Théorie et codage de l'information

Paul HONEINE
— H201 —

en collaboration avec Cédric Richard
(Université de Nice Sophia Antipolis)

Université de technologie de Troyes

– Automne 2014 –

Rappels de probabilités

Chapitre -1

Le modèle probabiliste

Espace probabilisable

Vocabulaire probabiliste.

- *Expérience aléatoire* : expérience dont le résultat est imprévisible
- *Espace fondamental* : ensemble des résultats possibles, noté Ω
- *Événement élémentaire* : élément de Ω
- *Événement* : assertion logique relative au résultat d'une expérience

Exemple. Deux jets successifs d'une pièce de monnaie.

- espace fondamental : $\Omega = \{(0,0);(0,1);(1,0);(1,1)\}$
- événement élémentaire : "obtenir deux faces", soit $\{(1,1)\}$
- événement : "obtenir au moins une face", soit $\{(0,1);(1,0);(1,1)\}$

Le modèle probabiliste

Espace probabilisable

notation	ensembliste	probabiliste
ω	point de Ω	événement élémentaire
A	sous-ensemble de Ω	événement
$\omega \in A$	ω appartient à A	ω réalise A
$A \subset B$	A est contenu dans B	A implique B
$A \cup B$	réunion de A et B	A ou B
$A \cap B$	intersection de A et B	A et B
$\overline{A}$	complémentaire de A	contraire de A
$\emptyset$	ensemble vide	événement impossible
Ω	ensemble plein	événement certain

TABLE : Correspondance entre les vocabulaires ensembliste et probabiliste.

Le modèle probabiliste

Espace probabilisable

Définition

Deux événements A et B sont incompatibles si la réalisation de l'un exclut celle de l'autre. Ceci signifie que les parties A et B de Ω sont disjointes, c'est-à-dire : $A \cap B = \emptyset$.

Définition

On dit que $A_1, A_2, \ldots, A_n$ forment un système complet d'événements si les parties $A_1, A_2, \ldots, A_n$ de Ω forment une partition de l'espace fondamental :

$$\begin{cases} A_i \cap A_j = \emptyset, & \forall i \neq j \\ \cup A_i = \Omega. \end{cases}$$

Le modèle probabiliste

Espace probabilisable

Tout événement peut se voir associé à un sous-ensemble de Ω.

Qu'en est-il de la réciproque ?

On suppose que l'ensemble des événements est une classe $\mathcal{A}$ de parties de Ω. Il est naturel d'exiger

- $\Omega \in \mathcal{A}$
- $A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A}$
- si A_i est un ensemble d'événements de $\mathcal{A}$, alors $\cup A_i \in \mathcal{A}$

Une telle classe est appelée tribu ou σ-algèbre de Boole

Définition

On appelle espace probabilisable le couple $(\Omega, \mathcal{A})$, où $\mathcal{A}$ constitue une tribu de parties de Ω.

Le modèle probabiliste

Espace probabilisé

La notion de probabilité a pour objectif de mesurer la chance qu'un événement a de se produire lors d'une expérience aléatoire, en lui associant un nombre réel de $[0, 1]$.

Définition

$(\Omega, \mathcal{A})$ étant un espace probabilisable, on appelle probabilité sur $(\Omega, \mathcal{A})$, toute application P définie sur $\mathcal{A}$ à valeurs dans $[0, 1]$ vérifiant les conditions suivantes :

- $P(\Omega) = 1$;
- *pour tout ensemble dénombrable d'événements incompatibles $\{A_i\}$, on a la relation $P(\cup A_i) = \sum P(A_i)$.*

Le triplet $(\Omega, \mathcal{A}, P)$ est appelé espace probabilisé.

Le modèle probabiliste

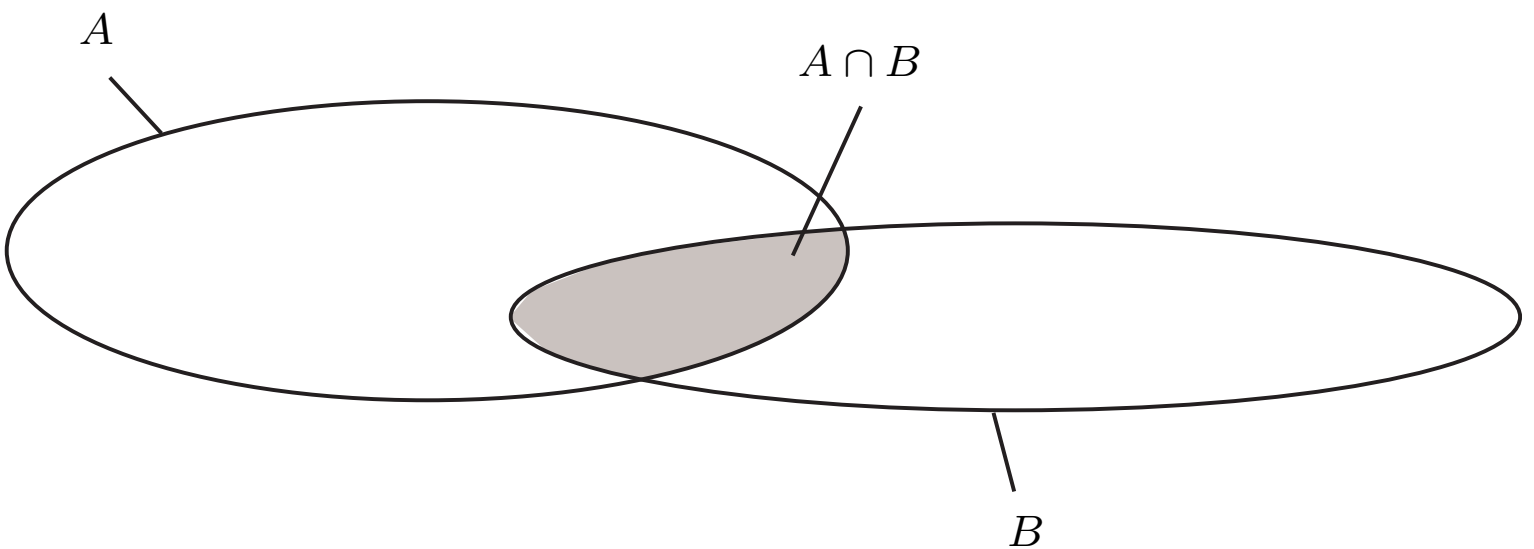
Probabilité conditionnelle, formules de Bayes

On considère la réalisation de A, sachant que B s'est produit.

Définition

B étant un événement de probabilité non nulle, on appelle probabilité conditionnelle de A sachant B le rapport

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



Le modèle probabiliste

Probabilité conditionnelle, formules de Bayes

Les relations de Bayes, qui visent à exprimer la probabilité conditionnelle $P(A|B)$ en fonction de $P(B|A)$, se déduisent directement de la définition précédente.

Théorème

Pour tout couple d'événements A et B, on a

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

Théorème

Si $B_1, B_2, \ldots, B_n$ forment un système complet d'événements, alors

$$P(B_i|A) = \frac{P(A|B_i)\,P(B_i)}{\sum_j P(A|B_j)\,P(B_j)}.$$

Le modèle probabiliste

Indépendance d'événements

On dit qu'un événement A est indépendant d'un événement B si l'information apportée sur B par A est nulle, c'est-à-dire $P(A|B) = P(A)$.

Définition

Deux événements A et B sont indépendants si

$$P(A \cap B) = P(A)\,P(B).$$

Cette définition peut être aisément étendue au cas de n événements.

Définition

Soient $A_1, A_2, \ldots, A_n$ des événements. Ils sont dits mutuellement indépendants si, pour toute partie I de $\{1, \ldots, n\}$, on a

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$

Variables aléatoires et lois de probabilité

Approche intuitive

Lorsqu'on s'intéresse à la somme des points marqués à la suite du lancé de deux dés, on définit implicitement une application X de l'espace fondamental

$$\Omega = \{(1,1),(1,2),\ldots,(6,6)\}$$

dans l'ensemble $\Omega' = \{2,3,\ldots,12\}$. Le concept de *variable aléatoire* vise précisément à formaliser cette notion consistant à associer le résultat d'une expérience aléatoire à un nombre.

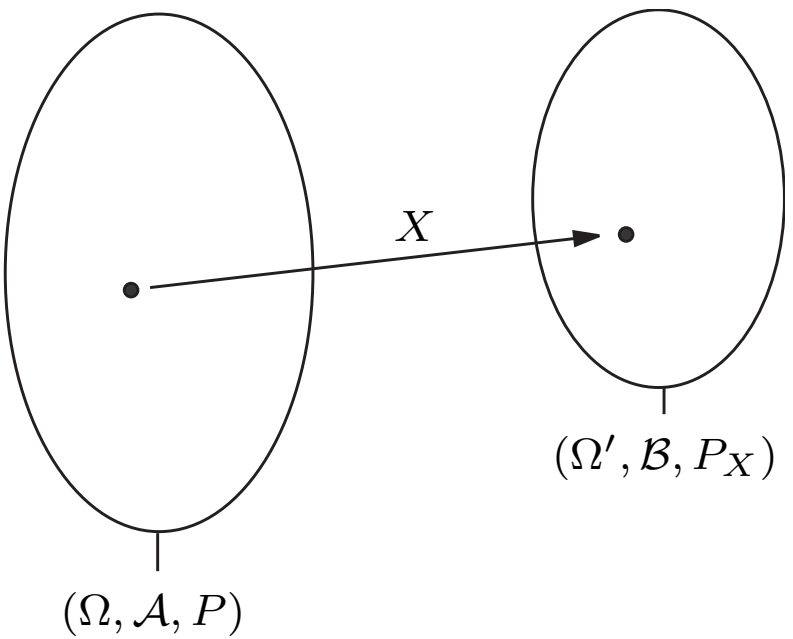
Variables aléatoires et lois de probabilité

definition

Définition

Une variable aléatoire X est une application mesurable de $(\Omega, \mathcal{A}, P)$ dans $(\Omega', \mathcal{B})$, c'est à dire telle que $\forall B \in \mathcal{B}, X^{-1}(B) \in \mathcal{A}$.

- *Si $\Omega' = \mathbb{R}$, X est appelée variable aléatoire réelle.*
- *Si Ω' est un sous-ensemble V fini ou dénombrable de $\mathbb{R}$, X est appelée variable aléatoire réelle discrète.*



Variables aléatoires et lois de probabilité

Variables aléatoires discrètes

Définition

La loi de probabilité d'une variable aléatoire discrète à valeurs dans $X(\Omega)$ est définie par l'ensemble des masses $p(x) = P(X = x)$, $x \in X(\Omega)$.

La loi de probabilité $p(x)$ vérifie les propriétés suivantes

- $p(x) \geq 0$
- $\sum_x p(x) = 1$
- $P_X(B) = \sum_{x \in B} p(x)$, $\forall\, B \in \mathcal{B}$

Variables aléatoires et lois de probabilité

Variables aléatoires continues

La notion de variable aléatoire absolument continue se confond avec celle de variable aléatoire admettant une densité de probabilité.

Définition

Une variable aléatoire réelle X admet pour densité f si, pour tout intervalle B de $\mathbb{R}$, on a :

$$P_X(B) = \int_B f(x)\,dx.$$

Parmi les propriétés satisfaites par toute densité de probabilité f, on peut citer :

- $f(x) \geq 0$
- $\int_{\mathbb{R}} f(x)\,dx = 1$
- $P(x < X < x + dx) = f(x)\,dx$
- $P(X = x) = 0$

Variables aléatoires et lois de probabilité

Espérance mathématique

Définition

L'espérance $E(X)$ d'une variable aléatoire discrète à valeurs dans $X(\Omega)$ est donnée par la relation suivante

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

sous réserve de convergence de cette série.

Définition

L'espérance $E(X)$ d'une variable aléatoire continue admettant une densité f est définie par

$$E(X) = \int_{\mathbb{R}} x\, f(x)\, dx,$$

sous réserve de convergence de cette intégrale.

Variables aléatoires et lois de probabilité

Propriété fondamentale de l'espérance mathématique

Le théorème suivant présente l'expression de l'espérance mathématique d'une fonction ϕ d'une variable aléatoire X.

Théorème

Soit ϕ une application de $\mathbb{R}$ dans $\mathbb{R}$. Si X désigne une variable aléatoire discrète à valeurs dans $X(\Omega)$ et de loi p, alors :

$$E(\phi(X)) = \sum_{x \in X(\Omega)} \phi(x)\, p(x).$$

Si X désigne une variable aléatoire continue de densité f, alors :

$$E(\phi(X)) = \int_{\mathbb{R}} \phi(x)\, f(x)\, dx.$$

Variables aléatoires et lois de probabilité

Variance

La variance d'une variable aléatoire réelle X, notée $\mathrm{var}(X)$ ou σ^2, caractérise sa dispersion autour de son espérance. Elle est définie comme suit.

Définition

La variance $\mathrm{var}(X)$ *d'une variable aléatoire est définie, si elle existe, par la relation :*

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

Les propriétés élémentaires qu'elle vérifie sont les suivantes :

- $E([X - a]^2) = \mathrm{var}(X) + [E(X) - a]^2$
- $\mathrm{var}(X) = E(X^2) - [E(X)]^2$
- $\mathrm{var}(aX + b) = a^2 \,\mathrm{var}(X)$

Vecteurs aléatoires

Loi jointe et lois marginales

Un vecteur aléatoire X est une application mesurable de $(\Omega, \mathcal{A}, P)$ dans $(\Omega', \mathcal{B}, P_X)$, où Ω' représente $\mathbb{R}^p$, ou toute partie de $\mathbb{R}^p$.

La loi de probabilité du vecteur aléatoire (X, Y) est généralement appelée *loi jointe*. Elle est donnée par

$$P_{XY}(B) = P(\{\omega | (X(\omega), Y(\omega)) \in B\}).$$

On désigne par *lois marginales* les lois de X et de Y considérés séparément.

Vecteurs aléatoires discrets

Loi jointe et lois marginales

Définition

La loi jointe d'un vecteur aléatoire discret à valeurs dans $V(\Omega)$ est entièrement définie par l'ensemble des valeurs

$$p_{XY}(x, y) = P(X = x\,;\, Y = y),$$

prises pour tous les couples (x, y) de $V(\Omega)$.

Les lois marginales de X et de Y sont respectivement définies par :

- $p_X(x) = \sum_{y \in V_y} p_{XY}(x, y)$ où V_Y est le domaine de variation de Y ;
- $p_Y(y) = \sum_{x \in V_x} p_{XY}(x, y)$ où V_X est le domaine de variation de X.

Vecteurs aléatoires continus

Loi jointe et lois marginales

Comme dans le cas des variables aléatoires, la notion de vecteur aléatoire continu se confond avec celle de vecteur aléatoire admettant une densité de probabilité jointe.

Définition

Un vecteur aléatoire réel (X, Y) admet pour densité jointe $f_{X,Y}$ si, pour tout ensemble $V_X \times V_Y$ de $\mathbb{R}^2$, on a :

$$P(X \in V_X \,;\, Y \in V_Y) = \int_{V_X \times V_Y} f_{XY}(x, y)\, dx\, dy.$$

Les lois marginales de X et de Y sont respectivement données par

- $f_X(x) = \int_{y \in \mathbb{R}} f_{XY}(x, y)\, dy$;
- $f_Y(y) = \int_{x \in \mathbb{R}} f_{XY}(x, y)\, dx$.

Vecteurs aléatoires continus

Lois conditionnelles

On sait que $P(A|B) = P(A \cap B)/P(B)$ si $P(A) > 0$. En considérant que A et B désignent les événements respectifs $X = x$ et $Y = y$, où X et Y sont des variables discrètes, alors :

$$P(X = x|Y = y) = \frac{p(x, y)}{p_Y(y)}.$$

On appelle cette dernière *fonction de probabilité conditionnelle* de X sachant Y.

Vecteurs aléatoires
Moments

Définition

L'espérance du vecteur aléatoire (X, Y) est le vecteur des espérances de X et Y :

$$E(X, Y) = (E(X), E(Y)).$$

Définition

Étant données deux variables aléatoires X et Y, on appelle covariance entre X et Y la quantité suivante :

$$\mathrm{cov}(X, Y) = E\left([X - E(X)]\,[Y - E(Y)]\right).$$

Vecteurs aléatoires
Indépendance de variables aléatoires

Théorème

Soient X et Y deux variables aléatoires discrètes. Celles-ci sont dites indépendantes si

$$P(X = x\,;\, Y = y) = P(X = x)\; P(Y = y).$$

Théorème

Soient X et Y deux variables aléatoires absolument continues de densité jointe $f_{XY}(x, y)$. Celles-ci sont dites indépendantes si

$$f_{XY}(x, y) = f_X(x)\, f_Y(y).$$

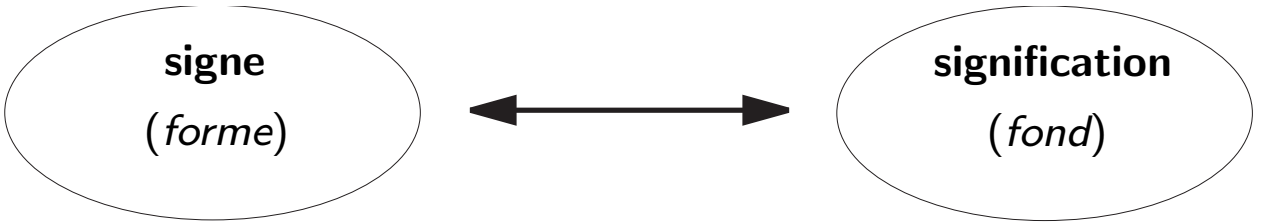
Objet de la théorie de l'information

Chapitre 0

De la notion d'information

Plusieurs conceptions

La notion d'information diffère selon qu'on se place du côté de la machine ou de l'individu. La conception analytique en rend compte.

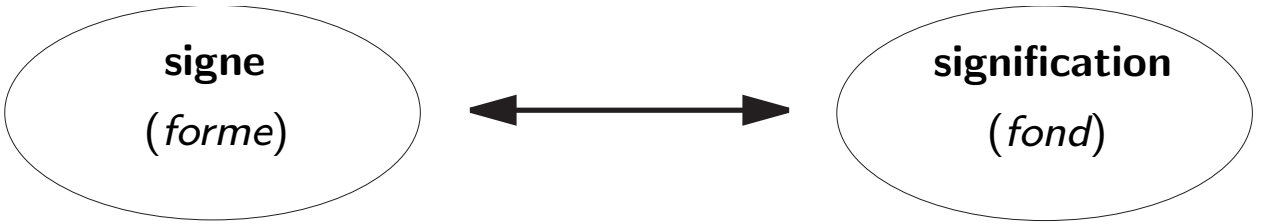


Conception analytique de l'information.

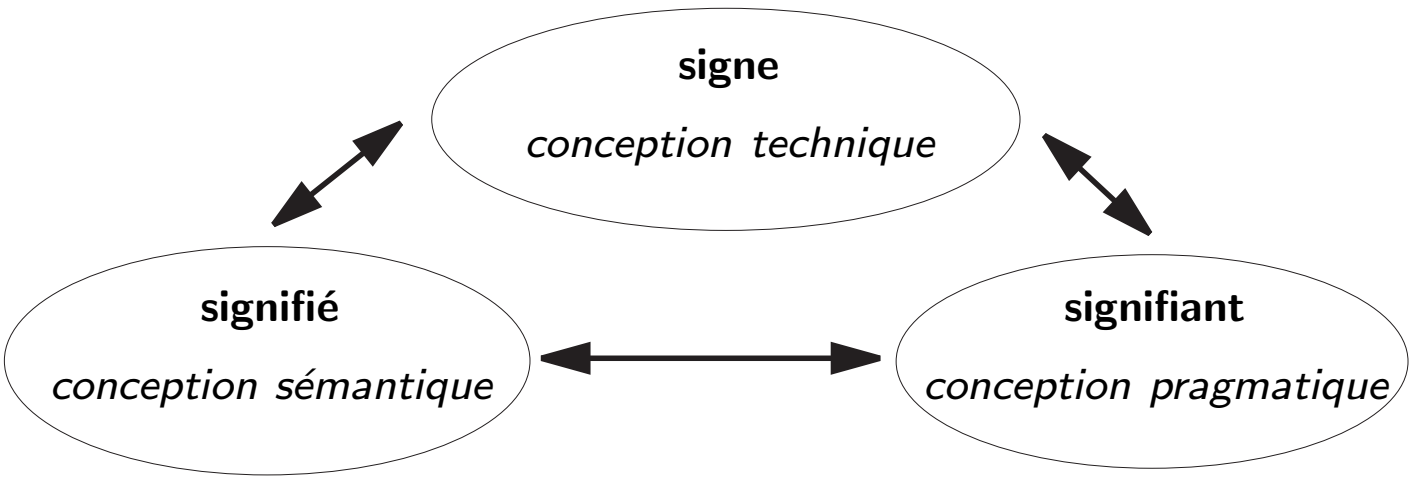
De la notion d'information

Plusieurs conceptions

La notion d'information diffère selon qu'on se place du côté de la machine ou de l'individu. La conception analytique en rend compte.



Conception analytique de l'information.



Conception systémique de l'information.

De la notion d'information

Objectifs de la théorie de l'information

Priorités du système informatique

Importance du signe prépondérante dans le traitement, le stockage et la transmission.

Priorités du système d'information

Aspects sémantiques et pragmatiques privilégiés.

Exemple : la facturation électronique

Remplace ou accompagne la facturation classique ?
Nombre de signes échangés, flux de données ?

La théorie de l'information s'intéresse au signe.

De la notion d'information

Les origines de la théorie de l'information (1928 - ...)

Travaux de H. Nyquist pour la théorie des communications

- Liens entre bande passante et vitesse d'émission.
- Etude des distorsions inter-symboles.

Travaux de R.V. Hartley

- Une définition de la notion d'information.

Oeuvre de C.E. Shannon

- Performances limites en présence de perturbations.
- Notions de source d'information et de canal de transmission.

