ELOSZLÁSOK ELTÉRÉSÉNEK INFORMÁCIÓ-TÍPUSÚ MÉRTÉKSZÁMAI, I.*

Írta: CSISZÁR IMRE

0. §. Bevezetés

Az információelméletben és a matemetikai statisztikában egyaránt fontos szerepet játszik a valószínűségeloszlások egymástól való eltérésének S. KULLBACK által bevezetett információelméleti mértékszáma, a „megkülönböztetési információ" vagy *I*-divergencia (l. [20], [21]); ugyanezt a mennyiséget sokszor információnyereségnek, relatív információnak, vagy általánosított entrópiának is nevezik. Az *I*-divergencia általános definíciójával, főbb tulajdonságaival, valamint különböző interpretációival számos szerző foglalkozott, így A. PEREZ [24], [25], M. SZ. PINSZKER [27], G. KALLIANPUR [17], VINCZE I. [37] stb. A *Shannon*-féle információmennyiségen alapuló *I*-divergencia általánosításaként RÉNYI A. [29], [30] bevezette az α-rendű *I*-divergencia (információnyereség) fogalmát is.

Az irányított divergencia jellegű *I*-divergencia helyett gyakran célszerűbb a belőle származtatható szimmetrikus mennyiség, az ún. *J*-divergencia használata; ez utóbbi mennyiség — az információelmélettől függetlenül — már JEFFREYS [16] könyvében is szerepel. A *J*-divergencia segítségével bizonyította be HÁJEK [13], hogy két *Gauss*-mérték mindig vagy ekvivalens vagy ortogonális egymásra.

Az *I*-divergencia alkalmazásai közül disszertációm témájával kapcsolatban az elégséges statisztika fogalmának KULLBACK és LEIBLER [20] által adott információelméleti jellemzését, valamint a határeloszlástételek elméletében való alkalmazásokat kell megemlítenem. Az a gondolat, hogy információelméleti fogalmak felhasználhatók határeloszlástételek bizonyítására, JU. V. LINNIKtől származik [22].

Bár a LINNIK által tárgyalt esetben (centrális határeloszlástétel) az információelméleti módszer a szokásos módszereknél lényegesen bonyolultabban vezet csak célhoz, pl. *Markov*-láncok ergodicitásának tárgyalására a módszer egyszerűen és szemléletes módon alkalmazható (erre RÉNYI Alfréd mutatott rá, l. [29], [30]); a végtelen *Markov*-láncok ergodicitására vonatkozó tételnek éppen az „információelméleti" bizonyítása a legegyszerűbb (D. G. KENDALL [18]).

Az *I*-divergencia előnyös tulajdonságainak jelentős része az $f(u)=u\log u$ függvény konvexitásának következménye. (Pl. a határeloszlástételek bizonyítására való alkalmazhatóságra vonatkozólag l. RÉNYI és KENDALL említett dolgozatait.) Ennek alapján disszertációmban bevezetem (1. fejezet) a valószínűségeloszlások egymástól való eltérésének egy általánosított mértékszámát, az *f*-eltérést, ahol $f(u)$ tetszőleges konvex függvény lehet.

Ebből a mértékszámból kiindulva természetes módon beszélhetünk az eloszlások *f*-környezeteiről és — speciális esetként — információs környezeteiről. Így az alapul

* A dolgozat a szerző kandidátusi disszertációja, amelyet 1966. november 23-án védett meg. Jelen közlemény a dolgozat 1. és 2. §-át, s a teljes irodalomjegyzéket tartalmazza.

vett mérhető téren értelmezett valószínűségeloszlások halmaza egy ún. *Fréchet*-féle (V)-térré válik (vö. [34]), értelmezhető tehát az eloszlások f-konvergenciájának és (speciálisan) információ-konvergenciájának a fogalma. Az f-konvergencia és az egyenletes konvergencia viszonyával, valamint azzal a kérdéssel, hogy az f-környezetek milyen esetben teszik az eloszlások halmazát topologikus térré, disszertációm 2. fejezetében foglalkozom. Az eredmények között szerepel egy PINSZKERTŐL [27] származó, az I-divergencia és az eloszlások variációs távolsága közötti egyenlőtlenség általános megfelelője. A 3. fejezet az f-eltéréseknek közvetett megfigyelések esetén való csökkenését, valamint az elégségesség, ill. ε-elégségesség f-eltérésekkel megadható feltételeit tárgyalja; itt az elégséges statisztika fogalmának általánosításaként bevezetem az elégséges csatorna fogalmát is. A 4. fejezet a kompakt csoportokon értelmezett valószínűségi mértékek konvolúcióhatványainak konvergenciájára vonatkozó tételnek egy, az f-eltérés felhasználásán alapuló egyszerű bizonyítását tartalmazza, s rámutat az f-eltérések néhány más alkalmazási lehetőségére is.

Disszertációm alapját a [4], [5], [6], [7] dolgozataimban publikált eredményeim képezik, de ezek jelentős részét az azóta talált — még nem publikált — általánosabb, ill. élesebb formában tárgyalom.

1. §. Az I-divergencia és az f-eltérés

Legyen $(X, \mathscr{X})$ tetszőleges mérhető tér, azaz X tetszőleges halmaz és $\mathscr{X}$ az X részhalmazainak valamely σ-algebrája. Az $(X, \mathscr{X})$-en értelmezett μ mértéket valószínűségi mértéknek, vagy más szóval valószínűségeloszlásnak nevezzük, ha $\mu(X)=1$.

Legyen μ_1 és μ_2 két valószínűségi mérték $(X, \mathscr{X})$-en és $\mathfrak{A}=(A_1, A_2, \dots, A_n)$ az X egy véges mérhető felosztása (azaz $A_i \in \mathscr{X}$, $i=1, \dots, n$; $A_i \cap A_j = \emptyset$, $i \neq j$; $\bigcup_{i=1}^{n} A_i = X$). A μ_1 és μ_2 eloszlás I-divergenciája az $\mathfrak{A}$ felosztásra, vagy másképpen az $\mathfrak{A}$ által generált $[\mathfrak{A}] \subset \mathscr{X}$ algebrára vonatkozólag definíció szerint

$$I_{\mathfrak{A}}(\mu_1 \| \mu_2) = I_{[\mathfrak{A}]}(\mu_1 \| \mu_2) = \sum_{k=1}^{n} \mu_1(A_k) \log \frac{\mu_1(A_k)}{\mu_2(A_k)}; \tag{1.1}$$

ahol megállapodás szerint $0 \log 0 = 0 \log \frac{0}{0} = 0$ és $0 \log \frac{a}{0} = +\infty$, ha $a>0$. A logaritmus alapjára vonatkozólag megjegyzendő, hogy az információelméletben a 2 alapú logaritmus használata a legmegszokottabb, de gyakori az e alapú logaritmus alkalmazása is; minthogy disszertációmban a 2 alapú logaritmus alkalmazása semmiféle előnnyel sem járna, egyszerűség kedvéért mindig a természetes logaritmust használom.

Ismeretes, hogy ha $\mathfrak{A}_2$ az $\mathfrak{A}_1$-nél finomabb véges mérhető felosztás, azaz, ha $[\mathfrak{A}_1] \subset [\mathfrak{A}_2]$, akkor

$$I_{\mathfrak{A}_1}(\mu_1 \| \mu_2) \leqq I_{\mathfrak{A}_2}(\mu_1 \| \mu_2). \tag{1.2}$$

Az (1.2) egyenlőtlenségnek, valamint az I-divergencia következőkben említendő (jól ismert) tulajdonságainak bizonyítását az 1. § további részében általánosabb formában, az f-eltérésekre megfogalmazva fogom megadni.

Az (1. 2) monotonitási tulajdonság alapján természetszerű a tetszőleges (nem szükségképpen véges) $\mathscr{X}'\subset\mathscr{X}$ algebrára vonatkozó I-divergencia következő definíciója:

$$I_{\mathscr{X}'}(\mu_1\|\mu_2) = \sup_{[\mathfrak{A}]\subset\mathscr{X}'} I_{\mathfrak{A}}(\mu_1\|\mu_2). \tag{1. 3}$$

Itt a felső határ az összes $[\mathfrak{A}]\subset\mathscr{X}'$ véges algebrákra vonatkozik, más szóval az összes olyan $\mathfrak{A}$ véges mérhető felosztásokra, melyeknek minden eleme $\mathscr{X}'$-höz tartozik.

(1. 3)-ból azonnal következik, hogy az $\mathscr{X}$ bármely két $\mathscr{X}'\subset\mathscr{X}''$ részalgebrája esetén is

$$I_{\mathscr{X}'}(\mu_1\|\mu_2) \leqq I_{\mathscr{X}''}(\mu_1\|\mu_2). \tag{1. 2'}$$

Az egész $\mathscr{X}$ σ-algebrára vonatkozó I-divergencia definíciója (1. 3) speciális eseteként adódik:

$$I(\mu_1\|\mu_2) = I_{\mathscr{X}}(\mu_1\|\mu_2) = \sup_{\mathfrak{A}} \sum_{k=1}^{n} \mu_1(A_k)\log\frac{\mu_1(A_k)}{\mu_2(A_k)}, \tag{1. 4}$$

ahol a szuprémum az X tér összes $\mathfrak{A}$ véges mérhető felosztásaira vonatkozik. Ehelyett elég volna olyan felosztásokra szorítkozni, melyeknek elemei valamely az X σ-algebrát generáló $\mathscr{X}'\subset\mathscr{X}$ algebrához tartoznak, ugyanis ismeretes, hogy bármely $\mathscr{X}'\subset\mathscr{X}$ algebrára vonatkozó I-divergencia megegyezik az $\mathscr{X}'$ által generált σ-algebrára vonatkozó I-divergenciával.

Az I-divergencia mindig nemnegatív és az $I_{\mathscr{X}'}(\mu_1\|\mu_2)=0$ egyenlőség akkor és csak akkor áll fenn, ha az $\mathscr{X}'$ algebrán $\mu_1=\mu_2$. Fontos továbbá az I-divergencia integrálelőállítására vonatkozó tétel (vö. [17], [24], [27]):

$$I(\mu_1\|\mu_2) = \begin{cases} \displaystyle\int \log\frac{\mu_1(dx)}{\mu_2(dx)}\,\mu_1(dx), & \text{ha} \quad \mu_1 \ll \mu_2 \\ +\infty, & \text{ha} \quad \mu_1 \not\ll \mu_2 \end{cases} \tag{1. 5}$$

vagy ekvivalens módon,

$$I(\mu_1\|\mu_2) = \int p_1(x)\log\frac{p_1(x)}{p_2(x)}\,\lambda(dx), \tag{1. 5'}$$

ahol λ tetszőleges σ-véges mérték, melyre vonatkozólag mind μ_1, mind μ_2 abszolút folytonos és $p_i(x)=\dfrac{\mu_i(dx)}{\lambda(dx)}$ $(i=1,2)$. Sokszor közvetlenül ezt a formát tekintik az I-divergencia definíciójának, mint pl. S. KULLBACK [21] könyvében.

Jegyezzük meg, hogy az I-divergencia bevezetésekor KULLBACK az információmennyiség *Shannon*-féle mértékszámából indul ki (SHANNON [33]), sőt ezen túlmenőleg, az I-divergenciát a *Shannon*-féle entrópia természetes általánosításának tekinti. A *Shannon*-féle entrópia ill. differenciális entrópia $H=-\sum p_k \log p_k$, ill. $H=-\int p(x)\log p(x)\,dx$ képlete azonban csak akkor tekinthető (előjeltől eltekintve) az (1.5) formula speciális esetének, ha az utóbbiban μ_2 gyanánt tetszőleges σ-véges mértéket is megengedünk (konkrét esetben azt a mértéket, melynél minden pont mértéke 1, ill. a *Lebesgue*-mértéket). Az ilyen — tetszőleges σ-véges μ_2-vel adódó — általánosított entrópia is alkalmazható az információelméletben, vö. PEREZ [24], [25], bár erre az I-divergencia előnyös tulajdonságainak egy része (pl. a nemnegativitás) már nem érvényes. Számos szerző azonban (pl. PINSZKER [27]) az általánosított entrópia elnevezéssel a disszertációmban I-divergenciának nevezett mennyiséget jelöli, tehát felteszi, hogy nemcsak μ_1, hanem μ_2 is valószínűségi mérték; így azonban a terminológia kissé következetlen, s ezért helyesebbnek tartom az I-divergencia elnevezést.

Az $I(\mu_1\|\mu_2)$ irányított divergenciának többféle interpretációja ismeretes. S. KULLBACK megkülönböztetési információként értelmezte (information for discrimination): ha μ_1 és μ_2 két hipotézis az $(X, \mathscr{X})$ téren érvényes valószínűség-eloszlásra vonatkozólag, és a μ_1 hipotézis a helyes, akkor X egy véletlenszerűen választott x elemének megfigyelése átlagosan $I(\mu_1\|\mu_2)$ információt szolgáltat a μ_1 hipotézis javára, a μ_2 ellenében. Ezt az interpretációt alátámasztja az a tény, hogy ha a μ_1 hipotézis és a μ_2 ellenhipotézis között n független megfigyelés alapján akarunk dönteni, az elsőfajú hiba valószínűségét tetszőlegesen rögzítve, és a legjobb próbát használva, a másodfajú hiba valószínűsége $\beta_n = e^{-nI(\mu_1\|\mu_2)+o(n)}$ (vö. [21], [24], [32]). Az I-divergencia tehát valóban a két eloszlás statisztikai megkülönböztethetőségét méri. Az (1. 2') egyenlőtlenség úgy értelmezhető, hogy egy durvább kísérlet nem adhat több információt a két eloszlás megkülönböztetésére, mint egy finomabb.

Felfogható az $I(\mu_1\|\mu_2)$ mennyiség egy olyan kísérlet által szolgáltatott információnyereség mértékszámaként is, melynek alapján a μ_2 a priori eloszlást a μ_1 a posteriori eloszlással helyettesíthetjük (vö. RÉNYI [29]). Egy harmadik lehetséges interpretációt (az $X = R^1$ esetre) VINCZE István adott [37], eszerint $I(\mu_1\|\mu_2)$ a μ_1 eloszláshoz tartozó információ mértékszáma, ha érdeklődésünk eloszlása a μ_2 mértéknek felel meg.

