Introduction

Le but de ce cours est d'étudier la transmission de données (le plus souvent sous la forme de bit) d'une source à un utilisateur récepteur par le biais d'un canal bruité (voir fig 1.1). Il faut savoir qu'aucun dispositif de transmission (ligne de transmission, ondes, ...) n'est parfait et qu'il peut donc inverser certains bits transmis et fausser l'information interceptée par l'utilisateur.
La théorie de l'information est une théorie mathématique fortement statistique, qui a été créée par Claude Shannon.

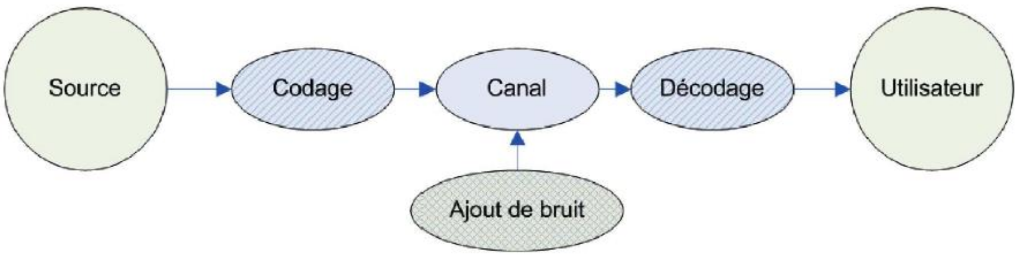

Fig 1.1 Cheminement de l'information.

Il existe deux types de communication :
Analogique : communique à travers une forme d'onde continue (ex: voltage d'un micro), via modulation d'amplitude (AM) ou modulation de fréquence (FM),...
-Utilise l'électronique analogique.
-La fidélité à la forme d'onde
Numérique : communique à travers un message formé par des symboles d'une source d'alphabet discrète généralement codé sous forme d'autres séquences de symboles adaptés au canal de transmission (ex: 0 et 1)
-Utilise la communication analogique pour traverser le canal
-Fidélité au message
-Approprié quant à la minimisation de l'énergie, transmission du BIG DATA, Débruitage, immunité contre les bruits, …
Cette dernière est celle qui nous intéresse dans ce cours.
Les sources transmettant de l'information constituent une suite d'évènements aléatoires, c'est-à-dire une suite d'expériences dont on ne connaît à priori pas le résultat.

1. Notion de l'information :

Il faut moins de bits pour écrire chien que mammifère. Pourtant l'indication Max est un chien contient bien plus d'information que l'indication Max est un mammifère : le contenu d'information d'un message dépend du contexte. En fait, c'est le couple message + contexte qui constitue le véritable porteur d'information.
Un évènement avec peu de probabilité représente beaucoup d'information.
Exemple : Il neige en janvier contient beaucoup moins d'information que Il neige en aout.
On voudrait quantifier l'« **information** » que contient chaque message émis. Par exemple :
-il est clair que si l'émetteur dit toujours **la même chose**, la quantité d'information apportée par une répétition supplémentaire est **nulle**.
- Lorsque la source transmet soit un ***oui*** soit un ***non***, on considère que le récepteur reçoit une unité d'information (**un *bit***).

Autrement dit : une unité d'information, c'est quand on a apriori un ensemble de deux possibilités équiprobables, et que l'une d'elles se réalise.

1.1 Incertitude

En théorie, l'information **diminue l'incertitude**. Plus un récepteur reçoit de l'information, plus son incertitude envers le message à transmettre diminue.
Avant le début de la transmission, le récepteur n'a aucune information sur le message : l'incertitude est maximale. A la fin de la transmission, son incertitude devient nulle.
Pendant la transmission, l'incertitude est dans un état intermédiaire équivalent (quantitativement) à la quantité d'information non encore reçue.
Exemple : Considérons par exemple une source qui peut produire trois symboles a, b et c. Quand le destinataire attend un symbole, il est dans l'incertitude quant au symbole que la source va engendrer. Lorsque le symbole apparait et qu'il arrive au destinataire, cette incertitude diminue.