De la notion d'information

Les origines de la théorie de l'information (1928 - ...)

Travaux de H. Nyquist pour la théorie des communications

- Liens entre bande passante et vitesse d'émission.
- Etude des distorsions inter-symboles.

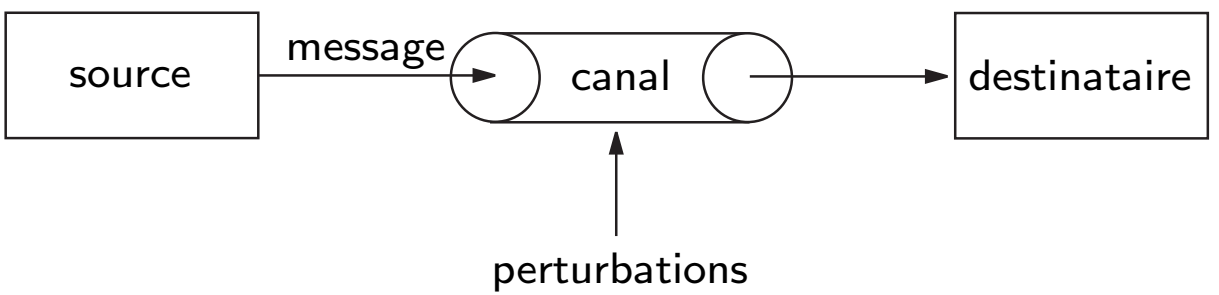
Travaux de R.V. Hartley

- Une définition de la notion d'information.

Oeuvre de C.E. Shannon

- Performances limites en présence de perturbations.
- Notions de source d'information et de canal de transmission.

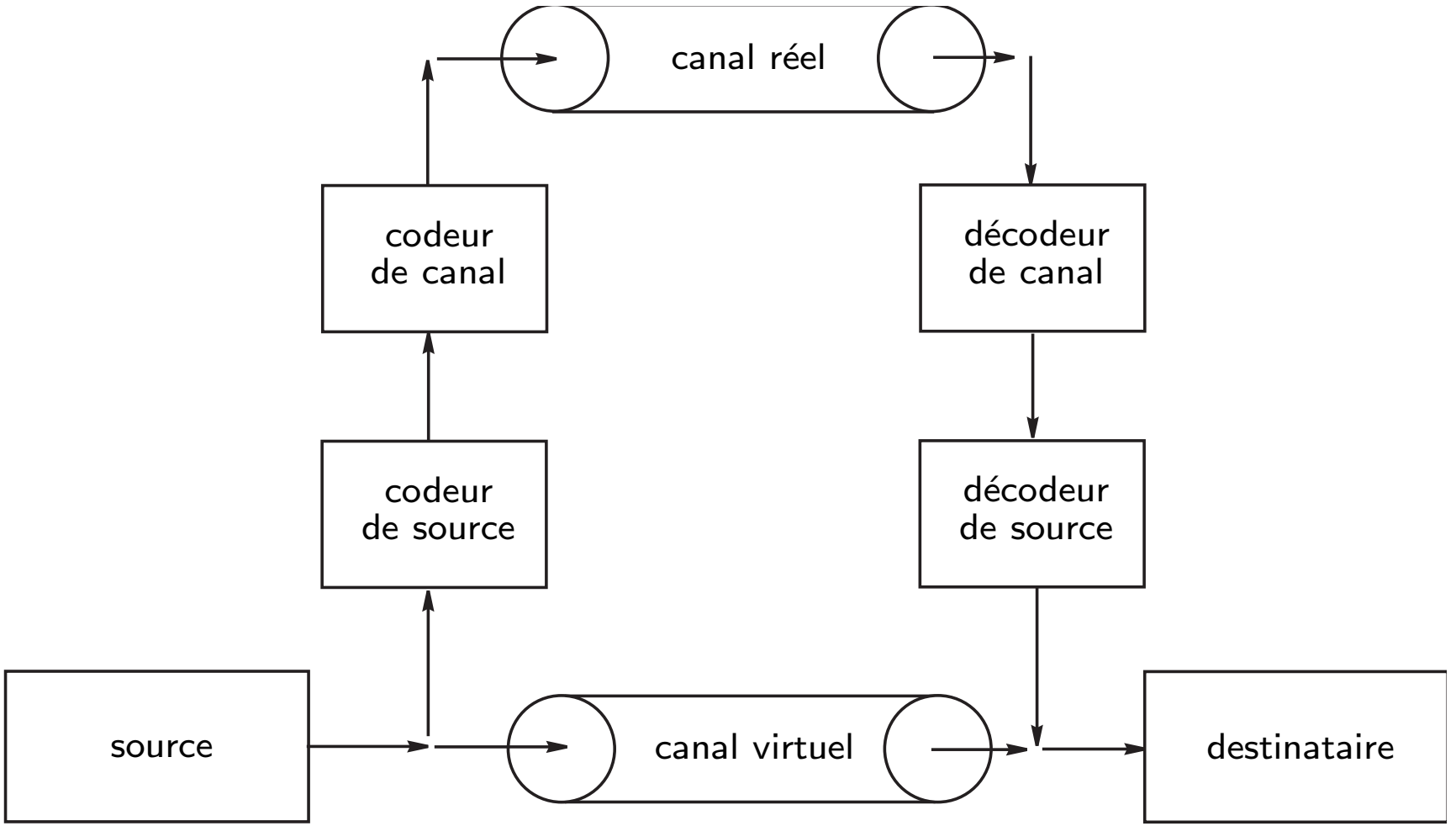
Paradigme de Shannon



- source : générateur de *message*.
- message : suite de *symboles* d'un *alphabet* donné.
- canal : vecteur de l'information entre *source* et *destinataire*.
- perturbations : stochastiques par nature.

Modèle de communication

Schéma général d'un système



"Théorie et codage de l'information"

Plan du cours et modalités de contrôle

- Chapitre 1 : Mesure quantitative de l'information
  → *entropie, information mutuelle, etc.*
- Chapitre 2 : Codage de source discrète
  → *modèles de source, théorème 1 de Shannon, techniques de codage, etc.*
- Chapitre 3 : Canaux discrets
  → *capacité d'un canal, théorème 2 de Shannon, décodage, etc.*
- Chapitre 4 : Techniques de codage de canal
  → *algèbre sur les corps finis, codes linéaires, etc.*
- Chapitre 5 : Mises en oeuvre de la théorie de l'information
  → *codage du son et de la parole, codage des images fixes et animées, etc.*

Outline

Mesure quantitative de l'information

Chapitre 1

Information propre et mutuelle

Quantité d'information propre d'un événement

Soit A un événement de probabilité $P(A)$ non-nulle.

L'information $h(A)$ apportée par la réalisation de A est d'autant plus grande qu'elle est improbable. Elle peut s'exprimer ainsi :

$$h(A) = f\left(\frac{1}{P(A)}\right).$$

Information propre et mutuelle

Quantité d'information propre d'un événement

Soit A un événement de probabilité $P(A)$ non-nulle.

L'information $h(A)$ apportée par la réalisation de A est d'autant plus grande qu'elle est improbable. Elle peut s'exprimer ainsi :

$$h(A) = f\left(\frac{1}{P(A)}\right).$$

La fonction $f(\cdot)$ vérifie les contraintes suivantes :

- $f(\cdot)$ est croissante
- info. apportée par 1 événement sûr est nulle : $\lim_{p\to 1} f(p) = 0$
- info. apportée par 2 événements indépendants : $f(p_1 \cdot p_2) = f(p_1) + f(p_2)$

Ceci nous conduit à utiliser la fonction logarithmique pour $f(\cdot)$

Information propre et mutuelle

Quantité d'information propre d'un événement

Lemme

La fonction $f(p) = -\log_b p$ est la seule qui soit à la fois positive, continue sur $]0,1]$, et qui vérifie $f(p1 \cdot p2) = f(p1) + f(p2)$.

Preuve. La démonstration comporte les étapes suivantes :

1. $f\left(p^n\right) = n\, f(p)$
2. $f\left(p^{1/n}\right) = \frac{1}{n}\, f(p)$ après avoir remplacer p par $p^{1/n}$
3. $f\left(p^{m/n}\right) = \frac{m}{n}\, f(p)$ en combinant le deux égalités précédentes
4. $f(p^q) = q\, f(p)$ où q désigne un nombre rationnel positif quelconque
5. $f\left(p^r\right) = \lim_{n\to+\infty} f\left(p^{q_n}\right) = \lim_{n\to+\infty} q_n\, f(p) = r\, f(p)$

Soient p et q appartenant à $]0,1[$. On peut écrire $p = q^{\log_q p}$, ce qui entraîne

$$f(p) = f\left(q^{\log_q p}\right) = f(q)\, \log_q p.$$

On aboutit finalement au résultat escompté, soit

$$\boldsymbol{f(p) = -\log_b p}$$

Information propre et mutuelle

Quantité d'information propre d'un événement

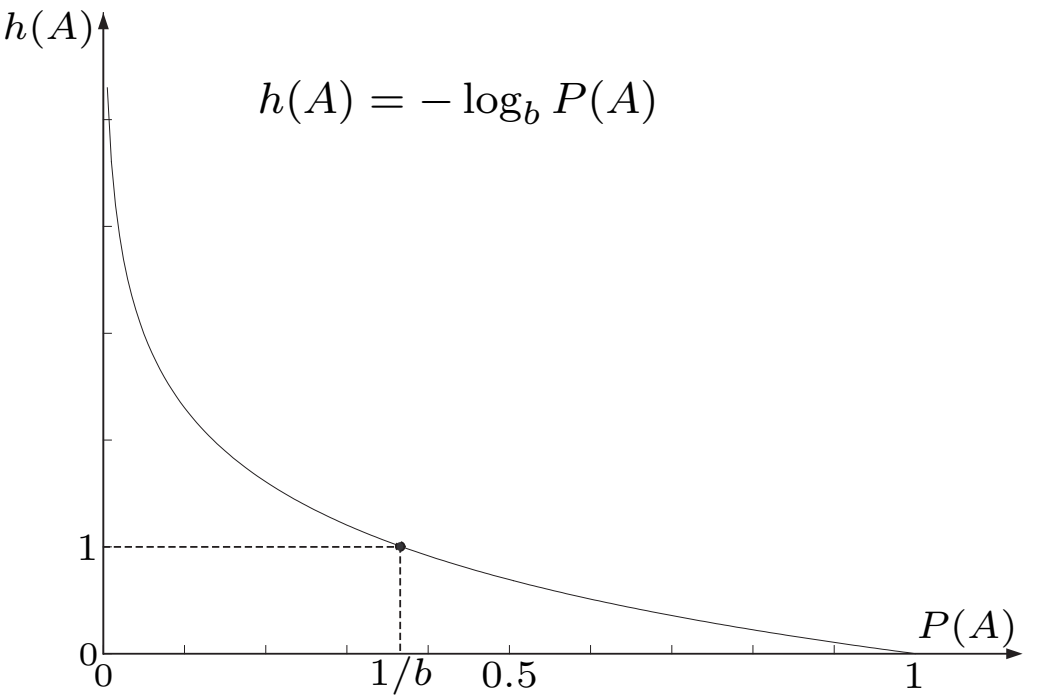
Définition

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et A un événement de $\mathcal{A}$ de probabilité $P(A)$ non-nulle. On associe à la réalisation de A la quantité d'information propre :

$$h(A) = -\log P(A).$$

Unités :
L'unité dépend de la base choisie pour le logarithme.

- $\log_2$: Shannon, bit (binary unit)
- $\log_e$: logon, nat (natural unit)
- $\log_{10}$: Hartley, decit (decimal unit)



Vocabulaire. $h(\cdot)$ est désigné par *incertitude* ou encore *quantité d'information*.

Information propre et mutuelle

Quantité d'information propre d'un événement

Exemple 1. Dans le cas d'une source binaire $\{0, 1\}$ telle que $P(0) = P(1) = 0.5$, l'information propre associée à chaque symbole binaire, ou bit au sens informatique du terme, vaut $h\left(\frac{1}{2}\right) = \log 2$, soit 1 bit ou Shannon.

Exemple 2. On considère une source S sélectionnant aléatoirement et indépendamment du passé chaque symbole émis parmi les 16 éléments d'un l'alphabet $\{s_0, \ldots, s_{15}\}$, tous équiprobables. L'information propre véhiculée par chacun d'eux est $\log 16$, soit 4 Shannon.

Attention !

Le bit informatique (*binary digit*) et le bit issu de la théorie de l'information (*binary unit*) sont deux notions distinctes.

Information propre et mutuelle

Quantité d'information conditionnelle

La définition de la quantité d'information propre s'applique à la réalisation conjointe de A et B. En remarquant que $P(A \cap B) = P(A)\,P(B|A)$, on obtient :

$$h(A \cap B) = -\log P(A \cap B) = -\log P(A) - \log P(B|A)$$

On note que $-\log P(B|A)$ correspond à la quantité d'information propre de B que ne fournit pas l'observation de A.

Définition

On appelle information conditionnelle de B sachant A la quantité :

$$h(B|A) = -\log P(B|A),$$

soit, en d'autres termes : $h(B|A) = h(A \cap B) - h(A)$.

Exercice. Étudier et interpréter le cas $A = B$, puis indépendance de A et B

Information propre et mutuelle

Quantité d'information mutuelle

La définition de l'information conditionnelle amène directement une autre définition, celle de l'information mutuelle, qui correspond à la part d'incertitude commune à deux événements.

Définition

On appelle information mutuelle entre A et B la quantité :

$$i(A,B) = h(A) - h(A|B) = h(B) - h(B|A).$$

Exercice. Étudier les cas $A = B$, $B \subset A$, enfin A et B indépendants.

Entropie d'une variable aléatoire

Définition de l'entropie

Soit une source S d'information sans mémoire sélectionnant aléatoirement un symbole parmi les n éléments d'un alphabet $\{s_1, \ldots, s_n\}$. Soit p_i la probabilité d'apparition de s_i. La quantité d'information moyenne associée à l'apparition de chaque symbole possible est donnée par :

$$H(S) = E\{h(s)\} = -\sum_{i=1}^{n} p_i \log p_i.$$

L'entropie est une quantité d'information moyenne.

Définition

Soit X une variable aléatoire à valeurs dans $\{x_1, \ldots, x_n\}$. L'entropie de X est définie comme suit :

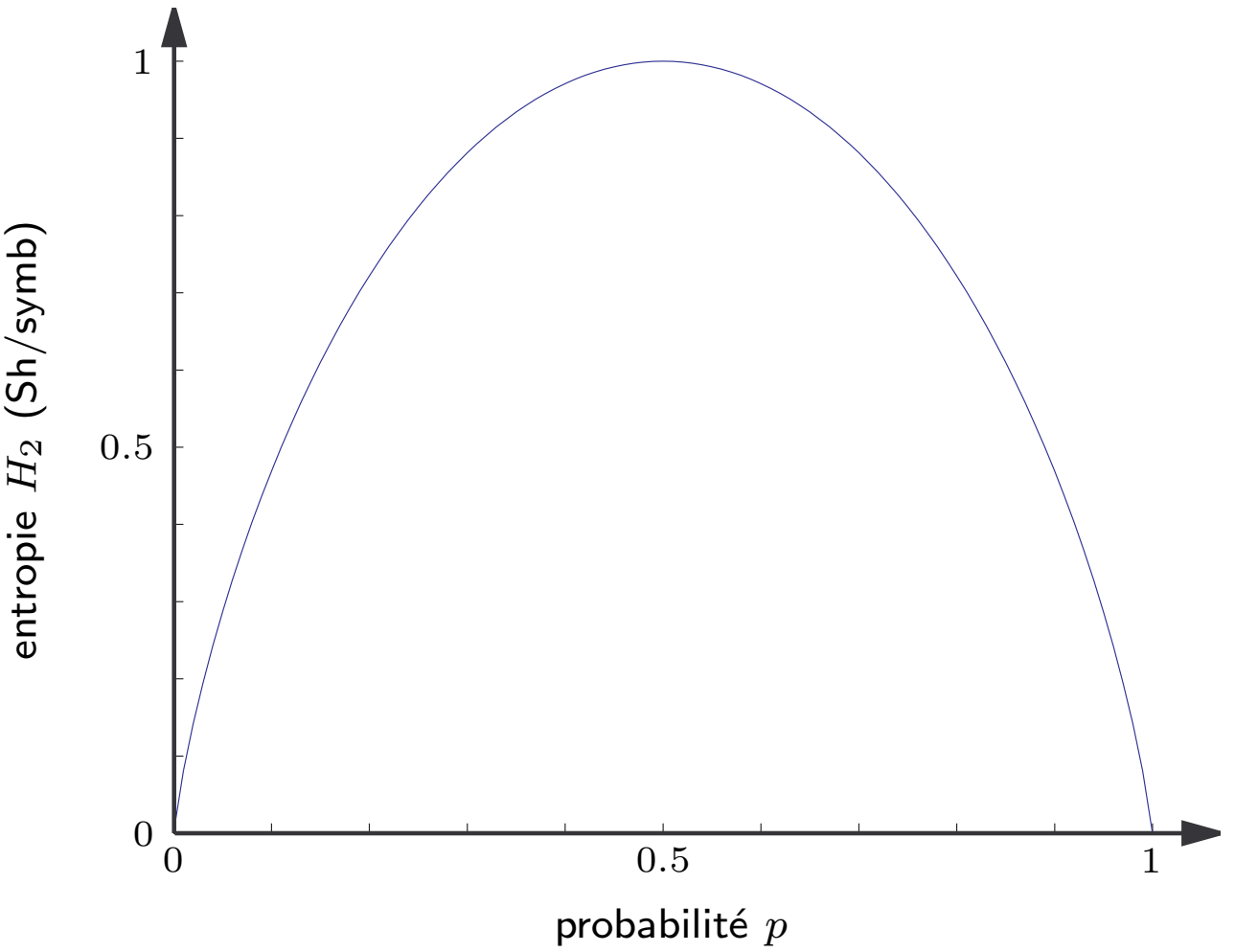
$$H(X) = -\sum_{i=1}^{n} P(X = x_i) \log P(X = x_i).$$

Entropie d'une variable aléatoire

Exemple d'une variable aléatoire binaire

L'entropie d'une variable aléatoire binaire est donnée par :

$$H(X) = -p \log p - (1-p) \log(1-p) \triangleq H_2(p).$$



Entropie d'une variable aléatoire

Notation et propriété préalables

Lemme (Inégalité de Gibbs)

Étant donné 2 distributions de probabilité discrètes $(p_1, \ldots, p_n)$ *et* $(q_1, \ldots, q_n)$ *sur un même univers fini, l'inégalité suivante est satisfaite :*

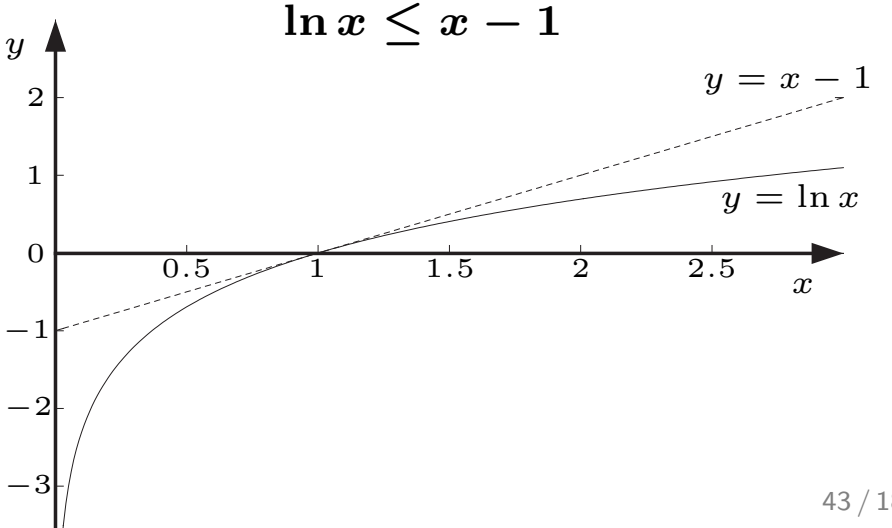
$$\sum_{i=1}^{n} p_i \log \frac{q_i}{p_i} \leq 0,$$

l'égalité étant obtenue lorsque $\forall i$ *:* $p_i = q_i$*.*

Preuve.
On effectue la démonstration dans le cas du logarithme népérien et on note que $\ln x \leq x - 1$, l'égalité étant obtenue pour $x = 1$. On pose $x = \frac{q_i}{p_i}$ et on a

$$\sum_{i=1}^{n} p_i \ln \frac{q_i}{p_i} \leq \sum_{i=1}^{n} p_i (\frac{q_i}{p_i} - 1) = 1 - 1 = 0.$$

Vérification graphique de l'inégalité $\ln x \leq x - 1$



Entropie d'une variable aléatoire

Quelques propriétés de l'entropie

Propriété

L'entropie vérifie l'inégalité suivante

$$H_n(p_1, \ldots, p_n) \leq \log n,$$

l'égalité étant réalisée dans le cas d'une loi uniforme, c'est-à-dire $\forall i : p_i = \frac{1}{n}$.

Preuve. A partir de l'inégalité de Gibbs, on pose $q_i = \frac{1}{n}$. L'incertitude sur le résultat d'une expérience est d'autant plus grande que tous les résultats possibles sont équiprobables.

Entropie d'une variable aléatoire

Quelques propriétés de l'entropie

Propriété

L'entropie augmente lorsque le nombre d'états du système augmente.

Preuve. Soit X une variable aléatoire discrète à valeurs dans $\{x_1, \ldots, x_n\}$ avec les probabilités $(p_1, \ldots, p_n)$. On suppose que l'état x_k est scindé en deux sous-états x_{k_1} et x_{k_2}, de probabilités respectives p_{k_1} et p_{k_2} non-nulles telles que $p_k = p_{k_1} + p_{k_2}$.

