ACCQ202 23 Septembre 2016

Cours 2

Enseignant: A. Tchamkerten *Crédit L. Groléaz and J. Parvillers*

Refs: Cover and Thomas Chapitres 2, 5.6-5.8, 2.5, 3, 7.6

1 Entropie et questionnement optimal

Exemple 1. $X \in \mathcal{X}$
But : Identifier X avec le moins de questions possibles.
Questions autorisées : $X \in A \subseteq \mathcal{X}$ *?*
Idée : Le questionnement équivaut à un code.
Ainsi le nombre moyen de questions est supérieur ou égal à H(X).
Une stratégie de questionnement optimal correspond donc à un code de Huffman.

2 Entropie et information mutuelle: propriétés

Definition 1.

$$\begin{aligned} H(X) &= -\sum_{x \in \mathcal{X}} p(x) \log p(x) \\ &= E_X(\log \frac{1}{p(X)}) \end{aligned}$$

Exemple 2.

$$X \in \{0, 1\}$$

avec $p(X = 0) = p = 1 - p(X = 1)$

$$H(X) = -p \log p - (1 - p) \log(1 - p)$$

que l'on note $H_b(p)$.

Theorem 2.

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

Preuve du théorème

$$\begin{aligned}
&\forall x \in \mathcal{X}\ 0 \leq p(x) \leq 1 \\
&\Rightarrow \forall x \in \mathcal{X}\ 0 \leq -\log(p(x)) \\
&\Rightarrow -\sum_{x\in\mathcal{X}} p(x) log(p(x)) \geq 0
\end{aligned}$$

Pour l'autre inégalité : On note $\forall x \in \mathcal{X}\ u(x) = \frac{1}{|\mathcal{X}|}$
alors

$$\begin{aligned}
H(X) &= -\sum_x p(x)\log(\frac{p(x)u(x)}{u(x)}) \\
&= -D(P||U) + \log(\mathcal{X})
\end{aligned}$$

or D(P||U) est positif. D'où le résultat.
Avec égalité si et seulement si P est uniforme, i.e. $D(P||U) = 0$.

Definition 3 (Entropie conjointe)**.** *On suppose* $(X, Y) \sim P_{X,Y}$

$$\begin{aligned}
H(X, Y) &= -\sum_{x,y} p(x, y) \log p(x, y) \\
&= E_{X,Y}(\log \frac{1}{p(X, Y)})
\end{aligned}$$

Definition 4 (Entropie conditionnelle)**.** *On suppose* $(X, Y) \sim P_{X,Y}$

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

$$\text{où} -\sum_x p(x|y) \log p(x|y) \stackrel{def}{=} H(X|Y = y)$$

Remarque 1.

$$\begin{aligned}
X = Y &\Rightarrow H(X|Y) = 0 \\
X \perp Y &\Rightarrow H(X|Y) = H(X)
\end{aligned}$$

Theorem 5 (Chain Rule). $H(X_1, X_1, ..., X_n) = \sum_i H(X_{i+1}|X^i)$ *où* $X^i = (X_1, X_2, ..., X_i)$

Preuve du théorème

$H(X_1, ..., X_n) = -E(\log(p(X_1, ..., X_n)))$
Or $p(X_1, ..., X_n) = \prod_{i=1}^{n} p(X_i|X^{i-1})$ et ainsi

$$\begin{aligned} H(X_1, ..., X_n) &= -\sum_{i=1}^{n} E(\log p(X_i|X^{i-1})) \\ &= \sum_{i=1}^{n} H(X_i|X^{i+1}) \end{aligned}$$

Definition 6 (Information mutuelle). $(X, Y) \sim P_{X,Y}$
$I(X, Y) = \sum_{x,y} p(x, y) \log \frac{p(x,y)}{p(x)p(y)}$
Où $p(x) = \sum_y p(x, y)$ *[Loi marginale]*

Theorem 7.

