

Nicolas Sendrier

Majeure d'informatique

Introduction la théorie de l'information

Cours n°6

Capacité d'un canal – Second Théorème de Shannon

Canal discret sans mémoire

Définition Un canal discret est défini par la donnée de

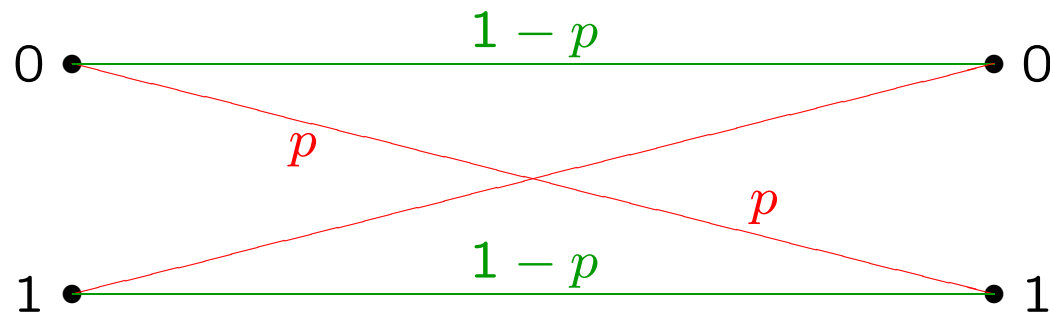
- un **alphabet d'entrée** $X = \{a_1, \dots, a_K\}$
- un **alphabet de sortie** $Y = \{b_1, \dots, b_J\}$
- une **loi de transition** $P_{Y|X}$, i.e. une **matrice stochastique**

$$\Pi = \begin{pmatrix} P(b_1 | a_1) & \dots & P(b_J | a_1) \\ \vdots & \ddots & \vdots \\ P(b_1 | a_K) & \dots & P(b_J | a_K) \end{pmatrix}$$

Nous parlerons du **canal** $\mathcal{T} = (X, Y, \Pi)$.

Le canal est **sans mémoire** si la loi de transition est constante au cours du temps. Nous étudierons principalement les canaux sans mémoire.

Exemple – Canal binaire symétrique

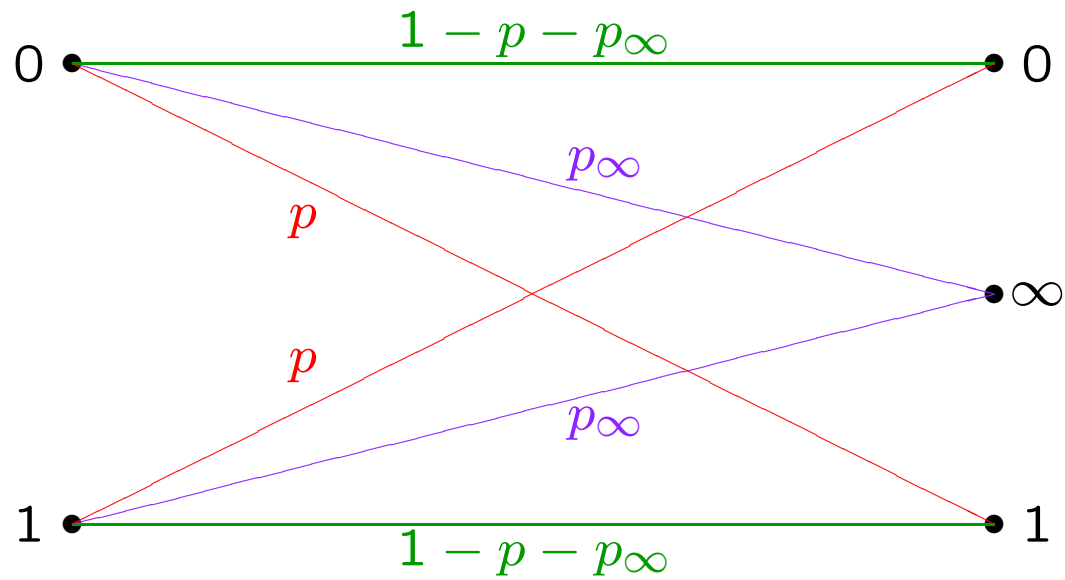


La matrice stochastique est

$$\begin{pmatrix} 1 - p & p \\ p & 1 - p \end{pmatrix}.$$

p est appelé *probabilité de transition* ou *probabilité d'erreur* du canal.

