THÉORIE DE L'INFORMATION

Antoine Chambert-Loir / Guillaume Garrigos

Examen du Vendredi 12 octobre 2018 (1 heure)

Aucune source d'information extérieure (document, notes de cours, appareil électronique) n'est autorisée.

QUESTION DE COURS

1 Soit p et q des lois de probabilité sur un ensemble (fini) A. Définir la *divergence* de la loi p par rapport à la loi q.

2 Soit X et Y des variables aléatoires discrètes (ne prenant qu'un nombre fini de valeurs). Donner la définition de l'*entropie* $\mathrm{H}(\mathrm{X})$ de X, de l'*entropie conditionnelle* $\mathrm{H}(\mathrm{X} \mid \mathrm{Y})$ et de l'*information relative* $\mathrm{I}(\mathrm{X}, \mathrm{Y})$.

EXERCICE 1

On lance successivement, de manière indépendante, une pièce déséquilibrée qui montre *pile* avec probabilité $t \in \left]0;1\right[$ et *face* avec probabilité $1-t$. Soit X la variable aléatoire donnant le nombre de lancers *pile* qu'il faut avoir vu avant d'obtenir *face* pour la première fois.

1 Calculer $\mathbf{P}(\mathrm{X}=0)$, $\mathbf{P}(\mathrm{X}=1)$, $\mathbf{P}(\mathrm{X}=2)$. Calculer $\mathbf{P}(\mathrm{X}=n)$ pour tout $n \in \mathbf{N}$.

2 Calculer l'espérance $\mathbf{E}(\mathrm{X})$. (*On pourra utiliser la relation* $\sum_{n=1}^{\infty} nt^n = t/(1-t)^2$, *valable pour* $|t|<1$.)

3 Démontrer que l'entropie $\mathrm{H}(\mathrm{X})$ de X est donnée par

$$\mathrm{H}(\mathrm{X}) = -\frac{t\log(t) + (1-t)\log(1-t)}{1-t}.$$

4 Soit m un nombre réel >0. Dans toute la suite, on choisit t de sorte que $\mathbf{E}(\mathrm{X}) = m$ et on note alors p la loi de X. (Que vaut t ?) Soit Y une variable aléatoire sur $\mathbf{N}$ telle que $\mathbf{E}(\mathrm{Y}) = m$; soit q sa loi. Démontrer que

$$\mathrm{H}(\mathrm{X}) - \mathrm{H}(\mathrm{Y}) = \mathrm{D}(q \mid p).$$

5 Démontrer que parmi toutes les lois sur $\mathbf{N}$ d'espérance m, la loi X est l'unique loi d'entropie maximale. Exprimer cette entropie en fonction de m.

EXERCICE 2

C'est l'histoire d'une étudiante qui travaille à la bibliothèque et qui, avec probabilité p, va prendre un café ; ceci fait, et avant de retourner travailler, elle peut, avec probabilité q, aller prendre l'air quelques minutes.

1 Décrire une chaîne de Markov à 3 états (qu'on pourra noter b, c, a) qui modélise cette histoire ; dessiner le graphe qui la représente.

2 Donner sa matrice de transitions. Vérifier qu'elle est stochastique.

3 Démontrer qu'elle possède une seule loi stationnaire et la déterminer.

4 Lorsque X_0 obéit à cette loi stationnaire, quel est le taux d'entropie du processus de Markov associé ?

5 On suppose que $0 < p < 1$ et $0 < q < 1$. Quel est le taux d'entropie du processus de Markov associé si X_0 obéit à la loi $\mathbf{P}(X_0 = b) = 1$?

Corrigé de l'examen du Vendredi 12 octobre 2018 (1 heure)
Sujet proposé (et corrigé) par A. Chambert-Loir

SOLUTION DE L'EXERCICE 1.