1. $I(X, Y) = I(Y, X)$
2. $I(X, Y) = H(X) - H(X|Y)$
3. $I(X, Y) = H(Y) - H(Y|X)$
4. $I(X, Y) = H(X) + H(Y) - H(X, Y)$
5. $I(X, X) = H(X)$
6. $I(X, Y) = D(P_{X,Y}||P_X.P_Y)$
7. $I(X, Y) = 0 \Leftrightarrow X \perp Y$
8. $H(Y) \geq H(Y|X)$
9. $H(X^n) \leq \sum_{i=1}^{n} H(X_i)$ *avec égalité ssi* X_i *sont iid*
10. $H(X)$ *est concave en* P_X
11. $\forall f \;\; H(X) \geq H(f(X))$

Preuve du théorème

1. Evident à partir de la définition
2. Développer H(X)-H(X|Y)
3. Immédiat avec 1) et 2)
4. Chain rule.
5. Avec 2) car $H(X|X) = 0$
6. Définition de $D(P_{X,Y}||P_X P_Y)$
7. $I(X,Y) = 0 \leftrightarrow D(P_{X,Y}||P_X P_Y) = 0 \Leftrightarrow P_{X,Y} = P_X P_Y \;\Leftrightarrow\; X \perp Y$
8. $H(Y) - H(Y|X) = I(X,Y) = D(P_{X,Y}||P_X P_Y) \geq 0$
9. $H(X^n) = \sum_{i=1}^n H(X_i|X^{i-1})$ et $H(X_i|X^{i-1}) \leq H(X_i)$ par (8).
10. Soit $X \sim P_X$, $Y \sim P_Y$, et $Z \sim P = \alpha P_X + (1-\alpha) P_Y$, $0 \leq \alpha \leq 1$. Il s'agit de vérifer que

$$H(Z) \geq \alpha H(X) + (1-\alpha) H(Y).$$

Soit T une variable aléatoire binaire, indépendante de X et de Y avec $Pr(T=1) = \alpha = 1 - Pr(T=0)$. Il suit que

$$Z = T \cdot X + (1-T) \cdot Y.$$

D'où en utilisant (8),

$$\begin{aligned} H(Z) &\geq H(Z|T) \\ &= H(Z|T=1)\alpha + H(Z|T=0)(1-\alpha) \\ &= H(X)\alpha + H(Y)(1-\alpha). \end{aligned}$$

11. $H(X) = H(X, f(X)) = H(f(X)) + H(X|f(X))$ et $H(X|f(X)) \geq 0$.

Définition 2.1. *Pour*

$$(X, Y, Z) \sim p(x, y, z).$$

On définit

$$I(X;Y|Z) = \mathbb{E}_{p(x,y,z)}\left[\log(\frac{\mathbb{P}(X,Y|Z)}{\mathbb{P}(X|Z)\mathbb{P}(Y|Z)})\right]$$

Propriété 2.1. (chain rule)

$$I(X_1...X_n;Y) = \sum\nolimits_{i\in[1,n]} I(X_i;Y|X^{i-1})$$

Preuve:

$$\begin{aligned} I(X^n;Y) &= H(X^n) - H(X^n|Y) \\ &= \sum\nolimits_{i\in[1,n]} H(X_i|X^{i-1}) - H(X_i|X^{i-1},Y) \\ &= I(X_i;Y|X^{i-1}) \end{aligned}$$

3 Typicalité

Rappel: lois des grands nombres Soit $(Z_i)_{i\in[1,n]}$ une famille de variables aléatoires i.i.d. Alors:

$$\frac{1}{n}\sum\nolimits_{i\in[1,n]} Z_i \underset{n\to+\infty}{\overset{\text{en probabilité}}{\longrightarrow}} \mathbb{E}(Z_1)$$

i.e. $\forall\ \epsilon > 0$, $\forall\ \delta > 0$, $\exists$ N $\in \mathbb{N}$, $\forall$ n $\geq$ N,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum\nolimits_{i\in[1,n]} Z_i - \mathbb{E}(Z_1)\right| \geq \epsilon\right) \leq \delta$$

4 Probabilité asymptotique d'équipartition

Théorème 4.1. (probabilité asymptotique d'équirépartition - A.E.P)
Soit $(X_i)_{i\in[1,n]}$ une suite de variables aléatoires i.i.d. Alors:

$$-\frac{1}{n}\log\left[p(X_1,...,X_n)\right] \underset{n\to+\infty}{\overset{\textit{en probabilité}}{\longrightarrow}} H(X)$$

