

ELOSZLÁSOK ELTÉRÉSÉNEK INFORMÁCIÓ-TÍPUSÚ MÉRTÉKSZÁMAI, II.*

ÍRTA: CSISZÁR IMRE

3. § Az f -eltérés csökkenése közvetett megfigyelés esetén

Legyen $(X, \mathcal{X})$ tetszőleges mérhető tér és μ_1, μ_2 két valószínűségeloszlás $(X, \mathcal{X})$ -en. Ha nem közvetlenül az X tér elemeit figyeljük meg, hanem csak valamilyen függvényüket, esetleg bizonyos megfigyelési hibával is terhelve, akkor szemléletesen nyilvánvaló, hogy a μ_1 és μ_2 által a közvetett megfigyelés kimeneteleinek $(Y, \mathcal{Y})$ terében indukált $\tilde{\mu}_1$ és $\tilde{\mu}_2$ eloszlások nem különbözhetnek egymástól jobban, mint az eredeti eloszlások. Ha tehát az f -eltérés megfelelő mértékszám a eloszlások egymástól való eltérésének, akkor teljesülnie kell a

$$(3.1) \quad \mathcal{I}_f(\tilde{\mu}_1, \tilde{\mu}_2) \leq \mathcal{I}_f(\mu_1, \mu_2)$$

egyenlőtlenségnek, és az egyenlőség fennállása azt kell hogy jelentse, hogy a közvetett megfigyelés valamilyen értelemben elégséges a (μ_1, μ_2) mértékpárra vonatkozólag.

A közvetett megfigyelés legegyszerűbb esete az, amikor x -et korlátozott pontossággal figyeljük meg, ami matematikailag azt jelenti, hogy az $\mathcal{X}$ σ -algebrát valamely $\mathcal{X}_0$ rész- σ -algebrájával helyettesítjük; ez gyakorlatilag sokszor úgy valósul meg, hogy valamilyen $y = T(x)$ statisztikát vizsgálunk, ahol T az $(X, \mathcal{X})$ -nek mérhető leképezése $(Y, \mathcal{Y})$ -ba. Foglalkozni fogok azonban azzal az általánosabb esettel is, amikor a közvetett megfigyelés y eredménye csak sztochasztikusan függ x -től, olyan értelemben, hogy minden $x \in X$ -nek megfelel egy $v(B|x)$ valószínűségeloszlás $(Y, \mathcal{Y})$ -on, ahol $v(B|x)$ minden rögzített $B \in \mathcal{Y}$ esetén $\mathcal{X}$ -mérhető. Az információelméletben szokásos kifejezéssel ilyenkor azt mondjuk, hogy egy (X, v, Y) megfigyelési csatorna van adva, melynek bemeneti tere $(X, \mathcal{X})$, kimeneti tere $(Y, \mathcal{Y})$ és átmenetfüggvénye $v(B|x)$ (vö. pl. PEREZ [25]). Természetesen ha a megfigyelő csatorna zaj nélküli, akkor az $y = T(x)$ esettel van dolgunk.

KULLBACK és LEIBLER [20] jól ismert tétele szerint (lásd még PEREZ [24]) az $y = T(x)$ esetben mindig

$$(3.2) \quad I(\tilde{\mu}_1 \| \tilde{\mu}_2) \leq I(\mu_1 \| \mu_2) \quad (\tilde{\mu}_i = \mu_i T^{-1}; \quad i = 1, 2),$$

és $I(\mu_1 \| \mu_2) < +\infty$ esetén az egyenlőség akkor és csak akkor áll fenn, ha T elégséges statisztika a μ_1, μ_2 mértékpárra vonatkozólag. Lényegében ismeretesnek tekinthető továbbá⁹, hogy a (3.2) egyenlőtlenség akkor is érvényben marad, ha a T statisztika

* A dolgozat első része, az 1. és 2. §, valamint a teljes irodalomjegyzék az MTA III. Oszt. Közlemények 17 (1967) számának 123–149. oldalán jelent meg.

⁹ Tudomásom szerint ebben az általános formában a (3.2) egyenlőtlenség explicite nem található meg az irodalomban, diszkrét kimeneti tér esetére azonban feladatként szerepel KULLBACK [21] könyvében (35. o. 8.37. feladat), egyébként pedig, mint látni fogjuk, a csatornákra vonatkozó (3.2) egyenlőtlenség közvetlen következménye a T statisztikákra vonatkozó (3.2) egyenlőtlenségnek.

kát egy (X, ν, Y) zajos megfigyelési csatornával pótoljuk; ekkor $\tilde{\mu}_i$ értelmezése

$$(3.3) \quad \tilde{\mu}_i(B) = \int \nu(B|x) \mu_i(dx) \quad (B \in \mathcal{Y}; \quad i = 1, 2).$$

A következőkben ennek a tételnek a tetszőleges f -feltérésekre vonatkozó analógját, valamint az egyenlőség feltétele stabilitásának kérdését tárgyalom. Külön hivatkozás nélkül használni fogom a feltételes valószínűségek és várható értékek következő jól ismert tulajdonságait (lásd pl. [11], [23]):

Egy $(\Omega, \mathcal{F}, P)$ valószínűségi mezőn értelmezett véges várható értékű ξ valószínűségi változónak valamely $\mathcal{F}_0 \subset \mathcal{F}$ σ -algebrára vonatkozó $M(\xi|\mathcal{F}_0)$ feltételes várható értékén olyan $\mathcal{F}_0$ -mérhető függvényt értünk, melyre bármely $\mathcal{F}_0$ -beli E halmaz esetén $\int_E M(\xi|\mathcal{F}_0) P(d\omega) = \int_E \xi P(d\omega)$. Az $M(\xi|\mathcal{F}_0)$ két különböző verziója csak nullmértékű halmazon különbözhet egymástól. A $P(A|\mathcal{F}_0) = M(\chi_A|\mathcal{F}_0)$ függvényt (ahol χ_A az $A \in \mathcal{F}$ halmaz indikátorfüggvénye) az A esemény $\mathcal{F}_0$ -ra vonatkozó feltételes valószínűségének nevezzük. Ha $A_1, A_2, \dots$ diszjunkt $\mathcal{F}$ -beli halmazok, akkor 1 valószínűséggel $P(\bigcup_{i=1}^{\infty} A_i|\mathcal{F}_0) = \sum_{i=1}^{\infty} P(A_i|\mathcal{F}_0)$. Bármely $\mathcal{F}_0$ -mérhető szorzó kiemelhető a feltételes várható értékből: ha η $\mathcal{F}_0$ -mérhető, akkor 1 valószínűséggel $M(\eta\xi|\mathcal{F}_0) = \eta M(\xi|\mathcal{F}_0)$.

3.1. definíció (HALMOS és SAVAGE [14]). Az $(X, \mathcal{X})$ mérhető térnek az $(Y, \mathcal{Y})$ mérhető térbe való T mérhető leképezését az $(X, \mathcal{X})$ -en értelmezett $\{\mu_\theta\}_{\theta \in \Theta}$ eloszláscsaládra vonatkozólag elégséges statisztikának nevezzük, ha a $\mu_\theta(A|y) = \mu_\theta(A|T^{-1}y)$ feltételes valószínűségeknek létezik egy közös (θ -tól nem függő) verziója, azaz, ha minden $A \in \mathcal{X}$ esetén van olyan $\pi_A(x)$ $\mathcal{Y}$ -mérhető függvény, melyre

$$(3.4) \quad \int_B \pi_A(y) \mu_\theta T^{-1}(dy) = \int_{T^{-1}B} \pi_A(Tx) \mu_\theta(dx) = \mu_\theta(A \cap T^{-1}B) \quad (B \in \mathcal{Y})$$

Hasonlóképpen egy $\mathcal{X}_0 \subset \mathcal{X}$ σ -algebrát akkor nevezünk elégségesnek $\{\mu_\theta\}_{\theta \in \Theta}$ -ra nézve, ha a $\mu_\theta(A|\mathcal{X}_0)$ ($A \in \mathcal{X}$) feltételes valószínűségeknek létezik közös (θ -tól független) $\pi_A(x)$ verziója:

$$(3.4') \quad \int_B \pi_A(x) \mu_\theta(dx) = \mu_\theta(A \cap B) \quad (B \in \mathcal{Y}, \quad \theta \in \Theta; \quad \pi_A(x) \text{ } \mathcal{X}_0 \text{ mérhető.})$$

Legyen μ_1 és μ_2 két tetszőleges eloszlás $(X, \mathcal{X})$ -en és $\mathcal{X}_0$ az $\mathcal{X}$ tetszőleges rész- σ -algebrája; jelölje $\bar{\mu}_i$ a μ_i $\mathcal{X}_0$ -ra való megszorítását ($i = 1, 2$).

Jelölje $C \in \mathcal{X}_0$ a $\bar{\mu}_1, \bar{\mu}_2$ -re nézve szinguláris komponensének hordozóját ($\bar{\mu}_1 \ll \bar{\mu}_2$ esetén $C = \emptyset$) és értelmezzük a μ_{12} mértéket a következőképpen:

$$(3.5) \quad \mu_{12}(A) = \int_{X-C} \mu_2(A|\mathcal{X}_0) \mu_1(dx) + \mu_1(A \cap C) \quad (A \in \mathcal{X}).$$

Itt $\mu_1(dx)$ helyett természetesen $\bar{\mu}_1(dx)$ is írható, minthogy $\mu_2(A|\mathcal{X}_0)$ definíció szerint $\mathcal{X}_0$ -mérhető. A (3.5) formula által egyértelműen definiált μ_{12} valószínűségi mértéknek a következőkben fontos szerepe lesz. A 3.1 definícióból közvetlenül következik, hogy ha $\mathcal{X}_0$ elégséges σ -algebra a μ_1, μ_2 eloszláspárra nézve, akkor $\mu_{12} = \mu_1$. Könnyen láthatóan a megfordítás is igaz, ugyanis $\mu_{12} = \mu_1$ esetén (3.5)-

ből következik, hogy egy $N_1 \cup N_2$ ($\mu_1(N_1) = \mu_2(N_2) = 0$) alakú halmaz kivételével (ahol $N_2 = C$) mindenütt $\mu_2(A|\mathcal{X}_0) = \mu_1(A|\mathcal{X}_0)$ s ezért a kétféle feltételes valószínűségnek van közös verziója.

Ha a μ_i ($i = 1, 2$) mértékek λ domináló mértéke olyan, hogy még a λ $\mathcal{X}_0$ -ra való $\bar{\lambda}$ megszorítása is σ -véges, akkor jelölje $p_i(x)$, ill. $\bar{p}_i(x)$ a μ_i , ill. $\bar{\mu}_i$ ($i = 1, 2$) mérték λ -ra, ill. $\bar{\lambda}$ -ra vonatkozó sűrűségfüggvényét (Radon—Nikodym-deriváltját). Ekkor (3. 5)-ben vehető $C = \{x: \bar{p}_2(x) = 0\}$, továbbá nyilván $\mu_{12} \ll \lambda$, végül μ_{12} λ -ra vonatkozó sűrűségfüggvénye a következő:

$$(3.6) \quad p_{12}(x) = \begin{cases} \frac{\bar{p}_1(x)}{\bar{p}_2(x)} p_2(x), & \text{ha } \bar{p}_2(x) > 0 \\ p_1(x), & \text{ha } \bar{p}_2(x) = 0. \end{cases}$$

Ezt elég arra az esetre igazolni, amikor λ maga is valószínűségi mérték, mert ha valamely λ domináló mértékre érvényes a (3. 6) előállítás, akkor nyilván minden vele ekvivalensre is, és minden σ -véges mértékhez található vele ekvivalens valószínűségi mérték. Mármost ha λ valószínűségi mérték, a λ szerint várható értéket jelöljük M -mel. Ekkor felhasználva, hogy az $\mathcal{X}_0$ szerinti feltételes várható értékből $\mathcal{X}_0$ -mérhető szorzó kihozható,

$$(3.7) \quad \begin{aligned} \int_A p_{12}(x) \lambda(dx) &= M \left(\frac{\bar{p}_1}{\bar{p}_2} \chi_{A-C} p_2 \right) + M(p_1 \chi_{A \cap C}) = \\ &= M \left(\frac{\bar{p}_1}{\bar{p}_2} \chi_{X-C} M(\chi_A p_2 | \mathcal{X}_0) \right) + \mu_1(A \cap C) = \\ &= M \left(\frac{\bar{p}_1}{\bar{p}_2} \chi_{X-C} \mu_2(A | \mathcal{X}_0) \bar{p}_2 \right) + \mu_1(A \cap C) = \\ &= \int_{X-C} \mu_2(A | \mathcal{X}_0) \bar{p}_1(x) \lambda(dx) + \mu_1(A \cap C) = \mu_{12}(A), \end{aligned}$$

azaz valóban $p_{12}(x) = \frac{\mu_{12}(dx)}{\lambda(dx)}$.

3. 1. LEMMA. Az $\mathcal{X}_0 \subset \mathcal{X}$ σ -algebra akkor és csak akkor elégséges a μ_1, μ_2 mértékpárra vonatkozólag, ha $\frac{p_1(x)}{p_2(x)}$ ($\mu_1 + \mu_2$)-majdnem mérhető $\mathcal{X}_0$ -ra vonatkozólag, azaz, ha ($\mu_1 + \mu_2$)-majdnem mindenütt értelmezve van és egyenlő egy $\mathcal{X}_0$ -mérhető függvényvel; itt megállapodásszerűen $\frac{a}{0} = +\infty$, ha $a > 0$, de $\frac{0}{0}$ nincs értelmezve.

