MÉRÉSI ADATOK KEZELÉSE ÉS ÉRTÉKELÉSE

Környezettudományi alapok tankönyvsorozat

A környezettan alapjai

A környezetvédelem alapjai

Környezetfizika

Környezeti áramlások

Környezeti ásványtan

Környezeti mintavételezés

Környezetkémia

Környezetminősítés

Környezettudományi terepgyakorlat

Mérések tervezése és kiértékelése

Talajtan környezettanosoknak

Enviromental Physics Methods Laboratory Practices

Eötvös Loránd Tudományegyetem
Természettudományi Kar

MÉRÉSI ADATOK KEZELÉSE ÉS ÉRTÉKELÉSE

Írta:
Havancsák Károly
egyetemi docens, Fizikai Intézet

Lektorálta:
Kardon Béla

TYPOTEX

2012

Lektorálta: Dr. Kardon Béla

ISBN 978-963-279-548-5
KÉSZÜLT: a Typotex Kiadó gondozásában
FELELŐS VEZETŐ: Votisky Zsuzsa

TÁMOGATÁS:
Készült a TÁMOP-4.1.2-08/2/A/KMR-2009-0047 számú, „Környezettudományi alapok tankönyvsorozat” című projekt keretében.

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

KULCSSZAVAK:

valószínűség, statisztika, mérési adatok, eloszlások, legkisebb négyzetek módszere, hisztogram, relatív gyakoriság, valószínűségi változó, várható érték, szórás, korreláció, normális eloszlás, nagy számok törvénye, mintavétel, empirikus jellemzők, becslési módszerek, kombinatorika, halmazelmélet

ÖSSZEFOGLALÁS:

Ebben a tankönyvben a véletlen jelenségek kezelésének alapjaival ismerkedhet meg az olvasó. A tankönyv fő részei: a mérési adatok leíró jellemzése a leíró statisztika módszereivel, a valószínűség-számítás eredményeinek alkalmazása a mérési adatok tulajdonságainak mélyebb megértése érdekében, a matematikai statisztika módszereinek segítségével, nagyszámú sokaság jellemzése kisebb számú mérési adat felhasználásával. Az anyag feldolgozása során a halmazelmélet és a kombinatorika fogalmaira és összefüggéseire is szükség van, ezért a függelékben e két témakör legfontosabb ismeretei is megtalálhatók. A tananyag feldolgozása során mindig az alkalmazhatóság a fő szempont, hiszen a tankönyv környezettudomány szakos hallgatóknak készült, akik a statisztikának nem művelői, hanem felhasználói lesznek. Ugyanakkor a matematika egy ágáról lévén szó, a szerző felhasználja a matematika jól bevált jelölésrendszerét, törekszik a szabatos fogalmazásra, és az esetek többségében az állítások (tételek) bizonyítását is megadja.

TARTALOMJEGYZÉK

A TANKÖNYV TARTALMÁRÓL

Bevezetés

A klasszikus fizika és egyéb klasszikus tudományok tanulmányozása során hozzászokunk ahhoz, hogy a jelenségek valamilyen meghatározó ok hatása alatt állnak, és ezért a folyamatok kimenetele egyértelműen meghatározott. Az ilyen folyamatokat *determinisztikus* (meghatározott) folyamatoknak nevezzük. Ha a determinisztikus folyamattal kapcsolatos kísérletet végzünk, akkor a folyamat mindig azonos módon megy végbe, és a kísérlet végeredménye mindig azonos lesz. Klasszikus példa az ilyen folyamatokra a szabadesés, amelynek végeredményét a Newton mozgástörvényei egyértelműen megadják. Persze ilyen esetekben is vannak zavaró körülmények, a klasszikus fizikának azonban az a módszere, hogy eltekint ezektől a zavaró körülményektől, amit azért lehet megtenni, mert ezek a hatások lényegesen kisebbek a jelenség lefolyását meghatározó fő hatásnál, és ezáltal csak kissé befolyásolják az eredményt.

Mindazonáltal nem minden folyamat ilyen. Vannak olyan folyamatok, amelyeknek végeredménye nem egyetlen meghatározott állapot, hanem több, esetleg végtelen sok lehetséges kimenetel közül az egyik. Az ilyen folyamatokra mindenki által ismert példa a szabályos játékkocka dobása, melynek hat lehetséges kimenetele van. Valahányszor feldobjuk a kockát, miután leesik, a kocka felső lapján hat szám közül az egyiket látjuk. Ha a kockadobást kísérletnek tekintjük, akkor azt mondhatjuk, hogy a kísérletnek hat lehetséges kimenetele van. Az ilyen kísérleteket, amelyeknek több lehetséges kimenetele van, és a kísérlet során ezek közül az egyik valósul meg, **véletlen** (vagy sztochasztikus) **kísérletnek** nevezzük. A kockadobáson kívül számos más példa is hozható a véletlen kísérletre. Ha egy érmét feldobunk, akkor az eredmény vagy fej, vagy írás lesz, azaz ennek a kísérletnek két lehetséges kimenetele van. Ha *90* szám közül ötöt kihúzunk (lottósorsolás), akkor a lehetséges kimenetelek száma *43 949 268*, amint az a kombinatorika módszereivel könnyűszerrel kiszámolható. Az eddig említett véletlen kísérletekben a végeredmény egy diszkrét sokaság értékei közül az egyik. Ha azonban például a klímaváltozás hatására vagyunk kíváncsiak, és mérjük a napi középhőmérséklet alakulását, akkor elvileg 0 K fok felett akármilyen értéket mérhetünk. A gyakorlatban természetesen szűkebb tartományban lévő értékeket mérünk, de mindenképpen egy folytonos sokaság közül kerül ki a mért hőmérséklet értéke.

Mi a különbség a determinisztikus és a véletlen folyamatok között? A véletlen kísérlet megnevezés semmi esetre sem jeleneti azt, hogy a véletlen folyamatoknak ne lenne oka. Azonban, míg a determinisztikus folyamatok esetén van egy meghatározó ok, ami megszabja a folyamat lefolyását, addig a véletlen folyamatok esetén több, sok esetben nagyon sok, közel egyenértékű ok vezet arra, hogy a végeredmény nem egyetlen jól meghatározott esemény. A determinisztikus folyamatra már említett példa a szabadesés, vagy a Föld keringése a Nap körül, ahol mindkét esetben a gravitációs erő a folyamatot meghatározó hatás. Az érme feldobásakor számos együttesen fellépő hatás az, amely meghatározza, hogy mi lesz a kísérlet kimenetele: az érmének adott felfelé irányuló kezdősebesség, a forgást előidéző forgatónyomaték, az oldalirányú kezdősebesség, a légmozgások stb.

Amikor a fentiekben azt állítottuk, hogy léteznek determinisztikus kísérletek, akkor a meghatározó hatás mellett ezekben a kísérletekben is jelenlévő, kicsiny hatásoktól eltekintettünk. Az elméleti megfontolások során az elhanyagolások helyénvalóak, és nagyon sikeresen vezetnek az alapjelenségek leíró egyenletek megtalálásához. Amikor azonban méréseket végzünk, akkor természetesen a kicsiny hatások is jelen vannak, aminek eredményeképpen a determinisztikusnak nevezett kísérletek eredménye, ha kismértékben is, de változó lesz. Ezt a jelenséget véletlen (statisztikus) mérési hibának nevezzük, és minden mérési folyamatban jelen van. Így, bár feltételezzük, hogy a mérendő mennyiségnek van meghatározott értéke, ez az érték soha nem mérhető meg teljes pontossággal, legfeljebb a statisztika módszereivel jó közelítéssel becsülhető.

Még inkább ilyen a helyzet az atomok, elemi részecskék világában, ahol a kvantummechanika törvényei írják le a jelenségek lefolyását. A kvantummechanika törvényei valószínűségi jellegűek. A mérések eredménye a lehetséges eredmények közül az egyik. Az elmélet a kimenetelek valószínűségét adja meg. Ha például azt mérjük, hogy egységnyi tömegű radioaktív anyag atomjai közül egységnyi idő alatt mennyi bomlik el, akkor az ismételt mérések során más-más eredményt kapunk. A kvantumtörvények lényegüknél fogva statisztikus jellegűek.

A fenti gondolatmenet során valami olyasmire jutottunk, hogy a determinisztikus folyamatok tulajdonképpen csak elméleti konstrukciók, és valójában, ha méréseket végzünk, akkor ilyen vagy olyan okok miatt, de mindig véletlen kísérlettel van dolgunk. A véletlen jelenségek vizsgálatával a statisztika és a valószínűség-számítás foglalkozik. A fenti bevezető sorok talán rávilágítottak arra, hogy e tudományágak eredményei nagy jelentőségűek a kísérletek tervezése és az eredmények értékelése során.

Ebben a tankönyvben a véletlen jelenségek kezelésének alapjaival fogunk megismerkedni. A tankönyv fő részei: a mérési adatok leíró jellemzése a leíró statisztika módszereivel, a valószínűség-számítás eredményeinek alkalmazása a mérési adatok tulajdonságainak mélyebb megértésére, a matematikai statisztika módszereinek segítségével nagyszámú sokaság jellemzése kisebb számú mérés felhasználásával. Minthogy az anyag feldolgozása során használjuk a halmazelmélet és a kombinatorika fogalmait és összefüggéseit, ezért a függelékben e két témakör legfontosabb ismereteit is összefoglaljuk.

A tananyag feldolgozása során mindig az alkalmazhatóságot tartjuk szem előtt, hiszen e tankönyv környezettudomány szakos hallgatóknak készül, akik a statisztikának nem művelői, hanem felhasználói lesznek. Ugyanakkor nem feledjük, hogy a matematika egy ágáról van szó, tehát felhasználjuk a matematika jól bevált jelölésrendszerét, törekszünk a szabatos fogalmazásra, és az esetek többségében az állítások (tételek) bizonyítását is megadjuk.

Történelmi áttekintés

A leíró statisztika tulajdonképpen régóta használatos eszköz nagyszámú adat tömörítésére és egyszerű kezelésére. Már az ókorban is voltak népszámlálások, egy-egy birodalom földművelésével, állatállományával kapcsolatos felmérések, ahol nagyszámú adatot kellett kezelni. A statisztika szó is a latin *status* (állam, állapot) szavakból ered. A valószínűség-elmélet *(sztochasztika)* a nagyszámú minta elemzése során tapasztalt törvényszerűségek absztrakt matematikai kezelésével foglalkozó tudományág. A sztochasztika szó görög eredetű (στοχος=ügyes találgatás, sejtés).

A valószínűséggel kapcsolatos matematikai megfontolásokkal a 15–16. században találkozunk először. A szerencsejátékok már akkor is rendkívül népszerűek voltak. A kockajátékokkal kapcsolatos, esetenként fogós kérdésekkel a kor ismert matematikusaihoz fordultak. Úgy tartják, hogy a valószínűség-számítás egyes kérdéseire Pascal figyelmét egy híres szerencsejátékos, de Méré lovag hozzá intézett kérdése fordította. A kérdés úgy hangzott, hogy miért valószínűbb, hogy egy kockával négyszer dobva legalább egyszer hatost dobunk, mint két kockával 24-szer dobva legalább egyszer dupla hatost dobni? Paradoxonnak tűnik a kérdés, hiszen a dupla hatosnak hatodannyi az esélye, mint az egyszeri hatos dobásnak, és a 24 éppen a 4 hatszorosa! A problémával Pascal (1623–1662) és Fermat (1600–1665) egyaránt foglalkozott, és különböző módszerekkel azonos eredményre jutottak. A hagyomány szerint a valószínűség matematikai megközelítése ennek a problémának a megoldásával kezdődhetett. A későbbiek során megoldjuk majd ezt a feladatot. Pascal és Fermat eredményeinek megismerését követően Huygens (1629–1695) is foglalkozott a valószínűség kiszámításának problémáival, és Pascal bátorítására könyvet is írt a valószínűség elméletéről. A kor eredményeit Jacob Bernoulli (1654–1705) foglalta össze *Ars Conjectandi* (A sejtés művészete) című könyvében. A 18. században a valószínűség-számítás már a gazdasági életben is fontos szerepet játszott. Életjáradékokkal és biztosítással kapcsolatos kérdésekben alkalmazták az eredményeit. A tudományban ekkor dolgozták ki a statisztikus gázelméletet, melynek során szintén a valószínűségelmélet eredményeit használták fel. A legfontosabb eredmények Laplace (1749–1825), Poisson (1781–1840), Bayes (1702–1761) és Gauss (1777–1855) nevéhez fűződnek. Gauss foglalkozott például a hibaszámítás elméletének kidolgozásával.

A 19. század második felében az orosz valószínűségi iskola nagyjai, Csebisev (1821–1894), Markov (1856–1922), Ljapunov (1857–1918) értek el jelentős eredményeket. A 20. század első felében a természettudományok, elsősorban a fizika forradalmi fejlődésen ment keresztül. Ebben a folyamatban jelentős mértékben alkalmazták a valószínűség-számítás korábban elérte eredményeit. Ugyanakkor a műszaki tudományok, a technika és a gazdaság fejlődése újabb és újabb alkalmazási területeket jelentettek. Az atomelmélet, a kvantummechanika, a telefonközpontok fejlesztése, a népesedési problémák, a genetika eredményei újszerű alkalmazási problémákat vetettek fel. A valószínűség-számítás új alapokra helyezése elkerülhetetlenné vált. Ezt a munkát Kolmogorov, orosz matematikus (1903–1987) végezte el, aki axiomatikus alapokra helyezte a valószínűségelméletet. Ennek eredményeképpen megszűnt az a korábbi bizonytalanság, amit a megfelelő alapok hiánya okozott. A valószínűségelmélet és a statisztika a tudományok rendkívül hasznos eszközévé válhatott.

I. A MÉRÉSI ADATOK LEÍRÓ JELLEMZÉSE

1. A MÉRÉSI ADATOK KEZELÉSE

1.1. Mérési adatok megjelenítése

Olyan jelenségekkel foglalkozunk tehát, amelyekkel kapcsolatban, ha méréseket végzünk, akkor általában különböző eredményeket kapunk. Felmerül a kérdés, hogy ha ilyen bizonytalan egy mérés eredménye, akkor tudományosan egyáltalán kezelhető-e ez a helyzet? A kérdés jogos, ugyanakkor van olyan tapasztalat, ami reménnyel tölthet el bennünket. Ha a véletlen jelenségekkel kapcsolatban nem egy, hanem több mérést végzünk, akkor felfigyelhetünk olyan szabályszerűségre, amely alapot adhat a kérdéskör tudományos kezelésére. Lássunk egy egyszerű példát! Ha egy érmét feldobunk, akkor kétféle végeredmény születhet: fej vagy írás. Az ilyen kísérletben hallgatólagosan mindig feltesszük, hogy az érme szabályos, tehát a kísérlet során egyforma eséllyel lehet fej vagy írás a végeredmény. Nézzük meg, hogy sokszor elvégezve a kísérletet, mit tapasztalunk? Legyen a kísérletek száma n. Az n kísérlet során a fejek száma legyen k_{fej}, az írásoké $k_{írás}$, ezek az n kísérlet során az adott esemény **gyakoriságát** mutató értékek. Az érmés kísérletben természetesen

$$k_{fej} + k_{írás} = n\,.$$

A tapasztalat az, hogy ha elég nagyszámú kísérletet végzünk, akkor az írások és a fejek gyakorisága közel azonos lesz, vagyis

$$\frac{k_{fej}}{k_{írás}} \approx 1\,.$$

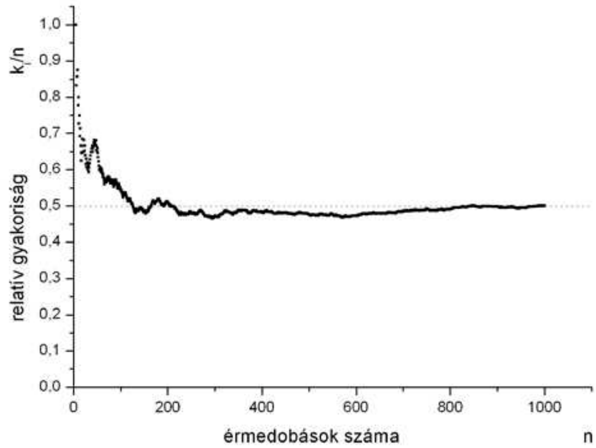
A nagyszámú kísérlet során szerzett kísérleti tapasztalatot kissé alaposabban is megvizsgáljuk. Ha a több kísérlet során a gyakoriság viselkedését akarjuk tanulmányozni, célszerű a

$$g_i = \frac{k_i}{n},$$

az ún. relatív gyakoriság vizsgálata, ahol i a lehetséges végeredmények közül az egyik. A relatív gyakoriság azt mutatja meg, hogy az n kísérlet során milyen arányban fordult elő az egyik lehetséges végeredmény. Könnyű belátni, hogy igazak az alábbi összefüggések:

$$0 \le k_i \le n; \text{ és } 0 \le \frac{k_i}{n} \le 1\,. \tag{1.1.1.}$$

Az 1.1. ábrán érmés kísérlet során a fej relatív gyakoriságának változást látjuk a kísérletszám függvényében, egészen *n=1000* kísérletig. Az ábrán az látszik, hogy ameddig a kísérletek száma kicsi, addig a relatív gyakoriság *0* és *1* között akármilyen értéket felvehet. Ahogyan azonban nő a kísérletek száma, a relatív gyakoriság értéke egyre kevésbé ingadozik, és nagy *n* értékekre állandó érték felé tart, ami jelen esetben *1/2*. Tehát, ahogyan a kísérletek száma nő, elegendően nagy *n* érték mellett a relatív gyakoriság stabilitást mutat. Más véletlen kísérlet kapcsán is hasonló stabilitást tapasztalnánk, esetleg másik érték körül. Ez a tapasztalat a **kísérleti nagy számok törvénye**. Erre a kísérleti tapasztalatra alapozódik a valószínűség-elmélet, és a későbbiek során visszatérünk még erre az eredményre.



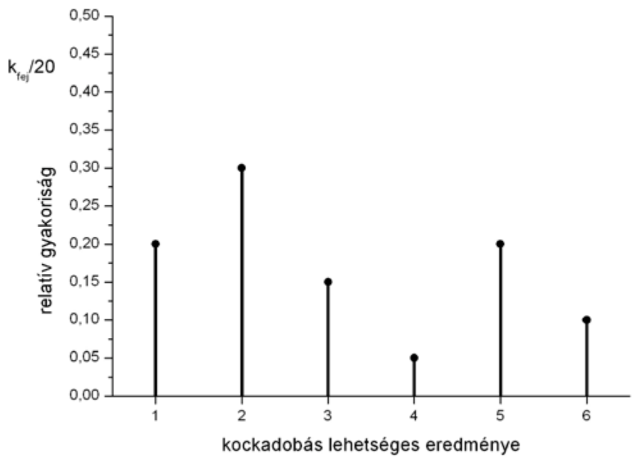
1.1. ábra: Érmedobások során a fej relatív gyakoriságának változása a kísérletszám függvényében

Foglalkozzunk most azzal a kérdéssel, hogy mérési adatainkat hogyan rögzítsük, és hogyan jelenítsük meg. Azt már láttuk, hogy véletlen kísérlet esetén nem elegendő egyetlen mérést végezni. Általában több, sokszor nagyon sok adattal van dolgunk. Ezeket az adatokat célszerű már az adatgyűjtés idején táblázatba foglalni. Ilyen *20* adatból álló adatsort látunk az 1.1. táblázat: Kockadobás eredménye n=20 kísérlet során, ahol a kockadobások során rögzítettük a kapott eredményeket, és gyakorisági táblázatot készítettünk.

lehetséges kimenet	**1**	**2**	**3**	**4**	**5**	**6**
gyakoriság k_{fej}	*4*	*6*	*3*	*1*	*4*	*2*
relatív gyakoriság $k_{fej}/20$	*0,20*	*0,30*	*0,15*	*0,05*	*0,20*	*0,10*

1.1. táblázat: Kockadobás eredménye n=20 kísérlet során

Az adatokat ábrán is szemléltethetjük. Az 1.2. ábra a táblázat relatív gyakoriság adatait mutatja. A vízszintes tengelyre a lehetséges kimenetek diszkrét értékeit rajzoltuk. A függőleges tengelyen pedig a relatív gyakoriság értékeket tüntettük fel. Amikor a véletlen kísérlet lehetséges kimenetelei diszkrét értékek, akkor az eredményeket gyakran ilyen, ún. pálcikaábrán szemléltetjük, ahol a pálcika hossza az adott kimenetel relatív gyakoriságát mutatja.



1.2. ábra: Kockadobás relatív gyakorisága 20 kísérlet során

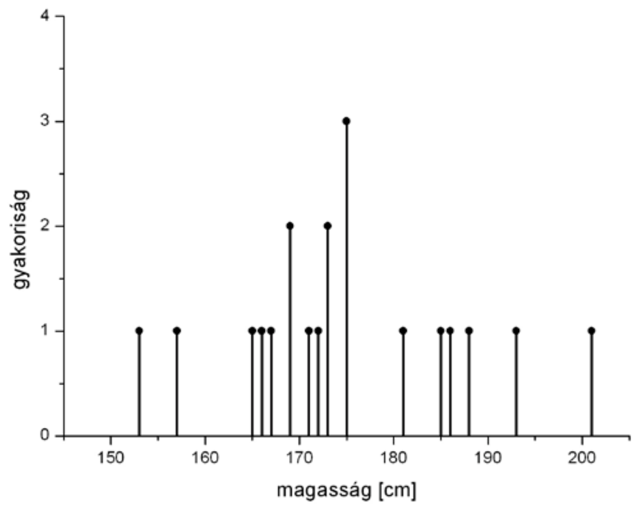
Kissé más a helyzet, amikor a lehetséges eredmények folytonos számhalmaz elemei lehetnek. Ilyen adatokat tartalmaz az 1.2. táblázat:, ahol *20* felnőtt magasságadatait láthatjuk.

sorszám	magasságadatok [cm]	rendezett magasságadatok [cm]
1	153	153
2	201	157
3	187	165
4	167	166
5	173	167
6	175	169
7	181	169
8	157	171
9	169	172
10	175	173
11	165	173
12	173	175
13	185	175
14	172	175
15	193	181
16	188	185
17	166	186
18	169	188
19	174	193
20	171	201

1.2. táblázat: Magasságadatok

A táblázat második oszlopa a nyers adatokat tartalmazza. A jobb áttekinthetőség érdekében a mérések elvégzése után célszerű nagyság szerint sorrendbe szedni az adatokat. A táblázatban ez a harmadik oszlopban látszik. Ilyen sorrendben könnyen felfedezhető, hogy vannak adatok, amelyek többször szerepelnek.

Próbáljuk meg pálcikaábrán ábrázolni az adatokat! Ezt mutatja az 1.3. ábra. Az ábra vízszintes tengelyén a magasság értékek, a függőleges tengelyén pedig egy-egy magassági érték gyakorisága szerepel.



1.3. ábra: Magasságadatok gyakorisága

Mivel a magasságadatok folytonosan helyezkednek el a számegyenesen, ezért gyakori az, hogy egy érték csak egyszer, vagy csak néhányszor szerepel. Ezért a gyakoriság sokszor csak *1*, és a gyakoriság ábra ilyen formában nem túl informatív. Az információt inkább az hordozza, hogy hol helyezkednek el sűrűn az adatok.

1.2. Hisztogram

Az előzőekben mondottak értelmében folytonos esetben nem a pálcikaábra a célravezető, hanem célszerű az adatok sűrűségét ábrázolni. De haladjunk sorjában! Első lépésként az adatokat osztályokba kell gyűjteni. Ez a jelen esetben azt jelenti, hogy az ésszerű módon kijelölt magasságtartományt intervallumokra osztjuk, és megszámoljuk az intervallumokba jutó magasságadatok számát (gyakoriságát). *200* magasságadatot tartalmazó osztályokra osztott adatsort tartalmaz az 1.3. táblázat:.

A táblázat első oszlopa az osztály i sorszámát jelöli. A táblázat második oszlopa az **osztályhatárokat** (x_i; x_{i+1}), a harmadik oszlop az osztályhatárok számtani közepét, az ún. **osztályközepet** ($(x_i+x_{i+1})/2$), a negyedik oszlop pedig az osztályba eső mérési adatok számát, azaz a **gyakoriságot** (k_i) mutatja. Megállapodhatunk abban, hogy ha egy adat az osztályhatárra esik, akkor a nagyobbik osztályba soroljuk. Az ilyen táblázat adatait általában **oszlopdiagrammal** ábrázoljuk, ahogyan ez a 1.4. ábrán látszik. Az oszlopdiagram hézagmentesen egymás mellé helyezett téglalapokból áll. A téglalapok szélessége megegyezik az osztályok szélességével, magassága pedig a gyakoriság, vagy a relatív gyakoriság értékével. A téglalap középvonala az osztályközép értékével esik egybe. A statisztikában az oszlopdiagramot **hisztogramnak** nevezik.

i **sorszám**	$(x_i;\ x_{i+1})$ **osztályhatárok [cm]**	$(x_i+x_{i+1})/2$ **osztályközép [cm]**	k_i **gyakoriság**	$\frac{k_i}{n}$ **relatív gyakoriság**	$\frac{k_i}{n\Delta x_i}$ **relatív gyakoriság sűrűség [1/cm]**	$\sum_{i=1}^{j}\frac{k_i}{n\Delta x_i}$ **kumulatív relatív gyakoriság [1/cm]**
1	(115; 125)	120	1	0,005	0,0005	0,0005
2	125; 135)	130	5	0,025	0,0025	0,0030
3	(135; 145)	140	3	0,015	0,0015	0,0045
4	(145; 155)	150	17	0,085	0,0085	0,0130
5	(155; 165)	160	47	0,235	0,0235	0,0365
6	(165; 175)	170	62	0,310	0,0310	0,0675
7	(175; 185)	180	41	0,205	0,0205	0,0880
8	(185; 195)	190	16	0,080	0,0080	0,0960
9	(195; 205)	200	6	0,030	0,0030	0,0990
10	(205; 215)	210	1	0,005	0,0005	0,0995
11	(215; 225)	220	1	0,005	0,0005	0,1000
12	(225; 235)	230	0	0,000	0,0000	0,1000

1.3. táblázat: Osztályokba rendezett magasságadatok

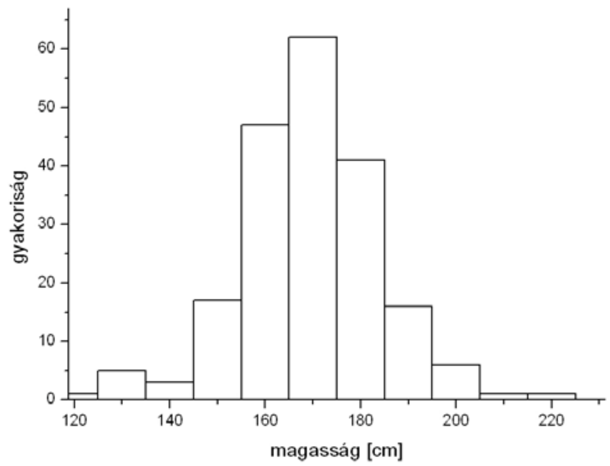
Ha a függőleges tengelyre a relatív gyakoriságot rajzoljuk, akkor ennek értéke a korábban mondottak értelmében:

$$g_i = \frac{k_i}{n}, \qquad i=0,\ 1,\ 2\ \dots m, \tag{1.2.1.}$$

ahol n az összes mért adatok száma, m pedig az osztályok száma. Természetesen igaz az, hogy

$$k_1 + k_2 + \dots + k_m = \sum_{i=1}^{m} k_i = n\,,$$

hiszen az egyes osztályokban elhelyezkedő gyakoriságok összege éppen a mérések számát adja.



1.4. ábra: 200 mérési adatot tartalmazó hisztogram

Az 1.4. ábrán az osztályközök azonos szélességűek. Ez nem kötelező, lehetnek különböző szélességű osztályok is. Ilyenkor azonban, hogy az ábra arányai ne torzuljanak, a függőleges tengelyre a **gyakoriság sűrűség** értékét, vagy a **relatív gyakoriság sűrűség** értékét rajzoljuk, vagyis az intervallum Δx_i hosszával elosztjuk a gyakoriság, vagy a relatív gyakoriság értékét. Tehát gyakoriság esetén a gyakoriság sűrűség h_i értéke:

$$h_i = \frac{k_i}{\Delta x_i}, \quad i=0,\ 1,\ 2\ ...m, \tag{1.2.2.}$$

a relatív gyakoriság esetén pedig a relatív gyakoriság sűrűség f_i értéke:

$$f_i = \frac{k_i}{n\Delta x_i}, \quad i=0,\ 1,\ 2\ ...m. \tag{1.2.3.}$$

Látszik, hogy ilyen esetben a gyakoriság, illetve a relatív gyakoriság értékét nem az oszlop magassága, hanem az oszlop területe jellemzi, hiszen (1.2.2.) átrendezésével:

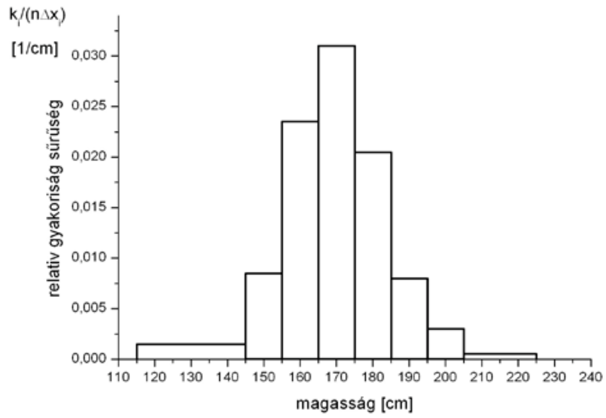
gyakoriság=oszlopmagasság · osztályszélesség,

illetve (1.2.3.) átrendezésével

relatív gyakoriság=oszlopmagasság · osztályszélesség.

Ezekben az esetekben **sűrűség hisztogramról** beszélünk. Ilyen sűrűség hisztogramot látunk a 1.5. ábrán, ahol az 1.4. ábrán látható adatokat relatív gyakoriság sűrűség diagramon ábrázoltuk. A görbe farkainál lévő osztályokat összevontuk, tehát az osztályközök most nem egyformák. Az adatsor azon részén célszerű szélesebb osztályközöket képezni, ahol kevesebb az adatok száma.

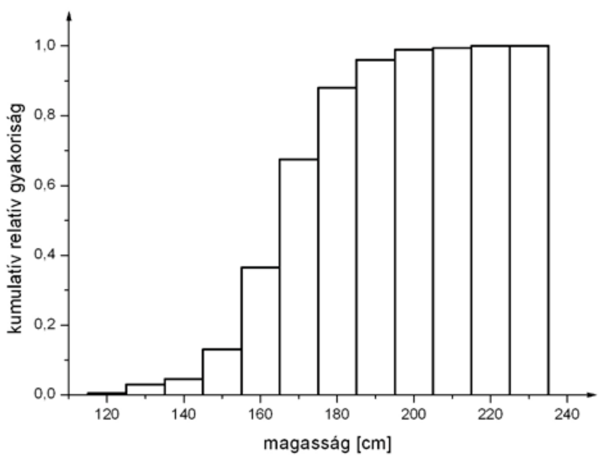
Könnyen belátható, hogy a relatív gyakoriság hisztogram görbe alatti területe egységnyi.



1.5. ábra: 200 ember magasságeloszlását ábrázoló hisztogram

1.3. Kumulatív gyakoriság

A statisztikában gyakorta az a kérdés, hogy adott értéknél kisebb adatok milyen gyakorisággal (relatív gyakorisággal) fordulnak elő a mérési adatok között. Ilyenkor beszélünk kumulatív adatokról. A kumulatív gyakorisági (relatív gyakorisági) görbét úgy szerkesztjük meg, hogy adott osztályközép fölé olyan magas téglalapot rajzolunk, hogy magassága megegyezzen az adott osztály és a megelőző osztályok gyakoriságának (relatív gyakoriságának) összegével. A korábbi példánkban a 1.3. táblázat: utolsó oszlopa tartalmazza a relatív gyakoriságok kumulatív értékét. A relatív gyakoriságok kumulatív értéke esetén a téglalapok magassága *0*-ról monoton növekszik, ameddig el nem éri az *1* értéket.



1.6. ábra: A magasságadatok kumulatív relatív gyakoriság görbéje

Gyors ellenőrző feladatok

1.1. Lássuk be, hogy valamennyi osztályra elvégezve a relatív gyakoriság sűrűség hisztogram téglalapjai területének összegzését, eredményül *1*-et kapunk. Ez azt jelenti, hogy a

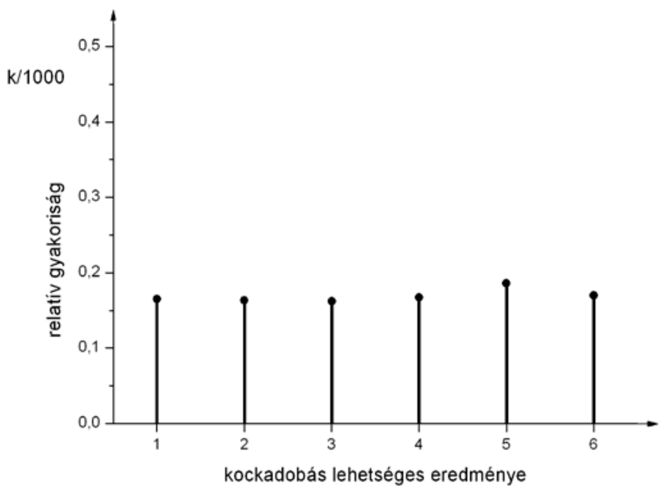
$$\sum_{i=1}^{m} \Delta x_i \frac{k_i}{n \Delta x_i} = 1$$

összefüggést kell igazolni.

1.2. Lássuk be, hogy az *m.* (jelen esetben az utolsó) osztály elérése esetén a kumulatív relatív gyakoriság oszlopmagassága egyenlő lesz *1*-el!

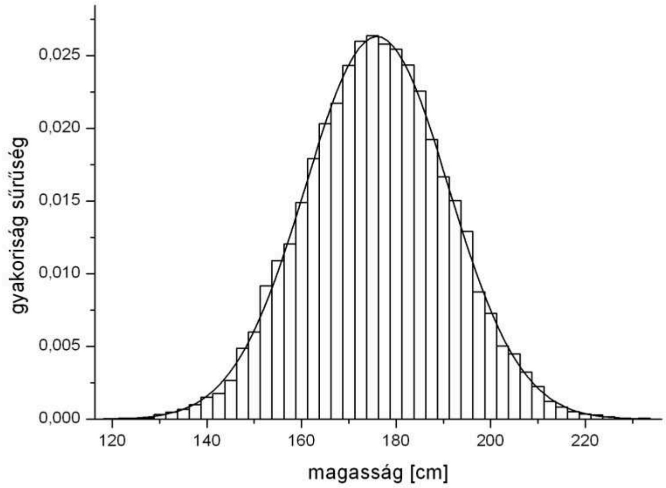
1.4. Relatív gyakoriság eloszlások

Már az 1.1. ábrán láttuk, hogy a kísérleti nagy számok törvénye értelmében nagyszámú kísérlet esetén a relatív gyakoriság stabilitást mutat. Ha elegendően nagyszámú kísérlet eredményét diszkrét eloszlás esetén, pálcikaábrákon rajzoljuk fel, akkor láthatjuk, hogy a relatív gyakoriság értékek hogyan oszlanak meg a lehetséges kísérlet kimenetelek között. A 1.7. ábrán 1000 kockadobás esetén a relatív gyakoriságok eloszlását rajzoltuk fel. (Ilyen kísérletet számítógépes szimulációval bárki könnyen elvégezhet.) Az ábrán az látszik, hogy szemben az 1.2. ábrán tapasztaltakkal, elegendően nagyszámú dobás esetén a különböző lehetséges kimenetelek relatív gyakoriságai egyenletesen oszlanak el. A várakozásunk is ez, hiszen ha a kocka szabályos, egyik oldal sincs kitüntetve.



1.7. ábra: Kockadobás relatív gyakoriságának eloszlása 1000 dobás esetén

Folytonos eloszlások esetén ezt a tulajdonságot jól tanulmányozhatjuk a gyakoriság sűrűség hisztogramon. Nézzük meg, hogyan változik meg az 1.4. ábrán látható hisztogram jellege, ha nem 200, hanem 1 000 adatból szerkesztjük meg a gyakoriság sűrűség hisztogramot. Az 1.8. ábrára rajzoltuk ezt a hisztogramot. Határozott tendenciát figyelhetünk meg az eloszlás menetében. Harang alakú eloszlást kapunk, amelyre akár folytonos függvényalakot is illeszthetünk. Ha tovább növeljük a mérések számát, a görbe már csak kismértékben változik, ami a relatív gyakoriság stabilitásának következménye. Az ábrán látható folytonos függvényalak nagyon jellegzetes, sok egymástól fizikailag különböző feladat esetében kapunk hasonló sűrűségeloszlás görbét. A görbét első alkalmazójáról Gauss-görbének nevezzük. A későbbiekben a Gauss-görbe matematikai alakját is megadjuk majd.



1.8. ábra: 1000 mérési adatból felrajzolt gyakoriság sűrűség hisztogram, és az illesztett folytonos görbe

Tulajdonképpen már kevesebb számú mérés esetén is felfedezhetjük a Gauss-alakú eloszlást. Lássunk egy egészen másfajta adatsort. Fizikatörténeti érdekessége van Michelson (1852–1931) 1879-ben végzett fénysebességmérésének. Michelson *100* mérést végzett. Adatait nagyság szerinti sorrendben a 1.4. táblázat:táblázat tartalmazza, ahol helytakarékosság miatt az adatoknak csak a 299 000 km/s feletti részét tüntettük fel.

	1	2	3	4	5	6	7	8	9	10
1	620	760	800	810	840	850	870	880	930	960
2	650	760	800	810	840	850	870	880	930	960
3	720	760	800	810	840	850	880	890	940	970
4	720	760	800	810	840	850	880	890	940	980
5	720	770	800	810	840	850	880	890	940	980
6	740	780	810	820	840	860	880	900	950	980
7	740	780	810	820	840	860	880	900	950	1000
8	740	790	810	830	850	860	880	910	950	1000
9	750	790	810	830	850	870	880	910	960	1000
10	760	790	810	840	850	870	880	920	960	1070

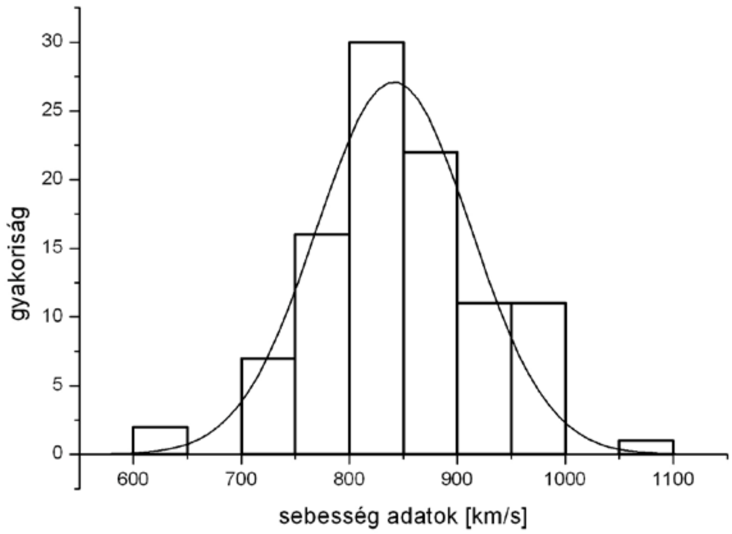
1.4. táblázat: Michelson fénysebesség mérési adatai. A táblázatbeli értéket km/s mértékegységűek, és a 299 000 km/s sebesség feletti értékeket mutatják

Az adatokat a következő osztályokba soroljuk: (600; 650), (650; 700), (700; 750), (750; 800), (800; 850), (850; 900), (900; 950), (950; 1000), (1000; 1050), (1050, 1100), ahol csak a 299 000 km/s feletti részt írtuk ki. Michelson osztályokba sorolt adatainak gyakoriság értékeit a 1.5. táblázat:táblázat mutatja.

osztályközép [km/s]	gyakoriság
299 625	*2*
299 675	*0*
299 725	*7*
299 775	*16*
299 825	*30*
299 875	*22*
299 925	*11*
299 975	*11*
300 025	*0*
300 075	*1*

1.5. táblázat: Michelson adatainak osztályba sorolása

A gyakoriság hisztogramot a 1.9. ábra mutatja. A vízszintes tengelyre most is csak a 299 000 feletti részt írtuk ki.

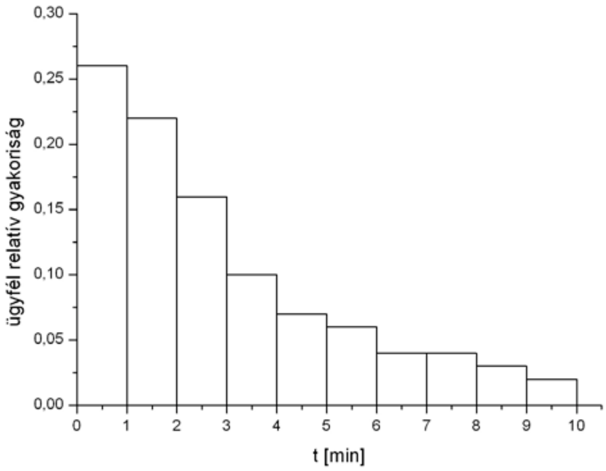


1.9. ábra: Michelson fénysebesség mérésének adataiból szerkesztett gyakoriság hisztogram

Ha a mérés abszolút pontosságú lenne (ilyen persze nem létezik) akkor minden egyes mérési adatnak azonosnak kellene lennie. Az adatok azonban szórnak, és ez a szórás a mérési hibával kapcsolatos. A mérési hibát a hosszmérés bizonytalansága, a hőmérséklet változása, a mérőberendezés forgó tükrének frekvenciabizonytalansága stb. okozták. Bár *100* mérés nem túl sok, de a hisztogramon már elég jól kirajzolódik, hogy a hibával terhelt adatok eloszlása a fentiekben megismert Gauss-alakú görbéhez közelít.

A Gauss-görbe típusú gyakoriság eloszlásokkal (sűrűségekkel) gyakran találkozunk, ennek okát az elméleti tárgyalás során majd elemezzük.

Nézzünk még egy példát, ahol a hisztogram alakja nem harang alakú görbe. Egy hivatalban megfigyeljük, hogy az ügyintéző mennyi ideig foglalkozik az ügyféllel. Megmérjük *100* ügyfél esetében az ügyintézés idejét. Az adatokat táblázatba rendezzük, és osztályokba soroljuk. Legyen az osztályszélesség *1 min*. Az 1.10. ábrán a mérés eredménye látszik relatív gyakoriság hisztogram formájában.



1.10. ábra: Az ügyfélszám relatív gyakoriság ügyintézés idejének függvényében 100 ügyfél esetén mérve

Az ábrán azonnal látszik, hogy az eloszlás maximum nélküli, és csökkenő tendenciájú. Az elméleti tárgyalás során fogjuk látni, hogy az időtartammal, élettartammal kapcsolatos folyamatok gyakorta ilyen ún. exponenciális eloszlással jellemezhetők.

Természetesen nagyszámú mérést követően a kumulatív gyakoriság görbék is felrajzolhatók, és tapasztalhatjuk, hogy elég nagy mérésszám esetén már lényegesen nem változik a görbe jellege.*

1.5. A mérési adatok egyszerűsített jellemzése

Méréseinket nem mindig akarjuk az összes adattal jellemezni. A gyors és egyszerű jellemzéshez az adatsort valamiképpen jellemző reprezentatív számértékre van szükség. Az, hogy milyen reprezentatív számértéket válasszunk az adatsor jellemzésére, a statisztika egyik fontos kérdése.

* Az itt letölthető szimuláció azt mutatja, hogy kockadobás esetén a relatív gyakoriság, a relatív gyakoriságok eloszlása és a kumulatív relatív gyakoriság hogyan változik, miközben a mérések száma növekszik.

Egy adatsor egyetlen számmal történő jellemzésére gyakran használjuk a középértéket, vagy átlagot. Többféle középérték létezik: számtani közép, mértani közép, harmonikus közép, négyzetes közép, medián, módusz. A feladat jellege dönti el, hogy melyik középérték jellemzi legjobban az adatsort. Az alábbiakban a középértékek használatával ismerkedünk meg.

1.6. Számtani közép

Leggyakrabban talán a számtani közepet használjuk.

Definíció. Legyen n darab mérési eredményünk:

$$x_1, x_2, x_3, \ldots, x_n .$$

Az n mérési eredmény $\overline{x}$ **számtani közepén** definíció szerint a következőt értjük:

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} . \tag{1.6.1.}$$

A definícióból leolvasható, hogy a számtani közép az a szám, amellyel ha az átlagolandó értékeket helyettesítjük, akkor az összegük változatlan marad.

Nézzünk egy konkrét példát! 10 darab kockadobás során a következő sorozatot kapjuk:

$$x_1 = 5;\ x_2 = 5;\ x_3 = 6;\ x_4 = 1;\ x_5 = 2;\ x_6 = 2;\ x_7 = 4;\ x_8 = 5;\ x_9 = 1;\ x_{10} = 2;$$

Az adatok számtani közepe:

$$\overline{x} = \frac{5+5+6+1+2+2+4+5+1+2}{10} = 3{,}3 . \tag{1.6.2.}$$

Mivel egy-egy szám többször is szerepel az adatsorban (kettő darab *1*-es, három darab *2*-es stb.), az átlagot másképpen is számíthatjuk:

$$\overline{x} = \frac{2 \cdot 1 + 3 \cdot 2 + 0 \cdot 3 + 1 \cdot 4 + 3 \cdot 5 + 1 \cdot 6}{10} = 3{,}3$$

Általánosíthatjuk is ezt a felismerést. Ha az azonos mérési eredmények gyakoriságát most is k_i-vel jelöljük, a különböző lehetséges eredmények számát N-nel, akkor az átlagszámolás képlete az alábbi lesz:

$$\bar{x} = \frac{k_1x_1 + k_2x_2 + k_3x_3 + \ldots + k_Nx_N}{n} = \frac{\sum_{i=1}^{N} k_i x_i}{\sum_{i=1}^{N} k_i}. \qquad (1.6.3.)$$

Az ilyen átlagszámítást *súlyozott átlag*nak, a k_i gyakoriság értékeket pedig *súlyfaktor*nak nevezzük.

Az (1.6.3) összefüggés alapján egy másik képletre is juthatunk. Mivel $\sum_{i=1}^{N} k_i = n$ a mérések száma, ami a méréssorozat esetén állandó, ezért felírható az alábbi összefüggés is:

$$\bar{x} = \frac{k_1}{n}x_1 + \frac{k_2}{n}x_2 + \frac{k_3}{n} + \ldots + \frac{k_N}{n}x_N = \sum_{i=1}^{N} \frac{k_i}{n}x_i = \sum_{i=1}^{N} g_i x_i, \qquad (1.6.4.)$$

ahol g_i most is a relatív gyakoriságot jelöli. Természetesen a számtani közép értéke a számolás módjától független. A számtani közepet a statisztikában **empirikus várható értéknek** is szokás nevezni.

A számtani közép tulajdonságai

1. A számtani közép definíciójából közvetlenül adódik az alábbi tulajdonság:

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0, \qquad (1.6.5.)$$

hiszen

$$\sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} x_i - n\bar{x} = 0.$$

Az így kapott összefüggést átrendezve éppen a számtani közép definíciójára jutunk. Ez az összefüggés tulajdonképpen azt jelenti, hogy a számegyenesen a számtani középtől balra és jobbra elhelyezkedő számok számtani középtől mért távolságainak összege megegyezik.

2. Ha az átlagolandó értékek mindegyikéhez hozzáadunk egy állandó számot, akkor az átlag ezzel az állandó értékkel változik meg:

$$\frac{1}{n}\sum_{i=1}^{n} (x_i + c) = \bar{x} + c. \qquad (1.6.6.)$$

Ez a tulajdonság az összegzés elvégzésével azonnal adódik. Az (1.6.6.) kifejezés természetesen kivonás esetén is igaz.

3. Ha az átlagolandó értékeket egy c állandóval megszorozzuk, akkor az átlag is megszorzódik ezzel az állandóval:

$$\frac{1}{n}\sum_{i=1}^{n} cx_i = c\overline{x}\,. \tag{1.6.7.}$$

Ez a tulajdonság az összegzés disztributív tulajdonságának következménye.

Az (1.6.6) és (1.6.7) tulajdonságok a számtani közép számolása esetén sokszor alkalmazhatóak. Az átlagolandó értékekből állandó értéket levonva, vagy állandóval osztva sokszor egyszerűbb számokhoz jutunk, majd a végén hozzáadással, vagy szorzással megkapjuk a helyes átlagértéket. Ha például 1200, 3600 és 4800 átlagát akarjuk számolni, akkor elegendő 12, 36 és 48 számtani közepét kiszámolni, majd a kapott értéket megszorozni 100-zal.

Haladóknak

4. Ha az átlagolandó értékekből levonunk egy számot, és a különbséget négyzetre emeljük, akkor a négyzetek összegének minimuma éppen az átlagértéknél lesz, azaz

$$\sum_{i=1}^{n}(x_i - x)^2 = f(x) \text{ minimális, ha } x = \overline{x}\,.$$

A tulajdonság úgy bizonyítható, ha megnézzük, hogy *f(x)*-nek x szerinti deriváltja milyen x értéknél egyenlő nullával. Tehát

$$f'(x) = -2\sum_{i=1}^{n}(x_i - x) = -2\sum_{-i=1}^{n} x_i - 2nx = 0\,,$$

ahonnan

$$x = \frac{1}{n}\sum_{i=1}^{n} x_i\,, \tag{1.6.8.}$$

és ez definíció szerint éppen a számtani közép.

Gyors ellenőrző feladatok

1.3. Ellenőrizzük, hogy a (1.6.8) kifejezés valóban minimum. Képezzük *f(x)* második deriváltját, és nézzük meg, hogy a kapott érték pozitív előjelű-e a szélsőérték helyén!

1.4. Számoljuk ki a következő adatok átlagát: 1400; 1200; 200; 500; 1100. Használjuk fel az összefüggések közül a megfelelőt!

1.5. Az 1.4. táblázat adatait felhasználva, alkalmazva az (1.6.6) és (1.6.7) összefüggéseket, számítsuk ki Michelson fénysebesség méréseinek számtani közepét (átlagát)! A feladat megoldása során célszerű táblázatkezelő programot használni!

1.7. A mértani közép

A számtani középpel nem mindig tudjuk kifejezni, amit az átlagtól, vagy a középérték fogalomtól elvárunk. Lássunk egy példát! A csapadék éves mennyisége az első évben 12%-ot, a második évben 1%-ot, a harmadik évben pedig 2%-ot növekedett. Mekkora a növekedés a három év alatt?

A növekedés:

$$1{,}12 \cdot 1{,}01 \cdot 1{,}02 = 1{,}1538\,,$$

vagyis, a növekedés a három év alatt *15,38%*-os.

Az átlagtól most is elvárjuk, hogy vele helyettesítve az átlagolandó értékeket ne változzék az összes növekedés mértéke. Tehát ha az átlagos éves növekedés p, akkor a három év alatti növekedés:

$$(1+p)^3 = 1{,}1538\,.$$

Megoldva p-re ezt az egyenletet, megkapjuk az átlagos évi csapadéknövekedést:

$$1+p = \sqrt[3]{1{,}1538} = 1{,}0488\,.$$

Az átlagos éves növekedés tehát *4,88%*. A három növekedés számtani közepe *5%*, tehát a jelen esetben a számtani közép helytelen eredményre vezetett volna.

Definíció. Általánosítva a mondottakat, ha $x_1, x_2, \ldots, x_n$ adatlista n darab nem negatív számból áll ($\forall x_i \geq 0$), akkor ezeknek a számoknak a **mértani közepe**:

$$\overline{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}\,. \qquad (1.7.1.)$$

A példából jól látszik, hogy a mértani középet akkor használjuk, amikor az átlagolandó számok szorzatának van értelme. Általában, ha éves növekedési rátákból számoljuk ki a növekedést (kamatok, népesség, árindex, szennyezettségváltozás, energianövekedés stb.), akkor az átlag kiszámításához mértani közepet használunk.

1.8. Harmonikus közép

Más átlagot használunk akkor, ha például átlagsebességet kell számolnunk. Vegyük a következő példát. Formula 1-es autó az s hosszúságú kört 200 km/h sebességgel teszi meg, míg a következő körben 300 km/h óra a sebessége. A kérdés az, hogy mekkora átlagsebességgel kellett volna mennie, hogy ugyanannyi idő alatt tegye meg a két kört? A számolást a következőképpen végezzük:

$$\overline{v} = \frac{2s}{t_1 + t_2} = \frac{2s}{\frac{s}{v_1} + \frac{s}{v_2}} = \frac{2}{\frac{1}{v_1} + \frac{1}{v_2}}.$$

A konkrét példában az így kiszámolt átlagsebesség *240* km/h. A számtani közép *250* km/h lett volna. Általánosítva a felismerést, az ún. harmonikus közép definíciója:

Definíció. Az $x_1, x_2, \ldots, x_n$ mérési adat **harmonikus középét** a

$$\overline{x}_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \tag{1.8.1.}$$

képlet alapján számoljuk. Vegyük észre, hogy ez a képlet valójában azt jelenti, hogy a harmonikus közép reciproka egyenlő a reciprokok számtani közepével, hiszen (1.8.1) kis átalakításával azt kapjuk, hogy

$$\frac{1}{\overline{x}_h} = \frac{\sum_{i=1}^{n} \frac{1}{x_i}}{n}.$$

Sok esetben a harmonikus közép adja a megfelelő átlagértéket. Lássunk néhány példát! A korábbi példában láttuk, hogy ha az út azonos részeit tesszük meg különböző sebességgel, akkor az átlagsebességet a sebességek harmonikus közepe adja meg. (Ha azonos idők alatt haladunk különböző sebességgel, akkor a sebességek számtani közepe ad helyes átlagot!). Ha különböző ellenállásokat kapcsolunk párhuzamosan, és ezeket azonos ellenállásokkal akarjuk helyettesíteni, akkor is az ellenállásértékek harmonikus közepe ad helyes értéket. Ha azonos tömegű, de különböző sűrűségű folyadékokat elegyítünk, akkor az átlagos sűrűséget a harmonikus közép alapján számolhatjuk.

Könnyen belátható, hogy ha a mérési adatok mind egyformák, akkor az eddig tárgyalt középértékek megegyeznek, és értékük megegyezik a mérési adatok értékével. Egyébként közöttük az alábbi nagysági viszony áll fenn:

$$\overline{x}_h \leq \overline{x}_g \leq \overline{x}.$$

Gyors ellenőrző feladatok

1.6. Magyarország népessége 2004-ben *10 116 742* fő volt [1]. A népesség éves csökkenését az alábbi táblázat tartalmazza:

év	2005	2006	2007	2008	2009
csökkenés	*1,897‰*	*2,0765‰*	*1,0344‰*	*2,0621‰*	*1,4361‰*

1.1. táblázat: Magyarország évenkénti népességcsökkenése

[1] Központi Statisztikai Hivatal: http://portal.ksh.hu

Mekkora volt Magyarország lakosságának létszáma 2009. év végén? Mekkora volt ebben az időszakban az átlagos népességcsökkenés?

1.7. Két, azonos tömegű, különböző sűrűségű folyadékot összekeverünk. Kérdés, hogy mekkora lesz a keverék sűrűsége?

1.8. $R_1=100\ \Omega$ és $R_2=200\ \Omega$ ellenállásokat párhuzamosan kapcsolunk. Kérdés, mekkora azonos ellenállásokkal kellene helyettesíteni a két különböző ellenállást, hogy az áramerősség az áramkörben ne változzon?

1.9. Medián

Az eddig tárgyalt középértékeknek az a tulajdonsága, hogy érzékenyek a kiugró értékekre. Ha például azt halljuk, hogy egy 25 főt foglalkoztató cégnél a bruttó átlagkereset 312 000 Ft, akkor ezt csábítónak érezzük. Azonban, ha utánanézünk a részleteknek, akkor kiderül, hogy a cégnél 24 fő átlagkeresete bruttó 200 000 Ft, és a vezető keresete bruttó 3 000 000 Ft. Helyesebb lenne ilyenkor azt mondani, hogy a beosztottak átlagkeresete 200 000 Ft és a vezető 3 000 000 Ft-ot kap. A számtani közép ilyen esetben félrevezető értéket ad, mert a kiugró értékre érzékeny.

Definíció. A nagyság szerint rendezett $x_1, x_2, \ldots, x_n$ mérési adatok **medián**ján az adatok középső értékét értjük. Ha n páratlan szám, akkor a medián az adatlista középső értéke. Ha n páros, akkor az adatlista két középső értékének számtani közepe adja meg a medián értékét.

A fenti fizetéses példában, ha valamennyi alkalmazottnak 200 000 Ft a fizetése, akkor a medián értéke is 200 000 Ft (függetlenül a vezető kiugróan magas fizetésétől). A definícióból látszik, hogy a medián olyan középérték, amelyik nem érzékeny a kiugró adatokra.

1.10. Az eloszlás módusza és terjedelme

Az eloszlás helyének jellemzésére használt paraméter a **módusz**.

Definíció. Diszkrét eloszlás esetén az eloszlás **módusz**a a leggyakrabban előforduló mért érték. Más szóval, a gyakoriság diagramnak a módusznál van a maximuma. Folytonos eloszlás esetén a módusz annak az osztálynak az osztályközepe, ahol a gyakoriságnak maximuma van.

Példaként, az 1.8. ábrán a módusz értéke *175 cm*.

A gyakoriság eloszlás másik fontos tulajdonsága az eloszlás szélessége. A szélességet többféleképpen jellemezhetjük. Az egyik lehetséges jellemzés az **eloszlás terjedelmének** a megadása.

Definíció. Az **eloszlás terjedelmén** a legnagyobb és a legkisebb adat különbségét értjük:

$$t = x_{max} - x_{min} .$$

Könnyen belátható, hogy a terjedelem érzékeny a kiugró adatokra. Ezért az x_i mérési adatok terjedelmét sokszor úgy szeretnénk jellemezni, hogy figyelembe vesszük valamenynyi adat eltérését valamelyik középértéktől. Leggyakrabban a számtani középtől való $x_i - \overline{x}$ eltérést vesszük. Minthogy azonban ezek összege a (1.6.5) összefüggés szerint nullát ad, ezért inkább az

$$s^* = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \overline{x}\right|$$

átlagos abszolút eltérést vehetnénk az eloszlás szélességének jellemzésére. Az abszolút értékkel azonban nehézkes a számolás, ezért ezt ritkán használjuk. A negatív értékű eltérésektől azonban nemcsak abszolút értékkel, hanem négyzetre emeléssel is meg lehet szabadulni. Ezért a leggyakrabban az **átlagos négyzetes eltérést** használjuk az eloszlás szélességének a mérésére.

1.11. Empirikus szórásnégyzet és szórás

Az eloszlás szélességét jellemző, leggyakrabban használt paraméter az empirikus szórás.

Definíció. Legyenek $x_1, x_2, ..., x_n$ tetszőleges valós számok a mérési adatlista elemei, amelyek számtani közepe $\overline{x}$. Ekkor az

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad (1.11.1.)$$

kifejezést **átlagos négyzetes eltérésnek,** vagy **empirikus** (tapasztalati) **szórásnégyzetnek** nevezzük. Az empirikus szórásnégyzetet a statisztikában **empirikus második centrális momentumnak** is nevezik. A kifejezésből látszik, hogy az empirikus szórásnégyzet négyzetes dimenziójú. Az eloszlás szélességét ezért jobban jellemzi a (1.11.1) kifejezés négyzetgyöke.

Definíció. Az **empirikus szóráson** az empirikus szórásnégyzet pozitív négyzetgyökét[2] értjük:

$$s = {}_{+}\sqrt{s^2} = {}_{+}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} . \qquad (1.11.2.)$$

[2] A négyzetgyökvonásnak két értéke van: egy pozitív és egy negatív. Mivel az empirikus szórás az eloszlás szélességét jellemzi, ezért a negatív megoldásnak itt nincs értelme.

Az empirikus szórásnégyzet számolása

Az empirikus szórásnégyzetet számolhatjuk a (1.11.1.) képlet alapján, de a gyakorlatban egyszerűbb a számolás az alábbi tétel alapján.

Tétel. Legyenek az adatlista elemei az $x_1, x_2, \ldots, x_n$ valós számok. Számtani közepük: $\overline{x}$. A szórásnégyzetre igaz az, hogy

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 . \qquad (1.11.3.)$$

Bizonyítás.

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\left(\sum_{i=1}^{n}(x_i^2 - 2\overline{x}x_i + \overline{x}^2)\right) = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} 2\overline{x}x_i + \sum_{i=1}^{n}\overline{x}^2\right) =$$
$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\overline{x}\frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}n\overline{x}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\overline{x}^2 + \overline{x}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 ,$$

vagyis az adatok négyzetének átlagából kivonva az adatok átlagának négyzetét megkapjuk az empirikus szórásnégyzetet.

