A Valószínűségszámítás II. előadássorozat hetedik előadása.
2002. október 29.

Határeloszlástételek független vektor értékű valószínűségi változókra.

Hangsúlyoztuk, hogy a Lindeberg féle centrális határeloszlástétel nemcsak azt állította, hogy független valószínűségi változóknak a Lindeberg feltételt teljesítő sorozatának normalizáltjai eloszlásban a normális eloszláshoz konvergálnak, hanem azt is, hogy a normális határeloszlástétel a ,,helyes szórással" rendelkezik. Tanulságos lehet látni olyan példát, melyben a Lindeberg feltételt nem teljesítő, de egyenletesen kicsi független valószínűségi változók normalizált részletösszegei a normális eloszláshoz konvergálnak, de a határeloszlás szórásnégyzete kisebb, mint a szabályos esetekben. A példa helyességének bizonyítása érdekében először belátunk egy önmagában is érdekes és hasznos lemmát, melyet az irodalomban Slutsky lemmának is szokás hívni.

Slutsky lemma. *Legyen adva valószínűségi változók két S_n és T_n, $n=1,2,\dots,$ sorozata, melyekre az S_n, $n=1,2,\dots,$ sorozat eloszlásban konvergál valamilyen F eloszláshoz, és a T_n, $n=1,2,\dots,$ sorozat sztochasztikusan konvergál nullához, azaz $P(|T_n|>\varepsilon)\to 0$ minden $\varepsilon>0$ számra, ha $n\to\infty$. Ekkor az S_n+T_n, $n=1,2,\dots,$ sorozat szintén konvergál eloszlásban az F eloszláshoz.*

A Slutsky lemma szemléletes tartalma világos. Azt mondja ki, hogy ha egy eloszlásban konvergens sorozat elemeit nagyon kicsit változatatjuk meg, akkor a konvergencia továbbra is érvényben marad.

A Slutsky lemma bizonyítása: Tekintsük az $F(\cdot)$ határeloszlásfüggvény egy x folytonossági pontját. Azt kell belátnuk, hogy a $\lim\limits_{n\to\infty} P(S_n+T_n<x)=F(x)$ reláció is teljesül. Ennek érdekében vegyünk tetszőles $\varepsilon>0$ számhoz olyan $\delta=\delta(\varepsilon,x)>0$ számot, melyre $F(x)-\frac{\varepsilon}{2}<F(x-\delta)<F(x)<F(x+\delta)<F(x+\delta)+\frac{\varepsilon}{2}$. Mivel az $F(\cdot)$ monoton függvénynek csak megszámlálhatóan sok szakadási pontja van, az általánosság megszorítása nélkül azt is feltehetjük, hogy az $x\pm\delta$ pontok folytonossági pontjai az $F(\cdot)$ függvénynek. A feltételek teljesülése esetén létezik olyan $n_0=n_0(x,\delta,\varepsilon)$ index, melyre $P(S_n<x+\delta)<F(x+\delta)+\frac{\varepsilon}{4}$, $P(S_n>x-\delta)<1-F(x-\delta)+\frac{\varepsilon}{4}$, és $P(|T_n|\ge\delta)<\frac{\varepsilon}{4}$, ha $n\ge n_0$. Ekkor

$$P(S_n+T_n<x)\le P(S_n<x+\delta)+P(|T_n|\ge\delta)<F(x+\delta)+\frac{\varepsilon}{2}<F(x)+\varepsilon,$$

ha $n\ge n_0(\varepsilon,\delta)$, mert $\{\omega\colon S_n(\omega)+T_n(\omega)<x\}\subset\{\omega\colon S_n(\omega)<x+\delta\}\cup\{\omega\colon |T_n(\omega)|\ge\delta\}$. Hasonlóan kapjuk, hogy

$$P(S_n+T_n\ge x)<P(S_n\ge x-\delta)+P(|T_n|\ge\delta)<1-F(x)+\varepsilon,$$

ha $n\ge n_0(\varepsilon,\delta)$, mert $\{\omega\colon S_n(\omega)\ge x\}\subset\{\omega\colon S_n(\omega)+T_n(\omega)\ge x-\delta\}\cup\{\omega\colon |T_n(\omega)|\ge\delta\}$.

Az első egyenlőtleségből következik, hogy $\limsup\limits_{n\to\infty} P(S_n+T_n<x)<F(x)+\varepsilon$, a második egyenlőtlenségből pedig az, hogy $\liminf\limits_{n\to\infty} P(S_n+T_n<x)>F(x)-\varepsilon$. Mivel ezek az egyenlőtlenségek tetszőleges $\varepsilon > 0$ számra igazak, innen következik a Slutsky lemma.

Most rátérünk a kívánt tulajdonságú példa ismertetésére.

1. Példa: *Legyen ξ_n, $n = 1, 2, \dots$, független valószínűségi változókból álló sorozatot, melyek elemeinek eloszlása a következő: $P(\xi_n = n) = P(\xi_n = -n) = \dfrac{1}{4n^2}$, $P(\xi_n = 1) = P(\xi_n = -1) = \dfrac{1}{4}$, és $P(\xi_n = 0) = \dfrac{1}{2} - \dfrac{1}{2n^2}$, $n = 1, 2, \dots$. Ekkor $E\xi_n = 0$, $E\xi_n^2 = 1$. Azt állítjuk, hogy az $S_n = \sum\limits_{k=1}^{n} \xi_k$, $n = 1, 2, \dots$, részletösszegek $\frac{1}{\sqrt{n}} S_n$, $n = 1, 2, \dots$, normalizáltjai eloszlásban konvergálnak egy nulla várható értékű, $\frac{1}{2}$ szórásnégyzetű normális valószínűségi változóhoz.*

