Cours de maths pour l'info 2 de S. Paños

FMdKdD
fmdkdd [à] free.fr

Université du Havre
Année 2008–2009

Table des matières

1 Combinatoire et probabilités **2**
1.1 Dénombrement 2
1.2 Probabilités 9

2 Codage et décodage **23**
2.1 Théorie des codes 23

Chapitre 1

Combinatoire et probabilités

1.1 Dénombrement

Définition 1.1. On dit qu'un ensemble E est fini et de cardinal n s'il existe une bijection de E dans $\mathbb{N}_n^*$.

On note $n = \operatorname{card} \mathrm{E}$ ou encore $n = |\mathrm{E}|$.

Remarque. Si deux ensembles E et F sont en bijection avec $\mathbb{N}_n^*$, alors ils sont en bijection entre eux : si $f : \mathrm{E} \to \mathbb{N}_n^*$ et $g : \mathrm{F} \to \mathbb{N}_n^*$ sont des bijections, alors $g^{-1} \circ f : \mathrm{E} \to \mathrm{F}$ est une bijection.

De même, si on connaît le cardinal d'un ensemble fini, il suffit de construire une bijection entre lui et un autre ensemble pour prouver que cet ensemble est fini et de même cardinal que le premier : si $f : \mathrm{E} \to \mathbb{N}_n^*$ et $g : \mathrm{E} \to \mathrm{F}$ sont bijectives, alors $f \circ g^{-1} : \mathrm{F} \to \mathbb{N}_n^*$ est bijective.

Théorème 1.1. *Soit* E *un ensemble fini de cardinal n. Si* $(\mathrm{A}_i)_{i \in \mathbb{N}_k^*}$ *est une partition de* E*, alors* $\operatorname{card} \mathrm{E} = \sum_{i=1}^{k} \operatorname{card} \mathrm{A}_i$.

Démonstration. Chaque élément A_i de la partition peut être considéré comme le tiroir d'une commode. Pour savoir combien il y a d'éléments dans la commode, il suffit de compter les objets de chaque tiroir, puis de faire la somme des nombres obtenus pour tous les tiroirs. □

Théorème 1.2.
1. *Soient* A *et* B *deux ensembles finis de cardinaux n et p. Alors leur produit cartésien* $\mathrm{A} \times \mathrm{B}$ *a pour cardinal np.*
2. *Soient* $\mathrm{A}_1, \ldots, \mathrm{A}_s$ *s ensembles finis de cardinaux respectifs* $n_1, \ldots, n_s$*. Alors leur produit cartésien a pour cardinal* $n_1 \cdot n_2 \cdot \cdots \cdot n_s$.

Démonstration. 1. On va désigner par $a_1, \ldots, a_n$ les éléments de A et par $b_1, \ldots, b_p$ les éléments de B. Alors $\mathrm{A} \times \mathrm{B} = \{(a, b) / a \in \mathrm{A} \text{ et } b \in \mathrm{B}\} =$

$\{(a_i, b_j)/i \in \mathbb{N}_n^*, j \in \mathbb{N}_p^*\}$. Posons $\mathrm{H}_i = \{(a_i, b_j)/j \in \mathbb{N}_p^*\}$ pour chaque i de $\mathbb{N}_n^*$. Alors $(\mathrm{H}_i)_{i\in\mathbb{N}_n^*}$ est une partition de $\mathrm{A}\times\mathrm{B}$ et $\forall i \in \mathbb{N}_n^*$, $\operatorname{card}\mathrm{H}_i = p$. D'après le théorème 1.1, $\operatorname{card}\mathrm{A}\times\mathrm{B} = \sum_{i=1}^{n}\operatorname{card}\mathrm{H}_i = np$.

2. Désignons par $\mathrm{B} = \mathrm{A}_2 \times \cdots \times \mathrm{A}_s$ et raisonnons par récurrence sur s. On suppose le théorème démontré pour tout produit cartésien de $(s-1)$ ensembles. Considérons un produit cartésien de s ensembles : $\mathrm{A}_1 \times \mathrm{A}_2 \times \cdots \times \mathrm{A}_s = \mathrm{A}_1 \times \mathrm{B}$. D'après le (1), $\operatorname{card}\mathrm{A}_1 \times \mathrm{B} = \operatorname{card}\mathrm{A}_1 \times \operatorname{card}\mathrm{B}$, et par hypothèse de récurrence, $\operatorname{card}\mathrm{B} = \operatorname{card}\mathrm{A}_2 \times \cdots \times \operatorname{card}\mathrm{A}_s$. D'où

$$\operatorname{card}\mathrm{A}_1 \times \mathrm{A}_2 \times \cdots \times \mathrm{A}_s = \operatorname{card}\mathrm{A}_1 \times \operatorname{card}\mathrm{A}_2 \times \cdots \times \operatorname{card}\mathrm{A}_s$$

□

Remarque. Cette formule est très utile pour dénombrer les façons de réaliser un processus en s étapes, chaque étape pouvant se réaliser de n_s façons différentes. La structure d'arbre est la mieux adaptée à la représentation d'un tel processus.

Exemple. Combien y a-t-il de chemins possibles pour un promeneur voulant aller de A à B sans jamais revenir en arrière ? C'est un processus à 3 étapes avec 3 choix à la première, 2 choix à la seconde, 4 à la troisième. Donc il y a 24 façons différentes d'aller de A à B.

Théorème 1.3. *Soit* E *un ensemble fini de cardinal* n *et* F *un ensemble fini de cardinal* p*. Alors l'ensemble des applications de* E *dans* F*,* $\mathscr{F}(\mathrm{E},\mathrm{F})$*, a pour cardinal* $(\operatorname{card}\mathrm{F})^{\operatorname{card}\mathrm{E}} = p^n$.

Démonstration. Réaliser une application de E dans F, c'est faire aboutir un processus de choix des n images des n éléments de E. Or chaque image peut être n'importe lequel des p éléments de F. On a donc n étapes avec p choix à chaque étape, c'est à dire

$$\underbrace{p \times p \times \cdots \times p}_{n \text{ fois}} = p^n$$

choix possibles, donc p^n applications de E dans F. □

Théorème 1.4. *Soit* E *un ensemble fini de cardinal* n*. Alors le cardinal de* $\mathscr{P}(\mathrm{E})$*, ensemble des parties de* E*, est égal à* 2^n.

Démonstration. Construire une partie de E consiste à prendre les n éléments de E un à un (n étapes), puis à décider de le mettre ou non dans la partie (2 choix). Il y a donc $\underbrace{2 \times 2 \times \cdots \times 2}_{n \text{ fois}}$ parties dans E. □

Remarque. On peut associer de façon bijective les parties de E et les applications de E dans $\{0,1\}$, puis utiliser le théorème 1.3.

Théorème 1.5. *1. Soient* E *et* F *deux ensembles finis de cardinaux respectifs* n *et* p $(n \leqslant p)$. *Le nombre d'injections de* E *dans* F *est égal à :* $p(p-1)\dots(p-n+1) = \frac{p!}{(p-n)!}$.

2. Si $n = p$, *le nombre de bijections de* E *dans* F *est* $p!$.

Démonstration. 1. Réaliser une injection de E dans F, c'est réaliser un processus en n étapes, chaque étape consistant à choisir un élément dans un ensemble ayant un élément de moins à chaque étape, puisque pour que l'application soit injective il faut que les n images des éléments de E soient toutes différentes. On a donc p choix à la première étape, $(p-1)$ à la seconde, $(p-2)$ à la troisième, …, $(p-n+1)$ à la n^{e}. On a donc $\frac{p!}{(p-n)!}$ injections de E dans F.

2. Dans le cas où $n = p$, l'ensemble des injections de E dans F est égal à l'ensemble des bijections de E dans F. Le nombre cherché est alors $\frac{p!}{(p-p)!} = \frac{p!}{0!} = p!$.

□

Définition 1.2. 1. On appelle permutation d'un ensemble E de cardinal n toute bijection de E dans E.

2. On appelle arrangement de n éléments pris p à p, sans répétition, tout p-uplet $(e_1, \dots, e_p)$ formé d'éléments deux à deux distincts pris parmi les n éléments.

3. On appelle combinaison, sans répétition, de n éléments pris p à p toute partie à p éléments de cet ensemble à n éléments.

Remarques. 1. Le cardinal de l'ensemble des permutations de E, S_E, est $(\operatorname{card} E)!$.

2. Un arrangement de n éléments pris p à p sans répétition est une suite de longueur p dont les éléments sont pris parmi les n. C'est donc une application de $\mathbb{N}_p^*$ dans cet ensemble de n éléments. Ces p éléments étant tous distincts deux à deux, cette application est injective. Si on note A_n^p le nombre de ces arrangements, le théorème 1.5 nous dit que $A_n^p = \frac{n!}{(n-p)!}$.

Théorème 1.6. *Le nombre de combinaisons de* n *éléments pris* p *à* p *est* $\binom{n}{p} = \frac{n!}{p!(n-p)!}$.

Démonstration. Par définition, $\binom{n}{p}$ est égal au nombre de parties à p éléments dans un ensemble à n éléments. Calculons A_n^p d'une façon différente de celle du théorème 1.5. Soit f une injection d'un ensemble à p éléments dans un ensemble à n éléments (avec $p \leqslant n$). Pour construire f procédons en deux étapes. D'abord choisissons une partie à p éléments parmi les n : il y a $\binom{n}{p}$ façons de le faire. Ensuite répartissons ces p objets en face des p éléments de l'ensemble de départ : il y a $p!$ façons de le faire. D'où $A_n^p = \binom{n}{p} p!$ et donc $\binom{n}{p} = \frac{A_n^p}{p!} = \frac{n!}{p!(n-p)!}$. □

Remarques. 1. Les nombres $\binom{n}{p}$ portent aussi le nom de coefficients binômiaux car ils figurent dans la formule du binôme de Newton.

$$(a+b)^n = \sum_{p=0}^{n} \binom{n}{p} a^p b^{n-p}$$

2. Ces coefficients vérifient les propriétés suivantes :

$$\begin{aligned}
\binom{n}{p} &= \binom{n}{n-p}, \\
\sum_{p=0}^{n} \binom{n}{p} &= 2^n, \\
\binom{n}{p} &= \frac{n}{p}\binom{n-1}{p-1} = \frac{n}{n-p}\binom{n-1}{p} = \frac{n-p+1}{p}\binom{n}{p-1}, \\
\text{et } \binom{n}{p} &= \binom{n-1}{p} + \binom{n-1}{p-1}
\end{aligned}$$

Définition 1.3. 1. On appelle arrangement de n éléments pris p à p, avec répétitions, tout p-uplet $(e_1, \ldots, e_p)$ formé d'éléments pris parmi les n, le même élément pouvant être choisi autant de fois que l'on veut.

2. On appelle combinaison de n éléments pris p à p avec répétitions toute liste non ordonnée de p élément pris parmi les n, les répétitions étant autorisées.

Théorème 1.7. *1. Le nombre d'arrangements de n éléments pris p à p avec répétitions est n^p.*

2. Le nombre de combinaisons de n éléments pris p à p avec répétitions est $\binom{n+p-1}{p}$.

Démonstration. 1. Pour réaliser un p-uplet avec les n éléments, le même élément pouvant être choisi autant de fois que l'on veut, on réalise un processus en p étapes avec n choix à chaque étape : il y a donc $\underbrace{n \times \cdots \times n}_{p \text{ fois}} = n^p$ réalisations possibles.

2. Considérons une combinaison de n éléments $a_1, \ldots, a_n$ pris p à p avec répétitions. Si x_i représente le nombre de a_i figurant dans la combinaison, celle-ci est parfaitement déterminée par la connaissance des x_i. Il y a donc autant de combinaisons à répétitions que de systèmes $x_1, \ldots, x_n$ tels que $\sum_{i=1}^{n} x_i = p$. Dénombrons ces solutions.

   On peut toujours représenter p par des sommes de paquets de « 1 » et toute solution de l'équation peut s'écrire sous forme de paquets de « 1 » séparés par $(n-1)$ signes « + ». Donc à toute solution de l'équation on peut associer une permutation des $(n+p-1)$ « 1 » et « + » dont p sont identiques à « 1 » et $(n-1)$ identiques à « + ». Il y a $\binom{n+p-1}{p}$ façons de réaliser ces permutations, car on réalise alors un processus en deux étapes :
   - On choisit p places pour les symboles « 1 » parmi les $(n+p-1)$ places vides. On a donc $\binom{n+p-1}{p}$ façons de le faire.
   - On comble les vides avec les $(n-1)$ symboles plus « + » restants. On a un seul choix.

