Információ- és kódelmélet

Fegyverneki, Sándor

Információ- és kódelmélet

Fegyverneki, Sándor

Miskolci Egyetem

Kelet-Magyarországi Informatika Tananyag Tárház

Kivonat

Nemzeti Fejlesztési Ügynökség http://ujszechenyiterv.gov.hu/ *06 40 638-638*

Lektor

Dr. Kálovics Ferenc

Miskolci Egyetem

A tananyagfejlesztés az Európai Unió támogatásával és az Európai Szociális Alap társfinanszírozásával a TÁMOP-4.1.2-08/1/A-2009-0046 számú Kelet-Magyarországi Informatika Tananyag Tárház projekt keretében valósult meg.

Tartalom

Az ábrák listája

1. fejezet - Bevezetés

A statisztikai hírközléselméletet három fő területre szokás osztani: információelmélet, jeldetektálas és sztochasztikus szűrés.

Jeldetektálás: Legyen $\{\xi_t, t \in T\}$ a megfigyelt sztochasztikus jel. A H_0 hipotézis esetén $\{\xi_t, t \in T\}$ egy mintafüggvény az N_t sztochasztikus zajból, míg a H_1 esetén az $S_t + N_t$ jel +zaj folyamatból. A megfigyelő dönt valamelyik hipotézis javára felhasználva egy megfelelő optimalitási kritériumot, pl. egy teszt statisztikát.

Sztochasztikus filtráció: ez nem más, mint a jelek, adatok szűrése, azaz a megfigyelt jel, adatsor transzformálása valamilyen szempontok szerint.

Az információ fogalma központi szerepet játszik az egyes ember és a társadalom életében, és a tudományos kutatásban. Mindennapi életünk minden pillanatában az információ megszerzés, továbbadás, tárolás problémájával vagyunk elfoglalva. Természetesen más és más a jelentése ugyanannak az információnak a különböző felhasználók számára. Hasonlókat mondhatunk az észlelés, tárolás, érték stb. esetében is. Az adott helyzettől függően szubjektíven döntünk, használjuk fel stb. Ezért nem foglalkozunk az információ fogalmával.

Az információelmélet szempontjából csak az információ mennyisége az érdekes, mint ahogy adattároláskor is mellékes, hogy honnan jöttek és mit jelentenek az adatok. Csak a célszerű elhelyezésükről kell gondoskodni.

Napjainkban már eléggé világos, hogy konkrét tartalmától, megjelenési formájától és felhasználásától elvonatkoztatva beszélhetünk az információ számszerű mennyiségéről, ami éppen olyan pontosan definiálható és mérhető, mint bármely más fizikai mennyiség. Hosszú volt azonban az út, amely ehhez a felismeréshez vezetett. Mindenekelőtt azt kell tisztázni, hogy mikor van egyáltalán a kérdésnek értelme. Persze mindenkinek van valamilyen – többé-kevésbé szubjektív – elképzelése az információ mennyiség fogalmáról, de a köznapi szóhasználatban ez általában az információ konkrét megjelenési formájának terjedelmességéhez, másrészt a hasznosságához és egyéb tulajdonságaihoz kapcsolódik. Ahhoz, hogy jól használható mérőszámot kapjunk, minden esetleges vagy szubjektív tényezőtől el kell vonatkoztatni. Ezek közé soroljuk az információ konkrét tartalmát, formáját és mindent, ami a köznyelvben az információ fogalmához kötődik. Ezt a könyörtelen absztrakciót az indokolja, hogy az információ megszerzésével, feldolgozásával, felhasználásával (tárolás, átalakítás, továbbítás) kapcsolatos gyakorlati problémák között nagyon sok olyan is akad, melynek megoldásához (pl. a kívánt berendezés vagy eljárás megtervezéséhez) az információ számos jellemzője közül kizárólag csak a mennyiséget kell figyelembe venni.

Az információ fogalma olyan univerzális, annyira áthatja a mindennapi életünket és a tudomány minden ágát, hogy e tekintetben csak az energiafogalommal hasonlítható össze. A két fogalom között több szempontból is érdekes párhuzamot vonhatunk. Ha végigtekintünk a kultúra, a tudomány nagy eredményein, a legnagyobb felfedezéseken, azoknak jelentős részét két világosan elkülöníthető osztályba sorolhatjuk.

Az egyik csoportba az energia átalakításával, tárolásával, továbbításával kapcsolatos felfedezések tartoznak. Pl. a tűz felfedezése, a víz- és szélenergia felhasználása, egyszerű gépek kostruálása, az elektromos energia hasznosítása stb.

A másik csoportba az információ átalakításával, tárolásával, továbbításával kapcsolatos felfedezések tartoznak. Pl. az írás, a könyvnyomtatás, a távíró, a fényképezés, a telefon, a rádió, a televízió és a számítógép stb.

Számos, az első csoportba tartozó felfedezésnek megvan a párja a második csoportban.

Még egy szempontból tanulságos párhuzamot vonni az energia- és az információfogalom között. Hosszú időbe telt, amíg kialakult az energiamennyiség elvont fogalma, amelynek alapján a különböző megjelenési formáit, mint pl. a mechanikai energiát, a hőenergiát, a kémiai energiát, az elektromos energiát stb. össze lehetett hasonlítani, közös egységgel lehetett mérni. Erre a felismerésre és egyben az energia-megmaradás elvének a meghatározására a XIX. század közepén jutott el a tudomány. Az információ fogalmával kapcsolatban a megfelelő lépés csak a XX. század közepén történt meg.

Mielőtt rátérnénk az információmennyiség mértékének kialakulására, történetére meghatározzuk, hogy mit is jelent az információ absztrakt formában.

Információn általában valamely véges számú és előre ismert lehetőség valamelyikének a megnevezését értjük.

Nagyon fontos, hogy információmennyiségről csak akkor beszélhetünk, ha a lehetséges alternatívák halmaza adott. De ebben az esetben is csak akkor beszélhetünk az információmennyiség definiálásáról, ha tömegjelenségről van szó, vagyis ha nagyon sok esetben kapunk vagy szerzünk információt arról, hogy az adott lehetőségek közül melyik következett be. Mindig ez a helyzet a híradástechnikában és az adatfeldolgozásban, de számos más területen is.

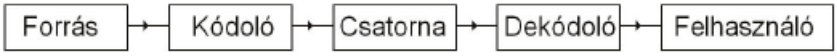
Az információmennyiség kialakulásához a kezdeteket a statisztikus fizika kutatói adták meg. Ebből adódik a fizikában használatos elnevezés (pl. entrópia): L. Boltzmann (1896), Szilárd L. (1929), Neumann J. (1932). Továbbá, a kommunikációelmélettel foglalkozók: H. Nyquist (1924), R.V.L. Hartley (1928).

A hírközlés matematikai elméletét C.E. Shannon (1948) foglalta össze oly módon, hogy hamarosan további, ugrásszerű fejlődés alakuljon ki ezen a területen. Már nemcsak az elmélet alapproblémáit fejti ki, hanem úgyszólván valamennyi alapvető módszerét és eredményét megadja.

Párhuzamosan fejlesztette ki elméletét N. Wiener (1948), amely erősen támaszkodott a matematikai statisztikára és elvezetett a kibernetikai tudományok kifejlődéséhez.

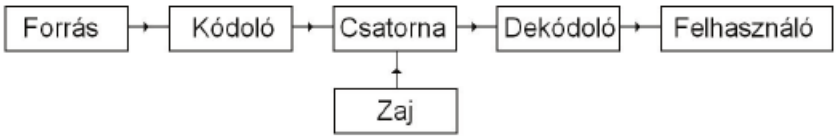
Shannon a következőképpen adta meg a zajmentes (egyirányú) hírközlési rendszer általános modelljét:

1.1. ábra - Az egyirányú hírközlési rendszer általános modellje (zajmentes)



Látható, hogy meg kell oldanunk a következő problémákat: Az üzenet lefordítása továbbítható formára. Az érkező jel alapján az üzenet biztonságos visszaállítása. A fordítás (kódolás) legyen gazdaságos (a dekódolás is) a biztonság megtartása mellett. Használjuk ki a csatorna lehetőségeit (sebesség, kapacitás).

1.2. ábra - Az egyirányú hírközlési rendszer általános modellje (zajos)



Természetesen ezek a problémák már a tervezési szakaszban felmerülnek. Viszont gyakran kerülünk szembe azzal, hogy a már meglévő rendszer jellegzetességeit, kapacitásait kell optimálisan kihasználni. Számos számítástechnikai példa van arra, hogy a biztonságos átvitel mennyire lelassítja az adatáramlást. Továbbá egy jó” kódolás hogyan változtatja az üzenet terjedelmét, a felhasználás gyorsaságát.

Az információelméletet két nagy területre bonthatjuk: az algebrai kódoláselmélet és a Shannon-féle, valószínűség-számításon alapuló, elmélet.

Az információelmélettel foglalkozók a következő három kérdés mennyiségi” vizsgálatával foglalkoznak: Mi az információ? Melyek az információátvitel pontosságának a korlátai? Melyek azok a módszertani és kiszámítási algoritmusok, amelyek a gyakorlati rendszerek esetén a hírközlés és az információtárolás a megvalósítás során megközelíti az előbb említet pontossági, hatékonysági korlátokat?

Az eddigiek alapján a jegyzet anyagát a következő témakörökben foglalhatjuk össze: Az információmennyiség mérése és ennek kapcsolata más matematikai területekkel. A hírközlési rendszerek matematikai modellje (zajos, zajmentes vagy diszkrét, folytonos). Kódoláselmélet (zajos, zajmentes; forrás, csatorna).

1. 1.1. A feldogozott területek címszavakban

Az egyirányú hírközlési rendszer általános modellje. Az információmennyiség Hartley-féle értelmezése, szemléletes jelentése, kapcsolata a blokkonkénti kódolással.

Az A esemény Shannon-féle információmennyisége, axiomatikus bevezetés (elvárt tulajdonság $x \ln(x)$ valószínűségi változó értéke által tartalmazott egyedi információmennyiség, Shannon-féle entrópia, az függvény tulajdonságai, Jensen-egyenlőtlenség, az entrópia tulajdonságai.

Információnyereség és várható értéke, Kullback-Leibler eltérés vagy I-divergencia, az entrópia axiomatikus származtatása, a sztochasztikus függőség mérése, teljes eseményrendszerek sztochasztikus függése, kölcsönös információmennyiség, az I-divergencia tulajdonságai.

Aszimptotikus Stirling-formula, az I-divergencia és a valószínűség kapcsolata, a kölcsönös információmennyiség és az entrópia kapcsolata, McMillan-felbontási (particionálási) tétel, a feltételes entrópia és tulajdonságai, Fano egyenlőtlenség.

Kódoláselméleti fogalmak: stacionaritás, betűnkénti és blokkonkénti kódolás, zajmentesség, emlékezetnélküliség, egyértelmű dekódolhatóság. Keresési stratégiák és prefix kódok, kódfa, átlagos kódhossz. Kraft-Fano egyenlőtlenség, prefix kódok átlagos kódhoszszára vonatkozó állítások. Hatásfok, maximális hatásfokú kód létezése, McMillan-dekódolási tétel (Karush-féle bizonyítás). Shannon-Fano-, Gilbert-Moore-, Huffman-féle kód. Az optimális kód tulajdonságai, a kódfához kapcsolódó tulajdonságok, az optimális kódolás első lépése.

Csatornakapacitás: emlékezetnélküli eset, zajmentes eset, bináris szimmetrikus csatorna. Nem azonos átviteli idő esete: információ átviteli sebesség, csatornakapacitás, optimális eloszlás. A kapacitás numerikus meghatározása, a módszer konvergenciája. Az átlagos időhossz, Kraft-Fano egyenlőtlenség.

Blokkonkénti kódolás, átlagos kódhossz és korlátai, stacionér forrás entrópiája, a zajmentes hírközlés alaptétele, McMillan-felbontási tétel és a zajos kódolás kapcsolata.

Zajos csatorna kódolása: bináris szimmetrikus csatorna, (k, n) kód, algebrai struktúrák, vektortér, a kizáró vagy művelete, norma, Hamming-távolság és tulajdonságai, maximum likelihood kódolás, a hibajavíthatóság és a kódtávolság kapcsolata, csoportkód, hibajelezhetőség, hibaáteresztés, lineáris kód, szisztematikus kód, paritásellenőrző mátrix, szindróma, részcsoport, mellékosztály és tulajdonságai, mellékosztályok és szindrómák kapcsolata, mellékosztályok táblázata, dekódolási táblázat, osztályelsők, a dekódolási táblázat távolság tulajdonsága.

Entrópia és I-divergencia folytonos esetben, tulajdonságok. Speciális eloszlások entrópiája. Entrópia maximalizálás, véges szórású eset.

2. 1.2. JAVA appletek a jegyzethez

A következő problémákhoz készült applet

1. Shannon-Fano kódolás

2. Snanon-Fano kódolás (additív költség esetén)

3. Gilbert-Moore kódolás

4. Gilbert-Moore kódolás (additív költség esetén)

5. Huffman kódolás

6. Csatornakapacitás számítása (additív költség esetén)

Appletek

2. fejezet - Az információmennyiség

1. 2.1. Egyedi információmennyiség, entrópia

A bevezetés alapján információn valamely véges számú és előre ismert lehetőség valamelyikének a megnevezését értjük.

Kérdés: Mennyi információra van szükség egy adott

$$X = \{x_1, x_2, \ldots, x_n\}$$

véges halmaz valamely tetszőleges elemének azonosításához vagy kiválasztásához?

Tekintsük például a jólismert hamis pénz problémát. Itt kétserpenyős mérleg segítségével kell kiválasztani a külsőre teljesen egyforma pénzdarabok közül a könnyebb hamisat. Ez úgy történhet, hogy azonos darabszámú csoportokat téve a mérlegre, megállapítjuk, hogy a keletkezett három csoportból melyikben van a hamis. Ha ugyanis a mérleg egyensúlyban van, akkor a maradékban van, ha nem, akkor a könnyebb csoportban. Ez az eljárás addig folytatódik, amíg megtaláljuk a hamis pénzdarabot.

Ha $n = 3^k$ alakú a pénzdarabok száma, akkor átlagosan k mérlegelésre van szükség, de átlagosan ennél kevesebb már nem vezethet mindig eredményre.

2.1. Megjegyzés. *Általában legalább*

$$\log_3 n$$

mérlegelésre van szükség, ami összefügg azzal, hogy egy mérlegelésnek 3 kimenetele van.

A probléma további vizsgálatára még visszatérünk, viszont előtte tekintsük a következő egyszerű problémát: *Hány bináris számjegy szükséges egy n elemű halmaz elemeinek azonosításához?*

2.1. Példa. Az amerikai hadseregnél állítólag úgy végzik a vérbajosok felkutatását, hogy az egész társaságtól vért vesznek, és a páciensek felének véréből egy részt összeöntve elvégzik a Wassermann-próbát. Amelyik félnél ez pozitív, ott a felezgetést tovább folytatják egész addig, amíg a betegeket ki nem szűrték. Ez a módszer nagyon gazdaságos, mert ha 1000 páciens között pontosan egy vérbajos van, akkor az 10 vizsgálattal lokalizálható, míg az egyenkénti vizsgálatnál – ami adminisztrációs szempontból persze sokkal egyszerűbb – átlagosan 500 próbára van szükség.

▲

Hartley(1928) szerint az n elemű X halmaz elemeinek azonosításához

$$I = \log_2 n$$

mennyiségű információra van szükség.

Ennek az a szemléletes tartalma, hogy ha $n = 2^k$ alakú, akkor $k = \log_2 n$ hosszúságú bináris sorozat szükséges. Ha $n \neq 2^k$ alakú, akkor $[\log_2 n] + 1$ ($[\cdot]$ az egészrészt jelöli) a szükséges bináris jegyek száma. Továbbá, ha azt tekintjük, hogy az általunk vizsgált esetek valamely tömegjelenséghez tartoznak, akkor az a kérdés, hogy az X elemeiből álló tetszőlegesen hosszú sorozatok hogyan írhatók le bináris sorozatokkal.

Tekintsük az m hosszúságú X elemeiből álló sorozatokat, akkor ezek száma n^m. Ha $2^{k-1} < n^m \leq 2^k$, akkor az X halmaz egy elemére eső bináris jegyek száma $\frac{k}{m}$. Ekkor

$$\log_2 n \leq \frac{k}{m} < \log_2 n + \frac{1}{m},$$

azaz m növelésével $\log_2 n$ tetszőlegesen megközelíthető.

Ezek szerint, Hartley formulája az információ mennyiségét a megadáshoz szükséges állandó hosszúságú bináris sorozatok alsó határaként definiálja.

Ennek megfelelően, az információmennyiség egységét *bit*nek nevezzük, ami valószínűleg a "binary digit" angol nyelvű kifejezés rövidítése. Hartley szerint a két elemű halmaz elemeinek azonosításához van szükség egységnyi ($1bit$) mennyiségű információra. Néhány szerző az e alapú természetes logaritmust preferálja, ekkor az egység a *nat*. A logaritmusok közötti átváltás alapján $1bit = \ln 2 nat$.

Hartley egyszerű formulája számos esetben jól használható, de van egy komoly hibája: nem veszi figyelembe, hogy – tömegjelenségről lévén szó – az egyes alternatívák nem feltétlenül egyenértékűek.

Például, nem sok információt nyerünk azzal, hogy ezen a héten sem nyertünk a lottón, mert ezt előre is sejthettük volna, hiszen rendszerint ez történik. Ezzel szemben az ötös találat híre rendkívül meglepő, mert igazán nem számíthatunk rá, ezért az sokkal több információt szolgáltat.

Ezt a nehézséget Shannon(1948) a valószínűség és az információ fogalmának összekapcsolásával oldotta meg. Shannon szerint egy $P(A)$ valószínűségű A esemény bekövetkezése

$$I = \log_2 \frac{1}{P(A)}$$

mennyiségű információt szolgáltat. Ez a mérőszám a Hartley-félénél sokkal árnyaltabb megkülönböztetést tesz lehetővé, és ha az n lehetőség mindegyike egyformán $\frac{1}{n}$ valószínűségű, akkor a Hartley-féle formulára redukálódik.

A továbbiakban először megvizsgáljuk, hogy mennyire természetes a Shannon által bevezetett mérőszám. Az eddigiek alapján a következő tulajdonságokat várjuk el az információmennyiség mérőszámától:

1. *Additivitás:* Legyen $n = NM$ alakú, azaz felírható két természetes szám szorzataként. Ekkor X felbontható N darab diszjunkt M elemű halmaz uniójára, azaz $X = \bigcup_{i=1}^{N} E_i$. Ez azt jelenti, hogy az azonosítása az elemeknek úgy is történhet, hogy először az E_i halmazok egyikét azonosítjuk, s utána az E_i halmazon belül történik az azonosítás. Emlékezzünk vissza a hamis pénz problémára. Ekkor elvárható, hogy a két számítási mód alapján az információmennyiségek megegyezzenek, azaz

$$I(NM) = I(N) + I(M).$$

2.2. Megjegyzés. *Ez a tulajdonság függetlenségként is felírható, mert két egymástól függetlenül elvégzett azonosítás összekapcsolásának felel meg.*

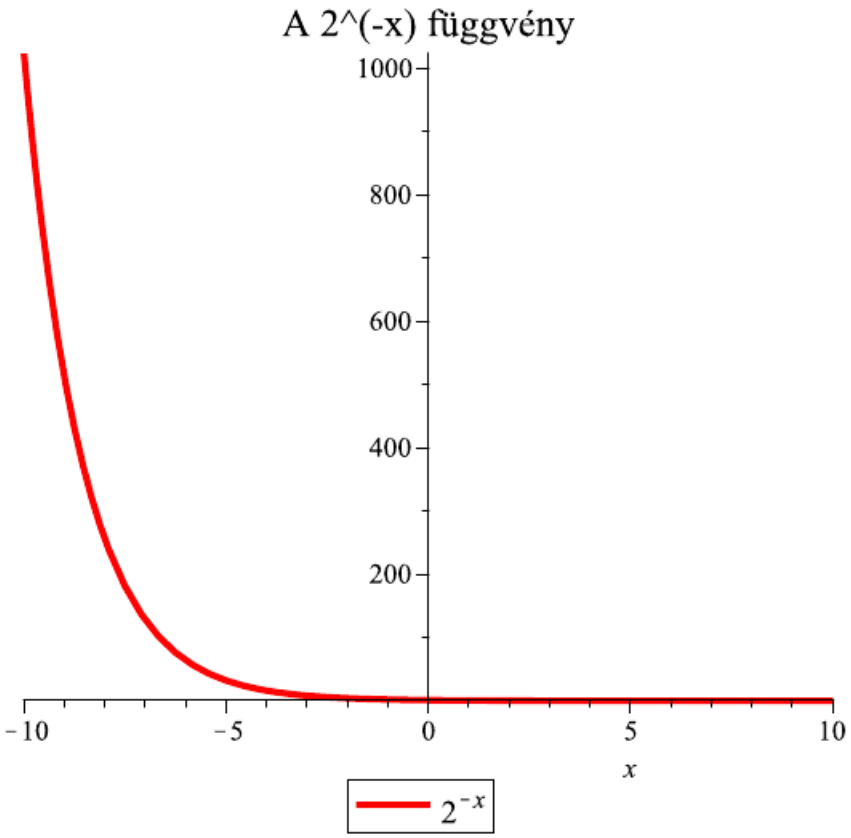
2. *Monotonitás:* A lottós példa alapján elvárható, hogy kisebb valószínűségű esemény bekövetkezése nagyobb információmennyiségű legyen. Ebből viszont rögtön következik, hogy az információmennyiség csak a valószínűségtől függ. Létezik f függvény, hogy az A esemény valószínűségéhez rendelt $I(A) = f(P(A))$. Hiszen $P(A) = P(B)$ esetén $I(A) = I(B)$, mert ha $P(A) \leq P(B)$, akkor $I(A) \geq I(B)$, míg ha $P(A) \geq P(B)$, akkor $I(A) \leq I(B)$.

3. *Normálás:* Legyen $I(A) = 1$, ha $P(A) = \frac{1}{2}$. Ez összhagban van azzal, hogy egy kételemű halmaz elemeinek az azonosításához pontosan $1bit$ információra van szükség.

2.3. Tétel. *Ha* $f : (0,1] \to \mathbf{R}$ *és (1)* $f(p) \geq f(q)$, *ha* $p \leq q$, *(2)* $f(pq) = f(p) + f(q)$, *(3)* $f(\frac{1}{2}) = 1$, *akkor* $f(p) = \log_2 \frac{1}{p}$.

Bizonyítás. Az $x = \log_2 \frac{1}{p}$ jelöléssel az állításunk alakja: $f(2^{-x}) = x$, ha $x \geq 0$. Ezt fogjuk bizonyítani.

2.1. ábra - A 2^{-x} függvény



A (2) feltétel alapján $f(p^n) = nf(p) (n \in \mathbf{N})$, ami teljes indukcióval egyszerűen belátható. Ezt alkalmazva a $p = \frac{1}{2}$ esetre kapjuk, hogy $f(2^{-n}) = n$. Továbbá,

$$2^{-n} = \left(2^{-\frac{n}{m}}\right)^m, \text{ azaz } f(2^{-n}) = mf\left(2^{-\frac{n}{m}}\right),$$

ekkor

$$f\left(2^{-\frac{n}{m}}\right) = \frac{n}{m}.$$

Tehát bármely $0 < x$ racionális számra $f(2^{-x}) = x$. Ha $x = 0$, akkor

$$1 = f(\frac{1}{2}2^0) = f(\frac{1}{2}) + f(2^0) = 1 + f(1), \quad \text{azaz} \quad f(1) = 0.$$

Ha $x > 0$ irracionális, akkor minden $m \in \mathbf{N}$ esetén létezik $n \in \mathbf{N}$, hogy

$$\frac{n}{m} \leq x < \frac{n+1}{m}.$$

Ekkor

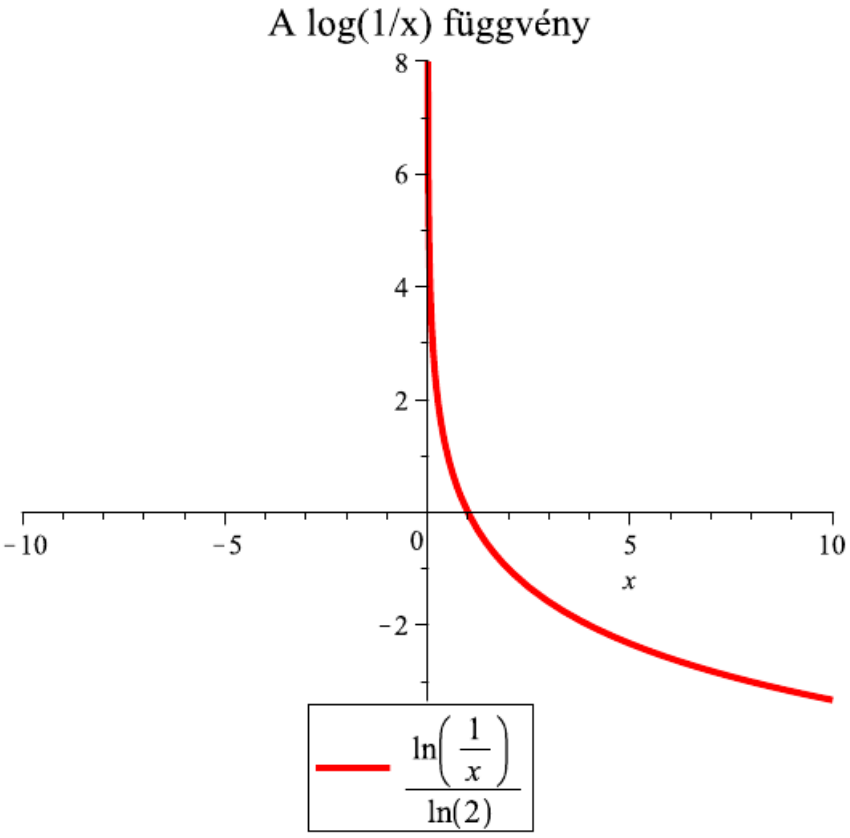
$$\frac{n}{m} = f\left(2^{-\frac{n}{m}}\right) \leq f(2^{-x}) \leq f\left(2^{-\frac{n+1}{m}}\right) = \frac{n+1}{m},$$

amelyből $m \to \infty$ esetén következik, hogy $f(2^{-x}) = x$, ha $x \geq 0$, azaz

$f(p) = \log_2 \frac{1}{p}$.

■

2.2. ábra - A reciprok logaritmusa



2.4. Megjegyzés. *Néhány alapvető irodalom, amelyben az alapfogalmak és tulajdonságaik megtalálhatóak: [3], [12], [7], [8].*

2.5. Definíció. *Az* $I(\xi = x) = \log_2 \frac{1}{P(\xi = x)}$ *mennyiséget a* ξ *valószínűségi változó* x *értéke által tartalmazott egyedi információmennyiségnek nevezzük.*

2.6. Definíció. *A*

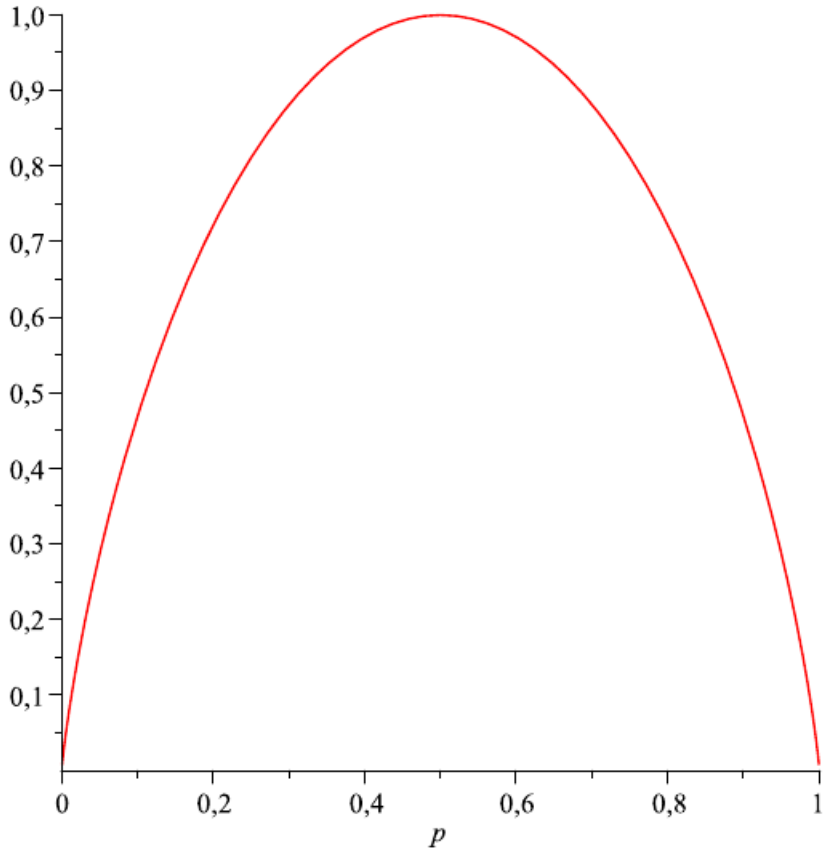
$\mathcal{P} = \{p_1, p_2, \ldots, p_n\}$

eloszlású ξ *valószínűségi változó Shannon-féle entrópiájának nevezzük a*

$H(\xi) = -\sum_{i=1}^{n} p_i \log_2 p_i$

mennyiséget.

2.3. ábra - Az entrópia függvény bináris esetben

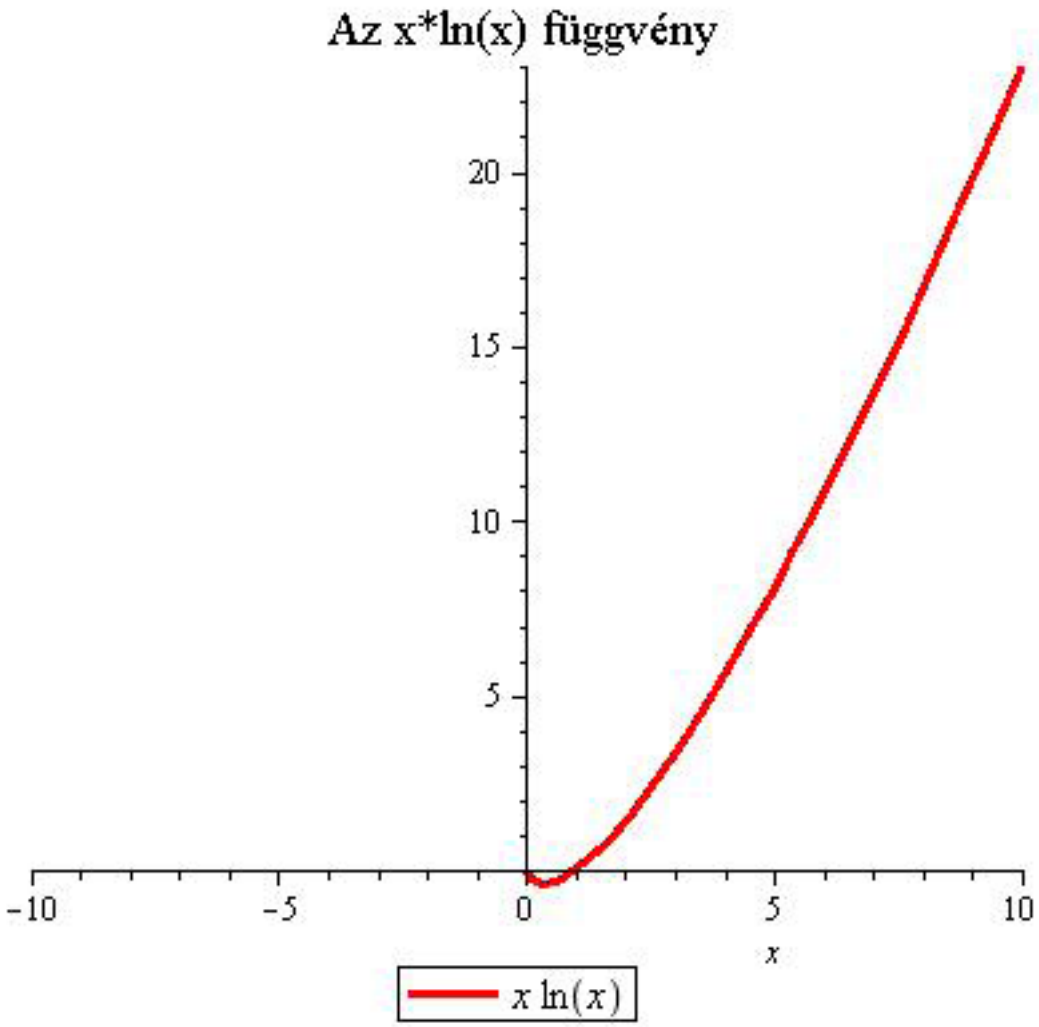


2.7. Megjegyzés. *A valószínűségek között a* 0 *is előfordulhat, így problémát okozhat, hiszen a logaritmus függvény csak pozitiv számokra értelmezett. Ezt azonban megoldja az, hogy az* $x \log_2 x$ *függvény folytonosan kiterjeszthető a nullára, mert*

$$\lim_{x \to 0+0} x \log_2 x = 0, \quad \text{azaz} \quad 0 \log_2 0 = -0 \log_2 \frac{1}{0} = 0$$

lehet definíció szerint.

2.4. ábra - Az $xln(x)$ **függvény**



Vegyük észre, hogy a $H(\xi)$ mennyiség nem más, mint az egyedi információmennyiség várható értéke.

Ha nem okoz zavart, akkor az entrópia jelölésére még a következőket is fogjuk használni:

$H(\xi) = H(\mathcal{P}) = H_n(p_1, p_2, \ldots, p_n) = H(p_1, p_2, \ldots, p_n).$

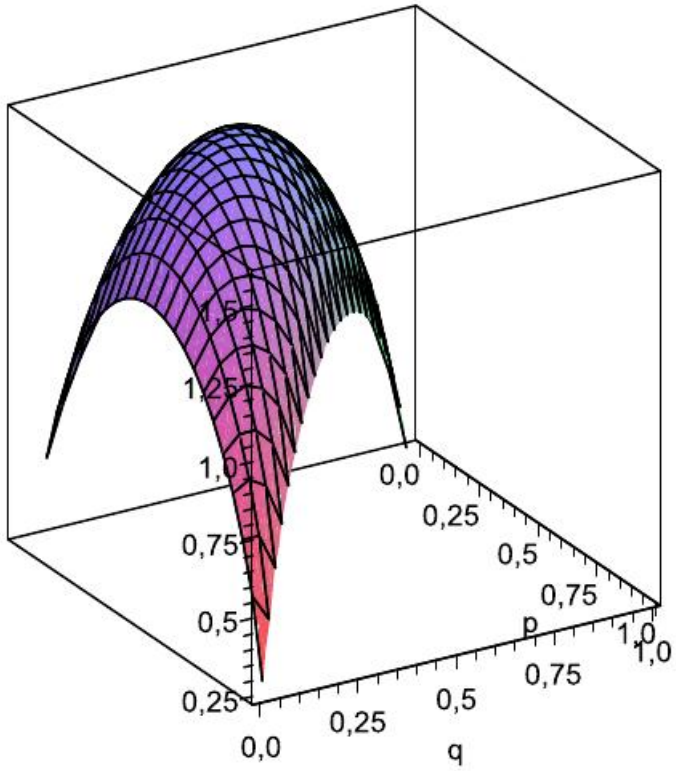
2. 2.2. Az entrópia tulajdonságai

1. $H_n(p_1, p_2, \ldots, p_n) \geq 0.$

Bizonyítás. Az összeg minden tagja nemnegatív.

■

2.5. ábra - Az entrópia függvény három elemű eloszlásra



2. Ha $p_k = 1$ és $p_i = 0 (1 \leq i \leq n, i \neq k)$, akkor $H_n(p_1, p_2, \ldots, p_n) = 0$.

3. $H_{n+1}(p_1, p_2, \ldots, p_n, 0) = H_n(p_1, p_2, \ldots, p_n)$.

4. $H_n(p_1, p_2, \ldots, p_n) \leq H_n\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) = \log_2 n$.

Bizonyítás. A $-\log_2 x$ konvex függvényre alkalmazzuk a Jensen-egyenlőtlenséget.

■

5. $H(\xi)$ folytonos függvény.

6. $H_n(p_1, p_2, \ldots, p_n)$ szimmetrikus a valószínűségekben.

7. Ha $q_n = p_1 + p_2 + \cdots + p_m$, akkor

$$H_{n+m-1}(q_1, q_2, \ldots, q_{n-1}, p_1, p_2, \ldots, p_m) =$$
$$= H_n(q_1, q_2, \ldots, q_n) + q_n H_m(\frac{p_1}{q_n}, \frac{p_2}{q_n}, \ldots, \frac{p_m}{q_n}).$$

Bizonyítás.

$$H_n(q_1,q_2,\dots,q_n)+q_nH_m(\frac{p_1}{q_n},\frac{p_2}{q_n},\dots,\frac{p_m}{q_n})=$$
$$=-\sum_{i=1}^{n}q_i\log_2 q_i-q_n\sum_{i=1}^{m}\frac{p_i}{q_n}\log_2\frac{p_i}{q_n}=$$
$$=-\sum_{i=1}^{n}q_i\log_2 q_i-\sum_{i=1}^{m}p_i(\log_2 p_i-\log_2 q_n)=$$
$$=-\sum_{i=1}^{n-1}q_i\log_2 q_i-q_n\log_2 q_n-\sum_{i=1}^{m}p_i\log_2 p_i+\log_2 q_n\sum_{i=1}^{m}p_i=$$
$$=-\sum_{i=1}^{n-1}q_i\log_2 q_i-\sum_{i=1}^{m}p_i\log_2 p_i=$$
$$=H_{n+m-1}(q_1,q_2,\dots,q_{n-1},p_1,p_2,\dots,p_m).$$

Tehát finomítás hatására az entrópia értéke nem csökkenhet.

