Chapitre 8

Codes correcteurs d'erreurs

8.1 Introduction : Le sport-toto

Le sport-toto est un jeu (de hasard !) où il faut deviner le score des matchs de football.

Imaginons un sport-toto à 4 matchs. Pour chaque match, on note 1 pour une victoire de l'équipe qui joue à domicile, 2 pour une victoire de l'équipe invitée et x pour un match nul.

On remplit une grille pour les 4 matchs. Par exemple on peut jouer la colonne suivante.

1er match	1
2e match	2
3e match	x
4e match	1

On peut se poser les questions suivantes.

1. Combien de grilles doit-on remplir pour être sûr de gagner (une grille a les 4 bons résultats) ?
2. Combien de grilles doit-on remplir pour être sûr d'avoir 3 matchs sur 4 avec les bons résultats ?

Les réponses sont les suivantes.

1. Cette réponse est facile, il y a [ ] possibilités et une seule combinaison est gagnante. Il faut donc jouer [ ] grilles.
2. Ici, c'est plus subtil, mais possible. Pour cela, on peut jouer les 9 colonnes suivantes.

x	x	x	1	1	1	2	2	2
x	1	2	x	1	2	x	1	2
x	1	2	1	2	x	2	x	1
x	1	2	2	x	1	1	2	x

En effet, pour chaque grille parmi toutes celles possibles (que l'on supposera être le résultat des 4 matchs), si on regarde combien de points on réalise avec chacune des 9 colonnes ci-dessus (un point pour un match juste), on verra que l'on a toujours une colonne ci-dessus qui livre 3 ou 4 points. C'est bien sûr une démonstration longue et ennuyeuse. Il y a une démonstration plus rapide.

Preuve

Créons un peu de vocabulaire. Disons que la *sphère d'influence centrée en une grille* est l'ensemble des colonnes qui ont au plus une différence avec cette grille. Il est facile de constater que chaque sphère d'influence a exactement 9 colonnes (la colonne au centre et les 8 colonnes qui ont exactement 1 différence (4 endroits et 2 possibilités)).

Regardons uniquement les sphères d'influence des 9 colonnes qui nous intéressent. Ces neufs colonnes ont la propriété essentielle suivante.

Il y a toujours 3 différences entre deux colonnes choisies parmi les 9.

Cela signifie que les 9 sphères d'influence sont disjointes.

On a donc 9 sphères d'influence disjointes contenant chacune 9 éléments. Ainsi ces 9 sphères contiennent au total $9 \cdot 9 = 81$ grilles. Il s'agit de toutes les grilles possibles.

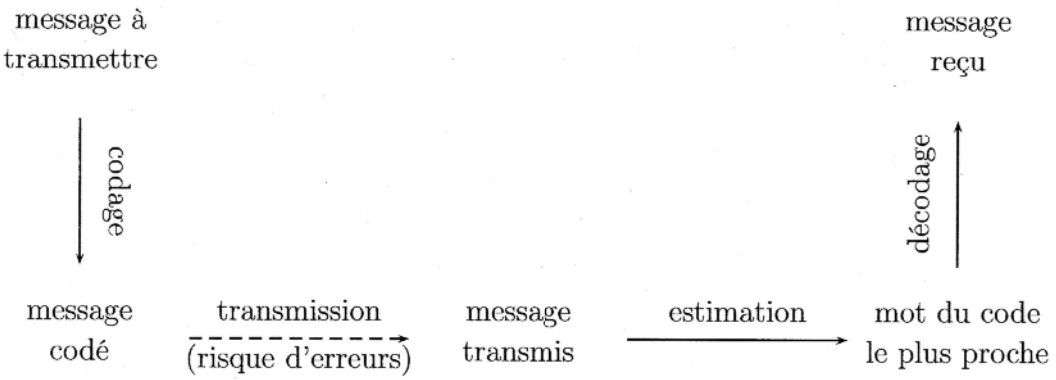
Ainsi, chaque grille (parmi les 81 possibles) se trouve à l'intérieur d'une seule sphère d'influence centrée en une grille parmi les 9 colonnes ci-dessus. La colonne correspondante est celle qui donne 3 ou 4 points. □

Bien évidemment les créateurs sont maintenant au courant de cette astuce et proposent des sport-toto à plus de 13 matchs (il y a aussi une technique similaire permettant d'assurer 12 points sur 13 matchs, mais elle coûte plus cher qu'elle ne rapporte).

8.2 Codes correcteurs d'erreurs

Contrairement aux codes en cryptographie (qui consistent à camoufler un message), les codes correcteurs d'erreurs ont été inventés pour pouvoir détecter et éventuellement corriger des erreurs qui s'y seraient glissées (de manière accidentelle). Voici une méthode bien connue des spécialistes du radar.