L'entropie de la variable aléatoire résultante X' s'écrit :

$$\begin{aligned} H(X') &= H(X) + p_k \log p_k - p_{k_1} \log p_{k_1} - p_{k_2} \log p_{k_2} \\ &= H(X) + p_{k_1}(\log p_k - \log p_{k_1}) + p_{k_2}(\log p_k - \log p_{k_2}). \end{aligned}$$

La fonction logarithmique étant strictement croissante, on a : $\log p_k > \log p_{k_i}$. Il en résulte directement que $H(X') > H(X)$.

Interprétation. Second Principe de la Thermodynamique

Entropie d'une variable aléatoire

Quelques propriétés de l'entropie

Propriété

L'entropie H_n est une fonction concave de $p_1, \ldots, p_n$.

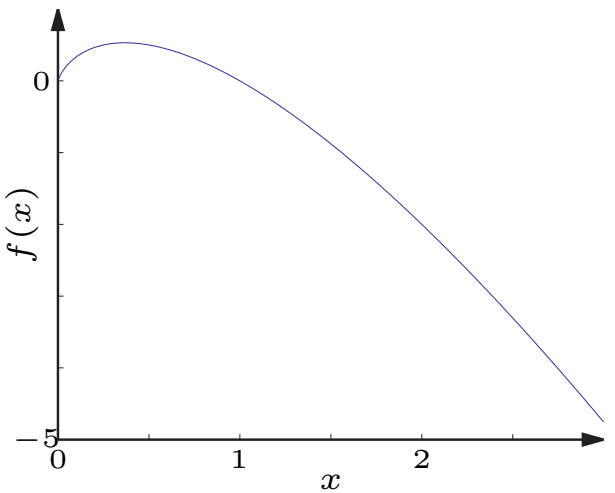
Preuve. Soient 2 distributions de probabilité discrètes $(p_1, \ldots, p_n)$ et $(q_1, \ldots, q_n)$. La concavité de la fonction entropie qu'il s'agit de démontrer se traduit par le fait que pour tout λ de l'intervalle $[0, 1]$, on a :

$$H_n(\lambda p_1 + (1-\lambda)q_1, \ldots, \lambda p_n + (1-\lambda)q_n) \geq \lambda H_n(p_1, \ldots, p_n) + (1-\lambda)H_n(q_1, \ldots, q_n).$$

En posant $f(x) = -x \log x$, on peut écrire que :

$$H_n(\lambda p_1 + (1-\lambda)q_1, \ldots, \lambda p_n + (1-\lambda)q_n) = \sum_{i=1}^{n} f(\lambda p_i + (1-\lambda)q_i).$$

Le résultat découle directement de la concavité de $f(x) = -x \log x$.



Entropie d'une variable aléatoire

Quelques propriétés de l'entropie

La convexité de H_n peut être généralisée à un nombre quelconque de distributions.

Propriété

Étant donné $\{(q_{1j}, \ldots, q_{nj})\}_{j=1}^{m}$ un ensemble fini de distributions de probabilité discrètes, l'inégalité suivante est satisfaite :

$$H_n(\sum_{j=1}^{m} \lambda_j \, q_{1j}, \ldots, \sum_{j=1}^{m} \lambda_j \, q_{mj}) \geq \sum_{j=1}^{m} \lambda_j \, H_n(q_{1j}, \ldots, q_{mj}),$$

avec $\{\lambda_j\}_{j=1}^{m}$ des éléments de $[0, 1]$ tels que $\sum_{j=1}^{m} \lambda_j = 1$.

Preuve. Comme dans le cas précédent, la démonstration de cette inégalité repose sur la concavité de la fonction $f(x) = -x \log x$. A charge du lecteur de le vérifier.

Couple de variables aléatoires

Entropie conjointe

Définition

Soient X et Y des variables aléatoires à valeurs dans $\{x_1, \ldots, x_n\}$ et $\{y_1, \ldots, y_m\}$, respectivement. L'entropie conjointe de X et Y est donnée :

$$H(X,Y) \triangleq -\sum_{i=1}^{n}\sum_{j=1}^{m} P(X=x_i, Y=y_j)\, \log P(X=x_i, Y=y_j).$$

▷ l'entropie conjointe est une grandeur symétrique : $H(X,Y) = H(Y,X)$

Exemple. Cas de variables aléatoires indépendantes ou strictement liées.

Couple de variables aléatoires

Entropie conditionnelle

Définition

Soient X et Y des variables aléatoires à valeurs dans $\{x_1, \ldots, x_n\}$ et $\{y_1, \ldots, y_m\}$. L'entropie conditionnelle de X sachant que $Y = y_j$ est donnée :

$$H(X|Y = y_j) \triangleq -\sum_{i=1}^{n} P(X = x_i|Y = y_j) \log P(X = x_i|Y = y_j).$$

▷ incertitude moyenne sur le résultat de X sachant que celui de Y est y_j

Définition

L'entropie conditionnelle moyenne de X sachant Y est donnée par :

$$H(X|Y) \triangleq \sum_{j=1}^{m} P(Y = y_j)\, H(X|Y = y_j),$$

Exemple. Cas de variables aléatoires indépendantes ou strictement liées.

Couple de variables aléatoires

Relations entre les entropies

La première relation énoncée lie ainsi les diverses entropies définies précédemment :

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y).$$

Ces égalités se démontrent aisément en écrivant d'abord

$$\log P(X = x, Y = y) = \log P(X = x|Y = y) + \log P(Y = y),$$

puis en prenant l'espérance mathématique de chacun des membres.

Propriété (règle de chaînage)

L'entropie jointe de de n variables aléatoires peut être évaluée selon la règle de chaînage suivante :

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i|X_1 \ldots X_{i-1}).$$

Chaque élément de $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$ est positif. On en déduit immédiatement que :

$$H(X) \leq H(X,Y)$$
$$H(Y) \leq H(X,Y)$$

Couple de variables aléatoires

Relations entre les entropies

A partir de la *convexité généralisée* de l'entropie, en posant $q_{ij} = P(X = x_i|Y = y_j)$ et $\lambda_j = P(Y = y_j)$, on aboutit à l'inégalité fondamentale suivante :

$$H(X|Y) \leq H(X)$$

Le conditionnement d'une variable aléatoire diminue son entropie. Sans démonstration, il se généralise ainsi :

Propriété (décroissance par conditionnement)

L'entropie d'une variable aléatoire décroît par conditionnements successifs, soit

$$H(X_1|X_2, \ldots, X_n) \leq \ldots \leq H(X_1|X_2, X_3) \leq H(X_1|X_2) \leq H(X_1),$$

où $X_1, \ldots, X_n$ désignent n variables aléatoires discrètes.

Soient X et Y des variables aléatoires à valeurs dans $\{x_1, \ldots, x_n\}$ et $\{y_1, \ldots, y_m\}$, respectivement. Les relations fondamentales vues précédemment se résument ainsi :

$$0 \leq H(X|Y) \leq H(X) \leq H(X, Y) \leq H(X) + H(Y) \leq 2H(X, Y).$$

Couple de variables aléatoires

Information mutuelle moyenne

Définition

L'information mutuelle moyenne de X et Y est définie par

$$I(X,Y) \triangleq H(X) - H(X|Y)$$

ou encore, de façon équivalente,

$$I(X,Y) \triangleq \sum_{i=1}^{n}\sum_{j=1}^{m} P(X=x_i, Y=y_j)\, \log \frac{P(X=x_i, Y=y_j)}{P(X=x_i)\,P(Y=y_j)}.$$

L'information mutuelle représente l'incertitude moyenne sur X diminuée de celle qui subsiste lorsque Y est connue.

Exercice. Cas de variables aléatoires indépendantes et strictement liées.

Autre interprétation : L'information mutuelle correspond à la distance de Kullback-Leibler entre la loi jointe et le produit des distributions marginales de X et Y.

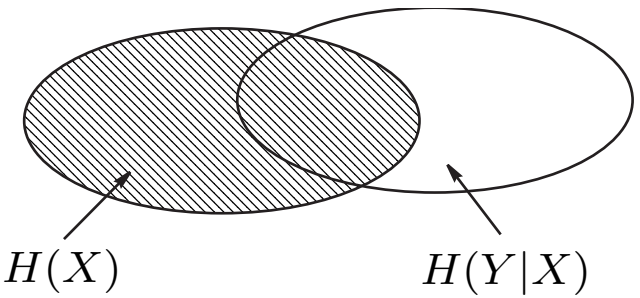
Rappel : La distance de Kullback-Leibler entre deux distributions P_1 et P_2, ici supposées discrètes, est donnée par

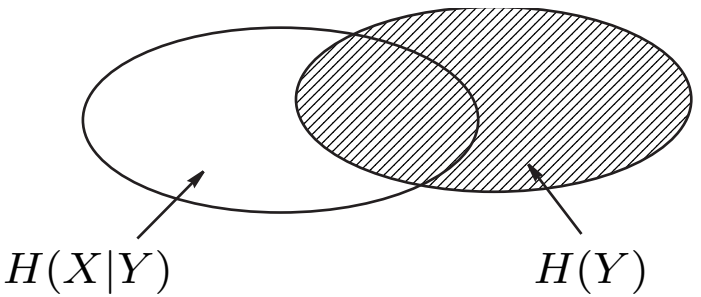
$$d(P_1, P_2) = \sum_{x \in X(\Omega)} P_1(X=x)\, \log \frac{P_1(X=x)}{P_2(X=x)}.$$

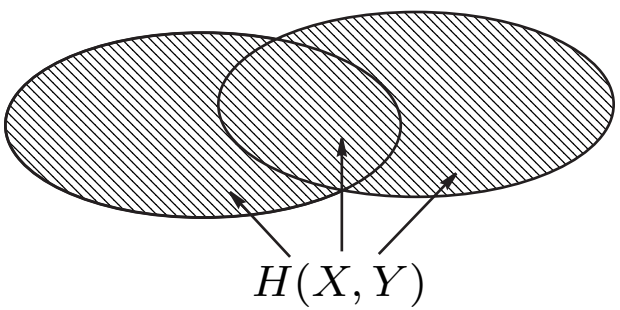
Couple de variables aléatoires

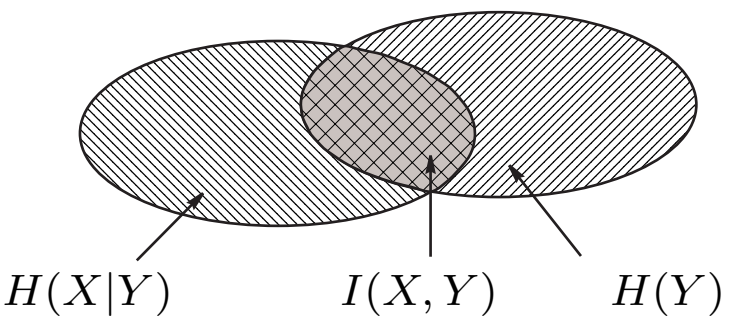
Diagramme de Venn

Le diagramme de Venn, ici à 2 variables, constitue un moyen mnémotechnique.









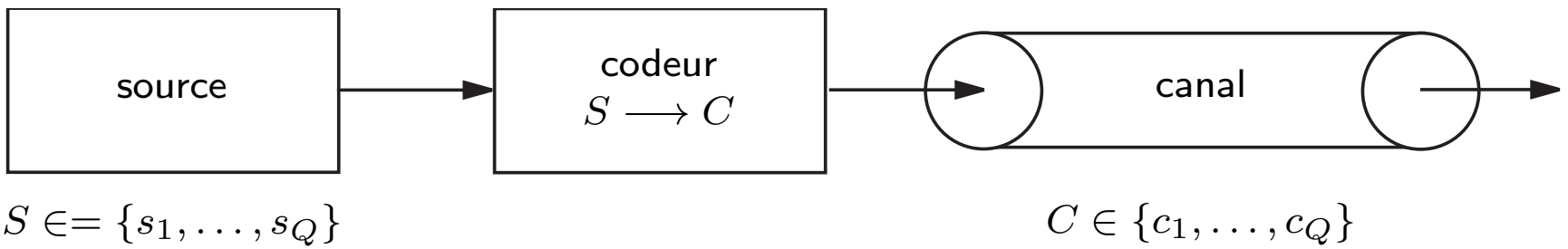
Outline

Codage de source discrète

Chapitre 2

Codage de source discrète

On associe à chacun des Q états s_i de la source un mot approprié, c'est-à-dire une suite de n_i symboles d'un alphabet q-aire. Ceux-ci constituent un code source que l'on note $\mathcal{C} = \{c_1, \ldots, c_q\}$.



Exemple. Le code Morse

- code quaternaire (point, trait, espace long, espace court)
- code de longueur variable
- la séquence la plus courte associée à "E"

Problématique

Adaptation d'une source à un canal non-bruité

Soit S une source caractérisée par un débit D_s (symbole Q-aire/seconde). Soit un canal non-bruité de débit maximal D_c (symbole q-aire/seconde). On définit

- *taux d'émission de la source :* $T \triangleq D_s\, H(S)$
- *capacité du canal :* $C \triangleq D_c \log q$

Si $\mathbf{T} > \mathbf{C}$: le canal ne peut écouler l'information

Si $\mathbf{T} \leq \mathbf{C}$: le canal peut en théorie écouler l'information

Si on dispose d'un code q-aire dont la longueur moyenne $\overline{n}$ des mots est telle $\overline{n}\, D_s \leq D_c$, alors celui-ci peut être utilisé pour la transmission.

Dans le cas contraire, comment coder les états de la source pour rendre leur transmission possible puisque rien ne s'y oppose en théorie ?

Le codage de source vise à éliminer la redondance d'information
SANS PERTE ! ! !

Source discrète en temps discret

Modèle général

Une source discrète est définie par un alphabet $\mathcal{A} = \{s_1, \ldots, s_Q\}$ et un mécanisme d'émission. Il s'agit d'un processus aléatoire en temps discret

$$S_1, \ldots, S_{i-1}, S_i, S_{i+1}, \ldots$$

caractérisé par les lois conjointes :

$$P(S_1, \ldots, S_n), \forall n \in \mathbb{N}^*$$

▷ modèle trop général pour donner lieu à des développements simples

Par simplification, on fait des hypothèses sur le modèle de source.

Source discrète en temps discret

Hypothèses complémentaires

Propriété (Processus stationnaire)

Un processus aléatoire S_i est dit stationnaire si les lois de probabilité qui le régissent sont indépendantes de l'origine des temps, c'est-à-dire

$$P(S_1 = s_{i_1}, \ldots, S_n = s_{i_n}) = P(S_{n_0+1} = s_{i_1}, \ldots, S_{n_0+n} = s_{i_n}),$$

pour tous n_0 et n positifs.

Exemple. Une *source sans mémoire* est caractérisée par des S_i i.i.d.. Il s'agit d'un processus stationnaire $P(S_1 = s_{i_1}, \ldots, S_n = s_{i_n}) = P(S = s_{i_1}) \ldots P(S = s_{i_n})$.

Source discrète en temps discret

Hypothèses complémentaires

Propriété (Processus stationnaire)

Un processus aléatoire S_i est dit stationnaire si les lois de probabilité qui le régissent sont indépendantes de l'origine des temps, c'est-à-dire

$$P(S_1 = s_{i_1}, \ldots, S_n = s_{i_n}) = P(S_{n_0+1} = s_{i_1}, \ldots, S_{n_0+n} = s_{i_n}),$$

pour tous n_0 et n positifs.

Exemple. Une *source sans mémoire* est caractérisée par des S_i i.i.d.. Il s'agit d'un processus stationnaire $P(S_1 = s_{i_1}, \ldots, S_n = s_{i_n}) = P(S = s_{i_1}) \ldots P(S = s_{i_n})$.

Propriété (Processus ergodique)

On dit qu'un processus aléatoire stationnaire S_i est ergodique si, pour tout $k = 1, 2, \ldots$, pour toute suite d'indices $i_1, \ldots, i_k$ et pour toute fonction bornée $f(\cdot)$ de $\mathcal{A}^k$ dans $\mathbb{R}$, on a

$$\frac{1}{n} \sum_{k=1}^{n} f(S_{i_1}, \ldots, S_{i_k}) \xrightarrow{p.s.} E\{f(S_{i_1}, \ldots, S_{i_k})\}.$$

Intérêt. Le processus considéré peut être étudié en observant une trajectoire quelconque mais suffisamment longue de celui-ci.

Source discrète en temps discret

Sources de Markov

Une source quelconque émet un symbole selon une loi qui peut dépendre des symboles qui l'ont précédé.

Définition (Source markovienne)

Une source S est dite markovienne si elle décrit une chaîne de Markov, soit

$$P(S_{n+1} = s_{i_{n+1}} | S_n = s_{i_n}, \ldots, S_1 = s_{i_1}) = P(S_{n+1} = s_{i_{n+1}} | S_n = s_{i_n})$$

pour tous symboles $s_{i_1}, \ldots, s_{i_{n+1}}$ issus de $\mathcal{A}$.

Il en résulte directement que $P(S_1, \ldots, S_n) = P(S_1)\, P(S_2|S_1)\, \ldots\, P(S_n|S_{n-1})$.

Source discrète en temps discret

Sources de Markov

Une source quelconque émet un symbole selon une loi qui peut dépendre des symboles qui l'ont précédé.

Définition (Source markovienne)

Une source S est dite markovienne si elle décrit une chaîne de Markov, soit

$$P(S_{n+1} = s_{i_{n+1}} | S_n = s_{i_n}, \ldots, S_1 = s_{i_1}) = P(S_{n+1} = s_{i_{n+1}} | S_n = s_{i_n})$$

pour tous symboles $s_{i_1}, \ldots, s_{i_{n+1}}$ issus de $\mathcal{A}$.

Il en résulte directement que $P(S_1, \ldots, S_n) = P(S_1)\, P(S_2|S_1)\, \ldots\, P(S_n|S_{n-1})$.

Définition (Invariance dans le temps)

Une source markovienne S est dite invariante dans le temps si, pour tout $n \in \{1, 2, \ldots\}$, on a

$$P(S_{n+1}|S_n) = P(S_2|S_1)$$

Une telle source est entièrement définie par un vecteur $p|_{t=0}$ de probabilités initiales et la matrice de transition Π dont les éléments sont

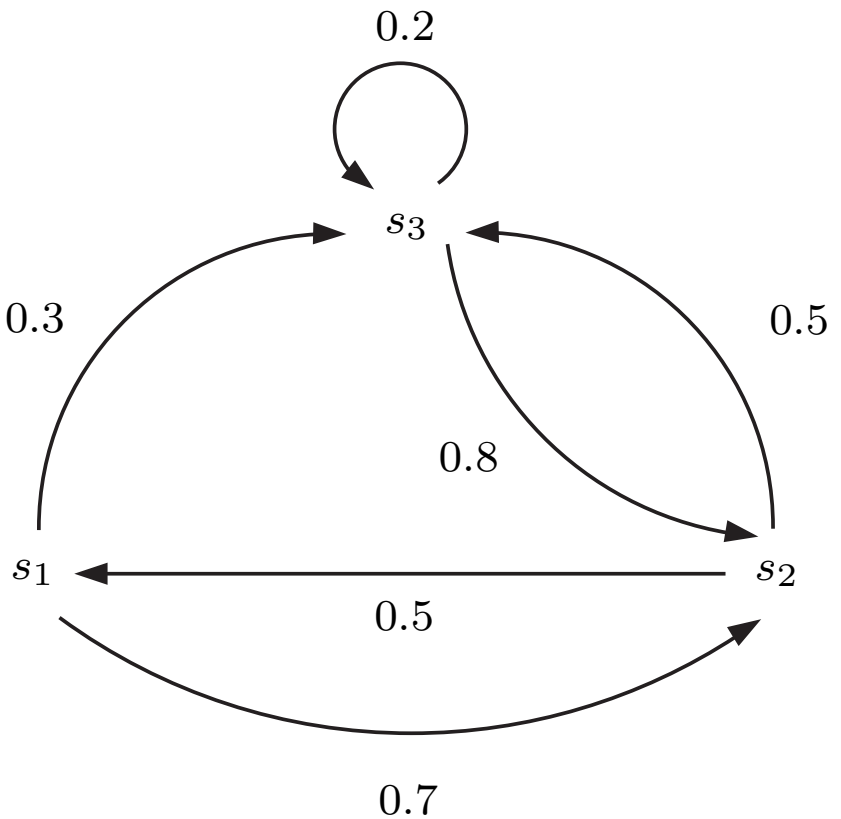
$$\Pi(i, j) = P(S_2 = s_j | S_1 = s_i)$$

Évidemment, on a $\sum_{j=1}^{q} \Pi(i, j) = 1$ et $\Pi(i, j) \geq 0$.

