ELŐSZÓ

Az MTA Matematikai Kutató Intézete 1972 folyamán tanfolyamot szervezett a "Számitástechnikai matematika alapjai" cimmel. Ennek keretében hangzott el Csiszár Imrének az információelmélet alapfogalmait és számitástechnikai alkalmazásait ismertető előadássorozata; ez a jegyzet ezeknek az előadásoknak az anyagát tartalmazza kissé kibővitett formában. Az információelmélet módszereinek magyar nyelvü ismertetését nagyon időszerüvé teszi az a körülmény, hogy ezek a módszerek mind jelentősebb szerepet játszanak különféle fontos gyakorlati problémák megoldásakor. Ezek közül máig is a hirközlési rendszerek tervezésével kapcsolatos alkalmazások a legfontosabbak - az ezirányu gyakorlati igény hozta létre az információelméletet huszegynéhány évvel ezelőtt. Azóta azonban az információelmélet módszerei számos más területre is behatoltak, például bonyolult adatfeldolgozó rendszerek kidolgozásakor sem nélkülözhetők. Természetesen, nem törekedhettünk teljességre. Például, az időben folytonos, vagy analóg jelekkel müködő távközlési csatornák - köztük a Gauss-csatornák - elméletének, valamint a szekvenciális dekódolás módszereinek tárgyalására a korlátozott terjedelem, és a szükséges matematikai fogalmak és eredmények bonyolultsága miatt nem kerülhetett sor. Mindenesetre, a feldolgozott anyag jó alapot biztosit az ilyen speciális jellegü kérdések tanulmányozásához.

A közölt eredmények és bizonyitások megértése különösebb matematikai előismereteket nem tételez fel; elég az analizis elemeinek, valamint a valószinüségszámitás legegyszerübb alapfogalmainak (diszkrét valószinüségi változók, feltételes valószinüség, várható érték) ismerete. A szokásos egyetemi előadások ezeknél lényegesen többet tartalmaznak. Meg kell jegyezni, hogy a 2.6, 2.7., valamint a 3.3. és a 3.4, 4.3. szakaszok anyaga a többinél nehezebben olvasható; első olvasáskor az itt közölt bizonyitásokat nyugodtan ki lehet hagyni.

A jegyzet öt fejezetből áll; az első bevezető jellegü, mig a többi a hirközléselmélet egy-egy területét ismerteti, különös tekintettel a számitástechnikai alkalmazásokra. A képletek számozása fejezetenként ujra kezdődik, de a tételek, lemmák, definiciók számozása folyamatos. A tételben, lemmában vagy definicióban kimondott állitások, valamint a bizonyitások végét +++ jelzi. Az irodalmi hivatkozások sorszáma szögletes zárójelbe került. Az irodalomjegyzék összeállitásakor nem az eredeti forrásmunkák, hanem a jól hozzáférhető tankönyvek és monográfiák felsorolása volt a fő cél. Az információelmélet rendkivül terjedelmes nemzetközi irodalmának tanulmányozásához szeretnénk segitséget nyujtani a szövegközi hivatkozásokkal.

Budapest, 1972. október 24.

Fritz József

T a r t a l o m

I. AZ INFORMÁCIÓ MENNYISÉGÉNEK MÉRŐSZÁMA

1.1. A Shannon-féle entrópia

Napjainkban már eléggé világos, hogy konkrét tartalmától, megjelenési formájától és felhasználhatóságától elvonatkoztatva, beszélhetünk az információ számszerű mennyiségéről, ami éppen olyan pontosan definiálható és mérhető, mint bármely más fizikai mennyiség. Hosszu volt azonban az ut, amely ehhez a felismeréshez vezetett; mindenekelőtt azt kellett tisztázni, hogy egyáltalán mikor van a kérdésnek értelme. Persze mindenkinek van valamilyen - többé kevésbé szubjektiv - elképzelése az információ mennyiségének fogalmáról, de a köznapi szóhasználatban ez általában az információ konkrét megjelenési formájának terjedelmességének, másrészt a hasznosságához, és egyéb tulajdonságaihoz kapcsolódik. Ahhoz, hogy jól használható mérőszámot kapjunk, minden esetleges vagy szubjektiv tényezőtől el kell vonatkoztatni, és ezek közé soroljuk az információ konkrét tartalmát, formáját is; jóformán mindent, ami a köznyelvben az információ fogalmához kapcsolódik. Ezt a könyörtelen absztrakciót az indokolja, hogy az információ megszerzésével, feldolgozásával és felhasználásával (tárolás, átalakitás, továbbitás) kapcsolatos gyakorlati problémák között nagyon sok olyan is akad, melynek megoldásához (például a kivant berendezés vagy eljárás megtervezéséhez) az információ számos jellemzője közül kizárólag csak a

mennyiséget kell figyelembe venni. Éppen ezekkel a kérdésekkel foglalkozik az információelmélet.

Ez az oka annak, hogy nem foglalkozunk magának az információnak a fogalmával; ez inkább a filozófia feladata lenne. Az információelmélet szempontjából csak az információ mennyisége az érdekes, mint ahogy adattároláskor is mellékes, hogy honnan jöttek és mit jelentenek a feldolgozandó adatok, csak a célszerü elhelyezésükről kell gondoskodni. Az információelmélet éppen a probléma ilyen absztrakt megfogalmazásának köszönheti széleskörü alkalmazhatóságát.

Információn általában valamely, véges számu és előre ismert lehetőség valamelyikének megnevezését értjük. A problémát többnyire ugy fogalmazzuk meg, hogy mennyi információra van szükség egy adott $X = \{x_1, x_2, \ldots, x_n\}$ véges halmaz valamely tetszőleges elemének azonositásához vagy kiválasztásához. Nagyon fontos, hogy információmennyiségről csak akkor beszélhetünk, ha a lehetséges alternativák X halmaza adott. De ebben az esetben is csak olyankor értelmes az információ mennyiségét definialni, ha tömegjelenségről van szó, vagyis ha nagyon sok esetben kapunk vagy szerzünk információt arról, hogy az adott lehetőségek közül aktuálisan melyik következett be. Mindig ez a helyzet a hiradástechnikában és az adatfeldolgozásban, de számos más területen is. Természetesen, az itt elmondottakat pontosabban is körül fogjuk irni.

Az n elemü X halmaz egyes elemeinek azonositásához R.V. Hartley 1928-ban bevezetett formulája szerint

(1) $$I = \log n$$

mennyiségü információra van szükség. Itt és a továbbiakban, $\log x$ az x pozitiv szám kettes alapu logaritmusát jelenti azzal a kiegészitéssel, hogy $0 \log \frac{0}{a} = 0 \log \frac{a}{0} = 0$ és $b \log \frac{b}{0} = +\infty$, $b \log \frac{0}{b} = -\infty$, ha $a \geq 0$, $b > 0$. Kényelmi okokból néha a természetes alapu logaritmust használják, de a hirközléselméletben a kettes alapu logaritmus az elterjedtebb.

Hartley formulájának szemléletes tartalma elég egyszerü. Ha ugyanis $n = 2^k$, akkor az X elemeinek reprezentásához éppen $k = \log n$ hosszuságu bináris sorozatokat célszerü használni, mivel a k hosszuságu bináris sorozatok száma éppen 2^k, tehát k-nál kevesebb bináris jegy már semmiképpen sem elég a kölcsönösen egyértelmü ábrázoláshoz. Perszé, ha $\log n$ nem egész szám, akkor a szükséges bináris jegyek száma a $\log n$ után következő egész szám lesz, mégis érdemes kitartani az $I = \log n$ képlet mellett. Gondoljunk csak arra, hogy ha az X elemeiből alkotható m hosszuságu sorozatokat kell bináris sorozatokkal reprezentálni (ami a hiradástechnikában mindennapos feladat, de az adatfeldolgozásban is elég gyakori), akkor ezek száma n^m, és igy olyan hosszuságu bináris sorozatokra van szükség, hogy

$2^{k-1} < n^m \leq 2^k$, vagyis az X halmaz egy elemére eső bináris jegyek $L = \frac{k}{m}$ száma $\log n$ és $\log n + \frac{1}{m}$ közé esik, azaz a $\log n$ alsó határ tetszőlegesen megközelithető.

Ezek szerint, Hartley formulája az információ mennyiségét a megadásához szükséges állandó hosszuságu bináris sorozatok hosszának alsó határaként definiálja. Ennek megfelelően, az információmennyiség egységét bit -nek nevezzük, ami valószinüleg a "binary digit" angol nyelvü kifejezés röviditése. Hartley szerint kételemü halmaz elemeinek azonositásához van szükség egységnyi (1 bit) mennyiségü információra.

