Az november 23-i szeminárium témája
Rövid összefoglaló

Miért fontos számunkra az előző gyakorlaton tárgyalt lineáris algebrai ismeretek felfrissítése?

Tekintsünk $\xi_1,\dots,\xi_k$ valószínűségi változókat, melyekre $E\xi_p^2<\infty$, $1\le p\le k$. Megjegyezzük, hogy ekkor $E|\xi_p\xi_q|<\infty$ minden $1\le p,q\le k$ esetén, mert ekkor $(E|\xi_p\xi_q|)^2<E\xi_p^2E\xi_q^2<\infty$ a Cauchy–Schwarz egyenlőtlenség alapján. Statisztika előadáson szerepelt egy k-dimenziós véletlen vektor kovariancia mátrixának a fogalma. A $(\xi_1,\dots,\xi_k)$ véletlen vektor kovariancia mátrixa az a $k\times k$ mátrix, melynek p-ik sorának és q oszlopának a metszetében a $\mathrm{Cov}\,(\xi_p,\xi_q)=E(\xi_p-E\xi_p)(\xi_q-E\xi_q)=E\xi_p\xi_q-E\xi_pE\xi_q$ elem áll. Jellemezni akarjuk a különböző véletlen vektorokhoz tartozó kovariancia mátrixokat. (A kovariancia mátrix fogalma tulajdonképpen a szórásnégyzet természetes több-dimenziós általánosítása.) Be fogjuk látni, hogy egy kovariancia mátrix szimmetrikus pozítiv definit mátrix. Továbbá igaz ennek az állításnak a megfordítása. Nevezetesen, minden $k\times k$-as szimmetrikus pozitív definit mátrixhoz létezik olyan k-dimenziós véletlen vektor, melynek ez a kovariancia mátrixa. Ez utóbbi állítás a most tárgyalandó eredmények közül a legnehezebb, viszont erre az eredményre lesz szükségünk a több-dimenziós normális eloszlások vizsgálatában. Annak érdekében, hogy a fenti kijelentések értelmét megértsük fel kell idéznünk a lineáris algebrában tanultakat.

A $(x_1,\dots,x_k)$ k hosszúságú sorozatokból álló sorozatok terét tekinthetjük euklideszi térnek, ha egyrészt bevezetjük a természetes (koordinátánként végzett) összeadást és (valós) számmal való szorzást, definiáljuk az $(x,y)=\sum\limits_{p=1}^k x_py_p$ skalárszorzatot. Ennek az euklideszi térnek egy (természetes) ortogonális bázisa az

$$e_p=(0,\cdots,0,1,0,\cdots,0),\quad 1\le p\le k,$$

vektorokból álló bázis,ahol az e_p vektor p-ik koordinátája 1, az összes többi koordinátája pedig nulla. Adva egy $(\xi_1,\dots,\xi_k)$ k-dimenziós valószínűségi vektor, mely elemeire teljesülnek az $E\xi_p=0$, $E\xi_p^2<\infty$, $1\le p\le k$ relációk vezessük be a $D(p,q)=E\xi_p\xi_q$ mennyiségeket, és definiáljuk a következő bilineáris függvényt a fenti euklideszi téren:

$$A(x,y)=E\left(\sum_{p=1}^k x_p\xi_k\right)\left(\sum_{q=1}^k y_q\xi_q\right)=\sum_{p=1}^k\sum_{q=1}^k D(p,q)x_py_q,$$

ahol $x=(x_1,\dots,x_k)$, $y=(y_1,\dots,y_k)$. Az A bilineáris függvény önadjungált és pozitív szemidefinit, ugyanis $A(x,y)=A(y,x)$, és $A(x,x)=E\left(\sum\limits_{p=1}^k x_p\xi_k\right)^2\ge 0$. Továbbá a fent definiált bilineáris függvény szoros kapcsolatban van az $(\xi_1,\dots,\xi_k)$ k-dimenziós véletlen vektor $D=(D(p,q))$, $1\le p,q\le k$, $D(p,q)=A(e_p,e_q)$, kovariancia mátrixával. Valóban, $A(x,y)=xDy^*$, ahol $x=(x_1,\dots,x_k)$, $y=(y_1,\dots,y_k)$.

