ACCQ202 Vendredi 16 octobre 2015

COURS n°4

Mathieu Ainsa & Marie de Faverges

Enseignant : Aslan Tchamkerten

1. Transmission d'information

Un canal est défini par la donnée de la probabilité d'obtenir la sortie y connaissant l'entrée x :

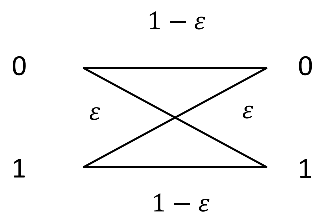
$$\Pr(y|x) = Q(y|x)$$

Par ailleurs :

$$\Pr(y^n|x^n) = \prod_{i=1}^{n} Q(y_i|x_i)$$

Exemple (BSC Canal binaire symétrique)

Ce canal est défini par : $Q(y|x) = \begin{cases} 1-\varepsilon \text{ si } y = x \\ \varepsilon \quad\quad \text{sinon} \end{cases}$ et peut se schématiser comme suit :



Définition (Capacité d'information)

Étant donné un canal $Q(y|x)$, on définit sa capacité d'information : $C = \max_{p(x)} I(p(x).\, Q(y|x))$. Cette définition formelle prendra sens lors de l'énoncé du théorème pour le codage de canal.

Propriétés.

- $C \geq 0$
- $C \leq \min\{\log\ |\mathcal{X}|, \log\ |\mathcal{Y}|\}$

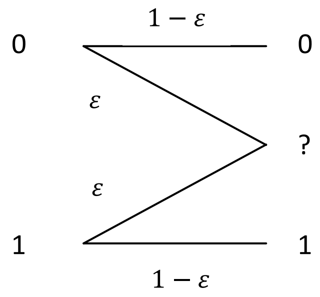
Exemple (Cas du BSC)

$I(X;Y) = H(X) - H(Y|X)$ et, de plus $H(Y|X) = H_b(\varepsilon)$; enfin $H(Y) \leq \log(2) = 1$. Donc $I(X;Y) \leq 1 - H_b(\varepsilon)$ et l'on a égalité dans le cas où X suit une loi de Bernoulli de paramètre $\frac{1}{2}$.

On en conclut alors que : $C = 1 - H_b(\varepsilon)$.

Exemple (Cas du BEC : Canal binaire à effacement)

Ce canal permet de modéliser la perte de paquets sur Internet et se schématise comme suit :



$$E = \begin{cases} 0 \text{ si } Y \neq ? \\ 1 \quad \text{sinon} \end{cases}$$

$$I(X;Y) = H(Y) - H(Y|X)$$

$$= H(Y) - H_b(\varepsilon)$$

$$= H(Y,E) - H_b(\varepsilon)$$

$$= H(E) + H(Y|E) - H_b(\varepsilon)$$

$$= H_b(\varepsilon) + H(Y|E) - H_b(\varepsilon)$$

$$= H(Y|E)$$

$$= H(Y|E = 0).P(E = 0) + H(Y|E = 1).P(E = 1)$$

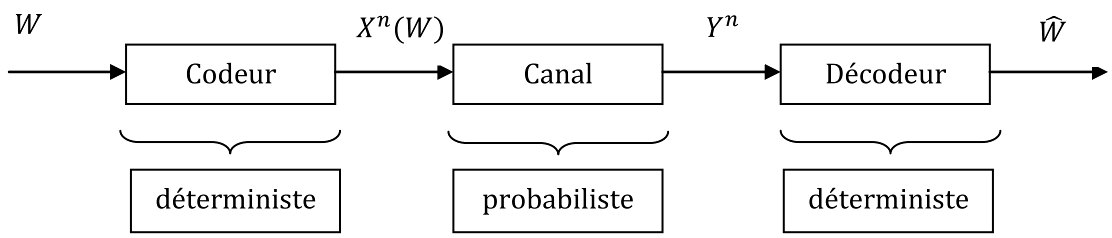
$$= (1 - \varepsilon).H_b(\pi).$$

avec $\pi = P(X = 0)$.

La valeur maximale de $I(X;Y)$ est obtenue pour $H_b(\pi) = 1$. On en déduit : $C = 1 - \varepsilon$.

2. Modélisation d'une chaîne de transmission

Soit $W \in \{1,2,\dots,M\}$ un message. On fait l'hypothèse que W suit une loi uniforme.



Définition ($(M,n) -$ code)

Un $(M,n) -$ code pour le canal $(\mathcal{X}, Q(y|x), \mathcal{Y})$ est la donnée :

- d'un alphabet $\{1,2, \dots, M\}$
- d'une fonction d'encodage X^n: $\{1,2, \dots, M\} \rightarrow \mathcal{X}^n$ renvoyant les mots-code $x^n(1)$, $x^n(2), \dots, x^n(M)$.
- d'une fonction de décodage $g : \mathcal{Y}^n \rightarrow \{1,2, \dots, M\}$ déterministe.

