3. Forrás- és hibavédő kódolás

3.1. Feladat

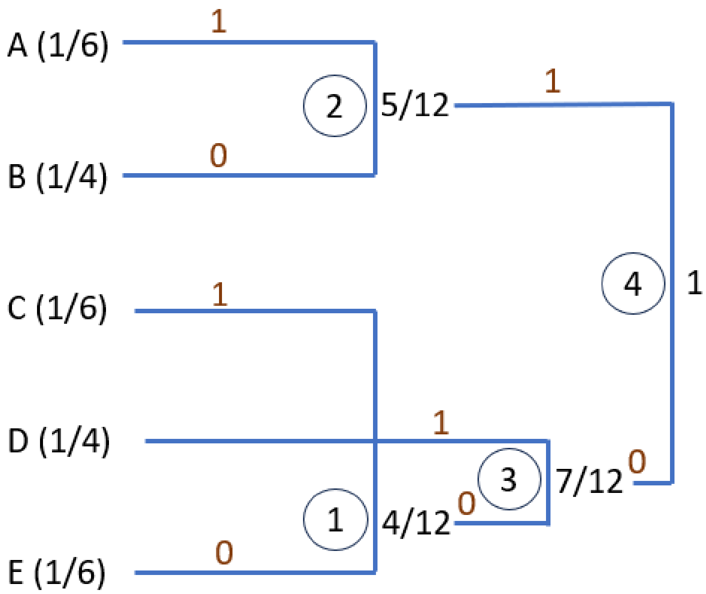
Egy memóriamentes forrás kimenetén öt különböző szimbólum (A, B, C, D és E) jelenik meg az alábbi táblázat szerinti valószínűségekkel (p_x oszlop). Négy különböző kódot készíttettünk a kódolására (α, β, γ, δ).

x	p_x	α	β	γ	δ
A	1/6	11	0	00	10000
B	¼	01	10	01	1000
C	1/6	001	110	11	100
D	¼	10	1110	10	10
E	1/6	000	1111	001	1

a) Mely kód(ok) egyértelműen dekódolhatóak?
b) Mely kód(ok) prefix(mentes) kódok?
c) Adja meg a fenti forrás esetében az alsó korlátot a minimális átlagos kódszóhosszra!
d) A felsoroltak közül mely kód(ok) átlagos kódszóhossza közelíti meg legjobban a teoretikus minimumot?

Megoldás:

a) A δ kódban minden kódszó 1-essel kezdődik, ez a detektálás szabálya, ezért egyértelműen dekódolható a kód. Hasonlóképp indokolható, hogy a β kód is egyértelműen dekódolható. A γ kódra nem teljesül a McMillan-egyenlőtlenség, ezért nem lehet egyértelműen dekódolható. Például a kódolt jel 001001 szakaszát értelmezhetjük 001 001-ként (EE), illetve 00 10 01-ként (ADB) is. Az α kódra teljesül a McMillan-egyenlőtlenség, ezért a dekódolhatóságáról nem mondhatunk róla semmit. Tervezzünk Huffman-kódot a táblázatban szereplő eloszláshoz:



Az így előállított kód pont az α kód. Ezért az α kód is egyértelműen dekódolható.

b) α, β

c) Az alsó korlátot a forrás entrópiája adja meg az alábbi képlet szerint

$$H(P) = \sum_{x \in \{A,B,C,D,E\}} p_x \cdot ld\left(\frac{1}{p_x}\right)$$

Ebbe behelyettesítve most 2.2924 adódik.

d) $L_\alpha = 2.333333$, $L_\beta = 2.833333$, $L_\delta = 3$

Az α kód egy szimbólumra eső átlagos kódszóhossza közelíti meg legjobban a teoretikus minimumot.

3.2. Feladat

Egy diszkrét, emlékezetnélküli, véletlen forrás szimbólumait a

$$P = \{0.5, 0.25, 0.15, 0.1\}$$

valószínűségeloszlás szerint szolgáltatja. Pistike alkot egy kódot, amelyben a kódszavak a szimbólumok fenti sorrendjében a következőek: $(01), (10), (011), (1011)$.

a) Egyértelműen dekódolható-e a fenti kód?

b) A fenti kódhosszúságokkal lehet-e prefixmentes kódot konstruálni?

Megoldás:

a) Tekintve, hogy az első kódszó előtagja a harmadiknak, s a második is a negyediknek, az a gyanúnk ébredhet, hogy az egyértelmű megfejthetőséggel is baj lehet. Valóban, ha például az adó kétszer egymás után a harmadik kódszót adja: 011 011, akkor a kódolt jelnek ezt a szakaszát akár 01 1011-nek is olvashatjuk. Pistike kódja tehát nem egyértelműen megfejthető kód. Megjegyezzük, sok olyan kód létezik, ami noha nem prefix (mentes), ennek ellenére egyértelműen megfejthető.

b) A Kraft egyenlőtlenség teljesül a 2, 2, 3, 4 számokra, hiszen 1/4+1/4+1/8+1/16 < 1, ezért e szóhosszakkal lehet prefixmentes kódot szerkeszteni, pl. pl. (00), (01), (100), (1100).

