

# DES NOMBRES

## À la fin de ce chapitre, je sais :

- ✎ distinguer les ensembles de nombres  $\mathbb{R}$ ,  $\mathbb{Q}$ ,  $\mathbb{D}$ ,  $\mathbb{Z}$  et  $\mathbb{N}$
- ✎ convertir un entier de la base 10 à la base 2 ou 16 et réciproquement
- ✎ expliquer pourquoi certains nombres ne peuvent pas être encodé en machine de manière exacte

## A Des cailloux à compter

Au commencement, il y a les entiers naturels, ces entiers qu'on manipule tous les jours, sans s'en rendre compte :  $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ . Il a fallu longtemps pour arriver à les abstraire comme nous le faisons aujourd'hui. Pour nos ancêtres, les entiers n'étaient souvent que des petits cailloux<sup>1</sup> que l'on manipulait pour compter des moutons, des mesures de blé ou des personnes.

C'est la théorie des ensembles qui nous a permis d'abstraire ces nombres et de les manipuler comme des ensembles. Cette théorie est née à la fin du XIX<sup>e</sup> siècle de l'audace de Cantor, l'audace d'avoir osé compter le nombre d'éléments d'un ensemble, même quand celui-ci était infini [milinowski\_uber\_1874].

■ **Définition 1 — Ensemble.** Un ensemble est une collection de choses qu'on appelle éléments. L'ensemble vide est noté  $\emptyset$ .

## B Cardinal d'un ensemble

■ **Définition 2 — Cardinal d'un ensemble fini.** Le cardinal d'un ensemble fini est son nombre d'éléments.

■ **Exemple 1** Soit l'ensemble  $A = \{\heartsuit, \diamondsuit, \spadesuit, \clubsuit, \star\}$ . On a  $|A| = 5$ .  
On a évidemment également  $|\emptyset| = 0$ .

S'il est facile de dire que le cardinal d'un ensemble fini est le nombre de ses éléments, la définition de cardinal d'un ensemble infini est plus délicate. D'ailleurs, plutôt que de le définir

1. C'est le mot caillou en latin *calculus* qui a donné le mot calcul en français.

directement, on le décrit plutôt [**devoldere\_cardinal\_2000**] en se donnant les moyens de :

1. dire si deux ensembles  $E$  et  $F$  ont le même cardinal (cf. définition 3). Il existe alors une bijection<sup>2</sup> entre les deux : on dit qu'ils sont équipotents.
2. comparer les cardinaux de deux ensembles  $E$  et  $F$ . S'il existe une injection de  $E$  dans  $F$ , alors le cardinal de  $E$  est inférieur ou égal au cardinal de  $F$ .

On peut définir également une relation d'ordre sur les cardinaux et ainsi tous les comparer.

■ **Définition 3 — Cardinal d'un ensemble.** Si deux ensembles peuvent être mis en bijection, on dit qu'ils ont le même cardinal, qu'ils sont équipotents.

## C Peut-on construire l'ensemble $\mathbb{N}$ ?

Les entiers jouent un rôle essentiel dans le cadre de la théorie de l'information en général et davantage encore dans le domaine de la cryptographie. La construction de l'ensemble  $\mathbb{N}$  peut être établie rigoureusement et simplement dans le cadre de la théorie des ensembles en utilisant l'ensemble vide<sup>3</sup> et trois axiomes : l'axiome de la paire, l'axiome de la réunion et l'axiome de l'infini. C'est la méthode de construction des ensembles dite de Von Neumann.

Pour construire l'ensemble des entiers naturels  $\mathbb{N}$  on peut :

1. choisir de noter 0 l'ensemble vide :  $0 = \emptyset$ ,
2. définir une fonction  $s$  *successeur de* en posant pour tout ensemble  $a$  :  $s(a) = a \cup \{a\}$ . Le successeur est donc obtenu en ajoutant à l'ensemble l'ensemble de départ comme élément. Si  $|a| = n$ , alors on voit immédiatement que  $|s(a)| = n + 1$

Le successeur de 0 s'écrit  $s(0) = 0 \cup \{0\} = \emptyset \cup \{\emptyset\} = \{\emptyset\}$ , ensemble que l'on peut noter 1 et dont le cardinal vaut 1. On remarque aussi que  $s(1) = 1 \cup \{1\} = \{\emptyset\} \cup \{\{\emptyset\}\} = \{\emptyset, \{\emptyset\}\} = \{0, 1\}$ , ensemble que l'on peut noter 2, dont le cardinal vaut bien 2. Ainsi de suite,  $n = \{0, 1, \dots, n-1\}$ , récursivement. Selon cette approche, on peut donc définir chaque nombre entier comme un ensemble.