Exemple – Canal binaire symétrique à effacement



$$\Pi = \begin{pmatrix} 1 - p - p_\infty & p_\infty & p \\ p & p_\infty & 1 - p - p_\infty \end{pmatrix}.$$

Capacité

La capacité d'un canal est la quantité maximale d'information pouvant transiter à travers le canal par unité de temps. Autrement dit :

Quelle quantité d'information puis-je obtenir au maximum sur X en observant Y ?

Cette quantité est l'**information mutuelle moyenne** de X et Y , et le **maximum** est pris par rapport à la seule chose susceptible de changer : la **loi d'émission**.

$$C = \max_{x \mapsto P(x)} I(X; Y)$$

On remarquera que $I(X; Y)$ peut s'écrire en fonction des seules lois de transition et d'émission :

$$I(X; Y) = \sum_{x,y} P(y | x) P(x) \log_2 \frac{P(y | x)}{P(y)} \quad \text{et} \quad P(y) = \sum_x P(y | x) P(x).$$

Canaux symétriques

Définition Un canal discret est dit *fortement symétrique* si les lignes et les colonnes de sa matrice stochastique sont égales à une permutation près.

Définition (Décomposition d'un canal) Nous dirons que le canal $\mathcal{T} = (X, Y, \Pi)$ se décompose en une combinaison des canaux $\mathcal{T}_i = (X, Y_i, \Pi_i)_{1 \leq i \leq L}$, si les Y_i sont disjoints et s'il existe des nombres réels positifs $q_1 + q_2 + \dots + q_L = 1$ tels que $\Pi = \left(q_1 \Pi_1 \mid \dots \mid q_L \Pi_L \right)$. Nous noterons formellement

$$\mathcal{T} = \sum_{i=1}^L q_i \mathcal{T}_i$$

Définition Un canal discret est dit *symétrique* s'il se décompose en une combinaison de canaux fortement symétriques.

Capacité d'un canal symétrique (1)

Nous allons utiliser l'identité

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

L'entropie conditionnelle $H(Y | X)$ ne dépend que du canal.

$$\begin{aligned} H(Y | X) &= - \sum_{x,y} P(x, y) \log_2 P(y | x) \\ &= - \sum_x P(x) \sum_y P(y | x) \log_2 P(y | x) \\ &= - \sum_x P(x) H(\Pi) = H(\Pi) \end{aligned}$$

où $H(\Pi) = \sum_y P(y | x) \log_2 P(y | x)$ est indépendant de x dans un canal symétrique. Et donc

$$\begin{aligned} C &= \max(H(Y) - H(Y | X)) \\ &= \max(H(Y)) - H(\Pi) \\ &\leq \log_2 |Y| - H(\Pi). \end{aligned}$$

Capacité d'un canal symétrique (2)

Proposition La capacité d'un canal fortement symétrique est atteinte pour une loi d'émission uniforme et vaut

$$C = \log_2 |Y| - H(\Pi)$$

Proposition Soit $\mathcal{T} = q_1\mathcal{T}_1 + \dots + q_L\mathcal{T}_L$ un canal symétrique. Sa capacité est atteinte lorsque la loi d'émission est uniforme et vaut

$$C = \sum_{i=1}^L q_i C_i$$

où les C_i sont les capacités des canaux fortement symétriques \mathcal{T}_i .