Définitions (Rendement et Probabilité d'erreur)

La performance d'un $(M,n) -$ code est donnée par :

- son rendement : $R = \frac{\log M}{n}$;
- sa probabilité d'erreur : $P(\hat{W} \neq W)$. Comme W suit une loi uniforme, on a :

$$P(\hat{W} \neq W) = \frac{1}{M} \sum_{w=1}^{M} P(\hat{W} \neq W | W = w)$$

Définition (Atteignabilité)

R est atteignable si pour tout $\varepsilon > 0$, il existe $(M = 2^{nR}, n)$ tel que $P(\hat{W} \neq W) < \varepsilon$.
Donc R est atteignable s'il existe une suite de codes à taux constant $\{(M = 2^{nR}, n), n \geq 1\}$ telle que $P_n(\hat{W} \neq W) \rightarrow 0$.

Théorème pour le codage de canal.

- $R < C$ est atteignable
- $R > C$ n'est pas atteignable

Autrement dit, C (la capacité) est la borne ultime de transmission fiable. Pour $R = C$, cela dépend du canal.

La preuve s'appuiera sur les 3 lemmes suivants.

Lemme 1 (Data-processing inequality)

(X, Y, Z) forment une chaîne de Markov, notée $X \rightarrow Y \rightarrow Z$, si $P(x, y, z) = p(x).p(y|x).p(z|y)$. Alors dans ce cas : $I(X; Y) \geq I(X; Z)$.

Preuve.

$I(X; (Y, Z)) = I(X; Y) + I(X; Z|Y)$. Comme $I(X; Z|Y) = \mathbb{E}\left[\log \frac{p(x,z|y)}{p(x|y).p(z|y)}\right] = 0$, il vient : $I(X; (Y, Z)) = I(X; Z) + I(X; Y|Z)$. On en déduit $I(X; Y) \geq I(X; Z)$.

♦

Lemme 2 (Fano).

Soient $(X, Y) \sim p(x, y)$ et $X \rightarrow Y \rightarrow \hat{X} = g(Y)$ t.q. $X \in \mathcal{X}$ et $\hat{X}$ est interprétée comme l'estimée de X sur la base de Y. Alors :

$$H(X, Y) \leq 1 + P\left(\hat{X} \neq X\right). \log\ |\mathcal{X}|$$

Preuve.

On pose : $E = \begin{cases} 1 \text{ si } \hat{X} \neq X \\ 0 \text{ si } \hat{X} = X \end{cases}$, de sorte que $\Pr(E = 1) = \ P(\hat{X} \neq X)$.

On a :

$$H\left(E, X \middle| \hat{X}\right) = H\left(E \middle| \hat{X}\right) + H\left(X \middle| E, \hat{X}\right) = H\left(X \middle| \hat{X}\right) + H(E | X, \hat{X})$$

De plus :

$$H\left(X \middle| E, \hat{X}\right) = H\left(X \middle| E = 0, \hat{X}\right). P(E = 0) + H\left(X \middle| E = 1, \hat{X}\right). P(E = 1) \leq 0 + \log|\mathcal{X}|. \Pr(E = 1)$$

et : $H(E|\hat{X}) \leq 1$.

Donc : $1 + \log|\mathcal{X}|. P(\hat{X} \neq X) \geq \ H\left(X \middle| \hat{X}\right)$. Or : $I(X; Y) \geq I(X; \hat{X})$ (Chaîne de Markov), *i.e.* :

$H(X) - H(X|Y) \geq H(X) - H(X|\hat{X})$, soit : $H\left(X \middle| \hat{X}\right) \geq H(X|Y)$. D'où le résultat.

♦

Lemme 3.

Soient $I(X^n; Y^n)$ l'information entre l'entrée et la sortie du canal et C sa capacité.

Alors : $I(X^n; Y^n) \leq nC$.

Preuve.

$$I(X^n; Y^n) = H(Y^n) - H(Y^n|X^n) = H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y^{i-1}, X^n)$$

$$= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i)$$

$$\leq \sum_{i=1}^{n} (H(Y_i) - \ H(Y_i|X_i))$$

car le canal est sans mémoire. Enfin : $H(Y_i) - \ H(Y_i|X_i) = I(X_i; Y_i) \leq C$, ce qui démontre le lemme.

♦

Référence

COVER & THOMAS, *Elements of Information Theory*, 2nd *edition*, Wiley, 2006.