Source discrète en temps discret

Exemple de source de Markov

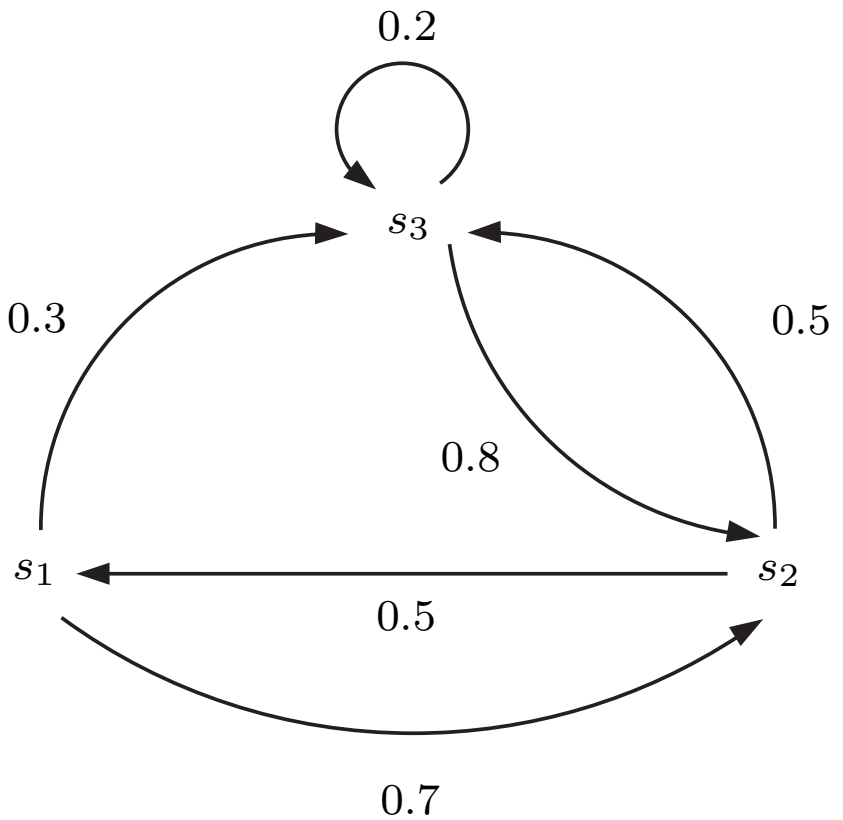
On considère la source markovienne suivante :



Source discrète en temps discret

Exemple de source de Markov

On considère la source markovienne suivante :



La matrice de transition de celle-ci s'écrit ainsi :

$$\Pi = \begin{pmatrix} 0 & 0.7 & 0.3 \\ 0.5 & 0 & 0.5 \\ 0 & 0.8 & 0.2 \end{pmatrix}$$

Source discrète en temps discret

Source de Markov en régime permanent

Définition (régime permanent - version 1)

Une source markovienne S atteint un régime permanent si

$$\lim_{n\to\infty} P(S_n = s_i)$$

existe pour tout $i \in \{1, \ldots, Q\}$.

On note $p|_{t\to\infty}$ la distribution limite si elle existe. Sachant que $p|_{t=n} = p|_{t=n-1}\,\Pi$, on a nécessairement

$$p|_{t\to\infty} = p|_{t\to\infty}\,\Pi$$

On dit que $p|_{t\to\infty}$ est une distribution stationnaire puisque l'initialisation de la chaîne de Markov avec celle-ci la rend stationnaire.

Inconvénient. Le régime permanent ainsi défini dépend de la distribution $p|_{t=0}$ initiale. D'autres définitions existent.

Source discrète en temps discret

Source de Markov en régime permanent

Définition (régime permanent - version 2)

Une source markovienne S atteint un régime permanent si

$$\lim_{n\to\infty} P(S_n = s_i | S_1 = j)$$

existe pour tous $i, j \in \{1, \ldots, Q\}$.

Avantage. Le comportement asymptotique de la source est indépendant de la distribution initiale.

Source discrète en temps discret

Source de Markov d'ordre m

Une source de Markov est caractérisée par une mémoire de taille $m = 1$. Ceci peut être généralisé à des mémoires de taille $m > 1$.

Définition (Source markovienne de taille m)

Une source S est dite markovienne de taille m si elle satisfait à

$$P(S_{n+1} = s_{i_{n+1}} | S_n = s_{i_n}, \ldots, S_1 = s_{i_1}) = P(S_{n+1} = s_{i_{n+1}} | S_{n-m} = s_{i_{n-m}}, \ldots, S_n = s_{i_n})$$

pour tous symboles $s_{i_1}, \ldots, s_{i_{n+1}}$ de $\mathcal{A}$.

Remarque. Une source markovienne de taille m peut être ramenée à une source markovienne de taille 1 en considérant *une extension d'ordre m* ou plus de celle-ci.

Modèles de source discrète

Entropie d'une source stationnaire

Une source quelconque émet un symbole selon une loi qui peut dépendre des symboles qui l'ont précédé. La définition de l'entropie doit en tenir compte.

Définition (Entropie d'une source stationnaire - version 1)

L'entropie d'une source S stationnaire est définie par :

$$H_0 \triangleq \lim_{n \to +\infty} H(S_n | S_1, \ldots, S_{n-1}).$$

Modèles de source discrète

Entropie d'une source stationnaire

Une source quelconque émet un symbole selon une loi qui peut dépendre des symboles qui l'ont précédé. La définition de l'entropie doit en tenir compte.

Définition (Entropie d'une source stationnaire - version 1)

L'entropie d'une source S stationnaire est définie par :

$$H_0 \triangleq \lim_{n \to +\infty} H(S_n | S_1, \ldots, S_{n-1}).$$

Validation de la définition. Il convient de s'assurer de l'existence de la limite.

Le conditionnement d'une variable aléatoire diminuant son entropie, on a :

$$0 \leq H(S_n | S_1, S_2, \ldots, S_{n-1}) \leq H(S_n | S_2, \ldots, S_{n-1}) \leq \ldots \leq H(S_n).$$

Puisque la source considérée est stationnaire, on peut écrire :

$$H(S_n) = H(S_1) \qquad H(S_n | S_{n-1}) = H(S_2 | S_1) \qquad \ldots$$

L'inégalité peut donc être remplacée par :

$$0 \leq H(S_n | S_1, \ldots, S_{n-1}) \leq H(S_{n-1} | S_1, \ldots, S_{n-2}) \leq \ldots \leq H(S_1).$$

La suite $\{H(S_n | S_1, \ldots, S_{n-1})\}_{n \geq 1}$ est décroissante et minorée. Elle est donc convergente, assurant la validité de la définition dans le cas stationnaire.

Modèles de source discrète

Entropie d'une source stationnaire

Définition (Entropie d'une source quelconque - version alternative)

L'entropie d'une source S stationnaire est définie par :

$$H_0 \triangleq \lim_{n \to +\infty} \frac{H(S_1, \ldots, S_n)}{n}.$$

Les deux définitions proposées sont équivalentes dans le cas stationnaire. En effet, il résulte de l'égalité suivante

$$H(S_1, \ldots, S_n) = H(S_1) + H(S_2|S_1) + \ldots + H(S_n|S_1, \ldots, S_{n-1})$$

que $H(S_1, \ldots, S_n)/n$ est la moyenne arithmétique des n premiers termes de la suite $H(S_1)$, $H(S_2|S_1)$, $\ldots$, $H(S_n|S_1, \ldots, S_{n-1})$. Le théorème présenté ci-dessous conduit directement au résultat.

Théorème de Cesaro. Si $a_n \underset{n \to \infty}{\longrightarrow} a$, alors $\frac{1}{n} \sum_{k=1}^{n} a_k \underset{n \to \infty}{\longrightarrow} a$

Modèles de source discrète

Entropie d'une source stationnaire

Définition (Entropie d'une source quelconque - version alternative)

L'entropie d'une source S stationnaire est définie par :

$$H_0 \triangleq \lim_{n \to +\infty} \frac{H(S_1, \ldots, S_n)}{n}.$$

Les deux définitions proposées sont équivalentes dans le cas stationnaire. En effet, il résulte de l'égalité suivante

$$H(S_1, \ldots, S_n) = H(S_1) + H(S_2|S_1) + \ldots + H(S_n|S_1, \ldots, S_{n-1})$$

que $H(S_1, \ldots, S_n)/n$ est la moyenne arithmétique des n premiers termes de la suite $H(S_1)$, $H(S_2|S_1)$, $\ldots$, $H(S_n|S_1, \ldots, S_{n-1})$. Le théorème présenté ci-dessous conduit directement au résultat.

Théorème de Cesaro. Si $a_n \underset{n \to \infty}{\longrightarrow} a$, alors $\frac{1}{n} \sum_{k=1}^{n} a_k \underset{n \to \infty}{\longrightarrow} a$

Exemple 1. Dans le cas d'une source sans mémoire, caractérisée par des S_i indépendants et distribués selon une même loi, on a : $H_0 = H(S_1)$

Exemple 2. Si S désigne une source markovienne invariante dans le temps, l'entropie de celle-ci est donnée par : $H_0 = H(S_2|S_1)$.

Caractérisation d'un codage

Vocabulaire

Le codage de source consiste à associer à chaque symbole s_i d'une source une séquence d'éléments de l'alphabet q-aire de destination, appelée *mot du code*.

Exemple 1. Codes ASCII (7 bits) et ASCII étendu (8 bits), code Morse, etc.

Définition

Régularité. *Un code est dit régulier, ou encore non-singulier si tous les mots de code sont distincts.*

Déchiffrabilité. *Un code régulier est dit déchiffrable, ou encore à décodage unique, si toute suite de mots de code ne peut être interprétée que de manière unique.*

> **Longueur fixe.** *Avec des mots de longueur fixe, on peut décoder tout message sans ambiguïté.*
>
> **Séparateur.** *On consacre un symbole de l'alphabet de destination comme séparateur de mot.*
>
> **Sans préfixe.** *On évite qu'un mot du code soit identique au début d'un autre mot. Un tel code est qualifié de code instantané.*

Exemple 2.	code A	code B	code C	code D	code E	code F	code G
s_1	1	0	00	0	0	0	0
s_2	1	10	11	10	01	10	10
s_3	0	01	10	11	011	110	110
s_4	0	11	01	110	0111	1110	111

Vers le premier théorème de Shannon

Inégalité de Kraft

On se propose de construire des codes déchiffrables, et plus particulièrement instantanés, aussi économiques que possible. L'inégalité de Kraft fournit une condition nécessaire et suffisante d'existence de codes instantanés.

Théorème (Inégalité de Kraft)

On note $n_1, \ldots, n_Q$ les longueurs des mots candidats pour coder les Q états d'une source dans un alphabet q-aire. Une condition nécessaire et suffisante d'existence d'un code instantané ayant ces longueurs de mots est donnée par :

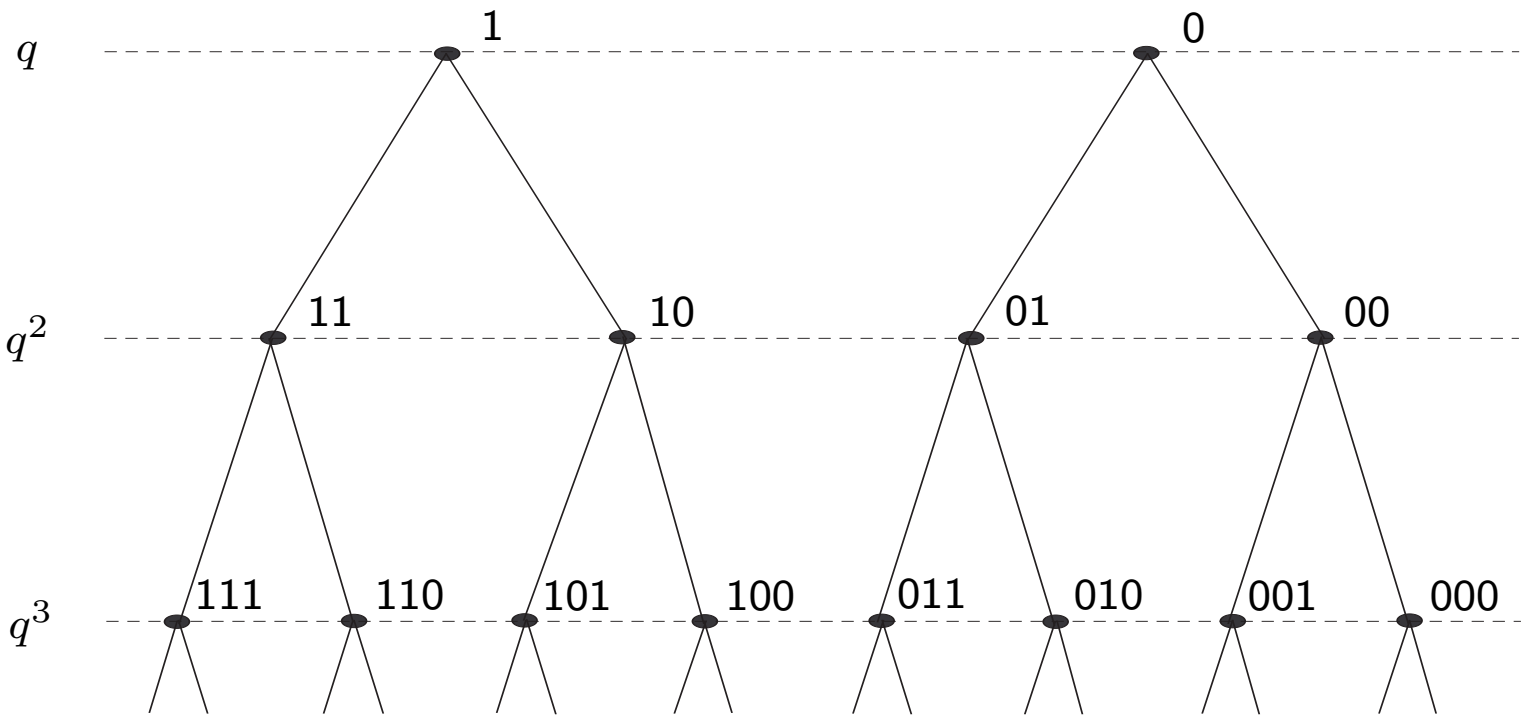
$$\sum_{i=1}^{Q} q^{-n_i} \leq 1.$$

Remarque. La même condition nécessaire et suffisante a été établie par McMillan pour les codes déchiffrables, antérieurement à l'inégalité de Kraft.

Vers le premier théorème de Shannon

Inégalité de Kraft

Preuve. La représentation graphique suivante, dans le cas d'un code binaire, rend la démonstration plus aisée.



Vers le premier théorème de Shannon

Inégalité de Kraft

On pose $n_1 \leq \ldots \leq n_Q$ et on considère un arbre q-aire de profondeur n_Q, comportant donc q^{n_Q} sommets terminaux.

Condition nécessaire. La condition du préfixe impose qu'un mot de longueur n_i exclut $q^{n_Q - n_i}$ sommets terminaux. Le nombre total de sommets exclus vaut donc :

$$\sum_{i=1}^{Q} q^{n_Q - n_i} \leq q^{n_Q}.$$

Condition suffisante. On sélectionne d'abord un nœud à la profondeur n_1, ce qui exclut $q^{n_Q - n_1}$ sommets terminaux. Il en existe toutefois encore car l'inégalité de Kraft entraîne $q^{n_Q - n_1} < q^{n_Q}$. Sur le trajet menant à l'un des sommets terminaux non-exclus, on sélectionne un nœud à la profondeur n_2...

Vers le premier théorème de Shannon

Inégalité de McMillan

L'inégalité de Kraft induit le caractère suffisant de l'inégalité de McMillan puisque tout code à préfixe est déchiffrable.

Condition nécessaire de l'inégalité de McMillan. On développe l'expression suivante selon

$$\left(\sum_{k=1}^{s} r_k\, q^{-k}\right)^m = \sum_{n=m}^{ms} \nu(n)\, q^{-n}$$

où $\nu(n) = \sum_{i_1+\dots i_m=n} r_{i_1} \dots r_{i_m}$. En interprétant r_k comme le nombre de mots de longueur k du code, $\nu(n)$ désigne le nombre de texte de longueur n. La condition de déchiffrabilité implique que $\nu(n) \leq q^n$. On a donc

$$\sum_{k=1}^{s} r_k\, q^{-k} \leq (m\, s)^{\frac{1}{m}},$$

ce qui conduit au résultat en passant à la limite.

Vers le premier théorème de Shannon

Code complet

Définition (Code complet)

Un code est dit complet s'il vérifie la relation

$$\sum_{i=1}^{Q} q^{-n_i} = 1.$$

Exemple	code A	code B	code C
s_1	00	0	0
s_2	01	100	10
s_3	10	110	110
s_4	11	111	11
$\sum_{i=1}^{4} 2^{-n_i}$	1	7/8	9/8

Vers le premier théorème de Shannon

Inégalité de McMillan

Corollaire (Conséquences de l'inégalité de McMillan)

Soit S une source sans mémoire à Q états. Soit p_i la probabilité d'apparition de s_i, auquel est associé un mot de code déchiffrable q-aire de longueur n_i. En posant $q_i = \frac{q^{-n_i}}{\sum_{j=1}^{Q} q^{-n_j}}$, puis en appliquant l'inégalité de Gibbs à p_i et q_i, on obtient alors

$$\sum_{i=1}^{Q} p_i \log \frac{1}{p_i} + \sum_{i=1}^{Q} p_i \log q^{-n_i} \leq \log \sum_{i=1}^{Q} q^{-n_i}.$$

En appliquant le théorème de McMillan au dernier membre de l'inégalité, il en résulte

$$H(S) - \overline{n} \log q \leq \log \sum_{i=1}^{Q} q^{-n_i} \leq 0,$$

où $\overline{n} = \sum_{i=1}^{Q} p_i \, n_i$ représente la longueur moyenne des mots du code.

Vers le premier théorème de Shannon

Conséquences de l'inégalité de McMillan

Théorème

La longueur moyenne $\overline{n}$ des mots de tout code déchiffrable est bornée inférieurement

$$\frac{H(S)}{\log q} \leq \overline{n}.$$

Condition d'égalité. L'inégalité ci-dessus se transforme en égalité à condition que $\sum_{i=1}^{Q} q^{-n_i} = 1$, c'est-à-dire si $p_i = q^{-n_i}$. Ceci signifie que

$$n_i = \frac{\log \frac{1}{p_i}}{\log q}.$$

Définition

Un code dont la longueur de chaque mot est telle que $n_i = \frac{\log \frac{1}{p_i}}{\log q}$ est dit absolument optimum.

Vers le premier théorème de Shannon

Conséquences de l'inégalité de McMillan

La condition d'égalité précédente n'est généralement pas vérifiée. Il est cependant possible de constituer un code tel que

$$\frac{\log \frac{1}{p_i}}{\log q} \leq n_i < \frac{\log \frac{1}{p_i}}{\log q} + 1.$$

En multipliant par p_i et en sommant sur i, ceci signifie que

$$\frac{H(S)}{\log q} \leq \overline{n} < \frac{H(S)}{\log q} + 1.$$

Définition (Codes compact et de Shannon)

Un code dont la longueur moyenne des mots vérifie la double inégalité présentée ci-dessus est dit compact. Plus particulièrement, on parle de code de Shannon lorsque

$$n_i = \left\lceil \frac{\log \frac{1}{p_i}}{\log q} \right\rceil.$$

Premier théorème de Shannon

Énoncé et démonstration

Les bornes qui viennent d'être établies vont nous permettre de démontrer le premier théorème de Shannon, qui s'énonce ainsi :

Théorème

Pour toute source stationnaire, il existe un procédé de codage déchiffrable où la longueur moyenne des mots est aussi voisine que l'on veut de sa borne inférieure.

Preuve pour une source sans mémoire. On considère la $k^{\text{ème}}$ extension de la source S. Dans le cas d'une source sans mémoire

$$\frac{kH(S)}{\log q} \leq \overline{n}_k < \frac{kH(S)}{\log q} + 1.$$

Dans cette expression, $\overline{n}_k$ désigne la longueur moyenne des mots de code utilisés dans le cadre de la $k^{\text{ème}}$ extension de S. On divise par k et on passe à la limite.

Premier théorème de Shannon

Énoncé et démonstration

Preuve pour une source stationnaire. On considère la $k^{\text{ème}}$ extension de la source S. Dans le cas d'une source sans mémoire

$$\frac{H(S_1,\ldots,S_k)}{k\,\log q} \leq \frac{\overline{n}_k}{k} < \frac{H(S_1,\ldots,S_k)}{k\,\log q} + \frac{1}{k}.$$

Dans cette expression, $\overline{n}_k$ désigne la longueur moyenne des mots de code utilisés dans le cadre de la $k^{\text{ème}}$ extension de S.
Dans le cas d'une source stationnaire, on sait que $\lim_{k\to\infty} H(S_1,\ldots,S_k)$ existe. En reprenant la notation conventionnelle H_0 de cette limite, on aboutit à

$$\lim_{k\to\infty} \frac{\overline{n}_k}{k} = \frac{H_0}{\log q}.$$

Remarque. D'un point de vue pratique, l'intérêt du Premier Théorème de Shannon est limité.

Techniques de codage binaire

Méthode directe

Le premier théorème de Shannon exprime une propriété asymptotique du langage, mais ne fournit aucune méthode pratique pour y parvenir.
Une technique de codage directe consiste à associer à chaque état de la source un nombre de symboles n_i tel que

$$n_i = \left\lceil \frac{\log \frac{1}{p_i}}{\log q} \right\rceil .$$

Remarque. Le code obtenu est un code de Shannon.

Techniques de codage binaire

Méthode directe

Le premier théorème de Shannon exprime une propriété asymptotique du langage, mais ne fournit aucune méthode pratique pour y parvenir.
Une technique de codage directe consiste à associer à chaque état de la source un nombre de symboles n_i tel que

$$n_i = \left\lceil \frac{\log \frac{1}{p_i}}{\log q} \right\rceil.$$

Remarque. Le code obtenu est un code de Shannon.

Exemple. On considère un système à 5 états $\{s_1, \ldots, s_5\}$ définis par les probabilités :

$$\begin{array}{lll} p_1 = 0.35 & -\log_2 p_1 = 1.51 & \longrightarrow \quad n_1 = 2 \\ p_2 = 0.22 & -\log_2 p_2 = 2.18 & \longrightarrow \quad n_2 = 3 \\ p_3 = 0.18 & -\log_2 p_3 = 2.47 & \longrightarrow \quad n_3 = 3 \\ p_4 = 0.15 & -\log_2 p_4 = 2.73 & \longrightarrow \quad n_4 = 3 \\ p_5 = 0.10 & -\log_2 p_5 = 3.32 & \longrightarrow \quad n_5 = 4. \end{array}$$

Il est aisé d'obtenir un code instantané vérifiant la condition précédente sur les n_i à l'aide d'un arbre. On obtient par exemple :

$$s_1 : 00 \qquad s_2 : 010 \qquad s_3 : 011 \qquad s_4 : 100 \qquad s_5 : 1010.$$

On aboutit à $\overline{n} = 2.75$, à comparer à $H(S) = 2.19$ Sh/symb.

Techniques de codage binaire

Code de Shannon-Fano

Le code de Shannon-Fano est le premier code à avoir exploité la redondance d'une source. On en expose à présent le principe.

