Gáll József

Pap Gyula

INFORMÁCIÓELMÉLET

EGYETEMI JEGYZET

mobiDIÁK könyvtár

Debreceni Egyetem

Informatikai Intézet

mobiDIÁK könyvtár

SOROZATSZERKESZTŐ

Fazekas István

mobiDIÁK könyvtár

Debreceni Egyetem

Informatikai Intézet

4010 Debrecen, Pf. 12.

http://mobidiak.inf.unideb.hu

A mű a *A mobiDIÁK önszervező mobil portál* (IKTA, OMFB-00373/2003) és a *GNU Iterátor, a legújabb generációs portál szoftver* (ITEM, 50/2003) projektek keretében készült.

Tartalomjegyzék

1. Alapfogalmak

Hírközlési rendszerek általános modellje:

$$\boxed{\text{FORRÁS}} \Longrightarrow \boxed{\text{KÓDOLÓ}} \Longrightarrow \boxed{\text{CSATORNA}} \Longrightarrow \boxed{\text{DEKÓDOLÓ}} \Longrightarrow \boxed{\text{FELHASZNÁLÓ}}$$

Példák: tam–tam dob, füstjelzés, távíró, telefon, rádió, TV, számítógéphálózat.

FORRÁS: egyenlő időközönként, folyamatosan sugároz jeleket, melyek csak **véges** sokfélék lehetnek; ezen jelek összessége: **forrásábécé**; elemei: **betűk**

CSATORNA: egyenlő időközönként, folyamatosan **kódjelek** vihetők át rajta; ezen jelek összessége a **csatornaábécé**, mely szintén csak **véges** sok jelből áll

KÖZLEMÉNY: a forrásábécé betűinek tetszőleges, de véges hosszúságú sorozata

KÓDKÖZLEMÉNY: a csatornaábécé jeleinek tetszőleges, de véges hosszúságú sorozata

KÓDOLÁSI ELJÁRÁS: a közlemények (végtelen) halmazát a kódközlemények (végtelen) halmazába képező függvény

EGYÉRTELMŰEN DEKÓDOLHATÓ KÓDOLÁSI ELJÁRÁS: injektív (azaz különböző közleményekhez különböző kódközleményeket rendel)

2. Egyértelműen dekódolható betűnkénti kódolási eljárások

Forrásábécé: $\mathcal{X} = \{x_1, x_2, \ldots, x_d\}$, ahol $d \geqslant 2$.

Eloszlása: $\mathcal{P} = \{p_1, p_2, \ldots, p_d\}$, ahol $p_i > 0$ minden $i = 1, 2, \ldots, d$ esetén, és $\sum_{i=1}^{d} p_i = 1$.

Csatornaábécé: $\mathcal{Y} = \{y_1, y_2, \ldots, y_D\}$, ahol $D \geqslant 2$.

Kód: $\mathcal{K} = \{K_1, K_2, \ldots, K_d\}$, melynek elemei a **kódszavak**, mégpedig K_i az x_i betű kódja, ami egy véges kódjelsorozat.

Kódszóhosszak: $\mathcal{L} = \{L_1, L_2, \ldots, L_d\}$.

A kódszóhossz várható (átlagos) értéke:

$$\sum_{i=1}^{d} p_i L_i =: \mathbb{E}(\mathcal{K}).$$

Betűnkénti kódolási eljárás: a közlemény betűihez rendelt kódszavak összefűzése .

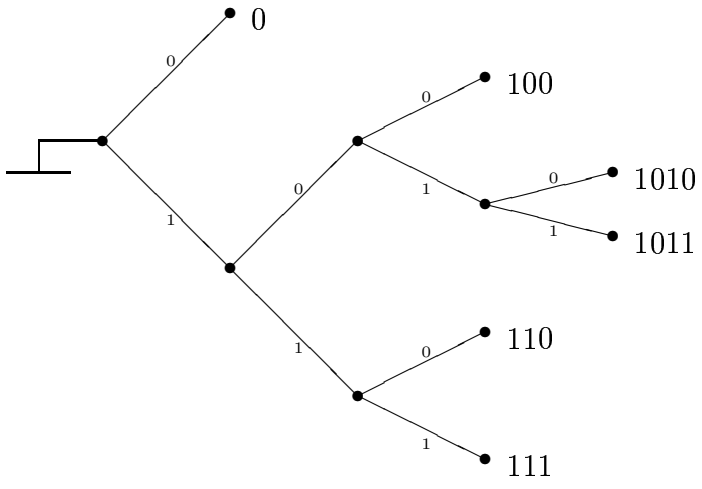
2.1 Állítás. *A $\mathcal{K}$ betűnkénti kódolási eljárás akkor és csak akkor egyértelműen dekódolható, ha tetszőleges $m, n \in \mathbb{N}$ és tetszőleges $K'_1, K'_2, \ldots, K'_m, K''_1, K''_2, \ldots, K''_n \in \mathcal{K}$ esetén $K'_1 K'_2 \ldots K'_m = K''_1 K''_2 \ldots K''_n$ csak úgy lehet, hogy $m = n$ és minden $i = 1, 2, \ldots, n$ esetén $K'_i = K''_i$.*

2.2 Definíció. *A $\mathcal{K}$ kódot **irreducibilisnek** nevezzük, ha a kódszavai mind különbözőek, és egyik kódszava sem folytatása a másiknak.*

2.3 Állítás. *Minden irreducibilis kód egyértelműen dekódolható.*

Bizonyítás. A kódközlemény összhossza szerinti teljes indukcióval. Ha $K'_1 K'_2 \ldots K'_m = K''_1 K''_2 \ldots K''_n$, akkor az irreducibilitás miatt $K'_1 = K''_1$, és így csökken a kódközlemény összhossza. □

Az irreducibilis kódok ábrázolhatók fagráfokkal. Például $\mathcal{K} = \{0, 100, 1010, 1011, 110, 111\}$:



Így például a bináris irreducibilis kódok és a bináris fagráfok között kölcsönösen egyértelmű kapcsolat van, stb.

2.4 Tétel. (Kraft–Fano egyenlőtlenség irreducibilis kódokra) *Tekintsünk egy olyan* $\mathcal{K} = \{K_1, K_2, \ldots, K_d\}$ *irreducibilis kódot, melynek kódszavai* D *számú kódjelből készültek, és a kódszóhosszak* $\mathcal{L} = \{L_1, L_2, \ldots, L_d\}$. *Ekkor*

$$D^{-L_1} + D^{-L_2} + \cdots + D^{-L_d} \leqslant 1.$$

Bizonyítás. Legyen $m := \max\limits_{1 \leqslant i \leqslant d} L_i$. Egészítsük ki a $\mathcal{K}$ kódszavait m hosszúságúra az összes lehetséges módon. Így egy L_i hosszúságú kódszót D^{m-L_i}–féleképpen lehet kiegészíteni. A kiegészítésekkel kapott kódjelsorozatok száma

$$D^{m-L_1} + D^{m-L_2} + \cdots + D^{m-L_d}.$$

Ez nyilván nem lehet több, mint az összes m hosszúságú különböző kódjelsorozatok száma, ami D^m. Tehát

$$D^{m-L_1} + D^{m-L_2} + \cdots + D^{m-L_d} \leqslant D^m.$$

□

2.5 Tétel. *Ha* D *és* $L_1, L_2, \ldots, L_d$ *olyan természetes számok, melyekre teljesül a*

$$D^{-L_1} + D^{-L_2} + \cdots + D^{-L_d} \leqslant 1$$

egyenlőtlenség, akkor létezik olyan D *kódjelből alkotott* $\mathcal{K} = \{K_1, K_2, \ldots, K_d\}$ *irreducibilis kód, melynél a kódszóhosszak:* $\mathcal{L} = \{L_1, L_2, \ldots, L_d\}$.

Bizonyítás. Rendezzük át az adott számokat úgy, hogy $L_1 \geqslant L_2 \geqslant \ldots \geqslant L_d$ teljesüljön. Legyen $m := \max\limits_{1 \leqslant i \leqslant d} L_i$. Legyen $\mathcal{K}^*$ az összes m hosszúságú különböző kódjelsorozatokból álló teljes kódfa. Ez D^m ágból áll, melyek m hosszúságúak. Tekintsük az első m hosszúságú ágat, és vegyük ennek az elején levő L_1 hosszúságú darabját; az ennek megfelelő kódszó legyen K_1, és töröljük le az összes m hosszúságú folytatását a teljes kódfából (mert

ezeket már nem használhatjuk); a letörölt ágak száma éppen D^{m-L_1}. Ha $d=1$, akkor készen vagyunk. Ha $d \geqslant 2$, akkor a feltétel alapján $D^{m-L_1} < D^m$, ezért még van szabad m hosszúságú ág a teljes kódfában, amelynek első L_2 hosszúságú darabja legyen K_2, és letöröljük ennek a folytatásait, melyeknek száma éppen D^{m-L_2}, stb. A feltétel garantálja, hogy az algoritmus nem szakad meg, mielőtt egy olyan fagráfot kapunk, melynek ágai éppen a kívánt hosszúságúak. □

2.6 Tétel. (Kraft–Fano egyenlőtlenség egyértelműen dekódolható kódokra) *Legyen* $\mathcal{K} = \{K_1, K_2, \ldots, K_d\}$ *olyan egyértelműen dekódolható kód, melynek kódszavai* D *számú kódjelből készültek, és a kódszóhosszak* $\mathcal{L} = \{L_1, L_2, \ldots, L_d\}$. *Ekkor*

$$D^{-L_1} + D^{-L_2} + \cdots + D^{-L_d} \leqslant 1.$$

Bizonyítás. Legyen $m := \max\limits_{1 \leqslant i \leqslant d} L_i$. Jelölje $N, L \in \mathbb{N}$ esetén $W(N,L)$ azon N hosszúságú közlemények számát, melyekhez tartozó kódközlemények hossza L. Nyilván $L > mN$ esetén $W(N,L) = 0$, továbbá

$$(D^{-L_1} + D^{-L_2} + \cdots + D^{-L_d})^N = \sum_{L=1}^{mN} W(N,L) D^{-L}.$$

Az egyértelmű dekódolhatóság miatt $W(N,L)$ nem lehet több, mint az összes L hosszúságú kódjelsorozatok száma, tehát $W(N,L) \leqslant D^L$, amiből $W(N,L)D^{-L} \leqslant 1$, így

$$\sum_{L=1}^{mN} W(N,L) D^{-L} \leqslant mN,$$

ezért

$$(D^{-L_1} + D^{-L_2} + \cdots + D^{-L_d})^N \leqslant mN.$$

Ebből már (indirekt úton) következik az állítás, ugyanis ha

$$c := D^{-L_1} + D^{-L_2} + \cdots + D^{-L_d} - 1 > 0$$

volna, akkor

$$(1+c)^N > \frac{N(N-1)}{2} c^2 > mN$$

lenne $N > 1 + 2m/c^2$ esetén, tehát ellentmondásra jutnánk. □

2.7 Lemma. (Alaplemma) *Ha* $\{p_1, p_2, \dots, p_d\}$ *és* $\{q_1, q_2, \dots, q_d\}$ *olyan pozitív számok, hogy* $\sum_{i=1}^{d} p_i = \sum_{i=1}^{d} q_i = 1$, *akkor*

$$\sum_{i=1}^{d} p_i \log \frac{q_i}{p_i} \leqslant 0,$$

és egyenlőség akkor és csak akkor teljesül, ha minden $i = 1, 2, \dots, d$ *esetén* $p_i = q_i$.

Bizonyítás. Mivel a természetes alapú logaritmus függvény (alulról) szigorúan konkáv, így $p, q, x, y > 0$, $p + q = 1$ esetén

$$p \log x + q \log y \leqslant \log(px + qy),$$

és egyenlőség akkor és csak akkor teljesül, ha $x = y$. Legyen most $a_1, a_2, b_1, b_2 > 0$, és alkalmazzuk ezt az egyenlőtlenséget $x := b_1/a_1$, $y := b_2/a_2$, $p := a_1/(a_1 + a_2)$, $q := a_2/(a_1 + a_2)$ szereposztással:

$$\frac{a_1}{a_1 + a_2} \log \frac{b_1}{a_1} + \frac{a_2}{a_1 + a_2} \log \frac{b_2}{a_2} \leqslant \log \left(\frac{a_1}{a_1 + a_2} \frac{b_1}{a_1} + \frac{a_2}{a_1 + a_2} \frac{b_2}{a_2} \right) = \log \frac{b_1 + b_2}{a_1 + a_2},$$

vagyis

$$a_1 \log \frac{b_1}{a_1} + a_2 \log \frac{b_2}{a_2} \leqslant (a_1 + a_2) \log \frac{b_1 + b_2}{a_1 + a_2},$$

és egyenlőség akkor és csak akkor teljesül, ha $b_1/a_1 = b_2/a_2$.

Teljes indukciót használunk d szerint. Ha $d = 2$, akkor alkalmazzuk a fenti egyenlőtlenséget $a_1 := p_1$, $a_2 := p_2$, $b_1 := q_1$, $b_2 := q_2$ szereposztással:

$$p_1 \log \frac{q_1}{p_1} + p_2 \log \frac{q_2}{p_2} \leqslant (p_1 + p_2) \log \frac{q_1 + q_2}{p_1 + p_2} = 0,$$

és egyenlőség akkor és csak akkor teljesül, ha $p_1 = q_1$, $p_2 = q_2$. Ha az egyenlőség teljesül $d-1$–re, akkor először alkalmazzuk az indukciós feltevést a $\{p_1 + p_2, p_3, \dots, p_d\}$ és $\{q_1 + q_2, q_3, \dots, q_d\}$ eloszlásokra:

$$(p_1 + p_2) \log \frac{q_1 + q_2}{p_1 + p_2} + p_3 \log \frac{q_3}{p_3} + \dots + p_d \log \frac{q_d}{p_d} \leqslant 0,$$

és egyenlőség akkor és csak akkor teljesül, ha $p_1 + p_2 = q_1 + q_2$, $p_3 = q_3$, $\dots$, $p_d = q_d$. Ezután alkalmazzuk a fenti egyenlőtlenséget $a_1 := p_1$, $a_2 := p_2$, $b_1 := q_1$, $b_2 := q_2$ szereposztással:

$$p_1 \log \frac{q_1}{p_1} + p_2 \log \frac{q_2}{p_2} \leqslant (p_1 + p_2) \log \frac{q_1 + q_2}{p_1 + p_2},$$

és egyenlőség akkor és csak akkor teljesül, ha $q_1/p_1 = q_2/p_2$, amiből $p_1 + p_2 = q_1 + q_2$ alapján következik $p_1 = q_1$, $p_2 = q_2$. □

2.8 Definíció. *A* $\mathcal{P} = \{p_1, p_2, \ldots, p_d\}$ *eloszlású* $\mathcal{X}$ *forrásábécé* ***entrópiája:***

$$H(\mathcal{X}) := \sum_{i=1}^{d} p_i \log_2 \frac{1}{p_i}.$$

Nyilván $H(\mathcal{X}) > 0$ mindig teljesül.

2.9 Tétel. (McMillan tétele) *Legyen* $\mathcal{K} = \{K_1, K_2, \ldots, K_d\}$ *olyan egyértelműen dekódolható kód, melynek kódszavai* D *számú kódjelből készültek, és a kódszóhosszak* $\mathcal{L} = \{L_1, L_2, \ldots, L_d\}$. *Ekkor*

$$\mathbb{E}(\mathcal{K}) = \sum_{i=1}^{d} p_i L_i \geqslant \sum_{i=1}^{d} p_i \log_D \frac{1}{p_i} = \frac{H(\mathcal{X})}{\log_2 D},$$

és egyenlőség akkor és csak akkor teljesül, ha minden $i = 1, 2, \ldots, d$ *esetén* $p_i = D^{-L_i}$.