A lineáris algebrában, ha rögzítjük egy k-dimenziós euklideszi tér egy $(e_1, \dots, e_k)$ ortonormált bázisát akkor természetes kapcsolatot tudunk teremteni a bilineáris függvények, lineáris transzformációk és mátrixok között. Nevezetesen, legyen $A(x, y)$ bilineáris függvény, és feleltessük meg neki az $D = D(p, q)$, $1 \le p, q \le n$, mátrixot, illetve az ebben a koordinátarendszerben az e mátrix által meghatározott lineáris operátort. Jegyezzük meg, hogy az ahogy az előző paragrafusban a kovariancia mátrix definicióját megadtuk egy bilineáris függvény segítségével ezt a konstrukciós módszert követi.

Egy bilineáris függvény által meghatározott mátrix függ a választott ortonormált bázistól, de be lehet látni, hogy az általa meghatározott operátor nem. Ezért a fenti konstrukció kölcsönösen egyértelmű lineáris operátorok és (rögzített ortogonális bázis esetén) $k \times k$-as mátrixok között. Ezért ha bevezetjük egy bilineáris operátor adjungáltját, az önadjungált, pozitív definit önadjungált, unitér operátorok fogalmát, akkor beszélhetünk a neki megfelelő lineáris operátorok adjungáltjáról, továbbá (pozitív definit) önadjungált és unitér tulajdonságáról. (Emlékeztetőül: Ha A bilineáris függvény, akkor ennek A^* adjungáltját az $(xA, y) = (x, yA^*)$ azonosság minden x és y vektorra formula határozza meg. Továbbá A önadjungált, ha $A = A^*$, unitér, ha $AA^* = A^*A = I$. Egy A önadjungált operátor pozitív definit, ha $(Ax, x) > 0$ minden $x \ne 0$ vektorra.) Nem magától értendő, hogy van értelme arról beszélni, hogy egy mátrix önadjungált vagy unitér. Ugyanis, különböző bilineáris függvényeknek ugyanaz a mátrix felelhet meg, ha más ortonormált bázisban írjuk fel a mátrixukat, tehát tisztázni kell azt, hogy a fenti fogalmak nem függnek attól, hogy melyik ortonormált bázisban dolgozunk. Be lehet látni, hogy van értelme mátrixok adjungáltjáról, (pozitív definit) öndjungált vagy unitér voltáról beszélni, csak ehhez kissé jobban meg kell értenünk a lineáris algebra néhány eredményét és fogalmát. Be lehet látni (nem nehéz a bizonyítás), hogy ha egy A bilineáris függvénynek egy $D = D(p, q)$, $1 \le p, q \le k$, mátrix felel meg, akkor az A^* adjungáltjának megfelelő $D^* = D^*(p, q)$ mátrixra $D^*(p, q) = D(q, p)$, tehát a D^* mátrix kiszámítható a D mátrix segítségével. A D mátrix akkor és csak akkor önadjungált, ha $D = D^*$, unitér, ha $DD^* = D^*D = I$, ahol a szokásos mátrix szorzást tekintjük, a D önadjungált mátrix akkor és csak akkor pozitív (szemi)definit, ha $xDx^* \ge 0$ minden x vektorra.

