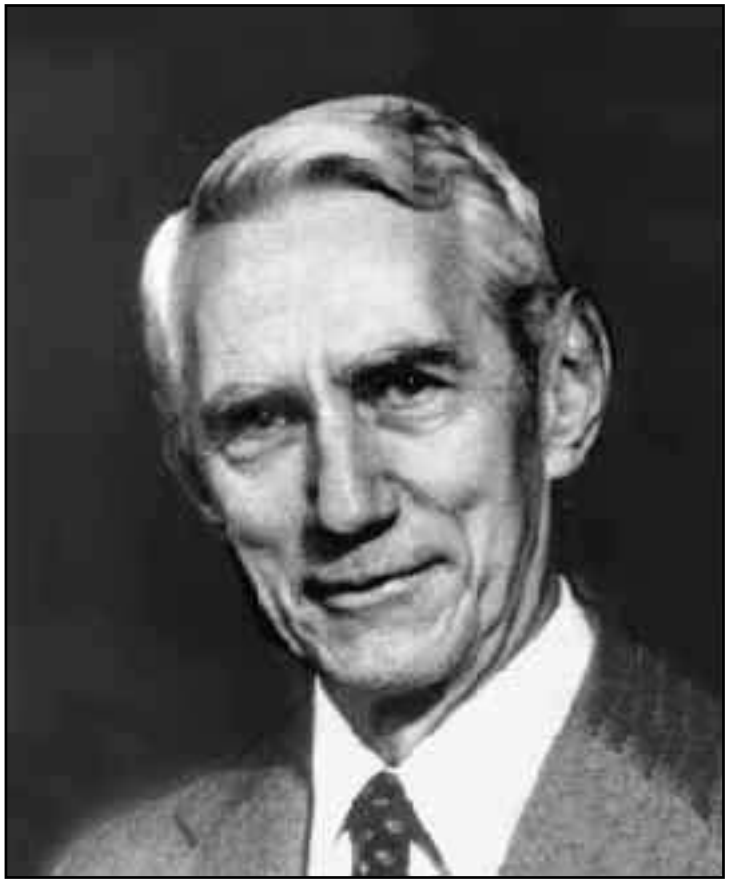


Claude Elwood SHANNON
30 Avril 1916 / 24 Février 2001

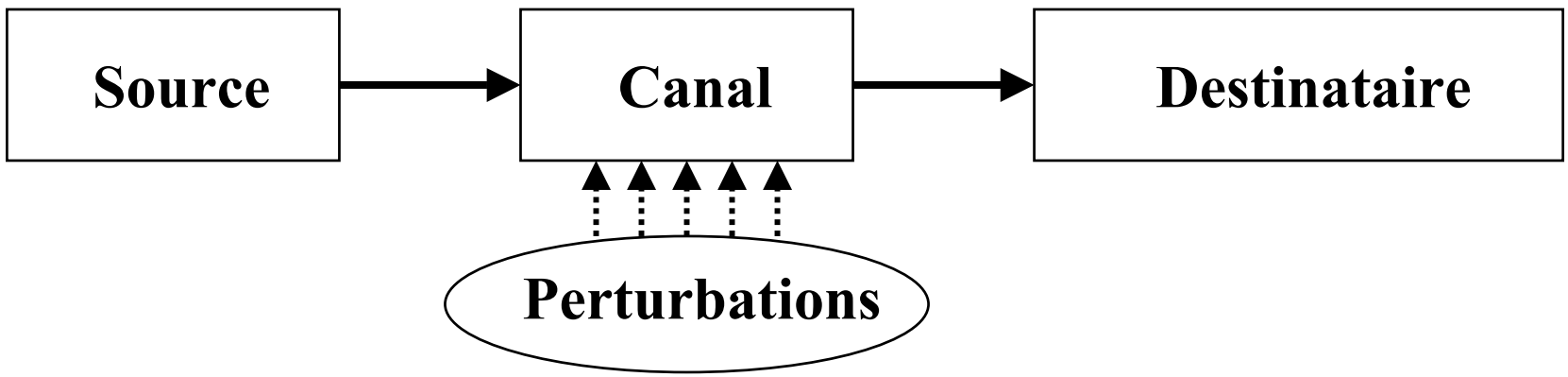
Introduction à la Théorie de l'Information

- **Information**
  - Renseignement obtenu de quelqu'un ou sur quelqu'un ou quelque chose - Nouvelle communiquée par une agence de presse, un journal, la radio, la télévision. Abrév.*(fam.) : info*.
  - Eléments de connaissance susceptibles d 'être codés pour être conservés, traités ou communiqués.