Ha a forrásábécé betűinek valószínűségei mind a D *szám negatív egész kitevős hatványai, akkor létezik olyan* $\mathcal{K}$ *irreducibilis kód, melynél*

$$\mathbb{E}(\mathcal{K}) = \sum_{i=1}^{d} p_i L_i = \frac{H(\mathcal{X})}{\log_2 D}.$$

Tetszőleges forrásábécé esetén létezik olyan $\mathcal{K}$ *irreducibilis kód, melynél*

$$\mathbb{E}(\mathcal{K}) = \sum_{i=1}^{d} p_i L_i < 1 + \frac{H(\mathcal{X})}{\log_2 D}.$$

Bizonyítás. A bizonyítandó egyenlőtlenség

$$\sum_{i=1}^{d} p_i \log \frac{1}{p_i} - (\log D) \sum_{i=1}^{d} p_i L_i \leqslant 0,$$

azaz

$$\sum_{i=1}^{d} p_i \log \frac{D^{-L_i}}{p_i} \leqslant 0.$$

(Itt tetszőleges, 1–nél nagyobb alapú logaritmust használhatunk.) Alkalmazzuk az alaplemmát a $\{p_1, p_2, \ldots, p_d\}$ és

$$q_i := \frac{D^{-L_i}}{\sum_{k=1}^{d} D^{-L_k}}, \qquad i = 1, 2, \ldots, d$$

eloszlásokra:

$$\sum_{i=1}^{d} p_i \log \frac{D^{-L_i}}{p_i \sum\limits_{k=1}^{d} D^{-L_k}} \leqslant 0,$$

azaz

$$\sum_{i=1}^{d} p_i \log \frac{D^{-L_i}}{p_i} \leqslant \log \left(\sum_{k=1}^{d} D^{-L_k} \right).$$

Mivel $\mathcal{K}$ egyértelműen dekódolható, így a Kraft–Fano egyenlőtlenség szerint

$$\log \left(\sum_{k=1}^{d} D^{-L_k} \right) \leqslant \log 1 = 0,$$

tehát a bizonyítandó egyenlőtlenséget beláttuk. Egyenlőség akkor és csak akkor teljesül, ha egyrészt $\log \left(\sum\limits_{k=1}^{d} D^{-L_k} \right) = 0$, másrészt minden $i = 1, 2, \ldots, d$ esetén $p_i = q_i$. Az első alapján $\sum\limits_{k=1}^{d} D^{-L_k} = 1$, és így a második alapján minden $i = 1, 2, \ldots, d$ esetén $p_i = D^{-L_i}$.

Ha léteznek olyan $\mathcal{L} = \{L_1, L_2, \ldots, L_d\}$ természetes számok, hogy minden $i = 1, 2, \ldots, d$ esetén $p_i = D^{-L_i}$, akkor $\sum\limits_{i=1}^{d} D^{-L_i} = \sum\limits_{i=1}^{d} p_i = 1$, ezért létezik olyan irreducibilis kód, melynél a kódszóhosszak $\mathcal{L} = \{L_1, L_2, \ldots, L_d\}$, így

$$\sum_{i=1}^{d} p_i L_i = \sum_{i=1}^{d} p_i \log_D \frac{1}{p_i} = \frac{H(\mathcal{X})}{\log_2 D}.$$

Tetszőleges $\mathcal{P} = \{p_1, p_2, \ldots, p_d\}$ eloszlású forrásábécé esetén legyen

$$L_i := \left\lceil \log_D \frac{1}{p_i} \right\rceil,$$

ahol $a \in \mathbb{R}$ esetén $\lceil a \rceil$ az a **felső egész részét** jelöli:

$$\lceil a \rceil := \min\{n \in \mathbb{Z} : n \geqslant a\},$$

ezért $a \leqslant \lceil a \rceil < a + 1$. Ekkor

$$\log_D \frac{1}{p_i} \leqslant L_i < 1 + \log_D \frac{1}{p_i},$$

így egyrészt $D^{-L_i} \leqslant p_i$ miatt teljesül $\sum\limits_{i=1}^{d} D^{-L_i} \leqslant \sum\limits_{i=1}^{d} p_i = 1$, ezért létezik olyan irreducibilis kód, melynél a kódszóhosszak éppen $\mathcal{L} = \{L_1, L_2, \ldots, L_d\}$, másrészt

$$\sum_{i=1}^{d} p_i L_i < \sum_{i=1}^{d} p_i \left(1 + \log_D \frac{1}{p_i} \right) = 1 + \frac{H(\mathcal{X})}{\log_2 D}.$$

□

2.10 Megjegyzés. A tétel harmadik állítása nem javítható, mert például ha $d = 2$ és $D = 2$, akkor tetszőleges egyértelműen dekódolható kód esetén $p_1L_1 + p_2L_2 \geqslant p_1 + p_2 = 1$, viszont a $H(\mathcal{X})$ entrópia tetszőlegesen kis pozitív szám lehet, hiszen ekkor

$$H(\mathcal{X}) = p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2} = p_1 \log_2 \frac{1}{p_1} + (1 - p_1) \log_2 \frac{1}{1 - p_1},$$

ezért

$$\lim_{p_1 \downarrow 0} H(\mathcal{X}) = 0.$$

2.11 Definíció. *A $\mathcal{P} = \{p_1, p_2, \ldots, p_d\}$ eloszlású $\mathcal{X}$ forrásábécé D számú kódjelből alkotott $\mathcal{L} = \{L_1, L_2, \ldots, L_d\}$ kódszóhosszakkal rendelkező $\mathcal{K} = \{K_1, K_2, \ldots, K_d\}$ kódjának* ***hatásfoka****:*

$$\eta(\mathcal{K}) := \frac{H(\mathcal{X})}{(\log_2 D)\mathbb{E}(\mathcal{K})} = \frac{H(\mathcal{X})}{(\log_2 D) \sum_{i=1}^{d} p_i L_i}.$$

Egyértelműen dekódolható $\mathcal{K}$ kód esetén a McMillan tétel alapján $\eta(\mathcal{K}) \in (0, 1]$, és $\eta(\mathcal{K}) = 1$ akkor és csak akkor teljesül, ha a $p_1, p_2, \ldots, p_d$ valószínűségek mind a D szám negatív egész kitevős hatványai, és $L_i = \log_D \frac{1}{p_i}$.

2.12 Tétel. *Tetszőleges forrásábécé esetén létezik maximális hatásfokú egyértelműen dekódolható kód.*

Bizonyítás. Legyen $p := \min_{1 \leqslant i \leqslant d} p_i$. Nyilván $p > 0$. Ha egy $\mathcal{K} = \{K_1, K_2, \ldots, K_d\}$ kód esetén az $\mathcal{L} = \{L_1, L_2, \ldots, L_d\}$ kódszóhosszakra

$$\max_{1 \leqslant i \leqslant d} L_i \geqslant \frac{1}{p} \left(1 + \frac{H(\mathcal{X})}{\log_2 D}\right)$$

teljesül, akkor $\mathcal{K}$ nem lehet maximális hatásfokú, hiszen ekkor egyrészt

$$\mathbb{E}(\mathcal{K}) = \sum_{i=1}^{d} p_i L_i \geqslant p \sum_{i=1}^{d} L_i \geqslant p \max_{1 \leqslant i \leqslant d} L_i \geqslant 1 + \frac{H(\mathcal{X})}{\log_2 D}$$

miatt

$$\eta(\mathcal{K}) \leqslant \frac{H(\mathcal{X})}{H(\mathcal{X}) + \log_2 D},$$

másrészt a McMillan–tétel szerint létezik olyan $\mathcal{K}' = \{K'_1, K'_2, \ldots, K'_d\}$ irreducibilis kód, melynél a kódszóhosszak $\mathcal{L}' = \{L'_1, L'_2, \ldots, L'_d\}$ és

$$\mathbb{E}(\mathcal{K}') = \sum_{i=1}^{d} p_i L'_i < 1 + \frac{H(\mathcal{X})}{\log_2 D},$$

így

$$\eta(\mathcal{K}') > \frac{H(\mathcal{X})}{H(\mathcal{X}) + \log_2 D} \geqslant \eta(\mathcal{K}).$$

Tehát maximális hatásfokú kód keresésénél elég azokat a kódokat figyelembe venni, melyeknél a kódszóhosszakra

$$\max_{1 \leqslant i \leqslant d} L_i < \frac{1}{p}\left(1 + \frac{H(\mathcal{X})}{\log_2 D}\right) =: m$$

teljesül; ilyen viszont csak véges sok van, mert a szóbajövő kódszavak száma legfeljebb D^m, tehát biztosan lesz olyan, amelyiknél a hatásfok maximális. □

Egy kód akkor és csak akkor maximális hatásfokú, ha az $\mathbb{E}(\mathcal{K})$ átlagos kódszóhossz minimális.

2.13 Definíció. *A maximális hatásfokú, egyértelműen dekódolható kódokat* ***optimális kódoknak*** *nevezzük.*

Mindig létezik **irreducibilis optimális kód**, ugyanis a Kraft–Fano egyenlőtlenség teljesül minden egyértelműen dekódolható kódra, és így a McMillan–tétel szerint ugyanilyen kódszóhosszakkal lehet irreducibilis kódot is konstruálni.

3. Irreducibilis kódok konstrukciója

3.1 Huffman–féle kódolási eljárás a $\mathbf{D = 2}$ (bináris) esetben: Rendezzük át a forrásábécé betűit úgy, hogy

$$p_1 \geqslant p_2 \geqslant \ldots \geqslant p_{d-2} \geqslant p_{d-1} \geqslant p_d$$

teljesüljön. Az alapötlet az, hogy a problémát visszavezetjük a két legkisebb valószínűségű betű összevonásával a

$$\{p_1, p_2, \ldots, p_{d-2}, p_{d-1} + p_d\}$$

eloszlású forrásábécéhez tartozó optimális kód meghatározására. Ha még ez is bonyolult, akkor tovább redukáljuk ugyanezzel a módszerrel a problémát (szükség szerint újrarendezve a valószínűségeket és összevonva a két legkisebb valószínűségű betűt). A $d = 2$ esetben a $\mathcal{K} := \{0, 1\}$ kód már eloszlástól függetlenül mindig optimális. Ezután visszafelé haladunk, és amikor egy betű „kettéosztódik", akkor úgy gyártunk új kódot, hogy ennek a betűnek a kódszavát kiegészítjük kétféle módon. Be fogjuk látni, hogy ez a kódolási eljárás mindig optimális, irreducibilis kódhoz vezet.

3.2 Állítás. *Tekintsük a $\mathcal{P} = \{p_1, p_2, \ldots, p_d\}$ eloszlású $\mathcal{X}$ forrásábécét, és tegyük fel, hogy $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_d$. Ekkor létezik $\mathcal{X}$–nek olyan irreducibilis, optimális bináris kódja, melynél a két utolsó betű kódszava, K_{d-1} és K_d maximális hosszúságú (azaz $L_{d-1} = L_d = \max_{1 \leqslant i \leqslant d} L_i$), és az utolsó kódjelüktől eltekintve azonosak.*

Bizonyítás. Vegyünk egy tetszőleges $\mathcal{K} = \{K_1, K_2, \ldots, K_d\}$ irreducibilis, optimális kódot. Legyen $m := \max_{1 \leqslant i \leqslant d} L_i$. Tekintsünk egy maximális hosszúságú K_j kódszót, azaz legyen $j \in \{1, 2, \ldots, d\}$ olyan, hogy $L_j = m$. Jelölje K' azt a kódjelsorozatot, melyet úgy kapunk, hogy a K_j kódszó utolsó jelét megváltoztatjuk. Ekkor $K' \in \mathcal{K}$ kell hogy legyen, mert ha ez nem teljesülne, akkor a K_j utolsó jelének elhagyásával kapott K'' kódszóra kicserélve a K_j kódszót egy olyan irreducibilis kódot kapnánk, amelynél az átlagos kódszóhossz kisebb lenne, mint $\mathcal{K}$ esetén; ez viszont ellentmondana annak, hogy $\mathcal{K}$ optimális. Tehát van egy olyan $k \in \{1, 2, \ldots, d\}$, $k \neq j$ index, melyre $K_k = K'$ teljesül. Tekintsük most azt a kódot, melyet úgy kapunk, hogy kicseréljük a K_j és K_k kódszavakat a K_d és K_{d-1} kódszavakkal, mégpedig úgy, hogy ha $p_j \geqslant p_k$, akkor a K_j kódszót a K_{d-1} kódszóval, egyébként a K_j kódszót a K_d kódszóval cseréljük ki. Az így keletkező $\mathcal{K}^*$ kód is irreducibilis lesz, és szintén

optimális, hiszen az átlagos kódszóhossz nem növekedhet, mert például $p_j \geqslant p_k$ esetén

$$\mathbb{E}(\mathcal{K}^*) - \mathbb{E}(\mathcal{K}) = (p_d + p_{d-1})m + p_j L_{d-1} + p_k L_d - p_{d-1}L_{d-1} - p_d L_d - (p_j + p_k)m$$

$$= (p_d - p_k)(m - L_d) + (p_{d-1} - p_j)(m - L_{d-1}) \leqslant 0,$$

hiszen nyilván $p_d \leqslant p_k$ és $p_{d-1} \leqslant p_j$. □

3.3 Tétel. *Tekintsük a* $\mathcal{P} = \{p_1, p_2, \ldots, p_d\}$ *eloszlású* $\mathcal{X} = \{x_1, x_2, \ldots, x_d\}$ *forrásábécét, és tegyük fel, hogy* $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_d$. *Tekintsük a* $\mathcal{P}' = \{p_1, p_2, \ldots, p_{d-2}, p_{d-1} + p_d\}$ *eloszlású* $\mathcal{X}' = \{x_1, x_2, \ldots, x_{d-2}, x'_{d-1}\}$ *forrásábécét. Legyen* $\mathcal{K}' = \{K'_1, K'_2, \ldots, K'_{d-1}\}$ *az* $\mathcal{X}'$*–nek egy optimális, irreducibilis bináris kódja. Legyen* $\mathcal{K} = \{K_1, K_2, \ldots, K_d\}$ *az* $\mathcal{X}$*–nek a következő módon konstruált bináris kódja:*

$$K_i := \begin{cases} K'_i & \text{ha } i = 1, 2, \ldots, d-2, \\ K'_{d-1}0 & \text{ha } i = d-1, \\ K'_{d-1}1 & \text{ha } i = d. \end{cases}$$

(Azaz K_{d-1} *és* K_d *a* K'_{d-1} *kétféle kiegészítése.) Ekkor* $\mathcal{K}$ *az* $\mathcal{X}$*–nek egy optimális, irreducibilis bináris kódja.*

Bizonyítás. Nyilván $\mathcal{K}$ irreducibilis. Jelölje a $\mathcal{K}$ kód kódszóhosszait $\mathcal{L} = \{L_1, L_2, \ldots, L_d\}$. Legyen $\ell := L_{d-1} = L_d$. Ekkor

$$\mathbb{E}(\mathcal{K}) = \sum_{i=1}^{d-2} p_i L_i + (p_{d-1} + p_d)\ell,$$

$$\mathbb{E}(\mathcal{K}') = \sum_{i=1}^{d-2} p_i L_i + (p_{d-1} + p_d)(\ell - 1),$$

tehát $\mathbb{E}(\mathcal{K}) = \mathbb{E}(\mathcal{K}') + p_{d-1} + p_d$. Ha $\mathcal{K}$ nem volna optimális, akkor volna olyan $\mathcal{K}^*$ irreducibilis, optimális kódja $\mathcal{X}$–nek, melyre $\mathbb{E}(\mathcal{K}^*) < \mathbb{E}(\mathcal{K})$, és az is elérhető az előző állítás szerint, hogy a két utolsó betű kódszava, K^*_{d-1} és K^*_d maximális hosszúságú, és az utolsó kódjelüktől eltekintve azonosak. Ha ezt a kódjelet elhagyjuk, akkor az $\mathcal{X}'$–nek egy olyan $\mathcal{K}^{*\prime}$ irreducibilis kódját kapnánk, melyre

$$\mathbb{E}(\mathcal{K}^{*\prime}) = \mathbb{E}(\mathcal{K}^*) - p_{d-1} - p_d < \mathbb{E}(\mathcal{K}) - p_{d-1} - p_d = \mathbb{E}(\mathcal{K}')$$

lenne, ami ellentmondana annak, hogy $\mathcal{K}'$ optimális. $\square$

Azt is be lehet látni, hogy tetszőleges irreducibilis, optimális kódot meg lehet kapni ezzel az eljárással.

Ha $D > 2$, akkor hasonló a konstrukció: minden lépésnél a D számú, legkisebb valószínűségű betűt vonjuk össze. Mivel azt szeretnénk elérni, hogy a végén D számú betű maradjon, és a betűk száma minden lépésben $D-1$–el csökken, így **az első lépésben r számú betűt kell összevonni**, ahol az $r \in \{2, \dots, D\}$ számot úgy kell megválasztani, hogy $r \equiv D \pmod{D-1}$ teljesüljön. Az előző módszerrel be lehet látni, hogy ez a kódolási eljárás is mindig optimális, irreducibilis kódhoz vezet.

3.4 Shannon–féle bináris kód. A konstrukció a következő: most is rendezzük át a forrásábécé betűit úgy, hogy

$$p_1 \geqslant p_2 \geqslant \dots \geqslant p_{d-2} \geqslant p_{d-1} \geqslant p_d$$

teljesüljön. Válasszuk meg az $\mathcal{L} = \{L_1, L_2, \dots, L_d\}$ kódszóhosszakat úgy, hogy $2^{-L_i} \leqslant p_i < 2^{-L_i+1}$ teljesüljön, azaz

$$L_i := \left\lceil \log_2 \frac{1}{p_i} \right\rceil .$$

Ekkor $2^{-L_i} \leqslant p_i$ miatt teljesül $\sum_{i=1}^{d} 2^{-L_i} \leqslant \sum_{i=1}^{d} p_i = 1$, ezért létezik olyan irreducibilis kód, melynél a kódszóhosszak éppen $\mathcal{L} = \{L_1, L_2, \dots, L_d\}$. Most egy ilyen kódot könnyen meg tudunk adni a következő módon. Legyen $r_1 := 0$, és $r_i := \sum_{k=1}^{i-1} p_k$, ha $i = 2, 3, \dots, d$. Jelölje $a \in \mathbb{R}$ esetén $\lfloor a \rfloor$ az a **alsó egész részét:**

$$\lfloor a \rfloor := \max\{n \in \mathbb{Z} : n \leqslant a\},$$

ezért $a - 1 < \lfloor a \rfloor \leqslant a$. A K_i kódszó legyen a $\lfloor 2^{L_i} r_i \rfloor$ szám 2–es számrendszerben felírt alakja úgy, hogy az pontosan L_i számjegyből álljon, vagyis ha rövidebb lenne, akkor 0–kat írunk elé. (Nyilván L_i–nél több számjegy nem lehet, mert $r_i < 1$ miatt $2^{L_i} r_i < 2^{L_i}$, és 2^{L_i} az első $L_i + 1$ jegyű szám.)

