INFORMÁCIÓELMÉLET

Összefoglaló – segédlet

Készítette: Fegyverneki Sándor

Miskolci Egyetem, 2002.

TARTALOMJEGYZÉK

1. 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 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 adattárolláskor 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 definiál-

ható 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 információmennyiségének 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.

Az egyik 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 adódík 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 tartalmazza.

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

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

$$\text{forrás} \rightarrow \text{kódoló} \rightarrow \text{csatorna} \rightarrow \text{dekódoló} \rightarrow \text{felhasználó}$$

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).

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 valószínűségszámításon 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 információmennyiségének mérése és ennek kapcso-

lata más matematikai területekkel. A hírközlési rendszerek matematikai modellje(zajos, zajmentes vagy diszkrét, folytonos). Kódoláselmélet (zajos, zajmentes; forrás, csatorna).

2. AZ INFORMÁCIÓMENNYISÉG

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.

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?*

Példa: Az amerikai hadseregnél állítólag úgy végzik a vérbajosok felkutatását, hogy az egész társaságtól vért vesznek, és a páciensek felé-

nek 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.

Tekintsünk 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 $\dfrac{k}{m}$. 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 ($1 bit$) mennyiségű információra. Néhány szerző az e alapú

természetes logaritmust preferálja, ekkor az egység a *nat.* A logaritmusok közötti átváltás alapján $1bit = \ln 2nat$.

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 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

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

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.

2.1.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.

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

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

2.2.Definíció: A $\mathcal{P} = \{p_1, p_2, \ldots, p_n\}$ eloszlású ξ valószínűségi változó *Shannon-féle entrópiá*jának nevezzük a

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

mennyiséget.

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 *(l. Függelék)*.

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.TÉTEL: **(Az entrópia tulajdonságai:)**

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

Bizonyítás: Az összeg minden tagja nemnegatív.

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 $q_n = p_1 + p_2 + \ldots + p_m$, 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:

Megjegyzés: Az entrópia axiomatikus származtatása: Ha a fenti tulajdonságok közül megköveteljük, hogy

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

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

(3) Ha $0 \leq \lambda \leq 1$, $\bar{\lambda} = 1 - \lambda$, akkor

$$H_{n+1}(p_1, p_2, \ldots, \lambda p_n, \bar{\lambda} p_n) = H_n(p_1, p_2, \ldots, p_n) + p_n H_2(\lambda, \bar{\lambda}).$$

3. Az I-divergencia

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 bizontalanság 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*ének vagy a $\mathcal{P}$ eloszlásnak a $\mathcal{Q}$ eloszlással való helyettesítésénél fellépő információnyereségnek nevezzük.

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

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 hipotézis-

vizsgálat felhasználásával. Szokásos elnevezés még az információ divergencia vagy I-divergencia.

3.1.TÉTEL: **(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:

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

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 például 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 nyilnánvaló. Ha $a \neq 0$, akkor legyen

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

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 feledatnak 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.

3.2 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.

3.1.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

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

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

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

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

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

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

3.2.Definíció: Ha $A_1, A_2, \ldots, A_n$ és $B_1, B_2, \ldots, B_m$ teljes eseményrendszerek, amelyekre $P(A_i) = q_i$ $(1 \le i \le n)$, $P(B_j) = r_j$ $(1 \le j \le m)$ 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.

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

3.3 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

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

3.3.Definíció: Legyen $\hat{p}_i = \dfrac{k_i}{K}$ $(1 \leq i \leq n)$, 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.

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

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

így

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

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.

3.2.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

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

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

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

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

4. FORRÁSKÓDOLÁS

A Shannon-féle egyirányú hirközlési modell általános alakja:

$$\text{forrás} \rightarrow \text{kódoló} \rightarrow \overset{\text{zaj}}{\text{csatorna}} \rightarrow \text{dekódoló} \rightarrow \text{felhasználó}$$

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.

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.

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

Forrás: $X = \{x_1, \ldots, x_n\}$ – véges halmaz, forrásábécé.

