Théorie de l’information

Marine Minier
INSA de Lyon / IF / CITI

Bibliographie

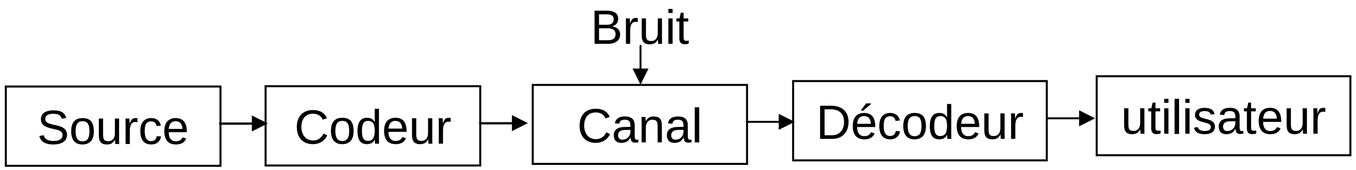
- Cours de N. Sendrier : https://www.rocq.inria.fr/secret/Nicolas.Sendrier/thinfo.pdf

- Elements of information theory (Wiley) de Thomas Cover et Joy A. Thomas

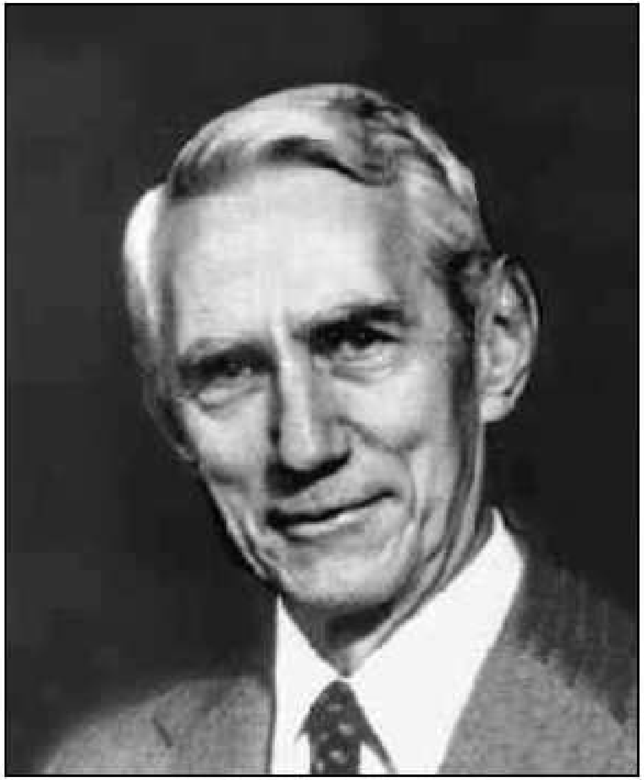
- Information theory, Inference, and Learning Algorithms (University Press, Cambridge) de David J.C. MacKay

Introduction (1/3)

- Théorie des communications : moyen de transmettre une information depuis une source jusqu'à un utilisateur
  - **Source** = voix, signal électromagnétique, séquences symboles binaires,…
  - **Canal** = ligne téléphonique, liaison radio, disque compact,…
  - **Bruit** = pertubateur du canal :
    - Perturbations électriques, rayures,…
  - **Codeur** = ens des opérations effectuées sur la sortie de la source avant transmission
    - modulation, compression,..
    - But = combattre le bruit
  - **Décodeur** = restituer l'information de la source



Introduction (2/3)

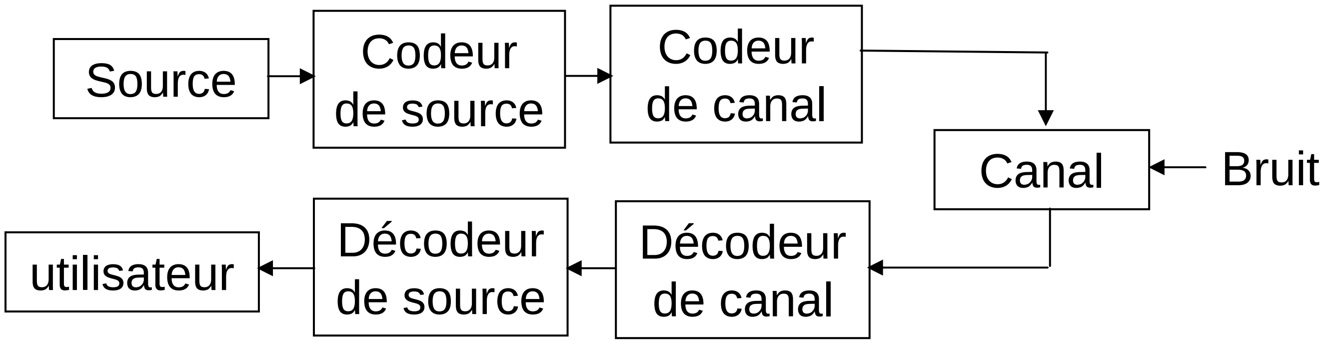


Claude Elwood SHANNON
30 Avril 1916 / 24 Février 2001

- Théorie de l'information
  - Shannon, 1948.
  - Modèles mathématiques (proba, codes,…) des systèmes de communications

Introduction (3/3)

- Simplification : séparation modèles de sources et modèles de canaux :



- Plan :
  - Mesure quantitative de l'information
  - Sources et codage de sources
  - Canal et codage de canal

Introduction : Mesure quantitative de l'information

- Comment mesurer la « Quantité d'information » d 'un message ?

- Message = événements aléatoires produits par la source
  - Exemple d'événements = émission d 'une suite de symboles discrets choisis dans un ensemble fini de symboles (alphabet)
  - Réalisation particulière parmi l'ensemble des données transmissibles

- La Quantité d 'information du message est proportionnelle à son degré d'incertitude (ou d'improbabilité)
  - Un événement certain ou connu à l 'avance du destinataire n'est pas très informatif ...

Introduction : Sources et codage de sources

- Entropie d 'une Source
  - Quantité d 'information moyenne qu'elle produit.

- Information mutuelle moyenne (ou information mutuelle)
  - Quantité d'information moyenne que la connaissance d 'un message reçu apporte sur le message émis.
    - Symétrique (Source -> Destination ou Destination -> Source)
    - Toujours inférieure à l 'entropie de la Source
    - Faibles perturbations -> Info. mutuelle proche de l'entropie de la Source
    - Fortes perturbations -> Forte diminution de l 'information mutuelle

Introduction : Canal et codage de canal

- Capacité du Canal
  - Maximum de l 'information mutuelle moyenne par rapport à toutes les sources possibles. Maximum de l 'information sur la Source que le canal peut transmettre au Destinataire

- Messages et procédés de codage :
  - Codage de Source: Concision maximale et suppression de redondance.

