

Móri Tamás

Főkomponens- és faktoranalízis

Elte Valószínűségelméleti és Statisztika Tanszék, 1999

Főkomponens- és faktoranalízis

Többdimenziós adatsor: sok változóra vonatkozóan vannak megfigyeléseink. A tárgyalandó többdimenziós módszerek céljai lehetnek például:

- A változók számának csökkentése, de úgy, hogy ezáltal a megfigyelésekben rejlő információ ne csökkenjen lényegesen; lényegkiemelés.
- Nehezen megadható fogalmak (pl. gazdasági fejlettség) definiálása összetett mutatórendszerrel való jellemzés útján.
- Osztályozási (csoportosítási) feladatok: a csoportképző ismérvek kijelölt változók nem függetlenek és nem azonos szórásúak, ezért nem lehet azonos súllyal venni figyelembe őket – a változókat kialakító közös faktorok alapján csoportosítunk..

A módszerek jellemzői:

- Számolásiigényesek, számítógépes programcsomagok segítségével hajthatók végre.
- Többdimenziós normális eloszlású megfigyelések esetén optimumtulajdonságokkal rendelkeznek, de bármely más, véges szórású mintaeloszlás esetén is igaz, hogy természetesen definiált célfüggvényeket optimalizálnak.
- A klasszikus módszerek nem robusztusak, érzékenyek a kiugró és extrém értékekre, de léteznek nemparaméteres, robusztusabb változatok is, amelyek rangstatisztikákkal dolgoznak.

A továbbiakban két népszerű módszerről lesz szó: a *főkomponens-analízisről* és a *faktoranalízisről*. Ezek közül a főkomponens-analízis matematikai háttere a kidolgozottabb, a faktoranalízisben számos kérdés máig sincs megnyugtatóan tisztázva. Ezért a főkomponens-analízis főbb állításainak a bizonyításai is szerepelni fognak, feltéve, ha nem túl bonyolultak, míg a faktoranalízisről szóló részben inkább csak a módszerek ismertetésére kerül sor.

Főkomponens-analízis

Legyen X $EX = 0$ várható értékű és $\text{Var}X = D$ kovarianciamátrixú n dimenziós véletlen vektor. Az a célunk, hogy az n dimenziós térben olyan új koordinátarendszert vezessünk be, amelyben a véletlen vektorunk koordinátái már korrelálatlanok. Ismeretes, hogy X -et az u_1, \dots, u_n ortonormált bázisban felírva, az i -edik koordinátája $Y_i = u_i^T X$ lesz, vagyis $Y = U^T X$, ahol U az u_1, \dots, u_n oszlopokból álló ortonormált mátrix. Másképpen mondva, $X = UY$, vagyis X -et egy korrelálatlan komponensű véletlen vektor elforgatottjaként kívánjuk előállítani.

Legyen $\text{Var}Y = \Lambda$ diagonális mátrix, akkor azt kapjuk, hogy $D = U\Lambda U^\top$, ez pedig éppen D spektrálelőállítás. Ennek alapján a következő elnevezéseket vezetjük be.

- Legyen $\text{Var}X$ spektrálelőállítása $D = U\Lambda U^\top = \sum_{i=1}^n \lambda_i u_i u_i^\top$, $\lambda_1 \geq \dots \geq \lambda_n \geq 0$. Ekkor az $Y = U^\top X$ véletlen vektort az X főkomponens-vektorának nevezzük, az Y vektor i -edik koordinátája, $Y_i = u_i^\top X$ pedig az X i -edik főkomponense. Világos, hogy a főkomponensek vektora szintén 0 várható értékű és a kovarianciamátrixa diagonális:

$$\text{Var}Y = U^\top D U = U^\top U \Lambda U^\top U = \Lambda,$$

vagyis a főkomponensek páronként korrelálatlanok és szórásaik, melyeket az X kanonikus szórásainak nevezünk, éppen D sajátértékeinek négyzetgyökei: $DY_i = \sqrt{\lambda_i}$.

- Vegyük észre, hogy a főkomponens-vektor forgatásinvariáns: ha V ortonormált mátrix, X és VX főkomponens-vektora megegyezik. Valóban, mivel $\text{Var}VX = VDV = VU\Lambda U^\top V^\top = W\Lambda W^\top$, ahol $W = VU$ is ortonormált, ezért VX főkomponens-vektora $W^\top VX = U^\top V^\top VX = U^\top X$.

- Az viszont már nem teljesül, hogy a főkomponens-vektor skálainvariáns is lenne (vagyis a mértékegység megváltoztatása általában már megváltoztatja a főkomponenseket). A skálatranszformáció egy Δ diagonális mátrixszal való szorzást jelent, és $U^\top X = U^\top \Delta^{-1} \Delta X$ miatt az invariancia azt jelentené, hogy $\text{Var} \Delta X = \Delta D \Delta$ spektrálelőállításában az ortonormált mátrix $\Delta^{-1} U$ lenne, ez utóbbi azonban rendszerint nem is forgatás.

- Ha $\text{rang } D = k \leq n$, akkor csak az első k főkomponens különbözik 0-tól. Legyen $v_i = \sqrt{\lambda_i} u_i$, $i = 1, \dots, k$, $V = [v_1, \dots, v_k]$ $n \times k$ méretű mátrix, és Z a $Z_i = Y_i / \sqrt{\lambda_i}$ koordináták-ból álló k dimenziós véletlen vektor. Ekkor $X = VZ$, Z szórás mátrixa az egység mátrix és V olyan mátrix, melynek oszlopai ortogonálisak és normáik rendre az X kanonikus szórásaival egyenlők. Ezt az előállítást X kanonikus előállításának nevezzük.

