Vektor értékű normális eloszlású valószínűségi változók és Wishart mátrixok.

E jegyzet témája a Bolla–Krámli könyv ötödik fejezete, amely a többváltozós normális eloszlásokkal kapcsolatos statisztikai problémákkal foglalkozik. E problémák szorosan kapcsolódnak az úgynevezett Wishart mátrixokhoz. A könyvben kidolgozott bizonyítások sok technikai részletét elhagyom, elsősorban arra koncentrálok, hogy a bizonyítások hátterét, és azokat a részleteit magyarázzam el, amelyek részletesebb kifejtést érdemelnek.

Először felidézek néhány olyan eredményt a vektor értékű normális eloszlásokról és a lineáris algebrából, amelyeket használni fogunk.

Egy normális eloszlású véletlen vektor eloszlását meghatározza várható érték vektora és kovarianciamátrixa. A kovarianciamátrix mindig pozitív szemidefinit, és akkor és csak akkor van egy normális eloszlású véletlen vektornak sűrűségfüggvénye, ha a kovarianciamátrixa szigorúan pozitív definit, vagy ami ezzel ekvivalens, invertálható. Ebben a jegyzetben csak invertálható kovarianciamátrixú normális eloszlású véletlen vektorokkal fogunk foglalkozni. Emlékeztetek arra is, hogy ha egy normális eloszlású véletlen vektor koordinátái korrelálatlanok, akkor függetlenek is. (Ennél valamivel élesebb állítás is igaz, amely a koordináták részhalmazai által indexelt vektorokra állítja ugyanezt.) Ugyancsak felidézem azt a tényt, hogy ha egy normális eloszlású véletlen vektort megszorzunk egy mátrixszal, akkor megint normális eloszlású véletlen vektort kapunk. Ez akkor is igaz, ha téglalap alakú mátrixszal szorzunk. Egyébként a könyvhöz hasonlóan $\mathcal{N}_p(\mathbf{m},\mathbf{C})$-vel fogjuk jelölni a p dimenziós normális eloszlást $\mathbf{m}$ várható érték vektorral és $\mathbf{C}$ kovarianciamátrixszal.

Alkalmazkodva a könyv jelöléséhez ebben a jegyzetben oszlopvektorokkal fogok dolgozni. A mátrixokkal kapcsolatban először azt az eredményt idézem fel, hogy egy $\mathbf{A}$ szimmetrikus mátrixnak van úgynevezett spektrál felbontása, azaz $\mathbf{A}=\mathbf{U}^T\mathbf{\Lambda}\mathbf{U}$ alakú előállítása, ahol $\mathbf{\Lambda}$ diagonális $\mathbf{U}$ pedig ortogonális mátrix. A $\mathbf{\Lambda}$ diagonális mátrixban az $\mathbf{A}$ mátrix sajátértékei, az $\mathbf{U}$ ortogonális mátrixban pedig az $\mathbf{A}$ mátrix hozzájuk tartozó 1 normájú sajátvektorai vannak.

Felmerül a kérdés, hogy mennyire egyértelmű szimmetrikus mátrixok spektrálfelbontása. A $\mathbf{\Lambda}$ mátrix egyértelműen meghatározott, eltekintve a mátrix diagonálisában levő λ_j sajátértékek sorrendjétől. Ezt rögzíthetjük például úgy, hogy a sajátértékeket monoton csökkenő sorrendben soroljuk fel a diagonálisban. Ha a sajátértékek között vannak megegyezőek, akkor az $\mathbf{U}$ ortogonális mátrix megadása nem egyértelmű. Ha az összes sajátérték különböző, akkor az $\mathbf{U}$ mátrix előállítása lényegében egyértelmű. Annyi szabadságunk van, hogy a benne szereplő sajátvektorokat beszorozhatjuk ± 1-gyel.

Fogunk dolgozni mátrixok nyomával (trace). Egy négyzetes, nem feltétlenül szimmetrikus $\mathbf{A}$ mátrixnak definiáljuk a nyomát, amit úgy jelölünk, hogy $\operatorname{tr} \mathbf{A}$. Ez egyenlő a mátrix diagonálisában levő elemek összegével. Egy mátrixok nyomáról szóló fontos azonosság az, hogy mátrixok szorzatának a nyoma nem változik, ha a a mátrixokat ciklikusan eltoljuk, azaz ha adva van n darab $\mathbf{A}_1, \dots \mathbf{A}_n$ mátrix, akkor

$$\operatorname{tr}(\mathbf{A}_1 \mathbf{A}_2 \cdots \mathbf{A}_n) = \operatorname{tr}(\mathbf{A}_k \mathbf{A}_{k+1} \cdots \mathbf{A}_n \mathbf{A}_1 \cdots \mathbf{A}_{k-1})$$

minden $k = 1, \dots, n$ indexre. Ebben az azonosságban téglalap alakú mátrixok is szerepelhetnek, az egyetlen megkötés az, hogy a mátrixok mérete olyan, hogy a szorzatuknak van értelme, és az összes mátrix szorzata négyzet alakú mátrix. Tehát az $\mathbf{A}_j$ mátrix mérete $N_j \times N_{j+1}$, $1 \le j \le n$, és $N_{n+1} = N_1$.

A fenti azonosság egyik következménye az, hogy szimmetrikus mátrix nyoma megegyezik a sajátértékei összegével, mert véve egy $\mathbf{A}$ szimmetrikus mátrix spektrálfelbontását azt kapjuk, hogy

$$\operatorname{tr} \mathbf{A} = \operatorname{tr}(\mathbf{U}^T \mathbf{\Lambda} \mathbf{U}) = \operatorname{tr}(\mathbf{U}\mathbf{U}^T \mathbf{\Lambda}) = \operatorname{tr} \mathbf{\Lambda}.$$

Felidézek még néhány eredményt (négyzetes) mátrixok determinánsáról is. A könyv jelölését követve egy $\mathbf{A}$ mátrix determinánsát úgy fogom jelölni, hogy $|\mathbf{A}|$. Egy fontos azonosság az, hogy $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$. Ezenkívül $|\mathbf{I}| = 1$, ahol $\mathbf{I}$ az identitás mátrix, és $|\mathbf{A}^T| = |\mathbf{A}|$. Innen $|\mathbf{U}| = \pm 1$ egy ortogonális mátrixra, mert $|\mathbf{U}|^2 = |\mathbf{U}\mathbf{U}^T| = |\mathbf{I}| = 1$. Egy $\mathbf{A}$ szimmetrikus mátrixra $\mathbf{A} = \mathbf{U}^T \mathbf{\Lambda} \mathbf{U}$ spektrálfelbontással $|\mathbf{A}| = |\mathbf{\Lambda}| = \lambda_1 \cdots \lambda_n$, ahol λ_j, $1 \le j \le n$, az $\mathbf{A}$ mátrix sajátértékei.

