A KÜLFÖLDI SZAKIRODALOMBÓL*

AZ INFORMÁCIÓ-TOVÁBBÍTÁS ELMÉLETE

A. N. KOLMOGOROV**

I. AZ ELMÉLET KELETKEZÉSE ÉS TARTALMA

Amidőn másodszor lépek Akadémiánk közgyűlése elé, azzal a megjegyzéssel szeretném kezdeni, hogy mai előadásom témája szorosan összefügg annak az előadásnak a témájával, amelyet az Akadémia közgyűlésén 1947-ben *„Folytonos spektrumú rezgések statisztikai elmélete"* címen tartottam.[1]

Valóban, az információelmélet tartalmilag leggazdagabb és — tisztán matematikai oldalról nézve — legérdekesebb részét: a stacionárius üzemben dolgozó, folytonos közléseket továbbító távközlési csatornák elméletét, — nem lehetett volna megalkotni, ha korábban nem dolgozták volna már ki a stacionárius sztochasztikus folyamatok elméletét, és speciálisan, ezen folyamatok spektrális elméletét.

A távközlés és információ-megőrzés különböző feladatai iránt való érdeklődés elég régóta tart. Lényegében már régen felmerült egy információ „mennyisége" értékelésének problémája. Az a probléma, hogy egy információ mennyiségére be lehet-e vezetni valamilyen univerzális számszerű mértéket, különösen fontos azokban az esetekben, amidőn egy bizonyos típusú információt át kell alakítani minőségileg másfajta információvá. — Információ átalakítására tipikus feladat folytonosan változó argumentumú folytonos függvények tabellázásának feladata. Ha például a kétváltozós $f(x, y)$ függvényt

* A Szerkesztőség, Kolmogorov cikkének közlésével, új rovatot nyit, amelyben a nemzetközi szakirodalom legjelentősebb és legmodernebb közleményeit kívánja magyar nyelven hozzáférhetővé tenni.

** (Сессия Академии Наук СССР по научным проблемам автоматизации производства 15—20 Октября 1956 г. Пленарные заседания. Изд. А. Н. СССР, Москва 1957. 66—99 стр.)

[1] L. [1]. Azon kérdések matematikai oldalának részletesebb kifejtése, amelyeknek ezt az 1947-es előadásomat szenteltem, a [2] cikkben található. Jelenleg analóg szerepet játszik a II. FEJEZET, amely kis változtatásokkal annak az előadásnak az összefoglalása, amelyet B. V. GNYEGYENKO nevemben felolvasott a *Rádiómérnökök Amerikai Intézete* (*Massachusetts Institute of Technology)* információelméleti symposiumán (1956. szeptember); ez sok részében azon az előadáson alapszik, amelyet I. M. GELFAND- és A. M. JAGLOMmal együtt az *Össz-szövetségi Matematikai Kongresszuson* olvastam fel 1956. júniusában.

$(0 \leq x \leq 1,\ 0 \leq y \leq 1)$ ε pontossággal kell megadni és tudjuk, hogy növekménye, Δf, abszolút értékben nem nagyobb ε-nál, amikor

$$|\Delta x| \leq \varepsilon_x, \qquad |\Delta y| \leq \varepsilon_y,$$

akkor nyilván elegendő az f függvényt x szerint ε_x és y szerint ε_y lépésekkel tabellázni, azaz közelítőleg

$$N \sim \frac{1}{\varepsilon_x \varepsilon_y}$$

számú értékét adni meg. Ha megengedjük, hogy $|f| \leq 1$, a függvény egy értékének rögzítéséhez közelítőleg elég

$$K \sim \log_{10}\left(\frac{1}{\varepsilon}\right)$$

tizedesjegy. Így, ha a most vázolt tervet követjük, a táblázatba összesen közelítőleg

$$NK \sim \frac{1}{\varepsilon_x \varepsilon_y} \log_{10}\left(\frac{1}{\varepsilon}\right)$$

számú tizedesjegyet kell felvenni (nem számítva az x és y változók értékeinek felírását, ami szabványosan megy mindazon f függvényeknél, amelyek a mondott megkötéseknek eleget tesznek). Mindnyájan tudjuk, hogy a dolog ilyen primitív kezelése, különösen többváltozós függvények esetében, óriási terjedelmű és gyakorlatilag kivitelezhetetlen táblázatokra vezetne. Valóban, elég „síma" függvények tabellázásakor a független változók lépéseit a táblázatban jelentősen nagyobbra választják és az argumentum közbenső értékei esetén a függvény értékének megkeresésére interpolációhoz folyamadnak, valamilyen p rendűig (többváltozós függvények táblázatai esetén gyakran a negyedik, vagy hatodik rendűig) terjedő differenciák segítségével. Ezenfelül $f(x, y)$ első jegyeit, amelyek közeli argumentumoknál meg szoktak ismétlődni, nem írják ki adott (x, y) értékpároknak megfelelő egyes táblázatrekeszekben, hanem pl. csak a sor elején írják ki, amely sok ilyen rekeszből áll.

A táblázat-szerkesztés leírt módjai általánosan és régóta ismeretesek; ennek ellenére az idevágó általános elméleti vizsgálódások, amelyeknek ki kellene mutatniok, mi az a *minimális* információmennyiség, amely szükséges ahhoz, hogy adott ε pontossággal lerögzítsünk egy tetszőleges, csupán bizonyos általános feltételeknek alávetett f függvényt, jelenleg még kezdeti stádiumban vannak. Például, csak nemrég közöltem [3] cikkemben explicit alakban a

$$H_\varepsilon(F^n_{p,\alpha}) \asymp \left(\frac{1}{\varepsilon}\right)^{\frac{n}{p+\alpha}} \tag{1}$$

formulát, amely azt mutatja, milyen rendben növekedik $\varepsilon \to 0$-nál az az információmennyiség, amely szükséges ahhoz, hogy fixálhassunk egy korlátos tartományban adott $F^n_{p,\alpha}$ n-változós függvényt, amelynek p-nél kisebb rendű deriváltjai korlátosak, p-edrendű deriváltjai pedig α kitevőjű Lipschitz-feltételnek tesznek eleget.[2]

Teljesen világos, hogy egy információ transzformálásának és megőrzésének kérdései döntő jelentőségűek a modern számológépek konstrukciójának és alkalmazási módjainak kidolgozásánál. Itt a tárolóberendezés „memóriakapacitását" az ezen berendezés által tárolható kettes számrendszerbeli jegyek (0 vagy 1) számával jellemezzük. Az a módszer, hogy információmennyiséget úgy mérünk, hogy tetszőleges információt összehasonlítunk valamilyen számú kettes számrendszerbeli számjegy alakjában megadott információval, ma már szabványossá vált az információelméletben.

Áttérve a „távközlési csatornák" elméletéből való feladatokra, abban a formájukban, amely a táviró vagy távbeszélővonalak és hozzájuk hasonló berendezések működésével kapcsolatos, kezdjük azzal az igen egyszerű feladattal, hogy továbbítani kell tízes számrendszerbeli jegyek

$$a_1, a_2, \ldots, a_n, \ldots \tag{2}$$

sorozatát, kettes számrendszerbeli számok

$$b_1, b_2, \ldots, b_n, \ldots \tag{3}$$

sorozatának segítségével. Engedjük meg, hogy a (2) sorozat az időben úgy lép fel, hogy időegység alatt egy jegy mutatkozik. Azt kérdezzük, milyen sebességgel kell továbbítani a (3) sorozat kettes számrendszerbeli jegyeit, ha azt kívánjuk, hogy a kiindulási (2) sorozatnak a (3) sorozat alapján való reprodukálása az idővel korlátlanul növekedő késés nélkül történjék. Ha minden a_n jegyet egyenként leképezünk kettes számrendszerbeli jegyek segítségével, egy a_n jegyre *négy* kettes számrendszerbeli jegyet kell számítanunk (mert $2^3 = 8 < 10$ és csak $2^4 = 16 > 10$). Ha a (2) sorozat minden egyes (a_{2n-1}, a_{2n}) jegypárját kell továbbítani kettes számrendszerbeli jegyek egy csoportja segítségével, ehhez *hét* kettes számrendszerbeli jegyet kell felhasználni (minthogy $2^6 = 64 < 100$ és csupán $2^7 = 128 > 100$); három tízes számrendszerbeli jegy továbbításához elég *tíz* kettes számrendszerbeli jegyet igénybe venni (minthogy $2^9 < 1000$, de $2^{10} = 1024 > 1000$) é. i. t. A (3) sorozatbeli jegyek képzésének sebessége, vagyis azon kettes számrendszerbeli jegyek

[2] (1) baloldalát illetőleg l. a dolgozat II. részét, különösen a II. 3. §-t. Ezt a képletet A. G. Vituskin képletének nevezhetjük, minthogy ő [4] dolgozatában más viszonylatban világítja meg az $n/(p+\alpha)$ kitevő szerepét: ugyanitt lényegében megtalálható az (1) képlet bizonyításának fele, nevezetesen azon tény bizonyítása, hogy $H_\varepsilon(F^n_{p,\alpha})$ rendje nem lehet *kisebb* az (1) képlettel adottnál.

száma, amelyet átlagban időegység alatt továbbítani kell, egyenlő (az említett átviteli szisztémáknak megfelelően) a következőkkel:

$$v_1 = 4, \quad v_2 = 7/2 = 3{,}5; \quad v_3 = 10/3 = 3{,}33\ldots$$

Könnyen bebizonyítható, hogy az itt felbukkanó sorozat a

$$v = \log_2 10 = 3{,}32\ldots$$

határértékhez tart. Egyszersmind ez a v alsó korlátja is a (3) sorozat jegyei azon előállítási sebességeinek, amelyek mellett szisztematikusan növekvő késlekedés nélkül lehet a (2) sorozatot továbbítani.

Már ezen a kétségtelenül naiv példán is megfigyelhetünk sok olyan típikus jelenséget, amelyekkel az információ-továbbítás modern elméletei kidolgozása során akkor találkoztak, amidőn ezeket a távközlési csatornák lényegesen általánosabb típusaira kezdték alkalmazni.

1. Az információ-előállítás sebessége (ha nem tekintjük annak *minőségi* különbségeit) megenged bizonyos észszerű *kvantitatív* mérést. Speciálisan — ha lerögzítjük, hogy az időegység alatt átlagosan v jel jön, amelyek közül mindegyik k értéket vehet fel — az információelőállítás sebességének mértékéül logikus a következő számot tekinteni:

$$\bar{H} = v \log_2 k.$$

(A mi példánkban

$$\bar{H} = \log_2 10 = 3{,}32\ldots$$

a (2) sorozat esetében és

$$\bar{H}' = v \log_2 2 = v$$

a (3) sorozat esetében, hogyha időegység alatt v jelet létesítünk).

