Lineáris módszerek a többdimenziós statisztikában.

E jegyzetben a Bolla–Krámli könyv hatodik fejezetének az eredményeit ismertetem. Ez egy felületes ismertetés, amelyben csak a legfontosabb eredményekről és fogalmakról írok. A bizonyításokat elhagyom.

Az első szekció témája a főkomponens analízis. Itt a kiinduló probléma a következő. Legyen adva egy p-dimenziós normális eloszlású $\mathbf{X}$ véletlen vektor $\mathbf{C}$ kovarianciamátrixszal és $\mathbf{m}$ várható értékkel. Írjuk ezt

$$\mathbf{X} = \mathbf{UZ} + \mathbf{m} \tag{1}$$

alakban, ahol $\mathbf{U}$ ortogonális mátrix, $\mathbf{Z}$ pedig olyan p-dimenziós normális eloszlású vektor, nulla várható értékkel, amelynek koordinátái függetlenek.

E probléma megoldásáról szól az 1.1 tétel. A bizonyítás a szimmetrikus mátrixok spektrálfelbontásán (diagonizálásán) alapul, illetve ennek kapcsolatán egy szimmetrikus mátrix sajávektoraival, és ezeknek a sajátvektoroknak a tulajdonságain. Az (1) képletben szereplő $\mathbf{Z}$ vektort hívják az irodalomban főkomponensvektornak, koordinátáit pedig főkomponenseknek. Az 1.2 állítás azt mondja ki, hogy ha a normális eloszlású $\mathbf{X}$ vektort megszorozzuk egy ortogonális vektorral (elforgatjuk), akkor főkomponensei nem változnak.

Az 1.3 tétel azzal a kérdéssel foglalkozik, hogy, ha adva van egy $\mathcal{N}_p(\mathbf{0}, \mathbf{C})$ eloszlású $\mathbf{X}$ vektor, akkor adott $k \leq p$ számra, mi az $\mathbf{X}$ vektor legjobb k-dimenziós közelítése. Azaz olyan k-dimenziós $\mathbf{A}$ lineáris transzformációt keresünk, amelyikre az $E\|\mathbf{X} - \mathbf{AX}\|^2$ kifejezés a lehető legkisebb. Az 1.3 tétel szerint ez az optimális k-dimenziós $\mathbf{A}$ leképezés a $\mathbf{C}$ kovarianciamátrix k legnagyobb sajátértékhez tartozó sajátvektor által kifeszített altérre való vetítés.

Az első szekció még egy eredmény megfogalmazását tartalmazza bizonyítás nélkül. Ebben az eredményben ismertetik azt a próbát, amely lehetővé teszi, hogy eldöntsük egy minta segítségével, hogy egy nulla várható értékű véletlen normális eloszlású vektor kovarianciamátrixának a k legkisebb sajátértéke megegyezik-e.

A második szekció témája a faktoranalízis. Azzal a problémával foglalkozik, hogy mikor és hogyan lehet egy $\mathcal{N}_p(\mathbf{C}, \mathbf{m})$ eloszlású $\mathbf{X}$ vektort felbontani viszonylag kis k-dimenziójú standard normális eloszlású véletlen $\mathbf{f}$ vektor lineáris transzformációja plusz egy tőle független $\mathbf{e}$ normális vektor

plusz egy determinisztikus $\mathbf{m}$ vektor összegére úgy, hogy az $\mathbf{e}$ normális vektor koordinátái függetlenek. Képletben kifejezve az $\mathbf{X}$ vektor következő alakú előállítását keressük.

$$\mathbf{X} = \mathbf{A}\mathbf{f} + \mathbf{e} + \mathbf{m}. \tag{2}$$

ahol $\mathbf{A}$ $p \times k$ méretű mátrix, $\mathbf{f}$ k-dimenziós statndard normális eloszlású valószínűségi változó, $\mathbf{f}$ az $\mathbf{e}$ vekortól független p-dimenziós normális eloszlású vektor, amelynek a koordinátái korrelálatlanok és 0 várható értékűek. Az $\mathbf{f}$ vektort *közös faktornak* az $\mathbf{e}$ vektort *egyedi faktornak* nevezik az irodalomban.

Az $\mathbf{e}$ és $\mathbf{f}$ vektorokra tett feltételek így fogalmazhatóak meg.

$$\begin{aligned} E\mathbf{f} &= 0, & E\mathbf{f}\mathbf{f}^T &= \mathbf{I}_k \\ E\mathbf{e} &= 0, & E\mathbf{e}\mathbf{e}^T &= \mathbf{D} \\ & & E\mathbf{e}\mathbf{f}^T &= \mathbf{0} \end{aligned}$$

ahol $\mathbf{D}$ diagonális mátrix. A (2) formulából következik, hogy

$$\mathbf{C} = \mathbf{A}\mathbf{A}^T + \mathbf{D}. \tag{3}$$

A 2.1 tétel szerint a (3) egyenlet akkor és csak akkor oldható meg egy $p \times k$ méretű $\mathbf{A}$ mátrixszal, ha létezik olyan $\mathbf{D}$ diagonális mátrix nem negatív elemekkel, amelyre $\mathbf{C} - \mathbf{D}$ pozitív szemidefinit mátrix, és rangja nem nagyobb, mint k.

A valódi feladat, nem egy ismert $\mathbf{C}$ kovarianciamátrix felbontása a (3) alakban, hanem az, hogy van egy $\mathcal{N}_p(\mathbf{C}, \mathbf{m})$ eloszlású mintánk ismeretlen $\mathbf{C}$ kovarianciamátrixszal és $\mathbf{m}$ várható értékkel. Ezután keressük a $\mathbf{C} = \mathbf{A}\mathbf{A}^T + \mathbf{D}$ alakú kovarianciamátrix maximum likelihood becslését (rögzített k paraméterrel.) Fel szokták tenni az egyértelmű előállítás kedvéért azt, hogy

$$\mathbf{A}^T\mathbf{D}^{-1}\mathbf{A} \text{ is diagonális mátrix.}$$

A könyv ismerteti ennek a feladatnak a megoldását bizonyítás nélkül, illetve javaslatot tesz arra, hogyan kell a számolásokat végrehajtani számítógép segítségével.

A harmadik szekció témája a többváltozós regresszióanalízis. A lineáris regresszióval foglalkozunk. E feladatban egy Y (függő) valószínűségi változónak keressük a legjobb lineáris becslését $X_1, \ldots, X_k$ (független) valószínűségi változók segítségével. Ismerjük az EY és EX_j, $1 \leq j \leq k$ várható értékeket

valamint a Var Y szórásnégyzetet és a Cov (X_j, Y) és Cov (X_i, X_j) kovarianciákat, $1 \leq i, j \leq k$.

