut dire *autant de chances* ? Cette dernière question est très compliquée. On ne peut se contenter de dire qu'il y a autant de possibilités de perdre que de gagner car dans tous les cas, la pièce n'a que deux possibilités : tomber sur *pile* ou tomber sur *face*. Une pièce sera considérée comme truquée simplement si elle a *plus de chances* de tomber sur une face que sur l'autre. Le problème de la définition de *truquée* et de *non truquée* est qu'elle fait intervenir un concept dont l'usage est courant mais la définition extrêmement difficile : le *hasard*.

La stratégie que nous allons suivre pour résoudre ce problème sera la suivante :

1. Elaborer une théorie mathématique modélisant notre expérience de pile ou face, et plus généralement les phénomènes régis par le hasard : *la théorie des probabilités*.
2. Exprimer dans cette théorie les deux situations possibles : la pièce est truquée ou non.
3. Elaborer une théorie mathématique permettant de déterminer *par l'expérience* si une pièce réelle suit plutôt la modélisation "pièce truquée" ou la modélisation "pièce non truquée", ce sont les *statistiques inférentielles*. En fait, cette théorie permet même de *quantifier* l'adéquation entre nos modélisations du comportement de la pièce et la réalité.

Avant de commencer la première étape de ce programme de travail, nous terminons cette introduction en tentant de répondre à la question suivante : *Qu'est-ce que le hasard ?*

3.1.2 Le hasard

A vrai dire, il n'y a aucun besoin de savoir ce qu'est le hasard pour faire des probabilités, tout comme il n'y a aucun besoin de savoir ce qu'est la définition d'un ensemble¹ pour faire de la théorie des ensembles. Cette section a simplement pour but de donner un aperçu de la richesse et la complexité de la notion de hasard mais aussi, et surtout, pour le mettre en regard avec la théorie des probabilités dans la suite du chapitre.

Note bibliographique. *Cette partie s'appuie sur – ou plutôt n'est qu'un résumé grossier de – la très bonne introduction du mémoire de M2 de Mathilde Escudero intitulé Le hasard : ni exclusivement subjectif, ni purement objectif, défense d'une conception « interventionniste ».*

Une définition possible, peut-être un peu naïve, est la suivante : *le hasard est la cause des évènements considérés comme imprévisibles*. Historiquement, il y a eu deux grandes manières de concevoir la notion de hasard en philosophie et en science : l'une *subjective* et l'autre *objective*.

Hasard subjectif. Le point de vue *subjectif* se décline en deux grandes conceptions : l'ignorance des causes de Laplace et le psychologisme de Bergson. Dans les deux cas, le hasard est conçu comme n'existant que pour l'observateur du phénomène fortuit. Dans le cadre de ce cours, nous n'aborderons que la conception de Pierre-Simon de Laplace (1749-1827), qui l'explique lui-même dans son *Essai philosophique sur les probabilités* de 1814.

Dans l'ignorance des liens qui les unissent au système entier de l'univers, on les a fait dépendre des causes finales ou du hasard, suivant qu'ils arrivaient et se succédaient avec régularité ou sans ordre apparent : mais ces causes imaginaires ont été successivement reculées avec les bornes de nos connaissances, et disparaissent entièrement devant la saine philosophie, qui ne voit en elles que l'expression de l'ignorance où nous sommes des véritables causes.

Ainsi, le hasard apparent auquel nous sommes confronté n'est qu'une illusion causée par notre propre ignorance de l'état du monde. Si nous connaissions parfaitement l'état du monde et les lois physiques qui régissent son évolution, tous les évènements deviendraient prévisibles et il n'y aurait pas de hasard. En des termes plus savants, Laplace défend un déterminisme métaphysique :

Nous devons donc envisager l'état présent de l'univers comme l'effet de son état antérieur, et comme la cause de celui qui va suivre. Une intelligence qui pour un instant

1. sous-entendu une définition qui en donnerait la nature.

donné connaîtrait toutes les forces dont la nature est animée et la situation respective des êtres qui la composent, si d'ailleurs elle était assez vaste pour soumettre ses données à l'analyse, embrasserait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome : rien ne serait incertain pour elle, et l'avenir comme le passé serait présent à ses yeux.

Hasard objectif. Le point de vue *objectif* se décline également en deux grandes conceptions : la recontre de séries causales indépendantes de Cournot, la disproportion entre les effets et les causes de Poincaré. Dans les deux cas, le hasard est conçu comme existant indépendamment d'un observateur, comme intrinsèque aux phénomènes considérés comme fortuits. Nous n'aborderons que celle de Henri Poincaré (1854 - 1912), qu'il présente lui-même dans son *Calcul des probabilités* de 1912.

Le premier exemple que nous allons choisir est celui de l'équilibre instable ; si un cône repose sur sa pointe, nous savons bien qu'il va tomber, mais nous ne savons pas de quel côté ; il nous semble que le hasard seul va en décider. Si le cône était parfaitement symétrique, si son axe était parfaitement vertical, s'il n'était soumis à aucune autre force que la pesanteur, il ne tomberait pas du tout. Mais le moindre défaut de symétrie va le faire pencher légèrement d'un côté ou de l'autre, et dès qu'il penchera, si peu que ce soit, il tombera tout à fait de ce côté. Si même la symétrie est parfaite, une trépidation très légère, un souffle d'air pourra le faire incliner de quelques secondes d'arc ; ce sera assez pour déterminer sa chute et même le sens de sa chute qui sera celui de l'inclinaison initiale.

Une cause très petite, qui nous échappe, détermine un effet considérable que nous ne pouvons pas ne pas voir, et alors nous disons que cet effet est dû au hasard. Si nous connaissions exactement les lois de la nature et la situation de l'univers à l'instant initial, nous pourrions prédire exactement la situation de ce même univers à un instant ultérieur. Mais, lors même que les lois naturelles n'auraient plus de secret pour nous, nous ne pourrions connaître la situation qu'approximativement. Si cela nous permet de prévoir la situation ultérieure avec la même approximation, c'est tout ce qu'il nous faut, nous disons que le phénomène a été prévu, qu'il est régi par des lois ; mais il n'en est pas toujours ainsi, il peut arriver que de petites différences dans les conditions initiales en engendrent de très grandes dans les phénomènes finaux ; une petite erreur sur les premières produirait une erreur énorme sur les derniers. La prédiction devient impossible et nous avons le phénomène fortuit.

Ainsi, Henri Poincaré conçoit le hasard comme provenant de la disproportion entre les causes et les effets des phénomènes perçus comme fortuits, on qualifie ces phénomènes de *chaotiques*. Pour ceux-là, si petite soit l'erreur sur la connaissance des conditions initiales, elle sera amplifiée en un phénomène non négligeable. Notre connaissance du système étant nécessairement imparfaite, les lois physiques ne nous sont d'aucun secours pour prédire l'évolution du système et le résultat de l'expérience nous apparaîtra toujours comme dû au hasard. On notera cependant que Poincaré défend lui aussi

un déterminisme métaphysique, qui n'est pas contradictoire avec l'impossibilité pour la science de prédire certains phénomènes.

Hasard quantique. J'ai volontairement mis de côté jusqu'à présent une révolution scientifique du XX^{ème} siècle ayant remis en cause² le déterminisme défendu par Laplace, Poincaré et sur lequel se fonde traditionnellement la démarche scientifique : la mécanique quantique.

Ne comprenant moi-même qu'assez mal la mécanique quantique, je ne me risquerai pas à rentrer dans les détails. Ce que je peux dire, c'est qu'à l'échelle microscopique, les particules peuvent être dans plusieurs états simultanément (par exemple à plusieurs endroits à la fois) et ce n'est que l'opération de mesure par l'observateur qui fige la particule dans l'un de ces états, avec une certaine probabilité pour chaque état.

Ceci ne tranche *a priori* pas entre hasard subjectif et objectif : l'électron a-t-il intrinsèquement un comportement probabiliste, ou nous manque-t-il des informations pour prédire sa position avant de la mesurer ? L'interprétation majoritairement partagée aujourd'hui est que le hasard intervenant dans l'opération de mesure n'est pas lié à l'ignorance par l'observateur de certaines données physiques, mais qu'il s'agit bien là d'un hasard *intrinsèque*, qui remet en cause la vision déterministe du monde si bien exprimée par Laplace.

En effet, John Bell (1928-1990) a montré que l'existence de variables *locales*³ cachées, qu'ignorerait l'observateur, était contradictoire. Ainsi, ou bien le hasard quantique est véritablement intrinsèque (c'est l'hypothèse la plus communément admise), ou bien il existe des variables cachées non locales — i.e. des choses qui vont plus vite que la lumière — et cette dernière possibilité serait une révolution pour la physique moderne.

3.2 Probabilités sur un univers fini

Notre but, dans cette partie, sera d'élaborer une théorie mathématique modélisant l'expérience de pile ou face, et plus généralement les phénomènes régis par le hasard n'ayant qu'un nombre fini d'issues possibles.