Bizonyítás. Ha $\mathcal{X}_0$ elégséges, vagyis, ha $\mu_{12} = \mu_1$, akkor

$$(3.8) \quad \frac{p_1(x)}{p_2(x)} = \frac{\bar{p}_1(x)}{\bar{p}_2(x)} \quad [\mu_1 + \mu_2]$$

ahol a jobb oldal definíció szerint $\mathcal{X}_0$ -mérhető. Valóban, (3. 6)-ból következik, hogy ha $p_{12}(x) = p_1(x)$ [λ] akkor $\bar{p}_2(y) > 0$ esetén, vagyis az $X - C$ halmazon ($\mu_2(C) = 0$) mindenesetre fennáll (3. 8); de ugyanez igaz a $C = \{\bar{p}_2(x) = 0\}$ halmazon

is (mindkét oldalon $+\infty$ -nel), mert $\bar{p}_2(x)=0$ esetén még inkább $p_2(x)=0$ [λ] és ugyanígy $p_1(x)>0$ -ból következik, hogy $\bar{p}_1(x)>0$ [λ]. Megfordítva, ha $\frac{p_1(x)}{p_2(x)} (\mu_1 + \mu_2)$ -majdnem mérhető $\mathcal{X}_0$ -ra nézve, akkor választhatjuk $p_1(x)$ -nek és $p_2(x)$ -nek olyan verzióját, hogy az $N = \{x: p_1(x) = p_2(x) = 0\} \in \mathcal{X}$ halmaz komplementerén értelmezett $\frac{p_1(x)}{p_2(x)}$ függvényt alkalmas módon kiterjesztve az egész X -re, $\mathcal{X}_0$ -mérhető függvényt kapunk. Így speciálisan az

$$N'_1 = \left\{ x: \frac{p_1(x)}{p_2(x)} = 0, x \in X - N \right\} \quad \text{és} \quad N'_2 = \left\{ x: \frac{p_1(x)}{p_2(x)} = +\infty, x \in X - N \right\}$$

halmazhoz létezik olyan $N''_1 \subset N$ és $N''_2 \subset N$ halmaz, hogy $N'_1 \cup N''_1 \in \mathcal{X}_0$ és $N'_2 \cup N''_2 \in \mathcal{X}_0$. Ha mármost N_0 -val jelöljük az N halmaz $\mathcal{X}_0$ -mérhető magját a $\bar{\lambda}$ mérték szerint, vagyis, ha $N_0 \in \mathcal{X}_0$ olyan halmaz, hogy $N - N_0$ minden $\mathcal{X}_0$ -beli részhalmaza nulla $\bar{\lambda}$ -mértékű ($\bar{\lambda}$ σ -végeessége miatt ilyen N_0 létezik), akkor a konstrukcióból nyilvánvaló, hogy $\bar{p}_i(x)$ alkalmas verziójára $N_i = \{x: \bar{p}_i(x) = 0\} = N'_i \cup N''_i \cup N_0$ ($i = 1, 2$). Ez azt jelenti, hogy az $X - C = X - N_2 \in \mathcal{X}_0$ halmazon $\mu_1 \ll \mu_2$, továbbá itt $\frac{\mu_1(dx)}{\mu_2(dx)}$ -nek van $\mathcal{X}_0$ -mérhető verziója, amely $x \notin N$ esetén $\frac{p_1(x)}{p_2(x)}$ -szel egyenlő, végül $X - C$ -n $\frac{\bar{\mu}_1(dx)}{\bar{\mu}_2(dx)} = \frac{\bar{p}_1(x)}{\bar{p}_2(x)}$ (μ_1 persze nem szükségképpen valószínűségi mérték $X - C$ -n). Ebből, kihasználva, hogy tetszőleges μ valószínűségi mértékre és $\nu \ll \mu$ véges mértékre — a feltételes várható érték definíciója szerint —

$$(3.9) \quad M_\mu \left(\frac{\nu(dx)}{\mu(dx)} \middle| \mathcal{X}_0 \right) = \frac{\bar{\nu}(dx)}{\bar{\mu}(dx)},$$

azonnal következik, hogy $X - C$ -n $\frac{p_1(x)}{p_2(x)} = \frac{\bar{p}_1(x)}{\bar{p}_2(x)}$ [μ_2]. Ez azonban (3. 6) szerint (figyelembe véve még, hogy $X - C$ -n $p_2(x)$ csak $p_1(x)$ -szel egyidejűleg tűnhet el, ti. ha $x \in N$) éppen azt jelenti, hogy $\mu_{12} = \mu_1$, azaz $\mathcal{X}_0$ elégséges, ami bizonyítandó volt.

MEGJEGYZÉS. A 3.1 lemma valójában nem más, mint az elégséges statisztikákra vonatkozó jólismert faktorizációs kritérium egy változata, a következők szempontjából célszerű megfogalmazásban.

3.1. TÉTEL. Ha μ_1 és μ_2 két tetszőleges valószínűségi mérték az $(X, \mathcal{X})$ mérhető téren és $\mathcal{X}_0$ az $\mathcal{X}$ tetszőleges rész- σ -algebrája, akkor bármely f konvex függvényre

$$(3.10) \quad \mathcal{J}_{f, \mathcal{X}_0}(\mu_1, \mu_2) \leq \mathcal{J}_f(\mu_1, \mu_2).$$

Ha f szigorúan konvex, és $\mathcal{J}_f(\mu_1, \mu_2) < +\infty$, (3.10)-ben akkor és csak akkor van egyenlőség, ha $\mathcal{X}_0$ elégséges σ -algebra a μ_1, μ_2 mértékpárra nézve.

Bizonyítás. Maga a (3. 10) egyenlőtlenség az 1. 1 tétel triviális következménye. Az egyenlőség feltételének meghatározásához azonban célszerű közvetlenül az 1. 1 definícióból kiindulni.

Ha $\mu_1 \ll \mu_2$, akkor az 1. 1 lemmához fűzött megjegyzés és (3. 9) felhasználásával $\mathcal{I}_f(\mu_1, \mu_2) < +\infty$ esetén

$$\begin{aligned} \mathcal{I}_f(\mu_1, \mu_2) &= Mf\left(\frac{p_1}{p_2}\right) = Mf\left(\frac{\mu_1(dx)}{\mu_2(dx)}\right) = \\ (3. 11) \quad &= M\left(M\left(f\left(\frac{\mu_1(dx)}{\mu_2(dx)}\right) \middle| \mathcal{X}_0\right)\right) \cong Mf\left(M\left(\frac{\mu_1(dx)}{\mu_2(dx)} \middle| \mathcal{X}_0\right)\right) = \\ &= Mf\left(\frac{\bar{p}_1}{\bar{p}_2}\right) = \mathcal{I}_f(\bar{\mu}_1, \bar{\mu}_2) = \mathcal{I}_{f; \mathcal{X}_0}(\mu_1, \mu_2) \end{aligned}$$

(a várható érték ebben a bizonyításban mindenütt a μ_2 mértékre vonatkozik), ahol az egyenlőség — feltéve, hogy f szigorúan konvex — akkor és csak akkor áll fenn, ha $\frac{p_1(x)}{p_2(x)}$ μ_2 -majdnem mérhető $\mathcal{X}_0$ -ra vonatkozólag, ami a 3. 1 lemma értelmében éppen azzal ekvivalens, hogy $\mathcal{X}_0$ elégséges σ -algebra μ_1, μ_2 -re nézve.

Az általános esetben jelölje μ'_1 , ill. μ''_1 a μ_1 -nek μ_2 -re abszolút folytonos, ill. szinguláris komponensét, $p'_1(x)$, ill. $p''_1(x)$ pedig ezek sűrűségfüggvényét (a λ domináló mértékre vonatkozólag). Ekkor $p_2(x) > 0$ esetén $p'_1(x) = p_1(x)$, $p_2(x) = 0$ esetén $p'_1(x) = 0$, továbbá $p'_1(x) + p''_1(x) = p_1(x)$. μ'_1 és μ''_1 $\mathcal{X}_0$ -ra való megszorítását, továbbá ezek sűrűségfüggvényét (feltesszük, hogy λ -nak $\mathcal{X}_0$ -ra való $\bar{\lambda}$ megszorítása is σ -véges) $\bar{\mu}'_1, \bar{\mu}''_1, \bar{p}'_1(x), \bar{p}''_1(x)$ -szel jelölve,

$$\frac{\mu'_1(dx)}{\mu_2(dx)} = \frac{p'_1(x)}{p_2(x)} \quad [\mu_2], \quad \frac{\bar{\mu}'_1(dx)}{\bar{\mu}_2(dx)} = \frac{\bar{p}'_1(x)}{\bar{p}_2(x)} \quad [\bar{\mu}_2]$$

és így (3. 11)-hez hasonlóan

$$\begin{aligned} \int p_2(x) f\left(\frac{p'_1(x)}{p_2(x)}\right) \lambda(dx) &= Mf\left(\frac{p'_1}{p_2}\right) \cong Mf\left(\frac{\bar{p}'_1}{\bar{p}_2}\right) = \\ (3. 11') \quad &= \int \bar{p}_2(x) f\left(\frac{\bar{p}'_1(x)}{\bar{p}_2(x)}\right) \lambda(dx), \end{aligned}$$

ahol az egyenlőség akkor és csak akkor áll fenn, ha

$$\frac{p'_1(x)}{p_2(x)} = \frac{\bar{p}'_1(x)}{\bar{p}_2(x)} \quad [\mu_2],$$

vagyis ha

$$(3. 12) \quad \frac{p_1(x)}{p_2(x)} = \frac{\bar{p}'_1(x)}{\bar{p}_2(x)} \quad \text{ha} \quad p_2(x) > 0 \quad [\lambda].$$

Mármost (3. 11')-ből, felhasználva az (1. 8) megállapodást és az (1. 2) lemma (1. 11) speciális esetét ($a_1 = \bar{p}_1'(x)$, $b_1 = 0$, $a_2 = \bar{p}_1'(x)$, $b_2 = \bar{p}_2(x)$ választással):

$$\begin{aligned}
 \mathcal{J}_f(\mu_1, \mu_2) &= \int_{p_2(x)=0} p_2(x) f\left(\frac{p_1(x)}{p_2(x)}\right) \lambda(dx) + \int_{p_2(x)>0} p_2(x) f\left(\frac{p_1(x)}{p_2(x)}\right) \lambda(dx) = \\
 &= \mu_1''(X) \lim_{u \rightarrow \infty} \frac{f(u)}{u} + \int p_2(x) f\left(\frac{p_1'(x)}{p_2(x)}\right) \lambda(dx) \cong \\
 (3.13) \quad &\cong \int 0 f\left(\frac{\bar{p}_1'(x)}{0}\right) \bar{\lambda}(dx) + \int \bar{p}_2(x) f\left(\frac{\bar{p}_1'(x)}{\bar{p}_2(x)}\right) \bar{\lambda}(dx) \cong \\
 &\cong \int \bar{p}_2(x) f\left(\frac{\bar{p}_1''(x) + \bar{p}_1'(x)}{\bar{p}_2(x)}\right) \lambda(dx) = \mathcal{J}_f(\bar{\mu}_1, \bar{\mu}_2) = \mathcal{J}_{f; \mathcal{X}_0}(\mu_1, \mu_2).
 \end{aligned}$$

Itt a második egyenlőtlenségben akkor és csak akkor van egyenlőség, ha $\bar{p}_2(x) > 0$ esetén $\bar{p}_1'(x) = 0$ [λ]; ez a feltétel egyrészt azt jelenti, hogy (3. 12)-ben $\bar{p}_1'(x)$ helyére $\bar{p}_1(x)$ írható, másrészt — minthogy $p_1(x) > p_2(x) = 0$ esetén $p_1''(x) = p_1'(x) > 0$, és így még inkább $\bar{p}_1''(x) > 0$ [λ] — azt is jelenti, hogy ha $p_1(x) > p_2(x) = 0$, akkor mindig $\bar{p}_2(x) = 0$ [λ], tehát $p_1(x) > p_2(x) = 0$ esetén is λ -majdnem mindenütt $\frac{p_1(x)}{p_2(x)} = \frac{\bar{p}_1(x)}{\bar{p}_2(x)}$ ($= +\infty$) [λ]. Ezt (3. 13)-mal egybevetve éppen (3. 8)-ra jutunk, tehát arra, hogy $\mathcal{X}_0$ elégséges. Megfordítva, ha $\mathcal{X}_0$ elégséges, akkor (3. 8) értelmében (3. 10)-ben biztosan egyenlőség van. Ezzel a 3. 1 tétel bizonyítását befejeztük.

KOROLLÁRIUM. Ha $y = T(x)$ ($X, \mathcal{X}$)-nek tetszőleges mérhető leképezése ($Y, \mathcal{Y}$)-ba, akkor — a $\tilde{\mu}_i = \mu_i T^{-1}$ jelöléssel —

$$(3.10a) \quad \mathcal{J}_f(\tilde{\mu}_1, \tilde{\mu}_2) \leq \mathcal{J}_f(\mu_1, \mu_2);$$

ha f szigorúan konvex és $\mathcal{J}_f(\mu_1, \mu_2) < +\infty$, (3. 10a)-ban akkor és csak akkor teljesül az egyenlőség, ha T elégséges statisztika μ_1, μ_2 -re nézve.

Bizonyítás. Akár az 1. 1 tételből, akár magából az 1. 1 definícióból nyilvánvaló, hogy $\mathcal{J}_f(\tilde{\mu}_1, \tilde{\mu}_2) = \mathcal{J}_{f; T^{-1}\mathcal{Y}}(\mu_1, \mu_2)$. Mivel a 3. 1 definíció értelmében T akkor és csak akkor elégséges statisztika a μ_1, μ_2 mértékpárra nézve, ha $T^{-1}\mathcal{Y}$ elégséges σ -algebra, a korollárium valóban következik a 3. 1 tételből.