Afin de garantir l'existence de l'ensemble  $\mathbb{N}$ , on a besoin de l'axiome de l'infini qui garantit qu'il existe un ensemble contenant 0 fermé pour l'opération successeur, c'est à dire que tout successeur d'un élément de  $I$  appartient à  $I$ . Grâce à cet axiome, l'ensemble des entiers naturels  $\mathbb{N}$  est un ensemble infini d'ensembles dont les cardinaux valent les nombres entiers.

## D Ensembles usuels

Dans ce document, on supposera donc que l'on sait construire les ensembles :

1.  $\mathbb{N} = \{0, 1, 2, 3 \dots\}$  les entiers naturels,
2.  $\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3 \dots\}$  les entiers relatifs,
3.  $\mathbb{Q}$  l'ensemble des nombres rationnels,
4.  $\mathbb{R}$  l'ensemble des nombres réels.

2. Il est intéressant de noter que Galilée déjà avait eu l'idée de faire un appariement bijectif.

3. dont le cardinal vaut 0.

On supposera également que  $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$  et que les opérations d'addition et de multiplication sont possibles sur tous ces ensembles. Enfin, on suppose pour l'instant que l'on peut comparer les nombres réels entre eux.

Ces ensembles sont représentés sur la figure 1. L'ensemble  $\mathbb{A}$  est l'ensemble des solutions des équations polynômiales à coefficients rationnels, c'est à dire l'ensemble des racines des polynômes de  $\mathbb{Q}[X]$ , par exemple  $\sqrt{2}$  qui est racine de  $X^2 - 2$ . L'ensemble  $\mathbb{D}$  est l'ensemble des nombres décimaux (cf. définition 6).