3.2.1 Définition

La définition moderne de *probabilité* est due au mathématicien russe Andreï Kolmogorov (1903 - 1987) qui fut le premier à formaliser la notion de probabilité comme une *mesure d'une quantité de chance* qu'un évènement se produise. Ce point de vue est très fructueux, comme nous allons le voir tout au long du chapitre.

Définition 3.1. Soit Ω un ensemble fini. On appelle *loi de probabilité sur Ω* , ou simplement *probabilité sur Ω* , toute application $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ telle que :

2. selon son interprétation majoritairement partagée.

3. qui sont conformes à l'hypothèse de localité : le fait qu'une interaction entre deux objets se propage de proche en proche à une vitesse inférieure ou égale à celle de la lumière.

- (i) (Positivité) Pour tout $A \in \mathcal{P}(\Omega)$, $\mathbb{P}(A) \geq 0$,
- (ii) (Masse unitaire) $\mathbb{P}(\Omega) = 1$,
- (iii) (Additivité) Pour tout $A \in \mathcal{P}(\Omega)$ et tout $B \in \mathcal{P}(\Omega)$ tels que $A \cap B = \emptyset$, on a $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Vocabulaire, intuition et exemples.

L'ensemble Ω est appelé *univers* et ses éléments sont appelés *issues*. Comme son nom l'indique, il représente l'ensemble des issues possible de l'expérience aléatoire.

Exemple 3.2.

1. Pour le lancer d'une pièce, les issues possibles de l'expérience aléatoire sont les deux faces de la pièce "pile" et "face", et l'univers est donc $\Omega = \{P, F\}$. Pour ne manipuler que des nombres, on pourrait tout aussi bien numéroter 0 et 1, de sorte que l'univers soit $\Omega = \{0, 1\}$, c'est ce que nous ferons dans la suite.
2. Pour le lancer d'un dé à six faces, les issues possibles de l'expérience aléatoire sont les nombres de 1 à 6. L'univers est donc : $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Les parties de Ω — donc des ensembles d'issues possibles — sont appelées *évènements*, ce sont eux dont on peut calculer la probabilité. Si $A \in \mathcal{P}(\Omega)$ est un évènement, alors l'image $\mathbb{P}(A)$ de A par \mathbb{P} est appelée *probabilité de l'évènement A* et mesure intuitivement la chance qu'à l'évènement A de se produire. On appelle *évènement élémentaire* un élément de la forme $\{\omega\}$ pour un certain $\omega \in \Omega$. Ce sont les évènements constitué d'une seule issue.

Exemple 3.3.

1. **Pièce :** Il n'y a ici que quatre évènements possibles, les quatre parties de Ω : \emptyset , $\{0\}$, $\{1\}$ et $\{0, 1\} = \Omega$. Ces évènements correspondent en langage courant aux évènements "la pièce n'est retombée sur aucune face", "la pièce est tombée sur pile", "la pièce est tombée sur face" et "la pièce est tombée sur pile ou sur face" (le ou est inclusif ici comme toujours en mathématiques). Si la pièce n'est pas truquée, la loi de probabilité que nous choisissons pour modéliser l'expérience du lancer de la pièce est la suivante :

$$\begin{array}{ll}
 \mathcal{P}(\Omega) & \longrightarrow \mathbb{R} \\
 \emptyset & \longmapsto 0 \\
 \{0\} & \longmapsto 1/2 \\
 \{1\} & \longmapsto 1/2 \\
 \{0, 1\} & \longmapsto 1
 \end{array}$$

(vous pouvez vérifier que c'est bien un probabilité au sens de la définition ci-dessus).

2. **Dé** : L'évènement "faire un nombre pair" est l'évènement $A = \{2, 4, 6\}$. Que vaut $\mathbb{P}(A)$? Cela dépend de la loi de probabilité que l'on choisit. Supposons par exemple que nous choisissons la loi de probabilité suivante :

$$\begin{aligned} \mathcal{P}(\Omega) &\longrightarrow \mathbb{R} \\ A &\longmapsto \begin{cases} 1 & \text{si } A = \{6\}, \\ 0 & \text{sinon.} \end{cases} \end{aligned}$$

qui correspondrait à la modélisation d'un dé truqué qui tombe sur 6 avec probabilité 1 (vous pouvez vérifier que c'est un bien une loi de probabilité sur Ω au sens de la définition précédente). Avec cette loi de probabilité, $\mathbb{P}(A) = \mathbb{P}(\{2, 4, 6\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = 0 + 0 + 1 = 1$ par additivité de \mathbb{P} . Le résultat du calcul est cohérent : si le dé tombe sur 6 avec probabilité 1, on a aussi probabilité 1 de faire un nombre pair.

Exercice 3.4. Définir la loi de probabilité qui modélise le lancer d'un dé équilibré (i.e. tel que toutes les faces ont même probabilité d'apparaître) à six faces. Correction : Définition/Proposition 3.9.

Pourquoi ces conditions ? Les trois conditions (i),(ii) et (iii) assurent que \mathbb{P} se comporte effectivement comme une mesure de la quantité de chance qu'un évènement se produise. En effet, la condition (i) assure que la quantité de chance \mathbb{P} doit être positive (ce qui est naturel pour une quantité).

La condition (ii) assure qu'une issue va arriver avec probabilité 1 puisque l'évènement Ω est simplement l'évènement "n'importe quelle issue parmi celles possibles". Autrement dit, cette condition indique que l'univers Ω est bien l'univers des issues possibles et que l'expérience aléatoire n'a aucune chance d'aboutir à une issue qui ne soit pas dans Ω .

La condition (iii) assure que \mathbb{P} se comporte bien comme une mesure : tout comme la quantité d'eau contenue dans l'union de deux récipients disjoints est la somme des quantités d'eau contenues dans chacun des récipients, les quantités de chance de deux évènements disjoints s'ajoute pour donner la quantité de chance que l'un ou l'autre (l'union des deux) se produise.

Vocabulaire 3.5. Soient A et B deux évènements.

1. On dit que A est *presque sûr* si $\mathbb{P}(A) = 1$.
2. On dit que A est *négligeable* si $\mathbb{P}(A) = 0$.
3. On appelle A et B l'évènement $A \cap B$.
4. On appelle A ou B l'évènement $A \cup B$.
5. On appelle *contraire de A* , et l'on note \overline{A} , l'évènement $\Omega \setminus A$.

3.2.2 Propriétés

Etudions les propriétés que notre modélisation du hasard possède.

Proposition 3.6. *Soit Ω un univers fini et soit \mathbb{P} une loi de probabilité sur Ω . On a :*

- (i) $\mathbb{P}(\emptyset) = 0$.
- (ii) *Pour tout $A \in \mathcal{P}(\Omega)$ et $B \in \mathcal{P}(\Omega)$,*
 - (a) *si $A \subseteq B$, alors $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$. En particulier, $\mathbb{P}(A) \leq \mathbb{P}(B)$.*
 - (b) $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$.
 - (c) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
- (iii) *Pour tout $A \in \mathcal{P}(\Omega)$, on a $0 \leq \mathbb{P}(A) \leq 1$.*

Intuition 3.7. On peut reformuler en langage courant les énoncés précédents :

- (i) La probabilité que l'issue de l'expérience aléatoire ne soit pas dans l'univers est nulle, i.e. l'univers modélise correctement l'expérience aléatoire.
- (ii) Soient A et B deux évènements. Ces propriétés sont des illustrations du fait qu'une loi de probabilité est une *mesure de la quantité de chance* qu'un évènement se produise :
 - (a) si les issues donnant l'évènement A donnent aussi l'évènement B , alors la quantité de chance que B arrive est plus grande que celle de A . Plus précisément, la quantité de chance pour que l'issue soit dans B mais pas dans A est précisément la quantité de chance d'être dans B moins celle d'être dans A .
 - (b) L'évènement contraire \overline{A} arrive avec "probabilité contraire", i.e. $1 - \mathbb{P}(A)$.
 - (c) C'est un raffinement de la propriété d'additivité : la quantité de chance que A ou B se produise est la somme des quantités de chance de A et de B si A et B sont disjoints. Sinon, la quantité de chance de $A \cup B$ est la quantité de chance d'être dans A mais pas dans B , plus celle d'être dans B mais pas dans A , plus celle d'être dans $A \cap B$. En additionnant $\mathbb{P}(A)$ et $\mathbb{P}(B)$, on a compté deux fois $A \cap B$: une fois avec A et une fois avec B . Il faut donc la soustraire une fois pour trouver la quantité de chance de $A \cup B$.
- (iii) Une quantité de chance est toujours comprise entre 0 et 1.

Démonstration.

- (i) Comme $\emptyset = \emptyset \cup \emptyset$ et $\emptyset = \emptyset \cap \emptyset$, on a par additivité de \mathbb{P} :

$$\mathbb{P}(\emptyset) = \mathbb{P}(\emptyset \cup \emptyset) = \mathbb{P}(\emptyset) + \mathbb{P}(\emptyset) = 2\mathbb{P}(\emptyset)$$

d'où $\mathbb{P}(\emptyset) = 0$.