■

2.8. Megjegyzés. *Az entrópia axiomatikus származtatása [1], [12]: Ha a fenti tulajdonságok közül megköveteljük, hogy*

(1) $H(\mathcal{P})$ *folytonos a* $\mathcal{P}$ *eloszlásban;*

(2) A $p_i=\frac{1}{n}(1\le i\le n)$ *esethez tartozó* H *monoton növekvő az* n *függvényében;*

(3) Ha $0\le\lambda\le 1$, *akkor*

$$H_{n+1}(p_1,p_2,\dots,\lambda p_n,(1-\lambda)p_n)=H_n(p_1,p_2,\dots,p_n)+p_nH_2(\lambda,1-\lambda).$$

3. 2.3. Feltételes entrópia

$X=\{x_1,\dots,x_n\},Y=\{y_1,\dots,y_m\}.$

Legyen

$(\xi,\eta):\Omega\to X\times Y$

véletlen vektor, melynek együttes eloszlása

$P(\xi=x_i,y_j)=p_{ij}.$

Mivel az entrópiát csak az eloszlás határozza meg, ezért rögtön adódik, hogy

$$H(\mathcal{P})=H(\xi,\eta)=-\sum_{i=1}^{n}\sum_{j=1}^{m}p_{ij}\log_2 p_{ij}.$$

2.9. Definíció. *A*

$$H(\xi|\eta)=\sum_{j=1}^{m}P(\eta=y_j)H(\xi|\eta=y_j)$$

mennyiséget a ξ *valószínűségi változó* η *valószínűségi változóra vonatkozó feltételes entrópiájának nevezzük, ahol*

$$H(\xi|\eta=y_j)=\sum_{i=1}^{n}P(\xi=x_i|\eta=y_j)\log_2\frac{1}{P(\xi=x_i|\eta=y_j)}.$$

2.10. Tétel.

$H(\xi,\eta) = H(\eta) + H(\xi|\eta).$

2.11. Tétel.

$H(\xi) \geq H(\xi|\eta)$

egyenlőség teljesül függetlenség esetén.

4. 2.4. Feladatok

1. n pénzdarab közül az egyik hamis, könnyebb, mint a többi. A többi mind egyenlő súlyú. Legalább hány mérésre van szükség ahhoz, hogy kétserpenyős mérleggel, súlyok nélkül minden esetben meg tudjuk határozni, melyik a hamis.

2. 12 pénzdarab közül az egyik hamis, de nem tudjuk könnyebb-e náluk vagy nehezebb. A többi mind egyenlő súlyú. Igazoljuk, hogy 3 mérés elég ahhoz, hogy kétserpenyős mérleggel, súlyok nélkül minden esetben meg tudjuk határozni, melyik a hamis! Általánosítsuk a feladatot n darabra!

3. Igazolja, hogy $H(\xi,\xi) = H(\xi)$!

4. Igazolja, hogy $H(\xi,\eta) = H(\eta) + H(\xi|\eta)$!

5. Igazolja, hogy $H(\xi) \geq H(\xi|\eta)$!

6. Igazolja, hogy $H(\xi,\eta) \leq H(\eta) + H(\xi)$! Mikor van egyenlőség?

7. Határozza meg az entrópiát a következő eloszláshoz:

$\mathcal{P} = \{0.5, 0.25, 0.125, 0.075, 0.05\}$!

5. 2.5. Önellenőrző kérdések

1. Ismertesse az egyirányú hírközlés általános modelljét!

2. Definiálja az egyedi információmennyiséget!

3. Definiálja az entrópiát!

4. Ismertesse az entrópia tulajdonságait!

5. Definiálja a feltételes entrópiát, ha adott az együttes eloszlás!

6. Adja meg az entrópiát meghatározó axiómákat!

7. Lehet-e az entrópia negatív?

8. Definiálja a feltételes entrópiát!

9. Ismertesse, hogy mely tulajdonságokból adódik a Shannon-féle információmennyiség!

10. Ismertesse a Jensen-egyenlőtlenséget!

3. fejezet - Az I-divergencia

1. 3.1. Információ és bizonytalanság

Egy véletlentől függő kimenetelű kísérlet eredménye több-kevesebb mértékben bizonytalan. A kísérlet elvégzésével ez a bizonytalanság megszűnik. A kísérlet eredményére vonatkozó, erdetileg fennálló bizonytalanságot mérhetjük azzal az információmennyiséggel, amit a kísérlet elvégzésével (átlagban) nyerünk. A bizonytalanságot tehát felfoghatjuk, mint információ hiányt, vagy megfordítva: az információt úgy, mint a bizonytalanság megszüntetését. Az információ betöltése ekvivalens a bizonytalanság megszüntetésével, azaz

információ betöltés= a-priori bizonytalanság – a-posteriori bizonytalanság.

A két fogalom viszonyát jól világítja meg a következő példa:

Ha egy A esemény valószínűsége eredetileg p, de a B esemény megfigyelése után q-ra változott (azaz $P(A) = p$ és $P(A|B) = q$), akkor

$$\log_2 \frac{1}{p} - \log_2 \frac{1}{q} = \log_2 \frac{q}{p}$$

információt nyertünk (vagy vesztettünk). Tehát $\log_2 \frac{q}{p}$ információt szereztünk A-ra nézve. Vegyük észre, hogy

$$\log_2 \frac{q}{p} = \log_2 \frac{P(A|B)}{P(A)} = \log_2 \frac{P(A \cap B)}{P(A)P(B)} = \log_2 \frac{P(B|A)}{P(B)}.$$

Továbbá, hogy az információnyereség 0, ha A és B függetlenek.

Egy $\mathcal{K}$ kísérlet lehetséges kimeneteleinek egy teljes eseményrendszere legyen az $A_1, A_2, \ldots, A_n$, amelyek (a-priori) valószínűsége $p_i = P(A_i)$ számok $(i = 1, 2, \ldots, n)$. Megfigyeltük egy B esemény bekövetkezését, amely kapcsolatban áll a $\mathcal{K}$ kísérlettel. Úgy azon feltétel mellett, hogy B bekövetkezett, az A_i események feltételes (a-posteriori) valószínűségei eltérnek ezek eredeti (a-priori) valószínűségeitől, mégpedig $P(A_i|B) = q_i$.

Kérdés: mennyi információt nyertünk a B esemény megfigyelése által a $\mathcal{K}$ kísérlet várható kimenetelére nézve?

Tudjuk, hogy $\mathcal{P} = \{p_i\}$ és $\mathcal{Q} = \{q_i\}$ eloszlások. Ha nem azonosak, akkor létezik olyan A_k esemény, amelyre $p_k > q_k$ (a bizonytalanság csökkent) és olyan is, amelyre $p_k < q_k$ (a bizonytalanság nőtt). Az információnyereség várható értéke:

$$D(\mathcal{Q}||\mathcal{P}) = \sum_{i=1}^{n} q_i \log_2 \frac{q_i}{p_i}.$$

Ezt a mennyiséget a B esemény megfigyelése által kapott, a $\mathcal{K}$ kísérletre vonatkozó, *Shannon-féle információmennyiség*nek vagy a $\mathcal{P}$ eloszlásnak a $\mathcal{Q}$ eloszlással való helyettesítésénél fellépő információnyereségnek nevezzük.

3.1. Példa. Egy választáson n párt indít jelöltet. Előzetes elképzelésünk az, hogy az egyes pártok jelöltjeire a leadott szavazatokból $p_1, p_2, \ldots, p_n$ rész esik. A választás után megismerjük a tényleges $q_1, q_2, \ldots, q_n$ szavazati arányokat. Az a hír, amely ezt az információt szállította információmennyiséget juttatta birtokunkba,

amely mennyiség jellemzi azt, hogy az eredeti elképzelésünktől milyen messze áll a valóság. Tehát felfogható a két eloszlás közötti eltérés mérőszámaként is.

▲

3.1. Megjegyzés. *Az eloszlások közötti eltérések mérőszámára sokféle próbálkozás történt (Hellinger(1926), Kolmogorov(1931), Mises(1931), Pearson(1905) stb.) Az információmennyiséghez kötődőt a*

$D(\mathcal{Q}||\mathcal{P})$

diszkrimináló információt Kullback és Leibler(1951) vezette be hipotézisvizsgálat felhasználásával. Szokásos elnevezés még az információ divergencia vagy I-divergencia.

2. 3.2. Az I-divergencia tulajdonságai

1. $D(\mathcal{Q}||\mathcal{P}) \geq 0$, egyenlőség akkor és csak akkor, ha $p_i = q_i \, (1 \leq i \leq n)$.

Bizonyítás.

$$D(\mathcal{Q}||\mathcal{P}) = \sum_{i=1}^{n} q_i \log_2 \frac{q_i}{p_i} \geq \frac{1}{\ln 2} \sum_{i=1}^{n} q_i \left(1 - \frac{q_i}{p_i}\right) = \sum_{i=1}^{n} q_i - \sum_{i=1}^{n} p_i = 0.$$

■

2. Ha $q_k = 1$ és $q_i = 0 \, (1 \leq i \leq n, i \neq k)$, akkor $D(\mathcal{Q}||\mathcal{P}) = \log_2 \frac{1}{p_k}$.

3. $D(\mathcal{Q}||\mathcal{P})$ nem szimmetrikus.

Bizonyítás. Tekintsük pédául azt az esetet, amikor

$$p_k = 1 \quad \text{és} \quad p_i = 0 \quad (1 \leq i \leq n, i \neq k).$$

■

4. $D(\mathcal{Q}||\mathcal{P})$ folytonos függvény.

5. $D(\mathcal{Q}||\mathcal{P})$ konvex függvénye a $\mathcal{P}$ eloszlásnak a $\mathcal{Q}$ rögzítése esetén.

6. $D(\mathcal{Q}||\mathcal{P})$ konvex függvénye a $\mathcal{Q}$ eloszlásnak a $\mathcal{P}$ rögzítése esetén.

7. Legyenek $\mathcal{Q}_1$ és $\mathcal{Q}_2$ illetve $\mathcal{P}_1$ és $\mathcal{P}_2$ függetlenek, ekkor

$$D(\mathcal{Q}_1 \times \mathcal{Q}_2||\mathcal{P}_1 \times \mathcal{P}_2) = D(\mathcal{Q}_1||\mathcal{P}_1) + D(\mathcal{Q}_2||\mathcal{P}_2).$$

8. Ha $q_k = \sum_{j=1}^{m_k} q_{kj}$ és $p_k = \sum_{j=1}^{m_k} p_{kj}$ $(k = 1, \ldots, n)$, akkor

$$\sum_{k=1}^{n} \sum_{j=1}^{m_k} q_{kj} \log_2 \frac{q_{kj}}{p_{kj}} \geq \sum_{k=1}^{n} q_k \log_2 \frac{q_k}{p_k},$$

azaz a felosztás (particionálás) finomítása nem csökkenti a diszkrimináló információt. Egyenlőség akkor és csak akkor, ha bármely k és j esetén

$$\frac{q_{kj}}{q_k} = \frac{p_{kj}}{p_k}.$$

Bizonyítás. Az ún. log-szumma egyenlőtlenség alapján bizonyítunk. Legyen $a_1, \ldots, a_n$ és $b_1, \ldots, b_n$ mindegyike nemnegatív, továbbá

$$\sum_{i=1}^{n} a_i = a \quad \text{és} \quad \sum_{i=1}^{n} b_i = b > 0,$$

ekkor

$$\sum_{i=1}^{n} a_i \log_2 \frac{a_i}{b_i} \geq a \log_2 \frac{a}{b}.$$

Egyenlőség akkor és csak akkor, ha bármely $1 \leq i \leq n$ esetén $\frac{a_i}{a} = \frac{b_i}{b}$.

Ha $a = 0$, akkor az állítás nyilvánvaló. Ha $a \neq 0$, akkor legyen

$$q_i = \frac{a_i}{a}, p_i = \frac{b_i}{b}, \quad \text{és} \quad \sum_{i=1}^{n} a_i \log_2 \frac{a_i}{b_i} - a \log_2 \frac{a}{b} = aD(\mathcal{Q}||\mathcal{P}) \geq 0.$$

■

3.2. Megjegyzés. *Legyen* $L(\mathcal{Q}||\mathcal{P}) = \sum_{i=1}^{n} q_i \log_2 p_i$. *Ekkor*

$$D(\mathcal{Q}||\mathcal{P}) = -H(\mathcal{Q}) - L(\mathcal{Q}||\mathcal{P}).$$

Ha $\mathcal{Q}$ rögzített, akkor $D(\mathcal{Q}||\mathcal{P})$ minimális, ha $L(\mathcal{Q}||\mathcal{P})$ maximális, ezért ezt maximum likelihood feladatnak nevezzük. Szokásos elnevezés $L(\mathcal{Q}||\mathcal{P})$ kifejezésre a likelihood illetve a $T(\mathcal{Q}||\mathcal{P}) = -L(\mathcal{Q}||\mathcal{P})$ kifejezésre az inakkurancia.

Ha $\mathcal{P}$ rögzített, akkor $D(\mathcal{Q}||\mathcal{P})$ minimalizálása a minimum diszkrimináló információ feladat.

3. 3.3. A sztochasztikus függőség mérése

A sztochasztikus függetlenség ellentéte a sztochasztikus függőség, ami azonban nem írható le olyan egyértelműen, mint az előbbi, hiszen nem csak egy eset lehetséges, ezért a függőség erősségének jellemzésére megpróbálunk bevezetni egy mérőszámot.

Legyen A és B két esemény, amelyre $P(A) = a$ és $P(B) = b$. Továbbá

$$C_1 = A \cap B, C_2 = A \cap \overline{B}, C_3 = \overline{A} \cap B, C_4 = \overline{A} \cap \overline{B}.$$

A $\{C_i\}$ teljes eseményrendszerhez kétféleképpen kapcsolunk valószínűségeket: a-priori feltételezzük, hogy függetlenek $(P(C_i) = p_i)$ és a-posteriori meghatározzuk (megfigyelés, becslés) a $P(C_i) = q_i$ valószínűségeket. Ekkor meg tudjuk határozni a két eloszlás eltérését.

3.3. Definíció. *Az A és B esemény függőségi mérőszámának nevezzük a*

$$D(\mathcal{Q}||\mathcal{P})$$

diszkrimináló információt.

Jele: $I(A \wedge B)$.

Ha A és B függetlenek, akkor

$$p_1 = ab, p_2 = a(1-b), p_3 = (1-a)b, p_4 = (1-a)(1-b).$$

Ha $P(A \cap B) = x$, akkor

$$q_1 = x, q_2 = a - x, q_3 = b - x, q_4 = 1 - a - b + x.$$

Tehát

$$I(A \wedge B) = x \log_2 \frac{x}{ab} + (a-x) \log_2 \frac{(a-x)}{a(1-b)} +$$
$$+ (b-x) \log_2 \frac{(b-x)}{b(1-a)} + (1-a-b+x) \log_2 \frac{(1-a-b+x)}{(1-a)(1-b)}.$$

Vizsgáljuk meg $I(A \wedge B)$ viselkedését!

1. $D(\mathcal{Q}||\mathcal{P}) \geq 0$, így $I(A \wedge B) \geq 0$.

2. $I(A \wedge B) = I(B \wedge A)$, azaz szimmetrikus.

3. Ha a és b rögzített, akkor

$$\max\{0, a+b-1\} \leq x \leq \min\{a, b\}.$$

Legyen $u = \max\{0, a+b-1\}$ és $v = \min\{a, b\}$, azaz $u \leq x \leq v$. Éz az intervallum sohasem üres, hiszen $ab \in [u, v]$. Innen az is következik, hogy x mindig megválasztható úgy, hogy $I(A \wedge B)$ minimuma elérhető legyen.

4. Legyen $f(x) = I(A \wedge B) \ln 2$, ekkor

$$f'(x) = \ln \frac{x(1-a-b+x)}{(a-x)(b-x)},$$
$$f''(x) = \frac{1}{x} + \frac{1}{1-a-b+x} + \frac{1}{a-x} + \frac{1}{b-x}.$$

Ebből adódik, hogy f konvex, f' monoton növekvő. Könnyen belátható, hogy

$$\lim_{x \to u+0} f'(x) = -\infty, \lim_{x \to v-0} f'(x) = +\infty \quad \text{és} \quad f'(ab) = 0.$$

3.4. Definíció. *Legyenek $A_1, A_2, \ldots, A_n$ és $B_1, B_2, \ldots, B_m$ teljes eseményrendszerek, amelyekre $P(A_i) = q_i$ $(1 \leq i \leq n)$, $P(B_j) = r_j$ $(1 \leq j \leq m)$ és $P(A_i \cap B_j) = p_{ij}$. Ekkor a $\{A_i\}$ és $\{B_j\}$ teljes eseményrendszerek sztochasztikus összefüggésének mérőszáma*

$$I(\{A_i\}, \{B_j\}) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log_2 \frac{p_{ij}}{q_i r_j}.$$

Ezt a mérőszámot kölcsönös információmennyiségnek nevezzük.

3.5. Megjegyzés. *A teljes eseményrendszerek alapján átírható valószínűségi változókra. Jele: $I(\xi, \eta)$.*

4. 3.4. Urnamodellek

Egy urnában n különböző fajtájú golyó van. Legyenek ezek a típusok $a_1, a_2, \ldots, a_n$. Az a_i típus kihúzása jelentse az A_i eseményt és tudjuk, hogy $P(A_i) = p_i$ $(1 \leq i \leq n)$. Húzzunk az urnából visszatevéssel K-szor. Ekkor

$$\Omega = \{\omega | \omega = (a_{i_1}, \ldots, a_{i_K})\} \quad \text{azaz} \quad |\Omega| = n^K.$$

3.6. Definíció. *Legyen $\hat{p}_i = \frac{k_i}{K}$ $(1 \leq i \leq n)$, ahol k_i az A_i esemény bekövetkezéseinek a száma egy adott $\omega \in \Omega$ elemi esemény(minta) esetén. Az $\omega \in \Omega$ minta (K, ε) tipikus (□jó"), ha $|\hat{p}_i - p_i| < \varepsilon$ minden $1 \leq i \leq n$ esetén.*

3.7. Megjegyzés. *A jó minták valószínűsége közel azonosnak tekinthető:*

$$P(\omega) = p_1^{k_1} p_2^{k_2} \cdot \ldots \cdot p_n^{k_n}.$$

$$\log_2 P(\omega) = \sum_{i=1}^{n} k_i \log_2 p_i =$$
$$= -K \sum_{i=1}^{n} \frac{k_i}{K} \log_2 \frac{1}{p_i} =$$
$$= -K \sum_{i=1}^{n} \hat{p}_i \log_2 \frac{1}{p_i}.$$

$$\left| \sum_{i=1}^{n} \hat{p}_i \log_2 \frac{1}{p_i} - \sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i} \right| = \left| \sum_{i=1}^{n} (\hat{p}_i - p_i) \log_2 \frac{1}{p_i} \right| < \varepsilon \left| \sum_{i=1}^{n} \log_2 \frac{1}{p_i} \right|,$$

ahol

$$\left| \sum_{i=1}^{n} \log_2 \frac{1}{p_i} \right|$$

egy korlátos mennyiség, így

$P(\omega) \approx 2^{-KH}$.

Felmerül a kérdés, hogy a tipikus minták mennyire töltik meg az elemi események terét.

Tekintsük rögzített K, n, ε esetén az összes tipikus mintát. Jelöljük ezt C-vel és jelölje B_i azt amikor az i-edik típusú golyó becslése (a relatív gyakoriság) ε-nál közelebb van a valószínűséghez. Ekkor

$$C = B_1 \cap B_2 \cap \cdots \cap B_n = \overline{\overline{B_1} \cup \overline{B_2} \cup \cdots \cup \overline{B_n}},$$

így

$$P(C) = 1 - P\left(\overline{B_1} \cup \overline{B_2} \cup \cdots \cup \overline{B_n}\right) \geq 1 - \sum_{i=1}^{n} P(\overline{B_i}),$$

de $P(B_i) \to 1$ a nagy számok törvénye értelmében. Tehát a □jó" minták összességének valószínűsége tart egyhez.

Az előzőek alapján heurisztikusan az várható, hogy Ω két részre bontható, amelyből az egyik valószínűsége kicsi, a másik pedig közel azonos valószínűségű elemekből áll.

3.8. Tétel. (McMillan felbontási tétel) *Legyen adott az előzőek szerint egy urnamodell. Rögzített $\delta > 0$ esetén létezik K_0, ha $K > K_0$, akkor*

$\Omega = F \cup \overline{F}$,

ahol

1. $P(\overline{F}) < \delta$.
2. Ha $\omega \in F$, akkor $\left| \frac{1}{K} \log_2 \frac{1}{P(\omega)} - H \right| < \delta$.
3. $(1-\delta)2^{K(H-\delta)} \leq |F| \leq 2^{K(H+\delta)}$.

Bizonyítás. Legyen

$$F = \{\omega \,||\log_2 \frac{1}{P(\omega)} - KH| < K\delta\},$$

azaz teljesítse a 2. feltételt. Tehát ha $\omega \in F$, akkor

$2^{-K(H+\delta)} \leq P(\omega) \leq 2^{-K(H-\delta)}$.

Legyen $\xi(\omega) = -\log_2 P(\omega)$, ekkor $E(\xi) = KH$ és a függetlenség miatt

$$D^2(\xi) = K\left(\sum_{i=1}^{n} p_i \left(\log_2 \frac{1}{p_i}\right)^2 - H^2\right).$$

Legyen $D^2(\xi) = K\sigma^2$, akkor a Csebisev-egyenlőtlenség alapján

$$P(\overline{F}) = P(|\xi - KH| > K\delta) \leq \frac{K\sigma^2}{K^2\delta^2} = \frac{1}{K}\frac{\sigma^2}{\delta^2} < \delta,$$

ha K_0 elég nagy.

A 3. rész bizonyításához vegyük észre, hogy

$$1 \geq \sum_{\omega \in F} P(\omega) \geq \sum_{\omega \in F} 2^{-K(H+\delta)} = |F| 2^{-K(H+\delta)},$$

amelyből adódik az állítás egyik fele. Másrészt $P(F) \geq 1 - \delta$, így rögtön következik a másik egyenlőtlenség is.

■

3.9. Megjegyzés. *Ha az urnamodellünk esetén nem a minták valószínűségét vizsgáljuk, hanem a gyakoriságok valószínűségét, akkor a következő érdekes eredményre jutunk.*

Ha az i-edik típus gyakorisága k_i, azaz a relatív gyakoriság

$$\frac{k_i}{K},$$

akkor a relatív gyakoriság közelítése (maximum likelihood becslése) az a-priori p_i valószínűségnek. Mivel a gyakoriságokat később ismerjük meg, így tekinthető a-posteriori valószínűségnek (eloszlásnak). Legyen az A esemény az, hogy a gyakoriságok pontosan

$$k_1, k_2, \ldots, k_n.$$

Tehát

$$P(A) = \frac{K!}{k_1! k_2! \cdots k_n!} p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n}.$$

Ekkor felhasználva az aszimptotikus Stirling-formulát (l. Függelék)

$$\begin{aligned}
\ln P(A) =& \ln K! - \sum_{i=1}^{n} \ln k_i! + \sum_{i=1}^{n} k_i \ln p_i \approx \\
\approx& K \ln K - K - \sum_{i=1}^{n} (k_i \ln k_i - k_i) + \sum_{i=1}^{n} k_i \ln p_i = \\
=& K \ln K - K - \sum_{i=1}^{n} k_i \ln k_i + \sum_{i=1}^{n} k_i + \sum_{i=1}^{n} k_i \ln p_i = \\
=& \sum_{i=1}^{n} \left(k_i \ln \frac{K}{k_i} p_i\right) = \\
=& -K \sum_{i=1}^{n} \left(\frac{k_i}{K} \ln \frac{\frac{k_i}{K}}{p_i}\right).
\end{aligned}$$

Ebből

$$\log_2 P(A) \approx -K \sum_{i=1}^{n} \left(\frac{k_i}{K} \log_2 \frac{\frac{k_i}{K}}{p_i}\right) = -KD(\hat{\mathcal{P}}||\mathcal{P}).$$

Rögzített K esetén, ha nagy az eltérés valószínűségben, akkor nagy az I-divergencia. Ekkor viszont kicsi az ilyen minta valószínűsége. Ezt fejezi ki lényegében a nagy számok törvénye.

5. 3.5. Fano-egyenlőtlenség

3.10. Lemma. *Ha $\eta = y$ esetén a ξ valószínűségi változó $m(y)$ számú értéket vehet fel pozitív valószínűséggel, akkor*

$$H(\xi\,|\eta) \leq \sum_{y\in Y} P(Y = y)\log_2 m(y).$$

Bizonyítás. Az entrópia maximumára vonatkozó egyenlőtlenség alapján

$$H(\xi\,|\eta = y) \leq \log_2 m(y),$$

amelynek várható értékét képezve kapjuk az állítást.

■

3.11. Tétel. (Fano-egyenlőtlenség) *Tegyük fel, hogy a ξ és az η valószínűségi változók ugyanazt az m értéket vehetik fel pozitív valószínűséggel, és legyen*

$$P_e = \sum_{x\neq y} P(\xi = x, \eta = y),$$

ekkor

$$H(\xi|\eta) \leq P_e \log_2(m-1) + H(P_e, 1-P_e).$$

Bizonyítás.

$$H(\xi|\eta) = \sum_{x\neq y} P(\xi = x, \eta = y)\log_2 \frac{P(\eta = y)}{P(\xi = x, \eta = y)} +$$

$$+ \sum_{y} P(\xi = y, \eta = y)\log_2 \frac{P(\eta = y)}{P(\xi = y, \eta = y)}.$$

Mivel

$$\sum_{x\neq y} P(\eta = y) = \sum_{x}(1 - P(\xi = x)) = m - 1,$$

a P_e definíciója alapján adódik, hogy

$$\sum_{x\neq y} P(\xi = x, \eta = y)\log_2 \frac{P(\eta = y)}{P(\xi = x, \eta = y)} \leq$$

$$\leq P_e \log_2 \frac{m-1}{P_e} = P_e \log_2(m-1) + P_e \log_2 \frac{1}{P_e}.$$

Másrészt a feltételes entrópia második tagjánál

$$\sum_{y} P(\xi = y, \eta = y) = 1 - P_e.$$

ezért

$$\sum_{y} P(\xi = y, \eta = y)\log_2 \frac{P(\eta = y)}{P(\xi = y, \eta = y)} \leq (1 - P_e)\log_2 \frac{1}{1 - P_e}.$$

A két felső becslés együttesen kiadja az állítást.

■

3.12. Megjegyzés. *Ha tehát a ξ valószínűségi változót az η valószínűségi változóval akarjuk helyettesíteni, akkor az itt elkövetett hibára alsó becslés adható a feltételes entrópia függvényeként. A Fano-egyenlőtlenség értéke éppen az, hogy a $P(\xi \neq \eta)$ hibavalószínűséget egy információelméleti mérőszámmal becsüli meg.*

6. 3.6. A kölcsönös információmennyiség tulajdonságai

3.13. Tétel. (A kölcsönös információmennyiség és az entrópia kapcsolata)

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

Bizonyítás. A definíció alapján a logaritmus felbontásával rögtön adódik:

$$I(\xi,\eta) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\log_2\frac{p_{ij}}{q_i r_j},$$

$$H(\xi,\eta) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\log_2 p_{ij},$$

$$H(\xi) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\log_2 q_i,$$

$$H(\eta) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\log_2 r_j.$$

■

3.14. Tétel. (A kölcsönös információmennyiség és a feltételes entrópia kapcsolata)

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

Bizonyítás. A feltételes entrópiáról tudjuk, hogy

$$H(\xi,\eta) = H(\eta) + H(\xi|\eta) = H(\xi) + H(\eta|\xi),$$

s így az előző tétel felhasználásával adódik állításunk.

■

7. 3.7. Feladatok

1. Határozza meg a

$$\mathcal{P} = \{2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}, 2^{-5}, 2^{-5}\}$$

eloszlás és a

$$\mathcal{Q} = \{3^{-1}, 3^{-1}, 3^{-2}, 3^{-2}, 3^{-3}, 2\cdot 3^{-3}\}$$

eloszlás Kullback-Leibler eltérését!

2. Igazolja a Bernoulli-féle nagy számok törvényét az I-divergencia felhasználásával!

3. Igazolja, hogy

$$\prod_{i=1}^{n} {x_i}^{p_i} \leq \sum_{i=1}^{n} x_i p_i,$$

ahol $\{p_1, p_2, \ldots, p_n\}$ eloszlás és $x_1, x_2, \ldots, x_n$ pozitív valós számok! Mikor van egyenlőség?

8. 3.8. Önellenőrző kérdések

1. Definiálja a Kullback-Leibler eltérést!

2. Ismertesse az I-divergencia tulajdonságait!

3. Bizonyítsa, hogy az I-divergencia nemnegatív!

4. Definiálja a kölcsönös információmennyiséget!

5. Definiálja teljes eseményrendszerek sztochasztikus függésének a merőszámát!

6. Ismertesse az aszimptotikus Stirling-formulát!

7. Ismertesse a Markov-egyenlőtlenséget!

8. Ismertesse a Csebisev-egyenlőtlenséget!

9. Bizonyítsa a McMillan-felbontási tételt!

10. Ismertesse a polinimiális eloszlást és tulajdonságait!

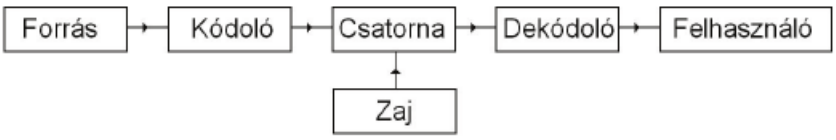
11. Ismertesse a Fano-egyenlőtlenséget!

4. fejezet - Forráskódolás

1. 4.1. Alapfogalmak

A Shannon-féle egyirányú hírközlési modell általános alakja [16], [3]:

4.1. ábra - Az egyirányú hírközlési rendszer általános modellje (zajos)



A hírközlés feladata eljuttatni az információt a felhasználóhoz. A távolságok miatt az információ továbbítására valamilyen eszközöket (csatornákat) használunk, amelyek néhány jól meghatározott típusú jelet tudnak továbbítani. Tehát a továbbításhoz az információt a csatorna típusának megfelelően kell átalakítani. Ez a kódolás, míg a továbbítás után vett jelekből az információnak a visszaalakítását dekódolásnak nevezzük.

További probléma forrása, hogy az átvitel során a továbbított jelek megváltozhatnak, azaz ún. zajos csatornával dolgozunk. Tehát olyan módszerekre is szükség van, melyekkel az ilyen zajos csatornákon is elég megbizhatóan vihető át az információ, és amellett az átvitel költségei, sebessége sem gátolja a használhatóságot.

Az információ ezer alakban jelenhet meg, ám minden csatorna csak jól meghatározott típusú, a csatornára nézve specifikus információkat tud továbbítani. Az üzenetet ezért mindig olyan jelekké kell átalakítanunk, amelyek a rendelkezésünkre álló csatornán átvihetők. A jelek átalakítását kódolásnak nevezzük. Ha egészen pontosak akarunk lenni, azt kell mondanunk, hogy a kommunikációban mindig átkódolást végzünk, sőt legtöbbször az üzenetet két-háromszor is át és visszaalakítjuk (transzformáljuk). Ha például az információ forrása az ember, az első átkódolás akkor zajlik le, amikor a gondolatainkat, amelyek az agynak nevezett információfeldolgozó és tároló berendezésben valamilyen formában el vannak raktározva, szabályos nyelvi formába öntjük. A második akkor, amikor beszédhanggá alakítjuk. Adott kommunikációs szituációban legtöbbször a kommunikációs láncnak csak egy szakaszát vizsgáljuk, s így teljes joggal beszélhetünk az illető szakaszra vonatkozó kódolásról.

4.1. Definíció. *A kódolás az az eljárás, amely egy nyelv véges ábécéjéből képzett szavakat kölcsönösen egyértelmű módon hozzárendeli egy másik nyelv meghatározott szavaihoz. A kódolással ellentétes eljárás a dekódolás.*

A csatornakapacitás egyik meghatározása: az az információmennyiség, amelyet egy adott csatornán optimális kódolás mellett az időegység alatt át lehet vinni. Shannon azt is megállapította, hogy alkalmas kódolási eljárással zaj jelenlétében is megvalósítható tetszőlegesen kis hibavalószínűségű információátvitel, ha az átvitel sebessége kisebb a csatorna kapacitásánál.

A kódolásnak az információátvitelben kettős célja van. Egyrészt az üzenetet a csatornán átvihető alakra kell hozni, másrészt ezt úgy kell végrehajtani, hogy az üzenet minél gazdaságosabban, minél rövidebb idő alatt és minél kevesebb veszteséggel jusson el a csatorna másik végébe. (Az adatfeldolgozásban a kódolásnak más céljai is vannak: az adatok tömörítése, titkosítása stb.).

4.2. Megjegyzés. *A kódolásnál elsőrendű követelmény a dekódolhatóság. Ha a vevő nem tudja az üzenetet, nem lehet szó kommunikációról. A megfejtés akkor lehetséges, ha az üzenet egyértelműen dekódolható. Ennek szükséges, de nem elégséges feltétele, hogy a különböző közleményekhez rendelt kódközlemények különbözőek legyenek.*

A legegyszerűbb kódolási eljárás a betűnkénti kódolás: a forrásközlemény minden betűjéhez hozzárendeljük az illető betű kódját. Hiába különböznek azonban az egyes betűk kódjai egymástól, az üzenet attól még nem lesz egyértelműen dekódolható. Ha ugyanis a jeleket egymás után írjuk, a jelsorozatot többféleképpen is felbonthatjuk. Ezen a bajon Morse úgy segített, hogy a betűk közé szünetet iktatott be. Az egyértelműséget azzal fizette meg, hogy hosszabbá tette az üzenetet. Baudot más megoldást választott: minden betűnek azonos hosszúságú kódjelet feleltetett meg. Így az üzenetet egyértelműen tagolni lehet, viszont ezzel a módszerrel is hosszabbá válik.

A változó kódhossz sokkal gazdaságosabb, mivel lehetőség van arra, hogy figyelembe vegyük a forrásábécé jeleinek gyakoriságát, s a gyakrabban előforduló jeleket rövidebb, a ritkábban előfordulókat hosszabb kódjelekkel kódoljuk. Ezt tette Morse is: az angol nyelv betűgyakorisága alapján állította össze ábécéjét. A gazdaságosságnak van még egy feltétele: az, hogy a betűk minden elválasztás nélkül egyértelműen dekódolhatók legyenek. Ez a feltétel csak akkor teljesül, ha úgynevezett prefix tulajdonságú kódot alkalmazunk.

A hírközlés matematikai modelljében szereplő résztvevők tulajdonságainak a leírására és a feladat megoldására használjuk a következő fogalmakat.

4.3. Definíció. *Legyen* $X = \{x_1, \ldots, x_n\}$, *amely véges halmaz tartalmazza a forrásábécé elemeit. Jelölje az* X*-ből készített véges sorozatok halmazát* $\mathcal{X}$. *Ennek elemeit nevezzük forrásüzeneteknek.*

4.4. Megjegyzés. *Tehát*

$$\mathcal{X} = \bigcup_{k=1}^{\infty} X^k.$$

Természetesen minden a forrás által kibocsátott minden elem tekinthető valószínűségi változónak, azaz a forrásüzenet az egy valószínűségi változó sorozat. Sztochasztikus tulajdonságainak leírásához meg kell adni a sorozat együttes eloszlását. Mint látni fogjuk, sokszor egyszerűsítjük ezt az általános esetet, hogy a véletlen leírása egyszerűbb legyen. A korábbi urnamodellünk is egy ilyen esetet ír le.

4.5. Definíció. *A forrást emlékezetnélkülinek nevezzük, ha a valószínűségi változó sorozat teljesen független.*

4.6. Definíció. *A forrást stacionáriusnak (stacionérnek) nevezzük, ha a valószínűségi változó sorozatban az eltolódással kapott véges dimenziós eloszlások megegyeznek, azaz* $(\xi_1, \xi_2, \ldots, \xi_m)$ *és* $(\xi_{k+1}, \xi_{k+2}, \ldots, \xi_{k+m})$ *véletlen vektorok együttes eloszlása minden* $m, k \in \mathbf{N}$ *esetén megegyezik.*

4.7. Definíció. *Legyen* $Y = \{y_1, \ldots, y_s\}$ $(s \geq 2)$, *amely halmaz tartalmazza a kódábécé vagy csatornaábécé elemeit. Jelölje az* Y*-ból készített véges sorozatok halmazát* $\mathcal{Y}$. *Ennek elemeit nevezzük kódüzeneteknek.*

4.8. Megjegyzés. *Tehát*

$$\mathcal{Y} = \bigcup_{k=1}^{\infty} Y^k.$$

A forrásüzenetet át kell alakítanunk olyan formára, hogy továbbítható legyen az ún. csatornán. Zajmentesnek nevezzük a csatornát, ha az üzenet továbbítás közben nem történik változás (hiba), ekkor általában nem szükséges további átalakítás. Viszont, ha a csatorna megváltoztathatja az üzenetet, akkor még történik egy átalakítás, hogy ez a változás jelezhető illetve javítható legyen. Ezt fogjuk csatornakódolásnak nevezni.

4.9. Definíció. *A kódolás az az eljárás, amely során a forrásüzenetekhez kódüzenetet rendelünk hozzá, azaz megadunk egy*

$$g : \mathcal{X} \to \mathcal{Y}$$

függvényt.

