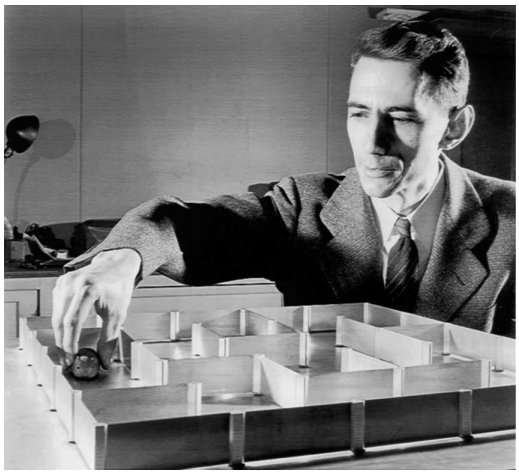
Chapitre 1 :

Théorie de l 'information

Claude Elwood Shannon (1916-2001)



Vue d 'ensemble de la théorie de l 'information

A côté de son usage devenu banal, le mot "information" a un contenu scientifique précis mais restrictif. La théorie de l'information élaborée et énoncée par l'ingénieur américain Claude Elwood Shannon en 1948, se présente comme un chapitre plutôt austère de la théorie des probabilités. Elle résume, en une magistrale synthèse, l'expérience théorique acquise avant et surtout pendant la Seconde Guerre mondiale sur les moyens de communication, en même temps qu'elle suggère des possibilités entièrement nouvelles. Elle affirme la possibilité paradoxale d'une communication sans erreur malgré des bruits perturbateurs affectant la transmission. pourvu qu'un codage approprié soit employé.

Vue d 'ensemble de la théorie de l 'information

Utile, indispensable même aux ingénieurs en tant que cadre conceptuel, elle n'a eu initialement qu'une faible influence directe sur les moyens de communication. Elle a pris de plus en plus d'importance à mesure qu'il devenait possible de réaliser des dispositifs complexes. Par une coïncidence qui fait rêver, 1948 est aussi l'année de l'invention du transistor. Le prodigieux développement de la technologie des semi-conducteurs a peu à peu fait entrer la théorie de l'information dans la pratique, et c'est peu dire qu'elle a fait désormais la preuve expérimentale de son utilité. La radiotéléphonie numérique et les CD seraient inconcevables sans les très efficaces procédés de codage qu'a directement suscités la théorie de l'information. Une immense expérience technique s'ajoute donc maintenant à la théorie proprement dite. Sa validité en est confirmée avec éclat et sa compréhension enrichie.

Vue d 'ensemble de la théorie de l 'information

Cette théorie est mal connue du public. L'une de ses caractéristiques fondamentale paraît si étrangère à la perception commune de l'information qu'elle étonne ou rebute et, en tout cas. fait obstacle à son assimilation : l'exclusion de la sémantique. La théorie de l'information est, en effet, indifférente à la signification des message. Au premier abord, la signification paraît l'essence même de l'information, au point que le refus de la sémantique semble la vider de tout contenu. Mais le point de vue de la théorie de l'information est modeste: celui d'un messager dont la fonction se limite au transfert d'un objet -une lettre par exemple- dont il n'a pas à connaître autre chose que le poids et les dimensions extérieures. L'information que peut porter cet objet n'a pas d'incidence sur les moyens de la transporter. Tel est aussi le point de vue de l'ingénieur en communications, seulement concerné par la quantité d'information qu'il doit transmettre, mesurable selon la théorie de Shannon.

G. Battail. Science et Avenir, hors série décembre 1999 janvier 2000, pp. 28-29

Quelle mesure quantitative pour l'information ?