- (ii) Soient A et B deux évènements.

- (a) Supposons que $A \subseteq B$. On peut alors décomposer $B = A \cup (B \setminus A)$ avec $A \cap (B \setminus A) = \emptyset$, de sorte que par additivité de \mathbb{P} , on ait $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$, et donc le résultat : $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$. En particulier, comme \mathbb{P} est positive, $\mathbb{P}(B \setminus A) \geq 0$ et donc $\mathbb{P}(B) \geq \mathbb{P}(A)$.
- (b) En prenant $B = \Omega$ dans la propriété précédente, on trouve ce résultat.
- (c) Ce point est démontré dans la feuille d'exercice (Exercice 3)
- (iii) Soit $A \in \mathcal{P}(\Omega)$. Comme $A \subseteq \Omega$, par le point (ii).(a) on a $\mathbb{P}(A) \leq \mathbb{P}(\Omega)$ et par la positivité et l'unitarité de la masse de \mathbb{P} , on a bien $0 \leq \mathbb{P}(A) \leq 1$.

□

3.2.3 Exemples

Lancer d'une pièce

Comme nous l'avons expliqué précédemment, l'univers modélisant le lancer d'une pièce est $\Omega = \{0, 1\}$. La loi de probabilité régissant le résultat de la pièce dépend de la pièce elle-même. Le comportement d'une pièce non truquée — on dit qu'elle est *équilibrée* — est modélisé par la loi suivante :

$$\begin{aligned} \mathcal{P}(\Omega) &\longrightarrow \mathbb{R} \\ \emptyset &\longmapsto 0 \\ \{0\} &\longmapsto 1/2 \\ \{1\} &\longmapsto 1/2 \\ \{0, 1\} &\longmapsto 1 \end{aligned}$$

Si, en revanche, la pièce est truquée, on peut modéliser son comportement par la loi suivante :

$$\begin{aligned} \mathcal{P}(\Omega) &\longrightarrow \mathbb{R} \\ \emptyset &\longmapsto 0 \\ \{0\} &\longmapsto q \\ \{1\} &\longmapsto p \\ \{0, 1\} &\longmapsto 1 \end{aligned}$$

où les paramètres p et q sont deux nombres réels entre 0 et 1 qui régissent le comportement de la pièce. Plus q est proche de 1 plus la pièce tombera sur pile avec forte probabilité, plus p est proche de 1, plus la pièce tombera sur face avec forte probabilité.

En fait, un seul paramètre suffit pour décrire le comportement de la pièce puisque $\mathbb{P}(\{0\}) = 1 - \mathbb{P}(\{1\})$, i.e. $q = 1 - p$. Ainsi, la modélisation du lancer d'une pièce (truquée

ou non) est constituée de l'univers $\{0, 1\}$ et de la loi de probabilité suivante :

$$\begin{aligned} \mathcal{P}(\Omega) &\longrightarrow \mathbb{R} \\ \emptyset &\longmapsto 0 \\ \{0\} &\longmapsto 1 - p \\ \{1\} &\longmapsto p \\ \{0, 1\} &\longmapsto 1 \end{aligned}$$

où p est le paramètre mesurant à quel point la pièce est truquée.

Remarque 3.8. Si p vaut $1/2$, la pièce est équilibrée.

Probabilité uniforme

Nous allons décrire ici de manière formelle une famille de lois de probabilité qui décrivent les phénomènes où toutes les issues ont même probabilité d'advenir.

Définition/Proposition 3.9. Soit Ω un univers fini non vide. Il existe une unique loi de probabilité \mathbb{P} sur Ω ayant même valeur sur tous les événements élémentaires, on l'appelle la *probabilité uniforme sur Ω* . Elle est donnée par la formule :

$$\begin{aligned} \mathcal{P}(\Omega) &\longrightarrow \mathbb{R} \\ A &\longmapsto \frac{\#A}{\#\Omega} . \end{aligned}$$

C'est la fameuse formule "nombre de cas favorables sur nombre de cas possibles", la loi de probabilité que l'on entend habituellement lorsque l'on dit que l'on choisit "au hasard" sans préciser la loi de probabilité régissant le hasard dont on parle.

Exemple 3.10. Vous pouvez vérifier que la loi que nous avons donnée précédemment pour modéliser le lancer d'une pièce équilibrée est la probabilité uniforme sur $\Omega = \{0, 1\}$.

Démonstration. On admet l'unicité et on montre l'existence. La probabilité \mathbb{P} qui fonctionne est déjà posée dans l'énoncé, on n'a plus qu'à vérifier que c'est bel est bien une loi de probabilité au sens de notre définition, et qu'elle donne bien même valeur à tous les événements élémentaires.

\mathbb{P} est une loi de probabilité.

Positivité : Pour tout $A \in \mathcal{P}(\Omega)$, on a $\#A \geq 0$. En particulier, on a donc aussi $\#\Omega \geq 0$, et ainsi $\mathbb{P}(A) \geq 0$.

Masse unitaire : $\mathbb{P}(\Omega) = \#\Omega/\#\Omega = 1$.

Additivité : Soient A et B deux événements tels que $A \cap B = \emptyset$. On admet le lemme intuitif suivant :

Lemme 3.11. Si A et B sont deux ensembles finis tels que $A \cap B = \emptyset$, alors $\#(A \cup B) = \#A + \#B$.

On peut donc calculer :

$$\mathbb{P}(A \cup B) = \frac{\#(A \cup B)}{\#\Omega} = \frac{\#A + \#B}{\#\Omega} = \frac{\#A}{\#\Omega} + \frac{\#B}{\#\Omega} = \mathbb{P}(A) + \mathbb{P}(B).$$

\mathbb{P} a même valeur sur tous les événements élémentaires. Soit $\omega \in \Omega$. On a alors $\mathbb{P}(\{\omega\}) = \#\{\omega\}/\#\Omega = 1/\#\Omega$. On a donc bien que la probabilité de n'importe quel événement élémentaire vaut $1/\#\Omega$. \square

3.3 Variables aléatoires

3.3.1 Vers les statistiques

Grâce à la partie précédente, on sait désormais formaliser mathématiquement notre expérience du lancer de pièce. Pour répondre à notre problème initial, i.e. pour déterminer si la pièce est truquée ou non, il nous faut déterminer si la pièce *réelle* suit notre modélisation ou non. Deux problèmes se posent alors :

1. *La pièce ne suit aucune modélisation.* Il est impossible de déterminer si une modélisation est plus qu'un simple mensonge qui ressemble à la réalité. Nous allons cependant devoir supposer que la pièce suit une modélisation (nous prendrons la modélisation décrite en Section 3.2.3) car c'est notre seule chance de pouvoir effectivement dire quelque chose de cette pièce.
2. *Le paramètre p de la modélisation reste inconnu.* Même en supposant que la pièce suit notre modélisation, le paramètre qui régit le comportement de la pièce reste inaccessible. Comment faire pour deviner quel p se cache derrière la pièce que je vous tends ?

La seule solution que vous avez, et la seule chose que n'importe qui ferait, est de lancer la pièce un certain nombre de fois pour en déduire une valeur probable de p . Cette démarche est celle des *statistiques inférentielles* :

1. Enregistrer plusieurs résultats de lancers de pièce : $r_1, \dots, r_n \in \{0, 1\}$. Ce sont les *données* ou l'*échantillon*.
2. Calculer à partir des données un *estimateur* \hat{p} censé approcher p .

Deux questions se soulèvent alors naturellement : quel estimateur prendre ? En quel sens est-il proche de p ?

Ici, un choix d'estimateur est assez intuitif : on peut considérer la proportion de 1 (face) dans nos lancers de pièce, i.e. le nombre de faces divisé par le nombre de lancers, i.e.

$$\hat{p} = \frac{\#\text{faces}}{\#\text{lancers}} = \frac{r_1 + \dots + r_n}{n}.$$

On voit l'intérêt de prendre la valeur 1 pour représenter l'issue "face" et 0 pour représenter l'issue "pile". Le nombre de faces dans une succession de lancers devient simplement $r_1 + \dots + r_n$.