  - Codage de Canal : Amélioration de la résistance aux perturbations.

- Antagonisme entre les deux codages précédents.

Mesure quantitative de l’information

Contexte (1/2)

- Envoi de messages par la source

- Message = suite finie de symboles appartenant à un ensemble fini, prédéterminé : l'alphabet
  - Alphabet = ens. fini de symboles
    - Lettres : a b c d e f…
    - Alphabet binaire : 0 1

- Exemple de messages :
  - « rendez-vous le 13 juin »
  - 0110100101010110001010001110100101101

- Source de messages = Ensemble de TOUS les messages susceptibles d'être formés à partir d'un alphabet

Contexte (2/2)

- Dans la suite : Sources discrètes et finies.

- Pour le destinataire, la source et le canal ont un comportement aléatoire, décrit en termes probabilistes.

- La communication n'a d 'intérêt que si le contenu du message est inconnu à priori.
  - « Plus un message est imprévu, improbable, plus il est informatif »

- Qualitativement, fournir une information = lever une partie de l'incertitude sur l'issue d'une expérience aléatoire

Description quantitative de l’information (1/2)

- L'information propre de *x : I(x)* doit être une fonction de sa probabilité :

$$I(x) = f(1/p(x))$$

- Pour définir $f()$ nous admettrons :
  - L'information propre de x est une **fonction décroissante** de $p(x)$: en effet un évènement certain n'apporte aucune information, alors qu'un évènement improbable en apportera beaucoup :

    $$f(1)=0 \text{ si } p(x)=1$$

  - **L'information propre est une grandeur additive** : si les évènements x et y sont statistiquement indépendants alors l'information totale qu'ils peuvent fournir est la somme des informations propres :

$$I(x,y)=f(1/p(x, y)) = f(1/p(x).1/p(y)) = f(1/p(x)) + f(1/p(y))=I(x)+I(y)$$

Description quantitative de l'information (2/2)

- On est donc tout naturellement conduit à choisir

$$f = \log$$

  - si $\log_2$ unité : bit ou Shannon (Sh)
  - si $\log_e$ unité : nat
  - si $\log_{10}$ unité : dit ou hartley

- Dans ce cours surtout $\log_2$

Exemple

- Soit une source dont l'alphabet de sortie $\{a_0,\ldots,a_{15}\}$ avec $P(a_k)=1/16$.
- L'information propre de l'une de ces sorties a_k est égale à : $I(a_k)=\log_2(16)=4$ bits
- Information : choisir k dans {0,…,15}
  => besoin de 4 bits
- Attention : ce résultat = vrai car équiprobabilité !

Définitions : Mesures quantitatives de l 'information par événements

- Information propre : $I(x) = log(1/p(x)) = -log(p(x))$
- Information conjointe : $I(x,y) = log\ (1/p(x, y)) = -\ log\ (p(x,y))$
- Information conditionnelle : $I(x \mid y) = log\ (1/p(x \mid y)) = -log(p(x \mid y))$
- La règle de Bayes : $P(x,y) = P(x|y).P(y) = P(y|x).P(x)$ donne
$$I(x,y) = I(y) + I(x|y) = I(x) + I(y|x)$$
- Information mutuelle :
$$I(x;y) = log(p(x \mid y)/p(x)) = log(p(x,y)/(p(x)p(y))) = log(p(y \mid x)/p(y)) = I(y;x)$$
$$I(x;y) = I(x) - I(x|y)$$
  - *Si* $I(x;y) > 0$ => si un des événements se réalise, alors la probabilité d'occurrence de l'autre augmente. (si $I(x;y) < 0$ *…)*
  - *Si* $I(x;y) = 0$ => les deux événements sont statistiquement indépendants.

Mesures quantitatives moyennes de l 'information : entropie

- Comportement probabiliste moyen de la source:
  - La source est une variable aléatoire X qui réalise les événements (émet les symboles) x_i.
  - Elle est discrète, finie et ... stationnaire :

$$p_i = P(X=x_i) \text{ pour i de 1 à n et } \Sigma\, p_i = 1$$

- La quantité d 'information moyenne pour chaque xi est la moyenne E[.] de l'information de chaque événement $X = x_i$:

$$H(X) = E[I(X)] = \sum_{i=1}^{n} p_i \, I(x_i) = \sum_{i=1}^{n} p_i \log(1/p_i)$$

- H(X) est **l 'entropie de la source X** (entropie moyenne par symbole)

Définitions

- Entropie conjointe de deux variables aléatoires X et Y qui réalisent les événements x_i et y_j :

$$H(X,Y) = E[I(X,Y)] = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j)\, I(x_i,y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j) \log(1/P(x_i,y_j))$$

- Entropie conditionnelle :

$$H(X/Y) = E[I(X/Y)] = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j)\, I(x_i/y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j) \log(1/P(x_i/y_j))$$

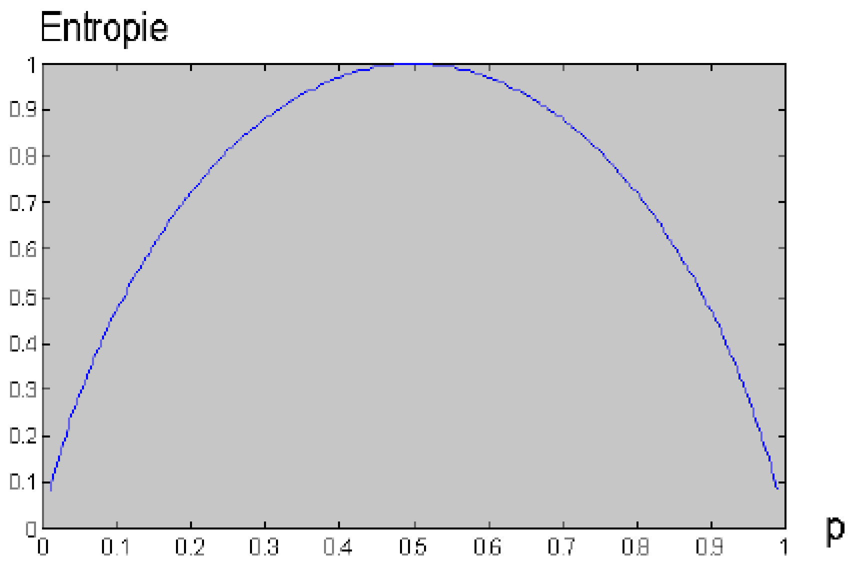
- Information mutuelle moyenne :

$$I(X;Y) = E[I(X;Y)] = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j)\, I(x_i;y_j)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j) \log(P(x_i,y_j)/P(x_i)P(y_j))$$

Propriétés (1/4)

- Exemple d'une variable aléatoire binaire X qui prend la valeur 1 avec proba p et 0 avec la proba (1-p).



