Matematika előadások

Alkalmazott közgazdaságtan szak

Tallos Péter

BCE, Matematika tanszék

2018. június

Előszó

Ez a tankönyv a Budapesti Corvinus Egyetem *Alkalmazott közgazdaságtan* alapszakának matematika tananyagát tartalmazza. Igyekszünk minden olyan matematikai fogalmat és eljárást bemutatni, amely a mikroökonómia, makroökonómia, statisztika tárgyakban előfordul, illetve felhasználásra kerül, továbbá biztos alapokat ad további haladó szintű közgazdaságtani tárgyak elsajátításához. Ezek a témakörök az egyváltozós analízis, a valószínűségszámítás elemei, a lineáris algebra és a többváltozós szélsőérték.

Az esetek többségében kerüljük az aprólékos bizonyításokat. Azokon a helyeken azonban, ahol ez segít a mélyebb megértésben és a készség kialakításában, megkíséreljük a tételeinket és állításainkat intuitív (és nem teljesen preciz) indoklással alátámasztani. Hangsúlyt helyezünk ugyanakkor a definíciók pontos megfogalmazására.

A megtárgyalt anyagrészeket bőséges példaanyaggal illusztráljuk. Ezek a példák segítenek a megértésben, bemutatják az alkalmazás módját és számos esetben rávilágítanak a tételeinkben megfogalmazott feltételek fontosságára. A FIGYELEM megjegyzéssel ellátott részek kicsit komplikáltabb gondolatokat takarnak, magyarázatuk a tanórákon hangzik el.

Minden egyes fejezet pontosan egy tanulmányi hét tanulnivalóit tartalmazza. Ennek megfelelően mindhárom félév tananyaga, 12 szorgalmi hetet feltételezve, 12 fejezetre oszlik. Mindegyik fejezet végén részletes útmutatót adunk az otthoni tanuláshoz, és a feladatokon keresztüli gyakorláshoz. Az útmutatókban szereplő feldolgozandó tankönyvi utalásokat a következőképpen kell értelmezni.

Tankönyv-1: K. Sydsaeter, P. Hammond: Matematika közgazdászoknak, Aula Kiadó, Budapest, 2005.

Feladatgyűjtemény-1: Ernyes Éva, Mala József, Orosz Ágota, Racsmány Anna, Szakál Szilvia: Matematikai alapok feladatgyűjtemény, Aula Kiadó, Budapest, 2007.

Tankönyv-2: Denkinger Géza: Valószínűségszámítás, Tankönyvkiadó, Budapest, 1999.

Feladatgyűjtemény-2: Denkinger Géza: Valószínűségszámítási gyakorlatok, Tankönyvkiadó, Budapest, 1999.

Köszönettel tartozom a könyv lektorának Komlósi Sándornak, valamint Puskás Csaba, Ernyes Éva és Fleiner Balázs tanszéki kollégáimnak, akik megjegyzéseikkel, javaslataikkal segítettek érthetőbbé tenni a tananyagot. A tananyag oktatásában szerzett tapasztalat, és a visszajelzések függvényében az internetes file folyamatosan frissül.

Budapest, 2018. június.

Tallos Péter

Tartalomjegyzék

I. rész

Első félév: Differenciál és integrálszámítás

1. fejezet

Sorozatok

1.1. A határérték definicíója

A természetes számok $\mathbb{N}$ halmazán értelmezett

$$a : \mathbb{N} \to \mathbb{R}$$

függvényt (végtelen) sorozatnak nevezzük. A sorozat n-ik elemére az a_n jelölést használjuk. Ha ez nem okoz félreértést, a sorozat jelölésére röviden az a_n szimbólumot használjuk.

1.1 Definíció. Azt mondjuk, hogy az a_n sorozat *konvergens*, és az A számhoz tart, jelölésben $a_n \to A$, vagy

$$\lim_{n\to\infty} a_n = A \ ,$$

ha bármely ε pozitív számhoz van olyan N index, hogy minden $n \geq N$ indexre

$$|a_n - A| < \varepsilon.$$

Ha ilyen A valós szám nincs, akkor azt mondjuk, hogy a sorozat *divergens*.

Konvergens sorozat esetében azt is mondjuk, hogy az A szám az a_n sorozat határértéke.

1.2 Példa. Tekintsük az $a_n = 1/n$ sorozatot. Ekkor tetszőleges $\varepsilon > 0$ mellett legyen N az $1/\varepsilon$ számnál nagyobb egész szám. Világos, hogy minden $n \geq N$ esetén

$$\frac{1}{n} < \varepsilon$$

ezért az 1.1 Definíció értelmében

$$\lim_{n\to\infty} \frac{1}{n} = 0 \ .$$

1.3 Példa. Hasonló módon más sorozatok határértékét is meghatározhatjuk. Tekintsük például az

$$a_n = \frac{2n^2+5}{n^2-6n+8}$$

sorozatot. Ha a számlálót és a nevezőt is n^2-tel osztjuk, akkor a sorozat n-ik eleme

$$a_n = \frac{2+5/n^2}{1-6/n+8/n^2}$$

alakba írható, ahol a számláló határértéke 2, a nevezőé pedig 1, így

$$\lim_{n\to\infty} a_n = 2\ .$$

Minden irracionális szám előáll racionális számok sorozatának határértékeként. Például az $a_1 = 1.4$, $a_2 = 1.41$, $a_3 = 1.414$, $a_4 = 1.4142\ldots$ sorozat esetén

$$\lim_{n\to\infty} a_n = \sqrt{2}$$

Valóban, a 1.1 Definíció értelmében, ha $\varepsilon = 10^{-N}$, akkor az $n \geq N$ indexekre $|a_n - \sqrt{2}| < \varepsilon$.

Tipikus példa olyan sorozatra, amelynek nincs határértéke:

$$a_n = (-1)^n$$

hiszen itt a páros indexű elemek értéke 1, a páratlan indexűeké pedig -1.

1.4 Tétel. *Legyenek a_n és b_n olyan sorozatok, amelyekre $\lim a_n = A$ és $\lim b_n = B$. Ekkor $\lim(a_n + b_n) = A + B$ és $\lim(a_n b_n) = AB$. Ha még egyik b_n sem nulla és $B \neq 0$, akkor $\lim(a_n/b_n) = A/B$.*

1.2. Végtelenbe tartó sorozatok

Vannak olyan sorozatok is, amelyeket nem nevezünk konvergensnek, de létezik határértékük. Vizsgáljuk meg például az

$$a_n = 2n + 5$$

(számtani) sorozatot. Ez a sorozat bármely előre megadott K valós számnál nagyobb értékeket vesz föl egy indextől kezdve. Ilyenkor azt mondjuk, hogy a sorozat határértéke $+\infty$.

1.5 Definíció. Azt mondjuk, hogy az a_n sorozat határértéke $+\infty$, jelölésben

$$\lim_{n\to\infty} a_n = +\infty \ ,$$

ha bármely előre megadott K valós számhoz található olyan N index, hogy minden $n \geq N$ mellett $a_n > K$.

Teljesen hasonló módon értelmezhetjük azt, hogy egy sorozat a $-\infty$-hez tart, azaz $\lim_{n\to\infty} a_n = -\infty$.

1.3. A rendőr-elv

Sorozatok határértéke gyakran meghatározható más, ismert sorozatok határértéke segítségével. Ezt fogalmazza meg a rendőr-elv.

1.6 Tétel. *Legyenek a_n, b_n és c_n olyan sorozatok, amelyekre minden n indexre*

$$a_n \leq b_n \leq c_n$$

továbbá az a_n és c_n sorozatoknak létezik határértéke és

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = A \ .$$

Akkor a b_n sorozatnak is létezik határértéke éspedig $\lim_{n\to\infty} b_n = A$.

1.7 Példa. Legyen $a > 1$ valós szám, és tekintsük a $b_n = \sqrt[n]{a}$ sorozatot. Mivel $a > 1$, azért a sorozat elemei

$$\sqrt[n]{a} = 1 + h_n$$

alakban írhatók fel, ahol $h_n > 0$. Innen a binomiális tétel szerint

$$a = (1 + h_n)^n > 1 + nh_n \ .$$

Az egyenlőtlenség átrendezésével

$$0 < h_n < \frac{a-1}{n} \ .$$

Mivel a jobb oldali kifejezés nullához tart, a rendőr-elv szerint $h_n \to 0$, azaz $\sqrt[n]{a} \to 1$.

Nyilván hasonló megállapítást tehetünk, ha $0 < a \leq 1$, hiszen akkor a sorozat elemeinek reciprokaira térhetünk át.

1.4. Korlátosság és monotonitás

Egy végtelenbe tartó sorozat elemei természetesen nem maradnak két fixen választott valós szám között. Bevezetjük az alábbi definíciót.

1.8 Definíció. Az a_n sorozatot felülről korlátosnak nevezzük, ha van olyan K valós szám, hogy $a_n \leq K$ minden n indexre. Hasonlóan értelmezzük az alulról korlátos sorozatokat. Egy sorozatot korlátosnak nevezünk, ha felülről és alulról is korlátos.

1.9 Példa. Döntsük el például, hogy az

$$a_n = \frac{2n}{\sqrt{4n^2+5}+8}$$

sorozat korlátos-e? Ha a számlálót és a nevezőt is $2n$-nel osztjuk, akkor

$$a_n = \frac{1}{\sqrt{1+5/4n^2}+8/2n}$$

aminek alapján $0 \leq a_n \leq 1$, ezért a sorozat alulról és felülről is korlátos. Az is világos, hogy e sorozat legkisebb felső korlátja 1, míg egy (de nem a legnagyobb) alsó korlátja 0.

Kitüntetett szerepe van a monoton sorozatoknak.

1.10 Definíció. Azt mondjuk, hogy az a_n sorozat monoton növő, ha minden n indexre $a_n \leq a_{n+1}$. Hasonlóan értelmezzük a monoton fogyó sorozatokat.

1.11 Példa. Tekintsük például az

$$a_n = \frac{2n-1}{n+2}$$

sorozatot. Egyszerű átalakítással látható, hogy

$$a_n = \frac{2n+4-5}{n+2} = 2 - \frac{5}{n+2}$$

azaz a 2-ből levont tört értéke csökken, ha n növekszik, ezért a sorozat monoton növő, tehát $a_n \leq a_{n+1}$ minden n indexre. Az is világos, hogy ez a sorozat felülről korlátos, és a legkisebb felső korlátja 2, valamint

$$\lim_{n\to\infty} a_n = 2\ .$$

Az alábbi tételünk azt mondja ki, hogy ez a tulajdonság monoton és korlátos sorozatokra tipikus jelenség.

1.12 Tétel. *Egy monoton növő és felülről korlátos sorozat konvergens.*

A tételt nem igazoljuk, csak megjegyezzük, hogy a valós számok azon tulajdonságán múlik, hogy a felső korlátok között mindig van legkisebb. Ezt a sorozat felső határának nevezzük. Nyilván analóg állítást fogalmazhatunk meg monoton fogyó és alulról korlátos sorozatokra.

1.5. Az Euler-féle e szám

Nevezetes, és gyakran előforduló sorozat az

$$a_n = \left(1 + \frac{1}{n}\right)^n . \tag{1.1}$$

Megmutatható, hogy ez a sorozat monoton növő és felülről korlátos, így konvergens is. Ehhez felhasználjuk a számtani és mértani közepek közötti egyenlőtlenséget. Nevezetesen, ha $x_1, \ldots, x_n$ pozitív számok, akkor

$$x_1 \ldots x_n \leq \left(\frac{x_1 + \ldots + x_n}{n}\right)^n$$

bármely n természetes számra, és egyenlőség akkor és csak akkor áll fenn, ha $x_1 = \ldots = x_n$, azaz mindegyik szám egyenlő.

1.13 Állítás. *Az (1.1) alatti a_n sorozat szigorúan monoton növő és felülről korlátos.*

Bizonyítás. Legyen adott egy n természetes szám. Először tekintsük az

$$x_1 = 1 + \frac{1}{n},\ \ldots,\ x_n = 1 + \frac{1}{n},\ x_{n+1} = 1$$

$n + 1$ darab pozitív számot, amelyek nem mind egyenlőek. Ezekre a számtani-mértani egyenlőtlenség az

$$\left(1 + \frac{1}{n}\right)^n < \left(\frac{n+1+1}{n+1}\right)^{n+1} = \left(1 + \frac{1}{n+1}\right)^{n+1}$$

alakot ölti, ahonnan látszik, hogy a sorozat szigorúan monoton növő.

Másodszor tekintsük az

$$x_1 = 1 + \frac{1}{n},\ \ldots,\ x_n = 1 + \frac{1}{n},\ x_{n+1} = \frac{1}{2},\ x_{n+2} = \frac{1}{2}$$

$n+2$ darab különböző pozitív számot. Ezekre a számtani-mértani egyenlőtlenség

$$\frac{1}{4} \cdot \left(1+\frac{1}{n}\right)^n < \left(\frac{n+1+1}{n+2}\right)^{n+2} = 1$$

alakban írható fel. Innen adódik, hogy $a_n < 4$, azaz a sorozat felülről korlátos, és ezért konvergens is. □

E sorozat határértékére az e jelölést használjuk. Pontosabb számítás azt mutatja, hogy e irracionális, és

$$e = 2.7182...$$

1.14 Állítás. *Legyen α tetszőleges valós szám. Akkor*

$$\lim_{n\to\infty} \left(1+\frac{\alpha}{n}\right)^n = e^\alpha$$

1.15 Példa. Tekintsük például az

$$a_n = \left(\frac{2n+1}{2n+3}\right)^n$$

sorozatot. Ekkor

$$a_n = \left(\frac{2n+1}{2n+3}\right)^n = \frac{\left(1+\frac{1/2}{n}\right)^n}{\left(1+\frac{3/2}{n}\right)^n} \to \frac{e^{1/2}}{e^{3/2}}$$

és így $\lim_{n\to\infty} a_n = e^{-1}$.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 I/1 szakasza kidolgozott példáinak feldolgozása.
2. Házi feladatok: az I/1 szakasz 1.1.5, 1.1.6, 1.1.8, 1.2.5, 1.2.6, 1.2.8, 1.3.6, 1.3.9, 1.6.2, 1.8.6, 1.8.9 feladatai.
3. Tankönyv-1 1. fejezet és 6.4 szakasz.

2. fejezet

Végtelen sorok

2.1. Végtelen sorok konvergenciája

Tekintsünk egy végtelen valós a_k sorozatot és képezzük a formális

$$\sum_{k=1}^{\infty} a_k \tag{2.1}$$

összeget. Ezt a szimbólumot *végtelen sornak* nevezzük.

Természetesen meg kell mondanunk, hogy mit értünk egy ilyen kifejezésen, hiszen végtelen sok tagú összeget eddig nem értelmeztünk.

Legyen n tetszőleges természetes szám és vezessük be a (2.1) sor n-ik részletösszegét az alábbi módon:

$$S_n = \sum_{k=1}^{n} a_k \tag{2.2}$$

Ilyen módon egy S_n valós sorozatot képeztünk.

2.1 Definíció. Azt mondjuk, hogy a (2.1) végtelen sor *konvergens*, és az összege az S valós szám, jelölésben

$$S = \sum_{k=1}^{\infty} a_k$$

ha az S_n számsorozat konvergens és határértéke S. Ellenkező esetben a sort divergensnek nevezzük.

Egy végtelen sor tehát divergens, ha az S_n sorozatnak nincs határértéke, de akkor is, ha ez a határérték létezik, de végtelen. Például ha $a_k = (-1)^k$ minden

k esetén, akkor

$$S_n = \sum_{k=1}^{n} (-1)^k = 0 \text{ ha } n \text{ páros és } S_n = \sum_{k=1}^{n} (-1)^k = -1 \text{ ha } n \text{ páratlan}$$

és ez az S_n sorozat nyilvánvalóan divergens.

2.2. A geometriai sor

2.2 Példa. (A geometriai sor) Legyen r valós szám, és tekintsük az r hányadosú geometriai sort:

$$\sum_{k=0}^{\infty} r^k$$

Ennek a sornak az n-ik részletösszege:

$$S_n = \sum_{k=0}^{n} r^k = \frac{1 - r^{n+1}}{1 - r}$$

Itt a sorozatokról tanultak alapján $r^n \to 0$ ha $|r| < 1$, minden más esetben a sorozat divergens. Tehát a geometriai sor akkor és csak akkor konvergens, ha $|r| < 1$, és ekkor

$$S = \sum_{k=0}^{\infty} r^k = \frac{1}{1 - r}$$

hiszen $|r| < 1$ esetén $r^n \to 0$.

2.3. Konvergencia a részletösszegek alapján

2.3 Példa. További példaként vizsgáljuk meg a

$$\sum_{k=2}^{\infty} \frac{1}{k(k-1)}$$

végtelen sort. Mivel

$$\frac{1}{k(k-1)} = \frac{1}{k-1} - \frac{1}{k}$$

azért a sor n-ik részletösszege az alábbi módon írható:

$$S_n = (1 - 1/2) + (1/2 - 1/3) + \ldots + (1/(n-1) - 1/n) = 1 - 1/n$$

ahol az egymást követő zárójelekben a negatív és pozitív tagok kiejtik egymást. Ennek a sorozatnak a határértéke 1, tehát a sor konvergens és az összege $S = 1$.

2.4. Feltételek konvergenciára

2.4 Tétel. (A konvergencia szükséges feltétele) *Tegyük fel, hogy a*

$$\sum_{k=1}^{\infty} a_k$$

sor konvergens. Akkor $\lim_{k\to\infty} a_k \to 0$.

2.5 Példa. Tételünk csak szükséges feltételt fogalmaz, de nem elégséges. Például megmutatható, hogy a

$$\sum_{k=1}^{\infty} \frac{1}{k}$$

sorra a szükséges feltétel teljesül, de a sor divergens. Ezt a sort *harmonikus sornak* nevezzük.

Valóban, legyen n adott természetes szám, és tekintsük a harmonikus sor 2^n-ik részletösszegét. Csoportosítsuk a tagokat a következő módon:

$$S_{2^n} = 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \ldots + \frac{1}{8}\right) + \ldots + \left(\frac{1}{2^{n-1}+1} + \ldots + \frac{1}{2^n}\right),$$

ahol mindegyik zárolejeles kifejezésben a következő 2 hatványig megyünk el. Könnyen látható, hogy mindegyik zárójelen belül a tagok összege több, mint 1/2, ezért

$$S_{2^n} > 1 + \frac{1}{2}n\,.$$

Innen adódik, hogy a részletösszegek sorozata nem korlátos, ezért a harmonikus sor divergens.

2.6 Tétel. (A konvergencia elégséges feltétele) *Tegyük fel, hogy minden k indexre $a_k \geq 0$ és a*

$$\sum_{k=1}^{\infty} a_k$$

sor konvergens. Ha minden k indexre $0 \leq b_k \leq a_k$, akkor a

$$\sum_{k=1}^{\infty} b_k$$

sor is konvergens.

Valóban, a feltételeink szerint az $S_n = \sum_{k=1}^{n} b_k$ részletösszegek sorozata egyrészt monoton növő, másrészt felülről korlátos, tehát konvergens is.

Hasonló módon nyerhetünk elégséges feltételt divergenciára is: egy nemnegatív tagú divergens sor tagjainál nagyobb tagokból álló sor nyilvánvalóan szintén divergens!

2.7 Példa. Példaként tekintsük a $\sum_{k=1}^{\infty} 1/k^2$ sort. Mivel minden $k > 1$ indexre

$$\frac{1}{k^2} < \frac{1}{k(k-1)}$$

azért az n-ik részletösszegre az adódik, hogy

$$S_n = \sum_{k=1}^{n} \frac{1}{k^2} < 1 + \sum_{k=2}^{n} \frac{1}{k(k-1)}$$

Tehát az elégséges feltételünk szerint a fenti sor konvergens és az összege $S < 2$.

Általában megmutatható, hogy a $\sum_{k=1}^{\infty} 1/k^{\alpha}$ sor divergens, ha $\alpha \leq 1$, és konvergens, ha $\alpha > 1$ (lásd a 9. fejezetben).

2.5. Abszolút konvergencia

Ebben a szakaszban olyan végtelen sorokat vizsgálunk, amelyekben pozitív és negatív tagok is előfordulhatnak. Tekintsük ezért a

$$\sum_{k=1}^{\infty} a_k \tag{2.3}$$

végtelen sort, ahol az a_k tagok nem feltétlenül mind nemnegatívak.

2.8 Definíció. Azt mondjuk, hogy a (2.3) sor *abszolút konvergens*, ha a

$$\sum_{k=1}^{\infty} |a_k|$$

sor konvergens.

2.9 Tétel. *Ha egy sor abszolút konvergens, akkor konvergens is.*

A bizonyítás részleteibe nem megyünk. Indoklásként azonban megjegyezzük a következőket. Jelentse S_n az első n tag abszolút értékeinek összegét, ez a feltételünk szerint konvergens, azaz

$$\lim \sum_{k=1}^{n} |a_k| = \lim S_n = S\,.$$

Jelentse továbbá R_n illetve T_n a $\sum_{k=1}^{\infty} a_k$ sor első n tagjából a negatív, illetve a nemnegatív tagok összegét. Ekkor R_n monoton fogyó, illetve T_n monoton növő, és mindkét sorozat korlátos, hiszen

$$R_n \geq -S \quad \text{illetve} \quad T_n \leq S\,.$$

Ezért mindkét sorozat konvergens, jelölésben $\lim R_n = R$, illetve $\lim T_n = T$. Tehát a sor n-ik részletösszegének határértéke

$$\lim \sum_{k=1}^{n} a_k = \lim(T_n + R_n) = T + R\,,$$

azaz a sor valóban konvergens.

Az alábbi példában megmutatjuk, hogy a fenti tételünk állítása nem fordítható meg.

2.10 Példa. Tekintsük az alábbi váltakozó előjelű sort:

$$\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}$$

Világos, hogy ez a sor nem abszolút konvergens, hiszen a tagok abszolút értékeiből álló sor éppen a harmonikus sor, amely divergens.

Megmutatjuk azonban, hogy a fenti sor konvergens. Valóban, a páros indexű részletösszegek sorozatára az adódik, hogy

$$\begin{aligned} S_{2n} &= \left(1 - \frac{1}{2}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \ldots + \left(\frac{1}{2n-1} - \frac{1}{2n}\right) = \\ &= \frac{1}{2} + \frac{1}{12} + \ldots + \frac{1}{2n(2n-1)}\,. \end{aligned}$$

A 2.3 Példa alapján ez a sorozat monoton növő, és felülről korlátos, hiszen $S_{2n} < 2$. Tehát konvergens is, jelölje a határértéket

$$\lim S_{2n} = S\,.$$

Másrészt a páratlan indexű részletösszegekre

$$S_{2n-1} = S_{2n} + \frac{1}{2n}$$

ezért $\lim S_{2n-1} = S$, tehát $\lim S_n = S$. Ez azt jelenti, hogy a sor konvergens.

2.6. Hányados-kritérium

Az alábbiakban egy elégséges feltételt feltételt fogalmazunk meg sorok konvergenciájára, illetve divergenciájára. Képezzük a

$$\sum_{k=1}^{\infty} a_k$$

sor szomszédos tagjainak hányadosait, és tegyük fel, hogy létezik a

$$\lim_{k\to\infty} \left| \frac{a_{k+1}}{a_k} \right| = \alpha$$

határérték.

2.11 Tétel. (Hányados-kritérium)

- *Ha $\alpha < 1$, akkor a sor abszolút konvergens.*
- *Ha $\alpha > 1$, akkor a sor divergens.*
- *ha $\alpha = 1$, akkor mindkét eset lehetséges.*

Bizonyítás. Ha $\alpha < 1$, akkor válasszunk egy β számot, amelyre $\alpha < \beta < 1$. Ekkor valamely N indextől kezdve

$$\left| \frac{a_{k+1}}{a_k} \right| < \beta$$

minden $k \geq N$ indexre. Innen lépesenként visszafelé haladva azt kapjuk, hogy

$$|a_{k+1}| < \beta |a_k| < \beta^2 |a_{k-1}| < \ldots < \beta^{k-N+1} |a_N|$$

Tehát az $n+1$-ik részletösszegre

$$S_{n+1} = \sum_{k=0}^{n} |a_{k+1}| < \sum_{k=0}^{N-1} |a_{k+1}| + |a_N| \cdot \sum_{k=N}^{n} \beta^{k-N+1}$$

ahol az utóbbi szumma $0 < \beta < 1$ miatt egy konvergens geometriai sor részletösszege, tehát korlátos, ha $n \to \infty$. Innen adódik az állítás.

Ha $\alpha > 1$, akkor a bizonyítás hasonlóan végezhető el, csak $1 < \beta < \alpha$ választással egy divergens geometriai sorra vezetjük vissza. $\square$

2.12 Példa. Ebben a példában megmutatjuk, hogy $\alpha = 1$ esetén a sor konvergenciájáról semmit sem mondhatunk a Hányados-kritérium alapján.

Valóban, ha a divergens harmonikus sort tekintjük, akkor $a_k = 1/k$ miatt

$$\frac{a_{k+1}}{a_k} = \frac{k}{k+1} \to 1 \quad \text{ha } k \to \infty\,.$$

Ha viszont azt a konvergens sort tekintjük, ahol $a_k = 1/k^2$, akkor

$$\frac{a_{k+1}}{a_k} = \left(\frac{k}{k+1}\right)^2 \to 1 \quad \text{ha } k \to \infty\,,$$

azaz valóban mindkét eset előfordulhat.

2.13 Példa. Vizsgáljuk meg, hogy konvergens-e a

$$\sum_{k=1}^{\infty} \frac{k^2 \cdot 2^k}{k!}$$

sor. Használjuk a Hányados-kritériumot:

$$\frac{a_{k+1}}{a_k} = \frac{(k+1)^2 2^{k+1}}{(k+1)!} \cdot \frac{k!}{k^2 2^k} = 2\left(\frac{k+1}{k}\right)^2 \cdot \frac{1}{k+1} \to 0$$

Tehát $\alpha = 0 < 1$, azaz a sor konvergens.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 I/2 szakasz kidolgozott példáinak feldolgozása.
2. Házi feladatok: az I/2 szakasz 2.1.7, 2.1.8, 2.1.9, 2.1.10, 2.1.11, 2.2.3, 2.2.4, 2.2.8, 2.2.9, 2.3.10, 2.3.11, 2.3.12, 2.3.15 feladatai.
3. Tankönyv-1 6.5 szakasza.

3. fejezet

Függvények határértéke és folytonosság

3.1. Függvények határértéke

Az alábbiakban $f : \mathbb{R} \to \mathbb{R}$ függvények határértékével foglalkozunk. Legyen x_0 olyan pont (lehet $\pm\infty$ is), amelyhez van olyan x_n sorozat az f értelmezési tartományából, hogy $x_n \neq x_0$ és $x_n \to x_0$.

3.1 Definíció. Azt mondjuk, hogy f határértéke az x_0 pontban A (ez lehet $\pm\infty$ is), jelölésben

$$\lim_{x \to x_0} f(x) = A$$

ha az értelmezési tartományból vett bármely $x_n \to x_0$ sorozatra amelyre $x_n \neq x_0$, $f(x_n) \to A$.

3.2 Példa. Határozzuk meg a

$$\lim_{x \to 2} \frac{x^2 - 4}{x - 2}$$

határértéket! Ez a függvény az $x = 2$ helyen nincs értelmezve, de minden $x \neq 2$ helyen $x + 2$-vel egyenlő. Ezért könnyen látható, hogy

$$\lim_{x \to 2} \frac{x^2 - 4}{x - 2} = 4 \; .$$

3.3 Példa. Tekintsük az $f(x) = 1/x$ függvényt. Ez a függvény az $x = 0$ helyen nincs értelmezve. Másrészt az értelmezési tartományból választott bármely $x_n > 0$, $x_n \to 0$ sorozatra $f(x_n) \to +\infty$, míg ugyanezen sorozat negatívjaira $f(-x_n) \to -\infty$. Tehát ennek a függvénynek az $x = 0$ helyen nincs határértéke, azaz

$$\lim_{x\to 0} \frac{1}{x}$$

nem létezik.

3.4 Példa. Tekintsük a következő határértéket:

$$\lim_{x\to+\infty} \frac{2x^4 - 5x^3 + x - 8}{8x^3 - x^2 + 12}$$

Ha a számlálóból és a nevezőből is x^3-t kiemelünk, akkor a

$$\frac{2x - 5 + 1/x^2 - 8/x^3}{8 - 1/x + 12/x^3}$$

kifejezéshez jutunk. Ekkor bármely $x_n \to +\infty$ sorozat esetén a számláló határértéke $+\infty$, míg a nevező határértéke 8, így a tört $+\infty$-hez tart.

Hasonlóan látható, hogy e tört határértéke $-\infty$, ha $x \to -\infty$.

3.5 Példa. Lássuk be, hogy

$$\lim_{x\to+\infty} (\sqrt{1+x^2} - x) = 0\,.$$

Valóban,

$$\sqrt{1+x^2} - x = \frac{1}{\sqrt{1+x^2} + x}$$

és a jobb oldali kifejezés 0-hoz tart, ha $x \to +\infty$.

Az 1.4 Tétel alapján fogalmazhatjuk meg a következő állítást.

3.6 Tétel. *Ha az f és g függvényekre*

$$\lim_{x\to x_0} f(x) = A \quad \textit{és} \quad \lim_{x\to x_0} g(x) = B\,,$$

akkor

$$\lim_{x\to x_0} (f(x) + g(x)) = A + B \quad \textit{és} \quad \lim_{x\to x_0} f(x)g(x) = AB\,.$$

Ha még $B \neq 0$ (ekkor g nem nulla az x_0 egy környezetében), akkor

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = \frac{A}{B}\,.$$

3.2. A rendőr-elv

A sorozatokhoz hasonlóan a függvények határértékére is érvényes a rendőr-elv.

3.7 Tétel. *Legyenek f, g és h olyan függvények, amelyekre minden x mellett*

$$f(x) \leq g(x) \leq h(x) \ ,$$

továbbá $\lim_{x\to x_0} f(x) = \lim_{x\to x_0} h(x) = A$. Akkor a g függvénynek is létezik határértéke az x_0 helyen, és

$$\lim_{x\to x_0} g(x) = A \ .$$

3.8 Példa. Határozzuk meg a

$$\lim_{x\to 0} \frac{\sin x}{x}$$

határértéket! Ez a függvény páros, így elég pozitív x-eket vizsgálni. Geometriai interpretáció mutatja, hogy minden $0 < x < \pi/2$ pontban

$$\sin x < x < \tan x$$

azaz a $\sin x$ kifejezéssel osztva és reciprokra térve

$$\cos x < \frac{\sin x}{x} < 1 \ .$$

Tehát a rendőr-elv alkalmazásával azt kapjuk, hogy

$$\lim_{x\to 0} \frac{\sin x}{x} = 1 \ .$$

3.3. Egyoldali határérték

3.9 Definíció. Azt mondjuk, hogy az f függvénynek az x_0 helyen létezik jobb oldali határértéke, és ez A, jelölésben

$$\lim_{x\to x_0+} f(x) = A$$

ha az értelmezési tartományból választott bármely $x_n \to x_0$, $x_n > x_0$ sorozatra $f(x_n) \to A$. Analóg módon értelmezzük a bal oldali határértéket is.

3.10 Példa. Vizsgáljuk meg az alábbi függvényt:

$$f(x) = \frac{2x+1}{x-2}$$

Nem hehéz belátni, hogy ha x_n jobbról tart 2-höz, akkor $f(x_n) \to +\infty$, míg ha x_n balról tart 2-höz, akkor $f(x_n) \to -\infty$. Ezért

$$\lim_{x \to 2-} f(x) = -\infty \quad \text{és} \quad \lim_{x \to 2+} f(x) = +\infty \ .$$

3.4. Folytonosság

Tekintsünk egy intervallumon értelmezett f függvényt.

3.11 Definíció. Azt mondjuk, hogy az f függvény az értelmezési tartomány valamely x_0 pontjában folytonos, ha

$$\lim_{x \to x_0} f(x) = f(x_0) \ .$$

Ha f az értelmezési tartomány valamely x_0 pontjában nem folytonos, akkor azt mondjuk, hogy ott szakadása van.

FIGYELEM! Folytonosságról csak az értelmezési tartomány pontjaiban beszélhetünk. Például az $f(x) = 1/x$ függvény az értelmezési tartomány (azaz $x \neq 0$) minden pontjában folytonos. Az $x_0 = 0$ pont nincs az értelmezési tartományban, így ott nem is beszélhetünk szakadásról.

Másrészt azonban nem is definiálhatjuk az f függvényt az $x_0 = 0$ pontban úgy, hogy ott folytonos legyen, hiszen ott a függvénynek nincs határértéke.

Általában elmondható, hogy a folytonos függvényekből kompozícióval, illetve az alapműveletekkel előállított függvények folytonosak, kivéve, ha egy hányados nevezője nulla.

3.12 Példa. Tekintsük például az alábbi függvényt:

$$f(x) = \begin{cases} \frac{1-\cos x}{x^2} & \text{ha } x \neq 0 \\ \frac{1}{2} & \text{ha } x = 0 \end{cases}$$

Világos, hogy ez a függvény az $x \neq 0$ helyeken folytonos, másrészt

$$\frac{1-\cos x}{x^2} = \frac{1-\cos^2 x}{(1+\cos x)x^2} = \left(\frac{\sin x}{x}\right)^2 \cdot \frac{1}{1+\cos x}$$

amiből látható, hogy a függvény határértéke a 0 helyen $1/2$. Tehát ez a függvény az egész számegyenesen folytonos.

3.5. Folytonos függvények tulajdonságai

Egy folytonos függvényt úgy képzelünk el, hogy a grafikonja folyamatos vonallal lerajzolható. Ezt fogalmazza meg Bolzano tétele.

3.13 Tétel. **(Bolzano-tétel)** *Legyen f folytonos függvény az $[a, b]$ intervallumon, és tegyük fel, hogy $f(a)$ és $f(b)$ ellenkező előjelűek. Akkor van olyan $a < c < b$ hely, amelyre $f(c) = 0$.*

A tételt nem bizonyítjuk, de egy rövid indoklást megmutatunk. Felezzük meg az $[a, b]$ intervallumot, és válasszuk azt a felét, amelynek végpontjaiban f különböző előjelű (ha nulla lenne, akkor kész a bizonyítás). A kiválasztott részintervallumot újra megfelezzük, és újra azt a felét választjuk, amelynek végpontjaiban f ellenkező előjelű, stb...

Az eljárást folytatva egymásba skatulyázott zárt intervallumok sorozatához jutunk, ahol az n-ik $[a_n, b_n]$ részintervallum hosszára azt kapjuk, hogy

$$b_n - a_n = \frac{b-a}{2^n}\,.$$

Úgy képzeljük, hogy ezen intervallumok metszete nem üres, és nyilván csak egyetlen pontot tartalmazhat. Jelölje $c \in [a, b]$ ezt a pontot, nevezetesen

$$\bigcap_{n=1}^{\infty}[a_n, b_n] = \{c\}\,.$$

Világos, hogy nem lehet $f(c) > 0$, hiszen akkor a folytonosság miatt a c egy kis környezetében is pozitív lenne, ami ellentmond a konstrukciónak. Hasonlóan nem lehet $f(c) < 0$ sem. Tehát $f(c) = 0$.

3.14 Példa. Vizsgáljuk meg, hogy vajon megoldható-e a

$$2x^5 - 18x^4 + 3x^3 + 20x - 13 = 0$$

egyenlet? Világos, hogy a bal oldalon álló kifejezés egy folytonos f függvényt definiál, amelyre

$$\lim_{x\to+\infty} f(x) = +\infty \quad \text{és} \quad \lim_{x\to-\infty} f(x) = -\infty$$

ezért f elég nagy x értékekre pozitív, míg elég kis x értékekre negatív értéket vesz fel. Tehát Bolzano tétele szerint az egyenletnek van legalább egy valós gyöke.

A szélsőértékek és optimalizálási feladatok szempontjából alapvető jelentőségű a folytonos függvények következő tulajdonsága.

3.15 Tétel. **(Weierstrass-tétel)** *Legyen f folytonos függvény az $[a, b]$ intervallumon. Akkor f ezen az intervallumon felveszi a maximumát és a minimumát.*

Nem bizonyítunk, csak heurisztikus gondolatmenetet adunk pl. a maximum létezésére. Először gondoljuk meg, hogy $[a, b]$ intervallumon folytonos függvény felülről korlátos. Ezt indirekt módon láthatjuk be.

Másrészt úgy képzeljük (NEM NYILVÁNVALÓ!), hogy a felső korlátok között van legkisebb, jelölje ezt M. Ekkor bármely n természetes számhoz van olyan $x_n \in [a, b]$, amelyre $f(x_n) > M - 1/n$. Ellenkező esetben ugyanis $M - 1/n$ lenne a legkissebb felső korlát.

Ha most az x_n sorozatnak valamely részsorozata a $c \in [a, b]$ számhoz tart, akkor a folytonosság miatt $f(x_n) \to f(c)$. Továbbá

$$M - \frac{1}{n} < f(x_n) \leq M$$

ezért a Rendőr-elv miatt csak $f(c) = M$ állhat fenn.

Hasonló gondolatmenet alkalmazható minimum esetére.

Például az

$$f(x) = \begin{cases} x & \text{ha } 0 \leq x \leq 1 \\ 3 - x & \text{ha } 1 < x \leq 2 \end{cases}$$

függvény nem veszi fel a maximumát a $[0, 2]$ intervallumon, de nem is folytonos az 1 helyen.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 I/4 és I/5 szakaszok kidolgozott példáinak feldolgozása.
2. Házi feladatok: az I/4 szakasz 4.1.4, 4.1.5, 4.2.5, 4.3.6, 4.3.9, 4.4.6, 4.4.7, továbbá az I/5 szakasz 5.1.7, 5.1.8, 5.2.1 feladatai.
3. Tankönyv-1 2. és 3. fejezetek, 6.1, 6.2, 6.3, 6.6, 6.7, 7.1 és 7.2 szakaszok.

4. fejezet

Függvények deriváltja

4.1. A derivált fogalma

Legyen f valamely intervallumon értelmezett függvény, és tegyük fel, hogy x_0 az intervallum belső pontja.

4.1 Definíció. Azt mondjuk, hogy f *differenciálható* az x_0 pontban, ha létezik és véges az alábbi határérték:

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

Ezt a határértéket az f deriváltjának nevezzük az x_0 pontban, jelölése $f'(x_0)$. Az f függvényt differenciálhatónak nevezzük egy intervallumban, ha annak minden belső pontjában differenciálható.

A fenti hányadost az f függvény *különbségi hányadosának* nevezzük az x_0 pontban.

4.2 Példa. Tekintsük az $f(x) = x^2$ függvényt. Az x_0 pontbeli különbségi hányados:

$$\frac{f(x_0 + h) - f(x_0)}{h} = \frac{(x_0 + h)^2 - x_0^2}{h} = 2x_0 + h$$

amelynek határértéke $2x_0$, ha $h \to 0$. Következésképpen

$$f'(x_0) = 2x_0 \ .$$

Teljesen hasonló módon látható, hogy $f(x) = x^n$ esetén, ahol n természetes szám (használjuk a Binomiális-tételt),

$$f'(x_0) = nx_0^{n-1} \ .$$

4.3 Állítás. *Ha f differenciálható az x pontban, akkor ott folytonos is.*

Bizonyítás. Legyen $h_n \to 0$ tetszőleges sorozat, akkor a differenciálhatóság folytán

$$\lim_{n\to\infty} \frac{f(x+h_n)-f(x)}{h_n} = f'(x)\ .$$

Ez csak úgy lehetséges, ha $\lim_{n\to\infty}(f(x+h_n)-f(x)) = 0$, azaz $\lim_{n\to\infty} f(x+h_n) = f(x)$. Ez éppen azt jelenti, hogy f folytonos az x pontban. □

FIGYELEM! Az állítás megfordítása általában nem igaz, amint azt a következő példa mutatja.

4.4 Példa. Tekintsük az $f(x) = |x|$ függvényt a számegyenesen, és vizsgáljuk meg az $x_0 = 0$ pontbeli különbségi hányadost. Világos, hogy

$$\frac{f(h)-f(0)}{h} = \frac{|h|}{h} = \begin{cases} 1 & \text{ha } h > 0 \\ -1 & \text{ha } h < 0 \end{cases}$$

ezért a határérték $h \to 0$ esetén nem létezik, hiszen a jobb oldali határérték $+1$, míg a bal oldali határérték -1. Tehát az f függvény a 0 pontban nem differenciálható.

Minden más pontban azonban igen, nevezetesen $f'(x) = 1$, ha $x > 0$, és $f'(x) = -1$, ha $x < 0$.

4.2. Görbék érintője

A geometriai interpretáció azt mutatja, hogy az $f'(x_0)$ derivált az f függvény grafikonjához az x_0 pontban húzott érintő meredekségét jelenti.

Ennek alapján meghatározhatjuk egy differenciálható függvény grafikonjához az x_0 pontban húzható érintő egyenletét:

$$y = f'(x_0)(x - x_0) + f(x_0)\ .$$

Például az $f(x) = x^3$ függvény grafikonjához az $x_0 = 1$ pontban húzható érintő egyenlete:

$$y = 3(x-1) + 1$$

4.5 Példa. Állítsuk elő az $f(x) = \sin x$ függvény grafikonjához az $x_0 = 0$ pontban húzott érintő egyenletét. Ekkor az érintő egyrészt átmegy az origón, másrészt a meredeksége:

$$f'(0) = \lim_{h\to 0} \frac{f(h)-f(0)}{h} = \lim_{h\to 0} \frac{\sin h}{h} = 1\,.$$

Ezért az érintő egyenlete $y = x$, amely az origóban átmetszi a grafikont.

4.3. Differenciálási szabályok

Tekintsük az f és g függvényeket, amelyek egyaránt differenciálhatók az x pontban. A határérték tulajdonságaiból adódnak az alábbi szabályok.

Összeg és számszoros deriváltja Ha α és β tetszőleges valós számok, akkor $\alpha f(x) + \beta g(x)$ is differenciálható az x pontban és

$$(\alpha f(x) + \beta g(x))' = \alpha f'(x) + \beta g'(x) \ ,$$

Szorzat deriváltja $f(x) \cdot g(x)$ is differenciálható és

$$(f(x) \cdot g(x))' = f'(x) \cdot g(x) + f(x) \cdot g'(x) \ ,$$

Hányados deriváltja ha $g(x) \neq 0$, akkor $f(x)/g(x)$ is differenciálható és

$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2} \ .$$

Példaként tekintsük a szorzat differenciálási szabályának igazolását.

$$\frac{f(x+h) \cdot g(x+h) - f(x) \cdot g(x)}{h} =$$

$$\frac{f(x+h) \cdot g(x+h) - f(x+h) \cdot g(x)}{h} +$$

$$\frac{f(x+h) \cdot g(x) - f(x) \cdot g(x)}{h} =$$

$$f(x+h)\frac{g(x+h) - g(x)}{h} + g(x)\frac{f(x+h) - f(x)}{h}$$

Itt az első tört határértéke $f(x)g'(x)$ az f folytonossága miatt, míg a második tört $f'(x)g(x)$-hez tart, ha $h \to 0$. Innen adódik az állítás. A többi szabály igazolása hasonlóan történik.

4.6 Példa. Mutassuk meg, hogy az $f(x) = 1/x$ függvény grafikonjának bármely pontjában húzott érintő a koordináta-tengelyekkel azonos területű háromszöget zár be.

A szimmetria miatt nyilván elég $x_0 > 0$ koordinátájú pontokra szorítkozni. A hányados deriválási szabálya alapján

$$f'(x_0) = -\frac{1}{x_0^2}$$

ezért az x_0 pontban húzható érintő egyenlete

$$y = -\frac{1}{x_0^2}(x - x_0) + \frac{1}{x_0}$$

Ennek az egyenesnek a tengelymetszetei:

$$x = 0 \text{ esetén az } y\text{-tengelyen } b = 2/x_0$$

illetve

$$y = 0 \text{ esetén az } x\text{-tengelyen } a = 2x_0.$$

Tehát a közbezárt háromszög területe

$$T = \frac{1}{2}ab = \frac{1}{2} \cdot 2x_0 \cdot \frac{2}{x_0} = 2$$

ami valóban független az x_0 pont választásától.

4.4. Függvények kompozíciója

Legyenek f és g egyaránt $\mathbb{R} \to \mathbb{R}$ függvények úgy, hogy g értékkészlete az f értelmezési tartományában fekszik. Ekkor az

$$x \to f(g(x))$$

hozzárendelést az f és g függvények kompozíciójának nevezzük. Jelölése $f \circ g$, azaz

$$f \circ g(x) = f(g(x)) \; .$$

Ha például $f(x) = \sqrt{x}$ és $g(x) = 1 + x^2$, akkor

$$f \circ g(x) = \sqrt{1 + x^2} \; .$$

Figyelem, a sorrend fontos!

Általában $f \circ g \neq g \circ f$. Ha a fenti példát tekintjük, akkor

$$g \circ f(x) = 1 + x$$

de ez a függvény csak $x \geq 0$ esetén van értelmezve!

Az is előfordulhat, hogy $f \circ g$ az egész nemnegatív félegyenesen értelmezve van, de $g \circ f$ nem is definiálható. Ha például

$$f(x) = -1 - x^4 \quad \text{ és } \quad g(x) = \sqrt{x} \; ,$$

akkor $f \circ g(x) = -1 - x^2$, ha $x \geq 0$, de $g \circ f(x) = \sqrt{-1 - x^4}$ egyetlen valós számra sincs értelmezve.

4.5. Láncszabály

A kompozíció-függvény differenciálhatóságáról szóló tételünk igen erős eszköz bonyolultabb függvények deriváltjának előállításhoz.

4.7 Tétel. (Láncszabály) *Tegyük fel, hogy g differenciálható az x pontban, és f differenciálható a $g(x)$ pontban, akkor $f \circ g$ is differenciálható az x pontban, és*

$$(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$$

Ha bevezetjük a $k = g(x+h) - g(x)$ jelölést, akkor az $f \circ g$ függvény különbségi hányadosa az x helyen a következő módon írható:

$$\frac{f(g(x+h)) - f(g(x))}{h} =$$

$$\frac{f(g(x)+k) - f(g(x))}{k} \cdot \frac{g(x+h) - g(x)}{h}$$

amennyiben $g(x+h) - g(x) \neq 0$. Ilyenkor $h \to 0$ esetén a g folytonossága miatt $k \to 0$, és így a jobb oldali kifejezés határértéke

$$f'(g(x)) \cdot g'(x)$$

Ez a gondolat nem működik akkor, ha $k = 0$. Ekkor a bizonyítás egy kicsit komplikáltabb, ezt nem részletezzük.

4.8 Példa. Legyen például

$$F(x) = (1 + 3x - x^2)^6 \ .$$

Ekkor a derivált a hatványozás elvégzése nélkül előállítható, ha észrevesszük, hogy az $f(x) = x^6$ és $g(x) = 1 + 3x - x^2$ jelölésekkel $F = f \circ g$. Tehát a tételünk szerint:

$$F'(x) = 6(1 + 3x - x^2)^5 \cdot (3 - 2x) \ .$$

4.9 Példa. Legyen most

$$F(x) = \left(\frac{2x+3}{5+x^2}\right)^3 \quad x \in \mathbb{R}$$

Ekkor a

$$g(x) = \frac{2x+3}{5+x^2} \quad \text{és} \quad f(x) = x^3$$

jelölésekkel $F = f \circ g$. Vegyük figyelembe, hogy a g hányadosként áll elő, így

$$F'(x) = f'(g(x)) \cdot g'(x) = 3\left(\frac{2x+3}{5+x^2}\right)^2 \cdot \frac{2(5+x^2) - 2x(2x+3)}{(5+x^2)^2}$$

amelyet még valamivel egyszerűbb alakra hozhatunk.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 I/3 és I/6 szakaszok kidolgozott példáinak feldolgozása.
2. Házi feladatok: az I/3 szakasz 3.1.4, 3.1.5, továbbá az I/6 szakasz 6.1.2, 6.1.4, 6.2.3, 6.2.7, 6.2.9 feladatai.
3. Tankönyv-1 4. fejezet, 5.2 és 5.6 szakaszok.

5. fejezet

A középérték-tétel

5.1. Az inverz függvény

Tekintsünk egy $f : \mathbb{R} \to \mathbb{R}$ függvényt, amely kölcsönösen egyértelmű valamely intervallumon. Ez például egy folytonos függvényre azt jelenti, hogy f vagy szigorúan monoton növő, vagy szigorúan monoton fogyó.

5.1 Definíció. Az f függvény inverzén azt az f^{-1} függvényt értjük, amelynek értelmezési tartománya az f értékkészlete, és értékkészlete az f értelmezési tartománya, továbbá

$$f^{-1} \circ f(x) = x$$

az f értelmezési tartományának minden pontjában.

Ezt a „fordított" hozzárendelést úgy állíthatjuk elő, hogy az

$$y = f(x)$$

egyenlőségből kifejezzük x-et y függvényeként:

$$x = f^{-1}(y) .$$

Ha például $f(x) = (2x + 5)^3$, akkor világos, hogy

$$f^{-1}(y) = \frac{\sqrt[3]{y} - 5}{2} .$$

Világos, hogy f^{-1} grafikonja és f grafikonja az $y = x$ egyenesre nézve tükrös helyzetűek.

5.2. Az inverz függvény differenciálhatósága

5.2 Tétel. *Tegyük fel, hogy f folytonos, szigorúan monoton valamely intervallumon, differenciálható annak valamely x belső pontjában és $f'(x) \neq 0$. Akkor f^{-1} is differenciálható az $y = f(x)$ pontban, és*

$$(f^{-1})'(y) = \frac{1}{f'(x)}\,.$$

Vázlatosan a következőről van szó. Tekintsük a különbségi hányadost:

$$\frac{f^{-1}(y+h) - f^{-1}(y)}{h}$$

Legyenek x és $x+k$ olyan pontok az f értelmezési tartományából, amelyekre $y = f(x)$ és $y+h = f(x+k)$. Akkor a különbségi hányados úgy írható, hogy

$$\frac{x+k-x}{f(x+k)-f(x)} = \frac{1}{\frac{f(x+k)-f(x)}{k}}$$

Ha itt $h \to 0$, akkor $k \to 0$ (FIGYELEM, nem nyilvánvaló!), és így a jobb oldali tört határértéke valóban $1/f'(x)$.

5.3 Példa. Határozzuk meg a

$$g(x) = \sqrt[n]{x}$$

függvény deriváltját valamely $x > 0$ pontban. Ekkor g éppen az $f(x) = x^n$ hatványfüggvény inverze a pozitív félegyenesen, azaz $g(y) = f^{-1}(y)$. Ezért

$$g'(y) = \frac{1}{f'(x)} = \frac{1}{nx^{n-1}} = \frac{1}{n} \cdot y^{\frac{1}{n}-1}$$

hiszen $y = x^n$ és így

$$x^{n-1} = y^{\frac{n-1}{n}}$$

Ezen példa alapján könnyen láthatjuk, hogy bármely r racionális kitevő esetén az $F(x) = x^r$ függvény bármely $x > 0$ pontban differenciálható, és

$$F'(x) = rx^{r-1}\,.$$

5.4 Példa. Állítsuk elő az

$$F(x) = \sqrt{1+x^4}$$

függvény deriváltját. Legyen $f(x) = \sqrt{x}$ és $g(x) = 1 + x^4$, ezekkel a jelölésekkel $F = f \circ g$. Ezért

$$F'(x) = f'(g(x)) \cdot g'(x) = \frac{4x^3}{2\sqrt{1+x^4}}$$

5.3. Az exponenciális és a logaritmus függvény

Tekintsük a számegyenesen az e alapú exponenciális függvényt, és annak inverzét, az e alapú logaritmus függvényt (ennek jelölésére az ln jel használatos):

$$f(x) = e^x \qquad f^{-1}(x) = \ln x \quad (x > 0) \ .$$

Ezeket természetes alapú exponenciális, illetve logaritmus függvénynek nevezzük. Ezen függvények deriváltjait szeretnénk előállítani. Ehhez felhasználjuk, hogy

$$\lim_{x\to\pm\infty} \left(1 + \frac{1}{x}\right)^x = e \ .$$

Határozzuk meg a természetes alapú logaritmus függvény deriváltját az $x_0 = 1$ helyen.

$$\frac{\ln(1+h) - \ln 1}{h} = \ln(1+h)^{1/h}$$

amelynek (feltételezve a logaritmus függvény folytonosságát) jobb és bal oldali határértéke egyaránt $\ln e$ a 0 helyen. Tehát a derivált értéke 1.

Az $f(x) = e^x$ függvény deriváltját a 0 helyen az inverz függvény deriváltjára vonatkozó tételünk alapján határozhatjuk meg:

$$f'(0) = \lim_{h\to 0} \frac{e^h - 1}{h} = \frac{1}{(\ln)'(1)} = 1 \ .$$

Innen már könnyen megkapjuk az exponenciális függvény deriváltját tetszőleges x pontban:

$$f'(x) = \lim_{h\to 0} \frac{e^{x+h} - e^x}{h} = e^x \cdot \lim_{h\to 0} \frac{e^h - 1}{h} = e^x$$

Ismét az inverz függvény deriválási szabályát alkalmazva adódik a logaritmus függvény deriváltja tetszőleges $x > 0$ pontban:

$$(f^{-1})'(x) = \frac{1}{e^{\ln x}} = \frac{1}{x} \ .$$

5.5 Példa. Alkalmazásképpen állítsuk elő az

$$f(x) = x^\alpha$$

általános hatványfüggvény deriváltját valamely $x > 0$ pontban, ahol α tetszőleges valós szám. Világos, hogy

$$f(x) = x^\alpha = e^{\alpha \ln x}$$

és így a kompozíció függvény deriválási szabálya folytán:

$$f'(x) = \alpha \frac{1}{x} e^{\alpha \ln x} = \alpha \frac{1}{x} x^\alpha = \alpha x^{\alpha - 1}$$

Ez azt jelenti, hogy a deriválás ugyanúgy végezhető el, mint racionális kitevő esetében.