En quel sens cet estimateur est-il proche de la probabilité de faire face ? On souhaite préciser mathématiquement, c'est-à-dire par une définition de "proche" et un théorème le démontrant, le fait que notre estimateur est proche du paramètre p . Notez que la notion de proche étant toute relative et n'ayant aucun sens absolu (il y a une infinité de nombres séparant 0 de 0,000000001), nous préciserons plus tard ce que nous entendons par "proche". Pour l'instant, voyons plutôt à quel théorème nous devons arriver. Informellement, nous pourrions l'écrire ainsi :

Théorème 3.12 (Informel 1). *Soient r_1, \dots, r_n les résultats de n lancers de la pièce. L'estimateur \hat{p} est proche de p .*

Une remarque très importante : il n'y a aucune chance que \hat{p} soit proche de p de manière certaine. Il est tout à fait possible de faire 1000 faces d'affilée avec une pièce équilibrée. La seule chose que nous avons une chance de pouvoir prouver est qu'il est *probable* que \hat{p} soit proche de p . Le théorème devrait donc plus ressembler à ceci :

Théorème 3.13 (Informel 2). *Soient r_1, \dots, r_n les résultats de n lancers de la pièce. La probabilité $\mathbb{P}(\hat{p} \text{ proche de } p)$ est proche de 1.*

Un nouveau problème se pose alors. L'estimateur \hat{p} est calculé à partir des résultats r_1, \dots, r_n . Une fois fixés par la phrase "*Soient r_1, \dots, r_n les résultats [...]*", il n'y a plus de hasard, et \hat{p} a une valeur déterminée par celles des résultats. Ainsi, parler de la probabilité $\mathbb{P}(\hat{p} \text{ proche de } p)$ n'a plus de sens, puisqu'on ne peut parler de probabilité que pour les situations où le hasard intervient.

Il nous faudrait pouvoir parler des résultats de la pièce sans fixer leurs valeurs. Autrement dit, il faudrait que r_1, \dots, r_n soient des "nombres dont la valeur n'est pas fixée et dépend des résultats de la pièce". En mathématiques, il existe des objets jouant le rôle de "nombres dont la valeur dépend du résultat d'une expérience aléatoire", et ces nombres s'appellent des *variables aléatoires*. Nous les définissons dans la section suivante.

Dans notre situation, en considérant les *variables aléatoires "résultats de la pièce"* R_1, \dots, R_n , on pourra alors définir de la même manière l'estimateur :

$$\hat{P} = \frac{R_1 + \dots + R_n}{n},$$

qui sera cette fois lui-même une variable aléatoire puisque sa valeur dépend des valeurs des variables "résultats". L'estimateur \hat{P} sera donc un "nombre dont la valeur dépend du hasard", et cela aura un sens de parler de $\mathbb{P}(\hat{P} \text{ proche de } p)$. Le théorème devrait donc ressembler à ceci :

Théorème 3.14 (Informel 3). *Soient R_1, \dots, R_n les variables aléatoires "résultats de la pièce". La probabilité $\mathbb{P}(\hat{P} \text{ proche de } p)$ est proche de 1.*

3.3.2 Définition

Dans toute cette section, on se fixe un univers fini Ω et une loi de probabilité \mathbb{P} sur Ω .

Définition 3.15. On appelle *variable aléatoire réelle (VAR)* sur Ω une application $X : \Omega \rightarrow \mathbb{R}$.

Intuition 3.16. On peut penser à une variable aléatoire comme à un *nombre* dont la valeur dépend de l'issue de l'expérience aléatoire modélisée par la loi \mathbb{P} sur Ω . Elles se manipulent comme n'importe quelle autre variable mathématique et, c'est un de leurs intérêts, on peut donc faire des calculs avec.

Un autre fruit de la notion de variable aléatoire est qu'elle donne un moyen simple de parler d'une grande diversité de quantités numériques dont la valeur dépend de l'issue d'une unique expérience aléatoire, comme l'illustrent les exemples suivants.

Exemple 3.17. Pour le jeu introductif modélisé par l'univers $\Omega = \{0, 1\}$, on peut définir par exemple :

1. la variable aléatoire *résultat de la pièce*, définie par :

$$\begin{array}{l} \Omega \longrightarrow \mathbb{R} \\ R : 0 \longmapsto 0 \text{ ,} \\ \quad 1 \longmapsto 1 \end{array}$$

2. ou encore la variable aléatoire *gain*, définie par :

$$\begin{array}{l} \Omega \longrightarrow \mathbb{R} \\ G : 0 \longmapsto -100 \text{ .} \\ \quad 1 \longmapsto 100 \end{array}$$

Exemple 3.18. Lorsque l'on tire au hasard uniformément une personne parmi les patients en réanimation en France, l'univers est l'ensemble des personnes en réanimation muni de la probabilité uniforme et l'on peut définir les variables aléatoires : l'âge de la personne tirée A , le poids de la personne tirée P , etc... Autant de quantité dépendant du hasard dont il pourrait être intéressant d'extraire des informations pour inférer des informations sur population en réanimation. L'avantage de la notion de variable aléatoire est qu'on ne change pas d'univers pour étudier ces différents quantités dépendant du hasard, on ne fait que changer de variable.

Exemple 3.19. Considérons le lancer de deux pièces. On peut modéliser cette expérience aléatoire par l'univers $\Omega = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ où l'on a noté chaque issue (x, y) où x est le résultat de la première pièce et y le résultat de la deuxième pièce. On peut alors définir deux variables aléatoires :

$$R_1 : \begin{array}{l} \Omega \longrightarrow \mathbb{R} \\ (x, y) \longmapsto x \text{ ,} \end{array}$$

le résultat de la première pièce et

$$R_2 : \begin{array}{ccc} \Omega & \longrightarrow & \mathbb{R} \\ (x, y) & \longmapsto & y \end{array} ,$$

le résultat de la deuxième pièce. Pour mieux comprendre le comportement de ces deux variables, on peut calculer leurs valeurs sur des exemples : $R_1(0, 1) = 0$, $R_1(1, 1) = 1$, $R_2(0, 1) = 1$ ou encore $R_2(1, 1) = 1$.

Exemple 3.20. Etant donné deux VAR X et Y sur Ω , on peut définir leur *somme*, notée $X + Y$, constituant une nouvelle variable aléatoire :

$$X + Y : \begin{array}{ccc} \Omega & \longrightarrow & \mathbb{R} \\ \omega & \longmapsto & X(\omega) + Y(\omega) \end{array} .$$

De même, on peut construire la variable aléatoire produit $X \cdot Y$, ou encore $2 \cdot X$, ou bien X^2 , etc...

Notation 3.21. Si X est une VAR sur Ω , on a un moyen très simple de noter les évènements qui vont nous intéresser à son propos. Par exemple, si $x, y, z \in \mathbb{R}$, on note :

$$\begin{aligned} \{X = x\} &= \{\omega \in \Omega \mid X(\omega) = x\}, \\ \{y < X < z\} &= \{\omega \in \Omega \mid y < X(\omega) < z\}, \end{aligned}$$

qui sont bien des évènements, i.e. des éléments de $\mathcal{P}(\Omega)$. Ceci nous permet d'écrire de manière suggestive $\mathbb{P}(X = 3)$ ou encore $\mathbb{P}(2 < X < 6)$.

Exemple 3.22. Prenons l'exemple 3.19 du lancer de deux pièces. On peut alors calculer les évènements suivants, pour $x \in \{0, 1\}$ et $y \in \{0, 1\}$:

$$\begin{aligned} \{R_1 = x\} &= \{(s, t) \in \Omega \mid R_1(s, t) = x\} \\ &= \{(s, t) \in \Omega \mid s = x\} \\ &= \{(x, 0), (x, 1)\} \end{aligned}$$

et

$$\begin{aligned} \{R_2 = y\} &= \{(s, t) \in \Omega \mid R_2(s, t) = y\} \\ &= \{(s, t) \in \Omega \mid t = y\} \\ &= \{(0, y), (1, y)\}. \end{aligned}$$

3.3.3 Indépendance

Lors de nos études statistiques, nous allons être amenés à répéter la même expérience aléatoire plusieurs fois en considérant que chaque répétition est indépendante de la précédente. Nous définissons ici en quel sens nous considérons que deux variables aléatoires sont indépendantes.

Définition 3.23. Soient X et Y deux VAR sur Ω . On dit que X et Y sont *indépendantes* si pour tout $x \in \mathbb{R}$ et tout $y \in \mathbb{R}$,

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y),$$

où l'on a noté : $\mathbb{P}(X = x, Y = y) = \mathbb{P}(\{X = x\} \cap \{Y = y\})$.

Intuition 3.24. La définition précédente est motivée par les considérations suivantes. Lorsque l'on lance deux fois de suite une pièce de monnaie, les résultats des deux lancers ne s'influencent pas mutuellement, et on a l'intuition que la probabilité de faire pile au premier *et* au deuxième lancer est le *produit* des deux probabilités. Si au contraire, nous tirions des boules dans une urne sans remettre les boules après le tirage, les résultats ne seraient pas indépendants les uns des autres⁴ et ceci s'exprimerait par le fait que la probabilité d'obtenir un certain numéro au premier et un certain autre au deuxième ne serait par le produit des probabilités.

