Hajba Tamás, Harmati István, Környei László, Szalay Krisztina
SZE-MTK, Matematika és Számítástudomány Tanszék

Valószínűség-számítás és matematikai statisztika

2013. május 31.

Műszaki és természettudományos alapismeretek
tananyagainak fejlesztése a mérnökképzésben
Pályázati azonosító: TÁMOP-4.1.2.A/1-11/1-2011-0054

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

IMPRESSZUM

Széchenyi István Egyetem, Műszaki Tudományi Kar, Matematika és Számítástudomány Tanszék
Lektor: Dr. Kovács Gergely;
Dr. Kárász Péter, Óbudai Egyetem Neumann János Informatikai Kar, egyetemi docens

ISBN 978-963-7175-95-4
Kiadó: Széchenyi István Egyetem, Műszaki Tudományi Kar

Támogatás:
Készült a TÁMOP-4.1.2.A/1-11/1-2011-0054 számú, "Műszaki és természettudományos alapismeretek tananyagainak fejlesztése a mérnökképzésben" című projekt keretében.

Kulcsszavak: *valószínűségszámítás, valószínűség-számítás, a valószínűség-számítás axiómái, feltételes valószínűség, diszkrét eloszlás, folytonos eloszlás, valószínűségi vektorváltozó, nagy számok törvényei, központi határeloszlás tétel, matematikai statisztika, hipotézisvizsgálat, statisztikai próba*

Tartalmi összefoglaló: A jegyzet a műszaki alapképzésben (BSC) oktatott Valószínűség-számítás és matematikai statisztika tantárgy anyagát öleli fel. A tananyag feldolgozásához szükséges a középszintű érettségi anyagán túl az Analízis, illetve a Lineáris algebra és többváltozós függvények tárgy ismerete. A jegyzet tagolása, a nagy számú részletesen kidolgozott feladat és önellenőrző kérdés elősegíti a tananyag önálló feldolgozását is.

Technikai megjegyzések a jegyzet használatához.

Ez a tananyag egy *elektronikus jegyzet*.

2013-ban, a megjelenés évében annyira elterjedtek az elektronikus tartalomfogyasztásra alkalmas eszközök, hogy bátran feltételezhetjük: az egyetemisták túlnyomó többsége rendelkezik saját számítógéppel, tablet-géppel vagy elektronikus könyvolvasóval. A tananyag elektronikus formája sok előnnyel rendelkezik a nyomtatotthoz képest:

- Aktív tartalmak: az elektronikus változatban belső kereszthivatkozások, külső linkek, mozgóképek, stb. helyezhetők el. A tartalomjegyzék fejezetszámai, az egyenlet- és ábrasorszámok automatikusan belső linket jelentenek, így biztosítják a kényelmes és gyors belső hivatkozást, de a Szerző tetszőleges helyre tud akár a dokumentum belsejébe, akár egy külső webhelyre mutató linket elhelyezni, ami a szokásos klikkentéssel aktivizálható.
- Rugalmasság: a nyomtatott könyv statikus, míg az elektronikus jegyzet esetében könnyű hibajavításokat, frissítéseket alkalmazni.
- Erőforrás-takarékosság, környezetvédelem: az elektronikus formában való terjesztés sokkal kisebb terhelést jelent a környezetre, mint a nyomtatott. Különösen igaz ez, ha a tananyagban sok a színes ábra.

A használt fájlformátum: *PDF*.

A Portable Document Format az Adobe által kifejlesztett formátum, mely igen széles körben elterjedt. Sok helyről szerezhetünk be programot, mely a PDF fájok olvasására alkalmas. Ezek egy része azonban nem tartalmazza a teljes szabvány minden elemét, ezért speciális tartalmak nem, vagy nem pontosan jelenhetnek meg, ha nem az Adobe olvasóját, az AdobeReader-t használjuk. (Letölthető innen.)

A legtöbb megjelenítőprogram jól fogja kezelni az alapszöveget, ábrákat és linkeket, de gondok lehetnek a speciálisabb funkciókkal, pl. a beágyazott dokumentumok kezelésével, az aktív tesztek, kérdőívek használatával.

A jegyzet *képernyőn való megjelenítésre* lett optimalizálva.

A jelenlegi általánosan elérhető könyvolvasó hardverek mérete és felbontása kisebb, mint a nyomtatott könyveké és a számítógépek monitorai általában fektetett helyzetűek. Ehhez igazítottuk a formátumot arra optimalizálva, hogy fektetett kijelzőn teljes képernyős üzemmódban lehessen olvasni. Ehhez állítottuk be a karaktertípust és -méretet valamint azt is, hogy csak kis margót hagyunk, minél több pixelt biztosítva ezzel a tartalomnak. Azért, hogy teljes képernyős üzemmódban is lehessen navigálni, a margón kis navigáló-ikonokat helyeztünk el, melyek a megszokott módon kezelhetők:

- Lapozás előre és hátra: a függőleges oldalak közepén elhelyezett, nyújtott nyilakkal.
- Címoldalra ugrás: kis házikó szimbólum a bal felső sarokban.
- Vissza és előreugrás a dokumentumban: két kicsi szimbólum a bal felső részen. Ezek nem azonosak a lapozással, hanem a web-böngészők vissza- és előrelépéséhez hasonlóan a hiperlinkeken való navigálást szolgálják.

A jegyzet *segítséget nyújt a tanulás ütemezésében*.

A megtanulandó tanagyag a szokásos fejezet-alfejezet felosztáson túl leckékre való bontást is tartalmaz. A leckék különböző számú alfejezetből állhatnak, de közös bennük, hogy a Szerző megítélés szerint egy lecke „együltő helyben” megtanulható, azaz várhatóan 1–1,5 óra alatt feldolgozható.

A leckék elején rövid leírás található a tárgyalt témakörökről, a szükséges előismeretekről, a végén pedig önellenőrző kérdések, melyek sok esetben a PDF fájlban (AdobeReader-rel) aktív tartalomként jelennek meg feleletkiválasztós teszt, számszerű vagy képletszerű kérdés formájában. Érdemes tehát leckénként haladni a tanulásban, mert ez segít az ütemezés tervezésében illetve a leckevégi ellenőrzések segítenek annak eldöntésében, tovább szabad-e haladni vagy inkább ezt vagy az előző leckéket kell újra elővenni.

Ha a tananyag indokolja, nagyobb egységeket „modulokba” szervezünk és a modulok végén a leckevégi önellenőrzéshez képest komolyabb feladatblokkot találhatunk.

Tartalom

II. MODUL | A valószínűségi változó és jellemzői

III. MODUL | Nevezetes eloszlások

IV. MODUL | Valószínűségi változók viszonyának jellemzése

V. MODUL | Egyenlőtlenségek

VI. MODUL | Határeloszlás tételek

1. Bevezető

A valós életben lépten-nyomon találkozunk olyan eseményekkel, melyek kimenetelét előre nem tudjuk teljes magabiztossággal megmondani, de a mindennapi életünket jelentős mértékben befolyásolják. Bizonytalan, előre pontosan nem meghatározható ideig tart egy vásárlás, nem tudjuk előre, hogy hányan fognak előttünk sorban állni a postán, a munkahelyre való eljutáshoz szükséges idő sem jósolható meg teljes biztonsággal, még ha a megszokott útvonalon haladunk is.

A műszaki alkalmazások területén is bőségesen előfordulnak a fentiekhez hasonló véletlen jelenségek. Ilyen lehet például egy termék élettartama, valamely mérés pontatlansága, egy forgalmi csomóponton adott idő alatt áthaladó járművek száma, helyi tömegközlekedésben résztvevő autóbusz utasainak száma egy adott pillanatban, az autóbusz tényleges menetideje, egy telefonvonal leterheltsége egy adott időszakban, a hibás pixelek száma egy monitoron stb.

A valószínűség-számítás témája véletlen tömegjelenségekre vonatkozó törvényszerűségek megállapítása. Véletlen jelenségne azt nevezzük, aminek a kimenetelét a tekintetbe vett, vagy az ésszerűség határain belül tekintetbe vehető, rendelkezésre álló feltételek még nem határozzák meg egyértelműen. Tömegjelenségen olyan jelenséget értünk, amely nagy számban megy végbe egyszerre, vagy (legalábbis elméletben) tetszőlegesen sokszor megismételhető. Az ezekből levonható törvényszerűségek statisztikai jellegűek, azaz nagy számú végrehajtás során átlagosan érvényes törvények.

A jegyzet használata

A jegyzet hét nagyobb részre, modulra tagolódik, melyek egy-egy nagyobb anyagrészt ölelnek fel. A modulok fejezetekre, azok pedig leckékre tagolódnak. Az egyes leckékben megjelenő tananyag nagyságát úgy alakítottuk ki, hogy annak megértése, feldolgozása másfél-két óra alatt lehetséges legyen. A matematikai témájú egyetemi tankönyvekben megszokott *Definíció*,*Tétel*, *Bizonyítás* hármas mellett a következő tartalmi elemek is megjelenek:

- *Önálló feladat*: A lecke mélyebb megértését elősegítő ellenőrző feladatok, melyek megoldása nem szerepel a jegyzetben.
- *Érdekesség*: A lecke elméleti anyagához kapcsolódó, a valós életből származó példák.
- *Aktivitás*: Önállóan (sokszor a gép mellől felkelve) elvégzendő, általában nem a szokásos matematikai feladatmegoldást igénylő, gyakran kísérletek, megfigyelések elvégzését követelő feladatok.

Minden lecke végén található egy-egy rövid feladatsor (*Ellenőrző kérdések*). A feleletválasztós kérdéseknél a helyesnek gondolt válasznak megfelelő négyzetbe kell kattintani, a válaszadás után látszik, hogy a válasz jó volt-e. A számítási feladatok végeredményét a feladat után szereplő téglapba kell beírni, a magyar szokásoktól eltérően nem tizedesvesszővel, hanem ponttal (ezután a biztonság kedvéért nyomjuk meg az Enter-t is). A helyes válasz az Ans gombra kattintva tekinthető meg.

A nagyobb anyagrészeket felölelő modulokat *Modulzáró feladatsor* zárja. A helyes válaszok itt már nem jelennek meg, csak a teszt kitöltése végén jelenik meg a helyes válaszok száma.

I. MODUL

A valószínűség-számítás alapjai

1. LECKE

Eseményalgebra

2. A valószínűség-számítás axiómái

2.1. Események, műveletek eseményekkel

Kísérlet alatt egy véletlen tömegjelenség megfigyelését értjük. Tekintsünk egy ilyen kísérletet. Ennek egy lehetséges kimenetele az *elemi esemény*. Az egy kísérlethez tartozó elemi események összessége az *eseménytér*, amelyet Ω-val jelölünk. Az eseménytér részhalmazait *esemény*eknek nevezzük. Az esemény definíciója szerint az elemi események egyetlen elemet tartalmazó események. Ha egy A eseményre vonatkozóan kísérletet végzünk, és a kísérlet során adódó a elemi esemény eleme az A-nak ($a \in A$), akkor azt mondjuk, hogy az A esemény bekövetkezik.

Példa: Dobjunk fel egy dobókockát. Ennek a kísérletnek hat lehetséges kimenetele van, így az eseménytér

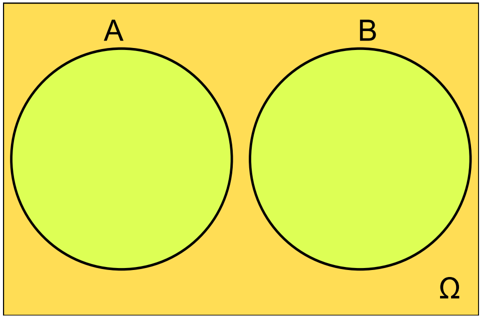
$$\Omega = \{1,2,3,4,5,6\}.$$

Legyen A az az esemény, hogy 7-nél kisebbet dobunk, B az, hogy nyolcast dobunk, C az, hogy 3-nál kisebbet dobunk, D az, hogy kettest dobunk, míg E az, hogy legalább négyest dobunk. Ekkor

$$A = \{1,2,3,4,5,6\},\ B = \emptyset,\ C = \{1,2\},\ D = \{2\},\ E = \{4,5,6\}.$$

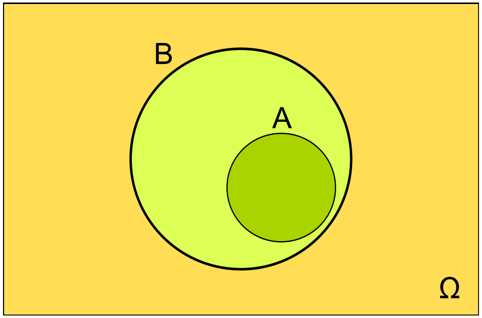
Látszik, hogy az A esemény a kísérlet minden kimenetele esetén bekövetkezik. Az ilyen eseményt biztos eseménynek nevezzük. Mivel a biztos esemény az összes elemi eseményt tartalmazza, ezért a biztos eseményt is Ω-val jelöljük. A B esemény viszont sosem következik be; az ilyen eseményt lehetetlen eseménynek nevezzük és (mivel egyetlen elemi eseményt sem tartalmaz) $\emptyset$-zal jelöljük.
Figyeljük meg továbbá, hogy a D esemény bekövetkezése esetén a C esemény is bekövetkezik, míg a D és az E események nem tudnak egyszerre bekövetkezni (kizárják egymást).

Ha az A és B esemény egyszerre sohasem következik be, akkor az A és B eseményeket egymást kizáró eseményeknek nevezzük. Halmazokkal kifejezve ez azt jelenti, hogy $A \cap B = \emptyset$.
Az előző oldali példában a C (3-nál kisebbet dobunk) és E (legalább 4-est dobunk) egymást kizáró események.



2.1. ábra. *Egymást kizáró események.*

Ha az A esemény bekövetkezésekor minden esetben egy másik, B esemény is bekövetkezik, akkor azt mondjuk, hogy az A esemény bekövetkezése maga után vonja a B esemény bekövetkezését. Halmazokkal kifejezve ez azt jelenti, hogy A részhalmaza B-nek, azaz $A \subset B$.
Az előző oldali példában a C esemény (kettest dobunk) bekövetkezése maga után vonja a D esemény (3-nál kisebbet dobunk) bekövetkezését.



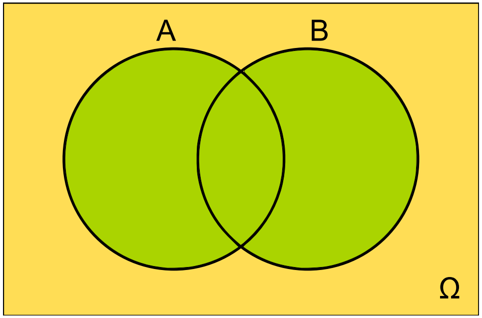
2.2. ábra. *Az A esemény maga után vonja B-t.*

Aktivitás: Írjon fel a füzetébe a kockadobással kapcsolatosan három, egymást páronként kizáró eseményt!

Az események között (a halmazműveletekhez hasonlóan) az alábbi műveleteket értelmezzük:

Az A és B események $A + B$ összege az az esemény, amely akkor következik be, ha az A és B események közül legalább az egyik bekövetkezik.
Példa: Legyen az A esemény, hogy párosat dobunk a dobókockával, a B pedig, hogy legalább ötöst dobunk. Ekkor

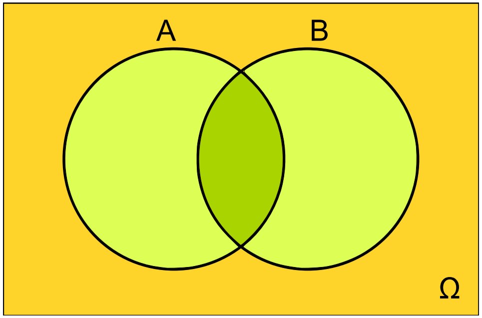
$$A + B = \{2,4,5,6\} .$$



2.3. ábra. *Az A és B események összege.*

Az A és B események $A \cdot B$ szorzata az az esemény, amely akkor következik be, ha A és B is bekövetkezik.
Legyen az A esemény, hogy párosat dobunk a dobókockával, a B pedig, hogy legalább ötöst dobunk. Ekkor

$$A \cdot B = \{6\} .$$



2.4. ábra. *Az A és B események szorzata.*

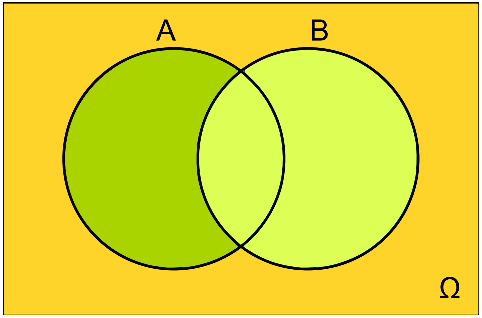
Önellenőrzés

1. Legyen az A esemény, hogy 3-mal osztható számot dobunk, a B pedig, hogy 4-nél kisebbet. Mivel egyenlő $A+B$?

☐ {1,2,3} ☐ {1,2,3,6} ☐ {3} ☐ {6}

Az A és B események $A-B$ különbsége az az esemény, amely akkor következik be, ha az A esemény bekövetkezik, de a B esemény nem.
Példa. Legyen a kockadobásnál az A esemény, hogy négyest vagy hatost dobunk, B pedig, hogy ötnél kisebbet dobunk. Ekkor

$$A-B=\{6\},\ \ B-A=\{1,2,3\}.$$



2.5. ábra. *Az A és B események különbsége.*

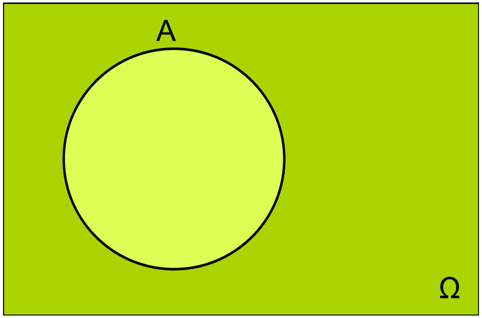
2-1. önálló feladat: Legyen A az az esemény, hogy az ötöslottó húzáson mind az öt kihúzott szám páros, míg B az az esemény, hogy legfeljebb egy páros szám van a kihúzottak között. Írja fel, hogy mit jelentenek az $A-B$, $B-A$, $\overline{A}$, $\overline{B}$ események!

2.1. feladat. Egy irodaházban három lift működik. Jelölje A_1,A_2, illetve A_3 azt az eseményt, hogy egy nap folyamán az első, második, illetve harmadik lift hibamentesen működik. Fogalmazzuk meg, hogy mit jelentenek a következő események:

a) $A_1\cdot A_2\cdot A_3$, *b)* $\overline{A_1}+\overline{A_2}+\overline{A_3}$, *c)* $\overline{A_1}\cdot\overline{A_2}\cdot\overline{A_3}$, *d)* $\overline{A_1\cdot A_2\cdot A_3}$, *e)* $A_1\cdot(\overline{A_2}+\overline{A_3})$, *f)* $\overline{A_1+A_2+A_3}$.

Az A esemény $\overline{A}$ ellentettje (komplementere) az az esemény, amely akkor következik be, ha az A esemény nem következik be. Az A esemény komplementere nem más, mint az Ω eseménytér és az A esemény különbsége: $\overline{A} = \Omega - A$.
Példa. Legyen a kockadobásnál az A esemény, hogy prímszámot dobunk. Ekkor

$$\overline{A} = \{1,4,6\}.$$



2.6. ábra. *Az A esemény komplementere.*

Megoldás:

a) Három esemény szorzatáról van szó, ezért a szorzat esemény azt jelenti, hogy mindhárom lift hiba nélkül működik.

b) Mivel itt összegről van szó, ezért a megadott esemény azt jelenti, hogy vagy az első, vagy a második, vagy a harmadik gép meghibásodik a nap folyamán, amit röviden úgy is mondhatunk, hogy legalább az egyik lift meghibásodik.

c) Az esemény azt jelenti, hogy mindegyik lift elromlik a nap folyamán.

d) Itt éppen az a, rész komplementer eseményéről van szó, azaz annak az ellentettjéről, hogy mindegyik lift jól működik. Ennek a komplementere az, hogy nem mindegyik lift működik jól, amit egyszerűbben úgy is mondhatunk, hogy legalább az egyik meghibásodik. Vegyük észre, hogy ez megegyezik a b-beli eseménnyel! (Vigyázat: gyakori hiba, hogy valaki a *mindegyik lift jól működik* ellentettjének a *mindegyik lift meghibásodik*-ot tartja; ne essünk ebbe a hibába!)

e) A megadott esemény azt jelenti, hogy az első gép jól működik, a másik kettőből viszont legalább az egyik meghibásodik.

f) Az $A_1+A_2+A_3$ esemény azt jelenti, hogy legalább az egyik lift jól működik; így ennek a komplementere az, hogy egyik lift sem működik jól. Vegyük észre, hogy a megadott esemény megegyezik a c-beli eseménnyel!

⇐2.1. feladat

2.2. feladat. Egy dolgozó minden reggel autóval megy a munkahelyére. Útja során három vasúti kereszteződésen kell áthajtania. Jelölje A_i azt az eseményt ($i = 1,2,3$), hogy az i. átjárón várakozás nélkül haladhat át (azaz nem jön a vonat). Fejezzük ki az A_i események segítségével a következő eseményeket:
a) mindhárom átjárónál pirosan villog a szemafor; *b)* a második átjárónál várakoznia kell, a másik kettőnél nem; *c)* legalább egy átjárónál várakoznia kell; *d*) pontosan egy átjárónál kell várakoznia.

Megoldás:

a) A három esemény egyike sem következik be, azaz a mindhárom eseménynek a komplementere következik be, vagyis az $\overline{A_1}$, $\overline{A_2}$, $\overline{A_3}$ szorzatáról van szó. Tehát a *mindhárom átjárónál várni kell* esemény a következő lesz

$$\overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}.$$

b) Az A_2 esemény nem következik be, míg A_1 és A_3 igen, így az A_1, $\overline{A_2}$, A_3 események szorzatát kell venni, vagyis a *második átjárónál várni kell, a másik kettőnél pedig nem* esemény az alábbi lesz:

$$A_1 \cdot \overline{A_2} \cdot A_3.$$

c) Vegyük észre, hogy a *legalább egy átjárónál várakozni kell* esemény nem más, mint a *mindegyik átjárón várakozás nélkül haladhatunk át* esemény komplementere, ezért a következő alakban írható:

$$\overline{A_1 \cdot A_2 \cdot A_3}.$$

d) A *pontosan egy átjárónál kell várnunk* esemény háromféleképpen következhet be: vagy az első átjárónál kell várnunk, és a másik kettőnél nem, vagy a másodiknál kell várnunk és a másik kettőnél nem, vagy a harmadiknál kell várnunk, és a másik kettőnél nem (ezek közül a második eseményt a b részben már felírtuk). Tehát ennek a három eseménynek kell vennünk az összegét, így a *pontosan egy átjárónál kell várnunk* esemény a következő alakban írható:

$$\overline{A_1} \cdot A_2 \cdot A_3 + A_1 \cdot \overline{A_2} \cdot A_3 + A_1 \cdot A_2 \cdot \overline{A_3}.$$

⇐2.2. feladat

Mivel az eseményeket halmazként értelmeztük, ezért az események közötti összeadás, szorzás ugyanolyan tulajdonságokkal rendelkezik, mint a halmazok körében az unió és a metszet, azaz

kommutatív:	$A + B = B + A$	$A \cdot B = B \cdot A$
asszociatív:	$(A + B) + C = A + (B + C)$	$(A \cdot B) \cdot C = A \cdot (B \cdot C)$
disztributív:	$A \cdot (B + C) = A \cdot B + A \cdot C$	$A + B \cdot C = (A + B) \cdot (A + C)$

Egy tetszőleges A esemény, a komplementere ($\overline{A}$), a biztos esemény (Ω) és a lehetetlen esemény ($\emptyset$) között érvényesek a következő egyszerű műveleti tulajdonságok:

$$A + A = A \qquad A \cdot A = A$$
$$A + \Omega = \Omega \qquad A \cdot \Omega = A$$
$$A + \emptyset = A \qquad A \cdot \emptyset = \emptyset$$
$$A + \overline{A} = \Omega \qquad A \cdot \overline{A} = \emptyset$$

Két esemény (A és B) különbsége átalakítható szorzattá:

$$A - B = A \cdot \overline{B}\,.$$

Itt is érvényesek a halmazok körében ismert De Morgan azonosságok:

$$\overline{A + B} = \overline{A} \cdot \overline{B} \qquad \overline{A \cdot B} = \overline{A} + \overline{B}\,.$$

2.3. feladat. Bizonyítsuk be, hogy tetszőleges A, B és C eseményekre teljesül az $(A - B) \cdot C = (A \cdot C) - B$ összefüggés!

Megoldás: 1. módszer:

$$(A - B) \cdot C = (A \cdot \overline{B}) \cdot C = (A \cdot C) \cdot \overline{B} = (A \cdot C) - B.$$

Az 1. és 3. egyenlőségnél az $A - B = A \cdot \overline{B}$ azonosságot, míg a második egyenlőségnél a szorzat kommutativitását és asszociativitását használtuk fel.

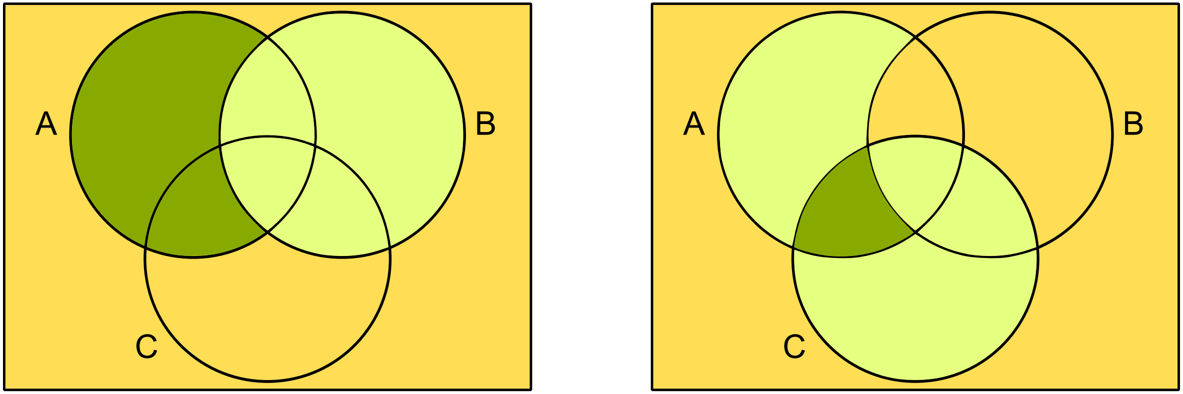
2.módszer: Az állítást Venn-diagramok segítségével szemléletesen is „bizonyíthatjuk".

Ábrázoljuk először a bal oldalt. Az $A - B$ eseménynek az ábrán az A és B halmazok különbsége felel meg; ezt láthatjuk a 2.7. ábra bal oldalán. Az $A - B$ és a C esemény szorzatának az ábrán az $A - B$ és a C halmazok metszete felel meg. Az így kapott, $(A - B) \cdot C$ eseménynek megfelelő halmazt a 2.7. ábra jobb oldali részén láthatjuk.

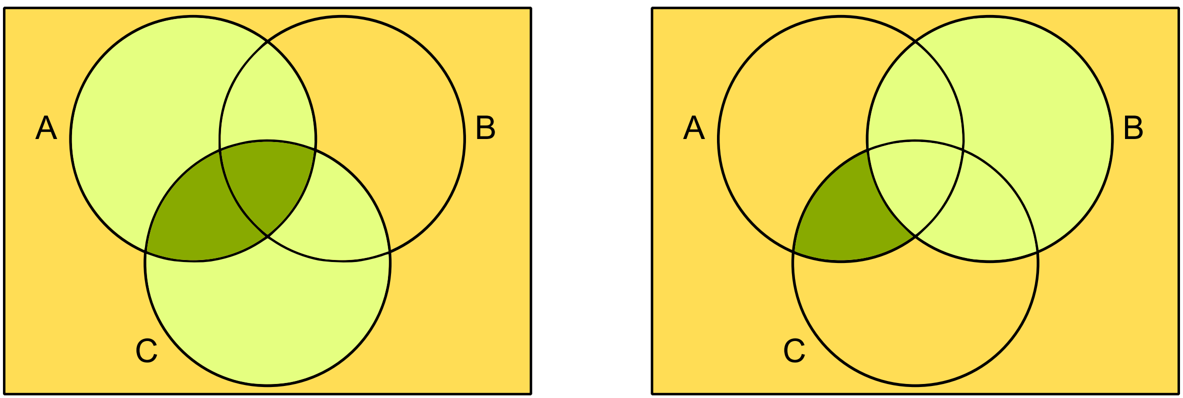
Most ábrázoljuk a bizonyítandó állítás jobb oldalát. Az $A \cdot C$ eseménynek az A és C halmazok metszete felel meg; ezt láthatjuk a 2.8. ábra bal oldalán. Az $(A \cdot C) - B$ eseménynek pedig az $A \cdot C$ és B halmazok különbsége felel meg; ezt láthatjuk a 2.8. ábra jobb oldalán. A két ábra jobb oldalát összehasonlítva valóban azt kapjuk, hogy

$$(A - B) \cdot C = (A \cdot C) - B.$$

⇐2.3. feladat



2.7. ábra. *Bal oldal: az* $A - B$ *esemény; jobb oldal: az* $(A - B) \cdot C$ *esemény (2.3. feladat).*

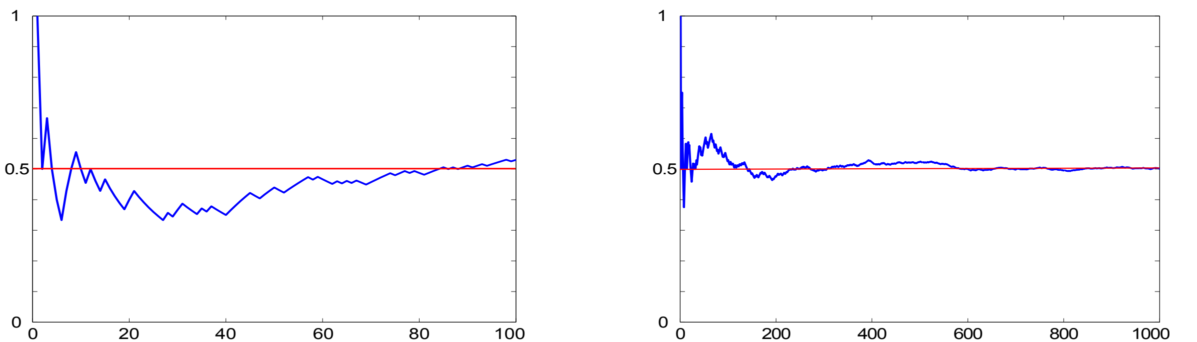


2.8. ábra. *Bal oldal: az* $A \cdot C$ *esemény; jobb oldal: az* $(A \cdot C) - B$ *esemény (2.3 feladat).*

2.2. A valószínűség-számítás axiómái

Ha egy véletlen tömegjelenséget (A) nagyon sokszor, azonos körülmények között megfigyelünk, akkor a tapasztalat szerint a bekövetkezések és az összes kísérletek számának aránya egy meghatározott számérték körül ingadozik, és az ingadozások a kísérletek számának növelésével általában egyre kisebbek lesznek (lásd az 2.9 ábrát). Azt a számot, amely körül ez az arány ingadozik, az esemény valószínűségének nevezzük és $P(A)$-val jelöljük.

2.1. definíció: Ha egy kísérletet n -szer azonos körülmények között megismételve az A esemény k_A esetben következik be, akkor ezt a k_A számot az A esemény gyakoriságának nevezzük. A gyakoriság és a kísérletek számának hányadosát, $\dfrac{k_A}{n}$-et pedig az A esemény relatív gyakoriságának hívjuk.



(a) $n = 100$ (b) $n = 1000$

2.9. ábra. *A fej relatív gyakoriságának időbeni alakulása egy 100 és egy másik, 1000 dobásból álló fej vagy írás kísérletsorozatban.*

Mivel a fentiek szerint az A esemény $\frac{k_A}{n}$ relatív gyakorisága az esemény $P(A)$ valószínűségéhez tart, ezért a relatív gyakoriság tulajdonságaiból következtethetünk a valószínűség tulajdonságaira is.

Ha egy kísérlettel kapcsolatos A esemény gyakorisága k_A , akkor nyilvánvaló, hogy az A esemény relatív gyakorisága 0 és 1 közötti érték, azaz

$$0 \leq \frac{k_A}{n} \leq 1\,.$$

Mivel az A esemény relatív gyakorisága az A esemény valószínűsége körül ingadozik, a

$$0 \leq P(A) \leq 1$$

feltételnek is igaznak kell lennie.

A biztos esemény mindig bekövetkezik, ezért relatív gyakorisága mindig 1, azaz

$$\frac{k_\Omega}{n} = 1\,.$$

Ezért a

$$P(\Omega) = 1$$

egyenlőségnek is teljesülnie kell.

Ha A és B egymást kizáró események, akkor az $A+B$ esemény $(k_A + k_B)$-szer következik be. A relatív gyakoriságokra áttérve innen azt kapjuk, hogy

$$\frac{k_{A+B}}{n} = \frac{k_A + k_B}{n} = \frac{k_A}{n} + \frac{k_B}{n}\,.$$

Mivel az $A+B$ esemény relatív gyakorisága az $A+B$ esemény valószínűsége körül ingadozik, az A, illetve a B esemény relatív gyakoriságai pedig az A, illetve a B esemény valószínűsége körül, ezért egymást kizáró események esetén a

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

egyenlőségnek is fenn kell állnia.

A relatív gyakoriság tulajdonságai alapján célszerű a következőket tekinteni a valószínűség-számítás axiómáinak:

1. Az adott Ω eseménytér minden egyes A eseményéhez tartozik egy 0 és 1 közé eső $P(A)$ szám, azaz

   $$0 \leq P(A) \leq 1,$$

   amelyet az A esemény valószínűségének (valószínűségi mértékének) nevezünk.

2. A biztos esemény valószínűsége 1 , azaz $P(\Omega) = 1$.

3. Az egymást páronként kizáró események összegének valószínűsége az egyes események valószínűségeinek összegével egyenlő, azaz ha az A_1,$A_2, \ldots$,$A_k, \ldots$ események esetén $A_i \cdot A_j = \emptyset$ (ha $i \neq j$), akkor

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

   Tömörebben:

   $$P\left(\sum_k A_k\right) = \sum_k P(A_k) \, .$$

2.3. Az axiómák egyszerűbb következményei

2.1. tétel: Az A esemény komplementerének valószínűsége $1 - P(A)$.

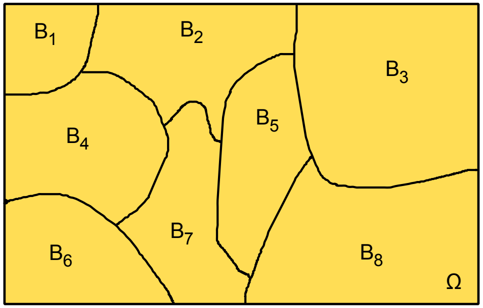
Bizonyítás: Tudjuk azt, hogy $\Omega = A + \overline{A}$, továbbá azt is, hogy A és $\overline{A}$ egymást kizárják ($A \cdot \overline{A} = \emptyset$), tehát alkalmazhatjuk a 3. axiómát:

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

Az első egyenlőségnél azt használtuk, hogy a biztos esemény valószínűsége 1 (2. axióma). Kifejezve a komplementer esemény valószínűségét azt kapjuk, hogy $P(\overline{A}) = 1 - P(A)$. □

2.2. definíció: Az $A_1, A_2, \ldots, A_n$ események teljes eseményrendszert alkotnak, ha egymást páronként kizárják és összegük a biztos esemény, azaz ha $A_i \cdot A_j = \emptyset$ (ha $i \neq j$), és $A_1 + A_2 + \ldots + A_n = \Omega$.

A teljes eseményrendszer tulajdonképpen az eseménytér felbontása olyan diszjunkt részhalmazokra, melyek együttesen lefedik a teljes eseményteret. Az ábrán a B_1, $B_2 \ldots, B_8$ események teljes eseményrendszet alkotnak.



Példa: Felírjuk egy-egy cédulára 1-től 8-ig az egész számokat, és a cédulákat beletesszük egy dobozba. Ezután kihúzunk a dobozból egy cédulát. A következő események teljes eseményrendszert alkotnak:

1. A_1: a kihúzott szám az egyes; A_2: a kihúzott szám nem az egyes. Ekkor A_1 és A_2 teljes eseményrendszert alkot, mert egyrészt kizárják egymást, hiszen egyszerre nem következhetnek be, másrészt a kettő közül az egyik biztosan bekövetkezik, így összegük a biztos esemény.

2. A_1: a kihúzott szám négynél kisebb; A_2: a kihúzott szám a négyes A_3: a kihúzott szám négynél nagyobb. Az A_1, A_2 és A_3 események most is teljes eseményrendszert alkotnak, hiszen egyszerre nem következhet be belőlük kettő (bármelyik kettő kizárja egymást), másrészt viszont valamelyik biztosan be fog közülük következni (összegük a biztos esemény).

Aktivitás: A fenti példában adjon meg legalább 2 további teljes eseményrendszert!

2.2. tétel: Ha az $A_1, A_2, \ldots, A_n$ események teljes eseményrendszert alkotnak, akkor valószínűségeik összege 1, azaz $P(A_1) + P(A_2) + \ldots + P(A_n) = 1$.

Bizonyítás: Mivel a teljes eseményrendszer tagjai egymást páronként kizárják, ezért a 3. axióma miatt az összegük valószínűsége megegyezik a valószínűségeik összegével. Másrészt a teljes eseményrendszer tagjainak összege a biztos esemény, melynek a valószínűsége a 2. axióma miatt 1. Ebből már követezik az állítás. Formálisan:

$$\begin{aligned} P(\Omega) &= P(A_1 + A_2 + \ldots + A_n) = P(A_1) + P(A_2) + \ldots + P(A_n) \\ 1 &= P(A_1) + P(A_2) + \ldots + P(A_n) \end{aligned}$$

□

2.3. tétel: Az A és B események $A - B$ különbségének valószínűsége

$$P(A - B) = P(A) - P(A \cdot B)\,.$$

Bizonyítás: Mivel az $A - B$ és $A \cdot B$ események összege A, továbbá az $A - B$ és $A \cdot B$ események kizárják egymást, ezért alkalmazhatjuk a 3. axiómát.

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

amiből átrendezéssel azt kapjuk, hogy $P(A - B) = P(A) - P(A \cdot B)$. □

Aktivitás: Venn-diagram segítségével győződjön meg róla, hogy az $A - B$ és $A \cdot B$ események egyrészt kizárják egymást, másrészt összegük A-val egyenlő!

A fenti tétel speciális esete a következő:

2.4. tétel: Ha $A \subset B$, akkor $P(B - A) = P(B) - P(A)$.

2.5. tétel: Az A és B események összegének valószínűsége

$$P(A + B) = P(A) + P(B) - P(A \cdot B)\,.$$

Bizonyítás: Az A és $B - A$ egymást kizáró események ($A \cdot (B - A) = \emptyset$), másrészt összegük az $A + B$ esemény. A 3. axióma alkalmazásával azt kapjuk, hogy

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

A $P(B - A)$ tagra alkalmazva a 2.3. tételt kapjuk, hogy

$$P(A + B) = P(A + (B - A)) = P(A) + P(B - A) = P(A) + P(B) - P(A \cdot B).$$

□

Aktivitás: Venn-diagram segítségével győződjön meg róla, hogy az A és $B - A$ események egyrészt kizárják egymást, másrészt összegük $A + B$-vel egyenlő!

2.4. feladat. Legyen az A esemény valószínűsége $0{,}4$, a B esemény valószínűsége $0{,}8$, az együttes bekövetkezés valószínűsége pedig $0{,}3$. Határozzuk meg az alábbi valószínűségeket:

a) $P(A + B)$ *b)* $P(A - B)$ *c)* $P(B - A)$ *d)* $P(\overline{A})$
e) $P(\overline{B})$ *f)* $P(\overline{A} \cdot \overline{B})$ *g)* $P(B - \overline{A})$ *h)* $P(\overline{A} + \overline{B})$.

Megoldás:

a)

$$P(A + B) = P(A) + P(B) - P(A \cdot B) = 0{,}4 + 0{,}8 - 0.3 = 0{,}9\,.$$

b)

$$P(A - B) = P(A) - P(A \cdot B) = 0{,}4 - 0{,}3 = 0{,}1\,.$$

c)

$$P(B - A) = P(B) - P(A \cdot B) = 0{,}8 - 0{,}3 = 0{,}5\,.$$

d)

$$P(\overline{A}) = 1 - P(A) = 1 - 0{,}4 = 0{,}6\,.$$

e)

$$P(\overline{B}) = 1 - P(B) = 1 - 0{,}8 = 0{,}2\,.$$

f) Vegyük észre, hogy a De-Morgan azonosság szerint $\overline{A} \cdot \overline{B} = \overline{A + B}$. Ezt felhasználva

$$P(\overline{A} \cdot \overline{B}) = P(\overline{A + B}) = 1 - P(A + B) = 1 - 0{,}9 = 0{,}1\,.$$

g) Használjuk fel a két esemény különbségére vonatkozó összefüggést!

$$P(B - \overline{A}) = P(B \cdot \overline{\overline{A}}) = P(B \cdot A) = 0{,}3\,.$$

h) A De-Morgan azonosság szerint $\overline{A} + \overline{B} = \overline{A \cdot B}$. Ezt felhasználva

$$P(\overline{A} + \overline{B}) = P(\overline{A \cdot B}) = 1 - P(A \cdot B) = 1 - 0{,}3 = 0{,}7\,.$$

⇐2.4. feladat

Megjegyzés: Az ilyen típusú feladatoknál hasznos lehet, ha Venn-diagrammal ábrázoljuk magunknak az eseményeket. Például az *f* részben rajzoljuk fel először az A esemény komplementerét (ami az A-nak megfelelő halmaz komplementere), utána rajzoljuk fel a B esemény komplementerét, majd vegyük a két esemény összegét, ami a két felrajzolt rész uniójának felel meg. Észrevehetjük, hogy pontosan az $A \cdot B$ rész maradt ki, vagyis azt kaptuk, hogy $\overline{A} + \overline{B} = \overline{A \cdot B}$. Így a feladatot akkor is meg tudjuk oldani, ha nem jut eszünkbe valamelyik azonosság!

2.5. feladat. Az A esemény bekövetkezése maga után vonja a B esemény bekövetkezését. Fejezzük ki az :

a) $P(A + B)$ *b)* $P(A \cdot B)$ *c)* $P(B - A)$ *d)* $P(A - B)$
e) $P(A \cdot \overline{B})$ *f)* $P(\overline{A} \cdot \overline{B})$ *g)* $P(B - \overline{A})$ *h)* $P(A - \overline{B})$.

valószínűségeket a $P(A)$ és $P(B)$ valószínűségek segítségével.

Megoldás: Az, hogy az A esemény bekövetkezése maga után vonja a B esemény bekövetkezését, pontosan azt jelenti, hogy $A \subset B$.

a) Mivel $A \subset B$, ezért $A + B = B$, így $P(A + B) = P(B)$.

b) Mivel $A \subset B$, ezért $A \cdot B = A$, így $P(A \cdot B) = P(A)$.

c) $P(B - A) = P(B) - P(A \cdot B) = P(B) - P(A)\,.$

d) Az $A - B$ esemény azt jelenti, hogy A bekövetkezik, B viszont nem. Mivel jelen esetben $A \subset B$, ezért ez sosem fordulhat elő, így $A - B = \emptyset$, így $P(A - B) = P(\emptyset) = 0$.

e) Mivel $A \subset B$, ezért ha A bekövetkezik, akkor B is bekövetkezik, így $P(A \cdot \overline{B}) = P(\emptyset) = 0$.

f) Mivel $A \subset B$, ezért $\overline{B} \subset \overline{A}$, így $P(\overline{A} \cdot \overline{B}) = P(\overline{B}) = 1 - P(B)$.

g) $P(B - \overline{A}) = P(B \cdot \overline{\overline{A}}) = P(B \cdot A) = P(A)$.

h) $P(A - \overline{B}) = P(A \cdot \overline{\overline{B}}) = P(A \cdot B) = P(A)$.

⇐2.5. feladat

2.6. feladat. Egy üzemben az elkészült termékek a minőség szempontjából első-, másod-, illetve harmadosztályúak lehetnek. Jelentse A azt az eseményt, hogy a raktárból véletlenszerűen kiválasztott termék elsőosztályú, B azt, hogy másodosztályú, C pedig azt, hogy harmadosztályú. Tudjuk,hogy $P(A) = 0{,}3$, míg $P(B) = 0{,}5$. Számoljuk ki a következő valószínűségek értékét!

a) $P(C)$ *b)* $P(A + B)$ *c)* $P(\overline{A} + \overline{B})$

Megoldás:

a) Mivel minőség szempontjából minden termék pontosan az egyik osztályba esik bele, így a három esemény teljes eseményrendszert alkot, ezért $P(A) + P(B) + P(C) = 1$. Ebből a C esemény valószínűségére $1 - 0{,}3 - 0{,}5 = 0{,}2$ adódik.

b) Az A és B egymást kizáró események, ezért $P(A \cdot B) = P(\emptyset) = 0$. Ebből azt kapjuk, hogy $P(A + B) = P(A) + P(B) - P(A \cdot B) = 0{,}3 + 0{,}5 - 0 = 0{,}8$.

c) Az $\overline{A} + \overline{B}$ esemény azt jelenti, hogy egy termék vagy nem elsőosztályú, vagy nem másodosztályú. Vegyük észre, hogy ez minden termékre igaz, vagyis az $\overline{A}$ és a $\overline{B}$ események összege a biztos esemény. Ebből azt kapjuk, hogy $P(\overline{A} + \overline{B}) = P(\Omega) = 1$.

Másik módon is megkaphattuk volna az eredményt. A De Morgan azonosság szerint $\overline{A}+\overline{B}=\overline{A\cdot B}$, amiből $P(\overline{A}+\overline{B})=P(\overline{A\cdot B})=1-P(A\cdot B)$ adódik. Mivel egy termék nem lehet egyszerre elsőosztályú és másodosztályú is, ezért $A\cdot B=\emptyset$. Így $P(\overline{A}+\overline{B})=1-P(\emptyset)=1-0=1$.

⇐2.6. feladat

Önellenőrzés

1. Legyen A, B és C három esemény. Mit jelent az $A \cdot (B + C)$ esemény?
   - [ ] Mindhárom esemény bekövetkezik.
   - [ ] Pontosan két esemény következik be.
   - [ ] Az A esemény bekövetkezik, és a B és C közül is bekövetkezik legalább az egyik.
   - [ ] Legalább az egyik esemény bekövetkezik.
2. Az alábbi események közül melyik fejezi ki azt, hogy az A és B események közül pontosan az egyik következik be?
   - [ ] $A + B$
   - [ ] $\overline{A + B}$
   - [ ] $A \cdot B$
   - [ ] $A \cdot \overline{B} + \overline{A} \cdot B$
3. Az alábbi események közül melyik fejezi ki azt, hogy az A, B, C események közül egyik sem következik be?
   - [ ] $\overline{A} + \overline{B} + \overline{C}$
   - [ ] $\overline{A + B + C}$
   - [ ] $\overline{A \cdot B \cdot C}$
   - [ ] $\overline{A} + \overline{B \cdot C}$
4. Ha $P(A) = 0{,}7$, $P(B) = 0{,}5$ és $P(A \cdot B) = 0{,}4$, akkor mennyi $P(A + \overline{B})$?

   Megold.
5. Tudjuk, hogy a B esemény maga után vonja az A eseményt, és $P(A) = 0{,}8$, $P(B) = 0{,}5$. Mennyi $P(A \cdot \overline{B})$ valószínűség értéke?

   Megold.
6. Az A, B és C események teljes eseményrendszert alkotnak. Tudjuk, hogy $P(A) = 0{,}1$ és $P(B) = 0{,}5$. Mennyi a $P(\overline{A + B})$ valószínűség értéke?

   Megold.

2. LECKE

A klasszikus valószínűségi mező

3. A klasszikus és a geometriai valószínűségi mező

3.1. A klasszikus valószínűségi mező

3.1. definíció: Ha egy kísérlettel kapcsolatban az elemi események száma véges (n), és minden elemi esemény valószínűsége egyenlő $\left(\frac{1}{n}\right)$, akkor a k féleképpen bekövetkező A esemény valószínűsége

$$P(A) = \frac{\text{kedvező esetek száma}}{\text{összes eset száma}} = \frac{k}{n}.$$

Ebben az esetben azt mondjuk, hogy az események és ezek valószínűségei klasszikus valószínűségi mezőt alkotnak.

Mint majd látni fogjuk, a kedvező, illetve az összes eset száma a legtöbb esetben kombinatorikus úton határozható meg a legegyszerűbben, ezért először példák segítségével áttekintjük a kombinatorika alapeseteit.

3.1.1. Kombinatorika

3.1. feladat. Az 1,2,3,4,5 számjegyek felhasználásával hány olyan ötjegyű szám készíthető, melyben minden számjegy csak egyszer szerepel?

Megoldás: Az első helyre még bármelyik számjegyet írhatjuk. Bámit is írunk az első helyre, utána a második helyre már csak 4, a harmadik helyre már csak 3, a negyedik helyre 2, az ötödikre pedig egyetlen lehetőségünk marad. Így összesen $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5! = 120$ szám készíthető.

⇐3.1. feladat

Teljesen hasonló gondolatmenettel belátható, hogy n különböző elem összes lehetséges sorrendjének (az n elem permutációinak) a száma:

$$n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1 = n!.$$

3.2. feladat. Hány különböző, nem feltétlenül értelmes tízbetűs szó (úgynevezett anagramma) készíthető a MATEMATIKA szó betűiből?

Megoldás: A feladat tehát az, hogy három A, két M, két T, és egy-egy E,I, illetve K betű felhasználásával hány tízbetűs szó készíthető. Első nekifutásra különböztessük meg egymástól az azonos betűket (például legyen egy piros, egy kék, és egy sárga A betű, stb). Ekkor a 10 betűből $10!$ szó készíthető. Így azonban minden egyes szót többször is megszámoltunk: az A betűk miatt minden szót $3!$-szor számoltunk meg, a T betűk miatt mindent $2!$-szor, míg az M betűk miatt mindent $2!$-szor. Ezért az összes különböző 10 betűs anagramma száma:

$$\frac{10!}{3! \cdot 2! \cdot 2!} = 151200.$$

⇐3.2. feladat

Ugyanígy látható be, hogy ha adott n elem, melyek között vannak megegyezőek is, nevezetesen az 1. típusú elemből n_1 , a második típusúból $n_2, \ldots$ a k. típusúból n_k darab van ($n_1 + n_2 + \ldots n_k = n$), akkor az n elem összes lehetséges sorrendjének (az ismétléses permutációinak) a száma:

$$\frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}.$$

3.3. feladat. Hány háromjegyű szám készíthető az 1,2,3,4,5, számjegyekből, ha minden számjegy legfeljebb egyszer szerepelhet?

Megoldás: Az első helyre még öt számjegy közül választhatunk. Ha az első számjegyet leírjuk, akkor a második helyre már csak négy lehetőségünk marad, végül az utolsó helyre már csak három számjegy közül választhatunk. Így összesen $5 \cdot 4 \cdot 3$ szám készíthető a megadott módon.

⇐3.3. feladat

Hasonlóan belátható, hogy az n elemből készíthető olyan k hosszú sorozatok száma ($k \leq n$), melyben egy elem legfeljebb egyszer szerepelhet:

$$n \cdot (n-1) \cdot \ldots \cdot (n-k+1).$$

3.4. feladat. Feldobunk nyolcszor egy dobókockát, és a kapott számokat egymás után írjuk. Hány különböző nyolcjegyű számot kaphatunk így?

Megoldás: Mivel mind a nyolc számjegy hatféle lehet, ezért $6^8 = 1679616$ féle nyolcjegyű számot kaphatunk.

⇐3.4. feladat

Hasonlóan belátható, hogy az n elemből készíthető olyan k hosszú sorozatok száma, melyben egy elem többször is szerepelhet:

$$n^k.$$

3.5. feladat. Az ötöslottón 90 számból sorsolnak ki ötöt. Hány különböző végeredménye lehet a sorsolásnak?

Megoldás: Az első húzásnál még 90 szám közül választhatunk, a másodiknál már csak 89 közül, a harmadiknál 88 közül, a negyediknél 87, míg az utolsónál 86 közül; ez összesen $90 \cdot 89 \cdot 88 \cdot 87 \cdot 86$ lehetőség. Igen ám, de így a húzás sorrendjét is figyelembe vettük, ezért minden egyes lehetőséget $5!$-szor megszámoltunk (hiszen 5 számot ennyiféleképpen lehet sorbarendezni). Ezért a különböző végeredmények száma:

$$\frac{90 \cdot 89 \cdot 88 \cdot 87 \cdot 86}{5!} = 43\,949\,268.$$

⇐3.5. feladat

Vezessük be a következő jelölést: legyen $0 \leq k \leq n$, és legyen $\binom{n}{k}$ (olvasd: „n alatt a k") azzal a számmal egyenlő, ahányféleképpen n különböző elemből ki lehet választani k darabot (az előző feladat megoldása tehát $\binom{90}{5}$). Az előző gondolatmenettel belátható, hogy n különböző tárgyból k darabot

$$\binom{n}{k} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k!}$$

féleképpen lehet kiválasztani.

3.6. feladat. : Három fajta fagyiból (eper, vanília, csoki) hányféleképpen állíthatunk össze egy négygombócos sétálókelyhet (egy fajtából természetesen többet is vehetünk)?

Megoldás: Minden sétálókehelyhez rendeljünk hozzá egy egyesekből és nullákból álló sorozatot a következő módon: először írjunk egy 1-est (ez felel meg az epernek), majd írjunk utána annyi nullát, ahány gombócot veszünk az eperből; ezután ismét írjunk le egy 1-est (ez felel meg a vaníliának), majd írjunk utána annyi 0-t, ahány gombócot veszünk a vaníliából; aztán megint írjunk egy 1-est (csoki), és utána annyi 0-t, amennyi gombócot a csokiból veszünk. Például egy 2 eperből és 2 csokiból álló kehelynek az 1001100 sorozat felel meg. Egy ilyen sorozatban három darab 1-es lesz (hiszen 3 fajta fagyi van), négy darab 0 lesz benne (hiszen 4 gombócot veszünk), és 1-essel fog kezdődni. Viszont minden ilyen tulajdonságú sorozatnak (amely 1-essel kezdődik, három darab 1-es és négy darab 0 van benne) egyértelműen meg tudunk feleltetni egy sétálókelyhet. Ez azt jelenti, hogy a különböző sétálókelyhek száma megegyezik az előbb felsorolt tulajdonságú sorozatok számával. Tehát azt kell megmondani, hogy hány olyan hét tagból álló, nullákat és egyeseket tartalmazó sorozat van, mely 1-essel kezdődik, és négy darab 0 van benne. Ezeknek a száma viszont $\binom{6}{4}$ (hiszen a négy darab 0-t hat hely valamelyikére kell elhelyeznünk), így a különböző sétálókelyhek száma $\binom{6}{4} = 15$. ⇐3.6. feladat

Ugyanezen gondolatmenet segítségével belátható, hogy ha n különböző elem közül k darabot akarunk kiválasztani úgy, hogy egy elemet többször is választhatunk, akkor a lehetséges kiválasztások száma

$$\binom{n+k-1}{k}.$$

A kombinatorikai alapesetek átnézése után lássunk most néhány klasszikus valószínűségi mezővel kapcsolatos feladatot.

3.7. feladat. Bankkártyánk PIN-kódjának egy valódi négyjegyű számot kell megadnunk. Véletlenszerűen választjuk ki a számot. Mi a valószínűsége, hogy

a) a kódunk mind a négy számjegye különböző;

b) a kódunkban van 1-es számjegy;

c) a kódunkban pontosan egy darab 6-os számjegy van?

Megoldás:

a) A valódi négyjegyű számok száma $9 \cdot 10^3$ (hiszen az első helyre 9, a másik három helyre 10 számjegy közül választhatunk), így az összes eset száma 9000. A jó esetek száma $9 \cdot 9 \cdot 8 \cdot 7$, hiszen az első helyre kilenc számjegy közül választhatunk, ezután a második helynél szintén kilenc (nem írhatjuk ugyanazt, amit az első helyre) lehetőségünk van, a harmadik helynél már csak 8, míg a negyediknél 7. Így a keresett valószínűség $\frac{9 \cdot 9 \cdot 8 \cdot 7}{9 \cdot 10^3} = 0{,}504$.

b) Ebben az esetben egyszerűbb a komplementer eseménynek kiszámolni a valószínűségét. A *kódban van 1-es* esemény komplementere a *kódban nincs 1-es* esemény. Mivel $8 \cdot 9 \cdot 9 \cdot 9$ olyan valódi négyjegyű szám van, melyben nincs egyes, így

$$P(\text{a kódban van 1-es}) = 1 - P(\text{a kódban nincs 1-es}) = 1 - \frac{8 \cdot 9 \cdot 9 \cdot 9}{9 \cdot 10^3} = 0{,}352.$$

c) Az, hogy a kódban pontosan egy darab hatos van, négyféleképpen valósulhat meg: vagy az első, vagy a második, vagy a harmadik, vagy a negyedik helyen van 6-os, a többi helyen pedig nem 6-os van. Így a jó esetek száma $1 \cdot 9 \cdot 9 \cdot 9 + 8 \cdot 1 \cdot 9 \cdot 9 + 8 \cdot 9 \cdot 1 \cdot 9 + 8 \cdot 9 \cdot 9 \cdot 1 = 2025$. Ebből a kérdéses valószínűség: $\frac{2025}{9000} = 0{,}225$.

⇐3.7. feladat

3.8. feladat. Egy rejtvénypályázatra 30 férfi és 20 nő küldött be jó megoldást. A helyes megfejtők között 3 díjat sorsolnak ki: egy TV-t, egy telefont és egy porszívót. Mindenki legfeljebb egy dolgot nyerhet. Mi a valószínűsége, hogy

a) csak női nyertes lesz; *b)* Kovács út nyer valamit?

Megoldás:

a) A sorsolás mindegyik lehetséges kimeneteléhez rendeljünk hozzá egy három tagból álló sorozatot: az első tag legyen a TV-t nyerő személy, a második a telefont nyerő személy, míg a harmadik a porszívót nyerő személy. Az összes eset száma megegyezik az 50 pályázóból képezhető, 3 tagból álló sorozatok számával, ezek száma pedig $50 \cdot 49 \cdot 48$. A jó esetek száma, tehát amikor csak női nyertes lesz $20 \cdot 19 \cdot 18$, így a kérdéses valószínűség $\frac{20 \cdot 19 \cdot 18}{50 \cdot 49 \cdot 48} \approx 0{,}058$.

b) Kétféle módon is eljárhatunk. 1. módszer: Az, hogy Kovács úr nyer valamit, háromféleképpen valósulhat meg: vagy a TV-t, vagy a telefont, vagy a porszívót nyeri meg. Számoljuk meg külön-külön a három lehetőséghez tartozó eseteket, majd ezek összegeként kapjuk a jó esetek számát. Ha Kovács nyeri a TV-t, akkor a telefont 49-en, míg a porszívót 48-an nyerhetik meg, így azon esetek száma, amikor Kovács nyeri a TV-t $1 \cdot 49 \cdot 48$. Hasonlóan számolhatjuk meg azon esetek számát, amikor Kovács úr telefont, illetve porszívót nyer. Ebből a kérdezett valószínűség:

$$P(\text{Kovács úr nyer valamit}) = \frac{1 \cdot 49 \cdot 48 + 49 \cdot 1 \cdot 48 + 49 \cdot 48 \cdot 1}{50 \cdot 49 \cdot 48} = 0{,}06.$$

2. módszer: Használjuk fel a $P(A) = 1 - P(\overline{A})$ összefüggést. Azon esetek száma, amikor Kovács Úr nem nyer semmit $49 \cdot 48 \cdot 47$, hiszen a TV-t 49, ezután a telefont 48, míg a porszívót 47 embernek adhatjuk oda.

$$P(\text{Kovács úr nyer valamit}) = 1 - P(\text{Kovács úr nem nyer semmit}) = 1 - \frac{49 \cdot 48 \cdot 47}{50 \cdot 49 \cdot 48} = 1 - 0{,}94 = 0{,}06.$$

⇐3.8. feladat

Aktivitás: Számolja ki az előző feladatbeli valószínűségeket abban az esetben is, ha egy ember több díjat is nyerhet!

3.1.2. Visszatevéses és visszatevés nélküli mintavétel

A különböző mintavételi eljárások számos területen játszanak fontos szerepet. Tegyük fel például, hogy minőségellenőrzés során egy cég meg akarja vizsgálni, hogy az aznap gyártott termékek hány százaléka hibás. A legtöbb esetben nincs lehetőség az összes termék megvizsgálására, mert az túl sok pénzbe és időbe kerülne. Ezért egy szokásos eljárás az, hogy véletlenszerűen kiválasztanak néhány terméket (mintát vesznek), és a bennük található hibás termékek számából „következtetnek” az összesben található hibások számára. A közvéleménykutatások hasonlóképpen működnek: egy párt támogatottságának megállapításához nem kérdezik meg az ország összes lakosát, hanem csak néhány kiválasztott embert (egy reprezentatív mintát), és az ő válaszaikból vonnak le az ország teljes lakosságára vonatkozó következtetést.

Aktivitás: Nézzen utána az interneten, hogy Magyarországon a közvéleménykutatásokkor körülbelül hány embert szoktak megkérdezni!

Visszatevéses mintavétel

Visszatevéses mintavétel esetén a minta elemeit egyesével választjuk ki, majd a vizsgálat után visszatesszük őket, és ezután vesszük a következőt, stb.

3.9. feladat. 100 termék között 20 selejtes van. Visszatevéses mintavétellel kiválasztunk 6 terméket. Mi a valószínűsége, hogy

a) az első 4 termék selejtes, a többi jó;

b) az első 2 termék jó, a többi selejtes;

c) pontosan 4 selejtes van a kiválasztott termékek közt?

Megoldás:

a) Mivel minden egyes húzásnál 100 termék közül választhatunk, így az összes eset száma 100^6. A jó eseteknél az első 4 húzásnál 20, az utolsó 2 húzásnál 80 termék közül választhatunk, így a jó esetek száma $20^4 \cdot 80^2$. Így a kérdezett valószínűség:

$$P(\text{első 4 selejt, utolsó 2 jó}) = \frac{\text{jó esetek száma}}{\text{összes eset száma}} = \frac{20^4 \cdot 80^2}{100^6} = \left(\frac{20}{100}\right)^4 \cdot \left(\frac{80}{100}\right)^2 = 0{,}001024.$$

b) A jó esetek száma $80^2 \cdot 20^4$, így a kérdéses valószínűség:

$$P(\text{első 2 jó, utolsó 4 selejt}) = \frac{\text{jó esetek száma}}{\text{összes eset száma}} = \frac{80^2 \cdot 20^4}{100^6} = \left(\frac{80}{100}\right)^2 \cdot \left(\frac{20}{100}\right)^4 = 0{,}001024.$$

c) A kihúzottak között pontosan 4 selejtes esemény felbontható egymást kizáró események összegére: (az első négy termék selejtes, a többi jó)+(az 1.,2.,3.,5. termék selejtes, a többi jó)+.... Az előző két rész alapján világos, hogy a felbontásban szereplő mindegyik esemény valószínűsége $\left(\frac{80}{100}\right)^2 \cdot \left(\frac{20}{100}\right)^4$. Már csak az a kérdés, hogy hány esemény van a felbontásban? Éppen annyi, ahányféleképpen kiválaszthatjuk annak a 4 húzásnak a sorszámát, amikor selejteset akarunk húzni; ezek száma $\binom{6}{4}$. Felhasználva, hogy egymást kizáró események összegének a valószínűsége a valószínűségek összege, azt kapjuk, hogy

$$P(\text{pontosan 4 selejtes van}) = \binom{6}{4}\left(\frac{80}{100}\right)^2 \cdot \left(\frac{20}{100}\right)^4 = 0{,}01536.$$

⇐3.9. feladat

A 3.9. feladat eredményét könnyen általánosíthatjuk: Legyen adott N termék, melyek között a selejtesek száma s. Visszatevéses mintavétellel veszünk egy n elemű mintát. Az egyszerűség kedvéért a selejtesek arányát jelöljük p-vel (azaz $p = \frac{s}{N}$). Ekkor a megoldás során használt módon belátható, hogy

$$P(\text{a mintában } k \text{ selejtes van}) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}.$$

Jó tanács: A képlet „bemagolása" helyett próbáljuk inkább azt megérteni és megjegyezni, hogy hogyan „jön ki". Ha ez sikerül, akkor mindig fel tudjuk majd idézni a formulát.

Visszatevés nélküli mintavétel

A visszatevés nélküli mintavételnél az elemeket kiválaszthatjuk egyszerre (nem számít a sorrend), vagy egyesével is, ügyelve arra, hogy a már kiválasztott elemeket ne tegyük vissza (ebben az esetben számít a sorrend). Oldjuk most meg a 3.9. feladatot visszatevés nélküli mintavétellel.

Adott tehát 100 termék ,melyek közül 20 selejtes. Visszatevés nélküli mintavétellel kiválasztunk 6 terméket. Azt akarjuk meghatározni, hogy mennyi annak a valószínűsége, hogy 4 selejtes van a kiválasztott termékek közt.

Tegyük fel először, hogy a termékeket egyszerre választjuk ki (a sorrend nem számít). Ekkor az összes eset száma $\binom{100}{6}$, hiszen ennyiféleképpen lehet 100 termékből 6-ot kiválasztani. Most nézzük a jó esetek számát: a kiválasztott termékek közt akkor lesz 4 selejtes, ha a selejtesekből 4-et, a jó termékekből pedig 2-t választunk ki, így a jó esetek száma $\binom{20}{4} \cdot \binom{80}{2}$. Ebből azt kapjuk, hogy

$$P(4 \text{ selejtes termék}) = \frac{\binom{20}{4} \cdot \binom{80}{2}}{\binom{100}{6}} \approx 0{,}0107.$$

Gondoljuk meg, hogy mi változik, ha az elemeket egyesével választjuk ki (vagyis a sorrend is számít). Tegyük fel például, hogy a mintavételnél az $1,2,3,4,5,6$ termékeket választottuk ki. Ha a kiválasztás sorrendje nem számít, akkor ez egy esetnek felel meg, míg ha sorrend is számít, akkor $6!$ -féleképpen választhattuk ki ezt a hat terméket. Általában is elmondható, hogy minden egyes, a kiválasztás sorrendjét nem számító esetnek $6!$ eset felel meg, ha a sorrendet is figyelembe vesszük. Ez azt jelenti, hogy a sorrendet figyelembe vévé a jó esetek száma $6!$-szorosa lesz a sorrendet nem számoló esetek jó esetek számának, míg a sorrendet figyelembe véve az összes eset száma $6!$-szorosa lesz a sorrendet nem számoló összes eset számának. Vagyis mind a jó esetek, mind az összes eset száma a $6!$-szorosára nő, így a hányadosuk nem változik, vagyis a kérdéses valószínűség marad ugyanaz.

Fogalmazzuk most meg az eredményeinket általánosan is: Ha N számú termékből s számú selejtes, akkor visszatevés nélküli mintavétellel kiválasztva a termékek közül n darabot, annak a valószínűsége, hogy a kiválasztott termékek közt k selejtes van

$$P(k \text{ selejtes}) = \frac{\binom{s}{k} \cdot \binom{N-s}{n-k}}{\binom{N}{n}},$$

és a valószínűség nem függ attól, hogy a termékeket egyszerre, vagy pedig egyesével választjuk ki. Természetesen a képlet bemagolása helyett inkább a képlet „logikájának” a megértése javallott!

3.10. feladat. Egy színházi előadás szünetében a jelenlevő 420 néző között kisorsolnak 10 darab színházbérletet a következő évadra. Mennyi a valószínűsége, hogy egy négytagú családnak legalább az egyik tagja nyer bérletet? (1 ember csak 1 bérletet nyerhet).

Megoldás: Az, hogy legalább az egyikük nyer, azt jelenti, hogy 1,2,3 vagy 4 családtag nyer. Egyszerűbb lesz a dolgunk, ha a komplementer esemény, azaz annak a valószínűségét számoljuk ki, hogy egyikük sem nyer bérletet. Mivel 420 ember közül sorsolnak ki 10-et, így az összes eset száma $\binom{420}{10}$. Az, hogy egyikük sem nyer semmit, azt jelenti, hogy mind a 10 nyertest a többi 416 ember közül sorsolják ki. Ezen esetek száma $\binom{416}{10}$.

Így azt kapjuk, hogy

$$P(\text{valamelyikük nyer}) = 1 - P(\text{egyikük sem nyer}) = 1 - \frac{\binom{416}{10}}{\binom{420}{10}} \approx 1 - 0{,}9078 = 0{,}0923.$$

⇐3.10. feladat

3.11. feladat. Oldjuk most meg az előző feladatot abban az esetben, ha egy ember több bérletet is nyerhet!

Megoldás: Miként előbb, itt is a komplementer esemény valószínűségéből fogjuk meghatározni annak valószínűségét, hogy a család legalább egy tagja nyer valamit. Mivel most 420 ember közül úgy sorsolnak ki 10-et, hogy egy ember akár többször is nyerhet, ezért az összes eset száma 420^{10}. Azon esetek száma, amikor egyikük sem nyer semmit (vagyis amikor a nyertesek a többi 416 ember közül kerülnek ki) 416^{10}. Így azt kapjuk, hogy

$$P(\text{valamelyikük nyer}) = 1 - P(\text{egyikük sem nyer}) = 1 - \frac{416^{10}}{420^{10}} = 1 - \left(\frac{416}{420}\right)^{10} \approx 1 - 0{,}9087 = 0{,}0913.$$

⇐3.11. feladat

A 3.10. és a 3.11. feladatokban kapott eredmények láthatólag igen közel vannak egymáshoz. Tudjuk azt, hogy ha visszatevéses modellel számolunk (3.11.), akkor a sorsolás folyamán végig állandó annak a valószínűsége, hogy nem a családhoz tartozó néző nyer. Mivel a kisorsolt bérletek száma (10) viszonylag kevés az összes néző számához (420) és a nem a családhoz tartozó nézők számához (416) képest, ezért ha visszatevés nélküli modellel számolunk (mint a 3.10. feladatban), a sorsolás folyamán akkor is csak kis mértékben változik meg ez a valószínűség, így a kapott végeredmény is igen közel lesz a visszatevéses modellel számított értékhez. Ezt részletesebben fogjuk majd tárgyalni a 14. leckében.

3.12. feladat. Egy kaparós sorsjegyen azt olvassuk, hogy minden negyedik sorsjegy nyer. Veszünk 3 sorsjegyet. Mennyi a valószínűsége, hogy pontoson két nyertes lesz köztük?

Megoldás: Mivel egy kiválasztott sorsjegyet nem adhatunk vissza, ezért itt visszatevés nélküli mintavételről van szó. Azonban visszatevés nélküli mintavételnél a kérdezett valószínűség kiszámításához ismernünk kellene az összes sorsjegy számát is! Mit lehet ilyenkor tenni? Először is feltételezhetjük, hogy az összes sorsjegy száma elég nagy, bizonyára milliós nagyságrendű. Mivel a kihúzott sorsjegyek száma mind a nyertes, mind a nyeretlen sorsjegyek számához képest kicsi, ezért akármilyen sorsjegyet húzunk elsőre, a nyertes szelvények aránya gyakorlatilag nem változik; ugyanez mondható el a többi húzás esetén is. Vagyis dolgozhatunk úgy, mintha a nyertes szelvények aránya minden egyes húzás során $1/4$ lenne. Ez pedig azt jelenti, hogy visszatevés nélküli mintavétel helyett visszatevésessel számolunk, természetesen így a kérdéses valószínűség egy közelítő értékét kapjuk.

A feladat szövege szerint minden negyedik sorsjegy nyer, ami azt jelenti, hogy a nyertes szelvények aránya $\frac{1}{4}$, azaz $p = \frac{1}{4}$. Így a kérdéses valószínűség közelítő értéke

$$P(2 \text{ nyertes sorsjegy}) \approx \binom{3}{2} \cdot \left(\frac{1}{4}\right)^2 \cdot \left(\frac{3}{4}\right)^1 = 0{,}140625.$$

Arra, hogy mikor alkalmazható ez módszer, a 14. leckében még visszatérünk.

⇐3.12. feladat

Önellenőrzés

1. A 0,1,2,3,4,5 számjegyekből hány valódi hatjegyű, 5-re végződő szám készíthető, ha minden számjegyet csak egyszer használhatunk fel? Megold.

2. Kitöltünk egy ötöslottó szelvényt. Mennyi a valószínűsége, hogy legalább két találatunk lesz?

   ☐ 0,0233 ☐ 0,225 ☐ 0,0314 ☐ 0,0195

3. Nyolcan moziba mennek. Egy nyolc székes sorba ülnek le, a jegyeket véletlenszerűen osztják el egymás közt. Mennyi a valószínűsége, hogy András és Viki egymás mellett fog ülni?

   Megold.

4. A 32 lapos magyar kártya csomagból visszatevéses mintavétellel kiválasztunk 3 lapot. Mennyi a valószínűsége, hogy pontosan egy piros lesz a kiválasztott lapok között? Az eredményt három tizedesjegy pontossággal adja meg! Megold.

5. Feldobunk egy dobókockát háromszor egymás után. Mi a valószínűsége, hogy a dobott számok közt lesz páros is? Megold.

3. LECKE

A geometriai valószínűségi mező

3.2. A geometriai valószínűségi mező

A klasszikus valószínűségi mező csak olyan esetekben használható, amikor véges sok elemi eseményünk van, és ezek valószínűsége megegyezik. Előfordulhat azonban, hogy egy kísérletben a szóba jöhető elemi események száma nem véges, de még csak nem is megszámlálhatóan végtelen. Ilyenkor a klasszikus értelmezés természetesen nem alkalmazható. Ennek ellenére bizonyos esetekben számolhatunk a $\frac{\text{jó esetek száma}}{\text{összes eset száma}}$ formulával analóg módon.

3.2. definíció: Ha egy kísérlettel kapcsolatos események egy geometriai alakzat részhalmazainak feleltethetők meg úgy, hogy az egyes események valószínűsége az eseményekhez rendelt részhalmaz geometriai mértékével (hosszúság, terület, térfogat) arányos, akkor az események és valószínűségeik geometriai valószínűségi mezőt alkotnak. Legyen A egy ilyen kísérlettel kapcsolatos esemény. A kísérlettel kapcsolatban szóba jövő teljes alakzat mértéke legyen M, az A eseménynek megfelelő részalakzaté pedig m. Az A esemény valószínűsége ekkor a következő módon számolható:

$$P(A) = \frac{m}{M}.$$

A definíció szerint tehát akkor beszélünk geometriai valószínűségi mezőről, ha a teljes eseménytérnek egy geometriai alakzat feleltethető meg, továbbá minden eseménynek ennek az alakzatnak egy részhalmaza feleltethető meg oly módon, hogy az esemény valószínűsége csak az alakzat mértékétől (hossz/terület/térfogat) függ, míg az alakzat elhelyezkedésétől és alakjától független.

Példa: Két települést két kilométer hosszú egyenes főút köt össze. A rendőrség az egyik nap sebesség ellenőrzést tart ezen az úton. Tegyük fel, hogy a traffipaxot az út mellett véletlenszerűen helyezik el, azaz annak a valószínűsége, hogy egy adott útszakaszon van traffipax, az útszakasz helyzetétől nem, csak az útszakasz hosszától függ. Ekkor annak a valószínűsége, hogy egy d kilométer hosszú ($0 \leq d \leq 2$) útszakaszon van traffipax, az útszakasz és a teljes út hosszának arányával egyezik meg, azaz

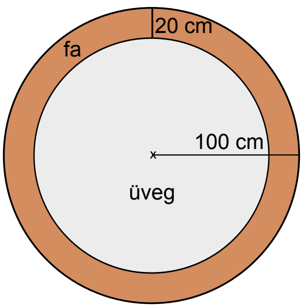
$$P(d \text{ hosszú úton van traffipax}) = \frac{d}{2}.$$

3.13. feladat. Egy kör alakú asztallap sugara 1 méter, az asztallap szélén egy 20 cm széles körgyűrű fából van, míg az asztal közepe üvegből. Egy pingponglabdát ejtünk le véletlenszerűen az asztallapra merőlegesen. Mennyi a valószínűsége, hogy a labda a fából készült részen pattan?

Megoldás:

⇐3.13. feladat

A véletlenszerűség azt jelenti, hogy annak a valószínűsége, hogy a labda egy kijelölt részre esik, az adott rész területével arányos. Így dolgozhatunk a geometriai valószínűségi mezővel. Az eseménytérnek a nagy kör felel meg, míg a jó eseményeknek a körgyűrű, a kérdezett valószínűség pedig a két alakzat területének hányadosával lesz egyenlő. A körgyűrű területét egyszerűen megkaphatjuk, ha a kör területéből kivonjuk a belső kisebb kör területét, így $P(\text{a labda a fa részre pattan}) = \frac{1^2 \cdot \pi - 0{,}8^2 \cdot \pi}{1^2 \cdot \pi} = 0{,}36$.

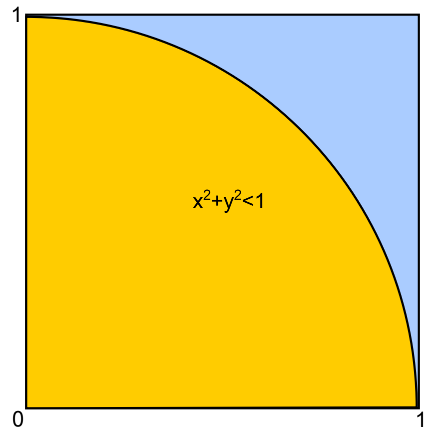


3-1. önálló feladat: Egy gyalogos-átkelőhelynél a lámpa 3 percig mutat pirosat és 20 másodpercig zöldet. A zebrán való átkelés 6 másodpercet vesz igénybe. Mi a valószínűsége, hogy az átkelőhöz érve a lámpa éppen zöldet mutat, és az átkelés végéig nem is vált pirosra?

3.14. feladat. Véletlenszerűen kiválasztjuk egy egységnyi oldalhosszúságú négyzet egy pontját. Mi a valószínűsége, hogy a választott pont a négyzet egy adott csúcsától 1 egységnél kisebb távolságra lesz?

Megoldás: Helyezzük el a négyzetet a koordináta-rendszerben úgy, hogy az adott csúcsa legyen az origó, továbbá egyik oldala az x tengely pozitív felére, egy másik oldala pedig az y tengely pozitív felére essen.

Egy (x,y) koordinátájú pontnak az origótól való távolsága $\sqrt{x^2+y^2}$, így a kiválasztott pont akkor lesz 1 egységnél kisebb távolságra az adott csúcstól (azaz az origótól, ha koordinátáira teljesül az $\sqrt{x^2+y^2}<1$ egyenlőtlenség, ami viszont ekvivalens azzal, hogy $x^2+y^2<1$. Azonban a koordinátasíkon azon pontok halmaza, melyek koordinátáira $x^2+y^2<1$, éppen az origó középpű, 1 sugarú körlap. Ezért a jó eseteknek megfelelő pontok halmaza az origó középpű, 1 sugarú körlap négyzetbe eső részével egyezik meg. Ismét használhatjuk a geometriai valószínűségi mezőt.



A kérdéses valószínűség a negyedkör és a négyzet területének a hányadosával lesz egyenlő, azaz

$$P(\text{a kiválasztott pont távolsága a négyzet adott csúcsától 1-nél kisebb }) = \frac{\frac{\pi}{4}}{1} = \frac{\pi}{4}\,.$$

⇐3.14. feladat

Érdekesség: Tudjuk, hogy egy kísérletet sokszor elvégezve, egy esemény relatív gyakorisága az esemény valószínűsége körül ingadozik. Így a 3.14. feladat szerint $\frac{\pi}{4}$ közelítését megkaphatjuk az alábbi módon: generáljunk véletlenszerűen „sok” pontot a négyzetben, számoljuk meg, hogy hány esik közülük a negyedkörbe, majd ezeknek a számát osszuk el az összes pont számával , így $\frac{\pi}{4}$ közelítését kapjuk. Ezzel a módszerrel tehát valószínűség-számítási megfontolások alapján kapjuk meg π egy közelítő értékét! Informatikusoknak érdemes egy kis programot írni, és elvégezni a kísérletet!

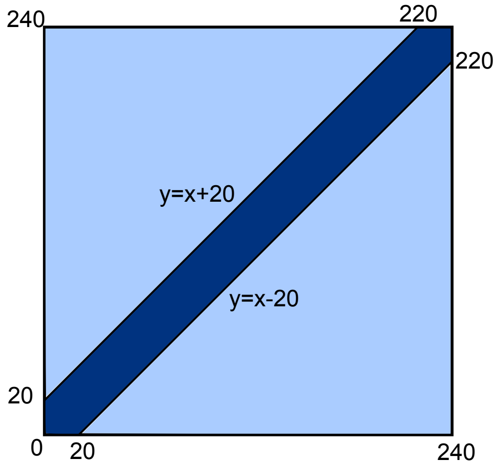
Az előző feladatoknál világos volt, hogy milyen alakzat felel meg az eseménytérnek, illetve a jó eseményeknek. Számos feladatnál azonban nem ez a helyzet, ilyenkor nekünk kell a megfelelő alakzatokat megtalálnunk.

3.15. feladat. Egy áruházba az egyik nap két kamion érkezik 8 és 12 óra között véletlenszerűen. Mindkét kamion kirakodása 20 percet vesz igénybe, azonban egyszerre csak az egyik kamionról lehet lerakodni az árukat, vagyis a később érkező kamionnak várnia kell, ha a korábbival még nem végeztek. Mi a valószínűsége, hogy a később érkező kamionnak várnia kell a pakolással?

Megoldás: Jelölje A azt az eseményt, hogy a később érkező kamionnak várakoznia kell. Jelölje továbbá x azt, hogy az egyik kamion 8 óra után hány perccel érkezik, és y azt, hogy a másik kamion hány perccel érkezik 8 után. Ekkor $0 \le x \le 240$ és $0 \le y \le 240$, az eseménytér pedig az összes ilyen (x,y) párból áll. Vegyük észre, hogy az ilyen párok éppen egy négyzetet határoznak meg, vagyis ez fog megfelelni az eseménytérnek. Határozzuk most meg a jó események halmazát. Két eset lehetséges:
1. Ha az első kamion érkezik előbb, azaz $x \le y$, akkor a második kamionnak pontosan akkor kell várakoznia, ha $x \le y \le x + 20$.
2. Ha a második kamion érkezik előbb, azaz $y \le x$, akkor a később érkező kamionnak pontosan akkor kell várakoznia, ha $y \le x \le y + 20$ teljesül.
A két esetnek a 3.1. ábrán a sötétebb rész felel meg. Ezek alapján az A esemény valószínűsége

$$P(A) = \frac{T_{\text{sötétebb}}}{T_{\text{négyzet}}} = \frac{240^2 - 2 \cdot \frac{220 \cdot 220}{2}}{240^2} \approx 0{,}1597\ .$$

A sötétebb rész területét úgy kaptuk meg, hogy a négyzet területéből levontuk a kimaradó két kis háromszög területét.



3.1. ábra. *A 3.15. („kamionos”) feladat jó eseményeinek halmaza egy hatszög.*

⇐3.15. feladat

Érdekesség: Számoljuk ki annak a valószínűségét, hogy az előző feladatban a két kamion egyszerre érkezik! Ekkor a kedvező eseteknek a négyzet azon pontjai felelnek meg, melyek x és y koordinátái megegyeznek, ez pedig éppen a négyzet origóból induló átlójának pontjaira teljesül. Mivel az átló (egy szakasz) területe 0, ezért annak a valószínűsége, hogy a két kamion egyszerre érkezik, $\frac{0}{240^2} = 0$. A két kamion egyszerre érkezik esemény természetesen nem lehetetlen esemény, a valószínűsége azonban mégis 0!

3.16. feladat. Egy két méteres rúd alkotója mentén véletlenszerűen kiválasztunk két pontot, és a rudat ezen a két helyen elfűrészeljük. Mennyi a valószínűsége, hogy a keletkezett 3 kis rúd mindegyike legalább 40 cm hosszú?

Megoldás: Legyen a két kiválasztott helynek a rúd bal oldali végpontjától vett távolsága x, illetve y centiméter. Ekkor az eseménytér az olyan (x,y) párokból áll, melyekre $0 \leq x \leq 200$ és $0 \leq y \leq 200$ teljesül. Ezen pontok összessége a síkban egy olyan négyzetnek felel meg, melynek az oldalának a hosszúsága 200 egység.

Jelölje A azt az eseményt, hogy mindhárom keletkező rúd hosszúsága legalább 40 cm. Két eset lehetséges:
1. $x < y$. Ekkor a kis rudak hossza rendre x, $y - x$ és $200 - y$, vagyis ebben az esetben a jó eseményekhez a négyzet azon pontjai tartoznak, melyek (x,y) koordinátáira az

$$x \geq 40 \qquad y - x \geq 40 \qquad 200 - y \geq 40$$

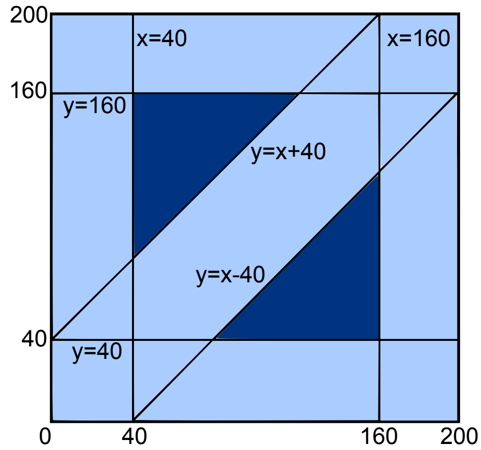
egyenlőtlenségek mindegyike teljesül.

2. $y < x$. Ekkor a kis rudak hossza rendre y, $x - y$ és $200 - x$, vagyis ebben az esetben a jó eseményekhez a négyzet azon pontjai tartoznak, melyek (x,y) koordinátáira az

$$y \geq 40 \qquad x - y \geq 40 \qquad 200 - x \geq 40$$

egyenlőtlenségek mindegyike teljesül.
A 3.2. ábrán a sötétebb rész mutatja a jó eseményeknek megfelelő halmazt. Látható, hogy a sötétebb rész két egybevágó, egyenlő szárú derékszögű háromszögből áll, ezért elegendő az egyik területét kiszámolni. A bal felső derékszögű háromszög derékszögnél lévő csúcsa az $x = 40$ és $y = 160$ egyenesek metszéspontja, így ennek a koordinátái $(40{,}160)$. Ugyanennek a háromszögnek egy másik pontja az $x = 40$ és $y = x + 40$ egyenesek metszéspontja, így koordinátáit ennek a két egyenletből álló egyenletrendszernek a megoldásaként kapjuk: $(40; 80)$. A derékszögű háromszög befogójának a hossza ennek a két pontnak a távolságával egyezik meg, ami $160 - 80 = 80$ cm. Így egy háromszög területe $\frac{80 \cdot 80}{2} = 3200$ cm^2. Ebből az A esemény valószínűsége:

$$P(A) = \frac{T_{\text{sötétebb}}}{T_{\text{négyzet}}} = \frac{2 \cdot 3200}{40000} = 0{,}08.$$



3.2. ábra. *A 3.16. („rudas") feladat jó eseményeinek halmaza két háromszög.*

⇐3.16. feladat

Önellenőrzés

1. Fogalmazza meg a saját szavaival, hogy mikor beszélhetünk geometriai valószínűségi mezőről!
2. Egy 4cm$\times$5cm-es téglalapnak véletlenszerűen kiválasztjuk egy pontját. Mi a valószínűsége, hogy a kiválasztott pont mindegyik oldaltól legalább 1cm távolságra van?

   ☐ 0,3 ☐ 0,4 ☐ 0,5 ☐ 0,6
3. Egy 5 cm sugarú, kör alakú céltáblát véletlenszerűen eltalál egy lövés. Mi a valószínűsége, hogy a találat helye a kör középpontjától legalább 2, de legfeljebb 3 cm-re van?

   ☐ 0,16 ☐ 0,2 ☐ 0,25 ☐ 0,3
4. Véletlenszerűen kiválasztunk a $[0,1]$ intervallumon 2 számot. Mennyi a valószínűsége, hogy a két szám összege legalább 1,2? Megold.
5. Egy 2 cm oldalhosszúságú négyzetben véletlenszerűen kiválasztunk egy pontot. Mi a valószínűsége, hogy a kiválasztott pont az átlók metszéspontjától kevesebb, mint $0,2$ cm-re van?

   ☐ 0,031 ☐ 0,068 ☐ 0,185 ☐ 0,263

4. LECKE

Feltételes valószínűség, függetlenség

4. Feltételes valószínűség, függetlenség

4.1. Feltételes valószínűség

Sokszor előfordul, hogy egy kísérlettel kapcsolatos A esemény valószínűségének meghatározásakor rendelkezünk valamilyen információval a kísérlet kimenetelével kapcsolatban, például tudjuk, hogy egy B esemény bekövetkezett. Ilyenkor azt mondjuk, hogy az A esemény valószínűségét a B feltétel mellett adjuk meg. Nézzünk most erre egy egyszerű példát.

Példa: Egy tárgyat 30 járműmérnök és 20 közlekedésmérnök hallgató vett fel. Év végén a járműmérnökök közül 24-en, a közlekedésmérnökök közül 15-en mentek át a vizsgán. Véletlenszerűen kiválasztunk a tárgyat felvettek közül egy hallgatót. Ekkor annak a valószínűsége, hogy átment a vizsgán

$$\frac{\text{átmentek száma}}{\text{összes hallgató száma}} = \frac{39}{50} = 0{,}78.$$

Tegyük most fel, hogy valaki elárulja nekünk, hogy járműmérnök hallgatót választottunk ki. Ekkor annak a valószínűsége, hogy átment a vizsgán

$$\frac{\text{átmentek száma a járműmérnökök közt}}{\text{járműmérnökök száma}} = \frac{24}{30} = 0{,}8.$$

A második valószínűség egy feltételes valószínűség volt, hiszen arra voltunk kíváncsiak, hogy mi a valószínűsége, hogy egy hallgató átment a vizsgán, feltéve, hogy a hallgató járműmérnök. Figyeljük meg, hogy a nevezőben a feltételnek (járműnérnök hallgató) megfelelő esetek száma áll, míg a számlálóban azoknak az eseteknek a száma, amikor a feltétel és a kérdezett esemény egyszerre bekövetkezik (járműmérnök és átment a vizsgán). Tulajdonképpen azt csináltuk, hogy a teljes eseményteret (hallgatók) leszűkítettük annak egy részhalmazára (járműmérnökök), és ezen az új, szűkebb eseménytéren határoztuk meg a kérdéses esemény valószínűségét.

A feltételes valószínűséget általános esetben is az előző példához hasonló módon, azaz a

$$\frac{\text{együttes bekövetkezés valószínűsége}}{\text{feltétel valószínűsége}}$$

képlettel határozhatjuk meg.

4.1. definíció: Ha A és B egy kísérlettel kapcsolatos két tetszőleges esemény és $P(B) > 0$, akkor az A esemény B -re vonatkozó feltételes valószínűségét a

$$P(A|B) = \frac{P(A \cdot B)}{P(B)}$$

kifejezéssel definiáljuk.

4.1. feladat. Legyen $P(A) = 0{,}4$, $P(B) = 0{,}8$ és $P(A|B) = 0{,}2$. Mennyi a $P(B|A)$ valószínűség értéke?

Megoldás: A definíció szerint, $P(B|A) = \frac{P(B \cdot A)}{P(A)}$, ezért meg kell határoznunk $P(B \cdot A)$-t. Ismét a definíciót és az adatokat használva a

$$0{,}2 = P(A|B) = \frac{P(A \cdot B)}{P(B)} = \frac{P(A \cdot B)}{0{,}8}$$

egyenletet kapjuk. Ezt átrendezve $P(A \cdot B) = 0{,}2 \cdot 0{,}8 = 0{,}16$ adódik. Így

$$P(B|A) = \frac{P(B \cdot A)}{P(A)} = \frac{0{,}16}{0{,}4} = 0{,}4.$$

⇐4.1. feladat

4.2. feladat. Feldobunk három dobókockát. Feltéve, hogy legalább 2 hatost dobunk, mi a valószínűsége, hogy mindhárom dobás hatos lesz?

Megoldás: A definíció szerint

$$P(3 \text{ hatos}|\text{ legalább 2 hatos}) = \frac{P(3 \text{ hatos és legalább 2 hatos})}{P(\text{legalább 2 hatos})} = \frac{P(3 \text{ hatos })}{P(\text{legalább 2 hatos})}.$$

Mivel az összes eset száma 6^3, és 3 hatost csak egyféleképpen dobhatunk, ezért $P(3\text{ hatos}) = \frac{1}{216}$. Az, hogy legalább 2 hatost dobunk, azt jelenti, hogy vagy pontosan 2 hatost dobunk, vagy 3 hatost dobunk. Pontosan 2 hatost úgy tudunk dobni, hogy az egyik kockával nem hatost dobunk, a másik kettővel pedig hatost; ezért ezeknek az eseteknek a száma $5 \cdot 1 \cdot 1 + 1 \cdot 5 \cdot 1 + 1 \cdot 1 \cdot 5 = 15$. Ebből

$$P(\text{legalább 2 hatos}) = \frac{15+1}{216} = \frac{16}{216},$$

amiből a kérdezett valószínűség:

$$P(3\text{ hatos} \mid \text{legalább 2 hatos}) = \frac{\frac{1}{216}}{\frac{16}{216}} = \frac{1}{16} = 0{,}0625.$$

⇐4.2. feladat

4.3. feladat. Tudjuk, hogy egy volt osztálytársunknak 2 gyereke van. Tegyük fel, hogy a fiúk és lányok születésének egyenlő a valószínűsége.

a) Mennyi a valószínűsége, hogy mindkét gyerek fiú?

b) Egy közösségi portálon látunk egy fotót a családról, amiből kiderül, hogy az egyik gyerek fiú. Mi a valószínűsége, hogy a másik is az?

Megoldás:

a) Jelölje F, illetve L a fiú, illetve lánygyereket. Az eseménytér, $\Omega = \{FF, FL, LF, LL\}$, ahol az első tag az elsőszülöttnek felel meg. Mivel a fiúk és lányok születésének egyenlő a valószínűsége, ezért alkalmazhatjuk a klasszikus valószínűségi mezőt:

$$P(\text{mindkét gyerek fiú}) = \frac{\text{jó esetek száma}}{\text{összes eset száma}} = \frac{1}{4}.$$

b) Ez a feladat az előadásokon rengeteg vitát szokott kiváltani. Sokan egyből rávágják, hogy a megoldás $\frac{1}{2}$ (hiszen a másik gyerek ugyanolyan valószínűséggel lehet fiú vagy lány), mások meg azt, hogy $\frac{1}{4}$ (hiszen a feladat ugyanaz mint az előbb). Nézzük meg, hogy valójában mi a helyzet!
Vegyük észre, hogy az előző részhez képest az az eltérés, hogy most egy feltételes valószínűségről van szó! A kérdés gyakorlatilag az, hogy mi a valószínűsége, hogy mindkét gyerek fiú, feltéve, hogy legalább az egyik fiú. A definíciót használva:

$$P(2 \text{ fiú} \,|\, \text{legalább 1 fiú}) = \frac{P(\text{2fiú és legalább 1 fiú})}{P(\text{legalabb 1 fiú})} = \frac{P(\text{2fiú})}{P(\text{legalabb 1 fiú})} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

Ha a kedves olvasó esetleg továbbra is hitetlenül csóválná a fejét, akkor gondoljuk meg a következőt: mivel tudjuk, hogy az egyik gyerek fiú, ezért a LL eset nem fordulhat elő, ezért az összes eset száma három (FF,FL,LF). A jó esetek száma egy (amikor mindkét gyerek fiú), így a valószínűség $\frac{1}{3}$.

⇐4.3. feladat

4.2. definíció: A feltételes valószínűség definícióját felhasználva $P(A \cdot B)$ kifejezhető

$$P(A \cdot B) = P(A|B) \cdot P(B)$$

alakban, amit a valószínűségek szorzási szabályának nevezünk.

4.4. feladat. Egy ládában 8 piros, 5 zöld és 3 kék golyó van. Visszatevés nélkül kihúzunk 3 golyót. Mennyi a valószínűsége, hogy elsőre pirosat, másodikra kéket, harmadikra pedig zöldet húzunk?

Megoldás: Legyen A az az esemény, hogy elsőre pirosat húzunk, B az, hogy másodikra kéket, C pedig az, hogy harmadikra zöldet. A feladat az $A \cdot B \cdot C$ esemény valószínűségének a kiszámítása. Többször alkalmazva a szorzási szabályt azt kapjuk, hogy

$$P(A \cdot B \cdot C) = P(C \cdot B \cdot A) = P(C|B \cdot A) \cdot P(B \cdot A) = P(C|B \cdot A) \cdot P(B|A) \cdot P(A).$$

Elsőre $\frac{8}{16}$ valószínűséggel húzhatunk pirosat (hiszen a 16 golyó közül 8 piros), így $P(A) = \frac{1}{2}$. Ha elsőre pirosat húzunk, akkor másodikra $\frac{3}{15}$ valószínűséggel húzhatunk kéket (hiszen ekkor már csak 15 golyó van, melyek közül 3 kék), azaz $P(B|A) = \frac{3}{15}$. Végül annak a valószínűsége, hogy harmadikra zöldet húzunk, feltéve, hogy elsőre pirosat, másodikra pedig kéket húztunk $\frac{5}{14}$ (a maradék 14 golyó között 5 zöld van), azaz $P(C|A \cdot B) = \frac{5}{14}$. Ebből a kérdéses valószínűség:

$$P(A \cdot B \cdot C) = \frac{8}{16} \cdot \frac{3}{15} \cdot \frac{5}{14} \approx 0{,}036.$$

⇐4.4. feladat

4.2. A teljes valószínűség tétele és a Bayes-tétel

Gyakran találkozunk olyan feladattal, amelyben ismerjük egy teljes eseményrendszerhez tartozó események valószínűségét, továbbá egy másik A eseménynek, a teljes eseményrendszer egyes eseményeire vonatkozó feltételes valószínűségeit. A következő tétel azt mutatja, hogy ezen adatok birtokában az A esemény valószínűsége is kiszámítható.

4.1. tétel: (A teljes valószínűség tétele) Ha a $B_1, B_2, \ldots, B_n$ események teljes eseményrendszert alkotnak, és $P(B_i) > 0$ minden $i = 1, 2, \ldots, n$ esetén, valamint A egy tetszőleges esemény, akkor

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i) .$$

Bizonyítás: Mivel a $B_1, B_2, \ldots, B_n$ események teljes eseményrendszert alkotnak, ezért $B_i \cdot B_j = \emptyset$ $(i \neq j)$, továbbá $B_1 + B_2 + \ldots + B_n = \Omega$. Ezt felhasználva átírhatjuk A-t más alakba:

$$A = A \cdot \Omega = A \cdot (B_1 + B_2 + \ldots + B_n) = A \cdot B_1 + A \cdot B_2 + \ldots + A \cdot B_n .$$

A kapott események páronként kizárják egymást, mivel

$$(A \cdot B_i) \cdot (A \cdot B_j) = A \cdot (B_i \cdot B_j) = A \cdot \emptyset = \emptyset\,.$$

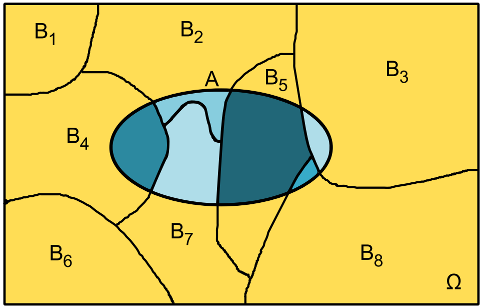
A 3. axióma alapján ezért

$$P(A) = P\left(A \cdot B_1 + A \cdot B_2 + \ldots + A \cdot B_n\right) = P(A \cdot B_1) + P(A \cdot B_2) + \ldots + P(A \cdot B_n)\,.$$

A szorzási szabály miatt azonban $P(A \cdot B_i) = P(A|B_i) \cdot P(B_i)$, így

$$\begin{aligned}P(A) &= P\left(A \cdot B_1 + A \cdot B_2 + \ldots + A \cdot B_n\right) = P(A \cdot B_1) + P(A \cdot B_2) + \ldots + P(A \cdot B_n) = \\ &= P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) + \ldots + P(A|B_n) \cdot P(B_n) = \\ &= \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)\,.\end{aligned}$$

$\square$



4.1. ábra. *A teljes valószínűségtétele szerint az A esemény valószínűsége kiszámítható az egymást kizáró $A \cdot B_i$ események valószínűségeinek összegeként.*

4.5. feladat. Egy üzemben három gép gyártja ugyanazt a terméket. Az első gép a teljes termelés 50%-át, a második gép a 30%-át, a harmadik pedig a 20%-át adja. Tapasztalataink szerint az első gép által gyártott termékek 5%-a, a második gép által gyártott termékek 3%-a, míg a harmadik gép által gyártott termékek 2%-a selejtes. A nap végén a termékeket beviszik a raktárba (itt már nem állapítható meg, hogy melyik gép gyártotta a terméket). Mi a valószínűsége, hogy a raktárból véletlenszerűen választott termék selejtes?

Megoldás: Jelölje A azt az eseményt, hogy a kiválasztott termék selejtes, B_1,B_2, illetve B_3 pedig azt, hogy a terméket az első, második, illetve a harmadik gép gyártotta. A feladat szövege alapján ismerjük a B_i események valószínűségét: $P(B_1) = 0{,}5$, $P(B_2) = 0{,}3$ és $P(B_3) = 0{,}2$. Ismerjük továbbá az A eseménynek a B_i feltételek melletti valószínűségét: $P(A|B_1) = 0{,}05$, $P(A|B_2) = 0{,}03$ és $P(A|B_3) = 0{,}02$. A teljes valószínűség tételét alkalmazva:

$$P(A) = \sum_{i=1}^{3} P(A|B_i) \cdot P(B_i) = 0{,}05 \cdot 0{,}5 + 0{,}03 \cdot 0{,}3 + 0{,}02 \cdot 0{,}2 = 0{,}38.$$

⇐4.5. feladat

4.6. feladat. Visszatevés nélkül kihúzunk a 32 lapos magyarkártya-csomagból 2 lapot. Mi a valószínűsége, hogy másodikra pirosat húzunk?

Megoldás: Legyen A az az esemény, hogy másodikra pirosat húzunk. Legyen a teljes eseményrendszer az, hogy elsőre milyen színű lapot húzunk: B_1 jelentse, hogy pirosat, B_2, hogy zöldet, B_3, hogy makkot, B_4, hogy tököt. Természetesen $P(B_i) = \frac{1}{4}$ mindegyik i-re. Ha elsőre nem pirosat húzunk, akkor annak a valószínűsége, hogy másodikra pirosat húzunk $\frac{8}{31}$, hiszen 31 lap közül 8 piros; így $P(A|B_2) = P(A|B_3) = P(A|B_4) = \frac{8}{31}$. Ha elsőre pirosat húzunk, akkor annak a valószínűsége, hogy másodikra is pirosat húzunk $\frac{7}{31}$, hiszen a 31 lap között 7 piros marad; így $P(A|B_1) = \frac{7}{31}$. A teljes valószínűség tételét alkalmazva

$$P(A) = \sum_{i=1}^{4} P(A|B_i) \cdot P(B_i) = \frac{7}{31} \cdot \frac{1}{4} + \frac{8}{31} \cdot \frac{1}{4} + \frac{8}{31} \cdot \frac{1}{4} + \frac{8}{31} \cdot \frac{1}{4} = 0{,}25.$$

A feladatot teljes valószínűség tételének felhasználása nélkül „józan paraszti ésszel” is megoldhattuk volna:

nyilvánvaló, hogy a feladatban a piros színnek semmilyen jelentősége nincs, így annak a valószínűsége, hogy másodikra pirosat húzunk, ugyanannyi, mint annak a valószínűsége, hogy másodikra zöldet/tököt/makkot húzunk. Mivel ezek egymást kizáró események és összegük a biztos esemény (azaz a 4 esemény egy teljes eseményrendszert alkot), ezért

$$P(\text{másodikra piros}) + P(\text{másodikra zöld}) + P(\text{másodikra makk}) + P(\text{másodikra tök}) = 1.$$

Mivel mind a négy valószínűség ugyanakkora, ezért annak a valószínűsége, hogy másodikra pirosat húzunk 0,25. Ugyanennyi a valószínűsége, hogy másodikra zöldet (makkot, tököt) húzunk. ⇐4.6. feladat

4.2. tétel: (Bayes-tétel) Ha a $B_1, B_2, \ldots, B_n$ események teljes eseményrendszert alkotnak és $P(B_i) > 0$ minden $i = 1,2,\ldots,n$ esetén, valamint A egy tetszőleges pozitív valószínűségű esemény, azaz $P(A) > 0$, akkor

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

Bizonyítás: Alakítsuk át a $P(B_k|A)$ valószínűséget a feltételes valószínűség definíciója szerint, majd a számlálóban alkalmazzuk a szorzási szabályt, a nevezőben pedig a teljes valószínűség tételét:

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

□

A Bayes-tétel segítségével meghatározhatjuk, hogy az A esemény bekövetkezése esetén milyen valószínűséggel következik be a teljes eseményrendszernek valamely eseménye.

4.7. feladat. Tapasztalatok szerint az egyik gyárban az elkészült termékek 0,1%-ának a felületén található valamilyen hiba (pl. apró karcolás). Az elkészült termékek felületét egy munkás ellenőrzi. A munkás egy olyan terméket, amelynek a felülete hibátlan, 2% valószínűséggel minősít tévesen hibásnak; míg egy hibás terméket 1% valószínűséggel minősít (szintén tévesen) hibátlannak.

a) Mennyi a valószínűsége, hogy a munkás hibásnak minősít egy terméket?

b) Egy terméket a munkás hibásnak minősít. Mennyi a valószínűsége, hogy a termék valóban hibás?

Megoldás:

a) Jelölje A azt az eseményt, hogy egy terméket a munkás hibásnak minősít. Jelölje továbbá B_1 azt az eseményt, hogy egy termék hibás, míg B_2 azt, hogy hibátlan. Ekkor B_1 és B_2 teljes eseményrendszert alkotnak. A feladat szövege alapján tudjuk, hogy $P(B_1) = 0{,}001$, amiből $P(B_2) = 1 - 0{,}001 = 0{,}999$. Mivel a munkás a hibás terméket 1% valószínűséggel jónak minősít, ezért 99% valószínűséggel a hibás teméket hibásnak minősíti, azaz $P(A|B_1) = 0{,}99$. A szövegből szintén kiolvasható, hogy $P(A|B_2) = 0{,}02$. Használjuk a teljes valószínűség tételét:

$$P(A) = P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) = 0{,}99 \cdot 0{,}001 + 0{,}02 \cdot 0{,}999 = 0{,}02097,$$

vagyis a munkás körülbelül 2 százalékos valószínűséggel minősít egy terméket hibásnak.

b) A kérdés az, hogy mennyi a valószínűsége annak, hogy egy termék hibás, feltéve, hogy a munkás hibásnak minősíti, vagyis a $P(B_1|A)$ valószínűséget kell kiszámolnunk. Írjuk fel a Bayes-tételt a kérdéses valószínűségre:

$$P(B_1|A) = \frac{P(A|B_1) \cdot P(B_1)}{P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2)} = \frac{0{,}99 \cdot 0{,}001}{0{,}99 \cdot 0{,}001 + 0{,}02 \cdot 0{,}999} \approx 0{,}047,$$

azaz körülbelül 4,7 százalék annak a valószínűsége, hogy egy, a munkás által hibásnak talált termék tényleg hibás.

⇐4.7. feladat

Azt kaptuk tehát, hogy ha egy terméket a munkás hibásnak minősít, akkor kicsi annak a valószínűsége (mindössze 4,7 százalék), hogy a termék tényleg hibás. Vajon mi értelme lenne akkor egy ilyen munkatársat alkalmazni?! A válaszhoz először gondoljuk meg a következőt:

Aktivitás: Számítsa ki annak a valószínűségét, hogy egy, a munkás által hibátlannak minősített termék valóban hibátlan!

Lassan körvonalazódik a munkatárs haszna. Képzeljük el, hogy létezik egy drága berendezés, amellyel pontosan meg lehet állapítani egy termékről, hogy van-e a felszínén hiba. Azonban a vizsgálat sokkal tovább tart, mint a munkás általi szemrevételezés, ezért az összes termék megvizsgálása gyakorlatilag lehetetlen. Ehelyett a következőképpen járunk el: a munkás megvizsgálja a terméket, és ha jónak találja, akkor szinte biztosak lehetünk benne, hogy a termék hibátlan (lásd az előző aktivitásbeli feladatot). Ha a munkás hibásnak találja a terméket, akkor megvizsgáljuk a berendezéssel is, így döntjük el, hogy valóban hibás-e a termék. Mivel egy terméket a munkás 0,02 valószínűséggel minősít hibásnak, így átlagosan csak minden 50. terméket kell a berendezéssel megvizsgálni, és így rengeteg időt (és vele pénzt) takarítunk meg.

Érdekesség: Az egészségügyben hasonló elven működik számos szűrővizsgálatra szolgáló teszt. A Down-szindrómát például a magzat kromoszómavizsgálatával ki lehet mutatni, ez azonban egy rendkívül kockázatos beavatkozást igényel, amely növeli a vetélés esélyét. Ezért ehelyett a Down-szindróma szűrésére különféle, kockázatmentes teszteket fejlesztettek ki (jelenleg az ún. kombinált teszt a legjobb). Ha a teszt negatív eredményt ad, akkor a születendő gyerek „szinte biztosan” nem lesz Down-szindrómás (közel 1 a valószínűsége). Ha a teszt pozitív eredményt ad, még akkor is kicsi a valószínűsége, hogy a születendő gyerek Down-szindrómás lesz (vagyis nem kell egyből megijedni!). Ilyen esetekben azonban a kismamáknak felajánlják, hogy elvégzik a magzati kromoszómavizsgálatot.

4.3. Események függetlensége

A hétköznapi életben akkor szoktunk két eseményt függetlennek tekinteni, ha egyiknek sincs semmilyen „hatása” a másikra. A valószínűség-számításban a függetlenséget másképp definiáljuk, de majd egy későbbi tétel biztosítja, hogy az így definiált függetlenség pontosan a hétköznapi értelemben elvárt függetlenségnek felel meg.