Legyen adva egy $\mathbf{A} = (\mathbf{a}_{i,j})$, $1 \le i, j \le n$, $n \times n$ méretű mátrix. Vezessük be a könyv jelöléséhez hasonlóan az $\mathbf{A}$ mátrix $\operatorname{adj}(\mathbf{A})$ adjungáltját, amit úgy definiálunk, hogy $\operatorname{adj}(\mathbf{A}) = ((-1)^{i+j} |\mathbf{A}_{j,i}|)$, $1 \le i, j \le n$, ahol $|\mathbf{A}_{i,j}|$ annak az $(n-1) \times (n-1)$ méretű mátrixnak a determinánsa, amelyet úgy kapunk, hogy az $\mathbf{A}$ mátrixnak elhagyjuk az i-edik sorát és j-edik oszlopát. Ismeretes, hogy

$$|\mathbf{A}| = \sum_{j=1}^{n} a_{i,j} (-1)^{(i+j)} |\mathbf{A}_{i,j}| \quad \text{minden } i = 1, \dots, n \text{ indexre.} \tag{1}$$

Ezenkívül

$$0 = \sum_{j=1}^{n} a_{i',j}(-1)^{(i+j)}|\mathbf{A}_{i',j}| \quad \text{ha } i \neq i'.$$

Innen nem nehéz belátni, hogy

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\text{adj}\,(\mathbf{A}).$$

Tekintsük az $|\mathbf{A}|$ determinánst, mint n^2 változós függvényt $a_{i,j}$, $1 \le i, j \le n$, argumentumokkal. Szükségünk lesz a következő eredményre is.

$$\frac{\partial|\mathbf{A}|}{\partial a_{i,j}} = (-1)^{(i+j)}|\mathbf{A}_{i,j}|.$$

Ez az azonosság egyszerűen következik az (1) formulából.

Használni fogjuk azt a tényt is, hogy ha $\mathbf{C}$ pozitív szemidefinit mátrix, akkor létezik olyan (egyértelműen meghatározott) pozitív szemidefinit $\mathbf{C}^{1/2}$ mátrix, amelyre $(\mathbf{C}^{1/2})^2 = \mathbf{C}$. Adva egy $\mathbf{X}$ $\mathcal{N}_p(\mathbf{m}, \mathbf{C})$ eloszlású p-dimenziós véletlen vektor invertálható $\mathbf{C}$ mátrixszal, definiáljuk az $\mathbf{Y} = (\mathbf{C}^{1/2})^{-1}(\mathbf{X} - \mathbf{m})$ véletlen vektort. Ez az $\mathbf{Y}$ vektor $\mathcal{N}_p(\mathbf{0}, \mathbf{I})$, azaz standard normális eloszlású. Ezt felhasználva nem nehéz belátni a az 5. fejezet alább megfogalmazott 1.3 Tételét.

1.3 Tétel. *Ha* $\mathbf{X}$ $\mathcal{N}_p(\mathbf{m}, \mathbf{C})$ *eloszlású véletlen vektor, és* $\mathbf{C}$ *invertálható mátrix, akkor az* $(\mathbf{X}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{X}-\mathbf{m})$ *valószínűségi változó eloszlása* $\chi^2(p)$.

Meg akarjuk találni az egydimenziós normális eloszlások statisztikai vizsgálatában használt eredmények többdimenziós megfelelőit. Az egyik fontos egydimenziós eredmény a következő. Legyenek $X_1, \dots, X_n$ független normális eloszlású valószínűségi változók m várható értékkel és σ^2 szórásnégyzettel. Ekkor az $\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$ átlag és $\frac{S}{n-1} = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X})^2$ empirikus szórásnégyzet független valószínűségi változók, $\frac{\sqrt{n}}{\sigma}(\bar{X} - m)$ standard normális, $\frac{S}{\sigma^2}$ pedig $\chi^2(n-1)$ eloszlású valószínűségi változó.

Ezen eredmény többdimenziós megfelelőjében független $\mathcal{N}_p(\mathbf{m}, \mathbf{C})$ eloszlású $\mathbf{X}_1, \dots, \mathbf{X}_n$ véletlen vektorokat tekintünk. Segítségükkel definiáljuk az

$$\bar{\mathbf{X}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{X}_j \tag{2}$$

véletlen vektort és

$$\mathbf{S} = \sum_{j=1}^{n} (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^T \tag{3}$$

véletlen mátrixot, és meg akarjuk határozni ezek együttes eloszlását. Az együttes eloszlás megadásában megjelennek a Wishart mátrixok. Ezek eloszlásai veszik át a χ^2 eloszlások szerepét a többváltozós eredményekben.

Wishart mátrix definiciója. *A $p \times p$ méretű véletlen mátrixot p dimenziós n szabadságfokú Wishart mátrixnak nevezünk $\mathbf{C}$ (mátrix)paraméterrel, ha az előáll $\mathbf{W} = \mathbf{X}\mathbf{X}^T$ alakban, ahol $\mathbf{X}$ olyan $p \times n$ méretű véletlen mátrix, amelynek oszlopai független $\mathcal{N}_p(\mathbf{0}, \mathbf{C})$ véletlen vektorok. E $\mathbf{W}$ vektor eloszlását p-dimenziós n szabadságfokú Wishart eloszlásnak hívjuk $\mathbf{C}$ paraméterrel, és úgy jelöljük, hogy $\mathcal{W}_p(n, \mathbf{C})$. A $\mathcal{W}_p(n, \mathbf{I}_p)$ eloszlást, ahol $\mathbf{I}_p$ a p-dimenziós identitás mátrix p-dimenziós n szabadságfokú standard Wishart eloszlásnak nevezzük.*

Megjegyzés. Ha egy Wishart mátrix $\mathbf{W} = \mathbf{X}\mathbf{X}^T$ alakban van megadva, ahol $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$, és $\mathbf{X}_1, \dots, \mathbf{X}_n$ független $\mathcal{N}_p(\mathbf{0}, \mathbf{C})$ eloszlású véletlen vektorok, akkor $\mathbf{W}$ felírható $\mathbf{W} = \sum_{j=1}^n \mathbf{X}_j \mathbf{X}_j^T$ alakban is. E képlet alapján tekinthetjük a Wishart eloszlásokat a $\chi^2(\cdot)$ eloszlások természetes többdimenziós általánosításainak.

Nem nehéz az 1.3 Tételhez hasonlóan belátni a következő a könyv 5. fejezetének 4.1 tételében megfogalmazott eredményt.

4.1 Tétel. *Legyen a $\mathbf{C}$ mátrix szigorúan pozitív definit. Egy $\mathbf{W}$ véletlen mátrix eloszlása akkor és csak akkor $\mathcal{W}_p(\mathbf{0}, \mathbf{C})$, ha $\mathbf{C}^{-1/2}\mathbf{W}\mathbf{C}^{-1/2}$ eloszlása $\mathcal{W}_p(\mathbf{0}, \mathbf{I}_p)$.*

Az alábbi a könyv 5. fejezet 4.3 tétele tartalmazza az előbb említett egy dimenziós állítás többdimenziós megfelelőjét.