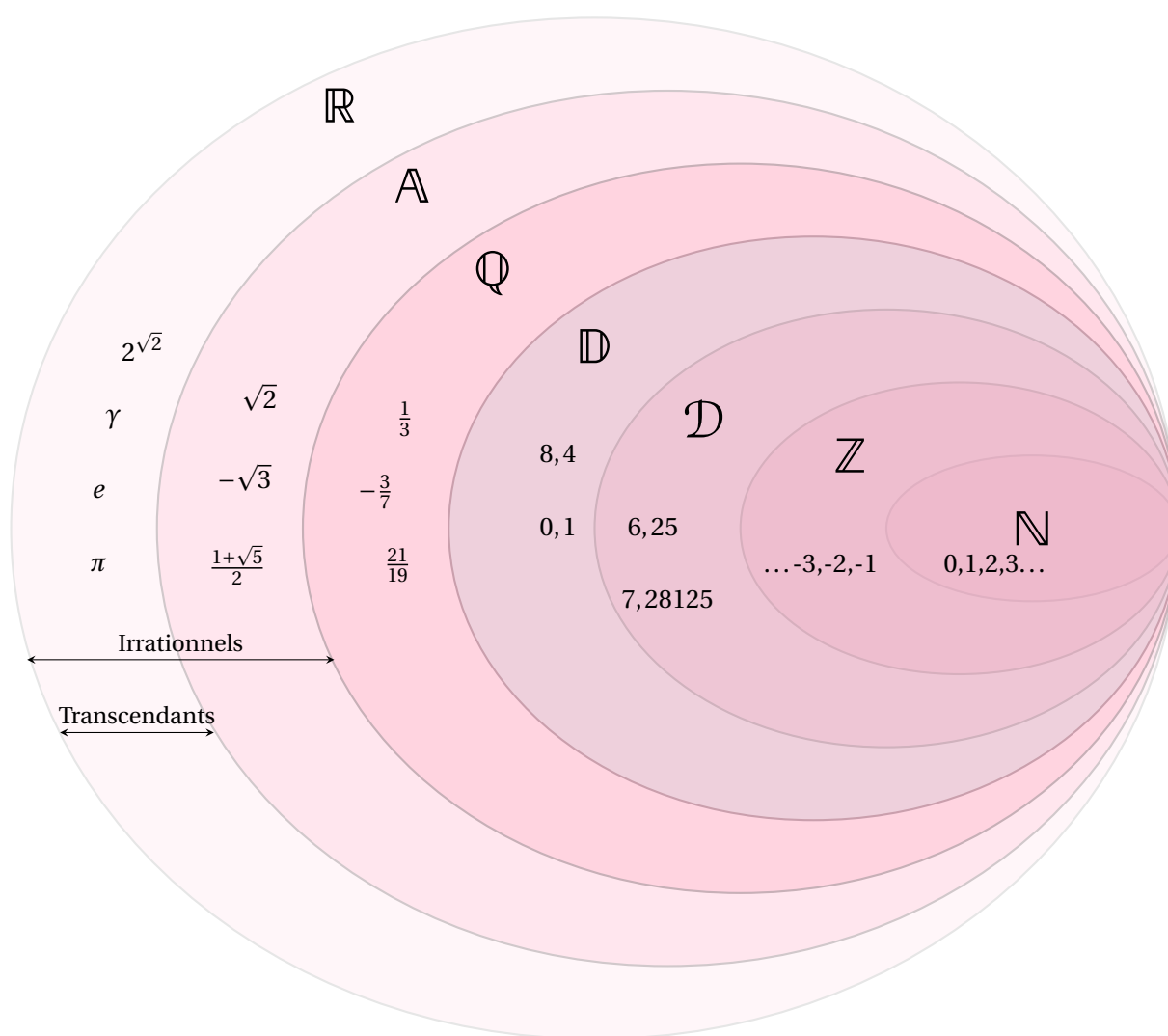


FIGURE 1 – Ensembles de nombres : naturels  $\mathbb{N}$ , relatifs  $\mathbb{Z}$ , dyadiques rationnels  $\mathbb{D}$ , décimaux  $\mathbb{D}$ , rationnels  $\mathbb{Q}$ , algébriques  $\mathbb{A}$  et réels  $\mathbb{R}$ .

## E Formalisation de la numération de position

■ **Définition 4 — Numération de position.** La numération de position est un principe de notation selon lequel la signification d'un chiffre dépend de sa position dans le nombre. Dans un tel système, chaque chiffre se voit affecter un «poids» dans un nombre, poids qui est un facteur multiplicatif et qui dépend de la position du chiffre dans ce nombre.

■ **Définition 5 — Base d'un système de numération de position.** Soit  $g$  un entier naturel fixé supérieur à 1.  $g$  est la base d'un système de numération de position, si, lors de l'écriture d'un nombre, une unité de chaque ordre vaut  $g$  unités de l'ordre précédent.

Même si le principe énoncé dans la définition 4 peut apparaître trivial, l'écriture des nombres n'a pas toujours suivi ce principe<sup>4</sup>. Aujourd'hui, nous utilisons la numération de position quelle que soit la base et les conventions prises sont la plupart du temps les suivantes :

- les positions des chiffres se comptent de droite à gauche,
- le chiffre le plus à droite représente les unités et possède l'indice 0,
- le nombre formé par 10 représente exactement la valeur de la base quelle que soit cette base, c'est-à-dire la base d'un système de numération se note toujours 10 dans cette base,
- si on ajoute un zéro à droite d'un nombre, cela revient à multiplier ce nombre par sa base.

Dans un nombre écrit avec le système de numération de position, un chiffre  $c$  en position  $p$  ne vaut pas la même chose qu'à la position  $p - 1$ . Si  $g$  est la base de ce système, le chiffre  $c$  en position  $p$  vaut  $g$  fois plus que s'il était placé en position  $p - 1$ .

La division euclidienne est une décomposition unique d'un nombre. C'est pourquoi, on montre dans la section suivante qu'un entier naturel  $a$  peut s'écrire d'une manière unique dans un système de numération de position en base  $g$  à l'aide d'un nombre minimal de chiffres  $n$ . Les coefficients  $a_i \in \{0, \dots, g - 1\}$  sont des entiers naturels strictement inférieurs à  $g$  (cf. théorème 1). On a alors :

$$a = a_{n-1}g^{n-1} + \dots + a_2g^2 + a_1g + a_0 = \sum_{k=0}^{n-1} a_k g^k. \quad (1)$$

La numération de position revient à représenter le nombre en écrivant seulement les coefficients de ce polynôme, en omettant la plupart du temps la base et en notant tous les coefficients nuls ou non, de manière à ce que leur place soit définie sans ambiguïté. On désigne alors un nombre  $a$ , qui possède  $n$  chiffres, par un  $n$ -uplet en séparant ou non les éléments du  $n$ -uplet :

$$a = (a_{n-1}, \dots, a_1, a_0) = (a_{n-1} \dots a_1 a_0)_g = a_{n-1} \dots a_1 a_0 \quad (2)$$

