# Összefoglaló valószínűségszámításból a Gépészmérnök Msc szak hallgatói számára

*Matematikai alapszöveg:*
**Bálint Péter**, BME Differenciálegyenletek Tanszék

*Konzultáció, kiegészítések gépészmérnöki szempontok szerint:*
**Halász Gábor**, BME Hidrodinamikai Rendszerek Tanszék

*Ábrák szerkesztése:*
**Varjú Tamás**, BME Differenciálegyenletek Tanszék

2011. február

**Kivonat**

Ez az összefoglaló a BME gépészmérnöki MSc szak hallgatói számára készül, a matematika tárgy valószínűségszámítás részéhez segédletként. Egyrészt igyekszünk – a terjedelmi és időbeli korlátok ellenére – egy matematikailag pontos, ugyanakkor mérnökök számára is szemléletes tárgyalást kialakítani. Ennek megfelelően a matematikai elmélet ismertetését helyenként mérnöki magyarázatokkal egészítjük ki. Másrészt célunk az ismeretek alkalmazása a gyakorlatban, a problémamegoldási készség fejlesztése. Így minden fejezet három alfejezetből áll; az elméleti összefoglalást (melyet esetenként tovább tagolunk) kidolgozott mintapéldák követik, végül pedig minden témakörhöz megadunk gyakorló feladatokat, melyek numerikus megoldását a függelékben lehet megtalálni.

# 1. Valószínűségi mező, feltételes valószínűség, függetlenség

## 1.1. Elméleti összefoglaló

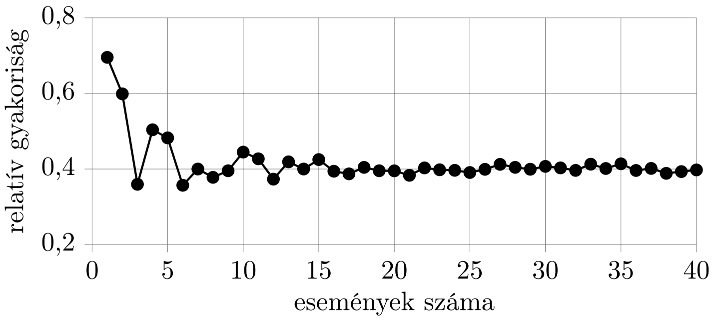
### 1.1.1. Véletlen és valószínűség

**Mérnöki bevezető a véletlen fogalmához.** A mérnöki tudás nagy része a különböző változók közötti kapcsolatok ismeretét jelenti, és a változók többnyire a klasszikus newtoni fizika változói. A közöttük lévő kapcsolat egy része mérnöki közelítésben determinisztikus kapcsolat, abban az értelemben, hogy egy összefüggés "bemenő" változói egyértelműen meghatározzák a "kimenő" változók értékeit. Például, ha ismerjük egy belső égésű motor fordulatszámát és a "gázpedál" helyzetét, akkor a motor jelleggörbéje egyértelműen meghatározza a motor nyomatékát és teljesítményét. A változók közötti kapcsolatok másik részében a bemenő változók nem határozzák meg ilyen egyértelműen a kimenő változó értékét, ugyan ahhoz a bemenő változóhoz esetről esetre különböző kimenő változó értéket tapasztalhatunk. Ez azért lehetséges, mert a figyelembe vett bemenő változók mellett még sok más, nehezen vagy alig figyelembe vehető hatás is befolyásolja a kimenő változó értékét. Ezek a hatások esetről esetre változhatnak. Tudjuk például, hogy a villamost szigorú menetrend szerint indítják a végállomásról, de a 3-4. megálló után már jelentős eltérést is tapasztalhatunk a beérkezés tényleges és a menetrend szerinti időpontja között. Nyilvánvaló, hogy a többi jármű forgalma, a le- és felszálló utasok száma, a jármű vezetőjének pillanatnyi vezetési "stílusa", az időjárás, stb. befolyásolja az érkezési időt, de e tényezők hatása nehezen számszerűsíthető. A figyelembe nem vett, de befolyással bíró paraméterek hatását nevezzük véletlen-nek, és mondjuk azt, hogy a villamos beérkezési idejét a menetrend és a véletlen együttesen határozzák meg. Ugyan ilyen értelemben mutat ingadozást egy automata gépsorról lekerülő darab mérete, egy termék élettartama, egy azonos körülmények között megismételt mérés eredménye is.

**Mérnöki bevezető a valószínűség fogalmához.** Tekintsük az utas számára kedvező $V$ eseménynek azt, ha a villamos a megállóba a menetrend szerinti időponthoz képest a $[0, +3]$ perc intervallumban érkezik. Tegyük fel, hogy rendszeresen, minden nap ugyanakkor közlekedünk, és megfigyeljük, hogy e kedvező $V$ esemény bekövetkezik, vagy sem. Mondjuk, hogy $n$ napon keresztül végezzük a megfigyelést, és ebből $k$ alkalommal következett be a számunkra kedvező esemény. Ez a $k$ szám az esemény bekövetkezésének gyakorisága, a $k/n$ hányados pedig a relatív gyakorisága. Ábrázoljuk grafikonban a $k/n$ relatív gyakoriságot az $n$ függvényében (lásd 1. ábra). Természetesen az ábra diszkrét pontsort mutat, csak azért kötöttük össze a pontokat, hogy a tendencia jól látható legyen. Tegyük fel, hogy a $V$ esemény bekövetkezését befolyásoló tényezők napról napra hasonlóak, akkor feltételezhetjük, hogy a relatív gyakoriság is bizonyos fokú stabilitást mutat: $n$ növekedésével $k/n$ értéke egyre kevésbé ingadozik. Ha $n$ elegendően nagy, és így az ingadozás kicsi, akkor jó mérnöki becsléssel berajzolhatunk az ábrába egy átlagot: e körül ingadozik a relatív gyakoriság. Ezt az átlagot tekintheti a mérnök a $V$ esemény $p$ valószínűségének, és úgy jelölik, hogy : $P(V) = p$. Tekintve, hogy $0 \leq k/n \leq 1$ , ezt a tulajdonságot örökli a valószínűség is: $0 \leq p \leq 1$. E magyarázat lényeges eleme, hogy hogyan definiáljuk a $V$ eseményt. Ha az automata gépsoron gyártott darab valamely méretét figyeljük meg, akkor a számunkra az a $V$ esemény kedvező, ha a gyártott darab mérete beleesik a névleges méret $\pm$ tűréshatár intervallumba. Hasonlóképpen fogalmazzuk meg a kedvező eseményt a termék élettartam vagy a mérési eredmény értékelése esetén is. A valószínűség fogalmának szemléletes bevezetése azt is sugallhatja, hogy ha $n \to \infty$, akkor $k/n \to p$. Ez a közönséges konvergencia, amelyet a soroknál ismertünk meg, a jelen esetben ilyen szigorúan nem igaz. Hogy mégis milyen értelemben beszélhetünk konvergenciáról, az a 4. fejezetben fog kiderülni.

**Alapfogalmak.** A valószínűségszámításban egy kísérlet lehetséges kimeneteleit *elemi eseményeknek* hívjuk. Az elemi események összessége az *eseménytér*, általában $\Omega$-val jelöljük. A kísérlet kimenetelével kapcsolatos kijelentéseinket $\Omega$ bizonyos részhalmazaival azonosíthatjuk – ezeket *eseményeknek* hívjuk, és nagybetűvel jelöljük – az $A$ esemény, mint részhalmaz, azokat (és csak azokat) az elemi eseményeket tartalmazza, amelyek az esemény bekövetkezését maguk után vonják.

Néhány jelölés: $\emptyset$ a lehetetlen esemény – az üres halmaz – ami sohasem következik be, $\Omega$ a biztos esemény – a teljes eseménytér – ami biztosan bekövetkezik; és néhány művelet: $A+B$ jelöli azt az eseményt,




1. ábra. Valószínűség, mint a relatív gyakoriság "határértéke"

hogy az $A$ és a $B$ események közül legalább az egyik bekövetkezik – halmazelméleti szempontból ez a két halmaz $A \cup B$ úniója, $AB$ pedig azt, hogy az $A$ és a $B$ esemény egyszerre bekövetkezik – halmazelméleti szempontból ez a két halmaz $A \cap B$ metszete, végül $\overline{A}$ az $A$ esemény komplementere, amely pontosan akkor következik be, amikor $A$ nem – halmazelméleti szempontból ez $A$ $\Omega$-ra vonatkozó komplementere, $\Omega \setminus A$. Hasonlóképpen definiálható pl. események egy $A_1, A_2, \ldots$ sorozatához az $A_1 + A_2 + \ldots$ esemény, ennek jelentése, hogy a sorozatban szereplő események közül legalább az egyik bekövetkezik. Itt is érvényesek a halmazműveleteknél már megszokott átalakítások, pl. az ún. de Morgan azonosságok: $\overline{A + B} = \overline{A} \cdot \overline{B}$, és $\overline{AB} = \overline{A} + \overline{B}$.

**1.1. Megjegyzés.** *A szóba jövő események összességét általában $\mathcal{F}$-fel jelöljük. Mélyebb matematikai oka van annak, hogy $\mathcal{F}$-be – a diszkrét $\Omega$ esetétől eltekintve – nem vesszük be $\Omega$ összes részhalmazát. $\mathcal{F}$ azonban mindig tartalmazza $\emptyset$-t, $\Omega$-t, és zárt a fenti műveletekre, pl. ha $A_1, A_2, \dots \in \mathcal{F}$, akkor $(A_1 + A_2 + \dots) \in \mathcal{F}$.*

Még a *valószínűség fogalmára* van szükségünk, ez egy $P : \mathcal{F} \to \mathbb{R}$ hozzárendelés, amely tehát minden szóba jövő $A$ eseményhez hozzárendel egy $P(A)$ számot, az esemény bekövetkezésének valószínűségét. A valószínűséget minden kísérletben máshogy kell meghatározni, intuíciónk alapján is posztulálhatjuk azonban a következő axiómákat, amelyek mindig teljesülnek: (I) minden $A$ eseményre $0 \leq P(A) \leq 1$, (II) $P(\Omega) = 1$, (III) ha az $A_1, A_2, \ldots$ események egymást páronként kizárják – azaz $A_iA_j = \emptyset$ minden $i \neq j$ esetben – akkor $P(A_1 + A_2 + \dots) = P(A_1) + P(A_2) + \dots$.