Említést érdemel végül, hogy az információelméletben központi szerepet játszó információ-mértékszám, a két valószínűségi változó $I(\xi, \eta)$ kölcsönös információja, az I-divergencia speciális esetének tekinthető. Valóban, ha a ξ, ill. η valószínűségi változó eloszlását μ_ξ-vel, ill. μ_η-val, együttes eloszlásukat pedig $\mu_{\xi\eta}$-val jelöljük, (ahol μ_ξ, ill. μ_η valamely $(X, \mathscr{X})$, ill. $(Y, \mathscr{Y})$ mérhető téren, $\mu_{\xi\eta}$ pedig az $(X\times Y, \mathscr{X}\times\mathscr{Y})$ szorzattéren értelmezett valószínűségi mérték), akkor, mint jól ismert (GELFAND—JAGLOM [12]; l. még [24], [10])

$$(1.5\text{a})\qquad I(\xi, \eta) = \begin{cases} \displaystyle\int \log \frac{\mu_{\xi\eta}(dx, dy)}{(\mu_\xi\times\mu_\eta)(dx, dy)}\, \mu_{\xi\eta}(dx, dy) & \text{ha}\quad \mu_{\xi\eta} \ll \mu_\xi\times\mu_\eta \\ +\infty & \text{ha}\quad \mu_{\xi\eta} \not\ll \mu_\xi\times\mu_\eta \end{cases}$$

azaz

$$I(\xi, \eta) = I(\mu_{\xi\eta}\|\mu_\xi\times\mu_\eta).$$

A kölcsönös információ ezenkívül — bizonyos regularitási feltételek teljesülése esetén — még az $I(\xi, \eta) = MI(\mu_{\xi|\eta}\|\mu_\xi) = MI(\mu_{\eta|\xi}\|\mu_\eta)$ alakban is kifejezhető az I-divergencia segítségével, vö. (3. 66).

Az I-divergencia valamennyi interpretációja megegyezik abban, hogy ez a mennyiség bizonyos értelemben a valószínűségeloszlások egymástól való eltérésének mértékszáma. A továbbiakban az I-divergenciának ezt az utóbbi tulajdonságát tartjuk csak szem előtt, s minthogy ez lényegében csak az $f(u) = u \log u$ függvény konvexitásán múlik, vizsgálni fogjuk azokat az analóg mennyiségeket is, melyek az $u \log u$ függvénynek tetszőleges konvex függvénnyel való helyettesítésével adódnak.

Bocsássuk előre a következő jól ismert lemmát (JENSEN-*egyenlőtlenség*):

1. 1. LEMMA. *Legyen ξ valamely $(\Omega, \mathscr{F}, P)$ valószínűségi mezőn értelmezett, véges várható értékű valószínűségi változó, melynek értékkészlete valamely $[a_1, a_2]$ (véges vagy végtelen) intervallumba esik és $f(u)$ az $[a_1, a_2]$ intervallumban értelmezett alulról konvex függvény, amely az intervallum belsejében folytonos (az a_1 és a_2 végpontokban $f(u)$ értéke lehet $+\infty$ is). Ekkor*

$$(1.6) \qquad Mf(\xi) \geqq f(M\xi).$$

Ha $f(u)$ az $u_0 = M\xi$ pontban szigorúan konvex (azaz, ha $f(u)$ az u_0 pont egyetlen környezetében sem lineáris), akkor (1. 6)-ban az egyenlőség csak $P\{\xi = M\xi\} = 1$ esetén áll fenn.

Bizonyítás. Legyen $M\xi = u_0$; feltehetjük, hogy $a_1 < u_0 < a_2$, mert ellenkező esetben biztosan $P\{\xi = M\xi\} = 1$. Az $f(u)$ függvény konvex volta miatt

$$(1.7) \qquad f(u) \geqq f(u_0) + b(u - u_0) \qquad (a_1 \leqq u \leqq a_2),$$

ahol b az $f(u)$ függvény $u_0 = M\xi$ pontbeli jobb és bal oldali deriváltjának számtani közepe. (Tehát ha $f(u)$ differenciálható az u_0 pontban, akkor $b = f'(u_0)$.) Az (1. 7)-et $u = \xi$-re alkalmazva és várható értéket véve (1. 6) adódik; ha $f(u)$ az $u_0 = M\xi$ pontban szigorúan konvex, akkor (1. 7)-ben $u \neq u_0$ esetén határozott egyenlőtlenség áll, amivel a lemma második állítását is igazoltuk.

MEGJEGYZÉS. Természetesen az 1. 1. lemma bármely pozitív valószínűségű feltétel melletti feltételes várható értékekre is alkalmazható, továbbá a fentiekhez teljesen hasonlóan igazolható ($u_0 = M(\xi|\mathscr{F}_1)$ választással), hogy ha $\mathscr{F}_1$ az $\mathscr{F}$ tetszőleges rész-σ-algebrája, akkor 1 valószínűséggel

$$(1.6') \qquad M(f(\xi)|\mathscr{F}_1) \geqq f(M(\xi|\mathscr{F}_1)),$$

és ha $f(u)$ szigorúan konvex, az egyenlőség akkor és csak akkor teljesül 1 valószínűséggel, ha $P\{\xi = M(\xi|\mathscr{F}_1)\} = 1$ (vagyis ha ξ majdnem mérhető $\mathscr{F}_1$-re vonatkozólag).

A továbbiakban jelöljön $f(u)$ egy, a $(0, +\infty)$ intervallumban értelmezett folytonos, alulról konvex függvényt. Legyen megállapodás szerint

$$f(0) = \lim_{u \to +0} f(u)$$

$$(1.8) \qquad 0 \cdot f\left(\frac{0}{0}\right) = 0$$

$$0 \cdot f\left(\frac{a}{0}\right) = \lim_{\varepsilon \to +0} \varepsilon f\left(\frac{a}{\varepsilon}\right) = a \lim_{u \to +\infty} \frac{f(u)}{u} \qquad (0 < a < +\infty)$$

(az $f(u)$ függvény konvexitása miatt a határértékek biztosan léteznek; értékük természetesen lehet $+\infty$ is, de $-\infty$ nem).

Az $(X, \mathscr{X})$ mérhető téren értelmezett μ_1 és μ_2 valószínűségi mértékeknek az X tér valamely $\mathfrak{A} = (A_1, \dots, A_n)$ véges mérhető felosztásához vagy más szóval az $[\mathfrak{A}] \subset \mathscr{X}$ véges algebrához tartozó f-eltérését a következőképpen definiáljuk:

$$(1.9) \qquad \mathscr{I}_{f;\mathfrak{A}}(\mu_1, \mu_2) = \sum_{k=1}^{n} \mu_2(A_k) f\left(\frac{\mu_1(A_k)}{\mu_2(A_k)}\right).$$

$f(u)=u \log u$ esetén (1.9) éppen az (1. 1) formulát adja. (1. 2)-höz hasonlóan érvényes a következő összefüggés:

$$\mathscr{I}_{f;\mathfrak{A}_1}(\mu_1, \mu_2) \leqq \mathscr{I}_{f;\mathfrak{A}_2}(\mu_1, \mu_2) \qquad ([\mathfrak{A}_1]\subset[\mathfrak{A}_2]\subset\mathscr{X}). \tag{1. 10}$$

Valóban, ha $\mathfrak{A}_1=(A_1^1, \ldots, A_n^1)$, $\mathfrak{A}_2=(A_1^2, \ldots, A_m^2)$,

$$A_k^1 = \bigcup_{i=i_k}^{i_{k+1}-1} A_i^2 \qquad (k=1, \ldots, n;\ 1=i_1<i_2<\ldots<i_{n+1}=m+1),$$

akkor a tetszőleges $a_1, \ldots, a_l$; $b_1, \ldots, b_l$ nemnegatív számokra érvényes

$$\sum_{k=1}^{l} b_k f\left(\frac{a_k}{b_k}\right) \geqq \left(\sum_{k=1}^{l} b_k\right) f\left(\frac{\sum_{k=1}^{l} a_k}{\sum_{k=1}^{l} b_k}\right) \tag{1. 11}$$

egyenlőtlenség értelmében

$$\sum_{i=i_k}^{i_{k+1}-1} \mu_2(A_i^2) f\left(\frac{\mu_1(A_i^2)}{\mu_2(A_i^2)}\right) \geqq \mu_2(A_k^1) f\left(\frac{\mu_1(A_k^1)}{\mu_2(A_k^1)}\right) \qquad (k=1, \ldots, n), \tag{1. 12}$$

amiből (1. 10) azonnal következik. Maga az (1. 11) egyenlőtlenség pozitív a_k és b_k számok esetén az 1. 1 lemma speciális esete, és az (1. 8) megállapodásokból könnyen következik, hogy az egyenlőtlenség akkor is érvényben marad, ha az a_k és b_k számok közül egy vagy több 0-val egyenlő. Általánosabban érvényes a következő:

1. 2. LEMMA. *Ha $\alpha(x)$ és $\beta(x)$ két nemnegatív integrálható függvény valamely $(X, \mathscr{X}, \lambda)$ mértéktéren, ahol λ tetszőleges σ-véges mérték, és $E\in\mathscr{X}$, $\lambda(E)>0$, akkor*

$$\int_E \beta(x) f\left(\frac{\alpha(x)}{\beta(x)}\right) \lambda(dx) \geqq \int_E \beta(x)\lambda(dx) f\left(\frac{\int_E \alpha(x)\lambda(dx)}{\int_E \beta(x)\lambda(dx)}\right). \tag{1. 13}$$

Az egyenlőtlenség bal oldalán álló integrál biztosan értelmezve van. Ha $\int_E \beta(x)\lambda(dx)>0$ és az $f(u)$ konvex függvény az $u_0=\int_E \alpha(x)\lambda(dx)\Big/\int_E \beta(x)\lambda(dx)$ pontban szigorúan konvex, az egyenlőtlenség akkor és csak akkor áll fenn, ha az E halmazon[1] *$\alpha(x)=u_0\beta(x)$ $[\lambda]$.*

Bizonyítás. Az $f(u)$ függvény konvex volta miatt található olyan A és B véges valós szám, hogy $f(u)\geqq A+Bu$ $(0\leqq u<+\infty)$. Ebből következik, hogy az (1. 13) bal oldalán álló integrál negatív része $\geqq A\int \beta(x)\lambda(dx)+B\int \alpha(x)\lambda(dx) > -\infty$, s ezért az integrál értelmezve van. Az (1. 13) egyenlőtlenség $\beta(x)>0$ esetén közvetlenül az 1.1. lemmából adódik (az $(E,\mathscr{X}_E,P_E)$ valószínűségi mezőn értelmezett $\xi(x)=\dfrac{\alpha(x)}{\beta(x)}$

[1] Itt és a továbbiakban azt, hogy valamilyen mérték szerint majdnem mindenütt, a szokott módon az illető mérték szögletes zárójelbe tételével jelöljük.

valószínűségi változóra alkalmazva, ahol $P_E(B)=\dfrac{\int_B \beta(x)\lambda(dx)}{\int_E \beta(x)\lambda(dx)}$, ha $B\in\mathscr{X}_E$, azaz $B\in\mathscr{X}$, $B\subset E$). A bizonyítás az általános esetben is az 1. 1. lemmáéhoz teljesen hasonló, csak azt kell megjegyeznünk, hogy az (1. 7) egyenlőtlenségben formálisan $u=\dfrac{\alpha(x)}{\beta(x)}$-et írva és $\beta(x)$-szel végigszorozva még $\beta(x)=0$ esetén is helyes egyenlőtlenséget kapunk, tekintettel az (1. 8) megállapodásokra (ha $\int_E \beta(x)\lambda(dx)=0$, akkor u_0 nincs értelmezve, de ekkor (1. 13) triviálisan teljesül).

Az (1. 10) monotonitási tulajdonság alapján, az I-divergencia esetéhez hasonlóan, természetes módon definiálhatjuk a μ_1 és μ_2 mérték tetszőleges $\mathscr{X}'\subset\mathscr{X}$ algebrához, illetőleg az egész $\mathscr{X}$-hez tartozó f-eltérését:

$$\tilde{\mathscr{I}}_{f;\mathscr{X}'}(\mu_1,\mu_2)=\sup_{\substack{[\mathfrak{A}]\subset\mathscr{X}' \\ \mathfrak{A}\text{ véges felosztás}}}\mathscr{I}_{f;\mathfrak{A}}(\mu_1,\mu_2), \tag{1. 14}$$

$$\tilde{\mathscr{I}}_f(\mu_1,\mu_2)=\tilde{\mathscr{I}}_{f;\mathscr{X}}(\mu_1,\mu_2)=\sup_{\mathfrak{A}}\sum_{k=1}^{n}\mu_2(A_k)f\left(\frac{\mu_1(A_k)}{\mu_2(A_k)}\right), \tag{1. 15}$$

ahol a felső határ az X tér összes $\mathfrak{A}$ véges mérhető felosztására vonatkozik. Célszerűbbnek látszik azonban az (1. 5), ill. (1. 5') integrálképlet analogonját választani kiindulásul:

1. 1. Definíció. *Az $(X,\mathscr{X})$ mérhető téren értelmezett μ_1 és μ_2 valószínűségi mértékek f-eltérésének ($f(u)$ tetszőleges konvex függvény) az*

$$\mathscr{I}_f(\mu_1,\mu_2)=\int p_2(x)f\left(\frac{p_1(x)}{p_2(x)}\right)\lambda(dx) \tag{1. 16}$$

mennyiséget nevezzük, ahol λ valamely, a μ_1, μ_2 mértékpárt domináló (azaz $\mu_1\ll\lambda$, $\mu_2\ll\lambda$) σ-véges mérték és $p_i(x)$ a μ_i mérték λ-ra vonatkozó sűrűségfüggvénye, vagyis Radon—Nikodym deriváltja [2]*:*

$$p_i(x)=\frac{\mu_i(dx)}{\lambda(dx)} \qquad (i=1,2).$$

Az 1. 2. lemma szerint az (1. 16) integrál mindig értelmezve van, és értéke nyilvánvalóan nem függ a λ domináló mérték választásától. Továbbá, ugyancsak az 1. 2. lemma szerint

$$\mathscr{I}_f(\mu_1,\mu_2)\geqq\int p_2(x)\lambda(dx)f\left(\frac{\int p_1(x)\lambda(dx)}{\int p_2(x)\lambda(dx)}\right)=f(1), \tag{1. 17}$$

[2] Az egyszerűség kedvéért itt és a továbbiakban mindig feltételezzük, hogy a szóban forgó *Radon—Nikodym* deriváltak úgy vannak rögzítve, hogy mindenütt (tehát nemcsak majdnem mindenütt) végesek és nemnegatívak. Az $\int$ jel az integrációs tartomány feltüntetése nélkül az egész térre vonatkozó integrált jelenti.

ahol az egyenlőség — feltéve, hogy $f(u)$ az $u_0=1$ pontban szigorúan konvex — csak $\mu_1=\mu_2$ esetén teljesül.