■ **Exemple 2**  $2021_{10} = 2 \times 10^3 + 0 \times 10^2 + 2 \times 10^1 + 1 \times 10^0 = 2021$

4. Le chiffres romains par exemple.

■ **Exemple 3**  $2021_3 = 2 \times 3^3 + 0 \times 3^2 + 2 \times 3^1 + 1 \times 3^0 = 61_{10}$  et ne se prononce pas «deux mille vingt et un»...

■ **Exemple 4**  $2021_{64} = 2 \times 64^3 + 0 \times 64^2 + 2 \times 64^1 + 1 \times 64^0 = 524417_{10}$

## F Écriture d'un entier dans base quelconque

**Théorème 1 — Décomposition d'un entier en base  $g$ .** Soit  $g \in \mathbb{N} \setminus \{0, 1\}$ . Pour tout  $a \in \mathbb{N}$ , il existe  $n \in \mathbb{N}$  et un  $n$ -uplet unique de chiffres  $(a_0, a_1, \dots, a_{n-1}) \in \llbracket 0, g-1 \rrbracket^n$  tels que :

$$a = \sum_{k=0}^{n-1} a_k g^k. \quad (3)$$

De plus, si  $a \in \mathbb{N}^*$ , on peut calculer le nombre de chiffres nécessaires pour représenter un nombre dans la base  $g$  :

$$n \leq \left\lfloor \log_g a \right\rfloor + 1 \quad (4)$$

*Démonstration.* 1. Unicité : on suppose qu'un développement tel que 3 existe. Alors, on peut écrire, en regroupant les puissances non nulles de  $g$  :

$$a = g A_1 + a_0$$

avec

$$A_1 = a_{n-1} g^{n-2} + \dots + a_1$$

Ainsi,  $a_0$  peut être vu comme le reste de la division euclidienne de  $a$  par  $g$ . Celle-ci étant unique, les coefficients  $a_0$  et  $A_1$  sont bien déterminés de manière unique. En considérant les termes  $A_k = a_{n-k} g^{n-k-1} + \dots + a_k$ , on trouve de même que  $a_k$  et  $A_{k+1}$  sont les restes et quotients de la division euclidienne de  $A_k$  par  $a$ . Par une récurrence immédiate, on montre ainsi que les  $a_k$  sont uniques.

2. Existence : on procède en construisant la solution, c'est à dire les coefficients entiers. On les choisit comme suit :

$$c_k = \left\lfloor \frac{a}{g^k} \right\rfloor - g \left\lfloor \frac{a}{g^{k+1}} \right\rfloor$$

D'après la définition de la partie entière,

$$\frac{a}{g^k} - 1 < \left\lfloor \frac{a}{g^k} \right\rfloor \leq \frac{a}{g^k}$$

et

$$-\frac{a}{g^{k+1}} \leq -\left\lfloor \frac{a}{g^{k+1}} \right\rfloor < -\frac{a}{g^{k+1}} + 1$$

En multipliant la dernière inéquation par  $g$  et en l'additionnant première, on obtient que :

$$-1 < c_k < g$$

Comme  $c_k$  est un entier, il appartient donc à l'ensemble  $\{0, \dots, g-1\}$ . Ces coefficients sont nuls à partir d'un certain rang. En effet, si  $k \geq \lfloor \log_g a \rfloor + 1$ , alors  $k > \lfloor \log_g a \rfloor$  et

$$\frac{a}{g^k} < \frac{a}{g^{\lfloor \log_g a \rfloor}} = \frac{a}{a} = 1$$

Ceci signifie que  $\lfloor \frac{a}{g^k} \rfloor = 0$  quelque soit  $k > \lfloor \log_g a \rfloor$ .