4.3 Tétel. *Legyenek $\mathbf{X}_1, \dots .\mathbf{X}_n$ független véletlen vektorok $\mathcal{N}_p(\mathbf{m}, \mathbf{C})$ eloszlással, és definiáljuk segítségükkel a (2) képletben szereplő $\bar{\mathbf{X}}$ véletlen vektort és a (3) képletben szereplő $\mathbf{S}$ véletlen mátrixot. Az alábbi állítások érvényesek.*

(1) $\bar{\mathbf{X}}$ $\mathcal{N}_p\left(\mathbf{m}, \frac{1}{n}\mathbf{C}\right)$ *eloszlású véletlen vektor.*

(2) $\mathbf{S}$ $\mathcal{W}_p(n-1, \mathbf{C})$ *eloszlású véletlen mátrix.*

(3) $\bar{\mathbf{X}}$ *és* $\mathbf{S}$ *függetlenek egymástól.*

A 4.3 tétel bizonyítása hasonló az egydimenziós esethez. Egy olyan $n \times n$ méretű ortogonális $\mathbf{Q}$ mátrixot veszünk amelynek utolsó sora az $\left(\frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}}\right)$ vektor, és tekintjük az $(\mathbf{Y}_1, \dots, \mathbf{Y}_n)^T = \mathbf{Q}(\mathbf{X}_1, \dots, \mathbf{X}_n)^T$ véletlen mátrixot. E véletlen mátrix $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ oszlopvektorainak a tulajdonságait vizsgáljuk. Először meghatározzuk e vektorok eloszlásait.

Némi számolás mutatja, hogy az $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ vektorok független normális eloszlású véletlen vektorok $\mathbf{C}$ kovarianciamátrixszal, a várható értékükre pedig teljesül az $E\mathbf{Y}_j = \mathbf{0}$, ha $j \neq n$, és $E\mathbf{Y}_n = \sqrt{n}\mathbf{m}$ azonosság. Ezen állítások igazolásában felhasználjuk, hogy a normális eloszlású vektorok tulajdonságai miatt elég ellenőrízni a korrelálatlanságot a függetlenség bizonyításához. Azokat a könybben megtalálható számolásokat kell elvégezni, amelyek szerint $E(\mathbf{Y}_i - E\mathbf{Y}_i)(\mathbf{Y}_i - E\mathbf{Y}_i)^T = \mathbf{C}$, $E(\mathbf{Y}_i - E\mathbf{Y}_i)(\mathbf{Y}_i - E\mathbf{Y}_i)^T = \mathbf{0}$, ha $i \neq j$, $E\mathbf{Y}_i = \mathbf{0}$, ha $1 \leq i \leq n-1$, $E\mathbf{Y}_n = \sqrt{n}\mathbf{m}$.

Továbbá, $\mathbf{Y}_n = \sqrt{n}\bar{\mathbf{X}}$, és $\mathbf{S} = \sum_{j=1}^{n-1} \mathbf{Y}_j \mathbf{Y}_j^T$. Az utóbbi állítás azért igaz, mert

$$\begin{aligned}
\sum_{j=1}^{n} \mathbf{Y}_j \mathbf{Y}_j^T &= (\mathbf{Y}_1, \dots, \mathbf{Y}_n)(\mathbf{Y}_1, \dots, \mathbf{Y}_n)^T \\
&= (\mathbf{X}_1, \dots, \mathbf{X}_n)\mathbf{Q}^T\mathbf{Q}(\mathbf{X}_1, \dots, \mathbf{X}_n)^T \\
&= (\mathbf{X}_1, \dots, \mathbf{X}_n)(\mathbf{X}_1, \dots, \mathbf{X}_n)^T = \sum_{j=1}^{n} \mathbf{X}_j \mathbf{X}_j^T.
\end{aligned}$$

Innen

$$\begin{aligned}
\sum_{j=1}^{n-1} \mathbf{Y}_j \mathbf{Y}_j^T &= \sum_{j=1}^{n} \mathbf{X}^j X_j^T - \mathbf{Y}_n \mathbf{Y}_n^T = \sum_{j=1}^{n} X^j \mathbf{X}_j^T - n\bar{\mathbf{X}}\bar{\mathbf{X}}^T \\
&= \sum_{j=1}^{n} (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j^T - \bar{\mathbf{X}}^T) = \mathbf{S}.
\end{aligned}$$

A fenti relációk segítségével nem nehéz belátni a 4.3 tételt.

Ezután a könyv a várható érték és a kovarianciamátrix maximum likelihood becslését határozza meg. Azt a feladatot vizsgálja, hogy adva egy $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ minta $\mathcal{N}_p(\mathbf{m}, \mathbf{C})$ eloszlással hogyan számoljuk ki az $\mathbf{m}$ várható érték vektor és $\mathbf{C}$ kovarianciamátrix maximum likelihood becslését, ha mindkettő ismeretlen. A következő eredményt bizonyítja be.

5.1 Tétel. *Legyenek $\mathbf{X}_1, \dots, \mathbf{X}_n$ független $\mathcal{N}_p(\mathbf{m}, \mathbf{C})$ eloszlású véletlen vektorok. Tegyük fel, hogy a $\mathbf{C}$ mátrix szigorúan pozitív definit, és $n > p$. Az* *ismeretlen* $\mathbf{m}$ *várható érték vektor és* $\mathbf{C}$ *kovarianciamátrix maximum likelihood becslése e minta alapján*

$$\hat{\mathbf{m}} = \bar{\mathbf{X}}, \quad \textit{és} \quad \hat{\mathbf{C}} = \frac{1}{n}\mathbf{S},$$

ahol $\bar{\mathbf{X}}$ *a (2) és* $\mathbf{S}$ *a (3) képletben van definiálva.*

E tétel bizonyításának csak a vázlatát ismertetem.

Első lépésben felírjuk a $\mathcal{N}_p(\mathbf{m}, \mathbf{C})$ eloszlású $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ minta

$$f_{\mathbf{m},\mathbf{C}}(\mathbf{X}_1, \dots, \mathbf{X}_n), \quad \mathbf{X}_j \in R^p, \ \ 1 \le j \le n,$$

sűrűségfüggvényét számunkra alkalmasabb alakban. A sűrűségfüggvény eredeti alakja

$$f_{\mathbf{m},\mathbf{C}}(\mathbf{X}_1, \dots, \mathbf{X}_n) = \frac{1}{(2\pi)^{np/2}|\mathbf{C}|^{n/2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{X}_j - \mathbf{m})^T\mathbf{C}^{-1}(\mathbf{X}_j - \mathbf{m})\right\} \tag{4}$$

Belátjuk, hogy ez

$$\begin{aligned} f_{\mathbf{m},\mathbf{C}}(\mathbf{X}_1, \dots, \mathbf{X}_n) = {} & \frac{1}{(2\pi)^{np/2}|\mathbf{C}|^{n/2}} \\ & \exp\left\{-\frac{1}{2}\mathrm{tr}\,(\mathbf{C}^{-1}\mathbf{S})\right\} \exp\left\{-\frac{1}{2}n(\bar{\mathbf{X}} - \mathbf{m})^T\mathbf{C}^{-1}(\bar{\mathbf{X}} - \mathbf{m})\right\} \end{aligned} \tag{5}$$

alakban írható. Ez következik az alábbi számolásból.