Természetesen két mérték f-eltérése $+\infty$ is lehet: (1. 8) értelmében biztosan ez a helyzet, ha $f(0)=+\infty$ és $\mu_2 \not\ll \mu_1$ vagy ha $\lim\limits_{u\to+\infty}\frac{f(u)}{u}=+\infty$ és $\mu_1 \not\ll \mu_2$. Ha azonban mind $f(0)$, mind $\lim\limits_{u\to+\infty}\frac{f(u)}{u}$ véges, akkor (1. 17) miatt $\mathscr{I}_f(\mu_1, \mu_2)$ minden esetben véges.

Az f-eltérés fenti definíciója alkalmazható az $\mathscr{X}$ σ-algebra bármely $\mathscr{X}_0$ rész-σ-algebrájára is, amely esetben $p_1(x)$ és $p_2(x)$ a vizsgált szűkebb σ-algebrára vonatkozó $\bar{p}_1(x)$ és $\bar{p}_2(x)$ *Radon—Nikodym* deriváltakkal helyettesítendő:

$$\mathscr{I}_{f;\mathscr{X}_0}(\mu_1, \mu_2) = \int \bar{p}_2(x) f\left(\frac{\bar{p}_1(x)}{\bar{p}_2(x)}\right)\lambda(dx). \tag{1. 16'}$$

Látható, hogy az $\mathscr{X}$ véges részalgebráira (melyek egyúttal rész-σ-algebrák is) így éppen az (1. 9) formulával definiált $\mathscr{I}_{f;\mathfrak{A}}(\mu_1, \mu_2)$ érték adódik. Mint látni fogjuk, az f-eltérés most adott definíciója az általános esetben is ekvivalens az előzőleg javasolttal, azaz $\mathscr{I}_f(\mu_1, \mu_2)=\tilde{\mathscr{I}}_f(\mu_1, \mu_2)$. Megjegyzem még, hogy az 1. 1. definíció tetszőleges σ-véges mértékekre is kiterjeszthető volna, feltéve, hogy az (1. 16) integrál értelmezve van (ami véges mértékek esetén automatikusan teljesül). A következőkben azonban mindig csak a valószínűségi mértékek esetére szorítkozom, bár a kapott eredmények egy része az általánosabb esetre is kiterjeszthető, vö. [6].

Az 1. 1. definíció az I-divergencia integrálelőállításának közvetlen analogonja; valóban, (1. 5') és (1. 16) alapján — az (1. 8) megállapodás figyelembevételével — nyilvánvaló, hogy

$$I(\mu_1\|\mu_2) = \mathscr{I}_{u\log u}(\mu_1, \mu_2) = \mathscr{I}_{-\log u}(\mu_2, \mu_1). \tag{1. 18}$$

Hasonlóképpen a RÉNYI Alfréd által ([29], [30]) bevezetett

$$I_\alpha(\mu_1\|\mu_2) = \frac{1}{\alpha-1}\log\int p_1^\alpha(x)p_2^{1-\alpha}(x)\lambda(dx) \qquad (0<\alpha<+\infty) \tag{1. 19}$$

α-adrendű I-divergencia (információnyereség) — melynek a közönséges (elsőrendű) I-divergencia az $\alpha\to 1$ határátmenettel adódó speciális esete — ugyancsak származtatható f-eltérésként:

$$I_\alpha(\mu_1\|\mu_2) = \begin{cases} \dfrac{1}{\alpha-1}\log|\mathscr{I}_{-u^\alpha}(\mu_1, \mu_2)| & (0<\alpha<1) \\ \dfrac{1}{\alpha-1}\log \mathscr{I}_\alpha(\mu_1, \mu_2) & (\alpha>1). \end{cases} \tag{1. 20}$$

Tehát az α-adrendű I-divergencia ($\alpha\neq 1$) az $f_\alpha(u)=u^\alpha \operatorname{sgn}(\alpha-1)$ konvex függvénynek megfelelő f-eltérés monoton függvénye. Jegyezzük meg, hogy míg $\alpha\geqq 1$ esetén $I_\alpha(\mu_1\|\mu_2)$ csak akkor lehet véges, ha $\mu_1\ll\mu_2$, a $0<\alpha<1$ esetben $I_\alpha(\mu_1\|\mu_2)$ mindig véges, kivéve, ha $\mu_1\perp\mu_2$ (tehát ha van olyan $E\in\mathscr{X}$, hogy $\mu_1(E)=\mu_2(X-E)=1$).

A mértékek egymástól való eltérésének legmegszokottabb mértékszáma, a

$$\varrho(\mu_1, \mu_2) = |\mu_1-\mu_2| = \int |p_1(x)-p_2(x)|\lambda(dx) \tag{1. 21}$$

variációs távolság is tekinthető f-eltérésnek. (Mint ismeretes, az $(X, \mathscr{X})$ mérhető téren értelmezett valószínűségi mértékek a variációs távolsággal teljes metrikus teret alkotnak.) Valóban, ha $f(u) = |u-1|$, akkor

$$(1.22) \qquad \mathscr{I}_f(\mu_1, \mu_2) = \mathscr{I}_{|u-1|}(\mu_1, \mu_2) = \int p_2(x) \left| \frac{p_1(x)}{p_2(x)} - 1 \right| \lambda(dx) = |\mu_1 - \mu_2|.$$

Ugyancsak az f-eltérések közé tartozik az eloszlások különbözőségének a matematikai statisztikában használt χ^2-mértékszáma is, melyet (a diszkrét esetben) a $\sum \frac{(p_i - q_i)^2}{q_i}$ képlet értelmez. Látható, hogy a χ^2-eltérés nem más, mint az $f(u) = (u-1)^2$ konvex függvénynek megfelelő f-eltérés. Egyébként (1. 20) alapján nyilvánvaló, hogy a χ^2-eltérés és a másodrendű I-divergencia között az $I_2(\mu_1 \| \mu_2) = \log(1 + \chi^2(\mu_1, \mu_2))$ függvénykapcsolat áll fenn.

Azt, hogy az 1. 1. definícióval meghatározott $\mathscr{I}_f(\mu_1, \mu_2)$ mennyiséget bármely konvex f esetén eltérés-mértékszámnak fogjuk fel, az indokolja, hogy (1. 17) értelmében $\mathscr{I}_f(\mu_1, \mu_2)$ a minimális értékét $\mu_1 = \mu_2$ esetén veszi fel (és csak ekkor, feltéve, hogy $f(u)$ az $u_0 = 1$ pontban szigorúan konvex), tehát a skála alkalmas eltolásával olyan mértékszámhoz jutunk, amely mindig nemnegatív és csak $\mu_1 = \mu_2$ esetén nulla. Ezzel szemben az f-eltérés általában nem szimmetrikus, azaz $\mathscr{I}_f(\mu_1, \mu_2) = \mathscr{I}_f(\mu_2, \mu_1)$ általában nem teljesül, s ugyancsak nem áll fenn általában a háromszög-egyenlőtlenség sem[3]. Ha $\mathscr{I}_f(\mu_1, \mu_2)$-ben μ_1 és μ_2 sorrendjét felcseréljük, egy másik, f^* konvex függvénynek megfelelő f-eltérést kapunk, nevezetesen (figyelembe véve az (1. 8) megállapodásokat is)

$$(1.23) \qquad \mathscr{I}_f(\mu_2, \mu_1) = \mathscr{I}_{f^*}(\mu_1, \mu_2),$$

ahol

$$(1.24) \qquad f^*(u) = uf\left(\frac{1}{u}\right).$$

Itt az (1. 24) egyenlőséggel definiált $f^*(u)$ függvény konvex volta az $f(u)$ konvexitásának közvetlen következménye, ui. az $f(u)$ konvex volta miatt bármely a és b pozitív számra és $0 < \alpha < 1$ számra

$$\frac{\alpha a}{\alpha a + (1-\alpha) b} f\left(\frac{1}{a}\right) + \frac{(1-\alpha) b}{\alpha a + (1-\alpha) b} f\left(\frac{1}{b}\right) \geqq f\left(\frac{1}{\alpha a + (1-\alpha) b}\right)$$

vagyis $\alpha f^*(a) + (1-\alpha) f^*(b) \geqq f^*(\alpha a + (1-\alpha) b)$.

Az $I(\mu_1 \| \mu_2)$ irányított divergencia helyett információelméleti különbözőség-mértékszámként szokták használni a $J(\mu_1, \mu_2) = I(\mu_1 \| \mu_2) + I(\mu_2 \| \mu_1)$ szimmetrikus kifejezést, az ún. J-divergenciát is, sőt ez utóbbi mennyiség bevezetése (az információelmélettől függetlenül) időben meg is előzte az I-divergenciát (l. JEFFREYS [16]). A J-divergencia, éppúgy, mint bármely $\mathscr{I}_f(\mu_1, \mu_2) + \mathscr{I}_f(\mu_2, \mu_1)$ alakú szimmetrikus mennyiség, szintén tekinthető f-eltérésnek, az $f + f^*$ konvex függvénnyel; azaz pl.

$$(1.25) \qquad J(\mu_1, \mu_2) = \mathscr{I}_{(u-1)\log u}(\mu_1, \mu_2).$$

[3] Ez azt jelenti, hogy az f-eltérés a topológiában megszokott értelemben nem „eltérés", hanem csak tágabb értelemben vett eltérés-mértékszám, éppúgy, mint az I-divergencia is.

Az $\mathscr{I}_f(\mu_1, \mu_2)$ és $\mathscr{I}_f(\mu_2, \mu_1)$ más szimmetrikus függvényeit (pl. szorzatát) is lehetne a két valószínűségeloszlás különbözősége mértékszámának tekinteni. A $J_\alpha(\mu_1, \mu_2) = I_\alpha(\mu_1 \| \mu_2) + I_\alpha(\mu_2 \| \mu_1)$ α-rendű J-divergencia pl. az $\mathscr{I}_{f_\alpha}(\mu_1, \mu_2)\mathscr{I}_{f_\alpha}(\mu_2, \mu_1)$ szorzat monoton függvényeként áll elő, ahol $f_\alpha(u) = u^\alpha \operatorname{sgn}(\alpha - 1)$. Az ilyen típusú mennyiségekkel azonban disszertációmban nem kívánok részletesebben foglalkozni.

Megemlítem még, hogy ha ξ és η két tetszőleges valószínűségi változó, a $\mu_{\xi\eta}$ együttes eloszlás és a marginális eloszlások $\mu_\xi \times \mu_\eta$ szorzatának f-eltérése a ξ és η közötti összefüggés mértékszámának tekinthető. Az $f(u) = u \log u$ esetben így éppen az $I(\xi, \eta)$ kölcsönös információ adódik, $f(u) = (u-1)^2$ (vagy $f(u) = u^2 - 1$) esetén pedig a $\varphi^2(\xi; \eta)$ négyzetes kontingencia.

MEGJEGYZÉS. Az információelméletben mind az entrópia, mind az I-divergencia elemi információmennyiségek középértékeként jelenik meg. A legmegszokottabb a közönséges lineáris közepelés, de, mint RÉNYI A. [29] kimutatta, az exponenciális közepelés is fontos információmennyiségekre vezet (α-rendű entrópia, ill. I-divergencia). Más függvényekkel való közepelésnél az információ additív tulajdonsága már veszendőbe megy. Mégis, mint ACZÉL J. és DARÓCZY Z. [1] dolgozatukban rámutattak, valamely $\mathscr{P} = (p_1, \ldots, p_n)$ diszkrét eloszláshoz tartozó információmennyiség mérésére bizonyos értelemben az $I_k = -\log p_k$ elemi információmennyiségek minden olyan

$$I(\mathscr{P}) = g^{-1}\left(\sum_{k=1}^{n} p_k g(-\log p_k)\right) \tag{1.26}$$

középértéke megfelel, melyben a g folytonos és szigorúan monoton függvény olyan, hogy

$$f(u) = ug(-\log u) \tag{1.27}$$

szigorúan konvex[4]. Ha az (1.26)-nak megfelelő I-divergencia jellegű mennyiséget akarjuk képezni — tetszőleges, de egyszerűség kedvéért egymásra nézve abszolút folytonos eloszlásokra — a $-\log p_k$ elemi információ helyett az $l(x) = \log \frac{\mu_1(dx)}{\mu_2(dx)}$ likelihood-függvény (-1)-szeresét kell átlagolnunk és az eredmény (-1)-szeresét vennünk (vö. [29], 262. o.); ekkor az 1. 1. definíció és (1.27) figyelembevételével a következő adódik:

$$I_g(\mu_1 \| \mu_2) = -g^{-1}\left[\int g(-\log l(x))\mu_1(dx)\right] = \tag{1.28}$$

$$= -g^{-1}\left[\int f\left(\frac{\mu_1(dx)}{\mu_2(dx)}\right)\mu_2(dx)\right] = -g^{-1}(\mathscr{I}_f(\mu_1, \mu_2)).$$

A $g_\alpha(u) = e^{(1-\alpha)u} \operatorname{sgn}(\alpha - 1)$ esetben (1.28) éppen (1.20)-at adja.

Másfelől, ha $f(u)$ tetszőleges folytonos konvex függvény $[0, +\infty]$-ben, amely a 0 pontban szigorúan konvex és $f(0) = 0$, akkor $f(u)/u$ szigorúan monoton növekedő, $g(u) = e^u f(e^{-u})$ pedig szigorúan monoton fogyó folytonos függvény és $f(u)$ és $g(u)$ eleget tesz (1.27)-nek. Ilyenkor tehát bizonyos értelemben célszerűbbnek látszana

[4] Az ilyen általános típusú információmértékszám bevezetését az [1] dolgozat szerzői Hincsinnek és A. Adamnak tulajdonítják; Hincsin ill. Adam általuk idézett munkáiban azonban ilyen jellegű általánosított információmértékszámokról nem esik szó.