Un constat :

la transmission d'un message certain est inutile

- source d'information : siège d'événements *aléatoires* qui constituent le message
- quantité d'information d'un *message* :

mesure de son *imprévisibilité*

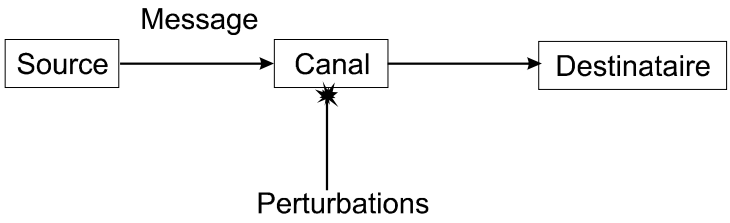
Voici la blague d'aujourd 'hui

- **Claude Shannon, avait l'habitude de faire jouer à ce petit jeu de société quand il était invité quelque part. Il prenait un livre au hasard, l'ouvrait au hasard, commençait à lire un paragraphe et s'arrêtait. Il demandait ensuite à l'assistance de deviner une à une les lettres suivantes. L'assistance se débrouillait bien et trouvait la lettre dans environ 75 % des cas. Shannon en déduisait que la langue anglaise possède un taux de redondance de 75 %.**
- **Quand nous manipulons du texte, les caractères que nous utilisons n'ont pas la même probabilité d'apparition. De plus il a une structure interne forte (la grammaire). Quand le mot arbre est au pluriel on peut aisément prédire la lettre qui suit le « e » final.**
- **Quand nous travaillons avec de la musique, la distribution des probabilités d'apparition des sons n'est pas uniforme non plus.**
- **Quand nous manipulons des images, elles possèdent également des régularités, elles ne sont pas « aléatoires ».**
- **C'est cette caractéristiques qui incite à compresser les données et c'est elle qui permet, souvent, de réussir.**

Théorie de l'information (Shannon 1948)

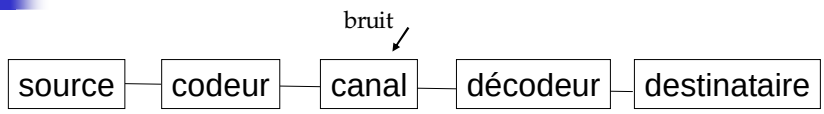
- Représentation efficace de l'information: le codage de source sans pertes (compaction de l'information)
- Théorie de la distorsion et codage de source avec pertes (compression de l'information)
- Capacité d'un canal de télécommunications et méthodes de codage de canal (transmission fiable de l'information à l'aide de codes correcteurs d'erreurs)
- Chiffrement et stratégies de cryptanalyse (confidentialité de l'information, authentification des utilisateurs, décryptement)

Modèle d 'un système de communication



Source = je parle
Canal = l'air ambiant
Perturbations = bruit sonore
Destinataire = tu écoutes

Modèle d 'un système de communication



- Th. Signaux
  - décrit messages et perturbations
- Modulation
  - modifie les signaux pour les propager
- Electronique
  - réalise les fonctions
- Th. Information
  - propose une mesure quantitative de l'**information** et étudie sa représentation, sa transmission, sa dégradation

Modèle d 'un système de communication

- **Source** : siège d'évènements aléatoires qui constituent le message émis → **Entropie**
- **Canal** : transmet et dégrade le message → **Capacité**

Des messages différents portent la même information, le **codage** cherche le message avec les meilleurs propriétés.

- Codage de source → supprime la redondance, reduit le coût
- Codage de canal → protège contre les perturbations
- Cryptage → protège contre les curieux

Information, grandeur mesurable ?

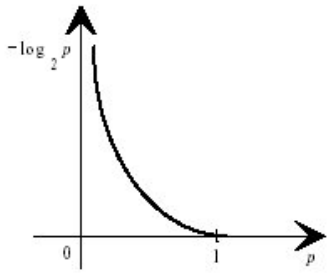
- **Aspects qualitatifs et quantitatifs de l 'information**
  -> apporter des valeurs chiffrées, mesurer
- Information liée à la nature aléatoire d 'un message
- Information + grandeur mesurable = probabilités
- communication = expérience aléatoire
- message = résultat de l expérience qui apporte l 'information
- Exemple : montant sur bulletin de paye

Information, grandeur mesurable ?

- Soit I(x), la quantité d'information apportée par le message x
- I(x) est une fonction f de $\frac{1}{p_x}$ Avec f croissante & f(1)=0
- I(x) doit être positive
- I(x) doit être additive : I(x+y) = I(x) + I(y)

Information, grandeur mesurable ?