5.4. A szélsőérték szükséges feltétele

Tekintsünk egy $f : \mathbb{R} \to \mathbb{R}$ függvényt.

5.6 Definíció. Azt mondjuk, hogy az értelmezési tartomány valamely x_0 pontja az f (globális) minimumhelye, ha $f(x_0) \leq f(x)$ az értelmezési tartomány bármely $x \neq x_0$ pontjára.

Azt mondjuk, hogy az értelmezési tartomány valamely x_0 pontja az f lokális minimumhelye, ha létezik olyan $\varepsilon > 0$ szám, hogy $f(x_0) \leq f(x)$ az értelmezési tartomány bármely olyan x pontjára, amelyre $0 < |x - x_0| < \varepsilon$.

Mindkét esetben szigorú minimumhelyről beszélünk, ha szigorú egyenlőtlenség teljesül.

Analóg definíció érvényes maximum esetében.

Nyilvánvaló, hogy egy globális minimumhely egyben lokális minimumhely is, a megfordítás azonban nem érvényes. Például az

$$f(x) = \begin{cases} (x+1)^2 & \text{ha } x < 0 \\ (x-1)^2 & \text{ha } x \geq 0 \end{cases}$$

függvénynek az $x = 0$ pontban lokális maximumhelye van (itt a függvény folytonos, de nem differenciálható, ellenőrizzük!), de a függvénynek nincs is globális maximumhelye, hiszen felülről nem korlátos.

Differenciálható függvények esetében a szélsőértékhely következő karakterizációját adhatjuk.

5.7 Tétel. *Tegyük fel, hogy f egy intervallumon értelmezett függvény, és ennek valamely x_0 belső pontjában f differenciálható. Ha x_0 az f lokális minimumhelye, akkor $f'(x_0) = 0$.*

Valóban, tekintsük a különbségi hányadost:

$$\frac{f(x_0+h) - f(x_0)}{h}.$$

Ha $h > 0$, akkor a különbségi hányados a h elég kicsi értékeire nem negatív, így a jobb oldali határérték is nem negatív. Másrészt ha $h < 0$, akkor hasonlóképpen a különbségi hányados nem pozitív, ezért a bal oldali határérték is nem pozitív. A differenciálhatósági feltétel folytán azonban a különbségi hányadosnak létezik határértéke, ami így csak nulla lehet. Tehát $f'(x_0) = 0$.

Ez a tételünk a szélsőérték szükséges feltételét fogalmazza meg, amely azonban nem elégséges. Például az $f(x) = x^3$ függvénynek nincs szélsőértékhelye, de $f'(0) = 0$.

Egy differenciálható függvény esetében azon x_0 pontokat, amelyben $f'(x_0) = 0$, a függvény kritikus pontjainak nevezzük.

Ezzel a szóhasználattal az értelmezési tartomány belső pontjai között minden szélsőértékhely kritikus pont, de ennek megfordítása nem érvényes.

5.5. Lagrange-féle középérték-tétel

A geometriai interpretáció alapján szemléletes állítást fogalmaz meg a Lagrange-féle középérték-tétel:

5.8 Tétel. *Legyen f folytonos az $[a,b]$ véges, zárt intervallumon és differenciálható az intervallum belsejében. Akkor található olyan $\xi \in (a,b)$ pont, amelyre*

$$f'(\xi) = \frac{f(b)-f(a)}{b-a}$$

Bizonyítás. Tekintsük ugyanis a

$$g(x) = f(x) - f(a) - \frac{f(b)-f(a)}{b-a}(x-a)$$

függvényt. Ez a függvény a feltételünk szerint folytonos az $[a,b]$ intervallumon, így Weierstrass tétele értelmében felveszi a minimumát és a maximumát az $[a,b]$ intervallumon. A minimum és a maximumhely közül legalább az egyik létezik az intervallum belsejében, hiszen

$$g(a) = g(b) = 0\ .$$

Ha ez a szélsőértékhely $\xi \in (a,b)$, akkor az előző tételünk értelmében $g'(\xi) = 0$. Ez éppen azt jelenti, hogy

$$f'(\xi) - \frac{f(b)-f(a)}{b-a} = 0\ .$$

Vegyük észre, hogy a tételünkben a folytonossági feltevés lényeges, készítsünk ábrát!

5.6. L'Hôpital-szabály

Az alábbi eljárás megkönnyíti viszonylag komplikált határértékek kiszámítását.

Legyenek f és g differenciálhatóak, továbbá f' és g' folytonosak az x_0 pont egy környezetében, és tegyük fel, hogy $f(x_0) = g(x_0) = 0$. Az alábbi határértéket szeretnénk meghatározni:

$$\lim_{x\to x_0} \frac{f(x)}{g(x)}$$

amely $0/0$, tehát „határozatlan” alakú.

A Lagrange-féle középérték-tétel szerint:

$$\frac{f(x)}{g(x)} = \frac{\frac{f(x)-f(x_0)}{x-x_0}}{\frac{g(x)-g(x_0)}{x-x_0}} = \frac{f'(\xi)}{g'(\eta)}$$

ahol ξ és η valamely x és x_0 között fekvő pontok. Ha most $x \to x_0$, akkor egyaránt $\xi \to x_0$ és $\eta \to x_0$. Tehát a deriváltak folytonossága alapján

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{x \to x_0} \frac{f'(x)}{g'(x)}$$

Ezt az egyenlőséget nevezzük L'Hôpital-szabálynak. Ha az így nyert határérték még mindig 0/0 alakú, akkor alkalmazzuk L'Hôpital'-szabályt újra, amíg egy "egyszerű" határértéket nem kapunk.

5.9 Példa. Tekintsük például a

$$\lim_{x \to 0} \frac{2 \sin x}{1 - \sqrt{1+x}}$$

határértéket. Ekkor a L'Hôpital-szabály szerint:

$$\lim_{x \to 0} \frac{2 \sin x}{1 - \sqrt{1+x}} = \frac{2 \cos 0}{-\frac{1}{2\sqrt{1+0}}} = -4$$

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 I/3, I/6 és I/7 szakaszok kidolgozott példáinak feldolgozása.
2. Házi feladatok: az I/3 szakasz 3.2.3, 3.2.4, 3.2.5, továbbá az I/6 szakasz 6.3.5, 6.3.6, 6.3.10, továbbá az I/7 szakasz 7.1.6, 7.1.8 és 7.1.9 feladatai
3. Tankönyv-1 5.1, 5.4, 7.5 és 7.6 szakaszok, 8. fejezet.

6. fejezet

A teljes függvényvizsgálat

6.1. Monoton függvények

6.1 Definíció. Azt mondjuk, hogy f valamely intervallumon monoton növő, ha az intervallum bármely két $x_1 < x_2$ pontjára $f(x_1) \leq f(x_2)$. Analóg a monoton fogyó függvény értelmezése.

Szigorú monotonitásról beszélünk, ha az utóbbi egyenlőtlenség szigorú formában teljesül.

6.2 Tétel. *Legyen f folytonos az $[a,b]$ véges, zárt intervallumon és differenciálható az intervallum belsejében. Ha minden belső pontban $f'(x) > 0$, akkor f az $[a,b]$ intervallumon szigorúan monoton növő.*

Valóban, ha $x_1 < x_2$ az $[a,b]$ intervallum két tetszőleges pontja, akkor a Lagrange-féle középérték-tétel szerint található olyan $x_1 < \xi < x_2$ pont, amelyre

$$f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1) \ .$$

A feltételünk szerint a jobb oldalon álló kifejezés pozitív, ezért

$$f(x_2) - f(x_1) > 0$$

azaz f valóban szigorúan monoton növő.

Vizsgáljunk most meg egy intervallumon monoton növő, differenciálható függvényt. Könnyen látható, hogy az intervallum bármely két különböző x és $x+h$ pontjára fennáll

$$\frac{f(x+h) - f(x)}{h} \geq 0$$

tehát a $h \to 0$ határértékre térve $f'(x) \geq 0$. Ezt az észrevételt az előző tételünkkel egybevetve a következő állítást fogalmazhatjuk meg:

6.3 Tétel. *Legyen f folytonos az $[a, b]$ intervallumon és differenciálható az intervallum belsejében. Az f függvény akkor és csak akkor monoton növő az intervallumon, ha $f'(x) \geq 0$ az intervallum minden belső pontjában.*

Hasonló állítás érvényes a monoton fogyó esetben is.

Az azonban nem igaz, hogy ha f szigorúan monoton növő, akkor $f'(x) > 0$ lenne minden x pontban. Például az $f(x) = x^3$ függvény az egész számegyenesen szigorúan monton növő, de $f'(0) = 0$.

6.2. A szélsőértékhely megkeresése

Tekintsünk egy $f : \mathbb{R} \to \mathbb{R}$ függvényt, és legyen x_0 az értelmezési tartomány belső pontja. Tegyük fel, hogy f differenciálható az x_0 pontban.

Amint láttuk, annak szükséges feltétele, hogy x_0 szélsőértékhely legyen, az, hogy $f'(x_0) = 0$. Kérdés, hogy milyen elégséges feltételt fogalmazhatunk meg a szélsőérték létezésére. Világos, hogy ha található olyan $\varepsilon > 0$ szám, amelyre f monoton fogyó az $[x_0 - \varepsilon, x_0]$ intervallumon, továbbá f monoton növő az $[x_0, x_0 + \varepsilon]$ intervallumon, akkor x_0 az f lokális minimumhelye.

Differenciálható függények esetében ezt az észrevételünket az alábbi tételben fogalmazzuk meg.

6.4 Tétel. *Tegyük fel, hogy f differenciálható egy intervallumban és x_0 az intervallum olyan belső pontja. Ha van olyan $\varepsilon > 0$ szám, hogy*

- $f'(x) \leq 0$, *ha* $x \in (x_0 - \varepsilon, x_0)$
- $f'(x) \geq 0$, *ha* $x \in (x_0, x_0 + \varepsilon)$

akkor x_0 az f lokális minimumhelye.

Nyilván analóg állítás érvényes lokális maximumhely esetére is.

6.5 Példa. Keressük meg az

$$f(x) = x^2 e^{-x}$$

függvény szélsőértékeit és monotonitási szakaszait. Könnyen látható, hogy

$$f'(x) = (2x - x^2)e^{-x}$$

amelynek előjele csak az első tényezőtől függ. Ennek alapján a következőt kapjuk.

- Ha $x \in (-\infty, 0)$, akkor $f'(x) < 0$, ezért itt f monoton fogyó
- Ha $x = 0$, akkor $f'(0) = 0$, ez kritikus pont
- Ha $x \in (0, 2)$, akkor $f'(x) > 0$, ezért itt f monoton növő
- Ha $x = 2$, akkor $f'(2) = 0$, ez kritikus pont
- Ha $x \in (2, +\infty)$, akkor $f'(x) < 0$, ezért itt f monoton fogyó

Az f' előjelváltásai alapján láthatjuk, hogy az $x = 0$ pont (globális) minimumhely, míg $x = 2$ lokális maximumhely.

6.6 Példa. Tekintsük például a számegyenesen az

$$f(x) = x + \sin x$$

függvényt. Mivel $f'(x) = 1 + \cos x$, világos, hogy a függvénynek az

$$x = (2k+1)\pi \qquad k = 0, \pm 1, \pm 2, \ldots$$

helyeken kritikus pontjai vannak. Ezek közül azonban egyikben sem lehet szélsőérték, ugyanis

$$x \neq (2k+1)\pi \quad \text{esetén} \quad f'(x) > 0\ ,$$

hiszen ezekben a pontokban $\cos x > -1$.

Ez a függvény tehát az egész számegyenesen szigorúan monoton növő.

6.3. Magasabbrendű deriváltak

Ha egy f függvény egy intervallumban mindenütt differenciálható, akkor az $x \to f'(x)$ hozzárendelést az f derivált-függvényének nevezzük. Ha f' differenciálható valamely x_0 pontban, akkor azt mondjuk, hogy f itt kétszer differenciálható. Ilyenkor az $(f')'(x_0)$ jelölés helyett az

$$f''(x_0)$$

jelölést használjuk, és ezt az f második deriváltjának nevezzük az x_0 pontban.

Teljesen hasonlóan tetszőleges n természetes szám esetén beszélhetünk az f n-ik deriváltjáról az x_0 pontban, ennek jelölése

$$f^{(n)}(x_0)\,.$$

Például az $f(x) = 1/x$ függvényre valamely $x_0 \neq 0$ pontban

$$f''(x_0) = \frac{2}{x_0^3} \qquad \text{valamint} \qquad f^{(n)}(x_0) = \frac{(-1)^n n!}{x^{n+1}}$$

minden n természetes számra.

6.7 Példa. Tekintsük az $f(x) = \sin x$ függvényt, és állítsuk elő a derivált-függvényét. Egyrészt

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{\sin(x+h) - \sin x}{h} = \lim_{h\to 0} \frac{\sin x \cos h + \cos x \sin h - \sin x}{h} \\ &= \sin x \cdot \lim_{h\to 0} \frac{\cos h - 1}{h} + \cos x \cdot \lim_{h\to 0} \frac{\sin h}{h} \end{aligned}$$

Itt a 3.12 Példa alapján az első határérték 0, míg a 3.8 Példa szerint a második határérték 1. Eszerint

$$f'(x) = \cos x$$

Teljesen hasonló eljárással mutatható meg, hogy

$$(\cos x)' = -\sin x$$

Tehát az $f(x) = \sin x$ függvény magasabbrendű deriváltjaira az alábbi eredmény adódik:

$$f^{(n)}(x) = \begin{cases} \cos x & \text{ha } n \text{ 4-gyel osztva 1 maradékot ad} \\ -\sin x & \text{ha } n \text{ 4-gyel osztva 2 maradékot ad} \\ -\cos x & \text{ha } n \text{ 4-gyel osztva 3 maradékot ad} \\ \sin x & \text{ha } n \text{ 4-gyel osztható} \end{cases}$$

6.4. Másodrendű feltételek

Előfordulhat, hogy olyan bonyolult függvényt kell vizsgálunk, amelynél a derivált előjelét nehéz meghatározni. Ilyen esetekben bizonyulhat hasznosnak a szélsőérték másodrendű (elégséges) feltétele.

6.8 Tétel. *Legyen f differenciálható valamely intervallumban, és tegyük fel, hogy az intervallum valamely x_0 belső pontjában f kétszer differenciálható.*

Ha $f'(x_0) = 0$ és $f''(x_0) > 0$, akkor x_0 az f lokális minimumhelye.

Valóban, a különbségi hányados vizsgálata alapján

$$\begin{aligned} f''(x_0) &= \lim_{h\to 0} \frac{f'(x_0+h) - f'(x_0)}{h} = \\ &= \lim_{h\to 0} \frac{f'(x_0+h)}{h} > 0 \end{aligned}$$

és ez azt jelenti, hogy az $f'(x_0+h)/h$ hányados pozitív, ha $0 < |h| < \varepsilon$ valamely $\varepsilon > 0$ mellett. Innen adódik, hogy

- ha $x \in (x_0 - \varepsilon, x_0)$, akkor $f'(x) < 0$,
- ha $x \in (x_0, x_0 + \varepsilon)$, akkor $f'(x) > 0$.

Tehát a 6.4 Tételünk alapján x_0 valóban lokális minimumhely.

Természetesen analóg elégséges feltételt fogalmazhatunk meg a lokális maximumhely esetére is.

Indirekt meggondolással azonnal kapjuk a szélsőérték másodrendű szükséges feltételét is.

6.9 Tétel. *Tegyük fel, hogy f kétszer differenciálható egy intervallumban, és legyen x_0 az intervallum valamely belső pontja.*

- *Ha x_0 lokális minimumhely, akkor $f'(x_0) = 0$, és $f''(x_0) \geq 0$.*
- *Ha x_0 lokális maximumhely, akkor $f'(x_0) = 0$, és $f''(x_0) \leq 0$.*

6.10 Példa. Legyen például $x > 0$ esetén

$$f(x) = x \ln x$$

Ekkor $f'(x) = 1 + \ln x$, tehát az f függvény egyetlen kritikus pontja $x = 1/e$. Mivel $f''(x) = 1/x$, ezen a helyen

$$f''(1/e) = e > 0 ,$$

tehát az f függvénynek az $x = 1/e$ helyen lokális minimumhelye van. Nem nehéz belátni, hogy ez egyben globális minimum is.

Vegyük észre, hogy az eddigi tételeink nem tartalmaznak információt egy olyan x_0 kritikus pont esetében, ahol

$$f''(x_0) = 0 .$$

Ennek az az oka, hogy ebben a „határesetben” bármi előfordulhat. Vizsgáljuk meg ugyanis az

$$f(x) = x^n \qquad (n \geq 3)$$

hatványfüggvény viselkedését az $x_0 = 0$ kritikus pontban. Világos, hogy itt $f'(0) = 0$ és $f''(0) = 0$. Másrészt

- ha n páros, akkor $x_0 = 0$ (globális) minimumhely,
- ha n páratlan, akkor $x_0 = 0$ nem szélsőértékhely.

Teljesen hasonlóan páros n esetén $x_0 = 0$ a $-f$ függvénynek (globális) maximumhelye.

6.5. Konvex és konkáv függvények

6.11 Definíció. Az f függvényt konvexnek nevezzük az $[a, b]$ intervallumon, ha az intervallum tetszőleges x_1 és x_2 pontjaira és bármely $0 \leq \alpha \leq 1$ valós számra

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2) .$$

Ez geometriailag azt jelenti, hogy a függvény grafikonjához rajzolt húr sehol sem lehet a függvény grafikonja alatt.

Konkáv függvényeket a fordított irányú egyenlőtlenséggel értelmezzük.

Kétszer differenciálható függvényekre a konvexitásnak egy igen egyszerű, és geometriailag is világos jellemzését adhatjuk.

6.12 Tétel. *Tegyük fel, hogy f folytonos valamely intervallumon és kétszer differenciálható az intervallum belsejében. Annak szükséges és elégséges feltétele, hogy f konvex az intervallumon:*

$$f''(x) \geq 0$$

az intervallum minden belső pontjában.

Ez azt jelenti, hogy konvex függvény esetében az érintő meredeksége monoton növő. Ezt úgy is szemléltethetjük, hogy egy konvex függvény esetében a függvény grafikonja sehol sincs az érintő alatt.

6.13 Példa. Vizsgáljuk meg az

$$f(x) = \frac{x}{1+x^2}$$

függvényt. Könnyen ellenőrizhetjük, hogy

$$f'(x) = \frac{1-x^2}{(1+x^2)^2} .$$

A derivált előjelének vizsgálata alapján

- f szigorúan monoton fogyó a $(-\infty, -1)$ intervallumon
- $x = -1$ (globális) minimumhely
- f szigorúan monoton növő a $(-1, 1)$ intervallumon
- $x = 1$ (globális) maximumhely

- f szigorúan monoton fogyó a $(1, +\infty)$ intervallumon

A konvexitást a második derivált előjele alapján vizsgálhatjuk:

$$f''(x) = \frac{2x^3 - 6x}{(1 + x^2)^3}$$

Világos, hogy a nevező minden pontban pozitív, ezért elég a

$$2x^3 - 6x = 2x(x^2 - 3)$$

számláló előjelét tekinteni. Tehát

- f konkáv a $(-\infty, -\sqrt{3})$ intervallumon
- f konvex a $(-\sqrt{3}, 0)$ intervallumon
- f konkáv a $(0, \sqrt{3})$ intervallumon
- f konvex a $(\sqrt{3}, +\infty)$ intervallumon

Vegyük észre, hogy $f''(-\sqrt{3}) = f''(0) = f''(\sqrt{3}) = 0$, és ezekben a pontokban a második derivált előjelet vált. Ezek a pontok tehát az f függvény konvex és konkáv szakaszait választják el. Az ilyen pontokat az f **inflexiós pontjainak** nevezzük.

Inflexiós pontban az érintő átmetszi a függvény grafikonját.

A konvex függvények egyik legfontosabb tulajdonsága, hogy a lokális és a globális minimumhely fogalma egybeesik.

6.14 Tétel. *Tekintsünk egy f kétszer differenciálható konvex függvényt egy intervallumon, és legyen x_0 az intervallum egy belső pontja. Ha x_0 lokális minimumhely, akkor globális minimumhely is.*

Valóban, a feltételeink szerint $f'(x_0) = 0$, másrészt a konvexitás miatt f' monoton növő. Ezért az intervallum belső pontjaiban

- $x < x_0$ esetén $f'(x) \leq 0$,
- $x > x_0$ esetén $f'(x) \geq 0$.

Innen az f' monotonitása miatt adódik az állítás.

Természetesen most is analóg állítást fogalmazhatunk meg konkáv függvények maximumhelyére vonatkozóan.

6.15 Példa. Tekintsük az a valós paraméterrel megadott

$$f(x) = ax + 2\ln x$$

függvényt az $x > 0$ pozitív félegyenesen. Az a paraméter milyen értékére lesz az f függvénynek globális maximuma az $x = 6$ pontban?

A szélsőérték szükséges feltétele alapján

$$f'(x) = a + \frac{2}{x} = 0$$

ahonnan $x = -2/a$. Innen a feltételünk szerint $a = -1/3$. Mivel az f második deriváltja

$$f''(x) = -\frac{1}{x^2} < 0\,,$$

azért a függvény (szigorúan) konvkáv az egész értelmezési tartományban, ezért $a = -1/3$ esetén az f függvénynek az $x = 6$ pontban globális maximumhelye van.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 I/6 és I/7 szakaszok kidolgozott példáinak feldolgozása.
2. Házi feladatok: az I/6 szakasz 6.6.3, 6.6.4, 6.6.6, 6.6.8, továbbá az I/7 szakasz 7.3.4, 7.3.7, 7.3.8 feladatai.
3. Tankönyv-1 9. fejezet.

7. fejezet

Integrálás

7.1. A határozatlan integrál fogalma

7.1 Definíció. Legyen f valamely I intervallumon értelmezett függvény. Azt mondjuk, hogy az I intervallumon értelmezett differenciálható F függvény az f *határozatlan integrálja*, vagy más elnevezéssel *primitív függvénye*, ha

$$F'(x) = f(x)$$

minden $x \in I$ esetén.

Világos, hogy a határozatlan integrál a deriválás fordított művelete. A határozatlan integrál fogalma azonban nem egyértelmű! Ha ugyanis egy F függvény az f határozatlan integrálja, akkor ahhoz bármely C konstanst adva ugyancsak határozatlan integrálhoz jutunk. Valóban:

$$(F(x) + C)' = F'(x) = f(x)$$

minden x pontban.

Beláthatjuk, hogy más típusú határozatlan integrál nem is létezhet.

7.2 Tétel. *Ha F az f határozatlan integrálja az I intervallumon, akkor az f minden határozatlan integrálja $F + C$ alakú valamely C konstans mellett.*

Bizonyítás. Valóban, ha a G differenciálható függvény is az f határozatlan integrálja az I intervallumon, akkor minden $x \in I$ pontban

$$(F(x) - G(x))' = f(x) - f(x) = 0$$

Ez azt jelenti, hogy $F - G$ deriváltja az I intervallumon zérus, ezért a Lagrange-féle középérték-tétel miatt $F - G$ konstans ezen az intervallumon.

A fenti tételünknek megfelelően a következő jelölésmódot használjuk a határozatlan integrálokra:

$$\int f(x)\,dx = F(x) + C$$

Például közvetlen deriválással ellenőrizhető, hogy

$$\int \cos x\,dx = \sin x + C$$

vagy teljesen hasonlóan

$$\int x^\alpha\,dx = \frac{x^{\alpha+1}}{\alpha+1} + C \qquad (\alpha \neq -1)$$

tetszőleges C konstans mellett. Ez azt mutatja, hogy ha egy függvénynek egy intervallumon van határozatlan integrálja, akkor végtelen sok is van.

7.2. Alapintegrálok

Számos határozatlan integrál kiszámításakor segít az alábbi szabály:

7.3 Tétel. $\int(\alpha f(x) + \beta g(x))\,dx = \alpha \int f(x)\,dx + \beta \int g(x)\,dx$

Ez a szabály természetesen tetszőleges véges tagú összegre általánosítható.

Figyelem: Nem minden függvénynek van határozatlan integrálja! Ha például az f függvénynek olyan szakadási pontja van, amelyben a jobb és bal oldali határértékek végesek, de különbözőek, akkor nem létezhet határozatlan integrál. Az alábbi tétel hasznos elégséges feltételt fogalmaz meg.

7.4 Tétel. *Ha f folytonos az I intervallumon, akkor létezik határozatlan integrálja.*

Határozatlan integrálok meghatározására szabályokat a deriválási szabályok megfordításával nyerhetünk. Az elemi függvények deriváltjait megadó formulákat megfordítva az elemi függvények határozatlan integráljaihoz jutunk, ezeket nevezzük alapintegráloknak. Az alapintegrálokat összefoglaló listát találunk a Feladatgyűjtemény-1-ben, TANULMÁNYOZZUK RÉSZLETESEN! Például:

$$\int \sin x\,dx = -\cos x + C$$

$$\int (2x^2 - 5x + 8)\,dx = \frac{2}{3}x^3 - \frac{5}{2}x^2 + 8x + C$$

$$\int e^{2x-1}\,dx = \frac{1}{2}e^{2x-1} + C$$

$$\int \frac{2x}{1+x^2}\,dx = \ln(1+x^2) + C$$

7.3. Kezdetiérték-feladatok

Az előzőekben láttuk, hogy egy függvénynek végtelen sok határozatlan integrálja lehet, amelyek csak konstansban különbözhetnek egymástól. Ha azonban a koordináta-rendszer egy pontját rögzítjük, azon már csak egyetlen határozatlan integrál halad át.

7.5 Példa. Keressük meg azt az F függvényt, amelyre

$$F'(x) = 2e^{-x} \qquad \text{és} \qquad F(0) = 1$$

Ekkor

$$F(x) = 2\int e^{-x}\,dx = -2e^{-x} + C$$

Innen a kezdeti értékre vonatkozó feltételből $C = 3$ adódik.

7.4. Határozott integrálok

Ebben a szakaszban röviden felvázoljuk azt, ahogy Bernhard Riemann német matematikus, a göttingeni egyetem professzora értelmezte a határozott integrál fogalmát a XIX. században. Ez az elképzelés az Arkhimédesz-féle kétoldali közelítés elvén alapul, ami az emberi gondolkodás egyik meghatározó eleme (és ez volt az az eljárás, ahogy az ókori Szirakuzában Arkhimédesz meghatározta a kör területét a belülről és kívülről közelítő szabályos sokszögek területei alapján).

Legyen f folytonos függvény az $[a, b]$ véges intervallumon, és tekintsük az intervallum egy

$$a = x_0 < x_1 < \ldots < x_n = b$$

felosztását n számú részintervallumra. Minden egyes $[x_{k-1}, x_k]$ részintervallumon jelentse m_k az f függvény legkisebb, és M_k a legnagyobb értékét, amelyek Weierstrass tétele értelmében léteznek. Készítsük el az

$$s_n = \sum_{k=1}^{n} m_k(x_k - x_{k-1})$$

úgynevezett *alsó összeget*, illetve az

$$S_n = \sum_{k=1}^{n} M_k(x_k - x_{k-1})$$

felső összeget. Ezen téglalapok területeinek összegei, az f grafikonja alatti területet alulról, illetve felülről közelítik. KÉSZÍTSÜNK ÁBRÁT!

Könnyen ellenőrizhetjük, hogy újabb osztópont beiktatásával s_n nem csökkenhet, míg S_n nem nőhet. Megmutatható, hogy a felosztás sűrítésével az alsóösszegek felső határa megegyezik a felső összegek alsó határával. Riemann értelmezésében ezt a közös S értéket az f függvény határozott integráljának nevezzük az $[a,b]$ véges intervallumon, jelölésben:

$$S = \int_a^b f(x)\,dx$$

ami az f függvény grafikonja alatti (előjeles!) területet jelenti.

Megfogalmazzuk a határozott integrál néhány fontos tulajdonságát, amelyeket természetesnek látunk az integrál geometriai interpretációja alapján.

7.6 Tétel. *Legyenek f és g olyan függvények, amelyeknek létezik integrálja.*

1. *Ha $f(x) \le g(x)$ az $[a,b]$ intervallumon, akkor*

$$\int_a^b f(x)\,dx \le \int_a^b g(x)\,dx$$

2. *Speciálisan, $|\int_a^b f(x)\,dx| \le \int_a^b |f(x)|\,dx$.*
3. *Speciálisan, ha $f(x) \le M$ az $[a,b]$ intervallumon (M konstans) akkor*

$$\int_a^b f(x)\,dx \le M(b-a)$$

4. *Ha f folytonos az $[a,b]$ intervallumon, akkor található olyan $\overline{x} \in [a,b]$, amelyre $\int_a^b f(x)\,dx = f(\overline{x})(b-a)$.*
5. *Megállapodás: $\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx$ ha $a \le b$.*
6. $\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx$

Készítsünk ábrákat, amelyeken értelmezzük a fenti állításokat!

7.5. Newton-Leibniz-formula

Ebben a szakaszban megvizsgáljuk, hogy a határozott integrál hogyan számítható ki a primitív függvény ismeretében.

7.7 Tétel. **(Newton-Leibniz-formula)** *Ha F az f folytonos függvény primitív függvénye az $[a,b]$ véges intervallumon, akkor*

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

Bizonyítás. Világos, hogy az állításunk nem függ attól, hogy melyik határozatlan integrált választottuk. Ha ugyanis G is az f függvény határozatlan integrálja, akkor

$$G(x) = F(x) + C$$

alakú az $[a, b]$ intervallumon, valamely C konstans mellett a 7.2 Tétel értelmében. Tehát ekkor

$$\int_a^b f(x)\,dx = [G(x)]_a^b = G(b) - G(a) = (F(b) + C) - (F(a) + C) = F(b) - F(a)\,.$$

Másrészt tekintsük tetszőleges $x \in [a, b]$ mellett az

$$F(x) = \int_a^x f(t)\,dt$$

integrált. Ekkor $F(a) = 0$, hiszen ekkor az integrálási út hossza zérus. Elég belátnunk, hogy F az f határozatlan integrálja.

Valóban, ekkor tetszőleges $a < x < b$ és olyan $h \neq 0$ esetén, amelyre $x + h \in [a, b]$, az integrál tulajdonságai alapján található olyan $\overline{x}$ az x és az $x+h$ pontok között, amelyre

$$\frac{1}{h}(F(x+h) - F(x)) = \frac{1}{h}\int_x^{x+h} f(t)\,dt = \frac{1}{h} f(\overline{x}) \cdot h$$

Ha most $h \to 0$, akkor $\overline{x} \to x$, ezért f folytonossága miatt $f(\overline{x}) \to f(x)$, azaz

$$\lim_{h\to 0} \frac{1}{h}(F(x+h) - F(x)) = F'(x) = \lim_{h\to 0} f(\overline{x}) = f(x)$$

tehát F valóban f primitív függvénye. □

Newton és Lebniz, és a korabeli matematika fantasztikus teljesítménye volt, hogy a határozott integrál geometriai fogalmát kapcsolatba tudta hozni a derivált fogalmával a tételünkben megfogalmazott módon.

Ez a felismerés hihetlen gyors fejlődést eredményezett először a fizika és a kémia, majd némi késéssel a biológia és a közgazdaságtan kvantitatív elemzésében is. Összességében azt mondhatjuk, hogy meghatározó szerepe volt a precíz tudományos nyelvezet létrejöttében minden tudományterületen.

A fenti bizonyítás egy következményeként fogalmazhatjuk meg a következő állítást.

7.8 Következmény. *Ha f folytonos egy intervallumon, akkor ott van primitív függvénye.*

Bizonyítás. Valóban, a fenti bizonyítás szerint az

$$F(x) = \int_a^x f(t)\,dt$$

függvény az f primitív függvénye. □

7.9 Példa. Számítsuk ki például az alábbi határozott integrált.

$$\int_1^2 \left(2x^3 + 1 + \frac{1}{x^2}\right) dx = \left[\frac{x^4}{2} + x - \frac{1}{x}\right]_1^2 = 9$$

Néhány további egyszerű példa:

$$\int_0^{\pi/2} \sin x \, dx = [-\cos x]_0^{\pi/2} = 1$$

$$\int_0^1 e^x \, dx = [e^x]_0^1 = e - 1$$

$$\int_0^4 \sqrt{x} \, dx = \left[\frac{2}{3} \cdot x^{3/2}\right]_0^4 = \frac{16}{3}$$

Ellenőrizzük ezeket az eredményeket.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 I/9 és I/10 szakaszai kidolgozott példáinak feldolgozása.
2. Házi feladatok: az I/9 szakasz 9.1.7, 9.1.10, 9.1.17, 9.2.9, 9.2.11, 9.2.14, valamint az I/10 szakasz 10.1.6, 10.1.9, 10.1.13 és 10.1.14 feladatai.
3. Tankönyv-1 10. fejezete.

8. fejezet

Integrálási technikák

8.1. Parciális integrálás

Ha f és g folytonosan differenciálható függvények valamely I intervallumon, akkor a szorzat differenciálási szabálya alapján:

$$\int f'(x)g(x)\,dx = f(x)g(x) - \int f(x)g'(x)\,dx$$

Ezt a formulát *parciális integrálásnak* nevezzük. Tekintsük például az

$$\int xe^{-x}\,dx$$

integrált, akkor az $f'(x) = e^{-x}$ és a $g(x) = x$ szereposztással (lehetne máshogy?):

$$\int xe^{-x}\,dx = -xe^{-x} + \int e^{-x}\,dx = -xe^{-x} - e^{-x} + C$$

8.1 Példa. Tekintsük példaként $n \neq -1$ mellett az

$$\int x^n \ln x\,dx$$

integrált. Integráljunk parciálisan az $f'(x) = x^n$ és $g(x) = \ln x$ szereposztással (mit kapnánk a fordított szereposztással?):

$$\int x^n \ln x\,dx = \frac{x^{n+1}}{n+1}\ln x - \int \frac{x^n}{n+1}\,dx = \frac{x^{n+1}}{n+1}\ln x - \frac{x^{n+1}}{(n+1)^2} + C$$

Speciálisan $n = 0$ esetén:

$$\int \ln x\,dx = x\ln x - x + C = x(\ln x - 1) + C$$

8.2. Parciális integrálás határozott integrálokra

A parciális integrálást használhatjuk határozott integrálok kiszámítására az alábbi módon:

$$\int_a^b f'(x)g(x)\,dx = [f(x)g(x)]_a^b - \int_a^b f(x)g'(x)\,dx$$

Például az $f'(x) = \sin x$ és $g(x) = x$ szereposztással (vajon a fordított jó lenne?):

$$\begin{aligned}\int_0^\pi x\sin x\,dx &= [-x\cos x]_0^\pi + \int_0^\pi \cos x\,dx \\ &= \pi + [\sin x]_0^\pi = \pi\end{aligned}$$

Ez az eljárás általában gyorsabb, mint először a határozatlan integrál kiszámítása parciális integrálással, majd ezután a határok behelyettesítése.

8.2 Példa. Néha szükség van a parciális integrálás többszöri elvégzésére. Tekintsük például az

$$\int x^2 e^{-\lambda x}\,dx$$

integrált, ahol $\lambda > 0$ adott paraméter. Használjuk az $f'(x) = e^{-\lambda x}$, illetve $g(x) = x^2$ szereposztást, akkor

$$\int x^2 e^{-\lambda x}\,dx = -\frac{1}{\lambda}x^2 e^{-\lambda x} + \frac{1}{\lambda}\int 2x e^{-\lambda x}\,dx$$

Az utóbbi integrált újabb parciális integrálással számíthatjuk ki.

Figyelem! Itt újra az $f'(x) = e^{-\lambda x}$ illetve $g(x) = x$ szereposztást kell választanunk, ellenkező esetben egy használhatatlan azonossághoz jutunk!

$$\int x^2 e^{-\lambda x}\,dx = -\frac{1}{\lambda}x^2 e^{-\lambda x} - \frac{2}{\lambda^2}x e^{-\lambda x} - \frac{2}{\lambda^3}e^{-\lambda x} + C$$

8.3 Példa. Számítsuk ki az alábbi határozott integrált:

$$\int_0^\pi e^x \sin x\,dx$$

Használjuk az $f'(x) = e^x$ és $g(x) = \sin x$ szereposztást, akkor kétszeri parciális integrálással:

$$\begin{aligned}\int_0^\pi e^x\sin x\,dx &= [e^x\sin x]_0^\pi - \int_0^\pi e^x\cos x\,dx \\ &= -[e^x\cos x]_0^\pi - \int_0^\pi e^x\sin x\,dx\end{aligned}$$

Fejezzük ki ebből az egyenlőségből a keresett integrált:

$$2\int_0^\pi e^x \sin x\, dx = -[e^x \cos x]_0^\pi$$

azaz

$$\int_0^\pi e^x \sin x\, dx = \frac{1}{2}(e^\pi + 1)$$

8.3. Integrálás helyettesítéssel

Az összetett függvény deriválási szabályából (láncszabály), annak integrálásával, adódik a következő formula:

$$\int f(g(t))g'(t)\, dt = \int f(x)\, dx$$

ahol $x = g(t)$ folytonosan differenciálható függvény egy adott intervallumon. Ezt a szabályt *helyettesítéses integrálásnak* nevezzük.

8.4 Példa. Határozzuk meg például az alábbi integrált:

$$\int 5t^3\sqrt{2+t^4}\, dt$$

integrált. Vegyük észre, hogy az $x = g(t) = t^4$ helyettesítéssel az integrál a következő alakba írható:

$$\int 5t^3\sqrt{2+t^4}\, dt = \frac{5}{4}\int \sqrt{2+x}\, dx = \frac{5}{4}\cdot\frac{2}{3}(2+x)^{3/2} + C$$

A visszehelyettesítést elvégezve:

$$\int 5t^3\sqrt{2+t^4}\, dt = \frac{5}{6}(2+t^4)^{3/2} + C$$

8.5 Példa. Lássunk olyan példát, ahol a fordított eljárás a célravezető:

$$\int e^x\sqrt{1+e^x}\, dx$$

Vezessük be az $x = g(t) = \ln t$ helyettesítést, ekkor $g'(t) = 1/t$, és így:

$$\int e^x\sqrt{1+e^x}\, dx = \int t\sqrt{1+t}\frac{1}{t}\, dt = \frac{2}{3}(1+t)^{3/2} + C$$

A $t = e^x$ visszahelyettesítéssel:

$$\int e^x\sqrt{1+e^x}\, dx = \frac{2}{3}(1+e^x)^{3/2} + C$$

8.4. Helyettesítés határozott integráloknál

Határozott integrálok kiszámításánál a visszahelyettesítésnél célravezetőbb és gyorsabb a határok megváltoztatása a helyettesítésnek megfelelően:

$$\int_a^b f(g(t))g'(t)\,dt = \int_{g(a)}^{g(b)} f(x)\,dx$$

8.6 Példa. Az alábbi példában az $x = g(t) = \cos t$, $g'(t) = -\sin t$ helyettesítést alkalmazzuk:

$$\begin{aligned} \int_0^{\pi/2} \frac{\sin 2t}{1+\cos^2 t}\,dt &= \int_0^{\pi/2} \frac{2\sin t \cos t}{1+\cos^2 t}\,dt \\ &= -\int_1^0 \frac{2x}{1+x^2}\,dx = \int_0^1 \frac{2x}{1+x^2}\,dx \\ &= [\ln(1+x^2)]_0^1 = \ln 2 \end{aligned}$$

8.7 Példa. Végül határozzuk meg az alábbi nevezetes integrált:

$$\int_0^1 \sqrt{1-x^2}\,dx$$

Vezessük be az $x = g(t) = \sin t$ helyettesítést, ekkor $g'(t) = \cos t$ és (figyeljük meg a határok megváltoztatását!):

$$\begin{aligned} \int_0^1 \sqrt{1-x^2}\,dx &= \int_0^{\pi/2} \cos^2 t\,dt \\ &= \frac{1}{2}\int_0^{\pi/2} (1+\cos 2t)\,dt = \frac{1}{2}\left[t + \frac{\sin 2t}{2}\right]_0^{\pi/2} = \frac{\pi}{4} \end{aligned}$$

Vegyük észre, hogy itt éppen az origó középpontú, egységsugarú kör területének egynegyedét határoztuk meg helyettesítéses integrálással.

8.5. Lineáris differenciálegyenlet

Differenciálegyenleten egy olyan egyenletet értünk, amelyben az ismeretlen függvény, és annak deriváltja is szerepel. Számos mikro és makroökonómiai probléma vezet ilyen feladatra. Egy tipikus ilyen feladat a lineáris differenciálegyenlet.

Legyenek a és b adott valós számok, és keressük azt az y ismeretlen differenciálható függvényt, amelyre

$$\begin{aligned} y' &= ay + b \\ y(0) &= y_0 \end{aligned} \tag{8.1}$$

ahol y_0 egy előre adott valós szám.

Itt az $y(0) = y_0$ egyenlőséget kezdeti feltételnek nevezzük. Azt mondjuk, hogy az y differenciálható függvény a fenti feladat megoldása, ha bármely $t \in \mathbb{R}$ esetén $y'(t) = ay(t) + b$, továbbá $y(0) = y_0$. Kérdés, hogy hogyan állítható elő a feladat megoldása?

Tegyük fel tehát a továbbaikban, hogy y megoldás. Szorozzuk meg az egyenlet mindkét oldalát az e^{-at} kifejezéssel, akkor átrendezés után azt kapjuk, hogy

$$y'(t)e^{-at} - ay(t)e^{-at} = be^{-at}$$

minden t valós számra. Vegyük észre, hogy a bal oldalon éppen az $y(t)e^{-at}$ szorzat deriváltja áll. Tehát mindkét oldalt a 0 ponttól t-ig integrálva (és az integrálási változót t-ről s-re cserélve)

$$\int_0^t (y'(s)e^{-as} - ay(s)e^{-as})\, ds = \left[y(s)e^{-as}\right]_0^t = \int_0^t be^{-as}\, ds$$

A határok behelyettesítésével az adódik, hogy

$$y(t)e^{-at} - y(0) = \int_0^t be^{-as}\, ds\,.$$

Átrendezve, és mindkét oldalt az e^{at} kifejezéssel szorozva az alábbi tételben fogalmazhatjuk meg az eredményünket.

8.8 Tétel. **(Cauchy-formula)** *A (8.1) feladat megoldása*

$$y(t) = e^{at}\left(y_0 + \int_0^t be^{-as}\, ds\right)$$

a számegyenesen.

Ebből azt is láthatjuk, hogy ha az $y(0) = y_0$ kezdeti feltételt nem írnánk elő, akkor a (8.1) differenciálegyenletnek végtelen sok megoldása lenne.

8.9 Példa. Ha például az

$$\begin{aligned} y' &= 2y + 5 \\ y(0) &= 3 \end{aligned}$$

lineáris differenciálegyenlet megoldását keressük, akkor a Cauchy-formula szerint

$$y(t) = e^{2t}\left(3 + \int_0^t 5e^{-2s}\,ds\right) = e^{2t}\left(3 - \frac{5}{2}\left[e^{-2s}\right]_0^t\right) = \frac{11}{2}e^{2t} - \frac{5}{2}$$

minden $t \in \mathbb{R}$ mellett.

Közvetlen behelyettesítéssel ellenőrizzük a megoldás helyességét!

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 I/9 és I/10 szakaszai kidolgozott példáinak feldolgozása.
2. Házi feladatok: az I/9 szakasz 9.3.4, 9.3.6, 9.3.8, 9.3.10, 9.3.13, 9.5.3, 9.5.6, valamint az I/10 szakasz 10.1.7, 10.1.9, 10.1.10 feladatai.
3. Tankönyv-1 11.1 és 11.2 szakaszai.

9. fejezet

Az integrálás kiterjesztése

9.1. Improprius integrálok

Tegyük fel, hogy f folytonos az $[a, +\infty)$ intervallumon. Ekkor bármely $b \geq a$ esetén létezik az $\int_a^b f(x)\,dx$ integrál.

9.1 Definíció. Azt mondjuk, hogy f improprius értelemben integrálható az $[a, \infty)$ intervallumon, ha létezik a $\lim_{b\to\infty} \int_a^b f(x)\,dx$ határérték és véges. Az improprius integrál értékét ekkor az

$$\int_a^\infty f(x)\,dx = \lim_{b\to\infty} \int_a^b f(x)\,dx$$

egyenlőséggel definiáljuk. Ha a fenti határérték nem véges, vagy nem létezik, akkor azt mondjuk, hogy az improprius integrál nem létezik (vagy nem konvergens).

Analóg módon értelmezzük a

$$\int_{-\infty}^a f(x)\,dx$$

alakú improprius integrálokat is.

9.2 Példa. Vizsgáljuk meg az

$$\int_1^\infty \frac{1}{x}\,dx$$

integrált. A definíció alapján

$$\int_1^b \frac{1}{x}\,dx = [\ln x]_1^b = \ln b$$

tehát ez az improprius integrál nem létezik, hiszen $b \to \infty$ mellett a határérték nem véges.

Ugyanakkor az

$$\int_1^\infty \frac{1}{x^2}\,dx$$

integrál létezik, ugyanis újra a definíció alapján

$$\lim_{b\to\infty}\int_1^b \frac{1}{x^2}\,dx = \lim_{b\to\infty}\left[-\frac{1}{x}\right]_1^b = 1$$

azaz az improprius integrál értéke 1.

A fenti gondolatmenet alapján könnyen látható, hogy az

$$\int_1^\infty \frac{1}{x^\alpha}\,dx$$

akkor és csak akkor létezik, ha $\alpha > 1$, és ilyenkor az improprius integrál értéke

$$\int_1^\infty \frac{1}{x^\alpha}\,dx = \frac{1}{\alpha - 1} \tag{9.1}$$

a felső határban ugyanis a határérték nulla.

9.3 Példa. Tekintsük az alábbi példát (az exponenciális eloszlás sűrűségfüggvénye):

$$\int_0^\infty \lambda e^{-\lambda x}\,dx$$

ahol $\lambda > 0$ adott konstans. Ekkor bármely $b > 0$ mellett:

$$\int_0^b \lambda e^{-\lambda x}\,dx = \left[-e^{-\lambda x}\right]_0^b = 1 - e^{-\lambda b}$$

Tehát

$$\int_0^\infty \lambda e^{-\lambda x}\,dx = \lim_{b\to\infty}(1 - e^{-\lambda b}) = 1$$

bármely $\lambda > 0$ konstans esetén.

9.2. Improprius integrálok a számegyenesen

9.4 Definíció. Azt mondjuk, hogy f improprius értelemben integrálható a számegyenesen, ha a

$$\int_{-\infty}^0 f(x)\,dx \quad \text{és} \quad \int_0^\infty f(x)\,dx$$

integrálok mindegyike létezik. Ekkor $\int_{-\infty}^{\infty} f(x)\,dx$ értéke e két integrál összege.

9.5 Példa. Például az

$$\int_{-\infty}^{\infty} \frac{2x}{1+x^2}\,dx$$

integrál nem létezik, jóllehet bármely $b > 0$ mellett

$$\int_{-b}^{b} \frac{2x}{1+x^2}\,dx = 0$$

hiszen az integrandus páratlan függvény. Azonban

$$\int_{0}^{b} \frac{2x}{1+x^2}\,dx = \ln(1+b^2)$$

és ennek határértéke $b \to \infty$ esetén $+\infty$, ezért a definíció értelmében a fenti improprius integrál nem létezik. Nyilván ugyanígy nem létezik az integrál a $(-\infty, 0]$ intervallumon sem.

9.6 Példa. Határozzuk meg az alábbi improprius integrált:

$$I = \int_{0}^{\infty} xe^{-cx^2}\,dx$$

ahol $c > 0$ adott konstans. Itt bármely $b > 0$ esetén

$$\int_{0}^{b} xe^{-cx^2}\,dx = \left[-\frac{1}{2c}e^{-cx^2}\right]_0^b$$

Innen adódik, hogy $I = 1/2c$, és ennek alapján

$$\int_{-\infty}^{\infty} xe^{-cx^2}\,dx = 0\,,$$

hiszen az integrandus páratlan függvény.

9.7 Példa. **(Gauss-féle integrál)** Fontos szerepet játszik a valószínűségszámításban az

$$I = \int_{-\infty}^{\infty} e^{-x^2}\,dx$$

improprius integrál, amely a normális eloszlás sűrűségfüggvényével van kapcsolatban. Ennek kiszámítása elég komplikált eszközöket igényel, amelyeket itt nem részletezünk. Ennek oka az, hogy a primitív függvényt nem tudjuk előállítani.

FIGYELEM! Ez nem azt jelenti, hogy nincs primitív függvény, hiszen az előző fejezetünk értelmében létezik, mert az integrandus folytonos. A probléma az, hogy ezt a primitív függvényt elemi függvények segítségével nem lehet zárt alakban előállítani.

Megmutatható, hogy

$$\int_0^\infty e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}$$

és innen az integrandus párossága folytán $I = \sqrt{\pi}$.

Egyszerű helyettesítéssel, ahol $x = t\sqrt{2}$, azt is láthatjuk, hogy

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-\frac{x^2}{2}}\,dx = 1 \tag{9.2}$$

Ezt az egyenlőséget felhasználjuk a valószínűségszámításban.

9.3. Parciális integrálás improprius integrálban

A következő példákban parciális integrálást használunk improprius integrálokban. A rövidség kedvéért a továbbiakban a $b \to +\infty$ határátmenet helyett a $+\infty$ felső határt használjuk.

9.8 Példa. Határozzuk meg adott λ pozitív konstans mellett az

$$\int_0^\infty \lambda x e^{-\lambda x}\,dx$$

improprius integrált. Az $f'(x) = \lambda e^{-\lambda x}$ és $g(x) = x$ szereposztással (azért, hogy az x szorzó eltűnjön), azt kapjuk, hogy

$$\begin{aligned}\int_0^\infty \lambda x e^{-\lambda x}\,dx &= \left[-xe^{-\lambda x}\right]_0^\infty - \int_0^\infty -e^{-\lambda x}\,dx \\ &= -\left[\frac{e^{-\lambda x}}{\lambda}\right]_0^\infty = \frac{1}{\lambda}\,.\end{aligned}$$

Ugyanis a kiintegrált rész nulla! Gondoljuk át: a L'Hôpital-szabály következménye.

9.9 Példa. Az előző példához hasonlóan határozzuk meg adott λ pozitív konstans mellett az

$$\int_0^\infty \lambda x^2 e^{-\lambda x}\,dx$$

improprius integrált. Az $f'(x) = \lambda e^{-\lambda x}$ és $g(x) = x^2$ szereposztással (azért, hogy az x^2 szorzó fokszáma csökkenjen), két egymás utáni parciális integrálással (ugyanazon szereposztással) azt kapjuk, hogy

$$\begin{aligned}
\int_0^\infty \lambda x^2 e^{-\lambda x}\, dx &= \left[-x^2 e^{-\lambda x}\right]_0^\infty - \int_0^\infty -2xe^{-\lambda x}\, dx \\
&= \left[\frac{-2xe^{-\lambda x}}{\lambda}\right]_0^\infty - \int_0^\infty -2\frac{e^{-\lambda x}}{\lambda}\, dx \\
&= \left[-2\frac{e^{-\lambda x}}{\lambda^2}\right]_0^\infty = \frac{2}{\lambda^2}
\end{aligned}$$

Ebben a példában két egymás utáni parciális integrálásra volt szükségünk, és az előző példához hasonlóan a kiintegrált részek nullák a L'Hôpital-szabály miatt.

9.10 Példa. Parciális integrálással keressük meg az

$$\int_{-\infty}^\infty x^2 e^{-x^2/2}\, dx$$

improprius integrál értékét. A tényezők szerepét ügyesen kiosztva a következőt kapjuk:

$$\int_{-\infty}^\infty (-x) \cdot \left(-xe^{-x^2/2}\right)\, dx = \left[-xe^{-x^2/2}\right]_{-\infty}^\infty + \int_{-\infty}^\infty e^{-x^2/2}\, dx = \sqrt{2\pi}$$

a (9.2) formula értelmében. Vegyük ugyanis észre, hogy a kiintegrált tagban mindkét határérték zérus a L'Hôpital-szabály miatt. Következésképpen

$$\int_{-\infty}^\infty x^2 e^{-x^2/2}\, dx = \sqrt{2\pi}\,. \tag{9.3}$$

9.4. Harmonikus sorok vizsgálata

A korábbiakban láttuk, hogy valamely $\alpha > 0$ valós kitevő mellett a

$$\sum_{k=1}^\infty \frac{1}{k^\alpha} \tag{9.4}$$

végtelen sor divergens, ha $\alpha \le 1$, illetve konvergens, ha $\alpha \ge 2$. Nem tudtunk azonban választ adni az $1 < \alpha < 2$ esetben. Erre most megoldást tudunk adni az improprius integrálok segítségével. Valóban, tekintsük a fenti sor n-ik részletösszegét

$$S_n = \sum_{k=1}^{n} \frac{1}{k^\alpha}$$

és rajzoljuk fel az

$$f(x) = \frac{1}{x^\alpha}$$

függvény grafikonját a pozitív félegyenesen. Vegyük fel a függvényértékeket az $1, \ldots, n$ egész helyeken, akkor a grafikont vizsgálva könnyen látható, hogy

$$S_n < 1 + \int_1^n \frac{1}{x^\alpha}\,dx$$

hiszen az f függvény szigorúan monoton fogyó.