3.3. Feladat

Egy diszkrét, emlékezetnélküli, véletlen forrás az A, B, C, D és E szimbólumait rendre a

$$P = \{0.35, 0.25, 0.15, 0.15, 0.1\}$$

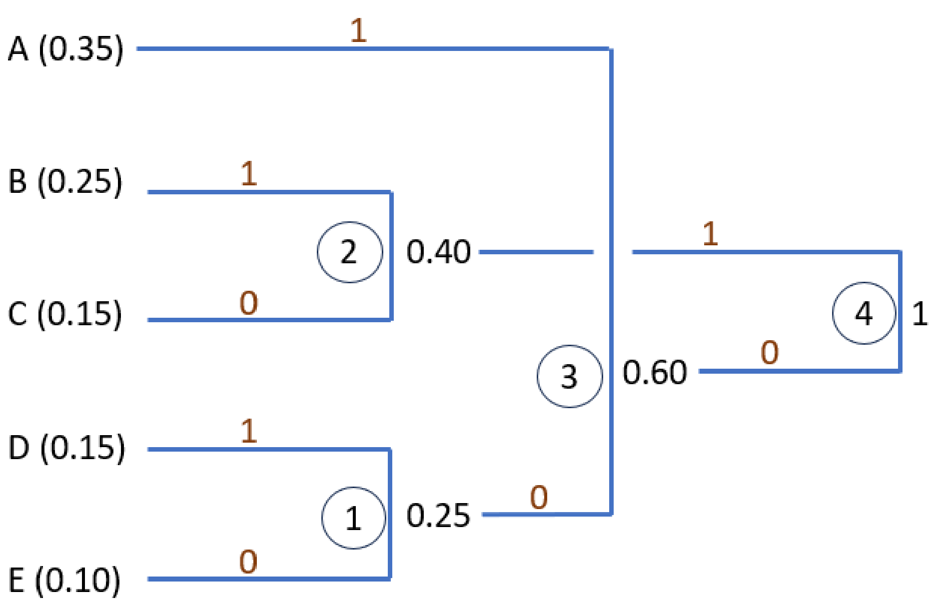
valószínűségeloszlás szerint szolgáltatja.

a) Tervezzen Huffman-kódot a forráshoz.

b) Határozza meg az Ön által tervezett kód (egy szimbólumra eső) várható kódszóhosszát! Milyen messze esik ez az érték a tömöríthetőség elvi alsó határától?

Megoldás:

a)



A: 01; B: 11; C: 10; D: 001; E: 000

b) Az (egy szimbólumra eső) átlagos kódszóhossz

$$L = 0.35 \cdot 2 + 0.25 \cdot 2 + 0.15 \cdot 2 + 0.15 \cdot 3 + 0.1 \cdot 3 = \ 2.25.$$

A tömöríthetőség alsó határát az eloszlás entrópiája adja, ez most

$$H(P) = 0.35 \cdot ld\left(\frac{1}{0.35}\right) + 0.25 \cdot ld\left(\frac{1}{0.25}\right) + 0.15 \cdot ld\left(\frac{1}{0.15}\right) + 0.15 \cdot ld\left(\frac{1}{0.15}\right) + 0.1 \cdot ld\left(\frac{1}{0.1}\right) \approx 2.1833.$$

3.4. Feladat

Egy bináris, lineáris blokk-kód a következő üzenet -> kódszó hozzárendeléssel adott:
00 -> 00000, 01 -> 01110, 10 -> 10101, 11 -> 11011.
a) Adja meg a kód generátormátrixát! Mi lehetett a legvalószínűbb küldött üzenet, amikor a vett blokk v = (01010)?
b) Adja meg a kód paritásellenőrző-mátrixát (vagy annak transzponáltját)! Számítsa ki, milyen szindrómát kapunk a fenti v esetén!
c) Hány bithiba detektálható ezzel a kóddal biztosan? Hány bithiba javítható ezzel a kóddal biztosan?

Megoldás:

a) $G = \begin{pmatrix} 10101 \\ 01110 \end{pmatrix}$

Először nézzük meg a vett szó Hamming-távolságát a kódszavaktól: d(v, 00000) = 2, d(v, 01110) = 1, d(v, 10101) = 5, d(v, 11011) = 2. Válasszuk a legkisebbet. Így a küldött üzenet a 01 lehetett.

b) $H^T = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \ \ s = v \ \cdot H^T = (100)$

c) A kód lineáris szisztematikus blokk-kód, így d_{min} a legkisebb súlyú nem nulla kódszó súlyával egyezik meg.

w(01110) = 3, w(10101) = 3, w(11011) = 4

Ezekből $d_{min} = 3$. Azaz a kód 1 bithiba javítására és 2 bithiba észlelésére alkalmas.

3.5. Feladat

Legyen adva a $G = \begin{pmatrix} 10100 \\ 01110 \end{pmatrix}$ generátormátrix, illetve a vele generált lineáris kód.

a) Írja fel a kód kódszavait!

b) Határozza meg az egyhibás átvitelhez tartozó szindrómákat!

c) Létrehozható -e egyetlen kódszó módosításával olyan lineáris kód, amely minden egyhibát javít?

Megoldás:

a) 00000, 10100, 01110 és 11010

b) Állítsuk elő a paritásellenőrző mátrix transzponáltját:

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

E mátrix sorai éppen az egyhibás hibamintákhoz tartozó szindrómákat tartalmazzák. Látható, hogy van két egyező sor, tehát van két olyan egyhibás helyzet, amelyeket ez a kód nem tud megkülönböztetni (első és harmadik pozíció hibája).

c) Nem. A csupa nulla kódszó eleme kell legyen a kódnak, hiszen a kód lineáris, tehát az nem módosítható. A maradék három kódszó bármelyikében bármit változtatunk, valamelyik másik kódszónak is változni kell, ha a kód linearitását meg akarjuk őrizni.