Le maximum d'entropie est atteint pour : p=0.5

Propriétés (2/4)

- L'entropie H est positive ou nulle : $H(p_1,\ldots,p_n) \geq 0$
- L'entropie H est nulle si l'un des événements est certain
- L'entropie H est maximale pour $p_i=1/n$
- Le remplacement de $p_1,\ldots,p_n$ par des moyennes $q_1,\ldots,q_n$ conduit à une augmentation de l'entropir (convexité de l'entropie)

Propriétés (3/4)

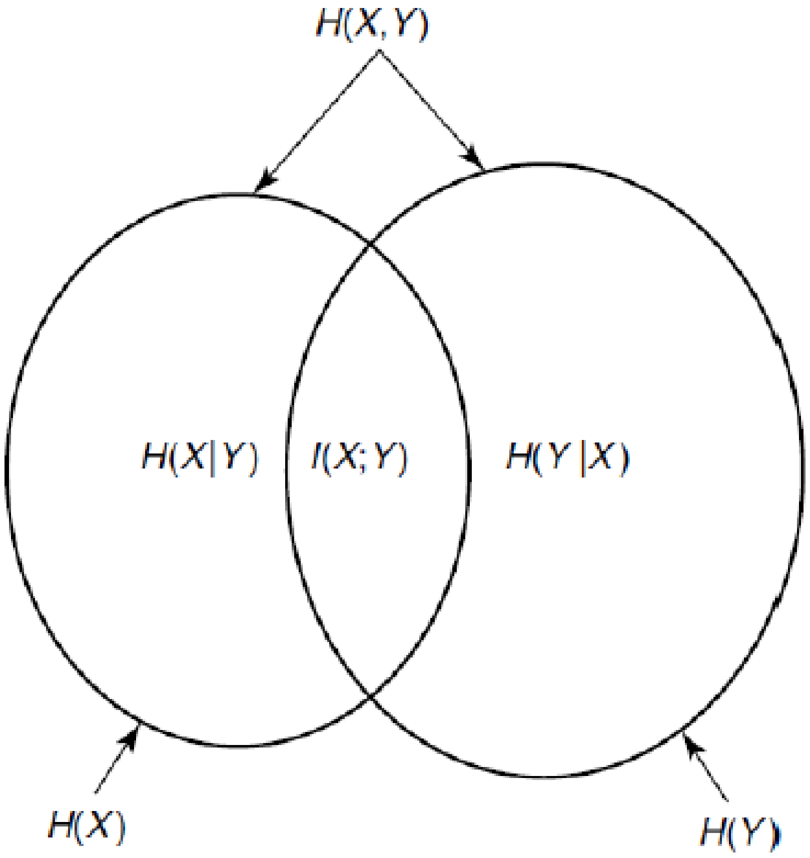
- $H(X,Y) = H(X) + H(Y|X)$
  $= H(Y) + H(X|Y)$

- $H(X,Y) \geq H(X)$ ou $H(Y)$
- $H(X|Y) \leq H(X)$
  (égalité ssi indépendance)

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

- $I(X;Y) = H(X) - H(X|Y)$
  $= H(Y) - H(Y/X)$
  $= H(X) + H(Y) - H(X,Y)$

- et donc $I(X;Y) \geq 0$
  (égalité ssi indépendance)



Propriétés (4/4) : Unicité

- Se démontre via :
  - Les axiomes de Khintchine
  - Les axiomes de Fadeev et Fenstein

- Ou
  - introduire l'information mutuelle moyenne de manière axiomatique,
    - en montrer l 'unicité
    - en déduire l 'entropie par : $I(X;X) = H(X)$
  - ou plus généralement $I(X;Y) = H(X)$ si $H(X|Y)=0$

Extension d’une source

- Soit une source codée par un alphabet Q-aire
  - par exemple Q = 2 donne l ’alphabet binaire [0,1]
- Une séquence de longueur k de symboles de cet alphabet constitue une nouvelle source S^k appelée k-ième extension de S
- Un bloc de k symboles de S est interprété comme un symbole de l ’alphabet Q^k-aire de S^k
- La fréquence d ’émission des symboles de S^k est 1/k fois celle de S
- Exemple :
  - Le code binaire correspondant à l'alphabet à 7 bits (0000000 à 1111111) est une extension de taille 7 de l'alphabet binaire.

Conclusion entropie

- Le mot information n 'a pas le sens du langage courant ...
  ... mais un sens technique lié au coût de transmission (temps)

- L 'entropie est ce qui caractérise le mieux un message dans un contexte de communication.

- L 'entropie ne dépend pas des symboles eux-mêmes, mais de l 'ensemble des probabilités associées à l 'alphabet en entier

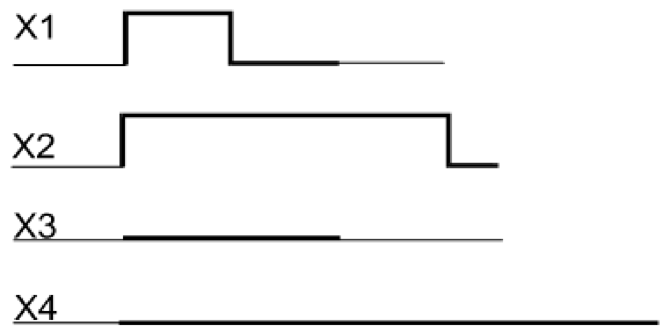
- L 'hypothèse de stationnarité de la source est essentielle à l'existence de l 'entropie de la source.

- Dans ce cadre, pas d 'adaptation ou d 'apprentissage ...

Sources et codage de sources

Introduction

Ex : Code morse, 4 symboles



- Source =
  - siège d'événements aléatoires qui constitue le message émis
  - caractérisation = **entropie**
- Codage de source
  - But : supprimer la redondance pour réduire le coût de transmission
  - 2 théorèmes fondamentaux de Shannon

Quels types de Sources ?