□

Théorème 1.8 (Formule de Sylvester)**.** *Soient* $\mathrm{A}_1, \ldots, \mathrm{A}_n$ *n parties d'un ensemble fini* E*. Alors*

$$\operatorname{card} \bigcup_{i=1}^{n} \mathrm{A}_i = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \operatorname{card} \bigcap_{l=1}^{k} \mathrm{A}_{i_l}$$

Démonstration. Opérons par récurrence sur n :

- Si $n = 1$, $\operatorname{card} \mathrm{A}_1 = (-1)^{1-1} \operatorname{card} \mathrm{A}_1$.
- Si $n = 2$, considérons une partition de $\mathrm{A}_1 \cup \mathrm{A}_2$:

$$\mathrm{A}_1 \cup \mathrm{A}_2 = (\mathrm{A}_1 \setminus \mathrm{A}_2) \cup (\mathrm{A}_2 \setminus \mathrm{A}_1) \cup (\mathrm{A}_1 \cap \mathrm{A}_2)$$

Or $(\mathrm{A}_1 \setminus \mathrm{A}_2) \cup (\mathrm{A}_1 \cap \mathrm{A}_2) = \mathrm{A}_1$ et donc $\operatorname{card}(\mathrm{A}_1 \setminus \mathrm{A}_2) = \operatorname{card} \mathrm{A}_1 - \operatorname{card}(\mathrm{A}_1 \cap \mathrm{A}_2)$.

De même, $\operatorname{card}(A_2 \setminus A_1) = \operatorname{card} A_2 - \operatorname{card}(A_1 \cap A_2)$. D'où

$$\begin{aligned}\operatorname{card}(A_1 \cup A_2) &= [\operatorname{card} A_1 - \operatorname{card}(A_1 \cap A_2)] + [\operatorname{card} A_2 - \operatorname{card}(A_1 \cap A_2)] \\ &\quad + \operatorname{card}(A_1 \cap A_2) \\ &= \operatorname{card} A_1 + \operatorname{card} A_2 - \operatorname{card}(A_1 \cap A_2) \\ &= \sum_{k=1}^{2} (-1)^{k-1} \sum_{1 \leqslant i_1 < i_2 \leqslant 2} \operatorname{card} \bigcap_{l=1}^{k} A_{i_k}\end{aligned}$$

– Supposons le résultat établi à l'ordre m,

$$\operatorname{card} \bigcup_{i=1}^{m} A_i = \sum_{k=1}^{m} (-1)^k \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant m} \operatorname{card} \bigcap_{l=1}^{k} A_{i_l}$$

et calculons $\operatorname{card} \bigcup_{i=1}^{m+1} A_i$. Posons $A = \bigcup_{i=1}^{m} A_i$; alors

$$\begin{aligned}\operatorname{card} \bigcup_{i=1}^{m+1} A_i &= \operatorname{card} \left[\left(\bigcup_{i=1}^{m} A_i \right) \cup A_{m+1} \right] \\ &= \operatorname{card} A \cup A_{m+1} \\ &= \operatorname{card} A + \operatorname{card} A_{m+1} - \operatorname{card} A \cap A_{m+1}\end{aligned}$$

Posons $A'_i = A_i \cap A_{m+1}$ pour $i \in \mathbb{N}^*_m$. Alors

$$\begin{aligned}(A_{i_1} \cup \cdots \cup A_{i_p}) \cap A_{m+1} &= A'_{i_1} \cup \cdots \cup A'_{i_p} \\ (A_{i_1} \cap \cdots \cap A_{i_p}) \cap A_{m+1} &= A'_{i_1} \cap \cdots \cap A'_{i_p}\end{aligned}$$

Or $\operatorname{card}(A \cap A_{m+1}) = \operatorname{card}(A'_1 \cup \cdots \cup A'_m)$ et par hypothèse de récurrence,

$$\begin{aligned}\operatorname{card} \bigcup_{i=1}^{m} A'_i &= \sum_{i=1}^{m} \operatorname{card} A'_i + \sum_{p=2}^{m} (-1)^{p-1} \sum_{1 \leqslant i_1 < \cdots < i_p < m} \operatorname{card} \bigcap_{j=1}^{p} A'_{i_j} \\ &= \sum_{i=1}^{m} \operatorname{card} A'_i + \sum_{p=2}^{m} (-1)^{p-1} \sum_{1 \leqslant i_1 < \cdots < i_p \leqslant m} \operatorname{card} \left[\left(\bigcap_{j=1}^{p} A_{i_j} \right) \cap A_{m+1} \right]\end{aligned}$$

Donc

$$\begin{aligned}\operatorname{card} \bigcup_{i=1}^{m+1} A'_i &= \sum_{i=1}^{m} \operatorname{card} A_i + \operatorname{card} A_{m+1} - \sum_{i=1}^{m} \operatorname{card} A'_i \\ &\quad + \sum_{p=2}^{m} (-1)^{p-1} \sum_{1 \leqslant i_1 < \cdots < i_p \leqslant m} \operatorname{card} \bigcap_{j=1}^{p} A_{i_j} \\ &\quad - \sum_{p=2}^{m} (-1)^{p-1} \sum_{1 \leqslant i_1 < \cdots < i_p \leqslant m} \operatorname{card} \left[\left(\bigcap_{j=1}^{p} A_{i_j} \right) \cap A_{m+1} \right]\end{aligned}$$

D'où

$$\operatorname{card}\bigcup_{i=1}^{m+1} A_i = \sum_{p=1}^{m+1}(-1)^{p-1} \sum_{1\leqslant i_1<\cdots<i_p\leqslant m+1} \operatorname{card}\bigcap_{j=1}^{p} A_{i_j}$$

□

Théorème 1.9. *Soient* E *et* F *deux ensembles de cardinaux respectifs* m *et* n*. Alors le nombre* S *de surjections de* E *dans* F *est*

$$S = \begin{cases} 0 & si\ m < n \\ n! & si\ n = m \\ \displaystyle\sum_{p=0}^{n}(-1)^p \binom{n}{p}(n-p)^m & si\ m > n \end{cases}$$

Démonstration. Si $m < n$, l'image de E par f ne peut contenir plus d'éléments que E. Donc si $m < n$, $\widehat{f}(E)$ sera strictement inclus dans F et f ne sera pas surjective.

Si $m = n$, toute surjection de E sur F est une bijection, et réciproquement toute bijection de E sur F est une surjection. Le nombre S est alors égal au cardinal de l'ensemble des bijections de E dans F, soit $n!$.

Si $m > n$, considérons $N = \{f \in \mathscr{F}(E,F)/f \text{ non surjective }\}$. Alors $S = \operatorname{card}\mathscr{F}(E,F) - \operatorname{card} N = n^m - \operatorname{card} N$. Calculons le cardinal de N. Désignons par $x_1,\ldots,x_n$ les éléments de F. Alors $(f \in N) \Rightarrow (\exists i \in \mathbb{N}/x_i \notin \widehat{f}(E))$. Soit $A_i = \{f \in \mathscr{F}(E,F)/x_i \notin \widehat{f}(E)\}$ pour $i \in \mathbb{N}_n^*$. Les $(A_i)_{n\in\mathbb{N}_n^*}$ forment un recouvrement de N : $\bigcup_{i=1}^{n} A_i = N$. D'après le théorème 1.8 :

$$\operatorname{card} N = \sum_{p=1}^{n}(-1)^{p-1} \sum_{1\leqslant i_1<\cdots<i_p\leqslant n} \operatorname{card}\bigcap_{j=1}^{p} A_{i_j}$$

Il faut donc évaluer les $\operatorname{card}\bigcap_{i=1}^{p} A_{i_j}$. Or,

$$\begin{aligned}
\operatorname{card}\bigcap_{i=1}^{p} A_{i_j} &= \operatorname{card}\left(\{f \in \mathscr{F}(E,F)/x_{i_1} \notin \widehat{f}(E)\} \cap \cdots \cap \{f \in \mathscr{F}(E,F)/x_{i_p} \notin \widehat{f}(E)\}\right) \\
&= \operatorname{card}\left(\{f \in \mathscr{F}(E,F)/\{x_{i_1},\ldots,x_{i_p}\} \not\subseteq \widehat{f}(E)\}\right) \\
&= \operatorname{card}\left(\{f \in \mathscr{F}(E,F)/\widehat{f}(E) \subseteq \complement_F\{x_{i_1},\ldots,x_{i_p}\}\right) \\
&= \operatorname{card}\left(\{f \in \mathscr{F}(E,F\setminus\{x_{i_1},\ldots,x_{i_p}\})\}\right) \\
&= \operatorname{card}\mathscr{F}(E,F\setminus\{x_{i_1},\ldots,x_{i_p}\}) \\
&= \operatorname{card}(F\setminus\{x_{i_1},\ldots,x_{i_p}\})^{\operatorname{card} E} \\
&= (n-p)^m
\end{aligned}$$

∎

D'où

$$\operatorname{card} \mathrm{N} = \sum_{p=1}^{n} (-1)^{p-1} \sum_{1 \leqslant i_1 < \cdots < i_p \leqslant n} (n-p)^m$$

La seconde somme est de termes constants. Il y a autant de termes que de choix possibles de $i_1, \ldots, i_p$ différents entre 1 et n. Toute partie à p éléments de $\mathbb{N}_n^*$ nous donne une telle suite et une seule. Il y a donc $\binom{n}{p}$ termes dans cette somme. D'où

$$\sum_{1 \leqslant i_1 < \cdots < i_p \leqslant n} (n-p)^m = \binom{n}{p} (n-p)^m$$

et donc

$$\operatorname{card} \mathrm{N} = \sum_{p=1}^{n} (-1)^{p-1} \binom{n}{p} (n-p)^m$$

Par conséquent,

$$\begin{aligned}
\mathrm{S} &= n^m - \operatorname{card} \mathrm{N} \\
&= (-1)^0 (n-0)^m - \sum_{p=1}^{n} (-1)^{p-1} \binom{n}{p} (n-p)^m \\
&= (-1)^0 \binom{n}{0} (n-0)^m + \sum_{p=1}^{n} (-1)^p \binom{n}{p} (n-p)^m \\
&= \sum_{p=0}^{n} (-1)^p \binom{n}{p} (n-p)^m
\end{aligned}$$

□

1.2 Probabilités

Définition 1.4. Un processus physique qui produit des effets aléatoires, c'est-à-dire qui ne peuvent être prédits avec certitude, est appelé une expérience aléatoire. Chaque effet résultant de l'expérience aléatoire est appelé résultat. On appelle univers et on désigne par Ω l'ensemble des résultats.

Remarque. Nous allons, pour définir la notion d'évènement lié à une expérience aléatoire, distinguer deux cas suivant que Ω est fini ou dénombrable, ou bien d'un cardinal supérieur ou égal à $\aleph_1$.

Dans le premier cas, toute partie de Ω sera un évènement lié à l'expérience, alors que ce ne sera pas le cas lorsque $\operatorname{card} \Omega \geqslant \aleph_1$: on choisira pour ensemble des évènements liés à l'expérience une partie de $\mathscr{P}(\Omega)$ ayant une structure particulière, celle de tribu ou de σ-algèbre.

Définition 1.5. Soit Ω un univers. On appelle tribu ou σ-algèbre sur Ω, tout ensemble $\mathscr{T}$ de parties de Ω, telles que :

1. $\Omega \in \mathscr{T}$
2. Si $A \in \mathscr{T}$ alors $\complement_\Omega A \in \mathscr{T}$.
3. Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'éléments de $\mathscr{T}$, alors $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{T}$.

Exemple. 1. $\mathscr{P}(\Omega)$ est une tribu sur Ω.

2. $\{\emptyset, \Omega\}$ est une tribu sur Ω.

Définition 1.6. On appelle espace probabilisable tout couple $(\Omega, \mathscr{T})$ constitué d'un univers Ω, et d'une tribu $\mathscr{T}$ sur Ω. Les éléments de $\mathscr{T}$ sont appelés évènements. $\emptyset$ est appelé évènement impossible. Ω est appelé évènement certain.

On dit que l'évènement A implique l'évènement B si $A \subseteq B$. On dit que $\complement A$, que l'on note $\overline{A}$, est l'évènement contraire de A. On dit que les éléments A et B sont incompatibles si $A \cap B = \emptyset$. On appelle système complet d'évènements de Ω toute partition $(A_n)_{n\in\mathbb{N}}$ de Ω dont les éléments sont dans $\mathscr{T}$. Dans le cas où $\operatorname{card}\Omega \leqslant \aleph_0$, $\mathscr{T}$ est toujours égal à $\mathscr{P}(\Omega)$ et tout singleton prend alors le nom d'évènement élémentaire.

Définition 1.7. Soit $(\Omega, \mathscr{T})$ un univers probabilisable. On appelle probabilité sur $(\Omega, \mathscr{T})$ toute application $P : \mathscr{T} \to [0,1] \subset \mathbb{R}$ telle que :

1. $P(\Omega) = 1$
2. Quelque soit la suite $(A_n)_{n\in\mathbb{N}}$ d'évènements deux à deux incompatibles, $P(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n=0}^{+\infty} P(A_n)$.