üzenet

emlékezetnélküli forrás

stacionárius forrás

csatornaábécé vagy kódábécé – $Y = \{y_1, \ldots, y_s\}$

csatorna

zajmentes csatorna

Kódolási eljárás:

$\mathcal{X}$ – az X-ből készített véges sorozatok halmaza

$\mathcal{Y}$ – az Y-ból készített véges sorozatok halmaza

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

– betűnkénti

– blokkonkénti

– állandó hossz

– változó hossz

A dekódolhatóság: egyértelmű dekódolhatóság, gazdaságossági szempontok. Egyértelműen dekódolható – különböző közlemények kódközleményei is különbözők.

Zajmentes csatorna + betűnkénti kódolás:

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

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

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

Megjegyzés: Az állandó kódhosszú kód mindig prefix, ha a kódszavai különbözőek.

4.1.ÁLLÍTÁS: Minden prefix kód egyértelműen dekódolható.

Sardinas-Patterson módszer:

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, $m = 0, 1, 2, \ldots$ hal-

mazokat. Legyen $S_0 = \mathcal{K}$. Az S_{m+1} halmazt az S_m halmaz szavai után következő szavak halmazaként definiáljuk.

4.2.ÁLLÍTÁS: A $\mathcal{X}$ kód akkor és csak akkor egyértelműen dekódolható, ha az S_i, $(i \geq 1)$ halmazok nem tartalmaznak kódszót, azaz S_0 egy elemét.

Keresési stratégiák és prefix kódok

4.1.TÉTEL: Az s alternatívás keresési stratégiára

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

Jelölések:

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

4.3.ÁLLÍTÁS: A kódfák és a prefix kódok között kölcsönös egy-egyértelmű megfeleltetés van.

Megjegyzés: A prefix kód átlagos kódhossza nem lehet kisebb, mint

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

4.2.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.

4.3.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.

4.4.TÉTEL: Létezik prefix kód, hogy

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

Bizonyítás: konstrukciós bizonyítás – Shannon-Fano kód.

Gilbert-Moore kód: Nincs szükség sorbarendezésre. Ekkor

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

4.x1.Definíció: Az egyértelműen dekódolható kód hatásfoka:

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

4.x2.Definíció: A kódot optimálisnak nevezzük, ha egyértelműen dekódolható és maximális hatásfokú.

4.4.ÁLLÍTÁS: Adott $\mathcal{P}$ eloszlású forrásábécé s számú kódjelből alkotott egyértelműen dekódolható kódjai között mindig van optimális.

Huffmann-kód – maximális hatásfokú prefix kód.

Tulajdonságok:

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

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.

Megjegyzés: Optimális, bináris kódfa teljes.

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

Megjegyzés: Összevonási algoritmus. Az optimális kódfa minden végponttól különböző szögpontjából s él indul ki, kivéve esetleg egy vépont 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.

4.5.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.

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

A bizonyítás lépései:

1. A kód egyértelműen dekódolható.

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

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

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

4.5.ÁLLÍTÁS: 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

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

Blokkos kódolás:

$$g : X^k \to \mathcal{Y}, \quad L = \frac{L^{(k)}}{k}.$$

Emlékezetnélküli, stacionárius forrás.

Az optimális kódra:

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

A függetlenségből

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

s így

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

Kompresszió:

$$X = Y, \qquad 1 \ge \frac{H(\mathcal{P})}{\log_2 s}.$$

Kölcsönös információmennyiség:

4.6.ÁLLÍTÁS:

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

Feltételes entrópia:

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

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

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

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

egyenlőség függetlenség esetén.

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

Stacionér forrás entrópiája:

4.7.ÁLLÍTÁS: $H(\xi|\eta) \le H(\xi|f(\eta))$.

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.

4.8.ÁLLÍTÁS: 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

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

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

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

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

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.