Exemple 3.25. Considérons le lancer de deux pièces (Exemple 3.19). Les variables aléatoires R_1 et R_2 sont-elles indépendantes? Cela dépend de la loi \mathbb{P} que nous choisissons sur Ω . Dans la situation la plus classique, on considère la probabilité uniforme sur Ω , qui modélise le lancer de deux pièces équilibrées. Dans ce cas, on peut vérifier que R_1 et R_2 sont indépendantes. Soit $x \in \mathbb{R}$ et $y \in \mathbb{R}$. Dans un premier temps, supposons que $x \in \{0, 1\}$ et $y \in \{0, 1\}$. Dans ce cas, on a montré précédemment que :

$$\begin{aligned} \{R_1 = x\} &= \{(x, 0), (x, 1)\}, \\ \{R_2 = y\} &= \{(0, y), (1, y)\}, \end{aligned}$$

et on peut alors vérifier que :

$$\{R_1 = x\} \cap \{R_2 = y\} = \{(x, y)\}.$$

Ainsi, on a :

$$\mathbb{P}(R_1 = x, R_2 = y) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = \mathbb{P}(R_1 = x) \times \mathbb{P}(R_2 = y),$$

en utilisant le fait que \mathbb{P} est la probabilité uniforme.

Si maintenant $x \notin \{0, 1\}$ (le cas $y \notin \{0, 1\}$ est symétrique), on a que $\{R_1 = x\} = \emptyset$ et $\{R_1 = x, R_2 = y\} = \emptyset$, d'où

$$\mathbb{P}(R_1 = x, R_2 = y) = 0 = 0 \times \mathbb{P}(R_2 = y) = \mathbb{P}(R_1 = x) \times \mathbb{P}(R_2 = y).$$

3.3.4 Variables aléatoires suivant une loi uniforme

On sera souvent amené à considérer des variables aléatoires qui peuvent prendre un certain nombre de valeurs avec égale probabilité, on dit qu'elles suivent la loi uniforme. Fixons dans cette sous-section une VAR X sur Ω .

4. puisqu'il y aurait une boule de moins à chaque tirage.

Définition 3.26. On appelle *univers image de X* l'ensemble :

$$X(\Omega) = \{x \in \mathbb{R} \mid \text{il existe } \omega \in \Omega, X(\omega) = x\}.$$

Intuition 3.27. L'univers image est l'ensemble des valeurs prises par X .

Exemple 3.28. On peut par exemple calculer dans l'exemple 3.19 :

$$R_1(\Omega) = \{0, 1\},$$

$$R_2(\Omega) = \{0, 1\}.$$

Remarque 3.29. Comme l'univers Ω est fini, l'univers image l'est aussi (admis). On peut numéroter leurs éléments ainsi :

$$\begin{aligned}\Omega &= \{\omega_1, \dots, \omega_n\}, \\ X(\Omega) &= \{x_1, \dots, x_m\},\end{aligned}$$

où $n = \#\Omega$ et $m = \#X(\Omega)$.

Définition 3.30. On dit que X suit une loi uniforme si pour tout $x \in X(\Omega)$, on a $\mathbb{P}(X = x) = \frac{1}{\#X(\Omega)}$.

Exemple 3.31. Dans l'exemple 3.19, la variable R_1 suit une loi uniforme. En effet, $\#R_1(\Omega) = 2$ et pour $x \in \{0, 1\}$, on a :

$$\mathbb{P}(R_1 = x) = \frac{\#\{(x, 0), (x, 1)\}}{\#\Omega} = \frac{2}{4} = \frac{1}{2} = \frac{1}{\#R_1(\Omega)}.$$

La variable R_2 suit-elle une loi uniforme ?

3.3.5 Variables aléatoires suivant une loi de Bernoulli

Dans les expériences statistiques que nous allons faire, les variables aléatoires qui nous intéressent ne prendront pour valeurs que 0 ou 1 pour représenter le fait que la pièce tombe soit sur pile, soit sur face. Ce sont des variables de Bernoulli.

Définition 3.32. Soit $p \in [0, 1]$. On dit qu'une VAR X sur Ω est une *variable de Bernoulli de paramètre p* , ou qu'elle *suit la loi de Bernoulli de paramètre p* , si :

- (i) $X(\Omega) = \{0, 1\}$,
- (ii) $\mathbb{P}(X = 1) = p$.

Remarque 3.33. Si X est une variable de Bernoulli de paramètre p , on a pour $x \in \mathbb{R}$,

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

Exemple 3.34. Le résultat du lancer d'une pièce de monnaie est une variable de Bernoulli. Si l'on note R ce résultat, on a bien :

$$\mathbb{P}(R = x) = \begin{cases} p & \text{si } x = 1, \\ 1 - p & \text{si } x = 0, \\ 0 & \text{sinon.} \end{cases}$$

3.4 Espérance

Dans cette section, nous introduisons une quantité calculée à partir d'une variable aléatoire et qui contient de l'information sur celle-ci. C'est une situation typique en mathématiques : pour appréhender un objet complexe contenant toute l'information sur notre problème (la variable aléatoire) les mathématiciens introduisent des objets plus simples (souvent des nombres) et facilement calculables (ici l'espérance en est un exemple) qui contiennent de l'information (souvent partielle, mais pertinente) et qui soit directement compréhensible, sur l'objet en question.

3.4.1 Introduction

L'espérance est une quantité qui apparaît naturellement lorsque l'on se pose la question suivante : comment savoir si un jeu d'argent (tel que celui proposé en introduction de ce chapitre) est avantageux ou non ? Considérons un jeu de lancer de pièce : pile rapporte 100 € et face fait perdre 100 €. Supposons cependant que nous puissions y jouer avec deux pièces différentes :

1. une pièce équilibrée pour laquelle $\mathbb{P}(G_1 = 100)$, noté p_1 , vaut $1/2$,
2. une pièce pour laquelle $\mathbb{P}(G_2 = 100)$, noté p_2 , vaut $1/3$,

où l'on a noté G_i le gain du jeu utilisant la pièce numéro i . Notez que pour une telle pièce, $\mathbb{P}(G_i = -100) = 1 - p_i$.

Notre intuition nous fait sentir que le jeu 2 est moins intéressant que le jeu 1. Plus précisément, on a le sentiment que si l'on joue un grand nombre de parties, on va perdre en moyenne environ 33 € au jeu 2 alors que le jeu 1 ne nous ferait ni perdre ni gagner en moyenne.

Faisons une petite expérience pour illustrer le bien fondé de cette intuition. Si l'on joue un grand nombre de parties d'un de ces jeux, le gain total aura pour formule :

$$\text{Gain} = 100\text{€} \times \text{Nbre piles} - 100\text{€} \times \text{Nbre faces}.$$

Le gain moyen aura donc pour expression :

$$\text{Gain moyen} = \frac{\text{Gain}}{\text{Nbre de lancers}} = 100\text{€} \times \frac{\text{Nbre piles}}{\text{Nbre de lancers}} - 100\text{€} \times \frac{\text{Nbre faces}}{\text{Nbre de lancers}}.$$

Quand le nombre de lancers augmentent, nous avons l'intuition que :

$$\frac{\text{Nbre piles}}{\text{Nbre de lancers}} \xrightarrow{\text{Nbre lancers} \rightarrow +\infty} \mathbb{P}(\text{faire pile}),$$

$$\frac{\text{Nbre faces}}{\text{Nbre de lancers}} \xrightarrow{\text{Nbre lancers} \rightarrow +\infty} \mathbb{P}(\text{faire face}).$$

C'est effectivement ce qu'il se passe en pratique, comme le montre les figures 3.1 et 3.2. Le gain moyen tend donc vers une valeur moyenne :

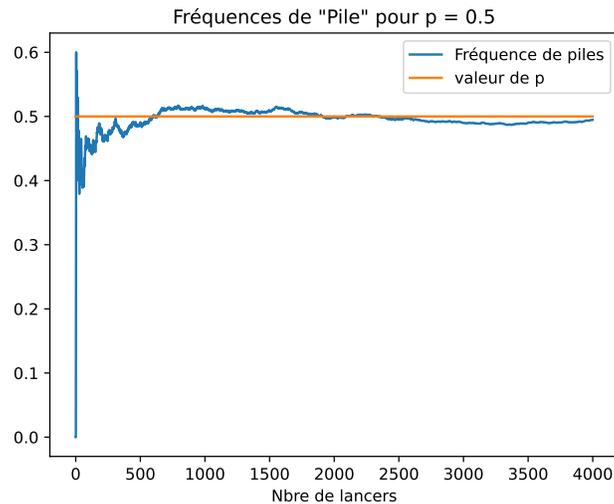


FIGURE 3.1 – Fréquences de faces en fonction du nombre de lancers pour le jeu 1.