3.5 Tétel. *A Shannon–féle bináris kód irreducibilis, és az átlagos kódszóhossza* $\mathbb{E}(\mathcal{K}) < H(\mathcal{X}) + 1$.

Bizonyítás. Legyen $1 \leqslant i < j \leqslant d$. Ekkor $p_i \geqslant p_j$, így $L_i \leqslant L_j$, ezért K_i nem lehet kiegészítése K_j–nek. Továbbá $p_i \geqslant 2^{-L_i}$ miatt $r_j \geqslant r_{i+1} = r_i + p_i \geqslant r_i + 2^{-L_i}$, így

$$\left\lfloor 2^{L_i} r_j \right\rfloor \geqslant \left\lfloor 2^{L_i}(r_i + 2^{-L_i}) \right\rfloor = \left\lfloor 2^{L_i} r_i \right\rfloor + 1 > \left\lfloor 2^{L_i} r_i \right\rfloor,$$

de a $\left\lfloor 2^{L_i} r_j \right\rfloor$ szám 2–es számrendszerbeli alakja éppen a $\left\lfloor 2^{L_j} r_j \right\rfloor$ első L_i jegye; vagyis K_i és K_j különböző, és K_j nem kiegészítése K_i–nek. □

3.6 Gilbert–féle bináris kód: Ennek a konstrukciónak az az előnye, hogy nem kell nagyság szerint sorbarendezni a valószínűségeket. Most válasszuk meg az $\mathcal{L} = \{L_1, L_2, \ldots, L_d\}$ kódszóhosszakat úgy, hogy $2^{-L_i+1} \leqslant p_i < 2^{-L_i+2}$ teljesüljön, azaz

$$L_i := \left\lceil \log_2 \frac{1}{p_i} \right\rceil + 1.$$

Legyen $r_1 := \frac{1}{2} p_1$, és $r_i := \frac{1}{2} p_i + \sum_{k=1}^{i-1} p_k$ ha $i = 2, 3, \ldots, d$. A K_i kódszó most is legyen a $\left\lfloor 2^{L_i} r_i \right\rfloor$ szám 2–es számrendszerben felírt alakja úgy, hogy az pontosan L_i számjegyből álljon.

3.7 Tétel. *A Gilbert–féle bináris kód irreducibilis, és az átlagos kódszóhossza* $H(\mathcal{X}) + 1 \leqslant \mathbb{E}(\mathcal{K}) < H(\mathcal{X}) + 2$.

Bizonyítás. Ha $i < j$, akkor $p_i \geqslant 2^{-L_i+1}$ és $p_j \geqslant 2^{-L_j+1}$ miatt

$$r_j = r_{j-1} + \frac{1}{2} p_{j-1} + \frac{1}{2} p_j \geqslant r_i + \frac{1}{2} p_i + \frac{1}{2} p_j \geqslant r_i + 2^{-L_i} + 2^{-L_j},$$

így

$$\left\lfloor 2^{L_i} r_j \right\rfloor \geqslant \left\lfloor 2^{L_i}(r_i + 2^{-L_i} + 2^{-L_j}) \right\rfloor \geqslant \left\lfloor 2^{L_i} r_i \right\rfloor + 1 > \left\lfloor 2^{L_i} r_i \right\rfloor$$

valamint

$$\left\lfloor 2^{L_j} r_j \right\rfloor \geqslant \left\lfloor 2^{L_j}(r_i + 2^{-L_i} + 2^{-L_j}) \right\rfloor \geqslant \left\lfloor 2^{L_j} r_i \right\rfloor + 1 > \left\lfloor 2^{L_j} r_i \right\rfloor,$$

így K_i és K_j különböző, és egyik sem lehet a másik kiegészítése. □

4. Az információ mennyiségének mérőszámai

Heurisztikusan: az információ mennyisége megegyezik a legtömörebb megfogalmazás terjedelmével. Ehhez pontosan körül kell határolni:

- az összes szóba jöhető információt: ez egy $\mathcal{X} = \{x_1, x_2, \ldots, x_d\}$ forrásábécé betűiből készíthető közlemények halmaza;
- azok megengedett megjelenési formáját: ez egy $\mathcal{Y} = \{y_1, y_2, \ldots, y_D\}$ csatronaábécé kódjeleinek felhasználásával, egyértelműen dekódolható kódolási eljárással készített kódközlemények halmaza;
- egy adott információ megjelenésének valószínűségét; ha a betűk egymástól függetlenül jönnek, akkor elegendő ismerni a forrásábécé betűinek $\mathcal{P} = \{p_1, p_2, \ldots, p_d\}$ előfordulási valószínűségeit.

Persze ha D növekszik, akkor egyre tömörebb lehet a megfogalmazás, ezért rögzítjük: $D := 2$. Tudjuk, hogy ekkor egyérteműen dekódolható betűnkénti kódolási eljárásnál a kódszóhossz várható értéke (azaz az egy betűre eső átlagos kódszóhossz) nem lehet kisebb, mint a $H(\mathcal{X})$ entrópia, és blokkonkénti kódolással ez a határ tetszőlegesen megközelíthető.

4.1 Definíció. *A $\mathcal{P} = \{p_1, p_2, \ldots, p_d\}$ eloszlású $\mathcal{X}$ forrásábécéből összeállított közlemények* ***egy betűre eső átlagos információmennyisége*** *a*

$$H(\mathcal{X}) = \sum_{i=1}^{d} p_i \log_2 \frac{1}{p_i}$$

entrópia.

Például ha $\mathcal{X} = \{0, 1\}$ és $\mathbb{P} = \{\frac{1}{2}, \frac{1}{2}\}$, akkor az egy jelre eső átlagos információ éppen **egységnyi**, hiszen $d = 2$, $p_1 = p_2 = \frac{1}{2}$, így $H(\mathcal{X}) = \frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2 = 1$; ezt nevezzük **1 bitnek**.

Kérdés: mennyi egy **egyedi közlemény** által tartalmazott információ mennyisége; például mennyi információt kapunk, amikor éppen az x_i betű kódszavát kapjuk?

Heurisztikusan: Ha az $\mathcal{X} = \{x_1, x_2, \ldots, x_d\}$ forrásábécének van 100 %–os hatásfokú, egyértelműen dekódolható bináris kódja, azaz a forrásábécé betűinek $\mathcal{P} = \{p_1, p_2, \ldots, p_d\}$ előfordulási valószínűségei éppen $p_i = 2^{-L_i}$ alakúak valamely $L_i \in \mathbb{N}$ számokkal, akkor az x_i betű kódszavának hossza éppen

$$L_i = \log_2 \frac{1}{p_i},$$

ezért ezt lehet az x_i betű által tartalmazott információ mennyiségének tekinteni.

A fenti kérdés a következő módon fogalmazható át: mennyi egy p valószínűségű A esemény megfigyelése által szerezhető egyedi információmennyiség, azaz mennyi az A eseménnyel kapcsolatos **bizonytalanság** mértéke? Azaz mennyi információra van szükségünk, hogy az A eseménnyel kapcsolatos bizonytalanságunkat eloszlassuk? Jelölje ezt $I(A)$. A következő természetes követelményeket kell kielégítenünk:

(i) Kevésbé valószínű esemény bekövetkezése több információt szolgáltat: ha $\mathbb{P}(A) \geqslant \mathbb{P}(A')$, akkor $I(A) \leqslant I(A')$. Ebből következik, hogy $I(A)$ csak a $\mathbb{P}(A)$ valószínűségtől függ! Ugyanis ha $\mathbb{P}(A) = \mathbb{P}(A')$, akkor egyrészt teljesülnie kell, hogy $I(A) \leqslant I(A')$, másrészt $I(A) \geqslant I(A')$, így $I(A) = I(A')$. Tehát $I(A) = f(\mathbb{P}(A))$ alakú, ahol $f : [0,1] \to \mathbb{R}$. Sőt az előzőek alapján f monoton fogyó: $p \leqslant q$ esetén $f(p) \geqslant f(q)$.

(ii) Független események együttes bekövetkezésének megfigyelése esetén az információ összeadódik: $I(A \cap B) = I(A) + I(B)$. Mivel ekkor $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$, így az kell, hogy $f(pq) = f(p) + f(q)$ ha $p, q \in [0,1]$.

(iii) Ha $\mathbb{P}(A) = \dfrac{1}{2}$, akkor $I(A) = 1$, azaz $f\left(\dfrac{1}{2}\right) = 1$.

4.2 Tétel. *Ha* $f : (0,1] \to \mathbb{R}$ *olyan függvény, melyre*

(i) $f(p) \geqslant f(q)$ *ha* $0 < p \leqslant q \leqslant 1$,

(ii) $f(pq) = f(p) + f(q)$, *ha* $0 < p, q \leqslant 1$,

(iii) $f\left(\dfrac{1}{2}\right) = 1$,

akkor $f(p) = \log_2 \frac{1}{p}$, $0 < p \leqslant 1$.

Bizonyítás. Helyettesítsük be a $p = q$ számokat (ii)–be: $f(p^2) = 2f(p)$. Ebből teljes indukcióval: $f(p^n) = nf(p)$, ha $n \in \mathbb{N}$. Ez (iii) felhasználásával $p = \frac{1}{2}$ esetén $f\left(\frac{1}{2^n}\right) = n$. Másrészt $m \in \mathbb{N}$ esetén $f(p) = \frac{1}{m} f(p^m)$, így ebbe a $p = 2^{-n/m}$ értéket helyettesítve

$$f(2^{-n/m}) = \frac{1}{m} f((2^{-n/m})^m) = \frac{1}{m} f(2^{-n}) = \frac{n}{m}.$$

Tehát tetszőleges x pozitív racionális szám esetén teljesül $f(2^{-x}) = x$. Sőt (ii) és (iii) alapján

$$1 = f\left(\frac{1}{2}\right) = f\left(\frac{1}{2} \cdot 1\right) = f\left(\frac{1}{2}\right) + f(1) = 1 + f(1),$$

ezért $f(1) = 0$, így a fenti teljesül $x = 0$ esetén is. Ha most $x > 0$ irracionális szám, akkor tetszőleges $m \in \mathbb{N}$ esetén létezik olyan n nemnegatív egész szám, hogy $\frac{n}{m} < x < \frac{n+1}{m}$. Ekkor (i) felhasználásával

$$\frac{n}{m} = f(2^{-n/m}) \leqslant f(2^{-x}) \leqslant f(2^{-(n+1)/m}) = \frac{n+1}{m},$$

ezért $0 \leqslant f(2^{-x}) - \frac{n}{m} \leqslant \frac{1}{m}$. Mivel $0 < x - \frac{n}{m} < \frac{1}{m}$, így $|f(2^{-x}) - x| < \frac{1}{m}$. Mivel $m \in \mathbb{N}$ tetszőleges lehet, így $f(2^{-x}) = x$ teljesül minden $x \geqslant 0$ esetén. Ebből $x = \log_2 \frac{1}{p}$ helyettesítéssel kapjuk az állítást. □

Tehát az egy betűre eső átlagos információ mennyiségét úgy is lehet tekinteni, mint az egyes betűk által szolgáltatott információ várható értékét.

Ha most egy valószínűségi változó értékeit figyeljük meg, akkor a bizonytalanságunk mértékét, vagyis az entrópiát, azaz a szerezhető átlagos információmennyiséget a következő módon lehet értelmezni:

4.3 Definíció. *Legyen* $\xi = (\xi_1, \xi_2, \ldots, \xi_m)$ *egy olyan diszkrét valószínűségi vektorváltozó, melynek lehetséges értékei* $\{x_1, x_2, \ldots, x_k\} \subset \mathbb{R}^m$, *és az* x_i *értéket* p_i *valószínűséggel veszi fel (azaz* $\mathbb{P}\{\xi = x_i\} = p_i$.*) Ekkor* ξ ***entrópiája***

$$H(\xi) := H(\xi_1, \xi_2, \ldots, \xi_m) := \sum_{i=1}^{k} \mathbb{P}\{\xi = x_i\} I\{\xi = x_i\} = \sum_{i=1}^{k} p_i \log_2 \frac{1}{p_i},$$

ahol $p_i = 0$ *esetén* $p_i \log_2 \frac{1}{p_i} := 0$.

4.4 Tétel. *Az entrópia tulajdonságai:*

(i) $H(\xi)$ *csak* ξ *eloszlásától függ, azaz ha* ξ *és* η *eloszlása megegyezik, azaz ugyanazokat az értékeket ugyanolyan valószínűséggel veszik fel, akkor* $H(\xi) = H(\eta)$.

(ii) $H(\xi) \geqslant 0$, *és* $H(\xi) = 0$ *akkor és csak akkor, ha* ξ *eloszlása elfajult, azaz* ξ *konstans, vagyis csak egy lehetséges értéket vehet fel.*

(iii) *Ha* ξ *lehetséges értékeinek száma* k, *akkor* $H(\xi) \leqslant \log_2 k$, *és* $H(\xi) = \log_2 k$ *akkor és csak akkor, ha* ξ *eloszlása egyenletes, azaz* $p_i = 1/k$, $i = 1, \ldots, k$.

(iv) $H(\xi, \xi) = H(\xi)$.

(v) $H(\xi, \eta) \leqslant H(\xi) + H(\eta)$, *és* $H(\xi, \eta) = H(\xi) + H(\eta)$ *akkor és csak akkor, ha* ξ *és* η *függetlenek.*

(vi) $H(\xi_1, \xi_2, \ldots, \xi_n) \leqslant H(\xi_1) + H(\xi_2) + \cdots + H(\xi_n)$, *és egyenlőség akkor és csak akkor áll fenn, ha* $\xi_1, \xi_2, \ldots, \xi_n$ *teljesen függetlenek.*

Bizonyítás. (i) következik a definícióból.

(ii) következik abból, hogy $x \log \dfrac{1}{x} > 0$ ha $x \in (0, 1)$.

(iii)–ban a bizonyítandó egyenlőtlenség $\log_2 k = \sum_{i=1}^{k} p_i \log_2 k$ alapján

$$0 \geqslant \sum_{i=1}^{k} p_i \left(\log_2 \frac{1}{p_i} - \log_2 k \right) = \sum_{i=1}^{k} p_i \log_2 \frac{1/k}{p_i},$$

így az állítás következik a 2.7 Alaplemmából $q_i := 1/k$ választással.

(iv) abból következik, hogy ha a ξ valószínűségi vektorváltozó lehetséges értékei $\{x_1, x_2, \ldots, x_k\}$, és az x_i értéket $p_i > 0$ valószínűséggel veszi fel, akkor a (ξ, ξ) valószínűségi vektorváltozó lehetséges értékei $\{(x_1, x_1), (x_2, x_2), \ldots, (x_k, x_k)\}$, és az (x_i, x_i) értéket p_i valószínűséggel veszi fel.