- Ha D sajátértékei mind különböznek, a főkomponensek előjeltől eltekintve egyértelműek. Ha azonban egy sajátérték többszörös, akkor a megfelelő sajátvektorok nem egyértelműek, a sajátaltérben tetszőlegesen elforgathatók.

- *A főkomponensek szemléletes jelentése.*

Legyen e tetszőleges egységvektor, ekkor X -nek az e irányára való vetülete $(e^\top X)e$, és az e irányába eső szórásnégyzete $D^2(e^\top X) = e^\top D e$. Ha e és f két egységvektor, az irányukra eső vetületek pontosan akkor korrelálatlanok, ha $e^\top D f = 0$, ugyanis $\text{cov}(e^\top X, f^\top X) = E(e^\top X X^\top f) = e^\top D f$. Ezért a következő állítás igazolható. A főkomponensek bevezetésénél tekintett új koordinátarendszer első tengelye irányában a legnagyobb az X szórása az összes lehetséges n dimenziós irány közül, és ez a maximum éppen az első kanonikus szórás. Ezt a tengelyt ezért *első főtengelynek* is nevezik. A második koordinátatengely iránya az elsőre merőleges irányok közül az, amelyikre nézve X szórása a legnagyobb, mégpedig éppen a második kanonikus szórással egyenlő, ez a irány a *második főtengely*, és így tovább. Az X vektor vetületei az az új koordinátatengelyekre a főkomponensek. Formálisan a következő összefüggéseket írhatjuk fel:

$$\begin{aligned} \max\{e^\top De : \|e\| = 1\} &= u_1^\top Du_1 = \lambda_1, \\ \max\{e^\top De : \|e\| = 1, e^\top u_j = 0, 1 \leq j < i\} &= u_i^\top Du_i = \lambda_i, \quad 1 < i \leq n. \end{aligned}$$

Ezek bizonyítása i szerinti teljes indukcióval történhet. Az első $i - 1$ u_j -re merőleges irányok mind benne vannak az u_i, \dots, u_n sajátvektorok által generált invariáns altérben, tehát

$$e = (e^\top u_i)u_i + \dots + (e^\top u_n)u_n, \quad (e^\top u_i)^2 + \dots + (e^\top u_n)^2 = \|e\|^2 = 1,$$

$$\begin{aligned} e^\top De &= \sum_{j=i}^n (e^\top u_j)u_j^\top \sum_{k=1}^n \lambda_k u_k u_k^\top \sum_{m=1}^n (e^\top u_m)u_m = \sum_{j=i}^n \sum_{k=1}^n \sum_{m=1}^n \lambda_k (e^\top u_j)(u_j^\top u_k)(u_k^\top u_m)(u_m^\top e) = \\ &= \sum_{j=i}^n \lambda_k (e^\top u_j)^2 \leq \lambda_i, \end{aligned}$$

és itt az egyenlőség teljesül, ha $e = u_i$. \square

- n dimenziós véletlen vektor közelítése alacsonyabb dimenziós változóval.

Legyen $k < n$ és \mathcal{L}_k az olyan n dimenziós Z valószínűségi változók tere, amelyek 1 valószínűséggel egy (legfeljebb) k dimenziós hipersíkban veszik fel az értékeiket (más szóval rang $\text{Var } Z \leq k$). Keressük azt a $Z \in \mathcal{L}_k$ valószínűségi vektorváltozót, amely négyzetes hibában a legjobban közelíti X -et:

$$\mathbb{E}\|X - Z\|^2 \rightarrow \min!$$

A várható érték minimumtulajdonsága miatt $\mathbf{E}Z = \mathbf{E}X = 0$, ezért Z egy legfeljebb k dimenziós altérbe esik. Mivel abban az altérben az X -hez legközelebbi pont PX , ahol P az altérre való merőleges vetítés, így szükségképpen $Z = PX$. Mivel $X - PX$ merőleges PX -re, a Pitagorasz-tétel miatt $\|X - PX\|^2 = \|X\|^2 - \|PX\|^2$, tehát a feladat:

$$\mathbb{E}\|PX\|^2 \rightarrow \max!$$

D spektrálelőállításával kifejezve

$$\text{Var}PX = P \text{Var}X P = \sum_{i=1}^n \lambda_i P u_i u_i^\top P,$$

ezért

$$\mathbb{E}\|PX\|^2 = \text{tr } \text{Var}PX = \sum_{i=1}^n \lambda_i \text{tr } P u_i (P u_i)^\top = \sum_{i=1}^n \lambda_i \|P u_i\|^2.$$

Itt $\|P u_i\|^2 \leq \|u_i\|^2 = 1$, és $I = U U^\top = \sum_{i=1}^n u_i u_i^\top$ miatt

$$\sum_{i=1}^n \|P u_i\|^2 = \sum_{i=1}^n \text{tr } P u_i (P u_i)^\top = \text{tr } P \left(\sum_{i=1}^n u_i u_i^\top \right) P = \text{tr } P \leq k.$$

