1 Théorie de l'information

Shannon s'est posé la question, en 1948, dans le cadre d'un ouvrage sur les communications s'est posé plusieurs questions :

- Peut-on définir la quantité d'information contenue dans un message ?
- Comment transmettre un message sans erreur à travers un canal bruité ?

Ces questions l'ont amené à introduire les notions d'entropie, d'information mutuelle, de capacité de canal et de codes correcteurs d'erreur, que l'on va présenter maintenant.

Définition 1.1 (Entropie)**.** L'entropie H_X d'une variable aléatoire discrète X prenant des valeurs x et ayant comme probabilité de distribution $p(x)$ est définie par :

$$H_X = -\sum_x \mathbb{P}(X = x) \log_2 \mathbb{P}(X = x)$$

Cette valeur est définie en bits.

Par exemple, si X prend M valeurs distinctes avec une probabilité $p(x) = \frac{1}{M}$ pour toute valeur x, alors $H_X = \log_2 M$. Ou encore, si $X \in \{0, 1\}$ et $p(0) = p$ et $p(1) = 1 - p$, alors $H_X = -p \log_2 p - (1 - p) \log_2(1 - p)$.

Définition 1.2 (Information mutuelle)**.** Etant données deux variables aléatoires discrètes X, Y, on définit $I_{X,Y}$ l'entropie mutuelle de X et Y :

$$I_{X,Y} = \sum_{x,y} p(X = x \cap Y = y) \log_2 \frac{\mathbb{P}(X = x \cap Y = y)}{\mathbb{P}(X = x)\mathbb{P}(Y = y)}$$

Si les variables sont indépendantes, on a alors $I_{X,Y} = 0$. L'entropie mutuelle est toujours positive [1] : $I_{X,Y} \geq 0$.

On souhaite transmettre un message sur un canal bruité, afin d'augmenter les chances de le recevoir intègre de l'autre côté, on code le message de manière à introduire de la redondance. L'idée étant que plus il y a de redondance, plus le message a de chances d'être intègre à l'arrivée.

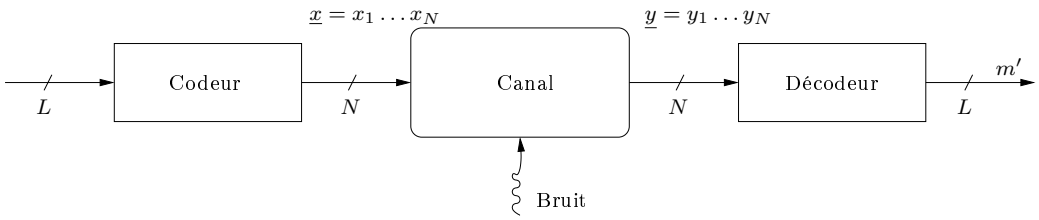

FIGURE 1 – Transmission d'un message sur un canal bruité.

Le canal sera décrit par une probabilité de transition $Q(\underline{y}|\underline{x}) = \Pi_i Q(y_i|x_i)$, qui est la probabilité que le message à l'arrivée soit $\underline{y}$ quand $\underline{x}$ était en entrée. On va donner des exemples de canaux :

Définition 1.3 (Binary Symetric Channel)**.** L'entrée x_i et la sortie y_i sont toutes deux dans $\{0, 1\}$ le canal est caractérisé par la probabilité p qu'un bit soit transmis comme étant son opposé.

Définition 1.4 (Binary Eraser Channel)**.** Dans ce cas, certains des bits de l'entrée sont effacés au lieu d'être corrompus. x_i est toujours dans $\{0, 1\}$, mais y_i est dans $\{0, 1, E\}$, où E veut dire que le bit a été effacé. Dans le cas où le canal est symétrique, il peut être caractérisé par la probabilité p qu'un bit soit effacé.

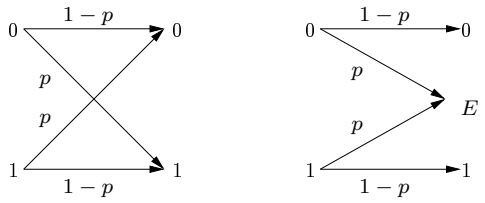

FIGURE 2 – Représentation des canaux BSC et BEC

Une des caractéristiques importantes d'un canal est sa *capacité*, que l'on définit ainsi :

Définition 1.5 (Capacité d'un canal)**.** La capacité C d'un canal est définie comme l'entropie mutuelle entre X le message envoyé et Y le message reçu :

$$C = \max_{\mathbb{P}(x)} I_{X,Y} = \max_{\mathbb{P}(x)} \sum_{x,y} \mathbb{P}(X = x \cap Y = y) \log_2 \frac{\mathbb{P}(X = x \cap Y = y)}{\mathbb{P}(X = x)\mathbb{P}(Y = y)}$$

Par exemple, les capacités des canaux BSC et BEC sont $C_{\text{BSC}} = 1 - H(p)$ et $C_{\text{BEC}} = 1 - p$.