A tetszőleges megfigyelési csatornákra vonatkozó analóg eredmény megfogalmazásához először is jegyezzük meg, hogy az ($X, \mathcal{X}$) bemeneti téren értelmezett bármely μ (σ -véges) mérték indukál az ($X \times Y, \mathcal{X} \times \mathcal{Y}$) szorzattéren egy μ^* (σ -véges) mértéket:

$$(3.14) \quad \mu^*(A \times B) = \int_A v(B|x) \mu(dx) \quad (A \in \mathcal{X}, \quad B \in \mathcal{Y})$$

és innen az egész $\mathcal{X} \times \mathcal{Y}$ -ra a μ^* mérték a szokásos kiterjesztéssel adódik; μ^* -nak az ($Y, \mathcal{Y}$)-ra való vetületét jelöljük $\tilde{\mu}$ -mal:

$$(3.15) \quad \tilde{\mu}(B) = \int v(B|x) \mu(dx) \quad (B \in \mathcal{Y}).$$

Célszerű továbbá az elégséges statisztika mintájára bevezetni az elégséges csatorna fogalmát.

3.2. definíció. Az (X, ν, Y) csatornát *elégségesnek* nevezzük az $(X, \mathcal{X})$ bemeneti téren értelmezett $\{\mu_\theta\}_{\theta \in \Theta}$ eloszláscsaládra nézve, ha a $\mu_\theta(A|y)$ ($A \in \mathcal{X}$) feltételes valószínűségeknek létezik θ -tól nem függő közös verziója, vagyis ha minden $A \in \mathcal{X}$ esetén van olyan $\pi_A(y)$ $\mathcal{Y}$ -mérhető függvény, melyre

$$(3.16) \quad \int_B \pi_A(y) \tilde{\mu}_\theta(dy) = \mu_\theta^*(A \times B) \quad (B \in \mathcal{Y}, \quad \theta \in \Theta).$$

Ez szemléletesen azt jelenti, hogy a közvetett megfigyelés (az y kimeneti jel megfigyelése) ugyanannyi információt nyújt az ismeretlen $\theta \in \Theta$ paraméterre vonatkozólag, mintha közvetlenül az x -et figyeltük volna meg; más szóval, ha a megfigyelési csatorna zajos is (pl. a megfigyelés csak valamilyen hibával történhetik), a zaj okozta információvesztés csak az ismeretlen paraméter meghatározása szempontjából lényegtelen információkat érinti.

Az elégségesség fogalmát BLACKWELL [3] még ennél is általánosabban, két tetszőleges kísérletre értelmezte. Eszerint — az általam használt jelölésekkel — tetszőleges $(Y, \mathcal{Y})$ kísérletet elégségesnek nevezünk az $(X, \mathcal{X})$ kísérletre nézve, a $\theta \in \Theta$ ismeretlen paraméterekre vonatkozólag, ha minden $\theta \in \Theta$ -nak megfelel az $(X \times Y, \mathcal{X} \times \mathcal{Y})$ szorzattéren egy μ_θ^* eloszlás (amely azonban nem szükségképpen (3.14) alakú) és ezek mindegyikére teljesül (3.16), alkalmas $\pi_A(y)$ függvényvel. Ez szemléletesen azt jelenti, hogy y megfigyelése legalább annyi információt szolgáltat a θ ismeretlen paraméterre vonatkozólag, mint az x megfigyelése.

Ez az általánosítás azonban már csak részben van összhangban az elégséges szemléletes fogalmával, mert ehhez, legalábbis véleményem szerint, az a kép fűződik, hogy a második kísérlet már eleve nem lehet jobb („informatívabb”) az eredetinel, hanem annak valamilyen értelemben „durvítása”, tehát csak azt kívánhatjuk, hogy ez a durvítás az ismeretlen paraméterre vonatkozólag rendelkezésre álló információt ne csökkentse.

3.2. TÉTEL. Az $(X, \mathcal{X})$ mérhető téren értelmezett bármely μ_1 és μ_2 eloszlásra, tetszőleges (X, ν, Y) megfigyelési csatorna és f konvex függvény esetén

$$(3.17) \quad \mathcal{J}_f(\tilde{\mu}_1, \tilde{\mu}_2) \leq \mathcal{J}_f(\mu_1, \mu_2).$$

Ha f szigorúan konvex és $\mathcal{J}_f(\mu_1, \mu_2) < +\infty$, (3.17)-ben akkor és csak akkor teljesül az egyenlőség, ha az (X, ν, Y) csatorna elégséges a μ_1, μ_2 mértékpárra vonatkozólag.

Bizonyítás. A 3.1 tétel szerint mindenestre

$$(3.18) \quad \mathcal{J}_f(\tilde{\mu}_1, \tilde{\mu}_2) \leq \mathcal{J}_f(\mu_1^*, \mu_2^*)$$

(ahol μ_i^* ($i=1, 2$) (3.14)-nek megfelelően van értelmezve), másrészt ugyancsak a (3.1) tétel szerint

$$(3.19) \quad \mathcal{J}_f(\mu_1, \mu_2) = \mathcal{J}_f(\mu_1^*, \mu_2^*),$$

minthogy az $\mathcal{X}' = \{A \times Y: A \in \mathcal{X}\}$ σ -algebra a csatorna definíciója értelmében biztosan elégséges az $(X \times Y, \mathcal{X} \times \mathcal{Y})$ szorzattéren értelmezett bármely eloszláscsaládra nézve. Természetesen a (3.19) egyenlőség közvetlenül is könnyen követ-

kezik az 1. 1 definícióból, mert μ_i^* -nak a (3. 14) szerinti λ^* -ra vonatkozó $p_i^*(x, y)$ sűrűségfüggvénye (3. 14)-ből következőleg λ^* -majdnem mindenütt $p_i(x)$ -szel egyenlő ($i=1, 2$). (3. 18) és (3. 19) együtt éppen a (3. 17) egyenlőtlenséget szolgáltatja. Az egyenlőség feltétele (szigorúan konvex f függvény és $\mathcal{J}_f(\mu_1^*, \mu_2^*) = \mathcal{J}_f(\mu_1, \mu_2) < +\infty$ esetén) a 3. 1 tétel értelmében az, hogy $\mathcal{Y}' = \{X \times B: B \in \mathcal{Y}\}$ elégséges σ -algebra legyen az $(X \times Y, \mathcal{X} \times \mathcal{Y})$ -on értelmezett μ_1^*, μ_2^* mértékpárra nézve, vagyis, hogy bármely $C \in \mathcal{X} \times \mathcal{Y}$ halmazhoz létezzék olyan $\mathcal{Y}$ -mérhető $\pi_C(y)$ függvény, hogy

$$(3.20) \quad \int_B \pi_C(y) \tilde{\mu}_i(dy) = \int_{X \times B} \pi_C(y) \mu_i^*(dx, dy) = \\ = \mu_i^*(C \cap (X \times B)) \quad (B \in \mathcal{Y}, \quad i = 1, 2).$$

Mivel az $A \times B_1$ ($A \in \mathcal{X}, B_1 \in \mathcal{Y}$) alakú halmazok generálják $\mathcal{X} \times \mathcal{Y}$ -t, (3. 20)-at elég a $C = A \times B_1$ esetre megkövetelni. Ebből nyilvánvaló, hogy az iménti feltétel éppen az (X, ν, Y) csatorna elégségeségével ekvivalens, ui., ha (3. 20) teljesül, akkor $\pi'_A(y) = \pi_{A \times Y}(y)$ eleget tesz (3. 16)-nak (ahol $\Theta = \{1, 2\}$), másrészt pedig, ha valamely $\pi'_A(y)$ függvényre (3. 16) teljesül, akkor $\pi_{A \times B_1}(y) = \pi'_A(y) \chi_{B_1}(y)$ választással (3. 20) is teljesül $C = A \times B_1$ -re (ahol $\chi_{B_1}(y)$ a $B_1 \in \mathcal{Y}$ halmaz indikátorfüggvényét jelöli).

Ezzel a 3. 2 tételt bebizonyítottuk.

A 3. 2 tétellel kapcsolatban felmerül az a kérdés, hogy hogyan jellemezhetők azok a csatornák, melyek valamely mértékpárra vonatkozólag elégségesek. Erre ad választ a következő

3. 2. LEMMA. Az (X, ν, Y) csatorna akkor és csak akkor elégséges az $(X, \mathcal{X})$ bemeneti téren értelmezett μ_1, μ_2 eloszláspárra vonatkozólag, ha Y előállítható

$$(3.21) \quad Y = \bigcup_{0 \leq s \leq +\infty} B_s \quad (B_{s_1} \cap B_{s_2} = \emptyset, \quad s_1 \neq s_2)$$

alakban, ahol a $[0, +\infty]$ intervallum minden $[a, b]$ részintervallumára $\bigcup_{a \leq s \leq b} B_s \in \mathcal{Y}$

és¹⁰ $\frac{p_1(x)}{p_2(x)} = s$ esetén $\nu(B_s|X) = 1$, ha $x \in X - X_0$, ahol $\mu_1(X_0) = \mu_2(X_0) = 0$. Itt a (3. 21) előállításban vehető $B_s = \left\{ y: \frac{\tilde{p}_1(y)}{\tilde{p}_2(y)} = s \right\}$ ($s > 0$) és $B_0 = \{y: \tilde{p}_1(y) = 0\}$.

Bizonyítás. Legyen

$$(3.22) \quad A_k^n = \left\{ x: \frac{k-1}{2^n} p_2(x) \leq p_1(x) < \frac{k}{2^n} p_2(x) \right\} \quad (k=1, 2, \dots) \\ B_k^n = \left\{ y: \frac{k-1}{2^n} \tilde{p}_2(y) \leq \tilde{p}_1(y) < \frac{k}{2^n} \tilde{p}_2(y) \right\} \quad (n=1, 2, \dots)$$

$$(3.23) \quad A_\infty = \{x: p_1(x) > p_2(x) = 0\} \\ B_\infty = \{y: \tilde{p}_1(y) > \tilde{p}_2(y) = 0\}$$

¹⁰ Azzal a megállapodással, hogy $a > 0$ esetén $\frac{a}{0} = +\infty$.

ahol a $p_i(x)$, ill. $\tilde{p}_i(y)$ sűrűségfüggvények ($i=1, 2$) tetszőleges λ , ill. $\tilde{\lambda}$ domináló mértékre vonatkoznak (feltesszük, hogy λ és $\tilde{\lambda}$ a (3. 15) kapcsolatban állnak egymással, bár még erre sem volna szükség). Ha az (X, ν, Y) csatorna elégséges a μ_1, μ_2 mértékpárra nézve, akkor a (3. 22) halmazokra (3. 14) és (3. 16) értelmében

$$(3. 24) \quad \frac{k-1}{2^n} \mu_2^*(A_k^n \times B_l^n) \leq \mu_1^*(A_k^n \times B_l^n) \leq \frac{k}{2^n} \mu_2^*(A_k^n \times B_l^n)$$

és

$$(3. 25) \quad \frac{l-1}{2^n} \mu_2^*(A_k^n \times B_l^n) \leq \mu_1^*(A_k^n \times B_l^n) \leq \frac{l}{2^n} \mu_2^*(A_k^n \times B_l^n),$$

ahol (mind (3. 24)-ben, mind (3. 25)-ben) a második egyenlőtlenségben csak akkor állhat egyenlőség, ha $\mu_i^*(A_k^n \times B_l^n) = 0$ ($i=1, 2$). (3. 24) és (3. 25) miatt $k \neq l$ esetén szükségképpen $\mu_i^*(A_k^n \times B_l^n) = 0$ ($i=1, 2$), s ennek, hasonló megfontolás alapján, $k = +\infty$, ill. $l = +\infty$ esetén is érvényben kell maradnia, ha $A_\infty^n = A_\infty$, $B_\infty^n = B_\infty$. Ez azonban (3. 14) értelmében azt jelenti, hogy $x \in A_k^n$ esetén $\nu(B_k^n | x) = 1$ ($k = 1, 2, \dots, +\infty$; $n=1, 2, \dots$), kivéve esetleg egy mind μ_1 , mind μ_2 szerint 0 mértékű X_0 halmazt. Ha tehát $B_s = \bigcap_{n=1}^{\infty} B_{[2^n s] + 1}^n = \left\{ y: \frac{\tilde{p}_1(y)}{\tilde{p}_2(y)} = s \right\}$ ($0 < s < +\infty$), $B_0 = \{y: \tilde{p}_1(y) = 0\}$ és B_∞ a (3. 23) szerinti, akkor $x \in X - X_0$ esetén valóban $\nu(B_s | x) = 1$. Megfordítva, ha Y előállítható a (3. 21) alakban, akkor $s(y)$ -nal jelölve azt az s értéket, amelyre $y \in B_s$, nyilvánvalóan

$$(3. 26) \quad s(y) = \frac{p_1(x)}{p_2(x)} [\mu_1^* + \mu_2^*].$$

Itt, a (3. 21) előállításra vonatkozó feltevésünk szerint $s(y)$ $\mathcal{Y}$ -mérhető, másrészt pedig, mint a 3. 2 tétel bizonyításában is megjegyeztük, (3. 14)-ből azonnal következik, hogy $p_i^*(x, y) = p_i(x) [\lambda^*]$. (3. 26) tehát azt jelenti, hogy $\frac{p_1^*(x, y)}{p_2^*(x, y)} (\mu_1^* + \mu_2^*)$ -majdnem mindenütt egyenlő egy $\mathcal{Y}'$ -mérhető függvénnyel, ahol $\mathcal{Y}' = \{X \times B: B \in \mathcal{Y}\}$. Ekkor azonban a 3. 1 lemma szerint $\mathcal{Y}'$ elégséges σ -algebra a μ_1^*, μ_2^* mértékpárra nézve, és így (X, ν, Y) elégséges csatorna.