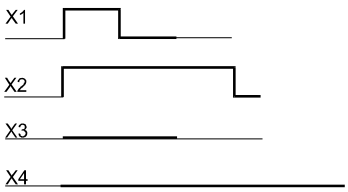
- Selon Shannon, $I(x) = \log(1/p(x)) = -\log(p(x))$
- Si log base 2, alors I(x) s 'exprime en bit
- $I(x_k)$ est aussi appelé Self-information de la source



Sources discrètes

- **Source discrète d'information** : suite de variables aléatoires discrètes X_1, X_2, … X_n
- **Symbole** ou **lettre** : élément fondamental irréductible contenant une information, cad réalisation particulière de la source d'information.
- **Mot** : succession finie de symboles
- **Alphabet** : totalité des D lettres
  [X] = [x_1,x_2, …., x_n]

Ex : Code morse, 4 symboles



Sources discrètes

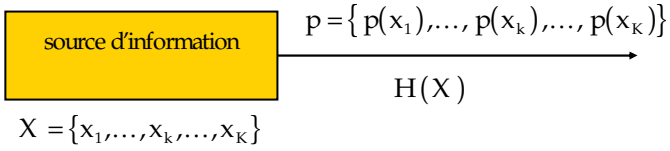
- **Source discrète sans mémoire** : source pour laquelle la probabilité d'apparition d'un symbole ne dépend pas des symboles précédents $p(x_{i_n} / x_{i_{n-1}}, x_{i_{n-2}}, ...) = p(x_{i_n})$
- **Source discrète à mémoire** : source pour laquelle la probabilité d'apparition d'un symbole dépend du ou des symboles précédents
- **Source stationnaire** : source pour laquelle les probabilités d'apparition des différents symboles ne dépendent pas de l'origine des temps $p(x_{i_n}) = p(x_{i_{n+k}}) \quad \forall k$
- **Source à débit contrôlable** : source pouvant générer des messages comme suite à une commande externe (Télégraphe, .)

Sources discrètes

- **Source à débit non contrôlable** : source générant des messages avec un débit fixé, propriété de la source (CD audio)
- **Source discrète à contraintes fixes** : source pour laquelle certains symboles ne peuvent être utilisés qu'en des conditions déterminées (Morse, …)
- **Source discrète à contraintes probabilistes** : source à mémoire. Dans un état, la source peut générer n'importe lequel des symboles avec une probabilité qui dépend des symboles précédents (texte …)
- **Source de Markov** : source pour laquelle la probabilité de générer un symbole ne dépend que du symbole à l'instant n-1

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

Entropie d'une source d'information



Hyp : source discrète finie stationnaire sans mémoire

Emission = variable aléatoire X

$p_i = p(X = x_i)$ pour $i = 1,2,...,n$

$\sum_{i=1}^{n} p_i = 1$

Entropie d ’une source d ’information

Quantité d'information moyenne associée à chaque symbole de la source = **entropie**

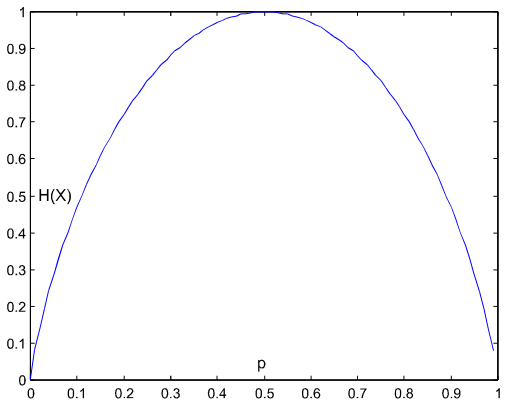
$$H(X) = E(I(X)) = \sum_{i=1}^{N} p_i.\log(1/p_i) = -\sum_{i=1}^{N} p_i.\log(p_i)$$

Entropie d ’une source binaire

$$H(X) = \begin{cases} -p.\log(p) - (1-p).\log(1-p) & \text{pour } 0 < p < 1 \\ 0 \quad \text{si} \quad p = 0 \text{ ou } 1 \end{cases}$$

$p(1) = p$

$p(0) = 1 - p$



Propriétés de l ’entropie

- **Additivité** : de part la définition de l'information propre.
- **Positive** : $H(X) = H(p_1, p_2, ..., p_n) \geq 0$
- **Bornée** : $H(X) \leq H(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}) = \log(n)$
- **Continuité** : l'entropie est une fonction continue de chaque variable p_i.