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

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

Csatornakapacitás (emlékezetnélküli eset):

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

Megjegyzés: 1. Legyen C a kapacitás, L az átlagos kódhossz. Ha $H(\mathcal{P}) \leq LC$, akkor továbbíthatjuk a forrás által szolgáltatott közleményeket.

2. Zajmentes és emlékezetnélküli csatornára:

$$C = \log_2 s.$$

3. Bináris szimmetrikus csatorna:

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

Milyenek a bemeneti illetve kimeneti eloszlások?

4.6.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

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

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.

5. CSATORNAKAPACITÁS

Az eddigiek alapján tudjuk, hogy zajmentes csatorna esetén az egy csatornajelre (kódábécébeli elemre) jutó átlagos információ átvitel megegyezik a kódábécé eloszlásához kapcsolódó entrópával, azaz $H(\mathcal{Q})$, 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.

3.1 Zajmentes csatorna kapacitása nem azonos átviteli idő esetén

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

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 $\mathcal{Q} = \{q_1, \ldots, q_s\}$. Ha a csatornaüzenet hossza N, akkor egy kiválasztott jel pl. y_i várhatóan Nq_i-szer fordul elő és a 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)

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

5.1.Definíció: Az információtovábbítás sebességének nevezzük a

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

mennyiséget.

Megjegyzés: Az előző definícióban szereplő mennyiség átlagsebesség.

5.2.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.

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

5.3.TÉTEL: Zajmentes, emlékezetnélküli (véges, diszkrét) csatorna esetén a csatornakapacitás a

$$\sum_{i=1}^{s} 2^{-vt_i} = 1 \tag{5.1}$$

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

$$q_i = 2^{-Ct_i} \qquad (i = 1, \ldots, s) \tag{5.2}$$

eloszlás realizálja.

Bizonyítás: Ha C megoldása az (5.1) egyenletnek, akkor az (5.2) szerint megadott sorozat 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 az (5.2) esetben teljesül.

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

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

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.

3.2 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:

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

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 csatorna mátrix 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 ugyanabból 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

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

Megjegyzés: Ha a kimeneti eloszlás egyenletes, akkor a bemeneti is.

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 alkalmazzuk. 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 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

$$2^{-\sum_{k=1}^{s} H_k \tau_{ki}} = r_i 2^{\frac{1}{\ln 2} - \lambda}, \quad i = 1, \ldots, s, \tag{5.2.1}$$

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

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

Az (5.2.1) egyenletrendszerből

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

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

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

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

Megjegyzés: 1. Ha létezik $q_k = 0$, akkor problémás a megoldás első része.

2. $I(\xi, \eta)$ 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 $q_k = 0$.

5.4.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.

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

5.5.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

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

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

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

A feltételek alapján viszont

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

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

Példa:

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

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

A parciális deriváltakból adódó egyenletrendszer.

$$-\frac{1}{\ln 2} - 0.9 \log_2 r_1 - 0.1 \log_2 r_2 - H_1 + \lambda = 0$$
$$-\frac{1}{\ln 2} - 0.2 \log_2 r_1 - 0.8 \log_2 r_2 - H_2 + \lambda = 0$$

Átalakítva

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

Legyen

$$u_2 = \frac{1}{\ln 2} - \lambda + \log_2 r_1$$
$$u_1 = \frac{1}{\ln 2} - \lambda + \log_2 r_2$$

Ekkor

$$u_2 \approx -0.7941945,$$
$$u_1 \approx -0.4328624,$$
$$C = \log_2 (2^{u_1} + 2^{u_2}) \approx 0.397754.$$

Továbbá

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

$$r_2 \approx 0.4377112,$$
$$r_1 \approx 0.5622888,$$
$$q_1 \approx 0.517555,$$
$$q_2 \approx 0.48445.$$

6. CSATORNAKÓDOLÁS

Szimmetrikus bináris csatorna esete:

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

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

Példa: A bit háromszorozós módszer(Commodore-64 kazettás egység):

$$0 \to 000$$

$$1 \to 111$$

Ha a dekódolás a több azonos bit szerint történik, akkor 2 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: üzenetszó(k bit)$\to$ kódszó(n bit) átalakítást(kódolást) (k, n) kódnak nevezzük.