Gyors ellenőrző feladatok

1.9. Az 1.6. alfejezetben kiszámoltuk 10 kockadobás számtani közepét és $\overline{x} = 3{,}3$ értéket kaptunk. Számoljuk ki a 10 adat empirikus szórásnégyzetét és empirikus szórását. Az adatlista:

$$x_1 = 5;\ x_2 = 5;\ x_3 = 6;\ x_4 = 1;\ x_5 = 2;\ x_6 = 2;\ x_7 = 4;\ x_8 = 5;\ x_9 = 1;\ x_{10} = 2.$$

1.10. Az adatlista elemei az $x_1, x_2, \ldots, x_n$ valós számok. Számtani közepük: $\overline{x}$. Készítsünk gyakoriság táblázatot az adatokból! Minthogy az adatok között vannak azonosak is, a táblázat elemei legyenek $a_1, a_2, \ldots, a_m$, ahol $m \le n$. Az a_i értékek gyakorisága k_i, relatív gyakorisága g_i. Lássuk be az alábbi összefüggéseket:

$$s^2 = \frac{1}{n}\sum_{i=1}^{m} k_i a_i^2 - \overline{x}^2 ,$$
$$s^2 = \sum_{i=1}^{m} g_i a_i^2 - \overline{x}^2 .$$

Az összefüggések alapján számoljuk ki ismét az 1.9 feladatban 10 kockadobás során az empirikus szórásnégyzet értékét.

2. ÖSSZEFÜGGÉSEK AZ ISMÉRVEK KÖZÖTT

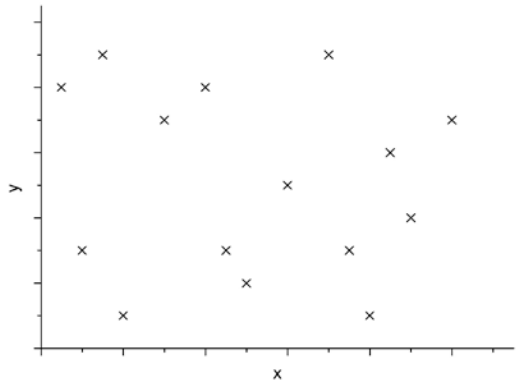
A statisztikában azt a rendszert, amit vizsgálunk, amelyen méréseket végzünk, statisztikai sokaságnak nevezzük. A korábbi fejezetben a sokasának csak egyetlen ismérvét tekintettük, és vizsgáltuk ennek az ismérvnek a gyakoriság eloszlását. Példaként vizsgáltuk az emberek egy csoportja magasságának gyakoriság eloszlását. A példában a sokaság az emberek csoportja, az ismérv pedig a magasság.

A sokaságnak azonban nem csak egy ismérve lehet. A fenti példában például a magasság mellett valamilyen mutatószám alapján vizsgálható az emberek önbizalma, mint a sokaság másik ismérve. Értelmes felvetni azt a kérdést, hogy vajon függetlenek-e egymástól ezek az ismérvek, vagy van-e közöttük valamilyen összefüggés? Tehát például függ-e az emberek önbizalma a magasságuktól? A tudományban gyakori az ilyen kérdésfelvetés. Az ismérvek közötti összefüggések tanulmányozása új felismerések forrása lehet.

Ebben a fejezetben a sokaság két ismérve közötti összefüggéseket vizsgáljuk. Ha az ismérvek számszerűsíthetők, azaz mérési adataink számok formájában jelennek meg, akkor az összefüggések vizsgálatára a regresszió- és korrelációszámítás jöhet szóba.

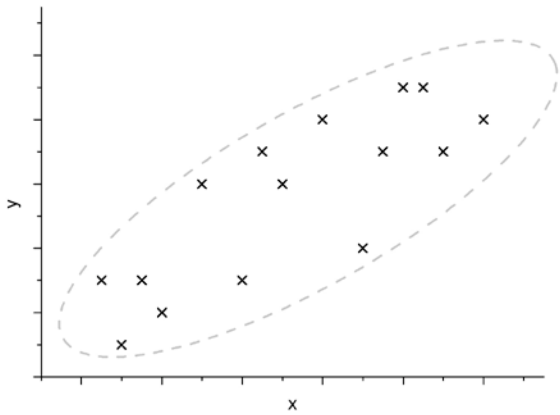
2.1. Pontdiagram

Ha a sokaság elemein mérést végzünk a két ismérvvel kapcsolatban, akkor (x_i,y_i) adatpárokhoz jutunk. Az adatokat általában táblázatba gyűjtjük. A táblázatban azonban nem látszanak a tendenciák. Sokkal szemléletesebb, ha az adatokat derékszögű koordináta-rendszerben ábrázoljuk. Minden értékpár egy-egy pontot eredményez a derékszögű koordináta-rendszer síkjában. Az ilyen ábrát **pontdiagram**nak nevezzük. A pontdiagramon sok esetben már szemre is kivehető, hogy van-e valamiféle összefüggés az ismérvek között. Ha a pontok eloszlása olyan, mint a 2.1. ábrán, akkor azt gondoljuk, hogy nincs összefüggés az adatok között. Ha pedig a pontok eloszlása olyan, hogy köréjük olyan ellipszis rajzolható, melynek főtengelye nullától különböző szöggel hajlik az x tengelyhez (2.2. ábra), akkor ez azt jelenti, hogy összefüggés tapasztalható az ismérvek között, azaz tendencia van a pontok eloszlásában.



2.1. ábra: Az ismérvek között nincs összefüggés

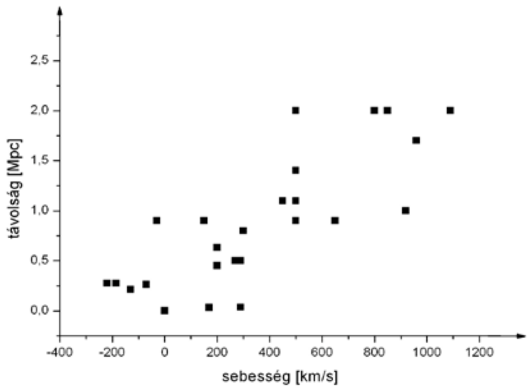
Az ismérveket jellemző adatok között gyakorta lineáris az összefüggés, vagy mint a későbbiekben látni fogjuk, az összefüggés könnyen lineárissá alakítható. Ezért először a lineáris függés jellemzőinek meghatározásával foglalkozunk.



2.2. ábra: Az ismérvek között összefüggés van

2.2. Lineáris regresszió

Nézzünk először példaként olyan pontdiagramot, amely a 20. század tudománytörténetének egyik jeles ábrája. Edwin Hubble és Milton Humason a galaxisokat tanulmányozta, és 1929-ben közölték azt a pontdiagramot, amely a galaxisok Földtől való távolságának és a Földtől távolodó sebességének adatait tartalmazza. A 2.3. ábrán látható ez a pontdiagram.

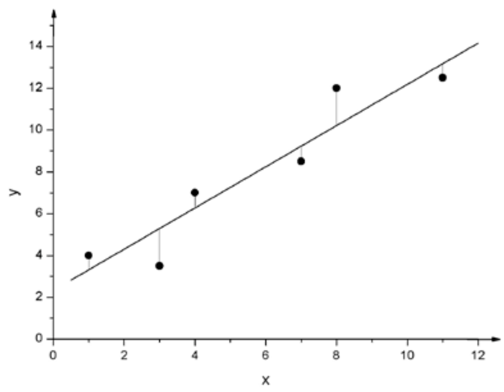


2.3. ábra: A galaxisok sebességének és Földtől való távolságának adatai

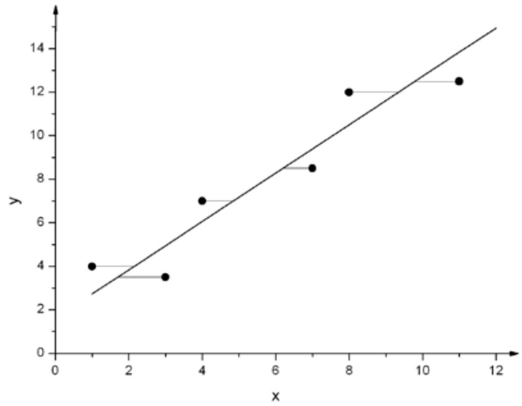
A sebesség mértékegysége *km/s*, a távolság millió (mega) *parsec*-ben szerepel (1 parsec=*3,26* fényév). A korábban mondottak szerint a pontok köré rajzolható ellipszis, tehát az ismérvek között van összefüggés. Az ábrán az látszik, hogy minél távolabb van egy galaxis a Földtől, annál nagyobb a Földtől távolodás sebessége. Ez a felismerés vezetett később a Nagy Bumm és a táguló világegyetem gondolatához, ami a modern asztronómia egyik alapfelismerése.

Ha lineáris összefüggést feltételezünk a sebesség és a távolság között, akkor felmerül a kérdés, hogy hogyan határozzuk meg ennek az egyenesnek az adatait. Az első gondolat az lehet, hogy szemre húzzuk be a trendvonalat. Ez nem rossz ötlet, sokszor ezt is tesszük, amikor a mérések után gyors értékelést végzünk. Hogyan húzunk vonalzóval a pontokra legjobban illeszkedő egyenest? Igyekszünk úgy elhelyezni a vonalzót, hogy a meghúzott vonal alatt és felett körülbelül azonos számú pont helyezkedjen el. A módszer nem rossz, hátránya azonban, hogy valahányszor szemre elvégezzük az illesztést, mindannyiszor kissé különböző eredményt kapunk.

Lehet objektívebb módszert is választani, amely tudományos módszerként jobban alkalmazható. Meghatározhatjuk például a pontok távolságainak összegét a lehetséges trendegyenesektől, és ezek közül azt választjuk, amely esetében a ponttávolságok összege a legkisebb. Milyen távolságokat válasszunk? Választhatjuk a pontok függőleges távolságát, ahogyan azt a 2.4. ábra mutatja, vagy a vízszintes távolságokat, ahogyan azt a 2.5. ábrán látjuk.



2.4. ábra: A pontok függőleges távolsága a trendegyenestől



2.5. ábra: A pontok vízszintes távolsága a trendegyenestől

Választható lenne még a pontok geometriai (merőleges) távolsága is, azzal azonban körülményes számolni, ezért azt nem szokták választani.

A távolság számolásnál abszolút értékekkel kellene dolgoznunk, ami matematikailag körülményes, ezért inkább a Gauss által javasolt négyzetösszegekkel dolgozunk, ahhoz hasonlóan, ahogyan azt az empirikus szórásnégyzet esetében tettük. A módszert legkisebb négyzetek módszerének nevezzük.

2.3. A legkisebb négyzetek módszere

A legkisebb négyzetek Gauss által javasolt módszerében a mérési pontok egy lehetséges trendegyenestől mért függőleges vagy vízszintes távolságainak négyzetösszegét számoljuk, majd megkeressük azt az egyenest, amely esetében ez a négyzetösszeg a legkisebb. Az alábbiakban matematikailag fogalmazzuk meg az itt leírt feltételt.

Definíció. A legkisebb négyzetek módszerével az n darab (x_i,y_i) pontpárra legjobban illeszkedő egyenest határozzuk meg. A még nem ismert, meghatározandó egyenes meredeksége legyen a, tengelymetszete b, tehát az egyenes egyenlete:

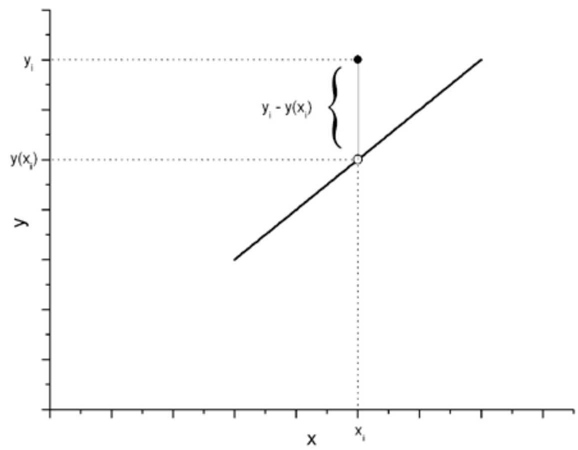
$$y = ax + b\,.$$

Az x_i mért értékhez y_i mért érték tartozik. Az x_i pontban a meghatározandó egyenes y koordinátája: $y(x_i)$, ahogyan azt a 2.6. ábrán láthatjuk. A mérési pont és az egyenesen lévő pont távolsága tehát:

$$y_i - y(x_i)\,.$$

A legkisebb négyzetek módszere szerint ezeknek a távolságoknak a négyzetösszegét kell számolni:

$$S(a,b) = \sum_{i=1}^{n} \left(y_i - y(x_i)\right)^2\,. \qquad (2.3.1.)$$



2.6. ábra: Magyarázó ábra a legkisebb négyzetek módszeréhez

Úgy kell megválasztani *a* és *b* értékét, hogy a négyzetösszeg minimális legyen, vagyis

$$S(a,b) = \sum_{i=1}^{n} \left(y_i - (ax_i + b)\right)^2 = min\,.$$

A fenti definíció alapján a számolás az alábbiak szerint történik. Az $S(a,b)$ függvény minimumát kell megkeresni, ami közismert módon a deriváltak nullhelyeinek meghatározását jelenti. Ki kell tehát számolnunk az alábbi két parciális derivált értékét:

$$\frac{\partial S(a,b)}{\partial a} = 0\,; \quad \frac{\partial S(a,b)}{\partial b} = 0\,.$$

Elvégezve a deriválás műveletét, két egyenletre jutunk:

$$\sum_{i=1}^{n} 2\left(y_i - (ax_i + b)\right)\left(-x_i\right) = 0\,, \tag{2.3.2.}$$

$$\sum_{i=1}^{n} 2\left(y_i - (ax_i + b)\right)\left(-1\right) = 0\,. \tag{2.3.3.}$$

Átrendezve az egyenleteket (2.3.2.)-ből és (2.3.3.)-ból azt kapjuk, hogy

$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i\,, \tag{2.3.4.}$$

$$\sum_{i=1}^{n} y_i = a\sum_{i=1}^{n} x_i + bn\,. \tag{2.3.5.}$$

A keresett a és b paraméter ebből az egyenletrendszerből, a mért x_i, y_i értékekkel kifejezhető.

Számolásra alkalmasabb és áttekinthetőbb formulát kapunk, ha bevezetjük a következő új változókat (ezek lényegében a mérési pontok koordinátáinak számtani közepei):

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \tag{2.3.6.}$$

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}. \tag{2.3.7.}$$

Ezekkel kifejezve a két keresett mennyiséget:

$$a = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}, \tag{2.3.8.}$$

$$b = \overline{y} - a\overline{x}. \tag{2.3.9.}$$

A második deriváltakkal belátható, hogy az így kapott a és b értékeknél $S(a,b)$-nek minimuma van.

A 2.4. ábrán a (2.3.8.) és a (2.3.9.) egyenletek felhasználásával kiszámolt egyenest rajzoltuk be. Ezt az eljárást lineáris regressziónak nevezik. A függőleges távolságok felhasználásával számolt legjobban illeszkedő egyenest pedig **első regressziós egyenesnek** szokták nevezni.

Természetesen a legkisebb négyzetek módszerével kiszámolható a vízszintes távolságok alapján is a legjobban illeszkedő egyenes. Ilyenkor az $x_i - x$ távolságok négyzetösszegét kell kiszámolni, és a minimumot megkeresni, ahol

$$x(y_i) = a^* y_i + b^*. \tag{2.3.10.}$$

Formálisan azonnal adódik, hogy a változók felcserélésével a megoldás alakja:

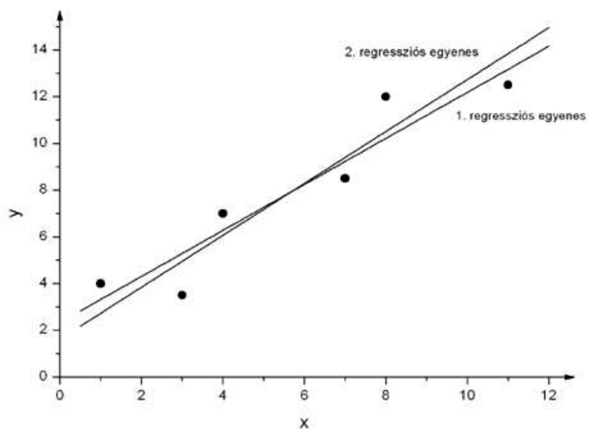
$$a^* = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} y_i^2 - n\overline{y}^2},$$

$$b^* = \overline{x} - a^*\overline{xy}.$$

Ha a szokásos alakra akarjuk hozni az egyenes egyenletét, akkor (2.3.10.) kis átalakításával azt kapjuk, hogy

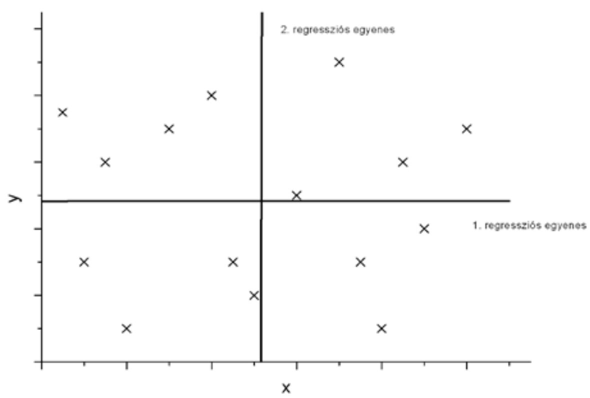
$$y = \frac{1}{a^*}x - \frac{b^*}{a^*}. \quad (2.3.11.)$$

A 2.5. ábrára az így kiszámolt paraméterű egyenest rajzoltuk. A vízszintes távolságok alapján számolt egyenest **második regressziós egyenesnek** nevezik. A 2.7. ábrára közös koordináta-rendszerbe rajzoltuk az első és második regressziós egyenest.



2.7. ábra: Az ábra pontjaihoz illeszkedő két regressziós egyenes

Általában a két egyenes különbözik egymástól. Mennél jobban illeszkednek a mérési pontok egy egyenesre, annál közelebb esik egymáshoz a két regressziós egyenes. Ha a pontok pontosan egy egyenesre esnek, akkor nyilvánvaló, hogy mindkét regressziós egyenes ezen az egyenesen halad, hiszen a pontok egyenestől való minimális távolsága így nulla. Ilyenkor tehát a két regressziós egyenes megegyezik. Mennél jobban szórnak a pontok, annál inkább különbözik a két regressziós egyenes. A 2.1. ábrán mutatott trend nélkül szóró pontok esetében az első regressziós egyenes vízszintes, a második regressziós egyenes pedig függőleges, ahogy az a 2.8. ábrán látható.



2.8. ábra: Trend nélkül szóró pontok esetében a két regressziós egyenes

2.4. Lineáris korreláció

A lineáris korrelációanalízis feladata az, hogy a korábban feltárt tulajdonságok felhasználásával jellemezze, és számszerűsítse azt, hogy a két mért ismérv között mennyire lineáris a kapcsolat, vagyis, hogy a pontok a diagramon mennyire illeszkednek egy egyenesre. Azt kell tehát vizsgálnunk, hogy a két regressziós egyenes mennyire tér el egymástól, azaz hogy az első regressziós egyenes meredeksége hogyan aránylik a második regressziós egyenes meredekségéhez. A korábbiakban már láttuk, hogy az első regressziós egyenes meredeksége

$$m_1 = a\,,$$

a második regressziós egyenes meredeksége pedig

$$m_2 = \frac{1}{a^*}\,.$$

A jellemzéshez használt hányados:

$$r^2 = \frac{m_1}{m_2} = \frac{a}{\frac{1}{a^*}} = aa^*\,. \qquad (2.4.1.)$$

Ennek gyöke

$$r = \sqrt{aa^*}\,. \qquad (2.4.2.)$$

Vizsgáljuk meg r abszolút értékének viselkedését! Ha a pontok pontosan egy egyenesre esnek, vagyis tökéletes a két ismérv között a lineáris kapcsolat, akkor a két regressziós egyenes meredeksége megegyezik, vagyis m_1=m_2. Ilyenkor tehát (2.4.1.)-ből az következik, hogy

$$|r| = 1\,. \qquad (2.4.3.)$$

A másik véglet az, amikor a két regressziós egyenes merőleges egymásra. Ilyenkor, mint láttuk, nincs lineáris kapcsolat az ismérvek között, és (2.4.1.)-ből az következik, hogy

$$|r| = 0\,. \qquad (2.4.4.)$$

Minden más esetben $|r|$ értéke 0 és 1 közötti.

Ha részletes felírjuk r^2 alakját, akkor az alábbi kifejezést kapjuk:

$$r^2 = \frac{\left(\sum_{i=1}^{n} x_i y_i - n\overline{xy}\right)\left(\sum_{i=1}^{n} x_i y_i - n\overline{xy}\right)}{\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right)\left(\sum_{i=1}^{n} y_i^2 - n\overline{y}^2\right)}.$$

Ennek négyzetgyöke:

$$r = \frac{\left(\sum_{i=1}^{n} x_i y_i - n\overline{xy}\right)}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right)\left(\sum_{i=1}^{n} y_i^2 - n\overline{y}^2\right)}}. \qquad (2.4.5.)$$

Ha most ezt tekintjük r definíciójának, akkor látszik, hogy r előjeles. A (2.4.5.) kifejezés részletesebb vizsgálata alapján megmutatható, hogy r akkor negatív, ha a regressziós egyenes meredeksége negatív, azaz az egyenes „lefelé” halad, és akkor pozitív, ha az egyenes meredeksége pozitív, vagyis az egyenes „emelkedik”. Az r mennyiség elnevezése **empirikus korrelációs együttható**.

A 2.4. ábrán mutatott adatok esetében $r=0,926$, ami még jó lineáris kapcsolatot mutat. Ellenben, ha r értéke például $0,4$, akkor gyenge a lineáris kapcsolat a két ismérv között, és a legkisebb négyzetek módszerével meghatározott egyenes paraméterei bizonytalan információval szolgálnak a trendet illetően. Az empirikus korrelációs együttható tehát arra használható, hogy az ismérvek közötti lineáris kapcsolat erősségét vizsgáljuk. Az r abszolút értéke minél közelebb van 1-hez, annál erősebb a lineáris kapcsolat.

Gyors ellenőrző feladatok

2.1. Lássuk be, hogy a korrelációs egyenes meredekségét leíró (2.3.8.) kifejezés nevezője mindig pozitív! Azt kell tehát belátni, hogy

$$\sum_{i=1}^{n} x_i^2 - n\overline{x}^2 > 0\,!$$

2.2. Bizonyítsuk be, hogy ha a regressziós egyenes meredeksége negatív ($a<0$), akkor ez r-re is igaz, vagyis ilyenkor $r < 0$.

2.3. Igazoljuk, hogy az r empirikus korrelációs együtthatóra talált (2.4.5.) összefüggés és az alábbiakban adott alak megegyezik egymással:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}. \qquad (2.4.6.)$$

2.5. Nemlineáris regresszió

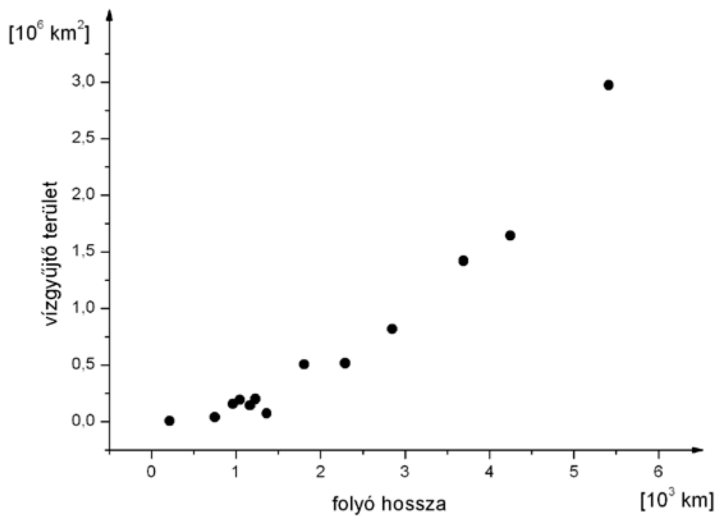
Linearizálási eljárás

Az ismérvek között sokszor nem lineáris a kapcsolat. Ez a mérési adatokat tartalmazó táblázat alapján ritkán derül ki, de vagy elméleti megfontolásokból tudjuk, vagy pedig a pontdiagram segítségével állapítjuk meg. Példaként lássuk a 2.1. táblázat:ot, amely eurázsiai folyók hosszát és a hozzájuk tartozó vízgyűjtő terület nagyságát mutatja be.

folyó	hossz [km]	vízgyűjtő terület [km^2]
Ob	5410	2 972 000
Irtis	4248	1 643 000
Volga	3692	1 420 000
Duna	2850	817 000
Dnyeper	2290	516 000
Káma	1805	507 000
Dnyeszter	1362	72 000
Rajna	1230	199 000
Elba	1165	144 000
Visztula	1047	194 000
Tisza	962	157 000
Dráva	749	40 000
Ipoly	212	5 000

2.1. táblázat: Eurázsiai folyók hossza és vízgyűjtő területeik nagysága

A táblázatból nem látszanak a tendenciák. Pontdiagramon ábrázolva a vízgyűjtő területet a folyóhossz függvényében, láthatóvá válik, hogy összefüggés van a két jellemző között, de az összefüggés nem lineáris. Ez látható a 2.9. ábrán.

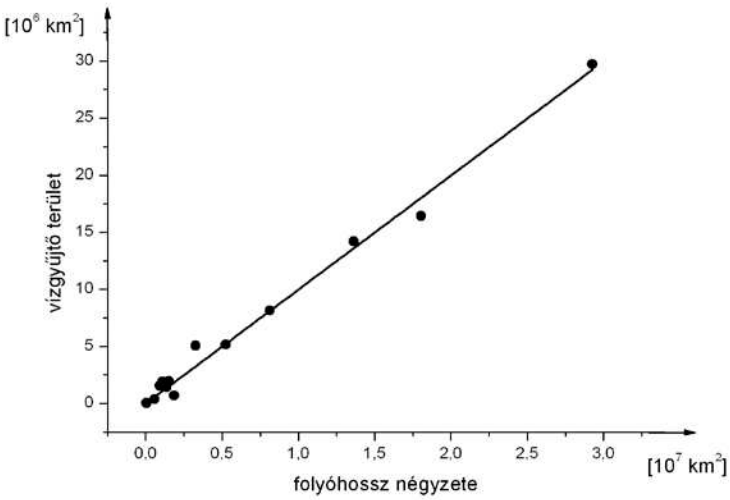


2.9. ábra: Eurázsiai folyók hossza és vízgyűjtő területük nagysága

Elméleti meggondolásokból sejthető, hogy a vízgyűjtő terület nagysága négyzetesen függ a folyó hosszától. Az is feltehető, hogy a *0* km hosszúságú folyóhoz *0* km^2 vízgyűjtő terület tartozik. A függvénykapcsolat jellege tehát

$$y = ax^2$$

alakú, ahol x folyó hosszát y pedig a vízgyűjtő terület nagyságát jelöli. Ha igaz a feltevés, akkor y-t x^2 függvényében ábrázolva egyenest kell kapnunk. Ezt az ábrázolás módot mutatja a 2.10. ábra.



2.10. ábra: A folyók vízgyűjtő területe a folyóhossz négyzetének függvényében

Ha tehát a sokaság jellemzői között nem lineáris a kapcsolat, akkor az ábrán sok esetben lineárissá tehető az összefüggés. Ezt követően pedig alkalmazható a korábban megismert lineáris regresszió a legkisebb négyzetek módszerével.

Ha elméleti megfontolásokból nem ismerjük a hatványfüggvény kitevőjét, akkor a következő eljárás vezethet sikerre. Legyen a függvény alakja

$$y = bx^a .$$

Képezzük az egyenlet mindkét oldalának természetes alapú logaritmusát:

$$ln\, y = ln\, b + a\, ln\, x .$$

Ha bevezetjük a következő új változókat:

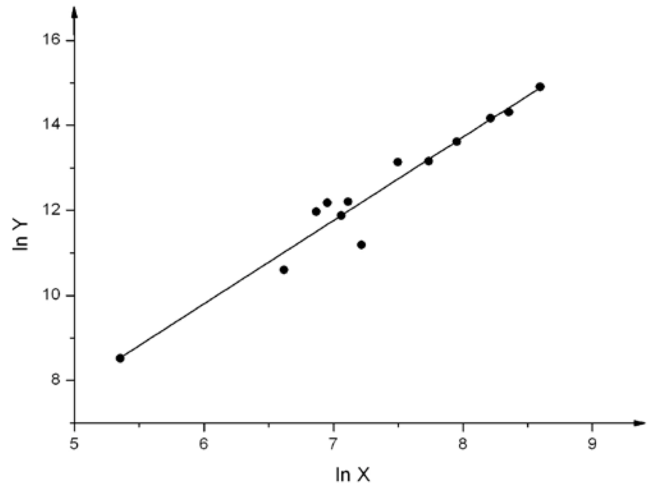
$$\hat{y} = ln\, y,\ \ \hat{b} = ln\, b;\ \ \hat{x} = ln\, x,$$

akkor azt kapjuk, hogy

$$\hat{y} = a\hat{x} + \hat{b} .$$

tehát, $\hat{y}$-ot ábrázolva $\hat{x}$ függvényében egyenest kapunk, amelynek meredeksége az ismeretlen kitevő.

Nézzük meg, hogyan alkalmazható a talált összefüggés például az előző, folyókkal kapcsolatos példában. Tegyük fel, hogy nem ismerjük a kitevőt, és meg szeretnénk határozni. Ehhez ábrázolni kell a folyók hosszának logaritmusát (*ln x*) a folyók vízgyűjtő területének logaritmusa (*ln y*) függvényében. Ez látható a 2.11. ábrán, ahol mindkét logaritmus természetes alapú.



2.11. ábra: A folyók vízgyűjtő területének logaritmusa a folyóhossz logaritmusának függvényében

Az ábrán a legkisebb négyzetek módszerével a pontokra illesztett egyenes is látszik. Az egyenes meredeksége:

$$a = 1{,}96\ ,$$

ami jó közelítéssel igazolja a meredekségre tett korábbi elméleti feltevésünket.

Gyakran exponenciális kifejezéssel van dolgunk (például élettartamok, radioaktív bomlások időpontja stb. során). Ilyenkor tehát az (x_i, y_i) jellemzők között

$$y = be^{-ax}\,; \quad x > 0$$

jellegű összefüggést keresünk. Meghatározandó az *a* és a *b* paraméter. A követhető eljárás az alábbi. Képezzük mindkét oldal természetes alapú logaritmusát:

$$\ln y = \ln b - ax\,.$$

Ha bevezetjük a

$$\hat{y} = \ln y,\ \ \hat{b} = \ln b$$

új változókat, és ábrázoljuk $\hat{y}$ -ot x függvényében, akkor egyenest kapunk, amelynek meredeksége a, tengelymetszete pedig $\ln b$.

A fentiekben vázolt linearizálási eljárás a legtöbb függvénytípus esetében sikeresen alkalmazható, és a gyakorlatban sokszor használjuk is ezt a módszert.

Haladóknak

Nemlineáris legkisebb négyzetek módszere

A jellemzők közötti nemlineáris kapcsolat esetén a fentiekben vázolt linearizáláson kívül alkalmazható a nemlineáris legkisebb négyzetek módszere. Az eljárást az alábbiakban vázoljuk.

Legyen a kísérletek során kapott n darab (x_i, y_i) $(i=1, 2, \ldots, n)$ értékpárunk. Elméleti megfontolásokból, vagy a linearizálási eljárás során megismert módon tudjuk, hogy a pontok között

$$y = \varphi(x; a_1, a_2, \ldots, a_m)$$

függvénykapcsolat van. Minthogy a mérés során statisztikus jellegű hibák terhelik a mérési pontokat, ezért általában a mért y_i és a a képlet adta $y = \varphi(x_i; a_1, a_2, \ldots, a_m)$ értékek eltérnek egymástól:

$$y_i - \varphi(x; a_1, a_2, \ldots, a_m) = \varepsilon_i, \quad (i=1, 2, \ldots, n).$$

A legkisebb négyzetek módszere szerint a feladat az, hogy a

$$S(a_1, a_2, \ldots, a_m) = \sum_{i=1}^{n} \varepsilon_i^2$$

összeget minimalizáljuk. Keressük tehát a

$$S(a_1, a_2, \ldots, a_m) = \sum_{i=1}^{n} \left[y_i - \varphi(x; a_1, a_2, \ldots, a_m) \right]^2$$

függvény minimumát az a_i-k függvényében. Amennyiben a φ függvény az a_i paraméterek szerint differenciálható, akkor a minimum szükséges feltétele:

$$\frac{\partial S}{\partial a_1} = 0; \; \frac{\partial S}{\partial a_2} = 0; \ldots; \frac{\partial S}{\partial a_m} = 0.$$

Az m számú egyenletből kiszámolható az m darab a_i paraméter értéke. Ezeket visszaírva a φ függvénybe, megkapjuk a pontokra legjobban illeszkedő függvényt.

Elméletileg belátható, hogy a linearizálással és a nemlineáris legkisebb négyzetek módszerével kapott függvények nem azonosak. Az eltérés általában nem nagy, és a gyakorlat számára általában elhanyagolható.

Fontos megjegyzés, hogy a legkisebb négyzetek módszerével kapott függvényalakok, bár jól illeszkednek a pontokra, de nem biztos, hogy a függvényalaknak van fizikai jelentése. Az így kapott görbe általában empirikus képletnek tekintendő!

II. A VALÓSZÍNŰSÉG-SZÁMÍTÁS ALAPJAI

3. A VALÓSZÍNŰSÉG FOGALMÁNAK BEVEZETÉSE

3.1. Az alapfogalmak bevezetése

A korábbi fejezetekben már láthattuk, hogy a kísérleti adatok kezeléséhez jól használhatók a leíró statisztika módszerei. Ahhoz azonban, hogy továbblépjünk és mélyebben megismerhessük a véletlen kísérletek sajátosságait, előbb meg kell ismernünk a valószínűségszámítás alapjait.

A leíró statisztika során találkoztunk már a véletlen kísérlet fogalmával. Definiáljuk most szabatosan ezt a fogalmat.

Definíció. Az olyan kísérletet nevezzük **véletlen kísérletnek**, melynek végeredményét az általunk figyelembe vett okok nem határozzák meg egyértelműen. A kísérlet kimenetele több, esetleg végtelen sok lehetséges eredmény közül az egyik.

Egyszerű példaként többször fogjuk emlegetni a kockadobást, vagy a pénzérmés kísérletet, amelyek kimenetele **diszkrét** elemekből álló halmaz egyik eleme. Másik lehetséges példa a napi középhőmérséklet alakulása, amikor **folytonos** halmaz egyik eleme a végeredmény.

Definíció. A véletlen kísérlet lehetséges kimeneteleit **elemi eseményeknek** nevezzük. A továbbiakban az elemi eseményeket ω_i-vel jelöljük, ahol az index a lehetséges elemek közül az i-ediket jelöli.

A következő fontos fogalom az eseménytér fogalma.

Definíció. Az elemi események összességét **eseménytérnek** nevezzük, és a továbbiakban az eseményteret Ω-val jelöljük, tehát

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}.$$

Látható, hogy az eseménytér egy halmaznak fogható fel, amelynek elemei az elemi események.

A véletlen eseményekkel kapcsolatban más események is megfogalmazhatók, például olyanok, amelyek több elemi eseményt tartalmaznak.

Definíció. Az elemi események egy halmazát tartalmazó eseményt **véletlen eseménynek** nevezzük. A továbbiakban a véletlen eseményeket *A, B, C* nagy latin betűkkel jelöljük. Igaz tehát az, hogy

$$A \subset \Omega,$$

vagyis halmazelméleti megfogalmazással, az *A* véletlen esemény az elemi események Ω terének egy részhalmaza.

Egy kísérlet során a véletlen eseményt akkor tekintjük megvalósultnak, ha a megvalósuló elemi esemény része az *A* halmaznak. Természetesen Ω is egy esemény, és mivel az eseménytérrel kapcsolatos mérés során Ω biztosan bekövetkezik, ezért ezt az eseményt **biztos eseménynek** nevezzük.

Például a kockadobás esetén az eseménytér:

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

A páros számokat tartalmazó véletlen esemény:

$$A = \{2, 4, 6\},$$

ahol az elemi események $\omega_1 = 2$, $\omega_2 = 4$, $\omega_3 = 6$. A kockadobásos kísérlet során az *A* esemény akkor valósul meg, ha valamelyik páros számot dobjuk.

Az eddigiekből már látszik, hogy az elemi halmazelmélet fogalmait és jelöléseit fogjuk használni. *Ha valaki ezekre nem emlékszik, akkori itt az ideje a téma átnézésének (lásd a Függelékben a 13. fejezetet)!*

Haladóknak

Felvetődhet az a kérdés, hogy egy eseménytérrel kapcsolatban hány esemény fogalmazható meg? Nézzünk egyszerű példát! Egy urnában négy golyó van, *1, 2, 3, 4* számokkal ellátva:

$$\Omega = \{1, 2, 3, 4\}.$$

A kísérlet abból állhat, hogy véletlenszerűen golyókat húzunk az urnából. Írjuk fel az összes lehetséges eseményt az elemi eseményeikkel:

$$\{0\};\ \{1\};\ \{2\};\ \{3\};\ \{4\};\ \{1,2\};\ \{1,3\};\ \{1,4\};\ \{2,3\};\ \{2,4\};\ \{3,4\};\ \{1,2,3\};\ \{1,2,4\};\ \{1,3,4\};\ \{2,3,4\};\ \{1,2,3,4\}.$$

A halmazelméletben szokásos, hogy a nulleseményt (jelen esetben, hogy nem húzunk golyót) minden halmaz részhalmazának tekintjük. Az összes véletlen események halmaza tehát ebben az esetben *16* eseményt tartalmaz.

Gyors ellenőrző feladat

3.1. Lássuk be, hogy általában igaz az, hogy ha Ω -nak n elemi eseménye van, akkor az összes lehetséges események száma: 2^n.

Műveletek eseményekkel

Láttuk, hogy az események halmazok, ezért minden olyan művelet elvégezhető az eseményekkel, amely a halmazok között definiálva van.

Tehát például a $\sum_{i=1}^{n} A_i$ összegesemény az az esemény, melynek során legalább az egyik A_i esemény megvalósul. Vagy, a $\prod_{i=1}^{n} A_i$ szorzatesemény az az esemény, melynek során valamennyi A_i megvalósul. Ha $\overline{A}$ (A komplementer eseménye) megvalósul, akkor A nem valósul meg stb.

Gyors ellenőrző feladatok

3.2. Legyen A és B két eseményhalmaz. Ezek olyan halmazok, amelyeknek van közös részük, vagyis $AB \neq \emptyset$. Írjuk fel összegüket két közös rész nélküli halmaz összegeként.
3.3. Milyen feltételekkel igaz az egyenlőség:

$$(A+B) - B = A.$$

3.2. Gyakoriság, relatív gyakoriság, empirikus nagy számok törvénye

A leíró statisztikáról szóló fejezetben már megismerkedtünk a gyakoriság és a relatív gyakoriság fogalmával. Elevenítsük fel, hogy a kísérletek során milyen tapasztalatokat szereztünk a gyakorisággal kapcsolatban, és nézzük meg a gyakoriságok néhány további tulajdonságát!

Legyen az eseményterünk:

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}.$$

Láttuk korábban, hogy ha véletlen kísérletet végzünk az eseménytéren, akkor a végeredmény az elemi események közül az egyik. Egyetlen kísérlet esetén ennél nagyobb bizonyosság nincs! Azonban, ha sokszor (n-szer) hajtjuk végre a kísérletet, és ennek során az A esemény gyakorisága k_A, a B esemény gyakorisága k_B, akkor a kísérleti tapasztalat azt mutatja, hogy a k_A/k_B hányados, elég nagy n esetén, viszonylagos stabilitást mutat. Ez azt jelenti, hogy a gyakoriságok hányadosa a kísérletszám függvényében már csak kicsit változik. Ezeket a kis változásokat statisztikus ingadozásoknak nevezzük. Ez a viszonylagos stabilitás az, amit **empirikus nagy számok törvényének** neveztünk. A tapasztalatai alapján mindenki által könnyen belátható példa az érmedobás esete. Nagyszámú dobás esetén azt várjuk, hogy közel azonos számú fej és írás lesz a végeredmény, azaz a két esemény gyakoriságának hányadosa:

$$\frac{k_A}{k_B} = \frac{k_{fej}}{k_{írás}} \approx 1 .$$

Az Ω biztos esemény gyakorisága megegyezik a mérések számával. Ha tehát a B esemény helyébe a biztos eseményt tesszük, akkor a hányados alakja:

$$\frac{k_A}{k_\Omega} = \frac{k_A}{n} .$$

Ezt a hányadost neveztük relatív gyakoriságnak. Az így kapott hányados is stabilitást mutat az n mérésszám növekedésével.

A gyakoriság száma nem lehet nagyobb a mérések számánál. Igaz tehát az, hogy

$$0 \le k_A \le n .$$

Ennek az egyenlőtlenségnek az átrendezésével azt kapjuk, hogy

$$0 \le \frac{k_A}{n} \le 1 . \qquad (3.2.1.)$$

A relatív gyakoriság tehát 0 és 1 közötti értéket vehet fel, a határokat is beleértve.

Mit tudunk mondani két esemény összegének relatív gyakoriságáról? Legyen a két esemény A és B. n kísérlet során a gyakoriságuk k_A és k_B. A két esemény összege $A+B$, az összeg gyakorisága k_{A+B}. Általában igaz az, hogy

$$k_{A+B} \le k_A + k_B , \qquad (3.2.2.)$$

hiszen, ha a két eseményhalmaznak van a nulleseménytől különböző közös része, akkor azt k_A és k_B értékébe is beleszámítjuk, ugyanakkor az ilyen események k_{A+B} számolásakor csak egyszer számolandók.

Nézzünk egy egyszerű példát!

3.1. Példa. A kockadobás estén az A esemény legyen a következő:

$$A = \{1,2,3\},$$

a B esemény pedig:

$$B = \{3,4,5\}.$$

Az $A+B$ esemény pedig a következő:

$$A+B=\{1,2,3,4,5\}.$$

Végezzünk el egy $n=10$ dobásból álló kockadobásos kísérletet, melynek során az alábbi elemi események következnek be:

$$1, 1, 3, 6, 5, 5, 3, 2, 2, 4.$$

Számoljuk meg, hányszor következett be az A, a B és az $A+B$ esemény! A következőket kapjuk:

$$k_A=6,\ k_B=5,\ k_{A+B}=8,$$

vagyis valóban a (3.2.2.) kifejezésben a kisebb reláció teljesül, melynek az oka az, hogy a közös elemeket az egyenlőtlenség jobb oldalán kétszer számoljuk.

Más a helyzet azonban akkor, ha a két halmaz diszjunkt, azaz nincs közös részük $(AB=\emptyset)$. Ekkor és csak ekkor, igaz az, hogy

$$k_{A+B}=k_A+k_B. \qquad (3.2.3.)$$

Ha a (3.2.4.) egyenlet mindkét oldalát osztjuk n-nel, akkor a relatív gyakoriságokra érvényes összefüggésre jutunk:

$$\frac{k_{A+B}}{n}=\frac{k_A}{n}+\frac{k_B}{n}. \qquad (3.2.4.)$$

Szavakban megfogalmazva: ha két eseménynek nincs közös része (diszjunkt halmazok), akkor összegük relatív gyakorisága megegyezik a relatív gyakoriságaik összegével.

Gyors ellenőrző feladat

3.4. Lássuk be, hogy a (3.2.3) összefüggés igaz tetszőleges számú, közös résszel nem rendelkező esemény gyakoriságára is! Alkalmazzuk a teljes indukció módszerét!

3.3. A valószínűség kísérleti meghatározása

Az előző fejezetben a relatív gyakorisággal kapcsolatban leírt törvényszerűségek alkalmasak a valószínűség fogalmának definiálására.

Mivel a relatív gyakoriság nagyszámú kísérlet esetén viszonylagos stabilitást mutat, ezért a véletlen kísérletek során célszerű ezt a mennyiséget az adott esemény jellemzésére használni. Azt mondjuk, hogy egy véletlen esemény valószínűségének kísérleti értéke az adott esemény nagyszámú kísérlet esetén bekövetkező relatív gyakorisága. Jelöljük az A esemény kísérleti valószínűségi értékét $P_{kis}(A)$-val, akkor

$$P_{kis}(A) = \frac{k_A}{n}\text{, ha } n \text{ elegendően nagy.} \qquad (3.3.1.)$$

Hogy mi az elegendően nagy n érték, azt adott kísérletben a pontosság igény szabja meg. A későbbiekben még visszatérünk erre a kérdésre. Az így definiált valószínűség értékre igaz, hogy

$$0 \le P_{kis}(A) \le 1\,, \qquad (3.3.2.)$$

vagyis, ha így definiáljuk a valószínűség kísérleti értékét, akkor egyúttal a skálát is megszabtuk. Nem feltétlenül kell azonban a skálának *0* és *1* közötti értékűnek lennie. A gyakorlatban sokszor százalékban fejezik ki a valószínűség értékét, ami azt jelenti, hogy a skála *0* és *100* közötti. Ilyenkor tulajdonképpen a (3.3.1.) definícióval kapott értéket százzal szorozzuk.

A definícióból az is látszik, hogy mivel n kísérlet során a biztos esemény n-szer következik be, ezért a biztos esemény kísérleti valószínűsége *1*.

Eddigi meggondolásaink kísérleti tapasztalatokra vonatkoztak. Ezek a megfigyelések teremtették meg az alapját annak, hogy a valószínűségelmélet axiomatikus felépítése megkonstruálható legyen.

3.4. A valószínűségelmélet axiómái

A. N. Kolmogorov orosz matematikus 1933-ban fektette le a valószínűségelmélet axiomatikus alapjait. Ez tette lehetővé, hogy a valószínűségelmélet a matematika biztos alapokkal rendelkező része lehessen. Az általa felírt három axióma az alábbi.

1. axióma. Minden $A \in \Omega$ véletlen eseményhez hozzárendelhető egy *P(A)* szám, amelyre igaz, hogy

$$0 \le P(A) \le 1\,. \qquad (3.4.1.)$$

A *P(A)* függvényt az *Ω* eseménytér *A* részhalmazain értelmezett valószínűség-eloszlásnak, vagy röviden az *A* esemény valószínűségének nevezzük.

2. axióma. A biztos esemény valószínűsége *1*, vagyis

$$P(\Omega) = 1\,. \qquad (3.4.2.)$$

3. axióma. Ha A_i véges sok, vagy megszámlálhatóan végtelen sok esemény, ahol $A_i A_j = \emptyset$ ha $i \ne j$, akkor

$$P(A_1 + A_2 + A_3 + ...) = P(A_1) + P(A_2) + P(A_3) + ...$$

Tömörebben felírva ugyanezt:

$$P\left(\sum_i A_i\right) = \sum_i P(A_i). \tag{3.4.3.}$$

A 3. axióma alakja két eseményre felírva: ha $A_1 A_2 = \emptyset$, akkor

$$P(A_1 + A_2) = P(A_1) + P(A_2).$$

Az axiómák lényegéből következik, hogy felírásuk önkényes. Azonban, ha a belőlük következő tételeket a valóság leírására kívánjuk felhasználni, akkor ez az önkényesség korlátozódik. Vegyük észre, hogy a valószínűség axiómái is korábbi kísérleti tapasztalatokra alapozódnak. A valószínűség axiómái olyan összefüggések, amelyek kísérletileg bizonyítható módon a relatív gyakoriságra igazak. Ha valószínűségnek azt az értéket tekintjük, ami körül nagy mérésszám esetén statisztikusan ingadozik a relatív gyakoriság, akkor az axióma megfogalmazásának újdonsága abban áll, hogy feltételezi ezeknek az összefüggéseknek az igazát erre az értékre is.

Látjuk tehát, hogy az axiómák az A eseményekhez egy-egy számot (a valószínűséget) rendelik hozzá. Ez az

$$A \to P(A)$$

hozzárendelés hasonló a valós számok közében megismert $x \to f(x)$ függvénykapcsolathoz. Az A eseményeket és a hozzájuk rendelt $P(A)$ számértékeket együttesen **valószínűségi mezőnek** nevezzük.

3.5. Az axiómák következményei

Az axiómákból a halmazelmélet és a matematikai logika segítségével olyan összefüggéseket vezetünk le, amelyeket a valószínűség-számítás alaptételeinek nevezhetünk. Mielőtt a tételekre rátérnénk, definiáljunk egy új fogalmat.

Definíció. Az $A_1, A_2, ..., A_n, ...$ események rendszerét **teljes eseményrendszernek** nevezzük, ha az események egymást páronként kizárják, és összegük a biztos eseményt adja. Más szóval

$$A_1 + A_2 + ... + A_n + ... = \Omega, \quad A_i A_k = \emptyset, \text{ ha } i \neq k. \tag{3.5.1.}$$

Megjegyzés: vegyük észre, hogy az eseményrendszer tartalmazhat véges számú eseményt, de megszámlálhatóan végtelen eleme is lehet.

Például a kockadobás esetén a páros számokat tartalmazó A eseményhalmaz, és a páratlan számokat tartalmazó B eseményhalmaz teljes eseményrendszert alkot:

$$A=\{2,4,6\} \text{ és } B=\{1,3,5\}.$$

Ezek után rátérhetünk az axiómákból következő egyszerű tételek megfogalmazására.

1. tétel. Ha az $A_1, A_2, \ldots, A_n$ események teljes eseményrendszert alkotnak, akkor igaz az, hogy:

$$P(A_1)+P(A_2)+\ldots+P(A_n)+\ldots=1. \qquad (3.5.2.)$$

Bizonyítás. A teljes eseményrendszer (3.5.1.) definíciója tartalmazza azt, hogy a rendszer elemei diszjunkt halmazok. Ezért alkalmazható a 3. axióma, vagyis

$$P(A_1+A_2+\ldots+A_n+\ldots)=P(A_1)+P(A_2)+\ldots+P(A_n)+\ldots=P(\Omega).$$

A második axióma szerint viszont $P(\Omega)=1$, tehát

$$P(A_1)+P(A_2)+\ldots+P(A_n)+\ldots=1,$$

és ez az, amit be kellett látnunk.

2. tétel. Az A esemény $\overline{A}$ komplementer eseményének valószínűsége:

$$P(\overline{A})=1-P(A). \qquad (3.5.3.)$$

Bizonyítás. A halmazelméletből ismeretes, hogy

$$A+\overline{A}=\Omega,$$

valamint

$$A\overline{A}=\emptyset.$$

Tehát, A és $\overline{A}$ teljes eseményrendszert alkotnak! Ezért az 1. tétel (3.5.2.) állításából következik, hogy

$$P(A)+P(\overline{A})=1.$$

Innen pedig már leolvasható a tétel állítása:

$$P(\overline{A}) = 1 - P(A).$$

3. tétel. A lehetetlen esemény valószínűsége nulla, vagyis

$$P(\emptyset) = 0. \tag{3.5.4.}$$

Bizonyítás. A halmazelméletből ismeretes, hogy

$$\overline{\Omega} = \emptyset,$$

tehát

$$P(\overline{\Omega}) = P(\emptyset)$$

A 2. tétel felhasználásával:

$$P(\emptyset) = P(\overline{\Omega}) = 1\text{-}P(\Omega) = 0.$$

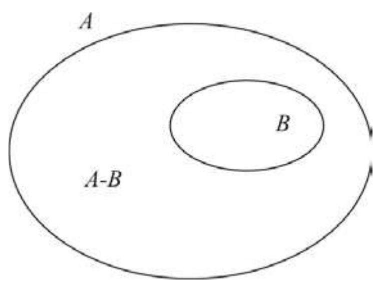
4. tétel. Ha $B \subset A$, akkor igaz az, hogy

$$P(A - B) = P(A) - P(B). \tag{3.5.5.}$$

Ez azt jelenti, hogy amennyiben *B* esemény része az *A* eseménynek, akkor (és csak akkor) a különbségük valószínűsége egyenlő a valószínűségeik különbségével.

Bizonyítás. A helyzetet jellemző Venn-diagramból (3.1. ábra) látszik, hogy $(A - B) + B = A$, valamint, az is látszik, hogy *B* és *A-B* közös rész nélküli halmazok, azaz

$$B(A - B) = \emptyset.$$



3.1. ábra: Segédábra a 4. tétel bizonyításához

Alkalmazható tehát a 3. axióma:

$$P(A-B)+P(B)=P(A),$$

ahonnan

$$P(A-B)=P(A)-P(B).$$

A tétel következménye az, hogy ha $B \subset A$, akkor $P(B) \leq P(A)$, hiszen $P(A-B) \geq 0$, tehát

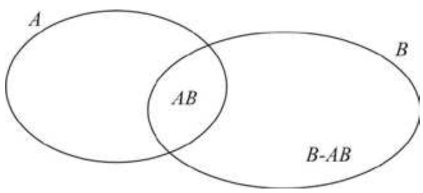
$$0 \leq P(A)-P(B),$$

ahonnan az állítás már leolvasható.

5. tétel. Ha *A* és *B* két tetszőleges véletlen esemény, akkor összegük valószínűsége:

$$P(A+B)=P(A)+P(B)-P(AB). \quad (3.5.6.)$$

Bizonyítás. Ismét rajzoljuk fel a Venn-diagramot az általános esetre. Ezt a 3.2. ábra mutatja.



3.2. ábra: Venn-diagram az 5. tétel bizonyításához

Az *A+B* halmaz felírható két közös rész nélküli halmaz összegeként, ahogy az a Venn-diagramon látszik:

$$A+B=A+(B-AB).$$

Mivel igaz az, hogy

$$A(B-AB)=\emptyset,$$

ezért alkalmazható a 3. axióma:

$$P(A+B)=P\big(A+(B-AB)\big)=P(A)+P(B-AB). \quad (3.5.7.)$$

Minthogy azonban $AB \subset B$, ezért a 4. tétel értelmében:

$$P(B - AB) = P(B) - P(AB).$$

Ha az így kapott összefüggést (3.5.7.)-be beírjuk, akkor a tétel állításához jutunk:

$$P(A + B) = P(A) + P(B) - P(AB).$$

Gyors ellenőrző feladatok

3.5. Az 5. tételhez hasonló tételt vezessünk le két tetszőleges eseményhalmaz különbségére! Lássuk be, hogy ha A és B két tetszőleges eseményhalmaz, akkor

$$P(A - B) = P(A) - P(AB)!$$

3.6. Lássuk be, hogy tetszőleges A és B eseményhalmazok esetén igaz az, hogy

$$P(A + B) \leq P(A) + P(B)!$$

A következő két alfejezetben egy-egy egyszerű valószínűségi mezővel ismerkedünk meg, amelyekben egyszerű elvek alapján számolható az események valószínűsége.

3.6. Klasszikus valószínűségi mező

Definíció. Klasszikus valószínűségi mezőnek nevezzük azt a valószínűségi mezőt, amelyben az elemi események száma véges, és valamennyinek azonos a valószínűsége.

Egyszerű példa klasszikus mezőre a pénzérme példája. Az elemi események (fej, írás) száma kettő, és ha szabályos az érme, akkor a két elemi esemény valószínűsége azonos.

Tétel. Ha a klasszikus valószínűségi mezőnek n elemi eseménye van, akkor ezek p valószínűsége:

$$p = \frac{1}{n}. \tag{3.6.1.}$$

Bizonyítás. A klasszikus mező elemi eseményei legyenek $\omega_1, \omega_2, \ldots, \omega_n$. A klasszikus mező definíciójából következik, hogy

$$P(\omega_1) = P(\omega_2) = \ldots = P(\omega_n) = p.$$

Továbbá, mivel az elemi események egymást kizáró események, vagyis $\omega_i\ \omega_j=\emptyset$, tehát alkalmazható a 3. axióma:

$$P(\Omega)=P(\omega_1+\omega_2+\ldots+\omega_n)=P(\omega_1)+P(\omega_2)+\ldots+P(\omega_n)=np=1.$$

Innen az utolsó egyenlőség felhasználásával kapjuk, hogy

$$p=\frac{1}{n}.$$

A klasszikus mező n elemi eseményt tartalmazó A eseményének valószínűségéről szól a következő tétel.

Tétel. Ha a klasszikus valószínűségi mezőben az A esemény k darab elemi eseményt tartalmaz, akkor ennek valószínűsége:

$$P(A)=\frac{k}{n}. \tag{3.6.2.}$$

Bizonyítás. Az előző tételhez hasonlóan kihasználjuk, hogy az elemi események egymást kizáró események, tehát alkalmazható a 3. axióma. Ha az A esemény, és a részét képező elemi események az alábbiak:

$$A=\omega_1+\omega_2+\ldots+\omega_k,$$

akkor

$$P(A)=P(\omega_1+\omega_2+\ldots+\omega_k)=P(\omega_1)+P(\omega_2)+\ldots+P(\omega_k)=k\frac{1}{n}=\frac{k}{n}.$$

A gyakorlatban, ha alkalmazni akarjuk a (3.6.2.) képletet, akkor természetesen meg kell győződni arról, hogy valóban klasszikus mezőről van-e szó, vagyis, hogy valami indokolja-e azt, hogy az elemi események valószínűségei azonosak.

3.1. Példa. Példaként számoljuk ki annak valószínűségét, hogy a játékkockával páratlan számot dobunk. Ha a játékkocka szabályos, akkor egyik oldala sincs kitüntetve a másikhoz képest. Ez indokolja azt, hogy klasszikus mezőről van szó, vagyis az elemi események számának végessége mellett valamennyi oldal azonos valószínűséggel kerülhet felülre. Az összes lehetőség száma *6,* és ezek között az A esemény részeként három páratlan szám szerepel. Alkalmazva a (4.17) képletet azt kapjuk, hogy

$$P(A)=\frac{\textit{a kedvező elemi események száma}}{\textit{az összes elemi esemény száma}}=\frac{3}{6}=\frac{1}{2}.$$

Gyors ellenőrző feladatok

3.7. Két érmét dobunk fel egyszerre. Mi a valószínűsége annak, hogy mind a két érme írással felfelé ér le?

3.8. Egy urnában három piros, két fehér és öt zöld golyó van. Mi a valószínűsége annak, hogy ha egy golyót véletlenszerűen húzunk ki az urnából, akkor az fehér lesz?

3.7. Geometriai valószínűségi mező

Sok esetben az Ω biztos esemény valamilyen geometriai alakzat (görbe, felület, térfogat, stb.). Ez nem klasszikus valószínűségi mező, hiszen az elemi események az Ω halmaz pontjai, ezek száma pedig végtelen. Az A események Ω részhalmazai ($A \subset \Omega$). Hogyan számoljuk ki az A esemény $P(A)$ valószínűségét?

A valószínűség kiszámításához feltételezzük, hogy az A esemény $P(A)$ valószínűsége arányos az A halmaz mértékével (a hosszával, a felület nagyságával, a köbtartalommal), azaz:

$$P(A) = c \cdot m(A), \tag{3.7.1.}$$

ahol c arányossági tényező, $m(A)$ pedig az A halmaz mértékét jelöli. Az axiómák alapján az arányossági tényező nagysága kiszámítható, hiszen a (3.7.1) összefüggésben A helyett Ω-t írva azt kapjuk, hogy

$$P(\Omega) = cm(\Omega).$$

A 2. axióma szerint $P(\Omega)=1$, tehát

$$cm(\Omega) = 1,$$

ahonnan

$$c = \frac{1}{m(\Omega)}.$$

Végeredményképpen tehát az A esemény valószínűsége:

$$P(A) = \frac{m(A)}{m(\Omega)}. \tag{3.7.2.}$$

Bár a geometriai valószínűségi mező az említett ok miatt nem klasszikus mező, de a valószínűségek kiszámítását illetően hasonló kifejezésre jutottunk, mint a klasszikus mező esetében, vagyis

$$valószínüség = \frac{kedvező\ esemény\ mértéke}{biztos\ esemény\ mértéke}.$$

Megjegyzés

A geometriai valószínűség esetén az elemi események az Ω eseményteret megjelenítő geometriai alakzat pontjai. Mekkora ezeknek az elemi eseményeknek (pontoknak) a valószínűsége? A (3.7.2.) képlet alapján látszik, hogy bármely pont valószínűsége *0,* hiszen a számlálóban szerepel az esemény (jelen esetben a pont) mértéke. A pont mértéke pedig zérus. Nullától különböző valószínűsége a végtelen sok pontot tartalmazó tartománynak van, ami lehet hosszúság, felület vagy térfogat.

Lássunk egy egyszerű példát a geometriai valószínűségre!

3.2. Példa. Tegyük fel, hogy véletlenszerűen lövünk egy R sugarú céltáblára. Feltéve, hogy a céltáblát biztosan eltaláljuk, mi a valószínűsége annak, hogy a középen lévő r sugarú *10*-es körbe találunk?

Megoldás:

$$m(A) = r^2\pi\,,\ m(\Omega) = R^2\pi\,.$$

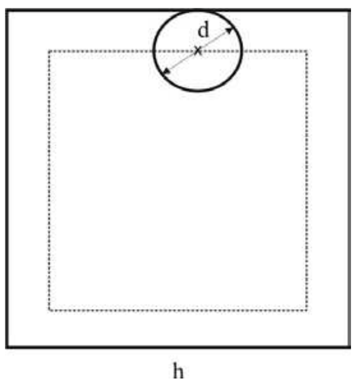
Tehát:

$$P(A) = \frac{m(A)}{m(\Omega)} = \frac{r^2\pi}{R^2\pi} = \frac{r^2}{R^2}\,. \tag{3.7.3.}$$

A 18. században Georges Buffon volt az, aki a geometriai valószínűség problémáival foglalkozott. Ő adta fel a következő példát.