4.10. Definíció. *Betűnkénti kódolásnak nevezzük, ha*

$$g : X \to \mathcal{Y},$$

azaz a hozzárendelés a forrásüzenethez elemenként történik. A

$$g(x_1), g(x_2), \ldots, g(x_n)$$

a forrásábécé elemeihez rendelt kódszavak.

4.11. Definíció. *Blokkonkénti kódolásnak nevezzük, ha*

$$g : X^k \to \mathcal{Y},$$

ahol $k \in \mathbf{N}$ rögzített, azaz a hozzárendelés a forrásüzenethez blokkonként történik.

4.12. Megjegyzés. *Az X^k halmaz tekinthető forrásábécének, s így tekinthető betűnkénti kódolásnak is. Továbbá, ha*

$$g : X^k \to Y^l,$$

ahol $l \in \mathbf{N}$ szintén rögzített, akkor blokkhoz blokkot rendelünk, ekkor a blokkokhoz rendelt kódszavak hossza megegyezik, azaz állandó hosszúságú kódról beszélünk. Speciális eset a betűnkénti eset $(k = 1)$.

4.13. Megjegyzés. *A kódok készítésénél természetesen sok szempontot szokás figyelembe venni, amelyek közül a legfontosabb a dekódolhatóság. Mi itt csak az egyértelműen dekódolható esetekkel foglalkozunk. Ha a g függvény kölcsönösen egyértelmű, akkor a kód egyértelműen dekódolható. Először a zajmentes csatorna esettel foglalkozunk, feltételezve, hogy a csatornán a betűk (jelek) azonos költséggel mennek át. Nem azonos költségű esetnek tekinthető például a Morse-kód, merta jelek nem azonos idejűek, azaz ha összeadjuk az időket, akkor azonos darabszám esetén is lehet az üzenet hossza különböző (additív költségű eset).*

Tehát a következő esetekben az ún. zajmentes csatorna + betűnkénti kódolás + azonos költségű esetekkel foglalkozunk, azaz

$$g : X \to \mathcal{Y}.$$

4.14. Megjegyzés. *$g(x_i) = K_i$ az x_i betűhöz rendelt kódszó. Jelölje $\mathcal{K}$ a kódszavak halmazát.*

2. 4.2. Sardinas-Patterson módszer

Először egy olyan módszerrel foglalkozunk, amely segítségével egy kódról eldönthető, hogy egyértelműen dekódolható.

Legyen $\mathcal{K}$ tetszőleges kód, amelyben a kódszavak különbözőek és nem üresek. A K'' szó a K' szó után következik, ha $K'' \neq \emptyset$ és létezik $K_i \in \mathcal{K}$, hogy $K'K'' = K_i$ vagy $K' = K_iK''$.

A $\mathcal{K}$ kódhoz rekurzíve megkonstruáljuk az $S_m, m = 0, 1, 2, \ldots$ halmazokat. Legyen $S_0 = \mathcal{K}$. Az S_{m+1} halmazt az S_m halmaz szavai után következő szavak halmazaként definiáljuk.

4.15. Tétel. *A $\mathcal{X}$ kód akkor és csak akkor egyértelműen dekódolható, ha az S_i, $(i \geq 1)$ halmazok nem tartalmaznak kódszót, azaz S_0 valamelyik elemét.*

4.16. Megjegyzés. *A tétel bizonyításával nem foglalkozunk, mert absztrakt algebra nélkül a bizonyítás hosszadalmas. Egy ilyen megtalálható a [3] könyvben. Az absztrakt algebrai bizonyítás pedig a [4] könyvben.*

4.1. Példa.

S_0	S_1	S_2	S_3	S_4
10	1	0	01	$\emptyset$
101		01		
001				

Tehát egyértelműen dekódolható.

▲

4.2. Példa.

S_0	S_1	S_2	S_3	S_4
0	10	1	01	1
010				0
101				

Tehát nem egyértelműen dekódolható hiszen a $0 \in S_4$ kódszó.

▲

4.3. Példa.

S_0	S_1	S_2	S_3	S_4	S_5
a	d	eb	de	b	ad
c	bb	cde			$bcde$
ad					
abb					
bad					
deb					
$bbcde$					

Tehát nem egyértelműen dekódolható hiszen az $ad \in S_5$ kódszó.

▲

4.4. Példa.

	S_0	S_1	S_2	S_3	S_4	S_5	S_6
x_1	010	1	100	11	00	01	0
x_2	0001		1110		110	011	10
x_3	0110		01011			110	001
x_4	1100					0	110
x_5	00011						0011
x_6	00110						0110
x_7	11110						
x_8	101011						

Tehát nem egyértelműen dekódolható hiszen az $x_3 = 0110 \in S_6$ kódszó.

A 000110101111000110 kódüzenet például kétféleképpen is dekódolható:

$x_2 x_8 x_4 x_3$

és

$x_5 x_1 x_7 x_6$.

▲

3. 4.3. Keresési stratégiák és prefix kódok

Elképzelhető olyan keresési feladat, ahol egyszerre legfeljebb $s \geq 2$ csoportról tudjuk egyetlen kísérletre eldönteni, hogy a keresett elem melyikben van.

Absztrakt megfogalmazás: Legyen a keresett ξ dolog az $X = \{x_1, \ldots, x_n\}$ véges halmaz valamelyik eleme. A ξ-t valószínűségi változónak tekintjük: $p_i = P(\xi_i = x_i)$ a ξ eloszlása. A kesesési stratégiák definiálásához és áttekintéséhez felhasználjuk a gráfos leírásukat.

Fának nevezzük az olyan irányított gráfokat, melyek egy kitüntetett szögpontjából, a kezdőpontból, ágak (irányított utak) indulnak ki, melyek a későbbi szögpontokban ismét elágaznak, de újra biztosan nem találkozhatnak. Azokat a szögpontokat, melyekből további élek már nem indulnak ki, végpontoknak nevezzük. Mivel az ágak újra nem találkozhatnak, a kezdőpontot mindegyik végponttal pontosan egy ág köti össze. Az ágat alkotó élek számát az ág hosszának nevezzük.

Tekintsünk egy n végpontú fát és rendeljük hozzá kölcsönösen egyértelmű módon a fa végpontjaihoz (ágaihoz) az n elemű X halmaz elemeit. Az ilyen módon cimkézett végpontú fát az X halmaz kódfájának nevezzük. Ha a kódfa minden szögpontjához az X halmaznak azt az A részhalmazát rendeljük hozzá, amely a szögponton áthaladó ágak végpontjaihoz tartozó elemekből áll, akkor olyan megfeleltetést kapunk a szögpontok és az X bizonyos részhalmazai között, hogy bármelyik szögponthoz rendelt halmaz a szögpontból kiinduló élek végpontjaihoz tartozó, páronként diszjunkt halmazok egyesítése. Látható, hogy az X halmaz olyan kódfája, ahol minden szögpontból legfeljebb s él indul ki egy s alternatívás keresési stratégiát definiál. Ez a megfeleltetés kölcsönösen egyértelmű.

Ha $\xi = x$, akkor a megtaláláshoz szükséges lépések száma az x végpontba vezető ág $L(x)$ hossza. A feladat az

$$L = \sum_{i=1}^{n} L(x_i)P(\xi = x_i)$$

várható lépésszám minimalizálása. Mivel egy lépéssel legfeljebb $\log_2 s$ mennyiségű információt nyerhetünk és a hiányzó információ $H(\xi)$, ezért várhatóan L nagyobb lesz, mint

$$\frac{H(\xi)}{\log_2 s}.$$

4.17. Tétel. *Az s alternatívás keresési stratégiára*

$$L = \sum_{i=1}^{n} L(x_i)P(\xi = x_i) \geq \frac{H(\xi)}{\log_2 s}.$$

Bizonyítás. Tekintsünk egy tetszőleges kódfát, és legyen A a kódfa olyan szögpontja, amely nem végpont. Jelölje $B_1, B_2, \ldots, B_j\,(j \leq s)$ az A-ból kiinduló élek végpontjait. A megfelelő halmazokat is jelöljük ugyanígy. Legyen

$$P(A) = \sum_{x \in A} P(\xi = x),$$

ekkor a $p_i = P(\xi = x_i)$ valószínűséget az x_i végponthoz vezető ág minden, a végponttól különböző szögpontjához odaírva, és áganként összegezve közvetlenül kapjuk, hogy

$$L = \sum P(A),$$

ahol az összegzést a végpontoktól különböző szögpontokra kell elvégezni.

Mivel $P(A)$ éppen annak a valószínűsége, hogy a keresés során eljutunk az A szögpontba, az A szögpontban végzendő kísérlet $B_1, B_2, \ldots, B_j$ kimeneteinek valószínűsége rendre $P(B_1|A), P(B_2|A), \ldots, P(B_j|A)$, ahol

$$P(B_i|A) = \frac{P(B_i)}{P(A)}.$$

Ennek a kísérletnek az entrópiája tehát

$$H_A = -\sum_{i=1}^{j} \frac{P(B_i)}{P(A)} \log_2 \frac{P(B_i)}{P(A)} =$$
$$= -\frac{1}{P(A)} \left(P(B_i)\log_2 P(B_i) - P(A)\log_2 P(A)\right).$$

Számoljuk ki a

$$\sum P(A)H_A$$

mennyiséget, ahol az összegzést a kódfa végponttól különböző szögpontjaira kell elvégezni. A felbontás azt mutatja, hogy ebben az összegben, a kezdőpont és a végpont kivételével minden szögponthoz a

$$P(C)\log_2 P(C)$$

kifejezés egyszer pozitív, egyszer negatív előjelű, mert C egyszer A típusú egyszer pedig B típusú.

Tehát az összegzés

$$\sum P(A)H_A = -\sum_{i=1}^{n} p_i \log_2 p_i + P(X)\log_2 P(X) = H(\xi).$$

Viszont az entrópia tulajdonságai alapján

$H_A \leq \log_2 s.$

Tehát

$$H(\xi) = \sum P(A) H_A \leq \log_2 s \sum P(A) = L \log_2 s,$$

amelyből adódik az állítás.

■

4.18. Megjegyzés. *Jól látható, hogy az alsó korlátot akkor közelítjük meg, ha minden lépésben az egyenletes elszláshoz közel eső s alternatívás lépést alkalmazunk.*

Jelölés: A g kódolás esetén $||g(x_i)|| = L_i$ a $g(x_i)$ kódszó hossza.

4.19. Definíció. *A $\mathcal{K}$ kód prefix, ha a kódszavak mind különbözőek és egyik kódszó sem folytatása a másiknak.*

4.20. Megjegyzés. *Az állandó kódhosszú kód mindig prefix, ha a kódszavai különbözőek. A prefix kódhoz kódfa rendelhető, s így közvetlen kapcsolatban van a keresési stratégiákkal.*

4.21. Tétel. *Minden prefix kód egyértelműen dekódolható.*

Bizonyítás. A $\mathcal{K}$ kód prefix, azaz a kódszavak mind különbözőek és egyik kódszó sem folytatása a másiknak. Tételezzük fel, hogy létezik egy üzenet, amely kétféleképpen dekódolható. Az ilyen üzenetek között van legrövidebb, s ekkor feltételezve, hogy létezik két különböző kódszavakra bontás, az első kódszavaknak rögtön különbözőknek kell lenniük. Viszont ekkor az egyik folytatása a másiknak (egyforma hosszúak nem lehetnek). Ez ellentmond annak, hogy a kód prefix.

■

4.22. Megjegyzés. *A Sardinas-Patterson módszer alkalmazása esetén prefix kódra $S_1 = \emptyset$.*

4.23. Lemma. *A kódfák és a prefix kódok között kölcsönös egy-egyértelmű megfeleltetés van.*

4.24. Megjegyzés. *A prefix kód átlagos kódhossza nem lehet kisebb, mint*

$$\frac{H(\xi)}{\log_2 s}.$$

4.25. Tétel. (Kraft-Fano egyenlőtlenség) *Ha $\mathcal{K} = \{K_1, \ldots, K_n\}$ s számú kódjelből készített prefix kód, akkor*

$$\sum_{i=1}^{n} s^{-L_i} \leq 1,$$

ahol L_i a K_i kódszó hossza.

Bizonyítás. Legyen $m = \max_{1 \leq i \leq n} L_i$. Minden kódszót egészítsünk ki m hosszúvá. L_i kiegészítése s^{m-L_i}-féleképpen lehetséges. Tehát

$$\sum_{i=1}^{n} s^{m-L_i} \leq s^m,$$

amelyből adódik az állítás.

■

4.26. Tétel. (Kraft-Fano egyenlőtlenség megfordítása) *Ha az*

$L_1, \ldots, L_n$

természetes számok eleget tesznek a

$$\sum_{i=1}^{n} s^{-L_i} \leq 1$$

egyenlőtlenségnek, ahol $s \geq 2$ *természetes szám, akkor létezik* s *kódjelből alkotott* $\mathcal{K} = \{K_1, \ldots, K_n\}$ *prefix kód, melynél a* K_i *kódszó hossza éppen* L_i.

Bizonyítás. Legyen $m = \max_{1 \leq i \leq n} L_i$, ekkor

$$\sum_{i=1}^{n} s^{m-L_i} \leq s^m.$$

Legyen a K^* teljes kódfa, amelyben minden ág m hosszú. Válasszunk ki egy ágat, amelyből $m - L_1$ élet elhagyunk stb. Teljes indukcióval adódik az állítás.

■

4. 4.4. Shannon-Fano kód

Bár Shannon nevét általában az információmennyiség meghatározásával kapcsolatban szokták legtöbbször emlegetni, információelméleti munkáiban a kódolás elvi kérdéseit is tisztázta, sőt eljárást is kidolgozott az optimális kódolásra. A shannoni tétel kimondja, hogy valamely meghatározott kódábécé esetén egy és csakis egy olyan ábrázolási mód van, amely adott mennyiségű információt a lehető legkevesebb jellel fejez ki. Ez az optimális kód. Ha az üzenetet több jellel fejezzük ki, redundánssá válik. Ez történik például, amikor egy olyan ábécét kell bináris kódra átírnunk, amelyben a betűk száma nem kettőnek egész számú hatványa. Egy 26 betűs ábécét csak 5 bináris számjeggyel írhatunk át. A látszólagos információmennyiség tehát 5 bit, holott egy betűhöz csak 4.65 bit tényleges információmennyiség tartozik. A parazita információk arányát csökkenthetjük, ha a betűket nem egyenként, hanem blokkonként, kettesével, hármasával stb. kódoljuk. Ekkor azonban a kód egyre bonyolultabbá válik és nő a kódolás költsége.

4.27. Tétel. *Létezik prefix kód, hogy*

$$L < \frac{H(\mathcal{P})}{\log_2 s} + 1.$$

Bizonyítás. A bizonyítás konstruktív és az elkészített kódot Shannon-Fano kódnak nevezzük.

Most pedig nézzük, hogy az algoritmus milyen lépésekből áll.

Legyen

$$\mathcal{P} = \{p_1, p_2, \ldots, p_n\}, p_i > 0, (i = 1, \ldots, n), \sum_{i=1}^{n} p_i = 1$$

tetszőleges forráseloszlás.

Ekkor a lépések a következők:

1. Rendezzük a valószínűségeket csökkenő sorrendbe:

   $$p_1^* \geq p_2^* \geq \cdots \geq p_n^* > 0.$$

   a. Képezzük az $x_i^* \, (i = 1, \ldots, n)$ értékeket a következőképpen:

   $$x_1^* = 0, x_2^* = p_1^*, x_3^* = p_1^* + p_2^*, \ldots, x_n^* = p_1^* + \cdots + p_{n-1}^*.$$

   b. Ábrázoljuk ezen értékeket a $[0, 1)$ intervallumon és osszuk fel a $[0, 1)$ intervallumot s egyenlő részre (s a kódábécé elemeinek száma).

c. Azokat az x_i^* intervallumokat, melyek egynél több x_i értéket tartalmaznak osszuk fel újra egészen addig míg mindegyik más intervallumba nem kerül.

d. A $g(x_i^*)$ kódszó az $s^{-1}, s^{-2}, \ldots, s^{-k}$ hosszúságú intervallumok megfelelő sorszámából áll, amelyekben x_i^* benne van, ahol k a kódszó hossza, illetve az osztáslépések száma.

A konstrukcióból látszik, hogy prefix kódot kapunk.

Megmutatjuk, hogy

$p_i^* < s^{-L_i+1}$,

ahol $L_i = ||g(x_i^*)||$.

Az x_i^* értéket tartalmazó utolsó előtti, s^{-L_i+1} hosszúságú intervallumban legalább még egy pont van, azaz az x_{i-1}^* és az x_{i+1}^* közül legalább az egyik. Mivel a $p_{i-1}^* \geq p_i^*$, így mindenképpen igaz, hogy

$p_i^* < s^{-L_i+1}$.

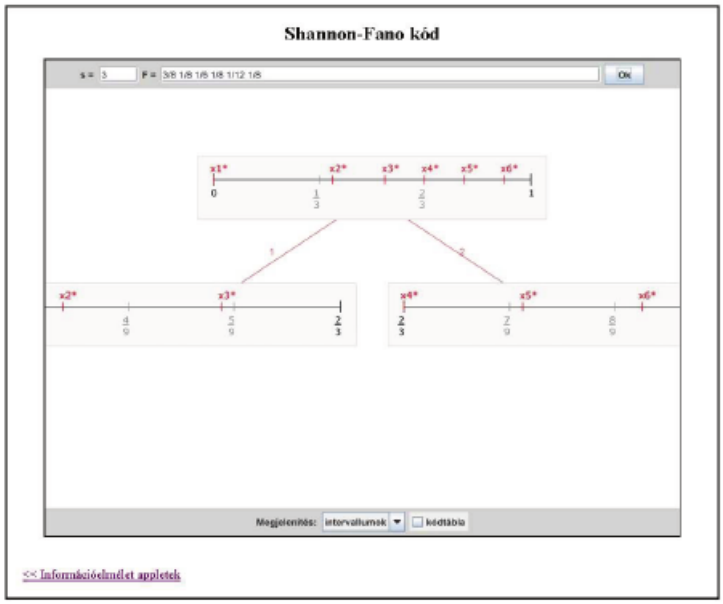
Képezzük mindkét oldal logaritmusát, majd negatívját és összegezzük minden i-re, akkor kapjuk, hogy

$$H(\mathcal{P}) = -\sum_{i=1}^{n} p_i \log_2 p_i > \log_2 s \sum_{i=1}^{n} p_i (L_i - 1) = (L-1)\log_2 s.$$

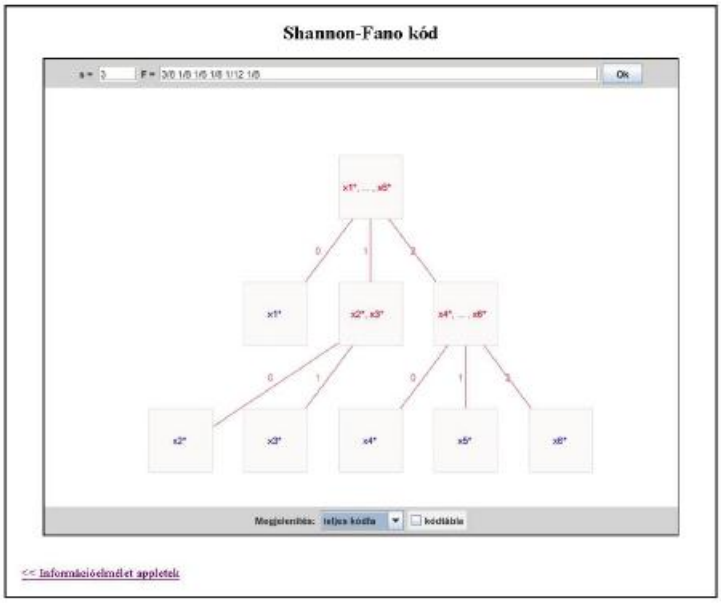
Ezzel az állítást bizonyítottuk.

■

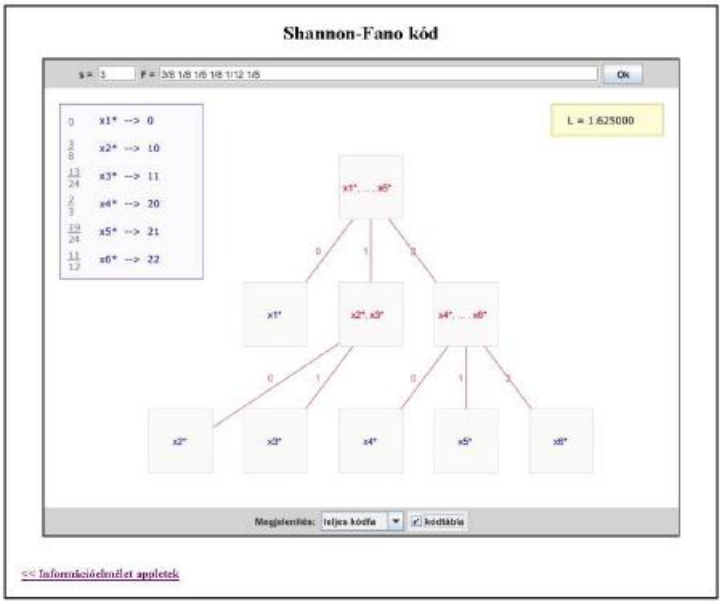
4.2. ábra - Példa a Shannon-Fano kódolásra (intervallumfelosztás)



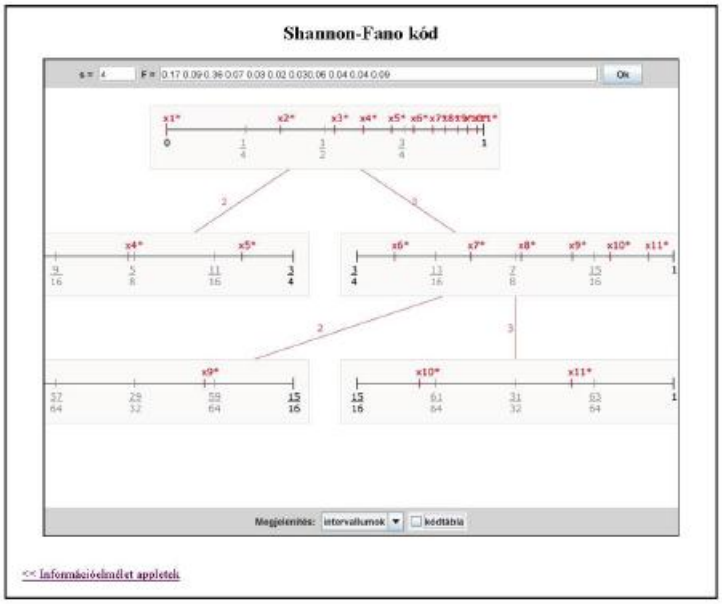
4.3. ábra - Példa a Shannon-Fano kódolásra (kódfa)



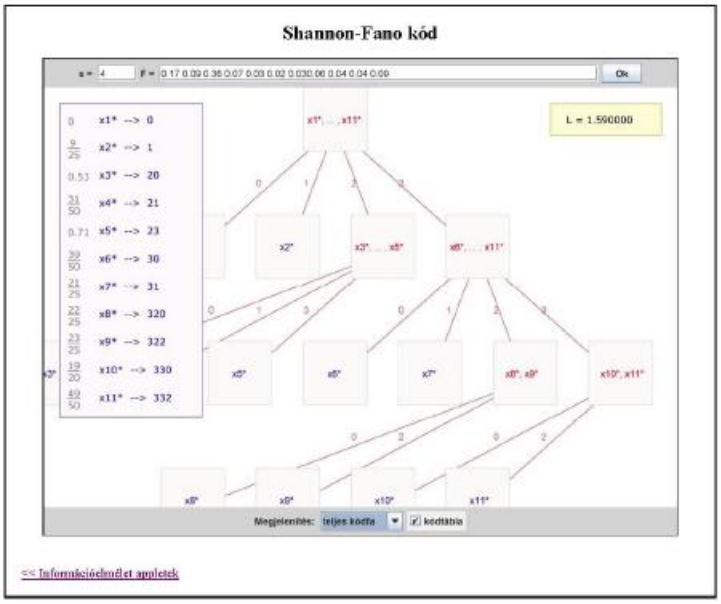
4.4. ábra - Példa a Shannon-Fano kódolásra (kód)



4.5. ábra - Példa a Shannon-Fano kódolásra (intervallumfelosztás)



4.6. ábra - Példa a Shannon-Fano kódolásra (kód)



5. 4.5. Gilbert-Moore kód

Nagyméretű forrásábécé esetén a sorbarendezés költsége magas lehet. Erre ad megoldást a következő kód, amelynél nincs szükség sorbarendezésre.

4.28. Tétel. *Létezik prefix kód, hogy*

$$L < \frac{H(\mathcal{P}) + 1}{\log_2 s} + 1.$$

Bizonyítás. A bizonyítás konstruktív és az elkészített kódot Gilbert-Moore kódnak nevezzük.

Most pedig nézzük, hogy az algoritmus milyen lépésekből áll.

Legyen

$$\mathcal{P} = \{p_1, p_2, \ldots, p_n\}, p_i > 0, (i = 1, \ldots, n), \sum_{i=1}^{n} p_i = 1$$

tetszőleges forráseloszlás.

Ekkor a lépések a következők:

1. Képezzük az $x_i^* (i = 1, \ldots, n)$ értékeket a következőképpen:

   $$x_1^* = \frac{p_1}{2}, x_2^* = p_1 + \frac{p_2}{2}, x_3^* = p_1 + p_2 + \frac{p_3}{2}, \ldots, x_n^* = p_1 + \cdots + p_{n-1} + \frac{p_n}{2}.$$

   a. Ábrázoljuk ezen értékeket a $[0, 1)$ intervallumon és osszuk fel a $[0, 1)$ intervallumot s egyenlő részre (s a kódábécé elemeinek száma).

   b. Azokat az intervallumokat, melyek egynél több x_i értéket tartalmaznak osszuk fel újra egészen addig míg mindegyik x_i^* más intervallumba nem kerül.

   c. A $g(x_i^*)$ kódszó az $s^{-1}, s^{-2}, \ldots, s^{-k}$ hosszúságú intervallumok megfelelő sorszámából áll, amelyekben x_i^* benne van, ahol k a kódszó hossza, illetve az osztáslépések száma.

A konstrukcióból látszik, hogy prefix kódot kapunk.

Megmutatjuk, hogy

$$\frac{p_i}{2} < s^{-L_i+1},$$

ahol $L_i = ||g(x_i^*)||$.

Az x_i^* értéket tartalmazó utolsó előtti, s^{-L_i+1} hosszúságú intervallumban legalább még egy pont van, azaz az x_{i-1}^* és az x_{i+1}^* közül legalább az egyik. Tehát mindenképpen igaz, hogy

$$\frac{p_i}{2} < s^{-L_i+1}.$$

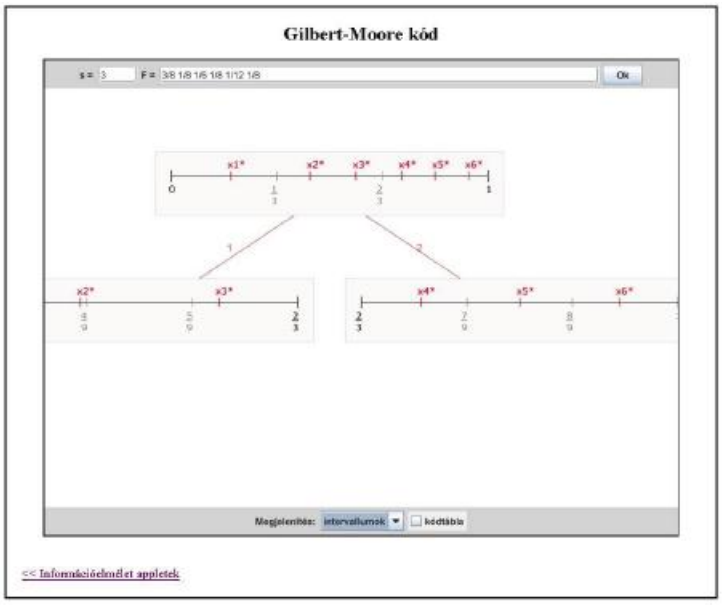
Képezzük mindkét oldal logaritmusát, majd negatívját és összegezzük minden i-re, akkor kapjuk, hogy

$$H(\mathcal{P}) = -\sum_{i=1}^{n} p_i \log_2 p_i > \log_2 s \sum_{i=1}^{n} p_i((L_i - 1) - 1) = (L - 1)\log_2 s - 1.$$

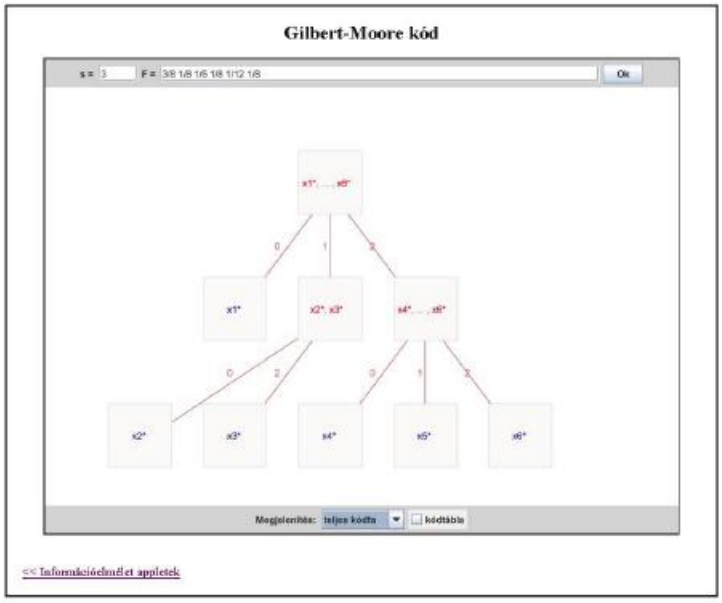
Ezzel az állítást bizonyítottuk.

■

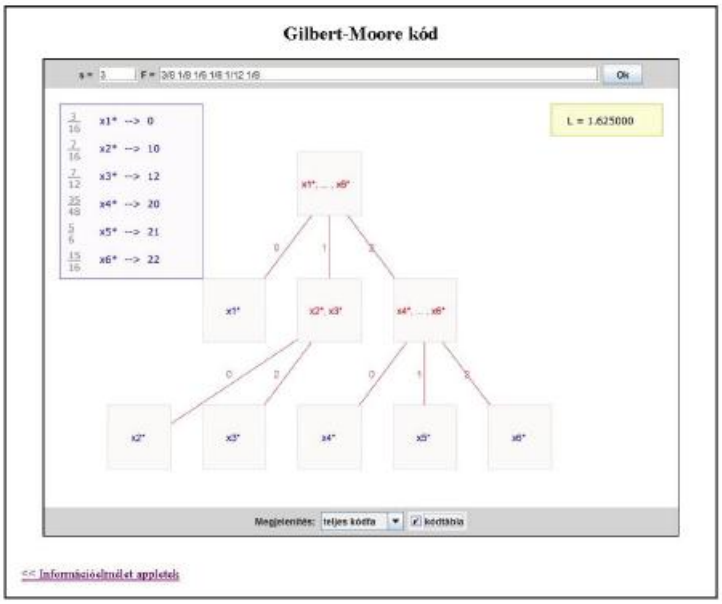
4.7. ábra - Példa a Gilbert-Moore kódolásra (intervallumfelosztás)



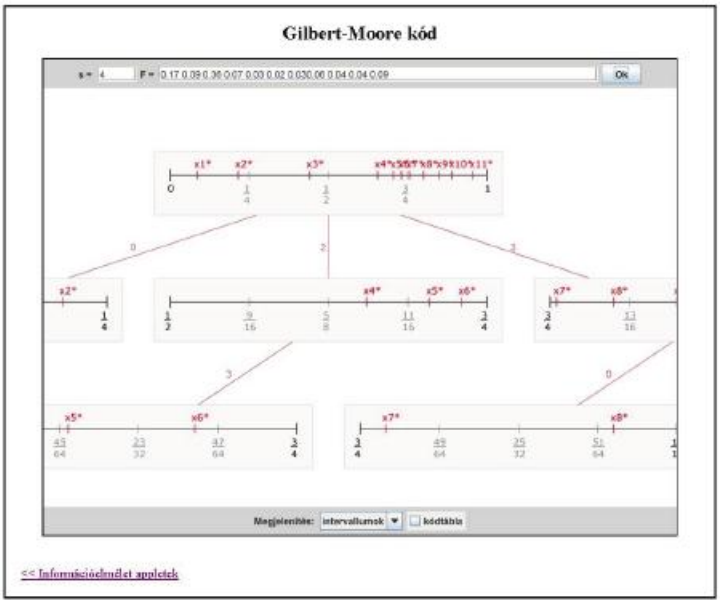
4.8. ábra - Példa a Gilbert-Moore kódolásra (kódfa)



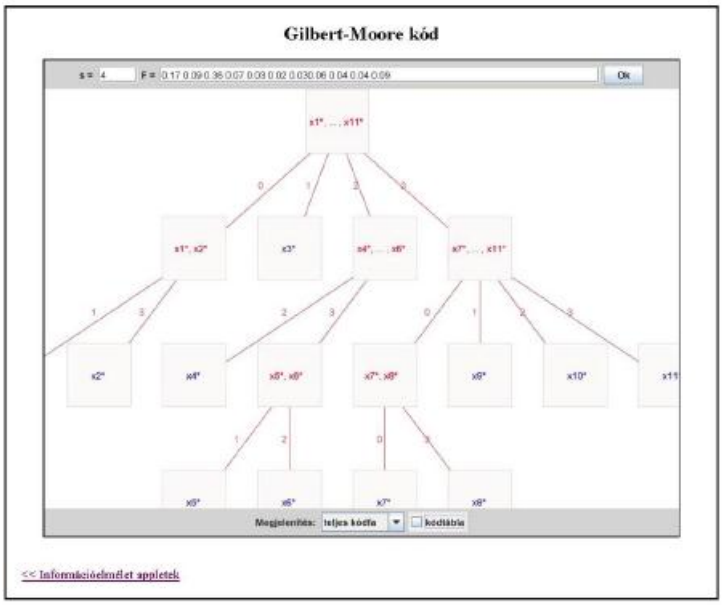
4.9. ábra - Példa a Gilbert-Moore kódolásra (kód)



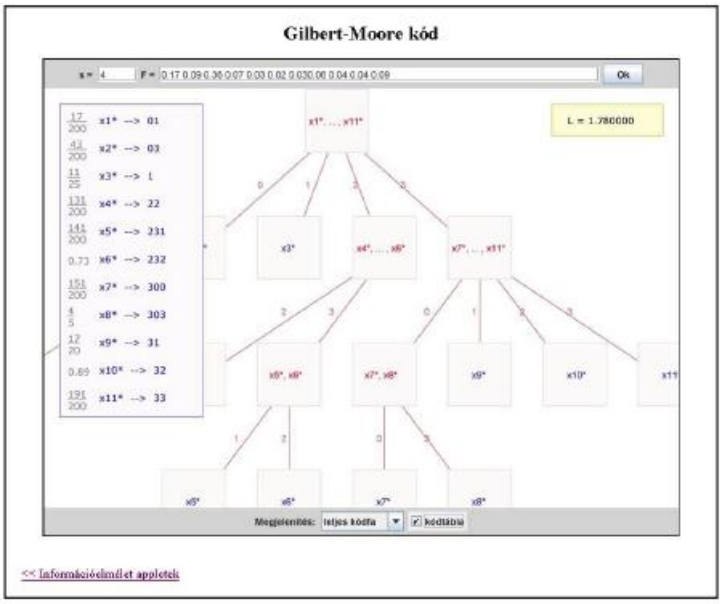
4.10. ábra - Példa a Gilbert-Moore kódolásra (intervallumfelosztás)



4.11. ábra - Példa a Gilbert-Moore kódolásra (kódfa)



4.12. ábra - Példa a Gilbert-Moore kódolásra (kód)



6. 4.6. Hatásfok

4.29. Definíció. *Az egyértelműen dekódolható kód hatásfoka:*

$$\gamma = \frac{H(\mathcal{P})}{L \log_2 s}.$$

4.34*A kódot optimálisnak nevezzük, ha egyértelműen dekódolható és maximális hatásfokú.*

4.30. Tétel. *Adott $\mathcal{P}$ eloszlású forrásábécé s számú kódjelből alkotott egyértelműen dekódolható kódjai között mindig van optimális.*

7. 4.7. Huffman-kód

Huffman-kód – maximális hatásfokú prefix kód.

Tulajdonságok:

1. Monotonitás. Ha $p_1 \geq p_2 \geq \cdots \geq p_n$, akkor $L_1 \leq L_2 \leq \cdots \leq L_n$.

2. A kódfa teljessége. Legyen $m = L_n$, ekkor minden $m-1$ hosszúságú kódjelsorozat ki van használva a kódolásnál, azaz maga is kódszó vagy egy rövidebb kódszó kiegészítéséből adódik, vagy pedig az egyik kódjel hozzáírásával valamelyik m hosszúságú kódszót kapjuk belőle. Ha volna kihasználatlan ág, akkor azt választva ismét prefix kódot kapnánk, melynek viszont kisebb az átlagos kódhossza.