MEGJEGYZÉS. A 3. 2 lemma értelmében, ha az (X, ν, Y) csatornához nem létezik n hosszúságú 0 hibavalószínűségű kód (azaz, ha nem léteznek olyan $x_1, x_2, \dots, x_n \in X$ elemek, hogy bármely két $\nu(\cdot | x_i)$ és $\nu(\cdot | x_k)$ ($i \neq k$) mérték ortogonális egymásra), akkor az (X, ν, Y) csatorna egyetlen olyan μ_1, μ_2 mértékpárra vonatkozólag se lehet elégséges, melyre a $p_1(x)/p_2(x)$ likelihood-hányados minden $X - X_0$ alakú halmazon ($\mu_1(X_0) = \mu_2(X_0) = 0$) legalább n különböző értéket vesz fel. Ha speciálisan még 2 hosszúságú 0 hibavalószínűségű kód sem létezik, akkor a csatorna egyetlen $\mu_1 \neq \mu_2$ mértékpárra vonatkozólag se lehet elégséges; ilyenkor tehát a 3. 2 tétel szerint az $\mathcal{J}_f(\tilde{\mu}_1, \tilde{\mu}_2) = \mathcal{J}_f(\mu_1, \mu_2) < +\infty$ egyenlőség csak $\mu_1 = \mu_2$ esetén állhat fenn (ha f szigorúan konvex).

Vizsgáljuk meg most a 3. 1 és 3. 2 tétellel kapcsolatban az egyenlőség feltétele stabilitásának kérdését, vagyis azt a kérdést, hogy milyen esetben érvényes a (3. 1) egyenlőtlenségben közelítőleg egyenlőség. Ehhez szükség lesz a 2. 1 lemma feltéte-

les várható értékekre vonatkozó megfelelőjére. (A jelölések ugyanazok, mint az 1. 1 és 2. 1 lemmában.)

3. 3. LEMMA. Ha az E Borel-halmaz minden pontja r_0 sugarú környezetének (a_1, a_2) -be eső részében $f''(u) \cong a > 0$, akkor bármely $\mathcal{F}_0 \subset \mathcal{F}$ σ -algebra esetén a

$$(3.27) \quad Mf(\xi) - Mf(M(\xi|\mathcal{F}_0)) \leq \delta$$

egyenlőtlenségből

$$(3.28) \quad M|\xi - M(\xi|\mathcal{F}_0)| \leq \sqrt{\frac{8\delta}{a}} + \int_{M(\xi|\mathcal{F}_0) \notin E} |\xi - M(\xi|\mathcal{F}_0)| P(d\omega)$$

következik, feltéve, hogy $\delta \leq \frac{1}{2}ar_0^2$.

Bizonyítás. A 2. 1 lemma bizonyítása szerint $0 < \varepsilon_1 \leq r_0$ esetén

$$(3.29) \quad f(u) \cong f(u_0) + b(u - u_0) + \frac{a}{2} \varepsilon_1 |u - u_0| \chi_{u_0, \varepsilon_1}(u) \chi_E(u_0),$$

ahol $\chi_{u_0, \varepsilon_1}$ az $(u - \varepsilon_1, u_0 + \varepsilon_1)$ intervallum komplementer halmazának, χ_E pedig az E halmaznak az indikátorfüggvényét jelöli; a b együttható értéke természetesen függ u_0 -tól. Ha (3. 29)-ben ε_1 -et $\sqrt{\frac{2\delta}{a}}$ -nak választjuk, az $u_0 = M(\xi|\mathcal{F}_0)$, $u = \xi$ helyettesítéssel, majd integrálással, (3. 27) figyelembevételével (3. 29)-ből

$$(3.30) \quad \int |\xi - M(\xi|\mathcal{F}_0)| P(d\omega) \leq \sqrt{\frac{2}{a}} \sqrt{\delta}$$

$$|\xi - M(\xi|\mathcal{F}_0)| \leq \sqrt{\frac{2\delta}{a}} \\ M(\xi|\mathcal{F}_0) \in E$$

adódik; ebből, minthogy

$$\int_{|\xi - M(\xi|\mathcal{F}_0)| < \sqrt{\frac{2\delta}{a}}} |\xi - M(\xi|\mathcal{F}_0)| P(d\omega) \leq \sqrt{\frac{2\delta}{a}},$$

éppen a (3. 28) becslést kapjuk.

Abban az esetben, ha $M(\xi|\mathcal{F}_0) \in E$ egy valószínűséggel teljesül, (3. 28)-ban a második tag természetesen eltűnik. Biztosan ez a helyzet pl. akkor, ha mindenütt $f''(u) \cong a > 0$; ez utóbbi esetben azonban még több is igaz, nevezetesen a 2. 1 lemmához fűzött 2. megjegyzés pontos analógiájára

$$(3.31) \quad M(\xi - M(\xi|\mathcal{F}_0))^2 \leq \frac{2}{a} \delta.$$

Említést érdemel továbbá a 3. 3 lemma alábbi következménye:

3. 4. LEMMA. Ha alkalmas $a > 0$ és $c^* > 0$ -ra $f''(u) \cong \min\left(a, \frac{c^*}{|u|}\right)$ ($a_1 < u < a_2$), akkor a (3. 27) egyenlőtlenségből bármely $\delta > 0$ esetén következik, hogy

$$(3.32) \quad M|\xi - M(\xi|\mathcal{F}_0)| \leq 4 \sqrt{\frac{M|\xi|}{c^*}} \delta.$$

Bizonyítás. Legyen

$$(3.33) \quad E_n = \left\{ u: |u| < n \frac{c^*}{2a} \right\} \quad (n=1, 2, \dots),$$

$E_0 = \emptyset$, és jelölje Ω_n az $M(|\xi| | \mathcal{F}_0) \in E_n - E_{n-1}$ eseményt, azaz

$$(3.34) \quad \Omega_n = \{ \omega: M(|\xi| | \mathcal{F}_0) \in E_n - E_{n-1}, \omega \in \Omega \} \quad (n=1, 2, \dots).$$

Ha $P(\Omega_n) > 0$, akkor tekintsük az $(\Omega_n, \mathcal{F}_n, P_n)$ valószínűségi mezőt (ahol $\mathcal{F}_n = \{F: F \in \mathcal{F}, F \subset \Omega_n\}$ és $P_n(F) = \frac{P(F)}{P(\Omega_n)}$, ha $F \in \mathcal{F}_n$); ezen a várható értéket M_n -nel jelölve, nyilvánvalóan

$$(3.35) \quad M_n(\xi | \mathcal{F}_{0,n}) = M(\xi | \mathcal{F}_0) \quad [P_n] \quad (\omega \in \Omega_n)$$

(ahol $\mathcal{F}_{0,n} = \mathcal{F}_0 \cap \mathcal{F}_n$; ξ -nek az Ω_n -re való megszorítását egyszerűség kedvéért nem jelöltem külön), továbbá

$$(3.36) \quad M_n |\xi - M_n(\xi | \mathcal{F}_{0,n})| = \frac{1}{P(\Omega_n)} \int_{\Omega_n} |\xi - M(\xi | \mathcal{F}_0)| P(d\omega).$$

Ezért, minthogy (3.33) szerint az E_n minden pontja $r_n = n \frac{c^*}{2a}$ sugarú környezetének

(a_1, a_2) -be eső részében $f''(u) \geq \frac{a}{2n}$, a 3.3 lemmából következik, $|M(\xi | \mathcal{F}_0)| \leq M(|\xi| | \mathcal{F}_0)$ figyelembevételével, hogy

$$(3.37) \quad \frac{1}{P(\Omega_n)} \int_{\Omega_n} |\xi - M(\xi | \mathcal{F}_0)| P(d\omega) \leq 2 \sqrt{\frac{2\delta_n \cdot 2n}{a}},$$

ha

$$\delta_n \leq \frac{1}{2} \cdot \frac{a}{2n} r_n^2 = \frac{n}{4} \cdot \frac{c^{*2}}{a},$$

ahol

$$(3.38) \quad \delta_n = M_n f(\xi) - M_n f(M_n(\xi | \mathcal{F}_{0,n})) = \frac{1}{P(\Omega_n)} \int_{\Omega_n} (f(\xi) - Mf(M(\xi | \mathcal{F}_0))) P(d\omega).$$

Másképpen viszont (3.33) és (3.34) miatt mindenképpen

$$(3.39) \quad \frac{1}{P(\Omega_n)} \int_{\Omega_n} |\xi - M(\xi | \mathcal{F}_0)| P(d\omega) \leq \frac{1}{P(\Omega_n)} \cdot 2 \int_{\Omega_n} M(|\xi| | \mathcal{F}_0) P(d\omega) \leq n \frac{c^*}{a},$$

ami azt jelenti, hogy (3.37) $\delta_n > \frac{n}{4} \frac{c^{*2}}{a}$ esetén is teljesül. Mármint (3.37)-ből és

(3. 38)-ból a Schwarz-egyenlőtlenség felhasználásával azt kapjuk, hogy

$$(3.40) \quad \begin{aligned} M|\xi - M(\xi|\mathcal{F}_0)| &= \sum_{n=1}^{\infty} \int_{\Omega_n} |\xi - M(\xi|\mathcal{F}_0)| P(dw) \leq \\ &\frac{4}{\sqrt{a}} \sum_{n=1}^{\infty} P(\Omega_n) \sqrt{n\delta_n} \leq \frac{4}{\sqrt{a}} \sqrt{\sum_{n=1}^{\infty} P(\Omega_n)n} \sqrt{\sum_{n=1}^{\infty} P(\Omega_n)\delta_n} \leq \\ &\leq \frac{2\sqrt{2}}{\sqrt{a}} \sqrt{1 + \frac{2a}{c^*} MM(|\xi|\mathcal{F}_0) \cdot \sqrt{\delta}} \leq 4 \sqrt{\left(\frac{M|\xi|}{c^*} + \frac{1}{2a}\right) \delta}. \end{aligned}$$

Ezzel a lemmát bebizonyítottuk.

3. 3. TÉTEL. Legyen μ_1 és μ_2 az $(X, \mathcal{X})$ téren értelmezett két valószínűségi mérték, legyen $\mathcal{X}_0$ az $\mathcal{X}$ valamely rész- σ -algebrája és tegyük fel, hogy az $f(u)$ konvex függvény második deriváltjának bármely véges intervallumban pozitív alsó korlátja van. Ekkor bármely $\varepsilon > 0$ -hoz megadható olyan, csak f -től és ε -tól függő $C_\varepsilon > 0$ és $\delta_\varepsilon > 0$, hogy

$$(3.41) \quad \mathcal{J}_f(\mu_1, \mu_2) - \mathcal{J}_{f; \mathcal{X}_0}(\mu_1, \mu_2) \leq \delta \leq \delta_\varepsilon \quad (\mathcal{J}_f(\mu_1, \mu_2) < +\infty)$$

esetén mindig¹¹

$$(3.42) \quad \mu_2 \left\{ \left\{ x: \left| \frac{p_1(x)}{p_2(x)} - \frac{\bar{p}_1(x)}{\bar{p}_2(x)} \right| \leq C_\varepsilon \sqrt{\delta} \right\} \right\} \geq 1 - \varepsilon.$$

Bizonyítás. A 3.1 tétel bizonyítását követjük. (3. 41) teljesülése esetén (3. 11')-ben mindenesetre

$$(3.43) \quad Mf\left(\frac{p'_1(x)}{p_2(x)}\right) - Mf\left(\frac{\bar{p}'_1(x)}{\bar{p}_2(x)}\right) \leq \delta,$$

ahol M a μ_2 szerinti várható értéket jelöli; ezért a 3. 3 lemmát valamely $E = (0, \gamma)$ intervallumra alkalmazva, nyerjük, hogy

$$(3.44) \quad \int_{\frac{\bar{p}_1(x)}{\bar{p}_2(x)} < \gamma} \left| \frac{p_1(x)}{p_2(x)} - \frac{\bar{p}_1(x)}{\bar{p}_2(x)} \right| p_2(x) \lambda(dx) \leq 2 \sqrt{\frac{2\delta}{a_{\gamma+r_0}}},$$

ha $\delta \leq \frac{a_{\gamma+r_0}}{2} r_0^2$, ahol $a_{\gamma+r_0} > 0$ az $f''(u)$ alsó határa a $(0, \gamma + r_0)$ intervallumban.

(3. 44)-ből természetesen következik, hogy bármely $\Gamma_1 > 0$ esetén

$$(3.45) \quad \mu_2 \left\{ \left\{ x: \left| \frac{p_1(x)}{p_2(x)} - \frac{\bar{p}_1(x)}{\bar{p}_2(x)} \right| > \Gamma_1 \sqrt{\delta}; \frac{\bar{p}_1(x)}{\bar{p}_2(x)} < \gamma \right\} \right\} \leq \frac{2}{\Gamma_1} \sqrt{\frac{2}{a_{\gamma+r_0}}}.$$

¹¹ A (3. 42)-ben szereplő halmaz értelmezése természetesen magában foglalja a $p_2(x)\bar{p}_2(x) > 0$ feltételt is.