Enfin, en notant  $m = 1 + \lfloor \log_g a \rfloor$  on obtient par télescopage :

$$\begin{aligned} \sum_{k=0}^{m-1} c_k g^k &= \left\lfloor \frac{a}{g^0} \right\rfloor - g \left\lfloor \frac{a}{g^1} \right\rfloor + \left( \left\lfloor \frac{a}{g^1} \right\rfloor - g \left\lfloor \frac{a}{g^2} \right\rfloor \right) g \\ &\quad + \left( \left\lfloor \frac{a}{g^2} \right\rfloor - g \left\lfloor \frac{a}{g^3} \right\rfloor \right) g^2 + \left( \left\lfloor \frac{a}{g^3} \right\rfloor - g \left\lfloor \frac{a}{g^4} \right\rfloor \right) g^3 \\ &\quad + \dots \\ &\quad + \left( \left\lfloor \frac{a}{g^{m-2}} \right\rfloor - g \left\lfloor \frac{a}{g^{m-1}} \right\rfloor \right) g^{m-2} + \left( \left\lfloor \frac{a}{g^{m-1}} \right\rfloor - g \left\lfloor \frac{a}{g^m} \right\rfloor \right) g^{m-1} \\ &= \left\lfloor \frac{a}{g^0} \right\rfloor - \left\lfloor \frac{a}{g^m} \right\rfloor g^m \\ &= a \end{aligned}$$

car  $m > \lfloor \log_g a \rfloor$  et  $\lfloor \frac{a}{g^m} \rfloor = 0$ . Les coefficients  $c_k$  conviennent donc bien pour les  $a_k$  du théorème. ■

**(R)** On peut définir un système de numération unaire, c'est à dire avec les seuls symboles 0 et 1, mais ce n'est pas l'objet ici. C'est pourquoi on a restreint  $g \in \mathbb{N} \setminus \{0, 1\}$ .

**Théorème 2** Si  $a \in \mathbb{N}$  s'écrit  $a_{n-1} \dots a_0$  dans la base  $g$ , alors  $g^{n-1} \leq a < g^n$ .

*Démonstration.* On le démontre par récurrence. Pour  $n = 1$ ,  $a = a_0 < g$  d'après le théorème 1. Supposons la proposition vraie pour les nombres à  $n$  chiffres et prenons un nombre  $c$  à  $n+1$  chiffres. Alors, on peut écrire  $c = c_n g^n + d$ , c'est à dire le nième chiffre multiplié par son poids dans la base plus un nombre  $d = d_{n-1} \dots d_0$  à  $n$  chiffres en base  $g$ . On applique l'hypothèse de récurrence au nombre  $d$ . Avec l'inégalité de droite, on obtient :

$$c < c_n g^n + g^n = g^n (c_n + 1) \leq g^{n+1}$$

puisque  $c_n \in \{0, \dots, g-1\}$ .

Avec l'inégalité de gauche, on obtient :

$$c \geq c_n g^n + g^{n-1} \geq c_n g^n \geq g^n$$

Donc,

$$g^n \leq c < g^{n+1}$$

Donc la proposition est vraie pour tout entier de  $n$  chiffres. ■

■ **Exemple 5 — en base 2.**  $1111_2 < 2^4$

■ **Exemple 6 — en base 10.**  $999 < 10^3$

**Théorème 3 — Moins on a de chiffres, plus on est petit .** Soient deux entiers écrits dans une même base  $g$  et dont le nombre de chiffres est différent, alors le plus petit est celui dont l'écriture possède le moins de chiffres.

*Démonstration.* Soient  $a$  et  $b$  deux entiers écrits respectivement  $a_{n-1} \dots a_0$  et  $b_{m-1} \dots b_0$  dans la base  $g$  et que  $a \neq b$ . Supposons que  $n < m$  alors  $n \leq m-1$  et puisque  $1 < g$ , la proposition 2 implique que  $a < g^n \leq g^{m-1} \leq b$ . Donc  $a < b$ . ■

## G Changement de base

Pour convertir en base 10 un nombre entier quelconque, il suffit d'appliquer la formule 1 comme dans l'exemple 3.

Pour faire l'opération inverse, c'est à dire convertir un nombre entier en base 10 vers une base quelconque, il faut remarquer que si  $a = (a_{n-1} \dots a_1 a_0)_g$  alors le quotient de la division euclidienne de  $a$  par  $g$  est égal à  $q = (a_{n-1} \dots a_1)_g$  et le reste à  $r = a_0$  puisque  $a = gq + r$  avec  $0 \leq r \leq g-1$ .