Hartley egyszerü formulája számos esetben jól használható, de van egy komoly hibája: nem veszi figyelembe, hogy - tömegjelenségről lévén szó - az egyes alternativák nem feltétlenül egyenértéküek. Például, nem sok információt nyerünk azzal, ha megtudjuk hogy ezen a héten sem nyertünk a lottón, mert ezt előre is sejthettük volna, hiszen rendszerint ez történik. Ezzel szemben az ötös találat nagyon rendkivül meglepő, mert igazán nem számithatunk rá, ezért az sokkal több információt szolgáltat. Ezt a nehézséget Shannon a valószinüség és az információ fogalmának összekapcsolásával oldotta meg. Shannon szerint egy $P(A)$ valószinüségü A esemény bekövetkezése

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

mennyiségü információt szolgáltat, vagyis az információ annál több, minél kevésbé valószinü. Ez a mérőszám a Hartley-félénél sokkal árnyaltabb megkülönböztetést tesz lehetővé, és ha n lehetőség mindegyike egyformán $\frac{1}{n}$ valószinüségü, akkor (2) a Hartley-formulára redukálódik.

Ez az elképzelés teljes összhangban van az eddig mondottakkal, mivel az információ mennyiségét tömegjelenségek kel kapcsolatban kivánjuk értelmezni, és ezek matematikai leirására éppen a valószinüségszámitás hivatott. Tételezzük fel tehát azt, hogy a vizsgált egyedi információk - vagyis az $X = \{x_1, x_2, \ldots, x_n\}$ véges halmaz elemei a $\mathcal{P} = \{p_1, p_2, \ldots, p_n\}$ eloszlásu ξ valószinüségi változó lehetséges értékei, ahol $P(\xi = x_i) = p_i$, $p_i \geq 0$ és $\sum_{i=1}^{n} p_i = 1$. A (2) képlet szerint ilyenkor a $\xi = x_i$ egyedi információ mennyisége $I = -\log p_i$ lesz. Minthogy előre nem tudhatjuk, hogy esetenként az egyes egyedi információk melyikére kellene felkészülnünk, azért hosszutávu tervezéskor a $-\log p_i$ egyedi információmennyiségek várható értékével kell számolni. Későbbi vizsgálatainkban betöltött alapvető funkciója igazolja majd, hogy valóban ez a helyes választás.

1. Definició: A $\mathcal{P} = \{p_1, p_2, \ldots, p_n\}$ eloszlásu ξ diszkrét valószinüségi változó Shannon-féle entrópiáját a

$$H = H(\xi) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$$

formulával definiáljuk. H az információ átlagos mennyiségét szintén bit-ben adja meg. Az

$$\iota_\xi = \log \frac{1}{p(\xi)}$$

valószinüségi változót - ahol $p(x) = P(\xi = x_i)$ ha $x = x_i$ - egyedi információnak, vagy entrópiasürüségnek nevezzük. (ι a görög ióta irásjel akar lenni.)

Közvetlenül látszik, hogy az entrópia értéke csak az eloszlástól függ, ugyanis különböző értékkészletü, de azonos eloszlásu valószinüségi változók entrópiája ugyanaz. Megjegyezzük, hogy az entrópia sohasem lehet negativ, és ha a ξ valószinüségi változó n számu különböző értéket vehet fel pozitiv valószinüséggel, akkor az entrópiája legfeljebb $\log n$ lehet; amint azt rövidesen be is bizonyitjuk.

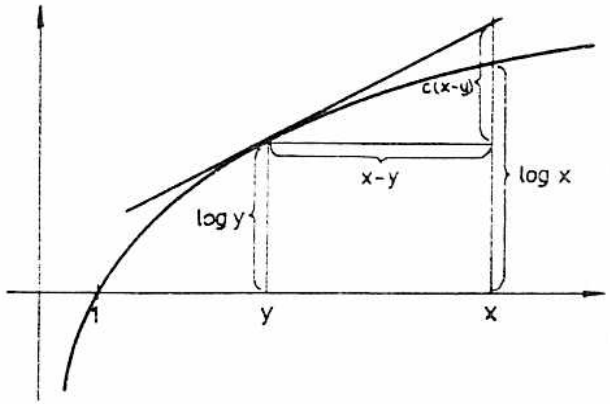
A Shannon-féle entrópia ugy is interpretálható, mint a szóbanforgó eloszlással leirt véletlen kisérlet aktuális kimenetelére vonatkozó bizonytalanságunk mértéke: a kisérlet előtt fennálló bizonytalanság eloszlatásához - vagyis a kisérlet aktuális eredményének megadásához - átlagosan ennyi információra van szükség. Ezt az elképzelést az entrópia tulajdonságai messzemenően alátámasztják; vizsgálatuk főként az alábbi lemmán alapul.

1. Lemma: Ha $a_i \geq 0$, $b_i > 0$; $i = 1,2,\ldots,n$ valós számok, és $a = \sum_{i=1}^{n} a_i$, $b = \sum_{i=1}^{n} b_i$, akkor

$$\sum_{i=1}^{n} a_i \log \frac{b_i}{a_i} \leq a \log \frac{b}{a},$$

ahol az egyenlőség feltétele $\frac{a_i}{b_i} = \text{const.}$ +++

Bizonyítás: A $\log x$ függvény konkáv, vagyis mindenütt az érintője alatt halad. (Lásd az 1. ábrát)



1. ábra

Ha tehát c jelöli az y pontban húzott érintő irány-tangensét, akkor

$$\log x \leq \log y + c(x-y) \tag{3}$$

minden x -re teljesül, és a szigoru konkávitás miatt az egyenlőség feltétele $x = y$. Ha ezt az egyenlőtlenséget az $x = \frac{b_i}{a_i}$, $y = \frac{b}{a}$ szereposztásban alkalmazzuk,

akkor

$$a_i \log \frac{b_i}{a_i} \leq a_i \log \frac{b}{a} + c\left(b_i - a_i \frac{b}{a}\right)$$

adódik, amiből az állitás összegezéssel következik. +++

A lemmából az $a_i = p_i$ és $b_i = 1$ helyettesitéssel azonnal megkapjuk a már emlitett $H(\xi) \leq \log n$ egyenlőtlenséget, és azt is látjuk, hogy egyenlőség csak az egyenletes eloszlás esetén teljesül, vagyis a bizonytalanság ilyenkor a legnagyobb. Ez érthető is, mert ha az eloszlás nem egyenletes, akkor azzal, hogy egyes lehetőségek valószinüsége kisebb, másoké pedig nagyobb, máris tudunk valamit a kisérlet várható eredményéről, tehát a bizonytalanságunk kisebb, mint a teljesen szimmetrikus egyenletes eloszlás esetén.

Természetesen semmiféle heurisztikus okoskodás nem indokolhatja meg kellőképpen azt, hogy miért éppen a Shannon-entrópiával kell mérni az információ mennyiségét. Ez a kérdés már csak azért is indokolt, mert számos más, a Shannon-entrópiához sokban hasonló mérőszám is definiálható, mint például a Rényi-féle

$$H_\alpha = \frac{1}{1-\alpha} \log \sum_{i=1}^{n} p_i^\alpha \tag{4}$$

α-entropia, ahol α tetszőleges pozitiv szám lehet. Megjegyezzük, hogy az $\alpha = 1$ esetben H_α éppen a Shannon-entrópiát adja. Pontosabban, $\lim H_\alpha = H$ ha α az 1-hez tart.

A Shannon-entrópia azért tekinthető az információmennyiség mértékének, mert számos gyakorlati probléma megoldásában kulcsszerepet játszik, és ezekben az esetekben mindig ilyen értelemben interpretálható. Az 1. Definició és a hozzá csatolt általános jellegü értelmezés ezeknek a konkrét eredményeknek a közös lényegét ragadja meg absztrakt formában.

A továbbiakban ilyen problémák megoldásán keresztül kivánjuk megmutatni, hogy valóban a Shannon-entrópia méri az információ (bizonytalanság) számszerü mennyiségét. Először az entrópia kereséselméleti értelmezésével foglalkozunk, ahol az információmennyiség ugy jelenik meg, mint az információ megszerzésének ára.

1.2. Keresési stratégiák és prefix kódok

Keresési problémára a legegyszerübb példa a jólismert Bar-Kochba játék. Itt a versenyzőnek véges számu, jól meghatározott dolog valamelyikét kell kitalálni a kérdéseire kapott válaszok alapján. Amennyiben kérésére csak igen-nem választ kaphat, ugy egy kérdéssel azt döntheti el, hogy a kérdéses dolog a lehetőségek egy tetszőlegesen választható részhalmazának eleme-e vagy nem. Ha a válasz igen, akkor ebben a halmazban folytatja a kérdezést, ha nem, akkor a komplementer halmazban, mindaddig, amig nem sikerül egyelemü halmazra redukálni a lehetőségek körét. A játék célja minimális számu kérdéssel kitalálni

a gondolt dolgot, vagyis a hiányzó információt a lehető legolcsóbban kell megszerezni. Persze, adott stratégia esetén a szükséges kérdések száma függhet a kitalálandó dologtól, ezért - pontosabb fogalmazásban - az átlagos kérdésszám minimalizálására kell törekedni. Világos, hogy átlagosan annál több kérdésre van szükség, minél nagyobb a megszerzendő információ mennyisége. Látni fogjuk, hogy ebben az összefüggésben információmennyiségen a H Shannon-entrópiát kell érteni. Arról van szó, hogy egy igen-nem válasz maximálisan 1 bit információt tartalmazhat, mig a hiányzó információ H bit. Nem várhatjuk tehát, hogy átlagosan H -nál kevesebb kérdés elég lehet, de azt is be fogjuk bizonyitani, hogy ez az elvi alsó korlát elég jól megközelithető.