- Sources discrètes :
  - débitant des messages sous forme discrètes
  - **Sources discrètes sans mémoire** = la probabilité d'apparition d'un symbole ne dépend pas des symboles précédents :

$$p(x_{i_n} / x_{i_{n-1}}, x_{i_{n-2}}, \ldots) = p(x_{i_n})$$

  - **Sources discrètes avec mémoire** = modélisation par un processus stochastique : **chaînes de Markov :**

$$p(x_{i_n} / x_{i_{n-1}}, x_{i_{n-2}}, \ldots) = p(x_{i_n} / x_{i_{n-1}})$$

  - **Sources discrètes stationnaires** = probabilité d'apparition des différents symboles indépendants du temps :

$$p(x_{in}) = p(x_{in+k})$$

Sources discrètes avec mémoire (1/2)

- Caractérisation de n états
- Chaîne d'ordre 1 : la dépendance au passé = état précédent
- Toute chaîne de Markov discrète finie est équivalente à une chaîne d'ordre 1

- A l'état x_i est associé une probabilité p_{ij} de transition vers x_j

- Matrice de transition Π = matrice des (p_{ij})

- L'évolution est décrite par $P_{t+1} = \Pi.P_t$

Sources discrètes avec mémoire (2/2)

- La chaîne est régulière ou complètement ergodique si : $P_\infty = \lim_{t->\infty} P_t = P_1.\underline{\Pi}$ avec $\underline{\Pi} = \lim_{k->\infty} \Pi^k$

existe et est indépendante de P_1

- Dans ce cas, les lignes de $\underline{\Pi}$ sont les mêmes
- et on a $P_\infty = [\pi_1, \ldots, \pi_n]$ = probas stationnaires de chaque état.

- On peut alors généraliser la notion d 'entropie par la moyenne des entropies associées à chaque état :

$$H = \sum_{i=1}^{n} \pi_i \sum_{j=1}^{n} p_{ij} \log(1/p_{ij})$$

Codage des sources discrètes (1/) Définitions

- On note :
  - A = alphabet discret (fini ou infini dénombrable)
  - A^l = ens. Des l-uplets de lettres de A
  - $A^* = \cup_{l \geq 1} A^l$
- Un code de A est une procédure qui associe à chaque lettre une séquence binaire finie appelée mot de code
- Code : application $\varphi : A^* \to \{0,1\}^*$ / $(x_1,\ldots,x_n) \to (\varphi(x_1),\ldots, \varphi(x_n))$

Exemple :

alphabet $\{A_1, A_2, A_3, A_4\}$

	Code 1	Code 2	Code 3
A_1	00	00	0
A_2	01	01	10
A_3	10	001	110
A_4	11	011	111

Codage des sources discrètes (2/) Définitions

- Un code = régulier si injectif
  - 2 lettres différentes se codent différemment
  - On considère cela dans la suite !
- Code = déchiffrable : les symboles du code se séparent sans ambiguïté
  - Condition du préfixe : aucun mot de code n'est le début d'un autre (CNS déchiff.)
  - Code préfixe = représentation en arbre binaire
- Code séparable : pas de signe de démarcation entre les mots
- Code à longueur variable / fixe

	Code 1	Code 2	Code 3
A_1	00	00	0
A_2	01	01	10
A_3	10	001	110
A_4	11	011	111

Code 1 et 3 : préfixe

Code 2 : pb 1 et 3 et 2 et 4

Codage des sources discrètes (3/) Longueur et efficacité

- Code de longueur variable
  - Source discrète sur un alphabet $A=\{a_1,\ldots,a_K\}$ de loi de proba p_x, d'entropie par message H
  - Longueur moyenne : $L = \Sigma_{i=1}^{K} p_x(a_i).l_i$ avec l_i nombre de symboles binaires pour coder a_i
  - On a : $\mathbf{L \geq H/log(K)=l_{min}}$
  - Efficacité du codage : **E = H/(L.log(K))**

- Code de longueur fixe
  - Source discrète sur un alphabet $A=\{a_1,\ldots,a_K\}$
  - Il existe un code régulier de longueur n / $\log_2 K \leq n \leq 1+ \log_2 K$

Codage des sources discrètes (4/) Théorème fondamental (Shannon)

- limite inférieure à la suppression de redondance par codage ?
  - On considère une source stationnaire (avec ou sans mémoire) d 'entropie par message H
  - Ces messages sont codés par des mots de longueur moyenne L , exprimée en nombre de symboles d 'un alphabet K-aire
  - Alors il existe un procédé de codage déchiffrable où L est aussi voisin que l 'on veut de la borne inférieure

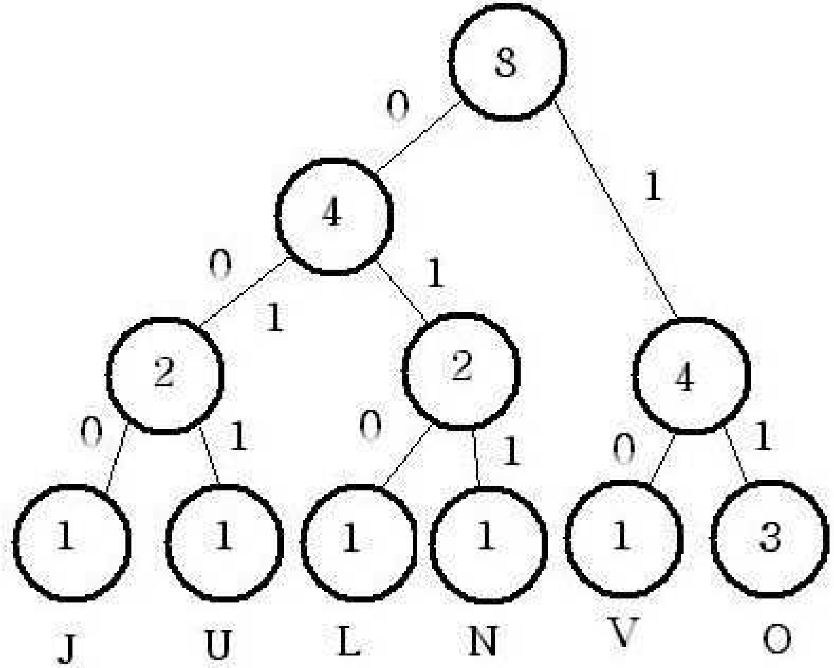
H / log(K)

Codage des sources discrètes (5/) Exemples

- Exemples de codages :
  - Codage sans perte :
    - Longueur fixe : Shannon-Fano, Huffman
    - Longueur variable : Lempel-Ziv (Lempel-Ziv-Welsh 84), Codage Arithmétique (Shannon-Fano-Elias)
    - Sont en + Adaptatif : Lempel-Ziv et Huffman
  - Codage avec pertes :
    - Images : JPEG (Huffman sur les plages de 0), MPEG …
    - Son : MP3
    - Zip, gzip => Lempel-Ziv 77

Codage binaire de Huffman

- Voir le cours de maths discrètes
- En résumé :
  - Algorithme de génération d'un **codage optimal** symbole par symbole.
  - Code à longueur variable => codes longs pour probas faibles
  - Algorithme :
    - Extraction des probabilités
    - Création de l'arbre
    - Création de la table d'Huffman
    - Codage
  - On transmet la table + les codes en binaire
    - Lecture de la table d'Huffman
    - Création de l'arbre de décodage
    - Lecture séquentielle et décodage