1. Ranger les états du système par probabilités décroissantes.
2. Subdiviser les états du système en 2 groupes G_0 et G_1 de probabilités voisines, *sans modifier l'ordre* dans lequel ils ont été rangés en 1.
3. Chaque groupe G_i est subdivisé en 2 sous-groupes G_{i0} et G_{i1} de probabilités aussi voisines que possibles, une fois encore *sans modifier l'ordre* des états.
4. La procédure s'arrête lorsque chaque sous-groupe est constitué d'un unique élément. L'indice du groupe donne le mot de code.

Techniques de codage binaire

Code de Shannon-Fano

Le code de Shannon-Fano est le premier code à avoir exploité la redondance d'une source. On en expose à présent le principe.

1. Ranger les états du système par probabilités décroissantes.
2. Subdiviser les états du système en 2 groupes G_0 et G_1 de probabilités voisines, *sans modifier l'ordre* dans lequel ils ont été rangés en 1.
3. Chaque groupe G_i est subdivisé en 2 sous-groupes G_{i0} et G_{i1} de probabilités aussi voisines que possibles, une fois encore *sans modifier l'ordre* des états.
4. La procédure s'arrête lorsque chaque sous-groupe est constitué d'un unique élément. L'indice du groupe donne le mot de code.

Exemple. Pour élaborer un code de Shannon-Fano, on procède ainsi :

état	p_i	étape 1	étape 2	étape 3	code
s_1	0.35	0	0		00
s_2	0.22	0	1		01
s_3	0.18	1	0		10
s_4	0.15	1	1	0	110
s_5	0.10	1	1	1	111

On aboutit à $\overline{n} = 2.25$, à comparer à $H(S) = 2.19$ Sh/symb.

Techniques de codage binaire

Code de Huffman

La méthode de Huffman fournit un code instantané compact de longueur moyenne minimale. Pour y parvenir, elle exploite la propriété suivante.

Lemme

Pour toute source, il existe un code instantané de longueur moyenne minimale satisfaisant les propriétés suivantes.

1. *Si* $P(S = s_i) > P(S = s_j)$, *alors* $n_i \leq n_j$.
2. *Les deux mots les plus longs, donc associés aux états les moins probables, ont même longueur et ne diffèrent que d'un bit.*

La méthode de Huffman consiste à regrouper les deux états les moins probables, puis à les traiter comme un seul en sommant leur probabilité. Cette technique est alors réitérée sur les états restants, jusqu'à ce qu'il n'en reste que deux.

Techniques de codage binaire

Code de Huffman

On construit un arbre en partant des feuilles les plus profondes, qui représentent les états de la source.

1. A chaque étape, on fusionne les feuilles les moins probables en une seule.
2. La procédure s'arrête lorsque on aboutit à une feuille unique constituée de tous les symboles.
3. Le parcours inverse de l'arbre fournit les mots du code.

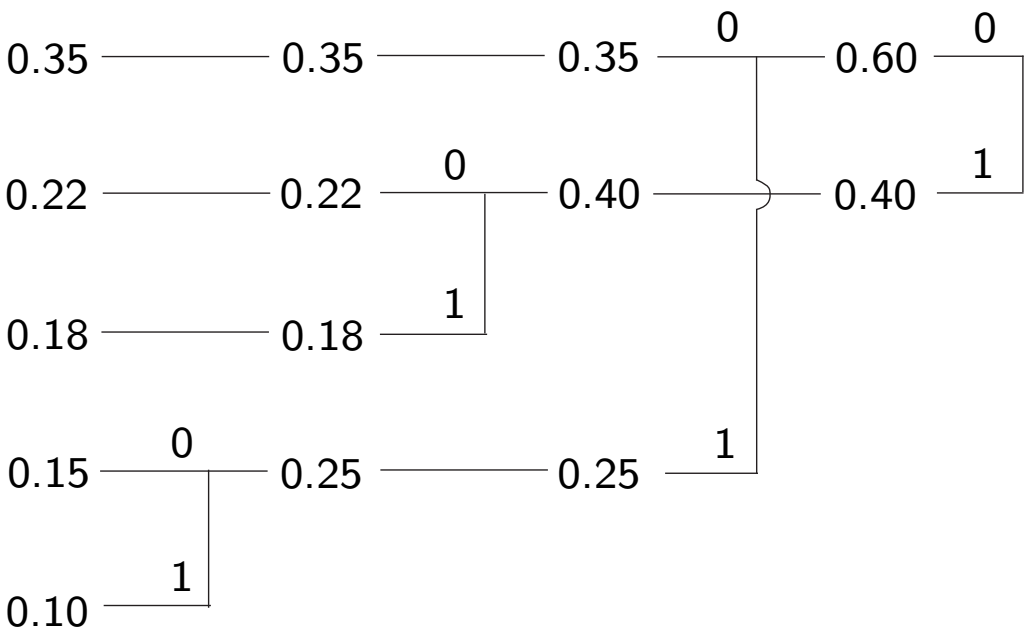
Techniques de codage binaire

Code de Huffman

On construit un arbre en partant des feuilles les plus profondes, qui représentent les états de la source.

1. A chaque étape, on fusionne les feuilles les moins probables en une seule.
2. La procédure s'arrête lorsque on aboutit à une feuille unique constituée de tous les symboles.
3. Le parcours inverse de l'arbre fournit les mots du code.

Exemple.



Finalement, le parcours inverse de l'arbre fournit le résultat suivant :

$$s_1 : 00 \qquad s_2 : 10 \qquad s_3 : 11 \qquad s_4 : 010 \qquad s_5 : 011.$$

On aboutit à $\overline{n} = 2.25$, à comparer à $H(S) = 2.19$ Sh/symb.

Outline

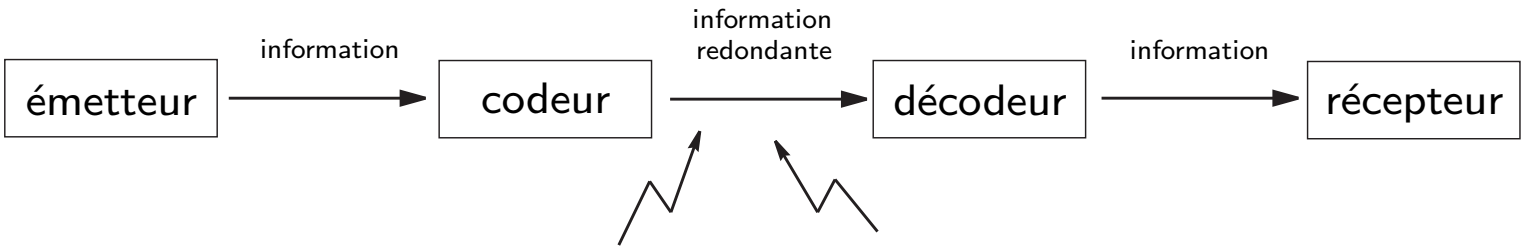
Codage de canal

Chapitre 3

Codage de canal

Motivations

Dans un système réel, le message reçu par le destinataire peut différer de celui qui a été émis par la source en raison de perturbations. On parle de *canal bruyant*.



Le codage de canal vise à introduire de la redondance dans le message $\longrightarrow$ compenser l'érosion de l'information due au canal.

Modèles de canal discret

Modèle général

Un canal discret est un système stochastique acceptant en entrée des suites de symboles définies sur un alphabet $\mathcal{X}$, et émettant en sortie des suites de symboles définies sur un alphabet de sortie $\mathcal{Y}$.

Entrées et sorties sont liées par un modèle probabiliste :

$$P(Y_1 = y_1, \ldots, Y_m = y_m | X_1 = x_1, \ldots, X_n = x_n)$$

▷ **modèle trop général pour donner lieu à des développements simples**

Modèles de canal discret

Propriétés

Par souci de simplification, on fait des hypothèses sur le modèle de canal.

Propriété (Canal causal)

Un canal est dit causal si

$$P(Y_1 = y_1, \ldots, Y_m = y_m | X_1 = x_1, \ldots, X_n = x_n) = P(Y_1 = y_1, \ldots, Y_m = y_m | X_1 = x_1, \ldots, X_m = x_m)$$

quels que soient m et n tels que $m \leq n$.

Conséquence. En sommant les 2 membres de l'égalité sur $Y_1, \ldots, Y_{m-1}$, on vérifie

$$P(Y_m = y_m | X_1 = x_1, \ldots, X_n = x_n) = P(Y_m = y_m | X_1 = x_1, \ldots, X_m = x_m)$$

$\longrightarrow$ toute sortie est indépendante des entrées futures

Modèles de canal discret

Propriétés

On peut être amené à faire l'hypothèse suivante sur le comportement du canal.

Propriété (Canal causal sans mémoire)

On dit qu'un canal causal est sans mémoire si, pour tout $k \geq 2$, on a :

$$P(Y_k = y_k | X_1 = x_1, \ldots, X_k = x_k, Y_1 = y_1, \ldots, Y_{k-1} = y_{k-1}) = P(Y_k = y_k | X_k = x_k).$$

Conséquence. La loi conditionnelle gouvernant le comportement du canal est entièrement déterminée par les lois conditionnelles instantanées :

$$P(Y_1 = y_1, \ldots, Y_m = y_m | X_1 = x_1, \ldots, X_n = x_n) = \prod_{k=1}^{m} P(Y_k = y_k | X_k = x_k).$$

$\longrightarrow P(Y_k = y_k | X_k = x_k)$ dépend éventuellement du temps

Modèles de canal discret

Propriétés

En remarquant que $P(Y_k = y_k | X_k = x_k)$ peut éventuellement dépendre du temps k, on est amené à introduire la propriété suivante.

Propriété (Canal sans mémoire stationnaire)

On dit d'un canal sans mémoire qu'il est stationnaire si, quel que soit $k \geq 1$, on a :

$$P(Y_k = y_k | X_k = x_k) = P(Y = y_k | X = x_k).$$

Notation. On note $(\mathcal{X}, \mathcal{Y}, \Pi)$ un canal discret sans mémoire, où Π est la matrice de transition définie par :

$$\Pi(i, j) = P(Y = y_j | X = x_i)$$

Modèles de canal discret

Propriétés

En remarquant que $P(Y_k = y_k | X_k = x_k)$ peut éventuellement dépendre du temps k, on est amené à introduire la propriété suivante.

Propriété (Canal sans mémoire stationnaire)

On dit d'un canal sans mémoire qu'il est stationnaire si, quel que soit $k \geq 1$, on a :

$$P(Y_k = y_k | X_k = x_k) = P(Y = y_k | X = x_k).$$

Notation. On note $(\mathcal{X}, \mathcal{Y}, \Pi)$ un canal discret sans mémoire, où Π est la matrice de transition définie par :

$$\Pi(i, j) = P(Y = y_j | X = x_i)$$

Définition (canal symétrique)

Un canal est dit symétrique si les lignes de sa matrice de transition sont formées des mêmes éléments à l'ordre près, tout comme ses colonnes.

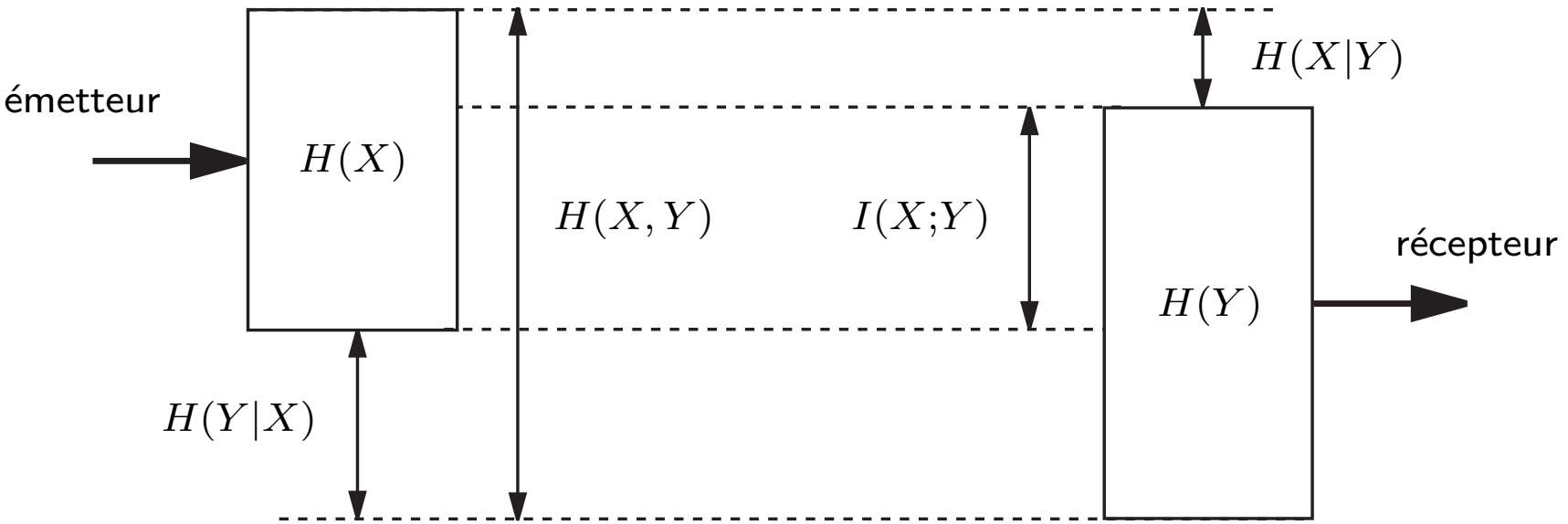
Exemples. Les matrices de transition suivantes sont celles de canaux symétriques.

$$\Pi = \begin{pmatrix} p & q & 1-p-q \\ q & 1-p-q & p \\ 1-p-q & p & q \end{pmatrix}, \qquad \Pi = \begin{pmatrix} p & 1-p-q & q \\ q & 1-p-q & p \end{pmatrix},$$

où p et q sont des éléments de l'intervalle $[0, 1]$.

Capacité d'un canal sans mémoire

Présentation intuitive



- $H(X)$ est la quantité d'information transmise par un canal sans bruit
- $H(X|Y)$ est l'information requise pour supprimer l'ambiguïté sur l'entrée
- $I(X;Y)$ est la quantité d'information transmise par le canal bruité

Capacité d'un canal sans mémoire

Définition

Définition

On définit la capacité en information par symbole d'un canal par :

$$C \triangleq \max_{P(X=x)} I(X;Y).$$

Précaution. On vérifie que $I(X,Y)$ est une fonction concave de la loi de X. En notant $f(x) = -x \log x$, on note qu'il s'agit d'une somme de fonctions concaves :

$$\begin{aligned} I(X;Y) &= \sum_i \sum_j p(i,j) \log \frac{p(i,j)}{p(i)\,p(j)} \\ &= \sum_i \sum_j p_i\, p_i(j) \log \frac{p_i(j)}{\sum_i p_i\, p_i(j)} \\ &= \sum_i p_i \left(\sum_j p_i(j) \log p_i(j) \right) + \sum_j f\left(\sum_i p_i p_i(j) \right). \end{aligned}$$

Capacité d'un canal sans mémoire

Calculs de capacités

Dans le cas général, le calcul direct de la capacité d'un canal s'avère compliqué. Toutefois, dans le cas d'un canal symétrique, le calcul s'effectue aisément.

Théorème

La capacité d'un canal symétrique $(\mathcal{X}, \mathcal{Y}, \Pi)$ est égale à $I(X;Y)$ dans le cas où l'entrée X suit une loi uniforme.

Démonstration.

L'entropie $H(Y|X = x_i) = -\sum_j p_i(j) \log p_i(j)$ est indépendante de i, les lignes i de Π étant formées des mêmes éléments : $H(Y|X)$ est donc indépendant de la loi de X. On vérifie aisément que Y suit une loi uniforme si celle de X l'est. En effet :

$$p_j = \sum_i p_i \, p_i(j) = \frac{1}{q} \sum_i p_i(j)$$

est indépendant de j car les colonnes de Π sont constituées des mêmes termes. □

Calculs de capacités

Exemples

Canal binaire sans bruit. Ce canal reproduit en sortie le symbole d'entrée. En conséquence, on a $I(X;Y) = H(X)$ car $H(X|Y) = 0$.

$$C = 1 \text{ Sh/symb}$$

Canal binaire en dysfonctionnement. Ce canal reproduit en sortie toujours le même symbole, indépendamment de l'entrée. En conséquence, l'information mutuelle $I(X;Y)$ est nulle puisque $H(Y) = H(Y|X) = 0$.

$$C = 0 \text{ Sh/symb}$$

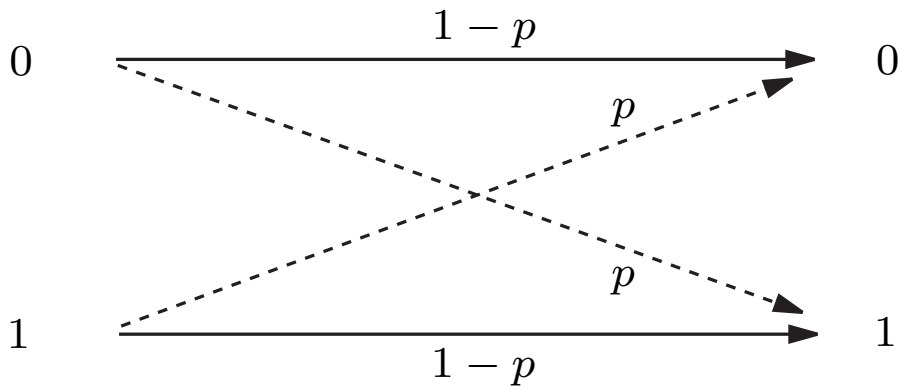
Calculs de capacités

Exemple du canal binaire symétrique

Le canal binaire symétrique est l'exemple le plus simple de canal bruyant. Sa matrice de transition est donnée par

$$\Pi = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$$

que l'on représente schématiquement ainsi



Afin d'évaluer la capacité en information de ce canal, calculons préalablement l'information mutuelle moyenne $I(X;Y)$:

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - P(X=0)H(Y|X=0) - P(X=1)H(Y|X=1). \end{aligned}$$

Or, un calcul simple permet de montrer que $H(Y|X=x) = H_2(p)$, avec $x \in \{0,1\}$, ce qui entraîne que :

$$I(X;Y) = H(Y) - H_2(p) \leq \log 2 - H_2(p).$$

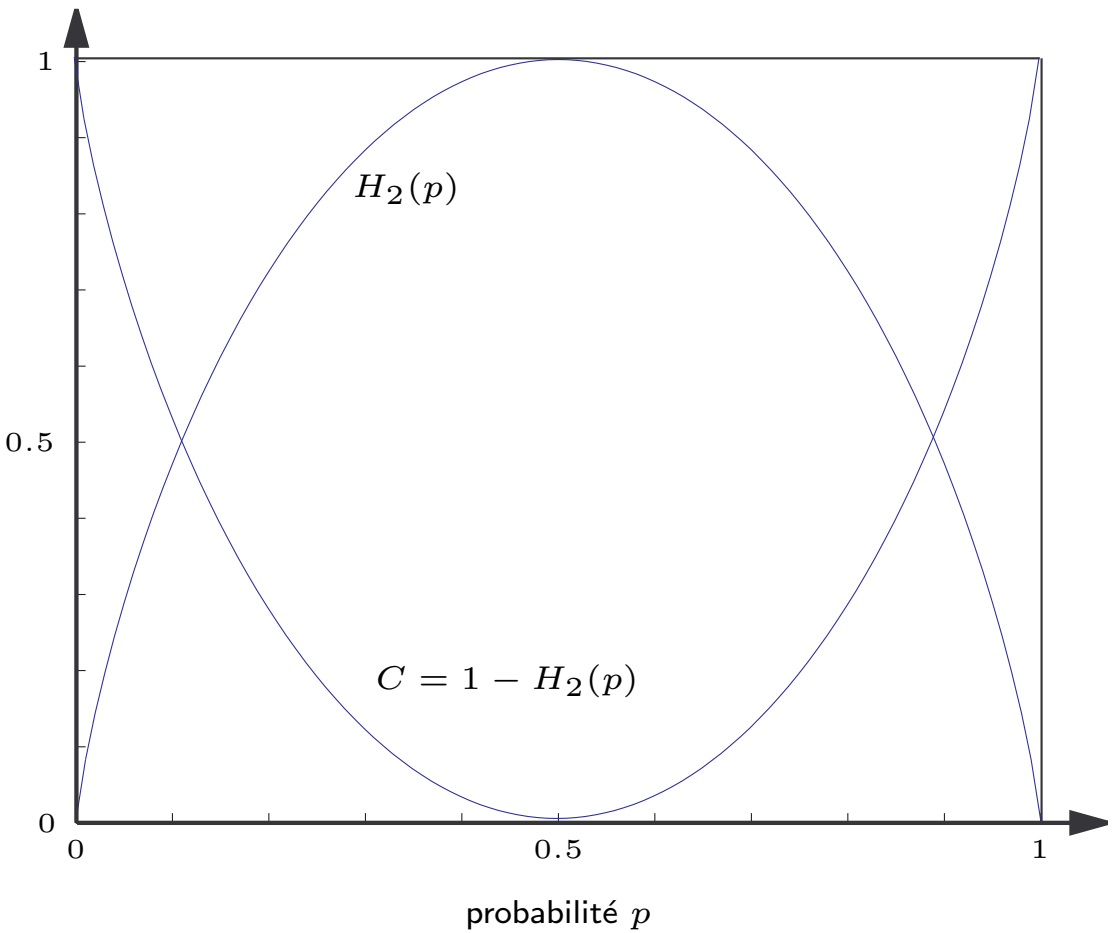
Calculs de capacités

Exemple du canal binaire symétrique

En conséquence, la capacité d'un canal binaire symétrique est donnée par :

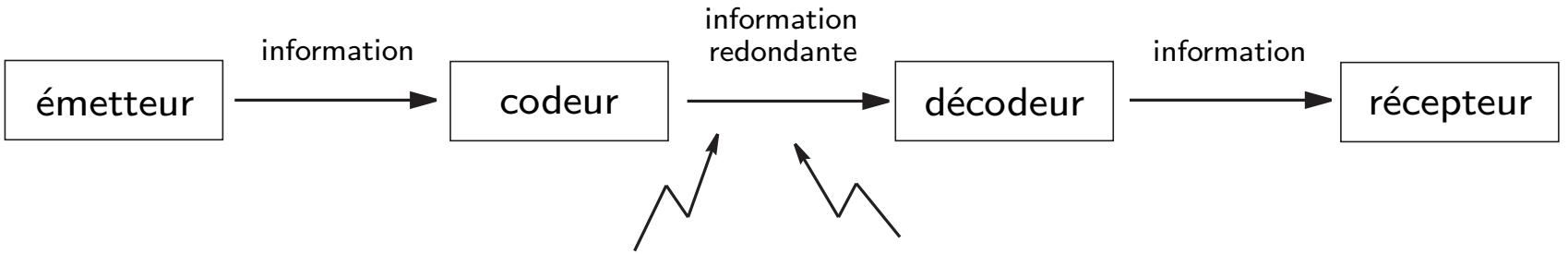
$$C = 1 - H_2(p) \text{ Sh/symb}$$

La capacité en information d'un canal binaire symétrique a pour allure :



Codage de canal

Définitions préalables



Afin de détecter et/ou corriger les erreurs transmises, il est nécessaire d'ajouter des *symboles de contrôle* selon une règle $\mathcal{C}$, appelée *codage*.