• **Redondance :**

$$R = H_{\max}(X) - H(X) \qquad \rho = 1 - \frac{H(X)}{H_{\max}(X)}$$

Entropie & Débit d ’information

- Le débit d'information d'**une source** est donné par le produit son entropie (valeur moyenne de l'info /symbole) par le nombre moyen de symboles par seconde, ce qui équivaut à :

$$D_X = \frac{H(X)}{\tau} \quad (bits.s^{-1}) \quad \text{avec } \tau \text{ durée moyenne d'un symbole}$$

• **Source Qaire :**

- **Source Q^{aire}** : source S dont l'alphabet possède Q éléments
- **k^{ième} extension** : source S^k dont l'alphabet est obtenu en groupant par bloc de k celui de la source S (ordre k)

Information mutuelle

$$I(x_k;y_k) = \log(p(x_k/y_k)/p(x_k))$$

Propriétés :

$I(x;y) = I(y;x)$

$I(x,y) = I(x/y) - I(x)$

$I(x/y) = I(x)$ si x et y indépendants

Règle de Bayes : $p(x,y) = p(x/y).p(y) = p(y/x).p(x) = p(y,x)$

Transinformation & entropies

- Quantité moyenne d 'information transmise par le canal :

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

$$I(X;Y) = \sum_{k=1}^{K}\sum_{j=1}^{J} p(x_k, y_j)\log\frac{p(y_j|x_k)}{p(y_j)} = \sum_{k=1}^{K}\sum_{j=1}^{J} p(x_k, y_j)\log\frac{p(x_k|y_j)}{p(x_k)}$$

- Entropie réunie ou conjointe

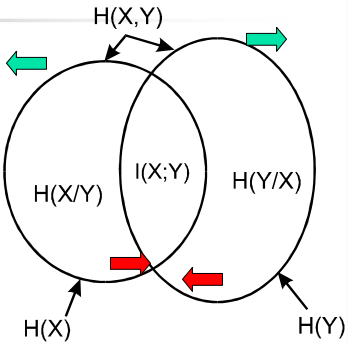
$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_j).\log(p(x_i,y_j))$$

- Entropie conditionnelle ou équivoque

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

Transinformation

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$
$$I(X;Y) = H(X) - H(X/Y) = H(Y) - H(Y/X)$$
$$0 \le I(X;Y) \le H(X)$$



- Canaux non perturbés ⬅
  - $H(X/Y) = H(Y/X) = 0$
  - $H(X,Y) = H(X) = H(Y)$
  - $I(X;Y) = H(X)$
- Canaux très perturbés ➡
  - $H(X/Y) = H(X)$ et $H(Y/X) = H(Y)$
  - $H(X,Y) = H(X) + H(Y)$
  - $I(X,Y) = 0$

Notion de Capacité d'un canal

Nous avons vu que :

- H(X) caractérise la source
- I(X;Y) dépend de la source à p(x)
- I(X;Y) dépend du canal à p(x/y) = P

- Cas extrêmes :
  - I(X;Y) = H(X) à canal non bruité
  - I(X;Y) = 0 à canal bruité

- I(X;Y) varie entre 0 ≤ I(X;Y) ≤ H(X) à on définit C

Capacité d'un canal

Capacité: quantité maximum d'information que l'on peut transmettre dans un canal de télécommunications avec une probabilité d'erreur arbitrairement faible

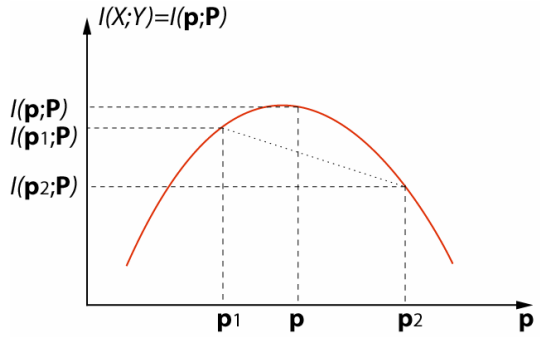
Autres définitions :

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.

