Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

Međuispit iz predmeta **TEORIJA INFORMACIJE**, 22. studenog 2017.

Pravilo bodovanja zadataka

Točno odgovoreni zadaci donose 5 bodova, netočno odgovoreni zadaci donose 2 negativna boda, a neodgovoreni zadaci donose 0 bodova.

1. zadatak: Mirna digitalizirana slika s bojama "1", "2", "3", …, "10" zadana je tablicom. Napomena: f_i označava frekvenciju pojavljivanja pojedine boje.

x_i	"1"	"2"	"3"	"4"	"5"	"6"	"7"	"8"	"9"	"10"
f_i	2500	500	1000	5000	4000	5000	3500	1500	2000	2000

Svaka boja ("1", "2", "3", …, "10") kodira se jednim simbolom iz skupa simbola $X = \{1, 2, 3, \ldots, 10\}$. Izračunajte minimalno vrijeme potrebno za prijenos dane slike od računala A do računala B modemom 56 kbit/s.

a) 3,0798 s

b) 1,4849 s

c) 0,4821 s

d) 2,7458 s

e) ništa od navedenog

Postupak rješavanja:

Kako bi odredili vjerojatnosti pojavljivanja pojedinog simbola $p(x_i)$, potrebno je izračunati ukupan broj odaslanih simbola, te pojedine frekvencije pojavljivanja simbola podijeliti tim brojem.

Ukupno imamo 10 simbola. Zbroj frekvencija pojavljivanja i pojedine vjerojatnosti dobivamo prema sljedećim izrazima:

$$\sum_{k=1}^{10} f_k = 27000$$

$$p(x_i) = \frac{f_i}{\sum_{k=1}^{10} f_k}; i = 1,2, \ldots ,10$$

Dakle,

$$[p(x_i)] = \begin{bmatrix} \frac{2500}{27000} & \frac{500}{27000} & \frac{1000}{27000} & \frac{5000}{27000} & \frac{4000}{27000} & \frac{5000}{27000} & \frac{3500}{27000} & \frac{1500}{27000} & \frac{2000}{27000} & \frac{2000}{27000} \end{bmatrix}$$

Tablica: Tablični prikaz podataka pojedine boje

x_i	"1"	"2"	"3"	"4"	"5"	"6"	"7"	"8"	"9"	"10"
f_i	2500	500	1000	5000	4000	5000	3500	1500	2000	2000
$p(x_i)$	0,926	0,019	0,037	0,185	0,148	0,185	0,13	0,056	0,074	0,074

$$H(X) = -\sum_{i=1}^{10} p(x_i)\log_2 p(x_i) = 3{,}0798\ \frac{\text{bit}}{\text{simbol}}$$

Minimalno vrijeme potrebno za prijenos slike zadanom brzinom može se odrediti prema:

$$t_{\min} = \frac{H(X)\sum_{k=1}^{10} f_k}{R_b}$$

$$t_{\min} = \frac{H(X)\sum_{k=1}^{10} f_k}{56\ \frac{\text{kbit}}{\text{s}}} = \frac{3{,}0798\ \cdot 27000}{56000}$$
$$= 1{,}4849\ \text{s}$$

2. zadatak: Na ulazu binarnog simetričnog kanala pojavljuju se dva simbola $X = \{0, 1\}$ s jednakom vjerojatnošću pojavljivanja. Vjerojatnog pogrešnog prijenosa kroz kanal dvostruko je manja od vjerojatnosti ispravnog prijenosa. Odredite srednji uzajamni sadržaj informacije (transinformacija) između ulaznog i izlaznog skupa simbola, tj. $I(X;Y)$.

a) 0,5 bit/simbol

b) 1 bit/simbol

c) 0,0817 bit/simbol

d) 0,0666 bit/simbol

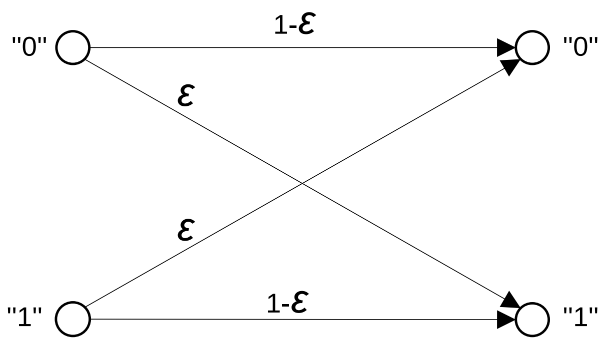
e) ništa od navedenog

Postupak rješeavanja:

Uz pomoć matrice uvjetnih vjerojatnosti prijelaza, gdje je s ε označena vjerojatnost pogrešnog prijenosa

$$\left[p(y_j \mid x_i)\right] = \begin{bmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix}$$

možemo skicirati kanal:



Slika: Binarni simetrični kanal iz zadatka

Imamo zadane vrijednosti ulaznih simbola $p_0 = p(0) = 1/2$ i $p_1 = p(1) = 1/2$ te možemo izračunati entropiju ulaznog skupa simbola:

$$H(X) = -\sum_{i=1}^{2} p(x_i) \cdot \log_2 p(x_i) = 1 \text{bit/simbol}$$

Poznavanjem vjerojatnosti pojavljivanja ulaznih simbola i matrice uvjetnih vjerojatnosti prijelaza moguće je dobiti matricu združenih vjerojatnosti, a iz nje (zbrajanjem po stupcima) vjerojatnosti pojavljivanja izlaznih simbola, a time i entropiju izlaznog skupa simbola *H*(*Y*).

$$\left[p(x_i, y_i)\right] = \begin{bmatrix} \frac{1}{2}(1-\varepsilon) & \frac{1}{2}\varepsilon \\ \frac{1}{2}\varepsilon & \frac{1}{2}(1-\varepsilon) \end{bmatrix} \Rightarrow p(y=0) = 1/2, p(y=1) = 1/2$$

$$H(Y) = 1 \text{bit/simbol}$$

Združena entropija *H*(*X*, *Y*) jednaka je

$$H(X,Y) = \sum_{i=1}^{2}\sum_{j=1}^{2} p(x_i, y_i) \log_2 p(x_i, y_i)$$
$$= -1(1-\varepsilon)\log_2\left[\frac{1}{2}(1-\varepsilon)\right] - \varepsilon \log_2(\frac{1}{2}\varepsilon)$$

$$H(X,Y) = 1 - \varepsilon \log_2 \varepsilon - (1-\varepsilon)\log_2(1-\varepsilon)$$

Izraz za transinformaciju u kanalu dobivamo pomoću dosad izračunatih entropija:

$$I(X;Y) = H(X) + H(Y) - H(X,Y) = 1 + \varepsilon \log_2 \varepsilon + (1-\varepsilon)\log_2(1-\varepsilon).$$

Iz uvjeta zadatka, znamo da je ε = 1/3. Konačno dobivamo,

$$I(X;Y) = 1 + \varepsilon \log_2 \varepsilon + (1-\varepsilon)\log_2(1-\varepsilon) = 0,0817 \text{ bit/simbol}$$

3. zadatak: Dan je skup simbola X = {1, 2, 3} s vjerojatnostima pojavljivanja p_1 = 0,8, p_2 = 0,02 i p_3 = 0,18 pri čemu su zadani kumulativni podskupovi [0, p_1), [p_1, $p_1 + p_2$) i [$p_1 + p_2$, 1). Dekodirajte primljenu kodiranu poruku 0,772352, kodiranu danim aritmetičkim kodom, u poruku duljine 4 simbola.

a) 1321

b) 1231

c) 1322

d) 1323

e) ništa od navedenog

Rješenje:

Odredimo tablicu kumulativnih podskupova:

Tablica1: Tablica kumulativnih podskupova

Simbol	Vjerojatnosti pojavljivanja	Komulativni podskupovi, [D_s, G_s)
1	0,8	[0, 0,8)
2	0,02	[0,8, 0,82)
3	0,18	[0,82, 1)

Dekodiranje poruke kodirane aritmetičkim kodom se provodi tako da se računaju pripadni podintervali za svaki simbol i provjerava kojem podintervalu kodirana poruka pripada.

Postupak se ponavlja onoliko puta koliko je poruka duga.