$$\begin{aligned} & \sum_{j=1}^{n}(\mathbf{X}_j - \mathbf{m})^T\mathbf{C}^{-1}(\mathbf{X}_j - \mathbf{m}) \\ & \quad = \sum_{j=1}^{n}(\mathbf{X}_j - \bar{\mathbf{X}})^T\mathbf{C}^{-1}(\mathbf{X}_j - \bar{\mathbf{X}}) + n(\bar{\mathbf{X}} - \mathbf{m})^T\mathbf{C}^{-1}(\bar{\mathbf{X}} - \mathbf{m}) \\ & \quad = \sum_{j=1}^{n}\mathrm{tr}\,(\mathbf{X}_j - \bar{\mathbf{X}})^T\mathbf{C}^{-1}(\mathbf{X}_j - \bar{\mathbf{X}}) + n(\bar{\mathbf{X}} - \mathbf{m})^T\mathbf{C}^{-1}(\bar{\mathbf{X}} - \mathbf{m}) \\ & \quad = \sum_{j=1}^{n}\mathrm{tr}\,\mathbf{C}^{-1}(\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^T + n(\bar{\mathbf{X}} - \mathbf{m})^T\mathbf{C}^{-1}(\bar{\mathbf{X}} - \mathbf{m}) \\ & \quad = \mathrm{tr}\,\mathbf{C}^{-1}\left(\sum_{j=1}^{n}(\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^T\right) + n(\bar{\mathbf{X}} - \mathbf{m})^T\mathbf{C}^{-1}(\bar{\mathbf{X}} - \mathbf{m}) \\ & \quad = \mathrm{tr}\,\mathbf{C}^{-1}\mathbf{S} + n(\bar{\mathbf{X}} - \mathbf{m})^T\mathbf{C}^{-1}(\bar{\mathbf{X}} - \mathbf{m}). \end{aligned}$$

Megjegyzem, hogy a sűrűségfüggvényt ebben a képletben a $\bar{\mathbf{X}}$ véletlen vektor és $\mathbf{S}$ véletlen mátrix által alkotott elégséges statisztika segítségével fejeztük ki.

A (5) formula segítségével nem nehéz belátni, hogy az $\mathbf{m}$ várható érték maximum likelihood becslése a $\hat{\mathbf{m}} = \bar{\mathbf{X}}$ kifejezés. Ez nem függ a $\mathbf{C}$ mátrix értékétől. Ezt a kifejezést behelyettesítjük a sűrűségfüggvénybe, és a maximum likelihood becslésben ennek a maximumát keressük a $\mathbf{C}^{-1}$ mátrix koordinátáinak a függvényében.

Jelölje $c_{i,j}$ a $\mathbf{C}$, és $c^{i,j}$ az inverz $\mathbf{C}^{-1}$ mátrix i-ik sorának j-ik elemét, és legyen $\mathbf{S} = (s_{i,j})$, $1 \le i, j \le p$, a (3) formulában definiált mátrix. A $\mathbf{C}$ mátrix maximum likelihood becslésében azzal a sűrűségfüggvénnyel kell dolgoznunk, amelyet úgy kapunk, hogy az $f_{\mathbf{m},\mathbf{C}}$ sűrűségfüggvényében elvégezzük az $\mathbf{m} = \hat{\mathbf{m}}$ helyettesítést. Ez a következő alakú.

$$f_{\hat{\mathbf{m}},\mathbf{C}}(\mathbf{X}_1, \dots, \mathbf{X}_n) = \frac{1}{(2\pi)^{np/2}|\mathbf{C}|^{n/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\,(\mathbf{C}^{-1}\mathbf{S})\right\}.$$

Felírva e sűrűségfüggvény logaritmusát, és azt deriválva a $c^{i,j}$ argumentumok szerint azt kapjuk, hogy a maximum likelihood egyenlet az $c^{i,j}$ változókra a következő.

$$n\frac{\partial \log |\mathbf{C}^{-1}|}{\partial c^{i,j}} = s_{i,j}, \quad 1 \le i, j \le p.$$

Másrészt,

$$\frac{\partial \log |\mathbf{C}^{-1}|}{\partial c^{i,j}} = \frac{1}{|\mathbf{C}^{-1}|}(-1)^{(i+j)}|\mathbf{C}^{-1}_{i,j}|,$$

ahol $\mathbf{C}^{-1}_{i,j}$ az a mátrix, amelyet úgy kapunk, hogy tekintük a $\mathbf{C}^{-1}$ mátrixot, és elhagyjuk annak az i-ik sorát és j-ik oszlopát. Továbbá, felhasználva, hogy $\mathbf{C}$ a $\mathbf{C}^{-1}$ mátrix inverze ki tudjuk fejezni a $\mathbf{C}$ mátrix $c_{j,i}$ elemét az inverziós formula segítségével azt kapjuk, hogy

$$c_{j,i} = \frac{1}{|\mathbf{C}^{-1}|}(-1)^{(i+j)}|\mathbf{C}^{-1}_{i,j}| = \frac{\partial \log |\mathbf{C}^{-1}|}{\partial c^{i,j}}.$$

Innen a maximum likelihood egyenlet $nc_{j,i} = s_{i,j}$, ahonnan a becslőfüggvény $\hat{c}_{j,i} = \frac{s_{i,j}}{n} = \frac{s_{j,i}}{n}$. Innen következik a tétel állítása. Megjegyzem, hogy a becsülendő paraméterek a $\mathbf{C}^{-1}$ inverz mátrix $c^{i,j}$ koordinátái voltak, de ezzel ekvivalens $\mathbf{C}$ mátrix $c_{i,j}$ koordinátáinak a maximum likelihood becslése.

Érdemes néhány megjegyzést tenni az előző számolásokkal kapcsolatban. A következő feladatot oldottuk meg. Tekintsük a $p \times p$ méretű pozitív determinánsú $\mathbf{C}$ mátrixokat, és ezek között határozzuk meg a $\mathbf{C}$ mátrix egy megadott függvényének a maximumát. Ezt, mint egy p^2 változós függvény maximumát találtuk meg parciális deriválások segítségével. A változók paraméterei az (i,j), $1 \le i,j \le p$, párok. Kiderült, hogy a megoldás speciális tulajdonságokkal rendelkezik. Az egy szimmetrikus (és pozitív definit) mátrix, amelynek koordinátáira $\hat{c}_{i,j} = \hat{c}_{j,i}$. Ilyen módon megtaláltuk a megoldást a minket érdeklő feladatra is, amelyikben az optimumot a szimmetrikus (és pozitív definit) mátrixok terében keressük, ami egy $\frac{p(p+1)}{2}$ dimenziós tér. Ha a szimmetrikus mátrixok terében dolgoznánk, akkor kissé másképpen kellene számolnunk. (Például a determináns deriválásáról szóló képlet megváltozik.)

A könyv 5. fejezetének a következő két tételét csak kimondom, mert a bizonyítás meglehetősen nehéz, és az eredmények nem tartoznak szorosan a könyv témaköréhez. Ezek az eredmények a Wishart mátrix sűrűségfüggvényéről szólnak, illetve a Wishart mátrix sajátértékeinek az eloszlásáról. Megjegyzem, hogy Lalley

https://galton.uchicago.edu/ lalley/Courses/386/ClassicalEnsembles.pdf

interneten elérhető jegyzetében e téma részletesebb kifejtése található, és hivatkozni fogok e jegyzet bizonyos eredményeire.

