MAXIMÁLKORRELÁCIÓ

Szakdolgozat

Írta : Fegyverneki Tamás
Matematika BSc
Alkalmazott matematikus szakirány

Témavezető :
Móri Tamás, egyetemi docens
Valószínűségelméleti és Statisztika Tanszék
Eötvös Loránd Tudományegyetem
Természettudományi Kar

Eötvös Loránd Tudományegyetem
Természettudományi Kar
2010

Tartalomjegyzék

1. fejezet

Bevezetés

Valószínűségi változók vizsgálatának fontosságát nem szükséges hangsúlyoznunk. Az előrejelzések fontos eszközei alapulnak különböző mért adatok, jellemzők változásainak megfigyelésén és ezek összefüggéseinek vizsgálatán, legyen szó gazdasági folyamatok összejátszásáról, tünetegyüttesek közös megjelenéséről vagy akár az időjárás előrejelzéséről. A mért ismeretek statisztikai feldolgozásánál természetesen nem csak a hipotetikus eloszlások önálló tulajdonságai érdekesek, hanem egymásra gyakorolt hatásuk, összefüggésük mértéke is. A kölcsönös, vagy éppen nem feltétlenül kölcsönös függés nagyságának kifejezésére a valószínűségszámítás és matematikai statisztika fejlődése közben számos jó és kevésbé jó, használható és kevésbé használható eszközt definiáltak. Ilyen összefüggőségi mérőszámok megalkotásában is vitathatatlan az – egyébként a világ első egyetemi statisztikai tanszékét megalapító – Karl Pearson munkássága, aki speciális alakban több ilyen, később a teljesség igénye nélkül ismertetett mennyiséget definiált, legalábbis speciális alakban. Az ő nevéhez fűződik a napjainkban is legelterjedtebb és legismertebb mérőszám, a korrelációs együttható, vagy egyszerűen korreláció megalkotása is, amely, mint ismert, a valószínűségi vátozók közötti lineáris összefüggés mértéke.

Ezt a mérőszámot használjuk leggyakrabban, az alkalmazás számos területén, holott hátrányai közismertek, gyakorlatban és az elméletben is természetesnek tűnő elvárásokat nem teljesít. Jelen dolgozat célja egy, a korreláció hiányosságait áthidaló mérőszám, a maximálkorreláció ismertetése, valamint néhány tulajdonságának körüljárása. A maximálkorreláció alapötlete Hans Gebelein *Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung* című 1941-es cikkéből származik, ahol speciális esetekben már definiált a mennyiség. A maximálkorreláció megfelel a "jó" mérőszámokról alkotott elképzeléseinknek, viszont általánosságban akadnak problémák kiszámíthatóságával. Látni fogjuk, hogy pontos értékének meghatározása korántsem olyan egyszerű, mint a klasszikus Pearson-féle korreláció esetében, ez az oka annak, hogy bár sok szempontból pontosabb képet ad a vizsgálandó összefüggések mértékéről, nem terjedt el a gyakorlatban. Mindenesetre a cél adott, minél jobb eszközöket alkotni a kiszámításra, vagy legalábbis a maximálkorreláció értékének jó becslésére vagy közelítésére, hogy elvi előnyei mellett alkalmazásszinten is használható legyen, jelentősen javítva ezzel sta-tisztikai vizsgálati, előrejelzési lehetőségeinket.

A dolgozat első részében röviden áttekintjük a használatos fogalmak egy részét, bevezetve későbbi jelöléseinket is, majd az összefüggőségi mérőszámok kérdését járjuk körül oly módon, hogy Rényi Alfréd[13] nyomán rögzítjük a velük szemben támasztott alapvető követelményeinket, majd néhány lehetséges mérőszámot tárgyalunk, vizsgálva ezek viszonyát elvárásainkkal. Természetesen a dolgozatban említetteknél sokkal több használatos mérőszám létezik összefüggés mérésére, elég például a Spearman-féle vagy a Kendall-féle rangkorrelációra gondolnunk. Az itt ismertetett mérőszámok a maximálkorreláció fogalmának kialakulásához vezető út egy-egy lépcsőfokának tekinthetők, másrészt a maximálkorreláció kiszámításához tartozó eszközök tárgyalásakor kapnak szerepet. A fejezet utolsó részében tovább bővítjük a felhasználandó eszköztárat a később használandó operátorok rövid vizsgálatával.

A következő fejezetben definiáljuk szintén Rényi nyomán a maximálkorrelációt, olyan esetekre is tárgyalva, amelyeket Gebelein még nem dolgozott fel. A fejezet első szakaszában a definiáláson túl belátjuk, hogy a maximálkorreláció egy jó mérőszám, olyan értelemben, hogy tényleg teljesíti elvi elvárásainkat. Ezek után a kiszámításhoz hasznos tulajdonságokat tárgyalunk, visszavezetve operátoregyenletek megoldásaira a problémát. Látni fogjuk, hogy a maximálkorreláció kiszámításához sok esetben adhat segítséget operátorok sajátértékeinek és sajátfüggvényeinek meghatározása. Az említett eszközök közül nem mindegyik használatára fogunk példát mutatni, van közöttük, amely egyszerűen csak hasznos segítségnek tűnhet egy-egy felmerülő speciális gyakorlati probléma esetén.

Ezek után a felépített módszerek egy részének segítségével bemutatunk néhány konkrét esetet, amikor a maximálkorreláció értékét pontosan meg tudjuk határozni, leginkább Csáki Péter és Fischer János[4][5][6] munkásságát felhasználva, valamint példát említve Székely Gábor és Móri Tamás[17] nyomán, legvégül pedig említésszinten bemutatunk két vázlatos algoritmust a fent említett optimális sajátfüggvények kiszámítására, melyek közül az egyik sokkal általánosabb esetben is működik függvénytérbeli legjobb transzformáló függvények közelítésére.

2. fejezet

A problémakör

2.1. Matematikai eszköztár

A fejezetben elsőként áttekintjük a továbbiakban használatos jelöléseket és terminológiát, majd az összefüggőségi mértékek vizsgálatának egy speciális funkcionálanalízisbeli megközelítésmódját taglaljuk, nevezetesen a probléma visszavezetését operátorok vizsgálatára.

Legyen $[\Omega, \mathcal{A}, P]$ egy valószínűségi mező, azaz Ω egy nemüres eseménytér, $\mathcal{A}$ Ω bizonyos részhalmazaiból álló σ-algebra, P pedig egy valószínűségi mérték $\mathcal{A}$-n.

Egy ξ valószínűségi változó várható értékét és szórását jelölje rendre $E\xi$ és $D\xi$. Ha ξ várható értéke és szórása létezik, valamint $D\xi > 0$, akkor legyen

$$\xi^* = \frac{\xi - E\xi}{D\xi} \tag{2.1}$$

a ξ valószínűségi változó standardizáltja. Egy valószínűségi változót standardnak nevezünk, ha $\xi = \xi^*$ teljesül rá.

A ξ és η valószínűségi változók együttes eloszlását jelölje $\mathbf{Q}_{\xi,\eta}$, eloszlásaikat rendre $\mathbf{Q}_\xi$ és $\mathbf{Q}_\eta$, valamint eloszlásaik direkt szorzatát $\mathbf{Q}_\xi \times \mathbf{Q}_\eta$, tehát a valós számegyenes bármely A és B részhalmazára

$$\mathbf{Q}_\xi \times \mathbf{Q}_\eta(A \times B) = P(\xi \in A)P(\eta \in B) \tag{2.2}$$

ahol $A \times B$ a halmazok direkt szorzata. A fenti definíció a mértékkiterjesztési tétel értelmében kiterjeszthető a sík Borel-halmazaira az ismert módon.

A ξ és az η közötti összefüggőséget *regulárisnak* nevezzük, ha az együttes eloszlásuk abszolút folytonos az eloszlásaik direkt szorzatára nézve, tehát ha $\mathbf{Q}_{\xi,\eta} \ll \mathbf{Q}_\xi \times \mathbf{Q}_\eta$.

$L^2 = L^2(\Omega, \mathcal{A}, P)$ jelenti azon ζ valószínűségi változóknak a terét, amelyekre $E(\zeta^2)$ véges. Ez Hilbert-tér, amelyben $\langle \zeta_1, \zeta_2 \rangle = E(\zeta_1 \zeta_2)$ a skaláris szorzata $\zeta \in L^2$-nek és $\zeta \in L^2$-nek, tehát $\|\zeta\| = \sqrt{E(\zeta^2)}$ adódik $\zeta \in L^2$ normájaként. Könnyen látható, hogy ez valóban skaláris szorzat és norma.

A ξ valószínűségi változó eloszlásfüggvénye legyen $F(x)$, ekkor legyen

$$L_F^2 = \left\{ f(x) : \int_{-\infty}^{\infty} f^2(x) dF(x) < \infty \right\}! \tag{2.3}$$

L_F^2 szeparábilis Hilbert-teret alkot az alábbi skaláris szorzattal :

$$\langle f_1(x), f_2(x) \rangle = \int_{-\infty}^{\infty} f_1(x) f_2(x) dF(x). \tag{2.4}$$

Legyen $f(x) \in L_F^2$, ekkor az $f = f(\xi)$ valószínűségi változóra $f \in L^2$. Az ilyen alakú valószínűségi változók terét jelölje L_ξ^2! Nyilvánvalóan ekkor fennáll $L_\xi^2 = L^2(\Omega, \mathcal{A}_\xi, P)$, ahol $\mathcal{A}_\xi$ jelöli a legszűkebb olyan σ-algebrát, ahol ξ mérhető. Funkcionálanalízisből ismert, hogy létezik kölcsönösen egyértelmű megfeleltetés a most definiált L_ξ^2 és L_F^2 terek között, amely skalárszorzattartó és természetesen normatartó is, így L_ξ^2 is szeparábilis Hilbert-tér lesz, amely ξ véges értékkészlete esetén véges dimenziós, speciálisan ha ξ egy diszkrét valószínűségi változó, amely n különböző értéket vehet fel, akkor L_ξ^2 euklideszi tér lesz, amely n-dimenziós.

Valójában a fent definiált terek valószínűségi változók és függvények ekvivalenciaosztályainak terei, L^2 és L_ξ^2 terekben a P-majdnem mindenütt egyenlőség, L_F^2 esetén pedig az F-majdnem mindenütt egyenlőség jelenti az ekvivalenciarelációt.

Jelölje L_0^2 az L^2-beli, 0 várható értékű és véges szórású valószínűségi változók alterét, és analóg módon jelölje $L_{F,0}^2$ és $L_{\xi,0}^2$ a megfelelő altereket!

Minden L^2-beli ζ valószínűségi változó egyértelműen felbontható $\zeta' + \zeta''$ alakban, ahol $\zeta' \in L^2_\xi$ és ζ'' ortogonális az L^2_ξ térre, tehát létezik pontosan egy $\zeta' \in L^2_\xi$, hogy

$$\forall f \in L^2_\xi \quad \langle \zeta', f \rangle = \langle \zeta, f \rangle \tag{2.5}$$

Ekkor természetesen $\zeta'' = \zeta - \zeta'$, és a merőleges vetületre teljesül, hogy $\min \|\zeta - f\| = \|\zeta - \zeta'\|$. Legyen $\mathbf{A}_\xi$ az L^2_ξ zárt altérre való ortogonális projekció. Ekkor minden $\zeta \in L^2$, $\ f \in L^2_\xi$ esetén

$$\langle \mathbf{A}_\xi \zeta, f \rangle = \langle \zeta, f \rangle. \tag{2.6}$$

A fenti ζ esetében $\mathbf{A}_\xi \zeta = E(\zeta|\xi)$ 1 valószínűséggel, azaz a ζ feltételes várható értéke ξ-re megegyezik ζ L^2_ξ-re vett ortogonális projekciójával, ugyanis ha veszünk egy $A \in \mathcal{A}_\xi$ halmazt, tehát létezik egy B Borel-halmaz, hogy $A = \{\xi \in B\}$, akkor speciálisan az

$$f = \chi_A = \begin{cases} 1, & \xi \in B \\ 0, & \xi \notin B \end{cases} \tag{2.7}$$

függvényre felhasználva (2.6) egyenlőséget:

$$\int_A \mathbf{A}_\xi \zeta dP = \int_A \zeta dP, \tag{2.8}$$

amiből következik, hogy $\mathbf{A}_\xi \zeta = E(\zeta|\xi)$ 1 valószínűséggel.

2.2. Összefüggőségi mérőszámok

Legyenek ξ és η valószínűségi változók az $[\Omega, \mathcal{A}, P]$ valószínűségi mezőn, mégpedig úgy, hogy egyikük se legyen P-majdnem mindenütt konstans! A statisztikai vizsgálatok szinte minden területén gyakran felmerülő probléma a két valószínűségi változó összefüggőségének, annak erősségének numerikus karakterizációja. Erre számos ismert és gyakorlatban is használatos mérőszám áll rendelkezésünkre. Egy ilyen definiálásánál viszont szükségszerűnek tűnik néhány megszorítást bevezetnünk, amelyek alapján eldönthető egy ilyen mennyiségről, mennyire jó, azaz mennyire ad pontos információt az összefüggés mértékéről. A követelményeket Rényi Alfréd[13] nevéhez köthetjük, mint egyfajta ésszerű elvárásokat az összefüggőségi mérőszámokkal szemben. Jelöljön esetünkben $\delta(\xi, \eta)$ egy ilyen mérőszámot!

Természetesnek tűnik, hogy ezt a [0,1] intervallumba eső értékekkel definiáljuk, az 1 értéket a szoros összefüggés, a 0 értéket a teljes függetlenség jellemzőjének tekintve. A $\delta(\xi, \eta)$-ra vonatkozó követelmények az alábbiak:

A) $\delta(\xi, \eta)$ legyen definiálva minden olyan ξ, η valószínűségiváltozó-párra, amelyek nem 1 valószínűséggel konstansok
B) $\delta(\xi, \eta) = \delta(\eta, \xi)$, azaz legyen szimmetrikus
C) $0 \leq \delta(\xi, \eta) \leq 1$
D) $\delta(\xi, \eta) = 0$ akkor és csak akkor, ha ξ és η függetlenek
E) $\delta(\xi, \eta) = 1$, ha ξ és η között egyenes összefüggés van, például ha $\xi = g(\eta)$ vagy $\eta = f(\xi)$ teljesül, ahol $f(x)$ és $g(x)$ Borel-mérhető függvények
F) Ha az $f(x)$ és $g(x)$ Borel-mérhető függvények kölcsönösen egyértelmű leképezések a valós tengelyről saját magára, akkor $\delta(f(\xi), g(\eta)) = \delta(\xi, \eta)$
G) Ha ξ és η együttes eloszlása normális, akkor $\delta(\xi, \eta) = |\mathbf{R}(\xi, \eta)|$, ahol $\mathbf{R}(\xi, \eta)$ a két valószínűségi változó korrelációs együtthatója

Az E) jelű feltétel esetében felmerülhet, hogy miért nem követelünk meg az D) esethez hasonlóan szükségességet és elégségességet is, ahogy az természetesnek tűnne. Ezt az elvárást Rényi túlságosan korlátozó erejűnek vélte, ezért vetette el.

A továbbiakban röviden áttekintjük – a teljesség igénye nélkül – az ismert és használatos mérőszámokat, valamint hogy mennyire felelnek meg a Rényi-féle követelményeknek.

