Examen du cours de Théorie de l'Information et Codage

31 mai 2011.

Temps: 3h.

Certains problèmes ont des questions indépendantes. Les 6 problèmes sont chacun notés sur 20/6.

1. Problème 1:

   (a) On considère un système d'encodage ayant la contrainte suivante: chaque mot code doit commencer par un symbole $\{A, B, C\}$ et ensuite utiliser les symboles binaires $\{0, 1\}$. On a donc un code ternaire pour le premier symbole puis binaire. En adaptant le codage de Huffman donner les mots code pour une distribution de probabilité:

   $$p = \left(\frac{16}{69}, \frac{15}{69}, \frac{12}{69}, \frac{10}{69}, \frac{8}{69}, \frac{8}{69}\right).$$

   Est-ce un code uniquement décodable optimal (i.e. ayant le nombre moyen minimum de symboles)?

   (b) Soit $\mathcal{X} = \{0, 1\}$ et $\mathcal{Y} = \{1, 2, 3\}$. Calculer les capacités des canaux ayant comme alphabet d'entrée $\mathcal{X}$ et de sortie $\mathcal{Y}$ et matrices de transition repectives:

   $$Q_1 = \begin{pmatrix} p_1 & p_2 & p_3 \\ p_1 & p_2 & p_3 \end{pmatrix} \quad \text{et} \quad Q_2 = \begin{pmatrix} p_1 & p_2 & p_3 \\ p_3 & p_2 & p_1 \end{pmatrix}$$

2. Problème 2: Canal avec mémoire. On considère un canal binaire symétrique avec $Y_i = X_i \oplus Z_i$ où $\oplus$ est l'addition modulo 2 et $X_i, Y_i \in \{0, 1\}$. Les $Z_1, Z_2, \ldots$ ne sont pas forcément indépendants, mais pour tout $n \geq 1$, $(Z_1, \ldots, Z_n)$ est indépendant de $(X_1, \ldots, X_n)$.

   (a) On suppose que $\mathbb{P}(Z_i = 1) = p = 1 - \mathbb{P}(Z_i = 0)$. Soit $C = 1 - H(p)$. Montrer que

   $$\max_{\mathbb{P}_{X_1,\ldots,X_n}} I(X_1, \ldots, X_n; Y_1, \ldots, Y_n) \geq nC.$$

   (b) On suppose que $\mathbb{P}(Z_1 = 0) = \mathbb{P}(Z_1 = 1) = 1/2$ et que pour $i \geq 1$, $\mathbb{P}(Z_{i+1} \neq Z_i | Z_1, \ldots, Z_i) = q \in [0, 1]$. Justifier:

   $$\begin{aligned} I(X_1, \ldots, X_n; Y_1, \ldots, Y_n) &= H(Y_1, \ldots, Y_n) - H(Z_1, \ldots, Z_n) && (1) \\ &= H(Y_1, \ldots, Y_n) - (1 + (n-1)H(q)) && (2) \\ &\leq (n-1)(1 - H(q)). && (3) \end{aligned}$$

   Trouver la distribution des $(X_1, \ldots, X_n)$, $\mathbb{P}_{X_1,\ldots,X_n}$ qui atteint cette borne supérieure. Qu'en déduire pour le canal avec ce bruit?

3. Problème 3: On considère un canal sans mémoire avec $Y = X \oplus Z$ où $\oplus$ est l'addition modulo 2 et $X, Z \in \{0, 1\}$ sont indépendants et $\mathbb{P}(Z = 1) = 1 - \mathbb{P}(Z = 0) = \epsilon \in [0, 1/2)$.

   (a) Quelle est la capacité de ce canal ? Pour quelle distribution d'entrée est-elle atteinte ?

(b) On considère un code avec uniquement deux mots-code: $\mathbf{u} = (u_1, \ldots, u_n)$ et $\mathbf{v} = (v_1, \ldots, v_n)$. Soit d la distance de Hamming entre $\mathbf{u}$ et $\mathbf{v}$. On suppose que le décodeur choisit le mot-code le plus proche pour la distance de Hamming de la suite $\mathbf{y}$ reçue. Si $\mathbf{y}$ est équidistant de $\mathbf{u}$ et $\mathbf{v}$, on suppose que le décodeur fait une erreur. Montrer que

$$\mathbb{P}(\text{ Error }) \leq (4\epsilon(1-\epsilon))^{d/2}.$$

Rappel: la borne de Chernoff valable pour des variables Z_i i.i.d. donne, pour tout $s \geq 0$:

$$\mathbb{P}\left(\sum_{i=1}^{d} Z_i \geq d\alpha\right) \leq \left(e^{-s\alpha}\mathbb{E}[e^{sZ_1}]\right)^d.$$

(c) Si d est impaire, montrer que:

$$\mathbb{P}(\text{ Error }) \geq \binom{d}{(d+1)/2} \epsilon^{(d+1)/2}(1-\epsilon)^{(d-1)/2}.$$

(d) Si les mots-code sont tirés au hasard de manière indépendante et de telle sorte que chaque entrée du mot-code est choisie indépendemment selon la distribution trouvée en (a), quelle est la valeur moyenne de d?

