Théorie de l'information – TD1 – Corrigé

Remarques générales

- Veillez à soigner la présentation et la rédaction. La clarté et la lisibilité des raisonnements sont primordiales en mathématiques.
- On n'utilise pas les symboles $\Leftarrow$ et $\Rightarrow$ au milieu d'une phrase avec des mots français. Le symbole $\Leftrightarrow$ s'utilise uniquement dans les raisonnements par équivalence : il n'y avait aucune raison de l'utiliser dans ce TD.
- Attention aux lettres muettes, variables etc. Si x est un réel fixé, l'expression $(1/(1-x))'$ n'a pas de sens.

Exercice 5 (Divergence)**.** *Soient p, q des distributions de probabilité sur un même ensemble fini S. On définit leur divergence de Kullback-Leibler par*

$$D_{\mathrm{KL}}(p\|q) = \sum_{\substack{x\in S \\ p(x)>0}} p(x)\log_2\left(\frac{p(x)}{q(x)}\right) \in [0,\infty]$$

où l'on pose, pour tout $x > 0$, $x\log_2(x/0) = \infty$.

1. *Donner un exemple montrant qu'en général, $D_{\mathrm{KL}}(p||q)$ et $D_{\mathrm{KL}}(q||p)$ ne sont pas égales.*

   Solution *Sur un ensemble à deux éléments, considérons les distributions de probabilité $p = (1/2, 1/2)$ et $q = (1/2, 3/4)$. Alors*

$$D_{\mathrm{KL}}(p\|q) = \tfrac{1}{2}\log_2\left(\tfrac{1/2}{1/4}\right) + \tfrac{1}{2}\log_2\left(\tfrac{1/2}{3/4}\right) = 1 - \tfrac{1}{2}\log_2(3) \tag{1}$$

$$D_{\mathrm{KL}}(q\|p) = \tfrac{1}{4}\log_2\left(\tfrac{1/4}{1/2}\right) + \tfrac{3}{4}\log_2\left(\tfrac{3/4}{1/2}\right) = -1 + \tfrac{3}{4}\log_2(3) \tag{2}$$

   et $D_{\mathrm{KL}}(p\|q) \neq D_{\mathrm{KL}}(q\|p)$ car $\log_2(3)$ n'est pas rationnel.

2. *En utilisant l'inégalité $\ln(x) \leqslant x - 1$ valable pour tout $x > 0$, montrer que $D_{\mathrm{KL}}(p\|q) \geqslant 0$.*

   Solution *Soit $x \in S$ tel que $p(x) > 0$ et $q(x) > 0$. Alors*

$$\log_2\left(\frac{q(x)}{p(x)}\right) \leqslant \frac{1}{\ln(2)}\left(\frac{q(x)}{p(x)} - 1\right) \tag{3}$$

   et en multipliant par $-p(x)$, on obtient :

$$p(x)\log_2\left(\frac{p(x)}{q(x)}\right) \geqslant \frac{1}{\ln(2)}(p(x) - q(x)). \tag{4}$$

Notons que cette inégalité est également valable pour tout $x \in S$ tel que $p(x) > 0$ et $q(x) = 0$. Alors

$$D_{\mathrm{KL}}(p\|q) \geqslant \frac{1}{\ln(2)} \sum_{\substack{x \in S \\ p(x)>0}} (p(x) - q(x)) \tag{5}$$

$$= \frac{1}{\ln(2)} \left(1 - \sum_{\substack{x \in S \\ p(x)>0}} q(x) \right) \geqslant 0. \tag{6}$$

3. *Montrer que $D_{\mathrm{KL}}(p\|q) = 0$ si et seulement si $p = q$.*

Solution $\Longrightarrow$ *Si $p = q$ alors pour tout $x \in S$ tel que $p(x) > 0$, on a $q(x) > 0$ et $\log_2(\frac{p(x)}{q(x)}) = 0$. Par conséquent, $D_{\mathrm{KL}}(p\|q) = 0$.*
$\Longleftarrow$ *Si $D_{\mathrm{KL}}(p\|q) = 0$, alors par (6) :*

$$\sum_{\substack{x \in S \\ p(x)>0}} q(x) = 1$$

ce qui montre en particulier que pour tout $x \in S$, $p(x) = 0$ si et seulement si $q(x) = 0$. De plus, par (5) :

$$D_{\mathrm{KL}}(p\|q) = \sum_{\substack{x \in S \\ p(x)>0}} (p(x) - q(x)).$$

Alors par (3) et (4), pour tout $x \in S$ tel que $p(x) > 0$, $\ln(\frac{p(x)}{q(x)}) = 1 - \frac{p(x)}{q(x)}$: ceci implique que $p(x) = q(x)$. On a donc montré que $p = q$.