Remarques. 1. Lorsque Ω est fini, $\Omega = \{\omega_1, \ldots, \omega_n\}$. Si P est une probabilité sur $(\Omega, \mathscr{P}(\Omega))$, posons $p_i = P(\{\omega_i\})$. Alors pour tout $i \in \mathbb{N}_n^*$, $0 \leqslant p_i \leqslant 1$ et $\sum_{i=1}^n p_i = 1$.
Réciproquement, si $(p_i)_{i\in\mathbb{N}_n^*}$ est une famille de réels vérifiant $\forall i \in \mathbb{N}_n^*$, $0 \leqslant p_i \leqslant 1$ et $\sum_{i=1}^n p_i = 1$ alors il existe une probabilité P et une seule sur $(\Omega, \mathscr{P}(\Omega))$ telle que $\forall i \in \mathbb{N}_n^*$, $P(\{\omega_i\}) = p_i$.

2. Une probabilité très utilisée sur les ensembles finis est l'équiprobabilité : si $\Omega = \{\omega_1, \ldots, \omega_n\}$ on pose $\forall i \in \mathbb{N}_n^*$, $P(\{\omega_i\}) = \frac{1}{n} = \frac{1}{\operatorname{card}\Omega}$. Alors $\forall A \in \mathscr{P}(\Omega)$, $P(A) = \frac{\operatorname{card} A}{\operatorname{card}\Omega}$. On dit aussi que Ω est muni de la probabilité uniforme.

3. Dans le cas d'un univers dénombrable $\Omega = \{\omega_1, \ldots, \omega_n, \ldots\}$, on a une propriété analogue : toute probabilité sur Ω est caractérisée par la donnée des probabilités des évènements élémentaires. Si $p_n = P(\{\omega_n\})$ alors $\sum_{n\in\mathbb{N}^*} p_n = 1$ et $\forall n \in \mathbb{N}^*$, $p_n \geqslant 0$. Par contre on ne peut retenir l'hypothèse d'équiprobabilité, car alors $\frac{\operatorname{card}\{\omega_n\}}{\operatorname{card}\Omega} = \frac{1}{\aleph_0} = 0$ et $\sum_{n\in\mathbb{N}^*} p_n = 0 \neq 1$.

Définition 1.8. On appelle espace probabilisé ou univers probabilisé tout triplet $(\Omega, \mathscr{T}, \mathrm{P})$ où $(\Omega, \mathscr{T})$ est un espace probabilisable et P une probabilité sur Ω.

Théorème 1.10. *Soit $(\Omega, \mathscr{T}, \mathrm{P})$ un univers probabilisé.*

1. $\mathrm{P}(\emptyset) = 0$
2. *Si* A *et* B *sont deux évènements incompatibles,* $\mathrm{P}(\mathrm{A} \cup \mathrm{B}) = \mathrm{P}(\mathrm{A}) + \mathrm{P}(\mathrm{B})$. *En général, si* $\mathrm{A}_1, \ldots, \mathrm{A}_n$ *sont n évènements incompatibles, on a* $\mathrm{P}(\bigcup_{i=1}^{n} \mathrm{A}_i) = \sum_{i=1}^{n} \mathrm{P}(\mathrm{A}_i)$.
3. *Si* A *est un évènement quelconque,* $\mathrm{P}(\overline{\mathrm{A}}) = 1 - \mathrm{P}(\mathrm{A})$.
4. *Si* A *et* B *sont deux évènements tels que* A *implique* B ($\mathrm{A} \subseteq \mathrm{B}$)*, alors* $\mathrm{P}(\mathrm{B} - \mathrm{A}) = \mathrm{P}(\mathrm{B}) - \mathrm{P}(\mathrm{A})$ *et* $\mathrm{P}(\mathrm{A}) \leqslant \mathrm{P}(\mathrm{B})$.
5. *Si* A *et* B *sont deux évènements quelconques,* $\mathrm{P}(\mathrm{A} \cup \mathrm{B}) = \mathrm{P}(\mathrm{A}) + \mathrm{P}(\mathrm{B}) - \mathrm{P}(\mathrm{A} \cap \mathrm{B})$. *Plus généralement,*

$$\mathrm{P}\left(\bigcup_{i=1}^{n} \mathrm{A}_i\right) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \mathrm{P}\left(\bigcap_{j=1}^{k} \mathrm{A}_{i_j}\right)$$

Formule de Poincaré.

Démonstration. 1. Considérons la suite $(\mathrm{A}_n)_{n \in \mathbb{N}}$ définie par $\forall n \in \mathbb{N}$, $\mathrm{A}_n = \emptyset$. Tous ces évènements sont incompatibles, donc d'après la définition 1.7,

$$\mathrm{P}\left(\bigcup_{n \in \mathbb{N}} \mathrm{A}_n\right) = \sum_{i=0}^{+\infty} \mathrm{P}(\mathrm{A}_n)$$

Pour que cette égalité ait un sens, il faut que la série du membre droit converge, donc que son terme général tende vers 0 quand n tend vers l'infini. D'où $\mathrm{P}(\emptyset) = 0$.

2. Considérons la suite $(\mathrm{A}_n)_{n \in \mathbb{N}}$ définie par $\mathrm{A}_0 = \mathrm{A}$, $\mathrm{A}_1 = \mathrm{B}$ et $\mathrm{A}_n = \emptyset$ pour $n \geqslant 2$. Alors

$$\mathrm{P}\left(\bigcup_{n \in \mathbb{N}} \mathrm{A}_n\right) = \sum_{i=0}^{+\infty} \mathrm{P}(\mathrm{A}_i) = \mathrm{P}(\mathrm{A}) + \mathrm{P}(\mathrm{B}) + \mathrm{P}(0) + \cdots = \mathrm{P}(\mathrm{A}) + \mathrm{P}(\mathrm{B})$$

De même, considérons la suite $(\mathrm{B}_n)_{n \in \mathbb{N}}$ définie par $\mathrm{B}_0 = \mathrm{A}_1$, $\mathrm{B}_1 = \mathrm{A}_2$, …, $\mathrm{B}_{n-1} = \mathrm{A}_n$, $\mathrm{B}_n = \emptyset = \mathrm{B}_{n+1} = \ldots$. Alors

$$\mathrm{P}\left(\bigcup_{n \in \mathbb{N}} \mathrm{B}_n\right) = \sum_{j=0}^{+\infty} \mathrm{P}(\mathrm{B}_j) = \mathrm{P}(\mathrm{A}_1 \cup \cdots \cup \mathrm{A}_n) = \sum_{j=0}^{n-1} \mathrm{P}(\mathrm{A}_{j+1}) + \sum_{j=n}^{+\infty} \mathrm{P}(\emptyset)$$

3. On applique le 2 aux deux évènements incompatibles A et $\overline{A}$ et on obtient :
$$P(\Omega) = P(A \cup \overline{A}) = P(A) + P(\overline{A}) = 1$$
D'où $P(\overline{A}) = 1 - P(A)$.
4. Soient A et B deux évènements tels que $A \subseteq B$. Posons $C = B - A$ ($= B \cap \complement_\Omega A$). A et C sont incompatibles, donc d'après le 1,
$$P(A \cup C) = P(A) + P(C) = P(A \cup (B - A)) = P(B)$$
d'où $P(C) = P(B - A) = P(B) - P(A)$. Donc $P(B) = P(A) + P(C) \geqslant P(A)$.
5. Le schéma de la démonstration est le même que pour la formule de Sylvester sur les cardinaux.

□

Définition 1.9. Soit $(\Omega, \mathscr{T}, P)$ un univers probabilisé. On dit que deux évènements A et B sont indépendants si et seulement si
$$P(A \cap B) = P(A)P(B)$$
On dit que les évènements $A_1, \ldots, A_n$ sont mutuellement indépendants si et seulement si
$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j})$$
pour tout k de $\mathbb{N}_n^*$, les $i_1, \ldots, i_k$ étant tous distincts dans $\mathbb{N}_n^*$.

Théorème 1.11. *Soient* $(\Omega, \mathscr{T}, P)$ *un univers probabilisé et* B *un évènement tel que* $P(B) \neq 0$. *Alors l'application*
$$\begin{aligned} P_B : \mathscr{T} &\to [0,1] \\ A &\mapsto P_B(A) = \frac{P(A \cap B)}{P(B)} \end{aligned}$$
est une probabilité sur l'espace probabilisable $(B, \mathscr{T} \cap B)$.

Démonstration. 1. $P_B(B) = \frac{P(B \cap B)}{P(B)} = 1$

2. Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'évènements deux à deux incompatibles. Alors les évènements $(A_n \cap B)_{n \in \mathbb{N}}$ sont également deux à deux incompatibles.

Alors

$$\begin{aligned}
P_B(\bigcup_{n=0}^{\infty} A_n) &= \frac{P[(\bigcup_{n=0}^{\infty} A_n) \cap B]}{P(B)} \\
&= \frac{P[(\bigcup_{n=0}^{\infty} (A_n \cap B))]}{P(B)} \\
&= \frac{\sum_{n=0}^{\infty} P(A_n \cap B)}{P(B)} \\
&= \sum_{n=0}^{\infty} \frac{P(A_n \cap B)}{P(B)} \\
&= \sum_{n=0}^{\infty} P_B(A_n)
\end{aligned}$$

□

Définition 1.10. La probabilité définie au théorème 1.11 s'appelle probabilité conditionnelle en B, ou probabilité conditionnelle sachant que B est réalisé.

Notation. Parfois, au lieu de la noter $P_B()$ on la note $P(|B)$.

Exemple. Un enquêteur recense le sexe des enfants dans les familles ayant deux enfants. Il y a quatre cas possibles (aîné, cadet) : (F,F), (F,G), (G,F), (G,G). Tous ont la même probabilité : $\frac{1}{4}$. Il sonne à une porte : une petite fille ouvre. Quelle est la probabilité pour que l'autre enfant soit un garçon ?

Les évènements sont réduits à (F,F), (F,G), (G,F). Dans deux cas sur trois il y a un garçon, donc la probabilité cherchée est $\frac{2}{3}$ alors qu'avant de sonner elle était de $\frac{1}{2}$.

Remarques. 1. On a toutes les propriétés des probabilités.

$$\begin{aligned}
P(\overline{A}|B) &= 1 - P(A|B) \\
P(C \cup D|B) &= P(C|B) + P(D|B) - P(C \cap D|B)
\end{aligned}$$

et si $A_1, \ldots, A_n$ sont incompatibles deux à deux,

$$P(A_1 \cup \cdots \cup A_n|B) = \sum_{i=1}^{n} P(A_i|B)$$

2. La notion d'indépendance entre A et B peut alors s'exprimer par $P_B(A) = P(A)$ car alors $\frac{P(A \cap B)}{P(B)} = P(A)$ et $P(A \cap B) = P(A)P(B)$.

3. Dans le cas d'une suite infinie $(A_n)_{n\in\mathbb{N}}$ d'évènements on dit que ces évènements sont mutuellement indépendants si et seulement si pour $n \in \mathbb{N}$, $A_0, \ldots, A_n$ sont mutuellement indépendants.
4. On montre que si $A_0, \ldots, A_n$ sont mutuellement indépendants, les évènements $B_0, \ldots, B_n$ obtenus en posant $B_k = A_k$ ou $\overline{A_k}$ ($k \in \mathbb{N}_n$) sont également mutuellement indépendants.