5.2 Tétel. *Legyenek $\mathbf{X}_1, \dots, \mathbf{X}_n$ $\mathcal{N}_p(\mathbf{0}, \mathbf{I})$, azaz p dimenziós standard normális eloszlású véletlen vektorok, és definiáljuk az $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ $p \times n$ méretű véletlen mátrixot és $\mathbf{W} = \mathbf{X}\mathbf{X}^T$ p-dimenziós, n szabadságfokú standard Wishart eloszlású véletlen mátrixot. Ennek sűrűségfüggvénye*

$$c_{n,p}|\mathbf{W}|^{(n-p-1)/2} \exp\left\{-\frac{tr\,\mathbf{W}}{2}\right\},$$

ha $n \ge p$, ahol $c_{n,p}$ egy csak az n és p paraméterektől függő normalizáló konstans.

Megjegyzés. A Lalley jegyzet kissé élesebb eredményét fogalmaztam meg. Ez az $n \ge p$ és nem az $n > p$ feltétel mellett mondta ki ezt az állítást. Az $n \ge p$ feltétel szükséges, mert különben p-dimenziós, $n < p$ vagy kisebb rangú, tehát szinguláris mátrixokra lenne koncentrálva a standard Wishart mérték.

Kiegészítés az 5.2 tétel jelentéséről. Az 5.2 Tételben szereplő p-dimenziós, n szabadságfokú standard Wishart eloszlású mátrixok úgy tekintendőek, mint olyan valószínűségi változók, amelyek értékeiket a $p \times p$ méretű szimmetrikus mátrixok terén veszik fel. Egy $\mathbf{A} = (a_{i,j})$, $1 \le i,j \le p$, $p \times p$ méretű szimmetrikus mátrix értékét meghatározzák az $a_{i,j}$, $1 \le i \le j \le p$, koordinátái. Ennek alapján azt mondjuk, hogy egy véletlen $\mathbf{A} = (a_{(i,j)})$, $1 \le i,j \le p$, szimmetrikus $p \times p$ méretű mátrixnak van $g(x_{(i,j)}$, $1 \le i \le j \le p)$, sűrűségfüggvénye, ha az $a_{(i,j)}$, $1 \le i \le j \le p$ koordinátáknak van $g(x_{(i,j)}$, $1 \le i \le j \le p)$, együttes sűrűségfüggvénye a Lebesgue mérték szerint a $p(p+1)/2$ dimenziós Euklideszi térben, amelynek koordinátáit az (i,j), $1 \le i \le j \le p$, számpárokkal indexeljük.

A következő 5.4 tétel megadja egy standard Wishart mátrix sajátértékeinek az eloszlását.

5.4 Tétel. *Legyen* $\mathbf{W}$ *egy* p *dimenziós* n *szabadságfokú standard Wishart mátrix,* $n \ge p$. *E véletlen mátrix* $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$ *sajátértékeinek az eloszlása*

$$\kappa_{n,p} \left(\prod_{j=1}^{p} \lambda_j \right)^{(n-p-1)/2} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{p} \lambda_j \right\} \prod_{1 \le j < k \le p} (\lambda_j - \lambda_k),$$

ahol $\kappa_{n,p}$ *alkalmas csak az* n *és* p *paraméterektől függő normáló konstans.*

Megjegyzem, hogy az 5.2 tétel alapján egy p dimenziós n szabadságfokú, $n \ge p$, $\mathbf{W}$ standard Wishart mátrix sűrűségfüggvényét a következő módon lehet felírni a $\mathbf{W}$ mátrix $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$ sajátértékeinek a segítségével.

$$g(\lambda_1, \dots, \lambda_p) = c_{n,p} \left(\prod_{j=1}^{p} \lambda_j \right)^{(n-p-1)/2} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{p} \lambda_j \right\}.$$

A Lalley jegyzet Theorem 1 eredménye ismerteti az úgynevezett Weyl integrációs formulát, amely megadja egy véletlen ortogonális mátrix sajátértékeinek az eloszlását abban az esetben, ha e véletlen mátrix eloszlásának van sűrűségfüggvénye, és az kifejezhető, mint a sajátértékek szimmetrikus függvénye. Az 5.2 tétel alapján e tétel feltételei teljesülnek a standard Wishart mátrixokra, és az 5.4 tétel a Weyl integrációs formula következménye.

Érdemes megfogalmazni az 5.4 tétel alábbi következményét.

Következmény. *Egy* p *dimenziós* n *szabadságfokú* $\mathbf{W}$ *standard Wishart mátrix az* $n \ge p$ *esetben 1 valószínűséggel invertálható.*

Valóban, az 5.4 tétel szerint ebben az esetben a $\mathbf{W}$ mátrixnak 1 valószínűséggel nincs 0 sajátértéke. Innen következik a Következmény állítása. Ha

$n < p$, akkor $\mathbf{W}$ 1 valószínűséggel nem invertálható, mert ekkor $\mathbf{W}$ egy p dimenziós mátrix, amelynek rangja kisebb vagy egyenlő, mint $n \le p$. Megjegyzem, hogy az $n \ge p$ esetben a $\mathbf{W}$ standard Wishart mátrix sajátértékei különbözőek.

Az 5. fejezet befejező része többdimenziós normális eloszlással kapcsolatos hipotézisvizsgálati problémákkal foglalkozik. Ezekben fontos szerepet játszik az alábbi 6.1 tétel. Ez az úgynevezett Hotteling-féle T^2 eloszlásoknak adja meg az explicit alakját. A tétel megfogalmazása előtt először a Hotteling eloszlás fogalmát ismertetem.

Hotteling-féle T^2 eloszlás definiciója. *Legyen adva egy $\mathcal{W}_p(n, \mathbf{C})$ Wishart mátrix, $n \ge p$, és egy tőle független $\xi = (\xi_1, \dots, \xi_p)^T$ $\mathcal{N}_p(\mathbf{m}, \mathbf{C})$ eloszlású véletlen vektor. E vektor $\mathbf{C}$ kovarianciamátrixa és a Wishart mátrix $\mathbf{C}$ paramétere megegyezik. Ekkor a*

$$T^2 = \xi^T \mathbf{W}^{-1} \xi$$

valószínűségi változót Hotteling eloszlásúnak nevezzük n, p, $\mathbf{C}$ és $\mathbf{m}$ paraméterekkel. Az n paramétert a Hotelling eloszlás szabadsági fokának nevezzük.

Érvényes a következő eredmény.

6.1 Tétel. *Ha $\mathbf{m} = \mathbf{0}$, akkor az n, p és $\mathbf{C}$ paraméterű T^2 eloszlás megegyezik az $\mathcal{F}(p, n-p+1)$ Fisher eloszlással, azaz egy olyan valószínűségi változó eloszlásával, amelyet úgy kapunk, hogy egy $\chi^2(p)$ eloszlású valószínűségi változót elosztunk egy tőle független $\chi^2(n-p+1)$ eloszlású valószínűségi változóval.*

A 6.1 tétel a következő két eredmény segítségével bizonyítható.