(Le Petit Larousse Illustré)

- **Message**
  - Information, nouvelle transmise à quelqu'un. (Le P.L.I.)
  - Lot d'informations formant un tout intelligible ou exploitable et transmis en une seule fois.

- **Signal**
  - Variation d'une grandeur de nature quelconque grâce à laquelle, dans un équipement, un élément en influence un autre.
  - Phénomène physique porteur d'une information et pouvant représenter des données.



- Schéma général de la « communication »
- Perturbations
  - Bruit thermique
  - Rayonnement électromagnétique
  - défauts des supports de stockage
  - …
- A compléter par les appareillages d 'adaptation de la *source* et du *destinataire* au *canal*

- **Le *Traitement du Signal* : étude des signaux qui se propagent dans le canal et des perturbations qu 'ils subissent.**

- **La *Théorie de l 'Information* : étude de l 'information transmise par ces signaux**
- **Mesure quantitative de l 'information des messages (et dégradations ...), indépendamment des codages ou signaux utilisés pour la transmission.**

- Seule la composante matérielle (formant) d'une information fait l'objet d'une communication: ce n'est pas le sens (formé) que l'on transmet
- L'information est en relation directe avec la notion d 'ordre (structure ordonnée), mesurée par la néguentropie (entropie -> désordre)

  Etymologie: *informare = donner une forme ...*

- 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)

- 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 ...

- *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

- *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.**

I. MESURE QUANTITATIVE DE L 'INFORMATION

Contexte :

- **L 'information est vue du point de vue des « *techniques de communications* »**
- **Un message est une suite de symboles appartenant à un ensemble fini, pré-déterminé, « *l 'alphabet* ».**
- **Alphabet:** Ensemble fini de symboles
  - **Lettres :** **a b c d e ...**
  - **Alphabet binaire :** **0 1**
- **Message:** Suite finie de symboles
  - **« A l'échelle cosmique, l'univers c'est tout petit ! »**
  - **01101001010101100010100011101001011101**
- **Source de messages:** Ensemble de TOUS les messages susceptibles d'être formés à partir d'un alphabet

- 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 a priori.

« Plus un message est imprévu, improbable, plus il est informatif »

- La quantité d 'information I(x) apportée par la réalisation d 'un événement x de probabilité P(x) :

  - est donc une fonction croissante f de son improbabilité :
    $I(x) = f(1/P(x))$
  - *nulle* si $P(x) = 1$. $f(1) = 0$
  - s 'ajoute à celle d 'un événement y indépendant de x :
    $I(x,y)=f(1/P(x).1/P(y)) = f(1/P(x)P(y)) = f(1/P(x)) + f(1/P(y)) = I(x)+I(y)$
  - toujours *positive* (et *additive*)

- 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
- On peut montrer que c 'est le choix le plus logique, car ...

  ... c 'est le seul possible !!!

Mesures quantitatives de l 'information PAR EVENEMENTS :

- **Quantité d 'information** $I(x) = \log \frac{1}{P(x)} = -\log P(x)$
- **Information conjointe** $I(x,y) = \log \frac{1}{P(x,y)} = -\log P(x,y)$
- **Information conditionnelle** $I(x/y) = \log \frac{1}{P(x/y)} = -\log P(x/y)$
- **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**

$$I(x;y) = \log \frac{P(x/y)}{P(x)} = \log \frac{P(x,y)}{P(x).P(y)} = \log \frac{P(y/x)}{P(y)} = I(y;x)$$

Mesures quantitatives MOYENNES de l 'information :

- **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.

$$p_i = P(X = x_i) \quad i = 1,2,\ldots,n \quad \text{et} \quad \sum_{i=1}^{n} p_i = 1$$

- **La quantité d 'information moyenne pour chaque x_i 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)

Mesures quantitatives MOYENNES de l 'information (suite):

