Informations- och kodningsteori: Tenta 2004-06-01

Skrivtid: 8-13
Hjälpmedel:
Miniräknare, läroboken, utdelade stenciler, egna anteckningar.

• 1. Betrakta följderna $(2,3,3,3,4,5,5,5,5)$, $(1,2,3,3,3)$ och $(1,2,3,4,5)$. Vilka av dessa kan beskriva ordlängderna till någon binär prefixfri kod? Ange en sådan kod i varje fall då det är möjligt!

• 2. Låt $\mathcal{S}$ vara en källa med 8 källsymboler med sannolikheter $0.6, 0.1, 0.1, 0.1, 0.025, 0.025, 0.025, 0.025$. Beräkna den ternära entropin $H_3(\mathcal{S})$ och medelordlängden i en optimal ternär kod för $\mathcal{S}$. Ange alla kodorden i en sådan kod!

• 3. Låt $\mathcal{S}$ vara en källa med 4 källsymboler med sannolikheter $0.4, 0.3, 0.2, 0.1$. Beräkna den binära entropin $H_2(\mathcal{S})$. Ange, med motivering, ett positivt heltal n sådant att det finns en prefixfri binär kod $\mathcal{C}$ för källan $\mathcal{S}^n$ vars medelordlängd som kodning av $\mathcal{S}$ är < 1.85, dvs., $L(\mathcal{C})/n < 1.85$.

• 4. Låt $\mathcal{C}$ vara den binära linjära kod som har paritets-check-matris

$$H = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}$$

Ange $\mathcal{C}$'s dimension, dess hastighet (dvs "rate"), och en generatormatris för $\mathcal{C}$. Ange även $\mathcal{C}$'s minimalavstånd.

• 5. Låt $\mathcal{A}$ och $\mathcal{B}$ vara källor med alfabeten $\{a_1, a_2\}$ och $\{b_1, b_2\}$, respektive. Låt Γ vara en kanal från $\mathcal{A}$ till $\mathcal{B}$ med kanalmatris

$$\begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}$$

Låt a_1, a_2 ha sannolikheterna 0.4, 0.6. Beräkna sannolikheterna för utdata-symbolerna b_1, b_2, de binära systementropierna $H(\mathcal{A}, \mathcal{B})$, $H(\mathcal{A}|\mathcal{B})$, $H(\mathcal{B}|\mathcal{A})$, och informationen $I(\mathcal{A}, \mathcal{B})$.

• 6. Finns det en binär Hammingkod som har dimension $k = 11$ (som linjär kod över $\mathbb{Z}_2$)? Om ja, ange en paritets-check-matris för en sådan kod.

Var god vänd!

• 7. Beräkna kapaciteten för kanalen med kanalmatris $\begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

• 8. Låt $\mathcal{C}$ vara en linjär kod av längd n över en ändlig kropp F, med minimalavstånd $d(\mathcal{C}) \geqq 2$, och låt $a \geqq 1$ vara ett heltal med $2a+1 < n$. Antag att det för varje *ord* $w \in F^n$ av vikt $\mathrm{wt}(w) = a+1$ gäller att det *kodord* $v \in \mathcal{C}$ som ligger närmast w har avstånd $d(v, w) = a$. Bevisa att koden $\mathcal{C}$ är perfekt.

LYCKA TILL!

Informations- och kodningsteori:
Lösningar till tenta 2004-06-01.

• 1. Enligt Krafts olikhet så finns en binär prefixfri kod med ordlängder $\ell_1, \ell_2, ..., \ell_r$ om och endast om $\sum_{j=1}^{r} 2^{-\ell_j} \leqq 1$. Vi beräknar:

$$2^{-2} + 2^{-3} + 2^{-3} + 2^{-3} + 2^{-4} + 2^{-5} + 2^{-5} + 2^{-5} + 2^{-5} = \frac{13}{16} < 1,$$

$$2^{-1} + 2^{-2} + 2^{-3} + 2^{-3} + 2^{-3} = \frac{9}{8} > 1,$$

$$2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} + 2^{-5} = \frac{31}{32} < 1.$$

Alltså finns det inte någon prefixfri binär kod med ordlängder $(1, 2, 3, 3, 3)$, men i de andra två fallen går det bra.