6.2 Lemma. *Legyen $\mathbf{X}$ $p \times n$ méretű mátrix, amelynek $X_{i,j}$, $1 \le i \le p$, $1 \le j \le n$, elemei független, standard normális eloszlású valószínűségi változók. Legyen $\mathbf{Q}$ az $\mathbf{X}$ mátrixtól független véletlen ortogonális $n \times n$ méretű mátrix. Ekkor az $\mathbf{XQ}$ és $\mathbf{X}$ mátrix eloszlása megegyezik, és az $\mathbf{XQ}$ mátrix is független $\mathbf{Q}$-tól.*

6.3 Lemma. *Legyen az $\mathbf{X}$ $p \times n$, $p \le n$, ugyanaz a véletlen mátrix, mint a 6.2 lemmában. Definiáljuk az $\mathbf{S} = \mathbf{X}\mathbf{X}^T$ Wishart mátrixot, és jelölje $\mathbf{S}_1$ azt a mátrixot, amelyet úgy kapunk, hogy vesszük az $\mathbf{S}$ mátrixot, és elhagyjuk annak utolsó sorát és utolsó oszlopát. Ekkor az $\frac{|\mathbf{S}|}{|\mathbf{S}_1|}$ valószínűségi változó eloszlása $\chi^2(n-p+1)$.*

A 6.2 lemma bizonyítása egyszerű. Azt kell kihasználni, hogy tetszőleges i indexre az $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,n})^T$ vektor eloszlásának sűrűségfüggvénye csak az $\mathbf{X}_i$ vektor hosszától függ, ezért minden $\mathbf{Q}$ ortogonális mátrixra $\mathbf{X}_i$ és $\mathbf{X}_i\mathbf{Q}$ eloszlása megegyezik. Innen következik, hogy ha a $\mathbf{Q}$ véletlen ortogonális mátrix független az $\mathbf{X}$ mátrixtól, akkor az $\mathbf{XQ}$ mátrix feltételes eloszlása feltéve $\mathbf{Q}$ értékét megegyezik az $\mathbf{X}$ mátrix eloszlásával. Ebből következik a 6.2 lemma.

A 6.3 lemma bizonyítása bonyolultabb. Csak a bizonyítás vázát írom le, illetve egy megjegyzést teszek arról, hogy hol kellett volna a könyvben részletesebb indoklást adni.

A bizonyítás lényeges része egy olyan indukciós állítás bizonyítása, amely szerint az $\frac{|\mathbf{S}|}{|\mathbf{S}_1|}$ tört eloszlása megegyezik akkor, ha $\mathbf{X}$ $p \times n$ mátrix és ha alkalmasan definiált $(p-1) \times (n-1)$ méretű mátrix. Ennek érdekében konstruálunk olyan az $(X_{1,1}, \dots, X_{1,n})$ koordinátáktól függő $\mathbf{Q}$ ortogonális mátrixot, amelyre $(X_{1,1}, \dots, X_{1,n})\mathbf{Q} = (R, 0, \dots, 0)$, ahol $R^2 = \sum_{j=1}^n X_{1,j}^2$. Ekkor $(\mathbf{XQ})(\mathbf{XQ})^T = \mathbf{XX}^T = \mathbf{S}$. Jelölje $\mathbf{X}^*$ azt a $(p-1) \times (n-1)$ méretű mátrixot, amelyet úgy kapunk, hogy az $\mathbf{XQ}$ mátrixnak elhagyjuk az első sorát és első oszlopát. A könyv megmutatja, hogy az $\mathbf{X}$ és $\mathbf{X}^*$ mátrix segítségével kiszámított $\frac{|\mathbf{S}|}{|\mathbf{S}_1|}$ hányados egyenlő. Ebből akarjuk levezetni az indukciós állítást. Ez lehetséges, de ehhez azt is észre kell venni, hogy a 6.2 lemma és a konstrukció alapján az $\mathbf{X}^*$ mátrix koordinátái független, standard normális eloszlású valószínűségi változók. (Ez a lépés hiányzik a könyv magyarázatából.)

A bizonyítás befejezéséhez még meg kell mutatni, hogy a $p = 2$ esetben, ha $\mathbf{X}$ $2 \times n$ méretű mátrix az $\frac{|\mathbf{S}|}{|\mathbf{S}_1|}$ hányados $\chi^2(n-1)$ eloszlású. (Ez az indukció kezdő lépése.) Ez hasonlóan történik az indukciós lépés bizonyításához. A technikai részleteket elhagyom.

A 6.1 tétel bizonyítása azon alapszik, hogy alkalmas transzformációkkal a $T^2 = \xi^T \mathbf{W}^{-1} \xi$ kifejezést, ahol $\mathbf{W} = \mathbf{XX}^T$, olyan alakra hozzuk, amely lehetővé teszi, hogy meghatározzuk az eloszlását a 6.1 és 6.2 lemma segítségével.

Először az 1.3 Tétel bizonyításához hasonló $\mathbf{C}^{-1/2}$ transzformációt alkalmazva mind a ξ vektorra, mind az $\mathbf{X}$ mátrixra a tételt redukáljuk arra a speciális esetre, amikor $\mathbf{C} = \mathbf{I}_p$, a p dimenziós identitás mátrix. Ezután veszünk egy olyan a ξ vektortól függő $\mathbf{Q}$ véletlen $p \times p$ méretű ortogonális mátrixot, amelyre $\xi^T \mathbf{Q} = (\xi_1, \dots, \xi_p)\mathbf{Q} = (0, \dots, 0, t)$ ahol $t^2 = \sum_{j=1}^p \xi_j^2$. Ekkor mivel $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_p$

$$T^2 = (\xi^T \mathbf{Q})(\mathbf{Q}^T \mathbf{W}^{-1} \mathbf{Q})(\mathbf{Q}^T \xi) = (0, \dots, 0, t)(\mathbf{Q}^T \mathbf{W}^{-1} \mathbf{Q})(0, \dots, 0, t)^T,$$

ami egyenlő t^2 szorozva a $\mathbf{Q}^T\mathbf{W}^{-1}\mathbf{Q}$ mátrix p-edik sorában és p-ik oszlopában álló taggal. Jelöljük ezt Z_p-vél. Továbbá t^2 $\chi^2(p)$ eloszlású valószínűségi változó, és

$$\mathbf{Q}^T\mathbf{W}^{-1}\mathbf{Q} = (\mathbf{Q}^T\mathbf{W}\mathbf{Q})^{-1} = (\mathbf{Q}^T\mathbf{X}\mathbf{X}^T\mathbf{Q})^{-1} = ((\mathbf{Q}^T\mathbf{X})(\mathbf{Q}^T\mathbf{X})^T)^{-1},$$

és $\mathbf{Q}^T\mathbf{X}$ olyan $p \times n$ méretű mátrix amelynek elemei standard normális eloszású valószínűségi változók, és a 6.2 lemma némi élesítése azt adja, hogy a $\mathbf{Q}^T\mathbf{X}$ mátrix független a t^2 valószínűségi változótól. Azt kell észrevenni, hogy a $\mathbf{Q}^T\mathbf{X}$ mátrix feltételes eloszlása az $\mathbf{X}$ mátrix eloszlásával egyenlő akkor is, ha a a ξ vektor értékét és a (tőle függő) $\mathbf{Q}$ mátrixot rögzítjük a feltételben. Ez azt jelenti,hogy Z_p független a t^2 valószínűségi változótól, és egy p-dimenziós n szabadságfokú standard Wishart mátrix inverzének a p-ik sorában és p-ik oszlopában levő elem. Ezért a mátrixok inverzének kiszámításáról szóló formula és a 6.3 lemma azt adja, hogy $\frac{1}{Z_p}$ eloszlása $\chi^2(n-p+1)$. Innen következik a 6.1 tétel.