- **Entropie conjointe :**

  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 :**

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

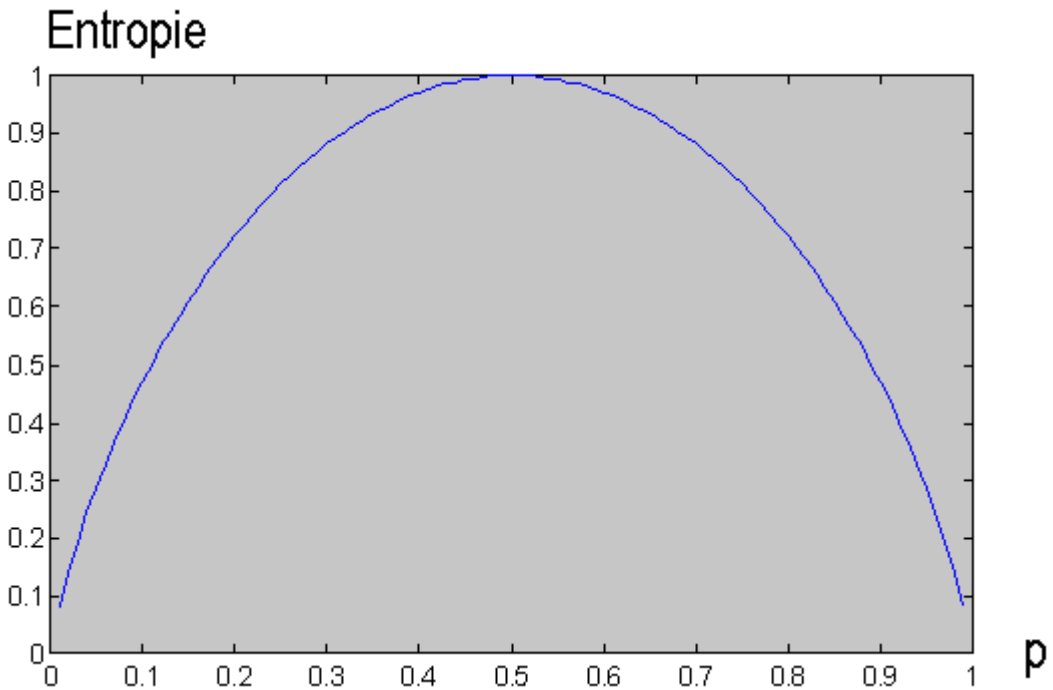
- **Information mutuelle moyenne**

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

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

Propriétés

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



Le maximum d 'entropie est atteint pour :

$p = 0.5$

Propriétés de l 'entropie $H(p_1,p_2, \ldots p_n)$

- L 'entropie H est non-négative : $H(p_1,p_2, \ldots p_n) >= 0$
- 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 $p_1,p_2, \ldots p_n$ par des moyennes $q_1,q_2, \ldots q_n$ conduit à une augmentation de l 'entropie (convexité de l 'entropie).

$$q_i = \sum_{j=1}^{n} a_{ij} p_j \quad \text{avec } a_{ij} \geq 0 \qquad \sum_{j=1}^{n} a_{ij} = 1, \quad \forall i \qquad \sum_{i=1}^{n} a_{ij} = 1, \quad \forall j$$

Propriétés des entropies conjointe, conditionnelle et de l 'information mutuelle

- H(X,Y) = H(X) + H(Y/X) = H(Y) + H(X/Y)
- H(X,Y) >= H(X) ou H(Y)
- H(X/Y) <= H(X) (égalité ssi indépendance)
- H(X,Y) <= H(X) + H(Y) <= 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) >=0 (égalité ssi indépendance)

Unicité de l 'expression de l 'entropie

- Axiomes de Khintchine :
  - La fonction $H(p_1,p_2, \ldots p_n)$ est maximale pour $p_i = 1/n$
  - $H(X,Y) = H(X) + H(Y/X) = H(Y) + H(X/Y)$
  - $H(p_1,p_2, \ldots p_n) = H(p_1,p_2, \ldots p_{n,}, 0)$

- Axiomes de Fadeev et Fenstein
  - si $p_1=p_2.=\ldots=.p_n= 1/n$, H est fonction monotone croissante de n
  - H(p, 1-p) est une fonction continue de p sur [0,1]
  - X et Y sources indépendantes, $p(x_i) = 1/n$ et $p(y_j) = 1/m$
    $H(\ldots,1/mn, \ldots) = H(\ldots,1/n, \ldots) + H(\ldots,1/m, \ldots)$
  - On n'apporte pas plus d'information en fractionnant les expériences .... (voir polycop) .