■ **Exemple 7 — Écrire  $61_{10}$  en base 3.**

$$61_{10} = 3 \times 20 + 1 \quad (5)$$

$$20 = 3 \times 6 + 2 \quad (6)$$

$$6 = 3 \times 2 + 0 \quad (7)$$

$$2 = 3 \times 0 + 2 \quad (8)$$

C'est pourquoi,  $61_{10} = 2021_3 = 2 \times 3^3 + 0 \times 3^2 + 2 \times 3^1 + 1 \times 3^0$

■ **Exemple 8 — Écrire  $61_{10}$  en base 2.**

$$61_{10} = 32 + 16 + 8 + 4 + 1 \quad (9)$$

$$= 1 \times 2^5 + 1 \times 2^4 + 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \quad (10)$$

$$= 111101_2 \quad (11)$$

$$= 0b111101 \quad (12)$$

■ Exemple 9 — Écrire  $61_{10}$  en base 16.

$$61_{10} = 3 \times 16 + D \quad (13)$$

$$= 3 \times 16^1 + D \times 16^0 \quad (14)$$

$$= 3D_{16} \quad (15)$$

$$= 0x3D \quad (16)$$

■ Exemple 10 — Écrire  $3CF_{16}$  en base 10.

$$3CF_{16} = 3 \times 16^2 + C \times 16^1 + F \quad (17)$$

$$= 3 \times 256 + C \times 160 + F \quad (18)$$

$$= 975_{10} \quad (19)$$

■ Exemple 11 — Écrire  $10110101_2$  en base 10.

$$10110101_2 = 1 \times 2^7 + 1 \times 2^5 + 1 \times 2^4 + 1 \times 2^2 + 1 \times 2^0 \quad (20)$$

$$= 128 + 32 + 16 + 4 + 1 \quad (21)$$

$$= 181_{10} \quad (22)$$

**(R)** En langage Python, on peut directement écrire du binaire en préfixant le nombre par **0b** ( `0b00011` ) ou de l'hexadécimal en préfixant par **0x** ( `0xF4E` ). En machine, un nombre est toujours codé en binaire, ces représentations nous sont donc destinées à nous, êtres humains. Les fonctions `bin`, `oct` et `hex` permettent de convertir directement des nombres en binaire, octal et hexadécimal. Inversement, l'instruction `int('2021', 3)` permet de convertir de la base 3 en décimal (on trouve 61).

**(R)** L'hexadécimal présente un intérêt fort car la conversion entre le binaire et l'hexadécimal est simple pour un être humain : chaque groupe de quatre bits représente un chiffre hexadécimal.

■ Exemple 12 — Écrire  $10110101_2$  en base 16.

$$10110101_2 = 1011_2 \times 16^1 + 0101_2 \times 16^0 \quad (23)$$

$$= B_{16} \times 16^1 + 5_{16} \times 16^0 \quad (24)$$

$$= B5_{16} \quad (25)$$

$$= 0xB5 \quad (26)$$



## H Nombres décimaux et dyadiques

■ **Définition 6 — Nombre décimal.** Un nombre décimal est un rationnel qui peut s'écrire sous la forme

$$\frac{a}{10^n}, a \in \mathbb{Z}, n \in \mathbb{N} \quad (27)$$

On note  $\mathbb{D} = \{\frac{a}{10^n}, a \in \mathbb{Z}, n \in \mathbb{N}\}$  l'ensemble des nombres décimaux.

■ **Exemple 13 — 8,4 est un nombre décimal.** En effet, on peut l'écrire  $\frac{84}{10}$ .

■ **Définition 7 — Nombre dyadique.** Un nombre dyadique est un rationnel qui peut s'écrire sous la forme

$$\frac{a}{2^n}, a \in \mathbb{Z}, n \in \mathbb{N} \quad (28)$$

On note  $\mathcal{D} = \{\frac{a}{2^n}, a \in \mathbb{Z}, n \in \mathbb{N}\}$  l'ensemble des nombres dyadiques.

■ **Exemple 14 — 6,25 est un nombre dyadique.** En effet, on observe que  $6,25_{10} = 110,01_2$ . Son développement binaire est fini. On peut l'écrire  $\frac{25}{2^2}$ .

**Théorème 4 — Caractérisation des dyadiques et des décimaux.** Un nombre est décimal (resp. dyadique) si son développement en base 10 (resp. 2) est fini.

■ **Exemple 15 —  $1/3$  n'est pas un nombre décimal.** En effet, son développement en base dix n'est pas fini, il se répète. On peut l'écrire  $\frac{1}{3} = 0,333333\dots$ . C'est un nombre rationnel.