Keresési problémára számos más példa is adható. Az amerikai hadseregnél állitólag ugy 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 pozitiv, ott a felezgetést tovább folytatják egész addig, amig a betegeket ki nem szürték. Ez a módszer nagyon gazdaságos, mert ha 1000 páciens között pontosan egy vérbajos van, akkor az 10 vizsgálattal lokalizálható, mig az egyenkénti vizsgálatnál - ami adminisztrációs szempontból persze sokkal egyszerübb - átlagosan 500 próbára van szükség.

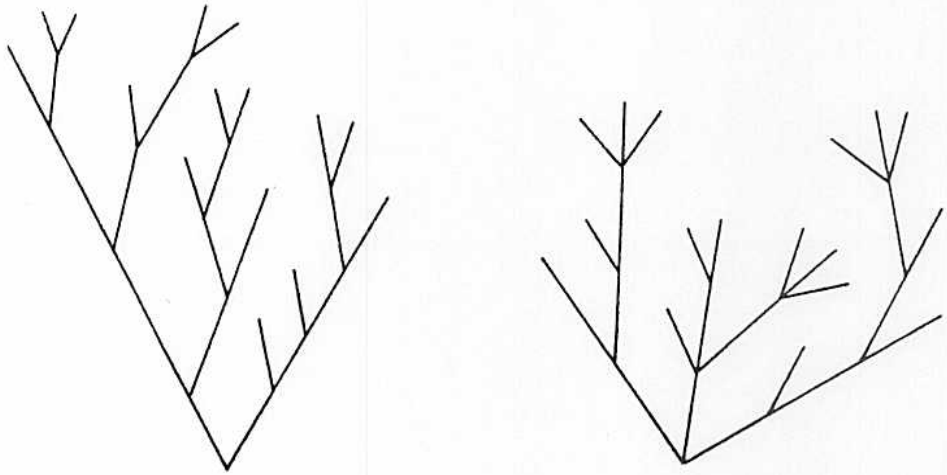
Hasonló feladatra vezet a jólismert hamispénz probléma is. Itt kétserpenyős mérleg segitségével kell kiválasztani a külsőre teljesen egyforma pénzdarabok közül a könnyebb hamisat. Ez ugy történhet, hogy azonos létszámu csoportokat téve a mérlegre, megállapithatjuk, hogy a keletkezett három csoportból melyikben van a hamis. Ha ugyanis a mérleg egyensulyban van, akkor a maradékban van, ha nem, akkor a könnyebb csoportban. Ezután az eljárás ugyanigy folytatódik. Mivel egy mérlegelésnek három kimenetele lehet, az maximálisan log 3 információt szolgáltathat. Ha tehát n pénzdarab közül mindegyik egyforma valószinüséggel lehet a hamis, akkora hiányzó $\log n$ információ megszerzéséhez átlagosan legalább $\frac{\log n}{\log 3}$ mérlegelésre van szükség. A közismert $n = 9$ esetben tényleg elég két mérlegelést végezni, de átlagosan ennél kevesebb már nem vezethet mindig eredményre.

Megjegyezzük, hogy a hamispénz probléma bonyolultabb az előzőeknél, mert ott mindig azonos létszámu csoportokat kell a mérlegre tenni, és igy nem választhatjuk meg tetszőlegesen azt a három csoportot, amelyről eldöntjük, hogy melyikben van a hamis.

Elképzelhető olyan keresési feladat is, ahol egyszerre legfeljebb $s \geq 2$ csoportról tudjuk egyetlen kisérlettel eldönteni, hogy a keresett elem melyikben van. A probléma általános megfogalmazása a következő: Legyen a keresett ξ dolog az $X = \{x_1, x_2, \ldots, x_n\}$ véges halmaz valamelyik eleme; a ξ -t valószinüségi

változónak tekintjük: $p_i = P(\xi = x_i)$ a ξ eloszlása, tehát $p_i \geq 0$, $\sum_{i=1}^{n} p_i = 1$. A szóbajöhető keresési stratégiák definiálását és áttekintését nagyon megkönnyiti a grafikus ábrázolásuk alábbi lehetősége.

Fának nevezzük az olyan irányitott gráfokat, melyek egy kitüntetett szögpontjából, a kezdőpontból, ágak (irányitott utak) indulnak ki, melyek a későbbi szögpontokban ismét elágazhatnak, de ujra már biztosan nem találkozhatnak. Azokat a szögpontokat, melyekből további élek már nem indulnak ki, végpontoknak nevezzük. Mivel az ágak ujra már nem találkoznak, a kezdőpontot mindegyik végponttal pontosan egy ág köti össze. Az ágat alkotó élek számát az ág hosszának nevezzük. (Lásd a 2. ábrát.)



2. ábra

Tekintsünk egy n végpontu fát, és rendeljük hozzá kölcsönösen egyértelmü módon a fa végpontjaihoz (ágaihoz) a fenti n elemü X halmaz elemeit. Az ilyen módon cimkézett végpontu fát az X halmaz kódfájának fogjuk nevezni. Ha most 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 egyesitése. Látható, hogy az X halmaz olyan kódfája, ahol minden szögpontból legfeljebb s él indul ki, s altern tivás keresési stratégiát definiál, és ez a megfeleltetés kölcsönösen egyértelmü. Ugyanis, a keresési eljárás a kódfa valamelyik ágának végigjárását jelenti: a végponthoz rendelt elem az, amit keresünk. Ez ugy történik, hogy ha a keresés során megállapitottuk, hogy ξ az A szögponthoz tartozó halmaz eleme, akkor a következő kisérlettel azt döntjük el, hogy az A-ból kiinduló élek $B_1, B_2, \dots, B_j$ $j \le s$ végpontjai közül melyikhez tartozik a ξ -t tartalmazó halmaz, és ennek az élnek az irányába haladunk tovább. Ha $\xi = x_i$, akkor a megtaláláshoz szükséges lépések száma az x_i végpontba vezető ág $L(x_i)$ hossza. A feladat az

$$L = \sum_{i=1}^{n} L(x_i)\, P(\xi = x_i) \tag{5}$$

átlagos lépésszám minimalizálása. Mivel egy kisérlettel legfeljebb log s mennyiségü információt nyerhetünk, és a hiányzó információ $H(\xi)$, nem számithatunk arra, hogy L kisebb lesz mint $\frac{H(\xi)}{\log s}$. Ezt most már be is bizonyitjuk.

1. Tétel: s alternativás keresési stratégia mindig eleget tesz az alábbi egyenlötlenségnek:

$$L = \sum_{i=1}^{n} L(x_i)\, P(\xi = x_i) \ge \frac{H(\xi)}{\log s} ,$$

vagyis kódfa ágainak átlagos hossza a fenti korlát felett marad. + + +

Bizonyitás: Tekintsünk egy tetszőleges kódfát, és legyen A a kódfa olyan szögpontja, amely nem végpont. Jelölje $B_1, B_2, \dots, B_j$, $j \le s$ az A -ból kiinduló élek végpontjait. Azonosan jelöljük a megfelelő halmazokat is. Legyen

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

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

$$L = \sum P(A), \tag{6}$$

ahol az összegezést a végpontoktól különböző szögpontokra kell elvégezni.

Mivel $P(A)$ éppen annak a valószinüsége, hogy a keresés során eljutunk az A szögpontba, az A szögpontban végzendő kisérlet $B_1, B_2, \dots, B_j$ kimeneteleinek valószinüségei rendre $P(B_1|A), P(B_2|A), \dots, P(B_j|A)$, ahol

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

Ennek a kisérletnek a bizonytalansága (entrópiája) tehát

$$H_A = -\sum_{i=1}^{j} \frac{P(B_i)}{P(A)} \log \frac{P(B_i)}{P(A)} =$$

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

Számoljuk ki a $\sum_A P(A)\, H_A$ mennyiséget, ahol az összegezést a kódfa végponttól különböző szögpontjaira kell elvégezni. A fenti felbontás azt mutatja, hogy ebben az összegben - a kezdőpont és a végpontok kivételével - minden szögponthoz a $P(C) \log P(C)$ kifejezés egyszer pozitiv, egyszer negativ előjellel lesz hozzárendelve, mert a C egyszer az "A", egyszer pedig valamelyik "B_i" szerepét játssza. Tehát az összegezés a

$$(7) \quad \sum_A P(A)\, H_A = -\sum_{i=1}^{n} p_i \log p_i + P(X) \log P(X) = H(\xi)$$

eredményre vezet, mivel a kezdőpont csak az "A" tipusu, a végpontok pedig csak a "B_i" tipusu szerepben fordulnak elő, és $P(X) = 1$. (X a kezdőpont).

