Tentamen Informatietheorie en Codes (kans A)

Opgave 1. *(7 punten)*
Gegeven is de kansverdeling $P = (0.5,\ 0.25,\ 0.1,\ 0.05.\ 0.05,\ 0.05)$.

(i) Bepaal de entropie $H(P)$ van P.

(ii) Laat zien dat er voor P twee echt verschillende binaire Huffman-coderingen bestaan (d.w.z. met verschillende codewoordlengtes) en geef de gemiddelde codewoordlengte L_H van deze coderingen aan.

(iii) Een *quaternaire* codering is een codering over een alfabet $\mathcal{A}$ met 4 elementen, bijvoorbeeld $\mathcal{A} = \{a, b, c, d\}$.

Bepaal een quaternaire Huffman-codering voor P en geef de gemiddelde codewoordlengte hiervan aan.

(iv) Voor een willekeurige stochast X laat zich een quaternaire codering $\mathcal{C}_4$ met gemiddelde codewoordlengte $L_{\mathcal{C}_4}(X)$ omzetten in een binaire codering $\mathcal{C}_2$ met gemiddelde codewoordlengte $L_{\mathcal{C}_2}(X) = 2L_{\mathcal{C}_4}(X)$ door de vier symbolen uit $\mathcal{A}$ te vervangen door de vier verschillende blokken van 2 bits, bijvoorbeeld $a \to 00$, $b \to 01$, $c \to 10$, $d \to 11$.

Bewijs dat er een quaternaire codering $\mathcal{C}_4$ bestaat zo dat voor de gemiddelde codewoordlengte $L_{\mathcal{C}_2}(X)$ van de hieruit verkregen binaire codering geldt dat

$$L_{\mathcal{C}_2}(X) \le L_H(X) + 1$$

voor $L_H(X)$ de gemiddelde codewoordlengte van de binaire Huffman-codering van X.

Hint: Verleng binaire codewoorden van oneven lengte tot codewoorden van even lengte en vat deze op als codewoorden van een quaternaire codering.

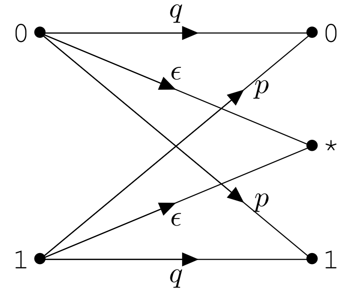
(v) Bewijs dat voor de gemiddelde codewoordlengtes L_{H_2} van de binaire en L_{H_4} van de quaternaire Huffman-codering geldt dat

$$L_{H_2}(X) \le 2L_{H_4}(X) \le L_{H_2}(X) + 1.$$

Opgave 2. *(6 punten)*
Een zeker kanaal heeft `0` en `1` als inputs en outputs, maar met een zekere kans ϵ is de output onleesbaar, dat noteren we met $\star$. De overgangsmatrix voor inputs in de volgorde `0`, `1` en outputs in de volgorde `0`, `1`, $\star$ is

$$\begin{pmatrix} 1-p-\epsilon & p & \epsilon \\ p & 1-p-\epsilon & \epsilon \end{pmatrix},$$

een schema voor dit kanaal is hieronder weergegeven, waarin $q = 1 - p - \varepsilon$.



z.o.z.

(i) Bereken voor een input verdeling $P(X) = \begin{pmatrix}\pi & \pi'\end{pmatrix}$ de output verdeling $P(Y)$ en leg uit waarom de entropie $H(Y)$ van de output verdeling maximaal is voor een uniforme verdeling op de inputs.

(ii) Bewijs dat dit kanaal capaciteit $C = (1-\epsilon)(1-\log(1-\epsilon)) + q\log q + p\log p$ heeft.

Hint: Dit is een *uniform dispersief* kanaal (d.w.z. iedere rij van de overgangsmatrix is een permutatie van de eerste rij) en we weten dat hiervoor $H(Y|X) = H(a_1, \dots, a_n)$ voor $a_1, \dots, a_n$ de eerste rij van de overgangsmatrix.

(iii) Geef de capaciteit van het kanaal aan voor de speciale gevallen

(a) $\epsilon = 0$;

(b) $p = 0$.

Kan je deze speciale gevallen met bekende kanalen identificeren?

Opgave 3. *(8 punten)*

Zij $\mathcal{C}$ de lineaire code over $\mathbb{F}_2 = \{0, 1\}$ met voortbrengermatrix $G = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$.

(i) Laat zien dat $\mathcal{C}$ minimum afstand $d = 3$ heeft.

(ii) Bepaal een check matrix H voor $\mathcal{C}$.

(iii) Voor de decodering gebruik je de methode van syndroom decodering (met restklassenrepresentanten van minimaal gewicht). Wegens $d = 3$ is het duidelijk dat je een enkele fout altijd kunt verbeteren.

(a) Hoeveel verschillende foutpatronen met 2 fouten kan je met deze code verbeteren?

(b) Je verstuurt de codewoorden van $\mathcal{C}$ via een binair symmetrisch kanaal (BSC) met omklapkans p.
Wat is de kans p_E op een decoderingsfout bij syndroom decodering?
Bepaal p_E concreet voor $p = 0.02 = 2\%$.

(iv) Omdat $\mathcal{C}$ drie informatiebits met drie checkbits aanvult, kan je middels deze code $2^3 = 8$ verschillende letters coderen. Je gebruikt de volgende idenficaties tussen blokken van drie informatiebits en letters:

$$\textit{spatie} \mathrel{\widehat{=}} 000,\ \mathtt{A} \mathrel{\widehat{=}} 100,\ \mathtt{D} \mathrel{\widehat{=}} 010,\ \mathtt{E} \mathrel{\widehat{=}} 001,\ \mathtt{K} \mathrel{\widehat{=}} 110,\ \mathtt{L} \mathrel{\widehat{=}} 101,\ \mathtt{N} \mathrel{\widehat{=}} 011,\ \mathtt{O} \mathrel{\widehat{=}} 111.$$

Bijvoorbeeld codeer je A als 100110 en K als 110011.

Je ontvangt de output

$$110001 \quad 111000 \quad 110101 \quad 100111 \quad 100101 \quad 110110.$$

Decodeer de oorspronkelijke boodschap (in letters) door eventuele transmissiefouten te verbeteren.

Succes ermee!