■ **Exemple 16 — 8,4 n'est pas un nombre dyadique.** En effet, on observe que  $8,4_{10} = 1000,011001100110011_2\dots$ . Son développement binaire n'est pas fini.

**Théorème 5 — Les dyadiques sont strictement inclus dans les décimaux.** On a  $\mathcal{D} \subsetneq \mathbb{D}$ . Ce qui signifie qu'il existe des décimaux qui ne sont pas des dyadiques.

*Démonstration.* On montre l'inclusion puis la stricte inclusion.

$\mathcal{D} \subset \mathbb{D}$  Soit un nombre dyadique  $\frac{a}{2^n}$ ,  $a \in \mathbb{Z}$ . Alors en multipliant en haut et en bas par  $5^n$  on obtient

$$\frac{a}{2^n} = \frac{a \times 5^n}{10^n}$$

ce qui prouve que c'est un nombre décimal puisque  $a \times 5^n \in \mathbb{Z}$ .

$\mathcal{D} \neq \mathbb{D}$  le développement binaire de 0,1 est infini. Ce nombre est décimal mais pas dyadique, ce qui montre l'inclusion stricte.

■

■ **Exemple 17 — Décimaux mais pas dyadiques.** 0,1, 0,2, 0,3 ou 8,4 sont des nombres décimaux mais pas dyadiques.

Ⓡ On a toujours  $\mathcal{D} \subset \mathbb{Q}$  et  $\mathbb{D} \subset \mathbb{Q}$  (cf. figure 1).

## I De l'écriture des nombres rationnels

On peut montrer qu'un nombre rationnel est un nombre :

- à représentation décimale répétitive. Par exemple,  $1/3 = 0,333\dots = 0.\overline{3}$  ou  $1/7 = 0,142857142857\dots = 0,\overline{142857}$ ,
- ou à représentation décimale terminale, lorsque la répétition est 0. Par exemple,  $1/2 = 0,5000\dots = 0,4999\dots = 0,5$ .

Il existe, en base 10, deux représentations équivalentes décimales terminales : soit avec des 0 soit avec des 9 (cf. proposition 6). C'est pourquoi, il faut être attentif à l'interprétation des résultats lorsque ces différentes représentations interviennent.

**Théorème 6 —**  $1 = 0,9999\dots$

*Démonstration.* On utilise le résultat de la somme des termes d'une suite géométrique.

$$0,999\dots = \frac{9}{10} + \frac{9}{100} + \frac{9}{1000} + \dots \quad (29)$$

$$= \frac{9}{10} + \frac{9}{10^2} + \frac{9}{10^3} + \dots \quad (30)$$

$$= 9 \left( \frac{1}{10} + \frac{1}{10^2} + \frac{1}{10^3} + \dots \right) \quad (31)$$

$$= 9 \left( -1 + \sum_{k=0}^{+\infty} \frac{1}{10^k} \right) \quad (32)$$

$$= 9 \left( -1 + \frac{1}{1 - \frac{1}{10}} \right) \quad (33)$$

$$= 9 \left( -1 + \frac{10}{9} \right) \quad (34)$$

$$= 9 \cdot \frac{1}{9} = 1 \quad (35)$$

■

## J Arrondis et erreurs

Il nous faut pouvoir estimer précisément les erreurs que l'on commet lorsqu'on choisit d'encoder en machine un nombre selon un format donné. Les définitions de l'arrondi, des erreurs absolues et relatives vont nous y aider.

■ **Définition 8 — Arrondir un nombre en base 10.** La plupart du temps, on choisit d'arrondir un nombre réel  $a$  à  $10^{-n}$  de la manière suivante :

$$\text{arrondi}(a, n) = \text{sgn}(a) \frac{\lfloor |a \times 10^n| + 0,5 \rfloor}{10^n} \quad (36)$$

■ **Exemple 18 — Arrondir à  $10^{-2}$ .** 5,456 sera arrondi à  $10^{-2}$  par la formule 36 à 5,46.

■ **Définition 9 — Erreur d'approximation absolue.** Soit  $v$  un nombre réel et  $\tilde{v}$  sa valeur approchée. L'erreur absolue est définie par :

$$\varepsilon_a = \tilde{v} - v \quad (37)$$

■ **Définition 10 — Erreur d'approximation relative.** Soit  $v$  un nombre réel et  $\tilde{v}$  sa valeur approchée. L'erreur relative est définie par :

$$\varepsilon_r = \frac{\tilde{v} - v}{|v|} \quad (38)$$