Université Paris 8 — Année 2022–2023

M1 Mathématiques et applications, parcours ACC

Théorie de l'information – Feuille de TD 1

21/09/2022

Le corrigé de certains exercices sera disponible à l'adresse suivante :

`www.math.univ-paris13.fr/~lavauzelle/teaching/2022-23/theorie-information.html`

($\star$) exercice fondamental ($\star\star$) pour s'entraîner ($\star\star\star$) pour aller plus loin 💻 sur machine

Exercice 1. ($\star$) Applications numériques.

Question 1.– Un tirage de la loterie consiste à choisir uniformément et sans remise 5 boules parmi 49, puis, de manière indépendante, à tirer un numéro complémentaire entre 1 et 10. Calculer l'entropie du tirage (on donnera une valeur approchée).

Question 2.– Comparer les entropies des variables aléatoires représentant :

(i) la valeur de 3 dés à 4 faces,
(ii) la valeur de 2 dés à 6 faces,
(iii) la valeur d'1 dé à 12 faces.

On supposera que les dés sont équilibrés.

Question 3.– On considère la distribution conjointe suivante :

$p_{X,Y}(x,y)$	x_1	x_2
y_1	$\frac{1}{4}$	0
y_2	$\frac{1}{8}$	$\frac{5}{8}$

Comparer les valeurs de $H(X)$, $H(Y)$, $H(X \mid Y)$ et $H(Y \mid X)$.

Exercice 2. ($\star$) Rappels sur les distributions.

Soit $X : \Omega \to \mathcal{X} = \{x_1, \ldots, x_n\}$ et $Y : \Omega \to \mathcal{Y} = \{y_1, \ldots, y_m\}$ deux variables aléatoires. On note $\boldsymbol{u} = (p_X(x_1), \ldots, p_X(x_n))$ et $\boldsymbol{v} = (p_Y(y_1), \ldots, p_Y(y_m))$ leurs distributions de probabilité, représentées sous la forme de vecteurs lignes.

Alors, la distribution de la variable conjointe X,Y peut être vue comme une matrice $\boldsymbol{M}$ de taille $n \times m$, où le coefficient d'indices (i,j) est

$$M_{i,j} = p_{X,Y}(x_i, y_j)\,.$$

De même, la distribution de la variable conditionnelle $X \mid Y$ peut être vue comme une matrice $\boldsymbol{N}$ de taille $n \times m$, où

$$N_{i,j} = p_{X|Y}(x_i, y_j)\,.$$

Question 1.– Parmi les assertions suivantes, lesquelles sont exactes (justifier) ?

1. Tous les coefficients de $\boldsymbol{M}$ sont dans $[0,1]$.
2. L'égalité $\boldsymbol{M} = \boldsymbol{u}^\top \cdot \boldsymbol{v}$ est toujours vérifiée.
3. La somme de tous les coefficients de $\boldsymbol{M}$ vaut 1.
4. La somme de tous les coefficients de $\boldsymbol{N}$ vaut 1.
5. La somme des coefficients de chaque colonne de $\boldsymbol{M}$ vaut 1.
6. La somme des coefficients de chaque colonne de $\boldsymbol{N}$ vaut 1.
7. Si $M_{i,j} = 0$, alors $N_{i,j} = 0$.
8. On a $M_{i,j} \leq N_{i,j}$ pour tous i,j.

Question 2.– On suppose maintenant que $m = n$. Démontrer que le vecteur $(1, 1, \ldots, 1)^\top$ est un vecteur propre de $\boldsymbol{N}^\top$. Quelle est la valeur propre associée ?

Exercice 3. ($\star\star$) Entropie et tirages uniformes.

Soit X une variable aléatoire de distribution $(p_1, p_2, p_3, p_4) = (\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8})$.

Question 1.– Quelle est l'entropie de X ? Quelle serait l'entropie maximale d'une variable aléatoire à image dans un ensemble à 4 élements ?

Soit U la variable aléatoire uniforme sur $\{0, 1\}$. On considère l'algorithme suivant :

1. Tirer $u_1 \leftarrow U$. Si $u_1 = 0$, retourner $x = 1$.
2. Sinon, tirer un nouvel élément $u_2 \leftarrow U$. Si $u_2 = 0$, retourner $x = 2$.
3. Sinon, tirer un nouvel élément $u_3 \leftarrow U$. Si $u_3 = 0$, retourner $x = 3$. Sinon, retourner $x = 4$.

On suppose que les tirages successifs de U sont indépendants.

Question 2.– Démontrer que l'algorithme retourne la valeur i avec probabilité p_i, où p_i est défini plus haut.

Question 3.– Quelle est le nombre moyen de tirages de U lors d'une exécution de l'algorithme ? Comparer avec l'entropie de X.

Exercice 4. ($\star\star$) Entropie conditionnelle.

On considère une rencontre sportive en deux manches gagnantes, entre deux adversaires A et B. On suppose que, pour toutes les manches de la rencontre, l'évènement « A gagne la manche » suit une loi de Bernoulli de paramètre $0 < t < 1$. On suppose également que ces variables sont indépendantes.

On note :

— Ω l'ensemble des scores possibles de la rencontre (qu'on peut écrire sous la forme de chaînes de 2 ou 3 caractères, comme $'aa'$ pour signifier que A a gagné la rencontre en deux manches), et p sa loi de probabilité associée,

— $\mathcal{X} = \{A, B\}$ l'ensemble des gagnant·e·s possibles de la rencontre, et X la variable associée,

— $\mathcal{N} = \{2, 3\}$ l'ensemble correspondant au nombre de manches du match, et N la variable associée,

— $\mathcal{Y} = \{a^*, b^*\}$ l'ensemble représentant le·a gagnant·e de la première manche, et Y la variable associée (a^* signifie que A a gagné la première manche).

Question 1.– Donner la distribution de probabilités de p, ainsi que des lois des variables X, Y et N.

Question 2.– Calculer et comparer les entropies $H(X)$, $H(Y)$ et $H(X,Y)$.

Question 3.– Rappeler comment se comparent génériquement $H(N)$ et $H(N \mid Y)$. Calculer $H(N)$, $H(N\,|\,Y = a^*)$ et $H(N\,|\,Y = b^*)$. Commenter.

Exercice 5. ($\star\star\star$) Entropie et probabilité maximale.

Soit X une variable aléatoire de distribution $(p_1, \ldots, p_n)$. On suppose que tous les p_i sont non-nuls et on note $p^* := \max\{p_i\}_{1\leq i\leq n}$.

Question 1.– Montrer que $H(X) \geq -\log p^*$.

Question 2.– Vérifier que cela ne contredit pas la borne $H(X) \leq \log n$ vue en cours.

Question 3.– Démontrer que, pour tous $a, b \geq 0$ avec $a + b > 0$, on a :

$$-(a+b)\log(a+b) \leq -a\log a - b\log b \leq -(a+b)\log\frac{a+b}{2}\,.$$

Démontrer également que la première égalité se produit si et seulement si $a = 0$ ou $b = 0$, et que la seconde se produit si et seulement si $a = b$.

Question 4.– Grâce à la question précédente, montrer que $H(X) \geq h_2(p^*)$, où $h_2(t) := -t\log_2 t - (1-t)\log_2(1-t)$ est la fonction d'entropie binaire.

Question 5.– En déduire que $H(X) \geq 2(1 - p^*)$.