

ENTRÓPIAMAXIMALIZÁLÁS ÉS ROKON MÓDSZEREK: AXIOMATIKA, ALGORITMUSOK¹

CSISZÁR IMRE
MTA Matematikai Kutató Intézet
1964 Budapest, Pf. 127

1. Bevezetés

Az entrópia eredetileg fizikai fogalom. A termodinamika második főtétele szerint zárt rendszerben az entrópia sohasem csökken, és az egyensúlyi állapotban éri el maximumát. Az információelmélet (Shannon [26]) az entrópia fogalmát új megvilágításba helyezte. Eszerint egy tetszőleges $P = (p_1, \dots, p_n)$ valószínűségeloszlás

$$H(P) = - \sum_{j=1}^n p_j \log p_j \quad (1.1)$$

entrópiája azt méri, hogy a P eloszlással jellemzett kísérlet kimenetele mennyire véletlen (bizonytalan), abban az értelemben, hogy a kísérlet elvégzése átlagosan hány bit információt szolgáltat. Pontosabban, a bitben mért információmennyiség akkor adódik, ha az (1.1) képletben a $\log 2$ alapú logaritmust jelent, az információelméletben szokásos konvenciónak megfelelően. Ebben a cikkben a szűkebb értelemben vett információelméleti szempontok nem játszanak szerepet, ezért egyszerűség kedvéért mindenütt természetes logaritmust használunk.

A fenti értelmezés természetesen sugallja az *entrópiamaximalizálás* elvét, melynek széleskörű alkalmazását elsősorban Jaynes kezdeményezte, vö. [19]: ha a P eloszlás nem ismert, csak az, hogy P egy \mathcal{P} eloszláshalmaz eleme, akkor az ismereteinkkel kompatibilis eloszlások közül a leginkább véletlenszerűt választjuk, vagyis a maximális entrópiájú $P \in \mathcal{P}$ eloszlást. Ha pl. bizonyos X_1, \dots, X_k valószínűségi változók $E(X_i) = b_i$ várható értékei ismertek, ahol X_i lehetséges értékei a_{i1}, \dots, a_{in} , akkor a \mathcal{P} eloszláshalmaz a

$$\sum_{j=1}^n a_{ij} p_j = b_i, \quad i = 1, \dots, k \quad (1.2)$$

¹Beérkezett: 1994. január 23.

egyenleteknek eleget tevő $P = (p_1, \dots, p_n)$ eloszlásokból áll. Ekkor az (1.1) entrópiafüggvénynek az (1.2) lineáris feltételek melletti maximalizálása adja az ismereteinkkel kompatibilis legvéletlenszerűbb eloszlást.

Az (1.1) entrópiával rokon mennyiség az

$$I(P||Q) = \sum_{j=1}^n p_j \log \frac{p_j}{q_j} \quad (1.3)$$

I -divergencia, amely több más néven is ismeretes, így relatív vagy kereszt-entrópia, diszkrimináló információ, vagy Kullback-Leibler információ; szokásos más jelölései $D(P||Q)$ (az információelméleti irodalomban) és $K(P, Q)$ (a valószínűségszámítási irodalomban). Az I -divergencia nem szimmetrikus információelméleti mértékszám, annak, hogy a $P = (p_1, \dots, p_n)$ eloszlás mennyire tér el a $Q = (q_1, \dots, q_n)$ eloszlástól. Statisztikai alkalmazását Kullback [23] alapozta meg, információelméleti alkalmazásairól l. Csiszár-Körner [13].

Az entrópiamaximalizálás elvének természetes általánosítása a *divergenciainimalizálás elve*: Ha a P ismeretlen eloszlásról az a priori elképzelés $P = Q$, de ez nem kompatibilis a rendelkezésre álló ismeretekkel (tehát $Q \notin \mathcal{P}$), akkor azt a $P \in \mathcal{P}$ eloszlást választjuk, melynek a Q -tól vett I -divergenciája minimális. Ez az entrópiamaximalizálással azonos eredményre vezet, ha a $Q_0 = (\frac{1}{n}, \dots, \frac{1}{n})$ egyenletes eloszlás az a priori elképzelés (ami természetes akkor, ha semmi előzetes információ nem áll rendelkezésre). Valóban, (1.1)-ből és (1.3)-ból látható, hogy minden P eloszlásra

$$I(P||Q_0) = \log n - H(P). \quad (1.4)$$

Az entrópiamaximalizálás elvének a „maximális véletlenszerűség” heurisztikánál mélyebb alátámasztását adja a következő, lényegében Boltzmanntól származó megfontolás (Boltzmann [2], Jaynes [19]). Egy n lehetséges kimenetelű kísérletet N -szer ismételve, ahol $N \gg n$, legyen $p_j = \frac{N_j}{N}$, ahol N_j a j -edik kimenetel bekövetkezéseinek száma. Feltesszük, hogy az N_j gyakoriságok nem figyelhetők meg, és a $P = (p_1, \dots, p_n)$ empirikus eloszlásról csak az állapítható meg, hogy $P \in \mathcal{P}$. Konkrétág kedvéért legyen pl.

$$\mathcal{P} = \left\{ P : \left| \sum_{j=1}^n p_j a_j - b \right| \leq \epsilon \right\} \quad (1.5)$$

ahol a_1, \dots, a_n, b és $\epsilon > 0$ adottak. Itt N_j jelentheti egy N részecskéből álló fizikai rendszer részecskéi közül a fázistér j -edik cellájába esők számát, a_j az energia értékét ebben a cellában, és b az ϵ hibakorlással mért átlagos energiát. Az N kísérlet kimeneteleinek lehetséges (i_1, \dots, i_N) konfigurációi közül (ahol

i_k a k -adik kísérlet kimenetelét jelenti)

$$\frac{N!}{\prod (Np_j)!} = e^{NH(P)+O(\log N)} \quad (1.6)$$

számúra adnak a $p_j = \frac{N_j}{N}$ relatív gyakoriságok egy meghatározott P empirikus eloszlást. Mivel a lehetséges kimenetel-konfigurációk száma n^N , míg a lehetséges P empirikus eloszlások száma N -nel csak polinomiálisan nő, (1.6)-ból következik, hogy a $P \in \mathcal{P}$ ismerettel kompatibilis (i_1, \dots, i_N) konfigurációk közt túlnyomó többséget alkotnak azok, melyekre az empirikus eloszlás entrópiája közel van a $\max_{P \in \mathcal{P}} H(P)$ maximumhoz. Az (1.5) szerinti konvex, zárt \mathcal{P} halmazon $H(P)$ maximuma egyetlen P^* eloszlásra éretik el, és $H(P)$ -nek $H(P^*)$ -hoz való közelségéből következik P -nek P^* -hoz való közelsége. Ezért a rendelkezésre álló ismerettel kompatibilis konfigurációk túlnyomó többségére a P empirikus eloszlás nem különbözhet számottevően az entrópiamaximalizálással kapott P^* eloszlástól. Ekvivalens megfogalmazással, elenyészően kicsi annak valószínűsége, hogy a tényleges P számottevően különbözik a maximális entrópiájú P^* -tól, feltéve, hogy a lehetséges konfigurációk a priori valószínűségei egyenlők, vagy legalábbis nem különböznek egymástól N -ben exponenciális faktorial. A divergenciaminimalizálás elve hasonló megfontolással támasztható alá, azzal a különbséggel, hogy ott az (i_1, \dots, i_N) konfigurációk a priori valószínűségeit nem egyenlőknek vesszük, hanem a \mathcal{Q} eloszlásból vett független mintavétel szerintinek.

A fenti elemi megfontolás és ennek matematikailag mélyebb változatai (pl. Csiszár [8], [9]) az entrópiamaximalizálás, illetve divergenciaminimalizálás elvének erős alátámasztását adják arra az esetre, amikor az ismeretlen eloszlás egy hosszú kísérletsorozatbeli empirikus eloszlásként, vagy bizonyos típusú feltételes eloszlásként interpretálható. Sajnos, a szokásos alkalmazások jelentős részében ez nem (vagy csak mesterként) lehetséges, sőt gyakran nem is valószínűségeloszlás meghatározása a feladat. Az utóbbi illusztrálására tekintsünk egy

$$\sum_{j=1}^n a_{ij} v_j = b_i, \quad i = 1, \dots, k, \quad k < n \quad (1.7)$$

egyenletrendszert, ahol a v_j ismeretlenek nemnegatívak, de (1.2)-től eltérőleg nincs kikötve, hogy az összegük 1. Az (1.7) egyenletrendszernek a divergenciaminimalizálás elve szerinti megoldásán azt értjük, hogy ha adott egy \mathbf{u} vektor, mint a megoldásra vonatkozó a priori elképzelés, az (1.7)-nek elegendő \mathbf{v} vektorok közül azt választjuk, melyre az I -divergencia

$$I(\mathbf{v}||\mathbf{u}) = \sum_{j=1}^n (v_j \log \frac{v_j}{u_j} - v_j + u_j) \quad (1.8)$$

által definiált általánosított változata minimális. A valószínűségeloszlások I -divergenciájától eltérőleg a tetszőleges nemnegatív komponensű vektorok I -divergenciájának nincs információelméleti jelentése, de az az értelmezés fenntartható, hogy $I(\mathbf{v}||\mathbf{u})$ a \mathbf{v} vektornak az \mathbf{u} -tól való különbözőségének mértékszámát. Valóban, $I(\mathbf{v}||\mathbf{u})$ nemnegatív és csak $\mathbf{v} = \mathbf{u}$ esetén nulla, mivel az (1.8) összeg minden tagja nemnegatív és $v_j \neq u_j$ esetén határozottan pozitív.

A következőkben feltételezzük, hogy az ismeretlen \mathbf{v} vektorról valószínűségi ismeretek nem állnak rendelkezésre, így az (1.7) egyenlet megoldására statisztikai jellegű módszerek nem alkalmazhatók.

Az ebben a cikkben ismertetett axiomatikus megközelítés (Csiszár [11]) fő motivációja az a kérdés volt, hogy ilyen általánosságban mennyire indokolt a divergenciáminimalizálás elvének alkalmazása. Ezt a kérdést sikerült (a szerző számára) megnyugtatóan tisztázni, amennyiben a divergenciáminimalizálás több szempontból is kitüntetett módszernek bizonyult arra, hogy az (1.7) típusú egyenletrendszer pozitív megoldásai közül kiválasszunk egyet. Korábban ilyen irányú eredményt ért el Shore és Johnson [27] és Skilling [28], de ők eleve feltételezték, hogy – a 2. pontban bevezetett terminológiával – a kiválasztási szabályt valamilyen függvény generálja. Az entrópiamaximalizálás módszerének ezen feltevés nélküli axiomatikus levezetését adták, a szerzővel egyidőben, Paris és Vencovská [25]. Az alábbi axiomatikus tárgyalás további lényeges eredménye, hogy elvezet alternatívaként számbajövő más módszerekhez is. Parallel módon vizsgálva azt az esetet, amikor a megoldásra nincs pozitivitási feltétel, az a frappáns, bár talán nem meglepő eredmény adódik, hogy ekkor a „legkisebb négyzetek” módszer a kitüntetett.