L_a = 0.772352 => očito je da kodirana poruka pripada podintervalu [0, 0.8), te je prvi simbol poruke `1`.

$$D = 0; \quad G = 0.8$$

1: $D' = D + (G - D)\cdot D_s = 0 + (0.8\text{-}0)\cdot 0 = 0$

$G' = D + (G - D)\cdot G_s = 0 + (0.8\text{-}0)\cdot 0.8 = 0.64$

2: $D' = 0 + (0.8 - 0)\cdot 0.8 = 0.64$

$G' = 0 + (0.8 - 0)\cdot 0.82 = 0.6560$

3: $D' = 0 + (0.8 - 0)\cdot 0.82 = 0.6560$

$G' = 0 + (0.8 - 0)\cdot 1 = 0.8$

Vidimo da kodirana poruka pripada podintervalu trećeg simbola te je drugi simbol poruke \`3\`.

$$D = 0.6560; \quad G = 0.8$$

1: $D' = 0.6560 + (0.8 - 0.6560)\cdot 0 = 0.6560$

$G' = 0.6560 + (0.8 - 0.6560)\cdot 0.8 = 0.7712$

2: $D' = 0.6560 + (0.8 - 0.6560)\cdot 0.8 = 0.7712$

$G' = 0.6560 + (0.8 - 0.6560)\cdot 0.82 = 0.7741$

Treći simbol poruke je simbol \`2\`.

$$D = 0.7712; \quad G = 0.7741$$

1: $D' = 0.7712 + (0.7741 - 0.7712)\cdot 0 = 0.7712$

$G' = 0.7712 + (0.7741 - 0.7712)\cdot 0.8 = 0.7735$

Četvrti simbol poruke je \`1\`, tj. poruka je

1321

4. zadatak: Uzimajući polazni rječnik D gdje je $D[0] = a$, $D[1] = b$ dekodirajte kodiranu poruku 0 1 0 4 5 koristeći algoritam LZW.

a) abbaaaa

b) abaaaab

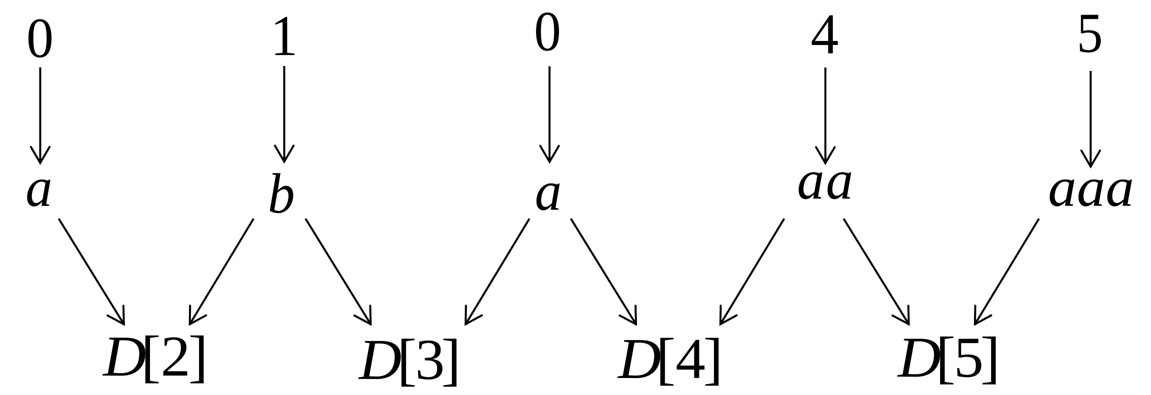
c) abaaaaba

d) abaaaaaa

e) ništa od navedenog

Postupak rješavanja:

Rječnik: $D[0] = a,\ D[1] = b,\ D[2] = ab,\ D[3] = ba,\ D[4] = aa,\ D[5] = aaa$



Dekodirana poruka: *abaaaaaa*

5. zadatak: Neki bezmemorijski izvor informacije generira simbole a i b. Razdioba vjerojatnosti je $P(a) = p$ i $P(b) = 1 - p$, $0 < p < 1$. Unutar jako dugačkog slučajnog slijeda simbola generiranog izvorom nalazi se niz sastavljen od m simbola a i n simbola b, pri čemu su $m, n \in \mathbf{N}$, $m \neq n$. Odredite za koju vjerojatnost p će količina informacije sadržana u tih $m + n$ simbola biti minimalna.

a) 1/2

b) $m/(m + n)$

c) $n/(m + n)$

d) m/n

e) ništa od navedenog

Postupak rješavanja:

Svaki simbol unutar niza doprinosi ukupnom sadržaju informacije proporcionalno svojoj vjerojatnosti, tj. sukladno svom vlastitom sadržaju informacije:

$$I(x_i) = \begin{cases} -\log_2 p & za\ x_i = a \\ -\log_2 (1-p) & za\ x_i = b \end{cases}, \ i = 1, 2, \ldots, m+n$$