Théorème 1.12 (Formule des probabilités composées)**.** *Soit* $(\Omega, \mathscr{T}, P)$ *un univers probabilisé et* $A_1, \ldots, A_n$ *n évènements tels que* $P(A_1 \cap \cdots \cap A_n) \neq 0$*. Alors*

$$P(A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\ldots P(A_n|A_1 \cap \cdots \cap A_{n-1})$$

Démonstration. P étant une fonction croissante, $P(A_1 \cap \ldots A_n) \neq 0$ nous assure que

$$P(A_1 \cap \cdots \cap A_{n-1}, \ldots, P(A_1 \cap A_2), P(A_1)$$

sont également différents de 0. Donc les écritures du second membre sont justifiées.

$$\begin{aligned}
&P(A_1)P(A_2|A_1)\ldots P(A_n|A_1 \cap \cdots \cap A_{n-1}) \\
&= P(A_1)\frac{P(A_2 \cap A_1)}{P(A_1)} \times \frac{P(A_3 \cap A_2 \cap A_1)}{P(A_2 \cap A_1)} \\
&\quad \times \cdots \times \frac{P(A_{n-1} \cap A_{n-2} \cap \cdots \cap A_1)}{P(A_{n-2} \cap \cdots \cap A_1)} \times \frac{P(A_n \cap A_{n-1} \cap \cdots \cap A_1)}{P(A_{n-1} \cap A_{n-2} \cap \cdots \cap A_1)} \\
&= P(A_1 \cap \cdots \cap A_n)
\end{aligned}$$

□

Théorème 1.13 (Formule de Bayes)**.** *Soit* $(\Omega, \mathscr{T}, P)$ *un univers probabilisé. Soit* $(A_n)_{n\in\mathbb{N}^*}$ *un système complet d'évènements. Alors*

$$\forall B \in \mathscr{T}, \forall i \in \mathbb{N}^*, P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{n=1}^{\infty} P(B|A_n)P(A_n)}$$

Démonstration.

$$\frac{P(B|A_i)P(A_i)}{\sum_{n=1}^{\infty} P(B|A_n)P(A_n)} = \frac{P(B \cap A_i)}{\sum_{n=1}^{\infty} P(B \cap A_n)}$$

Les évènements $(B \cap A_n)_{n\in\mathbb{N}^*}$ sont deux à deux incompatibles, donc

$$\begin{aligned}\frac{P(B\cap A_i)}{\sum_{n=1}^{\infty} P(B\cap A_n)} &= \frac{P(B\cap A_i)}{P(\bigcup_{n=1}^{\infty}(B\cap A_n))}\\ &= \frac{P(B\cap A_i)}{P(B\cap(\bigcup_{n=1}^{\infty} A_n))}\\ &= \frac{P(B\cap A_i)}{P(B\cap\Omega)}\\ &= \frac{P(B\cap A_i)}{P(B)}\\ &= P(A_i|B)\end{aligned}$$

□

Théorème 1.14 (Formule des probabilités totales)**.** *Soit* $(\Omega, \mathcal{T}, P)$ *un univers probabilisé et* $(A_n)_{n\in\mathbb{N}}$ *un système complet d'évènements. Soit* $B \in \mathcal{T}$ *alors*

$$P(B) = \sum_{n=1}^{\infty} P(B|A_n)P(A_n)$$

Démonstration. C'est l'évolution du dénominateur dans la démonstration précédente. □

Définition 1.11. Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé. Une application $X : \Omega \to \mathbb{R}$ est appelée variable aléatoire sur Ω si

$$\forall x \in \mathbb{R}, \{\omega \in \Omega / X(\omega) \leqslant x\} \in \mathcal{T}$$

Remarque. Si X est une variable aléatoire, on peut toujours calculer la probabilité

$$P(\{\omega \in \Omega / X(\omega) \leqslant x\})$$

Exemple.

1. On jette un dé : Ω représente l'ensemble des résultats $\{1, 2, 3, 4, 5, 6\}$ et on peut définir X comme l'indication de la parité du résultat : $X(\omega) = 0$ si ω est pair, $X(\omega) = 1$ si ω est impair.
2. On tourne une roue de loterie comportant 36 numéros. On peut définir une variable aléatoire X qui représente le numéro sorti et une variable aléatoire Y égale à l'angle θ de rotation effectué par la roue à chaque lancer.

Remarque. Si $X : \Omega \to \mathbb{R}$ est une variable aléatoire et I un intervalle de $\mathbb{R}$,

$$\widetilde{X}(I) = \{\omega \in \Omega / X(\omega) \in I\}$$

est un évènement lié à l'expérience aléatoire, c'est-à-dire $\widetilde{X}(I) \in \mathscr{T}$.

Au point de vue des notations, au lieu de noter $\{\omega \in \Omega / a \leqslant X(\omega) \leqslant b\}$ on notera $a \leqslant X \leqslant b$.

Définition 1.12.
1. Soit $(\Omega, \mathscr{T}, P)$ un espace probabilisé et X une variable aléatoire sur Ω. Si $\operatorname{card}\Omega \leqslant \aleph_0$ on dit que X est une variable aléatoire discrète.
2. Si X est une variable aléatoire discrète sur $(\Omega, \mathscr{T}, P)$ et si $X(\Omega) = \{x_1, \ldots, x_n\}$ on appelle loi de la variable aléatoire X, la famille $(p_k)_{k \in \mathbb{N}_n^*}$ définie par $p_k = P(X = x_k)$ pour tout k de $\mathbb{N}_n^*$.
3. Si X est une variable aléatoire discrète sur $(\Omega, \mathscr{T}, P)$ et si $X(\Omega) = \{x_1, \ldots, x_n\}$ on appelle loi de X la famille des réels $(p_k)_{k \in \mathbb{N}^*}$ définie par $P(X = x_k) = p_k$.

Remarque. Les $(p_k)_{k \in \mathbb{N}_n^*}$ vérifient $\forall k \in \mathbb{N}_n^*$, $p_k \geqslant 0$ et $\sum_{k=1}^{n} p_k = 1$. Respectivement, les $(p_k)_{k \in \mathbb{N}^*}$ vérifient $\forall k \in \mathbb{N}^*$, $p_k \geqslant 0$ et $\sum_{k=1}^{\infty} p_k = 1$.

La donnée d'une famille (p_k) vérifiant les deux propriétés ci-dessus permet de définir la loi d'une variable aléatoire.

Définition 1.13. Soit $(\Omega, \mathscr{T}, P)$ un espace probabilisé et X une variable aléatoire sur Ω. On appelle fonction de répartition de X l'application F_X de $\mathbb{R}$ dans $[0, 1]$ définie par :

$$\forall x \in \mathbb{R}, F_X(x) = P(X \leqslant x)$$

Remarque. La fonction F_X possède les propriétés suivantes :

1. F_X est croissante sur $\mathbb{R}$.
2. $\lim_{x \to +\infty} F_X(x) = 1$ et $\lim_{x \to -\infty} F_X(x) = 0$.
3. F_X est continue à droite en tout point de $\mathbb{R}$.
4. F_X est continue à gauche en tout point de $\mathbb{R} - X(\Omega)$. Si $x_0 \in X(\Omega)$,
$$F_X(x_0) - \lim_{x \to x_0^-} F_X(x) = P(X = x_0)$$
5. Si X est une variable aléatoire discrète, F_X est une fonction en escalier sur $\mathbb{R}$ constante entre deux valeurs prises par X.
6. $\forall x \in \mathbb{R}$,
$$F_X(x) = \sum_{x_k \in X(\Omega), x_k \leqslant x} P(X = x_k)$$

Définition 1.14. Soit X une variable aléatoire discrète définie sur $(\Omega, \mathscr{T}, \mathrm{P})$. On suppose que $\mathrm{X}(\Omega) = \{x_1, \ldots, x_n, \ldots\}$. Soit s un entier naturel, on dit que X admet un moment d'ordre s si la série $\sum_{k\geqslant 1} x_k^s \mathrm{P}(\mathrm{X} = x_k)$ est absolument convergente. Le moment d'ordre s de X est alors le réel

$$m_s(\mathrm{X}) = \sum_{k=1}^{\infty} x_k^s \mathrm{P}(\mathrm{X} = x_k)$$

On appelle espérance de X le réel

$$\mathrm{E}(\mathrm{X}) = \sum_{n=1}^{\infty} x_n \mathrm{P}(\mathrm{X} = x_n) = \sum_{n=1}^{\infty} x_n p_n$$

s'il existe.

Remarques. 1. Si $\mathrm{X}(\Omega)$ est fini, l'espérance existe et

$$\mathrm{E}(\mathrm{X}) = \sum_{k=1}^{n} x_k \mathrm{P}(\mathrm{X} = x_k)$$

2. Si $\mathrm{X}(\Omega) = \{\ldots, y_n, \ldots, y_2, y_1, x_1, \ldots, x_n, \ldots\}$ où $(x_n)_{n\in\mathbb{N}^*} \to +\infty$ et $(y_n)_{n\in\mathbb{N}^*} \to -\infty$, on dira que X admet une espérance si chacune des deux séries

$$\sum_{n\geqslant 1} x_n \mathrm{P}(\mathrm{X} = x_n) \quad \text{et} \quad \sum_{n\geqslant 1} y_n \mathrm{P}(\mathrm{X} = y_n)$$

est absolument convergente.

Exemples. 1. $\mathrm{X}(\Omega) = \mathbb{N}$ et $\forall k \in \mathbb{N}$, $\mathrm{P}(\mathrm{X} = k) = e^{-\lambda}\frac{\lambda^k}{k!}$.

$$\begin{aligned}
\sum_{k=0}^{\mathrm{N}} k\mathrm{P}(\mathrm{X} = k) &= \sum_{k=0}^{\mathrm{N}} k e^{-\lambda}\frac{\lambda^k}{k!} \\
&= \sum_{k=1}^{\mathrm{N}} e^{-\lambda}\frac{\lambda^k}{(k-1)!} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\mathrm{N}} \frac{\lambda^{k-1}}{(k-1)!} \\
\mathrm{E}(\mathrm{X}) &= \lim_{\mathrm{N}\to+\infty} \sum_{k=0}^{\mathrm{N}} k\mathrm{P}(\mathrm{X} = x_k) \\
&= \lim_{\mathrm{N}\to+\infty} \left(\sum_{k=1}^{\mathrm{N}} \frac{\lambda^{k-1}}{(k-1)!} \right) \lambda e^{-\lambda} \\
&= e^{\lambda}\lambda e^{-\lambda} \\
&= \lambda
\end{aligned}$$

2. $X(\Omega) = \mathbb{N}^*$, $\forall k \in \mathbb{N}^*$, $P(X = k) = \frac{1}{k(k+1)}$.

$$\begin{aligned}\sum_{k=1}^{N} kP(X=k) &= \sum_{k=1}^{N} k\frac{1}{k(k+1)} \\ &= \sum_{k=1}^{N} \frac{1}{k+1} \qquad\qquad = +\infty\end{aligned}$$

donc X n'a pas d'espérance.

Définition 1.15. Soit X une variable aléatoire définie sur $(\Omega, \mathcal{T}, P)$ et admettant une espérance. On suppose $X = \{x_1, \ldots, x_n, \ldots\}$ avec $(x_n)_{n\in\mathbb{N}^*}$ strictement croissante tendant vers $+\infty$.

Soit s un entier naturel. On dit que X admet un moment centré d'ordre s si la série

$$\sum_{k\geqslant 1}(x_k - E(X))^s P(X = x_k)$$

est absolument convergente. Le réel

$$\sum_{k\geqslant 1}(x_k - E(X))^s P(X = x_k)$$

sera noté $\mu_s(X)$ et appelé moment centré d'ordre s de X.

Le moment centré d'ordre 2, $\mu_2(X)$ est appelé variance de X et noté V(X). La racine carrée de V(X), $\sqrt{V(X)}$, est appelée écart type de X, et notée $\sigma(X)$.

Remarques. Il faut connaître deux propriétés importantes sur les moments.

1. Si X est une variable aléatoire discrète admettant une espérance,

$$E(aX + b) = aE(X) + b \quad (a, b) \in \mathbb{R}^2$$

2. Si X est une variable aléatoire discrète admettant une variance,

$$V(aX + b) = a^2 V(X) \quad (a, b) \in \mathbb{R}^2$$

Définition 1.16.
1. Une variable aléatoire X admettant une espérance est dite centrée si $E(X) = 0$.
2. Une variable aléatoire X admettant une variance est dite réduite si $V(X) = 1$.

3. Soit X une variable aléatoire discrète admettant une variance (et donc une espérance). On appelle variable centrée réduite associée à X, la variable aléatoire X^* définie par :

$$X^* = \frac{X - E(X)}{\sigma(X)}$$

où $\sigma(X)$ est l'écart type de X.

Théorème 1.15. *Soit* X^* *la variable aléatoire centrée réduite associée à* X*. Alors* $E(X^*) = 0$ *et* $V(X^*) = 1$.

Démonstration.

$$\begin{aligned}
E(X^*) &= E\left(\frac{X - E(X)}{\sigma(X)}\right) \\
&= E\left(\frac{1}{\sigma(X)}X - \frac{E(X)}{\sigma(X)}\right) \\
&= E\left(\frac{X}{\sigma(X)} - \frac{E(X)}{\sigma(X)}\right) \\
&= \frac{1}{\sigma(X)}E(X) - \frac{E(X)}{\sigma(X)} \\
&= 0 \\
V(X^*) &= V\left(\frac{X - E(X)}{\sigma(X)}\right) \\
&= V\left(\frac{1}{\sigma(X)}X - \frac{E(X)}{\sigma(X)}\right) \\
&= \frac{1}{\sigma(X)^2}V(X) \\
&= 1
\end{aligned}$$

□

Théorème 1.16 (Formule de Koenig-Huyghens)**.** *Soit* X *une variable aléatoire discrète définie sur* $(\Omega, \mathcal{T}, P)$ *et admettant une variance. Alors*

$$V(X) = E(X^2) - E(X)^2$$

Démonstration. Si $X(\Omega) = \{x_0, x_1, \ldots, x_n, \ldots\}$,

$$\begin{aligned}
V(X) &= \sum_{k=0}^{\infty} (x_k - E(X))^2 P(X = x_k) \\
&= \sum_{k=0}^{\infty} x_k^2 P(X = x_k) - 2\left(\sum_{k=0}^{+\infty} E(X) x_k P(X = x_k)\right) + \sum_{k=0}^{+\infty} E(X)^2 P(X = x_k) \\
&= E(X^2) - 2E(X) \sum_{k=0}^{+\infty} x_k P(X = x_k) + E(X)^2 \sum_{k=0}^{+\infty} P(X = x_k) \\
&= E(X^2) - E(X)^2
\end{aligned}$$

□

Définition 1.17. Soit X une variable aléatoire définie sur $(\Omega, \mathcal{T}, P)$. On dit que X est une variable aléatoire continue si et seulement s'il existe une fonction $f : \mathbb{R} \to \mathbb{R}$ telle que :

1. $f \geqslant 0$
2. f est continue sauf peut être en un nombre fini de points.
3. f admet aux points x_i, pour tout i de $\mathbb{N}_n^*$, une limite à gauche et une limite à droite finies ou égales à $+\infty$.
4. L'intégrale $\int_{-\infty}^{+\infty} f(x)dx$ converge et est égale à 1.
5. $\forall x \in \mathbb{R}$, $F_X(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(t)dt$.