A feladat a

$$E(Y - (a_1X_1 + \cdots + a_pX_p) - b)^2 \tag{4}$$

minimumának a meghatározása az $a_1, \ldots, a_p$ és b együtthatók függvényében. E feladat megoldását tartalmazza a 3.1 állítás. Ennek megfogalmazásához vezessük be a következő jelöléseket. Legyen $\mathbf{C} = (\text{Cov } (X_i, X_j))$, $1 \leq i, j \leq p$, $\mathbf{d} = (\text{Cov } (X_1, Y), \ldots, \text{Cov } (X_p, Y))^T$, $\mathbf{a} = (a_1, \ldots, a_p)^T$. Ezzel a jelöléssel igaz a következő állítás.

3.1 Állítás. *Legyen a* $\mathbf{C}$ *mátrix invertálható. Ekkor a (4) kifejezés optimumát az*

$$\mathbf{a} = \mathbf{C}^{-1}\mathbf{d},$$

és

$$b = EY - a_1EX_1 - \cdots - a_pEX_p$$

paraméterekre veszi fel.

Vezessük be az $l(\mathbf{X}) = a_1X_1 + \cdots + a_pX_p + b$ valószínűségi változót. Ennek segítségével definiáljuk az Y (függő) és $X_1, \ldots, X_p$ (független) változók közötti $r_{Y(X_1,\ldots,X_p)}$ többszörös korrelációt, mint az $r_{Y(X_1,\ldots,X_p)} = \text{Corr}(Y, l(\mathbf{X}))$ korrelációt. Ezzel jól lehet mérni az $l(\mathbf{X})$ közelítés $\varepsilon = Y - l(\mathbf{X})$ hibáját. Nevezetesen, mivel Var $(Y) =$ Var $(l(\mathbf{X}))$ + Var (ε),

$$\text{Var } (\varepsilon) = \text{Var } (Y)(1 - r^2_{Y(X_1,\ldots,X_p)}).$$

A 3.2 állítás a többszörös korreláció egy optimumtulajdonságát fogalmazza meg.

3.2 Állítás. Az $X_1, \ldots, X_p$ valószínűségi változók tetszőleges $h(\mathbf{X})$ lineáris kombinációjára

$$|r_{Y(X_1,\ldots,X_p)}| = |\text{Corr } (Y, l(\mathbf{X}))| \geq |\text{Corr } (Y, h(\mathbf{X}))|.$$

Megjegyzem, hogy a könyv hibás bizonyítást közöl erre az állításra. Olyan $h(x)$ függvényt kellett volna venni, amelyre $E(Y - h(\mathbf{X}))h(\mathbf{X}) = 0$. Ezután meg kell érteni a $h(\mathbf{X}) = l(\mathbf{X})$ függvénynek az optimum tulajdonságát.

A negyedik szekció témája szintén egy lineáris regresszió típusú probléma. Adva van egy p változós $y = a_1x_1 + \cdots + a_px_p$ lineáris leképezés, és az ebben a leképezésben szereplő $a_1, \ldots, a_p$ együtthatókat szeretnénk meghatározni

mérések segítségével. A mérések száma n, de az egyes méréseket csak hibával tudjuk elvégezni. Másrészt a determinisztikus p-dimenziós $(x_{j,1}, \ldots, x_{j,p})$, $1 \leq j \leq n$, mérési pontokat mi választhatjuk meg. Az egyes mérések ε_i, $1 \leq i \leq n$, hibája független nulla várható értékű σ^2 szórásnégyzetű normális eloszlású valószínűségi változó. Elvégzünk n mérést, és keressük az $a_1, \ldots, a_n$ paramétereknek azt a becslését, amelyre a hibák négyzetösszegének a várható értéke a lehető legkisebb. Feltesszük, hogy a mérések n számára $n \geq p$.

A feladat pontos matematikai megfogalmazása a következő. Legyen $\mathbf{a} = (a_1, \ldots, a_p)^T$, jelölje az n mérési eredményt illetve annak hibáját az

$$\mathbf{Y} = (Y_1, \ldots, Y_n)^T \quad \text{és} \quad \varepsilon = (\varepsilon_1, \ldots \varepsilon_n)^T$$

vektor. Legyen továbbá $\mathbf{X}$ az az $n \times p$ méretű mátrix, amelynek i-edik oszlopa az $\mathbf{x}_i = (x_{1,i}, \ldots, x_{n,i})^T$, $1 \leq i \leq p$, vektor. Akkor a méréseink eredményét az

$$\mathbf{Y} = \mathbf{X}\mathbf{a} + \varepsilon$$

képlet mutatja meg. Ekkor

$$E(\mathbf{Y} - \mathbf{X}\mathbf{a})(\mathbf{Y} - \mathbf{X}\mathbf{a})^T = E\varepsilon\varepsilon^T = \sigma^2 I_n,$$

és az $\mathbf{a}$ vektornak azt az $\hat{\mathbf{a}} = (\hat{a}_1, \ldots, \hat{a}_p)$ becslését keressük, amelyre az

$$\|\mathbf{Y} - \mathbf{X}\hat{\mathbf{a}}\|^2 = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{a}})(\mathbf{Y} - \mathbf{X}\hat{\mathbf{a}})^T$$

minimális. Ez a feladat megoldható, és azt kapjuk, hogy a keresett $\hat{\mathbf{a}}$ vektor az

$$\mathbf{X}^T\mathbf{X}\mathbf{a} = \mathbf{X}^T\mathbf{Y}$$

egyenlet megoldása. Ezt az egyenletet hívják Gauss–féle normálegyenletnek.

A könyv ismerteti ennek az egyenletnek egy geometriai levezetését is. Ennek részleteit nem írom le, de megadom azon mennyiségek definicióját, amelyek ebben az indoklásban megjelennek, mivel azok fontos szerepet játszanak későbbi formulákban is.

Legyen $F \subset R^n$ az az altere az R^n térnek, amelyet az $\mathbf{x}_1, \ldots, \mathbf{x}_p$ vektorok feszítenek ki, és legyen $\mathbf{P}$ az ortogonális vetítés az R^n térben az F altérre.

A Gauss-féle normálegyenletnek mindig van megoldása, de az nem feltétlenül egyértelmű. Egyértelmű a megoldás akkor, ha az $\mathbf{X}^T\mathbf{X}$ mátrix rangja $r = p$. Általában ezzel az esettel fogunk foglalkozni. Ekkor

$$\hat{\mathbf{a}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}. \tag{5}$$

A 4.1 állítás tartalma az, hogy az $r = p$ esetben az $\hat{\mathbf{a}}$ becslés torzítatlan, és ez az eredmény megadja e becslés kovarianciamátrixát is.