Ennélfogva akkor kapunk maximumot, ha $\|P u_i\|^2 = 1$, $i = 1, \dots, k$, és $P u_i = 0$ a többi sajátvektorra. Ez azt jelenti, hogy P éppen az u_1, \dots, u_k vektorok által kifeszített altérre

való vetítés: $P = \sum_{i=1}^k u_i u_i^T$, a legjobban közelítő Z valószínűségi vektorváltozó pedig az a vektor, amelynek első k koordinátája az új koordináta-rendszerben éppen az első k főkomponens, a többi pedig 0: $Z = \sum_{i=1}^k Y_i u_i$, végül a négyzetes hiba minimális értéke:

$$\mathbb{E}\|X - Z\|^2 = \lambda_{k+1} + \dots + \lambda_n.$$

- Az $\mathbb{E}\|X\|^2 = \text{tr } D = \lambda_1 + \dots + \lambda_n$ mennyiséget az X vektorváltozó *teljes varianciájának* nevezik. Ez megegyezik az $Y = U^T X$ főkomponens-vektor teljes varianciájával (hiszen a forgatás normatartó), és az első k főkomponens segítségével felírt legjobban közelítő k dimenziós Z vektorváltozó teljes varianciája $\mathbb{E}\|Z\|^2 = \lambda_1 + \dots + \lambda_k$. Ezt úgy mondják, hogy az első k főkomponens *ennyit magyaráz meg az X teljes varianciájából*. Ha ez az érték már az X teljes varianciájának a túlnyomó része, akkor a többi változó elhanyagolható. Ezért k megválasztásához a

$$\psi = \frac{\lambda_1 + \dots + \lambda_k}{\lambda_1 + \dots + \lambda_n}$$

hányados értékére támaszkodnak.

Amikor az első k főkomponenset választjuk ki a közelítéshez, akkor D -t két részre bontjuk: $D = D^{(1)} + D^{(2)}$, ahol $D^{(1)} = \text{Var } Z$ az ún. *reprodukált mátrix* (az elnevezés onnan ered, hogy ha az első k főkomponens már teljesen megmagyarázza X teljes varianciáját, akkor visszakapjuk D -t), $D^{(2)} = \text{Var}(X - Z)$ neve pedig: *reziduális kovarianciamátrix*.

- *Normális eloszlás esete*

Ha X többdimenziós normális eloszlású, akkor ugyanez igaz a főkomponens-vektorra is, és a főkomponensek korrelálatlanságából következik a függetlenségük is.

- *Tapasztalati főkomponensek*

Legyen adott egy n dimenziós eloszlásból vett N elemű minta: X_1, X_2, \dots, X_N ($N > n$). Definiáljuk a tapasztalati eloszlást a szokásos módon: ez az az n -dimenziós diszkrét eloszlás, amely $1/N - 1/N$ súlyt (valószínűséget) helyez a megfigyelések mindegyikére. A valószínűségeloszlásokra definiált szokásos mennyiségi jellemzőket (pl. várható érték, szórás, medián) erre a tapasztalati eloszlásra kiszámítva a mennyiségek ún. *tapasztalati megfelelőit* kapjuk. Általánosan alkalmazható szabály, hogy egy mennyiség várható értéke a tapasztalati eloszlás szerint nem más, mint a mennyiségnek a mintaelemeken vett átlaga. Ilyen módon a *tapasztalati várható érték* az \bar{X} -sal jelölt *mintaátlag*, a kovarianciamátrix tapasztalati megfelelője pedig az

$$S_N = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})(X_i - \bar{X})^T$$

tapasztalati kovarianciamátrix. A tapasztalati kovarianciamátrix $S_N = V_N \Lambda_N V_N^T$ spektráloroállítása segítségével definiálhatjuk a minta *tapasztalati főkomponenseit, kanonikus szórásait és fő tengelyeit*: most az i -edik tapasztalati főkomponens egy N dimenziós vektor, amelynek j -edik elemét úgy kapjuk, hogy a V_N mátrix i -edik oszlopának elemeivel súlyozva adjuk össze az X_j mintaelem koordinátáit. Ezért a V_N ortonormált mátrix neve: *átviteli (loading) mátrix*, oszlopait *súlyvektoroknak*, elemeit pedig *súlyoknak* nevezzük.

Természetesen igazak maradnak a főkomponensekről eddig kimondott tételeink tapasztalati megfelelői: az első fő tengely az az egyenes, amelynek irányában a legnagyobb a minta tapasztalati szórása, a második fő tengely az elsőre merőleges egyenesek közül rendelkezik ezzel a tulajdonsággal, és így tovább. Az első k tapasztalati főkomponens a minta teljes varianciájából, $\text{tr } S_N$ -ből

$$\hat{\psi} = \frac{\hat{\lambda}_1 + \dots + \hat{\lambda}_k}{\hat{\lambda}_1 + \dots + \hat{\lambda}_n}$$

hányadot magyaráz meg, ahol $\hat{\lambda}_i$ az S_N i -edik legnagyobb sajátértéke, és segítségükkel minimalizálható a $\sum_{i=1}^N \|X_i - Z_i\|^2$ négyzetösszeg az összes olyan Z_1, \dots, Z_N vektor- N -eseken, amelyek által \mathbb{R}^n -ben kifeszített alter dimenziója legfeljebb k .

