Tekst zadatka

Dan je skup simbola S s pripadajućim vjerojatnostima pojavljivanja :

$$S = \begin{pmatrix} s_1 & s_2 & \dots & s_m \\ p_1 & p_2 & & p_m \end{pmatrix}$$

Simboli su jednoznačno kodirani prefiksnim kodom. Ako je $m = 6$ i ako su zadane duljine kodnih riječi, odredite najmanji broj simbola abecede prefiksnog koda.

$$\{l_1, l_2, \dots, l_6\} = \{1,1,2,3,2,3\}$$

Postupak rješavanja

Za svaki prefiksni kod sa abecedom od d simbola i duljinama kodnih riječi $l_1, l_2, \ldots, l_n$ vrijedi Kraftova nejednakost:

$$\sum_{i=1}^{m} d^{-l_i} \leq 1$$

Dakle, za konkretan slučaj Kraftova nejednakost izgleda ovako:

$$\frac{1}{d} + \frac{1}{d} + \frac{1}{d^2} + \frac{1}{d^3} + \frac{1}{d^2} + \frac{1}{d^3} \leq 1$$

Postupak rješavanja

$$\frac{2}{d}+\frac{2}{d^2}+\frac{2}{d^3}\leq 1$$

$$2d^2+2d+2-d^3\leq 0$$

$$d\geq 2.9196$$

Konačno rješenje zadatka

Kako broj simbola abecede mora biti prirodan broj, najmanji cijeli broj koji zadovoljava Kraftovu nejednakost je :

$$d_{min} = \lceil 2.9196 \rceil = 3$$

$$d_{min} = 3$$

Tekst zadatka

Promatrani informacijski izvor bez memorije generira simbole x_i iz osmočlane abecede $X=\{x_1, x_2, \ldots, x_8\}$. Pripadajuća razdioba vjerojatnosti pojavljivanja simbola na izlazu izvora određena je sljedećim jednakostima:

$\sum_{i=1}^{8} p(x_i) = 1$, $P(x_1) = 1/4$, $P(x_2) = P(x_4) = 1/8$, $P(x_3) = P(x_5) = P(x_7) = P(x_8)$ $1/16$

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?

<u>NAPOMENA</u>: kodiranje Shannon – Fanovom tehnikom pojednostavljeno je provesti na način da se prije samog kodiranja simboli x_i poredaju sukladno vjerojatnosti $P(x_i)$, a 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.

Postupak rješavanja

- Entropija izvora koji generira simbole x_i:

$$H(X) = -\sum_{i=1}^{8} p(x_i)\log_2 p(x_i) \ [bit/simbol]$$

- Entropija na izlazu kodera:

$$H(B) = -p(b_0)\log_2 p(b_0) - p(b_1)\log_2 p(b_1) \ [bit/simbol]$$

Postupak rješavanja

- Kodiramo Shannon – Fanovom metodom:

x_i	$p(x_i)$	$c(x_i)$	$l(x_i)$
x_1	$1/4$	00	2
x_2	$1/4$	01	2
x_3	$1/8$	100	3
x_4	$1/8$	101	3
x_5	$1/16$	1100	4
x_6	$1/16$	1101	4
x_7	$1/16$	1110	4
x_8	$1/16$	1111	4

Postupak rješavanja

- Entropija izvora:

$$H(X) = -\sum_{i=1}^{8} p(x_i)\log_2 p(x_i) = 2.75 \ [bit/simbol]$$

- Srednja duljina kodnih riječi:

$$L\ (X) = \sum_{i=1}^{8} l_i \cdot P(x_i) = 2.75 \ [bit/simbol]$$

Tekst zadatka

- Ponašanje promatranog izvora s memorijom moguće je opisati homogenim markovljevim lancem prvog reda. Taj lanac predstavlja niz diskretnih slučajnih varijabli X_0, X_1, X_2, …, a ima dva stanja, 0 i 1. Razdioba slučajne varijable X_1 je $\mathbf{p}(1) = [P(X_1 = 0), P(X_1 = 1)] = [3/4, 1/4]$, slučajne varijable X_2 je $\mathbf{p}(2) = [P(X_2 = 0), P(X_2 = 1)] = [5/8, 3/8]$ i slučajne varijable X_3 je $\mathbf{p}(3) = [P(X_3 = 0), P(X_3 = 1)] = [9/16, 7/16]$. Odredite entropiju izvora.

Postupak rješavanja

- Za razdiobe slučajnih varijabli X_1, X_2 i X_3 vrijedi:

$$\mathbf{p}(2) = \mathbf{p}(1)\Pi$$

$$\mathbf{p}(3) = \mathbf{p}(2)\Pi$$

- pri čemu je matrica prijelaznih vjerojatnosti markovljevog lanca. Pretpostavimo da ta matrica ima oblik

$$\Pi = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix}$$

Postlupak rješavanja

- Dakle:

$$\mathbf{p}(2) = \mathbf{p}(1)\Pi \qquad \left[\frac{3}{4}, \frac{1}{4}\right] \cdot \begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix} = \left[\frac{5}{8}, \frac{3}{8}\right]$$

$$\mathbf{p}(3) = \mathbf{p}(2)\Pi \qquad \left[\frac{5}{8}, \frac{3}{8}\right] \cdot \begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix} = \left[\frac{9}{16}, \frac{7}{16}\right]$$

- Iz tog sustava jednadžbi odredimo α i β: $\alpha = 3/4 = \beta$. Sad je moguće odrediti stacionarne vjerojatnosti lanca: $\pi = [\pi_0, \pi_1]$ temeljem jednakosti $\Pi^T\pi = \pi$:

$$\begin{bmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{bmatrix}^T \cdot \begin{bmatrix} \pi_0 \\ \pi_1 \end{bmatrix} = \begin{bmatrix} \pi_0 \\ \pi_1 \end{bmatrix}$$

Postupak rješavanja

- Dakle, $\pi_0 = \pi_1 = 1/2$
- Sada je moguće odrediti entropiju izvora:

$$H = -\sum_{i=0}^{1}\pi_i \sum_{j=0}^{1} p_{ij}\log\left(p_{ij}\right) = -\frac{1}{2}2\left[\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}\right] = \frac{3}{4}\log_2\frac{4}{3} + \frac{1}{4}\log_2 4 =$$

$$= \frac{3}{4}\left[2 - \log_2 3\right] + \frac{1}{4}2 = \frac{6}{4} + \frac{2}{4} - \frac{3}{4}\log_2 3 = 2 - \frac{3}{4}\log_2 3 = 0{,}811\,\frac{\text{bit}}{\text{simbol}}$$

Tekst zadatka

Diskretno bezmemorijsko izvorište generira simbole iz skupa simbola $X = \{x_1, x_2, x_3, x_4\}$. 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 $x_1x_2x_1x_3$.

Postupak rješavanja

- Količina informacije koju svojom pojavom donosi simbol x_i:

$$I(x_i) = -\log_2 p(x_i)\ [bit / simbol]$$

- Količina informacije poruke $x_1 x_2 \ldots x_k$:

$$I(x_1 x_2 \ldots x_k) = -\log_2 (p(x_1) \cdot p(x_2) \cdot \ldots \cdot p(x_k))\ [bit / poruka]$$

Postupak rješavanja

- Zadane vjerojatnosti pojavljivanja simbola:

$$p(x_1) = 0{,}4$$
$$p(x_2) = 0{,}3$$
$$p(x_3) = 0{,}2$$
$$p(x_4) = 0{,}1$$

Konačno rješenje zadatka

- Količina informacije koja se prenosi u poruci $x_1x_2x_1x_3$:

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

Tekst zadatka

Binarni izvor generira slijed statistički neovisnih simbola A i B koji se pojavljuju s vjerojatnostima $p(A) = 1/3$ i $p(B) = 2/3$. Odredite optimalni ternarni kod za kodiranje poruka koje se sastoje od tri simbola odabranih iz skupa {A, B}. Izračunajte učinkovitost kreiranog ternarnog koda.

Postupak rješavanja

$p(A) = 1/3$ i $p(B) = 2/3$

$p(AAA)=1/27$

$p(AAB)=2/27$

$p(ABA)=2/27$

$p(ABB)=4/27$

$p(BAA)=2/27$

$p(BAB)=2/27$

$p(BBA)=4/27$

$p(BBB)=8/27$

Što se zbiva kada baza nije 2?