Érdeemes rámutatni, hogy eredményeink az I -divergencia, az euklideszi távolság és más (tágabb értelemben vett) távolságfüggvények axiomatikus karakterizációit is jelentik. Ezek abban különböznek az ismert axiomatikus karakterizációktól (melyeknek kiterjedt irodalma van, l. Aczél és Daróczy [1]), hogy axiómáink nem magára a távolságfüggvényre vonatkoznak, hanem az általa generált vetítési szabályra írnak elő bizonyos természetes követelményeket.

Röviden foglalkozni fogunk bizonyos numerikus módszerekkel is, konkrétan a RAS módszer néven is ismert iteratív arányos illesztési eljárással és az általánosított iteratív skálázás vagy SMART néven ismert módszerrel. Az entrópiamaximalizálás, illetve divergenciáminimalizálás numerikus módszereinek elég nagy irodalma van, melynek áttekintése meghaladná e cikk kereteit. Az említetteket elsősorban illusztrációként tárgyaljuk, mint matematikailag egyszerű és szemléletesen könnyen interpretálható iteratív algoritmusokat. Megmutatjuk, hogy konvergenciájuk következik egy geometriai jellegű általános tételből, amely bizonyos általános értelemben vett vetítések iterációjának

konvergenciájára vonatkozik.

Ez a cikk a szerzőnek a XXI. Magyar Operációkutatási Konferencián elhangzott előadásának kibővített változata. Az ismertetett eredmények főleg a szerző [11] dolgozatából valók. A munka az OTKA támogatásával készült, szerződésszám 1906.

2. Lineáris inverz problémák; kiválasztási és vetítési szabályok

Lineáris inverz problémáról beszélünk, ha egy ismeretlen f függvényt bizonyos Rf lineáris funkcionálok értékei alapján kell meghatározni. Ezek általában nem határozzák meg egyértelműen f -et, ezért szükség van valamilyen szabályra, amely a lehetséges f -ek közül kiválasztja a megoldásként elfogadandót. Annak a kérdésnek a matematikai vizsgálata, hogy mi a legjobb kiválasztási szabály, már az indulásnál akadályba ütközik: nem látszik lehetségesnek általános objektív mértékszámát adni egy kiválasztási szabály jóságának. Vannak azonban bizonyos természetes követelmények, melyek egy „jó” kiválasztási szabálytól megkívánhatók, mintegy annak logikailag konzisztens voltát jelentik. A következő két pontban megfogalmazzunk néhány ilyen követelményt, és – axiomatikus módszert alkalmazva – megvizsgáljuk, hogy milyen kiválasztási szabályok tesznek ezeknek eleget. Több gyakorlatilag fontos esetben a számításba jövő függvények halmazát nem lineáris egyenlőségek, hanem más típusú feltételek, például lineáris egyenlőtlenségek határozzák meg, l. a bevezetésben az (1.5) eloszláshalmazt. A lineáris inverz problémákra adódó „jó” kiválasztási szabályok kiterjesztése erre az általánosabb esetre nem okoz nehézséget, feltéve, hogy a számításba jövő halmaz konvex, vö. [12].

Következő vizsgálatainkban a lineáris inverz problémák diszkrét változótára szorítkozunk, amikor is valamely A $k \times n$ -es mátrix és $b \in \mathbb{R}^k$ vektor esetén az $A\mathbf{v} = \mathbf{b}$ lineáris egyenletrendszernek (l. (1.7)) eleget tevő $\mathbf{v} \in \mathbb{R}^n$ vektorok közül kell kiválasztani egyet. Ez nem jelenti az általánosság lényeges csorbítását, mert a „folytonos” probléma helyett mindig vehetjük annak diszkrét közelítését.

Tipikus példaként röviden vázoljuk a röntgen-tomográfia rekonstrukciós problémáját, vö. [17], [4]. A feladat az f röntgen-abszorpciós függvény meghatározása egy síkbeli tartománynak tekinthető vékony rétegben, ha ismertek ennek $R_i f$ vonalintegráljai bizonyos e_1, \dots, e_k egyenesek mentén. Az e_i egyenesek a vizsgálatban használt röntgensugár-utak, és az $R_i f$ értékek abból határozhatók meg, hogy az i -edik sugár intenzitásának csökkenési faktora $e^{-R_i f}$. Mármost bontsuk fel a vizsgált síkbeli tartományt n cellára,

melyekben az abszorpciós függvény már konstansnak tekinthető. Ekkor közelítőleg

$$f = \sum_{j=1}^n v_j f_j, \quad R_i f = \sum_{j=1}^n v_j R_i f_j, \quad (2.1)$$

ahol f_j a j -edik cella indikátorfüggvénye és v_j az f (konstansnak tekintett) értéke ebben a cellában. Így, bevezetve az $a_{ij} = R_i f_j$, $b_i = R_i f$ jelölést, a feladat arra redukálódik, hogy meghatározzuk a v_j komponensek alkotta \mathbf{v} vektort, ha a rendelkezésre álló ismeret az (1.7) egyenletek teljesülése.

A következőkben együtt tárgyaljuk azt a két esetet, amikor az (1.7) alakú egyenletrendszer megoldásaként csak a pozitív komponensű vektorok (mint a fenti példában), illetve tetszőleges $\mathbf{v} \in \mathbb{R}^n$ vektorok megengedettek. Ezekre pozitív és valós esetként fogunk utalni. Definícióink és állításaink, ha más nem mondunk, mindkét esetre vonatkoznak.

Jelölési konvenció: Jelölje S a pozitív esetben a pozitív komponensű $\mathbf{v} \in \mathbb{R}^n$ vektorok halmazát, valós esetben pedig a teljes \mathbb{R}^n teret. Itt $n \geq 3$ rögzített. \mathbf{u} és \mathbf{v} mindig S -beli vektort fog jelölni, ha külön nem mondjuk is.

2.1 Definíció *Affin altéren S -nek egy lineáris feltételekkel meghatározott nem üres részhalmazát értjük. Az affin alterek halmazát \mathcal{L} -el jelöljük, tehát $L \in \mathcal{L}$ akkor és csak akkor, ha valamely $k \leq n$ -re megadható olyan $k \times n$ -es \mathbf{A} mátrix és $\mathbf{b} \in \mathbb{R}^n$ vektor, hogy*

$$L = \{ \mathbf{v} : \mathbf{A}\mathbf{v} = \mathbf{b} \} \neq \emptyset. \quad (2.2)$$

Az $n - 1$ dimenziós (tehát egyetlen lineáris feltétellel meghatározott) affin altereket hipersíknak nevezzük.

Megjegyzés: A valós esetben a 2.1 definíció azonos az affin altér, illetve hipersík szokásos definíciójával. A pozitív esetben viszont egy közöséges értelemben vett affin altér, illetve hipersík S -sel való metszetéről van szó. A következőkben L (vagy L' , L'') mindig affin alteret fog jelölni, ha külön nem mondjuk is.

2.2 Definíció *Kiválasztási szabályon olyan $\pi : \mathcal{L} \rightarrow S$ leképezést értünk, melyre minden $L \in \mathcal{L}$ esetén $\pi(L) \in L$.*

A 2.1. definíció értelmében a teljes S halmaz is \mathcal{L} -be tartozik. $\pi(S)$ azt a \mathbf{v} vektort jelenti, melyet akkor választanánk, ha semmi információ nem állna rendelkezésre, tehát az ismeretlen vektorra vonatkozó a priori elképzelést. A bevezetésben mondtak értelmében az, hogy az $\mathbf{A}\mathbf{v} = \mathbf{b}$ egyenletnek eleget tevő \mathbf{v} vektorok közül melyiket választjuk, függhet a megoldásra vonatkozó a priori elképzeléstől. Ezt úgy formalizáljuk, hogy minden $\mathbf{u} \in S$ vektorhoz,

mint lehetséges a priori elképzeléshez, hozzárendelünk egy $\pi(\cdot|u)$ kiválasztási szabályt, melyre $\pi(S|u) = u$. Ekkor a $\pi(L|u)$ vektort u -nak az L affin altérre vett absztrakt vetületének tekinthetjük. Ez motiválja a következő definíciót:

2.3 Definíció *Vetítési szabályon kiválasztási szabályoknak olyan*

$$\{ \pi(\cdot|u), u \in S \}$$

családját értjük, melyre minden u esetén $\pi(S|u) = u$.

Kiválasztási szabály megadása történhet az S halmazon értelmezett valamely F függvény segítségével: legyen $\pi(L)$ az a $v \in L$ vektor, melyre $F(v)$ minimális az L halmazon. Egy adott F függvény akkor és csak akkor generál ily módon kiválasztási szabályt, ha F -nek L -en vett minimuma minden $L \in \mathcal{L}$ esetén eléretik, mégpedig egyetlen $v \in L$ -re. Speciálisan, szükséges feltétel, hogy F az S halmazon felvegye a minimumát, mégpedig egyetlen v^0 pontban. Az általánosság csorbítása nélkül feltehető, hogy itt $F(v^0) = 0$.

Hasonlóan, vetítési szabály megadása történhet egy $S \times S$ -en értelmezett $F(v|u)$ függvény segítségével, amely minden u -ra v függvényeként eleget tesz az előző bekezdésben említett feltételeknek, és minimumát $v = u$ esetén éri el. Az általánosság csorbítása nélkül feltehető, hogy ez a minimum 0 , így

$$F(v|u) > F(u|u) = 0 \quad \text{ha } v \neq u. \quad (2.3)$$

Egy (2.3)-nak eleget tevő F függvényt (tágabb értelemben vett) távolság-függvénynek nevezünk.

Megjegyzés: Természetesen nem minden kiválasztási, illetve vetítési szabály generálható függvénnyel. A következő pontban azonban látni fogjuk, hogy bizonyos egyszerű természetes követelmények már elégséges feltételei a függvénnyel való generálhatóságnak.

A kiválasztási szabályok legfontosabb példái a valós esetben az

$$F(v) = \|v\|^2 = \sum_{j=1}^n v_j^2 \quad (2.4)$$

által generált „legkisebb négyzetek” kiválasztási szabály, a pozitív esetben pedig az

$$F(v) = I(v||1) = \sum_{j=1}^n (v_j \log v_j - v_j + 1) \quad (2.5)$$

által generált „maximális entrópia” kiválasztási szabály (a (2.5) függvény minimalizálása olyan L affin altéren, melyen $\sum v_j$ konstans, ekvivalens a $-\sum_{j=1}^n v_j \log v_j$ maximalizálásával, tehát valószínűségeloszlások alkotta affin

altéren a $H(\mathbf{v})$ entrópia maximalizálásával). Vegyük észre, hogy a legkisebb négyzetek kiválasztási szabály a pozitív esetben nincs értelmezve, mert van olyan (pozitív komponensű vektorok alkotta) $L \in \mathcal{L}$, melyen a (2.4) függvény nem veszi fel a minimumát; másrészt a maximális entrópia kiválasztási szabály a valós esetben nincs értelmezve.