2.2.1. A korrelációs együttható

Tegyük fel, hogy $D\xi$ és $D\eta$ pozitív és véges, ekkor a klasszikus Pearson-féle korrelációs együttható :

$$\mathbf{R}(\xi, \eta) = \frac{E\big((\xi - E\xi)(\eta - E\eta)\big)}{D\xi D\eta} = E(\xi^* \eta^*) \tag{2.9}$$

Jól ismert, hogy a korrelációs együttható lineáris esetekben használható jól, mégis szokás bármilyen összefüggésben használni. A fenti elvárásoknak való megfelelés vizsgálatakor látható, hogy mint köztudott, a korreláció a [-1, 1] intervallumban veszi fel értékeit, tehát a C) pontnak csak az abszolútértéke felel meg, így eleve csak azt érdemes megnéznünk. Az abszolútérték nyilvánvalóan teljesíti a B), C) és G) követelményeket, de a többi pontot nem. Nyilván nem definiáljuk minden A)-beli esetben, hiszen csak pozitív és véges szórású valószínűségi változók esetében értelmeztük. Látható, hogy nem csak függetlenség esetében lesz 0, hiszen például ha ξ standard normális eloszlású, $\eta = \xi^2$, akkor $\mathbf{R}(\xi, \eta) = 0$, hiszen a standard normális eloszlás esetén a harmadik momentum 0. Ez jó példa arra is, hogy bár egyértelmű függvénykapcsolat van a két valószínűségi változó között, a korreláció mégsem lesz 1. Szintén jól ismert, hogy $|\mathbf{R}(\xi, \eta)| = 1$ akkor és csak akkor, ha lineáris kapcsolat áll fenn.

2.2.2. A korrelációs hányados

Tegyük fel, hogy η szórása véges és pozitív, ekkor az η valószínűségi változó ξ-re vett korrelációs hányadosa vagy korrelációs aránya

$$\Theta_\xi(\eta) = \frac{D(E(\eta|\xi))}{D\eta} \tag{2.10}$$

E mennyiség bevezetése speciális alakban szintén Pearson nevéhez fűződik, általánosan pedig Kolmogorov nevéhez köthető. Értékeit a [0,1] intervallumból veszi fel, de nem szimmetrikus, valamint a $\Theta_\xi(\eta) = 0$ egyenlőségből nem következik ξ és η függetlensége (lásd [14]). Természetesen ebből a mennyiségből származtatható szimmetrikus mérőszám,

$$\Theta(\xi, \eta) = \max(\Theta_\xi(\eta), \Theta_\eta(\xi)), \tag{2.11}$$

amely viszont továbbra sem teljesíti kimerítően Rényi feltételeit (lásd [13]). Belátható, hogy $\Theta_\xi(\eta)$ és $\Theta(\xi, \eta)$ teljesíti a G) jelű követelményt, valamint $\Theta_\xi(\eta) = \sup_{(f)} \mathbf{R}(f(\xi), \eta)$, ahol $f = f(x)$ végigfut az összes Borel-mérhető függvényen, amelynek szórásnégyzete véges és pozitív. Belátható továbbá (lásd [12]), hogy minden esetben

$$\exists f_0(x) : 0 < D^2(f_0) < \infty; \quad \Theta_\xi(\eta) = \mathbf{R}(f_0(\xi), \eta). \tag{2.12}$$

Felhasználva a korábban tárgyalt megfeleltetést a feltételes várható érték és az ortogonális projekció között, adódik, hogy

$$\Theta_\xi(\eta) = \|\mathbf{A}_\xi \eta^*\|, \tag{2.13}$$

innen pedig

$$\Theta_\xi^2(\eta) = \langle \mathbf{A}_\xi \eta^*, \mathbf{A}_\xi \eta^* \rangle = \langle \eta^*, \mathbf{A}_\xi \eta^* \rangle. \tag{2.14}$$

A két egyenletet elosztva egymással

$$\Theta_\xi(\eta) = \frac{\langle \eta^*, \mathbf{A}_\xi \eta^* \rangle}{\|\mathbf{A}_\xi \eta^*\|} = \mathbf{R}(\eta, \mathbf{A}_\xi \eta). \tag{2.15}$$

Amennyiben $\eta \in L^2$ és $f \in L^2_\xi$ teljesülnek, (2.6) egyenlőség miatt

$$\|\mathbf{A}_\xi \eta^* - \langle f^*, \eta^* \rangle f^*\|^2 = \|\mathbf{A}_\xi \eta^*\|^2 - \langle f^*, \eta^* \rangle^2, \tag{2.16}$$

így

$$\Theta_\xi^2(\eta) = \langle f^*, \eta^* \rangle^2 + \|\mathbf{A}_\xi \eta^* - \langle f^*, \eta^* \rangle f^*\|^2. \tag{2.17}$$

A korrelációs hányados ilyen megközelítésű vizsgálatának jelentős szerepe van a kétváltozós sztochasztikus kapcsolatok funkcionálanalízisbeli eszközökkel való vizsgálatában.

2.2.3. A kontingencia

A kontingenciát diszkrét valószínűségi változókra a korábbiakhoz hasonlóan Pearson vezette be, általánosan Rényi[13]. Ha a ξ és η közötti összefüggőség reguláris, akkor a Radon-Nikodym-tétel értelmében létezik egy Borel-mérhető $k(x,y) = \frac{d\mathbf{Q}_{\xi,\eta}}{d\mathbf{Q}_{\xi\times\eta}}$, hogy minden A és B Borel-halmazára a számegyenesnek

$$P(\xi \in A, \eta \in B) = \mathbf{Q}_{\xi,\eta}(A \times B) = \int_{x\in A}\int_{y\in B} k(x,y)dF(x)dG(y) \qquad (2.18)$$

ahol $F(x)$ és $G(y)$ rendre ξ és η eloszlásfüggvényei, $\mathbf{Q}_{\xi,\eta}$ az együttes eloszlás.

Ekkor legyen a két valószínűségi változó kontingenciája

$$\mathbf{C}(\xi,\eta) = \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\big(k(x,y)-1\big)^2 dF(x)dG(y)\right)^{\frac{1}{2}} \qquad (2.19)$$

A kontingencia szintén az összefüggés egy mérőszámának tekinthető, a B), D) és F) jelű tulajdonságokat teljesíti is. Értékkészlete a $[0,+\infty]$ tartomány, de ez egyszerűen transzformálható [0,1]-re, például ha $\mathbf{C}$ helyett a $\frac{\mathbf{C}(\xi,\eta)}{\sqrt{1+\mathbf{C}^2(\xi,\eta)}}$ mennyiséget tekintjük, ez teljesíti C) és G) pontokat is, de A) és E) továbbra sem teljesül.

Szintén a korábban tárgyalt vetítési operátorok segítségével definiálható a kontingencia a következő módon is :

Legyen f_n és g_n teljes ortonormált rendszer $L^2_{\xi,0}$-ban és $L^2_{\eta,0}$-ban! Ekkor definiálhatjuk a következő normákat:

$$|||\mathbf{A}_\eta|||^2 = \sum_i\sum_k \langle \mathbf{A}_\eta f_i, g_k\rangle^2 \qquad (2.20)$$

$$|||\mathbf{A}_\xi|||^2 = \sum_i\sum_k \langle f_i, \mathbf{A}_\xi g_k\rangle^2, \qquad (2.21)$$

ahol az $\mathbf{A}_\xi$ és $\mathbf{A}_\eta$ operátorok értelmezési tartományai rendre az L^2_η és L^2_ξ terek. Ekkor a kontingencia :

$$\mathbf{C}(\xi,\eta) = |||\mathbf{A}_\xi||| = |||\mathbf{A}_\eta||| \qquad (2.22)$$

Belátható, hogy ez a definíció megegyezik a korábbival (lásd [4]). A korrelációs hányados esetéhez hasonlóan ez a tárgyalásmód szintén lényeges szerepet kap az összefüggőség operátor-sajátértékekre vonatkozó problémaként való tárgyalásakor.

A kontingencia, a korrelációs hányados és a korreláció között fennáll a következő egyenlőtlenség (lásd [14]):

$$0 \le |\mathbf{R}(\xi,\eta)| \le \min\big(\Theta_\xi(\eta),\Theta_\eta(\xi)\big) \le \max\big(\Theta_\xi(\eta),\Theta_\eta(\xi)\big) \le \mathbf{C}(\xi,\eta) \quad (2.23)$$

Itt érdemes megjegyezni, hogy információelméleti meggondolások alapján Linfoot definiált egy összefüggőségi mérőszámot, az $\mathbf{L}(\xi,\eta) = (1-e^{-2\mathbf{I}(\xi,\eta)})^{\frac{1}{2}}$ mennyiséget, ahol $\mathbf{I}(\xi,\eta)$ a valószínűségi változók egymásra vonatkozó információtartalmát jelöli. Ha a ξ és η közötti kapcsolat reguláris, akkor $\mathbf{I}(\xi,\eta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} k(x,y)\log k(x,y)dF(x)dG(y)$ összefüggés igaz. Ez a mérőszám Rényi összes követelményét teljesíti.

2.3. Néhány tulajdonság az operátorokról

Ebben a szakaszban áttekintünk néhány eredményt, amelyek a későbbiekben hasznos eszközöket szolgáltatnak az összefüggőség és kiszámíthatóság problémájának operátor-sajátértékekre való visszavezetésekor.

Legyen ξ és η együttes eloszlásfüggvénye $F(x,y)$, amely meghatározza a P valószínűséget a síkon! Hasonló módon a peremeloszlásokhoz tartozó eloszlásfüggvények legyenek $F_1(x)$ és $F_2(y)$, amelyek rendre a P_1 és P_2 valószínűségi mértékeket határozzák meg a valós számegyenesen. A korábban tárgyalt értelemben L^2_ξ és $L^2_{F_1}$ izomorfak, hasonlóan L^2_η és $L^2_{F_2}$ is.

A korábban definiált $\mathbf{A}_\eta$ operátornak megfeleltethető egy $\mathbf{A}_1$ operátor, amely $L^2_{F_1}$-ből $L^2_{F_2}$-be képez, mégpedig ha $f(x) \in L^2_{F_1}$ és $f = f(\xi) \in L^2_\xi$, akkor a $g = \mathbf{A}_\eta f$ valószínűségi változóhoz egyértelműen létezik $g(y) \in L^2_{F_2}$, amelyre $g = g(\eta)$. Ekkor $\mathbf{A}_1$ az az operátor, ami $f(x)$-et $g(y)$-ba viszi. Teljesen hasonló módon $\mathbf{A}_\xi$-hez is létezik $\mathbf{A}_2$, ami $L^2_{F_2}$ elemeit transzformálja $L^2_{F_1}$-beliekké. A konstrukcióból látható, hogy analóg módon adódnak $\mathbf{A}_\xi$ és

$\mathbf{A}_\eta$ tulajdonságai $\mathbf{A}_2$-re és $\mathbf{A}_1$-re is. A kétváltozós függőség problémakörének néhány eszközéhez szükséges annak vizsgálata, hogy mikor lesz az imént definiált két operátor integráloperátor. Csáki és Fischer[4] a következőt látták be :

1.Állítás. *A fentiekben definiált $\boldsymbol{A}_1$ és $\boldsymbol{A}_2$ operátorok akkor és csak akkor integráloperátorok, ha $P \ll P_1 \times P_2$. Ebben az esetben létezik $k(x,y)$ függvény, amelyre*

$$P(B) = \iint_B k(x,y) dP_1 dP_2, \tag{2.24}$$

ahol B a sík egy tetszőleges mérhető halmaza; ezen kívül

$$\mathbf{A}_1 f(x) = \int_{-\infty}^{\infty} k(x,y) f(x) dF_1(x) \tag{2.25}$$

és

$$\mathbf{A}_2 g(y) = \int_{-\infty}^{\infty} k(x,y) g(y) dF_2(x) \tag{2.26}$$

teljesül.

Az 1.Állítás [4]-ben tárgyalt bizonyítása azt is megmutatja, hogy a tárgyalt operátorok közül vagy mindkettő integráloperátor vagy egyik sem. Az állítás-nak továbbá lényeges szerepe van annak bizonyításakor is, hogy a kontingencia két tárgyalt definíciója ekvivalens:

Legyenek $\{f_0(x), f_1(x), ...\}$ és $\{g_0(y), g_1(y), ...\}$ teljes ortonormált rendszerek rendre $L^2_{F_1}$-ben és $L^2_{F_2}$-ben úgy, hogy $f_0(x) \equiv g_0(y) \equiv 1$ teljesül! Ekkor $\{f_1(x), f_2(x), ...\}$ és $\{g_1(y), g_2(y), ...\}$ teljes ortonormált rendszerek $L^2_{F_1,0}$-ban és $L^2_{F_2,0}$-ban. Ekkor

$$\mathbf{C}^2(\xi,\eta) = |||\mathbf{A}_\eta|||^2 = \sum_{i\geq 1}\sum_{k\geq 1} \langle \mathbf{A}_1 f_i(x), g_k(y)\rangle^2 = \sum_{i\geq 0}\sum_{k\geq 0} \langle \mathbf{A}_1 f_i(x), g_k(y)\rangle^2 - 1 \tag{2.27}$$

teljesül, és $i \geq 1$ esetben

$$\langle \mathbf{A}_1 f_i(x), g_0(y)\rangle = \langle f_i(x), \mathbf{A}_2 g_0(y)\rangle = \langle f_i(x), f_0(y)\rangle = 0, \tag{2.28}$$

hasonlóan $k \geq 1$ esetben

$$\langle \mathbf{A}_1 f_0(x), g_k(y)\rangle = 0, \tag{2.29}$$

továbbá

$$\langle \mathbf{A}_1 f_0(x), g_0(y)\rangle = 1. \tag{2.30}$$

Ha $P \ll P_1 \times P_2$, az 1. Állítás szerint

$$|||\mathbf{A}_1|||^2 = \int_{-\infty}^{\infty} k^2(x,y) f(x) dF_1(x) = \sum_{i\geq 0}\sum_{k\geq 0} \langle \mathbf{A}_1 f_i(x), g_k(y)\rangle^2 \tag{2.31}$$

teljesül, ekkor viszont (2.27)-ből és (2.31)-ből azonnal adódik, hogy

$$\mathbf{C}^2(\xi,\eta) = |||\mathbf{A}_\eta|||^2 = |||\mathbf{A}_1|||^2 - 1, \tag{2.32}$$

és ez pontosan ugyanazt adja, mint (2.19).
Amennyiben $|||\mathbf{A}_1||| < \infty$ teljesül, akkor $\mathbf{A}_1$ egy folytonos integráloperátor, így az 1.Állítás szerint $P < P_1 \times P_2$ teljesül, ha nem véges, akkor pedig $\mathbf{C}(\xi,\eta) = \infty$, ami szintén a másik definíciót adja vissza. $\square$

Legyen P a már definiált valószínűség a síkon, amelyet az együttes eloszlás határoz meg, ekkor legyen a valós számegyenes egy rögzített mérhető A halmazára $P_A(B) = P(A \times B)$, ahol B egy tetszőleges mérhető halmaz. Hasonlóan definiáljuk rögzített B melletti tetszőleges A mérhető halmazokra $P^B(A) = P(A \times B)$-t! Mivel

$$P_A(B) \leq P_2(B); \quad P^B(A) \leq P_1(A), \tag{2.33}$$

következik, hogy $P_A \ll P_2$ és $P^B \ll P_1$. A Radon-Nikodym–tétel értelmében az ismert módon létezik $P_1(A|y)$ és $P_2(B|x)$, hogy

$$\begin{cases} P_A(B) = \int_B P_1(A|y) dP_2 \\ P^B(A) = \int_A P_2(B|x) dP_1. \end{cases} \tag{2.34}$$

Legyen $\chi_A(x)$ az A mérhető halmaz indikátora, ekkor

$$\int_B \mathbf{A}_1 \chi_A(x) dP_2 = \int_{-\infty}^{\infty} \chi_B(y) \mathbf{A}_1 \chi_A(x) dF_2(y) = \langle \mathbf{A}_1 \chi_A(x), \chi_B(y)\rangle = P_A(B), \tag{2.35}$$

tehát

$$\mathbf{A}_1\chi_A(x) = P_1(A|y) \tag{2.36}$$

P_2-majdnem minden y-ra. Hasonlóan P_1-majdnem minden x-re

$$\mathbf{A}_2\chi_B(y) = P_2(B|x). \tag{2.37}$$

Az abszolút folytonosság miatt, ha $P_1(A) = 0$, akkor

$$P_A(B) = \int_B P_1(A|y)dP_2 = 0, \tag{2.38}$$

amiből $P_1(A|y) = 0$ következik P_2-majdnem mindenütt, viszont ebből nem követezik, hogy $P_1(.|y) \ll P_1(.)$, erre Csáki és Fischer[4] következő állítása ad egy feltételt:

2.Állítás. *$P_1(.|y) \ll P_1(.)$ P_2-majdnem minden y-ra, $P_2(.|x) \ll P_2(.)$ P_1-majdnem minden x-re és $P \ll P_1 \times P_2$ közül vagy mindegyik teljesül, vagy egyik sem.*

Következésképp az 1.Állítás és a 2.Állítás miatt adódik, hogy az alábbi három tulajdonság egyenértékű:

1. $P_1(.|y) \ll P_1(.)$ P_2-majdnem minden y-ra,
2. $P \ll P_1 \times P_2$,
3. $\mathbf{A}_1$ integráloperátor.