- Moramo provjeriti imamo li "dovoljno" simbola za kodiranje!
  - ako nemamo dodajemo nove simbole s $p(x_i)=0$

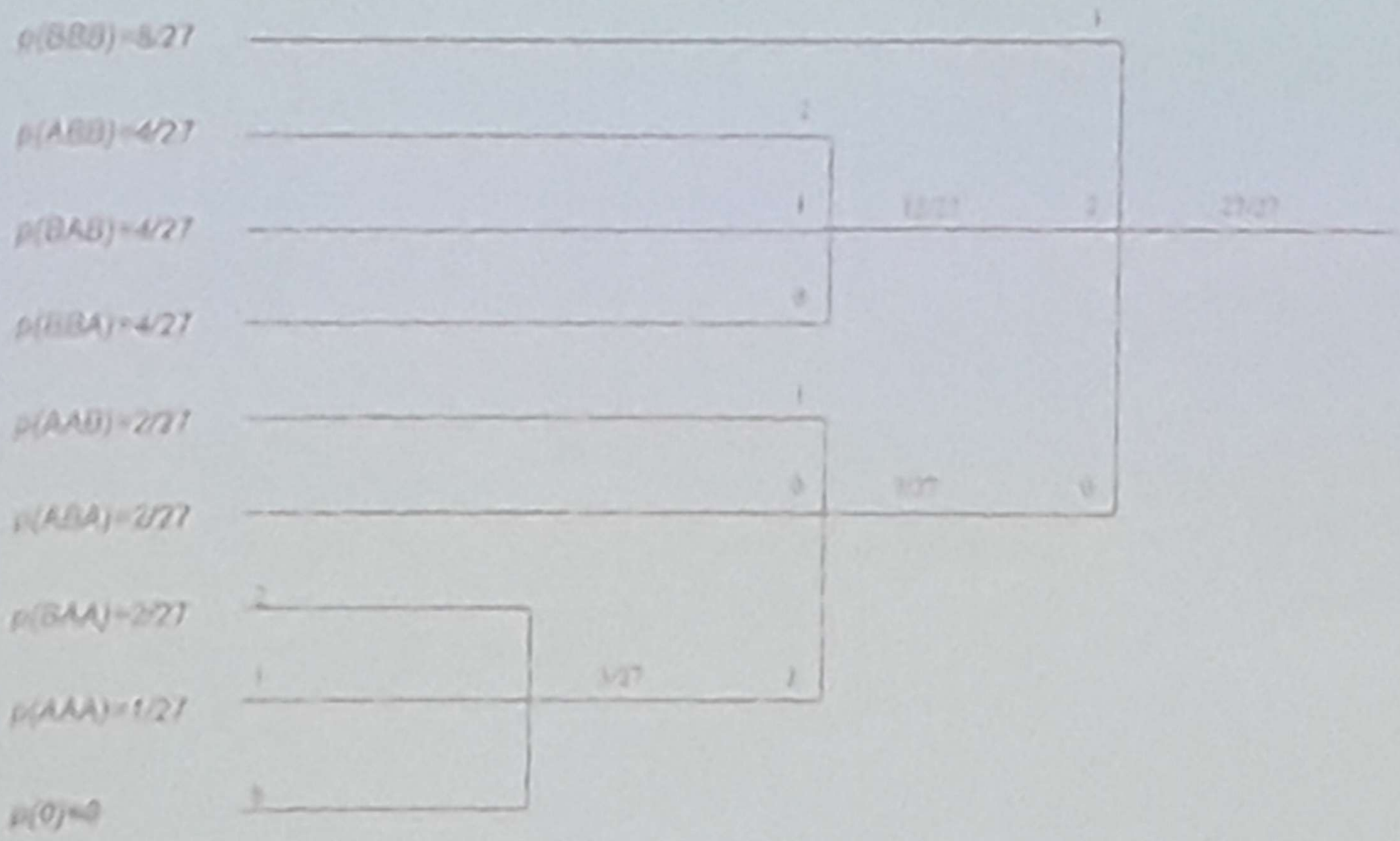
$$k' = \left\lceil \frac{N-1}{d-1} \right\rceil \qquad N' = k'(d-1)+1$$

N – broj simbola

d – baza u kojoj provodi kodiranje

Ako je $N' - N \neq 0$ treba uvesti nove simbole!

Postupak rješavanja



Postupak rješavanja

Huffmanovim kodiranjem smo dobili:

AAA -> 021	BAA -> 022
AAB -> 01	BAB -> 21
ABA -> 00	BBA -> 20
ABB -> 22	BBB -> 1

Postupak rješavanja

Duljina kodne riječi se računa preko formule:

$$L = \sum_{i=1}^{n} p(x_i) l(x_i) = \sum_{i=1}^{n} p_i l_i$$

$L = (1/27)\times 3 + (2/27)\times 2 + (2/27)\times 2 + (4/27)\times 2 + (2/27)\times 3 + (4/27)\times 2 + (4/27)\times 2 + (8/27)\times 1 = 1{,}8148$ [ternarni simbol/simbol]

$L = ?$ [bit/simbol]

Postupak rješavanja

$$\varepsilon_2 = \varepsilon_3$$

$$\frac{H(X)_2}{L_2} = \frac{H(X)_3}{L_3}$$

$$\frac{\sum_{i=1}^{n} p(x_i)\log_2[p(x_i)]}{L_2} = \frac{\sum_{i=1}^{n} p(x_i)\log_3[p(x_i)]}{L_3}$$

$$\log_a(x) = \frac{\log_b x}{\log_b a}$$

$$L_2 \times \log_{10} 2 = L_3 \times \log_{10} 3$$

$$L_2 = L_3 \times \frac{\log_{10} 3}{\log_{10} 2} = 2{,}8764[\frac{\text{bit}}{\text{simbol}}]$$

Postupak rješavanja

$$H(x_i) = -\sum_{i=1}^{n} p_i \log_2 \left(p_i \right) [\text{bit/simbol}]$$

$$H(X) = -[(1/27) \times \log_2(1/27) + 3 \times (2/27) \times \log_2(2/27) + 3 \times (4/27) \times \log_2(4/27) + (8/27) \times \log_2(8/27)]$$

$$H(X) = 2{,}7549 [\text{bit/simbol}]$$

Postupak rješavanja

Efikasnost se računa prema formuli:

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

$$\varepsilon = 0{,}9577$$

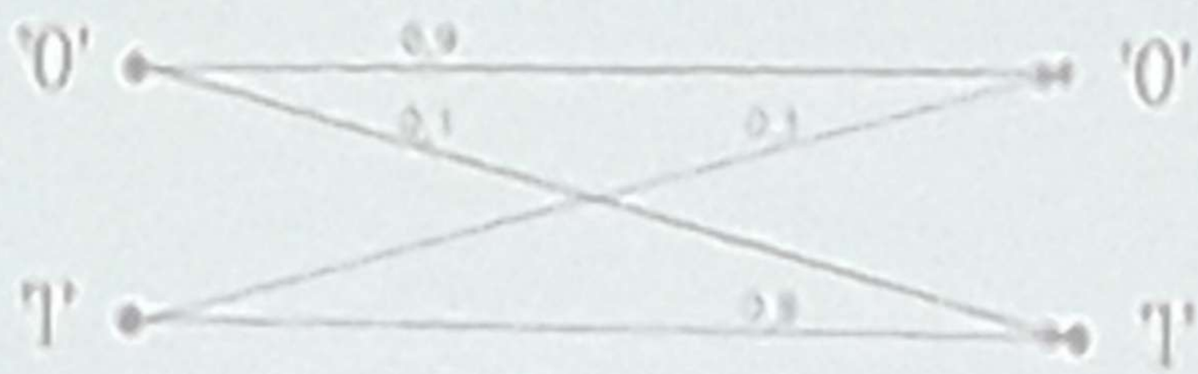
Konkretno rješenje zadatka

AAA → 021	BAA → 022
AAB → 01	BAB → 21
ABA → 00	BBA → 20
ABB → 22	BBB → 1

$$\varepsilon = 0{,}9577$$

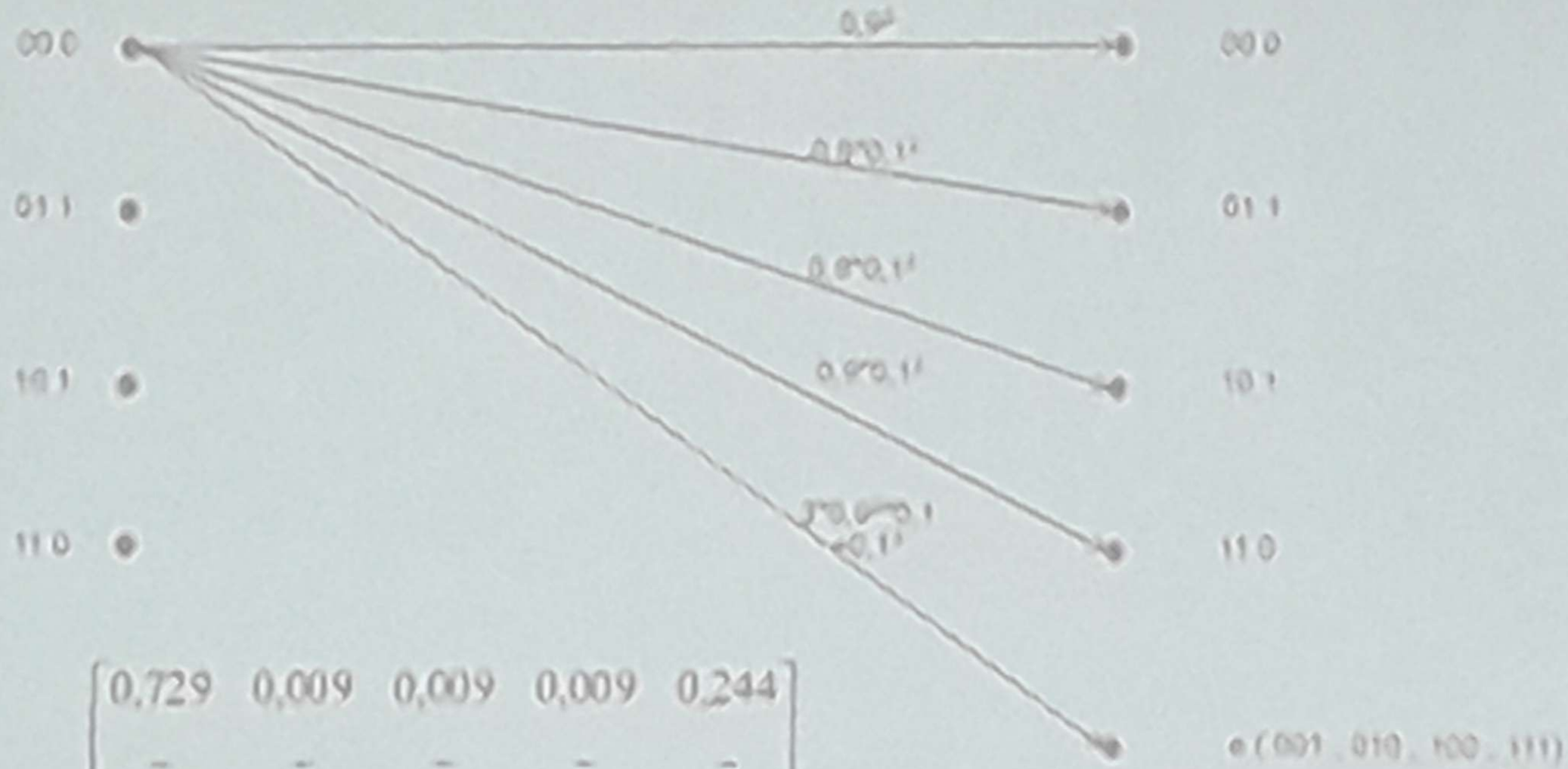
Tekst zadatka

- Četiri poruke, generirane iz skupa od četiriju jednako vjerojatnih simbola $X=\{x_1,\ldots,x_4\}$, kodirane binarnim kodom (x_1-'00'; x_2-'01; x_3-'10'; x_4-'11'), prenose se binarnim simetričnim kanalom (slika 1.). Izračunajte transformaciju u kanalu ako se u prijenosu kao zaštita poruka uvede jedan paritetni bit (parni paritet).



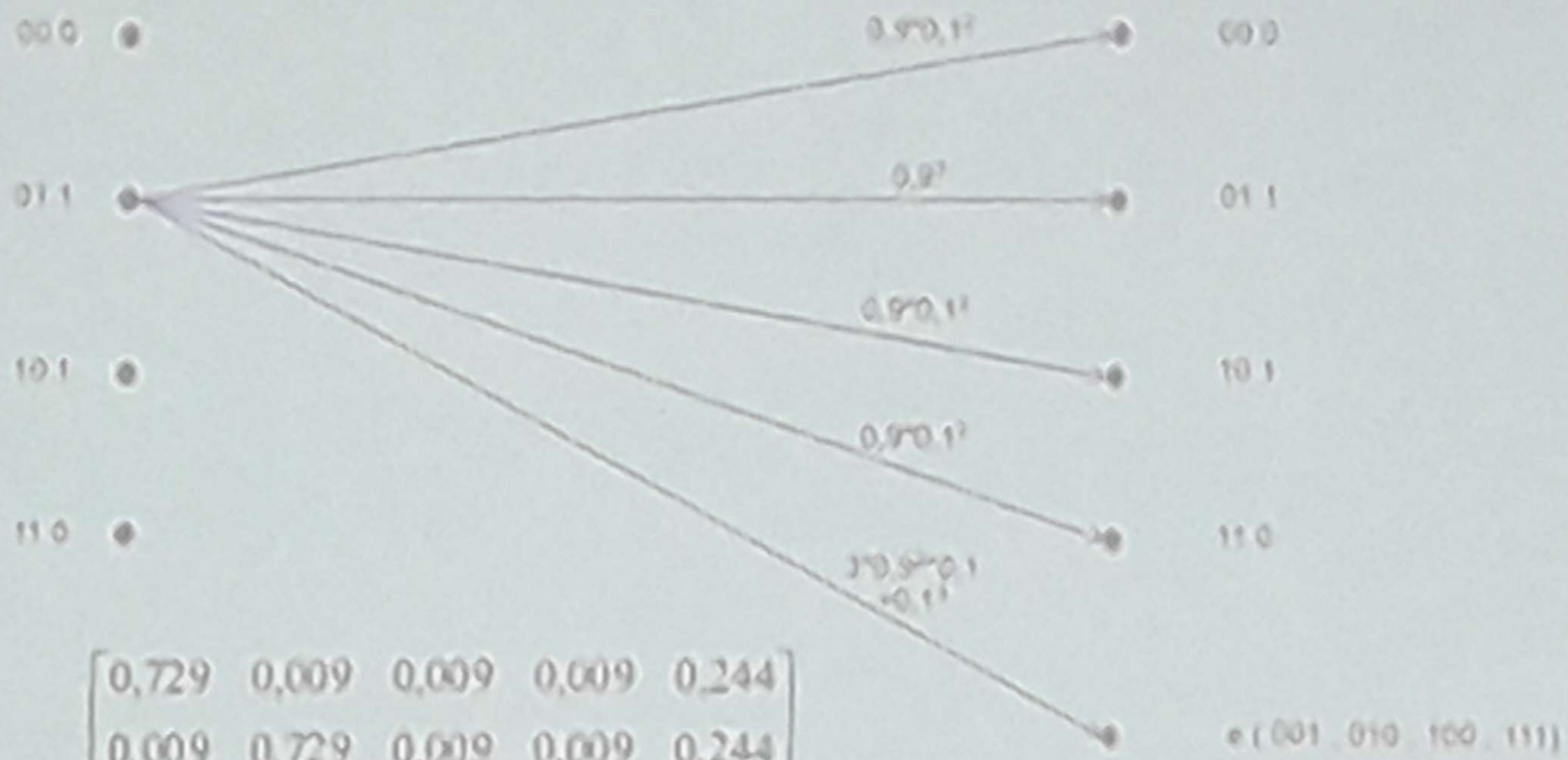
Slika 1.

Postupak rješavanja



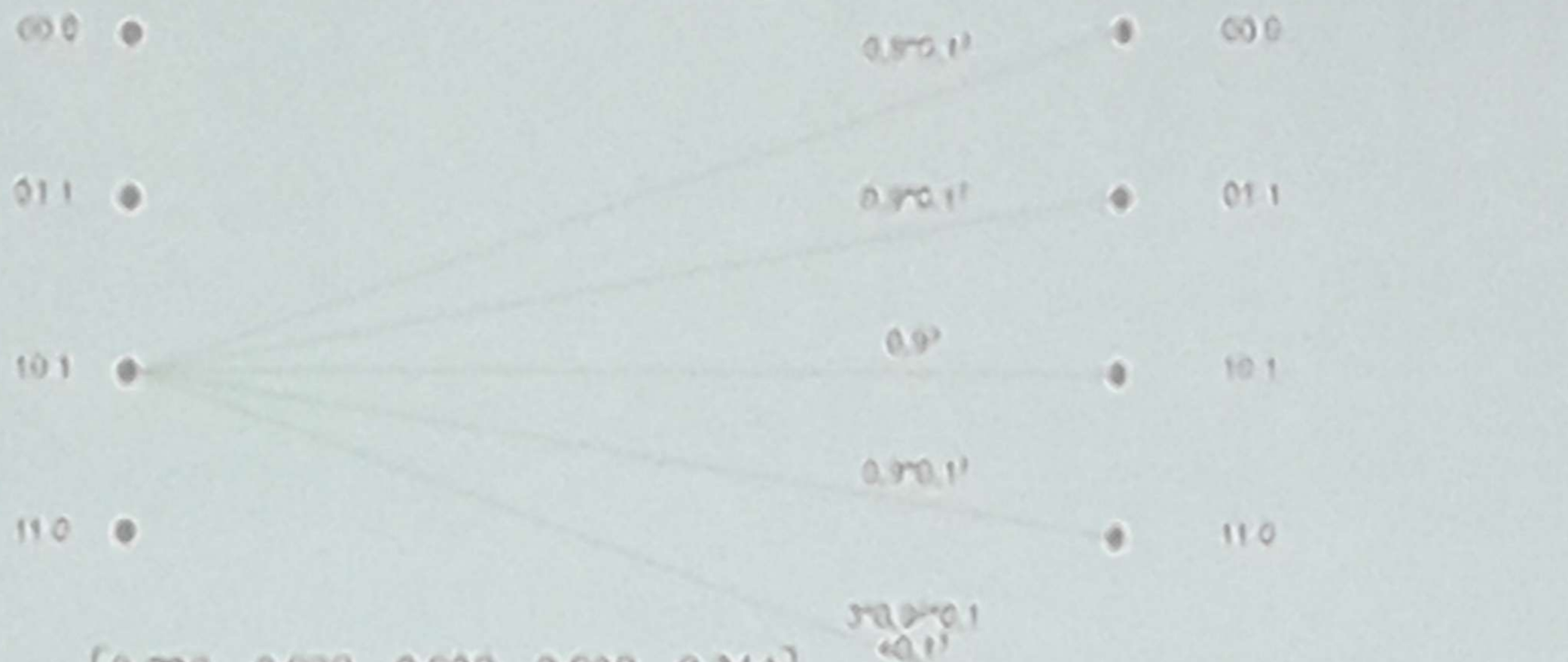
$$[p(y_i \mid x_i)] \begin{bmatrix} 0{,}729 & 0{,}009 & 0{,}009 & 0{,}009 & 0{,}244 \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \end{bmatrix}$$