(v) bizonyításához legyenek a ξ és η valószínűségi vektorváltozók lehetséges értékei $\{x_i : 1 \leqslant i \leqslant k\}$ illetve $\{y_j : 1 \leqslant j \leqslant \ell\}$. Jelölje $r_{i,j} := \mathbb{P}\{\xi = x_i,\, \eta = y_j\}$. Ekkor $1 \leqslant i \leqslant k$ esetén

$$p_i := \mathbb{P}\{\xi = x_i\} = \sum_{j=1}^{\ell} \mathbb{P}\{\xi = x_i,\, \eta = y_j\} = \sum_{j=1}^{\ell} r_{i,j},$$

és hasonlóan $1 \leqslant j \leqslant \ell$ esetén

$$q_j := \mathbb{P}\{\eta = y_j\} = \sum_{i=1}^{k} \mathbb{P}\{\xi = x_i,\, \eta = y_j\} = \sum_{i=1}^{k} r_{i,j}.$$

Azt kell megmutatni, hogy

$$\sum_{i=1}^{k}\sum_{j=1}^{\ell} r_{i,j} \log_2 \frac{1}{r_{i,j}} \leqslant \sum_{i=1}^{k} p_i \log_2 \frac{1}{p_i} + \sum_{j=1}^{\ell} q_j \log_2 \frac{1}{q_j},$$

azaz

$$0 \geqslant \sum_{i=1}^{k}\sum_{j=1}^{\ell} r_{i,j} \left(\log_2 \frac{1}{r_{i,j}} - \log_2 \frac{1}{p_i} - \log_2 \frac{1}{q_j} \right) = \sum_{i=1}^{k}\sum_{j=1}^{\ell} r_{i,j} \log_2 \frac{p_i q_j}{r_{i,j}}.$$

Mivel

$$\sum_{i=1}^{k}\sum_{j=1}^{\ell} p_i q_j = \left(\sum_{i=1}^{k} p_i \right) \left(\sum_{j=1}^{\ell} q_j \right) = \left(\sum_{i=1}^{k}\sum_{j=1}^{\ell} r_{i,j} \right)^2 = 1,$$

így a 2.7 Alaplemmát az $\{r_{i,j} : 1 \leqslant i \leqslant k,\, 1 \leqslant j \leqslant \ell\}$ és $\{p_i q_j : 1 \leqslant i \leqslant k,\, 1 \leqslant j \leqslant \ell\}$ eloszlásokra alkalmazva következik az egyenlőtlenség. Az egyenlőség feltétele az, hogy $r_{i,j} = p_i q_j$ teljesüljön minden $1 \leqslant i \leqslant k$ és $1 \leqslant j \leqslant \ell$ esetén, ami éppen a ξ és η függetlenségével ekvivalens. (Tulajdonképpen az Alaplemma alkalmazásánál csak az olyan (i, j) indexpárokat szabad figyelembe venni, melyekre $r_{i,j} > 0$ teljesül, ami egyébként azzal ekvivalens, hogy $p_i q_j > 0$.)

(vi) teljes indukcióval bizonyítható (v) felhasználásával. □

4.5 Definíció. *Legyenek ξ és η olyan diszkrét valószínűségi vektorváltozók, melyek lehetséges értékei $\{x_i : 1 \leqslant i \leqslant k\}$ illetve $\{y_j : 1 \leqslant j \leqslant \ell\}$. Jelölje $p_i := \mathbb{P}\{\xi = x_i\}$, $q_j := \mathbb{P}\{\eta = y_j\}$, $r_{i,j} := \mathbb{P}\{\xi = x_i,\, \eta = y_j\}$. Ekkor ξ* ***feltételes eloszlása az $\eta = y_j$ feltétel mellett:***

$$p_{i\,|\,j} := \mathbb{P}\{\xi = x_i \,|\, \eta = y_j\} = \frac{\mathbb{P}\{\xi = x_i,\, \eta = y_j\}}{\mathbb{P}\{\eta = y_j\}} = \frac{r_{i,j}}{q_j}.$$

*A ξ **feltételes entrópiája az $\eta = y_j$ feltétel mellett:***

$$H(\xi \,|\, \eta = y_j) := \sum_{i=1}^{k} p_{i\,|\,j} \log_2 \frac{1}{p_{i\,|\,j}} = \sum_{i=1}^{k} \frac{r_{i,j}}{q_j} \log_2 \frac{q_j}{r_{i,j}}.$$

*A ξ **feltételes entrópiája** azon feltétel mellett, hogy η adott:*

$$H(\xi \,|\, \eta) := \sum_{j=1}^{\ell} H(\xi \,|\, \eta = y_j)\mathbb{P}\{\eta = y_j\} = \sum_{j=1}^{\ell} q_j \sum_{i=1}^{k} p_{i\,|\,j} \log_2 \frac{1}{p_{i\,|\,j}} = \sum_{i=1}^{k}\sum_{j=1}^{\ell} r_{i,j} \log_2 \frac{q_j}{r_{i,j}}.$$

4.6 Állítás. $H(\xi,\eta) = H(\eta) + H(\xi \,|\, \eta) = H(\xi) + H(\eta \,|\, \xi)$, *azaz* $H(\xi \,|\, \eta) = H(\xi,\eta) - H(\eta)$, $H(\eta \,|\, \xi) = H(\xi,\eta) - H(\xi)$.

Bizonyítás.

$$\begin{aligned} H(\xi,\eta) - H(\xi \,|\, \eta) &= \sum_{i=1}^{k}\sum_{j=1}^{\ell} r_{i,j} \log_2 \frac{1}{r_{i,j}} - \sum_{i=1}^{k}\sum_{j=1}^{\ell} r_{i,j} \log_2 \frac{q_j}{r_{i,j}} = \sum_{i=1}^{k}\sum_{j=1}^{\ell} r_{i,j} \log_2 \frac{1}{q_j} \\ &= \sum_{j=1}^{\ell} \left(\log_2 \frac{1}{q_j} \right) \sum_{i=1}^{k} r_{i,j} = \sum_{j=1}^{\ell} q_j \log_2 \frac{1}{q_j} = H(\eta). \end{aligned}$$

□

Hasonlóan bizonyítható a következő „láncszabály”:

4.7 Állítás. $H(\xi_1, \xi_2, \ldots, \xi_n) = H(\xi_1) + H(\xi_2 \,|\, \xi_1) + \cdots + H(\xi_n \,|\, \xi_1, \ldots, \xi_{n-1})$.

4.8 Állítás. *Legyen f egy függvény az η értékkészletén értelmezve. Ekkor $H(\xi|\eta) \leqslant H(\xi|f(\eta))$.*

Bizonyítás. Legyen x és z a ξ illetve az $f(\eta)$ egy-egy lehetséges értéke és $I := \{y \mid f(y) = z, \mathbb{P}(\eta = y) > 0\}$, továbbá

$$p_y := \frac{\mathbb{P}(\xi = x, \eta = y)}{\mathbb{P}(\xi = x, f(\eta) = z)}, \qquad q_y := \frac{\mathbb{P}(\eta = y)}{\mathbb{P}(f(\eta) = z)}.$$

Vegyük észre, hogy $\{p_y \mid y \in I\}$ és $\{q_y \mid y \in I\}$ eloszlások, hiszen $\mathbb{P}(f(\eta) = z) = \sum_{y\in I} \mathbb{P}(\eta = y)$ és hasonlóan $\mathbb{P}(\xi = x, f(\eta) = z) = \sum_{y\in I} \mathbb{P}(\xi = x, \eta = y)$. Az alaplemma alapján azt kapjuk, hogy

$$0 \geqslant \sum_{y\in I} p_y \log_2 \frac{q_y}{p_y} = \sum_{y\in I} \frac{\mathbb{P}(\xi = x, \eta = y)}{\mathbb{P}(\xi = x, f(\eta) = z)} \log_2 \frac{\mathbb{P}(\eta = y)\mathbb{P}(\xi = x, f(\eta) = z)}{\mathbb{P}(f(\eta) = z)\mathbb{P}(\xi = x, \eta = y)}.$$

Innen $\mathbb{P}(\xi = x, f(\eta) = z)$-vel szorozva az egyenletet, majd átrendezve azt, kapjuk, hogy

$$\begin{aligned} c(x,z) &:= \sum_{y\in I} \mathbb{P}(\xi = x, \eta = y) \log_2 \frac{\mathbb{P}(\eta = y)}{\mathbb{P}(\xi = x, \eta = y)} \\ &\leqslant \sum_{y\in I} \mathbb{P}(\xi = x, \eta = y) \log_2 \frac{\mathbb{P}(f(\eta) = z)}{\mathbb{P}(\xi = x, f(\eta) = z)} \\ &= \mathbb{P}(\xi = x, f(\eta) = z \log_2 \frac{\mathbb{P}(f(\eta) = z)}{\mathbb{P}(\xi = x, f(\eta) = z)} =: d(x,z). \end{aligned}$$

Ennélfogva

$$H(\xi|\eta) = \sum_x \sum_z c(x,z) \leqslant \sum_x \sum_z d(x,z) = H(\xi|f(\eta)).$$

□

4.9 Definíció. *A ξ és η valószínűségi változók* ***kölcsönös információja:***

$$I(\xi,\eta) := H(\xi) + H(\eta) - H(\xi,\eta) = \sum_{i=1}^{k} \sum_{j=1}^{\ell} r_{i,j} \log_2 \frac{r_{i,j}}{p_i q_j}.$$

Ez annak az információnak a mennyisége, melyet ξ és η egymásra vonatkozóan tartalmaznak, hiszen $H(\xi,\eta) = H(\xi) + H(\eta) - I(\xi,\eta)$.

4.10 Állítás. *A kölcsönös információ tulajdonságai:*

(i) $I(\xi,\eta) = H(\xi) - H(\xi \,|\, \eta) = H(\eta) - H(\eta \,|\, \xi)$.

(ii) $I(\xi,\eta) \geqslant 0$, *és* $I(\xi,\eta) = 0$ *akkor és csak akkor, ha* ξ *és* η *függetlenek.*

(iii) $I(\xi,\eta) = I(\eta,\xi)$.

(iv) $I(\xi,\xi) = H(\xi)$.

(v) $I(\xi,\eta) \leqslant H(\xi)$, $I(\xi,\eta) \leqslant H(\eta)$, *így* $I(\xi,\eta) \leqslant \min\{H(\xi), H(\eta)\}$.

(vi) *A következő állítások ekvivalensek:*

- $I(\xi,\eta) = H(\xi)$
- $H(\xi \,|\, \eta) = 0$

- *létezik olyan* $f : \mathbb{R} \to \mathbb{R}$ *függvény, hogy* $\mathbb{P}\{\xi = f(\eta)\} = 1$.

Bizonyítás. Az (i)–(v) tulajdonságok könnyen levezethetők az entrópia tulajdonságaiból.

(vi) bizonyításához először tegyük fel, hogy létezik olyan $f : \mathbb{R} \to \mathbb{R}$ függvény, hogy $\mathbb{P}\{\xi = f(\eta)\} = 1$. Ekkor

$$p_{i\,|\,j} = \mathbb{P}\{\xi = x_i \,|\, \eta = y_j\} = \mathbb{P}\{f(\eta) = x_i \,|\, \eta = y_j\}$$

$$= \mathbb{P}\{f(y_j) = x_i \,|\, \eta = y_j\} = \begin{cases} 1 & \text{ha } x_i = f(y_j), \\ 0 & \text{ha } x_i \neq f(y_j), \end{cases}$$

ezért $H(\xi \,|\, \eta = y_j) = \sum_{i=1}^{k} p_{i\,|\,j} \log_2 \frac{1}{p_{i\,|\,j}} = 0$, így $H(\xi \,|\, \eta) = 0$, ezért $I(\xi, \eta) = H(\xi)$.

Fordítva: ha $I(\xi, \eta) = H(\xi)$, akkor $0 = H(\xi \,|\, \eta) = \sum_{j=1}^{\ell} H(\xi \,|\, \eta = y_j)\mathbb{P}\{\eta = y_j\}$. Ezért ha $j \in \{1, 2, \ldots, \ell\}$ olyan, hogy $\mathbb{P}\{\eta = y_j\} > 0$, akkor $0 = H(\xi \,|\, \eta = y_j) = \sum_{i=1}^{k} p_{i\,|\,j} \log_2 \frac{1}{p_{i\,|\,j}}$, amiből az következik, hogy a $\{p_{1\,|\,j}, p_{2\,|\,j}, \ldots, p_{k\,|\,j}\}$ eloszlás elfajult, azaz létezik olyan $i_j \in \{1, 2, \ldots, k\}$ index, melyre $p_{i_j\,|\,j} = 1$. Tekintsünk most egy olyan $f : \mathbb{R} \to \mathbb{R}$ függvényt, melyre teljesül az, hogy az ilyen j indexek esetén $x_{i_j} = f(j)$. Ekkor nyilván teljesül $\mathbb{P}\{\xi = f(\eta)\} = 1$. □

Az (i) állítás alapján egyébként az $I(\xi, \eta)$ kölcsönös információ úgy is interpretálható, mint annak a bizonytalanságnak a mértéke, melyet például az η értékének megadásával a ξ értékével kapcsolatosan eloszlatunk, hiszen ekkor ez a bizonytalanság $H(\xi)$–ről $H(\xi \,|\, \eta)$–ra csökken, ami éppen az $I(\xi, \eta)$ mennyiség.

5. Távközlési csatorna kapacitása

Tekintsünk egy $\mathcal{Y} = \{y_1, y_2, \ldots, y_D\}$ csatornaábécét. Legyen az η valószínűségi változó a csatorna bemenetén leadott kódjel, az $\widetilde{\eta}$ valószínűségi változó pedig a kimeneti oldalon vett kódjel. Ekkor $\widetilde{\eta}$ az η–ra vonatkozóan $I(\eta, \widetilde{\eta})$ mennyiségű információt szolgáltat, tehát egy kódjel $I(\eta, \widetilde{\eta})$ mennyiségű információt továbbít.

Ha **zajmentes** a csatorna, akkor $\eta = \widetilde{\eta}$, ezért $I(\eta, \widetilde{\eta}) = H(\eta)$. (Ekkor nyilván az $\mathcal{X} = \{x_1, x_2, \ldots, x_d\}$ forrásábécé $\mathcal{P} = \{p_1, p_2, \ldots, p_d\}$ eloszlása és a $\mathcal{K} = \{K_1, K_2, \ldots, K_d\}$ kódolási eljárás meghatározzák a $\mathbb{P}\{\eta = y_j\}$, $j = 1, 2, \ldots, D$ valószínűségeket is: ha a kódszóhosszak $\mathcal{L} = \{L_1, L_2, \ldots, L_d\}$, és a K_i kódszóban az y_j kódjel $N_{i,j}$ alkalommal fordul elő, akkor $q_j := \mathbb{P}\{\eta = y_j\} = \sum_{i=1}^{d} \frac{N_{i,j}}{L_i} p_i$.) Akkor lesz $H(\eta)$ maximális, ha η egyenletes eloszlású, azaz $q_1 = q_2 = \ldots = q_D = \frac{1}{D}$, és ekkor $H(\eta) = \log_2 D$, vagyis egy kódjel a lehetséges maximális információt továbbítja. Az x_i betű L_i hosszúságú K_i kódszavának továbbítása maximálisan $L_i \log_2 D$ mennyiségű információt jelent. Mivel ennek a betűnek az előfordulási valószínűsége p_i, így egy betű továbbításakor átvitt információ várható értéke maximálisan $(\log_2 D) \sum_{i=1}^{d} p_i L_i = (\log_2 D)\mathbb{E}(\mathcal{K})$. Tehát a $H(\mathcal{X}) \leqslant (\log_2 D)\mathbb{E}(\mathcal{K})$ egyenlőtlenség úgy is interpretálható, hogy a forrásábécé egy betűje által átlagosan szolgáltatott $H(\mathcal{X})$ mennyiségű információt csak akkor vihetjük át $\mathbb{E}(\mathcal{K})$ átlagos kódszóhosszú kódolási eljárás alkalmazásával a zajmentes csatornán, ha az nem haladja meg az egy kódszóval továbbítható információ lehetséges maximális $(\log_2 D)\mathbb{E}(\mathcal{K})$ mennyiségét! (Azt is tudjuk, hogy ügyesen megválasztott $\mathcal{K}$ irreducibilis kód alkalmazásával, tehát egyértelműen dekódolható módon megvalósítható $H(\mathcal{X})$ mennyiségű információ átvitele, ha $H(\mathcal{X}) > (\log_2 D)(\mathbb{E}(\mathcal{K}) - 1)$.)

Zajos csatornánál előfordulhat, hogy $\widetilde{\eta} \neq \eta$, emiatt az egy kódjellel továbbítható $I(\eta, \widetilde{\eta})$ mennyiségű információ kevesebb lesz $\log_2 D$–nél.

5.1 Definíció. *Tekintsünk egy távközlési csatornát $\mathcal{Y} = \{y_1, y_2, \ldots, y_D\}$ csatornaábécével. Azt mondjuk, hogy a csatorna* ***emlékezet nélküli****, ha a kimeneti oldalon vett $\widetilde{\eta}$ kódjel csak a csatorna bemenetén leadott η kódjeltől függ (vagyis nem befolyásolja az, hogy előtte milyen kódjeleket továbbítottunk a csatornán). Egy emlékezet nélküli csatorna viselkedését a következő* ***átmenetvalószínűségek*** *írják le:*

$$p_{i\,|\,j} := \mathbb{P}\{\widetilde{\eta} = y_i \,|\, \eta = y_j\}, \qquad 1 \leqslant i, j \leqslant D.$$

Egy emlékezet nélküli csatorna ***kapacitása:***

$$C := \sup_{\eta} I(\eta, \widetilde{\eta}),$$

ahol szuprémumot úgy értjük, hogy η *eloszlása az összes lehetséges* $\mathcal{Q} = \{q_1, q_2, \dots, q_D\}$ *eloszlást befutja.*

Vagyis egy emlékezet nélküli csatorna kapacitása az egy kódjellel átvihető információmennyiség várható értékének pontos felső határa. Ezért azt várjuk, hogy egy C kapacitású, emlékezet nélküli csatornán $\mathbb{E}(\mathcal{K})$ átlagos kódszóhosszú kód felhasználásával csak akkor továbbíthatjuk a forrás által szolgáltatott $H(\xi)$ mennyiségű információt, ha teljesül a $H(\mathcal{X}) \leqslant C\mathbb{E}(\mathcal{K})$ egyenlőtlenség.