Unicité de l 'expression de l 'entropie

- On peut même :
  - 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 $H(X/Y)=0$

Extensions 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.

Commentaires

- 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 ...

II. SOURCE ET CODAGE DE SOURCE

Quelques caractéristiques de sources:

- **Sources Stationnaires**

  Propriété suffisante à la définition d'une entropie

- **Sources Ergodiques**

- **Sources *sans* mémoire**

- **Sources *avec* mémoire**

- **Une modélisation possible des sources avec mémoire :**

  Les Chaînes de MARKOV

Les chaînes de MARKOV :

- Caractérisée par n états (ex: les lettres)
- Chaîne d'ordre 1 : la dépendance au passé se résume à la dépendance à l'état atteint
- 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
- On définit ainsi Π avec $\sum_{j=1}^{n} p_{ij} = 1$

- L'évolution est décrite par : $P_{t+1} = P_t \,.\, \Pi$ ou $P_{t+1} = P_1 \,.\, \Pi^t$

Les chaînes de MARKOV régulières:

- La chaîne est *régulière* ou *complètement ergodique* si :

$$P_\infty = \lim_{t\to\infty} P_t = P_1 .\overline{\Pi} \qquad \text{avec} \qquad \overline{\Pi} = \lim_{k\to\infty} \Pi^k$$

existe et est indépendante des conditions initiales P_1

- Dans ce cas, les lignes de $\overline{\Pi}$ sont toutes les mêmes

$$P_\infty = [\pi_1, \pi_2, \pi_3 ... \pi_n]$$ qui regroupe les probabilités « 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 de sources:

- **Objectif : supprimer la redondance de la source**

Propriétés a priori d 'un code :

- **Régularité** : deux messages différents se codent différemment
- **Déchiffrabilité** : les symboles du code se séparent sans ambiguïté
  - Symboles de longueur constante
  - Symbole de séparation
  - Un symbole n 'est pas le début d 'un autre : *codes irréductibles*

Théorème fondamental du codage de source

- Quelle est la 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 ñ , exprimée en nombre de symboles d 'un alphabet q-aire .

Il existe un procédé de codage déchiffrable où ñ est aussi voisin que l 'on veut de la borne inférieure

$$H / \log(q)$$

Exemples de codages :

- **Codage *sans* perte :**
  - **Longueur fixe : Shannon-Fano, Huffman**
  - **Longueur variable : Lempel / Ziv** (Lempel/Ziv/Welsh : LZW)
  - **Adaptatif : exploitation de la non stationnarité de la source** (LZ77).

    Les résultats peuvent être meilleurs que ce que prévoit la théorie pour les sources stationnaires
  - **Codage Arithmétique**

- **Codage *avec* pertes :**
  - **Théorie de la distorsion**
  - **Codages prédictifs (le plus souvent ...)**
  - **Images : JPEG, MPEG …**
  - **Son : MP3**

III. CANAL ET CODAGE DE CANAL

Définition de la *capacité* d 'un canal :

- **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.**

Autre définition :

- **La capacité C d 'un canal est le maximum de l'information mutuelle moyenne I(X;Y) avec X entrée, Y sortie.**

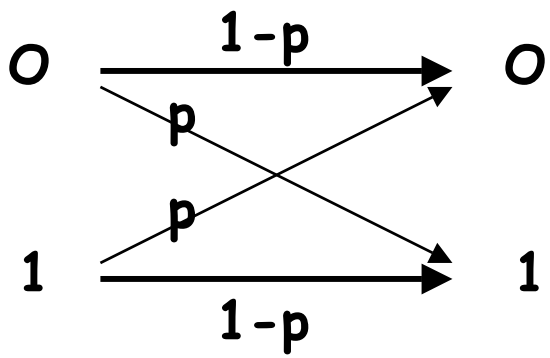
- **Remarque : I(X;Y) = H(X) – H(X/Y)**

  Ici H(X/Y) peut s'interpréter comme l ' *ambiguïté* à la réception, liée au canal (au bruit contenu dans le canal).

