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

Cours 1 — 8 février 2011

Enseignant: Marc Lelarge *Scribe: Marc Lelarge*

Pour information

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

Notations

Pour des variables aléatoires (v.a.) discrètes X et Y à valeurs dans $\mathcal{X}$ et $\mathcal{Y}$ resp., on utilisera les notations suivantes pour $x \in \mathcal{X}$ et $y \in \mathcal{Y}$:

$$\begin{aligned} p(x) &= P(X = x) \\ p(y) &= P(Y = y) \\ p(x, y) &= P(X = x, Y = y) \\ p(x|y) &= P(X = x|Y = y) = p(x, y)/p(y). \end{aligned}$$

Lorsque ces notations sont ambigues, on pourra écrire $p_X(x), p_Y(y), p_{X,Y}(x, y), p_{X|Y}(x|y)$.

1.1 Codage de source et test d'hypothèse

1.1.1 Entropie

Une source (discrète) émet une suite de v.a. $\{U_i\}_{i=1}^{\infty}$ à valeurs dans un ensemble fini $\mathcal{U}$ appelé l'alphabet de la source. Si les U_i sont indépendants et identiquement distribués (i.i.d.) de loi P, la source est dite sans mémoire de distribution P.

Définition 1.1.1 *Soit U une variable aléatoire à valeurs dans un ensemble fini $\mathcal{U}$, de distribution de probabilité :*

$$p(u) = P(U = u),\ u \in \mathcal{U}\,.$$

Son entropie est, par définition, la quantité

$$H(U) = -E\left[\log(p(U))\right] = -\sum_{u \in \mathcal{U}} p(u) \log p(u)\,,$$

avec la convention $0 \log 0 = 0$.

Le choix de la base du logarithme correspond à un choix d'unité. Dans ce premier cours, on choisit la base 2. L'entropie s'exprime alors en bits.

Un (k, n)-codage binaire est une paire de fonctions

$$f : \mathcal{U}^k \to \{0,1\}^n, \text{ et } \phi : \{0,1\}^n \to \mathcal{U}^k.$$

Pour une source donnée, la probabilité d'erreur du code (f, ϕ) est

$$e(f, \phi) := P(\phi(f(U^{(k)})) \neq U^{(k)}),$$

avec $U^{(k)} = (U_1, \ldots, U_k)$ les k premiers symboles émis par la source.

Le but est de trouver des codes avec un ratio n/k petit et une probabilité d'erreur petite. Plus précisément, pour tout k, soit $n(k, \epsilon)$ le plus petit entier n tel qu'il existe un (k, n)-code satisfaisant $e(f, \phi) \leq \epsilon$.

Théorème 1.1.1 *Pour une source discrète sans mémoire de distribution $P(U = u) = p(u)$, on a pour tout $\epsilon \in (0, 1)$:*

$$\lim_{k\to\infty} \frac{n(k, \epsilon)}{k} = H(U) = -\sum_{u\in\mathcal{U}} p(u) \log p(u).$$

Démonstration. L'existence d'un (k, n)-code binaire avec $e(f, \phi) \leq \epsilon$ est équivalente à l'existence d'un ensemble $A \subset \mathcal{U}^k$ avec $P(A) \geq 1 - \epsilon$ et $|A| \leq 2^n$. A est alors l'ensemble des suites $u^{(k)} \in \mathcal{U}^k$ reproduites de manière exacte, c.a.d. telles que $\phi(f(u^{(k)})) = u^{(k)}$.

Soit $s(k, \epsilon)$ la taille mimimale d'un ensemble $A \subset \mathcal{U}^k$ avec $P(A) \geq 1 - \epsilon$, c.a.d

$$s(k, \epsilon) = \min\{|A|;\ P(A) \geq 1 - \epsilon\}.$$

Pour prouver le théorème, il suffit de montrer que pour $\epsilon \in (0, 1)$,

$$\lim \frac{\log s(k, \epsilon)}{k} = H(U). \tag{1.1}$$

Pour ceci, on définit l'ensemble $B(k, \delta) \subset \mathcal{U}^k$ suivant

$$B(k, \delta) = \left\{u^{(k)} \in \mathcal{U}^k,\ 2^{-k(H(U)+\delta)} \leq p(u^{(k)}) \leq 2^{-k(H(U)-\delta)}\right\}.$$

Montrons tout d'abord que $\lim_{k\to\infty} P(B(k, \delta) = 1$ pour tout $\delta > 0$. Soit la v.a. réelle

$$Y_i = -\log p(U_i),$$

qui est bien définie avec probabilité 1 même s'il existe $u \in \mathcal{U}$ avec $p(u) = 0$. Les Y_i sont i.i.d. de moyenne $H(U)$. Par la loi faible des grands nombres, on a

$$\lim_{k\to\infty} P\left(\left|\frac{1}{k}\sum_{i=1}^{k} Y_i - H(U)\right| \leq \delta\right) = 1 \text{ pour tout } \delta > 0.$$