Postupak rješavanja



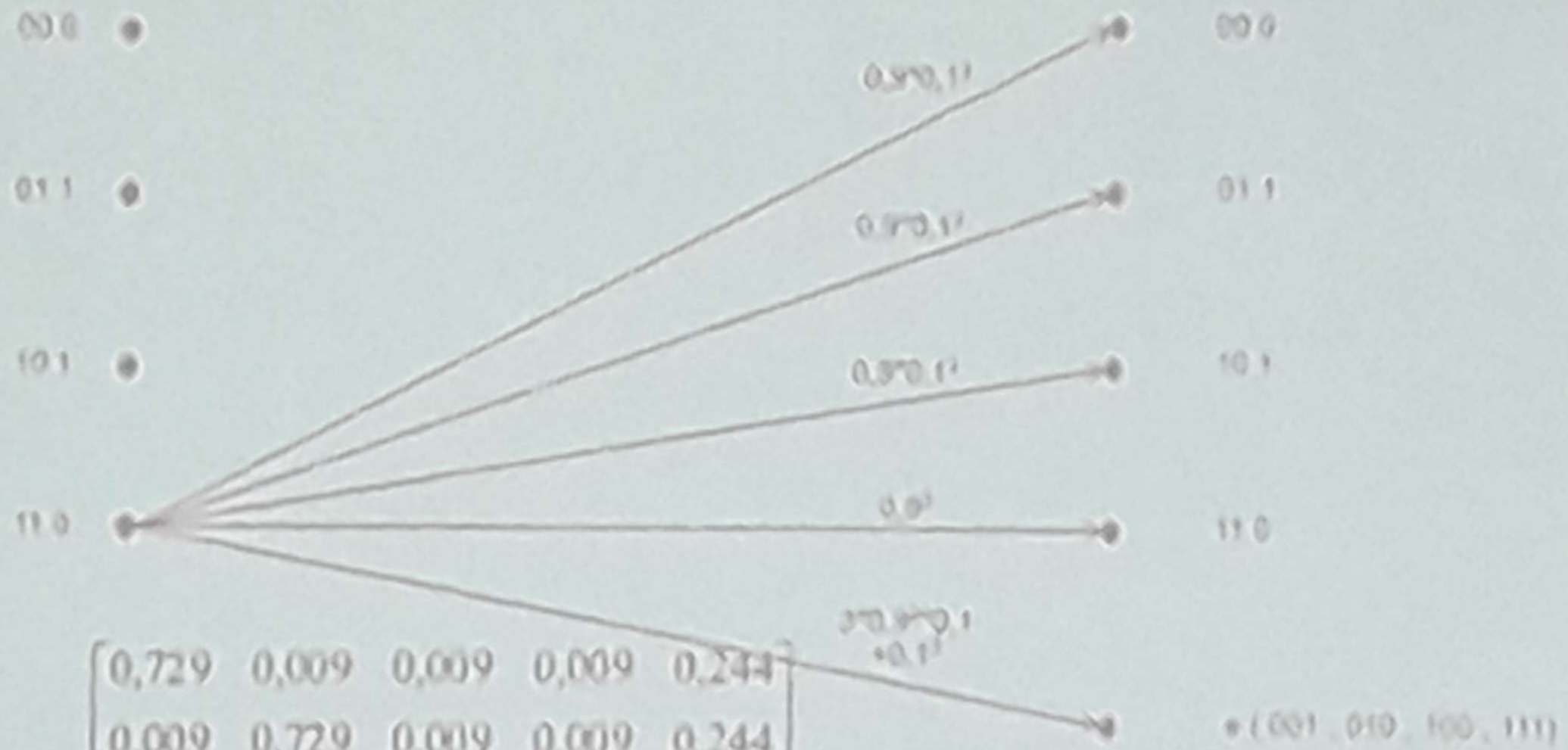
$$\lfloor p(y_j \mid x_i) \rfloor \begin{bmatrix} 0{,}729 & 0{,}009 & 0{,}009 & 0{,}009 & 0{,}244 \\ 0{,}009 & 0{,}729 & 0{,}009 & 0{,}009 & 0{,}244 \\ - & - & - & - & - \\ - & - & - & - & - \end{bmatrix}$$

Postupak rješavanja



$$[p(y_i \mid x_i)] \begin{bmatrix} 0{,}729 & 0{,}009 & 0{,}009 & 0{,}009 & 0{,}244 \\ 0{,}009 & 0{,}729 & 0{,}009 & 0{,}009 & 0{,}244 \\ 0{,}009 & 0{,}009 & 0{,}729 & 0{,}009 & 0{,}244 \\ - & - & - & - & - \end{bmatrix}$$

Postupak rješavanja



$$\{p(y_i \mid x_k)\} \begin{bmatrix} 0{,}729 & 0{,}009 & 0{,}009 & 0{,}009 & 0{,}244 \\ 0{,}009 & 0{,}729 & 0{,}009 & 0{,}009 & 0{,}244 \\ 0{,}009 & 0{,}009 & 0{,}729 & 0{,}009 & 0{,}244 \\ 0{,}009 & 0{,}009 & 0{,}009 & 0{,}729 & 0{,}244 \end{bmatrix}$$

Postupak rješavanja

$$[p(x_i, y_j)] = [p(x_i)][p(y_j \mid x_i)] =$$

$$= \begin{bmatrix} 0{,}25 & 0 & 0 & 0 \\ 0 & 0{,}25 & 0 & 0 \\ 0 & 0 & 0{,}25 & 0 \\ 0 & 0 & 0 & 0{,}25 \end{bmatrix} \begin{bmatrix} 0{,}729 & 0{,}009 & 0{,}009 & 0{,}009 & 0{,}244 \\ 0{,}009 & 0{,}729 & 0{,}009 & 0{,}009 & 0{,}244 \\ 0{,}009 & 0{,}009 & 0{,}729 & 0{,}009 & 0{,}244 \\ 0{,}009 & 0{,}009 & 0{,}009 & 0{,}729 & 0{,}244 \end{bmatrix}$$

$$[p(x_i, y_j)] = \begin{bmatrix} 0{,}18225 & 0{,}00225 & 0{,}00225 & 0{,}00225 & 0{,}061 \\ 0{,}00225 & 0{,}18225 & 0{,}00225 & 0{,}00225 & 0{,}061 \\ 0{,}00225 & 0{,}00225 & 0{,}18225 & 0{,}00225 & 0{,}061 \\ 0{,}00225 & 0{,}00225 & 0{,}00225 & 0{,}18225 & 0{,}061 \end{bmatrix}$$

$$\begin{array}{ccccc} \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ y_1 & y_2 & y_3 & y_4 & y_5 \end{array}$$

Postupak rješavanja

$$[p(x_i)] = [0{,}25 \quad 0{,}25 \quad 0{,}25 \quad 0{,}25]$$

$$H(X) = -\sum_{i=1}^{4} p(x_i)\log_2(p(x_i)) = 2\ \frac{bit}{simbol}$$

$$[p(y_j)] = [0{,}189 \quad 0{,}189 \quad 0{,}189 \quad 0{,}189 \quad 0{,}244]$$

$$H(Y) = -\sum_{j=1}^{5} p(y_j)\log_2(p(y_j)) = 2{,}31363\ \frac{bit}{simbol}$$

$$[p(x_i, y_j)] = \begin{bmatrix} 0{,}18225 & 0{,}00225 & 0{,}00225 & 0{,}00225 & 0{,}061 \\ 0{,}00225 & 0{,}18225 & 0{,}00225 & 0{,}00225 & 0{,}061 \\ 0{,}00225 & 0{,}00225 & 0{,}18225 & 0{,}00225 & 0{,}061 \\ 0{,}00225 & 0{,}00225 & 0{,}00225 & 0{,}18225 & 0{,}061 \end{bmatrix}$$

$$H(X,Y) = -\sum_{i=1}^{4}\sum_{j=1}^{5} p(x_i, y_j)\log_2(p(x_i, y_j)) = 3{,}01247\ \frac{bit}{simbol}$$

Konačno rješenje zadatka

$$I(X;Y) = H(X) + H(Y) - H(X,Y) =$$

$$= 2 + 2{,}31363 - 3{,}01247 =$$

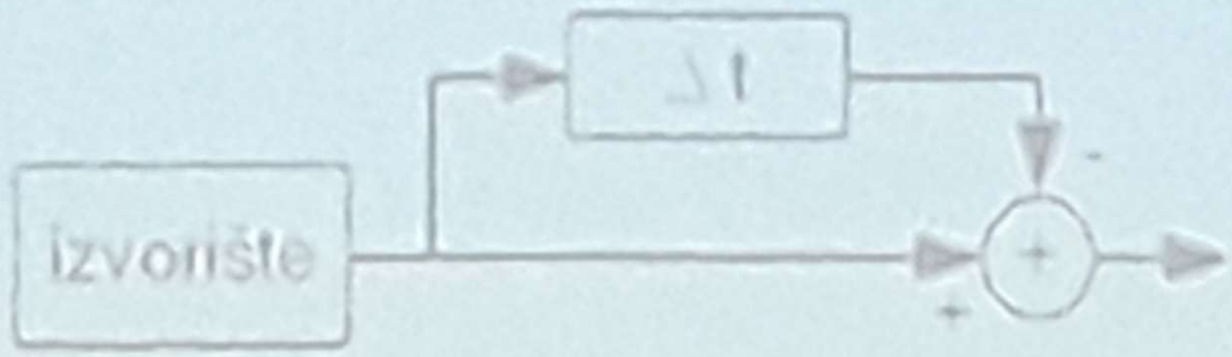
$$= 1{,}3012 \frac{bit}{simbol}$$

Zadatak:

- Diskretno izvorište generira simbole iz skupa simbola $X = \{4, 5, 6\}$. Statističke veze između dva uzastopna simbola koja izvorište generira dane su preko matrice združenih vjerojatnosti $[p(x_i, x_j)]$:

$$[p(x_i, x_j)] = \begin{bmatrix} 0.1172 & 0.1172 & 0.1563 \\ 0.0713 & 0.2138 & 0.0713 \\ 0.2023 & 0.0253 & 0.0253 \end{bmatrix}$$

- Na izvorište je priključen sklop (slika) koji na izlazu daje razliku između dva uzastopna simbola generirana na izvorištu



- Odredite entropiju skupa simbola na izlazu sklopa sa slike!