- *Főkomponensek becslése normális eloszlású minta esetén.*

Ismeretes, hogy ilyenkor az eloszlás D kovarianciamátrixának a maximum-likelihood becslése az S_N tapasztalati kovarianciamátrix. A maximum likelihood becslés invarianciájának következtében S_N sajátvektorai és sajátértékei a D sajátértékeinek és sajátvektorainak maximum likelihood becslései. Ha azonban feltételezzük, hogy a λ sajátérték multiplicitása többszörös: $\lambda_{q+1} = \dots = \lambda_{q+r}$, akkor emellett λ maximum likelihood becslése $\hat{\lambda} = (\hat{\lambda}_{q+1} + \dots + \hat{\lambda}_{q+r})/r$. Mivel ilyenkor a főkomponensek közül az ezeknek a sajátértékeknek megfelelőek nem egyértelműek, csak az $u_{q+1}Y_{q+1} + \dots + u_{q+r}Y_{q+r}$ összeg az (mégpedig X vetülete a megfelelő sajátaltérre), ezért a tapasztalati közelítéshez ezek a főkomponensek csak egyszerre használhatók.

Különösen fontos annak a hipotézisnek a vizsgálata, hogy az utolsó r sajátérték megegyezik-e. A $H_0 : \lambda_{n-r+1} = \dots = \lambda_n$ hipotézis tesztelésére *likekhood-hányados próbát* végezhetünk. A próbastatisztika:

$$Nr(\log a - \log g),$$

ahol a az S_N tapasztalati kovarianciamátrix legkisebb r sajátértékének számtani közepe, g pedig a mértani közepe. Megmutatható, hogy H_0 teljesülése esetén ennek $N \rightarrow \infty$ melletti határeloszlása $f = (r+2)(r-1)/2$ szabadságfokú χ^2 -eloszlás, ezért χ^2 -próba végezhető. A próba *Bartlett-féle módosítása* az, hogy a próbastatisztikában N helyett $N - (2n+1)/6$ írandó (ez a határeloszlást nem befolyásolja, de a próba kismintás tulajdonságait javítja).

A közelítéshez felhasznált főkomponensek számának megállapításához szükséges lehet a teljes varianciából magyarázott ψ hányadra vonatkozó hipotézisek vizsgálata, illetve – ami matematikailag ezzel egyenértékű – konfidenciaintervallum konstruálása. Belátható, hogy $\sqrt{N}(\hat{\psi} - \psi)$ határeloszlása 0 várható értékű és

$$\tau^2 = \frac{2 \operatorname{tr} D^2}{(\operatorname{tr} D)^2} (\psi^2 - 2h\psi + h)$$

szórásnégyzetű normális eloszlás, ahol

$$h = \frac{\lambda_1^2 + \dots + \lambda_k^2}{\lambda_1^2 + \dots + \lambda_n^2}.$$

Ennek alapján pl. a ψ -re vonatkozó, $1 - \alpha$ aszimptotikus megbízhatósági szintű *konfidenciaintervallum* végpontjai: $\hat{\psi} \pm c_{\alpha/2} \hat{\tau} (N-1)^{-1/2}$, ahol a $\hat{\tau}$ a megfelelő mennyiség tapasztalati becslését jelöli, $c_{\alpha/2}$ pedig a standard normális eloszlás felső $\alpha/2$ -kvantilise.

- *Főkomponensek standardizált mintából*

Előfordul, hogy a főkomponensek kiszámítása előtt a mintát (koordinátáinként) standardizálják: az X_i mintaelemeket úgy transzformálják, hogy az $X_i - \bar{X}$ koordinátáit rendre elosztják a tapasztalati kovarianciamátrix átlós elemeinek a négyzetgyökével. Ez annak felel meg, hogy nem a kovarianciamátrix, hanem a korrelációs mátrix spektrálfelbontását használjuk, vagyis átskálázást alkalmazunk, amelyről láttuk már, hogy nem hagyja a főkomponenseket változatlanul. Az így számított főkomponensekkel való közelítés nem a két minta megfelelő elemei közötti négyzetes eltérések összegét minimalizálja, hanem egy olyan súlyozott összeget, amelyben az i -edik koordinátára (változóra) vonatkozó hibanégyzet-összeget elosztjuk a szóbanforgó változóra vonatkozó megfigyelések tapasztalati szórásnégyzetével, vagyis a nagyobb szórású változónál fellépő eltéréseket kisebb súllyal vesszük figyelembe. Ennek a célfüggvénynek is lehet létjogosultsága.

Faktoranalízis

A faktoranalízis modelljében feltételezzük, hogy a megfigyelt korrelált változók nem megfigyelhető hipotetikus háttérváltozók, ún. *faktorok* lineáris kombinációjaként írhatók le. E faktorokat hatásuk szerint a következőképpen szokták osztályozni:

1. *közös faktor*: amely több megfigyelt változót is befolyásol. Ezek között lehetnek
 - 1.1 *általános faktorok*, ezek minden megfigyelt változót befolyásolják, illetve
 - 1.2 *csoporthfaktorok*, amelyek néhány (egynél több, de nem az összes) megfigyelt változóra hatnak.
2. *egyedi faktor*: csak egyetlen változót befolyásol. Ez részben vagy egészében származhat mérési hibából, amelyet *hibafaktornak* is szoktak nevezni, de nem különíthető el a más forrású egyedi faktortól.

Feltételezzük, hogy a faktorok korrelálatlanok és 0 várható értékűek.