Le but de la théorie de l'information est de mesurer cette incertitude avant réception.

Exemple On recherche une lettre dans une boite. Si on précise que la lettre se trouve dans une enveloppe rouge, on fournit une information qui diminuera le temps de recherche du fait que le nombre de lettres dans des enveloppes rouges est plus restreint.
Si on ajoute l'information que la lettre dans une grande enveloppe, on pourra abréger d'autant plus le temps de la recherche.

1.2 Quantité d'information

C'est une quantité positive ou nulle. Elle caractérise la diminution de l'incertitude apportée par la réalisation d'un événement N. Elle est définie par : $\mathbf{\log_2 (N/n)}$ exprimé en **bit** (ou logon, unité introduite par Shannon). Et cette éventualité a une probabilité **(n/N).**
N est le nombre d'évènements possibles.
n est le cardinal du sous-ensemble dénoté par l'information qui est exprimé en logon.
Observation La quantité d'information est une fonction croissante.
Exemple Dans une boite il y a N = 1050 lettres dont :
n1 = 500 lettres en enveloppes rouges
n2 = 250 en grandes enveloppes
n3 = 40 en grandes enveloppes rouges
L'information :
-La lettres est dans une enveloppe dans la boite vaut : $\log_2 (N/N) = \log_2(1) = 0$.
-La lettres est dans une enveloppe rouge vaut : $\log_2 (N/n1) = \log_2 (1050/500) = 1, 07$.
-La lettres est dans une grande enveloppe vaut : $\log_2 (N/n2)= \log_2 (1050/250) = 2, 07$.
-La lettres est dans une grande enveloppe bleue vaut $\log_2 (N/n3)= \log_2 (1050/40) = 3, 64$.

1.3 Entropie d'information

Mesurer la quantité d'information que contient un message a toujours été est un problème difficile. L'information n'est pas une entité physique mais un concept abstrait et difficile à quantifier surtout quant un facteur humain est inclus.
Pour surmonter ce problème, Shannon a donné l'idée de définir le contenu en information d'un événement comme une fonction *h(x)* qui dépend seulement de sa probabilité. Donc, soit deux

v.a.: X avec une distribution $pX(x)$ (ou seulement $p(x)$) et Y avec une distribution $pY(y)$. La v.a. X prend ces valeurs dans un alphabet **X** et Y prend ces valeurs dans **Y**.

-cette fonction $h(x)$ doit être décroissante. Parce que le plus un événement est probable moins il apporte de l'information.
- un événement sûr ne porte aucune information. Si $p(x)=1$, $h(x)=0$.
- pour deux événements indépendants, l'information totale est la somme de l'information de chaque événement.

$$h(x) = \log \frac{1}{p(x)} = -\log(p(x))$$

Donc pour deux v.a. indépendants x et y, l'information totale obtenue est :

$$h(x;y) = \log \frac{1}{p(x;y)} = \log \frac{1}{p(x)p(y)} = \log \frac{1}{p(x)} + \log \frac{1}{p(y)}$$

Donc elle satisfit la relation : $h(x;y) = h(x) + h(y)$.
L'entropie permet de mesurer la quantité d'information moyenne d'un ensemble d'évènements (en particulier de messages) et de mesurer son incertitude. On la note H.
L'entropie $H(X)$ d'une variable aléatoire discrète X est définie par :

$$H(X) = -\sum_{x\in\chi} p(x)\log(p(x))$$

Avec $p(x) = n / N$ la probabilité associée à l'évènement.
L'entropie de Shannon, est une fonction mathématique qui correspond à la quantité d'information moyenne contenue ou délivrée par une source d'information.
Remarque : L'entropie ne dépend pas des valeurs de la v.a. X mais seulement des probabilités.
Cette entropie peut être interprétée de différentes façons. Elle peut être vue comme :
- L'incertitude sur X.
- L'information nécessaire pour décrire X, c'est-à-dire pour lever l'incertitude.
- Le nombre de questions binaires (oui/non) nécessaires pour caractériser X.
- La longueur moyenne de la description la plus courte de X.

