Cours 7 — 29 mars 2011

Enseignant: Marc Lelarge *Scribe: Marc Lelarge*

Pour information

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

Sources sans mémoire et leurs fonctions taux-distorsion

7.1 La fonction taux-distorsion

Une source (discrète) émet une suite de symboles u dans un ensemble fini $\mathcal{U}$ appelé l'alphabet de la source. Une source sans mémoire est caractérisée par sa statistique $p(u)$: la suite de symboles est une suite de v.a. U_i indépendantes et identiquement distribuées (i.i.d.) de loi p.

Dans ce chapitre, nous considérons un encodage dans un alphabet de destination $\mathcal{V}$. Pour tout couple $(u, v) \in \mathcal{U} \times \mathcal{V}$, $d(u, v) \geq 0$ est l'erreur ou distortion associée. Pour des vecteurs $(\underline{u}, \underline{v}) \in \mathcal{U}^k \times \mathcal{V}^k$, on note $d(\underline{u}, \underline{v}) = \sum_{i=1}^k d(u_i, v_i)$.

Nous prendrons la convention : $\mathcal{U} = \{0, \dots, r-1\}$ et $\mathcal{V} = \{0, \dots, s-1\}$. La matrice de distorsion D est la matrice $r \times s$ à coefficients $d(u, v)$.

Étant donné $k > 0$, une source et une fonction d'encodage, la distorsion moyenne est alors :

$$E(d) \quad = \quad E[d(\underline{U}, \underline{V})] = \sum_{\underline{u},\underline{v}} p(\underline{u}) p(\underline{v}|\underline{u}) d(\underline{u}, \underline{v}).$$

On définit alors la fonction

$$R_k(\delta) = \min \{I(U; V),\ E(d) \leq k\delta\},$$

où la minimisation est faite sur toutes les paires $(\underline{U}, \underline{V})$ telles que $\underline{U} = (U_1, \dots U_k)$ et les $U_1, \dots U_k$ sont i.i.d. de loi $p(u)$.

Remarque 7.1.1 – *$p(\underline{v}|\underline{u})$ peut être vue comme la probabilité de transition d'un canal. On parlera alors de canal test.*
- *la fonction $p(\underline{v}|\underline{u}) \mapsto I(U; V)$ est continue donc elle atteint son minimum sur un compact.*
- *soit $\delta_{\min} = \sum_{u \in \mathcal{U}} p(u) \min_v d(v, u)$. On a $E(d) \geq k\delta_{\min}$, donc $R_k(\delta)$ est définie pour $\delta \geq \delta_{\min}$.*
- *la fonction $\delta \mapsto R_k(\delta)$ est clairement décroissante en δ.*

Définition 7.1.1 *La fonction taux-distorsion de la source est*

$$R(\delta) = \inf_k \frac{1}{k} R_k(\delta).$$

Théorème 7.1.1 *La fonction $\delta \mapsto R_k(\delta)$ est convexe pour $\delta \geq \delta_{\min}$.*

Démonstration. Soit $\alpha_1, \alpha_2 \geq 0$ avec $\alpha_1 + \alpha_2 = 1$. L'inégalité $R_k(\alpha_1\delta_1 + \alpha_2\delta_2) \leq \alpha_1 R_k(\delta_1) + \alpha_2 R_k(\delta_2)$ s'obtient facilement en considérant les probabilités de transition $p(\underline{v}|\underline{u}) = \alpha_1 p_1(\underline{v}|\underline{u}) + \alpha_2 p_2(\underline{v}|\underline{u})$ où p_i est un canal test atteignant $R_k(\delta_i)$ pour $i = 1, 2$ et en utilisant la convexité de $I(\underline{U}; \underline{V})$. $\square$

Théorème 7.1.2 *Pour une source discrète sans mémoire $R_k(\delta) = kR_1(\delta)$ pour tout k et $\delta \geq \delta_{\min}$.*

Démonstration. Soit $p(\underline{v}|\underline{u})$ une probabilité de transition atteignant $R_k(\delta)$, c'est à dire telle que

$$I(\underline{U}; \underline{V}) = R_k(\delta) \text{ et } E[d(\underline{U}, \underline{V})] \leq k\delta.$$

