ACCQ 202 - Information theory

1 Source coding

Soit X une variable aléatoire discrète d'alphabet $\mathcal{X}$ et de fonction de probabilité p telle que $\forall x \in \mathcal{X}, p(x) = \mathbf{P}(X = x)$. On note $p(x)$ plutôt que $p_X(x)$ par commodité, mais par $p(x)$ et $p(y)$ on fait référence à deux fonctions de probabilité distinctes.

Def. X est une **source d'information** si $|\mathcal{X}| < \infty$ et on note $\forall i \in [\![1\,;|\mathcal{X}|]\!], p_i = p(x_i) = \mathbf{P}(X = x_i)$.

Def. **Code** pour une source X : $\mathcal{C}\colon \mathcal{X} \to \{0,1\}^*$.

Def. **Longueur moyenne** d'un code $\mathcal{C}$: $\mathcal{L}(\mathcal{C}) = \sum_i p_i l_i$ avec l_i la longueur du i$^{\text{e}}$ mot codé.

Def. Un code est **non singulier** si $\forall x_i \neq x_j, \mathcal{C}(x_i) \neq \mathcal{C}(x_j)$.

Def. L'extension d'un code $\mathcal{C}$ est $\forall n, \forall x_1, , \dots, x_n, \mathcal{C}(x_1, \dots, x_n) \triangleq \mathcal{C}(x_1) * \mathcal{C}(x_2) \cdots * \mathcal{C}(x_n)$.

Def. Un code est **à décodage unique** si son extension est non singulière.

Def. Un code est dit **instantané** si aucun mot code n'est le préfixe d'un autre. On dit alors qu'il s'auto-ponctue car on peut décoder en temps réel, symbole par symbole.

Th (Inégalité de Kraft). *Soit $\mathcal{C}$ un code instantané avec longueurs (l_i). Alors $\sum_i l_i \leqslant 1$. Inversement, soit (l_i) une famille de longueurs. Si elle satisfait l'inégalité de Kraft alors il existe un code à décodage unique avec ces longueurs.*

Th (de **McMillan**). *Le théorème précédent reste valable si l'on remplace décodage instantané par décodage unique.*

Cor. $\min_{\mathcal{C} \text{ à décodage unique}} \mathcal{L}(\mathcal{C}) = \min_{\mathcal{C} \text{ à décodage instantané}} \mathcal{L}(\mathcal{C})$.

Th (Borne entropique). *Pour tout $\mathcal{C}$ à décodage unique, $\mathcal{L}(\mathcal{C}) \geqslant H(X)$, où $H(X) = -\sum_i p_i \log_2(p_i)$ est l'entropie de la source, avec égalité si et seulement si $\forall i, p_i = 2^{-l_i}$.*

Th (Inégalité de Jensen). *Si f est convexe, alors $\mathbf{E}(f(X)) \geqslant f(\mathbf{E}(X))$. Si la convexité est stricte alors $(\mathbf{E}(f(X)) \geqslant f(\mathbf{E}(X))) \iff (f$ est constante$)$.*

Def. La **divergence de Kullback-Leibler**, ou entropie relative, de deux probabilités P et Q est définie par $D_{KL}(P\|Q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right)$.

C'est une mesure de dissimilarité entre les deux distributions de probabilités.

Cor. *On a $D_{KL}(P\|Q) \geqslant 0$ avec égalité si et seulement si $\forall i, p_i = q_i$.*

Code de Shannon

On construit un code de Shannon en définissant les longueurs selon $l_i = \left\lceil \log\left(\frac{1}{p}\right) \right\rceil$, qui satisfait l'inégalité de Kraft et peut donc être utilisé pour produire un code instantané.

Prop. *Soit $\mathcal{C}$ un code de Shannon pour X. Alors $H(x) \leqslant \mathcal{L}(\mathcal{C}) \leqslant H(X) + 1$.*

Codage de Huffman

Pour construire un codage de Huffman on ordonne l'ensemble des p_i, puis l'on construit itérativement de nouveaux ensembles de probabilités en sommant à chaque étape les deux plus faibles. On repart ensuite à l'inverse : à partir de la dernière probabilité, égale à 1, on va re-diviser les probabilités de sorte à construire un arbre dont les feuilles correspondront aux p_i. La profondeur de chaque feuille i s'identifie alors à l_i.

Th. *Un code de Huffman minimise $\mathcal{L}(\mathcal{C})$.*

2 Entropie et questionnement

On remarque que $\mathcal{L}(\mathcal{C})$ s'identifie au nombre moyen de questions à poser pour identifier une valeur $X \in \mathcal{X}$.

Def (...).

Th. *On a $0 \leqslant H(X) \leqslant \log(|\mathcal{X}|)$.*

Def. Soit $(X,Y) \sim p(x,y)$. On a $H(X,Y) = -\sum_{x,y} p(x,y)\log(p(x,y)) = -\mathbf{E}_{p(x,y)}\left(\log(p(X,Y))\right)$. Et pour des v.a. $X_1, \dots, X_n$ il vient $H(X_1, \dots, X_n) = -\mathbf{E}_{p(x_1,\dots,x_n)}\left(\log(p(X_1, \dots, X_n))\right)$.

Def (Entropie conditionnelle). $H(Y \mid X) = \sum_x p(x) H(Y \mid X = x) = -\sum_{x,y} p(x,y)\log(p(y \mid x)) = -\mathbf{E}[\log(p(Y \mid X))]$

Th (*Chain rule*). $H(X_1, \dots, X_n) = \sum_{i=1}^n H(X_i \mid X^{i-1})$ *où* $X^i \triangleq X_1, \dots, X_i$.

Prop. *L'information mutuelle $I(X;Y) = \sum_{x,y} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right)$ vérifie*

- $I(X;Y) = H(X) + H(Y) - H(X,Y)$
- $I(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) = I(Y;X)$
- $I(X;X) = H(X)$
- $I(X;Y) = D_{KL}(p_{X,Y} \| p_X \cdot p_Y)$

- $I(X;Y) = 0 \iff X \perp\!\!\!\perp Y$
- $H(Y \mid X) \leqslant H(Y)$
- $H(X^n) \leqslant \sum_{i=1}^n H(X_i)$
- $H(X)$ *est concave en* p_X
- $H(f(X)) \leqslant H(x)$ *pour toute fonction* f *déterministe.*

Def. On définit $H(X;Y \mid Z) = \sum_{x,y,z} p(x,y,z) \log\left(\frac{p(x,y|z)}{p(x|z)p(y|z)}\right) = H(X \mid Z) - H(X \mid Y, Z)$.

Th. $I(X_1, \ldots, X_n; Y) = \sum_{i=1}^n I(X_i; Y \mid X^{i-1})$.

Def. Soit $X_1, \ldots, X_n$ i.i.d.. On appelle $A_\varepsilon^n = \left\{x^n \mid 2^{-n(H(x)+\varepsilon)} \leqslant p(x^n) \leqslant 2^{-n(H(x)-\varepsilon)}\right\}$ **ensemble typique**.

Th.
- *Pour* n *suffisamment grand,* $\mathbf{P}(A_\varepsilon^n) \geqslant 1 - \varepsilon$.
- $(1-\varepsilon)2^{-n(H(x)-\varepsilon)} \leqslant |A_\varepsilon^n| \leqslant 2^{-n(H(x)+\varepsilon)}$