- Ennek megfelelően a faktoranalízis modellje:

$$X = AY + W + m,$$

ahol X a megfigyelt változókból álló n dimenziós véletlen vektor, Y a közös faktorok k dimenziós véletlen vektora ($k < n$), W az egyedi faktorok n dimenziós véletlen vektora, $A = (a_{ij})$ egy $n \times k$ méretű rögzített mátrix, ennek a neve most is *átviteli (loading) mátrix*, elemei pedig a *faktorsúlyok*, végül m rögzített n dimenziós vektor. Feltételeink szerint $EY = 0$, $EW = 0$, továbbá $\text{Var}Y = I_k$ és $\text{Var}W = \Psi$ diagonális mátrix, átlójában a $\psi_i \geq 0$ elemekkel ($1 \leq i \leq n$), végül $\text{cov}(Y, W) = EYW^T = 0$. Azt mondjuk, hogy X *leírható a k -faktormodellel*, ha (eloszlása) előállítható a fenti alakban.

- Világos, hogy $EX = m$, és $\text{Var}X := D = AA^T + \Psi$. Speciálisan az i -edik változó szórásnégyzete $d_{ii} := D^2 X_i = h_i^2 + \psi_i$, ahol a

$$h_i^2 = \sum_{j=1}^k a_{ij}^2$$

mennyiséget *kommunalitásnak*, a ψ_i részt pedig *egyedi varianciának* nevezik. A továbbiakban feltesszük, hogy $m = 0$.

- X pontosan akkor írható le k -faktormodellel, ha létezik olyan nemnegatív elemű diagonális Ψ mátrix, amellyel $D - \Psi$ legfeljebb k rangú pozitív szemidefinit mátrix. Mint láttuk, ez a feltétel szükséges, de elegendő is, mert egy legfeljebb k rangú $n \times n$ -es pozitív szemidefinit mátrix mindig felírható AA^T alakban, ahol A $n \times k$ -as mátrix (ha ugyanis az előállítandó mátrix spektrálfelbontása

$$\lambda_1 u_1 u_1^T + \dots + \lambda_k u_k u_k^T,$$

akkor $v_i = \sqrt{\lambda_i} u_i$, $i = 1, \dots, k$ jelöléssel $A = [v_1, \dots, v_k]$ megfelelő lesz). Ezért ekkor olyan $n + k$ dimenziós Z véletlen vektort keresünk, amelynek koordinátái korrelálatlanok és egységnyi szórásúak, és X felírható $X = [A, \Psi^{1/2}]Z$ alakban. Ismeretes, hogy ez megtehető, mert $D = [A, \Psi^{1/2}][A, \Psi^{1/2}]^T$.

- Ha X leírható k -faktormodellel, akkor ez átskálázás után is igaz marad rá, ugyanis ha Δ diagonális mátrix, akkor $\Delta X = \Delta AY + \Delta W$, és itt ΔW koordinátái, lévén W koordinátáinak különféle számszorosai, továbbra is korrelálatlanok egymással és Y -nal.

Észrevehetjük, hogy a közös faktorok vektora és az átviteli mátrix nem egyértelmű: a közös faktorokat tetszőleges G ortonormált mátrixszal elforgatva GY továbbra is eleget tesz a közös faktoroktól megkövetelt kovarianciafeltételeknek, és az $X = AG^T GY + W$ előállítás is megfelel a k -faktormodellel, vagyis az elforgatott közös faktorokhoz „visszaforgatott” átviteli mátrix tartozik. A mátrixok *szingulárisérték-felbontását* alkalmazva viszont már megmutatható, hogy Ψ -t rögzítve A forogástól eltekintve egyértelműen meghatározott: ha A és B mérete megegyezik és $AA^T = BB^T$, akkor létezik olyan U ortonormált mátrix, amellyel $B = AU$.

Az egyértelműség eléréséhez A -ra további feltételeket szoktak tenni. *Lawley* javasolta, hogy olyan átviteli mátrixot keressünk, amelyre $A^T \Delta A$ diagonális valamilyen előre meg-

adott nemnegatív elemű Δ diagonális mátrixra. Δ -tól azt is megköveteljük, hogy az átló elemei monoton csökkenő sorrendben következzenek egymás után.

$\Delta = I_n$ esetén $A^T A$ diagonális mátrix, azaz A oszlopai páronként ortogonálisak, ilyenkor A -t *ortogonális átviteli mátrixnak* nevezzük. Ekkor *főfaktor-analízisről, főfaktor-modellről* beszélünk.

A $\Delta = \Psi^{-1}$ választás normális eloszlású k -faktormodell esetén egyszerűsíti a maximum likelihood becslés kiszámítását. (A megfigyelt változók sorrendjét permutálva elérhetjük, hogy $\psi_1 \leq \psi_2 \leq \dots \leq \psi_n$ teljesüljön.)

Másik lehetőség: Δ -t választhatjuk a D mátrix főátlójából képezett diagonális mátrix inverzének: $\Delta = \text{diag}(1/d_{11}, 1/d_{22}, \dots, 1/d_{nn})$. Mivel a faktormodell skálainvariáns, gyakori szokás a változókat standardizálni az elemzés előtt. Ez $\Delta^{1/2}$ -del való skálázást jelent, ezért az átskálázott átviteli mátrix ortogonalitása az eredeti modellben éppen az $A^T \Delta A$ mátrix diagonalitását jelenti.