A fenti eredményekből látható, hogy egy $(\xi_1, \dots, \xi_k)$ véletlen vektor $D = (D(p, q))$, $1 \le p, q \le k$, kovariancia mátrixa pozitív (szemi)definit mátrix, (és ennek az állításnak van értelme.) Be akarjuk látni, hogy tetszőleges D pozitív szemidefinit mátrix előáll mint alkalmas véletlen mátrix kovarianciamátrixa. Ehhez vegyük észre, hogy ha $(\xi_1, \dots, \xi_k)$ koordinátái függetlenek, (általában korrelálatlanok), $\operatorname{Var} \xi_p = 1$, $1 \le p \le k$, akkor ennek a vektornak a kovarianciamátrixa az I identitás mátrix, továbba, ha A tetszőleges ($k \times k$-as mátrix), akkor az $(\eta_1, \dots, \eta_k) = (\xi_1, \dots, \xi_k)A$ véletlen vektor kovariancia mátrixa az A^*A mátrix. Valóban, ha $(a_{p,q})$, $1 \le p, q \le k$, jelöli az A mátrixot, akkor $\operatorname{Cov}(\eta_p, \eta_q) = \operatorname{Cov}\left(\sum\limits_{u=1}^{k} a_{u,p}\xi_u, \sum\limits_{v=1}^{k} a_{v,q}\xi_v\right) = \sum\limits_{u=1}^{k} a_{u,p}a_{u,q}$, mivel $\operatorname{Cov}(\xi_u, \xi_u) = 1$ és $\operatorname{Cov}(\xi_u, \xi_v) = 0$, ha $u \ne v$. Ez a $\sum\limits_{u=1}^{k} a_{u,p}a_{u,q}$ kifejezés viszont megegyezik az A^*A mátrix p-ik p-ik sorában és q-ik oszlopában álló elemmel.

Ezért a kívánt állítás következik a következő (nem triviális) lineáris algebrai eredményből.

Lemma. *Tetszőleges A $k \times k$-as mátrixra a $D = A^*A$ mátrix $k \times k$-as szimmetrikus pozitív szemidefinit mátrix. Megfordítva, tetszőleges D $k \times k$-as pozitív szemidefinit mátrixhoz létezik olyan A $k \times k$-as mátrix, melyre $D = A^*A$. Sőt létezik olyan $k \times k$-as önadjungált pozitív szemidefinit A mátrix, melyre $D = AA = A^*A$.*

Megjegyzés: A D mátrix fenti előállítása nem egyértelmű, azaz a $D = A^*A$ egyenlet nem határozza meg egyértelműen az A mátrixot. Valóban, ha $D = A^*A$, U unitér mátrix és $\bar{A} = UA$, akkor $\bar{A}^*\bar{A} = (UA)^*UA = A^*U^*UA = AA^* = D$.

A lemma első (kevésbé fontos) állítása egyszerűen bizonyítható. Ugyanis, $(A^*A)^* = A^*(A^*)^* = A^*A$, azaz A^*A önadjungált mátrix, és $(xA^*A, x) = (xA^*, xA^*) \ge 0$, tehát D pozitív szemidefinit mátrix. A Lemma második, tartalmasabb állítása egyszerűen bizonyítható, abban a speciális esetben, ha a D mátrix diagonális, azaz a diagonálisban $(\lambda_1, \dots, \lambda_k)$ elemek vannak, és a mátrix összes többi eleme zéró. Ekkor ugyanis az, hogy a D mátrix pozitív szemidefinit ekvivalens azzal, hogy $\lambda_p \ge 0$ minden $1 \le p \le k$ számra. Ezért definiálhatjuk azt a pozitív szemidefinit diagonális A mátrixot, melynek diagonális elemei a $\left(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_k}\right)$ számok. Ekkor nyílvánvaló módon $D = AA = A^*A$.

Az általános eset visszavezethető erre az egyszerűbb speciális esetre a lineáris algebra azon alapvető eredménye alapján, mely szerint tetszőleges önadjungált lineáris operátor diagonizálható, azaz alkalmas ortonormált bázisban az operátor mátrixa diagonális. Itt az analízis egy alapvető és nagyon természetes gondolatát alkalmazzuk. Nevezetesen azt, hogy a megoldandó problémát tekintsük olyan koordinátarendszerben, melyben az a lehető legegyszerűbb alakú. Annak érdekében, hogy ezt jobban megértsük, tárgyaljuk meg ezt a diagonalizálási módszert a Lemmában megfogalmazott eredmény bizonyításában részletesebben.