Az 1. Példa igazolása: Vezessük be az $X_n = \xi_n I(|\xi_n| \le 2)$, és $Y_n = \xi_n I(|\xi_n| > 2)$ valószínűségi változókat, ahol $I(A)$ jelöli az A halmaz indikátorfüggvényét, és vezessük be az $\bar{S}_n = \sum\limits_{k=1}^{n} X_k$ és $T_n = \sum\limits_{k=1}^{n} Y_k$, $n = 1, 2, \dots$, részletösszegeket. Ekkor $\frac{1}{\sqrt{n}} \bar{S}_n$ sorozat eloszlásban konvergál a nulla várható értékű és $\frac{1}{2}$ szórásnégyzetű normális eloszláshoz, mert $EX_n = 0$, $EX_n^2 = \frac{1}{2}$, $s_n^2 = \sum\limits_{k=1}^{n} EX_k^2 = \frac{n}{2}$, és az X_n, $n = 1, 2, \dots$, valószínűségi változók teljesítik a Lindeberg feltételt, hiszen $\sum\limits_{k=1}^{n} EX_k^2 I(|X_k| \ge \varepsilon s_n) = 0$, ha $n \ge n_0(\varepsilon)$. Másrészt a $\frac{1}{\sqrt{n}} T_n$ kifejezések sztochasztikusan konvergálnak nullához, ha $n \to \infty$. Ez látható, ha észrevesszük, hogy $\sum\limits_{k=1}^{\infty} P(Y_k \ne 0) < \infty$, ezért a Borel–Cantelli lemma alapján egy valószínűséggel csak véges sok $Y_k(\omega)$ nem egyenlő nullával, és ezért $\sum\limits_{k=1}^{\infty} |Y_k(\omega)| \le K(\omega)$. Innen következik, hogy majdnem minden $\omega \in \Omega$ pontban $|T_n(\omega)| \le K(\omega)$ minden $n = 1, 2, \dots$ számra, ahol a jobboldalon szereplő becslés függhet az ω elemi eseménytől, de nem függ az n indextől. A Slutsky lemmából, az $\frac{1}{\sqrt{n}} S_n = \frac{1}{\sqrt{n}} \bar{S}_n + \frac{1}{\sqrt{n}} T_n$ azonosságból és a következik, hogy az $\frac{1}{\sqrt{n}} S_n$ és a $\frac{1}{\sqrt{n}} T_n$ sorozat sztochasztikus konvergeniájából nullához következik, hogy a $\frac{1}{\sqrt{n}} S_n$ és $\frac{1}{\sqrt{n}} \bar{S}_n$ sorozatoknak ugyanaz a határeloszlásuk, ezért igaz a példában megfogalmazott állítás.

Jegyezzük meg azt is, hogy a példában szereplő ξ_k valószínűségi változók teljesítik az egyenletes kicsiség feltételét, de nem teljesítik a Lindeberg feltételt. Valóban, $E\xi_k^2 = 1$, $s_n^2 = \sum\limits_{k=1}^{n} E\xi_k^2 = n$, ezért $\lim\limits_{n\to\infty} \sup\limits_{1 \le k \le n} \frac{E\xi_k^2}{s_n^2} = 0$, azaz az egyenletes kicsiség feltétele teljesül. Másrészt $\lim\limits_{n\to\infty} \sum\limits_{k=1}^{n} E\xi_k^2 I(|\xi_k| > \varepsilon s_n) = \lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{k=1}^{n} EY_k^2 = \frac{1}{2}$, tehát a Lindeberg feltétel ebben a példában nem teljesül.

Az egyforma eloszlású független valószínűségi változók normálizált összegeiről szóló centrális határeloszlástételben feltettük, hogy az összeadandóknak van véges második momentumuk. Ez természetes feltétel, mégis érdekes lehet megmutatni, hogy lehetséges olyan eset is, amikor végtelen szórásnégyzetű független, egyforma eloszlású valószínűségi vátltozók alkalmasan normalizált részletösszegei is konvergálnak eloszlásban a normális eloszláshoz. Ennek lehetőségét mutatjuk meg a következő példában, amelyik mögött hasonló gondolat van, mint az első példában.

2. Példa: *Legyen $\xi_1, \xi_2, \dots$ független egyforma eloszlású valószínűségi változók sorozata, mely sorozat tagjainak létezik sűrűségfüggvénye, és azt az $f(x) = \frac{1}{|x|^3}$, ha $|x| \ge 1$, $f(x) = 0$, ha $-1 < x < 1$ képlet adja meg. Ekkor $E\xi_1 = 0$, $E\xi_1^2 = \int_{x\colon\, |x|\ge 1} x^2 \frac{1}{|x|^3}\, dx = \infty$. Továbbá az $S_n = \sum\limits_{k=1}^n \xi_k$, $n = 1, 2, \dots$, részletösszegek $\frac{1}{\sqrt{n \log n}} S_n$ normalizáltjai eloszlásban konvergálnak a standard normális eloszláshoz.*

Az 2. Példa igazolása: Definiáljuk minden $n = 1, 2, \dots$ számra az $\eta_{k,n} = \xi_k I(|\xi_k| \le \sqrt{n \log\log n})$ és az $\zeta_{k,n} = \xi_k I(|\xi_k| > \sqrt{n \log\log n})$, $1 \le k \le n$ valószínűségi változókat, valmint ezek $\bar S_n = \sum\limits_{k=1}^n \eta_{k,n}$ és $T_n = \sum\limits_{k=1}^n \zeta_{k,n}$ összegeit. Ekkor

$$\frac{1}{\sqrt{n \log n}} S_n = \frac{1}{\sqrt{n \log n}} \bar S_n + \frac{1}{\sqrt{n \log n}} T_n$$

