Théorie de l'information et codage **2010/2011**

Cours 5 — 15 mars 2011

Enseignant: Marc Lelarge *Scribe: Nicolas Daviaud*

Pour information

- Page web du cours
  `http://www.di.ens.fr/~lelarge/info11.html`

Canaux discrets sans mémoire et leurs fonctions capcité-coût

5.1 Définitions et codes sans erreur

Un canal discret sans mémoire est caractérisé par :
- un alphabet d'entrée $\mathcal{X}$, de cardinal r
- un alphabet de sortie $\mathcal{Y}$, de cardinal s
- un coût $b(x)$ associé à chaque entrée x
- une probabilité de transition $p(y|x)$ qui définit une matrice stochastique (notée Q) de taille $r \times s$:

$$p(x|y) \geq 0, \quad \forall x \in \mathcal{X}, \sum_{y \in \mathcal{Y}} p(y|x) = 1.$$

Ce modèle de canal reçoit donc une entrée x à chaque unité de temps et émet y avec probabilité $p(y|x)$ pour un coût de $b(x)$.

Canal sans mémoire : s'il est utilisé n fois sur les entrées $x_1, \dots x_n$, la probabilité qu'il émette $y_1 \dots y_n$ est

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

et cela coute

$$b(\underline{x}) = \sum_{i=1}^{n} b(x_i).$$

Si les entrées sont des variables aléatoires (v.a.) $\underline{X} = (X_i, \dots, X_n)$ suivant une loi $p(\underline{X})$, alors le coût moyen est

$$\mathbb{E}\left[b(\underline{X})\right] = \sum_{\underline{x}} p(\underline{x}) b(\underline{x}).$$

Définition 5.1.1 *Un (M, n)-code est un sous-ensemble*

$$C = \{\underline{x}_1, \dots \underline{x}_M\} \subset \mathcal{X}^n.$$

La longueur d'un tel code est n.
Le taux du code est $R = \frac{\log_2 M}{n}$ (bits/symbole).
Un code est dit β-admissible si $\forall i, b(\underline{x}_i) \leq n\beta$.
Une règle de décodage pour ce code est une fonction

$$f : \mathcal{Y}^n \longrightarrow C \cup \{?\}$$

La *probabilité d'erreur* sur le mot code $\underline{x}_i$ est

$$\begin{aligned} P_E^{(i)} &= \mathbb{P}\left[f(\underline{y}) \neq \underline{x}_i | \underline{x}_i \text{ envoyé}\right] \\ &= \sum_{f(\underline{y}) \neq \underline{x}_i} p(\underline{y}|\underline{x}_i) \end{aligned}$$

Exemple 5.1.1: Prenons $\mathcal{X} = \left\{0, 1/2, 1\right\}$, $\mathcal{Y} = \{0, 1\}$ et $b(0) = b(1) = 1$, $b\left(1/2\right) = 0$ et les probabilités de transition suivantes :

$$Q = \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \\ 0 & 1 \end{pmatrix}.$$

Soit $k \leq n$. L'ensemble suivant est un $(2^k, n)$-code

$$C = \left\{\left(x_1 \cdots x_k, 1/2 \cdots 1/2\right) | x_i \in \{0, 1\}\right\}.$$

Ce code est β-admissible pour $\beta \geq \frac{k}{n}$ et son taux est $\frac{k}{n}$ bits/symbole.
En prenant la règle de décodage :

$$f(y_1, \dots, y_n) = \left(y_1, \dots, y_k, 1/2 \dots 1/2\right),$$

la probabilité d'erreur pour ce canal est nulle quelque soit le mot code envoyé.