Postupak rješavanja

- Sklop na izlazu daje razliku između dva uzastopna simbola generirana na izvorištu. Uzmemo li u obzir skup simbola izvorišta, X = {4, 5, 6} , svi mogući ishodi biti će unutar skupa Y= {0, 1, -1, 2, -2}, odnosno broj mogućih ishoda je 5. Ukupno će biti 3*3 = 9 mogućih parova iz skupa X, od kojih će svaki događaj (x_i, y_j) imati svoju vjerojatnost:

(x_i, y_i)	$p(x_i, y_i)$	$x_i - y_i$
4,4	0.1172	0
4,5	0.1172	-1
4,6	0.1563	-2
5,4	0.0713	1
5,5	0.2138	0
5,6	0.0713	-1
6,4	0.2023	2
6,5	0.0253	1
6,6	0.0253	0

- Definirajmo novu slučajnu varijablu Y koja se odnosi na sve moguće ishode (razlike elemenata x_i) na izlazu, odnosno:

$$Y \sim \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 0.1563 & 0.1172 + 0.0713 & 0.1172 + 0.2138 + 0.0253 & 0.0713 + 0.0253 & 0.2023 \end{pmatrix}$$

$$Y \sim \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 0.1563 & 0.1885 & 0.3563 & 0.0966 & 0.2023 \end{pmatrix}$$

Konačno rješenje zadatka

- Entropiju H(Y) računamo formulom:

$$H(Y) = -\sum_{j=1}^{n} p(y_j)\log_2 p(y_j) =$$

$$= -(0.1563\log_2 0.1563 + 0.1885\log_2 0.1885 +$$

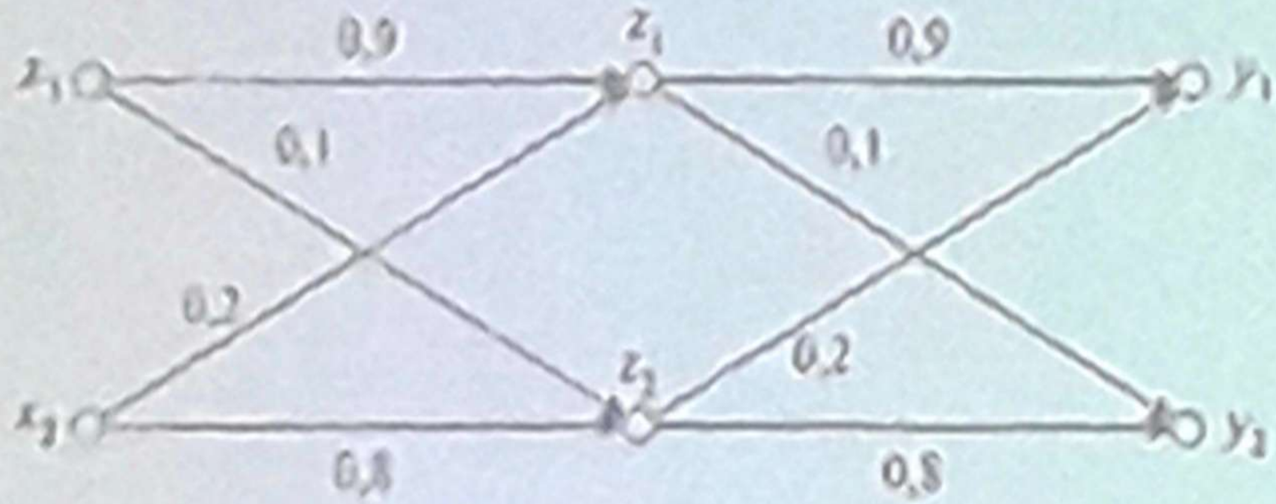
$$+ 0.3563\log_2 0.3563 + 0.0966\log_2 0.0966 +$$

$$+ 0.2023\log_2 0.2023) =$$

$$= 2.195 bit / simbol$$

Tekst zadatka

- Dva binarna kanala serijski su povezana kako je to predočeno na slici:



- Odredite srednji uzajamni sadržaj informacije u sustavu kanala ako je $p(x_1) = p(x_2) = 0,5$.

Postupak rješavanja

- Zadane vjerojatnosti pojavljivanja simbola:

$$p(x_1) = 0{,}5$$
$$p(x_2) = 0{,}5$$

Postupak rješavanja

- Iz kanala možemo iščitati matricu uvjetnih vjerojatnosti:

$$P(Z \mid X) = \begin{bmatrix} 0{,}9 & 0{,}1 \\ 0{,}2 & 0{,}8 \end{bmatrix}$$

- Iz te matrice možemo odrediti združenu vjerojatnost od (X,Z):

$$P(Z \mid X) = \frac{P(X,Z)}{P(X)}$$

$$P(X,Z) = P(X) \cdot P(Z \mid X)$$

Postupak rješavanja

- Suma stupaca u matrici P(X,Z) su vjerojatnosti z_1 i z_2:

$$P(X,Z) = \begin{bmatrix} 0{,}45 & 0{,}05 \\ 0{,}1 & 0{,}4 \end{bmatrix}$$

$$p(z_1) = 0{,}45 + 0{,}1 = 0{,}55$$

$$p(z_2) = 0{,}05 + 0{,}4 = 0{,}45$$

$$P(Z) = \begin{bmatrix} 0{,}55 & 0 \\ 0 & 0{,}45 \end{bmatrix}$$

Postupak rješavanja

- Matrica združenih vjerojatnosti (Z,Y):

$$P(Z,Y) = P(Z) \cdot P(Y \mid Z) = \begin{bmatrix} 0{,}495 & 0{,}055 \\ 0{,}09 & 0{,}36 \end{bmatrix}$$

- Suma stupaca u matrici P(Z,Y) su vjerojatnosti y1 i y2:

$$P(Y) = \begin{bmatrix} 0{,}585 & 0 \\ 0 & 0{,}415 \end{bmatrix}$$

Postupak rješavanja

- Vjerojatnost pojavljivanja y ako se na ulazu pojavio x:

$$P(Y \mid X) = \begin{bmatrix} 0{,}9 \cdot 0{,}9 + 0{,}1 \cdot 0{,}2 & 0{,}9 \cdot 0{,}1 + 0{,}1 \cdot 0{,}8 \\ 0{,}8 \cdot 0{,}2 + 0{,}2 \cdot 0{,}9 & 0{,}8 \cdot 0{,}8 + 0{,}2 \cdot 0{,}1 \end{bmatrix}$$

$$P(Y \mid X) = \begin{bmatrix} 0{,}83 & 0{,}17 \\ 0{,}34 & 0{,}66 \end{bmatrix}$$

Postupak rješavanja

- Matrica združenih vjerojatnosti (X,Y):

$$P(X,Y) = P(X) \cdot P(Y \mid X) = \begin{bmatrix} 0{,}415 & 0{,}085 \\ 0{,}17 & 0{,}33 \end{bmatrix}$$

Postupak rješavanja

- Entropija se računa kao:

$$H(A) = -\sum_{i=1}^{n} p(a_i) \cdot \log_2 \{p(a_i)\} \, [bit / simbol]$$

$$H(X) = 1$$

$$H(Y) = 0{,}97905$$

$$H(X,Y) = 1{,}7913$$

Konačno rješenje zadatka

- Srednji uzajamni sadržaj informacije je:

$$I(X;Y) = H(X) + H(Y) - H(X,Y)\,[bit/simbol]$$

$$I(X;Y) = 0{,}1878\,[bit/simbol]$$

Tekst zadatka

Neki komunikacijski kanal zadan je niže navedenom matricom [p(Y|X)], uz nepoznate vjerojatnosti pojavljivanja ulaznog skupa simbola X, pri čemu koristimo oznake pi = p(xi), i = 1, 2, …, 6. Ako je p1 = 1/16, p2 = 1/8, odredite vjerojatnost p3 za koju je moguće postići kapacitet kanala.