4.1 Állítás. *Ha $r = p$, és ε $\mathcal{N}_p(\mathbf{0}, \sigma^2\mathbf{I}_p)$ eloszlású, akkor $\hat{\mathbf{a}}$ $\mathcal{N}_p(\mathbf{a}, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1})$ eloszlású.*

A 4.2 tétel tartalma az, hogy $\hat{\mathbf{a}}$ az $\mathbf{a}$ vektor legjobb torzítatlan becslése.

4.2 Tétel (Gauss–Markov tétel). *Legyen $r = p$, és $\tilde{\mathbf{a}}$ az $\mathbf{a}$ vektor torzítatlan becslése. Ekkor*

$$Cov(\hat{\mathbf{a}}) \leq Cov(\tilde{\mathbf{a}}).$$

Ez az egyenlőtlenség azt jelenti, hogy $Cov(\hat{\mathbf{a}}) - Cov(\tilde{\mathbf{a}})$ pozitív definit mátrix.

A következő eredmény a Gauss–Markov tétel következménye.

4.3 Állítás. *Ha $r = p$, akkor tetszőleges $\mathbf{b} \in R^p$ vektorra az $\hat{\mathbf{a}}$ becslés segítségével definiált $\mathbf{b}^T\hat{\mathbf{a}}$ kifejezés a $\mathbf{b}^T\mathbf{a}$ paraméter torzítatlan becslése, és az ilyen becslések között a legkisebb szórásnégyzetű. (Ezt úgy hívják, hogy* **BLUE** *(Best linear Unbiased Estimate).*

A következő pontban a könyv a σ^2 szórásnégyzet becslésével foglalkozik. Bevezeti az

$$\mathbf{S}_\varepsilon^2 = \|\mathbf{Y} - \mathbf{X}\hat{\mathbf{a}}\|^2 = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{a}})^T(\mathbf{Y} - \mathbf{X}\bar{\mathbf{a}})$$

kifejezést, amit *reziduális varianciának* neveznek.

A könyv megmutatja természetes geometriai meggondolások segítségével, hogy σ^2 természetes becslése $r = p$ esetben

$$\hat{\sigma}^2 = \frac{\mathbf{S}_\varepsilon^2}{n - p},$$

és ez σ^2 torzítatlan becslése. Továbbá $\frac{\mathbf{S}_\varepsilon^2}{\sigma}$ $n-p$ paraméterű χ^2 eloszlású valószínűségi változó. Azt is állítja a könyv, hogy alkalmas feltevések esetén $\hat{\sigma}^2$ σ^2 legkisebb szórásnégyzetű torzítatlan becslése.

A következő 4.4 állítás kissé más problémával foglalkozik, mint ez a szekció. Azt a kérdést vizsgálja, milyen $\mathbf{b} \in R^p$ vektorokra létezik a $\mathbf{b}^T\mathbf{a}$ paraméterfüggvénynek torzítatlan becslése. Megmutatja, hogy akkor és csak akkor, ha a $\mathbf{b}$ vektor az $\mathbf{A}$ sorvektorai által kifeszített altérben van. Ez $r = p$ esetben minden $\mathbf{b} \in R^p$ vektorra teljesül.

A 4.5 állításban a könyv megadja az $\mathbf{a}$ vektor és σ^2 maximum likelihood becslését.

4.5 Állítás. *Ha* ε $\mathcal{N}_p(\mathbf{0}, \sigma^2\mathbf{I}_p)$ *eloszlású, akkor akkor* $\mathbf{a}$ *maximum likelihood becslése a (5) képletben megadott* $\hat{\mathbf{a}}$ *kifejezés, a* σ^2 *maximum likelihood becslése pedig*

$$\hat{\sigma}^2 = \frac{\mathbf{S}_\varepsilon^2}{n}.$$

A negyedik szekció utolsó témája annak a nullhipotézisnek az ellenőrzése, hogy

$$H_0: \qquad a_1 = a_2 = \cdots = a_p.$$

Meghatározza e nullhipotézis tesztelési eljárását, ha a likelihood hányados próbát alkalmazzuk. Azt kapja, hogy az eljárás a következő.

Vegyük az

$$F = \frac{\mathbf{Y}^T\mathbf{P}\mathbf{Y}}{\mathbf{Y}^T(\mathbf{I}-\mathbf{P})\mathbf{Y}}$$

hányadost, ahol $\mathbf{I}$ az identitás operátor, $\mathbf{P}$ pedig a korábban bevezetett $\mathbf{P}$ projekció. Ekkor F $\mathcal{F}(p, n-p)$ eloszlást követ, azaz eloszlása megegyezik két független p és $n-p$ paraméterű χ^2 eloszlású valószínűségi változó hányadosának az eloszlásával. Akkor utasítjuk el a nullhipotézist, ha $F \geq c_\varepsilon$ alkalmas c_ε számmal.

Az ötödik szekció témája a varianciaanalízis. Három fő részből áll. Az első rész témája az egyszempontos varianciaanalízis, a másodiké a kétszempontos varianciaanalízis interakció nélkül, a harmadiké a kétszempontos varianciaanalízis interakcióval. A könyv csak röviden említi a többszempontos varianciaanalízist.

A feladat a három esetben a következő:

a) Egyszempontos varianciaanalízis,

Van k csoport, az i-edik csoportban n_i megfigyelés. Ezek eredménye

$X_{i,j}$ $\mathcal{N}(b_i, \sigma^2)$ eloszlású valószínűségi változó, $1 \leq i \leq k$, $1 \leq j \leq n_i$.

Becsüljük meg a legkisebb négyzetek módszerével a b_i, $1 \leq i \leq k$, paramétereket. Vizsgáljuk a becslés tulajdonságait.

b) Kétszempontos varianciaanalízis interakció nélkül,

Vannak (i, j) pároknak csoportjai, $1 \leq i \leq k$, $1 \leq j \leq p$. Mindegyik (i, j) párra 1 megfigyelést végzünk. Legyen ez

$$X_{i,j} = m + a_i + b_j + \varepsilon_{i,j}, \quad (i = 1, \ldots, k,\ j = 1, \ldots, p),$$

olyan a_i és b_j számokkal, amelyekre

$$\sum_{i=1}^{k} a_i = 0, \quad \sum_{j=1}^{p} b_j = 0, \tag{6}$$

ahol az $\varepsilon_{i,j}$ valószínűségi változók függetlenek $\mathcal{N}(0, \sigma^2)$ eloszlással. (σ^2 ismeretlen.)