4.31. Megjegyzés. *Optimális, bináris kódfa teljes.*

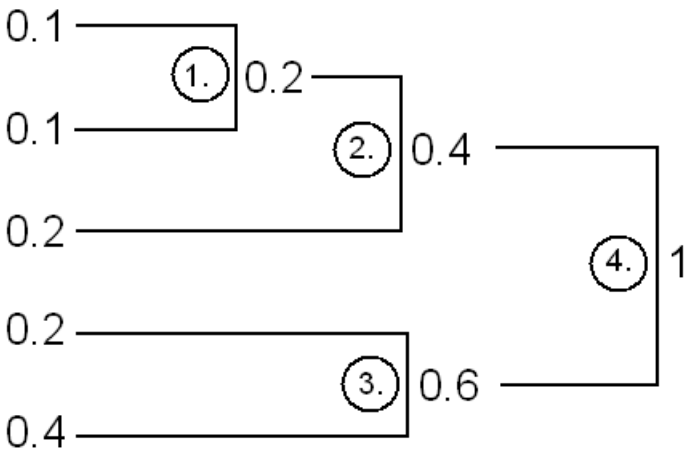
3. $L_n = L_{n-1}$ és az utolsó kódjelüktől eltekintve azonosak.

4.32. Megjegyzés. *Összevonási algoritmus. Az optimális kódfa minden végponttól különböző szögpontjából s él indul ki, kivéve esetleg egy végpont előtti szögpontot, amelyből r él megy tovább, ahol $2 \leq r \leq s$. Ekkor*

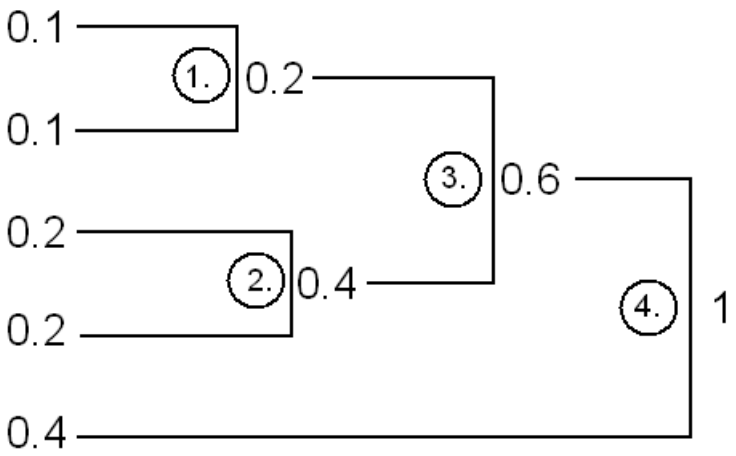
$$n = k(s-1) + r.$$

A teljes kódfánál $r = s$. Tehát az első összevonási lépésben az r legkevésbé valószínű elemet kell összevonni, míg az összes többiben az s legkevésbé valószínűt.

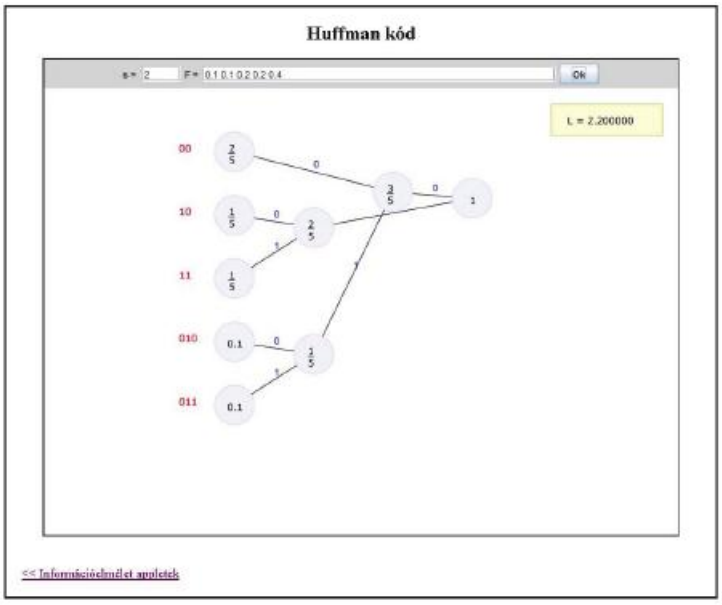
4.13. ábra - Példa Huffman-féle kódolásra 1. változat



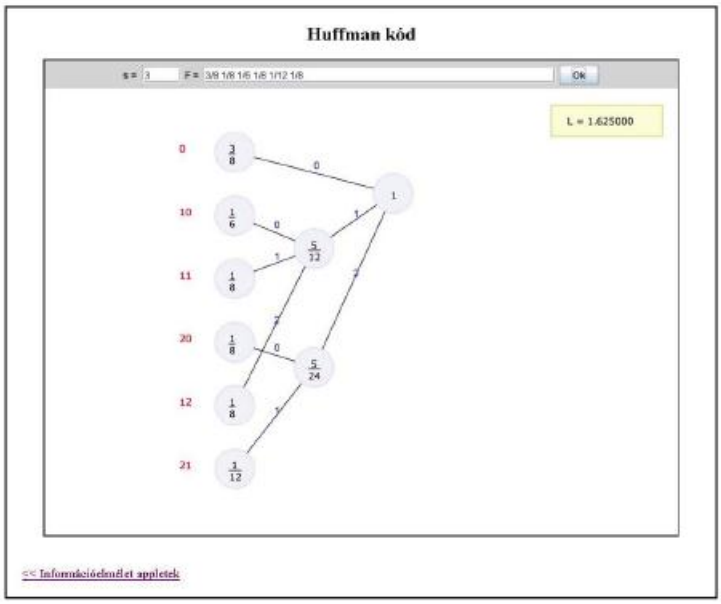
4.14. ábra - Példa Huffman-féle kódolásra 2. változat



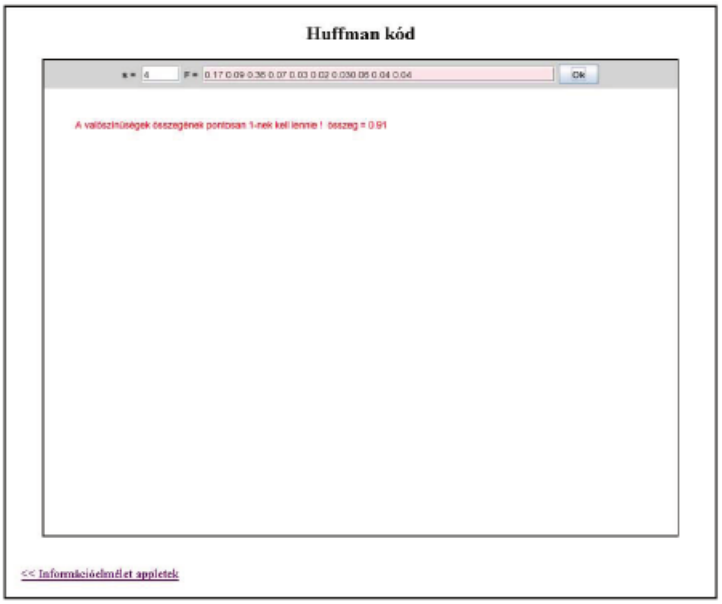
4.15. ábra - Példa Huffman-féle kódolásra 3. változat



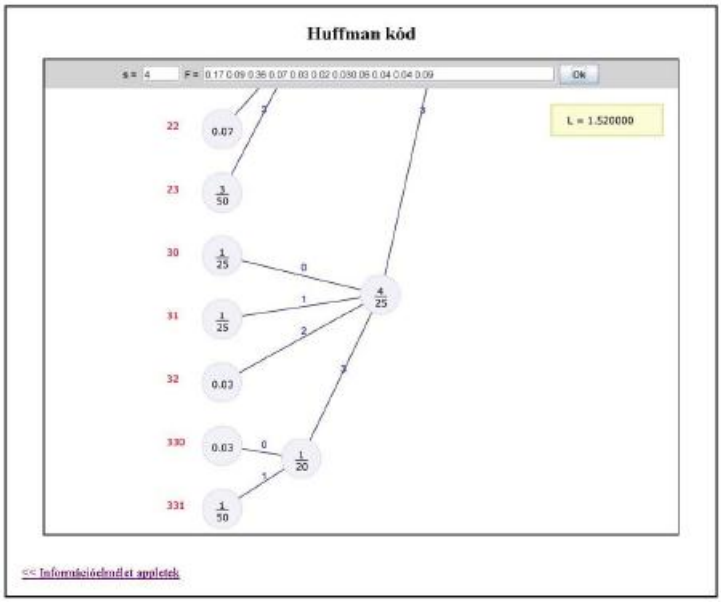
4.16. ábra - Példa a Huffman kódolásra



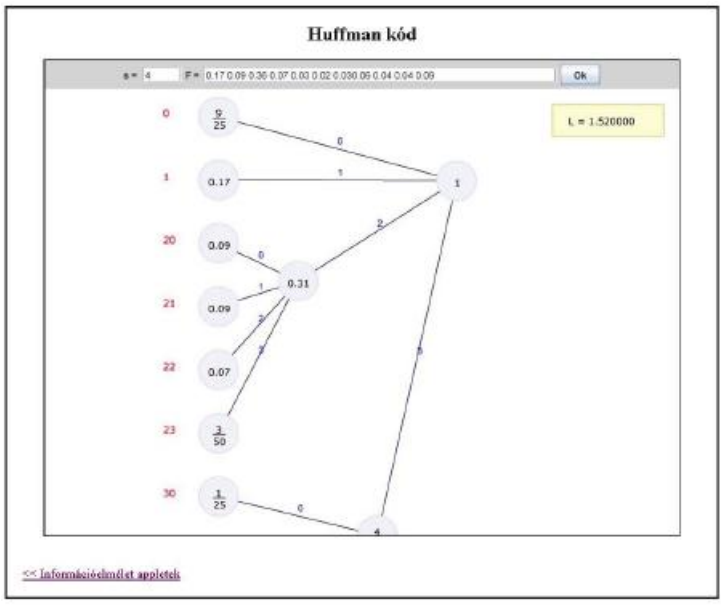
4.17. ábra - Példa a Huffman kódolásnál az eloszlás ellenőrzésére



4.18. ábra - Példa a Huffman kódolásra 1. rész



4.19. ábra - Példa a Huffman kódolásra 2. rész



8. 4.8. McMillan-dekódolási tétel

4.33. Tétel. (McMillan-dekódolási tétel) *Ha* $g : X \to \mathcal{Y}$ *egyértelműen dekódolható, akkor*

$$\sum_{i=1}^{n} s^{-||g(x_i)||} \leq 1.$$

Bizonyítás. Jelölje $W(N, L)$ azon N hosszúságú közleményeknek a számát, melyek kódközleményének a hossza éppen L.

$$m := \max_{1 \leq i \leq n} L_i.$$

A bizonyítás lépései (vázlat):

1. A kód egyértelműen dekódolható.

$W(N,L) \leq s^L$, azaz $W(N,L)s^{-L} \leq 1$.

Tehát

$$\sum_{L=1}^{mN} W(N,L)s^{-L} \leq mN.$$

2. Teljes indukcióval bizonyítjuk, hogy

$$\sum_{L=1}^{mN} W(N,L)s^{-L} = \left(\sum_{i=1}^{n} s^{-L_i}\right)^N.$$

3. Ha c és m adott pozitív számok, akkor az

$$(1+c)^N \leq mN$$

egyenlőtlenség nem teljesülhet minden N természetes számra, így

$$\sum_{i=1}^{n} s^{-L_i} \leq 1.$$

■

4.34. Tétel. *Egyértelműen dekódolható kód esetén*

$$L \geq \frac{H(\mathcal{P})}{\log_2 s}.$$

Bizonyítás. Legyen

$$q_i := \frac{s^{-L_i}}{\sum_{j=1}^{n} s^{-L_j}}.$$

Ekkor

$$0 \geq -D(\mathcal{Q}||\mathcal{P}).$$

■

4.35. Tétel. *Bármely egyértelműen dekódolható kód helyettesíthető egy másik ugyanolyan kódhosszúságú kóddal, amely viszont már prefix kód.*

Bizonyítás. A McMillan dekódolási tétel szerint egyértelműen dekódolható kódra teljesül a Kraft-Fano egyenlőtlenség. A Kraft-Fano megfordítása szerint viszont létezik olyan prefix kód amelyiknek pontosan ezek a kódhosszai.

■

4.36. Megjegyzés. *Az előző tétel szerint az optimális egyértelműen dekódolható kódhoz létezik ugyanilyen prefix, így az optimális prefix optimális az egyértelműen dekódolható kódok között is.*

9. 4.9. Blokkos kódolás, tömörítés, stacionér forrás entrópiája

Blokkos kódolás, azaz

$g : X^k \to \mathcal{Y}$,

esetén jelölje $\mathcal{P}^{(k)}$ az együttes eloszlást és $L^{(k)}$ az átlagos kódhosszot.

Az egy betűre jutó átlagos kódhossz pedig legyen

$$L = \frac{L^{(k)}}{k}.$$

Az optimális kódra:

$$\frac{H(\mathcal{P}^{(k)})}{\log_2 s} \leq L^{(k)} \leq \frac{H(\mathcal{P}^{(k)})}{\log_2 s} + 1.$$

Emlékezetnélküli, stacionárius forrás esetén a függetlenségből

$$H(\mathcal{P}^{(k)}) = kH(\mathcal{P}),$$

s így

$$\frac{H(\mathcal{P})}{\log_2 s} \leq L \leq \frac{H(\mathcal{P})}{\log_2 s} + \frac{1}{k}.$$

A következőkben egy stacionér forrás entrópiájával foglalkozunk, mert ennek segítségével tudjuk megadni a korlátokat általános esetben.

4.37. Tétel. $H(\xi|\eta) \leq H(\xi|f(\eta))$.

Bizonyítás. z jelölje $f(\eta)$ egy lehetséges értékét, és legyen

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

Ekkor p_y és q_y is egy eloszlást ad, ha y felveszi azokat az értékeket, amelyre $f(y) = z$, azaz $y \in f^{-1}(z)$. Az I-divergencia tulajdonságai alapján:

$$\sum_{f(y)=z} p_y \log_2 \frac{p_y}{q_y} \geq 0,$$

azaz

$$\sum_{f(y)=z} \frac{P(\xi = x, \eta = y)}{P(\xi = x, f(\eta) = z)} \log_2 \frac{P(\xi = x, \eta = y)P(f(\eta) = z)}{P(\xi = x, f(\eta) = z)P(\eta = y)} \geq 0.$$

Szorozzuk be a $P(\xi = x, f(\eta) = z)$ közös nevezővel és bontsuk fel a logaritmust a következőképpen:

$$\sum_{f(y)=z} P(\xi = x, \eta = y) \log_2 \frac{P(\xi = x, \eta = y)}{P(\eta = y)} +$$

$$+ \sum_{f(y)=z} P(\xi = x, \eta = y) \log_2 \frac{P(f(\eta) = z)}{P(\xi = x, f(\eta) = z)} \geq 0.$$

$$\sum_{f(y)=z} P(\xi = x, \eta = y) \log_2 \frac{1}{P(\xi = x|f(\eta) = z)} \geq$$

$$\geq \sum_{f(y)=z} P(\xi = x, \eta = y) \log_2 \frac{1}{P(\xi = x|\eta = y)}.$$

$$P(\xi = x, f(\eta) = z) \log_2 \frac{1}{P(\xi = x|f(\eta) = z)} \geq$$

$$\geq \sum_{f(y)=z} P(\xi = x, \eta = y) \log_2 \frac{1}{P(\xi = x|\eta = y)}.$$

Ez minden $\xi = x$ és $f(y) = z$ esetén teljesül. Végezzül el a $\sum_x \sum_z$ összegzést ezekre az egyenlőtlenségekre. Mivel

$$H(\xi|\eta) = \sum_y P(\eta = y) \sum_x P(\xi = x|\eta = y) \log_2 \frac{1}{P(\xi = x|\eta = y)},$$

így igazoltuk az állítást.

■

4.38. Tétel. *Stacionér forrás esetén a*

$$\lim_{k\to\infty} \frac{H(\mathcal{P}^{(k)})}{k}$$

határérték létezik.

Jele: $H^\star$

Bizonyítás. Tudjuk, hogy a forrás stacionér és

$$H(\xi, \eta) = H(\eta) + H(\xi|\eta),$$

ezért

$$\begin{aligned} H(\mathcal{P}^{(k)}) &= H(\xi_1, \ldots, \xi_k) = H(\xi_2, \ldots, \xi_k) + H(\xi_1|\xi_2, \ldots, \xi_k) = \\ &= H(\xi_k) + H(\xi_{k-1}|\xi_k) + \cdots + H(\xi_1|\xi_2, \ldots, \xi_k) = \\ &= H(\xi_1) + H(\xi_1|\xi_2) + \cdots + H(\xi_1|\xi_2, \ldots, \xi_k) = \\ &= H(\xi_1) + \sum_{i=2}^{k} H(\xi_1|\xi_2, \ldots, \xi_i). \end{aligned}$$

A $(\xi_2, \ldots, \xi_i)$ véletlen vektor függvénye a $(\xi_2, \ldots, \xi_{i+1})$ véletlen vektornak, ezért

$$H(\xi_1) \geq H(\xi_1|\xi_2) \geq \cdots \geq H(\xi_1|\xi_2, \ldots, \xi_k).$$

$$H(\mathcal{P}^{(k)}) = H(\xi_1) + \sum_{i=2}^{k} H(\xi_1|\xi_2, \ldots, \xi_i) \geq kH(\xi_1|\xi_2, \ldots, \xi_{k+1}).$$

Tehát

$$H(\mathcal{P}^{(k+1)}) = H(\mathcal{P}^{(k)}) + H(\xi_1|\xi_2, \ldots, \xi_{k+1}) \leq \frac{k+1}{k} H(\mathcal{P}^{(k)}).$$

Ekkor

$$\frac{H(\mathcal{P}^{(k+1)})}{k+1} \leq \frac{H(\mathcal{P}^{(k)})}{k}.$$

Tehát a sorozat monoton csökkenő és alulról korlátos, s így létezik a határérték.

■

4.39. Definíció. *A $H^\star$ mennyiséget a forrás átlagos entrópiájának nevezzük.*

4.40. Megjegyzés. *Emlékezetnélküli esetben $H^\star = H(\mathcal{P})$, egyébként $H^\star \leq H(\mathcal{P})$.*

Tekintsünk egy csatornát, amelybe bemennek a kódjeleink (jelölje általánosan ξ) és kijönnek a jelek (jelölje általánosan η). Kérdés mennyi információmennyiség érkezett meg az elküldöttből, azaz az η mennyit mond el a ξ-ről. Ez nyilván a kölcsönös információmennyiség. Ezután a csatornakapacitás (emlékezetnélküli eset):

$C = \max_{\varrho} I(\xi, \eta).$

4.41. Megjegyzés. *Mivel az eloszlások korlátos, zárt halmazt alkotnak, így a maximum létezik. Legyen C a kapacitás, L az átlagos kódhossz. Ha $H(\mathcal{P}) \leq LC$, akkor továbbíthatjuk a forrás által szolgáltatott közleményeket.*

4.42. Megjegyzés. *Zajmentes és emlékezetnélküli csatorná esetén $\xi = \eta$, ezért*

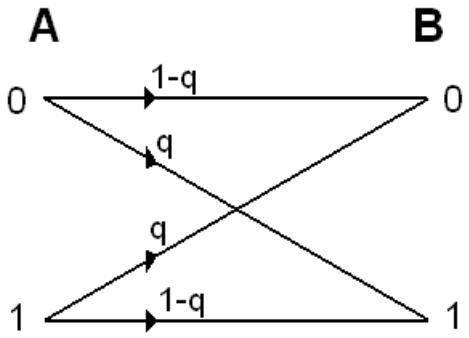
$C = \log_2 s.$

4.43. Megjegyzés. *Bináris szimmetrikus csatorna:*

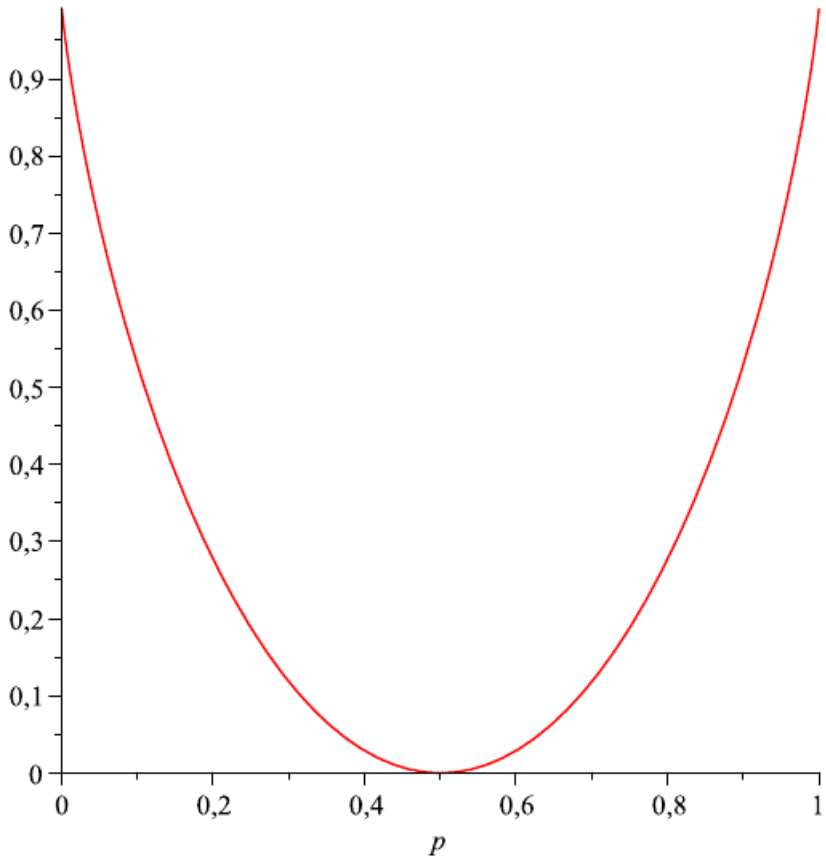
$C = 1 - H(p, q).$

Milyenek a bemeneti illetve kimeneti eloszlások?

4.20. ábra - Bináris szimmetrikus csatorna



4.21. ábra - Bináris szimmetrikus csatorna kapacitása a valószínűség függvényében



4.44. Tétel. (A zajmentes hírközlés alaptétele) *Ha a $H^\star$ entrópiájú stacionárius forrás közleményeit C kapacitású zajmentes csatornán továbbítjuk, akkor nincsen olyan egyértelműen dekódolható blokkonkénti kódolási eljárás, melynél*

$$L < \frac{H^\star}{C},$$

ha viszont $R > H^\star$, akkor létezik olyan blokkhossz, hogy

$$L < \frac{R}{C}.$$

4.45. Megjegyzés. *A McMillan-particionálási tétel szerepe: 1. Felhasználható állandó kódhosszú kód tervezéséhez. 2. Gyakorlati szempont: megfelelő az olyan kódolás is, amelynél annak valószínűsége, hogy egy kódszó dekódolásánál hibát követünk el kisebb, mint egy előre megadott $\delta > 0$ szám. Az ilyen kódolási eljárást $1 - \delta$ megbízhatósággal dekódolhatónak nevezzük.*

4.46. Megjegyzés. *A jegyzetnek nem feladata kompresszióval, tömörítéssel foglalkozni, de közvetlenül kapcsolódik a blokkos kódoláshoz. Kompresszióról beszélünk, ha a forrásüzenetet úgy kódoljuk, hogy a kódüzenet rövidebb, mint az eredeti. Erre biztosíték ha pl.*

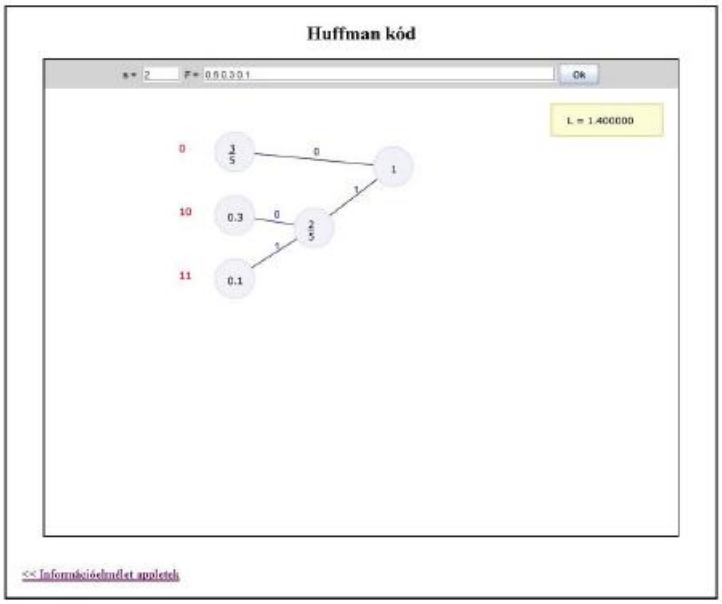
$$X = Y,$$

mert ekkor az entrópia tulajdonságai alapján

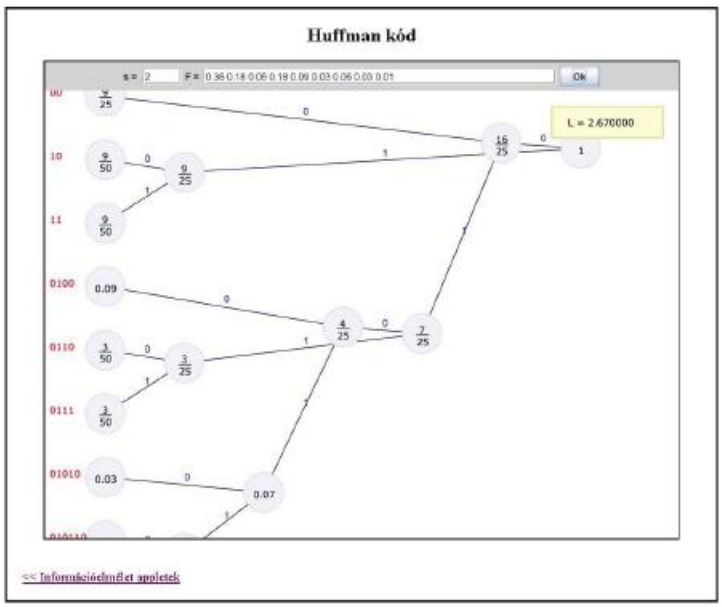
$$1 \geq \frac{H(\mathcal{P})}{\log_2 s}.$$

A következő példák mutatják, hogy milyen lehetőségeink vannak független (emlékezetnélküli) és függő esetben.

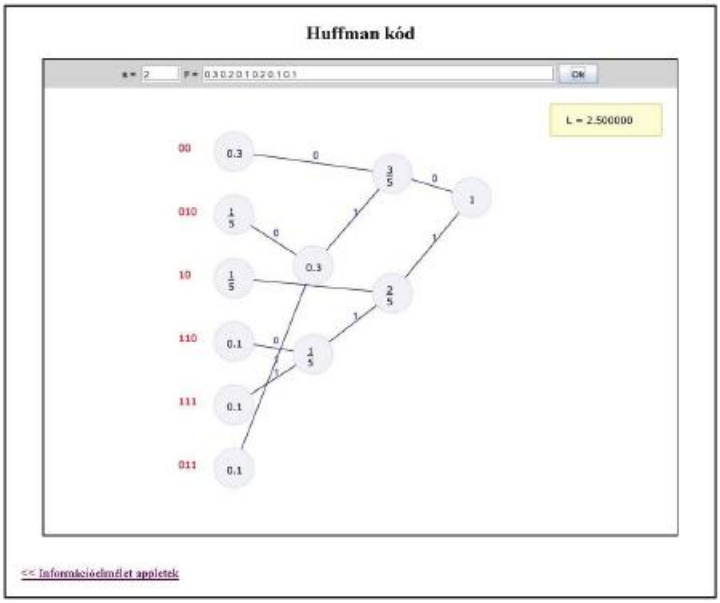
4.22. ábra - Példa blokkos kódoláshoz 1.



4.23. ábra - Példa blokkos kódoláshoz 2.



4.24. ábra - Példa blokkos kódoláshoz 3.



10. 4.10. Feladatok

1. Egyértelműen dekódolható-e az alábbi kód:

$$K = \{cdc, cccd, cddc, ddcc, cccdd, ccddc, ddddc, dcdcdd\},$$

ahol K a kódszavak halmaza? Ha nem, akkor adjon meg két forrásüzenetet, amelynek megegyezik a kódja!

2. Egyértelműen dekódolható-e az alábbi kód:

$$K = \{abc, abcd, e, dba, bace, ceac, ceab, eabd\},$$

ahol K a kódszavak halmaza?

3. Az n és s milyen pozítív egész értékeire teljesülhet az $1 = \sum_{i=1}^{n} s^{-L_i}$ egyenlőség, ahol az L_i értékek alkalmasan választott természetes számok?

4. A Kraft-Fano egyenlőtlenség alapján bizonyítsa be, hogy létezik prefix kód, amelyre az L átlagos kódhossz olyan, hogy

$$L \leq \frac{H(\xi)}{\log_2 s} + 1,$$

ahol $H(\xi)$ a forrásábécé eloszlásának az entrópiája és s a kódábécé elemeinek a száma!

5. Bizonyítsa be, hogy a $\mathcal{Q}$ eloszlású Z forrásábécé d számú kódjelből alkotott egyértelműen dekódolható kódolásai között mindig van optimális kódolás!

6. Egyértelműen dekódolható-e az alábbi kód:

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

ahol K a kódszavak halmaza?

7. A Kraft-Fano egyenlőtlenség alapján bizonyítsa be, hogy minden prefix kódra

$$L \geq \frac{H(\mathcal{P})}{\log_2 s},$$

ahol L az átlagos kódhossz, $H(\mathcal{P})$ a forrásábécé eloszlásának az entrópiája és s a kódábécé elemeinek a száma!

8. Egyértelműen dekódolható-e az alábbi kód:

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

ahol K a kódszavak halmaza?

9. A $\mathcal{P} = \{0.51, 0.18, 0.12, 0.09, 0.05, 0.04, 0.01\}$ eloszláshoz határozza meg a Shannon-Fano bináris kódot! Határozza meg az átlagos kódhosszot és hasonlítsa össze az entrópiából adódó alsó korláttal!

10. A $\mathcal{P} = \{0.31, 0.15, 0.18, 0.07, 0.08, 0.09, 0.12\}$ eloszláshoz határozza meg a Huffman-féle kódot és az optimális kód hatásfokát, ha a kódábécé $Y = \{1, 2, 3\}$!

11. A $\mathcal{P} = \{\frac{3}{8}, \frac{1}{8}, \frac{1}{6}, \frac{1}{8}, \frac{1}{12}, \frac{1}{8}\}$ eloszláshoz határozza meg a Gilbert-Moore kódot, ha a kódábécé $\{0, 1, 2\}$! Határozza meg a kód hatásfokát!

12. A $\mathcal{P} = \{0.51, 0.18, 0.12, 0.09, 0.05, 0.04, 0.01\}$ eloszláshoz határozza meg a Gilbert-Moore kódot és az átlagos kódhosszot, ha a csatornaábécé $\{0, 1\}$!

13. A

$$\mathcal{P} = \{0.2, 0.18, 0.1, 0.1, 0.1, 0.061, 0.059, 0.04, 0.08, 0.04, 0.03, 0.01\}$$

eloszláshoz határozza meg a Gilbert-Moore kód hatásfokát, ha a kódábécé $Y = \{1, 2, 3\}$!

14. A

$$\mathcal{P} = \{0.2, 0.18, 0.1, 0.1, 0.1, 0.061, 0.059, 0.04, 0.04, 0.04, 0.04, 0.03, 0.01\}$$

eloszláshoz határozza meg az optimális kód hatásfokát, ha a kódábécé $Y = \{1, 2, 3\}$!

15. A $\mathcal{P} = \{0.51, 0.18, 0.12, 0.09, 0.05, 0.04, 0.01\}$ eloszláshoz határozza meg az optimális kódot, ha a csatornaábécé $\{0, 1, 2\}$! Határozza meg az átlagos kódhosszot és hasonlítsa össze az entrópiából adódó alsó korláttal!

16. A $\mathcal{P} = \{0.51, 0.12, 0.04, 0.09, 0.05, 0.01, 0.18\}$ eloszláshoz határozza meg a Huffman-féle kódot és az optimális kód hatásfokát, ha a kódábécé $Y = \{1, 2, 3\}$!

17. Bizonyítsa be, hogy a

$$\mathcal{P} = \{3^{-1}, 3^{-1}, 3^{-2}, 3^{-2}, 3^{-3}, 3^{-3}, 3^{-3}\},$$

eloszlású forrásábécének a $\{0, 1, 2\}$ kódjelekből alkotott minden optimális prefix kódjára igaz az, hogy a csatorna kimenetelénél a kódjelek mind 3^{-1} valószínűséggel fordulnak elő!

18. Mutassa meg, hogy az $X = \{x_1, x_2, \ldots, x_8\}$ forrásábécének egyetlen olyan egyértelműen dekódolható bináris kódja van, ahol a maximális kódhossz három és határozza ezt meg!

19. Legyen az X forrásábécé eloszlása

$$\mathcal{P} = \{2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}, 2^{-5}, 2^{-5}\},$$

a kódolása pedig

$$K = \{1, 01, 001, 0001, 00001, 00000\}.$$

Igazolja, hogy egy véletlenszerűen választott közlemény kódolásához felhasznált 0-k és 1-esek számának várható értéke megegyezik!

20. Sorolja fel az összes olyan prefix és az összes olyan egyértelműen dekódolható bináris kódot, melyek kódhosszai 1, 2, 3, 3!

21. Adott egy információforrás, amely az $X = \{0, 1\}$ jeleket állítja elő $\mathcal{P} = \{\frac{1}{4}, \frac{3}{4}\}$ valószínűségekkel. Mennyi lesz az egy betűre jutó átlagos kódhossz 3 hosszúságú blokkok alkalmazása esetén? Milyen blokkhossz esetén lehetséges kompresszió?

22. Adott egy információforrás, amely az $X = \{a, b, c\}$ jeleket állítja elő $\mathcal{P} = \{\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\}$ valószínűségekkel. Mennyi lesz az egy betűre jutó átlagos kódhossz 2 hosszúságú blokkok alkalmazása esetén? Milyen blokkhossz esetén lehetséges kompresszió?

23. Adott egy információforrás, amely az $X = \{a, b, c\}$ jeleket állítja elő $\mathcal{P} = \{\frac{1}{3}, \frac{1}{6}, \frac{1}{2}\}$ valószínűségekkel. Mennyi lesz az egy betűre jutó átlagos kódhossz 2 hosszúságú blokkok alkalmazása esetén? Milyen blokkhossz esetén lehetséges kompresszió?

24. Legalább hány kódjelre van szükség az $X = \{x_1, x_2, \ldots, x_8\}$ forrásábécének olyan prefix kódjának az elkészítéséhez, melynek a kódhosszai rendre 2, 2, 2, 4, 4, 4, 4, 4.

25. A $\mathcal{P} = \{0.2, 0.15, 0.15, 0.1, 0.1, 0.1, 0.1, 0.1\}$ eloszlású nyolcelemű X forrásábécének készítse el a 0, 1 kódjelekből két olyan egyértelműen dekódolható optimális kódját, melyek kódhosszai különbözőek!

26. Sorolja fel az összes olyan prefix és az összes olyan egyértelműen dekódolható bináris kódot, melyek kódhosszai 1, 2, 3, 3!

27. Adott egy információforrás, amely az $X = \{0, 1\}$ jeleket állítja elő $\mathcal{P} = \{\frac{1}{3}, \frac{2}{3}\}$ valószínűségekkel. Mennyi lesz az egy betűre jutó átlagos kódhossz 3 hosszúságú blokkok alkalmazása esetén? Milyen blokkhossz esetén lehetséges kompresszió?

28. A $\mathcal{P} = \{0.47, 0.13, 0.12, 0.09, 0.09, 0.05, 0.04, 0.01\}$ eloszláshoz határozza meg az optimális kódot, ha a csatornaábécé $\{0, 1, 2\}$! Határozza meg az átlagos kódhosszot és a hatásfokot!

29. A $\mathcal{P} = \{0.47, 0.13, 0.12, 0.09, 0.09, 0.05, 0.04, 0.01\}$ eloszláshoz határozza meg a Shannon-Fano kódot, ha a csatornaábécé $\{0, 1, 2\}$! Határozza meg az átlagos kódhosszot és a hatásfokot!

30. A $\mathcal{P} = \{0.31, 0.15, 0.18, 0.07, 0.08, 0.09, 0.12\}$ eloszláshoz határozza meg a Huffman-féle kódot és az optimális kód hatásfokát, ha a kódábécé $Y = \{1, 2, 3\}$!

11. 4.11. Önellenőrző kérdések

1. Ismertesse a McMillan-dekódolási tételt!

2. Ismertesse a McMillan-felbontási tételt!

3. Mutassa meg, hogy az $X = \{x_1, x_2, \ldots, x_8\}$ forrásábécének egyetlen olyan egyértelműen dekódolható bináris kódja van, ahol a maximális kódhossz három!