Viszont az 1. Lemma következményeként tudjuk, hogy $H_A \le \log s$, mert egy szögpontból legfeljebb s él indulhat ki, tehát

$$H(\xi) = \sum_A P(A)\, H_A \le \log s \sum_A P(A) = L \log s$$

a (6) azonosság miatt, ahol az összegezést ismét a kódfa végpontoktól különböző szögpontjaira kell érteni. Innen a tétel állitása közvetlenül következik. +++

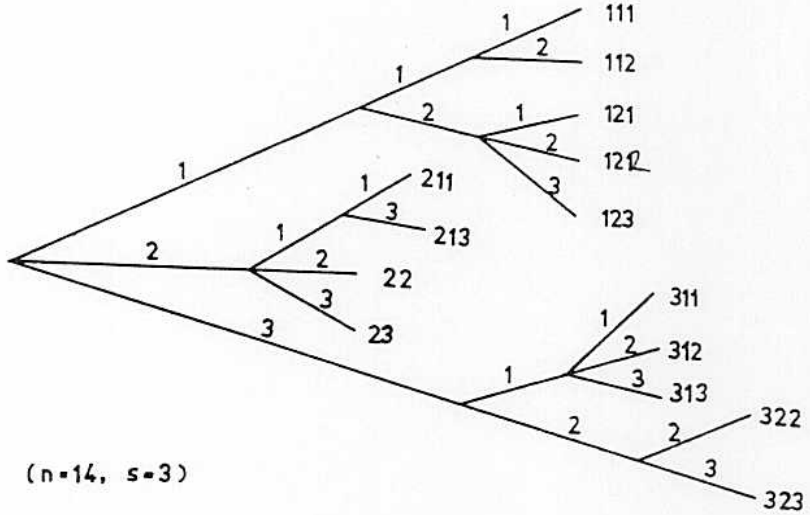
Nem tisztáztuk még azt a kérdést, hogy a $\frac{H}{\log s}$ elvi alsó korlát mennyire közelithető meg alkalmasan konstruált keresési stratégiákkal. Ehhez először a probléma egy ekvivalens átfogalmazását adjuk a hirközléselmélet és az adattárolás szempontjából alapvető fontosságu kódolás fogalmának a felhasználásával.

Az Y véges halmazból alkotható véges sorozatok halmazát a továbbiakban Y^+ jelöli, az Y^+ két elemét akkor tekintjük azonosnak, ha ugyanolyan hosszuak, és minden helyen megegyeznek. Az Y elemeit - kódolási problémákkal kapcsolatban - betüknek vagy kódjeleknek az Y^+ elemeit pedig szavaknak nevezzük. A 0 hosszuságu üres sorozatot is az Y^+ elemei közé soroljuk, bár ennek csak elvétve van jelentősége. Az Y halmazból alkotható n hosszuságu sorozatok halmazát általában Y^n jelöli; tehát Y^+ az $Y^0, Y^1, Y^2, \dots, Y^n, \dots$ halmazok egyesitése. Ha a v szó az u szóból bizonyos betük hozzáirásával keletkezik, vagy $u = v$, akkor azt mondjuk hogy v az u folytatása; ezt $u \prec v$ jelöli. Az X véges halmaznak az

Y^+ halmazba történő $g : X \longrightarrow Y^+$ leképezését az X kódjának nevezzük, a $g(x)$, $x \in X$ alaku szavakat pedig a g kód kódszavainak mondjuk. A $g(x)$ kódszó hosszát, vagyis az őt alkotó betük számát $\|g(x)\|$ jelöli. Általában fel fogjuk tételezni, hogy az s elemü Y kódábécé elemei az $1, 2, \ldots, s$ számok, ez az általánosságot semmiben sem csorbitja.

2. Definició: A $g: X \longrightarrow Y^+$ kódot prefix tulajdonságu kódnak nevezzük akkor, ha a kódszavai mind különbözők, és egyik sem folytatása a másiknak. + ++

További vizsgálataink kiindulópontja az az észrevétel, hogy a kódfák prefix kódokat ábrázolnak, és ez a megfeleltetés lényegében kölcsönösen egyértelmü. Ha ugyanis az $1, 2, \ldots, j$; $j \leq s$ számokkal rendre megszámozzuk egy kódfa azonos szögpontjaiból kiinduló éleit, és a kódszavakat az egyes ágakat alkotó élek sorszámainak egymás után irásával állitjuk elő, akkor nyilván prefix kódot kapunk, mert a kódfán belül az ágakhoz rendelt szavak az ágakat kölcsönösen egyértelmü módon reprezentálják, és egyik ág sem folytatása egy másiknak. Az is világos, hogy ezzel a módszerrel minden prefix kód ábrázolható kódfával (lásd a 3. ábrát), és ez az ábrázolás - az egyes szögpontokból kiinduló élek szamozásától eltekintve - kölcsönösen egyértelmü. Ez a többértelmüség persze egyáltalán nem zavaró, mert az egy szögpontból kiinduló élek egyenértékűek.



3. ábra

Amennyiben a kódolt X halmazon egy $P(x)$, $x \in X$ valószinüségeloszlás is adott, a g kód átlagos kódhosszát a

$$(8) \qquad L = \sum_{x \in X} \|g(x)\| P(x)$$

képlettel definiáljuk, ez megegyezik a megfelelő keresési stratégia átlagos lépésszámával. Eszerint az 1. Tételt ugy is fogalmazhatnánk, hogy s betüből álló kódábécé esetén prefix kód átlagos kódhossza nem lehet kisebb mint $\frac{H}{\log s}$. A következő szakaszban olyan prefix kódok konstrukciójával foglalkozunk, melyek átlagos kódhossza ezt jól megközeliti.

1.3. A Shannon-Fano és a Gilbert-Moore féle kód

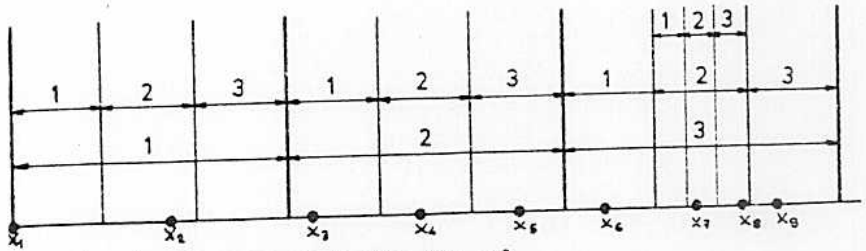
Legyen $\mathcal{P} = \{p_1, p_2, \dots, p_n\}$ tetszőleges eloszlás az X halmazon, melynek elemeit az Y -elemü ábécével kivánjuk kódolni. Éppugy mint az előző, az alábbi tétel is C. Shannontól származik.

2. Tétel: Ha az adott $\mathcal{P}$ eloszlás entrópiája H, akkor s kódjelből mindig készithető olyan prefix kód (illetve mindig van olyan s -alternativás keresési stratégia) hogy a L átlagos kódhosszra (átlagos lépésszámra)

$$L < \frac{H}{\log s} + 1$$

teljesül. +++

Bizonyitás: Egy kivánt tulajdonságu prefix kód a következő módszerrel konstruálható meg. Rendezzük át az X elemeinek számozását ugy hogy az elemek a valószinüségeik csökkenő sorrendjében legyenek megszámozva, vagyis $p_1 \geq p_2 \geq \dots \geq p_{n-1} \geq p_n$; nyugodtan feltehetjük, hogy p_n is pozitiv. Ezután mérjük fel balról jobbra haladva a $[0,1)$ intervallumra a $p_1, p_2, \dots, p_n$ szakaszokat; az i -ik szakasz kezdőpontját jelöle $x_i \in X$. (lásd a 4. ábrát)



A kód: {11, 12, 21, 22, 23, 31, 322, 323, 33}

(n=9, s=3)

4. ábra

Osszuk fel a $[0,1)$ alapintervallumot s egyenlő részre, majd a kapott - balról zárt, jobbról nyilt - intervallumok közül azokat, melyekben egynél több x_i pont van, osszuk ismét s egyenlő részre egészen addig, amig elérjük, hogy mindegyik x_i pont más intervallumban van. Végül, számozzuk meg a továbbosztott intervallumokon belül az s darab részintervallumot balról jobbra haladva az $1, 2, \dots, s$ számokkal; ezután a keresett g kód a következőképpen definiálható.