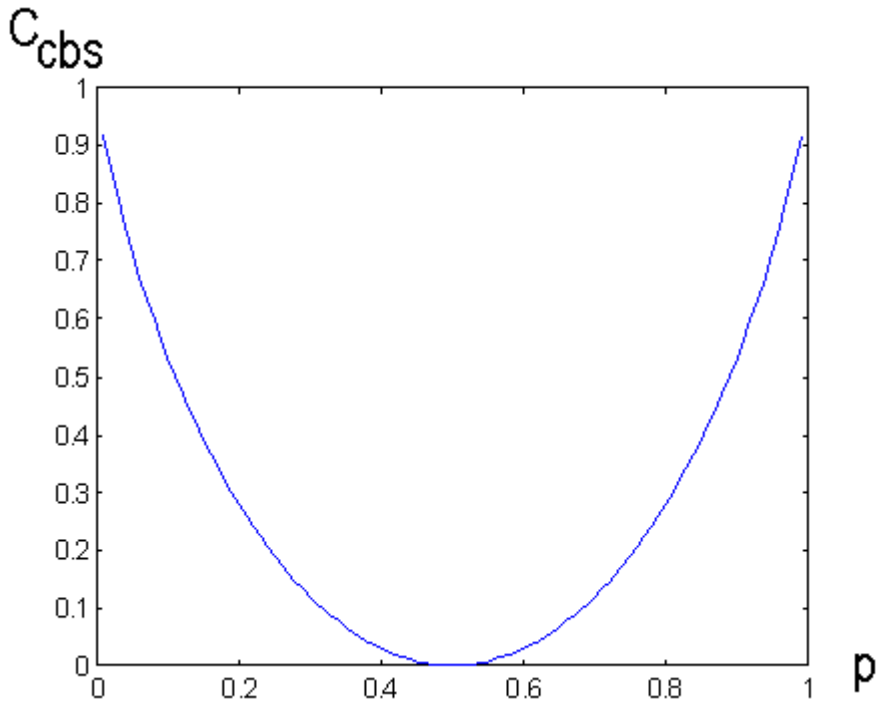
  Pour une communication effective, il faut H(X/Y) négligeable.

Exemple de Modélisation d'un canal :

Canal binaire symétrique (Canal stationnaire *sans* mémoire)