3.3. Példa. h oldalhosszúságú négyzetekre beosztott lapra d átmérőjű érmét dobunk, ahol $d < h$. A kérdés az, hogy mi a valószínűsége annak, hogy az érme hozzáér a négyzethálót alkotó vonalakhoz (legyen ez az A esemény)?

A 3.3. ábrára a h oldalhosszúságú négyzetháló egyetlen négyzetét rajzoltuk fel.



3.3. ábra: Ábra a négyzethálós példához

A probléma a geometriai valószínűség módszerével oldható meg. Nyilvánvaló, hogy az érme középpontja a négyzet bármely pontjára eshet, tehát az összes eset mértéke a négyzet területe, azaz: h^2. Az ábráról az is látszik, hogy az érme akkor nem érinti a négyzetháló vonalait, ha középpontja a szaggatott vonallal jelzett négyzeten belül tartózkodik. Annak valószínűsége tehát, hogy az érme nem érinti a vonalakat ($\overline{A}$ esemény):

$$P(\overline{A}) = \frac{(h-d)^2}{h^2} = \frac{h^2 - 2hd + d^2}{h^2} = 1 - \frac{2hd - d^2}{h^2}. \tag{3.7.4.}$$

Felhasználva a $P(A) = 1 - P(\overline{A})$ összefüggést, a keresett valószínűség:

$$P(A) = \frac{2hd - d^2}{d^2}.$$

Megjegyzés

A példában a megoldáshoz elegendő egyetlen négyzetet tekinteni, ugyanis akárhány négyzetet is tekintünk a (3.7.4.) kifejezésben ugyanakkora számmal szorozzuk a számlálót és a nevezőt is, tehát az eredmény változatlan marad.

Gyors ellenőrző feladat

3.9. Legyen a sík felosztva egymástól h távolságra lévő párhuzamos vonalakkal. Ejtsünk a síkra d átmérőjű érmét ($d < h$). Mi a valószínűsége annak, hogy az érme érinti a vonalak valamelyikét? (d/h).

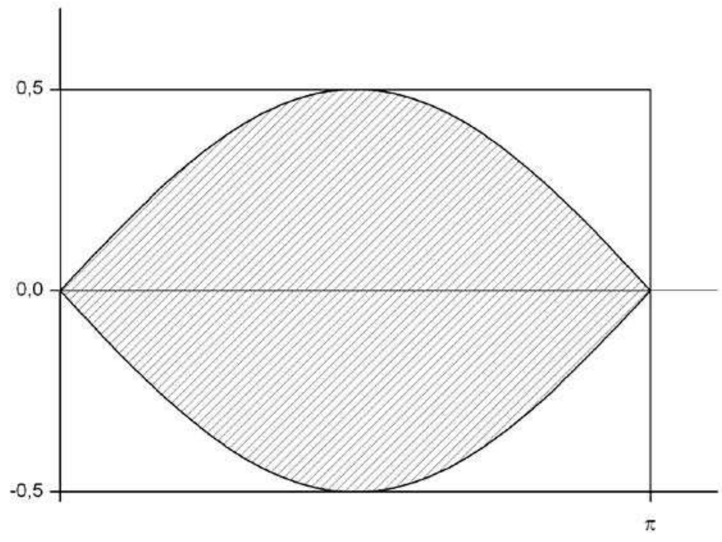
Haladóknak

3.4. Példa. Buffon a fentieknél érdekesebb feladatot is kitalált és megoldott. A feladat így szól: húzzunk a síkon egymástól egységnyi távolságra párhuzamos vonalakat. Ejtsünk erre az egyszerűség kedvéért egységnyi hosszúságú tűt. Mi a valószínűsége annak, hogy a tűnek van közös pontja valamelyik vonallal? Ez a probléma Buffon tűje néven vonult be az irodalomba.

A megoldás során ismét elegendő egyetlen vonalat tekinteni, és ennek a vonalnak két oldalról az $1/2$ szélességű környezetét. Az egyszerű leíráshoz két változót vezethetünk be. Az egyik a tű középpontjának távolsága a vonaltól. Legyen ez az x változó. Az x változó $-1/2$-től $+1/2$-ig változhat, A másik változó legyen a tű φ szöge a vonalhoz képest. A φ szög 0 radiántól π radiánig változhat. Adott $x > 0$ esetén az érintés feltétele az, hogy

$$\sin\varphi \geq \frac{x}{1/2}, \text{ azaz } \frac{1}{2}\sin\varphi \geq x. \tag{3.7.5.}$$

Ugyanez igaz negatív x-ekre is.



3.4. ábra: Ábra a Buffon-tű feladat megoldásához

A 4.4. ábra mutatja, hogy a „kedvező” terület a besatírozott rész. Ennek nagyságát a görbe alatti terület számításával határozhatjuk meg:

$$t = 2\int_0^{\pi} \frac{1}{2} \sin x \, dx = \left[-\cos x\right]_0^{\pi} = 2 .$$

A teljes terület nagysága pedig: π. A geometria valószínűség szerint annak valószínűsége, hogy a tű érintkezik vonalakkal (*A* esemény):

$$P(A) = \frac{\text{kedvező terület}}{\text{teljest terület}} = \frac{2}{\pi} = 0{,}636619\ldots . \qquad (3.7.6.)$$

Ez az eredmény érdekes lehetőséget ad a π értékének kísérleti becslésére, hiszen (3.7.6.) átalakításával azt kapjuk, hogy

$$\pi = \frac{2}{P(A)} . \qquad (3.7.7.)$$

Ha tehát kísérletileg meghatározzuk $P_{kis}(A)$ értékét, akkor (3.7.7.) kifejezésből adódik π értéke. Azóta többen elvégezték ezt a kísérletet, többek között 1901-ben Mario Lazzarini olasz matematikus *3 408* számú dobást végzett, amivel meglehetős pontossággal tudta a π értékét becsülni. Ma már nem kell tűk dobálásával kísérleteznünk, hiszen számítógépes szimulációval gyorsan elvégezhető a becsléshez szükséges nagyszámú dobás. *

3.8. Feltételes valószínűség

A feltételes valószínűség fogalmának definiálása előtt nézzünk egy egyszerű példát. Számoljuk ki, hogy mekkora a valószínűsége annak, hogy a játékkockával páratlan számot dobunk (*A* esemény). A klasszikus valószínűségnél mondottak alapján ezt egyszerűen ki tudjuk számolni:

$$P(A)=\frac{\text{kedvező események száma}}{\text{összes esemény száma}}=\frac{3}{6}=\frac{1}{2}.$$

Más a helyzet azonban, ha tudjuk azt, hogy a dobás során *4*-nél kevesebbet dobtunk (*B* esemény), vagy másként fogalmazva, csak az *4*-nél kisebb értékű dobásokat tekintjük. Ilyen feltétel mellett, mi a páratlan dobások (*A* esemény) valószínűsége? Az összes események száma most *3,* a kedvező események száma pedig *2*. Tehát, ha $P(A|B)$-vel jelöljük az *A* esemény valószínűségét abban az esetben, ha a *B* esemény biztosan bekövetkezik, akkor a keresett valószínűség:

$$P(A|B)=\frac{2}{3}. \tag{3.8.1.}$$

Általánosabban is felírhatjuk az eredményt, ha k_B-vel jelöljük a *B* esemény elemi eseményeinek számát, k_{AB}-vel pedig a *B* eseménnyel együtt bekövetkező *A* események számát, akkor a fenti képlet általános megfogalmazása:

$$P(A|B)=\frac{k_{AB}}{k_B}.$$

Ez a képlet nagyon hasonlít a klasszikus valószínűség kifejezésére, csak az összes elemi események számát átveszi a *B* esemény elemi eseményeinek száma.

Általánosítható kifejezésre jutunk, ha elosztjuk a számlálót és a nevezőt is *n*-el, ami az eseménytér összes elemi eseményének a száma. Ekkor azt kapjuk, hogy

$$P(A|B)=\frac{k_{AB}/n}{k_B/n}=\frac{P(AB)}{P(B)}.$$

Figyelem! A számlálóban és a nevezőben a hányadosok nem relatív gyakoriságok, hanem események elemi eseményeinek száma osztva az összes elemi esemény számával.

---

* A Buffon-tű problémát szimuláló animáció azt mutatja, hogy a kísérletek számának növekedésével a megoldás hogyan közelíti a π értékét.

A klasszikus valószínűségi mezőre vonatkozó könnyen átlátható feladat után most már nem meglepő a feltételes valószínűség definíciója.

Definíció. Tetszőleges Ω eseménytéren legyen két eseményünk A és B. Az A esemény B eseményre vonatkoztatott **feltételes valószínűségét** a következő képlet adja meg:

$$P(A|B)=\frac{P(AB)}{P(B)}\text{, ahol megköveteljük, hogy } \left(P(B)>0\right). \qquad (3.8.2.)$$

Haladóknak

Könnyen belátható, hogy $P(A|B)$ feltételes valószínűség egy valószínűség-eloszlást ír le a rögzített B esemény AB részhalmazain. $P(A|B)$ ugyanis eleget tesz az alábbi összefüggéseknek:

1. $0\le P(A|B)\le 1$.

Ez onnan következik, hogy $AB\subset B$, ezért a 4. tétel következménye alapján:

$$P(AB)\le P(B),$$

vagyis

$$0\le P(A|B)=\frac{P(AB)}{P(B)}\le\frac{P(B)}{P(B)}=1.$$

2. $P(B|B)=1$, vagyis a biztos esemény szerepét B veszi át, hiszen

$$P(B|B)=\frac{P(BB)}{P(B)}=\frac{P(B)}{P(B)}=1.$$

3. $P\left((A_1+A_2)|B\right)=P\left(A_1|B\right)+P\left(A_2|B\right)$, ha az A_1 és A_2 események kizárják egymást.

Bizonyítás. A bizonyítás első lépéseként lássuk be, hogy ha A_1 és A_2 események kizárják egymást, akkor A_1B és A_2B események is kizárják egymást. Felírható az alábbi azonosság

$$(A_1B)(A_2B)=B(A_1A_2)B, \qquad (3.8.3.)$$

ahol a halmaz szorzás kommutatív és asszociatív tulajdonságait használtuk ki. Mivel pedig $A_1A_2=\emptyset$, ezért

$$B(A_1A_2)B=B\emptyset B=\emptyset.$$

Ezek után felírhatjuk a feltételes valószínűség definíciója alapján azt, hogy

$$P\left((A_1+A_2)\middle|B\right)=\frac{P\left((A_1+A_2)B\right)}{P(B)}=\frac{P\left(A_1B+A_2B\right)}{P(B)}=\frac{P(A_1B)+P(A_2B)}{P(B)}=\frac{P(A_1B)}{P(B)}+\frac{P(A_2B)}{P(B)},$$

ahonnan már leolvasható az állítás.

Ez a három állítás lényegében megegyezik a valószínűség három axiómájával, ami azt is jelenti, hogy a feltételes valószínűség definíciójával valóban valószínűség eloszlást definiáltunk. A feltételes valószínűségre is érvényesek tehát a valószínűségre levezetett eddigi tételek.

Gyors ellenőrző feladatok

3.10. A harmadik állítás általánosításaként teljes indukcióval lássuk be, hogy ha A_1, A_2... véges számú, vagy végtelen sok esemény páronként kizárja egymást, akkor igaz az, hogy

$$P\left(\sum_i A_i \middle| B\right)=\sum_i P(A_i|B)\,!$$

3.11. Lássuk be, hogy

$$P(\overline{A}|B)=1-P(A|B)\,!$$

3.12. Lássuk be, hogy két tetszőleges A_1 és A_2 eseményre igaz, hogy

$$P\left((A_1+A_2)\middle|B\right)=P(A_1|B)+P(A_2|B)-P((A_1A_2)|B).$$

3.5. Példa. A sok kétgyermekes család közül véletlenszerűen választunk egyet. Mi a valószínűsége annak, hogy a családban van fiú is (A esemény), ha azt tudjuk, hogy legalább az egyik gyermek lány (B esemény)?

Két gyermek esetén a gyermekeket illetően az összes lehetséges a következő:

$$ff,\ fl,\ lf,\ ll.$$

A B esemény (van lány a családban) valószínűsége: $P(B)=\frac{3}{4}$. Annak valószínűsége, hogy lány és fiú is van a családban: $P(AB)=\frac{2}{4}=\frac{1}{2}$. A feltételes valószínűség képlete alapján tehát annak valószínűsége, hogy van fiú a családban, ha tudjuk, hogy van leány a családban:

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{\frac{2}{4}}{\frac{3}{4}}=\frac{2}{3}.$$

3.9. Szorzási szabály

A feltételes valószínűség alkalmazása során a feladat általában nem abban áll, hogy *P(AB)* és *P(B)* ismeretében meghatározzuk $P(A|B)$ feltételes valószínűséget. Éppen ellenkezőleg, általában sokkal egyszerűbb kiszámolni a feltételes valószínűségeket, és ebből számoljuk más események valószínűségeit.

Ha a (3.8.2.) összefüggés nevezőjével átszorzunk a másik oldalra, akkor azt kapjuk, hogy

$$P(AB) = P(A|B)P(B). \tag{3.9.1.}$$

Ez a kifejezés a **szorzási szabály,** amely lehetővé teszi, hogy *P(B)* és $P(A|B)$ ismeretében kiszámoljuk *P(AB)* értékét.

3.6. Példa. A 3.5. példa esetében könnyen kiszámolható közvetlenül is a $P(A|B) = \frac{2}{3}$ feltételes valószínűség. A *B* esemény valószínűsége (legalább az egyik gyermek lány): $P(B) = \frac{3}{4}$. Innen a szorzási szabállyal határozható meg az *AB* szorzatesemény valószínűsége:

$$P(AB) = \frac{2}{3}\frac{3}{4} = \frac{1}{2}.$$

3.10. A teljes valószínűség tétele

A teljes valószínűség tétele jelentőségének megértéséhez induljunk ki egy példából.

3.7. Példa. Tegyük fel, hogy a *H1N1* vírusfertőzés kimutatására kifejlesztenek egy teszteljárást. A teszteljárás bevizsgálása során arra jutnak, hogy vírusfertőzés megléte esetén a teszt *80%* valószínűséggel[1] mutatja ki a fertőzés meglétét. Ugyanakkor, vírusfertőzés hiánya esetén *10%*-os valószínűséggel pozitív eredményt mutat a teszt. Fordítsuk le ezt a valószínűség-számítás nyelvére! A teszt pozitív eredménye legyen a *T* esemény, a vírusfertőzés megléte pedig legyen a *V* esemény! Tehát, ha nincs vírusfertőzés, akkor ez a $\overline{V}$ (komplementer) esemény. Felírható, hogy a *T* esemény kétféleképpen következhet be:

$$T = TV + T\overline{V}. \tag{3.10.1.}$$

[1] A feladat kedvéért kitalált számérték ez és a többi valószínűség is ebben a feladatban.

Mivel V és $\overline{V}$ együttesen teljes eseményrendszert alkotnak $(V+\overline{V}=\Omega)$, ezért több lehetőség nincs. (3.10.1.) alapján felírjuk annak valószínűségét, hogy a teszt pozitív eredményt ad (T esemény):

$$P(T)=P(TV)+P(T\overline{V}). \qquad (3.10.2.)$$

A 3. axióma alkalmazható, hiszen V és $\overline{V}$ egymást kizáró események, tehát TV és $T\overline{V}$ is azok – lásd a (13.10.1.) bizonyítást!

Ezek után a szorzási szabály felhasználásával kapjuk az alábbi összefüggést:

$$P(T)=P(T|V)P(V)+P(T|\overline{V})P(\overline{V}). \qquad (3.10.3.)$$

Ha tudjuk azt, hogy a vírusfertőzöttség 2%-os valószínűségű, akkor

$$P(V)=0{,}02\,, \quad P(\overline{V})=0{,}98\,, \quad P(T|V)=0{,}8\,, \quad P(T|\overline{V})=0{,}1\,.$$

Innen (3.10.3.) felhasználásával kiszámíthatjuk annak valószínűségét, hogy a teszt során pozitív lesz az eredmény:

$$P(T)=0{,}8\cdot 0{,}02+0{,}1\cdot 0{,}98=0{,}114\,.$$

A (3.10.3.) összefüggés lényegében a teljes valószínűség tétele, alkalmazva egy konkrét példára. Most már megfogalmazhatjuk általánosan is a tételt.

Tétel

Legyen A az Ω eseménytéren egy tetszőleges véletlen esemény. Ugyanezen a téren a B_1, B_2, ..., B_n események alkossanak teljes eseményrendszert, tehát

$$B_1+B_2+\ldots+B_n=\Omega\,, \text{ és } B_iB_j=\varnothing\,, \text{ ha } i\neq j\,.$$

A **teljes valószínűség tétele** azt állítja, hogy ilyen eseményekre igaz az alábbi összefüggés:

$$P(A)=P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+\ldots+P(A|B_n)P(B_n)=\sum_{i=1}^{n}P(A|B_i)P(B_i). \qquad (3.10.4.)$$

Bizonyítás

Abból indulunk ki, hogy ha B_i és B_j események $i\neq j$ esetén kizárják egymást (diszjunkt halmazok), akkor AB_i és AB_j is ilyenek – lásd (3.10.1.).

Másrészt

$$A=A\Omega=A(B_1+B_2+\ldots+B_n)=AB_1+AB_2+\ldots+AB_n\,, \qquad (3.10.5.)$$

Tehát az A esemény valószínűsége a 3. axióma felhasználásával:

$$P(A) = P(AB_1) + P(AB_2) + ... + P(AB_n). \qquad (3.10.6.)$$

Innen a szorzási szabály alkalmazásával kapjuk a tétel állítását:

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + ... + P(A|B_n)P(B_n) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

Lássunk egy másik példát arra, hogy miként használható a teljes valószínűség tétele!

3.8. Példa. Három urnában fehér és fekete golyók vannak. Az első urnában *2* fehér és *3* fekete, a második urnában *3* fehér és *4* fekete, a harmadik urnában *4* fehér és *5* fekete golyó van. A kísérlet abból áll, hogy kiválasztunk egy urnát és a kiválasztott urnából húzunk egy golyót. A kérdés az, hogy mi a valószínűsége annak, hogy a húzás eredménye fehér golyó?

Jelöljük az egyes urnák kiválasztásának eseményeit B_1, B_2, és B_3 eseményekkel. Legyen az urnák kihúzásának egyforma a valószínűsége, vagyis

$$P(B_1) = \frac{1}{3},\ P(B_2) = \frac{1}{3},\ P(B_3) = \frac{1}{3}.$$

A fehér golyó kihúzásának eseményét jelöljük A-val. A feltételes valószínűség fogalmának felhasználásával könnyen megadható a fehér golyó kihúzásának valószínűsége az egyes urnákból:

$$P(A|B_1) = \frac{2}{5},\ P(A|B_2) = \frac{3}{7},\ P(A|B_3) = \frac{4}{10}.$$

A teljes valószínűség tételét felhasználva kiszámítható a keresett valószínűség:

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) = \frac{2}{5}\frac{1}{3} + \frac{3}{7}\frac{1}{3} + \frac{4}{10}\frac{1}{3} = 0{,}4095...\ .$$

3.11. Bayes tétele

A tétel bevezetéseként itt is a 3.7. példát említjük. Példánkban azt a kérdést tárgyaltuk, hogy mi a valószínűsége a pozitív tesztnek. A kérdés másként is feltehető. Pozitív teszt esetén mi a valószínűsége annak, hogy vírusfertőzött a vizsgált személy, vagyis a $P(V|T)$ valószínűséget keressük.

A feltételes valószínűség definíciója alapján felírhatjuk, hogy

$$P(V|T)=\frac{P(VT)}{P(T)}=\frac{P(T|V)P(V)}{P(T)} \qquad (3.11.1.)$$

A nevezőbe beírva *P(T)* (3.10.3.) kifejezését:

$$P(V|T)=\frac{P(T|V)P(V)}{P(T)}=\frac{P(T|V)P(V)}{P(T|V)P(V)+P(T|\overline{V})P(\overline{V})}. \qquad (3.11.2.)$$

Ez Thomas Bayes által felírt megoldása a konkrét feladatnak. A számadatokkal a példa eredménye: $P(V|T)=0{,}14...$

Ezután felírhatjuk Bayes tételének általános megfogalmazását.

Tétel. Ha A az Ω eseménytéren egy tetszőleges véletlen esemény, és ugyanezen a téren a $B_1, B_2, \ldots, B_n$ események teljes eseményrendszert alkotnak, akkor a **Bayes-tétel** állítása:

$$P(B_k|A)=\frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{n}P(A|B_i)P(B_i)}. \qquad (3.11.3.)$$

Bizonyítás. A bizonyítás lépései megegyeznek az előzőekben végigszámolt feladat megoldásának lépéseivel. A feltételes valószínűség definíciója alapján:

$$P(B_k|A)=\frac{P(A|B_k)P(B_k)}{P(A)}.$$

Ide beírva *P(A)* alakját (3.10.4.) alapján:

$$P(B_k|A)=\frac{P(A|B_k)P(B_k)}{P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+\ldots+P(A|B_n)P(B_n)},$$

és ez lényegében a (3.11.3.) bizonyítandó alak.

3.12. Események függetlensége

A hétköznapi életben akkor mondjuk, hogy két esemény független, ha egyik bekövetkezése nincs hatással a másik bekövetkezésére. A valószínűségelméletben az események függetlenségét a feltételes valószínűség segítségével definiáljuk.

Definíció. Az A és B eseményeket **függetlennek** mondjuk, ha

$$P(A|B) = P(A). \qquad (3.12.1.)$$

Vagyis az A eseménynek a B eseményre vonatkoztatott feltételes valószínűsége nem függ a B eseménytől.

Az alábbiakban a (3.12.1.) definíció következményeit tárgyaljuk. A definíció szerint

$$P(A|B) = \frac{P(AB)}{P(B)} = P(A), \qquad (3.12.2.)$$

ahonnan

$$P(AB) = P(A)P(B), \qquad (3.12.3.)$$

vagyis, ha az A esemény független a B eseménytől, akkor az AB szorzatesemény valószínűsége az A és B esemény valószínűségeinek szorzatával egyezik meg.

Mit tudunk mondani a B esemény A eseményre vonatkoztatott feltételes valószínűségről, ha A független B-től? A feltételes valószínűség definíciója szerint:

$$P(B|A) = \frac{P(BA)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B). \qquad (3.12.4.)$$

Arra jutottunk, hogy ha az A esemény független a B eseménytől, akkor a B esemény is független az A eseménytől, ahogyan azt el is várjuk.

Minthogy (3.12.2.) és (3.12.4.) egyaránt a (3.12.3.) összefüggésre vezet, általában a **függetlenség feltételeként** az A és B eseményekre szimmetrikus (3.12.3.) kifejezést szoktuk használni.

Sokszor nem kettő, hanem több esemény függetlenségének kérdése merül fel. Ezzel kapcsolatos a következő definíció.

Definíció. Az ugyanazon az eseménytéren értelmezett $A_1, A_2, \ldots, A_n$ eseményeket függetlennek nevezzük, ha közülük tetszőlegesen kiválasztva k eseményt, igaz az, hogy az események szorzatának valószínűsége egyenlő az események valószínűségeinek szorzatával.

Példaként A, B és C három esemény esetén a fenti definíció alapján a függetlenség feltételei:

$$P(AB) = P(A)P(B);\; P(AC) = P(A)P(C),\; P(BC) = P(B)P(C); \text{ és}$$
$$P(ABC) = P(A)P(B)P(C).$$

Megjegyzések

1. A függetlenség és a közös rész nélküliség (diszjunkt halmaz) fogalma nem keverendő! Tehát, ha $AB=\emptyset$, ebből nem következik, hogy a két esemény független, hiszen legyen A és B két olyan esemény, amelyekre igaz, hogy $AB=\emptyset$, tehát $P(AB)=0$, de $P(A)\neq 0$ és $P(B)\neq 0$. Ekkor a két esemény szorzatának valószínűségére azt kapjuk, hogy

$$P(AB)=0\neq P(A)P(B).$$

Hiszen két nullától különböző pozitív szám szorzata nem lehet nulla, vagyis a két esemény nem független. A két esemény akkor lenne független is, ha vagy $P(A)=0$ vagy $P(B)=0$.

2. Az előző megjegyzésben mondottakból következik, hogy az egymást kizáró A és $\overline{A}$ események nem függetlenek egymástól.

3. A gyakorlatban a kísérletek során általában nem ismertek az események valószínűségei, tehát a (3.12.3.) definíció alapján nem tudjuk eldönteni a függetlenség kérdését. Ilyenkor a valószínűség kísérleti definícióját hívhatjuk segítségül, vagyis a relatív gyakoriságok vizsgálata dönthet a függetlenségről.

4. A gyakorlati életben sokszor függetlennek mondunk két eseményt, amikor úgy érezzük, hogy egyik esemény nincs hatással a másikra. Például két egymás utáni kockadobást függetlennek gondolunk. Ilyen esetekben a relatív gyakoriságok alapján ellenőrizve azt tapasztaljuk, hogy valóban teljesül a matematikai függetlenség is. Ugyanakkor óvatosnak kell lennünk, mert lehetnek rejtett összefüggések, amelyek első ránézésre nem szembeötlőek.

5. Ha játékkockával dobunk, majd utána érmét dobunk fel, akkor a kockadobás végeredményének A eseményét és az érmedobás B eredményét olyannak gondoljuk, amelyek nincsenek hatással egymásra. Mi a helyzet a két esemény matematikai függetlenségével? A válasz azért nem magától értetődő, mert ez a két esemény nem ugyanannak az eseménytérnek az eleme. A függetlenség eddigi definíciói pedig ugyanazon az eseménytéren értelmezett eseményekre vonatkozott. A probléma feloldására megtehetjük azonban azt, hogy a két eseményteret egyesítjük. Az egyesített eseménytérnek minden elemi eseménye két elemi eseményt tartalmaz. Egyet az egyik, és egyet a második eseménytérből. A jelen példa esetén tehát az egyesített tér egy elemi eseménye egy *1–6* közötti számból, és a fej vagy írás egyikéből áll. Ezen az egyesített téren már értelmezhető a két esemény függetlenségének kérdése.

Lássuk egy példán, hogy a függetlenséggel kapcsolatos eddigiekben tanult összefüggéseket hogyan alkalmazhatjuk!

3.9. Példa. Lássuk be, hogy ha A és B események függetlenek egymástól, akkor $\overline{A}$ és $\overline{B}$ események is függetlenek!

A de Morgan-azonosságok alkalmazásával kapjuk azt, hogy

$$P(\overline{A}\,\overline{B})=P(\overline{A+B})=1-P(A+B)=1-P(A)-P(B)+P(AB).$$

A és B függetlensége miatt $P(AB) = P(A)P(B)$, tehát

$$P(\overline{A}\overline{B}) = 1 - P(A) - P(B) + P(A)P(B) = 1 - P(A) - P(B)(1 - PA)).$$

Innen azt kapjuk, hogy

$$P(\overline{A}\overline{B}) = 1 - P(A) - P(B)(1 - PA)) = (1 - P(A))(1 - P(B)) = P(\overline{A})P(\overline{B}),$$

ami éppen a feladat kérdésére adott válasz.

Gyors ellenőrző feladatok

3.13. Ha A és B események függetlenek egymástól, akkor lássuk be, hogy A és $\overline{B}$ is függetlenek!

3.14. Ha A és B események függetlenek egymástól, akkor lássuk be, hogy $\overline{A}$ és B is függetlenek!

3.13. A Bernoulli-kísérletsorozat

Tegyük fel, hogy olyan kísérletet végzünk, amelynek csak két kimenetele van, A és $\overline{A}$. A két esemény valószínűsége $P(A)=p$, és $P(\overline{A}) = 1 - p = q$. Ilyen kísérlettel gyakran találkozunk a gyakorlatban. A legegyszerűbb változata az ilyen kísérletnek az érmedobás. De kockadobásnál is felvetődhet az a kérdés, hogy 1-et, vagy pedig nem 1-et dobunk a kockával. Ez utóbbi esetben a valószínűségek:

$$P(i = 1) = p = \frac{1}{6};\ P(i > 1) = q = \frac{5}{6}.$$

Az ilyen kísérletben az eredmény gyakran *siker* vagy *kudarc* formájában jelentkezik. Ha az ilyen kísérletet egymásután n-szer megismételjük úgy, hogy az egyes kísérletek függetlenek egymástól, akkor Bernoulli-kísérletsorozatról beszélünk.

A Bernoulli-kísérletsorozat kapcsán gyakran felmerül a következő kérdés: mi a valószínűsége annak, hogy az n független kísérlet során pontosan k-szor következik be az A esemény? Ezt a feladatot gyakran **Bernoulli-problémának** nevezzük.

A feladat megoldása során először tegyük fel, hogy az n kísérlet során éppen azt kapjuk, hogy az első k kísérletben az A esemény, az azt követő n-k kísérletben az $\overline{A}$ esemény valósul meg. A független események kapcsán kimondott szorzási szabály alapján az egyesített kísérlet valószínűségét könnyen megadhatjuk:

$$P(\underbrace{AA\ldots A}_{k}\underbrace{\overline{A}\overline{A}\ldots\overline{A}}_{n-k}) = \underbrace{P(A)P(A)\ldots P(A)}_{k}\underbrace{P(\overline{A})P(\overline{A})\ldots P(\overline{A})}_{n-k} = p^k q^{n-k}.$$

Ez azonban még nem a probléma megoldása, hiszen az, hogy az A esemény k-szor valósul meg még más sorrendben is létrejöhet. A lehetőségek számát az adja meg, hogy n pozícióból hányféleképpen lehet pontosan k különböző pozíciót kiválasztani. A választ erre a kombinatorika segítségével kaphatjuk meg: ez éppen n elem k-ad osztályú kombinációja, azaz $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. A Bernoulli-probléma megoldása tehát:

$$B_p(n,k) = \binom{n}{k} p^k q^{n-k}. \tag{3.13.1.}$$

3.10. Példa. Tegyük fel, hogy egy urnában N darab golyó van. Az N golyó közül M fehér és N–M fekete golyó van. Kérdés, mi a valószínűsége annak, hogy n húzás során pontosan k alkalommal fehér golyót húzunk? A feladat csak akkor oldható meg a Bernoulli-probléma megoldás képletével, ha biztosítjuk az egymás utáni húzások függetlenségét. Ehhez minden húzás után vissza kell tenni a kihúzott golyót, és alaposan össze kell keverni a golyókat. Ebben az esetben a p valószínűség minden húzás során azonos lesz, és így p és q a klasszikus valószínűség képlete alapján kiszámolható:

$$p = \frac{M}{N};\ q = \frac{N-M}{N}.$$

Ezután alkalmazva a Bernoulli-probléma (3.13.1.) képletét, megkapjuk a feladat megoldását:

$$B_n(n,k) = \binom{n}{k}\left(\frac{M}{N}\right)^k\left(\frac{N-M}{N}\right)^{n-k}.$$

3.11. Példa. Most érkeztünk el oda, hogy megoldhatjuk a bevezetőben már említett problémát, amelynek megoldását de Méré lovag kérte Pascaltól. A kérdés az volt, hogy melyik a valószínűbb, az hogy egy kockával dobva négy dobás közül legalább egyszer hatost dobunk, vagy az, hogy két kockával dobva 24 dobás közül legalább egyszer tizenkettőt dobunk?

A megoldás során célszerű mindkét esetben a komplementer esemény valószínűségét kiszámolni, majd azt 1-ből kivonni. Tehát, annak valószínűsége, hogy egy kockával négyszer dobva egyszer sem dobunk hatost a Bernoulli-képlet alapján $B_p(4,0)$:

$$B_p(4,0) = \binom{4}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^4 = \left(\frac{5}{6}\right)^4 = 0{,}4822\ldots, \text{ és } 1 - B_p(4,0) = 0{,}5177\ldots .$$

Annak valószínűsége, hogy két kockával 24-szer dobva egyszer sem dobunk hatost szintén a Bernoulli-képletet használva $B_p(24,0)$:

$$B_p(24{,}0)=\binom{24}{0}\left(\frac{1}{36}\right)^0\left(\frac{35}{36}\right)^{24}=\left(\frac{35}{36}\right)^{24}=0{,}5085\ldots,\text{ és } 1-B_p(24{,}0)=0{,}4914\ldots .$$

(Célszerű a számítást logaritmus segítségével végezni!). Látszik, hogy a különbség meglehetősen kicsi. de Méré lovag igazán sokat kockázhatott, ha ilyen kis különbséget észrevett.

4. VALÓSZÍNŰSÉGI VÁLTOZÓ, VÁRHATÓ ÉRTÉK, SZÓRÁS

4.1. Valószínűségi változó

Az eddigiekben az események valószínűségeiről beszéltünk. A gyakorlatban azonban általában a véletlen esemény egy kísérlet végeredménye. A kísérletekben pedig arra törekszünk, hogy a végeredmény numerikus formában jelenjen meg. A véletlen események megvalósulásakor annak elemi eseményei közül az egyik jön létre. Tehát a kísérlet során az elemi eseményekhez rendelt számérték az, ami a kísérlet végeredményeként jelentkezik. A kockadobás esetén például az elemi esemény a kocka lapjaira írott szám. Persze a kockadobással kapcsolatban más események is megfogalmazhatók. Például az, hogy páros vagy páratlan számot dobunk. Ilyenkor magunk dönthetjük el, hogy milyen számot rendelünk a páros és milyen számot a páratlan esemény bekövetkeztéhez; mondjuk 0-t, ha páros, 1-et ha páratlan a végeredmény.

Általánosan fogalmazva az elemi eseményekhez a kísérlet igényeinek megfelelően számokat rendelhetünk hozzá. Minthogy a véletlen kísérlet során véletlenszerűen egyik vagy másik szám lesz a végeredmény, ezért az elemi eseményekhez így hozzárendelt számot **valószínűségi változónak** nevezzük. A továbbiakban a valószínűségi változót a szokásoknak megfelelően ξ (kszi), η (éta) stb. görög betűkkel fogjuk jelölni. Matematikai értelemben az elemi eseményekhez történő hozzárendelés $\xi=\xi(\omega)$ függvényszerű hozzárendelésnek felel meg, ahol a hozzárendelés értelmezési tartománya Ω. Innen jól látszik, hogy ξ véletlen változó, hiszen az elemi események is véletlen események.

A kísérletek során sok esetben az elemi események maguk is számértékek, és ilyenkor sokszor ezt tekintjük valószínűségi változónak is. A számok hozzárendelése az elemi eseményekhez azonban önkényes. A kísérletező maga dönti el, hogy a kísérlet során milyen valószínűségi változó értékeket alkalmaz. A kockadobás esetén sem kötelező a kockára írt számot választani. Mondhatjuk azt is, hogy a dobás során a felül lévő számhoz mint elemi eseményhez nem ezt a számot rendeljük hozzá valószínűségi változó értékként, hanem például a szám négyzetét. Sőt különböző elemi eseményekhez ugyanazt a számot is hozzárendelhetjük. Célszerű mindig a kísérlet követelményeinek legjobban megfelelő számértékeket választani.

Vannak olyan feladatok, amelyekben az elemi eseményekhez nem egy, hanem több számértéket célszerű hozzárendelni. Időjárás esetén például a napi középhőmérséklet, és a napi csapadékmennyiség. Ilyenkor a valószínűségi változó nem egy, hanem több számértékből áll, vagyis a valószínűségi változó lehet vektor jellegű is.

Foglalkozzunk egyelőre azzal az esettel, amikor a valószínűségi változónak minden lehetséges értéke egyetlen számból áll!

A valószínűségi változó lehet diszkrét, mint például a kockadobás esetén, és lehet folytonos is, mint a napi középhőmérséklet során. Definiáljuk pontosabban a diszkrét valószínűségi változót!

Definíció. A valószínűségi változót **diszkrétnek** nevezzük, ha ξ-nek véges sok, vagy megszámlálhatóan végtelen sok lehetséges értéke van. A továbbiakban ξ lehetséges értékeit általában az

$$x_1, x_2, \ldots, x_n, \ldots$$

betűkkel jelöljük.

Definíció. Ha a ξ valószínűségi változó a valós számok egy intervallumán minden értéket felvehet, akkor a valószínűségi változót **folytonosnak** nevezzük. Természetesen ez az intervallum lehet a teljes számegyenes is.

Van olyan kísérlet, amelyben csak az egyik elemi esemény bekövetkezése, vagy be nem következése érdekel bennünket. Például a kockadobás esetén hatost dobunk, vagy nem dobunk hatost. Ilyenkor célszerű az indikátorváltozó alkalmazása.

Definíció. Amennyiben egy véletlen kísérletben csak két esemény A és $\overline{A}$ fordulhat elő (vagy csak két esemény érdekel bennünket), és a hozzájuk rendelt valószínűségi változó az alábbi:

$$\xi = \begin{cases} 1, & \text{ha az } A \text{ esemény következik be} \\ 0, & \text{ha az } \overline{A} \text{ esemény következik be,} \end{cases}$$

akkor ezt a ξ valószínűségi változót **indikátorváltozónak** nevezzük.

Nézzünk néhány példát a valószínűségi változóra.

4.1. Példa. A kockadobás során az Ω eseménytér:

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Az elemi események maguk is számok. De nem ezek a valószínűségi változó értékek. A valószínűségi változó értékek hozzárendelése során természetesen választhatjuk ezeket a számokat is, de mint mondtuk, választhatunk más számokat is. Példánkban most ez utóbbi lehetőséget választjuk. A páros számokhoz rendeljünk 0-t, a páratlanokhoz pedig 1-et. Tehát

$$\xi(1) = 1;\ \xi(2) = 0;\ \xi(3) = 1;\ \xi(4) = 0;\ \xi(5) = 1;\ \xi(6) = 0.$$

Ez esetben ξ tulajdonképpen indikátorváltozó.

Lássunk példát a folytonos valószínűségi változóra is!

4.2. Példa. A céltáblás kísérletben az elemi események a céltábla (x,y) számpárral jellemzett pontjai. A pontokhoz hozzárendelhetünk egy olyan valószínűségi változót, amely megadja a pontnak a középponttól mért távolságát. Tehát, ha a kör alakú céltábla sugara R, akkor

$$\xi = r = \sqrt{x^2 + y^2}\text{; ahol } \xi \in [0, R] \text{ intervallumnak.}$$

Ebben az esetben ξ folytonos valószínűségi változó.

A céltábla esetében megtehetjük azt is, hogy az R sugarú céltáblát felosztjuk $R/10$, $2R/10$, ..., $10R/10$ sugarú körökkel. A ξ valószínűségi változó lehetséges értékei pedig legyenek a következők: a belső körhöz a $x_1=10$, a belülről számítva első körgyűrűhöz a $x_2=9$, ..., a legkülső körgyűrűhöz a $x_{10}=1$ értéket rendeljük hozzá! Ez a változat az, amit a gyakorlatban is alkalmazni szoktak. Ez utóbbi esetben diszkrét valószínűségi változót definiáltunk.

4.2. A diszkrét valószínűségi változó eloszlása

A valószínűségi változó definíciójából következik, hogy az elemi eseményekhez rendelt számérték. Az elemi események valószínűségét a korábbiakban definiáltuk. Tehát a hozzárendelés alapján a valószínűségi változókhoz is rendelhető valószínűségi érték.

Definíció. A valószínűségi változó lehetséges értékeihez rendelt valószínűség azon elemi események valószínűségeinek összegével egyezik meg, amelyekhez az adott értéket rendeltük, azaz

$$P(\xi = x_k) = \sum_i P(\omega_i),\ \forall i - \text{re, amelyre } \xi(\omega_i) = x_k. \tag{4.2.1.}$$

Látjuk, hogy összegzés akkor jelenik meg, ha különböző elemi eseményekhez azonos valószínűségi változó értéket rendelünk hozzá.

A valószínűségi változó lehetséges értékeihez rendelt valószínűség értékek együttesét a **valószínűségi változó eloszlásának** (vagy röviden valószínűségi eloszlásnak) nevezzük. Az értelmezési tartomány itt a valós számegyenesnek az a tartománya, ahol a ξ lehetséges értékei elhelyezkednek. Sokszor az egyszerűség kedvéért értelmezési tartományként az egész valós számegyenest tekintjük, és azokon a helyeken, ahol ξ-nek nincsenek értékei, a hozzárendelt valószínűség értéke nulla (lehetetlen esemény!). Ezt úgy is értelmezhetjük, hogy az Ω eseménytér szerepét most a számegyenes veszi át, az értékkészlet pedig a $[0,1]$ intervallumba eső valós számok köre. Az eloszlás elnevezést az indokolja, hogy a hozzárendelés azt mutatja meg, hogy a valószínűségi változóhoz rendelt valószínűségi értékek hogyan oszlanak meg a számegyenes pontjai között.

4.3. Példa. A 4.1. példában definiált valószínűségi változókhoz rendelt valószínűség értékek, mivel valamennyi elemi esemény valószínűsége 1/6:

$$P(\xi=1)=P(1)+P(3)+P(5)=\frac{3}{6}, \text{ és } P(\xi=2)=P(2)+P(4)+P(6)=\frac{3}{6}.$$

Minden ilyen esetben a valós számegyenes pontjaihoz rendelünk egy-egy számot, amely számok a [*0, 1*] intervallum pontjai. A 4.3. példában a *0* és az *1* ponthoz egyaránt a *3/6=1/2* pontot rendeltük hozzá. Tekinthetjük ezt úgy is, hogy a számegyenes többi pontjához pedig a *0* értéket rendelhetjük. A példán az is látszik, hogy a hozzárendelt értékek összege *1*.

A (4.2.1.) hozzárendelési szabályból látszik, hogy általában is igaz az, hogy a valószínűség értékek összege *1*, hiszen

$$\sum_k P(\xi=x_k)=\sum_k\sum_i P(\omega_i)=\sum_{j=1}^{n} P(\omega_j)=P\left(\sum_{j=1}^{n}\omega_j\right)=P(\Omega)=1,$$

ahol $n=k+i$ az összes elemi események száma.

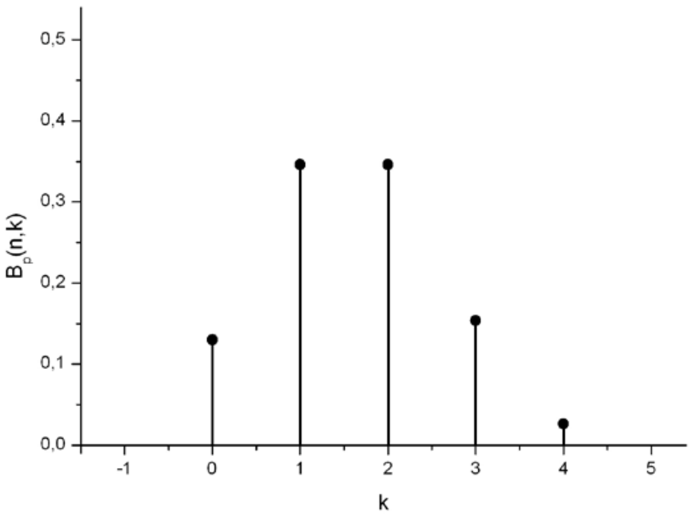
A valószínűségi változó bevezetésével lényegében az alábbi konstrukciót hoztuk létre:

$$\xi: x_1, x_2, ..., x_n, ...; \quad \text{ahol} \quad -\infty\le x_i\le\infty.$$

$$P(\xi):\ p(x_1), p(x_2), ..., p(x_n), ... \ ; \text{ ahol } 0\le p(x_i)\le 1.$$

Az x_i értékek függvényében ábrázolhatjuk a hozzájuk tartozó $p(x_i)$ értékeket. Diszkrét esetben ez az ábrázolás hasonló ahhoz, mint amit a diszkrét relatív gyakoriságok ábrázolása esetén tettünk. Eredményül pálcikaábrát kapunk, és ez az ábra tükrözi a valószínűség értékek eloszlását a ξ valószínűségi változó lehetséges x_i értékei között. Példa gyanánt tekintsük a Bernoulli-problémájaként megismert (3.13.1.) kifejezést.

4.4. Példa. Ha a (3.13.1.) kifejezést nem teljesen szabályos pénzérme esetére alkalmazzuk, ahol az írás valószínűsége $p=0,4$, akkor megkérdezhető, hogy mekkora a valószínűsége annak, hogy $n=4$ dobás során az írás $k=0$, $k=1$, ..., $k=4$ esetben le a dobás eredménye. Ebben a példában a k a valószínűségi változó, amelynek lehetséges értékei *0, 1, 2, 3, 4*. A (3.13.1) kifejezés alapján a hozzájuk tartozó valószínűségi értékek: $p(0)=0,13;$ $p(1)=0,346;$ $p(2)=0,346,$ $p(3)=0,154;$ $p(4)=0,026$. A 4.1. ábra mutatja a megfelelő pálcikaábrát.



4.1. ábra: A Bernoulli-eloszlás függvénye n=4 és p=0,4 értékekre

Tekinthetjük úgy ezt a függvényt, hogy a $(-\infty, +\infty)$ tartományon mindenhol nulla értékű, kivéve az ábrán megjelenített pontokban, ahol a függvényértékek az adott valószínűségi változóhoz rendelt valószínűség értékek.

Gyors ellenőrző feladatok

4.1. Szabályos kocka dobása során legyenek a valószínűségi változó értékek a kocka oldalaira írt számok. Rajzoljuk fel a feladathoz a valószínűség eloszlást ábrázoló pálcikaábrát!

4.2. Dobjunk egyszerre két szabályos érmével! Az eseménytér ekkor az alábbi:

$$\Omega = \{ff,\ fi,\ if,\ ii\}\ .$$

Az elemi eseményekhez rendelt ξ valószínűségi változó legyen az elemi esemény során a fejek száma, azaz

$$\xi(ff) = 2;\ \ \xi(fi) = 1;\ \ \xi(if) = 1;\ \ \xi(ii) = 0\,.$$

Adjuk meg a valószínűségi változó lehetséges értékeihez a valószínűség értékeket! Rajzoljuk fel a pálcikaábrát!

4.3. A folytonos valószínűségi változó esete

Folytonos valószínűségi változó esetén nem ennyire egyszerű a helyzet, ha grafikusan akarjuk jellemezni a valószínűség eloszlását. A korábbiakban láttuk már (a geometriai valószínűség kapcsán), hogy folytonos valószínűségi változó esetén egy pont valószínűsége

0. Ha tehát a pontok valószínűségeit ábrázolnánk, mint diszkrét esetben, akkor mindenütt *0* függvényt kapnánk, amivel nem sokra mennénk. Ehelyett nyilvánvalóan más utat kell választanunk. Választhatjuk például azt az ábrázolást, amit a kumulatív relatív gyakoriság esetében tettünk. Ábrázoljuk annak a valószínűségét, hogy a valószínűségi változó valamekkora x értéknél kisebb! Legyen az ezt megadó függvény neve valószínűségi eloszlás függvény, és jelöljük ezt a függvényt $F(x)$-el. Fogalmazzuk meg mindezt matematikai formában!

Definíció. Az

$$F(x) = P(\xi < x), \qquad -\infty < x < \infty \tag{4.3.1.}$$

függvényt a ξ valószínűségi változó **eloszlásfüggvényének** nevezzük. A definícióból következik, hogy $0 \le F(x) \le 1$ az egész értelmezési tartományon, hiszen $F(x)$ valószínűség. Folytonos esetben az $F(x)$ függvény is folytonos.

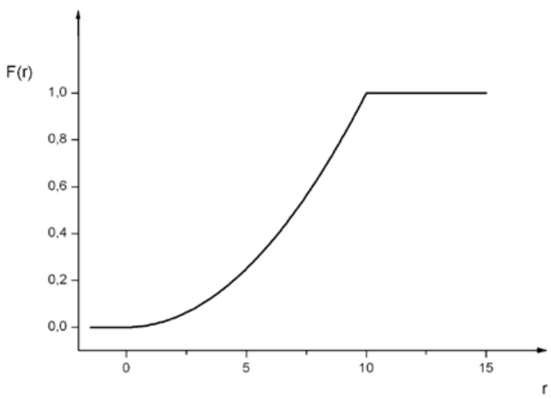
4.5. Példa. $R=10\ cm$ sugarú céltábla esetén adjuk meg az $F(r)$ eloszlásfüggvényt, amely megadja annak valószínűségét, hogy ha a céltáblát biztosan eltaláljuk, mi annak a valószínűsége, hogy a találat a r sugarú, a középponttal koncentrikus körön belül van? A 3.2. példában már kiszámítottuk ezt a valószínűséget, amely szerint

$$P(r < R) = \frac{r^2}{R^2}.$$

Az eloszlásfüggvény definíciója alapján tehát a teljes függvény:

$$F(r) = \begin{cases} 0, & ha\ r < 0 \\ \dfrac{r^2}{10^2}, & ha\ 0 < r < 10 \\ 1\ , & ha\ r > 10 \end{cases} . \tag{4.3.2.}$$

A (4.3.2) függvény grafikus alakja az 4.2. ábrán látható.



4.2. ábra: A (4.3.2) eloszlásfüggvény alakja

4.4. Az eloszlásfüggvény tulajdonságai

Az eloszlásfüggvény három tulajdonságát tételek formájában fogalmazzuk meg.

1. tétel. Az eloszlásfüggvény monoton növekvő (nem csökkenő) függvény, azaz

$$F(x_1) \le F(x_2), \text{ ha } x_1 < x_2. \tag{4.4.1.}$$

Bizonyítás. A tétel állítása a valószínűségelmélet 4. alaptételének következménye. Korábban bizonyítottuk, hogy ha $B \subset A$, akkor $P(B) \le P(A)$. Ha most a B eseményként a $\xi < x_1$ relációt, az A eseményként a $\xi < x_2$ relációt tekintjük, akkor $P(B) = F(\xi < x_1)$ és $P(A) = F(\xi < x_2)$ miatt a tétel állítása már következik.

A 4.2. ábrára pillantva láthatjuk, hogy a ($0,10$) tartományon a függvény szigorúan monoton emelkedik, míg az $x<0$ és az $x>10$ tartományokon az egyenlőség teljesül.

2. tétel. Az eloszlásfüggvényre igazak az alábbi határértékek:

$$\lim_{x\to-\infty} F(x) = 0, \qquad \text{és} \qquad \lim_{x\to\infty} F(x) = 1. \tag{4.4.2.}$$

Bizonyítás. Ez a tétel is a valószínűségelmélet alaptételeinek következménye, hiszen a

$$\lim_{x\to-\infty} F(x) = \lim_{x\to-\infty} P(\xi < x) = P(\xi < -\infty) = P(\emptyset) = 0,$$

hiszen a $\xi < -\infty$ esemény a lehetetlen esemény. Hasonló módon

$$\lim_{x\to\infty} F(x) = \lim_{x\to\infty} P(\xi < x) = P(\xi < \infty) = P(\Omega) = 1,$$

hiszen a $\xi < \infty$ esemény a biztos esemény.

3. tétel. Legyen a és b két valós szám. Annak a valószínűsége, hogy a ξ valószínűségi változó az (a,b) intervallumba esik, kiszámítható az *$F(x)$ eloszlásfüggvény* segítségével az alábbi módon:

$$P(a \le \xi < b) = F(b) - F(a). \tag{4.4.3.}$$

Bizonyítás. A $(\xi < b)$ esemény felbontható két közös rész nélküli esemény összegére:

$$(\xi < a) + (a \le \xi < b) = (\xi < b).$$

A 3. axióma szerint:

$$P(\xi < a) + P(a \le \xi < b) = P(\xi < b),$$

ahonnan átrendezéssel azt kapjuk, hogy

$$P(a \le \xi < b) = P(\xi < b) - P(\xi < a) = F(b) - F(a),$$

és ezzel az állítást bebizonyítottuk.

Gyors ellenőrző feladatok

4.1. A 4.5. példa adatait felhasználva számoljuk ki annak valószínűségét, hogy ha véletlenszerűen lövünk az $R=10\ cm$ sugarú céltáblára, mi a valószínűsége annak, hogy az $r=5\ cm$ és az $r=2\ cm$ sugarú körökkel határolt területre esik a lövés?

4.2. A számegyenesen az (a,b) intervallumban véletlenszerűen kijelölünk egy pontot. Tegyük fel, hogy annak valószínűsége, hogy az (a,b) szakasz valamelyik részintervallumára esik a kijelölt pont, arányos az adott részintervallum hosszával. Írjuk fel a probléma valószínűségi eloszlásfüggvényét! Legyen $a=2$, $b=3$! Rajzoljuk fel az eloszlásfüggvény alakját!

4.5. Az eloszlásfüggvény diszkrét valószínűségi változó esetén

Bár az eloszlásfüggvényt a folytonos valószínűségi változó kapcsán definiáltuk, azonban a definíció alapján diszkrét esetben is fel tudjuk rajzolni. Legyenek a ξ valószínűségi változó lehetséges értékei nagyság szerinti sorrendben $x_1, x_2, \ldots, x_n$, a hozzájuk rendelt valószínűség értékek pedig $p_1, p_2, \ldots, p_n$. Az $F(x)$ függvény (4.3.1) definíciója alapján határozzuk meg a függvény alakját!

Mivel $x<x_1$ esetben a valószínűségi változónak nincs értéke, ezért $P(\xi < x_1)=0$, tehát

$$F(x) = 0, \text{ ha } x \le x_1. \tag{4.5.1.}$$

Figyelem! A (4.5.1) kifejezés az $x=x_1$ pontban is érvényes (ezért szerepel a $\le$ jel), hiszen az eloszlásfüggvény definíciója szerint $F(x_1)=P(\xi < x_1)$.

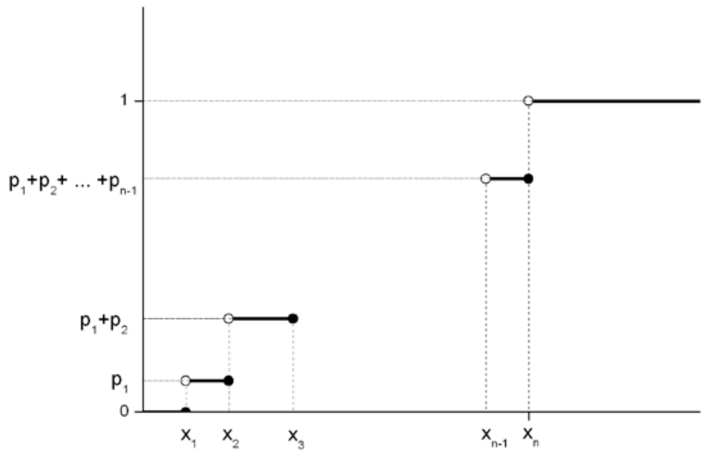
Ahogy átlépünk az x_1 ponton, minden $x_1 < x \le x_2$ pontban $P(\xi < x)=p_1$, tehát

$$F(x) = p_1, \text{ ha } x_1 < x \le x_2.$$

Az x_2 határra itt is érvényes az x_1 ponttal kapcsolatos megjegyzés. Amikor átlépjük az x_2 pontot, minden $x_2 < x \le x_3$ pontban $P(\xi < x)=p_1+p_2$, tehát

$$F(x) = p_1 + p_2, \text{ ha } x_2 < x \le x_3.$$

Most már látszik a konstrukció. A függvény minden x_i pontban **balról folytonos**, és ezekben a pontokban ugrása van a függvénynek. Az ugrás nagysága minden x_i pontban akkora, amekkora az adott ponthoz rendelt valószínűség értéke. A függvény alakját a 4.3. ábrán láthatjuk.



4.3. ábra: Diszkrét eloszlás eloszlásfüggvénye

Az $x>x_n$ pontokban a valószínűség értéke $P(\xi < x)=p_1+p_2+...+p_n=1$, tehát az eloszlásfüggvény értéke: $F(x)=1$.

Látjuk tehát, hogy diszkrét esetben az eloszlásfüggvény nem folytonos, hanem ξ minden lehetséges x_i értékénél szakadása (ugrása) van. Az ugrás értéke az x_i pontban éppen megegyezik a p_i valószínűséggel. Azt is láttuk, hogy az eloszlásfüggvény definíciójának az a következménye, hogy a függvény minden x_i pontban balról folytonos. Az ábrán a teli körök ezt jelzik.

Megjegyzés

Lehetne az $F(x)$ függvény definíciója a következő is:

$$F(x)=P(\xi \leq x).$$

Ha így definiálnánk az eloszlásfüggvényt, akkor diszkrét esetben, az ugráshelyeken $F(x)$ jobbról lenne folytonos.

Gyors ellenőrző feladatok

4.3. Rajzoljuk fel az eloszlásfüggvényt szabályos kockával történő dobás esetére!

4.4. Rajzoljuk fel az eloszlásfüggvényt a 4.4 példában meghatározott Bernoulli-eloszlás esetére!

4.6. A sűrűségfüggvény

Definíció. Legyen a folytonos ξ valószínűségi változó eloszlásfüggvénye $F(x)$. Az eloszlásfüggvény deriváltját a ξ valószínűségi változó **sűrűségfüggvényének** nevezzük, és $f(x)$-szel jelöljük, azaz

$$f(x) = F'(x). \tag{4.6.1.}$$

A sűrűségfüggvény tulajdonságait tételek formájában fogalmazzuk meg.

1. tétel. Az $f(x)$ függvényre igaz, hogy sehol nem negatív, azaz

$$f(x) \geq 0; \qquad -\infty < x < \infty. \tag{4.6.2.}$$

Bizonyítás. Az $F(x)$ függvényről tudjuk, hogy nem csökkenő függvény. A nem csökkenő függvények deriváltja nem lehet negatív, tehát ha

$$F(x_1) \leq F(x_2),\ \forall\ x_1 < x_2, \text{ akkor } F'(x) \geq 0.$$

2. tétel. Az $f(x)$ sűrűségfüggvény segítségével integrálással kiszámítható az $F(x)$ eloszlásfüggvény, azaz

$$F(x) = \int_{-\infty}^{x} f(x)dx. \tag{4.6.3.}$$

Bizonyítás. Az integrálás Newton–Leibnitz-szabályából következik, hogy

$$\int_{-\infty}^{x} f(x)dx = F(x) - F(-\infty).$$

Mivel $F(-\infty) = \lim\limits_{x\to-\infty} F(x) = 0$, ezért

$$\int_{-\infty}^{x} f(x)dx = F(x).$$

3. tétel. A sűrűségfüggvény egész számegyenesre vett integrálja 1, azaz

$$\int_{-\infty}^{\infty} f(x)dx = 1. \tag{4.6.4.}$$

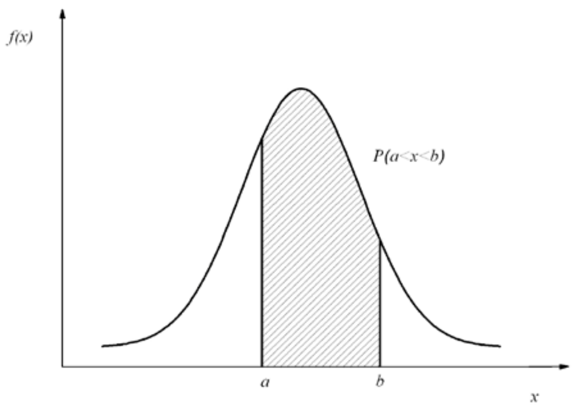
Ezt úgy is szokás mondani, hogy a sűrűségfüggvény *1*-re normált.

Bizonyítás.

$$\int_{-\infty}^{\infty} f(x)dx = F(\infty) - F(-\infty) = 1 - 0 = 1 .$$

4. tétel. A folytonos ξ valószínűségi változó *(a,b)* intervallumba esésének valószínűsége egyenlő a sűrűségfüggvényének *a*-tól *b*-ig terjedő görbe alatti területével, azaz:

$$P(a < \xi < b) = \int_a^b f(x)dx . \tag{4.6.5.}$$



4.4. ábra: A sűrűségfüggvény (a,b) tartományának görbe alatti területe

Bizonyítás. A Newton–Leibnitz-szabályból következik, hogy

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

A jobb oldalon szereplő különbségről pedig az eloszlásfüggvény tulajdonságai kapcsán beláttuk – lásd a (4.6.5) kifejezést –, hogy

$$F(b) - F(a) = P(a \le \xi < b) .$$

Folytonos esetben egy pont valószínűsége *0*, vagyis *P(a)=0*, tehát

$$F(b) - F(a) = P(a \le \xi < b) = P(a) + P(a < \xi < b) = P(a < \xi < b) .$$

Folytonos esetben ezzel beláttuk a (4.6.5) állítás igaz voltát.

Megjegyzés

A sűrűségfüggvény definíciója és a sűrűségfüggvény 2. tulajdonsága együtt azt is jelentik, hogy az eloszlás jellemzéséhez, elegendő vagy *F(x)*-et, vagy pedig *f(x)*-et ismernünk, hiszen egyik a másik ismeretében kiszámítható.

4.6. Példa. Határozzuk meg a 4.4 példában megismert céltáblás feladatban a valószínűségi változó sűrűségfüggvényét!

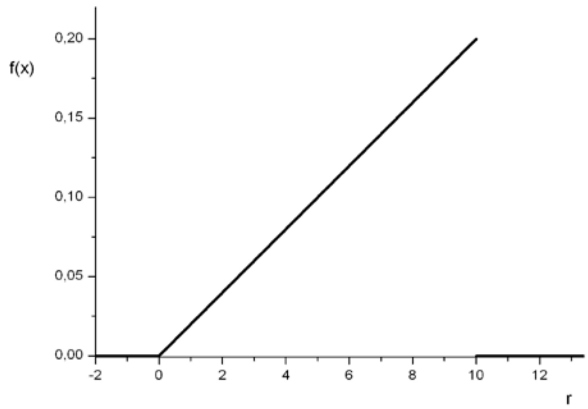
A 4.4. példában meghatározott eloszlásfüggvény:

$$F(r)=\begin{cases} 0 & \text{ha } r\le 0 \\ \dfrac{r^2}{10^2} & \text{ha } 0<r<10 \\ 1 & \text{ha } r>0 \end{cases}.$$

Ennek a függvénynek a deriváltja a tartományokon külön-külön létezik, éspedig:

$$f(x)=\begin{cases} 0 & \text{ha } r\le 0 \\ \dfrac{2r}{100} & \text{ha } 0<r<10 \\ 0 & \text{ha } r>0 \end{cases}.$$

A függvény grafikus alakja a 4.5. ábrán látszik.