• mot "JULONOVO"

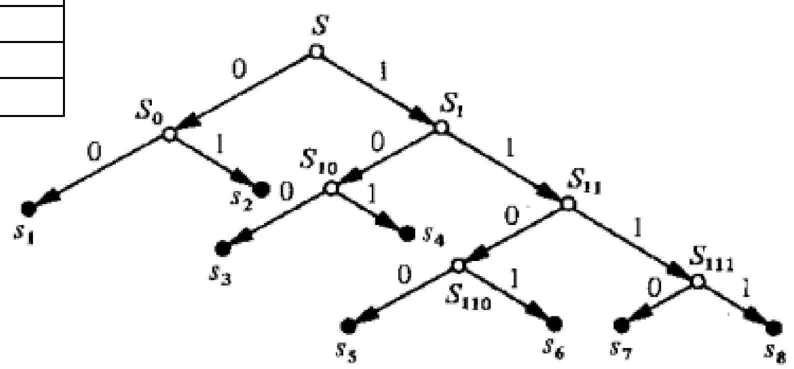
• "J, U, L, N, V" : 1 fois; "O" : 3 fois. Codage : J : 000 ; U : 001 ; L : 010 ; N : 011 ; V : 10 ; O : 11.

• "O" le plus fréquent = code le plus court

Codage de Shannon-Fano

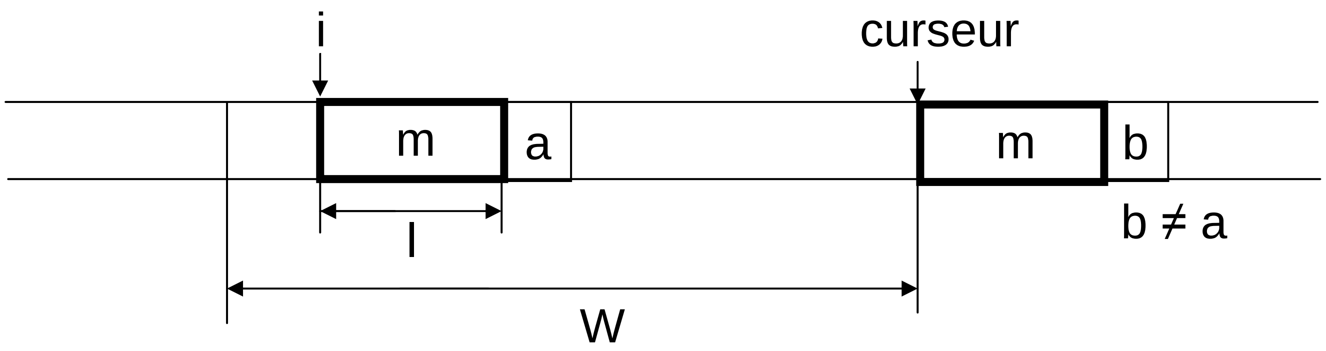
- Source discrète sans mémoire
- Algorithme de génération d'un **codage optimal absolu**, pour des sources divisibles récursivement (jusqu'à un symbole par ensemble) en deux sous-ensembles équiprobables.

Symboles s_k	Proba $p(s_k)$					Mots-codes c_k	Longueur l_k
s_1	0.25	0	0			00	2
s_2	0.25		1			01	2
s_3	0.125	1	0	0		100	3
s_4	0.125			1		101	3
s_5	0.0625		1	0	0	1100	4
s_6	0.0625				1	1101	4
s_7	0.0625			1	0	1110	4
s_8	0.0625				1	1111	4



Codage de Lempel-Ziv 77

- **Dictionnaire** de symboles incrémenté dynamiquement
- Fichier codé = suite des adresses des mots du dico
- Gérer l'incrément des bits d'adresse
- Implémentation :
  - Fenètre de taille W
  - Dictionnaire = mots de la fenêtre
  - Mot m codé par un couple (i,l)



(Entropie d'un langage)

- Alphabet usuel = 26 lettres
- I(chaque lettre) = $\log_2(26) \approx 4,7$ bits
- Si tient compte fréquence de chaque lettre :
  - Anglais : H(Anglais) ≈ 4,19 bits
  - Français : H(Français) ≈ 4,14 bits

- Ces valeurs diminuent si on tient compte des digrammes,…

- Entropie d'un langage L sur un alphabet A :

$$H(L)=\lim_{n \to \infty} (H(L)/n)$$

Canal et codage de canal

Introduction (1/2)

- Canal = endroit où on transmet et dégrade le message

  => **capacité**

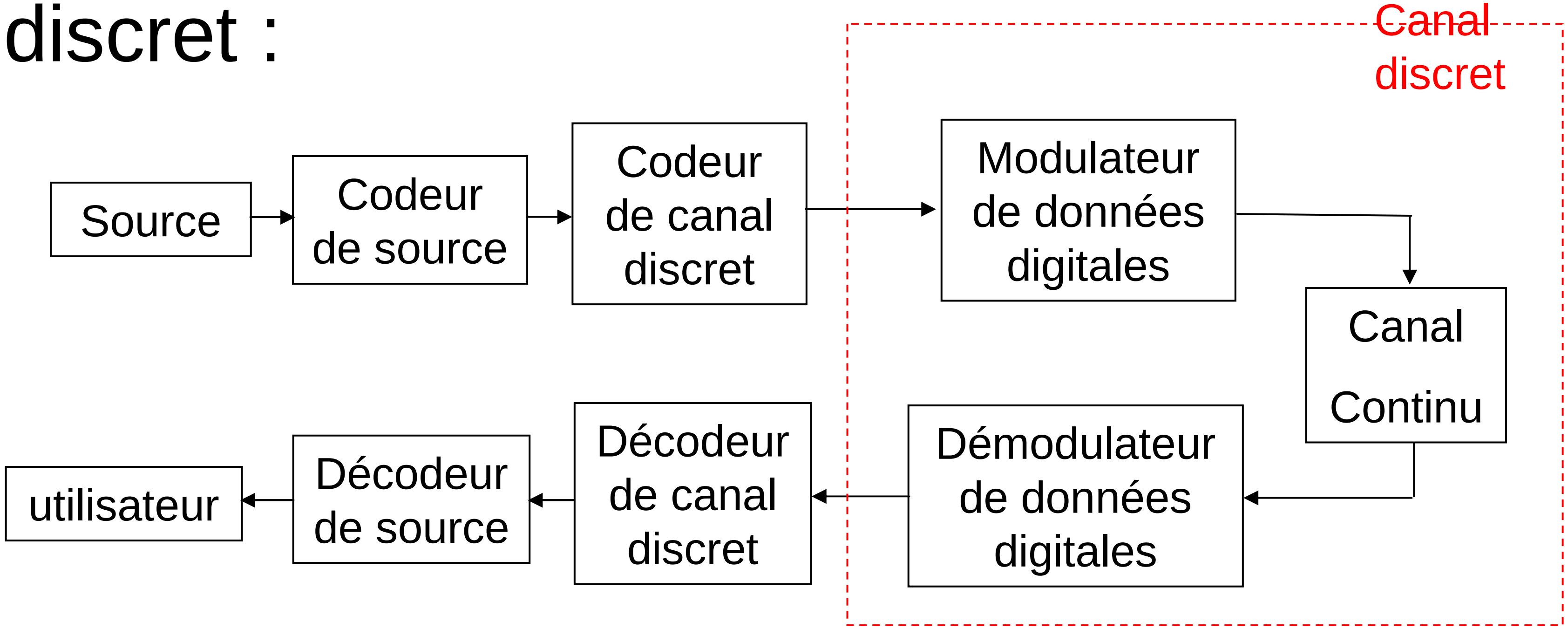
- Coder l'information de manière redondante pour rendre la détérioration négligeable