**1.2. Példa (Klasszikus valószínűségi mező).** *Ha (i) $\Omega$ egy véges halmaz, (ii) intuíciónk alapján az elemi eseményeket egyenlő valószínűségűeknek tekintjük. Ilyenkor tetszőleges $A$ esemény valószínűségét a $P(A) = \frac{|A|}{|\Omega|}$ képlettel határozhatjuk meg, itt a nevező a teljes eseménytér elemszáma – "összes lehetőség", a számláló pedig az $A$-t megvalósító elemi események száma – "kedvező lehetőségek".*

**1.3. Példa (Geometriai valószínűségi mező).** *Ha (i) $\Omega \subset \mathbb{R}^2$ valamely síkbeli tartomány, és (ii) intuíciónk alapján ezen a "valószínűség egyenletesen oszlik el". Ekkor minden szóba jövő $A \subset \Omega$ eseményre $P(A) = \frac{t(A)}{t(\Omega)}$, azaz $A$ területe (kedvező terület) osztva $\Omega$ területével (teljes terület).*

### 1.1.2. Feltételes valószínűség, függetlenség.

**Mérnöki bevezető a feltételes valószínűség fogalmához.** Folytassuk a megfigyeléseinket a villamosmegállóban, ahol esetenként találkozunk egy cimboránkkal, akivel együtt szoktunk utazni. Legyen továbbra is úgy, hogy az $n$ napon keresztül folytatott megfigyelésből $k$ alkalommal érkezik villamos a számunkra kedvező intervallumban. Ebből a $k$ alkalomból $m$ alkalommal cimboránk is megérkezik, és együtt utazunk. A cimboránk beérkezésének relatív gyakorisága, ha már beérkezett egy villamos: $m/k$. Ez a mennyiség egy feltételes relatív gyakoriság, hiszen feltételünk volt, hogy villamos már beérkezett a számunkra kedvező intervallumban. Az $m/k$ tört bővíthető: $(m/n)/(k/n)$ alakúra. Itt a számláló a villamos és cimboránk együttes beérkezésének relatív gyakorisága, a nevező a villamos beérkezésének relatív gyakorisága. Ez a hányados magyarázza a feltételes valószínűség fogalmát: Legyen $C$ esemény a cimboránk beérkezése, $P(C|V)$ jelölje cimboránk beérkezésének valószínűségét, feltéve, hogy a villamos is megjött. Ez a feltételes valószínűség (a relatív gyakoriságok példája alapján) egyenlő az együttes bekövetkezés valószínűsége osztva a villamos beérkezésének valószínűségével: $P(C|V) = P(C \cap V)/P(V)$.

**Mérnöki bevezető a függetlenség fogalmához.** Általában azt gondoljuk, hogy az előző példában a villamos beérkezése és cimboránk beérkezése nincs kapcsolatban egymással, függetlenek egymástól. Vagyis cimboránk beérkezésének valószínűsége, feltéve, hogy a villamos is beérkezett, ugyan akkora, mint cimboránk beérkezésének valószínűsége a feltétel nélkül: $P(C|V) = P(C)$. Ezt beírva a feltételes valószínűség definíciójába, azt kapjuk, hogy $P(C) \cdot P(V) = P(C \cap V)$. Szavakban: az együttes bekövetkezés valószínűsége egyenlő a valószínűségek szorzatával, ha a két esemény független.

Álljon itt még egy példa a feltételes valószínűség és a függetlenség fogalmának szemléltetésére. 365 reggelen keresztül figyeljük meg az időjárást Budapesten és Diósdon. Legyen $B$ illetve $D$ az az esemény, hogy Budapesten, illetve Diósdon esik az eső (az adott reggelen). A 365-ből rendre $k_B$, illetve $k_D$ alkalommal figyelünk meg esőt Budapesten, illetve Diósdon; ennek megfelelően az egyes valószínűségeket $P(B) \approx k_B/365$, illetve $P(D) \approx k_D/365$ közelíti. Ha $k_{BD}$-vel jelöljük azon reggelek számát, amikor Budapesten és Diósdon egyaránt esik az eső, akkor nyilván $k_{BD} \leq k_B$, de nem lehet $k_{BD}$ sokkal kisebb $k_B$-nél, hiszen ha Budapesten esik az eső, akkor általában Diósdon is esni szokott. Ennek megfelelően azt tapasztaljuk, hogy $k_B \cdot k_D < k_{BD}$, tükrözve, hogy $P(B \cap D) > P(B) \cdot P(D)$, illetve $P(D|B) > P(D)$ – ez a két esemény *nem* független egymástól. Ha $D$ helyett azt az $M$ eseményt tekintenénk, hogy Melbourneben esik az eső (ugyanazon a reggelen, ami ott estét jelent...) – akkor az $M$ és a $B$ eseményeket függetlennek találnánk, $k_B \cdot k_M \approx k_{BM}$ teljesülne, hiszen Budapest és Melbourne időjárását nagyjából függetlennek tekinthető körülmények alakítják, és így $P(B \cap M) = P(B) \cdot P(M)$-re, a két esemény függetlenségére számíthatunk.

**Feltételes valószínűség, függetlenség – matematikai tárgyalás.** Tegyük fel, hogy $P(B) \neq 0$, egyébként legyenek $A$ és $B$ tetszőleges események. Ekkor *$A$ feltételes valószínűsége $B$ bekövetkezése mellett*:

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

A fogalom jelentése: ha tudjuk, hogy $B$ bekövetkezett, mi a valószínűsége, hogy ($B$ mellett még) $A$ is bekövetkezett.
Az $A$ és $B$ események *függetlenek*, ha

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

*Egy fontos észrevétel.* Tegyük fel most, hogy $P(A) \neq 0$ és $P(B) \neq 0$. Ekkor

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

Ezek alapján: $A$ és $B$ függetlenek, ha *$A$ bekövetkezése nem befolyásolja $B$ valószínűségét.* Másképp fogalmazva, ha tudjuk, hogy $A$ bekövetkezett, semmi plusz információt nem nyerünk $B$ esélyére vonatkozóan (hiszen $B$ valószínűségét $A$ bekövetkezése mellett pontosan ugyanannyinak számolnánk, mint ezen információ nélkül).

**1.4. Megjegyzés.** *A kizáró események és a független események fogalmát ne keverjük össze! Ha $A$ és $B$ kizáró események, azaz $AB = \emptyset$, akkor $P(AB) = 0$ – vagyis a függetlenség semmiképp sem teljesülhet, ha pozitív valószínűségű eseményekről van szó. Másképp megfogalmazva: ha $A$ és $B$ kizáróak, $A$ bekövetkezésével nagyon sok információt nyerünk $B$ esélyeire vonatkozólag...*

**Bayes tétel.** Sokszor bizonyos feltételes valószínűségeket könnyebb kiszámolni, mint magukat a valószínűségeket. Egy a gyakorlatban előforduló szituáció: a $B$ eseményt különböző okok előzhetnek meg.

A kiváltó okokat tekinthetjük egy *teljes eseményrendszer* elemeinek: az $A_1, A_2, \ldots A_n$ események teljes eseményrendszert alkotnak, ha (i) $A_1 + A_2 + \cdots + A_n = \Omega$, (ii) $A_iA_j = \emptyset$ minden $i \neq j$ párra. Szavakban kifejezve, az $A_i$ események közül legalább egy biztosan bekövetkezik, de egyszerre kettő semmiképp sem következhet be. Ha könnyen ki tudjuk számolni a $P(A_i)$-ket (az egyes okok valószínűségeit), illetve a $P(B|A_i)$-ket ($B$ esélyét az egyes megelőző okok bekövetkezése mellett) akkor kézenfekvő a *teljes valószínűség tétele* néven ismert képlet:

$$P(B) = \left(\sum_{i=1}^{n} P(BA_i) =\right) \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

Az alkalmazások szempontjából legalább ennyire fontos a következő kérdés. Az $A_1, \ldots, A_n$ teljes eseményrendszer egyes elemeinek a valószínűségére vagyunk kiváncsiak, de azon információ birtokában, hogy a $B$ esemény bekövetkezett. Másképp fogalmazva: az $A_1, \ldots, A_n$ lehetőségek esélyeit miképp változtatja meg az a tény, hogy $B$ bekövetkezett? Vagyis $P(A_k|B)$-re vagyunk kiváncsiak. Némi számolással adódik *Bayes tétele*:

$$P(A_k|B) = \left(\frac{P(BA_k)}{P(B)} =\right) \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}.$$

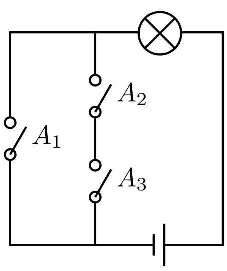
**1.5. Megjegyzés.** *Mielőtt elszörnyűlködnénk ezeken a képleteken, egy tanács a teljes vszg. tétele/Bayes tétel feladatok megoldásához. Készítsünk ágrajzot, ezen ábrázolva a $B$-t megvalósító $n$ lehetőséget. Az egyes ágakra írjuk rá felülre az előzmények valószínűségeit ($P(A_i)$), alulra pedig $B$ esélyét az adott előzmény mellett ($P(B|A_i)$). Az egyes ágak súlyát ezen számok szorzata adja. A teljes vszg tétel jelentése: mi az ágak súlya összesen? A Bayes tétel jelentése: mi a $k$-adik ág relatív súlya az összes ág súlyához képest?*

## 1.2. Kidolgozott példák

**1.1. Kidolgozott Feladat.** *A valószínűségszámítás egy fontos műszaki alkalmazása, amikor egy több részből álló rendszer megbízhatóságát vizsgáljuk, ismerve az egyes alkotóelemek működési valószínűségeit, valamint kapcsolódási hálójukat – gondolhatunk például egy elektromos áramkörre.*

(a) *A sorba kapcsolt $A$ és $B$ kapcsolókat az idő* 65*, illetve* 75 %*-ban zárjuk. Feltételezve, hogy a kapcsolók működtetése független, az idő hány százalékában folyik áram a teljes rendszeren?*

(b) *Helyezzük most el az $A_1$, $A_2$ és $A_3$ kapcsolókat a 2. ábra szerint, és tartsuk zárva ezeket egymástól függetlenül rendre $P(A_1) = 0,9$, $P(A_2) = 0,8$ és $P(A_3) = 0,7$ valószínűségekkel. Mi a teljes áramkör zárásának valószínűsége?*




2. ábra. Az 1.1. kidolgozott feladathoz

**Megoldás.**

(a) A következőképp gondolkodhatunk:

$$\begin{aligned} P(\text{áramkör zárva}) &= P(A \text{ zárva } \textbf{és } B \text{ zárva}) = P(A \text{ zárva}) \cdot P(B \text{ zárva}) = \\ &= 0,65 \cdot 0,75 = 0,4875, \end{aligned}$$

ahol az első egyenlőségnél a kapcsolók soros kapcsolását, a másodiknál függetlenségüket használtuk. Tehát az áramkör az idő $48,75$ százalékában van zárva.

(b) Használni fogjuk az alábbi, tetszőleges $A$ és $B$ események úniójának valószínűségére vonatkozó formulát (az ún. *szita formula* legegyszerűbb esete):

$$P(A+B) = P(A) + P(B) - P(A \cdot B). \tag{1}$$

Visszatérve a feladathoz, az áramkör akkor van zárva, ha vagy az $A_1$ kapcsolót zárjuk, vagy az $A_2$ és $A_3$ kapcsolókat egyszerre zárjuk. Ezt a tényt, majd az (1) formulát, végül pedig a kapcsolók függetlenségét használva:

$$\begin{aligned} P(\text{áramkör zárva}) &= P(A_1 \textbf{ vagy } (A_2 \textbf{ és } A_3)) = P(A_1) + P(A_2 \textbf{ és } A_3) - P(A_1 \textbf{ és } (A_2 \textbf{ és } A_3)) = \\ &= P(A_1) + P(A_2)P(A_3) - P(A_1)P(A_2)P(A_3) = \\ &= 0,9 + 0,8 \cdot 0,7 - 0,8 \cdot 0,8 \cdot 0,7 = 0,956. \end{aligned}$$

Ellenőrzésképp: a kapott valószínűség nagyobb $0,9$-nél, $A_1$ zárásának valószínűségénél, megfelelően annak, hogy ha $A_1$-t zárjuk, akkor az áramkör is biztosan zárul.