4.5. ábra: A céltáblás példafeladat valószínűség sűrűségfüggvénye

Az 4.5. ábráról is látszik, hogy a sűrűségfüggvény normáltságáról mondottak szerint a függvény alatti terület egységnyi.

Gyors ellenőrző feladatok

4.5. Állapítsuk meg, hogy az alábbi függvény λ milyen értéke mellett lehet sűrűségfüggvény:

$$f(x)=\begin{cases} 0 & x<0 \\ e^{-\lambda x} & x\geq 0 \end{cases}!$$

4.6. Határozzuk meg a 4.5. ellenőrző feladat sűrűségfüggvényéből a valószínűségi eloszlásfüggvényt!

4.7. A diszkrét valószínűségi változó függvénye

A gyakorlatban előfordulnak olyan feladatok, amelyekben a ξ valószínűségi változó függvényével kell dolgoznunk. Felvetődik a kérdés, hogy a valószínűségi változó függvénye tekinthető-e valószínűségi változónak, és ha igen, akkor milyen valószínűség értékek rendelhetők hozzá, vagyis milyen az eloszlása?

Először a diszkrét esettel foglalkozunk. Tekintsük az $y=\varphi(x)$ valós függvényt! Ez a függvény a ξ valószínűségi változóhoz az $\eta=\varphi(\xi)$ értéket rendeli hozzá. Ez azt jelenti, hogy az ω_i elemi eseményhez, most az $y_k=\varphi(\xi(\omega_i))$ értéket rendeljük hozzá. Mivel a valószínűségi változó definiálása során azt mondtuk, hogy a hozzárendelés önkényes, ezért η-t is tekinthetjük valószínűségi változónak, amelynek lehetséges értékei: $y_1, y_2, \dots, y_k\dots$

A kérdés ezek után az, hogy ha ismerjük a ξ valószínűségi változó $P(\xi=x_i)$ eloszlását, akkor abból hogyan kapjuk meg az η valószínűségi változó $P(\eta=y_k)$ eloszlását?

Egyszerű a helyzet, ha a $\varphi(x)$ függvény szigorúan monoton nő, vagy csökken, hiszen akkor minden x_i értékhez egyetlen y_i érték tartozik, amelynek valószínűsége megegyezik a megfelelő x_i érték valószínűségével, hiszen ugyanahhoz az elemi eseményhez tartozik. Nézzünk egy példát!

4.7. Példa. Legyen a ξ valószínűségi változó $P(\xi=x_i)$ eloszlása a következő:

$$x_1=-2;\ x_2=-1;\ x_3=0;\ x_4=1;\ x_5=2$$
$$P(x_1)=\frac{1}{6};\ P(x_2)=\frac{1}{12};\ P(x_3)=\frac{1}{4};\ P(x_4)=\frac{1}{3};\ P(x_5)=\frac{1}{6}.$$

Legyen a transzformáló függvény: $y=2x+1$! Innen az η valószínűségi változó lehetséges értékei és a valószínűség eloszlása:

$$y_1=-3;\ y_2=-1;\ y_3=-1;\ y_4=3;\ y_5=5;$$
$$P(y_1)=\frac{1}{6};\ P(y_2)=\frac{1}{12};\ P(y_3)=\frac{1}{4};\ P(y_4)=\frac{1}{3};\ P(y_5)=\frac{1}{6}.$$

A helyzet kissé bonyolódik akkor, ha a $\varphi(x_i)$ függvény különböző x_i értékekhez azonos y_k értéket rendel. Ilyenkor az azonos y_k értékhez több x_i érték is tartozik, tehát ezek valószínűségeit össze kell adni ahhoz, hogy megkapjuk az y_k-hoz rendelt valószínűség értéket. Ilyenkor tehát a valószínűség eloszlás:

$$P(\eta = y_k) = \sum_{\varphi(x_i)=y_k} P(\xi = x_i). \qquad (4.7.1.)$$

Erre az esetre is nézzünk egy példát!

4.8. Példa. Legyen a ξ valószínűségi változó $P(\xi=x_i)$ eloszlása azonos a 4.7. példában definiálttal. A transzformáló függvény most legyen: $y=x^2$. Innen az η változó lehetséges értékei:

$$y(-2)=4;\ y(-1)=1;\ y(0)=0;\ y(1)=1;\ y(2)=4.$$

Látjuk tehát, hogy az η valószínűségi változónak valójában három lehetséges értéke van: $y_1=0;\ y_2=1;\ y_3=4$. A transzformált változó eloszlása:

$$P(\eta=0)=\frac{1}{4};\ P(\eta=1)=\frac{1}{12}+\frac{1}{3}=\frac{5}{12};\ P(\eta=4)=\frac{1}{6}+\frac{1}{6}=\frac{1}{3}.$$

Haladóknak

4.8. A folytonos valószínűségi változó függvénye

A ξ folytonos valószínűségi változó $\eta=\varphi(\xi)$ függvénye is valószínűségi változónak tekinthető. A kérdés most is az, hogy amennyiben ismerjük a ξ változó $F(x)$ eloszlásfüggvényét, abból hogyan határozzuk meg az η változó $G(y)$ valószínűségi eloszlásfüggvényét?

Tétel. Csak azzal az esettel foglalkozunk, amikor az $y=\varphi(x)$ függvény szigorúan monoton nő, vagy csökken. Ha a ξ valószínűségi változó eloszlásfüggvénye $F(x)$, akkor az $\eta=\varphi(\xi)$ valószínűségi változó eloszlásfüggvénye:

$$G(y)=\begin{cases} F\left(\varphi^{-1}(y)\right) & \text{ha } \varphi(x) \text{ szigorúan monoton } \textit{nő}, \\ 1-F\left(\varphi^{-1}(y)\right) & \text{ha } \varphi(x) \text{ szigorúan monoton csökken,} \end{cases} \qquad (4.8.1.)$$

ahol $x=\varphi^{-1}(y)$ az $y=\varphi(x)$ függvény inverze.

Bizonyítás. Ha $\varphi(x)$ szigorúan monoton nő, akkor $G(y)$ definíció szerint a következő:

$$G(y)=P(\eta<y)=P(\varphi(\xi)<y)=P\left(\xi<\varphi^{-1}(y)\right)=F(\varphi^{-1}(y)).$$

Amennyiben $\varphi(x)$ szigorúan monoton csökken, akkor

$$1-G(y)=P(\eta>y)=P(\varphi(\xi)>y)=P\left(\xi>\varphi^{-1}(y)\right)=F(\varphi^{-1}(y)),\text{ tehát}$$
$$G(y)=1-F\left(\varphi^{-1}(y)\right).$$

A most belátott képletek alapján kiszámíthatjuk az η valószínűségi változó sűrűségfüggvényét is.

Tétel. Az $\eta=\varphi(\xi)$ valószínűségi változó $g(y)$ sűrűségfüggvénye, ha a ξ valószínűségi változó sűrűségfüggvénye $f(x)$:

$$g(y)=f\left(\varphi^{-1}(y)\right)\left|\frac{d\varphi^{-1}(y)}{dy}\right|. \tag{4.8.2.}$$

Bizonyítás. A sűrűségfüggvény definíciója, valamint az összetett függvény deriválási szabálya alapján, ha $\varphi(x)$ függvény – tehát $\varphi^{-1}(y)$ is – szigorúan monoton nő:

$$g(y)=\frac{dG(y)}{dy}=\frac{dF\left(\varphi^{-1}(y)\right)}{d\varphi^{-1}(y)}\frac{d\varphi^{-1}(y)}{dy}=f\left(\varphi^{-1}(y)\right)\frac{d\varphi^{-1}(y)}{dy}.$$

Ha $\varphi(x)$ függvény – tehát $\varphi^{-1}(y)$ is – szigorúan monoton csökken, akkor

$$g(y)=-f\left(\varphi^{-1}(y)\right)\frac{d\varphi^{-1}(y)}{dy}.$$

Minthogy ilyenkor

$$\frac{d\varphi^{-1}(y)}{dy}<0,$$

tehát, az abszolút értékkel felírt összefüggés mindkét esetben helyes.

4.9. Példa. Legyen a transzformáló függvény $y=ax+b$ alakú. Határozzuk meg az η valószínűségi változó $g(y)$ sűrűségfüggvényét, ha a ξ valószínűségi változó sűrűségfüggvénye:

$$f(x)=\begin{cases}0 & x<0,\\ e^{-x} & x\geq 0.\end{cases}$$

Megoldás. A $\varphi(x)$ függvény inverze, és annak deriváltja:

$$x = \frac{y-b}{a}, \qquad \frac{dx}{dy} = \frac{1}{a}.$$

Innen a (4.8.2) tétel alapján:

$$g(y) = f\left(\frac{y-b}{a}\right)\frac{1}{|a|} = e^{-\frac{y-b}{a}}\frac{1}{|a|}.$$

4.9. Várható érték diszkrét esetben

A kísérleti adatok eloszlásának jellemzése során már láttuk, hogy számtani középnek milyen szerepe van az eloszlás jellemzésében. Ha az $x_1, x_2, \ldots, x_n$ méréssorozatot egyetlen számmal akarjuk jellemezni, akkor legtöbbször a mérések számtani átlagát adjuk meg. Korábban már láttuk, hogy a számtani középet a következő képlettel számolhatjuk:

$$\overline{x} = \sum_{i=1}^{N} \frac{k_i x_i}{n}, \tag{4.9.1.}$$

ahol n a mérési eredmények száma, N a lehetséges végeredmények száma, x_i az összes lehetséges mérési eredmény, k_i pedig a lehetséges kimenetelek gyakorisága az adott méréssorozatban. A (4.9.1) kifejezésben a k_i/n hányados a relatív gyakoriság. A relatív gyakoriságról tudjuk, hogy elegendően nagy mérésszám esetén, a nagy számok törvénye alapján stabilitást mutat. Azt is mondtuk, hogy elegendően nagy mérésszám esetén a relatív gyakoriságot tekintjük a valószínűség kísérleti értékének. A valószínűség, mint elméleti konstrukció pedig nem más, mint az a relatív gyakoriság érték, amelyet végtelen nagy (a gyakorlatban természetesen nem megvalósítható) mérésszám esetén érnénk el a mérések során. (Erre a későbbiekben még visszatérünk). Tehát

$$\lim_{x\to\infty} \frac{k_i}{n} = p_i.$$

A várható érték a számtani középpel rokon fogalom. A fenti gondolatmenet után ez a rokonság látszik a várható érték alábbi definíciójából.

Definíció. Legyenek a ξ valószínűségi változó lehetséges értékei $x_1, x_2, \ldots, x_n$. A lehetséges értékekhez rendelt valószínűség értékek legyenek rendre $p_1, p_2, \ldots, p_n$. Ekkor a ξ valószínűségi változó $M(\xi)$ **várható értékét** a következő képlettel definiáljuk:

$$M(\xi) = \sum_{i=1}^{n} x_i p_i. \tag{4.9.2.}$$

Ha véges sok lehetséges értékünk van, akkor a (4.9.2) összeg mindig létezik. Ha megszámlálhatóan végtelen sok lehetséges x_i értékünk van, akkor $M(\xi)$ csak akkor létezik, ha a (4.9.2) összeg konvergens, és a sorösszeg független a tagok sorrendjétől. Ennek feltétele az, hogy a végtelen sor abszolút konvergens legyen, vagyis, hogy teljesüljön a

$$\sum_i |x_i| p_i < \infty$$

feltétel.

4.10. Példa. Határozzuk meg a kockadobásos kísérletben a valószínűségi változó várható értékét! A ξ valószínűségi változó lehetséges értékeit és a hozzájuk tartozó valószínűség értékeket a 4.1. táblázat tartalmazza.

i	**1**	**2**	**3**	**4**	**5**	**6**
x_i	*1*	*2*	*3*	*4*	*5*	*6*
p_i	*1/6*	*1/6*	*1/6*	*1/6*	*1/6*	*1/6*

4.1. táblázat: A kockadobás lehetséges értékei és a hozzájuk tartozó valószínűségek

A várható érték (4.9.2) alapján:

$$M(\xi) = \sum_{i=1}^{n} p_i x_i = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3{,}5 .$$

A példa egyúttal azt is mutatja, hogy a diszkrét esetben maga a várható érték nem feltétlenül szerepel a lehetséges értékek között.

Megjegyzés

A (4.9.2) definíciós kifejezés formailag azonos egyenes mentén elhelyezkedő tömegpontok súlypontját definiáló képlettel, ha x_i a tömegpontok origótól mért távolságát, a p_i pedig az *i.* tömegpont tömegének és az összes tömegpont együttes tömegének a hányadosát jelzi. Az analógiából következik, hogy a pontra szimmetrikus valószínűség-eloszlás várható értéke a szimmetriatengelyen van, mint az egyszerű 4.10. példában is. Látjuk tehát, hogy a várható érték a valószínűség-eloszlás „centruma”.

Gyors ellenőrző feladatok

4.7. Adjuk meg a karakterisztikus valószínűségi változó várható értékét!

4.10. A várható érték folytonos esetben

Definíció. Ha a ξ valószínűségi változó folytonos sűrűségfüggvénye *f(x)*, akkor várható értéke a következő kifejezéssel definiáljuk:

$$M(\xi) = \int_{-\infty}^{+\infty} xf(x)dx \,. \tag{4.10.1.}$$

A folytonos esetben is az egyértelműség érdekében meg kell követelni az abszolút integrálhatóságot, vagyis, hogy

$$\int_{-\infty}^{+\infty} |x| f(x)dx < \infty .$$

Könnyen belátható, hogy a folytonos esetre érvényes definíció nem más, mint a diszkrét eset definíciójának általánosítása. A valószínűségi sűrűségfüggvény kapcsán már láttuk, hogy annak a valószínűsége, hogy a ξ valószínűségi változó az *(a,b)* intervallumba esik, megegyezik az *f(x)* függvény *a-b* határokkal vett görbe alatti területével. Ha tehát a számegyenest felosztjuk kis Δx_i szakaszokra, akkor annak valószínűsége, hogy a ξ változó az *i.* intervallumon belül van, a következő:

$$P(\xi \in \Delta x_i) \approx f(x_i^*)\Delta x_i ,$$

ahol x_i^* a Δx_i intervallum egy tetszőleges belső pontja. Ha (4.9.2) definíció alapján az x_i^* változóra kiszámítjuk a várható értéket, akkor az alábbi kifejezésre jutunk:

$$M(\xi) = \sum_{i=1}^{n} x_i^* P(\xi \in \Delta x_i) \approx \sum_i x_i^* f(x_i^*)\Delta x_i .$$

Most ugyanúgy járhatunk el, ahogyan azt az integrál bevezetésénél tettük, tehát minden határon túl finomítjuk a számegyenes felosztását, azaz

$$M(\xi) = \lim_{\Delta x_i \to 0} \sum_i x_i^* f(x_i^*)\Delta x_i = \int_{-\infty}^{\infty} xf(x)dx .$$

Amennyiben a határérték létezik, akkor az *f(x)* integrálható, tehát megadtuk folytonos esetre a (4.9.2) kifejezés általánosítását.

Fontos megjegyzés

Az $M(\xi)$ várható értéket a ξ valószínűségi változó **első elméleti momentumának** is szokás nevezni. Láttuk, hogy a várható érték szoros rokonságban van a számtani középpel. Ha a ξ valószínűségi változóval jellemzett sokaságon méréseket végzünk, és a mérések eredményeinek számtani közepét képezzük, akkor, mint már említettük, az empirikus várható értéket kapjuk meg. Az empirikus várható értéket a statisztikában szokás az **első empirikus momentumnak** nevezni. A későbbiekben látni fogjuk majd, hogy a valószínűség-elmélet és a statisztika az elméleti és az empirikus momentumok kapcsolatát pontosabban is megadja.

4.11. Példa. A 4.6. feladatban kiszámítottuk a céltáblás kísérlet valószínűségi változójának sűrűségfüggvényét. Határozzuk meg a valószínűségi változó várható értékét!

A (6.3) definíció szerint a várható érték folytonos esetben:

$$M(\xi) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^{10} r\frac{2r}{100}dr = \frac{1}{100}\frac{2r^3}{3}\bigg|_0^{10} = 6{,}66...$$

4.11. A várható érték tulajdonságai

Mivel a várható értéket a számtani közép fogalmából származtattuk, ezért tulajdonságai is hasonlóak a számtani közép már megismert tulajdonságaival. A tételeket a diszkrét esetre látjuk be, de ahol szükséges, a tételt felírjuk folytonos esetre is. Mivel az integrált összegzés határértékeként vezettük be, ezért nem csoda az, hogy hasonló tulajdonságai vannak, mint az összegzésnek.

1. tétel. Konstans várható értéke maga a konstans, azaz

$$M(c) = c. \tag{4.11.1.}$$

Bizonyítás. Ha a ξ valószínűségi változónak egyetlen lehetséges értéke van, azaz $\xi=c$, akkor ennek az értéknek a valószínűsége: $P(\xi=c)=1$. Tehát a várható érték (4.9.2) alapján:

$$M(c) = 1 \cdot c = c.$$

2. tétel. A ξ valószínűségi változó állandószorosának várható értéke az $M(\xi)$ várható érték állandószorosa, azaz

$$M(c\xi) = cM(\xi). \tag{4.11.2.}$$

A tétel folytonos esetre is igaz.

Bizonyítás. A ξ valószínűségi változó állandószorosa azt jelenti, hogy lehetséges értékei szorzódnak az állandóval, azaz $c\xi$ lehetséges értékei: $cx_1, cx_2, ..., cx_n$. Tehát a várható érték:

$$M(c\xi) = \sum_{i=1}^{n} cx_i p_i = c\sum_{i=1}^{n} x_i p_i = cM(\xi).$$

Korábban már láttuk, hogy a ξ valószínűségi változó $\eta=\varphi(\xi)$ függvénye is valószínűségi változó. Az alábbi tétel az η változó várható értékéről szól.

3. tétel. A ξ valószínűségi változó lehetséges értékei legyenek $x_1, x_2, ..., x_n$, és vegye fel a változó rendre $p_1, p_2, ..., p_n$ valószínűséggel ezeket az értékeket. A ξ valószínűségi változónak létezik az $M(\xi)$ várható értéke. Továbbá legyen $y=\varphi(x)$ tetszőleges valós függvény. A tétel állítása az, hogy az $\eta=\varphi(\xi)$ transzformált változó várható értéke:

$$M(\eta)=M(\varphi(\xi))=\sum_{i=1}^{n}\varphi(x_i)p_i\,. \tag{4.11.3.}$$

Ha a ξ valószínűségi változó folytonos, akkor a $\eta=\varphi(\xi)$ új valószínűségi változó várható értéke:

$$M(\eta)=M(\varphi(\xi))=\int_{-\infty}^{\infty}\varphi(x)f(x)dx\,. \tag{4.11.4.}$$

Bizonyítás. A tételt könnyű belátni, ha a $\varphi(x)$ függvény szigorúan monoton függvény. Ebben az esetben ugyanis a $\varphi(x)$ transzformáció minden x_i lehetséges értékhez egyetlen $y_i=\varphi(x_i)$ értéket rendel hozzá. Mivel ez az új változó ugyanahhoz az ω_i elemi eseményhez tartozik, ezért az ehhez tartozó valószínűség érték is p_i. Tehát η várható értéke:

$$M(\eta)=\sum_i y_i P(\eta=y_i)=\sum_i \varphi(x_i)p_i\,.$$

Ha a transzformáló függvény nem szigorúan monoton függvény, akkor az η valószínűségi változó $y_i=\varphi(x_i)$ lehetséges értékei között vannak azonosak. Ilyenkor y_k valószínűségét úgy kapjuk meg, hogy összegezzük az összes ilyen x_i -hez tartozó p_i értéket, azaz

$$p_k=P(\eta=y_k)=\sum_{\varphi(x_i)=y_k}p_i\,. \tag{4.11.5.}$$

Tehát η várható értéke

$$M(\eta)=\sum_i y_i P(\eta=\varphi(x_i))=\sum_k y_k p_k\,. \tag{4.11.6.}$$

A (4.11.6) kifejezésben már csak a különböző y_k értékek szerepelnek. Ezzel ismét a (4.11.3) összefüggésre jutunk.

4.12. Példa. A 4.7. példához hasonlóan legyen a ξ valószínűségi változó $P(\xi=x_i)$ eloszlása a következő:

$$x_1=-2;\ x_2=-1;\ x_3=0;\ x_4=1;\ x_5=2$$
$$p_1=\frac{1}{6};\ p_2=\frac{1}{12};\ p_3=\frac{1}{4};\ p_4=\frac{1}{3};\ p_5=\frac{1}{6}.$$

A transzformáló függvény legyen a következő: $y=x^2$. Határozzuk meg az $\eta=\xi^2$ új valószínűségi változó várható értékét!

A (6.7) definíció szerint a várható érték:

$$M(\eta)=\sum_{i=1}^{n}\varphi(x_i)p_i=4\cdot\frac{1}{6}+1\cdot\frac{1}{12}+0\cdot\frac{1}{4}+1\cdot\frac{1}{3}+4\cdot\frac{1}{6}=$$
$$=4\left(\frac{1}{6}+\frac{1}{6}\right)+1\left(\frac{1}{12}+\frac{1}{3}\right)+0\left(\frac{1}{4}\right)=\frac{21}{12}=1{,}75.$$

4. tétel. Ha az új valószínűségi változót előállító transzformáló függvény több függvény összege, azaz $\varphi(x)=\varphi_1(x)+\varphi_2(x)$, akkor az $\eta=\varphi(\xi)=\varphi_1(\xi)+\varphi_2(\xi)=\eta_1+\eta_2$ új változó várható értéke:

$$M(\eta)=\sum_i\varphi_1(x_i)p_i+\sum_i\varphi_2(x_i)p_i=M(\eta_1)+M(\eta_2). \qquad (4.11.7.)$$

Bizonyítás. A tétel állítása a összegzés lineáris tulajdonságaiból következik. Hiszen

$$M(\eta)=\sum_i\varphi(x_i)p_i=\sum_i(\varphi_1(x)+\varphi_2(x))p_i=\sum_i\varphi_1(x_i)p_i+\sum_i\varphi_2(x_i)p_i=M(\eta_1)+M(\eta_2).$$

Az integrálás az összegzéshez hasonlóan lineáris operáció, tehát a bizonyítás teljesen hasonló módon történhet folytonos esetre is.

1. megjegyzés
Az 1., 2. és 4. tétel együttes alkalmazásából következik, hogy ha ismerjük a ξ valószínűségi változó $M(\xi)$ várható értékét, akkor az $\eta= ax+b$ új változó várható értéke:

$$M(\eta)=M(a\xi+b)=aM(\xi)+b.$$

A megjegyzés tulajdonképpen a ξ valószínűségi változó lineáris transzformáltjának várható értékéről szól.

2. megjegyzés
Teljes indukcióval belátható, hogy a 4. tétel nemcsak kettő, hanem több tag esetén is igaz.

4.13. Példa. Legyen a transzformáló függvény $\varphi(x)=a\xi^2+b\xi+c$ alakú! Határozzuk meg az $\eta=\varphi(\xi)$ új változó várható értékét!

Alkalmazva a (4.11.7) kifejezést és a várható érték korábban megismert tulajdonságait a megoldás a következő:

$$M(\eta)=M\left(a\xi^2+b\xi+c\right)=M(a\xi^2)+M(b\xi)+M(c)=aM(\xi^2)+bM(\xi)+c\,.$$

4.12. Szórás

Az 1.5. Mérési adatok egyszerű jellemzése című fejezetben bevezettük az empirikus szórásnégyzet és szórás fogalmát. A szórással a kísérleti adatoknak az empirikus várható értéktől való átlagos eltérését tudtuk jellemezni. Ezzel megegyező módon a valószínűségelméletben a szórás a valószínűségi változó lehetséges értékeinek a várható értéktől való eltérését jellemzi.

Definíció. A ξ valószínűségi változó **szórásnégyzetét** az alábbi kifejezés definiálja:

$$D^2(\xi)=M\left((\xi-M(\xi))^2\right). \tag{4.12.1.}$$

A szórásnégyzet pozitív négyzetgyöke pedig a szórást adja, azaz

$$D(\xi)={}_{+}\sqrt{M\left((\xi-M(\xi))^2\right)}\,. \tag{4.12.2.}$$

A definícióból látszik, hogy a szórásnégyzet a ξ valószínűségi változó várható értékétől való eltérése négyzetének a várható értéke. A szórásnégyzetet tehát a várható érték segítségével definiáltuk. A várható érték számos tulajdonságával az előző fejezetben megismerkedtünk, tehát nem lesz nehéz megismerni a szórásnégyzet és a szórás tulajdonságait.

4.13. Szórásnégyzet és szórás diszkrét esetben

A szórásnégyzet mint (4.12.1) definíciójából kitűnik, lényegében a ξ valószínűségi változó függvénye, így maga is valószínűségi változó. Ezért tehát a (4.9.2) képletből következik, hogy diszkrét esetben a szórásnégyzet konkrét alakja:

$$D^2(\xi)=\sum_{i=1}^{n}(x_i-M(\xi))^2\,p_i\,. \tag{4.13.1.}$$

A szórás a (4.13.1) kifejezés pozitív négyzetgyöke, azaz

$$D^2(\xi)={}_{+}\sqrt{\sum_{i=1}^{n}(x_i-M(\xi))^2\,p_i}\,. \tag{4.13.2.}$$

4.14. Szórásnégyzet és szórás folytonos esetben

A diszkrét esethez hasonló módon folytonos esetben is alkalmazható a valószínűségi változó függvényének várható értékét folytonos esetre meghatározó (4.10.5.) képlet. Ha a ξ valószínűségi változó sűrűségfüggvénye $f(x)$, akkor a szórásnégyzet:

$$D^2(\xi)=\int_{-\infty}^{\infty}(x-M(\xi))^2 f(x)dx\,. \tag{4.14.1.}$$

A szórás a (4.14.1.) kifejezés pozitív négyzetgyöke, azaz

$$D^2(\xi)={}_+\sqrt{\int_{-\infty}^{\infty}(x-M(\xi))^2 f(x)dx}\,. \tag{4.14.2.}$$

4.14. Példa. Határozzuk meg kockadobás esetén a valószínűségi változó szórását! A 4.10. példában már kiszámoltuk a várható értéket: $M(\xi)=3,5$. Alkalmazva a (4.13.1.) képletet:

$$D^2(\xi)=(1-3,5)^2\frac{1}{6}+(2-3,5)^2\frac{1}{6}+(3-3,5)^2\frac{1}{6}+(4-3,5)^2\frac{1}{6}+$$
$$+(5-3,5)^2\frac{1}{6}+(6-3,5)^2\frac{1}{6}=2,916...\ .$$

A szórás pedig: $D(\xi)=1,707...$

4.15. A szórás tulajdonságai

Az alábbiakban két tétellel ismerkedünk meg. Mindkettő a szórásnégyzet olyan tulajdonságaira mutat rá, amelyeket a gyakorlati számítások során gyakran alkalmazunk.

Tétel.

$$D^2(\xi)=M(\xi^2)-(M(\xi))^2=M(\xi^2)-M^2(\xi), \tag{4.15.1.}$$

vagyis a ξ valószínűségi változó szórásnégyzete kiszámolható a ξ^2 várható értéke és a ξ várható értéke négyzetének különbségeként.

Bizonyítás. A bizonyítás során a (4.13.1) definíciós képletből indulunk ki. A zárójelen belül elvégezzük a négyzetre emelést, és kihasználjuk a (4.11.7) tulajdonságot, miszerint a várható érték tagonként számolható:

$$D^2(\xi)=M\left((\xi-M(\xi))^2\right)=M\left(\xi^2-2\xi M(\xi)+(M(\xi))^2\right)=$$
$$=M(\xi^2)-2M(\xi)M(\xi)+M^2(\xi)=M(\xi^2)-M^2(\xi),$$

és ezzel beláttuk a (4.15.1.) kifejezés igaz voltát.

Az $M(\xi^2)$-et szokás a ξ valószínűségi változó elméleti második momentumának nevezni.

A gyakorlati számolások során sokszor kell a valószínűségi változó lineáris függvényének szórásnégyzetét számolni. Erről szól a következő képlet.

Tétel. Ha a ξ valószínűségi változó lineáris függvénye $\eta=a\xi+b$ alakú, akkor η szórásnégyzete a következő módon számolható:

$$D^2(\eta)=a^2D^2(\xi). \qquad (4.15.2.)$$

Bizonyítás. A bizonyítás során a (4.12.1) összefüggést és a várható érték számolásának szabályait alkalmazzuk.

$$D^2(\eta)=D^2(a\xi+b)=M\left[(a\xi+b-M(a\xi+b))^2\right]=M\left[(a\xi+b-aM(\xi)+b)^2\right]=$$
$$=M\left[(a\xi-aM(\xi))^2\right]=M\left[a^2(\xi-M(\xi))^2\right]=a^2M\left[(\xi-M(\xi))^2\right]=a^2D^2(\xi).$$

Gyors ellenőrző feladatok

4.8. Adjuk meg a karakterisztikus változó szórását!

5. TÖBB VALÓSZÍNŰSÉGI VÁLTOZÓ EGYÜTTES ELOSZLÁSA

A kísérleti adatok leírása részben már láttuk, hogy gyakran ugyanazon vizsgált jelenséggel kapcsolatban nem egyetlen jellemzőt vizsgálunk, hanem kettőt vagy többet. Sokszor a kérdés éppen az, hogy ezen jellemzők között van-e összefüggés. A valószínűségelméletben ez a probléma a valószínűségi változók együttes eloszlásának kérdéseként merül fel.

5.1. Diszkrét valószínűségi változók együttes eloszlása

Egy kísérlettel kapcsolatban fogalmazzunk meg két valószínűségi változót, és jelöljük ezeket ξ és η betűkkel. A ξ valószínűségi változó lehetséges értékei legyenek $x_1, x_2, \ldots, x_n, \ldots$, az η lehetséges értékei pedig $y_1, y_2, \ldots, y_n, \ldots$ Legyen r_{ij} annak valószínűsége, hogy az x_i és y_j értékek egyszerre valósulnak meg, azaz

$$r_{ij} = r(x_1, y_1) = P(\xi = x_i, \eta = y_j)\,. \qquad (5.1.1.)$$

A két valószínűségi változó lehetséges értékeinek együttes megvalósulása az x–y síkon egy-egy pontként ábrázolható. Ha az (x_i,y_j) pontok felett a z tengely irányában r_{ij} hosszúságú pálcikát rajzolunk, akkor kirajzolódik előttünk e két diszkrét valószínűségi változó **együttes eloszlása**.

Az $A_{ij}=(x_i,y_j)$ események egymást kizáró események, és miközben i és j végigfut az összes lehetséges értéken, megvalósul az összes esemény. A teljes valószínűség tétele értelmében tehát

$$\sum_{i,j} r_{ij} = \sum_i \sum_j r_{ij} = 1\,. \qquad (5.1.2.)$$

A grafikus ábrázolás helyett sokszor táblázatban jelenítjük meg a két változó lehetséges értékeit és a hozzájuk tartozó eloszlást.

ξ/η	y_1	y_2	...	y_m	Σ sor
x_1	r_{11}	r_{12}	.	r_{1m}	$p(x_1)$
x_2	r_{21}	r_{22}	.	r_{2m}	$p(x_2)$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
x_n	r_{n1}	r_{n2}	.	r_{nm}	$p(x_n)$
Σ oszlop	$q(y_1)$	$q(y_2)$		$q(y_m)$	1

5.1. táblázat: A ξ és η változók együttes eloszlása

5.2. Peremeloszlások diszkrét esetben

Az 5.1. táblázatban képezzük a sorok összegeit:

$$p_i = \sum_{j=1}^{m} r_{ij} . \tag{5.2.1.}$$

Ez a sorösszeg az 5.1. táblázatban az *i.* sor végén szerepel. Hasonló módon a *j.* oszlop alján az adott oszlopban szereplő elemek összege:

$$q_j = \sum_{i=1}^{n} r_{ij} . \tag{5.2.2.}$$

Ezeknek az összegeknek önálló jelentésük is van. Az (5.2.1.) p_i sorösszeg például megadja annak valószínűségét, hogy a ξ valószínűségi változó milyen valószínűséggel veszi fel az x_i értéket, függetlenül attól, hogy az η változónak milyen értéke valósul meg. Matematikai formába öntve tehát:

$$p_i = \sum_{j=1}^{m} r_{ij} = \sum_{j=1}^{m} P(\xi = x_i, \eta = y_j) = P(\xi = x_i).$$

Ezzel analóg módon a q_j oszlopösszeg megadja, hogy az η változó milyen valószínűséggel veszi fel az y_j értéket, függetlenül a ξ változó értékeitől, azaz

$$q_j = \sum_{i=1}^{n} r_{ij} = \sum_{i=1}^{n} P(\xi = x_i, \eta = y_j) = P(\eta = y_j).$$

A p_i és q_j eloszlásokat **peremeloszlásnak** nevezzük.

Természetesen igaz az, hogy ha az összes sorösszeget (peremeloszlás a sorok végén) vagy oszlopösszeget (peremeloszlás az oszlopok végén) összeadjuk, akkor az összeg 1, vagyis

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{m} P(\xi = x_i, \eta = y_j) = 1,$$

illetve

$$\sum_{i=1}^{n} q_i = \sum_{j=1}^{n}\sum_{i=1}^{m} r_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{m} P(\xi = x_i, \eta = y_j) = 1.$$

5.1. Példa. Háromszor dobunk egymás után egy érmével, és ezt tekintjük egy eseménynek. Az elemi események *3* darab *f* és *i*-ből állnak. Az alábbi táblázat első sora tartalmazza az összes elemi eseményt. A ξ valószínűségi változó lehetséges értékei az elemi eseményekben szereplő fejek száma, ezeket a táblázat második sora mutatja. Az η valószínűségi változó értéke egy, ha az írások száma kettő, minden más esetben *0*.

Ω =	{	*fff,*	*ffi ,*	*fif,*	*iff,*	*fii,*	*ifi,*	*iif,*	*iii*	}
ξ =	{	*3,*	*2,*	*2,*	*2,*	*1,*	*1,*	*1,*	*0*	}
η =	{	*0,*	*0,*	*0,*	*0,*	*1,*	*1,*	*1,*	*0*	}

5.2. táblázat: Az eseménytér és a valószínűségi változók értékei az 5.1. példához

Készítsünk táblázatot a ξ és η valószínűségi változók együttes eloszlásáról az 5.2. táblázatból kiszámolható valószínűség értékek felhasználásával! Az Ω eseménytérnek *8* elemi eseménye van, a klasszikus valószínűség alapján ezeknek egyenként *1/8*-ad a valószínűségük. A táblázatban azonban a valószínűségi változók együttes megvalósulásának valószínűségeit kell megadnunk. Példaként a $\xi=0$ és az $\eta=0$ értékek egyszerre csak egyszer szerepelnek lehetséges értékként, ezért ennek valószínűsége *1/8*. A $\xi=1$ és az $\eta=1$ háromszor rendeltük hozzá elemi eseményekhez, a táblázatban tehát *3/8* valószínűség szerepel stb.

ξ/η	***0***	***1***	Σ ***sor***
0	*1/8*	*0*	*1/8*
1	*0*	*3/8*	*3/8*
2	*3/8*	*0*	*3/8*
3	*1/8*	*0*	*1/8*
Σ ***oszlop***	*5/8*	*3/8*	*1*

5.3. táblázat: A két valószínűségi változó együttes eloszlása és a peremeloszlások az 5.1. példához

A sorok és oszlopok végén a táblázatban szerepelnek a peremeloszlások is. Az η valószínűségi változó peremeloszlását az utolsó sor adja. Például, az első oszlop alján szereplő 5/8 érték adja meg annak valószínűségét, hogy az η valószínűségi változó *0* értéket vesz fel, azaz: $P(\eta=0)=5/8$. Hasonló módon a ξ valószínűségi változó peremeloszlását az utolsó oszlop értékei adják. Látszik a táblázatból az is, hogy ha a peremeloszlás sorokban szereplő értékeket összeadjuk (az utolsó oszlop értékeinek összege), akkor *1*-et kapunk. Hasonló módon az utolsó sor összege is *1*.

Gyors ellenőrző feladatok

5.1. A ξ és η valószínűségi változók együttes eloszlását tartalmazza az alábbi táblázat.

ξ/η	**-1**	**0**	**1**
-1	*1/8*	*1/12*	*1/24*
1	*3/8*	*1/4*	*3/24*

Határozzuk meg a két változó eloszlását külön-külön!

5.2. Az alábbi táblázat két valószínűségi változó együttes eloszlását adja meg.

ξ/η	-1	0	1
-1	p	$3p$	$5p$
1	$2p$	$4p$	$6p$

Adjuk meg p értékét!

5.3. Diszkrét valószínűségi változók függetlensége

Ahogyan a bevezetőben említettük, a gyakorlatban sokszor felmerül a két változó függetlensége vagy függőségének a kérdése. Az alábbiakban két valószínűségi változó függetlenségének feltételét definiáljuk.

Definíció. A ξ és η diszkrét valószínűségi változók lehetséges értékei legyenek $x_1, x_2, \ldots, x_n, \ldots$ és $y_1, y_2, \ldots, y_n, \ldots$ A ξ és η változókat függetlennek mondjuk, ha minden i-re és j-re teljesül, hogy

$$P(\xi = x_i, \eta = y_j) = P(\xi = x_i)P(\eta = y_j), \qquad (5.3.1.)$$

vagy a fentiekben már használt egyszerűbb jelölésekkel:

$$r_{ij} = p_i q_j .$$

Könnyen belátható, hogy a fenti definíció összhangban van az események függetlenségére vonatkozó definícióval. Ha a $\xi=x_i$ eseményt A-val jelöljük, az $\eta=y_k$ eseményt pedig B-vel, akkor az (5.3.1) összefüggés eseményekkel felírva: $P(AB)=P(A)P(B)$. Ez pedig éppen az események valószínűségeivel megfogalmazott függetlenség feltétele. A valószínűségi változók függetlenségére most adott definíció tehát összhangban van a korábban az események függetlenségre adott definícióval.

Gyors ellenőrző feladatok

5.3. Állapítsuk meg, hogy az 5.1. példában felírt valószínűségi eloszlások függetlenek-e?

5.4. Állapítsuk meg, hogy az 5.1. gyors ellenőrző feladatban definiált két valószínűségi változó független-e?

Haladóknak

5.4. Feltételes eloszlások diszkrét esetben

Az események kapcsán már definiáltuk az feltételes valószínűség fogalmát. *A* esemény *B* eseményre vonatkoztatott feltételes valószínűségét úgy definiáltuk, hogy

$$P(A|B) = \frac{P(AB)}{P(B)}.$$

Ezzel a definícióval összhangban most a valószínűségi változók eloszlása kapcsán definiáljuk a feltételes eloszlás fogalmát.

Definíció. A ξ és η diszkrét valószínűségi változók lehetséges értékei legyenek $x_1, x_2, \ldots, x_n, \ldots$ **és** $y_1, y_2, \ldots, y_n, \ldots$ A ξ valószínűségi változó $\eta = y_j$ feltételre vonatkozó **feltételes valószínűség-eloszlásának** nevezzük az alábbi kifejezést:

$$P(\xi = x_i|\eta = y_j) = \frac{P(\xi = x_i, \eta = y_j)}{P(\eta = y_j)} = \frac{r_{ij}}{q_j}. \tag{5.4.1.}$$

A függetlenség definíciójával kapcsolatban mondottak most is érvényesek. Látszik, hogy összhangban vagyunk az események feltételes valószínűségével kapcsolatban korábban adott definícióval.

Egy konkrét példa kapcsán belátjuk, hogy az (5.4.1.) definíció valóban valószínűség-eloszlást definiál, miközben *i* és *j* az összes lehetséges értéket felveszi.

5.2. Példa. Az 5.3. táblázatban megadott valószínűség-eloszlás esetén kiszámítjuk az (5.4.1.) definíció alapján a $P(\xi = x_i|\eta = y_j)$ feltételes eloszlás táblázatát. A táblázat első eleme:

$$P(\xi = x_1|\eta = y_1) = \frac{r_{11}}{q_1} = \frac{\frac{1}{8}}{\frac{5}{8}} = \frac{1}{5}.$$

A táblázat második sorának második eleme:

$$P(\xi = x_2|\eta = y_2) = \frac{r_{22}}{q_2} = \frac{\frac{3}{8}}{\frac{3}{8}} = 1.$$

Hasonló módon számolható a többi elem is. Tehát a ξ valószínűségi változó η változó szerinti feltételes valószínűség-eloszlása:

ξ/η	**0**	**1**
0	*1/5*	*0*
1	*0*	*1*
2	*3/5*	*0*
3	*1/5*	*0*
Σ oszlop	*1*	*1*

5.5. Folytonos valószínűségi változók együttes eloszlása

Ha a ξ és η valószínűségi változók folytonosak, akkor az együttes eloszlásuk folytonos eloszlásfüggvénnyel jellemezhető, ahhoz hasonlóan, ahogyan egy változó esetében tettük.

Definíció. Az

$$F(x,y)=P(\xi<x,\eta<y) \tag{5.5.1.}$$

kétváltozós függvényt a ξ és η valószínűségi változók **együttes eloszlásfüggvényének** nevezzük.

Az egydimenziós eloszlásfüggvényéhez hasonlóan bebizonyítható, hogy a *F(x,y)* kétváltozós függvényre is érvényesek az alábbi tulajdonságok:

1. A *F(x,y)* függvény mindkét változójának monoton (nem csökkenő) függvénye, azaz ha $x_1 \le x_2$ és $y_1 \le y_2$, akkor

$$F(x_1,y_1) \le F(x_2,y_2)\,. \tag{5.5.2.}$$

2. A *F(x,y)* függvény határértékei:

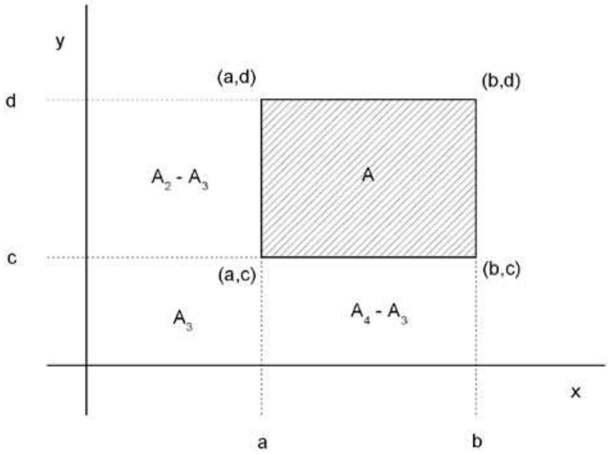
$$F(-\infty,y)=F(x,-\infty)=0;\;\; F(\infty,\infty)=1\,. \tag{5.5.3.}$$

Emlékeztetőül: az egyváltozós eloszlásfüggvény harmadik tulajdonsága arról szól, hogy a valószínűségi változó adott intervallumba esésének valószínűségét hogyan lehet kiszámítani az eloszlásfüggvény felhasználásával. A kétdimenziós esetben is megtehető ez, azonban az összefüggés kissé összetettebb, ezért itt ezt külön tételként adjuk meg.

Tétel. Ha a ξ és η valószínűségi változók együttes eloszlásfüggvénye *F(x,y)*, akkor a $P(a \le \xi \le b,\ c \le \eta \le d)$ valószínűség az alábbi módon számolható:

$$P(a \le \xi \le b;\, c \le \eta \le d) = F(b,d) - F(a,d) - F(b,c) + F(a,c). \qquad (5.5.4.)$$

Bizonyítás. A bizonyítás megkönnyítésére használjuk az 5.1. ábrát.



5.1. ábra: Ábra a kétváltozós eloszlásfüggvénnyel kapcsolatos tétel bizonyításához

Legyen A_1 esemény az, hogy a ξ és η változók a *(b,d)* csúccsal rendelkező síknegyedben vannak, vagyis

$$A_1 = (\xi < b,\ \eta < d).$$

Az eloszlásfüggvény definíciója alapján ennek az eseménynek a valószínűsége:

$$P(\xi < b, \eta < d) = P(A_1) = F(b,d).$$

Hasonló módon az az esemény, hogy a ξ és η változók a másik három csúccsal rendelkező térnegyedbe esnek rendre legyen a következő:

$$A_2 = (\xi < a,\ \eta < d); \quad A_3 = (\xi < a,\ \eta < c); \quad A_4 = (\xi < b,\ \eta < c).$$

Az eseményhalmazokra igaz, hogy

$$A = A_1 - (A_2 - A_3) - A_3 - (A_4 - A_3).$$

Az 5.1. ábrán látszik, hogy az A_2–A_3 és az A_4–A_3 síksávnak megfelelő, valamint az A_3 síknegyednek megfelelő események egymást kizáró események, valószínűségeikre tehát alkalmazható a 3. axióma. Ennek megfelelően:

$$P(A) = P(A_1) - P(A_2 - A_3) - P(A_3) - P(A_4 - A_3).$$

Mivel $A_3 \subset A_2$ és $A_3 \subset A_4$ ezért a különbséghalmazok valószínűsége egyenlő a valószínűségekkel, azaz

$$P(A) = P(A_1) - P(A_2) + P(A_3) - P(A_3) - P(A_4) + P(A_3) =$$
$$= P(A_1) - P(A_2) - P(A_4) + P(A_3).$$

Ha a fenti kifejezésben a valószínűségeket az eloszlásfüggvénnyel kifejezzük, akkor éppen a (5.5.4.) tétel állítására jutunk.

5.6. Együttes sűrűségfüggvény

Az egydimenziós eloszlások esetében a sűrűségfüggvényt az eloszlásfüggvény deriváltjaként definiáltuk. Ezzel analóg módon definiálható a kétdimenziós eloszlások esetén a sűrűségfüggvény.

Definíció. Ha a ξ és η együttes eloszlásfüggvénye $F(x,y)$, akkor az

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}, \tag{5.6.1.}$$

vegyes második parciális deriváltat a ξ és η együttes sűrűségfüggvényének nevezzük. Természetesen a definíció csak akkor érvényes, ha létezik ez a parciális derivált.

Megmutatható, hogy amennyiben létezik az $f(x,y)$ sűrűségfüggvény, akkor az $F(x,y)$ eloszlásfüggvény előállítható a sűrűségfüggvény kettős integráljaként, azaz

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y)\,dxdy. \tag{5.6.2.}$$

A kétdimenziós sűrűségfüggvényre is belátható, hogy a sík minden tartományán

$$f(x.y) \geq 0, \qquad -\infty < x < \infty, \quad -\infty < y < \infty,$$

és a normáltság is igaz, vagyis

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\,dxdy = 1.$$

Gyors ellenőrző feladatok

5.5. Legyen a ξ és η változók együttes eloszlásfüggvénye az alábbi:

$$F(x,y)=\begin{cases}\left(1-e^{-x}\right)\left(1-e^{-y}\right), & \text{ha} \quad x\geq 0, \quad y\geq 0,\\ 0, & \text{egyébként.}\end{cases}$$

Adjuk meg annak valószínűségét, hogy $0\leq\xi\leq 1$ és $0\leq\eta\leq 1$!

5.6. Adjuk meg az 5.5. feladatban definiált eloszlásfüggvény sűrűségfüggvényét!

5.7. Lássuk be, hogy az 5.6. feladatban kiszámított sűrűségfüggvény eleget tesz a sűrűségfüggvények tulajdonságainak!

5.7. Függetlenség folytonos valószínűségi változók esetén

Az események függetlenségét a $P(AB)=P(A)P(B)$ összefüggéssel definiáltuk. Ezzel teljes összhangban van a két folytonos valószínűségi változó függetlenségének definíciója.

Definíció. Az ξ és η valószínűségi változókat függetlennek nevezzük, ha az $A=(\xi<x)$ és a $B=(\eta<y)$ események függetlenek, azaz

$$P(\xi<x,\eta<y)=P(\xi<x)P(\eta<y).$$

Az egy és kétváltozós eloszlásfüggvények definícióit felhasználva ez a következőt jelenti:

$$F(x,y)=F(x)G(y), \tag{5.7.1.}$$

ahol $F(x,y)$ ξ és η együttes eloszlásfüggvénye, $F(x)$ a ξ változó, $G(y)$ pedig az η változó eloszlásfüggvénye.

A folytonos valószínűségi változók függetlenségének fenti definíciójából már következik függetlenség esetén a sűrűségfüggvényekre vonatkozó tétel.

Tétel. A ξ és η valószínűségi változók akkor és csak akkor függetlenek, ha a sűrűségfüggvényekre fennáll az alábbi összefüggés:

$$f(x,y)=f(x)g(y), \tag{5.7.2.}$$

ahol $f(x,y)$ a ξ és η változók együttes sűrűségfüggvénye, $f(x)$ és $g(y)$ pedig külön-külön a változók sűrűségfüggvénye.

Bizonyítás. A tételt csak arra az esetre bizonyítjuk, ha a függetlenség fennáll, akkor igaz (5.7.2). $F(x,y)=F(x)G(y)$ igaz, tehát

$$f(x,y)=\frac{\partial^2 F(x,y)}{\partial x\partial y}=\frac{\partial^2 F(x)G(y)}{\partial x\partial y}=\frac{\partial F(x)}{\partial x}\frac{\partial G(y)}{\partial y}=f(x)g(y).$$

Integrálással hasonlóan egyszerű belátni, hogy a tétel állítása visszafelé is igaz.

Gyors ellenőrző feladatok

5.9. Állapítsuk meg, hogy az 5.5. gyors ellenőrző feladatban megadott eloszlásfüggvény független valószínűségi változókat definiál-e!
5.10. Igaz-e az 5.5. feladat sűrűségfüggvényére az (5.7.2.) állítás?

5.8. Valószínűségi változók függvényének várható értéke

A valószínűségi változók függvényének várható értékéről már volt szó a 4.11. alfejezetben. A valószínűségi változók függvényének várható értékére vonatkozó tétel többdimenziós valószínűségi változóra is igaz.

Diszkrét változók esete

Legyen ξ és η két diszkrét valószínűségi változó, $x_1, x_2, \ldots, x_n, \ldots$ és $y_1, y_2, \ldots, y_n, \ldots$ lehetséges értékekkel. Legyen továbbá adott a két valószínűségi változó $P(\xi=x_i, \eta=y_j)=r_{ij}$ együttes eloszlása. Amennyiben a két változó függvénye $\zeta=\varphi(\xi,\eta)$, akkor ennek az új valószínűségi változónak a várható értéke:

$$M(\varphi(\xi,\eta))=\sum_i\sum_j \varphi(x_i,y_j)r_{ij}\,. \qquad (5.8.1.)$$

Folytonos változók esete

Legyen ξ és η két folytonos valószínűségi változó, amelyek együttes sűrűségfüggvénye $f(x,y)$. A két változó függvénye legyen most is $\zeta=\varphi(\xi,\eta)$. A (4.11.4) kifejezést alkalmazva a folytonos valószínűségi változók függvényének várható értéke:

$$M(\varphi(\xi,\eta))=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)f(x,y)dxdy\,. \qquad (5.8.2.)$$

Gyors ellenőrző feladatok

5.11. Számoljuk ki az 5.3. táblázatban megadott együttes eloszlás alapján a $\zeta=\varphi(\xi\eta)=\xi+\eta$ új változó várható értékét!

5.12. Számoljuk ki az 5.3. táblázatban megadott együttes eloszlás alapján a $\zeta=\varphi(\xi\eta)=\xi\,\eta$ szorzatfüggvény várható értékét!

5.9. Valószínűségi változók összegének várható értéke

Az előző alfejezetben általában beszéltünk arról, hogyan lehet kiszámolni valószínűségi változók függvényének várható értékét. Az általános képlet alapján összeg várható értéke is kiszámolható, ahogyan azt az 5.11. feladatban meg is tettük. Ugyanakkor, minthogy a valószínűségi változók összege gyakran szerepel a gyakorlatban, ezért most az összeg várható értékére egy egyszerűbb és gyorsabb megoldást is adunk.

Tétel. A ξ és η **valószínűségi változók összegének várható értéke** egyenlő a valószínűségi változók várható értékének összegével, azaz

$$M(\xi+\eta)=M(\xi)+M(\eta). \tag{5.9.1.}$$

Bizonyítás. A bizonyítást csak diszkrét esetre mutatjuk meg, folytonos esetre teljesen hasonló a menete. A bizonyítást az (5.8.1.) kifejezés alapján végezzük. Az általános képletbe a függvény helyére beírjuk a változók összegét:

$$M(\xi+\eta)=\sum_i\sum_j(x_i+y_j)r_{ij}=\sum_i\sum_j x_ir_{ij}+\sum_i\sum_j y_jr_{ij}\,.$$

A második tagban felcserélve az összegzés sorrendjét, valamint felhasználva az (5.2.1.) és (5.2.2.) azonosságokat, azt kapjuk, hogy

$$M(\xi+\eta)=\sum_i\sum_j x_ir_{ij}+\sum_i\sum_j y_jr_{ij}=\sum_i x_i\sum_j r_{ij}+\sum_j y_j\sum_i r_{ij}=$$
$$=\sum_i x_ip_i+\sum_i y_jq_j=M(\xi)+M(\eta).$$

Megjegyzés

Teljes indukcióval bebizonyítható, hogy ha ξ_1, ξ_2, ..., ξ_n valószínűségi változóknak létezik várható értéke, akkor összegük várható értéke:

$$M(\xi_1+\xi_2+....+\xi_n)=M(\xi_1)+M(\xi_2)+....+M(\xi_n). \tag{5.9.2.}$$

5.10. Valószínűségi változók szorzatának várható értéke

Az általános (5.8.1.) kifejezés alapján természetesen két valószínűségi változó szorzatának várható értéke is kiszámolható. Ha a valószínűségi változók nem függetlenek, akkor jobb megoldás nem is létezik. Ha azonban a szorzandó valószínűségi változók függetlenek, akkor a szorzat várható értékére egyszerűbb megoldás is adódik.

Tétel. Ha a ξ és η valószínűségi változók **függetlenek,** akkor a **szorzatuk várható értéke** egyenlő a várható értékeik szorzatával, azaz

$$M(\xi\eta) = M(\xi)\,M(\eta)\,. \tag{5.10.1.}$$

Bizonyítás. A bizonyítást most is diszkrét esetre végezzük, de az integrál és az összegzés hasonló tulajdonságai miatt folytonos esetre teljesen hasonló a bizonyítás menete. Ismét az (5.8.1.) általános képlet alapján indulunk el, tehát

$$M(\xi\eta) = \sum_i \sum_j x_i y_j r_{ij}\,. \tag{5.10.2.}$$

Most kihasználjuk a változók függetlenségét, tehát $r_{ij}=p_i q_j$. Ezt beírva (5.10.2)-be azt kapjuk, hogy

$$M(\xi\eta) = \sum_i \sum_j x_i y_j p_i q_j = \sum_i x_i p_i \sum_j y_j q_j = M(\xi)M(\eta)\,.$$

Megjegyzés

Teljes indukcióval bebizonyítható, hogy ha $\xi_1, \xi_2, \ldots, \xi_n$ valószínűségi változók függetlenek és létezik várható értékük, akkor szorzatuk várható értéke:

$$M(\xi_1\xi_2\cdot\ldots\cdot\xi_n) = M(\xi_1)M(\xi_2)\cdot\ldots\cdot M(\xi_n)\,. \tag{5.10.3.}$$

Gyors ellenőrző feladatok

5.13. Az 5.11. feladatban kitűzött számolást végezzük el az (5.9.1) összefüggés alapján is! Megnyugodhatunk, ha az eredmény egyezik az 5.11. feladat eredményével.

5.14. Az 5.12. feladat számolását végezzük el az (5.10.1) kifejezés alapján is! Az eredmény miért nem azonos az 5.12. feladat eredményével? Melyik a helyes eredmény?

5.11. Valószínűségi változók összegének szórása

A ξ és η valószínűségi változók összegének szórásnégyzete a definíciós egyenletből könnyen számolható.

Tétel. Ha ξ és η valószínűségi változók függetlenek, és létezik a szórásuk, akkor összegük szórásnégyzete megegyezik a szórásnégyzeteik összegével, azaz

$$D^2(\xi+\eta) = D^2(\xi) + D^2(\eta)\,. \tag{5.11.1.}$$

Bizonyítás. A bizonyításhoz először egy segédtételt látunk be, nevezetesen azt, hogy ha ξ és η valószínűségi változók függetlenek, akkor

$$M[(\xi - M(\xi))(\eta - M(\eta))] = 0\,. \tag{5.11.2.}$$

A zárójel felbontásával azt kapjuk, hogy

$$M[(\xi - M(\xi))(\eta - M(\eta))] = M[\xi\eta - M(\xi)\eta - \xi M(\eta) + M(\xi)M(\eta)].$$

Kihasználva a várható érték tulajdonságait, a tagonkénti várható érték képzés után adódik a segédtétel állítása, ha kihasználjuk a függetlenségből adódó $M(\xi\eta) = M(\xi)M(\eta)$ azonosságot.

Hozzáláthatunk most már az eredeti tétel bizonyításához! A szórásnégyzet definíciójából kiindulva:

$$D^2(\xi+\eta) = M\left[(\xi+\eta - M(\xi+\eta))^2\right].$$

Kihasználva a várható érték képzésének szabályát:

$$M(\xi+\eta) = M(\xi) + M(\eta),$$

majd beírva ezt a zárójelbe, azt kapjuk, hogy

$$D^2(\xi+\eta) = M\left[(\xi - M(\xi) + \eta - M(\eta))^2\right].$$

A szögletes zárójelen belül a négyzetre emelést végrehajtva:

$$M\left[(\xi - M(\xi) + (\eta - M(\eta))^2\right] = M\left[(\xi - M(\xi))^2 - 2((\xi - M(\xi)(\eta - M(\eta)) + (\eta - M(\eta)^2\right].$$

A várható érték képzést tagonként végrehajtva látjuk, hogy a középső tag a segédtétel állítása miatt zérus, tehát a végeredmény:

$$D^2(\xi+\eta) = M\left[(\xi - M(\xi))^2\right] + M\left[(\eta - M(\eta))^2\right] = D^2(\xi) + D^2(\eta),$$

és ez volt a bizonyítandó állítás.

Megjegyzés

1. Az (5.11.2.) összefüggés teljesüléséhez elegendő feltétel a függetlenség, de nem szükséges (azaz létezik más eset is, amikor fennáll). A fenti tétel igaz voltához tehát kevesebb is elegendő, mint a függetlenség. Elég, ha (5.11.2.) teljesül.

2. Teljes indukcióval belátható, hogy az összeg szórásnégyzetére vonatkozó (5.11.1.) tétel nemcsak kettő, hanem tetszőleges számú független valószínűségi változóra is igaz, vagyis

$$D^2(\xi_1+\xi_2+\ldots+\xi_n)=D^2(\xi_1)+D^2(\xi_2)+\ldots+D^2(\xi_n)\,. \qquad (5.11.3.)$$

Példa. Legyen *n* darab független valószínűségi változónk, amelyek szórásnégyzete azonos, vagyis

$$D^2(\xi_i)=\sigma^2\,, \qquad i=1,\ 2,\ \ldots,\ n.$$

Határozzuk meg a változók összegének szórásnégyzetét és szórását! Alkalmazzuk az (5.11.3.) azonosságot! Azt kapjuk, hogy

$$D^2(\xi_1+\xi_2+\ldots+\xi_n)=n\sigma^2\,. \qquad (5.11.4.)$$

Innen a szórás gyökvonással kapható:

$$D(\xi_1+\xi_2+\ldots+\xi_n)=\sqrt{n}\sigma\,. \qquad (5.11.5.)$$

Az eredményből látszik, hogy a szórás lassabban növekszik (gyökösen), mint a kísérletek száma. Ennek a gyakorlati alkalmazásokban jelentősége van!

6. KORRELÁCIÓ

6.1. Kovariancia

A mérési adatok közötti összefüggések tárgyalásakor már bevezettük az empirikus korrelációs együtthatót. Akkor láttuk, hogy ez a paraméter alkalmas a mérési jellemzők közötti lineáris kapcsolat jellemzésére. Most ugyanezt a problémát, nevezetesen a valószínűségi változók közötti kapcsolat problémáját fogjuk tárgyalni, a valószínűség-számítás eddig megismert apparátusa segítségével.

Mindenekelőtt emlékezzünk az előző fejezet (5.11.2.) segédtételére, ahol beláttuk, hogy amennyiben a ξ és η valószínűségi változók függetlenek, akkor igaz az alábbi összefüggés:

$$M[(\xi - M(\xi))(\eta - M(\eta))] = 0 \,.$$

Ez indokolja azt, hogy a változók közötti kapcsolat szorosságának mértékéül más esetben is ezt a szorzatot válasszuk.

Definíció. A

$$cov(\xi,\eta) = M[(\xi - M(\xi))(\eta - M(\eta))] \tag{6.1.1.}$$

várható értéket a ξ és η valószínűségi változók **kovarianciájának** nevezzük.