1 Les tirages étant indépendants, la probabilité d'obtenir n fois *pile* puis *face* est égale à $t^n(1-t)$.

2 Par définition de l'espérance, on a

$$\mathbf{E}(X) = \sum_{n=0}^{\infty} n\mathbf{P}(X=n) = \sum_{n=0}^{\infty} nt^n(1-t).$$

On rappelle la somme d'une série géométrique, pour $t \in]-1;1[$,

$$\sum_{n=0}^{\infty} t^n = \frac{1}{1-t}.$$

En dérivant par rapport à t, on trouve aussi

$$\sum_{n=1}^{\infty} nt^{n-1} = \frac{1}{(1-t)^2}.$$

Par suite,

$$E(X) = \frac{t}{1-t}.$$

3 Par définition de l'entropie, on a

$$\begin{aligned}
H(X) &= -\sum_{n\geqslant 0} \mathbf{P}(X=n)\log(\mathbf{P}(X=n)) \\
&= -\sum_{n\geqslant 0} t^n(1-t)\log(t^n(1-t)) \\
&= -(1-t)\log(t)\sum_{n\geqslant 0} nt^n - (1-t)\log(1-t)\sum_{n\geqslant 0} t^n \\
&= -(1-t)\log(t)\frac{t}{(1-t)^2} - (1-t)\log(1-t)\frac{1}{1-t} \\
&= -\frac{t}{1-t}\log(t) - \log(1-t) \\
&= \frac{h(t)}{1-t},
\end{aligned}$$

où $h(t) = -t\log(t) - (1-t)\log(1-t)$ est l'entropie du processus de Bernoulli de paramètre t.

4 Par définition, on a

$$H(X) - H(Y) = -\sum_{n\in\mathbf{N}} \mathbf{P}(X=n)\log(\mathbf{P}(X=n)) + \sum_{n\in\mathbf{N}} \mathbf{P}(Y=n)\log(\mathbf{P}(Y=n)).$$

Par ailleurs,

$$D(q \mid p) = \sum_{n\in\mathbf{N}} \mathbf{P}(Y=n)\log\left(\frac{\mathbf{P}(Y=n)}{\mathbf{P}(X=n)}\right).$$

Ainsi,

$$H(X) - H(Y) - D(q \mid p) = \sum_{n\in\mathbf{N}} (\mathbf{P}(Y=n) - \mathbf{P}(X=n))\log(\mathbf{P}(X=n)).$$

On a vu $\mathbf{P}(X=n) = t^n(1-t)$ pour $n \geqslant 0$, donc

$$\log(\mathbf{P}(X=n)) = n\log(t) + \log(1-t),$$

de sorte que

$$H(X) - H(Y) - D(q \mid p) = \sum_{n=0}^{\infty} n(\mathbf{P}(Y=n) - \mathbf{P}(X=n))\,\log(t) + \sum_{n=0}^{\infty} (\mathbf{P}(Y=n) - \mathbf{P}(X=n))\,\log(1-t).$$

Le premier terme vaut $(\mathbf{E}(\mathrm{Y}) - \mathbf{E}(\mathrm{X}))\log(t) = 0$, puisque X et Y ont même espérance. Le second terme vaut également 0 car $\sum_{n=0}^{\infty} \mathbf{P}(\mathrm{Y} = n) = 1 = \sum_{n=0}^{\infty} \mathbf{P}(\mathrm{X} = n)$.

5 Comme la divergence $\mathrm{D}(q \mid p)$ est positive ou nulle, on a $\mathrm{H}(\mathrm{X}) \geqslant \mathrm{H}(\mathrm{Y})$. Cette divergence n'est nulle que si $p = q$; c'est-à-dire si X et Y ont même loi.
Puisque $\mathbf{E}(\mathrm{X}) = t/(1-t) = m$, on a $t = m(1-t)$, d'où $t = m/(1+m)$. Alors, $1 - t = 1/(1+m)$ et

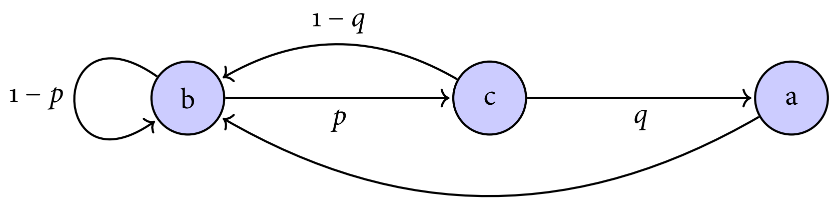
$$\begin{aligned} h(t) &= -t\log(t) - (1-t)\log(1-t) \\ &= -\frac{m}{1+m}\log(m/(1+m)) - \frac{1}{1+m}\log(1/(1+m)) \\ &= -\frac{m}{1+m}\log(m) + \frac{m}{1+m}\log(1+m) + \frac{1}{1+m}\log(1+m) \\ &= -\frac{m}{1+m}\log(m) + \log(1+m), \end{aligned}$$

de sorte que

$$\mathrm{H}(\mathrm{X}) = \frac{h(t)}{1-t} = -m\log(m) + (1+m)\log(1+m).$$

SOLUTION DE L'EXERCICE 2.