$\mathscr{I}_f(\mu_1, \mu_2)$ helyett az (1. 28) mennyiséggel foglalkozni. Valójában azonban — a következők szempontjából — az f-eltérés és annak bármely szigorúan monoton növekvő folytonos függvénye ekvivalens eltérésmértékszámnak tekinthető, ezért elég az egyszerűbben kezelhető f-eltérések vizsgálatára szorítkozni.

Az f-eltérés (1. 15) és (1. 16) definíciójának ekvivalenciája, vagyis az $\mathscr{I}_f(\mu_1, \mu_2) = \tilde{\mathscr{I}}_f(\mu_1, \mu_2)$ egyenlőség tetszőleges f konvex függvény esetére is lényegében ugyanúgy bizonyítható, mint az $f(u) = u \log u$ speciális esetben, csupán annak figyelembe vétele jelent némi nehézségtöbbletet, hogy az $f(u)$ konvex függvény 0, ill. $+\infty$-beli viselkedése többféle lehet. A következő bizonyítás, az említett okokból szükséges kisebb módosításokkal, DOBRUSINnak az $f(u) = u \log u$ esetre adott bizonyítása gondolatmenetét követi (vö. [10]).

Legyen

$$(1.29)\quad \begin{cases} A_0 = \{x\colon p_2(x) > p_1(x) = 0\}, & A_\infty = \{x\colon p_2(x) = 0\}, \\ A_\delta = \{x\colon 0 < p_1(x) < \delta p_2(x)\}, & A_M = \{x\colon p_1(x) \geqq M p_2(x) > 0\}, \\ A_k = \left\{x\colon h_k \leqq \dfrac{p_1(x)}{p_2(x)} < h_{k+1}\right\} & (k = 1, \dots, n), \end{cases}$$

ahol $\delta = h_1 < h_2 < \dots < h_{n+1} = M$ a $[\delta, M]$ intervallumnak olyan felosztása, hogy a $[h_k, h_{k+1}]$ intervallumok mindegyikében az $f(u)$ függvény ingadozása kisebb, mint ε. Ekkor $\mathfrak{A} = (A_0, A_\varepsilon, A_1, A_2, \dots, A_n, A_M, A_\infty)$ az X térnek egy véges mérhető felosztása.

(1. 29) valamint (1. 8) alapján nyilvánvaló, hogy

$$(1.30)\quad \begin{aligned} &\int_{A_0} p_2(x) f\left(\frac{p_1(x)}{p_2(x)}\right) \lambda(dx) = \mu_2(A_0) f\left(\frac{\mu_1(A_0)}{\mu_2(A_0)}\right); \\ &\int_{A_\infty} p_2(x) f\left(\frac{p_1(x)}{p_2(x)}\right) \lambda(dx) = \mu_2(A_\infty) f\left(\frac{\mu_1(A_\infty)}{\mu_2(A_\infty)}\right); \\ &\int_{A_k} p_2(x) f\left(\frac{p_1(x)}{p_2(x)}\right) \lambda(dx) \leqq \mu_2(A_k) \left[f\left(\frac{\mu_1(A_k)}{\mu_2(A_k)}\right) - \varepsilon\right] \qquad (k = 1, \dots, n). \end{aligned}$$

Jegyezzük meg, hogy $\delta \to 0$, ill. $M \to +\infty$ esetén $\mu_2(A_\delta) \to 0$ és $\mu_1(A_M) \to 0$ [5], és így, kihasználva, hogy az $f(u)$ függvény konvexitása miatt alkalmas K pozitív számra

$$(1.31)\qquad f(u) > -K, \quad uf\left(\frac{1}{u}\right) > -K \qquad (0 \leqq u \leqq 1),$$

az A_δ és A_M halmaz (1. 26) definíciója értelmében alkalmasan választott $\delta > 0$ és

[5] Ui. ha $\delta_j \downarrow 0$ ill. $M_j \uparrow +\infty$, akkor $\{A_{\delta_j}\}$ ill. $\{A_{M_j}\}$ üres metszettel bíró monoton csökkenő halmazsorozat, tehát $\mu_i(A_{\delta_j}) \to 0$, $\mu_i(A_{M_j}) \to 0$ $(i = 1, 2)$.

$M > 0$ esetén

$$\mu_2(A_\delta) f\left(\frac{\mu_1(A_\delta)}{\mu_2(A_\delta)}\right) > \mu_2(A_\delta)\cdot(-K) > -\varepsilon \tag{1.32}$$
$$\mu_2(A_M) f\left(\frac{\mu_1(A_M)}{\mu_2(A_M)}\right) > \mu_1(A_M)\cdot(-K) > -\varepsilon.$$

Az (1. 30) és (1. 32) összefüggések szerint

$$\int\limits_{X-A_\delta-A_M} p_2(x) f\left(\frac{p_1(x)}{p_2(x)}\right) \lambda(dx) \leqq \mathscr{I}_{f;\mathfrak{A}}(\mu_1, \mu_2) - 3\varepsilon. \tag{1.33}$$

Ha az (1. 33) egyenlőtlenség jobb oldalán $\mathscr{I}_{f;\mathfrak{A}}(\mu_1, \mu_2)$ helyett $\tilde{\mathscr{I}}_f(\mu_1, \mu_2)$-t írunk, az egyenlőtlenség még inkább igaz marad; sőt ekkor a bal oldalon az integrációs tartományt az egész X térnek is vehetjük, ui. ha δ_j ill. M_j monoton csökkenőleg, ill. növekedőleg tart 0-hoz, ill. $+\infty$-hez, akkor az $X - A_{\delta_j} - A_{M_j}$ halmazsorozat monoton növekedőleg tart X-hez. Így végül, minthogy a jobb oldalon szereplő $\varepsilon > 0$ tetszőleges volt, nyerjük, hogy

$$\mathscr{I}_f(\mu_1, \mu_2) \leqq \tilde{\mathscr{I}}_f(\mu_1, \mu_2). \tag{1.34}$$

Az ellenkező irányú egyenlőtlenség még egyszerűbben igazolható. Valóban, az 1. 2. lemma szerint bármely $\mathfrak{A} = (A_1, \dots, A_n)$ véges mérhető felosztásra

$$\int\limits_{A_k} p_2(x) f\left(\frac{p_1(x)}{p_2(x)}\right) \lambda(dx) \geqq \mu_2(A_k) f\left(\frac{\mu_1(A_k)}{\mu_2(A_k)}\right) \qquad (k = 1, \dots, n) \tag{1.35}$$

azaz összegezve

$$\mathscr{I}_f(\mu_1, \mu_2) \geqq \mathscr{I}_{f;\mathfrak{A}}(\mu_1, \mu_2). \tag{1.36}$$

Ezzel az $\mathscr{I}_f(\mu_1, \mu_2) = \tilde{\mathscr{I}}_f(\mu_1, \mu_2)$ egyenlőtlenséget bebizonyítottuk. Belátható továbbá, hogy az imént igazolt

$$\mathscr{I}_f(\mu_1, \mu_2) = \sup_{\mathfrak{A}} \sum_{k=1}^{n} \mu_2(A_k) f\left(\frac{\mu_1(A_k)}{\mu_2(A_k)}\right) \tag{1.37}$$

egyenlőség akkor is érvényben marad, ha a felső határt az X tér összes véges mérhető felosztásai helyett csak azokra a felbontásokra vesszük, melyeknek elemei egy rögzített, az $\mathscr{X}$-algebrát generáló $\mathscr{X}'$ algebrába tartozó halmazok, más szóval, hogy ilyenkor $\mathscr{I}_f(\mu_1, \mu_2) = \tilde{\mathscr{I}}_{f;\mathscr{X}'}(\mu_1, \mu_2)$. Valóban, ismét DOBRUSINnak az $f(u) = u \log u$ esetre vonatkozó bizonyítását követve, legyen $\mathfrak{A} = (A_1, \dots, A_n)$ X-nek egy rögzített véges (mérhető) felosztása; ekkor bármely $\delta > 0$ esetén megadhatók olyan $B_1, \dots, B_n$ diszjunkt halmazok, hogy $B_k \in \mathscr{X}'$ $(k = 1, \dots, n)$, továbbá[6]

$$\mu_1(A_k \circ B_k) \leqq n\delta, \quad \mu_2(A_k \circ B_k) \leqq n\delta \qquad (k = 1, \dots, n) \tag{1.38}$$

és

$$\mu_1(B_{n+1}) \leqq n\delta, \quad \mu_2(B_{n+1}) \leqq n\delta \qquad \left(B_{n+1} = X - \bigcup_{k=1}^{n} B_n\right) \tag{1.39}$$

[6] $A \circ B = (A - B) \cup (B - A)$.

(vö. [10], 54—55. o.). Az $f(u)$ függvény folytonossága miatt nyilvánvaló (figyelembe véve (1. 8)-at is), hogy bármely adott $\varepsilon>0$ és $M>0$ esetén $\delta>0$ megválasztható olyan kicsire, hogy

$$\mu_2(B_k)f\left(\frac{\mu_1(B_k)}{\mu_2(B_k)}\right)>\mu_2(A_k)f\left(\frac{\mu_1(A_k)}{\mu_2(A_k)}\right)-\varepsilon, \tag{1. 40}$$

valahányszor $\mu_2(A_k)f\left(\frac{\mu_1(A_k)}{\mu_2(A_k)}\right)<+\infty$ és

$$\mu_2(B_k)f\left(\frac{\mu_1(B_k)}{\mu_2(B_k)}\right)>M, \tag{1. 41}$$

ha

$$\mu_2(A_k)f\left(\frac{\mu_1(A_k)}{\mu_2(A_k)}\right)=+\infty;$$

végül — figyelembe véve (1. 31)-et —

$$\mu_2(B_{n+1})f\left(\frac{\mu_1(B_{n+1})}{\mu_2(B_{n+1})}\right)>-\varepsilon. \tag{1. 42}$$

Az (1. 40), (1. 41) és (1. 42) egyenlőtlenségek értelmében

$$\sum_{k=1}^{n+1}\mu_2(B_k)f\left(\frac{\mu_1(B_k)}{\mu_2(B_k)}\right)>\sum_{k=1}^{n}\mu_2(A_k)f\left(\frac{\mu_1(A_k)}{\mu_2(A_k)}\right)-(n+1)\varepsilon, \tag{1. 43}$$

ha a jobb oldal véges, ill. a bal oldal tetszőlegesen naggyá tehető, ha a jobb oldal végtelen. Ezzel a bizonyítást befejeztük.

Az (1. 10) monotonitási tulajdonságból nyilvánvalóan következik, hogy (1. 37)-ben a felső határ képzésénél az összes $\mathscr{X}'$-beli halmazokra való véges felosztások osztálya helyett szorítkozhatunk ennek bármely olyan részosztályára is, amely olyan tulajdonságú, hogy az előbbi osztály minden felosztásához található benne egy finomabb felosztás.

Eredményünket tétel formájában megfogalmazva:

1. 1. TÉTEL. *Legyen $\mathscr{X}'\subset\mathscr{X}$ tetszőleges olyan algebra, amely az $\mathscr{X}$ σ-algebrát generálja és jelölje R az X tér véges felosztásainak olyan összességét, hogy minden R-beli $\mathfrak{A}=(A_1,\ldots,A_n)$ felosztásra $A_k\in\mathscr{X}'$ $(k=1,\ldots,n)$, továbbá, hogy X-nek minden $\mathscr{X}'$-beli halmazokra való véges felosztásához található nála finomabb R-beli felosztás. Ekkor*

$$\mathscr{I}_f(\mu_1,\mu_2)=\int p_2(x)f\left(\frac{p_1(x)}{p_2(x)}\right)\lambda(dx)=\sup_{\mathfrak{A}\in R}\sum_{k=1}^{n}\mu_2(A_k)f\left(\frac{\mu_1(R_k)}{\mu_2(R_k)}\right). \tag{1. 44}$$

Ha az $\mathscr{X}$ σ-algebrát X részhalmazainak valamely megszámlálható rendszere generálja, akkor az R osztály választható egy alkalmas finomodó felosztássorozatnak és ilyenkor (1. 44)-ben sup helyett lim is írható. Ez a helyzet pl. akkor, ha X tetszőleges, a második megszámlálhatósági axiómának eleget tevő topologikus tér és $\mathscr{X}$ a *Borel*-halmazok σ-algebrája, azaz a nyílt halmazok összessége által generált σ-algebra.

2. §. Az f-eltérések által indukált topológiai struktúra vizsgálata

Minthogy az I-divergenciát és általánosabban az f-eltéréseket a valószínűségeloszlások egymástól való eltérésének mértékszámaként interpretáltuk, természetes módon felvetődik a kérdés, hogy ezek a mértékszámok milyen topológiai struktúrát indukálnak a valószínűségeloszlások terében.

Legyen $\mathcal{M}$ az $(X, \mathcal{X})$ mérhető téren értelmezett valószínűségeloszlások egy halmaza és $f(u)$ tetszőleges, $(0, +\infty)$-ben értelmezett folytonos konvex függvény.

2. 1. Definíció. *Valamely $\mu_0 \in \mathcal{M}$ eloszlás f-környezetein az*

$$U_f(\mu_0; \varepsilon) = \{\mu: \mathcal{I}_f(\mu, \mu_0) - f(1) < \varepsilon, \mu \in \mathcal{M}\} \tag{2. 1}$$

alakú eloszláshalmazokat értjük.

Ha speciálisan $f(u) = u \log u$ vagy $f(u) = u^\alpha \operatorname{sgn}(\alpha - 1)$, akkor (első, ill. α-rendű) információs környezetekről beszélünk, minthogy ezek a $\{\mu: I_\alpha(\mu \| \mu_0) < \varepsilon', \mu \in \mathcal{M}\}$ alakban is felírhatók.