Megjegyzés

Ha a lehető legszorosabb a kapcsolat két változó között, vagyis $\xi=\eta$, akkor a kovariancia a $D^2(\xi)$ szórásnégyzettel egyezik meg.

Tétel. A kovariancia számolására a definíciós összefüggésnél alkalmasabb az alábbi képlet:

$$cov(\xi,\eta) = M(\xi\eta) - M(\xi)M(\eta) \,. \tag{6.1.2.}$$

Bizonyítás. A bizonyítás során a kovariancia definíciójából indulunk ki, majd a szögletes zárójelen belül felbontjuk a zárójeleket:

$$cov(\xi,\eta) = M[(\xi - M(\xi))(\eta - M(\eta))] =$$
$$M[\xi\eta - \xi M(\eta) - M(\xi)\eta + M(\xi)M(\eta)] = M(\xi\eta) - M(\xi)M(\eta) \,.$$

A kovariancia értékének nagyságát jelentősen befolyásolják a két valószínűségi változó lehetséges értékei. Ennek kiküszöbölésére vezetjük be a korrelációs együtthatót.

6.2. Korrelációs együttható

Definíció. Ha a ξ és η valószínűségi változók kovarianciája és szórásai léteznek, akkor a ξ és η változó korrelációs együtthatóján a

$$R(\xi,\eta)=\frac{cov(\xi,\eta)}{D(\xi)D(\eta)}=\frac{M(\xi\eta)-M(\xi)M(\eta)}{D(\xi)D(\eta)} \qquad (6.2.1.)$$

képlettel meghatározott számértéket értjük. Ez az érték csak akkor létezik, ha a nevezőben szereplő szórások egyike sem zérus.

A korrelációs együttható is a ξ és η valószínűségi változók függőségét, kapcsolatának szorosságát méri valamilyen értelemben. Erre vonatkoznak az alábbi tételek.

Tétel. Ha ξ és η független valószínűségi változók, akkor $R=0$.

Bizonyítás. Az $R(\xi,\eta)$ definíciójából adódik, hiszen független változók esetén a számláló nulla.

Megjegyzés

Az állítás nem megfordítható, vagyis abból, hogy $R=0$ nem következik a két változó függetlensége. Ilyenkor csak annyit mondhatunk, hogy a két változó korrelálatlan.

Tétel. Ha a két valószínűségi változó között lineáris a kapcsolat, vagyis $\eta=a\xi+b$, akkor

$$R(\xi,\eta)=\pm 1. \qquad (6.2.2.)$$

Az előjel a előjelétől függ.

Bizonyítás. A kovariancia számolás (6.1.2.) képletéből indulunk ki.

$$cov(\xi,\eta)=M(\xi\eta)-M(\xi)M(\eta).$$

Látszik, hogy ki kell számolnunk $\xi\eta$ és $M(\xi\eta)$ értékeket, valamint meg kell adnunk $M(\eta)$ értékét is.

$$\xi\eta=a\xi^2+b\xi,$$
$$M(a\xi^2+b\xi)=aM(\xi^2)+bM(\xi),$$

$$M(\eta) = aM(\xi) + b\,.$$

Tehát:

$$cov(\xi,\eta) = aM(\xi^2) + bM(\xi) - aM^2(\xi) - bM(\xi) = a\left(M(\xi^2) - M^2(\xi)\right) = aD^2(\xi)\,.$$

Most számoljuk ki (6.2.1.) nevezőjében $D(\eta)$ értékét:

$$D^2(\eta) = a^2 D^2(\xi)\,,$$

ahonnan

$$D(\eta) = |a| D(\xi)\,.$$

Beírva a kapott értékeket (6.2.1.) kifejezésbe:

$$R = \frac{cov(\xi,\eta)}{D(\xi)D(\eta)} = \frac{aD^2(\xi)}{|a|D^2(\xi)} = \frac{a}{|a|} = \pm 1\,. \qquad (6.2.3.)$$

Megjegyzés

1. A fordított állítás is igaz, vagyis ha $R(\xi,\eta)=\pm 1$, akkor a két változó között lineáris a kapcsolat. Ezt nem bizonyítjuk.
2. Belátható, hogy általában igaz az, hogy $|R| \le 1$.
3. Ha a két változó között nem lineáris a kapcsolat, akkor $-1 < R < 1$ akármilyen lehet, még 0 is. A korrelációs együttható tehát nem azt mutatja meg, hogy van-e valamilyen függvénykapcsolat a két változó között, hanem azt, hogy van-e lineáris kapcsolat, illetve azt, hogy a kapcsolat milyen erősen lineáris jellegű.

6.3. Lineáris regresszió

A 2.1. és a 2.2. fejezetben már láttuk, hogy a mérési adatok között lehet összefüggés, és gyakorta ez a függés lineáris, vagy lineárisra visszavezethető. Nézzük meg, hogy mit tudunk mondani ugyanerről a témáról a valószínűség-elmélet eddig megismert tételei segítségével.

A probléma általánosan is kezelhető, mi azonban megmaradunk a lineáris összefüggés vizsgálatánál. Feltételezzük tehát hogy a ξ és η valószínűségi változó között az alábbi összefüggés érvényes:

$$\eta = a\xi + b\,,$$

de nem ismerjük az egyenes a és b paraméterének értékét. A valószínűség-számítás nyelvére lefordítva keressük az a és b paramétereknek azt az értékét amelyek mellett az $a\xi+b$ értéke a lehető legjobban megközelíti η értékét, vagyis legyen a következő várható érték minimális:

$$M\left[(\eta-(a\xi+b))^2\right]=S(a,b)=min\,.$$

A feladat megoldásának első lépéseként a zárójelen belül elvégezzük a kijelölt műveleteket, majd tagonként képezzük a várható értéket.

$$\begin{aligned}S(a,b)&=M\left[(\eta-a\xi-b)^2\right]=M\left[\eta^2-a\xi\eta-b\eta-a\xi\eta+a^2\xi^2+ab\xi-b\eta+ab\xi+b^2\right]=\\&=M\left(\eta^2\right)-aM\left(\xi\eta\right)-bM\left(\eta\right)-aM\left(\xi\eta\right)+a^2M\left(\xi^2\right)+abM\left(\xi\right)-bM\left(\eta\right)+abM\left(\xi\right)+b^2=\\&=M\left(\eta^2\right)-2aM\left(\xi\eta\right)-2bM\left(\eta\right)+a^2M\left(\xi^2\right)+2abM\left(\xi\right)-2bM\left(\eta\right)+b^2\,.\end{aligned}$$

A minimum megtalálásához képezni kell a és b szerinti parciális deriváltakat, amelyeket egyenlővé kell tenni nullával.

$$\frac{\partial S(a,b)}{\partial a}=-2M(\xi\eta)+2bM(\xi)+2aM(\xi^2)=0\,,$$

$$\frac{\partial S(a,b)}{\partial b}=2aM(\xi)-M(\eta)+2b=0\,.$$

Egyszerűsítés és rendezés után azt kapjuk, hogy

$$aM(\xi^2)+bM(\xi)=M(\xi\eta)\,,$$

$$aM(\xi)+b=M(\eta)\,.$$

A két egyenletből a és b kifejezhető:

$$a=\frac{M(\xi\eta)-M(\xi)M(\eta)}{M(\xi^2)-M^2(\xi)}=\frac{cov(\xi,\eta)}{D^2(\xi)}=R(\xi,\eta)\frac{D(\eta)}{D(\xi)}\,, \qquad (6.3.1.)$$

$$b=M(\eta)-aM(\xi))\,, \qquad (6.3.2.)$$

ahol felhasználtuk a kovariancia (6.1.2) és a korrelációs együttható (6.2.1) képleteit. Tehát, ha megbizonyosodunk, hogy a ξ és η változó között lineáris a kapcsolat, akkor az így megtalált a és b értékekkel kiszámolt az $a\xi+b$ valószínűségi változó közelíti meg legjobban η értékét. A kifejezésekben szereplő függvények (kovariancia, korrelációs együttható, szórások, várható értékek) a két változó együttes eloszlásából számolhatók.

7. NEVEZETES ELOSZLÁSOK

Az alábbiakban olyan valószínűségi eloszlásokat tárgyalunk, amelyek az adatok statisztikus kezelése és a mérési eredmények feldolgozása során gyakran előkerülnek. Többet közülük az eddigiek során már megismertük, most azonban rendezett formában ezeknek is áttekintjük a tulajdonságait. Amikor ábrázoljuk is az eloszlásokat, akkor a diszkrét eloszlásokat pálcikaábrákkal jelenítjük meg, a folytonos eloszlásokat pedig általában a sűrűségfüggvény ábrázolásával adjuk meg.

7.1. Az indikátorváltozó eloszlása

Korábban már megismerkedtünk az indikátorváltozóval. Most felidézzük a definíciót és megadjuk az eloszlás jellemzőit.

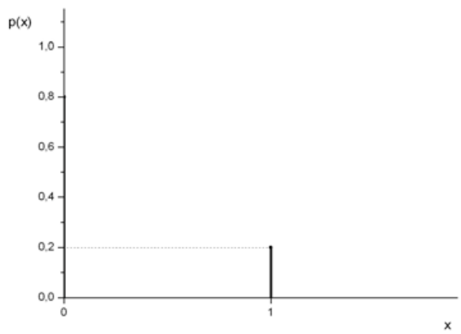
Legyen egy kísérlet két kimenetelű, melynek végeredménye A vagy $\overline{A}$. Az A esemény valószínűsége p, az $\overline{A}$ esemény valószínűsége $q=1-p$. A kísérlethez rendelt ξ valószínűségi változó, az ún. indikátorváltozó, amely az alábbi értékeket veszi fel:

$$\xi = \begin{cases} 1 & ha\, a \text{ } kísérletben \text{ } A \text{ } valósul\, meg, \\ 0 & ha\, a \text{ } kísérletben \text{ } \overline{A} \text{ } valósul\, meg. \end{cases} \qquad (7.1.1.)$$

A fenti értékekhez rendelt valószínűségek tehát:

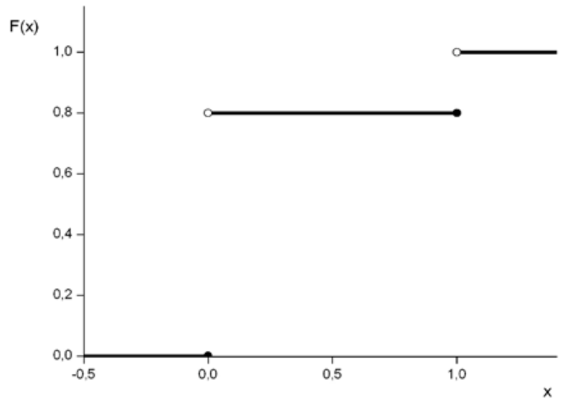
$$P(\xi = 0) = 1 - p = q\ ,$$
$$P(\xi = 1) = p\ .$$

Legyen $p=0,2$ és $q=0,8$. Az eloszlás pálcikaábrája a 7.1. ábrán látható.



7.1. ábra: A karakterisztikus változó pálcikaábrája p=0,2 esetén

A karakterisztikus változó eloszlásfüggvényét $p=0,2$ esetén a 7.2. ábrára rajzoltuk.



7.2. ábra: A karakterisztikus változó eloszlásfüggvénye p=0,2 esetén

A karakterisztikus valószínűségi változó várható értéke:

$$M(\xi) = 0q + 1p = p\,, \tag{7.1.2.}$$

szórásnégyzete pedig:

$$D^2(\xi) = M(\xi^2) - M^2(\xi) = 0q + 1^2 p - p^2 = p - p^2 = p(1-p) = pq\,. \tag{7.1.3.}$$

7.2. Az egyenletes eloszlás

Diszkrét eset

A ξ diszkrét valószínűségi változó **egyenletes eloszlású**, ha az $x_1, x_2, \ldots, x_n$ lehetséges értékeinek valószínűsége azonos. Az ilyen valószínűségi változót neveztük klasszikus változónak. A klasszikus változóval kapcsolatos meggondolásaink szerint a valószínűségek:

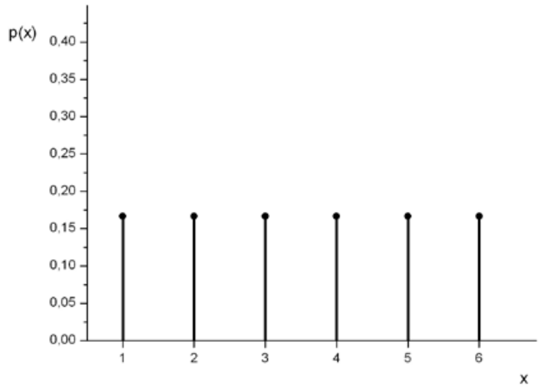
$$P(\xi = x_i) = \frac{1}{n}, \quad i = 1,2,\ldots,n\,. \tag{7.2.1.}$$

Az egyenletes eloszlás várható értéke és szórásnégyzete:

$$M(\xi) = \sum_{i=1}^{n} p_i x_i = \frac{1}{n}\sum_{i=1}^{n} x_i\,, \tag{7.2.2.}$$

$$D^2(\xi) = \frac{1}{n}\sum_{i=1}^{n}(x_i - M(\xi))^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^{\,2} - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2. \tag{7.2.3.}$$

7.1. Példa. A kockadobás esetén, ha a valószínűségi változó értékei a kocka lapjaira írt számok, akkor ennek a valószínűségi változónak egyenletes az eloszlása. Az eloszlás pálcikaábráját mutatja a 7.2. ábra.



7.3. ábra: Az egyenletes eloszlás pálcikaábrája kockadobás esetén

Gyors ellenőrző feladatok

7.1. A kockadobás esetére rajzoljuk fel az eloszlásfüggvényt!

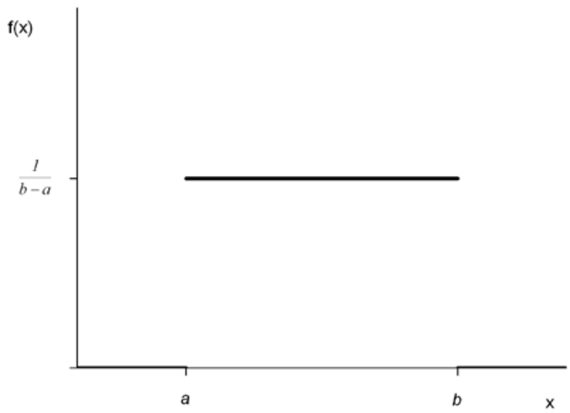
7.2. A kockadobás esetére számoljuk ki a várható értéket és a szórást!

Folytonos eset

A **folytonos** ξ valószínűségi változót **egyenletes eloszlásúnak** nevezzük az (a,b) intervallumon, ha sűrűségfüggvénye az alábbi:

$$f(x)=\begin{cases} \dfrac{1}{b-a} & \text{ha } a<x<b, \\ 0 & \text{egyébként.} \end{cases} \qquad (7.2.4.)$$

A sűrűségfüggvény alakja:

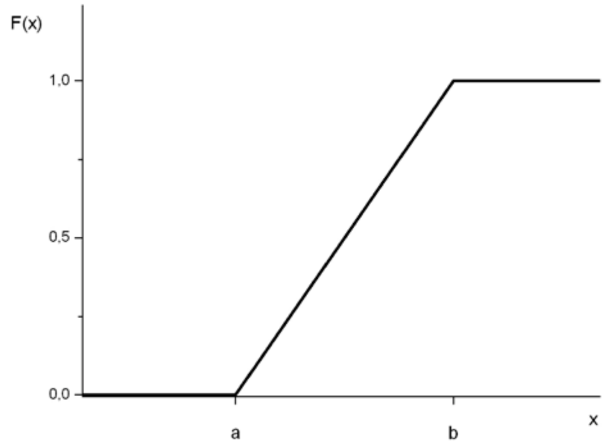


7.4. ábra: A folytonos egyenletes eloszlás sűrűségfüggvénye

A folytonos egyenletes eloszlás eloszlásfüggvénye, a definíció alapján:

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{a}^{b}\frac{1}{b-a}dx=\frac{1}{b-a}x\Big|_a^x=\frac{x-a}{b-a},\quad \text{ha } a\le x\le b \qquad (7.2.5.)$$

ha $x<a$, akkor $F(x)=0$, ha $x>a$, akkor $F(x)=1$, ahogyan azt a 7.5. ábra is mutatja.



7.5. ábra: A folytonos egyenletes eloszlás eloszlásfüggvénye

A várható érték és a szórásnégyzet folytonos esetben:

$$M(\xi)=\int_a^b\frac{x}{b-a}dx=\frac{b+a}{2}, \qquad (7.2.6.)$$

$$D^2(\xi)=\int_a^b\frac{x^2}{b-a}dx-\left(\frac{b+a}{2}\right)^2=\frac{(b-a)^2}{12}. \qquad (7.2.7.)$$

7.3. A Bernoulli-eloszlás

Mint a Bernoulli-kísérletsorozatban, a kísérletnek legyen két kimenetele, A és $\overline{A}$. Az A esemény valószínűsége: $P(A)=p$, az $\overline{A}$ esemény valószínűsége pedig: $P(\overline{A})=1-p=q$. Végezzük el a kísérletet n-szer, egymástól függetlenül. Legyen a ξ valószínűségi változó az n számú kísérlet közül azok száma, amelyekben A következett be. Korábban már láttuk, hogy ilyenkor annak valószínűsége, hogy n kísérlet során az A esemény k-szor következik be:

$$P(\xi=k)=B_p(n,k)=\binom{n}{k}p^k q^{n-k};\quad k=0,1,2,\dots,n. \qquad (7.3.1.)$$

Ezt az eloszlást **Bernoulli-eloszlásnak,** vagy sokszor **binomiális eloszlásnak** nevezzük, amely diszkrét eloszlás. Az alábbiakban az eloszlás tulajdonságait ismerjük meg részletesebben.

Mindenekelőtt be kell látnunk, hogy a (7.3.1) kifejezés valóban valószínűség eloszlás, azaz normált, ami azt jelenti, hogy az összes lehetséges értékre összegezve a valószínűség értékeket *1*-et kapunk.

Normáltság

$$\sum_{k=0}^{n} B_p(n,k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = 1.$$

Bizonyítás

A bizonyítás során felhasználjuk a binomiális tételt, miszerint

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} . \qquad (7.3.2.)$$

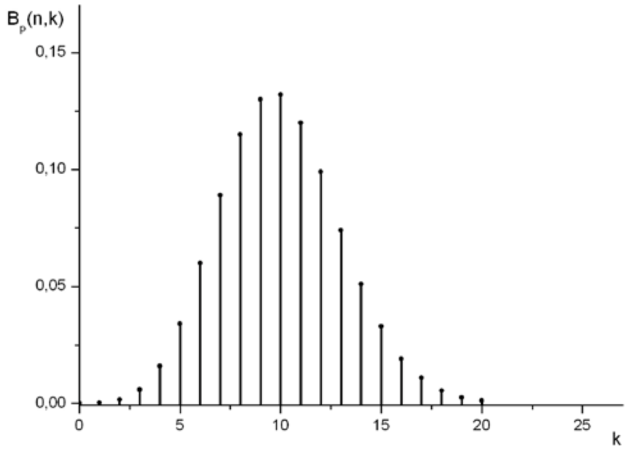
Ha most a (7.3.2) képletben végrehajtjuk az $a=p$ és $b=q$ helyettesítést, akkor azt kapjuk, hogy

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = (p+1-p)^n = 1,$$

és éppen ezt akartuk belátni.

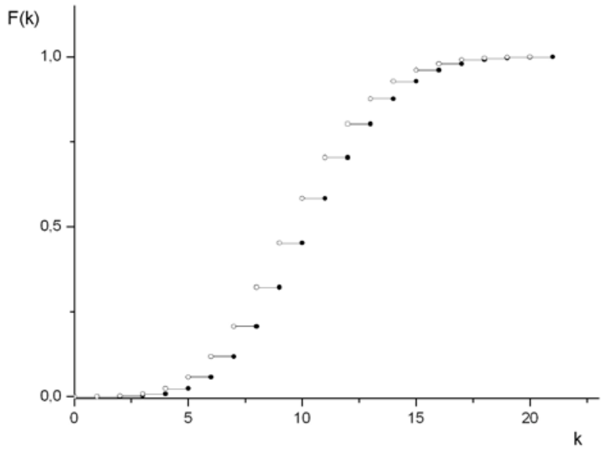
Az eloszlás alakja

Az alábbi ábra $p=0,1$ és $n=100$ esetén mutatja a Bernoulli-eloszlás alakját pálcikaábra formájában.



7.6. ábra: A Bernoulli-eloszlás alakja p=0,1 és n=100 esetén

A Bernoulli-eloszlás tehát diszkrét eloszlás, amelynek eloszlásfüggvényét a 7.7. ábra mutatja.



7.7. ábra: A Bernoulli-eloszlás eloszlásfüggvénye p=0,1 és n=100 esetén

Rekurziós képlet

A rekurziós képlet alapján az eloszlás *k-1.* értékéből kiszámolható a *k.* értéke, tehát

$$B_p(n,k) = \frac{n-k+1}{k}\frac{p}{1-p}B_p(n,k-1). \qquad (7.3.3.)$$

Bizonyítás

$$\frac{B_p(n,k)}{B_p(n,k-1)}=\frac{\binom{n}{k}p^k(1-p)^{n-k}}{\binom{n}{k-1}p^{k-1}(1-p)^{n-k+1}}=\frac{\frac{n!}{k!(n-k)!}}{\frac{n!}{(k-1)!(n-k+1)!}}\frac{p}{1-p}.$$

Az egyszerűsítések után éppen a bizonyítandó (7.3.3) összefüggésre jutunk.

A Bernoulli-eloszlás várható értéke

Tétel. A Bernoulli-eloszlás várható értéke

$$M(\xi)=np. \tag{7.3.4.}$$

Bizonyítás. A várható érték kiszámítására két módszert is megmutatunk.

a. A várható érték definíciója alapján a Bernoulli-eloszlás várható értéke:

$$M(\xi)=\sum_{k=0}^{n}k\binom{n}{k}p^kq^{n-k}=\sum_{k=0}^{n}k\frac{n!}{k!(n-k)!}p^kq^{n-k}.$$

Mivel $k=0$ esetén az első tag maga is nulla, ezért az összegzés $k=1$-től indulhat, tehát

$$M(\xi)=\sum_{k=1}^{n}k\frac{n!}{k!(n-k)!}p^kq^{n-k}=np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}q^{n-k},$$

ahol k-val egyszerűsítettünk, és az összegzésből kiemeltünk n-et és p-t. Vegyük észre, hogy az összegzésen belül most

$$\binom{n-1}{k-1}p^{k-1}q^{n-k}=B_p(n-1,k-1).$$

Ha most bevezetjük a $k^*=k-1$ új változót, akkor erre átírva az összegzést

$$M(\xi)=np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}q^{n-k}=np\sum_{k^*=0}^{n-1}\frac{(n-1)!}{k^*!(n-1-k^*)!}p^{k^*}q^{n-1-k^*}=$$

$$=np(p+q)^{n-1}=np.$$

Itt az utolsó lépésben alkalmaztuk a binomiális tételt, és azt, hogy $(p+q)^{n-1}=1$.

b. A várható érték kiszámítható az indikátorváltozók segítségével is. Az n mérés során minden egyes méréshez rendeljünk hozzá egy η_i indikátorváltozót, ahol $i=1, 2, ..., n$. Tehát

$$\eta_i = \begin{cases} 1, & \text{ha az } i. \text{ kísérletben az } A \text{ esemény valósul meg,} \\ 0, & \text{ha } i. \text{ kísérletben az } \overline{A} \text{ valósul meg.} \end{cases}$$

Mivel a kísérleteket egymástól függetlenül végezzük, ezért az η_i változók független valószínűségi változók. Adott kísérletsorozatban az indikátorváltozók értékeinek összege éppen k–t ad, tehát az η_i változók összege éppen ξ, vagyis

$$\xi = \eta_1 + \eta_2 + ... + \eta_n .$$

Számítsuk ki a várható értéket!

$$M(\xi) = M(\eta_1 + \eta_2 + ... + \eta_n) = M(\eta_1) + M(\eta_2) + ... + M(\eta_n) . \qquad (7.3.5.)$$

Az indikátorváltozó várható értékét már korábban kiszámítottuk (7.1.2), vagyis

$$M(\eta_i) = p \quad \forall\, i - re ,$$

tehát a (7.3.5) összege:

$$M(\xi) = M(\eta_1) + M(\eta_2) + ... + M(\eta_n) = np .$$

Ezzel ismét a korábban kapott várható értékhez jutottunk.

A Bernoulli-eloszlás szórása

Tétel. A Bernoulli-eloszlás szórását az alábbi összefüggés adja:

$$D(\xi) = \sqrt{npq} . \qquad (7.3.6.)$$

Bizonyítás. Természetesen most is választhatnánk a közvetlen bizonyítást, azaz a szórásnégyzet és a szórás definíciója alapján a szórás közvetlen kiszámítását. Ehelyett sokkal könnyebben célhoz érünk az indikátorváltozók segítségével. Ismét felhasználjuk, hogy ξ a fentiekben definiált n darab független valószínűségi változó összege. Szórásnégyzete tehát:

$$D^2(\xi) = D^2(\eta_1 + \eta_2 + ... + \eta_n) = D^2(\eta_1) + D^2(\eta_2) + ... + D^2(\eta_n) .$$

Az indikátorváltozó szórásnégyzetére vonatkozó (7.1.3) képlet szerint

$$D^2(\eta_i) = pq \quad \forall\ i - re ,$$

tehát

$$D^2(\xi) = D^2(\eta_1) + D^2(\eta_2) + \ldots + D^2(\eta_n) = npq\,.$$

A szórás pedig:

$$D(\xi) = \sqrt{npq}\,,$$

és ez az, amit be akartunk látni.

A Bernoulli-eloszlás módusza

A Bernoulli-eloszlásnak két paramétere van: n és p. Adott eloszlás esetén n és p konkrét számértékek. Miközben k változik, a hozzárendelt valószínűség értékek kezdetben növekszenek, majd csökkennek, ahogyan azt a 7.6. ábra esetén láthatjuk is. Kérdés, hogy milyen k értéknél lesz az eloszlásnak maximuma? Ameddig növekszenek a valószínűség értékek, addig igaz, hogy

$$B_p(n,k-1) < B_p(n,k)\,.$$

Amikor csökkennek a valószínűség értékek, akkor

$$B_p(n,k-1) > B_p(n,k)\,,$$

A csúcs közelében esetleg az egyenlőség is megengedett

$$B_p(n,k-1) = B_p(n,k)\,.$$

A fenti három összefüggés úgy is írható, hogy

$$\frac{B_p(n,k)}{B_p(n,k-1)} \lesseqqgtr 1,$$

ami a (7.3.3) rekurziós formula szerint megegyezik azzal, hogy

$$\frac{B_p(n,k)}{B_p(n,k-1)} = \frac{n-k+1}{k}\frac{p}{1-p} \lesseqqgtr 1.$$

Átrendezve az egyenlőtlenséget, azt kapjuk, hogy

$$(n+1)\,p \lesseqqgtr 1.$$

A maximum keresésekor, tehát azt kell vizsgálnunk, hogy k növekedésével mikor vált a reláció kisebbről nagyobbra. Két eset lehetséges: a. $(n+1)p$ nem egész szám, vagy pedig b. $(n+1)p$ egész szám.

Tekintsük először az a. esetet:

Miközben k növekszik, kezdetben az $(n+1)p>k$ egyenlőtlenség teljesül. Eddig a $B_p(n,k)>B_p(n,k-1)$ reláció teljesül. Ahogy k nő, van olyan értéke k-nak, amikor megfordul a reláció, azaz innentől kezdve a $B_p(n,k)<B_p(n,k-1)$ reláció lesz az igaz. Ez a k érték, ahol ez a váltás bekövetkezik az $(n+1)p$ egész részével lesz egyenlő, azaz

$$[(n+1)p]=k\,. \qquad (7.3.7.)$$

Ennél a k értéknél lesz az eloszlásnak a maximuma, vagyis ez a k az eloszlás módusza.

Tekintsük a b. esetet:

Ha történetesen az $(n+1)p$ érték egész szám, akkor van olyan k érték, ahol az egyenlőség teljesül, vagyis

$$(n+1)p=k\,.$$

Itt maximuma lesz az eloszlásnak, viszont ekkor igaz az is, hogy

$$B_p(n,k-1)=B_p(n,k)\,,$$

ami azt jelenti, hogy két maximális értéke is van az eloszlásnak a

$$k_1=(n+1)p \quad \text{és} \quad k_2=(n+1)p-1 \qquad (7.3.8.)$$

értékeknél. Ilyenkor bimodálisnak nevezzük az eloszlást.

7.2. Példa. Zárthelyin a hallgatók tesztet oldanak meg. 5 tesztkérdésre kell válaszolni, minden kérdésre három lehetséges megoldást kínál a teszt, de ezek közül csak egy a helyes. Ha valaki egyáltalán nem készült, és a kérdésekre véletlenszerűen válaszol, úgy hogy a válaszok függetlenek egymástól, akkor mi annak a valószínűsége, hogy legalább három kérdésre helyes választ ad?

Egy kérdésre a helyes válasz valószínűsége: $p=1/3$. Legalább három kérdésre adott helyes válasz azt jelenti, hogy vagy három, vagy négy, vagy mind az öt kérdésre helyes a válasz. A Bernoulli-eloszlás szerint, annak valószínűsége, hogy öt közül pontosan három kérdésre lesz helyes a válasz:

$$P(\xi=3)=B_p(5,3)=\binom{5}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^2=0{,}1646...$$

Hasonló módon annak valószínűsége, hogy pontosan négy feladat megoldása lesz helyes:

$$P(\xi=4)=B_p(5,4)=\binom{5}{4}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^1=0{,}0411\ldots,$$

Annak valószínűsége, hogy mind az öt megoldás helyes:

$$P(\xi=5)=B_p(5,5)=\binom{5}{5}\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right)^0=0{,}0041\ldots$$

Tehát, hogy legalább három feladat megoldása helyes, ennek a három valószínűségnek az összege, azaz

$$P(\xi\geq 3)=0{,}2098\ldots$$

Gyors ellenőrző feladatok

7.3. A 7.2. példa adatai alapján számítsuk ki annak valószínűségét, hogy véletlenszerű válaszadás esetén egyetlen feladat megoldása sem lesz helyes? (*0,13168*)

7.4. A 7.2. példában mi annak a valószínűsége, hogy legalább egy feladat megoldása helyes? (*1-0,13168=0,86832*)

7.4. A Poisson-eloszlás

A műszaki és tudományos gyakorlatban számos olyan véletlen folyamat van, amelynek leírását a Poisson-eloszlás adja (ejtsd Poászon!). A Poisson-eloszlás diszkrét eloszlás. Általában azt lehet mondani, hogy kis valószínűségű események térbeli elrendeződését, vagy időbeli lefolyását gyakran Poisson-eloszlás írja le. A térben ilyen jelenség például az erdőben adott területen a fák száma, a távcső látómezejében található csillagok száma, a vörös vérsejtek száma a mikroszkóp látómezejében, adott területen bizonyos növény egyedeinek a száma stb.. Időbeli lefolyásra példák: a telefonközpontba bizonyos idő alatt beérkező telefonhívások száma; adott idő alatt atomok radioaktív bomlásának száma; a Geiger-Müller-számlálóba adott idő alatt beérkező kozmikus részecskék száma stb. A Poisson-eloszlásnak eleget tevő időben lejátszódó folyamatokat **Poisson-folyamatnak** nevezzük. Az első ismert alkalmazása a Poisson-eloszlásnak a porosz hadseregben az egy évben lórúgásoktól meghalt katonák számának becslése volt.

Definíció. A ξ diszkrét valószínűségi változó eloszlását λ paraméterű Poisson-eloszlásnak nevezzük, ha a valószínűségi változó lehetséges értékei $k=0,\ 1,\ 2,\ \ldots$ egész számok, és a $\xi=k$ esemény valószínűsége:

$$P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \text{ahol} \quad \lambda > 0. \tag{7.4.1.}$$

Normáltság

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1,$$

ahol felhasználtuk az ismert hatványsor alakot:

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}, \quad \text{ha } \lambda > 0.$$

A Poisson-eloszlás várható értéke és szórása

Tétel. A λ paraméterű Poisson-eloszlás várható értéke

$$M(\xi) = \lambda. \tag{7.4.2.}$$

Bizonyítás.

$$M(\xi) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} \underbrace{\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}}_{e^{\lambda}} = \lambda,$$

hiszen $k - 1 = k^*$ helyettesítéssel:

$$\sum_{k^*=0}^{\infty} \frac{\lambda^{k^*}}{k^*!} = e^{\lambda}.$$

Tétel. A λ paraméterű Poisson-eloszlás szórásnégyzete

$$D^2(\xi) = \lambda, \tag{7.4.3.}$$

és szórása

$$D(\xi) = \sqrt{\lambda}. \tag{7.4.4.}$$

Bizonyítás.

$$D^2(\xi) = M(\xi^2) - M^2(\xi) = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} - \lambda^2,$$

$$\sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} = \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!} + \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} = \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} = \lambda^2 e^{\lambda} + \lambda e^{\lambda}$$

$$D^2(\xi) = \lambda^2 e^{\lambda} e^{-\lambda} + \lambda e^{\lambda} e^{-\lambda} - \lambda^2 = \lambda ,$$

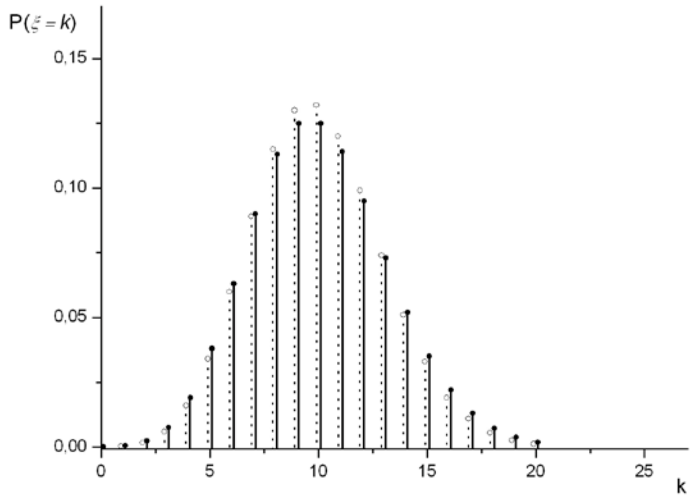
tehát a szórás:

$$D(\xi) = \sqrt{\lambda} .$$

Arra jutottunk tehát, hogy a Poisson-eloszlás λ paramétere egyúttal az eloszlás várható értéke és szórásnégyzete is.

Az eloszlás alakja

A Poisson-eloszlás alakja hasonló a Bernoulli-eloszlás alakjához. Mennél kisebb a p és mennél nagyobb n, miközben $np=\lambda$, a Poisson-eloszlás annál közelebb esik a Bernoulli-eloszláshoz. Ilyen feltételek mellett a Poisson-eloszlást a Bernoulli-eloszlás közelítésére is lehet használni. A 7.8. ábrán a $\lambda=10$ paraméterű Poisson-eloszlás (teli vonal), és a $p=0,1$, $n=100$ paraméterű Bernoulli-eloszlás (pontozott vonal) összehasonlítása látszik.



7.8. ábra: A Poisson-eloszlás (teli vonal) λ=10, és a Bernoulli-eloszlás (pontozott vonal) p=0,1, n=100 paraméterekkel

7.3. Példa. Radioaktív anyag bomlásának mérésekor a Geiger–Müller-számlálóval percenként átlagosan $n=10$ [db/min] beütésszámot mérünk. Mi annak a valószínűsége, hogy fél perc alatt nem érkezik beütés a számlálóba? Megoldás: fél perc alatt a beütés várható értéke: $\lambda=10$ [1/min] × $0,5$ [min]=5. A beütésszámra Poisson-eloszlást feltételezve, annak valószínűsége, hogy fél perc alatt egyetlen bomlás sem történik:

$$P(\xi = 0) = \frac{5^0}{0!} e^{-5} = 0,0067...$$

Gyors ellenőrző feladatok

7.5. A fenti példa adatait felhasználva adjuk meg a beütésszám szórását!

7.6. Adjunk rekurziós formulát a Poisson-eloszlásra!

7.5. A geometriai eloszlás

Egy kísérletnek két kimenetele van A és $\overline{A}$. Ezek valószínűségei: $P(A)=p$ és $P(\overline{A})=q=1-p$. A kísérletet egymástól függetlenül sokszor elvégezve, és ezt a sok kísérletet tekintsük egy kísérletnek. A kísérletet addig végezzük, ameddig az A esemény először bekövetkezik. A kísérlet kimenetele tehát $\overline{A}$-ból és A-ból álló sorozat. A ξ valószínűségi változó legyen annak a kísérletnek a sorszáma, amelyben először fordul elő az A esemény. Tehát ξ lehetséges értékei: *1, 2, 3, ...* Annak valószínűsége tehát, hogy $\xi=k$, vagyis, hogy az A esemény a k. kísérletben fordul elő először:

$$P_k = P(\xi = k) = q^{k-1}p \text{ , ahol } k=1,\ 2,\ \ldots, \qquad (7.5.1.)$$

Ezt az eloszlást **geometriai eloszlásnak** nevezzük.

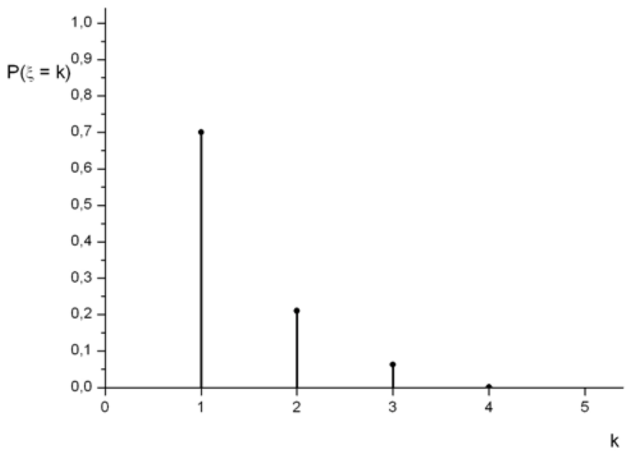
Normáltság

$$\sum_{k=1}^{\infty} P(\xi=k) = p\sum_{k=1}^{\infty} q^{k-1} = \frac{p}{q}\sum_{k=1}^{\infty} q^k = p\frac{1}{1-q} = 1 \ ,$$

ahol kihasználtuk a geometriai sorra vonatkozó összegképletet:

$$\sum_{k=1}^{\infty} q^k = \frac{q}{1-q} . \qquad (7.5.2.)$$

Az eloszlás alakja



7.9. ábra: A p=0,3 paraméterű geometriai eloszlás pálcikaábrája

A geometriai eloszlás várható értéke

Tétel. A geometriai eloszlás várható értéke:

$$M(\xi)=\frac{1}{p}. \tag{7.5.3.}$$

Bizonyítás

$$M(\xi)=\sum_{k=1}^{\infty}kq^{k-1}p=p\frac{1}{(1-q)^2}=\frac{1}{p},$$

ahol felhasználtuk azt a függvénysorokra érvényes szabályt, hogy

$$f(x)=\sum_{n=1}^{\infty}f_n(x)\ \rightarrow\ f'(x)=\sum_{n=1}^{\infty}f_n'(x).$$

A geometriai eloszlás szórásnégyzete

Tétel. A geometriai eloszlás szórásnégyzete:

$$D^2(\xi)=\frac{q}{p^2}. \tag{7.5.4.}$$

Bizonyítás

$$D^2(\xi)=M(\xi^2)-M^2(\xi)=\sum_{k=1}^{\infty}k^2q^{k-1}p-\frac{1}{p^2}=\frac{q}{p^2},$$

felhasználva, hogy

$$M(\xi^2)=\sum_{k=1}^{\infty}k^2pq^{k-1}=\frac{1+q}{(1-q)^2},$$

amit (7.5.2) kétszeres deriválása után kapott

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

kifejezés kis átalakításával nyerhetünk.

7.6. Az exponenciális eloszlás

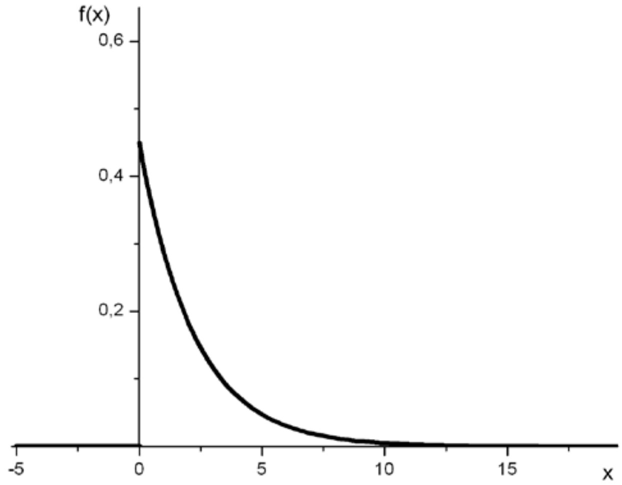
Definíció. A ξ folytonos valószínűségi változó λ paraméterű exponenciális eloszlású, ha sűrűségfüggvénye:

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & ha\ x\geq 0\\ 0, & ha\ x<0\,.\end{cases} \tag{7.6.1.}$$

A sűrűségfüggvény ismeretében az eloszlásfüggvény már számolható. Ennek megfelelően a ξ valószínűségi változó eloszlásfüggvénye:

$$F(x)=P(\xi<x)=\int_{-\infty}^{x}\lambda e^{-\lambda t}dt=\begin{cases}1-e^{-\lambda x}, & ha\ x\geq 0\\ 0, & ha\ x<0.\end{cases} \tag{7.6.2.}$$

Az eloszlás alakja

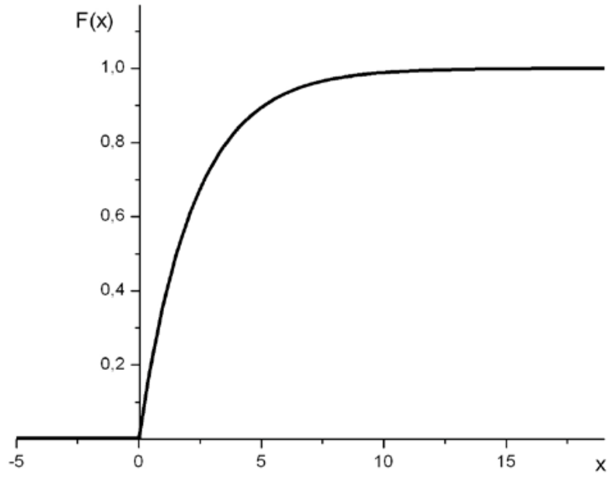


7.10. ábra: A λ=0,45 paraméterű exponenciális eloszlás sűrűségfüggvénye

Normáltság

Az hogy a (7.6.1) függvény valóban valószínűségi sűrűségfüggvény a normáltság ellenőrzésével látható be.

$$\int_{-\infty}^{\infty}f(x)dx=\int_{-\infty}^{0}0dx+\int_{0}^{\infty}\lambda e^{-\lambda x}dx=\left[-e^{-\lambda x}\right]_{0}^{\infty}=1\,.$$



7.11. ábra: A λ=0,45 paraméterű exponenciális eloszlás eloszlásfüggvénye

Normáltság

Az hogy a (7.6.1) függvény valóban valószínűségi sűrűségfüggvény a normáltság ellenőrzésével látható be.

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{\infty} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_{0}^{\infty} = 1 .$$

A exponenciális eloszlás várható értéke

$$M(\xi) = \int_{0}^{\infty} xf(x)dx = \lambda \int_{0}^{\infty} xe^{-\lambda x} dx = \frac{1}{\lambda}, \qquad (7.6.3.)$$

ahol az integrál parciális integrálással számolható ki.

Szórásnégyzet és szórás

$$D^2(\xi) = \int_{0}^{\infty} x^2 \lambda e^{-\lambda x} dx - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}. \qquad (7.6.4.)$$

Az integrál itt is parciális integrálással számolható. Tehát a szórás:

$$D(\xi) = \frac{1}{\lambda}. \qquad (7.6.5.)$$

Az exponenciális eloszlás élettartamokat leíró valószínűségi változók leírására alkalmas. Az ilyen élettartamokra az jellemző, hogy ha az élettartam x időpontig nem ért véget, akkor úgy tekinthető, mintha a folyamat x időpontban kezdődne. Sokszor ezt úgy fejezik ki, hogy elfelejtődik a múlt, és az élettartam a megfigyelés kezdetén kezdődik el, vagyis az ilyen valószínűségi változó *örökifjú tulajdonsággal* rendelkezik.

A fizikából ismeretes, hogy egy radioaktív atom élettartama exponenciális eloszlású. Hasonló módon exponenciális eloszlású egy lámpa vagy egy mikroprocesszor élettartama.

7.4. Példa. A Mössbauer-effektus mérése során használt Co^{57} atomok felezési ideje *0,75* év (*270* nap). Határozzuk meg, hogy egy atom élettartamának mekkora a várható értéke?

Megoldás: Az élettartam exponenciális eloszlást követ. Az eloszlás függvény tehát

$$F(x) = 1 - e^{-\lambda x}$$

alakú. A felezési idő azt jelenti, hogy az atom az x időpontig $F(x)=1/2$ valószínűséggel elbomlik. Tehát:

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

ahonnan

$$x = \frac{\ln 2}{\lambda}.$$

Ez a kifejezés adja meg a felezési idő és az eloszlás paramétere közötti kapcsolatot. A feladat adataival:

$$\lambda = \frac{\ln 2}{0{,}75 \text{ év}} = 0{,}924 \frac{1}{\text{év}}.$$

Mivel a várható érték $M(\xi)=1/\lambda$, innen a feladat megoldása, azaz egy atom élettartamának várható értéke: $M(\xi)=1{,}082$ év.

7.7. A normális eloszlás (Gauss-eloszlás)

A statisztikában az egyik leggyakoribb eloszlás a normális eloszlás, amelyet sokszor Gauss-eloszlásnak is nevezünk. A későbbiekben még visszatérünk arra, hogy miért fordul elő a normális eloszlás olyan gyakran.

Definíció. A ξ valószínűségi változó normális eloszlású, ha sűrűségfüggvénye:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad \text{ahol} \quad -\infty < x < \infty . \qquad (7.7.1.)$$

A eloszlásnak két paramétere van m és σ. Az m tetszőleges valós szám, σ pedig pozitív állandó. A normális eloszlás szokásos jelölése: $N(m,\sigma)$.

A normális eloszlás eloszlásfüggvénye

Az eloszlásfüggvény a sűrűségfüggvényből integrálással kapható meg:

$$F(x)=\int_{-\infty}^{x} f(t)dt=\int_{-\infty}^{x}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t-m)^2}{2\sigma^2}}dt . \qquad (7.7.2.)$$

Normáltság

A normális eloszlás folytonos eloszlás, tehát az alábbi integrálról kell belátni, hogy értéke egységnyi:

$$\int_{-\infty}^{\infty} f(x)dx=\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}dx .$$

Az integrált új változó bevezetésével hozzuk egyszerűbb alakra. Az új változó:

$$u=\frac{x-m}{\sigma}; \quad du=\frac{dx}{\sigma} .$$

Az új változóval kifejezve az integrált:

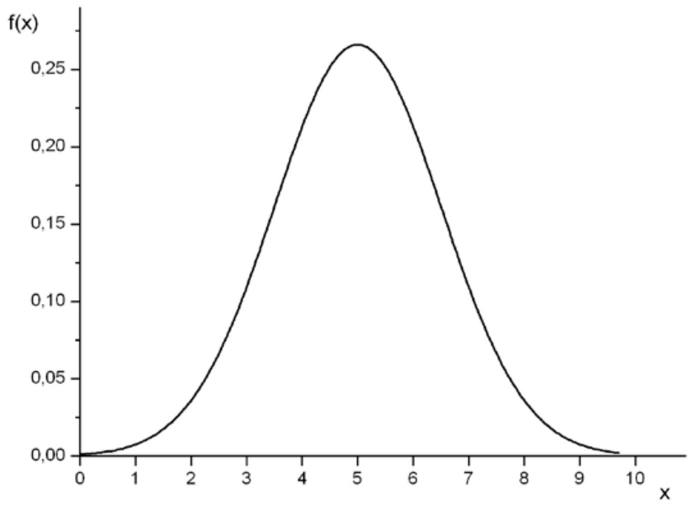
$$\int_{-\infty}^{\infty} f(x)dx=\frac{1}{\sigma\sqrt{2\pi}}\sigma\sqrt{2}\int_{-\infty}^{+\infty}e^{-u^2}du=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty}e^{-u^2}du=1 ,$$

és ez az amit be akartunk látni. Itt felhasználtuk az alábbi nevezetes integrál értéket:

$$\int_{0}^{+\infty}e^{-x^2}dx=\frac{\sqrt{\pi}}{2} . \qquad (7.7.3.)$$

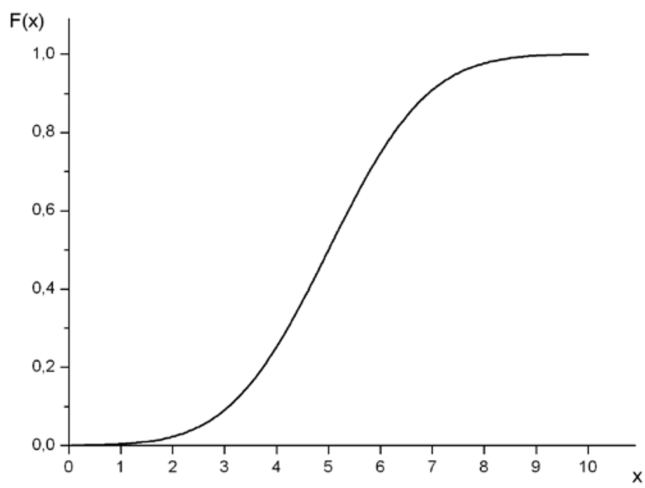
Az eloszlás alakja

Az $m=5$ és $\sigma=1,5$ paraméterű normális eloszlás sűrűségfüggvénye a 7.12. ábrán látható.



7.12. ábra: Az m=5 és σ=1,5 paraméterű normális eloszlás sűrűségfüggvénye

Az $m=5$ és $\sigma=1,5$ paraméterű normális eloszlás eloszlásfüggvénye a 7.13. ábrán látható.



7.13. ábra: Az m=5 és σ=1,5 paraméterű normális eloszlás eloszlásfüggvénye

Várható érték

Tétel. Az m és σ paraméterű normális eloszlás várható értéke $m,$ vagyis

$$M(\xi)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} xe^{-\frac{(x-m)^2}{2\sigma^2}}\,dx=m\,. \tag{7.7.4.}$$

Bizonyítás

Első lépésként ismét új változót vezetünk be, azaz

$$u = \frac{x-m}{\sigma}; \quad du = \frac{dx}{\sigma},$$

amivel az integrál egyszerűbbé válik:

$$M(\xi) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u e^{-\frac{u^2}{2}} du + \frac{m}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du = m, \qquad (7.7.5.)$$

hiszen (7.7.5)-ben az első integrál nulla, mert antiszimmetrikus függvény integráljáról van szó az egész valós tartományon. A második integrál értéke *m,* felhasználva a (7.7.3) azonosságot. Ezzel beláttuk, hogy igaz a várható értékre vonatkozó tétel.

Szórásnégyzet

Tétel

$$D^2(\xi) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \sigma^2. \qquad (7.7.6.)$$

Bizonyítás

Ismét bevezetjük a fentiekben már megismert új változót:

$$u = \frac{x-m}{\sigma}; \quad du = \frac{dx}{\sigma}.$$

Ezzel kifejezve a (7.7.6) kifejezés integrálját:

$$D^2(\xi) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u^2 e^{-\frac{u^2}{2}} du.$$

Most kihasználva azt, hogy

$$\left(-e^{-\frac{u^2}{2}}\right)' = \left(u e^{-\frac{u^2}{2}}\right),$$

majd parciálisan integrálva a kapott kifejezést, éppen a (7.7.6) kifejezésre jutunk.

Egyszerű függvényanalízissel belátható, hogy az $N(m,\sigma)$ normális eloszlás sűrűségfüggvényének maximuma m-nél van, a függvény inflexiós pontjai pedig az $m-\sigma$ és az $m+\sigma$ pontokban vannak. Az $F(x)=1/2$ egyenlet megoldása is $x=m$. Az m tehát nemcsak az eloszlás várható értéke, hanem szimmetriatengelye, módusza és mediánja is.

7.8. A standard normális eloszlás

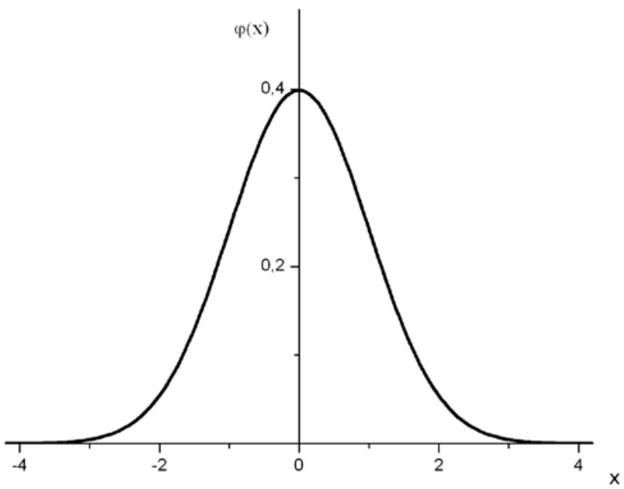
A statisztikában kitüntetett szerepe van annak a normális eloszlásnak, amelynek paraméterei: $m=0$ és $\sigma=1$. Az ilyen normális eloszlást standard normális eloszlásnak nevezzük és szokásos jelölése $N(0,1)$. A kitüntetett szerepnek megfelelően a sűrűségfüggvény jelölése $\varphi(x)$, az eloszlásfüggvényé pedig $\Phi(x)$.

A standard normális eloszlás sűrűségfüggvénye

A paraméterek behelyettesítésével azt kapjuk, hogy a standard normális eloszlás sűrűségfüggvénye:

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, \quad \text{ahol } -\infty < x < \infty . \tag{7.8.1.}$$

A sűrűségfüggvény alakja a 7.14 ábrán látható.



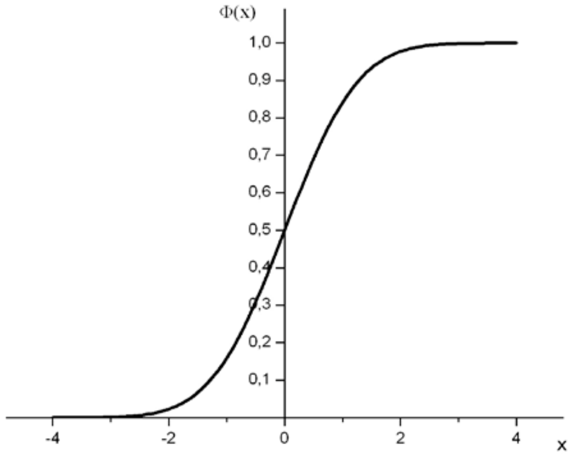
7.14. ábra: A standard normális eloszlás sűrűségfüggvénye

A standard normális eloszlás eloszlásfüggvénye

Az eloszlásfüggvény kifejezését megkapjuk az $m=0$ és a $\sigma=1$ paraméterek beírásával a (7.7.2) kifejezésbe:

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \, . \tag{7.8.2.}$$

Az eloszlásfüggvény alakját a 7.15 ábrára rajzoltuk.



7.15. ábra: A standard normális eloszlás eloszlásfüggvénye

A standard normális eloszlás tulajdonságai

A standard normális eloszlás várható értéke nulla, és sűrűségfüggvénye szimmetrikus függvény, azaz

$$\varphi(-x) = \varphi(x) \, , \tag{7.8.3.}$$

amiről behelyettesítéssel könnyen meggyőződhetünk.

Az eloszlásfüggvényre igaz az alábbi azonosság:

$$\Phi(-x) = 1 - \Phi(x) \, , \tag{7.8.4.}$$

ami szintén a sűrűségfüggvény szimmetrikus voltából következik, hiszen

$$\int_{-\infty}^{\infty} \varphi(t)dt = 1 = \int_{-\infty}^{-x} \varphi(t)dt + \int_{-x}^{x} \varphi(t)dt + \int_{x}^{\infty} \varphi(t)dt = 2\int_{-\infty}^{-x} \varphi(t)dt + \int_{-x}^{x} \varphi(t)dt \, .$$

Az integrálok kifejtésével azt kapjuk, hogy

$$1 = 2\Phi(-x) + \Phi(x) - \Phi(-x) \, ,$$

ahonnan átrendezéssel adódik a (7.8.4) azonosság.

A $\varphi(x)$ és $\Phi(x)$ függvények értékeit táblázatban szokták megadni. A szimmetriatulajdonság okán elegendő $x \geq 0$ értékekre megadni a táblázatot.

A standard normális eloszlásra igaz az alábbi tétel.

Tétel

Ha a ξ valószínűségi változó $N(0,1)$ eloszlású, akkor igaz az, hogy

$$P(-x \leq \xi \leq x) = 2\Phi(x) - 1. \tag{7.8.5.}$$

Bizonyítás

A tétel bizonyítása a (7.8.4) azonosság segítségével egyszerű, hiszen

$$P(-x \leq \xi \leq x) = \int_{-x}^{x} \varphi(t)dt = \Phi(x) - \Phi(-x) = \Phi(x) - (1 - \Phi(x)) = 2\Phi(x) - 1.$$

A normális eloszlás további tulajdonságai

Tétel. A normális eloszlás sűrűségfüggvénye kifejezhető a standard normális eloszlás sűrűségfüggvényével, az alábbiak szerint:

$$f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-m}{\sigma}\right). \tag{7.8.6.}$$

Bizonyítás

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}},$$

ahonnan $u = \dfrac{x-m}{\sigma}$ helyettesítéssel azt kapjuk, hogy

$$\varphi\left(\frac{x-m}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Innen látszik, hogy σ-val osztva éppen a normális eloszlás sűrűségfüggvényét kapjuk.

Hasonló módon a normális eloszlás eloszlásfüggvénye is kifejezhető a standard normális eloszlás eloszlásfüggvényével az alábbi tétel szerint.

Tétel

$$F(x)=\Phi\left(\frac{x-m}{\sigma}\right). \tag{7.8.7.}$$

Bizonyítás

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}}dt\,.$$

Innen $u=\dfrac{t-m}{\sigma}$ és $\sigma du=dt$ helyettesítéssel azt kapjuk, hogy

$$F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{u^2}{2}}du=\Phi\left(\frac{x-m}{\sigma}\right)-\Phi(-\infty)=\Phi\left(\frac{x-m}{\sigma}\right),$$

hiszen korábban már beláttuk, hogy $\Phi(-\infty)=0$.

A fenti két tételnek gyakorlati szempontból az a jelentősége, hogy elegendő a standard normális eloszlás táblázatokat megadni, amiből tetszőleges paraméterű normális eloszlás értékei már kiszámíthatók. A Függelék 14.4. fejezetében megtalálható a standard normális eloszlás eloszlásfüggvényének táblázta.

Egy további fontos tételt vezetünk le az $N(m,\sigma)$ eloszlásra vonatkozóan.

Tétel

Ha a ξ valószínűségi változó eloszlása $N(m,\sigma)$, akkor a várható értékre szimmetrikus $(m-\lambda\sigma,\ m+\lambda\sigma)$ tartományba esés valószínűsége a következőképpen számolható:

$$P(m-\lambda\sigma\le\xi\le m+\lambda\sigma)=2\Phi(\lambda)-1. \tag{7.8.8.}$$

Bizonyítás

$$P(m-\lambda\sigma\le\xi\le m+\lambda\sigma)=F(m+\lambda\sigma)-F(m-\lambda\sigma)=\Phi\left(\frac{m+\lambda\sigma-m}{\sigma}\right)-\Phi\left(\frac{m-\lambda\sigma-m}{\sigma}\right)=$$
$$=\Phi(\lambda)-\Phi(-\lambda)=2\Phi(\lambda)-1,$$

ahol kihasználtuk a (7.8.7) és a (7.8.4) azonosságokat.

A normális eloszlásnak a statisztikában központi szerepe van, amit a későbbiekben még tárgyalni fogunk.

Gyors ellenőrző feladatok

7.7. A standard normális eloszlás eloszlásfüggvényének táblázata segítségével számoljuk ki, hogy mekkora annak a valószínűsége, hogy az $N(m,\sigma)$ eloszlású ξ változó a várható érték körüli $(m-\lambda\sigma,\ m+\lambda\sigma)$ tartományba esik $\lambda=1$, $\lambda=2$ és $\lambda=3$ esetekben? (*0,6827, 0,9545, 0,9973*).

7.8. λ milyen értéke mellett esik a $N(m,\sigma)$ eloszlású ξ változó a várható érték körüli $(m-\lambda\sigma,\ m+\lambda\sigma)$ tartományba $p=0,99$ valószínűséggel? ($\lambda=2,575$)

7.9. Bizonyítsuk be, hogy ha ξ eloszlása $N(m,\sigma)$, akkor az

$$\eta = \frac{x-m}{\sigma}$$

új változó eloszlása $N(0,1)$. A η változót ezért standardizált változónak nevezzük. A bizonyítás során használjuk fel a (4.8.1) és (4.8.2) összefüggéseket.

A továbbiakban normális eloszlásból származtatott eloszlásokat vizsgálunk.