Becsüljük meg a legkisebb négyzetek módszerével az a_i, $1 \le i \le k$, és b_j, $1 \le j \le p$, paramétereket. Vizsgáljuk a becslés tulajdonságait.

c) Kétszempontos varianciaanalízis interakcióval,

Vannak (i, j) pároknak csoportjai, $1 \le i \le k$, $1 \le j \le p$. Mindegyik (i, j) párra végzünk n megfigyelést. Legyen ez

$$X_{i,j,l} = m + a_i + b_j + c_{i,j} + \varepsilon_{i,j,l}, \quad (i = 1, \dots, k,\ j = 1, \dots, p,\ l = 1, \dots, n),$$

olyan a_i, b_j és $c_{i,j}$ számokkal, amelyekre

$$\begin{aligned} &\sum_{i=1}^{k} a_i = 0, && \sum_{j=1}^{p} b_j = 0, \\ &\sum_{i=1}^{k} c_{i,j} = 0, \quad j = 1, \dots, p, \\ &\sum_{j=1}^{p} c_{i,j} = 0, \quad i = 1, \dots, k, \end{aligned} \tag{7}$$

ahol az $\varepsilon_{i,j,l}$ valószínűségi változók függetlenek $\mathcal{N}(0, \sigma^2)$ eloszlással. (σ^2 ismeretlen.)

Becsüljük meg a legkisebb négyzetek módszerével az a_i, b_j és $c_{i,j}$, $1 \le i \le k$, $1 \le j \le p$, paramétereket. Vizsgáljuk a becslés tulajdonságait.

Mindhárom esetben a keresett optimumot a Lagrange-féle multiplikátor módszer segítségével tudjuk kiszámolni. Megjegyzem, hogy a (6) feltétel a b) esetben, illetve a (7) feltétel a c) esetben nem jelent igazi megszorítást. Ugyanis, be lehet látni, hogy ha valamely a_i, b_j, m paraméterrendszer teljesíti a b) esetben felírt relációkat, akkor ezek alkalmas lineáris transzformációja teljesíti mind ezeket a relációkat mind a (6) feltételt. Hasonlóan, ha valamely a_i, b_j, $c_{i,j}$, m paraméterrendszer teljesíti a c) esetben felírt relációkat, akkor ezek alkalmas lineáris transzformációja teljesíti mind ezeket a relációkat mind

a (7) feltételt. Továbbá, ha alkalmazzuk a Lagrange-féle multiplikátor módszert, és elhagyjuk azokat a fölösleges egyenleteket, amelyek a többi egyenlet következményei, akkor egy olyan egyenletrendszert kapunk, amelyben a változók száma megegyezik az egyenletek számával. Megadom, hogy hogyan kell definiálni az a_i, b_j és $c_{i,j}$ mennyiségeket a c) esetben.

Ebben az esetben eredetileg olyan m' a'_i, b'_j és $c'_{i,j}$ mennyiségek vannak megadva, amelyekre

$$X_{i,j,l} = m' + a'_i + b'_j + c'_{i,j} + \varepsilon_{i,j,l}, \quad (i = 1, \dots, k,\ j = 1, \dots, p,\ l = 1, \dots, n).$$

Definiáljuk e mennyiségek segítségével az $a_\cdot = \frac{1}{k}\sum_{i=1}^{k} a'_i$, $b_\cdot = \frac{1}{p}\sum_{j=1}^{p} b'_j$, $c_{i\cdot} = \frac{1}{p}\sum_{j=1}^{p} c'_{i,j}$, $c_{\cdot j} = \frac{1}{k}\sum_{i=1}^{k} c'_{i,j}$ és $c_{\cdot\cdot} = \frac{1}{kp}\sum_{i=1}^{k}\sum_{j=1}^{p} c'_{i,j}$ kifejezéseket. Ekkor az $m = m' + a_\cdot + b_\cdot + c_{\cdot\cdot}$, $a_i = a'_i - a_\cdot + c_{i\cdot} - c'_{,,}$, $b_j = b'_j - b'_\cdot + c_{\cdot j} - c_{\cdot\cdot}$ és $c_{i,j} = c'_{i,j} - c_{i\cdot} - c_{\cdot j} + c_{\cdot\cdot}$ mennyiségek teljesítik mind az

$$X_{i,j,l} = m + a_i + b_j + c_{i,j} + \varepsilon_{i,j,l}, \quad (i = 1, \dots, k,\ j = 1, \dots, p,\ l = 1, \dots, n).$$

mind a (7) relációt.

Az a) esetben érdemes megfogalmazni a feladatot kissé különböző módon.

Írjuk fel a b_i számokat $b_i = a_i + m$ alakban, $1 \le i \le k$ úgy, hogy $\sum_{i=1}^{k} n_i a_i = 0$. Ekkor a megfigyeléseink

$$X_{i,j} = a_i + m + \varepsilon_{i,j}, \quad 1 \le i \le k, \quad 1 \le j \le n_i.$$

alakban írhatóak. A

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} \varepsilon_{i,j}^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} (X_{i,j} - a_i - m)^2$$

kifejezés minimumát keressük a $\sum_{i=1}^{k} n_i a_i = 0$ kényszerfeltétel mellett. A megoldás leírása érdekében vezessük be a következő mennyiségeket.

$$\bar{X}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{i,j}, \quad \bar{X}_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{i,j}.$$

Ezzel a jelöléssel a megoldás

$$\hat{m} = \bar{X}_{\cdot\cdot}, \text{ és } \hat{a}_i = \bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot}, \ \ i = 1, \dots, k.$$

A b) esetben a

$$\sum_{i=1}^{k}\sum_{j=1}^{p}(X_{i,j}-m-a_i-b_j)^2$$

kifejezés minimumát kell kiszámolni a (6) kényszerfeltétel mellett. A megoldás leírása érdekében vezessük be a következő mennyiségeket.

$$\begin{aligned}
\bar{X}_{i.} &= \frac{1}{p}\sum_{j=1}^{p}X_{i,j}, \quad i=1,\ldots,k,\\
\bar{X}_{.j} &= \frac{1}{k}\sum_{i=1}^{k}X_{i,j}, \quad j=1,\ldots,p,\\
\bar{X}_{..} &= \frac{1}{kp}\sum_{i=1}^{k}\sum_{j=1}^{p}X_{i,j}.
\end{aligned}$$

Ezzel a jelöléssel a megoldás

$$\hat{m}=\bar{X}_{..},\ \ \hat{a}_i=\bar{X}_{i.}-\bar{X}_{..},\ i=1,\ldots,k, \quad \text{és } \hat{b}_j=\bar{X}_{.j}-\bar{X}_{..},\ j=1,\ldots,p.$$