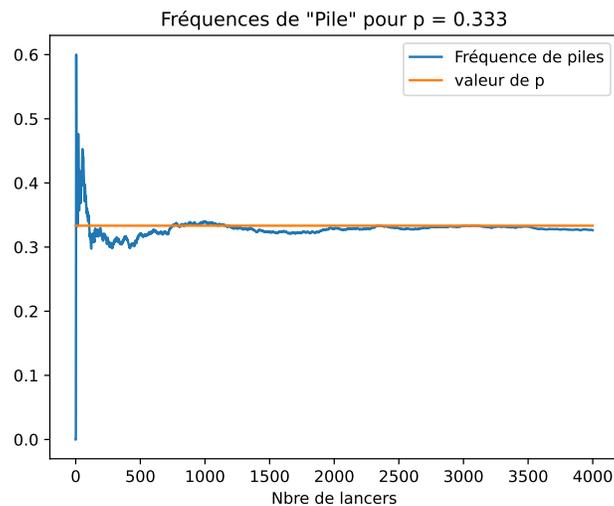


FIGURE 3.2 – Fréquences de faces en fonction du nombre de lancers pour le jeu 2.

$$\text{Gain moyen} \xrightarrow{\text{Nbre lancers} \rightarrow +\infty} 100\text{€} \times \mathbb{P}(\text{faire pile}) - 100\text{€} \times \mathbb{P}(\text{faire face}),$$

que l'on appelle l'*espérance de gain* et qui vaut bien 0 pour le jeu 1 et -33 pour le jeu 2. La notion d'espérance fait sens pour n'importe quelle variable aléatoire réelle d'univers image fini et est un indicateur de la *valeur moyenne* de la variable aléatoire. L'espérance est facile à calculer et contient de l'information pertinente sur la variable aléatoire considérée. Après avoir défini la notion d'espérance, nous montrerons un théorème précisant en quel sens le gain moyen tend vers l'espérance lorsque le nombre de lancers augmente, un théorème appelé *la loi des grands nombres*.

3.4.2 Définition

Soit X une VAR sur Ω . On note $X(\Omega) = \{x_1, \dots, x_m\}$ où $m = \#X(\Omega)$.

Définition 3.35. On appelle *espérance de X* , et l'on note $\mathbb{E}[X]$, le nombre réel défini par :

$$\mathbb{E}[X] = x_1 \cdot \mathbb{P}(X = x_1) + \dots + x_m \cdot \mathbb{P}(X = x_m).$$

Intuition 3.36. L'espérance d'une variable aléatoire représente la "valeur moyenne" de cette variable aléatoire, où l'on pondère chaque valeur possible par la probabilité que la variable prenne cette valeur. C'est un indicateur numérique simple qui donne de l'information sur une variable aléatoire.

Remarque 3.37. On peut également noter la somme $x_1 \cdot \mathbb{P}(X = x_1) + \dots + x_m \cdot \mathbb{P}(X = x_m)$ à l'aide de la notation suivante :

$$\sum_{x \in X(\Omega)} x \cdot \mathbb{P}(X = x),$$

de sorte que :

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

Remarque 3.38. La formule

$$\mathbb{E}[X] = x_1 \cdot \mathbb{P}(X = x_1) + \dots + x_m \cdot \mathbb{P}(X = x_m).$$

reste valable si l'on a seulement $X(\Omega) \subseteq \{x_1, \dots, x_m\}$, car si l'un des x_i n'est pas dans l'univers image, alors $\mathbb{P}(X = x_i) = 0$, et le terme n'intervient pas dans la formule.

Exemple 3.39. Si l'on considère l'exemple introductif de cette section, on peut calculer :

$$\begin{aligned} \mathbb{E}[G_i] &= 100 \times \mathbb{P}(G_i = 1) - 100 \times \mathbb{P}(G_i = -2) \\ &= 100 \times p_i - 100 \times (1 - p_i) \\ &= 100 \times (2p_i - 1), \end{aligned}$$

et le calcul donne donc :

$$\begin{aligned} \mathbb{E}[G_1] &= 0, \\ \mathbb{E}[G_2] &= -\frac{100}{3} \simeq -33,3. \end{aligned}$$

Vocabulaire 3.40. Une VAR X sur Ω telle que $\mathbb{E}[X] = 0$ est dite *centrée*. Si la variable aléatoire de gain d'un jeu est centré, le jeu est considéré comme équitable.

Un exemple très important.

Exemple 3.41 (Variable de Bernoulli). Si $p \in [0, 1]$ et X est une variable de Bernoulli sur Ω de paramètre p , on peut alors calculer :

$$\mathbb{E}[X] = 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) = p.$$

3.4.3 Propriétés

Proposition 3.42. Soit X une VAR sur Ω . On numérote les issues ainsi : $\Omega = \{\omega_1, \dots, \omega_n\}$ où $n = \#\Omega$.

(i) Si X est constante de valeur C , on a $\mathbb{E}[X] = C$.

(ii) (admis)

$$\mathbb{E}[X] = X(\omega_1) \cdot \mathbb{P}(\{\omega_1\}) + \dots + X(\omega_n) \cdot \mathbb{P}(\{\omega_n\}).$$

(iii) (Positivité) Si $X \geq 0$ ⁵, alors $\mathbb{E}[X] \geq 0$.

(iv) (Linéarité) Soient $\lambda \in \mathbb{R}$ et Y une VAR sur Ω . Alors,

$$\begin{aligned} \mathbb{E}[\lambda \cdot X] &= \lambda \cdot \mathbb{E}[X], \\ \mathbb{E}[X + Y] &= \mathbb{E}[X] + \mathbb{E}[Y]. \end{aligned}$$

Démonstration. Pour le point (i), si X est constante de valeur C , alors $X(\Omega) = \{C\}$ et $\{X = C\} = \Omega$ donc $\mathbb{P}(X = C) = \mathbb{P}(\Omega) = 1$, et ainsi $\mathbb{E}[X] = C \cdot \mathbb{P}(X = C) = C$.

On admet le point (ii), qui nécessiterait plus de technique de manipulation de sommes finies que nous n'en avons pour l'instant.

Montrons le (iii). Supposons que $X \geq 0$. On peut alors calculer par le (ii) :

$$\mathbb{E}[X] = X(\omega_1) \cdot \mathbb{P}(\{\omega_1\}) + \dots + X(\omega_n) \cdot \mathbb{P}(\{\omega_n\})$$

où chaque terme est positif car X est positive et \mathbb{P} est positive, l'espérance l'est donc aussi.

Montrons le (iv). On peut calculer :

$$\begin{aligned} \mathbb{E}[\lambda \cdot X] &= \lambda \cdot X(\omega_1) \cdot \mathbb{P}(\{\omega_1\}) + \dots + \lambda \cdot X(\omega_n) \cdot \mathbb{P}(\{\omega_n\}) \\ &= \lambda \cdot [X(\omega_1) \cdot \mathbb{P}(\{\omega_1\}) + \dots + X(\omega_n) \cdot \mathbb{P}(\{\omega_n\})] \\ &= \lambda \cdot \mathbb{E}[X]. \end{aligned}$$

Ainsi que :

$$\begin{aligned} \mathbb{E}[X + Y] &= (X(\omega_1) + Y(\omega_1)) \cdot \mathbb{P}(\{\omega_1\}) + \dots + (X(\omega_n) + Y(\omega_n)) \cdot \mathbb{P}(\{\omega_n\}) \\ &= X(\omega_1) \cdot \mathbb{P}(\{\omega_1\}) + Y(\omega_1) \cdot \mathbb{P}(\{\omega_1\}) + \dots \\ &\quad + X(\omega_n) \cdot \mathbb{P}(\{\omega_n\}) + Y(\omega_n) \cdot \mathbb{P}(\{\omega_n\}) \\ &= X(\omega_1) \cdot \mathbb{P}(\{\omega_1\}) + \dots + X(\omega_n) \cdot \mathbb{P}(\{\omega_n\}) \\ &\quad + Y(\omega_1) \cdot \mathbb{P}(\{\omega_1\}) + \dots + Y(\omega_n) \cdot \mathbb{P}(\{\omega_n\}) \\ &= \mathbb{E}[X] + \mathbb{E}[Y]. \end{aligned}$$

5. au sens où pour tout $\omega \in \Omega$, $X(\omega) \geq 0$.

□

Proposition 3.43 (admise). *Si X et Y sont deux VAR indépendantes sur Ω , alors $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$.*

3.4.4 Variance

Dans cette section, nous introduisons un nouveau descripteur numérique contenant d'une variable aléatoire. Il est assez clair que l'espérance est une connaissance très partielle du comportement d'une variable. Par exemple, une variable prenant avec équiprobabilité les valeurs 1 et -1 et une autre prenant avec équiprobabilité les valeurs 100 et -100 ont même espérance : 0. Il y a pourtant une différence très importante entre les deux : l'écart que ces variables ont en moyenne avec leur espérance est grand.

L'écart moyen à la moyenne est une donnée très importante sur la variable aléatoire : un élève qui a deux fois 10 et un élève qui a 0 puis 20 sont extrêmement différents, bien qu'ils aient la même moyenne. La différence est contenue s'exprime par le fait que l'écart à la moyenne est nul pour le premier et de 10 pour le second. Cet écart moyen est appelé *écart-type* et son carré est appelé *variance*. En voici les définitions formelles.

Définition 3.44. Soit X une VAR sur Ω .

- (i) On appelle *variance de X* , et l'on note $\mathbb{V}(X)$, le nombre $\mathbb{E}[(X - \mathbb{E}[X])^2]$.
- (ii) On appelle *écart-type de X* , et l'on note $\sigma(X)$, le nombre $\sqrt{\mathbb{V}(X)}$.