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

Capacité d'un canal

$$C = \max_{p} I(X;Y)$$



Capacité d'un canal

- Extensions d 'ordre n de la source :
  - On ajoute un buffer a la source qui attend d 'avoir reçu n symboles avant de transmettre
  - En sortie : M^n messages u possibles
  - Récepteur recoit les messages v

$$I(X^n,Y^n) = \sum_{X^n}\sum_{Y^n} p(u,v).\log_2(p(v/u)/p(v))$$

Capacité d'un canal

- Si les Symboles statistiquement indépendants
- Quantité moyenne fournie par un symbole

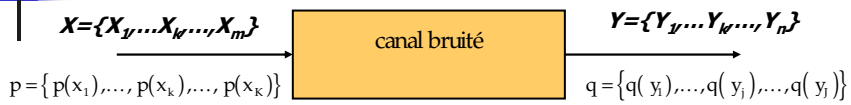
$$I(X;Y) = \frac{I(X^n,Y^n)}{n}$$

- $C = \max_{p(u),n} \frac{I(X^n,Y^n)}{n}$

Type de Canaux discrets

- **Canal** : milieu de transmission de l'information situé entre la source et la destination. Le canal opère une transformation entre l'espace des symboles à l'entrée et celui de la sortie.
- **Canal discret** : les espaces d'entrée et de sortie sont discrets
- **Canal continu** : les espaces d'entrée et de sortie sont continus
- **Canal sans mémoire** : si la transformation d'un symbole x à l'entrée en un symbole y en sortie ne dépend pas des transformations antérieures
- **Canal stationnaire** : si les transformations ne dépendent pas de l'origine des temps

Canaux discrets



$$P=\{p(y_1|x_1),\ldots,p(y_j|x_k),\ldots,p(y_J|x_K)\}$$

- **Matrice stochastique du canal :**

$$[P]=\begin{bmatrix} p(y_1/x_1) & p(y_1/x_2) & \ldots & p(y_1/x_m) \\ p(y_2/x_1) & p(y_2/x_2) & & p(y_2/x_m) \\ \ldots & & & \ldots \\ p(y_n/x_1) & p(y_n/x_2) & \ldots & p(y_n/x_m) \end{bmatrix}$$

Capacité d'un canal

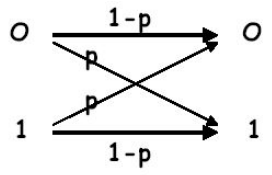
- Canal uniforme en entrée / Canal uniforme en sortie
- Canal uniforme en entrée et sortie :

$$C = \log_2 N + \sum_{j=1}^{n} p_j . \log_2 p_j$$

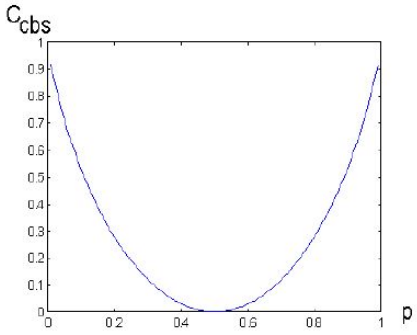
Capacité d'un canal

Exemple de Modélisation d'un canal :

Canal binaire symétrique (Canal stationnaire *sans* mémoire)



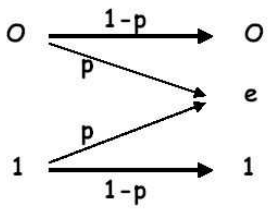
$C_{cbs} = 1 + (1-p) \log_2(1-p) + p \log_2(p)$



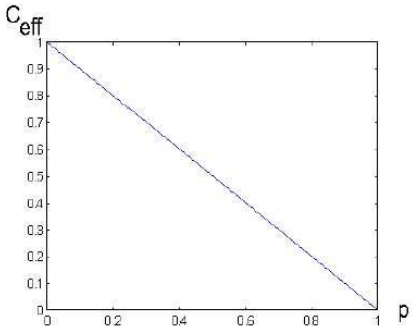
Capacité d'un canal

Exemple de Modélisation d'un canal :