Dakle, količina informacije sadržana u nizu od m simbola x_0 i n simbola x_1 dana je izrazom:

$$I_{\mathrm{u}} = \sum_{i=1}^{m+n} I(x_i) = -m \log_2 p - n \log_2 (1-p) [\mathrm{bit}]$$

Sada je potrebno utvrditi za koju vrijednost od p veličina I_{u} postiže svoj maksimum.

$$\frac{dI_u}{dp} = -m \frac{1}{\ln 2} \frac{1}{p} - n \frac{-1}{\ln 2} \frac{1}{(1-p)} = 0$$

$$\frac{m}{p} = \frac{n}{1-p} \Rightarrow p = \frac{m}{m+n}$$

$$\frac{d^2 I_u}{dp^2} = \frac{1}{\ln 2} \left(\frac{m}{p^2} + \frac{n}{(1-p)^2} \right) > 0$$

6. zadatak: Slijed od n, $n \in \mathbf{N}$, $n > 1$, uzastopnih simbola a na čijem je kraju znak ! kao oznaka kraja slijeda kodira se algoritmom LZ77 pri čemu je maksimalna duljina posmičnog prozora 2 simbola, a maksimalna duljina prozora za kodiranje $n - 1$ simbola. Koja od ponuđenih trojki (*pomak, duljina, sljedeći simbol*) se ne može pojaviti na izlazu kodera?

a) (0, 0, !)
b) (0, 0, a)
c) (1, $n - 1$, !)
d) (1, $n - 2$, a), $n > 2$
e) ništa od navedenog

Postupak rješavanja:

Za $n = 2$ koder generira sljedeće tri trojke: (0, 0, a), (0, 0, a) i (0, 0, !).

Za $n \geq 3$ koder generira sljedeće tri trojke: (0, 0, a), (1, $n - 2$, a) i (0, 0, !).

Dakle, na izlazu kodera se ne može pojaviti trojka oblika (1, $n - 1$, !).

7. zadatak: Neki komunikacijski kanal zadan je niže navedenom matricom $[P(Y|X)]$, pri čemu su vjerojatnosti pojavljivanja simbola na ulazu kanala označene kao $p_i = p(x_i)$, $i = 1, 2, \ldots, 6$. Uz pretpostavku da se simboli x_2 pojavljuju dvostruko češće od simbola x_1, odredite koliko iznosi vjerojatnost p_1 pri kojoj je transinformacija u kanalu maksimalna.

$$\left[p(Y|X)\right] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

a) 1/6

b) 1/8

c) 1/12

d) 1/4

e) ništa od navedenog

Postupak rješavanja:

Matrica združenih vjerojatnosti dana je sljedećim izrazom:

$$\left[P(x_i, y_j)\right] = \left[P(x_i)P(y_j|x_i)\right] = \begin{bmatrix} p_1 & 0 & 0 & 0 \\ p_2 & 0 & 0 & 0 \\ 0 & p_3 & 0 & 0 \\ 0 & p_4 & 0 & 0 \\ 0 & 0 & p_5 & 0 \\ 0 & 0 & 0 & p_6 \end{bmatrix}$$

Temeljem toga možemo odrediti vjerojatnosti $P(y_j)$, j = 1, 2, 3, 4:

$$\begin{bmatrix} P(y_1) & P(y_2) & P(y_3) & P(y_4) \end{bmatrix} = \begin{bmatrix} p_1 + p_2 & p_3 + p_4 & p_5 & p_6 \end{bmatrix}$$

Kapacitet kanala računa se prema izrazu:

$$C = \max_{\{P(x_i)\}} I(X;Y) = \max_{\{P(x_i)\}} \left[H(Y) - H(Y|X) \right]$$

S obzirom da je $H(Y|X) = 0$ (svi elementi matrice kanala su jednaki 1 ili 0), vrijedi:

$$C = \max_{\{P(x_i)\}} H(Y)$$

a taj se maksimum postiže kad vrijedi: $p_1 + p_2 = p_3 + p_4 = p_5 = p_6 = 1/4$. U tekstu zadatka je zadano da se simboli x_2 pojavljuju dvostruko češće od simbola x_1. Dakle, $p_2 = 2p_1$. Kombiniranjem tih jednadžbi dobivamo da je $p_1 = 1/12$.