**1.6. Megjegyzés.** *Az (1) formula levezetéséhez csak annyit kell használnunk, hogy kizáró események úniójára a valószínűség összeadódik:*

$$\begin{aligned} P(A+B) &= P(\overline{A} \cdot B) + P(A \cdot B) + P(A \cdot \overline{B}) = \\ &= \big(P(\overline{A} \cdot B) + P(A \cdot B)\big) + \big(P(A \cdot \overline{B}) + P(A \cdot B)\big) - P(A \cdot B) = \\ &= P(B) + P(A) - P(A \cdot B). \end{aligned}$$

**1.2. Kidolgozott Feladat.** *Feldobunk két szabályos dobókockát, egy pirosat és egy kéket. Tekintsük a következő eseményeket:*

$A = \{$ *A piros dobókockán páros szám áll.* $\}$ $B = \{$ *A két dobás összege tíz.* $\}$
$C = \{$ *A két dobás eredménye azonos.* $\}$ $D = \{$ *A piros dobókockán páratlan szám áll.* $\}$
$E = \{$ *A kék dobókockán páratlan szám áll.* $\}$

*(a)* $P(B) =$*?*, $P(C)$=*?*

*(b)* $P(A \mid B)$=*?*

*(c) A B, C, D és E események közül melyiktől független A, és melyiktől nem?*

**Megoldás.** A kísérlet eredményét megadhatjuk, ha megmondjuk, milyen szám áll az egyes kockákon. Ennek megfelelően az eseménytér:

$$\Omega = \{(i,j) | i = 1, \ldots, 6; j = 1, \ldots, 6\}$$

ahol $i$, illetve $j$ a piros, illetve a kék kocka eredménye. Tehát klasszikus valószínűségi mezővel van dolgunk, és a valószínűségek számolásánál a nevező minden esetben $|\Omega| = 6 \cdot 6 = 36$.

(a) $B = \{(4,6),(5,5),(6,4)\}$ tehát $|B| = 3$ és így $P(B) = 3/36 = 1/12$. $C = \{(1,1),\dots,(6,6)\}$ így $|C| = 6$ és $P(C) = 1/6$.

(b) Mivel $P(A \mid B) = \frac{P(AB)}{P(B)}$ és itt a nevezőt már az imént kiszámoltuk, ezért a számlálót kell meghatározni. Ehhez $AB = \{$ A piros kockán páros szám áll, és a dobások összege 10.$\} = \{(4,6),(6,4)\}$. Tehát $P(AB) = 2/36 = 1/18$ és $P(A \mid B) = \frac{1/18}{1/12} = 2/3$.

Másképp gondolkodva: a $B$-t megvalósító 3 elemi esemény közül pontosan 2 valósítja meg $A$-t is, így $A$ esélye, ha tudjuk, hogy $B$ bekövetkezett, 2/3.

(c) Ehhez először $A = \{(1,2),\dots,(6,2),(1,4),\dots,(6,4),(1,6),\dots,(6,6)\}$ tehát $|A| = 18$ és $P(A) = 1/2$.

Az előző részfeladatok alapján $P(A)P(B) \neq P(AB)$ tehát $A$ és $B$ *nem* függetlenek. Ugyenez látszik abból is, hogy $P(A \mid B) \neq P(A)$ – pozitív valószínűségű eseményekről van szó. (Ha $B$-ről tudjuk, hogy bekövetkezett, inkább fogadnánk $A$-ra, mint ha nem tudnánk semmit).

$AC = \{(2,2),(4,4),(6,6)\}$ és így $P(AC) = 1/12$, amiből $P(AC) = P(A)P(C)$, tehát $A$ és $C$ függetlenek.

$P(D) = 1/2$ és $P(E) = 1/2$ ugyanúgy adódik, mint $P(A)$. Ugyanakkor nyilván $AD = \emptyset$, tehát $P(AD) = 0$, vagyis $A$ és $D$ *nem* függetlenek (emlékeztető: pozitív vszg-ű kizáró események nem lehetnek függetlenek). Másrészt $AE = \{(1,2),(3,2),(5,2),(1,4),(3,4),(5,4),(1,6),(3,6),(5,6)\}$ és így $P(AE) = 1/4$, vagyis $P(A)P(E) = P(AE)$, tehát $A$ és $E$ függetlenek.

**1.3. Kidolgozott Feladat.** *Egy urnában tapintásra megkülönböztethetetlen,* 1*-től* 90*-ig számozott cédulák vannak, ezek közül a* 8, 19, 23, 64 *és* 74 *számúak feketék, a többi fehér. Belenyúlunk az urnába, és kihúzunk (visszatevés nélkül)* 5 *cédulát. Mi a valószínűsége, hogy a kihúzottak között pontosan k fekete* $(k = 0,1,\dots,5)$*? (Hogy hívják ezt a játékot más néven?)*

**Megoldás.** Először el kell döntenünk, mik legyenek az elemi események. Kézenfekvő választás, ha azt tekintjük egy elemi eseménynek, hogy megmondjuk, melyik a kihúzott öt cédula. Ezzel a konvencióval *eltekintünk a cédulák húzásának sorrendjétől.* Lényeges, hogy ezt végig tartsuk észben. Az elemi események tehát az $\{1,2,3,\dots,90\}$ halmaz ötelemű részhalmazai, azaz

$$\Omega = \{\{1,2,3,4,5\},\{1,2,3,4,6\},\dots,\{86,87,88,89,90\}\}.$$

Fontos látni, hogy pl. $\{1,3,6,8,10\}$ és $\{6,10,3,8,1\}$ ugyanaz az elemi esemény.

Jelöljük $A_k$-val azt az eseményt, hogy pontosan $k$ találatunk van $(k = 0,1,\dots,5)$.

Az összes lehetőségek száma, ahányféleképpen ki lehet választani 90 elemből 5-t, tehát $|\Omega| = \binom{90}{5}$. Minden esetben ez lesz a nevező.

A számlálók számolásához $|A_k|$-t kell meghatározni. Nyilván $A_5 = \{\{8,19,23,64,74\}\}$ és így $|A_5| = 1$, $P(A_5) = \binom{90}{5}^{-1}$. A többi $|A_k|$-hoz azt kell kiszámolni, hány olyan részhalmaza van az $\{1,2,\dots,90\}$ halmaznak, amely a $\{8,19,23,64,74\}$ halmazból pontosan $k$ elemet tartalmaz, ennek komplementeréből pedig pontosan $(5-k)$ elemet. Mivel egy ötelemű halmaz $k$ elemű részhalmazainak száma $\binom{5}{k}$, egy 85 elemű halmaz $(5-k)$ elemű részhalmazainak száma pedig $\binom{85}{5-k}$, adódik

$$|A_k| = \binom{5}{k}\binom{85}{5-k}; \qquad P(A_k) = \frac{\binom{5}{k}\binom{85}{5-k}}{\binom{90}{5}}.$$

**1.4. Kidolgozott Feladat.** *Válasszuk az*

$$A = \begin{bmatrix} a & 1 \\ 1 & d \end{bmatrix}$$

*szimmetrikus mátrix a és d elemeit egyenletes eloszlással, egymástól függetlenül az* $[1/2, 2]$ *intervallumból. Mi a valószínűsége, hogy* $\det A > 0$*?*

**Megoldás.** Mivel az $a$ és a $d$ elemeket függetlenül és egyenletes eloszlással választjuk, geometriai valószínűségi mezővel van dolgunk, az eseménytér:

$$\Omega = [1/2, 2] \times [1/2, 2]; \quad \text{ahol az} \quad (a, d) \in \Omega$$

pár első és második tagja az $A$ véletlen mátrix $(1/2 \leq) a (\leq 2)$ illetve $(1/2 \leq) d (\leq 2)$ diagonális elemeit jelöli. Jelöljük $C$-vel a $\det A > 0$ eseményt. $P(C)$ számolásához a teljes terület nyilván: $t(\Omega) = (2 - 1/2)^2 = 2,25$. A hasznos területhez a

$$C = \{(a, d) \in \Omega | \det A > 0\} = \{(a, d) \in [1/2, 2] \times [1/2, 2] \mid ad > 1\}$$

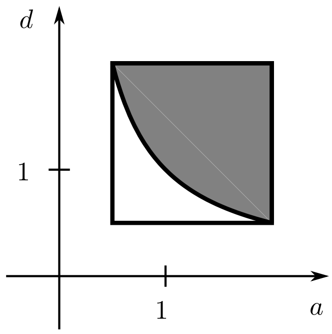
tartomány területét kell kiszámolni, egyszerűbb azonban $\overline{C}$ területének számolása, ugyanis ezt a halmazt az $1/2 \leq a \leq 2; 1/2 \leq d \leq 1/a$ egyenlőtlenségek jellemzik, és így integrálással:

$$t(\overline{C}) = \int_{0,5}^{2} \frac{1}{a}\, da - 1,5 \cdot 1/2 = 2 \ln 2 - 0,75.$$

Végül

$$P(C) = 1 - P(\overline{C}) = 1 - \frac{2 \ln 2 - 0,75}{2,25} \approx 0,7172.$$

A feladat megoldását a 3. ábra is szemlélteti.




3. ábra. Az 1.4. kidolgozott feladathoz

**1.5. Kidolgozott Feladat.** *Egy kikötőhöz 24 órás időtartamon belül véletlen időpontban két hajó érkezik. Az előbb érkezőn rögtön megkezdik a rakodást, amely két óráig tart. Ha a második hajó akkor érkezik, amikor az elsőn még rakodnak, akkor várakoznia kell a rakodás befejezéséig. Mi a valószínűsége, hogy szükség lesz várakozásra?*

**Megoldás.** Geometriai valószínűségi feladatról van most is szó: jelöljük $x$-szel és $y$-nal (óra egységekben) az egyes hajók érkezésének időpontját. Ekkor az elemi események leírhatók, mint $(x, y)$, $0 \leq x \leq 24$, $0 \leq y \leq 24$ pontpárok: az $\Omega$ eseménytér egy 24 egységnyi oldalhosszúságú négyzet. Jelöljük $A$-val azt az eseményt, hogy szükség van várakozásra. Ez akkor és csak akkor fordul elő, ha a két hajó érkezési időpontja között kevesebb, mint 2 óra a különbség, vagyis ha $|x - y| \leq 2$. Könnyen látható, hogy a komplementer tartomány előáll, mint két diszjunkt, derékszögű egyenlő szárú, 22 egységnyi befogójú háromszög. Így:

$$P(A) = 1 - \frac{22^2}{24^2} \approx 0,1597.$$

**1.6. Kidolgozott Feladat.** *6 doboz mindegyikében 6 cédula van, ezek közül rendre $1, 2, \dots, 6$ fekete, a többi fehér. Feldobunk egy szabályos dobókockát, ha a dobás eredménye $k$, a $k$-adik dobozból húzunk egy cédulát.*

*(a) Mi a valószínűsége, hogy a kihúzott cédula fekete?*

*(b) tegyük most fel, hogy fekete cédulát húztunk, és a szomszéd szobában levő barátunknak csak ennyit árultunk el a kísérletről. Barátunk szerint mi a valószínűsége annak, hogy hatost dobtunk?*

**Megoldás.** Vezessük be a következő eseményeket:

$A_k :=$ { a kockával $k$-t dobunk } $=$ {a $k$-adik dobozból húzunk}; $k = 1, \dots, 6$; továbbá $B :=$ {a kísérlet végén fekete cédulát húzunk.}

Az $A_k$; $k = 1, \dots, 6$ események teljes eseményrendszert alkotnak, és nyilván $P(A_k) = 1/6$ mindegyikükre. $P(B|A_k)$ annak valószínűsége, hogy a $k$-adik dobozból fekete cédulát húzunk: mivel az ebben a dobozban található 6 cédula közül pontosan $k$ fekete, könnyen adódik $P(B|A_k) = \frac{k}{6}$.