  => codes correcteurs d'erreurs

Canal

Introduction (2/2)

- Deux types de canaux : discret et continu
- Continu = peut être vu comme un canal discret :



Canal discret sans mémoire

- Défini par :
  - Un alphabet d'entrée $X=\{a_1,\ldots,a_K\}$
  - Un alphabet de sortie $Y=\{b_1,\ldots,b_J\}$
  - Une loi de transition de probabilité définie par $p(b_j \mid a_k)$
  - La matrice KxJ (matrice stochastique du canal) :

$$\Pi = \begin{pmatrix} P(b_1 \mid a_1) & \ldots & P(b_J \mid a_1) \\ \vdots & \ddots & \vdots \\ P(b_1 \mid a_K) & \ldots & P(b_J \mid a_K) \end{pmatrix}$$

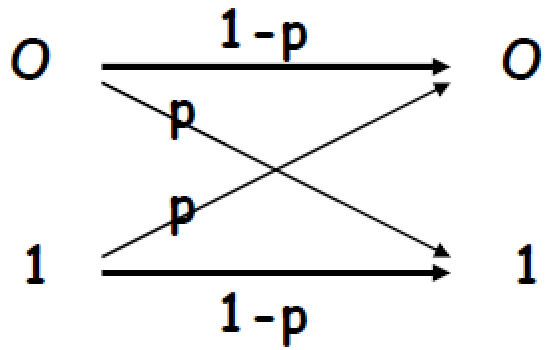
Capacité d’un canal

- Définition informelle :
  - La capacité C d ’un canal est la plus grande quantité d ’information moyenne qu ’il est capable de transmettre de son entrée à sa sortie.
  - On considère toutes les sources possibles à l’entrée.

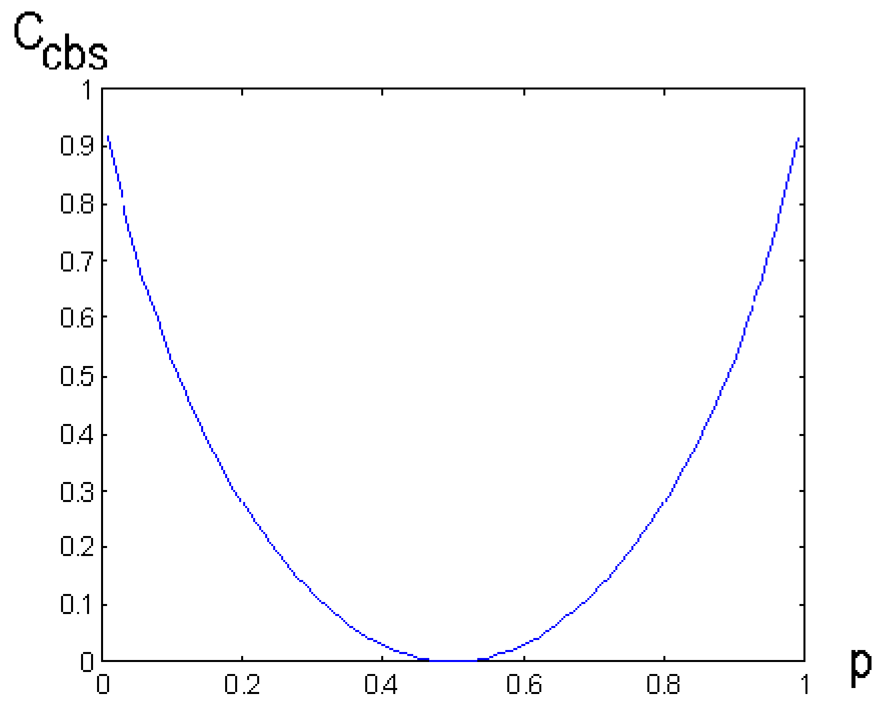
- Formellement :
  - La capacité C d ’un canal est le maximum de l’information mutuelle moyenne I(X;Y) avec X entrée, Y sortie.
  - Remarque : $I(X;Y) = H(X) – H(X/Y)$
    - Ici H(X/Y) peut s’interpréter comme l ’ ambiguïté à la réception, liée au canal (au bruit contenu dans le canal).
    - Pour une communication effective, il faut H(X/Y) négligeable.

Exemple de modélisation (1/2)

- Canal binaire symétrique (stationnaire sans mémoire)