A c) esetben a

$$\sum_{i=1}^{k}\sum_{j=1}^{p}\sum_{l=1}^{n}(X_{i,j,l}-m-a_i-b_j-c_{i,j})^2$$

kifejezés minimumát kell kiszámolni a (7) kényszerfeltétel mellett. A megoldás leírása érdekében vezessük be a következő mennyiségeket.

$$\begin{aligned}
\bar{X}_{i..} &= \frac{1}{pn}\sum_{j=1}^{p}\sum_{l=1}^{n}X_{i,j,l}, \quad i=1,\ldots,k,\\
\bar{X}_{.j.} &= \frac{1}{kn}\sum_{i=1}^{k}\sum_{l=1}^{n}X_{i,j,l}, \quad j=1,\ldots,p,\\
\bar{X}_{i,j.} &= \frac{1}{n}\sum_{l=1}^{n}X_{i,j,l}, \quad i=1,\ldots,k,\ \ j=1,\ldots,p,\\
\bar{X}_{...} &= \frac{1}{kpn}\sum_{i=1}^{k}\sum_{j=1}^{p}\sum_{l=1}^{m}X_{i,j,l}.
\end{aligned}$$

Ezzel a jelöléssel a megoldás

$$\hat{m}=\bar{X}_{...},\ \ \hat{a}_i=\bar{X}_{i..}-\bar{X}_{...},\ i=1,\ldots,k,\ \ \hat{b}_j=\bar{X}_{.j.}-\bar{X}_{...},\ j=1,\ldots,p$$
$$\text{és } c_{i,j}=\bar{X}_{i,j.}-\bar{X}_{i..}-\bar{X}_{.j.}+\bar{X}_{...}\ .$$

A most kapott kifejezések segítségével megadhatjuk a vizsgált modellekben természetesen felmerülő hipotézisek statisztikai vizsgálatát. Ezt csak röviden ismertetem bizonyítások nélkül.

A könyv először megemlíti, hogy az első lépésben ellenőrzi, hogy az adott modellekben az $\hat{m} = X_{\cdot}$ (vagy $\hat{m} = X_{\cdot\cdot}$ illetve $\hat{m} = X_{\cdots}$) valószínűségi változó várható értéke nulla-e (van-e főhatás), és csak azzal az esettel foglalkozik, amikor ez a várható érték nem nulla. Bár a könyv e várható érték becslésével nem foglalkozik, ez viszonylag egyszerű probléma. Azt kell ellenőrizni, hogy az $X_{\cdot}$ (vagy több pont van az indexben) normális eloszlású valószínűségi változó várható értéke nulla-e. Ezt könnyen ellenőrízhetjük, ha ismerjük az összeadandók szórásnégyzetét. Ha azt nem ismerjük, akkor azt is becsülni kell. Ez lehetséges a könyvben csoportokon belüli négyzetösszeg névvel definiált kifejezések segítségével.

Tekintsük először az a) esetet. Ekkor azt akarjuk ellenőrizni, hogy teljesüle az a H_0 nullhipotézis, amely szerint

$$H_0\colon \ a_1 = a_2 = \ldots = a_k = 0.$$

Be lehet látni, hogy a csoportokon belüli $Q_e = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{i,j} - X_{i.})^2$ négyzetösszegre $\frac{Q_e}{\sigma^2}$ χ^2 eloszlású $n-k$ szabadságfokkal, akár teljesül a nullhipotézis, akár nem. A csoportok közötti $Q_a = \sum_{i=1}^{k} n_i (X_{i.} - X_{..})^2$ négyzetösszegre $\frac{Q_a}{\sigma^2}$ a nullhipotézis teljesülése esetén a Q_e valószínűségi változótól független χ^2 eloszlású valószínűségi változó $k-1$ szabadságfokkal. E valószínűségi változó értéke a nullhipotézis nem-teljesülése esetén nagyobb, mert ekkor nem nulla várható értékű normális eloszlású valószínűségi változók négyzetösszegét kell tekinteni. Ezen észrevételek alapján el lehet készíteni a tesztet. Akkor fogadjuk el a nullhipotézist, ha a $\frac{Q_a}{Q_e}$ hányados nem túl nagy.

A b) és c) esetekben is hasonlóan érvelünk. Ezek pusztán jelzésszerű tárgyalásában nem írom le a könyvben megtalálható kifejezések pontos formáját, megelégszem azzal, hogy utalok azoknak a könyvben megtalálható nevére.

a b) esetben a $\frac{Q_e}{\sigma^2}$ normalizált véletlen hiba mindig χ^2 eloszlású $(k-1)(p-1)$ szabadságfokkal. Az a hatásoknak megfelelő Q_a négyzetösszegre $\frac{Q_a}{\sigma^2}$ a nullhipotézis teljesülése esetén a Q_e valószínűségi változótól független χ^2 eloszlású valószínűségi változó $k-1$ szabadságfokkal. Ez lehetővé teszi a H_{0a}

$$H_{0a}\colon \ a_1 = a_2 = \cdots = a_k = 0$$

nullhipotézis vizsgálatát az a) esethez hasonló módon.

A H_{0b}

$$H_{0b}\colon\ b_1 = b_2 = \cdots = b_p = 0$$

nullhipotézist hasonló módon vizsgáljuk, csak ekkor a b-hatásoknak megfelelő Q_b négyzetösszeg veszi át a Q_a négyzetösszeg szerepét.

Tekintsük a c) esetet. Ekkor vizsgáljuk a H_{0a}, H_{0b} és H_{0c} nullhipotéziseket, amelyeket a

$$H_{0a}\colon\ a_1 = a_2 = \cdots = a_k = 0,$$

$$H_{0b}\colon\ b_1 = b_2 = \cdots = b_p = 0,$$

$$H_{0ab}\colon\ c_{i,j} = 0. \quad i = 1, \ldots, k\ j = 1. \ldots, p,$$

feltevések jelentenek.

Ezek vizsgálatában felhasználjuk, hogy a $Q_e$véletlen hiba $\frac{Q_e}{\sigma^2}$ négyzetösszegének a $\frac{Q_e}{\sigma^2}$ normalizáltja mindig χ^2 eloszlású $kp(n-1)$ szabadságfokkal. A H_{0a} fennállása esetén az a-hatások Q_a négyzetösszegének $\frac{Q_a}{\sigma^2}$ normalizáltja χ^2 eloszlású $k-1$ szabadságfokkal, és független a Q_e valószínűségi változótól. Ezért a $\frac{Q_a}{Q_e}$ hányados viselkedése segítségével ellenőrizhetjük a H_{a0} hipotézist.