Comme $U^{(k)} \in B(k,\delta)$ ssi $\left|\frac{1}{k}\sum_{i=1}^{k} Y_i - H(U)\right| \leq \delta$, on a bien

$$\lim_{k\to\infty} P(B(k,\delta)) = 1 \text{ pour tout } \delta > 0. \tag{1.2}$$

La définition de $B(k,\delta)$ implique que $|B(k,\delta)| \leq 2^{k(H(U)+\delta)}$. Par (1.2), on a donc pour tout $\delta > 0$

$$\limsup_{k\to\infty} \frac{1}{k}\log s(k,\epsilon) \leq \limsup_{k\to\infty} \frac{1}{k}\log |B(k,\delta)| \leq H(U) + \delta. \tag{1.3}$$

Dans l'autre sens, pour tout $A \subset \mathcal{U}^k$ avec $P(A) \geq 1-\epsilon$, (1.2) implique pour k suffisament grand,

$$P(A \cap B(k,\delta)) \geq \frac{1-\epsilon}{2}.$$

On a donc par définition de $B(k,\delta)$,

$$|A| \geq |A \cap B(k,\delta)| \geq \sum_{u^{(k)} \in A\cap B(k,\delta)} p(u^{(k)}) 2^{k(H(U)-\delta)} \geq \frac{1-\epsilon}{2} 2^{k(H(U)-\delta)},$$

et donc pour tout $\delta > 0$,

$$\liminf_{k\to\infty} \frac{1}{k}\log s(k,\epsilon) \geq H(U) - \delta.$$

Ceci, avec (1.3), implique (1.1). □

Corollaire 1.1.1

$$0 \leq H(U) \leq \log |\mathcal{U}|$$

1.1.2 Information mutuelle

Nous généralisons maintenant l'approche précédente en donnant des poids différents aux éléments de $\mathcal{U}^k$. Plus précisément, nous considérons une suite de "fonctions masse" à valeur positive $M_1(u), M_2(u), \ldots$ sur $\mathcal{U}$ et on définit

$$M(u^{(k)}) := \prod_{i=1}^{k} M_i(u_i), \text{ pour } u^{(k)} = (u_1, \ldots, u_k),$$

et pour $A \subset \mathcal{U}^k$, $M(A) := \sum_{u^k \in A} M(u^k)$. On définit alors

$$s(k,\epsilon) = \min\{M(A);\ A \subset \mathcal{U}^k,\ P(A) \geq 1-\epsilon\}.$$

Théorème 1.1.2 *Si les U_i sont indépendants de loi p_i et $\max_{i\in[k],u\in\mathcal{U}} |\log M_i(u)| \leq c$ alors pour tout $\epsilon \in (0,1)$,*

$$\lim_{k\to\infty} \left(\frac{1}{k} \log s(k,\epsilon) - E_k \right) = 0,$$

avec

$$E_k := \frac{1}{k} \sum_{i=1}^{k} \sum_{u\in\mathcal{U}} p_i(u) \log \frac{M_i(u)}{p_i(u)}.$$

Démonstration. Soit $Y_i = \log \frac{M_i(U_i)}{p_i(U_i)}$. Les Y_i sont indépendants tels que $E\left[\frac{1}{k}\sum_{i=1}^{k} Y_i\right] = E_k$ donc l'inégalité de Chebyshev donne pour tout $\delta > 0$

$$P\left(\left| \frac{1}{k} \sum_{i=1}^{k} Y_i - E_k \right| \geq \delta \right) \leq \frac{1}{k^2\delta^2} \sum_{i=1}^{k} \text{Var}\,(Y_i) \leq \frac{1}{k\delta^2} \max_i \text{Var}\,(Y_i)$$

Ceci signifie que pour l'ensemble

$$B(k,\delta) = \left\{ u^{(k)} \in \mathcal{U}^k,\ E_k - \delta \leq \frac{1}{k} \log \frac{M(u^{(k)})}{p(u^{(k)})} \leq E_k + \delta \right\},$$

on a $P(B(k,\delta)) \geq 1 - \eta_k$ avec $\eta_k = \frac{1}{k\delta^2} \max_i \text{Var}\,(Y_i)$.

Par définition, on a

$$\begin{aligned} M(B(k,\delta)) &= \sum_{u^{(k)}\in B(k,\delta)} M(u^{(k)}) \\ &\leq \sum_{u^{(k)}\in B(k,\delta)} p(u^{(k)}) 2^{k(E_k+\delta)} \\ &\leq 2^{k(E_k+\delta)}. \end{aligned}$$

On en déduit que $\frac{1}{k}\log s(k,\epsilon) \leq \frac{1}{k}\log M(B(k,\delta)) \leq E_k + \delta$ si $\eta_k \leq \epsilon$.