(a) A teljes valószínűség tétele szerint:

$$P(B) = \sum_{i=1}^{6} P(B|A_i)P(A_i) = \frac{1 + 2 + \cdots + 6}{36} = \frac{6 \cdot (6+1)}{2 \cdot 36} = \frac{7}{12}.$$

(b) A feladat kérdése: $P(A_6|B) =?$ Ezt Bayes tételével számolhatjuk:

$$P(A_6|B) = \frac{P(B|A_6)P(P(A_6))}{P(B)} = \frac{1 \cdot 1/6}{7/12} = \frac{2}{7}.$$

**1.7. Kidolgozott Feladat.** *Magyarországon minden tízezredik lakos HIV fertőzött. A fertőzöttség szűrésére AIDS tesztet használnak, ami az esetek kis %-ában sajnos téved. Konkrétan, az egészséges emberek tesztjeinek 1%-a pozitív, illetve a fertőzött emberek tesztjeinek 1,5%-a negatív. Jancsi Bácsi tesztje sajnos pozitív. Mi a valószínűsége, hogy Jancsi Bácsi valóban fertőzött?*

**Megoldás.** Vezessük be a kövekező eseményeket:

$A_1 =$ { Jancsi bácsi HIV fertőzött. }, $A_2 = \overline{A_1} =$ { Jancsi bácsi *nem* HIV fertőzött. }

$B =$ {Jancsi bácsi tesztjének eredménye pozitív.} Végül $\overline{B} =$ {Jancsi bácsi tesztjének eredménye negatív.}

A feladat kérdése: $P(A_1|B) =?$ Bayes tételét fogjuk használni.

Nyilván $A_1$ és $A_2$ teljes eseményrendszert alkotnak. Mivel Jancsi bácsiról (a teszt elvégzése előtt) semmi információnk nincsen, így ő a magyar társadalom egy véletlenszerűen választott tagjának tekinthető, és így:

$$P(A_1) = 0,0001; \qquad P(A_2) = 1 - P(A_1) = 0,9999.$$

Meg kell még határozni $P(B|A_1)$-t és $P(B|A_2)$-t. $P(B|A_2)$ annak valószínűsége, hogy egy egészséges embert fertőzöttnek mutasson a teszt, a feladat szövege alapján $P(B|A_2) = 0,01$. A másik, feladat szövegéből közvetlenül kiolvasható valószínűség $P(\overline{B}|A_1) = 0,015$, annak esélye, hogy egy fertőzött embert egészségesnek mutat a teszt. Ebből:

$$P(B|A_1) = 1 - P(\overline{B}|A_1) = 0,985.$$

Most alkalmazhatjuk Bayes tételét:

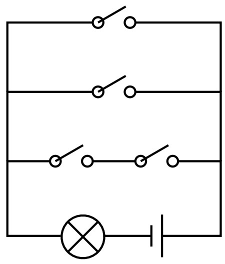
$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} = \frac{0,985 \cdot 0,0001}{0,985 \cdot 0,0001 + 0,01 \cdot 0,9999} \approx 0,00975;$$

tehát kevesebb, mint 1 %!

## 1.3. Gyakorló feladatok

**1.8. Feladat.** *Egy medencét egy hideg- és egy melegvizes csapon keresztül lehet feltölteni. A $H$ esemény jelentse azt, hogy a hidegvizes csapon keresztül folyik a víz, az $M$ pedig azt, hogy a melegvizes csapon keresztül. Írjuk le műveletekkel, melyik az az esemény, amikor (a) csak a hidegvizes csapon keresztül folyik a víz; (b) egyik csapon keresztül sem folyik víz; (c) pontosan egy csapon keresztül folyik a víz; (d) legalább az egyik csap el van zárva. (Esetenként több ekvivalens alak is lehetséges.)*

**1.9. Feladat.** *Tekintsük a 4. ábra áramkörét. A kapcsolókat egymástól függetlenül* $0,8$ *valószínűséggel zárjuk. Milyen valószínűséggel van zárva az áramkör?*



4. ábra. Az 1.9. feladathoz

**1.10. Feladat.** *Feldobunk 3 szabályos érmét: egy 20, egy 50 és egy 100 Ft-ost. Tekintsük a következő eseményeket:*

$A = \{$ *A 20 Ft-os Fej oldalra esik.* $\}$ $B = \{$ *Az eredmények között pontosan kettő Írás.* $\}$

$C = \{$ *A 100 Ft-os és az 50 Ft-os érme azonos oldalára esik.* $\}$

*(a)* $P(A) = ?$ *(b)* $P(A|B) = ?$ *(c) Független-e egymástól az* $A$ *és a* $C$ *esemény?*

**1.11. Feladat.** *32 lapos magyar kártyából húzunk két lapot.*

*(a) Feltéve, hogy a két lap egyike sem piros, mi a valószínűsége, hogy szerepel köztük a tök ász?*

*(b) Függetlenek-e az alábbi események? Indokoljuk a választ!*

$A = \{$*A két lap között szerepel piros.*$\}$ $B = \{$*A két lap között szerepel a tök ász.*$\}$

**1.12. Feladat.** *5 napos szabadságunk alatt a hűtőszekrény akkor és csak akkor működik, ha az* $A$ *és* $B$ *akkumulátorok közül legalább az egyik be van kapcsolva. Az* $A$ *akkumulátor a szabadság kezdetétől folyamatosan működik, majd a 3. napon 0 és 12 óra között véletlen időpontban kikapcsol. A* $B$ *akkumulátor ugyanebben az időintervallumban véletlen időpontban bekapcsol, és hazatértünkig folyamatosan működik. A hűtőben tárolt hús akkor és csak akkor romlik meg, ha legalább 6 óráig nincs hűtés. Mi a valószínűsége, hogy hazatérve romlott húst találunk a hűtőben?*

**1.13. Feladat.** *Zabhegyezéstanból – mint minden tárgyból – a puskázás csak rontja a vizsgázók esélyeit: akik csalnak, 30 % eséllyel mennek át a vizsgán, a becsületesek 60 % eséllyel sikeresek. Ennek ellenére a vizsgázóknak pontosan a fele puskázik. Kukutyin Kázmérról csak annyit tudunk, hogy átment a vizsgán. Mi a valószínűsége, hogy puskázott?*

# 2. Valószínűségi változók

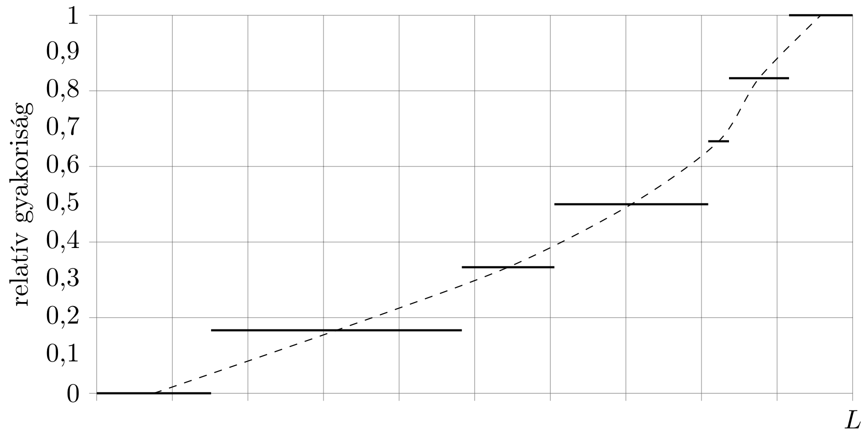
## 2.1. Elméleti összefoglaló

### 2.1.1. Valószínűségi változók.

**Mérnöki bevezető a valószínűségi változó fogalmához.** A véletlen fogalmának mérnöki megfogalmazásakor példaként használtuk azt az időpontot, amikor egy villamos beérkezik egy megállóba. A közvetlen megfigyelés alapján azt a következtetést vontuk le, hogy ezt az időpontot a véletlen és a menetrend együttesen határozzák meg. A mérnöki gyakorlat sok ilyen típusú változót ismer: például egy szerkezeti anyag szakítószilárdsága azonos gyártástechnológia mellett is ingadozik, szigorú minőségbiztosítási feltételek mellett előállított termék élettartama is változik, azonos körülmények között végrehajtott mérés eredménye is esetről esetre különböző lehet. Az olyan változókat, ahol a változó értékét a figyelembe vett körülmények nem határozzák meg egyértelműen, valószínűségi változóknak nevezzük. A korábban felsorolt változók egy intervallumon belül bármilyen értéket felvehetnek, ezeket folytonos valószínűségi változóknak hívjuk. De a műszaki gyakorlatban gyakoriak az olyan valószínűségi változók is, amelyek csak diszkrét értékeket vesznek fel. Egy műhelyben egy adott időszak alatt meghibásodó gépek száma, egy szállítási tételben a selejtes darabok száma, egy telefonközpontba percenként befutó hívások száma, egy weblapot időegység alatt felkeresők száma, stb. ilyen diszkrét változó. Ahhoz, hogy a véletlenszerűen ingadozó valószínűségi változókból a mérnöki gyakorlatban használható következtetéseket tudjunk levonni, más módszereket kell használnunk, mint a determinisztikus változók esetén. Ehhez szükséges megismernünk az eloszlás- és sűrűségfüggvény, továbbá a várható érték és a szórás fogalmát.

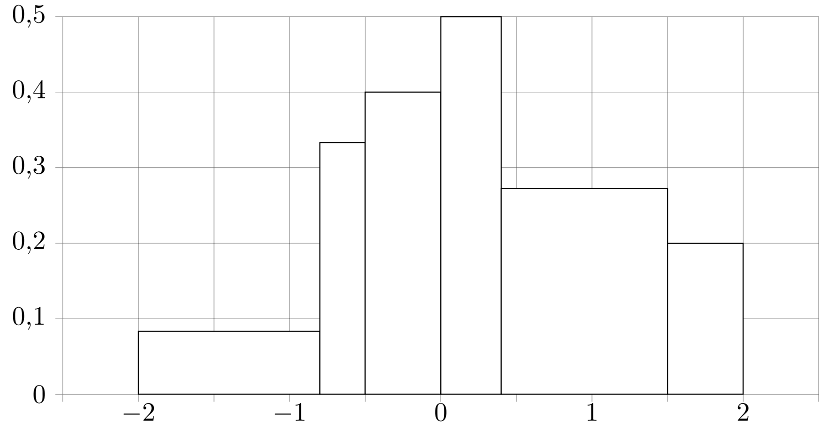
**Mérnöki bevezető az eloszlásfüggvény fogalmához.** Vegyünk egy $L$ hosszúságú, homogén, prizmatikus rudat, és a két végére ható húzóerővel szakítsuk el. Most ne foglalkozzunk a befogás környezetében törvényszerűen fellépő feszültségtorzulásokkal, és tekintsük úgy, hogy a rúd minden keresztmetszete ugyanolyan feszültségi állapotban van. A tett feltételek miatt minden keresztmetszet egyformán veszélyes, így a szakadás helye véletlenszerű. A rúd egyik végpontjától mérjük meg a szakadás helyét, legyen ez a $\xi$ jelű, távolság mértékegységű valószínűségi változó. Hajtsunk végre $n$ szakítási kísérletet, a szakadási helyek sorozata: $\xi_1, \xi_2, \ldots \xi_n$. Rajzoljuk meg a következő grafikont: az $x$ független változót mérjük a rúd hossztengelyének irányában: $0 \leq x \leq L$. Rögzítsünk egy tetszőleges $x$ értéket, és számoljuk össze, hogy hány rúd szakadt el a $[0, x)$ intervallumban (vagyis az $x$-től "balra"), jelölje ezt a számot $k_x$. Minden $x$ független változóhoz mérjük fel a $k_x/n$ relatív gyakoriságot. Nyilván, hogy ha $x < 0$, akkor $k_x/n = 0$, ha $x > L$, akkor $k_x/n = 1$. Egy adott méréssorozat ismeretében megrajzolhatjuk ezt az ábrát (lásd a 5. ábrát), amely nyilvánvalóan lépcsős függvény lesz, minden szakadási helyen áthaladva $k_x$ értéke ugrik. Feltételezhetjük, hogy ha a mérések száma növekszik, akkor a lépcsős függvény egy folytonos függvény felé tart. Ha $n \to \infty$ akkor $k_x/n$ közelíti a $p_x$ valószínűséget: annak az eseménynek a valószínűségét, hogy a szakadás helye kisebb, mint $x$, ezt úgy jelöljük, hogy $P(\xi < x)$. Az előbbiek alapján nyilvánvaló, hogy ez az érték $x$-től függ, tehát $P(\xi < x) = F(x)$. Ezt az $F(x)$ függvényt hívjuk eloszlásfüggvénynek. Az eloszlásfüggvény fontos tulajdonsága, hogy ismeretében meg tudjuk mondani egy adott $[a, b]$ intervallumba esés valószínűségét: $P(a \leq \xi \leq b) = F(b) - F(a)$.