Keressük ki az x_i pontot tartalmazó s^{-1} hosszuságu intervallumot, ennek sorszáma lesz a $g(x_i)$ kódszó első jegye. Ugyanigy, a második, ..., k -ik jegyet az x_i -t tartalmazó s^{-2}, ..., s^{-k} hosszuságu intervallum sorszámával adjuk meg, amennyiben ilyen még van. Ha a fentebb konstruál

intervallumok közül s^{-L_i} a legrövidebb olyan, amely x_i-t tartalmazza - és akkor ebben már x_i az egyetlen pont, de az ezt tartalmazó s^{L_i-1} hosszuságuban x_i-n kivül még más pont is van - akkor $\|g(x_i)\| = L_i$ lesz. A konstrukcióból közvetlenül látszik, hogy valóban prefix tulajdonságu kódot kaptunk. Megmutatjuk, hogy

(9) $$p_i < s^{-(L_i-1)}.$$

Azt tudjuk, hogy az x_i-t tartalmazó utolsóelőtti - $s^{-(L_i-1)}$ - hosszuságu intervallumban legalább még egy pont van, éspedig x_{i-1} és x_{i+1} közül valamelyik. A második esetben (9) helyessége nyilvánvaló, mivel x_i és x_{i+1} távolsága éppen p_i. Az első esetben viszont ugyanigy azt kapjuk, hogy

$$p_{i-1} < s^{-(L_i-1)},$$

amiből (9) $p_i \le p_{i-1}$ miatt következik.

(9) mindkét oldalának negativ logaritmusát képezve

$$-\log p_i > (L_i-1)\log s$$

adódik, amiből összegezéssel

$$H = -\sum_{i=1}^{s} p_i \log p_i > \log s \sum p_i (L_i-1) = (L-1)\log s.$$

következik. Ezzel a tételt bebizonyitottuk. +++

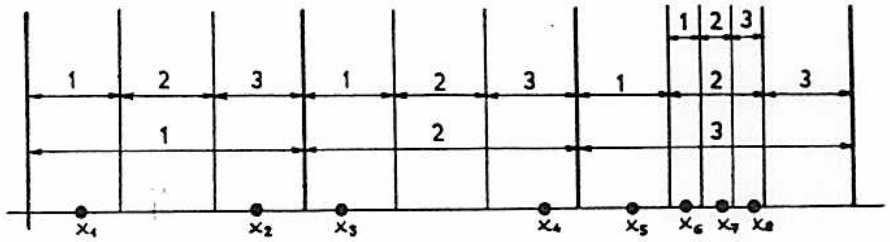
A tételben konstruált kódot Shannon-Fano féle kódnak nevezzük. Ezzel egyben elég jó tulajdonságu keresési stratégiák megkonstruálására is lehetőség nyilt.

A Shannon-Fano kód konstrukciója elég egyszerü, és az átlagos kódhossza is jó, de van egy komoly hátránya: először a valószinüségek csökkenő sorrendjében kell elrendezni a kódolandó jeleket, és ez az átrendezés - nagy elemszám esetén - rendkivül időrabló müvelet. Ezen a nehézségen segit az alábbi Gilbert-Moore féle kód, de ennek az átlagos kódhossza egy kicsit nagyobb lesz. Viszont számos olyan keresési probléma van, ahol éppen erre a kódra van szükség.

Ugyanazokat a jelöléseket használjuk, mint előbb, de most a p_i valószinüségekről nem tételezzük fel, hogy csökkenő sorrendben vannak megszámozva. Mérjük fel őket a $[0,1)$ intervallumra, és jelölje x_i az i-ik intervallum felezőpontját. Ezután osszuk s egyenlő részre a $[0,1)$ intervallumot, majd azokat a kapott intervallumokat osszuk tovább s egyenlő részre, melyekben egynél több x_i pont van, egészen addig, amig mindegyik x_i külön intervallumban lesz. A kódszavakat is ugyanugy definiáljuk, mint a Shannon-Fano kód esetén, tehát most is prefix kódot kapunk. Ha $L_i = \|g(x_i)\|$, akkor az x_i-t tartalmazó utolsó előtti - $s^{-(L_i-1)}$ hosszuságu - intervallum az x_{i-1} és az x_{i+1} pontok közül az egyiket biztosan tartalmazza, de a távolságuk x_i-től legalább $\frac{1}{2} p_i$, tehát most a (9) helyett

(10) $\frac{1}{2} p_i < s^{-(L_i-1)}$

teljesül. (Lásd az 5. ábrát)



A kód: {11, 13, 21, 23, 31, 321, 322, 323}

(n=8, s=3) 5. ábra

(10)-ből, mindkét oldal negativ logaritmusát véve

$$-\log p_i > (L_i-1)\log s - 1$$

következik, amiből

$$H = \sum_{i=1}^{n} p_i \log p_i > (L-1)\log s - 1,$$

tehát a Gilbert-Moore féle prefix kód átlagos kódhossza a

(11) $$\frac{H}{\log s} \leq L < \frac{H+1}{\log s} + 1$$

feltételnek tesz eleget, ez nem lényegesen rosszabb mint a Shannon-Fano kód esetében.

Ezzel szemben a Gilbert-Moore kódot olyan keresési feladatok megoldására is fel lehet használni, amikor a Shannon-Fano kód nem felel meg. Tegyük fel például, hogy egy elektromos berendezésekből álló lánc valamelyik elemének hibáját ugy tudjuk megállapitani, hogy a lánc egy szakaszának meg-

mérjük a vezetőképességet, és ebből kitünik, hogy a hiba a szóbanforgó részben van-e vagy sem. Ezután a hibakeresést a kiválasztott szakasz valamely részével folytatjuk a hiba lokalizálásáig. Ez a keresési stratégia nagyon hasonlit az előző szakaszban tárgyalthoz, de itt nem választhatjuk ki tetszőlegesen a megvizsgálandó csoportot, mert annak szomszédos elemekből kell állnia. Könnyü észrevenni, hogy a Gilbert-Moore féle kód éppen ilyen keresési stratégiákat definiál (itt most $s=2$) mivel a kódfáján az egy szögponton áthaladó ágak sorszámai (azaz a végpontjukhoz rendelt $x_i \in X$ elemek indexei) szomszédosak. Az ilyen kódot alfabetikus kódnak nevezzük. Nem bizonyitjuk részletesen, hogy a Gilbert-Moore kód tényleg alfabetikus, de a konstrukció alapján ez elég nyilvánvaló. Egyébként a Shannon-Fano kódnál is csak a kezdeti átrendezés rontja el ezt a tulajdonságot.

A kivánt keresési stratégia elkészitéséhez persze pontosan ismerni kell az eloszlást, mert a kód jó vagy rossz volta ettől függ, és persze a konstrukció is az eloszláson alapul. Sajnos, gyakorlati esetekben ez a feltétel nem mindig teljesül olyan szépen, mint az alábbi példában, amely a számológépes gyakorlatban elég sokszor előfordul. Tegyük fel, hogy a $N_1 < N_2 < \ldots < N_n < N_{n+1}$ valós számok által meghatározott n számu intervallumba kell besorolni a $[N_1, N_{n+1})$ intervallumból egyenletes eloszlás szerint

érkező számokat. Ez ugy történhet, hogy a beérkező számot összehasonlitgatjuk az N_i határokkal addig, amig a tartalmazó intervallumot nem sikerül $[N_i, N_{i+1})$ alakura redukálni. Itt az eloszlás pontosan ismert:

$$p_i = \frac{N_{i+1} - N_i}{N_{n+1} - N_1}$$

és igy a Gilbert-Moore féle konstrukcióval olyan besorolási stratégiát definiálhatunk, melynél a szükséges összehasonlitások átlagos száma közel van az optimumhoz. Ha n nagy, akkor ezzel jelentős megtakaritás érhető el. Érdemes észrevenni, hogy ez a módszer felhasználható adott $\{p_1, p_2, \ldots, p_n\}$ eloszlásu véletlen számsorozat generálására is. Ha ugyanis adott egy egyenletes eloszlásu véletlen számsorozat az $[N_1, N_{n+1})$ intervallumban, akkor az N_i határok alkalmas megválasztásával elérhetjük, hogy az osztályok valószinüségei éppen az előirt számok legyenek, és igy az osztály sorszáma (vagy valamilyen függvénye) lesz a kivánt eloszlásu valószinüségi változó.

1.4. A Huffman-féle optimális kód

A Shannon-Fano kód átlagos kódhossza elég közel van a $\frac{H}{\log s}$ alsó korláthoz, de általában nem éri azt el. Az 1. Tétel bizonyitásából az is látszik, hogy az

$$L = \frac{H}{\log s}$$

egyenlőség csak akkor teljesülhet, ha $H_A = \log s$ minden végponttól különböző A szögpontra. Ez annyit jelent, hogy a kódfa minden szögpontjából s él indul ki, és $P(B_i|A) = \frac{1}{s}$ mindegyik élre. Könnyü megmutatni, hogy ezek a feltételek csak akkor teljesülnek, ha a p_i valószinüségek mind s^{-L_i} alakuak, ahol L_i egész szám. Sőt, még az is igaz, hogy ilyenkor a Shannon-Fano kód megfelelő kódszavának hossza éppen L_i lesz, tehát $\log p_i = -L_i \log s$ felhasználásával

$$\sum_{i=1}^{n} L_i p_i = L = -\frac{1}{\log s} \sum_{i=1}^{n} p_i \log p_i = \frac{H}{\log s} .$$

Ebben a speciális esetben tehát ismerjük a lehető legjobb kódot, ami ugyan általában is létezik, csak nem olyan könnyü megtalálni.