Természetesen teljesen hasonló és ekvivalens állítás vonatkozik a P_2 valószínűségi mértékre és $\mathbf{A}_2$ operátorra is.

A fentiekben tárgyalt eszközökkel a továbbiakban már részletes áttekintést tudunk adni az összefüggőség egy olyan mérőszámáról, amely pótolja a gyakorlatban használatos ismertetett mérőszámok hiányosságait, és megfelel a bevezetett követelményeknek.

3. fejezet

A maximálkorreláció

3.1. Definíció és tulajdonságok

3.1.1. Definíció és Rényi követelményei

Korábban láttuk a korrelációs hányados esetében, hogy

$$\Theta_\xi(\eta) = \sup_{(f)} \mathbf{R}(f(\xi), \eta) \tag{3.1}$$

teljesül, így természetszerűleg adódik ennek további javítására a következő mennyiség :

$$\mathbf{S}(\xi, \eta) = \sup_{f,g} \mathbf{R}(f(\xi), g(\eta)), \tag{3.2}$$

ahol $f(x)$ és $g(x)$ végigfut minden olyan Borel-mérhető függvényen, amelyek szórásnégyzete pozitív és véges, azaz

$$\mathbf{S}(\xi, \eta) = \sup_{f \in L^2_\xi, g \in L^2_\eta} \mathbf{R}(f(\xi), g(\eta)) \tag{3.3}$$

Ez a mennyiség a ξ és η valószínűségi változók *maximálkorrelációja.* Ezt a mérőszámot H. Gebelein[8] vezette be olyan esetekben, ahol ξ és η eloszlása egyaránt diszkrét vagy egyaránt abszolút folytonos, általánosabb tárgyalása Rényi[13] és Sarmanov[15][16] nevéhez köthető.

Először belátjuk, hogy a maximálkorreláció rendelkezik a Rényi által előírt optimalitási tulajdonságokkal. Az A) tulajdonság nyilvánvaló, hiszen tetszőleges valószínűségi változókra értelmes a definíció, ugyancsak adódik a szimmetria és a [0,1]-beli értékek felvétele is (B) és C)). A D) tulajdonsághoz, nevezetesen hogy $\mathbf{S}(\xi,\eta)=0$ akkor és csak akkor, ha ξ és η függetlenek, tekintsük a következőket:

Ha a két valószínűségi változó független, a maximálkorreláció nulla a definíció alapján. Megfordítva feltesszük, hogy $\mathbf{S}(\xi,\eta)=0$, ekkor

$$\mathbf{R}(f(\xi),g(\eta))=0 \tag{3.4}$$

minden értelmes f-re és g-re.

Speciálisan válasszuk a következő függvényeket:

$$f_A(x)=\begin{cases}1, & X\in A\\ 0, & \text{egyébként}\end{cases} \qquad g_B(x)=\begin{cases}1, & X\in B\\ 0, & \text{egyébként,}\end{cases} \tag{3.5}$$

ahol A és B tetszőleges rögzített részhalmazai a valós számegyenesnek, és természetesen $0<P(\xi\in A)<1$ és $0<P(\eta\in B)<1$. Ekkor

$$\mathbf{R}(f_A(\xi),g_B(\eta))=\frac{P(\xi\in A,\eta\in B)-P(\xi\in A)P(\eta\in B)}{\sqrt{P(\xi\in A)(1-P(\xi\in A))P(\eta\in B)(1-P(\eta\in B))}}, \tag{3.6}$$

ebből pedig következik, hogy a maximálkorreláció nulla értéke mellett a számláló zérus, azaz tetszőleges A és B esetén

$$P(\xi\in A,\eta\in B)=P(\xi\in A)P(\eta\in B), \tag{3.7}$$

ez pedig pontosan ξ és η függetlenségét jelenti.

Az E) tulajdonság teljesülését vizsgáljuk általánosabban! Tegyük fel, hogy ξ és η között függvénykapcsolat áll fenn $f_0(\xi)=g_0(\eta)$ értelemben, ahol f és g Borel-mérhetőek. Ekkor legyen

$$f_1(\xi)=\frac{f_0(\xi)}{1+|f_0(\xi)|}=\frac{g_0(\eta)}{1+|g_0(\eta)|}=g_1(\eta)! \tag{3.8}$$

Mivel $f_1(\xi)$ és $g_1(\eta)$ korlátosak, szórásnégyzeteik léteznek és pozitívak,

$$\mathbf{R}(f_1(\xi),g_1(\eta))=1, \tag{3.9}$$

ennek speciális eseteként pedig az E) tulajdonság is teljesül. Ugyanakkor ez jól mutatja, hogy ha az eredeti feltételekben Rényi a túlzottan megszorítónak ítélt "akkor és csak akkor" kapcsolatot követeli meg, azt a maximálkorreláció nem teljesítené.

Az F) tulajdonság teljesülése triviális a definícióból, hiszen ha u és v kölcsönösen egyértelmű $\mathbb{R} \to \mathbb{R}$ Borel-mérhető leképezések, akkor

$$\mathbf{S}(u(\xi), v(\eta)) = \sup_{f \in L^2_\xi, g \in L^2_\eta} \mathbf{R}(f(u(\xi)), g(v(\eta))) =$$

$$= \sup_{f \in L^2_\xi, g \in L^2_\eta} \mathbf{R}(f(\xi), g(\eta)) = \mathbf{S}(\xi, \eta) \tag{3.10}$$

teljesül

Az utolsó követelményt a maximálkorreláció szintén teljesíti, azaz ha a két valószínűségi változó együttes eloszlása normális, akkor a maximálkorreláció értéke megegyezik a klasszikus Pearson-féle korreláció abszolútértékével. A tulajdonság bizonyítását már Lancaster[9] is közölte 1957-ben, Dembo, Kagan és Shepp[7] pedig példát közöltek olyan speciális esetre is, amikor az együttes eloszlás nem normális, de a maximálkorreláció szintén azonos a korreláció abszolútértékével. Ezt a példát a speciális esetek között tárgyaljuk. A Rényi-féle G) tulajdonság meglétének egy új bizonyítását közölte Yu[18] 2008-ban, a későbbiekben ezt a bizonyítást a további tulajdonságok között tárgyalt két állítás ismeretében át is tekintjük.

3.1.2. További tulajdonságok

Az előző fejezetben láttuk, hogy az összefüggőségi mérőszámokra vonatkozó Rényi-féle követelményeket a Gebelein által definiált maximálkorreláció hiánytalanul teljesíti. Ezeken az előnyös tulajdonságokon kívül áttekintünk néhány további összefüggést, amelyek hasznosak lehetnek a maximálkorreláció kiszámíthatóságával és közelíthetőségével kapcsolatban.

Tekintsük a korábban tárgyalt $\mathbf{A}_\xi$ és $\mathbf{A}_\eta$ operátorokat, leszűkítve értelmezési tartományukat rendre az $L^2_{\eta,0}$ és $L^2_{\xi,0}$ terekre! Ekkor a következő állítás igaz (lásd [4]):

3.Állítás.

$$\mathbf{S}(\xi,\eta) = \|\mathbf{A}_\eta\| = \|\mathbf{A}_\xi\| \tag{3.11}$$

Bizonyítás.

$$\sup_{f\in L^2_{\xi,0},\|f\|=1} \|\mathbf{A}_\eta f\| = \sup_{f\in L^2_{\xi,0},\|f\|=1} \left\langle \frac{\mathbf{A}_\eta f}{\|\mathbf{A}_\eta f\|}, f \right\rangle \le \mathbf{S}(\xi,\eta) \tag{3.12}$$

teljesül, továbbá ha $f \in L^2_{\xi,0}$, $\|f\| = 1$, $g \in L^2_{\eta,0}$, $\|g\| = 1$, akkor a (2.17) egyenlőség szerint

$$|\langle f, g\rangle| \le \left\langle \frac{\mathbf{A}_\eta f}{\|\mathbf{A}_\eta f\|}, f \right\rangle, \tag{3.13}$$

tehát

$$\mathbf{S}(\xi,\eta) \le \sup_{f\in L^2_{\xi,0},\|f\|=1} \left\langle \frac{\mathbf{A}_\eta f}{\|\mathbf{A}_\eta f\|}, f \right\rangle, \tag{3.14}$$

így az egyenlőség teljesül,

$$\mathbf{S}(\xi,\eta) = \sup_{f\in L^2_{\xi,0},\|f\|=1} \|\mathbf{A}_\eta f\| = \|\mathbf{A}_\eta\|. \tag{3.15}$$

Teljesen hasonlóan $\mathbf{A}_\xi$ esetén is igaz. □

Az előző állításhoz hasonlót bizonyít Rényi[13] is:

4.Állítás. *Az $L^2_{\xi,0}$ térben értelmezett $\mathbf{A}_\xi\mathbf{A}_\eta$ operátor önadjungált, pozitív szemidefinit és*

$$\mathbf{S}^2(\xi,\eta) = \sup_{f\in L^2_{\xi,0},\|f\|=1} \langle \mathbf{A}_\xi\mathbf{A}_\eta f, f\rangle = \|\mathbf{A}_\xi\mathbf{A}_\eta\|. \tag{3.16}$$

Az állítás bizonyítását a maximálkorreláció kiszámításásra vonatkozó elégséges feltételek tárgyalásánál tekintjük át.

Az ismertetett mérőszámok vizsgálatakor említett (2.23) egyenlőtlenséget kiegészíthetjük a maximálkorrelációval:

5.Állítás.

$$0 \leq |\mathbf{R}(\xi,\eta)| \leq \min\big(\Theta_\xi(\eta), \Theta_\eta(\xi)\big) \leq \Theta(\xi,\eta) \leq \mathbf{S}(\xi,\eta) \leq \mathbf{C}(\xi,\eta), \tag{3.17}$$

ahol $\Theta(\xi,\eta) = \max\big(\Theta_\xi(\eta), \Theta_\eta(\xi)\big)$.

Bizonyítás. Az egyenlőtlenségekből valójában az utolsót nem láttuk még. Az általánosság megszorítása nélkül feltehetjük, hogy $E(f(\xi)) = E(g(\eta)) = 0$ és $D(f(\xi)) = D(g(\eta)) = 1$ teljesül, különben vehetjük a standardizáltakat. A korrelációs együttható definíciójából és a feltételekből ekkor

$$\mathbf{R}(f(\xi), g(\eta)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)g(y)[k(x,y)-1]dF(x)dG(y), \tag{3.18}$$

a Cauchy-Schwarz-Bunyakovszkij egyenlőtlenség szerint pedig ekkor

$$\mathbf{R}(f(\xi), g(\eta)) \leq \mathbf{C}(\xi,\eta), \tag{3.19}$$

és ez minden f és g esetén fennáll, tehát

$$\mathbf{S}(\xi,\eta) \leq \mathbf{C}(\xi,\eta) \quad \Box \tag{3.20}$$

A következő két állítás felhasználásával adott Yu[18] rövid bizonyítást arra, hogy a maximálkorreláció teljesíti Rényi követelményei közül az utolsót:

6.Állítás. *Ha a nem* 1 *valószínűséggel konstans* ξ *és* η *valószínűségi változók feltételesen függetlenek adott* ζ *feltétel mellett, akkor*

$$\mathbf{S}(\xi,\eta) \leq \mathbf{S}(\xi,\zeta)\mathbf{S}(\zeta,\eta), \tag{3.21}$$

és egyenlőség áll fenn, ha (ξ,ζ) *és* (η,ζ) *azonos eloszlásúak.*

7.Állítás. *Ha a nem* 1 *valószínűséggel konstans* ξ *és* η *valószínűségi változók függetlenek és azonos eloszlásúak, és* $\zeta = f(\xi,\eta)$*, ahol* $f(x,y)$ *szimmetrikus függvénye* x*-nek és* y*-nak, akkor*

$$\mathbf{S}(\zeta,\xi) \leq 2^{-\frac{1}{2}} \tag{3.22}$$

Az előző két állítás segítségével most belátjuk az utolsó Rényi-tulajdonság meglétét:

Jelölje $\mathbf{S}(\rho)$ a maximálkorrelációját egy ρ korrelációjú valószínűségiváltozó-párnak, amelyek együttes eloszlása normális! Legyenek ξ, α, β független standard normális eloszlású valószínűségi változók és $-1 < \rho_1 < 1$ valamint $-1 < \rho_2 < 1$ mellett

$$\zeta = \rho_1\xi + \sqrt{1-\rho_1^2}\alpha,\ \eta = \rho_2\zeta + \sqrt{1-\rho_2^2}\beta! \tag{3.23}$$

Ekkor ζ és η is standard normális eloszlású valószínűségi változók, és a (ξ,η,ζ) véletlen vektor kovarianciamátrixa

$$\begin{pmatrix} 1 & \rho_1\rho_2 & \rho_1 \\ \rho_1\rho_2 & 1 & \rho_2 \\ \rho_1 & \rho_2 & 1 \end{pmatrix}. \tag{3.24}$$

Ebből ξ és η feltételes kovarianciamátrixa a ζ feltétel mellett

$$\begin{pmatrix} 1 & \rho_1\rho_2 \\ \rho_1\rho_2 & 1 \end{pmatrix} - \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} \begin{pmatrix} \rho_1, \rho_2 \end{pmatrix} = \begin{pmatrix} 1-\rho_1^2 & 0 \\ 0 & 1-\rho_2^2 \end{pmatrix}, \tag{3.25}$$

tehát ξ és η feltételesen függetlenek, mivel a feltételes eloszlás is normális. Ekkor alkalmazható a 6.Állítás, így

$$\mathbf{S}(\rho_1\rho_2) = \mathbf{S}(\xi,\eta) \le \mathbf{S}(\xi,\zeta)\mathbf{S}(\eta,\zeta) = \mathbf{S}(\rho_1)\mathbf{S}(\rho_2), \tag{3.26}$$

és $\rho = \rho_1 = \rho_2$ esetben, azaz ha (ξ,η) és (ξ,ζ) azonos eloszlásúak, akkor $\mathbf{S}(\rho^2) = \mathbf{S}^2(\rho)$. Az előző egyenlőtlenségből következik $\rho_1 = \rho'$ és $\rho_2 = \frac{\rho}{\rho'}$ mellett az is, hogy ha $|\rho| \le |\rho'|$, akkor $\mathbf{S}(\rho) \le \mathbf{S}(\rho')$, azaz a maximálkorreláció monoton függvénye $|\rho|$-nak.