On dit alors que X admet une densité, et f est appelée une densité de X.

Remarque. Toute fonction f possédant les propriétés 1 à 4 de la définition 1.17 est la densité d'une variable aléatoire X définie sur un univers probabilisé convenablement choisi.

Théorème 1.17. *Soit* X *une variable aléatoire continue de densité* f *et de fonction de répartition* F_X. *Alors :*

1. $\forall a, b \in \mathbb{R}$, $a \leqslant b$,
$$P(a < X \leqslant b) = \int_a^b f(t)dt$$
2. $\forall x \in \mathbb{R}$, $P(X = x) = 0$
3. $\forall a, b \in \mathbb{R}$, $a \leqslant b$,
$$P(a < X \leqslant b) = P(a \leqslant X \leqslant b) = P(a \leqslant X < b) = P(a < X < b)$$

Démonstration. 1.

$$\begin{aligned} \mathrm{P}(a < \mathrm{X} \leqslant b) &= \mathrm{F}_{\mathrm{X}}(b) - \mathrm{F}_{\mathrm{X}}(a) \\ &= \int_{-\infty}^{b} f(t)dt - \int_{-\infty}^{a} f(t)dt \\ &= \int_{a}^{b} f(t)dt \end{aligned}$$

2. Remarquons que

$$(\mathrm{X} = x) = \bigcap_{n \geqslant 1} (x - \frac{1}{n} < \mathrm{X} \leqslant x)$$

La suite $\left((x - \frac{1}{n} < \mathrm{X} \leqslant x)\right)_{n \geqslant 1}$ est une suite décroissante d'évènements. D'où

$$\begin{aligned} \mathrm{P}(\mathrm{X} = x) &= \lim_{n \to +\infty} \mathrm{P}(x - \frac{1}{n} < \mathrm{X} \leqslant x) \\ &= \lim_{n \to +\infty} \int_{x - \frac{1}{n}}^{x} f(t)dt \\ &= 0 \end{aligned}$$

3. Montrons l'une des égalités, les autres se montrant de façon identique.

$$\begin{aligned} \mathrm{P}(a \leqslant \mathrm{X} \leqslant b) &= \mathrm{P}(a < \mathrm{X} \leqslant b) + \mathrm{P}(\mathrm{X} = a) \\ &= \mathrm{P}(a < \mathrm{X} \leqslant b) \end{aligned}$$

□

Définition 1.18. 1. Soit X une variable aléatoire de densité f, définie sur $(\Omega, \mathscr{T}, \mathrm{P})$. Si $s \in \mathbb{N}^*$, on appelle moment d'ordre s le réel

$$m_s(\mathrm{X}) = \int_{-\infty}^{+\infty} t^s f(t)dt$$

sous réserve de convergence de cette intégrale.

2. On appelle espérance de X le moment d'ordre 1 de X et on le note E(X).

Définition 1.19. 1. Soit X une variable aléatoire continue de densité f. On suppose que X admet une espérance. Si $s \in \mathbb{N}^*$, on appelle moment centré d'ordre s le réel

$$\mu_s = \int_{-\infty}^{+\infty} (t - \mathrm{E}(\mathrm{X}))^s f(t)dt$$

sous réserve de convergence de cette intégrale.

2. Le moment centré d'ordre 2, $\mu_2(X)$, est appelé variance de X et est noté V(X).
3. On appelle écart type de X le réel $\sigma(X) = \sqrt{V(X)}$ si V(X) existe.

Remarques. 1. On a, comme pour les variables discrètes,

$$\forall a, b \in \mathbb{R}, E(aX + b) = aE(X) + b$$

2. La formule de Koenig-Huygens reste valable : $V(X) = E(X^2) - E(X)^2$.
3. Si $X = Y_1 + \cdots + Y_n$ où $Y_1, \ldots, Y_n$ sont toutes des variables aléatoires admettant une espérance, alors

$$E(X) = \sum_{i=1}^{n} E(Y_i)$$

4. $m_r(X) = E(X^r)$ et $\mu_r(X) = E[(X - E(X))^r]$.

Chapitre 2

Codage et décodage

2.1 Théorie des codes

Le schéma classique est le suivant

Définition 2.1. 1. Soit A un ensemble de q symboles. On appelle code sur A tout sous ensemble C de l'ensemble des suites finies d'éléments de A. Les éléments de C s'appellent les mots de code. On dit que C est un code q-aire.

2. On appelle longueur d'un mot de code le nombre de symboles qui y figurent. Chaque symbole étant compté autant de fois qu'il apparaît.

3. Si tous les mots d'un code C sont de longueur n, on dit que C est un code bloc de longueur n, ou encore code de longueur n.

Remarques. 1. Un code 2-aire se dit un code binaire.

2. La plupart du temps, si q est la puissance d'un nombre premier, on choisit comme alphabet l'unique corps commutatif de cardinal q (à un isomorphisme près), F_q. Un code q-aire de longueur n est alors simplement une partie de $(\mathrm{F}_q)^n$ sur (F_q).

3. Nous nous intéressons presque exclusivement aux codes binaires.

Exemples. 1. L'ensemble du vocabulaire français (écrit en majuscule) est un code 27-aire $\mathscr{A} = \{\mathrm{A}, \mathrm{B}, \ldots, \mathrm{Z}, -\}$.

2. Vous voulez aider un ami à vous rejoindre à travers un champ de mines. Vous communiquez avec un appareil capable d'émettre deux sons, un aigu représenté par $\not{1}$ et un grave représenté par $\not{0}$. Vous n'avez besoin que de quatre informations de base : N ($\uparrow$), S ($\downarrow$), E ($\rightarrow$), O ($\leftarrow$). On peut coder ces informations de plusieurs façons :

– C_1 : code de longueur 2 : N : $\not{0}\ \not{0}$, S : $\not{1}\ \not{0}$, O : $\not{0}\ \not{1}$, E : $\not{1}\ \not{1}$.
– C_2 : code de longueur 3 : N : $\not{0}\ \not{0}\ \not{0}$, S : $\not{1}\ \not{1}\ \not{0}$, O : $\not{0}\ \not{1}\ \not{1}$, E : $\not{1}\ \not{0}\ \not{1}$.
– C_3 : code de longueur 6 : N : $\not{0}\ \not{0}\ \not{0}\ \not{0}\ \not{0}\ \not{0}$, S : $\not{1}\ \not{1}\ \not{1}\ \not{0}\ \not{0}\ \not{0}$, O : $\not{0}\ \not{0}\ \not{1}\ \not{1}\ \not{1}\ \not{0}$, E : $\not{1}\ \not{1}\ \not{0}\ \not{0}\ \not{1}\ \not{1}$.

On souhaite bien sûr transmettre : $\leftarrow \leftarrow \uparrow \uparrow \rightarrow \rightarrow \uparrow \leftarrow \leftarrow \leftarrow \leftarrow \uparrow \rightarrow \rightarrow \rightarrow \rightarrow \uparrow \leftarrow \uparrow \uparrow \rightarrow \rightarrow \rightarrow \downarrow \downarrow \downarrow \downarrow \rightarrow \downarrow \downarrow \leftarrow \leftarrow \uparrow \uparrow \uparrow \uparrow \uparrow \uparrow \uparrow$. C_1 est très économique mais si on fait une erreur, ou s'il y a des parasites, notre ami risque de sauter. C_3 est plus coûteux mais plus sûr. S'il y a une erreur on peut la détecter et même la corriger. C_2 est plus économique, mais s'il y a une erreur on peut la détecter sans toutefois pouvoir la corriger. Ce type de code est valable si on peut demander confirmation à l'expéditeur.

Nous venons de voir ici les principaux problèmes de la théorie des codes. Essayons de les formaliser en utilisant principalement des codes binaires dont les symboles 0 et 1 de l'alphabet sont appelés *bits* (abréviation de « *binary digits* »).

Définition 2.2. 1. On appelle canal binaire tout dispositif permettant de transmettre des bits.

2. La probabilité qu'un bit soit mal transmis s'appelle la probabilité d'erreur du canal pour ce bit.
3. Si la probabilité d'erreur du 0 et du 1 sont les mêmes, on dit que le canal est symétrique.
4. Si la transmission de chaque bit est indépendante des autre, on dit que le canal est sans mémoire.

Remarques. 1. En général, la probabilité p de mauvaise transmission d'un bit est très petite (sinon on a un très mauvais canal). Mais même si le canal est bon, si on suppose qu'on émet un bit toutes les microsecondes, cela fait 3600000 par heure et sur ce laps de temps, le nombre de bits mal transmis n'est pas négligeable.

2. En général, les perturbations sont dues à des phénomènes électriques (réseau, orage) ou électroniques (proximité d'appareils envoyant eux-mêmes des signaux et créant des interférences avec votre canal).

Théorème 2.1. *Lorsqu'on transmet n bits sur un canal sans mémoire symétrique, si p est la probabilité d'erreur du canal et $q = 1 - p$, alors :*

1. *la probabilité que toutes les erreurs commises le soient sur u bits désignés à l'avance est $p^u q^{n-u}$,*
2. *la probabilité que le nombre de bits mal transmis soit r est $\binom{n}{r} p^r q^{n-r}$ (loi binomiale).*

Démonstration. 1. Pour tout $i \in \mathbb{N}_n^*$, soit A_i : « le i^{e} bit est bien transmis », et soit $K = \{i_1, \dots, i_u\}$ les u bits désignés à l'avance sur lesquels il y a une erreur de transmission. Alors l'évènement : « toutes les erreurs sont commises sur les u bits désignés à l'avance » peut s'écrire :

$$\left(\bigcap_{j=1}^{u} \overline{A_{i_j}}\right) \cap \left(\bigcap_{S \in \complement_{\mathbb{N}_n^*} K} A_s\right)$$

Le canal étant sans mémoire, tous les éléments de l'intersection sont indépendants d'où

$$\begin{aligned} P(E) &= \prod_{j=1}^{u} P(\overline{A_{i_j}}) \prod_{S \in \complement_{\mathbb{N}_n^*} K} P(A_S) \\ &= p^u (1-p)^{n-u} \\ &= p^u q^{n-u} \end{aligned}$$

2. Soit F l'évènement « le nombre de bits mal transmis est r ». Soit F_1 : « les r premiers bits ont été mal transmis et les autres ont été bien transmis ». D'après 1, $P(F_1) = p^r q^{n-r}$. Il en sera de même pour tout autre choix de r bits parmi les n. Soient $F_1, \dots, F_l$ les évènements distincts correspondant aux divers choix des r éléments mal transmis parmi les n. Il y en a autant que de parties à r éléments parmi les n, soit $\binom{n}{r}$. Donc $l = \binom{n}{r}$ et $F = \bigcup_{i=1}^{l} F_i$. Les $(F_i)_{i \in \mathbb{N}_l^*}$ sont disjoints (incompatibles), et on a

$$\begin{aligned} P(F) &= P(\bigcup_{i=1}^{l} F_i) \\ &= \sum_{i=1}^{l} P(F_i) \\ &= l p^r q^{n-r} \\ \binom{n}{r} p^r q^{n-r} & \end{aligned}$$

C'est une loi binomiale.

□

Exemple. Si $p = 0,01$ et qu'on envoie un message de 3 bits. La probabilité pour qu'il y ait

– 0 erreur est $0,97$.

- 1 erreur est $0,0294$.
- 2 erreurs est $0,000297$.
- 3 erreurs est $0,000001$.

Définition 2.3. Soit F_q l'unique corps commutatif à q éléments. Si x et y sont deux mots de $(\mathrm{F}_q)^n$, on appelle distance de Hamming entre x et y, et on note $\delta(x,y)$, le nombre de composantes sur lesquelles x et y diffèrent.

Exemple. Sur $(\frac{\mathbb{Z}}{2\mathbb{Z}})^6$, si $x = 110110$ et $y = 100101$, alors $\delta(x,y) = 3$.

Remarques.