▷ **le décodeur vérifie si la séquence reçue respecte $\mathcal{C}$**

Usage de redondance. On utilise des blocs de n symboles afin de transmettre k symboles d'information, avec $k < n$. Chaque bloc de longueur n est dit *mot du code*.

Codage de canal

Définitions préalables

Définition

Soit $\mathcal{A} = \{a_1, \ldots, a_q\}$ un ensemble fini dit alphabet du code. Soit $\mathcal{A}^n$ l'ensemble de toutes les chaînes de longueur n sur $\mathcal{A}$. Tout sous-ensemble non vide $\mathcal{C}$ de $\mathcal{A}^n$ est dit code en bloc q-aire. Chaque chaîne dans $\mathcal{C}$ sera dite mot du code.

Définition

Si $\mathcal{C} \subset \mathcal{A}^n$ contient M mots du code, on dit alors que $\mathcal{C}$ est de longueur n et de taille M. On parle alors de (n, M)-code.

Exemple. Le code $\mathcal{C}$ suivant est un (5,4)-code :

$$\mathcal{C} = \{11100, 01001, 10010, 00111\}$$

Codage de canal

Erreurs de détection

Définition

On parle d'erreur de détection lorsque le mot $c \in \mathcal{C}$ a été émis et que l'on reçoit le mot d, avec $c \neq d$.

Toute erreur de transmission ne peut être détectée que si le mot reçu n'est pas un autre mot du code. En conséquence, si $c \in \mathcal{C}$ est émis, on a :

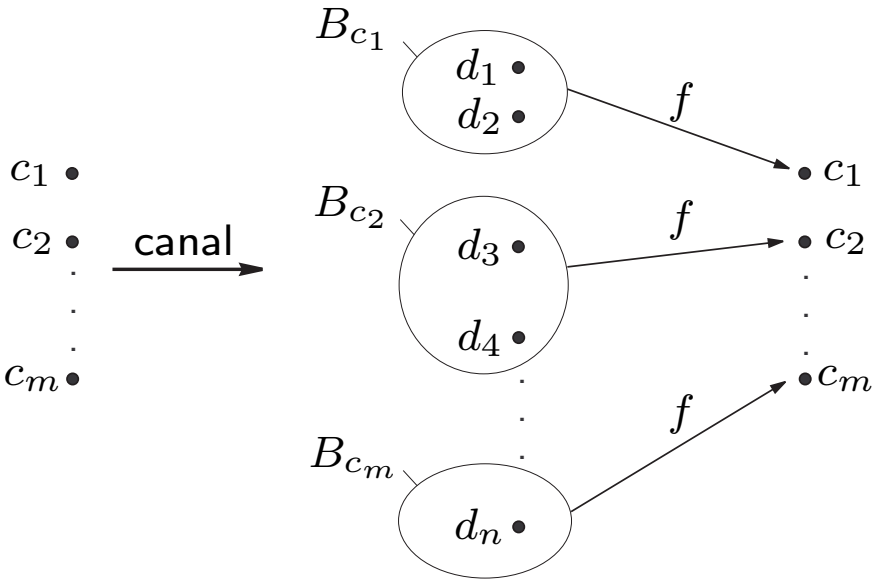
$$P(\text{erreur non détectée} \mid c \text{ est émis}) = \sum_{\substack{d \in \mathcal{C} \\ d \neq c}} P(d|c).$$

$$P(\text{erreur non détectée}) = \sum_{c \in \mathcal{C}} \sum_{\substack{d \in \mathcal{C} \\ d \neq c}} P(d|c)\ P(c).$$

Codage de canal

Erreurs de décision

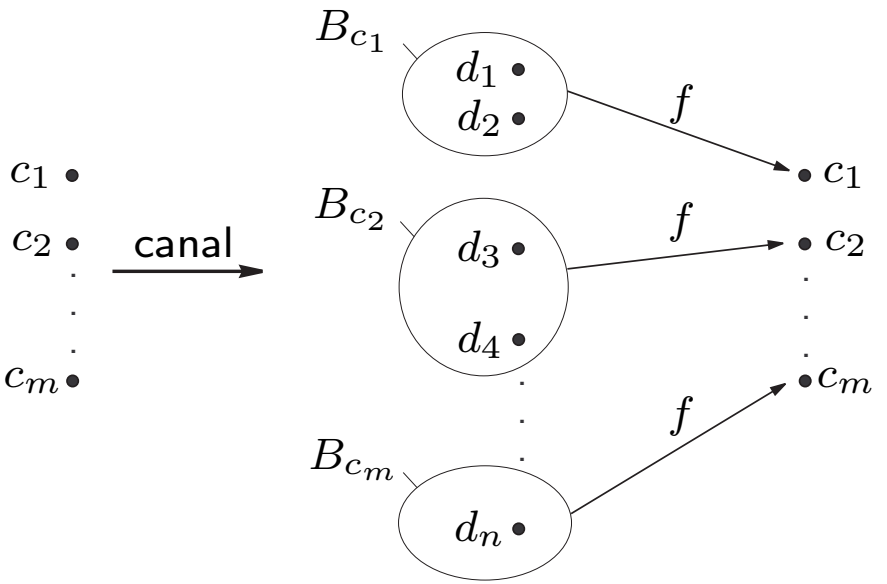
Définition :
Un schéma de décision est une fonction partielle f de l'ensemble des chaînes reçues vers l'ensemble des mots du code.



Codage de canal

Erreurs de décision

Définition :
Un schéma de décision est une fonction partielle f de l'ensemble des chaînes reçues vers l'ensemble des mots du code.



Définition

On parle d'erreur de décision lorsque le mot $c \in \mathcal{C}$ a été émis, que d a été reçu et qu'il a été décodé par $f(d) \neq c$.

La probabilité d'une erreur de décision sachant que c a été émis est définie par

$$P(\text{erreur de décodage} \mid c \text{ est émis}) = \sum_{\substack{d \in \mathcal{C} \\ d \notin f^{-1}(c)}} P(d|c),$$

et la probabilité d'une erreur de décodage est

$$P(\text{erreur de décodage}) = \sum_{c \in \mathcal{C}} P(\text{erreur de décodage} \mid c \text{ est émis})\, P(c).$$

Codage de canal

Second théorème de Shannon

En rappelant que la capacité d'un canal est la quantité maximale d'information qu'il peut transmettre, on peut énoncer le second théorème de Shannon.

Théorème (second théorème de Shannon)

Soit un canal discret et sans mémoire de capacité C. Pour tout nombre positif C' inférieur à C, il existe une suite $\mathcal{C}_k$ de codes r-aires associés aux schémas de décision f_k ayant les propriétés suivantes :

- *$\mathcal{C}_k$ est un code de longueur k et de taux de transmission supérieur ou égal à C' ;*
- *la probabilité max. d'erreur de décodage tend vers 0 lorsque k tend vers l'infini :*

$$\lim_{k \to +\infty} P_{max}(k) = 0,$$

avec $P_{max}(k) = \max_{c \in \mathcal{C}_k} P(\text{erreur de décodage} \mid c \text{ est émis})$.

▷ **aucune preuve constructive de ce théorème n'est connue à ce jour**

Codage de canal

Définition d'une métrique sur $\mathcal{C}$

Les codes détecteur/correcteur reposent sur une structure algébrique/géométrique.

Définition

Soient x et y des chaînes de même longueur sur le même alphabet. La distance de Hamming $d_{Ham}(x, y)$ entre x et y est par définition le nombre de positions pour lesquelles x et y diffèrent.

Exemple : $d_{Ham}(10112, 20110) = 2$

Codage de canal

Définition d'une métrique sur $\mathcal{C}$

Les codes détecteur/correcteur reposent sur une structure algébrique/géométrique.

Définition

Soient x et y des chaînes de même longueur sur le même alphabet. La distance de Hamming $d_{Ham}(x, y)$ entre x et y est par définition le nombre de positions pour lesquelles x et y diffèrent.

Exemple : $d_{Ham}(10112, 20110) = 2$

Théorème

L'espace $(\mathcal{A}^n, d_{Ham})$ est un espace métrique, autrement dit la distance de Hamming vérifie les propriétés suivantes pour tout x, y et z de $\mathcal{A}^n$:

1. $d_{Ham}(x, y) = 0 \iff x = y$
2. $d_{Ham}(x, y) = d_{Ham}(y, x)$
3. $d_{Ham}(x, y) \leq d_{Ham}(x, z) + d_{Ham}(z, y)$.

Codage de canal

Décodage par maximum de vraisemblance

On considère un canal binaire symétrique caractérisé par :

$$P(1|0) = P(0|1) = p \quad P(0|0) = P(1|1) = 1 - p \quad \text{avec :} \quad p < 0.5.$$

En notant c le mot envoyé et d le mot reçu, $d_{Ham}(c, d)$ correspond au nombre d'erreurs de symboles dues au canal. En conséquence :

$$P(d|c) = p^{d_{Ham}(c,d)} \ (1-p)^{n - d_{Ham}(c,d)}.$$

▷ **$P(d|c)$ est maximale lorsque $d_{Ham}(c, d)$ est minimale**

Théorème

Pour le canal binaire symétrique avec une probabilité d'erreur $p < 0.5$, la règle de décodage par maximum de vraisemblance est équivalente à la règle de décodage par minimum de distance.

Codage de canal

Distance minimale d'un code

Dans l'idée de pouvoir utiliser le décodage par minimum de distance, on est amené à poser les définitions suivantes.

Définition

La distance minimale du code $\mathcal{C}$ est définie par

$$d(\mathcal{C}) = \min_{x,y \in \mathcal{C}} d_{Ham}(x,y).$$

Définition

On parle de (n, M, d)-code pour évoquer un code de longueur n, de taille M et de distance minimale d.

Exemple : le code binaire $\mathcal{C} = \{11100, 01001, 10010, 00111\}$ est un (5,4,3)-code.

Codage de canal

Codes t-détecteurs d'erreurs

On définit un code t-détecteur d'erreurs ainsi.

Définition

Un code $\mathcal{C}$ est t-détecteur d'erreurs si, dès qu'au plus $t \geq 1$ erreurs se produisent dans un mot du code, le mot résultant n'est pas un mot du code.
Le code $\mathcal{C}$ est dit exactement t-détecteur lorsqu'il est t-détecteur mais pas $(t+1)$-détecteur.

On démontre aisément le résultat suivant :

Théorème

Un code $\mathcal{C}$ est exactement t-détecteurs d'erreurs si et seulement si

$$d(\mathcal{C}) = t + 1.$$

Codage de canal

Codes t-correcteurs d'erreurs

On définit un code t-correcteur d'erreurs ainsi.

Définition

Un code $\mathcal{C}$ est t-correcteur si le décodage par minimum de distance peut corriger les erreurs de taille inférieure ou égale à t dans tout mot du code.
Un code est dit exactement t-correcteur s'il est t-correcteur mais pas $(t+1)$-correcteur. Ceci signifie que toute erreur de taille t est corrigées mais qu'il existe au moins une erreur de taille $t+1$ qui est décodée incorrectement.

On démontre le résultat suivant :

Théorème

Un code $\mathcal{C}$ est exactement t-correcteur d'erreurs si et seulement si

$$d(\mathcal{C}) = 2t+1 \qquad \textit{ou} \qquad d(\mathcal{C}) = 2t+2$$

Outline

Éléments d'algèbre discrète

Chapitre 4

Algèbre discrète

Mise en oeuvre pour le codage de canal

D'après le second théorème de Shannon, le codage de canal permet de réduire arbitrairement les probabilités d'erreur de transmission. Cependant, aucune preuve constructive de celui-ci n'existe.

▷ **De multiples techniques de codage ont été proposées**

Les héritages d'algèbre discrète sont nombreux :

- représentation vectorielle des mots
- addition, produit
- distance
- ...

Structures algébriques

Groupes

Définition (loi de composition interne)

On appelle **loi de composition interne** *à un ensemble* E *une application* $\top$ *de* $E \times E$ *dans* E *:*

$$(x, y) \longmapsto z = x \top y.$$

On appelle **élément neutre** *de la loi* $\top$ *un élément* e *de* E *tel que :*

$$\forall x \in E, \quad x \top e = e \top x = x.$$

La loi $\top$ *est* **commutative** *si :*

$$\forall\, (x, y) \in E \times E, \quad x \top y = y \top x.$$

Structures algébriques

Groupes

Définition (groupe)

Un ensemble G muni d'une loi de composition interne $\top$ est appelé **groupe** *si :*

- *la loi $\top$ est* **associative**, *c'est-à-dire*

$$\forall\,(x,y,z)\in G^3,\quad (x\top y)\top z = x\top(y\top z)\,,$$

- *la loi $\top$ admet un* **élément neutre** *e dans G,*
- *tout élément de G admet un* **inverse** *pour la loi $\top$, c'est à dire*

$$\forall x\in G,\quad \exists x'\in G\quad :\quad x\top x' = x'\top x = e.$$

Notation. Lorsque la loi $\top$ est notée multiplicativement par $\times$, x' est souvent représenté par x^{-1} et e par 1. Lorsque la loi est notée additivement par $+$, x' est souvent représenté par $-x$ et e par 0.

Structures algébriques

Groupes

Définition

On peut distinguer les différents types de groupes suivants.

- *Si la loi $\top$ est commutative, le groupe $(G, \top)$ est dit* **abélien** *ou* **commutatif**.
- *S'il existe un élément x de G tel que*

$$\forall y \in G, \quad \exists i \in \mathbb{N} \quad : \quad y = x^i,$$

le groupe est dit **cyclique**. *On note $< x >$ le générateur du groupe. Ci-dessus, nous avons utilisé la notation $x^0 = e$ et, pour $i \neq 0$,*

$$x^i = \underbrace{x \top x \top \ldots \top x}_{i \text{ fois}}.$$

Définition

Soit G un groupe fini et $x \in G$. On appelle **ordre** *de x le plus petit entier positif n tel que $x^n = e$.*

Structures algébriques

Groupes

Exemple. $(\mathbb{Z}, +)$ est un groupe abélien.

Exercice 1. Montrer que $\{0, 1, 2\}$ muni de la loi $+$ définie par

"$i + j$ = reste de la division par 3 de la somme $(i + j)$ calculée dans $\mathbb{Z}$"

est un groupe abélien fini de cardinal 3.

Structures algébriques

Anneaux

Définition (anneau)

On appelle **anneau** *un ensemble A muni de deux lois de composition interne $+$ et $\times$ telles que :*

- $(A, +)$ *est un* **groupe commutatif**,
- *la loi $\times$ est* **associative**,
- *la loi $\times$ est* **distributive** *par rapport à $+$, c'est à dire*

$$\forall\,(x, y, z) \in A^3,\ x \times (y + z) = x \times y + x \times z \text{ et } (x + y) \times z = x \times z + y \times z.$$

Vocabulaire.

- Si la loi $\times$ est commutative, l'anneau A est dit **commutatif**.
- S'il existe un élément neutre, noté 1, pour la loi $\times$, l'anneau A est dit **unitaire**.
- Si quels que soient x et y des éléments de A, on a

$$x \times y = 0 \quad \Rightarrow \quad x = 0 \text{ ou } y = 0,$$

l'anneau A est dit **intègre** ou **anneau d'intégrité**.

Structures algébriques

Anneaux

Exemple. $(\mathbb{Z}, +, \times)$ est un anneau commutatif intègre.

Exercice 2. Reprenons l'exercice 1. Munissons $(\{0,1,2\}, +\}$ d'une loi "multiplicative" commutative définie par :

$$0 * i = 0,\ 1 * i = i,\ 2 * 2 = 1, \qquad i \in \{0,1,2\}$$

Montrer que $(\{0,1,2\}, +, *)$ est un anneau commutatif unitaire et intègre.

Structures algébriques

Corps

Définition (corps)

On appelle corps un ensemble $\mathbf{K}$ muni de deux lois de composition interne $+$ et $\times$ telles que :

- $(\mathbf{K}, +, \times)$ *est un anneau ;*
- *Si on note e l'élément neutre de la loi $+$, alors $(\mathbf{K} \setminus \{e\}, \times)$ est un groupe. Ceci implique que tout élément de $\mathbf{K}$ admet un inverse pour $\times$, à l'exception de e.*

$\triangleright$ Un corps est donc un anneau dans lequel tout élément non-nul est inversible.

Vocabulaire.

- Un corps est dit **fini** s'il admet un nombre fini d'éléments. Un corps fini à q éléments sera noté $\mathbf{F}_q$.
- Le corps $\mathbf{K}$ est dit **commutatif** si la loi $\times$ est commutative.

Structures algébriques

Caractéristique d'un corps

Définition (caractéristique)

Soit $(\mathbf{K}, +, \times)$ *un corps. Soit* r *le plus petit entier* k *tel que*

$$0 = \underbrace{1 \ + \ \ldots \ + \ 1}_{k \text{ fois}}.$$

r *est appelé* **caractéristique** *du corps* $\mathbf{K}$.

Propriété

La caractéristique d'un corps est zéro ou un nombre premier.

L'anneau $\mathbb{Z}/n\mathbb{Z}$

Rappels sur les entiers relatifs

Théorème (division euclidienne)

Soit $(a, b) \in \mathbb{Z} \times \mathbb{Z}$ tel que $b \neq 0$. Il existe un unique couple $(q, r) \in \mathbb{Z} \times \mathbb{Z}$ tel que $a = bq + r$ et $0 \leq r < |b|$, où $|b|$ désigne la valeur absolue de b.
r est le **reste** *et q le* **quotient** *de la division de a par b dans $\mathbb{Z}$.*

Théorème (Bezout)

Soient a et b deux éléments non nuls de $\mathbb{Z}$. Les éléments a et b sont premiers entre eux si et seulement s'il existe u et v de $\mathbb{Z}$ tels que

$$ua + vb = 1.$$

L'anneau $\mathbb{Z}/n\mathbb{Z}$

La congruence

Définition (congruence)

Soit n un entier relatif. Si a et b sont deux entiers relatifs tels que n divise $b-a$, ou $b-a$ est multiple de n, on dit que b est **congru à a modulo n.** *On note :*

$$b \equiv a \ (mod\ n) \text{ ou } b \equiv a\ [n].$$

La relation définie est une congruence modulo n.

Proposition

La relation de congruence est une **relation d'équivalence**, *c'est-à-dire qu'elle est*

symétrique : $b \equiv a\ (mod\ n) \Longrightarrow a \equiv b\ (mod\ n)$
reflexive : $a \equiv a\ (mod\ n)$
transitive : $b \equiv a\ (mod\ n)$ et $c \equiv b\ (mod\ n) \Longrightarrow c \equiv a\ (mod\ n)$

Notation. La classe d'équivalence de a pour cette relation, i.e. l'ensemble des entiers congrus à a modulo n, est la classe de congruence de a modulo n. Elle sera éventuellement notée dans la suite $\bar{a}$.

Remarque. L'ensemble des classes de congruence forme une partition de $\mathbb{Z}$.

L'anneau $\mathbb{Z}/n\mathbb{Z}$

La congruence

Proposition

Les éléments a et b sont congrus modulo n si et seulement s'ils ont même reste dans la division euclidienne par n.

Proposition

$(0, 1, \ldots, n-1)$ constitue un système de représentants de la congruence modulo n, chacun de ces entiers représentant une classe.

Proposition

La relation de congruence est compatible avec les opérations de $\mathbb{Z}$. Si $a \equiv b \ (mod\ n)$ et $a' \equiv b' \ (mod\ n)$, alors

$$\begin{aligned} a + a' &\equiv b + b' \ (mod\ n) \\ aa' &\equiv bb' \ (mod\ n) \end{aligned}$$

L'anneau $\mathbb{Z}/n\mathbb{Z}$

Définition

Définition

$\mathbb{Z}/n\mathbb{Z}$ est l'ensemble des **classes de congruence** *modulo n.*

Notation. La classe d'un entier a sera notée $\overline{a}$. Dans ces conditions, l'ensemble des éléments de $\mathbb{Z}/n\mathbb{Z}$ peut être écrit sous la forme $\overline{0}$, $\overline{1}$, $\ldots$, $\overline{n-1}$.

Définition (opérations)

Soient $\overline{a}$ et $\overline{b}$ deux classes de $\mathbb{Z}/n\mathbb{Z}$. On définit l'addition et la multiplication dans $\mathbb{Z}/n\mathbb{Z}$ ainsi :

$$\begin{aligned} \overline{a}+\overline{b} &= \overline{a+b} \\ \overline{a}\overline{b} &= \overline{ab} \end{aligned}$$

L'anneau $\mathbb{Z}/n\mathbb{Z}$

Propriété

Proposition

Muni des deux opérations définies précédemment, $\mathbb{Z}/n\mathbb{Z}$ est un anneau commutatif unitaire.