Intuition 3.45. La formule définissant la variance s'interprète ainsi : $\mathbb{E}[X]$ représentant la valeur moyenne de X , la variable aléatoire $(X - \mathbb{E}[X])^2$ représente l'écart *quadratique*⁶ (aléatoire) entre X et sa moyenne. La variance calcule la *moyenne* de cette écart quadratique. L'écart-type est la racine de la variance, c'est l'écart quadratique moyen de X à sa moyenne.

Considérer des écarts quadratiques a pour effet d'exacerber les gros écarts à la moyenne et de minimiser les petits. On dit que l'écart-type est la *moyenne quadratique* des écarts à la moyenne.

Proposition 3.46. *Soit X est une variable aléatoire réelle sur Ω . L'on a :*

- (i) $\mathbb{V}(X) \geq 0$,
- (ii) pour tout $\lambda \in \mathbb{R}$, on a $\mathbb{V}(\lambda X) = \lambda^2 \mathbb{V}(X)$.

Démonstration. Par positivité de l'espérance, une variance est toujours un nombre positif, car la variable aléatoire $(X - \mathbb{E}[X])^2$ est positive. De plus, par linéarité de l'espérance,

$$\begin{aligned} \mathbb{V}(\lambda X) &= \mathbb{E}[(\lambda X - \mathbb{E}[\lambda X])^2] \\ &= \mathbb{E}[(\lambda X - \lambda \mathbb{E}[X])^2] \\ &= \mathbb{E}[\lambda^2 (X - \mathbb{E}[X])^2] \\ &= \lambda^2 \mathbb{E}[(X - \mathbb{E}[X])^2]. \end{aligned}$$

6. élevé au carré

□

Exemple 3.47. Considérons deux variables aléatoires X_1 et X_2 sur Ω telles que $X_1(\Omega) = \{-10, 8, 10\}$ et $X_2(\Omega) = \{-10, 4, 10\}$, et :

$$\begin{aligned} \mathbb{P}(X_1 = -10) &= 0,2 & \mathbb{P}(X_1 = 8) &= 0,5 & \mathbb{P}(X_1 = 10) &= 0,3, \\ \mathbb{P}(X_2 = -10) &= 0,1 & \mathbb{P}(X_2 = 4) &= 0,5 & \mathbb{P}(X_2 = 10) &= 0,4. \end{aligned}$$

On peut alors calculer :

$$\begin{aligned} \mathbb{E}[X_1] &= -10 \times 0,2 + 8 \times 0,5 + 10 \times 0,3 = 5, \\ \mathbb{E}[X_2] &= -10 \times 0,1 + 4 \times 0,5 + 10 \times 0,4 = 5. \end{aligned}$$

Dans ce cas, les variables aléatoires $Y_i = (X_i - \mathbb{E}[X_i])^2$ ont pour univers image

$$\begin{aligned} Y_1(\Omega) &= \{9, 25, 225\}, \\ Y_2(\Omega) &= \{1, 25, 225\}, \end{aligned}$$

et prennent ces valeurs avec probabilités :

$$\begin{aligned} \mathbb{P}(Y_1 = 9) &= 0,5, & \mathbb{P}(Y_1 = 25) &= 0,3, & \mathbb{P}(Y_1 = 225) &= 0,2, \\ \mathbb{P}(Y_2 = 1) &= 0,5, & \mathbb{P}(Y_2 = 25) &= 0,3, & \mathbb{P}(Y_2 = 225) &= 0,2. \end{aligned}$$

On peut donc calculer

$$\begin{aligned} \mathbb{V}(X_1) &= 9 \times \mathbb{P}(Y_1 = 9) + 25 \times \mathbb{P}(Y_1 = 25) + 225 \times \mathbb{P}(Y_1 = 225), \\ \mathbb{V}(X_2) &= 1 \times \mathbb{P}(Y_2 = 1) + 25 \times \mathbb{P}(Y_2 = 25) + 225 \times \mathbb{P}(Y_2 = 225), \end{aligned}$$

et les variances et les écarts-types valent donc :

$$\begin{aligned} \mathbb{V}(X_1) &= 57, & \sigma(X_1) &\simeq 7,5, \\ \mathbb{V}(X_2) &= 53, & \sigma(X_2) &\simeq 7,3. \end{aligned}$$

L'écart-type (et la variance) de X_2 est plus faible que celui de X_1 et c'est bien normal car X_2 — qui vaut 4 avec probabilité 0,5 — est en moyenne plus proche de sa moyenne que X_1 — qui vaut 8 avec probabilité 0,5 (les autres probabilités des deux variables étant similaires par ailleurs).

Exemple 3.48. Soit X une variable aléatoire de Bernoulli de paramètre $p \in [0, 1]$. Calculons la variance et l'écart-type de X . Pour cela, et pour changer, utilisons la formule de König-Huygens (voir la feuille d'exercices) $\mathbb{V}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. De manière remarquable, la variable aléatoire X^2 est également une variable aléatoire de Bernoulli de même paramètre p que X . L'on a donc :

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p - p^2 = p(1 - p), \\ \sigma(X) &= \sqrt{p(1 - p)}. \end{aligned}$$

Proposition 3.49. *Si X et Y sont deux variables aléatoires réelles indépendantes sur Ω , alors*

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y).$$

Démonstration. On peut calculer, par la formule de König-Huygens,

$$\begin{aligned} \mathbb{V}(X + Y) &= \mathbb{E}[(X + Y)^2] - \mathbb{E}[X + Y]^2 \\ &= \mathbb{E}[X^2 + 2XY + Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - 2\mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]^2. \end{aligned}$$

Or, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ par indépendance de X et Y . Ainsi,

$$\begin{aligned} \mathbb{V}(X + Y) &= \mathbb{E}[X^2] + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - 2\mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]^2 \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 + \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 \\ &= \mathbb{V}(X) + \mathbb{V}(Y). \end{aligned}$$

□

3.4.5 Mesure de l'écart à la moyenne

Dans cette partie, nous allons montrer un résultat appelé *inégalité de Bienaymé-Tchebychev* — en hommage aux mathématiciens Irénée-Jules Bienaymé (1796-1878), un français, et Pafnouti Tchebychev (1821-1894), un russe — qui exprime que l'écart entre X et son espérance est mesurée par la variance de X .

Théorème 3.50 (Inégalité de Bienaymé-Tchebychev). *Soit X une variable aléatoire réelle sur Ω . Pour tout $\varepsilon > 0$, l'on a :*

$$1 - \frac{\mathbb{V}(X)}{\varepsilon^2} \leq \mathbb{P}(\mathbb{E}[X] - \varepsilon < X < \mathbb{E}[X] + \varepsilon) \leq 1.$$

Pour le démontrer, il nous faut tout d'abord l'énoncé suivant, portant le nom du mathématicien russe Andreï Markov (1856-1922).

Proposition 3.51 (Inégalité de Markov). *Soit X une variable aléatoire réelle positive sur Ω , i.e. telle que $X \geq 0$. Pour tout $\varepsilon > 0$, l'on a :*

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}[X]}{\varepsilon}.$$

Démonstration. On pose Y la variable aléatoire suivante :

$$Y : \begin{array}{l} \Omega \longrightarrow \mathbb{R} \\ \omega \longmapsto \begin{cases} \varepsilon & \text{si } X(\omega) \geq \varepsilon, \\ 0 & \text{sinon.} \end{cases} \end{array}$$

L'univers image de Y est alors $Y(\Omega) \subseteq \{0, \varepsilon\}$ et l'on a :

$$\{Y = \varepsilon\} = \{X \geq \varepsilon\}.$$

Ainsi, on peut calculer, grâce à la remarque 3.38 :

$$\mathbb{E}[Y] = \varepsilon \times \mathbb{P}(Y = \varepsilon) = \varepsilon \times \mathbb{P}(X \geq \varepsilon).$$

De plus, on peut remarquer que pour tout $\omega \in \Omega$, on a $X(\omega) - Y(\omega) \leq 0$. Ainsi, $X - Y$ est positive et par positivité de l'espérance, on $\mathbb{E}[X - Y] \geq 0$, i.e. $\mathbb{E}[X] \geq \mathbb{E}[Y]$, i.e. $\mathbb{E}[X] \geq \varepsilon \times \mathbb{P}(X \geq \varepsilon)$. Finalement, on a donc :