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

Postupak rješavanja

$$C = \max_{\{p(x_i)\}} I(X;Y) = \max_{\{p(x_i)\}} \left[H(Y) - H(Y \mid X) \right] \frac{bit}{simbol}$$

$H(Y) = ?$ $H(Y|X) = ?$

Tekst zadatka

$$[P(x_i)] = \left[\tfrac{1}{16}, \tfrac{1}{8}, P_3, P_4, P_5, P_6\right]$$

$$[P(Y \mid X)] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Postupak rješavanja

$$[P(X,Y)] = [P(x_i)]^{*}[P(Y \mid X)]$$

$$[P(X,Y)] = \begin{bmatrix} \frac{1}{16} & 0 & 0 & 0 \\ \frac{1}{8} & 0 & 0 & 0 \\ p_3 & 0 & 0 & 0 \\ 0 & p_4 & 0 & 0 \\ 0 & 0 & p_5 & 0 \\ 0 & 0 & 0 & p_6 \end{bmatrix}$$

$$[P(y_j)] = \begin{bmatrix} \frac{3}{16} + p_3 & p_4 & p_5 & p_6 \end{bmatrix}$$

Postupak rješavanja

$$C = \max_{\{p(x_i)\}} I(X;Y) = \max_{\{p(x_i)\}} \left[H(Y) - H(Y \mid X) \right] \frac{bit}{simbol}$$

$$H(Y) = -\sum_{i=1}^{6} p_i \log_2(p_i) \frac{bit}{simbol}$$

$$H(Y) = -\left(p_1 + \tfrac{1}{16}\right)\log_2\left(p_1 + \tfrac{1}{16}\right) - p_4 \log_2 p_4 - p_5 \log_2 p_5 - p_6 \log_2 p_6 \frac{bit}{simbol}$$

$$H(Y \mid X) = -\sum_{i=1}^{6} \sum_{j=1}^{m} p(x_i, y_j) \log_2 p(y_j \mid x_i) \frac{bit}{simbol}$$

$$H(Y \mid X) = -\left(p_1 \log_2 1 + p_2 \log_2 1 + p_3 \log_2 1 + p_4 \log_4 1 + p_5 \log_2 1 + p_6 \log_2 1\right) = 0 \frac{bit}{simbol}$$

Postupak rješavanja

$$C = \max\left(H(Y) - 0\right) = \log_2(4) = 2\ \frac{\text{bit}}{\text{simbol}}$$

- H(Y) je maksimalna kada su vjerojatnosti pojave simbola na izlazu jednake

Konačno rješenje zadatka

$$p_1 = \frac{1}{4} = 0.25 = p_3 + \frac{1}{16}$$

$$p_3 = \frac{1}{16}$$

Tekst zadatka

- Razmotrite izvor koji generira simbole s_i iz abecede S opisane tablicom. Vjerojatnosti $p(s_i)$ predstavljaju vjerojatnosti pojavljivanja pojedinih simbola na izlazu izvora. Pretpostavite da izvor generira poruku "bad_dab," (navodnici nisu dio poruke). Optimalno kodirajte zadanu poruku koristeći aritmetičko kodiranje i odredite interval u kojem se nalazi konačan broj čiji binarni ekvivalent predstavlja traženi kod za zadanu poruku. Koliko bita mora sadržavati konačno rješenje pa da binarna kodna riječ zadovoljava svojstvo prefiksnosti koda?

simbol s_i	$p(s_i)$
a	6/15
b	2/15
d	2/15
_	4/15
,	1/15

Postupak rješavanja

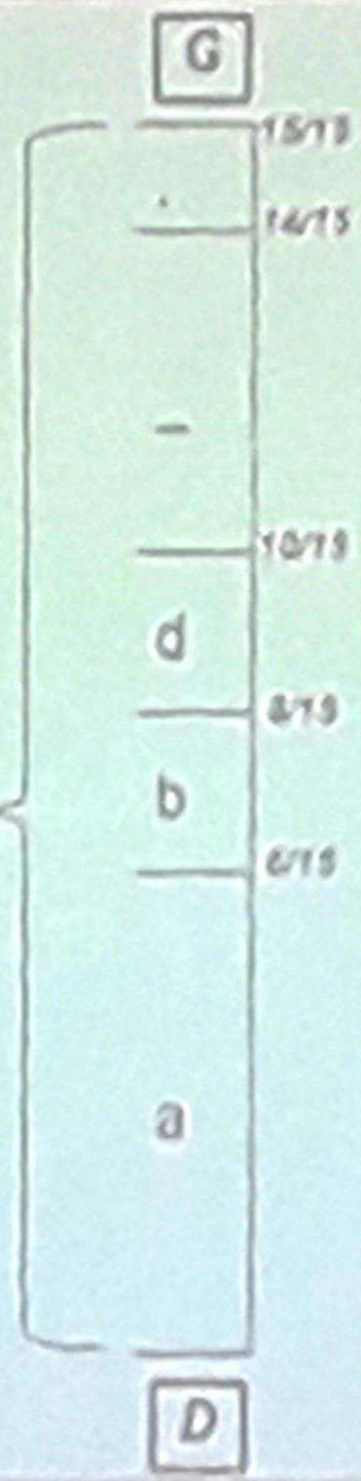
bad_dab,

simbol s_i	$p(s_i)$	Podskupovi (D_i, G_i)
a	6/15	[0, 6/15)
b	2/15	[6/15, 8/15)
d	2/15	[8/15, 10/15)
_	4/15	[10/15, 14/15)
.	1/15	[14/15, 15/15)

$$G = 1$$

$$D = 0$$

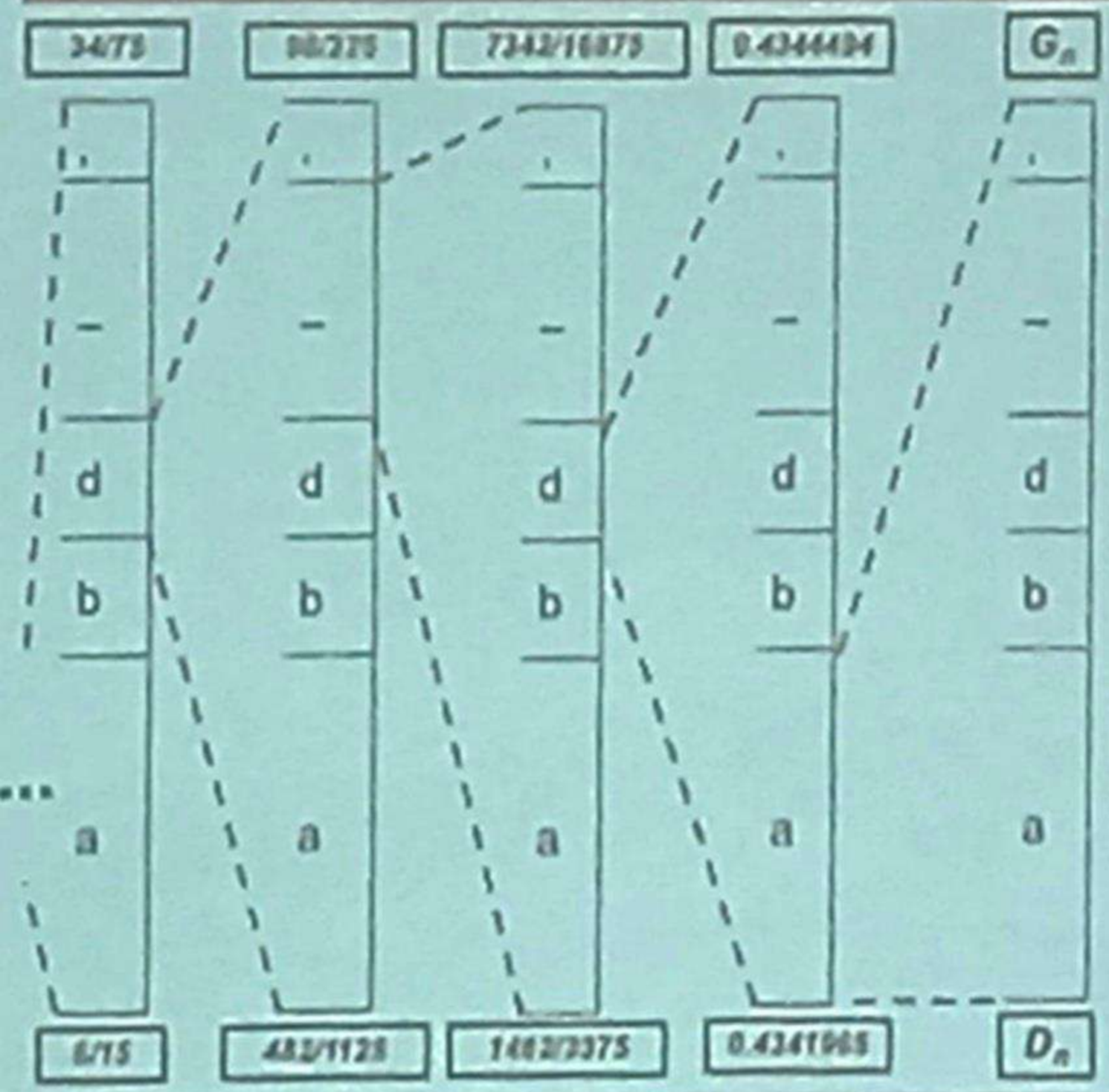
$$L = G - D = 1$$



Postupak rješavanja

[illegible]	[illegible]
[illegible]	[illegible]