4. Bizonyítsa, hogy létezik maximális hatásfokú prefix kód!

5. Ismertesse a Kraft-Fano egyenlőtlenséget!

6. Definiálja a prefix kódot!

7. Ismertesse az optimális kód tulajdonságait!

8. Ismertesse a zajmentes kódolás alaptételét!

9. Definiálja a stacionér forrás entrópiáját!

10. Definiálja egy kód hatásfokát!

11. Bizonyítsa a maximális hatásfokú kód létezését!

12. Bizonyítsa a stacionárius forrás entrópiájának a létezését!

5. fejezet - Csatornakapacitás

A csatorna tulajdonságai szempontjából az egyik legfontosabb tulajdonság a csatornakapacitás. A csatornakapacitás az elemenként (□betűnként") átvihető információ mennyiségével egyenlő (s így a csatornakapacitás lényegében sebesség jellegű mennyiség, ahol azonban a sebesség vonatkoztatási alapja nem az idő, hanem a □betű", noha a kettő ebből a szempontból összefügg). A csatornakapacitás fogalmával függ össze a redundancia, amely a betűnként továbbított átlagos információ mennyiségét és a csatornakapacitást hasonlítja össze mennyiségileg, vagyis azt mondja meg, hogy mennyivel terjengősebb egy közlemény az elvben lehetséges legrövidebb formánál. Maguk a kódok (szokás őket - mesterséges - nyelveknek is nevezni, így pl. idetartoznak a számítógépek programnyelvei), amelyekkel az információelmélet csökkenteni igyekszik a redundanciát, illetőleg növelni a kölcsönös információt, több típusba sorolhatók aszerint, hogy hány elemet (□betűt") használnak fel a közlemények összeállításánál. Az ún. bináris kódokban, amelyeket elterjedten használnak a digitális számítógépekben, két □betű" van csak: a 0 és az 1. Az ún. kódoláselmélet az információelméleten belül a különböző feltételeket teljesítő kódok konstruálásával foglalkozik. Ezekre igyekszik általános módszereket kidolgozni. Shannon, az információelmélet egyik úttörője általánosságban bebizonyította, hogy alkalmas kódolási eljárással zaj jelenlétében is megvalósítható a hibátlan (pontosabban előírhatóan kis hibavalószínűségű) információtovábbítás, ha sebessége kisebb a csatornakapacitásnál. Ez a tétel az információelmélet egyik alaptétele, maga a tétel azonban semmit sem tartalmaz a implementációra vonatkozóan. Az információelmélet gyakorlati feladata tehát az optimalizálásban ragadható meg. Ez egyrészt a költségek csökkentését, másrészt pedig a hibamentesség növelését jelenti.

Fizikai valójukban a csatornák nagyon sokfélék lehetnek: a levegő, a telefonvezeték, az optikai üvegszál, az élőlények idegszálai, a könyv, a CD stb. Osztályozni is több szempontból lehet őket. A térbeli csatornák a tér valamelyik pontjából egy vagy több másik pontjába, az időbeli csatornák a T időponttól a $(T+t)$ időpontba szállítják az információkat. Előbbiekre példa a telefonvezeték, utóbbiakra a CD. Természetesen ez a megkülönböztetés csak a lényegi jegyekre vonatkozik, mivel az információnak a térbeli csatornában is időre van szüksége, hogy célba jusson, a szóbeli csatornákon is lehet térben szállítani az információt.

Az volna az eszményi, ha a csatorna kimeneteli oldalán mindig azt az információt kapnánk meg, amely a másik oldalán belépett, azaz a belépő x_i jelnek a kimenetelnél mindig y_j jel felelne meg. Az ilyen - csak elméletben létező- ideális csatorna neve zajmentes csatorna. Sajnos a reális csatornák mindig zajosak, zaj minden olyan jelenség, amely a hírközlő csatornában □megtámadja" a hasznos információt, megcsonkítja, elnyomja, eltorzítja, legrosszabb esetben meg is semmisíti. Másképpen fogalmazva: zajos csatornánál a kilépő jel nem felel meg mindig a belépő jelnek, hamis jelek keverednek az igaziak közé. Zaj például az az elektromágneses rezgés, amely zavarja a rádióvételt, az utca zaja, amely elnyomja a beszélgetőtársunk hangját, a sajtóhiba. A zajokat két csoportra oszthatjuk. A rendszertorzítás azonos jel esetén mindig azonos, és elvileg teljesen kiküszöbölhető. A csatorna- vagy csőzaj független a jeltől, rendszertelen, statisztikus jellege van, és teljesen sohasem szüntethető meg. (Tulajdonképpen a zaj is információ, csak éppen nem az, amire szükségünk van, s nagyon sokszor a kódját sem ismerjük. Az is előfordulhat, hogy valamely jelenség zaj egy szempontból, s értékes információ egy másikból. Például a légköri elektromos jelenségek a rádióhallgató és a légkör fizikáját kutató tudós szempontjából.)

5.1. Definíció. *Az információtovábbítás eszközeként definiáljuk a diszkrét emlékezetnélküli csatornát (angol rövidítéssel DMC) a következőképpen:*

- a csatorna bemenetén ütemenként egy szimbólumot fogad, és ütemenként egy szimbólum jelenik meg a kimenetén (szinkron működés);

- a bemeneti és a kimeneti szimbólumkészlet nem feltétlenül azonos, de mindkettő rögzített és véges számú elemet tartalmaz (diszkrét);

- ha a bemeneti szimbólumok egymástól függetlenek, akkor a kimeneti szimbólumok is függetlenek lesznek (emlékezet nélküli).

Az eddigiek alapján tudjuk, hogy zajmentes csatorna esetén az egy csatornajelre (kódábécébeli elemre) jutó átlagos információ átvitel megegyezik a kódábécé eloszlásához kapcsolódó entrópával, azaz $H(Q)$, amely akkor maximális, ha a jelek eloszlása egyenletes. A forrás optimális kódolása ezt a maximális esetet próbálja közelíteni. A következő szakaszban arra próbálunk választ adni, mi történik akkor, ha a csatornajelek átviteli ideje nem azonos.

1. 5.1. Zajmentes csatorna kapacitása nem azonos átviteli idő esetén

Adott $Y = \{y_1, \dots, y_s\}$ csatornaábécé esetén feltételezzük, hogy a jelek átviteli ideje ismert illetve kísérleti úton megfelelő pontossággal meghatározható. Jelölje az időket $T = \{t_1, \dots, t_s\}$. Feltételezzük, hogy $t_i > 0$ $(i = 1, \dots, s)$.

Ha egy információforrás jeleket bocsát ki, akkor azt kódolva keletkezik egy kódüzenet (csatornaüzenet). Ekkor felmerülnek a következő kérdések: Egy adott kódolás esetén milyen gyorsan, milyen átlagos sebességgel továbbítja a csatorna az üzenetet? Van-e az információtovábbításnak felső határa és ha van, mennyi az?

Nyilván a sebesség függ a kódolástól (a kódüzenet elemeinek az eloszlásától), ezért az a célunk, hogy a kódot úgy válasszuk meg, hogy az információtovábbítás sebessége maximális legyen.

Legyen a csatornaábécé betűinek eloszlása $\mathcal{Q} = \{q_1, \dots, q_s\}$. Ha a csatornaüzenet hossza N, akkor egy kiválasztott jel, pl. y_i várhatóan Nq_i-szer fordul elő és várhatóan $-Nq_i \log_2 q_i$ mennyiségű információt továbbít. Ugyanez a teljes üzenetre

$$\sum_{i=1}^{s} Nq_i \log_2 \frac{1}{q_i} = NH(\mathcal{Q}).$$

Hasonlóan a várható átviteli idő:

$$\sum_{i=1}^{s} Nq_i t_i = N \sum_{i=1}^{s} q_i t_i,$$

ebből az egy betűre jutó várható átviteli idő (jelölje τ)

$$\tau = \sum_{i=1}^{s} q_i t_i.$$

5.2. Definíció. *Az információtovábbítás sebességének nevezzük a*

$$v(\mathcal{Q}) = \frac{H(\mathcal{Q})}{\tau}$$

mennyiséget.

5.3. Megjegyzés. *Az előző definícióban szereplő mennyiség átlagsebesség.*

5.4. Definíció. *Az információátviteli sebesség maximumát csatornakapacitásnak nevezzük (jele: C), azaz*

$$C = \max_{\mathcal{Q}} \quad v(\mathcal{Q}),$$

ahol $\mathcal{Q}$ a csatornaábécé lehetséges eloszlása.

5.5. Megjegyzés. *Az eloszlások összessége az előző definícióban kompakt halmazt alkot és $v(\mathcal{Q})$ folytonosan függ a $\mathcal{Q}$ eloszlástól, ezért létezik a szupremuma és azt fel is veszi.*

5.6. Tétel. *Zajmentes, emlékezetnélküli (véges, diszkrét) csatorna esetén a csatornakapacitás a*

$$\sum_{i=1}^{s} 2^{-vt_i} = 1$$

egyenlet egyetlen $v = C \geq 0$ megoldása. Ezt a

$$q_i = 2^{-Ct_i} \qquad (i = 1, \dots, s)$$

eloszlás realizálja.

Bizonyítás. Ha C megoldása az egyenletnek, akkor a tétel szerint megadott eloszlás és az átlagsebességre teljesül a következő:

$$v(\mathcal{Q}) = \frac{H(\mathcal{Q})}{\tau} = \frac{\sum_{i=1}^{s} q_i \log_2 \frac{1}{q_i}}{\sum_{i=1}^{s} q_i t_i} \leq \frac{\sum_{i=1}^{s} q_i \log_2 \frac{1}{2^{-Ct_i}}}{\sum_{i=1}^{s} q_i t_i} = \frac{\sum_{i=1}^{s} q_i C t_i}{\sum_{i=1}^{s} q_i t_i} = C,$$

ahol az egyenlőség csak a tételben megadott eloszlás esetén teljesül.

Tehát csak azt kell belátnunk, hogy C egyértelműen létezik. Az

$$f(v) = \sum_{i=1}^{s} 2^{-v t_i}$$

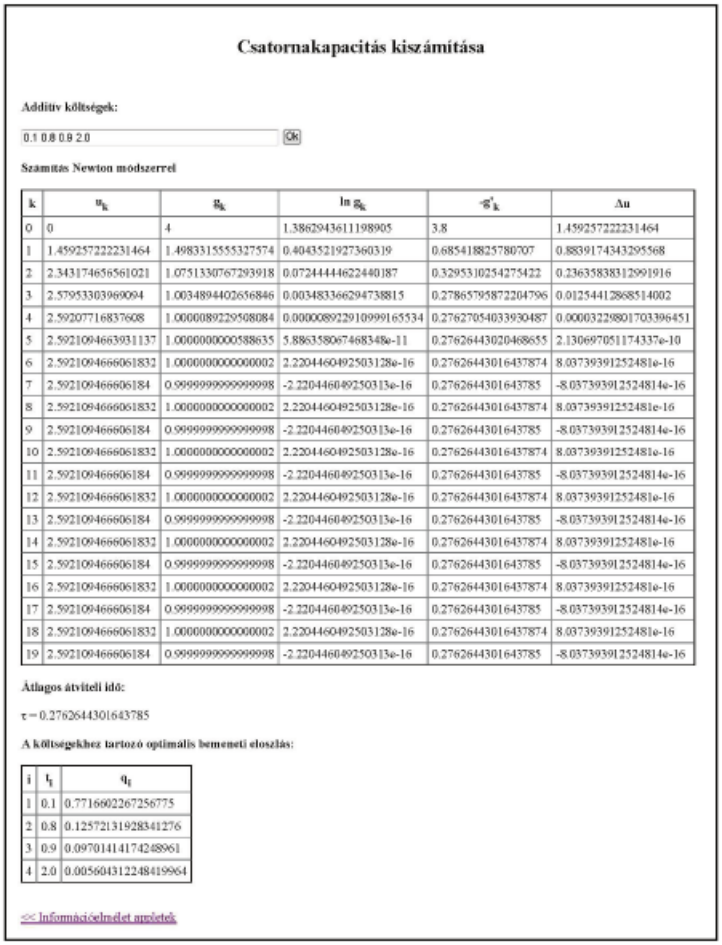
függvény szigorúan monoton csökkenő. Továbbá,

$$f(0) = s > 1 \quad \text{és} \quad \lim_{v \to +\infty} f(v) = 0.$$

A folytonosság miatt létezik $v = C$, ahol $f(C) = 1$ és ez egyértelmű a szigorú monotonitás miatt.

■

5.1. ábra - Példa csatornakapcitás numerikus meghatározására additív költség esetén



Csatornakapacitás kiszámítása

Additív költségek:

0.1 0.8 0.9 2.0 OK

Számítás Newton módszerrel

k	u_k	g_k	$\ln g_k$	$-g'_k$	Δu
0	0	4	1.38629436111989O5	3.8	1.459257222231464
1	1.459257222231464	1.4983315555327574	0.4043521927360319	0.685418825780707	0.8839174343295568
2	2.343174656561021	1.0751330767293918	0.07244444622440187	0.3295310254275422	0.236358383129919l6
3	2.579533039690941	1.0034894402656846	0.003483366294738815	0.2786579587220479O	0.012544128685140O2
4	2.592077168376O8	1.0000089229508084	0.000008922910999165534	0.27627054033930487	0.000032288017033964S1
5	2.592109466393113l	1.0000000000588635	5.886358067468348e-11	0.27626443020468655	2.13069705117433Te-10
6	2.592109466606183l	1.0000000000000002	2.220446049250312Be-16	0.27626443016437874	8.037393912524816-16
7	2.592109466606184	0.9999999999999998	-2.220446049250313e-16	0.27626443016437785	-8.037393912524814e-16
8	2.592109466606183l	1.0000000000000002	2.220446049250312Be-16	0.27626443016437874	8.037393912524816-16
9	2.592109466606184	0.9999999999999998	-2.220446049250313e-16	0.27626443016437785	-8.037393912524814e-16
10	2.592109466606183l	1.0000000000000002	2.220446049250312Be-16	0.27626443016437874	8.037393912524816-16
11	2.592109466606184	0.9999999999999998	-2.220446049250313e-16	0.27626443016437785	-8.037393912524814e-16
12	2.592109466606183l	1.0000000000000002	2.220446049250312Be-16	0.27626443016437874	8.037393912524816-16
13	2.592109466606184	0.9999999999999998	-2.220446049250313e-16	0.27626443016437785	-8.037393912524814e-16
14	2.592109466606183l	1.0000000000000002	2.220446049250312Be-16	0.27626443016437874	8.037393912524816-16
15	2.592109466606184	0.9999999999999998	-2.220446049250313e-16	0.27626443016437785	-8.037393912524814e-16
16	2.592109466606183l	1.0000000000000002	2.220446049250312Be-16	0.27626443016437874	8.037393912524816-16
17	2.592109466606184	0.9999999999999998	-2.220446049250313e-16	0.27626443016437785	-8.037393912524814e-16
18	2.592109466606183l	1.0000000000000002	2.220446049250312Be-16	0.27626443016437874	8.037393912524816-16
19	2.592109466606184	0.9999999999999998	-2.220446049250313e-16	0.27626443016437785	-8.037393912524814e-16

Átlagos átviteli idő:

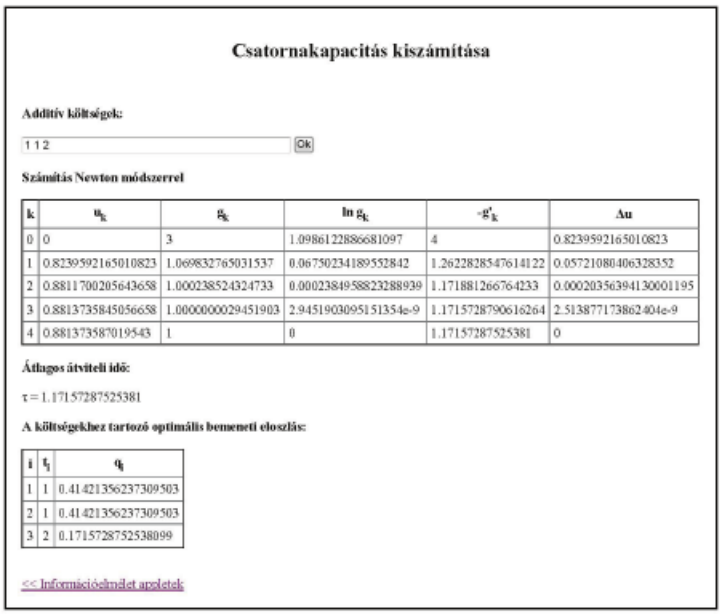
τ = 0.2762644301643785

A költségekhez tartozó optimális bemeneti eloszlás:

i	t_i	q_i
1	0.1	0.771660226725677S
2	0.8	0.12572131928341276
3	0.9	0.09701414174248961
4	2.0	0.005604312248419964

<< Információelmélet mvelletek

5.2. ábra - Példa csatornakapcitás numrikus meghatározására additív költség esetén



Csatornakapacitás kiszámítása

Additív költségek:

1 1 2 Ok

Számítás Newton módszerrel

k	u_k	g_k	$\ln g_k$	$-g'_k$	Δu
0	0	3	1.0986122886681097	4	0.8239592165010823
1	0.8239592165010823	1.0698327650315 37	0.06750234189552842	1.2622828547614122	0.05721080406328352
2	0.8811700205643658	1.000238524324733	0.00023849588232889 39	1.171881266764233	0.00020356394130001195
3	0.8813735845056658	1.0000000029451903	2.9451903095151354e-9	1.171572879061626 4	2.513877173862404e-9
4	0.8813735870195 43	1	0	1.171572875253 81	0

Átlagos átviteli idő:

τ = 1.171572875253 81

A költségekhez tartozó optimális bemeneti eloszlás:

i	t_i	q_i
1	1	0.4142135623730 9503
2	1	0.4142135623730 9503
3	2	0.1715728752538 099

<< Információelmélet appletek

2. 5.2. Shannon-Fano algoritmus, tetszőleges eloszlás esetén

Most nézzük, hogy mit kell tennünk, ha az intervallumok egyenletes felosztása helyett a felosztást tetszőleges módon végezzük el.

Legyen

$$\mathcal{P} = \{p_1, p_2, \ldots, p_n\}, p_i > 0, (i = 1, \ldots, n), \sum_{i=1}^{n} p_i = 1$$

tetszőleges forráseloszlás, továbbá legyen

$$\mathcal{Q} = (q_1, q_2, \ldots, q_k)$$

az optimális bemeneti eloszlás. Ekkor a lépések a következők:

1. Rendezzük a $\mathcal{P}$ eloszlás valószínűségeit csökkenő sorrendbe;
2. Képezzük az $x_i (i = 1, \ldots, n)$ értékeket a következőképpen:

   $x_1 = 0, x_2 = p_1, x_3 = p_1 + p_2, \ldots, x_n = p_1 + \cdots + p_{n-1}$;

   $(p_1 \geq p_2 \geq \cdots \geq p_n > 0)$
3. Ábrázoljuk ezen értékeket a $[0, 1)$ intervallumon. Majd osszuk fel az $[0, 1)$ intervallumot a q_i valószínűségek arányában $(i = 1, \ldots, s)$ (s a kódábécé elemeinek száma);
4. Azokat az intervallumokat, melyek egynél több x_i értéket tartalmaznak osszuk fel újra egészen addig míg mindegyik x_i más intervallumba nem kerül;
5. A $g(x_i)$ kódszó az $s^{-1}, s^{-2}, \ldots, s^{-k}$ hosszúságú intervallumok megfelelő sorszámából áll, amelyekben x_i benne van.

Mint észrevehetjük ez a megoldás csak az intervallumok felosztásának módjában tér el a korábban ismertetett eljárástól. Ezzel a módszerrel a kódolás additív költség esetén is elvégezhető.

Nézzünk egy példát az egyenletes esetre:

5.1. Példa. Legyen $K = \{A, B, C, D\}$ $(0.36, 0.10, 0.20, 0.08, 0.08, 0.04, 0.04, 0.02)$ bemeneti eloszlás, valamint a kódábécé legyen $\{\hat{A}, \hat{B}, \hat{C}, \hat{D}\}$. A rendezés után képezzük az x_i értékeket, amire kapjuk:

$$(0.0, 0.36, 0.56, 0.66, 0.74, 0.82, 0.90, 0.94, 0.98).$$

Ekkor az x_i értékekhez a generált kódszavak a következők:

$$\mathcal{K} = \begin{bmatrix} 0.0: & A \\ 0.36: & B \\ 0.56: & CA \\ 0.66: & CC \\ 0.74: & CD \\ 0.82: & BD \\ 0.90: & DC \\ 0.94: & DDA \\ 0.98: & DDC \end{bmatrix}.$$

▲

3. 5.3. Zajos csatorna kapacitása

Bemeneti ábécé: $Y = \{y_1, \ldots, y_s\}$.

Kimeneti ábécé: $Z = \{z_1, \ldots, z_m\}$.

$$\begin{aligned} q_j &= P(\xi = y_j), \quad j = 1, 2, \ldots, s, \\ r_i &= P(\eta = z_i), \quad i = 1, 2, \ldots, m, \\ t_{ij} &= P(z_i | y_j), \\ p_{ij} &= t_{ij} q_j, \\ r_i &= \sum_{j=1}^{s} t_{ij} q_j, \quad i = 1, 2, \ldots, m, \\ \mathcal{R} &= T\mathcal{Q}. \end{aligned}$$

A $T = (t_{ij})$ mátrixot adókarakterisztika vagy csatornamátrixnak nevezzük. Míg az együttes eloszlás $\mathcal{P} = (p_{ij})$ mátrixa az ún. átviteli mátrix.

Csatornakapacitás:

$$C = \max_{\mathcal{Q}} I(\xi, \eta).$$

5.7. Megjegyzés. $I(\xi, \eta) = H(\mathcal{R}) - H(\mathcal{R}|\mathcal{Q}) = H(\mathcal{Q}) - H(\mathcal{Q}|\mathcal{R})$.

A zajos csatornák osztályozása a csatornamátrix alapján:

1. $H(\mathcal{Q}|\mathcal{R}) = 0$ – veszteségmentes. A kimenet egyértelműen meghatározza a bemenetet. Minden i esetén létezik j, hogy $p_{ij} = r_i$.

2. $H(\mathcal{R}|\mathcal{Q}) = 0$ – determinisztikus. A bemenet egyértelműen meghatározza a kimenetet. Minden j esetén létezik i, hogy $p_{ij} = q_j$.

3. $H(\mathcal{R}|\mathcal{Q}) = H(\mathcal{Q}|\mathcal{R}) = 0$ – zajmentes. A bemenet és a kimenet egyértelműen meghatározzák egymást. Ekkor $t_{ij} = \delta_{ij}$.

4. $C = 0$, azaz $H(\mathcal{R}|\mathcal{Q}) = H(\mathcal{R})$ – használhatatlan. Pl. az oszlopok megegyeznek.

5. Szimmetrikus csatorna – a sorok és az oszlopok is ugyanazokból a vektorokból épülnek fel.

$$T = \begin{pmatrix} \frac{1}{6} & \frac{1}{3} \\ \frac{1}{6} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{6} \end{pmatrix}.$$

Szimmetrikus csatorna estén

$$H(\eta|\xi = y_j)$$

minden j esetén ugyanaz, így

$$H(\eta|\xi) = H(\eta|\xi = y_j).$$

Tehát

$$C = \max_{\mathcal{Q}} I(\xi, \eta) = \max_{\mathcal{Q}} H(\mathcal{R}) - H(\mathcal{R}|\mathcal{Q}) = \log_2 m - H(\mathcal{R}|\mathcal{Q}).$$

5.8. Megjegyzés. *Ha a kimeneti eloszlás egyenletes, akkor a bemeneti is az.*

Legyen $s = m$, azaz a bemeneti és kimeneti ábécé betűinek száma megegyezik. Jelölje H_k a csatornamátrix k-adik oszlopának entrópiáját, azaz $H_k = H(\eta|\xi = y_k)$. Ekkor a

$$C = \max_{\mathcal{Q}} I(\xi, \eta)$$

szélsőérték feladatot kell megoldanunk a

$$\sum_{j=1}^{s} q_j = 1$$

feltétel mellett, így a Lagrange-féle multiplikátoros módszert alkalmazhatjuk. A Lagrange-függvény

$$L(\mathcal{Q}, \lambda) = \sum_{i=1}^{s} r_i \log_2 \frac{1}{r_i} + \sum_{i=1}^{s} \sum_{j=1}^{s} p_{ij} \log_2 t_{ij} + \lambda \left(\sum_{j=1}^{s} q_j - 1 \right).$$

Határozzuk meg a deriváltakat:

1. $r_i = \sum_{j=1}^{s} q_j t_{ij}$, így

$$\frac{\partial H(\eta)}{\partial q_k} = \sum_{i=1}^{s} \frac{\partial H(\eta)}{\partial r_i} \frac{\partial r_i}{\partial q_k} =$$

$$= -\sum_{i=1}^{s} \left(\log_2 r_i + \frac{1}{\ln 2} \right) t_{ik} =$$

$$= -\frac{1}{\ln 2} - \sum_{i=1}^{s} t_{ik} \log_2 r_i.$$

2. $H(\eta|\xi) = \sum_{j=1}^{s} q_j H_j$, így

$$\frac{\partial H(\eta|\xi)}{\partial q_k} = H_k.$$

Tehát a Lagrange-függvény deriváltjaiból adódó egyenletrendszer

$$\frac{\partial L}{\partial q_k} = -\frac{1}{\ln 2} - \sum_{i=1}^{s} t_{ik} \log_2 r_i - H_k + \lambda = 0, \quad k = 1, \ldots, s,$$

$$\frac{\partial L}{\partial \lambda} = \sum_{j=1}^{s} q_j - 1 = 0.$$

1. Az első s egyenlet mindegyikét szorozzuk meg a megfelelő q_k valószínűséggel és adjuk őket össze. Ekkor

$$-\frac{1}{\ln 2} + \sum_{i=1}^{s} r_i \log_2 \frac{1}{r_i} - \sum_{k=1} q_k H_k + \lambda = 0.$$

Ebből

$$C = H(\eta) - H(\eta|\xi) = \frac{1}{\ln 2} - \lambda.$$

2. Meghatározzuk $C = \frac{1}{\ln 2} - \lambda$ értékét.

A Lagrange-függvény parciális deriváltjaiból adódó egyenleteket alakítsuk át a következőképpen.

$$-H_k = \sum_{i=1}^{s} \left(\frac{1}{\ln 2} - \lambda + \log_2 r_i \right) t_{ik}, \quad k = 1, \ldots, s.$$

Ez egy lineáris egyenletrendszernek tekinthető, amelynek az ismeretlenjei

$$u_i = \frac{1}{\ln 2} - \lambda + \log_2 r_i, \quad i = 1, \ldots, s.$$

A megoldás felírható

$$u_i = \frac{1}{\ln 2} - \lambda + \log_2 r_i = -\sum_{k=1}^{s} H_k \tau_{ki}, \quad i = 1, \ldots, s,$$

alakban, ahol a (τ_{ki}) mátrix a T mátrix inverze. Ekkor

$$2^{-\sum_{k=1}^{s} H_k \tau_{ki}} = r_i 2^{\frac{1}{\ln 2} - \lambda}, \quad i = 1, \ldots, s.$$

Összeadva az egyenleteket és alkalmazva a $\log_2$ függvényt azt kapjuk, hogy

$$\log_2 \sum_{i=1}^{s} 2^{-\sum_{k=1}^{s} H_k \tau_{ki}} = \frac{1}{\ln 2} - \lambda.$$

A lineáris egyenletrendszerből

$$2^{-C} \cdot 2^{-\sum_{k=1}^{s} H_k \tau_{ki}} = r_i, \quad i = 1, \ldots, s,$$

ahol $C = \frac{1}{\ln 2} - \lambda.$ Ezzel meghatároztuk az $\mathcal{R}$ kimeneti eloszlást. Az

$$\mathcal{R} = T\mathcal{Q}$$

lineáris egyenletrendszer megoldásával pedig meghatározható a $\mathcal{Q}$ bemeneti eloszlás.

5.9. Megjegyzés. *Ha létezik $q_k = 0$, akkor problémás a megoldás első része.*

5.10. Megjegyzés. *$I(\xi,\eta)$ maximális (és ezért egyenlő a csatornakapacitással) akkor és csak akkor, ha a $\mathcal{Q}$ bemeneti eloszlás olyan, hogy*

a) $\dfrac{\partial I}{\partial q_k} = \lambda$ *minden k esetén, amikor $q_k \neq 0$.*

b) $\dfrac{\partial I}{\partial q_k} \leq \lambda$ *minden k esetén, amikor $q_k = 0$.*

5.11. Tétel. *A létező megoldás egyértelmű és maximalizálja a kölcsönös információmennyiséget.*

Bizonyítás. A csatornamátrix rögzített, ezért $I(\xi,\eta)$ csak a bemeneti $\mathcal{Q}$ eloszlástól függ. Jelölje: $I(\mathcal{Q})$.

$I(\mathcal{Q})$ konkáv függvénye a $\mathcal{Q}$ eloszlásnak.

Legyen $\mathcal{Q}_1 = \{q_{11}, q_{12}, \ldots, q_{1s}\}, \mathcal{Q}_2 = \{q_{21}, q_{22}, \ldots, q_{2s}\}$, és

$$\mathcal{Q} = \vartheta\mathcal{Q}_1 + (1-\vartheta)\mathcal{Q}_2, \quad \text{ahol} \quad 0 \leq \vartheta \leq 1.$$

$$\begin{aligned} H(\mathcal{R}|\mathcal{Q}) = \sum_{j=1}^{s} q_j H_j &= \sum_{j=1}^{s} (\vartheta q_{1j} + (1-\vartheta)q_{2j})H_j = \\ &= \vartheta \sum_{j=1}^{s} q_{1j}H_j + (1-\vartheta)\sum_{j=1}^{s} q_{2j}H_j = \\ &= \vartheta H(\mathcal{R}_1|\mathcal{Q}_1) + (1-\vartheta)H(\mathcal{R}_2|\mathcal{Q}_2). \end{aligned}$$

Ebből adódóan

$$I(\mathcal{Q}) = H(\mathcal{Q}) - \vartheta H(\mathcal{R}_1|\mathcal{Q}_1) - (1-\vartheta)H(\mathcal{R}_2|\mathcal{Q}_2).$$

Viszont az entrópia konkáv, így $I(\mathcal{Q})$ is konkáv.

■

5.12. Lemma. *Tegyük fel, hogy $g : \mathbf{R}^n \to \mathbf{R}$ konkáv az*

$$S = \{(p_1, \ldots, p_n) | p_i \geq 0, \quad i = 1, 2, \ldots, n \text{ és } \sum_{i=1}^{n} p_i = 1\}$$

halmazon. Ha g folytonosan differenciálható S belsejében és létezik

$$p_i^* > 0 \quad (i = 1, 2, \ldots, n)$$

úgy, hogy

$$\frac{\partial g}{\partial p_i}|_{p=p^*} = 0, \quad i = 1, 2, \ldots, n \text{ és } p^* \in S,$$

ahol $p = (p_1, \ldots, p_n), p^* = (p_1^*, \ldots, p_n^*)$, akkor a g függvény abszolút maximuma az S halmazon $g(p^*)$.

Bizonyítás. Tegyük fel, hogy $g(p) > g(p^*)$. Legyen $0 < \vartheta \leq 1$, akkor

$$\begin{aligned} \frac{g((1-\vartheta)p^* + \vartheta p) - g(p^*)}{\vartheta} &\geq \frac{(1-\vartheta)g(p^*) + \vartheta g(p) - g(p^*)}{\vartheta} = \\ &= g(p) - g(p^*) > 0. \end{aligned}$$

A feltételek alapján viszont

$$\frac{\partial g}{\partial p_i}|_{p=p^*} = 0, \quad i = 1, 2, \ldots, n,$$

$g'(p) \leq g(p^*)$.ánymenti deriváltnak 0-hoz kellene tartani, ami ellentmondás. Tehát minden $p \in S$ esetén

■

5.2. Példa. Legyen a csatornamátrix

$$T = \begin{pmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{pmatrix}.$$

Ekkor

$$\begin{aligned} q_1 + q_2 &= 1, \\ r_1 &= 0.9q_1 + 0.2q_2, \\ r_2 &= 0.1q_1 + 0.8q_2, \\ H_1 &\approx 0.4689956, \\ H_2 &\approx 0.7219281. \end{aligned}$$

A parciális deriváltakból adódó egyenletrendszer:

$$\begin{aligned} -\frac{1}{\ln 2} - 0.9\log_2 r_1 - 0.1\log_2 r_2 - H_1 + \lambda &= 0, \\ -\frac{1}{\ln 2} - 0.2\log_2 r_1 - 0.8\log_2 r_2 - H_2 + \lambda &= 0. \end{aligned}$$

Átalakítva:

$$\begin{aligned} -H_1 &= \left(\frac{1}{\ln 2} - \lambda + \log_2 r_1\right) 0.9 + \left(\frac{1}{\ln 2} - \lambda + \log_2 r_2\right) 0.1, \\ -H_2 &= \left(\frac{1}{\ln 2} - \lambda + \log_2 r_1\right) 0.2 + \left(\frac{1}{\ln 2} - \lambda + \log_2 r_2\right) 0.8. \end{aligned}$$

Legyen

$$\begin{aligned} u_2 &= \frac{1}{\ln 2} - \lambda + \log_2 r_1, \\ u_1 &= \frac{1}{\ln 2} - \lambda + \log_2 r_2. \end{aligned}$$

Ekkor

$$\begin{aligned} u_2 &\approx -0.7941945, \\ u_1 &\approx -0.4328624, \\ C &= \log_2\left(2^{u_1} + 2^{u_2}\right) \approx 0.397754. \end{aligned}$$

Továbbá

$$\begin{aligned} 2^{-C+u_1} &= r_1 = 0.9q_1 + 0.2q_2, \\ 2^{-C+u_2} &= r_2 = 0.1q_1 + 0.8q_2, \end{aligned}$$

$$\begin{aligned} r_2 &\approx 0.4377112, \\ r_1 &\approx 0.5622888, \\ q_1 &\approx 0.517555, \\ q_2 &\approx 0.48445. \end{aligned}$$

▲

4. 5.4. Arimoto-Blahut algoritmus

Zajos csatorna kapacitásának kiszámítására általános módszert Arimoto [2] és Blahut [6] adtak először egymástól függetlenül 1972-ben. A módszer ismertetése előtt nézzük milyen kifejezésekre lesz szükségünk.

Legyen $\mathcal{P} = \{p_1, \ldots, p_n\}$ a csatorna egy lehetséges bemeneti eloszlása, valamint $\mathcal{Q} = \{q_1, \ldots, q_m\}$ eloszlás a csatorna kimenetén. Ha T az adókarakterisztika mátrix, akkor a $\mathcal{P}$ bemeneti eloszlásból a kimeneti eloszlást a $T\mathcal{P}$ mátrix-vektor szorzat adja. Bontsuk fel az adókarakterisztika mátrixot oszlopvektoraira és a T mátrix j-edik oszlopát jelölje T_j. Továbbá legyen

$$D_j(\mathcal{Q}) = D(T_j||\mathcal{Q}) = \sum_{i=1}^{m} t_{ij} \log \frac{t_{ij}}{q_i}.$$

Legyen $\mathcal{P}^{(0)} p_i > 0, (i = 1, \ldots, n)$ tetszőleges eloszlás. Képezzük az alábbi iterációs formula alapján eloszlásoknak egy sorozatát, valamint konstansok egy sorozatát:

$$p_j^{(N+1)} = \frac{1}{S^{(N)}} p_j^{(N)} e^{D_j(T\mathcal{P}^{(N)})}, \qquad j = 1, \ldots, m,$$

$$S^{(N)} = \sum_{j=1}^{m} p_j^{(N)} e^{D_j(T\mathcal{P}^{(N)})},$$

$$N = 0, 1, 2, \ldots$$

Ekkor a

$$\mathcal{P}^{(0)}, \quad \mathcal{P}^{(1)}, \quad \mathcal{P}^{(2)}, \quad \ldots$$

eloszlások sorozata konvergál valamely optimális csatornabemeneti eloszláshoz és a

$$\log S^{(0)}, \quad \log S^{(1)}, \quad \log S^{(2)}, \quad \ldots$$

konstansok sorozata alulról konvergál a csatorna C kapacitásához.

Bizonyítás. Legyen $\mathcal{P}^*$ egy tetszőleges bemeneti eloszlás és

$$R_Q(\mathcal{P}^*) = \sum_{l=1}^{m} p_j^* D(Q_j||\mathcal{P}^* Q) = C(Q).$$

Arimoto eredeti bizonyítása alapján kapjuk a következőt:

$$\begin{aligned}
D(\mathcal{P}^*||\mathcal{P}^{(k)}) - D(\mathcal{P}^*||\mathcal{P}^{(k+1)}) &= \sum_{l=1}^{m} p_j^* \log_2 \frac{p_j^{(k+1)}}{p_j^{(k)}} \\
&= \sum_{l=1}^{m} p_j^* D(Q_j||\mathcal{P}^{(k)} Q) - \log_2 \Big(\sum_{j=1}^{m} p_j^{(k)} 2^{D(Q_j||\mathcal{P}^{(k)} Q)} \Big) \\
&= D(\mathcal{P}^* Q||\mathcal{P}^{(k)} Q) + R_Q(\mathcal{P}^*) - \log_2 \Big(\sum_{j=1}^{m} p_j^{(k)} 2^{D(Q_j||\mathcal{P}^{(k)} Q)} \Big) \\
&\geq C(Q) - \log_2 \Big(\sum_{j=1}^{m} p_j^{(k)} 2^{D(Q_j||\mathcal{P}^{(k)} Q)} \Big) \\
&\geq C(Q) - R_Q(\mathcal{P}^{(k+1)}) \\
&\geq 0 \qquad k = 0, 1, 2, \ldots
\end{aligned}$$

Ebből következik, hogy

$$\sum_{k=1}^{m}[C(Q) - R_Q(\mathcal{P}^{(k)})] \leq D(\mathcal{P}^*||\mathcal{P}^{(0)}) \leq +\infty,$$

mert $C(Q) \geq R_Q(\mathcal{P}^{(k)})$ bármely $k \geq 0$-ra, és $\lim_{k\to\infty}[C(Q) - R_Q(\mathcal{P}^{(k)})] = 0$, ugyanis

$$\lim_{k\to\infty} R_Q(\mathcal{P}^{(k)}) = C(Q).$$

Ezenfelül az is látszik, hogy

$$D(\mathcal{P}^*||\mathcal{P}^{(k)}) \geq D(\mathcal{P}^*||\mathcal{P}^{(k+1)}) \geq 0 \qquad k = 0, 1, 2, \ldots \quad .$$

Ennek következtében fennáll a következő határérték

$$\lim_{k\to\infty} D(\mathcal{P}^*||\mathcal{P}^{(k)}) = \alpha^* \in R_+, \qquad \forall \mathcal{P}^* \in P'.$$

Legyen $\mathcal{P}^{(i_k)} i_k \geq k$ egy részsorozata a $\mathcal{P}^{(k)}$ sorozatnak $\forall k \geq 0$, melyre igaz a következő:

$$\lim_{k\to\infty} \mathcal{P}^{(i_k)} = \widehat{\mathcal{P}}.$$

Az $R_Q(p)$ függvény folytonosságából következik, hogy

$$R_Q(\widehat{\mathcal{P}}) = \lim_{k\to\infty} R_Q(\mathcal{P}^{(i_k)}) = C(Q).$$

A korábbi egyenlőségekből beláthatjuk, hogy

$$\lim_{k\to\infty} D(\widehat{\mathcal{P}}||\mathcal{P}^{(k)}) = \widehat{\alpha} \in R_+.$$

Az előző egyenlőségek alapján

$$\widehat{\alpha} = \lim_{k\to\infty} D(\widehat{\mathcal{P}}||\mathcal{P}^{(i_k)}) = D(\widehat{\mathcal{P}}||\widehat{\mathcal{P}}) = 0.$$

Ennélfogva azt kapjuk, hogy $D(\widehat{\mathcal{P}}||\mathcal{P}^{(k)}) \to 0$, továbbá $\mathcal{P}^{(k)} \to \widehat{\mathcal{P}}$, ha $k \to \infty$.