ezért a Slutsky lemma alapján elegendő belátni, hogy a $\frac{1}{\sqrt{n \log n}} \bar S_n$ valószínűségi változók eloszlásban konvergálnak a standard normális eloszláshoz, és a $\frac{1}{\sqrt{n \log n}} T_n$ valószínűségi változók sztochasztikusan konvergálnak nullához. Az első állítás igazolásához azt kell megmutatni, hogy a $\bar\xi_{k,n} = \frac{\xi_{k,n}}{\sqrt{n \log n}}$, $1 \le k \le n$, szériasorozatra alkalmazhatjuk a centrális határeloszlástételt. Ehhez vegyük észre, hogy $E\bar\xi_{k,n} = 0$,

$$E\bar\xi_{k,n}^2 = \int_{x\colon\, 1 \le |x| \le \sqrt{n \log\log n}} \frac{x^2}{n \log n} \frac{1}{|x|^3}\, dx = \frac{2 \log \sqrt{(n \log\log n)}}{n \log n} = \frac{\log(n \log\log n)}{n \log n},$$

ahonnan $\sum\limits_{k=1}^n E\bar\xi_{k,n}^2 = \frac{\log n + \log\log\log n}{\log n} \to 1$, ha $n \to \infty$. Ezért elég ellenőrizni, hogy a $\bar\xi_{k,n}$ valószínűségi változók teljesítik a Lindeberg feltételt, azaz minden $\varepsilon > 0$ számra

$$\sum_{k=1}^n E\bar\xi_{k,n}^2 I(|\bar\xi_{k,n}| > \varepsilon) = n \int_{x\colon\, \varepsilon\sqrt{n \log n} < |x| < \sqrt{n \log\log n}} \frac{x^2}{n \log n |x|^3}\, dx \to 0 \quad \text{ha } x \to \infty.$$

Ez az állítás viszont igaz, mivel elég nagy n-re az $\{x\colon \varepsilon\sqrt{n \log n} < |x| < \sqrt{n \log\log n}\}$ integrálási tartomány az üres halmaz, ezért a tekintett integrál nulla.

Be kell még látnunk, hogy a $\frac{1}{\sqrt{n \log n}} T_n$ valószínűségi változók sorozata sztochasztikusan nullához tart, azaz $\lim\limits_{n\to\infty} P\left(\left|\frac{1}{\sqrt{n \log n}} T_n\right| > \varepsilon\right) = 0$ minden $\varepsilon > 0$ számra.

Belátjuk, hogy az erősebb $\lim\limits_{n\to\infty} P(T_n\neq 0)=0$ reláció is teljesül. Valóban,

$$P(T_n\neq 0)\le\sum_{k=0}^n P(\zeta_{k,n}\neq 0)=n\int_{x\colon\, |x|>\sqrt{n\log\log n}}\frac{dx}{|x|^3}=\frac{1}{\log\log n}\to 0\quad\text{ha } n\to\infty.$$

Láttuk, hogy a határeloszlás szórásnégyzete lehet kisebb, mint az eloszlások szórásnégyzetének a limesze. Felmerül a kérdés, hogy lehet-e nagyobb. A válasz erre a kérdésre nemleges, és ezt mondja ki az alábbi lemma.

Lemma. *Konvergáljon F_n eloszlásfüggvények egy sorozata egy F eloszlásfüggvényhez. Ekkor $\liminf\limits_{n\to\infty}\int x^2F_n(\,dx)\ge\int x^2F(\,dx)$. Általánosabban, ha $g(x)$ tetszőleges folytonos, nem negatív (de nem feltétlenül korlátos függvény), akkor*

$$\liminf_{n\to\infty}\int g(x)F_n(\,dx)\ge\int g(x)F(\,dx).$$

A lemma bizonyítása. Először tekintsük azt az esetet, amikor $\int g(x)F(\,dx)<\infty$. Ekkor tetszőleges $\varepsilon>0$ számhoz létezik olyan $K=K(\varepsilon)>0$ szám, melyre a $g_K(x)=\min(K,g(x))$ korlátos és folytonos függvény teljesíti az $\int g_K(x)F(\,dx)\ge\int g(x)F(\,dx)-\varepsilon$ relációt. Ezért az F_n eloszlásfüggvények gyenge konvergenciájából az F eloszlásfüggvényhez következik, hogy

$$\liminf_{n\to\infty}\int g(x)F_n(\,dx)\ge\lim_{n\to\infty}\int g_K(x)F_n(\,dx)=\int g_K(x)F(\,dx)\ge\int g(x)F(\,dx)-\varepsilon,$$

és mivel ez az egyenlőtlenség minden $\varepsilon>0$ számra érvényes, innen következik a Lemma állítása ebben az esetben.

Ha $\int g(x)F(\,dx)=\infty$ akkor tetszőleges $L>0$ számhoz létezik olyan $K=K(L)>0$ szám, melyre a $g_K(x)=\min(K,g(x))$ korlátos és folytonos függvény teljesíti az $\int g_K(x)F(\,dx)\ge L$ relációt. Ezért az F_n eloszlásfüggvények gyenge konvergenciájából az F eloszlásfügggvényhez következik, hogy

$$\liminf_{n\to\infty}\int g(x)F_n(\,dx)\ge\lim_{n\to\infty}\int g_K(x)F_n(\,dx)=\int g_K(x)F(\,dx)\ge L,$$

és mivel ez az állítás minden $L>0$ számra érvényes, a Lemma állítása ebben az esetben is érvényes.

Megjegyzés: Az előző Lemma mutat némi hasonlóságot a mértékelméletben tanult Fatou lemmával, melyet ismertettem az ötödik előadás kiegészítésében. A két állítás bizonyítása is nagyon hasonló.

A fenti példák mutatják, hogy a centrális határeloszlástétel megfogalmazásában szereplő feltételek fontosak. Ezek nem teljesülése esetén másfajta jelenségek is felléphetnek. Természetes módon felmerült az az igény az elméletben, hogy adjunk teljes képet arról, hogy mikor lehetséges az, hogy független valószínűségi változók normalizált részletösszegei eloszlásban normális eloszláshoz konvergálnak, de a normálás lehet szokatlan is. Erről a problémáról sok érdekes és tartalmas eredményt bizonyítottak, de ezzel mi nem fogunk foglalkozni. Gyakorlati szempontból ugyanis csak a korábban tárgyalt eredmények fontosak, az itt tárgyalt eredmények csak nagyon speciális esetekben jelennek meg.