KÉSZÍTSÜNK ÁBRÁT!

Másrészt az f függvény pozitív és $\alpha > 1$ esetén az improprius integrálja véges az $[1, \infty)$ intervallumon, lásd a (9.1) egyenlőséget. Tehát

$$S_n < 1 + \int_1^n \frac{1}{x^\alpha}\,dx < 1 + \int_1^\infty \frac{1}{x^\alpha}\,dx = 1 + \frac{1}{\alpha - 1} = \frac{\alpha}{\alpha - 1}.$$

Innen adódik, hogy S_n felülről korlátos, másrészt nyilvánvalóan szigorúan monoton növő, következésképpen konvergens is. Ezt az eredményünket az alábbi állításban fogalmazzuk meg.

9.11 Állítás. *A (9.4) alatti végtelen sor akkor és csak akkor konvergens, ha $\alpha > 1$, és ebben az esetben*

$$\sum_{k=1}^{\infty} \frac{1}{k^\alpha} < \frac{\alpha}{\alpha - 1}$$

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 I/10 szakasza kidolgozott példáinak feldolgozása.
2. Házi feladatok: a Feladatgyűjtemény-1 I/10 szakaszának 10.2.10, 10.2.13, 10.2.14, 10.2.16, 1.6.2, 1.8.6, 1.8.9 feladatai.
3. Tankönyv-1 11.3 és 11.4 szakaszai.

10. fejezet

Hatványsorok

10.1. Hatványsorok összege

Ha $-1 < x < 1$ valós szám, akkor a geometriai sorokról láttuk, hogy a sor konvergens, és

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k .$$

Számos alkalmazásban fontos kérdés, hogy valamely f függvény megadható-e

$$f(x) = \sum_{k=0}^{\infty} a_k x^k \tag{10.1}$$

alakban alkalmas a_k együtthatókkal. Ilyenkor azt mondjuk, hogy f hatványsorba fejthető.

10.1 Definíció. A (10.1) egyenlőség jobb oldalán álló sort *hatványsornak* nevezzük, a bal oldalon álló f függvény a hatványsor *összegfüggvénye.*

Két kérdést vizsgálunk a fejezet hátralévő részében.

1. Milyen x értékekre konvergens egy hatványsor, és ott mi az f összegfüggvény.

2. Fordítva, egy adott f függvény esetén hogyan található meg az a hatványsor, amelynek f az összegfüggvénye.

Világos, hogy bármely hatványsor $x = 0$ esetén konvergens, és az összege a_0. Azon x értékek halmazát, amelyekre a hatványsor konvergens, konvergenciahalmaznak nevezzük.

10.2. A konvergencia-sugár

Egy hatványsor konvergencia-halmaza mindig egy origóra szimmetrikus intervallum, ezt fogalmazza meg a következő tételünk.

10.2 Tétel. **(Cauchy-Hadamard-tétel)** *A (10.1) hatványsorhoz található olyan R nemnegatív szám (lehet $R = 0$ vagy végtelen is), hogy a sor konvergens a $-R < x < R$ nyílt intervallumban, és divergens a $[-R, R]$ zárt intervallumon kívül.*

Bizonyítás. Csak arra az esetre szorítkozunk, amikor létezik a

$$\lim_{k\to\infty} \left| \frac{a_{k+1}}{a_k} \right| = r$$

határérték. Vezessük be a következő jelölést:

$$R = \begin{cases} 1/r & \text{ha } 0 < r < +\infty \\ +\infty & \text{ha } r = 0 \\ 0 & \text{ha } r = \infty \end{cases}$$

A Hányados-kritérium alapján a hatványsor konvergens, ha

$$\lim_{k\to\infty} \left| \frac{a_{k+1}x^{k+1}}{a_k x^k} \right| < 1$$

ez a fenti jelölésünkkel éppen azt jelenti, hogy $|x| < R$.

Teljesen hasonló módon láthatjuk be a hányados-kritérium felhasználásával, hogy a hatványsor divergens, ha $|x| > R$. □

10.3 Definíció. A fent bevezetett R számot a hatványsor *konvergencia-sugarának* nevezzük.

10.4 Példa. Tekintsük például a

$$\sum_{k=0}^{\infty} \frac{x^k}{k!}$$

hatványsort, itt

$$\lim_{k\to\infty} \frac{a_{k+1}}{a_k} = \lim_{k\to\infty} \frac{k!}{(k+1)!} = \lim_{k\to\infty} \frac{1}{k+1} = 0$$

ezért itt $R = \infty$, azaz a hatványsor az egész számegyenesen konvergens.

Egyik másik példaként vizsgáljuk meg a

$$\sum_{k=1}^{\infty} \frac{x^k}{k}$$

hatványsort. Világos, hogy ekkor

$$\lim_{k\to\infty} \frac{a_{k+1}}{a_k} = \lim_{k\to\infty} \frac{k}{k+1} = 1$$

tehát $R = 1$, ezért a hatványsor konvergens a $(-1,1)$ nyílt intervallumban, és divergens a $[-1,1]$ zárt intervallumon kívül. Másrészt látható, hogy $x = 1$ esetén a divergens harmonikus sorhoz jutunk, továbbá $x = -1$ esetén pedig a konvergens váltakozó előjelű sorhoz, lásd a 2.10 Példát. Tehát a

$$[-1,1)$$

balról zárt, jobbról nyílt intervallum a sor konvergencia-halmaza.

10.3. Hatványsor differenciálhatósága

Tekintsünk egy hatványsort, amelynek $R > 0$ a konvergencia-sugara, és f az összegfüggvénye, azaz

$$\sum_{k=0}^{\infty} a_k x^k = f(x)$$

minden $-R < x < R$ mellett.

10.5 Tétel. *A hatványsor f összegfüggvénye differenciálható, mégpedig*

$$f'(x) = \sum_{k=1}^{\infty} k a_k x^{k-1}$$

a $(-R,R)$ nyílt intervallumban.

A tétel bizonyítását nem végezzük el, csak megjegyezzük, hogy az úgynevezett "egyenletes konvergencia" fogalmán múlik. Néhány további megjegyzést könnyen kiolvashatunk a tétel állításából.

- A derivált a hatványsor tagonkénti deriválásával állítható elő. Ez nem nyilvánvaló, hiszen a tagonkénti deriválás szabályát csak véges összegre igazoltuk, végtelen összegre általában nem érvényes. KERESSÜNK PÉLDÁT!
- Vegyük észre, hogy a tagonkénti deriválással előállított hatványsor konvergencia-sugara továbbra is R. ELLENŐRIZZÜK!

- Mivel f' is egy hatványsor összege ugyanazon intervallumon, ezért a tétel többszöri alkalmazásával láthatjuk, hogy ilyenkor f akárhányszor differenciálható a $(-R, R)$ nyílt intervallumban.

10.6 Példa. Tekintsük például $-1 < x < 1$ esetén a

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$$

hatványsort. Vegyük észre, hogy a sor első tagja 1, amelynek deriváltja nulla. Igy a fenti tételünk szerint

$$\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$$

minden $-1 < x < 1$ esetén.

10.7 Példa. Keressük meg azt az f függvényt, amelyet a következő sor állít elő

$$f(x) = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{x^k}{k}$$

Egyszerű számolással adódik, hogy a konvergencia-sugár $R = 1$. Ekkor egyrészt $f(0) = 0$, másrészt a hatványsor differenciálhatósága alapján

$$f'(x) = \sum_{k=1}^{\infty} k \frac{(-x)^{k-1}}{k} = \sum_{k=1}^{\infty} (-x)^{k-1} = \frac{1}{1+x}$$

minden $-1 < x < 1$ esetén. Innen

$$f(x) = f(0) + \int_0^x \frac{1}{1+t}\, dt = [\ln(1+t)]_0^x = \ln(1+x)$$

a $(-1, 1)$ nyílt intervallumban. Sőt, az eredeti sor a 2.10 Példa szerint az $x = 1$ helyen is konvergens, ahonnan a nevezetes

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \ldots = \ln 2$$

azonosság adódik.

10.4. Az együtthatók meghatározása

Felmerül a kérdés, hogy egy adott f függvény előállítható-e valamely hatványsor összegeként egy adott intervallumban. A differenciálhatósági tétel szerint ilyenkor f szükségképpen akárhányszor differenciálható. Vajon hogyan kaphatjuk meg az a_k együtthatókat?

A differenciálhatósági tétel alapján a (10.1) alatti f függvényhez tartozó a_k együtthatókat meghatározhatjuk az f deriváltjai segítségével. Vegyük észre, hogy

$$f(0) = a_0, \quad f'(0) = a_1, \quad f''(0) = 2a_2, \quad \ldots$$

és általában tetszőleges k indexre

$$f^{(k)}(0) = k! \cdot a_k$$

Ha ezeket a kifejezéseket beírjuk az a_k együtthatók helyére, akkor az

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k$$

előállításhoz jutunk. Ezt az alakot az f függvény Taylor-sorának nevezzük.

10.5. Az exponenciális függvény hatványsora

Tekintsük ebben a szakaszban az e^x exponenciális függvényt. Ha ez a függvény valamely hatványsor összegeként áll elő, akkor a sor együtthatói csak az

$$a_k = \frac{1}{k!}$$

kifejezések lehetnek, hiszen a függvény akárhányadik deriváltja is e^x, amely az $x = 0$ helyen az 1 értéket veszi fel. Ez azt mutatja, hogy az e^x függvényhez rendelt hatványsor

$$\sum_{k=0}^{\infty} \frac{x^k}{k!}$$

amely a korábbi példánk szerint az egész számegyenesen konvergens.

Csak az a probléma, és ezért nem írtunk egyenlőséget, mert egyelőre nem világos, hogy az e^x függvény valóban előáll-e hatványsor összegfüggvényeként.

Ennek megválaszolásához tekintsük az

$$f(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

egyelőre ismeretlen függvényt a számegyenesen. Világos, hogy egyrészt $f(0) = 1$, másrészt a differenciálhatósági tételünk szerint

$$f'(x) = \sum_{k=1}^{\infty} k \frac{x^{k-1}}{k!} = \sum_{k=1}^{\infty} \frac{x^{k-1}}{(k-1)!} = f(x)$$

minden $-\infty < x < \infty$ esetén. Ez egy egyszerű lineáris differenciálegyenlet az ismeretlen f függvényre, amelynek megoldása

$$f(x) = e^x$$

az egész számegyenesen. Innen adódik a nevezetes azonosság

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{n!} + \ldots$$

az $x = 1$ behelyettesítésével.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 I/2 és I/8 szakaszai kidolgozott példáinak feldolgozása.
2. Házi feladatok: az I/2 szakasz 2.1.7, 2.1.9, 2.1.11, 2.2.3, 2.2.9, valamint az I/8 szakasz 8.1.8, 8.2.4, 8.2.5 és 8.3.3 feladatai.
3. Tankönyv-1 6.5 szakasza.

11. fejezet

Kétváltozós függvények deriválása

11.1. Parciális deriváltak

Tekintsünk egy kétváltozós $f : \mathbb{R}^2 \to \mathbb{R}$ függvényt. Rögzítsünk le egy $y = b$ pontot és vizsgáljuk az így keletkező

$$x \to f(x, b)$$

egyváltozós függvényt. Tegyük fel, hogy ez a függvény differenciálható valamely a helyen és határozzuk meg a deriváltját.

11.1 Definíció. A fenti deriváltat az f x változó szerinti *parciális deriváltjának* nevezzük az (a, b) helyen. Jelölése:

$$\frac{\partial f}{\partial x}(a, b) = f_1'(a, b)$$

Néha használatos az $f_x'(a, b)$ jelölés is.

11.2 Példa. Tekintsük például $f(x, y) = (x + 2y)e^{x+3y-1}$ függvényt és határozzuk meg az x változó szerinti parciális deriváltat az $(1, 1)$ pontban.

Ekkor $f(x, 1) = (x + 2)e^{x+2}$, amelynek deriváltja az x helyen

$$f_1'(x, 1) = e^{x+2} + (x + 2)e^{x+2} = (x + 3)e^{x+2}$$

Ennek az $x = 1$ helyen felvett értéke $f_1'(1, 1) = 4e^3$.

11.3 Példa. Elvileg azt is megtehetnénk, hogy formálisan meghatároznánk az f függvény x változó szerinti parciális deriváltját tetszőleges fix y mellett, majd ezt követően behelyettesítenénk az $x = a$ és $y = b$ értékeket. Ez azonban nem mindig célszerű, amint ezt a következő példán láthatjuk. Legyen

$$f(x, y) = \sqrt{x^2 + y^2 + 5} \cdot e^{-2x+y} \cdot \cos(y + \pi/2)$$

és határozzuk meg az x-szerinti parciális deriváltat az $(1, 0)$ pontban. Ekkor a deriválás elvégzése hosszadalmas lenne. Ha azonban a definíció szerint járunk el, akkor láthatjuk, hogy

$$f(x, 0) = 0$$

minden x értékre, ezért $f_1'(1, 0) = 0$.

Az

$$x \to \frac{\partial f}{\partial x}(x)\,, \quad x \in \mathbb{R}$$

hozzárendelést az f függvény x változó szerinti parciális deriváltfüggvényének nevezzük.

11.2. Érintősíkok

A parciális deriváltaknak (az egyváltozós esethez hasonlóan) geometriai értelmezés is adható. Tekintsünk egy $f : \mathbb{R}^2 \to \mathbb{R}$ kétváltozós függvényt. Ennek grafikonja egy felületet ad a háromdimenziós térben. Válasszuk ki e felületnek a

$$P(a, b, f(a, b))$$

pontját. Ha ezt a felületet metszi a P ponton átmenő $y = b$ sík, akkor egy görbéhez jutunk. E görbe érintőjének meredeksége a P pontban az $f_1'(a, b)$ parciális derivált. Teljesen hasonló interpretációt adhatunk az $x = a$ síkban fekvő érintő meredekségére is. A két érintő által meghatározott sík normálvektora tehát

$$v = (f_1'(a, b), f_2'(a, b), -1)$$

azaz a $c = f(a, b)$ jelöléssel a sík egyenlete

$$f_1'(a, b)(x - a) + f_2'(a, b)(y - b) - (z - c) = 0\,.$$

Ezt a síkot a felület P pontbeli érintősíkjának nevezzük.

11.4 Példa. Adjuk meg az p paraméter értékét úgy, hogy az

$$f(x, y) = px\sqrt{x^2 + y^2 + 1} - 7$$

függvény $a = 2$, $b = 2$, $c = f(2, 2)$ pontnál húzott érintősíkja átmegy a $Q(2, -1, 6)$ ponton.

Világos, hogy $f(2,2) = 6p-7$, azaz a $P(2,2,6p-7)$ ponthoz húzott érintősík egyenletét keressük. Mivel

$$\frac{\partial f}{\partial x}(2,2) = \frac{13}{3}p \quad \text{és} \quad \frac{\partial f}{\partial y}(2,2) = \frac{4}{3}p$$

azért az érintősík egyenlete a P pontban

$$\frac{13}{3}p(x-2) + \frac{4}{3}p(y-2) - (z-6p+7) = 0\,.$$

Ha az érintősík átmegy a Q ponton, akkor annak koordinátái kielégítik a sík egyenletét. Ez a következő egyenletet adja az ismeretlen p paraméterre:

$$-4p = 13 - 6p\,.$$

Ennek egyetlen megoldása $p = 13/2$.

11.3. A láncszabály

Tekintsük most az $f : \mathbb{R}^2 \to \mathbb{R}$ valamint a $g : \mathbb{R} \to \mathbb{R}^2$ függvényeket, ahol minden $t \in \mathbb{R}$ esetén

$$g(t) = (g_1(t), g_2(t))$$

Tegyük fel, hogy g értékkészlete az f értelmezési tartományában fekszik. Ekkor vizsgálhatjuk az

$$f \circ g : \mathbb{R} \to \mathbb{R}$$

kompozíció függvény differenciálhatóságát.

11.5 Tétel. **(Láncszabály)** *Ha g_1 és g_2 differenciálhatóak a t pontban és f parciális deriváltfüggvényei folytonosak a $g(t)$ pontban, akkor $f \circ g$ is differenciálható a t pontban, és*

$$(f \circ g)'(t) = \frac{\partial f}{\partial x}(g(t))g_1'(t) + \frac{\partial f}{\partial y}(g(t))g_2'(t)$$

Tételünk állítása az egyváltozós láncszabályhoz (lásd a 4. fejezetben) technikailag kicsit bonyolultabban, de teljesen hasonló elven igazolható.

11.6 Példa. Legyen például $f(x,y) = x^2 - xy + y^2$, továbbá

$$x = g_1(t) = \cos t \quad y = g_2(t) = \sin t$$

és tekintsük az $F(t) = (f \circ g)(t)$ összetett függvényt. Ekkor a láncszabály szerint

$$\begin{aligned} F'(t) &= (f \circ g)'(t) = \frac{\partial f}{\partial x}(g(t))g_1'(t) + \frac{\partial f}{\partial y}(g(t))g_2'(t) \\ &= (2\cos t - \sin t)(-\sin t) + (-\cos t + 2\sin t)\cos t = \sin^2 t - \cos^2 t \end{aligned}$$

minden $t \in \mathbb{R}$ esetén.

11.7 Példa. Tekintsünk egy $f : \mathbb{R}^2 \to \mathbb{R}$ függvényt, amelynek a parciális deriváltjai léteznek és folytonosak, és legyen $v = (v_1, v_2) \in \mathbb{R}^2$ adott vektor. Ekkor a $P(a, b) \in \mathbb{R}^2$ ponton átmenő, v irányvektorú egyenes egyenlete

$$g(t) = (a, b) + tv = (a + tv_1, b + tv_2)\,.$$

Ezekkel a jelölésekkel egyrészt $g_1'(t) = v_1$, $g_2'(t) = v_2$, másrészt az

$$F(t) = f \circ g(t) = f((a, b) + tv)$$

egyváltozós függvény deriváltja a láncszabály szerint:

$$F'(t) = \frac{\partial f}{\partial x}((a, b) + tv)v_1 + \frac{\partial f}{\partial y}((a, b) + tv)v_2$$

illetve speciálisan $t = 0$ esetén

$$F'(0) = \frac{\partial f}{\partial x}(a, b)v_1 + \frac{\partial f}{\partial y}(a, b)v_2$$

11.4. Lokális szélsőérték

A továbbiakban a kétdimenziós tér valamely $v = (x, y)$ vektorának abszolút értékét (az origótól mért távolságát) a következő módon értelmezzük:

$$\|v\| = (x^2 + y^2)^{1/2}$$

amelyet a v vektor normájának nevezünk.

11.8 Definíció. Az $\mathbb{R}^2$ tér origó középpontú egységsugarú környezetén a

$$B = \{v \in \mathbb{R}^2 : \|v\| \leq 1\}$$

halmazt értjük. Világos, hogy valamely $(a, b) \in \mathbb{R}^2$ pont körüli $r > 0$ sugarú környezete az

$$(a, b) + rB = \{v \in \mathbb{R}^2 : \|v - (a, b)\| \leq r\}$$

formulával adható meg.

Tekintsünk egy $f : \mathbb{R}^2 \to \mathbb{R}$ függvényt. Azt mondjuk, hogy az értelmezési tartomány valamely $P(a,b)$ pontja az f lokális minimumhelye, ha található olyan $\varepsilon > 0$, hogy

$$f(x,y) \geq f(a,b)$$

az értelmezési tartomány mindazon (x,y) pontjaiban, amelyekre $(x,y) \in (a,b)+\varepsilon B$, azaz $\|(x,y)-(a,b)\| \leq \varepsilon$.

Hasonlóan értelmezzük a lokális maximum fogalmát, és értelemszerűen fogalmazhatjuk meg a globális minimum és maximum definícióját is.

Tegyük fel a továbbiakban, hogy az $f : \mathbb{R}^2 \to \mathbb{R}$ függvény parciális deriváltjai léteznek és folytonosak.

11.5. Elsőrendű szükséges feltétel

11.9 Tétel. *Ha az $(a,b) \in \mathbb{R}^2$ pont az f lokális minimumhelye, akkor $f_1'(a,b) = 0$ és $f_2'(a,b) = 0$.*

Bizonyítás. Legyen ugyanis $v \in \mathbb{R}^n$ egy tetszőleges nem nulla vektor, és tekintsük az alábbi függvényt

$$F(t) = f((a,b) + tv)\,.$$

A lokális minimum definíciója alapján az F függvénynek lokális minimumhelye van a $t = 0$ pontban, másrészt a láncszabály szerint F differenciálható is, éspedig

$$F'(t) = \frac{\partial f}{\partial x}((a,b)+tv)v_1 + \frac{\partial f}{\partial y}((a,b)+tv)v_2$$

Az 5.7 Tételre tekintettel $F'(0) = 0$ bármely v vektor mellett, azaz

$$\frac{\partial f}{\partial x}((a,b))v_1 + \frac{\partial f}{\partial y}((a,b))v_2 = 0$$

bármely v_1 és v_2 valós számokra. Ez csak úgy lehetséges, ha

$$\frac{\partial f}{\partial x}((a,b)) = 0 \quad \text{és} \quad \frac{\partial f}{\partial y}((a,b)) = 0$$

amit igazolni akartunk. □

A fenti tétel szerint a parciális deriváltakra felírt egyenletrendszer megoldásai között kereshetjük a függvény szélsőértékhelyeit. Ez az egyenletrendszer azonban (az egyváltozós esethez hasonlóan) csak szükséges feltételt fogalmaz meg. Például az

$$f(x,y) = x^3y^2$$

függvény esetében az $f_1'(x,y) = f_2'(x,y) = 0$ egyenletrendszer egyik megoldása $(x,y) = (0,0)$. Ekkor

$$f(0,0) = 0$$

de ez nem lehet szélsőérték, hiszen f az origó körüli bármilyen sugarú gömb belsejében felvesz pozitív és negatív értékeket is.

11.10 Példa. Tekintsük az

$$f(x,y) = \frac{1}{x} + \frac{1}{y} + \frac{xy}{8}$$

függvényt a síkon, ahol $x \neq 0$ és $y \neq 0$, és keressük a szélsőértékeket. A fenti tételünk értelmében szélsőérték csak a parciális deriváltak zérushelyeinél lehet. Keressük meg tehát a

$$\begin{aligned} \frac{\partial f}{\partial x} &= -\frac{1}{x^2} + \frac{y}{8} = 0 \\ \frac{\partial f}{\partial y} &= -\frac{1}{y^2} + \frac{x}{8} = 0 \end{aligned}$$

egyenletrendszer megoldásait. Ennek az egyenletrendszernek az egyetlen megoldása

$$x = 2 \quad \text{és} \quad y = 2\,,$$

és az f függvénynek csak ebben a pontban lehet szélsőértéke.

Azt, hogy egy ilyen kritikus pont valóban szélsőérték-e, a Lineáris Algebra tárgyban kifejtett eszközökkel tudjuk majd megvizsgálni (lásd a harmadik félévben). Megjegyezzük, hogy jelen esetben a $P(2,2)$ pont a függvény minimumhelye. (Lásd a gyakorlat anyagát!)

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 III/1 szakasza kidolgozott példáinak feldolgozása.
2. Házi feladatok: a III/1 szakasz 1.1.4, 1.1.5, 1.1.6, 1.1.7, 1.4.3, 1.4.4, 1.4.5, 1.5.3, 1.5.4, 1.6.3, 1.6.5 feladatai.
3. Tankönyv-1 15.3, 15.4, 15.6, 16.1 és 16.2 szakaszai.

12. fejezet

Feltételes szélsőérték

12.1. Implicit függvények

A mikróökonómiában gyakran felmerülő probléma, hogy valamely

$$F(x, y) = 0$$

egyenletből az y változó mikor fejezhető ki egyértelműen az x függvényeként. Más szavakkal: található-e egyetlen olyan $y = g(x)$ függvény, amelyre fennáll az

$$F(x, g(x)) = 0$$

azonosság minden x pontban.

Ilyen függvény nem feltétlenül létezik. Például az

$$F(x, y) = x^2 + y^2 - 1 = 0$$

egyenlet esetében (az egységsugarú kör egyenlete) az y nem fejezhető ki egyértelműen az x függvényeként. Geometriailag ez azt jelenti, hogy az $F(x, y) = 0$ egyenletet kielégítő pontok a síkon nem alkothatják egy függvény grafikonját. Ennek az az oka, hogy a fenti görbét az y-tengellyel párhuzamos egyenesek két helyen is metszik.

Előfordulhat az is, hogy az implicit egyenletből az y változó algebrai átalakításokkal nem fejezhető ki. Például ilyen az

$$F(x, y) = e^{x+y} - 2\cos y + 1 = 0$$

egyenlet. Látható, hogy az egyenletet az $(x, y) = (0, 0)$ pont kielégíti, de az y változót nem tudjuk kifejezni.

Felmerül az is, hogy ha F differenciálható függvény, akkor mikor fejezhető ki a fenti egyenletből y az x differenciálható függvényeként. Erre ad választ az alábbi tétel.

12.1 Tétel. **(Implicitfüggvény-tétel)** *Tegyük fel, hogy valamely* (x_0, y_0) *pontban*

$$F(x_0, y_0) = 0$$

továbbá F *parciális deriváltjai folytonosak e pont egy környezetében, és*

$$F_2'(x_0, y_0) \neq 0$$

Akkor létezik egyetlen olyan folytonosan differenciálható g *függvény az* x_0 *pont egy környezetében, amelyre*

- $g(x_0) = y_0$
- $F(x, g(x)) = 0$ *minden* x *pontban*
- $g'(x) = -F_1'(x, g(x))/F_2'(x, g(x))$

Felhívjuk a figyelmet arra, hogy a parciális deriváltak folytonosságából adódik, hogy $F_2'(x, g(x)) \neq 0$ az x_0 egy kis környezetében.

Tételünk geometriailag azt fejezi ki, hogy ha az $F(x, y) = 0$ egyenletű síkbeli görbéhez az (x_0, y_0) pontban húzott érintő nem párhuzamos az y-tengellyel, akkor e pont egy környezetében (lokálisan) az y kifejezhető az x differenciálható függvényeként.

12.2 Példa. Tekintsük például az

$$F(x, y) = e^{x+y} + x + y - 1 = 0$$

implicit egyenletet. Ezt az egyenletet a $(0, 0)$ pont kielégíti. Másrészt ebben a pontban

$$F_2'(0, 0) = 2$$

Tehát F kielégíti az Implicitfüggvény-tétel feltételeit, ezért az egyenlet egyértelműen meghatároz egy $y = g(x)$ differenciálható függvényt. A tétel szerint

$$\begin{aligned} g'(x) &= -F_1'(x, g(x))/F_2'(x, g(x)) \\ &= -\frac{1}{e^{x+g(x)} + 1} \cdot (e^{x+g(x)} + 1) = -1 \end{aligned}$$

minden x pontban. Mivel $g(0) = 0$, ez azt jelenti, hogy

$$g(x) = -x$$

és ez az egyetlen megoldás.

12.3 Példa. Tekintsük a fentebb említett

$$F(x,y) = e^{x+y} - 2\cos y + 1 = 0$$

implicit egyenletet. Ezt az egyenletet a $(0,0)$ pont kielégíti. Másrészt ebben a pontban

$$F_2'(0,0) = 1$$

ezért fennállnak az Implicitfüggvény-tétel feltételei. Tehát ez az egyenlet egyértelműen meghatároz egy g differenciálható függvényt, bár ezt a függvényt az egyenletből algebrai átalakításokkal nem tudjuk előállítani.

12.2. Feltételes szélsőérték

Tekintsük az f és F egyaránt $\mathbb{R}^2 \to \mathbb{R}$ függvényeket, és tegyük fel a továbbiakban hogy léteznek és folytonosak a parciális deriváltjaik. Feltételes szélsőérték-feladaton az alábbi feldatot értjük:

$$\begin{aligned} f(x,y) &\to \min \\ F(x,y) &= c \end{aligned} \tag{12.1}$$

ahol c adott valós állandó. Más szavakkal az f minimumát (vagy máskor maximumát) keressük a

$$H = \{(x,y) \in \mathbb{R}^2 : F(x,y) = c\}$$

halmazon.

12.4 Definíció. Azt mondjuk, hogy az $(x_0, y_0) \in H$ pont a (12.1) feltételes szélsőérték-feladat megoldása, ha

$$f(x_0, y_0) \leq f(x,y)$$

minden $(x,y) \in H$ esetén. Analóg definíció érvényes maximum esetére.

12.5 Példa. Az alábbi példánk azt illusztrálja, hogy a korábbi szélsőértékre vonatkozó szükséges feltételeink az ilyen feladatokban nem működnek. Tekintsük például az

$$f(x,y) = x^2 + 2y, \quad F(x,y) = x + y = 0 \quad \text{azaz} \quad c = 0$$

feladatot. Ekkor az $x + y = 0$ feltételből $y = -x$, és így $f(x,y) = x^2 - 2x$. Ennek az $x = 1$ pontban van minimuma, és itt a feltétel szerint $y = -1$, azaz a feltételes minimum az

$$(x_0, y_0) = (1, -1)$$

pontban található. Ebben a pontban azonban a

$$\frac{\partial f}{\partial x} = 0, \quad \frac{\partial f}{\partial y} = 0$$

egyenlőségek egyike sem teljesül.

A fenti példában láthatjuk, hogy a feltételes szélsőérték-feladat feltétel nélkülivé alakítható át, ha az $F(x, y) = c$ feltételből az y változó kifejezhető. Nehezebb feladatoknál ezt azonban algebrai átalakításokkal nem tudjuk elvégezni.

12.3. Lagrange-multiplikátorok

Tekintsük a (12.1) feltételes szélsőérték-feladatot. Az implicitfüggvény-tétel segítségével feltételt adhatunk arra, hogy az $F(x, y) = c$ feltételből az y változót kifejezhessük, és így megoldhassuk a feladatot. Ezt fogalmazzuk meg az alábbiakban.

12.6 Definíció. A (12.1) feladat *Lagrange-függvényén* az

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda(F(x, y) - c)$$

függvényt értjük, ahol λ egyelőre tetszőleges valós szám.

12.7 Tétel. **(Lagrange-módszer)** *Tegyük fel, hogy (x_0, y_0) a (12.1) feladat megoldása, továbbá f és F parciális deriváltjai léteznek és folytonosak e pont környezetében. Ha*

$$F_2'(x_0, y_0) \neq 0\,, \tag{12.2}$$

akkor található pontosan egy olyan λ valós szám, amelyre

$$\frac{\partial \mathcal{L}}{\partial x}(x_0, y_0, \lambda) = 0\,, \quad \textit{és} \quad \frac{\partial \mathcal{L}}{\partial y}(x_0, y_0, \lambda) = 0$$

Bizonyítás. A (12.2) feltételünk miatt alkalmazhatjuk az implicitfüggvény-tételt. Eszerint található olyan g folytonosan differenciálható függvény, amelyre

- $g(x_0) = y_0$, és
- $F(x, g(x)) = c$ az x_0 egy környezetében, továbbá
- $g'(x_0) = -F_1'(x_0, y_0)/F_2'(x_0, y_0)$.

Ha (x_0, y_0) a (12.1) feladat megoldása, akkor az $x \to f(x, g(x))$ függvénynek minimuma van az x_0 pontban, ezért itt a deriváltja nulla. Ez a derivált a láncszabály szerint

$$f_1'(x_0, y_0) + f_2'(x_0, y_0)g'(x_0) = f_1'(x_0, y_0) - \frac{f_2'(x_0, y_0)}{F_2'(x_0, y_0)} F_1'(x_0, y_0) = 0\,.$$

Vezessük be a következő jelölést:

$$\lambda = \frac{f_2'(x_0, y_0)}{F_2'(x_0, y_0)}\,.$$

Ezzel a jelöléssel a fenti derivált úgy is írható, hogy:

$$\frac{\partial \mathcal{L}}{\partial x}(x_0, y_0, \lambda) = f_1'(x_0, y_0) - \lambda F_1'(x_0, y_0) = 0\,.$$

A tétel állításának második egyenlősége λ behelyettesítésével nyilvánvaló azonosság, hiszen:

$$\frac{\partial \mathcal{L}}{\partial y}(x_0, y_0, \lambda) = f_2'(x_0, y_0) - \lambda F_2'(x_0, y_0) = 0\,.\square$$

Tételünket ugyanígy fogalmazhatjuk meg maximum esetére is.

12.4. A szélsőérték-feladat megoldása

Az előző tételünk alapján a (12.1) feltételes szélsőérték-feladat megoldásának menete a következő.

1. Állítsuk elő a feladat Lagrange-függényét.
2. Határozzuk meg az x és y szerinti parciális deriváltakat, és tegyük őket nullával egyenlővé.
3. Vegyük figyelembe, hogy $F(x_0, y_0) = c$.
4. Oldjuk meg az így nyert háromismeretlenes egyenletrendszert.

A megoldásként kapott (x_0, y_0) kielégíti a szélsőérték szükséges feltételét. Az egyenletrendszer megoldásaként nyert λ valós számot a feladathoz tartozó Lagrange-multiplikátornak nevezzük.

12.8 Példa. Oldjuk most meg a 12.5 Példában vizsgált feladatot Lagrange módszerével. Ekkor a feladat Lagrange-függvénye:

$$\mathcal{L}(x, y, \lambda) = x^2 + 2y - \lambda(x + y)\,.$$

Egyenletrendszerünk tehát a következő alakot ölti:

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial x}(x_0, y_0, \lambda) &= 2x_0 - \lambda = 0 \\
\frac{\partial\mathcal{L}}{\partial y}(x_0, y_0, \lambda) &= 2 - \lambda = 0 \\
\frac{\partial\mathcal{L}}{\partial \lambda}(x_0, y_0, \lambda) &= x_0 + y_0 = 0
\end{aligned}$$

Innen azonnal adódik, hogy $\lambda = 2$, $x_0 = 1$ és $y_0 = -1$.

12.9 Példa. Tekintsünk most egy kicsit bonyolultabb, és a mikroökonómiában gyakran felmerülő feladatot. Keressük meg a fogyasztói keresletre vonatkozó

$$\begin{aligned}
x^\alpha y^\beta &\to \max \\
px + y &= m
\end{aligned} \tag{12.3}$$

feltételes maximumfeladat egyetlen lehetséges megoldását, ahol α, β, p és m adott pozitív állandók. Ekkor

$$f(x, y) = x^\alpha y^\beta \quad \text{és} \quad F(x, y) = px + y\,,$$

ezért a feladat Lagrange-függvénye

$$\mathcal{L}(x, y, \lambda) = x^\alpha y^\beta - \lambda(px + y - m)\,.$$

A Lagrange-módszerrel adódó egyenletrendszer tehát

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial x}(x_0, y_0, \lambda) &= \alpha \cdot x_0^{\alpha-1} y_0^\beta - \lambda p = 0 \\
\frac{\partial\mathcal{L}}{\partial y}(x_0, y_0, \lambda) &= \beta \cdot x_0^\alpha y_0^{\beta-1} - \lambda = 0 \\
\frac{\partial\mathcal{L}}{\partial \lambda}(x_0, y_0, \lambda) &= px_0 + y_0 - m = 0\,.
\end{aligned}$$

Ennek az egyenletrendszernek egyetlen megoldása

$$px_0 = \frac{\alpha}{\alpha + \beta} m \quad \text{és} \quad y_0 = \frac{\alpha}{\alpha + \beta} m\,,$$

a λ Lagrange-multiplikátor a második egyenletből meghatározható.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 III/1.8 és III/5 szakaszok kidolgozott példáinak feldolgozása.
2. Házi feladatok: a III/1 szakasz 1.8.3 és 1.8.5, továbbá az III/5 szakasz 5.1.5, 5.1.6, 5.1.7 és 5.2.3 feladatai.
3. Tankönyv-1 16.3, 18.1, 18.2, 18.3, 18.4, 18.5 és 18.6 szakaszok.

II. rész

Második félév: Valószínűségszámítás

13. fejezet

Valószínűség

13.1. Kísérletek

Az alábbiakban (véletlen) kísérletekkel és azok (véletlen) kimeneteleivel foglalkozunk. Olyan kísérleteket vizsgálunk, amelyek kimenetelét nem tudjuk előre meghatározni.

1. Földobunk egy dobókockát, és megnézzük hányast dobtunk.
2. Földobunk egy dobókockát kétszer egymás után.
3. Feldobunk egy dobókockát, majd egy érmét annyiszor, ahányat a kockával dobtunk.
4. Egy dobókockát addig dobunk, amíg először 6-ost kapunk.
5. Válasszunk az egységsugarú körlapon véletlenszerűen egy pontot.

További bonyolultabb példák:

- Egy kereszteződésen 10 és 11 óra között áthaladó autók száma.
- Egy telefonközpontba 8 és 9 között érkező hívások száma.
- Két hívás között eltelt időintervallum hossza.
- Egy részvény tőzsdei árfolyama a zárás időpontjában.
- Ügyfélszolgálati irodában a várakozási idő hossza.

13.2. Az eseménytér

13.1 Definíció. Jelölje a továbbiakban Ω egy kísérlet kimeneteleinek halmazát. Az Ω halmazt az adott kísérlethez tartozó *eseménytérnek* nevezzük.

Állítsuk elő az előző példáinkhoz tartozó eseménytereket. Ekkor sorrendben:

1. $\Omega = \{1, 2, 3, 4, 5, 6\}$
2. $\Omega = \{(1, 1), (1, 2), (2, 1), (1, 3), \dots, (6, 6)\}$
3. $\Omega = \{1F, 1I, 2FF, 2FI, 2IF, 2II, \dots\}$ (Kérdés: vajon hány elemből áll ez az eseménytér?)
4. Ω mindazon véges számsorozatokból áll, amelyek utolsó eleme 6, az előző elemek pedig az 1,2,3,4,5 számjegyek valamelyike.
5. $\Omega = \{(x, y) : x^2 + y^2 \leq 1\}$

13.3. Események

13.2 Definíció. Az eseménytér részhalmazait *eseményeknek* nevezzük.

Tekintsük néhány példát az előző eseményterekben.

1. Jelentse A azt az eseményt, hogy a dobott szám páros. Ekkor $A = \{2, 4, 6\}$.
2. Jelentse A azt az eseményt, hogy a dobott számok összege 7. Ekkor $A = \{(1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3)\}$.
3. Jelentse A azt az eseményt, hogy egyetlen Irást sem dobunk. Ekkor $A = \{1F, 2FF, 3FFF, 4FFFF, 5FFFFF, 6FFFFFF\}$.
4. Jelentse A azt az eseményt, hogy legfeljebb két dobásra van szükségünk. Ekkor $A = \{6, 16, 26, 36, 46, 56\}$.
5. Jelentse A azt az eseményt, hogy a választott pont a középponthoz 1/2-nél közelebb van. Ekkor $A = \{(x, y) : x^2 + y^2 < 1/4\}$.

13.4. Műveletek eseményekkel

Azt mondjuk, hogy az $A \subset \Omega$ esemény bekövetkezik, ha a kísérletnek olyan $\omega \in \Omega$ kimenetele következik be, amelyre $\omega \in A$.

A lehetetlen eseménynek egyetlen eleme sincs, jelölése: $\emptyset$ (üres halmaz). A biztos esemény: Ω (az egész eseménytér).

1. $A \cap B$ pontosan akkor következik be, ha A és B is bekövetkezik. Azt mondjuk, hogy A és B egymást kizáró események, ha $A \cap B = \emptyset$.
2. $A \cup B$ pontosan akkor következik be, ha vagy A vagy B bekövetkezik (vagy mindkettő).
3. $\overline{A}$ (az A ellentettje, komplementere) pontosan akkor következik be, ha A nem következik be.

Azt mondjuk, hogy az A esemény bekövetkezése maga után vonja B bekövetkezését (A implikálja B-t), ha $A \subset B$.

13.3 Tétel. (De Morgan formulák)

1. $\overline{A \cup B} = \overline{A} \cap \overline{B}$

2. $\overline{A \cap B} = \overline{A} \cup \overline{B}$

Ezek az azonosságok tetszőleges számú eseményre is érvényesek.

Bizonyítás. Az első egyenlőséget igazoljuk. Legyen $x \in \overline{A \cup B}$ tetszőleges. Ekkor

$$x \in \overline{A \cup B} \Rightarrow x \notin A \cup B \Rightarrow x \notin A \text{ és } x \notin B \Rightarrow x \in \overline{A} \text{ és } x \in \overline{B} \Rightarrow x \in \overline{A} \cap \overline{B}$$

Ez azt igazolja, hogy $\overline{A \cup B} \subset \overline{A} \cap \overline{B}$. A másik irányú tartalmazás (és így az egyenlőség) abból adódik, hogy a gondolatmenetben az implikációk mindegyike fordítva is igaz (és így ekvivalenciák). A második egyenlőség teljesen hasonlóan ellenőrizhető. $\square$

Egy kísérlet elvégzésekor nem biztos, hogy minden kimenetele megfigyelhető. Például ha két azonos kockát feldobunk, az $(1,2)$ és $(2,1)$ kimenetelekről nem dönthető el, hogy melyik következett be. Csak azt tudjuk megfigyelni, hogy az $\{(1,2),(2,1)\}$ esemény bekövetkezett.

13.4 Definíció. Jelentse $\mathcal{A}$ a *megfigyelhető események* halmazát, amelyről a következő tulajdonságokat tételezzük fel.

- Ha $A \in \mathcal{A}$, akkor $\overline{A} \in \mathcal{A}$ és $\Omega \in \mathcal{A}$.
- Ha $A_1, A_2, \ldots \in \mathcal{A}$, akkor $A_1 \cup A_2 \cup \ldots \in \mathcal{A}$.

13.5 Állítás. *Ha A és B megfigyelhető események, akkor $A \cap B$ is az.*

Bizonyítás. Valóban, ha A és B megfigyelhetőek, akkor

$$A \cap B = \overline{\overline{A} \cup \overline{B}} \in \mathcal{A}$$

is az a De Morgan formula alapján. □

A fenti állításunk a De Morgan azonosságra tekintettel megszámlálhatóan végtelen számú eseményre is érvényes.

13.6 Definíció. A továbbiakban *kísérleten* a $\mathcal{K} = (\Omega, \mathcal{A})$ kettőst értjük.

13.5. Valószínűségi mező

Tegyük fel, hogy egy $\mathcal{K}$ kísérletet n-szer egymás után elvégzünk és minden alkalommal megfigyeljük az $A \in \mathcal{A}$ esemény bekövetkezését. Ha az A k_n esetben következett be, akkor az A relatív gyakorisága:

$$\frac{k_n}{n}$$

A tapasztalat azt mutatja, hogy az n növekedésével a relatív gyakoriság egyre kisebb kilengésekkel egy bizonyos szám körül ingadozik. Ezt tekintjük az A esemény valószínűségének.

A továbbiakban a valószínűség fogalmára olyan axiomatikus bevezetést kívánunk felépíteni, amelyből ez a tapasztalati tény levezethető.

13.7 Definíció. **(Valószínűség axiómái)** Tekintsünk egy $\mathcal{K} = (\Omega, \mathcal{A})$ kísérletet. A *valószínűség* egy

$$P : \mathcal{A} \to [0, 1]$$

függvény, amely kielégíti a következő két axiómát:

1. $P(\Omega) = 1$
2. Ha $A_1, A_2, \ldots \in \mathcal{A}$, egymást páronként kizáró események, akkor

$$P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$$

Ebben az esetben az $(\Omega, \mathcal{A}, P)$ hármast *valószínűségi mezőnek* nevezzük.

Ez az axiomatikus felépítés A. N. Kolmogorovtól (1933) származik, és nagyjából akkortól számíthatjuk a modern valószínűségszámítás kialakulását.

Az axiómákból könnyen levezethetők a valószínűségi mező alábbi tulajdonságai.

13.8 Tétel.

1. *Bármely $A \in \mathcal{A}$ esetén*
$$P(\overline{A}) = 1 - P(A)$$
és ezért $P(\emptyset) = 0$.

2. *Ha $A, B \in \mathcal{A}$ és $A \subset B$, akkor*
$$P(A) \leq P(B)$$

3. *Ha $A, B \in \mathcal{A}$, akkor*
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Bizonyítás. 1. Mivel $A \cup \overline{A} = \Omega$, továbbá A és $\overline{A}$ egymást kizáró események, innen az axiómákból adódik az állítás.

2. Ha $A \subset B$, akkor $A \cup (B \cap \overline{A}) = B$, továbbá A és $B \cap \overline{A}$ egymást kizáró események, tehát az axiómák alapján
$$P(B) = P(A) + P(B \cap \overline{A}) \geq P(A)$$
hiszen $P(B \cap \overline{A}) \geq 0$.

A 3. állítást a következőképpen igazoljuk. Az $A \cup B$ eseményt diszjunkt részekre bontjuk a következő módon:
$$A \cup B = (A \cap \overline{B}) \cup (\overline{A} \cap B) \cup (A \cap B)\,.$$
Ekkor a valószínűség axiómái alapján
$$\begin{aligned} P(A \cup B) &= P(A \cap \overline{B}) + P(\overline{A} \cap B) + P(A \cap B) \\ &= P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) \end{aligned}$$
ahonnan azonnal adódik az állítás. □

13.9 Példa. Egy AK szakos évfolyam esetén annak a valószínűsége, hogy egy találomra választott hallgató levizsgázott matematikából 0.72, míg annak, hogy mikroökonómiából 0.66. Annak a valószínűsége, hogy mind a kettőből levizsgázott 0.54. Mi a valószínűsége, hogy egy véletlenszerűen választott hallgató

(a) legalább az egyik tárgyból levizsgázott,

(b) levizsgázott mikroökonómiából, de matematikából nem,

(c) egyik tárgyból sem vizsgázott le.

Jelölje A azt az eseményt, hogy a hallgató levizsgázott matematikából, és B azt az eseményt, hogy mikroökonómiából. Ekkor $P(A) = 0.72$, $P(B) = 0.66$ és $P(A \cap B) = 0.54$. Az A és B eseményekkel a kérdéses valószínűségek az alábbi módon adhatók meg.

(a) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.84$

(b) $P(\overline{A} \cap B) = P(B) - P(A \cap B) = 0.12$

(c) $P(\overline{A} \cap \overline{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 0.16$

Otthoni tanuláshoz

1. A Feladatgyűjtemény-2 II/ és III/3 szakaszainak feldolgozása.

2. Házi feladatok: a II szakasz 130, 131, 133, 134, 135, továbbá a III/3 szakasz 196, 198, 199 feladatai.

3. Tankönyv-2 2. fejezete és a 3.1, 3.2, 3.5 szakaszok. Továbbá: A KÖZÉPISKOLAI KOMBINATORIKA ALAPOS ÁTISMÉTLÉSE!

14. fejezet

Mintavételi eljárások

14.1. Klasszikus valószínűségi mezők

14.1 Definíció. Tekintsünk egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőt. Ezt *klasszikus valószínűségi mezőnek* nevezzük, ha

- Ω véges halmaz,
- Minden $\omega \in \Omega$ esetén $\{\omega\} \in \mathcal{A}$,
- Az Ω minden egyelemű részhalmaza azonos valószínűségű.

Világos, hogy ha Ω éppen n-elemű, akkor bármely $\omega \in \Omega$ esetén

$$P(\{\omega\}) = \frac{1}{n}$$

Nevezetesen, ha az $A \subset \Omega$ esemény k elemből áll, akkor

$$P(A) = \frac{k}{n}$$

Ezt úgy interpretálhatjuk, hogy az A valószínűsége:

$$P(A) = \frac{\text{kedvező esetek száma}}{\text{összes esetek száma}} \tag{14.1}$$

A (14.1) formulát a továbbiakban klasszikus képletnek nevezzük.

14.2 Példa. Egy szabályos dobókockát feldobunk egymás után kétszer. Mi a valószínűsége, hogy a dobott számok összege pontosan 7?

Jelentse A azt az eseményt, hogy az összeg 7. Látható, hogy Ω 36 elemű halmaz (összes esetek száma), míg A egy 6 elemű részhalmaz (kedvező esetek száma), amely az $(1,6),(6,1),(2,5),(5,2),(3,4),(4,3)$ elemekből áll. Ezért

$$P(A)=\frac{6}{36}=\frac{1}{6}$$

a (14.1) klasszikus képlet alapján.

14.3 Példa. Egy 52 lapos kártyapakliból kihúzunk véletlenszerűen 5 lapot. Mi a valószínűsége, hogy vagy mind az 5 lap treff, vagy van közte ász?

Vezessük be a következő jelöléseket:

$$A=\{\text{mind az öt lap treff}\} \qquad B=\{\text{van közte ász}\}$$

Világos, hogy $P(A\cup B)$ értékét keressük. Feltételünk szerint bármelyik 5 lap húzása egyformán valószínű, ezért:

$$P(A)=\frac{\binom{13}{5}}{\binom{52}{5}} \qquad P(B)=1-\frac{\binom{48}{5}}{\binom{52}{5}}$$

továbbá

$$P(A\cap B)=\frac{\binom{12}{4}}{\binom{52}{5}}$$

Innen

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

14.4 Példa. Egy kiárusitáson egy kosárban van 10 különböző pár cipő. Egy tolvaj véletlenszerűen elvisz 4 cipőt. Mi a valószínűsége, hogy van közte legalább egy pár?

Az alábbiakban két gondolatmenetet vázolunk, de csak az egyik vezet helyes eredményre.

- Először válasszunk ki egy párt, majd a másik kettő tetszőleges lehet, tehát vagy egy újabb pár, vagy két bármilyen cipő, azaz:

$$\frac{10\binom{18}{2}}{\binom{20}{4}}$$

- Nézzük meg mi a valószínűsége, hogy a kiválasztott cipők között egyetlen pár sincs. Ezt úgy érhetjük el, hogy egy kiválasztott cipő után annak párját félretesszük. Figyeljünk arra, hogy a kiválasztásnál a sorrend nem számít, így:

$$1-\frac{\frac{20\cdot 18\cdot 16\cdot 14}{4!}}{\binom{20}{4}}$$

Ellenőrizzük, hogy a fenti két érték nem egyenlő! Vajon melyik helyes?

14.5 Példa. Tegyük fel, hogy egy kockával addig dobunk, amíg először 6-os nem jön ki. Mi a valószínűsége, hogy páros sok dobásra van szükségünk?

Jelentse A azt az eseményt, hogy páros számú dobásra van szükségünk, illetve A_k azt az eseményt, hogy k számú dobásra van szükségünk. Világos, hogy

$$P(A_k) = \left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6}$$

ahol $k = 1, 2, \ldots$ Az A esemény így állítható elő:

$$A = A_2 \cup A_4 \cup \ldots = \bigcup_{k=1}^{\infty} A_{2k}$$

A jobb oldalon álló események páronként kizárják egymást, ezért

$$P(A) = \sum_{k=1}^{\infty} P(A_{2k}) = \sum_{k=1}^{\infty} \left(\frac{5}{6}\right)^{2k-1} \cdot \frac{1}{6} = \frac{5}{11}$$

14.2. Mintavétel visszatevés nélkül

Tekintsünk egy N számú objektumból álló halmazt, amelyek közül m számú selejtes. Válasszunk ki az egész halmazból véletlenszerűen egy n-elemű mintát visszatevés nélkül ($n \leq m$). Jelentse A_k azt az eseményt, hogy a minta pontosan k számú selejtes elemet tartalmaz ($0 \leq k \leq n$). Ekkor

$$P(A_k) = \frac{\binom{m}{k} \cdot \binom{N-m}{n-k}}{\binom{N}{n}}$$

amelyet a visszatevés nélküli mintavétel formulájának nevezünk.

14.6 Példa. Egy 52 lapos kártyapakliból válasszunk ki véletlenszerűen 5 lapot. Mi a valószínűsége, hogy a kiválasztott lapok között pontosan 2 treff van?

Jelentse A a kérdéses eseményt. A visszatevés nélküli mintavétel formulája alapján

$$P(A) = \frac{\binom{13}{2} \cdot \binom{39}{3}}{\binom{52}{5}} \,.$$

Itt értelemszerűen a treffek a "selejtes objektumok".

14.7 Példa. Keressük meg annak valószínűségét, hogy a hagyományos lottón egy véletlenszerűen kitöltött szelvénnyel legalább kettes találatunk van.

Jelölje A azt az eseményt, hogy legalább kettes találatunk van, illetve A_k azt, hogy pontosan k találatunk van. Világos, hogy az A_k események $k = 2, \ldots, 5$ esetén egymást páronként kizárják. Mivel $A = A_2 \cup A_3 \cup A_4 \cup A_5$, ezért

$$P(A) = \sum_{k=2}^{5} P(A_k) = \sum_{k=2}^{5} \frac{\binom{5}{k} \cdot \binom{85}{5-k}}{\binom{90}{5}}$$

hiszen a diszjunkt unió valószínűsége összegként áll elő.

14.8 Példa. Egy 52 lapos kártyapakliból véletlenszerűen kiveszünk 5 lapot visszatevés nélkül. Mi a valószínűsége, hogy a mintában mind a négy szín (treff, káró, kőr, pikk) előfordul?