5.1.1 Réciproquie du Théorème de codage de canal pour une probabilité d'erreur nulle

Supposons qu'on ait un $(2^{nR}, n)$-code avec $\forall i, P_E^{(i)} = 0$, et $\underline{X}$ est uniformément distribué. On a alors :

$$\begin{aligned}
nR = logM &= H(\underline{X}) \\
&= H(\underline{X}|\underline{Y}) + I(\underline{X};\underline{Y}) \\
&= 0 + \left(H(\underline{Y}) - \sum_{i=1}^{n} H\left(Y_i|Y_1,\ldots,Y_{i-1},\underline{X}\right)\right) \qquad \text{car } \forall i, P_E^{(i)} = 0, \\
&= H(\underline{Y}) - \sum_{i=1}^{n} H\left(Y_i|X_i\right) \qquad \text{car canal sans mémoire,} \\
&\leq \sum_{i=1}^{n} H\left(Y_i\right) - \sum_{i=1}^{n} H\left(Y_i|X_i\right) \\
&= \sum_{i=1}^{n} I\left(X_i;Y_i\right) \\
&\leq n \cdot max_{p(\underline{X})}\{I(\underline{X};\underline{Y})\} = nC,
\end{aligned}$$

Cette dernière quantité s'appelle la capacité du canal.

5.2 Inégalité de Fano et réciproque du théorème de codage de canal

Le calcul précédent a été fait pour une probabilité d'erreur nulle. Il est facile de le généralisé comme suit.

On Suppose toujours que $\underline{X}$ uniformément distribuée. Donc l'erreur moyenne est définie par : $\overline{P_E} = \frac{1}{2^{nR}} \sum_i P_E^{(i)}$. On a alors,

$$\begin{aligned}
nR &= H(\underline{X}) \\
&= H(\underline{X}|\underline{Y}) + I(\underline{X},\underline{Y}) \\
&\leq 1 + \mathbb{P}\left[f(\underline{Y}) \neq \underline{X}\right] nR + I(\underline{X};\underline{Y}) \qquad \text{par l'inégalité de Fano} \\
&\leq 1 + \overline{P_E} nR + I(\underline{X},\underline{Y}) \\
&\leq 1 + \overline{P_E} nR + nC
\end{aligned}$$

Ainsi, $\overline{P_E} \geq 1 - \frac{C}{R} - \frac{1}{nR}$. Par conséquent, si $R > C$, alors $\overline{P_E} > 0$ pour n suffisament grand et donc pour tout n (puisque dans le cas contraire, il suffirait alors de concatener un code court sans erreur pour obtenir une contradiction).

5.3 La fonction capacité-coût

Pour tout n, on définit la $n^{\text{ème}}$ fonction capacité coût par

$$C_n(\beta) = \max\left\{ I(\underline{X}, \underline{Y}) \,\middle|\, \mathbb{E}\left[b(\underline{X})\right] \leq \beta \text{ et } P(\underline{Y}|\underline{X}) = \prod\prod p(y_i|x_i) \right\}$$

Dans ce contexte, on dira que $\underline{X}$ est une source test et qu'elle est β-admissible si $E[b(\underline{X})] \leq n\beta$.

Remarque 5.3.1 *a) Pour $p(y|x)$ fixées $I(X,;Y)$ est une fonction continue de $p(x)$ donc la maximisation est faite sur un compacte donc atteinte.*
b) $C_n(\beta)$ est définit pour tout $\beta \geq \beta_{\min} = min\{b(x)|x \in \mathcal{X}\}$.
c) C_n est non-décroissante pour $\beta \geq \beta_{\min}$.

La fonction capacité-coût du canal est définie par

$$C_\beta = \sup_n \frac{1}{n} C_n(\beta)$$

Théorème 5.3.1 *C_n est concave en $\beta \geq \beta_{\min}$.*

Démonstration.
Soit $\alpha_1, \alpha_2 \geq 0$ tq $\alpha_1 + \alpha_2 = 1$.
On doit montrer que pour $\beta_1, \beta_2 \geq \beta_{\min}$, on a :

$$C_n(\alpha_1\beta_1 + \alpha_2\beta_2) \geq \alpha_1 C_n(\beta_1) + \alpha_2 C_n(\beta_2).$$