Érdemes megemlíteni, hogy az említett példákban a váratlan jelenségek oka az volt, hogy egy kis halmazon a valószínűségi változók nagyon nagy értéket vettek fel, ami alig befolyásolta összegeik eloszlását. Ezt a tényt a Slutsky lemma segítségével tudtuk megmutatni. Viszont egy kis halmazon felvett nagy értékek nagyon megváltoztatták a momentumokat. Minden szokatlan normálással érvényes határeloszlástétel hátterében ilyen jelenség van. Sőt, mint látni fogjuk a valószínűségszámítás egyéb problémáiban, mint például a később tárgyalandó nagy számok törvényének problémájában is hasonló problémák jelennek meg.

Határeloszlástételek vizsgálata vektorértékű független valószínűségi változók normalizált összegeire.

Tekintsük a következő két problémát

a.) Egy pénzdarabról el akarjuk dönteni, hogy szabályos-e. Ennek érdekében a pénzdarabot sokszor egymástól függetlenül feldobjuk, és feljegyezzük a kisérletek eredményét. Milyen eredménysorozat esetén tekinthetjük a pénzdarabot szabályosnak?

b.) Egy dobókockáról el akarjuk dönteni, hogy szabályos-e. Ennek érdekében a dobókockát sokszor egymástól függetlenül feldobjuk, és feljegyezzük a kisérletek eredményét. Milyen eredménysorozat esetén tekinthetjük a dobókockát szabályosnak?

Természetes úgy dönteni, hogy a pénzdarabot akkor tekintjük szabályosnak, ha a dobások körülbelül fele feje körülbelül fele pedig írásdobás. Hasonlóan, a dobókockát akkor tekintjük szabályosnak, ha mindegyik dobáseredmény az összes dobás körülbelül egyhatodában következik be. De mit jelent a „körülbelül” szó ezekben az állításokban?

Az a) probléma esetében a centrális határeloszlástétel segítségével adhatunk pontosabb választ erre a kérdésre. Végezzünk összesen n kisérletet, és legyen S_n a fedobások számát, akkor $\dfrac{S_n-\frac n2}{\frac{\sqrt n}{2}}$ a centrális határeloszlástétel következtében közel normális valószínűségi változó. Valóban vezessük be a következő ξ_k, $1\le k\le n$, valószínűségi változókat: $\xi_k=1$, ha a k-ik dobás eredménye fej, $\xi_k=0$, ha a k-ik dobás eredménye írás. Ekkor $S_n=\sum\limits_{k=1}^n\xi_k$, $E\xi_k=\frac12$, $\operatorname{Var}\xi_k=\frac14$, továbbá a ξ_k valószínűségi változók függetlenek. Ezért a centrális határeloszlástételből következik a megfogalmazott eredmény. Ennek segítségével kidolgozhatunk egy természetes stratégiát. Természetesen nem adhatunk olyan eljárást, amelyik biztosan helyes döntést ad. Próbáljunk olyan döntést hozni, amelyik egy szabályos pénzdarabot 0.9 valószínűséggel szabályosnak minősít. Természetes eljárás a következő: Keressük ki egy nomális eloszlástáblázat segítségével azt az x számot, melyre $\Phi(x) - \Phi(-x) = 0.9$, ahol $\Phi(x)$ jelöli a standard normális eloszlást. Ezután tekintsük a pénzdarabot szabályosnak, ha $-\frac{1}{2}\sqrt{n}x \le S_n - \frac{n}{2} \le \frac{1}{2}\sqrt{n}x$, ellenkező esetben pedig tekintsük a pénzdarabot szabálytalannak.

Lehet-e hasonló módszet megadni a b) feladat megoldására? A válasz erre a kérdésre igenlő, de ennek kidolgozásához be kell bizonyítani a centrális határeloszlástétel több-dimenziós változatát. Legyen ugyanis $S_n = (S_n^{(1)}, S_n^{(2)}, S_n^{(3)}, S_n^{(4)}, S_n^{(5)}, S_n^{(6)})$ az 1-es, 2-es, 3-as, 4-es, 5-ös és 6-os eredményű dobások száma, akkor ha n dobást hajtunk végre. A később tárgyalandó több-dimenziós centrális határeloszlástételben ezen véletlen vektor alkalmas normalizáltjának az aszimptotikus eloszlására kivánunk eredményt kapni. Vezessük be $\xi^{(k)} = (\xi_1^{(k)}, \xi_2^{(k)}, \xi_3^{(k)}, \xi_4^{(k)}, \xi_5^{(k)}, \xi_6^{(k)})$ valószínűségi változót, amelyiknek a j-ik koordinátája 1, ha a k-ik dobás eredménye j, és ebben az esetben az összes többi koordináta nulla. A $\xi^{(k)}$ vektor azt méri, hogy a k-ik dobásnak mennyi a hozadéka különböző dobáseredmények számához, ezért $S_n = \sum\limits_{k=1}^{n} \xi^{(k)}$. Ez azt jelenti, hogy a vizsgálandó S_n véletlen vektor előállítható, mint független véletlen vektorok összege. Ezért természetes azt vizsgálni, hogy a klasszikus centrális határeloszlástételt hogyan lehet általánosítani véletlen vektorokra. Mint a b) példa mutatja, egy ilyen eredmény hasznos lehet természetes gyakorlati problémák megoldásában is. Megjegyezzük, hogy a b) problémához hasonlóan tárgyalható az alábbi c) probléma:

c) Legyen adva k urna, és ezekbe egymástól függetlenül bedobunk n golyót egymástól függetlenül, amelyek ugyanolyan valószínűséggel esnek az egyes urnákba. Próbáljuk meg ellenőrízni azt, hogy igaz-e, hogy a golyók az egyes urnákba $p_1, \dots, p_k$ valószínűséggel esnek. (Feltesszük, hogy az elvégzett kisérletek n száma nagyon nagy.)