Az optimális (minimális átlagos kódhosszu) kód megkonstruálására először R.A. Huffman irt le általános módszert. Ez az eljárás tetszőleges eloszláshoz véges számu lépésben megadja az optimális kódok egyikét, melyet Huffman kódnak nevezünk. Részletesen csak a bináris kóddal foglalkozunk, az általános eset tárgyalása eléggé hasonlóan történik.

Legyen $\mathcal{P} = \{p_1, p_2, \ldots, p_n\}$ tetszőleges eloszlás az $X = \{x_1, x_2, \ldots, x_n\}$ véges halmazon, és tegyük fel, hogy $g : X \to Y^+$ az X halmaznak (a $\mathcal{P}$ eloszlásra vonatkozóan) optimális prefix kódja; itt Y

egyenlőre s elemü kódábécé, de figyelmünket már most az $Y = \{0,1\}$ bináris esetre koncentráljuk. Egyszerüség kedvéért feltételezzük, hogy a p_i valószinüségek csökkenő sorrendben vannak megszámozva, és a legkisebb is pozitiv:

$$p_1 \geq p_2 \geq \cdots \geq p_{n-1} \geq p_n > 0.$$

Megjegyezzük, hogy erre az átrendezésre ténylegesen nincs szükség, csak a jelölések egyszerüsitését szolgálja.

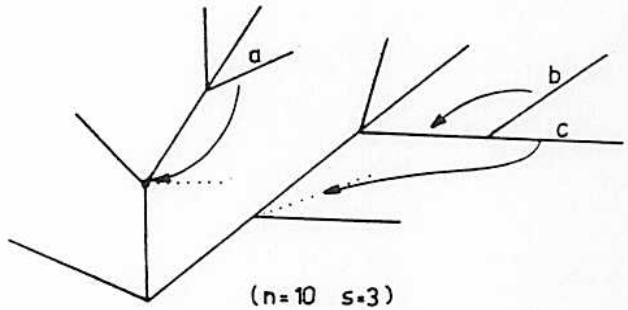
A g kód kódfájáról egyszerüen leolvashatóak az optimális kódok alábbi tulajdonságai:

A) Az $L_i = L(x_i)$ kodhosszak sorozata monoton növekvő, vagyis $L_1 \leq L_2 \leq \cdots \leq L_{n-1} \leq L_n$.

Ha ugyanis nem igy volna, akkor két kódszó felcserélésével az átlagos kódhosszt csökkenthetnénk, tehát g nem volna optimális.

B) A kódfa teljes abban az értelemben, hogy minden - végponttól különböző - szögpontból s számu él indul ki, kivéve esetleg egy végpont előtti szögpontot.

Ellenkező esetben ugyanis a maximális hosszuságu ágak utolsó élei közül egyeseket átültethetnénk a nem teljesen kihasznált szögpontokba vagy elhagyhatnánk, és ezzel az átlagos kódhossz kisebb lenne - vagy legalábbis nem nőne. (Lásd a 6. ábrát, ahol az a,b,c éleket lehet átültetni, illetve a b-t törölni).



6. ábra

C) $L_n = L_{n-1}$ és a bináris esetben ($s = 2$) minden szögpontból, ami nem végpont, két él indul ki. Feltehetjük tehát, hogy $g(x_n)$ és $g(x_{n-1})$ csak az utolsó kódjelben különbözik, vagyis x_{n-1} és x_n ugyanannak a (maximális hosszuságu) ágnak a két végpontja.

C) tulajdonképpen a B) következménye a binaris esetben; a Huffman kód konstrukciója ezen az észrevételen alapul.

Tételezzük fel, hogy a $\mathcal{P}' = \{p_1, p_2, \cdots, p_{n-2}, p_{n-1} + p_n\}$ eloszláshoz már megvan egy g' optimális bináris kód. Megmutatjuk, hogy a $p_{n-1} + p_n$ valószinüséghez rendelt kódszó kétféle kiegészitésével (ez a müvelet a kódfa megfelelő ágához két él hozzátoldását jelenti) kapott kód az eredeti eloszlás optimális kódja lesz. Ezzel a kiegészitéssel az L' átlagos kódhossz $L = L' + p_{n-1} + p_n$ -re nő; de ha az igy kapott kód nem volna optimális, akkor a

nála kisebb átlagos kódhosszu g^* prefix kód kódfájáról feltehetjük a C) tulajdonság értelmében, hogy az x_{n-1} és az x_n elemekhez rendelt ágak az utolsó élek kivételével megegyeznek, tehát ennek a két utolsó ágnak az elhagyásával a g' kód kódfájánál jobbat kapnánk, ami a g' optimális volta miatt lehetetlen, vagyis a g prefix kód valóban optimális.

Ennek alapján a binaris Huffman kódot ugy konstruálhatjuk meg, hogy a két legkisebb valószinüség összeadásával addig redukáljuk a problémát, amig meg nem tudjuk oldani. $n-2$ lépés után ez biztosan bekövetkezik, mert az $n=2$ eset már triviális. Ezután az elvégzett összevonások megforditásaként a megfelelő kódszót a kétféle kiegészitésével helyettesitve felépithetjük a kivánt optimális kódot.

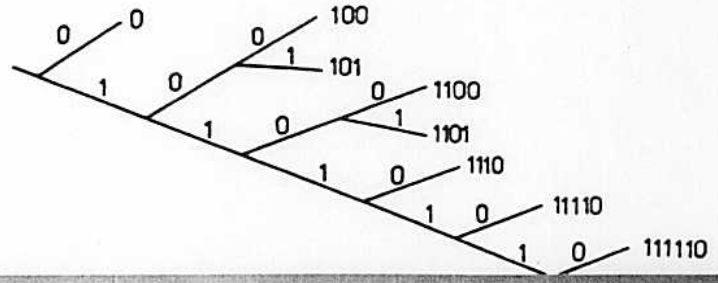
Az alábbi diagramokon a $\mathcal{P} = \{0{,}49\ ,\ 0{,}14\ ,\ 0{,}14\ ,\ 0{,}07\ ,\ 0{,}07\ ,\ 0{,}04,\ 0{,}02,\ 0{,}02\ ,\ 0{,}01\}$ eloszláshoz tartozó bináris Huffman kódot konstruáljuk meg.

I.	II.	III.	IV.	V.	VI.	VII.
0,49	0,49	0,49	0,49	0,49	0,49	0,49
0,14	0,14	0,14	0,14	0,14	0,14 →	0,28
0,14	0,14	0,14	0,14	0,14	0,14 ↗	0,23
0,07	0,07	0,07	0,07 →	0,14 →	0,23	
0,07	0,07	0,07	0,07 ↗	0,09 ↗		
0,04	0,04	0,04 →	0,09			
0,02	0,02 →	0,05 ↗				
0,02 →	0,03 ↗					
0,01 ↗						

Itt a lépésenként kapott eloszlások valószinüségei az oszlopokban szerepelnek, a redukció módját a nyilak jelzik. A VII. oszlopban kapott eloszlásnak $\{0\ ,\ 10\ ,\ 11\}$ kódszavakból álló prefix kód nyilván optimális kódja lesz; ebből kiindulva, a fenti táblázat nyilai szerint visszafelé haladva a következőképpen kapjuk meg az eredeti eloszlás Huffman-kódját; itt az oszlopokban a kódszavakat soroljuk fel, a nyilak pedig a kód bővitésének utját mutatják.

VII.	VI.	V.	IV.	III.	II.	I.
0	0	0	0	0	0	0
10 →	100	100	100	100	100	100
11 ↘	101	101	101	101	101	101
	11 →	110 →	1100	1100	1100	1100
	↘	111 ↘	1101	1101	1101	1101
			111 →	1110	1110	1110
			↘	1111 →	11110	11110
				↘	11111 →	111110
					↘	111111

A végeredmény az I. oszlopban áll, a kódfa a 7. ábrán látható.



Az általános esettel nem foglalkozunk részletesen, csak megjegyezzük, hogy a konstrukció lényegében ugyanigy történik, csak az egyes redukciós lépésekben az n és az s értékétől függően - az r legkevésbé valószinü elemet kell összevonni, és a valószinüségeiket összeadni. A tényleges konstrukció a redukciós lépések értelemszerü megforditásával történik. Az összeadandó valószinüségek r számát a B) tulajdonság alapján a következőképpen határozhatjuk meg. B) értelmében feltehető, hogy a keresett optimális kódfa minden végponttól különböző szögpontjából s él indul ki, kivéve esetleg egy végpont előtti szögpontot, amelyből r él megy tovább, ahol $2 \le r \le s$. Ezek az élek éppen a legkisebb valószinüségekhez tartozó végpontokhoz vezetnek, tehát a redukciónál ezeket kell törölni. Az r meghatározásához egészitsük ki $s-r$ él hozzávételével ezt a szögpontot is teljessé, az igy kapott, valóban teljes kódfa végpontjainak száma $s+m(s-1)$ alaku, ahol m egész szám. Ennek az az oka, hogy egy utolsó előtti szögpont, és a belőle kiinduló s él törlésével a végpontok száma $s-1$ -el csökken, és ezt az eljárást addig folytathatjuk, amig csak a kezdőpont, és a belőle kiinduló s él marad. Tehát szükségképpen

$$n + s - r = s + m(s-1),$$

amiből egyszerüsitéssel

$$n = m(s-1) + r$$

adódik, ami az n és s ismeretében az r, $2 \le r \le s$ egész számot egyértelmüen meghatározza. Azt, hogy az igy felépithető kód tényleg optimális lesz, ugyanugy bizonyithatjuk be, mint a bináris esetben.