A 2. 1. definícióval $\mathcal{M}$ ún. *Fréchet*-féle (V)-térré vált (vö. [34]), amely még azzal a tulajdonsággal is rendelkezik, hogy bármely $\mu_0 \in \mathcal{M}$ „pont" tetszőleges véges számú környezetének közös része maga is μ_0 környezete, továbbá, ha $f(u)$ az $u_0 = 1$ pontban szigorúan konvex, akkor (1. 17) értelmében bármely $\mu_0 \in \mathcal{M}$ összes f-környezetének egyetlen közös pontja μ_0. Ezen túlmenőleg az is belátható, hogy az f-környezetekre teljesül a *Hausdorff*-féle T_2 axióma is, azaz bármely két $\mu_1 \neq \mu_2$ eloszlásnak van egymást nem metsző f-környezete (l. a 2. 1. tételt).

Ismeretes, hogy (V)-terekben már sok topológiai jellegű fogalom értelmezhető, noha természetesen nem minden (V)-tér topologikus tér (látjuk majd, a valószínűségeloszlások halmaza sem válik minden esetben topologikus térré a környezetek 2. 1. definíciójával). Így minden (V)-térben beszélhetünk konvergenciáról; a 2. 1. definíciónak megfelelően egy $\mu_1, \mu_2, \ldots$ $(\mu_k \in \mathcal{M}, k = 1, 2, \ldots)$ eloszlássorozatról azt mondjuk, hogy f-konvergál a $\mu_0 \in \mathcal{M}$ valószínűségeloszláshoz, jelölve

$$\mu_n \xrightarrow{f} \mu_0 \qquad (n \to \infty) \tag{2. 2}$$

ha minden $\varepsilon > 0$-hoz megadható olyan $n_0 = n_0(\varepsilon)$, hogy $n \geqq n_0$ esetén $\mu_n \in U_f(\mu_0; \varepsilon)$, más szóval, ha

$$\lim_{n \to \infty} \mathcal{I}_f(\mu_n, \mu_0) = f(1). \tag{2. 3}$$

Speciálisan $f(u) = u \log u$, ill. $f(u) = u^\alpha \operatorname{sgn}(\alpha - 1)$ esetén (első-, ill. α-rendű) információkonvergenciáról fogunk beszélni. Nyilvánvaló ugyanis, hogy $f(u) = u^\alpha \operatorname{sgn}(\alpha - 1)$ esetén $\mathcal{I}_f(\mu_n, \mu_0) \to f(1)$ ekvivalens azzal, hogy $I_\alpha(\mu_n \| \mu_0) \to 0$.

MEGJEGYZÉSEK. *1.* A 2. 1. definíció a mértékek sorrendjét tekintve önkényes; ez azonban nem jelenti az általánosság korlátozását, mert a sorrend felcserélése (1. 23) és (1. 24) szerint az f konvex függvénynek az f^* konvex függvénnyel való helyettesítésével ekvivalens. Az f-konvergenciát definiálhatnánk úgy is, hogy mind (2. 3)-at, mind pedig a $\lim\limits_{n \to \infty} \mathcal{I}_f(\mu_0, \mu_n) = f(1)$ limeszrelációt feltesszük; az így adódó konvergenciafogalom azonban könnyen láthatóan az $\tilde{f} = f + f^*$ konvex függvénynek megfelelő $\tilde{f}$-konvergenciával ekvivalens, így pl. az $f(u) = u \log u$ esetben a $J(\mu_1, \mu_2) = \mathcal{I}_{(u-1)\log u}(\mu_1, \mu_2)$ J-divergencia nullához tartásával.

2. Az f-konvergencia természetesen nemcsak közönséges sorozatokra értelmezhető, hanem *Moore-Smith* féle általánosított sorozatokra, ill. az $\mathcal{M}$ tér részhalmazaiból álló rácsokra (filterbázisokra) is, és ekkor a konvergenciafogalom már egyértelműen meghatározza az $\mathcal{M}$ (V)-tér topológiai struktúráját. A továbbiakban azonban egyszerűség kedvéért csak sorozatkonvergenciáról fogunk beszélni.

3. Az információkonvergencia fogalmát [4] dolgozatomban vezettem be, ahol még a gyengébb „majdnem teljes" információkonvergenciával is foglalkoztam. Az ebben a fejezetben tárgyalt eredmények a [4] dolgozatomban az α-rendű I-divergenciára (vagyis az $f(u)=u^\alpha \operatorname{sgn}(\alpha-1)$, ill. $f(u)=u\log u$ esetre) kapott eredmények tetszőleges f-eltérésekre vonatkozó megfelelői. A 2. 1. tétel azonban lényegesen élesebb a [4] dolgozat megfelelő tételénél; ezt az élesítést az tette lehetővé, hogy a [4]-beli lemma helyett most az élesebb 2. 1. lemmát alkalmazom.

Az (1. 22) összefüggés szerint a variációs távolság értelmében vett konvergencia is az f-konvergencia speciális esete. A variációs távolság 0-hoz tartása a mértékek egyenletes konvergenciájával ekvivalens, vagyis azzal, hogy

$$\sup_{E\in\mathcal{X}} |\mu_1(E)-\mu_0(E)| \to 0 \qquad (n\to\infty). \tag{2.4}$$

Valóban, ha μ_1 és μ_2 két valószínűségi mérték $(X,\mathcal{X})$-en, akkor bármely $E\in\mathcal{X}$ halmazra $\mu_1(E)-\mu_2(E) = \mu_2(X-E)-\mu_1(X-E)$, tehát

$$\sup_{E\in\mathcal{X}}\big(\mu_1(E)-\mu_2(E)\big) = \sup_{E\in\mathcal{X}}\big(\mu_2(E)-\mu_1(E)\big) = \sup_{E\in\mathcal{X}}|\mu_1(E)-\mu_2(E)|$$

és így

$$|\mu_1-\mu_2| = \sup_{E\in\mathcal{X}}\big(\mu_1(E)-\mu_2(E)\big)+ \tag{2.5}$$

$$+\sup_{E\in\mathcal{X}}\big(\mu_2(E)-\mu_1(E)\big) = 2\sup_{E\in\mathcal{X}}|\mu_1(E)-\mu_2(E)|.$$

Meg fogjuk mutatni, hogy az f-konvergencia — feltéve, hogy az $f(u)$ függvény az $u_0=1$ pontban szigorúan konvex — mindig maga után vonja a valószínűségi mértékek egyenletes konvergenciáját. Ehhez felhasználjuk a következő, önmagában is érdekesnek látszó segédtételt, amely a *Jensen*-egyenlőtlenségben (1. 1. lemma) az egyenlőség közelítő fennállásának szükséges feltételére, azaz, szemléletesen szólva, az egyenlőség feltételének stabilitására vonatkozik (a jelölések ugyanazok, mint az 1. 1. lemmában).

2. 1. LEMMA. *Ha az $f(u)$ függvény az $u_0=M\xi$ pontban szigorúan konvex, akkor*

$$M|\xi-M\xi| \leqq \varphi_f\big(Mf(\xi)-f(M\xi)\big). \tag{2.6}$$

ahol $\varphi_f(v)$ egy (f-től függő) alkalmas monoton függvény, melyre $\lim_{v\to+0}\varphi_f(v)=\varphi_f(0)=0$ és $\varphi_f(v)>0$, ha $v>0$. Ha az $u_0=M\xi$ helyen $f(u)$ kétszer differenciálható és $f''(u_0)>$ $>a>0$, akkor a 0 elég kis δ_0 sugarú környezetében $\varphi_f(v)\leqq C\sqrt{v}$, ahol a C állandó választható pl. $C=\sqrt{\frac{8}{a}}$-nak. Ha $u_0=M\xi$ valamely r_0 sugarú környezetének (a_1,a_2)-be *eső részében mindenütt* $f''(u) \geqq a > 0$, *akkor választható* $\delta_0 = \frac{a}{2} r_0^2$, *tehát ekkor*

$$M|\xi - M\xi| \leqq C\sqrt{\delta} \quad ha \quad Mf(\xi) - f(M\xi) \leqq \delta \leqq \frac{a}{2} r_0^2. \tag{2.7}$$

Bizonyítás. Feltehetjük, hogy $a_1 < u_0 < a_2$; az $f(u)$ függvény konvex volta miatt $|u - u_0| \geqq \varepsilon_1 > 0$ esetén az (1. 7) egyenlőtlenségen túlmenőleg még

$$f(u) \geqq f(u_0) + b(u - u_0) + \frac{\varepsilon_2}{\varepsilon_1}|u - u_0| \tag{2.8}$$

is teljesül, ahol $\varepsilon_2 = \min\{f(u_0 + \varepsilon_1) - f(u_0) - b\varepsilon_1, f(u_0 - \varepsilon_1) - f(u_0) + b\varepsilon_1\}$; minthogy $f(u)$ az $u_0 = M\xi$ pontban szigorúan konvex, $\varepsilon_2 > 0$. (1. 7)-ből és (2. 8)-ból

$$Mf(\xi) \geqq f(M\xi) + \frac{\varepsilon_2}{\varepsilon_1} \int\limits_{|\xi - M\xi| \geqq \varepsilon_1} |\xi - M\xi| P(d\omega) \tag{2.9}$$

és így $Mf(\xi) - f(M\xi) \leqq \delta$ esetén

$$\begin{aligned} M|\xi - M\xi| = & \int\limits_{|\xi - M\xi| < \varepsilon_1} |\xi - M\xi| P(d\omega) + \\ & + \int\limits_{|\xi - M\xi| \geqq \varepsilon_1} |\xi - M\xi| P(d\omega) \leqq \varepsilon_1 + \frac{\varepsilon_1}{\varepsilon_2}\delta. \end{aligned} \tag{2.10}$$

Ha mármost az eddig tetszőleges $\varepsilon_1 > 0$ számot δ függvényében úgy választjuk, hogy $\delta \downarrow 0$ esetén mind ε_1, mind $\frac{\varepsilon_1}{\varepsilon_2}\delta$ monoton csökkenőleg nullához tartson, akkor (2. 10)-ben $\varepsilon_1 + \frac{\varepsilon_1}{\varepsilon_2}\delta = \varphi_f(\delta) \downarrow 0$; ezzel (2. 6)-ot igazoltuk. Az $f''(u_0) > a > 0$ esetben az

$$f(u_0) + b(u - u_0) + \frac{a}{2}(u - u_0)^2 \tag{2.11}$$

parabola egy elég kis intervallumban biztosan az $f(u)$ függvény görbéje alatt halad, így elég kis $\varepsilon_1 > 0$ esetén $\varepsilon_2 \geqq \frac{a}{2}\varepsilon_1^2$. Ezért pl. $\varepsilon_1 = \sqrt{\frac{2\delta}{a}}$ választással (2. 10)-ből $M|\xi - M\xi| < \sqrt{\frac{8}{a}}\sqrt{\delta}$ adódik (ha δ elég kicsi), tehát (2. 6)-ban valóban $\varphi_f(\delta) \leqq C\sqrt{\delta}$, ahol

$$C = \sqrt{\frac{8}{a}}. \tag{2.12}$$

Végül, ha u_0 valamely r_0 sugarú környezetének (a_1, a_2)-be eső részében $f''(u) \geqq a > 0$, akkor a (2. 11) parabola az $(u_0 - r_0, u_0 + r_0) \cap [a_1, a_2]$ intervallumban végig az $f(u)$ görbéje alatt halad, tehát az előbbi meggondolás $\varepsilon_1 \leqq r_0$, $\delta \leqq \frac{a}{2} r_0^2$ esetén biztosan alkalmazható. Ezzel a lemmát teljes egészében bebizonyítottuk.

MEGJEGYZÉSEK. *1.* Ha $f''(u_0)=0$ és az u_0 pontban a magasabbrendű deriváltak közül a $2k$-adik az első, amely nem tűnik el, akkor (2. 6)-ban $\varphi_f(v)$ választható $C_k\sqrt[2k]{v}$-nek, ha v elég kicsi. Ennek bizonyítása a fentihez teljesen hasonló. Ha az $f(u)$ görbének az $u_0=M\xi$ helyen törése van, már $\varphi_f(v)=C_0 v$ is megfelelő, ahol C_0 az u_0-beli jobb és bal oldali derivált különbségének felével egyenlő.

2. Ha $f''(u)\geqq a>0$ az (a_1, a_2) intervallumban mindenütt fennáll, akkor $f(u)$ görbéje az egész intervallumban a (2. 11) parabola fölött halad; ebből erre az esetre a (2. 7)-nél erősebb

$$D^2\xi = M|\xi-M\xi|^2 \leqq \frac{2}{a}\delta \tag{2. 13}$$

becslés adódik (a *Schwarz*-egyenlőtlenség alapján (2. 13)-ból következik (2. 7), mégpedig $C=\sqrt{\frac{2}{a}}$-val). Az általános esetben azonban hasonló becslés nem adható, sőt $D^2\xi$-nek nem is kell léteznie.

3. Könnyen látható, hogy ha $f''(u_0)>0$ létezik, az $M|\xi-M\xi|<C\sqrt{\delta}$ (ha δ elég kicsi) becslésben a $\sqrt{\delta}$ nagyságrend pontos. Valóban, $A>f''(u_0)$ esetén egy elég kis $[u_0-\varepsilon_0, u_0+\varepsilon_0]$ intervallumban $f(u)\leqq f(u_0)+b(u-u_0)+\frac{A}{2}(u-u_0)^2$; ezért, ha pl. ξ értéke $\frac{1}{2}-\frac{1}{2}$ valószínűséggel $u_0\pm\sqrt{\frac{2\delta}{A}}$, ahol $\delta\leqq\frac{A}{2}\varepsilon_0^2$ tetszőlegesen kicsi, akkor $Mf(\xi)-f(M\xi)\leqq\delta$ és ugyanakkor $M|\xi-M\xi|=\sqrt{\frac{2\delta}{A}}$.