A következők szempontjából megjegyezzük, hogy pontosan (hibák nélkül) dolgozó távközlési csatornák esetében azt a sebességet, amellyel ezek információkat továbbítanak, logikus „csatornakapacitásnak“ nevezni. A mi esetünkben a (3) sorozat csatornakapacitása — hogyha ezt távközlési csatornaként fogjuk fel — $\bar{C} = \bar{H}' = v$-vel egyenlő. Általános esetben a csatornakapacitást kicsit másképp definiálják; erre később még visszatérünk. — Ezen megjegyzés után fogalmazzuk meg a második általános elvet:

2. Annak a szükséges feltétele, hogy egy $\bar{H}$ sebességgel keletkező információt szisztematikusan növekvő késlekedés nélkül továbbítani lehessen egy olyan távközlési csatornán át, amelynek csatornakapacitása $\bar{C}$;

$$\bar{H} \leqq \bar{C}.$$

Az összesek közt a leglényegesebb azonban a harmadik elv, amelynek a tartalma az alább közlendő megfogalmazásban még nem eléggé határozott, mégis talán ebben található az a leglényegesebb újdonság, amelyet SHANNON információ távközlési csatornákon át való továbbításának elméletére vonatkozó munkáiban nyilvánosságra hozott:

3. Már abban az esetben is, amidőn a továbbításra kerülő információ lényeges minőségi különbséget mutat fel azon információhoz viszonyítva, amelyet a távközlési csatorna közvetlenül felvenni és továbbítani képes —

$$\bar{H} < \bar{C}$$

fennállása esetén elvileg lehetséges szisztematikusan növekedő késlekedés nélkül lefolytatni a továbbítást, hogyha a továbbításra kerülő információt a távközlési csatornára kerülése előtt alkalmas módon „kódoljuk“, — azután pedig, hogy az a csatornát elhagyta, „visszakódoljuk“. Mindamellett azonban általánosságban véve az a helyzet, hogy amidőn H $\bar{C}$-hoz közeledik, a kódolás és visszakódolás módjai elkerülhetetlenül bonyolódnak és (amidőn $\bar{H} \rightarrow \bar{C}$) egyre jobban és jobban mutatkozik a továbbítás késlekedése.

Megjegyezzük, hogy az a késlekedés, amelyről a 3. elvben szó van, $\bar{H} \rightarrow \bar{C}$ esetén növekedik, ha azonban az információforrás állandó, és $\bar{H} < \bar{C}$, akkor bizonyos korlát adható neki, amelyik megmarad akkor is, amidőn a csatorna működése korlátlanul folytatódik.

Következő példa gyanánt tekintsük egy olyan szövegnek a továbbítását, amely az orosz ábécé betűiből áll. Minthogy ábécénkben 33 különböző betű van, formálisan

$$N \quad 33^n$$

különböző, n hosszúságú „szöveget“ (vagyis n betűből álló sorozatot) lehet képezni. Az az információmennyiség, amelyet az ilyen módon meghatározott szövegek közül valamelyik tartalmaz, a megelőzők szerint:

$$I = n \log 33$$

(itt és a következőkben mindenütt 2 alapú logaritmust használunk anélkül, hogy ezt külön megjegyeznénk). A gyorsírás létezése mutatja azonban, hogy a tényleges, a nyelvben előforduló szövegeket lényegesen rövidebb módon is lehet továbbítani. Ez teljesen érthető: az n betűből álló „értelmes“ szövegek N^* száma nyilvánvalóan összehasonlíthatatlanul kisebb N-nél. Ezért elvileg lehetséges egy olyan írás-szisztéma („ideális“ gyorsírás), amelyik ezen N^* számú értelmes szöveg közül akármelyiknek a rögzítésére csupán

$$I^* \sim \log N^*$$

kettes számrendszerbeli jegyet kívánna. A „szöveg komprimálása“ ezen elvi lehetőségének teljes kihasználása, persze, bármilyen gyorsírási rendszer vagy szövegkódolás segítségével történő megvalósítás határain kívül fekszik (amenynyiben ezek a módszerek egyszerű formális szabályoknak tesznek eleget): a ténylegesen használt nyelvben előforduló szövegek megszerkesztésének törvényszerűségei aligha lesznek bármikor is teljes alapossággal megfogalmazhatók.

Fordítsuk most figyelmünket a nyelvbeli szövegek csupán egyetlen sajátságára: arra, hogy különböző betűk különböző gyakorisággal fordulnak bennük elő. Ha rögzítjük az

$$n_1, n_2, \ldots, n_k$$

számokat, amelyek azt mutatják, hogy egy n (betű) hosszúságú szövegben az

$$a_1, a_2, \ldots, a_k$$

betűk hányszor fordulnak elő (természetesen itt

$$n_1 + n_2 + \cdots + n_k = n),$$

akkor az n (betű) hosszúságú szövegek száma

$$N' = \frac{n!}{n_1!\, n_2! \ldots n_k!}$$

-ra csökken le. Ha felhasználjuk az úgynevezett Stirling-formulát, amely itt egyébként csak a gyengébb

$$\log (n!) \sim n \log n$$

alakjában szükséges, könnyen kiszámítható, hogy nagy n_i-k esetén

$$I' = \log N' \sim -n \sum_i p_i \log p_i,$$

ahol a

$$p_i = \frac{n_i}{n}$$

mennyiségek az egyes betűk előfordulásának gyakoriságai. A kapott eredményt így fejezhetjük ki: p_i gyakoriságú egyes betűk felhasználása esetén a továbbítható információ mennyisége „a szöveg egy betűjére számítva" a következő:

$$H = -\sum_i p_i \log p_i. \tag{4}$$

Egyenlő gyakoriságok esetén, vagyis ha

$$p_1 = p_2 = \cdots = p_k = \frac{1}{k},$$

újból megkapjuk a

$$H = \log k$$

összefüggést, tetszőleges más p_i gyakoriságok esetén pedig

$$H < \log k.$$

Könnyen megadhatók olyan kódolási módszerek, amelyek lehetővé teszik, hogy tetszőleges szöveget, amelyben az $a_1, a_2, \ldots, a_k$ betűk $p_1, p_2, \ldots, p_k$ gyakoriságokkal fordulnak elő, úgy továbbíthassunk, hogy (elég hosszú szövegek továbbítása esetén) az átlagban egy betűre eső kettes számrendszerbeli jegyek száma tetszőleges kevéssel különbözzék attól a H-tól, amelyet a (4) képlet határoz meg. Ebben mégegyszer kitűnik a fentebb megfogalmazott elvek közül a harmadik.

Most azonban abba fogjuk hagyni elemi példák vizsgálatát és rátérünk az általános elmélet alapjaira.[3]

A (4) képlet a tisztán statisztikai értelmezésen kívül (amelyben a p_i-k gyakoriságok) valószínűségi értelmezést is megenged. Tegyük fel, hogy adva van a ξ valószínűségi változó

$$\mathbf{P}(\xi = x_i) = p_i \tag{5}$$

valószínűségeloszlása. Ha csupán ez az eloszlás van megadva, akkor arra a kérdésre, hogy az

$$x_1, x_2, \ldots, x_k$$

lehetséges értékek közül melyikkel lesz ξ a valóságban egyenlő, a válasz bizonyos értelemben határozatlan marad. Ha közöljük, hogy ξ egyenlő valamelyik meghatározott x_i-vel, megszüntetjük ezt a határozatlanságot, vagyis bizonyos kiegészítő információt közlünk a ξ mennyiségre vonatkozólag. Az [5] eloszlás határozatlanságának mértékét ennek ún. *entrópiája* fejezheti ki, amelyet a már ismert (4) képlet szolgáltat. Ugyanez a formula fejezi ki azt az információmennyiséget is, amelyik szükséges ahhoz, hogy megszüntessük azt a határozatlanságot, amelyik akkor mutatkozik, amidőn ξ-t csupán az (5) eloszlással adjuk meg, vagyis azt az információmennyiséget, amely ξ értéke pontos közlésében foglaltatik benne.

Tegyük fel továbbá, hogy adva van két valószínűségi változó, ξ és η együttes valószínűségeloszlása,

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

Ha $\eta = y_j$ adva van, akkor ξ feltételes eloszlással:

$$\mathbf{P}(\xi = x_i \mid \eta = y_j) = p_{i|j}$$

-vel fog rendelkezni. Az az információmennyiség, amelyet ξ pontos értékének megadása képvisel, feltéve, hogy az $\eta = y_j$ érték már ismeretes, — a következő:

$$H(\xi \mid \eta = y_j) = -\sum_i p_{i|j} \log p_{i|j},$$

ennek átlaga pedig a következő:

$$\mathbf{M}H(\xi|\eta) = -\sum_j \mathbf{P}(\eta = y_j) \sum_i p_{i|j} \log p_{i|j}.$$

[3] A szerző az általános elmélet *összefoglalását* adja. Az az olvasó, aki ebben a tárgykörben még teljesen tájékozatlan, a részleteket megtalálhatja BALATONI J. és RÉNYI A.: *„Az entrópia fogalmáról“* c. dolgozatában (MTA Mat. Kut. Int. Közl. I. (1956) 1—2. füzet 9—40. o.) Ez a dolgozat A. JA. HINCSIN később megemlítendő [5] dolgozatának anyagát is felöleli és tovább fejleszti; általában kiegészíti a jelen munkát. (Ford.)

Kézenfekvő feltenni, hogy az

$$I(\eta, \xi) = H(\xi) - \mathbf{M} H(\xi \mid \eta) \tag{6}$$

különbség az a ξ-re vonatkozó információmennyiség, amelyik már benne van η megadásában. Könnyen kiszámítható, hogy $I(\xi, \eta)$-t a következő alakban írhatjuk:

$$I(\eta, \xi) = \sum_{i,j}{}' p_{ij} \log \frac{p_{ij}}{\mathbf{P}(\xi = x_i)\mathbf{P}(\eta = y_j)}, \tag{7}$$

amelyben ξ és η szerepe teljesen egyforma: a ξ-re vonatkozó információ mennyisége η-ban, valamint az η-ra vonatkozó információ mennyisége ξ-ben számszerint egymással egyenlők.

Könnyen bebizonyítható, hogy a

$$I(\xi, \xi) = H(\xi) \tag{8}$$

összefüggés mindig fennáll.