Másrészt a (3. 13) egyenlőtlenségből látható, hogy (3. 41) teljesülése esetén

$$(3.46) \quad \mu_2 \left\{ \left\{ x: 0f \left(\frac{\bar{p}_1''(x)}{0} \right) + \bar{p}_2(x)f \left(\frac{\bar{p}_1'(x)}{\bar{p}_2(x)} \right) - \bar{p}_2(x)f \left(\frac{\bar{p}_1(x)}{\bar{p}_2(x)} \right) \right\} \right\} \cong \frac{1}{\Gamma_2},$$

ahol $\Gamma_2 > 0$ tetszőleges. A (3. 46)-ban szereplő halmaz komplementerének azon a részén, ahol $\frac{\bar{p}_1(x)}{\bar{p}_2(x)} \cong \gamma$, a 2. 2 lemma értelmében (a lemmát két 1 mértékű pontból álló E halmazra alkalmazva)

$$(3.47) \quad \bar{p}_1''(x) \cong \sqrt{\frac{8\Gamma_2\delta}{a_{\gamma+r_0}}} \bar{p}_2(x),$$

$$\text{ha} \quad \Gamma_2 \delta \cong \frac{a_{\gamma+r_0}}{2} r_0^2.$$

Minthogy az $A_\gamma = \left\{ x: \frac{\bar{p}_1(x)}{\bar{p}_2(x)} \cong \frac{\bar{p}_1'(x)}{\bar{p}_2(x)} \cong \gamma \right\}$ halmazra nyilvánvalóan $\mu_2(A_\gamma) \cong \frac{1}{\gamma}$, (3. 45), (3. 46) és (3. 47)-ből következik, hogy az

$$A^* = \left\{ x: \left| \frac{p_1(x)}{p_2(x)} - \frac{\bar{p}_1'(x)}{\bar{p}_2(x)} \right| \cong \Gamma_1 \sqrt{\delta}, \frac{\bar{p}_1''(x)}{\bar{p}_2(x)} \cong \sqrt{\frac{8\Gamma_2\delta}{a_{\gamma+r_0}}}; p_2(x)\bar{p}_2(x) > 0 \right\}$$

halmaz μ_2 -mértéke $\mu_2(A^*) \cong 1 - \frac{1}{\gamma} - \frac{2}{\Gamma_1} \sqrt{\frac{2}{a_{\gamma+r_0}}} - \frac{1}{\Gamma_2}$. Ebből azonban a $\gamma = \Gamma_2 = \frac{3}{\varepsilon}$, $\Gamma_1 = \frac{6}{\varepsilon} \sqrt{\frac{2}{a_{\gamma+r_0}}}$ választással következik, hogy a $C_\varepsilon = \Gamma_1 + \sqrt{\frac{8\Gamma_2}{a_{\gamma+r_0}}}$ konstanssal érvényes a (3. 42) becslés, ha $\delta \cong \frac{\varepsilon a_{\gamma+r_0}}{6} r_0^2 = \delta_\varepsilon$ (ahol $r_0 > 0$ még tetszőlegesen választható).

MEGJEGYZÉS. Természetesen, ha a $\bar{\mu}_1$ és $\bar{\mu}_2$ olyan, hogy $(\bar{\mu}_1 + \bar{\mu}_2)$ -majdnem mindenütt $\gamma_1 \cong \frac{\bar{p}_1(x)}{\bar{p}_2(x)} \cong \gamma_2$ és a $(\gamma_1 - r_0, \gamma_2 + r_0) \cap (0, +\infty)$ intervallumban $f''(u) \cong a > 0$, akkor a fenti bizonyításból az is következik, hogy (3. 41) esetén mindig $\int \left| \frac{p_1(x)}{p_2(x)} - \frac{\bar{p}_1(x)}{\bar{p}_2(x)} \right| p_2(x) \lambda(dx) \cong \sqrt{\frac{8\delta}{a}}$, hacsak $\delta \cong \frac{a}{2} r_0^2$. Hasonló élesebb eredmény az f -re vonatkozó erősebb megszorítás esetén a μ_1 és μ_2 -re vonatkozó külön feltevés nélkül is igaz:

3. 4. TÉTEL. Legyen μ_1 és μ_2 az $(X, \mathcal{X})$ mérhető téren értelmezett két valószínűségi mérték, legyen $\mathcal{X}_0$ az $\mathcal{X}$ valamely rész- σ -algebrája és jelölje μ_{12} a (3. 5) képlettel értelmezett eloszlást. Tegyük fel továbbá, hogy az $f(u)$ konvex függvényre $f''(u) \cong \min \left(a, \frac{c^*}{u} \right) > 0$ ($0 < u < +\infty$). Ekkor

$$(3.48) \quad \mathcal{J}_f(\mu_1, \mu_2) - \mathcal{J}_{f; \mathcal{X}_0}(\mu_1, \mu_2) \cong \delta \quad (\mathcal{J}_f(\mu_1, \mu_2) < +\infty)$$

(ahol δ tetszőleges nemnegatív szám) csak úgy lehetséges, ha

$$(3.49) \quad |\mu_1 - \mu_{12}| = \int \left| \frac{p_1(x)}{p_2(x)} - \frac{\bar{p}_1(x)}{\bar{p}_2(x)} \right| p_2(x) \lambda(dx) \leq 4 \sqrt{\frac{2a + c^*}{2ac^*}} \delta.$$

Bizonyítás. Az f függvényre vonatkozó feltevésünkből következik, hogy $\lim_{u \rightarrow \infty} \frac{f(u)}{u} = +\infty$, ezért $\mathcal{J}_f(\mu_1, \mu_2)$ csak $\mu_1 \ll \mu_2$ esetén lehet véges. Ebben az esetben $|\mu_1 - \mu_{12}|$ (3. 6) szerint valóban a (3. 49)-beli integrállal egyenlő.

Az $(X, \mathcal{X}, \mu_2)$ valószínűségi mezőn értelmezett $\xi(x) = \frac{p_1(x)}{p_2(x)} = \frac{\mu_1(dx)}{\mu_2(dx)}$ valószínűségi változóra és az $\mathcal{X}_0$ σ -algebrára alkalmazhatjuk a 3. 4 lemmát, minthogy (3. 48) éppen (3. 27)-nek felel meg, lévén (3. 9) értelmében $M(\xi|\mathcal{X}_0) = \frac{\bar{\mu}_1(dx)}{\bar{\mu}_2(dx)} = \frac{\bar{p}_1(x)}{\bar{p}_2(x)}$. (3. 32) szerint — figyelembe véve (3. 6)-ot is — nyerjük, hogy

$$(3.50) \quad \begin{aligned} M|\xi - M(\xi|\mathcal{X}_0)| &= \int \left| \frac{p_1(x)}{p_2(x)} - \frac{\bar{p}_1(x)}{\bar{p}_2(x)} \right| p_2(x) \lambda(dx) = \\ &= |\mu_1 - \mu_{12}| \leq 4 \sqrt{\left(\frac{M|\xi|}{c^*} + \frac{1}{2a} \right)} \delta = 4 \sqrt{\frac{2a + c^*}{2ac^*}} \delta, \end{aligned}$$

ami bizonyítandó volt.

KOROLLÁRIUM. Ha a 3. 3, ill. 3. 4 tétel feltevései mellett valamely $y = T(x)$ statisztikára

$$(3.41a) \quad \mathcal{J}_f(\mu_1, \mu_2) - \mathcal{J}_f(\mu_1 T^{-1}, \mu_2 T^{-1}) \leq \delta,$$

akkor az $\mathcal{X}_0 = T^{-1}\mathcal{Y}$ σ -algebrát tekintve teljesülnie kell (3. 42)-nek, ill. (3. 49)-nek, az előbbi esetben feltételezve még azt is, hogy $\delta \leq \delta_e$.

A 3. 4 tétel bizonyításánál az $f''(u) \geq \min \left(a, \frac{c^*}{u} \right) > 0$ feltételt lényegesen kihasználtuk és enélkül általában nem is igaz az, hogy $\mathcal{J}_f(\mu_1, \mu_2) - \mathcal{J}_{f; \mathcal{X}_0}(\mu_1, \mu_2)$ csak úgy lehet kicsi, ha $|\mu_1 - \mu_{12}|$ is kicsi. Valóban, tekintsük pl. az $X = \{-1, 0, 1\}$ téren, ahol $\mathcal{X}$ az X összes részhalmazaiból áll, a $\mu_1 = (0, 0, 1)$ és $\mu_2 = \left(\frac{1}{2N}, 1 - \frac{1}{N}, 1 - \frac{1}{N} \right)$ eloszlást (tehát a $\lambda(\{x\}) = 1$, $x = 0, \pm 1$ által megadott domináló mértékre vonatkozó sűrűségfüggvények $p_1(-1) = p_1(0) = 0$, $p_1(1) = 1$, $p_2(-1) = p_2(1) = \frac{1}{2N}$, $p_2(0) = 1 - \frac{1}{N}$), továbbá a $T(x) = x^2$ statisztikát, illetőleg az ennek megfelelő $\mathcal{X}_0 = \{\emptyset, \{0\}, \{-1, 1\}, X\}$ σ -algebrát. Ekkor $\bar{p}_1(0) = 0$, $\bar{p}_1(-1) = \bar{p}_1(1) = \frac{1}{2}$ és $\bar{p}_2(x) = p_2(x)$ ($x = 0, \pm 1$), tehát $\mathcal{J}_f(\mu_1, \mu_2) - \mathcal{J}_{f; \mathcal{X}_0}(\mu_1, \mu_2) = \frac{1}{2N} (f(0) + f(2N) - 2f(N))$, másrészt pedig (3. 6) alapján $|\mu_1 - \mu_{12}| = \sum_{i=-1}^{+1} |p_1(i) - \bar{p}_1(i)| = 1$. Így pl. $\mathcal{J}_{-u^2}(\mu_1, \mu_2) -$

$-\mathcal{J}_{-u^*}; x_0(\mu_1, \mu_2)$ ($0 < \alpha < 1$) tetszőlegesen kicsi lehet, annak ellenére, hogy $|\mu_1 - \mu_{12}| = 1$.

Hasonlóan egyszerű ellenpéldával igazolható, hogy ha még az $f''(u) \geq a, > 0$ ($0 < u < \gamma$) feltétel se teljesül, akkor általában nem marad érvényben a 3.3 tétel állítása sem (vö. [6]).

A 3.3, ill. 3.4 tételből könnyen megkaphatjuk a tetszőleges (X, ν, Y) megfigyelési csatorna esetére vonatkozó analóg eredményt is.

Legyen μ_1 és μ_2 két tetszőleges eloszlás $(X, \mathcal{X})$ -en, jelölje λ egy σ -véges domináló mértéküket. Tekintsük az $(X \times Y, \mathcal{X} \times \mathcal{Y})$ szorzattéren a (3.14) szerint értelmezett $\mu_1^*, \mu_2^*, \lambda^*$ mértékeket és ezek (3.15) szerinti $\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\lambda}$ vetületét $(Y, \mathcal{Y})$ -ra; λ -t úgy választjuk, hogy $\tilde{\lambda}$ is σ -véges legyen. A megfelelő sűrűségfüggvényeket jelölje $p_i(x), p_i^*(x, y), \tilde{p}_i(y)$ ($i=1, 2$); mint már megjegyeztük, $p_i^*(x, y) = p_i(x)$ [λ] ($i=1, 2$). Az $(X \times Y, \mathcal{X} \times \mathcal{Y})$ szorzattéren értelmezzük a (3.5)-nek megfelelő μ_{12}^* mértéket, ahol $\mathcal{X}_0$ szerepét most az $X \times B$ ($B \in \mathcal{Y}$) alakú halmazok alkotta $\mathcal{Y}'$ σ -algebra játssza; (3.6) értelmében μ_{12}^* λ^* -ra vonatkozó sűrűségfüggvénye

$$(3.51) \quad p_{12}^*(x, y) = \begin{cases} \tilde{p}_1(y) p_2(x) & \text{ha } \tilde{p}_2(y) > 0 \\ \tilde{p}_2(y) & \text{ha } \tilde{p}_2(y) = 0 \\ p_1(x) & \end{cases}$$

(tekintettel arra, hogy $p_i^*(x, y) = p_i(x)$ [λ^*] ($i=1, 2$)). A $\mu_1 \ll \mu_2, \mu_1^* \ll \mu_2^*$ esetben természetesen $p_1(x) = p_1^*(x, y) = 0$, ha $\tilde{p}_2(y) = 0$ [λ^*].

A 3.3, ill. 3.4 tételt az $(X \times Y, \mathcal{X} \times \mathcal{Y})$ szorzattéren értelmezett μ_1^* és μ_2^* mértékekre és az $\mathcal{Y}'$ σ -algebrára alkalmazva, figyelembevéve (3.19)-et, a következő eredményt kapjuk:

3.5. TÉTEL. *Ha az f konvex függvény második deriváltjának minden véges intervallumban pozitív alsó korlátja van, akkor az*

$$(3.52) \quad \mathcal{J}_f(\mu_1, \mu_2) - \mathcal{J}_f(\tilde{\mu}_1, \tilde{\mu}_2) \leq \delta \quad (\mathcal{J}_f(\mu_1, \mu_2) < +\infty)$$

egyenlőtlenségből következik, hogy

$$(3.53) \quad \mu_2^* \left\{ \left\{ (x, y) : \left| \frac{p_1(x)}{p_2(x)} - \frac{\tilde{p}_1(y)}{\tilde{p}_2(y)} \right| \leq C_\varepsilon \sqrt{\delta} \right\} \right\} \geq 1 - \varepsilon,$$

ha δ elég kicsi ($\delta \leq \delta_\varepsilon$). Ha az f függvényre még $f''(u) \geq \min \left(a, \frac{c^}{u} \right) > 0$ ($0 < u < \infty$) is teljesül, akkor (3.52)-ből minden esetben*

$$(3.54) \quad |\mu_1^* - \mu_{12}^*| = \int_{p_2(x)\tilde{p}_2(y) > 0} \left| \frac{p_1(x)}{p_2(x)} - \frac{\tilde{p}_1(y)}{\tilde{p}_2(y)} \right| p_2(x) \lambda^*(dx, dy) \geq 4 \sqrt{\frac{2a + c^*}{2ac^*}} \delta$$

következik.

Természetesen a 3.5 tétel speciális esetként — $(Y, \mathcal{Y}) = (X, \mathcal{X}); \nu(B|x) = \chi_B(x)$ tartalmazza a 3.3 és 3.4 tételt. A 3.3 tételhez fűzött megjegyzés most is érvényes. A $\gamma_1 \leq \frac{\tilde{p}_1(y)}{\tilde{p}_2(y)} \leq \gamma_2$ feltétel biztosan teljesül pl. akkor, ha minden $x_1, x_2 \in X$ és $B \in \mathcal{Y}$ esetén $\gamma_1 \leq \frac{\nu(B|x_1)}{\nu(B|x_2)} \leq \gamma_2$.