Hasonlóképp a vetítési szabályok legfontosabb példái a valós esetben az $F(\mathbf{v}|\mathbf{u}) = \|\mathbf{v} - \mathbf{u}\|^2$ által generált *euklideszi vetítés*, a pozitív esetben pedig az $F(\mathbf{v}|\mathbf{u}) = I(\mathbf{v}|\mathbf{u})$ által generált *I-divergencia vetítés*. Nyilvánvaló, hogy az előbbinél $\pi(L|\mathbf{u})$ az \mathbf{u} vektornak az L affin altérre vett közönséges euklideszi vetülete; az utóbbinál a $\pi(L|\mathbf{u})$ vektort \mathbf{u} -nak L -re vett *I-vetületének* nevezzük.

3. A „legkisebb négyzetek” és „maximális entrópia” módszerek axiomatikus jellemzése

E pont két fő eredménye, hogy bizonyos alapvető axiómáknak csak összeg típusú függvények által generált kiválasztási és vetítési szabályok tesznek eleget, továbbá hogy hozzávéve egy természetes „kompozíciós konzisztencia” axiómát, a (2.4) vagy (2.5) által generált kiválasztási szabály, illetve az euklideszi vagy *I-divergencia vetítés* bizonyul az axiómáinknak eleget tevő egyedül lehetséges kiválasztási, illetve vetítési szabálynak, aszerint, hogy a valós vagy pozitív esetről van-e szó.

Alapvető axiómáink a 3.1 és 3.2 definíciók szerinti regularitás és lokalitás lesznek. Ezek megfogalmazása előtt rámutatunk az axiomatikus vizsgálat kiindulási pontját képező két implicit feltevésre. Egyrészt, hogy $\mathbf{v} \in L$ képviseli a \mathbf{v} ismeretlen vektorra vonatkozó összes ismeretünket (speciálisan, az $A\mathbf{v} = \mathbf{b}$ egyenletrendszer L megoldáshalmazából kiválasztásra kerülő \mathbf{v} vektor csak L -től függhet, és nem az egyenletrendszer konkrét alakjától), másrészt, hogy *minden* L affin altérből ki kell tudni választani egy \mathbf{v} vektort, tehát hogy a kiválasztási szabály értelmezési tartománya a teljes \mathcal{L} . A lineáris inverz problémák konkrét típusai szempontjából ez utóbbi feltevés túlzott idealizáció lehet, és kívánatos olyan kiválasztási szabályok vizsgálata is, melyek értelmezési tartománya \mathcal{L} -nek egy meghatározott részhalmaza. Például, ha valószínűségeloszlás választásáról van szó (l. a bevezetést), akkor csak olyan L affin alterekből való választásra van szükség, melyek vektorai eleget tesznek a $\sum v_j = 1$ feltételnek. Az utóbbi típusú affin alterek halmazán értelmezett kiválasztási és vetítési szabályok axiomatikus vizsgálata szintén megtörtént, és az alább ismertetendőkhez hasonló eredményekre vezetett, bár a technikai problémák nehezebbek voltak, l. Csiszár [11].

3.1. Definíció (a) A $\pi: \mathcal{L} \rightarrow S$ kiválasztási szabály regularis, ha

- (i) $L' \subset L$ affin alterekre $\pi(L) \in L'$ esetén $\pi(L') = \pi(L)$
- (ii) $L \neq L'$ hipersíkokra $\pi(L) \neq \pi(L')$, kivéve, ha L és L' tartalmazza $\pi(S)$ -t
- (iii) \mathcal{L} -nek bármely rögzített dimenziójú affin alterek alkotta részhalmazán $\pi(L)$ folytonosan függ L -től.
- (b) $A \{ \pi(\cdot|u), u \in S \}$ vetítési szabály reguláris, ha mindegyik $\pi(\cdot|u)$ kiválasztási szabály reguláris.

Diskusszió: Ha az ismeretlen v vektorra vonatkozó ismereteinket reprezentáló lineáris egyenletek L megoldáshalmazából a $\pi(L) = v^*$ vektort választottuk ki, és újabb információk az L halmazt L' -re szűkítik, akkor $v^* \in L$ esetén az új információ nem ad okot választásunk megváltoztatására. Az (i) követelmény ezt a természetes konzisztenciafeltételt formalizálja. Speciálisan, (i)-ből következik, hogy $\pi(S) \in L$ esetén mindig $\pi(L) = \pi(S)$. Így vetítési szabályokra (i)-ből következik az a szemléletesen elvárt tulajdonság, hogy $u \in L$ esetén mindig $\pi(L|u) = u$.

A (ii) követelmény technikai jellegű, szemléletesen kevésbé kényszerítő, mint (i) vagy (iii). További vizsgálat tárgya lehet, hogy valóban szükség van-e rá, illetve hogy helyettesíthető-e szemléletesen kényszerítőbb követelménnyel. A (iii) követelmény evidens szemléletes tartalmú technikai feltétel. Vetítési szabályok esetén természetesnek tűnne $\pi(L|u)$ -nak u -tól való folytonos függését is megkövetelni, de erre nem lesz szükség.

Megjegyzés: Bizonyos esetekben indokolt az (i) feltételt nem teljesítő kiválasztási szabályt használni, például ha az a cél, hogy a kiválasztott vektor valamilyen értelemben maximálisan reprezentatív eleme legyen az ismereteink által megengedett vektorok halmazának. Az ilyen esetek jelen vizsgálataink körén kívül esnek.

A következő definíció megfogalmazásához tetszőleges $v \in S$ vektor és tetszőleges $1 \leq j_1 < \dots < j_r \leq n$ egészek alkotta $J = \{j_1, \dots, j_r\}$ halmaz esetén jelölje v_J a v J -beli komponensei alkotta r -dimenziós vektort. Ha $\pi : \mathcal{L} \rightarrow S$ kiválasztási szabály, $\pi_J(L)$ a $(\pi(L))_J$ vektort fogja jelölni.

3.2 Definíció (a) A $\pi : \mathcal{L} \rightarrow S$ kiválasztási szabály lokális, ha minden olyan $L = L' \cap L''$ alakú affin altérre, melyre L' ill. L'' valamely $J \subset \{1, \dots, n\}$ esetén a $v_{J'}$ -re, illetve $v_{J''}$ -re vonatkozó lineáris egyenletekkel definiálható, tehát

$$L' = \{ v : A' v_{J'} = b' \}, \quad L'' = \{ v : A'' v_{J''} = b'' \}, \quad (3.1)$$

teljesül

$$\pi_J(L) = \pi_J(L'), \quad \pi_{J^c}(L) = \pi_{J^c}(L''). \quad (3.2)$$

(b) A $\{\pi(\cdot|\mathbf{u}), \mathbf{u} \in L\}$ vetítési szabály lokális, ha $\pi(\cdot|\mathbf{u})$ minden \mathbf{u} esetén lokális kiválasztási szabály, és ezenkívül a (3.1) feltétel mellett $\pi_J(L'|\mathbf{u})$ ill. $\pi_J(L''|\mathbf{u})$ csak \mathbf{u}_{J-n} , illetve \mathbf{u}_{J^c-n} keresztül függ \mathbf{u} -tól.

Diskusszió: A lokalitás azt jelenti, hogy ha az ismeretlen \mathbf{v} vektorra vonatkozó ismeretek két komplementer részből tevődnek össze, nevezetesen \mathbf{v} -nek a J -beli, ill. a J^c -beli komponenseire vonatkozó ismeretekből, akkor a kiválasztásra kerülő $\mathbf{v}^* = \pi(L)$ vektor J -beli, illetve J^c -beli komponenseit a komplementer halmazba tartozó komponensekre vonatkozó ismeretek nem befolyásolhatják. Ez a követelmény szemléletesen evidensnek tűnik. Például a 2. pontban vázolt tomográfiai problémában, ha az e_i utak halmaza két olyan részre osztható, hogy a vizsgált tartomány minden celláján csak az egyik halmazba tartozó utak haladnak át, akkor a tartományt két függetlenül vizsgált diszjunkt részből állónak tekinthetjük, így természetes, hogy az f abszorpciós függvény értékének meghatározásához mindkét tartománybeli celláknál csak az e_i utak megfelelő halmazához tartozó $R_i f$ értékeket használjuk. A tomográfiában az ilyen két részre bonthatóság csak elméleti lehetőség, a gyakorlatban nem fordul elő. Mégis, ha előfordulna, az f abszorpciós függvénynek két tartományban függetlenül történő meghatározásával szemben felhozható volna az a kifogás, hogy így elentmondásba kerülhetnénk az f -re vonatkozó esetleges simasági feltételekkel. A lokalitás követelményének ilyenfajta kritikája azonban – absztraktul megfogalmazva – azt feltételezi, hogy az ismeretlen \mathbf{v} vektorról mást is tudunk, mint hogy $\mathbf{v} \in L$, ez pedig ellentmond kiindulási feltevésünknek.

A kiválasztási, illetve vetítési szabályokra egy további természetes követelményt fogalmazhatunk meg, ha olyan lineáris inverz problémáról van szó, amely két (vagy több) tényező együttes hatásával kapcsolatos. Formálisan, tételezzük fel, hogy $n = lm$, és a $\mathbf{v} \in S$ vektorokat azonosítsuk $\{v_{ij}\}$ $l \times m$ -es mátrixokkal, pl. sorfolytonos reprezentációban.

Jelölje $\bar{\mathbf{v}} \in \mathbb{R}^l$ és $\bar{\bar{\mathbf{v}}} \in \mathbb{R}^m$ a \mathbf{v} mátrix marginálisait, melyek komponensei

$$\bar{v}_i = \sum_{j=1}^m v_{ij}, \quad i = 1, \dots, l; \quad \bar{\bar{v}}_j = \sum_{i=1}^l v_{ij}, \quad j = 1, \dots, m. \quad (3.3)$$

Pl. a 2. pontban vázolt tomográfiai problémában a vizsgált tartományt sorok és oszlopok szerint természetes cellákra osztani. Ekkor v_{ij} az abszorpciós függvény értékét jelöli az i -edik sor j -edik cellájában, \bar{v}_i és $\bar{\bar{v}}_j$ pedig a kumulatív érték az i -edik sorban illetve a j -edik oszlopban, tehát a „sor-hatás”, ill. „oszlop-hatás”.

Tetszőleges $\mathbf{v} \in S$ esetén jelölje $L_{\mathbf{v}}$ mindazon $\mathbf{w} \in S$ vektorok halmazát, melyekre, a vektorokat a fenti módon $l \times m$ -es mátrixokkal azonosítva, \mathbf{w}

marginálisai megegyeznek \mathbf{v} marginálisaiival, azaz

$$L_{\mathbf{v}} = \{ \mathbf{w} : \bar{\mathbf{w}} = \bar{\mathbf{v}}, \overline{\overline{\mathbf{w}}} = \overline{\overline{\mathbf{v}}} \}. \quad (3.4)$$

Nyilvánvalóan $L_{\mathbf{v}} \in \mathcal{L}$.