Vizsgáljuk meg a következő gondolatmenetet. Jelentse A azt az eseményt, hogy az 5 lapos mintában mind a négy szín előfordul. Bármelyik 5 lap választása egyformán valószínű, ezért klasszikus valószínűségi mezővel van dolgunk.

A kedvező esetek számának meghatározásához világos, hogy mindegyik színből 13-féleképpen választhatunk egy lapot. Ha ezt megtettük, akkor az ötödik lap már bármi lehet a maradék 48 lapból.

Az összes esetek száma: ahányféleképpen 5 lap kiválasztható 52 lapból. Tehát

$$P(A) = \frac{13^4 \cdot 48}{\binom{52}{5}}$$

Vajon helyes eredményre jutottunk-e? Ha nem, akkor hogyan javítható a megoldás?

14.3. Mintavétel visszatevéssel

Tekintsünk újra egy N számú objektumból álló halmazt, amelyek közül m számú selejtes. Válasszunk ki az egész halmazból véletlenszerűen, egyenként visszatevéssel egy n-elemű mintát. Jelentse A_k azt az eseményt, hogy a minta pontosan k-számú selejtes elemet tartalmaz.

Vizsgáljuk meg a különböző sorrendű húzásokat. Mivel bármelyik adott sorrendben k selejtes és $n-k$ nem selejtes kihúzásának valószínűsége éppen

$$\frac{m^k \cdot (N-m)^{n-k}}{N^n} = \left(\frac{m}{N}\right)^k \left(1 - \frac{m}{N}\right)^{n-k}$$

és ilyen húzásból pontosan $\binom{n}{k}$ darab van, és ezek páronként kizáróak, ezért

$$P(A_k) = \binom{n}{k} \left(\frac{m}{N}\right)^k \left(1 - \frac{m}{N}\right)^{n-k}$$

Ezt a visszatevéses mintavétel formulájának nevezzük.

14.9 Példa. Egy 52 lapos kártyapakliból húzzunk ki egymás után visszatevéssel véletlenszerűen 5 lapot (a kihúzott lapot húzás után mindig visszatesszük). Mi a valószínűsége, hogy

(a) pontosan 2 treffet húztunk,

(b) legalább 2 treffet húztunk.

Jelentse A_k azt az eseményt, hogy pontosan k treffet húztunk. Ekkor

$$(a) \quad P(A_2) = \binom{5}{2}\left(\frac{1}{4}\right)^2\left(1-\frac{1}{4}\right)^3$$

valamint

$$(b) \quad P(A_2 \cup \ldots \cup A_5) = \sum_{k=2}^{5}\binom{5}{k}\left(\frac{1}{4}\right)^k\left(\frac{3}{4}\right)^{5-k}$$

hiszen az $A_2, \ldots, A_5$ események páronként kizárják egymást.

14.4. A Bernoulli-kísérlet

A fenti gondolatmenet általánosítható is a következő módon. Tegyük fel, hogy egy kísérletben az A esemény bekövetkezésének valószínűsége valamely adott $0 \le p \le 1$ szám.

Tegyük fel, hogy elvégezzük ezt a kísérletet (egymástól függetlenül) egymás után n-szer, és minden esetben megfigyeljük, hogy az A esemény bekövetkezik-e vagy sem. Ezt a kísérletsorozatot Bernoulli-féle kísérletnek nevezzük.

Legyen $0 \le k \le n$ adott természetes szám. Jelentse A_k azt az eseményt, hogy az n számú kísérletből A pontosan k esetben következik be.

A visszatevéses mintavételnél követett gondolatmenetet alkalmazva láthatjuk, hogy a kérdéses valószínűség:

$$P(A_k) = \binom{n}{k} p^k (1-p)^{n-k}$$

bármely $k = 0, 1, \ldots, n$ esetén.

14.10 Példa. Egy hagyományos lottószelvény esetében azt mondjuk, hogy nyerő, ha legalább kéttalálatos. Vásárolunk 20 szelvényt, és véletlenszerűen (egymástól függetlenül) kitöltjük őket. Mi a valószínűsége, hogy legalább 5 nyerő szelvényünk lesz?

Világos, hogy egyetlen szelvény esetében annak valószínűsége, hogy nyerő:

$$p = \sum_{k=2}^{5} \frac{\binom{5}{k} \cdot \binom{85}{5-k}}{\binom{90}{5}}$$

Mivel ez mindegyik szelvényre érvényes, és a szelvényeket egymástól függetlenül töltöttük ki, eredeti feladatunk egy Bernoulli-féle kísérletnek tekinthető ezzel a p paraméterrel. Tehát a kérdéses valószínűség:

$$\sum_{k=5}^{20} \binom{20}{k} p^k (1-p)^{20-k}$$

ahol p a fenti valószínűség.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-2 III/1 szakaszának feldolgozása.
2. Házi feladatok: a III/3 szakasz 143, 144, 145, 153, 154, 157, 159, 165 és 173 feladatai.
3. Tankönyv-2 3.3 szakasza. Továbbá: A KÖZÉPISKOLAI KOMBINATORIKA ALAPOS ÁTISMÉTLÉSE!

15. fejezet

Feltételes valószínűség és Bayes-tétel

15.1. Feltételes valószínűség

Számos esetben kell az A esemény valószínűségét meghatározni olyan a priori feltétel mellett, hogy valamely B eseményről tudjuk, hogy bekövetkezett. Ilyen esetekben az eseménytér elemei közül csak azokat vesszük számításba, amelyek a B-nek is elemei.

Ez azt jelenti, hogy az Ω helyett az eseményteret a B eseményre szűkítjük, és erre vonatkozóan számítjuk ki egy A esemény (feltételes) valószínűségét.

15.1 Definíció. Tekintsünk egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőt és egy olyan $B \in \mathcal{A}$ eseményt, amelyre $P(B) \neq 0$. Az $A \in \mathcal{A}$ esemény B-re vonatkozó feltételes valószínűségén az alábbit értjük:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

15.2 Példa. Tegyük fel, hogy két kockával dobunk, de nem látjuk az eredményt. Valaki megmondta, hogy az egyik dobás 5-ös. Mi a valószínűsége, hogy a másik 6-os?

Figyelem, az eredmény nem $1/6$ lesz!

Jelentse A és B a következő eseményeket:

$$B = \{\text{az egyik dobás 5-ös}\} \qquad A = \{\text{a másik dobás 6-os}\}$$

Ekkor $P(B) = 11/36$ hiszen 11 olyan számpár van, amelyben az 5-ös szerepel. Másrészt $A \cap B = \{(5,6),(6,5)\}$, ezért $P(A \cap B) = 2/36$. Tehát

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2/36}{11/36} = \frac{2}{11}$$

15.3 Példa. Egy ismerősünket keressük az egyetemen, aki 5 különböző teremben lehet egyforma valószínűséggel. Annak valószínűsége, hogy egyáltalalán az egyetemen van $0 < p < 1$. Az 5 terem közül már 4-ben kerestük, de egyikben sem volt. Mi a valószínűsége, hogy az 5-ik teremben van?

Jelentse A_k azt az eseményt, hogy az ismerősünk a k-ik teremben van ($k = 1, \ldots 5$), ekkor $P(A_1 \cup \ldots \cup A_5) = p$. Mivel az A_k események egymást páronként kizárják, innen $P(A_k) = p/5$ mindegyik k indexre. Tehát a De Morgan azonosság alapján:

$$\begin{aligned} P(A_5|\overline{A_1} \cap \ldots \cap \overline{A_4}) &= P(A_5|\overline{A_1 \cup \ldots \cup A_4}) \\ &= \frac{P(A_5 \cap (\overline{A_1 \cup \ldots \cup A_4}))}{P(\overline{A_1 \cup \ldots \cup A_4})} \end{aligned}$$

Nyilvánvaló, hogy

$$A_5 \subset \overline{A_1 \cup \ldots \cup A_4}$$

és ezért

$$P(A_5 \cap (\overline{A_1 \cup \ldots \cup A_4})) = P(A_5)$$

Tehát a kérdéses valószínűség:

$$\begin{aligned} P(A_5|\overline{A_1} \cap \ldots \cap \overline{A_4}) &= P(A_5|\overline{A_1 \cup \ldots \cup A_4}) \\ &= \frac{P(A_5 \cap (\overline{A_1 \cup \ldots \cup A_4}))}{P(\overline{A_1 \cup \ldots \cup A_4})} \\ &= \frac{P(A_5)}{P(\overline{A_1 \cup \ldots \cup A_4})} = \frac{p/5}{1 - 4p/5} = \frac{p}{5 - 4p} \end{aligned}$$

15.2. Függetlenség

Tekintsük a következő nagyon egyszerű példát. Két kockával dobunk, de az eredményt nem látjuk. Valaki megmondta, hogy az első dobás páratlan. Mi a valószínűsége, hogy a dobott számok összege 7?

Jelentse A és B a következő eseményeket:

$$A = \{\text{az összeg } 7\} \qquad B = \{\text{az első dobás páratlan}\}$$

Ekkor a feltételes valószínűség definíciója szerint

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{3/36}{18/36} = \frac{1}{6}$$

Ez a korábbi példánkra tekintettel azt jelenti, hogy

$$P(A|B) = P(A)$$

azaz "a B esemény bekövetkezése nem befolyásolja az A valószínűségét". Ezt úgy fogalmazzuk meg, hogy A független a B eseménytől.

Világos, hogy $P(B) \neq 0$ esetén a $P(A|B) = P(A)$ feltétel ekvivalens a következővel:

$$P(A \cap B) = P(A) \cdot P(B) \tag{15.1}$$

Mivel úgy képzeljük, hogy a függetlenség szimmetrikus reláció (azaz ha A független B-től, akkor B is A-tól), és a fenti egyenlőség esetében ez nyilvánvalóan látható, célszerű ez utóbbi egyenlőséget a függetlenség definíciójául használni.

15.4 Definíció. Legyen $(\Omega, \mathcal{A}, P)$ adott valószínűségi mező és $A, B \in \mathcal{A}$ megfigyelhető események. Azt mondjuk, hogy A és B *függetlenek*, ha fennáll a (15.1) egyenlőség.

15.5 Példa. Egy 52 lapos kártyapakliból kihúzunk egymás után visszatevéssel két lapot. Mi a valószínűsége, hogy elsőre treffet, másodikra ászt húzunk?

Vezessük be a következő eseményeket:

$$A = \{\text{elsőre treffet húzunk}\} \qquad B = \{\text{másodikra ászt húzunk}\}$$

Ekkor

$$P(A \cap B) = \frac{13 \cdot 4}{52^2} = \frac{13}{52} \cdot \frac{4}{52} = P(A) \cdot P(B)$$

azaz az A és B események függetlenek.

Figyelem! Soha nem úgy érvelünk, hogy mivel A és B "láthatóan" függetlenek, azért $P(A \cap B) = P(A) \cdot P(B)$. Fordítva: az események függetlenségére úgy következtetünk, hogy ezt az egyenlőséget ellenőrizzük!

15.3. Teljes valószínűség tétele

15.6 Példa. Három egyforma boríték van előttünk,

1. az elsőben 2 darab ezres és 3 kétezres bankjegy,

2. a másodikban 5 ezres és 2 kétezres,
3. a harmadikban 5 kétezres bankjegy.

Véletlenszerűen kiválasztunk egy borítékot, majd abból találomra kihúzunk egy bankjegyet. Mi a valószínűsége, hogy kétezrest húzunk?

Jelentse A azt az eseményt, hogy kétezrest húzunk. Nyilván $P(A)$ könnyen meghatározható lenne, ha tudnánk, melyik borítékot választottuk. Nevezetesen, ha B_k jelöli azt az eseményt, hogy a k-ik borítékot választottuk, akkor a $P(A|B_k)$ feltételes valószínűségek rendre $3/5$, $2/7$ és 1.

Ez az észrevétel azonnal útmutatást is ad a megoldásra, ugyanis a B_k események egymást páronként kizárják, és egyesítésük a biztos esemény. Tehát:

$$A = A \cap \Omega = A \cap (B_1 \cup B_2 \cup B_3) = (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3)$$

Mivel a jobb oldali események egymást páronként kizárják:

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3) \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) \\ &= \frac{3}{5} \cdot \frac{1}{3} + \frac{2}{7} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} \end{aligned}$$

Ez a gondolatmenet nyilván továbbvihető tetszőleges számú B_k eseményre is, ezért vezessük be a következő definíciót.

15.7 Definíció. Azt mondjuk, hogy a $B_1, B_2, \ldots \in \mathcal{A}$ megfigyelhető események *teljes eseményrendszert* alkotnak, ha egyik sem nulla valószínűségű, továbbá

1. egymást páronként kizárják, azaz $B_i \cap B_j = \emptyset$ ha $i \neq j$,
2. egyikük biztosan bekövetkezik, azaz $B_1 \cup B_2 \cup \ldots = \Omega$.

A 15.6 Példában alkalmazott gondolatmenetet követve tetszőleges számú B_k eseményre adódik a következő tételünk.

15.8 Tétel. **(Teljes valószínűség tétele)** *Tegyük fel, hogy az* $(\Omega, \mathcal{A}, P)$ *valószínűségi mezőben a* $B_1, B_2, \ldots$ *nem nulla valószínűségű események teljes eseményrendszert alkotnak. Akkor tetszőleges* $A \in \mathcal{A}$ *eseményre*

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots$$

Bizonyítás. Valóban, ha a B_k események teljes eseményrendszert alkotnak, akkor

$$A = (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \cup \ldots$$

ahol az unió tagjai egymást páronként kizárják. Innen

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3) + \ldots$$

Mivel a feltételes valószínűség definíciója alapján minden k indexre

$$P(A \cap B_k) = P(A|B_k) \cdot P(B_k)$$

innen azonnal adódik a tétel állítása. □

15.9 Példa. Ha például annak valószínűsége, hogy egy telefonközpontba egy adott napon n hívás érkezik $0 < q_n < 1$, és mindegyik hívás adott $0 < p < 1$ valószínűséggel téves, akkor mi a valószínűsége, hogy ezen a napon éppen k számú téves hívás érkezik?

Vezessük be a következő jelöléseket. Jelentse A azt az eseményt, hogy központba k számú téves hívás érkezik, illetve B_n azt az eseményt, hogy az adott napon a beérkező hívások száma éppen n. Ekkor a B_n események teljes eseményrendszert alkotnak, ezért a teljes valószínűség tétele alapján

$$P(A) = \sum_{n=1}^{\infty} P(A|B_n) \cdot P(B_n) = \sum_{n=k}^{\infty} q_n \binom{n}{k} p^k (1-p)^{n-k}$$

ugyanis $n \geq k$ esetén a téves hívások egy Bernoulli-féle kísérlet eredményeként adódnak: n számú kísérletből hányszor következik be a téves hívás. Vegyük még figyelembe, hogy itt $P(A|B_n) = 0$, ha $n < k$.

15.4. Bayes-tétel

Térjünk vissza a 15.6 Példa vizsgálatára. Tegyük fel, hogy valaki elvégezte a húzást (mi nem láttuk), és közölte, hogy kétezres bankjegyet húzott. Mi a valószínűsége, hogy az első borítékból húzta?

Az ottani jelöléseinket használva a $P(B_1|A)$ feltételes valószínűségről van szó.

$$P(B_1|A) = \frac{P(A \cap B_1)}{P(A)} = \frac{P(A|B_1)P(B_1)}{P(A)}$$

Az utóbbi tört nevezője a Teljes valószínűség tételével határozható meg. Tehát:

$$\begin{aligned} P(B_1|A) &= \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)} \\ &= \frac{\frac{3}{5} \cdot \frac{1}{3}}{\frac{3}{5} \cdot \frac{1}{3} + \frac{2}{7} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} \end{aligned}$$

Ez a gondolatmenetünk általánosítható tetszőleges elemszámú teljes eseményrendszerre is.

15.10 Tétel. **(Bayes-tétel)** *Tegyük fel, hogy az $(\Omega, \mathcal{A}, P)$ valószínűségi mezőben a $B_1, B_2, \ldots$ nem nulla valószínűségű események teljes eseményrendszert alkotnak. Akkor tetszőleges $A \in \mathcal{A}$, $P(A) \neq 0$ eseményre és i indexre*

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots}$$

Bizonyítás. Valóban, a feltételes valószínűség definíciója alapján

$$P(B_i|A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{P(A)},$$

és innen a teljes valószínűség tétele alapján adódik az állítás. □

15.11 Példa. Például a telefonközpontos 15.9 példánkban annak valószínűsége, hogy az adott napon i hívás érkezett feltéve, hogy éppen k téves hívást regisztráltak $i \geq k$ esetén:

$$P(B_i|A) = \frac{q_i \binom{i}{k} p^k (1-p)^{i-k}}{\sum_{n=k}^{\infty} q_n \binom{n}{k} p^k (1-p)^{n-k}}$$

míg $i < k$ esetén ez a feltételes valószínűség 0.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-2 III/4 és III/5 szakaszainak feldolgozása.
2. Házi feladatok: a III/4 szakasz 228, 229, 232, 233, 237, 242, 251, 257 továbbá a III/4 szakasz 264, 265, 269, 270, 280 feladatai.
3. Tankönyv-2 3.6, 3.7, 3.8 és 3.9 szakaszai. Továbbá: A KÖZÉPISKOLAI KOMBINATORIKA ALAPOS ÁTISMÉTLÉSE!

16. fejezet

Valószínűségi változók és eloszlások

16.1. Valószínűségi változók

16.1 Definíció. Tekintsünk egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőt. Az

$$X : \Omega \to \mathbb{R}$$

függvényt *valószínűségi változónak* nevezzük, ha bármely $x \in \mathbb{R}$ esetén

$$\{X < x\} = \{\omega \in \Omega : X(\omega) < x\} \in \mathcal{A}$$

azaz minden ilyen nívóhalmaz megfigyelhető (és így van valószínűsége).

Az alábbi példákban állapítsuk meg a megadott valószínűségi változók R értékkészletét!

16.2 Példa.

1. Két kockával dobunk. Jelentse X a dobott számok összegét. Ekkor $R = \{2, 3, \ldots, 12\}$
2. Jelentse X a hagyományos lottón kihúzott öt nyerő szám közül a legkisebbet. Ekkor $R = \{1, 2, \ldots, 86\}$
3. Egy kockával addig dobunk, amíg először 6-os jön ki. Jelentse X a dobások számát. Ekkor $R = \mathbb{N}$.
4. Válasszunk az origó középpontú egységsugarú körlemezben véletlenszerűen egy pontot. Jelentse X a pont origótól mért távolságát. Ekkor $R = [0, 1]$.

16.3 Definíció. Egy valószínűségi változót *diszkrétnek* nevezünk, ha az értékkészlete véges vagy végtelen sorozatba rendezhető (azaz megszámlálható halmaz).

Példáinkban az első három változó diszkrét, míg a negyedik nem az.

16.2. Diszkrét valószínűségi változó eloszlása

16.4 Definíció. Legyen X diszkrét valószínűségi változó, amelynek értékkészlete $R = \{x_1, x_2, \ldots\}$. Ekkor a

$$p_k = P(X = x_k)\,, \qquad k = 1, 2, \ldots$$

sorozatot az X *eloszlásának* nevezzük.

16.5 Példa. Tekintsük a bevezető példáinkat diszkrét valószínűségi változókra.

1. Ha X a két dobott szám összegét jelenti, akkor az eloszlás az alábbi *táblázattal* adható meg:

x_k	2	3	4	...	12
p_k	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	...	$\frac{1}{36}$

2. Ha X a legkisebb kihúzott lottószámot jelenti, akkor az eloszlás az alábbi *formulával* adható meg:

$$p_k = \frac{\binom{90-k}{4}}{\binom{90}{5}} \qquad k = 1, 2, \ldots 86$$

3. Ha X az első 6-osig szükséges dobások száma, akkor X eloszlása:

$$p_k = \left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6} \qquad k = 1, 2, \ldots$$

   Az előző két példától eltérően az eloszlás itt végtelen sorozat.

Az eloszlás legfontosabb tulajdonságait a következő tételben foglaljuk össze.

16.6 Tétel. *Tekintsünk egy X diszkrét valószínűségi változót, amelynek értékkészlete $R = \{x_1, x_2, \ldots\}$ és eloszlása $p_k = P(X = x_k)$, $k = 1, 2, \ldots$. Ekkor*

- $0 \le p_k \le 1$ *minden* $k = 1, 2, \ldots$ *indexre.*

- $p_1 + p_2 + \ldots = 1$.
- *Ha* $a < b$ *tetszőleges valós számok, akkor*

$$P(a < X < b) = \sum_{a < x_k < b} p_k$$

ahol mindazon p_k *valószínűségeket összegezzük, amelyek indexeire fennáll az* $a < x_k < b$ *egyenlőtlenség. Itt bármelyik szigorú egyenlőtlenség helyett mindkét oldalon egyszerre* $\leq$ *is használható.*

16.3. Az eloszlásfüggvény

16.7 Definíció. Tekintsünk egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőt és egy $X : \Omega \to \mathbb{R}$ valószínűségi változót. Minden $x \in \mathbb{R}$ esetén legyen

$$F(x) = P(X < x)$$

Az így értelmezett $F : \mathbb{R} \to [0, 1]$ függvényt az X *eloszlásfüggvényének* nevezzük.

16.8 Példa. Könnyen ellenőrizhető, hogy a körlemezes példánkban (bevezető 4. példa) az X változó eloszlásfüggvénye

$$F(x) = \begin{cases} 0 & \text{ha } x \leq 0 \\ x^2 & \text{ha } 0 < x \leq 1 \\ 1 & \text{ha } x > 1 \end{cases} \tag{16.1}$$

Úgy képzeljük ugyanis, hogy annak valószínűsége, hogy a véletlenszerűen választott pont a körlemez egy adott részhalmazában van, arányos a részhalmaz területével. Nevezetesen például $P(0 \leq X < 1/2) = 1/4$.

A valószínűségszámítás és a statisztika számos feladatában van szükség valamely $P(a \leq X < b)$ alakú valószínűség meghatározására. Ehhez nyújt segítséget az eloszlásfüggvény, amelynek tulajdonságait az alábbi tételben foglaljuk össze.

16.9 Tétel. *Tekintsünk egy* X *valószínűségi változót és jelölje* F *az* X *eloszlásfüggvényét.*

- *Minden* $x \in \mathbb{R}$ *esetén* $0 \leq F(x) \leq 1$.
- F *monoton növő és minden pontban balról folytonos.*

- $\lim_{x\to-\infty} F(x) = 0$ *és* $\lim_{x\to+\infty} F(x) = 1$.
- *Ha $a < b$ tetszőlegesen adott valós számok, akkor*

$$P(a \leq X < b) = F(b) - F(a)\,.$$

Ha egy X diszkrét valószínűségi változó értékkészlete $R = \{x_1, x_2, \ldots\}$ ahol $x_1 < x_2 < \ldots$, és X ezeket az értékeket rendre $p_1, p_2, \ldots$ valószínűségekkel veszi fel, akkor az X eloszlásfüggvénye ilyen alakú:

$$F(x) = \begin{cases} 0 & \text{ha } x \leq x_1 \\ p_1 + \ldots + p_k & \text{ha } x_k < x \leq x_{k+1} \end{cases}$$

bármely $k = 1, 2, \ldots$ esetén. Készítsünk grafikont!

Ez azt jelenti, hogy ilyen esetben az eloszlásfüggvény szakaszonként konstans. Célszerű ilyenkor az $P(a \leq X < b) = F(b) - F(a)$ képlet alkalmazása helyett összegyűjteni az X értékkészletének mindazon elemeit, amelyek a és b közé esnek. Nevezetesen ha $P(X = x_k) = p_k$ minden k-ra:

$$P(a \leq X < b) = \sum_{a \leq x_k < b} p_k$$

Mivel itt csak a $P(X = x_k)$ valószínűségek szerepelnek, ezért kényelmesebb az X eloszlására hagyatkozni.

16.4. A sűrűségfüggvény

16.10 Definíció. Azt mondjuk, hogy X *folytonos eloszlású*, ha található olyan f integrálható függvény a számegyenesen, amelyre

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

minden $x \in \mathbb{R}$ mellett. Ilyenkor az f függvényt az X *sűrűségfüggvényének* nevezzük.

Például a (16.1) példában könnyen ellenőrizhető, hogy

$$f(t) = \begin{cases} 2t & \text{ha } 0 < t < 1 \\ 0 & \text{különben} \end{cases}$$

Ha az X valószínűségi változó folytonos eloszlású, akkor az F eloszlásfüggvény folytonos, és minden olyan x pontban, ahol az f sűrűségfüggvény folytonos F differencálható és:

$$F'(x) = f(x)$$

16.11 Tétel. *Ha X folytonos eloszlású és f a sűrűségfüggvénye, akkor bármely $a < b$ esetén*

$$P(a \leq X < b) = \int_a^b f(t)\,dt$$

Vajon milyen valószínűséggel vesz fel egy X valószínűségi változó egyetlen pontot? Legyen $a \in \mathbb{R}$ tetszőleges, ekkor belátható, hogy

$$\begin{aligned} P(X = a) &= P(\bigcap_{n=1}^{\infty} \{a \leq X < a + \frac{1}{n}\}) = \lim_{n\to\infty} P(a \leq X < a + \frac{1}{n}) \\ &= \lim_{n\to\infty} (F(a + \frac{1}{n}) - F(a)) = \lim_{x\to a+} F(x) - F(a) \end{aligned}$$

Tehát $P(X = a)$ az F a-pontbeli "ugrásával" egyenlő. **FIGYELEM:** Miért térhetünk limeszre a levezetés első sorában?

Előző gondolatmenetünk azt jelenti, hogy $P(X = a) = 0$ akkor és csak akkor áll fenn, ha F folytonos az a pontban. Nevezetesen, ha X folytonos eloszlású, akkor F folytonos az egész számegyenesen, tehát bármely $a < b$ valós számokra

$$P(a < X < b) = P(a \leq X \leq b)$$

Foglaljuk össze a sűrűségfüggvény tulajdonságait.

16.12 Tétel. *Ha f az X valószínűségi változó sűrűségfüggvénye, akkor*

1. *$f(x) \geq 0$ minden $x \in \mathbb{R}$ esetén*
2. $$\int_{-\infty}^{+\infty} f(x)\,dx = 1$$
3. *ha $a < b$ tetszőleges valós számok, akkor*
$$P(a < X < b) = P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

16.13 Példa. Legyen például az X sűrűségfüggvénye

$$f(x) = \begin{cases} x & \text{ha } 0 < x \leq 1 \\ 2 - x & \text{ha } 1 < x < 2 \\ 0 & \text{különben} \end{cases}$$

Könnyen ellenőrizhetjük, hogy f valóban kielégíti az előző tétel feltételeit. Ekkor például

$$\begin{aligned} P(0 \le X \le 3/2) &= P(0 < X < 3/2) = \int_0^{3/2} f(x)\,dx \\ &= \int_0^1 x\,dx + \int_1^{3/2} (2-x)\,dx \\ &= 1 - \int_{3/2}^2 (2-x)\,dx = \frac{7}{8} \end{aligned}$$

Otthoni tanuláshoz

1. A Feladatgyűjtemény-2 IV/1 szakaszának feldolgozása.
2. Házi feladatok: a IV/1 szakasz 299, 303, 306, , 314, 316, 323 és 324 feladatai.
3. Tankönyv-2 4.1, 4.2, 4.3, 4.4 szakaszai.

17. fejezet

A várható érték és a szórás

Hétköznapi szóhasználattal egy valószínűségi változó várható értéke az átlagértékét, a szórása pedig az átlagtól való átlagos eltérést jelenti. A pontos definíciókat az alábbiakban fogalmazzuk meg.

17.1. Diszkrét eloszlások várható értéke

17.1 Definíció. Tekintsünk egy X diszkrét valószínűségi változót, amelynek eloszlása

$$P(X = x_k) = p_k \qquad k = 1, 2, \ldots$$

Azt mondjuk, hogy az X-nek van várható értéke, ha a $\sum_{k=1}^{\infty} |x_k| \cdot p_k$ sor konvergens, és ekkor az

$$E(X) = \sum_{k=1}^{\infty} x_k \cdot p_k$$

sorösszeget az X *várható értékének* nevezzük.

Megjegyezzük, hogy a $\sum_{k=1}^{\infty} |x_k| \cdot p_k$ sor konvergenciája azért fontos feltétel, mert ellenkező esetben $E(X)$ értéke függhetne a sor tagjainak sorrendjétől.

17.2 Példa. Egy kockát feldobunk kétszer egymás után. Határozzuk meg a dobott számok összegének várható értékét.

Jelentse X a dobott számok összegét, akkor X eloszlását a 16.5 Példában láthatjuk. Tehát a definíció szerint a várható érték:

$$E(X) = \sum_{k=2}^{12} k p_k = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + \ldots + 12 \cdot \frac{1}{36} = 7\,.$$

17.3 Példa. Egy 52 lapos kártyapakliból véletlenszerűen válasszunk ki 5 lapot. Adjuk meg a kihúzott treffek számának várható értékét.

Jelentse X a kihúzott treffek számát, akkor a visszatevés nélküli mintavételi feladat alapján az X eloszlása:

$$P(X=k)=\frac{\binom{13}{k}\cdot\binom{39}{5-k}}{\binom{52}{5}}\qquad k=0,\dots,5$$

Ennek alapján a várható érték:

$$\begin{aligned} E(X) &= \sum_{k=0}^{5} kP(X=k)=\sum_{k=0}^{5} k\frac{\binom{13}{k}\cdot\binom{39}{5-k}}{\binom{52}{5}} \\ &= \frac{13}{\binom{52}{5}}\sum_{k=1}^{5}\binom{12}{k-1}\binom{39}{4-(k-1)}=\frac{13}{\binom{52}{5}}\cdot\binom{51}{4}=\frac{5}{4}. \end{aligned}$$

17.4 Példa. Tekintsük a 14.4 szakaszban vizsgált Bernoulli-kísérletet, és határozzuk meg, hogy n kísérletből várhatóan hányszor következik be az A esemény.

Jelentse X az A esemény bekövetkezéseinek számát, ekkor X eloszlása:

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}\qquad k=0,1,\dots,n$$

Tehát X várható értéke a binomiális tétel alapján

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} \\ &= np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}(1-p)^{n-k} \\ &= np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k}=np \end{aligned}$$

17.2. Végtelen elemű eloszlások várható értéke

Ebben a szakaszban olyan példákat vizsgálunk, ahol a diszkrét valószínűségi változó értékkészlete végtelen halmaz.

17.5 Példa. Egy kockát ismételten addig dobunk, amíg 6-os nem jön ki. Mennyi a dobások számának várható értéke?

Jelentse X a dobások számát, akkor X eloszlása

$$P(X=k)=\left(\frac{5}{6}\right)^{k-1}\cdot\frac{1}{6}\qquad k=1,2,\ldots$$

Tehát a várható érték

$$E(X)=\sum_{k=1}^{\infty}k\cdot\left(\frac{5}{6}\right)^{k-1}\cdot\frac{1}{6}=\frac{1}{6}\cdot\frac{1}{(1-5/6)^2}=6$$

17.6 Példa. Legyen λ adott pozitív szám, és X olyan valószínűségi változó, amelynek eloszlása

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}\qquad k=0,1,2,\ldots$$

Ekkor az exponenciális függvény hatványsora alapján az X várható értéke:

$$\begin{aligned}E(X) &= \sum_{k=0}^{\infty}k\frac{\lambda^k}{k!}e^{-\lambda}=\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}\\ &= \lambda e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!}=\lambda e^{-\lambda}e^{\lambda}=\lambda\end{aligned}$$

17.7 Példa. Egy dobozban 1 fekete és 1 fehér golyó van. Találomra kiveszünk egyet. Ha fekete, akkor visszatesszük, és hozzáteszünk még egy fekete golyót. A húzásokat addig folytatjuk, amíg a fehér golyót ki nem húzzuk. Adjuk meg a húzások számának várható értékét.

Jelentse X a húzások számát, akkor X eloszlása így adható meg: $P(X=1)=1/2$, és

$$P(X=k)=\frac{1}{2}\cdot\frac{2}{3}\cdot\frac{3}{4}\cdots\frac{k-1}{k}\cdot\frac{1}{k+1}=\frac{1}{k(k+1)},\quad k=2,3,\ldots$$

Tehát az X változó várható értékére az alábbi végtelen sor adódik:

$$E(X)=\sum_{k=1}^{\infty}kP(X=k)=\sum_{k=1}^{\infty}k\frac{1}{k(k+1)}=\sum_{k=1}^{\infty}\frac{1}{k+1}$$

Ez utóbbi sor azonban az első tag híján pontosan megegyezik a harmonikus sorral, ami divergens. Tehát ennek a valószínűségi változónak nincs várható értéke.

17.3. Folytonos eloszlások várható értéke

17.8 Definíció. Legyen X olyan folytonos eloszlású valószínűségi változó, amelynek sűrűségfüggvénye f. Azt mondjuk, hogy az X-nek van várható értéke, ha az $\int_{-\infty}^{\infty} |x| \cdot f(x)\,dx$ impropríus integrál konvergens, és ekkor az

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$$

integrál értékét az X *várható értékének* nevezzük.

17.9 Példa. Gondoljuk meg például, hogy az alábbi függvény sűrűségfüggvényt definiál

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \qquad -\infty < x < \infty$$

(ez az úgynevezett Cauchy-eloszlás), de nincs várható értéke, hiszen az

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x}{1+x^2}\,dx$$

impropríus integrál divergens, lásd a 9.5 Példát.

17.10 Példa. Legyen $[a, b]$ egy adott intervallum a számegyenesen, és tegyük fel, hogy az X valószínűségi változó sűrűségfüggvénye

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{ha } a < x < b \\ 0 & \text{különben} \end{cases}$$

Ellenőrizzük, hogy f valóban sűrűségfüggvény! Ekkor az X várható értéke

$$E(X) = \int_a^b \frac{x}{b-a}\,dx = \frac{1}{b-a} \cdot \frac{b^2-a^2}{2} = \frac{a+b}{2}$$

ami éppen az $[a, b]$ intervallum felezőpontja.

17.4. A várható érték tulajdonságai

Az X változó második momentumának nevezzük az $E(X^2)$ várható értéket (ha ez létezik). Megmutatható (FIGYELEM: helyettesítéses integrállal), hogy

$$E(X^2) = \begin{cases} \sum_k x_k^2 p_k & \text{ha } X \text{ diszkrét eloszlású} \\ \int_{-\infty}^{\infty} x^2 f(x)\,dx & \text{ha } X \text{ folytonos eloszlású} \end{cases}$$

Az alábbiakban összefoglaljuk a várható érték két fontos tulajdonságát.

17.11 Tétel.

1. *Ha az X változónak van várható értéke, akkor tetszőleges α és β valós számokra $E(\alpha X+\beta)=\alpha E(X)+\beta$.*
2. *Ha léteznek $E(X)$, $E(X^2)$, úgy $E(\alpha X^2+\beta X+\gamma)=\alpha E(X^2)+\beta E(X)+\gamma$.*

17.12 Példa. Legyen λ adott pozitív szám, és tegyük fel, hogy az X valószínűségi változó sűrűségfüggvénye

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & \text{ha } x>0\\ 0 & \text{különben.}\end{cases}$$

A 9.3 Példa alapján ez valóban sűrűségfüggvény, hiszen

$$\int_0^\infty f(x)\,dx=1\,,$$

másrészt a 9.8 Példa szerint

$$E(X)=\int_0^\infty xf(x)\,dx=\frac{1}{\lambda}\,.$$

Ennek a változónak a második momentuma a 9.9 Példa alapján

$$E(X^2)=\int_0^\infty x^2f(x)\,dx=\frac{2}{\lambda^2}\,.$$

17.5. A variancia és a szórás

Egy változó varianciáján a várható értéktől mért átlagos négyzetes eltérést értjük.

17.13 Definíció. Az X változó *varianciája* (ha ez létezik)

$$Var(X)=E((X-E(X))^2)$$

Ekkor az X változó *szórása:* $D(X)=\sqrt{Var(X)}$.

A variancia helyett néha használatos a szórásnégyzet elnevezés, és a $D^2(X)$ jelölés is.

A variancia a következő módon számítható ki egyszerűbb alakban:

$$\begin{aligned} Var(X) &= E((X - E(X))^2) = E(X^2 - 2E(X)X + E(X)^2) \\ &= E(X^2) - 2E(X)^2 + E(X)^2 = E(X^2) - E(X)^2 \end{aligned}$$

Fontos tulajdonságai:

$$Var(\alpha X + \beta) = \alpha^2 Var(X), \qquad D(\alpha X + \beta) = |\alpha| \cdot D(X)$$

Ellenőrizzük ezeket közvetlenül a definíció alapján.

17.14 Példa. Példaként határozzuk meg adott $\lambda > 0$ mellett a 17.12 Példában vizsgált folytonos eloszlású valószínűségi változó varianciáját és szórását.

$$Var(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

nevezetesen

$$D(X) = \frac{1}{\lambda}$$

17.15 Példa. Tekintsük most a 17.10 Példában vizsgált folytonos eloszlású valószínűségi változót. Ennek második momentuma a következőképpen számítható ki:

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_a^b \frac{x^2}{b-a}\, dx = \frac{1}{b-a} \left[\frac{x^3}{3}\right]_a^b \\ &= \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3} \end{aligned}$$

Tehát a varianciára az alábbi adódik:

$$Var(X) = E(X^2) - E(X)^2 = \frac{b^2 + ab + a^2}{3} - \frac{a^2 + 2ab + b^2}{4} = \frac{(b-a)^2}{12}$$

továbbá az X szórása ennek négyzetgyöke, azaz

$$D(X) = \frac{b-a}{2\sqrt{3}}\,.$$

Otthoni tanuláshoz

1. A Feladatgyűjtemény-2 IV/2 szakaszának feldolgozása.
2. Házi feladatok: a IV/2 szakasz 351, 352, 354, 355, 358, 361, 369, 374 és 375 feladatai.
3. Tankönyv-2 4.6 és 4.7 szakaszai.

18. fejezet

Nevezetes diszkrét eloszlások

Ebben a fejezetben a gyakorlatban leginkább előforduló diszkrét eloszlások tulajdonságait foglaljuk össze.

18.1. Karakterisztikus eloszlás

Legyen $(\Omega, \mathcal{A}, P)$ egy valószínűségi mező és tekintsünk egy $A \in \mathcal{A}$ eseményt, amelyre $P(A) = p$, ahol $0 < p < 1$. Ekkor az

$$X = \begin{cases} 1 & \text{ha } A \text{ bekövetkezik} \\ 0 & \text{ha } A \text{ nem következik be} \end{cases}$$

valószínűségi változó eloszlása

$$P(X = 0) = 1 - p \quad P(X = 1) = p$$

Ezt az A eseményhez tartozó *karakterisztikus eloszlásnak* nevezzük. Könnyen ellenőrizhetjük a következő tulajdonságokat.

18.1 Tétel.

- *Az eloszlás paramétere:* $0 < p < 1$.
- *Az eloszlás várható értéke:* $E(X) = p$
- *Az eloszlás varianciája:* $Var(X) = p(1 - p)$.

Bizonyítás. Csak a varianciát kell ellenőriznünk. Mivel a második momentum $E(X^2) = p$, innen azonnal adódik az állítás. $\square$

18.2. Binomiális eloszlás

Legyen újra $(\Omega, \mathcal{A}, P)$ egy valószínűségi mező és tekintsük a Bernoulli-kísérletet, ahol egymás után n független kísérletet végzünk, amelyek mindegyikében megfigyeljük, hogy egy bizonyos A esemény bekövetkezik-e vagy sem. Tegyük fel, hogy $P(A) = p$, $0 < p < 1$ adott. Jelentse X az A bekövetkezéseinek számát. A Bernoulli-kísérletről mondottak alapján az X változó eloszlása:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, 2, \ldots, n$$

ezt az eloszlást *binomiális eloszlásnak* nevezzük. Tulajdonságai az alábbiak (lásd a 17.4 Példát).

18.2 Tétel.

- *Az eloszlás paraméterei:* $n \in \mathbb{N}$ *és* $0 < p < 1$.
- *Az eloszlás várható értéke:* $E(X) = np$
- *Az eloszlás varianciája:* $Var(X) = np(1-p)$.

Bizonyítás. A 17.4 Példa szerint csak a varianciát kell ellenőriznünk. Ehhez meghatározzuk a második momentumot.

$$\begin{aligned}
E(X^2) &= \sum_{k=1}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} = \\
&= \sum_{k=2}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k} + \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} (1-p)^{n-k} + np = (n^2 - n)p^2 + np\,.
\end{aligned}$$

Tehát a variancia

$$Var(X) = E(X^2) - E(X)^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p)$$

ahol felhasználtuk, hogy a második sorban a második szumma éppen a várható érték. □

18.3. Hipergeometriai eloszlás

Nézzük most a következő visszatevés nélküli mintavételi kísérletet. Tekintsünk egy N számú objektumból álló halmazt, amelyek közül m számú selejtes. Válasszunk ki az egész halmazból véletlenszerűen egy n-elemű mintát visszatevés nélkül ($n \leq m$). Jelentse X a selejtes elemek számát a mintában. Ekkor X eloszlása:

$$P(X=k) = \frac{\binom{m}{k} \cdot \binom{N-m}{n-k}}{\binom{N}{n}} \qquad k = 0, 1, 2, \ldots, n$$

Ezt az eloszlást *hipergeometriai eloszlásnak* nevezzük. Ennek tulajdonságait az előző fejezet példája alapján a következő listában foglaljuk össze.

18.3 Tétel.

- *Az eloszlás paraméterei:* $N, m, n \in \mathbb{N}$.
- *Az eloszlás várható értéke:*

$$E(X) = n \cdot \frac{m}{N}$$

- *Az eloszlás varianciája:*

$$Var(X) = \frac{N-n}{N-1} \cdot n \cdot \frac{m}{N}\left(1 - \frac{m}{N}\right).$$

Bizonyítás. A 17.3 Példa gondolatmenetére tekintettel ismét csak a variancia szorul igazolásra. Számítsuk ki a második momentumot.

$$\begin{aligned} E(X^2) &= \sum_{k=1}^{n} k^2 \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} = \sum_{k=2}^{n} k(k-1)\frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} + \sum_{k=1}^{n} k\frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} \\ &= \frac{m(m-1)n(n-1)}{N(N-1)} \sum_{k=2}^{n} \frac{\binom{m-2}{k-2} \cdot \binom{N-m}{n-k+2}}{\binom{N-2}{n-2}} + n\frac{m}{N} \\ &= \frac{m(m-1)n(n-1)}{N(N-1)} + n\frac{m}{N}. \end{aligned}$$

Innen adódik, hogy

$$Var(X) = \frac{m(m-1)n(n-1)}{N(N-1)} + n\frac{m}{N} - n^2\frac{m^2}{N^2} = \frac{N-n}{N-1} \cdot n \cdot \frac{m}{N}\left(1 - \frac{m}{N}\right),$$

amit állítottunk. □

18.4. Geometriai eloszlás

Tekintsünk újra egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőt, és benne egy A eseményt, amelyre $P(A) = p$, ahol $0 < p < 1$ adott. A kísérletet egymás után függetlenül egészen addig ismételten elvégezzük, amíg az A esemény először bekövetkezik. Jelentse X a kísérletek számát. Ekkor X eloszlása:

$$P(X=k) = (1-p)^{k-1}p \qquad k = 1, 2, \ldots$$

Ezt az eloszlást *geometriai eloszlásnak* nevezzük.

18.4 Tétel.

- *Az eloszlás paramétere:* $0 < p < 1$.
- *Az eloszlás várható értéke:*
$$E(X) = \frac{1}{p}$$
- *Az eloszlás varianciája:*
$$Var(X) = \frac{1-p}{p^2}\,.$$

Bizonyítás. Mivel a várható érték a 17.5 Példa gondolatmenetét követve könnyen adódik, csak a varianciát kell igazolnunk. Esetünkben a második momentum a következő módon számolható ki. A hatványsor második deriváltját használva bármely $|x| < 1$ mellett

$$\sum_{k=2}^{\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3}$$

Ezt az azonosságot az $x = 1-p$ esetben felhasználva

$$\begin{aligned} E(X^2) &= \sum_{k=1}^{\infty} k^2(1-p)^{k-1}p = \sum_{k=2}^{\infty} k(k-1)(1-p)^{k-1}p + \sum_{k=1}^{\infty} k(1-p)^{k-1}p \\ &= p(1-p)\sum_{k=1}^{\infty} k(k-1)(1-p)^{k-2} + \frac{1}{p} = \frac{2p(1-p)}{p^3} + \frac{1}{p}\,. \end{aligned}$$

Innen azt kapjuk, hogy

$$Var(X) = E(X^2) - E(X)^2 = \frac{2p(1-p)}{p^3} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

amit igazolni akartunk. □

18.5. Poisson-eloszlás

Tegyük fel, hogy X olyan valószínűségi változó, amelynek értékkészlete $\{0\} \cup \mathbb{N}$ és eloszlása

$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda} \qquad k = 0, 1, 2, \ldots$$

ahol $\lambda > 0$ adott valós szám.

Nem nehéz belátni, hogy valóban eloszlást definiáltunk, ugyanis

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \cdot e^{\lambda} = 1$$

az exponenciális függvény hatványsora alapján. Ezt a végtelen elemű eloszlást *Poisson-eloszlásnak* nevezzük.

18.5 Tétel.

- *Az eloszlás paramétere:* $\lambda > 0$.
- *Az eloszlás várható értéke:* $E(X) = \lambda$,
- *Az eloszlás varianciája:* $Var(X) = \lambda$.

Bizonyítás. A 17.6 Példa gondolatmenetét követve ismét csak a variancia szorul igazolásra. Vizsgáljuk meg a második momentumot.

$$\begin{aligned}
E(X^2) &= \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \sum_{k=1}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} + \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} + \lambda = \lambda^2 + \lambda\,.
\end{aligned}$$

Innen a varianciára az adódik, hogy

$$Var(X) = E(X^2) - E(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

és ezzel a bizonyítás kész. □

Megjegyezzük, hogy a Poisson-eloszlás a binomiális eloszlás "határeloszlásaként" származtatható az alábbi értelemben.

18.6 Tétel. *Ha adott* $\lambda > 0$ *és* $0 < p_n < 1$ *olyan sorozat, amelyre* $np_n = \lambda$, *akkor*

$$\lim_{n \to \infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$$

bármely $k = 0, 1, 2, \ldots$ *esetén.*

Bizonyítás. Valóban, a feltételünk szerint bármely rögzített k index mellett

$$\begin{aligned}
\binom{n}{k} p_n^k (1 - p_n)^{n-k} &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n(n-1)\ldots(n-k+1)}{n^k} \cdot \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}
\end{aligned}$$

Itt a négy tényező határértékeit egyenként vizsgálva könnyen láthatjuk, hogy azok rendre 1, $\lambda^k/k!$, $e^{-\lambda}$ és 1. Innen már adódik a tétel állítása. $\square$

Ez a gyakorlatban azt eredményezi, hogy nagy n értékekre és kicsi p értékekre a binomiális eloszlás jól közelíthető a Poisson-eloszlással,

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

bármely $0 \leq k \leq n$ természetes számra.

18.7 Példa. Tegyük fel, hogy annak valószínűsége, hogy egy újonnan gyártott Suzuki Vitara gépkocsiban hibás a légzsák 0.002 egymástól függetlenül. A gyár visszahívást rendel el, ha egy adott hónapban legyártott 2000 gépkocsiban legalább 10 esetben hibás a légzsák. Mi a valószínűsége, hogy nem kell visszahívást elrendelni?

Jelölje X a hibás gépkocsik számát egy adott hónapban. Mivel a kísérletünk egy Bernoulli-kísérlet (hiszen a meghibásodás valószínűsége minden autóban 0.002 egymástól függetlenül), ezért X binomiális eloszlású $n = 2000$ és $p = 0.002$ paraméterekkel. Tehát a keresett valószínűség pontos értéke

$$P(X \leq 9) = \sum_{k=0}^{9} \binom{2000}{k} 0.002^k 0.998^{2000-k}$$

amelynek kiszámítása elég reménytelen feladat. Fenti tételünk értelmében azonban közelítő értéket adhatunk a Poisson eloszlás használatával, hiszen "n elég nagy és p elég kicsi", továbbá $\lambda = np = 4$, így

$$\sum_{k=0}^{9} \binom{2000}{k} 0.002^k 0.998^{2000-k} \approx \sum_{k=0}^{9} \frac{4^k}{k!} e^{-4} \approx 0.9919$$

Ez utóbbi értéket a Poisson-eloszlásra vonatkozó táblázat alapján állapíthatjuk meg, amely a Feladatgyűjtemény-2 332-ik oldalán található.

Annak vizsgálata, hogy ez a közelítés mennyire pontos, meghaladja e könyv kereteit.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-2 VI/1, VI/2, VI/3, VI/4, VI/5 és VI/6 szakaszainak feldolgozása.

2. Házi feladatok: a VI/1 szakasz 533, 534, VI/2 szakasz 546, 558, 559, VI/3 szakasz 566, 568, 572, VI/4 szakasz 581, 582, 583, VI/5 szakasz 588, 598, 600, 603, VI/6 szakasz 622, 624, 630 és 632 feladatai. feladatai.

3. Tankönyv-2 4.8 és 4.9 szakaszai.

19. fejezet

Nevezetes folytonos eloszlások

19.1. Egyenletes eloszlás

Legyen $[a,b]$ adott intervallum. Tekintsük azt az X valószínűségi változót, amelynek sűrűségfüggvénye:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{ha } a < x < b \\ 0 & \text{különben} \end{cases}$$

Ekkor az X változót az $[a,b]$ intervallumon *egyenletes eloszlásúnak* nevezzük. Az elnevezés onnan ered, hogy ilyenkor az $[a,b]$ valamely részintervallumába esés valószínűsége a részintervallum hosszával arányos.

19.1 Tétel.

- *Az eloszlás paraméterei: a és b, $a < b$.*
- *Az eloszlás várható értéke:*

$$E(X) = \frac{a+b}{2}$$

- *Az eloszlás varianciája:*

$$Var(X) = \frac{(b-a)^2}{12}\,.$$

Bizonyítás. Állításaink azonnal következnek a 17.10 és 17.15 Példák eredményeiből. □

19.2 Példa. Legyen X olyan egyenletes eloszlású változó, amelyre $E(X) = 5$ és $Var(X) = 3$. Határozzuk meg a $P(4 < X < 10)$ valószínűséget!

Az intervallum ismeretlen a és b végpontjaira a feltételek szerint az

$$\begin{aligned} \frac{a+b}{2} &= 5 \\ \frac{(b-a)^2}{12} &= 3 \end{aligned}$$

adódik, amelynek megoldásai $a = 2$ és $b = 8$. Tehát

$$P(4 < X < 10) = P(4 < X < 8) = \frac{2}{3}$$

hiszen a $[4, 8]$ intervallumon túlnyúló rész 0 valószínűségű.

19.2. Exponenciális eloszlás

Legyen $\lambda > 0$ adott valós szám. Tekintsük azt a valószínűségi változót, amelynek sűrűségfüggvénye

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{ha } x > 0 \\ 0 & \text{különben} \end{cases}$$

Ekkor az X változót *exponenciális eloszlásúnak* nevezzük. A 9.3 Példában elmondottak szerint f valóban sűrűségfüggvény.

19.3 Tétel.

- *Az eloszlás paramétere:* $\lambda > 0$.
- *Az eloszlás várható értéke:* $E(X) = 1/\lambda$,
- *Az eloszlás varianciája:* $Var(X) = 1/\lambda^2$.

Bizonyítás. Tételünk közvetlenül következik a 17.12 és 17.14 Példák egyenlőségeiből. □

19.4 Példa. Tekintsünk egy X adott $\lambda > 0$-paraméterű exponenciális eloszlású változót. Határozzuk meg a $P(X > E(X))$ valószínűséget!

Mivel tételünk szerint $E(X) = 1/\lambda$, azért

$$P(X > E(X)) = P\left(X > \frac{1}{\lambda}\right) = \int_{1/\lambda}^{\infty} \lambda e^{-\lambda x}\, dx = \left[-e^{-\lambda x}\right]_{1/\lambda}^{\infty} = \frac{1}{e}\,.$$

Az exponenciális eloszlást *memória nélkülinek* nevezzük az alábbi értelemben. Ha X valamely λ-paraméterű exponenciális eloszlású változó, továbbá $t, s > 0$ tetszőlegesek, akkor

$$P(X > t + s | X > t) = P(X > s)\,.$$

Valóban, az $\{X > t+s\}$ eseményből következik $\{X > t\}$, ezért a bal oldalon álló feltételes valószínűség a következőképpen írható

$$\begin{aligned} P(X > t+s | X > t) &= \frac{P(X > t+s)}{P(X > t)} = \frac{1 - \int_0^{t+s} \lambda e^{-\lambda x}\, dx}{1 - \int_0^t \lambda e^{-\lambda x}\, dx} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = 1 - \int_0^s \lambda e^{-\lambda x}\, dx = P(X > s)\,. \end{aligned}$$

Ha például X a várakozási időt jelenti két bekövetkezés (pl. két telefonhívás, vagy két ügyfél, stb) között, akkor a memória-nélküliség azt jelenti, hogy a további várakozás ideje nem függ attól, hogy előzetesen mennyit vártunk.

Fordítva, az is igazolható, hogy ha egy folytonos eloszlás memória nélküli, akkor az szükségképpen az exponenciális eloszlás.