MEGJEGYZÉS. Az f -eltérés „kis” csökkenésének szükséges feltételével először [6] dolgozatomban foglalkoztam, ahol bebizonyítottam a 3. 5 tétel első felét, amely ekvivalens a 3. 3 tétellel. A 3. 4 tétel (és a 3. 5 tétel második fele) ennek az eredménynek az élesítése az f függvényre vonatkozó erősebb megszorítás mellett. Az I -divergencia speciális esetére a 3. 4 tételhez hasonló éles becslést explicite először PEREZ [26] adott, bár ezt burkolt formában már PINSZKER [27] munkája is tartalmazza. Az I -divergencia esetére a bizonyítás egyszerűen azon az észrevételen alapszik, hogy — az általam használt jelölésekkel —

$$(3.55) \quad I(\mu_1 \| \mu_2) = I_{\mathcal{X}_0}(\mu_1 \| \mu_2) + I(\mu_1 \| \mu_{12})$$

és így közvetlenül alkalmazható egy (2. 20) típusú becslés; megjegyzendő, hogy burkoltan már a Kullback—Leibler-tétel eredeti bizonyítása is a (3. 55) összefüggésen alapult. Hasonló összefüggés azonban tetszőleges f -eltérésekre nem érvényes (valójában az $I(\mu_1 \| \mu_{12})$ tag nem más, mint a μ_1 és μ_2 mértékek átlagos feltételes I -divergenciája, tehát (3. 55) az információ additivitását fejezi ki), ezért az általános esetre vonatkozó bizonyításhoz új gondolatra volt szükség.

PEREZ [26] dolgozatában (301., majd 305. o.) bevezette az ε -elégesség fogalmát, az $\mathcal{X}_0 \subset \mathcal{X}$ σ -algebra ε -elégességén azt értve, hogy az $I_{\mathcal{X}_0}(\mu_1 \| \mu_2)$ I -divergencia ε -nál kevesebbel különbözik $I(\mu_1 \| \mu_2)$ -től. Megjegyzi, hogy itt a közösleges I -divergencia helyett a RÉNYI-féle α -rendű I -divergencia vagy általánosabban az általam vizsgált f -eltérések fogalmából is ki lehetne indulni. Természetes általánosításként persze beszélhetünk tetszőleges megfigyelési csatornák ε -elégességéről is. A PEREZ által tárgyalt — Bayes-féle döntési eljárásokra vonatkozó — alkalmazások elsősorban a $|\mu_1 - \mu_{12}|$ variációs távolság becslésére támaszkodnak, így a 3. 4 tétel értelmében ebből a szempontból az $f''(u) \equiv \min \left(a, \frac{c^*}{u} \right) > 0$ ($0 < u < \infty$) feltételnek eleget tevő f -eltérések tekinthetők „jóknak”.

A 3. 3—3. 5 tételeket az α -rendű I -divergencia esetére alkalmazva (figyelembe véve (1. 20)-at), megállapíthatjuk, hogy $1 \leq \alpha \leq 2$ esetén alkalmas C konstanssal érvényes a

$$(3.56) \quad |\mu_1 - \mu_{12}| \leq C \sqrt{I_\alpha(\mu_1 \| \mu_2) - I_{\alpha; \mathcal{X}_0}(\mu_1 \| \mu_2)},$$

illetőleg a

$$(3.57) \quad |\mu_1^* - \mu_{12}^*| = C \sqrt{I_\alpha(\mu_1 \| \mu_2) - I_\alpha(\tilde{\mu}_1 \| \tilde{\mu}_2)}$$

egyenlőtlenség. Ha azonban $0 < \alpha < 1$, akkor ilyen becslés általában már nem érvényes, csak az igaz, hogy az eredeti $\frac{p_1(x)}{p_2(x)}$ és a közvetett megfigyeléshez tartozó

$\frac{\bar{p}_1(x)}{\bar{p}_2(x)}$, ill. $\frac{\tilde{p}_1(x)}{\tilde{p}_2(x)}$ likelihood-hányados különbsége egy kis μ_2 (ill. μ_2^*) mértékű hal-

maztól eltekintve az információ csökkenés négyzetgyökével becsülhető. Végül, ha $\alpha > 2$, akkor általában még ez sem állítható. Persze, ha valamilyen — a μ_1 és μ_2 eloszlásokra és az $\mathcal{X}_0$ σ -algebrára, ill. a megfigyelési csatorna $v(B|x)$ átmenetfügg-

vényére vonatkozó — feltétel biztosítja azt, hogy $\frac{\bar{p}_1(x)}{\bar{p}_2(x)}$, ill. $\frac{\tilde{p}_1(y)}{\tilde{p}_2(y)}$ a $0 < \alpha < 1$ eset-

ben véges felső korláttal, vagy az $\alpha > 2$ esetben pozitív alsó korláttal rendelkezik, akkor (3. 56), ill. (3. 57) minden esetben teljesül.

A 3. fejezet befejezéséül szeretnék rámutatni arra, hogy eredményeim milyen összefüggésben vannak az információelméletnek a matematikai statisztikában a Bayes-féle felfogás alapján való alkalmazásaival (vö. pl. PEREZ [26], RÉNYI [31]). Legyen $\{\mu_\theta\}_{\theta \in \Theta}$ tetszőleges eloszláscsalád az $(X, \mathcal{X})$ téren és a θ paraméternek legyen valamilyen π a priori eloszlása, ahol tehát π valószínűségi mérték a Θ halmaz részhalmazainak egy olyan $\mathcal{T}$ σ -algebráján, amelyre vonatkozólag $\mu_\theta(A)$ minden rögzített $A \in \mathcal{X}$ esetén mérhető. Ekkor a μ_θ eloszláscsalád felfogható egy $(\Theta, \mathcal{T})$ bemeneti terű csatorna átmenetfüggvényének is, és beszélhetünk a θ és x együttes π^* eloszlásáról, amely a $(\Theta \times X, \mathcal{T} \times \mathcal{X})$ szorzattéren van értelmezve, valamint a π által az $(X, \mathcal{X})$ -en indukált $\tilde{\mu} = \mu$ eloszlásról is:

$$(3.58) \quad \pi^*(C \times A) = \int_C \mu_\theta(A) \pi(d\theta) \quad (C \in \mathcal{T}, \quad A \in \mathcal{X}),$$

$$(3.59) \quad \mu(A) = \tilde{\pi}(A) = \int_\Theta \mu_\theta(A) \pi(d\theta) \quad (A \in \mathcal{X}).$$

Ha x közvetlen megfigyelése helyett valamely (X, ν, Y) megfigyelési csatornát használunk, akkor ez az előbbi csatornával sorbakapcsoltnak tekintendő, miáltal egy újabb $(\Theta, \tilde{\mu}_\theta, Y)$ csatornát kapunk, melynek átmenetfüggvényét

$$(3.60) \quad \tilde{\mu}_\theta(B) = \int \nu(B|x) \mu_\theta(dx) \quad (B \in \mathcal{Y})$$

értelmezi. A π -nek megfelelő eloszlás $(\Theta \times Y, \mathcal{T} \times \mathcal{Y})$ -on, ill. $(Y, \mathcal{Y})$ -on

$$(3.61) \quad \tilde{\pi}^*(C \times B) = \int_C \tilde{\mu}_\theta(B) \pi(d\theta) = \int_C \int_X \nu(B|x) \mu_\theta(dx) \pi(d\theta) \quad (B \in \mathcal{Y}, \quad C \in \mathcal{T}),$$

ill.

$$(3.62) \quad \tilde{\mu}(B) = \int_\Theta \tilde{\mu}_\theta(B) \pi(d\theta) = \int_X \nu(B|x) \mu(dx) \quad (B \in \mathcal{Y}).$$

(Minthogy itt egyszerre több csatornáról van szó, a $*$ és $\sim$ jelölések már nem alkalmazhatók olyan magától értetődő következetességgel, mint korábban, de talán még így is elősegítik az áttekintést.)

A Bayes-féle felfogásnak az információelmélet statisztikai alkalmazásai szempontjából az az előnye, hogy lehetővé teszi az ismeretlen paraméterre vonatkozólag egy megfigyelés által átlagosan szolgáltatott információmennyiség értelmezését, a kölcsönös információ (1. 5a) képlete alapján. Esetünkben ez az információmennyiség közvetlen megfigyelés esetén $I(x, \theta) = I(\pi^* \| \pi \times \mu)$, közvetett megfigyeléskor pedig $I(y, \theta) = I(\tilde{\pi}^* \| \pi \times \tilde{\mu})$.

Mármost a 3. 2 tételből ($f(u) = u \log u$ választással) azonnal következik a szemléletes jelentésű

$$(3.63) \quad I(y, \theta) \leq I(x, \theta)$$

egyenlőtlenség; valóban, $\tilde{\pi}^*$ és $\pi \times \tilde{\mu}$ a π^* és $\pi \times \mu$ eloszlásból az immár negyedik $(\Theta \times X, \mathcal{T} \times \mathcal{X}, \Theta \times Y)$ megfigyelési csatorna alkalmazásával kapható meg, ahol

$$\nu(D|\theta, x) = \nu(D_\theta|x) \quad (D \in \mathcal{T} \times \mathcal{X}; \quad D_\theta = \{x: (\theta, x) \in D\}).$$

A (3. 63) egyenlőtlenség természetesen közvetlenül is könnyen következik a három valószínűségi változó információjára vonatkozó összefüggésből (vö. [10], [27]), ha figyelembe vesszük, hogy a θ , x , y valószínűségi változók *Markov*-láncot alkotnak.

Ha (3. 63)-ban a jobb és bal oldal különbsége ε -nál kisebb, akkor PEREZ [26] fogalomalkotásának természetes általánosításával az (X, ν, Y) csatornát ε -elégességnek nevezhetjük a π^* , $\pi \times \mu$ mértékpárra nézve, s a 3. 5 tétel felhasználásával megállapíthatjuk $\left(a \frac{\pi^*(d\theta, dx)}{(\pi \times \mu)(d\theta, dx)}, \text{ ill. } \frac{\tilde{\pi}^*(d\theta, dy)}{(\pi \times \tilde{\mu})(d\theta, dy)} \right)$ sűrűségfüggvényt $p(x, \theta)$ -val, ill. $\tilde{p}(y, \theta)$ -val jelölve), hogy az ilyen értelemben vett ε -elégesség esetén

$$(3. 64) \quad \int |p(x, \theta) - \tilde{p}(y, \theta)| (\pi \times \mu^*)(d\theta, dx, dy) \leq 4\sqrt{\varepsilon}$$

(sőt (3. 55) és (2. 20) felhasználásával itt a jobb oldalon $\sqrt{2\varepsilon}$ is írható).

A (3. 63) egyenlőtlenség még egy másik, szemléletesebb módon is megkapható a 3. 2 tételből, ha feltételezzük, hogy az $\mathcal{X}$ és $\mathcal{Y}$ σ -algebrát az X , ill. Y tér megszámlálható sok részhalmaza generálja. Valóban, az (1. 5a) képlet értelmében $I(x, \theta)$ csak akkor lehet véges, ha $\pi^* \ll \pi \times \mu$. Jelöljük ekkor a $\frac{\pi^*(d\theta, dx)}{(\pi \times \mu)(d\theta, dx)}$ sűrűségfüggvényt $p(x, \theta)$ -val, mikoris (3. 58) értelmében

$$\pi^*(C \times A) = \int_C \mu_\theta(A) \pi(d\theta) = \int_C \int_A p(x, \theta) \mu(dx) \pi(d\theta) \quad (A \in \mathcal{X}, C \in \mathcal{T}).$$

Ez azt jelenti, hogy bármely rögzített $A \in \mathcal{X}$ esetén

$$(3. 65) \quad \mu_\theta(A) = \int_A p(x, \theta) \mu(dx) \quad [\pi].$$

Minthogy feltevésünk szerint $\mathcal{X}$ -et X -nek megszámlálható sok részhalmaza generálja, (3. 65)-ben a kivételes null-mértékű halmaz A -tól függetlennek is választható, amiből következik, hogy $\mu_\theta \ll \mu$ és $\frac{\mu_\theta(dx)}{\mu(dx)} = p(x, \theta)$ $[\pi]$. Így tehát, (1. 5) és (1. 5a) szerint, figyelembe véve (3. 58)-at

$$(3. 66) \quad I(x, \theta) = \int \log p(x, \theta) \pi^*(d\theta, dx) = \int I(\mu_\theta \| \mu) \pi(d\theta).$$

Hasonlóképpen adódik, hogy

$$(3. 67) \quad I(y, \theta) = \int I(\tilde{\mu}_\theta \| \tilde{\mu}) \pi(d\theta).$$

MEGJEGYZÉS. Az $\mathcal{X}$, ill. $\mathcal{Y}$ σ -algebrára vonatkozó szeparabilitási feltétel nélkül a (3. 66), ill. (3. 67) egyenlőség már nem szükségképpen érvényes. Így pl. ha $X = \Theta = [0, 1]$ és $\mathcal{X} = \mathcal{T}$ a véges vagy megszámlálható halmazok és komplementeik által alkotott σ -algebra, továbbá μ_θ az $x = \theta$ pontban koncentrált eloszlás, π pedig az a mérték, melynél a megszámlálható halmazok nulla, komplementeik pedig egy mértékűek, akkor $I(x, \theta) = 0$, ugyanakkor pedig $\mu = \pi$, $\mu_\theta \ll \mu = \pi$, tehát

$$\int I(\mu_\theta \| \mu) \pi(d\theta) = +\infty.$$

A (3. 66) és (3. 67) összefüggés szerint a (3. 63) egyenlőtlenség a 3. 2 tételnek magára az (X, ν, Y) csatornára való alkalmazásával is kiadódik. S $I(x, \theta) < +\infty$ esetén az egyenlőségnek az a feltétele, hogy az (X, ν, Y) csatorna π -majdnem minden μ_θ, μ párra vonatkozólag elégséges legyen. Ez viszont — figyelembe véve (3. 59)-et és azt, hogy $\mu_\theta \ll \mu[\pi]$ — a 3. 2 definíció értelmében azzal ekvivalens, hogy az (X, ν, Y) csatorna elégséges a $\{\mu_\theta\}_{\theta \in \Theta}$ eloszláscsalád valamely $\{\mu_\theta\}_{\theta \in \Theta-N}$ részcsaládjára nézve, ahol $\pi(N)=0$. Természetesen ugyanezekre az eredményekre jutunk akkor is, ha (X, ν, Y) megfigyelési csatorna helyett valamely $y=T(x)$ statisztikát (tehát zaj nélküli megfigyelési csatornát) vizsgálunk.