Nous avons maintenant tout défini sauf la manière dont on code les messages, un *code* pour un mot m est un mot $\underline{x}(m)$, l'ensemble des mots de code est appelé le *dictionnaire*. Un dictionnaire de taille 2^L peut en fait être vu comme un ensemble de points de l'hypercube de taille 2^N.

On peut par exemple prendre le code qui consiste à répéter trois fois chaque bit sur un canal BSC, le décodage se faisant par la majorité sur ces trois bits. La probabilité de faire une erreur avec ce code est $\mathbb{P}_{err} = \sum_{r=2}^{3} \binom{3}{r} p^r (1-p)^{3-r} = p^3 + 3p^2(1-p)$ où p est la probabilité de faire une erreur sur un bit.

Comme on l'a dit précédemment, la redondance du code permet d'assurer une plus grande fiabilité de la transmission, on mesure cette redondance par le *taux de redondance* : $R = L/N$, plus le taux est bas, plus il y a de redondance dans le code.

La probabilité d'erreur lorsque l'on a un dictionnaire de taille 2^L est la suivante :

$$\mathbb{P}_{err} = \frac{1}{2^L} \sum_{m \in \{0,1\}^L} \Big[\sum_{y} Q(\underline{y}|\underline{x}(m)) \delta_{d(\underline{y}),m} \Big]$$

Shannon démontra qu'il est possible d'atteindre une probabilité d'erreur de zéro, tout en conservant un taux fini sur un canal de capacité finie C :

Théorème 1.6 (Shannon)**.** *Pour tout taux $R < C$, il existe une séquence de codes $\{C_n\}$ de taille N, de taux R_N et de probabilité d'erreur $\mathbb{P}_{B,N}$ tels que quand $N \to \infty$, $R_N \to R$ et $\mathbb{P}_{B,N} \to 0$. Inversement, pour une séquence de codes $\{C_N\}$, si on a $R_N \to R$ et $\mathbb{P}_{B,N} \to 0$ quand $N \to \infty$, alors $R < C$.*

Rappellons-nous que qu'un dictionnaire de taille 2^L est un hypercube de taille 2^N, prenons $L = RN$ et $L \to \infty$, $N \to \infty$ à $R = L/N$ fixé. Le principe des codes de Shannon est de prendre comme codes des mots les plus éloignés les uns des autres, le décodage consistant à prendre le mot le plus proche, c'est à dire le mot le plus probable, on appelle ce décodage Maximum A priori Probability (MAP) :

$$\mathbb{P}(\underline{x}|\underline{y}) \cong \Pi_i Q(y_i|x_i) \mathbb{I}(\underline{x} \in \text{Dictionnaire})$$

1. $-\log_2$ est convexe, on conclut par l'inégalité de Jensen.

Pour donner une idée de la géométrie du dictionnaire, on définit $\mathcal{N}\underline{x}(d)$ comme étant le nombre de mots de code à la distance d de $\underline{x}$, $\underline{x}^{(1)} = (0, \ldots, 0)$ et on pose $d = N\delta$, on a alors pour N grand :

$$\begin{aligned} \mathbb{E}(\mathcal{N}\underline{x}^{(1)}(d)) &= (2^L - 1)\frac{1}{2^N}\binom{N}{N\delta} \\ &= 2^{[R-1+H(\delta)]N} \end{aligned}$$

où $H(\delta) = -\delta \log_2 \delta - (1-\delta)\log_2(1-\delta)$. On note δ_{GV} la distance typique entre deux mots du dictionnaire, la courbe suivante décrit $H(\delta)$ et $R - 1 + H(\delta)$:

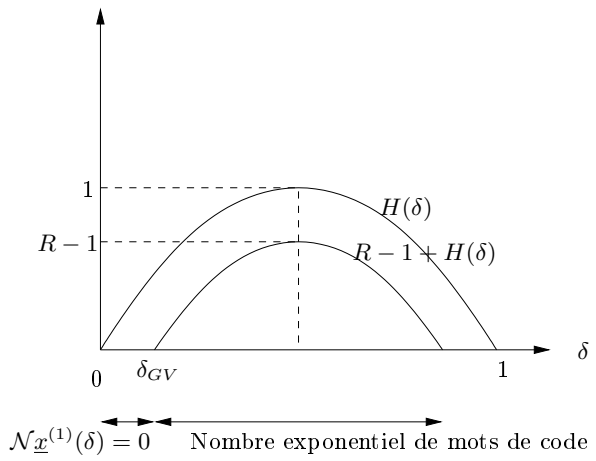

FIGURE 3 – Entropie en fonction de la distance.

Si $p < \delta_{GV}$, alors le décodage est parfait, si $p > \delta_{GV}$ le décodage est impossible. On assiste donc en quelque sorte à une transition de phase.