Annak érdekében, hogy a több-dimenziós centrális határeloszlástételt bebizonyítsuk bizonyos eredményeknek és fogalmaknak ki kell dolgozni a több-dimenziós megfelelőjét, így definiálnunk kell több-dimenziós eloszlások konvergenciáját es a karakterisztikus függvények több-dimenziós megfelelőjét. Ezután be lehet bizonyítani azt, hogy a több-dimenziós eloszlásokat is meghatározza azok karakterisztikus függvénye, és igaz az eloszlások konvergenciájank elaptételeként elnevezett eredmény több-dimenziós változata. Ezeket az eredményeket meg fogjuk fogalmazni, de a bizonyítások részleteit, melyek lényegében az egy-dimenziós esetben tárgyalt bizonyítások természetes módosításából állnak, elhagyjuk. Ezek az eredméyek lehetővé teszik, hogy a több-dimenziós centrális határeloszlástételt annak egy-dimenziós változatához hasonlóan bizonyítsuk be. Sőt, valójában a helyzet még ennél is egyszerúbb. Meg lehet mutatni, hogy a több-dimenziós centrális-határeloszlástétel közvetlen módon visszavezethető az egy-dimenziós esetre.

A fenti állítások részletesebb megtárgyalása lesz a következő programunk. De mielőtt ehhez hozzákezdenénk, megtárgyaljuk az egy-dimenziós valószínűségi változók vizsgálatában fontos szerepet játszó várható érték és szórásnégyzet fogalmának a több-dimenziós megfelelőit, a több-dimenziós várható érték és a kovarianciamátrix fogalmát. Ez utóbbi vizsgálatában fontos szerepet játszanak bizonyos alapvető lineáris algebrai ismeretek, mindenekelőtt a szimmetrikus mátrixok tulajdonságai, és azok úgynevezett főtengely transzformációja.

Több-dimenziós valószínűségi változók várható értéke és kovariancia mátrixa valamint azok legfontosabb tulajdonságai.

Először is ismertetem a tárgyalandó fogalmak definicióját.

Több-dimenziós valószínűségi változó várható értékének és kovariancia mátrixának a definiciója. *Legyen $Z = (Z_1, \dots, Z_k)$ k-dimenziós véletlen vektor. E véletlen vektor várható értéke az $EZ = (EZ_1, \dots, EZ_k)$ k-dimenziós vektor, feltéve, hogy mindegyik EZ_j, $1 \le j \le k$ várható érték létezik.*

Tegyük fel továbbá, hogy a $Z = (Z_1, \dots, Z_k)$ véletlen vektor minden koordinátája teljesíti az $EZ_j^2 < \infty$, $1 \le j \le k$, feltételt. Ekkor definiáljuk e véletlen vektor kovariancia mátrixát is, és az a $D = (d_{j,l})$, $1 \le j, l \le k$, $k \times k$ méretű mátrix, mely mátrix j-ik sorában és l-ik oszlopában lévő elem a $d_{j,l} = \operatorname{Cov}(Z_j, Z_l) = EZ_jZ_l - EZ_jEZ_l$ szám.

Megfogalmazom meg a vektorértékű valószínűségi változók várható értékének és kovariancia mátrixának néhány fontos tulajdonságát.

1. Tétel. *Legyenek $Z^{(j)} = \left(Z_1^{(j)}, \dots, Z_k^{(j)}\right)$, $1 \le j \le n$, véletlen k-dimenziós vektorok ugyanazon az $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn. Ekkor a $Z^{(1)} + \cdots + Z^{(n)}$ összeg várható értéke megegyezik a $Z^{(j)}$ vektorok várható értékeinek az összegével, azaz*

$$E\left(Z^{(1)} + \cdots + Z^{(n)}\right) = EZ^{(1)} + \cdots + EZ^{(n)}.$$

Ha egy $Z = (Z_1, \dots, Z_k)$, véletlen vektor várható értéke $M = (M_1, \dots, M_k)$, a tetszőleges valós szám, $x = (x_1, \dots, x_k)$ tetszőleges k-dimenziós vektor, akkor $E(Z + x) = EZ + x$, a $Z + x$ akkor $aZ = (aZ_1, \dots, aZ_k)$, véletlen vektor várható értéke aM, és $E(Z + x) = EZ + x$.

Ez az állítás következménye az egy-dimenziós valószínűségi változók várható értékéről tanult eredményeknek.

2. Tétel. *Ha az 1. Tételben szereplő $Z^{(j)}$, $1 \le j \le n$, véletlen vektorok függetlenek, vagy általánosabban a különböző vektorok koordinátái korrelálatlanok, ami azt jelenti, hogy $\operatorname{Cov}(Z_l^{(i)} Z_l^{(i')}) = 0$, ha $i \ne i'$, $1 \le j, l \le k$, akkor a kovariancia mátrix is additív. Részletesebben megfogalmazva: Ha a $Z^{(j)}$ mátrix kovariancia mátrixa a D_j mátrix, $1 \le j \le n$, akkor a $Z^{(1)} + \cdots + Z^{(n)}$ véletlen összeg kovariancia mátrixa a $D_1 + \cdots + D_n$ mátrix. Ha egy $Z = (Z_1, \dots, Z_k)$, véletlen vektor kovariancia mátrixa a D $k \times k$ méretű mátrix, a tetszőleges valós szám, akkor az $aZ = (aZ_1, \dots, aZ_k)$, kovariancia mátrixa az a^2D kovariancia mátrix. Továbbá, ha $x = (x_1, \dots, x_k)$ tetszőleges k-dimenziós vektor, akkor a $Z+x$ vektor kovariancia mátrixa megegyezik a Z vektor kovariancia mátrixával.*