Az 5. fejezet utolsó témája annak tesztelése egy minta alapján, hogy egy többváltozós normális eloszlású vektor várható értéke egy feltételezett $\mathbf{m}_0$ vektor-e, illetve annak tesztelése, hogy két többváltozós normális eloszlású vektornak, amelyek mindegyikéről egy mintával rendelkezünk, és amelyeknek megegyezik a kovariancamátrixuk megegyezik-e a várható értékük is. Külön vizsgálják azt az esetet, amikor a vizsgált véletlen vektor(ok) kovarianciamátrixának az értékét ismerjük, és amikor nem ismerjük azt. A feladat ismert kovarianciamátrixok esetén lényegesen egyszerűbb. Ennek tárgyalását, ami a könyvben megtalálható, elhagyom. Viszont ismertetem röviden az eredményeket, majd összehasonlítom őket azokkal az ezután tárgyalt eredményekkel, amelyeket akkor kapunk, ha e feladatok azon analogonjait tekintjük, amelyekben a kovarianciamátrix ismeretlen.

Az $\mathbf{m} = \mathbf{m}_0$ nullhipotézist az $n((\bar{\mathbf{X}} - \mathbf{m}_0)^T\mathbf{C}^{-1}(\bar{\mathbf{X}} - \mathbf{m}_0)$ statisztika segítségével vizsgáljuk, amely a nullhipotézis teljesülése esetén $\chi^2(p)$ eloszlású. Akkor fogadjuk el a nullhipotézist, ha e statisztika értéke egy az ε elsőfajú hibától függő számnál kisebb. Látni fogjuk, hogy ismeretlen kovarianciamátrix esetén hasonló eredmény érvényes, csak ekkor a felhasznált statisztikát megadó képletben a $\mathbf{C}$ kovarianciamátrixot annak $\frac{\mathbf{S}}{n}$ maximum likelihood becslésével helyettesítjük, és ez a statisztika Hotteling eloszlású alkalmas paraméterekkel.

A másik feladat az, hogy ha adva van két független p-dimenziós normális eloszlású vektorokból álló minta n és m elemszámmal, mind a kettő ismert $\mathbf{C}$

kovarianciamátrixszal, akkor hogyan ellenőrízzük azt a nullhipotézist, hogy a két minta elemeinek ugyanaz a várható értéke. Ezt a következő statisztika segítségével érdemes megtenni.

$$\frac{nm}{n+m}(\bar{\mathbf{X}}-\bar{\mathbf{Y}})^T\mathbf{C}^{-1}(\bar{\mathbf{X}}-\bar{\mathbf{Y}}),$$

ahol $\bar{\mathbf{X}}$ és $\bar{\mathbf{Y}}$ a két minta átlaga. Ez a statisztika $\chi^2(p)$ eloszlású a nullhipotézis teljesülése esetén. Ha a kovarianciamátrixot nem ismerjük, akkor olyan statisztikát veszünk amelyben a $\mathbf{C}$ kovarianciamátrixot annak természetes becslésével helyettesítjük, és ez a módosított statisztika Hotteling eloszlású.

Lényegesen nehezebb a feladat akkor, ha a kovarianciamátrixot nem ismerjük. Ismertetem annak a feladatnak a vizsgálatát, hogy hogyan ellenőrizzük egy n elemű minta segítségével azt, hogy egy ismeretlen kovarianciájú többváltozós normális eloszlású vektor várható értéke egy előírt $\mathbf{m}_0$ vektor-e. Először arról írok, hogy milyen matematikai feladatot kell megoldani egy jó ellenőrzési eljárás kidolgozásához.

A következő feladatot tekintjük. Vegyünk egy független p-dimenziós normális eloszlású vektorokból álló $\mathbf{X}_1,\dots,\mathbf{X}_n$ sorozatot $\mathbf{m}$ várható értékkel és $\mathbf{C}$ kovarianciamátrixszal, és tekintsük e sorozatnak az $f_{\mathbf{m},\mathbf{C}}(\mathbf{X}_1,\dots,\mathbf{X}_n)$ sűrűségfüggvényét. Ezt érdemes az (5) formában megadni.

Az úgynevezett likelihood hányados próbát alkalmazzuk. Ez a következőt jelenti.

Tekintsük az $f_{\mathbf{m},\mathbf{C}}(\mathbf{X}_1,\dots,\mathbf{X}_n)$ sűrűségfüggvény maximumát az $\mathbf{m}=\mathbf{m}_0$ esetben, mint a $\mathbf{C}$ kovarianciafüggvény maximumát, azaz az

$$L_0^*=L_0^*(\mathbf{X}_1,\dots,\mathbf{X}_n)=\sup_{\mathbf{C}} f_{\mathbf{m}_0,\mathbf{C}}(\mathbf{X}_1,\dots,\mathbf{X}_n)$$

függvényt, ahol a maximum a $\mathbf{C}$ kovarianciamátrixokra van véve. Definiáljuk hasonlóan az

$$L_1^*=L_1^*(\mathbf{X}_1,\dots,\mathbf{X}_n)=\sup_{\mathbf{m},\mathbf{C}} f_{\mathbf{m},\mathbf{C}}(\mathbf{X}_1,\dots,\mathbf{X}_n)$$

függvényt, ahol a maximum az $\mathbf{m}$ várható érték vektorra és $\mathbf{C}$ kovariancia mátrixokra van véve.

Számoljuk ki az $L_0^*(\mathbf{X}_1,\dots,\mathbf{X}_n)$ és $L_1^*(\mathbf{X}_1,\dots,\mathbf{X}_n)$ függvények értékét, és vegyük ezek

$$\lambda_n=\lambda_n(\mathbf{X}_1,\dots,\mathbf{X}_n)=\frac{L_0^*(\mathbf{X}_1,\dots,\mathbf{X}_n)}{L_1^*(\mathbf{X}_1,\dots,\mathbf{X}_n)}$$

hányadosát. Határozzuk meg a $\lambda_n(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ valószínűségi változó eloszlását. Az $\mathbf{m} = \mathbf{m}_0$ nullhipotézist akkor fogadjuk el, ha $\lambda_n > B(\varepsilon)$, ahol a $B(\varepsilon)$ számot úgy választjuk, hogy az elsőfajú hiba ε legyen. Ahhoz, hogy ezt az eljárást végrehajtsuk, ismernünk kell a λ_n valószínűségi változó eloszlását. Ezen eloszlás kiszámítását ismertetem.

Először meg kell határozni az $L_0^*(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ és $L_1^*(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ valószínűségi változók értékét. Ezt a könyv megadja. Azt állítja, hogy

$$L_0^* = (2\pi)^{-np/2} \left| \frac{1}{n} \sum_{j=1}^{n} (\mathbf{X}_j - \mathbf{m}_0)(\mathbf{X}_j - \mathbf{m}_0)^T \right|^{-n/2} e^{-np/2},$$

és

$$L_1^* = (2\pi)^{-np/2} \left| \frac{1}{n} \mathbf{S} \right|^{-n/2} e^{-np/2}.$$

Az L_1^*-re adott képlet egyszerűen következik a korábbi számolásokból, de az L_0^* mennyiségre adott képlet magyarázatra szorul.