Legyen $A = \{0, 1\}$ és adott a következő két műveleti tábla.

A "kizáró vagy" művelet (jele:$\vee$) vagy másképpen a modulo 2 összeadás (jele:$\oplus$), és a hagyományos szorzás a két 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 a $(A, \vee, \cdot)$ test felett.

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

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

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

6.4.ÁLLÍTÁS: A Hamming-féle távolság kielégíti a távolság tulajdonságait.

6.5.ÁLLÍTÁS: A Hamming-féle távolság invariáns az eltolásra, azaz

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

Bizonyítás:

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

6.6.Definíció: A $v_1, v_2, \ldots, v_m$ kódszavakból álló kód esetén a kódszavak távolságai közül a minimálisat kódtávolságnak nevezzük.

Megjegyzés: Legyen $d = \min\limits_{i \neq j} d(v_i, v_j)$, 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.

6.7.ÁLLÍTÁS: Legyen egy vett szóban a hibák száma legfeljebb r. Tetszőleges kódszó esetén a legfeljebb r számú hiba a minimális távolságon alapuló hibajavítás módszerével javítható akkor és csak akkor, 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

$$d(\tilde{v}, v_i) < d - r \Rightarrow d > 2r, \text{ azaz } d > 2r + 1.$$

6.8.Definíció: Ha a kódszavak csoportot alkotnak a kódot csoportkódnak nevezzük.

Példa: Adottak a következő (2,5) kódok:

6.9.ÁLLÍTÁS: Csoportkódban a kódszó alakú hibavektor esetén a hiba nem jelezhető és nem javítható. A nem kódszó alakú hiba legalább jelezhető.

6.10.ÁLLÍTÁS: Csoportkód esetén a hibaáteresztés valószínűsége megegyezik a csupa zérus kódszó alakú hibák valószínűségének az összegével.

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.

6.11.ÁLLÍTÁS: Egy (k,n) csoportkód esetén $d = \min_{v_i \neq 0} ||v_i||$.

6.12.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 eltolás invariáns, ezért

$$\begin{aligned} d(v_k \vee e, v_j) &= d((v_k \vee e) \vee (v_k \vee v_i), v_j \vee (v_k \vee v_i)) = \\ &= d(v_i \vee e, v_j \vee (v_k \vee v_i)) \geq d(v_i \vee e, v_i) = \\ &= d((v_i \vee e) \vee (v_k \vee v_i), v_i \vee (v_k \vee v_i)) = d(v_k \vee e, v_k). \end{aligned}$$

Megjegyzés: Hogyan lehetne automatizálni a következő problémákat? 1. A csoport tulajdonság ellenőrzése. 2. Tárolás, kódszó keresés. 3. Kódtávolság kiszámítás.

6.13.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.

Példa:

$$G = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

Éppen az (A) csoportkódot adja meg.

6.14.ÁLLÍTÁS: A lineáris kód csoportkód.

6.15.Definíció: Legyen E $k \times k$ típusú egységmátrix. A $G = (E|P)$ generátor mátrixú kódot szisztematikus kódnak nevezzük.

Néhány elnevezés: P paritásmátrix, $F = \begin{pmatrix} P \\ E \end{pmatrix}$ paritásellenőrző mátrix ($E_{(n-k)\times(n-k)}$), $u^T P$ paritásvektor.

6.16.Definíció: Legyen $\tilde{v}$ egy vett kódszó a csatornakimeneten. Az s vektort a $\tilde{v}$ vektorhoz tartozó szindrómának nevezzük, ha

$$s^T = \tilde{v}^T F.$$

6.17.ÁLLÍTÁS: A szindróma akkor és csak akkor zérusvektor, ha a vett szó kódszó.

Megjegyzés: A szindrómák egy osztályozást adnak.

6.18.ÁLLÍTÁS: 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 (osztályelsők-mimimális normájú osztályelemek kiválasztása).