Most már fáradság nélkül meghatározható egy információ olyan távközlési csatornán át való továbbításának sebessége, amely hibákkal dolgozik és egy ilyen csatorna csatornakapacitása (a megelőzők szerint a csatornakapacitást úgy definiáljuk, mint információtovábbítás lehetséges sebességeinek felső határát). Vegyünk még egy elemi példát. Tegyük fel, hogy a „bemenő oldalán" rávisszük egy csatornára az η jeleket, amelyek 0 vagy 1-gyel egyenlők, a „kimenő oldalán" pedig ezeknek megfelelően η' jeleket kapunk és emellett

$$\mathbf{P}(\eta' = 0) = 1 - \Delta, \quad \mathbf{P}(\eta' = 1) = \Delta, \quad \text{ha} \quad \eta = 0,$$
$$\mathbf{P}(\eta' = 0) = \Delta, \quad \mathbf{P}(\eta' = 1) = 1 - \Delta, \quad \text{ha} \quad \eta = 1.$$

Könnyen kiszámítható, hogy a

$$C = \sup I(\eta, \eta') = 1 + [\Delta \log \Delta + (1 - \Delta) \log (1 - \Delta)] \tag{9}$$

felső határt akkor érjük el, amikor

$$\mathbf{P}(\eta = 0) = \mathbf{P}(\eta = 1) = \frac{1}{2}.$$

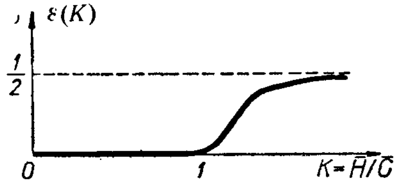
A (9) képlet szerint az az $I(\eta, \eta')$ információmennyiség, amelyet (egy betűre számítva) az említett csatornával továbbíthatunk, 0-sal egyenlő, hogyha $\Delta = \frac{1}{2}$. Ez teljesen érthető, minthogy valószínűségszámítási szempontból nézve ebben az esetben az η és η' jelek függetlenek. A csatornakapacitás $C = 1$ maximális értékét a $\Delta = 0$ és $\Delta = 1$ esetekben kapjuk meg. A $\Delta = 0$ esetben a csatorna hiba nélkül dolgozik: 1 valószínűséggel $\eta = \eta'$. Bár a második esetben 1 valószínűséggel $\eta \neq \eta'$, η ebben az esetben is könnyen rekonstruálható η' alapján: egyszerűen 0 helyett 1-et kell olvasni és megfordítva. Az összes fennmaradó esetekben

$$0 < C < 1.$$

Amikor SHANNON a negyvenes évek vége felé felvetette azt a gondolatot, hogy információmennyiséget a (4), (6), (7) képletekkel mérjenek és megfigyelte, hogy olyan távközlési csatornák esetén is, amelyek hibákkal dolgoznak, az információforrásra és a csatorna szerkezetére vonatkozó elég általános feltételek mellett is érvényben marad a fentebb megfogalmazott második és harmadik elv, — ezzel egyúttal a modern információelmélet alapjait is lefektette, olyan diszciplina formájában, amely szisztematikus fejlődésre képes.

Minden nagy felfedezésben vannak n e m v á r t elemek. Ebben is különbözik egy nagy felfedezés a mindennapos tudományos munka fokozatos eredmény-felhalmozódásától. Rámutatok itt arra, hogy mit látok én magam minőségileg újnak és nem vártnak, már az információelmélet fentebb vázolt elemeiben is:

1. Első vizsgálatra az „információ" nem szkaláris mennyiség. Információk megjelenési formái rendkívül különbözők lehetnek. Már korábban is várható volt, hogy ilyen vagy olyan módszerek jöhetnek szóba egy információ „mennyiségének" mérésére, nem volt azonban világos, lehetséges-e minőségileg különböző információkat (amelyekre ez vagy az a megfelelő kvantitatív mérték egyforma volt) távközlési csatornákon át való továbbításuk vagy memória-berendezésekben való megőrzésük nehézsége szempontjából ténylegesen is ekvivalenseknek tekinteni.



1. ábra

Bebizonyosodott, hogy létezik információ mennyiségének egy ilyen kiváltképpen „szabályos" mértéke, és ez lehetővé teszi egész sereg olyan probléma végleges megoldását, amelyeknél egyáltalában nem volt világos, hogy a megoldás apriori független-e az információ legfinomabb kvalitatív sajátságaitól.

2. Hibákkal dolgozó távközlési csatornák (vagy memóriaberendezések) esetében attól lehetett tartani, hogy elég kis valószínűséggel beálló hibával bíró továbbítás elérése a továbbítási sebesség igen nagymérvű csökkenését vonja maga után. Ezzel az aggodalommal szemben azonban kiderült, hogy a $\bar{H} < \bar{C}$ feltétel fennállása esetén rögtön lehetséges lesz információ továbbítása úgy, hogy hiba fellépésének valószínűsége tetszőleges kicsi legyen. Ennek a ténynek a megvilágítására a következőkre utalok. Legyen adva egy távközlési csatorna, amelynek a csatornakapacitása C. Tekintsük azt a problémát, hogy ezen át olyan, kettes számrendszerbeli jegyeket kell továbbítani, amelyek időegység alatt H számban keletkeznek. Akkor az egyes jegyek továbbításakor jelentkező hiba valószínűségének tetszőleges közelséggel elérhető minimuma univerzális $\varepsilon(K)$ függvénye lesz a $K = \bar{H}/\bar{C}$ hányadosnak (1. az 1. ábrán). $K < 1$ esetén $\varepsilon(K)$ értéke zérus, K további növekedésekor azonban nőni kezd és

$K \to \infty$ esetében 1/2-hez tart (analitikusan $K > 1$-nél az $\varepsilon(K)$ függvényt az $\frac{1}{K} = 1 + \varepsilon \log \varepsilon + (1-\varepsilon) \log (1-\varepsilon)$, $(\varepsilon < 1/2)$ egyenletből kell meghatározni). Függvényünk tehát $K \leq 1$-nél pontosan zérussal egyenlő értékekből kiindulva csupán $K \geq 1$-nél kezd el növekedni.

Persze, ahogy már korábban is megjegyeztük, C-hoz közeli H-knál az, hogy azt akarjuk elérni, hogy a hibának kis valószínűsége legyen, gyakran igen bonyolult szabályokat von maga után az információ kódolására vonatkozólag, és azt eredményezi, hogy a csatorna kimenő oldalánál az információ „kiadása" nagy késéssel fog történni. Elvileg fontos tény azonban a bonyolult kódolási eljárásoknak a hibák kiküszöbölése szempontjából kétségtelenül nagy hatóereje is; ennek az alkalmazásaira még visszatérünk.

Információmennyiség mérése SHANNON által javasolt módjainak (vagyis a (4) és (7) képleteknek) a megalapozása kétféleképp képzelhető el.

Először, kimondhatók ilyen vagy olyan kézenfekvő axiómatikus követelések, amelyeknek az információmennyiség mértékének eleget kell tennie. A $H(\xi)$ mennyiséggel kapcsolatban — vagyis ha a (8) képletet feleltetjük meg annak az információmennyiségnek, amelyet egy ξ valószínűségi változóra vonatkozólag magának a ξ változónak a megadása tartalmaz, — az információelmélet megalapozása problémájának ilyen tárgyalásmódja mintaszerűen világos formában található meg A. JA. HINCSIN [5] munkájában. Érdekes volna ennek kapcsán megalkotni egy ugyanannyira természetes axiómatikát egy valószínűségi változóban egy másik valószínűségi változóra vonatkozólag tartalmazott $I(\xi, \eta)$ információmennyiség általánosabb fogalmával kapcsolatban is. A nehézség ott van, (l. ennek az előadásnak a II. FEJEZETÉT), hogy a folytonos esetben $H(\xi)$ és $H(\eta)$, egyenként véve, általában végtelenek és $I(\xi, \eta)$-t nem lehet a (6) képlet szerint számítani ki.

Az axiómatikus irányzat kutatásainak mérlegeként lerögzíthetjük, hogy $I(\xi, \eta)$-n kívül nem létezhet más, valamennyire is természetes általános szkaláris jellemzője annak az információnak, amelyet egy ξ valószínűségi változó egy másik, η valószínűségi változóra vonatkozólag tartalmaz. Mindazonáltal, minthogy „információ" természeténél fogva nem kötelezően szkaláris mennyiség (a valóságban nem is az) semmiféle, az említett irányzatot követő axiómatikus vizsgálat nem adhat feleletet arra a kérdésre, milyen teljességgel jellemzi az $I(\xi, \eta)$ mennyiség a minket érdeklő „információt". Amint a megelőző fejtegetésekből érthető is, erre a kérdésre bizonyos feleletet azok a tételek adnak, amelyek alapjai az információtovábbítás elméletének fentebb megfogalmazott második és harmadik általános alapelve. Ezen tételek értelmében az esetek egész kiterjedt osztályánál a távközlési csatorna vagy a memóriaberendezés információ továbbítására vagy megőrzésére való alkalmasságát meglehetős teljességgel jellemezhetjük egyetlen C számmal (ha időegység alatti viszonyokról van szó, C-sal), magának az információnak továbbítására vagy megőrzésre való alkalmasságát pedig egyetlen H számmal (vagy, ha időegység alatti viszonyokról van szó $\bar{H}$-sal. Mindennél az információ kvalitatív sajátosságai lényegtelennek mutatkoznak. Nehéz kiértékelni ilyenfajta eredmények jelentőségét, ha számbavesszük, hogy milyen széles körben alkalmaznak a modern technikában pl. olyan átalakításokat, mint a televízióban továbbítandó képek átalakítása elektromos rezgésekké és megfordítva (az információtovábbítás elméletének terminológiája szerint ez speciális esete a „kódolásnak" és „visszakódolásnak") s. i. t. Egyúttal meg kell említeni, hogy bármennyire is elbűvölők az információelmélet elgondolásai, egy információ kvalitatív sajátosságainak említett háttérbe szorulása csak bizonyos közelítéssel és meghatározott feltételek mellett valósul meg. Ezek a feltételek az alaposan megvizsgált esetekben — durván szólva — abból állnak, hogy nagymennyiségű homogén információt kell felhalmozni úgy, hogy emellett bonyolult kódolási módszereknek is megvalósíthatók legyenek, és az információleadás utóbbinál fellépő késlekedésének sem szabad káros kihatásúnak lennie. Emellett azt is világosan kell látni, hogy ezen elképzelések realitásának pontos matematikai megalapozása egyelőre csak eléggé korlátozott feltevések mellett történt meg. Magának SHANNONnak a munkáiban is, melyek gondolatokban kétségtelenül igen gazdagok — a tárgyalásmód gyakran igen ködös. Elég általános feltételek mellett diszkrét jeleket továbbító stacionáriusan dolgozó távközlési csatornák esetében a „Shannon-féle tételeket" csak később bizonyították be kifogástalanul, tiszta matematikával foglalkozó kutatók, egész sor munkájukban. Az ilyen irányú munkák közt legjobban lezárt jellege A. JA. HINCSIN [6] munkájának van.

Az információelméletnek az alapvető matematikai kutatásokon jelentősen túlmenő tárgyalásával, amelyet az alkalmazási irányzat kutatói dolgoztak ki, a szovjet olvasó elég teljességgel megismerkedhetik a [7], [8], [9] könyvek alapján ([7]-ben benne foglaltatik SHANNON alapvető [10] munkájának a fordítása, GOLDMAN [9] könyvének orosz fordítása pedig nemsokára megjelenik).

Az információelméletnek a nehezebben kezelhető folytonos esetre való alkalmazását illetőleg az elmélet leglényegesebb elemei már SHANNON előtt is felbukkantak. Az információmennyiség egy olyan logaritmikus mértéke, amely teljesen analóg a Shannon-féle $H(\xi)$ kifejezéssel, alapja azoknak az aszimptotikus statisztikai módszereknek, amelyeket FISHER még 1921—1925-ben kidolgozott (l. [11] 32—33. fej.)[4]

[4] Általánosan ismert az a körülmény is, hogy a $H(\xi)$ kifejezés formálisan azonos az entrópiának a fizikában használt kifejezésével. Ezt az egybeesést én teljesen elegendőnek tartom annak indokolására, hogy a $H(\xi)$ mennyiséget az információelméletben is „entrópiá-

SHANNON munkáihoz még közelebb állnak V. A. KOTYELNYIKOV eredményei [12], amelyekhez még 1933-ban jutott. Ezekben fogalmazta meg ő információ folytonos jelek segítségével történő továbbítása spektrális elméletének alapgondolatát, amelyről részletesen fogok beszélni ennek az előadásnak a II. FEJEZETÉBEN.