Exempel på en prefixfri binär kod med ordlängder $(2, 3, 3, 3, 4, 5, 5, 5, 5)$:

$$00, 010, 011, 100, 1010, 10110, 10111, 11000, 11001.$$

Exempel på en prefixfri binär kod med ordlängder $(1, 2, 3, 4, 5)$:

$$0, 10, 110, 1110, 11110.$$

• 2. Ternär entropi:

$$H_3(\mathcal{S}) = -0.6 \cdot \log_3 0.6 - 3 \cdot 0.1 \cdot \log_3 0.1 - 4 \cdot 0.025 \cdot \log_3 0.025 = 1.24353...$$

Vi tar nu fram en optimal ternär kod för $\mathcal{S}$. Enligt en känd sats är varje ternär Huffmankod för $\mathcal{S}$ optimal; och en sådan Huffmankod kan tas fram så här (utöka först källan med en "dummy-symbol" så att totala antalet symboler blir $\equiv 1 \pmod{3-1}$:

$\mathcal{S}$				0.6	0.1	0.1	0.1	0.025	0.025	0.025	0.025	0
$\mathcal{S}'$:			0.05	0.6	0.1	0.1	0.1	0.025	0.025			
$\mathcal{S}''$:		0.1		0.6	0.1	0.1	0.1					
$\mathcal{S}'''$:	0.3			0.6	0.1							
$\mathcal{S}^{(4)}$:	1											
$\mathcal{C}$				1	2	01	02	001	002	0000	0001	*
$\mathcal{C}'$:			000	1	2	01	02	001	002			
$\mathcal{C}''$:		00		1	2	01	02					
$\mathcal{C}'''$:	0			1	2							
$\mathcal{C}^{(4)}$:	ε											

(Tabellen med kodord fylls i baklänges, efter att hela tabellen med sannolikheter har utarbetats.)
Medelordlängd: $L(\mathcal{C}) = 1 + 0.3 + 0.1 + 0.05 = 1.45$.
Kodord: $1, 2, 01, 02, 001, 002, 0000, 0001$.

• 3. Vi beräknar $H = H_2(\mathcal{S}) = -0.4 \cdot \log_2 0.4 - 0.3 \cdot \log_2 0.3 - 0.2 \cdot \log_2 0.2 - 0.1 \cdot \log_2 0.1 = 1.8464393....$ Alltså är $H_2(\mathcal{S}^n) = n \cdot H = n \cdot 1.8464393....$ Varje Shannon-Fano-kod $\mathcal{C}_n$ för $\mathcal{S}^n$ har alltså medelordlängd $L(\mathcal{C}_n) \leqq n \cdot H + 1$ så $L(\mathcal{C}_n)/n \leqq H + 1/n$. Alltså duger n säkert om $H + 1/n < 1.85$, dvs om $n > (1.85 - H)^{-1} = 280.8....$

Svar: T.ex. $n = 281$.

• 4. Vi ser från H's storlek att $\mathcal{C}$ är en linjär kod av längd $n = 8$, dvs. $\mathcal{C} \subset \mathbb{Z}_2^8$, och av dimension $k = n - 4 = 4$. Alltså är $\mathcal{C}$'s hastighet: $R = k/n = \frac{1}{2}$.

Vi vet att G är en generatormatris till $\mathcal{C}$ om och endast om G är en 4×8-matris vars rader utgör en bas i lösningsrummet till följande ekvationssystem (över kroppen $\mathbb{Z}_2$):

$$(x_1\, x_2\, ...\, x_8)H^T = (0\,0\,0\,0)$$

$$\iff \begin{cases} x_1 + x_2 + x_4 + x_6 = 0 \\ x_2 + x_3 + x_4 + x_8 = 0 \\ x_2 + x_5 + x_6 + x_8 = 0 \\ x_4 + x_6 + x_7 + x_8 = 0 \end{cases}$$

$$\iff \begin{cases} x_1 = x_2 + x_4 + x_6 \\ x_3 = x_2 + x_4 + x_8 \\ x_5 = x_2 + x_6 + x_8 \\ x_7 = x_4 + x_6 + x_8 \end{cases}$$

$$\iff (x_1\, x_2\, ...\, x_8) = x_2 \cdot (1,1,1,0,1,0,0,0) + x_4 \cdot (1,0,1,1,0,0,1,0) + x_6 \cdot (1,0,0,0,1,1,1,0) + x_8 \cdot (0,0,1,0,1,0,1,1).$$

Alltså kan vi ta generatormatris:

$$G = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

(Det finns många alternativa val!)

För att finna minimalavståndet kan vi utnyttja satsen som säger att minimalavståndet är likamed antalet element i den minsta delmängd

av kolumner i paritets-check-matrisen som är linjärt beroende. Alternativt kan vi räkna upp alla kodord i $\mathcal{C}$ (det är lätt när vi har G):

$$\begin{array}{llll} 00000000, & 11101000, & 10110010, & 01011010, \\ 10001110, & 01100110, & 00111100, & 11010100, \\ 00101011, & 11000011, & 10011001, & 01110001, \\ 10100101, & 01001101, & 00010111, & 11111111. \end{array}$$

Härur ser vi att *minimalavståndet är* $d(\mathcal{C}) = 4$ (enligt lemmat som säger att för varje linjär kod gäller $d(\mathcal{C}) = \min\{\mathrm{wt}(v) \mid v \in \mathcal{C}, v \neq 0\}$).