Soit $\underline{X}_1$ et $\underline{X}_2$ deux sources test de distribution p_1 et p_2 qui atteignent $C_n(\beta_1)$ et $C_n(\beta_2)$, c'est à dire telles que $\underline{Y}_1$ et $\underline{Y}_2$ soient les sorties associées et :
– $\underline{X}_i$ est β_i-admissible pour $i = 1, 2$;
– $I(\underline{X}_i; \underline{Y}_i) = C_n(\beta_i)$ pour $i = 1, 2$.
La source test $\underline{X}$ définie par la distribution $\alpha_1 p_1 + \alpha_2 p_2$ est $(\alpha_1\beta_1 + \alpha_2\beta_2)$-admissible, et donc $C_n(\alpha_1\beta_1 + \alpha_2\beta_2) \geq I(\underline{X}, \underline{Y})$, si on note $\underline{Y}$ la sortie associée. Comme $I(\underline{X}, \underline{Y})$ est concave en $p(x)$, on obtien :

$$\begin{aligned} I(\underline{X}, \underline{Y}) &\geq \alpha_1 I(\underline{X}_1, \underline{Y}_1) + \alpha_2 I(\underline{X}_2, \underline{Y}_2) \\ &= \alpha_1 C_n(\beta_1) + \alpha_2 C_n(\beta_2). \end{aligned}$$

□

Théorème 5.3.2 *Pour un canal discret sans mémoire,*

$$\forall n \geq 1, \forall \beta \geq \beta_{\min}, C_n(\beta) = n \cdot C_1(\beta)$$

Démonstration. Soit (X, Y) un couple de source test - sortie atteignant $C_1(\beta)$. En considérant la source test $\underline{X} = (X_1, \ldots, X_n)$ où $X, X_1, \ldots X_n$ sont i.i.d., on obtient facilement $C_n(\beta) \geq nC_1(\beta)$.

Pour ce qui est de l'autre sens de l'inégalité, soit $\underline{X}$ β-admissible qui atteint $C_n(\beta)$. Le canal n'ayant pas de mémoire, on a grâce au calcul fait en Section 5.1.1 :

$$C_n(\beta) = I(\underline{X}, \underline{Y}) \leq \sum_{i=1}^{n} I(X_i, Y_i).$$

Si on pose $\beta_i = \mathbb{E}[b(X_i]$, alors $\sum_i \beta_i \leq n\beta$.

Par définition, $\forall i, I(X_i, Y_i) \leq C_1(\beta_i)$, et on a donc :

$$\begin{aligned} C_n(\beta) &\leq n \cdot \frac{1}{n} \sum_{i=1}^{n} C_i(\beta_i) \\ &\leq n \cdot C_1\left(\frac{1}{n} \sum_{i=1}^{n} \beta_i\right) \qquad \text{(concavité de } C_1\text{)} \\ &\leq n \cdot C_1(\beta) \qquad \text{(croissance de } C_1\text{)} \end{aligned}$$

□

Corollaire 5.3.1 *Pour un canal discret sans mémoire (DMC en anglais discrete memoryless channel), $C(\beta) = C_1(\beta)$.*

Propriétés générales de $C(\beta)$ pour un DMC :

- $C(\cdot)$ est non-décroissante et concave pour $\beta \geq \beta_{\min}$. Donc $C(\cdot)$ est continue pour $\beta > \beta_{min}$. (En exercice, on montre $C(\beta)$ est continue en $\beta_{\min}$.
- La capacité du canal est le maximum de la fonction capacité-coût du canal (cela revient à accepter un coût infini). Si on note celle-ci $C_{\max}$ et $\beta_{\max} = \min\left\{ \mathbb{E}\left[b(\underline{X})\right] \middle| I(\underline{X}, \underline{Y}) = C_{max} \right\}$, alors on voit que $C(\beta) = C_{\max}$ pour $\beta \geq \beta_{\max}$ et $C(\beta) < C_{\max}$ pour $\beta < \beta_{\max}$. Donc $\beta \mapsto C(\beta)$ est strictement croissante sur $(\beta_{\min}, \beta_{\max})$. En particulier, on a pour $\beta \in [\beta_{\min}, \beta_{\max}]$:

$$C(\beta) = \max\left\{ I(X;Y) \middle| \mathbb{E}[b(X)] = \beta \right\}$$

Exemple 5.3.1: Canal binaire symmétrique

Prenons $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, $b(0) = 0$ et $b(1) = 1$ et les probabilités de transition suivantes :

$$Q = \begin{pmatrix} q & p \\ p & q \end{pmatrix},$$

avec $p, q \geq 0$ tq $p + q = 1$ et $p \leq 1/2$.