- Nem ismeretesek általános módszerek annak megállapítására, hogy egy n dimenziós X véletlen vektor mikor írható le k -faktormoddellel. A $D = AA^T + \Psi$ összefüggésben $n + nk$ ismeretlen és $n(n+1)/2$ egyenlet van, a Lawley-féle feltétel pedig további $k(k-1)/2$ egyenletet jelent. Ezért a felírható egyenletek és az ismeretlenek számának különbsége:

$$\frac{n(n+1) + k(k-1)}{2} - n(k+1) = \frac{1}{2}((n-k)^2 - (n+k)).$$

Ha ez pozitív, nem várható, hogy az egyenletrendszernek van megoldása.

- Másik fontos, de nem teljesen megoldott probléma az *identifikáció* kérdése. Azt mondjuk, hogy a modell identifikálható, ha adott k esetén D *egyértelműen* meghatározza Ψ -t és forgatástól eltekintve A -t. Egy részeredmény például a következő: ha az A mátrix bármely sorát elhagyva a megmaradt sorokat úgy lehet két csoportra osztani, hogy a nekik megfelelő két mátrix rangja egyaránt k , akkor $D = AA^T + \Psi$ identifikálható.

Ha konvergencia becslési eljárásokat (pl. maximum likelihood módszer) alkalmazunk, akkor elég nagy elemű mintára már elvárható, hogy A és Ψ becslései már kellően közel vannak a becslendő mátrixokhoz. Ezért az identifikálhatóság helyett megelégedhetünk A és Ψ *lokális identifikálhatóságával* is, ami azt jelenti, hogy van olyan (euklideszi) környezetük, amelyben már teljesül az egyértelműség. Erre vonatkozóan a következő mondható: ha $A^T \Psi^{-1} A$ diagonális és a $\Psi - A(A^T \Psi^{-1} A)^{-1} A^T$ mátrix elemeinek a négyzetéből álló mátrix nemszinguláris, akkor A és Ψ lokálisan identifikálható.

- *Maximum likelihood becslés normális eloszlású modellben*

Tegyük fel, hogy a (közös és az egyedi) faktorok függetlenek és normális eloszlásúak, ekkor X is normális eloszlású. Ebből az eloszlásból áll rendelkezésünkre egy N elemű minta: X_1, X_2, \dots, X_N ($N > n$). Feltesszük, hogy D és Ψ nemszinguláris, továbbá $A^T \Psi^{-1} A$ diagonális.

Ismert, hogy a 0 várható értékű és $D = AA^T + \Psi$ kovarianciamátrixú n dimenziós normális eloszlásból vett minta loglikelihood-függvénye

$$L(A, \Psi) = -\frac{N}{2} F(A, \Psi) + \text{konst.}$$

alakú, ahol $F(A, \Psi) = \log \det D + \text{tr}(D^{-1} S_N)$ és S_N a tapasztalati kovarianciamátrix. A maximum likelihood becsléshez $F(A, \Psi)$ -t kell minimalizálni. F -et A , illetve Ψ szerint differenciálva és a deriváltakat 0-val egyenlővé téve kapjuk a *likelihood-egyenleteket*:

$$\frac{\partial F}{\partial A} = \left(\frac{\partial F}{\partial a_{ij}} \right)_{n \times n} = 2D^{-1}(D - S_N)D^{-1}A, \quad \frac{\partial F}{\partial \psi_i} = [D^{-1}(D - S_N)D^{-1}]_{ii}, \quad 1 \leq i \leq n.$$

Ezeket iterációval oldják meg: előbb rögzített Ψ esetén az első egyenletből A -t fejezik ki, majd ezzel az A -val a második egyenletrendszerből meghatározzák Ψ -t, ezután ismét rögzített Ψ mellett minimalizálnak A -ban, stb.

Az első egyenlet megoldása (vagyis az átviteli mátrix maximum likelihood becslése ismert nonszinguláris Ψ mátrix esetén) a $\Psi^{-1/2} S_N \Psi^{-1/2}$ mátrix spektrálfelbontása segítségével kapható meg. Legyen

$$\Psi^{-1/2} S_N \Psi^{-1/2} = V \Lambda V^T,$$

ahol a Λ diagonális mátrix főátlójában a $\hat{\lambda}_i$ sajátértékek sorrendje monoton fogyó. Legyen $V_{(k)}$ a V első k oszlopából álló $n \times k$ méretű mátrix, és

$$Q_k = \text{diag}[(\hat{\lambda}_1 - 1)^+, \dots, (\hat{\lambda}_k - 1)^+].$$

Ezzel A maximum likelihood becslése:

$$\hat{A} = \Psi^{1/2} V_{(k)} Q_k.$$

Az iteráció konvergenciáját befolyásolja az induló Ψ megválasztása. *Jöreskog* javaslata:

$$\psi_i = \left(1 - \frac{k}{2n} \right) \frac{1}{s^{ii}},$$

ahol s^{ii} az S_N^{-1} mátrix főátlójának i -edik eleme.