8. zadatak: Promatrajte bacanje novčića uz zadane vjerojatnosti ishoda $P(\text{P}) = P(\text{G}) = 1/2$, pri čemu P označava pismo, a G glavu. Pretpostavimo da slučajna varijabla X predstavlja broj bacanja novčića do prve pojave pisma (na primjer, ako je slijed ishoda bacanja novčića GP, tada je vrijednost od $X = 2$). Odredite entropiju $H(X)$.

Pomoćni izrazi: $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$, $\sum_{n=0}^{\infty} n x^n = \frac{x}{(1-x)^2}$

a) 1 bit/simbol

b) 2 bit/simbol

c) 1,5 bit/simbol

d) 4 bit/simbol

e) ništa od navedenog

Postupak rješavanja:

U općenitom slučaju, neka je vjerojatnost da je u nekom bacanju bilo pismo p, a vjerojatnost da je bila glava $q = 1 - p$. Slučajna varijabla X podvrgava se geometrijskoj razdiobi s parametrom p, $P(X = n) = pq^{n-1}$, $n \in \mathbf{N}$. Sukladno tome, njenu je entropiju moguće izračunati na sljedeći način:

$$H(X) = -\sum_{n=1}^{\infty} pq^{n-1} \log\left(pq^{n-1}\right) = -\left[\sum_{n=0}^{\infty} pq^n \log p + \sum_{n=0}^{\infty} npq^n \log q \right] =$$
$$= -\frac{p \log p}{1-q} - \frac{pq \log q}{p^2} = \frac{-p \log p - q \log q}{p}$$

Uz razdiobu zadanu zadatkom vrijedi: $p = 1/2$ i $H(X) = 2$ bit/simbol, pri čemu su simboli u stvari nizovi bacanja novčića do pojave prvog pisma.

9. zadatak: Razmatrajte skup simbola x_i, $i = 1, .., 5$. Razdioba vjerojatnosti zadana je sljedećim izrazima: $P(x_1) = P(x_2) + P(x_3) + P(x_4)$, $P(x_2) = P(x_3) = P(x_4) = P(x_5)$, $\sum_{i=1}^{5} P(x_i) = 1$. Koder izvora kodira slučajan slijed simbola x_i binarnim kodom koristeći tehniku Shannon-Fano. Ta tehnika ne daje uvijek jedinstveni kôd. Drugim riječima, za neku zadanu razdiobu vjerojatnosti može proizvesti više od jednog koda, pri čemu ti kodovi mogu imati različite srednje duljine kodne riječi. Nadalje, pretpostavimo da koder kodira jako dugačak slijed simbola x_i. Odredite koliko će iznositi apsolutni iznos razlike u prosječnom broju poslanih binarnih simbola na 700 simbola x_i. ako se koriste najefikasnija i najmanje efikasna inačica koda.

a) 100 bita

b) 150 bita

c) 160 bita

d) 200 bita

e) ništa od navedenog

Postupak rješavanja:

Iz zadanih izraza možemo izračunati razdiobu vjerojatnosti nad skupom simbola x_i: $P(x_1) = 3/7$, $P(x_2) = P(x_3) = P(x_4) = P(x_5) = 1/7$. Prilikom kodiranja tehnikom Shannon-Fano potrebno je poredati te simbole po padajućim vjerojatnostima.

x_1 3/7 0 $l_1 = 1$ bit

x_2 1/7 1 0 0 $l_2 = 3$ bit

x_3 1/7 1 0 1 $l_3 = 3$ bit

x_4 1/7 1 1 0 $l_4 = 3$ bit

x_5 1/7 1 1 1 $l_5 = 3$ bit

Nazovimo ovaj kôd K_1. Njegova srednja duljina kodne riječi neka je L_1.

$$L_1 = \sum_{i=1}^{5} l_i P(x_i) = \frac{15}{7} \text{ bit/simbol}$$

Tehnika Shannon-Fano omogućava i drugačiji način kodiranja.

x_1 3/7 0 0 $l_1 = 2$ bit

x_2 1/7 0 1 $l_2 = 2$ bit

x_3 1/7 1 0 $l_3 = 2$ bit

x_4 1/7 1 1 0 $l_4 = 3$ bit

x_5 1/7 1 1 1 $l_5 = 3$ bit

Nazovimo ovaj kôd K_2. Njegova srednja duljina kodne riječi neka je L_2.