1. Si on note x le message émis, et x_r le message reçu, $\delta(x,x_r)$ représente le nombre de bits mal transmis. On a vu (2.1) que
$$\mathrm{P}[\delta(x,x_r) = k] = \binom{n}{k} p^k n^{n-k}$$
si le message comprend n bits.
2. δ possède toutes les propriétés des distances : pour tous x,y,z de $(\mathrm{F}_q)^n$.
   (a) $\delta(x,y) = 0 \iff x = y$
   (b) $\delta(x,y) = \delta(y,x)$
   (c) $\delta(x,z) \leqslant \delta(x,y) + \delta(y,z)$
3. On peut interpréter géométriquement la distance de Hamming. Si on représente $(\mathrm{F}_q)^n$ par son diagramme de Hasse, la distance de Hamming entre x et y est le nombre d'arêtes qu'il faut longer pour se rendre de x à y, par le chemin le plus court.
4. Lorsqu'un receveur reçoit un élément de $(\mathrm{F}_q)^n$ qui n'est pas un mot de code, il sait qu'une ou plusieurs erreurs ont été commises, mais il ne connaît pas ce nombre, ni les endroits précis où elles ont eu lieu. En général, il applique le principe du plus proche voisin, appelé aussi principe du maximum de vraisemblance. Si $r < s$, il est plus vraisemblable que r erreurs aient été commises plutôt que s.

Définition 2.4.

1. Soit C un code. On dit que C est un code k-détecteur s'il permet de détecter toutes les combinaisons de k (ou moins) erreurs. On dit que C est un code k-correcteur s'il permet de corriger toutes les combinaisons de k erreurs (ou moins).
2. Soit C un code de longueur n. On appelle distance minimum du code C, la distance de Hamming minimum entre tous les couples possibles de mots distincts de C. On la note $d(\mathrm{C})$.

Théorème 2.2. *1. Un code* C *est k-détecteur si et seulement si $d(\mathrm{C}) \geqslant k+1$.*

2. *Un code* C *est k-correcteur si et seulement si* $d(\mathrm{C}) \geqslant 2k+1$.

Démonstration. 1. Soient x et y deux mots de code du code C tels que $\delta(x,y) = d(\mathrm{C})$. L'erreur qui consiste à remplacer x par y ne peut être détectée. Il est donc nécessaire que $d(\mathrm{C}) > k$.

Réciproquement, supposons que l'on expédie le mot x et que l'on reçoive un mot y avec un nombre d'erreurs strictement inférieur à $d(\mathrm{C})$. Si cet y faux n'était pas détecté, c'est que y serait un mot de code. Ceci contredirait le caractère minimal de $d(\mathrm{C})$. Donc y ne peut être un mot de code et le receveur saura qu'il y a eu des erreurs.

2. Soient x le mot envoyé et y le mot reçu. Supposons que $\delta(x,y) \leqslant \frac{d(\mathrm{C})-1}{2}$, c'est à dire $d(\mathrm{C}) \geqslant 2\delta(x,y)+1$. Si z est un mot de code tel que $\delta(y,z) = \delta(x,y)$ alors l'inégalité triangulaire nous indique que

$$\begin{aligned}\delta(x,z) &\leqslant \delta(x,y) + \delta(y,z)\\ &\leqslant 2\delta(x,y)\\ &\leqslant d(\mathrm{C}) - 1\end{aligned}$$

D'où $\delta(x,z) < d(\mathrm{C})$, et par définition de $d(\mathrm{C})$, $\delta(x,z) = 0$. On a alors $z = x$ et y est le mot de code qui approche le mieux x. Le receveur remplace y par x et le message est décodé.

Réciproquement, soit m un entier vérifiant $\frac{d(\mathrm{C})-1}{2} < m \leqslant d(\mathrm{C})$ et soient x et y deux mots de code tels que $d(\mathrm{C}) = \delta(x,y)$. Supposons que l'on ait reçu z quand x a été envoyé, et qu'il se déduise de x en modifiant m des bits sur lesquels x et y diffèrent. Puisque $\delta(x,z) = m$, $\delta(y,z) = d(\mathrm{C}) - m$ et le mot de code y est plus près de z que de x. Le receveur se trompe alors en corrigeant. Donc si $m > \frac{d(\mathrm{C})-1}{2}$ il existe des messages mal corrigés comportant m erreurs.

□

Exemple. Soit $\mathrm{C} = \{00000, 11111\}$ un code sur B^5 (où $\mathrm{B} = \{0,1\}$. $d(\mathrm{C}) = 5$ et $\frac{d(\mathrm{C})-1}{2} = 2$. Si la transmission de x est entachée de 2 erreurs ($2 \leqslant 2$), par exemple $x = 00000$ devient 10010, le receveur corrigera en 00000, le mot de code le plus proche. S'il y a 3 erreurs ($3 > 2$), par exemple 10101, le receveur se trompera en corrigeant en 11111.

Définition 2.5. Soit C un code et $d(\mathrm{C})$ sa distance minimum. On appelle nombre d'erreurs corrigées de C, la partie entière de $\frac{d(\mathrm{C})-1}{2}$, c'est à dire le plus grand entier immédiatement inférieur à $\frac{d(\mathrm{C})-1}{2}$. On appelle nombre d'erreurs

détectées de C le nombre $d(\mathrm{C}) - 1$. On peut constituer un tableau.

$d(\mathrm{C})$	nombre d'erreurs détectées	nombre d'erreurs corrigées
1	0	0
2	1	0
3	2	1
4	3	1
5	4	2
6	5	2
7	6	3
8	7	3
9	8	4
10	9	4

Notation. On note (n, M, d)-code, tout code de longueur n contenant M mots de code et dont la distance minimale est d.

Exemples. 1. Dans l'exemple qui suit la définition 2.1,
- C_1 est un $(2, 4, 1)$-code.
- C_2 est un $(3, 4, 2)$-code.
- $\mathrm{D} = \{00000, 01101, 10110, 11011\}$ est un $(5, 4, 3)$-code.

2. Le code à répétition q-aire de longueur n est le code dont les mots sont

$$\begin{array}{rl} 000\ldots00 & n \text{ fois} \\ 111\ldots11 & n \text{ fois} \\ \vdots\ldots\vdots & \\ q-1q-1q-1\ldots q-1q-1 & n \text{ fois} \end{array}$$

C'est un (n, q, n)-code.

Remarque. Un bon (n, M, d)-code a un petit n (pour une transmission rapide des messages), un grand M (pour une grande variété de messages) et un grand d (pour corriger beaucoup d'erreurs).

Ces conditions se combattent entre elles et nous font découvrir le « principal problème de la théorie des codes », qui concerne l'optimisation de l'un des paramètres n, M, ou d, pour des valeurs fixées des deux autres.

La version la plus courante consiste à trouver le plus grand code (M) de longueur donnée ayant une distance minimum fixée.

Théorème 2.3. *Si on note par* $\mathrm{A}_q(n, d)$ *la plus grande valeur de* M *telle qu'il existe un* (n, M, d)*-code q-aire,*

1. $A_q(n,1) = q^n$
2. $A_q(n,n) = q$

Démonstration. 1. Si la distance minimale d'un code est 1, la seule exigence possible au niveau des mots de code est qu'ils soient distincts. Avec q symboles on peut écrire q^n mots distincts de longueur n.

2. Soit c un (n, M, n)-code q-aire. Alors deux mots de codes distincts diffèrent en n positions. C'est à dire que pour chaque position fixée, la première par exemple, des deux mots de code, les symboles sont différents. On a q symboles donc $M \leqslant q$. Donc $A_q(n,n) \leqslant q$. Or le code à répétition q-aire de longueur n, (voir ci-dessus) est un (n,q,n)-code q-aire. Donc $A_q(n,n) = q$.

□

Remarque. Essayons de calculer $A_2(5,3)$. Le code C_3 dans l'exemple qui suit la définition 2.1 est un $(5,4,3)$-code binaire, donc $A_2(5,3) \geqslant 4$. Si on veut vérifier qu'il existe des $(5,5,3)$-codes binaires, la méthode la plus bestiale consiste à examiner tous les sous ensembles de cardinal 5 de $\{0,1\}^5$. Or, il y en a 201376. Nous allons introduire une notion d'équivalence qui va permettre de réduire considérablement les recherches.

Définition 2.6. Deux codes q-aires sont dit équivalents si l'on peut obtenir l'un à partir de l'autre en combinant les opérations des types suivants :

1. Permutation des positions du code.
2. Permutation des symboles apparaissant dans une position donnée.

Si on représente un code par une matrice de M lignes et n colonnes dont les lignes sont les mots de code, une opération de type 1 est une permutation des colonnes et une opération de type 2 est une réaffectation des symboles figurant dans une colonne donnée.

Exemples. 1.

$$C = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix} \qquad D = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \end{pmatrix}$$

C équivaut à D par les opérations suivantes
– on échange les positions 2 et 4 de C (type 1).

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

– on permute 0 et 1 dans la colonne 3 (type 2).

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \end{pmatrix}$$

C est un ensemble de mots, donc l'ordre dans lequel on les écrit importe peu.

2. Le code ternaire $\mathrm{C} = \left(\begin{smallmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{smallmatrix}\right)$ est équivalent au code ternaire à répétition de longueur 3 $\left(\begin{smallmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 2 & 2 & 2 \end{smallmatrix}\right)$. On applique (021) à la position 2 puis (012) à la partition 3.

Remarque. Les distances entre les mots de code ne sont pas changées par les opérations de type 1 et 2, donc les codes obtenus auront les mêmes paramètres et corrigeront le même nombre d'erreurs.

Théorème 2.4. *Tout (n, M, d)-code q-aire sur un alphabet $\{0, 1, \ldots, q-1\}$ est équivalent à un (n, M, d)-code q-aire qui contient le mot $00\ldots0$.*

Démonstration. On choisit au hasard l'un des mots du code $x_1 x_2 \ldots x_n$ et chaque fois que $x_i \neq 0$, on applique $(x_i 0)$ en position i. □

Remarque. Terminons à présent le calcul de $\mathrm{A}_2(5, 3)$. Considérons un $(5, \mathrm{M}, 3)$-code C où $\mathrm{M} \geqslant 4$. Alors d'après le théorème 2.4, on peut supposer que C contient le mot 00000. Alors C contient au plus un mot ayant 4 ou 5 « 1 », sinon s'il existait deux tels mots x et y ils auraient au moins 3 « 1 » en position commune et $\delta(x, y) \leqslant 2$ contredisant $d(\mathrm{C}) = 3$.

Puisque 00000 appartient à C, on ne peut avoir de mots de code contenant exactement 1 ou 2 « 1 », et puisque $\mathrm{M} \geqslant 4$, il existe au moins deux mots de code qui ont exactement 3 « 1 ». En changeant les positions si nécessaire, on peut affirmer que C contient les mots de code 00000, 11100, 00111.

On montre facilement par « tâtonnement » que le seul autre mot de code possible est 11011. Donc $\mathrm{A}_2(5, 3) = 4$.

Si on se restreint aux codes binaires, on peut établir une table pour $n \leqslant 16$ et $d \leqslant 7$ grâce à Sloane et Mac Williams (1977) :

n	d = 3	d = 5	d = 7
5	4	2	-
6	8	2	-
7	16	2	2
8	20	4	2
9	40	6	2
10	72–79	12	2
11	144–158	24	4
12	256	32	4
13	512	64	8
14	1024	128	16
15	2048	256	32
16	2560–3276	256–340	36–37

Lecture de la table : $A_2(5,3) = 4$, $72 \leqslant A_2(10,3) \leqslant 79$.

On peut remarquer que la table ne donne que les valeurs impaires de d. Si d est un nombre pair, $A_2(n,d) = A_2(n-1,d-1)$.

Définition 2.7. Pour tout vecteur u de $(F_q)^n$ et tout entier $r \geqslant 0$, la sphère de rayon r et de centre u notée $S(u,r)$ est l'ensemble $\{v \in (F_q)^n / \delta(u,v) \leqslant r\}$.

Théorème 2.5. *Toute sphère de rayon r dans $(F_q)^n$, où $0 \leqslant r \leqslant n$, contient exactement*

$$\binom{n}{0} + \binom{n}{1}(q-1) + \binom{n}{1}(q-1)^2 + \cdots + \binom{n}{r}(q-1)^r$$

vecteurs.

Démonstration. Soit u un vecteur fixé de $(F_q)^n$. Cherchons combien de vecteurs v sont exactement à la distance m de u ($m \leqslant n$). On peut choisir les m positions où v va différer avec u de $\binom{n}{m}$ façons différentes.

Pour chacune des positions choisies, le symbole différent de celui de u peut être choisi de $(q-1)$ façons différentes. Donc les m symboles différents de ceux de u, pour chaque position, peuvent être choisis de $(q-1)^m$ façons différentes.

Le nombre de vecteurs v exactement à la distance m de u est donc $\binom{n}{m}(q-1)^m$.

Il est clair que le nombre total de vecteurs dans une sphère de rayon r centrée sur u est égal au nombre de vecteurs à la distance 0, augmenté du nombre de vecteurs à la distance 1, ..., finalement augmenté du nombre de vecteurs à la distance r :

$$\sum_{i=1}^{r} \binom{n}{i}(q-1)^i$$

□

Théorème 2.6 (Théorème de la borne de Hamming)**.** *Tout* $(n, \mathrm{M}, 2t+1)$*-code q-aire vérifie*

$$\mathrm{M}\left[\binom{n}{0} + \binom{n}{1}(q-1) + \cdots + \binom{n}{t}(q-1)^t\right] \leqslant q^n$$

Démonstration. Soit C un $(n, \mathrm{M}, 2t+1)$-code q-aire. Alors deux sphères de rayon t centrées sur des mots de code distincts n'ont aucun vecteur de $(\mathrm{F}_q)^n$ en commun.