Voici le schéma à avoir en tête lorsqu'on pense aux codes correcteurs.



8.2.1 La méthode des spécialistes du radar

Si lors d'une transmission horizontale, une vingtaine de caractères consécutifs sont perdus, il peut être très difficile de les retrouver !

```
Souvent, pour s'amuser, les hommes d'équipage
Prennent des albatros, vastes oiseaux des mers,
Qui suivent, indolents compagnons de voyage,
Le navire glissant sur les gouffres amers.
A peine les ont-ils déposés sur les planches,
Que ces rois de l'azur, maladroits et honteux,
********** leurs grandes ailes blanches
Comme des avirons traîner à coté d'eux.
Ce voyageur ailé, comme il est gauche et veule!
Lui, naguère si beau, qu'il est comique et laid!
L'un agace son bec avec un brûle-gueule,
L'autre mime, en boitant, l'infirme qui volait!
Le Poête est semblable au prince des nuées
Qui hante la tempête et se rit de l'archer;
Exilé sur le sol au milieu des huées,
Ses ailes de géant l'empêchent de marcher.
Charles Baudelaire (Les fleurs du mal)
```

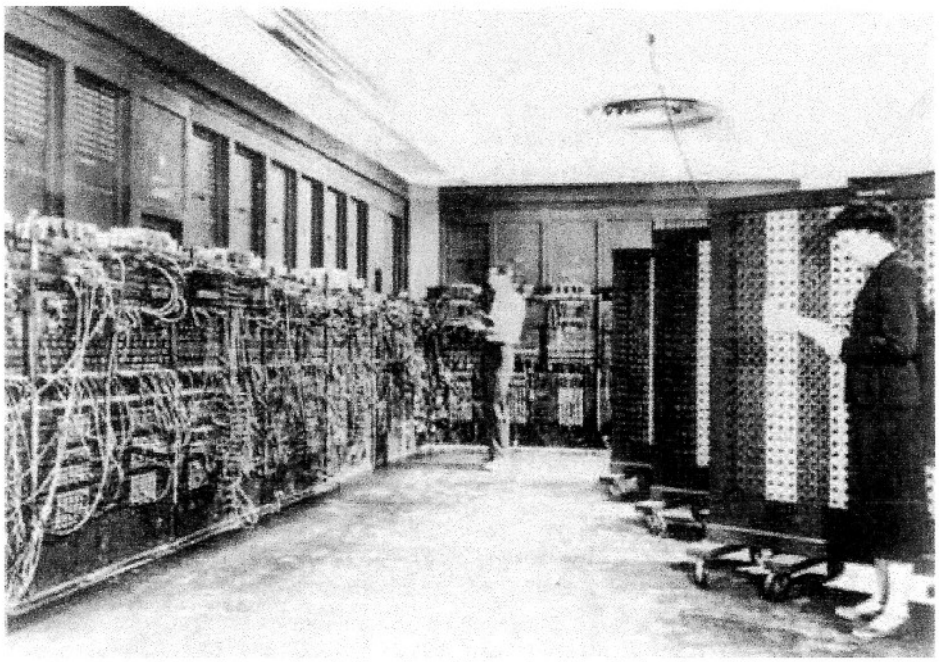
Par contre, si on transmet le texte verticalement, c'est un jeu d'enfant de retrouver vingt caractères consécutifs perdus.

```
Souvent, pour s'amu**r, les hommes d'équipage
Prennent des albatr**, vastes oiseaux des mers,
Qui suivent, indole**s compagnons de voyage,
Le navire glissant *ur les gouffres amers.
A peine les ont-ils*déposés sur les planches,
Que ces rois de l'a*ur, maladroits et honteux,
Laissent piteusemen* leurs grandes ailes blanches
Comme des avirons t*aîner à coté d'eux.
Ce voyageur ailé, c*mme il est gauche et veule!
Lui, naguère si bea*, qu'il est comique et laid!
L'un agace son bec *vec un brûle-gueule,
L'autre mime, en bo*tant, l'infirme qui volait!
Le Poête est sembla*le au prince des nuées
Qui hante la tempêt* et se rit de l'archer;
Exilé sur le sol au*milieu des huées,
Ses ailes de géant *'empêchent de marcher.
Charles Baudelaire *Les fleurs du mal)
```

Malheureusement, cette méthode ne fonctionne pas pour des messages composés de chiffres ou de lettres disposées de manière apparemment aléatoire (penser aux documents numériques : images, sons, musiques, vidéos).