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}[X]}{\varepsilon}.$$

□

Preuve de l'inégalité de Bienaymé-Tchebychev. L'inégalité de droite est vraie par la proposition 3.6.(iii). Pour celle de gauche, on pose $Y = (X - \mathbb{E}[X])^2$. Dans ce cas, Y est une variable aléatoire positive sur Ω , de sorte que par inégalité de Markov appliquée à Y et $\varepsilon^2 > 0$, l'on a :

$$\mathbb{P}(Y \geq \varepsilon^2) \leq \frac{\mathbb{E}[Y]}{\varepsilon^2}.$$

Or, $\mathbb{E}[Y] = \mathbb{V}(X)$ par définition de la variance, donc :

$$\mathbb{P}(Y \geq \varepsilon^2) \leq \frac{\mathbb{V}(X)}{\varepsilon^2}.$$

De plus, $\mathbb{P}(Y \geq \varepsilon^2) = 1 - \mathbb{P}(Y < \varepsilon^2)$, donc :

$$\mathbb{P}(Y < \varepsilon^2) \geq 1 - \frac{\mathbb{V}(X)}{\varepsilon^2}.$$

Pour conclure, il suffit maintenant de remarquer que

$$\{Y < \varepsilon^2\} = \{\mathbb{E}[X] - \varepsilon < X < \mathbb{E}[X] + \varepsilon\},$$

en utilisant le fait que pour tout nombre $x \in \mathbb{R}$ et $\alpha \geq 0$, on a que $x^2 \leq \alpha^2$ équivaut à $-\alpha \leq x \leq \alpha$. □

3.5 Statistiques et loi des grands nombres

Cette section n'est pas exigible à l'examen. Considérons à nouveau l'estimation de la probabilité de faire face avec une pièce inconnue. On exécute des lancers successifs de la pièce. Si l'on répète cette expérience un grand nombre de fois, nous avons l'intuition que :

$$\frac{\text{Nombre de face}}{\text{Nombre de lancers}} \xrightarrow{\text{Nombre lancers} \rightarrow +\infty} \mathbb{P}(\text{faire face}) = p.$$

Le membre de gauche de cette expression va nous servir d'estimateur pour p , c'est lui que nous calculerons en pratique à partir de nos résultats concrets et qui fournira une approximation de p .

Modélisation d'une répétition de lancers. L'expérience d'un unique lancer de pièce est modélisée par une variable de Bernoulli R de paramètre p , comme expliqué en Exemple 3.34. **On modélise la répétition de lancers indépendants par une suite de variables de Bernoulli indépendantes de même paramètre p .** Voici les définitions de ces notions.

Définition 3.52. On appelle *suite de variables aléatoires*, une application $R : \mathbb{N}^* \rightarrow \mathcal{F}(\Omega, \mathbb{R})$, où $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$ et $\mathcal{F}(\Omega, \mathbb{R})$ est l'ensemble des applications de Ω dans \mathbb{R} , i.e. l'ensemble des variables aléatoires réelles sur Ω .

Notation 3.53. Comme pour les suites réelles, on note R_n l'image de n par R . Avec nos notations, R peut s'écrire :

$$R : \begin{array}{ccc} \mathbb{N}^* & \longrightarrow & \mathcal{F}(\Omega, \mathbb{R}) \\ n & \longmapsto & R_n \end{array},$$

où pour chaque $n \in \mathbb{N}^*$, R_n est une variable aléatoire.

Définition 3.54. On dit qu'une suite R de variables aléatoires est une suite de variables *indépendantes* si pour tout $n \in \mathbb{N}^*$ et tout r_1, \dots, r_n des nombres réels, on a :

$$\mathbb{P}(R_1 = r_1, \dots, R_n = r_n) = \mathbb{P}(R_1 = r_1) \times \dots \times \mathbb{P}(R_n = r_n).$$

Remarque 3.55. Qu'est-ce qui vous assure qu'il existe une telle suite infinie de variables aléatoires indépendantes définies sur un même univers Ω ? Rien du tout, et en fait, c'est même impossible sur un univers fini. Heureusement, il existe un univers Ω infini sur lequel on peut définir une telle suite de variables aléatoires. La théorie des probabilités sur un univers infini étant significativement plus dure que dans le cas fini, nous admettrons ici l'existence d'un tel Ω et le fait que la théorie et les propriétés que vous connaissez se transposent au cas des variables aléatoires sur des univers infinis.

Remarque 3.56. Si R est une suite de variables de Bernoulli indépendantes de même paramètre p , on a donc pour tout $n \in \mathbb{N}^*$,

$$\begin{aligned} \mathbb{E}[R_n] &= p, \\ \mathbb{V}(R_n) &= p(1-p) \end{aligned}$$

De plus, la linéarité de l'espérance et une généralisation de la proposition 3.49 donne :

$$\begin{aligned} \mathbb{E}[R_1 + \dots + R_n] &= \mathbb{E}[R_1] + \dots + \mathbb{E}[R_n] = np, \\ \mathbb{E}[R_1 + \dots + R_n] &= \mathbb{V}(R_1) + \dots + \mathbb{V}(R_n) = np(1-p). \end{aligned}$$

Estimateur. On définit l'estimateur \hat{p}_n pour tout $n \in \mathbb{N}^*$ par l'expression suivante :

$$\hat{p}_n = \frac{R_1 + \cdots + R_n}{n}. \quad (3.1)$$

Le numérateur de la fraction représente bien le nombre de lancers étant tombés sur face après nos n premiers tirages puisque chaque R_i vaut 1 précisément lorsque la pièce tombe sur face.

On peut alors calculer l'espérance et la variance de notre estimateur par linéarité de l'espérance, l'indépendance des variables et la proposition 3.46 :

$$\begin{aligned} \mathbb{E}[\hat{p}_n] &= \frac{np}{n} = p \\ \mathbb{V}(\hat{p}_n) &= \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}. \end{aligned}$$

Estimation de la proportion. Grâce à l'inégalité de Bienaymé-Tchebychev, on peut alors calculer que pour tout $n \in \mathbb{N}^*$ et tout $\varepsilon > 0$, on a :

$$1 - \frac{p(1-p)}{n\varepsilon^2} \leq \mathbb{P}(p - \varepsilon < \hat{p}_n < p + \varepsilon) \leq 1. \quad (3.2)$$

On peut interpréter ce résultat ainsi. Le nombre $\varepsilon > 0$ représente l'erreur que l'on s'autorise sur le résultat calculé par notre estimateur \hat{p}_n et n est le nombre de lancers qu'il faut pour arriver à satisfaire cette marge d'erreur avec probabilité au moins $1 - \frac{p(1-p)}{n\varepsilon^2}$. Ainsi, si l'on choisit par exemple une marge de $\varepsilon = 0,1$ (donc une précision à plus ou moins 1 dixième), pour qu'il y ait 95% de chances que l'estimation empirique \hat{p}_n vaille p à plus ou moins 1 dixième, en supposant que $p = 0,43$, il faut que $n \geq 491$, i.e. que l'on lance au moins 491 fois la pièce.

Convergence. Nous aimerions bien pouvoir montrer un résultat du type :

$$\hat{p}_n \xrightarrow[n \rightarrow +\infty]{} p,$$

qui exprimerait que les proportions calculées sur un nombre n de tirages tendent vers la probabilité réelle p lorsque n tend vers $+\infty$. Cela nous assurerait que la méthode consistant à lancer de plus en plus de fois la pièce est bien fondée, puisqu'elle converge vers la bonne valeur. Cependant, la suite dont nous souhaitons exprimer la limite n'est pas une suite de nombres, mais une suite de variables aléatoires... nous n'avons pas défini de notion de *convergence* pour de tels objets.

Le théorème suivant exprime en quel sens on peut considérer que l'estimateur tend vers la probabilité théorique lorsque le nombre de tirages augmente, il exprime donc que notre estimateur estime bien la quantité p inconnue que nous cherchions.

Théorème 3.57 (Loi faible des grands nombres). *Soit R une suite de variables de Bernoulli indépendantes de même paramètre p et soit \hat{p} défini comme en (3.1). Alors, pour tout $\varepsilon > 0$,*

$$\mathbb{P}(p - \varepsilon < \hat{p}_n < p + \varepsilon) \xrightarrow[n \rightarrow +\infty]{} 1.$$

Intuition 3.58. Le sens que l'on donne à la convergence de l'estimateur est donc le suivant : on considère que \hat{p}_n converge vers p au sens où la probabilité pour que \hat{p}_n soit à distance plus petite que ε de p tend vers 1 quand n tend vers l'infini, et ce pour n'importe quel choix de distance ε .

Vocabulaire 3.59. On dit que \hat{p} est un estimateur *consistant*.

Démonstration. Pour tout $\varepsilon > 0$, on a déjà montré que :

$$1 - \frac{p(1-p)}{n\varepsilon^2} \leq \mathbb{P}(p - \varepsilon < \hat{p}_n < p + \varepsilon) \leq 1,$$

c'est l'équation (3.2). Or, par le théorème d'opération sur les limites,

$$\frac{p(1-p)}{n\varepsilon^2} \xrightarrow[n \rightarrow +\infty]{} 0,$$

et donc

$$1 - \frac{p(1-p)}{n\varepsilon^2} \xrightarrow[n \rightarrow +\infty]{} 1.$$

Par le théorème d'encadrement, on a enfin :

$$\mathbb{P}(p - \varepsilon < \hat{p}_n < p + \varepsilon) \xrightarrow[n \rightarrow +\infty]{} 1.$$

□