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



$C_{eff} = 1 - p$



D 'autres grandeurs

- **Efficacité d'un canal :** $\eta_c = \frac{I(X;Y)}{C}$
- **Taux d 'information (débit) :** $R_T = \frac{H(X)}{T_s}$
- **Capacité par unité de temps :** $C_T = \frac{C}{T_s}$
- **p_e : Probabilité moyenne d 'erreur :**

$$(R_t - C_T).T_s \leq H(p_e) + p_e.\log_2(M-1)$$

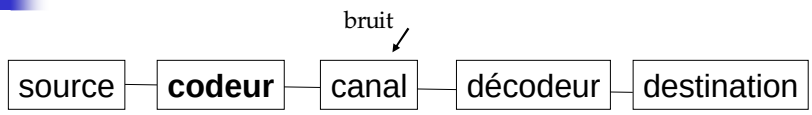
- **Entropie d 'erreur :**

$$H(p_e) = -p_e.\log p_e - (1-p_e).\log(1-p_e)$$

- **Théorème du codage source :**

Lorsque $R_t < C_t$ en utilisant une procédure adéquat de codage et décodage, on peut récupérer le message émis avec une probabilité d 'erreur relativement faible

Modèle d 'un système de communication



Alphabet du canal Z composé de D symboles

D < M (de la source) => codeur + décodeur

Message X_k devient Z_k de longueur n_k

$$n_m = \sum_{k=1}^{M} n_k . p(x_k)$$

On veut un codage pour lequel n_m est minimal

Modèle d 'un système de communication

Entropie max du codeur : $H(C)_{max} = \log D$

Entropie du codeur par symbole $H(C) = H(X) / n_m$

On peut définir l'efficacité du codeur : $e = H(C) / H(C)_{max}$

$e = H(X) / (n_m . \log D)$

e est maximum quand n_m est minimum

Modèle d 'un système de communication

Si n_k fixe alors il faut $D^{n_m} \geq M$ d 'ou $n_m \geq \log M / \log D$

Si symboles équiprobables on a H(X)= log M d 'ou $n_m \geq H(X) / \log D$

Théorème de Shannon :

n_m est borné et on peut toujours trouvé un codage optimal en essayant d'avoir :

$$n_m = H(X) / \log D$$

Pour des codes bien choisi, on peut obtenir $\lim_{N \to \infty} n_m = H(X) / \log D$

Et pour un signal continu ?

- On peut faire des raisonnements similaires mais c 'est beaucoup **moins simple !**

$$H(X) = \int_{-\infty}^{+\infty} f(x).\log(f(x))dx$$

$$I(X;Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y).\log(f(x,y)/(f_1((x).f_2(y))dx$$

$$\boxed{C = \max_{p} I(X;Y)}$$

  - Voir chapitre : Supports de transmission...

Cours de

Transmission de l 'information

Jean-Yves Ramel

Département Informatique de l 'EPU de TOURS
Laboratoire d 'Informatique – RFAI

http://www.rfai.li.univ-tours.fr/ramel/trans_info.html

Exemple

- Une source émet 8 lettres avec :
  p(a)=p(b)=1/4 p(c)=p(d)=1/8 p(e)=p(f)= p(e)=p(f)= 1/16
- Sur un canal binaire à nécessité d'un codeur
- 1ere solution de codage :
  - a à 000 b à 001 c à 010 d à 011
  - e à 100 f à 101 g à 110 h à 111
  - D'ou $n_{moy} = 3$ et $e_1 = H_X/3$
- 2e solution de codage :
  - a à 00 b à 01 c à 100 d à 101
  - e à 1100 f à 1101 g à 1110 h à 1111
  - D'ou $n_{moy} = 2,75$ et $e_2 = H_X/ 2,75$ à 2e solution mieux que 1ere
- Si destinataire reçoit 1100001001011111 à Pas d'ambiguïté car aucun code n'est le préfixe d'un autre à e a c d h

Et pour un signal continu ?

- **On se ramène la plupart du temps à :**
  - un signal limité dans le temps T, à une bande passante W
  - un canal soumis à un bruit blanc additif
  - Les répartitions en puissance du signal et du bruit suivent des distributions de probabilité gaussienne
  - Ce n 'est pas le cas dans la réalité mais cela fixe une référence dont on essaie de se rapprocher.