Comme les $U_1, U_2, \ldots, U_k$ sont indépendants, on a $I(\underline{U}; \underline{V}) \geq \sum_{i=1}^k I(U_i; V_i)$. Avec la notation $\delta_i = E[d(U_i, V_i)]$, on a $I(U_i; V_i) \geq R_1(\delta_i)$ donc $I(\underline{U}; \underline{V}) \geq \sum_{i=1}^k R_1(\delta_i)$. Par le résultat de convexité démontré ci-dessus, on obtient :

$$\sum_{i=1}^k R_1(\delta_i) \geq kR_1\left(\frac{1}{k}\sum_{i=1}^k \delta_i\right) \geq kR_1(\delta),$$

où on utilise la monotonie de R_1 dans la dernière inégalité. On a donc $R_k(\delta) \geq kR_1(\delta)$. Pour démontrer l'inégalité inverse, il suffit de considérer $p(\underline{v}|\underline{u}) = \prod_{i=1}^k p(v_i|u_i)$ où $p(v|u)$ atteint $R_1(\delta)$. $\square$

Corollaire 7.1.1 *Pour une source discrète sans mémoire $R(\delta) = R_1(\delta) = \min\{I(U; V),\ E[d] \leq \delta\}$.*

Propriétés de $R(\delta)$ pour une source DSM :

- la fonction $\delta \mapsto R(\delta)$ est décroissante, convexe donc continue pour $\delta > \delta_{\min}$. Elle est également continue en $\delta_{\min}$ (exo !).
- Soit $\delta_{\max} = \min_v \sum_u p(u)d(u, v)$ alors $R(\delta) = 0$ si et seulement si $\delta \geq \delta_{\max}$. En effet, soit $v^* = \arg\min \sum_u p(u)d(u, v)$ alors l'encodage déterministe $u \to v^*$ est tel que $I(U; v^*) = 0$ et $E[d] = \delta_{\max}$. Donc $R(\delta) = 0$ pour $\delta \geq \delta_{\max}$. Inversement si $R(\delta) = 0$ alors le canal test doit avoir U et V indépendants et donc

$$E[d] = \sum p(u)p(v)d(u, v) \geq \sum_v p(v)\delta_{\max} = \delta_{\max}.$$

- Comme $R(\delta)$ est décroissante, convexe pour $\delta \geq \delta_{\min}$ et constante pour $\delta \geq \delta_{\max}$, $R(\delta)$ est strictement décroissante pour $\delta_{\min} \leq \delta \leq \delta_{\max}$. En particulier, on a

$$R(\delta) = \min\{I(U;V),\ E[d] = \delta\} \text{ pour } \delta_{\min} \leq \delta \leq \delta_{\max}.$$

- Dans le cas particulier où chaque ligne de D a au moins un 0 et chaque colonne au plus un 0, on a $\delta_{\min} = 0$ et $E[d] = 0$ ssi $u \to v \in v(u) = \{v, d(u,v) = 0\}$ qui sont des ensembles disjoints par notre hypothèse. Donc $R(0) = I(U;V) = H(U) - H(U|V) = H(U)$.

Exemple 7.1.1:
$\mathcal{U} = \mathcal{V} = \{0,1\}$ avec $P(0) = p = 1 - P(1) \leq 1/2$ et matrice de distorsion

$$D = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

On a $\delta_{\min} = 0$, $\delta_{\max} = \min\{p, 1-p\} = p$ et $R(0) = H(p)$. Pour $0 < \delta < p$, soit (U,V) atteignant $R(\delta)$. On a $I(U,V) = H(U) - H(U|V) = H(p) - H(U|V)$ et $E[d] = P(U \neq V) = \delta$. Donc par l'inégalité de Fano, on a $H(U|V) \leq H(\delta)$ et donc $R(\delta \geq H(p) - H(\delta)$. Pour obtenir l'inégalité opposée, il faut trouver (U,V) tel que $E[d] = \delta$ et $I(U;V) = H(p) - H(\delta)$. Un candidat est de choisir $p(u|v) = \delta \mathbf{1}(u \neq v) + (1-\delta)\mathbf{1}(u = v)$ puisqu'on a alors $E[d] = \delta$ et $H(U|V) = H(\delta)$. Il faut donc vérifier qu'il est bien possible de choisir $\alpha = P(V = 0)$ tel que $P(U = 0) = p$. On calcule que $p = \alpha(1-\delta) + (1-\alpha)\delta$ soit $\alpha = \frac{p-\delta}{1-2\delta} \in [0,1]$ car $0 < \delta < p \leq 1/2$. Au final, on a montrer que

$$R(\delta) = \begin{cases} H(p) - H(\delta) & 0 \leq \delta \leq p, \\ 0 & \delta \geq p. \end{cases}$$