4.3. definíció: Két eseményt függetlennek nevezünk, ha az együttes bekövetkezésük valószínűsége egyenlő a valószínűségeik szorzatával, azaz

$$P(A \cdot B) = P(A) \cdot P(B) .$$

Definíció szerint tehát 2 esemény akkor független, ha az együttes bekövetkezésük valószínűsége egyenlő a külön-külön vett bekövetkezésük valószínűségeinek szorzatával.

4.8. feladat. Egy magyarkártya-pakliból kihúzunk egy lapot. Legyen A az az esemény, hogy a kihúzott lap piros, míg B, hogy a kihúzott lap ász. Független-e a két esemény?

Megoldás: Nyilván $P(A) = \frac{1}{4}$ és $P(B) = \frac{1}{8}$. Mivel az $A \cdot B$ esemény azt jelenti, hogy a kihúzott lap a piros ász, ezért $P(A \cdot B) = \frac{1}{32}$. De ekkor

$$P(A \cdot B) = \frac{1}{32} = \frac{1}{4} \cdot \frac{1}{8} = P(A) \cdot P(B),$$

ezért a definíció szerint A és B függetlenek.

⇐4.8. feladat

Természetesen nem csak két, hanem több esemény függetlenségéről is beszélhetünk. Az események függetlenségének a fenti definícióját kis módosítással kiterjeszthetjük több eseményre is.

4.4. definíció: Az $A_1, A_2, \ldots, A_n$ eseményeket (teljesen) függetleneknek nevezzük, ha bárhogyan is választunk ki közülük k darab eseményt ($k = 2, 3, \ldots, n$), ezek szorzatának valószínűsége egyenlő az egyes események valószínűségeinek szorzatával. Az $A_1, A_2, \ldots, A_n$ események függetlensége esetén tehát a

$$\begin{aligned} P(A_i \cdot A_j) &= P(A_i) \cdot P(A_j) && 1 \leq i < j \leq n \\ P(A_i \cdot A_j \cdot k) &= P(A_i) \cdot P(A_j) \cdot P(A_k) && 1 \leq i < j < k \leq n \\ &\vdots && \vdots \\ P(A_1 \cdot A_2 \cdot \ldots \cdot A_n) &= P(A_1) \cdot P(A_2) \cdot \ldots \cdot P(A_n) \end{aligned}$$

egyenlőségeknek mind teljesülniük kell.

Vagyis több esemény akkor független, ha bárhogy is választunk ki közülük néhányat, ezek szorzatának a valószínűsége meg kell, hogy egyezzen az egyes események valószínűségeinek a szorzatával. Vigyázat: vegyük észre, hogy itt összesen $\sum_{k=2}^{n} \binom{n}{k} = 2^n - n - 1$ feltételnek kell teljesülni!

Az alábbi tétel a feltételes valószínűség és az események függetlenségének a definíciójával egyszerűen belátható (a bizonyítást az olvasóra bízzuk).

4.3. tétel: Ha A és B független események, akkor

$$P(A|B) = P(A) \quad \text{és} \quad P(B|A) = P(B)\,.$$

A tételben szereplő két állítás felel meg az események hétköznapi értelemben használt függetlenségének, ui. a tétel azt mondja ki, hogy ha két esemény független, akkor egyik bekövetkezése sem változtatja meg a másik bekövetkezésének a valószínűségét (nincs rá hatással).

4.4. tétel: Ha A és B független események, akkor A és $\overline{B}$, $\overline{A}$ és B, valamint $\overline{A}$ és $\overline{B}$ is független események.

Bizonyítás: Az átalakítások során azt fogjuk kihasználni, hogy $A \cdot \overline{B} = A - B$, a különbség valószínűsége $P(A - B) = P(A) - P(A \cdot B)$, továbbá azt, hogy ha A és B függetlenek, akkor $P(A \cdot B) = P(A) \cdot P(B)$.

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

Azt kaptuk, hogy $P(A \cdot \overline{B}) = P(A) \cdot P(\overline{B})$, tehát A és $\overline{B}$ függetlenek. Hasonlóan látható be a tétel másik két állítása is. □

4.9. feladat. Egy irodában három új fénymásolót vesznek, amelyek aztán egymástól függetlenül működnek. Az elsőt 0,3, a másodikat 0,4, a harmadikat 0,5 valószínűséggel kell 1 éven belül javítani. Mennyi a valószínűsége, hogy

a) 1 éven belül egyikhez sem kell szerelőt hívni,

b) legalább az egyik elromlik 1 éven belül,

c) pontosan 2 romlik el egy éven belül?

Megoldás:

a) Jelölje A, B, illetve C azt az eseményt, hogy az első, a második, illetve a harmadik fénymásoló elromlik 1 éven belül. Ekkor felhasználva, hogy független eseményekről van szó:

$$P(\text{egyik sem romlik el 1 éven belül}) = P(\overline{A} \cdot \overline{B} \cdot \overline{C}) = P(\overline{A}) \cdot P(\overline{B}) \cdot P(\overline{C}) = 0{,}7 \cdot 0{,}6 \cdot 0{,}5 = 0{,}21.$$

b) A *legalább egyik elromlik esemény* komplementere az *egyik sem romlik el*, így felhasználva az előző rész eredményét:

$$P(\text{legalább 1 elromlik 1 éven belül}) = 1 - P(\text{egyik sem romlik el 1 éven belül}) = 1 - 0{,}21 = 0{,}79.$$

c) Az, hogy pontosan 2 gép romlik el, felbontható 3 esemény összegére: az első nem romlik el, a másik kettő igen+ a második nem romlik el, a másik kettő igen+ a 3. nem romlik el, a másik kettő igen. Így

$$P(\text{pontosan kettő romlik el}) = P(\overline{A} \cdot B \cdot C + A \cdot \overline{B} \cdot C + A \cdot B \cdot \overline{C}) = P(\overline{A}) \cdot P(B) \cdot P(C) +$$
$$+ P(A) \cdot P(\overline{B}) \cdot P(C) + P(A) \cdot P(B) \cdot P(\overline{C}) = 0{,}7 \cdot 0{,}4 \cdot 0{,}5 + 0{,}3 \cdot 0{,}6 \cdot 0{,}5 + 0{,}3 \cdot 0{,}4 \cdot 0{,}5 = 0{,}29.$$

⇐4.9. feladat

Önellenőrzés

1. Tudjuk, hogy $P(A) = 0{,}5$, $P(B) = 0{,}6$ és $P(A|B) = 0{,}7$. Mennyi a $P(A + B)$ valószínűség értéke?

Megold.

2. Feldobunk 2 dobókockát. Feltéve, hogy dobunk páratlan számot, mennyi a valószínűsége, hogy dobunk hatost?

☐ $\frac{6}{36}$ ☐ $\frac{27}{36}$ ☐ $\frac{3}{27}$ ☐ $\frac{6}{27}$

3. Egy városban a lakosság 48 százaléka férfi, 52 százaléka nő. A férfiak 65, a nők 70 százaléka ötven évnél fiatalabb. Véletlenszerűen kiválasztjuk a város egy lakosát. Mennyi a valószínűsége, hogy legalább 50 éves?

Megold.

4. Ha az előző feladatban egy véletlenszerűen kiválasztott lakos 50 évnél fiatalabb, akkor mi a valószínűsége, hogy férfi?

☐ $0{,}462$ ☐ $0{,}778$ ☐ $0{,}6$ ☐ $0{,}667$

5. Egy gyárban három gép működik egymástól függetlenül. Az első gép $0{,}03$, a második $0{,}05$, míg a harmadik $0{,}02$ valószínűséggel romlik el egy nap. Mennyi a valószínűsége, hogy egy nap legalább az egyik gép elromlik? Az eredményt öt tizedesjegy pontossággal adja meg!

Megold.

5. LECKE

Modulzáró feladatok 1.

5. Modulzáró feladatok

Start. ▷ Az alábbi kérdések megválaszolására 1 órája van:

1. Egy üzletközpontban két bankautomata és két mozgólépcső van. Jelölje A_1 és A_2, hogy az első, illetve második bankautomata működik egy nap; jelölje továbbá B_1 és B_2, hogy egy nap folyamán az első, illetve második mozgólépcső működik. Mit jelent az $(A_1 + A_2) \cdot (B_1 + B_2)$ esemény?

   ☐ Mindegyik mozgólépcső és bankautomata működik.

   ☐ Legalább az egyik bankautomata, és legalább az egyik mozgólépcső működik.

   ☐ A 4 berendezés közül legalább az egyik működik.

   ☐ A 4 berendezés közül legalább az egyik nem működik.

2. Az A és B események függetlenek. Tudjuk, hogy $P(A) = 0{,}4$ és $P(B) = 0{,}8$. Mennyi a $P(B - A)$ valószínűség értéke? ☐

3. Véletlenszerűen leültetünk nyolc embert egymás mellé egy sorba. Mennyi a valószínűsége, hogy András és Bea egymás mellett fog ülni? ☐

4. Véletlenszerűen kiválasztunk egy valódi ötjegyű számot. Mennyi a valószínűsége, hogy a kiválasztott szám számjegyei közt lesz legalább egy darab 1-es vagy 2-es?

   ☐ 0,615 ☐ 0,681 ☐ 0,734 ☐ 0,804

5. A 32 lapos magyarkártya-csomagból visszatevés nélkül kiválasztunk négy lapot. Mennyi a valószínűsége, hogy pontosan egy ász lesz a kiválasztottak közt?

   ☐ 0,258 ☐ 0,364 ☐ 0,415 ☐ 0,462

6. Egy üzembe két cég szállítja az alkatrészeket. Az első cég az alkatrészek 40, a másik pedig a 60 százalékát szállítja. Az első cégnél az alkatrészek 3 százaléka, a másodiknál az alkatrészek 2 százaléka hibás. Egy napon mindkét cég meghozza az árut, amit aztán bevisznek a raktárba Mennyi a valószínűsége, hogy egy, a raktárból véletlenszerűen választott alkatrész hibás?

7. Ha az előző feladatban egy, a raktárból véletlenszerűen választott alkatrész hibátlan, akkor mennyi a valószínűsége, hogy az első cégtől származik?

☐ 0,352 ☐ 0,365 ☐ 0,388 ☐ 0,398

8. Feldobunk három dobókockát. Feltéve, hogy a dobott számok szorzata páratlan, mennyi a valószínűsége, hogy a dobott számok szorzata prímszám? (Az 1 nem prímszám!)

☐ 0,222 ☐ 0,333 ☐ 0,444 ☐ 0,510

9. Egy ház egyik 5×3m-es (téglalap alakú) oldalfalán két $1{,}5 \times 1{,}5$m-es ablak van. Egy darázs véletlenszerűen rászáll a ház falára. Mi a valószínűsége, hogy az egyik ablakra fog rászállni?

▷ Stop.

Pontok:

II. MODUL

A valószínűségi változó és jellemzői

6. LECKE

A valószínűségi változó

6. A valószínűségi változó

Sok esetben egy kísérlet elvégzése során az elemi eseményekhez egy számérték is tartozik: például a kockadobásnál a pontok száma. Léteznek azonban olyan kísérletek is, ahol az eseménytér nem számokból áll: ha például arra vagyunk kíváncsiak, hogy mennyi annak a valószínűsége, hogy egy kétgyerekes családban mindkét gyerek lány, akkor az eseménytér az $\Omega = \{FF, FL, LF, LL\}$ halmaz (F a fiút, L a lányt jelöli). Ilyenkor is megtehetjük azonban, hogy minden elemi eseményhez hozzárendelünk egy számot: rendeljük hozzá mondjuk ebben az esetben minden elemi eseményhez a lányok számát (vagyis az FF-hez a 0-t, az FL, illetve LF-hez az 1-et, míg a LL-hez a 2-t rendeljük hozzá). Ily módon ebben az esetben is minden egyes elemi eseményhez tartozik pontosan egy számérték. Mivel ezen számérték megjelenése az elemi esemény bekövetkezésétől függ, ezért beszélhetünk a számérték bekövetkezéséről is.

6.1. A valószínűségi változó

6.1. definíció: Ha egy kísérlettel kapcsolatos elemi események mindegyikéhez egyértelműen hozzárendelünk egy-egy valós számot, akkor az elemi események Ω halmazán egy $X : \Omega \to \mathbb{R}$ függvényt értelmezünk. Ezt a függvényt valószínűségi változónak nevezzük.

Az, hogy milyen számot rendelünk az egyes elemi eseményekhez, valójában csak tőlünk függ, így- mint ahogy arra mindjárt példát is nézünk- ugyanahhoz a kísérlethez több valószínűségi változót is megadhatunk. A következő két dolgot érdemes azonban szem előtt tartani: ha az egyes elemi eseményekkel együtt valamilyen számértékek is adódnak és ezek mindegyike fontos a további vizsgálatok szempontjából, akkor érdemes az egyes elemi eseményekhez a velük együtt megjelenő értékeket rendelni. Ellenben, ha az elemi eseményekhez közvetlenül nem tartozik számérték, akkor az eseményekhez rendelt számértékeket úgy célszerű megválasztani, hogy a további elemzés, számolás során ne okozzanak felesleges bonyodalmakat.

Példa. Feldobunk két dobókockát. Ha számunkra csak az érdekes, hogy a dobott számok összege páros-e, akkor választhatjuk a következő valószínűségi változót:

$$X_1 = \begin{cases} 0, & \text{ha a dobott számok összege páratlan;} \\ 1, & \text{ha a dobott számok összege páros.} \end{cases}$$

Ha az érdekel minket, hogy hány hatost dobtunk, akkor választhatjuk a következő valószínűségi változót:

$$X_2 = \begin{cases} 0, & \text{ha nem dobtunk hatost;} \\ 1, & \text{ha 1 hatost dobtunk;} \\ 2, & \text{ha 2 hatost dobtunk.} \end{cases}$$

Végezetül ha számunkra az a fontos, hogy a két szám közül mekkora a nagyobbik, akkor tekinthetjük a következő valószínűségi változót:

$$X_2 = \begin{cases} 1, & \text{ha a dobott számok maximuma az 1;} \\ 2, & \text{ha a dobott számok maximuma a 2;} \\ 3, & \text{ha a dobott számok maximuma az 3;} \\ 4, & \text{ha a dobott számok maximuma a 4;} \\ 5, & \text{ha a dobott számok maximuma az 5;} \\ 6, & \text{ha a dobott számok maximuma a 6.} \end{cases}$$

Aktivitás: Írjon fel a füzetébe további két valószínűségi változót ugyanezzel a kísérlettel kapcsolatban!

Ha egy kísérlet során az a_i elemi esemény következik be és a valószínűségi változó megadásakor ehhez az x_i értéket rendeltük, akkor azt mondjuk, hogy az X valószínűségi változó az x_i értéket veszi fel, az x_i -t pedig az X egy lehetséges értékének nevezzük. Például a fenti X_1 valószínűségi változó a 0 és 1 értékeket veheti fel.

6.1.1. Diszkrét valószínűségi változó

6.2. definíció: Egy valószínűségi változót diszkrétnek nevezünk, ha a lehetséges értékeinek halmaza megszámlálható (azaz véges sok vagy megszámlálhatóan végtelen sok lehetséges értéke van).

Példa. Kitöltünk egy ötöslottó szelvényt. Legyen az X valószínűségi változó a találataink száma. Ekkor X lehetséges értékei: 0,1,2,3,4,5, így a lehetséges értékek száma véges.

Példa. Tamás és apukája megbeszélik, hogy minden este játszanak egy sakkpartit egészen addig, amíg Tamás nem nyer. Legyen a valószínűségi változó értéke, hogy hány partit fognak lejátszani. Ekkor a lehetséges értékek halmaza végtelen: $1,2\ldots$ (nincs felső korlát).

Ha egy Ω eseménytéren értelmezett X diszkrét valószínűségi változó lehetséges értékei az x_k, $(k = 1,2,\ldots)$ valós számok, és egy esemény bekövetkezésekor az X értéke x_k , akkor azt mondjuk, hogy az $\{X = x_k\}$ esemény következik be, és az esemény bekövetkezési valószínűségét $p_k = P(X = x_k)$-vel jelöljük.

6.3. definíció: Az X diszkrét valószínűségi változó lehetséges értékeihez tartozó bekövetkezési valószínűségek összességét X eloszlásának nevezzük. Azaz, ha X lehetséges értékei az x_k , $(k = 1,2,\ldots)$ számok, akkor X eloszlása a $p_k = P(X = x_k)$ $(k = 1,2,\ldots)$ bekövetkezési valószínűségek összessége.

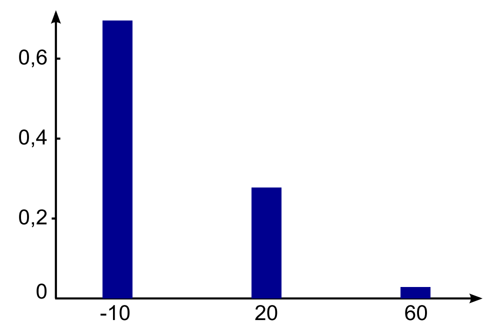
Egy diszkrét valószínűségi változó eloszlásának megadásához tehát egyrészt meg kell mondani, hogy milyen értékekeket vehet fel a valószínűségi változó, másrészt meg kell adni, hogy az egyes értékeket milyen valószínűséggel veszi fel.

6.1. feladat. András és Béla a következő játékot játsszák: mindketten feldobnak egy dobókockát. Ha egyikük sem dobott hatost, akkor Béla fizet Andrásnak 10 Ft-ot. Ha pontosan az egyikük dob hatost, akkor András fizet Bélának 20 Ft-ot, ha pedig mindketten hatost dobnak, akkor András 60 Ft-ot fizet Bélának. Legyen X Béla nyereménye. Írjuk fel X eloszlását.

Megoldás: A valószínűségi változó lehetséges értékei: $-10, 20, 60$. Számoljuk ki, hogy milyen valószínűségekkel veszi fel X ezeket az értékeket!
Béla akkor fizet Andrásnak 10 Ft-ot, ha egyikük sem dob hatost. Ennek a valószínűsége $\frac{25}{36}$ (hiszen az összes esetek száma $6 \cdot 6$, míg a jó esetek száma $5 \cdot 5$, így $P(X=-10) = \frac{25}{36}$.
Béla akkor nyer Andrástól 10 Ft-ot, ha pontosan egyikük dob hatost. Mivel ez tízféleképpen fordulhat elő, ezért ennek a valószínűsége $\frac{10}{36}$, így $P(X=10) = \frac{10}{36}$.
Végezetül Béla akkor nyer 60 Ft-ot, ha mindketten hatost dobnak. Ez csak egyféleképpen fordulhat elő, így ennek a valószínűsége $\frac{1}{36}$, azaz $P(X=60) = \frac{1}{36}$. A fentieket összefoglalva X eloszlását megadhatjuk az alábbi módon (az első sorban X lehetséges értékei, míg a második sorban az adott értékhez tartozó valószínűség szerepel):

$$X : \left\{ \begin{array}{ccc} -10 & 20 & 60 \\ \dfrac{25}{36} & \dfrac{10}{36} & \dfrac{1}{36} \end{array} \right. .$$

Megjegyzés: Arra a kérdésre, hogy kinek kedvező ez a játék, a későbbiekben visszatérünk.



6.1. ábra. *A feladatban szereplő eloszlás.*

⇐6.1. feladat

6.2. feladat. Egy urnában 3 piros és 7 kék golyó van. Kihúzunk egy golyót, majd visszatesszük az urnába, és egészen addig húzunk így visszatevéssel, amíg pirosat nem kapunk. Legyen az X valószínűségi változó a húzások száma. Adjuk meg X eloszlását!

Megoldás: Az X valószínűségi változó most végtelen sok értéket vehet fel, hiszen nincs felső korlátja annak, hogy hányadikra húzunk először pirosat, így X lehetséges értékei $1,2\dots$. Ahhoz, hogy az összes lehetséges értéknél megmondjuk, hogy X milyen valószínűséggel veszi fel az adott értéket, most célszerű egy képletet keresni. Nyilván annak a valószínűsége, hogy elsőre pirosat húzunk, $\frac{3}{10}$. Legyen most k egy 1-nél nagyobb egész szám. Ekkor az, hogy éppen a k. húzásra húzunk először pirosat, azt jelenti, hogy az első $k-1$ húzás során mindig kéket húztunk, míg a k. húzásra pirosat húztunk. Mivel ezek független események, ezért az együttes bekövetkezés valószínűsége egyenlő a külön-külön vett valószínűségek szorzatával, azaz

$$\begin{aligned} &P(k\text{-ra húzunk először pirosat}) = P(\text{elsőre kéket húzunk}) \cdot P(\text{másodikra kéket húzunk}) \cdot \ldots \\ &\cdot P((k-1)\text{-re kéket húzunk}) \cdot P(k\text{-ra pirosat húzunk}) = \underbrace{\frac{7}{10} \cdot \frac{7}{10} \cdot \ldots \cdot \frac{7}{10}}_{(k-1)\text{ darab}} \cdot \frac{3}{10} = \left(\frac{7}{10}\right)^{k-1} \cdot \frac{3}{10}. \end{aligned}$$

Vegyük észre, hogy a képlet $k = 1$ esetén is igaz, hiszen annak a valószínűsége, hogy elsőre pirosat húzunk, valóban

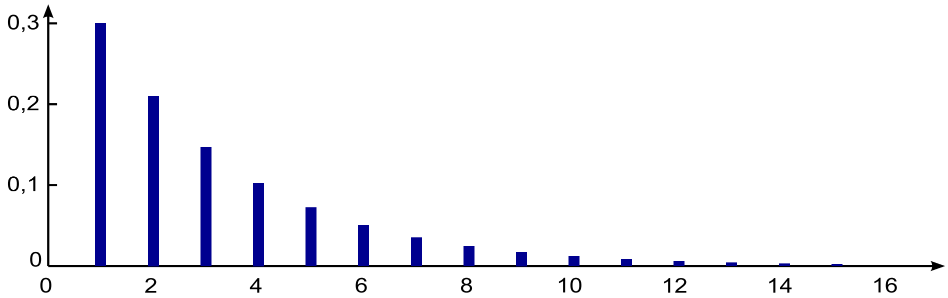
$$\left(\frac{7}{10}\right)^0 \cdot \frac{3}{10} = \frac{3}{10}.$$

Vagyis az X valószínűségi változó eloszlását a

$$P(X = k) = \left(\frac{7}{10}\right)^{k-1} \cdot \frac{3}{10} \qquad k = 1,2\ldots$$

valószínűségek határozzák meg. A 6.2. ábrán az eloszlás ábrázolását láthatjuk. ⇐6.2. feladat

Megjegyzés: A nemsokára ismertetésre kerülő nevezetes diszkrét valószínűségi változók eloszlását is képlet segítségével fogjuk megadni.



6.2. ábra. *A 6.2. feladatbeli eloszlás.*

6.1.2. Folytonos valószínűségi változó

Nem minden valószínűségi változónál tudjuk a lehetséges értékeket felsorolni. Például legyen az X valószínűségi változó értéke egy újonnan átadott négyemeletes ház esetén az első beázásig eltelt idő napokban számolva (ez természetesen lehet nem egész szám is). Ekkor a lehetséges értékek halmaza a $[0,\infty)$ félegyenes, így a valószínűségi változónak nem megszámlálhatóan végtelen sok értéke van.

6.4. definíció: Egy valószínűségi változót folytonosnak nevezünk, ha a lehetséges értékei egy vagy több intervallumot alkotnak (lehet nem korlátos is).

A folytonos valószínűségi változókat a hamarosan ismertetésre kerülő eloszlás-, illetve sűrűségfüggvényükkel lehet megadni.

Aktivitás: Találjon ki még legalább két folytonos valószínűségi változót!

6.2. Az eloszlásfüggvény

6.5. definíció: Az X valószínűségi változó eloszlásfüggvényének nevezzük azt az F függvényt, amely minden valós x értékhez hozzárendeli annak valószínűségét, hogy az X valószínűségi változó x-nél kisebb értéket vesz fel, azaz

$$F(x) = P(X < x) \qquad x \in \mathbb{R}.$$

6.3. feladat. Adott az X diszkrét valószínűségi változó eloszlása:

$$X : \begin{cases} -1 & 2 & 5 \\ 0{,}1 & 0{,}4 & 0{,}5 \end{cases}.$$

Írjuk fel, és ábrázoljuk x eloszlásfüggvényét!

Megoldás:

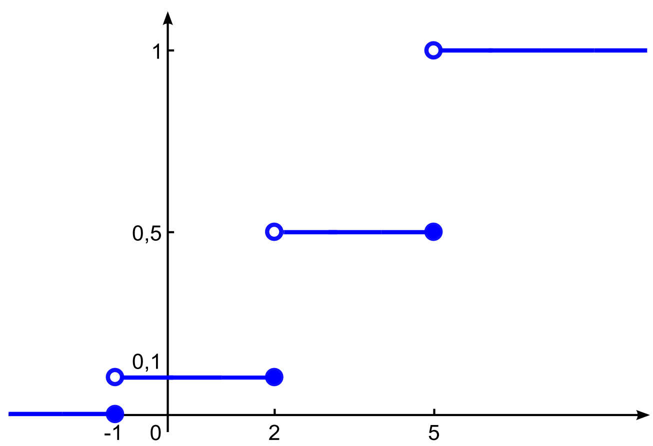
- Számoljuk ki a definíció alapján az eloszlásfüggény értékét néhány pontban! Definíció szerint $F(-2)$ annak a valószínűségével egyenlő, hogy a valószínűségi változó -2-nél kisebb értéket vesz fel. Mivel X sosem vesz fel -2-nél kisebb értéket, ezért $F(-2) = 0$. Hasonlóan kapjuk, hogy $F(-1{,}6) = 0$, sőt $F(x) = 0$ minden olyan x értékre, melyre $x \leq -1$.
- Definíció szerint $F(0)$ annak a valószínűségét adja meg, hogy X 0-nál kisebb értéket vesz fel. Ez csak úgy lehet, ha X a -1 értéket veszi fel. Mivel ennek a valószínűsége $0{,}1$, ezért $F(0) = 0{,}1$. Hasonlóan kapjuk, hogy $F(1{,}5) = 0{,}1$, sőt $F(x) = 0{,}1$ minden olyan x-re, melyre $-1 < x \leq 2$ teljesül.
- Definíció szerint $F(3)$ annak a valószínűségét adja meg, hogy X 3-nál kisebb értéket vesz fel. Ez úgy lehet, ha X a -1 vagy 2 értékeket veszi fel, ennek a valószínűsége $0{,}1+0{,}4 = 0{,}5$. Ezért $F(3) = 0{,}5$, sőt hasonlóan belátható, hogy $F(x) = 0{,}5$ minden olyan x esetén, melyre $2 < x \leq 5$ teljesül.

– $F(7) = 1$, hiszen X biztos, hogy 7-nél kisebb értéket fog felvenni. Hasonlóan adódik, hogy $F(x) = 1$ minden $5 < x$ esetén.

Összefoglalva: az X valószínűségi változó eloszlásfüggvénye a következő alakban adható meg:

$$F(x) = \begin{cases} 0, & \text{ha} \quad x \le -1; \\ 0{,}1, & \text{ha} \quad -1 < x \le 2; \\ 0{,}5, & \text{ha} \quad 2 < x \le 5; \\ 1, & \text{ha} \quad 5 < x. \end{cases}$$

⇐6.3. feladat



6.3. ábra. *Diszkrét eloszlású valószínűségi változó eloszlásfüggvénye „lépcsős” függvény (az ábrán a 6.3. feladatban meghatározott eloszlásfüggvény látható).*

Az eloszlásfüggvény definíciójából következnek az alábbi tulajdonságok:

6.1. tétel: Legyen X egy tetszőleges valószínűségi változó, az eloszlásfüggvénye pedig legyen $F(x)$. Ekkor $F(x)$ rendelkezik az alábbi tulajdonságokkal:

1. $0 \leq F(x) \leq 1$.
2. Monoton növő, azaz ha $a < b$, akkor $F(a) \leq F(b)$.
3. $\lim\limits_{x \to -\infty} F(x) = 0$ és $\lim\limits_{x \to \infty} F(x) = 1$.
4. $F(x)$ minden pontban balról folytonos, azaz $\lim\limits_{x \to a-0} F(x) = F(a)$.

Bizonyítás:

1. Mivel az eloszlásfüggvény értéke egy esemény valószínűségét adja meg, így $0 \leq F(x) \leq 1$.
2. Ha $a < b$, akkor $\{X < b\} = \{X < a\} + \{a \leq X < b\}$. Ezzel az $\{X < b\}$ eseményt egymást kizáró események összegeként írtuk fel, így a 3. axióma miatt a valószínűsége ezen két esemény valószínűségének az összege, azaz $P(X < b) = P(X < a) + P(a \leq X < b)$. Az eloszlásfüggvény definíciója alapján ez átírható $F(b) = F(a) + P(a \leq X < b)$ alakba. Mivel $P(a \leq X < b) \geq 0$, ezért $F(a) \leq F(b)$.

A 3. és a 4. tulajdonságot nem bizonyítjuk.

$\square$

Az eloszlásfüggvény definíciója alapján, ha az X valószínűségi változó eloszlásfüggvénye $F(x)$, akkor az $\{X < a\}$ esemény valószínűsége $F(a)$. Az alábbi tétel azt mutatja meg, hogy az $\{X \geq a\}$, $\{a \leq X < b\}$, $\{X > a\}$ és az $\{X = a\}$ események valószínűsége hogyan számítható ki az eloszlásfüggvény segítségével.

6.2. tétel: Legyen az X valószínűségi változó eloszlásfüggvénye $F(x)$. Ekkor

1. $P(X \geq a) = 1 - F(a)$.
2. $P(a \leq X < b) = F(b) - F(a)$.
3. $P(X > a) = 1 - \lim_{x \to a+0} F(x)$.
4. $P(X = a) = \lim_{x \to a+0} F(x) - F(a)$.

Bizonyítás:

1. Az $\{X \geq a\}$ esemény az $\{X < a\}$ esemény komplementere, ezért
$$P(X \geq a) = 1 - P(X < a) = 1 - F(a).$$
2. Az $\{a \leq X < b\}$ esemény az $\{X < b\}$ és az $\{X < a\}$ események különbsége, ráadásul e két esemény között fennáll az $\{X < a\} \subset \{X < b\}$ reláció is. Az események különbségének valószínűségére vonatkozó összefüggés alapján kapjuk, hogy
$$P(a \leq X < b) = P(X < b) - P(X < a) = F(b) - F(a).$$
3. $P(X > a) = \lim_{n \to \infty} P\left(X \geq a + \frac{1}{n}\right) = \lim_{n \to \infty} \left(1 - P\left(X < a + \frac{1}{n}\right)\right) = 1 - \lim_{x \to a+0} F(x)$.
4. Az $\{X = a\}$ esemény az $\{X \geq a\}$ és az $\{X > a\}$ események különbsége, továbbá teljesül az is, hogy $\{X > a\} \subset \{X \geq a\}$. Az események különbségének valószínűségére vonatkozó összefüggés és a tételbeli 1. és 3. pontok alapján kapjuk, hogy
$$P(X = a) = P(X \geq a) - P(X > a) = 1 - F(a) - \left(1 - \lim_{x \to a+0} F(x)\right) = \lim_{x \to a+0} F(x) - F(a).$$

□

Nézzük meg most egy konkrét példán, hogyan lehet az eloszlásfüggvény segítségével különféle valószínűségeket kiszámítani!

6.4. feladat. Adott az X valószínűségi változó eloszlásfüggvénye:

$$F(x) = \begin{cases} 1 - \frac{8}{x^3} & \text{ha } 2 < x \\ 0 & \text{különben} \end{cases}.$$

Határozzuk meg a következő valószínűségeket:

a) $P(X < 1)$ *b)* $P(X < 5)$ *c)* $P(X = 6)$
d) $P(X \leq 4)$ *e)* $P(X \geq 6)$ *f)* $P(X > 3)$
g) $P(6 \leq X < 9)$ *h)* $P(X > 8|X > 6)$ *i)* $P(X < 9|X > 3)$

Megoldás:

a) Az eloszlásfüggvény definíciója szerint $P(X < 1) = F(1) = 0$.

b) Ismét a definíciót használva $P(X < 5) = F(5) = 1 - \frac{8}{5^3} = 0{,}936$.

c) A 6.2. tétel szerint $P(X = 6) = \lim_{x \to 6+0} F(x) - F(6)$. Vegyük észre, hogy az eloszlásfüggvény folytonos, ezért $\lim_{x \to 6+0} = F(6)$, amiből

$$P(X = 6) = \lim_{x \to 6+0} F(x) - F(6) = F(6) - F(6) = 0.$$

Hasonlóan belátható, hogy ha az X valószínűségi változó eloszlásfüggvénye folytonos, akkor tetszőleges a valós számra $P(X = a) = 0$ teljesül.

d) Az előző pontbeli megjegyzésünk miatt $P(X=4)=0$. Ezt felhasználva azt kapjuk, hogy

$$P(X \le 4) = P(X < 4) + P(X = 4) = F(4) + 0 = 1 - \frac{8}{4^3} = 0{,}984375.$$

e) Az $X \ge 6$ esemény komplementere az $X < 6$ esemény. Használjuk fel a $P(A) = 1 - P(\overline{A})$ összefüggést:

$$P(X \ge 6) = 1 - P(X < 6) = 1 - F(6) = 1 - \left(1 - \frac{8}{6^3}\right) \approx 0{,}037.$$

f)

$$P(X > 3) = 1 - P(X \le 3) = 1 - [P(X < 3) + P(X = 3)] = 1 - (F(3) + 0) = 1 - \left(1 - \frac{8}{3^3}\right) \approx 0{,}296.$$

g) A 6.2. tétel szerint $P(a \le X < b) = F(b) - F(a)$, így

$$P(6 \le X < 9) = F(9) - F(6) = 1 - \frac{8}{9^3} - \left(1 - \frac{8}{6^3}\right) \approx 0{,}0033.$$

h) Vigyázzunk arra, hogy most egy feltételes valószínűséget kell kiszámítanunk! A feltételes valószínűség definíciója szerint:

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

A feltételes valószínűség definícióját, illetve a korábbi pontok eredményeit felhasználva:

$$P(X > 8 | X > 6) = \frac{P(X > 8 \text{ és } X > 6)}{P(X > 6)} = \frac{P(X > 8)}{P(X > 6)} = \frac{1 - F(8)}{1 - F(6)} = \frac{1 - \left(1 - \frac{8}{8^3}\right)}{1 - \left(1 - \frac{8}{6^3}\right)} \approx 0{,}422.$$

i)

$$P(X<9|X>3)=\frac{P(X<9 \text{ és } X>3)}{P(X>3)}=\frac{P(3<X<9)}{P(X>3)}=\frac{F(9)-F(3)}{1-F(3)}=\frac{1-\frac{8}{9^3}-\left(1-\frac{8}{3^3}\right)}{1-\left(1-\frac{8}{3^3}\right)}\approx 0{,}963.$$

⇐6.4. feladat

6.5. feladat. Egy újonnan átadott útszakasz állapotát különféle paraméterek alapján (pl. nyomvályú) rendszeresen osztályozzák. Ha az útszakasz állapota egy bizonyos szint alá süllyed, akkor valamilyen beavatkozást kell végrehajtani (pl. fel kell újítani). Az X valószínűségi változó legyen az első beavatkozásig eltelt idő évben számolva (természetesen X nem csak egész értékekeket vehet fel). Tegyük fel, hogy X eloszlásfüggvénye:

$$F(x)=\begin{cases} 0, & \text{ha } x\le 0; \\ \dfrac{x^3}{216}, & \text{ha } 0<x\le 6; \\ 1, & \text{ha 6<x.} \end{cases}$$

Mennyi a valószínűsége, hogy az első öt évben nem lesz szükség beavatkozásra?

Megoldás: Az, hogy az első 5 évben nincs szükség beavatkozásra, azt jelenti, hogy az első beavatkozásig eltelt idő legalább 5 év, azaz az $X\ge 5$ esemény valószínűségére vagyunk kíváncsiak. A korábbi ismereteinket felhasználva:

$$P(\text{első 5 évben nem kell beavatkozni})=P(X\ge 5)=1-F(5)=1-\frac{5^3}{216}\approx 0{,}421.$$

Vagyis körülbelül 42 százalék annak az esélye, hogy az első 5 évben nem kell semmilyen beavatkozást végrehajtani.

⇐6.5. feladat

Megjegyzés: Sok esetben az X valószínűségi változó egy alkatrész, termék, gép stb. élettartamát adja meg. Ekkor annak a valószínűsége, hogy a gép még t idő elteltével is működik, $1 - F(t)$-vel egyenlő, ahol F az X eloszlásfüggvénye. Éppen ezért ilyenkor a $G(t) = 1 - F(t)$ függvényt szokás túlélési függvénynek nevezni.

Önellenőrzés

1. Az $1,2\ldots,10$ számok közül véletlenszerűen kiválasztunk egyet. Az X valószínűségi változó értéke legyen a kiválasztott szám 6-tal való osztás utáni maradéka. Milyen értéket vesz fel az F eloszlásfüggvény a $3{,}5$ helyen, azaz mivel egyenlő $F(3{,}5)$? Megold.

2. Az X valószínűségi változó eloszlásfüggvénye:

$$F(x) = \begin{cases} 0, & \text{ha} \quad x \leq 0; \\ \dfrac{x^2+2x}{24}, & \text{ha} \quad 0 < x \leq 4; \\ 1, & \text{ha} \quad 4 < x. \end{cases}$$

Mennyi a $P(X > 5)$ valószínűség értéke? Megold.

3. A 2. önellenőrző feladat esetén mennyi a $P(2 < X < 3)$ valószínűség értéke?

☐ $\frac{17}{24}$ ☐ $\frac{15}{24}$ ☐ $\frac{7}{24}$ ☐ $\frac{9}{24}$

4. Az X valószínűségi változó eloszlásfüggvénye:

$$F(x) = \begin{cases} 0, & \text{ha} \quad x \leq 0; \\ 1 - e^{-2x}, & \text{ha} \quad 0 < x. \end{cases}$$

Mennyi a $P(X > 1)$ valószínűség értéke?

☐ 0,135 ☐ 0,368 ☐ 0,865 ☐ 0,632

5. A 4. önellenőrző kérdés esetén mennyi a $P(X > 2|X > 1)$ valószínűség értéke?

☐ 0,018 ☐ 0,245 ☐ 0,865 ☐ 0,135

7. LECKE

A sűrűségfüggvény

6.3. A sűrűségfüggvény

A valószínűségi változók között külön figyelmet érdemelnek azok, melyek eloszlásfüggvénye előáll egy másik függvény integráljaként.

6.6. definíció: Az X valószínűségi változót folytonos eloszlásúnak nevezzük, ha az eloszlásfüggvénye ($-\infty$-hez tartozó) integrálfüggvény, azaz van olyan f függvény, amelyre

$$\int_{-\infty}^{x} f(t)\,\mathrm{d}t = F(x) \qquad x \in \mathbb{R}.$$

A fenti definíció szerint adott F-hez keresünk megfelelő f-et. Azonban a definícióban szereplő f korántsem egyértelmű, hiszen ha egy, a feltételnek megfelelő f függvény értékeit megszámlálható (véges vagy megszámlálhatóan végtelen) sok helyen megváltoztatjuk, akkor az az integrál értékét nem befolyásolja. A definíciónak megfelelő f függvények közül olyat célszerű választani, ami a legkevesebb helyen nem folytonos. Tudjuk azt is, hogy abszolút folytonos függvény majdnem mindenütt differenciálható és egyenlő a deriváltjának határozatlan integráljával, így abszolút folytonos F esetén a $f = F'$ megfelel a fenti követelményeknek. A kérdés már csak az, hogy mit tegyünk azokban a pontokban, ahol F nem differenciálható. Megtehetnénk azt is, hogy ezeken a helyeken f-et nem definiáljuk, sőt akár tetszőleges értéket is adhatnánk neki, az az integrál értékét nem fogja befolyásolni. Mi itt azt a (műszaki alkalmazásokban elterjedt) megoldást választjuk, hogy ezeken a helyeken f értékét 0-nak definiáljuk.

6.7. definíció: Legyen az X valószínűségi változó eloszlásfüggvénye $F(x)$. Ha $F(x)$ abszolút folytonos, akkor legyen $f(x) = F'(x)$, ha pedig egy pontban $F(x)$ nem differenciálható, ott $f(x)$ értéke legyen 0. Az így definiált $f(x)$ függvényt az X valószínűségi változó sűrűségfüggvényének nevezzük.

A sűrűségfüggvény definíciójából következnek az alábbi tulajdonságok:

6.3. tétel: Legyen X egy tetszőleges folytonos eloszlású valószínűségi változó, az eloszlásfüggvénye legyen $F(x)$, a sűrűségfüggvénye pedig legyen $f(x)$. Ekkor $f(x)$ rendelkezik az alábbi tulajdonságokkal:

1. Nem negatív, azaz $f(x) \geq 0$.
2. $\int\limits_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$.

Bizonyítás:

1. Mivel az eloszlásfüggvény monoton növő, ezért deriváltja nem negatív azokon a helyeken, ahol F differenciálható, a többi pontban pedig a definíció szerint 0 a sűrűségfüggvény értéke.
2. Mivel a sűrűségfüggvény olyan függvény, amely megfelel a 6.6. definíció feltételének, ezért

$$\int\limits_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \lim_{x\to\infty} \int\limits_{-\infty}^{x} f(x)\,\mathrm{d}x = \lim_{x\to\infty} F(x) = 1\,.$$

□

Látható, hogy a sűrűségfüggvény hasonlóan viselkedik folytonos eloszlású valószínűségi változóknál, mint az eloszlás a diszkrét eloszlású valószínűségi változóknál. Ott az eloszlásban szereplő valószínűségek (melyek nyilvánvalóan nem negatívak) összege 1, míg itt a sűrűségfüggvény nem negatív és a teljes számegyenesen vett integrálja 1. Ha X diszkrét eloszlású valószínűségi változó, akkor az $\{X < a\}$, $\{X \geq a\}$ és az $\{a \leq X < b\}$ események valószínűsége a valószínűségi változónak a relációnak megfelelő értékeihez tartozó valószínűségek összegeként áll elő. Ezzel teljesen analóg módon határozhatóak meg a kérdéses valószínűségek folytonos eloszlású X esetén.

Az alábbi tétel azt mutatja meg, hogy az $\{X < a\}$, $\{X \geq a\}$ és az $\{a \leq X < b\}$ események valószínűsége hogyan számítható ki a sűrűségfüggvény segítségével.

6.4. tétel: Legyen X egy tetszőleges folytonos eloszlású valószínűségi változó, az eloszlásfüggvénye legyen $F(x)$, a sűrűségfüggvénye pedig legyen $f(x)$. Ekkor

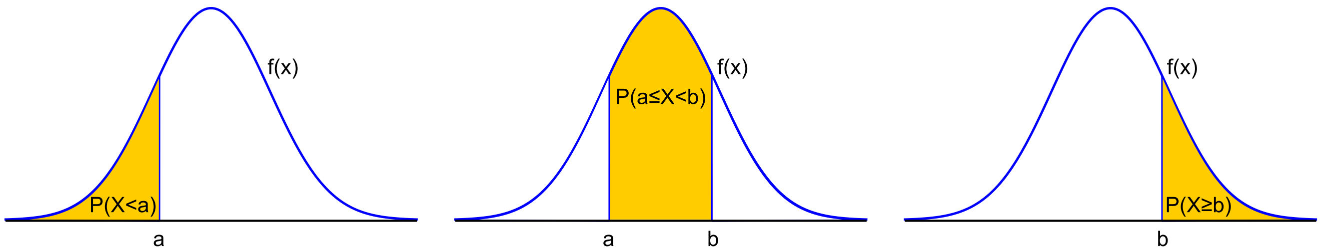
1. $P(X < a) = \int\limits_{-\infty}^{a} f(x)\,\mathrm{d}x.$
2. $P(X \geq a) = \int\limits_{a}^{\infty} f(x)\,\mathrm{d}x.$
3. $P(a \leq X < b) = \int\limits_{a}^{b} f(x)\,\mathrm{d}x.$

Bizonyítás:

1. $P(X < a) = F(a) = \int\limits_{-\infty}^{a} f(x)\,\mathrm{d}x.$
2. $P(X \geq a) = 1 - F(a) = \int\limits_{-\infty}^{\infty} f(x)\,\mathrm{d}x - \int\limits_{-\infty}^{a} f(x)\,\mathrm{d}x = \int\limits_{a}^{\infty} f(x)\,\mathrm{d}x.$
3. $P(a \leq X < b) = F(b) - F(a) = \int\limits_{-\infty}^{b} f(x)\,\mathrm{d}x - \int\limits_{-\infty}^{a} f(x)\,\mathrm{d}x = \int\limits_{a}^{b} f(x)\,\mathrm{d}x.$

A 6.3. tétel első pontja azt jelenti, hogy egy folytonos valószínűségi változó sűrűségfüggvénye nem vehet fel negatív értékeket, míg a második pont szerint a sűrűségfüggvénynek a mínusz végtelentől a végtelenig vett improprius integrálja 1-gyel egyenlő.

A 6.4. tétel szerint a $P(X < a)$, $P(X \geq b)$ és $P(a \leq X < b)$ valószínűségek értékét úgy kaphatjuk meg, hogy a sűrűségfüggvényt rendre integráljuk a megfelelő intervallumon. Ismert, hogy a határozott integrál a görbe alatti területet adja meg, így a $P(X < a)$, $P(X \geq b)$ és $P(a \leq X < b)$ valószínűségek értékei rendre megegyeznek a $(-\infty,a)$, (b,∞), és (a,b) intervallumokon a sűrűségfüggvény alatti terület mérőszámával.



6.4. ábra. *Valószínűségek számolása a sűrűségfüggvénnyel.*

A fentiek szerint, ha adott egy folytonos eloszlású valószínűségi változó sűrűségfüggvénye, akkor a különféle valószínűségek meghatározásához határozott, illetve improprius integrálokat kell kiszámítanunk, ezért szükségünk lesz az integrálási ismereteink felfrissítésére.

Aktivitás: Keresse meg és jegyzetelje ki a füzetébe az alapintegrálokat, illetve az integrálási szabályokat!

Érdekesség: A sűrűség fogalmával mindenki találkozott már az általános iskolai fizikaórákon. A mérnök hallgatók a fizika, kémia, mechanika tárgyakban további sűrűségfogalmakkal (pl. áramsűrűség) fognak találkozni.

6.6. feladat. Az X valószínűségi változó eloszlásfüggvénye az alábbi függvény:

$$F(x) = \begin{cases} 0, & \text{ha} \quad x \le 0; \\ \dfrac{\sqrt{x}}{2}, & \text{ha} \quad 0 < x \le 4; \\ 1, & \text{ha} \quad 4 < x. \end{cases}$$

Határozzuk meg X sűrűségfüggvényét!

Megoldás: Tudjuk, hogy a sűrűségfüggvény az eloszlásfüggvény deriváltja, ezért gyakorlatilag az $F'(x)$ függvényt kell meghatároznunk.

Mivel a $(-\infty,0)$ intervallumon F konstans, ezért $x < 0$ esetén $f(x) = F'(x) = 0$.

Ha $0 < x < 4$, akkor

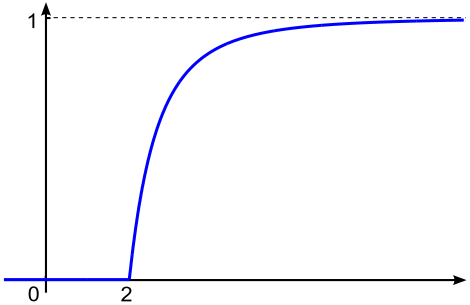
$$f(x) = F'(x) = \left(\frac{\sqrt{x}}{2}\right)' = \left(\frac{1}{2}x^{\frac{1}{2}}\right)' = \frac{1}{2} \cdot \frac{1}{2} \cdot x^{\frac{-1}{2}} = \frac{1}{4 \cdot \sqrt{x}}.$$

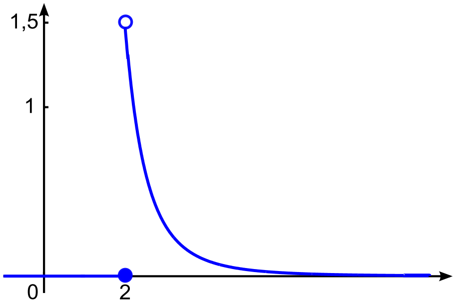
A $(4,\infty)$ intervallumon F konstans, ezért $4 < x$ esetén $f(x) = F'(x) = 0$. Az $x = 0$ és az $x = 4$ helyeken $F(x)$ nem deriválható, így a definíció szerint itt a sűrűségfüggvény értéke 0.

A fentieket összefoglalva X sűrűségfüggvénye az alábbi módon írható fel:

$$f(x) = \begin{cases} \dfrac{1}{4 \cdot \sqrt{x}}, & \text{ha } 0 < x < 4; \\ 0 & \text{különben.} \end{cases}$$

⇐6.6. feladat





6.5. ábra. *A 6.7. feladatban szereplő eloszlásfüggvény (bal oldal) és sűrűségfüggvény (jobb oldal).*

6.7. feladat. Az X valószínűségi változó eloszlásfüggvénye az alábbi függvény:

$$F(x) = \begin{cases} 0, & \text{ha} \quad x \leq 2; \\ 1 - \dfrac{8}{x^3}, & \text{ha} \quad 2 < x. \end{cases}$$

Határozzuk meg X sűrűségfüggvényét!

Megoldás: Tudjuk, hogy a sűrűségfüggvény az eloszlásfüggvény deriváltja, ezért gyakorlatilag az $F'(x)$ függvényt kell meghatároznunk. Mivel a $(-\infty,2)$ intervallumon F konstans, ezért $x < 2$ esetén $f(x) = F'(x) = 0$.

$F(x)$ nem deriválható az $x = 2$ helyen, ezért ott a sűrűségfüggvénye a definíció szerint 0. Ha $2 < x$, akkor

$$f(x) = F'(x) = \left(1 - \frac{8}{x^3}\right)' = \left(1 - 8 \cdot x^{-3}\right)' = -8 \cdot -3 \cdot x^{-4} = \frac{24}{x^4}\,.$$

A fentieket összefoglalva X sűrűségfüggvénye az alábbi módon írható fel:

$$f(x) = \begin{cases} 0, & \text{ha} \quad x \leq 2; \\ \dfrac{24}{x^4}, & \text{ha} \quad 2 < x. \end{cases}$$

⇐6.7. feladat

A 6.3. tétel szerint a sűrűségfüggvény csak nemnegatív értékeket vehet fel, és mínusz végtelentől végtelenig vett integrálja 1-gyel egyenlő. Belátható, hogy az állítás megfordítása is igaz, azaz ha egy, a valós számokon értelmezett függvény csak nemnegatív értékeket vesz fel, és a függvénynek a $(-\infty,\infty)$ intervallumon vett integrálja eggyel egyenlő, akkor a függvény egy valószínűségi változó sűrűségfüggvényének tekinthető.

6.8. feladat. Határozzuk meg a c paraméter értéket úgy, hogy a következő $f(x)$ függvény egy valószínűségi változó sűrűségfüggvénye legyen!

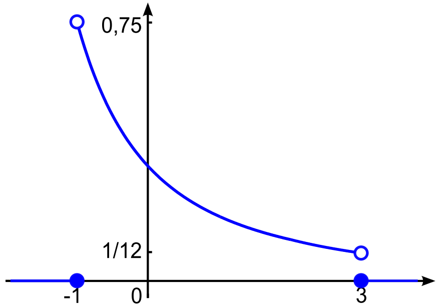
$$f(x) = \begin{cases} \dfrac{c}{(x+3)^2}, & \text{ha } -1 < x < 3; \\ 0 & \text{különben.} \end{cases}$$

Megoldás: Ahhoz, hogy a megadott f függvény sűrűségfüggvény legyen, egyrészt annak kell teljesülnie, hogy $f(x) \geq 0$ minden x esetén, másrészt f-nek a teljes számegyenesen vett integráljának 1-gyel kell egyenlőnek lennie.

Mivel $1/(x+3)^2$ sosem vehet fel negatív értéket, ezért az első feltételből azt kapjuk, hogy $c \geq 0$. A második feltételből kiszámítható c értéke. A feltétel szerint $\displaystyle\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$. Az integálásnál ügyelnünk kell arra, hogy a sűrűségfüggvény csak a $-1 < x < 3$ esetén egyezik meg $c/(x+3)^2$-nel, minden más x érték esetén $f(x) = 0$. Ezért az integrálban a $(-\infty,\infty)$ intervallumot szétbontjuk 3 részre.

$$1 = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \int_{-\infty}^{-1} f(x)\,\mathrm{d}x + \int_{-1}^{3} f(x)\,\mathrm{d}x + \int_{3}^{\infty} f(x)\,\mathrm{d}x = \int_{-\infty}^{-1} 0\,\mathrm{d}x + \int_{-1}^{3} \frac{c}{(x+3)^2}\,\mathrm{d}x + \int_{3}^{\infty} 0\,\mathrm{d}x =$$
$$= 0 + \left[\frac{-c}{x+3}\right]_{-1}^{3} + 0 = \frac{-c}{6} + \frac{c}{2} = \frac{1}{3} \cdot c$$

Azt kaptuk tehát, hogy $1 = c/3$, vagyis $c = 3$ esetén lesz a megadott függvény sűrűségfüggvény. ⇐6.8. feladat



6.6. ábra. *A 6.8. feladatban szereplő sűrűségfüggvény.*

6.9. feladat. Határozzuk meg a c paraméter értéket úgy, hogy a következő $f(x)$ függvény egy valószínűségi változó sűrűségfüggvénye legyen!

$$f(x) = \begin{cases} c \cdot x^5, & \text{ha } -2 < x < 4; \\ 0 & \text{különben.} \end{cases}$$

Megoldás: Ahhoz, hogy a megadott f függvény sűrűségfüggvény legyen, egyrészt annak kell teljesülnie, hogy $f(x) \geq 0$ minden x esetén, másrészt f-nek a teljes számegyenesen vett integráljának 1-gyel kell egyenlőnek lennie.

Vegyük észre, hogy az x^5 függvény a $(-2{,}4)$ intervallumban pozitív és negatív értékeket is felvesz; ezért ha $c \neq 0$, akkor például $f(-1) = -c$ és $f(2) = 32c$ biztosan különböző előjelű lesz. Azaz $c \neq 0$ esetben f negatív értéket is felvesz, ezért csak a $c = 0$ eset jöhet szóba. Ekkor azonban $f(x) = 0$, így $f(x)$ teljes számegyenesen vett integrálja is 0, vagyis a második feltétel nem teljesül. Azt kaptuk tehát, hogy nem létezik olyan c szám, mellyel az f függvény sűrűségfüggvény lenne.

⇐6.9. feladat

Azt már láttuk, hogy egy folytonos valószínűségi változó eloszlásfüggvényéből hogyan lehet a sűrűségfüggvényt meghatározni. Most nézzük meg, hogyan lehet a sűrűségfüggvényből az eloszlásfüggvényt megkapni!

6.10. feladat. Határozzuk meg az X valószínűségi változó F eloszlásfüggvényét, ha a sűrűségfüggvénye

$$f(x) = \begin{cases} \frac{x^2}{21}, & \text{ha } 1 < x < 4; \\ 0 & \text{különben} \end{cases} \quad !$$

Megoldás: A sűrűségfüggvény definíciója szerint

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t.$$

Ha $x \le 1$, akkor

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_{-\infty}^{x} 0\,\mathrm{d}t = 0.$$

Ha $1 < x \le 4$, akkor

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_{-\infty}^{1} f(t)\,\mathrm{d}t + \int_{1}^{x} f(t)\,\mathrm{d}t = \int_{-\infty}^{1} 0\,\mathrm{d}t + \int_{1}^{x} \frac{t^2}{21}\,\mathrm{d}t = 0 + \left[\frac{t^3}{63}\right]_1^x = \frac{x^3}{63} - \frac{1}{63}.$$

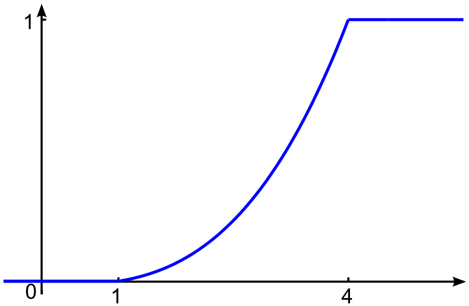
Ha $4 < x$, akkor

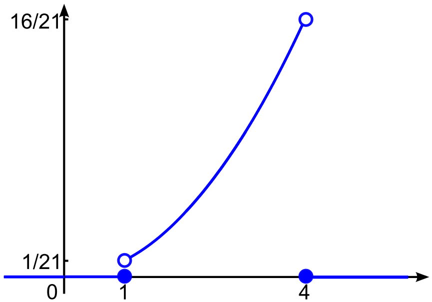
$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_{-\infty}^{1} f(t)\,\mathrm{d}t + \int_{1}^{4} f(t)\,\mathrm{d}t + \int_{4}^{x} f(t)\,\mathrm{d}t = \int_{-\infty}^{1} 0\,\mathrm{d}t + \int_{1}^{4} \frac{t^2}{21}\,\mathrm{d}t + \int_{4}^{x} 0\,\mathrm{d}t =$$
$$= 0 + \left[\frac{t^3}{63}\right]_1^4 + 0 = \frac{4^3}{63} - \frac{1}{63} = 1.$$

Tehát az X valószínűségi változó eloszlásfüggvénye:

$$F(x) = \begin{cases} 0, & \text{ha} \quad x \leq 1; \\ \dfrac{x^3 - 1}{63}, & \text{ha} \quad 1 < x \leq 4; \\ 1, & \text{ha} \quad 4 < x. \end{cases}$$

Az eloszlás-, illetve sűrűségfüggvény grafikonja a 6.7. ábrán látható. ⇐6.10. feladat





6.7. ábra. *A 6.10. feladatban szereplő eloszlásfüggvény (bal oldal) és sűrűségfüggvény (jobb oldal).*

6.11. feladat. Legyen az X valószínűségi változó sűrűségfüggvénye

$$f(x) = \begin{cases} 0, & \text{ha} \quad x \le 3; \\ \dfrac{18}{x^3}, & \text{ha} \quad 3 < x. \end{cases}$$

Számoljuk ki a következő valószínűségeket!

a) $P(X < 5)$, *b)* $P(X \ge 4)$, *c)* $P(3 \le X < 6)$, *d)*$(X > 4 | X < 8)$

Megoldás: A 6.4. tétel szerint az egyes valószínűségeket a sűrűségfüggvénynek a megfelelő intervallumon vett integrálásával számolhatjuk ki.

a)

$$P(X < 5) = \int_{-\infty}^{5} f(x)\,\mathrm{d}x = \int_{-\infty}^{3} f(x)\,\mathrm{d}x + \int_{3}^{5} f(x)\,\mathrm{d}x = \int_{-\infty}^{3} 0\,\mathrm{d}x + \int_{3}^{5} \frac{18}{x^3}\,\mathrm{d}x = 0 + \left[-\frac{9}{x^2}\right]_3^5 = -\frac{9}{25} + \frac{9}{9} = 0{,}64$$

b) Tudjuk, hogy $P(4 \le X) = \int_4^\infty f(x)\,\mathrm{d}x$. Mivel az integrálási tartomány felső határa a végtelen, ezért itt egy improprius integrállal van dolgunk.

$$P(4 \le X) = \int_{4}^{\infty} f(x)\,\mathrm{d}x = \int_{4}^{\infty} \frac{18}{x^3}\,\mathrm{d}x = \left[-\frac{9}{x^2}\right]_4^\infty = 0 + \frac{9}{16} = 0{,}5625.$$

c)

$$P(3 \le X < 6) = \int_{3}^{6} f(x)\,\mathrm{d}x = \int_{3}^{6} \frac{18}{x^3}\,\mathrm{d}x = \left[-\frac{9}{x^2}\right]_3^6 = -\frac{9}{36} + \frac{9}{9} = 0{,}75.$$

d) Írjuk fel először a megadott feltételes valószínűséget a definíció szerint!

$$P(X > 4|X < 8) = \frac{P(4 < X \text{ és } X < 8)}{P(X < 9)} = \frac{P(4 < X < 8)}{P(X < 8)}$$

A számlálót és a nevezőt külön-külön fogjuk kiszámolni. A számláló értéke

$$P(4 < X < 8) = \int\limits_4^8 f(x)\,\mathrm{d}x = \int\limits_4^8 \frac{18}{x^3}\,\mathrm{d}x = \left[-\frac{9}{x^2}\right]_4^8 = -\frac{9}{64} + \frac{9}{16} = 0{,}421875.$$

A nevező értéke

$$P(X < 8) = \int\limits_{-\infty}^8 f(x)\,\mathrm{d}x = \int\limits_{-\infty}^3 0\,\mathrm{d}x + \int\limits_3^8 \frac{18}{x^3}\,\mathrm{d}x = 0 + \left[-\frac{9}{x^2}\right]_3^8 = -\frac{9}{64} + \frac{9}{9} = 0{,}859375.$$

Így azt kapjuk, hogy

$$P(4 < X|X < 8) = \frac{0{,}421875}{0{,}859375} \approx 0{,}491.$$

⇐6.11. feladat

Önellenőrzés

1. A c paraméter mely értékére lesz az

$$f(x) = \begin{cases} \dfrac{c}{(x+1)^4}, & \text{ha } 1 < x; \\ 0, & \text{különben} \end{cases}$$

függvény sűrűségfüggvény? Megold.

2. Az X valószínűségi változó sűrűségfüggvénye:

$$f(x) = \begin{cases} 2\cos x, & \text{ha } 0 < x < \dfrac{\pi}{6}; \\ 0 & \text{különben.} \end{cases}$$

Mennyi a $P\left(X > \dfrac{\pi}{12}\right)$ valószínűség értéke?

- ☐ 0,518
- ☐ 0,482
- ☐ 0,500
- ☐ 0,420

3. Az X valószínűségi változó sűrűségfüggvénye:

$$f(x) = \begin{cases} \dfrac{\sqrt{x}+1}{180}, & \text{ha } 0 < x < 36; \\ 0 & \text{különben.} \end{cases}$$

Mennyi a $P(1 < X < 9)$ valószínűség értéke?

- ☐ 0,141
- ☐ 0,150
- ☐ 0,176
- ☐ 0,195

4. Az előző, 3. feladat esetén mivel egyenlő $P(X < 16)$?

- ☐ 0,182
- ☐ 0,211
- ☐ 0,415
- ☐ 0,326

5. A 3. önellenőrző kérdés esetén mivel egyenlő $P(X = 30)$? Megold.

8. LECKE

A valószínűségi változó várható értéke

7. A várható érték és a szórás

Az előző leckékben láttuk, hogy egy valószínűségi változó az eloszlás-, illetve folytonos esetben a sűrűségfüggvényével is megadható. Sok esetben azonban nincs szükségünk az eloszlás pontos ismeretére, inkább csak az érdekel minket, hogy átlagosan milyen értéket vesz fel a valószínűségi változó, illetve, hogy ehhez az átlagos értékhez képest mennyire ingadoznak a valószínűségi változó értékei. A várható érték és a szórás ezeket a mennyiségeket fogja megadni.

Egy befektetés kapcsán például a befektetőt főleg az érdekli, hogy várhatóan mekkora haszonra tesz szert, míg például egy biztosító céget az foglalkoztat, hogy egy adott típusú ember (pl. 30 éves, városban lakó, hat éve vezető férfi) várhatóan hány balesetet okoz egy évben.

7.1. Diszkrét valószínűségi változó várható értéke

A várható érték fogalmával valószínűleg már mindenki találkozott az élete folyamán. A televízióban például a lottósorsoláskor mindig elmondják, hogy mennyi a következő heti főnyeremény várható értéke, a születéskor várható élettartamról is bizonyára sokan hallottak már (2010-es adat alapján Magyarországon a férfiak átlagosan $70{,}3$, míg a nők 78,4 évig élnek). Nézzük most meg egy korábbi példa alapján, hogyan lehetne a várható érték matematikai definícióját megadni.

7.1. feladat. András és Béla a következő játékot játsszák: mindketten feldobnak egy dobókockát. Ha egyikük sem dobott hatost, akkor Béla fizet Andrásnak 10 Ft-ot. Ha pontosan az egyikük dob hatost, akkor András fizet Bélának 20 Ft-ot, ha pedig mindketten hatost dobnak, akkor András 60 Ft-ot fizet Bélának. Melyikük számára kedvező ez a játék?

Megoldás: Az 6.1. feladatban kiszámoltuk, hogy Béla $25/36$ valószínűséggel veszít 10 Ft-ot, $10/36$ valószínűséggel nyer 20 Ft-ot és $1/36$ valószínűséggel nyer 60 Ft-ot. Tegyük fel, hogy András és Béla sokszor, mondjuk n-szer játssza ezt a játékot. Ekkor a valószínűség értelmezésekor látottak szerint Béla körülbelül $n \cdot 25/36$ partiban veszít 10 Ft-ot, $n \cdot 10/36$ partiban nyer 20 Ft-ot, míg $n \cdot 1/36$ partiban nyer 60 Ft-ot. Így Béla átlagos, azaz egy partira eső nyereménye

$$\frac{-10 \cdot n \cdot \frac{25}{36} + 20 \cdot n \cdot \frac{10}{36} + 60 \cdot n \cdot \frac{1}{36}}{n} = -10 \cdot \frac{25}{36} + 20 \cdot \frac{10}{36} + 60 \cdot \frac{1}{36} = \frac{10}{36}.$$

Azt kaptuk tehát, hogy ha sok partit játszanak, akkor Béla átlagosan $10/36$ Ft-ot nyer partinként. Mivel Béla átlagos nyereménye pozitív, ezért a játék neki kedvező.

⇐7.1. feladat

Figyeljük meg, hogy az előző feladatban hogyan kaptuk meg Béla átlagos nyereményét! Ha X jelöli Béla nyereményét, akkor X lehetséges értékei -10, 20 és 30, és ezeket az értékeket X rendre $25/36$, $10/36$ és $1/36$ valószínűséggel veszi fel. A megoldásban szereplő képletet megnézve láthatjuk, hogy Béla átlagos nyereményét úgy kaptuk meg, hogy X lehetséges értékeit összeszoroztuk a megfelelő valószínűséggel, majd ezeket az értékeket összeadtuk. Diszkrét valószínűségi változó várható értékét pontosan így fogjuk definiálni.

7.1. definíció: Legyen X diszkrét valószínűségi változó, mely az $x_1, x_2, \ldots$ értékeket veheti fel. Legyen továbbá $p_i = P(X = x_i)$. Ekkor X várható értékén az $E(X) = \sum_i x_i \cdot p_i$ összeget értjük, amennyiben a $\sum_i |x_i| \cdot p_i$ összeg véges. Egyéb esetben azt mondjuk, hogy a várható érték nem létezik.

Érdekesség: A különböző szerencsejátékok (rulett, sorsjegyek, lottó, kenó stb.) esetén a szerencsejátékot működtető cég a játék árát, illetve a nyeremények felosztását úgy állapítja meg, hogy a játékos várható nyereménye negatív, ebből adódóan a cég várható nyeresége pozitív legyen. Nagyon ritkán előfordul azonban, hogy nem ez a helyzet. 1992-ben például néhányan észrevették, hogy az egyik ausztrál lottón az összes lehetséges kombinációt megjátszva komoly haszonra lehet szert tenni (összesen kb. hétmillió módon lehetett kitölteni egy szelvényt, egy szelvény ára egy dollár volt, míg a főnyeremény több mint 27 millió dollár). A rengeteg szelvény kitöltésére társakat is toboroztak. Bár határidőre még így sem tudták az összes lehetséges szelvényt feladni (csak kb. ötmilliót), de így is megnyerték a főnyereményt.

7.2. feladat. Egy urnában 10 golyó van, a golyók egytől tízig vannak számozva. Az urnából véletlenszerűen kiválasztunk egy golyót. Az X valószínűségi változó legyen a kihúzott golyóra írt szám. Számoljuk ki X várható értékét!

Megoldás: Mivel mindegyik golyót ugyanakkora valószínűséggel húzzuk ki, ezért mindegyik golyó kihúzásának $1/10 = 0{,}1$ a valószínűsége. X eloszlása tehát a szokásos jelöléssel (felső sorban a lehetséges értékek, az alsó sorban a megfelelő valószínűségek)

$$X : \begin{cases} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 0{,}1 & 0{,}1 & 0{,}1 & 0{,}1 & 0{,}1 & 0{,}1 & 0{,}1 & 0{,}1 & 0{,}1 & 0{,}1 \end{cases} .$$

A várható érték:

$$E(X) = 1 \cdot 0{,}1 + 2 \cdot 0{,}1 + \ldots + 10 \cdot 0{,}1 = 5{,}5.$$

⇐7.2. feladat

7.3. feladat. Egy kaparós sorsjegyből 100 ezer darabot hoznak forgalomba. 3 olyan van köztük, mellyel egymillió Ft-ot és 100 olyan, mellyel ötvenezer Ft-ot lehet nyerni. A sorsjegy ára 100 Ft. Egy sorsjegy vásárlása esetén mennyi a nettó nyeremény várható értéke?

Megoldás: Jelölje X a nettó nyereményt (azaz a nyereményből kivonva a sorsjegy árát). Ekkor X a 999900, 49900 és -100 értékeket veheti fel. X eloszlása:

$$X : \begin{cases} -100 & 49900 & 999900 \\ \frac{99897}{100000} & \frac{100}{100000} & \frac{3}{100000} \end{cases} .$$

Ezek alapján a nettó nyeremény várható értéke:

$$E(X) = -100 \cdot \frac{99897}{100000} + 49900 \cdot \frac{100}{100000} + 999900 \cdot \frac{3}{100000} = -\frac{1999700}{100000} = -19{,}997.$$

⇐7.3. feladat

A következő feladat azt mutatja, hogy nem mindegyik valószínűségi változónak létezik várható értéke.

7.4. feladat. Egy szabályos pénzérmét dobálgatunk. Ha k.-ra dobunk először fejet, akkor az X valószínűségi változó értéke legyen 2^k. Mennyi X várható értéke?

Megoldás: Határozzuk meg X eloszlását. Az, hogy a k. dobásra dobunk először fejet, pontosan azt jelenti, hogy az első $(k-1)$ dobásra írást dobtunk, a k.-ra pedig fejet. Mivel az egyes dobások eredménye egymástól független, ezért

$$P(X=2^k) = P(k\text{.-ra dobunk először fejet}) = \underbrace{\frac{1}{2}\cdot\frac{1}{2}\cdot\ldots\cdot\frac{1}{2}}_{k-1 \text{ darab}}\cdot\frac{1}{2} = \left(\frac{1}{2}\right)^k.$$

X eloszlása tehát a következőképpen írható fel:

$$X:\left\{\begin{array}{cccccc} 2^1 & 2^2 & \ldots & 2^k & \ldots \\ \left(\frac{1}{2}\right)^1 & \left(\frac{1}{2}\right)^2 & \ldots & \left(\frac{1}{2}\right)^k & \ldots \end{array}\right. .$$

Ebből X várható értéke:

$$E(X) = 2^1\cdot\left(\frac{1}{2}\right)^1 + 2^2\cdot\left(\frac{1}{2}\right)^2 + \ldots + 2^k\cdot\left(\frac{1}{2}\right)^k + \ldots = 1+1+1+\ldots = \infty,$$

vagyis a várható érték nem létezik.

⇐7.4. feladat

7.2. Folytonos valószínűségi változó várható értéke

Legyen X folytonos eloszlású valószínűségi változó, a sűrűségfüggvénye legyen $f(x)$. Legyen (a, b) az az intervallum, ahol $f(x)$ nem nulla. Osszuk fel ezt az intervallumot m részre, az i-edik részintervallum végpontjai legyenek t_{i-1} és t_i. Ekkor annak valószínűsége, hogy az X valószínűségi változó értéke az i-edik részintervallumba esik

$$P(t_{i-1} < X < t_i) = \int\limits_{t_{i-1}}^{t_i} f(x)\,\mathrm{d}x\,.$$

Végezzünk el n darab független kísérletet, az eredmények legyenek $x_1, x_2, \ldots, x_n$, és vizsgáljuk, hogy az egyes részintervallumokba hányszor esik az eredmény. Jelölje k_i az i-edik részintervallumba eső elemek számát. Az átlagot közelítsük úgy, hogy minden egyes részintervallumból választunk egy-egy számot, amivel helyettesítünk minden olyan értéket, ami az adott intervallumba esik (például, ha a $(3{,}4)$ intervallumot tekintjük, akkor vehetjük a $3{,}5$-et). Legyen az i-edik intervallumból választott érték τ_i. Ekkor az átlag közelítése:

$$\frac{1}{n}\sum_{i=1}^{n} x_i \approx \frac{1}{n}\sum_{i=1}^{n} k_i \cdot \tau_i = \sum_{i=1}^{n} \tau_i \cdot \frac{k_i}{n}\,.$$

Sok kísérletsorozat esetén a $\dfrac{k_i}{n}$ relatív gyakoriság az i-edik részintervallumba esés valószínűsége körül ingadozik, ezért

$$\frac{k_i}{n} \approx \int\limits_{t_{i-1}}^{t_i} f(x)\,\mathrm{d}x\,.$$

Ezt felhasználva az átlag közelítése:

$$\frac{1}{n}\sum_{i=1}^{n} x_i \approx \sum_{i=1}^{n} \tau_i \cdot \frac{k_i}{n} \approx \sum_{i=1}^{n} \tau_i \cdot \int\limits_{t_{i-1}}^{t_i} f(x)\,\mathrm{d}x \approx \sum_{i=1}^{n} \int\limits_{t_{i-1}}^{t_i} x \cdot f(x)\,\mathrm{d}x = \int\limits_{a}^{b} x \cdot f(x)\,\mathrm{d}x\,.$$

A fenti eljárás alapján definiálhatjuk a folytonos eloszlású valószínűségi változó várható értékét:

7.2. definíció: Legyen X folytonos eloszlású valószínűségi változó, a sűrűségfüggvénye legyen $f(x)$. Ekkor X várható értékén az $E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x$ integrált értjük, amennyiben az $\int_{-\infty}^{\infty} |x| \cdot f(x)\,\mathrm{d}x$ integrál véges. Egyéb esetben azt mondjuk, hogy a várható érték nem létezik.

7.5. feladat. Legyen az X valószínűségi változó sűrűségfüggvénye

$$f(x) = \begin{cases} \dfrac{c}{x^2}, & \text{ha } 2 < x < 4; \\ 0, & \text{különben.} \end{cases}$$

Mennyi a c értéke? Mennyi az X valószínűségi változó várható értéke?

Megoldás: Az f függvény akkor lesz sűrűségfüggvény, ha egyrészt csak nemnegatív értékeket vesz fel, másrészt a mínusz végtelentől végtelenig vett integrálja eggyel egyenlő. Az első feltételből azt kapjuk, hogy $0 \leq c$, míg a második feltétel szerint

$$1 = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \int_{2}^{4} \frac{c}{x^2}\,\mathrm{d}x = \left[-\frac{c}{x}\right]_2^4 = c \cdot \left(\frac{1}{2} - \frac{1}{4}\right) = \frac{c}{2}$$

amiből $c = 2$ adódik.

A várható érték definíciója szerint

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x = \int_{2}^{4} x \cdot \frac{2}{x^2}\,\mathrm{d}x = \int_{2}^{4} \frac{2}{x}\,\mathrm{d}x = \left[2 \cdot \ln x\right]_2^4 = 2 \cdot \ln 4 - 2 \cdot \ln 2 \approx 1{,}386.$$

⇐7.5. feladat

7.6. feladat. Az X valószínűségi változó eloszlásfüggvénye az alábbi függvény:

$$F(x) = \begin{cases} 0, & \text{ha} \quad x \leq 1; \\ \dfrac{x^2 - 1}{8}, & \text{ha} \quad 1 < x \leq 3; \\ 1, & \text{ha} \quad 3 < x. \end{cases}$$

Határozzuk meg X várható értékét!

Megoldás: A várható érték kiszámításához először a sűrűségfüggvényt kell meghatároznunk. Mivel a sűrűségfüggvény az eloszlásfüggvény deriváltja, ahol pedig az eloszlásfüggvény nem deriválható, ott a definíció szerint 0, ezért

$$f(x) = F'(x) = \begin{cases} \dfrac{2x}{8}, & \text{ha } 1 < x < 3; \\ 0, & \text{különben.} \end{cases}$$

Most már kiszámolhatjuk a várható értéket:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, \mathrm{d}x = \int_1^3 x \cdot \frac{2x}{8}\, \mathrm{d}x = \int_1^3 \frac{x^2}{4}\, \mathrm{d}x = \left[\frac{x^3}{12}\right]_1^3 = \frac{26}{12} \approx 2{,}167.$$

⇐7.6. feladat

7.1. tétel: Legyen X egy tetszőleges valószínűségi változó és legyen $Y = a \cdot X^2 + b \cdot X + c$, ahol a, b és c tetszőleges valós számok. Ekkor az Y valószínűségi változó várható értéke

$$E(Y) = a \cdot E(X^2) + b \cdot E(X) + c$$

amennyiben X és X^2 várható értéke létezik.

7.7. feladat. Feldobunk egy szabályos dobókockát. Az Y valószínűségi változó legyen a dobott szám háromszorosánál eggyel kisebb szám. Határozzuk meg Y várható értékét.

Megoldás: **1. módszer** Határozzuk meg Y eloszlását. Világos, hogy Y a $3\cdot 1-1, 3\cdot 2-1, \dots\ 3\cdot 6-1$ értékeket veheti fel, ráadásul mindegyiket egyforma valószínűséggel. Így Y eloszlása

$$Y:\left\{\begin{array}{cccccc} 2 & 5 & 8 & 11 & 14 & 17 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{array}\right. ,$$

amiből a várható értékre

$$E(Y)=2\cdot\frac{1}{6}+5\cdot\frac{1}{6}+8\cdot\frac{1}{6}+11\cdot\frac{1}{6}+14\cdot\frac{1}{6}+17\cdot\frac{1}{6}=\frac{57}{6}=9{,}5$$

adódik.

2. módszer A 7.1. tétel segítségével Y eloszlásának kiszámítása nélkül is megkaphatjuk a várható értéket. Legyen X a dobás során kapott érték, ekkor $Y=3X-1$. A 7.1. tétel alapján Y várható értéke kiszámítható X várható értékének segítségével. Mivel X az 1,2,3,4,5,6 értékeket veheti fel, továbbá mindegyik értéket $1/6$ valószínűséggel veszi fel, ezért

$$E(X)=1\cdot\frac{1}{6}+2\cdot\frac{1}{6}+3\cdot\frac{1}{6}+4\cdot\frac{1}{6}+5\cdot\frac{1}{6}+5\cdot\frac{1}{6}=\frac{21}{6}=3{,}5.$$

Ebből a tétel alapján

$$E(Y)=E(3X-1)=3\cdot E(X)-1=3\cdot\frac{21}{6}-1=\frac{57}{6}=9{,}5.$$

⇐7.7. feladat

Önellenőrzés

1. Az X valószínűségi változó eloszlása:

$$X : \begin{cases} -4 & -1 & 3 & 6 \\ 0{,}1 & 0{,}3 & 0{,}4 & 0{,}2 \end{cases}.$$

Mennyi X várható értéke? Megold.

2. Feldobunk három szabályos pénzérmét. Mennyi a fej dobások számának várható értéke? Megold.

3. Feldobunk két szabályos dobókockát. Mennyi a dobott számok összegének a várható értéke? Megold.

4. Az X valószínűségi változó sűrűségfüggvénye:

$$f(x) = \begin{cases} \dfrac{x^2}{18}, & \text{ha } -3 < x < 3; \\ 0, & \text{különben.} \end{cases}$$

Mennyi X várható értéke? Megold.

5. Az X valószínűségi változó eloszlásfüggvénye:

$$F(x) = \begin{cases} 0, & \text{ha} \quad x \le 2; \\ \dfrac{x-2}{8}, & \text{ha} \quad 2 < x \le 10; \\ 1, & \text{ha} \quad 10 < x. \end{cases}$$

Mennyi X várható értéke? Megold.

9. LECKE

A szórás

7.3. A szórás

Az előző leckében láttuk, hogy a várható érték gyakorlatilag a valószínűségi változó átlagos értékét adja meg. Sokszor azonban az is lényeges, hogy a valószínűségi változó értékei mennyire ingadoznak a várható érték körül.

Tegyük fel például, hogy a megspórolt pénzünket be akarjuk fektetni. Két lehetőség közül választhatunk: az elsőnél úgy gondoljuk, hogy 1 éven belül $0{,}7$ valószínűséggel nyerünk egymillió Ft-ot és $0{,}3$ valószínűséggel vesztünk hatszázezer Ft-ot, a másodiknál pedig a becslésünk szerint $0{,}5$ valószínűséggel nyerünk hatszázezer Ft-ot, és $0{,}5$ valószínűséggel nyerünk négyszázezer Ft-ot egy éven belül. Melyik befektetést válasszuk?

Jelölje X az első befektetés, Y pedig a második befektetés által elért nyereséget. Ekkor az X valószínűségi változó várható értéke $E(X) = 0{,}7 \cdot 1000000 - 0{,}3 \cdot 600000 = 500000$, míg az Y valószínűségi változó várható értéke $E(Y) = 0{,}5 \cdot 600000 + 0{,}5 \cdot 400000 = 500000$ Ft. Azt kaptuk tehát, hogy mindkét befektetéssel várhatóan ugyanannyit nyerünk, mégis mindenki számára világos, hogy a két valószínűségi változó jelentősen eltér egymástól. Az első valószínűségi változó lehetséges értékei a várható értékhez képest nagy, míg a másodiké csak kis ingadozást mutatnak. Arra, hogy melyiket érdemes választani, most nem lehet egyértelmű választ adni (ez függ attól, hogy mennyi pénzt fektetünk be, és attól is, hogy milyen a kockázattűrő képességünk). Ha a nagyobb nyereség reményében nem félünk kockáztatni azt, hogy sokat is veszíthetünk, akkor választhatjuk az első lehetőséget, de ha a célunk a kockázat nélküli biztos nyereség, akkor inkább a második lehetőséget célszerű választani.

A fentiek alapján célszerű lenne olyan mérőszámot keresnünk, amellyel jellemezhetjük a várható érték körül ingadozást, vagyis a valószínűségi változónak (X) és a várható értéknek ($E(X)$) az eltérését szeretnénk egyetlen számmal leírni. Gondolhatnánk arra, hogy a várható értéktől való eltérés átlaga megfelelő lenne erre a célra; ez éppen az $X - E(X)$ valószínűségi változó várható értékét jelentené, de

$$E[X - E(X)] = E(X) - E(X) = 0,$$

így ez egyáltalán nem jellemzi az ingadozást. Tekintsük ezért inkább az $X - E(X)$ különbség négyzetét. Ennek a valószínűségi változónak a várható értéke már megfelelő lesz az ingadozás jellemzésére. Pontosabban:

7.3. definíció: Az X valószínűségi változó szórásán az $(X - E(X))^2$ valószínűségi változó várható értékének négyzetgyökét értjük, amit $D(X)$-szel (vagy σ-val) jelölünk. Azaz X szórása

$$D(X) = \sqrt{E\left[(X - E(X))^2\right]}.$$

amennyiben X és $(X - E(X))^2$ várható értéke létezik.

7.4. definíció: A gyökvonás nélküli $D^2(X) = E\left[(X - E(X))^2\right]$ értéket az X valószínűségi változó szórásnégyzetének nevezzük. Általában először a szórásnégyzetet tudjuk meghatározni, és abból gyökvonással a szórást.

7.2. tétel: Ha az X valószínűségi változónak és annak a négyzetének is létezik a várható értéke, akkor létezik X szórása is, és

$$D(X) = \sqrt{E(X^2) - E^2(X)}.$$

Bizonyítás: A kifejezés szórásnégyzetre vonatkozó alakját fogjuk igazolni, abból pedig gyökvonással következik az állítás.

$$\begin{aligned} D^2(X) = E\left[(X - E(X))^2\right] &= E\left[X^2 - 2 \cdot X \cdot E(X) + E^2(X)\right] = \\ &= E(X^2) - 2 \cdot E(X) \cdot E(X) + E^2(X) = E(X^2) - E^2(X). \end{aligned}$$

□

Az előző tétel alapján tehát egy X valószínűségi változó szórásának kiszámításához elegendő X, illetve X^2 várható értékét kiszámítani.

7.8. feladat. Adott az X valószínűségi változó eloszlása:

$$X : \begin{cases} -2 & -1 & 1 & 2 \\ 0{,}1 & 0{,}3 & 0{,}4 & 0{,}2 \end{cases} .$$

Számoljuk ki X szórását!

Megoldás: Az előző tétel alapján a szórás kiszámításához először meghatározzuk X várható értékét, utána pedig X^2 várható értékét. A definíció alapján X várható értéke

$$E(X) = -2 \cdot 0{,}1 - 1 \cdot 0{,}3 + 1 \cdot 0{,}4 + 2 \cdot 0{,}2 = 0{,}3.$$

Az X^2 valószínűségi változó az 1 és 4 értékeket veheti fel. Mivel X^2 akkor veszi fel az 1 értéket, ha X a -1 vagy 1 értékeket veszi fel, ezért X^2 az 1 értéket $0{,}3 + 0{,}4 = 0{,}7$ valószínűséggel veszi fel. Hasonlóan látható, hogy X^2 a 4 értéket $0{,}3$ valószínűséggel veszi fel, így X^2 eloszlása

$$X^2 : \begin{cases} 1 & 4 \\ 0{,}7 & 0{,}3 \end{cases} .$$

Ebből a definíció szerint X^2 várható értéke

$$E(X^2) = 1 \cdot 0{,}7 + 4 \cdot 0{,}3 = 1{,}5.$$

A 7.2. tétel szerint X szórására

$$D(X) = \sqrt{E(X^2) - E^2(X)} = \sqrt{1{,}5 - 0{,}3^2} \approx 1{,}187$$

adódik.

⇐7.8. feladat

7.9. feladat. Ötször feldobunk egy szabályos pénzérmét. Számoljuk ki a fej dobások számának várható értékét és szórását!

Megoldás: Jelölje X a fej dobások számát. Határozzuk meg először X eloszlását! Mivel mindegyik dobásnak két kimenetele van (F,I), ezért az összes eset száma $2^5 = 32$. Az öt dobás során pontosan k fejet $\binom{5}{k}$ esetben dobhatunk, így $P(X = k) = \binom{5}{2} \cdot 1/32$. A fej dobások száma nullától ötig terjedhet, amiből a $P(X = k)$ képletben k helyére a lehetséges értékeket behelyettesítve kapjuk X eloszlását:

$$X : \begin{cases} 0 & 1 & 2 & 3 & 4 & 5 \\ \dfrac{1}{32} & \dfrac{5}{32} & \dfrac{10}{32} & \dfrac{10}{32} & \dfrac{5}{32} & \dfrac{1}{32} \end{cases} .$$

A várható érték definíciója alapján:

$$E(X) = 0 \cdot \frac{1}{32} + 1 \cdot \frac{5}{32} + 2 \cdot \frac{10}{32} + 3 \cdot \frac{10}{32} + 4 \cdot \frac{5}{32} + 5 \cdot \frac{1}{32} = \frac{80}{32} = 2{,}5.$$

A szórás kiszámolásához szükségünk van X^2 eloszlására:

$$X^2 : \begin{cases} 0 & 1 & 4 & 9 & 16 & 25 \\ \dfrac{1}{32} & \dfrac{5}{32} & \dfrac{10}{32} & \dfrac{10}{32} & \dfrac{5}{32} & \dfrac{1}{32} \end{cases} .$$

Ebből

$$E(X^2) = 0 \cdot \frac{1}{32} + 1 \cdot \frac{5}{32} + 4 \cdot \frac{10}{32} + 9 \cdot \frac{10}{32} + 16 \cdot \frac{5}{32} + 25 \cdot \frac{1}{32} = \frac{240}{32} = 7{,}5.$$

A 7.2. tétel szerint X szórása

$$D(X) = \sqrt{E(X^2) - E^2(X)} = \sqrt{7{,}5 - 2{,}5^2} = \sqrt{1{,}25} \approx 1{,}118.$$

⇐7.9. feladat

A későbbiekben majd igazoljuk, hogy ha X folytonos eloszlású valószínűségi változó, melynek $f(x)$ a sűrűségfüggvénye, akkor X^2 várható értéke

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)\,\mathrm{d}x$$

amiből X szórása (feltéve, hogy $E(X)$ és $E(X^2)$ is létezik)

$$D(X) = \sqrt{E(X^2) - (E(X))^2} = \sqrt{\int_{-\infty}^{\infty} x^2 \cdot f(x)\,\mathrm{d}x - \left(\int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x\right)^2}.$$

7.10. feladat. A folytonos eloszlású X valószínűségi változó sűrűségfüggvénye:

$$f(x) = \begin{cases} \dfrac{2-x}{2}, & \text{ha } 0 < x < 2; \\ 0, & \text{különben.} \end{cases}$$

Számoljuk ki X szórását!

Megoldás: A szórás kiszámításához az előző megjegyzés alapján először X várható értékét, majd X^2 várható értékét számoljuk ki.

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x = \int_0^2 x \cdot \frac{2-x}{2}\,\mathrm{d}x = \int_0^2 \frac{2x - x^2}{2}\,\mathrm{d}x = \left[\frac{1}{2} \cdot \left(x^2 - \frac{x^3}{3}\right)\right]_0^2 = \frac{2}{3} - 0 = \frac{2}{3}.$$

Most kiszámoljuk X^2 várható értékét.

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)\,\mathrm{d}x = \int_0^2 x^2 \cdot \frac{2-x}{2}\,\mathrm{d}x = \int_0^2 \frac{2x^2 - x^3}{2}\,\mathrm{d}x = \left[\frac{1}{2} \cdot \left(\frac{2x^3}{3} - \frac{x^4}{4}\right)\right]_0^2 = \frac{8}{12} - 0 = \frac{2}{3}.$$

Most már meg tudjuk határozni a szórást:

$$D(X) = \sqrt{E(X^2) - (E(X))^2} = \sqrt{\frac{2}{3} - \left(\frac{2}{3}\right)^2} = \frac{\sqrt{2}}{3}.$$

⇐7.10. feladat

7.11. feladat. A folytonos eloszlású X valószínűségi változó eloszlásfüggvénye:

$$F(x) = \begin{cases} 0, & \text{ha} \quad x \leq 3; \\ 1 - \dfrac{81}{x^4}, & \text{ha} \quad 3 < x. \end{cases}$$

Számoljuk ki X szórását!

Megoldás: Most is ugyanúgy járunk el, mint az előző feladatban, csak először ki kell számolnunk X sűrűségfüggvényét. Tudjuk, hogy a sűrűségfüggvény az eloszlásfüggvény deriváltjával egyezik meg. Mivel konstans deriváltja 0, ezért $x < 3$ esetén $f(x) = F'(x) = 0$. Ha $3 < x$, akkor

$$f(x) = F'(x) = \left(1 - \frac{81}{x^4}\right)' = \left(1 - 81 \cdot x^{-4}\right)' = (-4) \cdot (-81) \cdot x^{-5} = \frac{324}{x^5}.$$

Azt kaptuk tehát, hogy X sűrűségfüggvénye

$$f(x) = \begin{cases} 0, & \text{ha} \quad x \leq 3; \\ \dfrac{324}{x^5}, & \text{ha} \quad 3 < x. \end{cases}$$

Számoljuk most ki X várható értékét!

$$E(X)=\int_{-\infty}^{\infty} x\cdot f(x)\,\mathrm{d}x=\int_{3}^{\infty} x\cdot\frac{324}{x^5}\,\mathrm{d}x=\int_{3}^{\infty}\frac{324}{x^4}\,\mathrm{d}x=\int_{3}^{\infty}324\cdot x^{-4}\,\mathrm{d}x=\left[324\cdot\frac{x^{-3}}{-3}\right]_3^{\infty}=$$
$$=\left[\frac{-108}{x^3}\right]_3^{\infty}=0-\frac{-108}{27}=4.$$

Ezután kiszámoljuk X^2 várható értékét:

$$E(X^2)=\int_{-\infty}^{\infty} x^2\cdot f(x)\,\mathrm{d}x=\int_{3}^{\infty} x^2\cdot\frac{324}{x^5}\,\mathrm{d}x=\int_{3}^{\infty}\frac{324}{x^3}\,\mathrm{d}x=\int_{3}^{\infty}324\cdot x^{-3}\,\mathrm{d}x=\left[324\cdot\frac{x^{-2}}{-2}\right]_3^{\infty}=$$
$$=\left[\frac{-162}{x^2}\right]_3^{\infty}=0-\frac{-162}{9}=18.$$

Most már meg tudjuk határozni X szórását:

$$D(X)=\sqrt{E(X^2)-(E(X))^2}=\sqrt{18-4^2}=\sqrt{2}.$$

⇐7.11. feladat

7.3. tétel: Legyen X egy tetszőleges valószínűségi változó és legyen $Y=a\cdot X+b$. Ekkor Y szórása

$$D(Y)=D(a\cdot X+b)=|a|\cdot D(X)$$

amennyiben X szórása létezik.

A tétel alapján X szórásának ismeretében az $Y=a\cdot X+b$ valószínűségi változó szórása Y eloszlásának ismerete nélkül is kiszámolható.

7.12. feladat. Ötször feldobunk egy szabályos pénzérmét. Az Y valószínűségi változó értéke legyen a dobott fejek számának kétszeresénél néggyel kisebb szám. Számoljuk ki Y szórását!

Megoldás: Jelölje X a fej dobások számát. Ekkor $Y = 2X - 4$. Korábban a 7.9. feladatban már kiszámoltuk, hogy $D(X) = \sqrt{1{,}25}$. A 7.3. tétel szerint

$$D(Y) = D(2X - 4) = |2| \cdot \sqrt{1{,}25} \approx 2{,}236.$$

⇐7.12. feladat

7.13. feladat. A folytonos X valószínűségi változó sűrűségfüggvénye

$$f(x) = \begin{cases} \dfrac{3x^2}{7}, & \text{ha } 1 < x < 2; \\ 0 & \text{különben.} \end{cases}$$

Legyen $Y = -2 \cdot X + 1$. Számoljuk ki Y szórását!

Megoldás: Először kiszámoljuk X szórását, amiből a 7.3. tétel alapján megkapjuk Y szórását (azaz Y eloszlását nem kell felírnunk).

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x = \int_1^2 x \cdot \frac{3x^2}{7}\,\mathrm{d}x = \int_1^2 \frac{3}{7} \cdot x^3\,\mathrm{d}x = \left[\frac{3}{7} \cdot \frac{x^4}{4}\right]_1^2 = \frac{48}{28} - \frac{3}{28} = \frac{45}{28}.$$

Most számoljuk ki X^2 várható értékét.

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)\,\mathrm{d}x = \int_1^2 x^2 \cdot \frac{3x^2}{7}\,\mathrm{d}x = \int_1^2 \frac{3}{7} \cdot x^4\,\mathrm{d}x = \left[\frac{3}{7} \cdot \frac{x^5}{5}\right]_1^2 = \frac{96}{35} - \frac{3}{35} = \frac{93}{35}.$$

Ebből X szórása:

$$D(X) = \sqrt{E(X^2) - (E(X))^2} = \sqrt{\frac{93}{35} - \left(\frac{45}{28}\right)^2} \approx 0{,}272.$$

A 7.3. tétel szerint Y szórása:

$$D(Y) = D(-2X + 1) = |-2| \cdot 0{,}272 = 0{,}545.$$

⇐7.13. feladat

Önellenőrzés

1. Fogalmazza meg saját szavaival, hogy mit értünk egy valószínűségi változó szórásán!
2. Az X valószínűségi változó eloszlása:

$$X : \begin{cases} -3 & 1 & 3 & 4 \\ 0{,}1 & 0{,}3 & 0{,}4 & 0{,}2 \end{cases}.$$

Mennyi X szórásnégyzete? Megold.

3. Feldobunk két dobókockát. Mennyi a hatos dobások számának a szórása?

☐ 0,913 ☐ 0,527 ☐ 0,758 ☐ 0,651

4. Az X valószínűségi változó sűrűségfüggvénye:

$$f(x) = \begin{cases} \dfrac{2x}{81}, & \text{ha } 0 < x < 9; \\ 0 & \text{különben.} \end{cases}$$

Mennyi X szórásnégyzete? Megold.

5. Az X valószínűségi változó eloszlásfüggvénye:

$$F(x) = \begin{cases} 0, & \text{ha} \quad x \leq -1; \\ \dfrac{x+1}{13}, & \text{ha} \quad -1 < x \leq 12; \\ 1, & \text{ha} \quad 12 < x. \end{cases}$$

Mennyi X szórása?

☐ 13 ☐ 5,387 ☐ 3,753 ☐ 10

6. Az X valószínűségi változó eloszlása:

$$X : \begin{cases} -2 & -1 & 1 & 2 \\ 0{,}2 & 0{,}3 & 0{,}3 & 0{,}2 \end{cases} .$$

Legyen $Y = -3 \cdot X + 5$. Mennyi Y szórása?

☐ 6,600 ☐ $-4{,}450$ ☐ 9,450 ☐ 4,450

10. LECKE

Modulzáró feladatok 2.

8. Modulzáró feladatok

Start. Az alábbi kérdések megválaszolására 1 órája van:

1. Feldobunk egy dobókockát. Az X valószínűségi változó értéke legyen 1, ha párosat dobunk, legyen 2, ha egyest dobunk, és legyen 3, ha hármast vagy ötöst dobunk. Mennyi X várható értéke?

   ☐ 1,668 ☐ 1,752 ☐ 1,833 ☐ 1,985

2. Az előző feladatbeli X esetén mennyi lesz az $Y = -3X + 1$ valószínűségi változó szórása?

   ☐ −2,693 ☐ −1,693 ☐ 2,693 ☐ 3,693

3. Az X valószínűségi változó eloszlásfüggvénye:

$$F(x) = \begin{cases} 0, & \text{ha} \quad x \le 0; \\ \dfrac{x^2 + x^3}{36}, & \text{ha} \quad 0 < x \le 3; \\ 1, & \text{ha} \quad 3 < x. \end{cases}$$

   Mennyi a $P(2 < X)$ valószínűség értéke?

   ☐ 0,333 ☐ 0,667 ☐ 0,785 ☐ 0,236

4. Az X valószínűségi változó eloszlásfüggvénye:

$$F(x) = \begin{cases} 0, & \text{ha} \quad x \le 1; \\ 1 - \dfrac{1}{x^4}, & \text{ha} \quad 1 < x. \end{cases}$$

   Mennyi a $P(4 < X | 2 < X)$ valószínűség értéke? ______

5. Mennyi a várható értéke az előző feladatbeli X valószínűségi változónak?

☐ $-1{,}333$ ☐ $0{,}8$ ☐ 1 ☐ $1{,}333$

6. Az X valószínűségi változó sűrűségfüggvénye:

$$f(x) = \begin{cases} \dfrac{1}{4 \cdot \sqrt{x}}, & \text{ha } 4 < x < 16; \\ 0 & \text{különben.} \end{cases}$$

Mennyi a $P(2 < X < 9)$ valószínűség értéke? ☐

7. Mennyi az előző feladatban megadott X valószínűségi változó szórása?

☐ $3{,}477$ ☐ $4{,}507$ ☐ $2{,}831$ ☐ $5{,}678$

8. Az X valószínűségi változó sűrűségfüggvénye:

$$f(x) = \begin{cases} \dfrac{x}{18}, & \text{ha } 0 < x < 6; \\ 0 & \text{különben.} \end{cases}$$

Mennyi X szórásnégyzete? ☐

Stop.

Pontok:

III. MODUL

Nevezetes eloszlások

11. LECKE

Nevezetes diszkrét eloszlások I.

9. Nevezetes diszkrét eloszlások

9.1. Az indikátor változó eloszlása

Tegyük fel, hogy egyetlen kísérletet hajtunk végre, és csak arra vagyunk kíváncsiak, hogy egy adott valószínűségű esemény bekövetkezik vagy sem. Így kapjuk meg az indikátor valószínűségi változót.

9.1. definíció: Végezzünk el egyetlen kísérletet a p valószínűségű A esemény megfigyelésére. Az X valószínűségi vátozó értéke legyen 1, ha A bekövetkezik, és legyen 0, ha A nem következik be. Ekkor X az A esemény indikátor változója, eloszlása pedig az alábbi:

$$X : \begin{cases} 0 & 1 \\ 1-p & p \end{cases} .$$

A valószínűségi változó várható értéke:

$$E(X) = 0 \cdot (1-p) + 1 \cdot p = p\,.$$

Mivel $0^2 = 0$ és $1^2 = 1$, ezért a valószínűségi változó négyzetének eloszlása és várható értéke ugyanaz, mint X eloszlása és várható értéke:

$$X^2 : \begin{cases} 0 & 1 \\ 1-p & p \end{cases} \qquad \Longrightarrow \qquad E(X^2) = p\,.$$

A valószínűségi változó szórása:

$$D(X) = \sqrt{E(X^2) - E^2(X)} = \sqrt{p - p^2} = \sqrt{p(1-p)}\,.$$

Az indikátor eloszlás jelentősége abban áll, hogy ha ugyanazt a kísérletet többször elvégezzük, és azt figyeljük, hogy egy adott esemény hányszor következik be, akkor minden egyes kísérlethez hozzárendelhetünk egy-egy indikátor valószínűségi változót, így az esemény bekövetkezéseinek a számát ezen valószínűségi változók összege adja meg. Így kaphatjuk meg például a binomiális eloszlást is.

9.2. Binomiális eloszlás

Binomiális eloszlással adhatjuk meg, hogy adott számú független kísérlet elvégzése esetén a megfigyelt p valószínűségű esemény összesen hányszor következik be. Az előbbiek szerint a binomiális eloszlás megkapható úgy is, mint adott számú (n), független, azonos paraméterű (p) indikátor eloszlás összege. Binomiális eloszlással már a korábbiakban is találkoztunk: mivel a visszatevéses mintavételnél a körülmények minden egyes húzásnál teljesen azonosak, ezért a kihúzott selejtek száma binomiális eloszlással írható le.

9.2. definíció: Végezzünk el n darab független kísérletet a p valószínűségű A esemény megfigyelésére. Az X valószínűségi változó értéke legyen a sikeres kísérletek száma $(0,1,2,\dots,n)$. Ekkor X binomiális eloszlású valószínűségi változó az n és p paraméterekkel, eloszlása:

$$P(X=k)=\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}\qquad k=0,1,2,\dots,n\,.$$

Felhasználva, hogy a binomiális eloszlású valószínűségi változó tekinthető n darab független, karakterisztikus eloszlású valószínűségi változó összegének is, belátható, hogy a változó várható értéke és szórása:

$$E(X)=n\cdot p\qquad D(X)=\sqrt{np(1-p)}\,.$$

9.1. feladat. Egymás után feldobunk egy szabályos dobókockát hatszor.

a) Mennyi a valószínűsége, hogy pontosan kétszer dobunk hatost?

b) Mennyi a valószínűsége, hogy legfeljebb négyszer dobunk hatost?

c) Mennyi a hatos dobások számának várható értéke és szórása?

Megoldás: Jelölje X a hatos dobások számát. Mivel ugyanazt a kísérletet végezzük el hatszor, ezért X binomiális eloszlású valószínűségi változó, $n=6$ és $p=1/6$ paraméterekkel.

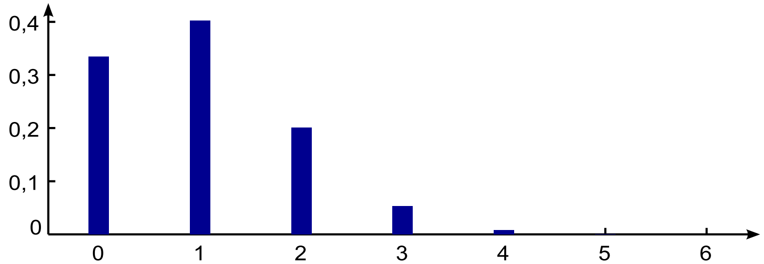
a)

$$P(X=2)=\binom{6}{2}\cdot\left(\frac{1}{6}\right)^2\cdot\left(\frac{5}{6}\right)^4\approx 0{,}201.$$

b) Az, hogy legfeljebb négyszer dobunk hatost, azt jelenti, hogy vagy nullaszor, vagy egyszer, vagy kétszer, vagy háromszor vagy négyszer dobunk hatost. Kevesebbet kell számolnunk, ha a komplementer esemény (ötször vagy hatszor dobunk hatost) valószínűségével számolunk.

$$P(X\le 4)=1-[P(X=5)+P(X=6)]=1-\left[\binom{6}{5}\cdot\left(\frac{1}{6}\right)^5\cdot\left(\frac{5}{6}\right)^1+\binom{6}{6}\cdot\left(\frac{1}{6}\right)^6\cdot\left(\frac{5}{6}\right)^0\right]$$
$$\approx 1-0{,}0007=0{,}9993.$$

c) A várható érték, $E(X)=n\cdot p=6\cdot 1/6=1$. A szórás, $D(X)=\sqrt{n\cdot p\cdot(1-p)}=\sqrt{6\cdot\frac{1}{6}\cdot\frac{5}{6}}\approx 0{,}913$.



9.1. ábra. *Binomiális eloszlás $n=6$ és $p=1/6$ paraméterekkel.*

⇐9.1. feladat

9.2. feladat. Egy cég 10 egyforma nyomtatót vásárol. Egy nyomtató $0{,}2$ valószínűséggel romlik el a garanciális idő alatt. Mennyi a

a) valószínűsége, hogy a garanciális idő alatt egyetlen nyomtató sem romlik el;

b) valószínűsége, hogy a garanciális idő alatt legalább három nyomtató elromlik;

c) valószínűsége, hogy a garanciális idő alatt kevesebb mint kettő nyomtató romlik el;

d) valószínűsége, hogy a garanciális idő alatt kettőnél több, de legfeljebb öt nyomtató romlik el?

e) garanciális idő alatt elromló nyomtatók számának várható értéke és szórása?

Megoldás: Legyen X a garanciális idő alatt elromló nyomtatók száma. Mivel a gépek egymástól függetlenül működnek, ezért X binomiális eloszlású valószínűségi változó $n = 10$ és $p = 1/5$ paraméterekkel.

a)

$$P(X = 0) = (1 - p)^{10} = \left(\frac{4}{5}\right)^{10} \approx 0{,}107.$$

b) Hamarabb célba érünk, ha a komplementer esemény valószínűségével számolunk:

$$P(3 \le X) = 1 - P(X < 3) = 1 - \Big[P(X = 0) + P(X = 1) + P(X = 2)\Big] = 1 - \left[\left(\frac{4}{5}\right)^{10} + \right.$$

$$\left. + \binom{10}{1} \cdot \left(\frac{1}{5}\right)^{1} \cdot \left(\frac{4}{5}\right)^{9} + \binom{10}{2} \cdot \left(\frac{1}{5}\right)^{2} \cdot \left(\frac{4}{5}\right)^{8}\right] \approx 0{,}322.$$

c)

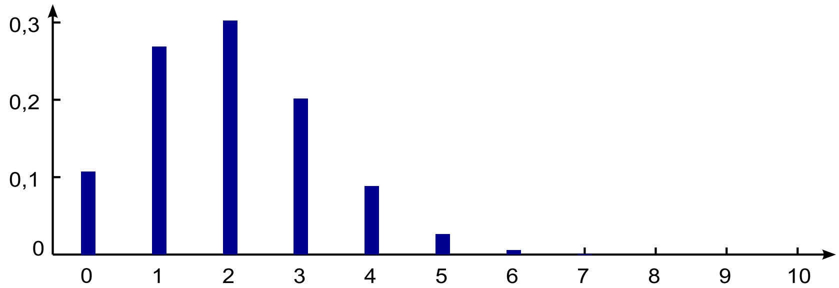
$$P(X < 2) = P(X = 0) + P(X = 1) = \left(\frac{4}{5}\right)^{10} + \binom{10}{1} \cdot \left(\frac{1}{5}\right)^{1} \cdot \left(\frac{4}{5}\right)^{9} \approx 0{,}376.$$

d)

$$P(2 < X \leq 5) = P(X=3) + P(X=4) + P(X=5) = \binom{10}{3} \cdot \left(\frac{1}{5}\right)^3 \cdot \left(\frac{4}{5}\right)^7 + \binom{10}{4} \cdot \left(\frac{1}{5}\right)^4 \cdot \left(\frac{4}{5}\right)^6 + \\ + \binom{10}{5} \cdot \left(\frac{1}{5}\right)^5 \cdot \left(\frac{4}{5}\right)^5 \approx 0{,}316.$$

e)

$$E(X) = n \cdot p = 10 \cdot 0{,}2 = 2; \quad D(X) = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{10 \cdot 0{,}2 \cdot 0{,}8} \approx 1{,}265.$$



9.2. ábra. *Binomiális eloszlás $n = 10$ és $p = 0{,}2$ paraméterekkel.*

⇐9.2. feladat

Az eloszlásból látható, hogy a legnagyobb valószínűsége (kb. 0,3) annak van, hogy a garanciális idő alatt 2 gép hibásodik meg; míg a legkisebb valószínűsége (kb. 10^{-7}) annak, hogy mindegyik gép meghibásodik. Kiszámolható az is, hogy annak a valószínűsége, hogy legalább hét gép meghibásodik a garanciális idő alatt, kevesebb mint 0,001. Mivel ez a valószínűség nagyon kicsi, ezért bár előfordulhat, hogy legalább hét gép hibásodik meg, azért ebben az esetben érdemes lenne arra gyanakodni, hogy nem igaz az, hogy a megadott idő alatt egy gép meghibásodásának a valószínűsége 0,2. Ilyen jellegű gondolatmenettel később a statisztikában a hipotézisvizsgálatoknál fogunk találkozni.

9.3. Hipergeometriai eloszlás

Korábban már volt szó a visszatevés nélküli mintavételről. A hipergeometriai eloszlással éppen a visszatevés nélküli mintavételnél a kihúzott selejtek számát adhatjuk meg. Később látni fogjuk, hogy a hipergeometriai eloszlás bizonyos körülmények között helyettesíthető (közelíthető) binomiális eloszlással.

9.3. definíció: Tekintsünk m darab elemet (bármit), amelyből s darabot valamilyen ok miatt megkülönböztetünk a többi $m-s$ darabtól (például selejtes). Ezután találomra kiválasztunk az m elemből n darabot visszatevés nélkül, ahol $n \le s$ és $n \le m-s$ (tehát előfordulhat az is, hogy minden kiválasztott selejtes, és az is, hogy egyik sem az). Legyen az X valószínűségi változó értéke az n darab kiválasztott között lévő megkülönböztetett elemek száma, így X lehetséges értékei $0,1,2,\ldots,n$. Ekkor X hipergeometriai eloszlású valószínűségi változó az m, s és n paraméterekkel, eloszlása:

$$P(X=k) = \frac{\binom{s}{k}\cdot\binom{m-s}{n-k}}{\binom{m}{n}} \qquad k=0,1,2,\ldots,n\,.$$

Jelölje p a selejtesek arányát, azaz legyen $p = s/m$. Belátható, hogy a valószínűségi változó várható értéke és szórása:

$$E(X) = n\cdot\frac{s}{m} = n\cdot p \qquad D(X) = \sqrt{np(1-p)\left(1-\frac{n-1}{m-1}\right)}\,.$$

Ahogy korábban is mondtuk, az eloszlás képletének bemagolása helyett célszerűbb annak a „logikáját” megjegyezni, azt, hogy hogyan jön ki. A nevezőben az összes eset száma (m tárgyból választunk ki n-et), míg a számlálóban a jó esetek száma (az s selejtesből választunk k darabot, míg az $m-s$ nem selejtesből a maradék $n-k$ darabot választjuk ki) áll.