A 2. Tétel is egyszerű következménye az egy-dimenziós valószínűségi változók várható értékének számításaról szóló állításoknak. Lássuk például azt, hogyan lehet kiszámítanai az összeg kovarianciamátrixának egy elemét

$$\mathrm{Cov}\,(Z_j^{(1)}+\cdots+Z_j^{(n)},Z_l^{(1)}+\cdots+Z_l^{(n)})=\sum_{i=1}^n\sum_{i'=1}^n EZ_j^{(i)}Z_l^{(i')}-\sum_{i=1}^n\sum_{i'=1}^n EZ_j^{(i)}EZ_l^{(i')}$$
$$=\sum_{i=1}^n \mathrm{Cov}\,(Z_j^{(i)}Z_l^{(i)})+\sum_{1\le i,i'\le n,\,i\ne i'}\mathrm{Cov}\,(Z_j^{(i)}Z_l^{(i')}).$$

Viszont a jobboldalon szereplő második tag nulla a $\mathrm{Cov}\,(Z_l^{(i)}Z_l^{(i')})=0$, ha $i\ne i'$, $1\le j,l\le k$, feltétel miatt, az első tag pedig megegyezik a $D_1+\cdots+D_n$ mátrix j-ik sorában és l-ik oszlopában szereplő taggal. A Tétel többi állításának a bizonyítása egyszerűbb.

Meg akarjuk adni, hogy milyen mátrixok jelenhetnek meg, mint alkalmas véletlen vektor kovariancia mátrixa. Ennek a kérdésnek a vizsgálatában fel kell használnunk a lineáris algebra néhány alapvető fogalmát és eredményét. Először idézzük fel a következő lineáris algebrai fogalmat.

Szimmetrikus és pozitív (szemi)definit mátrixok definiciója. *Legyen $D=(d_{j,l})$ egy $k\times k$ méretű mátrix. Azt mondjuk, hogy a D mátrix szimmetrikus, ha minden $1\le j,l\le k$ indexre $d_{j,l}=d_{l,j}$. (Pontosabban azt követeljük meg, ha nemcsak valós, hanem általános komplex értékű elemekkel rendelkező mátrixokat is tekintünk, ami ebben az előadásban nem fog előfordulni), hogy $d_{j,l}=\bar d_{l,j}$, ahol $\bar z$ a z komplex szám konjugáltja, azaz, ha $z=a+ib$, akkor $\bar z=a-ib$.) Egy $k\times k$ méretű szimmetrikus $D=(d_{j,l})$ mátrix pozitív szemidefinit, ha minden $x=(x_1,\dots,x_k)$ k-dimenziós vektorra $xDx^*=\sum\limits_{j=1}^k\sum\limits_{l=1}^k x_jd_{j,l}x_l\ge 0$ és (szigorúan) pozitív definit, ha a $xDx^*=\sum\limits_{j=1}^k\sum\limits_{l=1}^k x_jd_{j,l}x_l>0$ szigorú egyenlőtlenség is teljesül minden $(x_1,\dots,x_k)\ne(0,\dots,0)$ vektorra. (Ebben a formulában x^* jelöli az x vektor transzponáltját, azaz azt az oszlopvektort, melynek fölülről számíva l-ik eleme megegyezik az x vektor balról számított l-ik elemével. Ekkor xDx^* a szokásos vektor és mátrix szorzást jelöli.)*

Szükséges lesz a szimmetrikus mátrixok egyszerű reprezentációját és a pozitív szemidefinit mátrixok jellemzését. Ez az eredmény mártixok úgynevezett főtengelytranszformációinak megadása. Ennek az elnevezésnek az okáról a tétel után egy megjegyzésben írok. A tétel kimondása előtt idézzük fel az unitér mátrixok definicióját és legfontosabb fogalmait.

Egy U $k\times k$ méretű négyzetes mátrixot unitérnek nevezünk, ha teljesül az $UU^*=I$ azonosság, ahol I az identitás mátrix. Ez az azonosság akkor és csak akkor teljesül, ha érvényes az $U^*U=I$ azonosság. Az unitér mátrixok más ekvivelens jellemzése az, hogy sorai (és egyben az oszlopai is) ortonormált vektorok. Érdemes felidézni az unitér mátrixok geometriai tartalmát is. Egy U mátrix akkor és csak akkor unitér, ha az általa meghatározott transzformáció távolság (és ezért egyben szögtartó) transzformáció.

Szimmetrikus mátrixok főtengelytranszformációjáról szóló tétel. *Egy A szimmetrikus mátrix felírható $A=U\Lambda U^*$ alakban, ahol U unitér, Λ pedig diagonális mátrix.*

Egy $k \times k$ méretű szimmetrikus A mátrix $A = U\Lambda U^*$ előállítását meg lehet adni a következő módon: Létezik az A mátrixnak k ortonormált $u_1, \dots, u_k$ sajátvektora λ_1, … , λ_k sajátértékekkel, azaz olyan vektorok és számok melyekre teljesül az $u_j A = \lambda_j u_j$, $1 \le j \le k$, azonosság. Ekkor az U mátrixot megadhatjuk, mint azt a mátrixot, melynek j-ik oszlopa az u_j^* oszlopvektor, a Λ mátrix pedig ebben az esetben az a diagonális mátrix, melynek j-ik sorában és j-ik oszlopában a λ_j szám áll.

Egy A mátrix akkor és csak akkor pozitív szemidefinit, ha minden sajátértéke nem negatív, és akkor (szigorúan) pozitív definit, ha összes sajátértéke pozitív.