bad_dab,



$$L = 0,0002529$$

$$G_n = D + L \cdot G_0$$

$$G_n = 0,4342976$$

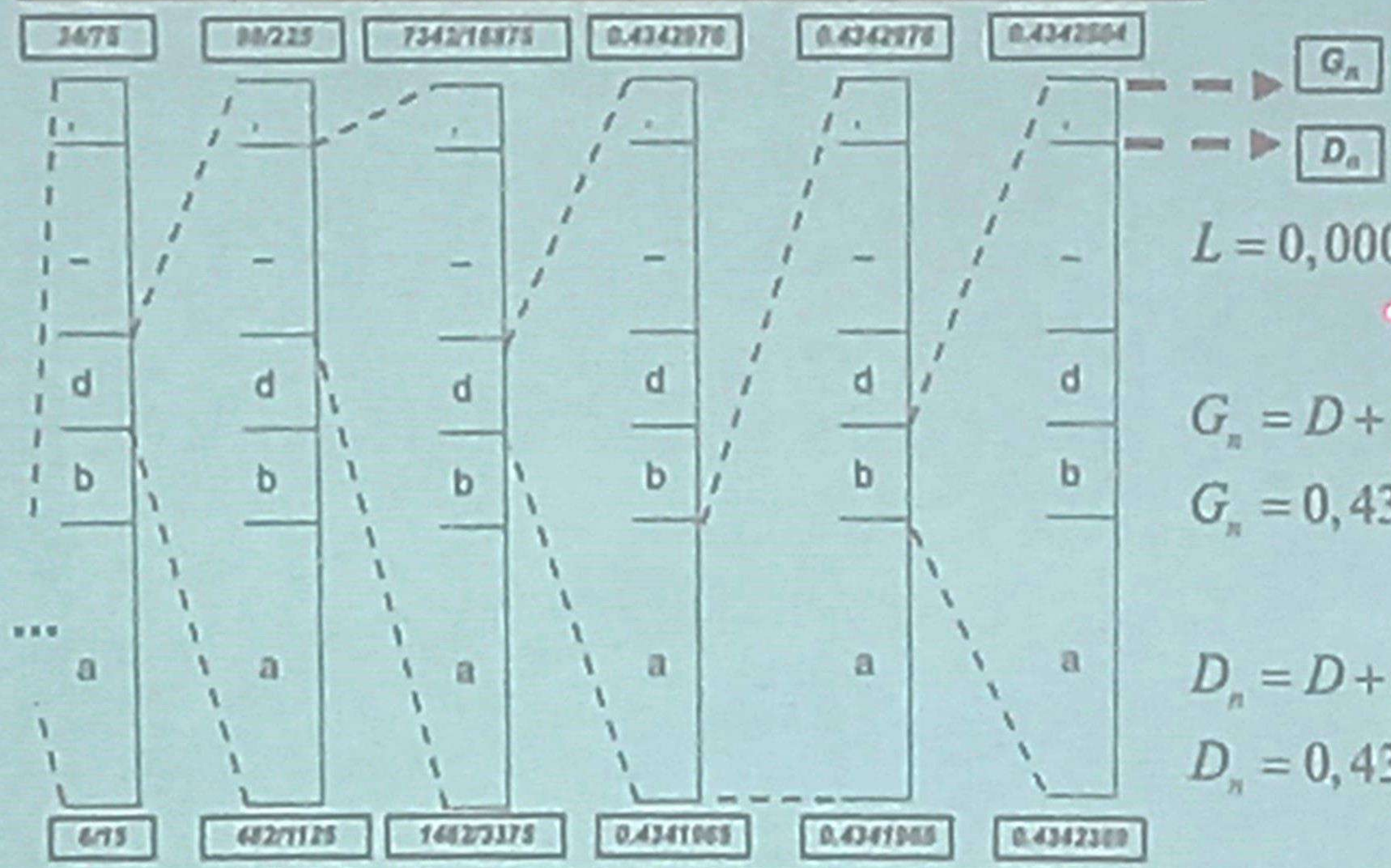
$$D_n = D + L \cdot D_0$$

$$D_n = 0,4341965$$

Postupak rješavanja

simbol	$[D_n, G_n]$
.	[14/15, 15/15]

bad_dab()



$$L = 0,0000135$$

$$G_n = D + L \cdot G_0$$

$$G_n = 0,4342504$$

$$D_n = D + L \cdot D_0$$

$$D_n = 0,4342495$$

Tekst zadatka

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.

Postupak rješavanja

- Na samom početku algoritma postoji samo okvir za kodiranje

aaaabbbccd*

- S obzirom da nema posmičnog prozora, nemamo gdje tražiti simbole,stoga je trojka za ovaj korak:

$$(0, 0, a)$$

Postupak rješavanja

- U sljedećem koraku dodajemo posmični prozor na prvi znak slijeda(iz posljednje trojke), a početak prozora za kodiranje pomičemo na prvi sljedeći znak:

$$\boxed{a}aaabb|bccd*$$

- Pomičemo se 1 korak(na početak posmičnog prozora), možemo triput pročitati znak 'a' jer se u prozoru za kodiranje nalaze tri znaka 'a' jedan za drugim. Sljedeći znak na koji se pomičemo je znak 'b':

$$(1, 3, b)$$

Postupak rješavanja

- Znak 'b' iz posljednje trojke sada postaje zadnji znak posmičnog prozora, a prozor za kodiranje počinje poslije tog znaka:

aaaab|bbccd*

- Vraćamo se 1 korak na znak 'b' u posmičnom prozoru te vidimo da ga dvaput možemo pročitati pa je sljedeći znak slijeda znak 'c' :

$$(1, 2, c)$$

Konačno rješenje zadatka

- Treća trojka kodirane poruke:

$$(1, 2, c)$$

Tekst zadatka

- 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.
- **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).

Ako izvor izgenerira točno 4951 slovo a, koliko će pri tome biti novih zapisa oblika (indeks, riječ) u rječniku?

Postupak rješavanja

- 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

Postupak rješavanja

- Broj uzastopnih slova **a** (lijevi stupac tablice), *Ni*(**a**), se povećava svaki puta za ukupan broj novih zapisa, *Xi*, tj. vrijedi:

$$N_i(\mathbf{a}) - N_{i-1}(\mathbf{a}) = Xi, \text{ za } i >= 2$$

Postupak rješavanja

- za ukupan broj uzastopno generiranih slova a, N_M, vrijedi

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

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

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

Konačno rješenje zadatka

- 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 $M = 100$.
- Ali: na kraju slijeda od 4951 slova "a" broj novih zapisa, x_M, je za jedan manji od M te iznosi 99!

Primjer: kodiranje i dekodiranje kodom LZW

Teorija informacije

Tekst zadatka

Primjer iz knjige "Uvod u teoriju informacije i kodiranje", 2. izdanje, str. 272.

Algoritmom LZW potrebno je prvo kodirati slijed od 9 slovnih simbola unutar navodnika: "A B B A B A B A C". Prije početka kodiranja u rječniku kodera nalaze se tri unosa, indeksirana kao (1), (2), odnosno (3), prikazanih u donjoj tablici Navedeni slijed slova potrebno je kodirati slijedom brojčanih indeksa koji označavaju unose u rječniku kodera, a nakon toga dekodirati primljeni slijed indeksa.

indeks	**riječ**
(1)	A
(2)	B
(3)	C

Postupak kodiranja

Prilikom kodiranja potrebno je primijeniti algoritam LZW, opisan pseudokodom u knjizi na str. 272.

U nastavku su korištene skraćenice: RR – radna riječ, NS – novi simbol, Rj – rječnik, Ul – Ulaz i Iz – izlaz

1. na početku slijeda RR = **A**

   NS = **B**

   RR + NS = AB

   nije u rječniku

   kod za RR = **A** → Iz

   AB → Rj, indeks **(4)**

   RR = B

Ul = **A** **B** B A B A B A C

indeks	riječ
(1)	A
(2)	B
(3)	C
(4)	**AB**

Iz = **(1)**

Postupak kodiranja

2. NS = **B**

RR + NS = BB

nije u rječniku

kod za RR = **B** → Iz

BB → Rj, indeks **(5)**

RR = B

Ul = A B **B** A B A B A C

indeks	riječ
(1)	A
(2)	B
(3)	C
(4)	AB
(5)	**BB**

Iz = (1) **(2)**

Postupak kodiranja

3. NS = **A**

RR + NS = BA

nije u rječniku

kod za RR = **B** → Iz

BA → Rj, indeks **(6)**

RR = A

Ul = A B B **A** B A B A C

indeks	riječ
(1)	A
(2)	B
(3)	C
(4)	AB
(5)	BB
(6)	**BA**

Iz = (1) (2) **(2)**

Postupak kodiranja

4. NS = **B**

RR + NS = AB

u rječniku je

RR = AB

NS = **A**

RR + NS = ABA

nije u rječniku

kod za RR = **AB** → Iz

ABA → Rj, indeks **(7)**

RR = A

Ul = A B B A **B** A B A C

Ul = A B B A B **A** B A C

indeks	riječ
(1)	A
(2)	B
(3)	C
(4)	AB
(5)	BB
(6)	BA
(7)	**ABA**

Iz = (1) (2) (2) **(4)**

Postupak kodiranja

5. NS = **B**
RR + NS = AB
u rječniku je
RR = AB
NS = **A**
RR + NS = ABA
u rječniku je
RR = ABA
NS = **C**
RR + NS = ABAC
nije u rječniku
kod za RR = **ABA** → Iz
ABAC → Rj, indeks **(8)**
RR = C