4. A (2. 7) becslésben a C állandó (2. 12) értéke általában javítható. Bár ennek a továbbiak szempontjából nincs jelentősége, a C pontos értékének meghatározása — adott $f(u)$ esetén — semmilyen elvi problémát sem jelent. Valóban, ha ξ tetszőleges $[a_1, a_2]$-beli értékkészletű, véges $M\xi=u_0$ várható értékű valószínűségi változó, akkor az $\Omega_1=\{\omega\colon \xi(\omega)\geqq u_0\}$, $\Omega_2=\{\omega\colon \xi(\omega)<u_0\}$, $\xi^*(\omega)=M(\xi|\Omega_i)$ ha $\omega\in\Omega_i$ $(i=1,2)$ jelöléssel (feltehetjük, hogy $P(\Omega_i)>0$, $i=1,2$, mert különben $M|\xi-M\xi|=0$), nyilvánvalóan $M\xi^*=M\xi=u_0$, $M|\xi-u_0|=M|\xi^*-u_0|$, végül a *Jensen*-egyenlőtlenség (1. 1 lemma) szerint $M(f(\xi)|\Omega_i)\geqq f(M(\xi|\Omega_i))$ $(i=1,2)$, amiből $Mf(\xi)\geqq Mf(\xi^*)$. Ezért a C állandó $C_{\min}$ pontos értékének megállapításához elegendő az olyan ξ^* valószínűségi változókra szorítkozni, melyek csak két különböző értéket vehetnek fel és így valamely adott $f(u)$ függvényhez és u_0 értékhez tartozó $C_{\min}$ meghatározása elemi szélsőértékfeladatra vezet. Pl. az $f(u)=u\log u$ esetben azt a minimális C értéket kell megkeresnünk, amelyre még teljesül a

$$C\sqrt{px\log x+(1-p)\frac{u_0-px}{1-p}\log\frac{u_0-px}{1-p}-u_0\log u_0} \geqq 2p|x-u_0|$$

egyenlőtlenség $\left(0\leqq p\leqq 1 \text{ és } 0\leqq x\leqq\frac{u_0}{p}\right)$ vagy x helyére $\frac{y}{p}$-t írva,

$$\psi(p,y) = y\log\frac{y}{p}+(u_0-y)\log\frac{u_0-y}{1-p}-u_0\log u_0-\frac{4}{C^2}(y-u_0p)^2 \geqq 0$$
$$(0\leqq p\leqq 1;\ 0\leqq y\leqq u_0);$$

itt

$$\frac{\partial\psi}{\partial p} = (y-u_0 p)\left(-\frac{1}{p(1-p)}+\frac{8u_0}{C^2}\right),$$

tehát ha $C \geqq \sqrt{2u_0}$, akkor $p<\frac{y}{u_0}$ esetén $\frac{\partial\psi}{\partial p} \leqq 0$, $p>\frac{y}{u_0}$ esetén pedig $\frac{\partial\psi}{\partial p} \geqq 0$ és így mindenképpen $\psi(p, y) \geqq \psi\left(\frac{y}{u_0}, y\right)=0$. Ha viszont $C<\sqrt{2u_0}$, akkor elég kis abszolút értékű ε-ra $\psi\left(\frac{1}{2}+\varepsilon, \frac{u_0}{2}\right)<0$, tehát az $f(u)=u \log u$ függvényre $C_{\min}=\sqrt{2u_0}$. Jegyezzük meg, hogy ebben az esetben a (2. 7) egyenlőtlenség (ahol $C=\sqrt{2u_0}$) δ-ra vonatkozó megszorítás nélkül érvényes, de kis δ értékekre szorítkozva se adható $C_{\min}=\sqrt{2u_0}$-nál jobb konstans.

A következőkben a 2. 1. lemma alábbi változatára lesz szükségünk (a jelölések ugyanazok, mint az 1. 2. lemmában):

2. 2. LEMMA. *Ha az $f(u)$ konvex függvény az u_0 pontban szigorúan konvex, akkor megadható olyan $\varphi_f(v)$ függvény, melyre $v \downarrow 0$ esetén $\varphi_f(v) \downarrow 0$, hogy ha*

$$\frac{\int\limits_E \alpha(x)\lambda(dx)}{\int\limits_E \beta(x)\lambda(dx)} = u_0 \tag{2. 14}$$

és

$$\frac{1}{\int\limits_E \beta(x)\lambda(dx)} \int\limits_E \beta(x) f\left(\frac{\alpha(x)}{\beta(x)}\right)\lambda(dx) \leqq f(u_0)+\delta, \tag{2. 15}$$

akkor

$$\int\limits_E |\alpha(x)-u_0\beta(x)|\,\lambda(dx) \leqq \varphi_f(\delta)\int\limits_E \beta(x)\lambda(dx). \tag{2. 16}$$

Speciálisan, ha $f''(u)$ létezik és pozitív: $f''(u_0)>a>0$, akkor elég kis $\delta>0$ esetén (2. 14) *és a* (2. 15) *egyenlőtlenség fennállásából*

$$\int\limits_E |\alpha(x)-u_0\beta(x)|\,\lambda(dx) \leqq C\sqrt{\delta}\int\limits_E \beta(x)\lambda(dx) \tag{2. 17}$$

következik, ahol a C állandó (2. 12) *szerint választható. Ha az u_0 pont valamely r_0 sugarú környezetében $f''(u) \geqq a>0$, akkor az előbbi állítás minden $\delta \leqq \frac{a}{2} r_0^2$ esetén igaz.*

Bizonyítás. Ha $\beta(x)$ az E halmazon mindenütt pozitív, a 2. 2. lemma éppen úgy speciális esete a 2. 1. lemmának, mint az 1. 2. lemma az 1. 1. lemmának. A bizonyítás az általános esetben is teljesen analóg a 2. 1. lemma bizonyításával, minthogy a (2. 8) egyenlőtlenség, az (1. 7)-hez hasonlóan, u helyére formálisan $\frac{\alpha(x)}{\beta(x)}$-et írva és $\beta(x)$-szel végigszorozva, akkor is érvényben marad, ha $\beta(x)=0$.

2. 1. TÉTEL. *Ha az f konvex függvény az $u_0=1$ pontban szigorúan konvex, akkor megadható olyan $\varphi_f(v)$ függvény, melyre $v\downarrow 0$ esetén $\varphi_f(v)\downarrow 0$, hogy*

(2. 18) $$\varrho(\mu_1,\mu_2)=|\mu_1-\mu_2|\leqq\varphi_f(\mathscr{I}_f(\mu_1,\mu_2)-f(1)).$$

Ha $f''(1)>a>0$, akkor elég kis δ esetén választható $\varphi_f(\delta)=C\sqrt{\delta}$; speciálisan, ha az $u_0=1$ hely valamely r_0 sugarú környezetében $f''(u)\geqq a>0$, akkor ez $\delta\leqq\frac{a}{2}r_0^2$ esetén biztosan megfelelő, azaz

(2. 19) $$\varrho(\mu_1,\mu_2)=|\mu_1-\mu_2|\leqq C\sqrt{\delta}\quad\text{ha}\quad\mathscr{I}_f(\mu_1,\mu_2)-f(1)\leqq\delta\leqq\frac{a}{2}r^2.$$

Itt a C állandó pl. (2. 12) szerint választható.

Bizonyítás. A tétel állítása közvetlenül következik a 2. 2. lemmából, ha E-nek az egész X teret választjuk és $\alpha(x)$, ill. $\beta(x)$ helyére a μ_1, ill. μ_2 mérték $p_1(x)$, ill. $p_2(x)$ sűrűségfüggvényét helyettesítjük, mikor is $u_0=1$.

KOROLLÁRIUM.

(2. 20) $$|\mu_1-\mu_2|\leqq\sqrt{2I(\mu_1\|\mu_2)}.$$

Bizonyítás. Minthogy $I(\mu_1\|\mu_2)=\mathscr{I}_{u\log u}(\mu_1,\mu_2)$ és $1\log 1=0$, $(u\log u)''=\frac{1}{u}$, a 2. 1. tétel értelmében $|\mu_1-\mu_2|\leqq C\sqrt{I(\mu_1\|\mu_2)}$. A 2. 1. lemmához fűzött 4. megjegyzés értelmében itt a C konstans lehető legjobb értéke $C_{\min}=\sqrt{2}$; a 2. 1. tétel közvetlen alkalmazásával csak az a gyengébb becslés adódik, hogy $I(\mu_1\|\mu_2)\leqq\frac{r_0^2}{2(1+r_0)}$ esetén $|\mu_1-\mu_2|\leqq\sqrt{8(1+r_0)}\sqrt{I(\mu_1\|\mu_2)}$ így speciálisan mindenképpen $|\mu_1-\mu_2|\leqq$ $\leqq 4\sqrt{I(\mu_1\|\mu_2)}$.

MEGJEGYZÉS. A $|\mu_1-\mu_2|\leqq C\sqrt{I(\mu_1\|\mu_2)}$ egyenlőtlenség lényegében PINSZKER-től származik, aki [27] munkájában kimutatta, hogy $\mu_1\ll\mu_2$, $\frac{\mu_1(dx)}{\mu_2(dx)}=a(x)$ esetén

$$\int|\log a(x)|\mu_1(dx)\leqq\int\log a(x)\,\mu_1(dx)+\Gamma\sqrt{\int\log a(x)\mu_1(dx)}=I(\mu_1\|\mu_2)+\Gamma\sqrt{I(\mu_1\|\mu_2)},$$

ahol Γ alkalmas abszolút konstans, továbbá, hogy $|\mu_1-\mu_2|\leqq 2\int|\log a(x)|\mu_1(dx)$. (A bizonyítást részletesen arra az esetre közli, amikor μ_1 valamely (ξ,η) valószínűségi változó pár együttes eloszlása, μ_2 pedig a marginális eloszlások direkt szorzata, amikor is $I(\mu_1\|\mu_2)$ az $I(\xi,\eta)$ kölcsönös információt adja, megjegyzi azonban, hogy az eredmény az általános esetben is igaz.)

A 2. 1. tétel szerint az ilyen típusú becslés sokkal általánosabban is érvényes, s az itt adott bizonyítás az $f(u)=u\log u$ speciális esetre szorítkozva is egyszerűbb az eredetinél, főleg ha (2. 20) helyett megelégszünk a gyengébb $|\mu_1-\mu_2|\leqq 4\sqrt{I(\mu_1\|\mu_2)}$ egyenlőtlenséggel. Megemlítem még, hogy az utóbbi esetre a (2. 20)-nál gyengébb $|\mu_1-\mu_2|\leqq 2\sqrt{I(\mu_1\|\mu_2)}$ becslés egyszerűen igazolható, de ugyancsak az $u\log u$ függvény speciális tulajdonságainak felhasználásával, l. [32], 33. o. (2. 19)-ben a $\sqrt{\delta}$ nagyságrend minden esetben pontos (ha $f''(1)>0$ létezik), mint ez a 2. 1. lemmához fűzött 3. megjegyzésből következik.

A 2. 1. tételből azonnal következik, hogy ha a valószínűségi mértékek valamely μ_n sorozata (vagy általánosabban egy $\{\mu_d\}_{d\in D}$ *Moore—Smith*-sorozata) f-konvergál egy μ_0 mértékhez, akkor az illető mértékek egyenletesen is konvergálnak μ_0-hoz, feltéve, hogy az $f(u)$ konvex függvény az $u_0=1$ pontban szigorúan konvex. Ha az $f(u)$ függvényről még azt is feltesszük, hogy $f(0)=\lim\limits_{u\to+0} f(u)$ és $0f\left(\frac{1}{0}\right)=\lim\limits_{u\to+\infty}\frac{f(u)}{u}$ véges, akkor megfordítva, az egyenletes konvergenciából is következik az f-konvergencia. Pontosabban, igaz a következő

2. 2. TÉTEL. *Ha mind* $\lim\limits_{u\to+0} f(u)$, *mind* $\lim\limits_{u\to+\infty}\frac{f(u)}{u}$ *véges, akkor*

$$\mathcal{I}_f(\mu_1,\mu_2)\leqq f(1)+C_1\sqrt{\varrho(\mu_1,\mu_2)}, \tag{2.21}$$

ahol C_1 *csak az* f *függvénytől függő állandó.*

Bizonyítás. A feltételből következik, hogy a (0, 1) intervallumban mind $f(0)$, mind $uf(1/u)$ korlátos, tehát alkalmas K_1 és K_2 állandóval

$$|f(u)-f(1)|<K_1,\quad u\left|f\left(\frac{1}{u}\right)-f(1)\right|<K_2\qquad (0<u\leqq 1). \tag{2.22}$$

Legyen $\varepsilon<1$ tetszőleges pozitív szám és vezessük be a

$$\begin{cases} E_\varepsilon^1=\{x\colon (1-\varepsilon)p_1(x)\geqq p_2(x);\ p_1(x)>0\}\\ E_\varepsilon^2=\{x\colon (1-\varepsilon)p_2(x)\geqq p_1(x)\},\\ E_\varepsilon\ =X-(E_\varepsilon^1\cup E_\varepsilon^2)=\left\{x\colon 1-\varepsilon<\dfrac{p_1(x)}{p_2(x)}<\dfrac{1}{1-\varepsilon}\right\}\end{cases} \tag{2.23}$$

jelölést. Ekkor $x\in E_\varepsilon$ esetén

$$\left|f\left(\frac{p_1(x)}{p_2(x)}\right)-f(1)\right|\leqq K_\varepsilon\cdot\varepsilon, \tag{2.24}$$

ahol

$$K_\varepsilon=\max\left\{\left|\frac{f(1)-f(1-E)}{\varepsilon}\right|,\ \left|\frac{f\left(\frac{1}{1-\varepsilon}\right)-f(1)}{\frac{1}{1-\varepsilon}-1}\right|\right\}$$

és ε csökkentésével K_ε is csökken (kihasználva $f(u)$ konvexitását). Jegyezzük meg továbbá, hogy (2. 22) és (2. 23) értelmében

$$\begin{cases}\displaystyle\int\limits_{E_\varepsilon^1}\left|f\left(\frac{p_1(x)}{p_2(x)}\right)-f(1)\right|\frac{p_2(x)}{p_1(x)}p_1(x)\lambda(dx)\leqq K_2\mu_1(E_\varepsilon^1),\\ \displaystyle\int\limits_{E_\varepsilon^2}\left|f\left(\frac{p_1(x)}{p_2(x)}\right)-f(1)\right|p_2(x)\lambda(dx)\leqq K_1\mu_2(E_\varepsilon^2)\end{cases} \tag{2.25}$$

és hogy (2. 23) szerint

$$\varrho(\mu_1, \mu_2) \geqq \int_{E_\varepsilon^i} |p_1(x)-p_2(x)|\lambda(dx) \geqq \varepsilon\mu_i(E_\varepsilon^i) \qquad (i = 1, 2). \tag{2.26}$$

A (2.24), (2. 25) és (2. 26) egyenlőtlenségből végül azt kapjuk, hogy

$$\mathcal{I}_f(\mu_1, \mu_2)-f(1) \leqq \int \left| f\left(\frac{p_1(x)}{p_2(x)}\right) - f(1)\right| p_2(x)\lambda(dx) = \tag{2.27}$$

$$= \int_{E_\varepsilon} + \int_{E_\varepsilon^1} + \int_{E_\varepsilon^2} \leqq K_\varepsilon \cdot \varepsilon + \frac{K_1+K_2}{\varepsilon}\,\varrho(\mu_1, \mu_2).$$

Innen, pl. az $\varepsilon = \frac{1}{2}\sqrt{\varrho(\mu_1, \mu_2)}$ választással éppen (2. 21) adódik, figyelembe véve, hogy ekkor $\varepsilon \leqq \frac{\sqrt{2}}{2}$ és így (2. 27)-ben $K_\varepsilon \leqq K_{\frac{\sqrt{2}}{2}}$.

