1. a) Izvor informacija bez memorije generiše dvije moguće poruke sa vjerovatnoćama pojavljivanja *p* i 1-*p* . Nacrtati zavisnost entropije izvora od *p* i odrediti njenu maksimalnu vrijednost.

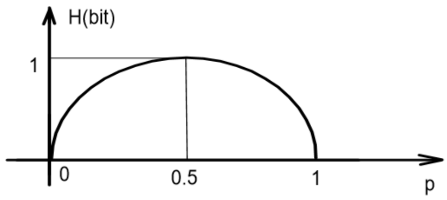
   b) Za izvor informacija sa M mogućih poruka odrediti verovatnoće p_k , k=0,1,2…M-1 , tako da entropija bude maksimalna. Odrediti njenu vrijednost.

<u>Rešenje:</u>

1. a) Entropija izvora sa dva moguća stanja može se napisati u obliku:

$$H = p\log_2 \frac{1}{p} + (1-p)\log_2 \frac{1}{1-p}$$

Maksimalnu vrijednost entropija ima za ono *p* za koje je $\frac{dH}{dp} = 0$, tj. $\log_2 \frac{1-p}{p} = 0$, odakle se dobija *p*=0.5. Maksimum iznosi $H_{\max} = 1 bit / simb$.
Zavisnost entropije od *p* prikazana je na Slici 1.



Slika 1. Zavisnost binarne entropije od vjerovatnoće p.

b) Problem se može riješiti Lagranžovim metodom, po kom se traži maksimum funkcije *F* :

$$F = H + \lambda\left(\sum_{k=0}^{M-1} p_k - 1\right)$$

gde je veličina λ Lagranžov multiplikator. Traži se *k* parcijalnih izvoda:

$\frac{dF}{dp_k} = \log_2 \frac{1}{p_k} - \frac{1}{\ln 2} + \lambda = 0$, sa rešenjima: $p_k = 2^{\lambda - \frac{1}{\ln 2}}$, *k*=0,1,2…M-1.

Pošto p_k u prethodnom izrazu očigledno ne zavisi od *k*, zaključuje se da su sve vrednosti p_k , *k*=0,1,2…M-1, međusobno jednake. Pošto je $\sum_k p_k = 1$, sve vjerovatnoće p_k imaju jednaku vrijednost i ona iznosi 1/M . Lako se izračunava da maksimalna entropija, za ovako određene vjerovatnoće, ima vrijednost:

$$H_{\max} = \log_2 M \ bit / simb$$

2. Posmatra se engleski jezik kao izvor informacija bez memorije a slova kao moguće poruke. Pretpostavlja se da su vjerovatnoće pojavljivanja "čistih" samoglasnika p_1, a ostalih karaktera p_2. Takođe važi i jednakost $p_1 = 5p_2$.

   a) Odrediti količinu informacija koju prenosi jedan samoglasnik.
   b) Odrediti entropiju izvora.

Rešenje:

a) U engleskom jeziku imamo 5 “čistih” samoglasnika, od ukupno 26 slova. Uz uslov koji je dat u tekstu zadatka i ograničenje po kome zbir svih vjerovatnoća mora biti jednak jedinici, dobijaju se dvije jednačine:

$$21p_2 + 5p_1 = 1$$
$$p_1 = 5p_2$$

Lako se izračunava da je: $p_2 \approx 0.0217$ i $p_1 \approx 0.108$.

a) Količina informacije koju nosi svaki samoglasnik je:

$$I_1 = -\log_2(p_1) = 3.21bit$$

Količina informacije za druga slova engleskog alfabeta u ovom slučaju je:

$$I_2 = -\log_2(p_2) = 5.526bit$$

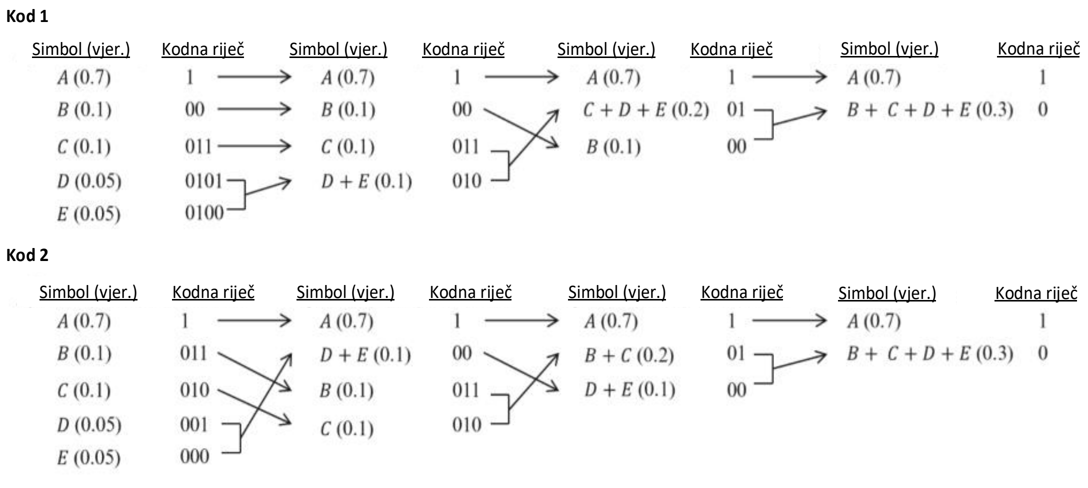
b) Entropija izvora iznosi:

$$\mathrm{H}{=}5\mathrm{p}_1 I_1 + 21p_2 I_2 = 25p_2 I_1 + 21p_2 I_2 = 4.251bit/simb$$

Stvarna entropija izvora je manja zbog postojanja memorije, odnosno zavisnosti između susjednih slova u tekstu.

