Théorie de l'information et du codage

TD n^{o}2 – Divergence de Kullback-Leibler

Soit $\mathcal{X}$ un ensemble fini, $\mathcal{P}(\mathcal{X})$ l'ensemble des mesures de probabilité sur $\mathcal{X}$.

Distances usuelles sur $\mathcal{P}(\mathcal{X})$

Il y a plusieurs manières naturelles de définir une distance sur $\mathcal{P}(\mathcal{X})$. Une première idée consiste à regarder $\mathcal{P}(\mathcal{X})$ comme un sous ensemble de $\mathbb{R}^{|\mathcal{X}|}$, et le faire ainsi hériter d'une des normes usuelles d'espace vectoriel (elles sont toutes équivalentes). Par exemple,

$$||\mu - \nu||_1 = \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)|.$$

On peut aussi voir les éléments de $\mathcal{P}(\mathcal{X})$ comme des fonctions de l'ensemble des parties de $\mathcal{X}$ dans $[0, 1]$, que l'on peut alors munir de la norme sup. C'est la distance en variation totale :

$$||\mu - \nu||_{\mathrm{VT}} = \sup_{A \subseteq \mathcal{X}} |\mu(A) - \nu(A)|.$$

Question 1. *Vérifier que pour tout $\mu, \nu \in \mathcal{P}(\mathcal{X})$, l'on a :*

$$||\mu - \nu||_{\mathrm{VT}} = \frac{1}{2}||\mu - \nu||_1.$$

Question 2. *Si μ' et ν' désignent les lois images de μ et ν par une fonction arbitraire $f \colon \mathcal{X} \to \mathcal{X}'$, comment se comparent $||\mu' - \nu'||_{\mathrm{VT}}$ et $||\mu - \nu||_{\mathrm{VT}}$? Quand y a-t'il égalité ?*

Divergence de Kullback-Leibler

En théorie de l'information comme en grandes déviations, il est en fait plus pertinent de mesurer la distance entre μ et ν par la quantité

$$D(\mu\|\nu) = \sum_{x\in\mathcal{X}} \mu(x)\log\frac{\mu(x)}{\nu(x)}.$$

C'est la *divergence de Kullback-Leibler* de μ par rapport ν.

Question 3. *Montrer que $D(\mu\|\nu) \geq 0$. Cas d'égalité ? Est-ce une distance ?*

Question 4. *Si μ' et ν' désignent les lois images de μ et ν par une fonction arbitraire $f\colon \mathcal{X} \to \mathcal{X}'$, comment se comparent $D(\mu'||\nu')$ et $D(\mu||\nu)$? Quand y a-t'il égalité ?*

Question 5. *Établir l'inégalité de Pinsker, valable pour tout $\mu, \nu \in \mathcal{P}(\mathcal{X})$:*

$$D(\mu\|\nu) \geq 2||\mu-\nu||_{\mathrm{TV}}^2.$$

Applications en théorie de l'information

Question 6. *Exprimer l'information mutuelle $I(X,Y)$ de deux variables aléatoires discrètes X et Y comme une divergence de Kullback-Leibler.*

Les codages étudiés jusqu'ici requièrent la connaissance exacte de la loi μ de la source, ce qui est irréalisable en pratique : on ne dispose que d'une certaine estimation à priori, notée ν.

Question 7. *Calculer la perte de performance induite dans le cas où ν est dyadique, puis en donner un encadrement dans le cas où ν est quelconque.*

Second principe thermodynamique

Soient $(X_n)_{n\geq 0}$ et $(Y_n)_{n\geq 0}$ deux chaînes de Markov de même matrice de transition sur $\mathcal{X}$.

Question 8. *Montrer que $D(\mathcal{L}(X_n)||\mathcal{L}(Y_n))$ décroît avec n. Qu'en déduire lorsque la chaîne admet la loi uniforme comme mesure stationnaire ?*

Question 9. *Lorsque la matrice de transition est irréductible et apériodique, retrouver que $\mathcal{L}(X_n)$ converge vers une unique loi stationnaire.*