4. Problème 4:

(a) On considère un système de cryptographie: P est le message à transmettre, C est le message crypté. Il est obtenu à partir de P et d'une clé aléatoire K, donc il existe une fonction déterministe telle que $C = f(P, K)$. Le récepteur a accès à la clé K et au message crypté C et doit pouvoir retrouver le message original. Il existe donc une fonction déterministe telle que $g(C, K) = P$. Pour un tiers qui ne possède pas la clé, le message crypté ne doit fournir aucune information sur le message à transmettre, donc P et C sont indépendants. Montrer que $H(K) \geq H(P)$ et en déduire que la clé K doit être constituée d'un certain nombre de bits au moins, que l'on exprimera en fonction de $H(P)$.

(b) Paradoxe de St Petersbourg: on considère le jeu suivant: pour un prix d'entrée de c euros, un joueur recoit 2^k euros avec probabilité 2^{-k}. Certains disent qu'un prix 'équitable' à payer pour jouer ce jeu est $c = \infty$. Pourquoi? On suppose maintenant que le joueur peut acheter une part du jeu. Par exemple, s'il investit $c/2$ euros dans le jeu, il recoit $X/2$ avec $\mathbb{P}(X = 2^k) = 2^{-k}$. On suppose les $X_1, X_2, \ldots$ i.i.d. et que le joueur est obligé de réinvestir la totalité de sa fortune à chaque étape. Soit $F_n(c)$ sa fortune au temps n. On suppose que $F_0(c) = 1 < c$. Montrer qu'il existe c^* tel que pour $c < c^*$ sa fortune tend vers l'infini et si $c > c^*$ elle tend vers 0. Calculer la valeur du prix 'équitable' c^*.

5. Problème 5:

(a) Montrer que le nombre moyen de 1 par mot-code (moyenné sur tous les mots-code) dans un code binaire linéaire de longueur N est au plus $N/2$.

(b) En déduire que la distance minimale d'un code binaire linéaire de longueur N ayant 2^L mots-code satisfait:

$$d_{\min} \leq \frac{2^{L-1}N}{2^L - 1}.$$

(c) Montrer que cette inégalité est valide pour tout code binaire (non nécessairement linéaire).

6. Problème 6: La classe des codes de Justesen est la seule classe de codes linéaires binaires explicitement connue contenant des codes $(\mathcal{C}_i)_{i\geq 1}$ dont les paramètres $(n_i, k_i, d_i)_{i\geq 1}$ satisfont:

$$n_i \xrightarrow[i\to\infty]{} +\infty, \qquad \liminf_{i\to\infty} \frac{k_i}{n_i} > 0 \qquad \text{et} \qquad \liminf_{i\to\infty} \frac{d_i}{n_i} > 0.$$

Notons $\mathcal{P}_r$ l'ensemble des polynômes de degré au plus r sur le corps fini $\mathbb{F}_{q^m}$ et soit $L = (\alpha_1, \ldots, \alpha_n)$ une famille de $n > r$ éléments 2 à 2 distincts de $\mathbb{F}_{q^m}$.

(a) À chaque $f \in \mathcal{P}_r$, on associe le vecteur de ses évaluations sur L, i.e. le mot-code

$$c(f) = (f(\alpha_1), \ldots, f(\alpha_n)) \in \mathbb{F}_{q^m}^n,$$

et l'on note $\mathcal{C}_{L,r}$ l'ensemble des vecteurs ainsi obtenus. Vérifier que $\mathcal{C}_{L,r}$ est un code linéaire, puis calculer sa dimension et sa distance.

(b) Montrer que cette famille de codes généralise celle des codes de Reed-Solomon.

(c) À chaque $f \in \mathcal{P}_r$, on associe à présent le mot-code

$$\widetilde{c}(f) = (f(\alpha_1), \alpha_1 f(\alpha_1), \ldots, f(\alpha_n), \alpha_n f(\alpha_n)) \in \mathbb{F}_{q^m}^{2n},$$

et l'on note $\widetilde{\mathcal{C}}_{L,r}$ l'ensemble des vecteurs ainsi obtenus. Quelles sont les longueur et dimension de $\widetilde{\mathcal{C}}_{L,r}$? Montrer qu'un mot-code non-nul contient toujours au moins $n - r$ couples $(f(\alpha_i), \alpha_i f(\alpha_i))$ 2 à 2 distincts.

(d) Expliquer comment transformer simplement un code linéaire q^m-aire de dimension k et de longueur n en un code linéaire $q-$aire de dimension mk et de longueur mn.

(e) Pour tout $m \geq 1$ et tout $\varrho \in [0, 1)$, on appelle code de Justesen d'ordre m et de paramètre ϱ le code binaire obtenu en appliquant la transformation de la question (d) au code de la question (c) avec $q = 2$, $r = \lfloor 2^m \varrho \rfloor$ et $L = \mathbb{F}_{2^m}$. Montrer que la classe des codes de Justesen vérifie bien la propriété annoncée.

Indication: on pourra démontrer le résultat suivant: si $x_1, \ldots, x_M$ sont des mots binaires 2 à 2 distincts de longueur N, alors la proportion totale de 1, $\gamma = \frac{1}{MN} \sum_{i=1}^{M} w(x_i)$, vérifie:

$$NH(\gamma) \geq \log_2(M).$$