7.9. Független normális eloszlások össze

Tétel. Ha ξ és η valószínűségi változók független valószínűségi változók, ξ eloszlása $N(m_1,\sigma)$, η eloszlása $N(m_2,\sigma)$, akkor a $\zeta=\xi+\eta$ valószínűségi változó eloszlása $N(m_1+m_2,\ \sqrt{\sigma_1^2+\sigma_2^2}\,)$. Az összeg valószínűségi változó sűrűségfüggvénye tehát:

$$f(x) = \frac{1}{(\sigma_1^2+\sigma_2^2)\sqrt{2\pi}} e^{-\frac{(x-m_1-m_2)^2}{2(\sigma_1^2+\sigma_2^2)}} . \qquad (7.9.1.)$$

Bizonyítás. Nem bizonyítjuk azt az állítást, hogy az eloszlás normális eloszlású. Viszont, ha már tudjuk, hogy az eloszlás normális, akkor az általános tételek alapján könnyen kiszámolhatjuk a várható értéket és a szórást. Tudjuk ugyanis, hogy a valószínűségi változók összegének várható értéke:

$$M(\zeta) = M(\xi) + M(\eta) = m_1 + m_2 ,$$

és azt is bebizonyítottuk, hogy a független valószínűségi változók összegének szórásnégyzete:

$$D^2(\zeta) = D^2(\xi) + D^2(\eta) = \sigma_1^2 + \sigma_2^2 .$$

Innen már következik a várható értékre és a szórásra vonatkozó állítás.

7.10. Logaritmikus normális eloszlás

Definíció. A ξ valószínűségi változót logaritmikus normális (röviden lognormális) eloszlásúnak nevezzük, ha a logaritmusa normális eloszlású, vagyis ha a belőle képezett

$$\eta = ln\,\xi$$

új változó normális eloszlású.

A definíció alapján kiszámolhatjuk a lognormális eloszlás eloszlásfüggvényét. Az η változó eloszlásfüggvénye normális, tehát

$$G(y) = P(\eta < y) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{(t-m)^2}{2\sigma^2}}\, dt\,.$$

Felírható viszont a következő összefüggés:

$$F(x) = P(\xi < x) = P(ln\,\xi < ln\,x) = P(\eta < y),$$

tehát

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{ln\,x} e^{-\frac{(t-m)^2}{2\sigma^2}}\, dt\,. \qquad (7.10.1.)$$

Megkaptuk tehát a lognormális eloszlás eloszlásfüggvényét. Innen deriválással jutunk a sűrűségfüggvényhez:

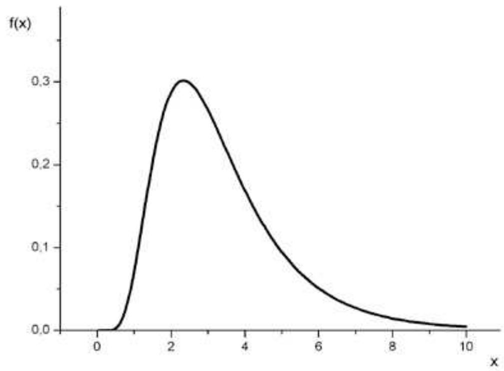
$$f(x) = F'(x) = \frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{(ln\,x-m)^2}{2\sigma^2}} \quad (x \ge 0) \qquad (7.10.2.)$$

A lognormális eloszlás várható értéke és szórásnégyzete

$$M(\xi) = \frac{1}{\sigma\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{(ln\,x-m)^2}{2\sigma^2}}\, dx = e^{m+\frac{\sigma^2}{2}}\,. \qquad (7.10.3.)$$

$$D^2(\xi) = M(\xi^2) - M^2(\xi) = \frac{1}{\sigma\sqrt{2\pi}} \int_{0}^{\infty} x e^{-\frac{(\ln x-m)^2}{2\sigma^2}}\, dx - e^{2m+\sigma^2} = e^{2m+\sigma^2}(e^{\sigma^2}-1). \qquad (7.10.4.)$$

A lognormális eloszlás alakja



7.16. ábra: A lognormális eloszlás sűrűségfüggvénye M(ξ)=3 és D(ξ)=0,5 paraméterekkel

A lognormális eloszlás véletlenszerű törési, osztódási, őrlési folyamatok esetén a végtermék valamelyik méretének, hosszának, térfogatának, súlyának eloszlása. Példaként a golyósmalomban őrölt anyag mérete, lebomló molekulák nagysága, a kőtöréskor keletkező törmelék mérete stb. sok esetben lognormális eloszlást követ.

Tétel. Ha a ξ valószínűségi változó lognormális eloszlású, akkor a belőle képezett $\eta = b\xi^a$ változó is lognormális eloszlású.

Bizonyítás. A lognormális eloszlás definíciója alapján az $ln\xi$ változó normális eloszlású. Ha η logaritmusát képezzük, akkor azt kapjuk, hogy

$$ln\eta = a\,ln\xi + ln b,$$

márpedig a normális eloszlás lineáris transzformáltja is normális eloszlás, ami azt jelenti, hogy az η valószínűségi változó is lognormális eloszlású.

A tételből következik, hogy ha egy termék lineáris mérete lognormális eloszlást követ, akkor a termék felszíne, térfogata, súlya is lognormális eloszlású, hiszen ezek a jellemzők a lineáris méret hatványfüggvényei.

Gyors ellenőrző feladat

7.10. Ha a ξ valószínűségi változó $N(m,\sigma)$ eloszlású akkor lássuk be, hogy az $\eta = a\xi + b$ új változó szintén normális eloszlású $m' = am + b$, $\sigma' = a\sigma$ paraméterekkel! Használjuk a (4.8.2) összefüggést!

8. SZÁRMAZTATOTT ELOSZLÁSOK

A statisztika gyakran használ olyan valószínűségi változókat, amelyeket a standard normális eloszlásból származtatunk. Ezeket az eloszlásokat nem tárgyaljuk olyan részletességgel, mint a korábbiakat. Megelégszünk a valószínűségi változó definíciójával, és a sűrűségfüggvény menetével. Az eloszlások sűrűségfüggvénye vagy eloszlásfüggvénye általában táblázatok formájában áll rendelkezésre, illetve a statisztikai számítógépes programok is szolgáltatják ezeknek a függvényeknek az értékeit.

8.1. A χ^2-eloszlás

Definíció. Legyen $\xi_1, \xi_2, ..., \xi_n$ *n* darab független, *N(0,1)* eloszlású valószínűségi változó. Ezekből képezzük a χ^2 valószínűségi változót, a

$$\chi^2 = \xi_1^2 + \xi_2^2 + \ldots + \xi_n^2 \qquad (8.1.1.)$$

definícióval. A χ^2-eloszlását *n* szabadsági fokú **χ^2-eloszlásnak** nevezzük. Az eloszlás sűrűségfüggvénye a 8.1. ábrán látható *n=5* és *n=15* értékek mellett. A χ^2 valószínűségi változónak csak $x \geq 0$ értékei lehetnek, ezért a sűrűségfüggvény értelmezési tartománya csak a pozitív valós számok halmaza, a *0*-t is beleértve.

A χ^2-eloszlás várható értéke

Mivel nem adjuk meg a sűrűségfüggvény alakját, ezért várható értékét is csak közöljük:

$$M\left(\chi^2\right) = n\,.$$

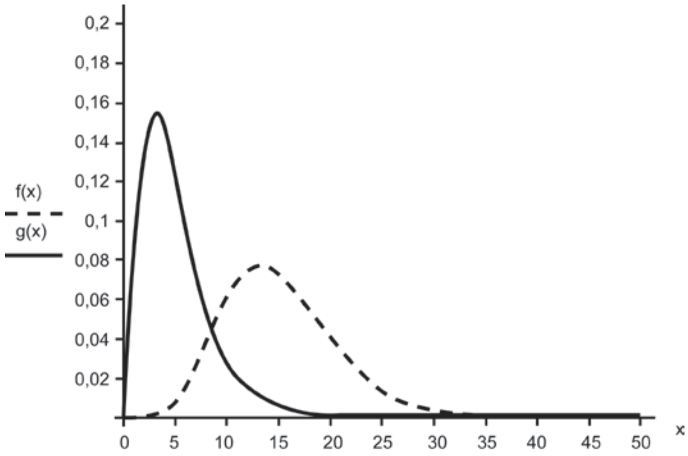
A 8.1. ábrán látszik, hogy *n* növekedésével a görbe (és vele a várható érték is) egyre inkább jobbra tolódik. Mivel a függvény nem szimmetrikus, ezért a várható érték és a módusz nem esik egybe.

A χ2-eloszlás szórásnégyzete

A χ^2-eloszlás szórásnégyzete:

$$D^2(\chi^2) = 2n\,.$$

A sűrűségfüggvény ábráján látszik, hogy növekvő *n*-el a függvény egyre szélesedik. A későbbiekben belátjuk majd, hogy ahogy *n* növekszik, a χ^2-eloszlás sűrűségfüggvénye egyre inkább közelít a normális eloszlás sűrűségfüggvényéhez.



8.1. ábra: A χ^2-eloszlás sűrűségfüggvénye n=5 (g(x)) és n=15 (f(x)) esetén

8.2. A χ-eloszlás

A χ^2-eloszlásból származtatjuk a χ-eloszlást.

Definíció. A χ^2 valószínűségi változóból létrehozott

$$\chi = \sqrt{\xi_1^2 + \xi_2^2 + \ldots + \xi_n^2} \tag{8.2.1.}$$

új valószínűségi változó eloszlását n szabadsági fokú χ-eloszlásnak nevezzük. A χ-eloszlás sűrűségfüggvényének menete hasonló, mint a χ^2-eloszlásé.

Azt gondolhatnánk, hogy a χ-eloszlás konstrukciója annyira mesterkélt, hogy távol áll a gyakorlati alkalmazásoktól. Ez nem így van. Példaként megemlítjük, hogy az ideális gáz sebességeloszlását leíró Maxwell-eloszlás $n=3$ szabadsági fokú χ-eloszlás. Ebben az esetben a ξ_1, ξ_2 és ξ_3 valószínűségi változók a sebességkomponensekkel kapcsolatosak

$$\xi_1 = \frac{v_x}{v_{\max}}; \quad \xi_2 = \frac{v_y}{v_{\max}}; \quad \xi_3 = \frac{v_z}{v_{\max}},$$

és $N(0,1)$ eloszlásúak. Ha a sebességet v-vel jelöljük, akkor

$$v = \sqrt{v_x^2 + v_y^2 + v_z^2},$$

v_{max} pedig az a sebesség, ahol az eloszlásnak maximuma van. Ha bevezetjük a

$$\frac{v}{v_{\max}} = x$$

jelölést, akkor a Maxwell-eloszlás sűrűségfüggvénye:

$$f(x) = \frac{dn}{n} = \frac{4}{\sqrt{\pi}} x^2 e^{-x^2},$$

amely megadja, hogy az összes n gázatomból dn-nek van a dv tartományba eső v sebessége. Megjegyzendő, hogy ugyanilyen eloszlása van egy hőtartállyal (moderátorral) egyensúlyban lévő neutronoknak, vagy szilárd testben az annihiláció előtt termikus egyensúlyba jutó pozitronoknak.

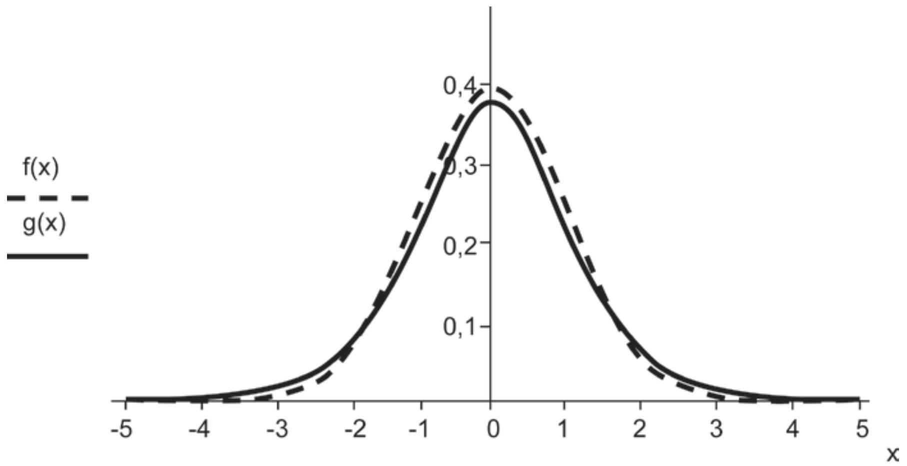
8.3. A Student-eloszlás

Definíció. Legyenek $\eta, \xi_1, \xi_2, \ldots, \xi_n$ független, $N(0,1)$ eloszlású valószínűségi változók. Ezekből képezünk egy új valószínűségi változót az alábbi képlet szerint:

$$t = \frac{\sqrt{n}\eta}{\sqrt{\xi_1^2 + \xi_2^2 + \ldots + \xi_n^2}}. \tag{8.3.1.}$$

A t valószínűségi változó eloszlását n szabadsági fokú **Student-eloszlásnak** nevezzük. Az eloszlást néha t-eloszlásnak is nevezik.

A Student-eloszlás sűrűségfüggvénye a 8.2. ábrán látható.



8.2. ábra: Az N(0,1) eloszlás (f(x)) és az n=5 szabadsági fokú Student-eloszlás (g(x))sűrűségfüggvénye

A Student-eloszlás a 0 pontra szimmetrikus eloszlás. Értelmezési tartománya az egész valós számegyenes. Az Student-eloszlás sűrűségfüggvénye egyre nagyobb n értékekre tart a standard normális eloszlás sűrűségfüggvényéhez. A 8.2. ábrán a Student-eloszlás és a standard normális eloszlás sűrűségfüggvényeinek összehasonlítása is látszik.

A Student-eloszlás várható értéke $M(t)=0$, de csak $n \geq 2$ esetén létezik. Belátható, hogy $n=1$ esetén nem létezik a várható érték.

A Student-eloszlás szórásnégyzete

$$D^2(t) = M(t^2) - M^2(t) = M(t^2) = \frac{n}{n-2}, \text{ ha } n \geq 3.$$

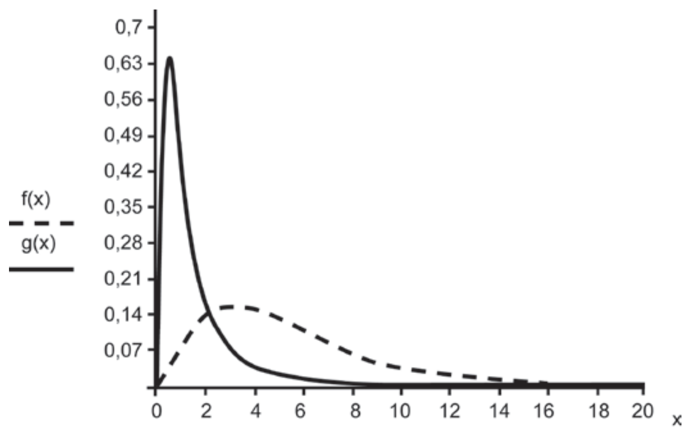
8.4. Az *F*-eloszlás

Definíció. Legyenek a $\xi_1, \xi_2, \ldots, \xi_m$ és az $\eta_1, \eta_2, \ldots, \eta_n$ valószínűségi változók függetlenek és $N(0,1)$ eloszlásúak. Az ezekből képezett F új valószínűségi változót az alábbi összefüggés definiálja:

$$F = \frac{\frac{1}{m}\sum_{i=1}^{m}\xi_i^2}{\frac{1}{n}\sum_{i=1}^{n}\eta_i^2}. \tag{8.4.1.}$$

Az F valószínűségi változó eloszlását m, n szabadsági fokú ***F*-eloszlásnak** nevezzük.

Az $n=5$, $m=5$ szabadsági fokú F-eloszlás sűrűségfüggvényének alakja a 8.3. ábrán látható. Az ábra az összehasonlítás kedvéért az $n=5$ szabadsági fokú χ^2-eloszlás sűrűségfüggvényét is mutatja. A függvény értelmezési tartománya az $x \geq 0$ valós számok köre.



8.3. ábra: Az $n=5$, $m=5$ szabadsági fokú F-eloszlás ($g(x)$) és az $n=5$ szabadsági fokú χ^2-eloszlás ($f(x)$) sűrűségfüggvénye

9. A NAGY SZÁMOK TÖRVÉNYEI

Ez a fejezet abban különbözik a többitől, hogy tételeket mondunk ki, de a bizonyítást mellőzzük. A tételek kijelentéseit azonban részletesen diszkutáljuk, mert mind elvileg, mind gyakorlatilag igen nagy jelentőségűek. A fejezet címe azért, a *Nagy számok törvényei,* mert olyan tételekről lesz szó, amelyekben nagyszámú valószínűségi változó szerepel.

9.1. A nagy számok törvénye (Bernoulli-törvény)

A véletlen jelenségek kapcsán már megismerkedtünk a kísérleti nagy számok törvényével. A kísérleti tapasztalatokra épülő állítás szerint elegendően nagyszámú kísérlet esetén a relatív gyakoriság stabilitást mutat, és egy állandó érték körül egyre kisebb ingadozásokat mutat. A nagy számok most ismertetendő törvénye azt mondja ki, hogy a valószínűség az az érték, amihez a relatív gyakoriság tart.

Tétel. Ha egy esemény valószínűsége *p,* és ε tetszőlegesen kicsiny szám, valamint n számú független kísérletet végezve a kísérletek során az esemény gyakorisága k_n, akkor igaz az, hogy

$$\lim_{n\to\infty} P\left(\frac{k_n}{n} - p\right) \geq \varepsilon = 0\,. \qquad (9.1.1.)$$

Szavakban megfogalmazva a tétel állítását, a kísérletek n számának növekedésével, egyre kisebb annak a valószínűsége, hogy a relatív gyakoriság és a valószínűség különbsége nagyobb legyen mint ε. Más szóval, a relatív gyakoriság nagy valószínűséggel a valószínűséghez tart.

Ez a tétel a nagy számok törvényének precíz megfogalmazása. Szokás ezt röviden úgy írni, hogy

$$\frac{k_n}{n} \Rightarrow p\,.$$

Ugyanakkor ez valószínűségi tétel, vagyis az állítás az, hogy n növekedtével a relatív gyakoriság *nagy valószínűséggel* tart a valószínűség értékéhez, és nem az, hogy a relatív gyakoriság tart a valószínűség értékéhez. Fontos azt látni, hogy a (9.1.1) kifejezésből nem következik az, hogy

$$\lim_{n\to\infty}\left(\frac{k_n}{n} - p\right) \geq \varepsilon = 0\,,$$

hiszen a tétel csak a zárójelben lévő kifejezés valószínűségére mond ki állítást. A valószínűség-számítás segítségével, csak valószínűségi tételekre juthatunk!

9.2. A számtani középről szóló nagy számok törvénye

Tétel. Ha $\xi_1, \xi_2, \ldots, \xi_n$ független valószínűségi változók, amelyeknek várható értéke és szórása azonos, azaz $M(\xi_i)=m$, $i=1, 2, \ldots, n,$ akkor

$$\frac{\xi_1+\xi_2+\ldots+\xi_n}{n} \Rightarrow m\,, \tag{9.2.1.}$$

vagyis az említett feltételek mellett a valószínűségi változók számtani közepe sztochasztikusan tart a közös várható értékhez.

Az 5.10. fejezetben szereplő példában beláttuk, hogy ha $M(\xi_i)=m$, $i=1, 2, \ldots, n$, akkor

$$M\left(\frac{\xi_1+\xi_2+\ldots+\xi_n}{n}\right)=m\,.$$

A (9.2.1) állítás azonban többet mond ennél. A valószínűségi változók számtani közepének nemcsak a várható értéke egyezik meg a közös várható értékkel, hanem az átlag nagy valószínűséggel tetszőleges közelségébe is jut a várható értéknek. Ez valóban többet mondó állítás, hiszen korábban a várható érték kapcsán már láttuk, hogy a várható érték képzésben résztvevő elemek akár távol is lehetnek a várható értéktől. Ezzel szemben most azt látjuk, hogy a valószínűségi változók számtani közepe, ha n elegendően nagy, akkor nagy valószínűséggel a várható érték tetszőleges közelségébe jut.

9.3. A központi határeloszlás tétel

Az alábbiakban kimondandó tétel kiemelkedően fontos szerepet játszik a valószínűségelméletben, ezért hívják központi határeloszlás tételnek (sokszor *centrális határeloszlás tétel* néven szerepel).

Tétel. Ha $\xi_1, \xi_2, \ldots, \xi_n$ azonos eloszlású független valószínűségi változók, amelyeknek várható értéke és szórása azonos, azaz $M(\xi_i)=m$ és $D(\xi_i)=\sigma$ ($i=1, 2, \ldots, n$), akkor

$$\lim_{n\to\infty} P\left(\frac{\xi_1+\xi_2+\ldots+\xi_n-nm}{\sigma\sqrt{n}}<x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}\,du=\Phi(x)\,. \tag{9.3.1.}$$

Bontsuk ki részletesebben a tétel állítását! A tétel bármilyen eloszlású valószínűségi változókból képezett új η_n^* valószínűségi változók sorozatáról tesz rendkívül fontos állítást, ahol

$$\eta_n^* = \frac{\xi_1 + \xi_2 + \ldots + \xi_n - nm}{\sigma\sqrt{n}}, \quad n=1,\ 2,\ \ldots \tag{9.3.2.}$$

Az állítás az, hogy az η_n^* valószínűségi változók sorozatának $F(\eta_n^*)$ eloszlásfüggvénye sztochasztikusan tart az $N(0,1)$ standard normális eloszlás $\Phi(x)$ eloszlásfüggvényhez, azaz

$$F(\eta_n^*) \Rightarrow \Phi(x).$$

Mivel a fenti tétel következményei a tudományokban nagy jelentőségűek, vizsgáljuk meg kissé részletesebben a tétel állítását.

A tételt ugyan nem bizonyítjuk, de azt könnyen beláthatjuk, hogy $M(\eta_n)=0$ és $D(\eta_n)=1$, hiszen

$$M(\eta_n^*) = \frac{M(\xi_1) + M(\xi_2) + \ldots + M(\xi_n) - nm}{\sigma\sqrt{n}} = \frac{nm - nm}{\sigma\sqrt{n}} = 0,$$

és

$$D^2(\eta_n^*) = D^2\left(\frac{\xi_1 + \xi_2 + \ldots + \xi_n - nm}{\sigma\sqrt{n}}\right) = \frac{1}{n\sigma^2}\left(D^2(\xi_1) + D^2(\xi_2) + \ldots + D^2(\xi_n)\right) = \frac{n\sigma^2}{n\sigma^2} = 1.$$

Ezzel legalább azt beláttuk, hogy teljesülnek az $N(0,1)$ eloszlás várható értékére és szórására vonatkozó értékek.

Az η_n^* változót az $\eta_n = \xi_1 + \xi_2 + \ldots + \xi_n$ összeg standardizáltjának nevezzük. Általában is igaz az, hogy a ξ valószínűségi változóból képezett

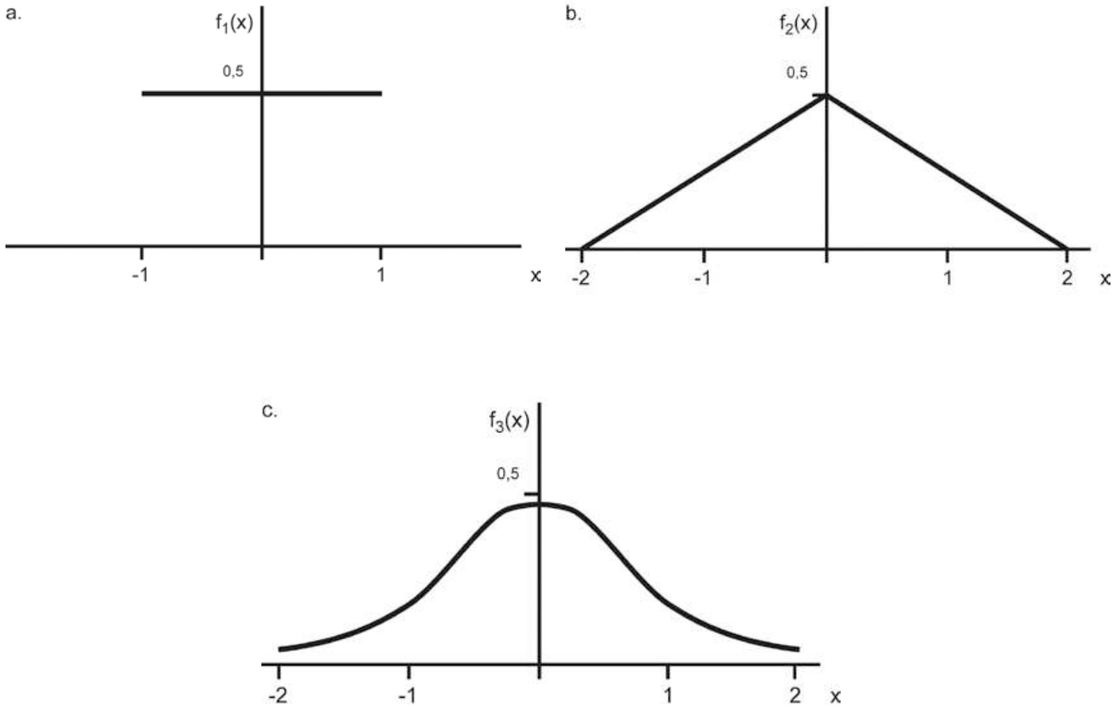
$$\xi^* = \frac{\xi - M(\xi)}{D(\xi)} \tag{9.3.3.}$$

új változó a ξ standardizáltja, amelyre igaz az, hogy $M(\xi^*) = 0$, és $D(\xi^*) = 1$.

Visszatérünk a központi határeloszlás tétel első ránézésre meglepő állítására, hogy tudniillik tetszőleges eloszlású ξ_i valószínűségi változók összegéből képezett η_n^* valószínűségi változó eloszlása standard normális eloszlás. Ennek az állításnak a következménye, hogy az

$$\eta_n = \xi_1 + \xi_2 + \ldots + \xi_n = \sigma\sqrt{n}\eta_n^* + nm$$

változó viszont közelítőleg $N(nm, \sigma\sqrt{n})$ normális (Gauss-) eloszlású (lásd a 7.10. gyors ellenőrző feladatot!). A gyakorlat azt mutatja, hogy ez az állítás már viszonylag kis n érték mellett is jó közelítéssel teljesül. Ezt mutatja a 9.1. ábra, ahol a $(-1,1)$ intervallumon egyenletes eloszlású ξ_1 valószínűségi változó f_1 sűrűségfüggvényét, illetve ugyanilyen eloszlású ξ_2 és ξ_3 változókkal képezett f_2 és f_3 sűrűségfüggvények látszanak, amelyek a $\xi_1+\xi_2$ és a $\xi_1+\xi_2+\xi_3$ változók sűrűségfüggvényei.



9.1. ábra: Az ábrasorozat egyenletes eloszlású valószínűségi változók összegének eloszlását mutatja[*]

Látszik tehát, hogy a központi határeloszlás tétel rendkívül fontos állítása az, hogy az azonos eloszlású valószínűségi változók összege közelítőleg normális eloszlású. A gyakorlatban ezért találkozunk olyan gyakran a normális (Gauss-) eloszlással, mert sok esetben fordul elő, hogy egy jelenség több folyamat összegének eredményeképpen alakul ki. Tipikus példája ennek a mérések statisztikus hibája. Már korábban volt szó arról, hogy a statisztikus hiba sok jelenség együttes hatásából (összegződése révén) okozza a mérések során a mérési adatok szórását. Michelson fénysebesség mérése során láttuk is (lásd 1.9. ábra), hogy *100* mérési adatból felrajzolt hisztogram közelíthető a Gauss-eloszlás harang alakú görbéjével.

A központi határeloszlás tételéből következik az is, hogy a binomiális eloszlás (amely a karakterisztikus valószínűségi változók összege) közelíthető a normális eloszlással. Hasonló módon a χ^2-eloszlás definíció szerint valószínűségi változók összege, tehát elég nagy szabadsági fok esetén a χ^2-eloszlás is közelíthető a normális eloszlással.

Vannak olyan folyamatok, ahol sok véletlen hatás érvényesül, de ezek a hatások nem összegződnek, hanem szorzódnak. Az így létrejövő valószínűségi változó logaritmusára

[*] Az itt letölthető szimuláció azt mutatja, hogy az egyre nagyobb számú egyenletes eloszlású valószínűségi változók összegének eloszlása hogyan közelít a normális eloszláshoz.

igaz, hogy közel normális eloszlású, hiszen a hatások logaritmusa összegződik. Tehát maga a változó logaritmikus normális eloszlású lesz.

Megjegyzés

A centrális határeloszlás tételének több megfogalmazása létezik. Ezek kissé különböző feltételekkel fogalmazzák meg azt az állítást, hogy a valószínűségi változók összege standardizált változatának eloszlása standard normális eloszlás. Egyes megfogalmazásokban az sem feltétel, hogy az összegben szereplő valószínűségi változók eloszlása azonos legyen.

Példa. A gyakorlatban sokszor felmerül az a kérdés, hogy hány mérést kell végeznünk ahhoz, hogy a mérési adatok átlaga a várható értéket előre megadott kis ε értéknél jobban megközelítse $1-p$ valószínűséggel (ahol p egy kis szám)?

Tekintsük a $\xi_1, \xi_2, \ldots, \xi_n$ mérési adatokat azonos eloszlású független valószínűségi változóknak, amelyek várható értéke $M(\xi_i)=m$, szórásuk pedig $D(\xi_i)=\sigma$ ($i=1, 2, \ldots n$). A 9.3.2. képlet szerint definiált η_n^* valószínűségi változó közelítőleg standard normális eloszlású, tehát (7.8.5) szerint

$$P(|\eta_n^*| \le \lambda) \approx 2\Phi(\lambda) - 1 .$$

Most azt szeretnénk, hogy

$$2\Phi(\lambda) - 1 = 1 - p$$

legyen, ahonnan

$$\Phi(\lambda) = 1 - \frac{p}{2} .$$

A standard normális eloszlás táblázatából a p értékének ismeretében λ visszakereshető. Tehát oda jutottunk, hogy

$$\left| \frac{\xi_1 + \xi_2 + \ldots + \xi_n - nm}{\sigma\sqrt{n}} \right| \le \lambda$$

teljesül $1-p$ valószínűséggel, ahonnan kis átalakítással

$$\left| \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n} - m \right| \le \frac{\lambda\sigma}{\sqrt{n}}$$

is $1-p$ valószínűséggel teljesül. Ha most az egyenlőtlenség jobb oldalára megköveteljük, hogy teljesüljön a

$$\frac{\lambda\sigma}{\sqrt{n}} \le \varepsilon$$

feltétel, akkor n-re azt kapjuk, hogy

$$n \ge \frac{\lambda^2\sigma^2}{\varepsilon^2}.$$

Tehát, ilyen számú mérés esetén a mérési adatok átlaga $1-p$ valószínűséggel a várható érték ε közelségébe kerül.

10. A MATEMATIKAI STATISZTIKA ELEMEI

A matematikai statisztika a valószínűség-számítás önálló fejezete, amely mérések eredményeiből, az ún. statisztikai adatokból következtet a véletlen események valószínűségeire, a valószínűségi változók ismeretlen eloszlásfüggvényeire és ezek paramétereire. A matematikai statisztika fejezetei: a mintavétel elmélete, becsléselmélet, hipotézisvizsgálat, korreláció- és regresszióanalízis, szóráselemzés, kísérletek tervezése, hibaszámítás.

10.1. Statisztikai mintavétel

A vizsgálat tárgyát képező elemek összességét a hozzájuk tartozó számértékekkel együtt **statisztikai sokaságnak** nevezzük. Például golyóscsapágy golyók halmaza, valamint a golyók átmérője. A statisztikai sokaság tartalmazhat véges vagy végtelen sok elemet.

A statisztikai sokaság felfogható valószínűségi mezőnek is. A ξ valószínűségi változó lehetséges értékei a sokaság elemeihez rendelt számértékek. ξ valószínűségi eloszlását a sokaság eloszlásának nevezzük.

A statisztikai vizsgálat célja az, hogy mintavétellel (kísérletek végzésével) a sokaság eloszlására vonatkozóan információt szerezzünk.

A mintavétel a következőket jelenti. A sokaságból *n* számú elemet véletlenszerűen kiválasztunk, és a kiválasztott elemeknek a bennünket érintő jellemzőjét megmérjük (például a golyóscsapágy golyók halmazából kiválasztott *n* számú golyó átmérőjét megmérjük). Legyenek az ezekhez tartozó számértékek a kiválasztás sorrendjében: $x_1, x_2, x_3, \ldots, x_n$. Ez egy *n* elemű minta.

Mivel a kiválasztás véletlenszerű egy következő kiválasztás más eredményt adat. Például $x_1^*, x_2^*, \ldots, x_n^*$. Emiatt $x_1, x_2, x_3, \ldots, x_n$ mintaelemek valószínűségi változónak tekinthetők. Vegyük észre, hogy az x_i jelölésnek két jelentése lehet. Vagy egy *n* elemű minta egy konkrét elemét jelenti, és ilyenkor egy számértéket helyettesít, vagy pedig mint valószínűségi változó az *n* elemű minta egyik elemét jelképezi. Ez a két jelentés nem keverendő, de a továbbiakban is meghagyjuk az azonos jelölést, hiszen a szövegösszefüggésből mindig világosan kiderül, hogy éppen melyik jelentést használjuk.

Választhatunk visszatevéssel vagy visszatevés nélkül. Ha visszatevéssel választunk (mindig ugyanolyan módon), akkor az $x_1, x_2, x_3, \ldots, x_n$ valószínűségi változók függetlenek, azonos eloszlásúak és eloszlásuk megegyezik a ξ valószínűségi változó eloszlásával.

Sok esetben, ha a kiválasztás visszatevés nélküli, akkor is teljesül a függetlenség. Például ha a sokaság elemeinek a száma olyan nagy, hogy kevés számú elem kiválasztása az eloszlást nem befolyásolja.

A statisztikai mintavétellel szemben alapvető követelmény, hogy reprezentatív mintavétel legyen, vagyis hűen tükrözze a sokaságot, amelyből való. Általában reprezentatív a mintavétel, ha a mintaelemek eloszlása azonos és az alapsokaságéval megegyező, továbbá ha az elemek független valószínűségi változók. Ez így kijelentve egyszerűnek tűnik, azonban a gyakorlatban gondosan kell ügyelni arra, hogy a reprezentativitás biztosítva legyen, és a rejtett függőségeket is elkerüljük. Ha például az ország lakosságának magasságeloszlását szeretnénk *300* elemű minta segítségével jellemezni, akkor nem a kosárlabdacsapatok tagjainak magasságát kell mintaelemeknek választani.

Ha a mintaelemeket megfelelően kiválasztottuk, akkor segítségükkel következtetni tudunk a sokaság eloszlására és az eloszlás paramétereire.

10.2. Empirikus eloszlásfüggvény

A sokaság eloszlásfüggvényét például közelíthetjük a mintaelemek segítségével létrehozott empirikus eloszlásfüggvénnyel.

Definíció. Tekintsük az $x_1, x_2, x_3, \ldots, x_n$ *n* elemű mintát. Legyen $F(x)$ a sokaság elméleti eloszlásfüggvénye. Ha az $x_1, x_2, x_3, \ldots, x_n$ pontok mindegyikéhez hozzárendelünk *1/n* valószínűséget, akkor diszkrét valószínűség-eloszlást kapunk. Az ehhez tartozó eloszlásfüggvény $F_n(x)$ empirikus eloszlásfüggvény, amit úgy rajzolunk fel, ahogy a diszkrét eloszlás esetén korábban eljártunk:

$$F_n(x) = \frac{k}{n}, \tag{10.2.1.}$$

ahol *k* azon x_i-k száma, melyekre $x_i < x$.

A definícióból látszik, hogy az $F_n(x)$ empirikus eloszlásfüggvény a $\xi<x$ esemény relatív gyakoriságát adja. Korábban a mérési adatok leíró jellemzése során már láttuk, hogy a kumulatív relatív gyakorisággal jellemezhetjük a mérési adatok eloszlását. A valószínűségelméleti ismeretink birtokában most még többet is állíthatunk. A (10.2.1) kifejezésből következik, hogy

$$nF_n(x) = k,$$

ami a $\xi<x$ esemény gyakoriságát adja.

Korábbi ismereteink alapján a $\xi<x$ esemény elméleti valószínűsége:

$$P(\xi < x) = F(x).$$

Ha most *k*-t mint valószínűségi változót tekintjük, akkor azt is tudjuk, hogy *k* binomiális eloszlású (Bernoulli-eloszlású) valószínűségi változó, melynek paramétere: $p=F(x)$. Innen a Bernoulli-eloszlás várható értékét felhasználva az következik, hogy

$$M(nF_n) = nF(x),$$

vagy n-el elosztva az egyenlet mindkét oldalát

$$M\big(F_n(x)\big) = F(x).$$

A nagy számok Bernoulli-féle törvényéből még az is következik, hogy

$$F_n(x) \Rightarrow F(x),$$

vagyis az empirikus eloszlásfüggvény sztochasztikusan tart az elméleti eloszlásfüggvényhez. Azt kaptuk tehát, hogy az empirikus eloszlásfüggvény olyan jó tulajdonságú statisztikai függvény, amellyel jól közelíthető az elméleti eloszlásfüggvény.

10.3. Empirikus sűrűségfüggvény

Az empirikus eloszlásfüggvényhez hasonlóan definiálhatjuk az empirikus sűrűségfüggvényt is.

Definíció. A valószínűségi sűrűségfüggvény is közelíthető a tapasztalati sűrűségfüggvénnyel, amelyet az ún. sűrűséghisztogrammal ábrázolható. Ismét egy $x_1, x_2, x_3, \ldots, x_n$ n elemű mintából indulunk ki. Osszuk fel azt az intervallumot, amelybe az x_i értékek esnek sok kis Δx hosszúságú szakasz összegére. A j-edik Δx_j szakasz fölé rajzoljunk téglalapot, amelyek magassága legyen:

$$\frac{k_j}{n\Delta x_j} \quad \forall\, j-re, \tag{10.3.1.}$$

ahol n a mintaelemek száma, k_j/n a j. intervallumba eső mintaelemek relatív gyakorisága, amelyet elosztva az intervallum hosszával sűrűség jellegű mennyiséget kapunk. Valamenynyi intervallum fölé emelt téglalapok együttese kirajzolja az $f_n(x)$ sűrűséghisztogramot.

A sűrűséghisztogram tulajdonságait már láttuk az 1.2. alfejezetben. A sűrűséghisztogramról az empirikus eloszlásfüggvényhez hasonlóan belátható, hogy sztochasztikusan tart az elméleti sűrűségfüggvényhez, vagyis $f_n(x) \Rightarrow f(x)$.

10.4. Empirikus várható érték

Az $x_1, x_2, x_3, \ldots, x_n$ n elemű minta segítségével közelíthetjük a sokaság más paramétereit is.

Definíció. Az elméleti várható érték közelítésére használatos az $\bar{x}$ empirikus várható érték, amelynek definíciója:

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}. \qquad (10.4.1.)$$

Ezt a definíciót a mérési adatok leíró jellemzése során már láttuk. Most azonban a valószínűség-számítás megismert módszereivel az empirikus paraméterek tulajdonságait mélyebben is megismerhetjük.

Mivel $x_1, x_2, x_3, \ldots, x_n$ mintaelemek valószínűségi változók, így a belőlük képezett függvények is valószínűségi változók, ezért az empirikus jellemzők valószínűségelméleti módszerekkel kezelhetők (pl. kiszámolható várható értékük, szórásuk, stb.).

Az empirikus várható érték elméleti várható értéke

Tegyük fel, hogy a sokaságnak létezik az elméleti várható értéke, vagyis $\exists\, M(\xi) = m$, amely az elméleti eloszlás várható értéke. Kérdés, hogy az empirikus várható érték várható értéke hogyan viszonyul az elméleti várható értékhez?

Tétel. Az $\overline{x}$ **empirikus várható érték várható értéke** megegyezik az eloszlás m elméleti várható értékével.

Bizonyítás. A várható érték képzésének szabályaival képezzük $\overline{x}$ várható értékét. A bizonyítás során használjuk ki, hogy valamennyi mintaelem eloszlása azonos, és várható értékük azonos, és megegyezik a sokaság elméleti várható értékével, vagyis $M(x_i) = m$ minden i értékre. Tehát

$$M(\overline{x}) = M\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n} M\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} M(x_i) = \frac{1}{n} n \cdot m = m, \qquad (10.4.2.)$$

és ezzel beláttuk a tétel állítását.

Az $\overline{x}$ valószínűségi változónak nem csak a várható értékét, de a szórását is kiszámolhatjuk.

Az empirikus várható érték elméleti szórása

Tegyük fel, hogy a sokaságnak létezik a szórásnégyzete, azaz $\exists\, D^2(\xi) = \sigma^2$. Kérdés, hogy mekkora az empirikus várható érték szórása?

Tétel. Az **empirikus várható érték szórása** a sokaság elméleti szórása osztva $\sqrt{n}$-el, ahol n a mintaelemek száma.

Bizonyítás. A szórásnégyzet képzés szabályait alkalmazzuk, és felhasználjuk, hogy a mintaelem szórásnégyzete azonos és megegyezik a sokaság szórásnégyzetével, vagyis $D^2(x_i)=\sigma^2$, minden i esetén.

$$D^2(\bar{x})=\frac{1}{n^2}D^2\left(\sum_{i=1}^{n}x_i\right)=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n},$$

és innen gyökvonással kapjuk az empirikus várható érték szórását:

$$D(\bar{x})=\frac{\sigma}{\sqrt{n}}. \qquad (10.4.3.)$$

10.5. Empirikus szórásnégyzet

Az empirikus várható értékhez hasonlóan az $x_1, x_2, x_3, \ldots, x_n$ n elemű minta elemeinek segítségével képezhető az empirikus szórásnégyzet is, ahogyan az már korábban láttuk a mérési adatok leíró jellemzése során. Most megismételjük a definíciót, majd a valószínűségelmélet módszereivel az empirikus szórásnégyzet újabb tulajdonságait mutatjuk meg.

Definíció. Az empirikus szórásnégyzet jele s^2, és a definíciója:

$$s^2=\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_n-\bar{x})^2}{n}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}. \qquad (10.5.1.)$$

Mivel a valószínűségi változóknak tekintett mintaelemekből képezett empirikus szórásnégyzet maga is valószínűségi változó, ezért képezhető az elméleti várható értéke.

Az empirikus szórásnégyzet elméleti várható értéke

Tétel. Az empirikus szórásnégyzet várható értéke a sokaság szórásnégyzetének *(n–1)/n*-szerese, vagyis

$$M(s^2)=\frac{n-1}{n}\sigma^2. \qquad (10.5.2.)$$

Bizonyítás. Az empirikus szórásnégyzettel kapcsolatban korábban már beláttuk (1.11.3) az alábbi összefüggést:

$$s^2=\sum_{i=1}^{n}\frac{1}{n}(x_i-\bar{x})^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2.$$

A továbbiakban bevezetjük a $z_i = x_i - m;\ i = 1, 2, \ldots, n$ mennyiséget, melynek átlaga $\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$. Felhasználva z_i definícióját, igaz az alábbi összefüggés is:

$$z_i - \bar{z} = x_i - m - \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{\cancel{n}}\cancel{n}m = x_i - \bar{x}.$$

Innen s^2 kifejezhető a z_i-kel:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2 = \frac{1}{n}\sum_{i=1}^{n} z_i^{\,2} - \bar{z}^2.$$

Innen már adódik a tétel bizonyítása:

$$M(s^2) = \frac{1}{n}M\left(\sum_{i=1}^{n} z_i^{\,2}\right) - M(\bar{z}^2) = \frac{1}{n}n\sigma^2 - \frac{1}{n^2}M\left[\left(\sum_{i=1}^{n} z_i\right)^2\right] =$$

$$= \sigma^2 - \frac{1}{n^2}M\left(\sum_{i=1}^{n} z_i^{\,2} + 2\sum_{i<j} z_i z_j\right) == \sigma^2 - \frac{1}{n^2}n\sigma^2 = \sigma^2\left(1 - \frac{1}{n}\right) = \sigma^2\frac{n-1}{n}.$$

Itt kihasználtuk, hogy z_i, z_j független valószínűségi változók $\forall\ i,j$-re. Ebből következik, hogy a zárójelben lévő kifejezés második tagjának várható értéke zérus, hiszen a zárójelben lévő második tag várható értéke éppen a két változó kovarianciája, amelyről tudjuk, hogy független változók esetén értéke zérus.

A tétel alapján tehát látjuk, hogy az empirikus szórásnégyzet „szépséghibája”, hogy várható értéke nem egyenlő az elméleti szórásnégyzettel. Ezért a gyakorlatban s^2 helyett az

$$s^{*2} = \frac{n}{n-1}s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{10.5.3.}$$

ún. korrigált empirikus szórásnégyzettel dolgozunk. Erre már teljesül, hogy

$$M(s^{*2}) = \sigma^2. \tag{10.5.4.}$$

Látszik, hogy nagy n-re s^{*2} és s^2 eltérése elhanyagolhatóvá válik. Megmutatható az is, hogy $s^{*2} \Rightarrow \sigma^2$, amiből viszont s^{*2} definíciója alapján következik, hogy $s^2 \Rightarrow \sigma^2$, hiszen határértékben s^{*2} és s^2 megegyezik.

Az empirikus várható érték és az empirikus szórásnégyzet definícióihoz hasonlóan az $x_1, x_2, x_3, \ldots, x_n$ n elemű minta elemeinek felhasználásával definiálható az empirikus medián, az empirikus terjedelem, a k. empirikus momentum stb. Például a k. empirikus momentum definíciója:

$$\frac{1}{n}\sum_{i=1}^{n} x_i^{\,k} . \tag{10.5.5.}$$

A *k.* empirikus centrális momentum definíciója pedig az alábbi:

$$\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^k . \tag{10.5.6.}$$

Az empirikus eloszlásfüggvény és az empirikus jellemzők a sokaság eloszlásfüggvényének és jellemző adatainak (elméleti jellemzőinek) közelítésére használatos.

Az alkalmazások során gyakran felmerül az a kérdés, hogy az empirikus jellemzőknek milyen az eloszlásuk. Általánosan nem válaszoljuk meg a kérdést, de a gyakorlati életben gyakori $N(m, \sigma)$ normális eloszlású sokaság esetén megadjuk $\bar{x}$ és s^2 eloszlását.

10.6. $\bar{x}$ és s^2 eloszlása normális eloszlás esetén

Tétel. $N(m,\sigma)$ eloszlású sokaság esetén $\bar{x}$ eloszlása: $N\left(m, \frac{\sigma}{\sqrt{n}}\right)$.

Bizonyítás. Az $\bar{x}$ eloszlását könnyű megtalálni, hiszen $\bar{x}$ független normális eloszlású valószínűségi változók összege, ami, mint korábban láttuk, maga is normális eloszlású. Korábban azt is láttuk, hogy $M(\bar{x}) = m$, $D(\bar{x}) = \frac{\sigma}{\sqrt{n}}$. Tehát $\bar{x}$ eloszlása:

$$N\left(m, \frac{\sigma}{\sqrt{n}}\right). \tag{10.6.1.}$$

Az s^2 eloszlásának megkeresése ennél kevésbé egyszerű, ezért itt bizonyítás nélkül adjuk meg.

Tétel. $N(m,\sigma)$ eloszlású sokaság esetén az $\frac{n}{\sigma^2}s^2$ valószínűségi változó *n–1* szabadsági fokú χ^2 eloszlású.

Megjegyzések:

1. Mivel $\frac{n}{\sigma^2}s^2 = \frac{n-1}{\sigma^2}s^{*2}$, ezért az $\frac{n-1}{\sigma^2}s^{*2}$ is *n–1* szabadsági fokú χ^2 eloszlású valószínűségi változó.
2. Belátható az is, hogy $\bar{x}$ éx s^2 független valószínűségi változók. Mivel s^2 és s^{*2} csak konstansban különböznek, ezért $\bar{x}$ és s^{*2} is függetlenek.

11. A BECSLÉSELMÉLET ELEMEI

A műszaki és tudományos gyakorlatban sokszor előfordul, hogy a vizsgálandó sokaság elméleti eloszlásfüggvényének alakja ismert (tudjuk például, hogy a sokaság normális eloszlású), de nem ismerjük az eloszlás paramétereit (pl. normális eloszlás esetén az m és σ paramétereket). A becsléselmélet tárgya az ismeretlen paraméterek becslése, amelyet a mintavétel során kapott adatok felhasználásával végzünk el. Ha az ismeretlen paramétert számértékkel becsüljük, akkor pontbecslésről beszélünk, ha pedig intervallumot adunk meg, amely a szóban forgó paramétert nagy valószínűséggel tartalmazza, akkor intervallumbecslésről beszélünk.

Az egyszerűség kedvéért tegyük fel, hogy az eloszlásfüggvény csak egy paramétertől függ. Ismert tehát az $F(x,a)$ eloszlásfüggvény, de nem ismerjük az a paraméter értékét.

Mint ahogy az előző fejezetben az empirikus momentumok esetében tettük, az a paraméter becslésére az $x_1, x_2, \ldots, x_n$ mintaelemeket használjuk. A mintaelemekből képzett

$$\hat{a} = \hat{a}(x_1, x_2, \ldots, x_n)$$

függvény szolgálhat az a paraméter becslésére. Általában a mintaelemekből képezett függvényeket statisztikai függvényeknek, vagy röviden statisztikának nevezzük. Mivel a statisztika valószínűségi változók függvénye, ezért maga is valószínűségi változó.

Kérdés azonban, hogy hogyan válasszuk meg a statisztikai függvényt? Az a paraméter jó becsléséhez használt statisztika nem lehet tetszőleges, hanem bizonyos jó tulajdonságokkal kell rendelkeznie. Hogyan juthatunk jó becslésekhez? Az eléggé nyilvánvaló, hogy az $\hat{a}$ statisztikát akkor tekintjük az a paraméter jó becslésének, ha eloszlása minél jobban koncentrálódik az a paraméter valódi értéke körül. A pontosabb fogalmazás érdekében felhasználjuk az alábbi definíciókat.

Definíció. Az a paraméter becslésére használt $\hat{a}$ statisztikát **torzítatlannak** nevezzük, ha várható értéke egyenlő a-val, vagyis

$$M(\hat{a}) = a\,.$$

Példaként, az elméleti várható értéknek az empirikus várható érték torzítatlan becslése. Az elméleti szórásnégyzetnek az empirikus szórásnégyzet torzított becslése, viszont a korrigált empirikus szórásnégyzet már torzítatlan becslés.

Definíció. Két $\hat{a}_1$ és $\hat{a}_2$ statisztika közül azt tekintjük *hatásosabbnak,* amelyik szórása kisebb. Tehát, ha

$$D^2(\hat{a}_1) < D^2(\hat{a}_2),$$

akkor az $\hat{a}_1$ becslés hatásosabb, mint az $\hat{a}_2$ becslés.

Definíció. Az a paraméternek egy $\hat{a}_1, \hat{a}_2, ..., \hat{a}_n$ becsléssorozatát konzisztensnek nevezzük, ha $\hat{a}_n$ sztochasztikusan tart *a*-hoz, vagyis:

$$\hat{a}_n \Rightarrow a .$$

Például az empirikus várható érték konzisztens becslése a várható értéknek, hiszen $\bar{x} \Rightarrow m$. Hasonló módon az empirikus szórásnégyzet, bár nem torzítatlan, de konzisztens becslése a szórásnégyzetnek, hiszen $s^2 \Rightarrow \sigma^2$.

Felvetődik a kérdés, hogyan lehet olyan becsléseket létrehozni, amelyek rendelkeznek a fenti tulajdonságokkal, és azok közül minél többel? Több ilyen módszer létezik, ezek közül az alábbiakban a momentumok módszerével és a maximum likelihood módszerrel ismerkedünk meg.

11.1. A momentumok módszere

Elméleti momentumok

A korábbiakban definiáltuk már a statisztikai sokaság elméleti és empirikus momentumait. Most összefoglaljuk a momentumokkal kapcsolatos eddigi ismereteinket.

A ξ valószínűségi változó *k.* elméleti momentuma:

$$\alpha_k = M(\xi^k). \qquad (11.1.1.)$$

A *k.* elméleti momentumot diszkrét esetben az alábbi képlet alapján számolunk:

$$\alpha_k = \sum_i x_i^k p_i, \qquad (11.1.2.)$$

ahol az x_i értékek a ξ valószínűségi változó lehetséges értékei, a p_i-számok pedig a hozzájuk rendelt valószínűségi értékek.

Folytonos esetben az elméleti *k.* momentum számolása az alábbi:

$$\alpha_k = \int_{-\infty}^{\infty} x^k f(x) dx. \qquad (11.1.3.)$$

A definíciós egyenletekből látszik, hogy az első elméleti momentum a valószínűségi változó várható értéke:

$$\alpha_1 = M(\xi).$$

Az elméleti *k.* centrális momentum definíciós egyenlete:

$$\mu_k = M\left(\xi - M(\xi)\right)^k. \tag{11.1.4.}$$

Diszkrét esetben az elméleti *k.* centrális momentum számolása:

$$\mu_k = \sum_i \left(x_i - M(\xi)\right)^k p_i. \tag{11.1.5.}$$

Folytonos esetben az elméleti *k.* centrális momentum számolása:

$$\mu_k = \int_{-\infty}^{\infty} \left(x - M(\xi)\right)^k f(x)dx. \tag{11.1.6.}$$

A definíciós egyenletből látszik, hogy az elméleti második centrális momentum a valószínűségi változó szórásnégyzete:

$$\mu_2 = D^2(\xi).$$

Gyors ellenőrző feladat

11.1. Mutassuk meg, hogy a μ_1 első centrális momentum értéke nulla!

11.2. Mutassuk meg, hogy a μ_2 második centrális momentum kifejezhető az α_1 és az α_2 momentumokkal!

Empirikus (tapasztalati) momentumok

Az $x_1, x_2, x_3, \ldots, x_n$ n elemű minta elemeinek felhasználásával definiálhatók az empirikus momentumok is. Mint már korábban láttuk a *k.* empirikus momentum definíciója:

$$\alpha_i^* = \frac{1}{n}\sum_{i=1}^{n} x_i^k. \tag{11.1.7.}$$

A *k.* empirikus centrális momentum definíciója pedig az alábbi:

$$\mu_i^* = \frac{1}{n}\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^k. \tag{11.1.8.}$$

Felhívjuk a figyelmet arra, hogy az elméleti momentumok esetében az x_i értékek a valószínűségi változó lehetséges értékeit jelölik, míg az empirikus momentumok esetén az x_i értékek az n elemű minta elemei.

A momentumok módszere

Ezek után rátérhetünk a momentumok módszerének tárgyalására. A módszer lényege az, hogy az elméleti momentumokat a keresett paraméterrel kifejezve egyenlővé tesszük a megfelelő empirikus momentumokkal. Az így kapott egyenletből általában kifejezhető a keresett paraméter a mintaelemek segítségével. A momentumok módszere tehát pontbecslés.

Példa

Határozzuk meg a momentumok módszerével az $f(x)=\lambda e^{-\lambda x}$ sűrűségfüggvénnyel jellemzett exponenciális eloszlás λ paraméterét.

Korábban már láttuk, hogy

$$M(\xi)=\int_0^{\infty} xf(x)dx=\frac{1}{\lambda}.$$

Ez az elméleti első momentum. A λ becslésére használt $\hat{\lambda}$-t tehát úgy keressük meg, hogy a $\hat{\lambda}$-pal kifejezett elméleti momentumot egyenlővé tesszük az empirikus első momentummal:

$$\frac{1}{\hat{\lambda}}=\sum_{i=1}^{n}\frac{1}{n}x_i=\bar{x}.$$

Tehát a $\hat{\lambda}$ paraméter

$$\hat{\lambda}=\frac{1}{\bar{x}}. \tag{11.1.9.}$$

Példa

A momentumok módszerével adjunk becslést a normális eloszlás m és σ^2 paraméterére!

A momentumok módszere alapján a várható értékre úgy adhatunk a becslést, hogy az első elméleti momentumot egyenlővé tesszük az első empirikus momentummal:

$$\alpha_1=\alpha_1^*,$$

és mivel

$$\alpha_1=m \text{ és } \alpha_i^*=\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i, \tag{11.1.10.}$$

ezért m becslésére a mintaelemek átlagát használhatjuk:

$$\hat{m} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \,. \qquad (11.1.11.)$$

A momentumok módszere alapján a σ^2 szórásnégyzetre úgy adhatunk becslést, hogy a második elméleti momentumot is egyenlővé tesszük a második empirikus momentummal. Tehát

$$\alpha_2 = \alpha_2^* \,. \qquad (11.1.12.)$$

Mivel

$$\mu_2 = \alpha_2 - \alpha_1^2 \,,$$

és innen

$$\hat{\sigma}^2 = \alpha_2^* - \alpha_1^{*2} = \sum_{i=1}^{n} x_i^2 p_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = s^2 \,.$$

Azt kaptuk tehát, hogy a momentumok módszere alapján a szórásnégyzetre az empirikus szórásnégyzet ad becslést.

11.2. A maximum likelihood módszer

A maximum likelihood (maximális valószínűség) módszert először diszkrét valószínűség-eloszláson mutatjuk be. Legyen

$$P(\xi = x) = p(x, a) \,,$$

vagyis szorítkozzunk egy a paraméter becslésére. Adott $x_1, x_2, \ldots, x_n$ n elemű minta esetén annak valószínűsége, hogy éppen ez a mintasorozat jön létre, a mintaelemek függetlensége miatt:

$$L(x_1, x_2, \ldots, x_n, a) = p(x_1, a)\, p(x_2, a) \cdots p(x_n, a) \,. \qquad (11.2.1.)$$

Az L függvényt likelihood függvénynek nevezzük. A módszer lényege az, hogy keressük az a paraméternek azt az $\hat{a}$ értékét, amely mellett L-nek maximuma van, vagyis a fent kapott minta megvalósulásának valószínűsége maximális. Látszik, hogy a maximum likelihood módszer is pontbecslést ad.

A maximum megkeresését általában a differenciálszámítás módszerei szerint végezzük el. Ismeretes, hogy a maximum megkereséséhez a függvény a paraméter szerinti deriváltját kell nullával egyenlővé tennünk. Azonban a (11.2.1) függvény szorzótényezőkből áll, a szorzat deriváltja pedig bonyolult kifejezés. Egyszerűbbé válik a feladat, ha nem L, hanem $\ln L$ maximumát keressük. A logaritmusképzés során a szorzat összeggé válik, és az összeg deriválása egyszerűbb kifejezésre vezet. Mivel a logaritmusfüggvény monoton növekvő, ezért a maximumhelyek a logaritmusképzés miatt nem változnak, tehát ugyanott vannak, ahol az L függvény maximumai. Ilyenkor tehát az

$$ln\, L(\hat{a}) = \sum_{i=1}^{n} ln\, p(x_i, \hat{a})$$

összeg maximumát keressük, azaz megoldjuk a

$$\frac{d\, ln\, L(\hat{a})}{d\hat{a}} = 0 \tag{11.2.2.}$$

egyenletet. Természetesen az egyenlet megoldását követően meg kell vizsgálni, hogy a talált szélsőérték helyek maximumok-e, vagyis, hogy teljesül-e a megoldás helyén a

$$\frac{d^2\, ln\, L(\hat{a})}{d^2 \hat{a}} < 0$$

feltétel.

Ha az L függvény több paramétertől függ, akkor a maximumkeresést valamennyi paraméter függvényében kell elvégeznünk, és ilyenkor a (11.2.3) kifejezésben parciális deriváltak szerepelnek.

Példa. A Poisson-eloszlás esetén adjunk becslést a λ paraméterre a maximum likelihood módszerrel. A Poisson-eloszlás esetén a valószínűség értékeket a

$$P(\xi, \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

kifejezés adja. Végezzünk n független kísérletet, melyek során a ξ változó mért értékei: k_1, k_2, ..., k_n. A likelihood függvény tehát:

$$L(k_1, k_2, ..., k_n, \hat{\lambda}) = \frac{(\hat{\lambda})^{k_1}}{k_1!} e^{-\lambda} \cdot \frac{(\hat{\lambda})^{k_2}}{k_2!} e^{-\lambda} \cdot ... \cdot \frac{(\hat{\lambda})^{k_n}}{k_n!} e^{-\lambda}.$$

A likelihood függvény logaritmusa:

$$\ln L(k_1, k_2, ..., k_n, \hat{\lambda}) = \sum_{i=1}^{n} \left(k_i \ln \hat{\lambda} - \ln k_i! - \hat{\lambda} \right).$$

A likelihood függvény logaritmusának deriváltja:

$$\frac{d\ln L(k_1,k_2,\dots,k_n,\hat\lambda)}{d\hat\lambda}=\sum_{i=1}^{n}\left(k_i\frac{1}{\hat\lambda}-1\right)=0,$$

ahonnan

$$\hat\lambda=\frac{\sum_{i=1}^{n}k_i}{n}.$$

Azt kaptuk tehát, hogy a Poisson-eloszlás λ paraméterére a likelihood becslés alapján kapott $\hat\lambda$ a mérési eredmények számtani közepe, amely a várható érték torzítatlan becslése.

A maximum likelihood módszer folytonos esetben

Folytonos eloszlás esetén a likelihood-függvény alakja:

$$L=f(x_1,a)f(x_2,a)\cdots f(x_n,a). \qquad (11.2.3.)$$

Egyébként az a paraméter becslésére használt $\hat a$ megkeresése ugyanúgy megy, mint diszkrét esetben.