II. AZ ELMÉLET FELÉPÍTÉSÉNEK ALAPELVEI FOLYTONOS KÖZLEMÉNYEK ESETÉBEN

Eléggé rávilágítottunk arra, hogy az információelméletben és közlemények diszkrét jelek segítségével való továbbításának elméletében mi a szerepe egy ξ valószínűségi változó

$$H(\xi) = -\sum_k p_k \log p_k$$

entrópiájának (ξ az $x_1, x_2, \ldots, x_n$ értékeket $p_1, p_2, \ldots, p_n$ valószínűségekkel veheti fel). A továbbiakban arra a gondolatra támaszkodom, hogy az az alapvető fogalom, amely tökéletesen tetszőleges folytonos közlemények és jelek esetére is általánosítható, nem közvetlenül maga az entrópia fogalma, hanem a ξ valószínűségi változóban foglalt, az η változóra vonatkozó $I(\xi, \eta)$ információmennyiség fogalma. A diszkrét esetben ezt a mennyiséget szabályszerűen kiszámíthatjuk SHANNON ismert

$$I(\xi, \eta) = H(\eta) - \mathbf{M} H(\eta \mid \xi)$$

képlete alapján.[5]

Sűrűségfüggvénnyel bíró véges dimenziós eloszlások esetére SHANNON az $I(\xi, \eta)$ mennyiséget a következő analóg képlettel definiálja:

$$I(\xi, \eta) = h(\eta) - \mathbf{M} h(\eta \mid \xi),$$

ahol $h(\eta)$ a „differenciális entrópia":

$$h(\eta) = -\int p_\eta(y) \log p_\eta(y)\, dy,$$

$h(\eta \mid \xi)$ pedig az analóg módon definiált feltételes differenciális entrópia. Általánosan ismert, hogy a $h(\xi)$ mennyiségnek nincs közvetlen reális értelme, amellett nem is invariáns az x-ek terében lefolytatott koordináta-transzfor-

nak" nevezik: az ilyen matematikai analógiákat mindig alá kell húzni, mert ha rájuk tereljük a figyelmet, azzal előmozdíthatjuk a tudomány fejlődését. Túlzás volna azonban úgy tekinteni a dolgokat, hogy azok a fizikai elméletek, amelyek az entrópia fogalmával kapcsolatosak, már kész formában magukban rejtik az információelmélet elemeit is: entrópiatípusú kifejezéseket információmennyiség mértékeként, — amennyire én tudom, — először FISHER használ előbb említett munkájában.

[5] Ezekben a számomra célszerűnek látszó jelölésekben $H(\eta \mid x)$ η feltételes entrópiája $\xi = x$ feltétel mellett, $\mathbf{M} H(\eta \mid \xi)$ pedig ezen feltételes entrópia várható értéke, változó ξ esetén.

mációval szemben. Végtelen dimenziós eloszlások esetére a $h(\xi)$ kifejezésnek általában nincs is analogonja.

Bizonyos sajátos értelemben a ξ folytonos eloszlású valószínűségi változó entrópiája mindig végtelen. A folytonos jelek csak azért nem szolgálhatnak korlátlanul nagyterjedelmű információ továbbítására, mert mindig csak korlátolt pontossággal figyelhetjük meg őket. Utóbbi folytán — miután már megadtuk a megfigyelés ε pontosságát — logikus definiálni a ξ valószínűségi változó ezen ε-nak megfelelő $H_\varepsilon(\xi)$ „ε-entrópiáját". Ezt tette SHANNON is (ő a „közlések létesítésének sebessége" elnevezést használta). Bár nem változtat a dolog lényegén, hogy erre a mennyiségre új elnevezést vezetünk be, legyen szabad javasolnom az említett elnevezést, mert az jobban kihangsúlyozza a fogalom széleskörű jelentőségét, valamint a közönséges pontos entrópiával való mély analógiáját. Már most rámutatok arra, hogy — amint arról a 3. §-ban szó lesz — az ε-entrópiára érvényben marad az a tétel, amely szerint a normális eloszlásnak (mind a végesdimenziós, mind pedig a végtelen dimenziós esetben) extrémális szerepe van. — Az 1—2 §-okban, nem is törekedve feltétlenül újat hozni, megadom az $I(\xi, \eta)$ mennyiség definíciójának és alapvető sajátságainak absztrakt megfogalmazását, valamint a Shannon-féle közleménytovábbítási elmélet kiindulási problematikájának áttekintését. A 3—6 §-okban bizonyos konkrét eredményekről lesz szó, amelyeket szovjet matematikusok a legutóbbi időben kaptak. Különösen hangsúlyozni akartam az ε-entrópia $\varepsilon \to 0$ esetén való aszimptotikus viselkedése vizsgálatának fontosságát. Az itt nyitott perspektívák megértéséhez hasznos szolgálatot tehet [3] dolgozatom, amelyben a tárgyalás más fogalmakkal történik.

1. §. Egy valószínűségi változóban egy másikra vonatkozólag bennfoglalt információ mennyisége

Legyenek ξ, ill. η valószínűségi változók, amelyek lehetséges értékei az X, ill. Y halmazokból valók, legyenek továbbá

$$P_\xi(A) = \mathbf{P}(\xi \in A), \quad P_\eta(B) = \mathbf{P}(\eta \in B)$$

a megfelelő valószínűségeloszlások és

$$P_{\xi\eta}(C) = \mathbf{P}((\xi, \eta) \in C)$$

a ξ és η valószínűségi változók együttes eloszlása. Definíció szerint a ξ valószínűségi változóban az η valószínűségi változóra vonatkozólag bennfoglalt információ mennyiségét adja meg az

$$(1) \qquad I(\xi, \eta) = \int_X \int_Y P_{\xi\eta}(dx\,dy) \log \frac{P_{\xi\eta}(dx\,dy)}{P_\xi(dx) P_\eta(dy)}$$

képlet. Ennek a képletnek a pontos értelme bizonyos magyarázatra szorul, mert az $I(\xi, \eta)$ mennyiségnek a továbbiakban közlendő általános tulajdonságai csak akkor érvényesek, ha bizonyos halmazelméleti jellegű korlátozásokat teszünk a P_ξ, P_η és $P_{\xi\eta}$ eloszlásokra; most azonban nem időzöm ennél a kérdésnél. Mindenesetre az általános elméletet különösebb fáradság nélkül fel lehet építeni olyan formában, hogy aztán egészen általános természetű ξ és η valószínűségi változókra (vektorokra, függvényekre, általánosított függvényekre s. i. t.) egyaránt alkalmazni lehessen. Az (1) definíciót SHANNON-nak lehet tulajdonítani, bár ő csak a

$$P_\xi(A) = \int_A p_\xi(x)\,dx, \quad P_\eta(B) = \int_B p_\eta(y)\,dy$$

$$P_{\xi\eta}(C) = \iint_C p_{\eta\xi}(x, y)\,dx\,dy$$

esettel foglalkozott, mikor is (1) az alábbiba megy át:

$$I(\xi, \eta) = \int_X \int_Y p_{\xi\eta}(x, y) \log \frac{p_{\xi\eta}(x, y)}{p_\xi(x)\,p_\eta(y)}\,dx\,dy.$$

Olykor hasznos lesz, ha a $P_{\xi\eta}$ eloszlást a

$$P_{\xi\eta}(C) = \iint_C a(x, y)\,P_\xi(dx)\,P_\eta(dy) + S(C) \tag{2}$$

alakban állítjuk elő, ahol az $S(C)$ függvény szinguláris a

$$P_\xi \times P_\eta$$

szorzatra vonatkozólag. Ha az S szinguláris komponens eltűnik, az

$$\alpha_{\xi\eta} = a(\xi, \eta) \tag{3}$$

képlet 1 valószínűséggel egyértelműen meghatározza az $\alpha_{\eta\xi}$ valószínűségi változót. Olykor hasznos lesz az I. M. GELFAND és A. M. JAGLOM által a [13]-ban megfogalmazott következő

TÉTEL: *Ha $S(X \times Y) > 0$ akkor $I(\xi, \eta) = \infty$. Ha $(X \times Y) = 0$, akkor*

$$\begin{aligned} I(\xi, \eta) &= \int_X \int_Y a(x, y) \log a(x, y)\,P_\xi(dx)\,P_\eta(dy) \\ &= \int_X \int_Y \log a(x, y)\,P_{\xi\eta}(dx\,dy) = \mathbf{M} \log \alpha_{\xi\eta}. \end{aligned} \tag{4}$$

Soroljuk fel az $I(\xi, \eta)$ mennyiség néhány alapvető tulajdonságát:

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

II. $I(\xi, \eta) \geqq 0$; $I(\xi, \eta) = 0$ csak akkor áll fenn, ha ξ és η függetlenek.

III. Ha a (ξ_1, η_1) és (ξ_2, η_2) párok függetlenek,

$$I((\xi_1, \xi_2), (\eta_1, \eta_2)) = I(\xi_1, \eta_1) + I(\xi_2, \eta_2).$$

IV. $I((\xi, \eta), \zeta) \geqq I(\eta, \zeta)$.

V. $I((\xi, \eta), \zeta) = I(\eta, \zeta)$ csakis akkor áll fenn, ha ξ, η, ζ Markov-féle sorozat, azaz ha ζ feltételes eloszlása, rögzített ξ és η mellett, csupán η-től függ.

A IV. tulajdonságot illetőleg hasznos megjegyezni a következőket:

A

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

entrópia esetében a (ξ, η) pár entrópiájára vonatkozó

$$H(\xi, \eta) \geqq H(\xi), \quad H(\xi, \eta) \geqq H(\eta)$$

alsó becsléseken kívül, amelyek I. és IV-ből következnek, érvényes a következő felső becslés is:

$$H(\xi, \eta) \leqq H(\xi) + H(\eta).$$

Arra az információra, amelyet ζ a (ξ, η) párra vonatkozólag tartalmaz, nincs analóg becslés:

$$I(\xi, \zeta) = 0, \quad I(\eta, \zeta) = 0$$

-ból még nem következik — amint az elemi példákkal igazolható —, hogy fennáll az

$$I((\xi, \eta), \zeta) = 0$$

egyenlőség.