Exemple. Les tables d'addition et de multiplication de $\mathbb{Z}/5\mathbb{Z}$ sont données par :

$+$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\times$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Exercice 3. Dans $\mathbb{Z}/13\mathbb{Z}$, par quoi peut-on remplacer $\overline{9}$, $\overline{29}$ et $\overline{-10}$?

L'anneau $\mathbb{Z}/n\mathbb{Z}$

Corps premier

Proposition

Un élément $\overline{a}$ de l'anneau $\mathbb{Z}/n\mathbb{Z}$ est inversible si et seulement si $pgcd(a, n) = 1$, c'est à dire a premier avec n.

Proposition

$\mathbb{Z}/n\mathbb{Z}$ est un corps si et seulement si n est premier.

Notation. Si $p \in \mathbb{N}$ est un nombre premier, le corps $\mathbb{Z}/p\mathbb{Z}$ est noté $\mathbf{F}_p$. On dit qu'il est un **corps premier**.

Polynômes

Rappels

Notation. Soit $\mathbf{K}$ un corps commutatif. On note $\mathbf{K}[X]$ l'anneau des polynômes à coefficients dans $\mathbf{K}$.

Proposition

$\mathbf{K}[X]$ *est un anneau commutatif intègre.*

Remarque. $\mathbf{K}$ peut être identifié à l'ensemble des polynômes constants de $\mathbf{K}[X]$.

Théorème (division euclidienne)

Soient $A \in \mathbf{K}[X]$ et $B \in \mathbf{K}[X]$, $B \neq 0$. Il existe un unique couple (Q, R) de $\mathbf{K}[X] \times \mathbf{K}[X]$ tel que

$$A = BQ + R$$

avec $d^\circ(R) < d^\circ(B)$ ou $R = 0$. Le polynôme Q est appelé le quotient et R le reste de la division euclidienne de A par B.

Remarque. D'un point de vue pratique, on pose la division suivant les puissances décroissantes. La division euclidienne dans $\mathbf{K}[X]$ est alors analogue à celle de $\mathbb{Z}$.

Polynômes

Rappels

Exercice 4. Effectuer dans $\mathbb{R}[X]$ la division euclidienne de $A(X) = 3X^5 + 2X^4 - 2X^3 - X^2 + 1$ et $B(X) = X^2 - 1$.

Exercice 5. Effectuer dans $\mathbf{F}_{31} = \mathbb{Z}/31\mathbb{Z}$ la division euclidienne de $A(X) = 12X^7 + 5X^3 - 21X^2$ et $B(X) = X^4 + X + 1$.

Définition

Soient A et B deux éléments de $\mathbf{K}[X]$ et D un polynôme unitaire $\mathbf{K}[X]$, i.e. dont le coefficient du terme de plus haut degré est égal à l'unité. On dit que D est **le plus grand commun diviseur** *des polynômes A et B si*

- *D divise A et D divise B,*
- *$G \in \mathbf{K}[X]$ divise A et B entraîne G divise D.*

Définition

Un polynôme P non nul de $\mathbf{K}[X]$ est dit **irréductible** *ou* **premier** *sur $\mathbf{K}$ si, dans $\mathbf{K}[X]$, ses seuls diviseurs sont les polynômes constants non nuls et les produits de P par des constantes non nulles.*

Polynômes

Rappels

Définition

Deux éléments non nuls de $\mathbf{K}[X]$ *sont premiers entre eux si et seulement si leur pgcd vaut* 1.

Calcul pratique du pgcd. A l'aide de la division euclidienne, on effectue les divisions successives

$$A = BQ_1 + R_1 \text{ avec } d^\circ(R_1) < d^\circ(B)$$
$$B = R_1Q_2 + R_2 \quad \dots$$

Le pgcd est le dernier reste unitaire non nul.

Exercice 6. Dans $\mathbf{F}_2[X]$, calculer le PGCD des polynômes suivants :
$A(X) = X^{16} + X^{12} + X^{11} + X^8 + X^4 + X + 1$
$B(X) = X^{13} + X^9 + X^8 + X^5 + X + 1.$

Polynômes

L'anneau des polynômes modulo P

Nous allons dans un premier temps définir une sous-structure particulière de la structure d'anneau appelée *idéal*.

Définition (Idéal)

Soit A un anneau et $\mathcal{I}$ une partie non vide de A. On dit que $\mathcal{I}$ est **un idéal** *de A si :*

- x *et* $y \in \mathcal{I} \Rightarrow x - y \in \mathcal{I}$, *c'est à dire* $(\mathcal{I}, +)$ *constitue un sous-groupe de* $(A, +)$.
- $\forall a \in A, \forall x \in \mathcal{I},\ a \times x \in \mathcal{I}$ *et* $x \times a \in \mathcal{I}$ *(idéal bilatère).*

Exemple. Soit A un anneau commutatif et $x \in A$. L'ensemble des multiples de x, noté $< x >$ ou xA, est un idéal.

Polynômes

L'anneau des polynômes modulo P

De même que nous avons défini les $\mathbb{Z}/n\mathbb{Z}$, nous allons définir des *anneaux quotients* de polynômes.

Théorème

Toute **relation d'équivalence** *sur un anneau A, compatible avec les opérations, est du type $(x-y) \in \mathcal{I}$ où $\mathcal{I}$ est un idéal. Ceci signifie que $\mathcal{I}$ est en fait la classe de l'élément nul de A.*

Remarque. Par **compatible avec les opérations**, nous entendons

$$x\mathcal{R}y \text{ et } x'\mathcal{R}y' \Longrightarrow (x+x')\mathcal{R}(y+y') \text{ et } xx'\mathcal{R}yy',$$

où $\mathcal{R}$ désigne la relation d'équivalence.

Polynômes

L'anneau des polynômes modulo P

Les notations utilisées ici sont les mêmes que celles introduites pour $\mathbb{Z}/n\mathbb{Z}$.

Notations.

- Toute relation de ce type s'appelle une congruence modulo $\mathcal{I}$.
- $x - y \in \mathcal{I}$ se note $x \equiv y \ (mod\, \mathcal{I})$ et se lit x **congru à** y **modulo** $\mathcal{I}$.
- L'ensemble des classes d'équivalence de A pour cette relation ou **ensemble quotient** est noté $A/\mathcal{I}$.
- La classe de x est notée $\overline{x}$. On peut remarquer que $\overline{x}$ est donnée par $x + \mathcal{I} = \{z = x + y \ : \ y \in \mathcal{I}\}$.

Polynômes

L'anneau des polynômes modulo P

Théorème

Soit $\mathbf{K}$ un corps commutatif. Soit $\mathcal{I}$ un sous-ensemble de $\mathbf{K}[X]$. Les 2 assertions suivantes sont équivalentes :

- *L'ensemble $\mathcal{I}$ est un idéal de $\mathbf{K}[X]$.*
- *Il existe un polynôme P de $\mathbf{K}[X]$ tel que $\mathcal{I}$ est l'ensemble des multiples de P dans $\mathbf{K}[X]$.*

Définition (équivalence modulo P)

Soit P un élément de $\mathbf{K}[X]$. Deux polynômes A et B de $\mathbf{K}[X]$ sont dit équivalents modulo P si et seulement si $A - B$ est un multiple de P.

Notation. On note $A \equiv B \pmod{P}$.

Polynômes

L'anneau des polynômes modulo P

Proposition

On a $A \equiv B \ (mod\, P)$ si et seulement si les polynômes A et B ont le même reste dans la division par le polynôme P. Dans ces conditions, pour tout C de $\mathbf{K}[X]$, les équivalences suivantes sont satisfaites :

$$\begin{aligned} A + C &\equiv B + C \ (mod\, P) \\ AC &\equiv BC \ (mod\, P). \end{aligned}$$

Corollaire

Soit S un polynôme non nul de $\mathbf{K}[X]$ de degré n, et soit $\mathcal{I} =< S >$ l'idéal engendré par S. Soit $\mathcal{P}_n$ l'ensemble des polynômes non nuls de degré strictement inférieur à n c'est-à-dire l'ensemble des polynômes sur $\mathbf{K}$ de la forme $a_0 + a_1 X + \ldots + a_{n-1} X^{n-1}$. Alors :

- *La classe de A pour l'équivalence modulo S est $A + \mathcal{I}$. C'est aussi $R + \mathcal{I}$ où R est le reste de la division de A par S.*
- *L'application $\phi : \mathbf{K}[X]/\mathcal{I} \longrightarrow \mathcal{P}_n$ définie par $\phi(A + \mathcal{I}) = R$ où R est le reste de la division de A par S est une bijection.*
- *Le cardinal de $\mathbf{K}[X]/\mathcal{I}$ est $|\mathbf{K}|^n$.*

▷ On définit ainsi un anneau commutatif qui sera noté $\mathbf{K}[X]/ < S >$.

Polynômes

L'anneau des polynômes modulo P

En pratique, dans $\mathbf{K}[X]/ < S >$, les polynômes et les résultats des opérations ordinaires sont remplacés par leurs restes dans la division par S.

Exemple. Prenons le cas $\mathbf{K} = \mathbb{Z}/2\mathbb{Z}$ et considérons le polynôme $S = X^2 + X + 1$. Les différents restes possibles de la division par S sont 0, 1, $A = X$ et $B = X + 1$. Les tables des opérations pour $\mathbf{K}[X] / < X^2 + X + 1 >$ sont les suivantes :

$+$	0	1	A	B
0	0	1	A	B
1	1	0	B	A
A	A	B	0	1
B	B	A	1	0

$\times$	0	1	A	B
0	0	0	0	0
1	0	1	A	B
A	0	A	B	1
B	0	B	1	A

Polynômes

L'anneau des polynômes modulo P

Théorème

L'anneau $\mathbf{K}[X] / < S >$ est **un corps** *si et seulement si le polynôme S est irréductible sur $\mathbf{K}$, c'est-à-dire si et seulement si les seuls diviseurs de S à coefficients dans $\mathbf{K}$ sont les constantes non nulles et lui même.*

$\triangleright$ **on parle alors de corps quotient**

Remarque. Si s est le degré de S et q le cardinal de $\mathbf{K}$ (en supposant $\mathbf{K}$ fini), alors $\mathbf{K}[X] / < S >$ contient q^s éléments.

Proposition

Tout corps fini est isomorphe à un corps $\mathbf{K}[X]/ < S >$ où $\mathbf{K} = \mathbb{Z}/p\mathbb{Z}$ avec p premier et S irréductible sur $\mathbf{K}$.

Notation. Un corps fini à q éléments est tel que $q = p^r$, avec p premier. On le note $\mathbf{F}_q$ ou $CG(q)$ (Corps de Galois à q éléments).

Les corps finis

Construction d'un corps fini

Afin de construire un corps fini, on utilise les propriétés suivantes :

Propriété (corps finis)

Soit $\mathbf{F}_q$ un corps fini à $q = p^r$ éléments. On a :

- *Si $\mathbf{F}_q = \mathbf{F}_p[X]/ < S >$, avec p premier et S irréductible sur $\mathbf{F}_p$, alors le polynôme S possède au moins une racine dans $\mathbf{F}_q$.*
- *Le groupe $(\mathbf{F}_q, \times)$ est un groupe cyclique, c'est-à-dire que les éléments non nuls de $\mathbf{F}_q$ sont les puissances d'un même élément générateur. Un tel élément s'appelle élément primitif.*
- *Soit α un élément primitif de $\mathbf{F}_q$, avec $q = p^r$. Alors, tout élément de $\mathbf{F}_q$ s'écrit de manière unique comme une combinaison linéaire de $1, \alpha, \alpha^2, \ldots, \alpha^{r-1}$. Autrement dit, si l'on considère $\mathbf{F}_q$ comme un $\mathbf{F}_p$ espace vectoriel, alors $\{1, \alpha, \alpha^2, \ldots, \alpha^{r-1}\}$ en est une base.*

Les corps finis

Construction d'un corps fini

Soit S un polynôme irréductible sur $\mathbf{F}_p$, unitaire et de degré r :

$$S = x^r + a_{r-1}x^{r-1} + \cdots + a_1 x + a_0.$$

Supposons que S possède comme racine dans $\mathbf{F}_q$ un élément primitif α de $\mathbf{F}_q$.
Puisque $f(\alpha) = 0$, il vient :

$$\alpha^r = -a_{r-1}\alpha^{r-1} - \cdots - a_1\alpha - a_0.$$

En multipliant l'égalité précédente par α, puis en remplaçant α^r par son expression, on obtient α^{n+1} sous la forme :

$$\alpha^{r+1} = -b_{r-1}\alpha^{r-1} - \cdots - b_1\alpha - b_0,$$

puis $\alpha^{r+2}, \ldots, \alpha^{p^r-2}$ comme combinaisons linéaires des éléments de la base $1, \alpha, \ldots, \alpha^{r-1}$. On obtient ainsi une expression de tous les éléments de $\mathbf{F}_q$ permettant d'effectuer les calculs.

Outline

Les codes linéaires

Chapitre 5

Principe

Définition d'un code linéaire

Soient p un nombre premier et s est un entier positif. Il existe un unique corps de taille $q = p^s$, noté $\mathbf{F}_q$. L'ensemble $(\mathbf{F}_q)^n$ de tous les n-uples formés d'éléments de $\mathbf{F}_q$ est un espace vectoriel sur $\mathbf{F}_q$.

Définition

$\mathcal{L}$ est un **code linéaire** *si $\mathcal{L}$ est un sous-espace vectoriel de $(\mathbf{F}_q)^n$. On dit que $\mathcal{L}$ est un $[n, k]$-code si $\dim(\mathcal{L}) = k$. Si la distance minimale de $\mathcal{L}$ est d, on parle de $[n, k, d]$-code.*

$\triangleright$ Attention aux notations "$(.)$-code" et "$[.]$-code" !

Principe

Poids d'un code linéaire

Définition

Le **poids** $\omega(x)$ *du mot* x *de* $(\mathbf{F}_q)^n$ *est le nombre de composantes non nulles de* x.
Le **poids minimal** $\omega(\mathcal{L})$ *du code* $\mathcal{L}$ *est le minimum des poids de tous les vecteurs non nuls de* $\mathcal{L}$.

Exemple

$\omega(1101) = 3$

$\mathcal{L} = \{00000, 10111, 11010, 01101\} \quad \longrightarrow \quad \omega(\mathcal{L}) = 3$

Principe

Poids d'un code linéaire

Définition

Le **poids** $\omega(x)$ *du mot* x *de* $(\mathbf{F}_q)^n$ *est le nombre de composantes non nulles de* x.
Le **poids minimal** $\omega(\mathcal{L})$ *du code* $\mathcal{L}$ *est le minimum des poids de tous les vecteurs non nuls de* $\mathcal{L}$.

Exemple

$\omega(1101) = 3$
$\mathcal{L} = \{00000, 10111, 11010, 01101\} \quad \longrightarrow \quad \omega(\mathcal{L}) = 3$

Définition

Si x *et* y *sont deux mots binaires, on appelle intersection de* x *et* y, *notée* $x \cap y$, *l'élément défini par* $(x \cap y)(i) = 1$ *si* $x(i) = y(i) = 1$, *et* 0 *sinon.*

En considérant les définitions ci-dessus, on peut démontrer les relations suivantes :

$$\forall x, y \in (\mathbf{F}_q)^n, d_{Ham}(x, y) = \omega(x - y),$$
$$\forall x, y \in (\mathbf{F}_2)^n, d_{Ham}(x, y) = \omega(x) + \omega(y) - 2\omega(x \cap y).$$

Théorème

Soit $\mathcal{L}$ *un code linéaire. On a :* $d(\mathcal{L}) = \omega(\mathcal{L})$.

Principe

Codes détecteurs et correcteurs linéaires

En utilisant ce qui a été vu au cours précédent, on a immédiatement que :

Théorème

En utilisant la règle de décodage par distance minimale, un code linéaire peut détecter jusqu'à t erreurs, avec $t = \omega(\mathcal{L}) - 1$.
De plus, il en corrige jusqu'à t', avec $\omega(\mathcal{L}) = 2t' + 1$ ou $2t' + 2$.

Exemple
$\mathcal{L} = \{00000, 10111, 11010, 01101\}$ est 2-détecteur et 1-correcteur.

Matrice génératrice

Définition

Définition

Soit $\mathcal{L}$ un $[n, k]$-code. Une matrice $\mathbf{G}$ de dimension $(k \times n)$ dont les lignes forment une base de $\mathcal{L}$ est une **matrice génératrice** *de $\mathcal{L}$. On a alors :*

$$\mathcal{L} = \{x\mathbf{G} \mid x \in (\mathbf{F}_q)^k\}.$$

Exemple

La matrice génératrice suivante permet d'encoder les mots de $(\mathbf{F}_2)^3$.

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

Matrice génératrice

Forme systématique

Définition (forme systématique)

Un $[n, k]$-code $\mathcal{L}$ est dit sous **forme systématique** *s'il existe k positions $i_1, \ldots, i_k$ telles que, par restriction des mots du code à ces k positions, on obtient les q^k mots q-aires possibles de longueur k.*

Exemple. $\mathcal{C} = \{0000, 0110, 1001, 1010\}$ est systématique sur les positions 1 et 3 :

$$\begin{aligned} 00 &\longrightarrow \underline{0}0\underline{0}0 \\ 01 &\longrightarrow \underline{0}1\underline{1}0 \\ 10 &\longrightarrow \underline{1}0\underline{0}1 \\ 11 &\longrightarrow \underline{1}0\underline{1}0 \end{aligned}$$

Matrice génératrice

Forme systématique

Définition (forme systématique)

Un $[n, k]$-code $\mathcal{L}$ est dit sous **forme systématique** *s'il existe k positions $i_1, \ldots, i_k$ telles que, par restriction des mots du code à ces k positions, on obtient les q^k mots q-aires possibles de longueur k.*

Exemple. $\mathcal{C} = \{0000, 0110, 1001, 1010\}$ est systématique sur les positions 1 et 3 :

$$\begin{array}{ccc} 00 & \longrightarrow & \underline{0}0\underline{0}0 \\ 01 & \longrightarrow & \underline{0}1\underline{1}0 \\ 10 & \longrightarrow & \underline{1}0\underline{0}1 \\ 11 & \longrightarrow & \underline{1}0\underline{1}0 \end{array}$$

Théorème (forme standard)

Tout $[n, k]$-code linéaire a une matrice génératrice de la forme $\mathbf{G} = (\mathbf{I}_k \mid \mathbf{A})$, dite **standard**, *où $\mathbf{I}_k$ désigne la matrice unité de dimension k.*

Exemple. La matrice génératrice suivante est sous forme standard.

$$\mathbf{G} = \left(\begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right).$$

Dual d'un code linéaire

Définition

Définition

Soit $\mathcal{L}$ un $[n,k]$-code.

$$\mathcal{L}^{\perp} = \{x \in (\mathbf{F}_q)^n \mid x.c = 0, \forall c \in \mathcal{L}\}$$

est appelé **code dual** *de $\mathcal{L}$.*

Théorème

Soit $\mathcal{L}$ un $[n,k]$-code linéaire et $\mathcal{L}^{\perp}$ son dual. On a :

1. $\mathcal{L}^{\perp} = \{x \in (\mathbf{F}_q)^n \mid x\mathbf{G}^{\top} = 0\}$ *où* $\mathbf{G}$ *est génératrice de* $\mathcal{L}$ *;*
2. $\mathcal{L}^{\perp}$ *est un* $[n, n-k]$*-code linéaire ;*
3. $\mathcal{L}^{\perp\perp} = \mathcal{L}$.

Dual d'un code linéaire

Matrice de test

Soit $\mathcal{L}$ un code linéaire de matrice génératrice $\mathbf{G} = (\mathbf{I}_k \mid \mathbf{A})$ de dimension $(k \times n)$. On pose $\mathbf{H} = (-\mathbf{A}^\top \mid \mathbf{I}_{n-k})$.

Définition

La matrice $\mathbf{H}$ *est dite* **matrice de contrôle** *ou* **matrice de test** *du code linéaire* $\mathcal{L}$.

On montre aisément que $\mathbf{H}$ est une matrice génératrice de $\mathcal{L}^\perp$. Cette matrice sera utilisée par la suite pour le décodage.

Théorème

Soit $\mathcal{L}$ *un code linéaire ayant* $\mathbf{H}$ *comme matrice de contrôle. Il existe un mot de poids* ω *si, et seulement si, il existe* ω *colonnes de* $\mathbf{H}$ *linéairement dépendantes.*

Exemple

Matrice génératrice et mots du code

On construit un $[6,3]$-code linéaire binaire en choisissant trois vecteurs linéairement indépendants de $(\mathbf{F}_2)^6$.

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \end{pmatrix}.$$

On obtient l'ensemble des mots du code $\mathcal{L}$ en calculant tous les produits $x\mathbf{G}$ avec $x \in (\mathbf{F}_2)^3$.

Exemple

Matrice génératrice et mots du code

On construit un $[6,3]$-code linéaire binaire en choisissant trois vecteurs linéairement indépendants de $(\mathbf{F}_2)^6$.

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \end{pmatrix}.$$

On obtient l'ensemble des mots du code $\mathcal{L}$ en calculant tous les produits $x\mathbf{G}$ avec $x \in (\mathbf{F}_2)^3$.

Les mots du code $\mathcal{L}$ ainsi que leur poids sont donnés par :

x	$\mathcal{L}$	ω
000	000000	0
001	110110	4
010	011101	4
011	101011	4
100	100101	3
101	010011	3
110	111000	3
111	001110	3

$\longrightarrow \mathcal{L}$ est 2-détecteur et 1-correcteur.