Nyilván a zajmentes csatornák emlékezet nélküliek, és

$$p_{i\,|\,j} = \begin{cases} 1 & \text{ha } \ i = j, \\ 0 & \text{ha } \ i \neq j, \end{cases}$$

és az előzőek alapján ekkor a kapacitás $C = \log_2 D$, és ezt egyenletes eloszlású bemeneti eloszlással el is lehet érni.

5.2 Állítás. *Ha* $\mathbb{P}\{\eta = y_j\} = q_j \ \ (j = 1, 2, \dots, D)$, *akkor*

$$I(\eta, \widetilde{\eta}) = \sum_{i=1}^{D} \sum_{j=1}^{D} p_{i\,|\,j} q_j \log_2 \frac{p_{i\,|\,j}}{\sum\limits_{i=1}^{D} p_{i\,|\,k} q_k}.$$

Bizonyítás. Nyilván

$$r_{i,j} := \mathbb{P}\{\widetilde{\eta} = y_i,\, \eta = y_j\} = \mathbb{P}\{\widetilde{\eta} = y_i \,|\, \eta = y_j\}\mathbb{P}\{\eta = y_j\} = p_{i\,|\,j} q_j,$$

$$p_i := \mathbb{P}\{\widetilde{\eta} = y_i\} = \sum_{j=1}^{D} r_{i,j} = \sum_{j=1}^{D} p_{i\,|\,j} q_j,$$

ezért

$$I(\eta, \widetilde{\eta}) = \sum_{i=1}^{D} \sum_{j=1}^{D} r_{i,j} \log_2 \frac{r_{i,j}}{p_i q_j} = \sum_{i=1}^{D} \sum_{j=1}^{D} p_{i\,|\,j} q_j \log_2 \frac{p_{i\,|\,j}}{\sum\limits_{i=1}^{D} p_{i\,|\,k} q_k}.$$

□

Egy csatorna kapacitásának meghatározásánál tehát tulajdonképpen a $G : [0,1]^D \to \mathbb{R}$,

$$G(q_1, q_2, \dots, q_D) := \sum_{i=1}^{D} \sum_{j=1}^{D} p_{i\,|\,j} q_j \log_2 \frac{p_{i\,|\,j}}{\sum\limits_{i=1}^{D} p_{i\,|\,k} q_k}$$

függvény feltételes szélsőértékét kell meghatározni azon feltétel mellett, hogy $\sum_{j=1}^{D} q_j = 1$.

5.3 Példa: Bináris szimmetrikus csatorna. Ekkor $D = 2$, és a jelek p valószínűséggel mennek át torzulás nélkül a csatornán, azaz

$$p_{i\,|\,j} = \begin{cases} p & \text{ha } i = j, \\ 1-p & \text{ha } i \neq j. \end{cases}$$

Ezért $j = 0, 1$ esetén

$$H(\widetilde{\eta}\,|\,\eta = y_j) = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p},$$

ami nem függ j–től, így

$$H(\widetilde{\eta}\,|\,\eta) = \sum_{j=0}^{1} H(\widetilde{\eta}\,|\,\eta = y_j)\mathbb{P}\{\eta = y_j\} = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$$

nem függ a bemeneti eloszlástól! Így a kölcsönös információ

$$I(\eta, \widetilde{\eta}) = H(\widetilde{\eta}) - H(\widetilde{\eta}\,|\,\eta) = H(\widetilde{\eta}) - \left(p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p} \right).$$

Mivel $\widetilde{\eta}$ kétféle értéket vehet fel, így $H(\widetilde{\eta}) \leqslant \log_2 2 = 1$, és egyenlőség pontosan akkor van, amikor a kimeneti eloszlás egyenletes. Ez létre is jön, mégpedig pontosan akkor, amikor a bemeneti eloszlás egyenletes, ugyanis $\mathbb{P}\{\widetilde{\eta} = 0\} = p\mathbb{P}\{\eta = 0\} + (1-p)\mathbb{P}\{\eta = 1\}$, $\mathbb{P}\{\widetilde{\eta} = 1\} = p\mathbb{P}\{\eta = 1\} + (1-p)\mathbb{P}\{\eta = 0\}$. Ezért a kapacitás

$$C = 1 - p \log_2 D\frac{1}{p} - (1-p) \log_2 D\frac{1}{1-p}.$$

Tehát $0 \leqslant C \leqslant 1$, és $C = 0$ akkor és csak akkor, ha $p = \frac{1}{2}$, valamint $C = 1$ akkor és csak akkor, ha $p = 1$ vagy $p = 0$.

6. Véletlen keresés

6.1 Példa: Hamis pénz keresése. Tudjuk, hogy 9 darab pénzből 1 hamis, és könnyebb, mint a többi. Kétserpenyős mérleggel (súly nélkül) a legkevesebb méréssel keressük meg a

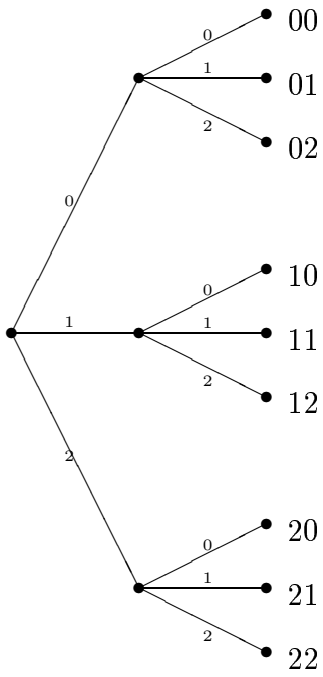
hamisat! Számozzuk meg a pénzdarabokat 1–től 9–ig. Jelölje a ξ valószínűségi változó a hamis pénz sorszámát. Mivel semmilyen előzetes információnk sincs, így $\mathbb{P}\{\xi = k\} = \frac{1}{9}$, azaz ξ egyenletes eloszlású. A ξ–vel kapcsolatos bizonytalanságunk mennyisége $H(\xi) = \log_2 9$, vagyis ennyi információt kell megszereznünk úgy, hogy a két serpenyőbe egyenlő számú pénzdarabot helyezünk, és így meg tudjuk állapítani, hogy a hamis pénz a mérlegre került–e, és ha igen, akkor melyik oldalon van. Jelölje A_1 és A_2 azon pénzdarabok halmazát, amelyeket a mérlegre felteszünk, és jelölje A_0 a kimaradó pénzdarabok halmazát. Jelölje az η valószínűségi változó annak a halmaznak az indexét, amelyikben a hamis pénzdarab van. Egy méréssel tehát az η valószínűségi változó értékét tudjuk megállapítani, vagyis az egy méréssel szerezhető információ mennyisége $I(\xi, \eta)$. Nyilván ξ értéke egyértelműen meghatározza η értékét, azaz létezik olyan $f : \mathbb{R} \to \mathbb{R}$ függvény, hogy $\eta = f(\xi)$, ezért $I(\xi, \eta) = H(\eta)$. Mivel η értékeinek száma 3, ezért $H(\eta) \leqslant \log_2 3$, és $H(\eta) = \log_2 3$ akkor és csak akkor, ha η egyenletes eloszlású, azaz $\mathbb{P}\{\eta = j\} = \frac{1}{3}$, $j = 0, 1, 2$. Az is nyilvánvaló, hogy η eloszlása csak attól függ, hogy hány pénzdarab kerül a mérlegre. Az alábbi táblázat tartalmazza azt, hogy $1-1$, $2-2$, ... pénzdarab mérlegre rakása esetén milyen lesz η eloszlása, és mennyi lesz az egy méréssel szerezhető információ $H(\eta)$ mennyisége:

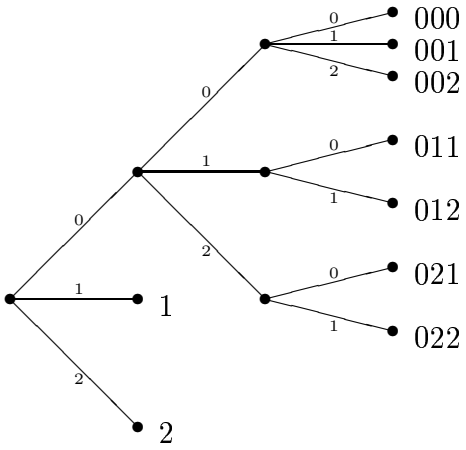
	$\mathbb{P}\{\eta = 0\}$	$\mathbb{P}\{\eta = 1\}$	$\mathbb{P}\{\eta = 2\}$	$H(\eta)$
$1-1$	$\frac{7}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$2\log_2 3 - \frac{7}{9}\log_2 7 \approx 0.99$
$2-2$	$\frac{5}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$2\log_2 3 - \frac{4}{9} - \frac{5}{9}\log_2 5 \approx 1.44$
$3-3$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\log_2 3 \approx 1.58$
$4-4$	$\frac{1}{9}$	$\frac{4}{9}$	$\frac{4}{9}$	$2\log_2 3 - \frac{16}{9} \approx 1.39$

Tehát ha $3-3$ pénzdarabot rakunk fel a mérlegre, akkor megszerezzük az egy méréssel megszerezhető maximális $\log_2 3$ információt. Mivel összesen $\log_2 9 = 2\log_2 3$ információra van szükségünk, így legalább két mérésre van szükség. Viszont két mérés elegendő, hiszen az első mérés után már csak 3 pénzdarabot kell vizsgálnunk, és ha a második mérésnél felteszünk

ezek közül $1-1$ pénzdarabot, akkor biztosan kiderül, melyik a hamis.

Ezzel a stratégiával mindig két mérés szükséges, és két méréssel biztosan el tudjuk dönteni, hogy melyik a hamis pénz. Ha az $1-1$ méréssel kezdünk, akkor előfordulhat, hogy nincs szükség több mérésre (ennek a valószínűsége $\frac{2}{9}$), viszont ha a kimaradó 7–es kupacban van a hamis pénz, akkor nem mindig lesz elegendő egy második mérés! Ha a második mérésnél $2-2$ pénzdarabot rakunk fel, akkor biztosan szükség lesz egy harmadik mérésre is, és így ekkor **a szükséges mérések számának várható értéke:** $\frac{2}{9}\cdot 1+\frac{7}{9}\cdot 3=\frac{23}{9}=2\frac{5}{9}>2$. Ha a második mérésnél $3-3$ pénzdarabot rakunk fel, akkor $\frac{1}{7}$ valószínűséggel előfordulhat, hogy nincs szükség harmadik mérésre, így ekkor a szükséges mérések számának várható értéke: $\frac{2}{9}\cdot 1+\frac{7}{9}\left(\frac{1}{7}\cdot 2+\frac{6}{7}\cdot 3\right)=\frac{22}{9}=2\frac{4}{9}>2$, ami egy kicsit kevesebb, mint az előbb, de még mindig több, mint amit az első startégiával értünk el. Ha a második mérésnél $1-1$ pénzdarabot rakunk fel, akkor az is előfordulhat, hogy a kimaradó 5–ös kupacban van a hamis pénz, és még egy negyedik mérésre is szükség van! Belátható, hogy ekkor a szükséges mérések számának várható értéke legalább $\frac{24}{9}>2$. A többi eset végigkövetésével ellenőrízhető, hogy az összes többi stratégia esetében is több a szükséges mérések számának várható értéke 2–nél. Az is észrevehető, hogy a stratégiákat lehet fagráffal ábrázolni! Például az első: $3-3$ majd $1-1$, illetve a második: $1-1$ majd $2-2$ stratégia fagráfja

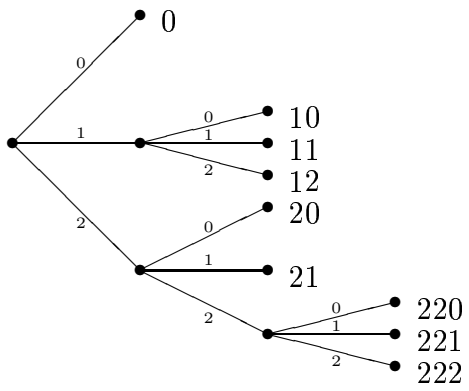




Ezt úgy kell érteni, hogy kiindulunk a gyökértől, és az első mérést az első elágazásnak megfelelően végezzük el: az egyik ágnak megfelelő pénzdarabok az egyik serpenyőbe kerülnek, egy másik ágnak megfelelő pénzdarabok a másik serpenyőbe kerülnek, és ha van még egy harmadik ág is, akkor azokat egyelőre félretesszük. Ezután annak megfelelően, hogy melyik csoportban van a hamis pénzdarab, követjük a fagráfot, és a következő elágazásnak megfelelő mérést hajtjuk végre. A szükséges mérések számának várható értéke akkor lesz a legkisebb, amikor a stratégiához tartozó fagráfnak megfelelő irreducibilis kód átlagos kódszóhossza a legkisebb, azaz amikor a kód optimális! Mivel 9 betűhöz és egyenletes eloszláshoz csak az első fagráf szolgáltat optimális irreducibilis kódot, így már érthető, hogy ha a szükséges mérések számának várható értékét akarjuk minimalizálni, akkor az első stratégia a legjobb.

Ezt a módszert akkor is lehet alkalmazni, amikor ξ nem egyenletes eloszlású, azaz van előzetes információnk arról, hogy melyik pénzdarab milyen valószínűséggel lehet hamis (például a színe alapján). Sajnos nem minden optimális irreducibilis kódhoz lehet kereső algoritmust hozzárendelni! Ugyanis ez csak pontosan abban az esetben megy, ha a fagráf elágazásai által kijelölt részhalmazok között mindig van két egyforma elemszámú! Például az alábbi fagráf egy 9 elemű, $\{3^{-1}, 3^{-2}, 3^{-2}, 3^{-2}, 3^{-2}, 3^{-2}, 3^{-3}, 3^{-3}, 3^{-3}, 3^{-3}\}$ eloszlású for-

rásábécé optimális kódja, de nem feleltethető meg ennek kereső algoritmus:

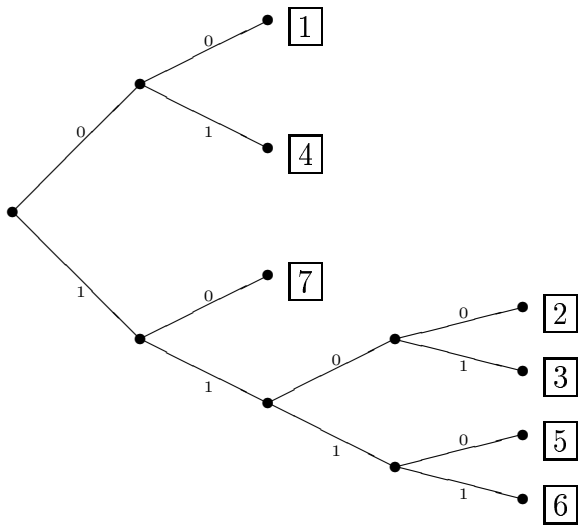


Mivel az összes optimális irreducibilis kód esetén ezek a kódhosszak szerepelnek, így be lehet látni, hogy nincs olyan optimális irreducibilis kód, melynek megfeleltethető lenne kereső algoritmus. Tehát ha 9 pénzdarabunk van, és annak a valószínűsége, hogyaz első, második, ...hamis, rendre $3^{-1}, 3^{-2}, 3^{-2}, 3^{-2}, 3^{-2}, 3^{-2}, 3^{-3}, 3^{-3}, 3^{-3}, 3^{-3}$, akkor végig kell próbálni, hogy melyik stratégia vezet a legkisebb várható mérésszámra.