3.3 Definíció (a) Egy $\pi : \mathcal{L} \rightarrow S$ kiválasztási szabály $l \times m$ kompozíció-konzisztens, ha $lm = n$ és a $\mathbf{v} \in S$ vektorokat sorfolytonosan reprezentált mátrixoknak tekintve teljesül $\pi(L_{\mathbf{v}}) = \mathbf{v}$, valahányszor \mathbf{v} megadható

$$v_{ij} = s_i + t_j \quad (3.5)$$

alakban (a valós esetben), illetve

$$v_{ij} = s_i t_j \quad (3.6)$$

alakban (a pozitív esetben).

(b) Egy vetítési szabály $l \times m$ kompozíció-konzisztens, ha $\pi(\cdot|\mathbf{u})$ kompozíció-konzisztens kiválasztási szabály minden (3.5), illetve (3.6) alakban megadható $\mathbf{u} \in S$ esetén, a valós, illetve pozitív esetnek megfelelően.

(c) Egy kiválasztási, illetve vetítési szabály kompozíció-konzisztens, ha valamely $l > 1$, $m > 1$ esetén $l \times m$ kompozíció-konzisztens.

Diskusszió: A \mathbf{v} mátrix (3.5) vagy (3.6) alakú voltát (a valós, illetve pozitív esetben) úgy interpretáljuk, hogy a sor- és oszlop-hatások között nincs interakció. Ezzel az interpretációval mindenesetre összhangban van az, hogy a (3.5) ill. (3.6) alakú mátrixokat marginálisaik egyértelműen meghatározzák. Valóban, (3.5) ill. (3.6)-ból egyszerűen következik, hogy

$$v_{ij} = \frac{1}{m} \bar{v}_i + \frac{1}{l} \bar{v}_j - \frac{1}{n} s(\mathbf{v}) \quad \text{ill.} \quad v_{ij} = \frac{1}{s(\mathbf{v})} \bar{v}_i \bar{v}_j,$$

ahol $s(\mathbf{v}) = \sum_{ij} v_{ij} = \sum_i \bar{v}_i = \sum_j \bar{v}_j$. Az említett interpretáció különösen evidens akkor, amikor a \mathbf{v} mátrix (kétdimenziós) valószínűségeloszlást reprezentál. Ekkor (3.6) azt fejezi ki, hogy két független valószínűségi változó együttes eloszlásáról van szó. Mármost $\pi(L_{\mathbf{v}}) = \mathbf{v}$ azt jelenti, hogy ha az ismeretlen mátrixról csak azt tudjuk, hogy marginálisai megegyeznek \mathbf{v} marginálisaiival, akkor kiválasztási szabályunk éppen a \mathbf{v} mátrixot eredményezi. Mivel a marginálisok ismerete semmit sem mond a sor- és oszlop-hatások közötti interakcióról, természetes követelmény, hogy olyan mátrixot kell választani, melynél ilyen interakció nincs. A 3.3. definíció ezt a követelményt formalizálja, a (b) esetben tekintetbe véve azt is, hogy az a priori elképzelés se utaljon interakcióra. Természetesen kompozíció-konzisztens kiválasztási, illetve vetítési szabályról csak akkor beszélhetünk, ha n nem prímszám.

3.1 Tétel (a) Minden reguláris, lokális kiválasztási szabály generálható

$$F(\mathbf{v}) = \sum_{j=1}^n f_j(v_j) \quad (3.7)$$

alakú függvénnyel, ahol – a valós ill. pozitív esetnek megfelelően – az f_j függvények a valós ill. pozitív számokon értelmezett nemnegatív, folytonosan differenciálható, konvex függvények, mindegyik f_j egyetlen v_j^0 pontban egyenlő nullával, és legfeljebb egyikük lehet nem szigorúan konvex; a pozitív esetben ezenkívül

$$\lim_{v \rightarrow 0} f_j'(v) = -\infty, \quad j = 1, \dots, n.$$

A fenti tulajdonságú $F(\mathbf{v})$ függvényt a vetítési szabály (pozitív) konstans faktortól eltekintve egyértelműen meghatározza, és minden ilyen F generál egy reguláris, lokális kiválasztási szabályt.

(b) Minden reguláris lokális vetítési szabály generálható egy

$$F(\mathbf{v}|\mathbf{u}) = \sum_{j=1}^n f_j(v_j|u_j) \quad (3.8)$$

alakú (tágabb értelemben vett) távolságfüggvénnyel, ahol az $f_j(\cdot|u_j)$ függvények minden rögzített \mathbf{u} -ra rendelkeznek az (a) pontbeli tulajdonságokkal, $v_j^0 = u_j$ -vel. A fenti tulajdonságú távolságfüggvényt a vetítési szabály (pozitív) konstans faktortól eltekintve egyértelműen meghatározza. Továbbá minden ilyen távolságfüggvény generál egy reguláris, lokális vetítési szabályt.

3.2 Tétel Tegyük fel, hogy n nem prímszám.

(a) A valós esetben a (2.4) függvény által generált „legkisebb négyzetek” kiválasztási szabály az egyetlen reguláris, lokális és kompozíció-konzisztens kiválasztási szabály, melyre $\pi(S) = 0$, a pozitív esetben pedig a (2.5) függvény által generált „maximális entrópia” kiválasztási szabály az egyetlen reguláris, lokális és kompozíció-konzisztens kiválasztási szabály, melyre $\pi(S) = 1$.

(b) Az egyetlen reguláris, lokális és kompozíció-konzisztens vetítési szabály a valós esetben az euklideszi vetítés, a pozitív esetben pedig az I-divergencia vetítés.

A 3.1 és 3.2 tételek bizonyítását l. [11]-ben. A 3.2 tétel egyszerű módon következik a 3.1 tételből, ennek illusztrálására vázoljuk az első állítás bizonyítását. Legyen a valós esetben π reguláris, lokális és $l \times m$ kompozíció-konzisztens kiválasztási szabály, melyre $\pi(S) = 0$. A 3.1 tétel szerint a π kiválasztási szabályt egy (3.7) alakú függvény generálja, melyet most, a

$\mathbf{v} \in S$ vektoroknak $l \times m$ -es mátrixokkal való azonosításának megfelelően,

$$F(\mathbf{v}) = \sum_{i=1}^l \sum_{j=1}^m f_{ij}(v_{ij}) \quad (3.9)$$

alakban írunk. Ha $\pi(L_{\mathbf{v}}) = \mathbf{v}$, azaz a (3.9) függvény a (3.6) halmazon a \mathbf{v} pontban éri el minimumát, (3.3) felhasználásával következik, hogy

$$f'_{ij}(v_{ij}) = \lambda_i + \mu_j, \quad i = 1, \dots, l; k = 1, \dots, m, \quad (3.10)$$

ahol λ_i és μ_j Lagrange multiplikátorok. Így minden i, j, k, l -re

$$f'_{ij}(v_{ij}) + f'_{kl}(v_{kl}) = f'_{il}(v_{il}) + f'_{kj}(v_{kj}). \quad (3.11)$$

Mivel feltevés szerint $\pi(L_{\mathbf{v}}) = \mathbf{v}$ mindig teljesül, ha \mathbf{v} (3.5) alakú, (3.11)-ből a következő függvényegyenlet-rendszert kapjuk:

$$f'_{ij}(s_i + t_j) + f'_{kl}(s_k + t_l) = f'_{il}(s_i + t_l) + f'_{kj}(s_k + t_j). \quad (3.12)$$

A $\pi(S) = 0$ feltevés értelmében az f_{ij} függvények (nullával egyenlő) minimumukat a 0 pontban érik el, így ott deriváltjuk 0. Ezért (3.12)-ből az $s_i = s$, $t_j = t$, $s_k = t_l = 0$ helyettesítéssel adódik, hogy

$$f'_{ij}(s + t) = f'_{il}(s) + f'_{kj}(t). \quad (3.13)$$

Ebből a $t = 0$ ill. $s = 0$ helyettesítéssel látható, hogy az f'_{il} és az f'_{kj} függvény is azonosan egyenlő az f'_{ij} függvénnyel. Következésképp az f'_{ij} függvények egyazon g folytonos függvénnyel egyenlők, amely (3.13) értelmében eleget tesz a

$$g(s + t) = g(s) + g(t)$$

Cauchy-féle függvényegyenletnek. Az utóbbiból $g(t) = ct$, tehát $f_{ij}(t) = ct^2$ minden i, j -re. Ezzel igazoltuk, hogy π a „legkisebb négyzetek” kiválasztási szabály.

4. Más szóhajövő vetítési szabályok

Ebben a pontban azt vizsgáljuk, hogy ha a kompozíció-konzisztencia követelményét mellőzzük, de helyette – a regularitás és lokalitás megtartásával – bizonyos más természetes követelményeket teszünk, akkor melyek lesznek az ezeknek eleget tevő vetítési szabályok. Az előző ponttól eltérőleg most individuális kiválasztási szabályokkal nem foglalkozunk, csak vetítési szabályokkal.

Legyen

$$L_{ij}(t) = \{ \mathbf{v} : v_i + v_j = t \}. \quad (4.1)$$

4.1 Definíció (a) Egy vetítési szabály kváziszimmetrikus, ha minden (4.1) alakú affin altér és olyan $\mathbf{u} \in S$ esetén, melyre $u_i = u_j$,

$$\mathbf{v}^* = \pi(L_{ij}(t)|\mathbf{u})\text{-ra} \quad v_i^* = v_j^* .$$

(b) A pozitív esetben, egy vetítési szabály statisztikai, ha minden (4.1) alakú affin altér és $\mathbf{u} \in S$ esetén

$$\mathbf{v}^* = \pi(L_{ij}(t)|\mathbf{u})\text{-ra} \quad \frac{v_i^*}{u_i} = \frac{v_j^*}{u_j} .$$

Diskusszió: Ha az ismeretlen \mathbf{v} vektorról csak annyit tudunk, hogy $v_i + v_j = t$, és az \mathbf{u} a priori elképzelés i -edik és j -edik komponense ugyanaz, akkor természetes szimmetria-kíváncsi olyan vektort választani, melynek i -edik és j -edik komponensei szintén azonosak. A pozitív esetben az is indokoltnak tűnik, hogy a $v_i + v_j = t$ ismeret alapján mindig olyan vektort válasszunk, melynek i -edik és j -edik komponensei az a priori elképzelés szerinti arányokkal arányosak.

Általánosabban, ha az $\{1, \dots, n\}$ halmaz valamely (J_1, \dots, J_k) partíciójára a $\sum_{j \in J_i} v_j$, $i = 1, \dots, k$ összegek ismeretesek (és más nem), akkor természetes azt a \mathbf{v}^* vektort választani, amely az a priori elképzelésnek ezen feltételekhez való arányos illesztésével kapható, tehát amelyre a $\frac{v_j^*}{u_j}$ hányadosok a J_i halmazokon belül állandók. Valószínűségeloszlások vonatkozásában ez az ún. Jeffrey-szabály általánosan elfogadott, vö. Diaconis-Zabell [16]. Erre utalva neveztük a (b) követelményt kielégítő vetítési szabályokat statisztikainak.