A továbbiak szempontjából emeljük ki azt a speciális esetet, amidőn ξ és η valószínűségi vektorok:

$$\xi = (\xi_1, \ldots, \xi_m),$$

$$\eta = (\eta_1, \ldots, \eta_n) = (\xi_{m+1}, \ldots, \xi_{m+n}),$$

a

$$\xi_1, \xi_2, \ldots, \xi_{m+n}$$

változók pedig normális eloszlásúak és második centrális momentumaik

$$s_{ij} = \mathbf{M}[(\xi_i - \mathbf{M}\xi_i)(\xi_j - \mathbf{M}\xi_j)].$$

Ha a

$$C = |s_{ij}| \qquad 1 \leqq i, j \leqq m+n$$

determináns zérustól különböző, akkor — ahogy ezt I. M. GELFAND és A. M. JAGLOM kiszámította —

$$I(\xi, \eta) = \frac{1}{2} \log \frac{AB}{C}, \tag{5}$$

ahol

$$A = |s_{ij}|_{1 \leqq i, j \leqq m}, \quad B = |s_{ij}|_{m < i, j \leqq m+n}.$$

Egyébként gyakran célszerű egy másik tárgyalásmód, amely akkor is alkalmazható, ha a $C > 0$ korlátozást elhagyjuk. Mint tudjuk [14], az X és Y tér koordinátáinak alkalmas lineáris transzformációjával elérhető, hogy az összes

s_{ij} második momentumok — azokat kivéve, amelyekre $i=j$ vagy $j=m+i$, — zérussá válnak. Ha így választjuk meg a koordinátákat,

(6) $$I(\xi, \eta) = -\frac{1}{2}\sum_k [1-r^2(\xi_k, \eta_k)],$$

ahol az összegezés azon

$$k \leq \min(m, n)$$

értékekre történik, amelyekre a korrelációs együttható

$$r(\xi_k, \eta_k) = \frac{s_{k,m+k}}{\sqrt{s_{k,k} \cdot s_{m+k,m+k}}}$$

kifejezésében levő nevező zérustól különböző.

2. §. Shannon elmélete alapjainak absztrakt tárgyalása

SHANNON a

$$\xi \to \eta \to \eta' \to \xi'$$

szkéma szerint lefolyó közlemény-továbbításokat vizsgálja, ahol az

$$\eta \to \eta'$$

„továbbító berendezést“ az η' „kimenő jelnek“ adott η „bemenő jel“ esetén fennálló

$$P_{\eta'|\eta}(B'|y) = \mathbf{P}(\eta' \in B' | \eta = y)$$

feltételes eloszlása jellemzi, valamint a bemenő jel P_η eloszlásának bizonyos

$$P_\eta \in V$$

korlátozása. A „kódolás“

$$\xi \to \eta$$

és a „visszakódolás“

$$\eta' \to \xi'$$

műveleteit a következő feltételes eloszlások jellemzik:

$$P_{\eta|\xi}(B|x) = \mathbf{P}(\eta \in B | \xi = x)$$
$$P_{\xi'|\eta'}(A'|y') = \mathbf{P}(\xi' \in A' | \eta' = y').$$

SHANNON alapproblémája a következőképp fogalmazható meg: Adva vannak a ξ „bemenő közlemények“, a ξ' „kimenő közlemények“, az η bemeneti jelek és az η' kimeneti jelek lehetséges értékeinek X, X', Y, Y' terei, adva vannak a továbbító berendezés jellemzői, vagyis a $P_{\eta'|\eta}$ feltételes eloszlások, és a bemenő jel számára megengedhető P_η eloszlások V osztálya; végül, adva van a bemenő oldalra jutó közlemény

$$P_\xi(A) = \mathbf{P}(\xi \in A)$$

eloszlása, valamint a „továbbítás pontosságának követelményei", $P_{\xi\xi'} \in W$, ahol W a bemenő oldalra jutó közlemény és kimenő oldalon kijövő közlemény

$$P_{\xi\xi'}(C) = \mathbf{P}((\xi, \xi') \in C)$$

együttes eloszlásának valamilyen osztálya.

Azt kérdezzük, lehetséges-e kódolási és visszakódolási szabályokat megadni és — ha igen — hogyan kell hangozniok ezeknek (vagyis milyeneknek kell lenniök a $P_{\eta|\xi}$ és $P_{\xi'|\eta'}$ feltételes eloszlásoknak), hogy amidőn a P_ξ, $P_{\eta|\xi}$, $P_{\eta'|\eta}$, $B_{\xi'|\xi}$ eloszlások alapján s azon feltétel mellett, hogy a

$$\xi, \eta, \eta', \xi'$$

sorozat Markov-féle, a $P_{\xi\xi'}$ eloszlást kiszámítjuk,

$$P_{\xi\xi'} \in W$$

-t kapjunk?

SHANNONnal megegyezésben definiáljuk a továbbító berendezés „csatornakapacitását" a következő képlettel:

$$C = \sup_{P_\eta \in V} I(\eta, \eta')$$

és vezessük be a

$$H_W(\xi) = \inf_{P_{\xi\xi'} \in W} I(\xi, \xi')$$

mennyiséget, amelyet SHANNON — időegységre vonatkoztatása esetén — a „közlemények létesítése sebességének" nevez. Ekkor az 1. §-beli V. tulajdonságból azonnal következik a továbbítás lehetőségének szükséges feltétele:

$$H_W(\xi) \leq C. \tag{7}$$

Mint már említettük, összehasonlíthatatlanul mélyebb SHANNONnak az az ötlete, hogy elég hoszú idő óta működő „távközlési csatornák" esetében a (7) feltétel bizonyos értelemben és bizonyos igen tág feltételek mellett egyszersmind „majdnem elégséges" is. Matematikai szempontból nézve itt arról van szó, hogy ilyen típusú határeloszlástételeket kell bebizonyítani: Tegyük fel, hogy az X, X', Y, Y' terek, a P_ξ és $P_{\eta'|\eta}$ eloszlások, a V és W osztályok és így a C és $H_W(\xi)$ mennyiségek is bizonyos T paramétertől függnek (az alkalmazásokban ez a továbbító berendezés működése időtartamának a szerepét játsza.) Bizonyos egészen általános jellegű feltételek mellett bebizonyítandó, hogy a

$$\liminf_{T \to \infty} \frac{C^T}{H_W^T(\xi)} > 1 \tag{8}$$

feltétel elég nagy T-k esetén elegendő ahhoz, hogy lehetséges legyen a fentebb kimondott feltételeknek eleget tevő továbbítás. Ilyen megfogalmazás mellett a feladat persze kissé határozatlan. Mindamellett szántszándékkal kerültem itt, hogy a stacionárius sztochasztikus folyamatok elméletének terminológiájához folyamodjam, minthogy a stacionárius jelleg feltételezése nélkül is lehetséges igen érdekes eredményeket kapni a körvonalazott problémakörben [15].

Az említett típusú határeloszlástételekkel a diszkrét esetben számos jelentős munka foglalkozik. Közülük különösen ki kell emelnem A. JA HINCSIN már említett [6] munkáját.

Az általánosságban vizsgált folytonos esettel — különös módon — még úgyszólván egyáltalán nem is foglalkoztak.

3. §. Az ε-entrópia kiszámítása és becslése egyes speciális esetekben

Ha a

$$P_{\xi\xi'} \in W$$

feltételt úgy fogalmazzuk meg, hogy ξ és ξ' 1 valószínűséggel pontosan egyezzék, vagyis, hogy

$$\mathbf{P}(\xi = \xi') = 1$$

legyen, akkor

$$H_W(\xi) = H(\xi).$$

Ennek megfelelően természetesen adódik, hogy az általános esetben $H_W(\xi)$-t „a ξ valószínűségi változóhoz W reprodukálási pontosság esetén tartozó entrópiának" nevezzük.

Tegyük fel most, hogy az X' tér egybeesik X-szel, vagyis, hogy azokat a módozatokat vizsgáljuk, amelyekkel a $\xi \in X$ pont helyzetére vonatkozó közlést ugyanezen X tér ξ' pontja megadása segítségével közelítőleg továbbíthatunk, — továbbá, hogy e térben bevezettünk bizonyos $\varrho(x, x')$ „távolságot", amely eleget tesz a „metrikus terek" szokásos axiómáinak. Kézenfekvő megkövetelni, hogy

$$\mathbf{P}\{\varrho(\xi, \xi') \leq \varepsilon\} = 1 \qquad (W_\varepsilon^0),$$

vagy, hogy

$$\mathbf{M}\varrho^2(\xi, \xi') \leq \varepsilon^2 \qquad (W_\varepsilon)$$

érvényes legyen.

A P_ξ eloszlás „ε-entrópiájának" ezt a két fajtáját

$$H_{W_\varepsilon^0}(\xi) = H_\varepsilon^0(\xi)$$

-vel, ill.

$$H_{W_\varepsilon}(\xi) = H_\varepsilon(\xi)$$

-vel fogjuk jelölni.

Ami a $H_\varepsilon^0(\xi)$ ε-entrópiát illeti, itt csak bizonyos becsléseket akarok ismertetni

$$H_\varepsilon^0(X) = \sup_{P_\xi} H_\varepsilon^0(\xi)\text{-re}$$

vonatkozólag, (a felső határ az X tér összes P_ξ valószínűségeloszlásaira értendő). $\varepsilon=0$-nál, mint ismeretes,

$$H_0^0(X)=\sup_{P_\xi} H(\xi)=\log N_X,$$

ahol N_X az X halmaz elemeinek a száma. $\varepsilon>0$ esetén

$$\log N_X^e(2\varepsilon)\leq H_\varepsilon^0(X)\leq \log N_X^a(\varepsilon),$$

ahol $N_X^a(\varepsilon)$ és N_X^e az X tér azon karakterisztikái, amelyeket [3] dolgozatomban vezettem be. Az $N_X(\varepsilon)$ függvények $\varepsilon\to 0$-nál mutatkozó aszimptotikus sajátságai, — amelyeket egész sor konkrét X tér esetére [3]-ban tanulmányoztam, — érdekes analogonjai a $H_\varepsilon(\xi)$ függvény továbbiakban kifejtendő aszimptotikus tulajdonságainak.

Tekintsük most a $H_\varepsilon(\xi)$ ε-entrópiát. Ha X az n-dimenziós euklidesi tér és

$$P_\xi(A)=\int_A p_\xi(x)\,dx_1, dx_2, \ldots, dx_n,$$

akkor — legalább is elég sima $p_\xi(x)$ függvény esetén — fennáll a jól ismert

$$H_\varepsilon(\xi)=n\log\frac{1}{\varepsilon}+[h(\xi)-n\log\sqrt{2\pi e}]+o(1) \tag{9}$$

formula, ahol

$$h(\xi)=-\int_X p_\xi(x)\log p_\xi(x)\,dx_1, dx_2, \ldots, dx_n$$

az ún. „differenciális entrópia", amely már SHANNON legelső munkáiban is szerepel. Így az n-dimenziós térben értelmezett elég sima folytonos eloszlások esetén $H_\varepsilon(\xi)$ aszimptotikus viselkedését e l s ő s o r b a n a tér dimenziója határozza meg és a $h(\xi)$ differenciális entrópia csak mint a $H_\varepsilon(\xi)$ kifejezés második tagja szerepel.