**Mérnöki bevezető a sűrűségfüggvény fogalmához** Az eloszlásfüggvény grafikonja nem nyújt szemléletes képet a valószínűségi változó "eloszlásáról": például ránézésre nem könnyű megállapítani, hogy azonos hosszúságú intervallumok közül, melyik intervallumba esik nagyobb valószínűséggel a változó. Szemléletesebb képet kapunk, ha sűrűségfüggvényt szerkesztünk az alábbi módon. Egy megfigyeléssorozat végén rendelkezésünkre áll a $\xi_1, \xi_2, \ldots \xi_n$ számsorozat, mint a $\xi$ valószínűségi változó megvalósult értékei (például a rúdszakadás helyei). Nevezzük ezt a számsorozatot mintának. A legkisebb és a legnagyobb mintaelemmel határolt intervallumot osszuk fel $\Delta x_1, \Delta x_2, \ldots, \Delta x_K$ részintervallumokra. Számoljuk meg, hogy a $\xi_1, \xi_2, \ldots \xi_n$ elemekből hány darab esik az első, a második, stb. a $K$-adik részintervallumba. Ezeket a számokat, az egyes a részintervallumokba esés gyakoriságait, a továbbiakban $\nu_1, \nu_2, \ldots \nu_K$ -val jelöljük. Rajzoljunk a részintervallumok fölé olyan lépcsős függvényt, hogy minden részintervallum feletti terület legyen egyenlő a részintervallumba esés relatív gyakoriságával (lásd a 6.




5. ábra. Az eloszlásfüggvény szemléletes származtatása

ábrát). Egy-egy terület nagysága $\nu_j/n$, a $\Delta x_j$ alaphossz fölé így $\nu_j/(n\Delta x_j)$ magasságú lépcsőt kell rajzolnunk. Az így kapott lépcsős függvényt nevezzük tapasztalati sűrűségfüggvénynek, jele $f_n(x)$. Ha mintaelemek számát növeljük és a $\Delta x$ részintervallumok hosszát csökkentjük, akkor a lépcsős függvény egy folytonos függvényt közelít. Nevezzük ezt a függvényt sűrűségfüggvénynek. Ez a függvény olyan tulajdonságú, hogy $[a, b]$ intervallum feletti integrálja egyenlő az intervallumba esés valószínűségével. Ezt összevetve az eloszlásfüggvény definíciójával, láthatjuk, hogy a sűrűségfüggvény az eloszlásfüggvény deriváltfüggvénye. Előnye, hogy szemléletes képet ad a valószínűségi változó eloszlásáról: ha egyenlő hosszúságú részintervallumokra osztjuk fel a valószínűségi változó értékkészletét, akkor a változó abba az intervallumba esik nagyobb valószínűséggel, amely feletti terület nagyobb.




6. ábra. A sűrűségfüggvény szemléletes származtatása

**Matematikai értelemben valószínűségi változó**knak hívjuk az $X : \Omega \to \mathbb{R}$ (mérhető) függvényeket. A gyakorlatban ez *véletlen számot* jelent. Természettudományos vagy műszaki háttérrel pedig úgy is gondolhatunk rá, mint egy mérési eredményre – egy mennyiségre, aminek az értéke függ a kísérlet kimenetelétől, és így a véletlentől. A valószínűségi változókat általában latin nagybetűkkel $(X, Y, Z, \dots)$ vagy görög kisbetűkkel $(\xi, \eta, \zeta, \dots)$ szoktuk jelölni.

Az **eloszlásfüggvény** egy minden valószínűségi változóra értelmezhető, azt jól jellemző $F : \mathbb{R} \to \mathbb{R}$

függvény, melyet a következőképp definiálunk:

$$F(x)(= F_\xi(x)) := P(\xi < x), \qquad \forall x \in \mathbb{R},$$

ahol $\xi$ a szóban forgó valószínűségi változó. Fontos látni a $\xi$ és az $x$ közti különbséget: $\xi$ egy véletlen szám, $x$ pedig egy (tetszőleges módon) rögzített, véletlentől nem függő érték. Az eloszlásfüggvény alábbi tulajdonságai intuíció alapján könnyen végiggondolhatóak:

(F1) $0 \leq F(x) \leq 1, \ \forall x \in \mathbb{R}$;

(F2) $F(x)$ monoton növő függvény: $\forall x_1 < x_2 : \ F(x_1) \leq F(x_2)$;

(F3) $\lim\limits_{x \to -\infty} F(x) = 0, \qquad \lim\limits_{x \to +\infty} F(x) = 1$;

(F4) $F(x)$ balról folytonos: $\lim\limits_{x \to a-0} F(x) = F(a), \quad \forall a \in \mathbb{R}$.

**2.1. Megjegyzés.** *Bizonyítható az is, hogy minden (F1)-(F4) tulajdonságokkal rendelkező $F : \mathbb{R} \to \mathbb{R}$ függvényhez található valószínűségi változó, amelynek $F(x)$ eloszlásfüggvénye. Tehát ezek a tulajdonságok pontosan karakterizálják az eloszlásfüggvényeket.*

Az eloszlásfüggvény jelentősége, hogy segítségével megfogalmazhatók a valószínűségi változóra vonatkozó kijelentések, például annak valószínűsége, hogy $\xi$ egy adott intervallumba essen:

$$P(a \leq \xi < b) = F_\xi(b) - F_\xi(a), \qquad \forall a < b.$$

**Diszkrét valószínűségi változó** olyan véletlen mennyiség, amely csak véges vagy megszámlálhatóan végtelen sok különböző értéket vehet fel. Legyenek ezek az értékek rendre az $x_1, x_2, \ldots$ valós számok. Diszkrét valószínűségi változó fontos jellemzője a *valószínűség-eloszlás*:

$$p_k := P(\xi = x_k), \ k = 1, 2, \ldots$$

melynek alaptulajdonságai:

(p1) $0 \leq p_k \leq 1, \quad k = 1, 2, \ldots$;

(p2) $\sum\limits_{i=1}^{\infty} p_k = 1$ (véges sok – $n$ –különböző érték esetén persze $\sum\limits_{i=1}^{n} p_k = 1$).

Diszkrét valószínűségi változó $F : \mathbb{R} \to \mathbb{R}$ eloszlásfüggvénye lépcsős függvény: azaz $F(x)$ a véges (vagy megszámlálható végtelen sok) $x_k$ értéktől eltekintve konstans, az $x_k$ pontokban $p_k$ nagyságú pozitív ugrásai vannak, és persze a fenti (F4) értelmében itt is balról folytonos.
**Abszolút folytonos valószínűségi változó** esetén ezzel szemben nemcsak hogy $F(x)$ folytonos $\forall x \in \mathbb{R}$-re, hanem (esetleg néhány kivételes ponttól eltekintve) differenciálható is. Pontosabban, létezik egy $f : \mathbb{R} \to \mathbb{R}$ függvény, a valószínűségi változó *sűrűségfüggvénye*, hogy:

$$F(x) = \int\limits_{-\infty}^{x} f(t)\, dt; \ \forall x \in \mathbb{R}.$$

A sűrűségfüggvény ugyanazt a szerepet tölti be abszolút folytonos esetben, mint a valószínűség-eloszlás diszkrét esetben. Alaptulajdonságai:

(f1) $0 \leq f(x), \ \forall x \in \mathbb{R}$;

(f2) $\int\limits_{-\infty}^{\infty} f(x)\, dx = 1$.

**2.2. Megjegyzés.** *Érdemes megjegyezni, hogy vannak olyan valószínűségi változók, amelyek se nem diszkrétek, se nem abszolút folytonosak. Ez nem csupán matematikai absztrakció, ilyen véletlen mennyiségek előfordulnak a természettudományos és a műszaki alkalmazásokban is – szoros kapcsolatuk van például a fraktálokkal – tárgyalásuk azonban meghaladja ennek az összefoglalónak a kereteit.*

### 2.1.2. Várható érték, szórás.

**Mérnöki bevezető a várható érték fogalmához.** Amint a pontos matematikai megfogalmazásból kiderült, egy valószínűségi változó minden jellemzője kiszámítható az eloszlásfüggvényből (vagy a sűrűségfüggvényből). Sajnos e két függvény megismerése, vagy statisztikai adatokból való becslése sok kísérletet és (az eloszlásra vonatkozó) ellenőrizendő feltevést kíván. Számos esetben meg kell elégednünk ennél kevesebb információval. Bizonyos esetekben elegendő, ha meg tudjuk mondani, hogy mely érték körül ingadozik véletlenszerűen a szóban forgó változó, és tudunk értéket mondani az ingadozás mértékére. Az ingadozás közepeként természetes módon a megfigyelt értékek átlagát szokás tekinteni. Az itt következő mérnöki gondolatmenet megmutatja, hogy hogyan juthatunk el az átlagtól a várható érték fogalmához.

A sűrűségfüggvény bevezetőjében mondottakhoz hasonlóan, mondjuk, hogy rendelkezésünkre áll a $\xi_1, \xi_2, \dots \xi_n$ számsorozat (minta). Ennek számtani átlaga:

$$\xi_a = \frac{1}{n}\sum_{i=1}^{n}\xi_i.$$

Egy $\Delta x_j$ intervallumba eső mintaelemek összegét közelítőleg úgy is kiszámíthatjuk, hogy az intervallum $x_j$ középértékét szorozzuk az intervallumba eső mintaelemek $\nu_j$ számával. Ezeket a szorzatokat összegezzük az összes intervallumra, így a mintaelemek összegének közelítését kapjuk.

$$\xi_a = \frac{1}{n}\sum_{i=1}^{n}\xi_i \approx \frac{1}{n}\sum_{j=1}^{K}\nu_j x_j = \sum_{j=1}^{K} x_j \frac{\nu_j}{n\Delta x_j}\Delta x_j = \sum_{j=1}^{K} x_j f_n(x_j)\Delta x_j.$$

A törtet bővítjük $\Delta x_j$-vel, az összefüggésben felismerhető az $f_n(x)$ tapasztalati sűrűségfüggvény. A kapott összefüggés egy improprius integrál közelítő összege, ezt az integrált hívjuk a valószínűségi változó várható értékének. A várható érték ebben a megfogalmazásban az átlag általánosítása, fizikai szempontból ugyanolyan jellegű mennyiség, mint maga a valószínűségi változó, ugyanazzal a mértékegységgel. Értéke egy adott valószínűségi változóra állandó, nem függ a véletlentől.