Hasonlóan a H_{0b} fennállása esetén a b-hatások Q_b négyzetösszegének $\frac{Q_b}{\sigma^2}$ normalizáltja χ^2 eloszlású $p-1$ szabadságfokkal, és független a Q_e valószínűségi változótól. Ezért a $\frac{Q_b}{Q_e}$ hányados viselkedése segítségével ellenőrizhetjük a H_{b0} hipotézist.

A H_{0ab} fennállása esetén az ab-interakció hatások Q_c négyzetösszegének $\frac{Q_c}{\sigma^2}$ normalizáltja χ^2 eloszlású $(k-1)(p-1)$ szabadságfokkal, és független a Q_e valószínűségi változótól. Ezért a $\frac{Q_c}{Q_e}$ hányados viselkedése segítségével ellenőrizhetjük a H_{ab0} hipotézist.

Valójában, ha a $H_{0,ab}$ nullhipotézist elfogadjuk, akkor a H_{0a} és H_{0b} nullhipotézisek vizsgálatát másképp is elvégezhetjük, és ezt a módszert alkalmazzák a gyakorlatban. Ekkor ugyanis, mivel Q_e és Q_c függetlenek, ezért a $\frac{Q_c+Q_e}{\sigma^2}$ összeg χ^2 eloszlású $knp-k-n+1$ szabadságfokkal. Ha a H_{0a} illetve H_{0b} nullhipotézis is teljesül, akkor ez független a Q_a illetve Q_b négyzetösszegtől is. Ezért ezeket a nullhipotézisket akkor fogadjuk el, ha a $\frac{Q_a}{Q_e+Q_c}$ illetve $\frac{Q_b}{Q_e+Q_c}$ viszonylag kicsi.

A hatodik szekció témája a kovarianciaanalízis. Ezt csak röviden ismertetem. A feladat a következő.

Megfigyelünk egy n elemű $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ véletlen vektort, amely a következő módon keletkezik.

Adott egy $\mathbf{B} = (b_{ij})$, $1 \le i \le n$, $1 \le j \le k$, $n \times k$ méretű struktúramátrix valamilyen $r \le k$ ranggal és egy $\mathbf{D} = (d_{ij})$, $1 \le i \le n$, $1 \le j \le l$, $n \times l$ méretű kisérőmátrix, amelynek rangja l. A $\mathbf{D}$ és $\mathbf{B}$ mátrixok értékét ismerjük. A megfigyelt $\mathbf{Y}$ (véletlen) vektor várható értéke

$$E\mathbf{Y} = \mathbf{Ba} + \mathbf{Dc},$$

valamilyen (ismeretlen) $\mathbf{a} = (a_1, \ldots, a_k)^T$ $\mathbf{c} = (c_1, \ldots, c_l)^T$ vektorokkal. Ezeket akarjuk megbecsülni a legkisebb négyzetek módszerével, azaz a

$$\sum_{i=1}^{n} (Y_i - b_{i1}a_1 - \cdots - b_{ik}a_k - d_{i1}c_1 - \cdots - d_{il}c_l)^2$$

kifejezés minimumának a meghatározásával az $a_1, \ldots, a_k$, $c_1, \ldots, c_l$ paraméterek függvényében. Ez a

$$\mathbf{B}^T\mathbf{Ba} + \mathbf{B}^T\mathbf{Dc} = \mathbf{B}^T\mathbf{Y} \quad (8)$$

$$\mathbf{D}^T\mathbf{Ba} + \mathbf{D}^T\mathbf{Dc} = \mathbf{D}^T\mathbf{Y} \quad (9)$$

egyenletrendszer megoldását jelenti.

Ennek érdekében vezessük be az $\hat{\mathbf{a}}_0$ vektort, mint a $\mathbf{B}^T\mathbf{Ba}_0 = \mathbf{B}^T\mathbf{Y}$ egyenletrendszer és az $\hat{\mathbf{a}}_i$, $i = 1, \ldots, l$, vektorokat, mint a $\mathbf{B}^T\mathbf{Ba}_i = \mathbf{B}^T\mathbf{d}_i$ egyenletrendszer megoldását, ahol $\mathbf{d}_i$ a $\mathbf{D}$ mátrix i-edik oszlopvektora.

A (8) egyenletet beszorozva a $\hat{\mathbf{a}}_i^T$ vektorral (balról), majd kivonva belőle a (9) egyenletrendszer i-edik sorát, és ezt elvégezve minden $i = 1, \ldots, l$ indexre l darab olyan lineáris egyenletet kapunk, amelyekben csak a c_i változók szerepelnek. (Az a_i változók hiányoznak.) Ezek segítségével ki lehet számolni a c_i együtthatókat, majd azok ismeretében az a_i együtthatókat is. Azt használjuk ki ebben a számolásban, hogy a (9) egyenletrendszer i-edik egyenletének első tagja $\mathbf{d}_i^T\mathbf{Ba} = \hat{\mathbf{a}}_i^T\mathbf{B}^T\mathbf{Ba}$. (A teljes bizonyításban még meg kell indokolni, hogy az $\hat{\mathbf{a}}_i$ mennyiségeket definiáló egyenleteknek van megoldásuk akkor is, ha $r = \text{rang}\,(\mathbf{B}) < k$. Feltesszük, hogy $n \ge \max(k, l)$.)

A könyv további képleteket tartalmaz e feladat megoldásáról illetve azokkal kapcsolatos teszt eljárásokról. Véleményem szerint ezek részletesebb magyarázatot igényeltek volna.

A hetedik szekció témája a kanonikus korrelációanalízis. Ez a következő problémával foglalkozik. Legyen $\mathbf{X} = (X_1, \ldots, X_p)$ és $\mathbf{Y} = (Y_1, \ldots, Y_q)$ két (együttesen) normális eloszlású véletlen vektor nulla várható értékkel. Jelölje $\mathbf{C}_{1,1}$ az $\mathbf{X}$ vektor, $\mathbf{C}_{2,2}$ az $\mathbf{Y}$ vektor kovarianciamátrixát, és legyen

$\mathbf{C}_{1,2} = (EX_iY_j)$, $1 \le i \le p$, $1 \le j \le q$, az $\mathbf{X}$ és $\mathbf{Y}$ vektorok közös kovarianciamátrixa. Definiáljuk az $\mathbf{X}$ és $\mathbf{Y}$ vektorok *kanonikus korrelációs együtthatóit*, és állapítsuk meg azok legfontosabb tulajdonságait. Látni fogjuk, hogy a kanonikus korrelációś együtthatók definiciója, illetve azok tulajdonságai szoros kapcsolatban vannak a (téglalap alakú) mátrixok szinguláris felbontásával.

