PARTIE II (6 POINTS)

Entropies

Soit p_1, p_2, $\ldots$, p_n une distribution de probabilité. La formule d'entropie de Shannon

$$H(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log(p_i)$$

est une mesure d'incertidude obtenue en imposant trois conditions "naturelles" à la notion d'incertitude. D'autres fonctions ont été introduites pour mesurer le degré d'incertitude. Notament, l'entropie minimale, et l'entropie de Renyi qui sont définies par

$$H_\infty(p_1, p_2, \ldots, p_n) = -\log(\max_{1 \leq i \leq n} p_i) \qquad \text{(entropie minimale)}$$

$$H_2(p_1, p_2, \ldots, p_n) = -\log(\sum_{i=1}^{n} p_i^2) \qquad \text{(entropie de Renyi)}$$

(0) Rappeler les propriétés caractéristiques de l'entropie de Shannon.
(1) Comparer deux à deux les fonctions H, H_∞ et H_2.
(2) Quelles sont les propriétés caractéristiques de l'entropie de Shannon qui sont contredites par l'entropie minimale ?
(3) Idem, avec l'entropie de Renyi.
(4) Soit n un entier positif. Calculer les entropies de Renyi, Shannon et minimale de la distribution $p_1 = p_2 = \ldots = p_{n-1}$ et $p_n = \frac{127}{128}$.
(5) Dans la situation précédente, à votre avis, quelle fonction d'entropie effectue la bonne mesure d'incertitude ?

Shannon vs Huffman

Soient α, β et γ trois nombres réels tels que

$$3\alpha + \beta + \gamma = 1, \quad \gamma > \alpha + \beta > 2\alpha > \beta > \alpha > 0$$

Une source est composé de cinq symboles a, b, c, d et e. La distribution de probabilité des symboles est de la forme $p_a = p_b = p_c = \alpha$,$p_d = \beta$, $p_e = \gamma$. On encode les symboles de la source avec un code préfixe binaire.

(1) Donner une formule du nombre de bits moyen par symbole.
(2) On utilise le code de Shannon-Fano. Donner une condition nécessaire et suffisante sur γ pour que les mots de codes soient de longeur au moins deux.
(3) Calculer le code de Shannon-Fano pour $\alpha = \frac{1}{8}$, $\beta = \frac{3}{16}$ et $\gamma = \frac{7}{16}$.
(4) Dans ce cas, le code de Shannon-Fano est-il optimal?