Az említett eredmények kapcsolatba hozhatók az elégségesség fogalmának egy RÉNYI Alfréd által adott új értelmezésével is. Eszerint valamely ξ valószínűségi változó egy $\eta=T(\xi)$ függvényét elégségesnek nevezzük a ζ valószínűségi változóra nézve, ha az $I(\eta, \zeta) \equiv I(\xi, \zeta)$ egyenlőtlenségben az egyenlőség érvényes. Ennek a fogalomalkotásnak természetes általánosításaként bármely három Markov-láncot alkotó ζ, ξ, η valószínűségi változó esetén mondhatjuk, hogy η ξ -elégséges a ζ -ra nézve, ha $I(\eta, \zeta) = I(\xi, \zeta)$. Az előbbieket szerint, az ilyen értelemben vett elégségesség $I(\xi, \zeta) < +\infty$ esetén azzal ekvivalens, hogy a ξ -t η -ba átvivő csatorna elégséges egy olyan eloszláscsaládra vonatkozólag, amely ξ -nek a ζ majdnem minden lehetséges értéke melletti feltételes eloszlásaiból áll. (Itt egyszerűség kedvéért közönséges értelemben vett, tehát valós vagy n -dimenziós vektor értékű valószínűségi változókra szorítkozunk, ahol az alapul vett σ -algebra a Borel-halmazok σ -algebrája; ekkor a kérdéses feltételes eloszlások biztosan léteznek, vö. pl. [11].) Ez a feltétel még úgy is megfogalmazható, hogy a kérdéses három valószínűségi változó nemcsak a ζ, ξ, η sorrendben alkot Markov-láncot, hanem a ζ, η, ξ sorrendben is.

Hasonló megfontolások érvényesek $f(u)=u \log u$ helyett más $f(u)$ szigorúan konvex függvényekre is, ha a kölcsönös információ helyére mindenütt az együttes eloszlásnak és a marginális eloszlások szorzatának megfelelő f -eltérését írjuk.

4. § Az f -eltérések alkalmazása határeloszlástételek bizonyítására

Az I -divergencia és általában az f -eltérés, mint az eloszlások egymástól való eltérésének mértékszám, felhasználható valószínűségeloszlások konvergenciájának bizonyítására¹²; a 2. §-ban igazoltuk, hogy az f -konvergencia (az $u_0=1$ pontban szigorúan konvex $f(u)$ függvény esetén) mindig maga után vonja az eloszlások egyenletes konvergenciáját. Markov-láncok esetén, ha létezik μ stacionárius eloszlás, a 3. 2 tételből következőleg az $\mathcal{J}_f(\mu_n, \mu)$ f -eltérések sorozata monoton csökkenő, ahol μ_n az n -edik lépésbeli eloszlást jelöli. Megfelelő további feltételek teljesülése esetén az is igazolható, hogy $\lim_{n \rightarrow \infty} \mathcal{J}_f(\mu_n, \mu) = f(1)$, amiből már következik, hogy μ_n egyenletesen tart a μ stacionárius eloszláshoz. Markov-láncok ergodicitására ilyen „információelméleti” bizonyítást először RÉNYI A. [29], majd később, általánosabb esetre, D. KENDALL [18] adott. Mint [6] dolgozatomban megmutattam, egyszerűbb esetekben a megfontolást lehet úgy módosítani, hogy maga a stacionárius eloszlás létezése is kiadódjék a módszerből; erre vonatkozó eléggé általános feltételeket azonban eddig még nem találtam.

¹² Az a gondolat, hogy bizonyos, az információelméletben használatos mennyiségek felhasználhatók határeloszlástételek bizonyítására, Ju. V. LINNIK-től származik, 1. [22].

Megjegyzendő, hogy a határeloszlástételek bizonyítására való alkalmazások szempontjából a különböző $f(u)$ (szigorúan konvex) függvényeknek megfelelő f -eltérések nem mindig egyenértékűek. Így pl. LINNIK [22] dolgozatában a logaritmus függvény konkrét tulajdonságai messzemenőleg ki vannak használva; természetesen nyitott kérdés marad azonban, hogy nem lehet-e találni olyan $f(u) \neq u \log u$ konvex függvényt, amellyel a centrális határeloszlástétel „információelméleti” bizonyítása egyszerűbben volna elvégezhető. A Markov-láncok ergodicitásának bizonyítására bizonyos esetekben minden f szigorúan konvex függvény egyformán jó, azonban pl. a KENDALL [18] által vizsgált esetben szükség volt az $f(u) > f(0) = 0$ feltételre is. Érdekes megjegyezni, hogy ez utóbbi feltétel az α -rendű I -divergenciák közül az $\alpha > 1$ esetet tünteti ki.

A következőkben a kompakt csoportokon értelmezett valószínűségeloszlások konvolúcióhatványainak konvergenciájára vonatkozó Urbanik—Klossz—Stromberg-féle tételre¹³ (l. [19], [35], [36]) adok egy új bizonyítást, amely az f -eltérések tulajdonságain alapszik; ez egyúttal azt is megmutatja, hogy az f -eltérések olyan határeloszlástételek bizonyítására is alkalmazhatók, melyeknél egyenletes konvergencia nem állítható, csak gyenge konvergencia.

Legyen G kompakt topologikus csoport, $\mathcal{B}$ a Borel-halmazok σ -algebrája; G -n értelmezett valószínűségeloszláson a $\mathcal{B}$ σ -algebrán értelmezett reguláris valószínűségi mértéket értünk. A $\{\mu_n\}_{n=1}^\infty$ eloszlássorozatról akkor mondjuk, hogy gyengén konvergál a μ határeloszláshoz, ha minden G -n értelmezett $h(x)$ folytonos valós függvényre

$$(4.1) \quad \lim_{n \rightarrow \infty} \int h(x) \mu_n(dx) = \int h(x) \mu(dx).$$

Két G -n értelmezett eloszlás, μ és ν , konvolúcióját a

$$(4.2) \quad (\mu \times \nu)(B) = \int \mu(Bx^{-1}) \nu(dx) = \int \nu(x^{-1}B) \mu(dx) \quad (B \in \mathcal{B})$$

vagy az ezzel ekvivalens

$$(4.2') \quad \int h(x) (\mu \times \nu)(dx) = \iint h(xy) \mu(dx) \nu(dy)$$

képlet értelmezi.

Ha μ tétszőleges valószínűségeloszlás G -n, akkor mindazon C kompakt halmazok C_μ metszete, melyekre $\mu(C) = 1$, maga is 1 valószínűségű: $\mu(C_\mu) = 1$. Ezt a C_μ halmazt nevezzük a μ eloszlás hordozójának. Ismeretes, hogy¹⁴

$$(4.3) \quad C_{\mu \times \nu} = C_\mu C_\nu$$

(vö. pl. B. M. KLOSSZ, [19]).

4. 1. TÉTEL. Valamely G -n értelmezett ν eloszlás ν^n konvolúcióhatványainak (ahol $\nu^1 = \nu$, $\nu^{n+1} = \nu^n \times \nu$, $n = 1, 2, \dots$) sorozata akkor és csak akkor konvergál gyenge értelemben valamely határeloszláshoz, ha G -nek a $C_{\nu^{-1}}$ tartalmazó H_ν leg-

¹³ Ez felfogható egy speciális MARKOV-lánc ergodicitási tételnek is, ahol a stacionárius eloszlás ismeretes.

¹⁴ A szorzás komplexusszorzat-értelemben veendő: $AB = \{ab : a \in A, b \in B\}$.

szűkebb kompakt részcsoporthoz nincs olyan valódi kompakt normálosztója, melynek C_v egyetlen mellékosztályában van; ekkor v^n a H_v csoporton értelmezett Haar-mértékhez (pontosabban ennek G -re való természetes kiterjesztéséhez) konvergál.

MEGJEGYZÉS. A v^n konvolúcióhatványok konvergenciájára 1956-ban PRÉKOPA, RÉNYI és URBANIK adott egy elégséges feltételt [28]. A szükséges és elégséges feltételt először K. URBANIK [36] és — tőle függetlenül — B. M. KLOSSZ [19] határozta meg, azonban, ITO és KAWADA [15] cikkének egy hibás eredményére támaszkodva, a feltételt $[C_v] = [C_v C_v^{-1}]$ alakban adták meg (ahol $[A]$ G -nek az $A \subset G$ halmazt tartalmazó legszűkebb kompakt részcsoporthat jelöli). A hibát K. STROMBERG [35] vette észre, megállapítván, hogy URBANIK, illetőleg KLOSSZ bizonyítása valójában a 4.1 tételt szolgáltatja. Az alábbi bizonyítás (vö. [7]) az f -eltérések tulajdonságain alapszik.

Bizonyítás. A (4.3) összefüggés értelmében $C_{v^n} \subset [C_v] = H_v$, ezért elegendő a H_v részcsoporthoz szorítkoznunk; más szóval, az általánosság megszorítása nélkül feltehetjük és a továbbiakban fel is tesszük, hogy $[C_v] = H_v = G$. A tétel szükségességi állítása triviális. Valóban, ha v^n (gyengén) konvergens, akkor bármely G/N faktor-csoportban is a megfelelő v'^n sorozat (gyengén) konvergens (ahol $v' = v\pi^{-1}$, π -vel jelölve a G -nek G/N -be való kanonikus leképezését; N a G tetszőleges kompakt normálosztója). Ha azonban $C_v \subset xN = x' \in G/N$, akkor v'^n konvergens volna azzal ekvivalens, hogy az x'^n sorozat konvergens a G/N csoportban, ami csak úgy lehetséges, ha x' a G/N egységeleme, tehát $C_v \subset N$, és ez (minthogy $[C_v] = G$) $N \neq G$ esetén lehetetlen. Ezért csak az elégségességet kell bizonyítani.

Jelöljük ω -val a Haar-mértéket G -n; az $\mathcal{J}_f(v^n, \omega)$ f -eltérések vizsgálata közvetlenül általában nem volna célravezető, ezért a v^n eloszlásokat először alkalmasan „kisimítjuk”. Legyen $h(x)$ a G -n értelmezett tetszőleges nemnegatív folytonos függvény, melynek az ω Haar-mérték szerinti integrálja 1-gyel egyenlő.

Jelöljük μ_0 -val a

$$(4.4) \quad \mu_0(A) = \int_A h(x) \omega(dx) \quad (A \in \mathcal{B})$$

által definiált eloszlást, és legyen

$$(4.5) \quad \mu_n = \mu_0 \times v^n \quad (n = 1, 2, \dots).$$

Ekkor nyilvánvalóan $\mu_n \ll \omega$ és μ_n -nek ω -ra vonatkozó sűrűségfüggvénye

$$(4.6) \quad p_n(x) = \int h(xy^{-1}) v^n(dy),$$

amiből speciálisan

$$(4.7) \quad p_{n+l}(x) = \int p_n(xy^{-1}) v^l(dy) \quad (l = 1, 2, \dots).$$

Tekintsük mármost az

$$(4.8) \quad \mathcal{J}_n = \mathcal{J}_f(\mu_n, \omega) = \int f(p_n(x)) \omega(dx) \quad (n = 1, 2, \dots)$$

f -eltéréseket, ahol $f(u)$ tetszőleges, $[0, +\infty)$ -ben folytonos, szigorúan konvex függvény (tehát $f(0) < +\infty$). A konvolúció (4.2) definíciója értelmében $\mu \times v$ úgy is

tekinthető, mint egy (G, ν, G) csatorna $(G, \mathcal{B})$ bemeneti terén értelmezett μ mértéknek az ugyancsak $(G, \mathcal{B})$ kimeneti téren megfelelő $\tilde{\mu}$ mérték (vö. (3. 15)), ahol a csatorna átmenetfüggvénye $\nu(B|x) = \nu(x^{-1}B)$. A 3. 2 tételt erre a csatornára alkalmazva, és figyelembe véve, hogy $\mu_n \times \nu = \mu_{n+1}$, $\omega \times \nu = \omega$, azt kapjuk, hogy $\mathcal{J}_{n+1} \subseteq \mathcal{J}_n$ ($n = 1, 2, \dots$). Ezért, figyelembe véve (1. 17)-et is, létezik az

$$(4. 9) \quad \mathcal{J} = \lim_{n \rightarrow \infty} \mathcal{J}_n$$

véges határérték. Igazolni fogjuk, hogy itt $\mathcal{J} = f(1)$, azaz $\mu_n \xrightarrow{f} \omega$ (vö. (2. 3)).

