Oulun yliopisto
Sovelletun matematiikan ja tilastotieteen yksikkö
Informaatioteoria 2016
Harjoitus 6

1. Olkoon koodille C

$$L_{100}(C) = \mathbb{E}\left[\ell(X)^{100}\right] = \sum_{x\in\mathcal{X}} p(x)\ell(x)^{100},$$

missä tavalliseen tapaan $\ell(x)$ on symbolin x koodisanan pituus. Määritellään

$$\begin{aligned} L_{100}^{(1)} &= \min\{L_{100}(C) \mid C \text{ välitön}\}, \\ L_{100}^{(2)} &= \min\{L_{100}(C) \mid C \text{ yksikäsitteisesti dekoodattavissa}\}. \end{aligned}$$

Mitä voit sanoa suureiden $L_{100}^{(1)}$ ja $L_{100}^{(2)}$ keskinäisestä suuruudesta?

2. Olkoon informaatiolähteen X viestiaakkosto $\mathcal{X} = \{a, b, c, d, e, f\}$. Oletetaan, että X:lle on konstruoitu yksikäsitteisesti dekoodattavissa oleva koodi koodiaakkostona $\mathcal{D} = \{0, 1, \ldots, D-1\}$ ja että symbolien a, b, c, d, e, f koodisanojen pituudet ovat 1,1,2,3,2 ja 3. Mitä D:n täytyy vähintään olla?

3. Mitkä allaolevista voivat olla koodisanojen joukkoja Huffmanin koodille?

(a) {0, 10, 11}

(b) {00, 01, 10, 110}

(c) {01, 10}

4. Konstruoi binäärinen Huffmanin koodi viestiaakkostolle $\mathcal{X} = \{a, b, c, d, e, f\}$, kun $p(a) = 0.4, p(b) = 0.3, p(c) = p(d) = 0.1, p(e) = 0.06$ ja $p(f) = 0.04$. Mikä on koodin keskimääräinen pituus? Entä lähteen entropia $H(X)$?

5. Tarkastellaan satunnaismuuttujaa X, jolle $\mathcal{X} = \{1, 2, \ldots, m\}$ ja $p_1 \geq p_2 \geq \cdots \geq p_m > 0$, missä $p_i = p(i)$, $i = 1, 2, \ldots, m$. Oletetaan, että $m > 1$. Määritellään

$$F_i = \sum_{k=1}^{i-1} p_k,$$

$i = 1, 2, \ldots, m$ ja $F_1 \equiv 0$. Merkitään $\ell_i = \lceil \log \frac{1}{p_i} \rceil$ ja määritellään koodi C s.e.

$$C(i) = F_i\text{:n binääriesityksen } \ell_i \text{ ensimmäistä bittiä pisteen jälkeen.}$$

Siis, kun esimerkiksi $F_i = 0.875$, niin $F_i = 0.111$ binäärisesti ja jos $\ell_i = 2$, on $C(i) = 11$. Jotta koodi olisi hyvin määritelty, sovitaan vielä, että binääriesityksissä ei saa esiintyä ykkösistä koostuvaa ääretöntä “häntää”, eli esimerkiksi desimaaliluvun 0.5 binääriesityksen $0.0111\cdots$ sijaan käytetään esitystä 0.1. Osoita, että koodi C on välitön ja että sille

$$H(X) \leq L(C) < H(X) + 1.$$

Sovella tätä koodausta tilanteeseen, jossa $(p_1, p_2, p_3, p_4) = (0.5, 0.25, 0.125, 0.125)$.