Először azt kell észrevenni, hogy a korábbi esethez hasonló számolással kapjuk a sűrűségfüggvényre (4) formulában megadott képlet segítségével, hogy az $\mathbf{m} = \mathbf{m}_0$ esetben a maximum a $\hat{\mathbf{C}} = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{X}_j - \mathbf{m}_0)(\mathbf{X}_j - \mathbf{m}_0)^T$ helyen vétetik fel. Ezután az L_0^*-ra adott formula könnyen kiszámítható a következő észrevétel segítségével.

$$\begin{aligned}
\sum_{j=1}^{n} (\mathbf{X}_j - \mathbf{m}_0)^T \hat{\mathbf{C}}^{-1} (\mathbf{X}_j - \mathbf{m}_0) &= \sum_{j=1}^{n} \operatorname{tr} (\mathbf{X}_j - \mathbf{m}_0)^T \hat{\mathbf{C}}^{-1} (\mathbf{X}_j - \mathbf{m}_0) \\
&= \sum_{j=1}^{n} \operatorname{tr} \hat{\mathbf{C}}^{-1} (\mathbf{X}_j - \mathbf{m}_0)(\mathbf{X}_j - \mathbf{m}_0)^T \\
&= \operatorname{tr} \hat{\mathbf{C}}^{-1} \left(\sum_{j=1}^{n} (\mathbf{X}_j - \mathbf{m}_0)(\mathbf{X}_j - \mathbf{m}_0)^T \right) = \operatorname{tr} \hat{\mathbf{C}}^{-1} (n\hat{\mathbf{C}}) = n \operatorname{tr} \mathbf{I}_p = np.
\end{aligned}$$

A következő számolások segítségével tudjuk meghatározni λ_n eloszlását.

$$\begin{aligned}
\frac{1}{\lambda_n} &= \frac{L_1^*}{L_0^*} = \left(\frac{|\mathbf{S}|}{\left| \sum_{j=1}^{n} (\mathbf{X}_j - \mathbf{m}_0)(\mathbf{X}_j - \mathbf{m}_0)^T \right|} \right)^{-n/2} \\
&= \left| \mathbf{S}^{-1} \sum_{j=1}^{n} (\mathbf{X}_j - \mathbf{m}_0)(\mathbf{X}_j - \mathbf{m}_0)^T \right|^{n/2}
\end{aligned}$$

$$
\begin{aligned}
&= \left| \mathbf{S}^{-1} \left[\sum_{j=1}^{n} (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^T + n(\bar{\mathbf{X}} - \mathbf{m}_0)(\bar{\mathbf{X}} - \mathbf{m}_0)^T \right] \right|^{n/2} \\
&= \left| \mathbf{S}^{-1} \left[\mathbf{S} + n(\bar{\mathbf{X}} - \mathbf{m}_0)(\bar{\mathbf{X}} - \mathbf{m}_0)^T \right] \right|^{n/2} \\
&= \left| \mathbf{I}_p + n\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mathbf{m}_0)(\bar{\mathbf{X}} - \mathbf{m}_0)^T \right|^{n/2}.
\end{aligned}
$$

Vegyük észre, hogy e formula utolsó sorában egy olyan mátrix determinánsa van, amelynek egy kivételével az összes sajátértéke 1, mivel az $\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mathbf{m}_0)(\bar{\mathbf{X}} - \mathbf{m}_0)^T$ mátrix rangja 1. Ezen 1 rangú mátrixnak a nem nulla sajátértéke $\operatorname{tr} \mathbf{S}^{-1}(\bar{\mathbf{X}} - \mathbf{m}_0)(\bar{\mathbf{X}} - \mathbf{m}_0)^T = \operatorname{tr} (\bar{\mathbf{X}} - \mathbf{m}_0)^T\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mathbf{m}_0) = (\bar{\mathbf{X}} - \mathbf{m}_0)^T\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mathbf{m}_0)$. Innen

$$
\frac{1}{\lambda_n} = \left[1 + n(\bar{\mathbf{X}} - \mathbf{m}_0)^T\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mathbf{m}_0) \right]^{n/2}.
$$

A 4.3 tételből következik, hogy az $n(\bar{\mathbf{X}} - \mathbf{m}_0)^T\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mathbf{m}_0)$ kifejezés teljesíti a 6.1 tétel feltételeit $n-1$ és p paraméterrel. Ezért az $\mathcal{F}(p, n-p)$ eloszlású. Ezen észrevétel segítségével a λ_n valószínűségi változó eloszlását is ki tudjuk számolni. Továbbá a λ_n-t az $(\bar{\mathbf{X}} - \mathbf{m}_0)^T\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mathbf{m}_0)$ valószínűségi változó monoton csökkenő függvényeként fejeztük ki, ezért a javasolt teszt eljárást így is megadhatjuk: Akkor fogadjuk el az $\mathbf{m} = \mathbf{m}_0$ nullhipotézist, ha $(\bar{\mathbf{X}} - \mathbf{m}_0)^T\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mathbf{m}_0) < C(\varepsilon)$, alkalmas az ε elsőfajú hibától függő $C(\varepsilon)$ konstanssal.

A könyv megemlíti az 5. fejezet végén azt a problémát, hogy hogyan ellenőrizzük azt a hipotézist, hogy két különböző $\mathcal{N}_p(\mathbf{m}, \mathbf{C})$ eloszlású független valószínűségi változókból álló minta elemeinek a várható értéke megegyezik, feltéve, hogy azok kovarianciamátrixa egyenlő. A feladat tehát a következő. Legyen adva független $\mathcal{N}_p(\mathbf{m}_0, \mathbf{C})$ eloszlású valószínűségi változók $\mathbf{X}_1, \dots, \mathbf{X}_n$ sorozata és független $\mathcal{N}_p(\mathbf{m}_1, \mathbf{C})$ eloszlású valószínűségi változók $\mathbf{Y}_1, \dots, \mathbf{Y}_m$ sorozata úgy, hogy ez a két sorozat egymástól is független. Ellenőrizzük eme minták alapján az $\mathbf{m}_0 = \mathbf{m}_1$ nullhipotézist.

A könyv megadja e feladat megoldását bizonyítás nélkül. A következő eljárást javasolja. Legyen

$$
\mathbf{S} = \mathbf{S}_1 + \mathbf{S}_2 = \sum_{j=1}^{n} (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^T + \sum_{j=1}^{m} (\mathbf{Y}_j - \bar{\mathbf{Y}})(\mathbf{Y}_j - \bar{\mathbf{Y}})^T.
$$

Be lehet látni, hogy a nullhipotézis teljesülése esetén

$$
T^2 = \frac{nm}{n+m} (\bar{\mathbf{X}} - \bar{\mathbf{Y}})\mathbf{S}^{-1}(\bar{\mathbf{X}} - \bar{\mathbf{Y}})^T
$$

Hotteling eloszlású p, $n+m-2$ és $\mathbf{C}$ paraméterekkel, ezért Fisher-féle $\mathcal{F}(p, n+m-p-1)$ eloszlású. Ennek alapján a null-hipotézist elfogadjuk, ha $T^2 < C(\varepsilon)$ alkalmas az ε elsőfajú hibától függő konstanssal, amit a Fisher-féle $\mathcal{F}(p, n+m-p-1)$ eloszlás segítségével határozunk meg.

A fent leírt módszer megalapozását nyújtó elmélet tárgyalását elhagyom. Az megtalálható Giri N. (1977) Multivariate Statistical inference. Acad. Press, New York. könyvében.