A kanonikus korrelációs együtthatók a $\mathbf{C}_{1,1}$, $\mathbf{C}_{2,2}$ és $\mathbf{C}_{1,2}$ kovarinciamátrixok függvényei. A következő kérdés az, hogy hogyan becsüljük meg (a maximum likelihood módszer segítségével a kanonikus korrelációs együtthatókat, ha a fenti kovarinciamátrixokat nem ismerjük, viszont van egy n-elemű mintánk az ($\mathbf{X}$, $\mathbf{Y}$) vektorpárokból. E kérdés folytatása annak a problémának a vizsgálata, hogy hogyan adjunk egy statisztikai proóbát annak ellenőrzésére, hogy csak a k legnagyobb kanonikus korrelációs együttható nem nulla.

A bizonyítások részleteit nem fogom kidolgozni. Egyébként a könyv is túlságosan tömören magyarázza ezt a témát.

Az első, legnagyobb kanonikus korrelációs együtthatót úgy határozzuk meg, mint a következő szélsőértékfeladat megoldását. Keressünk olyan $\mathbf{a}_1 \in R^p$ és $\mathbf{b_1} \in R^q$ vektorokat, amelyekre

$$\mathrm{Corr}\,(\mathbf{a}_1^T X, \mathbf{b}_1^T \mathbf{Y}) = \max_{(\mathbf{a},\mathbf{b})\in V_1} \mathrm{Corr}\,(\mathbf{a}^T X, \mathbf{b}^T \mathbf{Y}),$$

ahol

$$V_1 = \{(\mathbf{a}, \mathbf{b})\colon\ \mathbf{a} \in R^p,\ \mathbf{b} \in R^q\}.$$

Az $\mathbf{a}_1$ és $\mathbf{b}_1$ vektorok csak egy konstans szorzó erejéig vannak meghatározva. Pontosabban, a tekintett korreláció nem változik, ha e vektorokat (esetleg különböző) pozitív számokkal szorozzuk meg.

A $\mathrm{Corr}\,(\mathbf{a}_1^T X, \mathbf{b}_1^T \mathbf{Y})$ mennyiség az első, legnagyobb, kanonikus korrelációs együttható.

A k-ik kanonikus korrelációs együtthatót, és a hozzá tartozó $(\mathbf{a}_k, \mathbf{b}_k)$ vektorpárt definiáljuk minden $1 \le k \le \min(p, q)$ számra. Ezt indukcióval tesszük a következő módon. Ha definiáltuk a j-ik kanonikus korrelációs együtthatót, és a hozzá tartozó $(\mathbf{a}_j, \mathbf{b}_j)$ vektorpárt minden $1 \le j < k$ indexre, akkor legyen

$$\mathrm{Corr}\,(\mathbf{a}_k^T X, \mathbf{b}_k^T \mathbf{Y}) = \max_{(\mathbf{a},\mathbf{b})\in V_k} \mathrm{Corr}\,(\mathbf{a}^T X, \mathbf{b}^T \mathbf{Y}),$$

ahol

$$\begin{aligned} V_k\ = \{(\mathbf{a}, \mathbf{b})\colon\ & \mathbf{a} \in R^p,\ \mathrm{Corr}\,(\mathbf{aX}, \mathbf{a}_j\mathbf{X}) = 0,\ j = 1, \dots, k-1, \\ & \mathbf{b} \in R^q,\ \mathrm{Corr}\,(\mathbf{bY}, \mathbf{b}_j\mathbf{Y}) = 0,\ j = 1, \dots, k-1\}, \end{aligned}$$

és $(\mathbf{a}_k.\mathbf{b}_k) \in V_k$.

Az $\mathbf{a}_k$ és $\mathbf{b}_k$ vektorok is csak egy konstans szorzó erejéig vannak meghatározva. Az általuk meghatározott korreláció nem változik, ha e vektorokat pozitív számokkal szorozzuk meg. Ezt a tényt felhasználva jogunk van olyan $\mathbf{a}_k$ és $\mathbf{b}_k$ vektorokat tekinteni, amelyekre

$$\mathbf{a_k}^T\mathbf{C}_{1,1}\mathbf{a_k} = 1 \;\text{ és }\; \mathbf{b_k}^T\mathbf{C}_{2,2}\mathbf{b_k} = 1, \quad k = 1, \ldots, \min(p, q).$$

Tekintsünk ilyen $\mathbf{a}_k$ és $\mathbf{b}_k$ vektorokat, és definiáljuk segítségükkel az

$$\alpha_k = \mathbf{C}_{1,1}^{1/2}\mathbf{a_k} \;\text{ és }\; \beta_k = \mathbf{C}_{221}^{1/2}\mathbf{b_k}, \quad k = 1, \ldots, \min(p, q)$$

vektorokat. Az (α_k, β_k) vektorpárok meghatározása ekvivalens az $(\mathbf{a}_k, \mathbf{b}_k)$ vektorpárok meghatározásával. Az (α_k, β_k) párok meghatározása egy olyan szélsőérték feladat megoldását jelenti, amely ekvivalens a

$$\mathbf{D} = \mathbf{C}_{1,1}^{-1/2}\mathbf{C}_{1,2}\mathbf{C}_{22}^{-1/2}$$

mátrix $\mathbf{D} = \mathbf{VSU}^T$ szinguláris felbontásának megadásával. Ugyanis $\|\alpha_k\| = 1$, $\beta_k = 1$, és

$$\alpha_k\mathbf{D}\beta_k = \max_{\substack{\|\alpha\|=\|\beta\|=1 \\ \alpha^T\alpha_j=\beta^T\beta_j=0,\ j=1,\ldots,k-1}} \alpha_k\mathbf{D}\beta_k.$$

Itt a $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_p)$ és $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_q)$ mátrixok szinguláris vektorai közül az első $\min(p, q)$ tag adja az α_k, β_k vektorokat. Az $\mathbf{S}$ mátrixnak csak a diagonálisában vannak nem nulla elemek. Ezek a diagonálisban levő $s_1 \geq s_2 \geq \ldots \geq s_{\min(p,q)}$ számok megegyeznek a megfelelő kanonikus korrelációs együtthatókkal. Ha csak az első l elem nem nulla ezek közül az s_k együtthatók közül, akkor csak az első l kanonikus korrelációs együtthatóról érdemes beszélni.

Van a kanonikus korrelációs együtthatóknak egy másik hasznos jellemzése. Tekintsük az $(X_1, \ldots, X_p, Y_1, \ldots, Y_q)$ $p + q$ dimenziós vektort. Ennek kovarianciamátrixa a

$$\begin{pmatrix} \mathbf{C}_{1,1} & \mathbf{C}_{1,2} \\ \mathbf{C}_{2,1} & \mathbf{C}_{2,2} \end{pmatrix}$$

mátrix, ahol $\mathbf{C}_{2,1} = \mathbf{C}_{1,2}^T$. Megadjuk ennek egyszerűbb, úgynevezett kanonikus alakra hozását a $\mathbf{D} = \mathbf{C}_{1,1}^{-1/2}\mathbf{C}_{1,2}\mathbf{C}_{22}^{-1/2}$ mátrix $\mathbf{D} = \mathbf{VSU}^T$ szinguláris felbontásának segítségével.