**Mérnöki bevezető a szórás fogalmához.** A valószínűségi változó ingadozását a pillanatnyi érték és a várható érték különbsége mutatja. Ez is véletlen mennyiség, kérdés, hogy hogyan tudjuk ezt egyetlen számmal jellemezni. Kézenfekvő, hogy vegyük e különbség várható értékét. Belátható, hogy ez minden változóra zérus (a pozitív és negatív különbségek "kiegyenlítik" egymást), ezért ez nem jó jellemző. Ha a különbség négyzetét, vagy abszolút értékét vesszük, akkor a várható érték pozitív lesz. A gyakorlatban a négyzet mutatkozott hasznosabbnak, ezért a valószínűségi változó ingadozását az átlagos négyzetes eltéréssel jellemezzük: vesszük a változó és a várható érték különbségét, ezt négyzetre emeljük, és ennek vesszük a várható értékét. Ezt a mennyiséget nevezzük a valószínűségi változó szórásnégyzetének, és pozitív négyzetgyökét szórásnak. A várható értéknél tett megállapításaink a szórásra is vonatkoznak: adott valószínűségi változó szórása állandó, nem függ a véletlentől, mértékegysége megegyezik a valószínűségi változó mértékegységével. Megfigyelési értékekből (mintából) becsülhető a szórás értéke, a becslést tapasztalati szórásnak hívjuk. Ismerjük a $\xi_1, \xi_2, \dots \xi_n$ számsorozatot, ennek átlaga $\xi_a$, ezekből az adatokból az $s_\xi^2$ tapasztalati szórásnégyzet így számítható:

$$s_\xi^2 = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \xi_a)^2.$$

A **várható érték** és a **szórás** valószínűségi változók további fontos jellemzői, szokásos jelölésük $E\xi$ (expectation) illetve $D\xi$ (deviation). Gépészmérnökként hasznos analógia lehet: ha egy diszkrét, illetve folytonos eloszlásra, mint egydimenziós pontrendszerre, illetve folytonos inhomogén sűrűségű anyagra gondolunk, akkor a várható érték a súlypontnak, a szórásnégyzet pedig a (súlypontra vonatkoztatott) tehetetlenségi nyomatéknak felel meg. Másképp szólva, a várható érték az eloszlás "közepét", a szórás annak "szétkentségét" jellemzi. Számolásuk:

| | diszkrét | folytonos |
|---|---|---|
| várható érték, $E\xi$ | $\sum_{k=1}^{\infty} x_k p_k$ | $\int_{-\infty}^{\infty} x f(x)\, dx$ |
| szórásnégyzet, $D^2\xi$ | $\sum_{k=1}^{\infty} (x_k - E\xi)^2 p_k$ | $\int_{-\infty}^{\infty} (x - E\xi)^2 f(x)\, dx$ |

Definiálhatjuk a $\xi$ valószínűségi változótól (determinisztikusan) függő $t(\xi)$ mennyiség várható értékét is, $E(t(\xi)) = \sum_{k=1}^{\infty} t(x_k)p_k$ illetve $\int_{-\infty}^{\infty} t(x)f(x)dx$ alapján. A szórásnégyzetre könnyen adódik a $D^2\xi = E\xi^2 - (E\xi)^2$ hasznos alternatív képlet (vessük össze a Steiner tétellel!). Végül pedig a szórás: $D\xi = \sqrt{D^2\xi}$.

**2.3. Megjegyzés.** *Amennyiben az $E\xi$-t, illetve $D^2\xi$-t definiáló végtelen sor/improprius integrál nem abszolút konvergens, azt mondjuk, a várható érték, illetve a szórás nem létezik.*

### 2.1.3. Együttes eloszlások.

**Mérnöki bevezető a korrelációs együttható fogalmához.** A mérnöki gyakorlatban ritkán fordul elő, hogy egyetlen változót figyelünk meg, gyakran a változók közötti kapcsolatot vizsgáljuk: egyszerre figyelünk meg több változót. Egyes esetekben a vizsgált jelenség fizikai háttere nyilvánvalóvá teszi, hogy a megfigyelt változók között kapcsolat van: például a gépkocsira ható légellenállás függ a gépkocsi haladási sebességétől. Más esetekben a kérdés éppen a kapcsolat létezése: egy ötvöző anyag mennyiségének változása maga után vonja-e az acél szilárdsági tulajdonságainak változását, vagy egyetemi hallgatóknál összefügg-e a felvételi pontszám és a későbbi félévek görgetett átlaga. Ilyen típusú kérdés eldöntéséhez készítsünk megfigyeléssorozatot, amelynek végén rendelkezésünkre áll az összetartozó változó-párokból alkotott minta: $(\xi_1, \eta_1), (\xi_2, \eta_2), \ldots, (\xi_n, \eta_n)$. Számítsuk ki mindkét változó $(\xi_a, \eta_a)$ átlagát, és megfigyeléseinket ábrázoljuk olyan grafikonban, ahol a tengelyekre a változók átlagtól való eltérését rakjuk fel (lásd 7. ábra). Minden ábrázolt pontra számítsuk ki rendezők $c_i$ szorzatát:

$$c_i = (\xi_a - \xi_i)(\eta_a - \eta_i).$$

Ez a szorzat az első és a harmadik síknegyedben pozitív, a másik két síknegyedben negatív. Összegezzük ezeket a szorzatokat minden pontra.

$$C_n^* = \sum_{i=1}^{n} (\xi_a - \xi_i)(\eta_a - \eta_i).$$

Ha a pontok az első és a harmadik negyedben helyezkednek el, akkor $C_n^* > 0$, ha a másik két negyedben, akkor $C_n^* < 0$. Ha pedig a pontok rendezetlenül helyezkednek el az egész síkon, akkor $C_n^* \approx 0$ (lásd 8. ábra). Vagyis $C_n^*$ értéke információt tartalmaz a pontok elhelyezkedésére vonatkozóan. $C_n^*$ nagysága függ attól, hogy hány pontra végeztük az összegzést, és mennyi a változók szórása. Az összehasonlíthatóság érdekében szokás $C_n^*$ értékét a pontok számával és a két változó szórásával normálni.

$$C_n = \frac{\frac{1}{n}\sum_{i=1}^{n} (\xi_a - \xi_i)(\eta_a - \eta_i)}{s_\xi s_\eta}$$

Az eredményül kapott mennyiség a tapasztalati korrelációs együttható, általánosítása pedig a korrelációs együttható.

Ha a $(\xi, \eta)$ véletlen mennyiségeket egyszerre szeretnénk vizsgálni, a két valószínűségi változó **közös eloszlását** kell tekintenünk. Itt is megkülönböztethetünk diszkrét és abszolút folytonos eseteket. Diszkrét esetben $\xi$ illetve $\eta$ lehetséges értékei az $x_1, x_2, \ldots$ illetve $y_1, y_2, \ldots$ számok, az *együttes valószínűségeloszlás*, $p_{i,j} = P(\xi = x_i, \eta = y_j)$ értékeit sokszor táblázatba összefoglalva adjuk meg. Abszolút folytonos esetben az $f : \mathbb{R}^2 \to \mathbb{R}$ *közös sűrűségfüggvény* jellemzi a közös eloszlást, segítségével (mérhető) $T \subset \mathbb{R}^2$ tartományokra megadható annak valószínűsége, hogy a $(\xi, \eta)$ véletlen számpár a $T$ halmazba essen: $P((\xi, \eta) \in T) = \iint_T f(x, y)dxdy$. Ha csak önmagában $\xi$ (vagy $\eta$) viselkedésére vagyunk kiváncsiak, a peremeloszlásokat kell tekintenünk, ha pedig valamilyen, a $(\xi, \eta)$ pártól (determinisztikusan) függő $t(\xi, \eta)$ mennyiség várható értéke érdekes, értelemszerű szummázásokat/integrálásokat kell elvégeznünk. A közös eloszlás/sűrűségfüggvény alaptulajdonságai mellett ezeket a képleteket foglalja össze az alábbi táblázat:

| | diszkrét | folytonos |
|---|---|---|
| | $p_{i,j} \geq 0, \sum\limits_{i,j} p_{i,j} = 1$ | $f(x, y) \geq 0, \iint\limits_{\mathbb{R}^2} f(x, y)\, dxdy = 1$ |
| | peremeloszlások: $p_i^{(1)} = p(\xi = x_i) = \sum\limits_j p_{i,j}$ $p_j^{(2)} = p(\eta = y_j) = \sum\limits_i p_{i,j}$ | peremsűrűség-függvények: $f_\xi(x) = \int\limits_{-\infty}^{\infty} f(x, y)\, dy$ $f_\eta(y) = \int\limits_{-\infty}^{\infty} f(x, y)\, dx$ |
| $E(t(\xi, \eta)) =$ | $\sum\limits_{i,j} t(x_i, y_j)p_{i,j}$ | $\iint\limits_{\mathbb{R}^2} t(x, y)f(x, y)\, dxdy$ |

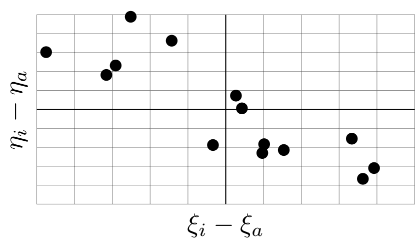
Amennyiben csak $\xi$-től (vagy $\eta$-tól) függő mennyiségeket, kijelentéseket akarunk vizsgálni, számolhatunk a peremeloszlások alapján, így pl. pusztán a peremeloszlásokból meghatározhatóak az $E\xi$ ($E\eta$) várható értékek és a $D\xi$ ($D\eta$) szórások. Általában a közös eloszlás lényegesen több információt hordoz, mint a peremeloszlások; fontos speciális eset azonban a következő: ha $p_{i,j} = p_i^{(1)} p_j^{(2)}$ minden $i, j$ párra, illetve $f(x, y) = f_\xi(x)f_\eta(y)$ minden $(x, y) \in \mathbb{R}^2$ pontra, akkor azt mondjuk, $\xi$ és $\eta$ **függetlenek**. Könnyen ellenőrizhető, hogy függetlenség estén tetszőleges $a < c$ és $b < d$ számpárokra az $A = \{a \leq \xi < c\}$ és a $B = \{b \leq \eta < d\}$ események függetlenek.

Különböző mennyiségek várható értékének, szórásának számításakor hasznos, könnyen ellenőrizhető összefüggések:

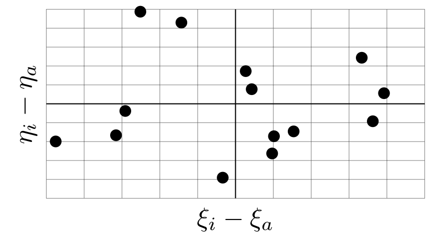
- $E(\xi + \eta) = E\xi + E\eta$ tetszőleges $(\xi, \eta)$ valószínűségi változókra;
- Ha $\xi$ és $\eta$ függetlenek, akkor $D^2(\xi + \eta) = D^2\xi + D^2\eta$;
- Ha $\xi$ és $\eta$ függetlenek, akkor $E(\xi\eta) = E\xi \cdot E\eta$.

