Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

Međuispit iz predmeta **TEORIJA INFORMACIJE**, 30. studenog 2015.

Pravilo bodovanja zadataka

Svaki točno odgovoreni zadatak donosi 5 bodova, svaki netočno odgovoreni -2 boda, a neodgovoreni zadaci donose 0 bodova.

Zadatak 1: Komunikacijskim kanalom prenose se tri jednostavne poruke, 'a', 'b' i 'c', generirane iz skupa simbola $X = \{a, b, c\}$. Vjerojatnosti pojavljivanja simbola su: $p(a) = p(b) = 2p(c)$. Matrica uvjetnih vjerojatnosti prijelaza u kanalu zadana je na sljedeći način:

$$\left[p\left(y_j|x_i\right)\right]=\begin{bmatrix}0,7 & 0,1 & 0,2\\ 0,2 & 0,7 & 0,1\\ 0,1 & 0,2 & 0,7\end{bmatrix}$$

Odredite transinformaciju u kanalu.

a) 0,4096 bit/simbol
b) 1,522 bit/simbol
c) 0,381 bit/simbol
d) 1,566 bit/simbol

Postupak rješavanja:
Iz uvjeta da vjerojatnosni skup bude potpun, slijedi:

$$p(a)=p(x_1)=0,4$$
$$p(b)=p(x_2)=0,4$$
$$p(c)=p(x_3)=0,2$$

odnosno

$$\left[p(x_i,y_j)\right]=\begin{bmatrix}0,28 & 0,04 & 0,08\\ 0,08 & 0,28 & 0,04\\ 0,02 & 0,04 & 0,14\end{bmatrix}$$

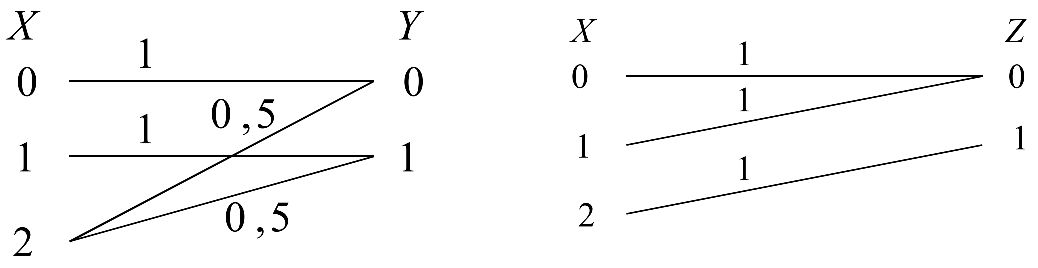
Zbrajanjem po stupcima matrice združenih vjerojatnosti dobivamo:

$$\left[p(y_j)\right]=\left[0,38 \quad 0,36 \quad 0,26\right]$$

Naposlijetku, transinformaciju dobivamo iz:

$$I(X;Y)=\sum_{i=1}^{3}\sum_{j=1}^{3}p(x_i,y_j)\log\frac{p(x_i,y_j)}{p(x_i)p(y_j)}=0,4096 \text{ bit/simbol}$$

Zadatak 2: Diskretno bezmemorijsko izvorište generira simbole iz skupa $X = \{0, 1, 2\}$ s vjerojatnostima pojavljivanja $p(0) = 0,25$, $p(1) = 0,25$ i $p(2) = 0,5$. Svaki od izvorišnih simbola istodobno se šalje dvama diskretnim komunikacijskim kanalima, prikazanim na slici, čiji su izlazi y_j , j = 1, 2, odnosno z_k, k = 1, 2. Odredite omjer transinformacije $I(X; Z)$ prema transinformaciji $I(X; Y)$.