■

Nézzünk pár egyszerű példát a csatornakapacitás meghatározására diszkrét, emlékezetnélküli csatorna esetén. A következő példák megtalálhatóak a [3], illetve a [18] irodalomban.

5.3. Példa. Legyen a csatorna adókarakterisztika mátrixa a következő:

$$T = \begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} & 0 & 0 & 1 \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} & 0 & 1 & 0 \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} & 1 & 0 & 0 \end{bmatrix}.$$

Ekkor $C(Q) = \log_2 3 \approx 1.5849625$ és az optimális bemeneti eloszlás pedig

$$p = \{0, 0, 0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}.$$

Az algoritmus pedig a következő eredményeket szolgáltatja 10^{-11} pontossággal:

$$C(Q) = 1.5849624079, \qquad p = \begin{bmatrix} 0.0000000161 \\ 0.0000000143 \\ 0.0000000064 \\ 0.3333333247 \\ 0.3333333219 \\ 0.3333333166 \end{bmatrix}$$

Az eredmények előállításához 17 iterációra volt szükség.

▲

5.4. Példa. Legyen a csatorna adókarakterisztika mátrixa

$$T = \begin{bmatrix} 0.1 & 0.3 & 0.4 & 0.2 \\ 0.3 & 0.4 & 0.2 & 0.1 \\ 0.4 & 0.2 & 0.1 & 0.3 \\ 0.2 & 0.1 & 0.3 & 0.4 \end{bmatrix}.$$

Ez egy gyengén szimmetrikus, diszkrét, emlékezetnélküli csatorna. A csatorna kapacitása

$$C(Q) = \log_2 4 - H(0.1, 0.2, 0.3, 0.4) = 2 + \frac{3}{10}\log_2 3 - \log_2 5 \approx 0.15356065.$$

Az optimális bemeneti eloszlásra pedig igaz a következő:

$$P = \{\mathcal{P} = (p_1, p_2, p_3, p_4) | p_1 + p_2 = 0.5, p_3 = p_1, p_2 = p_4\}.$$

Az Arimoto-Blahut algoritmussal kapott eredmények:

$$C(Q) = 0.1535606553, \qquad p^* = \begin{bmatrix} 0.3101020755 \\ 0.1899090077 \\ 0.3101020269 \\ 0.1898868898 \end{bmatrix}.$$

Az eredmények előállításához 116 iterációra volt szükség.

▲

5.5. Példa. Legyen a csatorna adókarakterisztika mátrixa

$$T = \begin{bmatrix} 0.04 & 0.11 & 0.05 & 0.05 & 0.23 & 0.07 & 0.03 & 0.02 & 0.05 & 0.16 \\ 0.06 & 0.21 & 0.13 & 0.19 & 0.04 & 0.18 & 0.23 & 0.24 & 0.14 & 0.33 \\ 0.26 & 0.04 & 0.20 & 0.22 & 0.17 & 0.39 & 0.17 & 0.15 & 0.37 & 0.03 \\ 0.40 & 0.40 & 0.36 & 0.42 & 0.08 & 0.07 & 0.50 & 0.33 & 0.32 & 0.17 \\ 0.24 & 0.24 & 0.26 & 0.12 & 0.48 & 0.29 & 0.07 & 0.26 & 0.12 & 0.31 \end{bmatrix}.$$

A csatorna kapacitása

$$C(Q) = 0.36217799,$$

az optimális bemeneti eloszlás

$$p = \{0, 0, 0, 0, 0.4954, 0, 0.5045, 0, 0, 0\}.$$

Az Arimoto-Blahut algoritmussal kapott eredmények pedig a következőek:

$$C(Q) = 0.3621779913,$$

$$p^* = \{0, 0, 0, 0, 0.4954129680, 0.0000025254, 0.5045845066, 0, 0, 0\}.$$

Az eredmények előállításához 3739 iterációra volt szükség.

▲

5.6. Példa. Legyen a csatorna adókarakterisztika mátrixa

$$T = \begin{bmatrix} 0 & \beta & 1-\beta \\ 0 & 1-\beta & \beta \\ 1 & 0 & 0 \end{bmatrix}.$$

A csatorna kapacitása

$$C(Q) = \log\left(1 + 2^{1-H(\beta,1-\beta)}\right).$$

Ha $\beta = 0.5$ akkor

$$\begin{aligned} C(Q) &= \log(1 + 2^{1-H(0.5,0.5)}) \\ &= \log(1 + 2^{1+(0.5\log 0.5 + 0.5\log 0.5)}) \\ &= \log(1 + 2^{1+(-1)}) \\ &= \log(1 + 2^0) = \log 2 = 1. \end{aligned}$$

Az Arimoto-Blahut algoritmussal kapott eredmények pedig a következőek:

$$C(Q) = 1,$$

$$p^* = \begin{bmatrix} 0.5 \\ 0.125 \\ 0.375 \end{bmatrix}.$$

Az eredmények előállításához 1 iterációra volt szükség.

▲

5.7. Példa.

$$T = \begin{bmatrix} \frac{p}{2} & \frac{1-p}{2} \\ \frac{p}{2} & \frac{1-p}{2} \\ \frac{1-p}{2} & \frac{p}{2} \\ \frac{1-p}{2} & \frac{p}{2} \end{bmatrix}.$$

A csatorna kapacitása $C(Q) = 1 - H(p, 1-p)$. Ha $p = 0.5$, akkor $C(Q) = 0$. Legyen $p = 0.2$, ekkor

$$\begin{aligned} C(Q) &= 1 - [-(0.2\log 0.2 + 0.8\log 0.8)] \\ &= 1 + (-0.464385 - 0.257542) \\ &= 1 - 0.721928 = 0,278071 \end{aligned}$$

Az Arimoto-Blahut algoritmussal kapott eredmények pedig a következőek:

$$C(Q) = 0.2780719051$$

és

$$p^* = [0.5, 0.5].$$

Az eredmények előállításához 1 iterációra volt szükség.

▲

5. 5.5. Iterációs módszer a relatív kapacitás meghatározására (kiegészítő tananyag)

A szakasz kitekintést ad arra, hogyan lehetne az Arimoto-Blahut algoritmust általánosítani, ha zajos csatorna esetén a jelek különböző idő alatt mennek át, azaz additív költség esetén [14].

5.13. Definíció. *Relatív kapacitásnak nevezzük a csatornán ténylegesen átvitt információ és a forrásentrópia hányadosát:*

$$C_R = \frac{I(\mathcal{P}, \mathcal{Q})}{H(\mathcal{P})}.$$

Legyen a P és P' bemeneti eloszlásvektorok halmaza a következőképpen definiálva:

$$P = \left\{ \mathcal{P} = (p_1, \ldots, p_n); p_i > 0 \quad (i = 1, \ldots, n), \sum_{i=1}^{n} p_i = 1 \right\}$$

és

$$P' = \left\{ \mathcal{P} = (p_1, \ldots, p_n); p_i \geq 0 (i = 1, \ldots, n), \sum_{i=1}^{n} p_i = 1 \right\},$$

valamint legyen

$$T = (t_{ij}) \qquad (i = 1, \ldots, m; j = 1, \ldots, n);$$

$$t_{ij} \geq 0, \qquad \sum_{i=1}^{n} t_{ij} = 1$$

az adókarakterisztika mátrix.

Ekkor a csatornán átvitt információmennyiséget a következőképpen számolhatjuk ki:

$$I(\mathcal{P}) = \sum_{i=1}^{m} \sum_{j=1}^{n} p_i t_{ij} \log \frac{t_{ij}}{\sum_{i=1}^{m} t_{ij}}, \qquad p \in P'.$$

Az egyszerűség kedvéért legyen

$$D_i(\mathcal{P}) = \sum_{j=1}^{n} t_{ij} \log \frac{t_{ij}}{\sum_{i=1}^{m} p_i t_{ij}} = \sum_{j=1}^{n} t_{ij} \log \frac{t_{ij}}{q_i},$$

ahol $\mathcal{Q} = (q_1, \ldots, q_n)$ kimeneti eloszlás, mely TP mátrix-vektor szorzással meghatározható.

Az jelölésbeni egyszerűsítés után az átvitt információmennyiség:

$$I(\mathcal{P}) = \sum_{i=1}^{m} p_i D_i(\mathcal{P}).$$

Ezek után vezessünk be egy $l_i > 0$ költségváltozót minden i-dik bemeneti szimbólumhoz, valamint legyen $l(\mathcal{P})$ költségfüggvény

$$l(\mathcal{P}) = \sum_{i=1}^{m} l_i p_i, \qquad p \in P'.$$

A relatív kapacitást a következőképpen definiálta Reza(1961):

$$C_R = \max_{p \in P'} \frac{I(\mathcal{P})}{l(\mathcal{P})}.$$

Az iterációs algoritmus a következő:

1. Legyen $a = \min_i l_i$.

2. Legyen $\mathcal{P}^{(0)} \in P$ tetszőleges bemeneti eloszlás.

3. A $\mathcal{P}^{(k)}$ valószínűségi vektor után határozzuk meg a $\mathcal{P}^{(k+1)}$ vektort a következőképpen

$$p_i^{(\acute{}k+1)} = p_i^{(k)} \exp\left[\frac{aD_i(\mathcal{P}^{(k)})}{l_i}\right], \qquad i = 1, \ldots, m;$$

$$w^{(k)} = \sum_{i=1}^{m} p_i^{(\acute{}k+1)};$$

$$p_i^{(k+1)} = \frac{p_i^{(\acute{}k+1)}}{w^{(k)}}, \qquad i = 1, \ldots, m.$$

5.14. Tétel. *Az*

$$\{\frac{I(\mathcal{P}^{(k)})}{l(\mathcal{P}^{(k)})}\}$$

sorozat monoton növekvő és konvergál a relatív kapacitáshoz.

Bizonyítás. Először is azt mutatjuk meg, hogy az eljárás által kapott

$$\{\frac{I(\mathcal{P}^{(k)})}{l(\mathcal{P}^{(k)})}\}$$

sorozat monoton növekvő. Jelöljük $\mathcal{P}$-vel és $\mathcal{R}$-rel a k-adik és a $(k+1)$-edik valószínűségi vektorokat, melyeket az előbbi iterációs formulával kapunk. Legyen

$$x_i = x_i(\mathcal{P}) = D_i(\mathcal{P})/l_i,$$
$$f_i = \exp[ax_i], \qquad (i = 1, \ldots, n)$$
$$w = \sum_{i=1}^{n} p_i f_i.$$

Ekkor nyilvánvalóan

$$r_i = \frac{p_i f_i}{w}.$$

Legyen $\mathcal{U} = (u_1, \ldots, u_n)$, ahol:

$$u_i = \frac{p_i l_i}{l(\mathcal{P})}.$$

■

A következő lemma garantálja, hogy az eljárás fenti sorozata monoton növekvő lesz.

5.15. Lemma.

$$\frac{I(\mathcal{P})}{l(\mathcal{P})} \leq \frac{\log \lambda}{a} \leq \frac{I(\mathcal{R})}{l(\mathcal{R})},$$

ahol

$$\lambda = \sum_{i=1}^{n} u_i f_i.$$

Egyenlőség akkor és csak akkor áll fenn mindkét oldalon, ha x_i konstans bármely $p_i > 0$ -ra.

Bizonyítás. A Jensen-egyenlőtlenséget alkalmazva kapjuk, hogy

$$\frac{\log \lambda}{a} \geq \frac{1}{a} \sum_{i=1}^{n} u_i \log f_i = \frac{I(\mathcal{P})}{l(\mathcal{P})}.$$

A második egyenlőtlenséghez elegendő belátni a következőt:

$$Q = aI(\mathcal{R}) - l(\mathcal{R}) \log \lambda \geq 0.$$

Legyen $\mathcal{S} = (s_1, \ldots, s_m)$ kimeneti valószínűségi vektor úgy, hogy

$$s_j = \sum_{i=1}^{n} r_i t_{ij} \qquad (j = 1, \ldots, m).$$

Az I-divergencia tulajdonságából könnyen beláthatjuk a következő egyenlőtlenséget:

$$D(\mathcal{R}||\mathcal{P}) \geq D(\mathcal{S}||\mathcal{Q}).$$

Figyelembe véve ezt az egyenlőtlenséget kapjuk a következőt:

$$\begin{aligned} I(\mathcal{R}) &= \sum_{i=1}^{m} r_i \sum_{j=1}^{n} t_{ij} \log \frac{t_{ij}}{s_j} \\ &= \sum_{i=1}^{m} r_i \sum_{j=1}^{n} t_{ij} \log \frac{t_{ij}}{q_j} - D(\mathcal{S}||\mathcal{Q}) \\ &\geq \sum_{i=1}^{m} r_i D_i(\mathcal{P}) - D(\mathcal{R}||\mathcal{P}) \\ &= \frac{1}{a} \sum_{i=1}^{m} r_i l_i \log f_i - D(\mathcal{R}||\mathcal{P}). \end{aligned}$$

Másfelől

$$\lambda = \sum_{i=1}^{m} u_i f_i = \frac{w l(\mathcal{R})}{l(\mathcal{P})}.$$

Ennélfogva

$$\begin{aligned} Q &\geq \sum_{i=1}^{m} r_i l_i \log f_i - aD(\mathcal{R}||\mathcal{P}) - l(\mathcal{R}) \log \lambda \\ &= \sum_{i=1}^{m} r_i l_i \log \frac{f_i}{w} - aD(\mathcal{R}||\mathcal{P}) - l(\mathcal{R}) \log \frac{l(\mathcal{R})}{l(\mathcal{P})} \\ &= \sum_{i=1}^{m} r_i (l_i - a) \log \frac{r_i}{p_i} - l(\mathcal{R}) \log \frac{l(\mathcal{R})}{l(\mathcal{P})} \\ &\geq (l(\mathcal{R}) - a) \log \frac{l(\mathcal{R}) - a}{l(\mathcal{P}) - a} l(\mathcal{R}) \log \frac{l(\mathcal{R})}{l(\mathcal{P})}. \end{aligned}$$

A lemmából azonnal következik, hogy ha $z \geq a > 0$ és $b \geq a$, akkor az

$$f(z) = (z - a) \log \frac{z - a}{b - a} - z \log \frac{z}{b}$$

függvény a $z = b$ esetben veszi fel a 0 értéket, és pozitív ha $z \neq b$. Könnyű belátni, hogy az egyenlőség feltétele, hogy x_i konstans legyen bármely $p_i > 0$-ra.

Most már készen vagyunk, hogy bebizonyítsuk, az iterációs eljárás által létrehozott

$$\{\frac{I(\mathcal{P}^{(k)})}{l(\mathcal{P}^{(k)})}\}$$

sorozat a relatív kapacitáshoz tart.

Legyen

$$x_i^{(k)} = \frac{D_i(\mathcal{P}^{(k)})}{l_i},$$
$$f_i^{(k)} = \exp[ax_i^{(k)}],$$
$$u_i^{(k)} = \frac{p_i^{(k)} l_i}{l(\mathcal{P}^{(k)})}, \quad (i = 1, \dots, m, k = 0, 1, \dots),$$
$$w^{(k)} = \sum_{i=1}^{m} p_i^{(k)} f_i^{(k)},$$
$$\lambda^{(k)} = \frac{\sum_{i=1}^{m} p_i^{(k)} l_i f_i^{(k)}}{l(\mathcal{P}^{(k)})} = \sum_{i=1}^{m} u_i^{(k)} f_i^{(k)}.$$

Mivel

$$p_i^{(k+1)} = \frac{p_i^{(k)} f_i^{(k)}}{w^{(k)}},$$

kapjuk, hogy

$$\frac{I(\mathcal{P}^{(k)})}{l(\mathcal{P}^{(k)})} \leq \frac{log\lambda^{(k)}}{a} \leq \frac{I(\mathcal{P}^{(k+1)})}{l(\mathcal{P}^{(k+1)})} \leq C_R, \quad (k = 0, 1, 2, \dots).$$

Ennélfogva a

$$\{\frac{I(\mathcal{P}^{(k)})}{l(\mathcal{P}^{(k)})}\}$$

és a

$$\{\frac{log\lambda^{(k)}}{a}\}$$

sorozatok monoton növekvőek és ugyanahhoz az értékhez tartanak. Vezessünk be egy $\mathcal{P}^*$ valószínűségi vektort mely elérte a relatív kapacitást és legyen

$$u_i^* = \frac{p_i^* l_i}{l(\mathcal{P}^*)}.$$

Ezenfelül vezessünk be egy

$$\mathcal{Q}^{(k)} \qquad (k = 0, 1, 2, \dots)$$

és $\mathcal{Q}^*$ kimeneti valószínűségi vektorokat. Ha figyelembe vesszük, hogy

$$\frac{u_i^{(k+1)}}{u_i^{(k)}} = \frac{f_i^{(k)}}{\lambda^{(k)}},$$

akkor a következőhöz jutunk

$$\begin{aligned} D(\mathcal{U}^*||\mathcal{U}^{(k)}) - D(\mathcal{U}^*||\mathcal{U}^{(k+1)}) &= \sum_{i=1}^{m} u_i^* \log \frac{u_i^{(k+1)}}{u_i^{(k)}} = a \sum_{i=1}^{m} u_i^* x_i^{(k)} - \log \lambda^{(k)} \\ &= \frac{a}{l(\mathcal{P}^*)} \Big(\sum_{i=1}^{m} p_i^* D_i(\mathcal{P}^*) + D(\mathcal{Q}^*||\mathcal{Q}^{(k)}) \Big) - \log \lambda^{(k)} \\ &= aC_R - \log \lambda^{(k)} + \frac{a}{l(\mathcal{P}^*)} D(\mathcal{Q}^*||\mathcal{Q}^{(k)}). \end{aligned}$$

Az egyenlőség jobboldalán lévő kifejezés nemnegatív, ezért

$$\frac{1}{a}\{D(\mathcal{U}^*||\mathcal{U}^{(k)}) - D(\mathcal{U}^*||\mathcal{U}^{(k+1)})\} \geq C_R - \frac{\log\lambda^{(k)}}{a} (\geq 0), \quad k = 0, 1, 2, \ldots.$$

Összeadva ezeket az egyenlőtlenségeket $k = 0$-tól $k = N - 1$-ig, kapjuk, hogy

$$\sum_{k=0}^{N-1} \left(C_R - \log\frac{\lambda^{(k)}}{a}\right) \leq \frac{1}{a}\{D(\mathcal{U}^*||\mathcal{U}^{(0)}) - D(\mathcal{U}^*||\mathcal{U}^{(N)})\}$$
$$\leq \frac{1}{a}\{D(\mathcal{U}^*||\mathcal{U}^{(0)})\}, \qquad N = 0, 1, 2, \ldots.$$

Az egyenlőtlenség jobb oldalán lévő kifejezés független N-től, és véges, a

$$\{\frac{\log\lambda^{(k)}}{a}\}$$

sorozat tart a C_R relatív kapacitáshoz. Ezzel bizonyítottuk az iterációs eljárás konvergenciáját.

■

5.16. Következmény. *A*

$$\{\mathcal{Q}^{(k)}\}$$

kimeneti eloszlások sorozata, hasonlóan az

$$\{\mathcal{U}^{(k)}\}$$

bemeneti eloszlások sorozatához, konvergens.

Bizonyítás. Az tételből kapjuk a

$$0 \leq \frac{a}{l(\mathcal{P}^*)} D(\mathcal{Q}^*||\mathcal{Q}^{(k)}) \leq D(\mathcal{U}^*||\mathcal{U}^{(k)}) - D(\mathcal{U}^*||\mathcal{U}^{(k+1)}), \qquad k = 0, 1, 2, \ldots.$$

egyenlőtlenséget. Összeadva ezeket az egyenlőtlenségeket $k = 0$-tól $k = N - 1$-ig, úgy, mint a korábbi egyenlőtlenség levezetésében, könnyen beláthatjuk a következmény helyességét.

■

A következő lemma a kovergencia sebességének kimondásához szükséges.

5.17. Lemma. *A közelítés*

$$e(k) = C_R - \frac{\log\lambda^{(k)}}{a}$$

hibája egy

$$\frac{const}{ka}$$

kifejezéssel határolható.

Bizonyítás. Az előzőek felhasználásával jutunk el a következő kifejezésig

$$e(k) \leq \frac{1}{ka} D(\mathcal{U}^*||\mathcal{U}^{(0)}) = \frac{const}{ka}.$$

■

A konvergencia sebessége:

Abban az esetben ha a $\mathcal{P}^*$ bemeneti eloszlás eléri a relatív kapacitást, akkor ez egyedi és $\mathcal{P}^* \in P$. Ekkor a konvergencia sebessége meglehetősen javul.

A következőkben szükségünk lesz az alábbi lemmára, mely könnyen levezethető a Kuhn-Tucker tételből.

5.18. Lemma. *Ha a $\mathcal{P}^*$ valószínűségi vektor eléri a relatív kapacitást, akkor*

$$\begin{aligned} \frac{D_i(\mathcal{P}^*)}{l_i} &= C_R \qquad \forall p_i^* > 0, \\ &\leq C_R \qquad \forall p_i^* = 0. \end{aligned}$$

5.19. Tétel. *Ha a bemeneti $\mathcal{P}^*$ valószínűségi vektor eléri a relatív kapacitást, akkor az egyedi és $\mathcal{P}^* \in P$. Ekkor létezik olyan pozitív egész N és egy konstans $0 < \sigma \leq 1$, amik kielégitik a következő egyenlőtlenséget*

$$e(k) \leq \frac{\sigma}{a}(1-\sigma)^{k-N} D(\mathcal{U}^* || \mathcal{U}^{(N)}) \qquad \forall k \geq N$$

és N független σ-tól.

Bizonyítás. Legyen $\mathcal{D} = \mathcal{P}^* - \mathcal{P}^k$, ekkor $\mathcal{D}$ tart a 0-hoz, ha $k \to \infty$.

Továbbá legyen

$$c_i = \frac{l_i}{l(\mathcal{P}^*)} = \frac{u_i^*}{p_i^*}.$$

Ekkor a következőkhöz jutunk:

$$\begin{aligned} D(\mathcal{U}^* || \mathcal{U}^{(k)}) &= -\sum_{i=1}^{m} u_i^* \log \frac{p_i^{(k)}}{p_i^*} + \log \frac{l(\mathcal{P}^{(k)})}{l(\mathcal{P}^*)} \\ &= -\sum_{i=1}^{m} u_i^* \log \left(1 - \frac{d_i}{p_i^*}\right) + \log \left(1 - \sum_{i}^{m} d_i c_i\right) \\ &= \frac{1}{2}\Big[\sum_{i=1}^{m} \frac{c_i}{p_i^*} d_i^2 - \Big(\sum_{i=1}^{m} d_i c_i\Big)^2\Big] + \sum_{i=1}^{m} O(d_i^3) \\ &= \frac{1}{2} d A d^T + \sum_{i=1}^{m} O(d_i^3), \end{aligned}$$

ahol $A = (a_{ij})$ egy $n \times n$ méretű szimmetrikus mátrix, úgy, hogy

$$\begin{aligned} a_{ij} &= \frac{c_i}{p_i^*} - c_i^2, \qquad \text{ha } i = j, \\ &= -c_i c_j, \qquad \text{ha } i \neq j. \end{aligned}$$

Eléggé nagy N esetén

$$D(\mathcal{U}^* || \mathcal{U}^{(k)}) \leq (1-\sigma)^{k-N} D(\mathcal{U}^* || \mathcal{U}^{(k)}), \qquad \forall k \geq N.$$

Az Arimoto-Blahut algoritmus bizonyításához hasonló érveléssel, kombinálva az egyenlőtlenségeket, bizonyítani tudjuk a tételt.

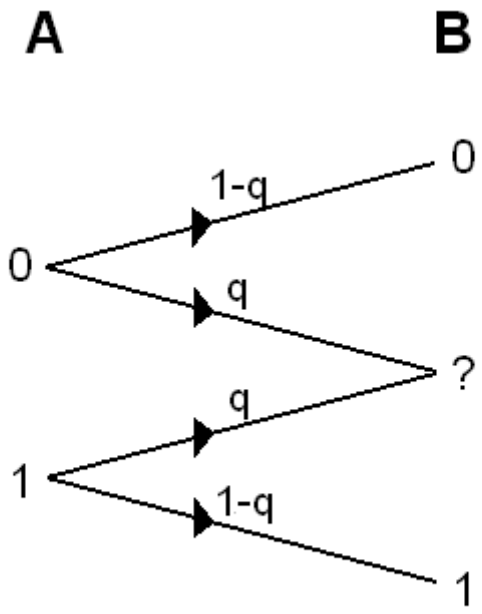
■

6. 5.6. Feladatok

1. Az $Y = \{0, 1, 2, 3\}$ csatornaábécéhez tartozó átviteli idők $T = \{1, 1, 1, 3\}$. Határozza meg a csatornakapacitást! Optimális kódolás esetén határozza meg az átlagos átviteli időt!

2. Bináris törlődéses csatornáról a következőt tudjuk: egy elküldött jel 0.9 valószínűséggel marad az eredeti és 0.1 valószínűséggel válik felismerhetetlenné. Határozza meg a csatornakapacitást!

5.3. ábra - Bináris törlődéses csatorna



3. Az

$Y = \{0, 1, 2, 3\}$

csatornaábécéhez tartozó átviteli idők

$T = \{0.5, 1.2, 1.3, 2.1\}.$

Határozza meg a csatornakapacitást! Optimális kódolás esetén határozza meg az átlagos átviteli időt!

4. Bináris csatorna esetén $p_{0|0} = 0.95, \quad p_{1|1} = 0.85, \quad p_{0|1} = 0.05, \quad p_{1|0} = 0.15.$ Határozza meg a csatornakapacitást!

5. Az $Y = \{0, 1, 2\}$ csatornaábécéhez tartozó átviteli idők $T = \{1, 1.2, 2.3\}.$ Határozza meg a csatornakapacitást! Optimális kódolás esetén határozza meg az átlagos átviteli időt!

6. Az

$Y = \{0, 1, 2, 3\}$

csatornaábécéhez tartozó átviteli idők

$T = \{0.5, 1.2, 1.3, 2.1\}.$

Határozza meg a csatornakapacitást! Optimális kódolás esetén határozza meg az átlagos átviteli időt!

7. A csatornamátrix

$$T = \begin{pmatrix} 0.9 & 0 \\ 0.1 & 0.1 \\ 0 & 0.9 \end{pmatrix}.$$

Határozza meg a csatornakapacitást!

8. A csatornamátrix

$$T = \begin{pmatrix} 0.1 & 0.4 \\ 0.4 & 0.1 \\ 0.4 & 0.1 \\ 0.1 & 0.4 \end{pmatrix}.$$

Határozza meg a csatornakapacitást!

9. Az $Y = \{0, 1, 2, 3\}$ csatornaábécéhez tartozó átviteli idők $T = \{0.6, 1, 1.2, 3\}$. Határozza meg a csatornakapacitást! Optimális kódolás esetén határozza meg az átlagos átviteli időt (Pontosság=0.00001)!

10. Az $Y = \{0, 1, 2, 3, 4\}$ csatornaábécéhez tartozó átviteli idők

$T = \{0.5, 0.8, 1.2, 1.3, 2.1\}$.

Határozza meg a csatornakapacitást! Optimális kódolás esetén határozza meg az átlagos átviteli időt!

11. Bináris csatorna esetén $p_{0|0} = 0.97, \quad p_{1|1} = 0.94, \quad p_{0|1} = 0.06, \quad p_{1|0} = 0.03.$ Határozza meg a csatornakapacitást!

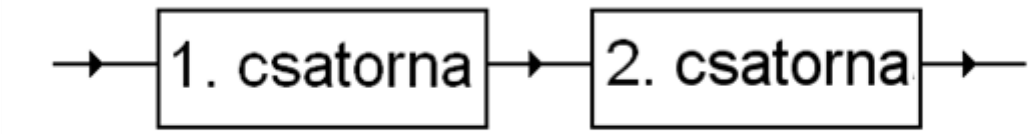
12. Bináris csatorna esetén $p_{0|0} = 0.9, p_{1|1} = 0.8, p_{0|1} = 0.1, p_{1|0} = 0.2.$ Határozza meg a csatornakapacitást!

13. Bináris szimmetrikus csatorna esetén $p_{0|0} = p_{1|1} = p$ és $p_{0|1} = p_{1|0} = q$. Határozza meg a csatornakapacitást, ha $p = 0.9$ és $q = 0.1$!

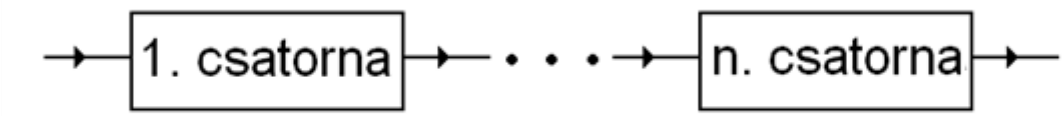
14. Az $Y = \{0, 1, 2, 3\}$ csatornaábécéhez tartozó átviteli idők $T = \{1, 1, 2, 3\}$. Határozza meg a csatornakapacitást! Optimális kódolás esetén határozza meg az átlagos átviteli időt!

15. Sorba kötött csatornák esetén határozza meg a csatornakapacitást!

5.4. ábra - Egymás után két csatorna (soros eset)

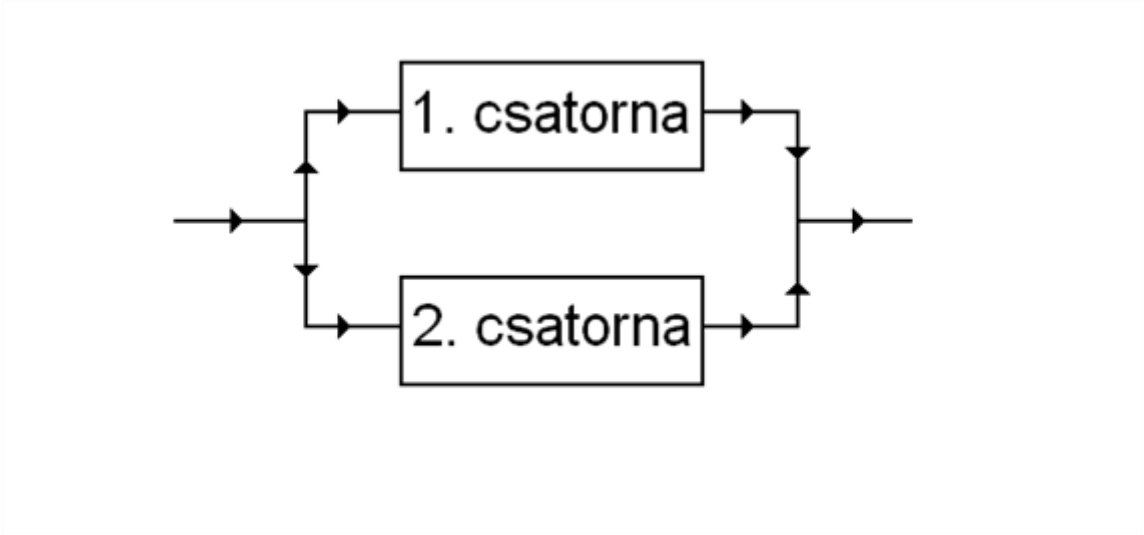


5.5. ábra - Egymás után több csatorna (soros eset)



16. Párhuzamos csatornák esetén határozza meg a csatornakapacitást!

5.6. ábra - Egymás mellett két csatorna (párhuzamos eset)



7. 5.7. Önellenőrző kérdések

1. Definiálja a feltételes entrópiát!

2. Definiálja a feltételes entrópiát, ha adott az együttes eloszlás!

3. Definiálja a kölcsönös információmennyiséget!

4. Definiálja a csatornakapacitást azonos átviteli idő esetén!

5. Vezesse le a szimmetrikus csatorna kapacitását!

6. Vezesse le a bináris törlődéses csatorna kapacitását!

7. Definiálja a csatornakapacitást nem azonos átviteli idő esetén!

8. Definiálja a csatornakapacitást additív költség esetén!

9. Definiálja az információ átviteli sebességét!

10. Bizonyítsa az adatátviteli lemmát!

11. Ismertesse és jellemezze a tanult csatorna típusokat!

12. Bizonyítsa a csatornakapacitás kiszámítására használt numerikus módszer konvergenciáját additív költség esetén!

13. Ismertesse az Arimoto-Blahut algoritmust!

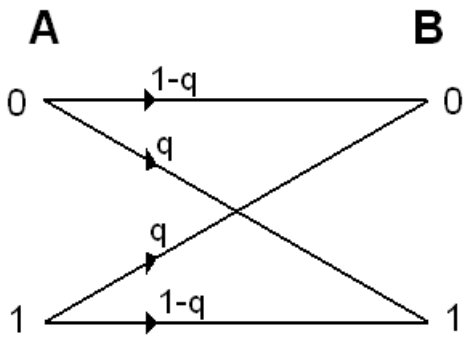
6. fejezet - Csatornakódolás

1. 6.1. Hibajavítás, kódtávolság

Szimmetrikus bináris csatorna esete, azaz a csatorna átviteli mátrixa legyen a következő:

$$T = \begin{pmatrix} p & q \\ q & p \end{pmatrix}.$$

6.1. ábra - Bináris szimmetrikus csatorna



Probléma: Milyen feltételek mellett és hogyan oldható meg a csatornában az átvitelnél keletkezett hibák jelzése és javítása?

6.1. Példa. A bit háromszorozás módszere (Commodore-64 kazettás egység):

$0 \to 000$

$1 \to 111$

Ha a dekódolás a több azonos bit szerint történik, akkor a legfeljebb két hiba jelezhető és egy hiba javítható.

Ha $p = 0.9$, akkor a helyes átvitel (javítással) valószínűsége

$$p^3 + 3p^2q = 0.972.$$

▲

6.1. Definíció. *Az üzenetszó (k bit) $\to$ kódszó (n bit) átalakítást (kódolást) (k, n) kódnak nevezzük.*

6.2. Megjegyzés. *Ez egy blokkos kódolás.*

Legyen $A = \{0, 1\}$ és adott a következő két művelet:

a □kizáró vagy" művelet (jele: $\vee$) vagy másképpen a modulo 2 összeadás (jele: $\oplus$), és a hagyományos szorzás a másik művelet.

Ekkor $(A, \vee)$ és $(A, \cdot)$ Abel-csoport. Továbbá, $(A, \vee, \cdot)$ test. Értelmezzük az A^n esetén az előbbi műveleteket bitenként, ekkor $(A^n, \vee)$ vektortér az $(A, \vee, \cdot)$ test felett.

6.3. Definíció. *Legyen $a \in A^n$. $||a||$ jelentse az egyes bitek számát.*

Ekkor $||\cdot|| : A^n \to \mathbf{Z}$ norma.

6.4. Definíció. *A* $d(a,b) = ||a \vee b||$ *mennyiséget Hamming-féle távolságnak nevezzük.*

6.5. Lemma. *A Hamming-féle távolság kielégíti a távolság tulajdonságait.*

6.6. Lemma. *A Hamming-féle távolság invariáns az eltolásra, azaz*

$$d(a,b) = d(a \vee c, b \vee c).$$

Bizonyítás.

$$d(a \vee c, b \vee c) = ||(a \vee c) \vee (b \vee c)|| = ||a \vee (c \vee c) \vee b|| = ||a \vee b|| = \\ = d(a,b).$$

■

6.7. Definíció. *A* $v_1, v_2, \ldots, v_m$ *kódszavakból álló kód esetén a kódszavak távolságai közül a minimálisat kódtávolságnak nevezzük, azaz a* d *kódtávolságra*

$$d = \min_{i \neq j} d(v_i, v_j).$$

6.8. Megjegyzés. *Legyen* $d = \min_{i \neq j} d(v_i, v_j)$, *a csatornaábécé* $Y = \{0,1\}$. *Jelölések:* $u \in Y^k$ *(az eredeti üzenet),* $v \in Y^n$ *(az* u*-nak megfelelő csatorna kódszó),* $\tilde{v} \in Y^n$ *(a* v*-nek megfelelő csatornán áthaladt jelsorozat, azaz az átvitelnél keletkezik). Ekkor*

$$P(\tilde{v}|v) = q^{d(\tilde{v},v)} p^{n-d(\tilde{v},v)}.$$

Ha a $\tilde{v}$ *eredetijének azt a* v *kódszót tekintjük, amelyre a* $P(\tilde{v}|v)$ *feltételes valószínűség a lehető legnagyobb, azt maximum likelihood kódolásnak nevezzük.*

Tegyük fel, hogy $0 < q < 0.5$, akkor

$$P(\tilde{v}|v) = \left(\frac{q}{p}\right)^{d(\tilde{v},v)} p^n$$

maximális, ha $d(\tilde{v}, v)$ minimális.