Legyenek μ és ν független standard normálisak, ekkor $\mathbf{R}(\mu+\nu,\mu) = 2^{-\frac{1}{2}}$, és a 7.Állítás szerint

$$\mathbf{S}(2^{-\frac{1}{2}}) = \mathbf{S}(\mu,\mu+\nu) \le 2^{-\frac{1}{2}} \tag{3.27}$$

Mivel $\mathbf{R} \le \mathbf{S}$, következik, hogy $\mathbf{S}(2^{-\frac{1}{2}}) = 2^{-\frac{1}{2}}$. Erre alkalmazva (3.26)-ot $(m-1)$ alkalommal, kapjuk a következőt:

$$2^{-\frac{m}{2}} \le \mathbf{S}(2^{-\frac{m}{2}}) \le \left\{\mathbf{S}(2^{-\frac{1}{2}})\right\}^m = 2^{-\frac{m}{2}}, \tag{3.28}$$

tehát itt végig egyenlőség teljesül. $\mathbf{S}(\rho^2) = \mathbf{S}^2(\rho)$ egyenletbe behelyettesítve $\rho = 2^{-\frac{m}{2^n}}$-t és iterálva visszavezetve $\frac{m}{2^n}$-t $\frac{m}{2}$-re:

$$\mathbf{R}(2^{-\frac{m}{2^n}}) = 2^{-\frac{m}{2^n}}. \tag{3.29}$$

Mivel minden $\rho \in (0,1)$ szám tetszőleges pontossággal közelíthető $q = \frac{m}{2^n}$ esetén 2^{-q} alakban, legyen

$$2^{-p_k} \leq \rho \leq 2^{-q_k}, \quad k = 1, 2, ..., \tag{3.30}$$

és kihasználva a maximálkorreláció monotonitását,

$$2^{-p_k} \leq \mathbf{S}(\rho) \leq 2^{-q_k}, \quad k = 1, 2, ..., \tag{3.31}$$

és $k \to \infty$ mellett $\mathbf{S}(\rho) = \rho$. Amennyiben $\rho < 0$, a szimmetria miatt $\mathbf{S}(\rho) = -\rho$ adódik. $\square$

Látható, hogy a maximálkorreláció rendelkezik mindazon tulajdonságokkal, amelyek elvben jobb mérőszámmá teszik a klasszikus korrelációnál és az említett további mérőszámoknál is. A gyakorlatban viszont fontos kérdés a mérőszám kiszámíthatósága, és ellentétben például a valamiképp hasonlóan definiált korrelációs hányados esetével, a maximálkorrelációnál nem tudunk (2.12)-höz hasonló állítást kimondani, azaz nem minden esetben létezik $f_0(x)$ és $g_0(x)$ függvény, hogy

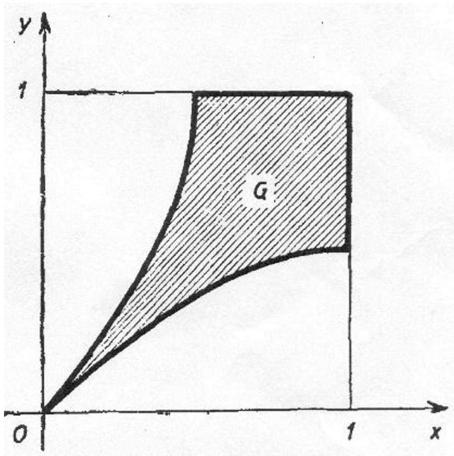
$$\mathbf{S}(\xi, \eta) = \mathbf{R}(f_0(\xi), g_0(\eta)). \tag{3.32}$$

A követező példa egy olyan esetet mutat be, amikor nem léteznek ilyen optimális függvények:

Legyen u konvex, kellően sima függvény, amelyre $u(0) = 0$, $u'(0) = 1$! Legyen $a = u''(0) > 0$! A G tartomány legyen

$$G = \{(x, y) \in [0, 1]^2 \quad : \quad v(x) \leq y \leq u(x)\}, \tag{3.33}$$

ahol $v = u^{-1}$!



Ezen kívül legyen

$$f(x) = g(x) = \begin{cases} 1, & 0 \leq x \leq \varepsilon \\ 0, & x > \varepsilon! \end{cases} \tag{3.34}$$

Ekkor belátjuk, hogy $\mathbf{R}(f(X), g(Y)) \to 1$, ha $\varepsilon \to 1$.

Legyen $\mathrm{A} = \{X \leq \varepsilon\}$ és $\mathrm{B} = \{Y \leq \varepsilon\}$, ekkor

$$\mathbf{R}(f(X), g(Y)) = \frac{P(\mathrm{A} \cap \mathrm{B}) - P(\mathrm{A})P(\mathrm{B})}{\sqrt{P(\mathrm{A})(1-P(\mathrm{A}))P(\mathrm{B})(1-P(\mathrm{B}))}} =$$

$$= \frac{P(\mathrm{A} \cap \mathrm{B}) - P(\mathrm{A})^2}{P(\mathrm{A})(1-P(\mathrm{A}))} = \frac{P(\mathrm{A}) - P(\mathrm{A} \setminus \mathrm{B}) - P(\mathrm{A})^2}{P(\mathrm{A})(1-P(\mathrm{A}))} = 1 - \frac{P(\mathrm{A} \setminus \mathrm{B})}{P(\mathrm{A})(1-P(\mathrm{A}))}. \tag{3.35}$$

Ekkor elegendő belátnunk, hogy $\frac{P(\mathrm{A} \setminus \mathrm{B})}{P(\mathrm{A})} \to 0$, mivel $\varepsilon \to 0$ esetén $P(\mathrm{A} \to 0)$. Ekkor tekintsük a következőket:

$P(\mathrm{A})$ arányos az AOE görbevonalú háromszög területével, $P(\mathrm{A} \setminus \mathrm{B})$ pedig arányos az ADB görbevonalú háromszög területével, a lenti ábra alapján. Az AD egyenes C pontban metszi az OB átlót. Ekkor a valószínűségek hányadosát felülről tudjuk becsülni valódi háromszögek területeivel:

$$\frac{P(\mathrm{A} \setminus \mathrm{B})}{P(\mathrm{A})} \leq \frac{T_{ADB}}{T_{ACB}} = \frac{BD}{EC} = \frac{\varepsilon - v(\varepsilon)}{\varepsilon - z}, \tag{3.36}$$

valamint ez a párhuzamos szelők tételéből következően pontosan $\frac{AB}{AE} = \frac{u(\varepsilon)-\varepsilon}{u(\varepsilon)-z}$. Tehát az előzőek alapján

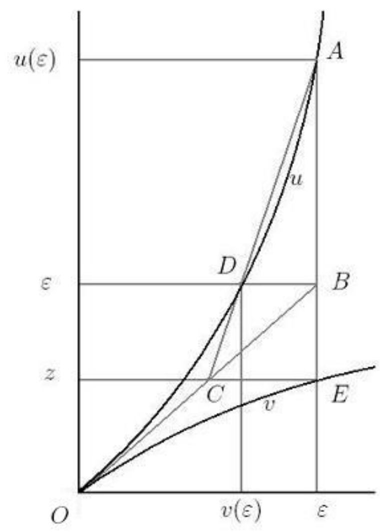
$$\frac{\varepsilon - v(\varepsilon)}{\varepsilon - z} = \frac{u(\varepsilon) - \varepsilon}{u(\varepsilon) - z}, \tag{3.37}$$

átalakítva kapjuk, hogy

$$z = \frac{u(\varepsilon)v(\varepsilon) - \varepsilon^2}{u(\varepsilon) + v(\varepsilon) - 2\varepsilon}, \tag{3.38}$$

és

$$\varepsilon - z = \frac{(u(\varepsilon) - \varepsilon)(\varepsilon - v(\varepsilon))}{u(\varepsilon) + v(\varepsilon) - 2\varepsilon}. \tag{3.39}$$



Ekkor tehát a felső becslésünk

$$\frac{P(\mathrm{A} \setminus \mathrm{B})}{P(\mathrm{A})} \leq \frac{u(\varepsilon) + v(\varepsilon) - 2\varepsilon}{\varepsilon - v(\varepsilon)}. \tag{3.40}$$

Az u függvény Taylor-sorfejtése szerint $u(\varepsilon) = \varepsilon + \frac{a}{2}\varepsilon^2 + o(\varepsilon^2)$, valamint v-re igaz, hogy $v(0) = 0$, $v'(0) = 1$,

$$v'(t) = \frac{1}{u'\big(v(t)\big)}, \tag{3.41}$$

$$v''(t) = \frac{-1}{\Big(u'\big(v(t)\big)\Big)^2} u''\big(v(t)\big)v'(t) \tag{3.42}$$

és így

$$v''(0) = \frac{-1}{1} \cdot a \cdot 1 = -a, \tag{3.43}$$

teljesül, hogy

$$v(\varepsilon) = \varepsilon - \frac{a}{2}\varepsilon^2 + o(\varepsilon^2). \tag{3.44}$$

Ezekből viszont a felső becslésre

$$\frac{u(\varepsilon) + v(\varepsilon) - 2\varepsilon}{\varepsilon - v(\varepsilon)} = \frac{o(\varepsilon^2)}{\frac{a}{2}\varepsilon^2 + o(\varepsilon^2)} \to 0, \tag{3.45}$$

ha $\varepsilon \to 0$, tehát ekkor

$$\mathbf{R}(f(X), g(Y)) = 1 - \frac{P(\mathrm{A} \setminus \mathrm{B})}{P(\mathrm{A})(1 - P(\mathrm{A}))} \to 1, \tag{3.46}$$

tehát a maximálkorreláció értéke 1.

Ekkor viszont nem létezik olyan f_0 és g_0, amelyekre (3.32) fennállna, mivel ebben az esetben $f_0(X) = g_0(Y)$ teljesülne, ebből pedig az adódna, hogy

$$\iint_G |f_0(x) - g_0(y)| dxdy = 0, \tag{3.47}$$

és így $f_0(x) \equiv g_0(x) \equiv c$, ahol c konstans, ilyen esetben viszont nemdefiniált az $\mathbf{R}(f_0(X), g_0(Y))$ mennyiség.

Elégséges feltételeket a maximálkorreláció kiszámíthatóságára a következő fejezetben tárgyalunk, Rényi[13], valamint Csáki és Fischer[4] munkái alapján, valamint bemutatunk speciális eseteket, amikor konkrét eredmény is adható.

3.2. A maximálkorreláció kiszámítása

3.2.1. Elégséges feltétel és operátorok

A szakasz címe valójában kissé félrevezető. Mint korábban említettük, a maximálkorreláció esetén nem tudunk mindig olyan optimális függvényeket mondani, mint hasonló esetben a korrelációs hányadosnál. A továbbiakban tárgyaltak olyan értelemben nem elégséges feltételei a kiszámíthatóságnak, hogy ha teljesülnek a tételekben állítottak, akkor pontosan ki tudjuk számolni a maximálkorreláció értékét, csupán az optimális függvények létére jelentenek elégségességet. A szakasz második részében lényegében az eszköztárat bővítjük ki néhány speciális eset tárgyalásához szükséges állításokkal.

Amennyiben a (3.32) egyenlőség fennáll, azaz a maximálkorreláció optimális függvényei f_0 és g_0, akkor feltéve, hogy az említett függvények várható értéke 0, szórásuk pedig 1, látható, hogy

$$E(f_0(\xi)|\eta) = \mathbf{S}(\xi,\eta)g_0(\eta) \tag{3.48}$$

és

$$E(g_0(\eta)|\xi) = \mathbf{S}(\xi,\eta)f_0(\xi), \tag{3.49}$$

Ekkor az optimáis függvényeket lényegében az egyenletrendszer megoldásaiként keressük, feltéve, hogy léteznek. Az egyenleteket más alakban felírva:

$$\begin{cases} E(E(f_0(\xi)|\eta)|\xi) = \mathbf{S}^2(\xi,\eta)f_0(\xi) \\ E(E(g_0(\eta)|\xi)|\eta) = \mathbf{S}^2(\xi,\eta)g_0(\eta). \end{cases} \tag{3.50}$$

Nyilván elegendő például a (3.50) egyenlőségből meghatározni az f_0 függvényt, a g_0 meghatározható az eredeti egyenletrendszerből f_0 ismeretében.

Láthatjuk, hogy a korábban definiált operátorokkal felírva az eredeti rendszer a következő:

$$\begin{cases} \mathbf{A}_\eta f_0 = \mathbf{S}(\xi,\eta)g_0 \\ \mathbf{A}_\xi g_0 = \mathbf{S}(\xi,\eta)f_0 \end{cases} \tag{3.51}$$

Ekkor legyen tetszőleges $f = f(\xi) \in L^2_\xi$ esetén

$$\mathbf{A}f = \mathbf{A}_\xi \mathbf{A}_\eta f = E(E(f(\xi)|\eta)|\xi), \tag{3.52}$$

így a (3.50) egyenlőséget átírhatjuk a

$$\mathbf{A}f_0 = \mathbf{S}^2(\xi,\eta)f_0 \tag{3.53}$$

alakba. Vizsgáljuk innentől az $\mathbf{A}$ operátort! Tudjuk, hogy ez egy korlátos lineáris transzformáció az $L^2_{\xi,0}$ térben, valamint önadjungált és pozitív szemidefinit.

Legyenek f és g egységnyi szórásúak, rendre az $L^2_{\xi,0}$ és $L^2_{\eta,0}$ terekből! Ekkor a teljes várható érték tétele és a Cauchy-Schwarz-Bunyakovszkij–egyenlőtlenség szerint

$$E^2\big(f(\xi)g(\eta)\big) = E^2\big(E(f(\xi)|\eta)g(\eta)\big) \leq E\big(E^2(f(\xi)|\eta)\big) = \langle \mathbf{A}f, f\rangle. \quad (3.54)$$

Legyen

$$\lambda = \|\mathbf{A}\| = \sup_{f\in L^2_{\xi,0}, \|f\|=1} \langle \mathbf{A}f, f\rangle, \quad (3.55)$$

és így

$$\mathbf{S}^2(\xi,\eta) \leq \lambda. \quad (3.56)$$

Másrészt ha $f \in L^2_{\xi,0}$ és $\|f\| = 1$, $g(\eta) = E(f(\xi)|\eta)$ helyettesítéssel kapjuk, hogy

$$D^2\big(g(\eta)\big) = E\big(f(\xi)g(\eta)\big) \leq \mathbf{S}(\xi,\eta)D\big(g(\eta)\big), \quad (3.57)$$

tehát $D\big(g(\eta)\big) \leq \mathbf{S}(\xi,\eta)$ teljesül, innen pedig

$$\langle \mathbf{A}f, f\rangle = E\big(g(\eta)E(f(\xi)|\eta)\big) = E\big(f(\xi)g(\eta)\big) \leq \mathbf{S}(\xi,\eta)D\big(g(\eta)\big) \leq \mathbf{S}^2(\xi,\eta), \quad (3.58)$$

és ebből következik, hogy $\lambda \leq \mathbf{S}^2(\xi,\eta)$, ehhez hozzávéve a (3.56) egyenlőtlenséget, adódik, hogy

$$\mathbf{S}^2(\xi,\eta) = \lambda = \sup_{f\in L^2_{\xi,0}, \|f\|=1} \langle \mathbf{A}f, f\rangle. \quad (3.59)$$

Ismert, hogy ha az $\mathbf{A}$ korlátos önadjungált operátor teljesen folytonos, azaz kompakt, akkor $\lambda = \sup_{\|f\|=1}\langle \mathbf{A}f, f\rangle$ az operátor legnagyobb sajátértéke, és létezik a sajátértékhez tartozó sajátfüggvény. Így a maximálkorrelációra beláttuk a következő tételt:

1.Tétel. *Ha a fenti módon definiált* $\mathbf{A}$ *operátor kompakt, akkor a* ξ *és* η *maximálkorrelációjára fennáll a (3.32) egyenlőség, ahol is* f_0 *az* $\mathbf{A}$ *legnagyobb sajátértékéhez,* $\mathbf{S}^2(\xi,\eta)$*-hoz tartozó sajátfüggvény és*

$$g_0 = \frac{1}{\mathbf{S}(\xi,\eta)} E\big(f_0(\xi)|\eta\big) \quad (3.60)$$

Az operátorra való feltétel teljesülésére Rényi[13] bizonyított egy hasznos tételt:

2.Tétel. *Ha a ξ és η közötti összefüggés reguláris és a kontingenciájuk véges, az* $\mathbf{A}$ *transzformáció egyenletesen folytonos, és így a maximálkorrelációhoz léteznek az* 1.Tétel*ben tárgyalt optimális transzformáló függvények.*

A 2.Tételben szereplő feltétel természetesen nem szükséges, csak elégséges, például ha $\xi = \eta$ teljesül, akkor a maximálkorreláció értéke nyilván 1, optimális függvényekre jó lesz $f_0(x) \equiv g_0(x) \equiv x$, ekkor viszont az összefüggés nem reguláris, így a kontingenciát nem is értelmezzük.