Érdekes kapcsolat van a Poisson-eloszlás és az exponenciális eloszlás között. Nevezetesen, ha a bekövetkezések közötti várakozási idők független, azonos λ-paraméterű exponenciális eloszlású változók, akkor egy egységnyi időintervallumban a bekövetkezések száma Poisson-eloszlású ugyanazon paraméterrel. Ezekkel a kérdésekkel a 22. fejezetben foglalkozunk részletesebben.

19.3. A standard normális eloszlás

A standard normális eloszlás centrális jelentősége miatt a valószínűségi változóra, annak sűrűség, illetve eloszlásfüggvényére külön jelölést használunk.

19.5 Definíció. Azt mondjuk, hogy a Z valószínűségi változó *standard normális eloszlású*, ha a sűrűségfüggvénye

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad -\infty < x < \infty$$

A (9.2) egyenlőség mutatja, hogy φ valóban sűrűségfüggvényt definiál. Gyakorlásképpen vizsgáljuk meg a φ függvényt, és mutassuk meg, hogy rendelkezik a következő tulajdonságokkal.

$$\lim_{x \to -\infty} \varphi(x) = \lim_{x \to +\infty} \varphi(x) = 0$$

továbbá φ szigorúan monoton növő a $(-\infty, 0)$ intervallumon, szigorúan monoton fogyó a $(0, \infty)$ intervallumon, és globális maximuma van az $x = 0$ helyen.

A második derivált vizsgálata alapján φ konvex a $(-\infty, 1)$ és $(1, +\infty)$ intervallumokon, valamint konkáv a $(-1, 1)$ intervallumon, és inflexiós pontjai vannak az $x = -1$ és $x = 1$ pontokban. KÉSZÍTSÜNK ÁBRÁT!

19.6 Tétel.

- *Az eloszlás paramétere: nincs paraméter,*
- *Az eloszlás várható értéke:* $E(Z) = 0$*,*
- *Az eloszlás varianciája:* $Var(Z) = 1$*.*

Bizonyítás. A 9.6 Példa szerint egyrészt $E(Z) = 0$, másrészt a (9.3) egyenlőség alapján a második momentum $E(Z^2) = 1$. Tehát

$$Var(Z) = E(Z^2) - E(Z)^2 = 1\,.$$

amit igazolnunk kellett. □

Jelölje a továbbiakban Φ a standard normális eloszlásfüggvényt, azaz

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)\, dt\,.$$

Ez a függvény rendelkezik az eloszlásfüggvények tulajdonságaival, de érdekessége, hogy nem lehet elemi függvényekkel, vagy azok véges kombinációival explicit alakban előállítani.

FIGYELEM! Lásd a 9. fejezetben mondottakat a φ primitív függvényéről!

Vegyük észre, hogy φ páros függvény, azaz szimmetrikus az y-tengelyre. Ebből adódik, hogy $\Phi(0) = 1/2$, valamint

$$\Phi(-x) = 1 - \Phi(x) \tag{19.1}$$

bármely x valós számra.

19.7 Példa. A statisztikában és további alkalmazásokban betöltött központi szerepe miatt a Φ eloszlásfüggény étékeire táblázatokat találhatunk a legtöbb valószínűségszámítással foglalkozó tankönyvben, illetve feladatgyűjteményben, továbbá a táblázatkezelő programokban, mint például az MS Windows Office Excel alkalmazásban is. Lásd a Feladatgyűjtemény-2 338–339-ik oldalain található táblázatokat!

Ha például Z standard normális eloszlású valószínűségi változó, úgy határozzuk meg a

$$P(-2 < Z < 2)$$

valószínűséget. A Feladatgyűjtemény-2 339-ik oldalán található táblázatot használva

$$\begin{aligned} P(-2 < Z < 2) &= \Phi(2) - \Phi(-2) = \Phi(2) - (1 - \Phi(2)) = 2\Phi(2) - 1 = \\ &= 2 \cdot 0.9772 - 1 = 0.9544 \end{aligned}$$

ahol kihasználtuk a (19.1) alatti szimmetriatulajdonságot is.

19.4. Normális eloszlás

19.8 Definíció. Legyenek m és σ adott valós számok, ahol $\sigma > 0$. Legyen Z standard normális eloszlású változó, ekkor az

$$X = \sigma Z + m$$

valószínűségi változót (m, σ)-*paraméterű normális eloszlásúnak* nevezzük.

A standard normális eloszlás, valamint a várható érték és szórás tulajdonságaiból (lásd a 17.11 Tételt) adódnak az (m, σ)-paraméterű normális eloszlású X változó tulajdonságai.

19.9 Tétel.

- *Az eloszlás paraméterei:* $m \in \mathbb{R}$ *és* $\sigma > 0$.
- *Az eloszlás várható értéke:* $E(X) = m$,
- *Az eloszlás varianciája:* $Var(X) = \sigma^2$.

Vajon hogyan állíthatjuk elő ennek az X változónak az eloszlás, és sűrűségfüggvényét? Jelölje F az X eloszlásfüggvényét, és legyen x tetszőleges valós szám. Ekkor

$$F(x) = P(X < x) = P(\sigma Z + m < x) = P\left(Z < \frac{x-m}{\sigma}\right) = \Phi\left(\frac{x-m}{\sigma}\right)$$

FIGYELEM! Fontos, hogy $\sigma > 0$, hiszen így pozitív számmal osztunk, ezért az egyenlőtlenség iránya nem változik.

Az X változó f sűrűségfüggvényét deriválással kapjuk: bármely $x \in \mathbb{R}$ esetén

$$f(x) = F'(x) = \frac{1}{\sigma}\varphi\left(\frac{x-m}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}$$

a láncszabály alapján. Ennek a függvénynek globális maximumhelye van az $x = m$ pontban, továbbá inflexiós pontjai vannak az $x = m - \sigma$ és $x = m + \sigma$ helyeken. KÉSZÍTSÜNK ÁBRÁT!

19.10 Példa. Egy (m, σ)-paraméterű normális eloszlású X változó esetén valamely intervallumba esés valószínűsége mindig kifejezhető a Φ eloszlásfüggvénnyel.

Legyenek ugyanis $a < b$ tetszőleges valós számok. Ekkor

$$P(a < X < b) = F(b) - F(a) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right).$$

Például egy $m = 10$, $\sigma = 2$ paraméterekkel rendelkező X normális eloszlású változóra

$$\begin{aligned} P(7 < X < 13) &= F(13) - F(7) = \Phi(1.5) - \Phi(-1.5) = 2\Phi(1.5) - 1 = \\ &= 2 \cdot 0.9332 - 1 = 0.8664 \end{aligned}$$

ahol felhasználtuk Φ szimmetriáját, és a Feladatgyűjtemény-2 339-ik oldalán található táblázatot is.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-2 VI/1, VI/2, VI/3, VI/4, VI/5 és VI/6 szakaszainak feldolgozása.
2. Házi feladatok: a VI/1 szakasz 533, 534, VI/2 szakasz 546, 558, 559, VI/3 szakasz 566, 568, 572, VI/4 szakasz 581, 582, 583, VI/5 szakasz 588, 598, 600, 603, VI/6 szakasz 622, 624, 630 és 632 feladatai. feladatai.
3. Tankönyv-2 4.8 és 4.9 szakaszai.

20. fejezet

Együttes eloszlások

20.1. Együttes eloszlásfüggvény

20.1 Definíció. Legyenek X és Y valószínűségi változók (nem feltételenül ugyanazon az eseménytéren). Tetszőleges x és y valós számokra az

$$F(x,y) = P(X < x, Y < y)$$

függvényt az X és Y *együttes eloszlásfüggvényének* nevezzük.

A definícióból adódik a következő állítás.

20.2 Állítás. *Ha F együttes eloszlásfüggvény, akkor*

$$\lim_{x\to-\infty} F(x,y) = \lim_{y\to-\infty} F(x,y) = 0$$

bármely rögzített y illetve x valós szám mellett, továbbá

$$\lim_{x,y\to+\infty} F(x,y) = 1$$

Az egydimenziós esethez hasonlóan itt is külön foglalkozunk a diszkrét és a folytonos eloszlásokkal.

20.2. Diszkrét együttes eloszlások

20.3 Definíció. Tegyük fel, hogy az X változó értékkészlete $\{x_1, x_2, \ldots\}$, illetve az Y változó értékkészlete $\{y_1, y_2, \ldots\}$. Ekkor X és Y együttes eloszlása

$$p_{ik} = P(X = x_i, Y = y_k) \qquad i = 1,2,\ldots \quad k = 1,2,\ldots$$

Ezeket az értékeket a következő táblázattal adhatjuk meg:

$y \setminus x$	x_1	x_2	x_3	$\cdots$
y_1	p_{11}	p_{21}	p_{31}	$\cdots$
y_2	p_{12}	p_{22}	p_{32}	$\cdots$
y_3	p_{13}	p_{23}	p_{33}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Világos, hogy mindegyik $p_{ik} \geq 0$ és $\sum_i \sum_k p_{ik} = 1$.

Legyen A a sík valamely részhalmaza. Vajon hogyan határozhatjuk meg a $P((X,Y) \in A)$ valószínűséget? Gyűjtsük össze mindazon x_i és y_k értékeket, amelyekre $(x_i, y_k) \in A$, ekkor

$$P((X,Y) \in A) = \sum_{(x_i,y_k)\in A} p_{ik}$$

20.4 Példa. Például az alábbi együttes eloszlás esetén

$y \setminus x$	0	1	2	3
0	0.1	0.08	0.13	0.04
1	0.04	0.2	0.08	0
2	0.03	0	0.05	0.25

az $A = \{(x,y) \in \mathbb{R}^2 : x + y \geq 3\}$ tartományba esés valószínűsége:

$$P(X + Y \geq 3) = 0.04 + 0.08 + 0.05 + 0.25 = 0.42$$

Felmerül a kérdés, hogy az együttes eloszlás ismeretében hogyan határozható meg X és Y eloszlása? A definícióból látható, hogy

$$p_i = P(X = x_i) = \sum_k p_{ik} = \sum_k P(X = x_i, Y = y_k) \qquad i = 1, 2, \ldots$$

Nevezetesen a $p_i = P(X = x_i)$ valószínűség az i-ik oszlop elemeinek összegeként nyerhető. Tehát az oszlopösszegek az X változó eloszlását adják.

Teljesen analóg módon

$$q_k = P(Y = x_k) = \sum_i p_{ik} = \sum_i P(X = x_i, Y = y_k) \qquad k = 1, 2, \ldots$$

azaz Y eloszlása a sorok összegeként kapható.

20.5 Definíció. X és Y eloszlásait az együttes eloszlás *peremeloszlásainak* nevezzük.

20.3. Folytonos együttes eloszlások

20.6 Definíció. Azt mondjuk, hogy X és Y együttes eloszása folytonos, ha létezik olyan f nemnegatív integrálható függvény a síkon, amelyre minden x és y valós számok mellett

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(t,s)\,ds\,dt$$

ahol F az X és Y valószínűségi változók együttes eloszlásfüggvénye. Ezt az f függvényt az X és Y *együttes sűrűségfüggvényének* nevezzük.

Világos, hogy ha f együttes sűrűségfüggvény, akkor nemnegatív a síkon, továbbá

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy=1$$

20.7 Példa. Legyen A a sík valamely részhalmaza. Hogyan határozható meg a $P((X,Y)\in A)$ valószínűség? Ha f az X és Y együttes sűrűségfüggvénye, akkor

$$P((X,Y)\in A)=\int\int_A f(x,y)\,dy\,dx$$

Például az alábbi együttes sűrűségfüggvény esetében

$$f(x,y)=\begin{cases}\frac{2}{3}(x+2y) & \text{ha } 0<x<1,\ 0<y<1\\ 0 & \text{különben}\end{cases} \tag{20.1}$$

az $A=\{(x,y)\in\mathbb{R}^2 : x<1/2,\ y<1/2\}$ tartományba esés valószínűsége

$$P(X<1/2,\ Y<1/2)=\frac{2}{3}\int_0^{1/2}\int_0^{1/2}(x+2y)\,dy\,dx=\frac{1}{8}$$

Felmerül a kérdés, hogy az együttes sűrűségfüggvény ismeretében hogyan határozható meg X és Y sűrűségfüggvénye? Belátható, hogy ha f_X az X sűrűségfüggvénye, akkor minden x pontban

$$f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,dy$$

és teljesen analóg módon

$$f_Y(y)=\int_{-\infty}^{\infty} f(x,y)\,dx$$

bármely x és y pontokban.

20.8 Definíció. Az f_X és f_Y függvényeket az együttes eloszlás *peremsűrűség-függvényeinek* nevezzük.

20.9 Példa. Például az előző példában vizsgált együttes sűrűségfüggvény esetén

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \begin{cases} \int_0^1 \frac{2}{3}(x+2y)\,dy & \text{ha } 0 < x < 1 \\ 0 & \text{különben} \end{cases}$$

Analóg módon állítható elő az Y peremsűrűségfüggvénye:

$$f_Y(y) = \begin{cases} \frac{1}{3}(4y+1) & \text{ha } 0 < y < 1 \\ 0 & \text{különben} \end{cases}$$

20.4. Függetlenség

20.10 Definíció. Legyenek X és Y valószínűségi változók, amelyek együttes eloszlásfüggvénye F. Jelölje F_X illetve F_Y az X illetve az Y eloszlásfüggvényét. Azt mondjuk, hogy X és Y *függetlenek*, ha

$$F(x,y) = F_X(x) \cdot F_Y(y)$$

bármely x, y valós számokra.

A definíciónk úgy is megfogalmazható, hogy X és Y függetlenek, ha

$$P(X < x, Y < y) = P(X < x) \cdot P(Y < y)$$

tetszőleges x, y valós számokra. Ennek a definíciónak könnyen ellenőrizhető ekvivalens alakjait fogalmazzuk meg az alábbiakban diszkrét illetve folytonos eloszlásokra

Legyenek X és Y diszkrét valószínűségi változók, amelyek együttes eloszlása

$$P(X = x_i, Y = y_k) = p_{ik} \qquad i = 1, 2, \ldots \quad k = 1, 2, \ldots$$

Tekintsük az X és Y peremeloszlásait:

$$P(X = x_i) = p_i \quad i = 1, 2, \ldots \qquad P(Y = y_k) = q_k \quad k = 1, 2, \ldots$$

20.11 Tétel. *X és Y akkor és csak akkor függetlenek, ha*

$$p_{ik} = p_i \cdot q_k$$

minden i és k indexekre.

Tételünk azt fogalmazza meg, hogy független változókra az együttes eloszlás a peremeloszlások szorzataként áll elő. Például a 20.4 Példában szereplő változók nem függetlenek, hiszen rögtön az együttes eloszlás első elemére

$$0.17 \cdot 0.35 = p_1 \cdot q_1 \neq p_{11} = 0.1\,,$$

ellenőrizzük!

Legyenek most X és Y folytonos eloszlású változók, amelyek együttes sűrűségfüggvénye f. Jelölje f_X az X, illetve f_Y az Y peremsűrűségfüggvényét.

20.12 Tétel. *X és Y akkor és csak akkor függetlenek, ha bármely x és y valós számokra*

$$f(x,y) = f_X(x) \cdot f_Y(y)$$

Bizonyítás. Könnyen adódik az $F(x,y) = F_X(x) \cdot F_Y(y)$ egyenlőségből. □

20.13 Példa. Például a (20.1) alatti példában szereplő változók nem függetlenek, hiszen

$$f_X(x) \cdot f_Y(y) \neq f(x,y)\,,$$

azaz a peremsűrűségfüggvények szorzata nem állítja elő az együttes sűrűségfüggvényt.

Könnyen látható azonban, hogy ha X és Y együttes sűrűségfüggvénye

$$f(x,y) = \begin{cases} 4xy & \text{ha } 0 < x < 1,\ 0 < y < 1 \\ 0 & \text{különben} \end{cases}$$

akkor X és Y függetlenek. Valóban, ekkor

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^1 4xy\,dy = \begin{cases} 2x & \text{ha } 0 < x < 1 \\ 0 & \text{különben}\ . \end{cases}$$

és f szimmetriája miatt f_Y ugyanilyen alakú az y változóban. Ezért

$$f(x,y) = f_X(x) \cdot f_Y(y)$$

minden x és y valós számra.

20.5. Feltételes eloszlások

Tekintsük az X és Y diszkrét valószínűségi változókat, amelyek együttes eloszlása $P(X = x_i, Y = y_k) = p_{ik}$, ahol $i = 1, 2, \ldots$ és $k = 1, 2, \ldots$.

20.14 Definíció. Tegyük fel, hogy valamely k indexre $P(Y = y_k) > 0$. Ekkor X *feltételes eloszlásán* az $Y = y_k$ feltétel mellett a

$$P(X = x_i|Y = y_k) = \frac{P(X = x_i, Y = y_k)}{P(Y = y_k)}, \quad i = 1, 2, \ldots$$

eloszlást értjük.

FIGYELEM! Ellenőrizzük közvetlenül a definíció alapján, hogy valóban eloszlást definiáltunk.

20.15 Definíció. Az X változó *feltételes várható értékén* az $Y = y_k$ feltétel mellett az

$$E(X|Y = y_k) = \sum_{i=1} x_i \cdot P(X = x_i|Y = y_k)$$

szummát értjük, amely véges vagy végtelen aszerint, hogy X értékkészlete véges vagy végtelen, ezért nem jelöltük a felső határt.

20.16 Példa. Tekintsük például a 20.4 Példában megadott együttes eloszlást. Ekkor $P(Y = 1) = 0.32$, továbbá az X feltételes várható értéke az $Y = 1$ feltétel mellett

$$E(X|Y = 1) = 0 \cdot 0.04 + 1 \cdot 0.2 + 2 \cdot 0.08 + 3 \cdot 0 = 0.36$$

Otthoni tanuláshoz

1. A Feladatgyűjtemény-2 V/1 és V/2 szakaszainak feldolgozása.
2. Házi feladatok: az V/1 szakasz 411, 412, 417, 420, 424 és 427, valamint az V/2 szakasz 435, 437 és 438 feladatai.
3. Tankönyv-2 6.1, 6.2, 6.3, 6.4, 6.5, 6.6 és 6.7 szakaszai.

21. fejezet

Kovariancia és korreláció

21.1. Összeg várható értéke

21.1 Tétel. *Ha az X és Y változóknak van várható értéke, akkor $X+Y$-nak is, és*

$$E(X+Y) = E(X) + E(Y)$$

Bizonyítás. Diszkrét esetben vázoljuk a bizonyítást, a folytonos esetben analóg módon megy.

$$\begin{aligned} E(X+Y) &= \sum_i \sum_k (x_i + y_k) P(X = x_i, Y = y_k) \\ &= \sum_i x_i \sum_k p_{ik} + \sum_k y_k \sum_i p_{ik} \\ &= \sum_i x_i P(X = x_i) + \sum_k y_k P(Y = y_k) = E(X) + E(Y) \qquad \square \end{aligned}$$

Tételünk természetesen tetszőleges véges összegre is érvényes marad.

21.2 Példa. Tegyük fel, hogy n számú papírlapra felírtuk egyenként az $1, \ldots, n$ számokat, majd a lapokat egy kalapba tettük. Találomra egymás után, visszatevéssel kiveszünk m számú lapot. Jelentse X a húzott számok összegét. Mennyi $E(X)$?

Kipróbálhatjuk, hogy X eloszlását igen komplikált előállítani.

Jelentsék $X_1, \ldots, X_m$ rendre a kihúzott számokat. A visszatevéses mintavétel miatt mindegyik X_k azonos eloszlású, nevezetesen

$$P(X_k = i) = \frac{1}{n} \qquad i = 1, \ldots, n$$

Ez azt jelenti, hogy

$$E(X_k) = \sum_{i=1}^{n} i \cdot \frac{1}{n} = \frac{n(n+1)}{2} \cdot \frac{1}{n} = \frac{n+1}{2}.$$

Másrészt nyilván $X = X_1 + \ldots + X_m$, ezért

$$E(X) = E(X_1) + \ldots + E(X_m) = m \cdot \frac{n+1}{2}$$

Tehát $E(X)$ megadható az X eloszlásának ismerete nélkül is!

21.2. Szorzat várható értéke

Ha X és Y diszkrét változók, amelyeknek van várható értékük, akkor

$$E(XY) = \sum_i \sum_k x_i y_k \cdot p_{ik}$$

ahol X értékkészlete $\{x_1, x_2, \ldots\}$, Y értékkészlete $\{y_1, y_2, \ldots\}$ és p_{ik} jelöli az együttes eloszlást.

Teljesen analóg módon, ha X és Y folytonos eloszlású változók, amelyeknek létezik várható értékük, és együttes sűrűségfüggvényük f, akkor

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f(x,y)\, dx\, dy$$

21.3 Tétel. *Ha X és Y függetlenek, akkor*

$$E(XY) = E(X) \cdot E(Y)$$

Bizonyítás. Vizsgáljuk meg a folytonos esetet, a diszkrét eset ugyanígy megy.

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f(x,y)\, dx\, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f_X(x) \cdot f_Y(y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} x f_X(x)\, dx \cdot \int_{-\infty}^{\infty} y f_Y(y)\, dy = E(X) \cdot E(Y) \end{aligned}$$

ugyanis a függetlenség miatt az együttes sűrűségfüggvény az $f(x,y) = f_X(x) \cdot f_Y(y)$ szorzat alakban áll elő. □

21.3. Összeg varianciája

21.4 Tétel. *Tegyük fel, hogy X és Y függetlenek, és mindkettőnek van varianciája. Akkor*

$$Var(X+Y) = Var(X) + Var(Y)$$

Az állítás tetszőleges véges összegre is érvényes.

Bizonyítás. Valóban, a szorzat várható értékére vonatkozó tételünk alapján:

$$\begin{aligned} Var(X+Y) &= E((X+Y-E(X+Y))^2) \\ &= E((X-E(X))^2) + E((Y-E(Y))^2) \\ &\quad +2E((X-E(X))(Y-E(Y))) \\ &= Var(X) + Var(Y) + 2(E(XY) - E(X)E(Y)) \\ &= Var(X) + Var(Y)\,. \qquad \square \end{aligned}$$

21.5 Példa. Vajon miért gondoljuk, hogy egy kísérlet többszöri elvégzésekor az eredmények átlagolásával várhatóan pontosabb eredményt kapunk?

Tegyük fel, hogy egy ismeretlen m mennyiség meghatározására n számú megfigyelést végzünk, amelyeknek eredményei $X_1, \ldots, X_n$ valószínűségi változók. Feltesszük, hogy a változók függetlenek, és azonos eloszlásúak, amelyekre

$$E(X_k) = m, \quad D(X_k) = \sigma, \quad k = 1, 2, \ldots, n\,.$$

Az a feltevésünk, hogy a változók azonos eloszlásúak, azt jelenti, hogy a megfigyeléseket (méréseket) azonos körülmények között végezzük. Ilyenkor σ jelenti a várható hibát. Átlagoljuk az eredményeinket, azaz vezessük be az

$$Y_n = \frac{X_1 + \ldots + X_n}{n}$$

valószínűségi változót. Ekkor világos, hogy $E(Y_n) = m$, továbbá a fenti tételünk értelmében

$$Var(Y_n) = Var\left(\frac{1}{n}(X_1 + \ldots + X_n)\right) = \frac{1}{n^2} n \cdot \sigma^2 = \frac{\sigma^2}{n}\,.$$

a függetlenség miatt. Innen az Y_n szórására az adódik, hogy

$$D(Y_n) = \frac{\sigma}{\sqrt{n}}$$

tehát $n \to \infty$ esetén $D(Y_n) \to 0$, azaz az n növelésével a várható hiba nullához tart.

21.4. Kovariancia és korreláció

Valószínűségi változók közötti összefüggés mérésére használjuk az alábbi fogalmakat.

21.6 Definíció. Az X és Y változók *kovarianciája*

$$Cov(X,Y) = E((X - E(X)) \cdot (Y - E(Y)))$$

és *korrelációs együtthatója*

$$Corr(X,Y) = \frac{Cov(X,Y)}{D(X) \cdot D(Y)}$$

Könnyen látható, hogy

$$Cov(X,Y) = E(XY - E(X)Y - E(Y)X + E(X)E(Y)) = E(XY) - E(X)E(Y)\,,$$

és általában ezt az egyszerűbb alakot használjuk a kovariancia meghatározására.

Megjegyezzük, hogy a kovariancia nem ad abszolút mérőszámot az összefüggés mérésére, hiszen bármely $\alpha \neq 0$ valós szám esetén

$$Cov(\alpha X, Y) = \alpha Cov(X,Y)$$

azaz a mérték függ a dimenziótól. Gondoljunk arra, hogy ha X és Y egy feladatban Euróban számított költségeket jelentenek, akkor Forintra áttérve a kovariancia kb. 350^2-szeresére változik. A korrelációs együttható azonban már független a dimenziótól, hiszen bármely $\alpha \neq 0$ és β valós számokra

$$Corr(\alpha X + \beta, \alpha Y + \beta) = Corr(X,Y)$$

azaz a korreláció invariáns a lineáris transzformációra. FIGYELEM! Ellenőrizzük ezt az egyenlőséget közvetlenül a definíció alapján!

21.7 Tétel.

1. *$-1 \leq Corr(X,Y) \leq 1$*
2. *Ha X és Y függetlenek, akkor $Cov(X,Y) = 0$*

Bizonyítás. Az első állítás bizonyításához legyen $t \in \mathbb{R}$ tetszőleges, és tekintsük a

$$W = [X - E(X) + t(Y - E(Y))]^2$$

valószínűségi változót. Mivel W nemnegatív, ezért várható értéke is nemnegatív. Ez azt jelenti, hogy

$$E(W) = E((X - E(X))^2) + 2tCov(X,Y) + t^2E((Y - E(Y))^2) \geq 0$$

minden t valós számra. Ez a kifejezés t-ben másodfokú, ezért csak úgy lehet nemnegatív, ha a diszkriminánsa nempozitív. Tehát

$$4Cov(X,Y)^2 - 4E((X-E(X))^2)E((Y-E(Y))^2) \leq 0\,.$$

A jobb oldalon éppen a varianciák szorzata áll. Rendezve és mindkét oldalból négyzetgyököt vonva azt kapjuk, hogy

$$|Cov(X,Y)| \leq D(X)D(Y)$$

amiből azonnal következik az állítás.

A második állításunk a 21.3 Tétel közvetlen következménye. □

21.8 Példa. FIGYELEM! Az alábbi példánk mutatja, hogy a tételünk második állítása nem fordítható meg. Feldobunk egy érmét kétszer egymás után, és legyen

$$X_k = \begin{cases} 0 & \text{ha a } k\text{-ik dobás fej} \\ 1 & \text{ha a } k\text{-ik dobás írás} \end{cases}$$

($k=1,2$). Tekintsük az $Y_1 = X_1 + X_2$ és $Y_2 = X_1 - X_2$. változókat. Ekkor

$Y_2 \setminus Y_1$	0	1	2
-1	0	0.25	0
0	0.25	0	0.25
1	0	0.25	0

Az együttes eloszlás vizsgálatával látható, hogy Y_1 és Y_2 nem függetlenek, de könnyen kiszámolható, hogy $Cov(Y_1,Y_2) = 0$

21.5. Teljes várható érték tétel

Tekintsük az X és Y diszkrét változókat, amelyek együttes eloszlása $P(X = x_i, Y = y_k) = p_{ik}$ és $P(Y = y_k) > 0$ az $i = 1,2,\ldots$ és $k = 1,2\ldots$ indexekre.

21.9 Definíció. Képezzük az X feltételes várható értékeit az $Y = y_k$ feltételek mellett, azaz

$$m_k = E(X|Y=y_k) = \sum_{i=1} x_i P(X = x_i | Y = y_k)$$

minden $k = 1,2\ldots$ esetén. Ezt a sorozatot az X változó Y-ra vonatkozó *feltételes várható értékének* nevezzük. Jelölése $E(X|Y)$.

Vegyük észre, hogy itt egy valószínűségi változót definiáltunk, nevezetesen

$$E(X|Y) = m_k\,, \quad \text{ha} \quad Y = y_k\,, k = 1,2,\ldots$$

Az alábbiakban megadjuk ennek a változónak a várható értékét. Ezt az eredményünket a Teljes valószínűség tétel általánosításának tekinthetjük.

21.10 Tétel. **(Teljes várható érték tétel)** $E(E(X|Y)) = E(X)$.

Bizonyítás. Valóban,

$$\begin{aligned} E(E(X|Y)) &= \sum_{k=1} m_k P(Y = y_k) = \sum_{k=1}\sum_{i=1} x_i P(X = x_i | Y = y_k) P(Y = y_k) \\ &= \sum_{i=1} x_i \sum_{k=1} P(X = x_i, Y = y_k) = \sum_{i=1} x_i P(X = x_i) = E(X) \end{aligned}$$

hiszen a második sorban éppen X peremeloszlását kapjuk. □

FIGYELEM! Vajon a második sorban miért cserélhető fel a két szumma?

21.11 Példa. Néha könnyebb $E(X)$ előállítása a fenti tételből, mint közvetlenül, amint azt az alábbi példában láthatjuk. Egy biztosítóhoz egy év alatt beérkező kárigények száma Poisson-eloszlású adott $\lambda > 0$ paraméterrel. Egy kárigényt a biztosító adott $p > 0$ valószínűséggel elutasít (egymástól függetlenül). Adjuk meg egy adott évben az elutasított kárigények számának várható értékét.

Jelentse X az elutasított kárigények számát, és Y az összes beérkező kárigény számát. Világos, hogy adott $n \in \mathbb{N}$ esetén az $Y = n$ feltétel mellett éppen a Bernoulli-kísérlettel állunk szemben, ezért

$$P(X = k | Y = n) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{ha} \quad n \geq k$$

illetve $P(X = k|Y = n) = 0$, ha $n < k$. Tehát a feltételes várható érték

$$m_n = E(X|Y = n) = np\,, \quad n = 1, 2, \ldots$$

Innen a Teljes várható érték tétel alapján

$$E(X) = E(E(X|Y)) = \sum_{n=1}^{\infty} np \frac{\lambda^n}{n!} e^{-\lambda} = \lambda p$$

FIGYELEM! Állítsuk elő a várható értéket közvetlenül X eloszlása alapján is (lásd a Gyakorlatok között)!

Otthoni tanuláshoz

1. A Feladatgyűjtemény-2 V/4 és VI/5 szakaszainak feldolgozása.
2. Házi feladatok: a V/4 szakasz 471, 473, 482, 485, 489, 504, 509, valamint a VI/5 szakasz 608, 609 és 6.10 feladatai.
3. Tankönyv-2 6.14 és 8.3 szakaszai.

22. fejezet

Változók összegének eloszlása

22.1. Diszkrét változók összegének eloszlása

Tegyük fel, hogy X és Y független Poisson-eloszlású változók, rendre $\lambda > 0$, illetve $\mu > 0$ paraméterekkel. Határozzuk meg $X + Y$ eloszlását. Ekkor tetszőleges k nemnegatív egész számra a függetlenség miatt

$$\begin{aligned} P(X+Y=k) &= \sum_{i=0}^{k} P(X=i, Y=k-i) = \sum_{i=0}^{k} P(X=i) \cdot P(Y=k-i) \\ &= \sum_{i=0}^{k} \frac{\lambda^i}{i!} e^{-\lambda} \cdot \frac{\mu^{k-i}}{(k-i)!} e^{-\mu} = \frac{e^{-(\lambda+\mu)}}{k!} \sum_{i=0}^{k} \binom{k}{i} \lambda^i \mu^{k-i} \\ &= \frac{(\lambda+\mu)^k}{k!} e^{-(\lambda+\mu)} \end{aligned}$$

Ez az eredményünk azt mutatja, hogy a két független változó összege is Poisson-eloszlású $\lambda + \mu$ paraméterrel.

Ezt az állításunkat indukcióval tetszőleges véges számú változóra is kiterjeszthetjük.

22.1 Tétel. *Tegyük fel, hogy $X_1, \ldots, X_n$ független Poisson-eloszlású valószínűségi változók rendre $\lambda_1, \ldots, \lambda_n$ pozitív paraméterekkel. Akkor az*

$$Y_n = X_1 + \ldots + X_n$$

változó is Poisson-eloszlású, amelynek paramétere $\lambda_1 + \ldots + \lambda_n$.

22.2. Folytonos változók összegének eloszlása

Legyenek most X és Y folytonos eloszlású független változók, amelyek sűrűségfüggvényei f illetve g. Jelölje F illetve G az eloszlásfüggvényeket. Jelentse H

az $X+Y$ eloszlásfüggvényét. Ennek előállításához válasszunk egy tetszőleges $x \in \mathbb{R}$ számot. Ábra készítésével láthatjuk, hogy

$$\begin{aligned} H(x) &= \iint_{t+s<x} f(s)g(t)\,dt\,ds = \int_{-\infty}^{\infty}\int_{-\infty}^{x-s} f(s)g(t)\,dt\,ds \\ &= \int_{-\infty}^{\infty} f(s)\left(\int_{-\infty}^{x-s} g(t)\,dt\right)ds = \int_{-\infty}^{\infty} f(s)G(x-s)\,ds\,. \end{aligned}$$

Innen az integrál deriválásával az $X+Y$ változó h sűrűségfüggvényére az adódik, hogy

$$h(x) = \int_{-\infty}^{\infty} f(s)g(x-s)\,ds$$

és ezt az f és g sűrűségfüggvények konvolúciós integráljának nevezzük.

FIGYELEM! Az integrál deriválása a fenti levezetésben nem egyszerű. Vizsgáljuk meg ezt a szabályt néhány egyszerű, könnyen kiintegrálható esetben!

22.2 Példa. Tekintsünk most azt a példát, ahol X és Y független, egyenletes eloszlású változók a $[0,1]$ intervallumon. Ekkor (X,Y) egyenletes eloszlású a sík egységnégyzetén. Ábra készítésével mutassuk meg, hogy ha h jelenti az $X+Y$ változó sűrűségfüggvényét, akkor

$$h(x) = \begin{cases} x & \text{ha } 0<x<1 \\ 2-x & \text{ha } 1<x<2 \\ 0 & \text{különben.} \end{cases}$$

22.3 Példa. Legyenek most X és Y független, azonos λ-paraméterű exponenciális eloszlású változók, és tekintsük az $X+Y$ változó h sűrűségfüggvényét. Ha f jelenti az exponenciális eloszlás sűrűségfüggvényét, akkor az előbbiek szerint a konvolúciós integrál

$$h(x) = \int_{-\infty}^{\infty} f(s)f(x-s)\,ds$$

alakú. Az exponenciális eloszlás sűrűségfüggvénye zérus a negatív félegyenesen, ezért ez az integrandus akkor csak akkor nem nulla, ha $s>0$ és $x-s>0$, azaz $0<s<x$. Tehát

$$h(x) = \int_0^x \lambda^2 e^{-\lambda s} e^{-\lambda(x-s)}\,ds = \lambda^2 \int_0^x e^{-\lambda x}\,ds = \lambda^2 x e^{-\lambda x}$$

bármely $x>0$ esetén, hiszen az integrandus nem függ az s változótól.

Teljes indukcióval igazolhatjuk a fenti eredményünk alábbi kiterjesztését.

22.4 Tétel. *Tegyük fel, hogy $X_1,\ldots,X_n$ független exponenciális eloszlású valószínűségi változók ugyanazon $\lambda>0$ paraméterrel. Akkor az*

$$Y_n = X_1 + \ldots + X_n$$

változó h_n sűrűségfüggvénye

$$h_n(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}$$

ha $x > 0$, és $h_n(x) = 0$, ha $x \leq 0$.

22.3. A Poisson-folyamat

Ebben a szakaszban az exponenciális eloszlás és a Poisson-eloszlás egy mélyebb összefüggésére világítunk rá.

Tekintsük a $T_1, T_2, \ldots$ valószínűségi változókat, amelyek egymás utáni várakozási időket jelentenek egymás utáni "bekövetkezések" között.

Gondoljunk például egymás utáni gépjárművek közötti időtartamra egy autóúton, egy biztosítóhoz egymás után beérkező káresemények közötti időre, ügyfélablaknál egymást követő várakozási időkre, call center-be beérkező hívások közötti időkre, stb.

Tegyük fel, hogy a $T_1, T_2, \ldots$ változók függetlenek és azonos $\lambda > 0$ paraméterű exponenciális eloszlásúak. A várható értékre tekintettel minél kisebb a λ, annál hosszabbak a várható várakozási idők. Megjegyezzük, hogy az exponenciális eloszlás memória-nélküli tulajdonságából adódik, hogy az eltelt várakozási időtől független a további várakozás időtartama.

Jelölje a továbbiakban $S_0 = 0$ és

$$S_n = T_1 + \ldots + T_n$$

amely a teljes várakozási időt jelenti az n-ik bekövetkezésig. Adott $t > 0$ esetén az

$$\{S_n \leq t\}$$

esemény azt jelenti, hogy az n-ik bekövetkezés a t időpont előtt történt. Ez azt jelenti, hogy a $[0, t]$ időintervallumban a bekövetkezések száma legalább n.

Jelentse tehát $N(t)$ a bekövetkezések számát a $[0, t]$ időintervallumban, ekkor az

$$\{N(t) \geq n\} = \{S_n \leq t\}$$

események megegyeznek. Minden $t > 0$ esetén $N(t)$ egy valószínűségi változó, ezt a hozzárendelést *Poisson-folyamatnak* nevezzük.

Kérdés, hogy adott $t > 0$ mellett hogyan határozhatjuk meg az $N(t)$ változó eloszlását? Az az esemény, hogy a $[0, t]$ intervallumban pontosan n bekövetkezés történik

$$\{N(t) = n\} = \{S_n \leq t\} \cap \overline{\{S_{n+1} \leq t\}} = \{S_n \leq t < S_{n+1}\}.$$

Mivel $\{S_{n+1} \le t\} \subset \{S_n \le t\}$, innen a valószínűségre azt kapjuk, hogy

$$P(N(t) = n) = P(S_n \le t) - P(S_{n+1} \le t)\,.$$

Jelölje most h_n az S_n változó, illetve h_{n+1} az S_{n+1} változó sűrűségfüggvényét. Mivel $T_1, T_2, \ldots$ független, azonos λ-paraméterű exponenciális eloszlású változók, azért a megelőző szakasz 22.4 Tétele alapján

$$h_n(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} \quad \text{illetve} \quad h_{n+1}(x) = \frac{\lambda^{n+1}}{n!} x^n e^{-\lambda x}$$

minden $x > 0$ esetén. Tehát

$$P(N(t) = n) = P(S_n \le t) - P(S_{n+1} \le t) = \int_0^t h_n(x)\, dx - \int_0^t h_{n+1}(x)\, dx\,.$$

Számítsuk ki a jobb oldalon álló első integrált parciális integrálással.

$$\begin{aligned}
\int_0^t h_n(x)\, dx &= \frac{\lambda^n}{(n-1)!} \int_0^t x^{n-1} e^{-\lambda x}\, dx \\
&= \frac{\lambda^n}{(n-1)!} \left[\frac{x^n}{n} e^{-\lambda x} \right]_0^t + \frac{\lambda^n}{(n-1)!} \int_0^t \frac{x^n}{n} \lambda e^{-\lambda x}\, dx \\
&= \frac{(\lambda t)^n}{n!} e^{-\lambda t} + \frac{\lambda^{n+1}}{n!} \int_0^t x^n e^{-\lambda x}\, dx\,.
\end{aligned}$$

Vegyük észre, hogy a legutolsó integrálban éppen h_{n+1} szerepel! Innen

$$P(N(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

22.5 Tétel. *A Poisson-folyamatban a $[0, t]$ időintervallumban történő bekövetkezések száma λt-paraméterű Poisson-eloszlású valószínűségi változó.*

22.4. Normális eloszlások összege

Legyenek Z_1 és Z_2 független, standard normális eloszlású valószínűségi változók, és határozzuk meg az

$$Y = Z_1 + Z_2$$

változó eloszlását. A konvolúciós integrál ebben az esetben

$$h(x) = \int_{-\infty}^{\infty} \varphi(s) \varphi(x - s)\, ds$$

ahol h az Y sűrűségfüggvénye. Ekkor

$$\begin{aligned} h(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-s^2/2}e^{-(x-s)^2/2}\,ds = \frac{1}{2\pi}e^{-\frac{x^2}{2}}\int_{-\infty}^{\infty} e^{xs-s^2}\,ds \\ &= \frac{1}{2\pi}e^{-\frac{x^2}{2}}\int_{-\infty}^{\infty} e^{-(s-x/2)^2}e^{x^2/4}\,ds = \frac{1}{2\pi}e^{-\frac{x^2}{4}}\int_{-\infty}^{\infty} e^{-(s-x/2)^2}\,ds \end{aligned}$$

A legutolsó integrál éppen a Gauss-integrál, amelynek értéke $\sqrt{\pi}$, tehát

$$h(x) = \frac{1}{2\sqrt{\pi}}e^{-\frac{x^2}{4}} \qquad -\infty < x < \infty$$

Ez éppen annak a normális eloszlásnak a sűrűségfüggvénye, amelynek paraméterei $m = 0$ és $\sigma = \sqrt{2}$.

Teljesen hasonló gondolatmenettel a következő eredményt fogalmazhatjuk meg.

22.6 Tétel. *Legyenek $Z_1, \ldots, Z_n$ független, standard normális eloszlású valószínűségi változók. Akkor az $Y = Z_1 + \ldots + Z_n$ változó is normális eloszlású, amelynek paraméterei $m = 0$ és $\sigma = \sqrt{n}$.*

22.5. Centrális határeloszlás-tétel

Képzeljük el, hogy valamilyen (ismeretlen) m mennyiség értékének meghatározására n számú független kísérletet végzünk. Az ismeretlen mennyiség közelítéséhez a kísérletek kimeneteleinek számtani átlagát használjuk.

Jelölje a kísérletek kimeneteleit rendre $X_1, \ldots, X_n$ és tegyük fel, hogy ezek független, azonos eloszlású változók, amelyekre

$$E(X_k) = m, \qquad D(X_k) = \sigma, \qquad k = 1, 2, \ldots, n$$

Vezessük be a következő jelölést a változók standardizált átlagára:

$$Y_n = \frac{\frac{1}{n}(X_1 + \ldots + X_n) - m}{\sigma/\sqrt{n}}$$

Tételeink szerint ennek az Y_n változónak a várható értéke 0 és a szórása 1.

Alekszandr Ljapunov orosz matematikus és a korabeli (XX.szd eleje) matematika csodálatos felismerése volt, hogy a fenti Y_n változó eloszlása tart a standard normális eloszláshoz.

22.7 Tétel. (Centrális határeloszlás-tétel) *A fenti feltételek mellett jelölje F_n az Y_n eloszlásfüggvényét. Ekkor minden $x \in \mathbb{R}$ esetén*

$$\lim_{n\to\infty} F_n(x) = \Phi(x)\,.$$

22.8 Példa. Egy kisforgalmú üzletbe egy adott napon 100 látogató érkezik. Mindegyik látogató (egymástól függetlenül) $p = 0.2$ valószínűséggel vásárol valamit. Mi a valószínűsége, hogy az adott napon az üzletben vásárlók száma 15 és 25 között lesz?

Jelölje X a vásárlók számát. Ekkor X binomiális eloszlású $n = 100$ és $p = 0.2$ paraméterekkel. Ha a k-ik vásárlóra bevezetjük az

$$X_k = \begin{cases} 0 & \text{ha nem vásárol} \\ 1 & \text{ha vásárol} \end{cases}$$

jelölést, akkor $X = X_1 + \ldots + X_{100}$. Könnyen látható, hogy minden k esetén $E(X_k) = 0.2$ és $Var(X_k) = 0.16$, és így $D(X_k) = 0.4$. Tehát

$$\begin{aligned} P(15 < X < 25) &= P\left(-\frac{5}{4} < \frac{X-20}{4} < \frac{5}{4}\right) \\ &= P\left(-\frac{5}{4} < \frac{\frac{1}{100}(X_1 + \ldots + X_{100}) - 0.2}{0.4/10} < \frac{5}{4}\right) \end{aligned}$$

A centrális határeloszlás-tétel alapján

$$\begin{aligned} P(15 < X < 25) &\approx \Phi(1.25) - \Phi(-1.25) \\ &= 2\Phi(1.25) - 1 = 0.7888 \end{aligned}$$

adódik a standard normális eloszlás táblázata alapján, a Feladatgyűjtemény-2 339-ik oldalán.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-2 V/3 és VI/5 szakaszainak feldolgozása.
2. Házi feladatok: a V/3 szakasz 445, 446, 447, 452, 453, 457, 468, valamint a VI/5 szakasz 608, 609 és 6.10 feladatai.
3. Tankönyv-2 6.10, 6.11 és 8.3 szakaszai.

23. fejezet

A nagy számok törvénye

23.1. Csebisev-egyenlőtlenség

Az eddigiekben számos esetben kellett meghatároznunk

$$P(a < X < b)$$

alakú valószínűségeket. Ez könnyen megtehető, ha ismert az X valószínűségi változó eloszlása. Nevezetesen diszkrét változó esetében

$$P(a < X < b) = \sum_{a < x_k < b} P(X = x_k)$$

míg egy f sűrűségfüggvénnyel rendelkező változóra

$$P(a < X < b) = \int_a^b f(x)\,dx$$

Vannak esetek, amikor így nem járhatunk el, például:

1. az X változó eloszlása nem ismert,
2. az X eloszlása ismert, de túl bonyolult a fenti valószínűség meghatározása.

Ilyen esetekben is megelégedhetünk azonban egy, a feladat szempontjából megfelelő becsléssel a fenti valószínűségre. Erre a kérdésre ad választ a Csebisev-egyenlőtlenség. Tekintsünk egy olyan X valószínűségi változót, amelynek van várható értéke és szórása.

23.1 Tétel. **(Csebisev-egyenlőtlenség)** *Legyen az X valószínűségi változó várható értéke $E(X) = m$ és szórása $D(X) = \sigma$. Ekkor*

$$P(|X - m| < k \cdot \sigma) \geq 1 - \frac{1}{k^2}$$

bármely $k > 0$ szám esetén.

Bizonyítás. A bizonyítást folytonos esetre mutatjuk meg, diszkrét esetben teljesen hasonlóan végezhető el. Legyen tehát f az X sűrűségfüggvénye, ekkor

$$\sigma^2 = \int_{-\infty}^{\infty} (x-m)^2 f(x)\, dx$$

Ha most $k > 0$ adott, akkor a jobb oldalon álló integrál értéke nem nőhet, ha kihagyjuk az $[m - k\sigma, m + k\sigma]$ intervallumot az integrálási útból. Nevezetesen

$$\sigma^2 \geq \int_{-\infty}^{m-k\sigma} (x-m)^2 f(x)\, dx + \int_{m+k\sigma}^{\infty} (x-m)^2 f(x)\, dx \tag{23.1}$$

hiszen az integrandus nemnegatív. Másrészt a $(-\infty, m - k\sigma]$ intervallum bármely x pontjában $(x-m)^2 \geq k^2\sigma^2$, ezért

$$\int_{-\infty}^{m-k\sigma} (x-m)^2 f(x)\, dx \geq \int_{-\infty}^{m-k\sigma} k^2\sigma^2 f(x)\, dx = k^2\sigma^2 P(X \leq m - k\sigma)\,.$$

Teljesen hasonló módon látható, hogy az $[m + k\sigma, \infty)$ intervallum bármely x pontjában $(x-m)^2 \geq k^2\sigma^2$, és ezért

$$\int_{m+k\sigma}^{\infty} (x-m)^2 f(x)\, dx \geq \int_{m+k\sigma}^{\infty} k^2\sigma^2 f(x)\, dx = k^2\sigma^2 P(X \geq m + k\sigma)\,.$$

Ha ezt a két utóbbi egyenlőtlenséget összevetjük a (23.1) alatti egyenlőtlenséggel, akkor az adódik, hogy

$$\sigma^2 \geq k^2\sigma^2 P(X \leq m - k\sigma) + k^2\sigma^2 P(X \geq m + k\sigma)\,.$$

Mindkét oldalt elosztva a $k^2\sigma^2$ pozitív kifejezéssel

$$\frac{1}{k^2} \geq P(X \leq m - k\sigma) + P(X \geq m + k\sigma) = P(|X - m| \geq k\sigma)\,.$$

Innen az ellentett eseményre térve adódik a tételünk állítása. $\square$

Megjegyezzük, hogy a tételünk irreleváns eredményt ad, ha $k \leq 1$, ezért az egyenlőtlenséget csak a $k > 1$ esetben fogjuk használni.

23.2 Példa. Például, ha egy X változó eloszlása nem ismert, de azt tudjuk, hogy a várható értéke $E(X) = 8$ és szórása $D(X) = 2$, akkor

$$P(2 < X < 14) \geq 1 - \frac{1}{9} \approx 0.8889$$

hiszen ekkor $k = 3$.

23.2. Csebisev-egyenlőtlenség ekvivalens alakban

Néha kényelmesebb lehet a Csebisev-egyenlőtlenséget a következő alakban használni:

$$P(|X - E(X)| < \varepsilon) \geq 1 - \frac{Var(X)}{\varepsilon^2}$$

ahol $\varepsilon > 0$. Valóban, ez az alak ekvivalens a tételünkkel, hiszen ha $k \cdot D(X) = \varepsilon > 0$, akkor

$$\frac{1}{k^2} = \frac{Var(X)}{\varepsilon^2}$$

Fogalmazzuk meg ezt az ekvivalens tételünket a következő alakban.

23.3 Tétel. *Tegyük fel, hogy az X valószínűségi változó várható értéke $E(X) = m$, és szórása $D(X) = \sigma$. Akkor bármely adott $\varepsilon > 0$ esetén*

$$P(|X - m| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2} \tag{23.2}$$

23.4 Példa. Képzeljük el, hogy egy forgalmas telefonközpontba egy óra alatt 2000 hívás érkezik. Mindegyik hívás (egymástól függetlenül) 0.002 valószínűséggel téves. Mi a valószínűsége, hogy legfeljebb 8 téves hívás érkezik?

Jelölje X a téves hívások számát. Világos, hogy X binomiális eloszlású, amelynek paraméterei $n = 2000$ és $p = 0.002$. A kérdéses valószínűség tehát:

$$P(X \leq 8) = \sum_{k=0}^{8} \binom{2000}{k} 0.002^k \cdot 0.998^{2000-k}$$

aminek kiszámítása elég reménytelen feladat (bár az eloszlás ismert).

Adhatunk azonban becslést a Csebisev-egyenlőtlenséggel. Itt $m = 4$ és $\sigma^2 = 4 \cdot 0.998 \approx 4$, ezért

$$P(X \leq 8) = P(|X - 4| < 5) \geq 1 - \frac{4}{25} = 0.84$$

23.3. Poisson-approximáció

23.5 Példa. Egy 2000 ágyas központi kórházban annak valószínűsége, hogy egy adott napon egy betegnek intenzív kezelésre van szüksége (a többiektől függetlenül) 0.002. Az igazgató egy új intenzív osztályt kíván létrehozni úgy, hogy annak valószínűsége, hogy egy olyan beteg, akinek intenzív ápolásra van szüksége, nem kap ágyat az intenzív osztályon, kisebb legyen, mint 0.01. Hány ágyas legyen a minimális költségű intenzív osztály?

Jelentse N az ágyak számát és X azon betegek számát, akiknek intenzív kezelés szükséges az adott napon, akkor (mivel X nyilvánvalóan binomiális eloszlású $m = 4$ várható értékkel, és $\sigma^2 = 4 \cdot 0.998 \approx 4$ varianciával),

$$P(X \leq N) = \sum_{k=0}^{N} \binom{2000}{k} 0.002^k 0.998^{2000-k} \geq 0.99$$

és ez az egyenlőtlenség megoldandó az egyenlőtlenség a legkisebb N-re (amely a legkisebb költséget jelenti).

Ez az az eset, amikor ugyan X eloszlása ismert, de kezelhetetlenül bonyolult ahhoz, hogy eredményre jussunk. Ehelyett használhatjuk a Csebisev-egyenlőtlenséget:

$$P(|X-4| < \varepsilon) \geq 1 - \frac{4}{\varepsilon^2} = 0.99$$

ahonnan $\varepsilon = 20$ és így $N = 23$ adódik az új intenzív osztály ágyainak számára.

Általában a Csebisev-egyenlőtlenségtől (hiszen minden eloszlásra érvényes) nem várhatunk pontos becslést. Sokkal pontosabb eredményt kaphatunk, ha a Poisson-approximációt használjuk. Az approximációt leíró tételt lásd a 18.5 szakaszban. Nevezetesen

$$\sum_{k=0}^{N} \binom{2000}{k} 0.002^k 0.998^{2000-k} \approx \sum_{k=0}^{N} \frac{4^k}{k!} e^{-4}$$

hiszen "$n = 2000$ elég nagy és $p = 0.002$ elég kicsi", továbbá $np = 4$. A Poisson-eloszlás táblázatát vizsgálva (lásd Feladatgyűjtemény-2, 332-ik oldal) azt láthatjuk, hogy a jobb oldali szumma $N = 9$-nél lépi át a 0.99 értéket. Ezért ezen approximáció alapján a feladatra az $N = 9$ megoldás adódik. (Azt nem vizsgáljuk, hogy ez mennyire pontos közelítés.)

23.4. Nagy számok törvénye

Képzeljük el, hogy egy kísérletet elvégzünk egymás után n esetben, egymástól függetlenül, és minden esetben megfigyeljük, hogy egy bizonyos A esemény bekövetkezik-e, vagy sem (Bernoulli-kísérlet).