A 2. 2. tétel feltételei mellett tehát — feltéve, hogy $f(u)$ az $u_0 = 1$ pontban szigorúan konvex — az f-környezetek (metrizálható) topologikus teret definiálnak, ugyanis az f-környezetek rendszere ekvivalens a ϱ metrika szerinti környezetek rendszerével (tehát az f-konvergencia a mértékek egyenletes konvergenciájával). Ha azonban $f(0) = +\infty$ vagy $\lim\limits_{u\to\infty} \frac{f(u)}{u} = +\infty$, akkor ez az ekvivalencia általában már nem teljesül, sőt, ekkor az eloszlások $\mathcal{M}$ halmazán az f-környezetekkel definiált (V)-tér általában nem is topologikus tér, még akkor sem, ha az X alaptér megszámlálható és $\mathcal{M}$-nek az összes olyan eloszlások halmazát vesszük, melyeknél X minden pontja pozitív mértékű.

Ha $\mathcal{M}$-ben az f-környezetek topológiát definiálnának, akkor minden $\mu_0 \in \mathcal{M}$ esetén teljesülnie kellene a következő (T) tulajdonságnak:

(T) *minden $U_f(\mu_0; \varepsilon)$ környezethez található olyan $U_f(\mu_0; \varepsilon')$ környezet, hogy minden $\mu \in U_f(\mu_0; \varepsilon')$ eloszlásnak van olyan $U_f(\mu; \varepsilon'')$ környezete, hogy $U_f(\mu; \varepsilon'') \subset U_f(\mu_0; \varepsilon)$.*

2. 3. TÉTEL. *Ha $\mathcal{M}$ valamely megszámlálhatóan végtelen X halmazon értelmezett összes olyan valószínűségi mértékek halmazát jelenti, melyeknél minden $x \in X$ pont pozitív mértékű és az $f(u)$ konvex függvényre vagy $f(0) = +\infty$ vagy $\lim\limits_{u\to\infty} \frac{f(u)}{u} = +\infty$, akkor a ($T$) tulajdonság egyetlen $\mu_0 \in \mathcal{M}$ eloszlásra sem teljesül.*

Bizonyítás. Az X megszámlálható halmazon minden valószínűségeloszlás egy $\mathcal{P} = (p_1, p_2, \ldots)$ számsorozattal adható meg, ahol p_i az $\{x_i\}$ ($x_i \in X$) egy pontból álló halmaz mértékét jelenti ($p_i \geqq 0$, $i = 1, 2, \ldots$; $\sum\limits_{i=1}^{\infty} p_i = 1$). Speciálisan az $\mathcal{M}$-beli eloszlásokat az jellemzi, hogy mindegyik $p_i > 0$. A $\mathcal{P} = (p_1, p_2, \ldots)$ és $Q = (q_1, q_2, \ldots)$ eloszlások f-eltérésére az 1. 1. definícióból a következő adódik:

$$\mathcal{I}_f(\mathcal{P}, Q) = \sum_{i=1}^{\infty} q_i f\left(\frac{p_i}{q_i}\right). \tag{2.28}$$

A tétel állításánál többet is fogunk bizonyítani, nevezetesen azt, hogy ha az $f(u)$ konvex függvényre $f(0) = \lim_{u\to+0} f(u) = +\infty$ vagy $\lim_{u\to+\infty} \frac{f(u)}{u} = +\infty$ (vagy mindkettő) teljesül, akkor az X-en értelmezett bármely $\mathcal{P}=(p_1, p_2, \dots)\in\mathcal{M}$ eloszláshoz és $\varepsilon>0$-hoz található olyan $Q\in\mathcal{M}$ eloszlás, hogy $\mathcal{I}_f(Q, \mathcal{P})<f(1)+\varepsilon$, továbbá ehhez és bármely $\varepsilon'>0$-hoz (ε' függhet Q-tól is) olyan $\mathcal{R}\in\mathcal{M}$ eloszlás, hogy $\mathcal{I}_f(\mathcal{R}, Q)<f(1)+$ $+\varepsilon'$, de $\mathcal{I}_f(\mathcal{R}, \mathcal{P}) = +\infty$. Valóban, ha $\sum_{k=1}^{\infty} \hat{q}_k$ tetszőleges pozitív tagú konvergens sor, melyre

$$\sum_{k=1}^{\infty} p_k f\left(\frac{\hat{q}_k}{p_k}\right) < +\infty, \tag{2.29}$$

továbbá $\sum_{k=1}^{\infty} \hat{r}_k$ olyan pozitív tagú konvergens sor, hogy

$$\sum_{k=1}^{\infty} \hat{q}_k f\left(\frac{\hat{r}_k}{\hat{q}_k}\right) < +\infty, \quad \sum_{k=1}^{\infty} p_k f\left(\frac{\hat{r}_k}{p_k}\right) = +\infty, \tag{2.30}$$

akkor a $Q = (q_1, q_2, \dots)\in\mathcal{M}$ eloszlást a

$$q_k = \begin{cases} c p_k & \text{ha} \quad k<N \\ \hat{q}_k & \text{ha} \quad k\geqq N \end{cases} \qquad c = \frac{1-\sum_{k=N}^{\infty}\hat{q}_k}{\sum_{k=1}^{N-1} p_k} \tag{2.31}$$

egyenlőséggel definiálva, elég nagy N esetén

$$\mathcal{I}_f(Q, \mathcal{P}) = \sum_{k=1}^{\infty} p_k f\left(\frac{q_k}{p_k}\right) = f\left(\frac{1}{c}\right)\sum_{k=1}^{N-1} p_k + \sum_{k=N}^{\infty} p_k f\left(\frac{\hat{q}_k}{p_k}\right) < f(1)+\varepsilon \tag{2.32}$$

adódik (kihasználva (2. 29)-et és azt, hogy $N\to\infty$ esetén $c\to 1$), és hasonlóképpen az

$$r_k = \begin{cases} c' q_k, & \text{ha} \quad k<N' \\ \hat{r}_k, & \text{ha} \quad k\geqq N' \end{cases} \qquad c' = \frac{1-\sum_{k=N'}^{\infty}\hat{r}_k}{\sum_{k=1}^{N'-1} q_k} \tag{2.33}$$

összefüggéssel definiált $\mathcal{R}=(r_1, r_2, \dots)\in\mathcal{M}$ eloszlásra elég nagy $N'>N$ esetén

$$\mathcal{I}_f(\mathcal{R}, Q) < f(1)+\varepsilon', \tag{2.34}$$

ugyanakkor azonban

$$\sum_{k=N'}^{\infty} p_k f\left(\frac{r_k}{p_k}\right) = \sum_{k=N'}^{\infty} p_k f\left(\frac{\hat{r}_k}{p_k}\right) = +\infty, \tag{2.35}$$

tehát $\mathcal{I}_f(\mathcal{R}, \mathcal{P}) = +\infty$.

Így tehát már csak azt kell igazolni, hogy mind az $f(0) = +\infty$, mind a $\lim_{u\to\infty} \frac{f(u)}{u} = +\infty$ esetben léteznek a (2.29) és (2. 30) feltételt kielégítő $\sum_{k=1}^{\infty} \hat{q}_k$ és $\sum_{k=1}^{\infty} \hat{r}_k$ konvergens sorok.

Ha $f(0) = +\infty$, akkor válasszuk meg az $\alpha_k \to +\infty$ számsorozatot úgy, hogy $\sum_{k=1}^{\infty} \alpha_k p_k$ még konvergens legyen és legyen $\hat{q}_k = p_k f^{-1}(\alpha_k)$ (ahol $f^{-1}(\alpha)$ az $f(u) = \alpha$ egyenletnek eleget tevő u értékek legkisebbikét jelöli; ha ilyen érték nincs, akkor legyen pl. $f^{-1}(\alpha) = 1$). Ekkor $\frac{\hat{q}_k}{p_k} \to 0$, tehát $\sum_{k=1}^{\infty} \hat{q}_k$ konvergens és (2. 29) teljesül. Ezután legyen $\beta_k \to +\infty$ olyan számsorozat, melyre $\sum_{k=1}^{\infty} \beta_k \hat{q}_k$ konvergens, de $\sum_{k=1}^{\infty} \beta_k p_k$ divergens; $\frac{\hat{q}_k}{p_k} \to 0$ miatt ilyen sorozat biztosan létezik. Ha mármost $\hat{r}_k = \hat{q}_k f^{-1}(\beta_k)$ (ahol f^{-1} értelmezése ugyanaz, mint fentebb), akkor (2. 30) is teljesül, figyelembe véve, hogy $\frac{\hat{q}_k}{p_k} \to 0$ miatt elég nagy k-ra $f\left(\frac{\hat{r}_k}{p_k}\right) > f\left(\frac{\hat{r}_k}{q_k}\right) = \beta_k$.

A $\lim\limits_{u\to\infty} \frac{f(u)}{u} = +\infty$ esetben a következő lemmából indulunk ki.

2. 3. LEMMA. *Ha $\sum_{k=1}^{\infty} a_k$ tetszőleges pozitív tagú konvergens sor és $\psi(u)$ tetszőleges, $u > 0$ esetén értelmezett pozitív értékű függvény, melyre $\lim\limits_{u\to\infty} \psi(u) = +\infty$, akkor létezik olyan $b_k \to +\infty$ számsorozat, hogy*

$$\sum_{k=1}^{\infty} a_k b_k < +\infty; \quad \sum_{k=1}^{\infty} a_k b_k \psi(b_k) = +\infty. \tag{2. 36}$$

A lemma bizonyítása. Legyen $\psi_1(u)$ olyan monoton növekedőleg végtelenhez tartó folytonos függvény, melyre $0 < \psi_1(u) \leqq \psi(u)$ $(0 < u < +\infty)$; ilyen ψ_1 nyilván létezik. Tekintsük a természetes számok olyan gyorsan növekvő n_l sorozatát, hogy a $\sum_{k=n_l}^{\infty} a_k = \frac{1}{c_l \psi_1(c_l)}$ összefüggéssel definiált c_l pozitív számokra $\sum_{l=1}^{\infty} \frac{1}{\psi_1(c_l)} < +\infty$ legyen. Ekkor a $b_k = \sum_{n_l \leqq k} c_l$ monoton növekedőleg végtelenhez tartó sorozat megfelel a (2. 36) követelményeknek, ugyanis c_l értelmezése szerint

$$\sum_{k=1}^{\infty} a_k b_k = \sum_{l=1}^{\infty} c_l \sum_{k=n_l}^{\infty} a_k = \sum_{l=1}^{\infty} \frac{1}{\psi_1(c_l)} < +\infty$$

és

$$\sum_{k=1}^{\infty} a_k b_k \psi(b_k) \geqq \sum_{l=1}^{\infty} c_l \psi_1(c_l) \sum_{k=n_l}^{\infty} a_k = \sum_{l=1}^{\infty} 1 = +\infty;$$

itt felhasználtuk, hogy $\psi(u) \geqq \psi_1(u)$ és $\psi_1(u)$ monoton volta miatt $b_k \psi(b_k) \geqq b_k \psi_1(b_k) \geqq \geqq \sum_{n_l \leqq k} c_l \psi_1(c_l)$.

Alkalmazzuk most a 2. 3. lemmát a $\sum p_k$ sorra (vagyis $q_k = p_k$), a $\psi(u) = f^{-1}\left(\frac{u}{f^{-1}(u)}\right)$ választással; itt, az előző esettől eltérőleg, $f^{-1}(u)$-n a legnagyobb olyan v értéket értjük, melyre $f(v) = u$ (ha ilyen v nincs, akkor legyen pl. $f^{-1}(u) = 1$.)

A lemma feltételei teljesülnek, mert a $\lim\limits_{u\to\infty}\frac{f(u)}{u}=+\infty$ feltétel értelmében $\lim\limits_{u\to\infty}\psi(u)=+\infty$.

Tekintsük a lemma szerinti b_k számokat és legyen

$$\hat{q}_k = p_k f^{-1}(b_k); \quad \hat{r}_k = \hat{q}_k f^{-1}\left(\frac{b_k}{f^{-1}(b_k)}\right) = \hat{q}_k\psi(b_k). \tag{2. 37}$$

Ekkor $f\left(\frac{\hat{q}_k}{p_k}\right)=b_k$, (ha k elég nagy), tehát (2. 36) első fele éppen (2. 29)-et adja. Továbbá (2. 37) második feléből $f\left(\frac{\hat{r}_k}{\hat{q}_k}\right)=\frac{b_k}{f^{-1}(b_k)}=b_k\frac{p_k}{q_k}$ és így (2. 36) első fele a (2. 30) első felét is szolgáltatja. Végül az $f(u)$ konvex volta és $\lim\limits_{u\to\infty}f(u)=+\infty$ miatt $f(u)$ görbéje elég nagy u_0 esetén biztosan az origót az $(u_0, f(u_0))$ ponttal összekötő egyenes fölött halad, ha $u>u_0$; ezt $u_0=\frac{\hat{q}_k}{p_k}$-ra és $u=\frac{\hat{r}_k}{p_k}$-ra alkalmazva, (2. 37) felhasználásával

$$f\left(\frac{\hat{r}_k}{p_k}\right) > \frac{f\left(\frac{\hat{q}_k}{p_k}\right)}{\frac{\hat{q}_k}{p_k}}\cdot\frac{\hat{r}_k}{p_k} = f\left(\frac{\hat{q}_k}{p_k}\right)\frac{\hat{r}_k}{\hat{q}_k} = b_k\psi(b_k) \tag{2. 38}$$

(ha k elég nagy), tehát (2. 30) második fele is következik (2. 36) második feléből. Ezzel a 2. 3. tétel bizonyítását befejeztük.