Beláttuk tehát, hogy ha tekintjük az alábbi, már vázolt egyenletrendszert, nevezetesen

$$\begin{cases} \mathbf{A}_\eta f = \mu g \\ \mathbf{A}_\xi g = \mu f \end{cases} \quad f \in L^2_{\xi,0}; g \in L^2_{\eta,0}, \tag{3.61}$$

tehát μ sajátérték és f, g egy sajátfüggvénypár az $\mathbf{A}_\eta, \mathbf{A}_\xi$ operátorpárhoz, akkor a maximálkorreláció definíciójában a supremum akkor és csak akkor maximum, ha $\mathbf{S}(\xi, \eta) = |\mu_0|$, ahol μ_0 a maximális abszolútértékű sajátérték.

Sok esetben a kiszámíthatóságra, valamint a merőlegesvetítés-operátorokra vonatkozó vizsgálatokat megkönnyíti a kétváltozós eloszlás változókban vett szimmetriája, ugyanis szimmetria esetén $\mathbf{Q}(A \times B) = \mathbf{Q}(B \times A)$, így tehát $\mathbf{Q}_1 = \mathbf{Q}_2$, ezért az alterekre $L^2_{F_1} = L^2_{F_2}$, tehát $\mathbf{A}_1 = \mathbf{A}_2$, tehát az operátoregyenletekre

$$\begin{cases} \mathbf{A}_1 f = \mu g \\ \mathbf{A}_1 g = \mu f \end{cases} \tag{3.62}$$

adódik. Ekkor ha $f(x) = g(x)$, akkor $f(x)$ sajátfüggvénye az operátornak, ha nem, akkor a két egyenletet kivonva egymásból $(f(x) - g(x))$ lesz a sajátfüggvény, tehát ilyen esetekben elegendő egy operátor sajátértékeit vizsgálnunk, ahelyett, hogy két operátor spektrálfelbontását végeznénk el.

Belátható, hogy minden kétváltozós eloszlás helyett bevezethető egy használható, változóiban szimmetrikus eloszlás, ugyanis ha az eredeti kétváltozós eloszlás $\mathbf{Q}$, akkor legyen a változóiban szimmetrikus eloszlás

$$\overline{\mathbf{Q}}(A \times B) = \langle \mathbf{A}_2\mathbf{A}_1\chi_A(x), \chi_B(x)\rangle_{L^2_{F_1}} = \langle \mathbf{A}_1\chi_A(x), \mathbf{A}_1\chi_B(x)\rangle_{L^2_{F_2}}, \quad (3.63)$$

ahol A és B a valós számegyenes tetszőleges mérhető halmazai. Ez valószínűségi mérték lesz a síkon, $\overline{\mathbf{Q}}(B\times A) = \overline{\mathbf{Q}}(A\times B)$, és a hozzá tartozó eloszlásfüggvényre $\overline{F}(x,y) = \overline{F}(y,x)$, valamint a peremeloszlásokra $\overline{\mathbf{Q}}_1 = \overline{\mathbf{Q}}_2 = \mathbf{Q}_1$, és így $L^2_{\overline{F_1}} = L^2_{\overline{F_2}} = L^2_{F_1}$.

A (3.63) egyenlőség utolsó tagjára

$$\langle \mathbf{A}_1\chi_A(x), \mathbf{A}_1\chi_B(x)\rangle_{L^2_{F_2}} = \int_{-\infty}^{\infty} \mathbf{Q}_1(A|y)\mathbf{Q}_1(B|y)dF_2(y), \quad (3.64)$$

így ésszerűnek látszik a $\overline{\mathbf{Q}}$ eloszlás azon megközelítése, hogy rögzítjük η egy tetszőleges értékét és függetlenül választunk ξ értékei közül kettőt, ekkor $\overline{\mathbf{Q}}$ az ilyen párok eloszlásainak keveréke, η feltétellel.

Ilyen transzformációkat javasol Gebelein[8] és Sarmanov[16] is, speciális esetekben. Csáki és Fischer[4] is belátták, hogy az $\mathbf{A}_1$ és $\mathbf{A}_2$ operátorok helyett érdemes az $\mathbf{A}_2\mathbf{A}_1$ operátor viselkedését vizsgálni, ha a $\overline{\mathbf{Q}}$ eloszlást vizsgáljuk, mivel igaz a következő összefüggés:

3.Tétel.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)g(y)d\overline{F}(x,y) = \langle \mathbf{A}_2\mathbf{A}_1 f(x), g(x)\rangle, \quad (3.65)$$

ahol $f(x) \in L^2_{F_1}$ *és* $g(x) \in L^2_{F_2}$.

Megjegyzés. A 3.Tételből következik, hogy

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)f(y)d\overline{F}(x,y) = \Theta^2_\eta\big(f(\xi)\big), \quad (3.66)$$

ahol $f(x) \in L^2_{F_1,0}$ és $\|f(x)\| = 1$.

A 3.Tétel követezményeképp látható, hogy az $\overline{\mathbf{A}} = \mathbf{A}_2\mathbf{A}_1$ operátor generálja a feltételes várható értéket a következő értelemben:

$$\overline{\mathbf{A}}f(x) = \int_{-\infty}^{\infty} f(x)d\overline{F}_1(x|y). \quad (3.67)$$

További következmény az alábbi állítás:

8.Állítás. *A maximálkorreláció $\overline{F}(x,y)$ eloszlásfüggvény mellett a négyzete az $F(x,y)$ mellettinek.*

A korábban említett állítás, mely szerint $\overline{\mathbf{Q}}$ mindkét peremeloszlása megegyezik az egyik eredeti peremeloszlással, eléggé használhatónak és hasznosnak tűnik olyan esetekben, ha az egyik marginális diszkrét vagy differenciálható.

Vizsgáljuk továbbra is a sajátfüggvényekre vonatkozó egyenletrendszerünket! Sok szempontból előnyös eset a lineáris összefüggés fennállása, így felmerül a feltételes várható értékek linearizálásának problémája. Erre a következőket mondhatjuk:

4.Tétel. *Az $f \in L^2_{\xi,0}$ és $g \in L^2_{\eta,0}$ standard valószínűségi változókra a következő állítások ekvivalensek:*
1. f és g sajátfüggvénypárt alkotnak,
2. $\Theta_\xi(g) = \Theta_\eta(f) = \langle f, g\rangle$,
3. f és g lineárisan korreláltak, és $\Theta_f(g) = \Theta_\xi(g)$, $\Theta_g(f) = \Theta_\eta(f)$

Ennek következményeként kimondhatjuk a következőt is:

5.Tétel. *Ha az $f = f(\xi)$ és $g = g(\eta)$ standard valószínűségi változók lineárisan korreláltak, és $f(x)$ és $g(y)$ injektív függvények, akkor f és g sajátfüggvénypárt alkot.*

Bizonyítás. Az injektivitás miatt $L^2_f = L^2_\xi$ és $L^2_g = L^2_\eta$, tehát az előző tétel 3. pontja teljesül $\square$

A maximálkorrelációhoz tartozó f és g standard sajátfüggvények távolságáról a következőket mondhatjuk:

Mivel $f \in L_{\xi,0^2}$ és $g \in L^2_{\eta,0}$, így

$$\|f - g\|^2 = 2\big(1 - \langle f, g\rangle\big), \tag{3.68}$$

ekkor pedig

$$\inf_{\substack{f \in L^2_{\xi,0},\, \|f\|=1 \\ g \in L^2_{\eta,0},\, \|g\|=1}} \|f-g\|^2 = 2\bigl(1 - \sup_{\substack{f \in L^2_{\xi,0},\, \|f\|=1 \\ g \in L^2_{\eta,0},\, \|g\|=1}} \langle f, g\rangle\bigr) = 2\bigl(1 - \mathbf{S}(\xi,\eta)\bigr). \quad (3.69)$$

Tekintsünk egy kissé módosított infimum-problémát, egészen pontosan a következőt:

$$\inf_{\substack{f \in L^2_{\xi,0},\, \|f\|=1 \\ g \in L^2_{\eta,0},\, \|g\|=1 \\ -\infty<\lambda<\infty}} \|f-\lambda g\|^2 = \inf_{\substack{f \in L^2_{\xi,0},\, \|f\|=1 \\ g \in L^2_{\eta,0},\, \|g\|=1 \\ -\infty<\lambda<\infty}} \bigl(1 - 2\lambda\langle f, g\rangle + \lambda^2\bigr)^2 = 1 - \mathbf{S}^2(\xi,\eta). \quad (3.70)$$

Nyilvánvaló, hogy a második esetben az infimum kisebb, mint az elsőben, kivéve $\mathbf{S}(\xi,\eta) = 1$ esetet, amikor egyenlőek. Az első összefüggésből követezik az is, hogy a maximálkorreláció pontosan akkor 1, ha az $L^2_{\xi,0}$ és az $L^2_{\eta,0}$ terek egységgömbjeinek távolsága zérus.

Ha ilyen esetben az egységgömbök diszjunktak, akkor a maximálkorrelációra nem léteznek optimális transzformáló függvények, azaz supremum nem lesz maximum. A korábban tekintett példa a függvények nemlétére pontosan ilyet ír le.

A (3.70)-ben tekintett infimum értéke megegyezik a sajátfüggvénypár feltételes szórásnégyzetével, mivel

$$\inf_{\substack{f \in L^2_{\xi,0} \\ \|f\|=1}} \|f - \mathbf{A}_\eta f\|^2 = \inf_{\substack{f \in L^2_{\xi,0} \\ \|f\|=1}} (1 - \|\mathbf{A}_\eta f\|^2) =$$

$$= \inf_{\substack{f \in L^2_{\xi,0} \\ \|f\|=1}} \bigl(1 - \Theta^2_\eta(f)\bigr) = 1 - \mathbf{S}^2(\xi,\eta). \quad (3.71)$$

Néhány szempontból előnyös feltenni a homoszkedaszticitást, azaz jelen esetben a feltételes szórásnégyzet állandóságát, egészen pontosan a ζ standard valószínűségi változóra a következő összefüggést:

$$\mathbf{A}_\xi \zeta^2 - (\mathbf{A}_\xi \zeta)^2 \equiv 1 - \Theta^2_\xi(\zeta). \quad (3.72)$$

Erre a következőt állíthatjuk:

6.Tétel. *A ξ és η lineárisan korrelált standard valószínűségi változókra a következőek egyenértékűek:*
1. ξ és η homoszkedasztikusan összefüggnek
2. $\xi^2 - 1$ és $\eta^2 - 1$ sajátfüggvénypárt alkotnak a $\langle \xi, \eta \rangle^2$ sajátértékkel.

Ennek következményeként pedig a következőt állíthatjuk:

7.Tétel. *Ha a $\xi = \xi^*$ és $\eta = \eta^*$ valószínűségi változók közötti kapcsolat lineáris és homoszkedasztikus, továbbá $0 < |\langle \xi, \eta \rangle| < 1$ teljesül, akkor ξ és η harmadik momentuma is zérus.*

Bizonyítás. Az előző tétel szerint

$$\langle \xi, \xi^2 - 1 \rangle = \langle \eta, \eta^2 - 1 \rangle = 0, \tag{3.73}$$

mivel különböző sajátértékekhez tartozó sajátfüggvények ortogonálisak. Ekkor pedig

$$E(\xi^3) = E(\xi) = E(\eta^3) = E(\eta) = 0. \qquad \square \tag{3.74}$$

Érdemes megjegyezni, hogy kétváltozós normális eloszlás esetén mind a 6.Tétel, mind a 7.Tétel feltételei teljesülnek.

Az eddig felépített eszköztár elég sok eszközt ad a kezünkbe, amelyekkel a maximálkorreláció pontos értékét ki tudjuk számítani, esetleg meg tudunk találni optimális transzformáló függvényeket. Maga a pontos kiszámítás, mint láttuk, nem egyszerű feladat, de bizonyos speciális esetekben akár nagyon egyszerű is lehet, erre ad egy következő jó lehetőséget az alábbi tétel:

8.Tétel. *Legyenek $\varphi_1, \varphi_2, \ldots$ és $\psi_1, \psi_2, \ldots$ rendre lineárisan független rendszerek az $L^2_{\xi,0}$ és $L^2_{\eta,0}$ terekben! Ekkor az*

$$f_n = \sum_{k=1}^{n} a_{kn} \varphi_k; \quad g_n = \sum_{k=1}^{n} a'_{kn} \psi_k; \qquad n = 1, 2, \ldots \tag{3.75}$$

függvények a sajátfüggvényei az $\mathbf{A}_\xi$, $\mathbf{A}_\eta$ *operátorpárnak akkor és csak akkor, ha léteznek olyan* b_{kn} *és* b'_{kn} *együtthatók, hogy*

$$\mathbf{A}_\eta \varphi_n = \sum_{k=1}^{n} b_{kn}\psi_k; \quad \mathbf{A}_\xi \psi_n = \sum_{k=1}^{n} b'_{kn}\varphi_k; \quad n = 1, 2, \ldots, \tag{3.76}$$

és ebben az esetben a pontos sajátértékek:

$$\lambda_n = \sqrt{b_{nn} b'_{nn}}; \quad n = 1, 2, \ldots, \tag{3.77}$$

továbbá az a_{kn} *és az* a'_{kn} *együtthatókra teljesülnek a*

$$\sum_{k=i}^{n} b_{ik} a_{kn} = \lambda_n a'_{in}; \quad \sum_{k=1}^{n} b'_{ik} a_{kn} = \lambda_n a_{in}; \quad i = 1, \ldots, n; \quad n = 1, 2, \ldots \tag{3.78}$$

egyenlőségek.

Bizonyítás. Lásd [5].

Amennyiben $\{\lambda_n\}$ a nem nulla sajátértékeket jelenti és a maximálkorreláció előáll maximumként, akkor

$$\mathbf{S}(\xi, \eta) = \max_n \sqrt{b_{nn} b'_{nn}} \tag{3.79}$$

fennáll. Ha az együttes eloszlás változóiban szimmetrikus, (3.76) és (3.78) egyenlőségei az $a'_{kn} = a_{kn}$, $b'_{kn} = b_{kn}$ alakra redukálódnak, ekkor értelemszerűen $\lambda_n = b_{nn}$ teljesül.