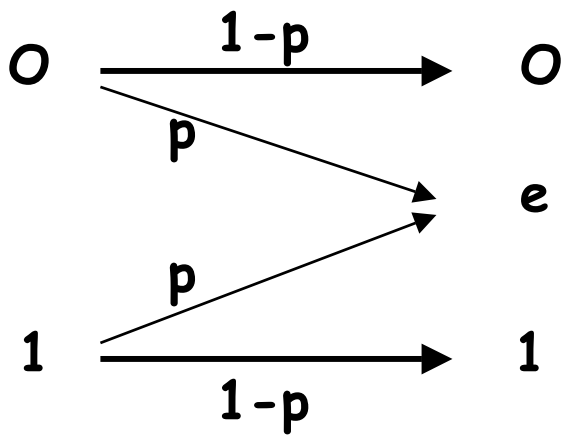


$$C_{cbs} = 1 + (1-p)\log_2(1-p) + p\log_2(p)$$

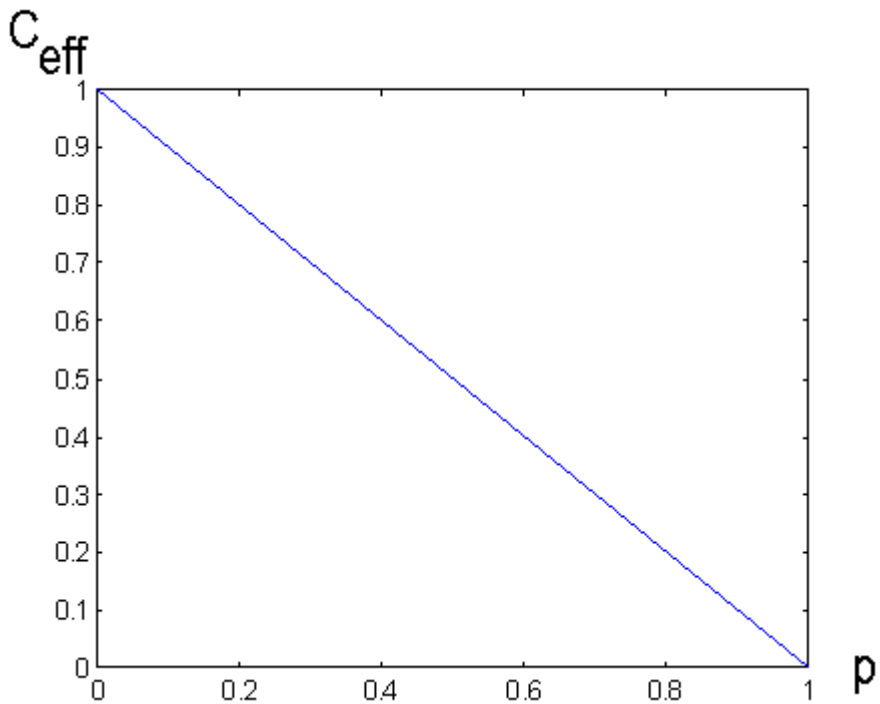


Exemple de Modélisation d'un canal :

Canal binaire à effacement (Canal stationnaire *sans* mémoire)



$C_{eff} = 1 - p$

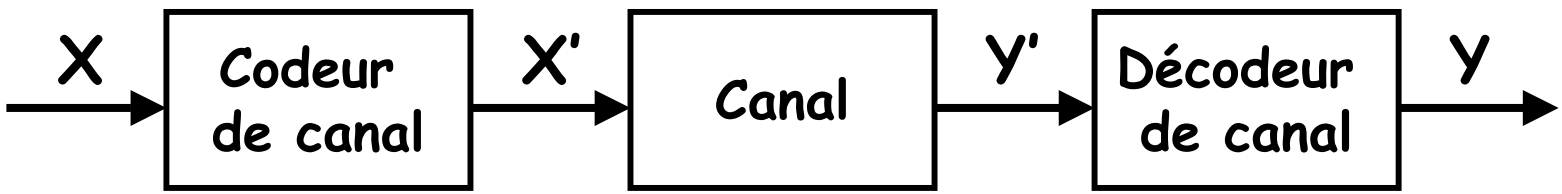


Autres Modélisations possibles :

- Canal stationnaire *sans* mémoire
  - Canal q-aire en entrée et en sortie
  - Canal q-aire en entrée et en sortie avec ambiguïté partielle
  - Canal binaire à bruit additif gaussien
  - ...
- Canal causal à mémoire finie

Codage de canal :

- Objectif : ré-introduire de la redondance pour permettre la détection / correction d 'une grande partie (toutes ?) des erreurs de transmission.



Théorème fondamental du codage de canal

- Quelles sont les limites du codage de canal pour la correction des erreurs par redondance ?

On considère un canal stationnaire causal de capacité ergodique C (et de mémoire finie m) et une source stationnaire et ergodique d 'entropie H < C, et $\varepsilon > 0$.

Il est alors possible de coder un message X de n symboles (si n assez grand !) par un message X' de n+m symboles qui sera émis. Le message Y' reçu (avec erreurs éventuelles) et décodé en Y permettra de retrouver X avec une probabilité supérieure à $1-\varepsilon$.

Codages détecteur / correcteur d 'erreurs :

- Ces codes utilisent la redondance pour cette détection et correction éventuelle.
- Un bloc de k symboles est codé par n symbole (k < n)
- La redondance sert à assurer la cohérence des mots du code à la réception (contrôle de parité).

- Métrique de Hamming :
  - La métrique de Hamming ou *distance de Hamming* joue un rôle clef dans l 'analyse de ces codes
  - d_H([010010],[001011]) = 3
  - d_H correspond bien à une distance

Codes linéaires par blocs :

- Ils sont définis par une *matrice génératrice* G de taille (k x n)
- Un mot $\underline{u}$ de k symboles est codé par un mot $\underline{v}$ du code de n symboles, avec :

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

- La détection des erreurs utilise la *matrice de contrôle* H construite à partir de G et telle que $G.H^t = [0]$.
- Pour les mots $\underline{v}$ du code, on a $\underline{v}.H^t = \underline{0}$
- Pour les mots $\underline{r} = \underline{v} + \underline{e}$ on a $\underline{r}.H^t = (\underline{v} + \underline{e}).H^t = \underline{r}.H^t = \underline{s}$.
- $\underline{s}$ est le syndrome d 'erreur.
- La table de configuration d 'erreurs permet de déduire $\underline{e}$ de $\underline{s}$