Példa. Tekintsük ugyanazt a feladatot, amelyet a momentumok módszerével korábban már megoldottunk. Tehát most a maximum likelihood módszerrel adjunk becslést az exponenciális eloszlás λ paraméterére. A sűrűségfüggvény alakja: $f(x)=\lambda e^{-\lambda x}$. A (11.2.3) kifejezés alapján a likelihood-függvény:

$$L=(\hat\lambda)^n e^{-\hat\lambda x_1}e^{-\hat\lambda x_2}\cdots e^{-\hat\lambda x_n}.$$

Vesszük a likelihood függvény természetes alapú logaritmusát:

$$\ln L=n\ln\hat\lambda-\hat\lambda\sum_{i=1}^{n}x_i .$$

Képezzük ennek a függvényben a deriváltját, és tegyük egyenlővé nullával:

$$\frac{\partial\ln L}{\partial\hat\lambda}=\frac{n}{\hat\lambda}-\sum_{i=1}^{n}x_i=0.$$

Innen kifejezhető a keresett $\hat\lambda$ paraméter:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{x}}.$$

Látjuk, hogy ugyanarra az eredményre jutottunk, mint a momentumok módszerével.

Lássunk egy példát két paraméter becslésére a maximum likelihood módszerrel.

Példa. Legyen a ξ valószínűségi változó $N(m,\sigma)$ eloszlású. n mérés alapján adjunk becslést az m várható értékre és a σ szórásra. Az eloszlás valószínűségi sűrűségfüggvénye

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

A likelihood-függvény tehát

$$L(x_1, x_2, ..., x_n, \hat{m}, \hat{\sigma}) = \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{(x_1-\hat{m})^2}{2\hat{\sigma}^2}} \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{(x_2-\hat{m})^2}{2\hat{\sigma}^2}} \cdot ... \cdot \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{(x_n-\hat{m})^2}{2\hat{\sigma}^2}}.$$

A likelihood-függvény logaritmusa:

$$\ln L(x_1, x_2, ..., x_n, \hat{m}, \hat{\sigma}) = -n \ln \hat{\sigma} - n \ln \sqrt{2\pi} - \sum_{i=1}^{n} \frac{(x_i - \hat{m})^2}{2\hat{\sigma}^2}.$$

A parciális derivált $\hat{m}$ szerint:

$$\frac{\partial \ln L(x_1, x_2, ..., x_n, \hat{m}, \hat{\sigma})}{\partial \hat{m}} = -\sum_{i=1}^{n} \frac{2(x_i - \hat{m})}{2\hat{\sigma}^2} = -\sum_{i=1}^{n} \frac{(x_i - \hat{m})}{\hat{\sigma}^2} = 0,$$

ahonnan $\hat{m}$ kifejezhető:

$$\hat{m} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

A várható értékre tehát az empirikus várható érték ad jó becslést a maximum likelihood módszer szerint.

A parciális derivált $\hat{\sigma}$ szerint:

$$\frac{\partial \ln L(x_1, x_2, ..., x_n, \hat{m}, \hat{\sigma})}{\partial \hat{\sigma}} = \frac{n}{\hat{\sigma}} - \frac{1}{\hat{\sigma}^3}\sum_{i=1}^{n}(x_i - \hat{m})^2 = 0,$$

ahonnan $\hat{\sigma}^2$ kifejezhető:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i - \hat{m})^2}{n} = s^2 .$$

A maximum likelihood módszer szerint tehát σ^2-re a legjobb becslést az empirikus szórásnégyzet adja. A szórás becslése tehát:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{m})^2}{n}} .$$

11.3. Intervallumbecslés

Az $\hat{a}$ statisztikai becslés és az elméleti a érték között még a legjobb tulajdonságú statisztikák esetén is van véletlen jellegű eltérés. Lehetőség van azonban az $x_1, x_2, ..., x_n$ mintára támaszkodva olyan $\hat{a}_1$ és $\hat{a}_2$ statisztikák létrehozására, amelyekre teljesül, hogy az a paraméter értéke nagy valószínűséggel az $(\hat{a}_1, \hat{a}_2)$ intervallumban található. Ezzel kapcsolatos a következő definíció.

Definíció. Legyen p nullához közeli, kis valószínűség. Az $x_1, x_2, ..., x_n$, n elemű minta segítségével általában létrehozható olyan $\hat{a}_1$ és $\hat{a}_2$ statisztika, amelyekre teljesül az, hogy

$$P(\hat{a}_1 \le a \le \hat{a}_2) = 1 - p .$$

Az $(\hat{a}_1, \hat{a}_2)$ véletlen helyzetű intervallumot **konfidencia** (megbízhatósági)-**intervallumnak** nevezzük. Az $(1–p)\cdot 100\%$-ot a megbízhatóság szintjének nevezzük. Az intervallum kezdő és végpontját konfidencia határoknak nevezzük. Általában $p=0,1$; $p=0,05$; $p=0,01$. Az alábbiakban néhány, a gyakorlatban gyakran előforduló példán keresztül mutatjuk be a konfidencia-intervallum keresésének módszereit.

Konfidencia-intervallum az m várható értékre N(m, σ) eloszlás esetén, ha σ ismert.

Legyen Az $x_1, x_2, ..., x_n$ n elemű minta, és vezessük be az u új valószínűségi változót:

$$u = \frac{\bar{x} - m}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\frac{\bar{x} - m}{\sigma} . \qquad (11.3.1.)$$

Korábbi ismereteink alapján $\overline{x}$ egy $N(m,\sigma/\sqrt{n})$ eloszlású valószínűségi változó. Azt is látjuk, hogy *u* standardizált valószínűségi változó, tehát *u* *N(0,1)* eloszlású valószínűségi változó. Az *N(0,1)* eloszlás táblázata alapján meghatározható az az u_p szám, amelyre teljesül az, hogy:

$$P(-u_p \le u \le u_p) = \frac{1}{\sqrt{2\pi}} \int_{-u_p}^{u_p} e^{-\frac{x^2}{2}} dx = 1 - p. \qquad (11.3.2.)$$

A korábban tanultak alapján

$$\frac{1}{\sqrt{2\pi}} \int_{-u_p}^{u_p} e^{-\frac{x^2}{2}} dx = \Phi(u_p) - \Phi(-u_p) = 2\Phi(u_p) - 1, \qquad (11.3.3.)$$

ahonnan (11.3.2) és (11.3.3) egybevetésével azt kapjuk, hogy

$$\Phi(u_p) = 1 - \frac{p}{2}.$$

Például $p=0,05$ esetén az $\Phi(x)$ táblázatából: $u_p=1,96$ (lásd a Függelék 15.4 fejezetét!). Mivel $1-p$ valószínűséggel érvényes az alábbi összefüggés:

$$-u_p \le \sqrt{n}\frac{\overline{x}-m}{\sigma} \le u_p,$$

a jobb és bal oldali egyenlőtlenség külön-külön megoldásával *m*-re a következő egyenlőtlenségekre jutunk:

$$\overline{x} - u_p \frac{\sigma}{\sqrt{n}} \le m \le \overline{x} + u_p \frac{\sigma}{\sqrt{n}}. \qquad (11.3.4.)$$

Ez azt jelenti, hogy az *m* várható érték $1-p$ valószínűséggel az

$$\left(\overline{x} - u_p \frac{\sigma}{\sqrt{n}}, \overline{x} + u_p \frac{\sigma}{\sqrt{n}}\right)$$

intervallumban van. Ez tehát *m*-re az $(1-p)\cdot 100\%$ szintű konfidencia-intervallum.

Sokszor úgy vetődik fel a kérdés, hogy előírt *p* esetén mekkorának kell lennie a minta elemszámának ahhoz, hogy a konfidencia intervallum félhossza legfeljebb *d* legyen. Ilyenkor a

$$d \ge u_p \frac{\sigma}{\sqrt{n}}$$

egyenlőtlenség megoldása megadja, hogy milyen számnál kell n-nek nagyobbnak lennie. Az egyenlőtlenség megoldásával azt kapjuk, hogy

$$n \geq u_p^2 \frac{\sigma^2}{d^2}. \tag{11.3.5.}$$

Konfidencia-intervallum az m várható értékre N(m, σ) eloszlás esetén, ha σ nem ismert.

Legyen $x_1, x_2, \ldots, x_n$ n elemű minta, és vezessük be a t új valószínűségi változót:

$$t = \sqrt{n}\frac{\bar{x} - m}{s^*}.$$

A t valószínűségi változó átírható a következő formára:

$$t = \sqrt{n-1}\frac{\dfrac{\bar{x} - m}{\sigma / \sqrt{n}}}{\dfrac{s^* \sqrt{n-1}}{\sigma}}.$$

A számlálóban lévő $\dfrac{\bar{x} - m}{\sigma / \sqrt{n}}$, mint láttuk $N(0,1)$ eloszlású változó. A nevezőben $\dfrac{s^* \sqrt{n-1}}{\sigma}$, a korábbiak alapján $n–1$ szabadsági fokú χ^2-eloszlású valószínűségi változó. Innen következik, hogy t $n–1$ szabadsági fokú Student-eloszlású valószínűségi változó.

Az $n–1$ szabadsági fokú Student-eloszlás $F(t)$ eloszlásfüggvényének ismeretében adott p-hez megadható az a t_p érték, amelyre

$$P(-t_p \leq t \leq t_p) = F(t_p) - F(-t_p) = 1 - p.$$

A számolások könnyítése érdekében általában nem az $F(t)$ eloszlásfüggvény táblázatát szokták megadni, hanem olyan táblázatot, amely p-hez t_p-t rendeli hozzá. A t változó definícióját figyelembe véve ez azt jelenti, hogy a táblázatból megkapjuk az a t_p értéket, amelyre a

$$-t_p \leq \sqrt{n}\frac{\bar{x} - m}{s^*} \leq t_p,$$

egyenlőtlenségek $1–p$ valószínűséggel teljesülnek.

A két egyenlőtlenségből az m várható értéket kifejezve azt kapjuk, hogy:

$$\bar{x} - t_p \frac{s^*}{\sqrt{n}} \leq m \leq \bar{x} + t_p \frac{s^*}{\sqrt{n}}. \tag{11.3.6.}$$

Ez azt jelenti, hogy az m várható érték $1–p$ valószínűséggel benne van az

$$\left(\bar{x} - t_p \frac{s^*}{\sqrt{n}}, \bar{x} + t_p \frac{s^*}{\sqrt{n}}\right) \tag{11.3.7.}$$

intervallumban, vagyis

$$P(\bar{x} - t_p \frac{s^*}{\sqrt{n}} \le m \le \bar{x} + t_p \frac{s^*}{\sqrt{n}}) = 1 - p .$$

Kaptunk tehát az m paraméterre egy konfidencia intervallumot az $(1–p)100\%$ megbízhatósági szinten.

Konfidencia intervallum a szórásra

A módszer lényege megegyezik a korábbiakkal. Legyen a sokaság $N(m, \sigma)$ eloszlású! Legyen p kicsiny szám, és keressünk a σ szórásra $(1–p)100\%$ megbízhatósági szenten konfidencia intervallumot.

Legyen $x_1, x_2, \ldots, x_n$ n elemű minta, és vezessük be a v új valószínűségi változót:

$$v = \frac{ns^2}{\sigma^2} .$$

Korábbi ismereteink alapján a v valószínűségi változó $n–1$ szabadsági fokú χ^2 eloszlású valószínűségi változó.

Adott p valószínűséghez az $n–1$ szabadsági fokú χ^2 eloszlás eloszlásfüggvénye segítségével találhatunk olyan intervallumot (ilyen sokféleképpen választható, tehát a választás nem egyértelmű), amelyre igaz, hogy

$$P(v_1 \le v \le v_2) = F(v_2) - F(v_1) = 1 - p .$$

A χ^2-eloszlás nem szimmetrikus, tehát a szokásos szimmetrikus intervallumválasztás nem megoldható. A szokásos intervallumválasztás az alábbi:

$$F(v_1) = \frac{p}{2}, \quad F(v_2) = 1 - \frac{p}{2} .$$

Az ilyen választás teljesíti azt a feltételt, hogy

$$P(v_1 \le \frac{ns^2}{\sigma^2} \le v_2) = F(v_2) - F(v_1) = 1 - p .$$

A szokásos táblázatok a p értékhez azt a v_i értéket adják meg, amelyre igaz, hogy

$$P(\nu < \nu_i) = 1 - p\,.$$

Ezért, ha a táblázatbeli értéket a szokásoknak megfelelően χ_i^2-vel jelöljük, akkor:

$$\nu_1 = \chi^2_{1-\frac{p}{2}} \text{ és } \nu_2 = \chi^2_{\frac{p}{2}}\,.$$

A zárójelen belül átrendezve az egyenlőtlenséget, arra az ekvivalens állításra jutunk, hogy

$$P\left(\frac{ns^2}{\chi^2_{\frac{p}{2}}} \le \sigma^2 \le \frac{ns^2}{\chi^2_{1-\frac{p}{2}}}\right) = 1 - p\,.$$

A szórásnégyzet konfidencia intervalluma tehát:

$$\left(\frac{ns^2}{\chi^2_{\frac{p}{2}}};\frac{ns^2}{\chi^2_{1-\frac{p}{2}}}\right),$$

a σ szórás konfidencia intervalluma négyzetgyökvonás után:

$$\left(\frac{\sqrt{n}s}{\sqrt{\chi^2_{\frac{p}{2}}}};\frac{\sqrt{n}s}{\sqrt{\chi^2_{1-\frac{p}{2}}}}\right).$$

11.4. Statisztikai hipotézisek vizsgálata

A statisztikai hipotézisvizsgálatok során feltevéseket teszünk az események valószínűségére, várható értékére, varianciájára, két változó függetlenségére stb. Ezeket a feltevéseket nevezzük statisztikai hipotéziseknek.

A hipotézisvizsgálat a hipotézisek elfogadásának vagy elvetésének módszereivel foglalkozik. A hipotéziseket statisztikai módszerekkel ellenőrizzük. Ezek az ún. statisztikai próbák.

Mielőtt a statisztikai próbákkal megismerkednénk, definiálnunk kell néhány új fogalmat.

Definíció. Azt a feltevést, amelyet igaznak tételezünk fel, nullhipotézisnek nevezzük, és H_o-lal jelöljük. Például, ha feltesszük, hogy a ξ valószínűségi változó várható értéke m_o, akkor a nullhipotézis:

$$H_o\colon M(\xi)=m_o. \qquad (11.4.1.)$$

Ezzel szemben az ún. ellenhipotézis, amelyet H-val jelölünk:

$$H\colon M(\xi)=m\neq m_o. \tag{11.4.2.}$$

A két hipotézist mindig úgy kell megalkotni, hogy egymást kizáró feltevések legyenek.

A hipotézisvizsgálat menete

Legyen a ξ valószínűségi változó n elemű statisztikai mintája $x_1, x_2, \dots, x_n$. Az a paraméter becslésére az $x_1, x_2, \dots, x_n$ mintaelemek egy

$$\hat{a}=\hat{a}(x_1,x_2,\dots,x_n)$$

statisztikai függvényét használjuk. A döntés úgy történik, hogy a megadott $0<p<1$ számhoz a statisztika összefüggései segítségével olyan $T\subset R$ elfogadási tartományt keresünk, amelyre igaz az, hogy a H_o nullhipotézis fennállása esetén az $\hat{a}$ statisztikai függvény értéke $1-p$ valószínűséggel benne van a T tartományban. Abban az esetben tehát, ha

$$P(\hat{a}\in T \mid H_o)\geq 1-p,$$

akkor $100(1-p)\%$ szignifikancia szinten elfogadjuk a H_o nullhipotézist. Ha $\hat{a}\notin T$, akkor elvetjük a H_o hipotézist, vagyis a H ellenhipotézist fogadjuk el.

Az egymintás u-próba

Adott az $N(m,\sigma)$ normális eloszlású sokaság. Ellenőrizni akarjuk, hogy m egyenlő-e adott m_o számmal. Ismerjük (például korábbi statisztikai vizsgálatokból) σ értékét. Kérdés, hogy az $\overline{x}$ mintaátlag mekkora eltérése esetén feltételezhetjük, hogy a várható érték m_o?

A nullhipotézis:

$$H_o\colon M(\xi)=m_o.$$

Az ellenhipotézis:

$$H\colon M(\xi)=m\neq m_o.$$

Az u-próba menete a következő. Készítünk egy próbafüggvényt (statisztikai függvényt):

$$u=\sqrt{n}\,\frac{\overline{x}-m}{\sigma},$$

amelyről a korábbiak alapján tudjuk, hogy u $N(0,1)$ standard normális eloszlású valószínűségi változó. A konfidencia intervallum kapcsán beláttuk, hogy az u valószínűségi változónak $1\!-\!p$ valószínűséggel megadható a konfidencia intervalluma, azaz

$$P(-u_p \leq \sqrt{n}\frac{\overline{x}-m}{\sigma} \leq u_p) = 1-p\,. \qquad (11.4.3.)$$

Az u_p érték a standard normális eloszlás eloszlásfüggvényének táblázatából meghatározható. Ha most feltesszük, hogy igaz a nullhipotézis, vagyis $m=m_o$, akkor u-ba behelyettesítve ezt az értéket, kiszámíthatjuk az alábbi u_s értéket:

$$u_s = \left|\sqrt{n}\frac{\overline{x}-m_o}{\sigma}\right|.$$

A nullhipotézis igaz volta esetén (11.4.3) szerint fenn kell állnia, hogy

$$u_s \leq u_p\,.$$

Ebben az esetben tehát 100$(1\!-\!p)\%$ biztonsági szinten elfogadjuk a nullhipotézist.

Ha azt találjuk, hogy

$$u_s > u_p\,,$$

akkor $100(1\!-\!p)\%$ biztonsági szinten elutasítjuk a nullhipotézist, vagyis a H ellenhipotézist fogadjuk el.

Mindkét döntésünk a véletlen folytán lehet hibás. Elsőfajú hibát követünk el, ha H_o igaz, de elvetjük, mert úgy találjuk, hogy $u_s > u_p$. Másodfajú hibát követünk el, ha a H_o hipotézis nem igaz, de a véletlen folytán elfogadjuk, mert úgy találjuk, hogy $u_s < u_p$.

Egymintás t-próba

A gyakorlatban általában a normális eloszlású változónak nemcsak a várható értéke, hanem a szórása is ismeretlen. Ilyen esetben ki tudjuk számítani a

$$s^{*2} = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}$$

korrigált szórásnégyzetet.

A nullhipotézis most is:

$$H_o\colon M(\xi)=m_o,$$

amelynek ellenőrzésére a

$$t = \sqrt{n}\,\frac{\overline{x} - m}{s*}$$

próbafüggvény használható. A korábbiak alapján tudjuk, hogy t egy $n–1$ szabadsági fokú Student-eloszlású valószínűségi változó. A Student-eloszlás táblázata alapján p-hez megadható az a t_p táblázati érték, amelyre igaz, hogy

$$P(-t_p \le t \le t_p) = 1 - p\,.$$

Ha most az u-próbához hasonlóan behelyettesítjük a próbafüggvénybe a nullhipotézis által feltételezett $m=m_o$ értéket, akkor t_s próbastatisztikát kapunk:

$$t_s = \left|\sqrt{n}\,\frac{\overline{x} - m_o}{s*}\right|,$$

amelyre H_o fennállása esetén igaznak kell lennie, hogy

$$t_s = \left|\sqrt{n}\,\frac{\overline{x} - m_o}{s*}\right| \le t_p\,. \qquad (11.4.4.)$$

Ha ez az egyenlőtlenség teljesül, akkor $100(1–p)\%$ biztonsági szinten elfogadjuk a H_o hipotézist.

Ha (11.4.4) nem teljesül, vagyis azt találjuk, hogy

$$t_s > t_p\,,$$

akkor a H_o nullhipotézist elutasítjuk, és a H ellenhipotézist fogadjuk el.

F-próba

Az F-próba alkalmazásával eldönthető, hogy két normális eloszlású, ismeretlen várható értékű statisztikai sokaság szórása azonos-e.

A két valószínűségi változó legye ξ és η. Legyen ξ eloszlása $N(m_1,\sigma_1)$, η eloszlása pedig $N(m_2,\sigma_2)$, továbbá legyen $x_1, x_2, \ldots, x_n$ a ξ változóhoz tartozó minta, és $y_1, y_2, \ldots, y_k$ az η változóhoz tartozó minta! A két minta legyen független egymástól!

A nullhipotézis:

$$H_o\text{: } D^2(\xi)=D^2(\eta) \text{ (azaz } \sigma_1^2 = \sigma_2^2\text{, vagyis } \sigma_1=\sigma_2\text{).}$$

Jelölje s_1^{*2} a ξ változóhoz tartozó minta empirikus szórásnégyzetét, és s_2^{*2} az η változóhoz tartozó minta empirikus szórásnégyzetét. A korábbiakban láttuk, hogy

$$\frac{n-1}{\sigma_1^2}s_1^{*2}$$

n–1 szabadságfokú χ^2-eloszlású valószínűségi változó, és hasonlóan

$$\frac{k-1}{\sigma_2^2}s_2^{*2}$$

k–1 szabadságfokú χ^2-eloszlású valószínűségi változó. Mivel a minták függetlenek, ezek a valószínűségi változók is függetlenek. Ebből a két valószínűségi változóból képezhető egy új *F* valószínűségi változó az alábbiak szerint:

$$F=\frac{k-1}{n-1}\frac{\frac{n-1}{\sigma_1^2}s_1^{*2}}{\frac{k-1}{\sigma_2^2}s_2^{*2}}=\frac{\sigma_2^2}{\sigma_1^2}\frac{s_1^{*2}}{s_2^{*2}}. \qquad (11.4.5.)$$

A korábbi ismereteink alapján az *F* valószínűségi változó *(n–1, k–1)* szabadsági fokú *F*-eloszlású. Az *F* eloszlás alapján meghatározhatók azok az F_1 és F_2 értékek, amelyekre igaz, hogy

$$P(F<F_1)=\frac{p}{2},$$

és

$$P(F>F_2)=\frac{p}{2}. \qquad (11.4.6.)$$

Tehát az *F* valószínűségi változó *p* valószínűséggel tartózkodik az (F_1,F_2) tartományon kívül, azaz *1–p* valószínűséggel tartózkodik a tartományon belül.

Ha most feltesszük a nullhipotézis érvényességét, vagyis hogy $\sigma_1=\sigma_2$, akkor (11.4.5) alapján az F_s próbastatisztika:

$$F_s=\frac{s_1^{*2}}{s_2^{*2}}. \qquad (11.4.7.)$$

Általában olyan táblázatunk van, amivel a (11.4.6) relációhoz tartozó értéket tudjuk meghatározni. Ilyen a Függelék 15.9. fejezetében található táblázat is. Ha most (11.4.7)-ben a számlálóba tesszük a nagyobb korrigált empirikus szórásnégyzet értéket, akkor F_s értéke nagyobb lesz *1*-nél. A táblázatból meghatározzuk F_2 értéket, és megnézzük, hogy teljesül-e az

$$F_s\le F_2$$

reláció. Ha teljesül, akkor ez már elegendő $100(1–p)\%$ szignifikancia szinten a H_o nullhipotézis elfogadásához. Ennek oka az, hogy a másik feltétel a következő:

$$P(F_s > F_1) = P\left(\frac{1}{F_s} < \frac{1}{F_1}\right). \qquad (11.4.8.)$$

Ha viszont F_s-t úgy választottuk, hogy $F_s \geq 1$, akkor $1/F_s \leq 1$. Ha megnézzük az F táblázatot, akkor láthatjuk, hogy a gyakorlatban használatos kis p értékek esetén a táblázati értékek mind nagyobbak, mint 1. Következésképpen $1/F_1>1$, tehát (11.4.8) a gyakorlat számára lényeges esetekben mindig teljesül.

Az F-próbát tehát az alábbiak szerint végezzük.

Első lépésben meghatározzuk F_s értékét. Az F_s értékét úgy kell vennünk, hogy a számlálóban van a nagyobb s_i^{*2} érték. Nem szabad eltéveszteni a szabadsági fokok sorrendjét. Ha $s_1^{*2} \geq s_2^{*2}$ és n az s_1^{*2} szabadsági foka, k pedig az s_2^{*2} szabadsági foka, akkor $n–1$, $k–1$ szabadsági fokról van szó.

Második lépésként az F-eloszlás táblázata alapján meghatározzuk a $100(1–p)\%$ szignifikancia szinthez tartozó F_2 értéket (a táblázat $k–1$. sorának és $n–1$. oszlopának értékét kell venni).

A harmadik lépésben összehasonlítjuk F_s és F_2 értékét. Ha igaz az, hogy

$$F_s \leq F_p,$$

akkor a H_o hipotézist $100(1–p)\%$ szignifikancia szinten elfogadjuk, vagyis elfogadjuk, hogy $\sigma_1=\sigma_2$. Ellenkező esetben a H_o hipotézist $100(1–p)\%$ szignifikancia szinten elvetjük, azaz $\sigma_1 \neq \sigma_2$, mert az eltérést szignifikánsnak tekintjük.

Illeszkedésvizsgálat χ^2-próbával

A matematikai statisztikában gyakran előfordul, hogy azt kell vizsgálni, valamely minta származhat-e adott, ismert paraméterekkel rendelkező eloszlásból. Ezt a kérdést vizsgáló statisztikai próbát illeszkedésvizsgálatnak nevezzük.

A próba az ismert eloszlás alapján várható gyakoriságok és a minta gyakorisága közötti eltérés vizsgálatából áll. Nézzük, hogyan megy ez diszkrét valószínűségi változó esetén.

Legyen $A_1, A_2, \ldots, A_r$ teljes eseményrendszer és vizsgálandó az, hogy igaz-e az események valószínűségeire $P(A_i) = p_i$, $i=1, 2, \ldots, r$.

A nullhipotézis:

$$H_o : P(A_i) = p_i \quad i = 1, 2, \ldots, r.$$

Az ellenhipotézis:

$$H:\exists i,\, amelyre\;\; P(A_i)\neq p_i\,.$$

Végezzük el a kísérletet n-szer (tehát készítsünk egy n elemű mintát!). Az eredmény szerint A_1 k_1-szer, A_2 k_2-ször ... A_r k_r-szer következik be. Nyilván

$$\sum_{i=1}^{r} k_i = n\,.$$

A k_i gyakoriságok valószínűségi változók, méghozzá korábban beláttuk, hogy binomiális eloszlást követnek. A binomiális eloszlás várható értéke alapján tehát ismerjük k_i elméleti várható értékét:

$$M(k_i) = np_i\,.$$

A megfigyelt és az elméletileg várt gyakoriság eltéréséből próbastatisztikát készítünk az alábbiak szerint:

$$\chi^2 = \sum_{i=1}^{r} \frac{(k_i - np_i)^2}{np_i}\,.$$

Belátható (ezt most nem tesszük meg), hogy ha n minden határon túl nő, akkor a χ^2-el jelölt valószínűségi változó $r–1$ szabadsági fokú χ^2-eloszlású (a gyakorlatban a megfelelő közelséghez már elegendő, ha $np_i \geq 10$ minden i-re).

A próbát ezek után a következő módon végezzük el. Megadjuk a kívánt p szignifikancia szintet és a χ^2 táblázatból kikeressük az ehhez tartozó χ^2_p értéket, amelyre

$$P(\chi^2 \leq \chi^2_p) = 1 - p\,.$$

Ha a mintából számolt

$$\chi^2_{sz} = \sum_{i=1}^{r} \frac{(k_i - np_i)^2}{np_i}$$

értékre igaz, hogy

$$\chi^2_{sz} \leq \chi^2_p\,,$$

akkor a H_o nullhipotézist elfogadjuk. Ellenkező esetben a a H_o nullhipotézist elvetjük, és a H ellenhipotézist fogadjuk el.

11.5. A regressziós egyenes becslése

Lineáris regresszió esetén a regressziós egyenesre a és b paraméterére kapott értékek a (6.3.1) és a (6.3.2) alapján a következők:

$$a = \frac{M(\xi\eta) - M(\xi)M(\eta)}{M(\xi^2) - M^2(\xi)} = \frac{cov(\xi,\eta)}{D^2(\xi)} = R(\xi,\eta)\frac{D(\eta)}{D(\xi)}, \qquad (11.5.1.)$$

$$b = M(\eta) - aM(\xi)). \qquad (11.5.2.)$$

Ha ismerjük a két változó együttes eloszlását, akkor a paraméterekben szereplő kifejezések kiszámolhatók. Sokszor a mérések során nem ismert az együttes eloszlás. Ebben az esetben a paraméterek becslése úgy végezhető, hogy az elméleti korrelációs együtthatót, az elméleti várható értékeket és a szórásokat becsüljük azok empirikus megfelelőivel. Ilyenkor megkapjuk az a paraméter egy $\hat{a}$ becslését, illetve a b paraméter egy $\hat{b}$ becslését. Emlékeztetőül felírjuk az empirikus korrelációs együtthatóra kapott (2.4.5) kifejezést:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}. \qquad (11.5.3.)$$

Az empirikus szórások pedig

$$s_x = \sqrt{\frac{1}{n}\sum_i (x_i - \overline{x})^2},$$

és

$$s_y = \sqrt{\frac{1}{n}\sum_i (y_i - \overline{y})^2}.$$

Ha most a (11.5.1) és a (11.5.2) kifejezésekbe beírjuk a megfelelő empirikus értékeket, akkor a kapott becsült paraméterek az alábbiak:

$$\hat{a} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}, \qquad (11.5.4.)$$

$$\hat{b} = \overline{y} - \hat{a}\overline{x}. \qquad (11.5.5.)$$

A statisztikai becslés alapján kapott paraméterek becsült értékei megegyeznek a korábban a legkisebb négyzetek módszerével kapott a (2.3.8) és b (2.3.9) értékekkel. Megnyugtató, hogy korábbi eredményünket visszakaptuk, de a statisztika segítségével a fenti becslés tulajdonságairól ennél többet is mondhatunk. Nem általánosan fogjuk a problémát kezelni, hanem olyan esetekre korlátozzuk a meggondolásainkat, amilyenekkel a kísérletek végzése során gyakorta találkozunk.

A kiinduló feltevés az, hogy az egyik változó szórása sokkal kisebb, mint a másik változóé. Legyen ez a ξ változó, és tekintsük úgy, hogy ennek x értékeit pontosan ismerjük. A kísérletek során gyakori helyzet, hogy a mérés során $Y=ax+b$ függést tételezünk fel, az x értéket beállítjuk (tehát pontosan ismerjük), és mérjük, hogy az adott x érték mellett y milyen értéket vesz fel. Bár a függvény alakját ismerjük az elméletből, de a mindig jelenlévő statisztikus hibák miatt a mért y érték szórást mutat.

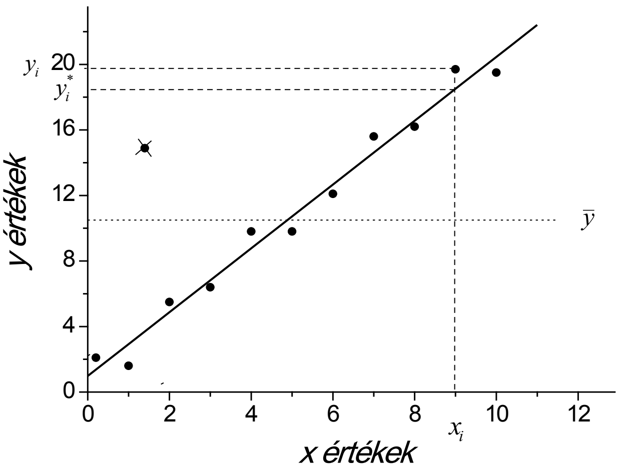
A fentiekben jellemzett helyzetet írjuk le a statisztikában definiált fogalmakkal. Tehát a ξ nem valószínűségi változó. A mérés során konkrét $x_1, x_2, \ldots, x_n$ értékeit mi választjuk meg, és pontosan ismerjük is ezeket az értékeket. Az $y = \varphi(x)$ függvény lineáris, vagyis: $Y = ax + b$. Nem ismerjük viszont az a és b paramétereket, de (11.5.4) és (11.5.5) alapján már elkészítettük becslésüket. Mivel az y_i-k valószínűségi változók, véletlen hibát tartalmaznak, ezért értékük eltér az elméleti $Y_i = ax_i + b$ értéktől. Igaz az, hogy

$$y_i = Y_i + \varepsilon_i ,$$

ahol az ε_i-k függetlenek, és $M(\varepsilon_i)=0$ $\forall i-$re. Ez úgy is írható, hogy

$$M(y_i | x_i) = Y_i .$$

A mérési hibák esetében a centrális határeloszlás tétel értelmében általában feltehető, hogy ε_i normális eloszlású σ_{y_i} szórással, tehát az eloszlás: $N(0,\sigma_{y_i})$. Tegyük fel azt is, hogy, $\sigma_{y_i} = \sigma_y$ $\forall i-$re, vagyis a szórás az egész tartományon állandó. Ha a mérési tartomány nem túl széles, akkor a szórásokra tett feltevés általában teljesül.



11.1. ábra: A legkisebb négyzetek módszerével kapott regressziós egyenes

Ilyen feltevések mellett a legkisebb módszerekkel a-ra és b-re kapott $\hat{a}$ és $\hat{b}$ becslések tulajdonságaira vonatkozóan a statisztika módszereivel információt kaphatunk. Az $\hat{a}$ és $\hat{b}$ valószínűségi változók, kiszámolható tehát várható értékük és szórásuk.

Az $\hat{a}$ és $\hat{b}$ várható értéke

Az $\hat{a}$ becslés várható értékének kiszámolásához felhasználjuk, hogy

$$y_i = Y_i + \varepsilon_i \text{ és } M(\varepsilon_i) = 0,$$

tehát

$$M(y_i) = M(Y_i) + M(\varepsilon_i) = ax_i + b,$$

hiszen x_i nem valószínűségi változó. Továbbá az is igaz, hogy

$$\bar{y} = \frac{1}{n}\left(\sum_{i=i}^{n} Y_i + \sum_{i=1}^{n} \varepsilon_i\right),$$

és

$$M(\bar{y}) = \frac{1}{n}\left(\sum_{i=i}^{n} M(Y_i) + \sum_{i=1}^{n} M(\varepsilon_i)\right) = \frac{1}{n}\sum_{i=i}^{n} M(Y_i) = a\bar{x} + b.$$

Tehát az $\hat{a}$ becslés várható értéke:

$$M(\hat{a}) = \frac{\sum_{i=1}^{n} x_i M(y_i) - na\bar{x}^2 - nb\bar{x}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i - na\bar{x}^2 - nb\bar{x}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = a. \quad (11.5.6.)$$

A $\hat{b}$ becslés várható értéke:

$$M(\hat{b}) = M(\bar{y}) - \bar{x}M(\hat{a}) = a\bar{x} + b - \bar{x}a = b. \quad (11.5.7.)$$

Vagyis $\hat{a}$ és $\hat{b}$ torzítatlan becslése a és b-nek.

Az $\hat{a}$ és $\hat{b}$ szórása

Ha az y_i értékek σ_y^2 szórásnégyzete valahonnan ismert (például onnan, hogy egy pontban sokszor mértünk, és az így meghatározott empirikus szórást felhasználjuk a szórás becslésére), és ez minden y_i pontban ugyanaz az állandó érték, akkor a szórásnégyzet számolás szabályai alapján (11.5.4)-ből és (11.5.5)-ből egyszerűen kiszámolhatjuk az $\hat{a}$ és $\hat{b}$ becsült értékek szórásnégyzetét:

$$D^2(\hat{a}) = \frac{\sigma_y^2}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}, \tag{11.5.8.}$$

felhasználva, hogy

$$D^2(y_i) = \sigma_y^2 \text{ és } D^2(\overline{y}) = \frac{1}{n^2}\sum_{i=1}^{n} D^2(y_i) = \frac{1}{n}\sigma_y^2 .$$

Továbbá

$$D^2(\hat{b}) = \sigma_y^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} \right). \tag{11.5.9.}$$

Mivel

$$\sum_i x_i^2 - n\overline{x}^2 = \sum_i (x_i - \overline{x})^2 , \tag{11.5.10.}$$

ezért látható, hogy n növekedtével $\hat{a}$ és $\hat{b}$ szórása nullához tart, tehát a becslés konzisztens.

A (11.5.8) és (11.5.9) kifejezésekből (11.5.10) figyelembe vételével az is látszik, hogy adott n mérésszám esetén a szórásnégyzetek annál kisebbek minél távolabb fekszenek a mérési pontok $\overline{x}$ értékétől. A kísérleti terv a hiba szempontjából tehát akkor optimális, ha az x_i pontok a vizsgált intervallum szélén helyezkednek el. Igaz, ilyenkor nem ellenőrizhető a vizsgált függvény lineáris jellege. Akkor célszerű így tervezni a mérést, ha a lineáris függvénykapcsolat fennállását korábban már ellenőriztük.

Megjegyzés

Ha az y_i mérési pontok σ_y szórása nem ismert, akkor ennek jó közelítése az

$$s_r^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2}, \tag{11.5.11.}$$

a különböző x_i pontokban mért y_i étékek alapján számolt ún. reziduális szórásnégyzet. A nevezőben itt azért szerepel $n–2$, mert a számlálóban szereplő n darab különbségnégyzet nem mind független, közöttük a (11.5.4) és (11.5.5) két egyenlet kapcsolatot teremt. A független adatok száma $n–2$.

FÜGGELÉK

12. A KOMBINATORIKA ALAPJAI

A kombinatorika a sorba rakás és kiválasztás kérdéseivel foglalkozik.

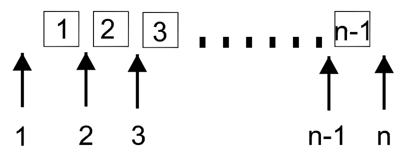
12.1. Permutációk (sorba rakás)

Egymástól különböző n elem meghatározott sorrendjét az n elem egy permutációjának nevezzük. Kérdés, hányféleképpen lehet n egymástól különböző elemet sorba rakni, vagyis mennyi n elem összes permutációinak P_n száma?

Tétel: $P_n=n!$, $(n!=1\cdot2\cdot3\cdot\ldots\cdot(n-1)\cdot n;\quad 0!=1)$.

Bizonyítás. A bizonyítás teljes indukcióval történhet.

1. Az első lépésben könnyen belátható, hogy ha $n=1$, akkor $P_n=1!=1$.
2. A második lépésben feltesszük, hogy $P_{n-1}=(n-1)!$.
3. A harmadik lépésben belátjuk, hogy P_n a P_{n-1}-ből úgy kapható, hogy az n. elemet egy permutáció minden lehetséges pozíciójába elhelyezzük. n ilyen pozíció van.



12.1. ábra: Ábra a permutációk számának meghatározásához

Ezt megtehetjük mind az $(n-1)!$ permutáció esetében, vagyis:

$$P_n = nP_{n-1} = n\cdot(n-1)! = n! \qquad (12.1.1.)$$

12.2. Ismétléses permutációk

Ha az n elem között k azonos, akkor az azonos elemek egymás közötti cseréje nem változtat az elrendezésen. Ha az n elem között k azonos elem van akkor ismétléses permutációról beszélünk, és ilyenkor a permutációk számát $P_n^{(k)}$-val jelöljük. Látszik, hogy

$$P_n^{(k)} < P_n \text{ (ha } k>1).$$

Nézzük az alábbi példát:

$$a\ a\ b;\quad a\ b\ a;\quad b\ a\ a.$$

Az $n=3$ elem közül kettő azonos, és a permutációk száma $P_n^{(k)}=3$, és nem $P_n=3!=6$, hiszen például a_1a_2b nem különbözik az a_2a_1b permutációtól.

Általánosan megfogalmazva az ismétléses permutációra vonatkozó tételt:

Tétel: Ha n elem közül k azonos, akkor az ismétléses permutációk száma:

$$P_n^{(k)}=\frac{n!}{k!}. \qquad (12.2.1.)$$

Bizonyítás. Az ismétléses esetet visszavezetjük az ismétlés nélküli esetre. Az azonos elemeket átmenetileg különböztessük meg valahogy (például sorszámozással). Ezek egymás között $k!$-féleképpen permutálhatók. Tehát ilyenkor minden $P_n^{(k)}$ permutációból $k!$ permutáció képezhető. Az összes ilyen megadja az ismétlés nélküli permutációk számát. Tehát

$$k!\,P_n^{(k)}=P_n,$$

vagyis

$$P_n^{(k)}=\frac{P_n}{k!}=\frac{n!}{k!}.$$

Tehát a fenti példában $n=3$, $k=2$, azaz $n!=6$, $k!=2$, tehát $P_n^{(k)}=3$.

Ha n elem közül több egymás között azonos ismétlődő elem is van, akkor az ismétléses permutációk számát az alábbi képlet alapján számoljuk.

Tétel. Ha n elem közül k_1 azonos, majd a fennmaradó $n–k$ elemből k_2 azonos, stb., akkor az ismétléses permutációk száma:

$$P_n^{(k_1,k_2,\dots,k_i)}=\frac{n!}{k_1!\,k_2!\dots k_i!},\quad \text{ahol } k_1+k_2+\dots+k_i\le n. \qquad (12.2.2.)$$

A bizonyítás teljes indukcióval történhet.

12.3. Kombinációk (kiválasztás, sorrend nélkül)

Ha n elem közül k elemet kiválasztunk, azt az n elem egy k-ad osztályú kombinációjának nevezzük. A kérdés az, hogy n elemnek hány k-ad osztályú kombinációja ($C_n^{(k)}$) van?

Tétel. n elem k-ad osztályú kombinációinak száma:

$$C_n^{(k)} = \binom{n}{k} = \frac{n!}{k!(n-k)!}. \qquad (12.3.1.)$$

Bizonyítás. A bizonyítás úgy történhet, hogy a problémát visszavezetjük az ismétléses permutációkra. Jelölje az n különböző elemet $a_1\ a_2\ a_3 \ldots a_{n-2}\ a_{n-1}\ a_n$. A kiválasztás jelölése úgy történik, hogy a kiválasztott k elem alá 1-et írunk. $n–k$ elem nincs kiválasztva, alájuk 0-t írunk.

$$\begin{array}{cccccc} a_1 & a_2 & a_3 \ldots a_{n-2} & a_{n-1} & a_n \\ 0 & 1 & 1 \quad\; 0 & 1 & 0. \end{array}$$

Minden kiválasztás megfelel az 1-esek és a 0-k egy elrendezésének (sorrendjének). Ez száma megegyezik n elem ismétléses permutációinak számával, ahol k és $n–k$ azonos elem van. Ezek száma az ismétléses permutáció (12.2.2) képlete alapján:

$$C_n^{(k)} = P_n^{(k,n-k)} = \frac{n!}{k!(n-k)!}.$$

12.4. Ismétléses kombinációk

Az ismétléses kombináció az a struktúra, amikor n különböző elemből k-t választunk ki, de egy elem akárhányszor felhasználható. Az ismétléses kombináció jelölése: $C_{n,ism}^{(k)}$. Például az a, b, c három elem esetén a 2-od osztályú ismétléses kombinációk: *a a, a b, a c, b b, b c, c c.* A harmad osztályú ismétléses kombinációk: *a a a, a a b, a a c, a b b, a b c, a c c, b b b, b b c, b c c, c c c.*

Kérdés, hogyan adhatjuk meg a k-ad osztályú ismétléses kombinációk számát?

Tétel. A k-ad osztályú ismétléses kombinációk száma:

$$C_{n,ism}^{(k)} = \binom{n+k-1}{k}. \qquad (12.4.1.)$$

Bizonyítás. Feltesszük, hogy a kombinációkat $1, 2, 3, \ldots, n$ elemekből (tehát számokból) készítjük. A kiválasztott elemeket egy kombináción belül rakjuk nagyság szerinti sorrendbe! (Ez mindig megtehető.) Tekintsünk egy ismétléses kombinációt, és az elemekhez adjunk rendre *0, 1, 2 ..., k–1*-et! Az így kapott k szám között nincs egyforma, és nagyság szerinti sorba rendezett. Tehát kombinációt kaptunk, de ismétlés nélkülit. Kérdés: mely elemek kombinációit? A legkisebb elem: $1+0=1$, a legnagyobb elem: $n+k-1$, és minden köztük lévő szám szerepel. Ilyen módon az $1, 2, ..., n$ elem ismétléses kombinációihoz egyértelműen hozzárendeltük $1, 2, ..., n+k-1$ elem egy k-ad osztályú ismétlés nélküli kombinációját. A megfeleltetés kölcsönösen egyértelmű, tehát

$$C_{n,ism}^{(k)} = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!} = \frac{n(n-1)...(n+k-1)}{k!}.$$

12.5. Variációk

Variációnak nevezzük azt a struktúrát, amikor n különböző elemből k-t választunk ki, de a sorrendet is figyelembe vesszük. Kérdés hogy mennyi a variációk száma?

Tétel. n elem k-ad osztályú variációinak száma

$$V_n^{(k)} = n(n-1)...(n-k+1). \qquad (12.5.1.)$$

Bizonyítás. Visszavezetjük a variációkat a kombinációkra. Láttuk, ha n elemből k elemet kiválasztunk, akkor ezt $C_n^{(k)} = \binom{n}{k}$-féleképpen tehetjük. Minden kombinációhoz vegyük a k elem valamennyi permutációját. Ez a (12.1.1) képlet szerint: $P_k=k!$. A sorrendet is figyelembe vevő kiválasztás (variáció), tehát az így kapott két szám szorzataként kapható meg.

Tehát:

$$V_n^{(k)} = C_n^{(k)} k! = \binom{n}{k} k! = \frac{n!}{k!(n-k)!} k! = n(n-1)\cdots(n-k+1).$$

12.6. Ismétléses variációk

Ez az eset abban különbözik az ismétlés nélkülitől, hogy a kiválasztás után az elemet visszatesszük, és megengedjük, hogy újból kiválasszuk. Kérdés, hogy ilyen módon hányféle különböző kiválasztás lehetséges?

Tétel. n elem k-ad osztályú ismétléses variációinak száma

$$V_{n,ism}^{(k)} = n^k \,. \tag{12.6.1.}$$

Bizonyítás. Van *1, 2, ..., n–1, n* elem, és van *1, 2,..., k* pozíció. Minden pozícióba választhatjuk az *n* elem valamelyikét. Az első pozícióba van *n* választási lehetőség, a másodikba ismét *n*, stb. Összesen tehát a kiválasztási lehetőségek száma: $n \cdot n \cdot n \cdots n = n^k$, vagyis

$$V_{n,ism}^{(k)} = n^k \,.$$

12.7. A binomiális tétel és a binomiális együtthatók

A kombinatorikában gyakran találkozunk a binomiális együtthatóval. Célszerű tehát áttekinteni a binomiális együttható tulajdonságait. Mindenekelőtt kezdjük a binomiális tétellel, ahol szintén megjelennek a binomiális együtthatók, és a bizonyításhoz is a kiválasztás törvényszerűségeinek ismerete szükséges.

Tétel. A binomiális tétel.

Ha *n* egész szám, *a* és *b* pedig tetszőleges valós (vagy komplex) számok, akkor

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k \,. \tag{12.7.1.}$$

Bizonyítás. Képezzük a $P=(a+b_1)(a+b_2)\ldots(a+b_n)$ segédszorzatot! Felbontjuk a zárójeleket, és a tagokat a hatványai szerint rendezzük. A tagok *n* tényezős szorzatok, melynek tényezői rendre *P* első, második stb. tényezőjéből valók. Tehát

$$P = a^n + B_1 a^{n-1} + B_2 a^{n-2} + \ldots + B_n \,.$$

Kérdés, hogy B_k tényezők milyen *b* értékeket tartalmaznak?

$B_1= b_1+ b_2+\ldots+b_n$ hiszen *n* darab a^{n-1}-et tartalmazó tag van.

$B_2=b_1b_2+b_1b_3+\ldots+b_1b_n+b_2b_3+\ldots+b_{n-1}b_n$, ahol *n–2* *a*-t veszünk, és minden lehetséges módon két b_i szorzatát.

.

.

.

$B_n=b_1 b_2\ldots b_n$, itt csak egy tag van.

A kérdés az, hogy B_k-ban hány tag van? Vegyük észre, hogy B_k-ban éppen annyi tag van ahányféleképpen az *n* elemből *k* elemet ki tudunk választani. Tehát a tagok száma:

$$C_n^k = \binom{n}{k}.$$

ha most a b értékeket nem különböztetjük meg, tehát ha $b_1 = b_2 = \ldots = b_n$, akkor

$$B_k = \binom{n}{k} b^k .$$

Innen következik, hogy:

$$(a+b)^n = \binom{n}{0} a^n + \binom{n}{1} a^{n-1} b^1 + \binom{n}{2} a^{n-2} b^2 + \ldots + \binom{n}{n} b^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k .$$

A binomiális tételhez hasonlóan belátható a polinomiális tétel, amely r elem összegének n-dik hatványát adja meg.

Tétel. A polinomiális tétel szerint:

$$(a_1 + a_2 + \ldots + a_r)^n = \sum_{k_1 + k_2 + \ldots + k_r = n} \frac{n!}{k_1! k_2! \cdots k_r!} a_1^{k_1} a_2^{k_2} \cdots a_r^{k_r} . \qquad (12.7.2.)$$

A bizonyítás elvégezhető a binomiális tételből kiindulva, teljes indukcióval.

12.8. A binomiális együtthatók néhány tulajdonsága

Az $\binom{n}{k}$ binomiális együtthatók az ún. Pascal-háromszögbe rendezhetők.

0. sor							**1**							
1. sor						**1**		**1**						
.					**1**		**2**		**1**					
.				**1**		**3**		**3**		**1**				
.			**1**		**4**		**6**		**4**		**1**			
		1		**5**		**10**		**10**		**5**		**1**		
	1		**6**		**15**		**20**		**15**		**6**		**1**	

1 7 21 35 35 21 7 1

12.2. ábra: A Pascal-háromszög első nyolc sora

Az együtthatók tulajdonságai leolvashatók a háromszögből:

a. Szimmetria, azaz

$$\binom{n}{k}=\binom{n}{n-k}.$$

b. Egy elem egyenlő a felette lévő két elem összegével:

$$\binom{n+1}{k+1}=\binom{n}{k}+\binom{n}{k+1}$$

c. Egy elem egyenlő a felette lévő sorok eggyel kisebb sorszámú elemeinek összegével:

$$\binom{n+1}{k+1}=\binom{n}{k}+\binom{n-1}{k}+\binom{n-2}{k}+\ldots+\binom{k}{k}.$$

d. A Pascal-háromszögben az n. sor elemeinek összege $=2^n$. Vagyis

$$\sum_{k=0}^{n}\binom{n}{k}=2^n .$$

Ez a binomiális tétel alapján is azonnal látszik, hiszen:

$$(a+b)^n=\sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k ,$$

és ha elvégezzük a következő helyettesítést: $a=b=1$, akkor azt kapjuk, hogy

$$(1+1)^n=\sum_{k=0}^{n}\binom{n}{k}.$$

13. HALMAZELMÉLETI ALAPFOGALMAK

A halmazokon végzett műveletek Boole-algebrának nevezzük, George Boole angol matematikus (1815–1864) tiszteletére.

13.1. A halmazok definíciója

A halmazok definíciója legyen az alábbi.

Definíció. Halmaznak az elemek összességét tekintjük. Halmazelem pedig lehet valamilyen objektum, személy, de akár esemény is. A halmazokat nagybetűvel, a halmazelemeket pedig kisbetűvel jelöljük, és kapcsos zárójelbe tesszük. Például ha az *A* halmaznak az elemei *a, b*, *c,* valamint *B* halmaznak elemei *a, b, c, d, e*, akkor ennek jelölése:

$$A = \{ a,b,c \},\ B = \{ a,b,c,d,e \},$$

továbbá $a \in A$, $d \notin A$ azt jelöli, hogy *a* eleme *A*-nak, de *d* nem eleme *A*-nak.

Az *A* halmaz részhalmaza *B*-nek, azt jelenti, hogy *A* halmaz minden eleme *B* halmaznak is eleme (mint a fenti példában). Ezt úgy jelölhetjük, hogy $A \subset B$.

A fenti definíciónak következménye az, hogy ha *A* halmaz *B* részhalmaza, és *B* halmaz a *C* halmaznak részhalmaza, akkor ebből következik, hogy az *A* halmaz a *C* halmaznak is részhalmaza. A jelölésekkel ugyanez az állítás:

$$\text{ha } A \subset B \text{ és } B \subset C \rightarrow A \subset C.$$

13.2. Halmazok összege

A halmazok összeadását az algebrában szokásos + jellel jelöljük, de felhívjuk a figyelmet, hogy a halmazösszeadás tulajdonságai némileg különböznek az algebrában megszokott összeadási tulajdonságoktól.

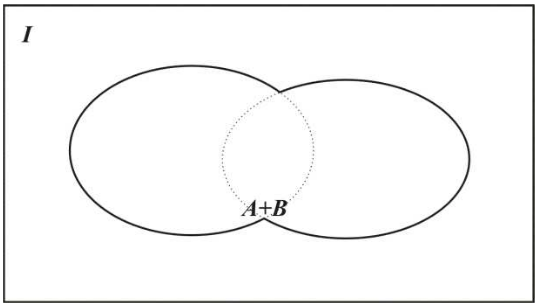
Definíció. Az *A+B* összeghalmaz azon elemek összessége, amelyek legalább *A*-nak vagy *B*-nek elemei.

Például legyen $A=\{1,2\}$; $B=\{2, 4, 6, 8\}$, ekkor az összeghalmaz

$$A+B=\{1, 2, 4, 6, 8\}.$$

A definícióból következik, hogy a *2* elemet nem vesszük kétszer az összeghalmazban!

Szokás a halmazokat ábrán, az ún. Venn-diagramon megjeleníteni. A halmazok összeadását Venn-diagramon a 13.1. ábra mutatja. Az ábrán az összeghalmazt a vastagon kihúzott vonal mutatja.



13.1. ábra: Az A+B halmaz ábrázolása Venn-diagramon

Az összeadás tulajdonságai

Az összeadás tulajdonságai a műveleti definícióból következnek:

$A+A=A$	indempotencia,
$A+B=B+A$	kommutativitás,
$A+(B+C)=(A+B)+C$	asszociativitás.

Az asszociativitásból következik, hogy az $A+B+C$ nem félrevezető írásmód, hiszen három halmaz bármilyen sorrendben összeadható, az eredményt a sorrend nem érinti.

A definíciónak az is következménye, hogy

ha $A \subset B$, akkor $A+B=B$.

Ha több halmazt adunk össze, akkor szokásos jelölési mód az alábbi:

$$A_1 + A_2 + \ldots + A_n = \sum_{i=1}^{n} A_i \, .$$

13.3. Halmazok szorzata

Definíció. Az AB szorzat azon elemek összessége, melyek A-nak és B-nek egyaránt elemei (halmazok közös része). A szorzás jele tehát a tényezők közötti pont, amit sokszor nem írunk ki, ahhoz hasonlóan, ahogyan azt az algebrában is tesszük. Mindazonáltal a szorzás tulajdonságai különböznek az algebrában megszokott tulajdonságoktól.

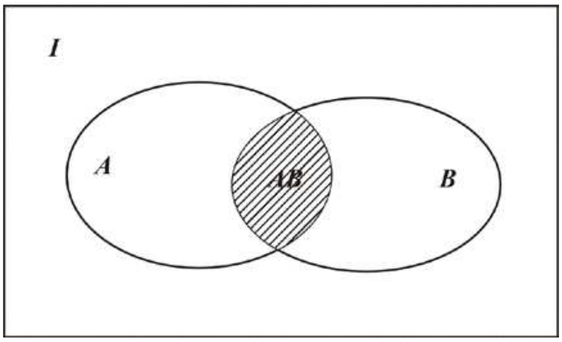
Például vegyük a következő A és B halmazok szorzatát:

$$A=\{1, 2, 3\};B=\{2, 3., 4, 5\}.$$

A szorzat a két halmaz közös elemi lesznek, tehát

$$AB=\{2, 3\}.$$

Ha A és B két halmaz szorzatát a 13.2. ábra Venn-diagramján láthatjuk.



13.2. ábra: Az A és B halmazok szorzata Venn-diagramon ábrázolva

A szorzás tulajdonságai a műveleti definícióból következnek:

$AA=A$	idempotencia,
$AB=BA$	kommutativitás,
$(AB)C=A(BC)$	asszociativitás.

A definíció következményei:

– Ha $A \subset B$, akkor $AB=A$.

– Az n tényezős szorzat esetén szokásos felírás:

$$\prod_{i=1}^{n} A_i = A_1 \cdot A_2 \cdot \ldots \cdot A_n .$$

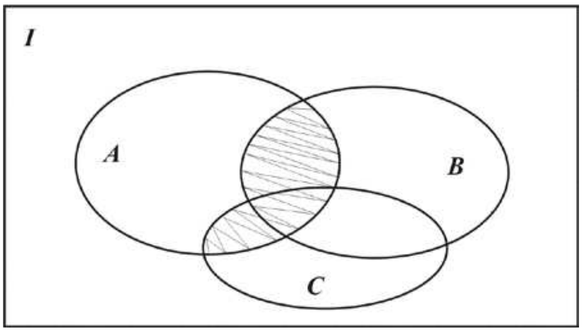
– Az összeadás és szorzás közös tulajdonsága a disztributivitás:

$$A(B+C)=AB+AC.$$

A disztributivitást szemléltető Venn-diagramon (13.3. ábra) látszik, hogy a jobb és bal oldal szerint elvégzett műveletek azonos halmazhoz vezetnek.

Több halmazra általánosítva a disztributivitás szabályát:

$$A\sum_{i=1}^{n} B_i = \sum_{i=1}^{n} AB_i .$$



13.3. ábra: A disztributivitást szemléltető ábra

Mielőtt továbbmennénk a műveletek sorában, ismerkedjünk meg új fogalmak definícióival.

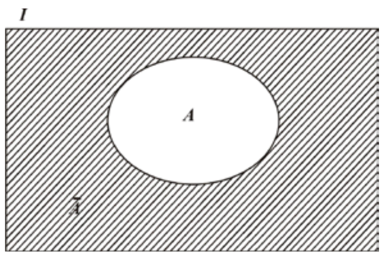
Definíció. Üres halmaznak nevezzük azt a halmazt, amelynek nincs eleme. Az üres halmaz jele: *Ø*. Megállapodás szerint az üres halmaz minden halmaz részhalmaza.

Az üres halmaz definíciójának következményei:

– Ha A és B halmazoknak nincs közös eleme, akkor $AB=Ø$. Az ilyen közös elem nélküli halmazokat *diszjunkt halmazoknak* nevezzük.

– $A+Ø=A$.
– $A\,Ø=Ø$.

Definíció. Az A halmaz I halmazra vonatkoztatott komplementer (kiegészítő) halmazának nevezzük azt az $\overline{A}$ halmazt, amely azon elemek összessége, amelyek az I halmaznak elemei, de nem tartoznak az A halmazhoz.



13.4. ábra: Az $\overline{A}$ komplementer halmaz ábrázolása Venn-diagramon

A komplementer halmaz definíciójának következményei:

– $A+\overline{A}=I$,

– $A\overline{A}=Ø$,

– $\overline{I}=Ø$,
– $\overline{Ø}=I$.

A komplementer halmaz fogalmához kötődnek a következő azonosságok, amelyeket *de Morgan* szabályoknak nevezünk:

$$\overline{A+B} = \overline{A}\,\overline{B}\,, \qquad (13.3.1.)$$
$$\overline{AB} = \overline{A} + \overline{B}\,. \qquad (13.3.2.)$$

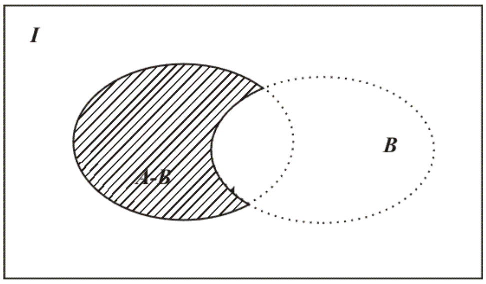
A *de Morgan* szabályokat könnyen igazolhatjuk a Venn-diagramon, hiszen a jobb és bal oldal ugyanarra a halmazra vezet.

13.4. Halmazok különbsége

Definíció. Az *A* és *B* halmaz különbsége alatt a halmazt értjük, amelynek elemei *A* halmazhoz tartoznak, de nem tartoznak a *B* halmazhoz. A különbségképzés jele a szokásos kivonásjel, tehát halmazok különbségét így jelöljük: *A–B*.

Példaként a két halmaz legyen a következő: *A={1, 2, 4, 6}; B={4, 6, 8, 10}*. A különbség halmaz: *A–B={1, 2}*.

Venn-diagramon is ábrázolhatjuk a különbséghalmazt, ahogyan azt a 13.5. ábra mutatja.



13.5. ábra: A különbséghalmaz ábrázolása Venn-diagramon

A különbségképzés definíciójának következményei:

– $A - B = A\overline{B}$ (ez lehetne a kivonás definíciója is).
– Általában *(A–B)+B≠A*! hanem *(A–B)+B=A+B*.
– Ellenben, ha $B \subset A$, akkor *(A–B)+B=A*.

14. A GYORS ELLENŐRZŐ FELADATOK MEGOLDÁSAI

14.1. Az 1. fejezethez

1.1.

$$\sum_{i=1}^{m} \Delta x_i \frac{k_i}{n\Delta x_i} = \sum_{i=1}^{m} \frac{k_i}{n} = \frac{1}{n}\sum_{i=1}^{m} k_i = 1, \text{ mivel } \sum_{i=1}^{m} k_i = n.$$

1.2.