1 Les trois états représentent la position de l'étudiante, b pour bibliothèque, c pour café, et a lorsqu'elle prend l'air. Soit X_n son état au « temps » n. Selon l'histoire en tête de l'exercice, lorsque l'étudiante est à la bibliothèque, elle part prend un café avec probabilité p, on a ainsi $\mathbf{P}(\mathrm{X}_{n+1} = \mathrm{c} \mid \mathrm{X}_n = \mathrm{b}) = p$, et $\mathbf{P}(\mathrm{X}_{n+1} = \mathrm{b} \mid \mathrm{X}_n = \mathrm{b}) = 1 - p$. Lorsqu'elle prend un café, elle retourne à la bibliothèque avec probabilité $1 - q$: $\mathbf{P}(\mathrm{X}_{n+1} = \mathrm{b} \mid \mathrm{X}_n = \mathrm{c}) = 1 - q$, et part prendre l'air avec probabilité q : $\mathbf{P}(\mathrm{X}_{n+1} = \mathrm{a} \mid \mathrm{X}_n = \mathrm{c}) = q$. Lorsqu'elle a pris l'air, elle retourne travailler à la bibliothèque : $\mathbf{P}(\mathrm{X}_{n+1} = \mathrm{b} \mid \mathrm{X}_n = \mathrm{a}) = 1$.



2 La matrice de transitions de cette chaîne de Markov est donnée par

$$\mathrm{P} = \begin{array}{c} \\ \mathrm{a} \\ \mathrm{b} \\ \mathrm{c} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{a} & \mathrm{b} & \mathrm{c} \end{array} \\ \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1-p & p \\ q & 1-q & 0 \end{pmatrix} \end{array}$$

On constate que la somme des coefficients de chaque ligne est égale à 1 ; cette matrice est donc stochastique.

3 Une loi stationnaire pour cette chaîne de Markov est donnée par un vecteur-ligne $\mathrm{M} = (a\ b\ c)$ à coefficients positifs, de somme 1, tel que $\mathrm{M} \cdot \mathrm{P} = \mathrm{M}$. Explicitons d'abord ce système linéaire ; on obtient :

$$qc = a, \quad a + (1-p)b + (1-q)c = b, \quad pb = c.$$

On a alors $a = qc = pqb$ et $c = pb$, et la dernière relation

$$pqb + (1-p)b + p(1-q)b = b$$

est automatiquement vérifiée. La relation $a + b + c = 1$ s'écrit alors $(pq + 1 + p)b = 1$, d'où $b = 1/(1 + p + pq)$. Il y a donc une unique loi stationnaire, donnée par le vecteur-ligne

$$\left(\frac{pq}{1+p+pq}\ \frac{1}{1+p+pq}\ \frac{p}{1+p+pq}\right).$$

4 Un théorème du cours donne l'entropie d'une chaîne de Markov homogène stationnaire : c'est la moyenne, pour la loi stationnaire, des entropies des lignes de sa matrice de transitions. Les entropies des lignes valent :

$$\begin{aligned} \mathrm{H_a} &= 0 \\ \mathrm{H_b} &= -(1-p)\log(1-p) - p\log(p) = h(p) \\ \mathrm{H_c} &= -q\log(q) - (1-q)\log(1-q) = h(q). \end{aligned}$$

(On a noté $h(p)$ l'entropie d'un processus de Bernoulli de paramètre p.) Alors,

$$\mathrm{H} = a\mathrm{H_a} + b\mathrm{H_b} + c\mathrm{H_c} = \frac{h(p) + ph(q)}{1 + p + pq}.$$

5 D'après un théorème de cours, ce calcul reste valable pour toute loi initiale si cette chaîne de Markov est irréductible apériodique. Le cycle b → c → a → b entraîne qu'on peut atteindre tout état depuis n'importe quel autre. Cette chaîne est donc irréductible. Par ailleurs, il y a un cycle b → b de longueur 1, de sorte que les pgcd des longueurs des cycles en chaque état sont égaux à 1. Ainsi, la chaîne est apériodique.