a) 0
b) 1
c) 2
d) 0,5

Postupak rješavanja:

$$\left[p(y_j \mid x_i)\right] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0,5 & 0,5 \end{bmatrix}$$

$$\left[p(x_i, y_j)\right] = \left[p(x_i)p(y_j \mid x_i)\right] = \begin{bmatrix} 0,25 & 0 \\ 0 & 0,25 \\ 0,25 & 0,25 \end{bmatrix}$$

$$\left[p(y_j)\right] = [0,5 \quad 0,5]$$

$$H(Y) = -\sum_{j=1}^{2} p(y_j)\log_2 p(y_j) = -(0,5\log_2 0,5 + 0,5\log_2 0,5) = 1\,\frac{\text{bit}}{\text{simbol}}$$

$$\left[p(z_k \mid x_i)\right] = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\left[p(x_i, z_k)\right] = \left[p(x_i)p(z_k \mid x_i)\right] = \begin{bmatrix} 0,25 & 0 \\ 0,25 & 0 \\ 0 & 0,5 \end{bmatrix}$$

$$\left[p(z_k)\right] = [0,5 \quad 0,5]$$

$$H(Z) = -\sum_{k=1}^{2} p(z_k)\log_2 p(z_k) = -(0,5\log_2 0,5 + 0,5\log_2 0,5) = 1\,\frac{\text{bit}}{\text{simbol}}$$

$$I(X;Y) = H(Y) - H(Y \mid X)$$

$$H(Y \mid X) = -\sum_{i=1}^{3}\sum_{j=1}^{2} p(x_i, y_j)\log_2 p(y_j \mid x_i) = -(-0,25 - 0,25) = 0,5\,\frac{\text{bit}}{\text{simbol}}$$

$$I(X;Y) = 1 - 0,5 = 0,5\,\frac{\text{bit}}{\text{simbol}}$$

$$I(X;Z) = H(Z) - H(Z \mid X)$$

$$H(Z \mid X) = -\sum_{i=1}^{3}\sum_{j=1}^{2} p(x_i, z_k)\log_2 p(z_k \mid x_i) = 0 \ \frac{\text{bit}}{\text{simbol}}$$

$$I(X;Z) = 1 - 0 = 1 \ \frac{\text{bit}}{\text{simbol}}$$

$$\frac{I(X;Z)}{I(X;Y)} = \frac{1}{0,5} = 2$$

Zadatak 3: Diskretno bezmemorijsko izvorište generira simbole iz skupa $X = \{x_1, x_2, x_3, x_4\}$. Vjerojatnosti pojavljivanja simbola su sljedeće: $p(x_1) = 0,4$, $p(x_2) = 0,3$, $p(x_3) = 0,2$ i $p(x_4) = 0,1$. Izračunajte količinu informacije koja se prenosi u poruci $x_1x_2x_1x_3$.

a) 6,70 bit/poruka
b) 7,386 bit/poruka
c) 2,043 bit/poruka
d) 3,35 bit/poruka

Postupak rješavanja:

$$I(x_i) = -\log_2 p(x_i) \ \text{bit/simbol}$$

$$I(x_1x_2x_1x_3) = -\log_2\left(p(x_1)\cdot p(x_2)\cdot p(x_1)\cdot p(x_3)\right)$$

$$I(x_1x_2x_1x_3) = -\log_2(0,4\cdot 0,3\cdot 0,4\cdot 0,2)$$

$$I(x_1x_2x_1x_3) = 6,70 \ \text{bit/poruka}$$

Zadatak 4: Zadano je diskretno bezmemorijsko izvorište koje generira simbole x_i, $i = 1, 2, ..., n$. Svi simboli su jednako vjerojatni i maksimalna entropija izvorišnog skupa simbola iznosi $H(X) = 3,4594$ bit/simbol. Odredite za koliko se promijeni efikasnost koda prilikom kodiranja skupa simbola ternarnim Huffmanovim kodom u odnosu na kodiranje istog skupa simbola binarnim Huffmanovim kodom.

a) smanji se za 0,9755
b) smanji se za 0,0151
c) poveća se za 0,9604
d) poveća se za 0,0151

Postupak rješavanja:

i) Potrebno je odrediti broj izvorišnih simbola n. Kako su svi simboli jednako vjerojatni možemo zapisati:

$$p(x_1) = p(x_2) = \cdots = p(x_n) = \frac{1}{n}$$

Iz prethodne jednakosti i poznate entropije na ulazu računamo broj simbola:

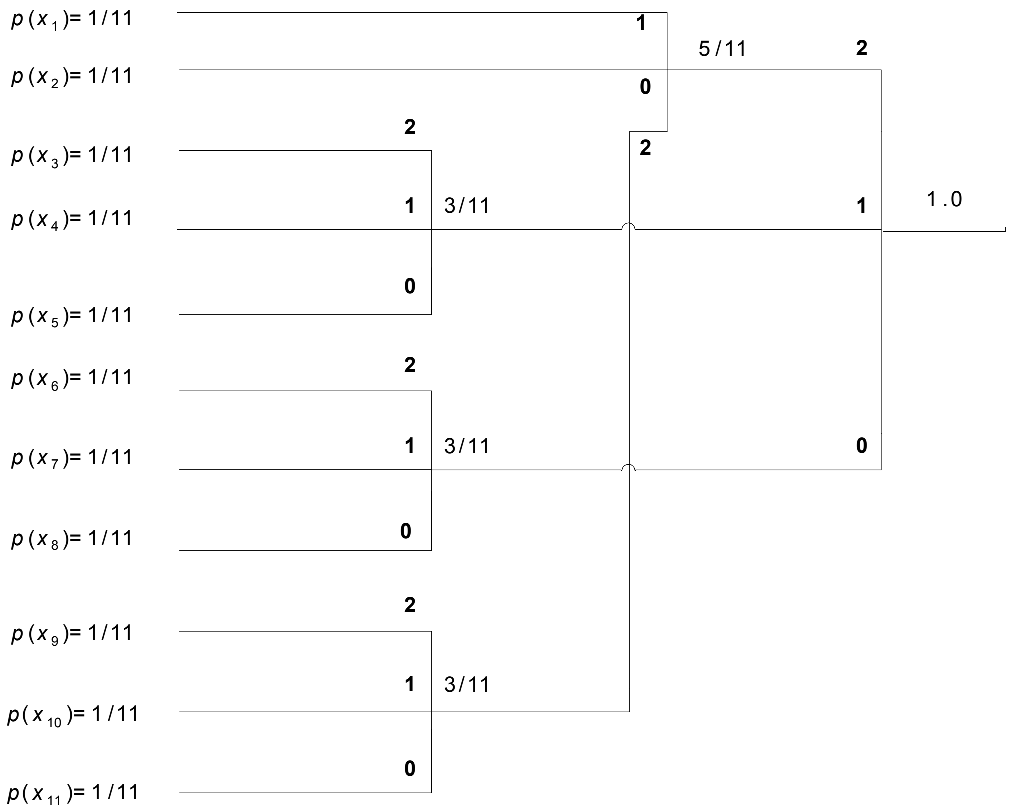
$$H(X) = 3,4594 \ \text{bit/simbol}$$

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) = -\sum_{i=1}^{n} \frac{1}{n}\log_2 \frac{1}{n} = -n \cdot \frac{1}{n}\log_2 \frac{1}{n} = \log_2 n$$

$$n = 2^{H(X)} = 2^{3,4594} = 11$$

Iz toga jasno slijedi :

$$p(x_1) = p(x_2) = \cdots = p(x_{11}) = \frac{1}{11}$$



U tablici su prikazani simboli sa pripadajućim kodnim riječima i njihovim duljinama za ternarno kodiranje

simbol (x_i)	vjerojatnost pojavljivanja $p(x_i)$	kodna riječ $C(x_i)$	duljina kodne riječi $l(x_i)$
x_1	1/11	21	2
x_2	1/11	20	2
x_3	1/11	12	2
x_4	1/11	11	2
x_5	1/11	10	2
x_6	1/11	02	2
x_7	1/11	01	2
x_8	1/11	00	2
x_9	1/11	222	3
x_{10}	1/11	221	3

x_{11}	1/11	220	3

Efikasnost koda računa se prema izrazu:

$$\varepsilon_{(3)} = \frac{H_{(3)}(X)}{L_{(3)}(X)}$$

Proračunajmo potrebne veličine:

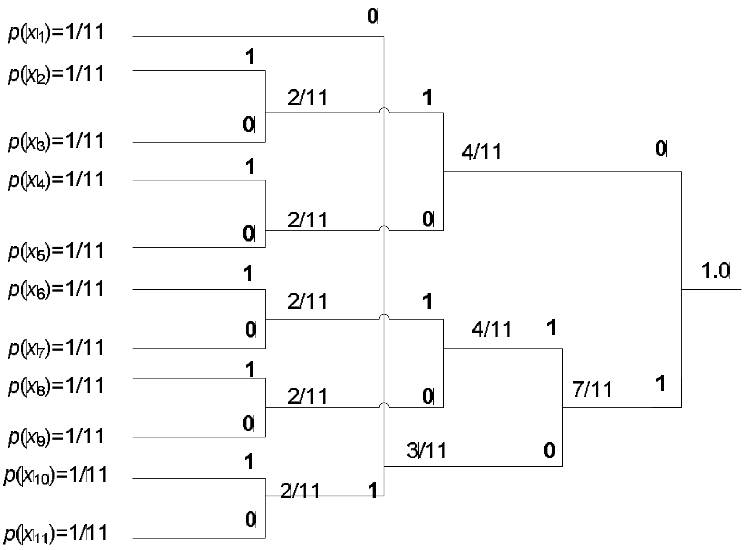
$$H_{(3)}(X) = -\sum_{i=1}^{11} p(x_i)\log_3 p(x_i) = -\log_3 \frac{1}{11} = 2{,}183\ \frac{\text{tern. simbola}}{\text{simbol}}$$

$$L_{(3)}(X) = \sum_{i=1}^{11} p(x_i) l_i = 2{,}273\ \frac{\text{tern. simbola}}{\text{simbol}}$$

Naposljetku:

$$\varepsilon_{(3)} = \frac{H_{(3)}(X)}{L_{(3)}(X)} = \frac{2{,}183}{2{,}273} = 0{,}9604$$

Potrebno je ponovno provesti postupak Huffmanovog kodiranja – ovaj put binarno, kako je prikazano na slici. U tablici su prikazane nove kodne riječi s pripadajućim duljinama.



simbol (x_i)	vjerojatnost pojavljivanja $p(x_i)$	kodna riječ $C(x_i)$	duljina kodne riječi $l(x_i)$
x_1	1/11	100	3
x_2	1/11	011	3
x_3	1/11	010	3
x_4	1/11	001	3
x_5	1/11	000	3

x_6	1/11	1111	4
x_7	1/11	1110	4
x_8	1/11	1101	4
x_9	1/11	1100	4
x_{10}	1/11	1011	4
x_{11}	1/11	1010	4

Efikasnost koda računamo prema izrazu:

$$\varepsilon = \frac{H(X)}{L(X)}$$

$$H(X) = -\sum_{i=1}^{11} p(x_i)\log_2 p(x_i) = -\log_2 \frac{1}{11} = 3{,}459\ \frac{\text{bit}}{\text{simbol}}$$

$$L(X) = \sum_{i=1}^{11} p(x_i) l_i = 3{,}546\ \frac{\text{bit}}{\text{simbol}}$$

$$\varepsilon = \frac{H(X)}{L(X)} = \frac{3{,}459}{3{,}546} = 0{,}9755$$

$$\Delta\varepsilon = \varepsilon - \varepsilon_{(3)} = 0{,}9755 - 0{,}9604 = 0{,}0151$$

Obzirom da vrijedi $\Delta\varepsilon > 0$, to je binarno kodiranje efikasnije od ternarnog.

Zadatak 5: Koristeći algoritam LZ77 kodirana je poruka aaaabbbccd*, uzimajući pri tome da je maksimalna duljina posmičnog prozora 6, a prozora za kodiranje 5 simbola. Odredite treću trojku kodirane poruke. **Napomena:** znak "*" označava kraj slijeda.