Tegyük fel, hogy az A esemény valószínűsége $P(A) = p$ (ahol $0 \leq p \leq 1$) és jelölje X_n azon kísérletek számát, amelyekben A bekövetkezik. Világos, hogy ekkor X_n/n az A esemény relatív gyakoriságát jelenti.

Azt szeretnénk megvizsgálni, hogy a relatív gyakoriság közelít-e a valódi valószínűséghez, ha növeljük a kísérletek számát, azaz $n \to \infty$?

Ez a kérdés elvi jelentőségű, hiszen a pozitív válasz azt mutatja, hogy az axiómáinkból egy tapasztalati tényt vezethetünk le. Ez azt jelenti, hogy az eddig felépített axiomatikus elméletünk a valóságot tükrözi, tehát helyesen választottuk meg az axiómáinkat.

Világos, hogy X_n binomiális eloszlású n és p paraméterekkel. Válasszunk egy $\varepsilon > 0$ számot és használjuk fel a Csebisev-egyenlőtlenséget.

$$P\left(\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right) = P(|X_n - np| \geq n\varepsilon)$$

Mivel $E(X_n) = np$ és $Var(X_n) = np(1-p)$, ez a Csebisev-egyenlőtlenség alapján úgy is írható, hogy

$$P(|X_n - np| \geq n\varepsilon) \leq \frac{np(1-p)}{n^2\varepsilon^2}$$

Mivel $p(1-p) \leq 1/4$ bármely p valós számra, innen

$$P\left(\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{1}{4n\varepsilon^2} \to 0$$

ha $n \to \infty$. Ezt az eredményünket az alábbi tételben fogalmazhatjuk meg.

23.6 Tétel. (Nagy számok Bernoulli-féle törvénye)

$$\lim_{n\to\infty} P\left(\left|\frac{X_n}{n} - p\right| < \varepsilon\right) = 1$$

bármely $\varepsilon > 0$ mellett.

23.7 Példa. Egy közvéleménykutató cég választási előrejelzést készít egy párt számára. Választók megkérdezésével kívánnak becslést adni a párt százalékos támogatottságára. Hány választót kell megkérdezniük, hogy a felmérésből adódó arány a valódi aránytól 90% valószínűséggel legfeljebb 1%-kal térjen el?

Jelölje $0 < p < 1$ az ismeretlen valódi arányt (a párt valóságos támogatottsága), erre kívánunk becslést kapni. Tegyük fel, hogy a minta (egyelőre ismeretlen) nagysága n és jelentse ezek közül X_n az adott párt támogatóinak számát.

Világos, hogy kísérletünk egy Bernoulli-kísérlet, ezért X_n binomiális eloszlású, így $E(X_n) = np$ és $Var(X_n) = np(1-p)$. Tehát a nagy számok törvénye szerint

$$P\left(\left|\frac{X_n}{n} - p\right| \leq 0.01\right) \geq 1 - \frac{1}{4n \cdot 10^{-4}}$$

A felmérésből adódó X_n/n arányra ezért a feltételünk miatt

$$1 - \frac{1}{4n \cdot 10^{-4}} = 0.90$$

teljesül, azaz $n = 25\,000$.

Tehát a cégnek a kívánt pontosság eléréséhez ekkora mintát kell választania a felméréshez. Természetesen a gyakorlatban bonyolultabb statisztikai módszerek alkalmazásával ennél kisebb minta is elegendő lehet. Ugyanakkor nehéz garantálni, hogy a megkérdezett választók halmaza homogén és reprezentatív legyen, abban az értelemben, hogy a mintából adódó arány tükrözze a teljes választói arányt.

A 23.6 Tétel (amelyet a nagy számok gyenge törvényének is neveznek) feltételei mellett az alábbi erősebb állítás is igazolható.

23.8 Tétel. *A 23.6 Tétel feltételei mellett*

$$P\left(\lim_{n\to\infty}\frac{X_n}{n}=p\right)=1$$

Szemléletesen megfogalmazva a 23.6 Tétel azt állítja, hogy a relatív gyakoriság nagyon valószínűen egyre közelebb lesz a p valószínűséghez. Azt azonban nem zárja ki, hogy bármilyen nagy n index fölött is lehetnek nagyobb kilengések, csak annyit mond, hogy ezek valószínűsége kicsi. A 23.8 Tétel viszont azt mondja, hogy az ilyen kilengések valószínűsége nulla. (A bizonyítás Ljapunovtól, illetve általánosabb formában Kolmogorovtól származik a múlt század harmincas éveiből.)

Otthoni tanuláshoz

1. A Feladatgyűjtemény-2 IV/3 szakaszának feldolgozása.
2. Házi feladatok: a IV/3 szakasz 396, 397, 398, 402, 403, 404, 408 és 409 feladatai.
3. Tankönyv-2 5.1, 5.2 és 5.3 szakaszai.

24. fejezet

A statisztika nevezetes eloszlásai

24.1. Kétdimenziós normális eloszlás

24.1 Definíció. Legyen r olyan adott valós szám, amelyre $|r| < 1$. Ha az X és Y változók együttes sűrűségfüggvénye

$$h(x,y) = \frac{1}{2\pi\sqrt{1-r^2}} e^{-\frac{x^2-2rxy+y^2}{2(1-r^2)}} \tag{24.1}$$

akkor azt mondjuk, hogy X és Y normális együttes eloszlású r paraméterrel.

24.2 Tétel. *Ha X és Y együttes sűrűségfüggvénye a (24.1) alatti függvény, akkor X és Y peremeloszlásai egyaránt standard normális eloszlások.*

Bizonyítás. Mivel a sűrűségfüggvény a két változóban szimmetrikus, ezért elég az állítást az X változó peremsűrűségfüggvényére igazolni.

Ha f jelenti az X sűrűségfüggvényét, akkor

$$f(x) = \int_{-\infty}^{\infty} h(x,y)\,dy = \frac{1}{2\pi\sqrt{1-r^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2-2rxy+y^2}{2(1-r^2)}}\,dy$$

Teljes négyzetté alakítással az integrál mögötti exponenciális függvény úgy írható, hogy

$$-\frac{x^2-2rxy+y^2}{2(1-r^2)} = -\frac{x^2}{2} - \frac{(y-rx)^2}{2(1-r^2)}\,.$$

Ha az integrálból kiemeljük az y-tól nem függő részt, akkor azt kapjuk, hogy

$$f(x) = \frac{1}{2\pi\sqrt{1-r^2}} e^{-\frac{x^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{(y-rx)^2}{2(1-r^2)}}\, dy$$

Vezessük be ebben az improprius integrálban a

$$t = \frac{y-rx}{\sqrt{1-r^2}}$$

helyettesítést, akkor az adódik, hogy

$$f(x) = \frac{1}{2\pi\sqrt{1-r^2}} e^{-\frac{x^2}{2}} \int_{-\infty}^{\infty} e^{-t^2/2} \cdot \sqrt{1-r^2}\, dt$$

A $\sqrt{1-r^2}$ kifejezéssel egyszerűsítve láthatjuk, hogy az utolsó integrál éppen a Gauss-integrál, amelynek értéke $\sqrt{2\pi}$, tehát

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad -\infty < x < \infty$$

ami éppen a standard normális sűrűségfüggvény. $\square$

24.2. Korrelálatlan normális eloszlások

24.3 Definíció. Az X és Y valószínűségi változókat *korrelálatlannak* nevezzük, ha $Corr(X,Y) = 0$.

Amint láttuk, a független változók korrelálatlanok, de példán láttuk, hogy ennek az állításnak a megfordítása általában nem érvényes. Normális eloszlású változók esetében azonban a két fogalom ekvivalens.

Tekintsük ebben a szakaszban újra a

$$h(x,y) = \frac{1}{2\pi\sqrt{1-r^2}} e^{-\frac{x^2-2rxy+y^2}{2(1-r^2)}}$$

kétdimenziós normális sűrűségfüggvényt, és határozzuk meg az X és Y változók $Corr(X,Y)$ korrelációs együtthatóját. Az előző szakaszban láttuk, hogy X és Y peremeloszlásai standard normális eloszlások, ezért

$$E(X) = E(Y) = 0 \quad \text{és} \quad D(X) = D(Y) = 1$$

A szorzat várható értékét az

$$E(XY) = \int_{-\infty}^{\infty} xyh(x,y)\, dx\, dy = \frac{1}{2\pi\sqrt{1-r^2}} \int_{-\infty}^{\infty} xye^{-\frac{x^2-2rxy+y^2}{2(1-r^2)}}\, dx\, dy$$

integrál szolgáltatja. Ezt az integrált kétváltozós helyettesítéssel lehet kiszámolni, ez az eljárás azonban túlmegy a tankönyvünk keretein. A részletes számítást megtaláljuk a Tankönyv-2 198-ik oldalán. Eszerint $E(XY) = r$, és innen azonnal következik, hogy

$$Corr(X, Y) = r\,.$$

Mivel $r = 0$ esetén a h együttes sűrűségfüggvény láthatóan két standard normális sűrűségfüggvény szorzatára bomlik, azaz X és Y függetlenek, megfogalmazhatjuk a következő állítást.

24.4 Tétel. *Ha az X és Y változók együttes eloszlását a (24.1) alatti együttes sűrűségfüggvény szolgáltatja, úgy X és Y akkor és csak akkor függetlenek, ha korrelálatlanok.*

24.3. Normálisból származtatott eloszlások

24.5 Példa. Állítsuk elő a standard normális eloszlású Z változó négyzetének eloszlását. Jelölje H, illetve h a Z^2 eloszlás, illetve sűrűségfüggvényét.

Ekkor tetszőleges $x > 0$ mellett

$$H(x) = P(Z^2 < x) = P(|Z| < \sqrt{x}) = P(-\sqrt{x} < Z < \sqrt{x}) = 2\Phi(\sqrt{x}) - 1$$

továbbá $H(x) = 0$, ha $x \leq 0$. Innen deriválással adódik, hogy

$$h(x) = \frac{1}{2\sqrt{x}} 2\varphi(\sqrt{x}) = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}} \quad \text{ha} \quad x > 0\,,$$

és $h(x) = 0$, ha $x \leq 0$. Ennek az eloszlásnak a várható értékére és szórására

$$E(Z^2) = 1\,, \quad Var(Z^2) = 2$$

adódik.

24.6 Példa. Legyenek most Z_1 és Z_2 független, standard normális eloszlású valószínűségi változók, és tekintsük a

$$\chi^2 = Z_1^2 + Z_2^2$$

összeget. Jelentse h a Z_1^2 és Z_2^2 változók sűrűségfüggvényét, továbbá k a χ^2 sűrűségfüggvényét. Ekkor az előző példa alapján a konvolúciós integrál a következő alakot ölti:

$$k(x) = \int_{-\infty}^{\infty} h(s)h(x-s)\,ds\,.$$

Az integrandus akkor és csak akkor nem nulla, ha $s > 0$ és $x - s > 0$, azaz ha $0 < s < x$. Tehát a konvolúciós integrál úgy írható, hogy

$$k(x) = \frac{1}{2\pi}\int_0^x s^{-\frac{1}{2}}e^{-\frac{s}{2}}(x-s)^{-\frac{1}{2}}e^{-\frac{x-s}{2}}\,ds = \frac{1}{2\pi}e^{-\frac{x}{2}}\int_0^x (sx-s^2)^{-\frac{1}{2}}\,ds$$

Ez utóbbi integrál elég komplikált, kiszámítása az úgynevezett Euler-féle béta-függvényhez vezet (lásd a Tankönyv-2 247-ik oldalát), és az értéke π. Ennek alapján

$$k(x) = \frac{1}{2}e^{-\frac{x}{2}} \quad \text{ha} \quad x > 0$$

és $k(x) = 0$, ha $x \le 0$, ami éppen az $1/2$-paraméterű exponenciális eloszlás. Ezért

$$E(\chi^2) = 2\,, \quad \text{és} \quad Var(\chi^2) = 4$$

adódik a várható értékre és a varianciára.

24.4. A χ^2-eloszlás, a t-eloszlás és az F-eloszlás

24.7 Definíció. Tekintsük a $Z_1, \dots Z_n$ független, standard normális eloszlású valószínűségi változókat. Ekkor a

$$\chi_n^2 = Z_1^2 + \ldots + Z_n^2$$

változót *n-szabadságfokú χ^2-eloszlásúnak* nevezzük.

Ennek a változónak a várható értéke és varianciája

$$E(\chi_n^2) = n \quad \text{és} \quad Var(\chi_n^2) = 2n$$

az előző szakasz példája alapján.

24.8 Definíció. Tekintsük a $Z_1, \dots Z_n$ és Z független, standard normális eloszlású valószínűségi változókat. Ekkor a

$$t_n = \frac{Z\sqrt{n}}{\sqrt{Z_1^2 + \ldots + Z_n^2}}$$

valószínűségi változót *n-szabadságfokú t-eloszlásnak* nevezzük.

Ennek az eloszlásnak $n = 1$ esetén nem létezik várható értéke, míg $n \ge 2$ mellett $E(t_n) = 0$. Az is megmutatható, hogy $n < 3$ esetén a variancia nem létezik, továbbá

$$Var(t_n) = \frac{n}{n-2}$$

ha $n \geq 3$.

24.9 Definíció. Legyenek most $X_1, \ldots, X_m$ és $Y_1, \ldots, Y_n$ független, standard normális eloszlású valószínűségi változók. Ekkor az

$$F_{m,n} = \frac{\frac{1}{m}\sum_{k=1}^{m} X_k^2}{\frac{1}{n}\sum_{k=1}^{n} Y_k^2}$$

változót (m,n)-*szabadságfokú* F-*eloszlásnak* nevezzük.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-2 VI/5 és VI/6 szakaszainak feldolgozása.
2. Házi feladatok: a VI/6 szakasz feladatai. Egész féléves ismétlés és az internetes oldalon kijelölt minta vizsgafeladatok gyakorlása.
3. Tankönyv-2 6.15, 9.5, 9.6 és 9.7 szakaszai.

III. rész

Harmadik félév: Lineáris algebra

25. fejezet

Vektorterek és alterek

25.1. Az $\mathbb{R}^n$ vektortér

Legyen n adott természetes szám. Az $\mathbb{R}^n$ halmazt az összes valós szám n-esek halmazaként értelmezzük, azaz:

$$\mathbb{R}^n = \left\{ x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} : x_1, \ldots, x_n \in \mathbb{R} \right\}$$

E halmaz elemeit vektoroknak, a komponenseit koordinátáknak nevezzük. A szokásos koordinátákra gondolva ez a halmaz $n = 2$ esetén a síkot, $n = 3$ esetén a háromdimenziós teret adja.

A továbbiakban a vektorokat általában kis latin betűkkel, a valós számokat, vagy más néven skalárokat, kis görög betűkkel jelöljük.

Az $\mathbb{R}^n$ vektorai között az alábbi műveleteket értelmezzük:

Vektorok összege

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \text{ és } y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \text{ esetén } x + y = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix} \in \mathbb{R}^n$$

Vektor szorzása skalárral

$$\alpha \in \mathbb{R} \text{ és } x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \text{ esetén } \alpha x = \begin{bmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{bmatrix} \in \mathbb{R}^n$$

Ezekkel a műveletekkel ellátva az $\mathbb{R}^n$ halmazt *vektortérnek* nevezzük.

25.1 Definíció. Legyenek $a_1, \ldots, a_k$ vektorok a vektortérben, és $\alpha_1, \ldots, \alpha_k$ tetszőleges valós számok (skalárok). Az

$$\alpha_1 a_1 + \ldots + \alpha_k a_k$$

vektort az $a_1, \ldots, a_k$ vektorok egy *lineáris kombinációjának* nevezzük.

25.2 Példa. Legyenek például

$$a_1 = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} \text{ és } a_2 = \begin{bmatrix} 3 \\ 0 \\ 4 \end{bmatrix} \text{ továbbá } \alpha_1 = 3 \text{ és } \alpha_2 = -2$$

akkor

$$\alpha_1 a_1 + \alpha_2 a_2 = \begin{bmatrix} 0 \\ -3 \\ 1 \end{bmatrix}$$

25.2. Alterek

25.3 Definíció. Az $\mathbb{R}^n$ vektortér egy M részhalmazát altérnek nevezzük, ha

- bármely $x, y \in M$ esetén $x + y \in M$, és
- bármely $x \in M$ és $\alpha \in \mathbb{R}$ esetén $\alpha x \in M$.

A definícióból világos, hogy egy altér a 0 vektort mindig tartalmazza. A legszűkebb altér az egyelemű $\{0\}$ altér, a legbővebb az egész vektortér.

25.4 Tétel. *Ha M altér, akkor bármely $a_1, \ldots, a_k \in M$ vektorok és bármely $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$ skalárok esetén*

$$\alpha_1 a_1 + \ldots + \alpha_k a_k \in M$$

Könnyen látható, hogy például az $\mathbb{R}^3$ vektortérben az origón átmenő egyenesek illetve síkok egyaránt alterek.

25.5 Példa. Tekintsük az $\mathbb{R}^n$ vektortér következő részhalmazát:

$$M = \left\{ x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n : x_1 + x_2 = 0 \right\}$$

Mutassuk meg, hogy M altér.

Valóban, ha $x, y \in M$, akkor $x_1 + x_2 = 0$ és $y_1 + y_2 = 0$, ezért az $x + y$ vektor első két koordinátájára is $(x_1 + y_1) + (x_2 + y_2) = 0$. Ez azt jelenti, hogy $x + y \in M$.

Teljesen hasonlóan, ha $x \in M$ és $\alpha \in \mathbb{R}$, akkor az $x_1 + x_2 = 0$ egyenlőségből adódik, hogy $\alpha(x_1 + x_2) = 0$, tehát $\alpha x \in M$.

Könnyen elképzelhető, hogy $n = 3$ estén a fenti M altér egy olyan, az alapsíkra merőleges sík, amely tartalmazza a harmadik tengelyt, és az alapsíkot a $-45°$-os egyenesben metszi. KÉSZÍTSÜNK ÁBRÁT!

Az is könnyen ellenőrizhető, hogy ha az M halmaz definíciójában $x_1 + x_2$ értékének bármilyen nullától különböző számot írtunk volna elő, akkor az M nem lett volna altér. Ekkor az összeadás és a skalárral való szorzás is kivezetett volna az M halmazból.

25.6 Tétel. *Alterek metszete is altér.*

Bizonyítás. Elég az állítást két altérre igazolni, a bizonyítás tetszőleges számú altérre hasonlóan megy.

Legyenek L és M alterek. Ha $x, y \in L \cap M$, akkor $x + y \in L$ és $x + y \in M$, mert mindketten alterek. Tehát $x + y \in L \cap M$.

Teljesen hasonlóan, ha $x \in L \cap M$ és $\alpha \in \mathbb{R}$, akkor $\alpha x \in L$ és $\alpha x \in M$, mert mindketten alterek. Ezért $\alpha x \in L \cap M$. □

25.3. Generált altér

A 25.6 Tétel alapján beszélhetünk adott vektorokat tartalmazó legszűkebb altérről. Ezt fogalmazza meg a következő definíció.

25.7 Definíció. Az $a_1, \ldots, a_k$ vektorokat tartalmazó legszűkebb alteret, jelölésben:

$$\mathrm{lin}\{a_1, \ldots, a_k\}$$

ezen vektorokat tartalmazó összes altér metszeteként értelmezzük, és az adott vektorok által *generált altérnek* nevezzük.

Mivel egy altér a vektorainak összes lineáris kombinációit tartalmazza, és ezek már önmagukban is alteret alkotnak, megfogalmazhatjuk az alábbi tételt.

25.8 Tétel. *Az $a_1, \ldots, a_k$ vektorok által generált altér ezen vektorok összes*

$$\alpha_1 a_1 + \ldots + \alpha_k a_k$$

lineáris kombinációinak halmaza, ahol $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$.

25.9 Példa. Például, ha az $\mathbb{R}^3$ vektortérben tekintjük az

$$a_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ és } a_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

vektorokat, akkor

$$\mathrm{lin}\{a_1, a_2\} = \left\{ x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3 : x_3 = 0 \right\}$$

azaz mindazon vektorok halmaza, amelyek harmadik koordinátája nulla. Gondoljuk meg, hogy ez a halmaz valóban altér!

25.4. Lineáris függetlenség

25.10 Definíció. Az $\mathbb{R}^n$ vektortér $a_1, \ldots, a_k$ vektorait *lineárisan függetlennek* nevezzük, ha valamely $\alpha_1, \ldots, \alpha_k$ skalárokra

$$\alpha_1 a_1 + \ldots + \alpha_k a_k = 0 \Rightarrow \alpha_1 = \ldots = \alpha_k = 0\,.$$

Ellenkező esetben a vektorokat *lineárisan összefüggőknek* nevezzük.

A lineáris függetlenség az algebra egyik legfontosabb alapfogalma, azt fogalmazza meg, hogy a vektorok egy lineáris kombinációja CSAK úgy lehet 0, ha minden együttható 0.

FIGYELEM! A definíció nem azt mondja, hogy ha minden együttható 0, akkor a lineáris kombináció is 0, ez ugyanis nyilvánvaló! Az implikáció fordított irányú!

Független vektorok közül egyik sem fejezhető ki a többi lineáris kombinációjaként. Ezt fogalmazza meg a következő tételünk.

25.11 Tétel. *Az $a_1, \ldots, a_k$ vektorok akkor és csak akkor lineárisan összefüggők, ha valamelyik kifejezhető a többi lineáris kombinációjaként.*

Bizonyítás. Ha valamelyik vektor, például a_1 kifejezhető a többi vektor lineáris kombinációjaként, akkor

$$a_1 = \alpha_2 a_2 + \ldots + \alpha_k a_k\,.$$

Ezt átrendezve azt kapjuk, hogy

$$-a_1 + \alpha_2 a_2 + \ldots + \alpha_k a_k = 0\,.$$

Ez azt jelenti, hogy a lineáris kombináció úgy nulla, hogy az első együttható nem nulla. Tehát a vektorok nem függetlenek.

Fordítva, tegyük fel, hogy a vektorok összefüggők. Ekkor van olyan lineáris kombináció, amelyre

$$\alpha_1 a_1 + \ldots + \alpha_k a_k = 0\,,$$

de nem minden együttható nulla, például éppen $\alpha_1 \neq 0$. Akkor

$$a_1 = -\frac{\alpha_2}{\alpha_1} a_2 - \ldots - \frac{\alpha_k}{\alpha_1} a_k\,,$$

azaz a_1 kifejezhető a többi vektor lineáris kombinációjaként. □

25.12 Példa. Tekintsük a következő vektorokat az $\mathbb{R}^3$ térben

$$a_1 = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} \quad a_2 = \begin{bmatrix} 3 \\ 0 \\ 4 \end{bmatrix} \quad a_3 = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$$

és döntsük el, hogy lineárisan függetlenek-e.

Egyszerű számolással láthatjuk, hogy $a_3 = 2a_1 - a_2$, ezért a fenti vektorok összefüggők.

Az alábbi egyszerű állítások könnyen ellenőrizhetők a definíció alapján.

25.13 Tétel. *Tekintsük az $a_1, \ldots, a_k$ vektorokat az $\mathbb{R}^n$ vektortérben.*

- *Ha a vektorok lineárisan függetlenek, akkor bármely részhalmazuk is az.*
- *Ha a vektorok között szerepel a 0, akkor lineárisan összefüggők.*
- *Ha a vektorok között szerepel két azonos, akkor lineárisan összefüggők.*
- *Ha a vektorok összefüggők, akkor bármilyen vektorral bővítve is összefüggők.*

Bizonyítás. Csak útmutatót adunk a bizonyítás elvégzéséhez, a részletek kidolgozása házi feladat.

- Tekintsük a részhalmaz egy lineáris kombinációját, és szerepeltessük a hiányzó vektorokat nulla együtthatóval.
- Tekintsük a vektorok egy olyan lineáris kombinációját, amelyben a nulla vektor az 1, az összes többi vektor nulla együtthatóval szerepel.
- Tekintsük a vektorok egy olyan lineáris kombinációját, amelyben a két azonos vektor +1 illetve −1 együtthatóval, az összes többi vektor nulla együtthatóval szerepelnek.

- Tekintsük a vektorok egy olyan lineáris kombinációját, amely nulla, de nem minden együttható nulla, és szerepeltessük a kibővített vektorokat nulla együtthatóval.

25.14 Példa. Tegyük fel, hogy az a, b és c vektorok lineárisan függetlenek. Vajon igaz-e, hogy az $a+b$, $b+c$, $c+a$ vektorok is lineárisan függetlenek?

Irjuk fel az adott vektorok egy lineáris kombinációját, amelyről tegyük fel, hogy nulla:

$$\alpha_1(a+b)+\alpha_2(b+c)+\alpha_3(c+a)=0\,.$$

Rendezve az egyenlőséget

$$(\alpha_1+\alpha_3)a+(\alpha_1+\alpha_2)b+(\alpha_2+\alpha_3)c=0$$

A függetlenség miatt innen az adódik, hogy

$$\alpha_1+\alpha_3=0 \quad \alpha_1+\alpha_2=0 \quad \alpha_2+\alpha_3=0$$

Ennek egyetlen megoldása $\alpha_1=\alpha_2=\alpha_3=0$, azaz az $a+b$, $b+c$ és $c+a$ vektorok is lineárisan függetlenek.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 II/1 és II/3 szakaszok kidolgozott példáinak feldolgozása.
2. Házi feladatok: a II/1 szakasz 1.1.3, 1.1.4, 1.2.2, 1.2.3, továbbá a II/3 szakasz 3.1.3, 3.1.4, 3.2.3, 3.2.4, 3.3.2 és 3.3.4 feladatai. feladatai.
3. Tankönyv-1 12.1, 12.2, 12.3, és 12.4 szakaszai.

26. fejezet

Lineáris függetlenség és bázis

26.1. Generátorrendszer

26.1 Definíció. Az $\mathbb{R}^n$ vektortérben az $a_1, \ldots, a_k$ vektorrendszert *generátorrendszernek* nevezzük, ha

$$\operatorname{lin}\{a_1, \ldots, a_k\} = \mathbb{R}^n\,,$$

azaz a tér minden vektora előáll az adott vektorok lineáris kombinációjaként.

26.2 Példa. Tekintsük a következő vektorokat az $\mathbb{R}^3$ térben

$$a_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad a_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad a_3 = \begin{bmatrix} 3 \\ -2 \\ 0 \end{bmatrix}$$

és döntsük el, hogy generátorrendszert alkotnak-e.

Könnyű meggondolni, hogy nem alkotnak generátorrendszert, hiszen egyetlen olyan vektort sem tudunk kifejezni, amelyiknek a harmadik koordinátája nem nulla. Ezek a vektorok nem is függetlenek, hiszen $a_3 = 3a_1 - 2a_2$.

Ugyanakkor a következő vektorrendszer

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

generátorrendszert alkot, hiszen bármely x vektor előállítható az alábbi módon:

$$x = x_1 e_1 + x_2 e_2 + x_3 e_3\,,$$

ahol x_1, x_2, x_3 az x vektor koordinátái. Egyszerűen ellenőrizhető, hogy ezek a vektorok egyébként lineárisan függetlenek is.

Teljesen hasonlóan értelmezhetjük az $\mathbb{R}^n$ valamely M alterének generátorrendszerét is.

26.3 Definíció. Azt mondjuk, hogy az $a_1, \ldots, a_k$ vektorok az M altér generátorrendszerét alkotják (vagy az M alteret generálják), ha az M bármely vektora előállítható az $a_1, \ldots, a_k$ vektorok lineáris kombinációjaként.

26.2. Bázis

26.4 Definíció. Az $\mathbb{R}^n$ vektortérben az $a_1, \ldots, a_k$ vektorrendszert *bázisnak* nevezzük, ha

- lineárisan független rendszer, és
- generátorrendszer.

Analóg módon értelmezzük valamely altér bázisát is.

26.5 Példa. Ahogy az előző példánkban láttuk, az $\mathbb{R}^3$ vektortérben az

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

vektorrendszer bázis, hiszen egyszerre lineárisan független és generátorrendszer. Teljesen hasonlóan látható, hogy az

$$a_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad a_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \quad a_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

vektorrendszert ugyancsak bázist alkot. (Ellenőrizzük!) Ugyanakkor a

$$a_1 = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} \quad a_2 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$$

vektorrendszer nem bázis, hiszen nem alkot generátorrendszert (bár lineárisan független).

A definíciója alapján ellenőrizhetjük a bázis következő tulajdonságait.

- Egy bázis maximális elemszámú lineárisan független rendszer.
- Egy bázis minimális elemszámú generátorrendszer.
- Egy vektortérben minden bázis azonos elemszámú. (FIGYELEM: NEM NYILVÁNVALÓ!)
- Egy vektortérben bármely vektor egyértelműen írható fel a bázisvektorok lineáris kombinációjaként.

26.6 Definíció. Az $\mathbb{R}^n$ vektortérben az alábbi vektorrendszert *standard bázisnak* nevezzük:

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad \dots \quad e_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

ezekre a vektorokra a továbbiakban mindig az e_k jelölést használjuk ($k = 1, \dots, n$).

Ellenőrizzük, hogy ezek a vektorok valóban bázist alkotnak! Bizonyos értelemben a vektortérben ez a "legegyszerűbb" bázis, hiszen bármely x vektor esetében

$$x = x_1 e_1 + \dots + x_n e_n \,.$$

ahol $x_1, \dots, x_n$ az x vektor koordinátáit jelölik.

26.3. Dimenzió

A bázis tulajdonságai alapján bevezethetjük az alábbi definíciót.

26.7 Definíció. Egy vektortér vagy altér dimenzióján a benne található maximális lineárisan független rendszer (azaz bázis) elemszámát értjük.

26.8 Példa. A 26.6 Példa értelmében bármely n természetes számra

$$\dim \mathbb{R}^n = n$$

hiszen az $e_1, \dots, e_n$ független rendszer maximális, azaz tovább nem bővíthető.

26.9 Példa. Tekintsük most az

$$a_1 = \begin{bmatrix} 2 \\ 2 \\ -3 \end{bmatrix} \quad a_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \quad a_3 = \begin{bmatrix} 5 \\ 3 \\ -5 \end{bmatrix}$$

vektorok által generált M alteret, és határozzuk meg a dimenzióját!

Könnyen ellenőrizhető, hogy a_1 és a_2 lineárisan függetlenek, de a_1, a_2, a_3 már nem, hiszen

$$2a_1 + a_2 = a_3\,.$$

Ezért a generált altér dimenziójára az adódik, hogy

$$\dim M = \dim \operatorname{lin}\{a_1, a_2, a_3\} = 2\,.$$

Az előző megállapításunk szerint persze az is igaz, hogy $\dim \operatorname{lin}\{a_1, a_2\} = 2$.

26.10 Definíció. Valamely $a_1, \ldots, a_k$ vektorok által generált altér dimenzióját a vektorrendszer *rangjának* is nevezzük, jelölésben:

$$\operatorname{rank}\{a_1, \ldots, a_k\} = \dim \operatorname{lin}\{a_1, \ldots, a_k\}\,.$$

26.4. Elemi bázistranszformáció

Ebben a szakaszban egy nagyon egyszerű és széleskörűen használható eljárást mutatunk be, amellyel gyorsan ellenőrizhető, hogy egy adott vektorrendszer lineárisan független-e.

Tekintsünk az $\mathbb{R}^n$ vektortérben egy a vektort, amely a standard bázissal az

$$a = \alpha_1 e_1 + \ldots + \alpha_n e_n \tag{26.1}$$

alakban írható fel. Tekintsünk egy másik b vektort is, amelynek előállítása

$$b = \beta_1 e_1 + \ldots + \beta_n e_n \quad \text{ahol} \quad \beta_1 \neq 0\,. \tag{26.2}$$

KÉRDÉS: Milyen lineáris kombinációval adható meg az a vektor, ha a standard bázis helyett a $b, e_2, \ldots, e_n$ bázist tekintjük, azaz az e_1 vektort a b vektorra cseréljük?

MEGJEGYZÉS: Könnyen látható, hogy a $\beta_1 \neq 0$ feltétel miatt a $b, e_2, \ldots, e_n$ rendszer bázis, hiszen egyrészt n-elemű, másrészt b független a többi vektortól (hiszen b előállításához szükség van az e_1 vektorra is).

Fejezzük ki az e_1 vektort a (26.2) egyenlőségből:

$$e_1 = \frac{1}{\beta_1} b - \frac{\beta_2}{\beta_1} e_2 - \ldots - \frac{\beta_n}{\beta_1} e_n$$

majd helyettesítsük ezt be e_1 helyére a (26.1) egyenlőségbe. Rendezés után azt kapjuk, hogy

$$a = \frac{\alpha_1}{\beta_1} b + \left(\alpha_2 - \frac{\alpha_1}{\beta_1}\beta_2\right) e_2 + \ldots + \left(\alpha_n - \frac{\alpha_1}{\beta_1}\beta_n\right) e_n\,. \tag{26.3}$$

Ez az eljárás, amelyet *elemi bázistranszformációnak* nevezünk, megadja az a vektor előállítását az új $b, e_2, \ldots, e_n$ bázisban.

26.11 Példa. Az elemi bázistranszformáció használatával vizsgáljuk meg, hogy az

$$a = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \quad b = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} \quad c = \begin{bmatrix} 3 \\ 3 \\ -4 \end{bmatrix}$$

vektorok lineárisan függetlenek-e. Az eljárás a következő:

$$\left|\begin{array}{rrr|rr|r} \boxed{1} & 2 & 3 & 2 & 3 & -1 \\ -1 & 1 & 3 & \boxed{3} & 6 & 2 \\ 2 & -1 & -4 & -5 & -10 & 0 \end{array}\right|$$

Ez a számolás azt mutatja, hogy a vektorok nem lineárisan függetlenek, nevezetesen a c vektor kifejezhető az a és b vektorok segítségével, éspedig $c = -a + 2b$. Ekkor az a, b, c vektorrendszer rangja 2, azaz

$$\dim \operatorname{lin}\{a, b, c\} = 2\,.$$

26.12 Példa. Tekintsük az $\mathbb{R}^4$ tér következő vektorait

$$a_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 2 \end{bmatrix} \quad a_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ -1 \end{bmatrix} \quad a_3 = \begin{bmatrix} -1 \\ -2 \\ -3 \\ 8 \end{bmatrix} \quad a_4 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 3 \end{bmatrix}$$

és határozzuk meg, hogy az $\mathbb{R}^4$ vektortérben hány dimenziós az a_1, a_2, a_3, a_4 vektorok által generált

$$M = \operatorname{lin}\{a_1, a_2, a_3, a_4\}$$

altér!

Végezzük el az elemi bázistranszformációt a megadott vektorokra, akkor azt kapjuk, hogy

$$\left|\begin{array}{rrrr|rrr|rr} \boxed{1} & 2 & -1 & 1 & 2 & -1 & 1 & 3 & -1 \\ 0 & 1 & -2 & 1 & \boxed{1} & -2 & 1 & -2 & 1 \\ -1 & 0 & -3 & 1 & 2 & -4 & 2 & 0 & 0 \\ 2 & -1 & 8 & -3 & -5 & 10 & -5 & 0 & 0 \end{array}\right|$$

Innen azt láthatjuk, hogy az a_3 és a_4 vektorok lineárisan függnek az a_1 és a_2 vektoroktól, nevezetesen

$$a_3 = 3a_1 - 2a_2$$

illetve

$$a_4 = -a_1 + a_2$$

Tehát az a_1, a_2, a_3, a_4 vektorrendszerben maximálisan két lineárisan független vektor van, ilyen pl. a_1 és a_2. Innen adódik, hogy

$$\dim M = \dim \operatorname{lin}\{a_1, a_2, a_3, a_4\} = 2\,.$$

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 II/3 szakasza kidolgozott példáinak feldolgozása.
2. Házi feladatok: a II/3 szakasz 3.1.4, 3.1.5, 3.2.4, 3.2.5, 3.3.2, 3.3.5 feladatai.
3. Tankönyv-1 14.1 szakasza.

27. fejezet

Lineáris leképezések és mátrixok

27.1. Lineáris leképezések

Legyenek n és m természetes számok, és tekintsük az $\mathbb{R}^n$ és $\mathbb{R}^m$ n, illetve m-dimenziós vektortereket.

27.1 Definíció. Az $A : \mathbb{R}^n \to \mathbb{R}^m$ leképezést *lineáris leképezésnek* nevezzük, ha

- $A(x + y) = Ax + Ay$
- $A(\alpha x) = \alpha Ax$

bármely $x, y \in \mathbb{R}^n$ vektorok és $\alpha \in \mathbb{R}$ skalár esetén.

Ha $n = m$, azaz a leképezés a vektorteret önmagába képezi, akkor *lineáris transzformációról* beszélünk.

Könnyen ellenőrizhetjük egy lineáris leképezés következő tulajdonságait:

- $A(\alpha x + \beta y) = \alpha Ax + \beta Ay$ bármely x, y vektorokra és α, β skalárokra.
- $A0 = 0$, azaz a 0 vektor képe mindig a 0 vektor.

27.2 Példa. Az alábbiakban megadunk olyan leképezéseket, amelyek az $\mathbb{R}^2$ síkot önmagába képezik.

1. Legyen A az a leképezés, amely bármely x vektorhoz a λ-szorosát rendeli, azaz $Ax = \lambda x$.

2. Legyen A az a leképezés, amely bármely x vektorhoz a vízszintes tengelyre vonatkozó tükörképét rendeli.
3. Legyen A az a leképezés, amely bármely vektorhoz az $y = x$ $45°$-os egyenesre vonatkozó vetületét rendeli.
4. Legyen A az a leképezés, amely bármely x vektorhoz az origó körül φ szöggel pozitív irányba történő elforgatottját rendeli.

Mutassuk meg, hogy a fenti példák mindegyikében A az $\mathbb{R}^2$ sík egy lineáris transzformációját definiálja, azaz kielégíti a definíció két egyenlőségét. Készítsünk ábrákat!

27.2. Leképezések mátrixa

Az alábbiakban azt vizsgáljuk meg, hogy ha egy lineáris leképezés esetén ismerjük a bázisvektorok képeit, akkor már tetszőleges vektor képét elő tudjuk állítani.

Legyen ugyanis $A : \mathbb{R}^n \to \mathbb{R}^m$ lineáris leképezés, továbbá legyen $e_1, \ldots, e_n$ a standard bázis az $\mathbb{R}^n$ vektortérben. Ha most $x \in \mathbb{R}^n$ egy tetszőleges vektor, akkor x felírható a standard bázis lineáris kombinációjaként:

$$x = x_1 e_1 + \ldots + x_n e_n \,.$$

Alkalmazzuk mindkét oldalra az A leképezést, akkor a linearitás tulajdonságát felhasználva

$$Ax = x_1 A e_1 + \ldots + x_n A e_n \,,$$

azaz Ax megadásához csak az $Ae_1, \ldots, Ae_n$ vektorokra van szükség.

Tegyük fel ezért, hogy $f_1, \ldots, f_m$ a standard bázis az $\mathbb{R}^m$ vektortérben, és írjuk fel az $Ae_1, \ldots, Ae_n$ vektorokat a bázis lineáris kombinációiként ebben a térben:

$$\begin{aligned} Ae_1 &= a_{11} f_1 + a_{21} f_2 + \ldots + a_{m1} f_m \\ Ae_2 &= a_{12} f_1 + a_{22} f_2 + \ldots + a_{m2} f_m \\ &\vdots \\ Ae_n &= a_{1n} f_1 + a_{2n} f_2 + \ldots + a_{mn} f_m \end{aligned}$$

Ezekben az egyenlőségekben álló együtthatókat egy táblázatba gyűjtve az alábbi $m \times n$-es *mátrixhoz* jutunk, amelyet az A leképezés *mátrixának* nevezünk.

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ & & \vdots & \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}$$

amelynek m sora és n oszlopa van. A mátrix j-ik oszlopa az Ae_j eleőállítása az $\mathbb{R}^m$ tér $f_1, \ldots, f_m$ bázisában. Ezáltal egy x vektor Ax képét úgy állíthatjuk elő, hogy az A mátrixát megszorozzuk az x vektor koordinátáival az alábbi módon:

$$Ax = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ & & \vdots & \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \ldots + a_{1n}x_n \\ a_{21}x_1 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n \end{bmatrix} \tag{27.1}$$

azaz, az Ax szorzat egy m-koordinátájú $\mathbb{R}^m$ térbeli vektor lesz.

Világos, hogy minden lineáris leképezés az adott bázisban mátrix-alakban írható fel. Ugyanakkor fordítva, az is világos, hogy a (27.1) egyenlőséggel a vektortéren egy lineáris leképezést értelmezünk. Tehát kölcsönösen egyértelmű megfeleltetést találtunk a lineáris leképezések és a mátrixok között.

FIGYELEM! A továbbiakban nem teszünk különbséget a lineáris leképezések és mátrixaik között.

27.3 Példa. Tekintsük a 27.2 Példában értelmezett lineáris transzformációkat, akkor azok mátrixai sorrendben:

1. $A = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$ 2. $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ 3. $A = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$
4. $A = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$

Készítsünk ábrát, és azon ellenőrizzük!

27.3. Mátrix rangja és szabadságfoka

27.4 Definíció. Tekintsünk egy $A : \mathbb{R}^n \to \mathbb{R}^m$ lineáris leképezést (azaz $m \times n$-es mátrixot). Az A értékészletét

$$\operatorname{im} A = \{y \in \mathbb{R}^m : \text{ van olyan } x \in \mathbb{R}^n, \text{ hogy } y = Ax\}$$

az A *képterének*, míg a

$$\ker A = \{x \in \mathbb{R}^n : Ax = 0\}$$

halmazt az A *magterének* nevezzük. Világos, hogy $\ker A$, illetve $\operatorname{im} A$ egyaránt alterek az $\mathbb{R}^n$ illetve $\mathbb{R}^m$ vektorterekben.

A definíció szerint az $\operatorname{im} A$ altér nem más, mint azon vektorok altere, amelyek kifejezhetők az A oszlopainak lineáris kombinációiként. Ha tehát $a_1, \ldots, a_n$ jelölik az A oszlopvektorait, akkor

$$\operatorname{im} A = \operatorname{lin}\{a_1, \ldots, a_n\}.$$

Teljesen hasonlóan a $\ker A$ altér azon x vektorokból áll, amelyekre

$$Ax = x_1 a_1 + \ldots + x_n a_n = 0\,,$$

ahol $x_1, \ldots, x_n$ az x vektor koordinátáit jelentik.

27.5 Példa. Tekintsük például a következő A mátrixot:

$$A = \begin{bmatrix} 1 & 3 & 1 \\ -2 & 1 & 5 \\ 2 & 2 & -2 \end{bmatrix}$$

és oszlopait jelölje rendre a_1, a_2 és a_3. Elemi bázistranszformációval láthatjuk, hogy az oszlopok nem függetlenek, hiszen $a_3 = -2a_1 + a_2$. Ezért A képtere, azaz az oszlopvektorok által generált altér:

$$\operatorname{im} A = \operatorname{lin}\{a_1, a_2\}\,.$$

Másrészt az előző egyenlőséget rendezve azt kapjuk, hogy

$$-2a_1 + a_2 - a_3 = 0\,.$$

Ez azt mutatja, hogy

$$\ker A = \operatorname{lin}\left\{\begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}\right\} = \left\{t \cdot \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix} \in \mathbb{R}^3 : t \in \mathbb{R}\right\},$$

azaz mindezen vektorokat A-val szorozva a nulla vektort kapjuk.

27.6 Definíció. Az A $m \times n$-es mátrix *rangján* a képtér dimenzióját értjük:

$$\operatorname{rank} A = \dim \operatorname{im} A$$

azaz az A lineárisan független oszlopainak maximális számát. Az A *szabadságfoka*

$$\deg A = \dim \ker A\,.$$

Például a 27.5 Példában vizsgált A mátrix esetében $\operatorname{rank} A = 2$ és $\deg A = 1$.

27.7 Tétel. *Bármely A $\mathbb{R}^n$-en értelmezett lineáris leképezésre* $\operatorname{rank} A + \deg A = n$

Bizonyítás. Legyenek $a_1, \ldots, a_k$ a $\ker A$ altér, illetve $Ab_1, \ldots, Ab_m$ az $\operatorname{im} A$ altér bázisvektorai. Megmutatjuk, hogy az

$$a_1, \ldots, a_k, b_1, \ldots, b_m$$

vektorok együttesen az $\mathbb{R}^n$ tér bázisát adják. Egyrészt lineárisan függetlenek, mert az

$$\alpha_1 a_1 + \ldots + \alpha_k a_k + \beta_1 b_1 + \ldots + \beta_m b_m = 0$$

egyenlőséget A-val szorozva a

$$\beta_1 A b_1 + \ldots + \beta_m A b_m = 0$$

egyenlőséghez jutunk, és innen $\beta_1 = \ldots = \beta_m = 0$ adódik. Ebből következik, hogy $\alpha_1 = \ldots = \alpha_k = 0$.

Másrészt generátorrendszert is alkotnak, ugyanis, ha $x \in \mathbb{R}^n$ tetszőleges, akkor $Ax \in \operatorname{im} A$, és így kifejezhető az $Ab_1, \ldots, Ab_m$ vektorokkal:

$$Ax = \beta_1 A b_1 + \ldots + \beta_m A b_m$$

Ekkor $x - (\beta_1 b_1 + \ldots + \beta_m b_m) \in \ker A$, ezért kifejezhető az $a_1, \ldots, a_k$ vektorok lineáris kombinációjával, azaz

$$x - (\beta_1 b_1 + \ldots + \beta_m b_m) = \alpha_1 a_1 + \ldots + \alpha_k a_k \,.$$

Következésképpen $k + m = n$. $\square$

27.4. Mátrixok szorzása

Tekintsük az alábbiakban az $A : \mathbb{R}^n \to \mathbb{R}^m$ és $B : \mathbb{R}^m \to \mathbb{R}^k$ lineáris leképezéseket. Ekkor képezhetjük a $B \circ A : \mathbb{R}^n \to \mathbb{R}^k$ kompozíció leképezést, amelyet lineáris leképezések esetén szorzatként jelölünk:

$$BA = B \circ A \,.$$

Könnyű ellenőrizni, hogy $BA : \mathbb{R}^n \to \mathbb{R}^k$ is lineáris leképezés, tehát a standard bázisban a mátrixa $k \times n$ méretű. Vajon hogyan állítható elő ez a mátrix?

Valamelyik j indexre tekintsük az Ae_j vektor képét a B leképezésnél, azaz a $B(Ae_j)$ vektort. Ennek a vektornak az i-ik koordinátája:

$$b_{i1} a_{1j} + \ldots + b_{im} a_{mj} \,,$$

ami a BA szorzatmátrix i-ik sorában és j-ik oszlopában álló elem lesz. Tehát a két mátrixot úgy szorozzuk össze, hogy B sorait a fenti szabály szerint megszorozzuk A oszlopaival.

FIGYELEM! A sorrend fontos, a BA szorzat nem egyezik meg AB-vel, sőt előfordulhat, hogy a fordított sorrend nincs is értelmezve!

27.8 Példa. A fenti szabály szerint ellenőrizzük a következő mátrixszorzást:

$$\begin{bmatrix} 2 & 3 & 1 & 0 \\ -1 & 0 & 2 & 1 \\ 1 & 1 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} -1 & 2 & -4 \\ 1 & -1 & 1 \\ 0 & 3 & -2 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 4 & -7 \\ 2 & 5 & -1 \\ 1 & -1 & -1 \end{bmatrix} .$$

azaz a szorzatmátrix 3×3 méretű.

Mátrixok szorzását leképezések kompozíciójaként tekintve azonnal adódik a következő asszociativitási tulajdonság:

$$C(BA) = (CB)A \tag{27.2}$$

mindazon esetekben, amikor a szorzás elvégezhető.

27.9 Példa. Tekintsük az α paraméterrel megadott

$$A = \begin{bmatrix} 1 & 2 & 5 \\ 2 & -1 & 0 \\ -3 & 1 & \alpha \end{bmatrix}$$

mátrixot, és határozzuk meg a rangját és szabadságfokát. Elemi bázistranszformációval

$$\begin{array}{|rrr|rr|r|} \boxed{1} & 2 & 5 & 2 & 5 & 1 \\ 2 & -1 & 0 & \boxed{\text{-5}} & -10 & 2 \\ -3 & 1 & \alpha & 7 & \alpha+15 & \alpha+1 \end{array}$$

Innen azt a következtetést vonhatjuk le, hogy

$$\operatorname{rank} A = \begin{cases} 3 & \text{ha } \alpha \neq -1 \\ 2 & \text{ha } \alpha = -1 \end{cases}$$

és ennek megfelelően, a 27.7 Tétel értelmében

$$\deg A = \begin{cases} 0 & \text{ha } \alpha \neq -1 \\ 1 & \text{ha } \alpha = -1\,. \end{cases}$$

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 II/2 és II/3 szakaszai kidolgozott példáinak feldolgozása.
2. Házi feladatok: a II/2 szakasz 2.1.2, 2.1.3, továbbá a II/3 szakasz 3.4.4, 3.4.6, 3.4.7, 3.4.8, 3.5.2, 3.6.2, 3.7.6, 3.7.7 feladatai.
3. Tankönyv-1 12.6, 12.7, 12.8 és 14.2 szakaszai.

28. fejezet

Lineáris egyenletrendszerek

28.1. Homogén lineáris egyenletrendszerek

Homogén lineáris egyenletrendszeren az alábbi rendszert értjük:

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n &= 0 \end{aligned}$$

ahol az a_{ij} együtthatók adott valós számok. Megoldandó az egyenletrendszer az $x_1, \ldots, x_n$ ismeretlenekre.

Ha összeállítjuk az együtthatók A mátrixát és az ismeretlenek x vektorát a következő módon:

$$A = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ & \vdots & \\ a_{m1} & \ldots & a_{mn} \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

akkor a fenti homogén rendszer a következő egyszerű alakban írható fel:

$$Ax = 0 \tag{28.1}$$

amely megoldásainak halmaza éppen a $\ker A$ altér. Az $x = 0$ vektor mindig megoldás, ha nem ez az egyetlen, akkor végtelen sok megoldás van.

28.1 Példa. Állítsuk elő az alábbi homogén lineáris egyenletrendszer általános megoldását (azaz az összes megoldást).

$$\begin{aligned} x_1 + 3x_2 - 3x_3 &= 0 \\ 2x_1 + x_2 + 4x_3 &= 0 \\ x_1 + 2x_2 - x_3 &= 0 \end{aligned}$$

Ekkor az egyenletrendszer együttható-mátrixa

$$A = \begin{bmatrix} 1 & 3 & -3 \\ 2 & 1 & 4 \\ 1 & 2 & -1 \end{bmatrix}$$

és az egyenletrendszer az $Ax = 0$ alakban írható fel. Alkalmazzuk az A mátrixra az elemi bázistranszformációt:

$$\left|\begin{array}{rrr} \boxed{1} & 3 & -3 \\ 2 & 1 & 4 \\ 1 & 2 & -1 \end{array}\right| \begin{array}{rr} 3 & -3 \\ \boxed{-5} & 10 \\ -1 & 2 \end{array} \left| \begin{array}{r} 3 \\ -2 \\ 0 \end{array} \right|$$

Ez azt jelenti, hogy A oszlopai lineárisan összefüggők, az oszlopokat az a_1, a_2, a_3 vektorokkal jelölve

$$a_3 = 3a_1 - 2a_2\,, \quad \text{azaz} \quad 3a_1 - 2a_2 - a_3 = 0\,.$$

Ezért $x_1 = 3$, $x_2 = -2$ és $x_3 = -1$ nyilvánvalóan megoldást szolgáltatnak, az általános megoldás pedig az

$$x = t \cdot \begin{bmatrix} 3 \\ -2 \\ -1 \end{bmatrix}, \quad t \in \mathbb{R}$$

alakban adható meg. Világos, hogy ebben az esetben

$$\deg A = 1 \quad \text{és} \quad \operatorname{rank} A = 2\,.$$

28.2. Inhomogén lineáris egyenletrendszerek

Tekintsük az A $m \times n$-es mátrixot, és legyen $b \in \mathbb{R}^m$ adott nem nulla vektor. Ekkor az

$$Ax = b \tag{28.2}$$

rendszert *inhomogén lineáris egyenletrendszernek* nevezzük.

28.2 Állítás. *A (28.2) rendszer akkor és csak akkor megoldható, ha $b \in \operatorname{im} A$, azaz a b vektor előállítható az A oszlopainak lineáris kombinációjaként.*

A rendszer megoldása csak akkor egyértelmű, ha az A oszlopai lineárisan függetlenek, azaz a szabadságfoka nulla.

Bizonyítás. Csak az egyértelműséget kell igazolnunk. Indirekt módon, ha lenne két különböző megoldás, x és y, akkor

$$A(x - y) = Ax - Ay = b - b = 0\,,$$

ekkor azonban A oszlopai lineárosan összefüggők. □

Tegyük fel, hogy ismerjük az inhomogén rendszer valamely $\bar{x}$ partikuláris megoldását. Ekkor a rendszer bármely megoldását előállíthatjuk a homogén rendszer megoldásainak ismeretében. Ezt fogalmazza meg a következő tételünk.

28.3 Tétel. *Legyen $\bar{x}$ az inhomogén rendszer valamely megoldása. Ekkor az inhomogén rendszer bármely megoldása előállítható az*

$$x = \bar{x} + x_0$$

alakban, ahol x_0 a homogén rendszer valamely megoldása. Fordítva, ha x_0 a homogén rendszer tetszőleges megoldása, akkor $\bar{x} + x_0$ kielégíti az inhomogén rendszert.