7.2 Théorème de codage de source de Shannon

Soit $U^{(k)} = (U_1, \ldots, U_k)$ un vecteur representant les k premiers symboles émis par une source. On suppose que ces k symboles sont 'compressés'en n bits $X^{(n)} = (X_1, \ldots, X_n)$ et qu'il est possible à partir de ces n bits de les 'décoder' en k symboles de l'alphabet de destination : $V^{(k)} = (V_1, \ldots, V_k)$ de telle sorte que $\sum_{i=1}^{k} E[d(U_i, V_i)] \leq k\delta$. Il est facile de voir que l'on a nécessairement :

$$\frac{n}{k} \geq R(\delta). \tag{7.1}$$

En effet, par définition, on a $I(U^{(k)}; V^{(k)}) \geq R_k(\delta)$. De plus par l'inégalité data processing, on a $I(U^{(k)}; V^{(k)}) \leq I(X^{(k)}; V^{(k)}) \leq H(X^{(k)}) \leq n$, ce qui implique que $R_k(\delta)/k \leq n/k$ et donc (7.1) découle (sans avoir fait l'hypothèse que la source est sans mémoire).

On voit donc qu'il faut au moins $R(\delta)$ bits pour représenter un symbole source si la distorsion moyenne doit être inférieure à δ. Nous allons maintenant montrer qu'il ne faut 'pas plus de $R(\delta)$ bits'.

Définition 7.2.1 *Un* **code de source** *de longueur k est un ensemble $C = \{\underline{v}_1, \ldots, \underline{v}_M\} \subset \mathcal{V}^k$. Le taux du code est $R = \frac{1}{k}\log_2 M$.*

Pour chaque suite de la source $\underline{u} = (u_1, \ldots, u_k)$, $f(\underline{u})$ est le mot code $\underline{v}_i$ le plus proche de $\underline{u}$:

$$d(\underline{u}, f(\underline{u})) \leq d(\underline{u}, \underline{v}_j) \quad \forall j \in \{1, \ldots, M\}.$$

La distorsion moyenne de C est $d(C) = \frac{1}{k}\sum_{\underline{u}\in\mathcal{U}^k} p(\underline{u})d(\underline{u}, f(\underline{u}))$, avec $p(\underline{u}) = p(u_1)\ldots p(u_k)$.

Exemple 7.2.1: code $(7,4)$ de Hamming : $\mathcal{U} = \mathcal{V} = \{0,1\}$ avec $p(0) = p(1) = 1/2$ et $D = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Chacun des 128 vecteurs binaires diffère d'au plus un bit par rapport à un mot code donc

$$d(C) = \frac{1}{7}\left(\frac{128-16}{128}\right) = \frac{1}{8}.$$

Théorème 7.2.1 *Soit $\delta \geq \delta_{\min}$. Pour tout $\delta' > \delta$ et $R' > R(\delta)$ pour k suffisament grand, il existe un codage de source C de longueur k avec M mots code tel que*
- *a) $M \leq 2^{\lfloor kR' \rfloor}$;*
- *b) $D(C) < \delta'$.*

Démonstration. Soient $R(\delta) < R'' < R'$ et $\delta < \delta'' < \delta'$. □

Pour un code $C = \{\underline{v}_1, \ldots, \underline{v}_M\}$ on définit les ensembles :

$$\begin{aligned} S &= \{\underline{u},\ d(\underline{u}, f(\underline{u})) \leq k\delta''\} \text{ (suites bien représentées par } C\text{)},\\ T &= \{\underline{u},\ d(\underline{u}, f(\underline{u})) > k\delta''\} \text{ (suites mal représentées par } C\text{)}.\end{aligned}$$

On a donc

$$d(C) \leq \underbrace{\frac{1}{k}\sum_{\underline{u}\in S} p(\underline{u})d(\underline{u}, f(\underline{u}))}_{\leq \delta''} + \underbrace{\frac{1}{k}\sum_{\underline{u}\in T} p(\underline{u})d(\underline{u}, f(\underline{u}))}_{\leq B\sum_{\underline{u}\in T} p(\underline{u})}$$

avec $B = \max d(u,v)$. On a donc

$$d(C) \leq \delta'' + BP\Big(d(\underline{u}, f(\underline{u})) > k\delta''\Big), \tag{7.2}$$

où la probabilité vient de l'aléa de la source. Pour être plus explicite, on définit

$$\Delta(\underline{u},\underline{v}) = \begin{cases} 1 & \text{si } d(\underline{u},\underline{v}) \leq k\delta'' \\ 0 & \text{sinon.} \end{cases}$$