6.18.ÁLLÍTÁS: A dekódolási táblázatban bármely szó távolsága a saját oszlopa tetején álló kódszótól nem nagyobb, mint bármely más kódszótól.

Megjegyzés: 1. A dekódolási táblázat alkalmas maximum likelihood kódolásra.

2. Az osztályelső alakú hibák javíthatók.

3. A helyes dekódolás valószínűsége megegyezik az osztályelső alakú hibák valószínűségeinek az összegével.

7. FOLYTONOS ESET

A ξ folytonos valószínűségi változó lehetséges értékeinek halmaza nemmegszámlálható, 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

$$P(\xi_\delta = n\delta) = P((n-1)\delta < \xi \leq n\delta) = \int_{(n-1)\delta}^{n\delta} f(x)dx = \delta\tilde{f}(n\delta),$$

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

$$\sum_{n=-\infty}^{+\infty} \delta\tilde{f}(n\delta) = \int_{-\infty}^{+\infty} f(x)dx = 1.$$

Ha $\delta \to 0$, akkor $-\ln\delta \to +\infty$, ezért

$$\lim_{\delta\downarrow 0}(H(\xi_\delta) + \ln\delta) = -\int_{-\infty}^{+\infty} f(x)\ln f(x)dx = H(\xi).$$

Példa: Legyen ξ a $0, \vartheta)$ intervallumon egyenletes eloszlású. Ekkor

$$H(\xi) = \ln \vartheta,$$

azaz jól látható, hogy a $H(\xi)$ lehet negatív is.

Megjegyzés: Az entrópia diszkrét esetben a bizonytalanságot méri, míg folytonos esetben csak a bizonytalanság változását.

Néhány fogalom folytonos esetben:

Entrópia, mint várható érték: $H(\xi) = E(-\ln f(\xi))$.

Példa: 1. Exponenciális eloszlásra: $H(\xi) = 1 - \ln \lambda$. 2. Normális eloszlásra: $H(\xi) = 0.5 + \ln(\sigma\sqrt{2\pi})$.

Együttes entrópia: $H(\xi, \eta) = E(-\ln f(\xi, \eta))$.

Példa: Normális eloszlásra: $H(\xi, \eta) = 1 + \ln(2\pi\sigma_1\sigma_2\sqrt{1 - r^2})$.

Kölcsönös információmennyiség:

$$I(\xi, \eta) = E\left(\ln \frac{f(\xi, \eta)}{f_\xi(\xi) f_\eta(\eta)}\right).$$

Példa: Normális eloszlásra: $I(\xi, \eta) = -0.5 \ln(1 - r^2)$.

I-divergencia:

$$D(\eta||\xi) = E\left(\ln \frac{f_\eta(\eta)}{f_\xi(\eta)}\right).$$

Megjegyzés: $D(\eta||\xi) \geq 0$.

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

$$\begin{aligned} H(\eta) &= -\int_{-\infty}^{+\infty} f_\eta(y) \ln f_\eta(y) dy = \\ &= -\int_{-\infty}^{+\infty} f_\xi(x) \ln \frac{f_\xi(x)}{|g'(x)|} dx = \\ &= -\int_{-\infty}^{+\infty} f_\xi(x) \ln f_\xi(x) dx + \int_{-\infty}^{+\infty} f_\xi(x) \ln |g'(x)| dx. \end{aligned}$$

Maximum entrópia módszer (MEM):

Entrópia maximalizálás feltételek mellett.

Példa: 1. 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:

$$-E(\ln g(\xi)) \leq \sum_{i=1}^{n} \lambda_i a_i - \ln A.$$

FÜGGELÉK

F1. VALÓSZÍNŰSÉGSZÁMÍTÁS

1.Definíció: Egy véletlen kísérlet lehetséges eredményeinek összességét *minta tér*nek nevezzük. Jele: Ω. Az Ω elemeit *elemi esemény*eknek nevezzük.