Bizonyítás. Valóban, ha x az inhomogén rendszer valamely megoldása, akkor tekintsük az $x_0 = x - \bar{x}$ vektort. Erre a vektorra

$$Ax_0 = A(x - \bar{x}) = Ax - A\bar{x} = b - b = 0\,,$$

azaz x_0 kielégíti a homogén rendszert, és $x = \bar{x} + x_0$.

Fordítva, ha x_0 a homogén rendszer egy tetszőleges megoldása, akkor az $x = \bar{x} + x_0$ vektorra

$$Ax = A(\bar{x} + x_0) = A\bar{x} + Ax_0 = b + 0 = b\,,$$

azaz x az inhomogén rendszer megoldását adja. □

28.4 Példa. Adjuk meg az alábbi inhomogén rendszer általános megoldását:

$$\begin{aligned} x_1 - x_2 + 3x_3 + 3x_4 &= 1 \\ x_1 - 2x_2 + x_3 - x_4 &= 4 \\ 2x_1 + x_2 - x_3 + 5x_4 &= 6\,. \end{aligned}$$

Ekkor a korábbi jelöléseinket használva

$$A = \begin{bmatrix} 1 & -1 & 3 & 3 \\ 1 & -2 & 1 & -1 \\ 2 & 1 & -1 & 5 \end{bmatrix} \quad \text{és} \quad b = \begin{bmatrix} 1 \\ 4 \\ 6 \end{bmatrix}.$$

Mivel az A mátrixnak 3 sora és 4 oszlopa van, ezért a szabadságfoka legalább 1, azaz a megoldás biztosan nem lesz egyértelmű. Elemi bázistranszformációval

$$\begin{array}{|cccc|c||ccc|c||cc|c||c|c|} \boxed{1} & -1 & 3 & 3 & 1 & -1 & 3 & 3 & 1 & 5 & 7 & -2 & 2 & 3 \\ 1 & -2 & 1 & -1 & 4 & \boxed{-1} & -2 & -4 & 3 & 2 & 4 & -3 & 2 & -1 \\ 2 & 1 & -1 & 5 & 6 & 3 & -7 & -1 & 4 & \boxed{-13} & -13 & 13 & 1 & -1 \end{array}$$

Tehát az A mátrix rangja 3, és a szabadságfoka 1. Ha az A mátrix oszlopaira rendre az a_1, a_2, a_3, a_4 jelölést használjuk, akkor ez a számolás azt mutatja (lásd az utolsó két oszlopot), hogy

$$a_4 = 2a_1 + 2a_2 + a_3 \quad \text{továbbá} \quad b = 3a_1 - a_2 - a_3 \,.$$

Igy megadhatjuk az inhomogén rendszer és a homogén rendszer egy-egy megoldását:

$$\bar{x} = \begin{bmatrix} 3 \\ -1 \\ -1 \\ 0 \end{bmatrix} \quad \text{és} \quad x_0 = \begin{bmatrix} 2 \\ 2 \\ 1 \\ -1 \end{bmatrix} .$$

Tehát a fenti tételünk értelmében az inhomogén rendszer általános megoldása:

$$x = \bar{x} + t \cdot x_0 = \begin{bmatrix} 3 \\ -1 \\ -1 \\ 0 \end{bmatrix} + t \cdot \begin{bmatrix} 2 \\ 2 \\ 1 \\ -1 \end{bmatrix} \qquad t \in \mathbb{R} \,.$$

28.5 Példa. Tekintsük az α és β valós paraméterekkel megadott $Ax = b$ inhomogén rendszert, ahol

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 2 & 1 & 3 \\ -1 & 1 & \alpha \end{bmatrix} \quad \text{és} \quad b = \begin{bmatrix} 1 \\ 2 \\ \beta \end{bmatrix} .$$

Vizsgáljuk meg, hogy a paraméterek milyen értékei mellett oldható meg az egyenletrendszer. Végezzük el az elemi bázistranszformációt:

$$\left|\begin{array}{rrr|r||rr|r||r|r} \boxed{1} & -1 & 0 & 1 & -1 & 0 & 1 & 1 & 1 \\ 2 & 1 & 3 & 2 & \boxed{3} & 3 & 0 & 1 & 0 \\ -1 & 1 & \alpha & \beta & 0 & \alpha & \beta+1 & \alpha & \beta+1 \end{array}\right|$$

Ennek alapján az alábbi következtetést vonhatjuk le:

- pontosan egy megoldás van, ha $\alpha \neq 0$ és β tetszőleges
- végtelen sok megoldás van, ha $\alpha = 0$ és $\beta = -1$
- nincs megoldás, ha $\alpha = 0$ és $\beta \neq -1$

Az A mátrix rangjára és szabadságfokára az adódik, hogy

$$\operatorname{rank} A = \begin{cases} 3 & \text{ha } \alpha \neq 0 \\ 2 & \text{ha } \alpha = 0 \end{cases}$$

illetve

$$\deg A = \begin{cases} 0 & \text{ha } \alpha \neq 0 \\ 1 & \text{ha } \alpha = 0 \,. \end{cases}$$

28.3. Inverz mátrix

Jelölje E azt az $n \times n$ méretű mátrixot, amelynek fődiagonálisában minden elem 1, azon kívül minden elem 0. Ezt a mátrixot $n \times n$-es *egységmátrixnak* nevezzük. Világos, hogy E mint lineáris leképezés, az identikus leképezés, hiszen minden x vektorra $Ex = x$.

28.6 Definíció. Tekintsünk egy $n \times n$-es A négyzetes mátrixot. Azt mondjuk, hogy A invertálható, ha van olyan $n \times n$-es A^{-1} mátrix, amelyre

$$A \cdot A^{-1} = E\,.$$

Ezt az A^{-1} mátrixot az A *inverzének* nevezzük.

Könnyen látható, hogy A^{-1} az inverz leképezés, azaz $AA^{-1}x = x$ minden $x \in \mathbb{R}^n$ esetén. Az is világos, hogy ilyenkor $A^{-1}A = E$ is érvényes.

Az inverz létezésének szükséges és elégséges feltétele, hogy a leképezés kölcsönösen egyértelmű legyen. Ezen az észrevételen alapul a következő tétel.

28.7 Tétel. *Egy A $n \times n$-es mátrixra a következő állítások ekvivalensek:*

1. *A invertálható*
2. *az A oszlopai lineárisan függetlenek*
3. $\ker A = \{0\}$
4. $\operatorname{im} A = \mathbb{R}^n$
5. $\operatorname{rank} A = n$
6. $\deg A = 0\,.$

A páronkénti ekvivalenciák ellenőrzése helyett az ilyen állításokat a matematikában úgy szokták bizonyítani, hogy sorrendben igazoljuk a következő implikációkat:

$$1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 5 \Rightarrow 6 \Rightarrow 1$$

Gondoljuk végig, hogy ezek az implikációk valóban könnyen ellenőrizhetők, és ezekből valóban következik a fenti állítások ekvivalenciája.

Invertálható mátrixok esetében az inhomogén lineáris egyenletrendszerek megoldását nagyon egyszerűen lehet megadni.

28.8 Tétel. *Legyen A $n \times n$-es invertálható mátrix. Akkor az*

$$Ax = b$$

inhomogén lineáris egyenletrendszer egyértelműen megoldható, és a megoldás az

$$x = A^{-1}b$$

alakban adható meg.

Érdemes megjegyezni, hogy ennek ellenére az egyenletrendszer megoldása az előző szakaszban tanult elemi bázistranszformációval általában kevesebb számolással jár, mint az inverz meghatározása. Az inverz kiszámítása akkor előnyös, ha az inhomogén egyenletrendszert többször, más és más b vektorral a rendszer jobb oldalán kell megoldani.

28.4. Az inverz mátrix meghatározása

Tekintsünk egy A $n \times n$-es invertálható mátrixot. Az inverz mátrix meghatározását úgy képzeljük el, hogy keressük azt az X $n \times n$-es mátrixot, amelyre $AX = E$. Ha az ismeretlen X mátrix oszlopait rendre $x_1, \ldots, x_n$ jelöli, akkor ez az n-számú

$$Ax_k = e_k$$

$(k = 1, \ldots, n)$ inhomogén lineáris egyenletrendszer megoldását jelenti. Ezt az eljárást az alábbi példán szemléltetjük (ahol egyszerre oldjuk meg az egyenletrendszereket).

28.9 Példa. Invertálható-e az alábbi A mátrix? Ha igen, állítsuk elő az inverzét:

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

Az $Ax = e_1$, $Ax = e_2$, $Ax = e_3$ egyenletrendszereket egyszerre megoldva:

$$\left|\begin{array}{ccc|ccc||cc|ccc||c|ccc||ccc|} \boxed{1} & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & -1 & 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & \boxed{1} & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 0 & 0 & 1 & -1 & 0 & -1 & 0 & 1 & \boxed{1} & -1 & 1 & 1 & -1 & 1 & 1\,. \end{array}\right.$$

Azt kaptuk, hogy

$$A^{-1} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ -1 & 1 & 1 \end{bmatrix},$$

amiről közvetlen szorzással ellenőrizhetjük, hogy $AA^{-1} = E$.

28.10 Példa. Legyenek A és B $n \times n$ méretű invertálható mátrixok. Az A és B leképezéseket tekintve világos, hogy az AB leképezés is invertálható.

FIGYELEM! Gondoljuk át, hogy ilyenkor az AB kompozíció leképezés is kölcsönösen egyértelmű!

Vajon hogyan határozható meg az $(AB)^{-1}$ leképezés mátrixa? Megmutatjuk, hogy

$$(AB)^{-1} = B^{-1}A^{-1}\,.$$

Valóban, a jobb oldalon álló mátrix az AB szorzat inverze, hiszen

$$(AB)B^{-1}A^{-1} = A(BB^{-1})A^{-1} = AEA^{-1} = E$$

a (27.2) asszociativitási tulajdonság miatt.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 II/4 és II/5 szakaszai kidolgozott példáinak feldolgozása.
2. Házi feladatok: a II/4 szakasz 4.1.5, 4.1.6, 4.2.2, 4.2.3, 4.3.4 továbbá a II/5 szakasz 5.1.3, 5.2.2, 5.3.6, 5.3.7, 5.4.2 feladatai.
3. Tankönyv-1 12.1, 13.7 és 14.2 szakaszai.

29. fejezet

Sajátérték, sajátvektor

29.1. Sajátérték, sajátvektor

29.1 Definíció. Tekintsük az $A : \mathbb{R}^n \to \mathbb{R}^n$ lineáris transzformációt, azaz $n \times n$-es mátrixot. Azt mondjuk, hogy a λ szám az A *sajátértéke*, ha van olyan $v \neq 0$ vektor, amelyre

$$Av = \lambda v\,.$$

Ilyenkor a v vektort a λ sajátértékhez tartozó *sajátvektornak* nevezzük.

29.2 Példa. Tekintsük például az $\mathbb{R}^3$ tér alábbi A transzformációját és a v vektort, ahol

$$A = \begin{bmatrix} 2 & 1 & -1 \\ 0 & 1 & 1 \\ 2 & 0 & -2 \end{bmatrix} \quad \text{és} \quad v = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}.$$

ekkor könnyen ellenőrizhető, hogy

$$Av = 2v$$

azaz $\lambda = 2$ az A sajátértéke, és v egy hozzá tartozó sajátvektor. Az is látható, hogy az

$$u = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

vektorra

$$Au = 0$$

tehát $\lambda = 0$ is sajátérték, és u egy hozzá tartozó sajátvektor. Ellenőrizzük, hogy $\lambda = -1$ is az A sajátértéke, és keressünk hozzá tartozó sajátvektort is!

FIGYELEM! Valamely λ sajátértékhez tartozó sajátvektor soha nem egyértelmű. Gondoljuk meg, hogy ha v sajátvektor, akkor annak bármely $\alpha \neq 0$ skalárszorosa is az. Valóban,

$$A(\alpha v) = \alpha A v = \alpha \cdot \lambda v = \lambda(\alpha v)\,.$$

29.3 Példa. Vizsgáljuk most meg a 27.2 Példában ismertetett síkbeli transzformációkat.

1. Ha A az α-szoros nyújtás, akkor α az A egyetlen sajátértéke, és a sík minden vektora sajátvektor.
2. Ha A a vízszintes tengelyre vonatkozó tükrözés, akkor $\lambda_1 = 1$ sajátérték, és az e_1 vektor egy megfelelő sajátvektor, továbbá $\lambda_2 = -1$ is sajátérték, és e_2 egy hozzá tartozó sajátvektor.
3. Ha A a $45°$-os szögfelezőre történő merőleges vetítés, akkor a sajátértékek és a megfelelő sajátvektorok a következők:

$$\lambda_1 = 1 \quad \text{és} \quad v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{illetve} \quad \lambda_2 = 0 \quad \text{és} \quad v_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

4. Ha A az origó körüli $0 \leq \varphi < 2\pi$ szöggel történő elforgatás, akkor $\varphi = 0$ esetén $\lambda = 1$, míg $\varphi = \pi$ esetén $\lambda = -1$ az egyetlen sajátérték. Mindkét esetben a sík minden vektora sajátvektor.

   Más φ szögekre az A elforgatásnak nincs valós sajátértéke.

29.2. Sajátaltér

29.4 Definíció. Tekintsük az $\mathbb{R}^n$ tér valamely A lineáris transzformációját, és tegyük fel, hogy λ az A sajátértéke. Ekkor mindazon v vektorok, amelyekre $Av = \lambda v$ alteret alkotnak, amelyet az A transzformáció λ sajátértékéhez tartozó *sajátaltérnek* nevezünk. Jelölésben

$$S_A(\lambda) = \{v \in \mathbb{R}^n : Av = \lambda v\}$$

29.5 Példa. Vizsgáljuk meg például a következő mátrixot

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Egyszerű számolással ellenőrizhetjük, hogy $\lambda = 2$ az A sajátértéke, és az e_2, e_3 vektorok egyaránt sajátvektorok, és lineárisan függetlenek.

Az is világos, hogy további független sajátvektorokat ehhez a sajátértékhez nem találhatunk, ezért a sajátaltérre

$$\dim S_A(2) = 2$$

továbbá e_2 és e_3 ennek az altérnek a bázisát szolgáltatják.

29.3. Sajátvektorok meghatározása

Tegyük fel ebben a szakaszban, hogy A olyan $n \times n$-es mátrix, amelynek a λ szám sajátértéke. Vajon hogyan határozható meg az összes megfelelő sajátvektor?

Jelölje E az $n \times n$-es egységmátrixot, és tegyük fel, hogy v valamelyik, a λ sajátértékhez tartozó sajátvektor. Ekkor

$$Av = \lambda v = \lambda E v$$

illetve az egyenlőséget átrendezve

$$(A - \lambda E)v = 0$$

tehát a v vektorra egy homogén lineáris egyenletrendszer adódik. Fogalmazzuk meg tételben ezt az észrevételünket.

29.6 Tétel. *Legyen A egy $n \times n$-es mátrix, és λ az A sajátértéke. Ekkor*

$$S_A(\lambda) = \ker (A - \lambda E)$$

azaz a sajátaltér egy homogén lineáris egyenletrendszer általános megoldása.

Külön érdekes a $\lambda = 0$ eset. Ha ugyanis ez sajátérték, akkor az $Av = 0$ homogén lineáris egyenletrendszernek van nem nulla megoldása, azaz A rangja nem lehet n. Ezt egy külön tételben is megfogalmazzuk.

29.7 Tétel. *A akkor és csak akkor invertálható, ha $\lambda = 0$ nem sajátérték.*

Általában jóval nehezebb feladat azonban egy A lineáris transzformáció sajátértékeinek megtalálása.

29.4. Lineárisan független sajátvektorok

Tekintsük az A transzformációt az $\mathbb{R}^n$ téren, és tegyük fel, hogy a $\lambda_1, \dots, \lambda_k$ különböző számok az A sajátértékei, továbbá $v_1, \dots, v_k$ az azonos indexű sajátértékekhez tartozó nem nulla sajátvektorok. Megmutatjuk, hogy ezek a vektorok lineárisan függetlenek.

29.8 Tétel. *Különböző sajátértékekhez tartozó nem nulla sajátvektorok lineárisan függetlenek.*

Bizonyítás. A tételt teljes indukcióval igazoljuk. Az állítás $k = 1$-re nyilvánvaló. Tegyük fel, hogy az állítás $k-1$-ig igaz, és indirekt módon tegyük fel, hogy $v_1, \dots, v_k$ összefüggők. Ezt jelenti, hogy van olyan lineáris kombinációjuk, amelyre

$$\alpha_1 v_1 + \dots + \alpha_k v_k = 0 \tag{29.1}$$

de nem minden együttható nulla, pl. az egyszerűség kedvéért $\alpha_1 \neq 0$. Alkalmazzuk mindkét oldalra az A transzformációt, akkor

$$\alpha_1 A v_1 + \dots + \alpha_k A v_k = \alpha_1 \lambda_1 v_1 + \dots + \alpha_k \lambda_k v_k = 0$$

Ha ez utóbbi egyenlőségből kivonjuk a (29.1) egyenlőség λ_k-szorosát, akkor utolsó tag kiesik, és azt kapjuk, hogy

$$\alpha_1(\lambda_1 - \lambda_k) v_1 + \dots + \alpha_{k-1}(\lambda_{k-1} - \lambda_k) v_{k-1} = 0\,.$$

Mivel a sajátértékek mind különbözőek, azért $\alpha_1 \neq 0$ miatt itt a v_1 együtthatója nem nulla, ami azt jelentené, hogy a $v_1, \dots, v_{k-1}$ vektorok lineárisan összefüggők. Ez azonban ellentmond az indukciós feltevésnek. □

29.9 Példa. Tekintsük a 29.2 Példában vizsgált A mátrixot. Egyszerű számolással láthatjuk, hogy $\lambda_1 = 2$, $\lambda_2 = 0$ és $\lambda_3 = -1$ egyaránt az A sajátértékei, és hozzájuk tartozó sajátvektorok rendre

$$v_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \quad \text{és} \quad v_3 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$$

Elemi bázistranszformációval könnyen ellenőrizhetjük, hogy a v_1, v_2, v_3 vektorok valóban lineárisan függetlenek.

29.5. Transzformációk diagonális alakja

Egy mátrixot diagonális alakúnak nevezünk, ha a fődiagonálisan kívül csak nulla elemei vannak. Az ilyen mátrixok kezelése (pl. szorzás, hatványozás) igen egyszerű, ezért hasznosak lehetnek a lineáris algebrában, és az alkalmazásaiban is. Érdekes kérdés tehát, hogy egy adott mátrixhoz találhatunk-e olyan bázist, amelyben az diagonális alakú lesz. Látni fogjuk, hogy ez éppen a sajátvektorokból álló bázis (amennyiben ez létezik).

Tegyük fel, hogy az $n \times n$-es A mátrix sajátértékei $\lambda_1, \dots, \lambda_n$ (nem feltétlenül mind különbözőek), és a hozzájuk tartozó sajátvektorok $v_1, \dots, v_n$, amelyekről tegyük fel, hogy lineárisan függetlenek. Mivel a számuk n, ezért ezek a vektorok ilyenkor az $\mathbb{R}^n$ bázisát alkotják.

Irjuk most fel az A transzformáció mátrixát a sajátvektorok bázisában. Jelölje $\hat{A}$ a mátrixot az új bázisban. Mivel a sajátvektorokra

$$Av_k = \lambda v_k \quad k = 1, \dots, n$$

azért $\hat{A}$ a következő alakot ölti:

$$\hat{A} = \begin{bmatrix} \lambda_1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & \lambda_n \end{bmatrix}$$

Vizsgáljuk meg, hogy milyen összefüggés van az A transzformáció $e_1, \dots, e_n$ bázisban, illetve a $v_1, \dots, v_n$ bázisban felírt mátrixai között. Jelentse S azt az $n \times n$ méretű mátrixot, amelynek oszlopai rendre a $v_1, \dots, v_n$ sajátvektorok, azaz

$$Se_k = v_k$$

Világos, hogy S invertálható, hiszen a feltételünk szerint az oszlopai lineárisan függetlenek. Továbbá

$$Av_k = S\hat{A}e_k$$

minden $k = 1, \dots, n$ esetén. Mindkét oldalt az S^{-1} mátrixszal szorozva

$$S^{-1}ASe_k = \hat{A}e_k$$

minden k index mellett, ezért

$$\hat{A} = S^{-1}AS$$

Ezt az eredményünket a következő tételben foglaljuk össze.

29.10 Tétel. *Tegyük fel, hogy A olyan $n \times n$-es mátrix, amelynek sajátértékei $\lambda_1, \dots, \lambda_n$ és a hozzá tartozó $v_1, \dots, v_n$ sajátvektorok a tér bázisát alkotják. Akkor a sajátvektorok bázisában A olyan $\hat{A}$ diagonális alakú mátrix, amelynek fődiagonálisában a sajátértékek állnak, továbbá*

$$\hat{A} = S^{-1}AS$$

ahol az S mátrix oszlopai a $v_1, \dots, v_n$ sajátvektorok.

29.11 Példa. Tekintsük újra a 29.2 Példában megadott A transzformációt. Ekkor a $\lambda_1 = 2$, $\lambda_2 = 0$ és $\lambda_3 = -1$ sajátértékekhez tartozó

$$v_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \quad \text{és} \quad v_3 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$$

sajátvektorok lineárisan függetlenek, ezért az $\mathbb{R}^3$ tér bázisát alkotják. Tehát a sajátvektorok által alkotott bázisban az A transzformáció mátrixa diagonális alakú lesz. A tételünk szerint az

$$S = \begin{bmatrix} 2 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & -1 & 2 \end{bmatrix} \quad \text{és} \quad \hat{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

jelölésekkel

$$\hat{A} = S^{-1} A S$$

Ellenőrizzük ezt az egyenlőséget közvetlenül is, az S^{-1} inverz mátrix meghatározásával, és a kijelölt műveletek elvégzésével!

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 II/6 szakasza kidolgozott példáinak feldolgozása.
2. Házi feladatok: a II/6 szakasz 6.4.1, 6.4.2, 6.4.3, 6.4.4, 6.4.5, 6.4.8, 6.4.9 és 6.4.10 feladatai.
3. Tankönyv-1 14.4 és 14.5 szakaszai.

30. fejezet

Determináns

30.1. Permutációk

Tekintsük az első n természetes szám $H = \{1, \ldots, n\}$ halmazát.

30.1 Definíció. Egy $p : H \to H$ kölcsönösen egyértelmű leképezést a H halmaz *permutációjának* nevezünk.

Szemléletesen egy permutáció a H halmaz elemeinek egy sorrendbe rakását jelenti. A H halmaz összes permutációinak száma $n!$ (n faktoriális).

30.2 Definíció. Tekintsük a H halmaz valamely p permutációját, amelyre

$$p(1) = i_1, \quad \ldots \quad p(n) = i_n$$

azaz az $\{i_1, \ldots, i_n\}$ sorrendet. Azt mondjuk, hogy az i_j és i_k elemek *inverziót* alkotnak, ha $j < k$ és $i_j > i_k$.

30.3 Példa. Például $n = 5$ esetén az

$$\{1, 3, 2, 4, 5\}$$

permutáció egyetlen inverziót tartalmaz, míg a

$$\{2, 3, 1, 5, 4\}$$

permutációban három inverzió található.

30.4 Definíció. A H halmaz p permutációját *páratlannak* nevezzük, ha az inverziók száma páratlan, ellenkező esetben a permutációt *párosnak* nevezzük.

30.2. A determináns fogalma

Tekintsük a következő $n \times n$ méretű A mátrixot

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

30.5 Definíció. Az A mátrix *determinánsán* az alábbi kifejezést értjük:

$$\det A = \sum (-1)^{\mathcal{P}} a_{1i_1} a_{2i_2} \cdots a_{ni_n}$$

ahol az összegzést az $\{1, \dots, n\}$ halmaz összes $\{i_1, \dots, i_n\}$ permutációjára végezzük, tehát a szumma $n!$ tagból áll. A (-1) kitevője aszerint páros vagy páratlan, hogy az adott $\{i_1, \dots, i_n\}$ permutáció páros vagy páratlan.

Szokásos még az alábbi jelölésmód is:

$$\det A = |A| = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}$$

Látható a definícióból, hogy a szumma mögötti szorzatokat úgy állítjuk össze, hogy a mátrix minden sorából és minden oszlopából pontosan egy tényezőt tartalmazzanak.

30.6 Példa. Ellenőrizzük közvetlenül a definíció alapján, hogy az

$$A = \begin{bmatrix} 2 & -1 \\ 3 & 4 \end{bmatrix} \quad \text{és} \quad B = \begin{bmatrix} 2 & 1 & -1 \\ 0 & 1 & 3 \\ 1 & 0 & -2 \end{bmatrix}$$

mátrixokra $\det A = 11$ és $\det B = 0$.

30.3. A determináns tulajdonságai

30.7 Állítás. $\det A = \det A^T$.

Valóban, ha a mátrix oszlopait és sorait felcseréljük, akkor az inverziók paritása nem változik.

30.8 Állítás. *Ha egy mátrix valamely sora csupa nullából áll, akkor a determinánsa zérus.*

Valóban, ekkor a szumma mögött mindegyik tagban szerepel zérus tényező.

30.9 Állítás. *Ha egy mátrix két sorát felcseréljük, akkor a determinánsa előjelet vált.*

Valóban, ilyenkor mindegyik tagban az inverziók számának paritása megváltozik.

30.10 Állítás. *Ha egy mátrixban két sor azonos, akkor a determinánsa zérus.*

Valóban, a két azonos sort felcserélve egyrészt a determináns előjelet vált, másrészt természetesen nem változik, azaz $\det A = -\det A$, és így $\det A = 0$.

30.11 Állítás. *Ha egy mátrix valamely sorát λ-val szorozzuk, akkor a determinánsa is λ-val szorzódik.*

Valóban, ilyenkor a szumma mögött minden tag λ-val szorzódik, hiszen minden tag minden sorból pontosan egy tényezőt tartalmaz.

30.12 Állítás. *Ha egy mátrixban valamelyik sor egy másik λ-szorosa, akkor a determinánsa zérus.*

Valóban, ha a λ szorzót az adott sorból kiemeljük, akkor olyan mátrixhoz jutunk, amelyben két azonos sor található.

30.13 Állítás. *Ha egy mátrixban egy sor minden eleme összegként áll elő, például*

$$a_{ij} = b_{ij} + c_{ij} \quad j = 1, \ldots, n$$

akkor a determinánsa azon két determináns összege, amelyeknél az i-ik sor rendre a b_{ij} illetve a c_{ij} elemekből áll.

Valóban, a szumma mögött minden szorzat pontosan ilyen összegre bomlik.

30.14 Állítás. *Ha egy mátrix valamely sora a többi sor lineáris kombinációja, akkor a determinánsa zérus.*

Valóban, bontsuk fel a determinánst összegre a megelező állítás értelmében. Ha most a skalár együtthatókat mindegyik determinánsból kiemeljük, olyan determinánsokhoz jutunk, amelyekben rendre két sor megegyezik, ezért mindegyik zérus.

30.15 Állítás. *Ha egy mátrixban valamely sorhoz hozzáadjuk egy másik sor λ-szorosát, akkor a determinánsa nem változik.*

Valóban, ekkor a determináns két olyan determináns összegére bomlik, ahol a második tag zérus.

30.16 Állítás. *Ha A ás B $n \times n$-es mátrixok, akkor* $\det(AB) = \det A \cdot \det B$.

Ez az állítás az előző állításaink felhasználásával, a mátrixszorzás lépésenkénti elvégzésével ellenőrizhető.

Végezetül állítasaink következményeként az alábbi fontos tételt fogalmazhatjuk meg.

30.17 Tétel. *Az A négyzetes mátrix oszlopai akkor és csak akkor lineárisan összefüggők, ha* $\det A = 0$.

A szükségesség a 30.14 Állítás közvetlen következménye. Az elégségesség abból adódik, hogy ha A oszlopai lineárisan függetlenek, akkor A invertálható, és ezért $AA^{-1} = E$. Tehát

$$\det(AA^{-1}) = \det A \cdot \det A^{-1} = \det E = 1$$

következésképpen $\det A \neq 0$.

30.4. A determináns kiszámítása

A definíció alapján világos, hogy egy 2×2-es mátrix determinánsa

$$\det A = a_{11}a_{22} - a_{12}a_{21}$$

Teljesen hasonlóan egy 3×3-as mátrix determinánsa a következő alakban írható

$$\det A = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Ha ezt az észrevételt induktívan többször alkalmazzuk, akkor a következő eredményhez jutunk.

30.18 Tétel. *Tekintsük az A $n \times n$-es mátrixot, és jelentse A_{1j} azt az $(n-1) \times (n-1)$-es mátrixot, amelyet az A első sorának és j-ik oszlopának elhagyásával* *nyerünk. Ekkor*

$$\det A = \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \det A_{1j}$$

Ezt az eljárást aldeterminánsokra történő felbontásnak nevezzük.

30.19 Példa. Használjuk az aldeterminánsokra történő felbontási eljárást az

$$A = \begin{bmatrix} 3 & 6 & 0 & 2 \\ 0 & 1 & 2 & 4 \\ 4 & 8 & 3 & 5 \\ 1 & 2 & 0 & 0 \end{bmatrix}$$

mátrixra. Ellenőrizzük lépésenként először 3×3-as, majd 2×2-es aldeterminánsokra való felbontással, hogy $\det A = -6$.

30.5. Sajátértékek meghatározása

Tegyük fel a továbbiakban, hogy A egy $n \times n$-es mátrix, amelynek λ sajátértéke, és $v \neq 0$ egy hozzá tartozó sajátvektor, azaz $Av = \lambda v$. Ezt úgy is írhatjuk, hogy

$$Av - \lambda v = (A - \lambda E)v = 0\,.$$

Ez az egyenlőség azt jelenti, hogy az $A - \lambda E$ mátrix oszlopai lineárisan összefüggők, hiszen a homogén lineáris egyenletrendszernek létezik nemtriviális megoldása. Tehát a 30.17 Tétel értelmében a következő eredményt fogalmazhatjuk meg.

30.20 Tétel. *A λ szám akkor és csak akkor az A sajátértéke, ha* $\det(A - \lambda E) = 0$.

Ez a szükséges és elégséges feltétel egy n-edfokú egyenletet ad az A sajátértékeire.

30.21 Példa. Tekintsük a következő mátrixot

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

és keressük meg a sajátértékeit. Állítsuk elő az $A - \lambda E$ mátrix determinánsát:

$$\det(A - \lambda E) = (\lambda + 1)^2 (\lambda - 2)$$

Ennek a harmadfokú polinomnak a gyökei $\lambda_1 = -1$ (kétszeres gyök), illetve $\lambda_2 = 2$, amelyek az A sajátértékei. Gyakorlásképpen keressük meg a megfelelő sajátvektorokat is. Ekkor az

$$(A - \lambda E)v = 0$$

homogén lineáris egyenletrendszer lineáriasan független megoldásaira például $\lambda = -1$ mellett

$$v_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \quad \text{és} \quad v_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

illetve $\lambda = 2$ mellett

$$v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

vektorok adódnak. Tehát a v_1, v_2, v_3 bázisban az A mátrix az

$$\hat{A} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

diagonális alakot ölti.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 II/2 és II/6 szakaszai kidolgozott példáinak feldolgozása.
2. Házi feladatok: a II/2 szakasz 2.3.2, 2.3.3 és 2.3.6, továbbá a II/6 szakasz 6.4.1, 6.4.2, 6.4.3, 6.4.4, 6.4.5, 6.4.8, 6.4.9 és 6.4.10 feladatai.
3. Tankönyv-1 13.1, 13.2, 13.3, 13.4 és 13.5 szakaszai.

31. fejezet

Skaláris szorzat

31.1. Skaláris szorzat

31.1 Definíció. Az x és y $\mathbb{R}^n$-beli vektorok *skaláris szorzatán* az

$$\langle x, y\rangle = x_1y_1 + \ldots + x_ny_n$$

kifejezést értjük, ahol a jobb oldalon a vektorok koordinátái szerepelnek.

Gondoljuk meg, hogy ez az értelmezés $n = 2$ esetében pontosan megegyezik a középiskolában tanult fogalommal.

31.2 Definíció. Az $x \in \mathbb{R}^n$ vektor *normáján* (vagy abszolút értékén) a következőt értjük:

$$\|x\| = (\langle x, x\rangle)^{1/2} = \sqrt{x_1^2 + \ldots + x_n^2}$$

amelyet a vektor hosszának is nevezünk.

Könnyen látható, hogy a Pitagorasz-tétel értelmében ez a megfogalmazásunk összhangban van az eddigi geometriai szemléletünkkel. Az is világos, hogy $\|x\| = 0$ akkor és csak akkor áll, ha $x = 0$.

31.3 Definíció. Az x és y vektorok *távolságán* az $\|x - y\|$ kifejezést értjük.

31.4 Példa. Például az

$$x = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} \text{ és } y = \begin{bmatrix} 4 \\ -3 \\ 0 \end{bmatrix}$$

vektorokra $\langle x, y\rangle = 14$, továbbá $\|x\| = 3$, illetve $\|y\| = 5$. A két vektor távolságára $\|x - y\| = \sqrt{6}$ adódik.

31.2. Vektorok szöge, merőlegesség

31.5 Tétel. Cauchy-Schwarz-egyenlőtlenség *Bármely $x, y \in \mathbb{R}^n$ vektorokra*

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$$

Bizonyítás. Legyen t tetszőleges valós szám, és tekintsük az alábbi másodfokú polinomot:

$$g(t) = \langle x + ty, x + ty\rangle$$

Először is a skaláris szorzat értelmezése alapján

$$g(t) = \|x\|^2 + 2t\langle x, y\rangle + t^2\|y\|^2$$

másrészt ez a kifejezés az $x + ty$ vektor normanégyzete, tehát nem negatív, azaz

$$g(t) \geq 0 \quad \text{minden } t \text{ esetén.}$$

Ha egy másodfokú polinom nem negatív, akkor a diszkriminánsa nem pozitív, azaz

$$4\left(\langle x, y\rangle\right)^2 \leq 4\|x\|^2 \cdot \|y\|^2$$

4-gyel egyszerűsítve, és mindkét oldalból négyzetgyököt vonva éppen azt kapjuk, hogy

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$$

amit igazolnunk kellett. □

31.6 Tétel. (Háromszög-egyenlőtlenség) $\|x + y\| \leq \|x\| + \|y\|$.

Bizonyítás. Valóban, a Cauchy-Schwartz-egyenlőtlenség folytán

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y\rangle = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\| \cdot \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2 \end{aligned}$$

ahonnan négyzetgyökvonással adódik a tétel állítása. □

31.7 Definíció. Az x és y nem nulla vektorok által bezárt szögön azt a $0 \leq \varphi \leq \pi$ szöget értjük, amelyre

$$\cos\varphi = \frac{\langle x, y\rangle}{\|x\| \cdot \|y\|}$$

Azt mondjuk továbbá, hogy x és y egymásra *ortogonálisak* (vagy merőlegesek), jelölésben $x \perp y$, ha

$$\langle x, y\rangle = 0$$

hiszen ilyenkor $\cos\varphi = 0$, azaz $\varphi = \pi/2$. A 0 vektor minden más vektorra merőleges.

FIGYELEM! Gondoljuk meg, hogy a vektorok szögét jól definiáltuk, ugyanis a Cauchy-Schwartz-féle egyenlőtlenség értelmében $-1 \leq \cos\varphi \leq 1$.

31.3. Ortogonális vektorok

31.8 Definíció. Azt mondjuk, hogy az $\mathbb{R}^n$ térben az $a_1, \ldots, a_k$ vektorok *ortogonális rendszert* alkotnak, ha egyikük sem nulla vektor, és a vektorok páronként merőlegesek, azaz

$$\langle a_i, a_j \rangle = 0$$

bármely $i \neq j$ indexekre.

31.9 Tétel. *Minden ortogonális rendszer lineárisan független.*

Bizonyítás. Tekintsük az $a_1, \ldots, a_k$ ortogonális rendszert, és tegyük fel, hogy

$$\alpha_1 a_1 + \ldots + \alpha_k a_k = 0\,.$$

Szorozzuk meg skalárisan mindkét oldalt az a_i vektorral. Ekkor a páronkénti ortogonalitás miatt az i-ik tag kivételével mindegyik szorzat nulla, azaz

$$\alpha_i \|a_i\|^2 = 0$$

adódik. Mivel $a_i \neq 0$, innen azt kapjuk, hogy $\alpha_i = 0$. Ezt a gondolatot mindegyik $i = 1, \ldots, k$ indexre alkalmazhatjuk, ezért $\alpha_1 = \ldots = \alpha_k = 0$. Ez éppen azt jelenti, hogy az $a_1, \ldots, a_k$ vektorok lineárisan függetlenek. □

Ez a tételünk azt mutatja, hogy az $\mathbb{R}^n$ térben maximum n-elemű ortogonális rendszer található. Egyúttal egy n-elemű ortogonális rendszer a tér bázisát alkotja.

31.10 Definíció. Az $\mathbb{R}^n$ tér egy *ortogonális bázisán* egy $a_1, \ldots, a_n$ ortogonális rendszert értünk. Azt mondjuk, hogy ez a bázis *ortonormált*, ha a bázisvektorok egységnyi normájúak, azaz $\|a_i\| = 1$ minden $i = 1, \ldots, n$ esetén.

Analóg módon értelmezzük egy tetszőleges M altér ortonormált bázisát is.

31.11 Példa. Könnyen ellenőrizhető például, hogy az

$$a_1 = \begin{bmatrix} \frac{\sqrt{3}}{2} \\ 0 \\ -\frac{1}{2} \end{bmatrix}, \quad a_2 = \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{\sqrt{3}}{2} \end{bmatrix}, \quad a_3 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$$

vektorok az $\mathbb{R}^3$ tér ortonormált bázisát szolgáltatják.

31.4. Gram-Schmidt-féle eljárás

Ortonormált bázisban felírt vektorokkal nagyon könnyű számolni, ezért természetes kérdésként merül fel, hogy vajon minden altérben létezik-e ortonormált bázis. Erre ad pozitív választ a Gram-Schmidt-módszer, amelynek segítségével elő is tudjuk állítani ezt a bázist.

Tekintsük az $\mathbb{R}^n$ tér egy tetszőleges M alterét, amelynek $a_1, \ldots, a_k$ bázisa. Megmutatjuk, hogy ebből kiindulva megszerkeszthetünk egy ortonormált bázist.

Legyen $b_1 = a_1$. Ezután legyen $b_2 = a_2 + \alpha_1 b_1$, ahol az ismeretlen α_1 együtthatót úgy választjuk meg, hogy b_2 merőleges legyen a b_1 vektorra. Tehát

$$\langle b_2, b_1\rangle = \langle a_2, b_1\rangle + \alpha_1\langle b_1, b_1\rangle = 0$$

Ebből az egyenlőségből azt kapjuk, hogy

$$\alpha_1 = -\frac{\langle b_1, a_2\rangle}{\|b_1\|^2}$$

ahonnan

$$b_2 = a_2 - \frac{\langle b_1, a_2\rangle}{\|b_1\|^2} b_1 \,.$$

Teljesen hasonló gondolatmenet alapján a b_3 vektort a

$$b_3 = a_3 + \beta_1 b_1 + \beta_2 b_2$$

alakban keressük úgy, hogy b_3 ortogonális legyen a b_1 és b_2 vektorok mindegyikére. Ekkor a két feltételből az együtthatókat meghatározva

$$b_3 = a_3 - \frac{\langle b_1, a_3\rangle}{\|b_1\|^2} b_1 - \frac{\langle b_2, a_3\rangle}{\|b_2\|^2} b_2 \,.$$

adódik. Az eljárást folytatva az M altér egy ortogonális bázisához jutunk. Ezt az eredményt a következő tételben fogalmazzuk meg.

31.12 Tétel. (Gram-Schmidt) *Az $\mathbb{R}^n$ tér bármely alterében van ortonormált bázis.*

Bizonyítás. A fenti eljárásban mindegyik b_i bázisvektort osszuk el a $\|b_i\|$ pozitív számmal, akkor egy ortonormált bázishoz jutunk. □

31.5. Az ortogonális komplementer

Tekintsünk egy tetszőleges M alteret az $\mathbb{R}^n$ vektortérben.

31.13 Definíció. Az M altér összes vektorára merőleges vektorok halmazát, jelölésben

$$M^\perp = \{y \in \mathbb{R}^n : \langle y, x\rangle = 0 \text{ minden } x \in M \text{ esetén }\}$$

az M *ortogonális komplementerének* nevezzük.

Könnyen ellenőrizhető, hogy $M^\perp$ is altér az $\mathbb{R}^n$ térben.

31.14 Példa. A három dimenziós térben egy origón átmenő egyenes merőleges kiegészítője az origón átmenő, az egyenesre merőleges sík. Fordítva, a sík merőleges kiegészítője a rá merőleges egyenes.

Ha például az M altér két adott vektor által generált altér, például

$$M = \operatorname{lin}\left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \right\}$$

akkor a merőleges kiegészítője az alábbi egyetlen vektor által generált altér:

$$M^\perp = \operatorname{lin}\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$$

Ez fordítva is igaz, ezt fogalmazza meg a következő tétel.

31.15 Tétel. *Bármely M altérre $(M^\perp)^\perp = M$.*

Az állítást könnyen ellenőrizhetjük közvetlenül a definíció alapján.

31.16 Tétel. *Legyen $a \in \mathbb{R}^n$ és M egy tetszőleges altér. Akkor van pontosan egy olyan $u \in M$ vektor, amelyre*

$$a - u \in M^\perp$$

Bizonyítás. Válasszunk egy $b_1, \ldots, b_k$ ortonormált bázist az M altérben. Keressük az u vektort az M altérben az

$$u = \alpha_1 b_1 + \ldots + \alpha_k b_k$$

alakban. Az ismeretlen együtthatókat úgy kell megválasztanunk, hogy $a - u$ mindegyik bázisvektorra ortogonális legyen. Ez azt jelenti, hogy

$$\langle b_i, a - u \rangle = \langle b_i, a \rangle - \alpha_i = 0$$

minden $i = 1, \ldots, k$ indexre, amivel az ismeretlen együtthatókat meghatároztuk.
□

Ezt az u vektort az a ortogonális vetületének nevezzük az M altérre.

31.17 Tétel. *Legyen M az $\mathbb{R}^n$ egy tetszőleges altere. Akkor bármely $a \in \mathbb{R}^n$ vektor egyértelműen írható fel*

$$a = u + v$$

alakban, ahol $u \in M$ és $v \in M^\perp$.

Bizonyítás. Jelentse $u \in M$ az a ortogonális vetületét az M altérre. Ekkor a $v = a - u$ ortogonális az M altérre, ezért $v \in M^\perp$.

Az egyértelműség abból következik, hogy ha

$$a = u' + v'$$

is ilyen előállítás lenne, akkor a két egyenletet kivonva $u - u' = v - v'$ adódna. Ez azt jelentené, hogy $u - u' \in M$ és $u - u' \in M^\perp$. Tehát $u - u'$ merőleges lenne saját magára is, azaz

$$0 = \langle u - u', u - u' \rangle = \|u - u'\|^2$$

és ez csak úgy lehetséges, ha $u - u' = 0$ és ugyanígy $v - v' = 0$. □

31.18 Tétel. *Legyen M az $\mathbb{R}^n$ tér egy tetszőleges altere. Akkor*

$$\dim M + \dim M^\perp = n\,.$$

Bizonyítás. Tekintsük az M valamely $u_1, \ldots, u_k$ ortonormált bázisát, illetve az $M^\perp$ valamely $v_1, \ldots, v_m$ ortonormált bázisát. Ekkor az előző tételünk szerint az

$$u_1, \ldots, u_k, v_1, \ldots, v_m$$

vektorrendszer a tér minden vektorát előállítja, tehát generátorrendszer, ezért $k + m \geq n$. Másrészt a vektorrendszer elemei páronként ortogonálisak, ezért lineárisan függetlenek, ezért $k + m \leq n$. Következésképpen $k + m = n$. □

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 II/1 szakasza kidolgozott példáinak feldolgozása.
2. Házi feladatok: a II/1 szakasz 1.3.2, 1.3.4, 1.4.3, 1.4.4, 1.5.3, 1.5.4, 1.6.3, 1.7.3, 1.8.4, 1.9.2, 1.9.8, 1.10.2 és 1.10.3 feladatai.
3. Tankönyv-1 12.4 és 12.5 szakaszai.

32. fejezet

A spektráltétel

32.1. Transzponált mátrix

Tekintsük az $\mathbb{R}^n$ tér egy A lineáris transzformációját, azaz $n \times n$-es mátrixot.

32.1 Definíció. Az A *transzponáltján* azt az A^T transzformációt értjük, amelyre

$$\langle A^T y, x \rangle = \langle y, Ax \rangle$$

minden $x, y \in \mathbb{R}^n$ esetén.

Vajon milyen alakú az A^T mátrix? Irjuk fel a definíciót speciálisan a bázisvektorokra, akkor

$$\langle A^T e_i, e_j \rangle = \langle e_i, Ae_j \rangle$$

bármely i és j indexekre. Az egyenlőség jobb oldalán a_{ij} azaz A i-ik sorának j-ik eleme áll, a bal oldalon A^T j-ik sorának i-ik eleme. Tehát az A^T mátrixot az A sorainak és oszlopainak felcserélésével kapjuk.

Úgy is fogalmazhatjuk, hogy az A^T mátrix az A elemeinek a fődiagonálisra tükrözésével nyerhető. Nyilvánvalóan $(A^T)^T = A$.

32.2 Tétel. *Bármely A $n \times n$-es mátrixra*

$$\ker A = \left(\operatorname{im} A^T\right)^{\perp}$$

Bizonyítás. Egyrészt, ha az x vektor ortogonális az $\operatorname{im} A^T$ altérre, akkor

$$0 = \langle A^T y, x \rangle = \langle y, Ax \rangle$$

minden y vektor esetén. Ez csak úgy lehetséges, ha $Ax = 0$, azaz $x \in \ker A$.

Fordítva, a fenti egyenlőség miatt $\ker A$ minden eleme ortogonális az $\operatorname{im} A^T$ altérre. □

Ebből az észrevételből könnyen következik a mátrixok rangtétele.

32.3 Tétel. **(Mátrixok rangtétele)** *Bármely A $n \times n$-es mátrixra*

$$\operatorname{rank} A = \operatorname{rank} A^T$$

Bizonyítás. Valóban, az előző tételünk és a 31.18 Tétel szerint

$$\dim \ker A + \dim \operatorname{im} A^T = n$$

és innen a 27.7 Tétel alapján $\dim \operatorname{im} A = \dim \operatorname{im} A^T$, ami az állításunkat igazolja. □

Ezt a tételünket úgy is megfogalmazhatjuk, hogy egy négyzetes mátrixban a lineárisan független oszlopok száma megegyezik a lineárisan független sorok számával.

32.2. Ortogonális mátrixok

32.4 Definíció. Az $\mathbb{R}^n$ tér egy S transzformációját *ortogonálisnak* nevezzük, ha invertálható, és $S^{-1} = S^T$.

Vajon milyen alakú S mátrixa? Az $S^T S = E$ egyenlőség azt jelenti, hogy az S i-ik oszlopának önmagával vett skaláris szorzata 1, továbbá $i \neq j$ esetén az i-ik és a j-ik oszlopok skaláris szorzata nulla. Tehát mindegyik oszlop egységnyi normájú, és a különböző oszlopok páronként ortogonálisok. Innen ered az elnevezés is.

32.5 Példa. Jelentse például S a sík vektorainak origó körüli elforgatását φ szöggel pozitív irányba. Amint már láttuk, ennek a transzformációnak a mátrixa

$$S = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$$

amelyről könnyen látható, hogy ortogonális, hiszen az oszlopok egységnyi normájúak, és a két oszlop skaláris szorzata nulla. Tehát

$$S^{-1} = \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix}$$

ami éppen a $-\varphi$ szöggel történő elforgatás mátrixa. Ellenőrizzük ezt közvetlenül az $S^T S$ szorzatmátrix kiszámolásával is.

32.6 Tétel. *Egy ortogonális transzformáció megtartja a vektorok hosszát.*

Bizonyítás. Valóban, ha S ortogonális transzformáció, akkor

$$\|Sx\|^2 = \langle Sx, Sx\rangle = \langle S^T Sx, x\rangle = \langle x, x\rangle = \|x\|^2$$

minden x vektor esetén. $\square$

32.7 Tétel. *Egy ortogonális transzformáció minden sajátértéke egységnyi abszolút értékű.*

Bizonyítás. Ha λ az S ortogonális transzformáció sajátértéke, és $v \neq 0$ egy hozzá tartozó sajátvektor, akkor

$$|\lambda|^2 \cdot \|v\|^2 = \langle \lambda v, \lambda v\rangle = \langle Sv, Sv\rangle = \langle S^T Sv, v\rangle = \|v\|^2$$

ahonnan adódik, hogy $|\lambda|^2 = 1$. $\square$

32.3. Szimmetrikus mátrixok

32.8 Definíció. Az $\mathbb{R}^n$ tér egy A transzformációját *szimmetrikusnak* nevezzük, ha $A = A^T$.

Világos, hogy az A mátrixa ilyenkor szimmetrikus (innen a név) a fődiagonálisra, azaz $a_{ij} = a_{ji}$ minden i és j indexre. A 32.2 Tétel speciális eseteként adódik az alábbi állítás.

32.9 Tétel. *Ha A szimmetrikus mátrix, akkor*

$$\ker A = (\operatorname{im} A)^{\perp}$$

Mit mondhatunk egy A $n \times n$-es szimmetrikus transzformáció sajátértékeiről és sajátvektorairól? Jelentse

$$P(\lambda) = \det(A - \lambda E)$$

az A (n-edfokú) karakterisztikus polinomját.

32.10 Tétel. *A P polinomnak van valós gyöke. Ezért az A szimmetrikus transzformációnak létezik valós sajátértéke, és hozzá tartozó nem nulla sajátvektora.*

A tétel bizonyítása elég komplikált számolást igényel, ezért azt elhagyjuk. Megjegyezzük, hogy az is igazolható, hogy a P polinomnak multiplicitással számolva pontosan n valós gyöke van.

32.11 Tétel. *Ha A szimmetrikus transzformáció, akkor az $\mathbb{R}^n$ térnek létezik az A sajátvektoraiból álló ortonormált bázisa.*

Bizonyítás. A bizonyítást teljes indukcióval végezzük az $1 \leq k \leq n$ értékekre. Az előző tétel alapján, ha $k = 1$, akkor az A transzformációnak van egy λ_1 valós sajátértéke, és egy hozzá tartozó v_1 sajátvektor, amelyre $\|v_1\| = 1$.

Tegyük fel, hogy már találtunk $v_1, \ldots, v_{k-1}$ ortonormált sajátvektort, és jelölje M az általuk generált alteret. Ekkor A az $M^\perp$ alteret önmagába képezi (invariáns), ha ugyanis $x \in M^\perp$ tetszőleges vektor, akkor

$$\langle v_i, Ax \rangle = \langle Av_i, x \rangle = \langle \lambda v_i, x \rangle = 0$$

minden $i = 1, \ldots, k-1$ esetén, tehát $Ax \in M^\perp$. Ha most az A szimmetrikus transzformációt az $M^\perp$ altérre leszűkítve tekintjük, akkor az előző tételünk szerint itt is van legalább egy valós λ_k sajátértéke, és hozzá tartozó v_k sajátvektora, amelyre $\|v_k\| = 1$.

A konstrukció folytán ez a v_k sajátvektor ortogonális a $v_1, \ldots, v_{k-1}$ sajátvektorokra, ezért $v_1, \ldots, v_k$ ortonormált rendszert alkotnak. □

32.4. Szimmetrikus mátrixok spektráltétele

Ebben a szakaszban azt mutatjuk meg, hogy hogyan végezhető el egy szimmetrikus mátrix diagonális alakra transzformálása.

Legyen tehát A egy adott $n \times n$-es szimmetrikus mátrix. A 32.11 Tétel értelmében található az $\mathbb{R}^n$ térnek olyan ortonormált bázisa, amely az A sajátvektoraiból áll. Jelentse S azt a mátrixot, amelynek oszlopai rendre ezek a sajátvektorok.

Világos, hogy S ekkor ortogonális mátrix, tehát $S^{-1} = S^T$. A 29.10 Tételt így speciálisan szimmetrikus mátrixokra a következő módon fogalmazhatjuk át.

32.12 Tétel. **(Szimmetrikus mátrixok spektráltétele)** *Legyen A $n \times n$-es szimmetrikus mátrix, és tekintsük a tér egy sajátvektorokból álló ortonormált bázisát. Jelentse S a sajátvektorokból alkotott mátrixot. Ekkor S ortogonális, továbbá A mátrixa a sajátvektorok bázisában*

$$\hat{A} = S^T A S$$

alakú, ahol $\hat{A}$ a következő alakú diagonális mátrix:

$$\hat{A} = \begin{bmatrix} \lambda_1 & \ldots & 0 \\ & \ddots & \\ 0 & \ldots & \lambda_n \end{bmatrix}$$

ahol a fődiagonálisban rendre a megfelelő sajátértékek állnak.

Nézzük meg, hogy egy adott A szimmetrikus mátrix esetén hogyan állítható elő az S mátrix.

Az egyszerű eset az, amikor az A mátrixnak n számú különböző sajátértéke van. Ilyenkor a megfelelő sajátvektorok automatikusan ortogonálisok egymásra. Ezért az S mátrixot úgy nyerjük, hogy az egymás utáni oszlopokba írjuk az egységnyi normájú sajátvektorokat.