Exemple :
Une v.a. X qui prend ces valeurs dans {0, 1}. X=1 avec probabilité p et X=0 avec probabilité 1-p.
On calcule l'entropie de X. En appliquant la définition, on obtient :

$$H(X) = -\sum_{x\in\chi} p(x)\log(p(x)) = -p\log(p) - (1-p)\log((1-p)) = H(p)$$

$H(p)$ est appelé fonction d'entropie binaire. Sa variation en fonction de la probabilité p est donnée par :

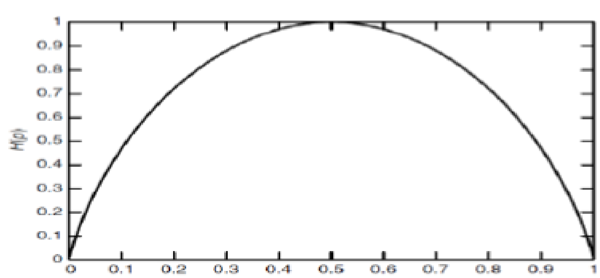

fig 1.2 fonction d'entropie binaire

Cette fonction varie en fonction de p. En particulier, si p vaut 1 ou 0, $H(p) = 0$. La fonction est symétrique par rapport à $p = 0.5$, et maximale =1 en cette même valeur.

Exemple : on considère une source qui produits 32 symboles avec une distribution de probabilités uniforme. L'entropie est donnée par:

$$H(X) = -\sum_{i=1}^{32} p(i)\log(p(i)) = -\sum_{i=1}^{32} \frac{1}{32}\log\left(\frac{1}{32}\right) = \log(32) = 5\,bits.$$

1.3.1 Propriétés

1- $H(X) \geq 0$. Démonstration :

$$0 \leq p(x) \leq 1 \Rightarrow \log\left(\frac{1}{p(x)}\right) \geq 0$$

2- L'entropie est maximale lorsque toutes les valeurs ont la même probabilité.

3- Changement de base :

$$H_b(X) = (\log_b(a))H_a(X)$$

Démonstration: $\log_b(p) = \log_b(a)\log_a(p)$

1.3.2 Entropie jointe et conditionnelle :

-Pour une paire (X,Y) distribuées comme $p(x,y)$, on définit l'entropie jointe comme étant :

$$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log_2 p(x,y) = H(Y,X)$$

- On définit l'entropie conditionnelle d'une v.a. Y connaissant une autre v.a. X comme la valeur moyenne des entropies des distributions conditionnelles, moyennées sur la v.a. de condition.

$$\begin{aligned} H(Y/X) &= \sum_{x\in\mathcal{X}} p(x)H(Y/X=x) \\ &= -\sum_{x\in\mathcal{X}} p(x)\sum_{y\in\mathcal{Y}} p(y/x)\log(p(y/x)) \\ &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log(p(y/x)) \\ &= -E\log(p(Y/X)) \end{aligned}$$

Cette grandeur est aussi appelée équivocation de Y par rapport à X. Elle mesure en effet l'incertitude résiduelle ou le caractère ambigu ou équivoque de Y, lorsqu'on connaît X.

1.3.3 Information mutuelle :

On l'interprète comme étant l'information partagée par X et Y, ou bien la réduction de l'incertitude sur X (ou Y) amenée par la connaissance de Y (ou X).

$$H(X:Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log_2\frac{p(x)p(y)}{p(x,y)}$$

_ L'entropie est séparable en 2 : $H(X) = H(X/Y) + H(X:Y)$ (de même pour H(Y))

_ L'entropie mutuelle est positive : $I(X:Y) \geq 0$

_ L'entropie est la "self-information" de X. En effet, $I(X:X) = H(X)$ ¡ $H(X/X) = H(X)$, puisque $H(X/X)$ est nulle par définition.