Le membre de gauche de l'inégalité représente le nombre total de vecteurs dans les M sphères de rayon t centrées sur les mots de code distincts et il est clair que ce nombre est plus petit que le nombre de mots dans $(\mathrm{F}_q)^n$: q^n. □

Remarques. 1. Pour les codes binaires,

$$\mathrm{M}\left[1 + \binom{n}{1} + \cdots + \binom{n}{t}\right] \leqslant 2^n$$

2. Pour des valeurs données de q, n et d, le théorème de la borne de Hamming donne une borne supérieure à $\mathrm{A}_q(n,d)$. Par exemple un $(5, \mathrm{M}, 3)$-code binaire doit satisfaire à $\mathrm{M}[1+5] \leqslant 32$. Donc $\mathrm{A}_2(5,3) \leqslant 5$.

   Attention : cela ne signifie pas qu'il existe un tel code où $n = 5$, $\mathrm{M} = 5$ et $d = 3$. Cela fournit juste une limite au delà de laquelle il ne faut pas chercher.

Définition 2.8. Si n et M sont fixés, on appelle borne de Hamming la borne supérieure fournie par l'inégalité de Hamming pour d.

Exemples. 1. Codage par bit de parité. On transmet des bits par paquets de 3. À la fin de chaque paquet, on rajoute un quatrième bit appelé bit de parité de façon à ce que le nombre de bits non nuls dans le paquet de 4 soit toujours pair.

   On a alors un $(4, 2^3, d)$-code. On a

$$2^3\left[\binom{4}{0} + \binom{4}{1} + \cdots + \binom{4}{t}\right] \leqslant 2^4$$

   donc $2^3\left[\binom{4}{0}\right] = 8 \leqslant 16$ et $2^3\left[\binom{4}{0} + \binom{4}{1}\right] = 40 > 16$ d'où $t = 0$ et $d = 1$.

2. Codage par répétition. On veut transmettre 0 ou 1 : on transmet 000 ou 111. C'est un $(3,2,3)$-code. L'inégalité de Hamming s'écrit

$$2\left[\binom{3}{0} + \binom{3}{1}\right] \leqslant 2^3$$
$$8 \leqslant 8$$

C'est une égalité. C'est vrai dans le cas des codes à répétitions de longueur n.

Définition 2.9. Si pour un (n, M, d)-code C, l'inégalité de Hamming est une égalité, on dit que le code est parfait.

Remarques.
1. Dans un code parfait $(n, \mathrm{M}, 2t+1)$ toute série de t erreurs sur un mot sera détectée et corrigée avec efficacité car tous les mots possibles de longueur n seront positionnés dans des sphères qui ne contiendront qu'un seul mot de code.
2. On montre (en M1 maths) que si q est un nombre premier ou une puissance d'un nombre premier, il existe un unique corps commutatif de cardinal q. On l'appelle corps de Galois de cardinal q. On le note F_q (*field*), ou $\mathrm{GF}(q)$. Ceci va nous permettre d'interpréter les suites de n éléments de ce corps comme les vecteurs d'un espace vectoriel de dimension n sur ce corps. Nous produirons alors des codes qui seront des sous espaces vectoriels que nous appellerons « codes linéaires ».

Définition 2.10. On considère F_q, où q est une puissance d'un nombre premier, comme alphabet. L'ensemble des suites de n éléments de F_q, $n \in \mathbb{N}^*$, $(\mathrm{F}_q)^n$, peut être muni d'une structure d'espace vectoriel de façon canonique. On appellera code linéaire sur F_q tout sous espace vectoriel de $(\mathrm{F}_q)^n$, c'est à dire toute partie non vide C de $(\mathrm{F}_q)^n$ telle que :

$$\begin{aligned} \forall u, v \in \mathrm{C} \quad & u + v \in \mathrm{C} \\ \forall u \in \mathrm{C}, \forall a \in \mathrm{F}_q \quad & au \in \mathrm{C} \end{aligned}$$

Notations. Au lieu de noter $(x_1, \ldots, x_n)$ un élément de $(\mathrm{F}_q)^n$, nous le noterons $x_1 x_2 \ldots x_n$.

Si C est un sous espace vectoriel de dimension k sur F_q, nous dirons que C est un $[n, k]$-code, ou encore que C est un $[n, k, d]$-code si on veut préciser sa distance minimale.

Remarques.
1. Un $[n, k, d]$-code q-aire est aussi un (n, q^k, d)-code q-aire, mais bien sûr la réciproque est fausse.
2. Le mot constitué uniquement de « 0 » figure dans tout code linéaire.
3. Dans certains ouvrages on emploie « code groupe » à la place de « code linéaire ».

Définition 2.11. Le poids $w(x)$ d'un vecteur x de $(\mathrm{F}_q)^n$ est le nombre de composantes non nulles de x.

Théorème 2.7. *Si x et y sont des mots dans $(\mathrm{F}_q)^n$, $\delta(x, y) = w(x - y)$.*

Démonstration. Le vecteur $(x-y)$ a des composantes non nulles uniquement là où x et y diffèrent. □

Remarque. Pour $q=2$, il est clair que $x-y=x+y$ et le théorème peut s'écrire $\delta(x,y)=w(x+y)$.

Théorème 2.8. *Soit* C *un code linéaire et* $w(\mathrm{C})$ *le plus petit poids atteint par un vecteur non nul de* C*. Alors*

$$d(\mathrm{C})=w(\mathrm{C})$$

Démonstration. Soient x et y deux mots de code tels que $\delta(x,y)=d(\mathrm{C})$. Alors d'après **??**,

$$d(\mathrm{C})=w(x-y)\geqslant w(\mathrm{C})$$

car C étant linéaire $x-y\in\mathrm{C}$. D'autre part il existe $z\in\mathrm{C}$ tel que

$$w(\mathrm{C})=w(z)=\delta(z,0)\geqslant d(\mathrm{C})$$

puisque $0\in\mathrm{C}$. D'où $d(\mathrm{C})=w(\mathrm{C})$. □

Remarques.
1. En général, quand on veut déterminer la distance minimale d'un code ayant M mots de code, il faut faire $\frac{\mathrm{M}(\mathrm{M}-1)}{2}$ comparaisons. Avec un code linéaire, $\mathrm{M}-1$ suffisent : on compare chaque mot à $00\ldots0$.
2. En général, pour indiquer un code, il faut donner la liste de tous les mots de code. Avec un code linéaire, l'énoncé d'une base suffit.

Définition 2.12. Une matrice à k lignes et n colonnes dont les lignes forment une base d'un $[n,k]$-code linéaire est appelée matrice génératrice de ce code.

Exemples.
1. Le code C_2 de l'exemple exemple 2 suivant la définition **??** est un $[3,2,2]$-code linéaire qui possède $\begin{smallmatrix}0&1&1\\1&0&1\end{smallmatrix}$ comme matrice génératrice.
2. Le code q-aire à répétition de longueur n sur F_q est $[n,1,n]$-code ayant $1\ 1\ \ldots\ 1$ pour matrice génératrice.

Remarques.
1. Les codes q-aire linéaires ne sont pas définis dans le cas où q n'est pas une puissance d'un nombre premier p. On peut cependant définir des codes q-aires restrictions de codes linéaires sur un alphabet plus grand (ex : le code ISBN est constitué de mots faits avec 10 symboles qui sont des mots faits avec 11 symboles dont un est absent).
2. Pour les codes linéaires, on peut affiner la relation d'équivalence de la définition **??**.

Définition 2.13. Deux codes linéaires sur F_q seront dits équivalents si on peut obtenir l'un à partir de l'autre en combinant les transformations de types suivants :

A Permutation des positions du code.

B Multiplication des symboles apparaissant en position fixe par un scalaire non nul.

Théorème 2.9. *Deux matrices* $(k \times n)$ *(k lignes et n colonnes) engendrent deux* $[n,k]$*-codes équivalents sur* F_q *si l'une des matrices peut être obtenue à partir de l'autre grâce aux opérations suivantes :*

L1 Permutation des lignes.

L2 Multiplication d'une ligne par un scalaire non nul.

L3 Addition à une ligne du produit d'une autre par un scalaire.

C1 Permutation des colonnes.

C2 Multiplication d'une colonne par un scalaire non nul.

Démonstration. (L1), (L2), (L3) sont des transformations qui ne changent pas le rang de la matrice donc qui préservent l'indépendance des vecteurs lignes de la matrice génératrice et donc transforme une base du code en une autre base du code.

(C1) et (C2) correspondent aux transformations de la définition **??**. □

Théorème 2.10. *Soit* G *une matrice génératrice d'un code linéaire de type* $[n,k]$*. En utilisant les transformations (L1), (L2), (L3), (C1) et (C2), on peut mettre* G *sous la forme standard :*

$$\left(\mathrm{I}_k \vdots \mathrm{A}\right)$$

où I_k *est la matrice unité d'ordre k et* A *une matrice de type* $k \times (n-k)$*.*

Démonstration. On applique la procédure suivante, en trois étapes, de $j = 1$ à k, la j^{e} application transformant la colonne C_j en $\begin{smallmatrix}0\\0\\\vdots\\0\\1\\0\\\vdots\\0\end{smallmatrix}$, le 1 étant sur la j^{e} ligne. Les coefficients g_{ij} représentent les coefficients de G à un moment donné de la transformation et non les coefficients de G au départ.

Supposons que nous ayons mis G sous la forme :

$$\begin{pmatrix} 1 & 0 & 0\ldots & 0 & g_{1j} & \ldots & g_{1n} \\ 0 & 1 & 0\ldots & 0 & g_{2j} & \ldots & g_{2n} \\ \vdots & & & & & & \\ 0 & 0 & 0\ldots & 1 & g_{j-1,j} & \ldots & g_{j-1,n} \\ 0 & 0 & 0\ldots & 0 & g_{jj} & \ldots & g_{jn} \\ \vdots & & & & & & \\ 0 & 0 & 0\ldots & 0 & g_{kj} & \ldots & g_{kn} \end{pmatrix}$$

Étape 1 – si $g_{jj} \neq 0$, passer à l'étape 2 ;
– si $g_{jj} = 0$ et si pour $i > j$, $g_{ij} \neq 0$ permuter L_j et L_i ;
– si $g_{jj} = 0$ et si pour tout $i > j$, $g_{ij} = 0$, alors choisissez h tel que $g_{ih} \neq 0$ et permuter C_j et C_h.

Étape 2 comme $g_{jj} \neq 0$, multipliez L_j par g_{jj}^{-1}.

Étape 3 comme $g_{jj} = 1$, pour tout $i \in \mathbb{N}_k^*$ tel que $i \neq j$, remplacez L_i par $L_i - g_{ij}L_j$. La colonne C_j a alors l'allure voulue.

□

Remarques. 1. Si on peut transformer G en une matrice standard G′ en jouant uniquement sur les lignes (ce qui sera le cas si et seulement si les k premières colonnes de G sont linéairement indépendantes) alors G et G′ engendreront le même code. Mais si on utilise (C1) et (C2), G′ engendrera un code équivalent mais pas nécessairement le même.

2. Dans la pratique, c'est souvent notre vue qui guide les opérations à effectuer pour la mise sous forme standard (voir les exemples ci-dessous).

3. La forme standard $(I_k \vdots A)$ d'une matrice génératrice n'est pas unique. Des permutations des colonnes de A donneront un code équivalent.

Exemples. 1. Dans l'exemple 1 suivant la définition **??**, le code C_2 a pour matrice génératrice $\begin{smallmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \end{smallmatrix}$. On la met facilement sous forme standard en permutant les deux lignes $\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{smallmatrix}$.

2. Soit C un code ayant pour matrice génératrice

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow[L_3 \to L_3 - L_1]{L_2 \to L_2 - L_1} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow[L_4 \to L_4 - L_2]{L_1 \to L_1 - L_2} \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

$$\xrightarrow{L_2 \to L_2 - L_3} \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

$$\xrightarrow{L_3 \to L_3 - L_4} \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

3. Considérons le $[6,3]$-code sur F_3 ayant pour matrice génératrice

$$G = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 2 \\ 1 & 0 & 2 & 0 & 1 & 1 \end{pmatrix} \xrightarrow{\text{en permutant } L_1 \text{ et } L_3} \begin{pmatrix} 1 & 0 & 2 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

$$\xrightarrow{\text{en permutant } C_3 \text{ et } C_4} \begin{pmatrix} 1 & 0 & 0 & 2 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

Définition 2.14. Soit C un $[n,k]$-code sur F_q, de matrice génératrice G. Par définition, C possède q^k mots de code et on peut envoyer q^k messages différents, que l'on identifie aux vecteurs de $(F_q)^k$. On appellera codage d'un message $u = u_1 u_2 \dots u_k$ le produit matriciel ${}^tUG$ où U est le vecteur colonne $\begin{matrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{matrix}$.