32.13 Példa. Tekintsük a következő szimmetrikus A mátrixot

$$A = \begin{bmatrix} 2 & 0 & 2 \\ 0 & -2 & 0 \\ 2 & 0 & 5 \end{bmatrix}$$

Határozzuk meg a sajátértékeket! Az A karakterisztikus polinomja

$$P(\lambda) = (\lambda^2 - 7\lambda + 6)(-2 - \lambda)$$

amelynek gyökei $\lambda_1 = 1$, $\lambda_2 = -2$ és $\lambda_3 = 6$. A megfelelő sajátvektorokat különböző λ sajátértékekre megkapjuk az $(A - \lambda E)x = 0$ homogén rendszer megoldásaiként. Ezek pl. rendre

$$v_1 = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} \quad v_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad v_3 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

Ezek a sajátvektorok egymásra ortogonálisok, amelyeket normálni kell, így az S ortogonális mátrix:

$$S = \begin{bmatrix} \frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{5}} & 0 & \frac{2}{\sqrt{5}} \end{bmatrix} \quad \text{és} \quad \hat{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

Ekkor

$$\hat{A} = S^T A S$$

amelyet ellenőrizzünk közvetlenül a kijelölt szorzások elvégzésével is.

Kicsit bonyolultabb a helyzet akkor, ha egy sajátértékhez több lineárisan független sajátérték is tartozik. Ilyenkor szükség lehet a Gram-Schmidt-féle eljárásra ahhoz, hogy ezek a sajátvektorok is ortogonálisok legyenek egymásra. Ezt az esetet mutatjuk be egy példán keresztül.

32.14 Példa. Módosítsuk az előző példát a következő módon:

$$A = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 6 & 0 \\ 2 & 0 & 5 \end{bmatrix}$$

ekkor a karakterisztikus polinom a következő alakot ölti

$$P(\lambda) = (6-\lambda)(\lambda^2 - 7\lambda + 6)$$

amelynek gyökei $\lambda_1 = 1$ és $\lambda_2 = 6$, ez utóbbi kétszeres gyök. A λ_1 sajátértékhez alkalmas sajátvektor az előző példa v_1 sajátvektora. A $\lambda_2 = 6$ sajátérték esetén az $(A - 6E)x = 0$ homogén rendszer szabadságfoka 2, ezért két lineárisan független megoldás van, például

$$v_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \quad \text{és} \quad v_3 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

Ez a két vektor azonban nem ortogonális, ezért alkalmazzuk a Gram-Schmidt-féle eljárást. Ekkor v_3 helyett az alábbi u_3 sajátvektor adódik:

$$u_3 = v_3 - \frac{\langle v_3, v_2 \rangle}{\|v_2\|^2} v_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Tehát ekkor a következő mátrixokhoz jutunk

$$S = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 0 \\ 0 & 0 & 1 \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} & 0 \end{bmatrix} \quad \text{és} \quad \hat{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

és ekkor

$$\hat{A} = S^T A S$$

amelyet újra ellenőrizzünk közvetlenül a kijelölt mátrixok összeszorzásának elvégzésével is.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 II/6 szakasza kidolgozott példáinak feldolgozása.
2. Házi feladatok: a II/6 szakasz 6.2.4, 6.3.2, 6.3.3, 6.3.4, 6.4.2, 6.4.3, 6.4.6, 6.4.7, és 6.4.10 feladatai.
3. Tankönyv-1 14.5 és 14.6 szakaszok.

33. fejezet

Kvadratikus alakok

33.1. Kvadratikus alakok

Az $\mathbb{R}^n$ téren értelemezett tisztán másodfokú függvényt, azaz kvadratikus alakot, az alábbi módon értelmezzük.

33.1 Definíció. A $Q : \mathbb{R}^n \to \mathbb{R}$ függvényt *kvadratikus alaknak* nevezzük, ha

$$Q(x) = Q(x_1, \dots, x_n) = a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2 + a_{13}x_1x_3 + \dots + a_{nn}x_n^2$$

alakú, azaz olyan hatványfüggvény, amelyben minden tag tisztán másodfokú.

Egy kvadratikus alakhoz mindig megadhatunk egy A $n \times n$-es mátrixot úgy, hogy Q a

$$Q(x) = \langle x, Ax \rangle \tag{33.1}$$

skaláris szorzat alakjában írható fel, amint azt a következő példa mutatja.

33.2 Példa. Tekintsük az $\mathbb{R}^3$ téren értelmezett

$$Q(x) = Q(x_1, x_2, x_3) = 3x_1^2 - 4x_1x_2 + 2x_1x_3 + x_2^2 + 6x_2x_3 - 5x_3^2$$

kvadratikus alakot. Az együtthatókból állítsuk össze az

$$A = \begin{bmatrix} 3 & -4 & 2 \\ 0 & 1 & 6 \\ 0 & 0 & -5 \end{bmatrix}$$

mátrixot. Ellenőrizzük, hogy erre az A mátrixra valóban $Q(x) = \langle x, Ax \rangle$. Ez azonban nem az egyetlen ilyen tulajdonságú mátrix. Ha ugyanis a

$$B = \begin{bmatrix} 3 & -2 & 1 \\ -2 & 1 & 3 \\ 1 & 3 & -5 \end{bmatrix}$$

mátrixot tekintjük, akkor erre is teljesül, hogy $Q(x) = \langle x, Bx \rangle$ minden $x \in \mathbb{R}^3$ mellett.

Ezen észrevétel alapján azt mondhatjuk, hogy végtelen sok olyan A mátrix található, amelyre fennáll a (33.1) azonosság, de csak egyetlen olyan B mátrix, amelynek elemei a főátlóra nézve szimmetrikusak. Ezt a szimmetrikus mátrixot bármely $Q(x) = \langle x, Ax \rangle$ kvadratikus alak esetén a

$$B = \frac{1}{2}\left(A + A^T\right)$$

egyenlőséggel adhatjuk meg, amelyre teljesül, hogy

$$Q(x) = \langle x, Ax \rangle = \langle x, Bx \rangle$$

minden $x \in \mathbb{R}^n$ esetén.

33.2. Kvadratikus alak mátrixa

Szimmetrikus mátrixokra megfogalmazhatjuk az előző szakasz észrevételét kvadratikus alakokra.

33.3 Tétel. *Tekintsünk egy $Q : \mathbb{R}^n \to \mathbb{R}$ kvadratikus alakot. Akkor található pontosan egy B $n \times n$-es szimmetrikus mátrix, amelyre*

$$Q(x) = \langle x, Bx \rangle$$

minden $x \in \mathbb{R}^n$ esetén. Fordítva, bármely B szimmetrikus mátrix a fenti skalárszorzattal egy kvadratikus alakot definiál.

Ez a tételünk azt fogalmazza meg, hogy a kavadratikus alakok és a szimmetrikus mátrixok között kölcsönösen egyértelmű megfeleltetés van.

33.4 Példa. Tekintsük például a

$$Q(x_1, x_2, x_3) = 2x_1^2 - 2x_1x_2 + 4x_1x_3 - x_2^2 + 8x_2x_3 + 3x_3^2$$

kvadratikus alakot, és állítsuk elő a neki megfelelő B szimmetrikus mátrixot.

Ekkor az előző szakasz példájához hasonlóan a vegyes szorzatok együtthatóit megfelezzük, és a következő szimmetrikus mátrixhoz jutunk:

$$B = \begin{bmatrix} 2 & -1 & 2 \\ -1 & -1 & 4 \\ 2 & 4 & 3 \end{bmatrix}$$

Ellenőrizzük, hogy ekkor valóban $Q(x) = \langle x, Bx \rangle$ minden $x \in \mathbb{R}^3$ esetén.

33.3. Kvadratikus alakok definitsége

A többváltozós szélsőérték megkereséséhez szükségünk lesz arra, hogy kvadratikus alakok milyen előjelű értékeket vesznek fel. Ehhez vezetjük be a következő definíciót.

33.5 Definíció. Azt mondjuk, hogy a Q kvadratikus alak

- pozitív definit, ha $Q(x) > 0$ minden $x \in \mathbb{R}^n$, $x \neq 0$ esetén,
- pozitív szemidefinit, ha $Q(x) \geq 0$ minden $x \in \mathbb{R}^n$ esetén, és van olyan $x_0 \neq 0$, hogy $Q(x_0) = 0$,
- negatív definit, ha $Q(x) < 0$ minden $x \in \mathbb{R}^n$, $x \neq 0$ esetén,
- negatív szemidefinit, ha $Q(x) \leq 0$ minden $x \in \mathbb{R}^n$ esetén, és van olyan $x_0 \neq 0$, hogy $Q(x_0) = 0$,
- indefinit, ha a fentiek egyike sem teljesül.

33.6 Példa. Például a háromváltozós

$$Q(x_1, x_2, x_3) = 2x_1^2 - 2x_1x_2 + x_2^2 + 3x_3^2$$

kvadratikus alak pozitív definit, hiszen a következőképpen alakítható át teljes négyzetek összegévé:

$$Q(x_1, x_2, x_3) = x_1^2 + (x_1 - x_2)^2 + 3x_3^2\,,$$

és ez a kifejezés minden $x \neq 0$ vektor esetén pozitív.

Ugyanakkor a

$$Q(x_1, x_2, x_3) = x_1^2 - 2x_1x_2 + x_2^2 + 3x_3^2 = (x_1 - x_2)^2 + 3x_3^2$$

kvadratikus alak csak pozitív szemidefinit, hiszen egyrészt $Q(x) \geq 0$ minden x vektor mellett, másrészt például $Q(1,1,0) = 0$, azaz található olyan nem nulla vektor, amelyen Q nulla értéket vesz fel.

Teljesen hasonlóan a

$$Q(x_1, x_2, x_3) = 2x_1^2 - 2x_1x_2 + x_2^2 - 3x_3^2$$

kvadratikus alak indefinit, hiszen egyaránt felvesz pozitív és negatív értékeket is. Behelyettesítéssel ellenőrizhető, hogy például $Q(1,1,0) = 1 > 0$, és $Q(0,0,1) = -3 < 0$.

A továbbiakban a definitség fogalmait ugyanilyen módon fogjuk használni a Q kvadratikus alaknak megfelelő B szimmetrikus mátrixra is.

33.4. Teljes négyzetté alakítás

Egy kvadratikus alak definitsége nagyon egyszerűen eldönthető, ha csak tiszta négyzetes tagokból áll (azaz nincs vegyesszorzat).

33.7 Példa. Tekintsük például az $\mathbb{R}^4$ téren a

$$Q(x) = 5x_1^2 + 3x_2^2 + 9x_3^2 + 2x_4^2$$

kvadratikus alakot. Ez nyilván pozitív definit, hiszen a négyzetösszeg pozitív, ha $x \neq 0$.

Az is könnyen látható, hogy a

$$R(x) = 3x_1^2 + 5x_2^2 - 2x_4^2$$

kvadratikus alak viszont indefinit, hiszen például az

$$x_1 = 1 \quad x_2 = 1 \quad x_3 = 0 \quad x_4 = 0$$

vektoron felvett értéke pozitív, ugyanakkor az

$$x_1 = 0 \quad x_2 = 0 \quad x_3 = 0 \quad x_4 = 1$$

választással R negatív értéket vesz fel.

Teljesen hasonló módon ellenőrizhetjük, hogy például a

$$P(x) = 3x_2^2 + x_3^2 + x_4^2$$

kvadratikus alak (FIGYELEM, x_1 hiányzik!) pozitív szemidefinit. Egyrészt ugyanis a négyzetösszeg nemnegatív, másrészt található olyan $x \neq 0$ vektor, nevezetesen

$$x_1 = 1 \quad \text{és} \quad x_2 = x_3 = x_4 = 0$$

amelyre P értéke nulla. Ezért P nem lehet pozitív definit.

A fenti példa észrevételeit a következő tételben fogalmazhatjuk meg.

33.8 Tétel. *Tegyük fel, hogy Q olyan kvadratikus alak, amely csak tisztán négyzetes tagokat tartalmaz, nevezetesen*

$$Q(x) = b_1x_1^2 + b_2x_2^2 + \ldots + b_nx_n^2$$

Akkor az együtthatók előjelei alapján a következő eseteket különböztetjük meg.

- *Ha minden k indexre $b_k > 0$, akkor Q pozitív definit.*
- *Ha mindegyik $b_k \geq 0$, és van olyan j, amelyre $b_j = 0$, akkor Q pozitív szemidefinit.*
- *Ha az együtthatók között előfordul pozitív és negatív is, akkor Q indefinit.*

Természetesen analóg eseteket fogalmazhatunk meg a negatív definit, illetve negatív szemidefinit esetekre is. Tételünk alapján a továbbiakban azt vizsgáljuk, hogy egy kvadratikus alak hogyan transzformálható olyanná, amelyben csak tiszta négyzetes tagok állnak (teljes négyzetté alakítás n-dimenzióban).

33.5. Definitség a sajátértékek alapján

Tekintsük a $Q : \mathbb{R}^n \to \mathbb{R}$ kvadratikus alakot, és jelölje B a hozzá tartozó $n \times n$-es szimmetrikus mátrixot.

A szimmetrikus mátrixok spektráltétele (lásd előző fejezet) szerint az $\mathbb{R}^n$ térnek létezik a B sajátvektoraiból álló

$$v_1, \dots, v_n$$

ortonormált bázisa, amelyre

$$Bv_1 = \lambda_1 v_1 \quad \dots \quad Bv_n = \lambda_n v_n \,.$$

Ebben a bázisban a B mátrix a következő $\hat{B}$ diagonális alakot ölti:

$$\hat{B} = \begin{bmatrix} \lambda_1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & \lambda_n \end{bmatrix}$$

Itt $\lambda_1, \dots, \lambda_n$ a B megfelelő sajátvektorai (nem feltétlenül mind különbözőek). Könnyen látható, hogy ezzel a diagonális mátrixszal a kvadratikus alak tiszta négyzetes alakú lesz. Ugyanis a tér bármely $y = y_1 v_1 + \dots + y_n v_n$ vektorára

$$\langle y, \hat{B} y \rangle = \lambda_1 y_1^2 + \dots + \lambda_n y_n^2$$

Ezen észrevétel alapján a következő tételt fogalmazzuk meg.

33.9 Tétel. *Tekintsük a Q kvadratikus alakot, és jelentse B a hozzá tartozó szimmetrikus mátrixot, azaz*

$$Q(x) = \langle x, Bx \rangle$$

minden $x \in \mathbb{R}^n$ esetén. Tekintsük a B sajátértékeit.

- *Ha minden sajátérték pozitív, akkor Q pozitív definit.*
- *Ha minden sajátérték nemnegatív, és van köztük nulla, akkor Q pozitív szemidefinit.*
- *Ha minden sajátérték negatív, akkor Q negatív definit.*
- *Ha minden sajátérték nempozitív, és van köztük nulla, akkor Q negatív szemidefinit.*
- *Ha van pozitív és negatív sajátérték is, akkor Q indefinit.*

Bizonyítás. Az eddigiek alapján csak annyit kell igazolnunk, hogy B és $\hat{B}$ definitsége megegyezik. Ha a korábbi jelöléseinket használva S jelenti a B sajátvektoraiból álló mátrixot, akkor

$$\langle y, \hat{B}y \rangle = \langle y, S^T BSy \rangle = \langle Sy, BSy \rangle = \langle x, Bx \rangle$$

a tér bármely y vektorára. Mivel S invertálható, azért $x = Sy$ a tér minden vektorát előállítja. □

33.10 Példa.

Állapítsuk meg a $Q(x) = 2x_1^2 + 5x_1x_3 + 5x_2^2 - x_3x_1 - x_3^2$ kvadratikus alak definitségét!

Világos, hogy a hozzá tartozó B szimmetrikus mátrix a következő:

$$B = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 5 & 0 \\ 2 & 0 & -1 \end{bmatrix}$$

amelynek karakterisztikus polinomja

$$\det(B - \lambda E) = (5 - \lambda)(\lambda - 3)(\lambda + 2)\,.$$

Ennek gyökei, azaz a B sajátértékei

$$\lambda_1 = 5 \quad \lambda_2 = 3 \quad \text{és} \quad \lambda_3 = -2$$

Ezek között pozitív és negatív is előfordul, ezért Q indefinit.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 II/1 és II/7 szakaszok kidolgozott példáinak feldolgozása.
2. Házi feladatok: a II/1 szakasz 1.6.4, 1.7.3, 1.8.3, 1.8.5, 1.9.3, 1.9.4, 1.9.5, továbbá a II/7 szakasz 7.1.4, 7.2.3, 7.3.4, 7.3.5, 7.4.4, 7.4.5 feladatai.
3. Tankönyv-1 12.5, 15.8 és 15.9 szakaszai.

34. fejezet

Többváltozós függvények deriválása

34.1. Parciális deriváltak

Tekintsünk egy $f : \mathbb{R}^n \to \mathbb{R}$ függvényt. Ezt úgy tekintjük, mint $f(x) = f(x_1, \ldots, x_n)$, azaz az $x \in \mathbb{R}^n$ vektor koordinátáinak n-változós függvényét.

34.1 Definíció. Azt mondjuk, hogy f *parciálisan differenciálható* a k-ik változó szerint az x pontban, ha az

$$F(t) = f(x + te_k)$$

függvény a t változó szerint differenciálható a $t = 0$ pontban, ahol e_k a k-ik standard bázisvektor. Ilyenkor jelölésben

$$F'(0) = f'_k(x) = \frac{\partial f}{\partial x_k}(x)$$

az f k-ik változó szerinti *parciális deriváltja* az x pontban.

Világos, hogy a definíció azt jelenti, hogy a k-ik változó szerinti parciális meghatározásához az összes többi változót állandónak tekintjük, és csak az x_k szerint deriválunk.

34.2 Példa. Tekintsük például az

$$f(x, y) = 5xe^{-2x+3y^2}$$

függvényt a síkon. Ekkor

$$\frac{\partial f}{\partial x}(x, y) = 5e^{-2x+3y^2} - 10xe^{-2x+3y^2}$$

a szorzat deriválási szabálya alapján, és teljesen hasonlóan

$$\frac{\partial f}{\partial y}(x,y) = 30xye^{-2x+3y^2}$$

bármely (x, y) pontban.

34.3 Példa. Egy adott pontbeli parciális derivált meghatározásához néha célszerűbb lehet az először állandónak tekintett változók közvetlen behelyettesítése, majd ezután a deriválás elvégzése. Tekintsük például a háromdimenziós téren az

$$f(x,y,z) = \sqrt{1+x^2+3y^2+2z^2} \cdot (5-x^2-y^2) \cdot e^{-x-2y-2z}$$

függvényt, és számítsuk ki a z változó szerinti parciális deriváltat a $P(2,1,2)$ pontban.

Természetesen megtehetjük, hogy formálisan előállítjuk a z-szerinti deriváltat, majd behelyettesítjük az adott P pont koordinátáit. Ez elég hosszadalmas számolással jár.

Sokkal gyorsabban kapunk eredményt, ha először beírjuk az $x = 2$, $y = 1$ értékeket, ekkor látható, hogy az

$$f(2,1,z) = 0$$

konstanshoz jutunk bármely z mellett. Tehát

$$\frac{\partial f}{\partial z}(2,1,2) = 0\,.$$

Természetesen a parciális derivált nulla bármely más $P(2,1,z)$ pontban is.

34.2. A derivált

34.4 Definíció. Tegyük fel, hogy az $f : \mathbb{R}^n \to \mathbb{R}$ függvénynek az x pontban léteznek a parciális deriváltjai (mindegyik változó szerint). Ekkor az

$$f'(x) = \left[\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x)\right]$$

kifejezést az f *deriváltjának* nevezzük az x pontban.

Ezt a vektort az irodalomban néha az f *gradiensének* is nevezik az x pontban.

34.5 Példa. Például a három dimenziós téren értelmezett alábbi f függvény:

$$f(x, y, z) = 2xy\sqrt{x^2 + y^2 + z^2}$$

esetében a $P(2, 1, 2)$ pontban

$$f'(2, 1, 2) = \left[\frac{26}{3}, \frac{40}{3}, \frac{8}{3}\right]$$

Ellenőrizzük ezt úgy, hogy először kiszámoljuk az $f'(x, y, z)$ vektort, majd behelyettesítjük a $P(2, 1, 2)$ pont koordinátáit!

34.6 Példa. Legyen B adott $n \times n$-es szimmetrikus mátrix, és tekintsük a

$$Q(x) = \langle x, Bx \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} x_i x_j$$

kvadratikus alakot, ahol b_{ij} a B mátrix i-ik sorának j-ik eleme. Határozzuk meg a Q kvadratikus alak x_k változó szerinti parciális deriváltját! Ekkor

$$\frac{\partial Q}{\partial x_k}(x) = \sum_{i=1}^{n} b_{ik} x_i + \sum_{j=1}^{n} b_{kj} x_j$$

hiszen mindazon tagok deriváltja nulla, amelyekben x_k nem szerepel. Ez a kifejezés a B mátrix szimmetriájára tekintettel (ugyanis $b_{ij} = b_{ji}$ minden indexre) úgy is írható, hogy

$$\frac{\partial Q}{\partial x_k}(x) = 2 \sum_{j=1}^{n} b_{kj} x_j$$

minden $k = 1, \ldots, n$ index esetén. A jobb oldalon éppen a $2Bx$ szorzatvektor k-ik eleme áll. Tehát a Q kavdratikus alak deriváltjára az adódik, hogy

$$Q'(x) = 2Bx$$

minden x mellett. Vegyük észre, hogy ez az eredmény teljesen analóg az egyváltozós másodfokú függvény deriváltjával!

34.3. Láncszabály

Legyen az alábbiakban $f : \mathbb{R}^n \to \mathbb{R}$ olyan függvény, amelynek léteznek a parciális deriváltjai, és azok folytonos függvények. Legyenek továbbá $g_1, \ldots, g_n$ differenciálható valós függvények a számegyenesen, és tekintsük az

$$F(t) = f(g_1(t), \ldots, g_n(t))$$

összetett függvényt. Az egyszerűbb írásmód kedvéért vezessük be a

$$g(t) = \begin{bmatrix} g_1(t) \\ \vdots \\ g_n(t) \end{bmatrix}$$

jelölést, akkor ez úgy is írható, hogy

$$F = f \circ g$$

a számegyenesen. Ekkor az egyváltozós Láncszabályhoz (lásd 4.7 Tétel) teljesen hasonlóan igazolható a következő tétel.

34.7 Tétel. **(Láncszabály)** *A fenti feltételek mellett F differenciálható, és*

$$F'(t) = \langle f'(g(t)), g'(t)\rangle = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(g(t))g_k'(t)$$

minden $t \in \mathbb{R}$ pontban.

34.8 Példa. Tekintsük például a síkon az

$$f(x,y) = x^2 - xy - 2y^2$$

függvényt, és legyenek

$$x = g_1(t) = \cos t \quad \text{valamint} \quad y = g_2(t) = \sin t$$

Ekkor az $F = f \circ g$ függvény deriváltja a Láncszabály szerint

$$F'(t) = -\frac{\partial f}{\partial x}(g(t))\sin t + \frac{\partial f}{\partial y}(g(t))\cos t = -4\sin t\cos t + \sin^2 t - \cos^2 t\,.$$

Ellenőrizzük, hogy g_1 és g_2 közvetlen behelyettesítésével, majd a kijelölt deriválás elvégzésével ugyanerre az eredményre jutunk.

34.9 Példa. Legyen ismét $f : \mathbb{R}^n \to \mathbb{R}$ olyan függvény, amelynek léteznek a parciális deriváltjai, és azok folytonos függvények. Legyen továbbá $v \in \mathbb{R}^n$ egy tetszőleges adott vektor, és tekintsük a

$$g(t) = x + tv$$

ahol x adott rögzített vektor. Határozzuk meg az $F = f \circ g$ deriváltját!

Mivel nyilvánvalóan $g'(t) = v$, azért a Láncszabály alapján

$$F'(t) = \langle f'(x + tv), v\rangle\,.$$

Nevezetesen a $t = 0$ helyen

$$F'(0) = \langle f'(x), v\rangle = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(x)v_k$$

ahol a v_k konstansok a v vektor koordinátái.

34.4. Másodrendű parciális deriváltak

Amennyiben az $f : \mathbb{R}^n \to \mathbb{R}$ függvény i-ik változó szerinti parciális deriváltfüggvénye parciálisan differenciálhat a j-ik változó szerint egy x pontban, akkor ott tekinthetjük az f másodrendű parciális deriváltját:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \text{ vagy } f''_{ij}(x) \quad \text{illetve } i = j \text{ esetén:} \quad \frac{\partial^2 f}{\partial x_i^2}(x) \text{ vagy } f''_{ii}(x)$$

jelöléseket használva. Az előbbit vegyes, az utóbbit tiszta másodrendű parciális deriváltnak nevezzük.

34.10 Példa. Például az

$$f(x, y) = x^2 - 3x^2y^2 + 2y^3$$

függvény esetében

$$\frac{\partial^2 f}{\partial x \partial y}(x, y) = -12xy \quad \text{illetve} \quad \frac{\partial^2 f}{\partial y^2}(x) = -6x^2 + 12y$$

minden x és y mellett.

34.11 Példa. Vizsgáljuk meg újra a 34.9 Példában definiált F függvényt, és adjuk meg az $F''(x + tv)$ második derivált értékét.

Mivel

$$F'(t) = \langle f'(x + tv), v\rangle = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x + tv)v_i$$

azért a deriválást újra elvégezve, a Láncszabály alapján az

$$F''(t) = \sum_{j=1}^{n}\sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(x + tv)v_i v_j$$

kifejezéshez jutunk. Speciálisan $t = 0$ esetén

$$F''(0) = \sum_{j=1}^{n}\sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(x)v_i v_j$$

Ez éppen egy olyan kvadratikus alak a v változóban, amelynek mátrixa

$$A = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix}$$

Ennek a mátrixnak a segítségével a fenti második derivált a következő alakba írható

$$F''(0) = \langle v, Av \rangle$$

minden v vektor mellett.

34.12 Definíció. A fenti $n \times n$-es A mátrixot az f *másodrendű deriváltjának* nevezzük az x pontban, jelölése:

$$f''(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix}$$

Az $f''(x)$ mátrixra használatos a *Hesse-mátrix* elnevezés is.

34.5. Young-tétel

34.13 Példa. Könnyen látható, hogy például az $f(x,y) = 2x^3+5x^2y^3-\ln(xy^2)$ függvény esetében

$$\begin{aligned} \frac{\partial f}{\partial x}(x,y) &= 6x^2 + 10xy^3 - 1/x \\ \frac{\partial f}{\partial y}(x,y) &= 15x^2y^2 - 1/y^2 \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2}(x,y) &= 12x + 10y^3 - 1/x^2 \\ \frac{\partial^2 f}{\partial y^2}(x,y) &= 30x^2y - 2/y^3 \\ \frac{\partial^2 f}{\partial x \partial y}(x,y) &= \frac{\partial^2 f}{\partial y \partial x}(x,y) = 30xy^2 \end{aligned}$$

A fenti példánkban azt láthatjuk, hogy az f függvény vegyes másodrendű parciális deriváltjai megegyeznek. A következő tételünk azt fogalmazza meg, hogy ez nem véletlen, viszonylag általános feltételek mellett ez mindig érvényes.

34.14 Tétel. **(Young)** *Ha az f n-változós függvény másodrendű parciális deriváltfüggvényei léteznek és folytonosak, akkor az $f''(x)$ Hesse-mátrix szimmetrikus, azaz*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$$

bármely $i, j = 1, 2, \ldots, n$ *indexekre.*

Bizonyítás. Nyilván elég a bizonyítást a kétváltozós esetre elvégezni. Tegyük fel, hogy $f : \mathbb{R}^2 \to \mathbb{R}$ a feltételeknek megfelelő. Legyen $v \in \mathbb{R}$ rögzített, és tekintsük az

$$F(t) = f(t, y+v) - f(t, y) , \qquad G(t) = f(x+v, t) - f(x, t)$$

függvényeket. A feltevésünk szerint ezek differenciálhatók az x, illetve az y pont egy környezetében, és

$$F(x+v) - F(x) = G(y+v) - G(y) . \tag{34.1}$$

A Lagrange-féle középértéktétel szerint található olyan $0 < t < 1$ szám, amelyre

$$F(x+v) - F(x) = F'(x+tv)v ,$$

azaz az F definíciójára tekintettel

$$\begin{aligned} F(x+v) - F(x) &= \left(f_1'(x+tv, y+v) - f_1'(x+tv, y)\right) v \\ &= \left(D_{12} f(x+tv, y)v + o(v)\right) v . \end{aligned}$$

Innen a második derivált folytonossága alapján

$$\lim_{v \to 0} \frac{F(x+v) - F(x)}{v^2} = \frac{\partial^2 f}{\partial x \partial y}(x, y) .$$

Teljesen hasonló gondolatmenettel az adódik, hogy

$$\lim_{v \to 0} \frac{G(y+v) - G(y)}{v^2} = \frac{\partial^2 f}{\partial y \partial x}(x, y) .$$

Ezért a (34.1) egyenlőségből azonnal következik, hogy

$$\frac{\partial^2 f}{\partial x \partial y}(x, y) = \frac{\partial^2 f}{\partial y \partial x}(x, y) ,$$

azaz a második derivált szimmetrikus mátrix. □

34.15 Példa. Legyen például $f : \mathbb{R}^3 \to \mathbb{R}$ az

$$f(x, y, z) = 2x^2 y + xyz - y^2 z^2$$

formulával értelmezett függvény. Ellenőrizzük, hogy a második derivált az (x, y, z) helyen az

$$f''(x, y, z) = \begin{bmatrix} 4y & 4x+z & y \\ 4x+z & -2z^2 & x-4yz \\ y & x-4yz & -2y^2 \end{bmatrix}$$

szimmetrikus mátrix.

34.16 Példa. Tekintsünk egy B $n\times n$-es szimmetrikus mátrixot, és határozzuk meg a

$$Q(x) = \langle x, Bx\rangle$$

kvadratikus alak második deriváltját.

Mivel $Q'(x) = 2Bx$, amelyben mindegyik koordináta elsőfokon szerepel, ezért könnyen látható, hogy a Hesse-mátrix a

$$Q''(x) = 2B$$

szimmetrikus mátrix minden x pontban.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 III/1 és III/2 szakaszai kidolgozott példáinak feldolgozása.
2. Házi feladatok: III/1 szakasz 1.1.4, 1.1.5, 1.1.6, 1.1.7, 1.4.3, 1.4.4, 1.4.5, 1.5.3, 1.5.4, 1.6.3, 1.6.5 feladatai, továbbá a III/2 szakasz 2.2.2, 2.2.3, 2.2.4, 2.2.6, 2.2.8, 2.2.10, 2.2.11, 2.2.14, 2.2.15, 2.2.16 feladatai.
3. Tankönyv-1 15.3, 15.4, 15.6, 16.1 és 16.2 szakaszai, valamint a 17. fejezet.

35. fejezet

Többváltozós függvények szélsőértéke

35.1. Lokális szélsőérték

35.1 Definíció. Az $\mathbb{R}^n$ tér origó középpontú egységgömbjén a

$$B = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$$

halmazt értjük. Világos, hogy valamely $a \in \mathbb{R}^n$ pont körüli $r > 0$ sugarú gömb az

$$a + rB = \{x \in \mathbb{R}^n : \|x - a\| \leq r\}$$

formulával adható meg.

Tekintsünk egy $f : \mathbb{R}^n \to \mathbb{R}$ függvényt. Azt mondjuk, hogy az értelmezési tartomány valamely a pontja az f lokális minimumhelye, ha található olyan $\varepsilon > 0$, hogy

$$f(x) \geq f(a)$$

az értelmezési tartomány minden olyan x pontjában, amelyre $x \in a + \varepsilon B$, azaz $\|x - a\| \leq \varepsilon$.

Hasonlóan értelmezzük a lokális maximum fogalmát, és értelemszerűen fogalmazhatjuk meg a globális minimum és maximum definícióját is.

35.2. Elsőrendű szükséges feltétel

Tegyük fel a továbbiakban, hogy az $f : \mathbb{R}^n \to \mathbb{R}$ függvény parciális deriváltjai léteznek és folytonosak az a pont egy környezetében.

35.2 Tétel. *Ha az $a \in \mathbb{R}^n$ pont az f lokális minimumhelye, akkor*

$$\frac{\partial f}{\partial x_1}(a) = \ldots = \frac{\partial f}{\partial x_n}(a) = 0\,.$$

Bizonyítás. Tekintsük ugyanis az $e_k \in \mathbb{R}^n$ egységvektorokat, és jelölje

$$F(t) = f(a + te_k)\,.$$

A lokális minimum definíciója alapján az F függvénynek lokális minimumhelye van a $t = 0$ pontban, másrészt a Láncszabály szerint F differenciálható is, éspedig

$$F'(t) = \langle f'(a + te_k), e_k \rangle\,.$$

Innen adódik, hogy minden $k = 1, \ldots, n$ indexre:

$$0 = F'(0) = \langle f'(a), e_k \rangle = \frac{\partial f}{\partial x_k}(a)$$

A fenti tétel szerint a parciális deriváltakra felírt egyenletrendszer megoldásai között kereshetjük a függvény szélsőértékhelyeit. Ez az egyenletrendszer azonban (az egyváltozós esethez hasonlóan) csak szükséges feltételt fogalmaz meg. Például az

$$f(x, y) = x^3 y^2$$

függvény esetében az $f_1'(x, y) = f_2'(x, y) = 0$ egyenletrendszer egyik megoldása $(x, y) = (0, 0)$. Ekkor

$$f(0, 0) = 0$$

de ez nem lehet szélsőérték, hiszen f az origó körüli bármilyen sugarú gömb belsejében felvesz pozitív és negatív értékeket is.

Szükségünk van tehát másodrendű (szükséges, illetve elégséges) feltételekre.

35.3. Másodrendű szükséges feltétel

A továbbiakban feltesszük, hogy $f : \mathbb{R}^n \to \mathbb{R}$ kétszer folytonosan differenciálható az a pont egy környezetében.

35.3 Tétel. *Tegyük fel, hogy a az f lokális minimuhelye. Akkor az f Hesse-mátrixa az a helyen pozitív szemidefinit.*

Bizonyítás. Legyen $v \in \mathbb{R}^n$ tetszőleges, és vezessük be a korábbiakban vizsgált

$$F(t) = f(a + tv)$$

függvényt. A feltételünk szerint F kétszer differenciálható, és ha a az f lokális minimumhelye, akkor 0 az F lokális minimumhelye, ezért $F''(0) \geq 0$. Ez azt jelenti, hogy

$$0 \leq F''(0) = \langle f''(a)v, v\rangle$$

Mivel $v \in \mathbb{R}^n$ tetszőleges volt, ez éppen azt jeleneti, hogy a Hesse-mátrix pozitív szemidefinit. □

Ez a tételünk még nem ad elégséges (csak szükséges) feltételt a szélsőértékre, elég, ha az

$$f(x, y) = x^5 + y^4$$

függvényre gondolunk. A $(0, 0)$ pontban a parciális deriváltak nullák és a Hesse-mátrix is a nulla mátrix (tehát egyszerre pozitív és negatív szemidefinit), de ez a pont nem szélsőérték.

Természetesen a tételünkkel analóg állítást fogalmazhatunk meg lokális maximumhely esetére is, ilyenkor a Hesse-mátrix negatív szemidefinit.

35.4. A szélsőérték elégséges feltétele

Tegyük fel, hogy az $f : \mathbb{R}^n \to \mathbb{R}$ függvény másodrendű parciális deriváltjai léteznek, és folytonosak az a pont egy környezetében.

35.4 Tétel. *Tegyük fel, hogy az a pontban az f parciális deriváltjai nullák, és itt az $f''(a)$ Hesse-mátrix pozitív definit. Akkor a az f lokális minimumhelye.*

Magától értetődően az $f''(a)$ negatív definitsége lokális maximumot jelent.

A tétel bizonyítása egy kicsit túlmegy a megtárgyalt anyagon, ezért azt elhagyjuk. Megjegyezzük azért, hogy természetes lenne arra gondolni, hogy az

$$F(t) = f(a + tv)$$

függvénynek legyen lokális minimuma van a $t = 0$ pontban bármely v vektor esetén. Ennek valóban elégséges feltétele az, hogy $F''(0) > 0$ bármely $v \neq 0$ esetén, ami azzal ekvivalens, hogy az $f''(a)$ Hesse-mátrix pozitív definit. Csakhogy a problémát az okozza, hogy az F lokális minimumhelye a $t = 0$ pontban bármely v vektor mellett még nem elegendő ahhoz, hogy az f függvénynek lokális minimumhelye legyen az a pontban. Ezt a következő példában mutatjuk meg.

35.5 Példa. Tekintsük síkon a következő függvényt:

$$f(x, y) = \begin{cases} x^2 & \text{ha } y = x^2 \text{ és } x > 0 \\ -x^2 & \text{ha } y = x^2 \text{ és } x < 0 \\ x^2 + y^2 & \text{különben} \end{cases}$$

Ebben az esetben az origó az f függvénynek nem szélsőértékhelye, hiszen annak bármely környezetében felvesz pozitív és negatív értékeket is.

Ha azonban bármely v nem nulla síkbeli vektort tekintünk, akkor az

$$F(t) = f((0,0) + tv)$$

függvénynek a $t = 0$ pont szigorú lokális minimumhelye, hiszen bármely origón átmenő v irányú egyenesnek van olyan, az origót tartalmazó szakasza, amely nem metszi az $y = x^2$ egyenletű parabolát a síkon.

A könnyebb megértés érdekében KÉSZÍTSÜNK ÁBRÁT!

35.5. A szélsőérték meghatározása

Egy n-változós függvény szélsőértékeinek meghatározásához tehát a következő lépéseket kell elvégeznünk:

1. Határozzuk meg a parciális deriváltakat
2. Mindegyik parciális deriváltat tegyük egyenlővé nullával, és oldjuk meg az így keletkező egyenletrendszert
3. Ezen helyek mindegyikén állítsuk elő a függvény Hesse-mátrixát
4. Ha egy adott helyen a Hesse-mátrix pozitív definit, akkor az a függvény lokális minimumhelye
5. Ha egy adott helyen a Hesse-mátrix negatív definit, akkor az a függvény lokális maximumhelye
6. Ha egy adott helyen a Hesse-mátrix indefinit, akkor itt a függvénynek nincs szélsőértéke
7. Ha egy adott helyen a Hesse-mátrix szemidefinit, akkor a feladat tételeink alapján nem oldható meg, az eset egyedi vizsgálatot kíván

35.6 Példa. Vegyük első példaként az

$$f(x,y) = x^4 + y^2$$

függvényt. Az egyetlen hely, ahol a parciális deriváltak nullák az origó. Ebben a pontban a Hesse-mátrix

$$H = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$$

amely nyilván pozitív szemidefinit. Világos azonban, hogy az origó a függvény (globális) minimumhelye.

Teljesen hasonló számítást hajthatunk végre, ha az

$$f(x,y) = -x^4 + y^2$$

függvényt vizsgáljuk. Ugyanúgy az origó az egyetlen kritikus pont, és a megfelelő Hesse-mátrix is azonos. Ekkor azonban az origó nem lehet szélsőértékhely, hiszen a függvény az origóhoz tetszőlegesen közel felvesz pozitív és negatív értékeket is.

Az ilyen pontot az f függvény nyeregpontjának nevezzük.

35.7 Példa. Keressük meg az alábbi függvény szélsőértékeit:

$$f(x,y,z) = (x^2 - 4y)e^{-(x+y+z^2)}$$

Ekkor az elsőrendű parciális deriváltakra az alábbi egyenletrendszert kapjuk:

$$\begin{aligned} f_1'(x,y,z) &= (2x - x^2 + 4y)e^{-(x+y+z^2)} = 0 \\ f_2'(x,y,z) &= (-4 - x^2 + 4y)e^{-(x+y+z^2)} = 0 \\ f_3'(x,y,z) &= -2z(x^2 - 4y)e^{-(x+y+z^2)} = 0 \end{aligned}$$

amelynek egyetlen megoldása $(x,y,z) = (-2,2,0)$

Alkalmazzuk ezután a másodrendű feltételünket. Ezen a helyen a Hesse-mátrix:

$$H = \begin{bmatrix} 6 & 4 & 0 \\ 4 & 4 & 0 \\ 0 & 0 & 8 \end{bmatrix}$$

A Hesse-mátrixnak megfelelő kvadratikus alak ezért

$$6x_1^2 + 8x_1x_2 + 4x_2^2 + 8x_3^2 = 2x_1^2 + 4(x_1 + x_2)^2 + 8x_3^2 \, .$$

Tehát a Hesse-mátrix pozitív definit, ezért a $(-2,2,0)$ pontban az f függvénynek lokális *minimuma* van.

35.6. A kétváltozós eset

Az elégséges feltételt megfogalmazó tételünk alapján könnyen adhatunk jól kezelhető feltételt lokális szélsőértékre a kétváltozós esetben. Ilyenkor ugyanis a definitség könnyen ellenőrizhető.

Ha $f : \mathbb{R}^2 \to \mathbb{R}$ a 35.4 Tétel feltételeit kielégítő függvény, továbbá $a \in \mathbb{R}^2$ olyan pont, ahol $f'(a) = 0$, akkor itt az f Hesse-mátrixa

$$f''(a) = \begin{bmatrix} f_{11}''(a) & f_{12}''(a) \\ f_{21}''(a) & f_{22}''(a) \end{bmatrix}$$

Tehát a Hesse-mátrix karakterisztikus polinomja

$$P(\lambda) = \lambda^2 - (f''_{11}(a) + f''_{22}(a))\lambda + f''_{11}(a)f''_{22}(a) - f''_{12}(a)^2$$

figyelembe véve a Hesse-mátrix szimmetriáját.

Ennek a másodfokú polinomnak csak valós gyökei vannak, hiszen a mátrix szimmetrikus. A gyökök és együtthatók összefüggése alapján tehát a következő állítást fogalmazhatjuk meg.

35.8 Tétel.

- *Ha* $f''_{11}(a)f''_{22}(a) - f''_{12}(a)^2 > 0$, *akkor az* $x = a$ *pont az* f *függvény lokális szélsőértékhelye. Ez a szélsőértékhely*
  - *lokális minimumhely, ha* $f''_{11}(a) > 0$.
  - *lokális maximumhely, ha* $f''_{11}(a) < 0$.
- *Nincs szélsőértékhely (nyeregpont), ha* $f''_{11}(a)f''_{22}(a) - f''_{12}(a)^2 < 0$.
- *Nem eldönthető, ha* $f''_{11}(a)f''_{22}(a) - f''_{12}(a)^2 = 0$.

Az utóbbi esetben a feladat további vizsgálatot igényel.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-1 III/2 szakasza kidolgozott példáinak feldolgozása.
2. Házi feladatok: a III/2 szakasz 2.2.2, 2.2.3, 2.2.4, 2.2.6, 2.2.8, 2.2.10, 2.2.11, 2.2.14, 2.2.15, 2.2.16 feladatai.
3. Tankönyv-1 17. fejezet.

36. fejezet

Legkisebb négyzetek módszere, regresszió

Ebben a fejezetben egy közelítő eljárást, a legkisebb négyzetek módszerét tárgyaljuk. Ez az eljárás a statisztika tananyagban szereplő regresszió matematikai alapja.

36.1. Legkisebb négyzetek módszere

Tegyük fel, hogy valamely kísérlet kimenetelére n számú megfigyelést végeztünk, és az $x_1, \ldots, x_n$ különböző helyeken az $y_1, \ldots, y_n$ értékek adódtak. Az az elképzelésünk, hogy ezekre a tapasztalati adatokra lineáris modell illeszthető, azaz egy olyan $y = mx + b$ egyenletű egyenest keresünk, amelyre

$$mx_1 + b = y_1 \qquad \ldots \qquad mx_n + b = y_n$$

Természetesen az adatok nem követik a mi hipotézisünket, ezért általában ilyen egyenes nem létezik. Ha ezt „mérési hibának” tudjuk be, és megelégszünk egy jó közelítéssel, akkor egy olyan egyenest keresünk, amely az adatainkat „jól” közelíti. Jó közelítésen azt értjük, hogy az

$$f(m, b) = \sum_{i=1}^{n} (mx_i + b - y_i)^2$$

négyzetösszeg minimális. Ezt a közelítő eljárást *legkisebb négyzetek* módszerének nevezzük.

FIGYELEM! Vajon miért nem szimplán az $mx_i - y_i$ eltérések összegét használjuk? Egyébként vizsgálhatnánk az $|mx_i - y_i|$ abszolút értékes eltérések összegét is. Ez teljesen korrekt lenne, de technikailag nagyon megnehezítené a számításokat.

36.2. Analitikus megoldás

Tekintsük tehát az

$$f(m,b) = \sum_{i=1}^{n} (mx_i + b - y_i)^2$$

függvényt, és adott $x_1, \ldots, x_n$ illetve $y_1, \ldots, y_n$ értékek mellett keressük meg m és b azon értékeit, amelyekre f minimális.

A minimumhelyre a parciális deriváltakból a

$$\begin{aligned} \frac{\partial f}{\partial m}(m,b) &= \sum_{i=1}^{n} 2x_i(mx_i + b - y_i) = 0 \\ \frac{\partial f}{\partial b}(m,b) &= \sum_{i=1}^{n} 2(mx_i + b - y_i) = 0 \end{aligned}$$

egyenletrendszer adódik. Innen a

$$\begin{aligned} \sum_{i=1}^{n} x_i y_i &= m\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} y_i &= m\sum_{i=1}^{p} x_i + bn \end{aligned} \tag{36.1}$$

lineáris egyenletrendszert kapjuk, amiből az m és b ismeretlenek már könnyen és egyértelműen meghatározhatók. Világos, hogy így minimumhoz jutunk, hiszen f teljes négyzetek összegeként áll elő.

Egyébként a másodrendű parciális deriváltak meghatározásával könnyen látható, hogy az f függvény Hesse-mátrixa állandó (m-től és b-től független), éspedig

$$f''(m,b) = \begin{bmatrix} \sum_{i=1}^{n} 2x_i^2 & \sum_{i=1}^{n} 2x_i \\ \sum_{i=1}^{n} 2x_i & 2n \end{bmatrix}$$

Világos, hogy a Hesse-mátrix pozitív definit szimmetrikus mátrix, hiszen a

$$\det(A - \lambda E) = \lambda^2 - \left(\sum_{i=1}^{n} 2x_i^2 + 2n\right) + 2n\sum_{i=1}^{n} 2x_i^2 - \left(\sum_{i=1}^{n} 2x_i\right)^2 = 0$$

egyenletnek csak pozitív gyökei lehetnek. Valóban, a számtani és négyzetes közepek közötti egyenlőtlenség alapján

$$\frac{1}{n}\sum_{i=1}^{n} x_i < \sqrt{\sum_{i=1}^{n} \frac{1}{n} x_i^2}$$

hiszen az x_i értékek nem mind azonosak. Innen adódik, hogy a fenti egyenlet konstans tagja pozitív.

36.3. Algebrai megoldás

Az előző szakaszban előállított minimumhelyet deriválás nélkül, pusztán algebrai eszözökkel is megkaphatjuk. Ha bevezetjük az

$$A = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}, \qquad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \qquad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \qquad z = \begin{bmatrix} m \\ b \end{bmatrix}$$

jelöléseket, akkor ellenőrizzük, hogy

$$Az - y = \begin{bmatrix} mx_1 + b - y_1 \\ \vdots \\ mx_n + b - y_n \end{bmatrix}$$

ezért az $f(m, b)$ függvény az

$$f(z) = \|Az - y\|^2$$

alakban írható fel. Olyan z vektort keresünk, amelyre az Az és y vektorok távolsága minimális. Más megfogalmazásban: keressük az $\operatorname{im} A$ altér azon elemét, amely az y vektorhoz legközelebb van. Ez a távolság nyilván pontosan akkor minimális, ha az $Az - y$ vektor ortogonális az $\operatorname{im} A$ altérre.

Ez azt jelenti, hogy az $\mathbb{R}^2$ mindkét e_i bázisvektorára

$$\langle y - Az, Ae_i \rangle = 0$$

Innen egyszerű átalakítással az

$$\left\langle A^T y, e_i \right\rangle = \left\langle A^T Az, e_i \right\rangle$$

egyenlet adódik $i = 1, 2$ mellett, amiből

$$A^T y = A^T Az \; .$$

Itt $A^T A$ nyilván invertálható, hiszen 2 rangú 2×2-es mátrix. Következésképpen

$$\begin{bmatrix} m \\ b \end{bmatrix} = z = \left(A^T A\right)^{-1} A^T y \; .$$

A kijelölt műveletek elvégzésével könnyen ellenőrizhető, hogy így is a (36.1) alatti egyenletrendszer megoldásához jutottunk. Gyakorlásképpen végezzük el ezt a számítást!

36.4. Regresszió

Legyenek X és Y valószínűségi változók, ahol az X értékkészlete az $\{x_1, \ldots, x_n\}$ halmaz. Tegyük fel, hogy az $x_1, \ldots, x_n$ pontokban az Y változó megfigyelt feltételes várható értékei

$$y_i = E(Y|X = x_i)$$

ahol $P(X = x_i) \neq 0$ minden $i = 1, \ldots, n$ esetén.

A kapott (x_i, y_i) pontokat a síkon ábrázolva lehet olyan elképzelésünk, hogy ezek közelítőleg egy adott függvény grafikonjára illeszkednek. Ha például ez a függvény az

$$y = mx + b$$

egyenes, akkor az m és b paramétereket úgy kívánjuk megválasztani, hogy ez a közelítés a lehető legjobb legyen, abban az értelemben, hogy az

$$E\left((mX + b - Y)^2\right)$$

várható érték minimális. Tekintsük tehát a

$$\begin{aligned} g(m, b) &= E\left((mX + b - Y)^2\right) \\ &= E\left(m^2X^2 + 2bmX + b^2 - 2mXY - 2bY + Y^2\right) \\ &= m^2E(X^2) + 2bmE(X) + b^2 - 2mE(XY) - 2bE(Y) + E(Y^2) \end{aligned}$$

függvényt. Ekkor a parciális deriváltakra a

$$\begin{aligned} \frac{\partial g}{\partial m}(m, b) &= 2mE(X^2) + 2bE(X) - 2E(XY) = 0 \\ \frac{\partial g}{\partial b}(m, b) &= 2mE(X) + 2b - 2E(Y) = 0 \end{aligned}$$

lineáris egyenletrendszer adódik. Ennek az egyenletrendszernek a megoldása

$$m = \frac{E(XY) - E(X)E(Y)}{E(X^2) - E(X)^2} = \frac{Cov(X, Y)}{Var(X)}$$

illetve

$$b = E(Y) - \frac{Cov(X, Y)}{Var(X)}E(X)$$

Könnyen látható, hogy ilyen módon valóban minimumhelyet kapunk, hiszen g olyan másodfokú függvény, amelyben a négyzetes tagok együtthatói pozitívak.

FIGYELEM!

Ellenőrizzük, hogy a g függvény Hesse-mátrixa pozitív definit! (És egyébként független az m és b változóktól.)

Az így nyert $y = mx + b$ függvényt (lineáris) *regressziós függvénynek* nevezzük. A statisztikában egyébként más (pl. kvadratikus vagy komplikáltabb) regressziós függvényeket is használnak.

Otthoni tanuláshoz

1. A Feladatgyűjtemény-2 V/4 szakasza feladatainak feldolgozása.
2. Házi feladatok: Feladatgyűjtemény-2, 515, 526 és 528 feladatai.
3. Tankönyv-1, 17. fejezet, és Tankönyv-2, 7. fejezet.

Irodalmi utalások, kitekintés

Természetesen tisztában vagyunk azzal, hogy egy három féléves matematika tárgyba jelen tananyagnál többet nem célszerű bepréselni. Ennek ellenére minden tanévben számos olyan érdeklődő hallgatónk akad, akik örömmel hallanának további érdekes feladatokról, nehezebb problémákról, és akik szívesen vállalnának még komolyabb kihívásokat. Számukra készült az alábbi útmutató, ahol megtalálhatják azokat a forrásokat, amelyek feldolgozása során kipróbálhatják magukat.

Analízis, többváltozós függvények: Walter Rudin: A matematikai analízis alapjai, Typotex, Budapest, 2012. (Talán a mai napig a világ legjobb klasszikus analízis tankönyve, nehéz és érdekes feladatokkal.)

Lineáris algebra: Dancs István és Puskás Csaba: Vektorterek, Aula Kiadó, 2005. (Az egyik legjobb magyar nyelvű lineáris algebra tankönyv, kifejezetten a matematikus-közgazdász hallgatók szempontjai szerint szerkesztve.)

Valószínűségszámítás: Medvedev Péter: Bevezetés a valószínűségszámításba. Letölthető: http://medvegyev.uni-corvinus.hu/kisvalszam.pdf (Szép, és nagyon igényes bevezetés, számos izgalmas közgazdasági, pénzügyi alkalmazással.)

Valószínűségszámítás: Kai Lai Chung: Elementary Probability Theory with Stochastic Processes, Springer-Verlag, New York, Heidelberg, Berlin, 1979. (Örök klasszikus, gyönyörű tárgyalásmód, gondolkodtató feladatok. Fontos magyar vonatkozása, hogy a szerző Pólya György diákja volt a Berkeley egyetemen.)

E könyv szerzője szívesen segít hallgatóinak a források beszerzésében, és nagy örömmel nyújt segítséget azok megértésében, feldolgozásában.