Dans l'autre sens, pour tout $A \subset \mathcal{U}^k$ ave $P(A) \geq 1 - \epsilon$, on a $P(A \cap B(k,\delta)) \geq 1 - \epsilon - \eta_k$ d'où

$$\begin{aligned} M(A) \geq M(A \cap B(k,\delta)) &\geq \sum_{u^{(k)}\in A\cap B(k,\delta)} p(u^{(k)}) 2^{k(E_k-\delta)} \\ &\geq (1 - \epsilon - \eta_k) 2^{k(E_k-\delta)}. \end{aligned}$$

On a donc $\frac{1}{k}\log s(k,\epsilon) \geq \frac{1}{k}\log(1-\epsilon-\eta_k) + E_k - \delta$. Ceci conclut la preuve en notant que l'hypothèse $\max_{i\in[k],u\in\mathcal{U}} |\log M_i(u)| \leq c$ implique que $\max_i \text{Var}\ Y_i$ est borné et donc que $\eta_k \to 0$ quand $k \to \infty$. □

Application: Test d'hypothèse.

Problème : décider entre deux distributions P et Q à partir d'un échantillon de taille k, c.a.d. le résultat de k tirages indépendants. Un test est défini par un ensemble $A \subset \mathcal{U}^k$: si l'échantillon $(U_1, \ldots, U_k)$ appartient à A alors le test retourne l'hypothèse P sinon Q.

On considère un scénario où les hypothèses ne sont pas symmétriques. On désire une probabilité d'erreur au plus ϵ si P est la vraie distribution, c.a.d. $P(A) \geq 1 - \epsilon$. Le but est alors de minimiser la probabilité d'erreur si l'hypothèse Q est vraie, c.a.d

$$\beta(k, \epsilon) = \min\{Q(A), \text{ t.q. } A \subset \mathcal{U}^k, \ P(A) \geq 1 - \epsilon\}.$$

Corollaire 1.1.2 *Pour tout $\epsilon \in (0, 1)$, on a*

$$\lim_{k \to \infty} \frac{1}{k} \log \beta(k, \epsilon) = \sum_{u \in \mathcal{U}} p(u) \log \frac{p(u)}{q(u)}.$$

Démonstration. Si $q(u) > 0$ pour tout $u \in \mathcal{U}$, il suffit d'appliquer le théorème précédent avec $P_i = P$ et $M_i = Q$.

S'il existe un u avec $p(u) > q(u) = 0$ alors la P-probabilité que l'échantillon de taille k contienne ce u tend vers 1 et $\beta(k, \epsilon) = 0$ pour k suffisament grand et dons l'assertion est encore valide. □

Définition 1.1.2 *L'entropie relative ou distance de Kullback-Leibler entre deux distributions p et q est définie par :*

$$D(p||q) = E_p\left[\log \frac{p(U)}{q(U)}\right] = \sum_{u \in \mathcal{U}} p(u) \log \frac{p(u)}{q(u)},$$

avec les conventions $0 \log \frac{0}{0} = 0$, $0 \log \frac{0}{q} = 0$ et $p \log \frac{p}{0} = \infty$.

L'entropie d'une paire (U, V) ne nécessite pas de nouvelle définition ! On notera $H((U, V)) = H(U, V)$. La différence $H(U, V) - H(U)$ mesure la quantité d'information supplémentaire sur le couple (U, V) donnée par V si U est déjà connu. Cette différence est appelée l'entropie conditionnelle de V sachant U et est notée :

$$H(V|U) = H(U, V) - H(U)$$

On a donc

$$H(V|U) = -\sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} p(u, v) \log \frac{p(u, v)}{p(u)} = \sum_{u \in \mathcal{U}} p(u) H(V|U = u),$$

avec $H(V|U = u) = -\sum_{v \in \mathcal{V}} p(v|u) \log p(v|u)$.

Lemme 1.1.1 *On a*

$$H(V|U) \leq H(V)$$

Démonstration.

$$\begin{aligned} H(V) - H(V|U) &= H(V) - H(U,V) + H(U) \\ &= \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} p(u,v) \log \frac{p(u,v)}{p(u)p(v)} \\ &= D(p_{U,V} \| p_U p_V) \geq 0. \end{aligned}$$

□

Définition 1.1.3 *L'information mutuelle entre U et V est définie par*

$$I(U;V) = H(V) - H(V|U) = H(U) - H(U|V) = H(V) + H(U) - H(U,V).$$

L'information mutuelle entre U et V correspond à la diminution d'incertitude sur V causée par la connaissance de U, c.a.d la quantité d'information sur V contenue dans U. Elle est symmétrique en U et V.