A 2. 3. tétel természetesen azt is jelenti, hogy az $f(0)+\lim\limits_{u\to\infty}\frac{f(u)}{u}=+\infty$ esetben az f-konvergencia (általános értelemben, azaz *Moore—Smith*-sorozatokra, ill. rácsokra értve) nem tekinthető topologikus térbeli konvergenciafogalomnak, illetőleg, hogy az eloszlássorozatok konvergenciája nem származtatható az első megszámlálhatósági axiómának eleget tevő topológiából. Nyitva maradó kérdés, hogy megadható-e egyáltalán $\mathscr{M}$-ben olyan topológia, amelyben az eloszlássorozatok konvergenciája ekvivalens az f-konvergenciával.

Láttuk, hogy $f(0)+\lim\limits_{u\to\infty}\frac{f(u)}{u}=+\infty$ esetén abból, hogy μ_n egyenletesen (más szóval a ϱ variációs távolság szerint) tart μ-höz, még nem következik a $\mu_n\to\mu$ f-konvergencia. Ha azonban azt az erősebb kikötést tesszük, hogy még $(\mu_n(A)-\mu(A))/ /(\mu_n(A)+\mu(A))$ is A-ban egyenletesen nullához tart $(A\in\mathscr{X},\ \mu_n(A)+\mu(A)>0)$, ami könnyen láthatólag azzal ekvivalens, hogy elég nagy n-re már $\mu_n\ll\mu$ és $p_n(x)=\frac{\mu_n(dx)}{\mu(dx)}$ egyenletesen tart 1-hez $[\mu]$, akkor $u_n\xrightarrow{f}\mu$ már biztosan teljesül. Ez egyszerűen abból következik, hogy $\mu_n\ll\mu$, vrai sup $|p_n(x)-1|<\delta$ esetén nyilvánvalóan ($\lambda=\mu$ választással)

$$\mathscr{I}_f(\mu_n,\mu) = \int f(p_n(x))\mu(dx)\leqq f(1)+\varepsilon, \tag{2. 39}$$

ahol[7]

$$\varepsilon = \max |f(u) - f(1)| \to 0, \quad \text{ha} \quad \delta \to 0.$$

Másképpen megfogalmazva, ha a valószínűségeloszlások terében bevezetjük pl. a

$$(2.40) \quad \varrho'(\mu_1, \mu_2) = \log \max (q(\mu_1, \mu_2), q(\mu_2, \mu_1)); \; q(\mu_i, \mu_j) = \operatorname{vrai\,sup} \frac{\mu_i(dx)}{\mu_j(dx)}$$

metrikát, mely szerint az $(X, \mathscr{X})$-en értelmezett összes, valamely rögzített μ_0 mértékkel ekvivalens μ eloszlások teljes metrikus teret alkotnak (ha eltekintünk attól, hogy $\varrho'(\mu_1, \mu_2)$ értéke $+\infty$ is lehet), akkor a $\mu_n \xrightarrow{\varrho'} \mu$ konvergenciából mindig következik a $\mu_n \xrightarrow{f} \mu$ f-konvergencia. Ugyanakkor azonban a 2. 2. és 2. 3. tételből nyilvánvaló (és közvetlenül is könnyen igazolható), hogy az f-konvergenciából semmilyen f esetén sem következik a ϱ'-konvergencia[8], feltéve, hogy a vizsgált eloszláshalmaz elég tág.

Eredményeinket az információkonvergencia esetére alkalmazva megállapíthatjuk, hogy a ϱ' metrika szerinti konvergenciából (vagyis $\frac{\mu_n(A) - \mu(A)}{\mu_n(A) + \mu(A)}$ egyenletesen 0-hoz való tartásából) mindig következik az információkonvergencia, de megfordítva nem. Az α-rendű információkonvergencia $0 < \alpha < 1$ esetén ekvivalens a mértékek variációs távolság szerinti (vagyis egyenletes) konvergenciájával, $\alpha \geqq 1$ esetén azonban, bár az információkonvergencia maga után vonja a mértékek egyenletes konvergenciáját, a megfordítás már nem érvényes. Ebből a szempontból tehát az α-rendű információ $0 < \alpha < 1$ esetén kedvezőbben viselkedik, mint a közönséges (elsőrendű) információ. Szemléletesen szólva ennek az az oka, hogy $0 < \alpha < 1$ esetén az α-rendű információ a kis valószínűségekre nem érzékeny, $\alpha \geqq 1$ esetén azonban igen; ez a tulajdonság, amely abban is megnyilvánul, hogy $\alpha \geqq 1$ esetén $I_\alpha(\mu_1 \| \mu_2)$ végtelenné válik, ha $\mu_1 \not\ll \mu_2$, $0 < \alpha < 1$ esetén viszont nem, ismét alátámasztja RÉNYI Alfrédnak [30] azt a megállapítását, hogy a különböző α-rendű információmennyiségek közötti választás lehetősége sokszor előnyös lehet és nem biztos, hogy mindig az elsőrendű információ használata a legcélszerűbb. Ezt a megállapítást kiegészíthetjük azzal, hogy az f-eltérések a választékot még inkább bővítik (bár ezek már nem rendelkeznek az információ additív tulajdonságával), ami egyes esetekben szintén hasznosnak bizonyulhat.

Az f-eltérésekkel kapcsolatban felvethető az a kérdés is, hogy található-e olyan $\varphi(t)$ szigorúan monoton növekedő folytonos függvény, hogy $\varphi(\mathscr{I}_f(\mu_1, \mu_2))$ kvázitávolság legyen $\mathscr{M}$-ben, azaz eleget tegyen a háromszögegyenlőtlenségnek. A 2. 3. tétel értelmében ilyen φ függvény létezése csak akkor várható, ha $f(0) < +\infty$, $\lim\limits_{u \to \infty} \frac{f(u)}{u} < +\infty$, így például az $I(\mu_1 \| \mu_2)$ I-divergenciának, illetőleg a $J(\mu_1, \mu_2)$ J-divergenciának semmilyen szigorúan monoton folytonos függvénye sem lehet kvázitávolság, ill. távolság (sőt jóval általánosabb típusú függvényei sem, vö. [5]).

[7] Ha $f''(1)$ létezik és $f''(1) < A$, akkor (2. 39)-ben ε választható $\frac{1}{2} A\delta^2$-nek is, ha δ elég kicsi; kihasználva, hogy ekkor $f(u) \leqq f(1) + b(u-1) + \frac{A}{2}(u-1)^2$, ha $|u-1| \leqq \delta$.

[8] KULLBACK [21] könyvében a 71. oldalon levő 2.1 lemma második állítása téves.

Az $f(u) = -u^\alpha$ $(0<\alpha<1)$ esetre vonatkozólag megemlítem azt a lényegében CZIPSZER Jánostól származó eredményt, hogy ilyenkor a $\varphi(t)=(1+t)^{\alpha'}$ $(\alpha'=\min(\alpha, 1-\alpha))$ függvény megfelelő, vagyis a

$$\Delta_\alpha(\mu_1, \mu_2) = \left[1 - \int p_1^\alpha(x) p_2^{1-\alpha}(x)\lambda(dx)\right]^{\alpha'}$$

$(0<\alpha<1; \alpha'=\min(\alpha, (\alpha, 1-\alpha))$ mennyiség eleget tesz a háromszögegyenlőtlenségnek, azaz kvázitávolság $\mathcal{M}$-ben. Speciálisan $\alpha=\frac{1}{2}$ esetén $\Delta_{\frac{1}{2}}(\mu_1, \mu_2)$ nem más, mint a *Bhattacharyya*-féle távolság. Ezzel a kérdéskörrel FISCHER Jánossal közös [9] dolgozatunkban foglalkoztunk.

IDÉZETT IRODALOM

[1] J. ACZÉL und Z. DARÓCZY: „Charakterisierung der Entropien positiver Ordnung und der Shannonschen Entropie". *Acta Math. Acad. Sci. Hung.* **14** (1963) 95—121.

[2] A. BHATTACHARYYA: „On some analogues of the amount of information and their use in statistical estimation". *Sankhya* **8** (1946) 1—14.

[3] D. BLACKWELL: „Equivalent comparisons of experiments". *Ann. Math. Stat.* **24** (1953) 265—272.

[4] I. CSISZÁR: „Informationstheoretische Konvergenzbegriffe im Raum der Wahrscheinlichkeitsverteilungen". *MTA Mat. Kut. Int. Közleményei* **7** (1962) 137—158.

[5] I. CSISZÁR: „Über topologische und metrische Eigenschaften der relativen Information der Ordnung α". *Transactions of the Third Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, Prague,* 1964; 63—73.

[6] I. CSISZÁR: „Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten". *MTA Mat. Kut. Int. Közleményei* **8** (1963) 85—108.

[7] I. CSISZÁR: „A note on limiting distributions on topological groups". *MTA Mat. Kut. Int. Közleményei* **9** (1964) 595—599.

[8] I. CSISZÁR: „On infinite products of random elements and infinite convolutions of probability distributions on locally compact groups". *Zeitschrift für Wahrscheinlichkeitstheorie* **5** (1966) 279—295.

[9] I. CSISZÁR—J. FISCHER: „Informationsentfernungen im Raum der Wahrscheinlichkeitsverteilungen". *MTA Mat. Kut. Int. Közleményei* **7** (1962) 159—180.

[10] R. L. DOBRUSIN: „A Shannon-féle alaptétel általános megfogalmazása az információelméletben". *MTA III. Oszt. Közleményei* **9** (1961) 427—456, **10** (1962) 51—103 és 141—167. (Az orosz eredeti: Успехи Математических Наук, **14** (1959) 3—104.)

[11] J. L. DOOB: *Stochastic Processes,* Wiley, New York, 1953.

[12] I. M. GELFAND—J. JAGLOM: „Über die Berechnung der Menge an Information über eine zufällige Funktion, die in einer anderen zufälligen Funktion enthalten ist". *Arbeiten zur Informationstheorie II. Deutscher Verlag der Wissenschaften, Berlin,* 1958, 7—56.

[13] Я. ГАЕК: „Об одном свойстве нормальных распределений произвольного стохастического процесса". *Чехосл. Мат. Журнал* **8** (1958) 610—619.

[14] P. R. HALMOS and L. J. SAVAGE: „Application of the Radon—Nikodym theorem to the theory of sufficient statistics". *Ann. Math. Stat.* **20** (1949) 225—241.

[15] K. ITO and Y. KAWADA: „On the probability distribution on a compact group". *Proc. Phys. Math. Soc. Japan,* **22** (1940) 977—998.

[16] H. JEFFREYS: *Theory of Probability, 2nd edition.* Clarendon Press, Oxford, 1948.

[17] G. KALLIANPUR: „On the amount of information contained in a σ-field". *Contributions to probability and statistics, the Harald Cramér volume, Stanford,* 1960. 265—271.

[18] D. G. KENDALL: „Information theory and the limit theorem for Markov chains and processes with a countable infinity of states". *Annals of the Institute of Statistical Mathematics* **15** (1964) 137—143.

[19] Б. М. КЛОСС: „О вероятностных распределениях на бикомпактных топологических группах". *Теория Вер. и Примен.* **4**/1959/255—290.

[20] S. KULLBACK and R. A. LEIBLER: „On information and sufficiency". *Ann. Math. Stat.* **22** (1951) 79—87.

[21] S. KULLBACK: *Information Theory and Statistics*. Wiley, New York, 1959.

[22] Ю. В. Линник: „Теоретико-информационное доказательство центральной предельной теоремы в условиях Линдеберга". *Теория Вер. и Примен.* 4/1959/ 311—321.

[23] M. LOÈVE: *Probability Theory*. Van Nostrand, 1955.

[24] A. PEREZ: „Notions généralisées d'incertitude, d'entropie et d'information du point de vue de la théorie de martingales". *Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, Prague*, 1957; 183—208.

[25] A. PEREZ: „Sur la théorie de l'information dans le cas d'un alphabet abstrait". *Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes. Prague*, 1957; 209—243.

[26] A. PEREZ: „Information, ε-sufficiency and data reduction problems". *Kybernetika (Ceskoslovenské Akademie Ved)* **1** (1965) 297—322.

[27] M. C. Пинскер: „Информация и информационная устойчивость случайных величин и процессов". *Пробл. пер. инф.* 7. Изд. А. Н. СССР., Москва, 1960.

[28] A. PRÉKOPA, A. RÉNYI, K. URBANIK: „О предельнам распределении для сумм независимых случайных величин на бикомпактных топологических группах". *Acta Math. Acad. Sci. Hung.* **7** (1956) 11—16.

[29] RÉNYI ALFRÉD: „Az információelmélet néhány alapvető kérdése". *MTA III. Oszt. Közleményei* **10** (1960) 251—282.

[30] A. RÉNYI: „On measures of entropy and information". *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability. Berkeley*, 1961; 541—561.

[31] A. RÉNYI: „On the amount of information concerning an unknown parameter in a sequence of observations". *MTA Mat. Kut. Int. Közleményei* **9** (1964) 617—624.

[32] M. SAKAGUCHI: *Information Theory and Decision Making*. Statistics Dept., George Washington University, 1964.

[33] C. E. SHANNON and W. WEAVER: *The Mathematical Theory of Communication*. Urbana, 1949.

[34] W. SIERPINSKI: *General Topology*. University of Toronto Press, Toronto, 1956.

[35] K. STROMBERG: „Probabilities on a compact group". *Transactions of the American Math. Soc.* **94** (1960) 295—309.

[36] K. URBANIK: „On the limiting probability distribution on a compact topological group". *Fundamenta Math.* **44** (1957) 243—261.

[37] VINCZE ISTVÁN: „Az információelmélet egy fogalmának értelmezéséről". *Matematikai Lapok* **10** (1959) 255—266.

(Beérkezett: 1966. I. 20.)