4.2 Definíció Egy vetítési szabály tranzitív, ha minden $L' \subset L$ és $\mathbf{u} \in S$ esetén

$$\pi(L'|\mathbf{u}) = \pi(L'|\pi(L|\mathbf{u})) . \quad (4.2)$$

Diskusszió: A tranzitivitás azt jelenti, hogy ha az \mathbf{u} a priori elképzelés és a $\mathbf{v} \in L$ ismeret alapján a $\mathbf{v}^* = \pi(L|\mathbf{u})$ vektort választottuk, és újabb információk a lehetséges \mathbf{v} -k halmazát L' -re szűkítik, akkor a $\mathbf{v} \in L'$ ismeret alapján való választásnál ugyanarra az eredményre jutunk, akár az eredeti \mathbf{u} -t akár \mathbf{v}^* -t használjuk a priori elképzelésként. Másrészt arra is gondolhatunk, hogy az ismereteinket reprezentáló lineáris egyenletrendszer két részből tevődik össze, melyek bizonyos L_1 és L_2 affin altereket határoznak meg. Ekkor az $L = L_1 \cap L_2$ által reprezentált ismereteinknek megfelelő

választás kétféleképp végezhető el két lépésben: először vagy L_1 -ből, vagy L_2 -ből választunk, majd a priori elképzelésként az eredeti \mathbf{u} helyett az így kapott vektort használva választunk második lépésben L -ből. A tranzitivitási követelmény biztosítja, hogy mindkét esetben ugyanazt az eredményt kapjuk, amely megegyezik $\pi(L|\mathbf{u})$ -val.

4.3 Definíció A $\{\pi(\cdot|\mathbf{u}), \mathbf{u} \in S\}$ vetítési szabály skála-invariáns, ha minden $\lambda > 0$, $L \in \mathcal{L}$, $\mathbf{u} \in S$ esetén

$$\pi(\lambda L|\lambda \mathbf{u}) = \lambda \pi(L|\mathbf{u}).$$

Továbbá, a valós esetben, a vetítési szabály eltolás-invariáns, ha minden $\mu \in \mathbb{R}$, $L \in \mathcal{L}$, $\mathbf{u} \in S$ esetén

$$\pi(L + \mu \mathbf{1}|\mathbf{u} + \mu \mathbf{1}) = \pi(L|\mathbf{u}) + \mu \mathbf{1}.$$

Itt λL ill. $L + \mu \mathbf{1}$ mindazon $\lambda \mathbf{v}$ ill. $\mathbf{v} + \mu \mathbf{1}$ vektorok halmazát jelöli, melyekre $\mathbf{v} \in L$. Tehát ha $L = \{\mathbf{v} : \mathbf{A}\mathbf{v} = \mathbf{b}\}$, akkor

$$\lambda L = \{\tilde{\mathbf{v}} : \mathbf{A}\tilde{\mathbf{v}} = \lambda \mathbf{b}\}, \quad \lambda L + \mu \mathbf{1} = \{\tilde{\mathbf{v}} : \mathbf{A}\tilde{\mathbf{v}} = \mathbf{b} + \mu \mathbf{A}\mathbf{1}\}. \quad (4.3)$$

Diskusszió: Ezek az invarianciatulajdonságok evidens szemléletes jelentésű, nyilvánvalóan kívánatos követelmények.

A 3.1 tétel szerint minden reguláris, lokális vetítési szabályt egy (3.8) alakú és a tételben leírt további tulajdonságokkal rendelkező távolságfüggvény generál. A következőkben F mindig ilyen távolságfüggvényt jelöl, anélkül, hogy ezt külön mondanánk. Alább olyan F -ek jellemzését adjuk meg, melyek a most megfogalmazott követelmények közül bizonyosoknak eleget tevő vetítési szabályokat generálnak. Korolláriumként az euklideszi vetítés és az I -divergencia vetítés újabb axiomaticus karakterizációját is kapjuk. A következő tételek a 3. pont végén vázolt bizonyításhoz hasonlóan igazolhatók, sőt a 4.1. tétel bizonyításához függvényegyenlet megoldására sincs szükség. A részletes bizonyításokat l. [11]-ben.

4.1 Tétel (a) F akkor és csak akkor generál kváziszimmetrikus vetítési szabályt, ha a (3.8)-beli $f_j(\mathbf{v}|\mathbf{u})$ függvények nem függenek j -től.

(b) A pozitív esetben, F akkor és csak akkor generál statisztikai vetítési szabályt, ha

$$F(\mathbf{v}|\mathbf{u}) = \sum_{j=1}^n u_j f\left(\frac{v_j}{u_j}\right), \quad (4.4)$$

ahol $f(t)$ folytonosan differenciálható szigorúan konvex függvény, melyre

$$f(1) = f'(1) = 0, \quad \lim_{t \rightarrow 0} f'(t) = -\infty.$$

Diskusszió: (a) azt mutatja, hogy (reguláris, lokális vetítési szabályok esetén) a kváziszimmetria már elégséges feltétele a szimmetriának, tehát annak, hogy a vetítési szabály invariáns legyen a vektorok komponenseinek permutációjára nézve. A statisztikai vetítési szabályokat generáló (4.4) alakú távolságfüggvények osztályát, más megfontolásból, a szerző vezette be [6] dolgozatában, *f*-divergenciák néven (valószínűségeloszlások közötti távolságfüggvényként). A statisztikában használatos távolságfüggvények jelentős része az *f*-divergenciák közé tartozik; részletes tárgyalásukat l. a [24] könyvben.

4.2 Tétel *F akkor és csak akkor generál tranzitív vetítési szabályt, ha*

$$F(\mathbf{v}|\mathbf{u}) = \sum_{j=1}^n (\varphi_j(v_j) - \varphi_j(u_j) - \varphi'_j(u_j)(v_j - u_j)), \quad (4.5)$$

ahol a φ_j függvények folytonosan differenciálhatóak és szigorúan konvexek, és a pozitív esetben $\lim_{v \rightarrow 0} \varphi'_j(v) = -\infty$.

Korollárium *A pozitív esetben az egyetlen reguláris, lokális, tranzitív és statisztikai vetítési szabály az I-divergencia vetítés.*

Diskusszió: (4.5) értelmében

$$F(\mathbf{v}|\mathbf{u}) = \Phi(\mathbf{v}) - \Phi(\mathbf{u}) - \nabla\Phi(\mathbf{u})(\mathbf{v} - \mathbf{u}), \quad (4.6)$$

ahol $\Phi(\mathbf{v}) = \sum_{j=1}^n \varphi_j(\mathbf{v})$. Tehát $F(\mathbf{v}|\mathbf{u})$ azt adja meg, hogy a Φ függvény gráfjának \mathbf{v} -hez tartozó pontja mennyivel van az \mathbf{u} -hoz tartozó pontban érintő hipersík felett. Az ilyen távolságfüggvények a *Bregman-divergenciák* l. Bregman [3]. Lineáris inverz problémákra való alkalmazásairól, a folytonos esetben, l. Jones és Byrne [20], Jones és Trutzer [21]. Az *f*-divergenciák és a Bregman-divergenciák családjának közös eleme csak az *I*-divergencia és konstansszorosai; ez a korollárium tartalma.

4.3 Tétel *(a) A valós esetben, F akkor és csak akkor generál eltolás- és skálainvariáns tranzitív vetítési szabályt, ha*

$$F(\mathbf{v}|\mathbf{u}) = \sum_{j=1}^n c_j (v_j - u_j)^2, \quad (4.7)$$

ahol a c_j -k pozitív konstansok.

(b) A pozitív esetben, F akkor és csak akkor generál skálainvariáns tranzitív vetítési szabályt, ha

$$F(\mathbf{v}|\mathbf{u}) = \sum_{j=1}^n c_j h_\alpha(v_j|u_j), \quad (4.8)$$

ahol a c_j -k pozitív konstansok, $\alpha \leq 1$, és

$$h_\alpha(v|u) = \begin{cases} v \log \frac{v}{u} - v + u, & \text{ha } \alpha = 1; \\ \log \frac{v}{u} + \frac{v}{u} - 1, & \text{ha } \alpha = 0; \\ \frac{1}{\alpha}(u^\alpha - v^\alpha) + u^{\alpha-1}(v - u), & \text{ha } \alpha < 1, \alpha \neq 0. \end{cases} \quad (4.9)$$

Korollárium (a) A valós esetben, az egyetlen reguláris, lokális, eltolás-és skálainvariáns, tranzitív és kváziszimmetrikus vetítési szabály az euklideszi vetítés.

(b) A pozitív esetben a reguláris, lokális, skálainvariáns, tranzitív és kváziszimmetrikus vetítési szabályok az

$$F_\alpha(\mathbf{v}|\mathbf{u}) = \sum_{j=1}^n h_\alpha(v_j|u_j), \quad \alpha \leq 1 \quad (4.10)$$

távolságfüggvények által generált vetítési szabályok, ahol a h_α függvények a (4.9) szerintiek.

Diskusszió: A (4.10) távolságfüggvény $\alpha = 1$ esetén az I -divergencia, $\alpha = 0$ esetén az Itakura-Saito távolság, l. [18]. Az α paraméter más értékeire adódó távolságoknak az irodalomban való előfordulásáról a szerzőnek nincs tudomása, kivéve Jones és Trutzer [21] munkáját, akik szerint egy ilyen alakú távolság folytonos megfelelője spektrumanalízis feladatoknál jobb eredményekre vezetett, mint az Itakura-Saito távolságot használó szokásos módszer. Mindenesetre a 4.3 tétel korolláriuma arra utal, hogy – a pozitív esetben – az I -divergencia vetítés alternatíváiként leginkább az $\alpha \neq 1$ paraméterű (4.10) távolságfüggvények szerinti vetítések jöhetnek szóba. Ezek közül azonban az Itakura-Saito távolság ($\alpha = 0$ eset) axiomatikus szempontból nincs kitüntetve.

5. Az I -vetület numerikus meghatározása

A lineáris inverz problémák egyik legegyszerűbb típusa, amikor egy nem-negatív elemű mátrixot kell rekonstruálni a marginálisai alapján, l. (3.3),

felhasználva a mátrixra vonatkozó valamilyen a priori elképzelést. Erre általában használt módszer az eredetileg különösebb elméleti megalapozás nélkül bevezetett *iteratív arányos illesztési eljárás*, amely RAS módszer néven is ismeretes. Az irodalomban a szerző tudomása szerint először Kruithof [22] dolgozatában fordult elő, telefonfoglalom becslésére; a statisztikában kontingenctáblázatok elemzésére először Deming és Stephan [15] alkalmazta.