3. Alfabet DMS izvora sastoji se od 5 simbola A, B, C, D, i E čije su vjerovatnoće 0.7, 0.1, 0.1, 0.05 i 0.05, respektivno. Pod pretpostavkom da se simboli kodiraju individualno (jedan po jedan), dizajnirati dva Hafmanova koda različitih varijansi i dokazati da oba koda imaju isti prosječni broj bita po simbolu.

Rešenje:



Slika 2. Hafmanov kod.

Na Slici 2 su prikazana dva Hafmanov stabla i tabela kodnih riječi za svako stablo. Prosječna dužina kodne riječi za oba Hafmanova koda je:

$$\overline{L}_1 = 0.7 \cdot 1 + 0.1 \cdot 2 + 0.1 \cdot 3 + 2 \cdot 0.05 \cdot 4 = 1.6\, bita$$
$$\overline{L}_2 = 0.7 \cdot 1 + (0.1 + 0.1 + 0.05 + 0.05) \cdot 3 = 1.6\, bita$$

Kao što je očekivano, oba Hafmanova koda imaju istu prosječnu dužinu kodne riječi. Međutim varijanse su različite:

$$\overline{\sigma}_1 = 0.7 \cdot (1-1.6)^2 + 0.1 \cdot (2-1.6)^2 + 0.1 \cdot (3-1.6)^2 + 2 \cdot 0.05 \cdot (4-1.6)^2 = 1.04$$
$$\overline{\sigma}_2 = 0.7 \cdot (1-1.6)^2 + (0.1 + 0.1 + 0.05 + 0.05) \cdot (3-1.6)^2 = 0.84$$

Hafmanov kod ima najveću efikasnost (prosječna dužina kodne riječi jednaka je entropiji) kada se je vjerovatnoća svakog simbola cjelobrojni stepen od ½.

4. Alfabet DMS izvora sastoji se od dva simbola A i B čije su vjerovatnoće 0.9 i 0.1, respektivno. Pronaći entropiju izvora. Dizajnirati Hafmanov kod za ovaj izvor i odrediti prosječni broj bita po simbiolu ukoliko su simboli kodirani:
   a) jedan po jedan (individualno),
   b) dva u isto vrijeme,
   c) tri u isto vrijeme,
   d) četiri u isto vrijeme.

Uporediti prosječnu dužinu kodne riječe u svakom od ova četiri slučaja sa entropijom.

<u>Rešenje</u>:

Entropija izvora je:

$$H = -0.1 \log_2 0.1 - 0.9 \log_2 0.9 = 0.469$$

Na Slici 3 su prikazane Hafmanove kodne riječi za slučajeve kada je primijenjen postupak proširenog kodiranja reda *n=1,2,3,4.*

a) Kada je n=1, prosječni broj bita po simbolu je:

   $$\overline{L}_1 = 1 \cdot 0.9 + 1 \cdot 0.1 = 1\, bit / simb$$

b) Kada se kodiraju dva simbola u isto vrijeme, važi:
   $\overline{L}_2 = 1 \cdot 0.81 + 2 \cdot 0.09 + 3 \cdot 0.09 + 3 \cdot 0.01 = 1.29$ bita za dva simbola, odnosno 0.645 bita po simbolu.

c) Kada se kodiraju tri simbola u isto vrijeme, važi:
   $\overline{L}_3 = 1 \cdot 0.729 + 3 \cdot 3 \cdot 0.081 + 3 \cdot 5 \cdot 0.009 + 5 \cdot 0.001 = 1.598$ bita za tri simbola, ili 0.533 bita po simbolu.

d) Kada se kodiraju četiri simbola u isto vrijeme, važi:

$\bar{L}_4 = 1\cdot 0.6561 + 3\cdot 3\cdot 0.0729 + 4\cdot 0.0729 + 1\cdot 6\cdot 0.0081 + 5\cdot 7\cdot 0.0081 + 3\cdot 9\cdot 0.0009 + 10\cdot 0.0009 + 10\cdot 0.0001 = 1.9702$ bita za 4 simbola, odnosno 0.493 bita po simbolu.

(a)

Simbol	Vjerovatnoća	Kodna riječ
A	0.9	1
B	0.1	0

(b)

Simbol	Vjerovatnoća	Kodna riječ
AA	0.81	1
AB	0.09	01
BA	0.09	001
BB	0.01	000

(c)

Simbol	Vjerovatnoća	Kodna riječ
AAA	0.729	1
AAB	0.081	001
BAA	0.081	000
ABA	0.081	011
BBA	0.009	01011
BAB	0.009	01010
ABB	0.009	01001
BBB	0.001	01000

(d)

Simbol	Vjerovatnoća	Kodna riječ
$AAAA$	0.6561	1
$AAAB$	0.0729	011
$AABA$	0.0729	010
$ABAA$	0.0729	001
$BAAA$	0.0729	0000
$AABB$	0.0081	000111
$ABAB$	0.0081	0001101
$BAAB$	0.0081	0001100
$ABBA$	0.0081	0001011
$BABA$	0.0081	0001010
$BBAA$	0.0081	0001001
$ABBB$	0.0009	000100011
$BABB$	0.0009	000100010
$BBAB$	0.0009	000100001
$BBBA$	0.0009	0001000001
$BBBB$	0.0001	0001000000

Slika 3.

Kodiranje na nivou više simbola instovremeno rezultuje manjim prosječnim brojem bita po simbolu. Sa druge strane, komplikuje se proces kodiranja i dekodiranja.