Ez azt jelenti, hogy bináris szimmetrikus csatorna esetén a minimális távolságon alapuló dekódolás (javítás) megegyezik a maximum likelihood kódolás alapján történővel.

6.9. Tétel. *Legyen egy vett szóban a hibák száma legfeljebb* r. *Tetszőleges kódszó esetén a legfeljebb* r *számú hiba a minimális távolságon alapuló hibajavítás módszerével akkor és csak akkor javítható, ha a kódtávolság* $d \geq 2r+1$.

Bizonyítás. *Elégségesség:* Ha $d \geq 2r+1$ és $||e|| \leq r$ (e a hibavektor), akkor

$$d(\tilde{v}, v_i) < d(\tilde{v}, v_j)$$

bármely $i \neq j$ esetén, ha $\tilde{v} = v_i \vee e$.

$$d = \min_{i \neq j} d(v_i, v_j) \leq d(v_i, v_j) \leq d(v_i, \tilde{v}) + d(\tilde{v}, v_j) \leq r + d(\tilde{v}, v_j),$$

azaz

$$d(\tilde{v}, v_j) \geq d - r \geq (2r+1) - r = r + 1.$$

Szükségesség: Ha $||e|| \leq r$ és $\tilde{v} = v_i \vee e$ minimális távolságon alapuló dekódolása mindig helyes eredményre vezet, akkor $d \geq 2r+1$.

$$d(v_i, v_j) \leq d(v_i, \tilde{v}) + d(\tilde{v}, v_j) \Rightarrow d(\tilde{v}, v_j) \geq d - r,$$

azaz a v_i-ből torzult $\tilde{v}$ szó a v_j-től legalább $d-r$ távolságra van. Mivel azt akarjuk, hogy a dekódolás v_i-be történjen, ezért

$$d(\tilde{v}, v_i) < d - r \Rightarrow d > 2r, \text{azaz} d > 2r + 1.$$

■

2. 6.2. Csoportkód

6.10. Definíció. *Ha a kódszavak csoportot alkotnak, a kódot csoportkódnak nevezzük.*

6.2. Példa. Adottak a következő (2,5) kódok:

		A	B	C
00	$\rightarrow$	00000	00100	00000
01	$\rightarrow$	01011	01011	01011
10	$\rightarrow$	10101	10101	10111
11	$\rightarrow$	11110	11110	11110

Az A-val jelzett oszlop csoportkód, míg a másik kettő nem, hiszen a B oszlopban nincs zérusvektor és a C oszlop esetén

$$01011 \vee 10111 = 11100.$$

Természetesen a vektorok oszlopvektorok, de az egyszerűség kedvéért, ha nem félreérthető, akkor csak sorban és egymás mellé írt bitsorozat lesz a vektor.

▲

6.11. Tétel. *Csoportkódban a kódszó alakú hibavektor esetén a hiba nem jelezhető és nem javítható. A nem kódszó alakú hiba legalább jelezhető.*

6.12. Tétel. *Csoportkód esetén a hibaáteresztés valószínűsége megegyezik a csupa zérus kódszó alakú hibák valószínűségének az összegével.*

6.3. Példa. Az (A) csoportkód esetén, ha $p = 0.9$, akkor a hibaáteresztés valószínűsége: $2q^3p^2 + q^4p = 0.0171$, a kódtávolság: 3, a dekódolói hiba (2 vagy több hiba): 0.08146.

▲

6.13. Tétel. *Egy* (k, n) *csoportkód esetén* $d = \min_{v_i \neq 0} ||v_i||$.

6.14. Tétel. *G csoportkód, $v_i \in G$ rögzített, e hibavektor. Ha a hiba javítható, akkor ez a tulajdonsága független v_i-től.*

Bizonyítás. Ha e javítható, akkor

$$d(v_i \vee e, v_i) < d(v_i \vee e, v_j) \quad i \neq j.$$

Azt kell belátni, hogy

$$d(v_k \vee e, v_k) < d(v_k \vee e, v_j) \quad k \neq j.$$

A Hamming-távolság eltolásra invariáns, ezért

$$\begin{aligned} d(v_k \vee e, v_j) &= d((v_k \vee e) \vee (v_k \vee v_i), v_j \vee (v_k \vee v_i)) = \\ &= d(v_i \vee e, v_j \vee (v_k \vee v_i)) \geq d(v_i \vee e, v_i) = \\ &= d((v_i \vee e) \vee (v_k \vee v_i), v_i \vee (v_k \vee v_i)) = d(v_k \vee e, v_k). \end{aligned}$$

■

6.15. Megjegyzés. *Hogyan lehetne automatizálni a következő problémákat?*

1. A csoport tulajdonság ellenőrzése.

2. Tárolás, kódszó keresés.

3. Kódtávolság kiszámítás.

3. 6.3. Lineáris kód

6.16. Definíció. *Legyen* $u \in A^k$, G $k \times n$ *típusú mátrix, ahol* $g_{ij} \in A$. *A kód lineáris, ha*

$$v^T = u^T G.$$

A G *mátrixot generáló mátrixnak nevezzük. A* v *vektor transzponáltjának a jele* v^T.

6.4. Példa.

$$G = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

Éppen az (A) csoportkódot adja meg.

▲

6.17. Tétel. *A lineáris kód csoportkód.*

Bizonyítás.

$G : A^k \to A^n$,

azaz mivel $(A^k, \vee)$ csoport, így van zérusvektor. Nem vezet ki a művelet a halmazból és létezik inverz elem hiszen minden elem a saját inverze.

■

6.18. Definíció. *Legyen* E $k \times k$ *típusú egységmátrix. A* $G = (E|P)$ *generátor mátrixú kódot szisztematikus kódnak nevezzük.*

Néhány elnevezés:

P paritásmátrix,

$F = \begin{pmatrix} P \\ \overline{E} \end{pmatrix}$ paritásellenőrző mátrix ($E_{(n-k)\times(n-k)}$),

$u^T P$ paritásvektor.

6.19. Definíció. *Legyen* $\tilde{v}$ *egy vett kódszó a csatornakimeneten. Az* s *vektort a* $\tilde{v}$ *vektorhoz tartozó szindrómának nevezzük, ha*

$$s^T = \tilde{v}^T F.$$

6.20. Tétel. *A szindróma akkor és csak akkor zérusvektor, ha a vett szó kódszó.*

6.21. Megjegyzés. *A szindrómák egy osztályozást adnak.*

6.22. Tétel. *A csatorna kimenetén vett azonos mellékosztályokba tartozó szavak szindrómája azonos, különböző mellékosztályokhoz tartozóké különböző.*

(k,n) szisztematikus kód esetén: $(A^k,\vee)$ csoport, a generálás után $(S,\vee)$ részcsoport $(S\subset A^n)$, S meghatároz egy mellékosztályra bontást. Készítsük el a mellékosztálytáblázatot, majd ebből a dekódolási táblázatot, azaz minden mellékosztályban kiválasztjuk a mimimális normájú osztályelemet. Ezzel az osztályelemmel generáljuk a mellékosztályt.

6.5. Példa. A

$$G=\begin{pmatrix}1&0&1&0&1\\0&1&0&1&1\end{pmatrix}$$

generátormátrixhoz készítsük el a mellékosztály táblázatot!

A következő mellékosztálytáblázat első sorában van a generált csoportkódunk, s a további sorokban egy-egy mellékosztály:

00000	01011	10101	11110
00001	01010	10100	11111
00010	01001	10111	11100
00100	01111	10001	11010
01000	00011	11101	10110
00101	01110	10000	11001
11001	10010	01100	00111
11000	10011	01101	00110

Egy mellékosztálytáblázat akkor jó dekódolási táblázatnak, ha minden sorban a legkisebb normájú elem az első. Ezek az ún. osztályelsők. Jól látható, hogy ez nem teljesül a 6. és a 7. sorban, így ezeket a sorokat újraszámoljuk.

Dekódolási táblázat:

00000	01011	10101	11110
00001	01010	10100	11111
00010	01001	10111	11100
00100	01111	10001	11010
01000	00011	11101	10110
10000	11011	00101	01110
10010	11001	00111	01100
11000	10011	01101	00110

Természetesen előfordulhat, hogy a normák megegyeznek, akkor választunk egyet.

A kódolás menete: kiválasztjuk a kódszót a táblázatban, majd hozzárendeljük az oszlop tetején lévő kódszót. Ehhez pedig az eredeti kódot.

Legyen $\tilde{v}^T=11011$, ami a hatodik sor második oszlopban található. Az oszlop tetején lévő elem: 01011, amelyhez eredetileg a 01 kódszó tartozik. Ekkor feltételeztük, hogy a hibavektor 10000, a hatodik sor első eleme.

▲

6.23. Tétel. *A dekódolási táblázatban bármely szó távolsága a saját oszlopa tetején álló kódszótól nem nagyobb, mint bármely más kódszótól.*

6.24. Megjegyzés. *A dekódolási táblázat alkalmas maximum likelihood kódolásra.*

6.25. Megjegyzés. *Az osztályelső alakú hibák javíthatók.*

6.26. Megjegyzés. *A helyes dekódolás valószínűsége megegyezik az osztályelső alakú hibák valószínűségeinek az összegével.*

4. 6.4. Hamming-kód

Sokféle Hamming-kód van. Itt a legegyszerűbb esetet vizsgáljuk.

6.27. Definíció. *Hamming-kódnak nevezzük azokat a szisztematikus kódokat, amelyek pontosan egy hibát tudnak javítani.*

6.28. Megjegyzés. *A lineáris (k,n)-kód 2^k k hosszúságú közleményhez rendeli hozzá a kódszavakat kölcsönösen egyértelmű módon. Bázistranszformációval könnyű megmutatni, hogy minden lineáris kód előállítható szisztematikus generátormátrixszal.*

$$s^T = \tilde{v}^T F = e^T F, \qquad \tilde{v} = v \vee e,$$

ahol e a hibavektor. Tehát a szindrómából az e hibavektor egyértelműen megadható. Ugyanis a dekódolás hibátlan lesz, hiszen a hibavektor ismeretében az üzenet is meghatározható.

Javítsunk ki minden egy hibát! Ha $e = e_l$, azaz az l-edik egységvektor, akkor $s^T = f_l^T$. Éppen a paritásellenőrző mátrix l-edik sora. Minden sornak különbözőnek kell lennie, azaz

$$n = 2^{n-k} - 1.$$

Ebből

$$k = 2^{n-k} - (n-k) - 1.$$

Néhány kód mérete kiszámolva:

$n-k$	2	3	4	5	6
k	1	4	11	26	57
n	3	7	15	31	63

Vegyük észre, hogy az $(1,3)$-kód éppen a bit háromszorozása.

5. 6.5. Feladatok

1. Legyen a szisztematikus kód paritásmátrixa

$$P = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

Dekódolja az 10111 vett kódszót a maximum likelihood dekódolás alapján!

2. Bináris szimmetrikus csatornán a hibás továbbítás valószínűsége $q = 0.02$. Legyen a szisztematikus kód paritásmátrixa

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Határozza meg a hibaáteresztés valószínűségét! Dekódolja az 101111 vett kódszót a maximum likelihood dekódolás alapján (Indoklás!)!

3. Bináris szimmetrikus csatornán a hibás továbbítás valószínűsége $q = 0.08$. Legyen a szisztematikus kód paritásmátrixa

$$P = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Határozza meg a hibaáteresztés valószínűségét! Dekódolja az 101101 vett kódszót a maximum likelihood dekódolás alapján (Indoklás!)!

4. Legyen a szisztematikus kód paritásmátrixa

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

Határozza meg a hibaáteresztés valószínűségét!

5. Bináris szimmetrikus csatornán a hibás továbbítás valószínűsége $q = \frac{1}{8}$ és az A, B, C, D betűket rendre 111, 100, 010, 001-gyel kódoljuk, és 110, 101, 011, 000 vétele esetén is A-t, B-t, C-t, D-t dekódolunk. Határozza meg a dekódolásnál elkövetett hiba valószínűségét!

6. A szisztematikus kód paritásmátrixa

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Határozza meg a szindrómákat!

7. Bináris szimmetrikus csatornán a szisztematikus kód paritásmátrixa

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Dekódolja az 111011 vett üzenetet!!

8. A lineáris kód generátormátrixa

$$G = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Határozza meg a hibaáteresztés valószínűségét, ha $q = 0.023$!

9. A szisztematikus kód paritásmátrixa

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Dekódolja az 11111 vett kódszót a maximum likelihood dekódolás alapján! Határozza meg a hibaáteresztés valószínűségét, ha p=0.98!

10. A szisztematikus kód paritásmátrixa

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Határozza meg a dekódolási táblázatot!

11. A szisztematikus kód paritásmátrixa

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Határozza meg a kódszavak részcsoportját és határozza meg a javítható hibák számát!

12. A szisztematikus kód paritásmátrixa

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Határozza meg a kódszavak részcsoportját és határozza meg a kódtávolságot!

13. A szisztematikus kód paritásmátrixa

$$P = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Határozza meg az 110011 vett kódszóhoz tartozó szindrómát, továbbá azokat a vett kódszavakat, amelyek szindrómája azonosan 0!

14. A szisztematikus kód paritásmátrixa

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

Határozza meg a hibaáteresztés valószínűségét és a $v^T = (0, 0, 1, 1, 1, 1)$ vektorhoz tartozó szindrómát és mellékosztályt!

15. Bináris szimmetrikus csatornán a hibás továbbítás valószínűsége $q = \frac{1}{8}$. A szisztematikus kód paritásmátrixa

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Határozza meg a hibaáteresztés valószínűségét! Dekódolja az 111011 vett üzenetet!

16. A hibajavító kód generátormátrixa

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

Határozza meg azt a mellékosztályt, amelyhez tartozó szindróma $s^T = 110$!

6. 6.6. Önellenőrző kérdések

1. Definiálja a Hamming-kódot!

2. Definiálja a mellékosztályt!

3. Ismertesse és bizonyítsa a maximum likelihood kódolás és a minimális kódtávolság ekvivalenciáját!

4. Igazolja, hogy a mellékosztályok diszjunktak!

5. Definiálja a lineáris kódot!

6. Ismertesse a Hamming-távolság tulajdonságait!

7. Igazolja, hogy a lineáris kód csoportkód!

8. Ismertesse és bizonyítsa a Hamming-távolság eltolásinvarianciáját!

9. Definiálja a szisztematikus kódot!

10. Bizonyítsa a hibajavíthatóság és kódtávolság kapcsolatára vonatkozó állítást!

11. Definiálja a szisztematikus kódot!

12. Definiálja a Hamming-féle távolságot!

13. Igazolja, hogy a szindróma jellemző a mellékosztályra!

14. Igazolja, hogy a dekódolási táblázatban bármely szó távolsága a saját oszlopa tetején álló kódszótól nem nagyobb, mint bármely más kódszótól!

15. Adott egy csoportkód és benne egy rögzített v kódszó. Igazolja, ha az e hiba javítható v-nél, akkor ez a tulajdonsága igaz bármely más kódszóra is!

7. fejezet - Bevezetés a folytonos esetbe

1. 7.1. Diszkretizálás

A ξ folytonos valószínűségi változó lehetséges értékeinek halmaza nem megszámlálható, ezért $H(\xi)$ definiálásához először elkészítjük a ξ_δ diszkrét valószínűségi változót, amely a ξ kerekítésének is tekinthető:

$$\xi_\delta = n\delta, \quad \text{ha} \quad (n-1)\delta < \xi \leq n\delta, \quad n \in \mathbf{Z}.$$

Ekkor

$$P(\xi_\delta = n\delta) = P((n-1)\delta < \xi \leq n\delta) = \int\limits_{(n-1)\delta}^{n\delta} f(x)dx = \delta\tilde{f}(n\delta),$$

ahol $m_n \leq \tilde{f}(n\delta) \leq M_n$, azaz a minimum és a maximum közötti érték az adott intervallumon.

$$H(\xi_\delta) = -\sum_{n=-\infty}^{+\infty} \delta\tilde{f}(n\delta)\ln(\delta\tilde{f}(n\delta)) =$$

$$= -\ln\delta - \sum_{n=-\infty}^{+\infty} \delta\tilde{f}(n\delta)\ln(\tilde{f}(n\delta)),$$

mert

$$\sum_{n=-\infty}^{+\infty} \delta\tilde{f}(n\delta) = \int\limits_{-\infty}^{+\infty} f(x)dx = 1.$$

Ha $\delta \to 0$, akkor $-\ln\delta \to +\infty$, ezért

$$\lim_{\delta\downarrow 0}(H(\xi_\delta) + \ln\delta) = -\int\limits_{-\infty}^{+\infty} f(x)\ln f(x)dx = H(\xi).$$

7.1. Példa. Legyen ξ a $(0, \vartheta)$ intervallumon egyenletes eloszlású. Ekkor

$$H(\xi) = \ln\vartheta,$$

azaz jól látható, hogy a $H(\xi)$ lehet negatív is.

▲

7.1. Megjegyzés. *Az entrópia diszkrét esetben a bizonytalanságot méri, míg folytonos esetben csak a bizonytalanság változását.*

2. 7.2. Néhány fogalom folytonos esetben

Entrópia, mint várható érték: $H(\xi) = E(-\ln f(\xi))$.

7.2. Példa. 1. Exponenciális eloszlásra: $H(\xi) = 1 - \ln\lambda$.

2. Normális eloszlásra: $H(\xi) = 0.5 + \ln(\sigma\sqrt{2\pi})$.

▲

Együttes entrópia: $H(\xi,\eta) = E(-\ln f(\xi,\eta))$.

7.3. Példa. Normális eloszlásra: $H(\xi,\eta) = 1 + \ln(2\pi\sigma_1\sigma_2\sqrt{1-r^2})$.

▲

Kölcsönös információmennyiség:

$$I(\xi,\eta) = E\left(\ln \frac{f(\xi,\eta)}{f_\xi(\xi) f_\eta(\eta)}\right).$$

7.4. Példa. Normális eloszlásra: $I(\xi,\eta) = -0.5\ln(1-r^2)$.

▲

I-divergencia:

$$D(\eta||\xi) = E\left(\ln \frac{f_\eta(\eta)}{f_\xi(\eta)}\right).$$

7.2. Megjegyzés. $D(\eta||\xi) \geq 0$.

7.3. Megjegyzés. *Transzformáció:* $\eta = g(\xi)$.

Diszkrét esetben $H(\eta) \leq H(\xi)$. *Egyenlőség akkor és csak akkor, ha* g *invertálható.*

Folytonos esetben $H(\eta) \leq H(\xi) + E(\ln|g'(\xi)|)$. *Egyenlőség akkor és csak akkor, ha* g *invertálható.*

Bizonyítás. Ha g invertálható, akkor a valószínűségi változók transzformációja alapján

$$\begin{aligned} H(\eta) &= -\int_{-\infty}^{+\infty} f_\eta(y)\ln f_\eta(y)dy = \\ &= -\int_{-\infty}^{+\infty} f_\xi(x)\ln\frac{f_\xi(x)}{|g'(x)|}dx = \\ &= -\int_{-\infty}^{+\infty} f_\xi(x)\ln f_\xi(x)dx + \int_{-\infty}^{+\infty} f_\xi(x)\ln|g'(x)|dx. \end{aligned}$$

■

3. 7.3. Maximum entrópia módszer (MEM)

Entrópia maximalizálás feltételek mellett.

7.5. Példa. Pénzfeldobás: $\xi = P(fej)$. Feltétel: A sűrűségfüggvény 0 a $[0,1]$ intervallumon kívül. Kérdés: $H(\xi)$ mikor lesz maximális?

Az I-divergencia nemnegativitása alapján tetszőleges g sűrűségfüggvény esetén

$$-\int_0^1 g(x)\ln g(x)dx \leq -\int_0^1 g(x)\ln f(x)dx = 0 = H(\xi),$$

ha ξ egyenletes eloszlású a $[0,1]$ intervallumon.

▲

Várható érték feltételek:

A ξ valószínűségi változó f sűrűségfüggvényét nem ismerjük. Viszont adott, hogy

$$E(g_i(\xi)) = a_i, \qquad (i = 1, 2, \ldots, n),$$

ahol a g_i függvények ismertek. Ekkor a MEM alapján

$$f(x) = A \exp\left(-\sum_{i=1}^{n} \lambda_i g_i(x) \right),$$

ahol a λ_i értékeket meghatározzák az adott várható értékek, míg az A értéke abból adódik, hogy f sűrűségfüggvény.

Bizonyítás. Ha $f(x)$ ebben a formában adott, akkor

$$E(\ln f(\xi)) = \ln A - \sum_{i=1}^{n} \lambda_i a_i.$$

Más sűrűségfüggvény esetén az I-divergencia alapján:

$$-E(\ln g(\xi)) \leq \sum_{i=1}^{n} \lambda_i a_i - \ln A.$$

■

4. 7.4. Feladatok

1. Határozza meg az entrópiát a Gamma-eloszláshoz!

2. Határozza meg az entrópiát a Cauchy-eloszláshoz!

3. Határozza meg az entrópiát az n-dimenziós normális eloszláshoz!

4. A ξ valószínűségi változó nemnegatív és $E(\xi) = 3.$ Határozza meg, hogy ilyen feltételek mellett melyik folytonos eloszlás esetén lesz az entrópia maximális?

5. A ξ valószínűségi változóra $D(\xi) = 3.$ Határozza meg, hogy ilyen feltételek mellett melyik folytonos eloszlás esetén lesz az entrópia maximális?

5. 7.5. Önellenőrző kérdések

1. Ismertesse az I-divergencia tulajdonságait!

2. Bizonyítsa, hogy az I-divergencia folytonos esetben is nemnegatív!

3. Definiálja a kölcsönös információmennyiséget!

4. Ismertesse és bizonyítsa a maximum entrópia módszert várható érték feltételek esetén!

5. Adjon meg olyan esetet, amikor az entrópia negatív (folytonos)!

6. Definiálja az entrópiát!

7. Definiálja a Kullback-Leibler eltérést!

8. Milyen a kapcsolat a valószínűségi változó és a transzformált valószínűségi változó entrópiája között?

8. fejezet - Függelék

1. 8.1. Jelölések

$\mathbf{N}$ – a természetes számok halmaza (pozitív egészek)

$\mathbf{R}$ – a valós számok halmaza

$\mathbf{R}^2$ – $\{(x, y)|x, y \in \mathbf{R}\}$

$A \subset B$ – az A részhalmaza a B-nek

$A \cap B$ – az A és B halmaz közös része

$A \cup B$ – az A és B halmaz összes eleme egy halmazban

$\overline{A}$ – az alaphalmaz A halmazon kívüli elemei

$A \backslash B$ – $A \cap \overline{B}$

$F(a + 0)$ – a jobboldali határérték, azaz $\lim_{x \to a+0} F(x)$

$F(a - 0)$ – a baloldali határérték, azaz $\lim_{x \to a-0} F(x)$

$\exp(x)$ – e^x

$f(\cdot) : D \to R$ – az f leképezés, D az értelmezési tartomány, a „pont” a változót helyettesíti

$f(D)$ – az f leképezés értékkészlete

2. 8.2. Konvex függvények

8.1. Definíció. *Legyen U egy intervallum (zárt, nyílt, félig zárt). Az $f : U \to \mathbf{R}$ konvex függvény, ha*

$$f(\lambda a + \mu b) \leq \lambda f(a) + \mu f(b),$$

ahol $a, b \in U, \lambda + \mu = 1, \lambda \geq 0$ és $\mu \geq 0$.

8.2. Tétel. *Ha f és g konvex függvény és $\alpha \geq 0, \beta \geq 0$, akkor $\alpha f + \beta g$ szintén konvex.*

8.3. Tétel. *Véges sok konvex függvény összege is konvex.*

8.4. Tétel. *Konvex függvények egy konvergens sorozatának a (pontonkénti) határa is konvex.*

8.5. Tétel. *Ha $f : U \to \mathbf{R}$ konvex függvény és $a < x < b$, akkor*

$$\frac{f(x) - f(a)}{x - a} \leq \frac{f(b) - f(a)}{b - a} \leq \frac{f(b) - f(x)}{b - x}.$$

Ha f szigorúan konvex, akkor az egyenlőtlenségek is szigorúak.

8.6. Tétel. *Ha $f : U \to \mathbf{R}$ konvex függvény és $a < c < b$, akkor létezik a bal- és jobboldali derivált minden c esetén. Továbbá, f'_- és f'_+ monoton nemcsökkenő és*

$$f'_-(c) \leq f'_+(c).$$

Ezenkívül minden $x \in U$ esetén

$$f(x) \geq f(c) + f'_-(c)(x - c), \quad f(x) \geq f(c) + f'_+(c)(x - c),$$

azaz a konvex függvény minden pontjához létezik egyenes (amely az adott ponton keresztül megy), amely a görbe alatt marad vagy legfeljebb érinti azt.

8.7. Tétel. *Az* $f : U \to \mathbf{R}$ *konvex függvény folytonos az intervallum minden belső pontjában.*

8.8. Tétel. *Legyen* U *nyílt és* f *kétszer differenciálható. Az* f *konvex akkor és csak akkor, ha* $f'' > 0$ *minden* $x \in U$.

8.9. Tétel. (Jensen-egyenlőtlenség) *Ha* f *konvex függvény és* ξ *olyan valószínűségi változó, amelyre létezik* $E(f(\xi))$ *és* $f(E(\xi))$, *akkor*

$$E(f(\xi)) \geq f(E(\xi)).$$

Bizonyítás. Legyen L a támasztóegyenes az f függvényhez az $(E(\xi), f(E(\xi)))$ pontban, akkor

$$E(f(\xi)) \geq E(L(\xi)) = L(E(\xi)) = f(E(\xi)).$$

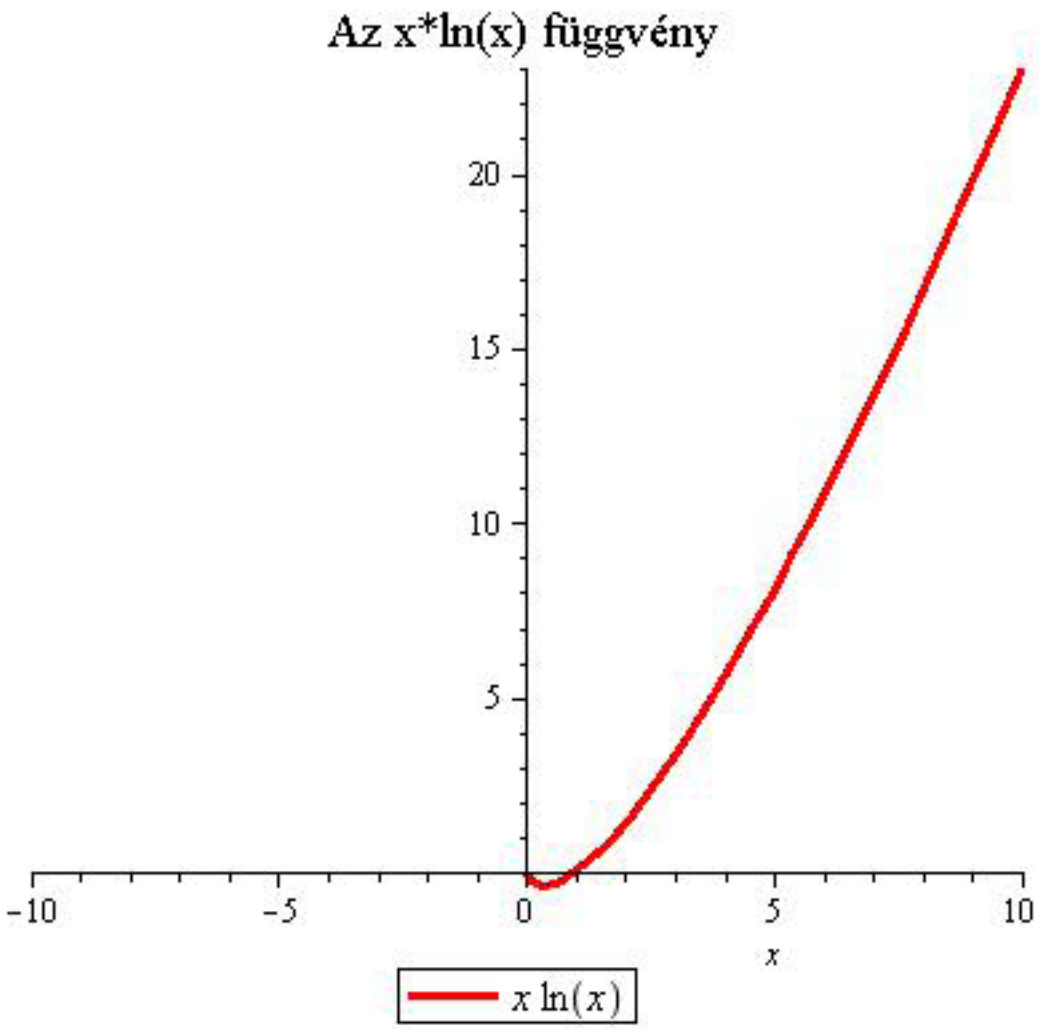
■

8.10. Megjegyzés. $E(\xi^2) \geq E^2(\xi)$.

3. 8.3. Az $x \ln x$ függvény vizsgálata

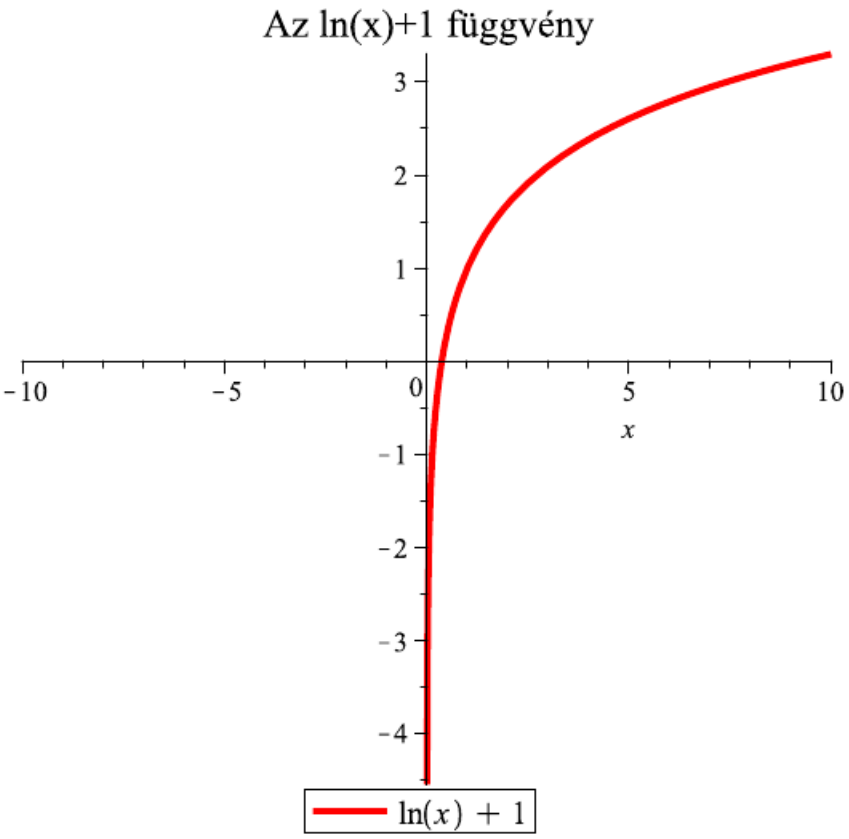
Az $f(x) = x \ln x$ csak $x > 0$ esetén értelmezett, viszont folytonosan kiterjeszthető az $x = 0$ esetre, azaz ha $x \to 0$, akkor létezik f határértéke.

8.1. ábra - Az $xln(x)$ függvény



$f'(x) = 1 + \ln x$, amiből látható, hogy $x < e^{-1}$ esetén $f'(x) < 0$, azaz f monoton csökkenő a $(0, e^{-1})$ szakaszon.

8.2. ábra - Az $xln(x)$ függvény deriváltja



Továbbá,

$$1 \leq \sqrt[n]{n} \leq \frac{2\sqrt{n} + (n-2)\cdot 1}{n} = \frac{n-2}{n} + \frac{2}{\sqrt{n}} < 1 + \frac{2}{\sqrt{n}},$$

így

$$\lim_{n\to+\infty} \sqrt[n]{n} = 1.$$

Tehát

$$\lim_{x\to 0+0} x \ln x = \lim_{n\to+\infty} \frac{1}{n} \ln \frac{1}{n} = \lim_{n\to+\infty} \left(-\ln \sqrt[n]{n}\right) = 0.$$

8.11. Tétel.

$$1 - \frac{1}{x} \leq \ln x \leq x - 1.$$

Bizonyítás. Az $\ln x$ függvény konkáv, így az $x = 1$ helyen felírt támasztó egyenesre igaz, hogy

$$\ln x \leq x - 1,$$

egyenlőség csak $x = 1$ esetén. Továbbá, ha $x > 0$, akkor

$$\frac{1}{x} > 0$$

is teljesül, azaz

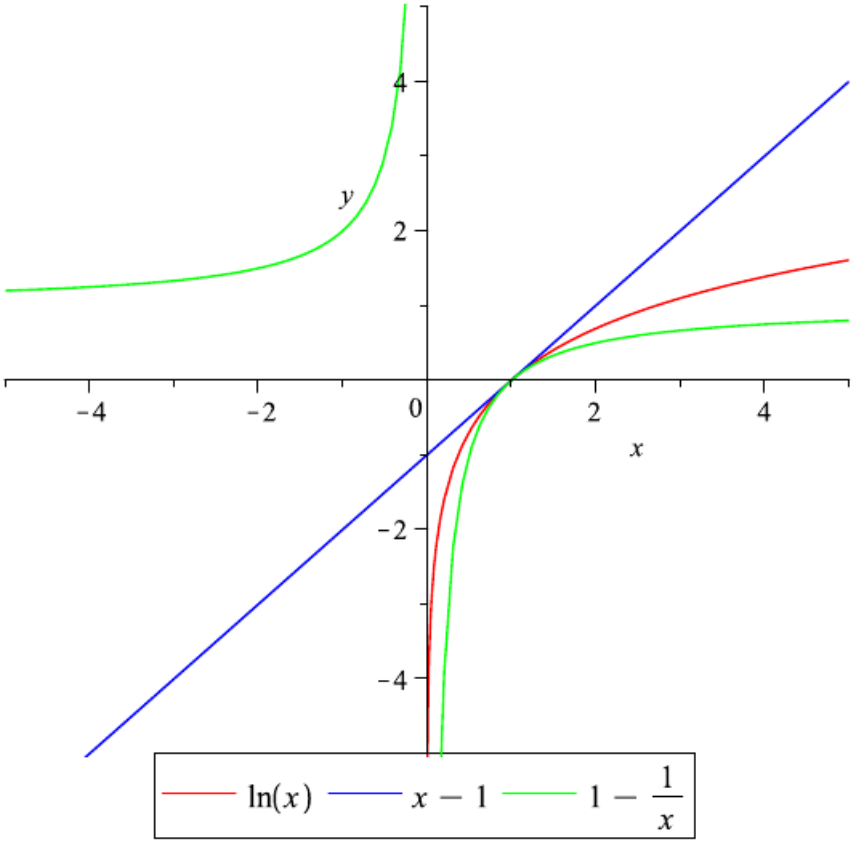
$$\ln \frac{1}{x} \leq \frac{1}{x} - 1,$$

ami ekvivalens azzal, hogy

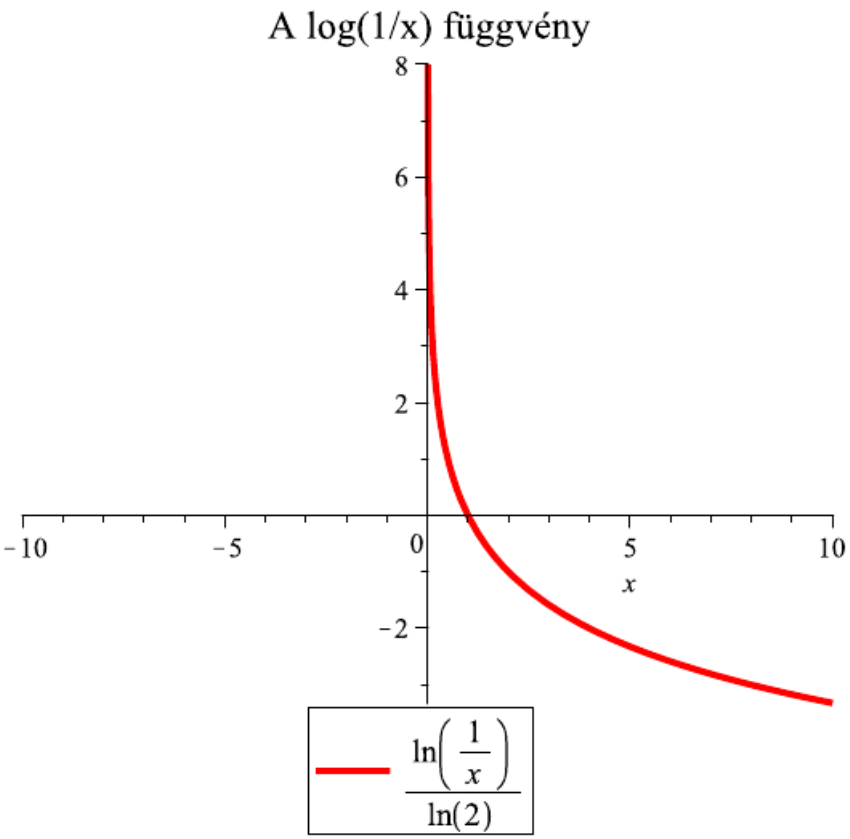
$\ln x \geq 1 - \frac{1}{x}$.

