Université Paris 8 — Année 2022–2023

M1 Mathématiques et applications, parcours ACC

Théorie de l'information – Feuille de TD 8

23/11/2022

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. 💻 ($\star\star$) Implantation de l'algorithme LZW.

Question 1.– Implanter l'algorithme de codage LZW vu dans le cours. L'algorithme prendra en entrée le message à coder, ainsi que l'alphabet de la source.

(si besoin, on pourra commencer par supposer que l'alphabet est $\{0,1\}$*)*

Question 2.– Tester l'algorithme sur les instances et exemples présentés en cours.

Question 3.– Implanter une source $\boldsymbol{X}$ sans mémoire sur $\mathcal{X} = \{0,1\}$ telle que pour tout $n \in \mathbb{N}$, on a $p(X_n = 1) = 0.1$. Puis, calculer la longueur moyenne du codage d'une suite N bits par LZW (avec N grand), et comparer avec $\overline{H}(\boldsymbol{X})$.

Question 4.– Implanter une source binaire $\boldsymbol{X}$ markovienne (pas nécessairement stationnaire) de son choix. Puis, calculer la longueur moyenne du codage d'une suite N bits par LZW (avec N grand), et comparer avec $\overline{H}(\boldsymbol{X})$.

Exercice 2. ($\star\star$) Entropie et codage par plage.

Soit $\boldsymbol{X} = (X_1, \dots, X_N)$ une séquence **finie** de $N \geq 1$ variables aléatoires binaires. On définit $\boldsymbol{R} = (R_1 \dots, R_M)$, où les variables R_m sont à valeurs sur $\mathbb{N}^+$, comme la séquence des longueurs de plages de symboles identiques dans $\boldsymbol{X}$. Notons que $M \leq N$, et que la valeur de M dépend de la réalisation des X_n. Par exemple, si $N = 13$ et si $\boldsymbol{X}$ se réalise comme (1100010000111), alors la variable $\boldsymbol{R}$ vaut $(2,3,1,4,3)$ et $M = 5$.

On note $H(\boldsymbol{X})$ l'entropie de la variable conjointe des $(X_1, \dots, X_N)$. De même, $H(\boldsymbol{R})$ est l'entropie de la variable conjointe des $(R_1, \dots, R_M)$.

Question 1.– Démontrer que $H(\boldsymbol{R}, X_1) = H(\boldsymbol{X})$.

Question 2.– En déduire que $H(\boldsymbol{X}) - H(X_1) \leq H(\boldsymbol{R}) \leq H(\boldsymbol{X})$.

<u>Exercice 3. ($\star\star$) Longueurs de plages.</u>

Soit $\boldsymbol{X} = (X_n)_{n\geq 1}$ un processus stochastique sans mémoire, dont les variables X_n sont binaires et uniformes. On note N la variable aléatoire comptant la longueur de la première plage de symboles identiques, c'est-à-dire telle que

$$X_{N+1} \neq X_N \quad \text{et} \quad \forall 2 \leq n \leq N,\, X_n = X_{n-1}\,.$$

On note enfin $\boldsymbol{Y} := (X_1, \ldots, X_N)$.

Question 1.– Déterminer le domaine des valeurs prises par les variables N et $\boldsymbol{Y}$?

Question 2.– Déterminer la loi de N, la loi de $\boldsymbol{Y}$ conditionnée à N puis en déduire la loi de $\boldsymbol{Y}$.

Question 3.– Démontrer que :

1. $H(\boldsymbol{Y}) = 3$,
2. $H(\boldsymbol{Y}\,|\,N) = 1$ (et interpréter),
3. $I(\boldsymbol{Y}\,;N) = 2$.