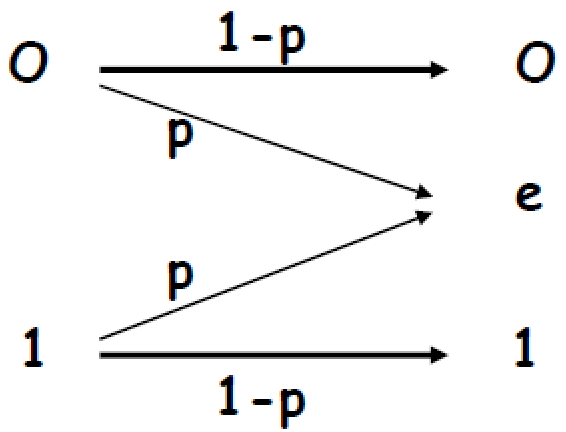


$$C_{cbs} = 1 + (1-p)\log_2(1-p) + p\log_2(p)$$

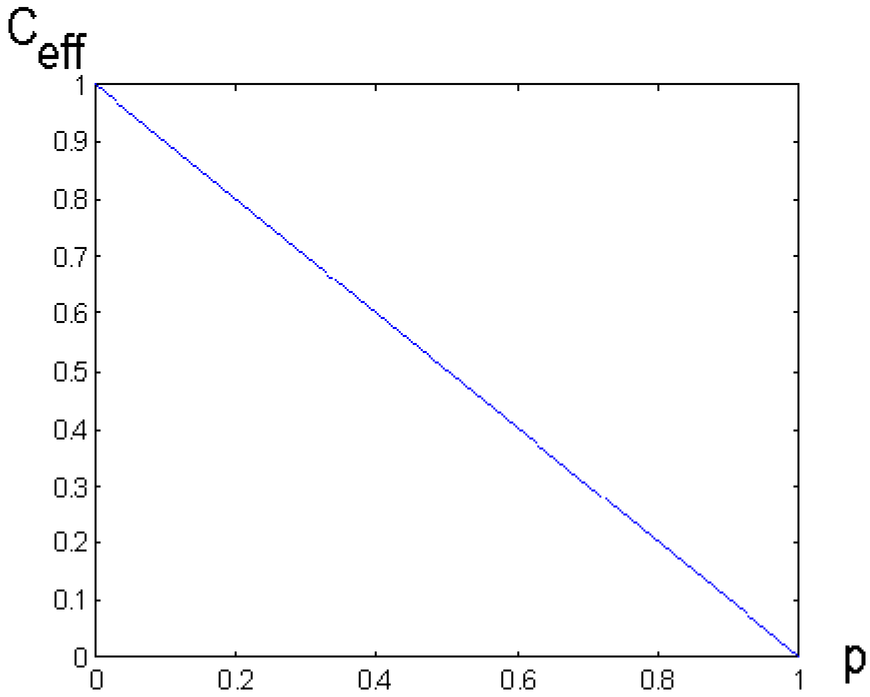


Exemple de modélisation (2/2)

- Canal binaire à effacement (Canal stationnaire sans mémoire)



$$C_{eff} = 1 - p$$



Taux de codage

- Le taux d'un code binaire de longueur n (après codage) et de cardinal M (taille de l'ens de départ) est égal à :

  $$\mathbf{R = (\log_2 M) / n} \text{ bits par transmission}$$

- Le **taux d'erreur résiduel P_e** d'un algorithme de décodage
  - = moyenne de $P(f(y) \neq x)$ avec y n-uplet reçu et x mot de code envoyé
  - Probabilité décodage incorrect après passage dans le canal

Théorème fondamental du codage de canal

- Soit un canal discret sans mémoire de capacité C. Pour tout taux R<C, il existe un code ayant des mots de longueur n, de sorte que le taux d'erreur résiduel P_e vérifie $P_e \leq 2^{-n.R}$

- Résultat inattendu !
- en pratique, si R<0,5.C, des codes existent avec P_e faible.

Codage de canal : codes correcteurs/détecteurs d'erreurs

Distance de Hamming

- A^n = ens. des mots de longueur n sur A
- $\underline{x} = (x_0,\ldots,x_{n-1})$ et $\underline{y} = (y_0,\ldots,y_{n-1})$ dans A^n
- La distance de Hamming entre $\underline{x}$ et $\underline{y}$ est

$$d_H(\underline{x},\underline{y}) = |\{i \;/\; 0 \leq i \leq n-1,\; x_i \neq y_i\}|$$

- Bien une distance !

Codes linéaires en blocs (1/2)

- Un code C sur A de dimension k et de longueur n code k symboles de A en n symboles de A avec k< n (redondance !).
- Les éléments de C sont appelés mots de code.
- Un code est linéaire possède une matrice génératrice G de taille k x n qui peut être mise sous forme systèmatique :

$$G = (I_k \mid P)$$

- On calcule les mots de code de la façon suivante :
  Soit $\underline{u}$ un mot de A^k, le mot de code correspondant à $\underline{u}$ est :

$$\underline{v}=\underline{u}.G$$

- La distance minimale d'un code C est :

$$d = \min\{d_H(\underline{x},\underline{y}) \ / \ \underline{x},\underline{y} \text{ dans C et } \underline{x}\neq\underline{y}\}$$

Codes linéaires en blocs (2/2)

- Un code C est donc défini par le quaduplet (A,n,k,d), on note C(A,n,k,d)
- Un code de distance minimale d peut corriger (d-1)/2 erreurs.
- Pour le décodage, on utilise la matrice de parité H de taille (n-k) x n qui est telle que :

$$H.{}^tG = 0 \text{ où } {}^tG \text{ est la transposée de } G$$

$$\text{On a : } H = (-P \mid I_{n-k})$$

- La détection des erreurs utilise la matrice de contrôle H (décodage au maximum de vraisemblance)
  - Pour les mots $\underline{v}$ du code, on a $\underline{v}.{}^tH = 0$
  - Pour les mots avec erreurs $\underline{r} = \underline{v} + \underline{e}$ on a $\underline{r}.{}^tH = (\underline{v} + \underline{e}).{}^tH = \underline{e}.{}^tH = \underline{s}$.
  - $\underline{s}$ est le syndrome d'erreur.
  - La table de configuration d 'erreurs permet de déduire $\underline{e}$ de $\underline{s}$.

Exemple

$$[G_2] = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

- k=2,n=3
- Alphabet : {0,1}
- G de taille 2x3

- H de taille 1x3 :

  [H]=[1 1 1]

$$[0 \quad 0] \times \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} = [0 \quad 0 \quad 0]$$

$$[0 \quad 1] \times \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} = [0 \quad 1 \quad 1]$$

$$[1 \quad 0] \times \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} = [1 \quad 0 \quad 1]$$

$$[1 \quad 1] \times \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} = [1 \quad 1 \quad 0]$$

Codes parfaits

- Un code parfait est un code tel que l'ensemble des boules de rayon (d-1)/2 centrées en tous les éléments du code forment un partition de A^n.

- Exemple : code de Hamming H_m
  - Code binaire linéaire de paramètres $n=2^m-1$, $k=2^m-m-1$, $d=3$).
  - La matrice de parité de ce code = 2^m-1 vecteurs colonnes

$$H_3 = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}$$

  - C'est un code parfait (démo !)

Codes détecteurs d'erreurs

- Un code de distance minimale d peut détecter d-1 erreurs.

- Exemple : code de parité de longueur n

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

  - Permet de détecter une erreur
  - Exple : Codage des 128 caractères ASCII sur des mots de longueur 8 avec bit de parité sur le dernier bit.