Az utóbbi összefüggés motiválja a **kovariancia** fogalmát tetszőleges valószínűségi változók esetén:

$$C(\xi, \eta) = E(\xi\eta) - E\xi \cdot E\eta.$$




7. ábra. Negatív korreláció




8. ábra. Nincs korreláció

Amennyiben $C(\xi, \eta) = 0$, azt mondjuk, $\xi$ és $\eta$ *korrelálatlanok.* A fentiek alapján: ha $\xi$ és $\eta$ függetlenek, akkor korrelálatlanok. Ennek megfordítása azonban nem igaz: általában a korrelálatlanságból nem következik a függetlenség. $C(\xi, \eta) \neq 0$ azt jelenti, hogy $\xi$ és $\eta$ kapcsolatában van valamilyen tendencia: ha $C(\xi, \eta) > 0$, "$\xi$ növelésével $\eta$ is nőni szeret", ha $C(\xi, \eta) < 0$, "$\xi$ növelésével $\eta$ csökkenni szeret".

A $|C(\xi, \eta)| \leq D(\xi) \cdot D(\eta)$ egyenlőtlenség mindig teljesül (megjegyezhetjük, ha felírjuk az $E\xi = E\eta = 0$ speciális esetben, ilyenkor látszik könnyen, hogy a lineáris algebrából és a függvényterek elméletéből jól ismert Cauchy-Schwartz egyenlőtlenségről van szó). Ez motiválja a *korrelációs együttható* bevezetését:

$$R(\xi, \eta) = \frac{C(\xi, \eta)}{D\xi \cdot D\eta}; \qquad \text{melyre} \qquad |R(\xi, \eta)| \leq 1.$$

Egyszerű számolással ellenőrizhető, hogy amennyiben $\xi$ és $\eta$ pontosan lineáris kapcsolatban állnak, $|R(\xi, \eta)| = 1$, vagyis a korreláció abszolút értéke a lehető legnagyobb.

## 2.2. Kidolgozott példák

**2.1. Kidolgozott Feladat.** *Bemegyek a kaszinóba, és elkezdem a rulettpörgetéseket figyelni.*[1] *Addig maradok, amíg első alkalommal fekete nem lesz a pörgetés eredménye.*

*(a) Mi a valószínűsége, hogy pontosan* 5 *pörgetést fogok megfigyelni?*

*(b) Mi a valószínűsége, hogy páros sok pörgetést fogok megfigyelni?*

*(c) Mi a megfigyelt pörgetések várható száma?*

**Megoldás.** Vezessük be a $\xi$ valószínűségi változót, amelynek értéke a szükséges pörgetések száma. Ez a diszkrét valószínűségi változó tetszőleges pozitív egész értéket felvehet. $\xi$ valószínűség-eloszlásához a $\{\xi = k\}$ esemény valószínűségét kell meghatározni ($k \in \mathbb{Z}^+$); ez pontosan akkor következik be, ha $k-1$ nem fekete pörgetést egy fekete pörgetés követ. A rulettkeréken 37 mező van, ezek közül 18 piros, 18 fekete és egy zöld (a nulla). Tehát minden egyes pörgetésre a fekete eredmény valószínűsége $(p =)\frac{18}{37}$, annak valószínűsége, hogy az eredmény nem fekete, $(q = 1 - p =)\frac{19}{37}$. Itt a $p$ és a $q$ paramétereket az egyszerűbb leírásmód, illetve az általánosíthatóság kedvéért vezettük be. Mivel az egyes pörgetések függetlennek tekinthetők, adódik:

$$p_k = P(\xi = k) = q^{k-1}p = \left(\frac{19}{37}\right)^{k-1} \frac{18}{37}. \tag{2}$$

(a) A feladat kérdése éppen $p_5$, amit (2) alapján könnyen számolhatunk:

$$p_5 = (19/37)^4 18/37 \approx 0,034.$$

[1]Tekintsünk egy idealizált kaszinót: tegyük fel, hogy a rulettkeréken szereplő 37 számmező bármelyikébe ugyanakkora eséllyel kerülhet a rulettgolyó, minden egyes pörgetésre.

(b) Ismét (2) alapján számolhatunk:

$$P(\xi \text{ páros}) = p_2 + p_4 + p_6 + \cdots = qp + q^3p + q^5p + \cdots = qp\frac{1}{1-q^2} = \frac{18}{37}\frac{19/37}{1-(18/37)^2} \approx 0,327.$$

Felhasználtuk a mértani sor összegképletét is.

(c) A kérdés a (2) eloszlás várható értéke, tehát

$$E\xi = \sum_{k=1}^{\infty} kp_k = \sum_{k=1}^{\infty} kpq^{k-1} = p\sum_{k=1}^{\infty} kq^{k-1}.$$

Ennek az összegnek a kiszámolásához tekintsük a

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

függvényt. Mivel $q < 1$, ez a mértani sor konvergens, a tagonkénti deriválással kapott sor is konvergens, és éppen a $g'(q)$ deriváltfüggvényt állítja elő. Ebből:

$$\sum_{k=1}^{\infty} kq^{k-1} = g'(q) = \frac{1}{(1-q)^2}.$$

Összefoglalva tehát az eddigieket:

$$E\xi = pg'(q) = \frac{p}{(1-q)^2} = \frac{1}{p} = \frac{37}{18} \approx 2,056.$$

Ennek a kérdésnek a megválaszolásához különösen érdemes volt a feladatot a $p$ parméterrel általánosítani.

**2.2. Kidolgozott Feladat.** *Kör alakú, 10 cm átmérőjű táblába pontszerű lövedék csapódik, a becsapódás helye egyenletes eloszlású a táblán. Jelöljük $\xi$-vel a becsapódás helyének távolságát a tábla középpontjától.*

*(a) Adjuk meg $\xi$ eloszlásfüggvényét!*

*(b) Számoljuk ki a $P(2 < \xi < 7)$ valószínűséget!*

*(c) $E\xi =$?*

**Megoldás.**

(a) $F_\xi(x)$ meghatározásához (a továbbiakban az alsó $\xi$ indexet elhagyjuk) első lépésként megállapíthatjuk, hogy $F(0) = P(\xi < 0) = 0$, hiszen távolság csak pozitív értékű lehet, valamint $F(10) = P(\xi < 10) = 1$, hiszen a táblát a lövedék biztosan eltalálja. $F(x)$ alaptulajdonságai miatt adódik:

$$F(x) = \begin{cases} 0 & \text{ha } x \le 0, \\ ? & \text{ha } 0 < x < 10, \\ 1 & \text{ha } x \ge 10; \end{cases}$$

ahol már csak a középső ágat kell meghatározni. A $P(\xi < x)$ esemény valószínűségére van szükség ($0 < x < 10$), és mivel a becsapódás helye egyenletes eloszlású, geometriai valószínűségi feladatról van szó. A teljes terület egy 10 cm sugarú körlap, a kedvező terület egy $x$ cm sugarú körlap. Tehát:

$$F(x) = \frac{x^2\pi}{10^2\pi} = \frac{x^2}{100} \qquad \text{ha} \qquad 0 < x < 10.$$

Az így adódó eloszlásfüggvényt a 9. ábra szemlélteti.


9. ábra. A 2.2. kidolgozott feladathoz.

(b) Az eloszlásfüggvény alapján könnyen számolható:

$$P(2 < \xi < 7) = P(2 \leq \xi < 7) = F(7) - F(2) = \frac{49-4}{100} = 0,45.$$

Kihasználtuk azt is, hogy folytonos eloszlásról van szó, $P(\xi = 2)$ valószínűsége 0.

(c) A várható érték számolásához szükség lesz a sűrűségfüggvényre:

$$f(x) = F'(x) = \begin{cases} 0 & \text{ha } x \leq 0, \\ \frac{x}{50} & \text{ha } 0 < x < 10, \\ 0 & \text{ha } x \geq 10; \end{cases}$$

ebből:

$$E\xi = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{0}^{10} x \frac{x}{50}\, dx = \left[\frac{x^3}{150}\right]_0^{10} = 20/3 \approx 6,667.$$

**2.3. Kidolgozott Feladat.** *Legyen a $\xi$ valószínűségi változó sűrűségfüggvénye*

$$f(x) = \begin{cases} \frac{A}{x^4} & \textit{ha } x > 1, \\ 0 & \textit{egyébként.} \end{cases}$$

*(a) $A =$?*

*(b) $P(10 < \xi) =$?*

*(c) Számoljuk ki $\xi$ várható értékét és szórását.*

**Megoldás.**

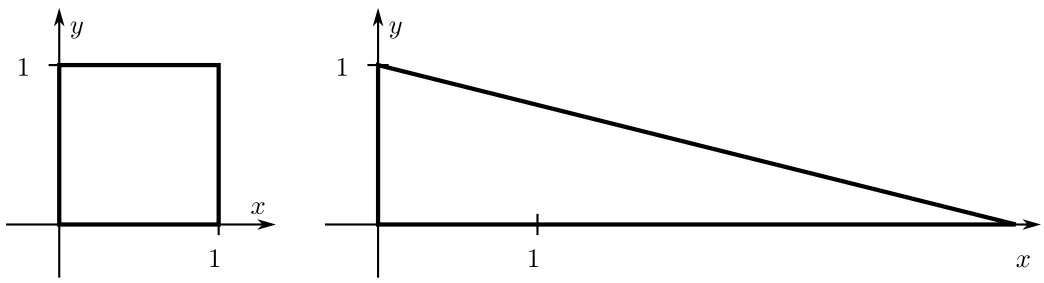
(a) Az $A$ paraméter meghatározásához ki kell használnunk, hogy $f(x)$ sűrűségfüggvény, és így:

$$1 = \int_{-\infty}^{\infty} f(x)\, dx = \int_{1}^{\infty} \frac{A}{x^4}\, dx = \lim_{T\to\infty} \int_{1}^{T} \frac{A}{x^4}\, dx = \lim_{T\to\infty} \left[\frac{-A}{3x^3}\right]_1^T = A/3,$$

tehát $A = 3$.

(b) Abszolút folytonos eloszlás esetén, kihasználva az eloszlásfüggvény és a sűrűségfüggvény közti kapcsolatot:

$$P(10 < \xi) = 1 - F(10) = \int_{10}^{\infty} f(x)\, dx = \lim_{T\to\infty} \int_{10}^{T} \frac{3}{x^4}\, dx = \lim_{T\to\infty} \left[\frac{-1}{x^3}\right]_{10}^T = 0,001.$$




10. ábra. A 2.4. kidolgozott feladathoz.