A fentiek következményeként állíthatjuk a következőt is:

9.Állítás. *A* 8.Tétel *alkalmazható, ha a lineárisan független függvényeket a következőképp választjuk:*

$$\varphi_n = \xi^n - E(\xi^n) \in L^2_{\xi,0}; \quad \psi_n = \eta^n - E(\eta^n) \in L^2_{\eta,0}; \quad n = 1, 2, \ldots \tag{3.80}$$

Ilyen speciális esetekben a 8.Tétel *szerint a sajátfüggvények polinomok akkor és csak akkor, ha minden* n *esetén az* n*-edik feltételes momentum legfeljebb* n*-edfokú polinomja a feltételben szereplő változónak.*

A 8.Tétel következményeként a sajátértékek minden további számítás nélkül megtalálhatóak, a sajátfüggvények együtthatói pedig lineáris egyenletrendszerből meghatározhatóak, feltéve, hogy a vizsgált terekben a lineárisan független rendszerek minden egyes függvénye olyan, hogy a tételbeli alakban felírható a feltételes várható értéke.

3.2.2. Speciális esetek

A következő szakaszban áttekintünk néhány pontosan kiszámítható speciális esetet az előzőekben tárgyaltak felhasználásával, bizonyos esetekben újabb szükséges tételek említésével.

Példa $\mathbf{S}(\xi,\eta) = \mathbf{R}(\xi,\eta)$ esetre

A következő példában a maximálkorreláció és a korreláció értéke megegyezik. Ezen az egyszerű példán – azon túl, hogy ilyen esetekben pontosan meg tudjuk határozni a maximálkorreláció értékét – jól bemutatható a 8.Tétel alkalmazhatósága.

Legyen az együttes sűrségfüggvény a következő:

$$f(x,y) = \begin{cases} \log \frac{1}{(1-x)y}, & 0 \le x \le y \le 1 \\ \log \frac{1}{x(1-y)}, & 0 \le y \le x \le 1. \end{cases} \tag{3.81}$$

Ekkor az együttes eloszlás változóiban szimmetrikus, elegendő a probléma egyoldali tárgyalása. A marginális és feltételes sűrűségfüggvényekre most

$$f_1(x) = 1, \quad 0 \le x \le 1 \tag{3.82}$$

és

$$f_1(x|y) = f(x,y), \quad 0 \le x \le 1, \quad 0 \le y \le 1. \tag{3.83}$$

Ekkor a 8.Tétel felhasználásával

$$\mathbf{A}_\eta \xi^n = \frac{1}{n+1}\left(\frac{1}{n+1} + \sum_{i=1}^{n} \frac{\eta^i}{i}\right); \quad n = 1,2,\ldots \tag{3.84}$$

és

$$\lambda_n = b_{nn} = \frac{1}{n(n+1)}; \qquad n = 1, 2, \ldots \tag{3.85}$$

adódik. A $\xi^n - E(\xi^n)$ függvények rendszere $L^2_{\xi,0}$-ban teljes lesz, így $\{\lambda_n\}$ az összes sajátérték halmaza, ebből pedig

$$\mathbf{S}(\xi, \eta) = \lambda_1 = \frac{1}{2}. \tag{3.86}$$

Trinomiális eloszlás

Legyen $\xi = i$ és $\eta = k$ együttes valószínűsége p_{ik}, a marginálisok pedig értelemszerűen $p_{i\cdot}$ és $p_{\cdot k}$! Ekkor a trinomiális eloszlás a következő:

$$p_{ik} = \frac{N!}{i!k!(N-i-k)!} p_1^i p_2^k (1 - p_1 - p_2)^{N-i-k}, \tag{3.87}$$

ahol $p_1 > 0$, $p_2 > 0$, $p_1 + p_2 < 1$; $i = 0, 1, \ldots, N$; $k = 0, 1, \ldots, N$; $i + k \leq N$. Ekkor a marginálisok:

$$p_{i\cdot} = \binom{N}{i} p_1^i (1 - p_1)^{N-i}; \quad i = 0, 1, \ldots, N \tag{3.88}$$

és

$$p_{\cdot k} = \binom{N}{k} p_2^k (1 - p_2)^{N-k}; \quad k = 0, 1, \ldots, N, \tag{3.89}$$

a feltételes eloszlások pedig

$$\frac{p_{ik}}{p_{\cdot k}} = \binom{N-k}{i} \Big(\frac{p_1}{1-p_2}\Big)^i \Big(\frac{1-p_1-p_2}{1-p_2}\Big)^{N-k-i}; \quad i = 0, 1, \ldots, N-k \tag{3.90}$$

és

$$\frac{p_{ik}}{p_{i\cdot}} = \binom{N-i}{k} \Big(\frac{p_2}{1-p_1}\Big)^k \Big(\frac{1-p_1-p_2}{1-p_1}\Big)^{N-i-k}; \quad k = 0, 1, \ldots, N-i. \tag{3.91}$$

Ekkor alkalmazzuk a 8.Tételt:

$$\mathbf{A}_\eta \xi^n = \sum_{j=1}^{n} a_{jn} \Big(\frac{p_1}{1-p_2}\Big)^j (N-\eta)(N-\eta-1)\ldots(N-\eta-j+1); \quad n = 1, \ldots, N, \tag{3.92}$$

$$\mathbf{A}_\xi \eta^n = \sum_{j=1}^{n} a_{jn} \Big(\frac{p_2}{1-p_1}\Big)^j (N-\xi)(N-\xi-1)\ldots(N-\xi-j+1); \quad n = 1, \ldots, N, \tag{3.93}$$

ahol $a_{nn} = 1$, így a sajátértékek

$$\lambda_n = \Big(\frac{p_1 p_2}{(1-p_1)(1-p_2)}\Big)^{\frac{n}{2}}; \quad n = 1, \dots, N, \tag{3.94}$$

és a maximálkorreláció

$$\mathbf{S}(\xi, \eta) = \sqrt{\frac{p_1 p_2}{(1-p_1)(1-p_2)}}. \tag{3.95}$$

Trihipergeometrikus eloszlás

Az előző esethez hasonló jelölésekkel a trihipergeometrikus eloszlás:

$$p_{ik} = \frac{\binom{N_1}{i}\binom{N_2}{k}\binom{N-N_1-N_2}{n-i-k}}{\binom{N}{n}}, \tag{3.96}$$

ahol n, N_1, N_2 és N pozitív egészek,

$$N_1 + N_2 < N; \quad i = 0, 1, \dots, n, \quad k = 0, 1, \dots, n; \quad i + k \le n.$$

A trinomiális esethez teljesen hasonlóan eljárva a következő sajátértékeket kapjuk:

$$\lambda_m = \left[\frac{\binom{N-N_1-m}{N_2-m}\binom{N-N_2-m}{N_1-m}}{\binom{N-N_1}{N_2}\binom{N-N_2}{N_1}}\right]^{\frac{1}{2}}, \tag{3.97}$$

és a maximumra most

$$\mathbf{S}(\xi, \eta) = \lambda_1 = \sqrt{\frac{N_1 N_2}{(N-N_1)(N-N_2)}}. \tag{3.98}$$

Észrevehető, hogy $p_1 = \frac{N_1}{N}$, $p_2 = \frac{N_2}{N}$ esetben a trinomiális eset ugyanezt adja eredményül.

Példa $\mathbf{S}(\xi, \eta) = \mathbf{C}(\xi, \eta)$ esetre

Legyen most a szimmetrikus sűrűségfüggvény

$$f(x, y) = p_1 a(x)a(y) + p_2\big(a(x)b(y) + b(x)a(y)\big) + p_3 b(x)b(y), \tag{3.99}$$

ahol $a(x)$ és $b(x)$ lineárisan függetlenek. Az egyszerűség kedvéért feltehetjük, hogy $a(x)$ és $b(x)$ is sűrűségfüggvény, $p_1 \geq 0$, $p_2 \geq 0$, $p_3 \geq 0$, $p_1+2p_2+p_3 = 1$. Ekkor a peremeloszlásokra

$$f_1(x) = pa(x) + qb(x); \quad p = p_1 + p_2; q = p_2 + p_3. \tag{3.100}$$

Az $f(x, y)$ sűrűségfüggvény ekkor olyan, hogy legfeljebb egy nem nulla sajátérték lehetséges, tehát

$$\mathbf{S}(\xi, \eta) = \mathbf{C}(\xi, \eta). \tag{3.101}$$

A korábban definiáltak miatt $\mathbf{C}^2(\xi, \eta)$ a következő kifejezés integrálásából fejezhető ki:

$$\frac{f(x, y) - f_1(x)f_1(y)}{\sqrt{f_1(x)f_1(y)}} = (p_1p_3-p_2^2)\frac{a(x) - b(x)}{\sqrt{pa(x) + qb(x)}}\cdot\frac{a(y) - b(y)}{\sqrt{pa(y) + qb(y)}}. \tag{3.102}$$

Innen

$$\mathbf{C}(\xi, \eta) = |p_1p_3 - p_2^2| \int_{-\infty}^{\infty} \frac{\big(a(x) - b(x)\big)^2}{pa(x) + qb(x)} dx =$$

$$= \frac{|p_1p_3 - p_2^2|}{pq}\left(1 - \int_{-\infty}^{\infty} \frac{a(x)b(x)}{pa(x) + qb(x)} dx\right). \tag{3.103}$$

Ekkor tehát

$$\mathbf{S}(\xi, \eta) = \frac{|p_1p_3 - p_2^2|}{pq}\left(1 - \int_{-\infty}^{\infty} F\Big(\frac{a(x)}{2q}, \frac{b(x)}{2p}\Big) dx\right), \tag{3.104}$$

ahol $F(x, y) = \frac{2}{\frac{1}{x}+\frac{1}{y}}$, a kapcsolódó sajátfüggvény pedig $\frac{a(x)b(x)}{pa(x)+qb(x)}$. Speciálisan $p_2 = 0$ esetben

$$\mathbf{S}(\xi, \eta) = 1 - \int_{-\infty}^{\infty} F\Big(\frac{a(x)}{2p_3}, \frac{b(x)}{2p_1}\Big) dx, \tag{3.105}$$

és $p_1 = p_3 = 0$, $p_2 = \frac{1}{2}$ esetben

$$\mathbf{S}(\xi, \eta) = 1 - \int_{-\infty}^{\infty} F\big(a(x), b(x)\big) dx. \tag{3.106}$$

A következőkben áttekintünk három példát a tárgyalt esetre:

a) Legyen a szimmetrikus sűrűségfüggvény most

$$f(x,y)=\frac{1}{2\pi}\left((\sqrt{2}e^{-\frac{x^2}{2}}-e^{x^2})e^{-y^2}+(\sqrt{2}e^{-\frac{y^2}{2}}-e^{y^2})e^{-x^2}\right)! \tag{3.107}$$

(lásd [14])

Ekkor a peremeloszlások standard normálisak, korrelálatlanok, de nem függetlenek. A fentebb tárgyaltakat nézve $p_1=p_3=0$, $p_2=\frac{1}{2}$ és

$$a(x)=\frac{1}{\sqrt{\pi}}(\sqrt{2}e^{-\frac{x^2}{2}}-e^{x^2}),\quad b(x)=\frac{1}{\sqrt{\pi}}e^{-x^2}, \tag{3.108}$$

és innen

$$\mathbf{S}(\xi,\eta)=1-\sqrt{\frac{2}{\pi}}\int_{-\infty}^{\infty}(\sqrt{2}e^{-x}-e^{-\frac{3x^2}{2}})dx=\frac{2}{\sqrt{3}}-1\approx 0,1547. \tag{3.109}$$

b) A következő speciális esetben legyen

$$f(x,y)=4p_1xy+2p_2(x+y)+p_3;\quad 0\le x\le 1;\;\; 0\le y\le 1, \tag{3.110}$$

ahol $a(x)=2x$ és $b(x)=1$ minden $[0,1]$-beli x-re. Ekkor szintén a korábbiak szerint a maximálkorreláció

$$\mathbf{S}(\xi,\eta)=\frac{|p_1p_3-p_2^2|}{p^3}(\text{arth}p-p). \tag{3.111}$$

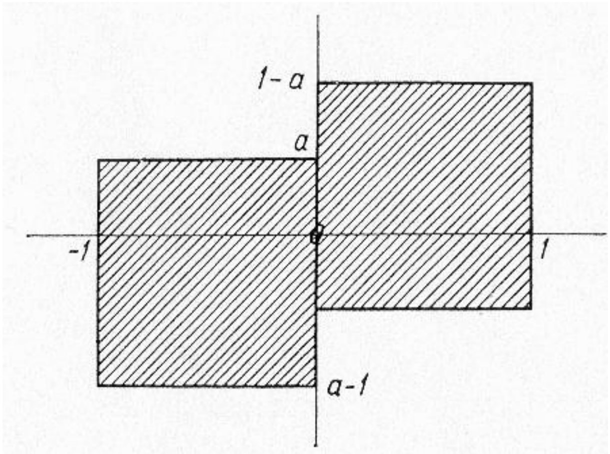
Speciálisan, ha $f(x,y)=x+y$ vagy $f(x,y)=2xy+\frac{1}{2}$, akkor

$$\mathbf{S}(\xi,\eta)=\log 3-1\approx 0,0986. \tag{3.112}$$

c) A következő példa jól mutatja a kapcsolatot az összefüggés erőssége és a maximálkorreláció értéke között. Legyen a T tartomány két egységnégyzet uniója, pontosabban

$$T=\{-1\le x\le 0;\;\; a-1\le y\le a\}\cup\{0\le x\le 1;\;\; -a\le y\le 1-a\}, \tag{3.113}$$

ahol $0\le a\le\frac{1}{2}$, mint a következő ábrán látható.



Legyen ξ és η együttes eloszlása egyenletes ezen a T-n! Ezt átalakíthatjuk változóiban szimmetrikus eloszlássá, ahogyan azt korábban tárgyaltuk, ekkor az együttes sűrűségfüggvény

$$\overline{f}(x,y) = \int_{-\infty}^{\infty} \frac{f(x,t)f(y,t)}{f_2(t)} dt. \tag{3.114}$$

Az új együttes eloszlásfüggvényhez tartozó sajátértékek a négyzetei az eredetiknek(lásd [4]), jelen esetben

$$p_1 = p_3 = \frac{1-a}{2}, \quad p_2 = \frac{a}{2}; \quad a(x) = 1 \quad -1 \le x \le 0, \quad b(x) = 1 \quad 0 \le x \le 1. \tag{3.115}$$

Ekkor a (3.104) egyenlőség szerint

$$\mathbf{S}(\xi,\eta) = \sqrt{\frac{|p_1p_3 - p_2^2|}{pq}} = \sqrt{1-2a}. \tag{3.116}$$

Többszörös sajátértékek

Tekintsük a következő együttes sűrűségfüggvényt:

$$f(x,y) = k(x-y) \qquad 0 \le x \le 2\pi; \quad 0 \le y \le 2\pi, \tag{3.117}$$

ahol a $k(x)$ függvény nemnegatív, 2π szerint periodikus, integrálható a $(0, 2\pi)$ intervallumon, $\int_0^{2\pi} k(x)dx = \frac{1}{2\pi}$ és $k(-x) = k(x)$! Ekkor a marginális és feltételes sűrűségfüggvényekre

$$f_1(x) = \frac{1}{2\pi}, \qquad 0 \le x \le 2\pi \tag{3.118}$$

és

$$f_1(x|y) = 2\pi k(x-y), \qquad 0 \le x \le 2\pi, 0 \le y \le 2\pi. \tag{3.119}$$

A sajátértékek ekkor

$$\lambda_n = 2\pi \int_0^{2\pi} k(x) \cos nx dx; \qquad n = 1, 2, \ldots, \tag{3.120}$$

és a megfelelő sajátfüggvények

$$f_{n,1} = \cos n\xi; \;\; f_{n,2} = \sin n\xi; \qquad n = 1, 2, \ldots, \tag{3.121}$$

így a λ_n-ek kétszeres sajátértékek,

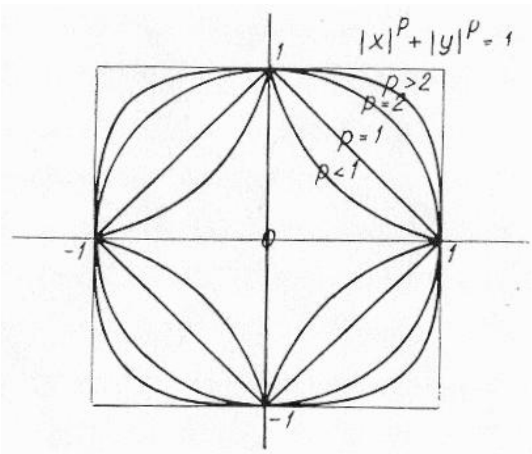
$$\mathbf{S}(\xi, \eta) = \max_n |\lambda_n|. \tag{3.122}$$

Bártfai publikálatlan példájának általánosítása

Legyen a síkon definiált T tartomány a következő:

$$T = \{(x, y) : |x|^p + |y|^q \le 1\} \quad p > 0, q > 0, \tag{3.123}$$

és legyen (ξ, η) egyenletes eloszlású T-n! A példa alapjai Bártfaitól erednek, nevezetesen a $p = q = 2$ esetre való megoldás, amit viszont soha nem publikált. Az általános esetből az is ki fog derülni, hogy $p = q$ esetben "keskenyebb" T esetén a maximálkorreláció értéke nagyobb.