- *A közös faktorok becslésére két módszer használatos.*

A *Bartlett-féle becslés* abból indul ki, hogy ismert A és Ψ esetén X feltételes eloszlása Y -ra nézve AY várható értékű és Ψ kovarianciamátrixú n dimenziós normális eloszlás. Ezért a feltételes loglikelihood-függvény

$$L(Y, X) = -\frac{1}{2} (X - AY)^T \Psi^{-1} (X - AY) + \text{konst.},$$

amelynek maximumhelyét Y szerinti deriválással kaphatjuk meg:

$$\frac{\partial L}{\partial Y} = (X - AY)^T \Psi^{-1} A = 0,$$

ebből

$$\hat{Y} = (A^T \Psi^{-1} A)^{-1} A^T \Psi^{-1} X.$$

(Ezt más úton is levezethetjük, ha észrevesszük, hogy Y értékét rögzítve a faktormodell korrelált hibájú normális lineáris modellbe megy át, ahol – mint ismeretes – az Y paraméter maximum likelihood becslése megegyezik a legkisebb négyzetes becsléssel.)

Látható, hogy a Bartlett-féle becslés torzítatlan.

A közös faktorok becslésének *Thompson-féle módszere* Y -t olyan paraméternek tekinti, amelynek a priori eloszlása standard normális, ezután Y -ra Bayes-becslés adható. Négyzetes veszteségfüggvény mellett a Bayes-becslés éppen az a posteriori eloszlás várható értéke. Az a posteriori sűrűségfüggvény a Bayes-tétellel

$$f(Y|X) = \text{konst}(X) \cdot \exp\left\{-\frac{1}{2}(X - AY)^T \Psi^{-1}(X - AY)\right\} \exp\left\{-\frac{1}{2}Y^T Y\right\},$$

ahol $\text{konst}(X)$ nem függ Y -tól. A kitevőben álló mennyiség (a $-1/2$ szorzó elhagyásával) a következőképpen alakítható át: bevezetve a $B = I_k + A^T \Psi^{-1} A$ jelölést,

$$\begin{aligned} (X - AY)^T \Psi^{-1}(X - AY) + Y^T Y &= Y^T B Y - X^T \Psi^{-1} A Y - Y^T A^T \Psi^{-1} X + X^T \Psi^{-1} X = \\ &= (Y - B^{-1} A^T \Psi^{-1} X)^T B (Y - B^{-1} A^T \Psi^{-1} X) + \text{konst}(X). \end{aligned}$$

Erről leolvasható, hogy az a posteriori eloszlás is n dimenziós normális, $B^{-1} A^T \Psi^{-1} X$ várható értékkel és B^{-1} kovarianciamátrixszal. Ezért Y Bayes-becslése

$$\hat{Y} = (I_k + A^T \Psi^{-1} A)^{-1} A^T \Psi^{-1} X.$$

Bár ez a Thompson-becslés torzított, az átlagos négyzetes hibája kisebb, mint a Bartlett-becslésé.

- *A faktorok forgatása*

Az átviteli mátrixnak Lawley módszerével történő egyértelművé tétele a becslési eljárások matematikai elemzését segíti, de az az ára, hogy a kapott közös faktorok gyakran csak nehezen értelmezhetők. Alkalmas elforgatással esetleg szemléletesebb jelentést adhatunk a faktoroknak. Ha például a faktorsúlyok között csak 0-hoz közeli vagy aránylag nagy értékek fordulnak elő, akkor a változók csoportosíthatók annak alapján hogy melyik faktor mely változóknak játszik fontos szerepet, szerencsés esetben a változók halmaza akár diszjunkt osztályokra is bontható.

A faktorsúlyok forgatására különféle módszereket használnak, a legismertebb ezek közül a *varimax forgatás*. Ennek az a célja, hogy minél több 0-hoz közeli faktorsúlyt állítson elő. A módszer lényege egy alkalmas kvadratikus célfüggvény maximalizálása, amit iterációval valósítanak meg. A feladat olyan $A = (a_{ij})$ átviteli mátrix előállítására egymás utáni forgatásokkal, amelyre a kommunalításokkal normált faktorsúly-négyzetek szórásnégyzete maximális, vagyis a maximalizálandó célfüggvény

$$\sum_{j=1}^k \sum_{i=1}^n (b_{ij} - \bar{b}_j)^2 = \sum_{j=1}^k \sum_{i=1}^n b_{ij}^2 - n \sum_{j=1}^k \bar{b}_j^2,$$

ahol

$$b_{ij} = \frac{a_{ij}^2}{h_i^2}, \quad \bar{b}_j = \frac{1}{n} \sum_{i=1}^n b_{ij}.$$

Mivel a forgatások az átviteli mátrix sorainak normáit nem változtatják meg, az egész eljárás során a kommunalítások változatlanok maradnak. A fokozatos maximalizálás történhet például úgy is, hogy minden lépésben csak egy-egy faktorpárt forgatunk el: ha a (j, m) párt választjuk ($1 \leq j < m \leq k$), akkor csak az a_{ij} és a_{im} faktorsúlyok változnak:

$$a'_{ij} = a_{ij} \cos \varphi - a_{im} \sin \varphi, \quad a'_{im} = a_{ij} \sin \varphi + a_{im} \cos \varphi,$$

ahol φ az elforgatás szöge; ez lépésenként csak egyváltozós szélsőérték-feladat megoldását jelenti. Minden ciklusban végighaladunk minden páron (összesen $k(k-1)/2$ pár van), és a ciklus végén ellenőrizzük a célfüggvény változását. Akkor állunk meg, ha már csak elhanyagolható mértékben változik.