A módszer, kissé általánosabb formában, akkor alkalmazható, ha a pozitív esetről van szó (tehát S a pozitív komponensű $\mathbf{v} \in \mathbb{R}^n$ vektorok halmaza), és az ismeretlen \mathbf{v} vektorról annyit tudunk, hogy az $\{1, \dots, n\}$ halmaz bizonyos $(J_1^1, \dots, J_{k_1}^1), \dots, (J_1^r, \dots, J_{k_r}^r)$ partícióira ismertek a

$$\sum_{j \in J_k^i} v_j = b_k^i, \quad 1 \leq i \leq r, \quad 1 \leq k \leq k_r \quad (5.1)$$

összegek. Az iteráció abban áll, hogy az adott $\mathbf{u} \in S$ a priori elképzelésből kiindulva, a $\mathbf{v}(t)$, $t = 0, 1, 2, \dots$ vektorokat a következőképp határozzuk meg: $\mathbf{v}(0) = \mathbf{u}$, és $t \geq 1$ esetén a $\mathbf{v}(t)$ vektort $\mathbf{v}(t-1)$ -ből az $i(t)$ -edik partíció által előírt $b_k^{i(t)}$ összegekhez való arányos illesztés származtatja, ahol $i(t) \equiv t \pmod{r}$. Másszóval

$$v_j(t) = \frac{b_k^{i(t)}}{\sum_{l \in J_k^{i(t)}} v_l(t-1)} v_j(t-1) \quad \text{ha} \quad j \in J_k^{i(t)}. \quad (5.2)$$

Speciálisan, ha \mathbf{v} mátrixot jelent, melynek $\bar{\mathbf{v}}$ és $\bar{\mathbf{v}}$ marginálisai vannak megadva, akkor itt $r = 2$, a két partíció a mátrix sorainak ill. oszlopainak felel meg, és az iteráció a megadott sor- és oszlopösszegekhez való arányos illesztések ciklikus ismétlése.

Az alábbi 5.1 tételből speciális esetként következik, hogy ha egyáltalán létezik az (5.1) feltételeket kielégítő $\mathbf{v} \in S$ vektor, akkor

$$\mathbf{v}^* = \lim_{t \rightarrow \infty} \mathbf{v}(t) \quad (5.3)$$

létezik, és egyenlő az \mathbf{u} I -vetületével az (5.1) feltételekkel definiált L affin altérre. Vegyük észre, hogy az (5.2) által definiált $\mathbf{v}(t)$ vektor nem más, mint $\mathbf{v}(t-1)$ I -vetülete az $L_{i(t)}$ affin altérre, ahol

$$L_i = \left\{ \mathbf{v} : \sum_{j \in J_k^i} v_j = b_k^i, \quad 1 \leq k \leq k_i \right\}. \quad (5.4)$$

Figyelemreméltó azonban, hogy míg az iteráció egyes lépései nemcsak az I -divergencia, hanem már f -divergenciák szerinti vetítésnek is tekinthetők, vö.

4.1 tétel, \mathbf{u} -nak az $L = \bigcap_{i=1}^r L_i$ affin altérre való vetületéről ez már nem állítható, és az (5.3) limesz éppen az I -vetülettel egyenlő. Ez azzal függ össze, hogy az iteratív vetítés konvergenciája a tranzitív vetítési szabályok tulajdonsága, és az f -divergenciák szerinti vetítések közül egyedül az I -divergencia vetítés tranzitív, l. a 4.2 tétel korolláriumát.

5.1 Tétel *Legyen $\{\pi(\cdot|\mathbf{u}), \mathbf{u} \in S\}$ reguláris, lokális, tranzitív vetítési szabály, és tegyük fel, hogy*

$$L = \bigcap_{i=1}^r L_i \neq \emptyset, \quad (5.5)$$

ahol L_1, \dots, L_r tetszőleges affin altérek. Legyen $\mathbf{v}(0) = \mathbf{u}$ és $t \geq 1$ esetén legyen $\mathbf{v}(t) = \pi(L_{i(t)}|\mathbf{v}(t-1))$, ahol $i(t) \equiv t \pmod{r}$. Ekkor

$$\lim_{t \rightarrow \infty} \mathbf{v}(t) = \pi(L|\mathbf{u}), \quad (5.6)$$

feltéve, hogy $\|\mathbf{v}(t)\|$ nem tart végtelenhez. Speciálisan, a következő két feltétel bármelyike elégséges (5.6) teljesüléséhez:

- (i) az L_i halmazok legalább egyike korlátos
- (ii) a vetítési szabályt generáló (4.5) távolságfüggvényben szereplő φ_j függvények mindegyikére

$$u\varphi'_j(\mathbf{u}) - \varphi_j(\mathbf{u}) \rightarrow \infty, \quad \text{ha } |\mathbf{u}| \rightarrow \infty. \quad (5.7)$$

Diskusszió: Az 5.1. tétel a pozitív és a valós esetben egyaránt érvényes (a pozitív esetben (5.7)-ben $|\mathbf{u}| \rightarrow \infty$ helyett $\mathbf{u} \rightarrow \infty$ értendő). Az (i) feltétel a pozitív esetben jön szóba, tipikus esete, amikor az L_i affin altérek egyike olyan \mathbf{v} vektorokból áll, melyek komponenseinek összege konstans. A (ii) feltétel teljesül pl. a 4.3 tételben szereplő vetítési szabályok közül az (a) esetbeliekre, és a (b) esetbeliekből azokra, amelyekre $0 \leq \alpha \leq 1$, így speciálisan az euklideszi vetítésre, az I -divergencia vetítésre és az Itakura-Saito távolság szerinti vetítésre.

Nyitott kérdés, hogy az 5.1. tételben nem állítható-e (5.6) minden további feltétel nélkül, tehát hogy $\|\mathbf{v}(t)\| \rightarrow \infty$ előfordulhat-e egyáltalán.

Az 5.1. tétel egyszerű következménye annak, hogy a tranzitív vetítési szabályokat generáló (4.5) alakú F távolságfüggvényekre teljesül a

Pitagorasz tétel: *Tetszőleges L affin altér, $\mathbf{u} \in S$ és $\mathbf{v} \in L$ esetén*

$$F(\mathbf{v}|\mathbf{u}) = F(\mathbf{v}|\mathbf{v}^*) + F(\mathbf{v}^*|\mathbf{u}) \quad \text{ahol } \mathbf{v}^* = \pi(L|\mathbf{u}). \quad (5.8)$$

Vegyük észre, hogy az $F(\mathbf{v}|\mathbf{u}) = \|\mathbf{v} - \mathbf{u}\|^2$ esetben (5.8) az \mathbf{u} , \mathbf{v}^* , \mathbf{v} által alkotott derékszögű háromszögre vonatkozó közöséges Pitagorasz tétel. Az $F(\mathbf{v}|\mathbf{u}) = I(\mathbf{v}|\mathbf{u})$ esetre vonatkozó (5.8) azonosság fontos szerepet játszik az I -divergencia statisztikai alkalmazásaiban, l. Kullback [23].

Az (5.8) Pitagorasz tétel bizonyításához jegyezzük meg, hogy mivel a vetítési szabályt az

$$F(\mathbf{v}|\mathbf{u}) = \sum_{j=1}^n (\varphi_j(v_j) - \varphi_j(u_j) - \varphi'_j(u_j)(v_j - u_j)) \quad (5.9)$$

távolságfüggvény generálja, (5.9) \mathbf{v} szerinti gradiensvektora a $\mathbf{v}^* = \pi(L|\mathbf{u})$ pontban merőleges L -re, így speciálisan a $\mathbf{v} - \mathbf{v}^*$ vektorra is. Tehát

$$\sum_{j=1}^n (v_j - v_j^*)(\varphi'_j(v_j) - \varphi'_j(u_j)) = 0, \quad (5.10)$$

ami ekvivalens (5.8)-cal.

Mármost az 5.1. tétel a következőképp bizonyítható, vö. Csiszár [7], ahol a tétel az I -divergencia vetítésre volt kimondva, valószínűségeloszlások esetére szorítkozva. Alkalmazzuk az (5.8) Pitagorasz tételt \mathbf{u} és L szerepében $\mathbf{v}(t-1)$ -re és $L_{i(t)}$ -re (az 5.1. tétel jelöléseivel) és legyen $\mathbf{v} \in \bigcap_{i=1}^r L_i$ tetszőleges. Az így adódó azonosságokat $t = 1$ -től N -ig összegezve azt kapjuk, hogy minden $\mathbf{v} \in L = \bigcap_{i=1}^r L_i$ esetén

$$F(\mathbf{v}|\mathbf{u}) = F(\mathbf{v}|\mathbf{v}(N)) + \sum_{t=1}^N F(\mathbf{v}(t)|\mathbf{v}(t-1)). \quad (5.12)$$

Speciálisan, minden $\mathbf{v} \in L$ esetén az $F(\mathbf{v}|\mathbf{v}(t))$, $t = 0, 1, \dots$ sorozat monoton csökkenő, és

$$F(\mathbf{v}|\mathbf{u}) = \lim_{t \rightarrow \infty} F(\mathbf{v}|\mathbf{v}(t)) + \sum_{t=1}^{\infty} F(\mathbf{v}(t)|\mathbf{v}(t-1)). \quad (5.13)$$

Ha $\|\mathbf{v}(t)\|$ nem tart végtelenhez, a $\mathbf{v}(t)$ sorozatból kiválasztható egy $\mathbf{v}(t_k) \rightarrow \mathbf{v}^*$ konvergens részsorozat. Itt $\mathbf{v}^* \in S$ a pozitív esetben is teljesül, mert ha \mathbf{v}^* valamelyik komponense 0 volna, akkor $\lim_{v \rightarrow 0} \varphi'_j(v) = -\infty$ miatt (l. 4.2 tétel) $F(\mathbf{v}|\mathbf{v}(t_k))$ végtelenhez tartana, ellentmondásban $F(\mathbf{v}|\mathbf{v}(t))$ monoton csökkenő voltával.

(5.13)-ból következik, hogy $F(\mathbf{v}(t)|\mathbf{v}(t-1)) \rightarrow 0$. Vegyük észre, hogy ha \mathbf{u} S -nek egy kompakt részhalmazából való, akkor $F(\mathbf{v}|\mathbf{u})$ tetszőlegesen

kicsi voltából következik $\|\mathbf{v} - \mathbf{u}\|$ tetszőlegesen kicsi volta. Ezért abból, hogy $\mathbf{v}(t_k) \rightarrow \mathbf{v}^* \in S$, következik $\mathbf{v}(t_k + 1) \rightarrow \mathbf{v}^*$, ..., $\mathbf{v}(t_k + r - 1) \rightarrow \mathbf{v}^*$ is. Így, mivel a $\mathbf{v}(t_k)$, $\mathbf{v}(t_k + 1)$, ..., $\mathbf{v}(t_k + r - 1)$ vektorok valamely ciklikus permutáció szerint rendre az L_1, \dots, L_r affin alterekből valók, szükségképpen $\mathbf{v}^* \in \bigcap_{i=1}^r L_i = L$. Felhasználva, hogy az $F(\mathbf{v}|\mathbf{v}(t))$ sorozat minden $\mathbf{v} \in L$ esetén monoton csökkenő, most már következik, hogy

$$\lim_{t \rightarrow \infty} F(\mathbf{v}^*|\mathbf{v}(t)) = \lim_{k \rightarrow \infty} F(\mathbf{v}^*|\mathbf{v}(t_k)) = 0, \quad (5.14)$$

tehát nemcsak $\mathbf{v}(t_k) \rightarrow \mathbf{v}^*$, hanem $\mathbf{v}(t) \rightarrow \mathbf{v}^*$ is teljesül. Ezután (5.13)-at $\mathbf{v} = \mathbf{v}^*$ -ra alkalmazva azt kapjuk, hogy

$$F(\mathbf{v}^*|\mathbf{u}) = \sum_{t=1}^{\infty} F(\mathbf{v}(t)|\mathbf{v}(t-1)); \quad (5.15)$$

ezt (5.13)-ba visszahelyettesítve látható, hogy $F(\mathbf{v}|\mathbf{u}) \geq F(\mathbf{v}^*|\mathbf{u})$ minden $\mathbf{v} \in L$ -re, tehát $\mathbf{v}^* = \pi(L|\mathbf{u})$.