Az együttes eloszlásfüggvény tehát $f(x,y) = \frac{1}{t}$, ha $(x,y) \in T$, ahol a következők igazak:

$$t = \iint_T dxdy = 4\int_0^1 \int_0^{(1-y^q)^{\frac{1}{p}}} dxdy = 4\int_0^1 (1-y^q)^{\frac{1}{p}} dy \tag{3.124}$$

A peremeloszlások sűrűségfüggvényei ekkor

$$f_1(x) = \frac{2}{t}(1-|x|^p)^{\frac{1}{q}}, \quad |x| \le 1, \tag{3.125}$$

$$f_2(y) = \frac{2}{t}(1-|y|^q)^{\frac{1}{p}}, \quad |y| \le 1, \tag{3.126}$$

és a feltételes sűrűségfüggvények

$$f_1(x|y) = \frac{1}{2(1-|y|^q)^{\frac{1}{p}}}, \quad |x| \le (1-|y|^q)^{\frac{1}{p}}, |y| \le 1, \tag{3.127}$$

$$f_2(y|x) = \frac{1}{2(1-|x|^p)^{\frac{1}{q}}}, \quad |y| \le (1-|x|^p)^{\frac{1}{q}}, |x| \le 1. \tag{3.128}$$

A 8.Tétel alkalmazásához szükséges lineárisan független függvényeket válasszuk a következőképp:

$$\varphi_n = |\xi|^{pn} - E(|\xi|^{pn}), \quad \psi_n = |\eta|^{qn} - E(|\eta|^{qn}) \quad n = 1,2,\dots \tag{3.129}$$

Tekintsük a következő egyenlőséget:

$$\int_{-1}^1 |x|^{pn} f_1(x|y)dx = \frac{1}{(1-|y|^q)^{\frac{1}{p}}} \int_0^{(1-|y|^q)^{\frac{1}{p}}} x^{pn}dx = \frac{1}{pn+1}(1-|y|^q)^n, \tag{3.130}$$

ebből a feltételes várható értékekre

$$\mathbf{A}_\eta \varphi_n = \frac{1}{pn+1}(1-|\eta|^q)^n - E(|\xi|^{pn}), \quad \mathbf{A}_\xi \psi_n = \frac{1}{qn+1}(1-|\xi|^p)^n - E(|\eta|^{qn}), \tag{3.131}$$

ekkor pedig az együtthatókra a 8.Tétel szerint

$$b_{nn} = \frac{(-1)^n}{pn+1}, \quad b'_{nn} = \frac{(-1)^n}{qn+1}; \quad n = 1,2,\dots, \tag{3.132}$$

így a sajátértékek a következőek:

$$\lambda_n = \frac{1}{\sqrt{(pn+1)(qn+1)}}; \quad n = 1,2,\dots \tag{3.133}$$

Az ilyen sajátértékekhez tartozó sajátfüggvények viszont nem minden esetben alkotnak teljes rendszert, így szükséges $\mathbf{C}(\xi,\eta)$ meghatározása. Belátható, hogy ezek a λ-k adják az összes nem nulla sajátértéket, mivel belátható (lásd [5]), hogy

$$1+\mathbf{C}^2(\xi,\eta)=1+\sum_{n=1}^{\infty}\lambda_n^2. \tag{3.134}$$

Ekkor tehát a maximálkorreláció értéke

$$\mathbf{S}(\xi,\eta)=\max_n\frac{1}{\sqrt{(pn+1)(qn+1)}}=\frac{1}{\sqrt{(p+1)(q+1)}}, \tag{3.135}$$

a kapcsolódó sajátfüggvények pedig

$$f_1=|\xi|^p-\frac{1}{p+2},\quad g_1=|\eta|^q-\frac{1}{q+2}. \tag{3.136}$$

Érdemes megjegyezni, hogy ebben a példában a korreláció és a korrelációs hányados is zérus, valamint a maximálkorreláció kiszámításakor akkor és csak akkor helyettesíthető ξ és η rendre $|\xi|$-vel és $|\eta|$-val, ha $\mathbf{S}(\xi,\eta)$-hoz páros sajátfüggvénypár tartozik. Ennek következtében nyilván a T tartomány helyett is elegendő csak a felső félsíkba vagy akár az első síknegyedbe eső részét vizsgálni.

Valószínűségi változók sorozatának részösszegei

Korábban említettük, hogy az együttes normális eloszlás nem szükséges követelménye annak, hogy a maximálkorreláció és a hagyományos korreláció értéke megegyezzen, erre most bemutatunk egy lehetőséget.

Dembo, Kagan és Shepp[7] belátták a következő állítást független azonos eloszlású valószínűségi változók sorozatára, felhasználva Efron és Stein egy állítását ilyen változók egy szimmetrikus függvényének felbontásáról:

10.Állítás. *Legyenek $\xi_1,\xi_2,\dots$ független, azonos eloszlású valószínűségi változók, legyen $S_k=\sum_{i=1}^{k}\xi_i$! Ha egy h függvényre $E\big(h(S_k)\big)^2<\infty$, akkor $l\le k$ számokra*

$$E\big(E(h(S_k)|S_l)\big)^2\le\frac{l}{k}E\big(h(S_k)\big)^2+\frac{1-l}{k}\big(E(h(S_k))\big)^2. \tag{3.137}$$

Az állítást valójában egy általánosabb alakban látták be, egészen pontosan ha $h(\xi_1, \dots, \xi_k)$ szimmetrikus függvény véges második momentummal, akkor

$$E\big(E(h(\xi_1, \dots, \xi_k)|\xi_1, \dots, \xi_l)\big)^2 \leq$$
$$\leq \frac{l}{k}E\big(h(\xi_1, \dots, \xi_k)\big)^2 + \frac{1-l}{k}\big(E(h(\xi_1, \dots, \xi_k))\big)^2 \qquad (3.138)$$

egyenlőtlenség teljesül. Ezt felhasználva bizonyíthatjuk a következő tételt:

9.Tétel. *Legyenek $\xi_1, \xi_2, \dots$ független, azonos eloszlású valószínűségi változók, $E(\xi^2) < \infty$, $S_k = \sum_{i=1}^k \xi_i$! Ekkor ha $m \leq n$ pozitív egészek, akkor S_m és S_n maximálkorrelációja megegyezik a klasszikus korrelációjukkal, és így nem függ a valószínűségi változók eloszlásától. A maximálkorreláció értéke ekkor*

$$\mathbf{S}(S_m, S_n) = \mathbf{R}(S_m, S_n) = \sqrt{\frac{m}{n}}, \qquad m \leq n. \qquad (3.139)$$

Bizonyítás. Válasszuk a megfelelő terekből f és g függvényeket úgy, hogy $E\big(f(S_m)\big) = 0$, $E\big(g(S_n)\big) = 0$, $E\big(f(S_m)^2\big) < \infty$ és $E\big(g(S_n)^2\big) < \infty$ teljesüljön! Ekkor a teljes várható érték tétele szerint

$$E\big(f(S_m)g(S_n)\big) = E\big(f(S_m)E(g(S_n)|S_m)\big) \qquad (3.140)$$

teljesül. A Cauchy-Schwarz-Bunyakovszkij egyenlőtlenség és az előző állítás felhasználásával ekkor

$$|E\big(f(S_m)g(S_n)\big)|^2 \leq E\big(f(S_m)\big)^2 E\big(E(g(S_n)|S_m)\big)^2 \leq \frac{m}{n}E\big(f(S_m). \qquad (3.141)$$

Mivel ez minden megfelelő f és g függvény esetén igaz, a maximálkorreláció definíciója szerint

$$\mathbf{S}^2(S_m, S_n) \leq \frac{m}{n}. \qquad (3.142)$$

A másik irányú egyenlőtlenséghez tekintsük először a hagyományos korrelációt, ennek értéke

$$\mathbf{R}(S_m, S_n) = \frac{E\big((S_m - E(S_m))(S_n - E(S_n))\big)}{D(S_m)D(S_n)} = \sqrt{\frac{m}{n}}, \qquad (3.143)$$

valamint a (3.17) egyenlőtlenség szerint

$$\mathbf{S}(S_m, S_n) \geq \mathbf{R}(S_m, S_n) = \sqrt{\frac{m}{n}}. \tag{3.144}$$

Mivel mindkét irányban megkaptuk az egyenlőtlenséget, a tétel igaz. □

Megjegyezendő, hogy $E(\xi_i^2) = \infty$ esetében a fenti bizonyításból is látható, hogy $\mathbf{S}(S_m, S_n) \leq \sqrt{\frac{m}{n}}$, ha $m \leq n$ pozitív egészek, a másik irány viszont nem látszik. Az állítás végtelen szórásnégyzetek esetén is igaz, egy bizonyítását közölte Novak[11] 2004-ben, majd 2005-ben Bryc, Dembo és Kagan[2] szintén tárgyalta a problémát. Az állítás nem azonos eloszlású valószínűségi változók esetén nem ismert.

Markov-láncok és maximálkorreláció

Markov-láncok konvergenciájának vizsgálatakor kiderül, hogy a konvergencia szoros kapcsolatban áll a lánc két állapotának korreláltságával. Ilyen irányú vizsgálatok esetén is érdemes a maximálkorrelációt vizsgálnunk, és bizonyos esetekben pontos értékét is meg tudjuk határozni.

Legyen $\{X_t\}$ egy irreducibilis, véges állapotterű folytonos Markov-lánc $\mathbf{p}(t) = \big(p_{ij}(t)\big), t \geq 0$ átmenetvalószínűségekkel, és a $\mathbf{Q} = (q_{ij})$ infinitezimális generátorral! Legyen a lánc π stacionárius eloszlása időben megfordítható, azaz teljesüljön

$$p_{ij}\pi_j = \pi_i p_{ij}! \tag{3.145}$$

Innentől eszerint az eloszlás szerint vizsgáljuk a dolgokat.

11.Állítás. *Ekkor a maximálkorreláció értéke*

$$\mathbf{S}(X_t, X_0) = e^{\lambda_1 t}, \qquad t \geq 0, \tag{3.146}$$

ahol $\lambda_1 < 0$ jelöli a $\boldsymbol{Q}$ mátrix legnagyobb nemtriviális sajátértékét.

Bizonyítás. A maximálkorreláció f és g transzformáló függvényeire igaz, hogy mivel definíció szerint $Ef = Eg = 0$ és $\|f\| = \|g\| = 1$,

$$\mathbf{R}(f(X_t), g(X_0)) = \langle \mathbf{p}(t) f, g \rangle, \tag{3.147}$$

és így létezik a térben $\{\varphi_i\}$ ortonormált bázis az átmenetmátrix sajátvektoraiból. Mivel az átmenetmátrixra

$$\mathbf{p}(t) = e^{\mathbf{Q}t} = \sum_{k=0}^{\infty} \frac{\mathbf{Q}^k t^k}{k!}, \tag{3.148}$$

az átmenetmátrix sajátértékei kifejezhetők a generátor sajátértékeiből, tehát ha $\mathbf{Q}$ sajátértékeit λ_i jelöli, akkor $\mathbf{p}(t)$ sajátértékei $e^{\lambda_i t}$ alakban állnak elő.

Az ortonormált bázisban írjuk fel a transzformáló függvényeket:

$$f = \sum_i \langle f, \varphi_i \rangle \varphi_i, \quad g = \sum_i \langle g, \varphi_i \rangle \varphi_i. \tag{3.149}$$

Ekkor látható, hogy speciálisan $f_0 = g_0 = \varphi_1$ esetén

$$\mathbf{R}(f_0(X_t), g_0(X_0)) = \langle \mathbf{p}(t) f_0, g_0 \rangle = \langle \mathbf{p}(t) \varphi_1, \varphi_1 \rangle = \langle e^{\mathbf{Q}t} \varphi_1, \varphi_1 \rangle =$$

$$= \langle e^{\lambda_1 t} \varphi_1, \varphi_1 \rangle = e^{\lambda_1 t} \langle \varphi_1, \varphi_1 \rangle = e^{\lambda_1 t}, \tag{3.150}$$

tehát léteznek olyan transzformáló függvények, amelyekkel a korreláció az állításbeli értéket veszi fel.

Általános, de természetesen standard esetben pedig a következő egyenlőtlenség igaz:

$$\mathbf{R}(f(X_t), g(X_0)) = \langle \mathbf{p}(t) f, g \rangle = \left\langle e^{\mathbf{Q}t} \sum_i \langle f, \varphi_i \rangle \varphi_i, \sum_i \langle g, \varphi_i \rangle \varphi_i \right\rangle =$$

$$= \left\langle \sum_i e^{\lambda_i t} \langle f, \varphi_i \rangle \varphi_i, \sum_i \langle g, \varphi_i \rangle \varphi_i \right\rangle = \sum_i \langle e^{\lambda_i t} \langle f, \varphi_i \rangle \varphi_i, \langle g, \varphi_i \rangle \varphi_i \rangle =$$

$$= \sum_i e^{\lambda_i t} \langle f, \varphi_i \rangle \langle g, \varphi_i \rangle \le \sum_i e^{\lambda_i t} \left(\frac{\langle f, \varphi_i \rangle^2 + \langle g, \varphi_i \rangle^2}{2} \right) \le$$

$$\le e^{\lambda_1 t} \sum_i \left(\frac{\langle f, \varphi_i \rangle^2 + \langle g, \varphi_i \rangle^2}{2} \right) = e^{\lambda_1 t} \left(\frac{\sum_i \langle f, \varphi_i \rangle^2 + \sum_i \langle g, \varphi_i \rangle^2}{2} \right) =$$

$$= e^{\lambda_1 t} \cdot \frac{1+1}{2} = e^{\lambda_1 t}, \tag{3.151}$$

tehát bármilyen más transzformáló függvények esetén sem lehet nagyobb a módosított korreláció, tehát a (3.146) egyenlőség igaz. □

Érdemes megjegyezni, hogy diszkrét paraméterű stacionárius lánc esetére Liu bizonyított hasonlót, szoros összefüggésben a lánc konvergenciájának vizsgálatával, visszavezetve a konvergenciasebesség kérdését a dolgozatban tárgyalt operátorproblémákhoz hasonló eszközökkel egy kovarianciaproblémára, majd felhasználva saját 1991-es eredményét, mely szerint a maximálkorreláció kifejezhető a következő alakban is:

$$\mathbf{S}(\xi, \eta) = \sum_{E(h)=0, D^2(h)=1} D^2\big(E(h(\xi), \eta)\big). \tag{3.152}$$

Ez természetesen teljesen hasonló az operátorok tárgyalásakor látott egyenlőségekhez, amelyek a feltételes várható érték operátorának normájáról szóltak.