- *Főfaktor-analízis*

A főfaktorok módszerének alkalmazásánál gyakran feltételezik az egyedi faktorok teljes hiányát. Így a feladat olyan közös faktorok meghatározására redukálódik, amelyek páronként ortogonálisak, a lehető legnagyobb mértékben járulnak hozzá az összes változó teljes kommunalitásához, azaz a

$$\sum_{i=1}^n h_i^2 = \mathbb{E}\|X\|^2$$

teljes varianciához, illetve amelyekkel a reprodukált mátrix lehető legjobban megközelíti D -t (a legkisebb négyzetek elv értelmében). Formálisan ez megegyezik a főkomponens-analízisben vizsgált feladattal, így most nem tárgyaljuk. Ha az egyedi faktorokat is belevesszük a modellbe, de valahonnan ismernénk Ψ -t, vagy ami ezzel egyenértékű, a kommunalításokat, akkor a feladat ismét főkomponens-analízis, de az $X - W$ véletlen vektoron (amit a $D - \Psi$ mátrix spektrálelőállításával segítségével végezhetünk el). Szokás ezért a kommunalításokat különféle módokon becsülni.

Amikor már előzetesen standardizált változókból indulnak ki, a D kovarianciamátrix voltaképpen korrelációs mátrix. A h_i^2 kommunalitás leggyakoribb becslése az i -edik változónak az összes többi változóra vonatkozó többszörös korrelációs együtthatójának négyzete. (Ez a többszörös korrelációs együttható definíció szerint a maximális korreláció, amely az i -edik változó és a többi változó tetszőleges lineáris kombinációja között előfordul). Igazolható, hogy ez a becslés a következő módon is megkapható:

$$\hat{h}_i^2 = 1 - \frac{1}{d^{ii}},$$

ahol d^{ii} a D^{-1} mátrix főátlójának i -edik eleme. Dwyer megmutatta, hogy ez mindig kisebb vagy egyenlő, mint a kommunalítás valódi értéke.

Irodalom

- Éltető Ödön, Meszéna György, Ziermann Margit: *Sztocasztikus módszerek és modellek*, Közgazdasági és Jogi Könyvkiadó, Budapest, 1982.
- Fazekas István (szerk.): *Bevezetés a matematikai statisztikába*, egyetemi jegyzet, Kossuth Egyetemi Kiadó, Debrecen, 1997.
- Füstös László, Meszéna György, Simonné Mosolygó Nóra: *A sokváltozós adatelemzés statisztikai módszerei*, Akadémiai Kiadó, Budapest, 1986.
- Móri F. Tamás, Székely J. Gábor. (szerk.): *Többváltozós statisztikai analízis*, Műszaki Könyvkiadó, Budapest, 1986.
- Dr. Sváb János: *Többváltozós módszerek a biometriában*, Mezőgazdasági Kiadó, Budapest, 1979.
- Ziermann Margit, Michaletzky György: Idősorok faktoranalízise. *Sigma*, **XXVI.** (1995) 77–91.

Függelék: Jelölések, tudnivalók

EX az X valószínűségi (vektor)változó *várható értéke*

D^2X az X valószínűségi változó *szórásnégyzete*

$\text{Var } X$ az X valószínűségi vektorváltozó *kovarianciamátrixa*

$\text{cov}(X, Y)$ az X és az Y valószínűségi (vektor)változók *kovarianciája*

a^\top, A^\top az a vektor, ill. az A mátrix *transzponáltja*

$\text{tr } A$ az A négyzetes mátrix *nyoma* (az átlóban álló elemek összege)

- a tr függvény lineáris: $\text{tr } cA = c \text{ tr } A$, $\text{tr } (A + B) = \text{tr } A + \text{tr } B$
- a tr függvény ciklikusan invariáns: $\text{tr } AB \dots Z = \text{tr } B \dots ZA$,
- szimmetrikus mátrix nyoma egyenlő a sajátértékeinek az összegével

I, I_n az n dimenziós *egységmátrix*

$\text{diag}[a_1, \dots, a_n]$ olyan (ún. *diagonális*) mátrix, amelynek főátlójában az a_1, \dots, a_n elemek állnak, a főátlón kívüli elemek pedig 0-val egyenlők.

projekció: lineáris altérre való merőleges vetítés, ill. annak mátrixa

- P pontosan akkor projekció, ha szimmetrikus ($P^\top = P$) és idempotens ($P^2 = P$)
- $\text{tr } P = \text{rang } P = \dim \text{Im } P$

Ortonormált mátrix (forgatás): U négyzetes mátrix, amelyre $U^{-1} = U^\top$, más szóval, U sorai, ill. oszlopai egyaránt ortonormált rendszert alkotnak

Spektrálelőállítás: Az A $n \times n$ -es szimmetrikus mátrix spektrálelőállítása:

$$A = U \Lambda U^\top = \sum_{i=1}^n \lambda_i u_i u_i^\top, \text{ ahol } U \text{ az } u_1, \dots, u_n \text{ oszlopokból álló ortonormált mátrix,}$$

Λ diagonális mátrix, az átlójában álló $\lambda_1, \dots, \lambda_n$ számok az A mátrix sajátértékei, u_i pedig a λ_i sajátértékhez tartozó sajátvektor.

Mivel egy k dimenziós altérre való vetítésnek k darab 1 sajátértéke van, a többi 0, ezért az altérre való P vetítés spektrálelőállítása:

$$P = \sum_{i=1}^k u_i u_i^\top, \text{ ahol } u_1, \dots, u_k \text{ az } \text{Im } P \text{ altér ortonormált bázisa.}$$