9.3. feladat. Egy áruháznak 80 tv-je van raktáron. Energiafogyasztás szempontjából 30 tv A+, míg 50 tv B besorolású. Minőségellenőrök véletlenszerűen kiválasztanak a raktáron levő tv-k közül ötöt. Mennyi

a) a valószínűsége, hogy a kiválasztott termékek közt pontosan két darab A+-os készülék lesz;

b) a valószínűsége, hogy a kiválasztott termékek közt B-s készülékből lesz több;

c) a kiválasztott termékek közt az A+-os készülékek számának a várható értéke, illetve a szórása?

Megoldás: A készülékeket az energiafogyasztás alapján két részre osztottuk. Mivel bennünket most a kiválasztott A+-os készülékek száma érdekel, ezért legyen X a kiválasztott A+-os készülékek száma. X hipergeometriai eloszlású valószínűségi változó, $m = 80$, $n = 5$ és $s = 30$ paraméterekkel.

a)

$$P(X = 2) = \frac{\binom{30}{2} \cdot \binom{50}{3}}{\binom{80}{5}} \approx 0{,}355.$$

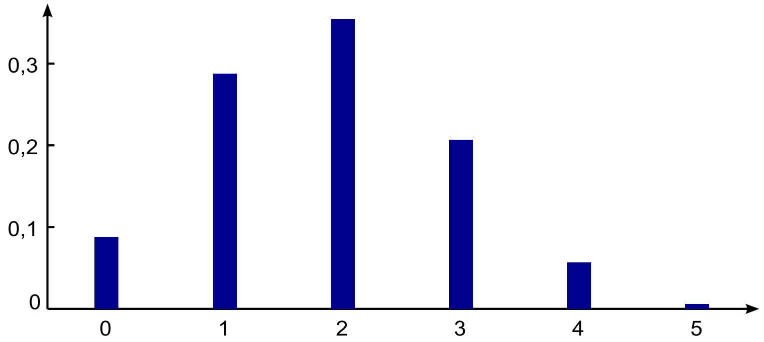
b) A B-s készülékek száma akkor lesz több, ha $X \leq 2$. Ezért a kérdéses valószínűség:

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{\binom{30}{0} \cdot \binom{50}{5}}{\binom{80}{5}} + \frac{\binom{30}{1} \cdot \binom{50}{4}}{\binom{80}{5}} + \frac{\binom{30}{2} \cdot \binom{50}{3}}{\binom{80}{5}} \approx 0{,}730.$$

c)

$$E(X) = n \cdot \frac{s}{m} = 5 \cdot \frac{30}{80} = \frac{15}{8} \qquad D(X) = \sqrt{np(1-p)\left(1 - \frac{n-1}{m-1}\right)} = \sqrt{5 \cdot \frac{3}{8} \cdot \frac{5}{8} \cdot \left(1 - \frac{5-1}{80-1}\right)} \approx 1{,}055.$$

⇐9.3. feladat



9.3. ábra. *A 9.3. feladatban szereplő eloszlás.*

Érdekesség: A háztartási gépeket energiahatékonyság alapján 7 osztályba sorolják. Az egyes osztályok hatékonyság alapján csökkenő sorrendben a következők: A+++, A++, A+, A, B, C és D. Az Európai Unióban 2012 júliusától már csak A+, vagy ennél jobb új hűtőgép hozható forgalomba.

9.4. feladat. Az 52 lapos franciakártya-csomagban a lapok színe négyféle lehet (káró, kőr, treff, pikk). Mindegyik színből 13, különböző értékű lap van a pakliban (kettestől az ászig). Véletlenszerűen kiválasztunk 5 lapot (visszatevés nélkül) egy csomagból.

a) Mennyi a valószínűsége, hogy a lapok között pontosan három treff lesz?

b) Mennyi a valószínűsége, hogy a lapok közt legalább három treff lesz?

c) Mennyi a valószínűsége, hogy mind a 4 ászt kihúzzuk?

d) Mennyi a valószínűsége, hogy mind az 5 lap ugyanolyan színű?

Megoldás:

a) A feladat szempontjából a kártyákat két részre oszthatjuk, treffekre és nem treffekre. Mivel visszatevés nélkül húzunk, ezért hipergeometriai eloszlással van dolgunk. A kihúzott 5 lap között pontosan akkor van 3 treff, ha a 13 treff lapból kiválasztunk hármat, a 39 nem treffből pedig ettől függetlenül kettőt, így a kérdezett valószínűség:

$$P(\text{pontosan három treff}) = \frac{\binom{13}{3} \cdot \binom{39}{2}}{\binom{52}{5}} \approx 0{,}082.$$

b) Az előző rész alapján:

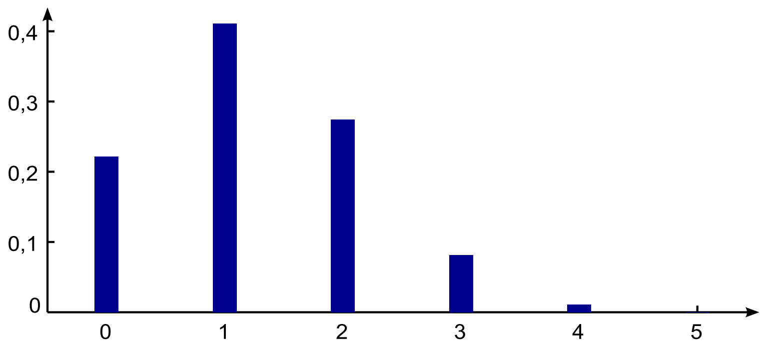
$$P(\text{legalább három treff}) = P(\text{pontosan három treff}) + P(\text{pontosan négy treff}) + P(\text{pontosan öt treff}) =$$
$$= \frac{\binom{13}{3} \cdot \binom{39}{2}}{\binom{52}{5}} + \frac{\binom{13}{4} \cdot \binom{39}{1}}{\binom{52}{5}} + \frac{\binom{13}{5} \cdot \binom{39}{0}}{\binom{52}{5}} \approx 0{,}093.$$

c) Ismét hipergeometriai eloszlással van dolgunk, hiszen a lapokat most is két részre osztjuk, csak más szempont alapján. Számunkra most csak az érdekes, hogy a kihúzott lap ász vagy nem. A kihúzott lapok között pontosan akkor van négy ász, ha mind a négy ászt kiválasztjuk, és a maradék 48 (nem ász) lapból választunk még egyet. Ezért a kérdéses valószínűség:

$$P(\text{pontosan négy ász}) = \frac{\binom{4}{4} \cdot \binom{48}{1}}{\binom{52}{5}} \approx 1{,}8 \cdot 10^{-5}.$$

d) Az, hogy mindegyik lap ugyanolyan színű, négyféleképpen valósulhat meg: vagy mindegyik treff, vagy mindegyik pikk, vagy mindegyik káró vagy mindegyik kőr. Mivel ez a négy esemény kizárja egymást, azért a kérdéses valószínűség ezen négy esemény valószínűségének összegével egyezik meg. Továbbá mivel a színeknek nincs jelentősége, ezért annak a valószínűsége, hogy mindegyik lap pikk, nyilván ugyanannyi, mint annak a valószínűsége, hogy mindegyik treff (káró, kőr). Ezeket felhasználva:

$$P(\text{mindegyik lap egyforma színű}) = P(\text{mind treff}) + P(\text{mind pikk}) + P(\text{mind káró}) + P(\text{mind kőr}) =$$
$$= 4 \cdot \frac{\binom{13}{5}}{\binom{52}{5}} \approx 0{,}002.$$



9.4. ábra. *A treff lapok számának eloszlása a 9.4. feladatban.*

⇐9.4. feladat

Érdekesség: A 9.4. feladat *c*, illetve *d* részében gyakorlatilag azt számoltuk ki, hogy az ötlapos pókerben mennyi az esélye annak, hogy osztásra ász pókert, illetve flusht (öt egyszínű lapot) kapunk.

Önellenőrzés

1. Egy kosárlabdázó 0,9 valószínűséggel dob be egy büntetőt. Mennyi a valószínűsége, hogy 14 büntetőből legfeljebb 11-t dob be?

☐ 0,114 ☐ 0,257 ☐ 0,842 ☐ 0,158

2. Az X binomiális eloszlású valószínűségi változó egyik paramétere $n = 10$. Tudjuk még, hogy $E(X) = 1$. Mennyi a $P(X \leq 1)$ értéke?

☐ 0,349 ☐ 0,736 ☐ 0,376 ☐ 0,268

3. Egy 32 lapos magyarkártya-csomagból visszatevéssel kihúzunk 6 lapot. Mennyi a valószínűsége, hogy a kihúzott lapok közt legalább 2 piros lesz?

☐ 0,297 ☐ 0,543 ☐ 0,356 ☐ 0,466

4. Mennyi az előző, 3. feladat eredménye, ha a lapokat visszatevés nélkül húzzuk ki?

☐ 0,476 ☐ 0,467 ☐ 0,365 ☐ 0,454

5. Egy vetélkedőben a játékosnak 15 táska közül kell hármat választania. Három táskában egymillió, a többi táskában százezer Ft van. Mennyi a valószínűsége, hogy a játékos legalább 2 millió Ft-t nyer?

☐ 0,079 ☐ 0,086 ☐ 0,068 ☐ 0,081

6. Egy gólyabálon hatszázan vesznek részt, közülük kétszázan elsősök. A résztvevők között visszatevés nélkül kisorsolnak három egyforma nyereményt. Mennyi a valószínűsége, hogy lesz elsős a nyertesek között?

☐ 0,296 ☐ 0,446 ☐ 0,654 ☐ 0,704

12. LECKE

Nevezetes diszkrét eloszlások II.

9.4. Geometriai eloszlás

Sokszor előfordul, hogy egy kísérletet hajtunk végre egymás után többször egészen addig, míg egy adott esemény be nem következik (például addig dobálunk a dobókockával, míg hatost nem dobunk). Ilyenkor az adott esemény bekövetkeztéig elvégzett kísérletek számát geometriai eloszlással írhatjuk le. Mint majd látni fogjuk, az eloszlás elnevezését a valószínűségek mértani (geometriai) sorozattal való kapcsolata indokolja.

9.4. definíció: Tegyük fel, hogy addig hajtunk végre független kísérleteket a p valószínűségű A esemény megfigyelésére, amíg az be nem következik. A szükséges kísérletek száma legyen X, a lehetséges értékek $1{,}2,\ldots$ (felső korlát nincs). Ekkor X geometriai eloszlású valószínűségi változó a p paraméterrel, eloszlása:

$$P(X=k)=(1-p)^{k-1}\cdot p \qquad k=1{,}2,\ldots.$$

Először is vegyük észre, hogy a $P(X=1)$, $P(X=2)$, $P(X=3),\ldots$ számok mértani (geometriai) sorozatot alkotnak, hiszen $P(X=k+1)=(1-p)\cdot P(X=k)$, azaz a mértani sorozat kvóciense éppen $(1-p)$. Most nézzük meg, hogyan kaptuk az eloszlás képletét: az A esemény pontosan akkor következik be először a k. kísérletben, ha az első $(k-1)$ kísérlet során nem következett be , míg a k. kísérletben bekövetkezett. Mivel az egyes kísérletek egymástól függetlenek, ezért

$$P(X=k)=P(\text{elsőre nem következik be})\cdot\ldots\cdot P((k-1)\text{.-re nem következik be})\cdot P(k\text{.-ra bekövetkezik})=$$
$$=\underbrace{(1-p)\cdot(1-p)\cdot\ldots\cdot(1-p)}_{k-1}\cdot p=(1-p)^{k-1}\cdot p$$

A geometriai eloszlású valószínűségi változó várható értéke és szórása a következőképpen számolható ki:

$$E(X)=\frac{1}{p} \qquad D(X)=\sqrt{\frac{1-p}{p^2}}.$$

9.5. feladat. Egy rádióadón minden nap telefonos játék zajlik. A játékosokat a betelefonálók közül véletlenszerűen választják ki, minden telefonáló $0{,}1$ valószínűséggel vehet részt a játékban. Kovács úr elhatározza, hogy a születésnapjától kezdve minden nap telefonál, egészen addig, míg játékba nem kerül.

a) Mi a valószínűsége, hogy a tizedik napon játszik először?

b) Mi a valószínűsége, hogy legalább hétszer kell telefonálnia, míg adásba kerül?

c) Mennyi Kovács úr telefonhívásainak a várható értéke?

Megoldás: Jelölje X azt, hogy hányadik napon kerül be először játékba Kovács úr. Mivel ugyanazt a kísérletet (Kovács úr betelefonál) hajtjuk végre addig, míg egy adott esemény (Kovács úr játékba kerül) be nem következik, ezért X geometria eloszlású valószínűségi változó, $p = 0{,}1$ paraméterrel.

a) Ha Kovács úr a tizedik napon kerül be először a játékba, akkor az első kilenc alkalommal nem járt sikerrel, míg a tizedik alkalommal igen, így

$$P(X = 10) = (1 - p)^9 \cdot p = (0{,}9)^9 \cdot 0{,}1 \approx 0{,}039.$$

b) Az, hogy Kovács úrnak legalább hétszer kell telefonálnia, pontosan azt jelenti, hogy az első hat alkalommal nem kerül be a játékba, ezért

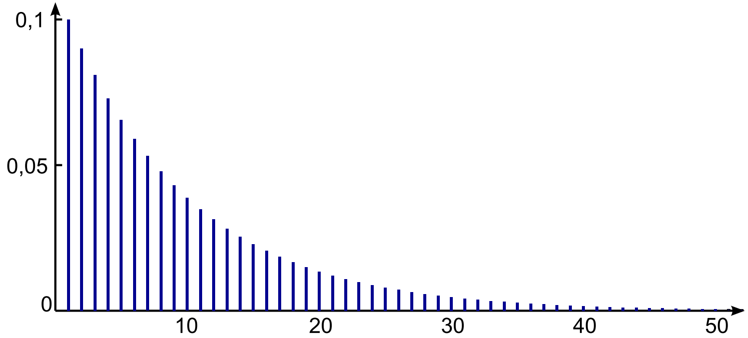
$$P(7 \leq X) = (1 - p)^6 = (0{,}9)^6 \approx 0{,}531.$$

c)

$$E(X) = \frac{1}{p} = \frac{1}{0{,}1} = 10,$$

azaz Kovács úr telefonálásainak várható értéke 10.

⇐9.5. feladat



9.5. ábra. *A 9.6. feladatban szereplő geometriai eloszlás ($p = 0{,}1$).*

Az előző feladathoz hasonlóan általánosan is igaz, hogy ha X geometria eloszlású, akkor $P(k \leq X) = (1-p)^{k-1}$. Ugyanis $k \leq X$ pontosan azt jelenti, hogy a megfigyelt A esemény az első $(k-1)$ kísérlet során nem következik be. Mivel a kísérletek egymástól függetlenek, ezért ennek a valószínűsége $(1-p)^{k-1}$.

9.6. feladat. Valaki a 18. születésnapját követően minden héten felad egy ötöslottó szelvényt ugyanazokkal a (családi) számokkal. Mennyi

a) a valószínűsége, hogy az ötödik héten nyer (azaz lesz legalább két találata) először;

b) a valószínűsége, hogy az első évben egyszer sem nyer;

c) az első nyerésig eltelt hetek számának a várható értéke?

Megoldás:

a) Jelölje X, hogy hányadik héten nyer az illető először. Mivel mindig ugyanazt a szelvényt adja fel (vagyis ugyanazt a kísérletet végezzük el), ezért X geometria eloszlású. Az eloszlás p paramétere azonban most nem ismert, először ezt kell kiszámolnunk.
Egy adott héten akkor nyerünk, ha 2,3,4 vagy 5 számot találunk el. A találatok száma hipergeometriai eloszlást követ, hiszen a számokat két csopotba oszthatjuk: van az általunk bejelölt 5 szám, és a maradék 85 (a két találat például azt jelenti, hogy a mi számainkból pontosan kettőt, a maradék 85-ből pedig pontosan hármat húznak ki). Ezért

$$p = P(\text{egy adott héten nyerünk}) = \frac{\binom{5}{2}\cdot\binom{85}{3}}{\binom{90}{5}} + \frac{\binom{5}{3}\cdot\binom{85}{2}}{\binom{90}{5}} + \frac{\binom{5}{4}\cdot\binom{85}{1}}{\binom{90}{5}} + \frac{\binom{5}{5}\cdot\binom{85}{0}}{\binom{90}{5}} \approx 0{,}023.$$

Egyelőre tehát meghatároztuk az X geometria eloszlású valószínűségi változó p paraméterét: $p = 0{,}023$. Most már megválaszolhatjuk a feladat kérdéseit:

$$P(\text{az ötödik héten nyer először}) = P(X = 5) = (1-p)^4 \cdot p = (1-0{,}023)^4 \cdot 0{,}023 \approx 0{,}021.$$

b) Az, hogy az első évben egyszer sem nyer az illető, azt jelenti, hogy az első 52 hét egyikén sem nyer, ezért

$$P(\text{az első évben nincs nyeremény}) = (1-p)^{52} = (1-0{,}023)^{52} \approx 0{,}298.$$

c)

$$E(X) = \frac{1}{p} = \frac{1}{0{,}023} \approx 43{,}5$$

azaz az első nyereményig eltelt hetek számának várható értéke 43,5.

⇐9.6. feladat

9.5. Negatív binomiális eloszlás

Előfordulhat, hogy egy kísérletet nem csak addig folytatunk, míg egy adott esemény először bekövetkezik, hanem például mondjuk addig, míg az esemény ötödször re be nem következik. Ilyenkor a szükséges kísérletek száma negatív binomiális eloszlással adható meg.

9.5. definíció: Tegyük fel, hogy addig hajtunk végre független kísérleteket a p valószínűségű A esemény megfigyelésére, amíg az n-szer be nem következik. A szükséges kísérletek száma legyen X, a lehetséges értékek $n, n+1, \ldots$ (felső korlát itt sincs). Ekkor X negatív binomiális eloszlású valószínűségi változó az n és p paraméterekkel, eloszlása:

$$P(X = n + k) = \binom{n+k-1}{n-1}(1-p)^k \cdot p^n \qquad k = 0,1,2,\ldots .$$

Nézzük ismét az eloszlás képletét! Az A esemény pontosan akkor következik n-edszerre az $(n+k)$-adik kísérletben, ha az első $(n+k-1)$ kísérletben pontosan $(n-1)$-szer következett be (aminek a valószínűsége $\binom{n+k-1}{n-1} \cdot p^{n-1} \cdot (1-p)^k$) és az $(n+k)$-adik kísérletben is bekövetkezik (aminek p a valószínűsége). Így valóban, annak a valószínűsége, hogy az A esemény éppen az $(n+k)$-adik kísérletben következik n-edszer:

$$\binom{n+k-1}{n-1} \cdot p^{n-1}(1-p)^k \cdot p = \binom{n+k-1}{n-1} \cdot (1-p)^k \cdot p^n.$$

Az n és p paraméterű negatív binomiális eloszlású valószínűségi változó tekinthető n darab független, p paraméterű geometriai eloszlású valószínűségi változó összegének is: az első sikeres kísérletig tart az első, itt újra kezdjük a számolást, a második sikeres kísérletig tart a második, stb. Ezt felhasználva belátható, hogy a várható értéke és szórása:

$$E(X) = n \cdot \frac{1}{p} \qquad D(X) = \sqrt{n \cdot \frac{1-p}{p^2}}.$$

9.7. feladat. Két testvér, András és Balázs a nyári szünet minden napján játszik egy asztalitenisz meccset. A tapasztalatok szerint 0,6 valószínűséggel András nyer meg egy meccset.

a) Mi a valószínűsége, hogy Balázs a hetedik napon szerzi meg a harmadik győzelmét?

b) Mi a valószínűsége, hogy Balázsnak több mint 10 napot kell várnia a harmadik győzelmére?

c) Mennyi a Balázs harmadik győzelméig eltelt napok számának a várható értéke ?

Megoldás:

a) Jelölje X Balázs harmadik győzelméig eltelt napok számát. Ekkor X negatív binomiális eloszlású valószínűségi változó az $n = 3$ és $p = 0{,}4$ paraméterekkel. Balázs pontosan akkor szerzi meg a harmadik győzelmét a hetedik napon, ha az első hat napon kétszer nyer, és a hetedik napon is nyer. A kérdezett valószínűség tehát:

$$P(X = 7) = \binom{6}{2} \cdot (0{,}6)^4 \cdot (0{,}4)^3 \approx 0{,}124.$$

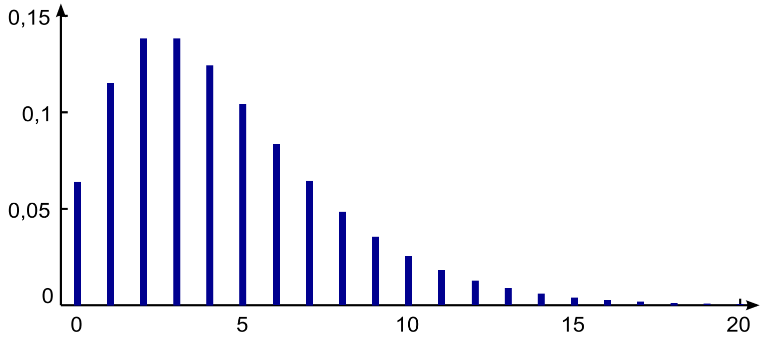
b) Balázsnak pontosan akkor kell több mint 10 napot várnia a harmadik győzelmére, ha az első 10 napon legfeljebb kétszer nyer. Így

$$P(10 < X) = P(\text{Balázs 0 meccset nyer tíz nap alatt}) + P(\text{Balázs 1 meccset nyer tíz nap alatt}) +$$
$$+ P(\text{Balázs 2 meccset nyer tíz nap alatt}) = \binom{10}{0} \cdot (0{,}6)^{10} \cdot (0{,}4)^0 + \binom{10}{1} \cdot (0{,}6)^9 \cdot (0{,}4)^1 +$$
$$+ \binom{10}{2} \cdot (0{,}6)^8 \cdot (0{,}4)^2 \approx 0{,}167.$$

c) Balázs harmadik győzelméig eltelt napok számának a várható értéke:

$$E(X) = \frac{n}{p} = \frac{3}{0{,}4} = 7{,}5.$$

⇐9.7. feladat



9.6. ábra. *A 9.7. feladatban szereplő negatív binomiális eloszlás* ($n = 3$, $p = 0{,}4$).

9.8. feladat. Egy boltba minden nap egy helyi sütöde szállítja a pékárut. A szerződés szerint az árunak legkésőbb reggel hétig meg kell érkeznie, továbbá az ötödik késés esetén a szerződés automatikusan megszűnik. Tudjuk, hogy az áru minden egyes nap 0,99 valószínűséggel időben megérkezik.

a) Mennyi a valószínűsége, hogy a szerződés (a késések miatt) a 200. napon szűnik meg ?

b) Várhatóan hány napig lesz érvényes a szerződés?

Megoldás:

a) Jelölje X a szerződés megszűnéséig eltelt napok számát. Ekkor X negatív binomiális eloszlású az $n = 5$ és $p = 0{,}01$ paraméterekkel. A szerződés pontosan akkor szűnik meg a 200. napon, ha a sütöde az első 199 napon pontosan négyszer késik, és a 200. napon is elkésik. Ezért

$$P(X = 200) = \binom{199}{4} \cdot (0{,}99)^{195} \cdot (0{,}01)^5 \approx 8{,}93 \cdot 10^{-4}.$$

b)

$$E(X) = \frac{n}{p} = \frac{5}{0{,}01} = 500,$$

azaz a szerződés megszűnéséig eltelt napok számának a várható értéke 500.

⇐9.8. feladat

9.6. Poisson-eloszlás

Előfordul, hogy sok, független, kis valószínűségű eseménnyel kapcsolatban nem az érdekel bennünket, hogy pontosan mely események következnek be, hanem csupán az, hogy közülük hány következik be. Ez a valószínűségi változó Poisson-eloszlásúnak tekinthető, még akkor is, ha az egyes események valószínűségei különbözőek. Események (pontok) térbeli vagy időbeli eloszlása is modellezhető Poisson-eloszlással, ha az egyes események egymástól függetlenek, továbbá minden térrészben (időintervallumban) a bekövetkezés valószínűsége arányos e részhalmaz méretével. Gyakran modellezhetőek Poisson-eloszlással például az alábbiak:

- adott időintervallumon belül a bekövetkezések száma (pl. egy óra alatt beérkező telefonhívások száma, tíz percen belül érkező járművek száma, öt perc alatt látott hullócsillagok száma, stb.),
- adott térrészben történő bekövetkezések száma (pl. sajtóhibák száma egy oldalon, hullócsillagok száma az égbolt egy adott részén),
- adott számú berendezés esetén adott idő alatt a meghibásodások száma.

9.6. definíció: Az X valószínűségi változó Poisson-eloszlású a $\lambda > 0$ paraméterrel, ha lehetséges értékei a nemnegatív egész számok $(0,1,\ldots)$, az eloszlása pedig

$$P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda} \qquad k = 0,1,2,\ldots .$$

A valószínűségi változó várható értéke és szórása:

$$E(X) = \lambda \qquad D(X) = \sqrt{\lambda}.$$

9.9. feladat. Egy vállalat naponta 8 és 16 óra közt tart telefonos ügyfélszolgálatot. Naponta átlagosan 120-an hívják fel az ügyfélszolgálatot. A hívások száma Poisson-eloszlású valószínűségi változónak tekinthető. Egy nap a telefonokat fogadó munkatársnak halaszthatatlan dolga akad, ezért 20 percre el kell mennie. Mennyi a valószínűsége, hogy ezen idő alatt

a) pontosan hat telefonáló lesz;

b) legalább négyen telefonálnak;

c) legfeljebb ketten telfonálnak?

Megoldás:

a) Jelölje X a 20 perc alatt betelefonálók számát. A feladat szövege alapján X Poisson-eloszlású. Mivel 8 óra alatt átlagosan 120-an telefonálnak, ezért 20 perc alatt átlagosan $120 \cdot 20/(8 \cdot 60) = 5$ telefonáló lesz, azaz $E(X) = 5$. Mivel a Possson-eloszlásnál $E(X) = \lambda$, ezért X egy $\lambda = 5$ paraméterű Poisson-eloszlású valószínűségi változó. Ebből az adódik, hogy

$$P(X = 6) = \frac{5^6}{6!} \cdot e^{-5} \approx 0{,}146.$$

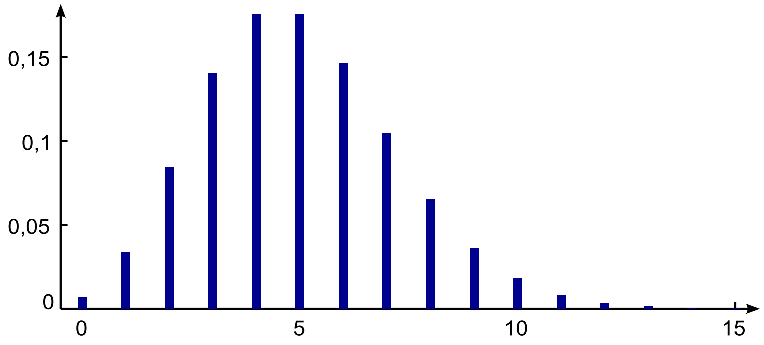
b) Az, hogy legalább négyen telefonálnak, azt jelenti, hogy vagy négyen, vagy öten, vagy hatan stb. telefonálnak. Mivel így végtelen sok valószínűséget kellene összeadni, ezért inkább a komplementer esemény valószínűségével fogunk számolni.

$$P(4 \leq X) = 1 - P(X < 4) = 1 - \Big[P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)\Big] =$$

$$= 1 - \left[\frac{5^0}{0!} \cdot e^{-5} + \frac{5^1}{1!} \cdot e^{-5} + \frac{5^2}{2!} \cdot e^{-5} + \frac{5^3}{3!} \cdot e^{-5}\right] \approx 0{,}735.$$

c)

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = \frac{5^0}{0!} \cdot e^{-5} + \frac{5^1}{1!} \cdot e^{-5} + \frac{5^2}{2!} \cdot e^{-5} \approx 0{,}125.$$



9.7. ábra. *A 9.9. feladatban szereplő Poisson-eloszlás ($\lambda = 5$).*

⇐9.9. feladat

9.10. feladat. Egy vasúti átjárón áthaladó autók száma Poisson-eloszlású valószínűségi változónak tekinthető. Tudjuk, hogy 0,0498 annak a valószínűsége, hogy két óra alatt egyetlen autó sem megy át az átjárón. Mennyi

a) annak a valószínűsége, hogy két óra alatt pontosan nyolc autó hajt át az átjárón;

b) annak a valószínűsége, hogy két óra alatt négynél több, de legfeljebb hét autó hajt át az átjárón;

c) az átjárón két óra alatt áthajtó autók számának a várható értéke?

Megoldás:

a) Tudjuk, hogy X Poisson-eloszlású, a λ paraméter azonban ismeretlen. Azonban ismerjük a $P(X = 0)$ valószínűség értékét, amiből a λ paramétert már ki tudjuk számolni.

$$0{,}0498 = P(X = 0) = \frac{\lambda^0}{0!} \cdot e^{-\lambda} = e^{-\lambda},$$

amiből $\lambda = -\ln 0{,}0498 = 3$ adódik. Most már válaszolhatunk a feladat kérdéseire.

$$P(X = 8) = \frac{3^8}{8!} \cdot e^{-3} \approx 0{,}008.$$

b)

$$P(4 < X \leq 7) = P(X = 5) + P(X = 6) + P(X = 7) = \frac{3^5}{5!} \cdot e^{-3} + \frac{3^6}{6!} \cdot e^{-3} + \frac{3^7}{7!} \cdot e^{-3} \approx 0{,}173.$$

c) $E(X) = \lambda = 3$, azaz két óra alatt várhatóan három autó halad át az átjárón.

⇐9.10. feladat

9.11. feladat. Kedvenc kuglófunkban a mazsolák száma Poisson-eloszlású valószínűségi változónak tekinthető. Az 1 kg-os kuglófban átlagosan 30 mazsola van. Egyik nap egy 25 dkg-os kuglófszeletet eszünk reggelire. Mennyi annak a valószínűsége, hogy

a) pontosan 6 mazsola lesz benne;

b) pontosan 6 mazsola lesz benne, feltéve, hogy hármat már meg is ettünk?

Megoldás:

a) Jelöljük X-szel a 25 dkg-os szeletben lévő mazsolák számát. Ekkor X Poisson-eloszlású, és mivel 1 kg-ban átlagosan 30 mazsola van, ezért X várható értéke, $E(X) = 30 \cdot 25/100 = 7{,}5$. Innen

$$P(X = 6) = \frac{7{,}5^6}{6!} \cdot e^{-7{,}5} \approx 0{,}137.$$

b) Vigyázzunk arra, hogy most egy feltételes valószínűséggel van dolgunk! Az, hogy három mazsolát már megettünk, azt jelenti, hogy a mazsolák száma legalább 3, így a feladat a $P(X = 6|X \geq 3)$ feltételes valószínűség kiszámolása. Használjuk először a feltételes valószínűség definícióját, majd utána a Poisson-eloszlás képletét:

$$P(X = 6|X \geq 3) = \frac{P(X = 6 \text{ és } X \geq 3)}{P(X \geq 3)} = \frac{P(X = 6)}{(X \geq 3)} = \frac{P(X = 6)}{1 - \Big[P(X = 0) + P(X = 1) + P(X = 2)\Big]}$$

$$= \frac{\frac{7{,}5^6}{6!} \cdot e^{-7{,}5}}{1 - \left[\frac{7{,}5^0}{0!} \cdot e^{-7{,}5} + \frac{7{,}5^1}{1!} \cdot e^{-7{,}5} + \frac{7{,}5^2}{2!} \cdot e^{-7{,}5}\right]} \approx 0{,}140.$$

⇐9.11. feladat

Önellenőrzés

1. Egy urnában 4 piros, 6 kék és 5 zöld golyó van. Visszatevéssel húzunk golyókat az urnából. Mennyi a valószínűsége, hogy az ötödik húzásra húzunk először pirosat?

☐ 0,289 ☐ 0,077 ☐ 0,364 ☐ 0,154

2. Az előző feladatban mennyi a valószínűsége, hogy hetedikre húzunk harmadszor pirosat?

☐ 0,082 ☐ 0,077 ☐ 0,125 ☐ 0,214

3. Egy szabályos dobókockát dobálunk. Mennyi a valószínűsége, hogy az első hatosig legalább négyet kell dobnunk?

☐ 0,196 ☐ 0,422 ☐ 0,579 ☐ 0,482

4. Egy szabályos dobókockát dobálunk. Mennyi a valószínűsége, hogy a tizedik dobásra dobunk másodszorra egyest?

☐ 0,032 ☐ 0,095 ☐ 0,085 ☐ 0,058

5. X egy $\lambda = 2$ paraméterű Poisson-eloszlású valószínűségi változó. Mennyi a $P(3 \leq X)$ valószínűség értéke?

☐ 0,677 ☐ 0,271 ☐ 0,323 ☐ 0,386

6. Egy üzletben a 8 és 12 óra közötti vásárlók száma Poisson-eloszlásúnak tekinthető, melynek várható értéke 26. Mennyi a valószínűsége, hogy 9 és 10 óra között pontosan hatan vásárolnak az üzletben?

☐ 0,157 ☐ 0,205 ☐ 0,260 ☐ 0,348

13. LECKE

Nevezetes folytonos eloszlások

10. Nevezetes folytonos eloszlások

10.1. Egyenletes eloszlás

Tegyük fel, hogy egy valószínűségi változó az $(a;b)$ intervallumból veheti fel az értékeit, és ha ezen belül veszünk egy $(c;d)$ részintervallumot, akkor annak a valószínűsége, hogy a valószínűségi változó ezen részintervallumba eső értéket vesz fel, a $(c;d)$ intervallum hosszával arányos, de nem függ attól, hogy az eredeti $(a;b)$-n belül hol helyezkedik el ez a részintervallum. Ebben az esetben azt mondjuk, hogy a valószínűségi változó egyenletes eloszlású az $(a;b)$ intervallumon. Ilyenkor tulajdonképpen geometriai valószínűségi mezőről van szó, hiszen a valószínűség a geometriai mérettel (hosszúsággal) arányos.

10.1. definíció: Az X valószínűségi változó egyenletes eloszlású az $(a;b)$ intervallumon, ha sűrűségfüggvénye

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{ha } a < x < b; \\ 0 & \text{különben.} \end{cases}$$

A sűrűségfüggvény alapján meghatározható az eloszlásfüggvénye

$$F(x) = \begin{cases} 0, & \text{ha } x \le a; \\ \dfrac{x-a}{b-a}, & \text{ha } a < x \le b; \\ 1, & \text{ha } x > b. \end{cases}$$

10.1. tétel: Ha X egyenletes eloszlású valószínűségi változó az $(a;b)$ intervallumon, akkor várható értéke és szórása:

$$E(X) = \frac{a+b}{2}, \qquad D(X) = \frac{b-a}{\sqrt{12}}.$$

10.1. feladat. Az egyik étteremnél házhozszállítás esetén a rendelés telefonos feladása és az étel megérkezése közt eltelt idő egyenletes eloszlású valószínűségi változónak tekinthető. A kiszállításig legalább 30, de legfeljebb 60 perc telik el. Jelölje X az étel megérkezéséig eltelt időt.

a) Írjuk fel X eloszlás-, illetve sűrűségfüggvényét!

b) Egy megrendelés esetén átlagosan mennyit kell várni a futár megérkezéséig?

c) Mennyi a valószínűsége, hogy a rendeléstől számítva 40 percen belül megkapjuk az ételt?

d) Mennyi a valószínűsége, hogy legalább 50 percet kell várnunk, mire meghozzák az ételt?

e) Az egyik nap pontban délben adjuk le a megrendelésünket. Háromnegyed 1-kor még nem jött meg az ebédünk. Mennyi a valószínűsége, hogy legkésőbb 12:53-ra megjön az étel?

Megoldás:

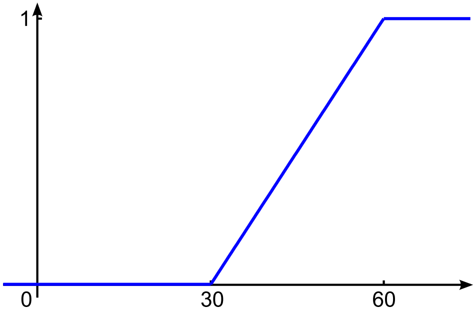
a) Mivel a megadott adatok alapján X egyenletes eloszlású a $(30; 60)$ intervallumon (azaz $a = 30$ és $b = 60$), ezért X eloszlásfüggvénye:

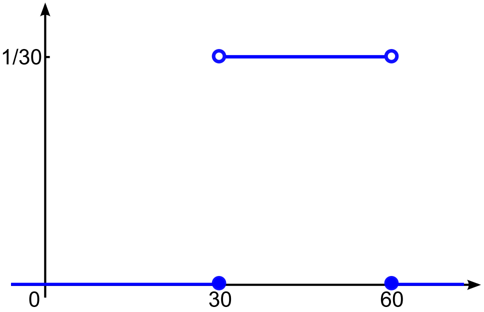
$$F(x) = \begin{cases} 0, & \text{ha } x \leq 30; \\ \dfrac{x-30}{60-30}, & \text{ha } 30 < x \leq 60; \\ 1, & \text{ha } x > 60. \end{cases}$$

X sűrűségfüggvénye:

$$f(x) = \begin{cases} \dfrac{1}{60-30}, & \text{ha } 30 < x < 60; \\ 0 & \text{különben.} \end{cases}$$

b) $E(X) = \dfrac{a+b}{2} = \dfrac{30+60}{2}$, azaz átlagosan 45 percet kell várnunk az étel megérkezéséig.





10.1. ábra. *A 10.1. feladatban szereplő, a* $(30;60)$ *intervallumon egyenletes eloszlás eloszlásfüggvénye (bal oldal) és sűrűségfüggvénye (jobb oldal).*

c) Az eloszlásfüggvényről szóló leckében tanultak szerint szerint $P(X<40)=F(40)$, így

$$P(X<40)=F(40)=\frac{40-30}{30}\approx 0{,}333.$$

d) Ismét alkalmazzuk az eloszlásfüggvényről korábban tanultakat:

$$P(50\le X)=1-P(X<50)=1-F(50)=\frac{50-30}{30}\approx 0{,}667.$$

e) Figyeljünk arra, hogy itt most egy feltételes valószínűséggel van dolgunk (a feltétel az, hogy $45\le X$).

$$P(X\le 53|45\le X)=\frac{P(X\le 53 \text{ és } 45\le X)}{P(45\le X)}=\frac{P(45\le X\le 53)}{P(45\le X)}=\frac{F(53)-F(45)}{1-F(45)}=\frac{\frac{23}{30}-\frac{15}{30}}{1-\frac{15}{30}}\approx 0{,}533.$$

⇐10.1. feladat

10.2. Exponenciális eloszlás

10.2. definíció: Az X valószínűségi változó $\lambda > 0$ paraméterű exponenciális eloszlású, ha sűrűségfügvénye

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda \cdot x}, & \text{ha } x > 0; \\ 0 & \text{különben.} \end{cases}$$

A sűrűségfüggvény alapján meghatározható az eloszlásfüggvénye:

$$F(x) = \begin{cases} 1 - e^{-\lambda \cdot x}, & \text{ha } x > 0; \\ 0, & \text{ha } x \leq 0. \end{cases}$$

10.2. tétel: Ha X egy λ paraméterű exponenciális eloszlású valószínűségi változó, akkor X várható értéke és szórása is a λ paraméter reciprokával egyezik meg, azaz

$$E(X) = D(X) = \frac{1}{\lambda}.$$

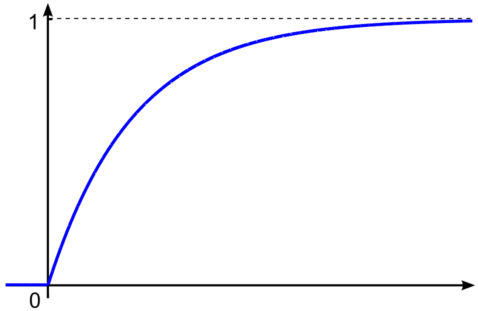
Exponenciális eloszlással modellezhető például a várakozási-, illetve a sorbanállási idő, továbbá bizonyos típusú berendezések, alkatrészek élettartama is.

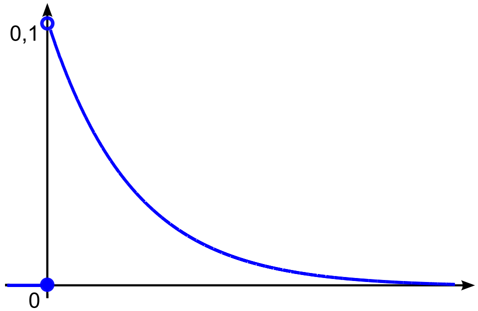
10.2. feladat. Az egyik autómosónál véletlenszerű érkezés esetén átlagosan 10 percet kell várnunk, míg sorra kerülünk. A várakozással töltött idő exponenciális eloszlású valószínűségi változónak tekinthető.

a) Jelölje X a várakozással töltött időt. Írjuk fel X eloszlás-, illetve sűrűségfüggvényét!

b) Mennyi a valószínűsége, hogy az érkezésünk után 5 percen belül sorra kerülünk?

c) Mennyi a valószínűsége, hogy legalább 14 percet kell várnunk, míg ránk kerül a sor?





10.2. ábra. *A* $\lambda = 0{,}1$ *paraméterű exponenciális eloszlás eloszlás- (bal oldal), illetve sűrűségfüggvénye (jobb oldal).*

Megoldás:

a) Mivel átlagosan 10 percet kell várnunk, ezért X várható értéke 10. Felhasználva az $E(X) = 1/\lambda$ összefüggést azt kapjuk, hogy $\lambda = \frac{1}{10} = 0{,}1$. Most már fel tudjuk írni az eloszlás-, és a sűrűségfüggvényt is. Az eloszlásfüggvény:

$$F(x) = \begin{cases} 1 - e^{-0{,}1 \cdot x}, & \text{ha } x > 0; \\ 0, & \text{ha } x \leq 0. \end{cases}$$

A sűrűségfüggvény:

$$f(x) = \begin{cases} 0{,}1 \cdot e^{-0{,}1 \cdot x}, & \text{ha } x > 0; \\ 0 & \text{különben.} \end{cases}$$

b) Használjuk az eloszlásfüggvény korábban megismert tulajdonságát!

$$P(X \leq 5) = F(5) = 1 - e^{-0{,}1 \cdot 5} \approx 0{,}393.$$

c)

$$P(X \geq 14) = 1 - P(X < 14) = 1 - F(14) = 1 - \left(1 - e^{-0{,}1 \cdot 14}\right) = e^{-1{,}4} \approx 0{,}247.$$

⇐10.2. feladat

10.3. feladat. Egy alkatrész várható élettartama 5 év. Az élettartam exponenciális eloszlású valószínűségi változónak tekinthető.

a) Mennyi a valószínűsége, hogy az alkatrész legalább 6 évig működőképes lesz?

b) Feltéve, hogy már 3 évet jól működött, mennyi a valószínűsége, hogy legalább 9 évig működőképes lesz?

Megoldás:

a) Jelölje X az alkatrész élettartamát. Ekkor egyrészt $E(X) = 5$, másrészt mivel X exponenciális eloszlású, ezért $E(X) = 1/\lambda$, amiből $\lambda = 1/5 = 0{,}2$ adódik. Írjuk fel X eloszlásfüggvényét:

$$F(x) = \begin{cases} 1 - e^{-0{,}2\cdot x}, & \text{ha } x > 0; \\ 0, & \text{ha } x \le 0. \end{cases}$$

Az, hogy az alkatrész legalább 6 évig működőképes, pontosan azt jelenti, hogy $6 \le X$. Az eloszlásfüggvény segítségével most már megkaphatjuk a kérdezett valószínűség értékét:

$$P(6 \le X) = 1 - P(X < 6) = 1 - F(6) = 1 - \left(1 - e^{-0{,}2\cdot 6}\right) = e^{-1{,}2} \approx 0{,}301.$$

b) Mivel tudjuk, hogy az alkatrész már három éve jól működik, ezért most egy feltételes valószínűséggel van dolgunk (az $X \ge 3$ esemény a feltétel). Írjuk fel először a megadott feltételes valószínűséget a definíció alapján, majd alkalmazzuk az eloszlásfüggvényről tanultakat.

$$P(X \ge 9 | X \ge 3) = \frac{P(X \ge 9 \text{ és } X \ge 3)}{P(X \ge 3)} = \frac{P(X \ge 9)}{P(X \ge 3)} = \frac{1 - F(9)}{1 - F(3)} = \frac{1 - \left(1 - e^{-0{,}2\cdot 9}\right)}{1 - \left(1 - e^{-0{,}2\cdot 3}\right)} = e^{-1{,}2} \approx 0{,}301.$$

⇐10.3. feladat

Az előző feladatot jobban megfigyelve, egy furcsaságra lehetünk figyelmesek. A második részben úgy is feltehettük volna a kérdést, hogy ha az alkatrész már 3 éve működik, akkor mi a valószínűsége, hogy még további 6 évig is működni fog. A megoldás során azt kaptuk, hogy ennek a valószínűsége megegyezik az első kérdésben feltett valószínűséggel, azaz azzal, hogy a berendezés legalább 6 évig működőképes lesz. Vagyis az alkatrész mintha nem vett volna tudomást az első 3 év elteltéről, más szóval nem öregedett.

A megoldásban bemutatott módon általánosan is belátható, hogy az exponenciális eloszlás rendelkezik az úgynevezett *örökifjú tulajdonsággal*. Ez annyit jelent, hogy ha egy alkatrész élettartama exponenciális eloszlású, akkor az alkatrész bizonyos értelemben véve nem öregszik. Ennek a pontos matematikai jelentése pedig az, hogy ha az alkatrész már t ideig működött, akkor annak valószínűsége, hogy még legalább s ideig fog működni (vagyis összesen legalább $t+s$ ideig fog működni), nem függ a megelőző t időtartamtól, ugyanakkora, mint annak valószínűsége, hogy a kezdettől számítva legalább s ideig működni fog. Tehát az eltelt t időtartam alatt az alkatrész nem használódott el, nem öregedett, ezért nevezzük ezt örökifjú tulajdonságnak.

10.3. tétel: Az exponencális eloszlás örökifjú tulajdonságú, azaz ha X exponenciális eloszlású valószínűségi változó, akkor

$$P(X \geq t+s | X \geq t) = P(X \geq s)\,.$$

Bizonyítás: Legyen X λ paraméterű exponenciális eloszlású valószínűségi változó. Ekkor a tételben szereplő feltételes valószínűség:

$$P(X \geq t+s | X \geq t) = \frac{P(X \geq t+s)}{P(X \geq t)} = \frac{1-F(s+t)}{1-F(t)} = \frac{1-(1-e^{-\lambda(s+t)})}{1-(1-e^{-\lambda t})} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} =$$
$$= e^{-\lambda s} = 1-(1-e^{-\lambda s}) = 1-F(s) = P(X \geq s)\,.$$

$\square$

10.3. Normális eloszlás

A normális eloszlás (vagy más néven Gauss-eloszlás) a legfontosabb valószínűség-eloszlás. A való életben számos véletlenszerű jelenség modellezhető normális eloszlással, például egy tömegtermelésben gyártott termék tömege és hossza, vagy egy méréssorozat esetén a mérés hibája. Általában, ha egy jelenséget nagy számú, egymástól független vagy csak kevéssé függő, véletlenszerű tényező határoz meg, továbbá az egyes tényezők egyenként csak kis mértékben járulnak hozzá a véletlentől függő ingadozásokhoz, és a tényezők hatásai összegződnek, akkor a jelenség eloszlása jól közelíthető normális eloszlással. Mint a későbbiekben látni fogjuk, a normális eloszlás kiemelkedő szerepet játszik a statisztikai vizsgálatok során is.

10.3. definíció: Az X folytonos eloszlású valószínűségi változót m, σ $(\sigma > 0)$ paraméterű normális eloszlásúnak nevezünk, ha a sűrűségfüggvénye

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} .$$

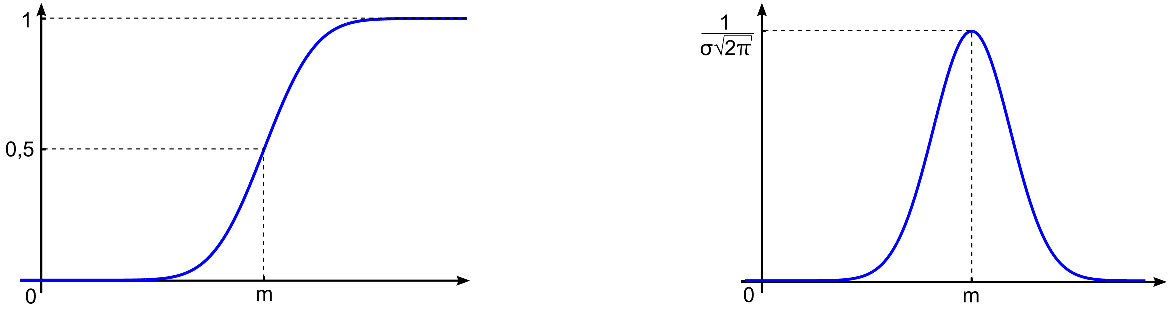
Eloszlásfüggvénye pedig

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}} \, \mathrm{d}t .$$

Várható értéke és szórása:

$$E(X) = m, \qquad D(X) = \sigma .$$

Aktivitás: Nézzen utána a Wikipédián, hogy mi az a Galton-deszka, és milyen kapcsolatban áll a normális eloszlással!



10.3. ábra. *A normális eloszlás eloszlás- (bal oldal), és sűrűségfüggvénye (jobb oldal).*

A normális eloszlás sűrűségfüggvénye az úgynevezett Gauss-görbe (szokás haranggörbének is nevezni). Mint a 10.3. ábrán látható, a sűrűségfüggvény grafikonja tengelyesen szimmetrikus az m várható értéknél az x tengelyre állított merőleges egyenesre, továbbá a függvény éppen a várható értéknél veszi fel a maximumát (a maximum értéke $\frac{1}{\sigma\cdot\sqrt{2\pi}}$). A görbe alakja (azaz, hogy mennyire csúcsos, illetve lapos) a σ szórástól függ. A haranggörbéhez hasonló grafikonokkal számos, a való életből vett statisztikánál is találkozhatunk. Például ezen az oldalon a 2009. évi középiskolai felvételi pontszámok hisztogramján is jól kivehető a haranggörbe alakja.

Sajnos az egyenletes és az exponenciális eloszlással ellentétben a normális eloszlás esetén az eloszlásfüggvény értékét egy adott pontban nem tudjuk egyszerűen kiszámolni, hiszen ehhez egy bonyolult improprius integrált kellene kiszámolnunk. Ez azt jelenti, hogy hiába ismerjük egy X normális eloszlású valószínűségi változó eloszlásfüggvényét, segítségével nem tudjuk meghatározni a $P(X < a)$ valószínűség értékét. Arra, hogy hogyan lehet mégis gyorsan meghatározni ezt a valószínűséget, a most következő részben kapunk választ.

10.3.1. Standard normális eloszlás

A normális eloszlások között kitüntetett szerep jut annak, amelynek várható értéke 0, szórása pedig 1, azaz $m = 0$ és $\sigma = 1$. Tetszőleges normális eloszlással kapcsolatos valószínűség meghatározását vissza lehet vezetni egy ilyen típusú normális eloszlással kapcsolatos kérdésre. Ezt standard normális eloszlásnak hívják, és fontossága miatt külön jelölése is van.

A sűrűségfüggvénye:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} .$$

Az eloszlásfüggvénye:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} \, \mathrm{d}t .$$

A standard normális eloszlás eloszlásfüggvényének ($\Phi(x)$) értékeit táblázatokban szokás közölni. Az 1. és 2. táblázatok a 0 és 4 közötti, két tizedesjegyű számokhoz tartozó Φ értékeket tartalmazzák. Mint látható, a táblázatok csak pozitív x-ekre tartalmazzák $\Phi(x)$ értékeit. Ennek oka a következő:

10.4. tétel: Legyen $\Phi(x)$ a standard normális eloszlás eloszlásfüggvénye. Ekkor:

$$\Phi(-x) = 1 - \Phi(x) .$$

Az előző tétel alapján ha X egy standard normális eloszlású valószínűségi változó, akkor tetszőleges a érték esetén a $P(X < a)$ valószínűség értéke, ami $\Phi(a)$-val egyenlő, a táblázatból kiolvasható. Mi a helyzet azonban olyankor, amikor X nem standard normális eloszlású? A következő tétel azt mutatja meg, hogy egy tetszőleges normális eloszlású valószínűségi változóból hogyan lehet standard normális eloszlásút „gyártani”.

10.5. tétel: Legyen X normális eloszlású valószínűségi változó, a várható értéke m, a szórása σ. Ekkor az

$$X^* = \frac{X - m}{\sigma}$$

valószínűségi változó standard normális eloszlású.

A tétel szerint tehát ha X egy (m,σ) paraméterű normális eloszlású valószínűségi változó, akkor a tételben leírt X^* valószínűségi változó standard normális eloszlású. X^*-ot az X standardizáltjának nevezzük. Nézzük most meg példákon keresztül, hogy ezen ismeret birtokában hogyan lehet például a $P(X < a)$ valószínűséget kiszámítani.

10.4. feladat. Egy gyárban az egyik gép által gyártott alkatrészek tömege normális eloszlású valószínűségi változónak tekinthető, melynek m várható értéke 30 dkg, σ szórása pedig 4 g. Mennyi a valószínűsége, hogy egy, a gép által gyártott alkatrész tömege

a) legfeljebb 302 g;

b) legalább 209 g;

c) pontosan 300 g;

d) legalább 208, de legfeljebb 303 g?

Megoldás:

a) Jelölje X a gép által gyártott alkatrészek grammban mért tömegét. Ekkor X egy $m = 300$ várható értékű és $\sigma = 4$ szórású normális eloszlású valószínűségi változó. A 10.5. tétel szerint az $X^* = \frac{X-300}{4}$ valószínűségi változó standard normális eloszlású. Ezt felhasználva a kérdezett valószínűség

$$P(X \le 302) = P\left(\frac{X - 300}{4} \le \frac{302 - 300}{4}\right) = P(X^* \le 0{,}5) = \Phi(0{,}5).$$

Az 1. táblázatból $\Phi(0{,}5)$ értékét a következőképpen kereshetjük ki: $0{,}5 = 0{,}5 + 0{,}00$, s ekkor $\Phi(0{,}5)$ éppen a táblázat $0{,}5$-es sorának és $0{,}00$-os oszlopának a „metszéspontjában” található érték. Így:

$$P(X \le 302) = \Phi(0{,}5) \approx 0{,}6915.$$

b)

$$P(209 \le X) = 1 - P(X < 209) = 1 - P\left(\frac{X-300}{4} < \frac{209-300}{4}\right) = 1 - P(X^* < -0{,}25) =$$
$$= 1 - \Phi(-0{,}25) = 1 - (1 - \Phi(0{,}25)) = \Phi(0{,}25).$$

A $\Phi(0{,}25)$ értéket ismét az 1. táblázatból kereshetjük ki: $0{,}25 = 0{,}2 + 0{,}05$, és ekkor a $\Phi(0{,}25)$ éppen a táblázat $0{,}2$-es sorának és $0{,}05$-ös oszlopának a „metszéspontjában" található érték. Ebből

$$P(209 \le X) = \Phi(0{,}25) \approx 0{,}5987.$$

c) Korábban a sűrűségfüggvénnyel foglalkozó leckében megtanultuk, hogy ha X folytonos eloszlású valószínűségi változó (azaz van sűrűségfüggvénye), akkor tetszőleges a érték esetén $P(X = a) = 0$. Mivel az alkatrészek tömege normális eloszlású (azaz folytonos) valószínűségi változónak tekinthető, ezért nulla annak a valószínűsége, hogy egy alkatrész tömege pontosan 30 dkg.

d)

$$P(208 \le X \le 303) = P\left(\frac{208-300}{4} \le \frac{X-300}{4} \le \frac{303-300}{4}\right) = P(-0{,}5 \le X^* \le 0{,}75) =$$
$$= \Phi(0{,}75) - \Phi(-0{,}5) = \Phi(0{,}75) - (1 - \Phi(0{,}5)) \approx 0{,}4649.$$

⇐10.4. feladat

10.5. feladat. Egy faüzemben készített deszkák hossza normális eloszlást követ 400 cm várható értékkel és 2 cm szórással. Egy deszka selejtesnek számít, ha a hossza legalább 5 cm-rel eltér 4 m-től.

a) Milyen selejtszázalékkal dolgozik az üzem?

b) Hogyan kellene változtatni a tűréshatárokat, hogy a selejtszázalék 0,5% legyen?

Megoldás:

a) A kérdés megválaszolásához azt kell kiszámítani, hogy mi a valószínűsége annak, hogy egy véletlenszerűen választott deszka selejtes. Legyen X a deszka hosszát jelentő normális eloszlású valószínűségi változó, $m = 400$, $\sigma = 2$. Egy deszka akkor nem selejtes, ha a hossza 395 és 405 cm közé esik, azaz

$$P(\text{nem selejt}) = P(395 < X < 405) = P\left(\frac{395-400}{2} < \frac{X-400}{2} < \frac{405-400}{2}\right) =$$
$$= P(-2{,}5 < X^* < 2{,}5) = \Phi(2{,}5) - \Phi(-2{,}5) = \Phi(2{,}5) - (1 - \Phi(2{,}5)) =$$
$$= 2 \cdot \Phi(2{,}5) - 1 \approx 2 \cdot 0{,}9938 - 1 = 0{,}9876\,.$$

Ezek szerint egy termék $0{,}9876$ valószínűséggel nem selejt, így $0{,}0124$ valószínűséggel selejt, tehát a selejtszázalék $1{,}24\%$.

b) Nyilván adott szórás mellett akkor lesz kisebb a selejtszázalék, ha a várható értéktől nagyobb eltérés is megengedett. Legyen ez a keresett eltérés d. Ekkor annak valószínűsége, hogy egy termék nem selejtes $99{,}5\%$, azaz $0{,}995$ (hiszen most csak $0{,}5\%$ selejt):

$$P(\text{nem selejt}) = P(400 - d < X < 400 + d) = P\left(\frac{-d}{2} < \frac{X-400}{2} < \frac{d}{2}\right) =$$
$$= P\left(\frac{-d}{2} < X^* < \frac{d}{2}\right) = \Phi\left(\frac{d}{2}\right) - \Phi\left(\frac{-d}{2}\right) = \Phi\left(\frac{d}{2}\right) - \left(1 - \Phi\left(\frac{d}{2}\right)\right) =$$
$$= 2 \cdot \Phi\left(\frac{d}{2}\right) - 1 = 0{,}995\,.$$

Ebből d már meghatározható:

$$2 \cdot \Phi\left(\frac{d}{2}\right) - 1 = 0{,}995$$

$$\Phi\left(\frac{d}{2}\right) = 0{,}9975$$

$$\frac{d}{2} = 2{,}81 \qquad \Longrightarrow \qquad d = 5{,}62\,.$$

Tehát a selejt százalék akkor lesz $0{,}5\%$, ha tűréshatár ± 5 cm helyett $\pm 5{,}62$ cm.

⇐10.5. feladat

10.6. feladat. Egy üdítőt gyártó üzemben gép végzi az üdítők üvegekbe töltését. Az üvegbe kerülő mennyiség normális eloszlású valószínűségi változónak tekinthető. Egy üvegbe átlagosan 1 liter üdítő kerül, továbbá $0{,}8413$ annak a valószínűsége, hogy egy véletlenszerűen választott üvegben legfeljebb $1{,}05$ liter folyadék van.

a) Mennyi a gép által az üvegekbe töltött folyadék szórása?

b) Az üvegek hány százalékában lesz legalább $0{,}99$ liter üdítő?

c) Az egyik üvegből kiöntünk 1 liter üdítőt egy pohárba. Mennyi a valószínűsége, hogy az üvegben marad még legalább $0{,}1$ liter üdítő?

Megoldás:

a) Jelölje X az üvegbe töltött üdítő literben mért mennyiségét. Az adatok alapján X normális eloszlású, a várható értéke $m = 1$. Ismert továbbá, hogy $P(X \leq 1{,}05) = 0{,}8413$. A σ szórás értékét ebből az egyenlőségből ki tudjuk számítani.

$$0{,}8413 = P(X \leq 1{,}05) = P\left(\frac{X-1}{\sigma} \leq \frac{1{,}05-1}{\sigma}\right) = P\left(X^* \leq \frac{0{,}05}{\sigma}\right) = \Phi\left(\frac{0{,}05}{\sigma}\right)$$

A táblázatból visszakeresve azt kapjuk, hogy $\dfrac{0{,}05}{\sigma} = 1$, amiből $\sigma = 0{,}05$, azaz az üvegbe töltött üdítő mennyiségének a szórása $0{,}05$ liter.

b)

$$P(0{,}99 \le X) = 1 - P(X < 0{,}99) = 1 - P\left(\frac{X-1}{0{,}05} < \frac{0{,}99-1}{0{,}05}\right) = 1 - P(X^* < -0{,}2) =$$
$$= 1 - \Phi(-0{,}2) = 1 - (1 - \Phi(0{,}2)) \approx 0{,}5793,$$

azaz körülbelül az üvegek $57{,}93$ százaléka tartalmaz legalább $0{,}99$ liter üdítőt.

c) Gyakorlatilag az a kérdés, hogy ha tudjuk, hogy egy üvegben legalább 1 liter üdítő van, akkor mennyi a valószínűsége, hogy legalább 1,1 liter üdítő van az üvegben. Vagyis egy feltételes valószínűséggel van dolgunk, melyben $1 \le X$ a feltétel. A feltételes valószínűség definícióját használva

$$P(X \ge 1{,}1 | X \ge 1) = \frac{P(X \ge 1{,}1 \text{ és } X \ge 1)}{P(X \ge 1)} = \frac{P(X \ge 1{,}1)}{P(X \ge 1)} = \frac{1 - P(X < 1{,}1)}{1 - P(X < 1)} = \frac{1 - P\left(\frac{X-1}{0{,}05} < \frac{1{,}1-1}{0{,}05}\right)}{1 - P\left(\frac{X-1}{0{,}05} < \frac{1-1}{0{,}05}\right)}$$
$$= \frac{1 - P(X^* < 2)}{1 - P(X^* < 0)} = \frac{1 - \Phi(2)}{1 - \Phi(0)} \approx 0{,}0456.$$

⇐10.6. feladat

Az előző feladatban definiált X valószínűségi változó, az üvegekbe töltött üdítő mennyisége, nyilván nem vehet fel negatív értékeket, sőt (az üvegek mérete miatt) akármilyen nagy értéket sem vehet fel. Ezzel szemben egy normális eloszlású valószínűségi változó bármekkora értéket felvehet. Vajon ez nem mond ellent annak, hogy feltettük, hogy az üvegbe töltött folyadék mennyisége normális eloszlású valószínűségi változóval modellezhető? A kérdés megválaszolásában segítségünkre lesz a következő feladat.

10.7. feladat. Legyen X egy normális eloszlású valószínűségi változó m várható értékkel és σ szórással. Számoljuk ki a következő két valószínűség értékét!

a) $P(|X - m| < 2 \cdot \sigma)$ *b)* $P(|X - m| < 4 \cdot \sigma)$

Megoldás:

a) Az abszolútérték definíciója szerint az $|X - m| < 2 \cdot \sigma$ feltétel azt jelenti, hogy

$$-2 \cdot \sigma < X - m < 2 \cdot \sigma,$$

amiből átrendezés után a vele ekvivalens

$$m - 2 \cdot \sigma < X < m + 2 \cdot \sigma$$

egyenlőtlenségrendszert kapjuk. Ezt felhasználva

$$P(|X - m| < 2 \cdot \sigma) = P(m - 2 \cdot \sigma < X < m + 2 \cdot \sigma) = P\left(\frac{m - 2 \cdot \sigma - m}{\sigma} < \frac{X - m}{\sigma} < \frac{m + 2 \cdot \sigma - m}{\sigma}\right) =$$
$$= P(X^* < 2) - P(X^* < -2) = \Phi(2) - \Phi(-2) = \Phi(2) - (1 - \Phi(2)) \approx 0{,}9544.$$

b) Az előző gondolatmenetet alkalmazva

$$P(|X - m| < 4 \cdot \sigma) = P(X^* < 4) - P(X^* < -4) = \Phi(4) - \Phi(-4) = \Phi(4) - (1 - \Phi(4)) \approx 0{,}999936.$$

Mivel a legtöbb táblázatban $\Phi(4)$ értéke már 1-re van kerekítve, ezért most $\Phi(4)$ értékét az EXCEL beépített STNORMELOSZL függvényének a segítségével határoztuk meg.

⇐10.7. feladat

Azt kaptuk tehát, hogy egy normális eloszlású valószínűségi változó értéke $0{,}9544$ valószínűséggel az $(m - 2 \cdot \sigma, m + 2 \cdot \sigma)$ intervallumba esik, míg $0{,}999936$ valószínűséggel az $(m - 4 \cdot \sigma, m + 4 \cdot \sigma)$ intervallumba esik. A 10.6. feladat esetében (melynél $m = 1$ és $\sigma = 0{,}05$) ez azt jelenti, hogy az üvegekbe töltött folyadék mennyisége $0{,}999936$ valószínűséggel legalább 8, de legfeljebb 12 deciliter. Vagyis gyakorlatilag nem kell amiatt aggódnunk, hogy normális eloszlást feltételezve az üvegekbe például negatív mennyiségű üdítő kerül.

Önellenőrzés

Önellenőrző kérdések

1. Legyen X egyenletes eloszlású valószínűségi változó az $(5; 15)$ intervallumon. Mennyi a $P(8 \leq X)$ valószínűség értéke ? Megold.
2. Az előző feladatban mennyi a $P(3 \leq X \leq 10)$ valószínűség értéke? Megold.
3. Egy berendezés átlagosan 6 évig működik; az élettartama exponenciális eloszlású valószínűségi változónak tekinthető. Mennyi a valószínűsége, hogy a berendezés legalább 8 évig működik?
   - ☐ 0,264 ☐ 0,736 ☐ 0,472 ☐ 0,528
4. Az előző feladatban mennyi a valószínűsége, hogy a berendezés legalább 3, de legfeljebb 8 évig működik?
   - ☐ 0,524 ☐ 0,264 ☐ 0,606 ☐ 0,343
5. X egy normális eloszlású valószínűségi változó, melynek a várható értéke 10, szórása pedig 0,8. Mennyi a $P(X \leq 9{,}76)$ valószínűség értéke?
   - ☐ 0,6179 ☐ 0,3821 ☐ 0,7881 ☐ 0,2119
6. Az előző feldatban mennyi a $P(10{,}2 < X \leq 10{,}4)$ valószínűség értéke?
   - ☐ 0,6915 ☐ 0,5987 ☐ 0,0928 ☐ 0,7881
7. Az X normális eloszlású valószínűségi változó várható értéke 50, szórása 4. Határozzuk meg azt a d értéket, melyre teljesül, hogy X értéke $0{,}95$ valószínűséggel $50 - d$ és $50 + d$ közé esik!
   - ☐ 0,975 ☐ 7,84 ☐ 1,96 ☐ 6,58

14. LECKE

Kapcsolatok a nevezetes diszkrét eloszlások között

11. Kapcsolatok a nevezetes diszkrét eloszlások között (határeloszlás tételek)

Előfordulhat, hogy egy véletlen jelenséggel kapcsolatos valószínűség meghatározásakor vagy nem ismerjük pontosan a jelenséget leíró valószínűségi eloszlást, vagy ismerjük ugyan, de még ennek a segítségével is csak nehezen tudjuk kiszámolni a valószínűség értékét. Ilyenkor megpróbálhatunk olyan egyszerűbb (kevesebb paramétert tartalmazó, vagy könnyebben számolható) eloszlást keresni, mellyel a kérdéses valószínűség egy jó közelítését kapjuk. Tekintsük például a következő feladatot.

11.1. feladat. Egy biztosító társaságnak legalább 50 ezer ügyfele van. Az ügyfelek 60 százaléka negyven évnél fiatalabb. Év végén a társaság kisorsol az ügyfelei között öt egyforma ajándékot. Mennyi a valószínűsége, hogy pontosan két nyertes lesz 40 évnél fiatalabb?

Megoldás: A 40 éven aluli nyertesek száma hipergeometriai eloszlású (hiszen gyakorlatilag visszatevés nélküli mintavételről van szó). Azonban nem ismerjük az eloszlás összes paraméterét, hiszen nem tudjuk, hogy pontosan hány ügyfele van a társaságnak. Így a kérdezett valószínűség értékét nem tudjuk pontosan meghatározni. Hogyan kaphatnánk meg azonban egy közelítő értéket? A következőképpen okoskodhatunk. Jelölje m az ügyfelek számát. A negyven éven aluli ügyfelek száma ekkor $0{,}6 \cdot m$. Ha elsőre negyven éven aluli ügyfelet sorsolunk ki, akkor a következő húzás előtt a negyven éven aluli ügyfelek aránya a maradék ügyfelek közt $(0{,}6m-1)/(m-1)$ lesz. Mivel m nagyobb, mint 50 ezer, ezért ez az arány gyakorlatilag $0{,}6$-nek tekinthető ($m = 50000$ esetén az arány $0{,}599992$). Hasonlóan kiszámolhatnánk, hogy ha elsőre negyven éven felülit húzunk, akkor a maradék ügyfelek között a negyven éven aluliak aránya továbbra is tekinthető 0,6-nek. Ugyanez a gondolatmenet elmondható a második, harmadik és negyedik húzás esetén is. Ezért aztán a kérdezett valószínűség egy jó közelítését kaphatjuk akkor, ha feltesszük, hogy minden egyes nyertes sorsolásakor a negyven éven aluliak aránya pontosan 60 százalék. Ez pedig nem mást jelent, mint hogy úgy tekintünk a feladatra, mintha a nyerteseket visszatevéses mintavétellel húznák ki, vagyis a tényleges hipergeometriai eloszlás helyett a binomiális eloszlással számoljuk ki a valószínűséget!

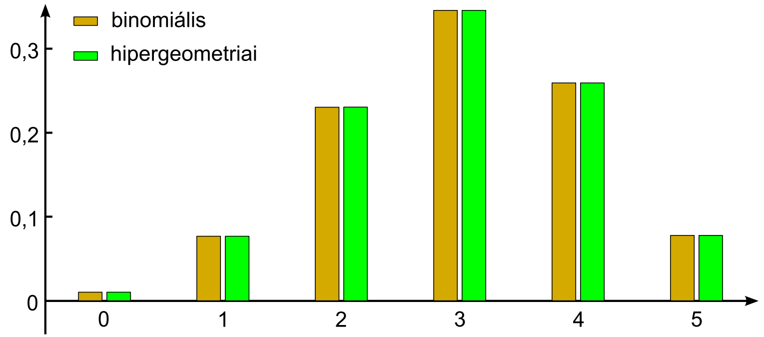
$$P(\text{pontosan két negyven év alatti nyertes}) \approx \binom{5}{2} \cdot 0{,}6^2 \cdot 0{,}4^3 = 0{,}2304.$$

⇐11.1. feladat

Tegyük fel, hogy az előző feladatban tudjuk, hogy a biztosítónak pontosan 50000 ügyfele van. Ekkor a negyven éven aluli ügyfelek száma 30000, így a hipergeometriai eloszlás segítségével most már kiszámolhatjuk a kérdezett valószínűség pontos értékét:

$$P(\text{pontosan két negyven év alatti nyertes}) = \frac{\binom{30000}{2} \cdot \binom{20000}{3}}{\binom{50000}{5}} \approx 0{,}230404.$$

A fenti tört értékét érdemes mindenkinek a saját számológépén is kiszámolni. Azt tapasztalhatjuk, hogy még (főként egyszerűbb) számológéppel sem könnyű az értékét kiszámolni (lehet, hogy valakinek nem is sikerül). Ezért még ha pontosan ismerjük is a hipergeometria eloszlás paramétereit, akkor is szóba jöhet a binomiális eloszlással való közelítése. Továbbá figyeljük meg, hogy 50000 ügyfél esetén a binomiális eloszlással számolva 5 tizedesjegyre pontosan megkapjuk az eredményt, vagyis igen jó közelítő megoldást kaptunk!



11.1. ábra. *A* $m = 50000$, $s = 30000$, $n = 5$ *paraméterű hipergeometria eloszlás, és az* $n = 5$, $p = 0{,}6$ *paraméterű binomiális eloszlás.*

11.1. A hipergeometriai eloszlás közelítése binomiális eloszlással

11.1. tétel: Legyen p egy 0 és 1 közé eső szám, legyen továbbá s_m olyan egész értékű sorozat, melyre

$$\lim_{m\to\infty} \frac{s_m}{m} = p\,.$$

Legyen n olyan m-től és s_m-től független egész állandó, melyre $n \le s_m$ és $n \le m - s_m$. Ekkor tetszőleges rögzített k esetén $(0 \le k \le n)$:

$$\lim_{m\to\infty} \frac{\binom{s_m}{k} \cdot \binom{m-s_m}{n-k}}{\binom{m}{n}} = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}\,.$$

Tehát az m, s_m és n paraméterű hipergeometriai eloszlások sorozata az n és p paraméterű binomiális eloszláshoz közelít.

Az előző tételt hétköznapi nyelvre lefordítva a következőt mondhatjuk: tegyük fel, hogy van összesen m termékünk, melyből s darabot megkülönbeztetünk a többitől; legyenek ezek például a selejtes termékek, a többi $m - s$ termék pedig legyen hibátlan. Válasszunk ki (visszatevés nélkül) az összes termékből n darabot. Ha a kiválasztások n száma mind a selejtes termékek s számához, mind a hibátlan termékek $m - s$ számához képest „kicsi”, akkor a kiválaszott termékek közti selejtes termékek száma jól közelíthető az (n,p) paraméterű binomiális eloszlással, melyben $p = s/m$.

11.2. feladat. Egy édességeket forgalmazó cég pályázatot hirdet. Aki egy borítékban a megadott címre elküldi a cég 5 különböző termékének a csomagolópapírját, az sorsoláson vesz részt, melyen tíz darab, egyenként 100 ezer Ft-os ajándékutalványt sorsolnak ki. Határidőre 500 ezer érvényes pályázat érkezik be, melyek közül 80 ezret Pest megyében adtak fel. Mennyi a valószínűsége, hogy

a) pontosan négy Pest megyei pályázat nyer;

b) legfeljebb három Pest megyei pályázat nyer?

Megoldás:

a) Jelölje X a nyertes Pest megyei pályázatok számát; X ekkor hipergeometriai eloszlású. Mivel a nyertesek száma (10) mind a Pest megyei pályázatok számához (80 ezer), mind a nem Pest megyei pályázatok számához képest is kicsi, ezért a nyertes Pest megyei pályázatok száma jól közelíthető az $n = 10$ és $p = 80000/500000 = 0{,}16$ paraméterű binomiális eloszlással. Ezek alapján

$$P(X = 4) = \binom{n}{4} \cdot p^4 \cdot (1-p)^6 = \binom{10}{4} \cdot 0{,}16^4 \cdot 0{,}84^6 \approx 0{,}048.$$

b) Az előző rész alapján

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = \binom{10}{0} \cdot 0{,}16^0 \cdot 0{,}84^{10} +$$
$$+ \binom{10}{1} \cdot 0{,}16^1 \cdot 0{,}84^9 + \binom{10}{2} \cdot 0{,}16^2 \cdot 0{,}84^8 + \binom{10}{3} \cdot 0{,}16^3 \cdot 0{,}84^7 \approx 0{,}939.$$

⇐11.2. feladat

11.3. feladat. Egy áruházban egy adott típusú termékből a raktáron lévő termékek 60 százaléka hazai gyártású. A termékek közül véletlenszerűen kiválasztunk hat darabot. Mennyi a valószínűsége, hogy pontosan három hazai gyártású termék van a kiválasztottak közt, ha

a) 2000 termék van raktáron;

b) negyven termék van raktáron?

Megoldás:

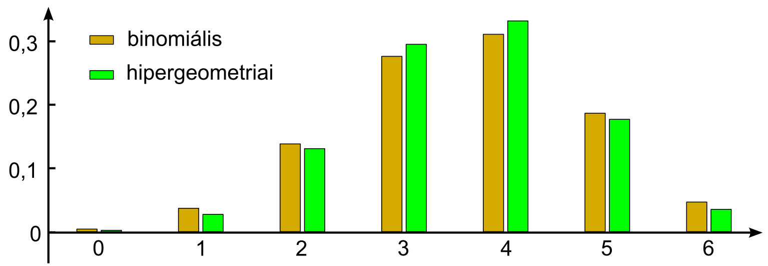
a) Jelölje X a kiválasztott hazai termékek számát. X ekkor hipergeometriai eloszlású, mivel azonban a kiválasztások száma (6), mind a hazai (1200), mind a külföld (800) gyártású termékek számához képest is kicsi, ezért X-et közelíthetjük egy $n = 6$ és $p = 1200/2000 = 0{,}6$ paraméterű binomiális eloszlással.

$$P(X = 3) = \binom{n}{3} \cdot p^3 \cdot (1-p)^3 = \binom{6}{3} \cdot 0{,}6^3 \cdot 0{,}4^3 \approx 0{,}276.$$

b) Mivel most a kiválasztások száma (6) a külföldi gyártású termékek számához (16) képest nem tekinthető kicsinek, ezért nem alkalmazhatjuk a binomiális közelítést. A hipergeometriai eloszlást használva:

$$P(X = 3) = \frac{\binom{24}{3} \cdot \binom{16}{3}}{\binom{40}{6}} \approx 0{,}295.$$

Vegyük észre, hogy ha most is binomiális eloszlással számoltunk volna, akkor ugyanazt az eredményt kaptuk volna, mint az előző részben (hiszen n és p nem változott volna). Vagyis binomiális eloszlással számolva közel $0{,}02$-dal kisebb értéket kaptunk volna a ténylegesnél!



11.2. ábra. *Az $m = 40$, $s = 24$, $n = 6$ paraméterű hipergeometria eloszlás, és az $n = 6$, $p = 0{,}6$ paraméterű binomiális eloszlás.*

⇐11.3. feladat

11.2. A binomiális eloszlás közelítése Poisson-eloszlással

Ahogy azt már korábban megtanultuk, ha egy kísérlet során egy adott esemény bekövetkezésének p a valószínűsége, akkor a kísérletet n-szer végrehajtva az adott esemény bekövetkezéseinek a száma egy n és p paraméterű binomiális eloszlással írható le. A következő tétel azt mutatja, hogy bizonyos esetekben a (két paramétertől függő) binomiális eloszlás közelíthető a nála egyszerűbb, egy paramétertől függő Poisson-eloszlással.

11.2. tétel: Legyen λ egy pozitív szám, legyen p_n $(n = 1,2,\ldots)$ 0 és 1 közötti számok olyan sorozata, amelynél

$$\lim_{n\to\infty} p_n \cdot n = \lambda\,.$$

Ekkor tetszőleges rögzített k esetén

$$\lim_{n\to\infty} \binom{n}{k} \cdot p_n^k \cdot (1-p_n)^{n-k} = \frac{\lambda^k}{k!} \cdot e^{-\lambda}\,.$$

Tehát az n és p_n paraméterű binomiális eloszlások sorozata a λ paraméterű Poisson-eloszláshoz közelít.

Hétköznapi nyelvre lefordítva a tétel úgy fogalmazható meg, hogy ha p elég kicsi és n nagy, akkor a binomiális eloszlást Poisson-eloszlással közelíthetjük, azaz

$$\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \approx \frac{(n\cdot p)^k}{k!} \cdot e^{-n\cdot p}\,.$$

11.4. feladat. Egy gazdasági konferenciára meghívják 600 cég, illetve vállalat vezetőjét. A vezetők egymástól függetlenül 0,002 valószínűséggel maradnak távol (pl. betegség miatt) a konferenciától. Mennyi a valószínűsége, hogy

a) pontosan 3 meghívott vezető nem megy el a konferenciára;

b) legalább 2 meghívott vezető nem vesz részt a konferencián;

c) kevesebb mint 4 meghívott vezető nem vesz részt a konferencián?

Megoldás:

a) Jelölje X azoknak a meghívott vezetőknek a számát, akik nem mennek el a konferenciára. Mivel a vezetők részvétele a konferencián egymástól független, ezért X binomiális eloszlású, $n = 600$ és $p = 0{,}002$ paraméterekkel. Mivel n nagy, p pedig kicsi, ezért X közelíthető egy $\lambda = n \cdot p = 600 \cdot 0{,}002 = 1{,}2$ paraméterű Poisson eloszlással.

$$P(X = 3) = \frac{\lambda^3}{3!} \cdot e^{-\lambda} = \frac{1{,}2^3}{3!} \cdot e^{-1{,}2} \approx 0{,}087.$$

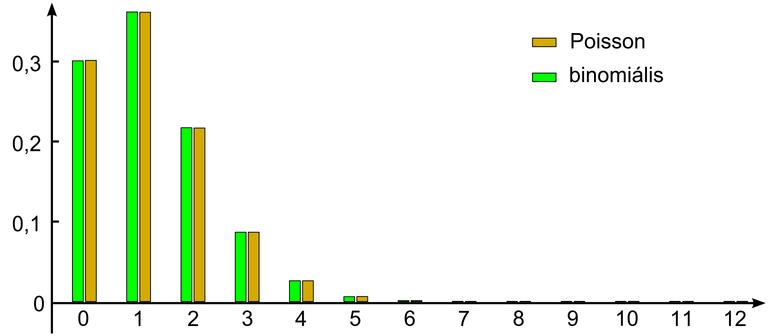
b) $P(2 \leq X) = 1 - [P(X = 0) + P(X = 1)] = 1 - \left[\frac{1{,}2^0}{0!} \cdot e^{-1{,}2} + \frac{1{,}2^1}{1!} \cdot e^{-1{,}2}\right] \approx 0{,}337.$

c)

$$P(X < 4) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = \frac{1{,}2^0}{0!} \cdot e^{-1{,}2} + \frac{1{,}2^1}{1!} \cdot e^{-1{,}2} +$$
$$+ \frac{1{,}2^2}{2!} \cdot e^{-1{,}2} + \frac{1{,}2^3}{3!} \cdot e^{-1{,}2} \approx 0{,}966.$$

⇐11.4. feladat

11.5. feladat. Gépelés során egy tapasztalt titkárnő minden egyes betű lenyomásakor a többi lenyomástól függetlenül 0,0005 valószínűséggel nyom le rossz billentyűt a klaviatúrán. Mennyi a valószínűsége, hogy egy 5000 betűt tartalmazó dokumentum gépelése során



11.3. ábra. *Az* $n = 600$, $p = 0{,}002$ *paraméterű binomiális eloszlás, és a* $\lambda = 1{,}2$ *paraméterű Poisson-eloszlás (11.4. feladat).*

a) pontosan 6 hibát vét;

b) legfeljebb 2 esetben üt le rossz billentyűt;

c) legalább kettő, de legfeljebb öt hibát követ el?

Megoldás:

a) Jelölje X a hibás leütések számát. A hibák egymástól függetlenek, ezért X binomiális eloszlású, $n = 5000$ és $p = 0{,}0005$ paraméterekkel. Mivel n nagy, p pedig kicsi, ezért X közelíthető egy $\lambda = n \cdot p = 5000 \cdot 0{,}0005 = 2{,}5$ paraméterű Poisson-eloszlással.

$$P(X = 6) = \frac{\lambda^6}{6!} \cdot e^{-\lambda} = \frac{2{,}5^6}{6!} \cdot e^{-2{,}5} \approx 0{,}028.$$

b)

$$P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{2{,}5^0}{0!} \cdot e^{-2{,}5} + \frac{2{,}5^1}{1!} \cdot e^{-2{,}5} + \frac{2{,}5^2}{2!} \cdot e^{-2{,}5} \approx 0{,}544.$$

c)

$$P(2 \leq X \leq 5) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) = \frac{2{,}5^2}{2!} \cdot e^{-2{,}5} + \frac{2{,}5^3}{3!} \cdot e^{-2{,}5} + $$
$$+ \frac{2{,}5^4}{4!} \cdot e^{-2{,}5} + \frac{2{,}5^5}{5!} \cdot e^{-2{,}5} \approx 0{,}671.$$

⇐11.5. feladat

Önellenőrzés

1. Fogalmazza meg saját szavaival, hogy a hipergeometriai eloszlás mikor közelíthető binomiális eloszlással!
2. Fogalmazza meg saját szavaival, hogy a binomiális eloszlás mikor közelíthető Poisson-eloszlással!
3. Egy utcai futóversenyen több mint 20 ezren indultak. Az indulók 63 százaléka volt férfi. A verseny végén az indulók közt kisorsolnak 8 egyforma szabadidő ruhát. Mennyi a valószínűsége, hogy pontosan 3 nő nyer a sorsoláson?

   ☐ 0,176 ☐ 0,358 ☐ 0,097 ☐ 0,282
4. Véletlenszerűen kiválasztunk 12 darab különböző, egymilliónál kisebb pozitív egész számot. Mennyi a valószínűsége, hogy négynél több, de nyolcnál kevesebb páros szám lesz köztük?

   ☐ 0,226 ☐ 0,612 ☐ 0,724 ☐ 0,564
5. Egy berendezés 3000 alkatrészből áll. Mindegyik alkatrész a többitől függetlenül 0,001 valószínűséggel romlik el fél év alatt. Mennyi a valószínűsége, hogy fél év alatt pontosan 5 alkatrész romlik el?

   ☐ 0,101 ☐ 0,187 ☐ 0,216 ☐ 0,158
6. Egy sportlövő egy lövés során 0,001 valószínűséggel lő a tábla mellé. Mennyi a valószínűsége, hogy ezer lövésből legalább háromszor lő a tábla mellé?

   ☐ 0,080 ☐ 0,061 ☐ 0,184 ☐ 0,218

15. LECKE

Modulzáró feladatok 3.

12. Modulzáró feladatok

Start. Az alábbi kérdések megválaszolására 1 órája van:

1. Feldobunk egy dobókockát tízszer egymás után. Mennyi a valószínűsége, hogy legfeljebb két alkalommal dobunk hármasnál kisebbet?

   ☐ 0,087 ☐ 0,195 ☐ 0,238 ☐ 0,299

2. Egy ötven fős évfolyamon 20 hallgatónak van már meg a nyelvvizsgája. Véletlenszerűen kiválasztjuk az évfolyam 4 hallgatóját. Mennyi a valószínűsége, hogy közülük legalább három már rendelkezik nyelvvizsgával?

   ☐ 0,149 ☐ 0,170 ☐ 0,359 ☐ 0830,

3. Egy focista 0,95 valószínűséggel lövi be a büntetőt. Az edzésen a büntetők gyakorlása során mennyi a valószínűsége, hogy a tizedik rúgásnál ront először?

   ☐ 0,382 ☐ 0,630 ☐ 0,599 ☐ 0,032

4. Az előző feladatnál mennyi a valószínűsége, hogy a harmadik rontásra a huszadik lövés után kerül sor?

   ☐ 0,794 ☐ 0,825 ☐ 0,867 ☐ 0,925

5. Az X valószínűségi változó Poisson-eloszlású, várható értéke pedig 2. Mennyi a valószínűsége, hogy X háromnál nagyobb értéket vesz fel?

   ☐ 0,677 ☐ 0,143 ☐ 0,581 ☐ 0,738

6. Egy berendezés várható élettartama 6 év; az élettartam exponenciális eloszlásúnak tekinthető. Mennyi a valószínűsége, hogy a berendezés legalább 5, de legfeljebb 7 évig működik?

☐ $0{,}123$ ☐ $0{,}311$ ☐ $0{,}432$ ☐ $0{,}491$

7. Az X valószínűségi változó egyenletes eloszlású a $[6; 14]$ intervallumon. Mennyi a $P(12 < X < 16)$ valószínűség értéke? ☐

8. Egy gép által gyártott alkatrész hossza normális eloszlású valószínűségi változónak tekinthető, melynek várható értéke 30 cm, szórása $0{,}4$ cm. Mennyi a valószínűsége. hogy egy alkatrész hossza legalább $29{,}2$ cm?

☐ $0{,}9744$ ☐ $0{,}0228$ ☐ $0{,}9772$ ☐ $0{,}9861$

9. Egy könnyűzenei koncerten több mint 40 ezren vettek részt; a résztvevők 60 százaléka budapesti volt. A koncert után a rendőrök a résztvevők közül véletlenszerűen választva igazoltatnak 5 embert. Mennyi a valószínűsége, hogy pontosan két budapesti van közöttük?

☐ $0{,}230$ ☐ $0{,}267$ ☐ $0{,}305$ ☐ $0{,}400$

Stop.

Pontok:

IV. MODUL

Valószínűségi változók viszonyának jellemzése

16. LECKE

Diszkrét valószínűségi változók függvényének leírása

13. Valószínűségi változó függvényének eloszlása

Előfordulhatnak olyan esetek, amikor egy véletlen eseményhez tartozó olyan X valószínűségi változót szeretnénk vizsgálni, mely egy másik, Y valószínűségi változó segítségével adható meg. Ekkor általában az Y eloszlása adott, vagy könnyen meghatározható. Az Y valószínűségi változó jellemzői segítségével az X eloszlása is meghatározható. Ebben a részben választ kaphatunk arra, hogyan.

Írjuk fel X-et Y függvényeként: $X = h(Y)$. Ez azt jelenti, hogy amikor az Y az y_0 értéket veszi fel, akkor az X értéke $h(y_0)$ lesz. Az X valószínűségi változót Y függvényének, vagy transzformáltjának mondjuk.

13.1. definíció: **Valószínűségi változó transzformáltja**
Legyenek X és Y valószínűségi változók. Tegyük fel, hogy létezik egy olyan $h : \mathbb{R} \to \mathbb{R}$ függvény, melyre $Y = y_0$ esetén $X = h(y_0)$ bármilyen y_0 esetén. Ekkor X-et az Y valószínűségi változó transzformáltjának hívjuk, s ezt az összefüggést az $X = h(Y)$ jelöléssel írjuk fel.

Figyeljük meg, hogy a $h(x)$ függvényt elég azoknál az értékeknél definiálni, melyeket a transzformálandó valószínűségi változó (a definícióban Y) felvesz. Nézzünk két példát arra, miképp jelenik meg egy valószínűségi változó függvénye:

1. A lottósorsolásnál a köznyelvben gyakran beszélünk arról, mennyi a valószínűsége annak, hogy kettesünk, hármasunk, netán ötösünk lesz. Természetes módon adódik az az Y valószínűségi változó, mely megadja, hogy egy adott szelvénnyel hány találatunk lesz a következő húzásnál. A nyeremény nagysága azonban jobban érdekelhet minket. Egyes helyeken előre kiadják, hogy hány találattal mennyi pénzt lehet nyerni: ezzel kaphatjuk meg a $h(x)$ transzformációs függvényt. Így az $X = h(Y)$ valószínűségi változó azt fogja megadni, mennyit fogunk nyerni az adott szelvénnyel a következő húzásnál.

Aktivitás: Keressen az internet segítségével olyan szerencsejátékot, ahol előre tudhatjuk, hogy adott tét mellett mekkora nyereményre számíthatunk!

2. Egy hosszú európai autós út alkalmával jó eséllyel ki fog fogyni a benzin vagy a gázolaj a járműből. Hogy hol, előre nem tudni, hiszen a fogyasztás sok tényezőtől függ. Jelölje az Y valószínűségi változó azt, hogy a tankolásra figyelmeztető jelző hány kilométer után gyullad ki. Az viszont jobban érdekel minket, hogy ezután hány kilométer múlva találjuk az első benzinkutat: ezt a távolságot jelölje az X valószínűségi változó, mely egyértelmű, hogy az Y-tól függ. Amennyiben biztosra tudjuk a benzinkutak helyzetét, azok éjjel-nappal nyitva tartanak, és útlezárással sem kell számolnunk, akkor az X és Y közötti összefüggés meghatározott: utunk minden egyes pontjához meg tudjuk adni, pontosan hány kilométer múlva lesz a következő benzinkút. Ez lesz a $h(x)$ függvény. Ilyenkor az X és az Y közötti összefüggés meghatározott, determinisztikus. Előfordulhat, hogy előre nem ismert útlezárás, a térkép bizonytalansága, illetve más véletlen tényezők miatt (betegség miatt zárva) az X és Y közötti összefüggést nem tudjuk függvényszerűen leírni. Erre az esetre a későbbiekben más eszközöket fogunk használni.

Aktivitás: Az M1 autópályán utazunk Budapestről Hegyeshalomba. Rajzolja fel erre az esetre a példában szereplő $h(x)$ függvényt!

13.1. Diszkrét valószínűségi változó függvényének eloszlása

Diszkrét valószínűségi változó transzformálása esetén az új valószínűségi változó is diszkrét lesz, hisz ha Y az $y_1, y_2, \dots$ értékeket veheti fel, akkor az $X = h(Y)$ értékei $h(y_1), h(y_2), \dots$, ami számosságra nem lehet több, mint az Y által felvett értékek számossága. Kevesebb viszont lehet. Ilyenkor az eredeti y_i értékek valószínűségei összeadódnak:

13.1. tétel: Legyenek az Y diszkrét eloszlású valószínűségi változó lehetséges értékei $y_1, y_2, \dots$ és ezek bekövetkezési valószínűségei pedig $p_1, p_2, \dots$. Az $X = h(Y)$ valószínűségi változó lehetséges értékei ekkor az $x_1 = h(y_1), x_2 = h(y_2), \dots$ számok, melyek között megegyezők is lehetnek. Ezek bekövetkezési valószínűségei a

$$P(X = x_k) = \sum_{h(y_i)=x_k} P(Y = y_i) = \sum_{h(y_i)=x_k} p_i \,.$$

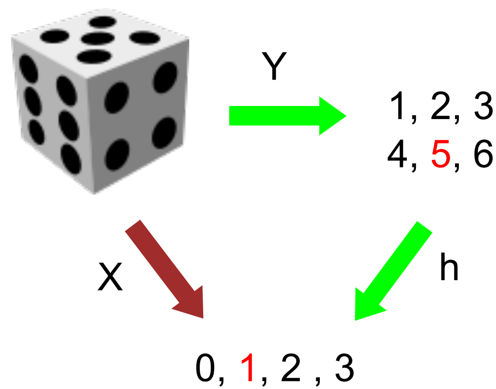
értékek, ahol az összegzés mindazon y_i-re vonatkozik, amelyre $h(y_i) = x_k$ fennáll.

Tehát az X transzformált egy adott x_k értékének valószínűségét úgy kaphatjuk meg, hogy összeadjuk Y azon y_i értékeinek valószínűségét, melyekre $x_k = h(y_i)$. Megjegyzendő, hogy amennyiben a $h(x)$ függvény szigorúan monoton, akkor az X valószínűségi változó $x_k = h(y_k)$, $(k = 1,2,\ldots)$ értékeihez tartozó valószínűség eloszlás megegyezik az Y valószínűségi változó eloszlásával: a konkrét valószínűség értékek megegyeznek, bár a lehetséges értékek eltérhetnek.

13.1. feladat. Dobókockával dobunk. Legyen X a dobott szám néggyel vett maradéka.

a) Határozzuk meg X eloszlását!

b) Mi a valószínűsége annak, hogy X kettőnél kevesebb?



13.1. ábra. *Az ötös dobáshoz, mint elemi eseményhez az $Y = 5$ dobott érték tartozik, melynek 1 a maradéka néggyel osztva.*

Megoldás:

a) Az X eloszlásának meghatározásához gondoljuk végig a következőt. Ha a kérdezett valószínűségi változó a dobókocka által dobott szám lenne, akkor annak eloszlását könnyen meg tudnánk határozni. Írjuk fel X-et valamilyen transzformációs függvény segítségével: Legyen az Y valószínűségi változó a kockával dobott szám értéke. Ekkor $X = Y \bmod 4$ és $h(x) = x \bmod 4$ a transzformáció függvénye. X, majd Y lehetséges értékeit felsorolva az alábbi táblázat alapján számolhatjuk X értékeinek valószínűségét:

a maradék lehetséges értékei (x_k)	a dobás lehetséges értékei (y_i)	a dobások valószínűsége (p_{y_i})	a maradékok valószínűsége (p_{x_k})
0	4	$1/6$	$1/6$
1	1 és 5	$1/6+1/6$	$1/3$
2	2 és 6	$1/6+1/6$	$1/3$
3	3	$1/6$	$1/6$

Így X eloszlását az alábbi formában adhatjuk meg:

$$X : \begin{cases} 0 & 1 & 2 & 3 \\ 1/6 & 1/3 & 1/3 & 1/6 \end{cases} .$$

Ezzel az X valószínűségi változó eloszlását meghatároztuk, azaz minden lehetséges értékére megállapítottuk, mekkora valószínűséggel veszi fel azt.

b) A valószínűség kiszámításához figyeljük meg, hogy míg Y az $1, 2, 3, 4, 5, 6$, addig X csak a $0, 1, 2, 3$ értékeket veheti fel. Ha azt szeretnénk, hogy X 2-nél kisebb legyen, akkor értéke csak 0 és 1 lehet. Így annak a valószínűsége, hogy $X < 2$, megegyezik azon valószínűségek összegével, melyekkel a nullát és az egyet felveszi:

$$P(X < 2) = P(X = 0) + P(X = 1) = 1/6 + 1/3 = 1/2 .$$

Tehát annak a valószínűsége, hogy X értéke kettőnél kisebb lesz, épp $1/2$. Ha megfigyeljük, X eloszlása szimmetrikus, és a feltételt épp az eloszlás egyik fele teljesíti.

⇐13.1. feladat

13-1. önálló feladat: Oldja meg az előző feladatot úgy is, hogy nem használunk transzformációt!

13.2. feladat.

Két dobókockával dobunk egyszerre. Legyen X a dobott számok összegének hattal vett maradéka.

a) Adjuk meg X eloszlását.

b) Mekkora valószínűséggel lesz X értéke páratlan?

Megoldás:

a) Az X eloszlását közvetlenül kiszámolni nem egyszerű. A dobott számok összegének eloszlásának felírása azonban már megoldható. Vegyünk egy új valószínűségi változót, Y-t. Legyen Y a kockákkal dobott számok összege, így $X = Y \bmod 6$. Ha meghatározzuk Y eloszlását, X eloszlását megadni már egyszerűbb lesz. Mivel a kockákon az értékek 1 és 6 között vannak, így Y lehetséges értékei a 2 és 12 közötti egész számok.

Az Y valószínűségi változó $P(Y = k)$ eloszlását úgy határozzuk meg, hogy mindegyik k értékhez kiszámoljuk, hányféleképpen lehet két kockadobás összegeként felírni. Adott k esetén ez lesz a jó esetek száma, míg az összes lehetséges dobás száma $6 \cdot 6 = 36$. Figyeljünk arra, hogy ilyenkor a kockadobások sorrendje számít, hiszen ekkor lesz klasszikus valószínűségi mezőnk!

$$2 = 1 + 1$$
$$3 = 1 + 2 = 2 + 1$$
$$4 = 1 + 3 = 2 + 2 = 3 + 1$$
$$5 = 1 + 4 = 2 + 3 = 3 + 2 = 4 + 1$$
$$6 = 1 + 5 = 2 + 4 = 3 + 3 = 4 + 2 = 5 + 1$$

$$\begin{aligned}
7 &= 1+6 = 2+5 = 3+4 = 4+3 = 5+2 = 6+1\\
8 &= 2+6 = 3+5 = 4+4 = 5+3 = 6+2\\
9 &= 3+6 = 4+5 = 5+4 = 6+3\\
10 &= 4+6 = 5+5 = 6+4\\
11 &= 5+6 = 6+5\\
12 &= 6+6
\end{aligned}$$

Minden egyes sorban megkapjuk, hányféleképpen tudjuk a sor eleji számot felírni két kockadobás összegeként. Mindegyik összeg $1/36$ valószínűséggel jön elő. Így Y eloszlását az alábbiak szerint tudjuk felírni:

$$Y : \begin{cases} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 1/36 & 2/36 & 3/36 & 4/36 & 5/36 & 6/36 & 5/36 & 4/36 & 3/36 & 2/36 & 1/36 \end{cases}$$

Az Y eloszlása segítségével már ki tudjuk számolni X eloszlását. Mivel X értékei hattal való osztás utáni maradékok, így X 0-tól 5-ig vehet fel egész értékeket:

a maradék lehetséges értékei (x_k)	a dobás lehetséges értékei (y_i)	a dobások valószínűsége (p_{y_i})	a maradékok valószínűsége (p_{x_k})
0	6 és 12	$5/36 + 1/36$	$1/6$
1	7	$6/36$	$1/6$
2	8 és 2	$1/36 + 5/36$	$1/6$
3	9 és 3	$2/36 + 4/36$	$1/6$
4	10 és 4	$3/36 + 3/36$	$1/6$
5	11 és 5	$4/36 + 2/36$	$1/6$

Ezek alapján az X valószínűségi változó eloszlását felírhatjuk:

$$X : \begin{cases} 0 & 1 & 2 & 3 & 4 & 5 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{cases} .$$

Ezzel az X valószínűségi változó eloszlását meghatároztuk, azaz minden lehetséges értékére megállapítottuk, mekkora valószínűséggel veszi fel azt.

b) Most annak a valószínűségét határozzuk meg, hogy X értéke páratlan lesz! Láthatjuk, hogy az X valószínűségi változó a $0, 1, 2, 3, 4, 5$ értékeket veheti fel. Közülük az 1, a 3 és az 5 páratlan. Így

$$P(X \text{ páratlan}) = P(X = 1) + P(X = 3) + P(X = 5) = 1/6 + 1/6 + 1/6 = 1/2 .$$

Tehát annak a valószínűsége, hogy X páratlan lesz, épp 50%. Ez nem meglepő, hiszen mindegyik értéket azonos valószínűséggel veszi fel, és a lehetséges értékeknek épp a fele páratlan.

⇐13.2. feladat

13-2. önálló feladat: Állítsa elő az előző feladat eloszlását úgy is, hogy csak egy kockadobás értékét használjuk fel! Írjuk fel a $h(x)$ függvényt!

Aktivitás:
Keressen olyan szerencsejátékot, melyben a nyereményünk két dobókocka együttes értékétől függ!

13.3. feladat. Nyolc cetlire az alábbi számokat írjuk sorba: $\pi/4$, $\pi/4$, $\pi/3$, $-\pi/3$, π, π, $-\pi$, $-\pi$. A cetliket egyenként golyókba rakjuk, majd a golyókat urnába rakva közülük véletlenszerűen húzunk. Megnézzük milyen szám van a cetlin, majd kiszámoljuk a koszinuszát. Legyen X a kapott érték!

a) Adjuk meg X eloszlását!

b) Mekkora valószínűséggel lesz X értéke nulla?

Megoldás:

a) X eloszlását ismét két lépésben határozzuk meg. Először annak az Y valószínűségi változónak az eloszlását írjuk fel, mely azt mondja meg, hogy a golyókból milyen számot húztunk. Ezek után kiszámolhatjuk X eloszlását is. Figyeljük meg Y lehetséges értékeit. Ebből nyolcnál kevesebb van, hiszen több papírra is ugyanazt írtuk.
Y öt lehetséges értéke: $-\pi$, $-\pi/3$, $\pi/4$, $\pi/3$ és π. Az urnában eltérő arányban fordulnak elő, rendre 2, 1, 2, 1 és 2 darab golyóban található az adott szám. Annak a valószínűsége, hogy Y egy konkrét értéket vesz fel, klasszikus valószínűségi mezőként írható fel, ahol az összes eset száma 8, a jó eseteké pedig rendre az előző értékek.
A fentiek alapján Y eloszlása felírható:

$$Y : \begin{cases} -\pi & -\pi/3 & \pi/4 & \pi/3 & \pi \\ 2/8 & 1/8 & 2/8 & 1/8 & 2/8 \end{cases} .$$

Így megkaptuk, hogy az egyes értékeket milyen valószínűséggel húzzuk az urnából. Az X valószínűségi változót felírhatjuk úgy, mint $X = \cos Y$. Lehetséges értékei meghatározásához ki kell számolnunk Y lehetséges értékeinek koszinuszait:

$$\cos(-\pi) = \cos(\pi) = -1$$
$$\cos(\pi/4) = \sqrt{2}/2$$
$$\cos(-\pi/3) = \cos(\pi/3) = 1/2$$

Láthatjuk, hogy X lehetséges értékei -1, $\sqrt{2}/2$ és $1/2$ lehetnek. Határozzuk meg, hogy az egyes értékeket milyen valószínűséggel veszi fel! Ehhez szükségünk lesz arra is, hogy az eredeti értékeket milyen valószínűséggel veszi fel Y:

$$\begin{aligned} P(X=-1) &= P(Y=-\pi)+P(Y=\pi)=2/8+2/8=1/2,\\ P(X=\sqrt{2}/2) &= P(Y=\pi/4)=2/8=1/4,\\ P(X=1/2) &= P(Y=-\pi/3)+P(Y=\pi/3)=1/8+1/8=1/4\,. \end{aligned}$$

Az X valószínűségi változó mindegyik lehetséges értékének meghatároztuk a valószínűségét. Így felírhatjuk az eloszlását:

$$X:\begin{cases} -1 & \sqrt{2}/2 & 1/2 \\ 1/2 & 1/4 & 1/4 \end{cases}.$$

b) Mekkora lehet annak a valószínűsége, hogy X értéke 0? Figyeljük meg, hogy X nem is veszi fel a 0 értéket! Ettől függetlenül ki tudjuk számolni a valószínűséget: ez egy lehetetlen esemény, melynek valószínűsége 0 lesz.

⇐13.3. feladat

Leckezáró önellenőrző kérdéssor

1. Válasszunk egy egész számot a $[-2;3]$ intervallumból. Legyen X a választott szám négyzete! Ekkor X lehetséges értékei

☐ $0, 1, 4, 9$ ☐ $0 \dots 9$ ☐ $1, 4, 9$ ☐ egyik sem

2. Legyen Y diszkrét eloszlású valószínűségi változó, és X pedig Y valamilyen tetszőleges $h(x)$ szerinti transzformáltja. Milyen eloszlású lesz X?

☐ folytonos ☐ diszkrét ☐ ismeretlen ☐ egyik sem

3. Nyolcoldalú dobókockával dobunk. Legyen X a dobott szám 5-tel vett maradéka. Mi a valószínűsége annak, hogy $X < 2$?

☐ 0 ☐ $3/8$ ☐ $3/4$ ☐ $1/2$

4. Két dobókockával dobunk. Legyen Y a dobott számok összege! Legyen $X = h(Y)$ az Y transzformáltja úgy, hogy $h(x) = x/12$. Mi a valószínűsége annak, hogy X értéke egész lesz?

☐ 0 ☐ $5/36$ ☐ $1/36$ ☐ 1

5. Legyen az Y valószínűségi változó eloszlása a következő: $Y : \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 2/3 & 1/6 & 1/6 \end{array} \right.$. Legyen $X = |Y|$. Ekkor X eloszlása

☐ $X : \left\{ \begin{array}{ccc} 1 & 0 & 1 \\ 2/3 & 1/6 & 1/6 \end{array} \right.$ ☐ $X : \left\{ \begin{array}{cc} 0 & 1 \\ 1/2 & 1/2 \end{array} \right.$ ☐ $X : \left\{ \begin{array}{cc} 0 & 1 \\ 1/6 & 5/6 \end{array} \right.$ ☐ nem határozható meg egyértelműen.

17. LECKE

Folytonos valószínűségi változók függvényének leírása

13.2. Folytonos valószínűségi változó függvényének eloszlása

Ebben a fejezetben azt vizsgáljuk, milyen módon viselkednek folytonos valószínűségi változók a transzformáció hatására. Ennek érdekében vegyünk egy Y folytonos eloszlású valószínűségi változót, melynek eloszlásfüggvénye $F_Y(x)$. Tegyük fel, hogy az X valószínűségi változó felírható az Y függvényeként, azaz $X = h(Y)$. Ahhoz, hogy X-et elemezni tudjuk, szükségünk lesz az eloszlásfüggvényére és a sűrűségfüggvényére. Ezeket az Y valószínűségi változó eloszlás- és sűrűségfüggvényének ismeretében meg tudjuk határozni.

A transzformációval kapcsolatban érdemes az alábbi módon gondolkodni: az X valószínűségi változóval kapcsolatos eseményt ki tudjuk fejezni Y-nal kapcsolatos eseményként. Amennyiben szigorúan monoton transzformációs függvényt használunk, az új valószínűségi változónk is folytonos lesz. A sűrűségfüggvényét az alábbi két tétel segítségével kaphatjuk meg:

13.2. tétel:

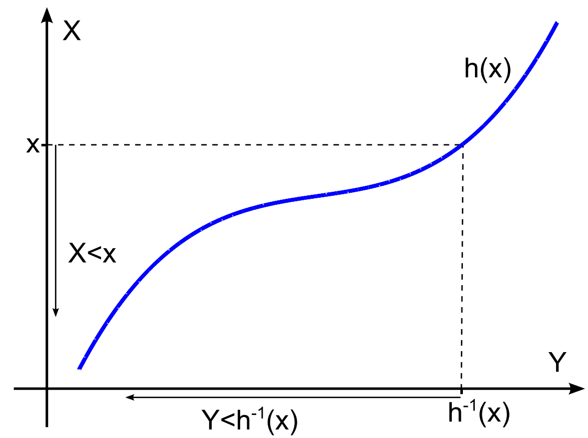
Legyen Y folytonos eloszlású valószínűségi változó $F_Y(x)$ eloszlás-, és $f_Y(x)$ sűrűségfüggvénnyel. Legyen h egy szigorúan monoton növő, differenciálható függvény, és $X = h(Y)$ pedig Y transzformáltja. Ekkor X eloszlásfüggvénye:

$$F_X(x) = F_Y\left(h^{-1}(x)\right), \text{ ha } x \in D_{h^{-1}},$$

és sűrűségfüggvénye:

$$f_X(x) = f_Y\left(h^{-1}(x)\right) \cdot \left(h^{-1}(x)\right)', \text{ ha } x \in D_{h^{-1}}.$$

Itt h^{-1} a h függvény inverzét, $D_{h^{-1}}$ pedig h^{-1} értelmezési tartományát, azaz h értékkészletét jelöli.