1.5. A feltételes entrópia és a kölcsönös információ

Eddigi eredményeink még nem jogositanak fel teljesen arra, hogy a Shannon-féle entrópiát az információmennyiség mértékének tekintsük. Az adott szituációban csak annyit állithatunk, hogy - a $p_i = s^{-l_i}$ speciális tipusu eloszlások kivételével - az entrópia elég jó alsó becslés az információ mennyiségére. A következő fejezetben bonyolultabb hirközlési problémák megoldásához kapcsolódóan megmutatjuk, hogy ez az esetlegesség a feladat nem elég általános megfogalmazásának következménye. A soronkövetkező 1.5 és 1.6 szakaszban az ezekhez a vizsgálatokhoz szükséges általános jellegü eredményeket bizonyitjuk.

Legyen ξ és η két diszkrét valószinüségi változó, melyek lehetséges értékei az X illetve Y véges halmaz elemei. A későbbiekben is széleskörüen használjuk majd a nagyon kényelmes

$$p(x,y) = P(\xi = x,\ \eta = y), \quad p(x) = P(\xi = x), \quad p(y) = P(\eta = y),$$

$$p(x|y) = P(\xi = x \mid \eta = y) \quad \text{ha} \quad p(y) > 0$$

jelöléseket, és analóg jelöléseket alkalmazunk több valószínüségi változó esetén is. Ez a jelölésmód nagyon ritkán lehet csak félrevezető, mivel számunkra legtöbbször csak az eloszlások érdekesek.

Megjegyezzük, hogy a (ξ, η) pár szintén véges értékkészletü valószínüségi változó, tehát az 1. Definició értelmében az entrópiáját

$$H(\xi, \eta) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{1}{p(x,y)}$$

definiálja, ezt a mennyiséget a ξ és η együttes entrópiájának nevezzük. Hasonlóan, a $\xi_1, \xi_2, \ldots, \xi_n$ valószínüségi változók együttes entrópiáját az alábbi formula definiálja:

$$H(\xi_1, \xi_2, \ldots, \xi_n) = \sum_{x_1 \in X_1} \sum_{x_2 \in X_2} \cdots \sum_{x_n \in X_n} p(x_1, \ldots, x_n) \log \frac{1}{p(x_1, \ldots, x_n)}$$

ahol $p(x_1, x_2, \ldots, x_n) = P(\xi_1 = x_1, \xi_2 = x_2, \ldots, \xi_n = x_n)$ a valószínüségi változók együttes eloszlása, X_i pedig a ξ_i véges értékkészletü valószínüségi változók lehetséges értékeinek halmaza.

Térjünk most vissza a ξ , η pár vizsgálatára. Ha tudjuk, hogy η értéke y , akkor a ξ viselkedését a $p(x|y)$, $x \in X$ eloszlás irja le, melynek entrópiája

$$H(\xi | \eta = y) = \sum_{x \in X} p(x|y) \log \frac{1}{p(x|y)} .$$

Persze többnyire az η értékét sem ismerjük, ilyenkor ennek a mennyiségnek a várható értéke adja meg azt az átlagos bizonytalanságot, amely az η értékének feltételezett ismeretében a ξ értékére vonatkozóan fennáll.

3. Definició. Ha ξ és η véges értékkészletü valószínüségi változók, akkor a ξ feltételes entropiáját azzal a feltétellel, hogy η adott, a

$$H(\xi|\eta) = \sum_{y \in Y} p(y) H(\xi|\eta)_y =$$

$$= \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{1}{p(x|y)}$$

képlet definiálja. Az

$$l_{\xi|\eta} = \log \frac{1}{p(\xi|\eta)}$$

valószínüségi változót feltételes entrópiasürüségnek nevezzük. +++

Természetesen, ez a definició is magában foglalja azt az esetet, amikor ξ és η vektorváltozók. Például, ha $\eta_1, \eta_2, \ldots, \eta_n$ mindegyikének értékei az Y halmazból valók, és az Y elemeiből alkotható n hosszuságu sorozatok halmazát Y^n jelöli, akkor

$$H(\xi|\eta_1, \eta_2, \ldots, \eta_n) = \sum_{x \in X} \sum_{v \in Y^n} p(x,v) \log \frac{1}{p(x|v)} ,$$

ahol például $p(x|v) = P(\xi = x | \eta_1 = y_1, \eta_2 = y_2, \dots, \eta_n = y_n)$ ha $(y_1, y_2, \dots, y_n) = v \in Y^n$. Analóg jelöléseket használunk majd minden olyan esetben, amikor egyszerre sok valószínüségi változó szerepel.

Mivel $H(\xi|\eta)$ annak az információnak az átlagos mennyisége, amely az η értékének az ismeretében a ξ megadásához még szükséges, azt várjuk, hogy

$$H(\xi,\eta) = H(\eta) + H(\xi|\eta). \tag{12}$$

Ennek igazolásához elég a feltételes valószinüség definiciójából adódó

$$-\log p(x,y) = \iota_{\xi,\eta} = \iota_\eta + \iota_{\xi|\eta}$$

egyenlőség mindkét oldalának a várható értékét képezni. Több valószinüségi változó esetére (12)-ből teljes indukcióval egyszerüen kapjuk, hogy

$$H(\xi_1, \xi_2, \dots, \xi_n) = H(\xi_1) + \sum_{k=2}^{n} H(\xi_k | \xi_1, \xi_2, \dots, \xi_{k-1}). \tag{13}$$

(12)-ből ugyanis

$$H(\xi_1, \xi_2, \dots, \xi_n) = H(\xi_1, \xi_2, \dots, \xi_{n-1}) + H(\xi_n | \xi_1, \xi_2, \dots, \xi_{n-1})$$

következik, és itt az első tagot tovább bontogatva kapjuk (13)-at.

A feltételes entrópia néhány fontos tulajdonságát foglalja össze a következő lemma.

2. Lemma: Tetszőleges $f: Y \to Z$ leképezés esetén

$$H(\xi|\eta) \le H(\xi|f(\eta)),$$

ahol az egyenlőség feltétele $p(x|y) = P(\xi = x | f(\eta) = z)$ minden olyan (x,y,z), $x \in X$, $y \in Y$, $z \in Z$ hármasra, melynél $f(y) = z$ és $p(y) > 0$. +++

Bizonyitás: Mivel $P(f(\eta) = z) = \sum_{f(y)=z} p(y)$, és ugyanigy, $P(\xi = x, f(\eta) = z) = \sum_{f(y)=z} p(x,y)$; az 1. Lemmából

$$\sum_{f(y)=z} p(x,y) \log \frac{p(y)}{p(x,y)} \le P(\xi = x, f(\eta) = z) \log \frac{P(f(\eta) = z)}{P(\xi = x, f(\eta) = z)}$$

ahol az egyenlőség feltétele $p(x|y) = \text{const}$ ha $f(y) = z$ és $p(y) > 0$, és ez a konstans csak $P(\xi = x | f(\eta) = z)$ lehet. Ezt az egyenlőtlenséget x és z szerint összegezve éppen a lemmát kapjuk. +++

A lemma néhány következményét tárgyaljuk. Legyen először $f = \text{const}$, ekkor a ξ -nek csak egy feltételes eloszlása van, és az az eredetivel azonos, tehát a lemmából

(14) $H(\xi|\eta) \le H(\xi)$

adódik. Az egyenlőség $p(x|y) = p(x)$ feltétele azt jelenti, hogy (14)-ben egyenlőség pontosan akkor áll, ha ξ és η független. (13) és (14) összehasonlításából - η -nak tagonként a $(\xi_1, \xi_2, \dots, \xi_{k-1})$ véges értékkészletü vektorváltozót választva - kapjuk a

(15) $$H(\xi_1, \xi_2, \dots, \xi_n) \le \sum_{i=1}^{n} H(\xi_i)$$

egyenlőtlenséget, ahol az egyenlőség feltétele a $\xi_1, \xi_2, \dots, \xi_n$ valószinüségi változók teljes függetlensége.

A 2. Lemmából az $f(y,z) = y$ választással (mivel η ilyen függvénye az (η, ζ) párnak) a

(16) $H(\xi|\eta, \zeta) \le H(\xi|\eta)$

becslés adódik, ahol az egyenlőség $p(x|y,z) = p(x|y)$ feltétele azt fejezi ki, hogy ξ és ζ feltételesen független ha az η értéke adott, vagyis ha ξ, η, ζ ebben a sorrendben Markov-láncot alkot. A (14) egyenlőtlenség szemléletesen azt fejezi ki, hogy ha valamit - jelen esetben az η értékét - megtudunk, akkor a bizonytalanság csökken, kivéve azt az esetet, amikor a kapott információ a hiányzóról semmit sem mond, vagyis a ξ és az η

független. A (16) reláció hasonlóképpen értelmezhető.