L'application $U \to {}^tUG$ de $(F_q)^k$ dans un sous espace de dimension k de $(F_q)^n$ s'appelle la fonction de codage.

Remarques. 1. Si $L_1, \dots, L_k$ sont les lignes de G,

$${}^tUG = \sum_{i=1}^{k} u_i L_i$$

2. La fonction de codage est une application linéaire injective de $(\mathrm{F}_q)^k$ dans $(\mathrm{F}_q)^n$.

3. Si G est sous forme standard $(\mathrm{I}_k \vdots \mathrm{A})$ où $\mathrm{A} = (a_{ij})$ est une matrice $(k \times (n-k))$, alors un message $u = u_1 \dots u_k$ est codé $x = {}^t\mathrm{UG} = u_1 \dots u_k x_{k+1} \dots x_n$. On appelle $u_1 \dots u_k$ les chiffres (*digits*) du message et $x_{k+1} \dots x_n$ les chiffres (*digits*) de redondance. Ils représentent ce que l'on ajoute au message pour le protéger du « bruit ».

Exemple. Si C est le $[7,4]$-code binaire ayant $\mathrm{G} = \begin{smallmatrix} 1&0&0&0&1&0&1 \\ 0&1&0&0&1&1&1 \\ 0&0&1&0&1&1&0 \\ 0&0&0&1&0&1&1 \end{smallmatrix}$ pour matrice génératrice, un message $u = u_1u_2u_3u_4$ sera codé $(u_1, u_2, u_3, u_4, u_1+u_2+u_3, u_2+u_3+u_4, u_1+u_2+u_4)$.

Par exemple, 0000 sera codé 0000000 ; 1000 sera codé 1000101 et 1110 sera codé 1110100.

Définition 2.15. Supposons que l'on envoie le mot de code $x = x_1 \dots x_n$ et que le receveur obtienne $y = y_1 \dots y_n$. On appelle vecteur d'erreur le vecteur $e = y - x = e_1 \dots e_n$.

Remarque. Le problème du décodeur est de trouver à partir de quel mot de code le message y a été envoyé et perturbé, ou encore à quel vecteur d'erreur il a à faire. Le principe du « plus proche voisin » se traduit de façon élégante dans les codes linéaires. Cette méthode, qui concerne l'aspect « groupe » des codes linéaires, a été mise en évidence par Slepian en 1960.

Théorème 2.11. *La relation binaire* R *définie sur* $(\mathrm{F}_q)^n$ *par :*

$$a\mathrm{R}b \iff a - b \in \mathrm{C}$$

où C *est un* $[n,k]$*-code linéaire, est une relation d'équivalence sur* $(\mathrm{F}_q)^n$*. D'après Lagrange, toutes les classes ont même cardinal, le cardinal de* C (q^k)*.*

Démonstration. Voir le cours d'algèbre générale 2. □

Remarque. Toute classe par rapport à R est de la forme $a+\mathrm{C}$, où $a \in (\mathrm{F}_q)^n$. Tout vecteur de $(\mathrm{F}_q)^n$ appartient à une classe et une seule. Le nombre de classes est q^{n-k}.

Exemple. Soit C le $[4,2]$-code binaire de matrice génératrice $\mathrm{G} = \begin{smallmatrix} 1&0&1&1 \\ 0&1&0&1 \end{smallmatrix}$. Alors $\mathrm{C} = \{0000, 1011, 0101, 1110\}$. Les classes par rapport à C sont

$$\begin{aligned} 0000 + \mathrm{C} &= \mathrm{C} \\ 1000 + \mathrm{C} &= \{1000, 0011, 1101, 0110\} \\ 0100 + \mathrm{C} &= \{0100, 1111, 0001, 1010\} \\ 0010 + \mathrm{C} &= \{0010, 1001, 0111, 1100\} \end{aligned}$$

Définition 2.16. 1. Soit C un $[n,k]$-code. Dans chaque classe par rapport à C, on appelle leader de classe un vecteur ayant un poids minimal.

2. Un tableau standard de Slepian pour un $[n,k]$-code est un tableau $q^{n-k} \times q^k$ de tous les vecteurs de $(\mathrm{F}_q)^n$ dont le premier rang est formé des vecteurs de C, le 0 de C étant à l'extrême gauche, les autres rangs étant les $a_i + \mathrm{C}$ où l'ordre de la première ligne est respecté, chaque a_i étant un leader de classe.

Exemple. Reprenons le code C de l'exemple précédent. Le tableau standard de Slepian pour C est

0000	1011	0101	1110
1000	0011	1101	0110
0100	1111	0001	1010
0010	1001	0111	1100

Les leaders de classe sont dans la première colonne.

Pour décoder en utilisant le tableau, on procède de la façon suivante. Si un mot de code est reçu, pas de problème on l'accepte. Sinon, supposons que l'on reçoive 1111. On le cherche dans le tableau et on l'interprète comme 1011, qui est le mot de code en haut de la colonne de 1111. C'est à dire que l'on décide que le vecteur d'erreur est le leader de classe à l'extrémité de la ligne de 1111, et on interprète le message reçu comme la diférence entre ce message et son leader de classe.

Remarque. Dans la pratique, cette méthode est trop lente pour les gros codes et trop coûteuse en termes de stockage d'informations. On utilise alors une autre méthode, le décodage par syndrôme que nous allons décrire maintenant. Pour cela, nous allons munir $(\mathrm{F}_q)^n$ d'un produit scalaire, le produit scalaire canonique : si $u = u_1 \dots u_n$ et $v = v_1 \dots v_n$ sont dans $(\mathrm{F}_q)^n$,

$$u.v = u_1 v_1 + \cdots + u_n v_n$$

Par exemple, dans $(\mathrm{F}_2)^4$,

$$(1001)(1101) = 0$$
$$(1111)(1110) = 1$$

et dans $(\mathrm{F}_3)^4$,

$$(2011)(1210) = 0$$
$$(1212)(2121) = 2$$

Définition 2.17. Soit C un $[n,k]$-code linéaire. On appellera code dual de C, le code, noté $C^{\perp}$, orthogonal de C.

$$C^{\perp} = \{v \in (F_q)^n / \forall u \in C, v.u = 0\}$$

Théorème 2.12. *Soit* C *un* $[n,k]$*-code de matrice génératrice* G*. Alors un vecteur* $v \in (F_q)^n$ *appartient à* $C^{\perp}$ *si et seulement si* v *est orthogonal à tout rang de* G *:*

$$v \in C^{\perp} \iff v^t G = 0$$

Donc $C^{\perp}$ *est un* $[n, n-k]$*-code.*

Démonstration. Simple application du cours d'algèbre linéaire 2. □

Exemples.
1. Si $C = \{0000, 1100, 0011, 1111\}$, $C^{\perp} = C$.
2. Si $C = \{000, 110, 011, 101\}$, $C^{\perp} = \{000, 111\}$.
3. Pour tout $[n,k]$-code, $(C^{\perp})^{\perp} = C$.

Définition 2.18. On appelle matrice de contrôle de parité pour un $[n,k]$-code C une matrice génératrice de $C^{\perp}$.

Remarques.
1. Si H est une matrice de contrôle de parité pour un $[n,k]$-code C, c'est une matrice du type $(n-k) \times n$ qui vérifie $G^t H = 0$. Alors
$$C = \{x \in (F_q)^n / x^t H = 0\}$$
2. Tout code linéaire est parfaitement déterminé par sa matrice de contrôle de parité.

Exemple. Si on reprend les codes de l'exemple précédent, pour le 1), $\left(\begin{smallmatrix}1100\\0011\end{smallmatrix}\right)$ est à la fois une matrice génératrice et une matrice de contrôle de parité pour C.
Pour le 2), (111) est une matrice de contrôle de parité.

Remarque. Les rangées d'une matrice de contrôle de parité sont appelées tests de parité sur les mots de code. Ils disent que certaines combinaisons linéaires des coordonnées de tout mot de code sont nuls. Un code est totalement déterminé par sa matrice de contrôle de parité : si $H = \left(\begin{smallmatrix}1100\\0011\end{smallmatrix}\right)$ alors C est le code $\{(x_1,x_2,x_3,x_4) \in (F_2)^4 / x_1 + x_2 = 0, x_3 + x_4 = 0\}$. Les équations $x_1 + x_2 = 0$ et $x_3 + x_4 = 0$ sont appellées équations de contrôle de parité.

Théorème 2.13. *Si* $G = (I_k \vdots A)$ *est une matrice génératrice sous forme standard d'un* $[n,k]$*-code* C*, alors* $H = (-{}^tA \vdots I_{n-k})$ *est une matrice de contrôle de parité pour* C*.*

Démonstration. Supposons

$$G = \begin{pmatrix} 1 & & 0 & \vdots & a_{11} & \dots & a_{1n-k} \\ & \ddots & & \vdots & \vdots & & \vdots \\ 0 & & 1 & \vdots & a_{k1} & \dots & a_{kn-k} \end{pmatrix}$$

et soit

$$H = \begin{pmatrix} -a_{11} & \dots & -a_{1n-k} & \vdots & 1 & & 0 \\ \vdots & & \vdots & \vdots & & \ddots & \\ -a_{k1} & \dots & -a_{kn-k} & \vdots & 0 & & 1 \end{pmatrix}$$

H a la bonne taille pour être une matrice de contrôle de C et ses lignes sont linéairement indépendantes. Il suffit donc de vérifier que chacune de ses lignes est orthogonale à chacune des lignes de G. Or le produit scalaire de la i^{e} ligne de G par la j^{e} ligne de H est :

$$0 + \cdots + 0 + (-a_{ij}) + 0 + \cdots + 0 + a_{ij} + 0 + \cdots + 0 = 0$$

□

Définition 2.19. Une matrice de contrôle de parité est dite en forme standard si $H = (B \vdots I_{n-k})$.

Remarques.

1. Si $H = (B \vdots I_{n-k})$ est une matrice de contrôle de parité d'un code linéaire C, le théoréme **??** nous dit que $G = (I_k \vdots {}^tB)$ est une matrice génératrice de C.
2. Si H est une matrice de contrôle de parité qui n'est pas en forme standard, on peut la mettre sous cette forme avec la même méthode que celle utilisée pour les matrices génératrices.

Définition 2.20. Soit H une matrice de contrôle de parité d'un $[n,k]$-code C. Si y est un vecteur quelconque de $(F_q)^n$, alors le vecteur uniligne

$$s(y) = y\,{}^tH$$

est appelé le syndrôme de y.

Remarques.

1. Si les lignes de H sont $h_1, \dots, h_{n-k}$ alors

$$s(y) = (yh_1, \dots, yh_{n-k})$$

2. $s(y) = 0 \iff y \in \mathrm{C}$.
3. Dans certains ouvrages, le syndrôme de y est défini comme le transposé du $s(y)$ que nous avons défini. C'est alors un vecteur unicolonne.

Théorème 2.14. *Dans* $(\mathrm{F}_q)^n$ *deux vecteurs* u *et* v *sont dans la même classe d'équivalence par rapport au code linéaire* C *si et seulement si ils ont même syndrôme.*

Démonstration. u et v sont dans la même classe par rapport à C équivant à $u + \mathrm{C} = v + \mathrm{C}$, ou encore $u - v \in \mathrm{C}$. Alors

$$\begin{aligned} (u-v)^t\mathrm{H} &= 0 \\ u^t\mathrm{H} - v^t\mathrm{H} &= 0 \\ u^t\mathrm{H} &= v^t\mathrm{H} \\ s(u) &= s(v) \end{aligned}$$

□

Corollaire 2.1. *Il existe une bijection entre les syndrômes et les classes d'un code linéaire* C.

Voyons comment utiliser cette propriété pour le décodage. Soit $\mathrm{G} = \begin{smallmatrix}1011\\0101\end{smallmatrix}$ la matrice génératrice sous forme standard d'un code linéaire C. Alors $\mathrm{H} = \begin{smallmatrix}1010\\1101\end{smallmatrix}$ est une matrice de contrôle de parité sous forme standard de C.

Les syndrômes des leaders de classe sont

$$s(0000) = 00, s(1000) = 11, s(0100) = 01, s(0010) = 10$$

Le tableau standard peut être résumé en un tableau réduit à deux colonnes :

syndrôme z	leader de classe de z
00	0000
11	1000
01	0100
10	0010

1. Lorsqu'on reçoit un message y, on calcule son syndrôme $s(y) = y^t\mathrm{H}$.
2. On repère $s(y) = z$ dans la première colonne du tableau réduit.
3. On décode le message y comme étant $y-$ le leader de classe correspondant.

Par exemple, on reçoit $1111 = y$. Alors $s(y) = 01$ et on décode y par $1111 - 0100 = 1011$.