Kézenfekvő azt várni, hogy a végtelen dimenziós tér tipikus eloszlásai esetén $H_\varepsilon(\xi)$ növekedése ($\varepsilon\to 0$ esetén) lényegesen gyorsabb lesz. Legegyszerűbb példaként tekintsük a Wiener-féle $\xi(t)$ sztochasztikus folyamatot, amely $0\leq t\leq 1$-re van definiálva, s amelynek a

$$\Delta\xi=\xi(t+\Delta t)-\xi(t)$$

független növekményei normális eloszlásúak, amellett

$$\xi(0)=0, \qquad \mathbf{M}\Delta\xi=0, \qquad \mathbf{M}(\Delta\xi)^2=\Delta t.$$

A. M. JAGLOM azt találta, hogy ebben az esetben az L^2 tér metrikájában

$$H_\varepsilon(\xi)=\frac{4}{\pi}\frac{1}{\varepsilon^2}+o\left(\frac{1}{\varepsilon^2}\right). \tag{10}$$

Egy a $t_0\leq t\leq t_1$ időintervallumon értelmezett diffúziós típusú Markov-folyamat esetére, amelynél

$$\mathbf{M}\Delta\xi = A(t, \xi(t))\Delta t + o(\Delta t); \quad \mathbf{M}(\Delta\xi)^2 = B(t, \xi(t))\Delta t + o(\Delta t)$$

bizonyos kézenfekvő feltételeket megengedve az előbbinél általánosabb

$$H_\varepsilon(\xi) = \frac{4}{\pi}\chi \cdot \frac{1}{\varepsilon^2} + o\left(\frac{1}{\varepsilon^2}\right) \tag{11}$$

képlet nyerhető, ahol

$$\chi(\xi) = \int_{t_0}^{t_1} \mathbf{M} B(t, \xi(t))\,dt.$$

Az n-dimenziós euklidesi, vagy Hilbert-féle térben értelmezett n-dimenziós normális eloszlás esetében a H_ε ε-entrópia pontosan kiszámítható: egy ξ n-dimenziós vektor a koordináták alkalmas ortogonális transzformációja után a

$$\xi = (\xi_1, \xi_2, \ldots, \xi_n)$$

alakot ölti, ahol a ξ_k koordináták kölcsönösen függetlenek és normális eloszlásúak. Adott ε mellett határozza meg a θ paramétert az

$$\varepsilon^2 = \sum \min(\theta^2, \mathbf{D}^2\xi_k)$$

egyenlet; ekkor, normális eloszlású ξ esetén

$$H_\varepsilon(\xi) = \frac{1}{2} \sum_{\mathbf{D}^2\xi_k > \theta^2} \log \frac{\mathbf{D}^2\xi_k}{\theta^2}. \tag{12}$$

A

$$\xi' = (\xi'_1, \xi'_2, \ldots, \xi'_n)$$

approximáló vektort úgy kell megválasztani, hogy $\mathbf{D}^2\xi_k \leq \theta^2$ esetén

$$\xi'_k = 0,$$

$\mathbf{D}^2\xi_k > \theta$ esetén pedig

$$\xi_k = \xi'_k + \Delta_k, \quad \mathbf{D}^2\Delta_k = \theta^2, \quad \mathbf{D}^2\xi'_k = \mathbf{D}^2\xi_k - \theta^2$$

álljon fenn és a ξ_k és Δ_k vektorok kölcsönösen függetlenek legyenek. A végtelen dimenziós eset semmiben sem különbözik a véges dimenzióstól.

Végül igen lényeges, hogy az (n-dimenziós vagy végtelen dimenziós) ξ vektorra — adott második centrális momentumok esetén — $H_\varepsilon(\xi)$ akkor vesz fel maximális értéket, ha ξ normális eloszlású.

Ez az eredmény akár közvetlenül is levezethető M. SZ. PINSZKER következő tételéből (l. [16]):

TÉTEL: *Legyen adva az s_{ij} $(0 \leq i, j \leq m + n)$ értékek pozitív definit szimmetrikus matrixa, valamint a*

$$\xi = (\xi_1, \xi_2, \ldots, \xi_m)$$

vektor P_ξ eloszlása, ξ második centrális momentumai legyenek egyenlők az

s_{ij}*-kel, ha* $0 \leq i, j \leq m$. *Tegyük fel, hogy a* ξ *vektor és a*

$$\xi' = (\xi_{m+1}, \xi_{m+2}, \ldots, \xi_{m+n})$$

vektor együttes $P_{\xi\xi'}$ *eloszlására szóló W feltétel abból áll, hogy a*

$$(\xi_1, \xi_2, \ldots, \xi_{m+n})$$

változók második centrális momentumai egyenlők az s_{ij}*-kel, ha* $0 \leq i, j \leq m+n$; *akkor*

$$H_W(\xi) \leq \frac{1}{2} \log \frac{AB}{C}. \tag{13}$$

A (13) képlet jelölései az 1. § tárgyalásmódjának felelnek meg. Az 1. §-ban mondottakkal való egybevetéséből világos, hogy a (13) egyenlőtlenségből egyenlőség lesz, ha P_ξ normális eloszlás.

A „közlemények létesítésének sebessége" kiszámításánál fellépő variációs feladatok megoldásának elveire SHANNON meglehetősen régen rámutatott már. [10]-ben SHANNON és WEAVER így ír: *„Sajnos, ezeket a formális megoldásokat speciális esetekben nehéz numerikusan kiértékelni és így értékük nem valami nagy".*[6] Mindazonáltal sok ilyen típusú feladat lényegében elég egyszerű, ahogy az a fentebbiekből is látható. Lehetséges, hogy az ez irányban végzett vizsgálatok lassú fejlődése azzal kapcsolatos, hogy nem értik át eléggé azt a körülményt, hogy a típikus esetekben a variációs feladatok megoldásai igen gyakran degeneráltak; például annál a fentebb vázolt feladatnál, amelyben egy normális eloszlású ξ vektor esetére kellett $H_\varepsilon(\xi)$-t kiszámítani, az n-dimenziós esetben gyakran megtörténik, hogy a ξ' vektor nem n-dimenziós, hanem csak $k (k < n)$ dimenziós, — a végtelen dimenziós esetben pedig a ξ' vektor mindig véges dimenziós.

4. §. Az információ mennyisége és a közlemények létesítésének sebessége stacionárius folyamatok esetében

Tekintsünk két stacionárius, egyszersmind stacionárius kapcsolatban levő folyamatot:

$$\xi(t), \quad \eta(t) \qquad -\infty < t < \infty.$$

Jelöljük ξ_T és η_T-vel a ξ és η folyamatok $0 < t \leq T$ időre vonatkozó szakaszait, ξ_- - és η_--szal pedig a ξ és η folyamatok lefolyását a $-\infty < t \leq 0$ negatív féltengelyen. A stacionárius kapcsolatban levő ξ és η folyamatok (ξ, η) párját megadni, ugyanazt jelenti, mint megadni egy a t tengely mentén történő eltolásra vonatkozólag invariáns $P_{\xi\eta}$ valószínűségeloszlást az $\{x(t), y(t)\}$

[6] [7] orosz fordításában: 28. §. 79. o.

függvénypárok terében. Ha ξ_--t rögzitjük, a $P_{\xi\eta}$ eloszlásból a

$$P_{\xi_T,\eta|\xi_-}(C|x_-) = \mathbf{P}\{(\xi_T, \eta) \in C | \xi_- = x_-\}$$

feltételes eloszlás lesz. Ennek az eloszlásnak a segítségével a 1. §-beliek alapján kiszámítható az

$$I(\xi_T, \eta | x)$$

feltételes információmennyiség.

Ha az

$$\mathbf{M} I(\xi_T, \eta | \xi_-)$$

várható érték véges valamilyen $T > 0$-ra, akkor az összes többi $T > 0$ értékekre is véges és

$$\mathbf{M} I(\xi_T, \eta | \xi_-) = T \vec{I}(\xi, \eta).$$

Az $\vec{I}(\xi, \eta)$ mennyiséget kézenfekvő így nevezni: „az η folyamatról szóló információ létesítési sebessége a ξ folyamat megfigyelésekor". Ha a ξ folyamat teljes pontossággal extrapolálható a multból a jövőbe, akkor

$$\vec{I}(\xi, \eta) = 0.$$

Ez lesz a helyzet, ha például a ξ folyamat spektruma korlátos. Általánosságban véve nem okvetlenül áll fenn az

$$\vec{I}(\xi, \eta) = \vec{I}(\eta, \xi) \tag{14}$$

egyenlőség. A ξ folyamat „regularitására" tett elég általános feltételek mellett fennáll azonban[7] az

$$\vec{I}(\xi, \eta) = \overline{I}(\xi, \eta)$$

egyenlőség, ahol

$$\overline{I}(\xi, \eta) = \lim_{T \to \infty} \frac{1}{T} I(\xi_T, \eta_T).$$

Minthogy $I(\xi_T, \eta_T) = I(\eta_T, \xi_T)$, ezért mindig fennáll, hogy

$$\overline{I}(\xi, \eta) = \overline{I}(\eta, \xi)$$

és ebből kifolyólag az $\vec{I}(\xi, \eta) = \overline{I}(\eta, \xi)$ és $\vec{I}(\eta, \xi) = \overline{I}(\eta, \xi)$ egyenletek együttes fennállása esetén (14) is érvényes. Tegyük most fel, hogy W két — ξ és ξ' — stacionárius kapcsolatban levő folyamat $P_{\xi\xi'}$ együttes eloszlásainak valamilyen osztálya. A

$$\vec{H}_W(\xi) = \inf_{P_{\xi\xi'} \in W} \vec{I}(\xi', \xi)$$

mennyiséget kézenfekvő így nevezni: „A közlemények létesítésének sebessége

[7] A folyamat regularitása itt és a továbbiakban azt jelenti, durván szólva, hogy a folyamat olyan szakaszai, amelyek a t tengely két, egymástól elég távol levő szakaszának felelnek meg, majdnem függetlenek. Gauss-folyamatok esetén itt alkalmazható a regularitásnak az a jól ismert definíciója, amelyet [17] munkámban vezettem be.

a ξ folyamatban, W reprodukálási pontosság esetében". A ξ folyamat reguláritására vonatkozó bizonyos feltételek mellett és a W-re tett feltevések bizonyos kézenfekvő típusaira bebizonyítható, hogy

$$\bar{H}_W(\xi) = \bar{\bar{H}}_W(\xi),$$

ahol

$$\bar{\bar{H}}_W(\xi) = \inf_{P_{\xi\xi'} \in W} \bar{I}(\xi, \xi').$$

5. §. Az információmennyiség és a közleménylétesítés sebességének kiszámítása és becslése a spektrum alapján

Abban az esetben, amidőn a $P_{\xi\eta}$ eloszlás normális és a ξ és η folyamatok közül legalább az egyik reguláris, érvényes az M. SZ. PINSZKER által [18]-ban közölt

$$\bar{I}(\xi, \eta) = -\frac{1}{4\pi} \int_{-\infty}^{\infty} \log\left[1 - r^2(\lambda)\right] d\lambda \tag{15}$$

képlet, ahol

$$r^2(\lambda) = \frac{|f_{\xi\eta}(\lambda)|^2}{f_{\xi\xi}(\lambda)\, f_{\eta\eta}(\lambda)}$$

($f_{\xi\xi}, f_{\xi\eta}, f_{\eta\eta}$ spektrális sűrűségek). Diszkrét t idejű folyamatok esetében a normális folyamat időegységre vonatkoztatott differenciális entrópiájára,

$$\bar{h}(\xi) = \lim_{T \to \infty} \frac{1}{T} h(\xi_1, \xi_2, \ldots, \xi_T)$$

-re ismeretes a következő egyenlőség:

$$\bar{h}(\xi) = \log(2\pi \sqrt{e}) + \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_{\xi\xi}(\lambda)\, d\lambda. \tag{16}$$

Mindazonáltal folytonos idő és nem korlátos spektrum esetében a $\bar{h}(\xi)$ kifejezésnek semmilyen analogonja sincs és a Pinszker-féle formulát külön le kell vezetni.