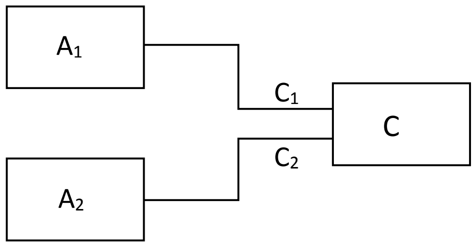
a) (1,2,c)
b) (1,1,a)
c) (1,2,a)
d) (1,1,b)

Postupak rješavanja:

aaaabbbccd*	(0,0,a)
aaaabbbccd*	(1,3,b)
aaaabbbccd*	(1,2,c)
aa**aabbbc**cd*	(1,1,d)
aaaa**bbbccd***	(0,0,*)

Zadatak 6: U kontrolnoj stanici za daljinska mjerenja nalaze se dva instrumenta A_1 i A_2. Instrument A_1 ima skalu sa 100 podjeljaka i njegovo se pokazivanje mijenja svakih 0,05 s. Instrument A_2 ima skalu s 10 podjeljaka, a njegovo se pokazivanje mijenja svakih 0,01 s. Sve

vrijednosti mjerenih veličina su jednako vjerojatne. Kolika je prosječna količina informacije (bit/s) koju iz kontrolne stanice treba prenijeti u centralnu stanicu C u jedinici vremena?



a)333 bit/s
b)100 bit/s
c)133 bit/s
d)466 bit/s

Postupak rješavanja:
Prosječna količina informacije koju treba prenijeti u centralnu stanicu jednaka je prosječnoj količini informacije koju generiraju instrumenti A_1 i A_2. Promatramo li instrumente kao izvore informacije moramo im pridijeliti nekakvu veličinu koja će ih opisati. To je srednji vlastiti sadržaj informacije koji glasi:

$$I(X)=\sum_{i=1}^{n}p(x_i)\cdot I(x_i)=-\sum_{i=1}^{n}p(x_i)\log_2 p(x_i)$$

Srednji vlastiti sadržaj informacije je težinska suma vlastitih sadržaja informacija svih pojedinačnih događaja pri čemu svaki događaj učestvuje s iznosom koji odgovara umnošku vjerojatnosti i vlastitog sadržaja informacije. Uz jednaku vjerojatnost svih simbola x_i dobije se pojednostavljeni izraz:

$$I(X)=\sum_{i=1}^{n}\frac{1}{n}\log_2(\frac{1}{n})=\log_2(n)$$

Prema tome, instrumentima A_1 i A_2 odgovaraju srednji vlastiti sadržaji informacije I_1, odnosno I_2:

$$I_1(X)=\log_2(100)=6,64 \text{ bit}$$

$$I_2(X)=\log_2(10)=3,33 \text{ bit}$$

Prosječna količina informacije koju u jedinici vremena generiraju instrumenti iznosi:

$$C_1=\frac{I_1(X)}{T_1}=\frac{6,64}{5\cdot 10^{-2}}=133 \text{ bit/s}$$

$$C_2=\frac{I_2(X)}{T_2}=\frac{3,33}{10^{-2}}=333 \text{ bit/s}$$

Ukupna prosječna količina informacije koju treba prenijeti u centralnu stanicu iznosi:

$$C=C_1+C_2=466 \text{ bit/s}$$

To je ujedno i najveći sadržaj informacije koji se može dobiti uz ovakve karakteristike instrumenata kao izvora informacije. (Neizvjesnost je najveća kada su svi događaji jednako vjerojatni.)

Zadatak 7: Prijenos poruka provodi se abecedom koju čini osam različitih simbola. Svaki je simbol fizikalno prikazan signalom trajanja τ = 1 ms. Vjerojatnosti predaje simbola su: $p(x_i)$ =

2^{-i}, $i = 1, 2, \ldots, 7$, $p(x_8) = p(x_7)$. Odredite kapacitet kanala za prijenos poruke bez prisustva smetnji.

a) 1980 bit/s
b) 3000 bit/s
c) 1,98 bit/s
d) 3 bit/s

Postupak rješavanja:

Entropija izvora iznosi:
Zanima nas maksimalna brzina prijenosa informacije od izvora do odredišta koja se može postići – kapacitet kanala. Sadržaj informacije koji se prenese jednak je transinformaciji, H_T. Ako s F označimo frekvenciju pojave signala: $F = 1/\tau$, kapacitet kanala možemo izraziti kao:

$$C = \max_{\{p(x_i)\}} F \cdot H_T = F \max_{\{p(x_i)\}} \left[H(X) - H(X \mid Y) \right] \text{ bit/s}$$

pri čemu se traži takva razdioba apriornih vjerojatnosti da se postigne maksimalna brzina prijenosa. Za idealni kanal, bez prisustva smetnji, $H(X \mid Y) = 0$, izraz za kapacitet kanala glasi:

$$C = F \max_{\{p(x_i)\}} \left[H(X) \right]$$

Maksimum funkcije $H(X)$ postiže se uz ravnomjernu razdiobu vjerojatnosti:

$$\max_{\{p(x_i)\}} H(X) = \log_2 8 = 3 \text{ bit/simbol}$$

tako da je kapacitet:

$$C = 3000 \text{ bit/s}$$

Zadatak 8: Promatrani informacijski izvor bez memorije generira simbole x_i iz osmočlane abecede $X = \{x_1, \ldots, x_8\}$. Pripadajuća razdioba vjerojatnosti pojavljivanja simbola na izlazu izvora određena je sljedećim jednakostima:

$$\sum_{i=1}^{8} p(x_i) = 1, \; P(x_1) = 1/4, \; P(x_3) = P(x_4) = 1/8, \; P(x_5) = P(x_6) = P(x_7) = P(x_8) = 1/16$$

Koder informacije, tj. koder izvora, kodira simbole x_i Shannon-Fanovom tehnikom. Simboli na izlazu kodera informacije su binarni simboli b_i iz skupa $B = \{0, 1\}$ čije su vjerojatnosti pojavljivanja $P(b_i = 0) = p_0$ i $P(b_i = 1) = p_1$. Koliko iznosi omjer entropije na izlazu kodera informacije i entropije na izlazu izvora, $H(B)/H(X)$, promatran na jako dugačkom slijedu generiranih simbola? **Napomena:** kodiranje Shannon-Fanovom tehnikom potrebno je provesti na način da se prije samog kodiranja simboli x_i poredaju sukladno vjerojatnosti $P(x_i)$, s time da se na višoj poziciji u početnoj grupi, koja obuhvaća sve simbole x_i, uvijek nalazi simbol čija je vjerojatnost veća ili jednaka vjerojatnosti simbola na nižoj poziciji u početnoj grupi.

a) 0,33 (simbol x_i)/(binarni simbol b_i)

b) 0,75 (simbol x_i)/(binarni simbol b_i)

c) 0,364 (simbol x_i)/(binarni simbol b_i)

d) 0,5 (simbol x_i)/(binarni simbol b_i)

Postupak rješavanja:

Na temelju zadanih vjerojatnosti i uvjeta vrijedi $P(x_2) = 1/4$. Poredamo simbole po padajućim vjerojatnostima i kodiramo Shannon-Fanovom tehnikom.

simbol	vjerojatnosti	1. korak	2. korak	3. korak	4. korak	$C(x_i)$	l_i
x_1	0,25	0	0			00	2
x_2	0,25	0	1			01	2
x_3	0,125	1	0	0		100	3
x_4	0,125	1	0	1		101	3
x_5	0,0625	1	1	0	0	1100	4
x_6	0,0625	1	1	0	1	1101	4
x_7	0,0625	1	1	1	0	1110	4
x_8	0,0625	1	1	1	1	1111	4

Entropija simbola na ulazu kodera, $H(X)$ iznosi:

$$H(X)=-\sum_{i=1}^{8}P(x_i)\log_2\left(P(x_i)\right)=-\left[2\cdot\frac{1}{4}\log_2\left(\frac{1}{4}\right)+2\cdot\frac{1}{8}\log_2\left(\frac{1}{8}\right)+4\cdot\frac{1}{16}\log_2\left(\frac{1}{16}\right)\right]=2\frac{3}{4}\frac{\text{bit}}{\text{simbol } x_i}$$

a, srednja duljina kodne riječi L_{sr} iznosi:

$$L_{\text{sr}}=\sum_{i=1}^{8}l_iP(x_i)=2\cdot2\cdot\frac{1}{4}+2\cdot3\cdot\frac{1}{8}+4\cdot4\cdot\frac{1}{16}=2\frac{3}{4}\frac{\text{bit}}{\text{simbol } x_i}$$