4. *Est-ce que D_{KL} définit une distance sur l'ensemble des distributions de probabilité sur S ?*

Solution *Non, car d'après la question 1, la divergence n'est pas symétrique.*

Exercice 6 (Loi géométrique)**.** *Soit X une variable aléatoire à valeurs dans $\mathbb{N}$ suivant la loi géométrique de paramètre $p \in [0,1]$. Cette loi vérifie pour tout $n \in \mathbb{N}$: $\mathbb{P}(X = n) = p^n(1-p)$. (Attention à cette convention ! Il y a au moins 4 définitions possibles de l'expression "loi géométrique".)*

▲ *On se convainc aisément que les résultats de l'exercice précédent sont encore valables pour une variable aléatoire à valeurs dans $\mathbb{N}$.*
▲ *Il vaut mieux supposer $p \in]0,1[$, car la formule ci-dessus ne définit pas une loi de probabilité pour $p = 1$, et définit une loi concentrée en 0 si $p = 0$.*

1. *Rappeler quel type de situation est modélisé par la loi géométrique.*

   Solution *On considère des tirages successifs d'une variable de loi de Bernoulli de probabilité de succès p, numérotés 0,1,2. . . . La loi géométrique décrit le nombre de succès avant le premier échec: plus précisément, $X = n$ si et seulement si les tirage numéroté $0, 1, \dots, n-1$ sont des succès et le tirage n est un échec.*

2. *Calculer l'espérance de X. La série entière $S = \sum_{n\geqslant 0} x^k = \frac{1}{1-x} \in \mathbb{R}((x))$ a un rayon de convergence 1. Par conséquent, sa dérivée S' a également rayon de convergence 1, et*

$$S'(x) = \sum_{n\geqslant 1} nx^{n-1} = \frac{1}{(1-x)^2}.$$

   On en déduit que X admet une espérance qui vaut

$$\begin{aligned} \mathbb{E}(X) &= \sum_{n=1}^{\infty} np^n(1-p) && (7)\\ &= p(1-p)\sum_{n=1}^{\infty} np^{n-1} && (8)\\ &= \frac{p}{1-p}. && (9) \end{aligned}$$

3. *Calculer l'entropie de X.*

$$\begin{aligned} H(X) &= -\sum_{n\geqslant 0} p^n(1-p)\log_2(p^n(1-p)) && (10)\\ &= -(1-p)\left[\log_2(p)\sum_{n\geqslant 0} np^n + \log_2(1-p)\sum_{n\geqslant 0} p^n\right] && (11)\\ &= -\log_2(p)\frac{p}{(1-p)} - \log_2(1-p) && (12)\\ &= -\mathbb{E}(X)\log_2(p) + \log_2(1-p). && (13) \end{aligned}$$

4. *Soit Y une variable aléatoire à valeurs dans $\mathbb{N}$ d'espérance $\mathbb{E}(X)$. Exprimer $H(X) - H(Y)$ en fonction de la divergence de Kullback-Leibler des lois de X et Y, et en déduire que $H(Y) \leqslant H(X)$.*

Solution *Notons p_X, p_Y les lois respectives de X et Y. Par définition,*

$$
\begin{aligned}
D_{\mathrm{KL}}(p_Y \| p_X) &= \sum_{n | p_Y(n) > 0} p_Y(n) \log_2(p_Y(n)) - \sum_{n | p_Y(n) > 0} p_Y(n) \log_2(p_X(n)) \\
&= -H(Y) - \sum_{n=0}^{\infty} p_Y(n) \log_2(p^n(1-p)) \\
&= -H(Y) - \log_2(p) \sum_{n=0}^{\infty} n p_Y(n) - \log_2(1-p) && (\text{car } \textstyle\sum_n p_Y(n) = 1) \\
&= -H(Y) - \log_2(p)\mathbb{E}(Y) - \log_2(1-p) \\
&= H(X) - H(Y). && (\text{car } \mathbb{E}(Y) = \mathbb{E}(X))
\end{aligned}
$$

Par positivité de la divergence, on conclut que $H(X) \geqslant H(Y)$.