13.2. ábra. *Amennyiben $h(x)$ szigorúan monoton nő, X akkor lesz kisebb, mint x, ha Y kisebb mint $h^{-1}(x)$.*

Bizonyítás: Az $\{X < x\}$ esemény (ami tulajdonképpen a $\{h(Y) < x\}$ esemény) megegyezik az $\{X < h^{-1}(x)\}$ eseménnyel, ha x benne van h értékkészletében. Így X eloszlásfüggvénye:

$$F_X(x) = P(X < x) = P(h(Y) < x) = P(Y < h^{-1}(x)) = F_Y\left(h^{-1}(x)\right).$$

Ebből X sűrűségfüggvénye az összetett függvény deriválásával megkapható:

$$f_X(x) = \left[F_Y\left(h^{-1}(x)\right)\right]' = f_Y\left(h^{-1}(x)\right) \cdot \left(h^{-1}(x)\right)'.$$

□

13.4. feladat. Legyen Y egyenletes eloszlású az $[1;2]$ intervallumon. Legyen ezen kívül $X = Y^2$. Adjuk meg X

a) eloszlás- és

b) sűrűségfüggvényét!

Megoldás:

a) Először az eloszlásfüggvényt számoljuk ki. Ismerjük fel, hogy a feladatban szereplő transzformáció függvénye a $h(x) = x^2$, amely a megadott intervallumon szigorúan monoton nő és differenciálható. A tétel alapján a feladat megoldásához az inverzét kell kiszámolni. A $h(x)$ függvény inverze a $h^{-1}(x) = \sqrt{x}$ lesz. Ismerjük Y eloszlásfüggvényét is:

$$F_Y(x) = \begin{cases} 0, & \text{ha } x \le 1; \\ x-1, & \text{ha } 1 < x \le 2; \\ 1, & \text{ha } x > 2; \end{cases}.$$

Így a tétel segítségével meghatározhatjuk X eloszlásfüggvényét:

$$F_X(x) = F_Y(h^{-1}(x)) = F_Y(\sqrt{x}) = \begin{cases} 0, & \text{ha } x \le 1; \\ \sqrt{x}-1, & \text{ha } 1 < x \le 4; \\ 1, & \text{ha } x > 4. \end{cases}$$

Így megkaptuk X eloszlásfüggvényét. Figyeljük meg, hogy ugyan az eloszlásfüggvény alakja nagyon hasonlít az egyenletes eloszláséhoz, X eloszlása *nem* egyenletes lesz.

b) A sűrűségfüggvény kiszámításához a transzformációs függvény inverzének deriváltjára is szükségünk lesz:

$$\left[h^{-1}(x)\right]' = \left[\sqrt{x}\right]' = \frac{1}{2\sqrt{x}}.$$

Az egyenletes eloszlás sűrűségfüggvényét pedig felírhatjuk Y-ra:

$$f_Y(x) = \begin{cases} 1, & \text{ha } 1 < x < 2; \\ 0 & \text{egyébként}. \end{cases}$$

Alkalmazzuk a tételben szereplő sűrűségfüggvény számítást:

$$f_X(x) = f_Y\left(h^{-1}(x)\right) \cdot \left(h^{-1}(x)\right)' = f_Y\left(\sqrt{x}\right) \cdot \frac{1}{2\sqrt{x}}$$

Ebből az X sűrűségfüggvénye:

$$f_X(x) = \begin{cases} \dfrac{1}{2\sqrt{x}}, & \text{ha } 1 < x < 4; \\ 0 & \text{egyébként}. \end{cases}$$

Vegyük észre, hogy a sűrűségfüggvény értéke a szorzás miatt változott, a határok pedig amiatt, hogy az eredeti függvényt a $\sqrt{x}$ helyen néztük. Ismét láthatjuk, hogy X eloszlása nem egyenletes, akkor sem, ha sűrűségfüggvénye alakilag hasonlónak látszik. Lényeges különbség, hogy az egyenletes eloszlás sűrűségfüggvénye mindig lépcsős: az egyes szakaszokon belül felvett érték nem függhet x-től.

⇐13.4. feladat

13-3. önálló feladat: Számolja ki a fenti példa X valószínűségi változójának várható értékét, és hasonlítsa össze az Y valószínűségi változó várható értékének négyzetével!

13.3. tétel: Legyen Y folytonos eloszlású valószínűségi változó $F_Y(x)$ eloszlás-, és $f_Y(x)$ sűrűségfüggvénnyel. Legyen h egy szigorúan monoton csökkenő, differenciálható függvény, és $X = h(Y)$ pedig Y transzformáltja. Ekkor X eloszlásfüggvénye:

$$F_X(x) = 1 - F_Y\left(h^{-1}(x)\right), \text{ ha } x \in D_{h^{-1}},$$

és sűrűségfüggvénye:

$$f_X(x) = -f_Y\left(h^{-1}(x)\right) \cdot \left(h^{-1}(x)\right)', \text{ ha } x \in D_{h^{-1}}.$$

Itt h^{-1} a h függvény inverzét, $D_{h^{-1}}$ pedig az inverz értelmezési tartományát, azaz h értékkészletét jelöli.

Bizonyítás: Az $\{X < x\}$ esemény (ami most is a $\{h(Y) < x\}$ esemény) megegyezik az $\{Y > h^{-1}(x)\}$ eseménnyel, ha x benne van h értékkészletében. Így ebben az esetben X eloszlásfüggvénye:

$$F_X(x) = P(X < x) = P(h(Y) < x) = P(Y > h^{-1}(x)) = 1 - F_Y\left(h^{-1}(x)\right).$$

Ebből X sűrűségfüggvénye az összetett függvény deriválásával megkapható:

$$f_X(x) = \left[1 - F\left(h^{-1}(x)\right)\right]' = -f\left(h^{-1}(x)\right) \cdot \left(h^{-1}(x)\right)'.$$

$\square$

Szigorúan monoton növő, illetve csökkenő függvény inverzének deriváltja szigorúan pozitív, illetve negatív. Ebből adódóan a fenti két tételből kapott sűrűségfüggvény transzformáltja egybe foglalható:

$$f_X(x) = f\left(h^{-1}(x)\right) \cdot \left|\left(h^{-1}(x)\right)'\right|, \text{ ha } x \in D_{h^{-1}}.$$

Vegyük észre, hogy az abszolútértékjel épp azt teszi, amit a tételekben szereplő előjel: szigorúan monoton növő transzformációs függvénynél a pozitív előjelet helyben hagyja, szigorúan csökkenőnél pedig az ellentettjére, épp pozitívra változtatja.

Aktivitás: Keressen a való életben olyan dolgokat, melyeknek két különböző paraméterük között nem lineáris az összefüggés (pl. pizza átmérője, tömege, ára)!

A fenti tételek esetén sem szükséges, hogy a $h(x)$ függvény a teljes számegyenesen értelmezve legyen. Elég egy olyan intervallumon megadni, ahol a transzformálandó Y valószínűségi változó felveszi az értékeit. Ennek megfelelően a transzformált X eloszlás-, és sűrűségfüggvényét ki kell egészíteni. Lássunk erre egy példát:

13.5. feladat. Válasszunk a $[0;1]$ intervallumon egy valós számot. A választott számmal, mint sugárral rajzoljunk kört. Legyen X a kör területe!

a) Adjuk meg X eloszlás-, és sűrűségfüggvényét!

b) Mi a valószínűsége annak, hogy a kapott kör területe legalább $\pi/4$?

Megoldás:

a) Az eloszlásfüggvény meghatározásához először írjuk fel az X valószínűségi változót az intervallumon választott szám függvényeként. Ez utóbbi legyen az Y valószínűségi változó. Ekkor vegyük észre, hogy a kör területének képletét használva felírhatjuk X-et:

$$X = Y^2 \cdot \pi,$$

így a transzformáció függvénye a $h(x) = x^2 \cdot \pi$ lesz.
Amennyiben ezt a függvényt csak a $[0;1]$ intervallumon definiáljuk, ahol Y az értékeit felveheti, akkor szigorúan monoton növekvő lesz, így alkalmazhatjuk rá az előbb megismert tételt. Az eloszlásfüggvény meghatározásához szükségünk lesz a $h(x)$ függvény inverzére, ez az alábbi lesz:

$$h^{-1}(x) = \sqrt{\frac{x}{\pi}}.$$

Az Y eloszlása viszont egyenletes, így $F_Y = \frac{x-0}{1-0} = x$, ha $0 < x \leq 1$, azaz

$$F_Y(x) = \begin{cases} 0, & \text{ha } x \leq 0; \\ x, & \text{ha } 0 < x \leq 1; \\ 1, & \text{ha } x > 1\,. \end{cases}$$

Ebből X eloszlásfüggvénye:

$$F_X(x) = F_Y\left(\sqrt{\frac{x}{\pi}}\right) = \sqrt{\frac{x}{\pi}}, \quad \text{ha } 0 < \sqrt{\frac{x}{\pi}} \leq 1, \quad \text{azaz } 0 < x \leq \pi\,.$$

Mivel X értéke 0-nál kisebb és π-nél nagyobb értéket nem vehet fel, így a teljes eloszlásfüggvény az alábbiak szerint írható fel:

$$F_X(x) = \begin{cases} 0, & \text{ha } x \leq 0; \\ \sqrt{\frac{x}{\pi}}, & \text{ha } 0 < x \leq \pi; \\ 1, & \text{ha } x > \pi\,. \end{cases}$$

A sűrűségfüggvény számításához ki kell számolnunk h inverzének deriváltját:

$$\left(h^{-1}(x)\right)' = \left(\sqrt{\frac{x}{\pi}}\right)' = 1/\left(2\sqrt{x\pi}\right)$$

Mivel Y sűrűségfüggvénye $f_Y(x) = \begin{cases} 1, & \text{ha } 0 < x < 1; \\ 0 & \text{egyébként} \end{cases}$, az X sűrűségfüggvényét az alábbi alakban írhatjuk fel:

$$f_X(x) = f_Y\left(h^{-1}(x)\right) \cdot \left(h^{-1}(x)\right)' = f_Y\left(\sqrt{\frac{x}{\pi}}\right) \cdot \frac{1}{2\sqrt{x\pi}}\,.$$

Így felírhatjuk X sűrűségfüggvényét:

$$f_X(x) = \begin{cases} \frac{1}{2\sqrt{x\pi}}, & \text{ha } 0 < x < \pi; \\ 0 & \text{egyébként}\,. \end{cases}$$

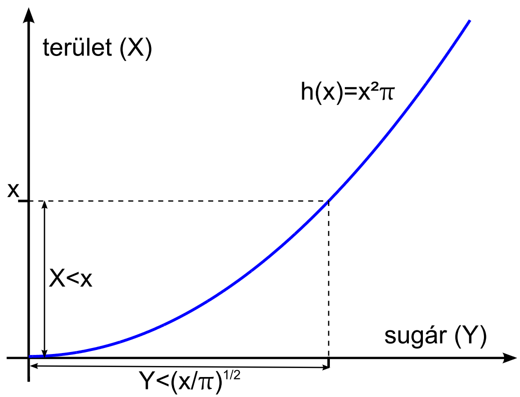
Ez az eredmény mind a tétel alkalmazásával, mind az $F_X(x)$ közvetlen deriválásával megkapható. Figyeljük meg, hogy a transzformációs függvény négyzetes: Y értékeit az intervallum elejéről és végéről is az intervallum hosszával arányos valószínűséggel választjuk. Az X esetén viszont megváltozik a helyzet: az intervallum elejéről lényegesen gyakrabban jönnek elő az értékek.

b) Most számoljuk ki annak a valószínűségét, hogy a kör területe legalább $\pi/4$. Ezt X eloszlásfüggvényének segítségével számíthatjuk a már jól ismert módon:

$$P(X \geq \pi/4) = 1 - F_X(\pi/4) = 1 - \sqrt{\frac{\pi/4}{\pi}} = 1 - 1/2 = 1/2\,.$$

Tehát annak a valószínűsége, hogy a kör területe legalább $\pi/4$, $1/2$ lesz. Vegyük észre, hogy X lehetséges értékei 0 és π között vannak, mégis az intervallum első negyedének értékeit ugyanakkora valószínűséggel veszi fel, mint a többit. Ez is azt mutatja, hogy X eloszlása nem lehet egyenletes.

⇐13.5. feladat



13.3. ábra. *Az ábrán a feladatban szereplő transzformációs függvény ábrázolását láthatjuk. A véletlenszerűen választott sugár 0 és 1 közötti értékeket vehet fel, ezáltal a szerkesztett kör területe 0 és π között fog alakulni. Ha a sugár kisebb, mint x, akkor a terület kisebb, mint $x^2\pi$, így ha a terület kisebb, mint x, akkor a sugár kisebb, mint* $\sqrt{x/\pi}$.

13.2. definíció: Lineáris transzformáció
Amennyiben a valószínűségi változó transzformálását végző függvény lineáris, azaz $h(x) = a \cdot x + b$ alakú, lineáris változótranszformációról beszélünk.

13.4. tétel: Legyen Y egy folytonos eloszlású, f_Y sűrűségfüggvényű valószínűségi változó. Ekkor az $X = a \cdot Y + b$ $(a \neq 0)$ valószínűségi változó sűrűségfüggvénye:

$$f_X(x) = \frac{1}{|a|} \cdot f_Y\left(\frac{x-b}{a}\right).$$

Bizonyítás: Tekintsük az előző tételt a $h(x) = a \cdot x + b$ függvénnyel. Ekkor $h(x)$ inverze $h^{-1}(x) = \dfrac{x-b}{a}$, ennek deriváltja pedig $\left(h^{-1}(x)\right)' = \dfrac{1}{a}$, így adódik a tétel állítása. $\square$

13.6. feladat. Lássuk be, hogy egy egyenletes eloszlású folytonos valószínűségi változó lineáris transzformáltja is egyenletes eloszlású!

Megoldás: Ahhoz, hogy erről bebizonyosodjunk, vegyünk egy tetszőleges egyenletes eloszlású valószínűségi változót, melyet jelöljünk Y-nal. Y az értékeit c és d között veszi fel. Figyeljünk a jelölésre, mert a c és a d nem szokványos, és nem összekeverendő a lineáris transzformáció két paraméterével a-val és b-vel. Tehát Y egyenletes eloszlású valószínűségi változó c és d között, X-et pedig az $X = a \cdot Y + b$ lineáris transzformációval írhatjuk fel. Ekkor a transzformációs függvény $h(x) = ax + b$ lesz.

Az eloszlásfüggvény meghatározásához fel kell írnunk a transzformációs függvény inverzét:

$$h^{-1}(x) = \frac{x-b}{a}.$$

Ekkor X eloszlásfüggvénye:

$$F_X(x) = F_Y\left(\frac{x-b}{a}\right) = \begin{cases} 0, & \text{ha} \quad \frac{x-b}{a} \le c; \\ \dfrac{\frac{x-b}{a} - c}{d-c}, & \text{ha} \quad c < \frac{x-b}{a} \le d; \\ 1, & \text{ha} \quad \frac{x-b}{a} > d; \end{cases} =$$

$$= \begin{cases} 0, & \text{ha} \quad x \le ac+b; \\ \dfrac{x-(ac+b)}{(ad+b)-(ac+b)}, & \text{ha} \quad ac+b < x \le ad+b; \\ 1, & \text{ha} \quad x > ad+b; \end{cases} \quad = \begin{cases} 0, & \text{ha} \quad x \le c'; \\ \dfrac{x-c'}{d'-c'}, & \text{ha} \quad c' < x \le d'; \\ 1, & \text{ha} \quad x > d'. \end{cases}$$

Vegyük észre, hogy X eloszlásfüggvényének alakja pontosan megegyezik az egyenletes eloszlás eloszlásfüggvényének alakjával! Lássuk meg azt is, hogy X értékeit $c' = ac+b$ és $d' = ad+b$ között fogja felvenni. Ezzel beláttuk, hogy X eloszlása egyenletes lesz, azaz bármilyen egyenletes eloszlású valószínűségi változó lineáris transzformáltja is egyenletes eloszlású lesz.

⇐13.6. feladat

Aktivitás: Gondolja végig, milyen tantárgyakból és milyen környezetben tanultak még lineáris transzformációról!

A lineáris transzformálást leggyakrabban akkor használjuk, amikor egy normális eloszlású valószínűségi változót standardizálunk. Az előző tétel alapján egyszerűen beláthatjuk, hogy a standardizálással kapott valószínűségi változó valóban standard normális eloszlású.

13.5. tétel: Legyen X normális eloszlású valószínűségi változó az m és σ paraméterekkel. Ekkor X standardizáltja, vagyis az $X^* = \dfrac{X-m}{\sigma}$ valószínűségi változó standard normális eloszlású.

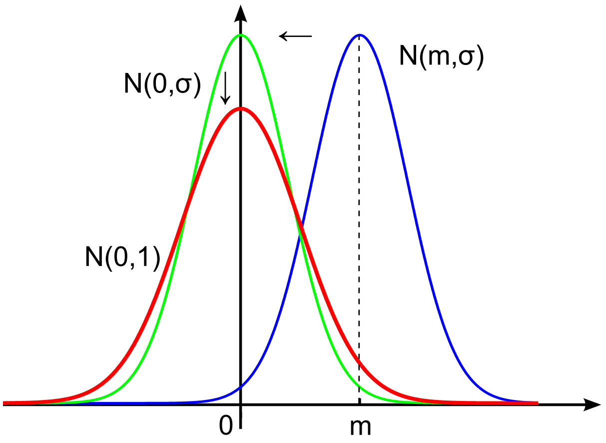
Bizonyítás: A bizonyításhoz elég azt megmutatni, hogy X^* sűrűségfüggvénye $\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$. Mivel X normális eloszlású, ezért a sűrűségfüggvénye

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Mivel $X^* = \frac{X-m}{\sigma} = \frac{1}{\sigma} \cdot X - \frac{m}{\sigma}$, ezért alkalmazhatjuk az előző tételt az $a = \frac{1}{\sigma}$ és $b = -\frac{m}{\sigma}$ választással. Így X^* sűrűségfüggvénye:

$$\frac{1}{|a|} \cdot f\left(\frac{x-b}{a}\right) = \sigma \cdot f\left(\frac{x - \frac{-m}{\sigma}}{\frac{1}{\sigma}}\right) = \sigma \cdot f(\sigma \cdot x + m) =$$
$$= \sigma \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{((\sigma \cdot x + m) - m)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}.$$

□



13.4. ábra. *A standardizálás során először a várható értéket állítjuk 0-ra, majd a szórást 1-re.*

13-4. önálló feladat: Lássa be az előző tételek segítségével, hogy ha X normális eloszlású valószínűségi változó m és σ paraméterekkel, akkor eloszlásfüggvénye

$$F_X(x) = \Phi\left(\frac{x-m}{\sigma}\right).$$

Leckezáró önellenőrző kérdéssor

1. Legyen Y exponenciális eloszlású, várható értéke 1, és $X = e^Y$. Ekkor az X valószínűségi változó $F_X(x)$ eloszlásfüggvénye:

☐ $\begin{cases} 1 - \frac{1}{e^x}, & \text{ha } x > 0; \\ 0 & \text{egy.} \end{cases}$ ☐ $\begin{cases} 1 - \ln x, & \text{ha } x > 1; \\ 0 & \text{egy.} \end{cases}$ ☐ $\begin{cases} 1 - e^{-x}, & \text{ha } x > 0; \\ 0 & \text{egy.} \end{cases}$ ☐ $\begin{cases} 1 - \frac{1}{x}, & \text{ha } x > 1; \\ 0 & \text{egy.} \end{cases}$

2. Legyen Y egyenletes eloszlású a $[0; \pi]$ intervallumon, és $X = \cos Y$. Használhatjuk-e a leckében tanultakat X sűrűségfüggvényének kiszámításához?

☐ Nem, mert a $\cos(x)$ nem monoton függvény.
☐ Igen, mert a $\cos(x)$ monoton részekre bontható.
☐ Nem, mert a $\cos(x)$ deriváltja a két végpontban 0.
☐ Igen, mert a $\cos(x)$ szigorúan monoton csökkenő ott, ahol az Y az értékeit felveszi.

3. Az alábbi transzformációk közül melyikre igaz, hogy az egyenletes eloszlású Y valószínűségi változóból egyenletes eloszlásút csinál?

☐ $h(x) = x^2 - 1$ ☐ $h(x) = 2 \cdot x - 1$ ☐ mindkettőre ☐ egyikre sem

4. Józsika egy valós számra gondol 0 és 100 között. Amint kitalálta, elosztja kettővel, kivon belőle ötöt, majd megszorozza hattal, lecseréli az előjelét, és lecsökkenti hússzal. Mi a valószínűsége annak, hogy Józsika negatív számot kapott?

☐ $2/15$ ☐ $29/30$ ☐ $6/17$ ☐ 0

5. Mennyi lesz a 13.5 feladatban a kérdezett valószínűség, ha a sugarat az $[1; 2]$ intervallumon választjuk?

☐ $1/4$ ☐ 1 ☐ nem változik ☐ egyik sem

18. LECKE

Valószínűségi változó függvényének várható értéke és szórása

13.3. Valószínűségi változó függvényének várható értéke és szórása

Az $X = h(Y)$ valószínűségi változó várható értékét kiszámíthatjuk diszkrét esetben az eloszlásából, folytonos esetben pedig a sűrűségfüggvényéből. Erre a megfelelő módszereket és képleteket az eddigi fejezetekben megismerhettük. Ha azonban csak X várható értéke érdekel minket, az viszont nem, hogy az eloszlása, eloszlásfüggvénye mi lesz, akkor jóval egyszerűbb számításokat is alkalmazhatunk. Külön vizsgáljuk a diszkrét és a folytonos valószínűségi változókat.

Az alábbi tétel segítségével láthatjuk, hogy az eredeti valószínűségi változó eloszlását ismerve meghatározható a transzformált várható értéke is.

13.6. tétel: Legyen Y diszkrét eloszlású valószínűségi változó, lehetséges értékei $y_1, y_2, \ldots$, a hozzájuk tartozó valószínűségek pedig $p_{y_1}, p_{y_2}, \ldots$. Legyen továbbá $X = h(Y)$. Ekkor

1. X várható értéke
$$E(X) = \sum_i h(y_i) \cdot p_{y_i}\,.$$
2. X^2 várható értéke
$$E(X^2) = \sum_i h^2(y_i) \cdot p_{y_i}\,.$$
3. X szórásnégyzete
$$D^2(X) = \left[\sum_i h^2(y_i) \cdot p_{y_i}\right] - \left[\sum_i h(y_i) \cdot p_{y_i}\right]^2 .$$

Bizonyítás:

1. Az X valószínűségi változó várható értéke:
$$E(X) = \sum_k x_k \cdot p_{x_k} \qquad \text{, ahol } p_{x_k} = \sum_{h(y_i)=x_k} p_{y_i}.$$

Ezzel

$$E(X)=\sum_k x_k\cdot p_{x_k}=\sum_k x_k\cdot\sum_{h(y_i)=x_k} p_{y_i}=\sum_k\sum_{h(y_i)=x_k} x_k\cdot p_{y_i}=\sum_k\sum_{h(y_i)=x_k} h(y_i)\cdot p_{y_i}=$$
$$=\sum_i h(y_i)\cdot p_{y_i}\,.$$

2. Az $X^2=h^2(Y)$ transzformációt használva, az előző tételből adódik.
3. Az X^2 várható értékéből és a szórásnégyzet definíciójából adódik.

□

Figyeljük meg a tételben bemutatott képleteket. Az eredeti y_i értékek helyett most a $h(y_i)$ értékek lesznek súlyozva aszerint, mekkora a valószínűsége y_i-nek. Láthatjuk, hogy több azonos $h(y_i)$ értékek esetén a súlyok összeadódnak.

13.7. feladat.

Egy hatoldalú dobókockával dobunk. Legyen X a dobott szám néggyel vett maradéka. Adjuk meg X

a) várható értékét

b) és szórását!

Megoldás:

a) A feladat megoldásához érdemes az X valószínűségi változót egy másik valószínűségi változó függvényeként felírni. Természetes módon adódik, hogy legyen Y a kockadobás során dobott szám. Ekkor felírhatjuk X-et Y segítségével:

$$X=Y \bmod 4\,.$$

Így a transzformáció függvénye a $h(x)=x \bmod 4$ lesz. Írjuk fel táblázatba az Y eloszlását, majd minden Y értékhez számoljuk ki a transzformáltat is:

a dobások lehetséges értékei (y_i)	a dobás valószínűsége (p_{y_i})	a dobás maradéka ($h(y_i)$)	$h(y_i) \cdot p_{y_i}$	a maradék négyzete ($h^2(y_i)$)	$h^2(y_i) \cdot p_{y_i}$
1	$1/6$	1	$1/6$	1	$1/6$
2	$1/6$	2	$2/6$	4	$4/6$
3	$1/6$	3	$3/6$	9	$9/6$
4	$1/6$	0	0	0	0
5	$1/6$	1	$1/6$	1	$1/6$
6	$1/6$	2	$2/6$	4	$4/6$

A negyedik oszlop összege lesz az X valószínűségi változó várható értéke:

$$E(X) = (1+2+3+0+1+2)/6 = 9/6 = 1{,}5\,.$$

Figyeljük meg, hogy a táblázatban nem szerepel közvetlen X eloszlása, s amikor értékeit súlyozzuk, azok többször is előfordulnak, mivel a transzformációs függvény több y_i értéknél is azonos értékeket vesz fel. A második és harmadik oszlop segítségével meghatározhatjuk X eloszlását:

$$X : \begin{cases} 0 & 1 & 2 & 3 \\ 1/6 & 1/3 & 1/3 & 1/6 \end{cases}$$

Az eloszlás ismeretében a várható érték közvetlen is meghatározható:

$$E(X) = 0 \cdot 1/6 + 1 \cdot 1/3 + 2 \cdot 1/3 + 3 \cdot 1/6 = 9/6 = 1{,}5\,.$$

Tehát a két módszerrel ugyanaz az eredmény jött ki. Az értékek megértésének érdekében számoljuk ki Y várható értékét is! Mivel hatoldalú kockával dobunk, a dobás értékének várható értéke:

$$E(Y) = \frac{6+1}{2} = 3{,}5\,.$$

Ebből is látszik, hogy $E(h(Y)) \neq h(E(Y))$, legalábbis ebben az esetben. Néhány kivételes eset kivételével nem igaz az, hogy a transzformált várható értéke megegyezik a várható érték transzformáltjával.

b) A hatodik oszlop összege pedig az X^2 várható értéke:

$$E(X^2) = (1+4+9+0+1+4)/6 = 19/6 \approx 3{,}1667\,.$$

Határozzuk meg X eloszlása segítségével is X^2 várható értékét! Ehhez első lépésben írjuk fel X^2 eloszlását:

$$X^2 : \begin{cases} 0 & 1 & 4 & 9 \\ 1/6 & 1/3 & 1/3 & 1/6 \end{cases}$$

Az eloszlásból a várható érték is meghatározható:

$$E(X^2) = 0\cdot 1/6 + 1\cdot 1/3 + 4\cdot 1/3 + 9\cdot 1/6 = 19/6 \approx 3{,}1667\,.$$

Láthatjuk, hogy ismét mindkét módszerrel ugyanazt az eredményt kaptuk. A szórást X^2 számításának módszerétől függetlenül, mindkét esetben ugyanúgy számoljuk:

$$D(X) = \sqrt{\frac{19}{6} - \left[\frac{9}{6}\right]^2} = \frac{\sqrt{33}}{6} \approx 0{,}9574\,.$$

Tehát az X tényleges értékeinek szóródása a várható érték $(1{,}5)$ körül $\approx 0{,}9574$. Nézzük meg a kockadobás értékének szórását is:

$$D(Y) = \sqrt{\frac{6^2-1}{12}} \approx 1{,}7078$$

Láthatjuk, hogy a kockadobás értékei jóval nagyobb mértékben fognak a várható érték $(3{,}5)$ körül ingadozni. Lássuk be azt is, hogy ennek a két értéknek sincs semmi közvetlen köze a néggyel való oszthatósághoz.

⇐13.7. feladat

Aktivitás: Gondolja végig, milyen esetben érdemes a transzformált valószínűségi változó várható értékét a saját, és milyen esetben az eredeti valószínűségi változó eloszlásának segítségével kiszámolni!

13.7. tétel: Legyen Y folytonos eloszlású valószínűségi változó, a sűrűségfüggvénye legyen $f_Y(x)$, és legyen $X = h(Y)$. Ekkor

1. X várható értéke
$$E(X) = \int_{-\infty}^{\infty} h(x) \cdot f_Y(x)\,\mathrm{d}x,$$
2. X^2 várható értéke
$$E(X^2) = \int_{-\infty}^{\infty} h^2(x) \cdot f_Y(x)\,\mathrm{d}x,$$
3. X szórásnégyzete
$$D^2(X) = \left[\int_{-\infty}^{\infty} h^2(x) \cdot f_Y(x)\,\mathrm{d}x\right] - \left[\int_{-\infty}^{\infty} h(x) \cdot f_Y(x)\,\mathrm{d}x\right]^2.$$

Bizonyítás:

1. Mivel minden függvény felbontható szigorúan monoton függvények összegére[1], ezért a tételt elég szigorúan monoton h-ra bizonyítani.

$$E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} x \cdot f_Y\left(h^{-1}(x)\right) \cdot \left(h^{-1}(x)\right)'\,\mathrm{d}x\,.$$

[1] Pontosabban minden olyan függvény, amely folytonos és minden véges intervallumon korlátos variációjú.

Használjuk fel, hogy az inverz függvényre vonatkozó deriválási szabály miatt: $\left(h^{-1}(x)\right)' = \dfrac{1}{h'\left(h^{-1}(x)\right)}$, s ezután pedig végezzük el az $y = h^{-1}(x)$ helyettesítést. Ekkor $x = h(y)$, $\mathrm{d}x = h'(y)\mathrm{d}y$. Ezzel

$$E(X) = \int_{-\infty}^{\infty} x \cdot f_Y\left(h^{-1}(x)\right) \cdot \frac{1}{h'\left(h^{-1}(x)\right)}\,\mathrm{d}x = \int_{-\infty}^{\infty} h(y) \cdot f_Y(y) \cdot \frac{1}{h'(y)} \cdot h'(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} h(y) \cdot f_Y(y)\,\mathrm{d}y\,.$$

Ha h szigorún monoton csökkenő, akkor a levezetés teljesen hasonlóan elvégezhető. Ha pedig h nem szigorúan monoton, akkor felbontható szigorúan monoton függvények összegére.

2. Az $X^2 = h^2(Y)$ transzformációt használva, az előző tételből adódik.
3. Az X^2 várható értékéből és a szórásnégyzet definíciójából adódik.

□

13.8. feladat. Legyen az Y valószínűségi változó eloszlásfüggvénye a következő:

$$F_Y(x) = \begin{cases} 1 - \dfrac{1}{x^2}, & \text{ha } x > 1; \\ 0 & \text{egyébként}\,. \end{cases}$$

Vizsgáljuk meg, hogy az alábbi transzformációk milyen hatással vannak Y várható értékére:

a) $h(x) = \sqrt{x}$, *b)* $h(x) = 1/x$, *c)* $h(x) = 4 \cdot x - 1$.

Megoldás: A várható érték vizsgálatához először számoljuk ki magának Y-nak a várható értékét. Ehhez Y sűrűségfüggvényére lesz szükségünk, melyet az eloszlásfüggvény differenciálásával megkaphatunk:

$$f_Y(x) = [F_Y(x)]' = \begin{cases} \dfrac{2}{x^3}, & \text{ha } x > 1; \\ 0 & \text{egyébként}\,. \end{cases}$$

Az eloszlásfüggvény első sorát az alábbiak szerint deriváltuk:

$$\left[1-\frac{1}{x^2}\right]' = \left[-x^{-2}\right]' = 2\cdot x^{-3}\,.$$

Y sűrűségfüggvényének segítségével felírható a várható értéke:

$$E(Y) = \int_{-\infty}^{\infty} x\cdot f_Y(x)\,\mathrm{d}x = \int_{1}^{\infty} x\cdot\frac{2}{x^3}\,\mathrm{d}x = \int_{1}^{\infty} 2\cdot x^{-2}\,\mathrm{d}x = 2\cdot\left[\frac{x^{-1}}{-1}\right]_1^{\infty} = 2\,.$$

a) Legyen $X_a = \sqrt{Y}$. Ebben az esetben X_a várható értékét az alábbiak szerint számolhatjuk:

$$E(X_a) = \int_{-\infty}^{\infty} h(x)\cdot f_Y(x)\,\mathrm{d}x = \int_{1}^{\infty} \sqrt{x}\cdot\frac{2}{x^3}\,\mathrm{d}x = \int_{1}^{\infty} 2\cdot x^{-5/2}\,\mathrm{d}x = 2\cdot\left[\frac{x^{-3/2}}{-3/2}\right]_1^{\infty} = \frac{4}{3}.$$

Láthatjuk, hogy ugyan a várható érték csökkent, a tényleges értéknek nem sok köze van az eredeti várható érték gyökéhez. Figyeljük meg, hogy az integrálás határa úgy változik, hogy x csak olyan értéket vehet fel, ahol a sűrűségfüggvény nem nulla.

b) A következő esetben vegyük az $X_b = 1/Y$ valószínűségi változót. Az X_b várható értékét így számolhatjuk:

$$E(X_b) = \int_{-\infty}^{\infty} h(x)\cdot f_Y(x)\,\mathrm{d}x = \int_{1}^{\infty} \frac{1}{x}\cdot\frac{2}{x^3}\,\mathrm{d}x = \int_{1}^{\infty} 2\cdot x^{-4}\,\mathrm{d}x = 2\cdot\left[\frac{x^{-3}}{-3}\right]_1^{\infty} = \frac{2}{3}.$$

Bár a várható érték nem lett az eredeti reciproka, a megfelelő intervallumba került: míg az eredeti valószínűségi változó az $[1;\infty]$ intervallumból kapta az értékeket, addig a reciproka a $[0;1]$-ből kapja. Természetesen a várható értéknek is ebben az intervallumban kell lennie.

c) Utoljára vizsgáljuk meg, mi történik, ha a transzformáció lineáris. Legyen $X_c = 4 \cdot x - 1$. Számoljuk ki várható értékét:

$$E(X_c) = \int_{-\infty}^{\infty} h(x) \cdot f_Y(x)\,\mathrm{d}x = \int_{1}^{\infty} (4 \cdot x - 1) \cdot \frac{2}{x^3}\,\mathrm{d}x = \int_{1}^{\infty} 8 \cdot x^{-2} - 2 \cdot x^{-3}\,\mathrm{d}x = \left[8 \cdot \frac{x^{-1}}{-1} - 2 \cdot \frac{x^{-2}}{-2} \right]_1^{\infty} = 7\,.$$

Az eddigi esetekkel ellentétben itt viszont a várható érték is transzformálódik. Ismerjük a tulajdonságait, melyek alapján X_c várható értéke lineáris transzformációval is számolható:

$$E(X_c) = E(4 \cdot Y - 1) = 4 \cdot E(Y) - 1 = 7\,.$$

Így egy gyakorlati példán is beláttuk, hogy a várható érték csak nagyon ritkán számolható úgy, ahogy a valószínűségi változó felvett értékei: csak a lineáris átalakítás transzformálja ugyanúgy a várható értéket.

⇐13.8. feladat

13.8. tétel: Legyen Y egyenletes eloszlású a és b között, és $X = h(Y)$. Ekkor

1. X várható értéke:

$$E(X) = \frac{1}{b-a} \int_a^b h(x)\,\mathrm{d}x,$$

2. és X^2 várható értéke:

$$E(X^2) = \frac{1}{b-a} \int_a^b h^2(x)\,\mathrm{d}x\,.$$

Bizonyítás:

1. Mivel Y sűrűségfüggvénye $f_Y(x) = \begin{cases} \dfrac{1}{b-a}, & \text{ha } a < x < b; \\ 0 & \text{egyébként;} \end{cases}$ így az előző tétel alapján

$$E(X) = \int_{-\infty}^{\infty} h(x) f_Y(x)\,\mathrm{d}x = \int_a^b h(x)\frac{1}{b-a}\,\mathrm{d}x = \frac{1}{b-a}\int_a^b h(x)\,\mathrm{d}x\,.$$

2. Az $X^2 = h^2(Y)$ transzformációt használva adódik.

□

Aktivitás: Egy folytonos valószínűségi változó várható értéke meghatározható a sűrűségfüggvényének görbe alatti területének súlypontjával. Tud-e valami hasonlót mondani a $h(x)$ függvényről és a transzformált várható értékéről?

13.9. feladat. A $[0; 1]$ intervallumból válasszunk egy valós számot. A számmal mint sugárral rajzoljunk kört. Adjuk meg a kapott kör területének

a) várható értékét és

b) szórását!

Megoldás:

a) A kör területének, mint valószínűségi változónak a sűrűségfüggvénye felírható, amiből a várható érték meghatározható. Jelen esetben azonban célravezetőbb a területet a sugár mint valószínűségi változó függvényeként felírni. Legyen az intervallumból választott szám Y, ez lesz a kör sugara, és a rajzolt kör területe X. Ekkor $X = Y^2 \cdot \pi$, azaz a transzformáció függvénye $h(x) = x^2 \cdot \pi$. Figyeljük meg, hogy most a várható érték közbenső lépések nélkül kiszámolható:

$$E(X) = \frac{1}{b-a}\int\limits_a^b h(x)\,\mathrm{d}x = \frac{1}{1-0}\int\limits_0^1 x^2 \cdot \pi\,\mathrm{d}x = \pi\int\limits_0^1 x^2\,\mathrm{d}x = \pi\left[\frac{x^3}{3}\right]_0^1 = \frac{\pi}{3} \approx 1{,}0472\,.$$

Így a kör területének várható értéke $\approx 1{,}0472$ lesz. Jegyezzük meg, hogy amennyiben Y várható értékével mint sugárral rajzolunk kört, akkor a kör területe $\pi/4$ lesz.

b) Ahhoz, hogy a kör területének szórását meg tudjuk határozni, először ki kell számolnunk négyzetének várható értékét:

$$E(X^2) = \frac{1}{b-a}\int\limits_a^b h^2(x)\,\mathrm{d}x = \frac{1}{1-0}\int\limits_0^1 \left[x^2\cdot\pi\right]^2\,\mathrm{d}x = \pi^2\int\limits_0^1 x^4\,\mathrm{d}x = \pi^2\left[\frac{x^5}{5}\right]_0^1 = \frac{\pi^2}{5} \approx 1{,}9739\,.$$

Így, hogy megvan X és X^2 várható értéke, a szórás már könnyen számítható:

$$D(X) = \sqrt{E(X^2) - E^2(X)} = \sqrt{\frac{\pi^2}{5} - \left[\frac{\pi}{3}\right]^2} = \pi\sqrt{\frac{1}{5} - \frac{1}{9}} = \frac{2\pi}{\sqrt{45}} \approx 0{,}9366\,.$$

Tehát a kör alapterületének szórása $\approx 0{,}9366$ lesz.

⇐13.9. feladat

13-5. önálló feladat: Oldja meg a fenti feladatot úgy, hogy a választott sugárral gömböt rajzolunk, s a gömb térfogatának, illetve felületének várható értékét vizsgáljuk!

Leckezáró önellenőrző kérdéssor

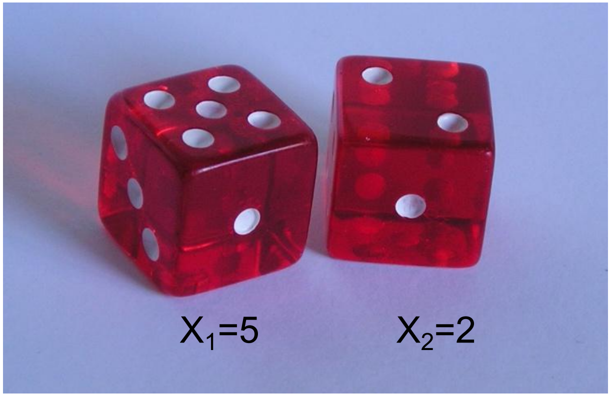
1. Dobókockával dobunk. Legyen X a dobott szám négyzete. X várható értéke ekkor
   - ☐ a dobott érték várható értékének négyzete.
   - ☐ nagyobb lesz, mint a dobott érték várható értékének négyzete.
   - ☐ kisebb lesz, mint a dobott érték várható értékének négyzete.
   - ☐ véletlenszerűen ingadozni fog.
2. Nyolcoldalú dobókockával dobunk. Legyen X a dobott szám néggyel vett maradéka! Számoljuk ki X várható értékét!
   - ☐ 2
   - ☐ 0
   - ☐ 2,3
   - ☐ 1,5
3. Az X valószínűségi változót egy olyan $h(x)$ függvény állítja elő, melynek értékkészlete az $[1;\ 3]$ intervallum. Ekkor X várható értéke
   - ☐ nem lehet pozitív,
   - ☐ lehet 3-nál nagyobb,
   - ☐ 1-nél csak nagyobb lehet,
   - ☐ lehet 1, vagy akár 3 is.
4. Egy dobókockával dobunk. Legyen X a dobott szám néggyel vett maradéka. Mi a valószínűsége annak, hogy X értéke legfeljebb a szórással tér el a várható értékétől?
   - ☐ $2/3$
   - ☐ $1/3$
   - ☐ $1/6$
   - ☐ $1/2$
5. A legutolsó kidolgozott feladatban a választott számot a kör rajzolása előtt a duplájára növeljük. Hányszorosára nő a kör területének várható értéke?
   - ☐ kétszeresére,
   - ☐ négyszeresére,
   - ☐ nyolcszorosára,
   - ☐ nem fog nőni.

19. LECKE

Több valószínűségi változó együttes eloszlása. Diszkrét valószínűségi változók együttes eloszlása és függetlensége.

14. Több valószínűségi változó együttes eloszlása

Sok gyakorlati problémánál előfordul, hogy a jelenség leírásához egyszerre több valószínűségi változót használunk. Nézzünk erre egy pár példát:



14.1. ábra. *Ha a két dobókockával 5-öt és 2-t dobunk, akkor $X_1 = 5$ és $X_2 = 2$ lesz.*

1. Két dobókockával dobunk egyszerre. Az eredmény számszerűsítésére két valószínűségi változót használunk: legyen X_1 az első és X_2 a második kockával dobott szám. A két érték nyilván nem fog egymástól függeni.
2. Egyidejűleg mérjük egy adott embercsoport magasságát és testtömegét. X fogja leírni a véletlenszerűen választott ember magasságát, Y pedig a testtömegét. Ez a két érték nyilván összefügg egymással, de mégsem mondhatjuk, hogy függvényszerű kapcsolat áll fenn közöttük, hiszen nem csak magas és nehéz, illetve alacsony és könnyű, hanem sovány magasak és testesebb alacsony növésűek is. Ilyen esetben a teljes valószínűségi leíráshoz ismernünk kell azt, hogy a két valószínűségi változó együttesen hogyan viselkedik.
3. Egy fizikai kísérlet során valamely paraméter (mondjuk a hőmérséklet) értékét mérjük. Az eljárást háromszor ismételve a mért paraméter három értéke X_1, X_2 illetve X_3. Ez három olyan valószínűségi változó, melyek eloszlása azonos, valamint – amennyiben a kísérleti beállítást sikerül teljesen visszaállítani – függetlennek tekinthetőek.
4. Olyan valószínűségi változókat is vizsgálhatunk együtt, melyek közül az egyik a másik transzformáltja. Például ha X_1 az éves szinten aszálykár miatt tönkrement kukoricaföldek nagysága hektárban, X_2 lehet az aszálykár miatt tönkrement termék tömege kilogrammban, X_3 pedig a keletkezett kár forintban.

Ahhoz, hogy több valószínűségi változót együtt tudjunk kezelni matematikailag, érdemes bevezetni egy erre megfelelő objektumot. Számok rendezett halmazát vektorként kezeljük, ehhez hasonlóan kezelhetjük a valószínűségi változókat egy úgynevezett valószínűségi vektorváltozóként. A valószínűségi vektorváltozót az alábbiak szerint definiáljuk:

14.1. definíció: **Valószínűségi vektorváltozó**
Az $X_1, X_2, \ldots, X_n$ együtt megfigyelt valószínűségi változók rendezett összességét n dimenziós valószínűségi vektorváltozónak nevezzük, és az $(X_1, X_2, \ldots, X_n)$ szimbólummal jelöljük.

Természetesen itt mindegyik komponensnek lehet különböző az eloszlása, viszont esetenként akár az összes eloszlása is lehet egyforma. Erre megkötés nincs. A valószínűségi vektorváltozó értéke minden egyes elemi esemény esetén egy valós számokból álló vektor. A vektor azon értékekből áll, melyeket az egyes valószínűségi változók az adott elemi esemény mellett felvesznek. Meg kell említeni, hogy vektorváltozó esetén az egyes valószínűségi változók alaphalmazai nem feltétlen egyeznek meg. Gondoljuk át, hogy a fenti példák esetén hogy alakulnak az alaphalmazok:

1. Amennyiben két kockával dobunk egyszerre, akkor az első alaphalmazunk az első kockával történő dobás lehetséges kimenetelei, ebből hatféle van. A másik kockával is hatfélét dobhatunk, így ezen dobáshoz tartozó alaphalmaz is hatelemű. A vektorváltozó esetén bármely variáció előjöhet, így egy 36 elemű alaphalmazt kapunk.

14-1. önálló feladat: Írja fel az első példában szereplő valószínűségi vektorváltozóhoz tartozó alaphalmazt!

2. A második példánál az elemi eseményünk ahhoz a személyhez kapcsolódik, akinek a magasságát vagy a súlyát mérjük. A két alaphalmaz megegyezik, így a vektorváltozónak is ez lesz az alaphalmaza. Gondoljuk meg, hogy ugyan az előző feladatban is mintha össze tudnánk vonni a két alaphalmazt, de az tönkretenné a modellt. Ebben az esetben épp az tenné tönkre, ha a két halmazt nem vonnánk össze.

3. Az egymás utáni mérésnek nyilván úgy van értelme, ha ezeket ténylegesen el is végezzük. Itt sem tudjuk épp ezért összevonni az alaphalmazokat. A három alaphalmaz a három kísérlet lehetséges kimenetelei lesznek.
4. Gondoljuk át, hogy itt ismét arról az esetről van szó, amikor egy adott elemi eseményhez (azaz egy adott káreseményhez) adunk meg több valószínűségi változóval értéket: a károsodott terület nagyságát, a kiesett termény tömegét, valamint értékét. Ennek ismét úgy van csak értelme, ha azonos az alaphalmaz mind a három valószínűségi változónál, mind a vektorváltozónál.

Aktivitás: Keressen a környezetben a fenti két típusra valós példákat!

A valószínűségi vektorváltozók között is két speciális típussal foglalkozunk. A folytonos és a diszkrét valószínűségi vektorváltozó minden egyes komponense azonos típusú. Ennek megfelelően az alábbiak szerint definiáljuk őket:

14.2. definíció: **Diszkrét (folytonos) valószínűségi vektorváltozó**
Egy valószínűségi vektorváltozót diszkrétnek (folytonosnak) nevezünk, ha valamennyi komponense diszkrét (folytonos).

Az előbbi példákra visszatérve vizsgáljuk meg, hogy ott diszkrét, avagy folytonos valószínűségi vektorváltozókat adtunk-e meg. Természetesen előfordulhat az is, hogy egy adott valószínűségi vektorváltozó egyik kategóriába sem tartozik: például ha komponensei vegyesen diszkrét és folytonos valószínűségi változók.

1. Két kockával dobva mindegyik kockával 1 és 6 közötti egész számokat kaphatunk, így mindegyik valószínűségi változó 6 értéket vehet fel, azaz diszkrét lesz mindkettő, így a vektorváltozó is.
2. Jelenleg alig több, mint hét milliárd ember él a földön. Akármekkora embercsoportot is vizsgálunk, a valószínűségi változók által felvett értékek száma nem haladhatja meg ezt a hétmilliárdot. Ezek alapján

ez a vektorváltozó mindenképp diszkrét lesz. Természetesen több egyéb tényezőt is figyelembe vehetünk. Egyrészt az emberek magassága és tömege – bizonyos élettani funkcióknak megfelelően – változik. A magasság mindenképp folytonosan. Másrészt ebben az esetben is számolnunk kell egyes mérési hibákból adódó ingadozásokkal. Amennyiben az egyidejűséget és a mérési precizitást nem tudjuk biztosítani, és az embercsoport számossága elég nagy, az eloszlást kezelhetjük folytonosként.

3. A hőmérséklet alapvetően folytonos skálán mozog. Azt várjuk, hogy amikor egy kísérletet többször, azonos felállásban elvégzünk, akkor az általunk érdekesnek vélt hőmérséklet mindhárom kísérletnél azonos lesz. A valóságban azonban azt tapasztaljuk, hogy számos olyan tényező befolyásolhatja az eredményt, amire hatással nem tudunk lenni. Ezek a hatások lehetnek a rendszer kísérleti mivoltjából adódó fizikai tényezők, illetve a megfigyelési pontatlanságból adódó mérési hibák. A ténylegesen mért értékek így a várt körüli, valamilyen folytonos skálán fognak mozogni, így a valószínűségi vektorváltozónk folytonos lesz.
   Bele kell gondolnunk azonban abba is, hogy amennyiben a mérést valamilyen digitális eszközzel mérjük, a mért értékek számossága – akármilyen finom felbontással is dolgozunk, az eszköz érzékenységének és méréshatárának véges volta miatt – véges. Gondoljunk arra a digitális lázmérőre, ami 35 és 41 Celsius fok között tized fok pontossággal mér. Ez összesen kevesebb, mint 65-féle érték.

4. Az aszálykáros kukoricaföldek példájában mind az aszálykárt szenvedett terület nagysága, mind a kárba veszett kukorica tömege, valamint annak értéke is folytonos értéket vehet fel. Gondoljuk meg, hogy maga a földterület nagysága természetesen akármilyen kis részekre osztható. Fix hektáronkénti terméshozammal, illetve fix terményárakkal számolva a másik két valószínűségi változó is folytonos lesz, így a valószínűségi vektorváltozó is folytonos lesz.

Aktivitás: Keressen olyan példákat a való életből, amikor egy látszólag folytonos eloszlású véletlen mennyiség valamilyen anomália miatt mégis diszkrét értékeket vesz fel.

A továbbiakban a definíciókat, tételeket két valószínűségi változóra (X és Y) mondjuk ki. A teljes indukció elve alapján ugyanis a legtöbb állítás természetes módon általánosítható tetszőleges, véges számú valószínűségi változó esetére. Egy későbbi leckében mutatjuk be n darab valószínűségi változó együttes kezelését.

14.1. Két diszkrét valószínűségi változó együttes eloszlása

Diszkrét valószínűségi változók esetén a változók viselkedését elsősorban az eloszlás segítségével írjuk le. A valószínűségek felsorolásával jellemezzük a vektorváltozókat is, melyet a komponensek együttes eloszlásnak hívunk, s az alábbiak szerint definiálunk:

14.3. definíció: **Együttes eloszlás**
Az X és Y diszkrét valószínűségi változók együttes eloszlásán az (x_i,y_k) számpárok és a hozzájuk tartozó

$$p_{ik} = P(X = x_i, Y = y_k) = P\left(\{X = x_i \text{ és } Y = y_k\}\right)$$

valószínűségek összességét értjük.

Miként írhatjuk fel az együttes eloszlást? Erre két diszkrét valószínűségi változó esetén két lehetőségünk van. A példákban feltesszük, hogy X-nek n, Y-nak pedig m-féle értéke lehet:

1. Felírhatjuk úgy, mint ahogy az egyszerű diszkrét valószínűségi változó eloszlását felírjuk. Ekkor a nem nulla valószínűséggel előforduló értékpárokat soroljuk fel, alattuk a hozzájuk tartozó valószínűségeket felsorolva:
$$\begin{pmatrix} X \\ Y \end{pmatrix} : \left\{ \begin{array}{cccccc} (x_0,y_0) & (x_0,y_1) & \ldots & (x_i,y_k) & \ldots & (x_n,y_m) \\ p_{00} & p_{01} & \ldots & p_{ik} & \ldots & p_{nm} \end{array} \right.$$
Figyeljük meg, hogy ebben az esetben X és Y értékeinek nem mindegyik párosítása jelenik meg a felsorolásban, csak azok, melyek egyszerre előfordulhatnak.
2. A valószínűségeket táblázatos formában is felírhatjuk. A táblázat oszlopait X, sorait pedig Y értékeivel jelöljük. Adott oszlopba és sorba annak a p_{ik} valószínűsége kerül, hogy X az adott oszlopot jelölő x_i, Y pedig az adott sort jelölő y_k értéket veszi fel:

$$\begin{array}{c|cccccc}
X\backslash X & y_0 & y_1 & \dots & y_k & \dots & y_m \\
\hline
x_0 & p_{00} & p_{01} & \dots & p_{0k} & \dots & p_{0m} \\
x_1 & p_{10} & p_{11} & \dots & p_{1k} & \dots & p_{1m} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
x_i & p_{i0} & p_{i1} & \dots & p_{ik} & \dots & p_{im} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
x_n & p_{n0} & p_{n1} & \dots & p_{nk} & \dots & p_{nm}
\end{array}$$

Itt azon (x_i,y_k) párok is szerepelnek, melyek egyébként együtt nem fordulhatnak elő. Ekkor a táblázatba 0 kerül. Ilyen módszerrel csak két valószínűségi változó esetén tudjuk az együttes eloszlást felírni.

14.1. feladat. Egy pakli magyar kártyából húzunk egy kártyát. Legyen X a húzott királyok, Y pedig a húzott ászok száma. Adjuk meg X és Y együttes eloszlását!

Megoldás: Mindkét valószínűségi változó lehetséges értéke 0 és 1 lehet, és az (X,Y) vektorváltozó lehetséges értékei $(0,0)$, $(0,1)$ és $(1,0)$. Egyszerre nem lehet mindkét változó értéke 1, hiszen nem húzhatunk egyszerre királyt és ászt is. Az eloszlást két valószínűségi változó esetén a fentiek szerint kétféleképpen adhatjuk meg.

1. Az első felírás szerint a felsorolt lehetséges értékpárok alá annak a valószínűségét írjuk, hogy a valószínűségi vektorváltozó az adott értékpárt veszi fel. Számoljuk ki most azokat a valószínűségeket, melyekkel a vektorváltozó az egyes értékpárokat felveszi.
   (a) Ha $X=0$ és $Y=0$, akkor nem húzunk se királyt, se ászt. Ekkor $32-4-4=24$ lap lesz jó a 32-ből, tehát
   $$P(X=0,Y=0)=\frac{24}{32}=\frac{3}{4}\,.$$
   (b) Ha $X=1$ és $Y=0$, akkor királyt húzunk, de ászt nem. Ekkor 4 lap lesz jó a 32-ből, tehát
   $$P(X=1,Y=0)=\frac{4}{32}=\frac{1}{8}\,.$$

(c) Ha $X = 0$ és $Y = 1$, akkor ászt húzunk, de királyt nem. Ekkor is 4 lap lesz jó a 32-ből, tehát

$$P(X = 0, Y = 1) = \frac{4}{32} = \frac{1}{8}.$$

Egyszerre ászt és királyt nem húzhatunk, tehát $P(X = 1, Y = 1) = 0$. Ezek alapján a vektorváltozó eloszlását az alábbiak szerint tudjuk felírni:

$$\begin{pmatrix} X \\ Y \end{pmatrix} : \begin{cases} (0,0) & (0,1) & (1,0) \\ 3/4 & 1/8 & 1/8 \end{cases}$$

2. Az eloszlást táblázatos formában is felírhatjuk. Mivel 0 vagy 1 ászt vagy királyt húzhatunk, így mind X, mind Y lehetséges értékei 0 és 1. A táblázatunknak így a sor és oszlopcímkék mellett 2 sora, és 2 oszlopa lesz. A cellákba írandó valószínűségeket az előző pontban már kiszámoltuk, így a táblázatot felírhatjuk:

$X \backslash Y$	0	1
0	$6/8$	$1/8$
1	$1/8$	0

Ezzel az eloszlást kétféleképpen is megadtuk. Bár az első felírás hasonlít jobban ahhoz, amit a hétköznapi valószínűségi változó eloszlásánál megismerhettünk, a táblázatos felírás mégis áttekinthetőbb, és esetenként célravezetőbb ezt használni.

⇐14.1. feladat

Valamely diszkrét valószínűségi vektorváltozó együttes eloszlását ismerve megkaphatjuk annak komponenseinek eloszlását is, melyet peremeloszlásnak hívunk:

14.4. definíció: **Peremeloszlás**

Az (X, Y) valószínűségi vektorváltozó komponenseinek eloszlását peremeloszlásoknak nevezzük.

A peremeloszlást az alábbi tétel alkalmazásával határozhatjuk meg:

14.1. tétel: Legyen (X,Y) kétdimenziós diszkrét valószínűségi vektorváltozó, a $P(X = x_i, Y = y_k)$ valószínűséget pedig jelölje p_{ik}. Ekkor a peremeloszlásokat (X és Y eloszlását) az alábbi formulákkal kapjuk meg:

$$p_{x_i} = P(X = x_i) = \sum_k p_{ik} \qquad \text{és} \qquad p_{y_k} = P(Y = y_k) = \sum_i p_{ik}\,.$$

Tehát a peremeloszlást, azaz a valószínűségi vektorváltozó egyes komponenseinek eloszlását úgy kaphatjuk meg, ha az együttes eloszlás megfelelő értékeit összeadjuk. Ezek az értékek táblázatos felírás esetén egy sorban, illetve egy oszlopban vannak: ezeket összeadva kaphatjuk meg azon valószínűségeket, melyekkel az egyes komponens valószínűségi változók értékeiket felveszik.

14.2. feladat. Egy pakli magyar kártyából húzunk egy kártyát. Legyen X a húzott királyok, Y pedig a húzott ászok száma. Adjuk meg az (X, Y) vektorváltozó komponenseinek peremeloszlásait!

Megoldás: Az előző feladatban az eloszlást már meghatároztuk, s táblázatos formában fel is írtuk. Annyi feladatunk van pluszban, hogy az oszlopokban és a sorokban szereplő valószínűség értékeket összegezzük:

$X\backslash Y$	0	1	p_{x_k}
0	$6/8$	$1/8$	$7/8$
1	$1/8$	0	$1/8$
p_{y_k}	$7/8$	$1/8$	

A táblázatban az összegzés után megjelennek az X és az Y valószínűségi változókhoz tartozó eloszlások valószínűségei. Így felírhatjuk X és Y eloszlását külön külön is:

$$X : \begin{cases} 0 & 1 \\ 7/8 & 1/8 \end{cases} \qquad \text{és} \qquad Y : \begin{cases} 0 & 1 \\ 7/8 & 1/8 \end{cases}.$$

Láthatjuk, hogy a két eloszlás megegyezik. Természetesen ez nem meglepő, hiszen a feladatban az ász és a király szerepe felcserélhető: ugyanakkora valószínűséggel húzzuk őket.

⇐14.2. feladat

Aktivitás: Írjon fel olyan valószínűségi változó párokat, melyeknél bizonyos értékpárok kizárják egymást!

14.2. Diszkrét valószínűségi változók függetlensége

Emlékezzünk vissza a függetlenség definíciójára: az A és B eseményeket akkor nevezzük függetlennek, ha teljesül a $P(A \cdot B) = P(A) \cdot P(B)$ egyenlőség, azaz az együttes bekövetkezés valószínűsége a valószínűségek szorzata. Két diszkrét valószínűségi változót akkor nevezünk függetlennek, ha ez minden velük kapcsolatos eseménypárra teljesül, pontosabban:

14.5. definíció: **Diszkrét valószínűségi változók függetlensége**
Az X és Y diszkrét eloszlású valószínűségi változók függetlenek, ha

$$P(X = x_i, Y = y_k) = P(X = x_i) \cdot P(Y = y_k),$$

azaz

$$p_{ik} = p_{x_i} \cdot p_{y_k}$$

teljesül minden x_i, y_k értékpár esetén.

A fenti definíció azt jelenti, hogy két független valószínűségi változó esetén amennyiben tudjuk az egyik változó felvett értékét, nem jutunk új információhoz azzal kapcsolatban, milyen értéket fog a másik valószínűségi változó felvenni.

14.3. feladat. A 32 lapos magyarkártya-pakliból húzunk egy kártyát. Legyen X a húzott királyok, Y pedig a húzott ászok száma. Igaz-e hogy X és Y független?

Megoldás: A függetlenség vizsgálatakor a következőt kell meggondolni: amennyiben ismerjük az egyik valószínűségi változó bármilyen értékét, megváltoznak-e azok a valószínűségek, melyekkel a másik valószínűségi változó az értékeit felveszi? Ebben az esetben természetesen igen. Ha X értéke 1, azaz királyt húztunk, akkor Y mindenképp 0 lesz, hiszen nem húzhatunk ászt is. Tehát a két valószínűségi változó nem lesz független.

A választ úgy is megkaphatjuk, hogy összehasonlítjuk az együttes eloszlás p_{ik} valószínűségeit a peremeloszlásokból nyert $p_{x_i} \cdot p_{y_k}$ valószínűségekkel:

$$\begin{aligned}
3/4 = P(X=0, Y=0) &\neq P(X=0)\cdot P(Y=0) = 7/8 \cdot 7/8 = 49/64\\
1/8 = P(X=1, Y=0) &\neq P(X=1)\cdot P(Y=0) = 1/8 \cdot 7/8 = 7/64\\
1/8 = P(X=0, Y=1) &\neq P(X=0)\cdot P(Y=1) = 7/8 \cdot 1/8 = 7/64
\end{aligned}$$

Ahhoz, hogy a függetlenség megbukjon, elég egy értékpárt vizsgálni. Itt viszont egyik értékpárra sem igaz a függetlenség feltétele. Így beláttuk számolással is, hogy X és Y nem független. ⇐14.3. feladat

14-2. önálló feladat: A 32 lapos magyarkártya-pakliból *két* lapot húzunk visszatevéssel. Legyen X a húzott királyok, Y a húzott ászok száma. Írja fel X és Y együttes eloszlását, határozza meg a peremeloszlásokat, majd vizsgálja meg, független lesz-e X és Y.

Leckezáró önellenőrző kérdéssor

1. Egy dobókockát elgurítunk. Legyen X a dobott szám, Y pedig a kocka megállásáig eltelt idő, amit stopperrel mérünk. Ebben az esetben az (X,Y) valószínűségi vektorváltozó

☐ folytonos, ☐ diszkrét, ☐ mindkettő, ☐ egyik sem.

2. Magyarország lakosságának vagyonát és hitelállományát vizsgáljuk. Legyen X egy adott pillanatban véletlenszerűen választott polgár vagyona, Y pedig adósságainak nagysága. Ekkor az (X, Y) valószínűségi vektorváltozó

☐ diszkrét, mert az alaphalmaz véges elemszámú

☐ folytonos, mert a vagyon folytonos értékeket vesz fel

☐ diszkrét, ha X és Y is diszkrét

☐ kizárólag folytonos eloszlású lehet.

3. A 32 lapos magyarkártya-pakliból egy lapot húzunk. Legyen X a kihúzott ászok, Y pedig a kihúzott pirosak száma. Mekkora a valószínűsége annak, hogy X és Y is 1-et vesz fel?

☐ 1 ☐ $1/8$ ☐ $1/32$ ☐ 2

4. A 32 lapos magyarkártya-pakliból egy lapot húzunk. Legyen X a kihúzott ászok, Y pedig a kihúzott pirosak száma. Igaz-e, hogy X és Y független?

☐ Igen, mert a kártyalapokat függetlenül húzzuk.

☐ Nem, mert ha királyt húzunk, akkor nem húzhatunk pirosat.

☐ Igen, mert az, hogy ászt húzunk-e vagy nem, nem befolyásolja annak a valószínűségét, hogy a húzott kártya piros.

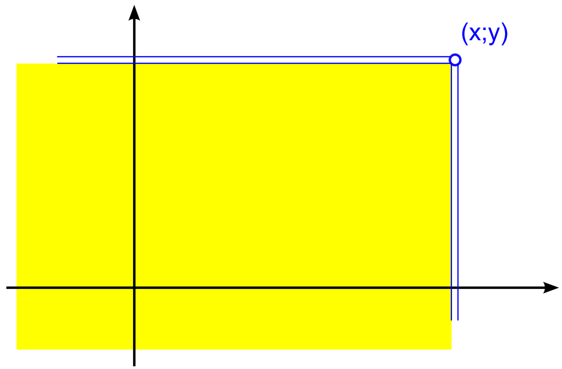
☐ Nem, mert a kártyát nem visszatevéssel húztuk.

20. LECKE

Két folytonos valószínűségi változó együttes eloszlásfüggvénye, peremeloszlásaik

14.3. Két valószínűségi változó együttes eloszlásfüggvénye

Egyetlen folytonos valószínűségi változó esetében a vele kapcsolatos valószínűségeket általában az eloszlásfüggvénye alapján határoztuk meg. Az X valószínűségi változó $F(x)$ eloszlásfüggvénye annak valószínűségét adta meg, hogy $X < x$. Ilyenkor X értékei az x-től balra eső félegyenesen helyezkednek el. Nem nulla valószínűségű események pedig a valós tengely bizonyos részintervallumai lehetnek: annak a valószínűsége, hogy X egy adott pontra vagy egy megszámlálható számosságú ponthalmazba essen, nulla.



14.2. ábra. *Az X és Y együttes eloszlásfüggvénye x és y pontban annak a valószínűségét adja meg, hogy az (X, Y) pontpár a koordinátasíkon az (x, y) pontpártól balra és lefele veszi fel az értékét. A negyedsík a tartománya határát nem tartalmazza.*

Két (X és Y) valószínűségi változó esetén is definiálhatjuk az eloszlásfüggvény fogalmát. Ezt az $X < x$ és $Y < y$ események együttes bekövetkezésének valószínűségével adjuk meg. Ekkor az (X,Y) számpárok lehetséges értékei az x-től balra és az y-tól lefelé elhelyezkedő negyedsíkon helyezkednek el. Folytonos valószínűségi vektorváltozó esetén nem nulla valószínűsége olyan tartományoknak lehet, melynek a síkbeli területe sem nulla: pontokra, megszámlálható ponthalmazokba, vonalakra, egyenesekre s ezek megszámlálható halmazaiba akkor is 0 valószínűséggel esik a vektorváltozó, ha ezek a lehetséges értékek tartományán belül vannak.

A valószínűségi vektorváltozó eloszlásfüggvényéről abban az esetben is beszélhetünk, amikor a vektorváltozó nem mindegyik komponense folytonos. A gyakorlatban az eloszlásfüggvényt viszont elsősorban ezek leírására használjuk.

14.6. definíció: Valószínűségi vektorváltozó eloszlásfüggvénye (együttes eloszlásfüggvény)
Azt a függvényt, amely minden valós (x,y) számpárhoz hozzárendeli annak valószínűségét, hogy az $X < x$ és az $Y < y$ események együttesen bekövetkeznek, az (X,Y) valószínűségi vektorváltozó eloszlásfüggvényének (az X és Y valószínűségi változók együttes eloszlásfüggvényének) nevezzük. Azaz

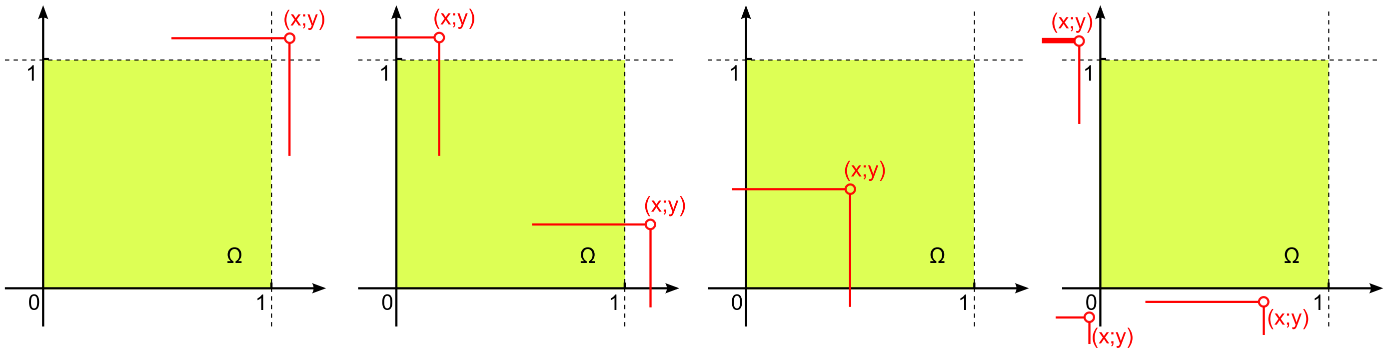
$$F(x,y) = P(X < x, Y < y)\,.$$

14.4. feladat. Vegyük a koordinátasíkon a $(0,0)$, $(0,1)$, $(1,1)$ és az $(1,0)$ pontok által meghatározott négyzetet. Válasszunk a négyzeten véletlenszerűen egy pontot. Legyen X a választott pont x, és Y az y koordinátája. Adjuk meg az (X,Y) valószínűségi vektorváltozó eloszlásfüggvényét!

Megoldás: A pontot a négyzeten belül véletlenszerűen választjuk. Ebben az esetben annak a valószínűsége, hogy egy adott négyzeten belüli területen belül választunk, arányos lesz a terület nagyságával. Az eloszlásfüggvény felírásához meg kell vizsgálni, hogy az egyes (x,y) számpárok esetén milyen valószínűséggel lesz együttesen X értéke kisebb, mint x, és Y kisebb, mint y. Az (x,y) által meghatározott síknegyed egy részt metszhet ki a négyzetből. Annak a valószínűsége, hogy a választott pont ide esik, arányos lesz a rész területével. Geometriai valószínűségi mezőről beszélünk, tehát

$$F(x,y) = P(X < x, Y < y) = \frac{\text{a síknegyed és a négyzet metszetének területe}}{\text{négyzet területe}}\,.$$

Ahhoz, hogy a valószínűséget megkapjuk, el kell osztanunk a rész területét a négyzet területével. Többféleképpen is vághatunk ki a négyzetből a síknegyed segítségével. Ezeket az eseteket az alábbi ábra segítségével elemezhetjük:



Nézzük az ábra egyes részeit:

bal szélső:
Amennyiben mind x, mind y nagyobb, mint 1, annak a valószínűsége, hogy mind X, mind Y kisebb lesz náluk, 1 lesz, hiszen a teljes négyzet a síknegyeden belül van.

bal középső:
Amennyiben y még mindig nagyobb, mint 1, de $0 < x < 1$, akkor Y mindig kisebb lesz, mint y, de annak a valószínűsége, hogy $X < x$ legyen, konkrétan x lesz: ekkora méretű részt vág ki a síknegyed a négyzetből. Hasonlóan gondoljuk meg, hogy az $x > 1$ és $0 < y < 1$ esetén az eloszlásfüggvény értéke y lesz.

jobb középső:
Amennyiben az (x,y) pont a négyzeten belül esik, a síknegyed és a négyzet metszete egy x és y oldalú téglalap lesz az origó mellett, melynek területe, s ezáltal az eloszlásfüggvény értéke $x \cdot y$ lesz.

jobb szélső:
Egyik más esetben sincs a négyzetnek és a síknegyednek közös pontja, így az eloszlásfüggvény értéke 0.

A határesetekben bármely érintett eset logikáját követhetjük. Ezek alapján az eloszlásfüggvény felírható:

$$F(x,y) = \begin{cases} 1, & \text{ha} \quad x,y > 1; \\ x, & \text{ha} \quad 0 < x \le 1, \quad y > 1; \\ y, & \text{ha} \quad 0 < y \le 1, \quad x > 1; \\ xy, & \text{ha} \quad 0 < x,y \le 1; \\ 0 & \text{egyébként}. \end{cases}$$

Ezzel az (X,Y) valószínűségi vektorváltozó eloszlásfüggvényét meghatároztuk. Ez a függvény mindegyik változójában monoton növekvő, ezen kívül akkor lesz 1, ha mindkét változó elég nagy. Viszont amint akár csak az egyik változót 0 alá visszük, az eloszlásfüggvény értéke is 0 lesz. Az eloszlásfüggvény nem csak ott fog változni, ahol a vektorváltozó értékeket felvehet: akkor is változik, ha például $x > 1$, y pedig 0 és 1 között mozog. Pedig a valószínűségi változónk nem vehet fel ezen a tartományon értéket. ⇐14.4. feladat

14-3. önálló feladat: Vegyük a koordinátasíkon a $(0,0)$, $(0,1)$ és az $(1,0)$ pontok által meghatározott *háromszöget*. Válasszunk a háromszögön véletlenszerűen egy pontot. Legyen X a választott pont x, és Y az y koordinátája. Adja meg az (X,Y) valószínűségi vektorváltozó eloszlásfüggvényét!

A valószínűségi vektorváltozó eloszlásfüggvénye magában foglalja az egyes komponensek leírásához szükséges információkat. Az eloszlásfüggvény segítségével meghatározható az egyes komponensek eloszlásfüggvénye:

14.2. tétel: Az X és az Y valószínűségi változó F_X illetve F_Y eloszlásfüggvénye felírható az (X,Y) valószínűségi vektorváltozó eloszlásfüggvényének segítségével az alábbi formákban:

$$F_X(x) = \lim_{y\to\infty} F(x,y)$$

$$F_Y(y) = \lim_{x\to\infty} F(x,y)$$

Bizonyítás: Amennyiben csak X eloszlását vizsgáljuk, Y értéke bármennyi lehet, azaz $Y < \infty$. Így az eloszlásfüggvény definícióját alkalmazva

$$F_X(x) = P(X < x) = P(X < x, Y < \infty) = \lim_{y\to\infty} F(x,y)\,.$$

A másik eloszlásfüggvényre hasonlóan megy a bizonyítás. □

A fenti tételből adódóan definiálhatjuk a perem-eloszlásfüggvény fogalmát:

14.7. definíció: **Perem-eloszlásfüggvény**
Az (X,Y) valószínűségi vektorváltozó első (második) komponensének eloszlásfüggvényét az X-hez (Y-hoz) tartozó perem-eloszlásfüggvénynek nevezzük, és F_X-szel (F_Y-nal) jelöljük.

Az együttes eloszlásfüggvény rendelkezik néhány érdekes tulajdonsággal, melyek közvetlen következményei a definíciónak. Például az előző tétel alapján a függvény tartalmazza mindegyik komponensének eloszlásfüggvényét. Beláthatjuk azt is, hogy az együttes eloszlásfüggvény határértéke nulla, ha valamelyik változója $-\infty$-hez tart. Az alábbiak szerint gondolkodjunk:

$$F(x,y) = P(X < x, Y < y) \le P(X < x, Y < \infty) = F_X(x),$$

tehát bármi is y értéke,

$$F(x,y) \le F_X(x)\,.$$

Mivel $F_X(x)$ határértéke $x \to -\infty$-ben nulla, $F(x,y)$ értékét „maga alatt tolva” szükségszerűen tetszőleges y-ra

$$\lim_{x\to-\infty} F(x,y) = 0\,.$$

Tehát akár ha az egyik változója is $-\infty$-be tart, $F(x,y)$ a 0-hoz fog közelíteni.

Foglaljuk össze az együttes eloszlásfüggvény tulajdonságait.

14.3. tétel: Tetszőleges kétdimenziós (X,Y) valószínűségi vektorváltozó F eloszlásfüggvénye rendelkezik az alábbi tulajdonságokkal:

1. $0 \leq F(x,y) \leq 1$.
2. Mindkét változójában monoton növő, azaz

$$\text{ha } x_1 < x_2, \text{ akkor } F(x_1,b) \leq F(x_2,b)$$
$$\text{ha } y_1 < y_2, \text{ akkor } F(a,y_1) \leq F(a,y_2)\,.$$

3. Értéke nullához tart, ha bármelyik változója $-\infty$-hez tart, azaz

$$\lim_{x\to-\infty} F(x,y) = 0 \quad \text{és} \quad \lim_{y\to-\infty} F(x,y) = 0\,.$$

4. Értéke egyhez tart, ha mindkét változója ∞-hez tart, azaz

$$\lim_{x\to\infty}\lim_{y\to\infty} F(x,y) = \lim_{y\to\infty}\lim_{x\to\infty} F(x,y) = 1\,.$$

5. Ha az egyik változóját rögzítjük és a másik ∞-hez tart, akkor az értéke a rögzített változó által meghatározott perem-eloszlásfüggvény értékéhez tart, azaz

$$\lim_{x\to\infty} F(x,y) = F_Y(y) \quad \text{és} \quad \lim_{y\to\infty} F(x,y) = F_X(x)\,.$$

14-4. önálló feladat: Ellenőrizze az eddig kiszámolt eloszlásfüggvényeken a fenti tulajdonságokat!

14.5. feladat. Vegyük a koordinátasíkon a (0,0), (0,1), (1,1) és az (1,0) pontok által meghatározott négyzetet. Válasszunk a négyzeten véletlenszerűen egy pontot. Legyen X a választott pont x, és Y az y koordinátája. Határozzuk meg X és Y eloszlásfüggvényét úgy, hogy az (X,Y) valószínűségi vektorváltozó együttes eloszlásfüggvényéből a perem-eloszlásfüggvényeit kiszámítjuk.

Megoldás: Ahhoz, hogy az X-hez tartozó peremeloszlást meghatározzuk, meg kell nézni mi történik az együttes eloszlásfüggvénnyel, ha az y változó ∞-be tart:

$$F_X(x) = \lim_{y\to\infty} F(x,y) = \lim_{y\to\infty} \begin{cases} 1, & \text{ha} \quad x,y > 1; \\ x, & \text{ha} \quad 0 < x \le 1, \quad y > 1; \\ y, & \text{ha} \quad 0 < y \le 1, \quad x > 1; \\ xy, & \text{ha} \quad 0 < x,y \le 1; \\ 0 & \text{egyébként}\,. \end{cases}$$

Mivel $y \to \infty$, y nem lehet 0 és 1 között, azaz a harmadik és a negyedik sorban leírt eset eset nem állhat fenn. A többi eset viszont nem függ y értékétől, így:

$$F_X(x) = \begin{cases} 1, & \text{ha} \quad x > 1; \\ x, & \text{ha} \quad 0 < x \le 1; \\ 0 & \text{egyébként}\,. \end{cases}$$

Hasonlóan az Y eloszlásfüggvényéhez, a közös eloszlásfüggvény x változója ∞-be tart, azaz nem lehet 0 és 1 között, amit a második és a negyedik sor ír le. A többi sor értéke nem függ x-től, így:

$$F_Y(y) = \begin{cases} 1, & \text{ha} \quad y > 1; \\ y, & \text{ha} \quad 0 < y \le 1; \\ 0 & \text{egyébként}\,. \end{cases}$$

Ezzel mind az X, mind az Y eloszlásfüggvényét meghatároztuk.

⇐14.5. feladat

Egyetlen valószínűségi változó esetén az $\{a \le X < b\}$ esemény valószínűsége $F(b) - F(a)$, vagyis az adott intervallumba esés valószínűsége meghatározható az eloszlásfüggvénynek az intervallum végpontjaiban felvett értékeiből. Ennek megfelelően két valószínűségi változó esetén is megmondhatjuk az együttes eloszlásfüggvény segítségével, mekkora annak a valószínűsége, hogy (X,Y) egy adott téglalapba esik. Az együttes eloszlásfüggvénynek a téglalap csúcsaiban felvett értékeit kell ehhez felhasználnunk az alábbi tétel szerint:

14.4. tétel: Legyen az $F(x,y)$ az X és Y valószínűségi változók együttes eloszlásfüggvénye. Ekkor

$$P(a_1 \le X < a_2\,, b_1 \le Y < b_2) = F(a_2,b_2) - F(a_2,b_1) - F(a_1,b_2) + F(a_1,b_1)\,.$$

Bizonyítás: A bizonyításhoz első lépésként az $\{X < a_2, Y < b_2\}$ eseményt fogjuk diszjunkt események összegére bontani:

$$\{X < a_2,Y < b_2\} = \{a_1 \le X < a_2,b_1 \le Y < b_2\} + \{X < a_1,Y < b_2\} + \{a_1 \le X < a_2,Y < b_1\}\ .$$

Az utolsó tagot még tovább lehet bontani, ezzel:

$$= \{a_1 \le X < a_2,b_1 \le Y < b_2\} + \{X < a_1,Y < b_2\} + \{X < a_2,Y < b_1\} - \{X < a_1,Y < b_1\}\ .$$

A valószínűségszámítás 3. axiómája szerint diszjunkt események összegének valószínűsége megegyezik a valószínűségeik összegével. A fent szereplő események valószínűsége kifejezhető az együttes eloszlásfüggvénnyel:

$$\begin{aligned} P(X < a_2,Y < b_2) &= F(a_2,b_2), & \qquad P(X < a_1,Y < b_2) &= F(a_1,b_2),\\ P(X < a_2,Y < b_1) &= F(a_2,b_1), & \qquad P(X < a_1,Y < b_1) &= F(a_1,b_1)\,. \end{aligned}$$

Rendezés után azt kapjuk a valószínűségre, hogy

$$\begin{aligned} &P(a_1 \le X < a_2\,, b_1 \le Y < b_2) = \\ &\qquad = P(X < a_2,Y < b_2) - P(X < a_1,Y < b_2) - P(X < a_2,Y < b_1) + P(X < a_1,Y < b_1) = \\ &\qquad = F(a_2,b_2) - F(a_1,b_2) - F(a_2,b_1) + F(a_1,b_1)\,. \end{aligned}$$

Aktivitás: Vegyen elő négy színes (pl. két piros és két kék) papírlapot, és illessze oda egy papírra felrajzolt téglalap négy sarkához úgy, hogy azok a négy eloszlásfüggvény értéket szimbolizálják. A piros lapok a tételnek megfelelően a hozzáadott, a kékek a kivont értékeket szimbolizálják.

14.6. feladat. Vegyük a koordinátasíkon a $(0{,}0)$, $(0{,}1)$, $(1{,}1)$ és az $(1{,}0)$ pontok által meghatározott négyzetet. Válasszunk a négyzeten véletlenszerűen egy pontot. Legyen X a választott pont x, és Y az y koordinátája.

a) Számítsuk ki annak a valószínűségét, hogy mindkét koordináta kisebb, mint $1/3$!

b) Feltéve, hogy X nem kisebb $1/2$-nél, mi a valószínűsége annak, hogy Y viszont kisebb mint $1/2$?

Megoldás:

a) Az első kérdés megoldásához a $P(X < 1/3, Y < 1/3)$ valószínűséget kell kiszámolni. Ezt a valószínűséget az együttes eloszlásfüggvény definíciójának közvetlen alkalmazásával meghatározhatjuk:

$$P\left(X < \frac{1}{3}, Y < \frac{1}{3}\right) = F\left(\frac{1}{3}, \frac{1}{3}\right) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}.$$

Tehát annak a valószínűsége, hogy mindkét valószínűségi változó $1/3$ alatt lesz, $1/9$ lesz. Épp a bal alsó kilencedet fogja kimetszeni a negyedsíkunk a négyzetből.

b) A második kérdés megválaszolásához a $P(Y < 1/2 | X \geq 1/2)$ valószínűséget kell meghatároznunk. Alkalmazzuk feltételes valószínűség definícióját:

$$P\left(Y < \frac{1}{2} \middle| X \geq \frac{1}{2}\right) = \frac{P\left(X \geq \frac{1}{2}, \text{ és } Y < \frac{1}{2}\right)}{P\left(X \geq \frac{1}{2}\right)}.$$

Ezek után használjuk ki, hogy a két valószínűségi változó folytonos, továbbá hogy 0 és 1 közötti értékeket vehet fel:

$$\frac{P(X \geq \frac{1}{2}, Y < \frac{1}{2})}{P(X \geq \frac{1}{2})} = \frac{P(\frac{1}{2} \leq X < 1 \text{ és } 0 \leq Y < \frac{1}{2})}{P(X \geq \frac{1}{2})}.$$

Az előző tétel alapján a fenti kifejezés az eloszlásfüggvények segítségével felírható:

$$\frac{P(\frac{1}{2} \leq X < 1 \text{ és } 0 \leq Y < \frac{1}{2})}{P(X \geq \frac{1}{2})} = \frac{F(1,\frac{1}{2}) - F(1,0) - F(\frac{1}{2},\frac{1}{2}) - F(\frac{1}{2},0)}{1 - F_X(\frac{1}{2})} =$$

$$= \frac{\frac{1}{2} - 0 - \frac{1}{2} \cdot \frac{1}{2} - 0}{1 - \frac{1}{2}} = \frac{1}{2}.$$

Így amennyiben tudjuk, hogy X nem kisebb, mint $1/2$, annak a valószínűsége, hogy Y viszont kisebb lesz, épp $1/2$.

14-5. önálló feladat: Számolja ki annak a valószínűségét, hogy $Y < 1/2$, és hasonlítsuk össze az előbbi eredménnyel!

⇐14.6. feladat

Leckezáró önellenőrző kérdéssor

1. Két valószínűségi változó együttes eloszlásfüggvényét vizsgáljuk. Milyen értékhez fog tartani a függvény, ha mindegyik változójával $-\infty$ felé tartunk?

- ☐ $-\infty$
- ☐ 0
- ☐ a közös perem-eloszlásfüggvényhez
- ☐ 1/2

2. Az X és Y folytonos valószínűségi változók együttes eloszlásfüggvényének egyik perem-eloszlásfüggvényét vizsgáljuk. Hány változós ez a függvény?

- ☐ 4
- ☐ 3
- ☐ 2
- ☐ 1

3. Vegyük a koordinátasíkon a (0,0), (0,1), (1,1) és az (1,0) pontok által meghatározott négyzetet. Válasszunk a négyzeten véletlenszerűen egy pontot. Legyen X a választott pont x, és Y az y koordinátája. Mennyi lesz a $P(X = 1/2|Y = 1/4)$ valószínűség értéke?

- ☐ 0
- ☐ 1
- ☐ Nem kiszámítható.
- ☐ 2/3

4. Vegyük a koordinátasíkon a (0,0), (0,1) és az (1,0) pontok által meghatározott háromszöget. Válasszunk a háromszögön véletlenszerűen egy pontot. Legyen X a választott pont x, és Y az y koordinátája. Mennyi lesz a $P(X > 1/2|Y > 1/2)$ valószínűség értéke?

- ☐ 0
- ☐ 1
- ☐ Nem kiszámítható.
- ☐ 2/3

5. Vegyük a koordinátasíkon az (1,1), (0,1) és az (1,0) pontok által meghatározott háromszöget. Válasszunk a háromszögön véletlenszerűen egy pontot. Legyen X a választott pont x, és Y az y koordinátája. Mennyi lesz a $P(X > 1/2|Y > 1/2)$ valószínűség értéke?

- ☐ 0
- ☐ 1
- ☐ Nem kiszámítható.
- ☐ 2/3

21. LECKE

Együttes sűrűségfüggvény és függetlenség

14.4. Két folytonos valószínűségi változó együttes sűrűségfüggvénye

Bár az együttes eloszlásfüggvényt elég egyszerűen definiáljuk mind folytonos, mind diszkrét esetben, mégis csak ritkán használjuk közvetlenül valószínűségek meghatározására. Diszkrét esetben az együttes eloszlás átláthatóbb és jobban alkalmazható. Folytonos esetben viszont, amennyiben nem téglalap alakú tartomány valószínűségére vagyunk kíváncsiak, az előző tétel keveset segít. Már ebben a viszonylag egyszerű esetben is elég összetetten fejezhető ki a valószínűség az együttes eloszlásfüggvénnyel, de ha a tartomány bonyolultabb, akkor az eloszlásfüggvénnyel történő megadás már technikailag nagyságrendekkel körülményesebb.

Folytonos valószínűségi vektorváltozó esetén inkább az együttes sűrűségfüggvényt használjuk, mely az egyváltozós esethez hasonlóan definiálható:

14.8. definíció:
Az (X,Y) valószínűségi vektorváltozót folytonos eloszlásúnak nevezzük, ha van olyan f függvény, melyre

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(t,s)\,\mathrm{d}t\,\mathrm{d}s\,.$$

Miként az egyváltozós esetben, itt is igaz az, hogy ha van ilyen tulajdonságú f függvény, akkor végtelen sok van. Itt is célszerű olyat választani, ami a lehető legkevesebb helyen nem folytonos. Abszolút folytonos függvény majdnem mindenütt differenciálható és egyenlő a deriváltjának határozatlan integráljával, így abszolút folytonos $F(x,y)$ esetén az $f(x,y) = \dfrac{\partial^2 F(x,y)}{\partial x \partial y}$ függvény megfelel a fenti követelményeknek. Azokon a helyeken pedig, ahol $\dfrac{\partial^2 F(x,y)}{\partial x \partial y}$ nem létezik, válasszuk $f(x,y)$ értékét nullának (ezek az integrál értékét nem fogják befolyásolni).

14.9. definíció: Valószínűségi vektorváltozó sűrűségfüggvénye
Legyen az (X,Y) folytonos eloszlású valószínűségi vektorváltozó eloszlásfüggvénye $F(x,y)$. Azokon a helyeken, ahol a parciális derivált létezik legyen

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

a többi helyen pedig legyen $f(x,y) = 0$. Az így definiált $f(x,y)$ függvényt az (X,Y) valószínűségi vektorváltozó sűrűségfüggvényének (együttes sűrűségfüggvényének) nevezzük.

A valószínűségszámítás axiómáiból és az együttes sűrűségfüggvény definíciójából azonnal következnek az együttes sűrűségfüggvény egyszerűbb tulajdonságai:

Aktivitás: Keressen meg az internet segítségével a többváltozós normális eloszlást. Keressünk olyan ábrákat is, melyek az együttes sűrűségfüggvényt ábrázolják!

14.5. tétel: Amennyiben a folytonos eloszlású (X,Y) valószínűségi vektorváltozónak létezik f sűrűségfüggvénye, az az alábbi tulajdonságokkal rendelkezik:

1. Értékei nem negatívak, azaz $f(x,y) \geq 0$.
2. Ha mindegyik változójában $-\infty$-től ∞-ig integrálunk, akkor az integrál értéke egy:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y\,\mathrm{d}x = 1\,.$$

Tekintsünk egy kísérletet, melynek kimenetele az (X,Y) valószínűségi vektorváltozó egy értéke, azaz egy számpár. Ekkor minden egyes eseménynek megfeleltethető a sík pontjainak olyan halmaza (A), melynél a pontok koordinátái az eseménynek megfelelő számpárok. Ilyenkor az esemény valószínűségét az együttes sűrűségfüggvény A tartomány feletti integrálásával kaphatjuk meg.

14.6. tétel: Legyen az X és Y valószínűségi változók együttes sűrűségfüggvénye f. Ekkor annak valószínűsége, hogy (X,Y) értéke egy adott A tartományba esik:

$$P(A) = \iint_A f(x,y)\,\mathrm{d}x\,\mathrm{d}y\,.$$

14-6. önálló feladat: Legyen X egy véletlenszerűen választott valós szám a $[0,\ 1]$ intervallumból, valamint $Y = 1 - X$. Lássa be, hogy az $(X,\ Y)$ valószínűségi vektorváltozónak nincs sűrűségfüggvénye!

14.7. feladat. Vegyük a koordinátasíkon a $(0,0)$, $(0,1)$, $(1,1)$ és az $(1,0)$ pontok által meghatározott négyzetet. Válasszunk a négyzeten véletlenszerűen egy pontot. Legyen X a választott pont x, és Y az y koordinátája.

a) Adjuk meg az (X,Y) valószínűségi vektorváltozó sűrűségfüggvényét, majd

b) számoljuk ki annak a valószínűségét, hogy az (X,Y) értékpár az origótól maximum egy egység távolságra van!

Megoldás:

a) A sűrűségfüggvény meghatározásához a már kiszámolt eloszlásfüggvényt deriváljuk. Az eloszlásfüggvényünk úgy van megadva, hogy x és y értékeitől függően más-más kifejezés segítségével adjuk meg az értéket. Ilyen esetben a deriválást úgy végezzük el, hogy az egyes eseteket külön-külön, azaz soronként deriváljuk.

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{\partial}{\partial x}\left[\frac{\partial}{\partial y} F(x,y)\right]$$

Először az y szerinti deriválást végezzük el ott, ahol a függvény deriválható:

$$\frac{\partial}{\partial y}F(x,y)=\frac{\partial}{\partial y}\begin{cases}1, & \text{ha} \quad x,y>1;\\ x, & \text{ha} \quad 0<x\leq 1,\quad y>1;\\ y, & \text{ha} \quad 0<y\leq 1,\quad x>1;\\ xy, & \text{ha} \quad 0<x,y\leq 1;\\ 0 & \text{egyébként};\end{cases}=\begin{cases}0, & \text{ha} \quad x,y>1;\\ 0, & \text{ha} \quad 0<x<1,\quad y>1;\\ 1, & \text{ha} \quad 0<y<1,\quad x>1;\\ x, & \text{ha} \quad 0<x,y<1;\\ 0 & \text{egyébként}.\end{cases}$$

Az első két 0-ás sort elhagyhatjuk, így

$$\frac{\partial}{\partial y}F(x,y)=\begin{cases}1, & \text{ha} \quad 0<y\leq 1,\quad x>1;\\ x, & \text{ha} \quad 0<x,y\leq 1;\\ 0 & \text{egyébként}.\end{cases}$$

Végezzük el most az x szerinti deriválást is:

$$\frac{\partial}{\partial x}\left[\frac{\partial}{\partial y}F(x,y)\right]=\frac{\partial}{\partial x}\begin{cases}1, & \text{ha} \quad 0<y<1,\quad x>1;\\ x, & \text{ha} \quad 0<x,y<1;\\ 0 & \text{egyébként};\end{cases}=\begin{cases}0, & \text{ha} \quad 0<y<1,\quad x>1;\\ 1, & \text{ha} \quad 0<x,y<1;\\ 0 & \text{egyébként}.\end{cases}$$

Tehát összességében:

$$f(x,y)=\frac{\partial}{\partial x}\left[\frac{\partial}{\partial y}F(x,y)\right]=\begin{cases}1, & \text{ha} \quad 0<x,y<1;\\ 0 & \text{egyébként}.\end{cases}$$

Láthatjuk, hogy az első sort ismét elhagyhatjuk. Az eloszlásfüggvény öt esete helyett itt két eset van. Figyeljük meg, hogy a sűrűségfüggvény abban a környezetben vesz fel nem nulla értéket, ahol a vektorváltozó ténylegesen értéket vehet fel. Vegyük észre, hogy egy egyenletes eloszlás sűrűségfüggvényéről van szó, így a függvény értéke konstans azon a területen, ahol a valószínűségi változó értéket vehet fel. A felvett konstans épp a terület nagyságának reciproka lesz. A $0<x,y<1$ feltétel épp egy egység területű négyzetet határoz meg, így a felvett konstans érték is 1 lesz.

b) A valószínűség meghatározásához definiáljuk a nekünk jó tartományt. Legyen az A tartomány azon pontok halmaza, melyek az origótól maximum 1 egységre vannak. Ekkor az A egy origó középpontú, 1 egység sugarú körlap. Annak a valószínűsége, hogy a valószínűségi vektorváltozó ide esik, kiszámolható úgy, hogy a sűrűségfüggvényt ezen a tartományon integráljuk:

$$P(A) = \iint_A f(x,y)\,\mathrm{d}x\mathrm{d}y = \iint_{A'} 1\,\mathrm{d}x\mathrm{d}y\,.$$

Itt az A' tartomány az eredeti kör első koordinátanegyedbeli része. Ez épp egy negyedkör lesz. Az integrálást ezen a negyedkörön kell végrehajtani. Mivel az integrandus a konstans 1 függvény, az integrál értéke ennek a negyedkörnek a területe, azaz $P(A) = \pi/4$.
Tehát annak a valószínűsége, hogy a választott pont 1 egységnél közelebb lesz az origóhoz, körülbelül 0,7845 lesz.

⇐14.7. feladat

Aktivitás: Gondolja át, hogy az előző feladatban keresett valószínűséget miért nem tudjuk az eloszlásfüggvény segítségével közvetlenül kiszámolni.

14-7. önálló feladat: Vegyük a koordinátasíkon a $(0{,}0)$, $(0{,}1)$ és az $(1{,}0)$ pontok által meghatározott *háromszöget*. Válasszunk háromszögön véletlenszerűen egy pontot. Legyen X a választott pont x, és Y az y koordinátája. Határozza meg X és Y közös sűrűségfüggvényét!

A perem-eloszlásfüggvényhez hasonlóan az együttes sűrűségfüggvény segítségével meghatározható az egyes komponensek sűrűségfüggvénye külön-külön. Az így kapott sűrűségfüggvényt perem-sűrűségfüggvényként definiáljuk:

14.10. definíció: **Perem-sűrűségfüggvény**
Az (X,Y) folytonos eloszlású valószínűségi vektorváltozó első (második) komponensének sűrűségfüggvényét az X-hez (Y-hoz) tartozó perem-sűrűségfüggvénynek nevezzük, és f_X-szel (f_Y-nal) jelöljük.

A perem-sűrűségfüggvényt az alábbi tétel segítségével határozhatjuk meg:

14.7. tétel: Ha X és Y együttes sűrűségfüggvénye $f(x,y)$, akkor

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y \quad \text{és} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x\,.$$

Bizonyítás:

A bizonyítást csak az egyik változóra mutatjuk be, a másikra természetesen ugyanúgy működik. Az együttes sűrűségfüggvény definíciója alapján:

$$F_X(x) = \lim_{y\to\infty} F(x,y) = \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} f(s,y)\,\mathrm{d}y \right) \mathrm{d}s\,.$$

Képezzük mindkét oldal x szerinti deriváltját:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y\,. \qquad \square$$

14.8. feladat. Vegyük a koordinátasíkon a $(0,0)$, $(0,1)$, $(1,1)$ és az $(1,0)$ pontok által meghatározott négyzetet. Válasszunk a négyzeten véletlenszerűen egy pontot. Legyen X a választott pont x, és Y az y koordinátája. Az együttes eloszlásfüggvényből kiindulva határozzuk meg a második komponens sűrűségfüggvényét *a)* előbb a perem-eloszlásfüggvény,

b) majd az együttes sűrűségfüggvény segítségével!

Megoldás:

a) Először a perem-eloszlásfüggvény segítségével határozzuk meg a perem-sűrűségfüggvényt. A második komponens eloszlásfüggvényét egy előző feladat során meghatároztuk, ebből deriválás segítségével írhatjuk fel a második komponens sűrűségfüggvényét:

$$f_Y(y) = \frac{\partial}{\partial y} F_Y(y) = \frac{\partial}{\partial y} \begin{cases} 1, & \text{ha} \quad y > 1; \\ y, & \text{ha} \quad 0 < y < 1; \\ 0 & \text{egyébként}; \end{cases} = \begin{cases} 1, & \text{ha } 0 < y < 1; \\ 0 & \text{egyébként}. \end{cases}$$

Látható, hogy az Y eloszlása egyenletes a $[0;1]$ intervallumon, és mind az F_Y eloszlásfüggvény, mind az f_Y sűrűségfüggvény ennek megfelelő.

b) Az együttes sűrűségfüggvény használata során Y peremsűrűségét meghatározhatjuk úgy, hogy a többi – jelen esetben az X – valószínűségi változóhoz kapcsolt változó (most az x) szerint integráljuk:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x$$

Az $f(x,y)$ együttes sűrűségfüggvény kizárólag akkor vesz fel nem nulla értéket, ha $0 < x < 1$. Ezért az integrálást elég ezen a tartományon elvégezni:

$$\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x = \int_{0}^{1} f(x,y)\,\mathrm{d}x$$

Azt is figyeljük meg, hogy az együttes sűrűségfüggvény csak abban az esetben vesz fel nem nulla értéket, ha $0 < y < 1$:

$$f_Y(y) = \int_0^1 f(x,y)\,\mathrm{d}x = \int_0^1 1\,\mathrm{d}x = [x]_0^1 = 1 \qquad \text{ha } 0 < y < 1\,.$$

Minden más esetben $f(x,y)$ mindenhol 0, így az integrál értéke is 0:

$$f_Y(y) = \int_0^1 f(x,y)\,\mathrm{d}x = \int_0^1 0\,\mathrm{d}x = [0]_0^1 = 0 \qquad \text{ha } y < 0 \text{ vagy } y > 1\,.$$

Így a sűrűségfüggvényt meghatároztuk:

$$f_Y(y) = \begin{cases} 1, & \text{ha } 0 < y < 1; \\ 0 & \text{egyébként}\,. \end{cases}$$

Figyeljük meg, hogy mind a két módszerrel ugyanarra az eredményre jutottunk. Természetesen a deriválási technika egyszerűbbnek és gyorsabbnak bizonyulhat.

⇐14.8. feladat

Aktivitás: Gondolja át, milyen módszerekkel számolhatjuk ki a közös eloszlásfüggvényből a perem-sűrűségfüggvényeket. Hasonlítsa össze a kapott módszereket!

14.5. Folytonos valószínűségi változók függetlensége

Két valószínűségi változót függetlennek tekintünk, ha az egyik értékéről kapott információ alapján nem tudunk meg többet a másik értékével kapcsolatban. Tehát az egyik értékével kapcsolatos események függetlenek a másik értékével kapcsolatos eseményektől. Itt elégséges az eloszlásfüggvényben használt $\{X < x\}$ típusú eseményeket vizsgálni:

14.11. definíció: **Valószínűségi változók függetlensége**
Az X és Y valószínűségi változókat függetlennek nevezzük, ha az $A = \{X < x\}$ és a $B = \{Y < y\}$ események minden x és y esetén függetlenek.

A függetlenség definíciójából azonnal következik az alábbi tétel:

14.8. tétel: Az X és Y valószínűségi változók akkor és csak akkor függetlenek, ha az együttes eloszlásfüggvény az egyes komponensek eloszlásfüggvényeinek a szorzata, azaz

$$F(x,y) = F_X(x) \cdot F_Y(y)\,.$$

Bizonyítás: Legyen x és y tetszőleges. Ekkor X és Y függetlensége esetén:

$$\begin{aligned} F(x,y) &= P(X < x, Y < y) = (\text{mert } X \text{ és } Y \text{ függetlenek}) = P(X < x) \cdot P(Y < y) = \\ &= F_X(x) \cdot F_Y(y)\,. \end{aligned}$$

Visszafelé, amennyiben az eloszlásfüggvény a perem-eloszlásfüggvények szorzataként felírható:

$$P(X < x, Y < y) = F(x,y) = F_X(x) \cdot F_Y(y) = P(X < x) \cdot P(Y < y)\,.$$

$\square$

Hasonló igaz független folytonos valószínűségi változóknál az együttes sűrűségfüggvényre is:

14.9. tétel: AZ X és Y folytonos valószínűségi változók akkor és csak akkor függetlenek, ha az együttes sűrűségfüggvény az egyes komponensek sűrűségfüggvényeinek a szorzata, azaz

$$f(x,y) = f_X(x) \cdot f_Y(y) \,.$$

Bizonyítás: A bizonyításban az előző tételt használjuk fel, valamint azt, hogy az együttes sűrűségfüggvény az együttes eloszlásfüggvényből a változók szerinti parciális deriválással kapható meg:

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{\partial^2 \left[F_X(x) \cdot F_Y(y)\right]}{\partial x \partial y} = \frac{\partial \left[F_X(x) \cdot f_Y(y)\right]}{\partial x} = f_X(x) \cdot f_Y(y) \,.$$

Visszafelé az együttes sűrűségfüggvény definícióját használhatjuk:

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(t,s) \, \mathrm{d}t\mathrm{d}s = \int_{-\infty}^{x} \int_{-\infty}^{y} f_X(t) f_Y(s) \, \mathrm{d}t\mathrm{d}s =$$

$$= \int_{-\infty}^{x} f_X(t) \, \mathrm{d}t \cdot \int_{-\infty}^{y} f_Y(s) \, \mathrm{d}s = F_X(x) \cdot F_Y(y) \,.$$

□

14.9. feladat. Vegyük a koordinátasíkon a $(0,0)$, $(0,1)$, $(1,1)$ és az $(1,0)$ pontok által meghatározott négyzetet. Válasszunk a négyzeten véletlenszerűen egy pontot. Legyen X a választott pont x, és Y az y koordinátája. Igaz-e, hogy X és Y független?

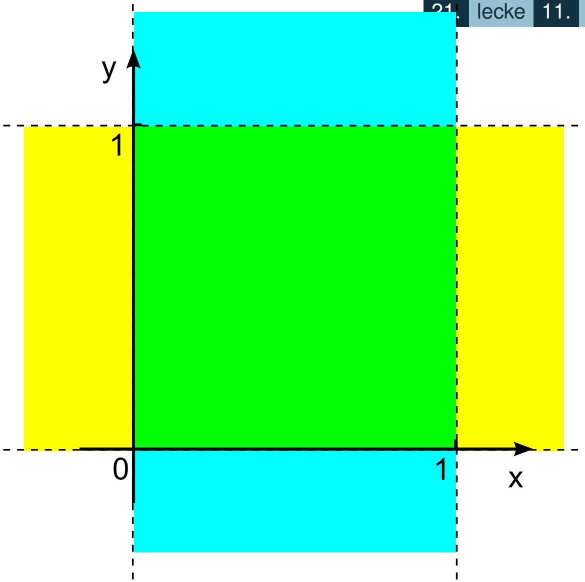
Megoldás:

A függetlenség belátásához elég megvizsgálni azt, hogy előállítható-e az együttes sűrűségfüggvény a perem-sűrűségfüggvények szorzataként:

$$f_X(x)\cdot f_Y(y) =$$
$$= \begin{cases} 1, & \text{ha } 0<x<1; \\ 0 & \text{egyébként;} \end{cases} \cdot \begin{cases} 1, & \text{ha } 0<y<1; \\ 0 & \text{egyébként;} \end{cases} =$$
$$= \begin{cases} 1, & \text{ha } 0<x<1 \text{ és } 0<y<1; \\ 0 & \text{egyébként;} \end{cases} =$$
$$= f(x,y)\,.$$

Tehát X és Y független.

⇐14.9. feladat



14.3. ábra. *A kék illetve sárga tartomány azokat a területeket jelzi, ahol az $f_X(x)$ illetve az $f_Y(y)$ függvények nem nulla értéket vesznek fel. Szorzatuk így épp a zölddel jelölt közös részen lesz nem nulla.*

14-8. önálló feladat: Vegyük a koordinátasíkon a $(0,0)$, $(0,1)$ és az $(1,0)$ pontok által meghatározott *háromszöget*. Válasszunk háromszögön véletlenszerűen egy pontot. Legyen X a választott pont x, és Y az y koordinátája. Igaz-e, hogy X és Y független?

Leckezáró önellenőrző kérdéssor

1. Két folytonos valószínűségi változó együttes eloszlásfüggvényének mi lesz az értelmezési tartománya, és az értékkészlete?
   - ☐ Az értelmezési tartomány a valós számok, az értékkészlete a $[0;1]$ intervallum.
   - ☐ Az értelmezési tartomány a valós számpárok halmaza, az értékei pozitív valós számok.
   - ☐ Az értelmezési tartomány a valós számpárok halmaza, az értékkészlete a $[0;1]$ intervallum.
   - ☐ Az értelmezési tartomány a valós számpárok halmaza, az értékei nem negatív valós számok.
2. Válasszunk 0 és 1 között véletlenszerűen egy számot. A választott számmal, mint sugárral, rajzoljunk kört. Legyen X a rajzolt kör kerülete, Y pedig a területe. Független lesz-e X és Y?
   - ☐ Igen, mert míg a kerület arányos a sugárral, a terület a sugár négyzetével lesz arányos.
   - ☐ Nem, mert X az Y transzformáltja.
   - ☐ Igen, mert Y az X transzformáltja.
   - ☐ Nem tudjuk, mert az együttes sűrűségfüggvényt nem lehet kiszámítani.
3. Legyen X_1, X_2 és X_3 három folytonos valószínűségi változó. Tudjuk, hogy X_1 és X_2 függetlenek, és azt is, hogy X_2 és X_3 is függetlenek. Igaz-e, hogy ebben az esetben X_1 és X_3 is függetlenek?
   - ☐ Igen, mert a többszörös függetlenség tranzitív módon megerősödik.
   - ☐ Nem, mert a függetlenség függetlensége függőséget indukál.
   - ☐ Igen, például ha a valószínűségi változók három kockadobás eredményei.
   - ☐ Nem, például ha $X_1 = X_3$.
4. Egységnyi hosszúságú szakaszon választunk véletlenszerűen egy pontot, mely a szakaszt két részre osztja. Legyen X az egyik, Y a másik szakasz hossza. Milyen lesz a közös eloszlásfüggvény alakja a $0 < x,y < 1$ tartományban?
   - ☐ $\sqrt{xy}$
   - ☐ $x + y - 1$, ha $x + y > 1$; 0 egyébként
   - ☐ $\sqrt{x^2y^2}$
   - ☐ $(x + y)|x - y|$

22. LECKE

Kettőnél több valószínűségi változó együttes eloszlása

14.6. Több valószínűségi változó együttes eloszlása

Mindegyik valószínűségi vektorváltozó esetén definiálható az együttes eloszlásfüggvény, valamint az egyes valószínűségi változókhoz tartozó perem-eloszlásfüggvény, legyen ez akár folytonos, vagy diszkrét.

14.12. definíció: **Valószínűségi vektorváltozók eloszlásfüggvénye**
Azt a függvényt, amely minden valós $(x_1,x_2,\ldots,x_n)$ szám n-eshez hozzárendeli annak valószínűségét, hogy az

$$X_1 < x_1, \quad X_2 < x_2, \quad \ldots \quad X_n < x_n$$

események együttesen bekövetkeznek, az $(X_1,X_2,\ldots,X_n)$ valószínűségi vektorváltozó eloszlásfüggvényének nevezzük. Azaz

$$F(x_1,x_2,\ldots,x_n) = P(X_1 < x_1, X_2 < x_2, \ldots, X_n < x_n)\,.$$

14.10. feladat. A $[0;2]$ intervallumon három pontot választunk véletlenszerűen. Az (X_1,X_2,X_3) valószínűségi vektorváltozó komponensei legyenek rendre a választott értékek. Írjuk fel a valószínűségi vektorváltozóhoz tartozó együttes eloszlásfüggvényt!