Végül még azt kell belátni, hogy az 5.1 tételbeli (i) ill. (ii) feltétel kizárja a $\|\mathbf{v}(t)\| \rightarrow \infty$ lehetőséget. Az (i) feltételnél ez nyilvánvaló, (ii)-nél pedig abból következik, hogy $\|\mathbf{v}(t)\| \rightarrow \infty$ esetén az $F(\mathbf{v}|\mathbf{v}(t))$ sorozat nem lehetne monoton csökkenő, sőt végtelenhez tartana. Valóban, a (ii) feltétel mellett egyszerűen igazolható, hogy bármely rögzített $\mathbf{v} \in S$ esetén $F(\mathbf{v}|\mathbf{u}) \rightarrow \infty$ ha $\|\mathbf{u}\| \rightarrow \infty$.

Az 5.1. tétel elvi lehetőséget ad bármely L affin altérre való I -vetület, vagy általánosabban, tetszőleges reguláris, lokális, tranzitív vetítési szabály szerinti $\pi(L|\mathbf{u})$ vetület kiszámítására, legalábbis az (i) vagy (ii) feltétel teljesülése esetén. Valóban, minden L affin altér előállítható hipersíkok metszeteiként. Így $\pi(L|\mathbf{u})$ kiszámítása történhet olyan iterációval, melynek minden lépése hipersíkra, azaz egyetlen lineáris feltétellel meghatározott L_i affin altérre való vetítés. Ezek a lépések azonban általában lényegesen számításigényesebbek az (5.4) speciális alakú affin alterekre való I -divergencia vetítésnél, ami egyszerűen csak arányos illesztést jelent.

A következő módszerrel minden olyan L affin altérre vett I -vetület kiszámítható az iteratív arányos illesztési eljárásához mérhető számítási komplexitással, amely L -re a $\mathbf{v} \in L$ vektorok komponenseinek összege állandó. Mivel az I -divergencia vetítés skálainvariáns, az általánosság csorbítása nélkül feltehetjük, hogy ez az összeg 1, azaz, hogy L -et valószínűségeloszlások alkotják.

Legyen tehát L a pozitív komponensű $\mathbf{v} \in \mathbb{R}^n$ vektorok S halmazának olyan affin altére, hogy minden $\mathbf{v} \in L$ vektorra $\sum_{j=1}^n v_j = 1$. Könnyű belátni, hogy ekkor L előállítható $L = \{\mathbf{v} : \mathbf{A}\mathbf{v} = \mathbf{b}\}$ alakban olyan \mathbf{A} $k \times n$ -es mátrixszal és $\mathbf{b} \in \mathbb{R}^k$ vektorral, hogy \mathbf{A} elemei nemnegatívak és oszlopösszegei 1-gyel egyenlők, továbbá \mathbf{b} elemei pozitívak és összegük 1.

Egy ilyen előállítást véve alapul, legyen $\mathbf{v}(0) = \mathbf{u}$ (az a priori elképzelést reprezentáló vektor), és a $\mathbf{v}(t)$, $t = 1, 2, \dots$ vektorokat definiálja

$$v_j(t) = v_j(t-1) \prod_{i=1}^k \left(\frac{b_i}{\sum_{l=1}^n a_{il} v_l(t-1)} \right)^{a_{ij}} \quad (5.16)$$

Ezt az iterációt először Darroch és Ratcliff [14] javasolták, és bebizonyították, hogy $\mathbf{v}(t)$ konvergál \mathbf{u} -nak L -re vett I -vetületéhez. Az I -vetület kiszámításának ez a módja az irodalomban SMART algoritmus néven is ismeretes.

Az (5.16) iterációval származtatott $\mathbf{v}(t)$ sorozat konvergenciája \mathbf{u} -nak L -re vett I -vetületéhez egyszerűen levezethető az 5.1 tételből (l. [10]). Ehhez jelölje M azoknak az (i, j) pároknak a halmazát, melyekre $a_{ij} \neq 0$, és jelölje \tilde{S} a $\{\tilde{v}_{ij} : (i, j) \in M\}$ alakban írt pozitív komponensű $\tilde{\mathbf{v}} \in \mathbb{R}^m$ vektorok halmazát, ahol $m \leq nk$ az M elemszáma. Legyen \tilde{L} mindazon $\tilde{\mathbf{v}} \in \tilde{S}$ vektorok halmaza, melyek komponensei $\tilde{v}_{ij} = a_{ij} v_j$ alakúak valamely $\mathbf{v} \in L$ vektorra, és legyen $\tilde{\mathbf{u}} = \{a_{ij} u_j, (i, j) \in M\}$. Ekkor $\tilde{\mathbf{v}} \in \tilde{L}$ esetén $I(\tilde{\mathbf{v}}|\tilde{\mathbf{u}}) = I(\mathbf{v}|\mathbf{u})$, ezért $\tilde{\mathbf{u}}$ \tilde{L} -ra vett I -vetülete az \mathbf{u} L -re vett \mathbf{v}^* I -vetületének megfelelő $\tilde{\mathbf{v}}^*$ -gal egyenlő. Mivel $\tilde{L} = \tilde{L}_1 \cap \tilde{L}_2$, ahol

$$\tilde{L}_1 = \{ \tilde{\mathbf{v}} : \sum_j \tilde{v}_{ij} = b_i, \quad i = 1, \dots, k \} \quad (5.17)$$

$$\tilde{L}_2 = \{ \tilde{\mathbf{v}} : \sum_l \tilde{v}_{lj} - \frac{1}{a_{ij}} \tilde{v}_{ij} = 0, \quad (i, j) \in M \} \quad (5.18)$$

(itt az összegek azokra a j ill. l indexekre vonatkoznak, melyek a szóbanforgó i ill. j indexszel M -beli párt alkotnak), az 5.1. tétel szerint $\tilde{\mathbf{v}}^*$ megkapható az \tilde{L}_1 -ra ill. \tilde{L}_2 -ra való vetítések ciklikus ismétlésével származtatott $\tilde{\mathbf{v}}(t)$ vektorok sorozatának limeszeként. Egyszerű számolással adódik, hogy ez az iteráció ekvivalens az (5.16) iterációval, pontosabban az (5.16) szerinti $\mathbf{v}(t)$ vektorok a $\tilde{\mathbf{v}}(2t)$ vektoroknak felelnek meg.

6. Kiegészítő megjegyzések

A lineáris inverz problémáknak a 2. pont szerintnél általánosabb változatában az ismeretlen f függvényt bizonyos Rf funkcionálok értékeinek *pontatlan* ismerete alapján kell meghatározni. Formálisan, a diszkrét esetben, az ismeretlen \mathbf{v} vektorról rendelkezésre álló ismereteket (1.7) helyett egy

$$\sum_{j=1}^n a_{ij} v_j + e_i = b_i, \quad i = 1, \dots, k \quad (6.1)$$

alakú egyenletrendszer írja le, ahol e_1, \dots, e_k az ismeretlen e hibavektor komponensei. Ezeket szokásos ismert vagy részben ismert együttes eloszlású (legtöbbször független normális) valószínűségi változóknak tekinteni, amikor is a feladat a statisztika módszereivel közelíthető meg. Az eddigiekkel összhangban gondoljunk ehelyett arra az esetre, amikor az e hibavektor nem véletlen (vagy ha igen, eloszlásáról nincs információ), így statisztikai módszerek megalapozottan nem alkalmazhatók. Pl. a 2. pontban vázolt tomográfiai problémában, ha azt hiba megengedésével modellezzük, e a diszkrét közelítésből és a berendezés geometriájának pontatlan ismeretéből származó szisztematikus hibát fogja jelenteni, mert a véletlen mérési hibák ehhez képest elhanyagolhatók. Ha az e hibavektorról feltehető, hogy $|e_i| \leq \epsilon$, $i = 1, \dots, k$, akkor mondhatjuk, hogy a v vektort egy affin altér helyett a

$$\{v : |\sum_j a_{ij} v_j - b_i| \leq \epsilon, i = 1, \dots, k\} \quad (6.2)$$

konvex halmazból kell kiválasztani, vö. (1.5); a dolgozatban tárgyalt axiomatikus eredmények egyszerűen kiterjeszthetők úgy, hogy erre a feladatra is alkalmazhatók legyenek, l. [12]. A (6.2) alakú halmazra való vetítés numerikus módszereit illetően az [5] cikkre utalunk, amely bőséges irodalomjegyzéket is tartalmaz.

A (6.1) probléma más megközelítése, hogy azt olyan szűkebb értelemben vett lineáris inverz problémának tekintjük, melyben a (v, e) pár az ismeretlen, és ennek megválasztására alkalmazunk valamilyen „jó” kiválasztási szabályt. Így a pozitív esetben számításba jön $\|v\|^2$ és $\|e\|^2$ valamely lineáris kombinációjának minimalizálása a (6.1) feltételek mellett, vö. [4], [17].

A klasszikus legkisebb négyzetek módszere a $\sum_j a_{ij} v_j = b_i$, $i = 1, \dots, k$ egyenletrendszer közelítő megoldására szolgál abban a tipikusan $k > n$ -re előforduló esetben, amikor az egyenletek inkompatibilisek; ilyenkor a (6.1) egyenletrendszernek azt a megoldását fogadjuk el, amelyre $\|e\|^2$ minimális. Az, hogy $\|v\|^2$ és $\|e\|^2$ valamely lineáris kombinációját minimalizáljuk, természetes közös általánosítása a 2. pontbeli és a klasszikus értelemben vett legkisebb négyzetek módszerének. A (6.1) egyenletrendszer ily módon történő megoldásának axiomatikusan kitüntetett voltát azonban nem sikerült bizonyítani többek között azért, mert a (6.1) alakú egyenletek nagyon speciálisak, így szó sincs arról, hogy a (v, e) párokból képezhető minden affin altérből ki kell tudni választani egy (v, e) párt.