A G csoport kompakt volta miatt a $h(x)$ folytonos függvény egyenletesen is folytonos, és így (4. 6)-ból következik, hogy a $p_n(x)$ sűrűségfüggvények egyenletesen egyenlő mértékben folytonosak. Következésképpen, ismét használva G kompakt-ságát, a $p_n(x)$ ($n = 1, 2, \dots$) függvényt-sorozat bármely részsorozatából kiválasztható egyenletesen konvergens részsorozat. Ha sikerül megmutatni, hogy ezek mindegyike az azonosan 1 függvényhez konvergál, ez azt fogja jelenteni, hogy

$$(4. 10) \quad \lim_{n \rightarrow \infty} p_n(x) = 1 \quad \text{egyenletesen.}$$

Mármost, ha

$$(4. 11) \quad \lim_{k \rightarrow \infty} p_{n_k}(x) = \tilde{p}(x) \quad \text{egyenletesen,}$$

akkor — felhasználva a (4. 7) összefüggést — bármely $l \geq 1$ egész számra

$$(4. 12) \quad \lim_{k \rightarrow \infty} p_{n_k+l}(x) = \int \tilde{p}(xy^{-1}) \nu^l(dx) = \tilde{p}^{(l)}(x) \quad \text{egyenletesen.}$$

Ezért, $\tilde{\mu}$ -mal, ill. $\tilde{\mu}^{(l)}$ -lel jelölve a $\tilde{p}(x)$, ill. $\tilde{p}^{(l)}(x)$ sűrűségfüggvényű eloszlást (ahol a sűrűségfüggvények az ω Haar-mértékre vonatkoznak),

$$(4. 13) \quad \mathcal{J}_f(\tilde{\mu}, \omega) = \mathcal{J}_f(\tilde{\mu}^{(l)}, \omega) = \mathcal{J}.$$

Minthogy azonban $\tilde{\mu}^{(l)} = \tilde{\mu} \times \nu^l$, $\omega = \omega \times \nu^l$, (4. 13)-ból a 3. 2 tétel szerint az következik, hogy a ν^l -lel való konvolúciónak megfelelő csatorna (melynek átmenetfüggvénye $\nu^{(l)}(B|x) = \nu^l(x^{-1}B)$) elégséges a $\tilde{\mu}, \omega$ mértékpárra nézve. Ezért, a 3. 2 lemma értelmében, a $B_s = \{y: \tilde{p}^{(l)}(y) = s\}$ páronként diszjunkt kompakt halmazokra

$$(4. 14) \quad \nu^l(x^{-1}B_{\tilde{p}(x)}) = 1,$$

eltekintve esetleg bizonyos $x \in G_0$ pontoktól, ahol $\omega(G_0) = 0$. Itt azonban valójában $G_0 = \emptyset$. Ha ui. valamely $x_0 \in G$ esetén $\nu^l(x_0^{-1}B_{\tilde{p}(x_0)}) < 1$, akkor a ν^l mérték regularitása miatt még egy alkalmas $U \supset x_0^{-1}B_{\tilde{p}(x_0)}$ nyílt halmaz ν^l -mértékre is 1-nél kisebb. Ekkor azonban a G egységelemének alkalmas V (nyílt) környezetére $Ux_0 \supset V B_{\tilde{p}(x_0)}$ és így $V_1^{-1}V_1 \subset V$ esetén $U \subset (V_1x_0)^{-1}(V_1B_{\tilde{p}(x_0)})$. Ha mármost az egységelem $V_2 \subset V_1$ környezetét úgy választjuk, hogy $x \in V_2x_0$ esetén $|\tilde{p}(x) - \tilde{p}(x_0)| < \min_{y \in V B_{\tilde{p}(x_0)}} |\tilde{p}^{(l)}(y) - \tilde{p}^{(l)}(x_0)|$ legyen (itt a jobboldal $\tilde{p}^{(l)}(x)$ folytonossága és $G - V B_{\tilde{p}(x_0)}$ kompakt volta miatt létezik és pozitív), akkor a V_2x_0 minden x pontjára $x^{-1}B_{\tilde{p}(x_0)} \subset U$, tehát $V_2x_0 \subset G_0$, ami ellentmond annak, hogy $\omega(G_0) = 0$.

(4. 14)-ből a G_s halmazok kompaktsága miatt $x^{-1}B_{\tilde{p}(x)} \supset C_{\nu^l}$, azaz $x C_{\nu^l} \subset B_{\tilde{p}(x)}$ következik. Ezért $x C_{\nu^l} \cap y C_{\nu^l} \neq \emptyset$, azaz $y \in x C_{\nu^l} C_{\nu^l}^{-1}$ esetén szükségképpen $\tilde{p}(y) =$

$=\tilde{p}(x)$ és indukcióval ugyanez következik

$$y \in x C_{v^i}^{-1} C_{v^i}^{-1} C_{v^i}^{-1} \dots C_{v^i}^{-1} C_{v^i}^{-1} \quad (n=1, 2, \dots)$$

esetére is. Így tehát $\tilde{p}(x)$ -nek állandónak kell lennie a

$$H^{(l)} = \bigcup_{n=1}^{\infty} C_{v^i}^{-1} C_{v^i}^{-1} C_{v^i}^{-1} \dots C_{v^i}^{-1} C_{v^i}^{-1}$$

csoport bármely rögzített baloldali mellékosztályán, speciálisan magán H^l -en is ($l=1, 2, \dots$). Mármint a (4.3) szerinti $C_{v^i} = \overset{1}{C_v} \overset{2}{C_v} \dots \overset{n}{C_v}$ összefüggésből közvetlenül következik, hogy bármely

$$(4.15) \quad x = x_1 x_2 \dots x_n \quad (x_i \in C_v, \quad i = 1, \dots, n; \quad n = 1, 2, \dots)$$

alakú elemre $x N_0 x^{-1} \subset N_0$, ahol $N_0 = \bigcup_{l=1}^{\infty} H^{(l)}$ és így természetesen az N_0 csoport N zárt burkára is $x N x^{-1} \subset N$. Másrészt azonban a $[C_v] = G$ feltevés értelmében, kihasználva még, hogy a G kompakt csoport bármely x_i eleme x_i^n hatványainak az egységelem torlódási pontja, a (4.15) alakú elemek halmaza mindenütt sűrű G -ben; tehát N (kompakt) normálosztója G -nek. Ezért, minthogy C_v az N -nek egyetlen mellékosztályába esik (ha ui. egy tetszőleges G csoport valamely C nem üres részhalmazára és H részcsoportjára $CC^{-1} \subset H$, akkor bármely $x_0 \in C$ -re $C \subset CC^{-1} x_0 \subset H x_0$), feltételünk értelmében $N = G$, azaz $N_0 = \bigcup_{l=1}^{\infty} H^{(l)}$ mindenütt sűrű G -ben és így a $\tilde{p}(x)$ folytonos függvény az egész G csoporton állandó. Ez csak úgy lehetséges, ha $\tilde{p}(x)$ az azonosan 1 függvénnyel egyenlő, amivel (4.10)-et igazoltuk, s ezzel egyúttal azt is, hogy $\mathcal{J} = f(1)$, azaz $u_n \xrightarrow{\mathcal{J}} \omega$. Ebből a $v^n \rightarrow \omega$ gyenge konvergencia már közvetlenül következik, ugyanis tetszőleges G -n értelmezett $h(y)$ folytonos valós függvény előállítható $h(y) = \alpha h_1(y^{-1}) - \beta h_2(y^{-1})$ alakban, ahol $h_1(x)$ és $h_2(x)$ nemnegatív folytonos függvények, melyeknek az ω Haar-mérték szerinti integrálja 1-gyel egyenlő. Ekkor pedig az imént bizonyított (4.10) összefüggés értelmében (x -nek a G egységelemét választva)

$$\begin{aligned} & \lim_{n \rightarrow \infty} \int h(y) v^n(dy) = \\ (4.16) \quad & = \lim_{n \rightarrow \infty} \left(\alpha \int h_1(y^{-1}) v^n(dy) - \beta \int h_2(y^{-1}) v^n(dy) \right) = \\ & = \alpha - \beta = \int h(y^{-1}) \omega(dy) = \int h(y) \omega(dy). \end{aligned}$$

Természetesen ha a v mérték abszolút folytonos az ω Haar-mértékre nézve és sűrűségfüggvénye folytonos függvény, akkor a μ_n „kisimított mértékek” bevezetésére nincs szükség, hanem magukra a v^n mértékekre következik, hogy ω -ra vonatkozó sűrűségfüggvényük egyenletesen tart 1-hez; ilyenkor tehát v^n még a (2.40) metrika értelmében is tart ω -hoz (ami persze közvetlenül is egyszerűen igazolható). Megjegyzendő, hogy az iménti bizonyítás szempontjából minden $f(u)$ szigorúan konvex függvény egyformán „jó”, még az $f(u)$ 0-pontbeli folytonosságának fel-

tevése sem volt lényeges, mert a (4. 4)-beli $h(x)$ függvényt vehettük volna szigorúan pozitívnak is.

Érdekes nyitott kérdés, hogy lokálisan kompakt csoportok esetén mikor konvergál a v^n sorozat a (nem normálható) egyenletes eloszláshoz, abban az értelemben, hogy $\lim_{n \rightarrow \infty} \frac{v^n(A)}{v^n(B)} = \frac{\omega(A)}{\omega(B)}$, ha A és B pozitív és véges Haar-mértékű Borel-halmazok, melyeknek a határa nulla Haar-mértékű. Ezt a problémát ilyen általánosságban RÉNYI Alfréd vetette fel. Lehetséges, hogy az f -eltérések erre az esetre is alkalmazhatók lesznek, eddig azonban ebben az irányban nem értem el eredményt. Megemlítem végül, hogy a lokálisan kompakt csoportokon értelmezett tetszőleges $\mu_1, \mu_2, \dots$ eloszlássorozatokból képezett $\mu_1 \times \dots \times \mu_n$ konvolúciósorozatok aszimptotikus viselkedésével [8] dolgozatomban foglalkoztam, itt azonban más módszert használtam. Mindazonáltal valószínűnek látszik, hogy az f -eltérések különböző eloszlások konvolúcióinak határeloszlására vonatkozó tételek bizonyítására is alkalmazhatók.

INFORMATION-TYPE MEASURES OF DIVERGENCE OF PROBABILITY DISTRIBUTIONS

BY I. CSISZÁR

Summary

Let $(X, \mathcal{X})$ be a measurable space and λ a σ -finite measure on $(X, \mathcal{X})$. Probability measures on $(X, \mathcal{X})$ and their densities (with respect to λ) are denoted by μ and p , respectively, both with the same indices (if any).

Definition 1. 1. The f -divergence of two probability distributions μ_1 and μ_2 is

$$(1.16) \quad \mathcal{J}_f(\mu_1, \mu_2) = \int p_2(x) f\left(\frac{p_1(x)}{p_2(x)}\right) \lambda(dx)$$

Here $f(u)$ stands for an arbitrary convex function defined for $0 < u < +\infty$ and expressions of form $f(0)$ and $0f\left(\frac{a}{0}\right)$ are understood as in (1. 8); the value of $\mathcal{J}_f(\mu_1, \mu_2)$ does not depend on the choice of λ .

An equivalent definition of $\mathcal{J}_f(\mu_1, \mu_2)$ is indicated by theorem 1. 1; in (1. 44) R stands for an arbitrary class of finite measurable partitions of $(X, \mathcal{X})$ that spans $\mathcal{X}$.

The class of f -divergences includes the variation distance

$$|\mu_1 - \mu_2| = \int |p_1(x) - p_2(x)| \lambda(dx) \quad (f(u) = |u - 1|), \quad \text{the } X^2\text{-divergence}$$

$$X^2(\mu_1, \mu_2) = \int \frac{(p_1(x) - p_2(x))^2}{p_2(x)} \lambda(dx) \quad (f(u) = (u - 1)^2), \quad \text{Kullback's } I\text{-divergence}$$

$$I(\mu_1 \| \mu_2) = \int p_1(x) \log \frac{p_1(x)}{p_2(x)} \lambda(dx) \quad (f(u) = u \log u), \quad \text{Jeffrey's } \mathcal{J}\text{-divergence}$$

$$\mathcal{J}(\mu_1, \mu_2) = I(\mu_1 \| \mu_2) + I(\mu_2 \| \mu_1) \quad \text{etc.}$$

One motivation of the paper is to extend known results concerning Kullback's I -divergence to the class of f -divergences.

In section 2 we define f -neighbourhoods in an arbitrary set $\mathcal{M}$ of probability distributions on $(X, \mathcal{X})$, see (2. 1). The topological structure obtained in this way on $\mathcal{M}$ is finer than the one induced

by the variation distance (Theorem 2. 1), but in general, it is not a topology in the usual sense, unless both $f(0)$ and $\lim_{u \rightarrow \infty} \frac{f(u)}{u}$ are finite (Theorem 2. 3). In the latter case, however, the f -neighbourhoods define the topology of the variation distance on $\mathcal{M}$ (Theorem 2. 2).

In section 3 we consider indirect observations characterized either by a sub- σ -algebra $\mathcal{X}_0$ of $\mathcal{X}$ or, more generally, by a stochastic kernel ν from $(X, \mathcal{X})$ to another measurable space $(Y, \mathcal{Y})$. If $\tilde{\mu}_i$ denotes the restriction of μ_i to $\mathcal{X}_0$ or the measure on $(Y, \mathcal{Y})$ induced by μ_i and ν , we have $\mathcal{J}_f(\tilde{\mu}_1, \tilde{\mu}_2) \leq \mathcal{J}_f(\mu_1, \mu_2)$ (theorems 3. 1 and 3. 2), the equality holds if and only if (provided that f is strictly convex) the indirect observation is sufficient with respect to the pair μ_1, μ_2 (sufficiency of observation channels is defined analogously to sufficiency of sub- σ -algebras and of statistics). Also some necessary conditions of „small” decrease of the f -divergence are presented, that can be interpreted as „ ε -sufficiency (Theorems 3. 3—3. 5). Section 3 concludes with a discussion of the implications of the results for the amount of information from a statistical observation with respect to the unknown parameter, if the Bayesian point of view is adopted.

Section 4 contains a new proof of the well-known necessary and sufficient condition of convergence of the convolution powers of a probability distribution on a compact group, based on properties of f -divergences. •