Megoldás: Láthatjuk, hogy a valószínűségi vektorváltozó mindegyik komponense 0 és 2 között veheti fel az értékeit. Ez azt jelenti, hogy azt kell megvizsgálni, mi történik, ha x, y és z értékei ezen intervallumba beleesnek, vagy annak jobb illetve bal oldalán tartózkodnak. Ennek megfelelően kell az $F(x,y,z)$ eloszlásfüggvény felírásához meghatározni annak a valószínűségét, hogy $X_1 < x$ és $X_2 < y$ és $X_3 < z$. A három választott pont egy ponthármast ad, mely a $[0;2]\cdot[0;2]\cdot[0;2]$ kockába fog esni. Azt kell vizsgálni, milyen térfogatba esnek azok a pontok, melyek koordinátáinak mindegyike rendre x, y és z-nél is kisebbek.

a) Mi a helyzet akkor, ha x, y és z közül *bármelyik* 0 alatt van? Ha például $x<0$, akkor annak az az esemény hogy $X_1 < x < 0$, nem következhet be, azaz a valószínűsége 0 lesz, így ekkor $F(x,y,z)=0$.

b) Mi a helyzet, ha x, y és z közül *mindegyik* 2 felett van? Ekkor az $X_1 < 2 < x$, $X_2 < 2 < y$ és $X_3 < 2 < z$ események közül mindegyik biztosan teljesül, tehát $F(x,y,z)=1$.

c) A harmadik eset akkor jön elő, ha az egyik komponens, mondjuk x, 0 és 2 között van, a többi felette. Ekkor $X_2 < 2 < y$ és $X_3 < 2 < z$ mindig teljesül. Annak a valószínűségét kell kiszámolnunk viszont, hogy $X_1 < x$. Mivel a kockán belül a pontot véletlenszerűen választjuk, az esemény valószínűségét úgy tudjuk kiszámolni, hogy az (x,y,z) pontból kiinduló térnyolcad metszetét vizsgáljuk a $[0;2] \cdot [0;2] \cdot [0;2]$ kockával. A térnyolcad azon (a,b,c) pontokat fogja tartalmazni, melyekre $a \leq x$, $b \leq y$ és $c \leq z$. Ebben az esetben a térnyolcad a kockából egy $x \cdot 2 \cdot 2$ térfogatú részt metsz ki, így $F(x,y,z) = (x \cdot 2 \cdot 2)/(2 \cdot 2 \cdot 2) = x/2$ lesz.

d) Ha x, y és z közül ketten, mondjuk x és y lesznek 0 és 2 között, és $z > 2$, akkor az $X_3 < z$ esemény mindig teljesül. Az $X_1 < x$ és az $X_2 < y$ esemény akkor következik be együttesen, ha a választott ponthármas abba a térfogatba kerül, amit az (x,y,z) térnyolcad kimetsz. A kimetszett rész térfogata $x \cdot y \cdot 2$ lesz, így ebben az esetben $F(x,y,z) = (x \cdot y \cdot 2)/(2 \cdot 2 \cdot 2) = xy/4$ lesz.

e) Ha F mindhárom változója 0 és 2 között van, akkor a térnyolcad épp egy $x \cdot y \cdot z$ térfogatú részt vág ki a kockából. Így ebben az esetben $F(x,y,z) = (x \cdot y \cdot z)/(2 \cdot 2 \cdot 2) = xyz/8$ lesz.

Összefoglalva, az eloszlásfüggvényt az alábbi formában írhatjuk fel:

$$F(x,y,z) = \begin{cases} xzy/8, & \text{ha} \quad 0 < x,y,z < 2; \\ xy/4, & \text{ha} \quad 0 < x,y < 2 \quad \text{és} \quad z > 2; \\ xz/4, & \text{ha} \quad 0 < x,z < 2 \quad \text{és} \quad y > 2; \\ yz/4, & \text{ha} \quad 0 < y,z < 2 \quad \text{és} \quad x > 2; \\ x/2, & \text{ha} \quad 0 < x < 2 \quad \text{és} \quad y,z > 2; \\ y/2, & \text{ha} \quad 0 < y < 2 \quad \text{és} \quad x,z > 2; \\ z/2, & \text{ha} \quad 0 < z < 2 \quad \text{és} \quad x,y > 2; \\ 0 & \text{egyébként}. \end{cases}$$

Ezzel az eloszlásfüggvényt megadtuk. Láthatjuk, hogy a valószínűségi vektorváltozó komponenseinek számát növelve az eloszlásfüggvény formája lényegesen összetettebb lehet.

⇐14.10. feladat

Aktivitás: Papír és ceruza, vagy számítógépes szoftver segítségével ábrázolja a fenti öt esetben a kocka és a térnyolcad viszonyát!

14.13. definíció: Perem-eloszlásfüggvény
Az $(X_1,X_2,\ldots,X_n)$ valószínűségi vektorváltozó i-edik komponensének eloszlását az X_i-hez tartozó peremeloszlásnak nevezzük, eloszlásfüggvényét pedig F_i-vel jelöljük.

14-9. önálló feladat: Gondolja végig, miképp kaphatjuk meg az együttes eloszlásfüggvényből az egyes perem-eloszlásokat!

Sűrűségfüggvényről és perem-sűrűségfüggvényről csak folytonos vektorváltozó esetén beszélünk:

14.14. definíció: Valószínűségi vektorváltozó sűrűségfüggvénye
Amennyiben az $(X_1,X_2,\ldots,X_n)$ valószínűségi vektorváltozó eloszlásfüggvénye integrálfüggvény, azaz van olyan f függvény, amelyre

$$F(x_1,x_2,\ldots,x_n) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\ldots\int_{-\infty}^{x_n} f(t_1,t_2,\ldots,t_n)\,\mathrm{d}t_n\ldots\,\mathrm{d}t_2\,\mathrm{d}t_1,$$

akkor ezt az f függvényt a valószínűségi vektorváltozó sűrűségfüggvényének nevezzük. A választott sűrűségfüggvénynek az alábbi két kitételt kell teljesítenie:

1. Csak ott lehet nem folytonos, ahol a fenti feltételt teljesítő függvénysereg egyike sem.
2. Ahol a fenti feltételt teljesítő függvénysereg egyike se folytonos, ott 0 értéket vesz fel.

14-10. önálló feladat: A $[0;2]$ intervallumon három pontot választunk véletlenszerűen. Az (X_1,X_2,X_3) valószínűségi vektorváltozó komponensei legyenek rendre a választott értékek. A kiszámolt eloszlásfüggvény segítségével határozza meg a valószínűségi vektorváltozó sűrűségfüggvényét!

14.15. definíció: Perem-sűrűségfüggvény
Az $(X_1,X_2,\ldots,X_n)$ folytonos eloszlású valószínűségi vektorváltozó i-edik komponensének sűrűségfüggvényét az X_i-hez tartozó perem-sűrűségfüggvénynek nevezzük, és f_i-vel jelöljük.

14-11. önálló feladat: Határozza meg az előző feladatban szereplő valószínűségi vektorváltozóhoz tartozó perem-sűrűségfüggvényeket. A feladatot oldja meg úgy is, hogy

a) a perem-eloszlásfüggvények differenciálhányadosait számoljuk ki,

b) illetve az együttes sűrűségfüggvényt integráljuk!

A feladat megoldásához kövesse a két valószínűségi változó esetél leírtakat!

Több valószínűségi változó függetlenségét az alábbiak szerint definiáljuk:

14.16. definíció: Valószínűségi változók függetlensége
Az $X_1,X_2,\ldots,X_n$ valószínűségi változókat függetleneknek nevezünk, ha a $X_1 < x_1$, $X_2 < x_2$, …, $X_n < x_n$ események minden $x_1,x_2,\ldots,x_n$ esetén függetlenek.

A függetlenség definíciójából közvetlenül adódik az alábbi két állítás.

14.10. tétel: Az $(X_1,X_2,\ldots,X_n)$ valószínűségi vektorváltozó komponensei akkor és csak akkor függetlenek, ha az együttes eloszlásfüggvény az egyes komponensek eloszlásfüggvényeinek a szorzata, azaz

$$F(x_1,x_2,\ldots,x_n) = F_1(x_1)\cdot F_2(x_2)\cdot\ldots\cdot F_n(x_n)\,.$$

14.11. tétel: A folytonos eloszlású $(X_1,X_2,\ldots,X_n)$ valószínűségi vektorváltozó komponensei akkor és csak akkor függetlenek, ha az együttes sűrűségfüggvény az egyes komponensek sűrűségfüggvényeinek a szorzata, azaz

$$f(x_1,x_2,\ldots,x_n) = f_1(x_1)\cdot f_2(x_2)\cdot\ldots\cdot f_n(x_n)\,.$$

14-12. önálló feladat: A $[0;2]$ intervallumon választunk három pontot. A választott pontokhoz tartozó értékeket vegye fel rendre az X_1, X_2 és X_3 valószínűségi változó. Ellenőrizze a leírt két módszerrel, hogy a valószínűségi változók függetlenek-e:

a) Igaz-e, hogy a közös eloszlásfüggvény a perem-eloszlásfüggvények szorzata?

b) Igaz-e, hogy a közös sűrűségfüggvény a perem-sűrűségfüggvények szorzata?

Leckezáró önellenőrző kérdéssor

1. Egy valószínűségi vektorváltozó az X_1, X_2, Y_1 és az Y_2 komponensekből áll. Hány változós lesz ekkor a közös eloszlásfüggvény?

☐ 2 ☐ 4 ☐ 3 ☐ a komponensektől függ

2. Egy valószínűségi vektorváltozó az X_1, X_2, Y_1 és az Y_2 komponensekből áll. A közös sűrűségfüggvény létezik, ebből szeretnénk a harmadik komponens perem-sűrűségfüggvényét előállítani. Hány integrálást kell ehhez végeznünk?

☐ 2 ☐ 4 ☐ 3 ☐ a komponensektől függ

3. Egy többkomponensű valószínűségi vektorváltozó komponensei függetlenek. Melyik együttes függvény írható fel a peremfüggvényeinek szorzataként, ha létezik?

☐ az eloszlás ☐ az eloszlásfüggvény ☐ a sűrűségfüggvény ☐ mindhárom

4. Egységnyi oldalhosszúságú kockában választunk véletlenszerűen egy pontot. Legyen a $(X_1, X_2, \ldots, X_6)$ valószínűségi vektorváltozó komponensei rendre a pont távolságai az oldallapoktól. Melyik állítás igaz a következők közül?

☐ A komponensek függetlenek, mert az együttes eloszlásfüggvény felírható a komponensek eloszlásfüggvényeinek szorzataként.

☐ A komponensek nem függetlenek, mert az együttes eloszlásfüggvény nem írható fel a komponensek eloszlásfüggvényeinek szorzataként.

☐ A komponensek függetlenek, mert az együttes sűrűségfüggvény felírható a komponensek sűrűségfüggvényeinek szorzataként.

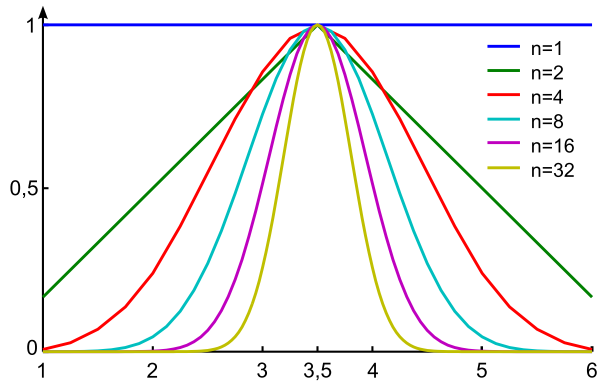
☐ A komponensek nem függetlenek, mert az együttes sűrűségfüggvény nem írható fel a komponensek sűrűségfüggvényeinek szorzataként.

23. LECKE

Valószínűségi változók szorzatának, összegének, átlagának várható értéke

15. Valószínűségi változók összege és átlaga

Sokféle kérdésre választ kaphatunk az $Y_n = X_1 + X_2 + \ldots + X_n$ valószínűségi változó vizsgálatával. Mint látható, az Y_n valószínűségi változó n darab valószínűségi változó összegeként áll elő, így a teljes valószínűségi leíráshoz szükségünk lenne $X_1, X_2, \ldots, X_n$ együttes eloszlására (eloszlásfüggvényére). Az alkalmazásokban azonban gyakorta találkozunk olyan esetekkel, amikor nincs szükség az általános n dimenziós valószínűségi modell vizsgálatára, annak néhány jellemzője (várható értéke, szórása) felhasználásával már kielégítő választ kaphatunk a kérdésekre. A következő néhány leckében valószínűségi változók összegének várható értékéről, szórásáról, sűrűségfüggvényéről lesz szó. Ezek után pedig látni fogjuk, hogy valószínűségi változók összegének eloszlása bizonyos feltételek mellett normális eloszlással közelíthető.



15.1. ábra. *Hatoldalú dobókockával dobva, a dobások átlaga a dobás várható értékéhez közelít. Minél többet dobunk, annál kevésbé ingadozik az átlag értéke* $3{,}5$ *körül.*

A statisztikai alkalmazásokban elengedhetetlenül fontos az átlag (számtani közép) használata. Független, azonos várható értékkel és szórással rendelkező valószínűségi változók esetén megmutatható, hogy az átlag várható értéke ugyanannyi, mint egyetlen valószínűségi változó várható értéke. Ha viszont a szórást vizsgáljuk, az drasztikusan (a $\sqrt{n}$-ed részére) csökken egyetlen valószínűségi változó szórásához képest. A méréstechnikában ez az egyik alapvető eszköz. A pontatlanságokból, és az esetleges precizitás hiányából adódóan a tényleges mért érték ingadozást mutat. Azonban ha a mérést többször elvégezzük, és az eredményeket átlagoljuk, az átlag ingadozása már jóval kisebb lesz. Ez azt jelenti, hogy minél több valószínűségi változót átlagolunk, a kapott átlagértékek annál jobban koncentrálódnak a várható érték körül.

15.1. Valószínűségi változók összegének és átlagának várható értéke

15.1. tétel: Ha az X és Y valószínűségi változók várható értéke létezik, akkor létezik az összegük várható értéke is, és

$$E(X+Y)=E(X)+E(Y)\,.$$

Bizonyítás: A bizonyítást külön-külön végezzük el diszkrét, illetve folytonos esetre.

1. Diszkrét esetben:

$$\begin{aligned}E(X+Y)&=\sum_i\sum_k(x_i+y_k)\cdot p_{ik}=\sum_i\sum_k x_i\cdot p_{ik}+\sum_i\sum_k y_k\cdot p_{ik}=\\&=\sum_i x_i\sum_k p_{ik}+\sum_k y_k\sum_i p_{ik}=\sum_i x_i\cdot p_i+\sum_k y_k\cdot p_k=E(X)+E(Y)\,.\end{aligned}$$

2. Folytonos esetben:

$$E(X+Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)\cdot f(x,y)\,\mathrm{d}x\,\mathrm{d}y=$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x\cdot f(x,y)\,\mathrm{d}x\,\mathrm{d}y+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}y\cdot f(x,y)\,\mathrm{d}x\,\mathrm{d}y=$$

$$=\int_{-\infty}^{\infty}x\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}y\,\mathrm{d}x+\int_{-\infty}^{\infty}y\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}x\,\mathrm{d}y=\int_{-\infty}^{\infty}x\cdot f_X(x)\,\mathrm{d}x+\int_{-\infty}^{\infty}y\cdot f_Y(y)\,\mathrm{d}y=$$

$$=E(X)+E(Y)\,.$$

$\square$

15.1. feladat. Hatoldalú dobókockával dobunk. Legyen X a dobott szám 2-vel, Y pedig a 4-gyel való osztás utáni maradéka.

a) Írjuk fel X és Y eloszlását!

b) Számoljuk ki X és Y várható értékét!

c) Írjuk fel X és Y együttes eloszlását!

d) Írjuk fel $X+Y$ eloszlását, és számoljuk ki a várható értékét!

Megoldás:

a) Először X eloszlását határozzuk meg. X értéke akkor lesz 0, ha párosat dobunk. Az érték 1 lesz, ha a dobás páratlan. Így az eloszlás könnyen felírható:

$$X : \begin{cases} 0 & 1 \\ 1/2 & 1/2 \end{cases}$$

Az Y eloszlásának felírásához vegyük észre, hogy Y értékei 0, 1, 2, és 3 lehetnek. Vegyük végig a hat dobható számot. Két-két esetben lesz Y értéke 1 illetve 2, egy-egy esetben veszi pedig fel a 0 és a 3 értékeket. Így az eloszlás:

$$Y : \begin{cases} 0 & 1 & 2 & 3 \\ 1/6 & 1/3 & 1/3 & 1/6 \end{cases}$$

b) Az eloszlások alapján, definíció szerint számolhatjuk ki X és Y várható értékét:

$$E(X) = \sum_i x_i p_i = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2},$$

valamint:

$$E(Y) = \sum_k y_k p_k = 0 \cdot \frac{1}{6} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{6} = \frac{3}{2}.$$

Tehát az X valószínűségi változó várható értéke $1/2$, az Y-é pedig $3/2$.

c) Ahhoz, hogy az $X+Y$ valószínűségi változó várható értékét meghatározhassuk, előbb $X+Y$ eloszlását kell felírnunk. Ehhez előbb írjuk fel X és Y együttes eloszlását:

$X\backslash Y$	0	1	2	3
0	$1/6$	0	$1/3$	0
1	0	$1/3$	0	$1/6$

d) Az $X+Y$ valószínűségi változó eloszlásának felírásához először gondoljuk meg, milyen értékeket vehet fel a két valószínűségi változó összege. A táblázatból azonnal látjuk, hogy a 0 és 4 közötti egész számok jöhetnek szóba. Az együttes eloszlás alapján azonban azonnal látszik, hogy az összeg 1 és 3 nem lehet. Az oszthatóság szabályai miatt belegondolhatunk ennek az okába. Annak a valószínűsége, hogy az összeg 0, 2 vagy 4, a táblázat alapján meghatározható, így az összeg eloszlása az alábbi szerint írható fel:

$$X+Y : \begin{cases} 0 & 2 & 4 \\ 1/6 & 2/3 & 1/6 \end{cases}$$

e) Az összeg eloszlása alapján annak várható értékét a definíció alapján meghatározhatjuk:

$$E(X+Y) = \sum_i (x+y)_i p_i = 0 \cdot \frac{1}{6} + 2 \cdot \frac{2}{3} + 4 \cdot \frac{1}{6} = 2\,.$$

Tehát $X+Y$ várható értéke 2. Ez megegyezik X és Y várható értékeinek összegével:

$$2 = E(X+Y) = E(X) + E(Y) = \frac{1}{2} + \frac{3}{2}.$$

$X+Y$ várható értékét így kétféleképpen is meghatároztuk, és ellenőriztük, hogy mindegyik esetben 2 lett az eredmény.

⇐15.1. feladat

15.2. feladat. Egy hat- és egy nyolcoldalú dobókockával dobunk. Adjuk meg a dobott számok összegének várható értékét!

Megoldás:

15.2. ábra. *A dobókockák szabályos testek segítségével készülnek: a hatoldalú egy kocka, a nyolcoldalú pedig egy oktaéder.*

Először kiszámoljuk, hogy egyenként a dobókockákkal várhatóan mennyit fogunk dobni. Amikor ez megvan, már csak a kapott értékeket össze kell adni. A hatoldalú dobókocka mindegyik oldalára $1/6$ valószínűséggel gördül. Jelöljük X_6-tal a dobott szám értékét. Ekkor X_6 várható értéke:

$$\begin{aligned}E(X_6) &= \sum_i x_i \cdot p_i = \\ &= 1\cdot\frac{1}{6}+2\cdot\frac{1}{6}+3\cdot\frac{1}{6}+4\cdot\frac{1}{6}+5\cdot\frac{1}{6}+6\cdot\frac{1}{6} \\ &= \frac{1+2+3+4+5+6}{6} = \frac{6\cdot(6+1)}{2\cdot 6} \\ &= \frac{6+1}{2} = 3{,}5\,.\end{aligned}$$

Jelöljük X_8-cal a nyolcoldalú dobókockával dobott számot. Vegyük észre, hogy mindegyik számot $1/8$ valószínűséggel dobhatjuk! Így X_8 várható értéke hasonlóan az előzőekhez felírható:

$$E(X_8) = \frac{1+2+3+4+5+6+7+8}{8} = \frac{8\cdot(8+1)}{2\cdot 8} = \frac{8+1}{2} = 4{,}5\,.$$

Így a dobott számok összegének várható értéke

$$E(X_6+X_8) = E(X_6)+E(X_8) = 3{,}5+4{,}5 = 8\,.$$

Tehát ha a egy nyolc- és egy hatoldalú dobókockával dobunk, a dobott számok összegének várható értéke 8 lesz.

⇐15.2. feladat

15-1. önálló feladat: Két darab, hatoldalú dobókockával dobunk. Számoljuk ki a dobott számok összegének várható értékét kétféleképpen. Először írjuk fel az összeg eloszlását, és a már tanult módon számítsuk ki a várható értéket. Ezután alkalmazzuk az előző tételt, és két várható érték összegeként írjuk fel, majd ellenőrizzük, egyforma eredményt kaptunk-e!

Belátható a tétel több valószínűségi változóra:

15.2. tétel: Amennyiben az $X_1, X_2, \ldots, X_n$ valószínűségi változók várható értéke létezik, akkor

1. az összegük várható értéke is létezik, és
$$E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n),$$
2. az átlaguk várható értéke is létezik, és
$$E\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{E(X_1) + E(X_2) + \ldots + E(X_n)}{n}.$$

Bizonyítás:

1. Az állítás teljes indukcióval egyszerűen belátható: kéttagú összegre beláttuk, az n tagú összeget pedig felírhatjuk, mint $X_1 + \ldots + X_{n-1}$ és X_n összege, amit ismét kéttagú összegként kezelhetünk.
2. Az állítás belátásához kihasználjuk, hogy bármilyen Y valószínűségi változóra teljesül, hogy $E(c \cdot Y) = c \cdot E(Y)$. Ez azt jelenti, hogy $E\left(\frac{Y}{n}\right) = \frac{E(Y)}{n}$, így ha Y helyére a valószínűségi változók összegét írjuk, épp az állítást kapjuk meg.

□

15.3. feladat.

Egy vödörben két hatoldalú, három nyolcoldalú és egy tizenkét-oldalú dobókocka van. A vödört kiborítva adjuk meg a dobott számok

a) összegének és

b) átlagának várható értékét!

Megoldás:

A vödörben összesen 6 dobókocka van, jelölje X_1 és X_2 a hat, X_3, X_4 és X_5 a nyolc-, az X_6 pedig a tizenkét-oldalú dobókockákkal dobott értéket. Mind az átlag, mind az összeg várható értékének számításához előbb egyenként kell kiszámolni a valószínűségi változók várható értékét. Ezért kezdjük ezzel:

A hatoldalú dobókockákat egyformának vehetjük, így X_1 és X_2 eloszlása és várható értéke megegyezik. Így:

$$E(X_1) = E(X_2) = \frac{6+1}{2} = 3{,}5\,.$$

Hasonlóan azonos eloszlást követ a három, nyolcoldalú dobókockát leíró valószínűségi változó. Ezek várható értéke is megegyezik:

$$E(X_3) = E(X_4) = E(X_5) = \frac{8+1}{2} = 4{,}5\,.$$

A tizenkét-oldalú dobókockával dobott érték várható értékét az előbbiekhez hasonlóan számíthatjuk:

$$E(X_6) = \frac{12+1}{2} = 6{,}5\,.$$

A hatoldalúval dobott szám várható értéke $3{,}5$, a nyolcoldalúval $4{,}5$, a tizenkét-oldalúval pedig $6{,}5$.

a) A dobott értékek összegének várható értékét megkapjuk úgy, ha a valószínűségi változók várható értékét összegezzük:

$$E(X_1 + X_2 + X_3 + X_4 + X_5 + X_6) =$$
$$E(X_1) + E(X_2) + E(X_3) + E(X_4) + E(X_5) + E(X_6) =$$
$$3{,}5 + 3{,}5 + 4{,}5 + 4{,}5 + 4{,}5 + 6{,}5 = 2 \cdot 3{,}5 + 3 \cdot 4{,}5 + 6{,}5 = 27$$

Tehát amennyiben ezt a hat dobókockát elgurítjuk, akkor a dobott számok összegének várható értéke 27 lesz.

b) Az átlag számítása hasonló módon történik:

$$E\left(\frac{X_1 + X_2 + X_3 + X_4 + X_5 + X_6}{6}\right) =$$
$$\frac{E(X_1) + E(X_2) + E(X_3) + E(X_4) + E(X_5) + E(X_6)}{6} =$$
$$\frac{3{,}5 + 3{,}5 + 4{,}5 + 4{,}5 + 4{,}5 + 6{,}5}{6} = \frac{2 \cdot 3{,}5 + 3 \cdot 4{,}5 + 6{,}5}{6} = \frac{27}{6} = 4{,}5\,.$$

Tehát a kiborított hat dobókockán szereplő hat számot átlagoljuk, akkor annak várható értéke $4{,}5$ lesz. A két számítási mód kizárólag abban különbözik, hogy a második végig van osztva egy 6-ossal.

⇐15.3. feladat

Aktivitás: Vegyen elő két dobókockát, és gurítsa el őket többször. Adja össze minden dobásnál a dobott értékeket! Jegyezze le, készítsen statisztikát, és nézze meg, tényleg a várható értéke körül ingadozik-e a legtöbb összeg!

A tárgyalt tételnek van egy speciális esete. Legyen azonos az egyes valószínűségi változók várható értéke. Ebben az esetben az alábbi tétel szerint, többet is mondhatunk:

15.3. tétel: Amennyiben az $X_1, X_2, \ldots, X_n$ valószínűségi változók várható értéke létezik és megegyezik, azaz $E(X_1) = E(X_2) = \ldots = E(X_n) = m$, akkor

1. az összegük várható értéke is létezik, és

$$E(X_1 + X_2 + \ldots + X_n) = n \cdot m,$$

2. az átlaguk várható értéke is létezik, és

$$E\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = m\,.$$

Bizonyítás:

1. Azt kell felhasználni, hogy összeg várható értéke a várható értékek összege, így

$$E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n) = m + m + \ldots + m = n \cdot m\,.$$

2. Az előző eredmény mindkét oldalát n-nel osztva, és a várható érték linearitását kihasználva az állítás azonnal adódik. □

15.4. feladat. Száz dobókockát dobunk a levegőbe, majd megszámoljuk, hány esett a 6-os számmal felfele. Mennyi lett a kapott szám várható értéke?

a) Számoljunk először az előző tétel segítségével, nevezetes eloszlás használata nélkül!

b) Számoljuk ki az eredményt nevezetes eloszlással is, majd hasonlítsuk össze az eredményeket!

Megoldás:

a) Ahhoz, hogy az előző tételt használni tudjuk, látnunk kell, hogy a 100 kockadobás egymástól független, de azonos eloszlás szerint viselkedik: $1/6$ lesz a hatos dobás valószínűsége. Ha X_1, X_2, …, X_{100}-al jelöljük az egyes dobókockák által dobott hatosok számát, akkor láthatjuk, hogy ezen valószínűségi változók eloszlása azonos lesz:

$$X_1, X_2, \ldots, X_{100} : \begin{cases} 0 & 1 \\ 5/6 & 1/6 \end{cases} .$$

Így mindegyik valószínűségi változónak egyforma lesz a várható értéke:

$$E(X_1) = E(X_2) = \ldots = E(X_{100}) = 0 \cdot \frac{5}{6} + 1 \cdot \frac{1}{6} = \frac{1}{6} .$$

A fenti tételt alkalmazva láthatjuk, hogy $n = 100$ darab azonos eloszlású valószínűségi változóról van szó, melyek mindegyikének várható értéke $m = 1/6$. Így összegük várható értéke:

$$E(X_1 + X_2 + \ldots X_{100}) = n \cdot m = 100 \cdot \frac{1}{6} = \frac{50}{3} \approx 12{,}6667 .$$

b) A dobott hatosok Y száma binomiális eloszlást követ. Ebben az esetben a binomiális eloszlás paraméterei $n = 100$ és $p = 1/6$, hiszen 100 független kísérletet végzünk, s mindegyik $1/6$ valószínűséggel teljesül. Így

$$E(Y) = n \cdot p = 100 \cdot \frac{1}{6} = \frac{50}{3} \approx 16{,}6667 .$$

Megállapíthatjuk, hogy mindkét módszerrel azonos értéket kaptunk a várható értékre. A száz feldobott dobókocka közül a hatos dobások számának várható értéke 16,6667 lesz. ⇐15.4. feladat

Aktivitás: Keressen olyan tanult nevezetes eloszlást, mely több azonos eloszlású, másik nevezetes eloszlás összegeként írható fel.

15.5. feladat.

Valamikor a nem túl távoli jövőben minden mosógéphez örök garancia fog járni: az elromlott gépeket azonnal cserélik. Egy olyan típust veszünk, melynél – a gyártó hirdetése szerint – 99% annak az esélye, hogy a mosógép élettartama eléri a három évet. Feltéve, hogy az állítás igaz, valamint hogy mindegyik mosógép élettartama exponenciális eloszlású, számoljuk ki, várhatóan mikor fog elromlani a második cseregép?

Megoldás: Legelsőnek azt gondoljuk meg, hogy amikor a második cseregép elromlik, akkor ez a harmadik mosógép lesz, ami tönkremegy. Jelölje X_1, X_2 és X_3 a három gép élettartamát. Mindhárom valószínűségi változó eloszlása exponenciális lesz, megegyező λ paraméterrel. Először ezt a λ paramétert számoljuk ki!

A λ paraméter megállapításához használjuk fel azt az információt, hogy 99% annak a valószínűsége, hogy a mosógép élettartama eléri a 3 évet. Ezt az alábbiak szerint írhatjuk fel:

$$P(X_1 > 3) = 0{,}99$$

Itt elég X_1-gyel számolni, mivel a másik két valószínűségi változó eloszlása is ugyanilyen lesz. Használjuk ki azt, hogy az eloszlás exponenciális:

$$P(X_1 < x) = F_{X_1}(x) = \begin{cases} 1 - e^{-\lambda \cdot x}, & \text{ha } x > 0; \\ 0 & \text{egyébként}. \end{cases}$$

Felírhatjuk ez alapján a $P(X > 3)$ valószínűséget:

$$P(X_1 > 3) = 1 - F_{X_1}(3) = 1 - \left[1 - e^{-3\lambda}\right] = e^{-3\lambda},$$

azaz

$$0{,}99 = P(X_1 > 3) = e^{-3\lambda},$$

amiből $\lambda = -\dfrac{\ln 0{,}99}{3}$. Mivel a valószínűségi változók eloszlása egyforma, így mindegyik exponenciális eloszlás paramétere ez lesz. Várható értékük is meg fog egyezni, azaz

$$E(X_1) = E(X_2) = E(X_3) = -\frac{3}{\ln 0{,}99} = 298{,}50\,.$$

Ez azt jelenti, hogy egy gép nagyjából átlagosan 300 évig bírja. Mivel mindegyik valószínűségi változó várható értéke megegyezik, ezért használhatjuk az előző tételt az összeg várható értékének megállapításához. Három gép esetén $n = 3$ lesz, illetve $m = -\dfrac{3}{\ln 0{,}99}$ a közös várható érték. Így az összeg várható értéke

$$E(X_1 + X_2 + X_3) = n \cdot m = 3 \cdot \left(-\frac{3}{\ln 0{,}99}\right) \approx 895{,}49\,.$$

Azt kaptuk, hogy várhatóan körülbelül 895 és fél év múlva romlik el a második cseregép.

⇐15.5. feladat

Aktivitás: Gondoljon bele milyen gazdasági, és környezeti hatásai lehetnek annak, ha egy mosógép átlagosan 300 évig működőképes. Járjon utána, hogy mi nehezíti meg azt, hogy egy termék ennyire nehezen menjen tönkre.

Leckezáró önellenőrző kérdéssor

1. Hat független valószínűségi változó várható értékei egyenként 1, 2, 3, 4, 5 és 6. Mennyi lesz a valószínűségi változók összegének várható értéke?

☐ 21 ☐ 3,5 ☐ 3 ☐ 25

2. Tizenhat független, egyenletes eloszlású valószínűségi változó mindegyikének várható értéke 4. Mennyi lesz az átlaguk várható értéke?

☐ $4 \cdot 16$ ☐ $4/16$ ☐ 4 ☐ $\frac{4+16}{2}$

3. Hat *valós* számot választunk véletlenszerűen 10 és 20 között. Mekkora lesz a számok átlagának várható értéke?

☐ 90 ☐ 15 ☐ 60 ☐ $25/6$

4. Hat *egész* számot választunk véletlenszerűen 10 és 20 között. Mekkora lesz a számok összegének várható értéke?

☐ 90 ☐ 15 ☐ 60 ☐ $25/6$

5. Egy szabályos érmét dobálunk fel addig, amíg egy fej dobás után közvetlenül egy írást nem dobunk. Mennyi a szükséges dobások várható értéke?

☐ 12 ☐ 6 ☐ 4 ☐ 2

24. LECKE

Független valószínűségi változók szorzatának várható értéke

15.2. Független valószínűségi változók szorzatának várható értéke

15.4. tétel: Ha az X és Y *független* valószínűségi változók várható értéke létezik, akkor létezik a szorzatuk várható értéke is, és

$$E(X \cdot Y) = E(X) \cdot E(Y)\,.$$

Bizonyítás: A bizonyítás most is külön-külön végezzük el diszkrét, illetve folytonos esetre.

1. Diszkrét esetben:

$$E(X \cdot Y) = \sum_i \sum_k (x_i \cdot y_k) \cdot p_{ik} = \sum_i \sum_k (x_i \cdot y_k) \cdot (p_i \cdot q_k) = \sum_i x_i \cdot p_i \sum_k y_k \cdot q_k = \\ = E(X) \cdot E(Y)\,.$$

2. Folytonos esetben:

$$E(X \cdot Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x \cdot y) \cdot f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x \cdot y) \cdot (f_X(x) \cdot f_Y(y))\,\mathrm{d}x\,\mathrm{d}y = \\ = \int_{-\infty}^{\infty} x \cdot f_X(x)\,\mathrm{d}x \int_{-\infty}^{\infty} y \cdot f_Y(y)\,\mathrm{d}y = \\ = E(X) \cdot E(Y)\,.$$

$\square$

Teljes indukcióval belátható, hogy ha az $X_1, X_2, \ldots, X_n$ független valószínűségi változók várható értéke létezik, akkor a szorzatuk várható értéke is létezik, és

$$E(X_1 \cdot X_2 \cdot \ldots \cdot X_n) = E(X_1) \cdot E(X_2) \cdot \ldots \cdot E(X_n)\,.$$

15.6. feladat. Dobókockával dobunk. Legyen X a dobott szám kettővel, Y pedig dobott szám *hárommal* vett maradéka.

a) Független lesz-e X és Y?

b) Igaz-e, hogy X és Y szorzatának várható értéke megegyezik a várható értékeik szorzatával?

Megoldás: Mindkét feladat megoldásához előbb fel kell írni X és Y eloszlását, valamint az együttes eloszlást is. Ezek segítségével közvetlenül megállapítható a függetlenség. Az együttes eloszlás segítségével felírható a szorzat eloszlása is, így mind X, Y és $X \cdot Y$ várható értéke is kiszámolható.

Az X valószínűségi változó eloszlásához gondoljuk meg, hogy a dobókocka hat száma közül az 1, a 3 és az 5 páratlan, a 2, a 4 és a 6 pedig páros. Páratlan dobás esetén X értéke 1, páros esetén pedig 0. Így X eloszlása:

$$X : \begin{cases} 0 & 1 \\ 1/2 & 1/2 \end{cases} .$$

Hasonlóan meghatározhatjuk Y eloszlását is. A hárommal való osztás maradéka háromféle lehet: 0, 1 és 2. A hatoldalú dobókockán 3 és 6 lesz hárommal osztható, így 0 maradékot ad, az 1 és a 4 ad 1 maradékot, a 2 és az 5 adja a 2-t. Mindegyik értéknél két jó eset lesz a hatból, így Y eloszlása:

$$Y : \begin{cases} 0 & 1 & 2 \\ 1/3 & 1/3 & 1/3 \end{cases} .$$

Az együttes eloszlás felírásához gondoljuk át a következőket:

$$1 = 0 \cdot 2 + 1 = 0 \cdot 3 + 1 \qquad 4 = 2 \cdot 2 + 0 = 1 \cdot 3 + 1$$
$$2 = 1 \cdot 2 + 0 = 0 \cdot 3 + 2 \qquad 5 = 2 \cdot 2 + 1 = 1 \cdot 3 + 2$$
$$3 = 1 \cdot 2 + 1 = 1 \cdot 3 + 0 \qquad 6 = 3 \cdot 2 + 0 = 2 \cdot 3 + 0$$

A fentiek alapján az együttes eloszlás könnyen felírható:

$X\backslash Y$	0	1	2
0	$1/6$	$1/6$	$1/6$
1	$1/6$	$1/6$	$1/6$

a) Vizsgáljuk meg X és Y függetlenségét! X és Y mindegyik értékpárját csak egy bizonyos dobás esetén veszi fel. Ez azt jelenti, hogy mindegyik értékpár azonos valószínűséggel fordul elő. Az egyes peremeloszlásokban is azonos valószínűséggel fordulnak elő X és Y értékei, így a függetlenség vizsgálatát egyszerre vizsgálhatjuk mindegyik érték esetén:

$$P(X = x_i) \cdot P(Y = y_k) = 1/2 \cdot 1/3 = 1/6 = P(X = x_i, Y = y_k)$$

A fenti számítás X mindegyik x_i és Y mindegyik y_k értékére működik. Így a két valószínűségi változó független lesz.

b) Most vizsgáljuk meg a várható értékeket. $X \cdot Y$ várható értékéhez meg kell állapítanunk az eloszlását, ami az együttes eloszlás alapján egyszerű lesz. $X \cdot Y$ értéke 0 lesz, ha bármelyik értéke 0. Egyébként az 1 és a 2 értéket veheti fel az alábbi esetekben:

$$\begin{aligned} P(X \cdot Y = 1) &= P(X = 1, Y = 1) = 1/6 \\ P(X \cdot Y = 2) &= P(X = 1, Y = 2) = 1/6 \\ P(X \cdot Y = 0) &= 2/3\,. \end{aligned}$$

A maradék 4 esetben lesz $X \cdot Y$ értéke 0. Így az eloszlását fel is írhatjuk:

$$X \cdot Y : \begin{cases} 0 & 1 & 2 \\ 2/3 & 1/6 & 1/6 \end{cases}.$$

Most, hogy kiszámoltuk X, Y és $X \cdot Y$ eloszlását, a várható értékeiket definíció szerint számolhatjuk:

$$E(X) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2},$$
$$E(Y) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} = 1,$$
$$E(X \cdot Y) = 0 \cdot \frac{2}{3} + 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} = \frac{1}{2}.$$

Így a valószínűségi változók várható értékeit kiszámoltuk. Az *a)* feladat alapján tudhatjuk, hogy a két valószínűségi változó független. Ellenőrizzük le, hogy a két várható érték szorzata egyezik-e a szorzatuk várható értékével:

$$E(X) \cdot E(Y) = \frac{1}{2} \cdot 1 = \frac{1}{2} = E(X \cdot Y)$$

Így a tétel állítását ezen az egyszerű példán szemléltethettük.

⇐15.6. feladat

15-2. önálló feladat: Dobókockával dobunk. Legyen X a dobott szám kettővel, Y pedig dobott szám *néggyel* vett maradéka.

a) Független lesz-e X és Y?

b) Igaz-e, hogy X és Y szorzatának várható értéke megegyezik a várható értékeik szorzatával?

15.7. feladat. Legyen Y diszkrét valószínűségi változó az alábbi eloszlással:

$$Y : \begin{cases} 0 & \pi/2 & \pi & 3\pi/2 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{cases}$$

Definiáljuk X_1-et és X_2-t mint Y transzformáltját az alábbi módon:

$$X_1 = \sin Y \qquad\qquad X_2 = \cos Y\,.$$

a) Igaz-e, hogy X_1 és X_2 független?

b) Igaz-e, hogy X_1 és X_2 szorzatának várható értéke megegyezik várható értékük szorzatával?

Megoldás: Mindkét kérdés megválaszolásához előbb fel kell írni X_1 és X_2 eloszlását, valamint az együttes eloszlást. Nézzük meg először, milyen értékeket vehetnek fel ezek a valószínűségi változók Y egyes értékei esetén:

$$\begin{aligned} \cos(0) &= 1 & \qquad \sin(0) &= 0 \\ \cos(\pi/2) &= 0 & \sin(\pi/2) &= 1 \\ \cos(\pi) &= -1 & \sin(\pi) &= 0 \\ \cos(3\pi/2) &= 0 & \sin(3\pi/2) &= -1 \end{aligned}$$

Ez azt jelenti, hogy mind X_1, mind X_2 lehetséges értékei -1, 0 és 1. Megfigyelhetjük, hogy eloszlásuk is azonos, hiszen mindegyik értéket ugyanannyiszor veszik fel, Y pedig mindegyik értékét azonos valószínűséggel veszi fel:

$$X_1 : \begin{cases} -1 & 0 & 1 \\ 1/4 & 1/2 & 1/4 \end{cases}, \qquad X_2 : \begin{cases} -1 & 0 & 1 \\ 1/4 & 1/2 & 1/4 \end{cases}.$$

Az együttes eloszlás is egyszerűen előállítható a fenti értékek alapján:

$X_1 \backslash X_2$	-1	0	1
-1	0	$1/4$	0
0	$1/4$	0	$1/4$
1	0	$1/4$	0

a) Először X_1 és X_2 függetlenségét vizsgáljuk meg. A két eloszlásból és az együttes eloszlás alapján ez könnyen cáfolható. Elég egy (x_i,x_j) értékpárt vizsgálni, hiszen a függetlenséghez bármelyik (x_i,x_j) értékpárra igaznak kell lennie, hogy

$$P(X_1 = x_i, X_2 = x_j) = P(X_1 = x_i) \cdot P(X_2 = x_j)\,.$$

Legyen most x_i és x_j is 0. Nem lehet két valószínűségi változó egyszerre 0, hiszen az együttes eloszlásból kiolvasható:

$$P(X_1 = 0, X_2 = 0) = 0\,.$$

Azonban külön-külön mindegyik valószínűségi változó $1/2$ valószínűséggel veszi fel a 0 értéket. Ebből adódik, hogy:

$$P(X_1 = 0, X_2 = 0) = 0 \neq 1/2 \cdot 1/2 = P(X_1 = 0) \cdot P(X_2 = 0)\,.$$

Ebből azt a következtetést vonhatjuk le, hogy X_1 és X_2 *nem* független.

b) Vizsgáljuk meg most a várható értékeket. X_1 és X_2 eloszlásából könnyen meghatározható azok várható értéke:

$$E(X_1) = -1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} = 0,$$
$$E(X_2) = -1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} = 0\,.$$

Mivel a két valószínűségi változó eloszlása azonos, így várható értékük is megegyezik, mindkettőjüknek 0. X_1X_2 várható értékének meghatározásához előbb vizsgáljuk meg a lehetséges értékeit. X_1 és X_2 együttes eloszlása alapján megfigyelhetjük, hogy a két valószínűségi változóból alkotott (X_1, X_2) valószínűségi vektorváltozó értéke négyféle lehet: $(0, -1)$, $(0{,}1)$, $(-1{,}0)$ és $(1{,}0)$. Ez azt jelenti, hogy az X_1X_2 szorzatnak kizárólag egyféle értéke lehet:

$$P(X_1X_2 = 0) = 1\,.$$

Természetesen ez azt jelenti, hogy a várható értéke is 0 lesz, sőt ez meg fog egyezni a két komponens várható értékének szorzatával:

$$E(X_1X_2) = 0 = 0 \cdot 0 = E(X_1) \cdot E(X_2)\,.$$

Figyeljük meg, hogy most egy olyan példát láthattunk, ahol két valószínűségi változó szorzatának várható értéke megegyezik a valószínűségi változók várható értékének szorzatával. Ennek ellenére a két valószínűségi változó *nem* független.

⇐15.7. feladat

Aktivitás: Keressen példát a való életből olyan jelenségekre, melyeket egy olyan $X \cdot Y$ alakú valószínűségi változóval írhatunk le, ahol az X karakterisztikus eloszlású. Például egy útkereszteződésben bekövetkező balesetet így írhatunk le: X legyen 1, ha egy adott napon történik a kereszteződésben baleset, Y pedig az adott kereszteződésben bekövetkezett balesetben a sebesültek száma. Ekkor $X \cdot Y$ lesz az adott napon az adott kereszteződésben balesetben megsérültek száma.

Az alábbi feladat segítségével arra kaphatunk példát, hogy amennyiben a függetlenség adott és természetes, milyen jelentős mértékben egyszerűsödik le a várható érték számítása.

15.8. feladat.

Három dobókockával dobunk egyszerre. Összeadjuk a dobott számokat, de úgy, hogy előbb egy érmét dobunk fel: ha fejet dobunk, hozzászámoljuk a dobott értéket az összeghez, ha írást, akkor nem. Mekkora lesz az így kapott összeg várható értéke?

Megoldás:

Legelőször azt vegyük észre, hogy összesen hat kísérlet történik: három kockadobás és három érmedobás. Arra is figyeljünk fel, hogy ez a hat kísérlet független egymástól. Jogosan bízhatunk abban, hogy ilyen esetben a kért várható érték egyszerűen számítható.

Legelőször próbáljuk meg felírni azt a valószínűségi változót, aminek a várható értékét keressük. Ehhez vezessünk be hat valószínűségi változót, ami a hat kísérlet kimenetelét jellemzi: legyen X_1, X_2 és X_3 rendre a három dobókocka által dobott szám. Ezen kívül legyenek Y_1, Y_2 és Y_3 valószínűségi változók, melyek értéke legyen 1, ha rendre elsőre, másodikra avagy harmadikra fejet, és 0, ha írást dobtunk.

Gondoljuk át: az első dobókocka által dobott X_1 értéket akkor adjuk az összeghez, ha Y_1 értéke 1. Tehát lényegében az $X_1 \cdot Y_1$ típusú szorzatokat összegezzük. Tehát a keresett valószínűségi változónk az $X_1 \cdot Y_1 + X_2 \cdot Y_2 + X_3 \cdot Y_3$ lesz. Mivel a hat valószínűségi változó egymástól független, a kérdéses változó várható értéke kiszámolható:

$$\begin{aligned} E(X_1Y_1 + X_2Y_2 + X_3Y_3) &= E(X_1 \cdot Y_1) + E(X_2 \cdot Y_2) + E(X_3 \cdot Y_3) = \\ &= E(X_1)E(Y_1) + E(X_2)E(Y_2) + E(X_3)E(Y_3) = \\ &= 3{,}5 \cdot 1/2 + 3{,}5 \cdot 1/2 + 3{,}5 \cdot 1/2 = 5{,}25\,. \end{aligned}$$

Tehát három kockadobás esetén ha mindegyik dobott értéket csak 50% valószínűséggel számoljuk az összeghez, akkor az várhatóan $5{,}25$ lesz. Ez épp a fele annak, amit akkor várnánk, amikor a három kockadobás eredményét egyszerűen összeadnánk.

⇐15.8. feladat

15-3. önálló feladat: 216 darab kis dobókockából egy nagy, $6 \times 6 \times 6$-os kockát építünk úgy, hogy a kis dobókockák pöttyei azonos irányba állnak. A nagy kocka mindegyik oldalát pirosra festjük. Miután a festék megszáradt, a kis dobókockákat egy vödörbe szedjük, majd abból kigurítjuk őket. Számolja össze a dobott piros számokat! Mekkora lesz az összeg várható értéke?

Leckezáró önellenőrző kérdéssor

1. Négy dobókockát elgurítunk. Mennyi lesz a dobott számok szorzatának várható értéke?

☐ $\frac{6^4}{2}$ ☐ $\frac{7^4}{2^2}$ ☐ $\frac{7}{2^4}$ ☐ $\frac{7^4}{16}$

2. A $[2;4]$ intervallumon választunk két számot véletlenszerűen, amit összeszorzunk. Mennyi lesz a szorzat várható értéke?

☐ 3 ☐ 12 ☐ 9 ☐ 8

3. Legyen X egy véletlenszerűen választott szám a -1, 0 és 1 számok közül, s legyen $Y = -X$. Igaz-e, hogy X és Y szorzatának várható értéke megegyezik várható értékük szorzatával?

☐ Igen, mert az állításhoz nem kell, hogy X és Y független legyen.

☐ Nem, mert ebben az esetben biztos, hogy X és Y nem függetlenek.

☐ Igen, mert ebben az esetben X és Y független.

☐ Nem, mert X és Y várható értéke 0, szorzatuk várható értéke viszont negatív.

4. Melyik középpel igaz az állítás független valószínűségi változókra?

☐ Számtani közepük várható értéke megegyezik a várható értékek számtani közepével.

☐ Mértani közepük várható értéke megegyezik a várható értékek mértani közepével.

☐ Egyik középpel sem igaz az állítás.

☐ Mindegyik középpel igaz az állítás.

25. LECKE

Valószínűségi változók összegének és átlagának szórása

15.3. Független valószínűségi változók összegének és átlagának szórása

A várható érték után a szórás a második fontos paraméter, amivel egy valószínűségi változót jellemezhetünk. Több valószínűségi változó esetén a szórás vizsgálatánál a legfontosabb kérdés, hogy a valószínűségi változók függetlenek-e vagy nem. Független valószínűségi változók esetén ezek összegének és átlagának szórása viszonylag könnyen kiszámítható. Ennek az alapja az, hogy független valószínűségi változók összegének szórásnégyzete megegyezik a szórásnégyzetük összegével. Ezt tételként az alábbiak szerint mondjuk ki:

15.5. tétel:

Ha az X és Y *független* valószínűségi változók szórásnégyzete létezik, akkor létezik az összegük szórásnégyzete is, és

$$D^2(X+Y) = D^2(X) + D^2(Y)\,.$$

Bizonyítás:

$$D^2(X+Y) = E[(X+Y)^2] - E^2[X+Y] = E[X^2 + 2X \cdot Y + Y^2] - (E(X)+E(Y))^2 =$$

Kihasználva azt, hogy összeg várható értéke a várható értékek összege, valamint azt, hogy függetlenek szorzatának várható értéke a várható értékek szorzata:

$$\begin{aligned} &= E(X^2) + 2 \cdot E(X) \cdot E(Y) + E(Y^2) - E^2(X) - 2 \cdot E(X) \cdot E(Y) - E^2(Y) = \\ &= E(X^2) - E^2(X) + E(Y^2) - E^2(Y) = \\ &= D^2(X) + D^2(Y)\,. \end{aligned}$$

□

A tétel segítségével, gyökvonás után a feltételeket teljesítő X és Y összegének szórását is kiszámolhatjuk:

$$D(X+Y) = \sqrt{D^2(X) + D^2(Y)}\,.$$

15.6. tétel: Ha az $X_1, X_2, \dots, X_n$ független valószínűségi változók szórásnégyzete létezik, akkor az összegük szórásnégyzete is létezik, és

$$D^2(X_1 + X_2 + \ldots + X_n) = D^2(X_1) + D^2(X_2) + \ldots + D^2(X_n)\,.$$

Bizonyítás: Az állítás teljes indukció segítségével belátható. □

Hasonlóan, a tétel feltételeit teljesítő $X_1, X_2, \dots, X_n$ valószínűségi változók összegének szórása is felírható:

$$D(X_1 + X_2 + \ldots + X_n) = \sqrt{D^2(X_1) + D^2(X_2) + \ldots + D^2(X_n)}\,.$$

15.9. feladat. Legyen X és Y független, azonos eloszlású valószínűségi változó, melyeknek várható értéke 10, szórása 5.

a) Számoljuk ki $X + Y$ várható értékét és szórását!

b) Számoljuk ki $2 \cdot X$ várható értékét és szórását, s hasonlítsuk össze az előző eredménnyel!

Megoldás:

a) $X + Y$ várható értéke a 23. lecke alapján egyszerűen a várható értékek összege lesz:

$$E(X + Y) = E(X) + E(Y) = 10 + 10 = 20\,.$$

Tehát $X + Y$ várható értéke 20. A szórásának kiszámításához a most bizonyított tételt használhatjuk, mivel X és Y független:

$$D(X + Y) = \sqrt{D^2(X) + D^2(Y)} = \sqrt{5^2 + 5^2} = 5\sqrt{2} \approx 7{,}0711\,.$$

Így az $X + Y$ valószínűségi változó szórása egy kicsivel több, mint 7.

b) Vizsgáljuk meg most a $2 \cdot X$ várható értékét. Használjuk ki a várható érték azon tulajdonságát, hogy konstansszorzót kiemelhetünk!

$$E(2 \cdot X) = 2 \cdot E(X) = 2 \cdot 10 = 20\,.$$

Tehát ha X-et kettővel szorozzuk, a várható értéke is duplájára nő. Nézzük, mi történik a szórással. Itt is használjuk ki, hogy a konstanst kiemelhetjük:

$$D(2 \cdot X) = |2| \cdot D(X) = 2 \cdot 5 = 10\,.$$

X duplájának szórása is a kétszeresére nő. Figyeljük meg, hogy mennyire más, ha két azonos eloszlású valószínűségi változót összeadunk, mint ha egyszerűen kettővel megszorozzuk. Ugyan a várható értékük meg fog egyezni, de a másodiknak jóval nagyobb lesz a szórása.

⇐15.9. feladat

Aktivitás: Gondolja meg, mi okozhatja a második esetben a nagyobb szórást!

15.10. feladat. Dobjunk egy hat-, egy nyolc- és egy tizenkét-oldalú dobókockával egyszerre.

a) Számoljuk ki a dobott számok összegének szórását!

b) Számoljuk ki a dobott számok átlagának szórását is!

Megoldás: Mind az összeg, mind az átlag szórásához szükség lesz az egyes dobókockák által dobott értékek szórására. A hatoldalú dobókockán 1-től 6-ig, a nyolcoldalún 1-től 8-ig, a tizenkét-oldalún pedig 1-től 12-ig szerepelnek a számok. Három szórás helyett elég egyet is számolni, ha azt egy n paraméter segítségével tesszük. Képzeljünk el egy olyan dobókockát, melyiken 1-től n-ig vannak a számok, és mindegyik oldalára $1/n$ valószínűséggel gurul. Jelölje X_n a dobókocka által dobott számot. Az összeg és az átlag szórásnégyzetének számításánál figyelembe vehetjük, hogy az egyes dobókockákkal dobott értékel függetlenek egymástól.

Először számoljuk ki X_n várható értékét! Ezt legegyszerűbben a definíció segítségével tehetjük meg:

$$E(X_n) = \sum_i x_i p_i = 1 \cdot \frac{1}{n} + 2 \cdot \frac{1}{n} + \ldots + n \cdot \frac{1}{n} = \frac{1+2+\ldots+n}{n}.$$

Használjuk ki, hogy az első n pozitív egész szám összege $n \cdot (n+1)/2$, tehát

$$E(X_n) = \frac{1+2+\ldots+n}{n} = \frac{n \cdot (n+1)}{2 \cdot n} = \frac{n+1}{2}.$$

Hasonlóan számoljuk ki a dobott szám négyzetének várható értékét is:

$$E(X_n^2) = \frac{1^2+2^2+\ldots+n^2}{n} = \frac{n(n+1)(2n+1)}{6n} = \frac{(n+1)(2n+1)}{6}.$$

Itt kihasználtuk, hogy az első n pozitív egész szám négyzetének összege $\dfrac{n \cdot (n+1) \cdot (2n+1)}{6}$. Ha n értéke rendre a 6, 8 és 12 számokat veszi fel, akkor X_6, X_8 és X_{12} jelöli rendre a hat, a nyolc és a tizenkét-oldalú dobókockával dobott értéket. A kockadobásoknak egymásra nincs hatásuk, így X_6, X_8 illetve X_{12} függetlennek tekinthető. Írjuk fel a bevezetett valószínűségi változók és négyzeteik várható értékét:

$$E(X_6) = \frac{6+1}{2} = \frac{7}{2}, \qquad E(X_6^2) = \frac{(6+1)\cdot(2\cdot 6+1)}{6} = \frac{91}{6},$$

$$E(X_8) = \frac{8+1}{2} = \frac{9}{2}, \qquad E(X_8^2) = \frac{(8+1)\cdot(2\cdot 8+1)}{6} = \frac{153}{6},$$

$$E(X_{12}) = \frac{12+1}{2} = \frac{13}{2}, \qquad E(X_{12}^2) = \frac{(12+1)\cdot(2\cdot 12+1)}{6} = \frac{325}{6}.$$

A valószínűségi változók és négyzeteik várható értékeinek segítségével ki tudjuk számolni a dobott számok szórásnégyzeteit:

$$D^2(X_6) = E(X_6^2) - E^2(X_6) = \frac{91}{6} - \left[\frac{7}{2}\right]^2 = \frac{35}{12},$$

$$D^2(X_8) = E(X_8^2) - E^2(X_8) = \frac{153}{6} - \left[\frac{9}{2}\right]^2 = \frac{63}{12},$$

$$D^2(X_{12}) = E(X_{12}^2) - E^2(X_{12}) = \frac{325}{6} - \left[\frac{13}{2}\right]^2 = \frac{143}{12}.$$

a) Az előbb kimondott tétel elsősorban a szórásnégyzetről szól: független valószínűségi változók összegének szórásnégyzete megegyezik szórásnégyzeteik összegével. Ez azt jelenti, hogy mivel X_6, X_8 és X_{12} független, az $X_6 + X_8 + X_{12}$ valószínűségi változó szórásnégyzete felírható az előbb számoltakat felhasználva:

$$D^2(X_6 + X_8 + X_{12}) = D^2(X_6) + D^2(X_8) + D^2(X_{12}) = \frac{35 + 63 + 143}{12} = \frac{241}{12}.$$

Az összeg szórásnégyzete így $\frac{241}{12}$ lesz, ebből a szórás gyökvonással számítható:

$$D(X_6 + X_8 + X_{12}) = \sqrt{\frac{241}{12}} \approx 4{,}4814\,.$$

Azaz a dobott számok összegének szórása körülbelül 4 és fél lesz.

b) Az átlag szórásnégyzeténél az elemek száma a nevezőből négyzetként kiemelhető, mivel konstans. Az összeg szórásnégyzete – független valószínűségi változókról lévén szó – a szórásnégyzetek összege lesz:

$$D^2\left(\frac{X_6+X_8+X_{12}}{3}\right)=\frac{D^2(X_6)+D^2(X_8)+D^2(X_{12})}{3^2}=\frac{35+63+143}{12\cdot 9}=\frac{241}{108}.$$

Az átlag szórásnégyzete így $\frac{241}{108}$ lesz, ebből a szórás gyökvonással számítható:

$$D\left(\frac{X_6+X_8+X_{12}}{3}\right)=\sqrt{\frac{241}{108}}\approx 1{,}4938\,.$$

Az átlag szórását úgy is megkaphatjuk, ha kihasználjuk a szórás azon tulajdonságát, hogy a konstansszorzó abszolút értéke kiemelhető:

$$D\left(\frac{X_6+X_8+X_{12}}{3}\right)=\left|\frac{1}{3}\right|\cdot D(X_6+X_8+X_{12})\approx\frac{1}{3}\cdot 4{,}4814\approx 1{,}4938\,.$$

Tehát az átlag szórása nagyjából másfél lesz. Ez épp az összeg szórásának harmada, mint ahogy azt láthattuk.

⇐15.10. feladat

15-4. önálló feladat: Képzeljük el az előző feladatban leírt virtuális dobókockát, amivel 1-től n-ig dobhatunk számokat úgy, hogy mindegyik számot azonos valószínűséggel dobjuk. Lássa be, hogy a dobott szám, mint valószínűségi változó, szórásnégyzete $\frac{n^2-1}{12}$ lesz! Hasonlítson össze egy ilyen kockadobást azzal, amikor az $[1;n]$ intervallumból húz véletlenszerűen egy valós számot!

Ha több független valószínűségi változónak azonos az eloszlása, akkor az előbbi tételt egyszerűbb formában is alkalmazni tudjuk. A most kimondott állítások elsősorban a statisztikai vizsgálatok során fontosak, amikor független mintavételt végzünk, hiszen tudjuk, hogy az egyes minták vizsgált paramétere azonos eloszlású lesz. Magát a tételt elég olyan független valószínűségi változókra kimondani, melyek szórása megegyezik.

15.7. tétel: Legyenek $X_1, X_2, \ldots, X_n$ független valószínűségi változók, azonos szórással, azaz $D(X_1) = D(X_2) = \ldots = D(X_n) = \sigma$. Ekkor

1. Az összegük szórása $\sigma\sqrt{n}$, azaz
$$D(X_1 + X_2 + \ldots + X_n) = \sigma\sqrt{n}\,.$$
2. Az átlaguk szórása $\dfrac{\sigma}{\sqrt{n}}$, azaz
$$D\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{\sigma}{\sqrt{n}}\,.$$

Bizonyítás:

1. Azt kell felhasználni, hogy függetlenek összegének szórása a szórásnégyzetek összegének négyzetgyöke, így
$$\begin{aligned} D(X_1 + X_2 + \ldots + X_n) &= \sqrt{D^2(X_1) + D^2(X_2) + \ldots + D^2(X_n)} = \\ &= \sqrt{\sigma^2 + \sigma^2 + \ldots + \sigma^2} = \sqrt{n \cdot \sigma^2} = \sigma\sqrt{n}\,. \end{aligned}$$
2. Az előző pont eredményének mindkét oldalát n-nel osztva, s a konstans kiemelhetőségét kihasználva:
$$D\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{1}{n} \cdot D(X_1 + X_2 + \ldots + X_n) = \frac{1}{n} \cdot \sigma\sqrt{n} = \frac{\sigma}{\sqrt{n}}\,.$$

Foglaljuk össze, hogy több (n darab) azonos eloszlású független valószínűségi változó összegének és átlagának várható értéke és szórása hogyan viselkedik a valószínűségi változókat egyenként jellemző m várható értékhez és σ szóráshoz képest. A változások „növekvő sorrendben” lesznek feltüntetve:

1. Átlaguk szórása az eredeti szórás $\sqrt{n}$-ed részére csökken.
2. Átlaguk várható értéke megegyezik az eredeti várható értékkel.
3. Összegük szórása az eredeti szórás $\sqrt{n}$-szeresére nő.
4. Összegük várható értéke az eredeti várható érték n-szeresére nő.

15-5. önálló feladat: Pozitív értékeket felvevő valószínűségi változók jellemzésére szokták még használni az úgynevezett relatív szórást. Ez a paraméter a valószínűségi változó szórásának és várható értékének hányadosa. Vizsgálja meg, hogyan viselkedik a relatív szórás azonos eloszlású, független valószínűségi változók összegzése és átlagolása esetén!

15.11. feladat.

Egy vödörben száz darab hatoldalú dobókocka van. A vödört kiborítva összeadjuk a dobókockák által mutatott számokat.

a) Adjuk meg az eredmény várható értékét és szórását!

b) Hogyan változik az eredmény, ha a dobókockák felét nyolcoldalúra cseréljük?

Megoldás:

a) Jelölje $X_1, X_2, \ldots, X_{100}$ az egyes dobókockák által dobott számokat. Feltehetjük, hogy mindegyik dobókocka szabályos, így a valószínűségi változók eloszlása azonos lesz, így várható értékük és szórásuk megegyezik. A feladat megoldásához először ezt az m várható értéket és σ szórást, azaz az egyes dobókockák által dobott számok várható értékét és szórását határozzuk meg.
Láthattuk korábban, hogy ha egy dobókockán 1-től n-ig vannak a számok, melyeket azonos valószínűséggel vesz fel, akkor a dobott szám várható értéke $\frac{n+1}{2}$, szórásnégyzete pedig $\frac{n^2-1}{12}$. Így a közös m várható érték és σ szórás az alábbiak szerint számolható:

$$m = \frac{6+1}{2} = 3{,}5, \qquad \sigma = \sqrt{\frac{6^2-1}{12}} \approx 1{,}7078\,.$$

A vödörben a dobókockák száma $n = 100$, így a dobott számok összegének várható értéke az alábbiak szerint alakul:

$$E(X_1 + X_2 + \ldots + X_{100}) = n \cdot m = 100 \cdot 3{,}5 = 350\,.$$

Így 100 darab dobókockát elgurítva a dobások összegének várható értéke 350 lesz. Az összeg szórását is egyszerűen meghatározhatjuk, mivel a dobások eredményei függetlenek egymástól:

$$D(X_1 + X_2 + \ldots + X_{100}) = \sqrt{n} \cdot \sigma \approx \sqrt{100} \cdot 1{,}7078 = 17{,}078\,.$$

A dobott értékek összegének szórása így körülbelül 17 lesz.

b) Mint az előbbi részfeladatban láttuk, előbb érdemes külön kiszámolni az egyes dobókockák által dobott értékek várható értékét és szórását. Az egyértelmű jelölés miatt legyen most m_6 a hatoldalú és m_8 a nyolcoldalúval dobott szám várható értéke, σ_6 és σ_8 pedig a szórása. Ezek a fenti képletek segítségével meghatározhatóak:

$$m_6 = \frac{6+1}{2} = 3{,}5, \qquad \sigma_6 = \sqrt{\frac{6^2-1}{12}} \approx 1{,}7078,$$

$$m_8 = \frac{8+1}{2} = 4{,}5, \qquad \sigma_8 = \sqrt{\frac{8^2-1}{12}} \approx 2{,}2913\,.$$

Mivel a dobókockák felét cseréltük ki, ezért mindegyik típusból $n = 50$ darab lesz. A dobott számok összegének várható értékét könnyen számolhatjuk. Jelölje $X_1, X_2, \ldots, X_{50}$ a hatoldalúak, $Y_1, Y_2, \ldots, Y_{50}$ pedig a nyolcoldalúak által dobott számokat:

$$\begin{aligned} E(X_1 + X_2 + \ldots + X_{50} + Y_1 + Y_2 + \ldots + Y_{50}) =& E(X_1 + X_2 + \ldots + X_{50}) + E(Y_1 + Y_2 + \ldots + Y_{50}) \\ =& n \cdot m_6 + n \cdot m_8 = 50 \cdot 3{,}5 + 50 \cdot 4{,}5 = 400\,. \end{aligned}$$

Tehát ha a dobókockák felét nyolcoldalúra cseréljük, akkor a dobott számok összegének várható értéke 350-ről 400-ra növekszik.

⇐15.11. feladat

15.12. feladat. Egységnyi oldalhosszúságú négyzeten választunk egy pontot. Legyen az X_1, X_2, X_3 és X_4 valószínűségi változó a pont távolsága a négy oldaltól.

a) Számoljuk ki a valószínűségi változók várható értékét és szórását!

b) Számoljuk ki a valószínűségi változók összegének várható értékét és szórását!

Megoldás:

a) Mivel a négyzet szimmetrikus, a négy valószínűségi változó eloszlása megegyezik: el tudjuk forgatni a négyzetet úgy, hogy az egyik valószínűségi változó szerepét lényegében egy másik veszi át. Így elég lesz például az X_1 várható értékét és szórását meghatározni. Könnyen átgondolható az is, hogy X_1, s így a többi eloszlása egyenletes lesz, méghozzá a $[0;1]$ intervallumon. Ha m-mel jelöljük a közös várható értéket, σ-val pedig a közös szórást, akkor ezeket az egyenletes eloszlásnál tanultak alapján könnyen számolhatjuk:

$$m = \frac{0+1}{2} = \frac{1}{2}, \qquad \sigma = \frac{1-0}{\sqrt{12}} \approx 0{,}2887\,.$$

Tehát X_1, X_2, X_3 és X_4 mindegyikének várható értéke $1/2$, és szórásuk is megegyezik, ami $0{,}2887$ lesz.

b) Az összegük várható értéke könnyen számítható, hisz $n = 4$ azonos eloszlású valószínűségi változó esetén ez a várható érték négyszerese lesz:

$$E(X_1 + X_2 + X_3 + X_4) = n \cdot m = 4 \cdot \frac{1}{2} = 2\,.$$

A szórás kiszámításánál figyelembe kell venni azt, hogy ezek a valószínűségi változók *nem* függetlenek. Ha X_1 és X_2 szemközti oldalaktól mért távolságot jelöl – ekkor X_3 és X_4 is szemközti oldalaktól mért távolságot jelöl – akkor $X_1 + X_2 = 1$ lesz, akárhogyan is választom a pontot. Hasonlóan $X_3 + X_4 = 1$ lesz minden esetben. Ez azt jelenti, hogy

$$D(X_1 + X_2 + X_3 + X_4) = D(2) = 0,$$

hiszen az összeg mindig 2, így a szórása is 0 lesz. Vegyük észre, hogy a várható érték viszont még az ilyen furcsán függő helyzetekben is megtartja azt a tulajdonságát, hogy az összeg várható értéke a várható értékek összege lesz.

⇐15.12. feladat

Független valószínűségi változók esetén láthattuk, miként viselkedik összegük várható értéke, és szórása. Amennyiben azt is tudjuk, hogy ezek a valószínűségi változók valamilyen meghatározott eloszlást követnek, többet is tudunk mondani:

15.8. tétel: Legyen X és Y két független, normális eloszlású valószínűségi változó. Ebben az esetben az $X+Y$ valószínűségi változó is normális eloszlású.

15-6. önálló feladat: A tételben szereplő X és Y valószínűségi változók paramétereiből számolja ki az $X+Y$ paramétereit!

15.9. tétel: Legyen $X_1, X_2, \ldots, X_n$ független, normális eloszlású valószínűségi változó. Ekkor az $Y = X_1 + X_2 + \ldots + X_n$ valószínűségi változó is normális eloszlású.

15-7. önálló feladat: A normális eloszlásról tanultak segítségével mondjon ki a fenti tételekhez hasonlót független, normális eloszlású valószínűségi változók átlagáról!

15.13. feladat. Négy független, normális eloszlású valószínűségi változó mindegyikének $m = 6$ a várható értéke, és $\sigma = 8$ a szórása. Mekkora a valószínűsége annak, hogy

a) összegük nagyobb, mint 32,

b) átlaguk kisebb, mint 10?

Megoldás:

a) Normális eloszlású valószínűségi változók X_Σ összege is normális eloszlású lesz. A valószínűség meghatározásához az összeg eloszlásának paramétereit, azaz várható értékét és szórását kell meghatároznunk. A valószínűségi változók eloszlása egyforma, így

$$E(X_\Sigma) = n \cdot m = 4 \cdot 6 = 24,$$

valamint
$$D(X_\Sigma) = \sqrt{n} \cdot \sigma = \sqrt{4} \cdot 8 = 16\,.$$
A kért valószínűség így az alábbiak szerint számítható:
$$P(X_\Sigma > 32) = 1 - \Phi\left(\frac{32-24}{16}\right) = 1 - \Phi(0{,}5) \approx 1 - 0{,}6915 = 0{,}3085\,.$$
Tehát annak a valószínűsége, hogy a vizsgált valószínűségi változók összege nagyobb, mint 32, kerekítve 0,3085.

b) Hasonlóan az összeghez, a vizsgált normális eloszlású valószínűségi változók $\overline{X}$ átlaga is normális eloszlású lesz. Számoljuk ki most az átlag várható értékét, és szórását:
$$E(\overline{X}) = m = 6,$$
valamint
$$D(\overline{X}) = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{4}} = 4\,.$$
A kért valószínűség így az alábbiak szerint számítható:
$$P(\overline{X} < 10) = \Phi\left(\frac{10-6}{4}\right) = \Phi(1) \approx 0{,}8413\,.$$
Tehát annak a valószínűsége, hogy a vizsgált valószínűségi változók átlaga kisebb, mint 10, kerekítve 0,8413.

⇐15.13. feladat

Leckezáró önellenőrző kérdéssor

1. Adott több független, azonos eloszlású valószínűségi változó. Melyik az a mennyiség, amit a valószínűségi változók száma nem befolyásol?

- [ ] összegük várható értéke
- [ ] összegük szórása
- [ ] átlaguk várható értéke
- [ ] átlaguk szórása

2. Hat független, Poisson-eloszlású valószínűségi változó várható értéke rendre 1, 2, 3, 4, 5 és 6. Mekkora lesz az összegük szórása?

- [ ] 21
- [ ] $4{,}5826$
- [ ] $10{,}8318$
- [ ] $1{,}8708$

3. Tizenhat független, de nem azonos eloszlású valószínűségi változó mindegyikének várható értéke 5, szórása pedig 20. Mekkora lesz az átlaguk szórása?

- [ ] 1
- [ ] 4
- [ ] 5
- [ ] 20

4. Egységnyi hosszúságú szakaszon választunk véletlenszerűen két pontot. A két pont a szakaszt három részre osztja. Legyen az X_1, X_2 és X_3 valószínűségi változó a három szakasz hossza. Számoljuk ki a három valószínűségi változó átlagának szórását!

- [ ] $0{,}5736$
- [ ] $0{,}0833$
- [ ] 1
- [ ] 0

5. Harminc független, exponenciális eloszlású valószínűségi változó mindegyikének szórása 10. Az alábbiak közül melyik lesz a legnagyobb?

- [ ] összegük várható értéke
- [ ] összegük szórása
- [ ] átlaguk várható értéke
- [ ] átlaguk szórása

26. LECKE

Valószínűségi változók kovarianciája

15.4. Valószínűségi változók kovarianciája

Általános esetben, ha X és Y nem feltétlenül függetlenek, akkor az összegük szórásnégyzete a következőképpen alakul:

15.10. tétel: Ha az X és Y valószínűségi változók szórásnégyzete létezik, akkor létezik az összegük szórásnégyzete is, és

$$D^2(X+Y) = D^2(X) + D^2(Y) + 2\cdot(E(X\cdot Y) - E(X)\cdot E(Y)) \,.$$

Bizonyítás:

$$\begin{aligned} D^2(X+Y) &= E\left[(X+Y-E(X+Y))^2\right] = E[(X+Y)^2] - E^2(X+Y) = \\ &= E(X^2+Y^2+2\cdot X\cdot Y) - E^2(X) - E^2(Y) - 2\cdot E(X)\cdot E(Y) = \\ &= E(X^2) - E^2(X) + E(Y^2) - E^2(Y) + 2\cdot(E(X\cdot Y) - E(X)\cdot E(Y)) = \\ &= D^2(X) + D^2(Y) + 2\cdot(E(X\cdot Y) - E(X)\cdot E(Y)) \,. \end{aligned}$$

□

15.1. definíció: **Valószínűségi változók kovarianciája**
A fenti tételben szereplő $E(X\cdot Y) - E(X)\cdot E(Y)$ kifejezést az X és Y valószínűségi változók kovarianciájának nevezzük, jelölése $\mathrm{cov}(X,Y)$. Azaz

$$\mathrm{cov}(X,Y) = E(X\cdot Y) - E(X)\cdot E(Y)\,.$$

Figyeljük meg, hogy $\operatorname{cov}(X,X) = D^2(X)$! A kovariancia használatával az X és Y valószínűségi változók összegének szórásnégyzete a következő alakban is felírható:

$$D^2(X+Y) = D^2(X) + D^2(Y) + 2\cdot\operatorname{cov}(X,Y)\,.$$

15.11. tétel: Ha az X és Y valószínűségi változók függetlenek, akkor kovarianciájuk nulla, azaz $\operatorname{cov}(X,Y) = 0$, feltéve persze, hogy az ehhez szükséges várható értékek léteznek.

Bizonyítás: A bizonyításhoz azt kell felhasználni, hogy ha X és Y függetlenek, akkor $E(X\cdot Y) = E(X)\cdot E(Y)$.

$$\operatorname{cov}(\mathrm{X},\mathrm{Y}) = E(X\cdot Y) - E(X)\cdot E(Y) = E(X)\cdot E(Y) - E(X)\cdot E(Y)\,.$$

$\square$

A fenti tétel megfordítása már nem igaz: ha X és Y kovarianciája 0, abból nem következik, hogy függetlenek.

15.14. feladat. Hatoldalú dobókockával dobunk. Legyen Y a dobott szám. Legyen ezen kívül X_2 a dobott szám kettővel, X_3 a dobott szám hárommal, és X_4 a dobott szám néggyel való osztás utáni maradéka, azaz ezek legyenek Y transzformáltjai:

$$X_2 = Y \bmod 2, \qquad X_3 = Y \bmod 3, \qquad X_4 = Y \bmod 4\,.$$

a) Vizsgáljuk meg X_2, X_3 és X_4 függetlenségi viszonyát!

b) Számoljuk ki X_2, X_3 és X_4 páronkénti kovarianciáját!

c) Számoljuk ki $X_3 + X_4$ szórását!

Megoldás: A feladat megoldásához az alábbi lépéseket fogjuk elvégezni:

a) Felírjuk a három X_i valószínűségi változó eloszlását és a későbbiekben szükséges várható értékét, majd a páronkénti együttes eloszlást. Az együttes eloszlás segítségével megnézzük, melyik páros lesz független.

b) A független páros(ok) kovarianciája nulla. A többinél fel kell írni a szorzat eloszlását, abból a várható értékét, s az előzőleg kiszámolt egyenkénti várható értékekkel megállapíthatóak a kovarianciák.

c) A két valószínűségi változó összegének szórása a kovarianciájuk segítségével felírható.

Nézzük akkor sorban az egyes részfeladatok megoldását:

a) Az első részfeladat első lépéseként a három valószínűségi változó eloszlását kell felírni. Azt kell végiggondolni, hogy a hatoldalú dobókocka dobásának hatféle kimenetele van: az egész számok 1-től 6-ig. Az egyes valószínűségi értékeket klasszikus valószínűségi mező segítségével számoljuk, tehát elég megszámolni az adott maradékot adó dobások számát. Táblázat segítségével foglalhatjuk ezeket össze:

A maradék	2-vel osztva	3-mal osztva	4-gyel osztva
0	2, 4 és 6	3 és 6	4
1	1, 3 és 5	1 és 4	1 és 5
2	nem lehet	2 és 5	2 és 6
3	nem lehet	nem lehet	3

A táblázat adott oszlopában és sorában azok a dobásértékek szerepelnek, melyek az oszlop szerinti osztás után a sor szerinti maradékot adják. A táblázat alapján az eloszlások könnyen felírhatók:

$$X_2 : \begin{cases} 0 & 1 \\ 1/2 & 1/2 \end{cases}, \qquad X_3 : \begin{cases} 0 & 1 & 2 \\ 1/3 & 1/3 & 1/3 \end{cases}, \qquad X_4 : \begin{cases} 0 & 1 & 2 & 3 \\ 1/6 & 1/3 & 1/3 & 1/6 \end{cases}.$$

Az eloszlások alapján a várható értékek már definíció szerint számolhatóak. Ezekre később, a kovariancia számításánál szükség lesz:

$$
\begin{aligned}
E(X_2) &= 0 \cdot 1/2 + 1 \cdot 1/2 = 1/2,\\
E(X_3) &= 0 \cdot 1/3 + 1 \cdot 1/3 + 2 \cdot 1/3 = 1,\\
E(X_4) &= 0 \cdot 1/6 + 1 \cdot 1/3 + 2 \cdot 1/3 + 3 \cdot 1/6 = 3/2\,.
\end{aligned}
$$

A közös eloszlást a leggyorsabban a következőféleképpen írhatjuk fel. Felrajzoljuk a táblázatos formát úgy, hogy mindenhova 0 értéket írunk. Végigmegyünk Y lehetséges értékein 1-től 6-ig, meghatározzuk a maradékokat, és a megfelelő maradékokkal jelzett sorban és oszlopban lévő értékhez $1/6$-ot adunk.
X_2 és X_3 közös eloszlása az alábbiak szerint fog alakulni:

$X_2\backslash X_3$	0	1	2
0	$1/6$	$1/6$	$1/6$
1	$1/6$	$1/6$	$1/6$

A táblázat mindegyik értéke $1/6$. Ahhoz, hogy X_2 és X_3 független legyen, elő kell állnia $1/6$-nak a peremértékek, azaz $1/2$ és $1/3$ szorzataként. Ez természetesen igaz is, így X_2 és X_3 független lesz.
Nézzük most meg X_2 és X_4 közös eloszlását.

$X_2\backslash X_4$	0	1	2	3
0	$1/6$	0	$1/3$	0
1	0	$1/3$	0	$1/6$

Vegyük észre, hogy amennyiben egy ilyen felírásnál a táblázatban 0 szerepel, a két valószínűségi változó nem lehet független. Itt is ha $X_2 = 0$ és $X_4 = 1$, akkor $P(X_2 = 0) = 1/2$ és $P(X_4 = 1) = 1/3$, de az együttes bekövetkezés valószínűsége $P(X_2 = 0, X_4 = 1) = 0$, és nem pedig $1/2 \cdot 1/3$. Tehát X_2 és X_4 *nem* lehet független.

Végül vizsgáljuk meg X_3 és X_4 együttes eloszlását:

$X_3 \backslash X_4$	0	1	2	3
0	0	0	1/6	1/6
1	1/6	1/6	0	0
2	0	1/6	1/6	0

Figyeljük meg, hogy ebben a táblázatban is szerepelnek 0 értékek, így az előző gondolatmenetet követve, X_3 és X_4 *sem* független.

b) Számoljuk ki most az egyes valószínűségi változók közötti kovarianciát. Mivel X_2 és X_3 független, így kovarianciájuk nulla:

$$\operatorname{cov}(X_2, X_3) = 0\,.$$

A másik két páros kovarianciájának számításához előbb a szorzatuk várható értékét, ahhoz pedig szorzatuk eloszlását kell felírnunk. Nézzük először $X_2 \cdot X_4$ és $X_3 \cdot X_4$ lehetséges értékeit. Az együttes eloszlás alapján látható, hogy az előbbi lehetséges értékei a 0, 1 és a 3, míg az utóbbié pedig a 0, 1, 2 és 4. Így a szorzatok eloszlása az alábbiak szerint alakul.

$$X_2 \cdot X_4 : \left\{ \begin{array}{ccc} 0 & 1 & 3 \\ 1/2 & 1/3 & 1/6 \end{array} \right. , \qquad X_3 \cdot X_4 : \left\{ \begin{array}{cccc} 0 & 1 & 2 & 4 \\ 1/2 & 1/6 & 1/6 & 1/6 \end{array} \right. .$$

Az eloszlásokból a várható értékek a definíció szerint gyorsan számolhatóak:

$$E(X_2 \cdot X_4) = 0 \cdot 1/2 + 1 \cdot 1/3 + 3 \cdot 1/6 = 5/6,$$
$$E(X_3 \cdot X_4) = 0 \cdot 1/2 + 1 \cdot 1/6 + 2 \cdot 1/6 + 4 \cdot 1/6 = 7/6\,.$$

Most, hogy megvannak a szorzatok várható értéke, így a kovarianciákat – a már számolt egyenkénti várható értékek segítségével – kiszámolhatjuk:

$$\operatorname{cov}(X_2, X_4) = E(X_2 \cdot X_4) - E(X_2) \cdot E(X_4) = 5/6 - 1/2 \cdot 3/2 = 10/12 - 9/12 = 1/12,$$
$$\operatorname{cov}(X_3, X_4) = E(X_3 \cdot X_4) - E(X_3) \cdot E(X_4) = 7/6 - 1 \cdot 3/2 = 7/6 - 9/6 = -1/3\,.$$

Összefoglalva arra az eredményre jutottunk, hogy X_2 és X_3 kovarianciája 0, mivel függetlenek, X_2 és X_4 kovarianciája $1/12$. Így a két valószínűségi változó értéke nem lehet független. Ez látszik is, hiszen X_4 értékéből tudunk X_2 értékére következtetni. A kovariancia pozitív értéke jelzi, hogy a két valószínűségi változó a várható értékeik körül inkább azonos, mint ellentétes irányban mozdulnak el. A $(0{,}0)$ és $(1{,}3)$ értékek inkább alátámasztják, míg a másik két értékpár inkább ellentmond a fenti magyarázatnak. A kovariancia kapott értéke a két hatás aktuális súlyozásából adódik. Figyeljük meg ezt az állítást a közös eloszlás táblázatán! Végül X_3 és X_4 kovarianciája viszont $-1/3$ lesz. A kovariancia értéke negatív is lehet! Ez azt jelenti, hogy amennyiben az egyik értéke a várható értékénél nagyobb lesz, akkor a másik inkább a várható értékénél *kisebb* értéket fog felvenni.

c) Végezetül számoljuk ki $X_3 + X_4$ szórását! Pontosabban az egyszerűség kedvéért előbb az összeg szórásnégyzetét számítjuk ki. A szórásnégyzet számításához előbb X_3 és X_4 szórásnégyzetére lesz szükségünk, ezekhez pedig előbb X_3^2 és X_4^2 várható értéke szükséges. Számítsuk ki előbb ezeket a már felírt eloszlások alapján!

$$
\begin{aligned}
E(X_3^2) &= 0^2 \cdot 1/3 + 1^2 \cdot 1/3 + 2^2 \cdot 1/3 = 5/3, \\
E(X_4^2) &= 0^2 \cdot 1/6 + 1^2 \cdot 1/3 + 2^2 \cdot 1/3 + 3^2 \cdot 1/6 = 19/6\,.
\end{aligned}
$$

A négyzetek várható értékeiből a szórásnégyzetek kiszámíthatóak:

$$
\begin{aligned}
D^2(X_3) &= 5/3 - 1^2 = 2/3, \\
D^2(X_4) &= 19/6 - 9/4 = 38/12 - 27/12 = 11/12\,.
\end{aligned}
$$

Mivel X_3 és X_4 *nem* függetlenek, az összeg szórásnégyzetét a kovariancia segítségével írhatjuk fel:

$$
\begin{aligned}
D^2(X_3 + X_4) &= D^2(X_3) + D^2(X_4) + 2 \cdot \operatorname{cov}(X_3, X_4) \\
&= 5/3 + 19/6 + 2 \cdot (-1/3) = 25/6\,.
\end{aligned}
$$

Tehát X_3 és X_4 összegének szórása $\sqrt{25/6} \approx 2{,}0412$ lesz. Vegyük észre, hogy mivel a kovariancia negatív, így az összeg szórása kisebb lesz, mint független esetben. Tehát ha az egyik valószínűségi változó a várható értékénél nagyobb lesz, és ezt a másik egy a vártnál kisebb értékkel kompenzálja, akkor ez a kompenzáció az összeg ingadozását is csökkenteni fogja.

⇐15.14. feladat

Aktivitás: Keressen példát olyan véletlen jelenségekre, ahol a jelenségeket leíró valószínűségi változók kovarianciája negatív, és olyat is, ahol a kovariancia pozitív.

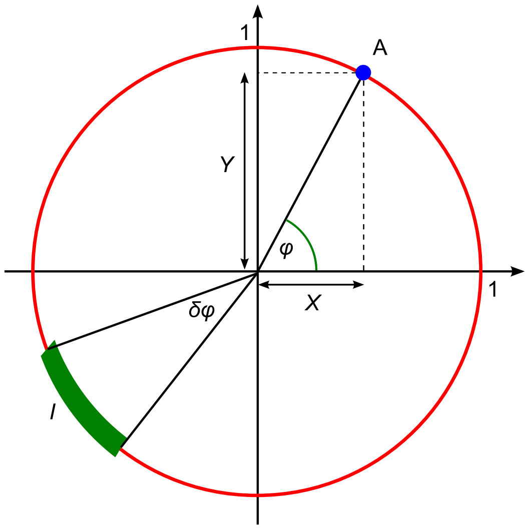
15.15. feladat. A koordináta-rendszerben lévő origó középpontú egység sugarú kör kerületén válasszunk véletlenszerűen egy pontot. Legyen X és Y a pont x és y koordinátája.

a) Számoljuk ki a két valószínűségi változó kovarianciáját!

b) Igaz-e, hogy a két változó független?

Megoldás:

Vegyük észre, hogy most is két olyan valószínűségi változót kell vizsgálnunk, melyek egy harmadik transzformáltjai. Ennek meglátásához kössük össze a választott pontot az origóval. Legyen φ a kapott szakasz és az x tengely által bezárt szög. Mivel a pontot a körvonalon választjuk véletlenszerűen, így az adott körívre való esés valószínűsége arányos lesz a körív hosszával, s ezáltal a körívhez tartozó központi szög nagyságával. A választott ponthoz tartozó φ szög, mint valószínűségi változó, egyenletes eloszlású lesz 0 és 2π között.



15.3. ábra. *A véletlenszerűen választott A pont koordinátái X és Y. Iránya φ szöget zár be*

A bevezetett φ valószínűségi változó segítségével írhatjuk fel mind X-et, mind Y-t az alábbi módon:

$$X = \cos\varphi \qquad \text{és} \qquad Y = \sin\varphi\,.$$

Nem csak a két valószínűségi változó, hanem a szorzatuk is felírható ebben az esetben transzformáltként:

$$X \cdot Y = \cos\varphi\sin\varphi\,.$$

a) A kovariancia számításához ki kell számolni X, Y és a szorzatuk várható értékét. Mivel φ egyenletes eloszlású, transzformáltjainak várható értéke az alábbi, a 18. leckében megismert képlet szerint számítható:

$$E(h(\varphi)) = \frac{1}{2\pi}\int_0^{2\pi} h(\varphi)\,\mathrm{d}\varphi$$

Először számoljuk ki X várható értékét:

$$E(X) = E(\cos\varphi) = \frac{1}{2\pi}\int_0^{2\pi} \cos\varphi\,\mathrm{d}\varphi = \frac{1}{2\pi}\left[\sin\varphi\right]_0^{2\pi} = 0\,.$$

Hasonlóan számíthatjuk Y várható értékét:

$$E(Y) = E(\sin\varphi) = \frac{1}{2\pi}\int_0^{2\pi} \sin\varphi\,\mathrm{d}\varphi = \frac{1}{2\pi}\left[-\cos\varphi\right]_0^{2\pi} = 0\,.$$

A szorzat várható értékének számításánál is alkalmazhatjuk a transzformált várható értékének képletét:

$$E(X\cdot Y)=E(\cos\varphi\sin\varphi)=\frac{1}{2\pi}\int_0^{2\pi}\cos\varphi\sin\varphi\,\mathrm{d}\varphi=$$

$$=\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{2}\sin 2\varphi\,\mathrm{d}\varphi=\frac{1}{4\pi}\left[-\frac{\cos 2\varphi}{2}\right]_0^{2\pi}=0\,.$$

Megállapíthatjuk, hogy nem meglepő módon X, Y és szorzatuk várható értéke is 0 lett. Kovarianciájuk is 0 lesz így, hiszen:

$$\mathrm{cov}(\mathrm{X,Y})=E(X\cdot Y)-E(X)\cdot E(Y)=0-0\cdot 0=0\,.$$

b) Vizsgáljuk meg most a két változó függetlenségét! Vegyük észre, hogy mind X, mind Y -1 és 1 között veszi fel az értékeit. Láthattuk az előző példánál, hogy attól, hogy ugyanabból a valószínűségi változóból lettek transzformálva, még lehetnek függetlenek. Itt viszont az ábrát megfigyelve a következőket vehetjük észre: Legyen $x=y=-\frac{\sqrt{2}}{2}$. Ebben az esetben nem lehet egyszerre $X=\cos\varphi$ és $Y=\sin\varphi$ is kisebb, mint $-\frac{\sqrt{2}}{2}$, azaz $P(X<x$ és $Y<y)=0$. Viszont külön-külön lehetnek, így $P(X<x)\cdot P(Y<y)>0$. Ez azt jelenti, hogy X és Y nem független.
A függetlenséget másképp is meggondolhatjuk. Következtethetünk-e X valamilyen értékéből Y értékére? Mi van akkor, ha X értéke az 1-hez közel van? Ekkor bizony Y értéke pedig a 0-hoz lesz közel. Így a két változó nem lehet független.

⇐15.15. feladat

15-8. önálló feladat: Legyen Y egyenletes eloszlású a $[0;1]$ intervallumon. Vegyük az Y három transzformáltját: Legyen $X_1=2\cdot Y$, $X_2=1-Y$, illetve $X_3=Y^2$. Számolja ki Y kovarianciáját a transzformáltakkal!

Leckezáró önellenőrző kérdéssor

1. Legyenek X és Y olyan valószínűségi változók, melyek kovarianciája 0. Ebben az esetben
   - ☐ X és Y összegének szórásnégyzete megegyezik a szórásnégyzeteik összegével.
   - ☐ X és Y összegének szórásnégyzete kisebb, mint a szórásnégyzeteik összege.
   - ☐ X és Y összegének szórásnégyzete nagyobb, mint a szórásnégyzeteik összege.
   - ☐ X és Y függetlenek.

2. Legyenek X és Y olyan valószínűségi változók, melyek csak pozitív értéket vehetnek fel. Ekkor a kovarianciájuk
   - ☐ nem lehet negatív,
   - ☐ nem meghatározható,
   - ☐ csak pozitív lehet,
   - ☐ lehet pozitív és negatív és 0 is.

3. A 32 lapos magyarkártya-pakliból húzunk egy kártyát. Legyen X a húzott királyok, Y a húzott ászok száma. Mekkora lesz X és Y kovarianciája?
   - ☐ 0
   - ☐ $1/64$
   - ☐ $-1/64$
   - ☐ $1/8$

4. Tegyük fel a legutolsó kidolgozott feladatban, hogy ismerjük X értékéről, hogy kisebb mint $\frac{\sqrt{2}}{2}$. Mekkora a valószínűsége annak, hogy Y értéke is kisebb lesz, mint $\frac{\sqrt{2}}{2}$?
   - ☐ 0
   - ☐ 1
   - ☐ $1/2$
   - ☐ Nem kiszámítható.

27. LECKE

Modulzáró feladatok 4.

16. Modulzáró feladatok

Start.

A modulzáró megoldására 1 órája van. Legalább 4 tizedesjegy pontossággal adja meg az eredményeket.

1. Egy dobókockát kétszer elgurítunk. Legyen X a dobott számok összegének hattal való osztás utáni maradéka. Mennyi lesz X várható értéke?
2. Egységnyi sugarú félköríven választunk véletlenszerűen egy pontot. Legyen X_1 a választott pont távolsága a félkörív egyik végpontjától. Mennyi lesz X_1 várható értéke?
3. Az előző feladatban legyen X_2 a választott pont távolsága a félkörív másik végpontjától. Mennyi lesz $X_1^2 + X_2^2$ szórása?
4. Egység oldalú négyzet kerületén választunk egy pontot véletlenszerűen. Legyen X_1 és X_2 a pont távolsága egy előre kiválasztott oldal két végpontjától. Mennyi a valószínűsége annak, hogy $X_1 < 1$ és $X_2 < 2$?
5. Az előző feladatban tegyük fel, hogy X_1 kisebb mint 1. Mekkora lehet annak a valószínűsége, hogy X_2 is kisebb lesz mint 1?
6. Egy vödörben négy darab hat- és ugyanennyi nyolcoldalú dobókocka van. A vödörből véletlenszerűen húzunk két dobókockát, majd elgurítjuk őket. Adjuk meg a dobott számok összegének várható értékét!
7. Mekkora lesz az előző feladatban a dobott számok átlagának várható értéke?
8. Négy Poisson, és hat exponenciális eloszlású valószínűségi változó független, s mindegyik várható értéke 4. Adjuk meg összegük szórását!
9. Két hatoldalú dobókockával dobunk. Adjuk meg a dobott számok összegének és átlagának kovarianciáját!

Stop.

Pontok:

V. MODUL

Egyenlőtlenségek

28. LECKE

A Markov-egyenlőtlenség

17. A Markov- és a Csebisev-egyenlőtlenség

Sok esetben előfordul, hogy nem áll minden információ rendelkezésre ahhoz, hogy egy valószínűséget pontosan meg tudjunk határozni. Ilyenkor a valószínűségi változó néhány jellemzőjének (várható érték, szórás) ismeretében csupán a valószínűség becslésére szorítkozhatunk.

Ebben a fejezetben először annak valószínűségére mutatunk felső becslést, hogy egy nem negatív értékű valószínűségi változó értéke nagyobb vagy egyenlő, mint egy adott szám. Látni fogjuk, hogy ez a valószínűség nem lehet akármekkora, a várható érték alapján adható rá egy felső korlát. Ez azt is jelenti, hogy a valószínűségi változó csak viszonylag kis valószínűséggel vehet fel a várható értékhez képest nagy értékeket.

Ezután annak valószínűségére adunk becslést, hogy a valószínűségi változó a várható értéktől legalább egy adott értékkel eltér. Ez is azt mutatja, hogy túlságosan nem lehet eltérni a várható értéktől, azaz a nagy eltérés valószínűsége kicsi.

17.1. A Markov-egyenlőtlenség

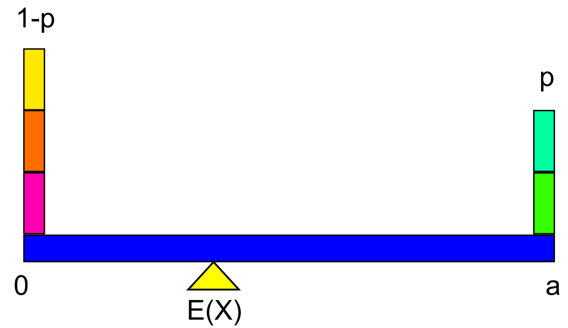
17.1. tétel:

Ha X olyan nem negatív értékeket felvevő valószínűségi változó, melynek van várható értéke, és az a egy tetszőleges pozitív valós szám, akkor

$$P\left(X \geq a\right) \leq \frac{E(X)}{a}.$$

Bizonyítás:

Az állítást külön-külön bizonyítjuk diszkrét, illetve folytonos eloszlású valószínűségi változóra.



17.1. ábra. *A Markov-egyenlőtlenség mérlege egyensúlyban van. Ha egy súlyt jobbra szeretnénk tolni, akkor egy másikat balra kell.*

1. Legyen X a feltételeknek megfelelő diszkrét eloszlású valószínűségi változó. A lehetséges értékei legyenek $x_1, x_2, \ldots$, az ezekhez tartozó valószínűségek pedig $p_1, p_2, \ldots$. A bizonyításhoz felírjuk X várható értékét, majd azt alulról becsüljük:

$$E(X) = \sum_{k=1}^{\infty} x_k \cdot p_k \geq \sum_{x_k \geq a} x_k \cdot p_k \geq \sum_{x_k \geq a} a \cdot p_k = a \cdot \sum_{x_k \geq a} p_k = a \cdot P(X \geq a)\,.$$

Az alábbi egyenlőtlenséget kaptuk, melyből átrendezéssel adódik a tétel állítása:

$$E(X) \geq a \cdot P(X \geq a) \qquad \Longrightarrow \qquad P\left(X \geq a\right) \leq \frac{E(X)}{a}\,.$$

2. Legyen most X a feltételeknek megfelelő folytonos eloszlású valószínűségi változó, a sűrűségfüggvénye pedig legyen $f(x)$. Ugyanazt az eljárást fogjuk követni, mint diszkrét esetben:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x = \int_{0}^{\infty} x \cdot f(x)\,\mathrm{d}x \geq \int_{a}^{\infty} x \cdot f(x)\,\mathrm{d}x \geq \int_{a}^{\infty} a \cdot f(x)\,\mathrm{d}x =$$

$$= a \cdot \int_{a}^{\infty} f(x)\,\mathrm{d}x = a \cdot P(X \geq a)\,.$$

Ugyanolyan egyenlőtlenséget kaptuk, mint az előbb, most is átrendezéssel adódik a tétel állítása:

$$E(X) \geq a \cdot P(X \geq a) \qquad \Longrightarrow \qquad P\left(X \geq a\right) \leq \frac{E(X)}{a}\,.$$

$\square$

A becslést az $X > a$ típusú, szigorú egyenlőtlenségre is alkalmazni tudjuk, hiszen

$$P(X > a) \leq P(X \geq a) \leq \frac{E(X)}{a}\,.$$

17.1. feladat. Egy izzó élettartamát vizsgáljuk. Az élettartam eloszlásáról nem tudunk semmit, csak hogy várható értéke 1000 óra.

a) Mekkora lehet annak a valószínűsége, hogy az izzó eléri a 10000 óra élettartamot?

b) Hogyan változik az eredmény, ha feltételezhetjük, hogy az élettartam eloszlása exponenciális?

c) Mit tudunk akkor mondani, ha tudjuk, hogy az élettartam eloszlása egyenletes?

Megoldás:

a) Jelöljük X-szel az izzó élettartamát. Nem ismerjük X eloszlását, így a kérdésben szereplő $X \geq 10000$ valószínűségét csak becsülni tudjuk. Ahhoz, hogy a Markov-egyenlőtlenség segítségével becsülni tudjunk, két feltételnek kell teljesülnie. Egyrészt X nem vehet fel negatív értéket, másrészt ismerni kell X várható értékét. Mindkét feltétel teljesül, hiszen élettartamról lévén szó $X \geq 0$, valamint $E(X) = 1000$.
Annak a valószínűségét, hogy $X \geq 10000$, így meg tudjuk becsülni az alábbiak szerint:

$$P(X \geq 10000) \leq \frac{E(X)}{10000} = \frac{1000}{10000} = 0{,}1\,.$$

Tehát annak a valószínűsége, hogy az izzó legalább 10000 óráig ég, legfeljebb 0,1. A kapott 0,1 csak egy felső korlát a valószínűségre: ennél lényegesen kisebb valószínűségek is előfordulhatnak.

b) Most vizsgáljuk meg azt az esetet, amikor az élettartam eloszlása exponenciális. Mivel X várható értéke ismert, az őt leíró exponenciális eloszlás λ paraméterét is kiszámolhatjuk:

$$E(X) = 1000, \qquad \text{tehát} \qquad \lambda = \frac{1}{1000}\,.$$

Az ismert λ paraméter segítségével felírhatjuk X eloszlásfüggvényét:

$$F(x) = \begin{cases} 1 - e^{-\frac{x}{1000}}, & \text{ha } x > 0; \\ 0 & \text{egyébként}\,. \end{cases}$$

Így, hogy X eloszlása ismert, ki tudjuk számolni az $X \geq 10000$ esemény valószínűségét:

$$P(X \geq 10000) = 1 - F(10000) = 1 - \left[1 - e^{-\frac{10000}{1000}}\right] = e^{-10} \approx 4{,}54 \cdot 10^{-5}\,.$$

Tehát annak a valószínűsége, hogy az izzó élettartama eléri a 10000 órát, körülbelül 4 és fél ezred százalék. Vessük össze ezt az értéket a Markov-egyenlőtlenség segítségével kapott 10%-os értékkel. Ha a valószínűségi változónk eloszlását nem ismerjük, akkor akár több ezerszeres különbségek is előjöhetnek a felső korlát, és a tényleges valószínűség között.

c) Mit tudunk mondani abban az esetben, ha tudjuk, hogy X eloszlása egyenletes? Bár az adott eloszlás élettartam esetén meglepő lehet, mégis vizsgáljuk meg a lehetőségeinket ebben az esetben. Ahhoz, hogy egy egyenletes eloszlású valószínűségi változót leírjunk, két paraméter megadására van szükség: az a és a b paraméter adja meg annak az $[a;b]$ intervallumnak az alsó és felső határát, ahonnan a valószínűségi változónkat véletlenszerűen választjuk. A jelenlegi X valószínűségi változónak csak egy paraméterét, a várható értékét ismerjük, így a és b között az alábbi összefüggést tudjuk felírni:

$$1000 = E(X) = \frac{a+b}{2}, \qquad \text{tehát} \qquad a = 2000 - b\,.$$

Jelen formában, a szórás ismerete nélkül pontosabb információnk nincs a és b-ről. Két dolgot azonban fel kell ismerni: egyrészt mivel élettartamról van szó, X értéke nem lehet negatív, azaz szükségszerűen $a \geq 0$. Ezen kívül mivel a a bal, és b a jobb oldali intervallum végpont, így $a \leq b$. A két egyenlőtlenséget összevetve kapjuk, hogy

$$0 \leq a \leq 1000, \qquad \text{valamint} \qquad 1000 \leq b \leq 2000\,.$$

Mit mond a legutolsó egyenlőtlenség? A b paraméter nem lehet 2000-nél nagyobb. Ez azt jelenti, hogy X-et semmiképp nem választhatjuk 2000 feletti számnak, hiszen nincs olyan, a feltételeknek megfelelő $[a;b]$ intervallum, amiből ezt megtehetjük. Ebből viszont közvetlen következik, hogy

$$P(X \geq 10000) = 0\,.$$

Tehát 0 annak a valószínűsége, hogy ebben az esetben az izzó élettartama eléri a 10000 órát.

⇐17.1. feladat

Aktivitás: Állítsunk össze egy gondolatkísérletet annak meghatározására, hogy a háztartásunkban fellelhető izzók élettartamainak eloszlásait valamilyen módon jellemezni tudjuk! Figyeljük meg, hogy az egyes izzótípusok esetén a gyártók miben mérik az élettartamot!

17.2. feladat. A helyi élelmiszer-ellátó üzletben sajtos baromfipárizsit vásárolunk. Az eladótól 15 dekagrammnyi felvágottat kérünk. Tapasztalatból tudjuk, hogy a mérlegre került darabok össztömegének várható értéke 18 dekagramm, de előfordul, hogy 20 dekánál is több kerül oda.

a) Adjunk becslést annak a valószínűségére, hogy a kimért sajtos baromfipárizsi tömege eléri a 20 dekagrammot!

b) Hogyan becsülhetjük ezt a valószínűséget, ha tudjuk, hogy a kiadott áru tömege *mindig* eléri a 15 dekagrammot?

c) Számoljunk utána az eredménynek, ha tudjuk, hogy a boltos *mindig* 18 szeletet vág, s az egyes szeletek tömege független, és normális eloszlást követ 2 gramm szórással.

Megoldás:

a) Jelölje az X valószínűségi változó az eladó által kimért áru tömegét. Láthatjuk, hogy X nem lehet negatív, hiszen a tömeg sosem negatív. Mivel nem ismerjük X eloszlását, de ismerjük a várható értékét, becsülhetjük az $X \geq 20$ esemény valószínűségét a Markov-egyenlőtlenség segítségével:

$$P(X \geq 20) \leq \frac{18}{20} = 0{,}9\,.$$

Ez azt jelenti, hogy annak a valószínűsége, hogy 20 dekánál több párizsit kapunk, legfeljebb 90% lehet. Természetesen ahhoz, hogy ilyen nagyságrendű valószínűséget elérjünk, néha nagyon kevés párizsit kell kapnunk, amit azért nehezen lehet elképzelni.

b) Mi a helyzet abban az esetben, ha tudjuk, hogy *mindig* megkapjuk legalább az általunk kért 15 dekagramm párizsit. Miként tudjuk ezt az új információt felhasználni arra, hogy a becslésünket pontosítsuk? Hívjunk segítségül egy új valószínűségi változót. Legyen

$$Y = X - 15, \qquad \text{azaz} \qquad X = Y + 15\,.$$

Mivel $X \geq 15$, így $Y \geq 0$, azaz Y-ra használhatjuk a Markov-egyenlőtlenséget. Számítsuk ki Y várható értékét:

$$E(Y) = E(X - 15) = E(X) - 15 = 3\,.$$

A várható érték ismeretében már alkalmazhatjuk a Markov-egyenlőtlenséget:

$$P(X \geq 20) = P(Y + 15 \geq 20) = P(Y \geq 5) \leq \frac{3}{5} = 0{,}6\,.$$

Láthatjuk, hogy egy új információ tudatában a becslést pontosítani tudjuk. Annak a valószínűsége, hogy 20 dekánál több felvágottat kapunk, legfeljebb 60% lehet.

c) A feladat utolsó részében a valószínűség számítását egy pontosabb modell szerint végezzük. Az eseményeket most nem egy, hanem 18 valószínűségi változó segítségével írjuk le. Az X valószínűségi változó lesz a 18 összege. Ismerjük X várható értékét, ami 18. Szórását is kiszámolhatjuk, hiszen minden egyes szelet tömege független és szórása $\sigma = 0{,}2$ dekagramm, így összegük szórása:

$$D(X) = \sqrt{n} \cdot \sigma = \sqrt{18} \cdot 0{,}2 \approx 0{,}8485\,.$$

Normális eloszlású valószínűségi változók összegének eloszlása is normális lesz. Ez azt jelenti, hogy X is normális eloszlású lesz 18 várható értékkel, és $\sqrt{18} \cdot 2$ szórással. Ki tudjuk ekkor számolni pontosan a kérdezett valószínűséget:

$$P(X \geq 20) \approx 1 - \Phi\left(\frac{20 - 18}{0{,}8485}\right) \approx 1 - \Phi(0{,}2357) \approx 0{,}0092\,.$$

Tehát annak a valószínűsége, hogy 20 dekagrammnál több felvágottat kapunk, kicsit kevesebb, mint 1%. Ez lényegesen kevesebb, mint a becsléssel kapott két felső határ, bár itt nem feltételeztük azt, hogy a 15 dekagrammot mindenképp megkapjuk.

⇐17.2. feladat

Aktivitás: Figyeljük meg, és modellezzük a boltban történő kiszolgálást a felvágottas pultnál! Jegyezzük le a kapott áruk tömegét, s becsüljük meg a tömeg várható értékét, s annak a valószínűségét, hogy egy nagyobb értéket átlép. Hasonlítsuk össze a kapott eredményeket a feladat eredményeivel!

A Markov-egyenlőtlenség segítségével felülről tudjuk becsülni annak a valószínűségét, hogy a nemnegatív valószínűségi változónk egy bizonyos értéket elér. Azáltal, hogy ezt a $P(X \geq a)$ valószínűséget felülről tudjuk becsülni, a $P(X < a)$ valószínűséget is becsülhetjük, méghozzá alulról:

17.2. tétel:

Ha X olyan nem negatív értékeket felvevő valószínűségi változó, melynek van várható értéke, és a egy tetszőleges pozitív valós szám, akkor

$$P\left(X < a\right) \geq 1 - \frac{E(X)}{a}.$$

Bizonyítás: A tétel bizonyításához a Markov-egyenlőtlenség mindkét oldalát -1-gyel szorozzuk meg. Figyeljünk arra, hogy ekkor az egyenlőtlenség iránya megváltozik:

$$-P\left(X \geq a\right) \geq -\frac{E(X)}{a}.$$

Adjunk hozzá mindkét oldalhoz 1-et, s alakítsuk át a bal oldalt.

$$P\left(X < a\right) = 1 - P\left(X \geq a\right) \geq 1 - \frac{E(X)}{a}.$$

Az egyenlőtlenség két végét összehasonlítva a tétel állítását kapjuk. □

17-1. önálló feladat: Lássuk be, hogy az $X \leq a$ esemény valószínűségét is $1 - \dfrac{E(X)}{a}$-val tudjuk alulról becsülni!

17.3. feladat. Az X valószínűségi változó várható értéke 8, szórása 2.

a) Adjunk becslést annak a valószínűségére, hogy X^2 értéke kisebb, mint 144.

b) Mi a válasz, ha tudjuk, hogy X eloszlása normális?

c) Mi a helyzet, ha X eloszlása egyenletes?