■

8.3. ábra - A logaritmus függvény konvexitásának bemutatása



8.4. ábra - A reciprok logaritmusa



4. 8.4. Az aszimptotikus Stirling-formula

8.12. Tétel. *Ha* $\{a_n > 0\}$ *(valós) számsorozat és* $a_n \to a$, *akkor*

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} a_k = a \quad \text{és} \quad \lim_{n\to+\infty} \sqrt[n]{\prod_{k=1}^{n} a_k} = a.$$

8.13. Tétel.

$$\lim_{n\to+\infty} \frac{n}{\sqrt[n]{n!}} = e.$$

8.14. Tétel. (Aszimptotikus Stirling-formula)

$$\lim_{n\to+\infty} \frac{\ln n!}{n\ln n - n} = 1.$$

5. 8.5. Valószínűség-számítás összefoglaló

5.1. 8.5.1. A valószínűség fogalma

8.15. Definíció. *Egy véletlen kísérlet lehetséges eredményeinek összességét* ***eseménytér****nek (mintatér) nevezzük. Jele:* Ω. *Az* Ω *elemeit* ***elemi esemény****eknek nevezzük.*

8.16. Definíció. *Az* Ω *részhalmazainak egy* $\mathcal{F}$ *rendszerét* σ***-algebrá****nak nevezzük, ha*

(1) $\Omega \in \mathcal{F}$,

(2) $A \in \mathcal{F}$, *akkor* $\overline{A} \in \mathcal{F}$,

(3) $A_1, A_2, \dots \in \mathcal{F}$, *akkor* $A_1 \cup A_2 \cup \dots \in \mathcal{F}$.

*Az $\mathcal{F}$ elemeit pedig **esemény**eknek nevezzük.*

8.17. Megjegyzés. *Ha $A, B \in \mathcal{F}$, akkor $A \cap B \in \mathcal{F}$.*

8.18. Definíció. *Az Ω-t szokás **biztos esemény**nek, az $\emptyset$-t pedig **lehetetlen esemény**nek nevezni. Továbbá, az A esemény **bekövetkezik**, ha a kísérlet eredménye eleme az A halmaznak.*

8.19. Megjegyzés. *Az $A \cup B$ esemény bekövetkezik, ha legalább az egyik közülük bekövetkezik, míg az $A \cap B$ esemény akkor következik be, ha mind a kettő bekövetkezik.*

8.20. Definíció. *A $P : \mathcal{F} \to \mathbf{R}$ nemnegatív leképezést valószínűségnek nevezzük, ha*

(1) $P(\Omega) = 1,$

(2) $A \cap B = \emptyset,$ *akkor* $P(A \cup B) = P(A) + P(B),$

(3) $A_1, A_2, \ldots$ *egymást kölcsönösen kizáró események (azaz* $A_i \cap A_j = \emptyset,$ *ha* $i < j$ *és* $i, j = 1, 2, \ldots$*), akkor*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

8.21. Lemma.

(1) $P(\overline{A}) = 1 - P(A).$

(2) $P(\emptyset) = 0.$

(3) $P(B \backslash A) = P(B) - P(A \cap B).$

(4) Ha $A \subset B,$ *akkor* $P(A) \leq P(B).$

(5) $P(A \cup B) = P(A) + P(B) - P(A \cap B).$

(6) Ha $B_{n+1} \subset B_n$ *és* $\bigcap_{i=1}^{\infty} B_n = \emptyset,$ *akkor* $\lim_{n \to \infty} P(B_n) = 0.$

8.22. Definíció. *Az $(\Omega, \mathcal{F}, P)$ hármast **valószínűségi mező**nek nevezzük.*

8.23. Definíció. *Ha az elemi események száma véges és valószínűségük megegyezik, akkor a valószínűségi mezőt **klasszikus**nak nevezzük.*

8.24. Megjegyzés. *Legyen $|\Omega| = n$ és jelölje az elemi eseményeket ω_i $(i = 1, 2, \ldots, n).$ Ekkor*

$$1 = P(\Omega) = P\left(\bigcup_{i=1}^{n} \{\omega_i\}\right) = \sum_{i=1}^{n} P(\{\omega_i\}) = nP(\{\omega_i\}).$$

Tehát $P(\{\omega_i\}) = \frac{1}{n}$ $(i = 1, 2, \ldots, n).$

8.25. Definíció. ***Bernoulli kísérletsorozat**nak nevezzük azt, ha adott $A \in \mathcal{F}$ és egymástól függetlenül, azonos körülmények között elvégezzük ugyanazt a kísérletet, s "csak" azt figyeljük, hogy az A esemény bekövetkezett-e vagy sem.*

8.1. Példa. *Visszatevéses mintavétel:* Adott N darab különböző objektum, amelyek közül s darab rendelkezik egy bizonyos tulajdonsággal, például selejt. Visszatevéssel kiveszünk n darabot. Legyen a kivett selejtek száma ξ. Mennyi a valószínűsége, hogy $\xi = k,$ ahol $0 \leq k \leq n.$

$$P(\xi = k) = \frac{\binom{n}{k} s^k (N - s)^{n-k}}{N^n}.$$

▲

8.2. Példa. *Visszatevés nélküli mintavétel:* Adott N darab különböző objektum, amelyek közül s darab rendelkezik egy bizonyos tulajdonsággal, például selejt. Visszatevés nélkül kiveszünk n darabot. Legyen a kivett selejtek száma ξ. Mennyi a valószínűsége, hogy $\xi = k$, ahol $0 \leq k \leq \min\{n, s\}$.

$$P(\xi = k) = \frac{\binom{s}{k}\binom{N-s}{n-k}}{\binom{N}{n}}.$$

▲

8.26. Tétel. (Poincaré) *Az $A_1, A_2, \ldots, A_n$ eseményekre*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{i_1<i_2<\cdots<i_k} P\left(\bigcap_{j=1}^{k} A_{i_j}\right),$$

ahol az összegzést az összes lehetséges $\{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots, n\}$ esetre tekintjük.

8.27. Definíció. *Az A esemény B feltétel melletti **feltételes valószínűség**ének nevezzük a*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

mennyiséget, ha $P(B) > 0$.

8.28. Megjegyzés. *A $P(\cdot|B) : \mathcal{F} \to \mathbf{R}$ leképezés tényleg valószínűség.*

8.29. Lemma. *Ha az $A_1, A_2, \ldots, A_n$ eseményrendszerre $P(\bigcap_{i=1}^{n-1} A_i) > 0$, akkor*

$$P(\bigcap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)\cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

8.30. Definíció. *Az $A_1, A_2, \ldots$ eseményrendszert **teljes eseményrendszer**nek nevezzük, ha*

$A_i \cap A_j = \emptyset$,

($i < j$ és $i, j = 1, 2, \ldots$) és

$$\bigcup_{i=1}^{\infty} A_i = \Omega.$$

8.31. Tétel. (teljes valószínűség) *Ha $A_1, A_2, \ldots$ teljes eseményrendszer és $P(A_i) > 0$, ha $i = 1, 2, \ldots$, akkor tetszőleges B esemény esetén*

$$P(B) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i).$$

8.32. Tétel. (Bayes) *Ha $A_1, A_2, \ldots$ teljes eseményrendszer és $P(A_i) > 0$, ha $i = 1, 2, \ldots$, akkor tetszőleges pozitív valószínűségű B esemény esetén*

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{\infty} P(B|A_i)P(A_i)}.$$

8.33. Megjegyzés. *A Bayes-tételhez kapcsolódóan bevezethetjük a következő elnevezéseket: $P(A_i)$ az ún.* a-priori *valószínűség és $P(A_i|A)$ az ún.* a-posteriori *valószínűség.*

8.34. Definíció. *Az A és B eseményt **sztochasztikusan független**nek nevezzük, ha*

$P(A \cap B) = P(A)P(B).$

*Az $A_1, A_2, \ldots, A_n$ eseményeket **páronként sztochasztikusan független**nek nevezzük, ha*

$P(A_i \cap A_j) = P(A_i)P(A_j) \qquad (1 \leq i < j \leq n).$

*Az $A_1, A_2, \ldots, A_n$ eseményeket **teljesen sztochasztikusan független**nek nevezzük, ha*

$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}),$

ahol $1 \leq i_1 < \cdots < i_k \leq n, \quad 2 \leq k \leq n.$

8.35. Lemma. *Ha az A és B események függetlenek, akkor $\overline{A}$ és B, A és $\overline{B}$ és $\overline{A}$ és $\overline{B}$ is függetlenek.*

8.36. Lemma. *Ha $A_1, A_1, \ldots, A_n$ független események és $P(A_i) < 1 (i = 1, 2, \ldots, n)$, akkor*

$P(\bigcup_{i=1}^{n} A_i) < 1.$

Bizonyítás.

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\overline{\overline{\bigcup_{i=1}^{n} A_i}}\right) = 1 - P\left(\overline{\bigcup_{i=1}^{n} A_i}\right) =$$
$$= 1 - P\left(\bigcap_{i=1}^{n} \overline{A_i}\right) = 1 - \prod_{i=1}^{n} P(\overline{A_i}).$$

■

5.2. 8.5.2. A valószínűségi változó

8.37. Definíció. *A $\xi : \Omega \to \mathbf{R}$ leképezést **valószínűségi változó**nak nevezzük, ha*

$\{\xi < x\} = \{\omega | \omega \in \Omega, \quad \xi(\omega) < x\} \in \mathcal{F}, \qquad \forall x \in \mathbf{R}.$

8.38. Definíció. *Az $F(x) = P(\xi < x)$ formulával meghatározott valós függvényt a ξ valószínűségi változó **eloszlásfüggvény**ének nevezzük.*

8.39. Tétel. *Az F valós függvény akkor és csak akkor lehet eloszlásfüggvény, ha*

1. $\lim_{x \to -\infty} F(x) = 0,$
2. $\lim_{x \to \infty} F(x) = 1,$
3. $F(x_1) \leq F(x_2)$, *ha* $x_1 < x_2$, *azaz monoton növekvő,*
4. $\lim_{x \to x_0 - 0} F(x) = F(x_0), \forall x_0 \in \mathbf{R}$, *azaz balról folytonos.*

8.40. Tétel. *Legyen F a ξ valószínűségi változó eloszlásfüggvénye és $a, b \in \mathbf{R}$, ekkor*

1. $P(a \leq \xi < b) = F(b) - F(a),$
2. $P(\xi = a) = F(a + 0) - F(a).$

8.41. Definíció. *A ξ valószínűségi változót **diszkrét**nek nevezzük, ha a lehetséges értékek $\xi(\Omega)$ halmazának számossága legfeljebb megszámlálhatóan végtelen.*

8.42. Megjegyzés. *Diszkrét valószínűségi változó esetén a lehetséges értékek felírhatók egy sorozatként.*

$(i = 1, 2, \dots)$ió. *Legyen a ξ valószínűségi változó lehetséges értekeinek sorozata $x_1, x_2, \dots$. A $p_i = P(\xi = x_i)$ valószínűségek sorozatát* ***eloszlás****nak nevezzük.*

8.44. Tétel. *Ha $p_1, p_2, \dots$ eloszlás, akkor*

$$p_i \geq 0 \quad (i = 1, 2, \dots) \quad \text{és} \quad \sum_{i=1}^{\infty} p_i = 1.$$

8.45. Definíció. *Ha létezik f nemnegatív valós függvény, melyre*

$$F(x) = \int_{-\infty}^{x} f(t)dt, \quad \forall x \in \mathbf{R},$$

akkor f az F eloszlásfüggvényhez tartozó ***sűrűségfüggvény****.*

8.46. Megjegyzés. *A sűrűségfüggvény nem egyértelmű.*

8.47. Tétel. *Az f valós függvény akkor és csak akkor lehet sűrűségfüggvény, ha nemnegatív és*

$$\int_{-\infty}^{+\infty} f(t)dt = 1.$$

8.48. Definíció. *A valószínűségi változót* ***folytonos****nak nevezzük, ha létezik a sűrűségfüggvénye.*

8.49. Tétel. *Legyen a ξ folytonos valószínűségi változó f sűrűségfüggvénnyel és $a, b \in \mathbf{R}$, ekkor $P(\xi = a) = 0$, és $P(a \leq \xi < b) = \int_a^b f(x)dx$.*

8.50. Definíció.

1. Ha a ξ diszkrét valószínűségi változó lehetséges értékeinek a száma véges, azaz a lehetséges értékek

$$x_1, x_2, \dots, x_n \quad \text{és} \quad p_i = P(\xi = x_i) \quad (i = 1, 2, \dots, n),$$

akkor a

$$\sum_{i=1}^{n} x_i p_i$$

mennyiséget várható ***érték****nek nevezzük.*

2. Ha a ξ diszkrét valószínűségi változó lehetséges értékeinek számossága megszámlálhatóan végtelen, azaz a lehetséges értékek

$$x_1, x_2, \dots, \quad \text{és} \quad p_i = P(\xi = x_i) \quad (i = 1, 2, \dots),$$

akkor a

$$\sum_{i=1}^{\infty} x_i p_i$$

mennyiséget ***várható érték****nek nevezzük, ha* $\sum_{i=1}^{\infty} |x_i| p_i < +\infty.$

3. Ha ξ folytonos valószínűségi változó f sűrűségfüggvénnyel, akkor a

$$\int_{-\infty}^{+\infty} x f(x)dx$$

*mennyiséget **várható érték**nek nevezzük, ha*

$$\int_{-\infty}^{+\infty} |x| f(x) dx < +\infty.$$

A ξ valószínűségi változó várható értékének a jele: $E(\xi)$.

8.51. Tétel.

1. $E(a\xi + b) = aE(\xi) + b, \forall a, b \in \mathbf{R}$.

2. Ha $m \leq \xi \leq M$, *akkor* $m \leq E(\xi) \leq M$.

8.52. Definíció. *Legyen ξ valószínűségi változó és g valós függvény. Ha az $\eta = g(\xi)$ függvény valószínűségi változó, akkor a ξ **transzformált**jának nevezzük.*

8.53. Megjegyzés. *A transzformált eloszlásfüggvénye*

$$F_\eta(y) = P(\{\omega | g(\xi(\omega)) < y\}).$$

8.54. Tétel. *Ha g differenciálható és $g'(x) \neq 0$, akkor ξ folytonos valószínűségi változó esetén $\eta = g(\xi)$ folytonos valószínűségi változó, melynek sűrűségfüggvénye*

$$f_\eta(y) = \begin{cases} f_\xi(g^{-1}(y)) |\dfrac{d}{dy} g^{-1}(y)|, & \text{ha} a < y < b, \\ 0, & \text{egyébként}, \end{cases}$$

ahol

$$a = \min(\lim_{x \to -\infty} g(x), \lim_{x \to +\infty} g(x)), \quad b = \max(\lim_{x \to -\infty} g(x), \lim_{x \to +\infty} g(x)).$$

8.55. Tétel. *Ha $\eta = g(\xi)$ a ξ valószínűségi változó transzformáltja, akkor*

$$E(\eta) = \begin{cases} \sum_{i=1}^{\infty} g(x_i) P(\xi = x_i), & \text{ha} \xi \text{diszkrét}, \\ \int_{-\infty}^{+\infty} g(x) f_\xi(x) dx, & \text{ha} \xi \text{és} \eta \text{folytonos}. \end{cases}$$

8.56. Definíció. *Az $E((\xi - E(\xi))^2)$ mennyiséget a ξ valószínűségi változó **szórásnégyzet**ének nevezzük. Jele: $D^2(\xi)$.*

8.57. Definíció. *A $\sqrt{E((\xi - E(\xi))^2)}$ mennyiséget a ξ valószínűségi változó **szórás**ának nevezzük. Jele: $D(\xi)$.*

8.58. Definíció. *Az $E(\xi^k)$ mennyiséget a ξ valószínűségi változó k**-adik momentum**ának nevezzük.*

8.59. Definíció. *Az $E((\xi - E(\xi))^k)$ mennyiséget a ξ valószínűségi változó k**-adik centrális momentum**ának nevezzük.*

8.60. Tétel.

1. $D(a\xi + b) = |a| D(\xi), \forall a, b \in \mathbf{R}$.

2. $\min_{a \in \mathbf{R}} E((\xi - a)^2) = D^2(\xi)$, *és ekkor* $a = E(\xi)$.

3. $D^2(\xi) = E(\xi^2) - E^2(\xi)$.

5.3. 8.5.3. Néhány diszkrét eloszlás és jellemzői

1. BINOMIÁLIS ELOSZLÁS

Legyen $n \in \mathbf{N}$, $A \in \mathcal{F}$ és végezzünk el egy n hosszúságú Bernoulli kísérletsorozatot. Továbbá, legyen ξ az A esemény bekövetkezéseinek a száma. Ekkor ξ eloszlása

$$P(\xi = k) = \binom{n}{k} p^k q^{n-k}, \qquad (k = 0, 1, \dots, n),$$

ahol $P(A) = p$ és $q = 1 - p$.

$E(\xi) = np, D^2(\xi) = npq.$

8.61. Megjegyzés. *A visszatevéses mintavétel binomiális eloszláshoz vezet.*

2. POISSON-ELOSZLÁS

Legyen $\lambda > 0$ és $\lambda = np_n$, ekkor

$$\lim_{\substack{n\to\infty, \\ \lambda = np_n}} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad \text{ahol} \quad k = 0, 1, \dots.$$

A ξ valószínűségi változót Poisson-eloszlásúnak nevezzük $\lambda > 0$ paraméterrel, ha eloszlása

$$P(\xi = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad \text{ahol} \quad k = 0, 1, \dots.$$

$E(\xi) = \lambda, D^2(\xi) = \lambda.$

3. GEOMETRIAI ELOSZLÁS

A binomiális eloszlás bevezetésekor használt jelölések mellett a ξ valószínűségi változó jelentse az A esemény első bekövetkezéséhez szükséges kísérletek számát. A ξ eloszlása

$$P(\xi = k) = pq^{k-1}, \qquad \text{ahol} \quad k = 1, 2, \dots.$$

$$E(\xi) = \frac{1}{p}, D^2(\xi) = \frac{q}{p^2}.$$

8.62. Megjegyzés. *A $\eta = \xi - 1$ valószínűségi változót is szokás geometriai eloszlásúnak nevezni. Az η eloszlása*

$$P(\eta = k) = pq^k, \qquad \text{ahol} \quad k = 0, 1, 2, \dots.$$

$$E(\eta) = \frac{q}{p}, D^2(\eta) = \frac{q}{p^2}.$$

5.4. 8.5.4. Néhány folytonos eloszlás és jellemzői

1. EGYENLETES ELOSZLÁS

Legyen $a, b \in \mathbf{R}$ és $a < b$. A ξ egyenletes eloszlású az (a, b) intervallumon, ha a sűrűségfüggvénye

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{ha} a < x < b, \\ 0, & \text{egyébként.} \end{cases}$$

$E(\xi) = \dfrac{a+b}{2}, D^2(\xi) = \dfrac{(b-a)^2}{12}.$ Az eloszlásfüggvény

$$F(x) = \begin{cases} 0, & \text{ha} x \le a, \\ \dfrac{x-a}{b-a}, & \text{ha} a < x \le b, \\ 1, & \text{ha} x > b. \end{cases}$$

2. EXPONENCIÁLIS ELOSZLÁS

A ξ exponenciális eloszlású $\lambda > 0$ paraméterrel, ha a sűrűségfüggvénye

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{ha} x \geq 0, \\ 0, & \text{egyébként.} \end{cases}$$

$E(\xi) = \frac{1}{\lambda}, D^2(\xi) = \frac{1}{\lambda^2}$. Az eloszlásfüggvény

$$F(x) = \begin{cases} 0, & \text{ha} x \leq 0, \\ 1 - e^{-\lambda x}, & \text{ha} x > 0. \end{cases}$$

Örökifjú tulajdonság: $P(\xi \geq a + b | \xi \geq a) = P(\xi \geq b)$, ahol $a > 0, b > 0$.

3. NORMÁLIS ELOSZLÁS

Legyen $m \in \mathbf{R}, \sigma > 0$. Az η normális eloszlású, ha a sűrűségfüggvénye

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \qquad x \in \mathbf{R}.$$

$E(\xi) = m$, $D^2(\xi) = \sigma^2$. Ha $m = 0$ és $\sigma = 1$, akkor a valószínűségi változót standard normális eloszlásúnak nevezzük. Jelölje a sűrűségfüggvényét φ és az eloszlásfüggvényét Φ. Ha ξ standard normális eloszlású, akkor az $\eta = \sigma\xi + m$ valószínűségi változó F eloszlásfüggvényére jellemző, hogy

$$F(x) = \Phi\left(\frac{x-m}{\sigma}\right).$$

8.63. Megjegyzés. *A φ függvény írja le a Gauss-görbét (harang görbét). $\Phi(0) = 0.5$ és $\Phi(-x) = 1 - \Phi(x)$.*

4. CAUCHY-ELOSZLÁS

Legyen $c \in \mathbf{R}, s > 0$. Az η Cauchy-eloszlású, ha a sűrűségfüggvénye

$$f(x) = \frac{1}{\pi s\left[1 + \left(\frac{x-c}{s}\right)^2\right]}, \qquad x \in \mathbf{R}.$$

Nem létezik a várható érték. Az eloszlásfüggvény

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x-c}{s}\right).$$

8.64. Megjegyzés. *Szokás csak a $c = 0, s = 1$ esetet (standard) Cauchy-eloszlásnak nevezni.*

5.5. 8.5.5. A véletlen vektorok

8.65. Definíció. *A $(\xi, \eta) : \Omega \to \mathbf{R}^2$ leképezést **(kétdimenziós) véletlen vektor**nak nevezzük, ha*

$$\{\xi < x, \eta < y\} = \{\omega | \omega \in \Omega, \quad \xi(\omega) < x, \eta(\omega) < y\} \in \mathcal{F}, \qquad \forall x, y \in \mathbf{R}.$$

8.66. Definíció. *Az $F(x, y) = P(\xi < x, \eta < y)$ formulával meghatározott valós értékű függvényt a (ξ, η) véletlen vektor **együttes eloszlásfüggvény**ének nevezzük. Az*

$$F_\xi(x) = \lim_{y \to +\infty} F(x, y), \quad F_\eta(y) = \lim_{x \to +\infty} F(x, y)$$

*függvényeket **peremeloszlásfüggvény**nek nevezzük.*

8.67. Tétel. *Az F függvény akkor és csak akkor lehet együttes eloszlásfüggvény, ha*

1. $\lim_{x\to-\infty} F(x,y)=0,\ \lim_{y\to-\infty} F(x,y)=0,$

2. $\lim_{\substack{x\to\infty\\ y\to\infty}} F(x,y)=1,$

3. F mindkét változójában balról folytonos,

4. $F(b,d)-F(b,c)-F(a,d)+F(a,c)\geq 0,\quad \forall a<b,\quad c<d$ *esetén, azaz teljesül az ún. "téglalap" tulajdonság.*

8.68. Megjegyzés. *A téglalap tulajdonságból következik, hogy F mindkét változójában monoton növekvő.*

8.69. Definíció. *A (ξ,η) véletlen vektort **diszkrét**nek nevezzük, ha a lehetséges értékek számossága legfeljebb megszámlálhatóan végtelen.*

8.70. Definíció. *Legyen a ξ, illetve η valószínűségi változó lehetséges értekeinek sorozata $x_1,x_2,\dots$, illetve $y_1,y_2,\dots$. A $P(\xi=x_i,\eta=y_j)=p_{ij}\,(i,j=1,2,\dots)$ valószínűségek sorozatát **együttes eloszlás**nak nevezzük. A*

$$q_i=\sum_{j=1}^{\infty} p_{ij},\quad (i=1,2,\dots),$$

$$r_j=\sum_{i=1}^{\infty} p_{ij},\quad (j=1,2,\dots)$$

*valószínűség sorozatokat **peremeloszlás**nak nevezzük. Minden $r_j>0$ esetén a ξ **feltételes eloszlás**a adott $\eta=y_j$ mellett*

$$P(\xi=x_i|\eta=y_j)=\frac{p_{ij}}{r_j}.$$

Az

$$E(\xi\,|\eta=y_j)=\sum_{i=1}^{\infty} x_i\frac{p_{ij}}{r_j}$$

*mennyiséget **feltételes várható érték**nek nevezzük. Az*

$$E(\xi\,|\eta=y_j)=m_2(y_j)$$

*függvényt a ξ-nek az η-ra vonatkozó **regressziós függvény**ének nevezzük.*

8.71. Tétel. *Ha $p_{ij}\,(i,j=1,2,\dots)$ együttes eloszlás, akkor*

$$p_{ij}\geq 0\quad (i,j=1,2,\dots)\quad \text{és}\quad \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p_{ij}=1.$$

8.72. Definíció. *Ha létezik f nemnegatív valós értékű függvény, melyre*

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dvdu,\quad \forall x,y\in\mathbf{R},$$

*akkor f az F eloszlásfüggvényhez tartozó **együttes sűrűségfüggvény**. Az*

$$f_\xi(x)=\int_{-\infty}^{+\infty} f(x,y)dy,\quad f_\eta(y)=\int_{-\infty}^{+\infty} f(x,y)dx$$

függvényeket ***peremsűrűségfüggvény****nek nevezzük.*

8.73. Tétel. *Az* f *függvény akkor és csak akkor lehet együttes sűrűségfüggvény, ha nemnegatív és*

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dydx = 1.$$

8.74. Definíció. *A* (ξ,η) *véletlen vektort* ***folytonos****nak nevezzük, ha létezik az együttes sűrűségfüggvénye.*

8.75. Definíció. *A* ξ *és* η *valószínűségi változót* ***független****nek nevezzük, ha*

$$F(x,y) = F_\xi(x)F_\eta(y), \qquad \forall x,y \in \mathbf{R}.$$

8.76. Megjegyzés. *A függetlenség megfelelői diszkrét illetve folytonos esetben:*

$$p_{ij} = q_i r_j, \quad (i,j = 1,2,\dots),$$

$$f(x,y) = f_\xi(x)f_\eta(y), \quad \forall x,y \in \mathbf{R}.$$

8.77. Definíció. *Legyen* (ξ,η) *véletlen vektor. Az* $F(x|y)$ *a* ***feltételes eloszlásfüggvény****e a* ξ*-nek* $\eta = y$ *esetén, ha*

$$F(x|y) = P(\xi < x|\eta = y) = \lim_{h\to 0+0} P(\xi < x|y \le \eta < y+h).$$

8.78. Megjegyzés. *Ha léteznek a feltételes valószínűségek.*

8.79. Definíció. *Ha létezik* $f_{\xi|\eta}$ *nemnegatív valós értékű függvény, melyre*

$$F(x|y) = \int_{-\infty}^{x} f_{\xi|\eta}(u|y)du, \quad \forall x,y \in \mathbf{R},$$

akkor $f_{\xi|\eta}$ *a* ξ*-nek az* η*-ra vonatkozó* ***feltételes sűrűségfüggvény****e.*

8.80. Megjegyzés.

$$f_{\xi|\eta}(x|y) = \frac{f(x,y)}{f_\eta(y)}.$$

8.81. Definíció. *A feltételes sűrűségfüggvény segítségével meghatározott feltételes várható értéket* ***regressziós függvény****nek nevezzük, azaz az*

$$\int_{-\infty}^{+\infty} f_{\xi|\eta}(x|y)dx = m_2(y)$$

függvényt a ξ*-nek az* η*-ra vonatkozó regressziós függvényének nevezzük.*

8.82. Megjegyzés. *Ha* (ξ,η) *véletlen vektor és* $g : \mathbf{R}^2 \to \mathbf{R}$ *olyan függvény, hogy* $g(\xi,\eta)$ *valószínűségi változó, akkor*

$$E(g(\xi,\eta)) = \begin{cases} \sum_{i,j} g(x_i,y_j)p_{ij}, & \text{ha } (\xi,\eta) \text{ diszkrét,} \\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)dydx, & \text{ha } (\xi,\eta) \text{ folytonos.} \end{cases}$$

8.83. Definíció. *A*

$$cov(\xi,\eta) = E((\xi - E(\xi))(\eta - E(\eta)))$$

mennyiséget ***kovarianciá****nak nevezzük. Az*

$$r(\xi,\eta)=\frac{cov(\xi,\eta)}{D(\xi)D(\eta)}$$

mennyiséget pedig ***korrelációs együtthatónak*** *nevezzük.*

8.84. Tétel.

1. $E(\xi+\eta)=E(\xi)+E(\eta)$.

2. $D^2(\xi+\eta)=D^2(\xi)+D^2(\eta)+2cov(\xi,\eta)$.

3. $E(E(\xi|\eta=y))=E(\xi)$.

4. $|cov(\xi,\eta)|\le D(\xi)D(\eta)$, *azaz* $|r(\xi,\eta)|\le 1$.

8.85. Megjegyzés. *A véletlen vektor és a hozzákapcsolódó fogalmak definícióját csak kétdimenziós esetben adtuk meg, de nagyon egyszerűen kiterjeszthetőek véges sok valószínűségi változó esetére. Például, a* $\xi_1,\xi_2,\dots,\xi_n$ *valószínűségi változókat függetlennek nevezzük, ha*

$$F(x_1,x_2,\dots,x_n)=F_{\xi_1}(x_1)F_{\xi_2}(x_2)\cdots F_{\xi_n}(x_n),\quad \forall x_1,x_2,\dots,x_n\in\mathbf{R}.$$

8.86. Tétel. *Az* $F(x_1,x_2,\dots,x_n)$ *függvény akkor és csak akkor együttes eloszlásfüggvény, ha minden változójában balról folytonos, és*

$$\lim_{x_i\to-\infty}F(x_1,x_2,\dots,x_n)=0,\qquad (i=1,2,\dots,n),$$

$$\lim_{x_i\to+\infty(i=1,2,\dots,n)}F(x_1,x_2,\dots,x_n)=1,$$

$$\sum_{K=e_1+e_2+\cdots+e_n}(-1)^K F(e_1a_1+(1-e_1)b_1,\dots,e_na_n+(1-e_n)b_n)\ge 0,$$

$$\forall a_i\le b_i,\qquad (i=1,2,\dots,n)$$

és az összegzést $\forall K$ *esetében vesszük, ahol az*

$$e_1,e_2,\dots,e_n$$

értéke 0 *és* 1 *lehet.*

8.87. Tétel. *Legyenek* $\xi_1,\xi_2,\dots,\xi_n$ *független valószínűségi változók, melyeknek rendre* $F_{\xi_1},F_{\xi_2},\dots,F_{\xi_n}$ *az eloszlásfüggvénye. Ekkor*

(a) az $\eta(\omega)=max\{\xi_1(\omega),\dots,\xi_n(\omega)\}\ (\forall\omega\in\Omega)$ *valószínűségi változó eloszlásfüggvénye*

$$F_\eta(y)=F_{\xi_1}(y)F_{\xi_2}(y)\cdots F_{\xi_n}(y).$$

(b) az $\eta(\omega)=min\{\xi_1(\omega),\dots,\xi_n(\omega)\}\ (\forall\omega\in\Omega)$ *valószínűségi változó eloszlásfüggvénye*

$$F_\eta(z)=1-(1-F_{\xi_1}(z))(1-F_{\xi_2}(z))\cdots(1-F_{\xi_n}(z)).$$

5.6. 8.5.6. Néhány többdimenziós eloszlás

A (ξ,η) véletlen vektor

(i) normális eloszlású, ha

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp[-Q],$$

$$Q = \frac{1}{2(1-\rho^2)}\left[(\frac{x-m_1}{\sigma_1})^2 - 2\rho(\frac{x-m_1}{\sigma_1})(\frac{y-m_2}{\sigma_2}) + (\frac{y-m_2}{\sigma_2})^2\right],$$

ahol $\sigma_1 > 0, \sigma_2 > 0, -1 < \rho < 1.$

(ii) egyenletes eloszlású az $A \subset \mathbf{R}^2$ tartományon, ha

$$f(x,y) = \begin{cases} \frac{1}{|A|}, & \text{ha } (x,y) \in A, \\ 0, & \text{egyébként.} \end{cases}$$

5.7. 8.5.7. Néhány alapvető tétel

8.88. Tétel. (Markov-egyenlőtlenség) *Legyen a ξ nemnegatív valószínűségi változó, melynek létezik a várható értéke, ekkor $\forall c > 0$ esetén*

$$P(\xi \geq c) \leq \frac{E(\xi)}{c}.$$

8.89. Tétel. (Csebisev-egyenlőtlenség) *Ha a ξ valószínűségi változónak létezik a szórásnégyzete, akkor $\forall \varepsilon > 0$ esetén*

$$P(|\xi - E(\xi)| \geq \varepsilon) \leq \frac{D^2(\xi)}{\varepsilon^2}.$$

8.90. Tétel. (nagy számok gyenge törvénye) *Legyen $\xi_1, \xi_2, \dots$ független, azonos eloszlású valószínűségi változók sorozata. Létezik a szórásnégyzet. Ekkor tetszőleges $\varepsilon > 0$ esetén*

$$\lim_{n\to+\infty} P\left(|\frac{\xi_1 + \dots + \xi_n}{n} - E(\xi_1)| \geq \varepsilon\right) = 0.$$

8.91. Megjegyzés. *Legyen A esemény és S_n az A esemény gyakorisága az első n kísérletből egy Bernoulli kísérletsorozatnál. Ekkor tetszőleges $\varepsilon > 0$ esetén*

$$\lim_{n\to+\infty} P\left(|\frac{S_n}{n} - P(A)| \geq \varepsilon\right) = 0.$$

8.92. Tétel. (centrális határeloszlás-tétel) *Legyen $\xi_1, \xi_2, \dots$ független, azonos eloszlású valószínűségi változók sorozata és létezik az $E(\xi_i) = \mu$ és $D^2(\xi_i) = \sigma^2 > 0$. Ha $S_n = \sum_{k=1}^{n} \xi_k$, akkor*

$$\lim_{n\to+\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} < x\right) = \Phi(x), \qquad x \in \mathbf{R},$$

ahol Φ a standard normális eloszlásfüggvény.

8.93. Tétel. (Moivre-Laplace) *Legyen a ξ valószínűségi változó binomiális eloszlású n és p paraméterrel és $0 \leq a < b \leq n$ egész, akkor*

$$P(a \leq \xi \leq b) = \sum_{k=a}^{b} \binom{n}{k} p^k q^{n-k} \approx$$

$$\approx \Phi\left(\frac{b - np + \frac{1}{2}}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np - \frac{1}{2}}{\sqrt{npq}}\right).$$

Irodalomjegyzék

[1] J. Aczél, Z. Daróczy. *On Measures of Information and Their Characterization*. Academic Press, New York. 1975.

[2] S. Arimoto. *An algorithm for calculating the capacity of an arbitrary discrete memoryless channel*. IEEE Trans. Inform. Theory, IT-18. 1972. 14–20.

[3] R. B. Ash. *Information Theory*. Interscience, New York. 1965.

[4] J. Berstel, D. Perrin. *Theory of Codes*. Academic Press, New York. 2002.

[5] G. Birkhoff, T.C. Bartee. *A modern algebra a számítógéptudományban*. Műszaki Könyvkiadó, Budapest. 1964.

[6] R. Blahut. *Computation of channel capacity and rate distortion functions*. IEEE Trans. Inform. Theory, IT-18. 1972. 460–472.

[7] T. M. Cover, J.A. Thomas. *Elements of information theory*. Wiley, New York. 1991.

[8] Csiszár I., Fritz József. *Információelmélet*. Tankönyvkiadó, Budapest. 1980.

[9] Fritz József. *Bevezetés az információelméletbe*. Tankönyvkiadó, Budapest. 1971.

[10] Fritz József. *Információelmélet*. Mat.Kut.Int., Budapest. 1973.

[11] Fülöp Géza. *Az információ*. Eötvös Loránd Tudományegyetem Könyvtártudományi - Informatikai Tanszék, Budapest. 1996.

[12] S. Guiasu. *Information theory with applications*. McGRAW-HILL, New York. 1977.

[13] Sz. V. Jablonszkij, O.B. Lupanov. *Diszkrét matematika a számítástudományban*. Műszaki Könyvkiadó, Budapest. 1980.

[14] M. Jimbo, K. Kunisawa. *An Iteration Method for Calculating the Relative Capacity*. Department of Information Sciences, Faculty os Sience and Technology, Sience University of Tokyo, Noda City Chiba 278, Japan.

[15] F. M. Reza. *Bevezetés az információelméletbe*. Műszaki Könyvkiadó, Budapest. 1963.

[16] C. E. Shannon, W.Weaver. *A kommunikáció matematikai elmélete*. OMIKK, Budapest. 1986.

[17] Vassányi István. *Információelmélet*. Veszprémi Egyetem, Műszaki Informatika Szak. 2002-2005.

[18] Xue-Bin Liang. *An Algebraic, Analytic and Algorithmic Investigation on the Capacity and Capacity-Achieving Input Probability Distributions of Finite-Input Finite-Output Discrete Memoryless Channels*. Department of Electrical and Computer Engineering Louisiana State University, Baton Rouge, LA 70803. 2004.