On peut donc écrire (7.2) comme suit :

$$d(C) \leq \delta'' + B \underbrace{\sum_{\underline{u}} p(\underline{u}) \prod_{i=1}^{M} \left(1 - \Delta(\underline{u},\underline{v}_i)\right)}_{K(C)}.$$

Il faut donc trouver un code de source C de longueur k avec au plus $2^{\lfloor kR' \rfloor}$ mots code et tel que $K(C) < \frac{\delta' - \delta''}{B}$. Nous utilisons à nouveau un argument de **random coding**. Soit $p(u,v)$ la probabilité sur $\mathcal{U} \times \mathcal{V}$ qui atteint $R(\delta)$: $I(U;V) = R(\delta)$ et $E[d(U,V)] \leq \delta$. On note les marginales par $p(u) = \sum_v p(u,v)$ et $p(v) = \sum_u p(u,v)$. On définit alors une probabilité sur $\mathcal{U}^k \times \mathcal{V}^k$ par $p(\underline{u}) = \prod_{i=1}^k p(u_i)$, $p(\underline{v}|\underline{u}) = \prod_{i=1}^k p(v_i|u_i)$ de telle sorte que $p(\underline{u},\underline{v}) = \prod_{i=1}^k p(u_i,v_i)$. On définit alors une probabilité sur les codes de longueur k et ayant M mots code (vu comme des M-uplets de vecteurs de dimensions k) par :

$$p(C) = \prod_{i=1}^{M} p(\underline{v}_i) \text{ avec } p(\underline{v}_i) = \prod_{j=1}^{k} p(v_{ij}).$$

Nous calculons la moyenne de $K(C)$ sous cette probabilité :

$$\begin{aligned} E[K] &= \sum_{\underline{v}_1 \cdots \underline{v}_M} p(\underline{v}_1) \ldots p(\underline{v}_M) \sum_{\underline{u}} p(\underline{u}) \prod_{i=1}^{M} \left(1 - \Delta(\underline{u},\underline{v}_i)\right) \\ &= \sum_{\underline{u}} p(\underline{u}) \left(\sum_{\underline{v}} p(\underline{v})(1 - \Delta(\underline{u},\underline{v}) \right)^M. \end{aligned}$$

On définit maintenant

$$\Delta_0(\underline{u},\underline{v}) = \begin{cases} 1 & \text{si } d(\underline{u},\underline{v}) \leq k\delta'' \text{ et } I(\underline{u};\underline{v}) \leq kR'', \\ 0 & \text{sinon.} \end{cases}$$

avec $I(\underline{u};\underline{v}) = \log_2\left(\frac{p(\underline{v}|\underline{u})}{p(\underline{v})}\right)$. Si $\delta_0(\underline{u},\underline{v}) = 1$ alors $p(\underline{v}) \geq 2^{-kR''} p(\underline{v}|\underline{u})$ et donc

$$E[K] \leq \sum_{\underline{u}} p(\underline{u}) \left(1 - 2^{-kR''} \sum_{\underline{v}} p(\underline{v}|\underline{u}) \Delta_0(\underline{u},\underline{v})\right)^M,$$

et donc en utilisant l'inégalité $(1-xy)^M \leq 1 - x + e^{-yM}$ valide pour $0 \leq x, y \leq 1$ et $M > 0$ avec $x = \sum_{\underline{v}} p(\underline{v}|\underline{u})\Delta_0(\underline{u},\underline{v})$ et $y = 2^{-kR''}$, on obtient

$$\begin{aligned} E[k] &\leq 1 - \sum_{\underline{u},\underline{v}} p(\underline{u},\underline{v})\Delta_0(\underline{u},\underline{v}) + \exp\left(-2^{-kR''}M\right) \\ &\leq P(d(\underline{U},\underline{V}) > k\delta'') + P(I(\underline{U};\underline{V}) > kR'') + \exp\left(-2^{-kR''}M\right). \end{aligned}$$

Le dernier terme tend vers 0 quand k tend vers l'infini car $M = 2^{\lfloor kR' \rfloor}$ avec $R' > R''$. Le fait que les deux premiers temes tendent également vers 0 découle de la loi faible des grands nombres. En effet, il suffit d'écrire : $d(\underline{U},\underline{V}) = \sum_{i=1}^k d(U_i, v_i)$ qui est une somme de k v.a. i.i.d. de moyenne $E[d(U,V)] \leq \delta < \delta''$. De même $I(\underline{U};\underline{V}) = \sum_{i=1}^k I(U_i; V_i)$ est une somme de v.a. i.i.d. de moyenne $R(\delta) < R''$.