

JY / Matematiikan ja tilastotieteen laitos
MATA2500 Informaatioteoria, Syksy 2019
Harjoitus 4

Tämän viikon demoissa puhutaan koodausteoriasta.

1. Olkoon X TN-jakauma $(\{a, b, c, d, e, f, g, h\}, \{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{3}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}\})$.
 - (a) Laske jakauman entropia.
 - (b) Määritä $|S_\delta|$ sekä $H_\delta(X)$ kaikilla $0 < \delta < 1$.
 - (c) Muodosta jakauman Huffman-koodi, ja laske Huffman-koodin odotuspituus.
2. Olkoon X TN-jakauma $(\{0, 1\}, \{0.8, 0.2\})$.
Muodosta jakaumista X^2 , X^3 ja X^4 niiden Huffman-koodit. Laske myöskin kunkin jakauman entropia sekä jakaumaa vastaavan Huffman-koodin odotuspituus.
3. Tutki seuraavista koodeista ovatko ne yksikäsitteisesti purettavissa taikka etuliitevapaita.
 - (a) $\{a \mapsto 00, b \mapsto 11, c \mapsto 0101, d \mapsto 111, e \mapsto 1010, f \mapsto 100100, g \mapsto 0110\}$
 - (b) $\{x \mapsto 00, y \mapsto 11, z \mapsto 1101, w \mapsto 0010\}$

Osoita, että etuliitevapaa koodi on aina yksikäsitteisesti purettavissa.

Syventävät tehtävät

Seuraavat tehtävät palautetaan kirjallisina harjoitusten pitäjälle, kurssin palautuslokeroon tai sähköpostitse. Deadline 03.12.2019 klo 08.30.

4. Tutkitaan kahden alkion tasajakaumaa, $X = \{(0, 1), (\frac{1}{2}, \frac{1}{2})\}$.
Hahmottele funktion $\delta \mapsto H_\delta(X^N)$ kuvaajaa kun $N = 1, 2, 4, 1000$.
Vihje: Normaali-jakauma.
5. Olkoon X TN-jakauma ja merkitään tämän jakauman Huffmanin koodia c_H . Osoita, että c_H on optimaalinen etuliitevapaiden täydellisten koodien joukossa, eli että kaikilla muilla etuliitevapailta koodeilla c joilla $\sum_j 2^{-\ell_i} = 1$ pätee

$$L(c_H, X) \leq L(c, X).$$

(Vihje: Jos on olemassa koodi jolla on pienempi odotuspituus, niin löytyy optimaalinen koodi jolla on pienempi odotuspituus. Tutki askelta, jossa Huffman asettaa kahdelle vähiten todennäköiselle alkiolle a ja b yhtä pitkät koodit. Oleta, että c antaa näille eri mittaiset koodit $\ell(c(a)) < \ell(c(b))$ ja etsi alkio x jolle $\ell(c(c)) \geq \ell(c(b))$. Katso mitä koodin c odotuspituudelle käy kun vaihdat koodit $c(b)$ ja $c(x)$ keskenään.)

6. Olkoon X TN-jakauma missä $\Omega = \{a_1, \dots, a_n\}$, $p_j = \frac{1}{j}$ kaikilla j ja oletetaan että n ei ole muotoa 2^k millään $k \in \mathbb{N}$. Olkoon c optimaalinen koodi (eli Huffmanin koodi) jakaumalle X .

Merkitään $l^+ = \lceil \log_2(n) \rceil$, ja asetetaan

$$f^+ := \frac{\#\{x \in \Omega \mid \ell(x) = l^+\}}{n}.$$

- (a) Osoita, että $f^+ = 2 - (2^{l^+})/n$ ja $L(c, X) = l^+ - 1 + f^+$.
- (b) Merkitään edelleen $\Delta L := L(c, X) - H(X)$. Osoita että

$$\Delta L \leq 1 - \frac{\ln(\ln 2)}{\ln 2} - \frac{1}{\ln(2)} \approx 0.086.$$

(Vihje: a-kohdan nojalla voit kirjoittaa muutoksen ΔL parametrin n muuttujana. Korvaa n jatkuvalla muuttujalla x ja tutki derivaattaa.)

- (c) Laske ΔL tilanteessa, jossa X on yhdentoista alkion tasajakauma ja c sen Huffmanin koodi.