Az idézett diagonalizálhatósági eredmény egyik lehetséges, leggyakrabban megfogalmazott alakja a következő: Egy $k \times k$-as A önadjungált mátrixnak létezik k darab $\bar{e}_1, \dots, \bar{e}_k$ sajátvektora $\lambda_1, \dots, \lambda_k$ sajátértékkel, azaz $\bar{e}_pA = \lambda_p\bar{e}_p$, $1 \le p \le k$. Továbbá a sajátértékek valósak és a sajátvektorok ortongonálisak. (A teljesség kedvéért jegyezzük meg, hogy abban az esetben, ha többszörös sajátértékek vannak, akkor a többszörös sajátértékekhez tartozó sajátvektorok megválasztása nem egyértelmű, és ekkor jól kell definiálni ezeket a sajátvektorokat.) Ez azt jelenti, hogy ha az önadjungált operátort ebben a koordinátarendszerben írjuk fel, akkor az diagonális, ezért a fenti diagonális mátrixokra vonatkozó érvelés elegendő a feladat bizonyításához.

A fenti módszer korrekt bizonyítást ad, ha világosan kidolgozzuk azokat a részleteket, melyek lehetővé teszik, hogy mátrixok helyett a nekik megfelelő és koordinátarendszertől független bilineáris függvényekkel illetve lineáris operátorokkal dolgozhassunk. Mégis tanulságos lehet végiggondolni azt, hogy hogyan lehet ezt a módszert az eredeti koordinátarendszerben alkalmazni.

Legyen $\bar{e}_1, \dots, \bar{e}_k$ egy D önadjungált mátrix sajátvektorokból álló ortonormált bázisa, és tekintsük az euklideszi tér eredeti $e_1, \dots, e_k$ ortonormált bázisát. Legyenek $\lambda_1, \dots, \lambda_k$ az $\bar{e}_1, \dots, \bar{e}_k$ sajátvektorokhoz tartozó sajátértékek, továbbá Λ az a diagonális mátrix, melynek diagonálisában a $\lambda_1, \dots, \lambda_k$ számok állnak. Definiáljuk az U lineáris mátrixot az $\bar{e}_p U = e_p$, $1 \le p \le k$, relációval. Vegyük észre, hogy U unitér operátor, és annak U^* adjungáltját az $e_p U^* = \bar{e}_p$, $1 \le p \le k$, reláció definiálja. (Érdemes megérteni az unitér operátorok geometriai tartalmát. Az, hogy U unitér operátor, az geometriailag azt jelenti, hogy távolság és szögtartó leképezés, tehát forgatás és esetleg utána egy tengelyre való tükrözés alkalmazása. Be lehet látni, hogy U akkor és csak akkor unitér, ha egy rögzített ortonormált bázist egy (általában) másik ortonormált bázisba visz.) Ennek az állításnak a formális bizonyítása azon múlik, hogy $(\bar{e}_p U, e_q) = (\bar{e}_p, e_q U^*)$ minden $1 \le p, q \le k$ indexre, mert az azonosság mind a két oldala nulla vagy egy, attól függően, hogy $p \ne q$ vagy $p = q$, $1 \le p, q \le k$. Innen következik, hogy a definiált U^* leképezés valóban az U adjungáltja. Végül $UU^* = U^*U = I$, mert $\bar{e}_p UU^* = e_p U^* = \bar{e}_p$, és hasonlóan $e_p U^* U = e_p$ minden $1 \le p \le k$ indexre. Végül mutassuk meg, hogy a fenti jelölésekkel $D = U\Lambda U^*$. Ez a reláció ekvivalens a $U^* D U = \Lambda$ azonossággal, vagy az $(e_p U^* D U, e_q) = (e_p \Lambda, e_q)$, minden $1 \le p, q \le k$ azonossággal. Viszont $(e_p U^* D U, e_q) = (e_p U^* D, e_q U^*) = (\bar{e}_p D, \bar{e}_q) = \lambda_p(\bar{e}_p, \bar{e}_q) = \lambda_p \delta_{p,q}$, ahol $\delta_{p,q} = 1$, ha $p = q$, és $\delta_{p,q} = 0$, ha $p \ne q$. Az $(e_p \Lambda, e_q) = \lambda_p \delta_{p,q}$ azonosság nyílvánvaló.