2.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) $A_1, A_2, \ldots \in \mathcal{F}$, akkor $A_1 \cup A_2 \cup \ldots \in \mathcal{F}$.

Az $\mathcal{F}$ elemeit pedig *esemény*eknek nevezzük.

Megjegyzés: Ha $A, B \in \mathcal{F}$, akkor $A \cap B \in \mathcal{F}$.

3.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.

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.

4.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

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

1.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) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

1.TÉTEL: **(Poincaré)** Az A_1, $A_2, \ldots,$ A_n, eseményekre

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{i_1 < i_2 < \ldots < i_k} P(\bigcap_{j=1}^{k} A_{i_j}),$$

ahol az összegzést az összes lehetséges $\{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots, n\}$ esetre tekintjük.

5.Definíció: Az A esemény B feltétel melletti *feltételes valószínűség*ének nevezzük a

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

mennyiséget, ha $P(B) > 0$.

Megjegyzés: A $P(\cdot|B) : \mathcal{F} \to \mathbf{R}$ leképezés tényleg valószínűség.

2.LEMMA: Ha az $A_1, A_2, \ldots, A_n$ eseményrendszerre $P(\bigcap_{i=1}^{n-1} A_i) > 0$, akkor

$$P(\bigcap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1) \ldots P(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1}).$$

6.Definíció: Az $A_1, A_2, \ldots$ eseményrendszert *teljes eseményrendszer*nek nevezzük, ha $A_i \cap A_j = \emptyset$, ha $i < j$ és $i, j = 1, 2, \ldots$, és $\bigcup_{i=1}^{\infty} A_i = \Omega$.

2.TÉTEL: Ha $A_1, A_2, \ldots$ teljes esményrendszer és $P(A_i) > 0$, ha $i = 1, 2, \ldots$, akkor tetszőleges B esemény esetén

$$P(B) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i).$$

3.TÉTEL: **(Bayes)** Ha $A_1, A_2, \ldots$ teljes esményrendszer és $P(A_i) > 0$, ha $i = 1, 2, \ldots$, akkor tetszőleges pozitív valószínűségű B esemény esetén

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{\infty} P(B|A_i)P(A_i)}.$$

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.

7.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

$$P(A_i \cap A_j) = P(A_i)P(A_j) \qquad (1 \le i < j \le n).$$

Az $A_1, A_2, \ldots, A_n$ eseményeket *teljesen sztochasztikusan független*nek nevezzük, ha

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \ldots P(A_{i_k})$$
$$(1 \le i_1 < \ldots < i_k \le n, \quad 2 \le k \le n).$$

Az $\{A_1, A_2, \ldots, A_n\}$ és $\{B_1, B_2, \ldots, B_m\}$ eseményrendszereket *sztochasztikusan független*nek nevezzük, ha

$$P(A_i \cap B_j) = P(A_i)P(B_j)$$

$$(1 \le i \le n, \quad 1 \le j \le m).$$

Példa: Ha az A és B események függetlenek, akkor $\overline{A}$ és B, A és $\overline{B}$ és $\overline{A}$ és $\overline{B}$ is függetlenek, azaz az $\{A, \overline{A}\}$ és $\{B, \overline{B}\}$ eseményrendszerek is függetlenek.

3.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:

$$P(\bigcup_{i=1}^{n} A_i) =$$

$$= P(\overline{\bigcup_{i=1}^{n} A_i}) = 1 - P(\overline{\bigcup_{i=1}^{n} A_i}) = 1 - P(\bigcap_{i=1}^{n} \overline{A_i}) = 1 - \prod_{i=1}^{n} P(\overline{A_i}).$$

8.Definíció: A $\xi :\rightarrow \mathbf{R}$ leképezést diszkrét *valószínűségi változó*nak nevezzük, ha $\xi(\Omega) = \{x_1, x_2, \ldots\}$ lehetséges értékek halmaza legfeljebb megszámlálhatóan végtelen számosságú és

$$\xi^{-1}(x_i) = \{\omega | \omega \in \Omega, \quad \xi(\omega) = x_i\} \in \mathcal{F} (i = 1, 2, \ldots).$$

9.Definíció: A $\{p_i = P(\xi^{-1}(x_i)), i = 1, 2, \ldots\}$ számok összességét a ξ valószínűségi változó *eloszlás*ának nevezzük.