$$L_2 = \sum_{i=1}^{5} l_i P(x_i) = \frac{16}{7} \text{ bit/simbol}$$

Konačno, postoji i treća inačica koda.

x_1 3/7 0 0 $l_1 = 2$ bit

x_2 1/7 0 1 $l_2 = 2$ bit

x_3 1/7 1 0 0 $l_3 = 3$ bit

x_4 1/7 1 0 1 $l_4 = 3$ bit

x_5 1/7 1 1 $l_5 = 2$ bit

Nazovimo ovaj kôd K_3. Njegova srednja duljina kodne riječi neka je L_3.

$$L_3 = \sum_{i=1}^{5} l_i P(x_i) = \frac{16}{7} \text{ bit/simbol}$$

Očito vrijedi $L_2 = L_3 > L_1$. S obzirom da je entropija $H(X)$ jednaka u sva tri slučaja, najefikasniji kôd je onaj s najmanjom srednjom duljinom kodne riječi, tj. kôd K_1, a najmanje efikasni K_2, odnosno K_3. Dakle, slijed duljine 700 simbola x_i bit će prosječno kodiran s 1500 bita ako primijenimo kôd K_1, odnosno s 1600 bita ako primijenimo kôd K_2 ili K_3. Konačno, razlika iznosi 100 bita.

10. zadatak: Razmatrajte skup od četiri simbola koje generira bezmemorijski izvor, $X = \{ x_1, x_2, x_3, x_4\}$. Razdioba vjerojatnosti simbola, $P(x_i) = p_i$, $i = 1, 2, 3, 4$, zadovoljava sljedeće izraze:

$$\sum_{i=1}^{4} p_i = 1$$
$$1 > p_4 \geq p_3 \geq p_2 \geq p_1 > 0$$
$$p_1 + p_2 \leq p_3$$
$$p_1 + p_2 + p_3 \leq p_4$$

Odredite koliko iznosi najveća prosječna duljina kodne riječi ako su simboli kodirani binarnim Huffmanovim kodom.

a) 1,75 bit/simbol

b) 1 bit/simbol

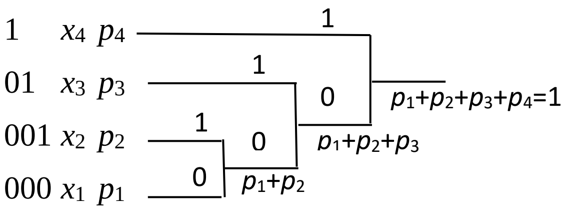
c) 1,5 bit/simbol

d) 2 bit/simbol

e) ništa od navedenog

Postupak rješavanja:

Poredamo li simbole x_i po padajućim vjerojatnostima i primijenimo Huffmanov kod sukladno zadatkom zadanim pravilima, dobivamo sljedeće kodne riječi:



Za srednju duljinu kodne riječi, L, vrijedit će sljedeće:

$$L = \sum_{i=1}^{4} l(x_i) p_i = 3p_1 + 3p_2 + 2p_3 + p_4 =$$
$$= (p_1 + p_2 + p_3 + p_4) + (p_1 + p_2 + p_3) + (p_1 + p_2) =$$
$$= 1 + (p_1 + p_2 + p_3) + (p_1 + p_2) [\text{bit/simbol}]$$

pri čemu su $l(x_i)$ duljine kodnih riječi x_i: $l(x_1) = l(x_2) = 3$, $l(x_3) = 2$, $l(x_4) = 1$.

Sada je potrebno uključiti zadane nejednakosti:

$$p_4 + p_3 + p_2 + p_1 = 1$$

$$\left.\begin{aligned} p_4 &= 1 - \left(p_3 + p_2 + p_1\right) \\ p_4 &\geq p_3 + p_2 + p_1 \end{aligned}\right\} \Rightarrow 1 - \left(p_3 + p_2 + p_1\right) \geq p_3 + p_2 + p_1 \Rightarrow p_3 + p_2 + p_1 \leq 0,5$$

$$p_3 \geq p_2 + p_1 \Rightarrow \left(p_2 + p_1\right) \leq 0,25$$

Sada možemo zaključiti da će za srednju duljinu kodne riječi vrijediti $L \leq 1 + 0,5 + 0,25 =$ 1,75bit/simbol.