Codes cycliques : corps finis (1/6)

- Groupe : un ensemble muni d'une loi (en général notée +) qui est :
  - Commutativité : x+y = y+x mod n
  - Associativité : (x+y)+z = x+(y+z) mod n
  - Élément neutre : 0+x=x+0=x mod n
  - Existence d'un opposé : x-x=0 mod n

- Exemple ($\mathbb{Z}$**/n**$\mathbb{Z}$**,**+) est un groupe

Codes cycliques : corps finis (2/6)

- Anneau : un ensemble muni de deux lois (en général notées + et .) :
  - Groupe pour la première loi (+)
  - La multiplication (.,2ème loi) conserve :
    - La commutativité
    - L'associativité
    - L'élément neutre 1
    - L'élément absorbant 0
    - La distributivité par rapport à l'addition
    - PAS L'EXISTENCE D'UN INVERSE

- Exemple ($\mathbb{Z}/n\mathbb{Z}$,+,.) est un anneau
- Si existence d'un inverse => CORPS
- Exemple ($\mathbb{Z}/n\mathbb{Z}$,+,.) est un corps ssi n est premier

Codes cycliques : corps finis (3/6)

- Soit F un corps fini de cardinal q > 1. Alors
  - $q=p^m$ où p est un nb premier et m entier positif
  - F est unique à isomorphisme près.

- Propriétés :
  - Si p premier alors $\mathbb{F}_p$ est égal à $\mathbb{Z}/p\mathbb{Z}$
  - Si $q=p^m$ $(m>1)$, $\mathbb{F}_q$ est une extension de degré m de $\mathbb{F}_p$.

Codes cycliques : corps finis (4/6)

- Soit $\mathbb{F}_q$ un corps fini. Soit $p_m(x)$ un polynôme irréd. de $\mathbb{F}_q[x]$ de degré m. Soit $(p_m(x))$ l'ens. des polynômes multiples de $p_m(x)$.
  - Le quotient $\mathbb{F}_q[x]/(p_m(x))$ est un corps fini de cardinal q^m.
  - Il existe toujours un élément de $\mathbb{F}_{q^m}$ tel que $p_m(a)=0$.

- Un polynôme $p_m(x)$ irréductible est dit primitif si l'ens. des restes de x^i par $p_m(x)$ sont tous distincts pour $0 \leq i < q^m-1$.

- Un élément a tel que $p_m(a)=0$ est dit primitif.

Codes cycliques : corps finis (5/6)

- Soit a un élément primitif de $\mathbb{F}_{q^m}$, on notera $\mathbb{F}_{q^m} = \mathbb{F}_q[a]/(p_m(a))$
  - Le corps fini à q^m éléments = $\{0,1, a, a^2,\dots, a^{q^m-2}\}$.
  - Le corps fini à q^m éléments = l'ens des polynômes en a de degré < m
  - L'addition = l'addition de deux polynômes de $\mathbb{F}_q[a]$.
  - La multiplication de deux éléments sera le reste de la division par $p_m(a)$ de la multiplication de deux polynômes de $\mathbb{F}_q[a]$.

Codes cycliques : corps finis (6/6)

- Exemple l'AES
  - Corps de base $\mathbb{F}_2=\{0,1\}$ muni de + (ou-exclusif) et de . (AND)
  - Extension : $\mathbb{F}_{2^8} = \mathbb{F}_2[x]/(x^8+x^4+x^3+x+1)$
  - Addition
    - $(x^6+x^4+x^2+x+1)+(x^7+x+1) = x^7 + x^6 + x^4 + x^2$ (notation polynomiale)
    - $\{01010111\} \oplus \{10000011\} = \{11010100\}$ (notation binaire)
    - $\{57\} \oplus \{83\} = \{d4\}$ (notation hexadecimale).
  - Multiplication
    - $\{57\} \cdot \{83\} = \{c1\}$ car :

$$\begin{aligned}(x^6+x^4+x^2+x+1)\,(x^7+x+1) &= x^{13}+x^{11}+x^9+x^8+x^7+ \\ &\quad x^7+x^5+x^3+x^2+x+ \\ &\quad x^6+x^4+x^2+x+1 \\ &= x^{13}+x^{11}+x^9+x^8+x^6+x^5+x^4+x^3+1\end{aligned}$$

$$\begin{aligned}x^{13}+x^{11}+x^9+x^8+x^6+x^5+x^4+x^3+1 \ \mathtt{modulo}\ (x^8+x^4+x^3+x+1) \\ = x^7+x^6+1.\end{aligned}$$

Codes cycliques

- Alphabet : $\mathbb{F}_q$, m un entier, $n=q^m-1$.
- On note R= $\mathbb{F}_q[x]/(x^n-1)$= anneau de polynômes à coeff dans $\mathbb{F}_q$.

- Code cyclique = code linéaire + invariance par permutation circulaire dans la base $(1,x,x^2,\ldots,x^{n-1})$

- Un code cyclique :
  - $g(x)$ facteur de x^n-1. Appelé polynôme générateur.
  - Un code C est tel que : Tout c(x) de C s'écrit de façon unique $c(x)=f(x)g(x)$ dans $\mathbb{F}_q[x]$.
  - La dimension de C est n-r avec r = deg(g)

Codes cycliques : exemple

- Sur $\mathbb{F}_2$, si n =7, k=4
  - $1+x^7 = (1+x)\times(1+x^2+x^3)\times(1+x+x^3)$
  - g(x) est de degré 3 donc :

    $g(x) = (1+x^2+x^3)$ ou $g(x) = (1+x+x^3)$
  - Matrice génératrice :

$$G_{(k,n)} = \begin{bmatrix} g(x) \\ x.g(x) \\ \ldots \\ x^{k-1}.g(x) \end{bmatrix} \qquad G_{(4,7)} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

Codes cycliques

- Essentiellement deux types de codes cycliques :
  - Codage par multiplication
  - Codage par division (voir CRC)
  - Décodage par division : calcul du reste de v'(x) reçu par g(x)
    - Reste = 0 => pas d'erreur
    - Sinon erreur

Codes cycliques : exemple (CRC)

- Une séquence de k symboles binaires est représenté par un polynôme i(x).
  - g(x) poly de degré s
  - Mot de code correspondant à i(x) est :

    $c(x) = x^{s} \cdot i(x) + r(x)$ avec $r(x)$ reste de $x^{s} \cdot i(x) / g(x)$
  - Détection de toute rafale d'erreurs ≤ s

- CRC-16 : $x^{16} + x^{12} + x^{5} + 1$ sur $\mathbb{F}_2$ (norme CCITT N°41)

- CRC-32 : $x^{32} + x^{26} + x^{23} + x^{22} + x^{16} + x^{12} + x^{11} + x^{10} + x^{8} + x^{7} + x^{5} + x^{4} + x^{2} + x + 1$ (IEEE 802 réseau locaux)

Autres codes

- Codes convolutifs
  - Codage en flux continu
  - Décodage par l’algorithme de Viterbi
- LDPC
- …

Conclusion générale

- Théorie de l’information = méthode pour modéliser les communications

- Mais aussi pour la crypto !
  - Entropie maximale pour le téléphone rouge
  - Confusion/Diffusion