Megoldás:

a) Az $X^2 < 144$ esemény valószínűségét kérdezi a feladat. Mivel az X^2 mint valószínűségi változó nem lehet negatív, ezért a valószínűség becslésére alkalmazhatjuk a Markov-egyenlőtlenséget. Viszont az egyenlőtlenség alkalmazásához előbb X^2 várható értékét kell kiszámolnunk X szórásának és várható értékének segítségével. Elevenítsük fel a szórásnégyzet definícióját:

$$D^2(X) = E(X^2) - E^2(X)\,.$$

A definícióból X^2 várható értéke kifejezhető:

$$E(X^2) = E^2(X) + D^2(X) = 8^2 + 2^2 = 68\,.$$

Mivel X^2 várható értékét kiszámoltuk, alkalmazhatjuk az $X^2 < 144$ egyenlőtlenség valószínűségének becslésére a Markov-egyenlőtlenséget:

$$P(X^2 < 144) \geq 1 - \frac{68}{144} = \frac{19}{36} \approx 0{,}5278\,.$$

Tehát annak a valószínűsége, hogy X^2 értéke kisebb, mint 144, nem lehet kevesebb, mint $52{,}78\%$. A konkrét valószínűség értékek ennél csak nagyobbak lehetnek.

b) Mi a helyzet akkor, ha tudjuk, hogy X eloszlása normális? Ebben az esetben X^2 eloszlását is meghatározhatjuk, de egyszerűbb, ha belegondolunk abba, milyen X értékek esetén lesz X^2 értéke kisebb, mint 144: X ekkor nem lehet 12-nél nagyobb, de -12-nél kisebb sem! Tehát $-12 < X < 12$ kell, hogy legyen. Mivel X eloszlása normális annak a valószínűségét kiszámolhatjuk, hogy X -12 és 12 közé esik:

$$P(X^2 < 144) = P(-12 < X < 12) = \Phi\left(\frac{12-8}{2}\right) - \Phi\left(\frac{-12-8}{2}\right)$$
$$= \Phi(2) - \Phi(-10) \approx 0{,}9772 - 0 = 0{,}9772\,.$$

Tehát annak a valószínűsége, hogy normális eloszlás esetén X^2 értéke kisebb, mint 144, majdnem 98%. Vessük össze a becslés értékével: a kapott valószínűség jóval nagyobb $0{,}5278$-nál.

c) Vizsgáljuk meg most azt az esetet, amikor X eloszlásáról tudjuk, hogy egyenletes. Mivel az eloszlás várható értékét és szórását is ismerjük, az egyenletes eloszlás felírásához szükséges a-t és b-t meghatározhatjuk úgy, hogy a várható érték és a szórás képleteit felírjuk:

$$8 = E(X) = \frac{a+b}{2}, \qquad 2 = D(X) = \frac{b-a}{\sqrt{12}}\,.$$

A két felírt egyenletben csak a és b ismeretlen, így azok meghatározhatóak:

$$a = 8 - 2\sqrt{3} = 4{,}5359, \qquad b = 8 + 2\sqrt{3} = 11{,}4641\,.$$

Ez azt jelenti, hogy mivel X egyenletes eloszlású, $4{,}5$ és $11{,}5$ között választjuk véletlenszerűen. Ebben az esetben viszont X mindenképp -12 és 12 között lesz:

$$P(X^2 < 144) = P(-12 < X < 12) = 1\,.$$

Tehát annak a valószínűsége, hogy X^2 értéke 144 alatt lesz ebben az esetben 1. Láthatjuk, hogy ugyan a Markov-egyenlőtlenséggel végzett becslés itt is teljesül, a tényleges valószínűség nagyon messze van a becslés által szabott határtól.

⇐17.3. feladat

Leckezáró önellenőrző kérdéssor

1. Az $X > 10$ esemény valószínűségét Markov-egyenlőtlenség segítségével becsültük. Az eredmény -2 lett.
   - ☐ Az eredmény nem jó, mert a valószínűség mindig 0 és 1 közötti szám.
   - ☐ Az eredmény jó, mert a Markov-egyenlőtlenség a vizsgált eseményre egy alsó korlátot ad.
   - ☐ Az eredmény jó, ha a Markov-egyenlőtlenség a vizsgált eseményre egy felső korlátot ad.
   - ☐ Az eredmény nem jó, mert a Markov-egyenlőtlenség a vizsgált eseményre egy felső korlátot ad.

2. Egy termék élettartamát egy valószínűségi változóval próbáljuk leírni. A valószínűségi változó eloszlásáról nincs pontosabb információnk. Markov-egyenlőtlenséggel becsülve mekkora lehet annak a valószínűsége, hogy a terméket a várható élettartamának legalább duplájáig fogjuk tudni használni?

   ☐ ≤ 2 ☐ $\geq \frac{1}{2}$ ☐ $\leq \frac{1}{2}$ ☐ ≥ -2

3. Egy lakás téli fűtésköltségének várható értéke 200000 forint. Markov-egyenlőtlenséggel becsülve mekkora lehet annak a valószínűsége, hogy egymillió forintból ki tudjuk fizetni a jövő téli fűtést, ha a költség nagyságának eloszlását nem ismerjük?

   ☐ ≈ 1 ☐ < 5 ☐ $\geq \frac{4}{5}$ ☐ $\leq \frac{1}{5}$

4. Adjunk becslést a Markov-egyenlőtlenség segítségével annak a valószínűségére, hogy hat kockadobás eredményének összege legalább 30! Vegyük figyelembe, hogy mennyi a legkisebb dobás értéke!

   ☐ $\leq \frac{7}{10}$ ☐ $\geq \frac{3}{10}$ ☐ $\leq \frac{5}{8}$ ☐ $\leq \frac{3}{7}$

5. Tizenhat számot választunk egymástól függetlenül, véletlenszerűen a $[0; 2]$ intervallumból. Mekkora lehet annak a valószínűsége, hogy a választott számok átlaga 2-nél kisebb lesz.

   ☐ $\leq \frac{1}{8}$ ☐ $\geq \frac{7}{8}$ ☐ $\leq \frac{1}{2}$ ☐ $= 1$

29. LECKE

A Csebisev-egyenlőtlenség

17.2. A Csebisev-egyenlőtlenség

A Markov-egyenlőtlenséget akkor használhattuk bizonyos valószínűségek becslésére, ha az ismeretlen eloszlású valószínűségi változó nem vehetett fel negatív értéket, valamint ismertük a várható értékét. Előfordulhat, hogy az ismeretlen eloszlású valószínűségi változónak nem csak a várható értékét, hanem a szórását is ismerjük. Ebben az esetben tudunk korlátokat adni bizonyos valószínűségekre akkor is, ha a valószínűségi változó értéke negatív értéket is felvehet. Erre szolgál a most tárgyalandó Csebisev-egyenlőtlenség. Emlékezzünk vissza: a valószínűségi változó értékeinek a várható értéktől való eltérését a szórással jellemezzük. A következőkben azt fogjuk látni, hogy a várható értéktől a szórás sokszorosával való eltérés valószínűsége nem lehet bármekkora, a nagyobb eltérés valószínűségének felső korlátja van.

17.3. tétel: Legyen X olyan valószínűségi változó, melynek van várható értéke és szórása, legyen továbbá λ tetszőleges pozitív szám. Ekkor

$$P\left(|X - E(X)| \geq \lambda \cdot D(X)\right) \leq \frac{1}{\lambda^2}.$$

Bizonyítás:

A bizonyításhoz a Markov-egyenlőtlenséget fogjuk felhasználni. Az ott szereplő nemnegatív Y valószínűségi változó legyen az X transzformáltja:

$$Y = (X - E(X))^2 .$$

Y-nak létezik a várható értéke, hiszen

$$E(Y) = E\left[(X - E(X))^2\right] = D^2(X).$$

Erről tudjuk, hogy létezik, hiszen X-nek van szórása, így szórásnégyzete is. Alkalmazhatjuk tehát a Markov-egyenlőtlenséget. A Markov-egyenlőtlenségben lévő a szám legyen:

$$a = \lambda^2 \cdot D^2(X).$$

Így felírva a Markov-egyenlőtlenséget:

$$P(Y > a) = P\left((X - E(X))^2 \geq \lambda^2 \cdot D^2(X)\right) \leq \frac{E(Y)}{a} = \frac{E[(X - E(X))^2]}{\lambda^2 \cdot D^2(X)} = \frac{D^2(X)}{\lambda^2 \cdot D^2(X)} = \frac{1}{\lambda^2}\,.$$

A bal oldal átalakításhoz használjuk fel, hogy

$$(X - E(X))^2 \geq \lambda^2 \cdot D^2(X) \qquad \text{akkor és csak akkor teljesül, amikor} \qquad |X - E(X)| \geq \lambda \cdot D(X)\,.$$

A jobb oldal egyszerűsítéséhez pedig felhasználhatjuk azt, hogy

$$E[(X - E(X))^2] = D^2(X)\,.$$

A tétel állítása így adódik. □

A Csebisev-egyenlőtlenséget gyakran alkalmazzuk annak becslésére, hogy az X valószínűségi változó értéke milyen valószínűséggel esik egy adott, várható érték körüli szimmetrikus intervallumba. Ez az esemény a Csebisev-egyenlőtlenségben szereplőnek a komplementere, így az egyenlőtlenséget az alábbi, az eredeti alakkal ekvivalens formában használjuk:

$$P\left(|X - E(X)| < \lambda \cdot D(X)\right) \geq 1 - \frac{1}{\lambda^2}\,.$$

A Csebisev-egyenlőtlenséget fel tudjuk írni olyan formában is, amikor a várható értéktől való eltérés nem a szórás többszöröseként van megadva, hanem önmagában. Ekkor az $\varepsilon = \lambda \cdot D(X)$ helyettesítést használva az egyenlőtlenség alábbi formáit kaphatjuk meg:

$$P\left(|X - E(X)| \geq \varepsilon\right) \leq \frac{D^2(X)}{\varepsilon^2},$$

$$P\left(|X - E(X)| < \varepsilon\right) \geq 1 - \frac{D^2(X)}{\varepsilon^2}\,.$$

17-2. önálló feladat: Lássuk be, hogy a Csebisev-egyenlőtlenség segítségével a $>$ ε és $\leq$ ε típusú egyenlőtlenségek valószínűségét is becsülni tudjuk!

17.4. feladat. Az X valószínűségi változó várható értéke 10, szórása pedig 5. Adjunk becslést annak a valószínűségére, hogy

a) X értéke legfeljebb -5, vagy legalább 25,

b) X értéke 5 és 15 között van,

c) X eltérése a várható értékétől legalább 20,

d) X eltérése a várható értékétől 20-nál kevesebb,

e) X eltérése a várható értékétől legalább a szórásának háromszorosa,

f) X eltérése a várható értékétől legfeljebb a szórásának másfélszerese.

Számoljuk ki a valószínűségeket azokban az esetekben is, amikor X eloszlása normális, illetve egyenletes!

Megoldás: Először X eloszlása legyen ismeretlen. Mivel X várható értékét és szórását ismerjük, alkalmazni tudjuk a Csebisev-egyenlőtlenséget. Mivel X értékeiről nem tudjuk, lehetnek-e negatívak, nem alkalmazhatjuk a Markov-egyenlőtlenséget. A Csebisev-egyenlőtlenség alkalmazásához a becsülendő valószínűségek formája csak $P(|X - E(X)| < \ldots)$ vagy $P(|X - E(X)| \geq \ldots)$ alakú lehet.

a) A feladat azt kérdezi, mikor teljesül az $X < -5$ vagy az $X > 25$ események közül az egyik. A -5-nél kisebb és a 25-nél nagyobb számok 10-től legalább 15 egységnyi távolságra vannak, azaz

$$P(X < -5 \text{ vagy } X > 25) = P(|X - 10| > 15) = P(|X - E(X)| > 15)\,.$$

Ezt a valószínűséget viszont már Csebisev-egyenlőtlenség segítségével becsülni tudjuk. Az $\varepsilon = 15$ és $D(X) = 5$, így

$$P\left(|X - E(X)| > 15\right) \leq \frac{D^2(X)}{\varepsilon^2} = \frac{5^2}{15^2} = \frac{1}{9}\,.$$

Kaptunk tehát egy felső korlátot annak a valószínűségére, hogy X értéke legfeljebb -5 , vagy legalább 25. Ez a valószínűség nem lehet $\frac{1}{9}$-nél nagyobb. Tudatosítsuk azt, hogy ismeretlen eloszlás esetén a vizsgált esemény valószínűségét nem fogjuk tudni meghatározni. Az $\frac{1}{9}$ *nem* az esemény valószínűsége, hanem arra egy felső korlát.

b) A következő kérdés esetén annak a valószínűségét vizsgáljuk, hogy az X valószínűségi változó 5 és 15 közé esik. Tehát annak a valószínűségét keressük, hogy X az $[5; 15]$ intervallumba esik. Amennyiben sikerült megoldani az önálló feladatot, láthatjuk, hogy lényegtelen az, hogy a vizsgált intervallum nyitott lesz-e vagy zárt, a becslés eredményén ez nem változtat. Ha X ebbe az intervallumba esik, akkor a várható értékétől, 10-től maximum 5 egységre kerülhet, tehát

$$P(5 < X < 15) = P(|X - E(X)| < 5)\,.$$

Így sikerült a kérdéses kifejezést olyan formára hozni, melyet már alkalmazhatunk a Csebisev-egyenlőtlenségnél:

$$P(|X - E(X)| < 5) \geq 1 - \frac{D^2(X)}{\varepsilon^2} = 1 - \frac{5^2}{5^2} = 0\,.$$

Tehát azt kaptuk, hogy annak a valószínűsége, hogy X értéke 5 és 15 között van, legalább 0. Ezt eddig is tudtuk, tehát a Csebisev-egyenlőtlenség alkalmazásával *nem* kaptunk új információt.

c) Adjunk becslést most annak a valószínűségére, hogy X eltérése a várható értéktől legalább 20. A feladat szövege alapján rögtön olyan formában tudjuk felírni a kérdéses valószínűséget, melyre a Csebisev-egyenlőtlenség közvetlenül alkalmazható:

$$P(|X - E(X)| \geq 20) \leq \frac{5^2}{20^2} = \frac{1}{16}.$$

Tehát annak a valószínűsége, hogy a valószínűségi változó eltérése a várható értéktől legalább 20, *legfeljebb* $6{,}25\%$ lehet. Vegyük észre, hogy itt nem használtuk fel azt az információt, hogy X várható értéke 10.

d) Keressünk most becslést annak a valószínűségére, hogy X eltérése a várható értékétől 20-nál kevesebb. Ismét olyan formában tudjuk felírni a kérdéses valószínűséget, melyből a becslés közvetlen elvégezhető.

$$P(|X - E(X)| < 20) \geq 1 - \frac{5^2}{20^2} = \frac{15}{16}.$$

Itt most felső korlát helyett alsó korlátot találtunk. Annak a valószínűsége, hogy a valószínűségi változónk eltérése a várható értéktől kisebb, mint 20, *legalább* $93{,}75\%$. A valószínűségi változónk várható értékét ebben az esetben sem használtunk fel. Ez az eset az előző eset komplementere: míg az előző eseményt felülről becsültük $1/16$-dal, addig az ellentettjét alulról becsüljük $15/16$-dal.

e) Vizsgáljuk most annak a valószínűségét, hogy a valószínűségi változó eltérése a várható értékétől legalább a szórás háromszorosa. Most ismét pontosan olyan típusú esemény valószínűségét keressük, melyet Csebisev-egyenlőtlenséggel becsülni tudunk:

$$P(|X - E(X)| \geq 3 \cdot D(X)) \leq \frac{1}{3^2} = \frac{1}{9}.$$

Itt a tételben szereplő λ értéke 3 lesz, s a valószínűségre azt mondhatjuk, hogy *legfeljebb* $1/9$. Ebben az esetben az X valószínűségi változónak sem a várható értékét, sem a szórását nem használtuk fel. Jegyezzük meg viszont, hogy az egyenlőtlenség *csak* akkor működik, ha mind a várható érték, mind a szórás létezik.

f) Végezetül becsüljük meg annak a valószínűségét, hogy X eltérése a várható értékétől legfeljebb szórásának másfélszerese. Ismét tudjuk használni a Csebisev-egyenlőtlenséget, $\lambda = 1{,}5$, és sem a várható értéket, sem a szórást nem használjuk:

$$P(|X - E(X)| \leq 1{,}5 \cdot D(X)) \geq 1 - \frac{1}{1{,}5^2} = \frac{5}{9}.$$

Tehát *legalább* $55{,}56\%$ annak a valószínűsége, hogy X eltérése a várható értékétől legfeljebb a szórásának másfélszerese.

Nézzük meg a feladat megoldását akkor, ha tudjuk, hogy X eloszlása normális. Mivel normális eloszlás esetén a valószínűségek számítását már bővebben tárgyaltuk, itt csak vázlatosan írjuk fel az eredményeket:

a) $P(X < -5 \text{ vagy } X > 25) = P(X < -5) + P(X > 25) = \Phi\left(\dfrac{-5-10}{5}\right) + 1 - \Phi\left(\dfrac{25-10}{5}\right) =$

$$= \Phi(-3) + 1 - \Phi(3) = 2 - 2\cdot\Phi(3) \approx 2 - 2\cdot 0{,}99865 = 0{,}0027\,.$$

b) $P(5 < X < 15) = \Phi\left(\dfrac{15-10}{5}\right) - \Phi\left(\dfrac{5-10}{5}\right) =$

$$= \Phi(1) - \Phi(-1) = 2\cdot\Phi(1) - 1 \approx 2\cdot 0{,}8413 - 1 = 0{,}6827\,.$$

A $P(|X - E(X)| < \ldots)$ vagy $P(|X - E(X)| \geq \ldots)$ alakú valószínűségeket vissza kell alakítani a fenti két formára, így tudjuk normális eloszlás esetén az értékeiket kiszámolni:

c) $P(|X - E(X)| \geq 20) = P(X < -10 \text{ vagy } X > 30) = P(X < -10) + P(X > 30) =$

$$= \Phi\left(\frac{-10-10}{5}\right) + 1 - \Phi\left(\frac{30-10}{5}\right) = \Phi(-4) + 1 - \Phi(4) =$$

$$= 2 - 2\cdot\Phi(4) \approx 2 - 2\cdot 1 = 0\,.$$

d) $P(|X - E(X)| < 20) = P(-10 < X < 30) =$

$$= \Phi\left(\frac{30-10}{5}\right) - \Phi\left(\frac{-10-10}{5}\right) = \Phi(4) - \Phi(-4) =$$

$$= 2\cdot\Phi(4) - 1 \approx 2\cdot 1 - 1 = 1\,.$$

e) $P(|X - E(X)| \geq 3 \cdot D(X)) = P(X < -5 \text{ vagy } X > 25) = P(X < -5) + P(X > 25) =$

$$= \Phi\left(\frac{-5-10}{5}\right) + 1 - \Phi\left(\frac{25-10}{5}\right) = \Phi(-3) + 1 - \Phi(3) =$$

$$= 2 - 2 \cdot \Phi(3) \approx 2 - 2 \cdot 0{,}99865 = 0{,}0027\,.$$

f) $P(|X - E(X)| < 1{,}5 \cdot D(X)) = P(2{,}5 < X < 17{,}5) =$

$$= \Phi\left(\frac{17{,}5-10}{5}\right) - \Phi\left(\frac{2{,}5-10}{5}\right) = \Phi(1{,}5) - \Phi(-1{,}5) =$$

$$= 2 \cdot \Phi(1{,}5) - 1 \approx 2 \cdot 0{,}9332 - 1 = 0{,}8664\,.$$

Az utolsó két részfeladatnál az eredmény független a várható érték tényleges értékétől!

17-3. önálló feladat: Oldjuk meg az utolsó két részfeladatot X várható értékének és szórásának felhasználása nélkül!

A feladat utolsó részében a valószínűségértékeket egyenletes eloszlás esetén határozzuk meg. Emlékezzünk arra, hogy egyenletes eloszlás esetén egy adott $[a;b]$ intervallumból választunk egy számot véletlenszerűen. Mivel tudjuk a valószínűségi változónk várható értékét és szórását, a és b értékét meghatározhatjuk:

$$10 = E(X) = \frac{a+b}{2}, \qquad 5 = D(X) = \frac{b-a}{\sqrt{12}}\,.$$

A két felírt egyenletben csak a és b ismeretlen, így azok meghatározhatóak:

$$a = 10 - 5\sqrt{3} = 1{,}3397, \qquad b = 10 + 5\sqrt{3} = 18{,}6603\,.$$

Figyeljük meg a következőt: a és b távolsága a várható értéktől (10) egyaránt a szórás (5) konstans-, pontosabban $\sqrt{3} \approx 1{,}7321$-szorosa. Az eredmények segítségével az alábbi részfeladatokat azonnal meg tudjuk oldani:

a) $P(X < -5 \text{ vagy } X > 25) = 0\,.$

c) $P(|X - E(X)| \geq 20) = P(X \leq -10 \text{ vagy } X \geq 30) = 0\,.$

d) $P(|X - E(X)| \leq 20) = P(-10 < X < 30) = 1\,.$

e) $P(|X - E(X)| \geq 3 \cdot D(X)) = P(X \leq -5 \text{ vagy } X \geq 25) = 0\,.$

A maradék két esetben a vizsgált intervallum az $[a;b]$ intervallumon belülre esik, így geometriai valószínűségi mező segítségével meg tudjuk oldani a feladatot. A halmaz mérete, ahonnan X értékét választjuk, $b - a = 5\sqrt{12} = 10\sqrt{3}$ lesz, ezért

b) $P(5 < X < 15) = \dfrac{15-5}{b-a} = \dfrac{1}{\sqrt{3}} \approx 0{,}5774\,.$

f) $P(|X - E(X)| < 1{,}5 \cdot D(X)) = P(2{,}5 < X < 17{,}5) = \dfrac{17{,}5-2{,}5}{b-a} = \dfrac{\sqrt{3}}{2} \approx 0{,}8660$

Foglaljuk össze táblázatban a kapott eredményeket:

	esemény \ eloszlás	normális	egyenletes	ismeretlen
a)	$X < -5$ vagy $X > 25$	0,0027	0	$\leq 1/9$
b)	$5 < X < 15$	0,6827	0,5774	≥ 0
c)	$\lvert X - E(X)\rvert \geq 20$	0	0	$\leq 1/16$
d)	$\lvert X - E(X)\rvert < 20$	1	1	$\geq 15/16$
e)	$\lvert X - E(X)\rvert \geq 3 \cdot D(X)$	0,0027	0	$\leq 1/9$
f)	$\lvert X - E(X)\rvert < 1{,}5 \cdot D(X)$	0,8664	0,8660	$\geq 5/9$

Figyeljük meg a kapott eredményeket. Láthatjuk, hogy a Csebisev-egyenlőtlenség által szabott korlátot mindegyik valószínűség betartja. Azt is figyeljük meg, hogy ez a korlát általában csak egy nagyon durva becslésre ad lehetőséget.

⇐17.4. feladat

Aktivitás: Dobókocka segítségével ellenőrizzük le a Csebisev-egyenlőtlenséget a $\lambda = 1{,}5$ esetben. Legyen X négy egymás után dobott szám átlaga. Készítsünk mérési jegyzőkönyvet, és vizsgáljuk meg azt, hogy a tanultak teljesülnek-e.

A Csebisev-egyenlőtlenséget nem csak a valószínűségi változó várható értékére szimmetrikus intervallumba történő esés vizsgálatánál tudjuk alkalmazni. A kérdéses intervallumot alkalmasan szimmetrikussá bővítve vagy szűkítve becsülhető annak a valószínűsége, hogy a valószínűségi változó beleesik.

17.5. feladat. Az X valószínűségi változó várható értéke 10, szórása pedig 5. Adjunk becslést annak a valószínűségére, hogy

a) X értéke 0 és 20 közé esik,

b) X értéke 0 és 100 közé esik,

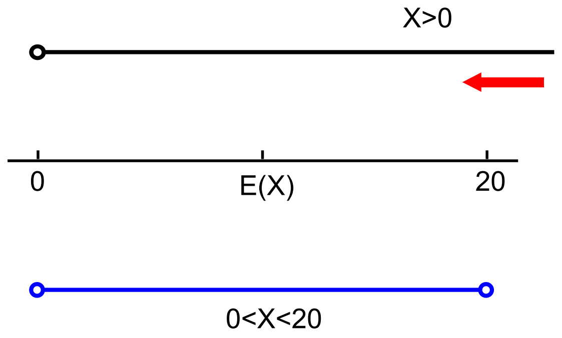
c) X értéke pozitív!

Megoldás:

a) Az első részfeladatban egy olyan intervallumot vizsgálunk, mely szimmetrikus a várható értékre. Ebben az esetben a Csebisev-egyenlőtlenség további becslés nélkül alkalmazható:

$$P(0 < X < 20) = P(|X - E(X)| < 10) \geq 1 - \frac{5^2}{10^2} = 0{,}75\,.$$

Tehát annak a valószínűsége, hogy X az adott intervallumba kerül, *legalább* 75%.



17.2. ábra. *A várható értékre aszimmetrikus* $(0;\infty)$ *intervallumot behúzzuk a* $(0;20)$*-ra, ami a várható értékre már szimmetrikus. Mindig a várható értéktől távolabbi véget* (∞) *hozzuk közelebb.*

b) Adjunk becslést most annak a valószínűségére, hogy X értéke 0 és 100 közé esik. Láthatjuk, hogy a $[0;100]$ intervallum nem szimmetrikus a várható értékre, hiszen az egyik végpont 10, a másik 90 egységre van a várható értéktől, a 10-től. A Csebisev-egyenlőtlenséggel csak akkor tudunk dolgozni, ha ez az intervallum szimmetrikus. Figyeljük meg azt, hogy mivel a várható érték az intervallumba beleesik, ezért a kérdéses valószínűségre alsó korlátot fogunk találni. Az intervallumot ezért úgy érdemes a várható értékre szimmetrikussá tenni, hogy az új intervallum az eredeti intervallum része legyen. El kell azt kerülnünk azonban, hogy a csökkentett intervallum túl kicsi legyen. A cél érdekében a vizsgált tartományt úgy alakítjuk át, hogy a várható értéktől távolabbi végpontját olyan közel hozzuk a várható értékhez, mint a másik végpont. Így a két végpont egyenlő távolságra lesz a várható értéktől. Tehát a 100-as végpontot, ami 90 egységnyi távolságra van, 10 egység közel kell hoznunk. Ezzel viszont az intervallum méretét csökkentettük, annak a valószínűsége így, hogy X beleesik, csökkenhet:

$$P(0 < X < 100) \geq P(0 < X < 20) \geq 0{,}75\,.$$

Annak a valószínűségét, hogy X a korrigált $[0;20]$ intervallumba esik, már megbecsültük, és egy alsó korlátot kaptunk rá, ami 75% volt. Ebben az esetben sem kapunk jobb eredményt, az alsó korlát itt is 75% lesz.

c) Annak a valószínűségét, hogy X pozitív, az előzőekhez hasonlóan tudjuk becsülni. Itt viszont nem egy véges, hanem egyik irányban végtelen $[0;\infty)$ intervallumba való kerülés valószínűségét vizsgáljuk. Hasonló módszerrel élve most a ∞-ben lévő pont lesz a távolabbi, ezt hozzuk 10 egység távolra a várható értékhez. Az intervallumunk saját részintervallumára csökkent, tehát

$$P(X > 0) = P(0 < X < \infty) \geq P(0 < X < 20) \geq 0{,}75\,.$$

Láthatjuk, hogy annak a valószínűségét, hogy X pozitív, hasonlóan 75%-kal tudjuk becsülni alulról. Ugyan a három vizsgált intervallum hosszúsága lényegesen eltér, mégis csak ugyanazt a becslést tudjuk arra adni a Csebisev-egyenlőtlenség segítségével, hogy a valószínűségi változó beleesik.

⇐17.5. feladat

17-4. önálló feladat: Az X valószínűségi változó várható értéke 8, szórása pedig 6. Adjunk becslést annak a valószínűségére, hogy X értéke kisebb, mint -4!

17.6. feladat. A helyi élelmiszer-ellátó üzletben sajtos baromfipárizsit vásárolunk. Az eladótól 15 dekagrammnyi felvágottat kérünk. Tapasztalatból tudjuk, hogy a mérlegre került darabok össztömegének várható értéke 18 dekagramm, de előfordul, hogy 20 dekánál is több kerül oda.

a) Adjunk becslést annak a valószínűségére, hogy a kimért sajtos baromfiparizsi tömege eléri a 20 dekagrammot!

b) Számoljunk utána az eredménynek, ha tudjuk, hogy a boltos *mindig* 18 szeletet vág, s az egyes szeletek tömege független, de *ismeretlen* eloszlást követ 2 gramm szórással.

Megoldás:

a) Az első részfeladatot az előző leckében tanult Markov-egyenlőtlenség segítségével oldhatjuk meg, mivel a kapott áru X tömegének szórását nem ismerjük, viszont tudjuk, hogy mindenképp pozitív lesz:

$$P(X > 20) \leq \frac{18}{20} = 0{,}9\,.$$

Tehát azt kapjuk, hogy annak a valószínűsége, hogy 20 dekagrammnál több párizsit kapunk, *legfeljebb* 90% lehet.

b) A második részfeladatban többletinformációkat kaptunk. Egyrészt 18 szelet párizsit kapunk, másrészt mindegyik szelet párizsinak azonos a tömegeloszlása, és a tömeg szórása is megegyezik: 2 gramm. Ebből a szeletek össztömegének szórását is kiszámolhatjuk, hiszen független valószínűségi változók összegéről van szó, $n = 18$, és $\sigma = 0{,}2$, így

$$D(X) = \sqrt{18} \cdot 0{,}2 \approx 0{,}8485\,.$$

Így már X egy olyan valószínűségi változó lesz, melynek mind a várható értékét, mind a szórását ismerjük. Ebben az esetben a $P(X > 20)$ valószínűséget becsülhetjük Csebisev-egyenlőtlenség segítségével. A vizsgált esemény azt nézi, hogy mikor esik X a $(20; \infty)$ intervallumba. Ez nem lesz szimmetrikus a várható értékre, ami 18. A várható érték az intervallumon kívül esik, így a valószínűséget felülről fogjuk becsülni, ezért a kérdéses intervallumot növelve, a növeltet kell majd felülről becsülnünk. Ahhoz, hogy a $(20; \infty)$ szimmetrikus legyen a 18-ra, hozzá kell venni a $(-\infty; 16)$ intervallumot. Így a vizsgált intervallumba való esés valószínűsége növekedhetett:

$$P(X > 20) \leq P(X > 20 \text{ vagy } X < 16) = P(|X - E(X)| > 2)\,.$$

A kapott formára már közvetlen alkalmazni tudjuk a Csebisev-egyenlőtlenséget:

$$P(|X - E(X)| > 2) \leq \frac{D^2(X)}{2^2} = \frac{18 \cdot 0{,}2^2}{2^2} = 0{,}18$$

Az egyenlőtlenség elejét és végét összeolvasva azt kapjuk, hogy

$$P(X > 20) \leq 0{,}18\,.$$

Tehát annak a valószínűsége, hogy az eladó 20 dekagrammnál több felvágottat ad, *nem lehet több*, mint 18%. Gondoljunk bele, hogy a modellünk kis pontosítással lényegesen erősebb felső korlátot adott (85% helyett 18%) a kérdéses valószínűségre.

⇐17.6. feladat

17-5. önálló feladat: Vizsgáljuk meg, miként változik az előző feladat *b)* részének megoldása, ha a szeletek számát változtatjuk úgy, hogy a többi paraméter fixen marad.

17.7. feladat. Az X valószínűségi változó várható értéke és szórása 2. Mekkora lehet annak a valószínűsége, hogy X^2 legalább 16?

a) Adjunk becslést Markov-egyenlőtlenség segítségével!

b) Készítsünk becslést a Csebisev-egyenlőtlenséget használva!

c) Vizsgáljuk meg, mi a helyzet, ha az eloszlás exponenciális!

Megoldás:

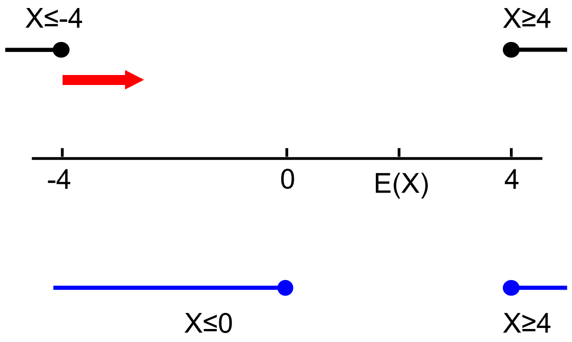
a) Először végezzük el a becslést a Markov-egyenlőtlenséggel. A Markov-egyenlőtlenséghez olyan valószínűségi változóra van szükségünk, ami nem vehet fel negatív értéket. Az X^2 ilyen. Ki kell még számolnunk X^2 várható értékét, amit X várható értékének és szórásának segítségével megtehetünk:

$$E(X^2) = E^2(X) + D^2(X) = 2^2 + 2^2 = 8\,.$$

Mivel $X^2 \geq 0$, és $E(X^2) = 8$, alkalmazhatjuk a Markov-egyenlőtlenséget az $X^2 > 16$ esemény valószínűségének becslésére:

$$P(X^2 > 16) \leq \frac{E(X^2)}{16} = \frac{8}{16} = \frac{1}{2}\,.$$

Tehát annak a valószínűsége, hogy X négyzete 16 felett lesz, *legfeljebb* 50% lehet.



17.3. ábra. *A várható értékre aszimmetrikus két intervallum közül a távolabbit ($X < -4$) behúzzuk úgy, hogy a várható értékre már szimmetrikusak legyenek. Mindig a várható értéktől távolabbi véget (-4) hozzuk közelebb.*

b) Most próbáljunk a Csebisev-egyenlőtlenséggel becsülni. Vegyük észre, hogy X^2 akkor lesz 16-nál nagyobb, ha $X > 4$ vagy $X < -4$. Ez a két intervallum nem lesz szimmetrikus a 2-re, a várható értékre, de ha a távolabbi pontot, a -4-et közelebb húzzuk, akkor az lesz. Így a vizsgált intervallumunk növekedni fog:

$$P(X^2 > 16) = P(X > 4 \text{ vagy } X < -4) \leq P(X > 4 \text{ vagy } X < 0) = P(|X - E(X)| > 2)\,.$$

Erre a formára viszont már alkalmazhatjuk a Csebisev-egyenlőtlenséget:

$$P(|X - E(X)| > 2) \leq \frac{2^2}{2^2} = 1\,.$$

A Csebisev-egyenlőtlenség segítségével tehát azt kaptuk, hogy annak a valószínűsége, hogy $X^2 > 16$, *legfeljebb* 1. Ezzel nem kaptunk új információt, így a Markov-egyenlőtlenség által adott $\leq 50\%$ az erősebb korlát.

c) Oldjuk meg a kérdést egy olyan esetben is, amikor X eloszlása ismert. Legyen X eloszlása exponenciális, így $1/\lambda = E(X) = D(X) = 2$, tehát $\lambda = 0{,}5$. Az $X^2 > 16$ valószínűségét az előbbieknek megfelelően X segítségével is felírhatjuk:

$$P(X^2 > 16) = P(X > 4 \text{ vagy } X < -4)\,.$$

Természetesen most nincs szükségünk szimmetrikus intervallumra. Ellenben exponenciális eloszlásról lévén szó, X értéke nem lehet negatív, így az $X < -4$ sosem teljesül, tehát

$$P(X > 4 \text{ vagy } X < -4) = P(X > 4)\,.$$

Ezt a valószínűséget pedig az exponenciális eloszlás eloszlásfüggvényének segítségével könnyen kiszámíthatjuk:

$$P(X > 4) = 1 - F(4) = 1 - \left[1 - e^{-0{,}5 \cdot 4}\right] = e^{-2} \approx 0{,}1353\,.$$

Tehát exponenciális eloszlás esetén a keresett valószínűség körülbelül 13 és fél százalék. Ez lényegesen kevesebb, mint a két egyenlőtlenség által szolgáltatott felső korlátok.

⇐17.7. feladat

Leckezáró önellenőrző kérdéssor

1. Az X valószínűségi változó eloszlása ismeretlen, várható értéke és szórása pedig 6. Mekkora lehet annak a valószínűsége, hogy X^2 értéke legalább 144? Számoljunk mindkét tanult egyenlőtlenséggel, és az erősebb korlátot adjuk meg!

   - [ ] 1
   - [ ] < 1
   - [ ] $\leq 1/2$
   - [ ] $\geq 1/2$

2. Legyen X normális eloszlású valószínűségi változó, melynek várható értéke 40, szórása 5. Mekkora lehet annak a valószínűsége, hogy X legfeljebb szórása kétszeresével tér el a várható értékétől?

   - [ ] $0{,}9544$
   - [ ] $\leq 0{,}9544$
   - [ ] $\leq 0{,}25$
   - [ ] $\geq 0{,}75$

3. Mekkora lehet annak a valószínűsége, hogy egy ismeretlen eloszlású valószínűségi változó legalább a szórása tízszeresével eltér a várható értékétől?

   - [ ] $\leq 1\%$
   - [ ] $\geq 0{,}1\%$
   - [ ] $\approx 0{,}01$
   - [ ] $0{,}01\%$

4. Egy ismeretlen eloszlású valószínűségi változó várható értéke 10, szórása 1. Becsülni szeretnénk annak a valószínűségét, hogy X^2 nagyobb mint 144. Melyik egyenlőtlenség adja a szigorúbb becslést?

   - [ ] Markov
   - [ ] Csebisev
   - [ ] Egyformán jók
   - [ ] Egyik sem ad értékelhető eredményt

5. Az X ismeretlen eloszlású valószínűségi változó várható értéke 40, szórása 5. Mekkora lehet annak a valószínűsége, hogy X értéke 30 és 55 között van?

   - [ ] legalább 90%
   - [ ] legfeljebb 75%
   - [ ] $\geq 3/4$
   - [ ] $< 1/4$

30. LECKE

A nagy számok törvényei: a nagy számok törvénye az átlagra

18. A nagy számok törvényei

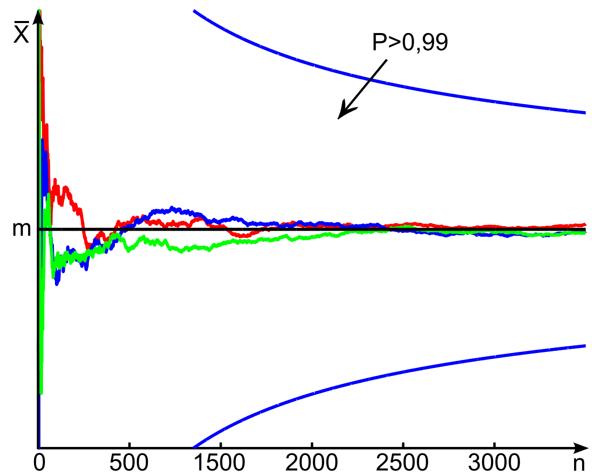
A köznyelvben gyakran előfordulnak homályos utalások valamiféle nagy számok törvényére. Sokak fejében ez nagyjából úgy él, hogy ha sokáig játszunk, akkor egyszer, valamikor majd *biztosan* nyerünk. Ennek ellenére a nagy számok törvénye nem egészen arról szól, hogy ha valamit sokszor próbálunk, akkor sikerülni fog. Inkább azt fogalmazza meg, amit az empirikus valószínűség leírásánál már megfogalmaztunk: egy esemény bekövetkezésének relatív gyakorisága az esemény valószínűségéhez tart; valamint hogy adott valószínűségi változó több felvett értékét vizsgálva azok átlaga a valószínűségi változó várható értékéhez tart. Az itt tárgyalt törvények ennél többet is mondanak: egyfajta valószínűségi felső korlátot adnak arra, mekkora eséllyel térhet el a tapasztalt értékünk az elvárttól. Ez nem a klasszikus értelemben vett konvergencia, ahol egy adott küszöbszám felett a határértéktől való eltérés *biztos*, hogy egy megadott felső korlát alatt lesz.

18.1. ábra. *Vajon hány héten át kell játszani, hogy legalább 50% valószínűséggel ötösünk legyen az adott időszakban?*

Ebben a fejezetben a nagy számok törvényeiről lesz szó matematikai, valószínűség elméleti értelemben. Általánosan nagy számok törvényének nevezünk minden olyan tételt, amely bizonyos valószínűségi változók számtani közepének (átlagának) stabilitását mondja ki, amennyiben a valószínűségi változók száma minden határon túl nőhet. A valószínűség fogalmának szemléletes megalapozáskor láttuk, hogy ha nagyon sokszor, azonos körülmények között megismétlünk egy kísérletet, akkor a sikeres kísérletek relatív gyakorisága bizonyos stabilitást mutat, egy adott érték körül ingadozik, melyet az esemény valószínűségének tekintünk. Hasonlóan igaz az is, hogy nagy számú kísérlet esetén az egyes kísérletek eredményeinek átlaga az elméleti várható érték körül ingadozik. Érdemes átgondolni azt, hogy az egész eddig felállított matematikai modell ezen törvények által lesznek empirikusan ellenőrizhetőek.

18.1. A nagy számok törvénye az átlagra

A következőkben a Csebisev-egyenlőtlenség alapján az átlagra, majd a relatív gyakoriságra bizonyítjuk be a nagy számok törvényét. A nagy számok törvényének Csebisev-féle alakja független, azonos várható értékű és szórású valószínűségi változók átlagának a közös várható értéktől való adott eltérésének valószínűségére ad becslést. Lényegileg arról van szó, hogy az átlag túlságosan nem térhet el a várható értéktől, olyan értelemben, hogy a nagy eltérés valószínűsége kicsi.



18.2. ábra. *Azonos eloszlású valószínűségi változók átlaga annál jobban közelít a közös várható értékhez, minél több változót vizsgálunk. Megadható az a sáv, melyen belül az átlagok értékei adott valószínűségnél gyakrabban fordulnak elő.*

18.1. tétel:

Legyenek $X_1,X_2,\ldots,X_n$ független valószínűségi változók, azonos várható értékkel és szórással, azaz $E(X_i)=m$ és $D(X_i)=\sigma$ $(i=1,2,\ldots,n)$. Ekkor

$$P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-m\right|\geq\varepsilon\right)\leq\frac{\sigma^2}{\varepsilon^2\cdot n}.$$

Bizonyítás:

A bizonyításhoz a Csebisev-egyenlőtlenséget kell alkalmazni az átlagra. Vezessünk be egy új valószínűségi változót az átlagra:

$$Y = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Ekkor a függetlenek átlagának várható értékére és szórására vonatkozó tétel alapján:

$$E(Y) = E\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = m\,,$$
$$D(Y) = D\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{\sigma}{\sqrt{n}}\,.$$

Ezzel a Csebisev-egyenlőtlenség a következő alakot ölti:

$$P\left(|Y - m| \geq \lambda \cdot \frac{\sigma}{\sqrt{n}}\right) \leq \frac{1}{\lambda^2}\,.$$

Legyen $\lambda = \dfrac{\varepsilon \cdot \sqrt{n}}{\sigma}$. Ezzel

$$\lambda \cdot \frac{\sigma}{\sqrt{n}} = \frac{\varepsilon \cdot \sqrt{n}}{\sigma} \cdot \frac{\sigma}{\sqrt{n}} = \varepsilon \qquad \text{és} \qquad \frac{1}{\lambda^2} = \frac{1}{\left(\frac{\varepsilon \cdot \sqrt{n}}{\sigma}\right)^2} = \frac{\sigma^2}{\varepsilon^2 \cdot n},$$

amiből a tétel állítása már közvetlenül adódik. □

1. A Csebisev-féle nagy számok törvényét gyakran annak becslésére használjuk, hogy a valószínűségi változók milyen valószínűséggel közelítik meg a várható érték bizonyos környezetét. Ehhez a tételt az alábbi formában alkalmazzuk:

$$P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - m\right| < \varepsilon\right) \geq 1 - \frac{\sigma^2}{\varepsilon^2 \cdot n}\,.$$

2. Vizsgáljuk meg mi történik akkor, ha független, azonos eloszlású valószínűségi változók egyre nagyobb halmazára alkalmazzuk a tételt. Az átlag határértékben a közös várható értékhez „tart”, hiszen adott σ és ε esetén

$$\lim_{n\to\infty} \frac{\sigma^2}{\varepsilon^2 \cdot n} = 0 .$$

Ez azt jelenti, hogy egyre több valószínűségi változót átlagolva az átlag értéke egyrészt egyre biztosabban kerül a várható érték egy bizonyos környezetébe, másrészt adott bizonyossággal a várható érték egyre kisebb környezetébe kerül.

3. Az egyenlőtlenséget összegre is megfogalmazhatjuk. Ekkor a Csebisev egyenlőtlenségben az átlag helyett az összeg várható értéke és szórása szerepel:

$$P\left(|X_1 + X_2 + \ldots + X_n - n \cdot m| \geq \varepsilon\right) \leq \frac{\sigma^2 \cdot n}{\varepsilon^2} .$$

Jegyezzük meg azonban, hogy amíg az átlag eloszlása egyre jobban a várható érték körül csúcsosodik ki, addig az összeg eloszlása inkább egyre jobban szétterül.

18.1. feladat. Egy Y valószínűségi változó szórása 10, várható értékét nem ismerjük. 16 független értékének átlagát vizsgáljuk.

a) Mekkora lehet a valószínűsége annak, hogy a kapott átlag legalább 5-tel eltér a várható értéktől?

b) Hogyan változik az eredmény, ha vizsgált értékek számát megduplázzuk 32-re?

c) Mit tudunk akkor mondani, ha a minimális eltérést duplázzuk meg 10-re?

d) Mi a helyzet abban az esetben, ha Y – és ezáltal az értékek átlaga is – normális eloszlást követ?

Megoldás:

a) Nem ismerjük az Y valószínűségi változó várható értékét, és azt sem, hogy milyen eredményt kapunk a 16 értéket átlagolva. Bevezethetünk a 16 kapott értékre 16 darab független, Y-nal azonos eloszlású valószínűségi változót: $X_1, X_2, \ldots, X_{16}$. Ezen valószínűségi változók átlagára felírhatjuk a Csebisev-féle nagy számok törvényét:

$$P\left(\left|\frac{X_1 + X_2 + \ldots + X_{16}}{16} - m\right| \geq 5\right) \leq \frac{\sigma^2}{\varepsilon^2 \cdot n} = \frac{10^2}{5^2 \cdot 16} = 0{,}25\,.$$

A képletben szereplő m jelöli Y várható értékét, s ezáltal a többi valószínűségi változó, valamint az átlag várható értékét is. Az egyenlőtlenség azt mondja, hogy annak a valószínűsége, hogy az átlag a várható értéktől legalább 5-tel eltér, *legfeljebb* 25% lehet. Természetesen a konkrét valószínűség függ Y tényleges eloszlásától, ami a kapott eredménynél akár lényegesen kisebb is lehet, de több nem.

b) A vizsgált értékek számát duplázva a vizsgált valószínűségi változók száma is duplázódik. Így $X_1, X_2, \ldots, X_{32}$-re írjuk fel a Csebisev-féle nagy számok törvényét:

$$P\left(\left|\frac{X_1 + X_2 + \ldots + X_{32}}{32} - m\right| \geq 5\right) \leq \frac{\sigma^2}{\varepsilon^2 \cdot n} = \frac{10^2}{5^2 \cdot 32} = 0{,}125\,.$$

Láthatjuk, hogy míg a vizsgált értékek száma megduplázódik, a szélsőséges átlagértékek gyakoriságának felső korlátja feleződik. Ez nem azt jelenti, hogy a tényleges valószínűség is feleződik, kizárólag a felső korlátot tudtuk szűkebbre venni.

c) Ha a vizsgált eltérés nagyságát duplázzuk, a vizsgált valószínűségi változók száma nem változik. Így újra $X_1, X_2, \ldots, X_{16}$-ra írjuk fel a Csebisev-féle nagy számok törvényét:

$$P\left(\left|\frac{X_1 + X_2 + \ldots + X_{16}}{16} - m\right| \geq 10\right) \leq \frac{\sigma^2}{\varepsilon^2 \cdot n} = \frac{10^2}{10^2 \cdot 16} = 0{,}0625\,.$$

Tehát amennyiben a várható értéktől való távolság határát megduplázzuk, negyedére esik vissza annak a felső korlátja, hogy az értékek átlaga a határon kívül esik. Vigyázzunk, mert ismét csak a felső korlátról beszélhetünk csak, mivel a konkrét valószínűségeket nem ismerjük.

d) Legyen Y most normális eloszlású valószínűségi változó. Mivel meg van adva az eloszlás típusa, pontos értékeket kaphatunk azokra a valószínűségekre, amiket az előbb csak felső korláttal tudtunk becsülni. Az egyszerűség kedvéért jelöljük az átlagot $\overline{X}$-sal:

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_{16}}{16}.$$

Mivel $\overline{X}$ várható értéke is m lesz, így $\overline{X} - m$ olyan normális eloszlású valószínűségi változó lesz, melynek várható értéke 0, szórása pedig az átlag szórása alapján számolható:

$$D(\overline{X}) = \frac{10}{\sqrt{16}} = 2{,}5\,.$$

Ezen információk segítségével az *a)* és a *c)* feladat már kiszámítható, hiszen

a) $$P\left(\left|\frac{X_1 + X_2 + \ldots + X_{16}}{16} - m\right| \geq 5\right) = P\left(\overline{X} - m \leq -5 \text{ vagy } \overline{X} - m \geq 5\right) =$$
$$= \Phi\left(-\frac{5}{2{,}5}\right) + 1 - \Phi\left(\frac{5}{2{,}5}\right) = 2 - 2 \cdot \Phi(2) \approx 4{,}55 \cdot 10^{-2}\,.$$

c) $$P\left(\left|\frac{X_1 + X_2 + \ldots + X_{16}}{16} - m\right| \geq 10\right) = P\left(\overline{X} - m \leq -10 \text{ vagy } \overline{X} - m \geq 10\right) =$$
$$= \Phi\left(-\frac{10}{2{,}5}\right) + 1 - \Phi\left(\frac{10}{2{,}5}\right) = 2 - 2 \cdot \Phi(4) \approx 6{,}33 \cdot 10^{-5}\,.$$

b) A részfeladat megoldásához előbb vegyük észre, hogy $\overline{X}$-t meg kell változtatnunk, hiszen már 32 darab valószínűségi változót átlagolunk. Az új valószínűségi változó átlaga nem változik, viszont szórása igen:

$$\overline{X}' = \frac{X_1 + X_2 + \ldots + X_{32}}{32}, \qquad \text{és} \qquad D(\overline{X}') = \frac{10}{\sqrt{32}} = 1{,}25 \cdot \sqrt{2} \approx 1{,}7678\,.$$

Ebből:

$$P\left(\left|\frac{X_1+X_2+\ldots+X_{32}}{32}-m\right|\geq 5\right)=P\left(\overline{X}'-m\leq -5 \text{ vagy } \overline{X}'-m\geq 5\right)\approx$$
$$\approx \Phi\left(-\frac{5}{1{,}7678}\right)+1-\Phi\left(\frac{5}{1{,}7678}\right)\approx 2-2\cdot\Phi(2{,}8284)\approx 4{,}68\cdot 10^{-3}\,.$$

Leellenőrizhetjük, hogy normális eloszlás esetén a Csebisev-féle nagy számok törvényéből kapott becslések rendben vannak, hiszen

a) $4{,}55\cdot 10^{-2}\leq 0{,}25,$ *b)* $4{,}68\cdot 10^{-3}\leq 0{,}125,$ *c)* $6{,}33\cdot 10^{-5}\leq 0{,}0625\,.$

Vegyük azonban észre, hogy míg a becslések részfeladatról részfeladatra feleződtek, addig a tényleges valószínűségi értékek több nagyságrendet estek. A feladat megmutatta tehát, hogy a kapott felső korlát még nagyságrendileg sem sejtetheti, hogy mekkora lesz a tényleges valószínűség, amit csak becsülünk az egyik oldalról.

⇐18.1. feladat

Aktivitás: Ebben a leckében a nagy számok úgynevezett gyenge törvényét tanuljuk. Keressen rá az Internet segítségével a nagy számok erős törvényére. Fogalmazza meg saját szavainkkal, mint mond ki az erős törvény az előző feladatra alkalmazva.

18.2. feladat. Egy kísérletsorozat során a víz forráspontját mérjük. A berendezés beállításai miatt a mért hőmérséklet egy $373{,}15$ K várható értékű, 1 K szórású valószínűségi változó lesz. A mérést négyszer ismételjük meg, s a mért értékeket átlagoljuk. Tételezzük fel, hogy sikerül az egyes méréseket egymástól függetlenül elvégezni.

a) Adjunk meg egy olyan, a várható értékre szimmetrikus intervallumot, melybe a négy mérés átlaga legalább 95%-os valószínűséggel beleesik.

b) Miként változik ez az intervallum, ha a mérések számát megduplázzuk?

c) Mi a válasz, ha feltételezhetjük, hogy az egyes mért értékek – és így az átlaguk is – normális eloszlást követnek?

Megoldás:

Az első két részfeladatnál a valószínűségi változók eloszlását nem ismerjük. A nagy számok törvénye alapján viszont arra gondolhatunk, hogy minél több kísérletet végzünk, annál pontosabban ki tudjuk mérni a keresett 373,15 Kelvines értéket.

a) Az első részfeladatban a várható értéknek egy olyan ε környezetét keressük, amelybe az átlag legalább 95%-os valószínűséggel esik. A kísérletek száma $n = 4$, az egyes mérésekhez tartozó szórás pedig $\sigma = 1$ K. Jelölje X_1, X_2, X_3 és X_4 az egyes kísérletek kimenetelét, s írjuk fel a Csebisev-féle nagy számok törvényét az alábbi módon:

$$P\left(\left|\frac{X_1 + X_2 + X_3 + X_4}{4} - m\right| \le \varepsilon\right) \ge 1 - \frac{\sigma^2}{\varepsilon^2 \cdot n} \ge 0{,}95\,.$$

Figyeljük meg azt, hogy amennyiben ε elég nagy, akkor 1-ből elég kicsi számot fogunk kivonni ahhoz, hogy a kifejezés értéke 0,95 felett legyen, viszont ezáltal a kívánt valószínűség is 0,95 felett lesz, s épp ezt szeretnénk elérni. Tehát azt szeretnénk, hogy

$$1 - \frac{\sigma^2}{\varepsilon^2 \cdot n} \ge 0{,}95 \qquad \text{legyen, tehát} \qquad \varepsilon^2 \ge \frac{20 \cdot \sigma^2}{n} = \frac{20 \cdot 1^2}{4} = 5 \quad [\mathrm{K}^2]\,.$$

Ebből $\varepsilon \approx 2{,}2361$ Kelvin lesz, azaz a mért értékek átlaga legalább 95%-os valószínűséggel esik a $[370{,}91\,\mathrm{K}; 375{,}39\,\mathrm{K}]$ intervallumba. Ne csodálkozzunk azonban, ha a mért átlag ennél gyakrabban esik az adott intervallumba, hiszen a valószínűségre csak egy alsó korlátot állapítottunk meg.

b) Ha a mérések számát megduplázzuk, azaz 8 kísérletet végzünk, akkor is végigjárhatjuk a fenti gondolatmenetet, tehát azt kell elérnünk, hogy

$$1-\frac{\sigma^2}{\varepsilon^2\cdot n}\geq 0{,}95 \qquad \text{legyen, azazt} \qquad \varepsilon^2\geq\frac{20\cdot\sigma^2}{n}=\frac{20\cdot 1^2}{8}=2{,}5 \quad [\mathrm{K}^2]\,.$$

Ebből $\varepsilon \approx 1{,}5811$ Kelvin lesz, azaz a mért értékek átlaga legalább 95%-os valószínűséggel esik a $[371{,}57\,\mathrm{K}; 374{,}73\,\mathrm{K}]$ intervallumba. Figyeljük meg, hogy a kísérletek számának duplázása mellett a vizsgált intervallum hossza kb. $4{,}5$-ről kb. 3-ra csökkent (a $\sqrt{2}$-ed részére).

ca) Vizsgáljuk meg most azt, mi történik akkor, ha feltételezhetjük, hogy az egyes kísérletek eredményei normális eloszlást követnek. Ekkor a mért értékek átlaga is normális eloszlást követ. Az X_1, X_2, X_3 és X_4 valószínűségi változók $\overline{X}$ átlagának várható értéke megegyezik az eredeti várható értékkel, szórása viszont $\sqrt{n}$-ed részére csökken, azaz

$$E(\overline{X})=373{,}15\,\mathrm{K}, \qquad \text{és} \qquad D(\overline{X})=\frac{1}{\sqrt{4}}=0{,}5\,\mathrm{K}\,.$$

A feladat megoldásához egy olyan ε értéket kell keresnünk, melyre

$$P(|\overline{X}-m|<\varepsilon)\geq 0{,}95, \qquad \text{azaz} \qquad P(m-\varepsilon<\overline{X}<m+\varepsilon)\geq 0{,}95\,.$$

Abba gondoljunk bele, hogy mivel $\overline{X}$ folytonos eloszlású, ezért a vizsgált valószínűség ε-nal folytonosan növekedni fog. Tehát a „$\geq$” reláció helyett egyenlőséggel számolva, ha kapunk egy ε értéket, akkor ez a kapott érték lesz a legkisebb, melyre a kívánt feltétel teljesül. Tehát az alábbi egyenletből ε-t kell kifejezni:

$$P(m-\varepsilon<\overline{X}<m+\varepsilon)=0{,}95\,.$$

A kifejezés bal oldalát az alábbiak szerint alakíthatjuk át, mivel tudjuk, hogy $\overline{X}$ eloszlása normális:

$$P(m-\varepsilon<\overline{X}<m+\varepsilon)=\Phi\left(\frac{m+\varepsilon-m}{D(\overline{X})}\right)-\Phi\left(\frac{m-\varepsilon-m}{D(\overline{X})}\right)=2\cdot\Phi\left(\frac{\varepsilon}{D(\overline{X})}\right)-1$$

Mivel a kapott kifejezés 0,95-tel egyenlő, így

$$2 \cdot \Phi\left(\frac{\varepsilon}{D(\overline{X})}\right) - 1 = 0{,}95, \qquad \text{azaz} \qquad \Phi\left(\frac{\varepsilon}{D(\overline{X})}\right) = 0{,}975\,.$$

Amennyiben $\Phi(x)$ értéke $\approx 0{,}975$, akkor táblázatból visszakereshetjük, hogy x értéke $1{,}96$ lesz. Ebből következik, hogy

$$\varepsilon \approx 1{,}96 \cdot D(\overline{X}) = 0{,}98\,[\mathrm{K}]\,.$$

Tehát amennyiben ismerjük a mérés eredményének az eloszlását, s ez a normális eloszlás, akkor ki tudjuk számolni, milyen intervallumba fog a mért értékek átlaga esni. Ez a $[372{,}17\,\mathrm{K}; 374{,}13\,\mathrm{K}]$ intervallum. Láthatjuk, hogy ez egy szűkebb intervallum, mint az, amit ismeretlen eloszlás esetén számoltunk. Annak a valószínűsége, hogy normális eloszlás esetén a megelőző részfeladatban számolt $[370{,}91\,\mathrm{K}; 375{,}39\,\mathrm{K}]$ intervallumba essen a mérési eredmények átlaga így nyilván 95%-nál nagyobb lesz, azaz a mostani eredményünk konzisztens az előző eredménnyel.

cb) Hogyan változik az előző eredmény, ha 4 helyett 8 mérést végzünk? Ekkor $\overline{X}$ várható értéke nem változik, viszont szórása kisebb lesz:

$$D(\overline{X}) = \frac{1}{\sqrt{8}} \approx 0{,}3536\,[\mathrm{K}]\,.$$

Vegyük észre, hogy az előző részfeladatban leírt gondolatmenetet lépésről lépésre követhetjük, azaz

$$\varepsilon \approx 1{,}96 \cdot D(\overline{X}) = 0{,}6931\,[\mathrm{K}]\,.$$

Így most a mért értékek átlaga a $[372{,}46\,\mathrm{K}; 373{,}84\,\mathrm{K}]$ intervallumba fog esni. Vegyük észre, hogy az előző részfeladathoz képest az intervallumunk hossza most is $\sqrt{2}$-ed részére csökkent. Ugyanez a megállapítás érvényes mind a normális, mind az ismeretlen eloszlás esetén.

⇐18.2. feladat

18-1. önálló feladat:
Írjunk fel összefüggést az előző feladatban bemutatott intervallum hossza és az átlagolt valószínűségi változók száma között normális eloszlás és ismeretlen eloszlás esetén is.

18.3. feladat. Szabályos, hatoldalú dobókockával dobott szám várható értékét szeretnénk kísérleti módszerrel meghatározni. A cél érdekében a kockát többször elgurítjuk, s kiszámoljuk a dobott értékek átlagát. Hányszor kell elgurítanunk a kockát, ha azt szeretnénk, hogy a kapott átlag legalább 95%-os valószínűséggel

a) $0{,}05$-nél jobban,

b) $0{,}005$-nél jobban megközelítse a várható értéket?

Megoldás:

a) Becsüljük a várható értéktől való eltérést a nagy számok törvénye segítségével. Jelölje $\overline{X}$ a dobott számok átlagát. Egy kockadobás értékének eloszlása egyazon X valószínűségi változó eloszlását követi. Jelölje m az X várható értékét, és σ pedig a szórását. Egy kockadobás esetén mind a várható értéket, mind a szórást előző feladatok során már meghatároztuk:

$$m = E(X) = \frac{6+1}{2} = 3{,}5, \qquad \text{és} \qquad \sigma = D(X) = \sqrt{\frac{6^2-1}{12}} = 1{,}7078\,.$$

A vizsgált $\overline{X}$ valószínűségi változó több, X-szel azonos eloszlású valószínűségi változó átlaga. Azt szeretnénk, hogy $\overline{X}$ eltérése a várható értékétől – ami megegyezik X várható értékével – legfeljebb $0{,}05$ legyen 95% valószínűséggel. Írjuk fel a nagy számok törvényét az alábbi alakban:

$$P\left(\left|\overline{X} - m\right| \le \varepsilon\right) \ge 1 - \frac{\sigma^2}{\varepsilon^2 \cdot n} \ge 0{,}95\,.$$

A felírt képletben mind az ε, mind a σ értéke adott. Az átlagolt valószínűségi változók számát növelve viszont a középső tagot 0,95 felé növelhetjük, ezáltal a keresett valószínűség is legalább 0,95 lesz. Tehát olyan n értéket keresünk, melyre

$$1 - \frac{\sigma^2}{\varepsilon^2 \cdot n} \geq 0{,}95 \qquad \text{azaz} \qquad n \geq \frac{20 \cdot \sigma^2}{\varepsilon^2} = \frac{20 \cdot 35}{12 \cdot 0{,}05^2} = 23333{,}3\,.$$

Ez azt jelenti, hogy ha 23334-szer vagy többször elgurítjuk a dobókockát, s a kapott értékeket átlagoljuk, akkor annak a valószínűsége, hogy a kapott átlag 5 századnál közelebb van a várható értékhez, *legalább* 95%.

b) Most azt vizsgáljuk meg, hány gurítás szükséges ahhoz, hogy legalább 95%-os valószínűséggel 0,005-nél közelebb legyünk a várható értékhez. Vegyük észre, hogy az előző részfeladat gondolatmenetét kell újra végigjátszanunk, azaz olyan n értéket keresünk, melyre

$$1 - \frac{\sigma^2}{\varepsilon^2 \cdot n} \geq 0{,}95 \qquad \text{azaz} \qquad n \geq \frac{20 \cdot \sigma^2}{\varepsilon^2} = \frac{20 \cdot 35}{12 \cdot 0{,}005^2} = 2333333{,}3\,.$$

Tehát ha *legalább* 2333334-szer dobunk, akkor annak a valószínűsége, hogy a dobott értékek átlaga a 3,5-et 5 ezrednél jobban megközelíti, *legalább* 95%. Figyeljük meg, hogy ahhoz, hogy tízszer pontosabb becslést kapjunk, százszor annyi gurítást kell végeznünk.

⇐18.3. feladat

18-2. önálló feladat: Legyen az X valószínűségi változó szórása 4. X hány független értékét kell átlagolni ahhoz, hogy az átlag eltérése X várható értékétől 1-nél kisebb legyen? Mi a válasz, ha X normális eloszlású?

Leckezáró önellenőrző kérdéssor

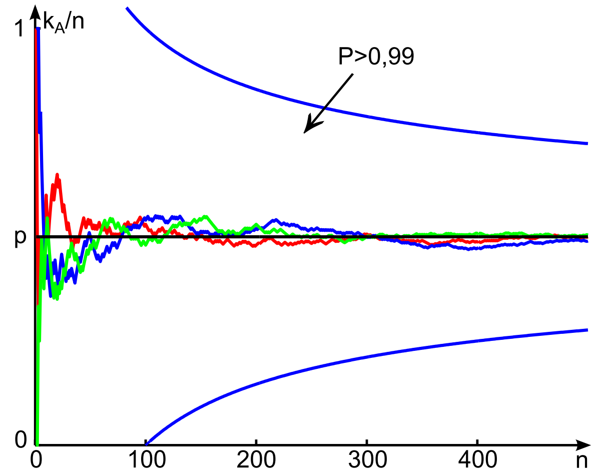
1. A kimondott Csebisev-féle nagy számok törvénye esetén az átlagolandó valószínűségi változók
   - ☐ száma nagy,
   - ☐ eltérése a várható értéktől kicsi,
   - ☐ szórása azonos,
   - ☐ átlaga állandó.
2. Kilenc független, azonos eloszlású valószínűségi változó mindegyikének várható értéke 10, szórásuk 3. A Csebisev-féle nagy számok törvénye segítségével becsüljük meg, mekkora a valószínűsége annak, hogy az átlaguk értéke a várható értéktől legfeljebb 4 egységgel tér el?
   ☐ $\approx 0{,}75$ ☐ $\geq 0{,}25$ ☐ $\geq 0{,}9375$ ☐ $\geq 0{,}75$
3. Négy független, normális eloszlású valószínűségi változó mindegyikének várható értéke és szórása 6. Mekkora lehet annak a valószínűsége, hogy a valószínűségi változók összege legfeljebb kilenccel tér el a várható értékétől?
   ☐ $\approx 0{,}5468$ ☐ $\geq 0{,}6667$ ☐ $\leq 0{,}3333$ ☐ $\geq 0{,}8167$
4. Egy méréssorozat során ismeretes, hogy a mért értékek szórása $0{,}1$ egység. Hány mérést kell elvégezni ahhoz, hogy legalább 90% legyen annak a valószínűsége, hogy a mért értékek átlaga legfeljebb $0{,}01$-dal térjen el annak várható értékétől?
   ☐ ≥ 1000 ☐ ≥ 4000 ☐ ≈ 100 ☐ ≈ 40000

31. LECKE

A nagy számok törvénye a relatív gyakoriságra

18.2. A nagy számok törvénye a relatív gyakoriságra

A nagy számok Bernoulli-féle törvénye adott számú független kísérlet esetén a relatív gyakoriságnak a valószínűségtől való adott eltérésének valószínűségére ad becslést. Kicsit pontatlanul megfogalmazva azt mondja ki a tétel, hogy a relatív gyakoriság nagyon nem térhet el az elméleti valószínűség értéktől, olyan értelemben, hogy a nagy eltérés valószínűsége kicsi.



18.3. ábra. *Adott esemény relatív gyakorisága annál jobban közelít az esemény valószínűségéhez, minél több kísérletet vizsgálunk. Megadható az a sáv, melyen belül a relatív gyakoriság értékei adott valószínűségnél gyakrabban fordulnak elő.*

18.2. tétel: Végezzünk el n darab független kísérletet a p valószínűségű A esemény megfigyelésére. Tegyük fel, hogy a kísérletek során az A esemény k-szor következett be. Ekkor tetszőleges $\varepsilon > 0$ esetén

$$P\left(\left|\frac{k}{n} - p\right| \geq \varepsilon\right) \leq \frac{p\cdot(1-p)}{\varepsilon^2 \cdot n}.$$

Bizonyítás: Rendeljünk minden egyes kísérlethez egy-egy karakterisztikus eloszlású valószínűségi változót:

$$X_i = \begin{cases} 1, & \text{ha az } A \text{ esemény bekövetkezett} \\ & \text{az } i\text{-edik kísérletben;} \\ 0 & \text{különben.} \end{cases}$$

Ekkor X_i eloszlása, várható értéke és szórása:

$$X_i : \begin{cases} 0 & 1 \\ 1-p & p \end{cases}, \qquad E(X_i) = p, \qquad D(X_i) = \sqrt{p\cdot(1-p)}.$$

Vegyük észre, hogy a sikeres kísérletek száma (k) az $X_1, X_2, \ldots, X_n$ valószínűségi változók összege, így a $\frac{k}{n}$ hányados ezen valószínűségi változók átlaga. Alkalmazhatjuk tehát a nagy számok törvényének Csebisev-féle alakját az $m = p$ és $\sigma = \sqrt{p \cdot (1-p)}$ értékekkel, amiből a tétel állítása azonnal adódik. $\square$

1. A nagy számok Bernoulli-féle törvényét gyakran annak becslésére használjuk, hogy a relatív gyakoriság milyen valószínűséggel közelíti meg az adott A esemény p valószínűségét. Ehhez a tételt az alábbi formában alkalmazzuk:
$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{p \cdot (1-p)}{\varepsilon^2 \cdot n}.$$
2. A relatív gyakoriságok határértéke az esemény p valószínűségével egyezik meg, hiszen adott p és ε esetén
$$\lim_{n\to\infty} \frac{p \cdot (1-p)}{\varepsilon^2 \cdot n} = 0.$$
Pontosabban fogalmazva, a relatív gyakoriság sztochasztikusan konvergál az esemény p valószínűségéhez.
3. A nagy számok Bernoulli-féle törvényét akkor is használhatjuk a valószínűség becslésére, ha a megfigyelt A esemény p valószínűsége nem ismert. Mivel tetszőleges $0 \leq p \leq 1$ esetén teljesül, hogy $p \cdot (1-p) \leq \frac{1}{4}$, ezért
$$P\left(\left|\frac{k}{n} - p\right| \geq \varepsilon\right) \leq \frac{p \cdot (1-p)}{\varepsilon^2 \cdot n} \leq \frac{1}{4 \cdot \varepsilon^2 \cdot n}.$$
Vagyis ismeretlen p esetén az alábbi alakokban használhatjuk a tételt:
$$P\left(\left|\frac{k}{n} - p\right| \geq \varepsilon\right) \leq \frac{1}{4 \cdot \varepsilon^2 \cdot n},$$
$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{1}{4 \cdot \varepsilon^2 \cdot n}.$$

4. Vegyük észre, hogy a Bernoulli-féle nagy számok törvényében szereplő k egy valószínűségi változó. Ráadásul ismerjük is k eloszlását, hiszen azt számolja meg, hogy n darab független kísérlet esetén a p valószínűséggel bekövetkező A esemény hányszor következik be. Tehát k binomiális eloszlású. Ez azt jelenti, hogy az előző leckében tárgyalt Csebisev-féle nagy számok törvényével ellentétben – ahol a valószínűségi változók eloszlását nem ismertük – az itt felírt törvény egy konkrét értékre próbál becslést adni.

Aktivitás: Keressen egy olyan szoftvert, mely alkalmas a nagy számok törvényében szereplő

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right)$$

típusú valószínűségeket kiszámolni abban az esetben, ha k eloszlása binomiális.

18.4. feladat. A frissen kapott ösztöndíjunkat kaszinóba megyünk kamatoztatni. Rulettezni fogunk úgy, hogy mindig csak egy számot teszünk meg. A rulettkeréken 0-tól 36-ig szerepelnek a számok, s a behelyezett golyó ezek közül választ véletlenszerűen.

a) Adjunk becslést annak a valószínűségére, hogy 100 játék után a nyertes játékok relatív gyakoriságának eltérése a nyerés valószínűségétől maximum 0,05.

b) Hogyan változik a becsült valószínűség, ha az egész éves kaszinózásunk 10000 darab tétjére szeretnénk inkább kiszámolni?

c) Számoljuk ki a valószínűségek értékét pontosan is!

Megoldás: A feladatban azt vizsgáljuk, hogy $n = 100$ kísérlet esetén hányszor következik be egy $p = 1/37$ valószínűségű esemény. A számítandó valószínűséget becsülni tudjuk a Bernoulli-féle nagy számok törvényének segítségével.

a) Az első részfeladatban annak a valószínűségét vizsgáljuk, hogy a relatív gyakoriság a valószínűségtől maximum $\varepsilon = 0{,}05$-tel tér el. Közvetlenül alkalmazhatjuk a Bernoulli-féle nagy számok törvényét:

$$P\left(\left|\frac{k}{n} - p\right| < 0{,}05\right) \geq 1 - \frac{\frac{1}{37}\cdot\frac{36}{37}}{100 \cdot 0{,}05^2} \approx 0{,}8948\,.$$

Tehát annak a valószínűsége, hogy a relatív gyakoriság a valószínűségtől $0{,}05$-nál kevesebbel tér el, *legalább* $89{,}48\%$.

ca) Ne feledjük el, hogy ezt a valószínűséget *pontosan* is kiszámolhatjuk, hiszen k binomiális eloszlású valószínűségi változó, $n = 100$ és $p = 1/37$ paraméterekkel:

$$P\left(\left|\frac{k}{n} - p\right| < 0{,}05\right) = P\left(|k - np| < 0{,}05 \cdot n\right) = P\left(\left|k - \frac{100}{37}\right| < 0{,}05 \cdot 100\right) =$$

$$= P(-5 < k - 2{,}7 < 5) = P(-2{,}3 < k < 7{,}7) = P(0 \leq k \leq 7)\,.$$

Tudjuk, hogy k binomiális eloszlású, tehát 0 és n közötti egész értékeket vehet fel. Ismerjük az eloszlás pontos alakját, így a kérdéses 8 értéket is kiszámolhatjuk. Ne keverjük az itt használt k szimbólumot azzal, amit a nevezetes eloszlásoknál használtunk!

$$P(k = i) = \binom{n}{i} \cdot p^i \cdot q^{n-i}$$

$$P(k = 0) = \binom{100}{0} \cdot \left[\frac{1}{37}\right]^0 \cdot \left[\frac{1}{37}\right]^{100} \approx 0{,}064577, \qquad P(k = 1) = \binom{100}{1} \cdot \left[\frac{1}{37}\right]^0 \cdot \left[\frac{1}{37}\right]^{99} \approx 0{,}17938$$

$$P(k = 2) = \binom{100}{2} \cdot \left[\frac{1}{37}\right]^0 \cdot \left[\frac{1}{37}\right]^{98} \approx 0{,}246648 \qquad P(k = 3) = \binom{100}{3} \cdot \left[\frac{1}{37}\right]^0 \cdot \left[\frac{1}{37}\right]^{97} \approx 0{,}22381$$

$$P(k=4) = \binom{100}{4} \cdot \left[\frac{1}{37}\right]^0 \cdot \left[\frac{1}{37}\right]^{96} \approx 0{,}150761 \qquad P(k=5) = \binom{100}{5} \cdot \left[\frac{1}{37}\right]^0 \cdot \left[\frac{1}{37}\right]^{95} \approx 0{,}080406$$

$$P(k=6) = \binom{100}{6} \cdot \left[\frac{1}{37}\right]^0 \cdot \left[\frac{1}{37}\right]^{94} \approx 0{,}035364 \qquad P(k=7) = \binom{100}{7} \cdot \left[\frac{1}{37}\right]^0 \cdot \left[\frac{1}{37}\right]^{93} \approx 0{,}013191$$

A számolt valószínűségek összege, így a keresett valószínűség $0{,}9941$ lesz. Természetesen a Bernoulli-féle nagy számok törvénye által adott $\geq 0{,}8948$ becslés jó, de elég durva.
Emlékezzünk vissza a Csebisev-féle nagy számok törvényére. Amikor adva volt a vizsgált valószínűségi változó eloszlása (pl. normális), akkor tudtunk konkrét valószínűséget számolni. Vegyük észre, hogy itt is ez történik, hiszen a Bernoulli-féle nagy számok törvényében épp a Csebisev-féle formában szereplő valószínűségi változók eloszlását adjuk meg. S így ki is tudjuk – bizonyos esetekben akár kézzel is – számolni a törvényben vizsgált valószínűséget pontosan.

b) Adjunk most becslést abban az esetben is, ha a játékok, azaz a kísérletek száma $n = 10000$. A Bernoulli-féle nagy számok törvényét az előzőekhez hasonlóan alkalmazhatjuk:

$$P\left(\left|\frac{k}{n} - p\right| < 0{,}05\right) \geq 1 - \frac{\frac{1}{37} \cdot \frac{36}{37}}{10000 \cdot 0{,}05^2} \approx 0{,}998948\,.$$

Tehát annak a valószínűsége, hogy a relatív gyakoriság a valószínűségtől $0{,}05$-nál kevesebbel tér el, *legalább* $99{,}8948\%$. Figyeljük meg, hogy a 8948 számjegyek elé nem véletlen kerül be a két darab kilences. A megjelenő kilencesek száma megegyezik az n értéke után írt nullák számával.

cb) A pontos valószínűséget ebben az esetben is kiszámolhatjuk. A k valószínűségi változó itt is binomiális eloszlású, de $n = 10000$ és $p = 1/37$ paraméterekkel:

$$P\left(\left|\frac{k}{n} - p\right| < 0{,}05\right) = P\left(|k - np| < 0{,}05 \cdot n\right) = P\left(\left|k - \frac{10000}{37}\right| < 0{,}05 \cdot 10000\right) =$$
$$= P(-500 < k - 270{,}27 < 500) = P(0 \leq k \leq 770)\,.$$

Az azért nem várható el senkitől, hogy a fenti 771 darab binomiális eloszlás értéket kiszámolja. Megfelelő szoftverrel azonban ez bizonyos pontosságig meghatározható.

$$P(0 \leq k \leq 770) \approx 1\,.$$

Azt azért tudni kell, hogy a számítógép minden egyes valószínűség értéket 10^{-15} pontossággal számolt, azaz ha majdnem 1000 számot is adott össze, legalább 12 darab kilences lesz a tizedesvessző után az eredményben (a Bernoulli-féle nagy számok törvényében kettő volt).

⇐18.4. feladat

Aktivitás: Írjon programot, vagy alkalmazzon szoftvert az előző példa utolsó részfeladatának megoldásához.

18.5. feladat. Egy hatoldalú dobókockát dobálunk. A dobott hármasok számát számoljuk.

a) Adjunk becslést arra, hogy legalább hányszor kell feldobni a kockát ahhoz, hogy a dobott hármasok számának relatív gyakorisága legalább 95% valószínűséggel 0,02-nál kevesebbel térjen el az elméleti 1/6 valószínűségtől.

b) Hogyan változik a becslés, ha 99%-ra akarjuk a kérdéses valószínűséget növelni?

c) Számítsuk ki a valószínűségek pontos értékét.

Megoldás:

a) Minél többször dobjuk fel a kockát, annál nagyobb valószínűséggel fog a relatív gyakoriság a tényleges valószínűséghez közelebb kerülni. A relatív gyakoriság, mint valószínűségi változó, eloszlása egyre jobban a hármas dobásának valószínűsége köré fog húzódni. A Bernoulli-féle nagy számok törvényével alsó becslést adhatunk annak a valószínűségére, hogy a relatív gyakoriság a valószínűségnek, mint várható értékének, adott környezetébe kerül. Ebben az esetben nem ismerjük a dobások, azaz a kísérletek számát, n-et, viszont azt szeretnénk elérni, hogy a kapott valószínűség legalább 95% legyen. Írjuk fel a nagy számok törvényének Bernoulli-féle alakját az alábbi formában:

$$P\left(\left|\frac{k}{n}-p\right|<\varepsilon\right) \geq 1-\frac{p\cdot(1-p)}{\varepsilon^2\cdot n} \geq 0{,}95\,.$$

Figyeljük meg a képletben, hogy ha n elég nagy, akkor az egyből kivont kifejezés elég kicsi lesz ahhoz, hogy a különbség $0{,}95$ felett legyen. Tehát az alábbi kifejezésből n minimális értékére vagyunk kíváncsiak:

$$1-\frac{p\cdot(1-p)}{\varepsilon^2\cdot n} \geq 0{,}95, \qquad \text{amiből} \qquad n \geq \frac{20\cdot p\cdot(1-p)}{\varepsilon^2} = \frac{20\cdot\frac{1}{6}\cdot\frac{5}{6}}{0{,}02^2} \approx 6944{,}4\,.$$

Azt az eredményt kaptuk, hogy ha *legalább* 6945-ször elgurítjuk a dobókockát, akkor a hármas dobások számának relatív gyakorisága *legalább* 95%-os valószínűséggel közelebb lesz $1/6$-hoz, mint $0{,}02$. A becslés azt mondja, hogy ez a valószínűség 6945 darab gurítás esetén már biztos legalább 95% lesz. Azonban az is lehet, hogy ennél kevesebbre van szükségünk.

ca) Számoljuk ki, hogy az előző feladatban mennyi lesz pontosan ez a valószínűség. Különböző n értékek esetén kell a

$$P\left(\left|\frac{k}{n}-p\right|<\varepsilon\right) = P\left(|k-np|<n\cdot\varepsilon\right) = P\left(\lceil np-n\cdot\varepsilon\rceil \leq k \leq \lfloor np+n\cdot\varepsilon\rfloor\right)$$

valószínűséget vizsgálni. Természetesen ezt érdemes szoftver segítségével megtenni. Emlékezzünk rá, hogy a k még mindig egy binomiális eloszlású valószínűségi változó. Próbálgatással rájöhetünk, hogy $n = 1333$ esetén a kérdéses valószínűség $94{,}68\%$, $n = 1334$ esetén pedig $95{,}28\%$. Tehát az *a)* feladatra adott pontos válasz az, hogy legalább 1334-szer kell elgurítanunk a kockát ahhoz, hogy 95%-nál nagyobb valószínűséggel közelítse meg $0{,}02$-nél jobban a hármas dobás valószínűségét. Összefoglalva az $n \geq 6945$ egy elégséges feltétel a feladatban szereplő kritériumhoz, de nem szükséges. A most számolt $n \geq 1334$ viszont szükséges is és elégséges is a kritérium teljesítéséhez.

b) Most vizsgáljuk meg azt, miképp változik az eredmény akkor, ha 95% helyett legalább 99% valószínűséggel szeretnénk, ha a relatív gyakoriság a valószínűség $0{,}02$ sugarú környezetébe esne. Az első részfeladat gondolatmenetét követhetjük, tehát n értékére leszünk kíváncsiak:

$$1 - \frac{p \cdot (1-p)}{\varepsilon^2 \cdot n} \geq 0{,}99, \qquad \text{amiből} \qquad n \geq \frac{100 \cdot p \cdot (1-p)}{\varepsilon^2} = \frac{100 \cdot \frac{1}{6} \cdot \frac{5}{6}}{0{,}02^2} \approx 34722{,}2\,.$$

Azaz a Bernoulli-féle nagy számok alapján, ha *legalább* 34723 gurítást végzünk, akkor annak a valószínűsége, hogy a relatív gyakoriság eltérése a hármas dobás valószínűségétől legfeljenn $0{,}02$, legalább 99%. Láthatjuk, hogy ez az érték az előző érték ötszöröse.

cb) Most is számoljuk ki, mennyi lesz pontosan ez a valószínűség. Ha újra alkalmazzuk a megfelelő szoftvert, akkor láthatjuk, hogy a keresett valószínűség $n = 2308$ esetén $98{,}98\%$, $n = 2309$ esetén pedig $99{,}06\%$. Ez az előbbi számított érték nagyjából $1{,}73$-szorosa.

⇐18.5. feladat

Aktivitás: Alkalmas szoftver segítségével számoljuk ki az előző feladat *c)* részeinek eredményeit.

18.6. feladat.

Közvélemény-kutatás alapján vizsgálják, legyen-e a sör adó- és járulékmentes. A megkérdezett 10040 ember közül 8012 ezt támogatja, a maradék nem.

a) Mi a valószínűsége annak, hogy az igenek aránya legfeljebb $0{,}01$ távol esik attól az aránytól, amit akkor kapnánk, ha mind a tízmillió polgárt megkérdeznénk?

b) Legalább hány embert kell megkérdezni ahhoz, hogy ez a valószínűség 99%-nál nagyobb legyen?

Megoldás: A tízmillió lakosból mindenki biztos abban, támogatja-e a kezdeményezést, vagy nem. A problémát az okozza, hogy nem tudni, melyik 10040 embert sikerült a közvélemény-kutató cégnek megkeresnie. Így az igenek számában ingadozások lehetnek, különbözhet a teljes lakosság igeneinek arányától. Figyelembe kell vennünk azt is, hogy az eddigiekkel ellentétben ez egy hipergeometriai eloszlás, hiszen nem kérdezzük meg kétszer ugyanannak az embernek a véleményét. Azonban mivel a megkérdezettek aránya a lakosságon belül nagyon kicsi, így elhanyagolható hibával számolunk, ha binomiális eloszlást használunk.

Vegyük észre azonban, hogy most nem ismerjük annak a p valószínűségét, hogy a teljes lakosságból véletlenszerűen választott polgár támogatja-e a kezdeményezést. A megkeresések száma adott, $n = 10040$.

a) Az első kérdéshez azt kell megvizsgálnunk, hogy mi a valószínűsége annak, hogy az igenek számának kapott relatív gyakorisága az igenek teljes arányától kevesebb mint $0{,}01$-dal tér el. A Bernoulli-féle nagy számok törvénye segítségével ennek a valószínűségnek adhatunk egy alsó korlátot. Mivel az egyenlőtlenségben szereplő p értéket nem ismerjük, ezért a becsléshez a p-től független formát használjuk:

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{1}{4 \cdot \varepsilon^2 \cdot n} = 1 - \frac{1}{4 \cdot 0{,}01^2 \cdot 10040} = 0{,}7510\,.$$

Ez azt jelenti, hogy annak a valószínűsége, hogy a megkérdezettek közötti igenek arányának eltérése a teljes lakosság körében az igenek arányától legfeljebb $0{,}01$, *legalább* $75{,}1\%$. Vegyük észre, hogy mivel p értékét nem ismerjük, így a tényleges valószínűséget kiszámolni nem tudjuk.

b) A megkérdezett emberek számának növelésével növekszik az az alsó korlát, amit annak a valószínűségének szabunk meg, hogy a relatív gyakoriság a teljes aránytól abszolút értékben legfeljebb $0{,}01$-dal tér el. Írjuk fel a nagy számok törvényét az alábbi formában:

$$P\left(\left|\frac{k}{n}-p\right|<\varepsilon\right)\geq 1-\frac{1}{4\cdot\varepsilon^2\cdot n}\geq 0{,}99\,.$$

Figyeljük meg, hogy elég nagy n esetén egyből elég kicsi számot vonunk ki ahhoz, hogy a különbség $0{,}99$ felett legyen. Ezáltal a keresett valószínűségünk is 99% felett lesz. Tehát fejezzük ki n-et az alábbi egyenlőtlenségből. A jobboldali két kifejezést átrendezve:

$$1-\frac{1}{4\cdot\varepsilon^2\cdot n}\geq 0{,}99 \qquad \text{azaz} \qquad n\geq\frac{100}{4\cdot\varepsilon^2}=\frac{100}{4\cdot 0{,}01^2}=250000\,.$$

Tehát ha legalább 250000 embert megkérdezünk, el tudjuk mondani, hogy annak a valószínűsége, hogy a közvélemény-kutatással kapott igenek relatív gyakorisága $0{,}01$-nél közelebb van a teljes populációban lévő igenek gyakoriságához, legalább 99%. Ez az eddig megkérdezettek számának majdnem 25-szöröse.

⇐18.6. feladat

18-3. önálló feladat: Egy nyerőautomata minden egyes játék esetén egymástól függetlenül 10^{-6} valószínűséggel adja oda a játékosnak a főnyereményt. A nagy számok törvénye alapján azt gondolja az ember, hogy egyszer biztosan nyerni fog. Adjunk becslést arra, hány játékot kell játszani ahhoz, hogy legalább 95% valószínűsége legyen annak, hogy megüssük a főnyereményt!

Leckezáró önellenőrző kérdéssor

1. A Bernoulli-féle nagy számok törvényének felírt formájában melyik szimbólum lesz valószínűségi változó?

☐ k ☐ n ☐ p ☐ egyik sem

2. Milyen eloszlású lesz az előbb kiválasztott valószínűségi változó?

☐ karakterisztikus ☐ normális ☐ binomiális ☐ egyik sem

3. Melyik állítás fogalmazza meg legjobban a tanult Bernoulli-féle nagy számok törvényét?

☐ Ha valamit elég sokszor próbálunk, akkor egyszer biztos sikerülni fog.

☐ Ha valamit sokszor próbálunk, akkor egyre nagyobb valószínűséggel fog sikerülni.

☐ Akármilyen pici esély is van valamire, ha minél többet próbálkozunk, annál nagyobb a valószínűsége annak, hogy az egyik próbálkozás sikerrel jár a sok közül.

☐ Akármilyen pici esély is van valamire, ha elég sokszor próbáljuk, az egyre nagyobb valószínűséggel fog bekövetkezni.

4. A vonatok késési gyakoriságát vizsgáljuk egy adott vonalon. Azt tapasztaljuk, hogy az előző év 100 utazásából a vonat 23 esetben késett. Adjunk becslést a Bernoulli-féle nagy számok törvény segítségével arra a valószínűségre, hogy a kapott arány maximum $0{,}1$-del tér el a tényleges valószínűségtől. Feltételezhetjük, hogy a vonatok mindig azonos valószínűséggel, egymástól független késnek.

☐ $< 0{,}5$ ☐ $\approx 0{,}23$ ☐ $\leq 0{,}25$ ☐ $\geq 0{,}75$

5. Egy pénzérméről szeretnénk eldönteni azt, hogy szabályos vagy nem. Ezer feldobás után 564 fejet kaptunk. A nagy számok törvényének segítségével adjunk becslést annak a valószínűségére, hogy a kapott relatív gyakoriság eltérése a fejdobás valószínűségétől legalább $0{,}064$?

☐ $\approx 28\%$ ☐ $\geq 93{,}9\%$ ☐ $\geq 36\%$ ☐ $\leq 6{,}1\%$

32. LECKE

Modulzáró feladatok 5.

19. Modulzáró feladatok

Start.

A modulzáró megoldására 1 órája van. Legalább 4 tizedesjegy pontossággal adja meg az eredményeket.

1. Legyen az X valószínűségi változó várható értéke 6, szórása pedig 12. Az $X^2 > 400$ valószínűséget szeretnénk becsülni, Markov-, vagy Csebisev-egyenlőtlenséggel.

   - [ ] A Markov-egyenlőtlenség ad erősebb alsó korlátot.
   - [ ] A Markov-egyenlőtlenség ad erősebb felső korlátot.
   - [ ] A Csebisev-egyenlőtlenség ad erősebb alsó korlátot.
   - [ ] A Csebisev-egyenlőtlenség ad erősebb felső korlátot.

2. Az előző feladatban meghatározott korlát értéke

3. A valószínűségszámítás zárthelyi dolgozatot a diákok 5 grammos $A4$-es papírra írják. Az egyes papírok tömege egymástól független, a tömeg szórása 40 mg. Adjunk korlátot annak a valószínűségére, hogy 400 darab dolgozat tömege a 2 kilót legalább 2 grammal meghaladja.

   - [ ] A Markov-egyenlőtlenség ad erősebb alsó korlátot.
   - [ ] A Markov-egyenlőtlenség ad erősebb felső korlátot.
   - [ ] A Csebisev-egyenlőtlenség ad erősebb alsó korlátot.
   - [ ] A Csebisev-egyenlőtlenség ad erősebb felső korlátot.

4. Az előző feladatban meghatározott korlát értéke

5. A zárthelyi dolgozatok úgy lettek kalibrálva, hogy egy dolgozat eredményének várható értéke 6 pont legyen, 2 pont szórással. Adjunk becslést annak a valószínűségére, hogy az idei 400 dolgozat átlagpontszáma meghaladja a tavalyi 5,63-at.

- ☐ A Markov-egyenlőtlenség használható, ez ad erősebb alsó korlátot.
- ☐ A Markov-egyenlőtlenség használható, ez ad erősebb felső korlátot.
- ☐ A Csebisev-egyenlőtlenség vagy a nagy számok törvénye használható, ez ad erősebb alsó korlátot.
- ☐ A Csebisev-egyenlőtlenség vagy a nagy számok törvénye használható, ez ad erősebb felső korlátot.

6. Az előző feladatban meghatározott korlát értéke

7. Egy termék gyártása során keletkezett selejtek arányát vizsgáljuk. Adjunk becslést a nagy számok törvénye alapján, hogy hány terméket kell megvizsgálnunk annak érdekében, hogy annak a valószínűsége, hogy a vizsgálat során kapott selejtek relatív gyakorisága a tényleges selejtaránytól való eltérése legalább 0,05, maximum 5% legyen.

Stop. Pontok:

VI. MODUL

Határeloszlás tételek

20. Határeloszlás tételek

A normális eloszlás a gyakorlatban (és az elméletben is) az egyik legtöbbet használt eloszlás. Ennek egyik oka a központi határeloszlás tétel, amely azt fejezi ki, hogy sok, független valószínűségi változó összege igen általános feltételek mellett közelítőleg normális eloszlású. Erre egy példa a mérések pontatlansága: a teljes mérési hiba ugyanis sok apró hibából áll össze. A központi határeloszlás tétel miatt így a teljes mérési hiba normális eloszlású, ezért a normális eloszlást normális hibatörvénynek is szokás nevezni.

A tétel nem keveset állít: az eddig tanult nevezetes diszkrét, vagy folytonos, vagy akár teljesen ismeretlen eloszlások is – amennyiben a kritériumokat teljesítik – elég nagy mennyiségben hasonlóan viselkednek. A tétel egyfajta rendet mutat ki a káoszban, hiszen nem csupán arról van szó, hogy a véletlent vizsgáljuk nagy mennyiségben, hanem épp a mennyiség növekedésével lesz a tétel egyre pontosabb.

A központi határeloszlás tételt centrális határeloszlás tételnek is szokás nevezni, innen ered a gyakori „CHT" rövidítés. Érdekes megemlíteni, hogy a tétel első formája Abraham de Moivre-tól származik még 1733-ból. De Moivre azt becsülte a normális eloszlás segítségével, hogy egy érmét sokszor feldobva hány dobás eredménye lesz fej. Pierre-Simon Laplace terjesztette ki a tételt a binomiális eloszlásra majd egy évszázaddal később, 1812-ben. A XX. század elején, 1901-ben fogalmazta meg Alekszandr Ljapunov a centrális határeloszlás tételt a tanult formában. A Ljapunov feltétel teljesülése esetén a határeloszlás tétel akkor is működik, ha a valószínűségi változók eloszlása nem azonos. Szükséges és elégséges feltételt a finn Jarl Waldemar Lindeberg publikált 1920-ban, amit azóta Lindeberg feltételnek neveznek.

A centrális határeloszlás tétele a független valószínűségi változók szórásának létét követeli meg, ebben az esetben a valószínűségi változók átlagának eloszlása a normális eloszláshoz közelít. Léteznek más határeloszlás tételek, amelyeket akkor használhatunk, ha nincs az adott eloszlásnak véges szórása vagy várható értéke. Ekkor a független valószínűségi változók átlagának eloszlása egy a normálishoz hasonló úgynevezett stabil eloszláshoz konvergál.

33. LECKE

A de Moivre–Laplace-tétel

20.1. A de Moivre–Laplace-tétel

Először a nagy kísérletszámú binomiális eloszlásra kimondott de Moivre–Laplace-tételt nézzük.

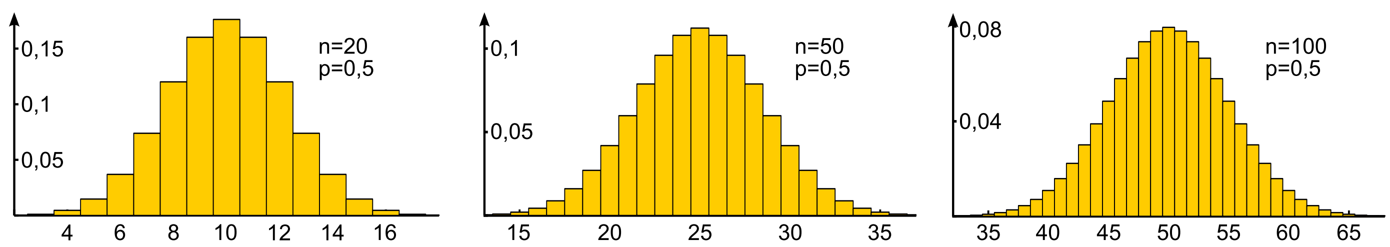
20.1. tétel: Legyen X binomiális eloszlású valószínűségi változó az n és p paraméterekkel (ebben az esetben X várható értéke $n \cdot m$, szórása $\sqrt{np(1-p)}$). Ekkor elég nagy n (nagyobb mint 30) esetén

1. annak valószínűsége, hogy X értéke pontosan k, közelíthető egy azonos várható értékkel és szórással bíró normális eloszlás sűrűségfüggvényének segítségével (a de Moivre–Laplace-tétel lokális alakja):

$$P(X=k) \approx \frac{1}{\sqrt{np(1-p)}} \cdot \varphi\left(\frac{k-n\cdot p}{\sqrt{np(1-p)}}\right) = \frac{1}{\sqrt{2\pi np(1-p)}} \cdot e^{-\frac{(k-n\cdot p)^2}{2np(1-p)}}.$$

2. annak a valószínűsége, hogy X értéke a és b közé esik, becsülhető egy azonos várható értékkel és szórással bíró normális eloszlás eloszlásfüggvényének segítségével (a de Moivre–Laplace-tétel globális alakja):

$$P(a \le X \le b) \approx \Phi\left(\frac{b-n\cdot p}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-n\cdot p}{\sqrt{np(1-p)}}\right).$$



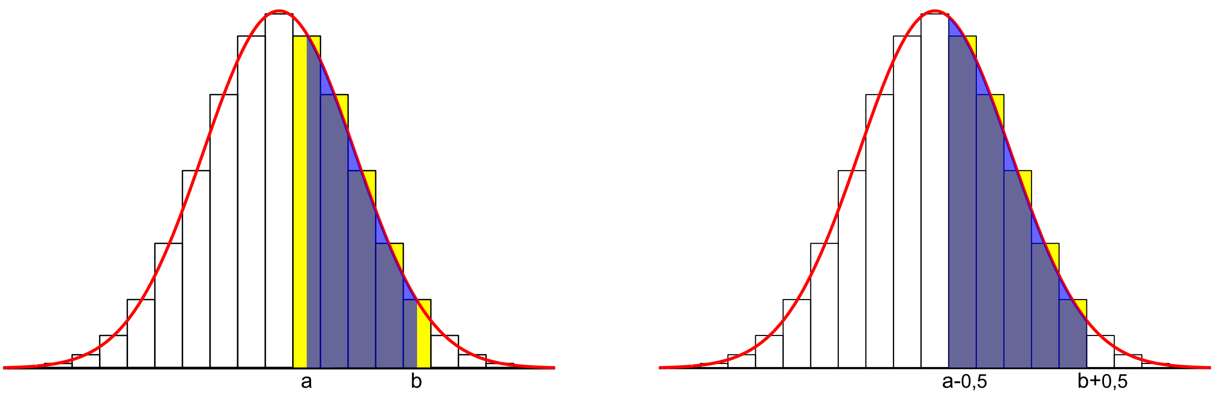
20.1. ábra. *Az ábrákon a binomiális eloszlást figyelhetjük meg* $p = 0{,}5$ *mellett úgy, hogy* n *értéke egyre nagyobb lesz. Bal oldalon* $n = 20$, *majd* $n = 50$ *középen, s a jobb szélen* $n = 100$. *Figyeljük meg az eloszlások alakját, ahogy balról jobbra haladva egyre jobban hasonlít egy haranggörbéhez!*

A becslés annál pontosabb, minél nagyobb az n paraméter, s minél közelebb van a p valószínűség az $1/2$-hez. Természetesen az ember azt várja, hogy a globális alak valamennyire visszaadja a lokális alakot. Ezzel szemben $a = b$ esetén a globális alak 0-nak becsüli a $P(X = a)$ valószínűséget. Ugyancsak problémás a becslés, amennyiben a és b közel van egymáshoz.

A de Moivre–Laplace-tételnél jobb becslést ad a tétel úgynevezett $1/2$-del korrigált változata:

20.2. tétel: Legyen X binomiális eloszlású valószínűségi változó az n és p paraméterekkel. Ekkor amennyiben n elég nagy (nagyobb, mint 30), annak a valószínűsége, hogy X valamely a és b egész számok által határolt intervallumba esik, becsülhető a normális eloszlás eloszlásfüggvényével:

$$P(a \leq X \leq b) \approx \Phi\left(\frac{b + 0.5 - n \cdot p}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - 0.5 - n \cdot p}{\sqrt{np(1-p)}}\right).$$



20.2. ábra. *Az ábrákon a de Moivre–Laplace-tétel eredeti (balra), és $1/2$-del korrigált közelítését (jobbra) látjuk. A binomiális eloszlás értékei oszlopdiagrammal, a normális eloszlás görbével van ábrázolva. A $P(a \leq x \leq b)$ valószínűség pontos értéke az oszlopdiagram alatti sárga, a normális eloszlással közelített érték pedig a görbéje alatti kék terület. A jobb oldali közelítés nem csak a két oldalt látható pontatlanságokat korrigálja, hanem az oszlopdiagram cikázottságait is első rendben közelíti.*

A becslés annál pontosabb, minél nagyobb az n paraméter, s minél közelebb van a p valószínűség az $1/2$-hez. A becslést és a korrekció hatását a fenti ábrán figyelhetjük meg.

20.1. feladat. Egy szabályos pénzérmét százszor dobunk fel.

a) A de Moivre–Laplace-tétel segítségével adjunk becslést annak a valószínűségére, hogy épp annyi fejet dobunk, amennyi a fejdobások számának várható értéke!

b) Számoljuk ki a pontos értéket is, és hasonlítsuk össze a határeloszlás tétel által adott becsléssel!

c) Adjunk meg a határeloszlás tétel segítségével egy olyan, a várható értékre szimmetrikus intervallumot, ahova a dobott fejek száma 99%-os valószínűséggel beleesik!

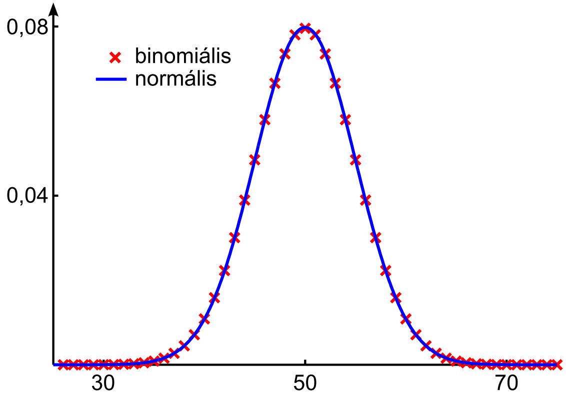
d) Adjunk becslést egy alkalmas tanult egyenlőtlenség segítségével annak a valószínűségére, hogy a fej dobások száma a számolt intervallumba kerül!

e) Számoljuk ki az előző feladatban becsült valószínűség pontos értékét is, és hasonlítsuk össze az eddig kapott értékekkel!

Megoldás:

A dobott fejek X száma binomiális eloszlást fog követni, ahol $p = 1/2$ a fej dobás bekövetkezésének valószínűsége és $n = 100$ a dobások, azaz a kísérletek száma. Így a dobott fejek számának várható értéke $E(X) = m = n \cdot p = 100 \cdot 1/2 = 50$, szórása pedig $D(X) = \sigma = \sqrt{np(1-p)} = \sqrt{100 \cdot 1/2 \cdot 1/2} = 5$.

a) Először adjunk becslést a de Moivre–Laplace-tétel segítségével arra, mekkora annak a valószínűsége, hogy épp annyi fejet dobunk, mint a fej dobások számának várható értéke. A fej dobások számának várható értéke 50, így azt kell becsülnünk, mekkora valószínűséggel lesz X értéke 50. A határeloszlás tétel elsőnek megfogalmazott becslése alapján a valószínűséget egy olyan X_N, normális eloszlású valószínűségi változó sűrűségfüggvényének segítségével becsüljük, amelynek várható értéke és szórása megegyezik X várható



20.3. ábra. *Az ábra az $n = 100$, $p = 1/2$ binomiális eloszlást és az őt közelítő normális eloszlást mutatja.*

értékével és szórásával, azaz $X_N \sim N(50; 5)$. X_N sűrűségfüggvénye az alábbiak szerint írható fel:

$$f_{X_N}(x) = \frac{1}{\sigma} \cdot \varphi\left(\frac{x-m}{\sigma}\right),$$

ahol φ a standard normális eloszlás sűrűségfüggvénye. Így a becslést az alábbiak szerint tehetjük meg:

$$P(X = 50) \approx f_{X_N}(50) = \frac{1}{\sigma} \cdot \varphi\left(\frac{50-m}{\sigma}\right) = \frac{1}{5} \cdot \varphi(0) = \frac{1}{5\sqrt{2\pi}} \cdot e^0 \approx 0{,}079788\,.$$

Tehát körülbelül 8% annak a valószínűsége, hogy 100 dobásból épp annyi fejet dobunk, mint amennyi a fej dobások számának várható értéke, azaz 50-et.

b) Az előbb becsült eredményt közvetlenül is kiszámolhatjuk, hiszen X eloszlása binomiális, és ismerjük n-et és p-t is. A számolást az alábbiak szerint végezhetjük el:

$$P(X = 50) = \binom{100}{50} \cdot \frac{1}{2^{50}} \cdot \frac{1}{2^{50}} \approx 0{,}079589\,.$$

Hasonló eredményt kaptunk, mint az előbb, tehát körülbelül 8% a keresett valószínűség a binomiális eloszlás segítségével számolva is. Észrevehetjük azonban, hogy az értékek közötti különbség $\approx 0{,}0002$. Ez a hiba viszont kisebb, mint a pontos valószínűség értékének fél százaléka. Így a becslést akár pontosnak is nevezhetjük.

c) Keressünk most a határeloszlás tétel segítségével egy olyan, a várható értékre szimmetrikus intervallumot, amelybe a dobott fejek száma legalább 99%-os valószínűséggel beleesik. Jelöljük x-szel az intervallum végpontjainak távolságát a várható értéktől. Így az intervallumot felírhatjuk az $[m-x; m+x]$ alakban. Úgy kell meghatároznunk x-et, hogy $P(m-x \leq X \leq m+x) = 0{,}99$ legyen.
Figyeljünk azonban a következőre: mivel X eloszlása diszkrét, így az intervallum határának folytonos mozgatása esetén a kérdéses valószínűség nem folytonosan, hanem ugrásszerűen fog mozogni, így azt 99%-ra csak körülbelül tudjuk beállítani. Ezt a bizonytalanságot növeli az a tény is, hogy a határeloszlás tétel alkalmazásával csak közelíteni tudja a binomiális eloszlás konkrét értékeit, így az eredményünk is csak annyira lesz jó, amennyire ez a közelítés jó.
Írjuk fel a vizsgált valószínűséget úgy, hogy az X_N, normális eloszlású valószínűségi változóval közelítjük:

$$0{,}99 = P(m-x \leq X \leq m+x) \approx P(m-x-0{,}5 < X_N < m+x+0{,}5) =$$
$$= \Phi\left(\frac{m+x+0{,}5-m}{\sigma}\right) - \Phi\left(\frac{m-x-0{,}5-m}{\sigma}\right) = 2\Phi\left(\frac{x+0{,}5}{\sigma}\right) - 1\,.$$

Az egyenlet két végét összevetve kapjuk, hogy

$$0{,}99 \approx 2\Phi\left(\frac{x+0{,}5}{\sigma}\right) - 1 \qquad \text{azaz} \qquad \Phi\left(\frac{x+0{,}5}{\sigma}\right) \approx 0{,}995\,.$$

A normális eloszlás táblázat segítségével ki tudjuk keresni azt az értéket, melynél Φ értéke $0{,}995$ lesz. Ez az érték a $2{,}5758$ lesz. Tehát

$$\frac{x+0{,}5}{\sigma} = 2{,}5758 \qquad \text{amiből} \qquad x \approx 5 \cdot 2{,}5758 - 0{,}5 \approx 12{,}379\,.$$

Mivel X eloszlása diszkrét, várható értéke is egész számra esik, így x tizedes utáni jegyei nem számítanak. Kerekítsünk a közelebbi egészhez, $x = 12$-re. Ekkor a keresett intervallumunk az $[50-12; 50+12]$, azaz a $[38; 62]$ lesz. Úgy becsüljük, ide fog beleesni nagyjából 99% valószínűséggel a dobott fejek száma.

d) Adjunk becslést alkalmas tanult egyenlőtlenség segítségével arra, hogy a dobott fejek X száma a most számolt $[38; 62]$ intervallumba esik. A becsléshez alkalmazhatjuk a Bernoulli-féle nagy számok törvényét is, viszont a Csebisev-egyenlőtlenséget közvetlenül felírhatjuk, hiszen ismerjük X várható értékét és szórását is. Vegyük figyelembe, hogy mivel X diszkrét, érdemesebb lesz a $[38; 62]$ zárt intervallum helyett a $(37; 63)$ intervallummal felírni a becslést:

$$P(38 \leq X \leq 62) = P(37 < X < 63) = P(|X-m| < 13) \geq 1 - \frac{5^2}{13^2} \approx 0{,}8521\,.$$

Tehát annak a valószínűsége, hogy X a $[38; 62]$ zárt intervallumba esik, *legalább* 85,21%. Ez az eredmény nem mond ellent annak, hogy olyan intervallumot keressünk, ahova a dobott fejek száma 99% valószínűséggel esik bele, hiszen $0{,}99 \geq 0{,}8521$.

e) Végül határozzuk meg pontosan annak a valószínűségét, hogy a dobott fejek száma a $[38; 62]$ intervallumba esik. Írjuk fel az alábbi módon a keresett valószínűséget:

$$P(38 \leq X \leq 62) = P(38 \leq X \leq 49) + P(X = 50) + P(51 \leq X \leq 62)\,.$$

Vegyük észre, hogy mivel $p = 0{,}5$, ezért X eloszlása szimmetrikus a várható értékre, tehát $P(X = 50 - i) = P(X = 50 + i)$, ebből következik, hogy $P(38 \leq X \leq 49) = P(51 \leq X \leq 62)$. Ezen kívül a $P(X = 50)$ valószínűséget már kiszámoltuk. A nem ismert valószínűséget kézzel, vagy szoftver segítségével számolhatjuk. Összegezve:

$$P(38 \leq X \leq 62) \approx 0{,}454189 + 0{,}079589 + 0{,}454189 = 0{,}987967\,.$$

Tehát annak a valószínűsége, hogy X a $[38; 62]$ intervallumba esik, 98,79%. Láthatjuk, hogy ez kicsit 99% alatt van. Ezért ellenőrizzük annak a valószínűségét is, hogy X a $[37; 63]$ intervallumba esik. Ez 99,33% lesz. Láthatjuk, hogy kizárólag a most említett két intervallumot fogadhatjuk el, mint a feladat eredménye, s a *c)* feladatban megjelölt megoldás esetén közelebb van a vizsgált valószínűség a 99%-hoz.

⇐20.1. feladat

Aktivitás: Alkalmas szoftver segítségével számolja ki, majd ábrázolja grafikonon, hogy miként függ az előző feladatban vizsgált szimmetrikus intervallum hossza attól a valószínűségtől, hogy a dobott fejek száma beleesik.

Vizsgáljuk meg miként viselkedik a normális eloszlással való közelítés akkor, amikor a p valószínűségünk kicsi.

20.2. feladat.