Megjegyzés: A főtengelytranszformációnak szemléletes tartalma a következő. Egy A mátrixnak (egy rögzített ortonormált bázis esetén) megfelel egy lineáris transzformáció. Ha a mátrix szimmetrikus, (amelyik tulajdonság megfogalmazható a neki megfelelő A lineáris transzformáció segítségével is, mint az $(xA, y) = (yA, x)$ azonosság), akkor létezik olyan ortonormált bázis, melyben az A mátrixnak megfelelő transzfomáció diagonális. Az erre a koordinátarendszerre való áttérést nevezik főtengelytranszformációnak. A tételben megfogalmazott eredmény tulajdonképpen azt írja le, hogy a transzformációnak ez a „szép” koordinátarendszerben megadott diagonális alakja hogyan látszik az eredeti koordinátarendszerben.

Ezután a lineáris algebrai előkészíts után meg tudjuk fogalmazni és be tudjuk bizonyítani a kovarianciamátrixok jellemzését kimondó tételt.

Tétel a kovariancia mátrixok jellemzéséről. *Legyen $Z = (Z_1, \dots, Z_k)$ egy k-dimenziós véletlen vektor. Ekkor a Z vektor kovariancia mátrixa szimmetrikus és pozitív szemidefinit mátrix. Megfordítva, tetszőleges D szimmetrikus, pozitív szemidefinit mátrixhoz létezik olyan $Z = (Z_1, \dots, Z_k)$ véletlen vektor, melynek ez a D mátrix a kovariancia mátrixa. Sőt igaz a következő tartalmasabb állítás is: Legyen $Y = (Y_1, \dots, Y_k)$ olyan véletlen vektor, melynek a kovariancia mátrixa az identitás mátrix, azaz $\mathrm{Var}\, Y_j = 1$, $1 \le j \le k$, $\mathrm{Cov}\,(Y_j, Y_l) = 0$, ha $1 \le j, l \le k$, és $j \ne l$. (Ez a helyzet például akkor, ha az Y_j, $1 \le j \le k$ valószínűségi változók függetlenek, és $\mathrm{Var}\, Y_j = 1$.) Ekkor létezik olyan $A = (a_{j,l})$ $k \times k$ méretű mátrix, melyre igaz, hogy a $Z = (Z_1, \dots, Z_k) = (Y_1, \dots, Y_k)A = \left(\sum\limits_{p=1}^{k} a_{1,p} Y_p, \dots, \sum\limits_{p=1}^{k} a_{k,p} Y_p \right)$ véletlen vektor kovariancia mátrixa a D mátrix.*

A bizonyítás felhasználja a következő lineáris algebrai eredményt.

Tétel a lineáris algebrából. *Legyen D pozitív szemidefinit mátrix. Ekkor létezik olyan A mátrix, melyre érvényes a $D = A^* A$ azonosság, ahol A^* az A mátrix transzponáltját jelöli. Sőt, azt is feltehetjük, hogy az A mátrix önadjungált és pozitív szemidefinit.*

A tétel bizonyítása a lineáris algebráról kimondott tétel segítségével. Tekintsünk először egy $Z = (Z_1, \dots, Z_k)$ egy k-dimenziós véletlen vektort és annak $D = (d_{j,l})$, $d_{j,l} = \mathrm{Cov}\,(Z_j, Z_l)$, $1 \le j, l \le k$, kovariancia mátrixát. Ekkor D szimmetrikus mátrix, mert $d_{j,l} = d_{l,j}$, azaz $\mathrm{Cov}\,(Z_j, Z_l) = \mathrm{Cov}\,(Z_l, Z_j)$. Másrészt tetszőleges $x = (x_1, \dots, x_k)$

k-dimenziós vektorra $\mathrm{Var}\left(\sum\limits_{j=1}^k x_jZ_j\right)\ge 0$. Ezenkívül

$$\begin{aligned}\mathrm{Var}\left(\sum_{j=1}^k x_jZ_j\right)&=\sum_{j=1}^k\sum_{l=1}^k E\left(x_jx_l(Z_jZ_l-EZ_jEZ_l)\right)=\sum_{j=1}^k\sum_{l=1}^k x_jx_l\mathrm{Cov}\,(Z_j,Z_l)\\&=\sum_{j=1}^k\sum_{l=1}^k x_jx_ld_{j,l}=xDx^*.\end{aligned}$$

Innen következik, hogy $xDx^*\ge 0$ tetszőleges $x=(x_1,\dots,x_k)$ k-dimenziós vektorra, azaz D szimmetrikus pozitív szemidefinit mátrix.

Megfordítva, legyen D pozitív szemidefinit mátrix, és $Y=(Y_1,\dots,Y_k)$ olyan véletlen vektor, melynek a kovariancia mátrixa az identitás mátrix, azaz $\mathrm{Var}\,Y_j=1$, $1\le j\le k$, $\mathrm{Cov}\,(Y_j,Y_l)=0$, ha $1\le j,l\le k$, és $j\ne l$. A kimondott lineáris algebrai eredmény szerint létezik olyan $A=(a_{j,l})$, $1\le j,l\le k$ $k\times k$ méretű mátrix, melyre $D=A^*A$. Azt állítom, hogy a $Z=(Z_1,\dots,Z_k)=YA$, azaz a $Z=(Z_1,\dots,Z_k)$, $Z_j=\sum\limits_{p=1}^k a_{p,j}Y_p$, $1\le j\le k$, véletlen vektor kovariancia mátrixa a D mátrix. Innen következik a feladat második állítása is. Ennek az azonosságnak a bizonyítása érdekében számoljuk ki a $\mathrm{Cov}\,(Z_j,Z_l)$, $1\le j,l\le k$, kovarianciákat. Azt kapjuk, hogy $\mathrm{Cov}\,(Z_j,Z_l)=\mathrm{Cov}\left(\sum\limits_{p=1}^k a_{p,j}Y_p,\sum\limits_{q=1}^k a_{q,l}Y_q\right)$, ezért $\mathrm{Cov}\,(Z_j,Z_l)=\sum\limits_{p=1}^k\sum\limits_{q=1}^k a_{p,j}a_{q,l}\mathrm{Cov}\,(Y_p,Y_q)$, ahonnan mivel $\mathrm{Cov}\,(Y_p,Y_q)=0$, ha $p\ne q$, és $\mathrm{Cov}\,(Y_p,Y_p)=1$, $\mathrm{Cov}\,(Z_j,Z_l)=\sum\limits_{p=1}^k a_{p,j}a_{p,l}=d_{j,l}$, ahol $d_{j,l}$ a $D=A^*A$ mátrix j-ik sorában és l-ik oszlopában szereplő konstans. Ezt kelllett bebizonyítanunk