Exemples

Capacité du canal binaire symétrique :

$$C = 1 + H_2(p)$$

Capacité du canal binaire symétrique à effacement :

$$C = (1 - p_\infty) \left(1 + H_2 \left(\frac{p}{1 - p_\infty} \right) \right)$$

$$H_2(x) = -x \log_2(x) - (1 - x) \log_2(1 - x)$$

Codage de canal

Nous considérons un canal discret $\mathcal{T} = (X, Y, \Pi)$

Définition Un *code en bloc* de *longueur* n et de *cardinal* M est M séquences de n lettres de X . Nous parlerons de code (M, n) . Le *taux de transmission* d'un code est égal à

$$R = \frac{\log_{|X|} M}{n} \leq 1$$

Un code va permettre de « coder » une quantité d'information égale à $\log_2 M$ bits. En pratique un code binaire ($|X| = 2$) de longueur n et de cardinal $M = 2^k$ transforme un bloc de k bits d'information en un bloc de n symboles binaires, dans ce cas $R = k/n$.

Un *codeur* est une procédure qui associe à toute séquence binaire finie une séquence finie de lettres de X .

Exemples

Code à répétition de longueur 3

$$C = \{000, 111\}$$

Code de parité de longueur 4

$$C = \{0000, 0011, 0101, 0110, 1001, 1010, 1100, 1111\}$$

Code de Hamming de longueur 7

$$C = \{0000000, 1101000, 0110100, 0011010, \\ 0001101, 1000110, 0100011, 1010001, \\ 1111111, 0010111, 1001011, 1100101, \\ 1110010, 0111001, 1011100, 0101110\}$$

Performance d'un code – Décodage

Soit \mathcal{C} un code en bloc (M, n) utilisé dans un canal discret (X, Y, Π)

Définition Un *algorithme de décodage* de \mathcal{C} est une procédure qui a tout bloc de n lettres de Y associe un mot de code de \mathcal{C} .

L'événement « *mauvais décodage* » pour un algorithme de décodage et un canal donné est défini par :

Un mot de code $\mathbf{x} \in \mathcal{C} \subset X^n$ est transmis à travers le canal, le mot $\mathbf{y} \in Y^n$ est reçu et est décodé en $\tilde{\mathbf{x}} \neq \mathbf{x}$.

Définition Le *taux d'erreur de \mathcal{C}* (dans le canal considéré) noté $P_e(\mathcal{C})$ est le minimum de la probabilité de mauvais décodage pour tous les algorithmes de décodage.

Second théorème de Shannon

Théorème Soit un canal discret sans mémoire de capacité C . Pour tout $R < C$, il existe une suite de codes en bloc $(\mathcal{C}_n(M, n))_{n>0}$ de taux de transmission R_n telle que

$$\lim_{n \rightarrow \infty} R_n = R \quad \text{et} \quad \lim_{n \rightarrow \infty} P_e(\mathcal{C}_n) = 0$$

Théorème (réciproque) Soit un canal discret sans mémoire de capacité C . Tout code \mathcal{C} de taux de transmission $R > C$ vérifie $P_e(\mathcal{C}) > K(C, R)$, où $K(C, R) > 0$ dépend du canal et du taux de transmission mais est indépendant de la longueur du code.

AEP conjointe

Définition Ensemble des séquences typiques conjointes

$$A_\varepsilon^{(n)} = \left\{ (\vec{x}, \vec{y}) \in \mathcal{X}^n \times \mathcal{Y}^n, \left| \frac{1}{n} \log_2 \frac{1}{P(\vec{x})} - H(X) \right| \leq \varepsilon, \right. \\ \left. \left| \frac{1}{n} \log_2 \frac{1}{P(\vec{y})} - H(Y) \right| \leq \varepsilon, \left| \frac{1}{n} \log_2 \frac{1}{P(\vec{x}, \vec{y})} - H(X, Y) \right| \leq \varepsilon \right\}$$

Le processus $X \times Y$ vérifie l'AEP conjointe si

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} \Pr(A_\varepsilon^{(n)}) = 1.$$