6.2 Példa: Hibakeresés. Egy tenger alatti kábel víz alatti szakaszán 7 automatikus erősítő van; egy elromlott, amit meg kell keresni. A keresés úgy történik, hogy a kábelt kiemelik a vízből két erősítő között, és megvizsgálják, melyik fele nem vezeti az áramot. Így tehát meg tudjuk állapítani, hogy a kábel még szóba jövő szakaszának melyik felén van a hibás erősítő. Tudjuk, hogy az erősítők meghibásodási esélyei: $\{\frac{1}{4}, \frac{1}{16}, \frac{1}{16}, \frac{1}{4}, \frac{1}{16}, \frac{1}{16}, \frac{1}{4}\}$. Most bináris fagráf feleltethető meg minden kereső algoritusnak, és most is az a cél, hogy az átlagos kódszóhosszat minimalizáljuk. A Huffman–féle kódolási eljárás a következő optimális irreducibilis kódhoz vezet:

K_1	K_2	K_3	K_4	K_5	K_6	K_7
00	1100	1101	01	1110	1111	10

Ennek a gráfja:

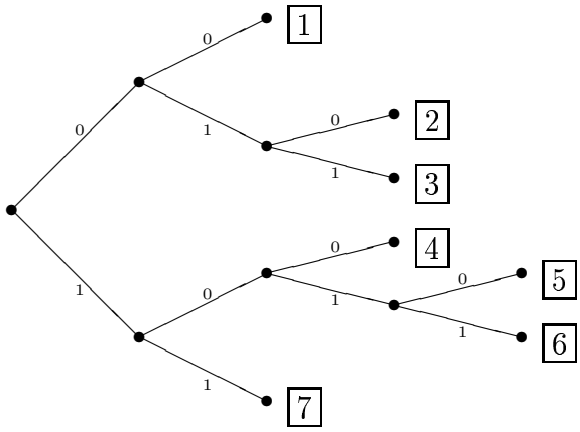


Ez nem felel meg, mert már az első elágazásnál az $\{1,4\}$ és a $\{2,3,5,6,7\}$ csoportokról kellene dönteni, ami a feladat speciális jellege miatt nem lehetséges! A Gilbert–Moore–féle kódolási eljárás viszont megfelelő kódra vezet, mert nem kell nagyság szerint rendezni a valószínűségeket. A kapott kód:

K_1	K_2	K_3	K_4	K_5	K_6	K_7
001	01001	01011	100	10101	10111	111

Ezt nyilván lehet javítani úgy, hogy a felesleges „ágacskákat” levágjuk:

K'_1	K'_2	K'_3	K'_4	K'_5	K'_6	K'_7
001	010	011	100	1010	1011	11



Ez megvalósítható kereső algoritust szolgáltat: először az első elágazás alapján a [3] és [4] erősítők között vizsgáljuk meg a kábelt; ezután az eredménytől függően, a következő elágazást követve vagy az [1] és [2], vagy pedig a [6] és [7] erősítők között végzünk vizsgálatot, stb.

7. Blokkonkénti kódolás

Ebben a fejezetben a 2. fejezetben az (optimális) egyértelmű kódeljárásokra kapott eredmények alapján fogalmazunk meg néhány további fontos állítást. A megközelítés lényege, hogy az eddigi betűnkénti kódeljárás helyett blokkonkénti kódeljárásokkal foglalkozunk és vezetjük le a 2. fejezetbeli egyik legfőbb eredménynek, a McMillan tételnek egyfajta általánosítását, melyet az információelmélet (egyik) alaptételének is szokás nevezni.

Ehhez a 2. fejezethez hasonlóan abból indulunk ki, hogy adottak az alábbiak:

- forrásábécé: $\mathcal{X} = \{x_1, x_2, \ldots, x_d\}$, ahol $d \geqslant 2$, melynek
- eloszlása: $\mathcal{P} = \{p_1, p_2, \ldots, p_d\}$, ahol $p_i > 0$ minden $i = 1, 2, \ldots, d$ esetén, és $\sum_{i=1}^{d} p_i = 1$,
- csatornaábécé: $\mathcal{Y} = \{y_1, y_2, \ldots, y_D\}$, ahol $D \geqslant 2$.

A bolokkonkénti kódolás alapja, hogy a forrásból érkező közleményt n hosszúságú részekre, ún. blokkokra vágjuk szét, ahol n pozitív egész. A kódeljárás nem a forrásábécé betűit kódolja közvetlenül, hanem a lehetséges n hosszúságú közleményrészeket.

Legyen

$$\mathcal{X}^{(n)} := \{(x'_1, x'_2, \ldots, x'_n) \mid x'_i \in \mathcal{X},\ i = 1, 2, \ldots, n\}.$$

Ekkor egy blokkonkénti kódolási eljárás nem más, mint az $\mathcal{X}^{(n)}$ ábécén, mint forrásábécén definiált kódolási eljárás. A továbbiakban feltételezzük, hogy a forrás által kibocsátott X_1, X_2, ..., X_k, ... végtelen jelsorozatra (betűsorozatra) teljesül, hogy adott n pozitív egész esetén az (X_{m+1}, X_{m+2}, ..., X_{m+n}) vektor eloszlása minden $m \geqslant 0$ egész esetén azonos, azaz nem függ m-től. Egy forrát **emlékezet nélkülinek** nevezünk, ha a kibocsátott X_1, X_2, ..., X_k, ... betűk független eloszlásúak. Ez annyit jelent tehát, hogy a soron következő betűt soha nem befolyásolja a korábbi betűk sorozata.

A blokkonkénti kódeljárás $\mathcal{X}^{(n)}$ forrásábécéjének entrópiája nem más, mint

$$H(\mathcal{X}^{(n)}) = H(X_{m+1}, X_{m+2}, \ldots, X_{m+n}) = H(X_1, X_2, \ldots, X_n).$$

A McMillan tétel szerint egy $\mathcal{K}$ blokkonkénti kódeljárás $\mathbb{E}\mathcal{K}$ átlagos kódhosszára teljesül, hogy

$$\mathbb{E}\mathcal{K} \geqslant \frac{H(\mathcal{X}^{(n)})}{\log_2 D}.$$

Innen pedig látható, hogy az $\mathcal{X}$ forrásábécé egy betűjére jutó átlagos kódhossz

$$\bar{\mathbb{E}}\mathcal{K} := \frac{\mathbb{E}\mathcal{K}}{n} \geqslant \frac{H(\mathcal{X}^{(n)})}{n \log_2 D}.$$

A forrás által sugárzott egy betűre eső átlagos információmennyiséget $H(\mathcal{X}^{(n)})/n$ fejezi ki. A következő lemma ennek tulajdonságait foglalja össze.

7.1 Lemma. *Az ezen fejezetben megfogalmazott feltételek mellett a $H(\mathcal{X}^{(n)})/n$ sorozat monoton csökkenő, $H(\mathcal{X}^{(n)})/n \leqslant H(\mathcal{X})$, és létezik határértéke.*

Bizonyítás. Az entrópia tulajdonságaiból adódóan a $H(\mathcal{X}^{(n)})$ sorozat nemnegatív. Továbbá

$$H(\mathcal{X}^{(n)}) = H(X_1, X_2, \ldots, X_n) \leqslant H(X_1) + H(X_2) + H(X_n) = nH(X_1) = nH(\mathcal{X}^{(n)}).$$

A monotonitás igazolásához vegyük észre, hogy

$$\begin{aligned} H(\mathcal{X}^{(n)}) &= H(X_1, X_2, \ldots, X_n) \\ &= H(X_n) + H(X_{n-1}|X_n) + H(X_{n-2}|X_{n-1}, X_n) + \ldots + H(X_1|X_2, \ldots, X_n) \\ &= H(X_1) + H(X_1|X_2) + H(X_1|X_2, X_3) + \ldots + H(X_1|X_2, \ldots, X_n), \end{aligned} \tag{7.2}$$

másrészt azt, hogy a 4.8 Állítás alapján

$$H(X_1) \geqslant H(X_1|X_2) \geqslant H(X_1|X_2, X_3) \geqslant \ldots \geqslant H(X_1|X_2, \ldots, X_n).$$

Ennélfogva (7.2) alapján

$$\begin{aligned} H(X_1|X_2, \ldots, X_{n+1}) &\leqslant \frac{H(X_1) + H(X_1|X_2) + H(X_1|X_2, X_3) + \ldots + H(X_1|X_2, \ldots, X_{n+1})}{n} \\ &= \frac{1}{n} H(\mathcal{X}^{(n)}). \end{aligned}$$

Így az egy betűre jutó entrópia csökkennő (hiszen egy csökkenő sorozat átlagaként adódik):

$$\frac{1}{1+n} H(\mathcal{X}^{(n+1)}) = \frac{n\frac{1}{n} H(\mathcal{X}^{(n)}) + H(X_1|X_2, \ldots, X_{n+1})}{n+1} \leqslant \frac{1}{n} H(\mathcal{X}^{(n)}).$$

Tehát a szóbanforgó sorozat alulról korlátos és monoton csökkenő, így konvergens. □

7.3 Definíció. *A fenti lemmában szereplő* $\bar{H} := \lim_{n\to\infty} H(\mathcal{X}^{(n)})$ *mennyiséget* **a forrás átlagos entrópiájának** *nevezzük.*

Vegyük észre, $\bar{H} \leqslant H(\mathcal{X})$. A 7.1 Lemma bizonyításából az is látszik, hogy $\bar{h} = H(\mathcal{X})$ pontosan akkor teljesülhet, ha a csatorna emlékezet nélküli (hiszen ekkor a bizonyításban az összes egyenlőtlenség helyett írhatunk egyenlőséget a függetlenség miatt).

Az fenti előkészítés után már kimondhatjuk az alaptételt.

7.4 Tétel. (Alaptétel) *Az ezen fejezetben megfogalmazott feltételek mellett minden egyértelműen dekódolható blokkonkénti kódeljárás esetén*

$$\mathbb{E}\bar{\mathcal{K}} \geqslant \frac{\bar{H}}{\log_2 D}.$$

Bármely $\varepsilon > 0$ *esetén létezik olyan blokkonkénti kódolás, melyre*

$$\mathbb{E}\bar{\mathcal{K}} < \frac{\bar{H}}{\log_2 D} + \varepsilon.$$

Bizonyítás. A bizonyítás közvetlenül adódik a McMillan tételből. Hiszen láttuk, hogy bármely $\mathcal{K}$ blokkonkénti kód esetén

$$\mathbb{E}\bar{\mathcal{K}} = \geqslant \frac{H(\mathcal{X}^{(n)})}{n \log_2 D} \geqslant \frac{\bar{H}}{\log_2 D}.$$

Másrészt, ha n elég nagy, akkor $\frac{H(\mathcal{X}^{(n)})}{n} + \frac{1}{n} < \bar{H} + \varepsilon$, hiszen $\frac{H(\mathcal{X}^{(n)})}{n} + \frac{1}{n}$ monoton csökkenő és határértéke $\bar{H}$. Egy ilyen n esetén a McMillan tétel szerint létezik olyan $\mathcal{K}'$ egyértelműen dekódolható blokkonkénti kódeljárás, melyre

$$\mathbb{E}\mathcal{K}' < H(\mathcal{X}^{(n)}) + 1 < n(\bar{H} + \varepsilon),$$

amelyből adódik a bizonyítandó állítás. □

8. Entrópia folytonos változók esetén

A korábbiakban rendre olyan problémákkal foglalkoztunk, amelyekben diszkrét valószínűségi változók szerepeltek. Átekintettük ilyen valószínűségi változók esetén, hogy hogyan lehet az információt (a bizonytalanságot) az entrópia segítségével mérni és ismertettük az entrópia és az ehhez kapcsolódó fogalmak legfontosabb tulajdonságait. Az alábbiakban (abszolút) folytonos valószínűségi változók esetén tekintjük át az analóg fogalmakat. Látni fogjuk, hogy a folytonos változók esetén nem öröklődik minden tulajdonság, melyet a diszkrét esetben ismertettünk. Ám a most következő definíciók mégis lehetővé teszik a legfontosabb tulajdonságok biztosítását.

8.1 Definíció. *Legyen ξ egy abszolút folytonos valószínűségi változó egy $(\Omega, \mathcal{F}, \mathbb{P})$ valószínűségi mezőn, melynek sűrűségfüggvénye f_ξ. Legyen $f_\xi(x)\log f_\xi(x) := 0$, ha $f_\xi(x) = 0$, $x \in \mathbb{R}$. Ekkor a ξ változó* **entrópiája** *alatt a*

$$H(\xi) := -\int_{\mathbb{R}} f_\xi(x)\ln f_\xi(x)dx$$

mennyiséget értjük, amennyiben a fenti integrál létezik[1].

Hasonlóan, ha a $\xi = (\xi_1, \xi_2, \ldots, \xi_n) : \Omega \to \mathbb{R}^n$ valószínűségi vektorváltozó ($n \in \mathbb{N}$) abszolút folytonos egy $f_\xi : \mathbb{R}^n \to \mathbb{R}$ sűrűségfüggvénnyel, akkor ξ entrópiája alatt a

$$H(\xi) := -\int_{\mathbb{R}}\int_{\mathbb{R}}\ldots\int_{\mathbb{R}} f_\xi(x_1, x_2, \ldots, x_n)\ln f_\xi(x_1, x_2, \ldots, x_n)dx_1dx_2\ldots dx_n$$

mennyiséget értjük, amennyiben a fenti integrál létezik.

Diszkrét esetben láttuk, hogy az entrópia nemnegatív és az egyenletes eloszlás esetén veszi fel maximális értékét. A következőekben látni fogjuk, hogy folytonos változók esetén a helyzet nem ilyen egyszerű. Az alábbiakban ismertetünk néhány példát, amelyek bemutatják, hogy tetszőleges valós számot felvehet az entrópia egy alkalmas valóínűségi változó esetén, sőt, tetszőleges $\bar{\mathbb{R}}$-beli értéket felvehet, ahol $\bar{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}$.

[1]Itt jegyezzük meg, hogy elég, ha a Lebesgue mérték szerint integrálható $f_\xi \ln f_\xi$.

Vegyük észre, hogy az entrópia nem változik eltolás esetén, azaz $H(\xi + b) = H(\xi)$, ahol $\xi : \Omega \to \mathbb{R}^n$, $b \in \mathbb{R}^n$, $n \in \mathbb{N}$.

8.2 Példa. Legyen ξ egyenletes eloszlású a $[0, b]$ intervallumon, ahol $b > 0$. Ekkor tudjuk, hogy

$$f_\xi(x) = \begin{cases} 1/b & \text{ha } x \in [0, b] \\ 0 & \text{egyébként.} \end{cases}$$

Ezért

$$H(\xi) = \int_0^b \frac{1}{b} \ln b \, dx = \ln b,$$

amiből látszik, hogy b-t alkalmasan választva tetszőleges valós értéket felvehet az entrópia.

8.3 Példa. Legyen a ξ valószínűségi változó sűrűségfüggvénye

$$f_\xi(x) = \begin{cases} \dfrac{1}{x \left(\ln x\right)^2} & \text{ha } x \geqslant \mathbf{e} \\ 0 & \text{egyébként.} \end{cases}$$

Könnyű látni, hogy a fentiekben valóban sűrűségfüggvényt adtunk meg, hiszen az nemnegatív és

$$\int_{\mathbb{R}} f_\xi(x) dx = \int_{\mathbf{e}}^{\infty} \frac{(\ln x)'}{(\ln x)^2} dx = \int_1^{\infty} \frac{1}{y^2} \mathbf{e}^{-y} \mathbf{e}^y dy = \left[-\frac{1}{y} \right]_1^{\infty} = 1,$$

ahol az $y = \ln x$ ($\frac{dx}{dy} = \mathbf{e}^y$) helyettesítést alkalmaztuk. Ugyanezen helyettesítést alkalmazva kapjuk, hogy ξ entrópiája végtelen, hiszen

$$\begin{aligned} H(\xi) &= \int_{\mathbf{e}}^{\infty} \frac{1}{x \left(\ln x\right)^2} \left(\ln x + \ln \left(\ln x\right)^2 \right) dx \geqslant \int_{\mathbf{e}}^{\infty} \frac{1}{x \left(\ln x\right)^2} \ln x \, dx \\ &= \int_1^{\infty} \frac{1}{y} \mathbf{e}^{-y} \mathbf{e}^y dy = \left[\ln y \right]_1^{\infty} = \infty. \end{aligned}$$

8.4 Példa. Tegyük fel, hogy adottak a valós számegyesen az I_n diszjunkt intervallumok minden $n \geqslant 2$, $n \in \mathbb{N}$ esetén úgy, hogy az intervallumok hossza rendre

$$|I_n| = \frac{1}{(n \ln n)^2} \qquad (\forall n = 2, 3, \ldots).$$

Legyen $K := \sum_{n=2}^{\infty} |I_n|$. Az előző példa számításaiból rögtön adódik, hogy K véges, hiszen

$$K = \sum_{n=2}^{\infty} \frac{1}{(n \ln n)^2} \leqslant \int_1^{\infty} \frac{1}{x\,(\ln x)^2} dx \leqslant \infty.$$

Definiáljuk a ξ valószínűségi változó sűrűségfüggvényét az alábbi módon:

$$f_\xi(x) = \begin{cases} \dfrac{n}{K} & \text{ha } x \in I_n,\ \forall n = 2, 3, \ldots, \\ 0 & \text{egyébként.} \end{cases}$$

Könnyű látni, hogy f_ξ valóban sűrűségfüggvény, hiszen az nemnegatív és

$$\int_{\mathbb{R}} f_\xi(x) dx = \sum_{n=2}^{\infty} \frac{1}{n^2 \ln^2 n} \cdot \frac{n}{K} = \frac{K}{K} = 1.$$

Jegyezzük meg továbbá, hogy $\sum_{n=2}^{\infty} \frac{1}{n \ln n} = \infty$, hiszen

$$\sum_{n=2}^{\infty} \frac{1}{n \ln n} \geqslant \int_1^{\infty} \frac{1}{x \ln x} dx = \int_{\ln 2}^{\infty} \frac{1}{y} \mathbf{e}^{-y} \mathbf{e}^{y} dy = \infty,$$

ahol az $y = \ln x$ ($\frac{dx}{dy} = \mathbf{e}^y$) helyettesítést alkalmaztuk. Végül láthatjuk, hogy $H(\xi) = -\infty$, hiszen

$$\begin{aligned} H(\xi) &= \sum_{n=2}^{\infty} \frac{n}{K} \ln\left(\frac{K}{n}\right) \frac{1}{n^2 \ln^2 n} \\ &= \frac{\ln K}{K} \sum_{n=2}^{\infty} \frac{1}{n \ln^2 n} - \frac{1}{K} \sum_{n=2}^{\infty} \frac{1}{n \ln n} = \ln K - \infty = -\infty. \end{aligned}$$

A diszkét esetben láttuk, hogy számos állítás bizonyításában az Alaplemma játszott központi szerepet. Most ennek ismertetjük folytonos megfelelőjét.