10.Definíció: A $\sum_{i=1}^{\infty} x_i p_i$ mennyiséget várható értéknek nevezzük, ha $\sum_{i=1}^{\infty} |x_i|\, p_i < +\infty$. Jele: $E(\xi)$.

11.Definíció: *Bernoulli kísérletsorozat*nak nevezzük, 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.

F2. KONVEX FÜGGVÉNYEK

k1.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 $\mu \geq 0$.

k2.TÉTEL: 1. Ha f és g konvex függvény és $\alpha \geq 0$, $\beta \geq 0$, akkor $\alpha f + \beta g$ szintén konvex.

2. Véges sok konvex függvény összege is konvex.

3. Konvex függvények egy konvergens sorozatának a (pontonkénti) határa is konvex.

4. 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 azok.

5. 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.

6. Az $f : U \to \mathbf{R}$ konvex függvény folytonos az intervallum minden belső pontjában.

7. Legyen U nyílt és f kétszer differenciálható, akkor f konvex akkor és csak akkor, ha $f'' > 0$ minden $x \in U$.

k3.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

$$E(f(\xi)) \geq E(L(\xi)) = L(E(\xi)) = f(E(\xi)).$$

Megjegyzés: $E(\xi^2) \geq E^2(\xi)$.

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.

$f'(x) = 1 + \ln x$, amiből látható, hogy $x < e^{-1}$ esetén $f'(x) < 0$, azaz f monoton csökkenő a $(0, e^{-1})$ szakaszon. 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

$$\lim_{x \to 0+0} x \ln x = \lim_{n \to +\infty} \frac{1}{n} \ln \frac{1}{n} = \lim_{n \to +\infty} -\ln \sqrt[n]{n} = 0.$$

k4.ÁLLÍTÁS:

$$1 - \frac{1}{x} \leq \ln x \leq x - 1.$$

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 $\dfrac{1}{x} > 0$ is teljesül, azaz

$$\ln \frac{1}{x} \leq \frac{1}{x} - 1,$$

ami ekvivalens azzal, hogy

$$\ln x \geq 1 - \frac{1}{x}.$$

k5.ÁLLÍTÁS: Ha $\{a_n > 0\}$ sorozat és $a_n \to a$, akkor

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k=1}^{n} a_k = a \quad \text{és} \quad \lim_{n\to+\infty} \sqrt[n]{\prod_{k=1}^{n} a_k} = a.$$

k6.ÁLLÍTÁS:

$$\lim_{n\to+\infty} \frac{n}{\sqrt[n]{n!}} = e.$$

k7.ÁLLÍTÁS: **(Aszimptotikus Stirling-formula)**

$$\lim_{n\to+\infty} \frac{\ln n!}{n \ln n - n} = 1.$$

IRODALOMJEGYZÉK

[1] G. Birkhoff–T.C.Bartee: *A modern algebra a számítógéptudományban,* Műszaki Könyvkiadó, Budapest, 1964.

[2] Csiszár I.– Fritz József: *Információelmélet,* Tankönyvkiadó, Budapest, 1980.

[3] Fritz József: *Bevezetés az információelméletbe,* Tankönyvkiadó, Budapest, 1971.

[4] Fritz József: *Információelmélet,* Mat.Kut.Int., Budapest, 1973.

[5] Sz.V. Jablonszkij–O.B. Lupanov: *Diszkrét matematika a számítástudományban,* Műszaki Könyvkiadó, Budapest, 1980.

[6] C.E. Shannon–W.Weaver: *A kommunikáció matematikai elmélete,* OMIKK, Budapest, 1986.

[7] F.M. Reza: *Bevezetés az információelméletbe,* Műszaki Könyvkiadó, Budapest, 1963.