Alors $\beta_{\min} = 0$ et $C_{\min} = 0$.

Soit X une source test qui atteint $C(\beta)$ pour $0 \leq \beta \leq \beta_{\max}$, cette dernière quantité étant pour le moment inconnue. On a en particulier

$$\begin{aligned} P(X=1) &= b(1)P(X=1) + b(0)P(X=0) \\ &= \mathbb{E}\left[b(X)\right] \\ &= \beta \end{aligned}$$

et

$$\begin{aligned} C(\beta) &= I(X,Y) \\ &= H(Y) - H(Y|X) \\ &= H\left(Ber_{(1-\beta)p+\beta q}\right) - H\left(Ber_p\right) \\ &\leq 1 - H\left(Ber_p\right) \end{aligned}$$

avec égalité pour $\beta_{\max} = 1/2$, donc $C_{\max} = 1 - H\left(Ber_p\right) = 1 - H(p)$. Au final, on a :

$$C(\beta) = \begin{cases} H((1-\beta)q + \beta p) - H(p) & 0 \leq \beta \leq 1/2 \\ 1 - H(p) & \beta > 1/2. \end{cases}$$

Exemple 5.3.2: Reprenons l'exemple avec $\mathcal{X} = \left\{0, 1/2, 1\right\}$, $\mathcal{Y} = \{0,1\}$ et $b(0) = b(1) = 1$, $b\left(1/2\right) = 0$ et les probabilités de transition suivantes :

$$Q = \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \\ 0 & 1 \end{pmatrix}.$$

Alors $\beta_{min} = 0$ et $C_{min} = 0$.

Soit X une source test qui atteint $C(\beta)$ pour $0 \leq \beta \leq \beta_{\max}$. On a $\beta = \mathbb{E}\left[b(X)\right] = p(0) + p(1)$ et $I(X;Y)$ est concave en $p(0), p(1/2), p(1)$, donc par symmétrie $p(0) = p(1) = \beta_2$, on a alors :

$$C(\beta) = I(X,Y) = H(Y) - H(Y|X) = 1 - (1-\beta) = \beta.$$

Donc $\beta_{\max} = 1$ et

$$\begin{aligned} \forall \beta \in [0,1] \quad & C(\beta) = \beta \\ \forall \beta \in [1,\infty[\quad & C(\beta) = 1. \end{aligned}$$

EXEMPLE 5.3.3:

Prenons $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$, et $b(0) = b(1) = 1$, $b(2) = 4$ et les probabilités de transition suivantes :

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Alors $\beta_{min} = 1$ et $C_{min} = 1$.

Soit X une source test qui atteint $C(\beta)$, notons $\alpha_i = P(X = i)$. On a alors

$$\begin{aligned} C(\beta) &= H(Y) - H(Y|X) \\ &= H(Y) \qquad \text{car Y = X} \\ &= H(X) \\ &= H(\alpha_0, \alpha_1, \alpha_2) \end{aligned}$$

Cette quantité est maximale pour $\alpha_0 = \alpha_1 = \alpha_2 = 1/3$, et on obtient ainsi $C_{max} = \log_2 3$ et $\beta_{max} = 2$.

En maximisant la quantité $H(\alpha_0, \alpha_1, \alpha_2)$ sous la contrainte $1 \cdot \alpha_0 + 1 \cdot \alpha_1 + 4 \cdot \alpha_2 = \beta$, on obtient finalement

$$\forall \beta \in [1, 2], C(\beta) = H\left(2/3 - \beta/6, 2/3 - \beta/6, \beta/3 - 1/3\right).$$