Preuve:

$$\begin{aligned} -\frac{1}{n}\log\left[p(X_1,...,X_n)\right] &= -\frac{1}{n}\log\left[\prod\nolimits_{i\in[1,n]} p(x_i)\right] \\ &= -\frac{1}{n}\sum\nolimits_{i\in[1,n]} \log\left[p(x_i)\right] \\ &\underset{n\to+\infty}{\longrightarrow} \mathbb{E}\left[\log\left(\frac{1}{X}\right)\right] = H(X) \end{aligned}$$

Définition 4.1. (ensemble typique)
L'ensemble typique par rapport à une distribution $p(x)$ est défini par:

$$\mathcal{A}_\epsilon^n = \left\{x^n : 2^{-n(H(X)+\epsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\epsilon)}\right\}$$

Propriété 4.1.
$\forall \epsilon > 0, \exists N \in \mathbb{N}, \forall n \geq N,$

$$\begin{aligned} \mathbb{P}(\mathcal{A}_\epsilon^n) &\geq 1-\epsilon \\ |\mathcal{A}_\epsilon^n| &\leq 2^{n(H(X)+\epsilon)} \\ |\mathcal{A}_\epsilon^n| &\geq (1-\epsilon)2^{n(H(X)-\epsilon)} \end{aligned}$$

Preuve:
$\forall \epsilon > 0$, *et pour n suffisamment grand, l'A.E.P donne:*
$\mathbb{P}(\mathcal{A}_\epsilon^n) = \mathbb{P}\left(\left|\frac{1}{n}\log\left(\frac{1}{p(x_i)}\right) - H(X)\right| \leq \epsilon\right) \geq 1-\epsilon$

$$\begin{aligned} 1 = \sum\nolimits_{x^n \in [1,n]} p(x^n) &\geq \sum\nolimits_{x^n \in \mathcal{A}_\epsilon^n} p(x^n) \\ &\geq \sum\nolimits_{x^n \in \mathcal{A}_\epsilon^n} 2^{-n(H(X)+\epsilon)} \\ D'o\grave{u} |\mathcal{A}_\epsilon^n| &\leq 2^{n(H(X)+\epsilon)} \end{aligned}$$

Pour n suffisamment grand,

$$\begin{aligned} 1-\epsilon &\leq \mathbb{P}(\mathcal{A}_\epsilon^n) \\ &= \sum\nolimits_{x^n \in \mathcal{A}_\epsilon^n} p(x^n) \\ &\leq |\mathcal{A}_\epsilon^n|\, 2^{-n(H(X)-\epsilon)} \end{aligned}$$

D'où $(1-\epsilon)2^{n(H(X)-\epsilon)}$
Remarque: *on notera dès lors "$a_n \doteq b_n$" lorsque* $\forall \epsilon > 0, \exists N \in \mathbb{N}, \forall n \geq N,$

$$\left|\frac{1}{n}\log\left(\frac{a_n}{b_n}\right)\right| \leq \epsilon$$

5 Codage de source revisité

Un code $\mathcal{C}$ est tel que:

$$\mathcal{C} : x^n = (x_1, ..., x_n) \mapsto \mathcal{C}(x_1, ..., x_n)$$

$$\left\{\begin{array}{c} \text{Si } x^n \in \mathcal{A}_\epsilon^n \text{ alors: } \mathcal{C}(x^n) \triangleq c_1 ... c_k \\ \text{Si } x^n \notin \mathcal{A}_\epsilon^n \text{ alors: } \mathcal{C}(x^n) \triangleq c_1 ... c_l \end{array}\right.$$

Avec:

$$\left\{\begin{array}{c} k \leq n(H(X)+\epsilon)+1 \\ l \leq n\log|\chi|+1 \end{array}\right.$$

Et alors:

$$\begin{array}{rcl} L(\mathcal{C}) & = & \sum_{x\in\chi} l(x^n)p(x^n) \\ & = & \sum_{x^n\in\mathcal{A}_\epsilon^n} l(x^n)p(x^n) + \sum_{x^n\notin\mathcal{A}_\epsilon^n} l(x^n)p(x^n) \\ & \leq & n(H(X)+\epsilon)+1+\epsilon(n\log|\chi|+1) \triangleq n(H(X)+\epsilon'(n)) \end{array}$$

où $\epsilon'(n)\epsilon(1+\log|\chi|+1/n)+{}^1\!/_n$

6 Typicalité conjointe

Définition 6.1.