Legyen $\mathbf{A} = \mathbf{C}_{1,1}^{-1/2}\mathbf{V}$ és $\mathbf{B} = \mathbf{C}_{2,2}^{-1/2}\mathbf{U}$. Definiáljuk segítségükkel az $\mathbf{A}^T\mathbf{X}$ és $\mathbf{B}^T\mathbf{Y}$ vektorokat. Írjuk fel ezek együttes eloszlásának a kovarianciamátrixát. Ez

$$\begin{pmatrix} \mathbf{A}^T\mathbf{C}_{1,1}\mathbf{A} & \mathbf{A}^T\mathbf{C}_{1,2}\mathbf{B} \\ \mathbf{B}^T\mathbf{C}_{2,1}\mathbf{A} & \mathbf{B}^T\mathbf{C}_{2,2}\mathbf{B} \end{pmatrix} = \begin{pmatrix} \mathbf{V}^T\mathbf{V} & \mathbf{V}^T\mathbf{D}\mathbf{U} \\ \mathbf{U}^T\mathbf{D}^T\mathbf{V} & \mathbf{U}^T\mathbf{U} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_p & \mathbf{S} \\ \mathbf{S}^T & \mathbf{I}_q \end{pmatrix}.$$

Vegyük észre, hogy nemcsak az $\mathbf{I}_p$ és $\mathbf{I}_q$ mátrixok diagonálisak, hanem az $\mathbf{S}$ mátrix is. Ennek diagonális elemei a kanonikus korrelációs együtthatók. Ezek mérik, hogy milyen erős a kapcsolat az $\mathbf{X}$ és $\mathbf{Y}$ vektorok között.

A következő probléma az, hogy hogyan becsüljük meg a kanonikus korrelációs együtthatókat, ha nem ismerjük a $\mathbf{C}_{1,1}$, $\mathbf{C}_{2,2}$ és $\mathbf{C}_{1,2}$ kovarianciamátrixokat, viszont van egy n-elemű $(\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_n, \mathbf{Y}_n)$ mintánk az $(\mathbf{X}, \mathbf{Y})$ vektorra. A természetes hozzáállás az, hogy megbecsüljük a minta segítségével a $\mathbf{C}_{1,1}$, $\mathbf{C}_{2,2}$ és $\mathbf{C}_{1,2}$ kovarianciamátrixokat, és ezekkel a becslésekkel számolunk.

A könyv jelöléséhez alkalmazkodva legyen $\mathbf{F} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)^T$, és $\mathbf{G} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)^T$. Ekkor az $\mathbf{C}_{1,1}$, $\mathbf{C}_{2,2}$ és $\mathbf{C}_{1,2}$ mátrixok maximum likelihood becslése

$$\hat{\mathbf{C}}_{1,1} = \frac{1}{n}\mathbf{F}^T\mathbf{F}, \quad \hat{\mathbf{C}}_{2,2} = \frac{1}{n}\mathbf{G}^T\mathbf{G}, \quad \hat{\mathbf{C}}_{1,2} = \frac{1}{n}\mathbf{F}^T\mathbf{G}.$$

Ennek segítségével meg tudjuk határozni a $\mathbf{D} = \mathbf{C}_{1,1}^{-1/2}\mathbf{C}_{1,2}\mathbf{C}_{22}^{-1/2}$ mátrix $\hat{\mathbf{D}} = \hat{\mathbf{C}}_{1,1}^{-1/2}\hat{\mathbf{C}}_{1,2}\hat{\mathbf{C}}_{22}^{-1/2}$ maximum likelihood becslését. Készítsük el a $\mathbf{D} = \mathbf{V}\mathbf{S}\mathbf{U}^T$ szinguláris felbontáshoz hasonlóan a $\hat{\mathbf{D}} = \hat{\mathbf{V}}\mathbf{R}\hat{\mathbf{U}}^T$ szinguláris felbontást, ahol $\mathbf{R}$ felel meg az $\mathbf{S}$ diagonális mátrixnak. Feleljen meg az $\mathbf{S}$ mátrix s_k elemének a $\mathbf{R}$ mátrix r_k eleme.

Ezen kifejezések segítségével el tudjuk készíteni a likelihood hányados próbát arra az állításra, hogy összesen $k \leq \min(p, q)$ nem zéró kanonikus korrelációs együttható van. Ez ekvivalens azzal az állítással, hogy $\mathbf{S}$ diagonális mátrix diagonálisában $s_k \neq 0$, és $s_j = 0$, ha $j > k$.

E feladat megoldásának érdekében először felírjuk az

$$((\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_n, \mathbf{Y}_n))$$

minta sűrűségfüggvényét. Ezt meg tudjuk tenni, felhasználva, hogy meg tudjuk adni az $\mathbf{Y}_i$ véletlen vektor feltételes eloszlását a $\mathbf{X}_i = \mathbf{x}_i$ feltétel mellett. Ez a feltételes eloszlás a normális eloszlás $\mathbf{Q}\mathbf{x}_i$ várható értékkel, és $\mathbf{C}_{2,2,1}$ kovarianciamátrixszal, ahol

$$\mathbf{Q} = \mathbf{C}_{2,1}\mathbf{C}_{1,1}^{-1}, \quad \mathbf{C}_{2,2,1} = \mathbf{C}_{2,2} - \mathbf{C}_{1,2}^T\mathbf{C}_{1,1}^{-1}\mathbf{C}_{1,2}.$$

(Lásd például a könyv 5.1.7 Állítását.)

Érdemes lenne az

$$((\mathbf{A}^T\mathbf{X}_1, \mathbf{B}^T\mathbf{Y}_1), \ldots, (\mathbf{A}^T\mathbf{X}_n, \mathbf{B}^T\mathbf{Y}_n))$$

minta sűrűségfüggvényét is megadni, de a könyv ezt nem teszi. Viszont azt állítja, hogy e sűrűségfüggvény maximuma akkor, ha $s_j \neq 0$, $j \leq k$-ra, és $s_j = 0$ $j > k$-ra megegyezik az általa megadott képlettel. Ezt érdemes lett volna részletesebben megindokolni. Ha ezt az eredményt elfogadjuk, akkor nem nehéz látni, hogy a maximum likelihood hányados próba a kanonikus korrelációfüggvény viselkedéséről szóló állításra megegyezik a könyvben leírt eljárással.