Az utolsó oszlop magassága:

$$\sum_{i=1}^{m} \Delta x_i \frac{k_i}{n\Delta x_i},$$

erről pedig az előző feladatban beláttuk, hogy egyenlő 1-el.

1.3.

$$f''(x) = 2n, \text{ ha } n > 0 \text{ akkor } f'' > 0.$$

1.4. Az átlagszámoláshoz célszerűen az (1.6.6) kifejezés használható, azaz $c=100$-al elosztjuk az átlagolandó értékeket, majd a végeredményt szorozzuk 100-al.

$$\frac{14+12+2+6+11}{5} = \frac{45}{5} = 9,$$

tehát az átlag értéke: $\bar{x} = 9 \cdot 100 = 900$.

1.5. A feladatot úgy célszerű elvégezni, hogy *299 000 km/h* értéket levonunk minden adatból, és csak a táblázatbeli adatokat adjuk össze (például Excel programmal). Az eredményt elosztjuk az adatok számával (jelen esetben *100*-al), majd a végeredményhez hozzáadjuk az előzőleg levont *299 000* számot. Az így kapott átlag *299 852 km/h*.

1.6. Ha a táblázatbeli csökkenés értékeket p_i-vel, a népességszámot 2004-ben L-el jelöljük, akkor a számítás módja a következő:

$$L(1-p_1)(1-p_2)(1-p_3)(1-p_4)(1-p_5)=L\cdot 0{,}991522=10\ 030\ 976\ .$$

Az átlagos népességcsökkenés p értéke az öt év alatt:

$$(1-p)^5=0{,}991522\,,$$
$$p=1-\sqrt[5]{0{,}991522}=1{,}7013\text{‰}\,.$$

A gyökvonást célszerű logaritmus segítségével végezni.

1.7. Az átlagot a harmonikus közép segítségével számolhatjuk ki. Tehát az átlagos sűrűség:

$$\overline{\rho}=\frac{2}{\frac{1}{\rho_1}+\frac{1}{\rho_2}}\,.$$

1.8. Párhuzamos ellenállások esetén is a harmonikus középpel számolható az átlag:

$$\overline{R}=\frac{2}{\frac{1}{R_1}+\frac{1}{R_2}}=\frac{2}{\frac{1}{100}+\frac{1}{200}}=133{,}33\ \Omega\,.$$

1.9. Alkalmazva az (1.11.3) összefüggést, a végeredmény:

$$s^2=3{,}21\ldots;\ s=1{,}79\ldots$$

1.10. Az eredménynek meg kell egyeznie az előző feladat eredményével.

14.2. A 2. fejezethez

1.1. Azt már korábban beláttuk (1.11.3), hogy

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\overline{x}^2=s^2\geq 0\,.$$

$n>0$-val megszorozva az egyenlet mindkét oldalát, azt kapjuk, hogy

$$\sum_{i=1}^{n}(x_i-\overline{x})^2=\sum_{i=1}^{n}x_i^2-n\overline{x}^2\,.$$

Ez a kifejezés csak akkor 0, ha $x_i = \bar{x} \quad \forall i$ -re. Ilyenkor azonban az egyenesnek nincs sok értelme, ezért ezt kizárva, csak a > jel az érvényes.

1.2. Mivel az előző feladatban beláttuk, hogy r (2.4.5) és a (2.3.8) nevezője mindig pozitív, ezért amennyiben a<0, akkor az azt jelenti, hogy számlálója negatív, ez viszont megegyezik r számlálójával, ami emiatt szintén negatív.

1.3. A számlálót átalakítva azt kapjuk, hogy

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}(x_iy_i-x_i\bar{y}-\bar{x}y_i+\overline{xy})=\sum_{i=1}^{n}x_iy_i-n\overline{xy}-\bar{x}n\bar{y}+n\overline{xy}=\sum_{i=1}^{n}x_iy_i-n\overline{xy}.$$

A nevezőben:

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^2-n\bar{x}^2 \quad \text{és} \quad \sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}y_i^2-n\bar{y}^2 .$$

14.3. A 3. fejezethez

3.1. A feladat lényegében az, hogy n elemből összesen hányféleképpen lehet kiválasztani k=0, 1, … n elemet. A binomiális tételt felhasználva:

$$\sum_{k=0}^{n}\binom{n}{k}=(a+b)^n=2^n \text{, ha } a=b=1.$$

3.2. $A–AB$ és B nem rendelkezik közös résszel, azaz $(A-AB)B=\emptyset$. Ugyanakkor $(A–AB)+B=A+B$.

3.3. Az előző feladat eredményét felhasználva:

$$(A+B)-B=(A-AB)+B-B=A \text{, ha } AB=\emptyset.$$

3.4. A teljes indukció módszerét alkalmazzuk. A módszer három lépésből áll. Az első lépésben be kell látni, hogy ha $A_1A_2=\emptyset$, akkor

$$k_{A_1+A_2}=k_{A_1}+k_{A_2}.$$

Ezt korábban már beláttuk.

A második lépésben feltesszük, hogy $n–1$ diszjunkt halmazra igaz, hogy

$$k_{A_1+A_2+\ldots+A_{n-1}} = k_{A_1} + k_{A_2} + \ldots + k_{A_{n-1}}.$$

A harmadik lépésben belátjuk, hogy ha a feltevés igaz, akkor n diszjunkt halmazra is igaz az állítás. Ehhez, tegyük fel, hogy

$$B = A_1 + A_2 + \ldots + A_{n-1},$$

és legyen A_n olyan halmaz, amelyre igaz, hogy $A_nB = \emptyset$. Ekkor viszont az 1. lépés szerint

$$k_{B+A_n} = k_B + k_{A_n}.$$

Innen a 2. lépés miatt azt kapjuk, hogy

$$k_{B+A_n} = k_{A_1+A_2+\ldots+A_{n-1}+A_n} = k_{A_1} + k_{A_2} + \ldots + k_{A_{n-1}} + k_{A_n}.$$

Ezzel az állítást bebizonyítottuk.

3.5. Be kell látnunk, hogy

$$P(A-B) = P(A) - P(AB).$$

Azt tudjuk, hogy $A–B=A–AB$. Ugyanakkor $AB \subset A$, tehát alkalmazható (3.5.5), tehát

$$P(A-B) = P(A-AB) = P(A) - P(AB).$$

3.6. Be kell látni, hogy $P(A+B) \le P(A) + P(B)$. Korábban már beláttuk, hogy

$$P(A+B) = P(A) + P(B) - P(AB).$$

Mivel $P(AB) \ge 0$, ezért

$$P(A+B) \le P(A) + P(B).$$

Az egyenlőség akkor áll fenn, ha $A_nB = \emptyset$, hiszen ekkor $P(AB)=0$.

3.7. Két érme esetén z összes lehetséges esemény halmaza:

$$\Omega = \{ff, ii, if, fi\}.$$

Klasszikus valószínűségi mező lévén innen $P(A = ii) = \frac{1}{4}$.

3.8. A kedvező elemi események száma 2. Az összes elemi esemény száma 10. A klasszikus valószínűség számolása szerint tehát az esemény valószínűsége:

$$P(A) = \frac{2}{10} = \frac{1}{5}.$$

3.9. A feladat a geometriai valószínűséggel oldható meg. Elegendő két vonalat tekintenünk L hosszúságon. Az A esemény legyen az, hogy az érme érinti az egyik vonalat. Tehát az $\overline{A}$ az az esemény, hogy az érme az egyik vonalat sem érinti. Ennek az a feltétele, hogy az érme középpontjának távolsága mindegyik vonaltól nagyobb legyen, mint $d/2$. Az összes esemény valószínűsége arány Lh-val, a kedvező esemény valószínűsége pedig $L(h\text{-}d)$-vel. A geometriai valószínűség szerint tehát,

$$P(\overline{A}) = \frac{L(h-d)}{Lh} = \frac{h-d}{h} = 1 - \frac{d}{h}.$$

Innen kapjuk meg az A esemény valószínűségét a $P(A) = 1 - P(\overline{A})$ összefüggés alapján:

$$P(A) = 1 - (1 - \frac{d}{h}) = \frac{d}{h}.$$

3.10. A feladat teljes indukcióval hasonló módon oldható meg, mint ahogyan azt a 3.4. feladat során tettük.

3.11. Alkalmazva a feltételes valószínűségre vonatkozó 3. tételt:

$$P\left((A + \overline{A}) \middle| B\right) = P(A|B) + P(\overline{A}|B),$$

hiszen $A\overline{A} = \emptyset$. Másrészről viszont

$$P\left((A + \overline{A}) \middle| B\right) = P(\Omega|B) = \frac{P(\Omega B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Alkalmazva ezt a két összefüggést, azt kapjuk, hogy

$$P(\overline{A}|B) = 1 - P(A|B),$$

amit be kellett látni.

3.12. Be kell látni, hogy,

$$P\big((A_1 + A_2)\big|B\big) = P\big(A_1\big|B\big) + P\big(A_2\big|B\big) + P\big(A_1A_2\big|B\big).$$

Ehhez felírjuk a feltételes valószínűség definíciója alapján, hogy

$$P\big((A_1 + A_2)\big|B\big) = \frac{P\big((A_1 + A_2)B\big)}{P(B)} = \frac{P(A_1B + A_2B)}{P(B)} = \frac{P(A_1B) + P(A_2B) - P(A_1A_2B)}{P(B)}.$$

A törteket külön-külön felírva már következik az állítás.

3.13. A feladat megoldásához a következő összefüggéseket használhatjuk fel:

$$(A + \overline{A})\overline{B} = \Omega\overline{B} = \overline{B},$$
$$(A + \overline{A})\overline{B} = A\overline{B} + \overline{A}\overline{B} \text{ és}$$
$$(A\overline{B})(\overline{A}\overline{B}) = \emptyset.$$

Tehát,

$$P\big((A + \overline{A})B\big) = P(A\overline{B}) + P(\overline{A}\overline{B}) = P(\overline{B}),$$

innen pedig

$$P(A\overline{B}) = P(\overline{B}) - P(\overline{A}\overline{B}).$$

Felhasználva a 3.9. példa eredményét:

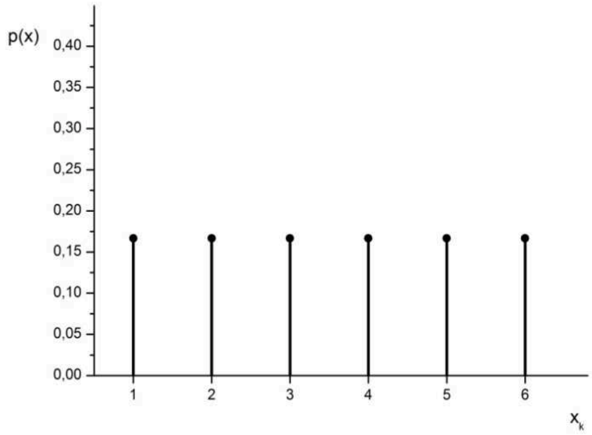
$$P(A\overline{B}) = P(\overline{B}) - P(\overline{A})P(\overline{B}) = P(\overline{B})\big(1 - P(\overline{A})\big) = P(A)P(\overline{B}),$$

ami éppen azt jelenti, hogy A és $\overline{B}$ függetlenek egymástól.

3.14. Ezt a feladatot az előző feladat módszerével kell megoldani.

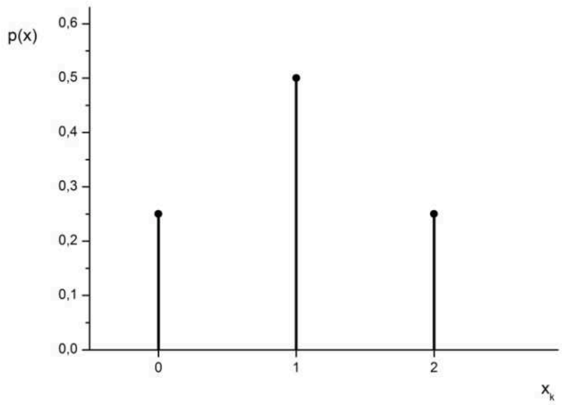
14.4. A 4. fejezethez

4.1.



4.2.

$$P(0)=\frac{1}{4};\ P(1)=\frac{2}{4}=\frac{1}{2};\ P(2)=\frac{1}{4}.$$

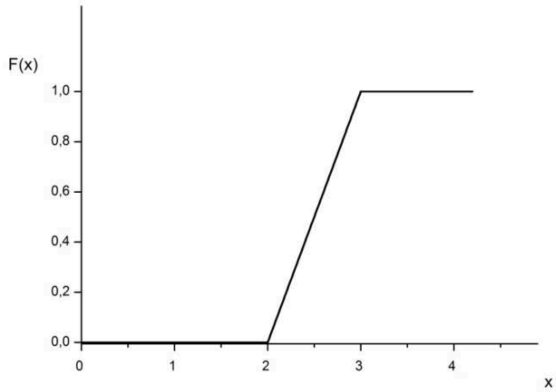


4.3.

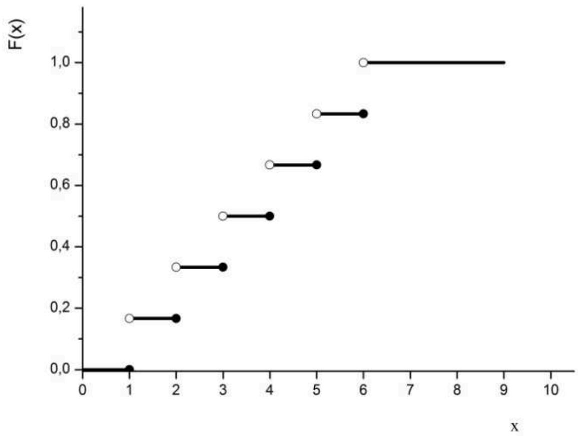
$$P(2\le r\le 5)=P(0\le r\le 5)-P(0\le r\le 2)=\frac{5^2}{10^2}-\frac{2^2}{10^2}=\frac{21}{100}.$$

4.4. A probléma a geometriai valószínűség segítségével oldható meg. Ha feltesszük, hogy a kijelölt pont csak a (*2, 3*) intervallumba eshet, akkor az eloszlásfüggvény alakja:

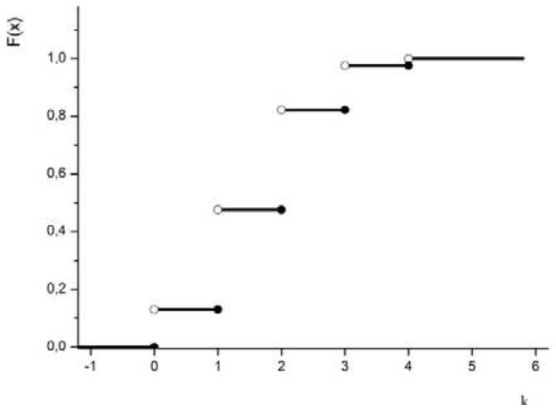
$$F(x)=\begin{cases} 0 & \text{ha } x \le 2 \\ x-2 & \text{ha } 2 < x < 3 \\ 1 & \text{ha } x \ge 3 \end{cases}.$$



4.5.



4.6.



4.7. A normáltság szükséges feltétel ahhoz, hogy az *f(x)* függvény sűrűségfüggvény legyen, tehát

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

vagyis $\lambda = 1$.

4.8.

$$F(x) = 0\text{, ha } x < 0.$$

$$F(x) = \int_{-\infty}^{x} e^{-t}dt = \left[-e^{-t}\right]_0^x = -e^{-x} + 1 = 1 - e^{-x}\text{, ha } x \geq 0.$$

4.9. A karakterisztikus változó lehetséges értékei: $x_1=0$ *és* $x_2=1$. A hozzájuk tartozó valószínűségi értékek: $P(0)=q=p-1$, $P(1)=p$. A definíciója alapján kapjuk a várható értéket:

$$M(\xi) = 0\cdot q + 1\cdot p = p.$$

4.10. A szórásnégyzetet az alábbi képlettel számolhatjuk:

$$D^2(\xi) = M(\xi^2) - M^2(\xi).$$

$$M(\xi^2) = 0^2\cdot(1-p) + 1^2 p = p.$$

Tehát a szórásnégyzet:

$$D^2(\xi) = M(\xi^2) - M^2(\xi) = p - p^2 = p(1-p) = pq.$$

Innen a szórás:

$$D(\xi) = \sqrt{pq}\ .$$

14.5. Az 5. fejezethez

5.1. A feladat lényegében a peremeloszlások meghatározása.

ξ/η	***1***	***0***	***1***	***Σ sor***
-1	*1/8*	*1/12*	*1/24*	*6/24=1/4*
1	*3/8*	*1/4*	*3/24*	*18/24=3/4*
Σ oszlop	*4/8=1/2*	*4/12=1/3*	*4/24=1/6*	*1*

Tehát:

$$P(x=-1)=\frac{1}{4};\quad P(x=1)=\frac{3}{4};$$
$$P(y=-1)=\frac{1}{2};\ P(y=0)=\frac{1}{3},\quad P(y=1)=\frac{1}{6}.$$

5.2. Határozzuk meg először a peremeloszlásokat!

ξ/η	***1***	***0***	***1***	***Σ sor***
-1	*p*	*3p*	*5p*	*9p*
1	*2p*	*4p*	*6p*	*12p*
Σ oszlop	*3p*	*7p*	*11p*	*21p*

Tudjuk, hogy a jobb alsó sarokban lévő cellájában *1*-nek kell állnia, tehát

$21p = 1$, azaz

$$p=\frac{1}{21}.$$

5.3. A függetlenség feltételét az (5.3.1) kifejezés definiálja. Ez azt jelenti, hogy a két valószínűségi változó együttes eloszlását leíró táblázat sorvégi és oszlop végi peremeloszlásainak szorzata meg kell egyezzen a megfelelő sor és oszlop találkozási cellájában lévő elem értékével. Az 5.3. táblázat elemeire ez nem igaz. Például az első oszlop első sorában lévő elem értéke 1/8, ugyanakkor az oszlop alján 5/8, a sor végén 1/8 perem-eloszlásbeli érték szerepel. E két szám szorzata nem egyezik a táblázatbeli 1/8 értékkel. A két valószínűségi változó tehát nem független.

5.4. A két valószínűségi változó független, hiszen az (5.3.1) feltétel a táblázat valamennyi elemére teljesül.

5.5. Az (5.5.4) összefüggés alapján adható meg a keresett valószínűség. Tehát:

$$P(0 \le \xi \le 1; 0 \le \eta \le 1) = F(1,1) - F(0,1) - F(1,0) + F(0,0) = \left(1 - \frac{1}{e}\right)\left(1 - \frac{1}{e}\right),$$

hiszen $F(1,1) = \left(1 - \frac{1}{e}\right)\left(1 - \frac{1}{e}\right)$; $F(0,1) = F(1,0) = F(0,0) = 0$.

5.6. A sűrűségfüggvény az eloszlásfüggvényből számolható az (5.6.1) kifejezés alapján.

$$\frac{\partial F(x,y)}{\partial x} = \left(1 - e^{-y}\right)e^{-x}; \quad f(x,y) = \frac{\partial F(x,y)}{\partial x \partial y} = e^{-x}e^{-y}.$$

5.7. Az előző feladatban megtalált *f(x,y)* sűrűségfüggvényről könnyű belátni, hogy az $x \ge 0$; $y \ge 0$ tartományban $f(x,y) \ge 0$, és

$$\int_0^\infty \int_0^\infty e^{-x}e^{-y} = 1.$$

5.8. Ha $F(x) = 1 - e^{-x}$ és $G(y) = 1 - e^{-y}$, akkor igaz, hogy $F(x,y) = F(x)G(y)$, és ez a két változó függetlenségének feltétele.

5.9. Az 5.5. gyors ellenőrző feladatban definiált eloszlásfüggvény nem független valószínűségi változókat definiál, hiszen

$$f(x) = \frac{\partial F(x)}{\partial x} = e^{-x}; g(y) = \frac{\partial G(y)}{\partial y} = e^{-y}$$

$$f(x,y) = \frac{\partial F(x,y)}{\partial x \partial y} = e^{-x}e^{-y} = f(x)g(y).$$

5.10. Az összeg várható értékét az (5.8.1) képlet alapján a következőképpen számolhatjuk:

$$M(\xi + \eta) = \sum_{i=1}^{3} \sum_{j=1}^{2} (x_i + y_j) r_{ij} = \frac{15}{8}.$$

5.11. A szorzat várható értékét az (5.8.1) képlet alapján az alábbiak szerint számolhatjuk:

$$M(\xi\eta)=\sum_{i=1}^{3}\sum_{j=1}^{2}(x_i y_j)r_{ij}=\frac{3}{8}.$$

5.12. Az (5.9.1) összefüggés szerint számolva:

$$M(\xi)=0\cdot\frac{1}{8}+1\cdot\frac{3}{8}+2\cdot\frac{3}{8}+3\cdot\frac{1}{8}=\frac{12}{8};\quad M(\eta)=0\cdot\frac{5}{8}+1\cdot\frac{3}{8}=\frac{3}{8}.$$

$$M(\xi+\eta)=M(\xi)+M(\eta)=\frac{12}{8}+\frac{3}{8}=\frac{15}{8}.$$

5.13. Az (5.9.1) összefüggés szerint számolva:

$$M(\xi)M(\eta)=\frac{3}{8}\frac{15}{8}=\frac{45}{64}\neq M(\xi\eta).$$

Ez az érték azért nem egyezik a várható érték (5.8.1) képlet alapján számolt értékével, mert nem teljesül a két valószínűségi változó függetlenségének feltétele, ami pedig az (5.9.1) kifejezés alkalmazásának előfeltétele.

14.6. A 7. fejezethez

7.1. Lásd a 4.5. gyors ellenőrző feladat megoldását.

7.2. A várható érték számolása a (7.2.2) képlet alapján:

$$M(\xi)=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{1}{6}(1+2+3+4+5+6)=3{,}5.$$

A szórásnégyzet számolás a (7.2.3) alapján:

$$D^2(\xi)=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2=2{,}916\ldots\ .$$

Innen a szórás:

$$D(\xi)=1{,}70\ldots$$

7.3. Annak valószínűsége, hogy a 7.2. példában leírt helyzetben egyetlen feladat megoldása sem lesz helyes, a Bernoulli-eloszlás alapján:

$$B_p(5,0)=\binom{5}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^5=0,131\ldots$$

7.4. Annak valószínűsége, hogy a 7.2. példában leírt helyzetben legalább egy feladat megoldása helyes:

$$P(A)=1-0,131\ldots=0,868\ldots$$

7.5. A fél perc alatt beérkező beütésszám szórása a Poisson-eloszlás szórása alapján:

$$D(\xi)=\sqrt{5}=2,23\ldots$$

7.6. A rekurziós formula a Poisson-eloszlásra:

$$\frac{P(\xi=k)}{P(\xi=k-1)}=\frac{\frac{\lambda^k}{k!}e^{-\lambda}}{\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}}=\frac{\lambda^k}{\lambda^{k-1}}\frac{(k-1)!}{k!}=\frac{\lambda}{k},$$

tehát

$$P(\xi=k)=\frac{\lambda}{k}P(\xi=k-1).$$

7.7. A táblázatból vesszük az alábbi értékeket:

$$\Phi(1)=0,8413\,;\ \Phi(2)=0,9772\,;\ \Phi(3)=0,9987\,.$$

Innen a (7.8.8) képlet alapján:

$$P(m-\sigma\le m\le m+\sigma)=2\Phi(1)-1=0,6826\,,$$
$$P(m-2\sigma\le m\le m+2\sigma)=2\Phi(2)-1=0,9544\,.$$
$$P(m-3\sigma\le m\le m+3\sigma)=2\Phi(3)-1=0,9974\,.$$

7.8.

$$2\Phi(\lambda)-1=0,99\,,$$
$$\Phi(\lambda)=0,995\,.$$

A standard normális eloszlás táblázatából visszakeresve, és interpolálva, azt kapjuk, hogy

$$\lambda = 2{,}575 .$$

7.9. Az új változó alakja:

$$\eta = \frac{\xi - m}{\sigma} .$$

Innen az új változó lehetséges értékei:

$$y = \frac{x - m}{\sigma} .$$

Az inverz függvény:

$$x = \sigma y + m .$$

A normális eloszlás sűrűségfüggvényét kifejezve az inverz függvénnyel:

$$f\left(\varphi^{-1}(y)\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\sigma y + m - m)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2}} .$$

Az inverz függvény deriváltja y szerint:

$$\left| \frac{d\varphi^{-1}(y)}{dy} \right| = \sigma .$$

Az új változó sűrűségfüggvénye a (4.8.2) kifejezés alapján:

$$g(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2}} \cdot \sigma = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} .$$

Ez pedig nem más, mint a standard normális eloszlás sűrűségfüggvénye.

7.10. Alkalmazzuk a (4.8.2) **Hiba! A hivatkozási forrás nem található.**összefüggést!

$$y = ax + b ,$$

tehát az inverz függvény:

$$x = \frac{y - b}{a} ,$$

és ennek y szerinti deriváltja:

$$\left|\frac{d\varphi^{-1}(y)}{dy}\right| = \left|\frac{1}{a}\right|.$$

Az $N(m,\sigma)$ eloszlás sűrűségfüggvényét kifejezve y-nal:

$$f\left(\varphi^{-1}(y)\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\left(\frac{y-b}{a}-m\right)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-b-am)^2}{2a^2\sigma^2}}.$$

Tehát a (4.8.2) összefüggés alapján az új sűrűségfüggvény:

$$g(y) = \frac{1}{|a|\sigma\sqrt{2\pi}} e^{-\frac{(y-b-m)^2}{2a^2\sigma^2}}.$$

Ez azt jelenti, hogy az új valószínűségi változó szintén normális eloszlású, azonban az eloszlás új paraméterei:

$$m' = am + b\,;\;\; \sigma' = |a|\sigma\,.$$

14.7. A 11. fejezethez

11.1. A első centrális momentum diszkrét valószínűségi változó esetén:

$$\mu_1 = \sum_i \left(x_i - M(\xi)p_i\right) = \sum_i x_i p_i - \sum_i M(\xi)p_i = M(\xi) - M(\xi)\sum_i p_i = 0\,,$$

mivel $\sum_i p_i = 1$.

Hasonló módon belátható az állítás folytonos esetben is.

11.2. A második centrális momentum képlete:

$$\mu_2 = \sum_i \left(x_i - M(\xi)\right)^2 p_i = \sum_i \left(x_i^2 p_i - 2M(\xi)x_i p_i + M^2(\xi)p_i\right) =$$
$$= \sum_i x_i^2 p_i - 2M(\xi)\sum_i x_i p_i + M^2(\xi)\sum_i p_i == \sum_i x_i^2 p_i - M^2(\xi) = \alpha_2 - \alpha_1^2\,.$$

15. TÁBLÁZATOK

15.1. A Poisson-eloszlás táblázatának használata

A Poisson-eloszlás táblázat a

$$P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

képlet alapján a *P(ξ=k)* értékeit adja. A λ értékek a táblázat felső sorában, a k értékek a táblázat első oszlopában találhatók.

15.2. A Poisson-eloszlás táblázata

k \ λ	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0	0,90484	0,81873	0,74082	0,67032	0,60653	0,54881	0,49659	0,44933	0,40657
1	0,09048	0,16375	0,22225	0,26813	0,30327	0,32929	0,34761	0,35946	0,36591
2	0,00452	0,01637	0,03334	0,05363	0,07582	0,09879	0,12166	0,14379	0,16466
3	0,00015	0,00109	0,00333	0,00715	0,01264	0,01976	0,02839	0,03834	0,04940
4	0,00000	0,00005	0,00025	0,00072	0,00158	0,00296	0,00497	0,00767	0,01111
5	0,00000	0,00000	0,00002	0,00006	0,00016	0,00036	0,00070	0,00123	0,00200
6	0,00000	0,00000	0,00000	0,00000	0,00001	0,00004	0,00008	0,00016	0,00030
7	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00001	0,00002	0,00004
8	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000

k \ λ	1	2	3	4	5	6	7	8	9	10
0	0,36788	0,13534	0,04979	0,01832	0,00674	0,00248	0,00091	0,00034	0,00012	0,00005
1	0,36788	0,27067	0,14936	0,07326	0,03369	0,01487	0,00638	0,00268	0,00111	0,00045
2	0,18394	0,27067	0,22404	0,14653	0,08422	0,04462	0,02234	0,01073	0,00500	0,00227
3	0,06131	0,18045	0,22404	0,19537	0,14037	0,08924	0,05213	0,02863	0,01499	0,00757
4	0,01533	0,09022	0,16803	0,19537	0,17547	0,13385	0,09123	0,05725	0,03374	0,01892
5	0,00307	0,03609	0,10082	0,15629	0,17547	0,16062	0,12772	0,09160	0,06073	0,03783
6	0,00051	0,01203	0,05041	0,10420	0,14622	0,16062	0,14900	0,12214	0,09109	0,06306
7	0,00007	0,00344	0,02160	0,05954	0,10444	0,13768	0,14900	0,13959	0,11712	0,09008
8	0,00001	0,00086	0,00810	0,02977	0,06528	0,10326	0,13038	0,13959	0,13176	0,11260
9	0,00000	0,00019	0,00270	0,01323	0,03627	0,06884	0,10140	0,12408	0,13176	0,12511
10	0,00000	0,00004	0,00081	0,00529	0,01813	0,04130	0,07098	0,09926	0,11858	0,12511
11	0,00000	0,00001	0,00022	0,00192	0,00824	0,02253	0,04517	0,07219	0,09702	0,11374
12	0,00000	0,00000	0,00006	0,00064	0,00343	0,01126	0,02635	0,04813	0,07277	0,09478
13	0,00000	0,00000	0,00001	0,00020	0,00132	0,00520	0,01419	0,02962	0,05038	0,07291
14	0,00000	0,00000	0,00000	0,00006	0,00047	0,00223	0,00709	0,01692	0,03238	0,05208
15	0,00000	0,00000	0,00000	0,00002	0,00016	0,00089	0,00331	0,00903	0,01943	0,03472
16	0,00000	0,00000	0,00000	0,00000	0,00005	0,00033	0,00145	0,00451	0,01093	0,02170
17	0,00000	0,00000	0,00000	0,00000	0,00001	0,00012	0,00060	0,00212	0,00579	0,01276
18	0,00000	0,00000	0,00000	0,00000	0,00000	0,00004	0,00023	0,00094	0,00289	0,00709
19	0,00000	0,00000	0,00000	0,00000	0,00000	0,00001	0,00009	0,00040	0,00137	0,00373
20	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00003	0,00016	0,00062	0,00187
21	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00001	0,00006	0,00026	0,00089
22	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00002	0,00011	0,00040
23	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00001	0,00004	0,00018
24	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00002	0,00007
26	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00001

15.3. A standard normális eloszlás táblázat használata

A 15.4. alfejezetben látható táblázat az $N(0,1)$ standard normális eloszlás eloszlásfüggvényének értékeit mutatja a $[0; 3,8]$ tartományban. A táblázatbeli számok az alábbi integrál x helyen felvett értékei:

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt\ .$$

Az $x<0$ értékeket nem kell a táblázatba foglalni, hiszen ezek az értékek a

$$\Phi(-x) = 1 - \Phi(x)$$

kifejezés alapján számolhatjuk.

Az $x>3,8$ esetén már $\Phi(x) = 1$ vehető.

Ha m és σ értékekkel rendelkező normális eloszlás értékét akarjuk kiszámítani az x helyen, akkor a

$$F(x) = \Phi\left(\frac{x-m}{\sigma}\right)$$

kifejezés alapján az $(x-m)/\sigma$ helyen fogjuk ezt az értéket megtalálni.

15.4. A standard normális eloszlás eloszlásfüggvényének táblázata

x	Φ(x)		x	Φ(x)		x	Φ(x)		x	Φ(x)
0,00	0,5000		**0,34**	0,6331		**0,68**	0,7517		**1,02**	0,8461
0,01	0,5040		**0,35**	0,6368		**0,69**	0,7549		**1,03**	0,8485
0,02	0,5080		**0,36**	0,6406		**0,70**	0,7580		**1,04**	0,8508
0,03	0,5120		**0,37**	0,6443		**0,71**	0,7611		**1,05**	0,8531
0,04	0,5160		**0,38**	0,6480		**0,72**	0,7642		**1,06**	0,8554
0,05	0,5199		**0,39**	0,6517		**0,73**	0,7673		**1,07**	0,8577
0,06	0,5239		**0,40**	0,6554		**0,74**	0,7704		**1,08**	0,8599
0,07	0,5279		**0,41**	0,6591		**0,75**	0,7734		**1,09**	0,8621
0,08	0,5319		**0,42**	0,6628		**0,76**	0,7764		**1,10**	0,8643
0,09	0,5359		**0,43**	0,6664		**0,77**	0,7794		**1,11**	0,8665
0,10	0,5398		**0,44**	0,6700		**0,78**	0,7823		**1,12**	0,8686
0,11	0,5438		**0,45**	0,6736		**0,79**	0,7852		**1,13**	0,8708
0,12	0,5478		**0,46**	0,6772		**0,80**	0,7881		**1,14**	0,8729
0,13	0,5517		**0,47**	0,6808		**0,81**	0,7910		**1,15**	0,8749
0,14	0,5557		**0,48**	0,6844		**0,82**	0,7939		**1,16**	0,8770
0,15	0,5596		**0,49**	0,6879		**0,83**	0,7967		**1,17**	0,8790
0,16	0,5636		**0,50**	0,6915		**0,84**	0,7995		**1,18**	0,8810
0,17	0,5675		**0,51**	0,6950		**0,85**	0,8023		**1,19**	0,8830
0,18	0,5714		**0,52**	0,6985		**0,86**	0,8051		**1,20**	0,8849
0,19	0,5753		**0,53**	0,7019		**0,87**	0,8078		**1,21**	0,8869
0,20	0,5793		**0,54**	0,7054		**0,88**	0,8106		**1,22**	0,8888
0,21	0,5832		**0,55**	0,7088		**0,89**	0,8133		**1,23**	0,8907
0,22	0,5871		**0,56**	0,7123		**0,90**	0,8159		**1,24**	0,8925
0,23	0,5910		**0,57**	0,7157		**0,91**	0,8186		**1,25**	0,8944
0,24	0,5948		**0,58**	0,7190		**0,92**	0,8212		**1,26**	0,8962
0,25	0,5987		**0,59**	0,7224		**0,93**	0,8238		**1,27**	0,8980
0,26	0,6026		**0,60**	0,7257		**0,94**	0,8264		**1,28**	0,8997
0,27	0,6064		**0,61**	0,7291		**0,95**	0,8289		**1,29**	0,9015
0,28	0,6103		**0,62**	0,7324		**0,96**	0,8315		**1,30**	0,9032
0,29	0,6141		**0,63**	0,7357		**0,97**	0,8340		**1,31**	0,9049
0,30	0,6179		**0,64**	0,7389		**0,98**	0,8365		**1,32**	0,9066
0,31	0,6217		**0,65**	0,7422		**0,99**	0,8389		**1,33**	0,9082
0,32	0,6255		**0,66**	0,7454		**1,00**	0,8413		**1,34**	0,9099
0,33	0,6293		**0,67**	0,7486		**1,01**	0,8438		**1,35**	0,9115

x	Φ(x)		x	Φ(x)		x	Φ(x)		x	Φ(x)
1,36	0,9131		**1,70**	0,9554		**2,08**	0,9812		**2,76**	0,9971
1,37	0,9147		**1,71**	0,9564		**2,10**	0,9821		**2,78**	0,9973
1,38	0,9162		**1,72**	0,9573		**2,12**	0,9830		**2,80**	0,9974
1,39	0,9177		**1,73**	0,9582		**2,14**	0,9838		**2,82**	0,9976
1,40	0,9192		**1,74**	0,9591		**2,16**	0,9846		**2,84**	0,9977
1,41	0,9207		**1,75**	0,9599		**2,18**	0,9854		**2,86**	0,9979
1,42	0,9222		**1,76**	0,9608		**2,20**	0,9861		**2,88**	0,9980
1,43	0,9236		**1,77**	0,9616		**2,22**	0,9868		**2,90**	0,9981
1,44	0,9251		**1,78**	0,9625		**2,24**	0,9875		**2,92**	0,9982
1,45	0,9265		**1,79**	0,9633		**2,26**	0,9881		**2,94**	0,9984
1,46	0,9279		**1,80**	0,9641		**2,28**	0,9887		**2,96**	0,9985
1,47	0,9292		**1,81**	0,9649		**2,30**	0,9893		**2,98**	0,9986
1,48	0,9306		**1,82**	0,9656		**2,32**	0,9898		**3,00**	0,9987
1,49	0,9319		**1,83**	0,9664		**2,34**	0,9904		**3,05**	0,9989
1,50	0,9332		**1,84**	0,9671		**2,36**	0,9909		**3,10**	0,9990
1,51	0,9345		**1,85**	0,9678		**2,38**	0,9913		**3,15**	0,9992
1,52	0,9357		**1,86**	0,9686		**2,40**	0,9918		**3,20**	0,9993
1,53	0,9370		**1,87**	0,9693		**2,42**	0,9922		**3,25**	0,9994
1,54	0,9382		**1,88**	0,9699		**2,44**	0,9927		**3,30**	0,9995
1,55	0,9394		**1,89**	0,9706		**2,46**	0,9931		**3,35**	0,9996
1,56	0,9406		**1,90**	0,9713		**2,48**	0,9934		**3,40**	0,9997
1,57	0,9418		**1,91**	0,9719		**2,50**	0,9938		**3,45**	0,9997
1,58	0,9429		**1,92**	0,9726		**2,52**	0,9941		**3,50**	0,9998
1,59	0,9441		**1,93**	0,9732		**2,54**	0,9945		**3,55**	0,9998
1,60	0,9452		**1,94**	0,9738		**2,56**	0,9948		**3,60**	0,9998
1,61	0,9463		**1,95**	0,9744		**2,58**	0,9951		**3,65**	0,9999
1,62	0,9474		**1,96**	0,9750		**2,60**	0,9953		**3,70**	0,9999
1,63	0,9484		**1,97**	0,9756		**2,62**	0,9956		**3,75**	0,9999
1,64	0,9495		**1,98**	0,9761		**2,64**	0,9959		**3,80**	0,9999
1,65	0,9505		**1,99**	0,9767		**2,66**	0,9961			
1,66	0,9515		**2,00**	0,9772		**2,68**	0,9963			
1,67	0,9525		**2,02**	0,9783		**2,70**	0,9965			
1,68	0,9535		**2,04**	0,9793		**2,72**	0,9967			
1,69	0,9545		**2,06**	0,9803		**2,74**	0,9969			

15.5. A Student-eloszlás táblázatának használata

A Student-eloszlás táblázata adott p valószínűség értékhez megad egy t_p számot. Ez a szám az n szabadsági fokú Student-eloszlás sűrűségfüggvénye alatti terület $-t_p$ és t_p határait jelöli ki, úgy hogy a görbe alatti terület $1-p$ legyen. A p értékek a legfelső sorban, az n szabadsági fok értékek az első oszlopban találhatók.

Ha $N(m, \sigma)$ eloszlású sokaság és ismeretlen σ esetén konfidencia intervallumot keresünk az m várható értékre, akkor a következőképpen kell eljárni. n elemű minta esetén az $n-1$. sorban az adott p értékhez megkeressük a táblázatbeli t_p értéket. A t_p és a mintaelemekből képezett

$$\left(\overline{x} - t_p \frac{s^*}{\sqrt{n}}, \overline{x} + t_p \frac{s^*}{\sqrt{n}} \right)$$

intervallum $100(1-p)\%$ szintű konfidencia intervallum lesz az m várható érték számára.

Hasonló módon keressük meg a t próba esetén a t_p értéket.

A Student-eloszlás $n \to \infty$ esetén a standard normális eloszláshoz tart. A táblázat utolsó sorában tehát a standard normális eloszláshoz tartozó értékek találhatók. Ha tehát $N(m, \sigma)$ eloszlású sokaság és ismeret σ esetén konfidencia intervallumot keresünk az m várható értékre, akkor itt találjuk meg a konfidencia intervallum kiszámolásához szükséges u_p értéket.

15.6. A Student-eloszlás táblázata

n \ p	0,200	0,100	0,050	0,020	0,010	0,005	0,002	0,001
1	3,078	6,314	12,706	31,820	63,657	127,321	318,309	636,619
2	1,886	2,920	4,303	6,965	9,925	14,089	22,327	31,599
3	1,638	2,353	3,182	4,541	5,841	7,453	10,215	12,924
4	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	1,476	2,015	2,571	3,365	4,032	4,773	5,893	6,869
6	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	1,397	1,860	2,306	2,897	3,355	3,833	4,501	5,041
9	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	1,345	1,761	2,145	2,625	2,977	3,326	3,787	4,140
15	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	1,337	1,746	2,120	2,584	2,921	3,252	3,686	4,015
17	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965
18	1,330	1,734	2,101	2,552	2,878	3,197	3,610	3,922
19	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
20	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,850
21	1,323	1,721	2,080	2,518	2,831	3,135	3,527	3,819
22	1,321	1,717	2,074	2,508	2,819	3,119	3,505	3,792
23	1,319	1,714	2,069	2,500	2,807	3,104	3,485	3,768
24	1,318	1,711	2,064	2,492	2,797	3,090	3,467	3,745
25	1,316	1,708	2,060	2,485	2,787	3,078	3,450	3,725
26	1,315	1,706	2,056	2,479	2,779	3,067	3,435	3,707
27	1,314	1,703	2,052	2,473	2,771	3,057	3,421	3,690
28	1,313	1,701	2,048	2,467	2,763	3,047	3,408	3,674
29	1,311	1,699	2,045	2,462	2,756	3,038	3,396	3,659
30	1,310	1,697	2,042	2,457	2,750	3,030	3,385	3,646
∞	1,282	1,645	1,960	2,326	2,576	2,807	3,090	3,291

15.7. A χ^2-eloszlás táblázatának használata

A χ^2-eloszlás táblázatok általában adott p valószínűség és n szabadsági fok értékekhez a táblázatbeli χ_p^2 értéket rendelik hozzá, amelyre igaz, hogy

$$P\left(\chi^2 > \chi_p^2\right) = p\,,$$

vagy, ami ugyanez

$$P\left(\chi^2 \le \chi_p^2\right) = 1 - p\,.$$

Tehát, ha a statisztikai vizsgálathoz meg akarjuk keresni azt a valószínűséget, hogy

$$P\left(\chi^2 \le \chi_{\frac{p}{2}}^2\right) = \frac{p}{2}\,,$$

azt a táblázatnak az $1–p/2$ értékű oszlopában kell keresni. Tehát ha $p=0{,}01$, akkor a $p/2$-hez tartozó χ_p^2 értéket a $p=1–0{,}005=0{,}995$ oszlopban, és az n szabadsági foknak megfelelő sorban találjuk meg.

Ha olyan értéket kell keresnünk, ami a táblázatban nem található, akkor a táblázatbeli értékek között lineáris interpolációt kell alkalmazni.

15.8. A χ^2-eloszlás táblázata

n \ p	0,995	0,99	0,95	0,9	0,5	0,1	0,05	0,01	0,005
1	0,0000	0,0002	0,0039	0,0158	0,4549	2,7055	3,8415	6,6349	7,8794
2	0,0100	0,0201	0,1026	0,2107	1,3863	4,6052	5,9915	9,2103	10,5966
3	0,0717	0,1148	0,3519	0,5844	2,3660	6,2514	7,8147	11,3449	12,8382
4	0,2070	0,2971	0,7107	1,0636	3,3567	7,7794	9,4877	13,2767	14,8603
5	0,4117	0,5543	1,1455	1,6103	4,3515	9,2364	11,0705	15,0863	16,7496
6	0,6757	0,8721	1,6354	2,2041	5,3481	10,6446	12,5916	16,8119	18,5476
7	0,9893	1,2390	2,1674	2,8331	6,3458	12,0170	14,0671	18,4753	20,2777
8	1,3444	1,6465	2,7326	3,4895	7,3441	13,3616	15,5073	20,0902	21,9550
9	1,7349	2,0879	3,3251	4,1682	8,3428	14,6837	16,9190	21,6660	23,5894
10	2,1559	2,5582	3,9403	4,8652	9,3418	15,9872	18,3070	23,2093	25,1882
11	2,6032	3,0535	4,5748	5,5778	10,3410	17,2750	19,6751	24,7250	26,7569
12	3,0738	3,5706	5,2260	6,3038	11,3403	18,5494	21,0261	26,2170	28,2995
13	3,5650	4,1069	5,8919	7,0415	12,3398	19,8119	22,3620	27,6883	29,8195
14	4,0747	4,6604	6,5706	7,7895	13,3393	21,0641	23,6848	29,1412	31,3194
15	4,6009	5,2294	7,2609	8,5468	14,3389	22,3071	24,9958	30,5779	32,8013
16	5,1422	5,8122	7,9617	9,3122	15,3385	23,5418	26,2962	31,9999	34,2672
17	5,6972	6,4078	8,6718	10,0852	16,3382	24,7690	27,5871	33,4087	35,7185
18	6,2648	7,0149	9,3905	10,8649	17,3379	25,9894	28,8693	34,8053	37,1565
19	6,8440	7,6327	10,1170	11,6509	18,3377	27,2036	30,1435	36,1909	38,5823
20	7,4338	8,2604	10,8508	12,4426	19,3374	28,4120	31,4104	37,5662	39,9969
21	8,0337	8,8972	11,5913	13,2396	20,3372	29,6151	32,6706	38,9322	41,4011
22	8,6427	9,5425	12,3380	14,0415	21,3370	30,8133	33,9244	40,2894	42,7957
23	9,2604	10,1957	13,0905	14,8480	22,3369	32,0069	35,1725	41,6384	44,1813
24	9,8862	10,8564	13,8484	15,6587	23,3367	33,1962	36,4150	42,9798	45,5585
25	10,5197	11,5240	14,6114	16,4734	24,3366	34,3816	37,6525	44,3141	46,9279
26	11,1602	12,1982	15,3792	17,2919	25,3365	35,5632	38,8851	45,6417	48,2899
27	11,8076	12,8785	16,1514	18,1139	26,3363	36,7412	40,1133	46,9629	49,6449
28	12,4613	13,5647	16,9279	18,9392	27,3362	37,9159	41,3371	48,2782	50,9934
29	13,1212	14,2565	17,7084	19,7677	28,3361	39,0875	42,5570	49,5879	52,3356
30	13,7867	14,9535	18,4927	20,5992	29,3360	40,2560	43,7730	50,8922	53,6720

15.9. Az *F*-eloszlás táblázatának használata

A következő alfejezetben a $p=0,01$, $p=0,025$, $p=0,05$ és a $p=0,1$ értékekhez tartozó táblázatokat találjuk meg. A táblázat értékei megadják azt az F_2 értéket, amelyre igaz, hogy

$$P(F > F_2) = p\,.$$

A táblázat első sorában és első oszlopában a szabadsági fokok szerepelnek.

Ha tehát az F próbához adott p-hez meg akarjuk keresni a táblázatbeli értéket, amelyre igaz, hogy

$$P(F > F_2) = \frac{p}{2}\,,$$

akkor megkeressük a $p/2$ értékhez tartozó táblázatot. Ezt követően megnézzük, hogy a nagyobbik empirikus szórásnégyzethez hány mintaelem tartozik. Ha ez n, akkor az $n-1.$ oszlopban keresünk. Ha a kisebbik empirikus szórásnégyzethez k mintaelem tartozik, akkor a $k-1.$ sorban keresünk. Az így megtalált értéket használjuk az F próba összehasonlításában.

15.10. *F*-eloszlás táblázatok

p=0,01

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12
1	4052,18	4999,50	5403,35	5624,58	5763,65	5858,99	5928,36	5981,07	6022,47	6055,85	6106,32
2	98,503	99,000	99,166	99,249	99,299	99,333	99,356	99,374	99,388	99,399	99,416
3	34,116	30,817	29,457	28,710	28,237	27,911	27,672	27,489	27,345	27,229	27,052
4	21,198	18,000	16,694	15,977	15,522	15,207	14,976	14,799	14,659	14,546	14,374
5	16,258	13,274	12,060	11,392	10,967	10,672	10,456	10,289	10,158	10,051	9,888
6	13,745	10,925	9,780	9,148	8,746	8,466	8,260	8,102	7,976	7,874	7,718
7	12,246	9,547	8,451	7,847	7,460	7,191	6,993	6,840	6,719	6,620	6,469
8	11,259	8,649	7,591	7,006	6,632	6,371	6,178	6,029	5,911	5,814	5,667
9	10,561	8,022	6,992	6,422	6,057	5,802	5,613	5,467	5,351	5,257	5,111
10	10,044	7,559	6,552	5,994	5,636	5,386	5,200	5,057	4,942	4,849	4,706
11	9,646	7,206	6,217	5,668	5,316	5,069	4,886	4,744	4,632	4,539	4,397
12	9,330	6,927	5,953	5,412	5,064	4,821	4,640	4,499	4,388	4,296	4,155
13	9,074	6,701	5,739	5,205	4,862	4,620	4,441	4,302	4,191	4,100	3,960
14	8,862	6,515	5,564	5,035	4,695	4,456	4,278	4,140	4,030	3,939	3,800
15	8,683	6,359	5,417	4,893	4,556	4,318	4,142	4,004	3,895	3,805	3,666
16	8,531	6,226	5,292	4,773	4,437	4,202	4,026	3,890	3,780	3,691	3,553
17	8,400	6,112	5,185	4,669	4,336	4,102	3,927	3,791	3,682	3,593	3,455
18	8,285	6,013	5,092	4,579	4,248	4,015	3,841	3,705	3,597	3,508	3,371
19	8,185	5,926	5,010	4,500	4,171	3,939	3,765	3,631	3,523	3,434	3,297
20	8,096	5,849	4,938	4,431	4,103	3,871	3,699	3,564	3,457	3,368	3,231
21	8,017	5,780	4,874	4,369	4,042	3,812	3,640	3,506	3,398	3,310	3,173
22	7,945	5,719	4,817	4,313	3,988	3,758	3,587	3,453	3,346	3,258	3,121
23	7,881	5,664	4,765	4,264	3,939	3,710	3,539	3,406	3,299	3,211	3,074
24	7,823	5,614	4,718	4,218	3,895	3,667	3,496	3,363	3,256	3,168	3,032
25	7,770	5,568	4,675	4,177	3,855	3,627	3,457	3,324	3,217	3,129	2,993
26	7,721	5,526	4,637	4,140	3,818	3,591	3,421	3,288	3,182	3,094	2,958
27	7,677	5,488	4,601	4,106	3,785	3,558	3,388	3,256	3,149	3,062	2,926
28	7,636	5,453	4,568	4,074	3,754	3,528	3,358	3,226	3,120	3,032	2,896
29	7,598	5,420	4,538	4,045	3,725	3,499	3,330	3,198	3,092	3,005	2,868
30	7,562	5,390	4,510	4,018	3,699	3,473	3,304	3,173	3,067	2,979	2,843
40	7,314	5,179	4,313	3,828	3,514	3,291	3,124	2,993	2,888	2,801	2,665
60	7,077	4,977	4,126	3,649	3,339	3,119	2,953	2,823	2,718	2,632	2,496
120	6,851	4,787	3,949	3,480	3,174	2,956	2,792	2,663	2,559	2,472	2,336
∞	6,635	4,605	3,782	3,319	3,017	2,802	2,639	2,511	2,407	2,321	2,185

p=0,025

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12
1	647,79	799,50	864,16	899,58	921,85	937,11	948,22	956,66	963,28	968,63	976,71
2	38,506	39,000	39,166	39,248	39,298	39,332	39,355	39,373	39,387	39,398	39,415
3	17,443	16,044	15,439	15,101	14,885	14,735	14,624	14,540	14,473	14,419	14,337
4	12,218	10,649	9,979	9,605	9,365	9,197	9,074	8,980	8,905	8,844	8,751
5	10,007	8,434	7,764	7,388	7,146	6,978	6,853	6,757	6,681	6,619	6,525
6	8,813	7,260	6,599	6,227	5,988	5,820	5,696	5,600	5,523	5,461	5,366
7	8,073	6,542	5,890	5,523	5,285	5,119	4,995	4,899	4,823	4,761	4,666
8	7,571	6,060	5,416	5,053	4,817	4,652	4,529	4,433	4,357	4,295	4,200
9	7,209	5,715	5,078	4,718	4,484	4,320	4,197	4,102	4,026	3,964	3,868
10	6,937	5,456	4,826	4,468	4,236	4,072	3,950	3,855	3,779	3,717	3,621
11	6,724	5,256	4,630	4,275	4,044	3,881	3,759	3,664	3,588	3,526	3,430
12	6,554	5,096	4,474	4,121	3,891	3,728	3,607	3,512	3,436	3,374	3,277
13	6,414	4,965	4,347	3,996	3,767	3,604	3,483	3,388	3,312	3,250	3,153
14	6,298	4,857	4,242	3,892	3,663	3,501	3,380	3,285	3,209	3,147	3,050
15	6,200	4,765	4,153	3,804	3,576	3,415	3,293	3,199	3,123	3,060	2,963
16	6,115	4,687	4,077	3,729	3,502	3,341	3,219	3,125	3,049	2,986	2,889
17	6,042	4,619	4,011	3,665	3,438	3,277	3,156	3,061	2,985	2,922	2,825
18	5,978	4,560	3,954	3,608	3,382	3,221	3,100	3,005	2,929	2,866	2,769
19	5,922	4,508	3,903	3,559	3,333	3,172	3,051	2,956	2,880	2,817	2,720
20	5,872	4,461	3,859	3,515	3,289	3,128	3,007	2,913	2,837	2,774	2,676
21	5,827	4,420	3,819	3,475	3,250	3,090	2,969	2,874	2,798	2,735	2,637
22	5,786	4,383	3,783	3,440	3,215	3,055	2,934	2,839	2,763	2,700	2,602
23	5,750	4,349	3,751	3,408	3,184	3,023	2,902	2,808	2,731	2,668	2,570
24	5,717	4,319	3,721	3,379	3,155	2,995	2,874	2,779	2,703	2,640	2,541
25	5,686	4,291	3,694	3,353	3,129	2,969	2,848	2,753	2,677	2,614	2,515
26	5,659	4,266	3,670	3,329	3,105	2,945	2,824	2,729	2,653	2,590	2,491
27	5,633	4,242	3,647	3,307	3,083	2,923	2,802	2,707	2,631	2,568	2,469
28	5,610	4,221	3,626	3,286	3,063	2,903	2,782	2,687	2,611	2,547	2,448
29	5,588	4,201	3,607	3,267	3,044	2,884	2,763	2,669	2,592	2,529	2,430
30	5,568	4,182	3,589	3,250	3,027	2,867	2,746	2,651	2,575	2,511	2,412
40	5,424	4,051	3,463	3,126	2,904	2,744	2,624	2,529	2,452	2,388	2,288
60	5,286	3,925	3,343	3,008	2,786	2,627	2,507	2,412	2,334	2,270	2,169
120	5,152	3,805	3,227	2,894	2,674	2,515	2,395	2,299	2,222	2,157	2,055
∞	5,024	3,689	3,116	2,786	2,567	2,408	2,288	2,192	2,114	2,048	1,945

p=0,05

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12
1	161,448	199,500	215,707	224,583	230,162	233,986	236,768	238,883	240,543	241,882	243,906
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385	19,396	19,413
3	10,128	9,552	9,277	9,117	9,014	8,941	8,887	8,845	8,812	8,786	8,745
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	5,912
5	6,608	5,786	5,410	5,192	5,050	4,950	4,876	4,818	4,773	4,735	4,678
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060	4,000
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637	3,575
8	5,318	4,459	4,066	3,838	3,688	3,581	3,501	3,438	3,388	3,347	3,284
9	5,117	4,257	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137	3,073
10	4,965	4,103	3,708	3,478	3,326	3,217	3,136	3,072	3,020	2,978	2,913
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854	2,788
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753	2,687
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714	2,671	2,604
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646	2,602	2,534
15	4,543	3,682	3,287	3,056	2,901	2,791	2,707	2,641	2,588	2,544	2,475
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538	2,494	2,425
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494	2,450	2,381
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456	2,412	2,342
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423	2,378	2,308
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393	2,348	2,278
21	4,325	3,467	3,073	2,840	2,685	2,573	2,488	2,421	2,366	2,321	2,250
22	4,301	3,443	3,049	2,817	2,661	2,549	2,464	2,397	2,342	2,297	2,226
23	4,279	3,422	3,028	2,796	2,640	2,528	2,442	2,375	2,320	2,275	2,204
24	4,260	3,403	3,009	2,776	2,621	2,508	2,423	2,355	2,300	2,255	2,183
25	4,242	3,385	2,991	2,759	2,603	2,490	2,405	2,337	2,282	2,237	2,165
26	4,225	3,369	2,975	2,743	2,587	2,474	2,388	2,321	2,266	2,220	2,148
27	4,210	3,354	2,960	2,728	2,572	2,459	2,373	2,305	2,250	2,204	2,132
28	4,196	3,340	2,947	2,714	2,558	2,445	2,359	2,291	2,236	2,190	2,118
29	4,183	3,328	2,934	2,701	2,545	2,432	2,346	2,278	2,223	2,177	2,105
30	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,211	2,165	2,092
40	4,085	3,232	2,839	2,606	2,450	2,336	2,249	2,180	2,124	2,077	2,004
60	4,001	3,150	2,758	2,525	2,368	2,254	2,167	2,097	2,040	1,993	1,917
120	3,920	3,072	2,680	2,447	2,290	2,175	2,087	2,016	1,959	1,911	1,834
∞	3,842	2,996	2,605	2,372	2,214	2,099	2,010	1,938	1,880	1,831	1,752

p=0,1

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12
1	39,863	49,500	53,593	55,833	57,240	58,204	58,906	59,439	59,858	60,195	60,705
2	8,526	9,000	9,162	9,243	9,293	9,326	9,349	9,367	9,381	9,392	9,408
3	5,538	5,462	5,391	5,343	5,309	5,285	5,266	5,252	5,240	5,230	5,216
4	4,545	4,325	4,191	4,107	4,051	4,010	3,979	3,955	3,936	3,920	3,896
5	4,060	3,780	3,619	3,520	3,453	3,405	3,368	3,339	3,316	3,297	3,268
6	3,776	3,463	3,289	3,181	3,108	3,055	3,014	2,983	2,958	2,937	2,905
7	3,589	3,257	3,074	2,961	2,883	2,827	2,785	2,752	2,725	2,703	2,668
8	3,458	3,113	2,924	2,806	2,726	2,668	2,624	2,589	2,561	2,538	2,502
9	3,360	3,006	2,813	2,693	2,611	2,551	2,505	2,469	2,440	2,416	2,379
10	3,285	2,924	2,728	2,605	2,522	2,461	2,414	2,377	2,347	2,323	2,284
11	3,225	2,860	2,660	2,536	2,451	2,389	2,342	2,304	2,274	2,248	2,209
12	3,177	2,807	2,606	2,480	2,394	2,331	2,283	2,245	2,214	2,188	2,147
13	3,136	2,763	2,560	2,434	2,347	2,283	2,234	2,195	2,164	2,138	2,097
14	3,102	2,726	2,522	2,395	2,307	2,243	2,193	2,154	2,122	2,095	2,054
15	3,073	2,695	2,490	2,361	2,273	2,208	2,158	2,119	2,086	2,059	2,017
16	3,048	2,668	2,462	2,333	2,244	2,178	2,128	2,088	2,055	2,028	1,985
17	3,026	2,645	2,437	2,308	2,218	2,152	2,102	2,061	2,028	2,001	1,958
18	3,007	2,624	2,416	2,286	2,196	2,130	2,079	2,038	2,005	1,977	1,933
19	2,990	2,606	2,397	2,266	2,176	2,109	2,058	2,017	1,984	1,956	1,912
20	2,975	2,589	2,380	2,249	2,158	2,091	2,040	1,999	1,965	1,937	1,892
21	2,961	2,575	2,365	2,233	2,142	2,075	2,023	1,982	1,948	1,920	1,875
22	2,949	2,561	2,351	2,219	2,128	2,061	2,008	1,967	1,933	1,904	1,859
23	2,937	2,549	2,339	2,207	2,115	2,047	1,995	1,953	1,919	1,890	1,845
24	2,927	2,538	2,327	2,195	2,103	2,035	1,983	1,941	1,906	1,877	1,832
25	2,918	2,528	2,317	2,184	2,092	2,024	1,971	1,929	1,895	1,866	1,820
26	2,909	2,519	2,307	2,174	2,082	2,014	1,961	1,919	1,884	1,855	1,809
27	2,901	2,511	2,299	2,165	2,073	2,005	1,952	1,909	1,874	1,845	1,799
28	2,894	2,503	2,291	2,157	2,064	1,996	1,943	1,900	1,865	1,836	1,790
29	2,887	2,495	2,283	2,149	2,057	1,988	1,935	1,892	1,857	1,827	1,781
30	2,881	2,489	2,276	2,142	2,049	1,980	1,927	1,884	1,849	1,819	1,773
40	2,835	2,440	2,226	2,091	1,997	1,927	1,873	1,829	1,793	1,763	1,715
60	2,791	2,393	2,177	2,041	1,946	1,875	1,819	1,775	1,738	1,707	1,657
120	2,748	2,347	2,130	1,992	1,896	1,824	1,767	1,722	1,684	1,652	1,601
∞	2,706	2,303	2,084	1,945	1,847	1,774	1,717	1,670	1,632	1,599	1,546