Egy irodaházban 100 telefonkészülék üzemel. Ezek naponta, egymástól független 4% valószínűséggel romlanak el. A hibás készülékeket még aznap este cserélik.

a) Mi a valószínűsége annak, hogy egy adott nap négynél kevesebb készülék romlik el?

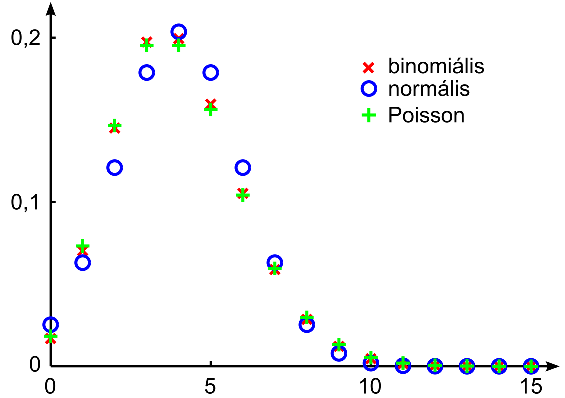
b) A pontos érték mellett adjunk becslést Poisson- és normális eloszlással való közelítés alapján is!

Megoldás:

a) Jelölje X az adott nap elromlott készülékek számát, az X valószínűségi változó binomiális eloszlású lesz, $p = 0{,}04$ és $n = 100$ paraméterekkel. A feladat a $P(X < 4)$ valószínűség kiszámítását kéri. Mivel a binomiális eloszlás diszkrét, és csak egész értékeket vehet fel 0 és n között, a kért valószínűséget az alábbi módon számíthatjuk ki:

$$\begin{aligned}
P(X<4) &= P(0 \le X \le 3) = P(X=0)+P(X=1)+P(X=2)+P(X=3)\\
&= \binom{100}{0}\cdot 0{,}04^0 \cdot 0{,}96^{100} + \binom{100}{1}\cdot 0{,}04^1 \cdot 0{,}96^{99} + \binom{100}{2}\cdot 0{,}04^2 \cdot 0{,}96^{98} + \binom{100}{3}\cdot 0{,}04^3 \cdot 0{,}96^{97}\\
&\approx 0{,}01687 + 0{,}07029 + 0{,}14498 + 0{,}19733 = 0{,}4295\,.
\end{aligned}$$

Tehát annak a valószínűsége, hogy egy adott napon négynél kevesebb készülék romlik el, $42{,}95\%$ lesz. Az eredményt tekinthetjük négy tizedesjegyig pontosnak, ha a számolást megfelelő pontossággal végeztük.



20.4. ábra. *Az ábra az $n = 100$, $p = 0{,}04$ binomiális eloszlást és az őt közelítő Poisson-, illetve normális eloszlást mutatja.*

b) A vizsgált valószínűséget becsüljük most Poisson-eloszlással. Ehhez meg kell választani a becslésre használt X_P valószínűségi változó λ paraméterét. Úgy kell választani, hogy X és X_P várható értéke megegyezzen, így

$$\lambda = n \cdot p = 4\,.$$

Tehát X_P egy $\lambda = 4$ paraméterű Poisson-eloszlású valószínűségi változó lesz. A becslést az alábbi módon tehetjük meg:

$$\begin{aligned} P(X < 4) \approx P(X_P < 4) = P(X_P = 0) + P(X_P = 1) + P(X_P = 2) + P(X_P = 3) = \\ = \frac{4^0}{0!} \cdot e^{-4} + \frac{4^1}{1!} \cdot e^{-4} + \frac{4^2}{2!} \cdot e^{-4} + \frac{4^3}{3!} \cdot e^{-4} \approx \\ \approx 0{,}01832 + 0{,}07326 + 0{,}14652 + 0{,}19537 = 0{,}4335\,. \end{aligned}$$

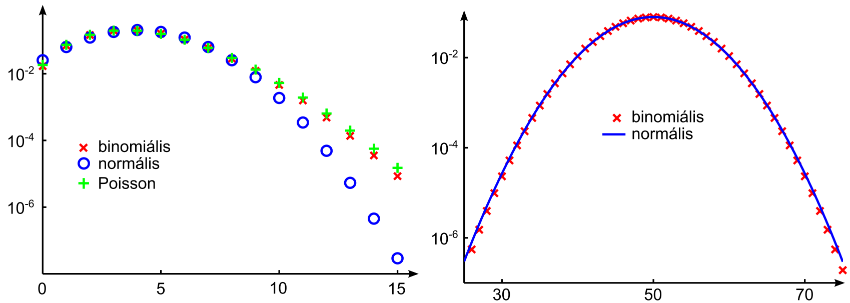
Tehát a vizsgált valószínűséget Poisson-eloszlás segítségével számolva annak a valószínűsége, hogy az adott nap 4-nél kevesebb készülék romlik el 43,35% lesz. Láthatjuk, hogy a közelítés 4 ezredre pontos, azonban a számérték kiszámítása kevesebb, és egyszerűbb műveleteket igényel.

c) Végül normális eloszlás segítségével is számítsuk ki a valószínűség értékét. A binomiális eloszlású X-et a de Moivre–Laplace-tétel segítségével a normális eloszlású X_N-nel közelítjük. X_N várható értékét és szórását úgy kell megválasztani, hogy az megegyezzen X várható értékével és szórásával, azaz

$$m = E(X_N) = E(X) = 4, \qquad \text{valamint}$$
$$\sigma = D(X_N) = D(X) = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{100 \cdot 0{,}04 \cdot 0{,}96} \approx 1{,}9595\,.$$

Így a kérdéses valószínűséget a de Moivre–Laplace-tétel segítségével az alábbi módon becsülhetjük:

$$P(X < 4) = P(0 \le X \le 3) \approx P(-0{,}5 < X_N < 3{,}5) =$$
$$= \Phi\left(\frac{3{,}5-4}{\sqrt{3{,}84}}\right) - \Phi\left(\frac{-0{,}5-4}{\sqrt{3{,}84}}\right) \approx 1 - \Phi(0{,}2552) - [1 - \Phi(2{,}2965)] = 0{,}3885\,.$$



20.5. ábra. *A normális eloszlás nem csak kis valószínűségek esetén pontatlan. A várható értéktől távolodva több nagyságrendben rosszabb becslést ad a Poissonnál ($n = 100$, $p = 0{,}04$, balra). A pontatlanság megfigyelhető az ideális valószínűség esetén is az eloszlás szélein ($n = 100$, $p = 0{,}5$, jobbra).*

A határeloszlás tétel segítségével kapott becslés messze nem olyan pontos, mint a Poisson-eloszlás segítségével végzett becslés. Majdnem 10%-os eltérés van a két érték között.

A feladat alapján láthatjuk, hogy kis valószínűségek esetén a binomiális eloszlást pontosabban közelíthetjük Poisson-eloszlás segítségével.

⇐20.2. feladat

20-1. önálló feladat: Vizsgálja meg, miként változik az előző feladatban leírt közelítések pontossága, ha az $n \cdot p = \lambda$ értéket fixen tartjuk, és közben n értékét minden lépésben 10-szeresére növeljük.

20.3. feladat. Árpád, a hatvanas születésű hajóskapitány, 400 férőhelyes sétahajót üzemeltet a Balatonon. A tapasztalat szerint, ha Árpád épp annyi foglalást fogad, amennyi a férőhelyek száma, akkor átlagosan 350 utas jelenik meg ténylegesen a hajó indulásakor. Feltehetjük, hogy az utasok egymástól függetlenül, azonos valószínűséggel döntenek arról, megjelennek-e az utazáson. Árpád a bevétel növelése érdekében több foglalást akar fogadni.

a) Mi a valószínűsége annak, hogy a megjelenő utasok felférnek a hajóra, ha Árpád 10%-kal több foglalást fogad, mint a hajó kapacitása.

b) Hány foglalást fogadhat maximum, ha Árpád azt akarja, hogy 95% valószínűséggel mindegyik utas felférjen, aki megjelenik az induláskor?

Megoldás: A feladatban az utazásra ténylegesen megjelenő utasok száma véletlen, így ezt jelöljük az X valószínűségi változóval. Állapítsuk meg X eloszlását. Egy adott útra n utas ad le foglalást, s mindegyikük egymástól függetlenül, azonos p valószínűséggel jelenik meg. Ezért X eloszlása binomiális.

A feladat megoldásához először meg kell állapítani az X valószínűségi változó eloszlásának n és p paraméterét. Az n paraméter részfeladatonként változik, azonban annak a valószínűsége, hogy egy adott utas eljön, állandó lesz. Tudjuk, hogy $n = 400$ esetén $E(X) = 350$ lesz. Így, mivel az eloszlás binomiális:

$$350 = E(X) = n \cdot p = 400 \cdot p, \qquad \text{azaz} \qquad p = \frac{E(X)}{n} = \frac{350}{400} = 0{,}875\,.$$

Tehát az egyes utasok egymástól függetlenül $87{,}5\%$ valószínűséggel jelennek meg. Ez az érték független a foglalások számától.

a) Mi a helyzet akkor, ha a hajón 10%-os túlfoglalás van, azaz Árpád 400 helyett 440 foglalást fogad? Ebben az esetben $n = 440$, s annak a valószínűségét kell kiszámolnunk, hogy X, azaz a megjelenő utasok száma maximum 400. Mivel X ebben az esetben összesen 401-féle értéket vehet fel 0-tól 400-ig, a binomiális eloszlás közvetlen alkalmazása technikailag korlátozott. Érdemes így a valószínűség becslésére a de Moivre–Laplace-tételt alkalmazni. Az X_N normális eloszlású valószínűségi változó, mellyel X-et közelítjük, várható értékben és szórásban meg fog egyezni az eredeti eloszlás várható értékével és szórásával:

$$\begin{aligned} m &= E(X_N) = E(X) = n \cdot p = 440 \cdot 0{,}875 = 385, \qquad \text{valamint} \\ \sigma &= D(X_N) = D(X) = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{440 \cdot 0{,}875 \cdot 0{,}125} \approx 6{,}9372\,. \end{aligned}$$

A vizsgált valószínűséget így az alábbiak szerint becsülhetjük:

$$P(X \leq 400) \approx P(X_N < 400{,}5) = \Phi\left(\frac{400{,}5 - 385}{6{,}9372}\right) \approx \Phi(2{,}2343) \approx 0{,}9873\,.$$

Tehát úgy becsültük, hogy ha Árpád 400 helyett 440 foglalást ad ki, akkor $98{,}73\%$ valószínűséggel minden megjelenő utas fel is fér a hajóra. Binomiális eloszlással és megfelelő szoftver alkalmazásával számolva ez a valószínűség $98{,}98\%$. Az eredmény pontosnak látszik, viszont ha azt kérdezzük, hogy mi a valószínűsége annak, hogy *nem* fér fel valaki, akkor a binomiális eloszlással számolt érték $1{,}02\%$, a határeloszlással számolt becslés pedig $1{,}27\%$.

b) Láthatjuk az előző részfeladatból, hogy 440 utas esetén több mint 95% a valószínűsége annak, hogy minden utas felfér a hajóra. Ha több foglalást ad ki Árpád, akkor ez a valószínűség csökkenni fog. Kérdés, hogy meddig mehetünk még el úgy, hogy 95% felett maradjon annak a valószínűsége, hogy mindenki felfér. A kiadott foglalások számát nem ismerjük, így az X valószínűségi változó n paramétere ismeretlen. Ez azt jelenti, hogy az őt közelítő X_N normális eloszlású valószínűségi változó várható értékét és szórását csak n függvényeként tudjuk megadni:

$$m = E(X_N) = E(X) = n \cdot p = 0{,}875 \cdot n, \qquad \text{valamint}$$

$$\sigma = D(X_N) = D(X) = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{n \cdot 0{,}875 \cdot 0{,}125} \approx 0{,}33072 \cdot \sqrt{n}\,.$$

Azt vizsgáljuk, hogy milyen valószínűséggel férnek fel az utasok a hajóra, azaz mekkora valószínűséggel lesz X értéke legfeljebb 400, hiszen azt szeretnénk, hogy ez 95% legyen:

$$95\% = P(X \le 400) \approx P(X_N < 400{,}5) = \Phi\left(\frac{400{,}5 - 0{,}875 \cdot n}{0{,}33072 \cdot \sqrt{n}}\right).$$

A fenti kifejtésben X eloszlását X_N eloszlásával közelítettük. Vegyük észre, hogy a jobb oldali Φ függvény értéke a bal szélen $0{,}95$, amit Φ az $1{,}645$ helyen vesz fel, így

$$\frac{400{,}5 - 0{,}875 \cdot n}{0{,}33072 \cdot \sqrt{n}} = 1{,}645, \qquad \text{azaz} \qquad 0{,}875 \cdot n + 0{,}54403 \cdot \sqrt{n} - 400{,}5 = 0\,.$$

Vegyük észre, hogy a jobb oldali egyenlet egy $\sqrt{n}$-re másodfokú egyenlet. Ha $\sqrt{n}$ helyére x-et írunk, akkor

$$0{,}875 \cdot x^2 + 0{,}54403 \cdot x - 400{,}5 = 0, \quad \text{melynek a megoldásai} \quad x_1 \approx -21{,}707 \quad \text{és} \quad x_2 \approx 21{,}086\,.$$

Mivel $x_1 < 0$, ezért ő nem lehet $\sqrt{n}$ értéke, így $\sqrt{n} \approx 21{,}086$ és $n \approx 444{,}6$ lesz. Azt kaptuk, hogy ha a megjelenő utasok számát de Moivre–Laplace-tétel szerint közelítjük, akkor legfeljebb 444 foglalás esetén fog több mint 95% valószínűséggel felférni a hajóra mindenki. Közvetlenül a binomiális eloszlással, szoftver segítségével kiszámíthatjuk, hogy 444 foglalás esetén $96{,}1\%$, 445 foglalás esetén pedig $94{,}8\%$ valószínűséggel fog minden megjelenő utas felférni a hajóra. Tehát a határeloszlás tételt alkalmazva sikerült a pontos eredményt meghatározni.

⇐20.3. feladat

20-2. önálló feladat: Hogyan változnak az előző feladat eredményei akkor, ha minden utas kizárólag párosával utazik, így mind a foglalás, mind a megjelenés páronként történik?

Aktivitás: Nézzen utána az interneten, hogy az utaztatással, esetleg vendéglátással foglalkozó cégek mekkora túlfoglalással szoktak dolgozni!

Leckezáró önellenőrző kérdéssor

1. A de Moivre–Laplace-tételt szeretnénk arra használni, hogy egy szabályos érmét sokszor feldobva a fej dobások számát vizsgáljuk. A fej dobások számának becsléséhez használt normális eloszlás melyik paramétere fog változni, ha növeljük az érmedobások számát?

☐ várható érték ☐ szórás ☐ mindkettő ☐ egyik sem

2. Annak a valószínűsége, hogy egy adott nap esik az eső, egy állandó p értéknek tekinthető. Alkalmazhatjuk-e a de Moivre–Laplace-tételt az egy évben előforduló esős napok számának leírására?

☐ Igen, mert $n = 365$ és a p bekövetkezés fix.

☐ Nem, mert az esős napok számának eloszlása binomiális.

☐ Igen, mert az esős napok számának eloszlása binomiális.

☐ Nem, hacsak nem mutatható meg, hogy az esős napok egymástól függetlenül jönnek, így az esős napok száma binomiális lesz.

3. A de Moivre–Laplace-tétel segítségével becsüljük meg annak a valószínűségét, hogy egy szabályos pénzérmét 100-szor feldobva egyszer sem dobunk írást.

☐ $\approx 7 \cdot 10^{-31}$ ☐ $\approx 2 \cdot 10^{-23}$ ☐ $\approx 10^{-28}$ ☐ 0

4. A de Moivre–Laplace-tétel segítségével becsüljük meg annak a valószínűségét, hogy 100 kockadobásból legalább 20 hatos lesz.

☐ $\approx 15\%$ ☐ $\leq 1/6$ ☐ $\geq 5/6$ ☐ $\approx 22\%$

34. LECKE

A központi határeloszlás tétel

20.2. A központi határeloszlás tétel

Az alábbi tétel a központi határeloszlás tétel legegyszerűbben megfogalmazott alakja. A tétel legfontosabb feltételei a függetlenség és a szórások végessége.

20.3. tétel: Legyenek $X_1, X_2, \ldots, X_n, \ldots$ független, azonos eloszlású valószínűségi változók, létező és azonos várható értékkel és szórással, azaz mindegyik X_i valószínűségi változóra $E(X_i) = m$ és $D(X_i) = \sigma$. Ebben az esetben az X_i valószínűségi változók összegének standardizáltja határesetben (azaz ha $n \to \infty$) standard normális eloszlású. Ez azt jelenti, hogy

$$\lim_{n\to\infty} P\left(\frac{X_1 + X_2 + \ldots + X_n - n\cdot m}{\sigma\sqrt{n}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,\mathrm{d}t = \Phi(x)\,.$$

A tétel segítségével az alábbi állításokat fogalmazhatjuk meg nem csak az összegre, hanem az átlagra is:

1. Elég nagy mennyiségű ($n > 30$), független, azonos eloszlású valószínűségi változó *összegének standardizáltja* közelítőleg standard normális eloszlású, azaz
$$P\left(\frac{X_1 + X_2 + \ldots + X_n - n\cdot m}{\sigma\sqrt{n}} < x\right) \approx \Phi(x)\,.$$
2. Elég nagy mennyiségű, független, azonos eloszlású valószínűségi változó *összege* olyan normális eloszlással közelíthető, melynek várható értéke $n \cdot m$, szórása pedig $\sigma\sqrt{n}$, azaz
$$P(X_1 + X_2 + \ldots + X_n < x) \approx \Phi\left(\frac{x - n\cdot m}{\sigma\sqrt{n}}\right)\,.$$

3. Elég nagy mennyiségű, független, azonos eloszlású valószínűségi változó *átlagának standardizáltja* közelítőleg standard normális eloszlású, azaz

$$P\left(\frac{\frac{X_1+X_2+\ldots+X_n}{n}-m}{\frac{\sigma}{\sqrt{n}}}<x\right)\approx\Phi(x)\,.$$

4. Elég nagy mennyiségű, független, azonos eloszlású valószínűségi változó átlaga olyan normális eloszlással közelíthető, melynek várható értéke m, szórása pedig $\frac{\sigma}{\sqrt{n}}$, azaz

$$P\left(\frac{X_1+X_2+\ldots+X_n}{n}<x\right)\approx\Phi\left(\frac{x-m}{\sigma/\sqrt{n}}\right)\,.$$

5. Mivel a binomiális eloszlású valószínűségi változó tekinthető n darab független, indikátor eloszlású valószínűségi változó összegének, ezért a de Moivre–Laplace-tétel valójában a központi határeloszlás tétel speciális esete, mikor az X_i-k karakterisztikus eloszlásúak, és mindegyik X_i valószínűségi változóra $E(X_i)=p$ és $D(X_i)=\sqrt{p(1-p)}$.

20.4. feladat. 400 darab független, azonos eloszlású valószínűségi változó mindegyikének várható értéke 12, szórása pedig 4.

a) Mi a valószínűsége annak, hogy átlaguk 11,5 és 12,5 között lesz?

b) Adjunk becslést a Csebisev-egyenlőtlenség segítségével is, majd hasonlítsuk össze az eredményeket!

Megoldás: Vegyük észre, hogy a valószínűségi változók eloszlása nincs megadva, csak a várható értékük és a szórásuk. Ahhoz, hogy a centrális határeloszlás tételt alkalmazni tudjuk, a valószínűségi változók várható értékére és szórására lesz szükségünk. Mivel ezek adottak, hiszen $m=12$ és $\sigma=4$, valamint végesek, így a határeloszlás tételt alkalmazhatjuk. Mind a határeloszlás, mind az egyenlőtlenség alkalmazásához szükségünk van a valószínűségi változók $\overline{X}$ átlagának várható értékére és szórására.

$$E(\overline{X}) = m = 12, \qquad \text{valamint} \qquad D(\overline{X}) = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{400}} = 0{,}2\,.$$

Tehát $\overline{X}$ várható értéke 12, szórása pedig $0{,}2$ lesz.

a) A centrális határeloszlás tétele alapján az átlag eloszlását egy olyan X_N normális eloszlású valószínűségi változóval tudjuk közelíteni, melynek várható értéke és szórása megegyezik az $\overline{X}$ várható értékével és szórásával. Így a határeloszlás tételt az alábbi módon tudjuk alkalmazni:

$$P(11{,}5 < \overline{X} < 12{,}5) \approx P(11{,}5 < X_N < 12{,}5) = \Phi\left(\frac{12{,}5-12}{0{,}2}\right) - \Phi\left(\frac{11{,}5-12}{0{,}2}\right) = 2\Phi(2{,}5) - 1 \approx 0{,}9876\,.$$

Tehát a centrális határeloszlás tételével becsülve annak a valószínűsége, hogy a valószínűségi változók átlaga $11{,}5$ és $12{,}5$ között lesz, körülbelül $98{,}76\%$.

b) A feladat második részében a valószínűség értékét a Csebisev-egyenlőtlenség segítségével becsüljük. Az $\overline{X}$ valószínűségi változó várható értékét és szórását már kiszámoltuk, az egyenlőtlenség felírásához ezek kellenek:

$$P(11{,}5 < \overline{X} < 12{,}5) = P\left(\left|\overline{X} - E(\overline{X})\right| < 0{,}5\right) \geq 1 - \frac{0{,}2^2}{0{,}5^2} = 0{,}84\,.$$

Tehát azt kaptuk, hogy annak a valószínűsége, hogy az átlag $11{,}5$ és $12{,}5$ között van, legalább 84%. Figyeljük meg, hogy ez az érték jelentősen eltér attól, mint amint az előző részfeladatban kaptunk. Az előző feladat eredménye viszont egy becslés, a mostani pedig egy szigorú alsó korlát.

⇐20.4. feladat

Aktivitás: Írjon fel legalább 2-3 olyan példát a való életből, ahol valamilyen véletlen mennyiséget egy haranggörbe formájú eloszlással tudunk legjobban jellemezni. Vizsgálja meg, hogy igaz-e, hogy a véletlen ingadozás több, apró hatás összességéből adódik.

20.5. feladat. Az e-mail fiókban naponta talált kéretlen reklámlevelek száma Poisson-eloszlást követ. Hétköznap átlagosan 4, a hétvégén pedig átlagosan 2 kéretlen levél jön egy nap alatt. Tegyük fel azt, hogy az egyes napokon kapott kéretlen levelek száma független egymástól.

a) Számoljuk ki az 52 hét alatt érkező kéretlen reklámlevelek számának várható értékét és szórását.

b) Mi a valószínűsége annak, hogy az 52 hét alatt több mint 1300 kéretlen levél érkezik?

c) Hány hét alatt gyűlik össze több mint 95% valószínűséggel legalább 1000 kéretlen levél?

Megoldás:

a) Először az 52 hét alatt összegyűlő reklámlevelek számának várható értékét és szórását számoljuk ki. Több Poisson-eloszlású valószínűségi változó összege is Poisson-eloszlású lesz, bár az összeg várható értékének és szórásának kiszámításához ez nem kell. Az 52 hét alatt hetente 5 hétköznap van, ami összesen 260 hétköznap, amelyeken 4 e-mailt kapunk átlagosan. A szombatok és vasárnapok összesen 104 napot adnak ki, ezeken a napokon 2 kéretlen e-mail jön átlagosan. Jelölje X a 364 nap alatt érkezett levelek számát. Ekkor X várható értéke:

$$E(X) = 52 \cdot 5 \cdot 4 + 52 \cdot 2 \cdot 2 = 1248\,.$$

A szórás kiszámítása előtt X szórásnégyzetét számítsuk ki. A λ paraméterű Poisson-eloszlású valószínűségi változó szórásnégyzete is λ. Mivel az X valószínűségi változó független, Poisson-eloszlású valószínűségi változók összege, így szórásnégyzete is a szórásnégyzetek összege.

$$D^2(X) = 52 \cdot 5 \cdot 4 + 52 \cdot 2 \cdot 2 = 1248\,.$$

Az eredmény abból is következik, hogy X eloszlása is Poisson, $\lambda = 1248$-as paraméterrel. Így X szórása $D(X) = \sqrt{1248} \approx 35{,}327$ lesz.

b) Számoljuk ki most annak a valószínűségét, hogy 52 év alatt több mint 1300 levél érkezik. A tényleges valószínűséget kiszámolni nem tudjuk egyszerűen, a határeloszlás tétel segítségével azonban becsülhetjük. Vigyázzunk azonban, mivel a centrális határeloszlás tételének használatához több mint 30 független, azonos eloszlású valószínűségi változóra lesz szükségünk. Mivel azonban hétköznap és hétvégén más a kéretlen levelek eloszlása, így a tételt más valószínűségi változókra kell alkalmaznunk.
Legyenek $X_1, X_2, \ldots X_{52}$ új valószínűségi változók, melyek az egyes heteken belül, tehát 5 munkanap, valamint szombat és vasárnap alatt összegyűlt kéretlen levelek számát jelzik. Ezekre a valószínűségi változókra már alkalmazhatjuk a centrális határeloszlás tételt, mivel függetlenek, azonos eloszlásúak, és elég sokan vannak. Mivel az összegük épp X, így X-et közelíthetjük normális eloszlással:

$$X \approx X_N \sim N(1248, \sqrt{1248})\,.$$

A feladatban kérdezett $X > 1300$ esemény valószínűségét az alábbiak szerint becsülhetjük:

$$P(X > 1300) \approx P(X_N > 1300{,}5) = 1 - \Phi\left(\frac{1300{,}5 - 1248}{\sqrt{1248}}\right) \approx 1 - \Phi(1{,}4861) \approx 1 - 0{,}9314 = 0{,}0686\,.$$

Tehát nagyjából $6{,}8\%$ annak a valószínűsége, hogy az 52 hét alatt kapott kéretlen levelek száma meghaladja az 1300-at. Vegyük észre, hogy mivel X eloszlása diszkrét, és csak egész értékeket vehet fel, a de Moivre–Laplace-tételnél alkalmazott $1/2$-es korrekciót itt is alkalmaznunk kell.

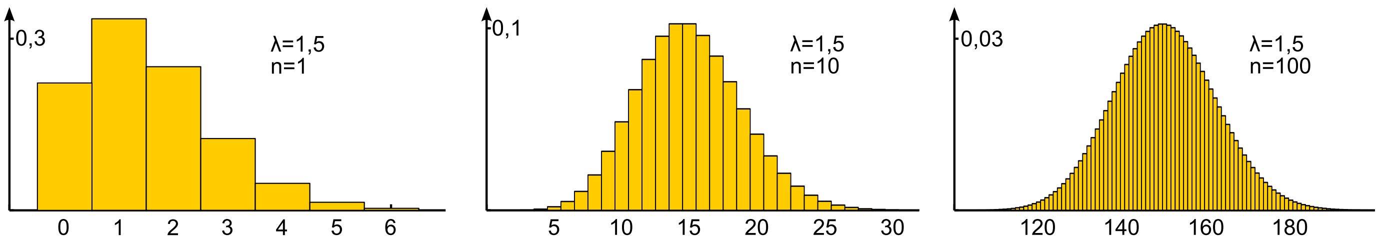
c) Az utolsó feladatban azt vizsgáljuk, hány hét alatt fog több mint 95% valószínűséggel legalább 1000 levél összegyűlni. Az előző részfeladatokhoz képest most nem ismerjük az X pontos eloszlását, viszont n héttel számolva n függvényeként fel tudjuk írni várható értékét és szórásnégyzetét:

$$E(X) = D^2(X) = n \cdot 5 \cdot 4 + n \cdot 2 \cdot 2 = 24 \cdot n\,.$$

Tehát az n hét alatt kapott kéretlen reklámlevelek számának várható értéke $24 \cdot n$, szórása pedig $\sqrt{24 \cdot n}$. Minél több hetet vizsgálunk, annál nagyobb lesz annak a valószínűsége, hogy összejön az 1000 levél, így ha meghatározzuk azt a nem feltétlen egész n értéket, melyre $P(X \geq 1000) = 0{,}95$, akkor felfele kerekítés után megkaphatjuk a választ a feladatra. Az X valószínűségi változó eloszlását a vele megegyező várható értékkel és szórással bíró X_N normális eloszlású valószínűségi változó eloszlásával közelítjük. A vizsgált valószínűséget az alábbiak szerint számolhatjuk ki:

$$0{,}95 = P(X \geq 1000) \approx P(X_N > 999{,}5) = 1 - \Phi\left(\frac{999{,}5 - 24 \cdot n}{\sqrt{24 \cdot n}}\right) = \Phi\left(\frac{24 \cdot n - 999{,}5}{\sqrt{24 \cdot n}}\right).$$

Az utolsó lépésben a $\Phi(-x) = 1 - \Phi(x)$ azonosságot alkalmaztuk, hiszen Φ értékeit akkor tudjuk táblázatból visszakeresni, ha $0{,}5$ és 1 között vannak. Itt a $\Phi(x)$ értéke $0{,}95$ lesz, ebből x értéke $1{,}645$. Az egyenletet így



20.6. ábra. *Az ábra különböző Poisson-eloszlásokat mutat. A bal oldali ábrán a* $\lambda = 1{,}5$*, középen a* $\lambda_{10} = 15$*, jobb oldalon pedig* $\lambda_{100} = 150$*. Mivel a független Poisson-eloszlású valószínűségi változók összege is Poisson-eloszlású, így a bal oldalon* $n = 1$*, középen* $n = 10$*, jobb oldalon* $n = 100$ *darab, azonos,* $\lambda = 1{,}5$ *paraméterű, Poisson-eloszlású valószínűségi változó összegét láthatjuk. Figyeljük meg, hogy jobbra haladva az ábrák egyre jobban hasonlítanak a haranggörbére.*

visszavezethetjük egy $\sqrt{n}$-ben másodfokú egyenletre:

$$\frac{24 \cdot n - 999{,}5}{\sqrt{24 \cdot n}} = 1{,}645, \qquad \text{amiből}$$

$$24 \cdot n - 1{,}645 \cdot \sqrt{24} \cdot \sqrt{n} - 999{,}5 = 0\,.$$

Amennyiben $\sqrt{n}$-et x-szel helyettesítjük, egy másodfokú egyenletet kapunk:

$$24 \cdot x^2 - 1{,}645 \cdot \sqrt{24} \cdot x - 999{,}5 = 0\,.$$

Az egyenlet megoldásai $x_1 = 6{,}6234$ és $x_2 = -6{,}2876$. Mivel $x = \sqrt{n}$, ezért x_2 nem lehet megoldás, hisz negatív. Tehát $n = x_1^2 = 43{,}87$. Ez azt jelenti, hogy annak a valószínűsége, hogy 44 hét alatt legalább 1000 kéretlen reklámlevél érkezik, legalább 95%.

⇐20.5. feladat

Aktivitás: Alkalmas szoftver segítségével számítsa ki az előző feladat kérdéseire a pontos válaszokat.

20.6. feladat. 2030-ban megváltoznak a tenisz szabályai. A bonyolult pontozás helyett egyszerűsítik a rendszert: a játék addig tart, amíg valaki összesen 100 adogatást meg nem nyer. Két rangos játékos, A és B mérkőzése következik. Az előzetes statisztikák alapján tudhatjuk, hogy A klasszisokkal jobb, és minden egyes szervát 60%-os valószínűséggel nyer meg. Kimutatták azt is, hogy a siker minden labdaváltásnál független egymástól, és állandó valószínűséggel következik be.

a) Legyen X_A értéke folyamatos játék esetén a szervák száma addig, amíg az A játékos 100 szervát meg nem nyer. Számoljuk ki X_A várható értékét és szórását!

b) Mi a valószínűsége annak, hogy az esélyek ellenére B mégis nyer?

c) Milyen erős ellenfelet válasszunk A-nak, ha azt szeretnénk, hogy a játékos 75% valószínűséggel nyerjen?

Megoldás:

a) A feladatok megoldásához tegyük fel azt, hogy a játék legalább addig tart, amíg mindkét játékos el nem éri a 100 pontot. Természetesen ez a feltétel a nyertes személyén nem változtat, hiszen aki ezt előbb eléri, az nyer. X_A jelöli a szervák számát addig, amíg az A játékos 100 játékot meg nem nyer. Ennek a valószínűségi változónak az eloszlása negatív binomiális lesz $p = 0{,}6$ és $r = 100$ paraméterekkel. Így a nevezetes eloszlás alapján a várható értéket és a szórást kiszámolhatjuk:

$$E(X_A) = \frac{r}{p} = \frac{100}{0{,}6} \approx 166{,}667, \qquad \text{valamint}$$

$$D(X_A) = \frac{\sqrt{rq}}{p} = \frac{\sqrt{100 \cdot 0{,}4}}{0{,}6} \approx 10{,}541\,.$$

Tehát addig, amíg az A játékosnak össze nem gyűlik a 100 pontja, a szervák számának várható értéke 166,667, szórása pedig 10,541 lesz.

b) Vizsgáljuk meg, mekkora lehet annak a valószínűsége, hogy B mégis nyer? Ehhez az kell, hogy B előbb érje el a 100 nyertes szervát, mint A. Ne számoljuk azt, hogy mikor szedi össze B a 100 pontot, mivel ez nem lesz független X_A-tól (ezt el szeretnénk kerülni), inkább legyen az Y_B valószínűségi változó B pontszáma akkor, amikor A eléri a 100 pontot. Y_B könnyen kifejezhető X_A segítségével:

$$Y_B = X_A - 100,$$

hiszen az A játékos az X_A darab adogatásból 100 pontot nyert, a maradék pontot pedig B. Abban az esetben fog B nyerni, ha eddigre már sikerült legalább 100 pontot gyűjteni, így a keresett valószínűséget az alábbi módon írhatjuk fel:

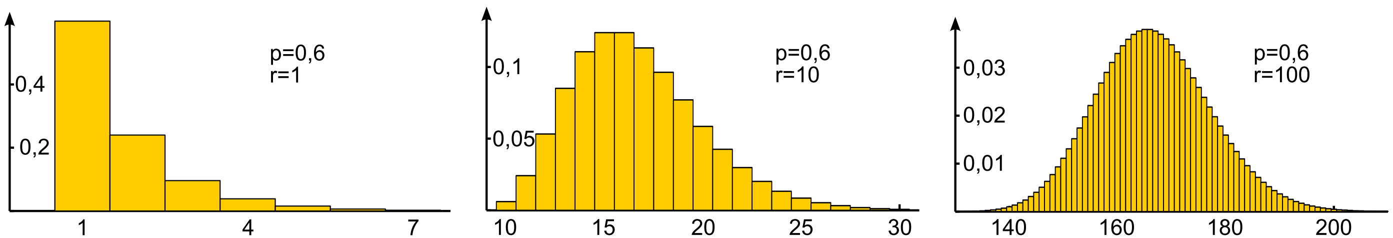
$$P(Y_B \geq 100) = P(X_A - 100 \geq 100) = P(X_A \geq 200)\,.$$

Használhatjuk-e a határeloszlás tételt a valószínűség kiszámításához? Felírható-e X_A mint független, azonos eloszlású valószínűségi változók összege? Igen, felírható, hiszen az $r = 100$ paraméterű negatív binomiális eloszlás épp 100 darab, $p = 0{,}6$ paraméterű geometriai eloszlás összege. Az egyes valószínűségi változók ebben az esetben azt jelölik, hogy hány adogatás kellett az egyes pontok megszerzéséhez. Használhatjuk így a centrális határeloszlás tételt, azaz X_A-t egy normális eloszlású X_N valószínűségi változóval közelíthetjük, melynek várható értéke és szórása megegyezik X várható értékével és szórásával. A vizsgált valószínűséget az alábbi módon írhatjuk fel:

$$P(X_A \geq 200) \approx P(X_N > 199{,}5) = 1 - \Phi\left(\frac{199{,}5 - 166{,}667}{10{,}541}\right) \approx 1 - \Phi(3{,}1148) \approx 0{,}00092\,.$$

Annak a valószínűsége, hogy mégis B nyeri a mérkőzést, elég kicsi, alig 92 század ezrelék. Figyeljük meg azt, hogy ez lényegesen erősebb különbség, mint az eredeti $60 - 40\%$-os arány.

c) Az utolsó részfeladat azt kérdezi, milyen módon hangoljuk A és B adogatás nyerési esélyeit annak érdekében, hogy annak a valószínűsége, hogy A nyeri az egész meccset, 75% legyen. Az előző részfeladatok logikáját követve előbb felírjuk X_A várható értékét és szórását, majd az X_A segítségével megfogalmazott és felírt nyerés valószínűségét közelítjük egy olyan X_N valószínűségi változó segítségével, mely várható



20.7. ábra. *Az ábra különböző negatív binomiális eloszlásokat mutat. Mindegyik eloszlásnál a p paraméter megegyezik, $0{,}6$ valószínűséggel következik be a vizsgált esemény. A bal oldali ábránál $r = 1$, így az eloszlás geometriai, középen $r = 10$, jobb oldalon pedig $r = 100$. Figyeljük meg, hogy jobbra haladva az ábrák egyre jobban hasonlítanak a haranggörbére.*

értékben és szórásban megegyezik X_A-val. Az X_A valószínűségi változó várható értéke és szórása az alábbi módon számítható:

$$E(X_A) = \frac{r}{p} = \frac{100}{p}, \qquad \text{valamint}$$
$$D(X_A) = \frac{\sqrt{rq}}{p} = \frac{\sqrt{100 \cdot (1-p)}}{p}.$$

Vegyük észre azt, hogy most nem ismerjük azt a p-vel jelölt valószínűséget, hogy az egyes adogatásokat A milyen valószínűséggel nyeri meg. Épp ezt az értéket keressük. Ismerjük viszont annak a valószínűségét, hogy A megnyeri a meccset. Ezt a valószínűséget az előző részfeladatban leírt gondolatmenet alapján az alábbiak szerint fogalmazhatjuk meg:

$$P(X_A < 200) = 0{,}75\,.$$

A bal oldalon felírt valószínűséget becsülhetjük egy olyan X_N normális eloszlású valószínűségi változó segítségével, melynek várható értéke és szórása megegyezik X_A-val. Így az alábbiak szerint számolhatunk:

$$P(X_A < 200) \approx P(X_N < 199{,}5) = \Phi\left(\frac{199{,}5 - \frac{100}{p}}{10 \cdot \frac{\sqrt{1-p}}{p}}\right) = \Phi\left(\frac{199{,}5 \cdot p - 100}{10 \cdot \sqrt{1-p}}\right).$$

Mivel a fenti kifejezésben a Φ függvény értéke körülbelül $0{,}75$, táblázat segítségével meghatározhatjuk, hogy Φ ezt az értéket $0{,}6745$-nél veszi fel, így

$$\frac{199{,}5 \cdot p - 100}{10 \cdot \sqrt{1-p}} = 0{,}6745, \qquad \text{azaz} \qquad 199{,}5 \cdot p - 6{,}745 \cdot \sqrt{1-p} - 100 = 0\,.$$

Láthatjuk, hogy $\sqrt{1-p}$-re másodfokú egyenletet kaptunk. Legyen $x = \sqrt{1-p}$, ebben az esetben $p = 1 - x^2$. Helyettesítsük be ezeket az értékeket az egyenletbe:

$$199{,}5 \cdot (1 - x^2) - 6{,}745 \cdot x - 100 = 0, \qquad \text{azaz } 199{,}5 \cdot x^2 + 6{,}745 \cdot x - 99{,}5 = 0\,.$$

A másodfokú egyenlet megoldásai $x_1 = -0{,}7233$ és $x_2 = 0{,}6895$, melyek közül az első kizárható, hiszen $\sqrt{1-p}$ értéke nem lehet negatív. Mivel így $x = 0{,}6895$, ezért $p = 0{,}5246$ lesz. Tehát a tenisz új szabályai szerint, amennyiben az A játékos $52{,}46\%$ valószínűséggel nyer meg egy adogatást, akkor 75% a valószínűsége annak, hogy megnyeri az egész meccset. Vegyük észre, hogy ugyan A alig jobb B-nél, a játék kimenetelének esélyei ezt a különbséget azonban felnagyítják.

⇐20.6. feladat

Aktivitás: Alkalmas szoftver segítségével számolja ki és ábrázolja annak a valószínűségét, hogy az A játékos nyeri meg a meccset, annak a valószínűségnek a függvényében, hogy egy-egy adogatást nyer. Az ábrázolás során vizsgálja meg azt is, mi történik, ha a nyeréshez szükséges nyert labdaváltások számát 100-ról nagyságrendekkel lefele (1 vagy 10), illetve felfele (1000 stb.) változtatjuk

Leckezáró önellenőrző kérdéssor

1. 10000 darab független, azonos eloszlású valószínűségi változó mindegyikének várható értéke 8, szórásuk 10. Összegük standardizáltját olyan normális eloszlással közelítjük, melynek
   - ☐ várható értéke 80000, szórása 1000,
   - ☐ várható értéke 800, szórása 100000,
   - ☐ várható értéke 0, szórása 1,
   - ☐ várható értéke 8, szórása 0,1.
2. 10000 darab független, azonos eloszlású valószínűségi változó mindegyikének várható értéke 8, szórásuk 10. Átlagukat olyan normális eloszlással közelítjük, melynek
   - ☐ várható értéke 80000, szórása 1000,
   - ☐ várható értéke 800, szórása 100000,
   - ☐ várható értéke 0, szórása 1,
   - ☐ várható értéke 8, szórása 0,1.
3. 100 darab független, Poisson-eloszlású valószínűségi változó mindegyikének szórása 2. Adjunk becslést annak a valószínűségére, hogy összegük kisebb, mint 400.

   ☐ $\approx 0{,}5$ ☐ $\leq 0{,}75$ ☐ $\approx 0{,}49$ ☐ ≈ 1

4. Egy dobókockát 64-szer elgurítunk. Kiszámoljuk a dobott számok átlagát. Mi a valószínűsége annak, hogy a kapott átlag 0,1-nál kevesebbel tér el az átlag várható értékétől

   ☐ $\approx 0{,}3657$ ☐ $\leq 0{,}3333$ ☐ $\approx 0{,}9490$ ☐ $\geq 0{,}9490$

35. LECKE

Modulzáró feladatok 6.

21. Modulzáró feladatok

Start.

A modulzáró megoldására 1 órája van. Legalább 4 tizedesjegy pontossággal adja meg az eredményeket.

1. 100 darab független, azonos eloszlású folytonos valószínűségi változó mindegyikének várható értéke 8, szórásuk 2. A határeloszlás tételek segítségével adjunk becslést annak a valószínűségére, hogy összegük 840-nél nagyobb!
2. Egy szabályos, hatoldalú dobókockát százszor elgurítunk. A határeloszlás tételek segítségével adjunk becslést annak a valószínűségére, hogy a dobott számok átlaga annak várható értékét 1%-os pontosságon belül megközelíti!
3. Egy szabályos pénzérmét 400-szor feldobunk. Adjunk becslést a határeloszlás tételek segítségével arra, mekkora lehet annak a valószínűsége, hogy legalább 195 fejet dobunk!
4. Egy légitársaság TU-154-es gépére 167 utas fér fel. A tapasztalat szerint feltehetjük, hogy minden utas, aki foglalást adott le, a többiektől függetlenül 90%-os valószínűséggel jelenik meg beszálláskor. Hány jegyet lehet eladni, ha azt szeretnénk, hogy a beszálláskor megjelent utasok 99% valószínűséggel felférjenek a gépre?
5. A kontinentális viharok során két villámcsapás közötti eltelt idő exponenciális eloszlást követ, egy perc várható értékkel. Mi a valószínűsége annak, hogy legalább két óra eltelik a századik villámcsapásig?
6. 64 független, azonos eloszlású valószínűségi változó mindegyikének szórása 4. Adjunk becslést a határeloszlás tételek segítségével annak a valószínűségére, hogy átlaguk 0,5 egységnél közelebb kerül az ismeretlen várható értékhez!

Stop. Pontok:

VII. MODUL

Matematikai statisztika

A matematikai statisztika módszerei széles körben elterjedtek, a mérnöki tudományoktól a természettudományokon át a gazdaság- és társadalomtudományokig. A hétköznapokban is gyakran találkozunk vele, elegendő a közvéleménykutatásokra gondolni. A statisztikai szoftverek egyre szélesebb elterjedése pedig azok számára is lehetővé teszi a statisztika módszereinek használatát, akik egyébként nem érdeklődnek a matematikai háttér iránt. A következő fejezetekben áttekintjük a matematikai statisztika alapfogalmait, illetve megismerkedünk a statisztikai próbák menetével, legfőbb típusaival.

36. LECKE

A matematikai statisztikában használatos eloszlások

22. A matematikai statisztikában használatos eloszlások

A matematikai statisztikában a leggyakrabban használt eloszlás a már ismert standard normális eloszlás. Az itt szereplő többi nevezetes eloszlás független, standard normális eloszlású valószínűségi változókból származtatható. Emlékezzünk rá, hogy a standard normális eloszlással kapcsolatos feladatokat az eloszlásfüggvény értékeit tartalmazó táblázat segítségével oldottuk meg. A belőle származtatott, még bonyolultabb eloszlások esetén szintén a táblázatokba foglalt értékek alapján dolgozhatunk.
A standard normális eloszlásból származtatott eloszlások jellemzéséhez szükségünk lesz az úgynevezett Euler-féle gammafüggvényre.

22.1. definíció: A $\Gamma(x)$ függvényt Euler-féle gammafüggvénynek nevezzük, és a következőképpen definiáljuk:

$$\Gamma(x) = \int_0^\infty e^{-t} \cdot t^{x-1} \mathrm{d}t \,, \qquad x > 0 \,.$$

A gammafüggvény rendelkezik az alábbi tulajdonságokkal:

1. $\Gamma(x+1) = x \cdot \Gamma(x)$, $x > 0$.
2. Ha $n \in \mathbb{N}$, akkor $\Gamma(n+1) = n!\,$.
3. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

22.1. A χ^2 eloszlás

22.2. definíció: Legyenek $X_1, X_2, \ldots, X_n$ független, standard normális eloszlású valószínűségi változók. Ekkor a belőlük képzett

$$\chi_n^2 = X_1^2 + X_2^2 + \ldots + X_n^2$$

valószínűségi változó eloszlását n szabadsági fokú χ^2 (khí négyzet)-eloszlásnak nevezzük.

Az n szabadsági fokú χ^2-eloszlás sűrűségfüggvénye:

$$f_n(x) = \begin{cases} \dfrac{x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \cdot \Gamma\left(\frac{n}{2}\right)} & \text{ha } x > 0, \\ 0 & \text{ha } x \leq 0. \end{cases}$$

Az n szabadsági fokú χ^2-eloszlású valószínűségi változó várható értéke és szórásnégyzete:

$$E(\chi_n^2) = n \qquad \text{és} \qquad D^2(\chi_n^2) = 2n\,.$$

Az eloszlás első látásra bonyolultnak tűnik, használata azonban nagyon egyszerű. A χ^2-eloszlás esetében általában nem az eloszlásfüggvény adott helyen felvett értékére vagyunk kíváncsiak, hanem arra, hogy egy adott értéket hol vesz fel az eloszlásfüggvény (azaz az eloszlásfüggvény inverzének értéke). (Emlékezzünk a valószínűségszámítás tanulmányainkból, hogy az eloszlásfüggvény adott helyen felvett értéke megegyezett a sűrűségfüggvény adott helytől balra eső görbe alatti területével.) A standard normális eloszláshoz hasonlóan a χ^2-eloszlás értékeit is táblázatból olvashatjuk ki. A χ^2-eloszlás inverzének gyakrabban használt értékei a 4. táblázatban megtalálhatóak. Nézzük, hogyan használjuk a táblázatot!

22.1. feladat: Adott az n szabadsági fokú χ^2-eloszlású valószínűségi változó. Határozzuk meg, hogy az eloszlásfüggvénye hol veszi fel a p értéket, ha

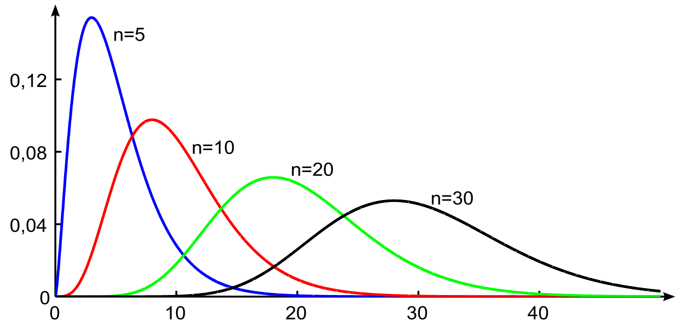
a) $n = 5$, p=0,95,

b) $n = 20$, p=0,98!

Megoldás: A 4. táblázat első sorában az eloszlásfüggvény értékei, első oszlopában pedig a szabadsági fok található. Mindösszesen annyi a dolgunk, hogy az első oszlopból kikeressük a megadott szabadsági fokot, az első sorból pedig a megadott valószínűséget. A sor és oszlop metszetéből pedig leolvassuk a keresett értéket (amelyet igazodva a későbbi feladatainkhoz, jelöljünk χ_t^2-vel).

a) $\chi_t^2 = 11{,}07$,

b) $\chi_t^2 = 35{,}02$.



22.1. ábra. *A χ^2-eloszlás sűrűségfüggvénye különféle szabadsági fokok (n) esetén.*

22.2. A Student-eloszlás

A Student-eloszlás (vagy másnéven t-eloszlás) adja a matematikai statisztikában használatos t-próba alapját.

22.3. definíció: Legyenek Y és $X_1, X_2, \ldots, X_n$ független, standard normális eloszlású valószínűségi változók. Ekkor a belőlük képzett

$$t_n = \frac{Y}{\sqrt{\dfrac{X_1^2 + X_2^2 + \ldots + X_n^2}{n}}}$$

valószínűségi változó eloszlását n szabadsági fokú Student-eloszlásnak (t-eloszlásnak) nevezzük.

Az n szabadsági fokú t-eloszlás sűrűségfüggvénye:

$$f(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}.$$

Az n szabadsági fokú t-eloszlású valószínűségi változó várható értéke:

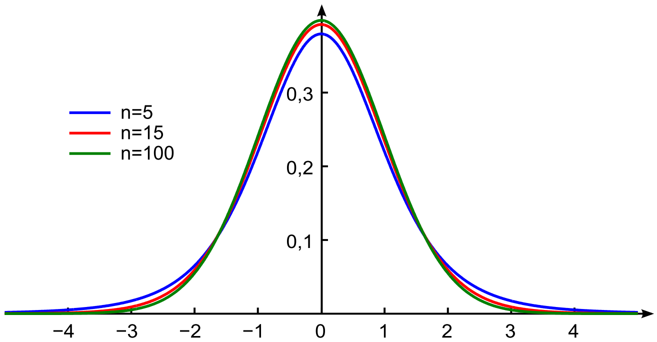
$$E(t_n) = \begin{cases} 0 & \text{ha } n \geq 2, \\ \text{nem létezik} & \text{ha } n = 1. \end{cases}$$

Szórásnégyzete:

$$D^2(t_n) = \begin{cases} \dfrac{n}{n-2} & \text{ha } n \geq 3, \\ \text{nem létezik} & \text{ha } n = 1, 2. \end{cases}$$

A t-eloszlás esetében (a χ^2-eloszláshoz hasonlóan) elsősorban az érdekel bennünket, hogy egy adott értéket hol vesz fel az eloszlásfüggvény (azaz megint az eloszlásfüggvény inverzének értékére). A t-eloszlás értékeit is

táblázatból olvashatjuk ki, leggyakrabban használt értékei a 3. táblázatban találhatóak. A táblázat használata megegyezik a χ^2 eloszlásnál látottal.
Nézzük meg a Student-eloszlás sűrűségfüggvényét a 22.2. ábrán! A szabadsági fok növelésével a görbe egyre inkább hasonlít a standard normális eloszlás haranggörbéjéhez (a Student-eloszlás határeloszlása a standard normális). Ezen hasonlóság alapján mi a Student-eloszlás táblázatot $n = 30$-ig fogjuk használni, $n > 30$ esetén standard normális eloszlással közelítjük a Student-eloszlást.



22.2. ábra. *A Student-eloszlás sűrűségfüggvénye különféle szabadsági fokok (n) esetén.*

22.2. feladat: Adott az n szabadsági fokú t-eloszlású valószínűségi változó. Határozzuk meg, hogy az eloszlásfüggvénye hol veszi fel a p értéket, ha

a) $n = 8$, p=0,95,

b) $n = 20$, p=0,99!

Megoldás: A 3. táblázat első sorában az eloszlásfüggvény értékei, első oszlopában pedig a szabadsági fok található. Újra csak annyi a dolgunk, hogy az első oszlopból kikeressük a megadott szabadsági fokot, az első sorból pedig a megadott valószínűséget. (A keresett értéket jelöljünk t_t-vel).

a) $t_t = 1{,}86$,

b) $t_t = 2{,}53$.

Aktivitás: Hasonlítsa össze a t-eloszlás táblázat értékeit $n > 30$ esetén a standard normális eloszlásfüggvény inverzének értékeivel adott p esetén!

22.3. F-eloszlás

22.4. definíció: Legyenek $X_1, X_2, \ldots, X_m$ és $Y_1, Y_2, \ldots, Y_n$ független, standard normális eloszlású valószínűségi változók. Ekkor a belőlük képzett

$$F_{m,n} = \frac{n}{m} \cdot \frac{X_1^2 + X_2^2 + \ldots + X_m^2}{Y_1^2 + Y_2^2 + \ldots + Y_n^2}$$

valószínűségi változó eloszlását (m,n) szabadsági fokú F-eloszlásnak nevezzük.

Az (m,n) szabadsági fokú F-eloszlás sűrűségfüggvénye:

$$f_{m,n}(x) = \begin{cases} \left(\frac{m}{n}\right)^{\frac{m}{2}} \cdot \dfrac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \cdot \Gamma\left(\frac{n}{2}\right)} \cdot \dfrac{x^{\frac{m}{2}-1}}{\left(1 + \frac{m}{n}x\right)^{\frac{m+n}{2}}} & \text{ha } x > 0, \\ 0 & \text{ha } x \le 0. \end{cases}$$

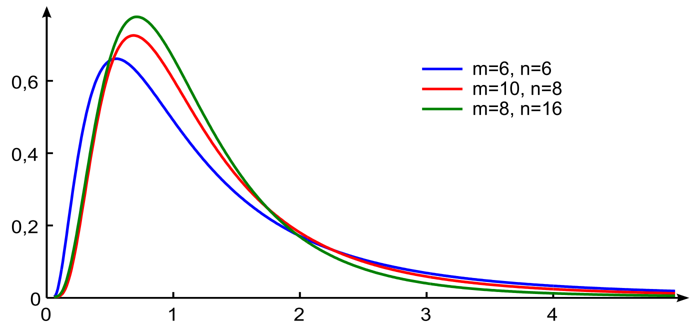
Az (m,n) szabadsági fokú F-eloszlású valószínűségi változó várható értéke:

$$E(F_{m,n}) = \begin{cases} \dfrac{n}{n-2} & \text{ha } n \ge 3, \\ \text{nem létezik} & \text{ha } n = 1, 2. \end{cases}$$

Szórásnégyzete:

$$D^2(F_{m,n}) = \begin{cases} 2\left(\dfrac{n}{n-2}\right)^2 \dfrac{(m+n-2)}{m \cdot (n-4)} & \text{ha } n \ge 5, \\ \text{nem létezik} & \text{ha } n = 1, 2, 3, 4. \end{cases}$$

Az F-eloszlás esetében is az eloszlásfüggvény inverzének értékeit táblázatból olvashatjuk ki. Az 5. táblázat használata során az (m,n) szabadsági fokokat a táblázat első sorában, illetve első oszlopában találjuk meg. Az eloszlásfüggvény értéke ebben a táblázatban $p = 0{,}95$. (Ennek kitüntett szerepéről később lesz szó.)



22.3. ábra. *Az F-eloszlás sűrűségfüggvénye különféle szabadsági fokok $(m,\ n)$ esetén.*

22.3. feladat: Adott az $(m = 10, n = 8)$ szabadsági fokú F eloszlású valószínűségi változó. Határozzuk meg, hogy az eloszlásfüggvénye hol veszi fel a $p = 0{,}95$ értéket!

Megoldás: Az 5. táblázat első sorában m, első oszlopában pedig n értékei találhatók. A keresett értéket F_t-vel jelölve: $F_t = 3{,}35$.

Önellenőrzés

Önellenőrző kérdések

1. Hol veszi fel a 12 szabadsági fokú χ^2-eloszlású valószínűségi változó eloszlásfüggvénye a 0,99 értéket?
   - ☐ 2,68 ☐ 24,72 ☐ 26,22 ☐ 24,05
2. Hol veszi fel a 11 szabadsági fokú t-eloszlású valószínűségi változó eloszlásfüggvénye a 0,98 értéket?

☐ 2,76 ☐ 2,76 ☐ 2,33 ☐ 2,36

3. Hol veszi fel a (8,7) szabadsági fokú F-eloszlású valószínűségi változó eloszlásfüggvénye a 0,95 értéket?

☐ 3,44 ☐ 3,5 ☐ 3,73 ☐ 2,71

37. LECKE

A matematikai statisztika alapfogalmai

23. A matematikai statisztika alapfogalmai

Tekintsünk egy (vagy több) valószínűségi változót, amely egy sokaság minden elemén felvesz egy értéket. A valószínűségi változó eloszlása általában nem ismert, ennek hiányában sem a várható értékét, sem a szórását, sem adott intervallumba esésének valószínűségét nem tudjuk megmondani. Például tekintsük a magyarországi felsőoktatási intézmények hallgatóit, mint sokaságot, a valószínűségi változó pedig legyen az IQ-juk. Az alábbi kérdésekre szeretnénk választ kapni:

– Mennyi a magyarországi felsőoktatási intézmények hallgatóinak IQ-jának várható értéke?
– Mennyi a magyarországi felsőoktatási intézmények hallgatóinak IQ-jának szórása?
– A magyarországi felsőoktatási intézmények hallgatóinak IQ-jának eloszlása tekinthető-e normálisnak?
– Igaz-e, hogy a magyarországi felsőoktatási intézmények hallgatóinak IQ-jának várható értéke magasabb az európai átlagnál?
– Van-e kapcsolat az iskolai végzettség és az IQ között?
– Stb.

Hasonló kérdéseket szeretnénk a sokaságból véletlenszerűen kiválasztott elemek alapján megválaszolni.
A gyakorlati feladatok esetében az ismeretlen valószínűségre legfeljebb előfeltételezéseink vannak. A matematikai statisztika feladata, hogy a véletlen tömegjelenséghez tartozó megfigyeléssorozat alapján a jelenséghez tartozó valószínűséget megfelelő pontossággal tudjuk közelíteni. A matematikai statisztika fogalomköre és módszertana a valószínűség-számításon alapul, szemben a (főként társadalomtudományokban használt) leíró statisztikával, ahol a következtetéseket heurisztikus módszereket is alkalmazva hozzák meg.
A statisztika tárgya tehát véletlen tömegjelenségek viselkedésének leírása, véges sok kísérlet eredményének felhasználásával.

23.1. Statisztikai minta, statisztikai függvények

A mintavételnél fontos, hogy a sokaság egyes elemeit azonos valószínűséggel válasszuk ki, hiszen ekkor fogja a minta a legjobban visszaadni a sokaság jellemzőit.

23.1. definíció: A megfigyelt X valószínűségi változóval azonos eloszlású, független $X_1, X_2, \ldots, X_n$ valószínűségi változók (mintavételi változók vagy mintaelemek) összességét statisztikai mintának nevezzük.

Az X valószínűségi változó megfigyelésére elvégzett n számú kísérlet során $X_1, X_2, \ldots, X_n$ rendre az $x_1, x_2, \ldots, x_n$ konkrét értékeket veszi fel. (Az $x_1, x_2, \ldots, x_n$-et a statisztikai minta egy realizációjának is nevezzük.)
A bevezetőben említett példára tekintsünk vissza. Válasszunk ki véletlenszerűen $n = 100$ felsőoktatási hallgatót és mérjük meg mindegyikük IQ-ját. Ezzel kapunk egy $n = 100$ elemű statisztikai mintát. Az azonos eloszlás ebben az esetben azt jelenti, hogy annak valószínűsége, hogy a kiválasztott hallgató IQ-ja pl. magasabb 130-nál, minden egyes hallgató esetében ugyanannyi. A függetlenség pedig azt, hogy az egyes hallgatók mért IQ-ja nem befolyásolja annak valószínűségét, hogy a többiek esetében az intelligenciahányados pl. alacsonyabb vagy magasabb 130-nál.
Bizonyos esetekben megkönnyíti a dolgunkat, ha a mintavételi változókat nagyság szerint növekvő sorrendbe rendezzük. Az így kapott mintát rendezett mintának nevezzük.

23.2. definíció: Ha az $X_1, X_2, \ldots, X_n$ valószínűségi változók nagyság szerint növekvő értékei közül az i-ediket jelöljük X_i^*-vel ($i = 1, 2, \ldots, n$), akkor az $X_1^*, X_2^*, \ldots, X_n^*$ statisztikai mintát rendezett mintának nevezzük. Ekkor $X_1^* \leq X_2^* \leq \ldots \leq X_n^*$.

Nyilvánvaló, hogy ha egy valószínűségi változó megfigyelésére több méréssorozatot hajtunk végre, akkor más és más adatokat kapunk. A mintavétel után a mintaelemeket felhasználva következtethetünk a megfigyelt valószínűségi változó főbb paramétereire, így a várható értékre és a szórásra is. A következtetéseinket a statisztikai mintából számolt értékek segítségével hozzuk meg oly módon, hogy a mintaelemeket bizonyos formulákba helyettesítjük. Gyakorlatilag egy többváltozós függvény helyettesítési értékeit kapjuk. Ez történik például akkor, mikor az adott minta elemeinek átlagát számítjuk ki. A mintavételi változókon tehát értelmezünk egy függvényt (statisztikát), amely maga is valószínűségi változó, értéke minden minta esetében különböző (lehet).

23.3. definíció: Az $X_1, X_2, \ldots, X_n$ valószínűségi változókat a valós számok halmazára képező

$$\hat{\alpha}_n : \mathbb{R}^n \to \mathbb{R}$$

függvényt statisztikai függvénynek (statisztikának) nevezzük.

Felvetődik a kérdés, milyen alkalmasan választott statisztikai függvényekkel tudjuk a legjobb következtetéseket levonni a sokaság paramétereire vonatkozóan. Hogyan becsüljük mondjuk a várható értéket? A mintaelemek átlagával? A minta legkisebb és legnagyobb elemének átlagával? A rendezett minta középső elemével? Hogyan becsüljük a szórást?

Az alábbiakban áttekintjük a leggyakrabban előforduló statisztikákat!

23.4. definíció: Az $X_1, X_2, \ldots, X_n$ mintaelemek mintaátlaga (empirikus közepe) az

$$\overline{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

valószínűségi változó.

fenti statisztika adott mintán kiszámított

$$\hat{m} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

átlaggal becsülhetjük a megfigyelt X valószínűségi változó $E(X) = m$ várható értékét. A következő statisztikai függvények esetében az egyszerűség kedvéért nem különítjük el a konkrét minta alapján számított értéket.

23.5. definíció: Az $X_1, X_2, \ldots, X_n$ mintaelemek tapasztalati (vagy empirikus) szórásnégyzetén a mintaelemek mintaátlagtól való eltéréseinek négyzetes közepét, azaz a

$$\hat{\sigma}_n^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \hat{m}_n)^2 = \frac{(X_1 - \hat{m}_n)^2 + (X_2 - \hat{m}_n)^2 + \ldots + (X_n - \hat{m}_n)^2}{n}$$

értéket értjük.

A tapasztalati szórásnégyzet kiszámítására (az elméleti szórásnégyzetnél tanultakkal analóg módon) az alábbi formulát is használhatjuk:

$$\hat{\sigma}_n^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i^2 - \frac{1}{n} \cdot \left(\sum_{i=1}^{n} X_i \right)^2$$

(Ezzel a kevesebb kerekítés miatt jobb a számítási pontosság.)

Vegyük észre, hogy tapasztalai szórás formulájában a minta átlagos eltérését a várható érték helyett a saját mintaátlagától számítjuk, amihez 'közelebb' van. Ezért szükséges az alábbi korrekció.

23.6. definíció: Az $X_1, X_2, \ldots, X_n$ mintaelemek korrigált tapasztalati szórásnégyzetén az

$$\hat{s}_n^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \hat{m}_n)^2 = \frac{(X_1 - \hat{m}_n)^2 + (X_2 - \hat{m}_n)^2 + \ldots + (X_n - \hat{m}_n)^2}{n-1}$$

értéket értjük.

A formulákból látható, hogy a tapasztalati és a korrigált tapasztalati szórások közötti összefüggés:

$$\hat{s}_n = \sqrt{\frac{n}{n-1}} \cdot \hat{\sigma}_n.$$

A medián a rendezett minta középső értéke. Páratlan elemszámú minta esetében nem okoz gondot a kiválasztása, páros elemszám esetén azonban pontosabb meghatározásra van szükség.

23.7. definíció: Az $X_1^*, X_2^*, \ldots, X_n^*$ rendezett minta mediánja

$$\begin{cases} \dfrac{X_k^* + X_{k+1}^*}{2} & \text{ha } n = 2 \cdot k, \\ X_{k+1}^* & \text{ha } n = 2 \cdot k + 1. \end{cases}$$

23.8. definíció: A módusz a statisztikai minta leggyakrabban előforduló mintaeleme.

Nem létezik módusz, ha nincs leggyakrabban előforduló mintaelem (minden mintaelem ugyanannyiszor fordul elő a mintában). Ugyanakkor előfordulhat, hogy több mintaelem előfordulási száma is azonos a mintában. Ekkor a mintának több módusza is van: a legnagyobb előfordulási számhoz tartozó mintaelem. A módusz a mintából könnyen számítható, a szélsőséges mintaelemek hatását csökkenti, ugyanakkor jelentősége az alkalmazásokban kicsi.
A statisztikai minta legnagyobb és legkisebb elemének különbsége adja a minta terjedelmét.

23.9. definíció: Az $X_1^*, X_2^*, \ldots, X_n^*$ rendezett minta terjedelmén az $X_n^* - X_1^*$ értéket értjük.

Érdemes még megemlíteni a legnagyobb és legkisebb mintaelem számtani közepét.

23.10. definíció: Az $X_1^*, X_2^*, \dots, X_n^*$ rendezett minta középpontján az $\dfrac{X_n^* + X_1^*}{2}$ értéket értjük.

Ahhoz, hogy a fenti statisztikai függvények közül ki tudjuk választani a megfelelőt, el kell döntenünk, hogy mit várunk a becsléstől. Általában azt várjuk, hogy a legnagyobb valószínűséggel a legközelebb legyen a becsülni kívánt paraméterhez. Ehhez az kell, hogy a becslés várható értéke megegyezzen a becsült paraméter értékével (ekkor torzítatlan becslésről beszélünk), illetve a szórása minimális legyen (ekkor a becslés hatásos).
Belátható, hogy a mintaátlag a várható értéknek hatásos és torzítatlan becslése. Ugyanakkor a tapasztalati szórás nem, a korrigált tapasztalati szórás viszont torzítatlan becslése a szórásnak.

23.2. Hisztogramok

23.11. definíció: Az $X_1,X_2,\dots,X_n$ minta tapasztalati eloszlásfüggvényének nevezzük az

$$F_n(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} k_{X_i<x}$$

függvényt, ahol $k_{X_i<x}$ az x-nél kisebb mintaelemek számát jelenti.

A gyakorlatban ez azt jelenti, hogy a tapasztalati eloszlásfüggvény értéke bármely $x \in \mathbb{R}$ esetén megegyezik az x-nél kisebb mintaelemek számának relatív gyakoriságával. A tapasztalati eloszlásfüggvényolyan lépcsős függvény, melynek ugráshelyei a mintaelemektől függenek, az ugrások nagysága pedig különböző mintaelemek esetén $\frac{1}{n}$.

23.1. tétel: (Glivenko tétele)
Az $F_n(x)$ tapasztalati eloszlásfüggvény az egész számegyenesen 1 valószínűséggel, egyenletesen konvergál az $F(x)$ elméleti eloszlásfüggvényhez, azaz

$$P\left(\lim_{n\to\infty} (\sup|F_n(x) - F(x)|) = 0\right) = 1.$$

Ez azt jelenti, hogy ha elegendően nagy számú mintát veszünk, akkor tetszőleges pontossággal tudjuk közelíteni az elméleti eloszlásfüggvényt.

A statisztikai minta szemléltetésének leggyakoribb módja a különféle hisztogramokkal való ábrázolás. Első lépésként a mintaelemeket valamilyen módon csoportosítani kell. Ehhez az $X_1^*,X_2^*,\dots,X_n^*$ rendezett mintára lesz szükségünk. Foglaljuk a mintaelemeket egy $[a,b)$ intervallumba oly módon, hogy $a \leq X_1^*$ illetve $X_n^* < b$ teljesüljenek. Ezután osszuk fel az intervallumot r részintervallumra (osztályra). Az osztópontok elhelyezése lehet egyenközű (ekkor minden részintervallum azonos hosszúságú), de akár tetszőleges hosszúságú részintervallumokat is megadhatunk. Az osztópontokat jelölje rendre $x_0,x_1,\dots,x_r$, melyekre teljesüljön:

$$a = x_0 < x_1 < x_2 < \dots < x_r = b.$$

Jelölje k_i az i-edik ($[x_{i-1},x_i)$) ($i = 1,2,\dots,r$) intervallumba eső mintaelemek számát.

Gyakorisági hisztogram esetén az $[x_{i-1},x_i)$ (i-edik) részintervallumhoz tartozó téglalap magasságát az intervallumba eső mintaelemek számának és az intervallum hosszának hányadosa adja: $\dfrac{k_i}{x_i - x_{i-1}}$.
Könnyen belátható, hogy a kapott téglalapok területének összege megegyezik a minta elemszámával, azaz n-nel.

Sűrűséghisztogramot úgy kaphatunk, ha a gyakorisági hisztogramban az egyes téglalapok magasságát a minta elemszámával, azaz n-nel osztjuk. Az $[x_{i-1},x_i)$ intervallumhoz tartozó téglalap magassága tehát: $\dfrac{k_i}{(x_i - x_{i-1}) \cdot n}$
A téglalapok összterülete ekkor 1.
Vegyük észre, hogy a gyakorisági és a sűrűséghisztogram csak a téglalapok magasságában tér el!

Adott minta esetén a gyakorisági- és a sűrűséghisztogram az intervallum beosztásától, a részintervallumok számától és hosszától nagymértékben függ. Ezek megválasztása természetesen az adott feladattól függ, de általánosságban azt mondhatjuk, hogy célszerű egyenközű (vagy ahhoz közeli) beosztást választani, a részintervallumok számát pedig $\log_2(n)$ (kis elemszám esetén $\sqrt{n}$) közelinek választani.

Sokszor a mintát úgy kapjuk meg, hogy a mintaelemeket már osztályokba sorolták. Amellett, hogy a hisztogramok létrehozására különösen ügyelni kell, felvetődik egy kérdés. Hogyan számoljuk ebben az esetben a korábban felsorolt statisztikákat (mintaátlag stb.), ahol a konkrét mintaelemekre van szükség a számításhoz? Az egyik legegyszerűbb és leggyakoribb megoldás, ha az egyes mintaelemeket annak az osztálynak az osztályközepével (az adott intervallum felezőpontjával) helyettesítjük, amelybe tartoznak. (Ennek a nyilvánvaló pontatlanságnak a csökkentésére vannak kísérletek, ezzel mi nem foglalkozunk.)

A következő feladatok a statisztikai függvények kiszámítását mutatják adott mintából. Az első feladatban a mintaelemeket felsorolva, a másodikban osztályokba sorolva adtuk meg. Nézzük, ez a különbség mit jelent a számításoknál!

23.1. feladat: Adott az alábbi 8 elemű statisztikai minta:

102	108	95	98	104	90	94	104

Adjuk meg a minta

a) átlagát,
b) tapasztalati szórását,
c) korrigált tapasztalati szórását,
d) terjedelmét,
e) mediánját,
f) móduszát,
g) középpontját!

Megoldás: A feladat megoldásához a korábbi formulákba kell behelyettesíteni a minta elemeit. (X_i a minta i-edik elemét jelöli.)

a)

$$\hat{m}_8 = \frac{1}{8} \cdot \sum_{i=1}^{8} X_i = \frac{X_1 + X_2 + \ldots + X_8}{8} = \frac{102 + 108 + \ldots + 104}{8} = \frac{795}{8} = 99{,}375.$$

b)

$$\hat{\sigma}_8 = \sqrt{\frac{1}{8} \cdot \sum_{i=1}^{8} (X_i - \hat{m}_8)^2} = \sqrt{\frac{(X_1 - \hat{m}_8)^2 + (X_2 - \hat{m}_8)^2 + \ldots + (X_8 - \hat{m}_8)^2}{8}} =$$

$$= \sqrt{\frac{(102 - 99{,}375)^2 + (108 - 99{,}375)^2 + \ldots + (104 - 99{,}375)^2}{8}} = \sqrt{\frac{261{,}875}{8}} = \sqrt{32{,}7344} = 5{,}7214.$$

c)

$$\hat{s}_8 = \sqrt{\frac{1}{8-1} \cdot \sum_{i=1}^{8} (X_i - \hat{m}_8)^2} = \sqrt{\frac{(X_1 - \hat{m}_8)^2 + (X_2 - \hat{m}_8)^2 + \ldots + (X_8 - \hat{m}_8)^2}{8-1}} =$$

$$= \sqrt{\frac{(102 - 99{,}375)^2 + (108 - 99{,}375)^2 + \ldots + (104 - 99{,}375)^2}{8-1}} = \sqrt{\frac{261{,}875}{7}} = \sqrt{37{,}4107} = 6{,}1164.$$

A tapasztalati szórás ismert értékét felhasználva rövidebben jutunk eredményre :

$$\hat{s}_n = \sqrt{\frac{n}{n-1}} \cdot \hat{\sigma}_n = \sqrt{\frac{8}{7}} \cdot 5{,}7214 = 6{,}1164.$$

A továbbiakban szükségünk lesz a rendezett mintára:

90	94	95	98	102	104	104	108

d) Terjedelem: $X_8^* - X_1^* = 108 - 90 = 18$.

e) A minta elemszáma páros (8), ezért a medián: $\dfrac{X_4^* + X_5^*}{2} = \dfrac{98 + 102}{2} = 100$.

f) A leggyakoribb elem a mintában (kétszer fordul elő) a módusz: 104.

g) A minta középpontja: $\dfrac{X_1^* + X_8^*}{2} = \dfrac{90 + 108}{2} = 99$.

23.2. feladat: Egy kis faluban a felnőtt lakosság életkor szerinti eloszlását vizsgálták. Az alábbi eredményeket kapták:

Életkor (év)	14-20	20-30	30-40	40-50	50-60	60-90
Emberek száma	12	14	26	35	23	6

Adjuk meg a

a) mintaátlagot,

b) korrigált tapasztalati szórást,

c) sűrűséghisztogramot,

d) tapasztalati eloszlásfüggvényt!

Megoldás: Ebben a feladatban az $n = 116$ mintaelemet a mintavételkor $r = 6$ osztályba sorolták. Becsüljünk minden mintaelemet a neki megfelelő osztályközéppel (pl. a tizenkettő darab, 14 és 20 év közötti embert egyaránt 17 évesnek becsüljük). Ezzel sok azonos mintaelemet kapunk, amelyeknek a gyakoriságára (hányszor fordulnak elő a mintában), illetve relatív gyakoriságára (a mintában való előfordulási arányukra) lesz szükségünk a megoldáshoz. Az egyszerűbb és áttekinthetőbb számoláshoz az adatainkat táblázatba foglaljuk.

Osztály: $[x_{i-1}, x_i)$	Osztályközép: $\frac{x_i + x_{i-1}}{2}$	Gyakoriság: k_i	Relatív gyakoriság: $\frac{k_i}{n}$
14-20	17	12	$12/116$
20-30	25	14	$14/116$
30-40	35	26	$26/116$
40-50	45	35	$35/116$
50-60	55	23	$23/116$
60-90	75	6	$6/116$
		$\sum_{i=1}^{6} k_i = 116$	$\sum_{i=1}^{r} \frac{k_i}{n} = 1$

A táblázatból a számításainkhoz szükséges összes érték kiolvasható. A megfelelő formulákba helyettesítünk:

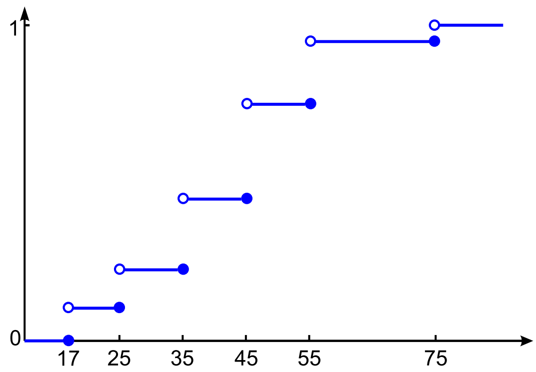
a) $$\hat{m}_{116} = \frac{1}{n} \sum_{i=1}^{6} \frac{x_i + x_{i-1}}{2} \cdot k_i = \frac{17 \cdot 12 + 25 \cdot 14 + \ldots + 75 \cdot 6}{116} = 40{,}9828.$$

b) $$\hat{s}_{116} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{6} \left(\frac{x_i + x_{i-1}}{2} - \hat{m}_{116} \right)^2 \cdot k_i} =$$

$$= \sqrt{\frac{12 \cdot (17 - 40{,}9828)^2 + 14 \cdot (25 - 40{,}9828)^2 + \ldots + (80 - 40{,}9828)^2}{116 - 1}} = 14{,}2755$$

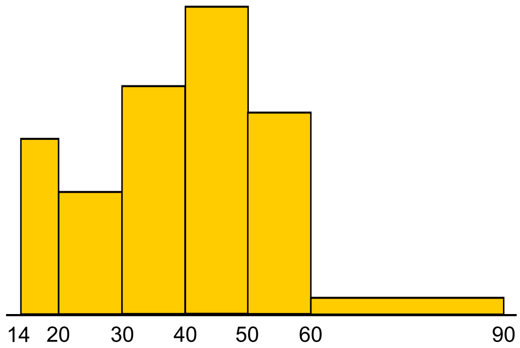
c) A tapasztalati eloszlásfüggvény:

$$F_{116}(x) = \begin{cases} 0 & \text{ha } x \leq 17, \\ \dfrac{12}{116} & \text{ha } 17 < x \leq 25, \\ \dfrac{26}{116} & \text{ha } 25 < x \leq 35, \\ \vdots & \\ 1 & \text{ha } x > 75, \end{cases}.$$



23.1. ábra. *A 23.2. feladathoz tartozó tapasztalati eloszlásfüggvény.*

d) A sűrűséghisztogram: Intervallum-felosztásnak elfogadjuk a feladatban megadott osztályokat. Annyi dolgunk van, hogy a táblázatban megadott intervallumokat a hozzájuk tartozó relatív gyakoriságokkal ábrázoljuk. A táblázatból kiolvasható a részintervallumok hossza $(x_i - x_{i-1})$, az intervallumba eső mintaelemek száma (k_i). A téglalapok alapja a megfelelő részintervallum, területe pedig az adott intervallumhoz tartozó relatív gyakoriság.



23.2. ábra. *A 23.2. feladathoz tartozó sűrűséghisztogram.*

23.3. Konfidencia (megbízhatósági) intervallumok

A matematikai statisztika egyik feladata, hogy a statisztikai minta elemeinek segítségével becslést adjon a megfigyelt valószínűségeloszlás ismeretlen paramétereinek értékére. Alapvetően két típusba sorolhatjuk a becsléseket.

– **Pontbecslések**
  Ebben az esetben a valószínűségi változó értékét a mintából számított konkrét értékkel becsüljük. Ide sorolható az egyik legismertebb módszer, a legnagyobb valószínűség (maximum likehood) elve. Ekkor azt keressük, hogy az eloszlás mely paraméterei mellett kapható meg a kísérlet során legnagyobb valószínűséggel az adott minta. (A becslés nagyban függ az adott mintától.)
  (Vegyük észre, hogy pl. az $\hat{m}_n$ mintaátlag az m várható értékre vonatkozóan, az $\hat{s}_n$ korrigált tapasztalati szórás a σ szórásra nézve pontbecslés.)

– **Intervallumbecslések**
  Olyan intervallumot keresünk, amelybe előre adott, nagy valószínűséggel beleesik a becsült paraméter valódi értéke. A becsült paraméter értékére tehát egy intervallumot adunk meg, ezzel az adott mintából származó bizonytalanságot csökkentjük.

A továbbiakban csak az intervallum-becsléssel foglalkozunk. Azt az intervallumbecslést, amely az esetek nagy, $(1-\varepsilon)\cdot 100\%$-ában tartalmazza a becsült α paramétert, $(1-\varepsilon)\cdot 100\%$ megbízhatósági szintű konfidencia-intervallumnak nevezzük. Az adott megbízhatósági szintű konfidencia-intervallum határai maguk is valószínűségi változók, melyek értéke az egyes minták esetében más és más.

23.12. definíció: A (c_1,c_2) intervallumot az X valószínűségi változó α paraméterére vonatkozó $(1-\varepsilon)\cdot 100\%$ megbízhatósági szintű konfidencia-intervalumnak nevezzük, ha

$$P(c_1 < \alpha < c_2) = 1-\varepsilon$$

teljesül.

Nézzük, hogyan határozhatunk meg a várható értékre vonatkozó konfidencia-intervallumot (újra kiemelve, hogy adott megbízhatósági szint mellett különböző minták esetén a határai változnak)!
A definícióból adódik, hogy a $(c_1;c_2)$ intervallum nem egyértelmű. Olyan intervallumot keresünk, amelyre teljesül, hogy $P(m < c_1) = P(m > c_2) = \dfrac{1-p}{2}$. Ekkor szimmetrikus konfidencia-intervallumról beszélhetünk.
Tekintsük az ismeretlen m várható értékű és **ismert** σ **szórású** valószínűségi változót. Az m várható érték becslésére egy **nagy** (n) **elemszámú** $X_1,X_2,\ldots,X_n$ mintát veszünk.
A centrális határeloszlás tétel szerint

$$\hat{m}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

közelítőleg normális eloszlású valószínűségi változó, melynek várható értéke m, szórása $\dfrac{\sigma}{\sqrt{n}}$. Transzformáció után az

$$\frac{\hat{m}_n - m}{\frac{\sigma}{\sqrt{n}}}$$

közel standard normális eloszlású valószínűségi változó.
Keressük azt a $(-u,u)$ intervallumot, amelybe $p = 1 - \varepsilon$ valószínűséggel esik ez a transzformált változó.

Az u értékének meghatározásához felhasználjuk a standard normális eloszlás Φ eloszlásfüggvényéről a valószínűség-számítás részben tanultakat.

$$P\left(-u < \frac{\hat{m}_n - m}{\frac{\sigma}{\sqrt{n}}} < u\right) = p.$$

$$\Phi(u) - \Phi(-u) = p.$$

Tudjuk, hogy $\Phi(-u) = 1 - \Phi(u)$, így :

$$2\Phi(u) - 1 = p.$$

Rendezve az egyenletet:

$$\Phi(u) = \frac{1+p}{2}.$$

Ha p adott, u értéke a standard normális eloszlás táblázatból kiolvasható. Nézzük meg, hogy ha az $\hat{m}_n$ mintaátlag adott, akkor milyen intervallumba esik p valószínűséggel az m várható érték. A korábbi egyenlőtlenséget az alábbiak szerint alakíthatjuk:

$$P\left(-u\frac{\sigma}{\sqrt{n}} < \hat{m}_n - m < u\frac{\sigma}{\sqrt{n}}\right) = p.$$

Szorozzuk az egyenlőtlenség mindhárom oldalát $\frac{\sigma}{\sqrt{n}}$-nel. (Az ekvivalens átalakítás nem változtatja meg a valószínűséget.)

$$P\left(-u\frac{\sigma}{\sqrt{n}} < \hat{m}_n - m < u\frac{\sigma}{\sqrt{n}}\right) = p.$$

Az egyenlőtlenségek iránya a (-1)-gyel szorzás után megfordul:

$$P\left(u\frac{\sigma}{\sqrt{n}} > m - \hat{m}_n > -u\frac{\sigma}{\sqrt{n}}\right) = p.$$

Hozzáadunk $\hat{m}_n$-et mindhárom oldalhoz:

$$P\left(\hat{m}_n + u\frac{\sigma}{\sqrt{n}} > m > \hat{m}_n - u\frac{\sigma}{\sqrt{n}}\right) = p.$$

A várható érték tehát az adott p valószínűséggel esik az

$$\left(\hat{m}_n - u\frac{\sigma}{\sqrt{n}}, \hat{m}_n + u\frac{\sigma}{\sqrt{n}}\right)$$

intervallumba, ahol

$$\Phi(u) = \frac{1+p}{2}.$$

23.2. tétel: Az X ismeretlen m várható értékű, de ismert σ szórású valószínűségi változó megfigyelésére nagy n elemszámú mintát veszünk, melyből a mintaátlag $\hat{m}_n$. Ekkor az m várható értékre vonatkozó $p = 1 - \varepsilon$ megbízhatósági szintű konfidencia-intervallum:

$$\left(\hat{m}_n - u\frac{\sigma}{\sqrt{n}}, \hat{m}_n + u\frac{\sigma}{\sqrt{n}}\right),$$

ahol

$$\Phi(u) = \frac{1+p}{2}.$$

Megjegyzés:

1. Ha X normális eloszlású, kis mintaelemszám esetén is ugyanezt a konfidencia-intervallumot kapjuk a várható értékre.
2. Nagy mintaelem-szám esetén, ha X normális eloszlású, ismeretlen σ szórású valószínűségi változó, akkor a fenti formulában σ-t az $\hat{s}_n$ korrigált tapasztalati szórással becsülhetjük.
3. A megbízhatósági szint emelése az intervallum sugarának növekedését vonja maga után, ugyanakkor a minta elemszámának növelése fordított hatású, csökkenti az intervallum sugarát.

23.3. feladat: Egy normális eloszlásból származó 120 elemű mintát vizsgálva a mintaátlag 8,4. A szórás ismert, 1,44.

a) Határozzunk meg a várható értékre vonatkozó 97%-os megbízhatósági szintű konfidencia-intervallumot!

b) Határozzunk meg a várható értékre vonatkozó 95%-os megbízhatósági szintű konfidencia-intervallumot!

Megoldás:

a) A 23.3. tétel formuláit akarjuk felhasználni. Ehhez a feladatból kiolvassuk az adatainkat:

$$n = 120,\ \hat{m}_n = 8{,}4,\ \sigma = 1{,}44,\ p = 0{,}97,$$

$$\Phi(u) = \frac{1 + 0{,}97}{2} = 0{,}985.$$

A standard normális eloszlás táblázatból visszakeresve:

$$u = 2{,}17.$$

Ezek után nincs más dolgunk, mint behelyettesíteni az

$$\left(\hat{m}_n - u\frac{\sigma}{\sqrt{n}}, \hat{m}_n + u\frac{\sigma}{\sqrt{n}}\right)$$

formulába.

Felhasználva az adatainkat:

$$\left(8{,}4 - 2{,}17\frac{1{,}44}{\sqrt{120}};\ 8{,}4 + 2{,}17\frac{1{,}44}{\sqrt{120}}\right).$$

Elvégezve a műveleteket, a kapott konfidencia-intervallum:

$$(8{,}1147;\ 8{,}6853).$$

Azaz:

$$P\left(8{,}1147 < m < 8{,}6853\right) = 0{,}97.$$

b) Az adataink közül csak a megbízhatósági szint változott: $p = 0{,}95$, így:

$$\Phi(u) = \frac{1 + 0{,}95}{2} = 0{,}975.$$

A standard normális eloszlás táblázatból kapjuk, hogy

$$u = 1{,}96.$$

Helyettesítés után a konfidencia-intervallum:

$$\left(8{,}4 - 1{,}96\frac{1{,}44}{\sqrt{120}};\ 8{,}4 + 1{,}96\frac{1{,}44}{\sqrt{120}}\right) = (8{,}1424;\ 8{,}6576).$$

Vegyük észre, hogy a megbízhatósági szint csökkentése az intervallum sugarának csökkenését vonta maga után (az intervallum hossza csökkent). Ezzel pontosabb, de kevésbé biztos becslést kaptunk, mivel az intervallum kisebb valószínűséggel tartalmazza az ismeretlen várható értéket.

Önellenőrzés

Önellenőrző kérdések

1. Adott a

176	181	179	181	184	179

statisztikai minta.

(a) Adjuk meg a mintaátlag értékét 4 tizedesjegy pontossággal! Megold.

(b) Adjuk meg a korrigált tapasztalati szórás értékét 4 tizedesjegy pontossággal! Megold.

(c) Adjuk meg a medián értékét! Megold.

2. Egy középiskola 10 tanulójának testmagasságát megmérve az alábbi adatokat kapták (cm-ben):

180	176	164	167	170	174	172	178	186	181

(a) Adjuk meg a tapasztalati szórás értékét 4 tizedesjegy pontossággal! Megold.

(b) Adjuk meg a mintaközép értékét 4 tizedesjegy pontossággal! Megold.

(c) Adjuk meg a minta terjedelmét! Megold.

3. Egy szabályos dobókockát 30-szor feldobva az alábbi dobáseredmények születtek:

Dobás értéke	1	2	3	4	5	6
Dobások száma	3	4	6	5	7	5

(a) Adjuk meg a mintaátlag értékét 4 tizedesjegy pontossággal!

Megold.

(b) Adjuk meg a korrigált tapasztalati szórás értékét 4 tizedesjegy pontossággal!

Megold.

4. 80 izzó élettartamát vizsgálva az alábbi eredményeket kapták (órában):

Élettartam (óra)	660-670	670-680	680-690	690-700	700-710	710-720	720-730	730-740
Izzók száma	4	5	12	24	15	10	7	3

(a) Adjuk meg a mintaátlag értékét 4 tizedesjegy pontossággal!

Megold.

(b) Adjuk meg a korrigált tapasztalati szórás értékét 4 tizedesjegy pontossággal!

Megold.

5. Egy $n = 10$ elemű statisztikai minta vizsgálata során a tapasztalati szórás értékére 2,5412 adódott. Mennyi a minta korrigált tapasztalati szórása?

☐ 2,4108 ☐ 2,5412 ☐ 2,6787 ☐ 2,5824

6. 150 speciális liszteszsák tömegét megmérve a mintaátlag 748 gramm. A szórást 3,6 grammnak feltételezve határozzuk meg a liszteszsák tömegének várható értékére

(a) 95%-os konfidencia-intervallumot!

☐ (747,4239; 748,5761) ☐ (747,3159; 748,6841) ☐ (747,4240; 748,6840) ☐ (747,5162; 748,4838)

(b) 98%-os konfidencia-intervallumot!

☐ (747,4240; 748,6840) ☐ (747,4239; 748,5761) ☐ (747,5169; 748,4831) ☐ (747,3151; 748,6849)

38. LECKE

A hipotézisvizsgálat

24. Hipotézisvizsgálat

Szinte mindannyian találkoztunk valamilyen formában már piackutatással,- illetve közvéleménykutatással. Előbbi esetén a vásárlási szokásokat vizsgálják, utóbbi esetében bizonyos kérdésekről akarják megismerni az emberek véleményét. A közös bennük, hogy a mintavétellel nyert véges számú adatból kell következtetéseket levonni az adott mintánál lényegesen nagyobb sokaságra vonatkozóan. A tudományos kutatások, gazdasági modellek felállítása során is gyakran merülnek fel olyan szakmai kérdések, melyek megválaszolása statisztikai hipotézisvizsgálat segítségével történhet. Az alábbi kérdések is ilyenek:

– Egy adott termék tömege megfelel-e az előírtnak?
– Hatásos-e egy bizonyos diéta?
– Összefügg-e az IQ és a kreativitás?
– Adott populáció kor szerinti eloszlása normális-e?
– Stb.

24.1. Statisztikai hipotézisek

24.1. definíció: Statisztikai hipotézisnek egy vagy több valószínűségeloszlásra vonatkozó feltevést nevezünk.

A statisztikai hipotézis vonatkozhat egy-, vagy több sokaságra, a feltételezett eloszlásra, illetve annak valamely paramétereire. Érvényességét a sokaságból véletlenszerűen vett minták segítségével vizsgálhatjuk. Nézzünk egy egyszerű példát. Egy újfajta fogyasztószerről szeretnénk eldönteni, hogy a gyártó állításának megfelelően valóban hatásos-e. Feltételezve, hogy a szer ténylegesen fogyasztó hatású, a hipotézisünk az, hogy a szer megfelelő használata fogyást eredményez.

24.2. definíció: Statisztikai próbának nevezzük az eljárást, amelynek segíségével a hipotézis elfogadásáról, illetve elutasításáról döntünk.

Fontos megjegyezni, hogy a hipotézisvizsgálat során soha nem arról döntünk, hogy a felállított hipotézisünk igaz, vagy nem igaz. A döntésünk csak az lehet, hogy az adott minta (minták) alapján a hipotézisünk elfogadható, hihető-e, vagy sem (nagy valószínűséggel igaz-e, vagy sem).

24.2. A statisztikai próba menete

A statisztikai próbák menete minden hipotézisvizsgálat során ugyanaz, az egyes próbák csak technikai részletekben térnek el egymástól.

24.2.1. A statisztikai próba elméleti lépései

1. A feltételezett eloszlásra, illetve annak valamely paraméterére felállítunk egy H_0 nullhipotézist (amely szinte mindig egyenlőség).
   Mivel előfordulhat, hogy vizsgálódásaink eredményeként a nullhipotézist elutasítjuk, szükséges a H_1 alternatív vagy ellenhipotézis meghatározása. A H_0 null- és a H_1 ellenhipotézist úgy kell megfogalmazni, hogy egymást kizárják. Közülük pontosan egyet fogadunk el, vagyis ha a nullhipotézist elutasítjuk, azzal az ellenhipotézist elfogadjuk és fordítva. A H_0 és a H_1 megfogalmazása attól függ, mi a vizsgálat célja. (Gyakran az érdekel bennünket, hogy a H_1 fennáll-e.) Általánosságban az mondható, hogy ha lehet, a null- és ellenhipotézist úgy fogalmazzuk meg, hogy **a nullhipotézis elvetése legyen fontos**.

24.1. feladat: Egy üzemben 100 g névleges tömegű csokoládészeleteket gyártanak. Azt akarjuk eldönteni, hogy jól van-e beállítva a gyártó automata. Állítsuk fel a null- és az ellenhipotézist!

Megoldás: A nullhipotézis ekkor az, hogy az automata jól van beállítva, azaz a csokoládészeletek tömegének várható értéke 100 g. Az ellenhipotézis ekkor természetesen adódik: az automata rosszul van beállítva, tehát a tömeg nem 100 g.
Ekkor hipotéziseink az alábbi alakba írhatók :

$$H_0 : E(X) = 100 \quad \text{(A csokiszeletek tömegének várható értéke } m_0 = 100.)$$
$$H_1 : E(X) \neq 100 \quad \text{(A csokiszeletek tömegének várható értéke nem } 100.)$$

Az ellenhipotézis a feladattól függően természetesen akár egyenlőtlenség is lehet, ahogy azt az alábbi példában láthatjuk.

24.2. feladat: Kicsit módosítsuk az előző feladatot! Az üzemben 100 g névleges tömeg gyártanak. A Fogyasztóvédelmi Felügyelet azt vizsgálja, hogy a gyártó megkárosítja- null-, illetve ellenhipotézist tudunk felállítani?

Megoldás: A feltételezésünk az, hogy a gyártó nem károsítja meg a vásárlóit, ekkor a csokiszeletek várható tömege legalább 100 g. Az ellenhipotézis pedig, hogy a gyártó becsapja a vásárlóit azzal, hogy a csokiszeletek várható tömege nem éri el a 100 g-ot.

$$H_0 : E(X) \geq 100 \quad \text{(A csokiszeletek tömegének várható értéke legalább } m_0 = 100.)$$
$$H_1 : E(X) < 100 \quad \text{(A csokiszeletek tömegének várható értéke kevesebb, mint } 100.)$$

Ha a feltételezésünk csak egyenlőtlenséget tartalmaz, akkor azt az ellenhipotézisben fogalmazhatjuk meg:

24.3. feladat: Hogyan fogalmazhatjuk meg a null- és ellenhipotézist, ha azt akarjuk vizsgálni, hogy a korábban említett fogyasztószer hatásos-e?

Megoldás: Ha a fogyasztó tabletta hatásos, megfelelő alkalmazása esetén súlycsökkenés várható. A feltételezésünk tehát az, hogy a súlyvesztés (a fogyókúra előtti, illetve utáni súly különbségének várható értéke) pozitív. Ezt az állítást csak az ellenhipotézisben fogalmazhatjuk meg. A nullhipotézis ennek az ellentettjét tartalmazza, azaz súlycsökkenés nem történt (ebbe az is belefér, hogy akár súlynövekedés várható).

$H_0: E(X) \leq 0$ (A súlycsökkenés várható értéke legfeljebb $m_0 = 0$, nem hatásos a fogyókúrás szer.)
$H_1: E(X) > 0$ (A súlycsökkenés várható értéke nagyobb, mint 0, hatásos a fogyasztó tabletta.)

A gyakorlatban azt mondhatjuk, hogy a **nullhipotézis**t úgy kell felállítani, hogy **mindig tartalmazza az egyenlőséget**, az **ellenhipotézis** pedig soha **ne tartalmazzon egyenlőséget**.

2. A próbafüggvény kiválasztása.
   A próbafüggvény egy alkalmasan megválasztott statisztikai függvény, amellyel H_0 fennállását vizsgáljuk. Gyakorlatilag a próbafüggvény egy valószínűségi változó, mely az egyes minták esetében különböző értéket vesz fel. A statisztikai próbák száma több százra tehető, a mi dolgunk kiválasztani azt a próbafüggvényt, amely a feladat feltételeinek megfelelő.
   Az előző példák esetében, ismert elméleti szórás (σ) mellett, n elemű mintát tekintve a megfelelő próbafüggvény:
   $$\alpha_n = \frac{\hat{m}_n - m_0}{\frac{\sigma}{\sqrt{n}}}$$
3. A próbafüggvény eloszlásának meghatározása.
   Ez a konkrét próba kiválasztása esetén a nullhipotézis fennállása mellett adott.
   A fenti példa esetében pl. standard normális eloszlás. (Emlékezzünk vissza a konfidencia-intervallum meghatározására!)

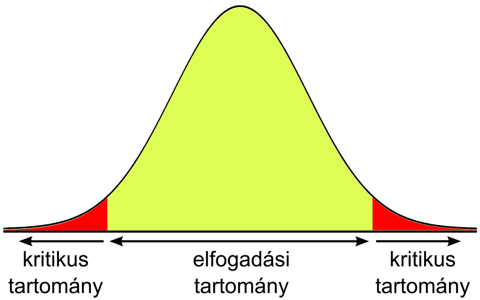
4. A próba megbízhatósági vagy szignifikanciaszintjének ($(1-\varepsilon)\cdot 100\%$) megadása (általában százalékos alakban).
   A szignifikanciaszint általában nagy, így ε értéke kicsi. Ha nincs egyéb kikötés, a szignifikanciaszintet 95%-nak ($0{,}95$-nek) tekinthetjük. (Ennek okára később visszatérünk.)

5. Kritikus vagy elutasítási tartomány kijelölése.
   Ebben a lépésben gyakorlatilag a próbafüggvény teljes értékkészletét bontjuk két tartományra. Kritikus (vagy elutasítási) tartománynak azt a részintervallumot nevezzük, ahova $(1-\varepsilon)\cdot 100\%$ szignifikanciaszint mellett a próbastatisztika értéke ε valószínűséggel esik. Az intervallumok határait minden esetben a próbafüggvény eloszlásának ismeretében határozhatjuk meg. A kritikus tartomány komplementere az elfogadási tartomány.
   A kritikus tartomány megválasztása azért nagyon fontos, mert ettől függ a hipotéziseinkről való döntés.

Az ellenhipotézistől függően a kritikus tartomány kijelölésében az alábbi eseteket különböztetjük meg . (A várható értékre vonatkozó nullhipotézis esetén mutatjuk be a tartomány kijelölését.)

(a) Kétoldali próba.
Az ellenhipotézis a feltételezett értéktől bármilyen irányú eltérést tartalmaz.

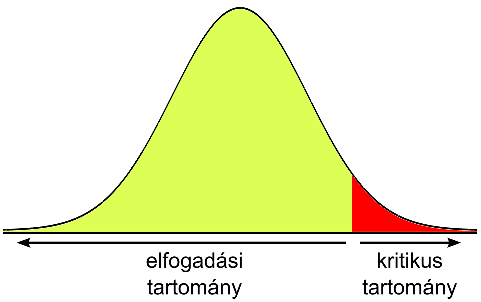
$$H_0: E(X) = m_0$$
$$H_1: E(X) \neq m_0$$



24.1. ábra. *Kritikus tartomány kijelölése kétoldali próba esetén.*

(b) Egyoldali próba.
Az ellenhipotézis a feltételezett értéktől csak egyirányú eltérést tartalmaz. Vegyük észre, hogy ekkor a kritikus tartomány kijelölése az ellenhipotézis egyenlőtlenségének irányától függ.
jobb oldali próba:

$$H_0: E(X) \leq m_0$$
$$H_1: E(X) > m_0$$

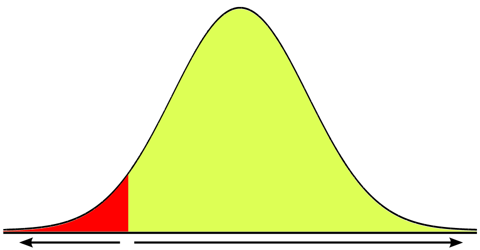


24.2. ábra. *Kritikus tartomány kijelölése jobb oldali próba esetén.*

A legtöbb próba egyoldali és kétoldali kritikus tartomány esetén is elvégezhető, de néhány speciális esetben a próba csak jobb oldali lehet. Ilyen lesz később például az F- és a χ^2-próba is.

bal oldali próba:

$$H_0 : E(X) \geq m_0$$
$$H_1 : E(X) < m_0$$



24.3. ábra. *Kritikus tartomány kijelölése bal oldali próba esetén.*

6. Az $\hat{\alpha}_n$ próbafüggvény értékének (számított érték) meghatározása az adott minta alapján.
   A minta adatait egyszerűen behelyettesítjük a próbastatisztika formulába.

7. Döntés.
   Attól függően, hogy a próbafüggvény számított értéke a kritikus vagy az elfogadási tartományba esik, döntünk a H_0 hipotézis elfogadásáról illetve elvetéséről. H_0-t elfogadjuk, ha a számított érték az elfogadási tartományba esik, és elvetjük, ha a számított érték a kritikus tartományba esik. A H_0 hipotézisről való döntéssel egyben a H_1 hipotézisről is döntünk. Ha H_0-t elfogadjuk, egyben H_1-et elvetjük, ha H_0-t elvetjük, egyben H_1-et elfogadjuk.
   Megjegyzés: Vegyük észre, hogy a hipotézisvizsgálat során gyakorlatilag arról döntünk, hogy az ismert eloszlásból nagy, $1 - \varepsilon$ valószínűséggel kijöhet-e az adott minta.

24.2.2. A statisztikai próba gyakorlati lépései

1. A H_0 null- és a H_1 ellenhipotézis felállítása.
2. A próbafüggvény kiválasztása és számított értékének meghatározása az adott minta alapján.
3. A kritikus- és elfogadási tartomány kijelölése a próbafüggvény eloszlása és a szignifikanciaszint ismeretében.
4. Döntés a H_0 hipotézis elfogadásáról vagy elvetéséről. (A gyakorlati kérdés megválaszolása szavakkal!)

24.3. Hibalehetőségek

A statisztikai döntéshozatal során (megfelelően lefolytatva a próbát) kétféle hibát követhetünk el. Elsőfajú a hiba, ha a H_0 hipotézis igaz, ennek ellenére elutasítjuk. Ha a rossz H_0 hipotézist elfogadjuk, másodfajú hibát követünk el. Összefoglalva:

	H_0 igaz	H_0 hamis
H_0-t elfogadjuk	Jó döntés	Másodfajú hiba
H_0-t elutasítjuk	Elsőfajú hiba	Jó döntés

Az elsőfajú hiba elkövetésének valószínűsége megegyezik ε-nal, így ennek változtatásával befolyásolhatjuk a hiba valószínűségét (például gyógyszerkísérletek esetén a szignifikanciaszintet $(1-\varepsilon)$ magasabbra állítják, hogy a hiba (ε) valószínűségét csökkentsék). A másodfajú hiba valószínűsége a minta elemszámának növelésével csökkenthető. Arra azonban ügyelni kell, hogy adott elemszám mellett az elsőfajú hiba csökkenésével nő a másodfajú hiba valószínűsége. A gyakorlatban legtöbbször használt 95%-os szignifikanciaszint kompromisszumnak tekinthető a kétfajta hiba nagyságát illetően.

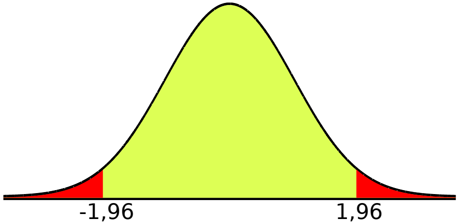
Nézzünk példákat a kritikus tartomány kijelölésére egy- és kétoldali próba esetén, ha a próbastatisztika standard normális eloszlású:

24.4. feladat: Határozzuk meg a kritikus tartományt kétoldali próba esetén, 95%- os szignifikanciaszint mellett, ha a próbastatisztika standard normális eloszlású!

Megoldás: Kétoldali próba esetén a hipotéziseink az alábbi alakba írhatók:

$$H_0: E(X) = m_0$$
$$H_1: E(X) \neq m_0$$

A kritikus tartomány határait jelölje $-u_{kr}$, illetve u_{kr}. A kritikus tartomány ekkor $(-\infty; -u_{kr}) \cup (u_{kr}; \infty)$. Az u_{kr} érték meghatározásához vezessük be a $p = 1 - \varepsilon$ jelölést, azaz a szignifikanciaszintet jelölje p. Ekkor keressük azt az u_{kr} értéket, melyre teljesül, hogy $\Phi(u_{kr}) = p + \frac{1-p}{2} = \frac{1+p}{2} = \frac{1+0{,}95}{2} = 0{,}975$. A standard normális eloszlás eloszlásfüggvényének értékeit tartalmazó táblázatból kiolvasható, hogy $u_{kr} = 1{,}96$.



24.4. ábra. *Kritikus tartomány kijelölése a 24.4. feladathoz.*

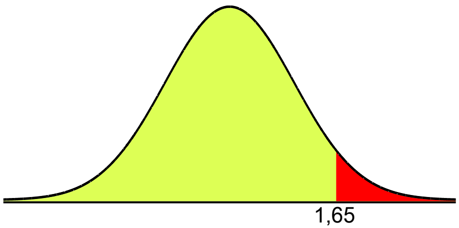
24.5. feladat: Határozzuk meg a kritikus tartományt egyoldali próba esetén, 95%- os szignifikanciaszint mellett, ha a próbastatisztika standard normális eloszlású!

Megoldás: Az ellenhipotézistől függően két esetet kell megkülönböztetnünk:

1. eset:

$$H_0 : E(X) \leq m_0$$
$$H_1 : E(X) > m_0$$

A kritikus tartomány ekkor $(u_{kr}; \infty)$. Az előző feladat jelölését használva $\Phi(u_{kr}) = 1 - \varepsilon = p = 0{,}95$. A táblázatból $u_{kr} = 1{,}65$.

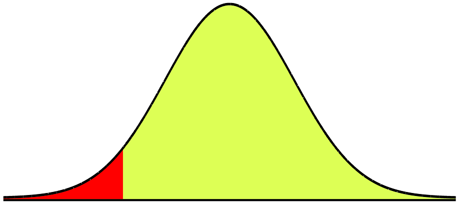


24.5. ábra. *jobb oldali kritikus tartomány kijelölése a 24.5. feladathoz.*

2. eset:

$$H_0: E(X) \geq m_0$$
$$H_1: E(X) < m_0$$

A kritikus tartomány ekkor $(-\infty, -u_{kr})$. A standard normális eloszlás szimmetriáját kihasználva $\Phi(u_{kr}) = 1 - \varepsilon = p = 0{,}95$. Ekkor (a táblázatból kiolvasott érték előjelét meg kell változtatnunk) $u_{kr} = -1{,}65$. A kritikus tartomány tehát a $(-\infty; -1{,}65)$.



24.6. ábra. *bal oldali kritikus tartomány kijelölése a 24.5. feladathoz.*

Önellenőrzés

Önellenőrző kérdések

1. Mit értünk első,- illetve másodfajú hibán hipotézisvizsgálat esetén?
2. Sejtésünk szerint a szerviz javítja az autók fogyasztását (azaz szerviz után csökken az autók fogyasztása). Hipotézisvizsgálattal akarjuk ellenőrizni a feltételezésünket. Fogalmazzuk meg az ehhez szükséges null- és ellenhipotézist!

   ☐ $H_0: E(X) = 0,$ $H_1: E(X) > 0$

   ☐ $H_0: E(X) \geq 0,$ $H_1: E(X) < 0$

   ☐ $H_0: E(X) < 0,$ $H_1: E(X) \geq 0$

   ☐ $H_0: E(X) = 0,$ $H_1: E(X) < 0$
3. Határozzuk meg a kritikus tartományt 98% -os szignifikanciaszint és standard normális eloszlású próbafüggvény esetén, ha hipotéziseink

$$H_0: E(X) = m_0$$
$$H_1: E(X) \neq m_0$$

   alakba írhatók!

   ☐ $(-\infty; -2{,}05) \cup (2{,}05; \infty)$

   ☐ $(-\infty; -2{,}33) \cup (2{,}33; \infty)$

   ☐ $(2{,}33; \infty)$

   ☐ $(-2{,}05; 2{,}05)$

4. Határozza meg a kritikus tartományt 96% -os szignifikanciaszint és standard normális eloszlású próbafüggvény esetén, ha hipotéziseink

$$H_0: E(X) \geq m_0$$
$$H_1: E(X) < m_0$$

alakba írhatók!

☐ $(2{,}06; \infty)$
☐ $(-\infty; -2{,}32)$
☐ $(-\infty; -2{,}06) \cup (2{,}06; \infty)$
☐ $(-\infty; -1{,}75)$

5. Határozza meg a kritikus tartományt 99% -os szignifikanciaszint és standard normális eloszlású próbafüggvény esetén, ha hipotéziseink

$$H_0: E(X) \leq m_0$$
$$H_1: E(X) > m_0$$

alakba írhatók!