8.3 Le code de Hamming

En 1947, Richard W. Hamming avait accès à un ordinateur de l'armée seulement pendant les week-ends. Les ordinateurs de l'époque étaient très grands (voir photo ci-dessous) et extrêmement lents par rapport à ceux d'aujourd'hui.



Cette photo provient de l'armée américaine et est dans le domaine publique.

L'ordinateur sur lequel Hamming travaillait avait un code détecteur d'erreur, appelé 2-sur-5. On disposait les nombres de 0 à 9 sur des rampes de 5 lampes dont 2 étaient allumées et 3 étaient éteintes.

1	1	1	0	0	0
2	1	0	1	0	0
3	0	1	1	0	0
4	1	0	0	1	0
5	0	1	0	1	0
6	0	0	1	1	0
7	1	0	0	0	1
8	0	1	0	0	1
9	0	0	1	0	1
0	0	0	0	1	1

On voit que toutes les combinaisons possibles de deux lampes allumées sont représentées. Ainsi, si on voit qu'il n'y a pas exactement deux lampes allumées, on sait qu'une erreur s'est produite. Les opérateurs pouvaient retrouver l'erreur (en examinant ce qu'il s'était passé avant), mais ils n'étaient pas présent le week-end et l'ordinateur devait être redémarré (en perdant beaucoup de temps).

Après qu'un calcul a été stoppé de cette manière deux week-ends consécutifs, Hamming était frustré et ennuyé et il s'est demandé pourquoi si l'ordinateur pouvait détecter, il ne pouvait pas trouver sa position et la corriger.

Il inventa ainsi le premier code correcteur de l'Histoire en 1947.

Il s'est placé dans l'anneau $\mathbb{Z}_2 = \{0, 1\}$ avec les règles d'addition suivantes.

$$0+1=1=1+0 \qquad 0+0=0 \qquad 1+1=0$$

Si on écrit les nombres de 0 à 9 en base 2, on a besoin de 4 lampes.

	0	1	2	3	4	5	6	7	8	9
1	0	1	0	1	0	1	0	1	0	1
2	0	0	1	1	0	0	1	1	0	0
4	0	0	0	0	1	1	1	1	0	0
8	0	0	0	0	0	0	0	0	1	1

Et il eu l'idée d'écrire le tableau suivant.

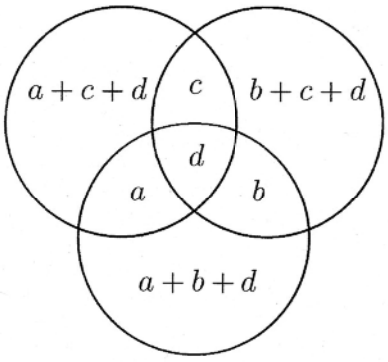
a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	$a+b+c+d$

Cela donnait une application de codage où chaque chiffre était représenté par un mot (a, b, c, d) avec a, b, c et $d \in \mathbb{Z}_2$. On y associait le mot codé

$$(a,\ b,\ c,\ d,\ \underbrace{a+b,\ c+d,\ a+c,\ b+d,\ a+b+c+d}_{\text{caractères de contrôle}})$$

Si un des neufs éléments de ce mot est changé (un 1 en un 0 ou l'inverse), alors on peut dire qu'il y a une erreur et on peut même situer où elle se trouve. Ainsi faisant, on s'aperçoit (comme Hamming) qu'on peut se passer du caractère de contrôle $a+b+c+d$. Ainsi, on a besoin de 8 lampes pour corriger une erreur (4 pour le chiffre auxquelles on en ajoute 4 pour les caractères de contrôle).

Mais Hamming a réussi à faire encore mieux, grâce à l'idée suivante (Leonhard Euler a eu l'idée d'utiliser les diagrammes de Venn).



On a donc l'application de codage suivante (ordre alphabétique).

$$(a, b, c, d) \mapsto (a,\ b,\ c,\ d,\ \underbrace{a+b+d,\ a+c+d,\ b+c+d}_{\text{caractères de contrôle}})$$

Ce code plus astucieux permet de corriger une erreur et de n'utiliser que 7 lampes. Il a été démontré qu'on ne peux pas corriger une erreur avec moins de 7 lampes.

Proposition

S'il existe un code binaire 1-correcteur d'erreur de longueur n systématique sur les r premières positions (cela signifie que sur les r premières positions on retrouve le message à coder : le code de Hamming ci-dessus est de longueur 7 et systématique sur les 4 premières positions). Alors

$$2^n \geqslant (n+1)2^r$$

Preuve

Il y a 2^r mots du code (un par message à coder) et 2^n mots de longueur n (potentiellement recevable après la transmission et ses multiples erreurs possibles).

