École Normale Supérieure
16 avril 2015

Traitement du signal

Feuille d'exercices n$^{\text{o}}$9

Exercice 1

Soit X une variable aléatoire prenant la valeur 0 (respectivement 1) avec probabilité p (respectivement $1-p$). Soit T_ϵ^n l'ensemble typique associé pour n tirages indépendants de X.
Estimer le nombre de « 1 » contenus dans les suites de T_ϵ^n, sous l'hypothèse $p \neq 1/2$.

Exercice 2

On considère une variable aléatoire discrète, à valeurs dans {M, P, R, U, Y, Z}, chaque symbole ayant la probabilité suivante :

M	P	R	U	Y	Z
1/16	1/16	1/16	1/4	1/2	1/16

1. Calculer l'entropie de cette distribution.

2. a) Calculer le code de Huffman associé.
b) Pour ce code, quel est le nombre moyen de bits par lettre ?

3. Pour ce code, montrer que, quelle que soit $\epsilon_1, ..., \epsilon_n$ une suite de 0 et de 1, il existe un mot dont le code commence par $\epsilon_1 ... \epsilon_n$.

4. a) Déterminer le code du mot YUM.
b) Si on change le premier bit de ce code, le code que l'on obtient correspond-il encore à un mot ? Si oui, lequel ?

Exercice 3 : longueur moyenne d'un code quelconque

On note $\mathcal{B}$ l'ensemble des mots de longueur finie composés uniquement des symboles 0 et 1. On note $l : \mathcal{B} \to \mathbb{N}$ la fonction qui, à un mot, associe sa longueur.
Soit X une variable aléatoire à valeurs dans $\mathcal{E} = \{1, ..., n\}$. Soit $\phi : \mathcal{E} \to \mathcal{B}$ injective.

1. Montrer qu'il existe une permutation $\pi : \mathcal{E} \to \mathcal{E}$ telle que :

$$\mathbb{E}(l(\phi(X))) \geq \sum_{k=1}^{n} P(X = \pi(k))[\log_2(k)]$$

(où la notation [.] désigne la partie entière et P la probabilité)

2. Montrer que $\mathbb{E}(l(\phi(X))) \geq H(X) - 1 - \log_2(1 + \ln(n))$.
[Indication : utiliser la concavité de $\log_2$ et le fait que $\sum_{k=1}^{n} \frac{1}{k} \leq 1 + \ln(n)$.]

3. Soit $X_1, X_2, ...$ une suite de variables aléatoires indépendantes et de même loi, à valeurs dans $\mathcal{E}$. Pour tout k, soit $\phi_k : \mathcal{E}^k \to \mathcal{B}$ injective. Montrer que, pour tout $\epsilon > 0$:

$$\frac{1}{k}\mathbb{E}(l(\phi_k(X_1, ..., X_k))) \geq H(X_1) - \epsilon \quad \text{pour tout } k \text{ assez grand}$$

Exercice 4 : codage arithmétique

Soit X une variable aléatoire à valeurs dans $\{1, ..., K\}$. On note $p_k = P(X = k)$ et on définit :

$$a_s = \sum_{k=1}^{s} p_k \quad \forall s = 0, ..., K$$

On code un mot $x_1...x_n \in \{1, ..., K\}^n$ de la manière suivante :

- On définit des suites $(y_k^{\text{inf}})_{k=0,...,n}$, $(y_k^{\text{sup}})_{k=0,...,n}$ telles que $y_0^{\text{inf}} = 0, y_0^{\text{sup}} = 1$ et :

$$y_k^{\text{inf}} = y_{k-1}^{\text{inf}} + a_{x_k-1}(y_{k-1}^{\text{sup}} - y_{k-1}^{\text{inf}}) \qquad y_k^{\text{sup}} = y_{k-1}^{\text{inf}} + a_{x_k}(y_{k-1}^{\text{sup}} - y_{k-1}^{\text{inf}})$$

- On utilise comme code le réel (ou l'un des réels) de l'intervalle $[y_n^{\text{inf}}; y_n^{\text{sup}}[$ dont le développement binaire est fini et de longueur minimale ; le code est l'écriture en binaire de ce réel.

1. Décrire un processus de décodage (en supposant n connu).

2. On suppose qu'on encode des mots dont chaque lettre est une réalisation indépendante de X. On note l_n l'espérance de la longueur du code d'un mot de n lettres. Montrer que, pour tout $\epsilon > 0$:

$$\frac{1}{n} l_n \leq H(X) + \epsilon \quad \text{pour tout } n \text{ assez grand}$$

3. [Comparaison avec Huffman] Soit X une variable aléatoire binaire valant 0 avec probabilité $1 - \epsilon$ et 1 avec probabilité ϵ.
a) Calculer l'entropie de cette distribution.
b) Calculer le code de Huffman associé à cette distribution.
c) Pour ce code, quelle est la longueur moyenne du code d'un mot de n lettres ?

Exercice 5

On veut coder une variable aléatoire X à images dans $\mathbb{N}^*$, de loi $P(X = k) = 2^{-k}$.

1. Calculer l'entropie $H(X)$.

2. Trouver un code optimal possédant la propriété du préfixe (en s'inspirant du code de Huffman).

3. On coupe les mots de plus de N bits en n'envoyant que les N premiers bits. Donner les expressions du nombre moyen de bits et de l'erreur quadratique moyenne.