A felhasznált lineáris algebrai tétel bizonyítása. Írjuk fel az D mátrixot $D=U\Lambda U^*$ alakban, ahol U unitér, Λ pedig diagonális mátrix. Az, hogy az D mátrix pozitív szemidefinit azt jelenti, hogy a Λ diagonális mátrix λ_j, $1\le j\le k$, elemei nem negatívak. Ezért létezik az $\sqrt{\Lambda}$ diagonális mátrix, melynek j-ik sorának j-ik oszlopában a $\sqrt{\lambda_j}$ elem áll. Definiáljuk az $A=U\sqrt{\Lambda}U^*$ mátrixot. Ekkor A szimmetrikus mátrix, mert $A^*=(U\sqrt{\Lambda}U^*)^*=U\sqrt{\Lambda}U^*=A$, valamint $A^*A=A^2=U\sqrt{\Lambda}U^*U\sqrt{\Lambda}U^*=U\Lambda U^*=D$. Az A mátrix pozitív szemidefinit is egyben, ha a λ_j számok pozitív négyzetgyökeit vesszük.)

Megjegyzés a lineáris algebrai tételhez: A nem negatív számoknak a pozitív szemidefinit mátrixok felelnek meg magasabb dimenzióban. A megfogalmazott eredmény azt mondja ki, hogy pozitív szemidefinit mátrixokból, hasonlóan a nem negatív számokhoz lehet négyzetgyököt vonni. Ez a négyzetgyökvonás nem egyértelmű. Például egy megoldás ismeretében a következő módon lehet új megoldásokat kapnin. Ha $D=A^*A$ és U unitér transzformáció, akkor az $\bar A=UA$ mátrixra is teljesül, hogy $\bar A^*\bar A=A^*U^*UA=A^*A=D$, tehát A^* is megoldása a tekintett egyenletnek. Jegyezzük meg, hogy valós számok között is csak akkor egyértelmű a gyökvonás, ha csak a pozitív gyököt tekintjük.

Ennek az állításnak érvényes a következő több-dimenziós általánosítása: A $D = A^*A$ egyenletnek pontosan egy szimmetrikus pozitív szemidefinit megoldása van. Ennek bizonyítása nem nehéz, de mivel erre nem lesz szükségünk, ezért a bizonyítást elhagyon.

Valós értékű valószínűségi változók szórásnégyzete nem-negatív, tehát nem zártuk ki annnak a lehtőségét sem, hogy nulla. De ez csak abban az elfajult esetben lehetséges, ha a tekintett valószínűségi változó konstans, azaz 1 valószínűséggel ugyanazt az értéket veszi fel. Hasonlóan, a több-dimenziós esetben csak azt követeltük meg, hogy a véletlen vektor kovarianciamátrixa pozitív szemidefefinit legyen, de megendedtük, hogy ne legyen (szigorúan) pozitív definit. De ha a kovarianciamátrix nem pozitív definit az bizonyos elfajultságot jelent, és akkor a véletlen vektozoknak van olyan nem triviális lineáris kombinációja, amelyik egy valószínűséggel konstans. Ezt a tényt fogalmazzuk meg az alábbi Lemmában.

Lemma. *Egy $\xi = (\xi_1, \dots, \xi_k)$ véletlen vektor $D = (d_{i,j})$ kovariancimátrixa akkor és csak akkor nem (szigorúan) pozitív definit, ha léteznak olyan a_j, $1 \le j \le k$, számok, melyek nem mindegyike nulla, és $\sum\limits_{j=1}^{k} a_j\xi_j = K$ valamilyen K (determinisztikus) konstanssal egy valószínűséggel.*

Bizonyítás: Ha léteznek olyan $a_1, \dots, a_k$ számok, melyek nem mindegyike nulla, és $\sum\limits_{j=1}^{k} a_j\xi_j = K$ egy valószínűséggel, akkor

$$0 = \operatorname{Var}\left(\sum_{j=1}^{k} a_j\xi_j\right) = \sum_{i=1}^{k}\sum_{j=1}^{k} a_ia_j \operatorname{Cov}(\xi_i, \xi_j) = \sum_{i=1}^{k}\sum_{j=1}^{k} a_ia_jd_{i,j},$$

ami azt jelenti, hogy a $D = (d_{i,j})$ mátrix nem (szigorúan) pozitív definit.

Megfordítva, ha a $D = (d_{i,j})$ mátrix nem (szigorúan) pozitív definit, akkor léteznek olyan $a_1, \dots, a_k$ számok, melyek nem mindegyike nulla és $\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{k} a_ia_jd_{i,j} = 0$. Ekkor a $\sum\limits_{i=1}^{k} a_i\xi_k$ valószínűségi változóra

$$\operatorname{Var}\left(\sum_{j=1}^{k} a_j\xi_j\right) = \sum_{i=1}^{k}\sum_{j=1}^{k} a_ia_j \operatorname{Cov}(\xi_i, \xi_j) = \sum_{i=1}^{k}\sum_{j=1}^{k} a_ia_jd_{i,j} = 0,$$

ahonnan $\sum\limits_{i=1}^{k} a_i\xi_k = K$ valamilyen K konstansra egy valószínűséggel.