Érdemes elgondolkodni a

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

reláció jelentésén is. Arról van itt szó, hogy ha az η már ismert, akkor a (ξ, η) pár megadásához már kevesebb információra van szükség, mint magának a ξ -nek a megadásához: az η értéke információt szolgáltat a ξ értékére vonatkozóan, kivéve azt az esetet, amikor a ξ és az η független. Ez motiválja a következő definiciót.

<u>4. Definició:</u> Az $I(\xi,\eta) = H(\xi) - H(\xi|\eta)$ mennyiséget a ξ és η véges értékkészletü valószinüségi változók kölcsönös információjának nevezzük. +++

Egyszerü számolás mutatja, hogy

$$I(\xi,\eta) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} ,$$

tehát a kölcsönös információ szimmetrikus mennyiség, nem ugy mint a feltételes entrópia. Mivel a feltételes entrópia sem lehet negativ, (14)-ből következik, hogy

(17) $0 \le I(\xi,\eta) \le H(\xi)$.

Megjegyezzük, hogy $I(\xi,\eta) = 0$ csak akkor teljesül, ha ξ és η független, mig $I(\xi,\eta) = H(\xi)$ csak akkor, ha ξ az η függvénye. Az első állitás a

(14)-hez füzött megjegyzésből következik, a második pedig közvetlenül is könnyen verifikálható.

(12) egyszerü következménye, hogy

$$(18)\qquad I(\xi,\eta) = H(\xi) - H(\xi|\eta) = H(\xi) + H(\eta) - H(\xi,\eta) = \\ = H(\eta) - H(\eta|\xi) = I(\eta,\xi).$$

Szükségünk lesz a következő egyszerü észrevételre is:

3. Lemma: A ξ és η tetszőleges f illetve g függvényével

$$I(f(\xi), g(\eta)) \le I(\xi,\eta). \quad +++$$

Bizonyitás: A 2. Lemma és a (18) reláció alkalmazásával

$$I(\xi,\eta) = H(\xi) - H(\xi|\eta) \ge H(\xi) - H(\xi|g(\eta)) = I(\xi, g(\eta)) =$$
$$= I(g(\eta),\xi) = H(g(\eta)) - H(g(\eta)|\xi) \ge H(g(\eta)) - H(g(\eta)|f(\xi)) =$$
$$= I(g(\eta), f(\xi)) = I(f(\xi), g(\eta)). \quad +++$$

A kölcsönös információ fogalmának igazi jelentőségét a III. fejezetben látjuk majd, de már most megjegyezzük, hogy vannak akik a kölcsönös információt tekintik elsődleges információmennyiségnek, és az entrópiát a $H(\xi) = I(\xi,\xi)$ azonossággal definiálják.

1.6. A feltételes entrópia maximuma és a Fano-egyenlőtlenség

Ebben a szakaszban két technikai jellegü segédeszközt, - két egyenlötlenséget - tárgyalunk, melyek a későbbiekben

kerülnek majd felhasználásra. A következő lemma a $H(\xi) \le \log n$ becslés általánositása.

4. Lemma: Ha $\eta = y$ esetén a ξ valószinüségi változó $m(y)$ számu értéket vehet fel pozitiv valószinüséggel, akkor

$$H(\xi|\eta) \le \sum_{y\in Y} p(y) \log m(y). \quad +++$$

Bizonyitás: Azt az 1. Lemmából tudjuk, hogy

$$H(\xi|\eta=y) \le \log m(y),$$

aminek a várható értékét képezve éppen a lemmát kapjuk. +++

A bizonyitásból az is kiolvasható, hogy egyenlőség csak akkor állhat, ha a $p(x|y)$, $x\in Y$ feltételes eloszlások mindegyike egyenletes eloszlás. Ebből az eredményből is következik, hogy ha η értéke a ξ értékét egyértelmüen meghatározza, vagyis $m(y) = 1$ ha $p(y) > 0$, akkor $H(\xi|\eta) = 0$.

Ennél lényegesen mélyebb összefüggést fejez ki a következő becslés, amely R.M. Fanotól származik.

3. Tétel: Tegyük fel, hogy a ξ és η valószinüségi változók ugyanazt az m értéket veszik fel pozitiv valószinüséggel, és legyen

$$P_e = \sum_x \sum_{y\ne x} p(x,y)$$

annak a valószinüsége, hogy $\xi \neq \eta$. Ekkor

$$H(\xi|\eta) \leq P_e \log(m-1) + P_e \log \frac{1}{P_e} + (1-P_e) \log \frac{1}{1-P_e} \quad . \; +++$$

Bizonyitás. Vegyük észre, hogy

$$H(\xi|\eta) = \sum_x \sum_{y \neq x} p(x,y) \log \frac{p(y)}{p(x,y)} + \sum_y p(y,y) \log \frac{p(y)}{p(y,y)} \quad .$$

Mivel

$$\sum_x \sum_{y \neq x} p(y) = \sum_x (1-p(x)) = m-1 ,$$

a P_e definiciójának figyelembevételével az 1. Lemmából azt kapjuk, hogy

$$\sum_x \sum_{y \neq x} p(x,y) \log \frac{p(y)}{p(x,y)} \leq P_e \log \frac{m-1}{P_e} =$$

$$= P_e \log(m-1) + P_e \log \frac{1}{P_e} \quad .$$

Másrészt, a második tagnál

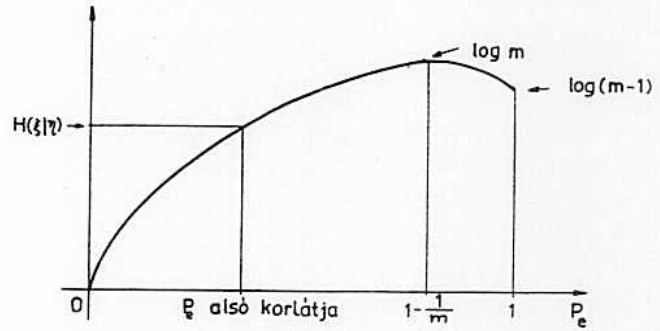
$$\sum_y p(y) = 1 \quad \text{és} \quad \sum_y p(y,y) = P(\xi = \eta) = 1 - P_e ,$$

tehát ugyancsak az 1. Lemmával

$$\sum_y p(y,y) \log \frac{p(y)}{p(y,y)} \leq (1-P_e) \log \frac{1}{1-P_e}$$

adódik. Ez a két becslés együttesen éppen a tétel állitását adja. +++

Nagyon fontos, hogy a Fano-egyenlőtlenség jobboldala, mint a P_e függvénye folytonos, a $[0, 1-\frac{1}{m})$ intervallumban szigoruan monoton, és csak $P_e = 0$ esetén 0 az értéke. (Lásd a 9. ábrát)



9. ábra

Ha tehát a ξ valószinüségi változót az η -val akarjuk helyettesiteni, akkor az itt elkövetett P_e hibára $H(\xi|\eta)$ függvényeként alsó becslés adható az ábrából leolvasható módon. Ez a becslés ugyan nem tulságosan

éles, de céljainknak meg fog felelni. A Fano-egyenlőtlenség értéke éppen az, hogy a $P_e = P(\xi \neq \eta)$ hibavalószínüséget éppen egy információelméleti mérőszámmal - a feltételes entrópiával becsüli meg.

II. INF[illegible] ZAJMENTES KÓDOLÁSA

2.1. A hirkö[illegible] matematikai modellje

A hirközlés [illegible] fogalmazva az, hogy a rendelkezésre álló i[illegible] ell juttatni a felhasználóhoz. Nagy távolságok, [illegible] tségek esetén ez persze közvetlenül nem lehetsé[illegible] különböző műszaki berendezések - távközlési csator[illegible] tására van szükség. Az első probléma onnan adódi[illegible] itandó információ megjelenési formája nagyon k[illegible] viszont a csatornán csak néhány, jól meghatározot[illegible] ehet átvinni; például bináris jegyeket, illetv[illegible] entáló elektromágneses jelformákat. Ezért a t[illegible] információt a csatorna típusának megfelelő formába kell alakítani. Ezt a műveletet kódolásnak nevezzük, míg a továbbitás után vett jelekből az információ eredeti (vagy ahhoz közelálló) formában történő reprodukálását dekódolásnak. A második probléma forrása az a körülmény, hogy a távközlési csatornák általában nem teljesen megbizhatóan működnek, és igy az átvitel során a továbbitott jelek megváltozhatnak. Olyan kódolási és dekódolási módszerekre van tehát szükség, melyekkel az ilyen zajos csatornákon is elég megbizhatóan vihető át az információ, és amellett az átvitel költségei sem lesznek nagyobbak a kelleténél.

A felvetett matematikai problémák szempontjából a hirközlési rendszerek 10. ábrán látható blokkdiagramja jól kifejezi