• 5. Utdata-sannolikheterna blir:

$$\begin{pmatrix} 0.4 & 0.6 \end{pmatrix} \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix} = \begin{pmatrix} 0.5 & 0.5 \end{pmatrix}.$$

(dvs ut-källan $\mathcal{B}$ har $slh(b_1) = 0.5$, $slh(b_2) = 0.5$).

Bakåt-sannolikheterna Q_{ij} ges av Bayes formel:

$$Q_{ij} = slh(a = a_i \mid b = b_j) = \frac{p_i}{q_j} P_{ij}.$$

Alltså:

$$Q_{11} = \frac{0.4}{0.5} \cdot 0.8 = 0.64, \qquad Q_{12} = \frac{0.4}{0.5} \cdot 0.2 = 0.16,$$
$$Q_{21} = \frac{0.6}{0.5} \cdot 0.3 = 0.36, \qquad Q_{22} = \frac{0.6}{0.5} \cdot 0.7 = 0.84.$$

Systementropier (vi använder basen 2, och vi skriver som vanligt $H_2(x) = -x \log_2 x - (1-x)\log_2(1-x)$):

$$\begin{aligned} H_2(\mathcal{A}|\mathcal{B}) &= 0.5 \cdot H_2(\mathcal{A}|b = b_1) + 0.5 \cdot H_2(\mathcal{A}|b = b_2) \\ &= 0.5 \cdot H_2(Q_{11}) + 0.5 \cdot H_2(Q_{12}) = 0.788496... \\ H_2(\mathcal{B}|\mathcal{A}) &= 0.4 \cdot H_2(\mathcal{B}|a = a_1) + 0.6 \cdot H_2(\mathcal{B}|a = a_2) \\ &= 0.4 \cdot H_2(0.8) + 0.6 \cdot H_2(0.3) = 0.817545... \\ H_2(\mathcal{A}, \mathcal{B}) &= H_2(\mathcal{B}) + H_2(\mathcal{A}|\mathcal{B}) = H_2(0.5) + H_2(\mathcal{A}|\mathcal{B}) = 1.788496... \end{aligned}$$

Alternativ beräkning: Vi beräknar förenade sannolikheterna $R_{ij} = p_i P_{ij}$; $R_{11} = 0.32$, $R_{12} = 0.08$, $R_{21} = 0.18$, $R_{22} = 0.42$, härur följer $H(\mathcal{A}, \mathcal{B}) = -\sum_{i,j} R_{ij} \log_2 R_{ij} = 1.788496....$

Slutligen, informationen:

$$I_2(\mathcal{A}, \mathcal{B}) = H_2(\mathcal{A}) - H_2(\mathcal{A}|\mathcal{B}) = 0.182452...$$

• 6. För Hammingkoder över $\mathbb{Z}_2$ gäller $n = 2^c - 1$, $c = n - k$; detta ger $k = 2^c - 1 - c$; alltså fås $k = 11$ för $c = 4$, dvs det *finns* en Hammingkod av sökt typ!

En paritets-check-matris för denna kod är (t.ex.):

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

(Det finns väldigt många olika möjliga val; det viktiga är att varje möjlig kolumn utom noll-kolumnen ska förekomma precis en gång.)

• 7. Antag att inkällan har sannolikhetsvektorn (x_1, x_2, x_3), där $x_1 + x_2 + x_3 = 1$. Då har utkällan sannolikhetsvektorn $(\frac{1}{2}(x_1 + x_2), \frac{1}{2}(x_1 + x_2), x_3)$. Alltså är

$$\begin{aligned} H(\mathcal{B}|\mathcal{A}) &= x_1 \cdot H(\mathcal{B}|a = a_1) + x_2 \cdot H(\mathcal{B}|a = a_2) + x_3 \cdot H(\mathcal{B}|a = a_3) \\ &= (x_1 + x_2) \cdot (-2 \cdot \tfrac{1}{2} \cdot \log \tfrac{1}{2}) + x_3 \cdot 0 = (x_1 + x_2) \log 2, \end{aligned}$$

och därmed är

$$\begin{aligned} I(\mathcal{A}, \mathcal{B}) &= H(\mathcal{B}) - H(\mathcal{B}|\mathcal{A}) \\ &= \Big(-2 \cdot \frac{1}{2}(x_1 + x_2) \log \tfrac{1}{2}(x_1 + x_2) - x_3 \log x_3\Big) - (x_1 + x_2) \log 2 \\ &= -(x_1 + x_2) \log(x_1 + x_2) - x_3 \log x_3 \\ &= -x_3 \log x_3 - (1 - x_3) \log(1 - x_3). \end{aligned}$$

Vi vet att detta uttryck ("entropin för en binär källa med sannolikheter x_3 och $1 - x_3$") är maximalt då $x_3 = \frac{1}{2}$. Om vi använder bas 2 är detta maximum $I_2(\mathcal{A}, \mathcal{B}) = 1$.