A fenti azonosság lehetővé teszi a Lemma egyszerű bizonyítását. Ha D pozitív szemidefinit mátrix, akkor tekintsük annak a fenti $D = U\Lambda U^*$ előállítását. Legyen $\sqrt{\Lambda}$ az a diagonális mátrix, melynek koordinátáiban a λ_p sajátértékek $\sqrt{\lambda_p}$ négyzetgyökei vannak, $1 \le p \le q$. Ekkor az $A = U\sqrt{\Lambda}U^*$ mátrixra $A = A^*$ és $AA = U\sqrt{\Lambda}U^*U\sqrt{\Lambda}U^* = U\Lambda U^* = D$, és ezt kellett bizonyítanunk.

A fenti eredmények segítségével beláthatjuk, hogy a normális eloszlású valószínűségi változók alább megadandó három lehetséges definiciója ekvivalens. A definiciók kimondása előtt idézzük fel, hogy a korábban tárgyalt eredmények alapján független egyforma eloszlású nulla várható értékű véletlen vektorok normalizált részletösszegei egy olyan eloszlás, melynek karakterisztikus függvénye a $(t_1, \dots, t_k)$ pontban a

$$\varphi(t_1, \dots, t_k) = \exp\left\{ -\sum_{p=1}^{k} \sum_{q=1}^{l} D(p,q) t_p t_q \Big/ 2 \right\}$$

függvény, ahol a $D(p,q)$ számok a tekintett véletlen vektorok kovariancia mátrixának megfelelő elemei. A több-dimenziós normális eloszlás különböző egymással ekvivalens definiciója az ilyen karakterisztikus függvénnyel rendelkező valószínűségi vektorokat írja le, ha esetleg még egy konstans vektort adunk hozzá.

Normális eloszlású valószínűségi vektorok fogalma I.: *A $(\xi_1, \dots, \xi_k)$ valószínűségi vektor akkor normális eloszlású, ha a $\sum\limits_{p=1}^{k} a_p \xi_p$ egydimenziós valószínűségi változó normális eloszlású tetszőleges $a_1, \dots, a_k$ valós számokkal.*

Normális eloszlású valószínűségi vektorok fogalma II.: *A $(\xi_1, \dots, \xi_k)$ valószínűségi vektor akkor normális eloszlású, ha annak karakterisztikus függvénye*

$$\varphi(t_1, \dots, t_k) = Ee^{i(t_1\xi_1 + \dots + \xi_k)} = \exp\left\{ -\frac{1}{2} \sum_{p=1}^{k} \sum_{q=1}^{k} D(p,q) t_p t_q + i \sum_{p=1}^{k} m_p t_p \right\}$$

alakú, ahol $D(p,q)$, $1 \le p,q \le k$, pozitív szemidefinit szimmetrikus mátrix, és m_p, $1 \le p \le k$, tetszőleges valós számok. Ekkor $D(p,q) = \operatorname{Cov}(\xi_p - E\xi_p)(\xi_q - E\xi_q)$, *$1 \le p,q \le k$, és $m_p = E\xi_p$, $1 \le p \le k$.*