A ξ stacionárius folyamatnak a stacionárius és ξ-vel stacionárius kapcsolatban levő ξ' folyamat segítségével történő reprodukálása pontosságát kézenfekvő a

$$\sigma^2 = \mathbf{M}[\xi(t) - \xi'(t)]^2$$

mennyiséggel jellemezni, és

$$\sigma^2 \leq \varepsilon^2$$

típusú W feltétel esetén kézenfekvő a

$$\bar{H}_\varepsilon(\xi) = \bar{H}_W(\xi)$$

mennyiséget a ξ folyamat időegységre vonatkozó ε-entrópiájának nevezni, — azon feltétel mellett pedig, hogy

$$\bar{H}_W(\xi) = H_W(\xi),$$

úgy hívni, hogy: „a közlemények létesítésének sebessége a ξ folyamatban, a továbbítás ε átlagos pontossága esetén". A véges dimenziós eloszlásokra vonatkozó megfelelő feltételekből (1.3. §) levezethető, hogy adott $f_{\xi\xi}(\lambda)$ spektrális sűrűség esetén a $\bar{H}_\varepsilon(\xi)$ mennyiség akkor éri el maximumát, ha ξ normális folyamat. A normális esetben a $\bar{H}(\xi)$ mennyiség könnyen kiszámítható az $f_{\xi\xi}(\lambda)$ spektrális sűrűség alapján, teljesen analóg módon ahhoz, ahogy azt a 3. §-ban kifejtettük, amikor n-dimenziós eloszlások esetére foglalkoztunk a $H_\varepsilon(\xi)$ mennyiséggel. A θ paramétert a következő egyenlet definiálja:

$$\varepsilon^2 = \int_{-\infty}^{\infty} \min\left(\theta^2, f_{\xi\xi}(\lambda)\right) d\lambda. \tag{17}$$

Ennek a paraméternek a segítségével a $\bar{H}_\varepsilon(\lambda)$ mennyiséget a

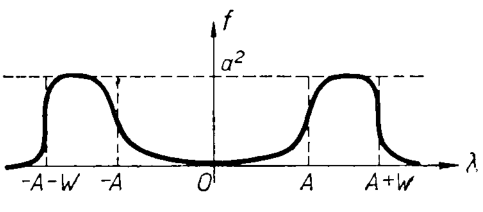
$$\bar{H}_\varepsilon(\xi) = \frac{1}{2} \int\limits_{f_{\xi\xi}(\lambda) > \theta^2} \log \frac{f_{\xi\xi}(\lambda)}{\theta^2}\, d\lambda \tag{18}$$

képlet alapján számíthatjuk ki.

Gyakorlati érdekessége van a 2. ábrán feltüntetett spektrális sűrűségeknek, amelyeket jól lehet approximálni a

$$\varphi(\lambda) = \begin{cases} a^2, & \text{ha} \quad A \leq |\lambda| \leq A + W \\ 0 & \text{egyébként} \end{cases}$$

fügvénnyel. Könnyen kiszámítható, hogy ebben az esetben *nem túlságosan kicsi* ε értékek esetén normális folyamatra közelítőleg:



2. ábra

$$\theta^2 \sim \frac{\varepsilon^2}{2W}, \qquad \bar{H}_\varepsilon(\xi) \sim W \log \frac{2Wa^2}{\varepsilon^2}. \tag{19}$$

A (19) képlet, persze, nem más, mint SHANNON jól ismert

$$R = W \log \frac{Q}{N} \tag{20}$$

formulája. Mindazonáltal van ebben bizonyos elvi újdonság, éspedig az, hogy most látjuk, miért és milyen határok közt (nem túl kicsiny ε-nál) alkalmazható ez a képlet nem korlátos spektrumú folyamatokra — a közlemények továbbítása elméletében szereplő összes, minket valóban érdeklő folyamatok pedig ilyenek.

Ha (19)-et

(21) $$H_\varepsilon(\xi) \approx 2W \left[\log (a \sqrt{2W}) + \log \frac{1}{\varepsilon} \right]$$

alakban írjuk és összevetjük (9)-cel, látjuk, hogy a felhasznált frekvenciasáv $2W$ kétszeres szélessége játsza a dimenziószám szerepét. Ezt a gondolatot — ti., hogy a frekvencia kétszeres szélessége ekvivalens a bizonyos értelemben az időegységre jutó dimenziószámmal — minden bizonnyal V. A. KOTYELNYIKOV mondta ki először (l. [12]). Ennek a gondolatnak a megalapozásaként KOTYELNYIKOV arra a körülményre utalt, hogy egy függvényt, amelynek a spektruma a $2W$ szélességű sávban helyezkedik el, egyértelműen meghatároznak a függvénynek a

$$\ldots, -\frac{2}{2W}, -\frac{1}{2W}, 0, \frac{1}{2W}, \frac{2}{2W}, \ldots, \frac{k}{2W}, \ldots$$

pontokban felvett értékei. Ugyanez a bizonyításmenet SHANNONnál is megvan, aki a segítségével kapott eredményeket a (20) képlet levezetésénél is felhasználta. Mivel egy korlátos spektrumú függvény a [17]-beli értelemben mindig szinguláris és egy ilyen függvény megfigyelése egyáltalán nincs kapcsolatban újabb információ stacionárius „odaáramlásával", ilyenfajta okoskodás értelme nem volt teljesen világos, úgy hogy a (21) közelítő formula előbb bemutatott új levezetésében — azt hiszem — van valami érdekesség.

Kis ε-oknál tetszőleges normális eloszlású reguláris sztochasztikus folyamat esetén $\bar{H}_\varepsilon(\xi)$ növekedése ε csökkenésekor lényegesen gyorsabban folyik le, mint ahogy az a (21) képletből adódnék. Speciálisan, ha $f_{\xi\xi}(\lambda)$ $\lambda \to \infty$ esetén $\lambda^{-\beta}$ rendű, akkor $H_\varepsilon(\xi)$ $\varepsilon^{-2(\beta-1)}$ rendű.

6. §. A csatornakapacitás kiszámítása és becslése egyes speciális esetekben

Vizsgáljuk először azt az esetet, amikor a bemenő jel egy

$$\eta_i = (\eta_{i1}, \ldots, \eta_{im})$$

m-dimenziós vektor, a kimenő jel pedig egy

$$\eta_i' = (\eta_{i1}', \ldots, \eta_{in}')$$

n-dimenziós vektor. Ahogy már mondtuk, az $\eta_i \to \eta_i'$ továbbító berendezést jellemzik a $P_{\eta_i'|\eta_i}$ feltételes valószínűségeloszlás, valamint az η_i bemenő jelre tett bizonyos megkötések. Tegyük fel, hogy η_i' η_i-tól való függése lineáris normális korreláció jellegű, vagyis, hogy

(22) $$\eta_i' = A\eta_i + \zeta,$$

ahol az A operátor lineáris, a ζ vektor pedig független η_i-tól és n-dimenziós

Gauss-eloszlású. Ami a bemenő jelre vonatkozó feltételt illeti, engedjük meg, hogy ez

(23) $$\mathbf{M}\, Q(\eta) \leqq \varepsilon^2$$

típusú, ahol a Q az $\eta_1, \ldots, \eta_m$ koordináták valamilyen pozitív definit kvadratikus alakja. Ha az Y és Y' terekben alkalmas módon választunk meg lineáris koordináta-transzformációkat, az általános eset arra az esetre vezethető vissza, amidőn

(24) $$Q(\eta) = \Sigma \eta_i^2$$

és ekkor a (22) összefüggés a következő alakban írható:

(25) $$\begin{aligned} \eta_i' &= a_i \eta_i + \zeta_i, \quad a_i \neq 0 \\ \eta_i' &= \zeta_i \end{aligned} \quad , \text{ ha } \begin{cases} 1 \leqq i \leqq k \\ i > k, \end{cases}$$

ahol $k = \min(m, n)$.

A tett feltevések mellett a csatornakapacitás, vagyis az $I(\eta, \eta')$ információmennyiségnek a tett feltevésekkel összhangban álló C felső korlátja könnyen kiszámítható. Ezt a korlátot akkor érjük el, ha az η_i-k kölcsönösen független Gauss-eloszlású valószínűségi változók, és

(26) $$\mathbf{D}^2 \eta_i = \theta^2 - \frac{\mathbf{D}^2 \zeta_i}{a_i}, \text{ ha } i \leqq k \text{ és } a_i \theta^2 < \mathbf{D}^2 \zeta_i,$$

$$\mathbf{D}^2 \eta_i = 0 \qquad \text{minden más } i\text{-re},$$

ahol a θ^2 konstanst, mint egyszerűen belátható, egyértelműen meghatározzák a (26) megkötések és

(27) $$\sum_i \mathbf{D}^2 \eta_i^2 = \varepsilon^2.$$

Az ennek megfelelő $C = I(\eta, \eta')$ érték (vö. 3. §.)

(28) $$C = \frac{1}{2} \sum_{a_i^2 \theta^2 > \mathbf{D}^2 \zeta_i} \log \frac{a_i^2 \theta^2}{\mathbf{D}^2 \zeta_i}.$$

Analóg helyzettel van dolgunk stacionárius lineáris távközlési csatornák esetében, ha „a zaj Gauss-szerű“, vagyis ha

(29) $$\eta' = A \eta(t) + \zeta(t),$$

ahol A lineáris operátor, $\zeta(t)$ pedig egy, az $\eta(t)$-től független Gauss-féle stacionárius sztochasztikus folyamat. Mint tudjuk, a megfelelő spektrális sűrűségek között a következő kapcsolat áll fenn:

(30) $$f_{\eta'\eta'}(\lambda) = a^2(\lambda) f_{\eta\eta}(\lambda) + f_{\zeta\zeta}(\lambda).$$

Azt a feltételt, amely korlátot szab meg a csatorna bemenő teljesítményére, a következő alakban fogalmazzuk meg:

(31) $$\int_{-\infty}^{\infty} f_{\eta\eta}(\lambda)\, d\lambda \leqq \varepsilon^2.$$

A feladat formális megoldása teljesen analóg a fentebb vizsgált véges dimenziós feladat megoldásával. A C maximális csatornakapacitást akkor érjük el, ha

$$
(32)\qquad \begin{aligned} f_{\eta\eta}(\lambda) &= \theta^2 - \frac{f_{\zeta\zeta}(\lambda)}{a^2(\lambda)} \\ f_{\eta\eta}(\lambda) &= 0 \end{aligned}\quad,\ \text{ha}\ \left\{ \begin{aligned} \frac{f_{\zeta\zeta}(\lambda)}{a^2(\lambda)} &< \theta^2 \\ \frac{f_{\zeta\zeta}(\lambda)}{a^2(\lambda)} &\geqq \theta^2. \end{aligned} \right.
$$

A megfelelő $\bar{I}(\eta, \eta')$ érték a (15) képlet szerint a következő:

$$
(33)\qquad \bar{C} = \frac{1}{4\pi} \int\limits_{a^2(\lambda)\theta^2 > f_{\zeta\zeta}(\lambda)} \log \frac{a^2(\lambda)\theta^2}{f_{\zeta\zeta}(\lambda)}\, d\lambda.
$$

Mindjárt meg kell jegyezni, hogy a gyakorlatilag érdekes esetekben a (32) képletek szerint kiszámított $f_{\eta\eta}(\lambda)$ spektrális sűrűség zérussá válik a λ tengely bizonyos korlátos szakaszán kívül és így egy szinguláris folyamat spektrális sűrűsége lesz. Az $\eta'(t)$ folyamat ekkor (azon kézenfekvő feltevés mellett, hogy a $\zeta(t)$ folyamat reguláris) — vegyes lesz, ti. az $A\eta$ szinguláris komponens és a ζ reguláris komponens összegéből fog állani. Az η' folyamat megfigyelésekor a η folyamatról kapott információ létesítésének $\vec{I}(\eta', \eta)$ sebességét ilyen esetben egyáltalán nem adja meg az $\vec{I}(\eta', \eta) = \bar{C}$ képlet: az ti. zérussal egyenlő (l. 5. §.).