8.5 Lemma. (Alaplemma, folytonos eset.) *Tegyük fel, hogy p és q sűrűségfüggvények.*

(a) *Ha $\int_{\mathbb{R}} p(x) \ln q(x)\, dx$ létezik és véges, akkor $\int_{\mathbb{R}} p(x) \ln p(x)\, dx$ is létezik és*

$$-\int_{\mathbb{R}} p(x) \ln p(x)\, dx \leqslant \int_{\mathbb{R}} p(x) \ln q(x)\, dx. \tag{8.6}$$

Továbbá (8.6)-ban pontosan akkor teljesül az egyenlőség, ha $p(x) = q(x)$ majdnem minden[2] *$x \in \mathbb{R}$ esetén.*

[2] a Lebesgue mérték szerint

(b) *Ha* $\int_{\mathbb{R}} p(x)\ln p(x)\,dx$ *létezik és véges, akkor* $\int_{\mathbb{R}} p(x)\ln q(x)\,dx$ *is létezik és teljesül (8.6). Továbbá (8.6)-ban pontosan akkor teljesül az egyenlőség, ha* $p(x)=q(x)$ *majdnem minden* $x\in\mathbb{R}$ *esetén.*

Bizonyítás. A logaritmus függvény szigorú konkávságából adódóan

$$\ln b \leqslant b-1 \qquad b\in(0,b),$$

továbbá $\ln b = b-1$ pontosan akkor teljesül, ha $b=1$. Jegyezzük meg, hogy a $b\mapsto b-1$ egyenes éppen a logaritmus függvény érintője a $b_0=1$ pontban. Legyen $\frac{q(x)}{p(x)} := 0$ ha $q(x)=p(x)=0$. A fentiek miatt

$$p(x)\ln\frac{q(x)}{p(x)} \leqslant q(x)-p(x), \qquad x\in\mathbb{R}, \tag{8.7}$$

és a (8.7) egyenlőtlenségben az egyenlőség pontosan akkor teljesül, ha $q(x)=p(x)$. Ekkor közvetlenül adódik, hogy

$$\int_{\mathbb{R}} p(x)\ln\frac{q(x)}{p(x)}\,dx \leqslant \int_{\mathbb{R}} q(x)\,dx - \int_{\mathbb{R}} p(x)\,dx = 1-1=0 \tag{8.8}$$

és a (8.8) egyenlőtlenségben pontosan akkor teljesül az egyenlőség, ha $q(x)=p(x)$ majdnem minden $x\in\mathbb{R}$ esetén.

Nézzük most az **(a)** állítás bizonyítását. Majdnem minden $x\in\mathbb{R}$ esetén

$$-p(x)\ln p(x) = p(x)\ln\frac{q(x)}{p(x)} - p(x)\ln q(x), \tag{8.9}$$

hiszen a fenti felbontást csak akkor nem használhatjuk, ha $p(x)>0$ és $q(x)=0$, mert ekkor a bal oldalon $\infty-\infty$ lenne. Ám ez csak egy nullmértékű halmazon teljesülhet, hiszen $\int_{\mathbb{R}} p(x)\ln q(x)\,dx$ véges. A (8.8) egyenlőtlenségből és $\int_{\mathbb{R}} p(x)\ln q(x)\,dx$ végességéből (8.9) alapján adódik, hogy $\int_{\mathbb{R}} p(x)\ln p(x)\,dx$ létezik. Ugyanezekből egyben adódik (8.6) is. Végül azt is láthatjuk, hogy (8.6)-ban egyenlőség pontosan akkor teljesülhet, ha (8.8)-ban egyenlőség van, azaz amikor $q(x)=p(x)$ majdnem minden $x\in\mathbb{R}$ esetén.

A **(b)** állítás bizonyítása teljesen analóg módon adódik. Ekkor az egyetlen különbség, hogy $\int_{\mathbb{R}} p(x)\ln q(x)\,dx$ végességének belátásához a a (8.8) egyenlőtlenség mellett a

$$-p(x)\ln p(x) - p(x)\ln\frac{q(x)}{p(x)} = -p(x)\ln q(x)$$

egyenlőséget kell alkalmaznunk (8.9) helyett. □

8.10 Definíció. *Tegyük fel, hogy $(\xi, \eta) : \Omega \to \mathbb{R}^2$ abszolút folytonos valószínűségi vektorváltozó egy $(\Omega, \mathcal{F}, \mathbb{P})$ valószínűségi mezőn az $f_{(\xi,\eta)}$ sűrűségfüggvénnyel és jelölje ξ-nek η-ra vett feltételes sűrűségfüggvényét $(x, y) \in \mathbb{R}^2$ pontban $f_{(\xi|\eta)}(x|y)$. Ekkor ξ* **feltételes entrópiája** *$\eta = y$* **feltételre nézve**

$$H(\xi|\eta = y) := -\int_{\mathbb{R}} f_{(\xi,\eta)}(x, y) \ln f_{(\xi|\eta)}(x|y)\, dx,$$

míg ξ **feltételes entrópiája** *η-***ra** *nézve*

$$H(\xi|\eta) := -\int_{\mathbb{R}} H(\xi|\eta = y)\, dy = -\int_{\mathbb{R}}\int_{\mathbb{R}} f_{(\xi,\eta)}(x, y) \ln f_{(\xi|\eta)}(x|y)\, dx\, dy,$$

amennyiben a fenti integrálok léteznek.

Mint látni fogjuk a következő tételben, az entrópia és feltételes entrópia legfontosabb tulajdonságai folytonos változók esetén megegyeznek a diszkrét esetben leírtakkal.

8.11 Tétel. *Tegyük fel, hogy $(\xi, \eta) : \Omega \to \mathbb{R}^2$ abszolút folytonos valószínűségi vektorváltozó egy $(\Omega, \mathcal{F}, \mathbb{P})$ valószínűségi mezőn az $f_{(\xi,\eta)}$ sűrűségfüggvénnyel és jelölje ξ-nek η-ra vett feltételes sűrűségfüggvényét $(x, y) \in \mathbb{R}^2$ pontban $f_{(\xi|\eta)}(x|y)$. Továbbá jeölje ξ illetve η (marginális) sűrűségfüggvényeit f_ξ és f_η. Tegyük fel, hogy $H(\xi)$ és $H(\eta)$ véges.*

(a) *$H(\xi, \eta)$ létezik és $H(\xi, \eta) \leqslant H(\xi) + H(\eta)$, továbbá ebben az egyenlőtlenségben pontosan akkor teljesül az egyenlőség, ha ξ és η függetlenek.*

(b) *Létezik $H(\xi|\eta)$, $H(\eta|\xi)$, és $H(\xi, \eta) = H(\xi) + H(\eta|\xi)$.*

(c) *$H(\eta|\xi) \leqslant H(\eta)$ és pontosan akkor teljesül az egyenlőség, ha ξ és η függetlenek.*

Bizonyítás. (a) Mivel az $(x, y) \mapsto f_\xi(x) \cdot f_\eta(y)$ függvény is sűrűségfüggvény, így alkalmazhatjuk az Alaplemmát, amely szerint

$$\begin{aligned} H(\xi, \eta) &= -\int_{\mathbb{R}} \int_{\mathbb{R}} f_{(\xi,\eta)}(x, y) \ln f_{(\xi,\eta)}(x, y)\, dx\, dy \\ &\leqslant -\int_{\mathbb{R}} \int_{\mathbb{R}} f_{(\xi,\eta)}(x, y) \ln f_\xi(x) \cdot f_\eta(y)\, dx\, dy \\ &= -\int_{\mathbb{R}} \ln f_\xi(x) \left(\int_{\mathbb{R}} f_{(\xi,\eta)}(x, y)\, dy \right) dx - \int_{\mathbb{R}} \ln f_\eta(y) \left(\int_{\mathbb{R}} f_{(\xi,\eta)}(x, y)\, dx \right) dy \\ &= H(\xi) + H(\eta) \end{aligned}$$

és egyenlőség pontosan akkor lehetséges, ha $f_{(\xi,\eta)}(x, y) = f_\xi(x) \cdot f_\eta(y)$ teljesül majdnem mindenütt, azaz pontosan akkor, ha ξ és η függetlenek. Az Alaplemmából látszik az is, hogy $H(\xi, \eta)$ létezik, hiszen a fenti egyenlőtlenség bal oldala létezik és véges a feltételek szerint. Jegyezzük meg, hogy az Alaplemmát egydimenziós eloszlások sűrűségfüggvényeire mondtuk ki, ám a bizonyításból látszik, hogy ugyanúgy teljesül, ha többdimenziós eloszlásokra alkalmazzuk.

(b) Rögtön adódik a feltételes sűrűségfüggvény definíciójából, hogy:

$$\begin{aligned} H(\xi|\eta) &= -\int_{\mathbb{R}} \int_{\mathbb{R}} f_{(\xi,\eta)}(x, y) \ln f_{(\xi|\eta)}(x|y)\, dx\, dy \\ &= -\int_{\mathbb{R}} \int_{\mathbb{R}} f_{(\xi,\eta)}(x, y) \ln f_{(\xi,\eta)}(x, y)\, dx\, dy + \int_{\mathbb{R}} \ln f_\xi(x) \left(\int_{\mathbb{R}} f_{(\xi,\eta)}(x, y)\, dy \right) dx \\ &= H(\xi, \eta) - H(\xi). \end{aligned}$$

Az egyenlet utolsó sorában levő mennyiségek léteznek, az egyik véges, így a jobb oldal is létezik.

(c) Az előző két állításból közvetlenül adódik. □

A fenti tétel lehetővé teszi az alábbi definíció bevezetését.

8.12 Definíció. *Tegyük fel, hogy $(\xi, \eta) : \Omega \to \mathbb{R}^2$ abszolút folytonos valószínűségi vektorváltozó egy $(\Omega, \mathcal{F}, \mathbb{P})$ valószínűségi mezőn. Tegyük fel, hogy $H(\xi)$ és $H(\eta)$ véges. Ekkor a ξ és*

η kölcsönös információja

$$I(\xi,\eta) := H(\xi) - H(\xi|\eta) = H(\eta) - H(\eta|\xi).$$

8.13 Következmény. *Tegyük fel, hogy $(\xi,\eta) : \Omega \to \mathbb{R}^2$ abszolút folytonos valószínűségi vektorváltozó egy $(\Omega,\mathcal{F},\mathbb{P})$ valószínűségi mezőn. Tegyük fel, hogy $H(\xi)$ és $H(\eta)$ véges. Ekkor*

(a) $I(\xi,\eta) = H(\xi) + H(\eta) - H(\xi,\eta)$,

(b) $I(\xi,\eta) \geqslant 0$,

(c) $I(\xi,\eta) = 0$ *pontosan akkor teljesül, ha ξ és η függetlenek,*

(d) $I(\xi,\eta) = I(\eta,\xi)$.

Bizonyítás. A 8.11 Tétel tulajdonságaiból a bizonyítás pontosan úgy adódik, mint diszkrét változók esetén. □

Diszkrét változók esetén láttuk, hogy a diszkrét egyenletes eloszlás volt maximális entrópiájú. Az alábbiakban hasonló állításokat mondunk ki folytonos esetre.

8.14 Tétel. *Legyen X normális eloszlású σ^2 szórásnégyzettel, továbbá Y egy olyan valószínűségi változó, melyre $D^2Y \leqslant \sigma^2$. Ekkor*

(a) $H(X) = \frac{1}{2}\left(1 + \ln 2\pi\sigma^2\right)$,

(b) $H(Y) \leqslant H(Y)$.

Bizonyítás. Jelölje q az Y sűrűségfüggvényét és legyen $\mathbb{E}Y = \mu$, $D^2Y = \bar{\sigma}^2$, továbbá

$$p(x) := \frac{1}{\sqrt{2\pi}\bar{\sigma}}\mathbf{e}^{-\frac{(x-\mu)}{2\bar{\sigma}^2}}, \qquad x \in \mathbb{R}.$$

Az Alaplemma miatt

$$\begin{aligned} H(Y) &= -\int_{\mathbb{R}} q(x)\ln q(x)\,dx \leqslant -\int_{\mathbb{R}} q(x)\ln p(x)\,dx \\ &= -\ln\left(\sqrt{2\pi}\bar{\sigma}^2\right)\int_{\mathbb{R}} q(x)\,dx + \int_{\mathbb{R}} q(x)\frac{(x-\mu)}{2\bar{\sigma}^2}\,dx \\ &= \frac{1}{2}\ln\left(2\pi\bar{\sigma}^2\right) + \frac{D^2Y}{2\bar{\sigma}} = \frac{1}{2}\left(1+\ln 2\pi\bar{\sigma}^2\right) \leqslant \frac{1}{2}\left(1+\ln 2\pi\sigma^2\right). \end{aligned}$$

Jegyezzük meg, hogy az Alaplemma egyben azt is eredményezi, hogy $\int_{\mathbb{R}} q(x)\ln q(x)\ dx$ létezik (hiszen $\int_{\mathbb{R}} q(x)\ln p(x)\ dx$ –a fenti levezetésből láthatóan– véges). Speciálisan, ha q helyébe a

$$g(x) := \frac{1}{\sqrt{2\pi}\sigma}\mathbf{e}^{-\frac{(x-\mu)}{2\sigma^2}}, \qquad x \in \mathbb{R}.$$

sűrűségfüggvényét írjuk, akkor a fenti levezetében egyenlőséget kapunk végig, amelyből adódik mindkét állítás. Hiszen, bár g nem feltétlenül az X sűrűségfüggvénye, hanem annak egy eltoltja, de ekkor $H(X) = -\int_{\mathbb{R}} g(x)\ln g(x)\ dx$. □

8.15 Tétel. *Legyen X exponenciális eloszlású $\lambda > 0$ paraméterrel, továbbá Y egy olyan nemnegatív valószínűségi változó, melyre $\mathbb{E}Y \leqslant \frac{1}{\lambda}$. Ekkor*

(a) $H(X) = 1 - \ln\lambda$,

(b) $H(Y) \leqslant H(Y)$.

Bizonyítás. Jelölje q az Y sűrűségfüggvényét és legyen $\mathbb{E}Y = 1/\bar{\lambda}$, továbbá

$$p(x) := \begin{cases} \bar{\lambda}\mathbf{e}^{-\bar{\lambda}x} & x \geqslant 0, \\ 0 & \text{egyébként.} \end{cases}$$

Az Alaplemma miatt

$$\begin{aligned} H(Y) &= -\int_{\mathbb{R}} q(x)\ln q(x)\ dx \leqslant -\int_{\mathbb{R}} q(x)\ln p(x)\ dx \\ &= -\ln\bar{\lambda}\int_{\mathbb{R}} q(x)\ dx + \bar{\lambda}\int_{\mathbb{R}} xq(x)\ dx = -\ln\bar{\lambda} + 1 \leqslant -\ln\lambda + 1 \end{aligned}$$

Jegyezzük meg, hogy az Alaplemma egyben azt is eredményezi, hogy $\int_{\mathbb{R}} q(x)\ln q(x)\ dx$ létezik (hiszen $\int_{\mathbb{R}} q(x)\ln p(x)\ dx$ –a fenti levezetésből láthatóan– véges). Speciálisan, ha q helyébe az X sűrűségfüggvényét írjuk, akkor a fenti levezetében egyenlőséget kapunk végig, amelyből adódik mindkét állítás. □

8.16 Tétel. *Legyen X egyenletes eloszlású az $[a,b]$ intervallumon ($a < b$), továbbá Y egy olyan valószínűségi változó, melyre $\mathbb{P}(Y \in [a,b]) = 1$. Ekkor*

(a) $H(X) = \ln(b-a)$,

(b) $H(Y) \leqslant H(Y)$.