☐ $(-\infty; -2{,}05) \cup (2{,}05; \infty)$
☐ $(2{,}33; \infty)$
☐ $(2{,}05; \infty)$
☐ $(-2{,}33, \infty)$

39. LECKE

Egymintás próbák

25. Paraméteres próbák

Paraméteres próbák esetében a H_0 hipotézis és a H_1 ellenhipotézis a megfigyelt valószínűségi változó valamely paraméterére vonatkozik. Mi a továbbiakban a várható értékre, illetve a szórásra kimodott hipotézisek ellenőrzésére alkalmas legfontosabb próbákat tekintjük át.
Elsőként nézzük azokat a próbákat, amelyek egyetlen minta vizsgálatán alapulnak. A továbbiakban őket összefoglalóan egymintás próbáknak nevezzük.

25.1. Várható értékre vonatkozó egymintás próbák

Ismeretlen várható értékű, normális eloszlású valószínűségi változó várható értékére vonatkozó H_0: $E(X) = m_0$ hipotézis helyességét akarjuk ellenőrizni.

25.1.1. Egymintás u-próba

Azt szeretnénk eldönteni, hogy egy n elemű, $\hat{m}_n$ mintaátlagú statisztikai minta elméleti várható értéke lehet-e m_0.

1. Legyen a valószínűségi változó σ **szórása ismert**:
   Ekkor, feltételezve, hogy a H_0 hipotézis helyes, a mintaelemek m_0 várható értékű, σ szórású valószínűségi változók. Valószínűségszámítás tanulmányaink alapján az $\hat{m}_n$ valószínűségi változó normális eloszlású, m várható értékkel és $\dfrac{\sigma}{\sqrt{n}}$ szórással. A transzformáció után kapott

   $$u_p = \frac{\hat{m}_n - m_0}{\dfrac{\sigma}{\sqrt{n}}}$$

   próbastatisztika standard normális eloszlású.
   Megjegyzés: A centrális határeloszlás tétele miatt nagy elemszámú minta esetén u_p a megfigyelt valószínűségi változó eloszlásától függetlenül standard normális eloszlásúnak tekinthető.

2. Most nézzük azt az esetet, amikor a valószínűségi változó σ **szórása nem ismert**, de a minta **elemszáma nagy** (mi a továbbiakban nagy elemszámon az $n \geq 30$ esetet értjük, de ettől természetesen felfelé el lehet térni):
A centrális határeloszlás tétele alapján a megfigyelt valószínűségi változó eloszlásától függetlenül a mintaátlag közel normális eloszlású. Szeretnénk standard normális eloszlásúvá transzformálni, de ehhez nem ismerjük a szórását. Az $\hat{s}_n$ korrigált tapasztalati szórást felhasználva azonban a mintaátlag szórását $\dfrac{\hat{s}_n}{\sqrt{n}}$-nel becsülhetjük. Az

$$u_p = \frac{\hat{m}_n - m_0}{\dfrac{\hat{s}_n}{\sqrt{n}}}$$

próbastatisztika közel standard normális eloszlású.
Megjegyzés: A korábbiakban említettük a korrigált és a tapasztalati szórás közötti összefüggést: $\hat{\sigma}_n = \sqrt{\dfrac{n}{n-1}} \cdot \hat{s}_n$. Ha n nagy, akkor $\sqrt{\dfrac{n}{n-1}} \approx 1$, tehát $\hat{s}_n \approx \hat{\sigma}_n$. Nagy elemszámú minta esetén tehát a korrigált tapasztalati szórás helyettesíthető a tapasztalati szórással.

25.1. feladat: Egy normális eloszlású valószínűségi változó megfigyelésére $n = 8$ mérést végeztek. Az eredményeket az alábbi táblázat tartalmazza:

102	108	95	98	104	90	94	104

Feltételezve, hogy a valószínűségi változó szórása $\sigma = 6$, 95%-os szignifikanciaszinten igaz-e, hogy a valószínűségi változó várható értéke 100?

Megoldás: Kövessük a megoldás során a hipotézisvizsgálat 24.2.2. fejezetben megismert gyakorlati lépéseit!

1. *A null- és ellenhipotézis kimondása.*
   A feltételezésünk az, hogy a várható érték $m_0 = 100$. A feladatban semmilyen megszorítás nincs az ettől való eltérés irányára, az ellenhipotézis tehát, hogy a várható érték nem 100. Formálisan a hipotéziseink:

$$H_0 \colon E(X) = 100$$

$$H_1 \colon E(X) \neq 100$$

   Az ellenhipotézis alapján a próba kétoldali.

2. *A próba kiválasztása és a minta alapján a próbastatisztika értékének kiszámítása.*
   Mivel a minta normális eloszlásból származik, és a szórás ismert, u-próbát alkalmazunk. Az $u_p = \dfrac{\hat{m}_n - m_0}{\frac{\sigma}{\sqrt{n}}}$ próbastatisztika standard normális eloszlású valószínűségi változó. Ahhoz, hogy az adott mintából számított értékét meghatározzuk, a feladatban megadott $n = 8$ és $\sigma = 6$ mellett ki kell számítanunk a mintaátlagot: $\hat{m}_8 = \dfrac{102 + 108 + \ldots + 104}{8} = \dfrac{795}{8} = 99{,}375.$
   A próbastatisztika formulába helyettesítve:

$$u_p = \frac{\hat{m}_n - m_0}{\frac{\sigma}{\sqrt{n}}} = \frac{99{,}375 - 100}{\frac{6}{\sqrt{8}}} = -0{,}2946.$$

3. *A kritikus érték meghatározása.*
   A próba kétoldali u-próba, $p = 0{,}95$ szignifikanciaszinttel. A korábbiakban láttuk a kritikus tartomány kijelölését kétoldali próba esetén, ha a próbastatisztika standard normális eloszlású.

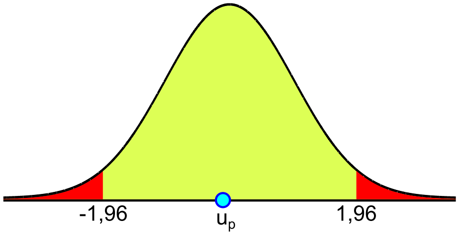
A kritikus érték meghatározásához a

$$\Phi(u_{kr}) = \frac{1+0{,}95}{2} = 0{,}975$$

egyenlőségnek kell teljesülni. A standard normális eloszlás táblázatából u_{kr} értéke kiolvasható:

$$u_{kr} = 1{,}96.$$

Ennek alapján az alábbi ábrán látható a kritikus, illetve az elfogadási tartomány, melyben elhelyeztük a próbastatisztika értékét is (kék pont).



25.1. ábra. *Kritikus tartomány kijelölése a 25.1. feladathoz.*

4. *A döntés.*
   A próbastatisztika értéke az elfogadási tartományba esik, ezért a H_0 hipotézist elfogadjuk, 95%-os szignifikanciaszinten a minta nem cáfolja, hogy a valószínűségi változó várható értéke 100.

25.2. feladat: A tapasztalatok szerint egy nyelvi teszt eredménye normális eloszlású valószínűségi változó, 70 pont várható értékkel. Sajnos a szórás nem ismert. Egy évfolyam hallgatói közül véletlenszerűen kiválasztottak 36-ot, ők kitöltötték a tesztet. Az átlagpontszámuk 68,5 lett, a korrigált tapasztalati szórás 6. Elfogadható-e 95%-os szignifikanciaszinten, hogy ez az évfolyam gyengébben teljesít az átlagosnál ezen a teszten?

Megoldás: A megszokott jelöléseinket használva a feladatból kiolvashatóak a következő adatok: $n = 36$, $\hat{m}_n = 68{,}5$, $\hat{s}_n = 6$. Lássuk a hipotézisvizsgálat lépéseit!

1. *A null- és ellenhipotézis kimondása.*
   Ebben a feladatban a minket érdeklő kérdést (gyengébben teljesít az évfolyam az átlagosnál) az ellenhipotézisben tudjuk csak megfogalmazni. (Emlékezzünk, a nullhipotézis mindig tartalmazza az egyenlőséget!) A nullhipotézis ennek az ellentettje, az évfolyam legalább átlagosan teljesít a teszten (a pontszámuk várható értéke legalább 70), az ellenhipotézis pedig, hogy ez az évfolyam gyengébben teljesít a teszten, azaz a pontszámuk várható értéke kevesebb, mint 70.

$$H_0\colon E(X) \geq 70$$

$$H_1\colon E(X) < 70$$

   Az ellenhipotézis alapján a próba egyoldali.

2. *A próba kiválasztása és a minta alapján a próbastatisztika értékének kiszámítása.*
   A minta normális eloszlásból származik, a szórás nem ismert, de a minta elemszáma nagy ($36 \geq 30$), ezért u-próbát alkalmazunk. A próbastatisztika értéke a minta adatait felhasználva:

$$u_p = \frac{\hat{m}_n - m_0}{\frac{\hat{s}_n}{\sqrt{n}}} = \frac{68{,}5 - 70}{\frac{6}{\sqrt{36}}} = -1{,}5.$$

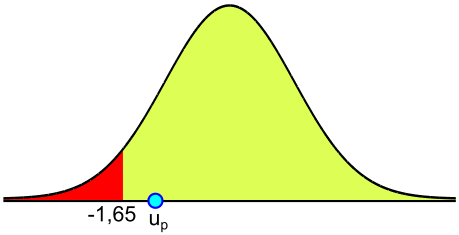
3. *A kritikus érték meghatározása.*
   A próba egyoldali (bal oldali) u-próba, a kritikus érték ennek megfelelően az

$$\Phi(u_{kr}) = 0{,}95$$

egyenlőségből számítható. A standard normális eloszlás táblázatából u_{kr} értéke kiolvasható:

$$u_{kr} = -1{,}65.$$

Ennek alapján az alábbi ábrán látható a kritikus, illetve az elfogadási tartomány, melyben elhelyeztük a próbastatisztika értékét is (kék pont).



25.2. ábra. *Kritikus tartomány kijelölése a 25.2. feladathoz.*

4. *A döntés.*
   A próbastatisztika értéke az elfogadási tartományba esett, a nullhipotézist elfogadjuk 95%-os szignifikanciaszinten. Ez egyben azt jelenti, hogy az ellenhipotézist elvetjük, tehát az adott szignifikanciaszinten nem fogadható el, hogy ez az évfolyam gyengébben teljesít, a minta alapján nem szignifikáns az eltérés a 70-től.

Aktivitás: A számított érték nagyon közel van a kritikus értékhez. Vizsgálja meg, hogy a szignifikanciaszintet 90%- ra, illetve 98%-ra változtatva hogyan változik a döntésünk!

A H_0 hipotézis elvetésével az esetlegesen elkövetett hiba tehát elsőfajú lehet. A szignifikanciaszint emelésével csökkenthető az elsőfajú hiba valószínűsége.

25.1.2. Egymintás t-próba

Továbbra is azt szeretnénk eldönteni, hogy egy n elemű, $\hat{m}_n$ mintaátlagú statisztikai minta elméleti várható értéke lehet-e m_0.
A korábbiakban azokat az eseteket tárgyaltuk, amikor vagy a valószínűségi változó σ szórása ismert, vagy a minta elemszáma nagy.
Most nézzük azt az esetet, amikor a valószínűségi változó σ **szórása nem ismert** és a minta **elemszáma kicsi** $(n < 30)$.
Ekkor, feltételezve, hogy a H_0: $E(X) = m_0$ hipotézis helyes, a

$$t_p = \frac{\hat{m}_n - m_0}{\frac{\hat{s}_n}{\sqrt{n}}}$$

próbastatisztika $(n-1)$ szabadsági fokú, Student-t-eloszlású.
Megjegyzés: A korrigált és a tapasztalati szórás közötti összefüggés természetesen továbbra is érvényes: $\hat{\sigma}_n = \sqrt{\frac{n}{n-1}} \cdot \hat{s}_n$. Azonban ha n kicsi, akkor $\sqrt{\frac{n}{n-1}}$ durván közelíti az 1-et, ezért az $\hat{s}_n$ korrigált tapasztalati szórás ebben az esetben nem helyettesíthető a $\hat{\sigma}_n$ tapasztalati szórással.
(Az eddigieket figyelembe véve biztosan nem követünk el hibát, ha a formulákban σ közelítésére mindig az $\hat{s}_n$ korrigált tapasztalati szórást használjuk!)

Az alábbi táblázat összefoglalja az egymintás u- és t-próba alkalmazását.

	σ ismert	σ nem ismert
$n < 30$	u-próba: $u_p = \frac{\hat{m}_n - m_0}{\frac{\sigma}{\sqrt{n}}}$	t-próba: $t_p = \frac{\hat{m}_n - m_0}{\frac{\hat{s}_n}{\sqrt{n}}}$
$n \geq 30$	u-próba: $u_p = \frac{\hat{m}_n - m_0}{\frac{\sigma}{\sqrt{n}}}$	u-próba: $u_p = \frac{\hat{m}_n - m_0}{\frac{\hat{s}_n}{\sqrt{n}}}$

25.3. feladat: Egy ismeretlen szórású, normális eloszlású valószínűségi változó megfigyelésére $n = 5$ mérést végeztek. Az eredményeket az alábbi táblázat tartalmazza:

1,51	1,49	1,54	1,52	1,54

95%-os szignifikanciaszinten elfogadható-e, hogy a valószínűségi változó várható értéke 1,50?

Megoldás: Ahogy korábban említettük, a hipotézisvizsgálat lépései minden próba esetén hasonlóak. Kövessük ezeket a lépéseket a t-próba esetén!

1. *A null- és ellenhipotézis kimondása.*
   A feltételezésünk az, hogy a várható érték $m_0 = 1{,}50$, ettől semmilyen irányú eltérést nem engedünk meg.

   $$\begin{aligned} &H_0: E(X) = 1{,}50 \quad &&\text{(A várható érték 1,50.)} \\ &H_1: E(X) \neq 1{,}50 \quad &&\text{(A várható érték nem 1,50.)} \end{aligned}$$

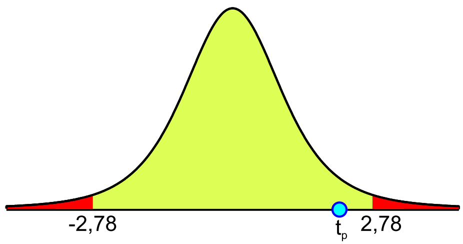
   Az ellenhipotézis alapján a próba kétoldali.

2. *A próba kiválasztása és a minta alapján a próbastatisztika értékének kiszámítása.*
   Mivel a szórás nem ismert, a minta elemszáma pedig kicsi, t-próbát alkalmazunk. A próbastatisztika formulába helyettesítsük a minta adatait, így megkapjuk a próbastatisztika számított értékét.

   $$t_p = \frac{\hat{m}_n - m_0}{\frac{\hat{s}_n}{\sqrt{n}}} = \frac{1{,}52 - 1{,}50}{\frac{0{,}0212}{\sqrt{5}}} = 2{,}1082.$$

   A számításhoz felhasználtuk, hogy
   $\hat{m}_5 = \dfrac{1.51 + 1.49 + \ldots + 154}{5} = \dfrac{7{,}6}{5} = 1{,}52$, és
   $\hat{s}_5 = \sqrt{\dfrac{(1{,}51 - 1{,}52)^2 + (1{,}49 - 1{,}52)^2 + \ldots + (1{,}54 - 1{,}52)^2}{5 - 1}} = \sqrt{\dfrac{0{,}0018}{4}} = \sqrt{0{,}00045} = 0{,}0212.$

3. *A kritikus érték meghatározása.*
A próbastatisztika Student-eloszlású, t_{kr} értéke a 4-szabadsági fokú Student-eloszlás táblázatából $p = \dfrac{1+0{,}95}{2} = 0{,}975$ értéknél:
$$t_{kr} = 2{,}78.$$
Ennek alapján az alábbi ábrán látható a kritikus, illetve az elfogadási tartomány, melyben elhelyeztük a próbastatisztika értékét is (kék pont).



25.3. ábra. *Kritikus tartomány kijelölése a 25.3. feladathoz.*

4. *A döntés.*
A próbastatisztika számított értéke az elfogadási tartományba esik, H_0 hipotézist elfogadjuk, 95%-os szignifikanciaszinten elfogadható, hogy a valószínűségi változó várható értéke 1,50.

25.4. feladat: Egy vizsgálatban a vegetariánusok és a húsevők koleszterinszintjét akarták összehasonlítani. Ehhez 8 véletlenszerűen kiválasztott vegetariánus koleszterinszintjét mérték meg. Az alábbi értékeket kapták (mmol/l):

3,1	2,8	1,5	1,7	2,4	2,0	3,3	1,6

A húsevők koleszterinszintjének várható értéke 3,1. Feltételezve, hogy a minta normális eloszlásból származik, 98%-os szignifikanciaszinten igazolható-e, hogy a vegetariánusok koleszterinszintjének várható értéke alacsonyabb, mint a húsevőké?

Megoldás: A minta elemszáma: $n = 8$.

1. *A null- és ellenhipotézis kimondása.*
   A kérdés most arra irányul, hogy a vegetariánusok koleszterinszintjének várható értéke kevesebb-e, mint $m_0 = 3{,}1$. Emlékezzünk, hogy a nullhipotézisnek mindig tartalmaznia kell egyenlőséget. A feltevésünkre (a vegetariánusok koleszterinszintjének várható értéke kevesebb, mint 3,1), ami formálisan $E(X) < 3{,}1$ alakba írható, ez nem teljesül, így csak az ellenhipotézisben fogalmazhatjuk meg. Ekkor a nullhipotézis az, hogy a vegetariánusok koleszterinszintjének várható értéke legalább 3,1.

   $H_0:\ E(X) \geq 3{,}1$ (A vegetariánusok koleszterinszintjének várható értéke legalább akkora, mint a húsevőké.)

   $H_1:\ E(X) < 3{,}1$ (A vegetariánusok koleszterinszintjének várható értéke alacsonyabb, mint a húsevőké.)

   A próba egyoldali.

2. *A próba kiválasztása és a minta alapján a próbastatisztika értékének kiszámítása.*
Ismeretlen szórás és kis elemszám mellett t-próbát alkalmazunk. Az előző feladathoz hasonlóan a mintaátlagot és a korrigált tapasztalati szórást a mintából kell számítanunk.
$\hat{m}_8 = \dfrac{3{,}1 + 2{,}8 + \ldots + 1{,}6}{8} = \dfrac{18{,}4}{8} = 2{,}3$, és
$\hat{s}_8 = \sqrt{\dfrac{(3{,}1 - 2{,}3)^2 + (2{,}8 - 2{,}3)^2 + \ldots + (1{,}6 - 2{,}3)^2}{8 - 1}} = \sqrt{\dfrac{3{,}48}{7}} = \sqrt{0{,}4971} = 0{,}7051.$
A kapott adatokat a próbastatisztika formulába helyettesítve kapjuk a próbastatisztika számított értékét.

$$t_p = \frac{\hat{m}_n - m_0}{\frac{\hat{s}_n}{\sqrt{n}}} = \frac{2{,}3 - 3{,}1}{\frac{0{,}7061}{\sqrt{8}}} = -3{,}2091.$$

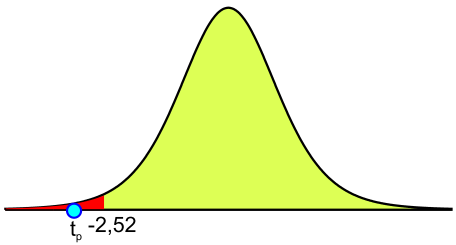
3. *A kritikus érték meghatározása.*
A próba egyoldali, t_{kr} értéke a 7-szabadsági fokú, Student-eloszlás táblázatából a $0{,}98$ értéknél:

$$t_{kr} = -2{,}52,$$

kihasználva, hogy a próba bal oldali.

Ennek alapján az alábbi ábrán látható a kritikus, illetve az elfogadási tartomány, melyben elhelyeztük a próbastatisztika értékét (kék pont).

4. *A döntés.*
A próbastatisztika értéke a kritikus tartományba esik, H_0 hipotézist elutasítjuk, a H_1 ellenhipotézist elfogadjuk. 98%-os szignifikanciaszinten elfogadható, hogy a vegetariánusok koleszterinszintje alacsonyabb, mint a húsevőké.
Vegyük észre, hogy a szignifikanciaszint növelésével csökkenthetjük az elsőfajú hiba valószínűségét.



25.4. ábra. *Kritikus tartomány kijelölése a 25.4. feladathoz.*

Megjegyzés: Vegyük észre, milyen sok múlik a kérdés megfogalmazásán! Ha a 25.4. feladatban azt kérdezzük, hogy 98%-os szignifikanciaszinten igazolható-e, hogy a vegetariánusok koleszterinszintjének várható értéke legfeljebb akkora, mint a húsevőké, akkor a hipotéziseket másképp kell kimondani. Ebben az esetben ugyanis az a feltevésünk, hogy a vegetariánusok koleszterinszintjének várható értéke legfeljebb $3{,}1$ $(E(X) \leq 3{,}1)$, ami tartalmazza az egyenlőséget, tehát nullhipotézisként kell megfogalmazni.

Aktivitás: Mondja ki a null- és az ellenhipotézist a 25.4. feladat fenti módosításához, majd végezze el a hipotézisvizsgálatot! Hasonlítsa össze a kapott eredményt a 25.4. feladat eredményével!

25.5. feladat: 10 emberen vizsgálták egy fogyasztótabletta hatását. A vizsgálat kezdetén megmérték a súlyukat, majd két héten át napi rendszerességgel szedték a tablettát. A második hét végén ismét megmérték a súlyukat. Az eredményeket (kg) az alábbi táblázat tartalmazza. (Feltesszük, hogy az adatok normális eloszlásból származnak.)

Diéta előtt	88	74	92	81	73	96	89	102	98	85
Diéta után	84	68	85	83	74	90	88	98	99	85

Csökkentette-e a súlyukat a fogyasztó tabletta?

Megoldás: A minták nem függetlenek, mivel ugyanannak a 10 embernek a súlyát mérték meg, ezért egymintás próbát kell alkalmaznunk. A kérdés valójában a két mérés között bekövetkezett súlyváltozásra vonatkozik, így a hipotézisvizsgálatot az alábbi mintára végezzük el:

Súlyváltozás a második hét végére (kg)	4	−6	−7	2	1	−6	−1	−4	1	0

A minta elemszáma $n = 10$.

1. *A null- és ellenhipotézis kimondása.*
   Most a feltevésünk az, hogy a tabletta hatására csökkent a testsúly, vagyis a tabletta hatására bekövetkező testsúlycsökkenés várható értéke nagyobb, mint 0. Az előző feladathoz hasonlóan megint csak az ellenhipotézisként fogalmazhatjuk meg ezt az állításunkat (egyenlőséget nem tartalmaz).

   $$H_0: E(X) \geq 0 \quad \text{(A fogyasztótabletta hatására nem változott a testsúly.)}$$
   $$H_1: E(X) < 0 \quad \text{(A fogyasztótabletta hatására csökkent a testsúly.)}$$

2. *A próba kiválasztása és a minta alapján a próbastatisztika értékének kiszámítása.*
   Mivel a szórás nem ismert, a minta elemszáma pedig kicsi, t-próbát alkalmazunk.

   $$t_p = \frac{\hat{m}_n - m_0}{\frac{\hat{s}_n}{\sqrt{n}}} = \frac{-2{,}4 - 0}{\frac{3{,}3731}{\sqrt{10}}} = -2{,}25.$$

A számításhoz felhasználtuk, hogy

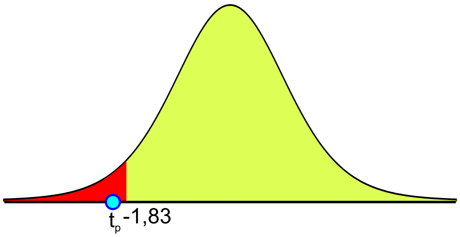
$$\hat{m}_{10} = \frac{-4 + (-6) + \ldots + 0}{10} = -2{,}4, \text{ és}$$

$$\hat{s}_{10} = \sqrt{\frac{(-4-(-2{,}4))^2 + (-6-(-2{,}4))^2 + \ldots + (0-(-2{,}4))^2}{10-1}} = \sqrt{\frac{102{,}4}{9}} = \sqrt{11{,}3778} = 3{,}3731.$$

3. *A kritikus érték meghatározása.*
Ahogy korábban említettük, ha a szignifikanciaszintre nincs egyéb kikötés, tekintsük $0{,}95$-nek.
A próba egyoldali, t_{kr} értéke a 9-szabadsági fokú, Student-eloszlás táblázatból :

$$t_{kr} = -1{,}83.$$

Ennek alapján az alábbi ábrán látható a kritikus, illetve az elfogadási tartomány, melyben elhelyeztük a próbastatisztika értékét is (kék pont).



25.5. ábra. *Kritikus tartomány kijelölése a 25.5. feladathoz.*

4. *A döntés.*
   A próbastatisztika számított értéke a kritikus tartományba esik, a H_0 hipotézist elvetjük, a H_1 ellenhipotézist elfogadjuk. 95%-os szignifikanciaszinten elfogadható, hogy a fogyasztótabletta csökkenti a testsúlyt.

Aktivitás: Oldjuk meg a feladatot abban az esetben, ha a szignifikanciaszintet $97{,}5\%$-ra változtatjuk. Hogyan változik a döntés?

Az egymintás próbák zárásául nézzük meg az alábbi feladatot.

25.6. feladat: Egy valószínűségszámítás kurzus egyik gyakorlatán 10 hallgató próbavizsgát írt. Az eredményeket látva az oktató azt feltételezi, hogy ez az évfolyam gyengébb a korábbiaknál, amelyeknél a pontszám várható értéke 45 volt, 14 pont szórással.

a) Elfogadva, hogy a szórás most is 14 pont, valóban gyengébben teljesít ez az évfolyam a korábbiaknál?

b) Mi a válasz az előző kérdésre, ha a szórást ismeretlennek tekintjük?

c) Az oktató ezek után több gyakorló feladatot ad, majd a 10 hallgató 'éles' vizsgán elért eredményeit összehasonlítja a próbavizsgán elértekkel. Ennek alapján állíthatjuk-e, hogy az évfolyam teljesítménye javult a gyakorló feladatoknak köszönhetően?

d) Végezetül mit mondhatunk, ez az évfolyam valóban gyengébben teljesít, mint a korábbiak?

A próba,- illetve az „éles" vizsga eredményeit az alábbi táblázat tartalmazza.

Próba vizsga	45	58	57	25	31	43	36	13	30	32
„Éles" vizsga	51	56	60	35	28	48	41	12	33	30

Megoldás:

a) A próbavizsga eredményei alapján azt a hipotézist akarjuk ellenőrizni, hogy a jelen évfolyam hallgatói által a valószínűségszámítás vizsgán elért pontszám várható értéke kevesebb, mint $m_0 = 45$ pont (ellenhipotézis). Mivel ismert a szórás ($\sigma = 14$), a kis ($n = 10$) elemű minta ellenére egymintás u-próbát alkalmazunk. A próbavizsga eredményeit felhasználva a mintaátlag $\hat{m}_0 = 37$.

1. A null- és ellenhipotézis.

$$H_0\colon E(X) \geq 45$$

$$H_1\colon E(X) < 45$$

2. u-próbát alkalmazunk.

$$u_p = \frac{\hat{m}_n - m_0}{\frac{\sigma}{\sqrt{n}}} = \frac{37 - 45}{\frac{14}{\sqrt{10}}} = -1{,}8070.$$

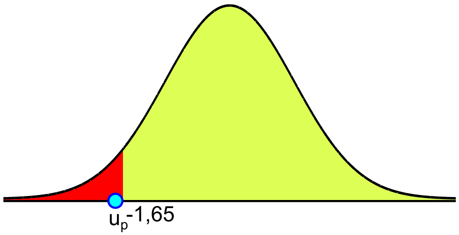
3. A kritikus érték meghatározása.

A próba egyoldali, $p = 0{,}95$ szignifikanciaszinttel.

$$\Phi(u_{kr}) = 0{,}95,$$

$$u_{kr} = -1{,}65.$$

Ennek alapján az alábbi ábrán látható a kritikus, illetve az elfogadási tartomány, melyben elhelyeztük a próbastatisztika értékét is (kék pont).



25.6. ábra. *Kritikus tartomány kijelölése a 25.6.a) feladathoz.*

4. A döntés.
A H_0 hipotézist elvetjük, H_1-et elfogadjuk. 95%-os szignifikanciaszinten elfogadjuk, hogy ez az évfolyam gyengébben teljesít a korábbiaknál.

b) Ha nem fogadjuk el a korábbi évek tapasztalata alapján a szórás értékét 14-nek, akkor a kis mintaelemszám alapján t-próbát kell alkalmaznunk. A mintából $\hat{s_n} = 14{,}0317$.

1. A null- és ellenhipotézis.

$$H_0: E(X) \geq 45$$

$$H_1: E(X) < 45$$

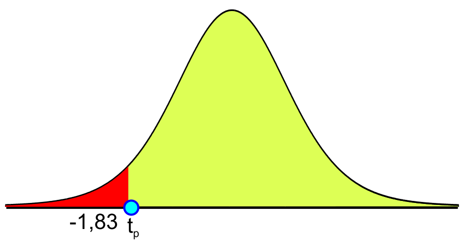
2. t-próbát alkalmazunk.

$$t_p = \frac{\hat{m}_n - m_0}{\frac{\hat{s_n}}{\sqrt{n}}} = \frac{37-45}{\frac{14{,}0317}{\sqrt{10}}} = -1{,}8030$$

3. A kritikus érték meghatározása.
A próba egyoldali, $p = 0{,}95$ szignifikanciaszinttel. A szabadsági fok 9.

$$t_{kr} = -1{,}83.$$

Az ábrán látható a kritikus, illetve az elfogadási tartomány, melyben a próbastatisztika értéke (kék pont) az elfogadási tartományba esik.



25.7. ábra. *Kritikus tartomány kijelölése a 25.6.b) feladathoz.*

4. A döntés.
A H_0 hipotézist elfogadjuk, H_1-et elvetjük. 95%-os szignifikanciaszinten nem fogadjuk el, hogy ez az évfolyam gyengébben teljesít a korábbiaknál.

c) Ahhoz, hogy el tudjuk dönteni, hogy történt-e javulás a gyakorló feladatok hatására, azt kell megvizsgálnunk, hogy mennyit változott a hallgatók pontszáma. Ehhez az 'éles' vizsga eredményeiből kivonjuk a próba vizsga eredményeit, és a kapott mintára alkalmazzunk t-próbát (az ismeretlen szórás miatt). A vizsga eredmények változása tehát:

Pontszám növekedése	6	−2	3	10	−3	5	5	−1	3	−2

Feltételezzük, hogy nem történt javulás, a pontszám növekedésének várható értéke ≤ 0. A mintaátlag $\hat{m_n} = 2{,}4$, a korrigált tapasztali szórás $\hat{s_n} = 4{,}2740$ (a mintából számolva).
1. A null- és ellenhipotézis.

$$H_0 : E(X) \leq 0$$

$$H_1 : E(X) > 0$$

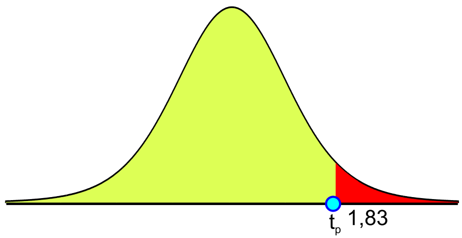
2. t-próbát alkalmazunk.

$$t_p = \frac{\hat{m_n} - m_0}{\frac{\hat{s_n}}{\sqrt{n}}} = \frac{2{,}4 - 0}{\frac{4{,}274}{\sqrt{10}}} = 1{,}78.$$

3. A kritikus érték meghatározása.
A kritikus érték megegyezik az előző feladatrészével:

$$t_{kr} = 1{,}83.$$

Az alábbi ábrán látható a kritikus tartomány.



25.8. ábra. *Kritikus tartomány kijelölése a 25.6.c) feladathoz.*

4. A döntés.
A H_0 hipotézist elfogadjuk, H_1-et elvetjük. 95%-os szignifikanciaszinten nem fogadható el, hogy a gyakorló feladatok hatására javult a teljesítmény.

d) Az eredményeink meglehetősen ellentmondásosnak tűnnek. Az a) feladatrész szerint gyengébben teljesít ez az évfolyam, és a gyakorló feladatok hatására sem mutatkozik javulás. Csupán akkor fogadjuk el, hogy mégsem gyengébb ez az évfolyam, ha a szórást ismeretlennek tekintjük (a becsléséből adódó bizonytalanság miatt a kritikus tartomány szűkebb) . Vegyük ugyanakkor észre, hogy a kritikus és a számított értékek mindegyik megoldásrészben közel vannak egymáshoz. Hogyan döntsünk? Meglepő módon a számításaink alapján azt kell mondanunk, hogy nem gyengébb ez az évfolyam a korábbiaknál. Ennek oka pedig az, hogy ha csak az 'éles' vizsga eredményét tekintjük (ahol a mintaátlag 39,4 pont, magasabb, mint a próbavizsgáé), és arra alkalmazzuk az a) és b) megoldást, akkor a nullhipotézist el kell fogadjuk.

Aktivitás: Mutassa meg, hogy ha az oktató csak a valódi vizsga eredményeit tekintette volna, akkor rögtön arra a következtetésre jut, hogy nem gyengébb ez az évfolyam a korábbiaknál!

Önellenőrzés

Önellenőrző kérdések

1. Egy normális eloszlású valószínűségi változó szórása 1,2, egy 12 elemű mintából meghatározott mintaátlaga 11,3. Azt vizsgáljuk, elfogadható-e 98%-os szignifikanciaszinten, hogy a valószínűségi változó várható értéke 11?
   (a) Milyen próbát használjunk a kérdés eldöntésére?
   ☐ egymintás u-próba ☐ egymintás t-próba
   (b) Mennyi a kritikus érték?
   ☐ 0,8660 ☐ 2,33 ☐ 2,05 ☐ 1,96
   (c) Mi a döntés? Elfogadható-e 98%-os szignifikanciaszinten, hogy a valószínűségi változó várható értéke 11?
   ☐ nem fogadható el ☐ elfogadható
2. Egy termék hossza normális eloszlást követ, 12 cm szórással. A termék névleges hossza 420 cm. A gyártósor karbantartása után 100 véletlenszerűen kiválasztott termék hosszának átlaga 423 cm. Azt vizsgáljuk, elfogadható-e 95%-os szignifikanciaszinten, hogy a termék hosszának várható értéke 420 cm?
   (a) A kérdés eldöntésére használt próba:
   ☐ egyoldali ☐ kétoldali
   (b) Mennyi a próbastatisztika mintából számított értéke?
   ☐ 1,65 ☐ 1,12 ☐ 2,50 ☐ -1,12
   (c) Mennyi a kritikus érték?
   ☐ 1,96 ☐ 2,50 ☐ 1,64 ☐ 2,05
   (d) Mi a döntés? Igaz-e 95%-os szignifikanciaszinten, hogy a termék hosszának várható értéke 420 cm?

☐ elfogadható

☐ nem fogadható el

3. Egy csavar átmérője normálist eloszlást követ. Egy 5 elemű mintát vizsgálva a csavarok átmérőjére cm-ben az alábbi eredményeket kaptuk:

1,51	1,49	1,54	1,52	1,54

Azt vizsgáljuk, elfogadható-e 95%-os szignifikanciaszinten, hogy a csavarok átmérőjének várható értéke 1,50 cm?

(a) A kérdés eldöntésére használt próba:

☐ t-próba ☐ u-próba

(b) Mennyi a kritikus érték?

☐ 1,96 ☐ 1,65 ☐ 2,78 ☐ 2,33

(c) Mi a döntés? Igaz-e 95%-os szignifikanciaszinten, hogy a csavarok átmérőjének várható értéke 1,50 cm?

☐ nem fogadható el ☐ elfogadható

4. Egy adott autótípus fogyasztási adatait vizsgálták. 10 autó alapján az átlagfogyasztás 8.1 liter/100 km. Feltételezzük, hogy az adatok normális eloszlásból származnak és a tapasztalati szórás 1,2 liter/100 km. Azt vizsgáljuk, 95 $\%$-os szignifikanciaszinten elfogadható-e, hogy az autók fogyasztásának várható értéke legalább 8 liter/100 km?

(a) A kérdés eldöntésére használt próba:

☐ t-próba ☐ u-próba

(b) Mennyi a kritikus érték?

☐ -1,83 ☐ -2,33 ☐ 1,96 ☐ 1,65

(c) Elfogadható-e 95 $\%$-os szignifikanciaszinten, hogy az autók fogyasztásának várható értéke legalább 8 liter/100 km?

☐ nem fogadható el ☐ elfogadható

5. Egy fagylaltárus által kiadott gombócok tömegét vizsgáltuk, és a következő eredményt kaptuk:

Tömeg (dkg)	4,1	4,6	4,5	3,9	4,2	4,5	3,8	4,3	3,6	3,8

Tételezzük fel, hogy a gombócok tömege normális eloszlást követ. Azt vizsgáljuk, elfogadható-e 98%-os szignifikanciaszinten, hogy a gombócok tömegének várható értéke kevesebb, mint 4 dkg?

(a) A kérdés eldöntésére használt próba:

☐ t-próba ☐ u-próba

(b) Mennyi a kritikus érték?

☐ -2,40 ☐ 1,32 ☐ 2,36 ☐ -2,05

(c) Mi a döntés? Igaz-e 98%-os szignifikanciaszinten, hogy a gombócok tömegének várható értéke kevesebb, mint 4 dkg?

☐ elfogadható ☐ nem fogadható el

6. Az alábbi minta 8 autó fogyasztási adatait tartalmazza, szerviz előtt és szerviz után(liter/100 km):

Szerviz előtt	6,8	7,4	8,1	7,6	6,4	6,9	8,0	7,3
Szerviz után	6,5	7,2	7,8	7,6	6,0	6,8	7,6	7,0

Azt vizsgáljuk, csökkentette-e a szerviz a fogyasztást?

(a) A kérdés eldöntésére melyik próbát alkalmazzuk?

☐ egymintás t-próba ☐ kétmintás t-próba

(b) Csökkentette-e a szerviz a fogyasztást?

☐ igen ☐ nem

40. LECKE

Kétmintás próbák

Gyakran a probléma nem egy mintához kapcsolódik, hanem két minta összehasonlítása a feladat. Például, ha össze szeretnénk hasonlítani két fogyasztótabletta hatását, és ennek érdekében két tesztcsoporton végzünk méréseket, akkor gyakorlatilag a két szer hatására bekövetkező fogyások várható értékeit hasonlítjuk össze, független minták alapján. Ha a minták függetlenek, kétmintás próbát alkalmazhatunk.
A továbbiakban két normális eloszlású valószínűségi változó várható értékét akarjuk összehasonlítani. Attól függően, hogy a valószínűségi változók szórása ismert-e, illetve, hogy mekkora a megfigyelésükből kapott minták nagysága, az alábbi eseteket különböztetjük meg.

25.1.3. Kétmintás u-próba

Az X és az Y **normális eloszlású**, $E(X) = m_1$ és $E(Y) = m_2$ ismeretlen várható értékű, $D(X) = \sigma_1$ és $D(Y) = \sigma_2$ szórású valószínűségi változók várható értékének különbségére vonatkozó

$$H_0\colon m_1 - m_2 = m_0$$

hipotézis helyességét akarjuk ellenőrizni.

1. Tekintsük először azt az esetet, mikor a valószínűségi változók **szórása**, azaz $D(X) = \sigma_1$ és $D(Y) = \sigma_2$ **ismert**.
   A próbastatisztika formula:
   $$u_p = \frac{\hat{m}_{n_1} - \hat{m}_{n_2} - m_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}.$$
   A próbastatisztika H_0 fennállása esetén standard normális eloszlású.
   n_1 az X-re vonatkozó, n_2 az Y-ra vonatkozó független minták elemszáma.

Megjegyzés: A fenti próbastatisztika formula egyszerűbb alakba írható, ha az **ismert szórások** egyben **azonosak**, azaz $\sigma_1 = \sigma_2 = \sigma$. Ekkor a próbastatisztika formula:

$$u_p = \frac{\hat{m}_{n_1} - \hat{m}_{n_2} - m_0}{\sigma \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}.$$

Az eloszlás természetesen továbbra is standard normális.

2. Most nézzük azt az esetet, mikor a valószínűségi változók **szórása azonos** ($\sigma_1 = \sigma_2 = \sigma$), **de** a közös szórás (σ) **nem ismert**, továbbá **mindkét minta elemszáma nagy**.
A próbastatisztika formula:

$$u_p = \frac{1}{\sqrt{\dfrac{\hat{s}_{n_1}^2 \cdot (n_1 - 1) + \hat{s}_{n_2}^2 \cdot (n_2 - 1)}{n_1 + n_2 - 2}}} \frac{\hat{m}_{n_1} - \hat{m}_{n_2} - m_0}{\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}},$$

eloszlása standard normális, amennyiben H_0 fennáll.

25.7. feladat: Két normális eloszlású valószínűségi változó várható értékét szeretnénk összehasonlítani. Az első szórása 5,48, a második szórása 6,32. Az első esetben 10 elemű mintát vizsgáltunk, a mintaátlag 40,1. A második esetben 8 elemű mintát vizsgáltunk, a mintaátlag 38,3. 95%-os szignifikanciaszinten elfogadható-e, hogy a két valószínűségi változó várható értéke megegyezik?

Megoldás: A két minta adatai (minták elemszáma, a várható értékek és szórások) a feladatból kiolvashatóak.

Első minta:	Második minta:
$n_1 = 10$	$n_2 = 8$
$\hat{m}_1 = 40{,}1$	$\hat{m}_2 = 38{,}3$
$\sigma_1 = 5{,}48$	$\sigma_2 = 6{,}32$

1. *A null- és ellenhipotézis kimondása.*
   A feltételezésünk, hogy a várható értékek megegyeznek, vagyis a várható értékek különbsége $m_0 = 0$.

   $$H_0: m_1 - m_2 = 0 \quad \text{(Feltételezzük, hogy a várható értékek megegyeznek.)}$$
   $$H_1: m_1 - m_2 \neq 0 \quad \text{(A várható értékek nem egyeznek meg.)}$$

   A próba kétoldali.

2. *A próba kiválasztása és a minta alapján a próbastatisztika értékének kiszámítása.*
   Mivel a szórások ismertek, a minták függetlenek, kétmintás u-próbát alkalmazunk. Helyettesítsük a minta adatait a próbastatisztika formulába:

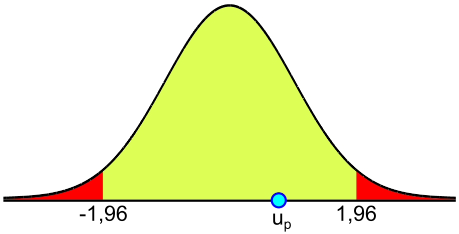
   $$u_p = \frac{\hat{m}_{n_1} - \hat{m}_{n_2} - m_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{40{,}1 - 38{,}3 - 0}{\sqrt{\frac{5{,}48^2}{10} + \frac{6{,}32^2}{8}}} = 0{,}6366.$$

3. *A kritikus érték meghatározása.*
   A próba kétoldali, így $\Phi(u_{kr}) = \dfrac{1 + 0{,}95}{2} = 0{,}975$.
   A standard normális eloszlás eloszlásfüggvényének értékeit tartalmazó táblázatból

   $$u_{kr} = 1{,}96.$$

   Ennek alapján az alábbi ábrán látható a kritikus, illetve az elfogadási tartomány, a próbastatisztika számított értékével (kék pont), amely az elfogadási tartományba esik.

4. *A döntés.*
A H_0 hipotézist elfogadjuk, 95%-os szignifikanciaszinten elfogadható, hogy a valószínűségi változók várható értéke azonos.



25.9. ábra. *Kritikus tartomány kijelölése a 25.7. feladathoz.*

25.8. feladat: 12 éves általános iskolások matematikai tudását mérték fel. Véletlenszerűen kiválasztottak 225 fiút és 250 lányt, akik ugyanazt a matematikai tesztet írták meg. A fiúk átlagosan 57 pontot értek el, a lányok 60 pontot. A szórás értéke a fiúk esetében 12, a lányok esetében 15. Normális eloszlást feltételezve, 98%-os szignifikanciaszinten állíthatjuk-e, hogy a 12 éves lányok jobbak matematikából az azonos korú fiúknál?

Megoldás: A minták adatai:

Fiúk:	Lányok:
$n_1 = 225$	$n_2 = 250$
$\hat{m}_1 = 57$	$\hat{m}_2 = 60$
$\sigma_1 = 12$	$\sigma_2 = 15$

Kövessük a statisztikai hipotézisvizsgálat szokásos lépéseit!

1. *A null- és ellenhipotézis kimondása.*
   A korábban megoldott egymintás-próbákhoz hasonlóan azt az állítást kell megfogalmazni a nullhipotézisben, amelyik tartalmazza az egyenlőséget. Ennél a feladatnál a kérdésben megfogalmazott feltevés nem ilyen (a 12 éves lányok jobban teljesítenek matematikából, mint a velük egykorú fiúk), ezért az ellenhipotézisben írjuk fel. A nullhipotézis az lesz, hogy a fiúk teljesítményének várható értéke (m_1) legalább akkora, mint a lányoké (m_2), azaz $m_1 \geq m_2$. (Ez pedig átrendezés után azt jelenti, hogy $m_1 - m_2 \geq 0$).

   $H_0 : m_1 - m_2 \geq 0$ (Feltételezzük, hogy a fiúk teljesítménye legalább akkora, mint a lányoké.)
   $H_1 : m_1 - m_2 < m_0 = 0$ (A lányok jobban teljesítenek.)

   Az ellenhipotézis alapján a próba egyoldali.

2. *A próba kiválasztása és a minta alapján a próbastatisztika értékének kiszámítása.*
   Mivel a szórások ismertek, a minták függetlenek, kétmintás u-próbát alkalmazunk. A próbastatisztika formulába helyettesítve a minták adatait, megkapjuk a számított értéket.

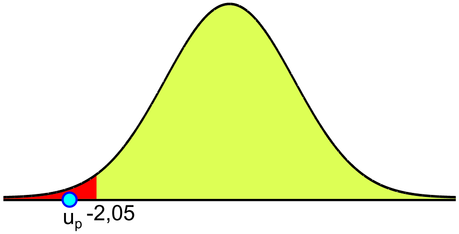
$$u_p = \frac{\hat{m}_{n_1} - \hat{m}_{n_2} - m_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{57 - 60 - 0}{\sqrt{\frac{12^2}{225} + \frac{15^2}{250}}} = -2{,}4175.$$

3. *A kritikus érték meghatározása.*
   A próba egyoldali, így $\Phi(u_{kr}) = 0{,}98$.
   A standard normális eloszlás eloszlásfüggvényének értékeit tartalmazó táblázatból

$$u_{kr} = -2{,}05.$$

   A következő ábrán látható a kritikus, illetve az elfogadási tartomány, melyben a próbastatisztika értéke (kék pont) az elutasítási tartományba esik.

4. *A döntés.*
   A H_0 hipotézist tehát elutasítjuk, ami azt jelenti, hogy a H_1 hipotézist elfogadjuk. 98%-os szignifikanciaszinten elfogadható, hogy az adott populációban a 12 éves lányok jobban teljesítenek matematikából, mint az azonos korú fiúk.



25.10. ábra. *Kritikus tartomány kijelölése a 25.8. feladathoz.*

25.9. feladat: Két golfőrült, Kovács úr és Kiss úr azt akarja összehasonlítani, hogy a golfütővel milyen messzire tudják ütni a golflabdát. Kovács úr állítja, hogy ő több, mint 4 méterrel távolabb képes ütni a golflabdát, mint Kiss úr. A vitát a golfpályán döntik el. Kovács úr 40 ütésből átlagosan 102 méterre juttatta a labdát, Kiss úr 35 ütéssel átlagosan 95 métert ért el. Tegyük fel, hogy az adatok normális eloszlásból származnak és az ütések szórása mindkét esetben azonos, 5,648 méter. 99%-os szignifikanciaszinten igazolható-e Kovács úr állítása?

Megoldás: Adataink:

Kovács úr:

$n_1 = 40$

$\hat{m}_1 = 102$

$\sigma_1 = 5{,}648 = \sigma$

Kiss úr:

$n_2 = 35$

$\hat{m}_2 = 95$

$\sigma_2 = 5{,}648 = \sigma$

1. *A null- és ellenhipotézis kimondása.*
   Azt akarjuk igazolni, hogy Kovács úr ütéshosszának várható értéke (m_1) több, mint 4 m-rel haladja meg Kiss úr ütéshosszának várható értékét (m_2-t). Ez azt jelenti, hogy $m_1 - m_2 > 4$. Ezt az állítást az ellenhipotézben fogalmazhatjuk meg, a nullhipotézis az egyenlőséget is megengedő ellentettje kell legyen. A nullhipotézis ekkor az lesz, hogy Kovács úr teljesítménye pont 4 m-rel, vagy annál kevesebbel haladja meg Kiss úrét.

   $H_0: m_1 - m_2 \leq 4$ (Kovács úr teljesítménye legfeljebb $m_0 = 4$ méterrel jobb.)
   $H_1: m_1 - m_2 > 4$ (Kovács úr teljesítménye több, mint 4 méterrel haladja meg Kiss úrét.)

   Az ellenhipotézis alapján a próba egyoldali.

2. *A próba kiválasztása és a minta alapján a próbastatisztika értékének kiszámítása.*
   Mivel a szórások ismertek és azonosak, a minták függetlenek, kétmintás u-próbát alkalmazunk.

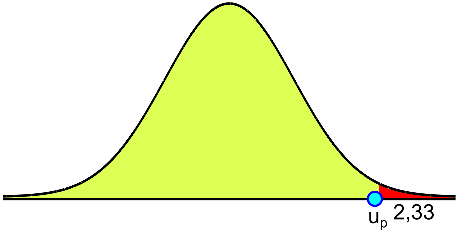
$$u_p = \frac{\hat{m}_{n_1} - \hat{m}_{n_2} - m_0}{\sigma \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{102 - 95 - 4}{5{,}648 \cdot \sqrt{\frac{1}{40} + \frac{1}{35}}} = 2{,}2949$$

3. *A kritikus érték meghatározása.*
   A próba egyoldali, így $\Phi(u_{kr}) = 0{,}99$.
   A standard normális eloszlás eloszlásfüggvényének értékeit tartalmazó táblázatból

$$u_{kr} = 2{,}33.$$

   A következő ábrán látható, hogy a számított érték az elfogadási tartományba esik.

4. *A döntés.*
   A H_0 hipotézist elfogadjuk, H_1-et így elutasítjuk. 99%-os szignifikanciaszinten nem fogadható el, hogy Kovács úr legalább 4 m-rel hosszabb ütésekre képes.



25.11. ábra. *Kritikus tartomány kijelölése a 25.9. feladathoz.*

Megjegyzés: Az előző feladatban a számított- és a tesztérték nagyon közel esik egymáshoz, ami felveti, hogy a döntésünk valóban helytálló-e. (Az elsőfajú hiba elkövetésének valószínűsége most 0,01.) Növeljük az elsőfajú hiba elkövetésének valószínűségét 0,02-re (a szignifikanciaszintet 98%-ra csökkentjük) és nézzük meg ekkor a feladat megoldását!

25.10. feladat: Hogyan változna a 25.9. feladat megoldása, ha a szignifikanciaszintet 99%-ról 98%-ra változtatjuk?

Megoldás: A szignifikanciaszint változtatása az első két lépést nem befolyásolja, a 3. lépésben a kritikus tartomány kijelölésére van hatással.
3. *A* teszt statisztika értékének meghatározása.
A próba továbbra is egyoldali, csak a szignifikanciaszint változott, így $\Phi(u_{kr}) = 0{,}98$.

$$u_{kr} = 2{,}05.$$

4. A próbastatisztika értéke most a kritikus tartományba esik, a H_0 hipotézist tehát elvetjük. 98%-os szignifikanciaszinten igazolható, hogy Kovács úr legalább 4m-rel hosszabb ütésekre képes.

25.1.4. Kétmintás t-próba

Az X és az Y normális eloszlású, $E(X) = m_1$ és $E(Y) = m_2$ ismeretlen várható értékű, ismeretlen, de azonos $D(X) = \sigma_1 = D(Y) = \sigma_2 = \sigma$ szórású valószínűségi változók várható értékének különbségére vonatkozó

$$H_0 : m_1 - m_2 = m_0$$

hipotézis helyességét akarjuk ellenőrizni.
A próbastatisztika formula:

$$t_p = \frac{1}{\sqrt{\dfrac{\hat{s}_{n_1}^2 \cdot (n_1 - 1) + \hat{s}_{n_2}^2 \cdot (n_2 - 1)}{n_1 + n_2 - 2}}} \frac{\hat{m}_{n_1} - \hat{m}_{n_2} - m_0}{\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

A próbastatisztika $(n_1 + n_2 - 2)$ szabadsági fokú Student-eloszlású H_0 fennállása esetén.
n_1 az X-re vonatkozó, n_2 az Y-ra vonatkozó független minták elemszáma.
Megjegyzés: Ha n_1 és n_2 nagy, akkor az eloszlás közel standard normálisnak tekinthető, megegyezik az u-próba megfelelő alakjával.

25.11. feladat: Tegyük fel, hogy két valószínűségi változó normális eloszlású és szórásuk egyenlő. Igaz- e 95%-os szignifikanciaszinten, hogy a két valószínűségi változó várható értéke egyenlő? A mérési adatok:

A	10,2	10,1	9,8	9,4	10,4	11,0	10,5	9,9	10,6	9,7	10,3
B	10,6	10,3	9,9	10,2	9,9	10,1	10,9	10,5	-	-	-

Megoldás: A mintákból kiszámoljuk a szükséges statisztikákat:

$$n_1 = 11$$
$$n_2 = 8$$
$$\hat{m}_1 = \frac{1}{n_1} \cdot \sum_{i=1}^{n_1} X_i = \frac{111{,}9}{11} = 10{,}1727$$
$$(\hat{s}_1)^2 = \frac{1}{n_1 - 1} \cdot \sum_{i=1}^{n_1} (X_i - \hat{m}_1)^2 = \frac{2{,}08182}{10} = 0{,}2082$$
$$\hat{m}_2 = \frac{1}{n_2} \cdot \sum_{i=1}^{n_2} Y_i = \frac{82{,}4}{8} = 10{,}3$$
$$(\hat{s}_2)^2 = \frac{1}{n_2 - 1} \cdot \sum_{i=1}^{n_2} (Y_i - \hat{m}_2)^2 = \frac{0{,}86}{7} = 0{,}1229$$

A minták adatai így az alábbiak:

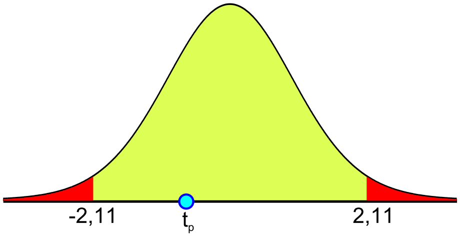
A	B
$n_1 = 11$	$n_2 = 8$
$\hat{m}_1 = 10{,}1727$	$\hat{m}_2 = 10{,}3$
$\hat{s}_1 = 0{,}4563$	$\hat{s}_2 = 0{,}3505$

1. *A null- és ellenhipotézis kimondása.*
   Feltételezzük, hogy a két valószínűségi változó várható értéke azonos ($m_1 = m_2$), így a különbségük 0 ($m_1 - m_2 = 0$). A feltevésünk nullhipotézisként mondható ki.
   $$H_0: m_1 - m_2 = 0 \quad \text{(A várható értékek egyenlőek.)}$$
   $$H_1: m_1 - m_2 \neq 0 \quad \text{(A várható értékek nem egyenlőek.)}$$
   A próba kétoldali.
2. *A próba kiválasztása és a minta alapján a próbastatisztika értékének kiszámítása.*
   Mivel a szórások nem ismertek, de azonosak, a minták elemszáma pedig kicsi, kétmintás t-próbát alkalmazunk.
   $$t_p = \frac{1}{\sqrt{\dfrac{\hat{s}^2_{n_1}\cdot(n_1-1)+\hat{s}^2_{n_2}\cdot(n_2-1)}{n_1+n_2-2}}}\frac{\hat{m}_{n_1}-\hat{m}_{n_2}-m_0}{\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} = \frac{1}{\sqrt{\dfrac{0{,}2082\cdot 10+0{,}1229\cdot 7}{11+8-2}}}\frac{10{,}1727-10{,}3-0}{\sqrt{\dfrac{1}{11}+\dfrac{1}{8}}} =$$
   $$= -0{,}6586.$$
3. *A kritikus érték meghatározása.*
   A próba kétoldali. A kritikus érték a 17 szabadsági fokú Student-eloszlás eloszlásfüggvényének inverzének értéke a $\dfrac{1+0.95}{2} = 0{,}975$ helyen.

Ahogy azt a Student-eloszlásnál korábban megtanultuk, a tesztérték a t-eloszlás táblázatból kiolvasható:

$$t_{kr} = 2{,}11$$

A következő ábrán látható, hogy a próbastatisztika értéke az elfogadási tartományba esik.



25.12. ábra. *Kritikus tartomány kijelölése a 25.11. feladathoz.*

4. *A döntés.*
   A H_0 hipotézist elfogadjuk, 95%-os szignifikanciaszinten a várható értékek egyenlősége elfogadható.

25.2. Szórások egyenlőségére vonatkozó próba

A kétmintás t-próba alkalmazásának egyik feltétele, hogy a vizsgált valószínűségi változók szórása egyenlő legyen. A konkrét feladatokban gyakran semmiféle információnk nincs a szórásokról, az adott mintákból kell eldöntenünk, hogy tekinthetjük-e a szórásokat azonosnak.

25.2.1. F-próba

Az X és az Y **normális eloszlású**, $D(X) = \sigma_1$ és $D(Y) = \sigma_2$ **ismeretlen szórású** valószínűségi változók **szórásainak egyenlőségére** vonatkozó

$$H_0\colon \sigma_1 = \sigma_2$$

hipotézis helyességét akarjuk ellenőrizni.
A próbastatisztika formula:

$$F_p = \frac{\hat{s}_1^2}{\hat{s}_2^2}.$$

A próbastatisztika $(n_1 - 1, n_2 - 1)$ szabadsági fokú F-eloszlású, ha H_0 fennáll.
n_1 az X-re vonatkozó, n_2 az Y-ra vonatkozó független minták elemszáma.

Megjegyzés: Az, hogy az $\frac{\hat{s}_1^2}{\hat{s}_2^2}$ hányados értéke 1-nél nagyobb vagy kisebb, attól függ, milyen sorrendben dolgozzuk fel a mintákat. Az F-eloszlás táblázatának használatát megkönnyítendő a számlálóba mindig annak a mintának a korrigált tapasztalati szórásnégyzete kerüljön, amelynél ez az érték nagyobb. Ekkor a hányados értéke nem lehet 1-nél kisebb. Technikailag ez azt jelenti, hogy a próbastatisztikát úgy választjuk, hogy értéke legalább 1 legyen:

$$F_p = \max\left(\frac{\hat{s}_1^2}{\hat{s}_2^2}, \frac{\hat{s}_2^2}{\hat{s}_1^2}\right) \geq 1.$$

Ennek az a következménye, hogy F-próba esetén az elfogadási tartomány mindig az egyoldali próbának megfelelő $(0, F_{kr})$ lesz.

25.12. feladat: Adott az alábbi két normális eloszlású, független minta:

A	11,9	12,1	12,8	12,2	12,5	11,9	12,5	11,8	12,4	12,9
B	12,1	12,0	12,9	12,2	12,7	12,6	12,6	12,8	12,0	13,1

95%-os szignifikanciaszinten tekinthetjük-e azonosnak a szórásokat?

Megoldás: A mintákból kiszámoljuk a szükséges statisztikákat:

$$
\begin{aligned}
n_1 &= 10 \\
n_2 &= 10 \\
\hat{m}_1 &= \frac{1}{n_1} \cdot \sum_{i=1}^{n_1} X_i = \frac{123}{10} = 12{,}3 \\
(\hat{s}_1)^2 &= \frac{1}{n_1 - 1} \cdot \sum_{i=1}^{n_1} (X_i - \hat{m}_1)^2 = \frac{1{,}32}{9} = 0{,}1467 \\
\hat{m}_2 &= \frac{1}{n_2} \cdot \sum_{i=1}^{n_2} Y_i = \frac{125}{10} = 12{,}5 \\
(\hat{s}_2)^2 &= \frac{1}{n_2 - 1} \cdot \sum_{i=1}^{n_2} (Y_i - \hat{m}_2)^2 = \frac{1{,}42}{9} = 0{,}1578
\end{aligned}
$$

1. *A nullhipotézis kimondása.*

$$H_0 : \sigma_1 = \sigma_2 \qquad (\text{ A szórások egyenlőek.})$$

   Az ellenhipotézis F-próba esetében szükségtelen, a próba egyoldali (jobb oldali kritikus tartománnyal).

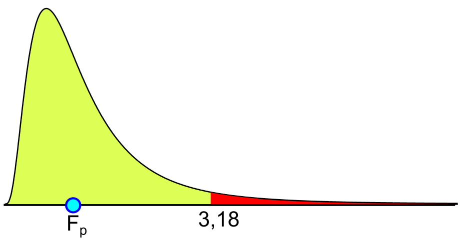
2. *A próba kiválasztása és a minta alapján a próbastatisztika értékének kiszámítása.*
   A szórások egyenlőségének vizsgálatára az F-próbát alkalmazzuk.

$$F_p = \max\left(\frac{\hat{s}_1^2}{\hat{s}_2^2}, \frac{\hat{s}_2^2}{\hat{s}_1^2}\right) = \frac{1{,}42}{1{,}32} = 1{,}0758.$$

3. *A kritikus érték meghatározása.*
   Az F-próba mindig egyoldali, tesztértéke az $(n_1 - 1, n_2 - 1) = (9{,}9)$ szabadsági fok és a szignifikanciaszint $(0{,}95)$ ismeretében az F-eloszlás táblázatból kiolvasható. (A táblázat használatát az F-eloszlásnál megmutattuk.)

$$F_{kr} = 3{,}18.$$

   Az elfogadási tartomány tehát $(0;\ 3{,}18)$.
   F_p=1,0758 beleesik ebbe a tartományba.



25.13. ábra. *Kritikus tartomány kijelölése a 25.12. feladathoz.*

4. *A döntés.*
   A H_0 hipotézist elfogadjuk. 95%-os szignifikanciaszinten elfogadható, hogy a megfigyelt valószínűségi változók szórása egyenlő.
   Megjegyzés: A szórások egyenlőségét most már elfogadhatjuk. Annak a kérdésnek az eldöntésére, hogy a megfigyelt valószínűségi változók várható értéke megegyezik-e, teljes joggal alkalmazhatjuk a kétmintás t-próbát. (Ez egy jó gyakorló feladat!)

Önellenőrzés

Önellenőrző kérdések

1. Két különböző (A és B) gyáregységben termelt azonos célt szolgáló termék selejtarányát mérték (ezrelékben). Az eredményeket az alábbi táblázat tartalmazza :

A	8,6	7,9	8,8	9,6	8,4	8,9	8,2
B	8,9	8,2	7,6	7,9	8,0	8,8	9,6

Feltételezzük, hogy az adatok normális eloszlásból származnak és a selejtarányok szórása azonos a két gyáregységben.

(a) Elfogadható-e 95%-os szignifikanciaszinten, hogy a két gyárban termelt selejtarány várható értéke megegyezik? A kérdés eldöntésére használt próba:

☐ kétmintás t-próba ☐ kétmintás u-próba

(b) Elfogadható-e 95%-os szignifikanciaszinten, hogy a két gyárban termelt selejtarány várható értéke megegyezik?

☐ nem fogadható el ☐ elfogadható

2. Két normális eloszlású valószínűségi változó várható értékét szeretnénk összehasonlítani. Az egyik szórása 2,5, a másik szórása 2,4. A első esetben 12, a második esetben 20 elemű mintát vizsgáltunk, a mintaátlagok rendre 42,5 és 44,8. Elfogadható-e 95%-os szignifikanciaszinten, hogy a két valószínűségi változó várható értéke megegyezik?

(a) A kérdés eldöntésére használt próba:

☐ kétmintás u-próba ☐ kétmintás t-próba

(b) Elfogadható-e 95%-os szignifikanciaszinten, hogy a két gyárban termelt selejtarány várható értéke megegyezik?

- [ ] nem fogadható el
- [ ] elfogadható

41. LECKE

A χ^2-próba

26. Nem-paraméteres próbák

nem-paraméteres próbák esetében a H_0 hipotézis a megfigyelt valószínűségi változó ismeretlen eloszlására vonatkozik.

26.1. A χ^2-próba

A leggyakrabban használt nem-paraméteres próba a χ^2-próba. Fontos megjegyezni, hogy míg a korábban ismertetett próbák kis és közepes minták esetében is jól használhatóak, a χ^2-próba csak nagy elemszámú minták esetében ad megbízható eredményt.
Tekintsük át a χ^2-próba három legfontosabb alkalmazási területét!

26.1.1. Illeszkedésvizsgálat

Akkor alkalmazzuk, ha azt akarjuk ellenőrizni, hogy a minta adott eloszlású populációból származik-e. A gyakorlatban ez azt jelenti, hogy azt vizsgáljuk, hogy a mintából származó gyakoriságok mennyire illeszkednek az eloszlásból származó becsült gyakoriságokra. Adott egy n elemű minta, a mintaelemeket osszuk r (páronként diszjunkt) csoportba, azaz tekintsük az $A_1, A_2, \ldots, A_r$ teljes eseményrendszert. Az A_i esemény valószínűségét p_i, gyakoriságát (hány mintaelem esik az i-edik csoportba) pedig μ_i jelölje ($i = 1,2,\ldots,r$ esetén).
Az eloszlásra vonatkozó $H_0 : P(A_i) = p_i$ $(i = 1,2,\ldots,r)$ nullhipotézist akarjuk ellenőrizni, azaz azt, hogy a megfigyelt valószínűségi változó adott eloszlású.
A próbastatisztika:

$$\chi_p^2 = \sum_{i=1}^{r} \frac{(\mu_i - n \cdot p_i)^2}{n \cdot p_i}$$

Eloszlása: $r-1$ szabadsági fokú χ^2-eloszlás, ha H_0 fennáll.
(Pontosabban $n \to \infty$ esetén $r-1$ szabadsági fokú χ^2-eloszlással közelíthető. A gyakorlatban azonban a fenti próbastatisztikát $r-1$ szabadsági fokú χ^2-eloszlásúnak tekintjük.)
A próba mindig egyoldali, jobb oldali kritikus tartománnyal. (Vegyük észre, ha a feltételezett eloszlás különbözik

a megfigyelttől, akkor a próbastatisztika értéke a formulából adódóan nagy. Minél nagyobb a számított érték, annál nagyobb ez az eltérés.)

26.1. feladat: Kockajáték közben felmerül annak gyanúja, hogy a kocka nem szabályos. Gyanúnk ellenőrzésére 300-szor feldobjuk a dobókockát. A dobások eredményét az alábbi táblázat foglalja össze:

Dobás értéke	1	2	3	4	5	6	Összes
Dobások száma	45	42	56	55	58	44	300

Vizsgáljuk meg 95%-os szignifikanciaszinten, hogy a dobókocka szabályos-e!

Megoldás: A dobott szám értékét tekintve az X valószínűségi változónak, annak eloszlása diszkrét egyenletes, ha a kocka szabályos. Ekkor azonos a dobások valószínűsége, minden esetben $p_i = \frac{1}{6}$ $(i = 1,2,\ldots,6)$. A mintaelemek száma $n = 300$ ($r = 6$ csoportba osztva), ezért minden dobásérték várt előfordulási száma (feltételezett gyakorisága) $np_i = 50$. A hipotézisvizsgálat során tehát azt kell eldöntenünk, hogy a megfigyelt valószínűségi változó (dobás értéke) a minta alapján tekinthető-e diszkrét egyenletes eloszlásúnak. Ennek alapján a próba illeszkedésvizsgálat.

Az áttekinthetőbb számoláshoz az alábbi táblázatot hozzuk létre:

Dobás értéke	1	2	3	4	5	6
μ_i	45	42	56	55	58	44
np_i	50	50	50	50	50	50
$(\mu_i - np_i)^2$	25	64	36	25	64	36

A hipotézisvizsgálat lépései a megszokottak.

1. *A nullhipotézis kimondása.*
   A feltételezésünk most az, hogy a kocka szabályos, a dobott szám értéke diszkrét egyenletes eloszlást követ (tehát minden dobásértékhez azonos valószínűség tartozik).

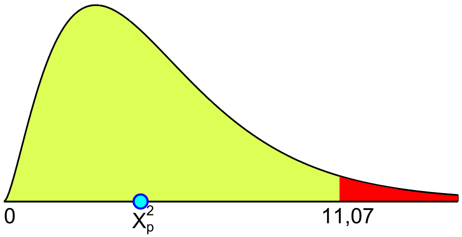
2. *A próba kiválasztása és a minta alapján a próbastatisztika értékének kiszámítása.*
   A próba illeszkedésvizsgálat, χ^2-próbával. Korábban említettük, hogy a próba mindig egyoldali (jobb oldali kritikus tartománnyal). A próbastatisztika mintából számított értéke, felhasználva a táblázat adatait:

$$\chi_p^2 = \sum_{i=1}^{6} \frac{(\mu_i - n \cdot p_i)^2}{n \cdot p_i} = \frac{25 + 64 + 36 + 25 + 64 + 36}{50} = 5$$

3. *A kritikus tartomány kijelölése.*
   A próba egyoldali χ^2-próba. Keresem azt a kritikus értéket, amely mellett az $r - 1 = 6 - 1 = 5$ szabadsági fokú χ^2-eloszlás eloszlásfüggvénye $0{,}95$ értéket vesz fel. Ez a 4. táblázat ötödik sor, első oszlopban található

$$\chi_{kr}^2 = 11{,}07$$

   érték. A kritikus tartomány ábráján a korábban megszokott jelölésekkel látható, hogy a kritikus érték az elfogadási tartományba esik.



26.1. ábra. *Kritikus tartomány kijelölése a 26.1. feladathoz.*

4. *A döntés.* $4{,}6 < 11{,}07$, a próbastatisztika értéke az elfogadási tartományba esik, a nullhipotézist elfogadjuk 95%-os szignifikanciaszinten. Ez egyben azt jelenti, hogy az adott minta alapján elfogadjuk, hogy a kocka szabályos. (Minden dobásértékhez azonos valószínűség tartozik, a dobott számot tekintve valószínűségi változónak, annak eloszlása diszkrét egyenletes.)

26.1.2. Homogenitásvizsgálat

Akkor alkalmazzuk, ha azt akarjuk eldönteni, hogy két független minta azonos eloszlásból származik-e. (Anélkül, hogy a konkrét eloszlást meghatároznánk.)
Jelölje X és Y a megfigyelt valószínűségi változókat. Az X-re vonatkozó minta elemszáma legyen m, az Y-ra vonatkozó minta elemszáma n. Soroljuk valamilyen szempont szerint r csoportba a mintaelemeket. Jelölje μ_i az X, ν_i az Y megfigyeléséből származó minta i-edik csoportba eső mintaelemeinek számát (azaz az A_i, illetve B_i események gyakoriságát). $\left(\sum\limits_{i=1}^{r} \mu_i = m, \sum\limits_{i=1}^{r} \nu_i = n\right)$.
A nullhipotézisünk az, hogy a két minta azonos eloszlásból származik.
A próbastatisztika:

$$\chi_p^2 = \frac{1}{m \cdot n} \sum_{i=1}^{r} \frac{(n \cdot \mu_i - m \cdot \nu_i)^2}{\mu_i + \nu_i}$$

Eloszlása: H_0 fennállása esetén $r-1$ szabadsági fokú χ^2-eloszlás.
A próba most is egyoldali, jobb oldali kritikus tartománnyal. (Vegyük észre, ha a két minta eloszlása különböző, akkor a próbastatisztika értéke a formulából adódóan nagy. Minél nagyobb ez a számított érték, annál nagyobb ez az eltérés.)

26.2. feladat: Egy tejüzemben különböző technológiával csomagolják a fél-, illetve egyliteres dobozos tejeket. A csomagolásokra érkező minőségi kifogások miatt vizsgálat indult a cégnél. Mindkét kiszerelésű, hibás csomagolású termékből mintát vettek, majd a hibák jellege szerint csoportosították azokat.

Hiba jellege	Gyenge ragasztás	Leszakadó fül	Hiányzó fül	Kiszakadó zárókupak	Egyéb	Összes
Félliteres kiszerelés	51	64	26	18	21	180
Literes kiszerelés	72	51	33	23	21	200

95%-os szignifikanciaszinten igazolható-e, hogy a hibák aránya megegyezik a kétfajta kiszerelés esetén?

Megoldás: Azt akarjuk eldönteni, hogy a két minta eloszlása azonos-e (ebben az esetben lesz a hibákhoz tartozó valószínűség azonos). A minták függetlenek, ezért ezt homogenitásvizsgálattal dönthetjük el.

1. *A nullhipotézis kimondása.*
   A feltételezésünk most az, hogy az egyes hibák aránya azonos a kétfajta kiszerelésnél, a nullhipotézis tehát, hogy a hibák típusa azonos eloszlású a két mintában.

2. *A próba kiválasztása és a minta alapján a próbastatisztika értékének kiszámítása.*
   A próba homogenitásvizsgálat, χ^2-próbával. A próba továbbra is egyoldali (jobb oldali kritikus tartománnyal), ahogy korábban megállapítottuk. A számításhoz alkalmazzuk a homogenitásvizsgálatnál használt jelöléseinket az adott mintára. A mintaelemeket a hibák típusától függően $r = 5$ csoportba osztották. A félliteres csomagoláshoz tartozó mintaelemek száma $m = 180$, az egyliteres csomagoláshoz tartozóké $n = 200$. Az egyes csoportokba tartozó mintaelemek száma a táblázatból kiolvasható, μ_i értékei a második sorból, ν_i értékei a harmadik sorból.

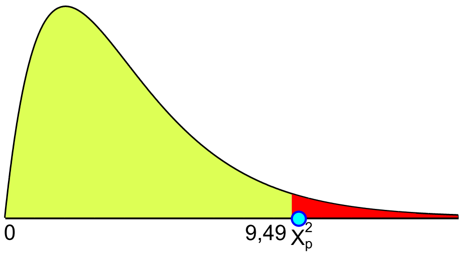
A próbastatisztika mintából számított értéke:

$$\chi_p^2 = \frac{1}{180 \cdot 200} \sum_{i=1}^{5} \frac{(180 \cdot \mu_i - 200 \cdot \nu_i)^2}{\mu_i + \nu_i} \chi_p^2 =$$
$$= \frac{1}{180 \cdot 200} \Big(\frac{(180 \cdot 51 - 200 \cdot 72)^2}{51 + 72} + \frac{(180 \cdot 64 - 200 \cdot 51)^2}{64 + 51} +$$
$$+ \frac{(180 \cdot 26 - 200 \cdot 33)^2}{26 + 33} + \frac{(180 \cdot 18 - 200 \cdot 23)^2}{18 + 23} + \frac{(180 \cdot 21 - 200 \cdot 21)^2}{21 + 21} \Big) =$$
$$= 9{,}68.$$

3. *A kritikus tartomány kijelölése.*
A próba egyoldali χ^2-próba. Az előző feladatnak megfelelően keressük azt a kritikus értéket, amely mellett az $r - 1 = 5 - 1 = 4$ szabadsági fokú χ^2-eloszlás eloszlásfüggvénye $0{,}95$ értéket vesz fel. Ez a 4. táblázat negyedik sor, első oszlopban található

$$\chi_{kr}^2 = 9{,}49$$

érték. A kritikus tartomány az alábbi ábrán látható.



26.2. ábra. *Kritikus tartomány kijelölése a 26.2. feladathoz.*

4. *A döntés.*
   $9{,}49 < 9{,}68$, a próbastatisztika számított értéke az elutasítási tartományba esik, a nullhipotézist 95%-os szignifikanciaszinten nem tudjuk elfogadni. Ez egyben azt jelenti, hogy az adott minta alapján nem teljesül a két eloszlás azonossága.

26.3. feladat: Tekintsük a 26.2. feladatot, de változtassuk meg a szignifikanciaszintet 98%-ra.

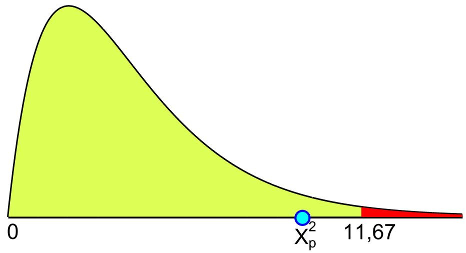
Megoldás: Vegyük észre, hogy a szignifikanciaszint megváltoztatásával a feladat megoldásában csak a kritikus érték meghatározása és a döntés változik meg.

1. Nullhipotézis kimondása. (Lásd. 26.2. feladat megoldása.)
2. Próbastatisztika értéke: $\chi_p^2 = 9{,}68$. (Lásd. 26.2. feladat megoldása.)
3. Kritikus tartomány kijelölése.
   Most azt a kritikus értéket keresem, amelynél az $r - 1 = 5 - 1 = 4$ szabadsági fokú χ^2-eloszlás eloszlásfüggvénye $0{,}98$ értéket vesz fel. Ez a 4. táblázat negyedik sor, második oszlopban található

   $$\chi_{kr}^2 = 11{,}67$$

   érték.

A kritikus tartomány az alábbi ábrán látható.



26.3. ábra. *Kritikus tartomány kijelölése a 26.3. feladathoz.*

4. lépés A döntés.
   $9{,}68 < 11{,}67$, a próbastatisztika számított értéke az elfogadási tartományba esik, a nullhipotézist 98%-os szignifikanciaszinten elfogadjuk. Ez egyben azt jelenti, hogy az adott minta alapján elfogadjuk, hogy a két eloszlás azonos.

26.1.3. Függetlenségvizsgálat

Azt akarjuk eldönteni, hogy két valószínűségi változó, X és Y függetlenek-e. Az (X,Y) valószínűségi változó megfigyelésére n elemű mintát veszünk. Az X valószínűségi változó értékeit r, Y értékeit s csoportba soroljuk, azaz létrehozzuk X-re vonatkozóan az $A_1,A_2,\ldots,A_r$, Y-ra vonatkozóan a $B_1,B_2,\ldots,B_s$ teljes eseményrendszereket. Jelölje μ_{i*} az A_i $(i = 1,2,\ldots,r)$, ν_{*j} a B_j $(j = 1,2,\ldots,s)$ bekövetkezésének gyakoriságát, μ_{ij} pedig az (A_i,B_j) együttes bekövetkezésének gyakoriságát. Természetesen ekkor teljesül, $\sum\limits_{i=1}^{r}\sum\limits_{j=1}^{s}\mu_{ij} = 1$,

valamint, hogy $\sum_{i=1}^{r} \mu_{ij} = \nu_{*j}$, és $\sum_{j=1}^{s} \nu_{ij} = \mu_{i*}$. (Ez utóbbiakhoz idézzük fel a valószínűségi vektorváltozók peremeloszlásait diszkrét esetben!)
A feltevésünk, hogy a két valószínűségi változó független, vagyis $(A_1,A_2,\ldots,A_r)$ és $(B_1,B_2,\ldots,B_s)$ teljes eseményrendszerek függetlenek. A nullhipotézis formálisan az, hogy $P(A_i \cdot B_j) = P(A_i) \cdot P(B_j)$ $(i = 1,2,\ldots,r,$ $j = 1,2,\ldots,s$ esetén).
A próbastatisztika:

$$\chi_p^2 = \frac{1}{n} \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(n \cdot \mu_{ij} - \mu_{i*} \cdot \nu_{*j})^2}{\mu_{i*} \cdot \nu_{*j}}$$

Eloszlása: $(r-1)\cdot(s-1)$ szabadsági fokú χ^2-eloszlás, amennyiben H_0 fennáll.
A próba egyoldali, jobb oldali kritikus tartománnyal, hasonlóan az illeszkedés,- illetve a homogenitásvizsgálathoz.

26.4. feladat: 3 televíziós csatorna nézettségének életkor szerinti eloszlását vizsgálták. Ehhez 200 különböző életkorú, televíziót néző embert kérdeztek meg arról, hogy egy átlagos héten melyik csatornát nézték legtöbbet. A válaszokat az alábbi táblázat tartalmazza, az életkor szerinti eloszlással.

	0-18 éves	18-49 éves	49 év feletti
A csatorna	24	13	42
B csatorna	12	8	23
C csatorna	17	31	30

Van-e összefüggés a televíziós csatorna választása és az életkor között?

Megoldás: A feladat függetlenségvizsgálat, mivel azt akarjuk eldönteni, függ-e a televíziós csatorna választása az életkortól, vagy attól független.

1. *A nullhipotézis kimondása.*
   A feltételezésünk, egyben a nullhipotézis most az, hogy az életkor és a csatornaválasztás független.
2. *A próba kiválasztása és a minta alapján a próbastatisztika értékének kiszámítása.*
   A próba függetlenségvizsgálat, χ^2-próbával, ami egyoldali (jobb oldali kritikus tartománnyal). A számításhoz a fenti táblázatot egészítsük ki az alábbiak szerint.

	0-18 éves	18-49 éves	49 év feletti	Összesen
A csatorna	24	13	42	79
B csatorna	12	8	23	43
C csatorna	17	31	30	78
Összesen	53	52	95	200

A számításhoz alkalmazzuk a homogenitásvizsgálatnál használt jelöléseinket az adott mintára. Mind a tv-csatornák, mind az életkor szerint $r = 3$, $s = 3$ csoportba osztották az adatainkat. A mintaelemek száma $n = 200$. μ_{i*} értékei rendre megadják, hogy az egyes (i-edik) csatornákat hányan részesítik előnyben, azaz értékei az utolsó oszlopból leolvashatók. ν_{*j} értékei azt adják meg, hogy az egyes korosztályokból hány embert kérdeztek meg, azaz értékei az utolsó sorból leolvashatók. μ_{ij} értékei rendre az i-edik sor és j-edik oszlop elemével egyeznek meg. Például $\mu_{1*} = 79$, $\nu_{*2} = 52$, $\mu_{2,3} = 23$.

A próbastatisztika mintából számított értéke tehát:

$$\chi_p^2 = \frac{1}{200} \sum_{i=1}^{3} \sum_{j=1}^{3} \frac{(200 \cdot \mu_{ij} - \mu_{i*} \cdot \nu_{*j})^2}{\mu_{i*} \cdot \nu_{*j}} =$$

$$= \frac{1}{200} \Bigg(\frac{(200 \cdot 24 - 79 \cdot 53)^2}{79 \cdot 53} + \frac{(200 \cdot 13 - 79 \cdot 52)^2}{79 \cdot 52} + \frac{(200 \cdot 42 - 79 \cdot 95)^2}{79 \cdot 95} +$$

$$+ \frac{(200 \cdot 12 - 43 \cdot 53)^2}{43 \cdot 53} + \frac{(200 \cdot 8 - 43 \cdot 52)^2}{43 \cdot 52} + \frac{(200 \cdot 23 - 43 \cdot 95)^2}{43 \cdot 95} +$$

$$+ \frac{(200 \cdot 17 - 78 \cdot 53)^2}{78 \cdot 53} + \frac{(200 \cdot 31 - 78 \cdot 52)^2}{78 \cdot 52} + \frac{(200 \cdot 30 - 78 \cdot 95)^2}{78 \cdot 95} \Bigg) =$$

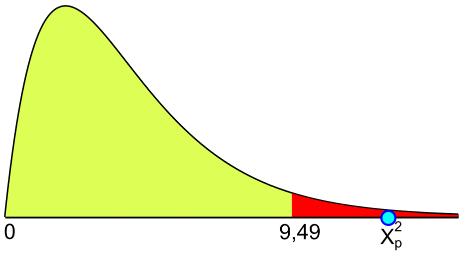
$$= 12{,}67.$$

3. *A kritikus tartomány kijelölése.*
   Nincs a feladatban megadva szignifikanciaszint, tehát az általánosan elfogadott 95%-kal számolunk.
   A próba egyoldali χ^2-próba. A kritikus tartomány kijelöléséhez keresem azt a kritikus értéket, amely mellett az $(r-1) \cdot (s-1) = (3-1) \cdot (3-1) = 4$ szabadsági fokú χ^2-eloszlás eloszlásfüggvénye $0{,}95$ értéket vesz fel. Ez a 4. táblázat negyedik sor, első oszlopban található

$$\chi_{kr}^2 = 9{,}49$$

   érték.

A kritikus tartomány az alábbi ábrán látható.



26.4. ábra. *Kritikus tartomány kijelölése a 26.4. feladathoz.*

4. *A döntés.*
   $9{,}49 < 12{,}67$, a próbastatisztika számított értéke az elutasítási tartományba esik, a nullhipotézist 95%-os szignifikanciaszinten elvetjük. Ez egyben azt jelenti, hogy az adott minta alapján nem független a tv-csatorna választás az életkortól.

Önellenőrzés

Önellenőrző kérdések

1. Mikor használunk illeszkedésvizsgálatot?
2. Mikor használunk függetlenségvizsgálatot?

3. 1000 háztartást a havi nettó jövedelme alapján három csoportba osztottak, majd megnézték, hogy milyen tv-készülék található az egyes háztartásokban. A kapott eredményeket az alábbi táblázat tartalmazza:

	Alacsony jövedelem	Közepes jövedelem	Magas jövedelem
Hagyományos TV	56	51	93
Síkképernyős TV	118	207	375
Nincs tv	26	42	32

95%-os szignifikanciaszinten igazolható-e, hogy nincs összefüggés a tv fajtája és a jövedelem között?

(a) A kérdés eldöntésére használt próba:

☐ függetlenségvizsgálat χ^2-próbával ☐ illeszkedésvizsgálat χ^2-próbával

(b) 95%-os szignifikanciaszinten igazolható-e, hogy nincs összefüggés a tv fajtája és a jövedelem között?

☐ nem igaz ☐ igaz

4. Két dobókockával egyaránt 120-szor dobtunk és feljegyeztük, hogy az egyes dobásértékek hányszor fordultak elő a két dobássorozatban.Az eredményeket az alábbi táblázatba foglaltuk.

Dobott szám	1	2	3	4	5	6
Első kocka	17	22	23	16	17	25
Második kocka	19	25	15	27	15	19

95%-os szignifikanciaszinten azonosnak tekinthető-e a két eloszlás?

(a) A kérdés eldöntésére használt próba:

☐ függetlenségvizsgálat χ^2-próbával ☐ homogenitásvizsgálat χ^2-próbával

(b) 95%-os szignifikanciaszinten azonosnak tekinthető-e a két eloszlás?

☐ igen ☐ nem

42. LECKE

Modulzáró feladatok 7.

27. Modulzáró feladatok

Start. Az alábbi kérdések megválaszolására 90 perce van:

1. Hipotézisvizsgálat esetén akkor követünk el elsőfajú hibát, ha

- [ ] rosszul választjuk ki az alkalmazott próbát,
- [ ] elfogadjuk a rossz nullhipotézist,
- [ ] elutasítjuk a jó nullhipotézist,
- [ ] elutasítjuk a jó ellenhipotézist.

2. Egymintás t-próba esetén milyen eloszlású a próbastatisztika, ha a minta elemszáma n?

- [ ] n szabadsági fokú Student-eloszlású.
- [ ] $(n-1)$ szabadsági fokú Student-eloszlású.
- [ ] Normális eloszlású.
- [ ] Standard normális eloszlású.

3. Egy termék élettartamának vizsgálatára 200 mérést végeztek. A mérési adatok alapján $\hat{m}_{200} = 984{,}6$ óra és $\hat{s}_{200} = 72{,}1$ óra. Elfogadható-e 98%-os szignifikanciaszinten, hogy a termék élettartamának várható értéke 1000 óra?

- [ ] Egymintás u-próba alapján elfogadható.
- [ ] Egymintás u-próba alapján nem fogadható el.
- [ ] Egymintás t-próba alapján elfogadható.
- [ ] Egymintás t-próba alapján nem fogadható el.

4. Egy normális eloszlású valószínűségi változó megfigyelésére $n = 8$ mérést végeztek. Az eredményeket az alábbi táblázat tartalmazza:

21	28	17	19	24	16	22	27

Ha a szórás értéke $\sigma = 4$, 95%-os szignifikanciaszinten elfogadható-e, hogy a valószínűségi változó várható értéke több, mint 20?

- ☐ Egymintás u-próba alapján elfogadható.
- ☐ Egymintás u-próba alapján nem fogadható el.
- ☐ Egymintás t-próba alapján elfogaható.
- ☐ Egymintás t-próba alapján nem fogadható el.

5. Egy csavar átmérője normális eloszlású valószínűségi változó. 15 véletlenszerűen kiválasztott csavar átmérőjét megmértük. A kapott eredmények (mm): $\hat{m}_{15} = 18{,}6$, $\hat{s}_{15} = 1{,}9$.
90%-os szignifikanciaszinten elfogadható-e, hogy a csavar átmérőjének várható értéke 20 mm?

- ☐ Egymintás u-próba alapján elfogadható.
- ☐ Egymintás u-próba alapján nem fogadható el.
- ☐ Egymintás t-próba alapján elfogadható.
- ☐ Egymintás t-próba alapján nem fogadható el.

6. Egy üzemben deszkákat készítenek, a deszkák hossza normális eloszlást követ, 3 cm szórással. Az üzemben 18 elemű mintát veszünk. A mintaátlag 1,99 méter.
95%-os szignifikanciaszinten elfogadható-e, hogy a deszkák hosszának várható értéke kevesebb, mint 2 méter?

- [ ] Egymintás u-próba alapján elfogadható.
- [ ] Egymintás u-próba alapján nem fogadható el.
- [ ] Egymintás t-próba alapján elfogadható.
- [ ] Egymintás t-próba alapján nem fogadható el.

7. Egy új edzésmódszert teszteltek gyanútlan hosszútávfutókon. A kísérlet elején két 150 fős csoportba osztották a résztvevőket. Az egyik (A) csoportot a hagyományos, a másik (B) csoportot az új módszer szerint edzették. A kísérlet végén mindkét csoportot egy 10 km-es futással tesztelték. Tegyük fel, hogy mindkét csoport eredményei normális eloszlásúak, a szórások pedig $\sigma_A = 1{,}22$, $\sigma_B = 1{,}31$. A tesztfutáson elért eredmények (percben) $m_A = 32{,}10$, $m_B = 31{,}34$. Elfogadható-e 95%-os szignifikanciaszinten, hogy az új módszer javította az eredményeket?

- [ ] Kétmintás u-próba alapján elfogadható.
- [ ] Kétmintás u-próba alapján nem fogadható el.
- [ ] Kétmintás t-próba alapján elfogadható.
- [ ] Kétmintás t-próba alapján elfogadható.

8. Egy területi bajnokságban játszó futbalcsapat 100 mérkőzésén elért góljainak számát tartalmazza az alábbi táblázat:

Rúgott gólok száma	0	1	2	3	4	5	6	7
Mérkőzések száma	18	14	29	18	10	5	5	1

Milyen próbát alkamazna annak eldöntésére, hogy a csapat által rúgott gólok száma Poisson-eloszlást követ-e?

☐ Homogenitásvizsgálat χ^2-próbával.
☐ Illeszkedésvizsgálat χ^2-próbával.
☐ Függetlenségvizsgálat χ^2-próbával.

9. Egy nagyvállalatnál felmérést készítettek arról, hogy a munkából teljes napot hiányzó munkavállalók a hét mely napjain hiányoznak. Ehhez megnézték, hogy egy adott időszakban 500 távollét milyen napra esett. A kapott eredményeket az alábbi táblázat tartalmazza:

Hét napja	Hétfő	Kedd	Szerda	Csütörtök	Péntek	Összesen
Hiányzások száma	121	87	87	91	114	500

95%-os szignifikanciaszinten elfogadható-e, hogy a hiányzások a hét napjain egyenletesen oszlanak el?

☐ Nem fogadható el.
☐ Elfogadható.

Stop.

Pontok:

43. LECKE

Táblázatok

1. táblázat. *A standard normális eloszlásfüggvény ($\Phi(x)$) értékei I. (x=0... 1,99)*

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767

2. táblázat. *A standard normális eloszlásfüggvény ($\Phi(x)$) értékei II. (x=2...3,99)*

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

3. táblázat. *A Student-eloszlásfüggvény inverzének értékei t-próbához*

f	0,90	0,95	0,975	0,98	0,99	0,995
1	3,08	6,31	12,71	15,89	31,82	63,66
2	1,89	2,92	4,30	4,85	6,96	9,92
3	1,64	2,35	3,18	3,48	4,54	5,84
4	1,53	2,13	2,78	3,00	3,75	4,60
5	1,48	2,02	2,57	2,76	3,36	4,03
6	1,44	1,94	2,45	2,61	3,14	3,71
7	1,41	1,89	2,36	2,52	3,00	3,50
8	1,40	1,86	2,31	2,45	2,90	3,36
9	1,38	1,83	2,26	2,40	2,82	3,25
10	1,37	1,81	2,23	2,36	2,76	3,17
11	1,36	1,80	2,20	2,33	2,72	3,11
12	1,36	1,78	2,18	2,30	2,68	3,05
13	1,35	1,77	2,16	2,28	2,65	3,01
14	1,35	1,76	2,14	2,26	2,62	2,98
15	1,34	1,75	2,13	2,25	2,60	2,95
16	1,34	1,75	2,12	2,24	2,58	2,92
17	1,33	1,74	2,11	2,22	2,57	2,90
18	1,33	1,73	2,10	2,21	2,55	2,88
19	1,33	1,73	2,09	2,20	2,54	2,86
20	1,33	1,72	2,09	2,20	2,53	2,85

f	0,90	0,95	0,975	0,98	0,99	0,995
21	1,32	1,72	2,08	2,19	2,52	2,83
22	1,32	1,72	2,07	2,18	2,51	2,82
23	1,32	1,71	2,07	2,18	2,50	2,81
24	1,32	1,71	2,06	2,17	2,49	2,80
25	1,32	1,71	2,06	2,17	2,49	2,79
26	1,31	1,71	2,06	2,16	2,48	2,78
27	1,31	1,70	2,05	2,16	2,47	2,77
28	1,31	1,70	2,05	2,15	2,47	2,76
29	1,31	1,70	2,05	2,15	2,46	2,76
30	1,31	1,70	2,04	2,15	2,46	2,75
35	1,31	1,69	2,03	2,13	2,44	2,72
40	1,30	1,68	2,02	2,12	2,42	2,70
45	1,30	1,68	2,01	2,12	2,41	2,69
50	1,30	1,68	2,01	2,11	2,40	2,68
60	1,30	1,67	2,00	2,10	2,39	2,66
70	1,29	1,67	1,99	2,09	2,38	2,65
80	1,29	1,66	1,99	2,09	2,37	2,64
90	1,29	1,66	1,99	2,08	2,37	2,63
100	1,29	1,66	1,98	2,08	2,36	2,63
200	1,29	1,65	1,97	2,07	2,35	2,60

4. táblázat. *A χ^2-eloszlásfüggvény inverzének néhány értéke*

f	0,95	0,98	0,99
1	3,84	5,41	6,63
2	5,99	7,82	9,21
3	7,81	9,84	11,34
4	9,49	11,67	13,28
5	11,07	13,39	15,09
6	12,59	15,03	16,81
7	14,07	16,62	18,48
8	15,51	18,17	20,09
9	16,92	19,68	21,67
10	18,31	21,16	23,21
11	19,68	22,62	24,72
12	21,03	24,05	26,22
13	22,36	25,47	27,69
14	23,68	26,87	29,14
15	25,00	28,26	30,58
16	26,30	29,63	32,00
17	27,59	31,00	33,41
18	28,87	32,35	34,81
19	30,14	33,69	36,19
20	31,41	35,02	37,57

f	0,95	0,98	0,99
21	32,67	36,34	38,93
22	33,92	37,66	40,29
23	35,17	38,97	41,64
24	36,42	40,27	42,98
25	37,65	41,57	44,31
26	38,89	42,86	45,64
27	40,11	44,14	46,96
28	41,34	45,42	48,28
29	42,56	46,69	49,59
30	43,77	47,96	50,89
31	44,99	49,23	52,19
32	46,19	50,49	53,49
33	47,40	51,74	54,78
34	48,60	53,00	56,06
35	49,80	54,24	57,34
36	51,00	55,49	58,62
37	52,19	56,73	59,89
38	53,38	57,97	61,16
39	54,57	59,20	62,43
40	55,76	60,44	63,69

f	0,95	0,98	0,99
50	67,50	72,61	76,15
60	79,08	84,58	88,38
70	90,53	96,39	100,43
80	101,88	108,07	112,33
90	113,15	119,65	124,12
100	124,34	131,14	135,81
110	135,48	142,56	147,41
120	146,57	153,92	158,95
130	157,61	165,22	170,42
140	168,61	176,47	181,84
150	179,58	187,68	193,21
160	190,52	198,85	204,53
170	201,42	209,98	215,81
180	212,30	221,08	227,06
190	223,16	232,15	238,27
200	233,99	243,19	249,45
300	341,40	352,42	359,91
400	447,63	460,21	468,72
500	553,13	567,07	576,49
1000	1074,68	1093,98	1106,97

5. táblázat. *Az F-eloszlásfüggvény inverzének néhány értéke $p = 0{,}95$ esetén*

	A számláló szabadsági foka (f_1)									
f_2	1	2	3	4	5	6	7	8	9	10
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35

Feladatok megoldása

7.1. tétel bizonyítása:

1. Ha X diszkrét valószínűségi változó, akkor

$$E(Y)=\sum_i y_i\cdot p_i=\sum_i\left(a\cdot x_i^2+b\cdot x_i+c\right)\cdot p_i=$$
$$=a\cdot\sum_i x_i^2\cdot p_i+b\cdot\sum_i x_i\cdot p_i+c\cdot\sum_i p_i=a\cdot E(X^2)+b\cdot E(X)+c\,.$$

2. Ha X folytonos eloszlású valószínűségi változó, akkor

$$E(Y)=\int\limits_{-\infty}^{\infty}y(x)\cdot f(x)\,\mathrm{d}x=\int\limits_{-\infty}^{\infty}\left(a\cdot x^2+b\cdot x+c\right)\cdot f(x)\,\mathrm{d}x=$$
$$=a\cdot\int_{-\infty}^{\infty}x^2\cdot f(x)\,\mathrm{d}x+b\cdot\int\limits_{-\infty}^{\infty}x\cdot f(x)\,\mathrm{d}x+c\cdot\int\limits_{-\infty}^{\infty}f(x)\,\mathrm{d}x=$$
$$=a\cdot E(X^2)+b\cdot E(X)+c\,.$$

$\square$

7.3. tétel bizonyítása:

$$\begin{aligned} D^2(Y) &= D^2(a\cdot X+b) = E\left[(a\cdot X+b)^2\right] - E^2\left[a\cdot X+b\right] = \\ &= E\left[a^2\cdot X^2 + 2ab\cdot X + b^2\right] - \left[a\cdot E(X)+b\right]^2 = \\ &= a^2\cdot E(X^2) + 2ab\cdot E(X) + b^2 - a^2\cdot E^2(X) - 2ab\cdot E(X) - b^2 = \\ &= a^2\cdot E(X^2) - a^2\cdot E^2(X) = a^2\cdot\left(E(X^2)-E^2(X)\right) = a^2\cdot D^2(X)\,. \end{aligned}$$

Azt kaptuk tehát, hogy $D^2(Y) = a^2\cdot D^2(X)$, amiből gyökvonással következik a tétel állítása. $\square$

10.1. tétel bizonyítása: A várható érték:

$$E(X) = \int\limits_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x = \int\limits_{a}^{b} x \cdot \frac{1}{b-a}\,\mathrm{d}x = \frac{1}{b-a} \cdot \left[\frac{x^2}{2}\right]_a^b = \frac{1}{b-a} \cdot \frac{b^2-a^2}{2} = \frac{a+b}{2}.$$

A szórás meghatározásához először ki kell számítanunk X^2 várható értékét:

$$E(X^2) = \int\limits_{-\infty}^{\infty} x^2 \cdot f(x)\,\mathrm{d}x = \int\limits_{a}^{b} x^2 \cdot \frac{1}{b-a}\,\mathrm{d}x = \frac{1}{b-a} \cdot \left[\frac{x^3}{3}\right]_a^b = \frac{1}{b-a} \cdot \frac{b^3-a^3}{3} = \frac{a^2+ab+b^2}{3}.$$

Így a szórásnégyzet:

$$D^2(X) = E(X^2) - E^2(X) = \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} =$$
$$= \frac{4a^2+4ab+4b^2}{12} - \frac{3a^2+6ab+3b^2}{12} = \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12}.$$

Tehát a valószínűségi változó szórása:

$$D(X) = \sqrt{\frac{(b-a)^2}{12}} = \frac{b-a}{\sqrt{12}} = \frac{b-a}{2\sqrt{3}}.$$

□

10.2. tétel bizonyítása: A valószínűségi változó várható értékét parciális integrálással határozhatjuk meg az $u = x \cdot \lambda$ és $v' = e^{-\lambda \cdot x}$ választással:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x = \int_{0}^{\infty} x \cdot \lambda \cdot e^{-\lambda \cdot x}\,\mathrm{d}x = \left[x \cdot \lambda \cdot \frac{e^{-\lambda \cdot x}}{-\lambda} \right]_0^{\infty} - \int_0^{\infty} \lambda \cdot \frac{e^{-\lambda \cdot x}}{-\lambda}\,\mathrm{d}x =$$

az első kifejezés határértéke ∞-ben 0 (beláthatjuk a L'Hospital-szabállyal), 0-ban a helyettesítési értéke szintén 0, így

$$= 0 - \int_{0}^{\infty} \lambda \cdot \frac{e^{-\lambda \cdot x}}{-\lambda}\,\mathrm{d}x = \int_{0}^{\infty} e^{-\lambda \cdot x}\,\mathrm{d}x = \left[\frac{e^{-\lambda \cdot x}}{-\lambda} \right]_0^{\infty} = 0 - \frac{-1}{\lambda} = \frac{1}{\lambda}\,.$$

A szórás kiszámításához X^2 várható értékét is meg kell határozni. Ez az előbbihez hasonlóan történik, azzal a különbséggel, hogy most kétszer kell parciálisan integrálni. Így a következőt kapjuk:

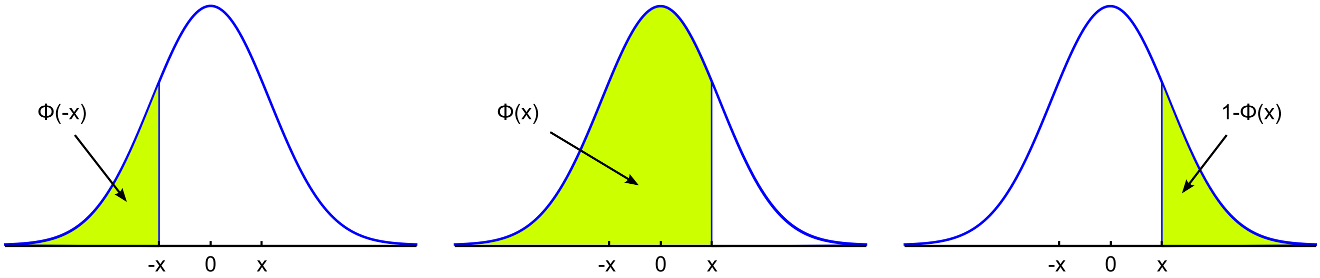
$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)\,\mathrm{d}x = \int_{0}^{\infty} x^2 \cdot \lambda \cdot e^{-\lambda \cdot x}\,\mathrm{d}x = \frac{2}{\lambda^2}\,.$$

Most már meg tudjuk határozni a szórásnégyzetet, majd abból a szórást:

$$D^2(X) = E(X^2) - E^2(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2} \qquad \Longrightarrow \qquad D(X) = \frac{1}{\lambda}\,.$$

Tehát exponenciális eloszlás esetén a várható érték és szórás egyenlő, mindkettő a λ paraméter reciproka. $\square$

10.4. tétel bizonyítása: Mivel $\Phi(x)$ egy folytonos eloszlású valószínűségi változó eloszlásfüggvénye, ezért az értéke a sűrűségfüggvény integrálásával adódik. Így $\Phi(-x)$ értéke szemléletesen a sűrűségfüggvény alatti terület a $(-\infty, -x)$ intervallumon, $\Phi(x)$ értéke pedig sűrűségfüggvény alatti terület a $(-\infty, x)$ intervallumon. Mivel a standard normális eloszlás sűrűségfüggvénye szimmetrikus a függőleges tengelyre, ezért a függvény alatti, $-x$-tól balra és x-től jobbra eső területek nagysága megegyezik. Mivel a teljes görbe alatti terület 1 (hiszen sűrűségfüggvény), ezért az utóbbi terület nagysága éppen $1 - \Phi(x)$, ami igazolja a tétel állítását.



27.1. ábra. *A standard normális eloszlás sűrűségfüggvényének szimmetriája miatt* $\Phi(-x) = 1 - \Phi(x)$.

$\square$

10.5. tétel bizonyítása: Azt fogjuk belátni, hogy X^* eloszlásfüggvényének deriváltja (azaz a sűrűségfüggvénye) a standard normális eloszlás sűrűségfüggvénye, így X^* nyilvánvalóan standard normális eloszlású. Legyen X^* eloszlásfügvénye $F^*(x)$. Az eloszlásfüggvény definíciója miatt ez a következő valószínűséggel egyenlő:

$$F^*(x) = P(X^* < x)\,.$$

Használjuk fel azt, hogy $X^* = \dfrac{X-m}{\sigma}$, és alakítsuk át a fenti formulát:

$$F^*(x) = P(X^* < x) = P\left(\frac{X-m}{\sigma} < x\right) = P(X < x\cdot\sigma + m) = F(x\cdot\sigma+m)\,.$$

Azt kaptuk, hogy X^* eloszlásfüggvénye nem más, mint X eloszlásfüggvénye az $x\cdot\sigma+m$ helyen. A sűrűségfüggvényt ennek deriválásával kapjuk. Az összetett függvény deriválási szabálya alapján:

$$(F^*(x))' = (F(x\cdot\sigma+m))' = \sigma\cdot F'(x\cdot\sigma+m) = \sigma\cdot f(x\cdot\sigma+m),$$

ahol f az X sűrűségfüggvénye, azaz

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}\,.$$

Elvégezve a műveleteket:

$$(F^*(x))' = \sigma\cdot f(x\cdot\sigma+m) = \sigma\cdot\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x\cdot\sigma+m-m)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,.$$

Tehát X^* sűrűségfüggvénye a standard normális eloszlás sűrűségfüggvénye, vagyis X^* valóban standard normális eloszlású. □

11.1. tétel bizonyítása: Először alakítsuk át a hipergeometriai kifejezést (az egyszerűség kedvéért s_m helyett s-et írunk):

$$\frac{\binom{s}{k}\cdot\binom{m-s}{n-k}}{\binom{m}{n}} = \frac{\frac{s!}{k!\cdot(s-k)!}\cdot\frac{(m-s)!}{(n-k)!\cdot(m-s-(n-k))!}}{\frac{m!}{n!\cdot(m-n)!}} =$$

$$= \frac{n!}{k!\cdot(n-k)!}\cdot\frac{s\cdot(s-1)\cdot\ldots\cdot(s-(k-1))\cdot(m-s)\cdot(m-s-1)\cdot\ldots\cdot(m-s-(n-k-1))}{m\cdot(m-1)\cdot\ldots\cdot(m-(n-1))}$$

Vegyük észre, hogy a hosszú tört számlálója és nevezője is összesen n tényezőből áll

$$= \binom{n}{k}\cdot\frac{s}{m}\cdot\frac{s-1}{m-1}\cdot\ldots\cdot\frac{s-(k-1)}{m-(k-1)}\cdot\frac{m-s}{m-k}\cdot\frac{m-s-1}{m-(k+1)}\cdot\ldots\cdot\frac{m-s-(n-k-1)}{m-(n-1)}.$$

Nézzük meg az egyes tényezők határértékét:

- $\lim\limits_{m\to\infty}\frac{s-\text{konstans}}{m-\text{konstans}} = \lim\limits_{m\to\infty}\frac{s}{m} = p.$
- $\lim\limits_{m\to\infty}\frac{m-s-\text{konstans}}{m-\text{konstans}} = \lim\limits_{m\to\infty}\frac{m-s}{m} = 1-p.$

Az első típusból k darab, a másodikból pedig $n-k$ darab van, így a teljes kifejezés határértéke:

$$\lim_{m\to\infty}\frac{\binom{s}{k}\cdot\binom{m-s}{n-k}}{\binom{m}{n}} = \binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}.$$

$\square$

Ha m, s és $m-s$ nagy az alapsokaság n elemszámához képest, akkor a hipergeometriai eloszlás közelíthető a $p = \dfrac{s}{m}$ paraméterű binomiális eloszlással. Ekkor tehát

$$\frac{\binom{s}{k} \cdot \binom{m-s}{n-k}}{\binom{m}{n}} \approx \binom{n}{k} \cdot \left(\frac{s}{n}\right)^k \cdot \left(1 - \frac{s}{m}\right)^{n-k} = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} .$$

$\square$

11.2. tétel bizonyítása: A bizonyításhoz először át fogjuk alakítani a binomiális eloszlás kifejezését:

$$\begin{aligned}\binom{n}{k}\cdot p_n^k\cdot(1-p_n)^{n-k} &= \frac{n!}{k!\cdot(n-k)!}\cdot p_n^k\cdot(1-p_n)^{n-k} = \\ &= \frac{n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-(k-1))}{k!}\cdot p_n^k\cdot(1-p_n)^{n-k}\,.\end{aligned}$$

Ezután meghatározzuk az egyes tényezők határértékét:

$$\lim_{n\to\infty}(1-p_n)^{n-k} = \lim_{n\to\infty}(1-p_n)^n\cdot(1-p_n)^{-k} = \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n\cdot\left(1-\frac{\lambda}{n}\right)^{-k} = e^{-\lambda}\cdot 1 = e^{-\lambda}\,.$$

$$\begin{aligned}&\lim_{n\to\infty}\frac{n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-(k-1))}{k!}\cdot p_n^k = \\ &\quad= \lim_{n\to\infty}\frac{1}{k!}\cdot\frac{n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-(k-1))}{n^k}(p_n\cdot n)^k = \\ &\quad= \lim_{n\to\infty}\frac{1}{k!}\cdot\frac{n}{n}\cdot\frac{n-1}{n}\cdot\frac{n-2}{n}\cdot\ldots\cdot\frac{n-(k-1)}{n}(p_n\cdot n)^k = \frac{1}{k!}\cdot 1\cdot 1\cdot\ldots\cdot 1\cdot\lambda^k = \frac{\lambda^k}{k!}\,.\end{aligned}$$

Azt kaptuk tehát, hogy

$$\lim_{n\to\infty}\binom{n}{k}\cdot p_n^k\cdot(1-p_n)^{n-k} = \frac{\lambda^k}{k!}\cdot e^{-\lambda}\,.$$

Tehát a fenti feltételek mellett a binomiális eloszlások sorozata Poisson-eloszláshoz tart. $\square$