Bizonyos kiegészítő vizsgálatok segítségével lehetséges azonban kimutatni, hogy ha az $\eta(t)$ bemenő jel reguláris folyamat, — ami elvileg gyakorlatilag megvalósítható — és valóban szállít az idő folyamán keletkező információt, kaphatunk olyan $\vec{I}(\eta', \eta) = \vec{I}(\eta, \eta')$ információ-létesítési sebességet, amely tetszőlegesen közel van $\bar{C}$-hoz. Ez véglegesen igazolja a (33) formulát.

III. AZ ALKALMAZHATÓSÁG HATÁRAI

Már említettem, hogy az „információ", ha elsődleges tartalmát nézzük, nem szkaláris mennyiség és általában az a helyzet, hogy egyetlen egy szám, „az információ mennyisége" segítségével nem lehet teljesen jellemezni. Egy közvetlenül meg nem figyelhető θ objektumról a megfigyelhető objektumban tartalmazott „információnak" (és nem információmennyiségnek) az előbbinél sokkal szélesebb körű, kvalitatív fogalma — bizonyos feltételeket megengedve — szintén formális matematikai tárgyalás alá vonható; már FISHER megkezdte ezt, a jelen dolgozat elején említett munkáiban. Ez sajátságos téma, amelyről most nem beszélhetek részletesen. Ehelyett tekintsünk egy példát, amelyik azt mutatja, mennyire hibás volna azt gondolni, hogy a Shannon-féle „információmennyiség" kimerítően jellemzi mindazon kapcsolatokat, amelyek felhasználhatók arra, hogy egyes objektumokat megfigyelve, másokról is ítéletet mondhassunk. Tegyük fel, hogy a ξ, η, ζ változók bármelyike csak két értéket vehet fel, 0-t és 1-et. Tegyük fel, hogy mind a négy lehetséges értékkombinációjuk,

$$\begin{matrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{matrix}$$

egyenként 1/4—1/4 valószínűségű. Könnyen igazolható, hogy — valószínűségszámítási értelemben véve — a (ξ, η), (η, ζ), (ζ, ξ) párok mindegyikében a változók kölcsönösen függetlenek. Ezért

$$I(\xi, \zeta) = I(\eta, \zeta) = 0.$$

Másrészt, ha ξ és η értékei ismertek, ζ értéke egyértelműen meghatározható a következő táblázat szerint:

$\eta \diagdown \xi$	0	1
0	1	0
1	0	1

Ezért

$$I[(\xi, \eta), \zeta] = H(\zeta) = 1.$$

Az a ζ-ra vonatkozó információmennyiség tehát, amelyet a ξ, ill. η valószínűségi változó tartalmaz, egyenként mindegyiknél zérus, (ξ, η) azonban már meghatározott (sőt kimerítő) információt tartalmaz ζ-ra vonatkozólag, amelynek a mennyiségét az 1 szám fejezi ki.

Fentebb már szó volt arról, hogy sok esetben, amidőn homogén adatok hosszú sorozatait kell továbbítani, adott távközlési csatornán át való továbbításuk elvi lehetősége meglehetős teljességgel jellemezhető a bennük tartalmazott információmennyiséggel. Rámutattunk, hogy ezen alapgondolatok alkalmazhatósági határainak felkutatása az egyik alapvető feladat az elmélet további kidolgozásában. Pontos információ továbbítása esetére (akár hibákkal dolgozó csatornán keresztül is) már sok vizsgálat történt e téren és további pozitív eredmények is remélhetők. *Nem pontos* információ továbbítása esetében azonban másként áll a dolog. Képzeljük el, hogy a fentebb vizsgált

$$\xi \to \eta \to \eta' \to \xi'$$

szkémában a bemenő oldali ξ közlés kettős számrendszerbeli jegyek sorozata, amely egy oly másik ξ_0 sorozatból származik, mely egymástól független, 1/2—1/2 valószínűséggel fellépő zérusokból és egyesekből áll, emellett a

ξ sorozat minden egyes jegye csupán $1-J<1$ valószínűséggel esik egybe a ξ_0 sorozat megfelelő jegyével. Ekkor a ξ sorozatban a ξ_0 sorozatra vonatkozólag tartalmazott információmennyiséget (egy számjegyre vonatkoztatva) a már ismert

$$I=\bar{I}(\xi, \xi_0)=1+J\log J+(1-J)\log(1-J)$$

képlet szolgáltatja. Most úgy akarjuk továbbítani a ξ sorozatot, az $\eta \to \eta'$ távközlési csatornán keresztül, hogy olyan ξ' sorozatot kapjunk, amely tetszőleges pontossággal tájékoztat mind a ξ_0, mind a ξ sorozat „állományáról"; ezt a követelést az

$$\bar{I}(\xi', \xi_0)=\bar{I}(\xi, \xi_0)$$

alakban formulázzuk meg. Bebizonyítható, hogy ennek szükséges feltétele az, hogy a ξ sorozatot *pontosan* továbbítsuk, azaz $C=1$ csatornakapacitású csatornát használjunk, ámbár a nekünk szükséges hasznos (egy betűre eső) információ csupán $I<1$-gyel egyenlő. Ennek a megjegyzésnek, úgy látom, van bizonyos jelentősége az alkalmazások szempontjából is, például amikor megérthetetlen vagy már eltorzult beszéd távbeszélő vonalakon át való továbbítása pontosságára vonatkozó követelményeket kell megállapitani.

Még folytathatnánk a hasonló megjegyzéseket. Az információelmélet ma még fejlődése kezdeti stádiumában van. Nagyon valószínű, hogy további fejlődése során azt a jelenleg uralkodó kizárólagos törekvést, hogy minden problémát információmennyiség kiszámítására vezessünk vissza, fel fogja váltani az a célkitűzés, hogy a különböző információtipusokhoz a jelenleginél teljesebb matematikai jellemzést találjunk, és ne ignoráljuk teljesen kvalitatív sajátosságaikat.

Záradékul még egy gondolatot szeretnék igen határozottan kimondani: Amidőn az információelmélet lehetséges alkalmazásait vizsgáljuk, ne tegyünk olyan a priori korlátozásokat, amelyek nem a dolog lényegéből következnek. Az a nálunk egy idő óta megfigyelhető törekvés, hogy az információelméletet olyan segédtudományként kezeljük, amely csak a technika pontosan meghatározott területein vehető igénybe, igen siralmas következményeket vonhat maga után. Már az információelmélet fogalmainak a nemrég megjelent [19] cikkben foglalt optikai alkalmazásai sem jelenhettek volna meg a maguk idejében, ha az információelmélet sokkal szélesebb körű alkalmazásairól szóló elgondolások nem érnek el hozzánk a külföldi irodalom utján. Megjegyzem, hogy — véleményem szerint — az információelmélet fogalmainak a valóságos „memóriaberendezésekre" való alkalmazása — amikor ti. az idegrendszer működését és az átöröklési jelenségeket tanulmányozzuk — teljesen szolid alapon áll és bizonyára igen lényeges lesz a tudomány említett fejezeteinek fejlődése szempontjából.

Fordította: *Medgyessy Pál*
A MTA Matematikai Kutató Intézete munkatársa

IRODALOM

[1] А. Н. Колмогоров: Статистическая теория колебаний с непрерывным спектром, Материалы общего собрания АН СССР, 1947 г.

[2] А. Н. Колмогоров: Статистическая теория колебаний с непрерывным спектром. Юбилейный сборник А. Н. СССР, Т. I. 1947, 242—254.

[3] А. Н. Колмогоров: О некоторых асимптотических характеристиках вполне ограниченных метрических пространств. Д. А. Н. СССР 108, Но. 3. 1956, 385—388.

[4] А. Г. Витушкин: К тринадцатой проблеме Гильберта. ДАН СССР, 94, 4, 1954, 701—704.

[5] А. Я. Хинчин: Понятие энтропии в теории вероятностей. УМН, VIII. вып. 3 (55). 1953, 3—20.

[6] А. Я. Хинчин: Об основных теоремах теории информации. УМН, XI вып. 1 (67), 1956, 17—75.

[7] „Теория передачи электрических сигналов при наличии помех." Сборник переводов. М., 1953.

[8] А. А. Харкевич: Очерик общей теории связы. Гос. Изд.—во Техн.—Теор. Литературы. М., 1955.

[9] Goldman, Stanford: Information Theory, New York 1953.

[10] C. E. Shannon and W. Weaver: The matematical theory of communication, *Univ. of Illinois Press*, 1949, 3—89. (orosz fordítását l.: [7]-ben.)

[11] Г. Крамер: Математические методы статистики. М., 1948.

[12] В. А. Котельников: Материалы к I-му Всесоюзному съезду по вопросам реконструкции связы. 1933.

[13] А. Н. Колмогоров, А. Н. Яглом, И. М. Гельфанд: Количество информации и энтропия дла непрерывных распределений. Доклад на Третьем всесоюзном математическом съезде. 1956.

[14] А. М. Обухов: Корреляция векторов. Ученные залиски МГУ, вып. 45. 1940. 73—92

[15] Р. М. Розенблат: Труды Третьего всесоюзного математического съезда. Т. II. 1956, 132—133.

[16] М. С. Пинскер: Труды Третьего всесоюзного математического съезда. Т. I. 125,

[17] А. Н. Колмогоров: Бюллетень МГУ, т. 2, вып. 6. 1941.

[18] М. С. Пинскер: ДАН СССР, 98, 1954, 213—216.

[19] С. Г. Раутиап: О мере разрешающей способности оптического прибора ДАН СССР, 109, (1956), 743—745.