Za određivanje entropije simbola na izlazu kodera, $H(B)$ potrebno je poznavati vjerojatnosti p_0 i p_1. Ako se radi o jako dugačkom nizu generiranih simbola, te vjerojatnosti određujemo sljedećim izrazima:

$$p_0=\frac{\sum_{i=1}^{8}n_{0i}P(x_i)}{L_{\text{sr}}}=\frac{2\cdot\frac{1}{4}+1\cdot\frac{1}{4}+2\cdot\frac{1}{8}+1\cdot\frac{1}{8}+2\cdot\frac{1}{16}+1\cdot\frac{1}{16}+1\cdot\frac{1}{16}}{2\frac{3}{4}}=\frac{\frac{22}{16}}{\frac{11}{4}}=\frac{1}{2},$$

$$p_1=\frac{\sum_{i=1}^{8}n_{1i}P(x_i)}{L_{\text{sr}}}=\frac{1\cdot\frac{1}{4}+1\cdot\frac{1}{8}+2\cdot\frac{1}{8}+2\cdot\frac{1}{16}+3\cdot\frac{1}{16}+3\cdot\frac{1}{16}+4\cdot\frac{1}{16}}{2\frac{3}{4}}=\frac{\frac{22}{16}}{\frac{11}{4}}=\frac{1}{2}$$

pri čemu za broj binarnih nula, n_{0i}, odnosno broj binarnih jedinica, n_{1i}, u svakoj kodnoj riječi $C(x_i)$, i = 1, …, 8, vrijedi: $n_{0i}+n_{1i}=l_i$. Dakle, $p_0=p_1=1/2$, pa $H(B)$ iznosi:

$$H(B)=\sum_{i=0}^{1}p_i\log_2(p_i)=1\frac{\text{bit}}{\text{simbol } b_i}$$

Dakle, $H(B)/H(X)$ = 1/(11/4) = 4/11 ≈ 0,364 (simbol x_i)/(binarni simbol b_i).

Zadatak 9: Promatrajte informacijski izvor bez memorije koji generira poruke sastavljene od simbola x_i iz N-članog skupa simbola $X=\{x_1,\ldots,x_N\}$ čije su vjerojatnosti pojavljivanja $P(x_i)$. Te su vjerojatnosti određene sljedećim izrazima:

$$\sum_{i=1}^{N}P(x_i)=1,$$

$$P(x_1)=\frac{1}{2},\frac{P(x_i)}{P(x_{i+1})}=2,\forall i=1,\ldots,N-2$$

Poruke sastavljene od sljedova simbola x_i kodiraju se aritmetičkim kodom C. Pri kodiranju se koristi sljedeće pravilo: za svaku poruke duljine M simbola, p_{Mj}, $j = 1, \ldots, N^M$, $M \in \mathbf{N}$, u podintervalu dobivenom postupkom kodiranja, čija je duljina $d(p_{Mj})$, uvijek se odabire vrijednost na sredini tog podintervala, te se tako dobivena vrijednost pretvara u binarni broj od kojeg se za kodnu riječ $C(p_{Mj})$ uzimaju znamenke desno od decimalnog zareza. Ako pretpostavimo da je kôd C ujedno i prefiksni, odredite koliko binarnih znamenaka sadrži najdulja kodirana poruka sastavljena od M simbola x_i.

a) $N \cdot M$ bita

b) $N \cdot M + 1$ bita

c) $(N - 1) \cdot M$ bita

d) $(N - 1) \cdot M + 1$ bita

Postupak rješavanja:

Iz zadanih uvjeta očito je da vrijedi $P(x_{N-1}) = P(x_N) = 1/(2^{N-1})$. Prilikom aritmetičkog kodiranja poruka sastavljenih od M simbola x_i, koje jednoznačno zapisujemo kao $p_{Mj} = (x_{1i}, x_{2i}, \ldots, x_{Mi})$, $x_{ki} \in X$, $\forall k = 1, \ldots, M$, $\forall i = 1, \ldots, N$, $\forall j = 1, \ldots, N^M$, duljina $d(p_{Mj})$ podintervala pridruženog poruci p_{Mj} određuje se množenjem vjerojatnosti simbola od kojih je poruka sastavljena. Nadalje, ako pretpostavimo da je kôd C prefiksni, tada se prilikom formiranja kodne riječi za poruku p_{Mj} uzima broj bita $l(p_{Mj})$ određen sljedećim izrazom:

$$l\left(p_{Mj}\right) = \left\lceil \log_2\left(\frac{1}{d\left(p_{Mj}\right)}\right)\right\rceil + 1 \text{ bit}$$

Duljina poruke će očito biti maksimalna, ako poruka sadrži isključivo simbole čije su vjerojatnosti pojavljivanja $1/(2^{N-1})$, tj. najmanje moguće, a to su simboli x_k, $k \in \{N\text{-}1, N\}$. Dakle, vrijedi:

$$\max_j l\left(p_{Mj}\right) = \max_j\left(\left\lceil \log_2\left(\frac{1}{d\left(p_{Mj}\right)}\right)\right\rceil + 1\right) = \left\lceil \log_2\left(\frac{1}{\left(1/2^{N-1}\right)^M}\right)\right\rceil + 1 = (N-1)\cdot M + 1 \text{ bit}$$

Zadatak 10: Informacijski izvor generira simbole hrvatske abecede. Slijed se kodira kodom LZW. Prije početka kodiranja u memoriju kodera pohranjeni su zapisi oblika (indeks, znak) za sva slova abecede i za znakove koji se u tekstu mogu pojaviti. Razmotrimo poseban slučaj (npr. izvor je u kvaru) kad izvor generira isključivo slovo **a**. Ako izvor izgenerira točno 4951 slovo **a**, koliko će pri tome biti novih zapisa oblika (indeks, riječ) u rječniku? **Pojašnjenje:** svaki slijed sastavljen od uzastopnih slova **a**, npr. **aaa**, a koji nije od ranije sadržan u rječniku, smatra se jednim novim zapisom u rječnik, oblika (indeks, riječ).

a) 98

b) 100

c) 102

d) 4950

Postupak rješavanja:

Ako izvor kodira slijed uzastopnih slova **a** moguće je zamijetiti sljedeću pravilnost:

1) kad izvor kodira prvo slovo **a** (na samom početku dugačkog slijeda), ne zapisuje ništa u rječnik jer za slovo **a** već postoji zapis u rječniku;

2) zatim uzima sljedeće slovo **a** i u rječnik zapisuje simbol **aa** te njemu pripadajući indeks (to je jedan novi zapis);

3) zatim uzme to drugo slovo **a** i spaja ga s naredna dva slova **a** te u rječnik zapisuje **aaa**

4) itd.

Ako to prikažemo tablicom, vidljiva je sljedeća pravilnost:

iteracija i	broj uzastopnih slova **a** $N_i(\mathbf{a})$	ukupan broj novih zapisa u rječniku x_i
1	1	0
2	2	1
3	4	2
4	7	3
5	11	4
6	16	5

Broj uzastopnih slova **a** (lijevi stupac tablice), $N_i(\mathbf{a})$, se povećava svaki puta za ukupan broj novih zapisa, x_i, tj. vrijedi: $N_i(\mathbf{a}) - N_{i-1}(\mathbf{a}) = x_i$, za $i \geq 2$. Iz toga možemo zaključiti da vrijedi:

$$N_M = \sum_{i=2}^{M} x_i + N_1 = \sum_{i=1}^{M-1} i + 1 = 1 + \frac{M \cdot (M-1)}{2}$$

$$M^2 - M - 2(N_M - 1) = 0$$

$$M_{1,2} = \frac{1 \pm \sqrt{1 + 8(N_M - 1)}}{2}$$

Očito je da samo jedno rješenje ima smisla, pa konačno vrijedi:

$$M = \frac{1 + \sqrt{1 + 8(N_M - 1)}}{2}$$

Dakle, ako je $N_M = 4951$, tada je $M = 100$.