Exemple

Matrice de contrôle et mots du code dual

On écrit la matrice $\mathbf{G}$ sous forme standard (pivot de Gauss) :

$$\mathbf{G} \to \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} \cdots \to \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{array} \right) = (\mathbf{I}_3 \mid \mathbf{A}).$$

La matrice de contrôle $\mathbf{H} = (-\mathbf{A}^\top \mid \mathbf{I}_3)$ de $\mathcal{L}$ est donnée par :

$$\mathbf{H} = \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{array} \right).$$

Exemple

Matrice de contrôle et mots du code dual

On écrit la matrice $\mathbf{G}$ sous forme standard (pivot de Gauss) :

$$\mathbf{G} \to \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} \cdots \to \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{array} \right) = (\mathbf{I}_3 \mid \mathbf{A}).$$

La matrice de contrôle $\mathbf{H} = (-\mathbf{A}^\top \mid \mathbf{I}_3)$ de $\mathcal{L}$ est donnée par :

$$\mathbf{H} = \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{array} \right).$$

A partir de $\mathbf{H}$, on obtient les mots du code dual $\mathcal{L}^\perp$:

x	$\mathcal{L}^\perp$	ω
000	000000	0
001	110001	3
010	011010	3
011	101011	4
100	101100	3
101	011101	4
110	110110	4
111	000111	3

Décodage par tableau standard

Notion de syndrome

Nous allons à présent exposer une règle de décodage à distance minimale reposant sur un tableau, dit *tableau standard*.

Définition

Soit $\mathcal{L}$ un $[n,k]$-code linéaire de matrice de contrôle $\mathbf{H}$. Soit x un élément de $(\mathbf{F}_q)^n$. le mot $x\mathbf{H}^\top$ est appelé syndrome de x.

Cette définition permet d'introduire l'application linéaire s suivante :

$$\begin{array}{rccc} \mathrm{s}: & (\mathbf{F}_q)^n & \longrightarrow & (\mathbf{F}_q)^{n-k} \\ & x & \longrightarrow & \mathrm{s}(x) = x\mathbf{H}^\top \end{array}$$

Soit x un élément de $\mathcal{L}$. On note que $\mathrm{s}(x) = 0$, ce qui signifie que

$$\mathcal{L} = \ker(s).$$

Décodage par tableau standard

Classes d'équivalence

On définit la classe de x, notée C_x ou $x + \mathcal{L}$, par :

$$\mathrm{C}_x = \{x + I \mid I \in \mathcal{L}\}.$$

Théorème

L'ensemble des classes C_x, $x \in (\mathbf{F}_q)^n$, forme une partition de $(\mathbf{F}_q)^n$.

Théorème

Soit $\mathcal{L}$ un $[n, k]$-code. Les éléments x et y de $(\mathbf{F}_q)^n$ ont le même syndrome si et seulement si ils définissent la même classe.

Décodage par tableau standard

Classes d'équivalence

On définit la classe de x, notée C_x ou $x + \mathcal{L}$, par :

$$\mathrm{C}_x = \{x + I \mid I \in \mathcal{L}\}.$$

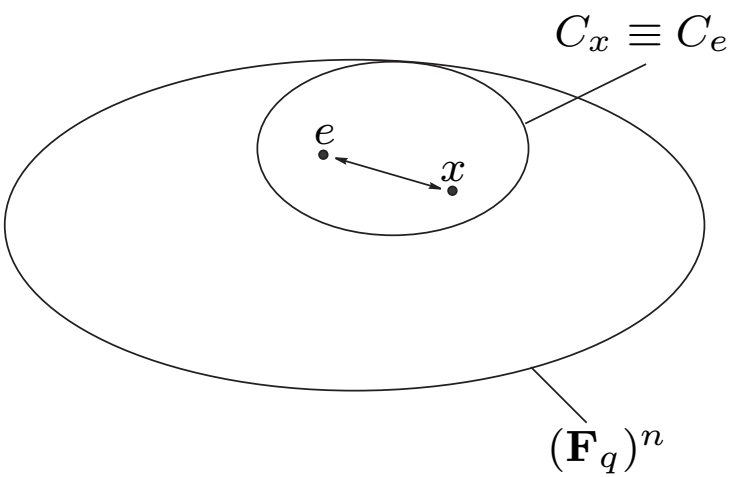
Théorème

L'ensemble des classes C_x, $x \in (\mathbf{F}_q)^n$, forme une partition de $(\mathbf{F}_q)^n$.

Théorème

Soit $\mathcal{L}$ un $[n, k]$-code. Les éléments x et y de $(\mathbf{F}_q)^n$ ont le même syndrome si et seulement si ils définissent la même classe.

Soit x le mot reçu. Le décodage par distance minimale requiert que l'on décode x par le mot de code c pour lequel $e = x - c$ a le poids le plus faible. Étant donné que c varie dans $\mathcal{L}$, e est un élément de C_x. Le vecteur d'erreur e a donc le même syndrome que x.



Décodage par tableau standard

Construction du tableau

Mode de construction

1. La première ligne comporte les mots de $\mathcal{L}$.
2. La ligne j est constituée des $e_j + \mathcal{L}$, où e_j est un mot sélectionné de plus petit poids ne se trouvant pas dans les $j-1$ lignes précédentes.
3. Le processus est itéré jusqu'à ce que tous les mots de $(\mathbf{F}_q)^n$ soient représentés dans le tableau.

0	c_1	c_2	$\dots$	c_m		
e_1	$c_1 + e_1$	$c_2 + e_1$	$\dots$	$c_m + e_1$	$\rightarrow$	$e_1 + \mathcal{L}$
e_2	$c_1 + e_2$	$c_2 + e_2$	$\dots$	$c_m + e_2$	$\rightarrow$	$e_2 + \mathcal{L}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	
e_s	$c_1 + e_s$	$c_2 + e_s$	$\dots$	$c_m + e_s$	$\rightarrow$	$e_s + \mathcal{L}$

Décodage par tableau standard

Exemple

Soit $\mathcal{L}$ un $[4, 2]$-code linéaire binaire défini par la matrice génératrice

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

On a donc :

x	$\mathcal{L}$	ω
00	0000	0
01	0100	1
10	1101	3
11	1001	2

Décodage par tableau standard

Exemple

Soit $\mathcal{L}$ un $[4, 2]$-code linéaire binaire défini par la matrice génératrice

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

On a donc :

x	$\mathcal{L}$	ω
00	0000	0
01	0100	1
10	1101	3
11	1001	2

Ceci nous conduit au tableau standard suivant :

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

Inconvénient : espace mémoire occupé (16 Go pour un [32,6]-code !).

$\Longrightarrow$ **décodage par syndrome**

Décodage par syndrome

Principe

Les mots d'une même ligne du tableau standard ont même syndrome.

$\Rightarrow$ **construction d'une table de représentants et de syndromes**

0	c_1	c_2	$\dots$	c_m	$\mathrm{s}(e_i)$
e_1	$c_1 + e_1$	$c_2 + e_1$	$\dots$	$c_m + e_1$	$e_1\mathbf{H}^\top$
e_2	$c_1 + e_2$	$c_2 + e_2$	$\dots$	$c_m + e_2$	$e_2\mathbf{H}^\top$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
e_s	$c_1 + e_s$	$c_2 + e_s$	$\dots$	$c_m + e_s$	$e_s\mathbf{H}^\top$

Le calcul du syndrome d'un mot reçu x désigne son représentant e_i. Le décodage s'effectue en calculant $x - e_i$.

Décodage par syndrome

Exemple

On considère le code linéaire $\mathcal{L}$ défini par la matrice génératrice $\mathbf{G}$:

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \rightsquigarrow \mathbf{G}' = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \rightsquigarrow \mathbf{H} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Représentant e_i	Syndrome $e_i\mathbf{H}^\top$
0000	00
1000	01
0010	10
1010	11

Décodage par syndrome

Exemple

On considère le code linéaire $\mathcal{L}$ défini par la matrice génératrice $\mathbf{G}$:

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \rightsquigarrow \mathbf{G}' = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \rightsquigarrow \mathbf{H} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Représentant e_i	Syndrome $e_i\mathbf{H}^\top$
0000	00
1000	01
0010	10
1010	11

On suppose que $x = 1110$ a été reçu. Le calcul de son syndrome donne :

$$x\mathbf{H}^\top = \begin{pmatrix} 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}^\top = \begin{pmatrix} 1 & 1 \end{pmatrix}.$$

Le représentant de la classe est donc 1010 et l'on obtient

$$c = 1110 - 1010 = 0100.$$

Les codes de Hamming

Principe

La distance minimale d'un code linéaire $\mathcal{L}$ est le plus petit nombre de colonnes linéairement dépendantes dans sa matrice de test $\mathbf{H}$. Pour un $[n, k, 3]$-code, aucune colonne de $\mathbf{H}$ n'est multiple d'une autre.

Construction

Les codes de Hamming sont des $[n, k, 3]$-codes construits ainsi :

1. choix d'un vecteur-colonne c_1 non-nul dans $(\mathbf{F}_q)^r$
2. choix d'un vecteur-colonne c_2 dans $(\mathbf{F}_q)^r - \{\alpha.c_1 : \alpha \in \mathbf{F}_q^\star\}$
3. réitération jusqu'à ce qu'il n'y ait plus de c_i non-nul

Les codes de Hamming

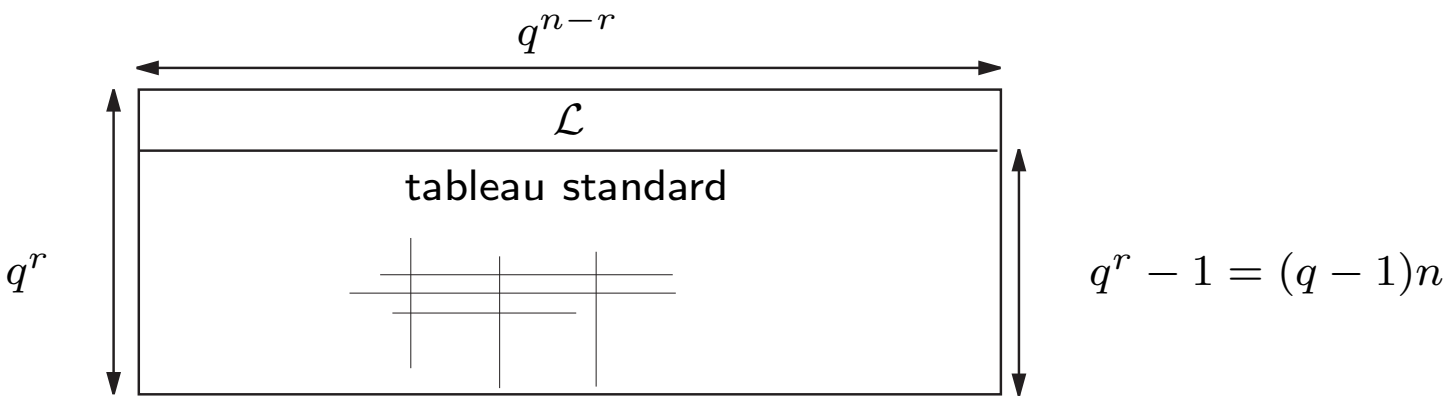
Intérêts

Notations

Un $[n, k, d]$-code de Hamming q-aire d'ordre r, noté $\mathcal{H}_q(r)$, est tel que :

$$n = (q^r - 1)/(q - 1) \quad ; \quad k = n - r \quad ; \quad d = 3$$

Décodage des codes de Hamming



On constate que $(q-1)n$ représente aussi le nombre d'erreurs possibles de poids 1 !

▷ le syndrome de e_i est donc égal à la $i^{\text{ème}}$ colonne de $\mathbf{H}$.

Décodage des codes de Hamming $\mathcal{H}_2(r)$

Exemple

Les colonnes de la matrice de contrôle $\mathbf{H}$ sont simplement les représentations binaires des $2^r - 1$ premiers nombres positifs non-nuls.

$$\triangleright \text{ syndrome} = \text{position de l'erreur à corriger}$$

Exemple : $\mathcal{H}_2(3)$

$$\mathbf{H} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

En supposant qu'une unique erreur s'est produite à la position 3, ce qui correspond au vecteur d'erreur donné par $e_3 = 0010000$, le syndrome du mot reçu est égal à $e_3\mathbf{H}^\top = 011$. Ce nombre donne la position de l'erreur.

Décodage des codes de Hamming $\mathcal{H}_3(r)$

Exemple

Comme pour $\mathcal{H}_2(r)$, on choisit les colonnes de $\mathbf{H}$ comme l'expression des premiers nombres dans une base ternaire, en s'assurant que la première composante non-nulle de ces nombres est 1.

Exemple : $\mathcal{H}_3(3)$

$$\mathbf{H} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ 1 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 \end{pmatrix}$$

Si une erreur apparaît à la position i, le vecteur d'erreur est de la forme $\alpha\, e_i$, avec $\alpha \in \{1, 2\}$. Le syndrome résultant $\alpha\, e_i \mathbf{H}^\top$.

$\triangleright$ on détermine la position de l'erreur et la correction à apporter.

Principe des codes cycliques

Définition

Les codes cycliques $\mathcal{C}$ constituent l'une des classes les plus importantes parmi les codes linéaires.

Définition

Un code $\mathcal{C}$ est dit cyclique s'il est linéaire et s'il vérifie la propriété suivante :

$$(c_0 \dots c_{n-1}) \in \mathcal{C} \iff (c_{n-1} c_0 \dots c_{n-2}) \in \mathcal{C}.$$

La permutation circulaire des composantes est appelée *shift*. On peut dire que $(c_{n-1} c_0 \dots c_{n-2})$ est le shift de $(c_0 \dots c_{n-1})$.

Exemples :
Les codes suivants sont des exemples de codes cycliques, qui ne présentent pas tous un intérêt pratique :

- $\{0\}$ et $(\mathbf{F}_q)^n$
- $\mathcal{C} = \{000, 101, 011, 110\}$
- Soit $\mathcal{C}$ le code dont la matrice génératrice est définie par

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{matrix} \longrightarrow \mathbf{G}^{(1)} \\ \longrightarrow \mathbf{G}^{(2)} \\ \longrightarrow \mathbf{G}^{(3)} \end{matrix}$$

Représentation polynômiale

Intérêt

Il est commode d'utiliser la représentation polynômiale suivante :

$$(c_0 c_1 \dots c_{n-1}) \longleftrightarrow m(x) = c_0 + c_1 x + \dots + c_{n-1} x^{n-1}.$$

En effet, le polynôme associé au mot shifté $(c_{n-1} c_0 \dots c_{n-2})$ est celui que l'on obtient en évaluant $x.m(x)$ modulo $(x^n - 1)$:

$$\begin{aligned} c_{n-1} + c_0 x + \dots + c_{n-2} x^{n-1} &= x(c_0 + \dots + c_{n-1} x^{n-1}) - c_{n-1}(x^n - 1) \\ &\equiv x.m(x) \text{ modulo } (x^n - 1). \end{aligned}$$

Représentation polynômiale

Cadre algébrique

Définition

Soit $\mathbf{F}_q$ un corps fini et soit n un entier non-nul. On appelle représentation polynômiale de $(\mathbf{F}_q)^n$ l'application

$$\theta : (\mathbf{F}_q)^n \longrightarrow \mathbf{F}_q[x]/ < x^n - 1 >$$

telle que $\theta(c_0 c_1 \dots c_{n-1}) = c_0 + c_1 x + \dots + c_{n-1} x^{n-1}$.

Définition

On appelle représentation polynômiale de $\mathcal{C}$ l'ensemble des représentations polynômiales des mots de $\mathcal{C}$, que l'on note $\theta(\mathcal{C})$.

Exemple

Si $\mathcal{C} = \{000, 101, 011, 110\}$, alors $\theta(\mathcal{C}) = \{0, 1 + x^2, x + x^2, 1 + x\}$ où 0 désigne ici le polynôme nul.

Structure algébrique de $\theta(\mathcal{C})$

Codes cycliques et idéaux

La définition d'un code cyclique nous amène directement à :

Théorème

Le code $\mathcal{C}$ est cyclique si et seulement si $\mathcal{C}$ est un sous-espace vectoriel de $(\mathbf{F}_q)^n$ et si tout multiple modulo $(x^n - 1)$ d'un polynôme de $\theta(\mathcal{C})$ est aussi un polynôme de $\theta(\mathcal{C})$.

En se rappelant de la définition d'un *idéal bilatère*, on obtient :

Théorème

Soit $\mathcal{C}$ un code linéaire de longueur n sur $\mathbf{F}_q$. Alors $\mathcal{C}$ est un code cyclique si et seulement si sa représentation polynômiale est un idéal bilatère de l'anneau $\mathbf{F}_q[x]/ < x^n - 1 >$.

Structure algébrique de $\theta(\mathcal{C})$

Polynôme générateur

Après avoir montré que tout idéal de l'anneau $\mathbf{F}_q[x]/ < x^n - 1 >$ est engendré par un même polynôme, dit *polynôme générateur*, on montre :

Théorème

Chaque code cyclique $\mathcal{C}$ de longueur n sur $\mathbf{F}_q$, et non réduit à l'élément nul, possède un polynôme générateur unitaire et un seul qui est diviseur de $(x^n - 1)$ dans $\mathbf{F}_q[x]$.

Exemple Soit $\mathcal{C}$ le code cyclique tel que :

$$\theta(\mathcal{C}) = \{0, 1 + x, x + x^2, 1 + x^2\}.$$

On constate que le polynôme $(1 + x)$ est le polynôme générateur de $\mathcal{C}$.

Il est maintenant possible d'exhiber tous les codes cycliques de longueur n grâce à la recherche de tous les diviseurs de $(x^n - 1)$:

Théorème

Soit $\mathcal{C}$ un code cyclique de longueur n et $g(x)$ son polynôme générateur tel que $\mathrm{d}^\circ(g) = t$. La famille suivante

$$\{g(x), x.g(x), \ldots, x^{n-t-1}.g(x)\}$$

est une base de $\theta(\mathcal{C})$ et la dimension du code est $n - t$.

Construction d'un code cyclique

Exemple

On veut construire un code cyclique de longueur 7 sur $\mathbf{F}_2$. On montre que $(x^7-1)=(x-1)(x^3+x+1)(x^3+x^2+1)$, ce qui nous conduit à :

$$
\begin{aligned}
\mathcal{C}_0 : \quad & g_0(x) && = x^7-1 \equiv 0 \\
\mathcal{C}_1 : \quad & g_1(x) && = x-1 \\
\mathcal{C}_2 : \quad & g_2(x) && = x^3+x+1 \\
\mathcal{C}_3 : \quad & g_3(x) && = x^3+x^2+1 \\
\mathcal{C}_4 : \quad & g_4(x) && = g_1(x).g_2(x) = x^4+x^3+x^2+1 \\
\mathcal{C}_5 : \quad & g_5(x) && = g_1(x).g_3(x) = x^4+x^2+x+1 \\
\mathcal{C}_6 : \quad & g_6(x) && = g_2(x).g_3(x) = x^6+x^5+x^4+x^3+x^2+x+1
\end{aligned}
$$

Matrice génératrice

Construction à partir du polynôme générateur

La famille $\{g(x), x.g(x), \ldots, x^{n-t-1}.g(x)\}$ est une base de $\theta(\mathcal{C})$. Il suffit donc de choisir les mots associés à cette base pour construire $\mathbf{G}$.

Théorème

Soit $g(x) = g_0 + g_1 x + \ldots + g_t x^t$ le polynôme générateur d'un code cyclique $\mathcal{C}$ de longueur n sur $\mathbf{F}_q$. La matrice $\mathbf{G}$ constituée de $n-t$ lignes et n colonnes suivante est génératrice.

$$\mathbf{G} = \begin{pmatrix} g_0 & g_1 & \ldots & g_t & 0 & \ldots & 0 \\ 0 & g_0 & g_1 & 0 & g_t & \ldots & 0 \\ & & \ldots & & \ldots & & \\ 0 & \ldots & 0 & g_0 & g_1 & \ldots & g_t \end{pmatrix}.$$

Exemple :
Considérons le code cyclique $\mathcal{C}$ de $(\mathbf{F}_2)^7$ engendré par le polynôme générateur $g(x) = 1 + x^2 + x^3$. D'après le théorème précédent, une matrice génératrice de ce code est donnée par :

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

Chaque ligne de $\mathbf{G}$ peut être obtenue par un shift de la précédente.

Matrice de contrôle

Définition du polynôme de contrôle

On définit un polynôme de contrôle ainsi :

Définition

Soit $\mathcal{C}$ un $[n,k]$-code cyclique de polynôme générateur $g(x)$. Le polynôme $h(x)$ vérifiant $g(x).h(x) = (x^n - 1)$ est dit polynôme de contrôle.

On peut montrer le résultat suivant :

Théorème

Soit $\mathcal{C}$ un $[n,k]$-code cyclique dont le polynôme de contrôle est $h(x)$. On a la relation suivante :

$$p(x) \in \theta(\mathcal{C}) \Leftrightarrow p(x).h(x) = 0.$$

Matrice de contrôle

Définition du polynôme de contrôle

On définit un polynôme de contrôle ainsi :

Définition

Soit $\mathcal{C}$ un $[n,k]$-code cyclique de polynôme générateur $g(x)$. Le polynôme $h(x)$ vérifiant $g(x).h(x) = (x^n - 1)$ est dit polynôme de contrôle.

On peut montrer le résultat suivant :

Théorème

Soit $\mathcal{C}$ un $[n,k]$-code cyclique dont le polynôme de contrôle est $h(x)$. On a la relation suivante :

$$p(x) \in \theta(\mathcal{C}) \Leftrightarrow p(x).h(x) = 0.$$

Déterminons l'expression de la matrice de contrôle à partir du polynôme de contrôle.

Théorème

Soit $\mathcal{C}$ un $[n,k,d]$-code cyclique de polynôme de contrôle $h(x) = h_0 + h_1 x + \ldots + h_k x^k$. La matrice $\mathbf{H}$ suivante est une matrice de test de $\mathcal{C}$:

$$\mathbf{H} = \begin{pmatrix} h_k & h_{k-1} & \ldots & h_0 & 0 & \ldots & 0 \\ 0 & h_k & h_{k-1} & \ldots & h_0 & \ldots & 0 \\ & & \ldots & & \ldots & & \\ 0 & \ldots & 0 & h_k & h_{k-1} & \ldots & h_0 \end{pmatrix}.$$