Alltså är (binära) kapaciteten $C_2(\Gamma) = 1$.

• 8. Antag, i syfte att erhålla en motsägelse, att det finns något kodord $v \in \mathcal{C}$, $v \neq 0$, med $\mathrm{wt}(v) \leqq 2a$.

Fall 1: $\mathrm{wt}(v) \geqq a + 1$. Genom att ändra några icke-noll-positioner i v till noll så kan vi då finna ett ord $w \in F^n$ med $\mathrm{wt}(w) = a + 1$ och $d(v, w) = \mathrm{wt}(v) - \mathrm{wt}(w) \leqq 2a - (a + 1) = a - 1 < a$. Detta är en motsägelse mot förutsättningarna.

Fall 2: $\mathrm{wt}(v) \leqq a$. Genom att ändra några noll-positioner i v till icke-noll kan vi då finna ett ord $w \in F^n$ med $\mathrm{wt}(w) = a + 1$ och $d(v, w) = \mathrm{wt}(w) - \mathrm{wt}(v)$. Eftersom $d(\mathcal{C}) \geqq 2$ så måste $\mathrm{wt}(v) \geqq 2$, och därmed $d(v, w) = \mathrm{wt}(w) - \mathrm{wt}(v) \leqq a + 1 - 2 = a - 1 < a$. Detta är en motsägelse, precis som ovan.

Vårt antagande leder alltså säkert till en motsägelse, och måste därför vara felaktigt. Detta visar att det inte finns något kodord $v \in \mathcal{C}$, $v \neq 0$, med $\mathrm{wt}(v) \leqq 2a$. Eftersom $\mathcal{C}$ är linjär så har alltså $\mathcal{C}$ minimalavstånd $d(\mathcal{C}) \geqq 2a + 1$.

Å andra sidan, om $w \in F^n$ är ett godtyckligt ord av vikt $\mathrm{wt}(w) = a+1$ så finns enligt förutsättningarna ett kodord $v \in \mathcal{C}$ med $d(v,w) = a$; då är $\mathrm{wt}(v) = d(v,0) \leqq d(v,w) + d(w,0) = a + (a+1) = 2a+1$. Alltså är faktiskt $\mathcal{C}$'s minimalavstånd $d(\mathcal{C})$ *likamed* $2a+1$:

$$d(\mathcal{C}) = 2a+1.$$

Låt nu $w \in F^n$ vara ett godtyckligt ord. Låt $v \in \mathcal{C}$ vara ett kodord som ligger så nära w som möjligt. Antag att $d(v,w) \geqq a+1$. Genom att eventuellt ändra w "mot v" i vissa av de positioner där v och w skiljer sig åt kan vi då hitta ett ord w' som uppfyller $d(v,w') = a+1$ och $d(v,w) = d(v,w') + d(w',w)$. (Om $d(v,w) = a+1$ så blir $w' = w$.) Nu är $\mathrm{wt}(w'-v) = d(v,w') = a+1$ så enligt förutsättningarna finns ett kodord $u \in \mathcal{C}$ med $d(u, w'-v) = a$. Då är $u+v \in \mathcal{C}$ och $d(u+v, w') = d(u, w'-v) = a$; alltså är $d(u+v,w) \leqq d(u+v,w') + d(w',w) = a + (d(v,w) - d(v,w')) = a + d(v,w) - (a+1) = d(v,w) - 1$. Detta är en motsägelse mot att v var ett kodord som ligger så nära w som möjligt (ty $u+v$ ligger ju närmare!). Alltså kan omöjligt $d(v,w) \geqq a+1$ gälla; det måste istället vara så att $d(v,w) \leqq a$.

Vi har därmed visat att det för varje $w \in F^n$ finns ett kodord $v \in \mathcal{C}$ med $d(v,w) \leqq a$. Detta betyder att "bollarna"

$$S_a(v) = \{ w \in F^n \mid d(v,w) \leqq a \}$$

täcker hela F^n då v löper genom $\mathcal{C}$. Detta betyder att vi har *likhet* i Hammings sfärpackningsgräns (ty obs $t = [(d(\mathcal{C})-1)/2] = a$, pga $d(\mathcal{C}) = 2a+1$), det vill säga att koden $\mathcal{C}$ är perfekt.

V.S.B.