Egyenletes eloszlású mintaelemek

A következő példa elsősorban egy felső becslést ad bizonyos esetben a maximálkorreláció értékére, valamint speciális esetben egyenlőséget is láthatunk. A becslést adó tételt speciálisan kételemű mintára, ahola mintaelemek eloszlása folytonos, Terrell látta be 1983-ban, a következő tétel ennek egy általánosabb formája, Székely és Móri[17] tétele:

10.Tétel. *Legyenek $X_1, X_2, \ldots, X_n$ független, azonos eloszlású valószínűségi változók, és jelölje $\hat{X}_1 \le \hat{X}_2 \le \ldots \le \hat{X}_n$ a belőlük álló rendezett statisztikát. Rögzített $1 \le i \le j \le n$ mellett tegyük fel, hogy $\hat{X}_i$ és $\hat{X}_j$ véges szórásúak. Ekkor*

$$\mathbf{R}(\hat{X}_i, \hat{X}_j) \le \left(\frac{i(n+1-j)}{j(n+1-i)}\right)^{\frac{1}{2}}, \tag{3.153}$$

és egyenlőség akkor és csak akkor áll fenn, ha az X_i valószínűségi változók egyenletes eloszlásúak.

A tételt a szerzők ennél általánosabban bizonyítják, valójában ezt a felső becslést a maximálkorrelációra is belátják. A tétel bizonyításából kiderül, hogy mivel a (0,1) intervallumon egyenletes eloszlásból vett minta esetén a rendezett mintaelemek béta-eloszlásúak, azaz sűrűségfüggvényeik

$$f_{\hat{X}_i}(x) = B(i, n+1-i)^{-1} x^{i-1}(1-x)^{n-i}, \quad 0 < x < 1 \tag{3.154}$$

és

$$f_{\hat{X}_j}(y) = B(j, n+1-j)^{-1} y^{j-1}(1-y)^{n-j}, \quad 0 < y < 1, \tag{3.155}$$

az együttes eloszlásfüggvény pedig

$$f_{\hat{X}_i,\hat{X}_j}(x,y) = B(i, j-i, n+1-j)^{-1} x^{i-1}(y-x)^{j-i-1}(1-y)^{n-j}, \quad 0 < x < y < 1, \tag{3.156}$$

ahol

$$B(t_1, t_2, \ldots, t_k) = \frac{\Gamma(t_1)\Gamma(t_2)\cdots\Gamma(t_k)}{\Gamma(t_1 + t_2 + \ldots + t_k)}, \tag{3.157}$$

azaz a béta-eloszlásnál megismert normáló konstans, ha a ξ és η valószínűségi változók együttes eloszlása ilyen, akkor a tétel szerint teljesül az egyenlőség, tehát ha

$$f_{\xi,\eta}(x,y) = f_{\alpha\beta\gamma}(x,y) = B(\alpha,\beta,\gamma)^{-1} x^{\alpha-1}(y-x)^{\beta-1}(1-y)^{\gamma-1},$$

$$0 < x < y < 1, \alpha, \beta, \gamma > 0, \tag{3.158}$$

akkor

$$\mathbf{S}(\xi,\eta) = \left(\frac{\alpha\gamma}{(\alpha+\beta)(\gamma+\beta)}\right)^{\frac{1}{2}}. \tag{3.159}$$

Ebből az általános esetből következik szintén a fent említett szerzők példája:

Legyenek az $X_1, X_2, \ldots, X_n$ valószínűségi változók egyenletes eloszlásúak a következő tartományon:

$$\left\{ x \in \mathbb{R}^n \quad \sum_{i=1}^{n} |x|^{\frac{1}{t_i}} \leq 1 \right\}, \quad 0 < t_i < \infty, \tag{3.160}$$

ekkor

$$\mathbf{S}(X_1, X_2) = \mathbf{R}(|X_1|^{\frac{1}{t_1}}, 1 - |X_2|^{\frac{1}{t_2}}) = \left(\frac{t_1 t_2}{(1 + T - t_1)(1 + T - t_2)} \right)^{\frac{1}{2}}, \quad (3.161)$$

ahol $T = \sum_{i=1}^{n} t_i$. Ez pedig onnan adódik, hogy $|X_1|^{\frac{1}{t_1}}$ és $1 - |X_2|^{\frac{1}{t_2}}$ együttes sűrűségfüggvénye pontosan olyan, mint az előbb tárgyalt általánosabb esetben, $\alpha = t_1$, $\beta = 1 + T - t_1 - t_2$ és $\gamma = t_2$ helyettesítésekkel, valamint belátható, hogy $\mathbf{S}(X_1, X_2) = \mathbf{S}(|X_1|, |X_2|)$. Látható, hogy ez az eset a Bártfai példájánál leírtak általánosítása.

3.2.3. Algoritmikus közelítés

A konkrétan kiszámítható esetek mellett meglehetősen sokan tárgyalnak módszereket az optimális transzformáló függvények vagy a sajátértékek közelítésére. Nyilvánvalóan cél lehet egy általános érvényű módszer megalkotása, amellyel kiszámítható esetben a létező optimális függvényeket tetszőlegesen jól közelíti, és olyan esetekben, amikor nem léteznek legjobb transzformáló függvények, akkor is valamilyen szempontból optimális eredményt ad. A következő szakaszban áttekintünk két lehetőséget röviden a maximálkorreláció iteratív közelítésére.

Az első algoritmus tárgyalásához tekintsünk Sarmanov[16] eredeti megközelítését a maximálkorrelációra, teljesen analóg módon az eddigiekkel:

Legyen $f(x, y)$ a korrelációt meghatározó sűrűségfüggvény ξ és η valószínűségi változókra, amelyeket a

$$T = \{a \leq \xi \leq b; \quad c \leq \eta \leq d\} \quad (3.162)$$

tartományon tekintünk, ami akár végtelen is lehet. A valószínűségi változók peremeloszlásaira ekkor

$$f_1(x) = \int_c^d f(x, y) dy; \quad f_2(y) = \int_a^b f(x, y) dx. \quad (3.163)$$

Tegyük fel, hogy a $k(x,y) = \frac{f(x,y)}{\sqrt{f_1(x)f_2(y)}}$ magfüggvény négyzetesen integrálható mindkét változó szerint! Az $f(x,y)$ sűrűségfüggvény meghatároz két szimmetrikus sűrűségfüggvényt, mégpedig

$$g_1(x,y) = \int_c^d \frac{f(x,t)f(y,t)}{f_2(t)}dt \quad \text{és} \quad g_2(x,y) = \int_a^b \frac{f(t,x)f(t,y)}{f_1(t)}dt \tag{3.164}$$

függvényeket, és két szimmetrikus magfüggvényt, nevezetesen

$$k_1(x,y) = \frac{g_1(x,y)}{\sqrt{f_1(x)f_1(y)}} \quad \text{és} \quad k_2(x,y) = \frac{g_2(x,y)}{\sqrt{f_2(x)f_2(y)}} \tag{3.165}$$

függvényeket. Ezek a magfüggvények pozitívak és ugyanazok a sajátértékeik, jelölje ezeket $1 < \lambda_1^2 \leq \lambda_2^2 \leq \ldots$, valamint a két magfüggvényhez tartozó sajátfüggvényeket rendre $|\varphi_i(x)|$ és $|\psi_i(x)|$, ahol $i = 1, 2, \ldots$, ekkor természetesen a sajátfüggvényekről nem állíthatjuk általánosságban, hogy megegyeznének. Ekkor a korábban definiáltakkal teljesen analóg módon az $f(x,y)$ együttes sűrűségfüggvénnyel meghatározott eloszláshoz tartozó maximálkorrelációra $\mathbf{S}(\xi,\eta) = \frac{1}{|\lambda_1|}$ teljesül, és teljesen hasonlóan a korábbiakhoz a 8.Állítás igaz, miszerint a definiált szimmetrikus eloszlásokhoz tartozó maximálkorreláció értéke a négyzete az eredeti eloszláshoz tartozónak, azaz értéke $\frac{1}{\lambda_1^2}$.

A bevezetett tárgyalásmódra nézve a maximálkorreláció kiszámítására a következő közelítő eljárást adhatjuk:

Legyen az $r_0(y)$ kezdő közelítésünk egy tetszőleges függvény, aminek van szórása és várható értéke 0! Ekkor a következő iterációt tekintjük:

$$r_{2k+1}(x) = \int_c^d r_{2k}(y)\frac{f(x,y)}{f_1(x)}dy \tag{3.166}$$

és

$$r_{2k}(y) = \int_a^b r_{2k+1}(x)\frac{f(x,y)}{f_2(y)}dx, \tag{3.167}$$

ahol $k = 1, 2, \ldots$

Sarmanov bizonyította a módszer konvergenciáját, valamint hogy az első sajátfüggvények meghatározhatók normáló konstansokkal való szorzás erejéig, mint határérték, egészen pontosan

$$a_1\psi_1(y) = \lim_{k\to\infty} r_{2k}(y)\lambda_1^{2k} \tag{3.168}$$

és
$$b_1\varphi_1(x) = \lim_{k\to\infty} r_{2k+1}(x)\lambda_1^{2k}. \tag{3.169}$$
Így ha k elég nagy, akkor a szimmetrikus eloszlásokhoz tartozó maximálkorreláció értéke
$$\frac{1}{\lambda_1^2} \approx \frac{r_{2k}(y)}{r_{2k-2}(y)} \approx \frac{r_{2k+1}(x)}{r_{2k-1}(x)}. \tag{3.170}$$

A következő módszer nem a maximálkorreláció értékét, hanem kiszámításához az optimális transzformáló függvényeket közelíti. A módszer alapvetően sokkal általánosabb esetben ad jó algoritmust optimális transzformáló függvényekre. Valójában a regresszióanalízisben adódó feltételes várható értékek esetén a változókat gyakran valamilyen függvénybe helyettesítve módosítják a feladatot, hogy optimalizálják a problémát, például Y valószínűségi változó és a feltételt képező X_i, $i = 1, \ldots, p$ valószínűségi változók esetén inkább $\theta(Y)$ és $\phi_i(X_i)$ változókat vizsgálják, minimalizálandó a
$$\frac{E\big((\theta(Y) - \sum_{i=1}^{p} \phi_i(X_i))^2\big)}{D^2\big(\theta(Y)\big)} \tag{3.171}$$
kifejezést. Breiman és Friedman[1] algoritmusa bizonyos feltételek mellett optimális függvényeket kiszámító módszert ad a problémára. Látható, hogy $p = 1$ esetben ez pontosan a maximálkorreláció optimális függvényeire ad közelítést. Az eddigi jelöléseinkkel tehát minimalizálni akarjuk a következő kifejezést:
$$E\big(f(\xi) - g(\eta)\big)^2, \tag{3.172}$$
ahol persze feltehetünk minden korábbit a várható értékekre és szórásokra. Ekkor rögzített $g(\eta)$ mellett, megtartva $E(f^2(\xi)) = 1$-et, az f-re vett minimalizálásra adódik
$$f'(\xi) = \frac{E(g(\eta)|\xi)}{\|E(g(\eta)|\xi)\|}, \tag{3.173}$$
azaz egy normált feltételes várható érték, ami szintén analóg a korábban tárgyaltakkal. Természetesen a másik irányú minimalizálásra teljesen hasonlóan adható g' képlete, de ezt elegendő normálás nélkül tekintenünk.

Ekkor a módszer úgynevezett alternáló feltételes várható értékeken alapuló algoritmus(ACE), és legalapvetőbb formája a következő:

Legyen $f(\xi) = \frac{\xi}{\|\xi\|}$! Egy iterációs lépés legyen az alábbi:
$g'(\eta) = E(f(\xi)|\eta)$, *helyettesítsük* g *helyére* g'*-t;*
$f'(\xi) = E(g(\eta)|\xi)$, *helyettesítsük* f *helyére* f'*-t!*

Az iterációs lépést addig ismételgetjük, amíg $E\big(f(\xi) - g(\eta)\big)^2$ csökken, mikor nem csökken tovább, megállunk, és a kapott f és g függvényeket választjuk.

Breiman és Friedman belátták, hogy az algoritmus valóban a kérdéses optimális transzformáló függvényeket adja meg, természetesen általánosabb esetben is, nem csak két változóra, továbbá beláttak Rényi, Csáki és Fischer eredményeihez teljesen hasonlókat az optimális függvények létezésére, valamint kapcsolatukra a feltételesvárhatóérték-operátorokkal. A fent említett módszert tárgyalja Buja[3] is, emellett egy úgynevezett ALS-algoritmust, azaz alternáló legkisebb négyzetes közelítésekkel dolgozó algoritmust is.

Köszönetnyilvánítás

Szeretnék köszönetet mondani témavezetőmnek, Móri Tamásnak türelméért, segítségéért a téma szakirodalmának összegyűjtésében, a LaTeX használatában való javaslatokért valamint a dolgozat felépítésében és szerkesztésében adott hasznos tanácsokért és iránymutatásért.

Irodalomjegyzék

[1] Breiman, L., Friedman J. : Estimating optimal transformations for multiple regression and correlation. *J. Amer. Stat. Assoc.* **80** (1985) 580-598.

[2] Bryc, W., Dembo, A., Kagan, A. : On the maximal correlation coefficient *Theory Probab. Appl.* **49** (2005) 132-138.

[3] Buja, A. : Remarks of functional canonical variates, alternating least squares methods and ACE. *Annals of Statistics* **18** (1990) 1032-1069.

[4] Csáki P., Fischer J. : On bivariate stochastic connection. *MTA Mat. Kut. Int. Közl.* **5** (1960) 311-323.

[5] Csáki P., Fischer J. : Contributions to the problem of maximal correlation. *MTA Mat. Kut. Int. Közl.* **5** (1960) 325-337.

[6] Csáki P., Fischer J. : On the general notions of maximal correlation. *MTA Mat. Kut. Int. Közl.* **8** (1963) 27-51.

[7] Dembo, A., Kagan, A., Shepp, L. A. : Remarks on the maximum correlation coefficient. *Bernoulli* **7** (2001) 343-350.

[8] Gebelein, H. : Das statistische Problem der Korrelation als Variation- und Eigenwertproblem und sein Zusammen-

hang mit Ausgleichsrechnung. *Zeitschrift für angewandte Mathematik und Mechanik* **21** (1947) 364-379.

[9] Lancaster, H. O. : Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika* **44** (1957) 289-292.

[10] Linfoot, E. H. : An informational measure of dependence. *Information and Control* **1** (1957) 85-89.

[11] Novak, S. Y. : On Gebelein's correlation coefficient. *Statistics and Probability Letters* **69** (2004) 299-303.

[12] Rényi A. : New version of the probabilistic generalization of the large sieve. *Acta Math. Acad Sci Hung* **10** (1959) 218-226.

[13] Rényi A. : On measures of depence. *Acta Math. Acad. Sci. Hung.* **10** (1959) 441-451.

[14] Rényi A. : Valószínűségszámítás. *Tankönyvkiadó* (1981)

[15] Sarmanov, O. V. : The maximal correlation coefficient (symmetric case) *Dok. Akad. Nauk USSR* **120** (1958) 715-718.

[16] Sarmanov, O. V. : The maximal correlation coefficient (nonsymmetric case) *Dok. Akad. Nauk USSR* **121** (1958) 52-55.

[17] Székely G. J., Móri T. F. : An extremal property of rectangular distributions. *Statistics and Probability Letters* **3** (1985) 107-109.

[18] Yu, Y. : On the maximal correlation coefficient. *Statistics and Probability Letters* **78** (2008) 1072-1075.