Bizonyítás. Jelölje q az Y sűrűségfüggvényét és

$$p(x) := \begin{cases} \frac{1}{b-a} & x \in [a,b], \\ 0 & \text{egyébként.} \end{cases}$$

Az Alaplemma miatt

$$H(Y) = -\int_{\mathbb{R}} q(x) \ln q(x)\ dx \leqslant -\int_{\mathbb{R}} q(x) \ln p(x)\ dx = \ln(b-a) \int_{\mathbb{R}} q(x)\ dx.$$

Jegyezzük meg, hogy az Alaplemma egyben azt is eredményezi, hogy $\int_{\mathbb{R}} q(x) \ln q(x)\ dx$ létezik (hiszen $\int_{\mathbb{R}} q(x) \ln p(x)\ dx$ –a fenti levezetésből láthatóan– véges). Speciálisan, ha q helyébe is p-t írjuk, azaz X sűrűségfüggvényét, akkor a fenti levezetében egyenlőséget kapunk végig, amelyből adódik mindkét állítás. □

Összefoglalva tehát azt mondhatjuk, hogy: a véges szórású (azaz véges második momentummal rendelkező) eloszlások között a normális eloszlás entrópiája a maximális; a nemnegatív, véges várható értékű eloszlások között az exponenciális eloszlásé maximális; végül a korlátos eloszlások között az egyenletes eloszlásé maximális.

8.17 Példa. (Információátvitel zajos csatornán additív Gauss zaj esetén)
Tekintsünk egy csatornát, ahol a bemenő jelet X jelöli, míg a kimenő jelet Y. Feltesszük, hogy a csatorna a bemenő jelre egy független additív Gauss –azaz normális eloszlású– zajt tesz rá. Tehát feltevésünk szerint $Y = X + Z$, ahol Z eloszlása $\mathcal{N}(0, \sigma_Z)$. Feltesszük továbbá, hogy X abszolút folytonos és $H(X)$ véges.

Az alábbiakban egy ilyen csatornán vizsgáljuk meg, hogy mennyi információ vihető át, azaz célunk a csatorna kapacitásának meghatározása.

Először vegyük észre, hogy

$$\begin{aligned} H(Y|X) &= -\int_{\mathbb{R}}\int_{\mathbb{R}} f_{(Y,X)}(x,y)\ln f_{(Y|X)}(y|x)\ dy\ dx \\ &= -\int_{\mathbb{R}}\int_{\mathbb{R}} f_{(Y|X)}(x|y)\ln f_{(Y|X)}(y|x)\ dy\ f_X(x)\ dx \\ &= -\int_{\mathbb{R}}\int_{\mathbb{R}} f_{(Z}(z)\ln f_{(Z)}(z)\ dz\ f_X(x)\ dx \\ &= -\int_{\mathbb{R}} H(Z)\ dx = H(Z). \end{aligned}$$

Ennélfogva a kölcsönös információ az alábbi módon egyszerűsödik le:

$$I(X,Y) = H(Y) - H(Y|X) = H(Y) - H(Z) = H(Y) - \frac{1}{2}\ln\left(2\pi\mathbf{e}\sigma_Z^2\right).$$

A csatornakapacitást azzal a feltétellel számoljuk, hogy a bemenő jel szórása nem halad meg egy adott $\sigma > 0$ szintet.

Ekkor $D^2Y = D^2X + D^2Z = \sigma^2 + \sigma_Z^2$. Legyen Y_0 eloszlása $\mathcal{N}(0, \sigma^2 + \sigma_Z^2)$. Jegyezzük meg, hogy Y_0 egy lehetséges kimeneti jel, méghozzákkor, ha a bemeneti jel eloszlása $\mathcal{N}(0,\sigma^2)$. Továbbá azt is tudjuk, hogy a legfeljebb $\sigma^2 + \sigma_Z^2$ szórásnégyzetű valószínűségi változók között Y_0 entrópiája maximális. Így végül azt kapjuk, hogy

$$\begin{aligned} \sup_{X,\ D^2X \leqslant \sigma^2} I(X,Y) &= H(Y_0) - H(Z) = \frac{1}{2}\ln\left(2\pi\mathbf{e}(\sigma^2+\sigma_Z^2)\right) - \frac{1}{2}\ln\left(2\pi\mathbf{e}\sigma_Z^2\right) \\ &= \frac{1}{2}\ln\left(\frac{\sigma^2+\sigma_Z^2}{\sigma_Z^2}\right). \end{aligned}$$

9. Feladatok

A.1 Döntsük el, hogy az alábbi táblázatban megadott betűnkénti kódolás közül melyek egyértelműen dekódolhatók.

Ábécé	1. kód	2. kód	3. kód	4. kód	5. kód	6. kód
A:	0	1	0	111	1	0
B:	10	011	10	110	01	01
C:	110	010	110	010	0011	011
D:	1110	001	1110	100	0010	0111
E:	1011	000	11110	011	0001	01111
F:	1101	110	111110	101	0000	011111

A.2 Adjuk meg az A, B, C, D, E, F, G betűkből álló ábácének a 0,1,2 kódjelekből készített olyan irreducibilis kódját, ahol az A és B kódszóhossza 1, C és D kódszóhossza 2, a többi betűé pedig 3. Hány ilyen kód van összesen?

A.3 Mutassuk meg, hogy ha az A, B, C, D betűk előfordulási valószínűségei rendre 1/2, 1/4, 1/8, 1/8, akkor az állandó kódszóhosszúságú bináris kódnál jobb változó kódszóhosszúságú bináris kód is van.

A.4 Döntsük el, hogy az alábbi kódok közül melyek az egyértelműen dekódolhatók.

$\mathcal{K}_1 = \{00,\ 001,\ 010, 00110,\ 011,\ 111,\ 11100,\ 011000\}$

$\mathcal{K}_2 = \{00,\ 001,\ 010,\ 00110,\ 011,\ 111,\ 1110,\ 0110,\ 011100101\}$

$\mathcal{K}_3 = \{123,\ 321,\ 32,\ 011,\ 02,\ 023,\ 33,\ 3323,\ 22\}$

$\mathcal{K}_4 = \{10,\ 110,\ 20,\ 220,\ 2222,\ 1,\ 222\}$

$\mathcal{K}_5 = \{0,\ 10,\ 11,\ 21,\ 20,\ 12,\ 220,\ 2220,\ 22210,\ 2222\}$

$\mathcal{K}_6 = \{0,\ 10,\ 11,\ 21,\ 20,\ 220,\ 2220,\ 12200,\ 12210,\ 121,\ 120\}$

$\mathcal{K}_7 = \{0,\ 10,\ 11,\ 21,\ 20,\ 120,\ 1201,\ 11201\}$

A.5 Soroljuk fel az összes olyan irreducibilis bináris kódot, melyek kódszóhosszai 1, 2, 3, 3.

A.6 Mik lehetnek a kódszóhosszai az $\mathcal{X} = \{A, B, C, D\}$ forrásábécé legfeljebb 3 kódszóhosszúságú irreducibilis bináris kódjainak?

A.7 Készítsünk el a $\mathcal{P} = \{2^{-1}, 2^{-2}, 2^{-3}, 2^{-3}, 2^{-4}, 2^{-4}, 2^{-4}, 2^{-4}\}$ eloszlású ábécéhez egy minimális átlagos kódszóhosszúságú irreducibilis bináris kódot.

A.8 Készítsünk el a $\mathcal{P} = \{3/8, 1/8, 1/6, 1/8, 1/8, 1/12\}$ eloszlású ábécéhez egy minimális átlagos kódszóhosszúságú irreducibilis kódot a 0, 1, 2 kódjelekből.

A.9 Konstruáljunk olyan példát, amikor az optimális kód hatásfoka kisebb mint 1 %.

A.10 Készítsünk el a $\mathcal{P} = \{0.25, 0.33, 0.28, 0.09, 0.03, 0.02\}$ eloszlású ábécéhez egy bináris Huffman–kódot. Mennyi a hatásfoka? Milyen eloszlás esetében lenne ez a kód 100 %–os?

A.11 Bizonyítsuk be, hogy optimális bináris kód kódszóhosszait a forrásábécé eloszlása nem határozza meg egyértelműen.

A.12 Készítsünk bináris Huffman–kódot a következő forrásábécé–eloszlásokhoz:

$$\mathcal{P}_1 = \{0.68,\, 0.17,\, 0.04,\, 0.04,\, 0.03,\, 0.03,\, 0.01\}$$

$$\mathcal{P}_2 = \{0.68,\, 0.27,\, 0.01,\, 0.01,\, 0.01,\, 0.01,\, 0.01\}$$

$$\mathcal{P}_3 = \{0.31,\, 0.18,\, 0.15,\, 0.12,\, 0.09,\, 0.08,\, 0.07\}$$

$$\mathcal{P}_4 = \{0.28,\, 0.21,\, 0.16,\, 0.14,\, 0.13,\, 0.06,\, 0.02\}$$

$$\mathcal{P}_5 = \{0.17,\, 0.16,\, 0.15,\, 0.14,\, 0.13,\, 0.13,\, 0.12\}$$

$$\mathcal{P}_6 = \{0.15,\, 0.15,\, 0.14,\, 0.14,\, 0.14,\, 0.14,\, 0.14\}$$

A.13 Készítsünk bináris Shannon–kódot és Gilbert–kódot a fenti forrásábécé–eloszlásokhoz.

A.14 Átlagosan mennyi információt tartalmaz egy autórendszám, ha az három betűből és három számjegyből áll, ha mind a 26 betű és mind a 10 számjegy egyformán valószínű?

A.15 Maximálisan mennyi információt sugározhat egy TV-adó másodpercenként, ha egy kép 520000 képpontból áll, amelyek háromféle színűek lehetnek, és az adó másodpercenként 50 képet sugároz? Maximálisan mennyi információt tartalmazhat egy 500 betűs szöveg, ha csak a 26 ékezet nélküli betűt vesszük figyelembe? Maximálisan mennyi információt tartalmazhat egy CD-ROM?

B.1 Legyen $\mathcal{K} = \{K_1, K_2, \ldots, K_d\}$ egy olyan irreducibilis, optimális kódja a $\mathcal{P} = \{p_1, p_2, \ldots, p_d\}$ eloszlású forrásábécének, mely D számú kódjelből készült. Jelölje a kószóhosszakat $\mathcal{L} = \{L_1, L_2, \ldots, L_d\}$.

(i) Bizonyítsuk be, hogy nagyobb valószínűségű betűhöz nem tartozhat hosszabb kódszó, azaz ha $p_i > p_j$, akkor $L_i \leqslant L_j$.

(ii) Bizonyítsuk be, hogy $\mathcal{K}$ kódfájának minden olyan pontjából D elágazás indul ki, amely nem végpont, és nem egy maximális hosszúságú ág utolsó előtti pontja. Bizonyítsuk be, hogy egy maximális hosszúságú ág utolsó előtti pontjából legalább 2 és legfeljebb D elágazás indul ki.

(iii) $D = 2$ esetén minden olyan pontból, mely nem végpont, 2 elágazás indul ki.

(iv) $D = 3$ esetén legfeljebb egy olyan pont lehet, mely egy maximális hosszúságú ág utolsó előtti pontja, és nem 3 elágazás indul ki belőle, hanem csak 2.

(v) $D \geqslant 4$ esetén lehet több olyan pont is, mely egy maximális hosszúságú ág utolsó előtti pontja, és nem D elágazás indul ki belőle.

B.2 Legyenek egy forrásábécé valószínűségei nagyság szerint rendezve: $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_d$. Legyen a csatornaábécé $\mathcal{Y} = \{y_1, y_2, \ldots, y_D\}$. Legyen $r \in \{2, \ldots, D\}$ az a szám, melyre $r \equiv d \pmod{D-1}$ teljesül.

(i) Bizonyítsuk be, hogy létezik olyan $\mathcal{K} = \{K_1, K_2, \ldots, K_d\}$ egy olyan irreducibilis, optimális kód, melynél az r utolsó betű kódszava, $K_{d-r+1}, \ldots, K_{d-1}, K_d$ csak az utolsó kódjelükben különböznek.

(ii) Tekintsük a $\mathcal{P}' = \{p_1, p_2, \ldots, p_{d-r}, p_{d-r+1} + \cdots + p_{d-1} + p_d\}$ eloszlású $\mathcal{X}' = \{x_1, x_2, \ldots, x_{d-r}, x'_{d-r+1}\}$ forrásábécét.

- Legyen $\mathcal{K}' = \{K'_1, K'_2, \ldots, K'_{d-r}, K'_{d-r+1}\}$ az $\mathcal{X}'$–nek egy optimális, irreducibilis kódja. Ekkor a K'_{d-r+1} kódszó r–féle kiegészítésével kapott $\mathcal{K} = \{K'_1, K'_2, \ldots, K'_{d-r}, K'_{d-r+1}y_1, \ldots, K'_{d-r+1}y_r\}$ kód $\mathcal{X}$–nek egy optimális, irreducibilis kódja.
- Legyen $\mathcal{K} = \{K_1, K_2, \ldots, K_d\}$ az $\mathcal{X}$–nek egy olyan optimális, irreducibilis kódja, melynél az utolsó r kódszó, $K_{d-r+1}, \ldots, K_{d-1}, K_d$ csak az utolsó kódjelükben különböznek, azaz létezik olyan $\widetilde{K}$ kódjelsorozat, melynek különböző kiegészítései. Ekkor a $\mathcal{K}' = \{K_1, K_2, \ldots, K_{d-r}, \widetilde{K}\}$ kód $\mathcal{X}'$–nek egy optimális, irreducibilis kódja.

B.3 Egy fagráf akkor és csak akkor lehet egy D számú kódjelből alkotott 100 %–os hatásfokú irreducibilis kód kódfája, ha minden olyan pontból, mely nem végpontja egy ágnak, pontosan D számú elágazás indul ki. Határozzuk meg $D = 2$ és $d \leqslant 6$,

illetve $D = 3$ és $d \leqslant 9$ esetén az összes olyan 100 %–os hatásfokú irreducibilis kód kódfáját, melynél $L_1 \leqslant L_2 \leqslant \ldots \leqslant L_d$. Adjuk meg a megfelelő eloszlást is!

B.4 Egy fagráf akkor és csak akkor lehet egy D számú kódjelből alkotott optimális irreducibilis kód kódfája, ha pontosan D számú elágazás indul ki minden olyan pontból, mely nem végpontja egy ágnak, és nem egy maximális hosszúságú ág utolsó előtti pontja, továbbá egy maximális hosszúságú ág utolsó előtti pontjából legalább 2 és legfeljebb D elágazás indul ki. Bizonyítsuk be, hogy $D = 2$ esetén az optimális irreducibilis kódok kódfáinak halmaza és a 100 %–os hatásfokú irreducibilis kódok kódfáinak halmaza egybeesik. Határozzuk meg $D = 3$ és $d \leqslant 7$ esetén az összes olyan optimális irreducibilis kód kódfáját, melynél $L_1 \leqslant L_2 \leqslant \ldots \leqslant L_d$. Adjunk megfelelő eloszlást is!

B.5 Bizonyítsuk be, hogy ha az $L_1, L_2, \ldots, L_d$ természetes számok olyanok, hogy $\sum_{i=1}^{d} D^{-L_i} = 1$ teljesül valamely $D \geqslant 2$ természetes szám esetén, akkor a $D^{-L_1}, D^{-L_2}, \ldots, D^{-L_d}$ számokat be lehet osztani D számú csoportba úgy, hogy az egyes csoportokba tartozó számok összege minden csoportban egyformán $1/D$.

B.6 Bizonyítsuk be, hogy ha a Shannon–féle bináris kódolási eljárásban az r_i számok helyett az

$$s_i := \sum_{k=1}^{i-1} 2^{-L_k}$$

számokat használjuk, akkor is irreducibilis kódot kapunk, melynek a hatásfoka ugyanaz, de a konstrukció kevesebb számolással jár. Általánosítsuk a Shannon–féle kódolási eljárást $D > 2$ esetre is.

B.7 Bizonyítsuk be, hogy ha a Gilbert–féle bináris kódolási eljárásban az r_i számok helyett az

$$s_i := 2^{-L_i} + 2\sum_{k=1}^{i-1} 2^{-L_k}$$

számokat használjuk, akkor is irreducibilis kódot kapunk, melynek a hatásfoka ugyanaz, de a konstrukció kevesebb számolással jár. Általánosítsuk a Gilbert–féle kódolási

eljárást $D > 2$ esetre is.

B.8 Egy amerikai kisváros diákjainak 75 %–a átment az évvégi vizsgán, 25 %–a megbukott. A sikeres vizsgázók 10 %–ának, míg a sikertelen vizsgázók 50 %–ának van saját kocsija. Mennyi információt jelent egy diák vizsgaeredményére vonatkozóan az, ha tudjuk, hogy saját kocsija van? És ha azt tudjuk, hogy nincs saját kocsija? Átlagosan mennyi információt szolgáltat az, ha megtudjuk, hogy van–e, vagy nincs saját kocsija?

B.9 Egy meteorológus a napok 5/12 részében jósolt esőt, és a prognózisa az esetek 2/5 részében vált be, míg a száraz időre vonatkozó jóslása az esetek 10/11 részében vált be. Ráérő emberek kiszámították, hogy a meteorológus jóslása 12/16 valószínűséggel vált be, míg ha mindig száraz időt jósolt volna, akkor 13/16 valószínűséggel helyes lett volna a prognózisa, tehát az utóbbit kellet volna tennie. Miért nincs igazuk?