Visszatérve a 2. pontbeli értelemben vett inverz problémákra, egy „folytonos” probléma diszkrét közelítése nem csak az ott vázolt módon lehetséges. Általánosabban, a (2.1) közelítésben az f_j függvények egy felosztás celláinak indikátorfüggvényei helyett más bázisfüggvények is lehetnek, vö. [4]. Ekkor azonban a diszkrét közelítés már nem tükrözi közvetlenül a folytonos probléma

geometriáját, és ezért a lokalitás (3.2 definíció) nem tűnik kényszerítő követelménynek. Ilyenkor számításba jöhet olyan Bregman-divergencia – azaz (4.6) alakú távolságfüggvény – által generált vetítési szabályt alkalmazni, ahol a $\Phi(\mathbf{v})$ függvény (szigorúan konvex és folytonosan differenciálható, de) nem $\sum_{j=1}^n \varphi_j(v_j)$ alakú. Ezekre az (5.8) Pitagorasz-tétel és következképpen a tranzitivitás változatlanul teljesül, és az 5.1. tétel is kiterjeszthető erre az esetre. Nyitott kérdés, hogy a tranzitivitás követelménye önmagában vagy esetleg valamilyen egyszerű technikai feltétellel együtt elégséges-e az ilyen vetítési szabályok axiomatikus jellemzéséhez.

Vizsgálatainkban fontos szerepe volt annak a feltevésnek, hogy a számításba vehető vektorok S halmaza vagy a teljes \mathbb{R}^n , vagy ennek a pozitív komponensű vektorokból álló részhalmaza. Kívánatos volna hasonló eredményeket bizonyítani más S alaphalmaz esetén is; ez sikerült arra az esetre, amikor S mindazon vektorok halmaza, melyek komponensei pozitívak és összegük 1, l. [11].

Egyszerű észrevétel, hogy a pozitív komponensű vektorok alkotta S esetre értelmezett π reguláris, lokális kiválasztási szabályok közül bizonyosaknak triviális módon megfeleltethető olyan $\tilde{\pi}$ kiválasztási szabály, melynek alaphalmaza S lezártja. Nevezetesen, ha a 3.1 tétel szerinti $F(\mathbf{v})$ függvény olyan $f_j(v_j)$ függvények összege, melyekre az $f_j(0) = \lim_{v \rightarrow 0} f_j(v)$ határértékek végesek, $F(\mathbf{v})$ folytonos kiterjesztése $\text{cl}(S)$ -re egy $\tilde{\pi}$ kiválasztási szabályt generál $\text{cl}(S)$ affin altereinek (vagyis (2.2) alakú részhalmazainak) halmazán. Természetesen ha L affin altere S -nek, akkor $\text{cl}(L)$ affin altere $\text{cl}(S)$ -nek és

$$\tilde{\pi}(\text{cl}(L)) = \pi(L). \quad (6.2)$$

Az utóbbi észrevételhez kapcsolódik az 5.2. tétel következő általánosítása, a pozitív esetben. Tegyük fel, hogy az adott tranzitív vetítési szabályt generáló (4.5) távolságfüggvényben szereplő φ_j függvényekre a $\varphi_j(0) = \lim_{v \rightarrow 0} \varphi_j(v)$ határértékek végesek. Ekkor az 5.2 tételben a $\bigcap_{i=1}^r L_i \neq \emptyset$ feltétel helyett elég azt feltenni, hogy

$$\bar{L} = \bigcap_{i=1}^r \text{cl}(L_i) \neq \emptyset, \quad (6.3)$$

ha (5.6)-ban $\pi(L|\mathbf{u})$ helyett $\tilde{\pi}(\bar{L}|\mathbf{u})$ -t írunk, ahol $\tilde{\pi}(\bar{L}|\mathbf{u})$ az előző bekezdés szerint értendő. Valóban, $F(\mathbf{v}|\mathbf{u})$ értelmezésének $\mathbf{v} \in \text{cl}(S)$ -re való természetes kiterjesztésével az 5.2 tétel bizonyítása változatlanul alkalmazható, figyelembe véve, hogy az (5.8) Pitagorasz tétel nyilvánvalóan $\mathbf{v} \in \text{cl}(L)$ esetén is érvényes.

Az 5.2 tétel fenti általánosabb változata érvényes speciálisan az I -divergencia vetítésre. Ezt alkalmazva, az (5.2) iteratív arányos illesztési algoritmus konvergenciája nemcsak akkor állítható, ha létezik az (5.1) feltételeket

kielégítő pozitív komponensű \mathbf{v} vektor, hanem az is igaz, hogy az (5.3) limesz létezik és egyenlő \mathbf{u} -nak az (5.1) feltételeket kielégítő nemnegatív komponensű vektorok halmazára vett I -vetületével, ha ez a halmaz nem üres. Hasonlóképp, az (5.16) iteráció mindig konvergál \mathbf{u} -nak az $\mathbf{A}\mathbf{v} = \mathbf{b}$ feltételt kielégítő nemnegatív komponensű vektorok halmazára vett I -vetületéhez, ha ez a halmaz nem üres, akkor is, ha nem tartalmaz pozitív komponensű vektort (az \mathbf{A} mátrixra és a \mathbf{b} vektorra az iteráció definiálása előtt megfogalmazott feltételek mellett). Végül érdekességként bizonyítás nélkül megemlítjük, hogy az (5.16) iteráció akkor is konvergens, ha nem létezik az $\mathbf{A}\mathbf{v} = \mathbf{b}$ feltételt kielégítő nemnegatív komponensű vektor. Ekkor $\mathbf{v}(t)$ az \mathbf{u} -nak azon nemnegatív komponensű \mathbf{v} vektorok halmazára vett I -vetületéhez tart, melyekre az $I(\mathbf{A}\mathbf{v}|\mathbf{b})$ I -divergencia minimális.

Irodalom

1. Aczél, J. – Daróczy, Z. *On Measures of Information and Their Characterizations*. Academic, New York, 1975.
2. Boltzmann, L. Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärme Gleichgewicht. *Wien. Ber.* 76 (1877) 373–435.
3. Bregman, L. M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. and Math. Phys.* 7 (1967) 200–217.
4. Censor, Y. Finite series-expansion reconstruction methods. *Proc. IEEE* 71 (1983) 409–419.
5. Censor, Y., De Pierro, A., Elfving, V., Herman, G. T., Iusem, A. N. On iterative methods for linearly constrained entropy maximization. *Numerical Analysis and Mathematical Modelling*. Banach Center Publications, Vol. 24, 145–163. Polish Scientific Publishers, Warsaw 1990.
6. Csiszár, I. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publ. Math. Inst. Hungar. Acad. Sci.* 8 (1963) 85–108.
7. Csiszár, I. I-divergence geometry of probability distributions and minimization problems. *Ann. Probab.* 3 (1975) 146–158.
8. Csiszár, I. Sanov property, generalized I-projection, and a conditional limit theorem. *Ann. Probab.* 12 (1984) 768–793.
9. Csiszár, I. An extended maximum entropy principle and a Bayesian justification. *Bayesian Statistics 2* (Eds. J. M. Bernardo etc.), 83–89. North Holland, Amsterdam, 1985.
10. Csiszár, I. A geometric interpretation of Darroch and Ratcliff's generalized iterative scaling. *Ann. Statist.* 17 (1989) 1409–1413.
11. Csiszár, I. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Ann. Statist.* 19 (1991), 2032–2066.

12. Csiszár, I. New axiomatic results on inference for inverse problems. *Studia Sci. Math. Hungar.* 26 (1991) 207–237.
13. Csiszár, I. – Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Akadémiai Kiadó, Budapest – Academic, New York, 1981.
14. Darroch, J. N. – Ratcliff, D. Generalized iterative scaling for log-linear models. *Ann. Math. Statist.* 43 (1972) 1470–1480.
15. Deming, W. E. – Stephan, F. F. On a least square adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.* 11 (1940) 427–444.
16. Diaconis, P. – Zabell, S. L. Updating subjective probability. *J. Amer. Statist. Assoc.* 7 (1982) 831–834.
17. Herman, G. T. – Lent, A. Iterative reconstruction algorithms. *Computers in Biology and Medicine* 6 (1976) 273–294.
18. Itakura, F. – Saito, S. Analysis synthesis telephony based on the maximum likelihood method. *Reports of the Sixth International Congress on Acoustics* (Ed. Y. Kohasi), 17–20. Tokyo, Japan, 1968.
19. Jaynes, E. T. On the rationale of maximum entropy methods. *Proc. IEEE* 70 (1982) 939–952.
20. Jones, L. K. – Byrne, C. L. General entropy criteria for inverse problems, with applications to data compression, pattern classification and cluster analysis. *IEEE Trans. Inform. Theory* IT-36 (1990) 23–30.
21. Jones, L. – Trutzu, V. Computationally feasible high-resolution minimum-distance procedures which extend the maximum-entropy method. *Inverse Problems* 5 (1989) 749–766.
22. Kruihof, J. Telefoon verkeerrekening. *Het Engenieur* 52 (1937) E15–E16.
23. Kullback, S. *Information Theory and Statistics*. Wiley, New York, 1959.
24. Liese, F. – Vajda, I. *Convex Statistical Distances*. Teubner, Leipzig, 1987.
25. Paris, J. B. – Vencovská, A. A note on the inevitability of maximum entropy. *Internat. J. Inexact Reasoning* 4 (1990) 183–223.
26. Shannon, C. E. A mathematical theory of communication. *Bell System Tech. J.* 27 (1948) 379–423, 623–656.
27. Shore, J. E. – Johnson, R. W. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inform. Theory* IT-26 (1980) 26–37. Correction: IT-29 (1983) 942–943.
28. Skilling, J. The axioms of maximum entropy. *Maximum Entropy and Bayesian Methods in Science and Engineering*, Vol. 1, 173–187. Kluwer, Amsterdam, 1988.

MAXIMUM ENTROPY AND RELATED METHODS: AXIOMS, ALGORITHMS

In this survey, for systems of k linear equations with n unknowns, where $n \geq 3$ is fixed and $k < n$ is arbitrary, we consider rules to select a distinguished solution vector $\mathbf{v} \in \mathbb{R}^n$. Two cases are treated: (i) the positive case, when solution vectors with positive components are assumed to exist and the distinguished \mathbf{v} is selected from those (ii) the real case, when there are no such constraints. Certain natural axioms are imposed on the permissible rules of selection. A set of such axioms imply that the selected \mathbf{v} must be the feasible solution of maximum entropy resp. minimum Euclidean norm, according to the positive resp. real case, or if a prior guess $\mathbf{u} \in \mathbb{R}^n$ is given, the feasible solution closest to \mathbf{u} in the sense of I-divergence (1.8) resp. of Euclidean distance. Some weaker axioms permit using other distances, viz. the f-divergences (4.4) or the Bregman distances (4.5). Out of algorithms for minimizing I-divergence, we consider interactive scaling (RAS) and generalized iterative scaling (SMART). Both are particular cases of iterative Bregman distance minimization, whose convergence follows from the Pythagorean theorem for Bregman distances.