Ul = A B B A B A **B** A C

Ul = A B B A B A B **A** C

Ul = A B B A B A B A **C**

indeks	riječ
(1)	A
(2)	B
(3)	C
(4)	AB
(5)	BB
(6)	BA
(7)	ABA
(8)	**ABAC**

Iz = (1) (2) (2) (4) **(7)**

Postupak kodiranja

6. nema više novih simbola
   RR + NS = C
   u rječniku je
   kod za RR = **C** → Iz
   nema unosa u rječnik
   KRAJ

Ul = A B B A B A B A C

indeks	riječ
(1)	A
(2)	B
(3)	C
(4)	AB
(5)	BB
(6)	BA
(7)	ABA
(8)	ABAC

Iz = (1) (2) (2) (4) (7) **(3)**

Postupak dekodiranja

1. na početku slijeda RR je prazna

   NS = **(1)**

   Iz = **A**

   RR + NS = A

   A je u Rj

   RR = A

Ul = **(1)** (2) (2) (4) (7) (3)

indeks	riječ
(1)	A
(2)	B
(3)	C

Iz = **A**

Postupak dekodiranja

2. NS = **(2)**

   Iz = **B**

   RR + NS = AB

   nije u Rj

   AB → Rj, indeks **(4)**

   RR = NS = B

Ul = (1) **(2)** (2) (4) (7) (3)

indeks	riječ
(1)	A
(2)	B
(3)	C
(4)	**AB**

Iz = A **B**

Postupak dekodiranja

3. NS = **(2)**

   Iz = **B**

   RR + NS = BB

   nije u Rj

   BB → Rj, indeks **(5)**

   RR = NS = B

Ul = (1) (2) **(2)** (4) (7) (3)

indeks	riječ
(1)	A
(2)	B
(3)	C
(4)	AB
(5)	**BB**

Iz = A B **B**

Postupak dekodiranja

4. NS = **(4)**

   Iz = **AB**

   NS1 = A, NS2 = B

   RR + NS1 = BA

   BA nije u Rj

   BA → Rj, indeks **(6)**

   RR = A

   RR + NS2 = AB

   AB je u Rj

   RR = AB

Ul = (1) (2) (2) **(4)** (7) (3)

indeks	riječ
(1)	A
(2)	B
(3)	C
(4)	AB
(5)	BB
(6)	**BA**

Iz = A B B **A B**

Postupak dekodiranja

5. NS = **(7)** UI = (1) (2) (2) (4) **(7)** (3)

   Iz = **?**

   problem: s ulaza je došao simbol koji još nije upisan u rječnik

   trenutno RR = AB, u rječnik će svakako biti upisana kombinacija RR + x, x je prvi simbol od (7)

   zašto? s obzirom da je AB već u rječniku, sljedeći unos u rječnik koji počinje s AB očito će imati tri slovna simbola

   zaključak: novi i nepoznati simbol (7) imat će tri simbola, xyz

   očito mora vrijediti: ABx = xyz → x = A, y = B i z = x = A

   rezultat: (7) = ABA

Postupak dekodiranja

5. NS = **(7)**
   Iz = **ABA**
   NS1 = A, NS2 = B, NS3 = A
   RR + NS1 = ABA
   ABA nije u Rj
   ABA → Rj, indeks **(7)**
   RR = A
   RR + NS2 = AB
   AB je u Rj
   RR = AB
   RR + NS3 = ABA
   ABA je u Rj
   RR = ABA

indeks	riječ
(1)	A
(2)	B
(3)	C
(4)	AB
(5)	BB
(6)	BA
(7)	**ABA**

Iz = A B B A B **A B A**

Postupak dekodiranja

6. NS = **(3)**

   Iz = **C**

   RR + NS = ABAC

   ABAC nije u Rj

   ABAC → Rj, indeks **(8)**

   RR = C

Ul = (1) (2) (2) (4) (7) **(3)**

indeks	riječ
(1)	A
(2)	B
(3)	C
(4)	AB
(5)	BB
(6)	BA
(7)	ABA
(8)	**ABAC**

Iz = A B B A B A B A **C**

Postupak dekodiranja

7. nema više novih simbola

   RR + NS = C

   C je u Rj

   KRAJ

Ul = (1) (2) (2) (4) (7) (3)

indeks	riječ
(1)	A
(2)	B
(3)	C
(4)	AB
(5)	BB
(6)	BA
(7)	ABA
(8)	ABAC

Iz = A B B A B A B A C

Primjer: Shannon-Fanovo kodiranje ternarnim simbolima

Teorija informacije

Tekst zadatka

Zadan je skup poruka $M = \{m_1, m_2, m_3, m_4, m_5, m_6\}$. Pripadajuće vjerojatnosti zadane su na sljedeći način:

$P(m_1) = 3/8$, $P(m_2) = 1/6$, $P(m_3) = P(m_4) = P(m_5) = 1/8$, $P(m_6) = 1/12$

Zadani skup poruka potrebno je kodirati ternarnim kodom koristeći tehniku Shannon-Fano i simbole iz skupa {0, 1, 2}. Potrebno je odrediti srednju duljinu kodne riječi te efikasnost kodiranja. Također je potrebno odrediti vjerojatnost s kojom će se ternarni simboli pojavljivati u jako dugačkom slučajnom slijedu poruka iz skupa M.

Postupak rješavanja

Prvo je potrebno sve simbole poredati po padajućim vjerojatnostima. Nakon toga potrebno je simbole podijeliti u tri grupe. U svakoj grupi zbroj vjerojatnosti simbola iznosit će P_i, $i \in \{1, 2, 3\}$. Pravilo raspoređivanja u grupe je takvo da zbroj svih apsolutnih razlika oblika:

$|P_i - P_j|$, $i, j \in \{1, 2, 3\}$, $i \neq j$

bude minimalan.

Postupak rješavanja

Nakon što su simboli podijeljeni u grupe, simbolima prve grupe (promatrano od simbola veće vjerojatnosti prema onima manje vjerojatnosti) potrebno je dodijeliti simbol 0, simbolima druge grupe (sljedeće grupe odmah do prve) potrebno je dodijeliti simbol 1, a simbolima posljednje (najniže, tj. treće grupe) potrebno je dodijeliti simbol 2.

Postupak rješavanja

simbol m_i	vjerojatnost $P(m_i)$	ternarni simboli		kodna riječ	duljina kodne riječi, l_i
m_1	3/8	0		0	1
m_2	1/6	1	0	10	2
m_3	1/8	1	1	11	2
m_4	1/8	2	0	20	2
m_5	1/8	2	1	21	2
m_6	1/12	2	2	22	2

srednju duljinu kodne riječi moguće je odrediti kao:

$$\overline{L} = \sum_{i=1}^{6} l_i \cdot P\left(m_i\right) = \frac{39}{24} = 1,625\,\text{ts/poruka}$$

pri čemu skraćenica ts označava ternarne simbole

Postupak rješavanja

Efikasnost dobivenog ternarnog koda određujemo sljedećim izrazom:

$$\varepsilon = \frac{H_3(M)}{\overline{L}} = \frac{-\sum_{i=1}^{6} P(m_i)\log_3 P(m_i)}{\overline{L}} = \frac{-\frac{1}{\log_2 3}\cdot\sum_{i=1}^{6} P(m_i)\log_2 P(m_i)}{\overline{L}} =$$

$$= \frac{H_2(M)}{\overline{L}\cdot\log_2 3} \approx 0,926$$

Postupak rješavanja

Vjerojatnosti pojavljivanja ternarnih simbola u jako dugačkom slučajnom slijedu poruka iz skupa M određujemo sljedećim izrazima:

$$\frac{\sum_{i=1}^{6} l_i P(m_i)}{\sum_{i=1}^{6} l_i P(m_i)} = \frac{\sum_{i=1}^{6} (l_{i0} + l_{i1} + l_{i2}) P(m_i)}{\sum_{i=1}^{6} l_i P(m_i)} =$$

u svakoj ternarnoj kodnoj riječi duljine l_i ima l_{i0} nula, l_{i1} jedinica i l_{i2} dvojki

$$= \frac{\sum_{i=1}^{6} l_{i0} P(m_i) + \sum_{i=1}^{6} l_{i1} P(m_i) + \sum_{i=1}^{6} l_{i2} P(m_i)}{\sum_{i=1}^{6} l_i P(m_i)} =$$

$$= \frac{\sum_{i=1}^{6} l_{i0} P(m_i)}{\sum_{i=1}^{6} l_i P(m_i)} + \frac{\sum_{i=1}^{6} l_{i1} P(m_i)}{\sum_{i=1}^{6} l_i P(m_i)} + \frac{\sum_{i=1}^{6} l_{i2} P(m_i)}{\sum_{i=1}^{6} l_i P(m_i)} = P(0) + P(1) + P(2)$$

$$P(0) = \frac{16}{39}$$

$$P(1) = \frac{13}{39}$$

$$P(2) = \frac{10}{39}$$