Normális eloszlású valószínűségi vektorok fogalma III.: *A $(\xi_1,\dots,\xi_k)$ valószínűségi vektor akkor normális eloszlású, ha létezik olyan független $(\eta_1,\dots,\eta_k)$ standard normális eloszlású valószínűségi változókból álló véletlen vektor, valamint $A = a(p,q)$, $1 \le p,q \le k$, $k \times k$ mátrix és $(m_1,\dots,m_k)$ k-dimenziós valós számokból álló vektor, melyekre az $(\eta_1,\dots,\eta_k)A + (m_1,\dots,m_k)$ azaz a $(\zeta_1,\dots,\zeta_k)$, $\zeta_p = \sum\limits_{q=1}^{k} a(q,p)\eta_q + m_p$, $1 \le p \le k$ vektor eloszlása megegyezik a $(\xi_1,\dots,\xi_k)$ véletlen vektor eloszlásával. Jelölje ebben az esetben $D(p,q)$ az A^*A mátrix p-ik sorának q-ik elemét. Ekkor Ekkor $D(p,q) =$* $\operatorname{Cov}(\xi_p - E\xi_p)(\xi_q - E\xi_q)$, *és $1 \le p,q \le k$, $m_p = E\xi_p$, $1 \le p \le k$.*

Lássuk be először az alábbi definiciók ekvivalenciája közül a legnehezebbet, a Definició II. $\Rightarrow$ Definició III. állítást.

A Definició II.-ben szereplő $D(p,q)$ számok által meghatározott $D = (D(p,q))$, $1 \le p,q \le q$ $k \times k$-as mátrix önadjungált és pozitív szemidefinit. Ezért az ezen gyakorlaton tárgyalt lemma alapján felírható $D = A^*A$ alakban. Ekkor a $(\xi_1,\dots,\xi_k) = (\eta_1,\dots,\eta_k)A + (m_1,\dots,m_k)$ k-dimenziós vektor, ahol $(\eta_1,\dots,\eta_k)$ független standard normális eloszlású valószínűségi változók, olyan véletlen vektor, melynek a karakterisztikus függvényét a Definició II. képletben megadott képlet adja meg. Ugyanis tetszőleges $t_1,\dots,t_k$ valós számokra a $t_1\xi_1 + \cdots + t_k\xi_k$ valószínűségi változó normális eloszlású (emlékeztetőül: független (egydimenziós) normális eloszlású valószínűségi változók lineáris kombinációja is normális eloszlású) $\operatorname{Var}(t_1\xi_1 + \cdots + t_k\xi_k) = \sum\limits_{p=1}^{k}\sum\limits_{q=1}^{k} t_p t_q \operatorname{Cov}(\xi_p\xi_q) = tA^*At^* = tDt^*$ szórásnégyzettel, ahol $t = (t_1,\dots,t_k)$ és $\sum\limits_{p=1}^{k} m_p t_p$ várható értékkel. Mivel egy valószínűségi vektor eloszlását meghatározza annak karakterisztikus függvénye, ez azt jelenti, hogy a fenti reprezentáció egy a harmadik definicióban szereplő valószínűségi vektorral azonos eloszlású valószínűségi vektort állít elő, azaz teljesíti a harmadik definiciót.

A Definició III. $\Rightarrow$ Definició I. állítás bizonyítása egyszerű. Ha a $(\xi_1,\dots,\xi_k)$ véletlen vektor eloszlása megegyezik egy a Definició III.-ban megadott független standard eloszlású valószínűségi változók lineáris transzformációjával (plusz egy determinisztikus vektor), akkor a $(\xi_1,\dots,\xi_k)$ véletlen vektor valamilyen lineáris transzformációjának az eloszlása megegyezik független normális eloszlású valószínűségi változók (általában egy másik) lineáris transzformációjával, ezért az is normális eloszlású.

A Definició I. $\Rightarrow$ Definició II. állítás bizonyítása is egyszerű. Ha a $(\xi_1,\dots,\xi_k)$ véletlen vektor teljesíti a Definició I. állítását, akkor tetszőleges $t_1\xi_1 + \cdots + t_k\xi_k$ valószínűségi változó normális eloszlású, melynek ki tudjuk számítani a várható értékét és szórásnégyzetét. Innen látható, hogy a $(\xi_1,\dots,\xi_k)$ véletlen vektor karakterisztikus függvénye olyan, mint azt a Definició II.-ben megadtuk.