Si le code est 1-correcteur, cela signifie que les sphères d'influence des mots du code sont disjointes (rappelons que la sphère d'influence d'un mot du code contient tous les mots qui ont au plus une différence par rapport au mode du code).

Comme on parle de code binaire de longueur n, chaque sphère d'influence d'un mot du code contient $(n+1)$ mots (n modifications possibles d'un zéro ou d'un un et le mot du code lui-même).

On a ainsi

$$\overbrace{2^n}^{\text{nombre de mots de longueur } n} \geqslant \underbrace{2^r}_{\text{nombre de sphères d'influences centrées en les mots du code}} \underbrace{(n+1)}_{\text{nombre de mots dans chaque sphère}}$$

$$\underbrace{\qquad\qquad}_{\text{nombre de mots dans toutes les sphères}}$$

On remarque, en bonus, qu'on a l'égalité lorsque tout mot de longueur n se trouve dans une unique sphère d'influence centrée en un mot du code. □

Cas d'égalité

L'égalité de l'équation de la proposition se produit pour

1. $n = 3$ et $r = 1$.

   Il s'agit du code 1-correcteur élémentaire donné par l'application de codage

   $$(a) \mapsto (a, a, a)$$

   En effet, si on triple l'information, on a un code qui corrige une erreur.

2. $n = 7$ et $r = 4$.

   C'est le code de Hamming vu précédemment.

3. $n = 15$ et $r = 11$.

   Il existe un tel code qui a été utilisé à l'époque dans les transmissions US.

4. $n = 31$ et $r = 26$.

   Il est théoriquement possible qu'un tel code existe, mais pour en avoir la certitude, il faudrait l'exhiber. Or même s'il existait, un tel code ne serait pas si utile car il ne permettrait que de corriger une erreur sur les 31 positions possibles.

5. Pour tous les nombres n entre 2 et 31 qui n'apparaissent pas ci-dessus, la valeur de r n'est pas entière. On est donc sûr qu'il n'y a pas de codes binaires de ces longueurs n.

8.4 Les codes ISBN

En 1972, on a commencé à assigner un numéro ISBN (International Standard Book Number) aux livres (certaines informations y sont cachées, comme la zone linguistique, etc). En 2007, le code ISBN-13 est apparu pour principalement deux raisons : augmenter la capacité de numérotation des ouvrages et s'aligner avec les codes barres.

8.4.1 Le code ISBN-10

On note $\mathcal{N} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ l'ensemble des chiffres. Le code ISBN-10 est un code de longueur 10 sur l'alphabet $\mathcal{A} = \mathcal{N} \cup \{\mathrm{X}\}$.

L'application de codage est la suivante.

$$(x_1, \ldots, x_9) \mapsto (x_1, \ldots, x_9, \underbrace{x_{10}}_{\text{caractère de contrôle}})$$

Le caractère de contrôle x_{10} est le reste de division par 11 du nombre $\sum_{i=1}^{9} i \cdot x_i$.
Autrement dit :

$$x_{10} \equiv \sum_{i=1}^{9} i \cdot x_i \pmod{11}$$

On écrit X à la place de x_{10} si x_{10} vaut 10. Ainsi, seul le 10-ième caractère peut être un X.

Le code ISBN permet de coder 10^9 livres.

8.4.2 Le code ISBN-13

On note $\mathcal{N} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ l'ensemble des chiffres. Le code ISBN-13 est un code de longueur 13 sur l'alphabet $\mathcal{N}$.

L'application de codage est la suivante.

$$(x_1, \ldots, x_{12}) \mapsto (x_1, \ldots, x_{12}, \underbrace{x_{13}}_{\text{caractère de contrôle}})$$

Le caractère de contrôle x_{13} est le reste de division par 10 du nombre $-\sum_{\substack{i \text{ impair} \\ i \neq 13}} x_i \; - \; 3 \sum_{i \text{ pair}} x_i$.
Autrement dit :

$$\sum_{i \text{ impair}} x_i \; + \; 3 \sum_{i \text{ pair}} x_i \equiv 0 \pmod{10}$$

Le code ISBN permet de coder 10^{12} livres.

Compatibilité en ISBN-10 et ISBN-13

Pour passer d'un code ISBN-10 à un code ISBN-13, on enlève le caractère de contrôle, on ajoute 978 (pour la plupart des ouvrages) et avec le code à 12 chiffres obtenus, on calcule le caractère de contrôle en suivant la méthode du code ISBN-13. Par exemple, on a

ISBN-10	étape 1	étape 2	ISBN-13
047144779X	047144779	978047144779	9780471447795
2980859737	298085973	978298085973	9782980859731

On ne peut pas faire la démarche à l'envers, car certains nouveaux ouvrages n'ont pas 978 au début de leur code ISBN-13.