$$\tilde{\mathcal{A}}_\epsilon^n = \left\{(x^n,y^n), \left\{\begin{array}{c} \left|-\frac{1}{n}\log\left(p(x^n)\right)-H(X)\right| \leq \epsilon \\ \left|-\frac{1}{n}\log\left(p(y^n)\right)-H(Y)\right| \leq \epsilon \\ \left|-\frac{1}{n}\log\left(p(x^n,y^n)\right)-H(X,Y)\right| \leq \epsilon \end{array}\right.\right\}$$

Théorème 6.1.
Si $(X^n,Y^n) \sim \prod_{i\in[1,n]} p(x_i,y_i)$, *alors* $\forall\epsilon>0, \exists N\in\mathbb{N}, \forall n\geq N$*:*

$$\begin{array}{rcl} \mathbb{P}\left((X^n,Y^n)\in\tilde{\mathcal{A}}_\epsilon^n\right) & \underset{n\to+\infty}{\longrightarrow} & 1 \\ |\mathcal{A}_\epsilon^n| & \leq & 2^{n(H(X,Y)+\epsilon)} \end{array}$$

Si $\left(\bar{X}^n,\bar{Y}^n\right) \sim \prod_{i\in[1,n]} p(x_i)p(y_i)$ *alors* $\forall\epsilon>0, \exists N\in\mathbb{N}, \forall n\geq N$*:*

$$\begin{array}{rcl} \mathbb{P}\left(\left(\bar{X}^n,\bar{Y}^n\right)\in\tilde{\mathcal{A}}_\epsilon^n\right) & \leq & 2^{-n(I(X,Y)-3\epsilon)} \\ \mathbb{P}\left(\left(\bar{X}^n,\bar{Y}^n\right)\in\tilde{\mathcal{A}}_\epsilon^n\right) & \geq & (1-\epsilon)\,2^{-n(I(X,Y)+3\epsilon)} \end{array}$$

Preuve:

$$\tilde{\mathcal{A}}_\epsilon^n = \mathcal{A}_1\cap\mathcal{A}_2\cap\mathcal{A}_3 \quad \left\{\begin{array}{c} \mathbb{P}(\mathcal{A}_1) \geq 1-{}^\epsilon\!/_3 \\ \mathbb{P}(\mathcal{A}_2) \geq 1-{}^\epsilon\!/_3 \\ \mathbb{P}(\mathcal{A}_3) \geq 1-{}^\epsilon\!/_3 \end{array}\right.$$

$$
\begin{aligned}
\mathbb{P}\left(\tilde{\mathcal{A}}_{\epsilon}^{n}\right) &= \mathbb{P}\left(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3\right) \\
&= 1 - \mathbb{P}\left(\left(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3\right)^c\right) \\
&= 1 - \mathbb{P}\left(\mathcal{A}_1^c \cup \mathcal{A}_2^c \cup \mathcal{A}_3^c\right) \\
&\geq 1 - \epsilon
\end{aligned}
$$

$$
\begin{aligned}
1 &= \sum\nolimits_{(x^n, y^n) \in \mathcal{X} \times \mathcal{Y}} p(x^n, y^n) \\
&\geq \sum\nolimits_{(x^n, y^n) \in \mathcal{A}_n^{\epsilon}} p(x^n, y^n) \\
&\geq 2^{-n(H(X,Y)+\epsilon)} |\tilde{\mathcal{A}}_n^{\epsilon}|
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{P}\left(\left(\bar{X}^n, \bar{Y}^n\right) \in \mathcal{A}_{\epsilon}^{n}\right) &= \sum\nolimits_{(x^n, y^n) \in \mathcal{A}_n^{\epsilon}} p(x^n) p(y^n) \\
&\leq \sum 2^{-n(H(X)-\epsilon)} 2^{-n(H(Y)-\epsilon)} \\
&\leq 2^{n(H(X,Y)-H(X)-H(Y)+3\epsilon)} \\
&= 2^{n(I(X,Y)+3\epsilon)}
\end{aligned}
$$