(c) A definíciók alapján:

$$E\xi = \int_{-\infty}^{\infty} x f(x)\, dx = \lim_{T\to\infty} \int_1^T \frac{3}{x^3}\, dx = \lim_{T\to\infty} \left[\frac{-3}{2x^2}\right]_1^T = \frac{3}{2};$$

$$E\xi^2 = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \lim_{T\to\infty} \int_1^T \frac{3}{x^2}\, dx = \lim_{T\to\infty} \left[\frac{-3}{x}\right]_1^T = 3;$$

$$D^2(\xi) = E\xi^2 - (E\xi)^2 = 3 - (3/2)^2 = 0,75; \qquad \Longrightarrow \qquad D(\xi) = \sqrt{D^2(\xi)} \approx 0,866.$$

**2.4. Kidolgozott Feladat.** *A $(\xi, \eta)$ valószínűségi változó-pár egyenletes eloszlású a $D \subset \mathbb{R}^2$ síkbeli tartományon, ha az együttes sűrűségfüggvény:*

$$f(x,y) = \begin{cases} const. = \frac{1}{t_D} & ha\ (x,y) \in D, \\ 0 & ha\ (x,y) \notin D. \end{cases}$$

*Itt $t_D$ a $D$ tartomány területe, ezzel a választással érjük el, hogy $\iint_{\mathbb{R}^2} f(x,y)\, dxdy = 1$. Számoljuk ki a peremsűrűség-függvényeket, a kovarianciát, a korrelációs együtthatót, és ezek alapján döntsük el, $\xi$ és $\eta$ függetlenek, illetve korrelálatlanok-e, ha*

*(a) $D$ a $(0,0)$, $(1,0)$, $(1,1)$, $(0,1)$ csúcspontú négyzet;*

*(b) $D$ a $(0,0)$, $(4,0)$, $(0,1)$ csúcspontú háromszög.*

*A két esetet a 10. ábra szemlélteti.*

**Megoldás.**

(a) A fenti képlet alapján

$$f(x,y) = \begin{cases} 1 & \text{ha } 0 \le x \le 1 \text{ és } 0 \le y \le 1, \\ 0 & \text{egyébként.} \end{cases}$$

Így a peremsűrűség-függvények:

$$f_\xi(x) = \int_{-\infty}^{\infty} f(x,y)\, dy = \begin{cases} 0 & \text{ha } x < 0, \\ \int_0^1 1\, dy = 1 & \text{ha } 0 \le x \le 1, \\ 0 & \text{ha } x > 1. \end{cases}$$

$$f_\eta(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \begin{cases} 0 & \text{ha } y < 0, \\ \int_0^1 1\,dx = 1 & \text{ha } 0 \le y \le 1, \\ 0 & \text{ha } y > 1. \end{cases}$$

Könnyen ellenőrizhető, hogy $f(x,y) = f_\xi(x) \cdot f_\eta(y)$ minden $(x,y) \in \mathbb{R}^2$ esetén, és így $\xi$ és $\eta$ függetlenek. Így $\xi$ és $\eta$ korrelálatlanok is, tehát $C(\xi,\eta) = 0$ és $R(\xi,\eta) = 0$.

(b) Ezúttal

$$f(x,y) = \begin{cases} 1/2 & \text{ha } 0 \le x \le 4 \text{ és } 0 \le y \le 1 - x/4, \\ 0 & \text{egyébként.} \end{cases}$$

Így a peremsűrűség-függvények:

$$f_\xi(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \begin{cases} 0 & \text{ha } x < 0, \\ \int_0^{1-x/4} 1/2\,dy = 1/2 - x/8 & \text{ha } 0 \le x \le 4, \\ 0 & \text{ha } x > 4. \end{cases}$$

$$f_\eta(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \begin{cases} 0 & \text{ha } y < 0, \\ \int_0^{4-4y} 1/2\,dx = 2 - 2y & \text{ha } 0 \le y \le 1, \\ 0 & \text{ha } y > 1. \end{cases}$$

Mivel például $f(3,1/2) = 0$ de $f_\xi(3) = 1/8$ és $f_\eta(1/2) = 1$, $\xi$ és $\eta$ *nem* függetlenek.

A kovariancia számolásához:

$$\begin{aligned} E(\xi\eta) &= \iint_{\mathbb{R}^2} xyf(x,y)\,dxdy = \int_0^4 \int_0^{1-x/4} \frac{xy}{2}\,dydx = \int_0^4 \frac{x(1-x/4)^2}{4}\,dx = \\ &= \int_0^1 4t^2 - 4t^3\,dt = \left[\frac{4t^3}{3} - t^4\right]_0^1 = 1/3; \end{aligned}$$

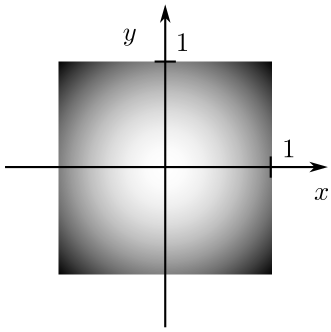
$$E\xi = \int_{-\infty}^{\infty} xf_\xi(x)\,dx = \int_0^4 x(1/2 - x/8)\,dx = \left[x^2/4 - x^3/24\right]_0^4 = 4/3;$$

$$E\eta = \int_{-\infty}^{\infty} yf_\eta(y)\,dy = \int_0^1 y(2-2y)\,dx = \left[y^2 - \frac{2y^3}{3}\right]_0^1 = 1/3;$$

és ezekből adódik:

$$C(\xi,\eta) = E(\xi\eta) - E\xi E\eta = 1/3 - 4/9 = -1/9.$$

A korrelációs együtthatóhoz szükségünk van a szórásokra is:




11. ábra. A 2.5. kidolgozott feladathoz.

$$E\xi^2 = \int_{-\infty}^{\infty} x^2 f_\xi(x)\,dx = \int_0^4 x^2(1/2 - x/8)\,dx = \left[x^3/6 - x^4/32\right]_0^4 = 8/3;$$

$$D^2(\xi) = E\xi^2 - (E\xi)^2 = 8/3 - (4/3)^2 = 8/9; \qquad \Longrightarrow \qquad D(\xi) = \sqrt{D^2(\xi)} \approx 0,943.$$

$$E\eta^2 = \int_{-\infty}^{\infty} y f_\eta(y)\,dy = \int_0^1 y^2(2-2y)\,dx = \left[\frac{2y^3}{3} - y^4/2\right]_0^1 = 1/6;$$

$$D^2(\eta) = E\eta^2 - (E\eta)^2 = 1/6 - (1/3)^2 = 1/18; \qquad \Longrightarrow \qquad D(\xi) = \sqrt{D^2(\xi)} \approx 0,553.$$

Mindezek alapján:

$$R(\xi,\eta) = \frac{C(\xi,\eta)}{D(\xi)D(\eta)} = \frac{C(\xi,\eta)}{\sqrt{D^2(\xi)D^2(\eta)}} = -1/2.$$

**2.5. Kidolgozott Feladat.** *Legyen a $(\xi,\eta)$ valószínűségi változó-pár együttes sűrűségfüggvénye:*

$$f(x,y) = \begin{cases} A(x^2+y^2) & \textit{ha } |x| \le 1 \textit{ és } |y| \le 1, \\ 0 & \textit{egyébként.} \end{cases}$$

*Ahogyan azt a 11. ábra is szemlélteti, az eloszlás ezúttal nem egyenletes, a négyzet sarkai felé növekszik a sűrűségfüggvény értéke.*

*(a) $A =$?*

*(b) Független-e $\xi$ és $\eta$?*

*(c) Korrelálatlan-e $\xi$ és $\eta$?*

**Megoldás.**

(a) Ismét a sűrűségfüggvény alaptulajdonságát használjuk fel:

$$\begin{aligned} 1 &= \iint_{\mathbb{R}^2} f(x,y)\,dxdy = \int_{-1}^{1}\int_{-1}^{1} A(x^2+y^2)\,dydx = \int_{-1}^{1}\left[Ax^2y + \frac{Ay^3}{3}\right]_{y=-1}^{y=1} dx = \\ &= \int_{-1}^{1}\left(2Ax^2 + \frac{2A}{3}\right) dx = \left[\frac{2Ax^3}{3} + \frac{2Ax}{3}\right]_{x=-1}^{x=1} = \frac{8A}{3} \end{aligned}$$

és így $A = 3/8$.

(b) A kérdés megválaszolásához határozzuk meg a peremsűrűség-függvényeket:

$$f_\xi(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \begin{cases} 0 & \text{ha } x < -1, \\ \int_{-1}^{1} \frac{3}{8}(x^2+y^2)\,dy = \frac{1}{4} - \frac{3x^2}{4} & \text{ha } -1 \le x \le 1, \\ 0 & \text{ha } x > 1. \end{cases}$$

$$f_\eta(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \begin{cases} 0 & \text{ha } y < -1, \\ \int_{-1}^{1} \frac{3}{8}(x^2+y^2)\,dx = \frac{1}{4} - \frac{3y^2}{4} & \text{ha } -1 \le y \le 1, \\ 0 & \text{ha } y > 1. \end{cases}$$

$\xi$ és $\eta$ nem függetlenek, mert például $f(0,0) = 0$, de $f_\xi(0) = f_\eta(0) = 1/4$.

(c) Az alábbi integrálok mindegyike 0, mert páratlan függvényt integrálunk origóra szimmetrikus intervallumon:

$$E(\xi\eta) = \iint_{\mathbb{R}^2} xyf(x,y)\,dxdy = \int_{-1}^{1}\int_{-1}^{1} \frac{3}{8}xy(x^2+y^2)\,dydx = 0;$$

$$E\xi = \int_{-\infty}^{\infty} xf_\xi(x)\,dx = \int_{-1}^{1} \frac{1}{4}x(1-3x^2)\,dx = 0;$$

$$E\eta = \int_{-\infty}^{\infty} yf_\eta(y)\,dy = \int_{-1}^{1} \frac{1}{4}y(1-3y^2)\,dy = 0.$$

Így tehát $C(\xi,\eta) = E(\xi\eta) - (E\xi)(E\eta) = 0$, vagyis $\xi$ és $\eta$ korrelálatlanok.

## 2.3. Gyakorló feladatok

**2.6. Feladat.** *Feldobunk két szabályos dobókockát. Jelöljük $\xi$-vel a dobott számok összegét. Határozzuk meg a $\xi$ valószínűségi változó várható értékét és szórását!*

**2.7. Feladat.** *A $\xi$ folytonos valószínűségi változó sűrűségfüggvénye:*

$$f(x) = \begin{cases} B\sin x & \textit{ha} \ \ 0 < x < \frac{\pi}{2}, \\ 0 & \textit{egyébként.} \end{cases}$$

*(a) Mennyi $B$ értéke?* *(b) Számoljuk ki a $P\left(0 < \xi < \frac{\pi}{3}\right)$ valószínűséget!*
*(c) Mennyi $\xi$ várható értéke?* *(d) Mennyi $\xi$ szórása?*

**2.8. Feladat.** *A $(\xi, \eta)$ diszkrét valószínűségi változó–pár közös eloszlását az alábbi táblázat írja le.*

| | $\xi = -1$ | $\xi = 0$ | $\xi = 1$ |
|---|---|---|---|
| $\eta = 0$ | *0,1* | $p$ | *0,2* |
| $\eta = 1$ | $p$ | *0,3* | $2p$ |

*(a) $p =$?* *(b) Független–e $\xi$ és $\eta$?* *(c) Számoljuk ki az $R(\xi, \eta)$ korrelációs együtthatót!*

**2.9. Feladat.** *A $(\xi, \eta)$ valószínűségi változók együttes eloszlását az*

$$f(x, y) = \begin{cases} B\,(x^4 + y^6) & \textit{ha} \ \ -1 \le x \le 1 \ \ \textit{és} \ -1 \le y \le 1, \\ 0 & \textit{egyébként.} \end{cases}$$

*közös sűrűségfüggvény írja le.*
*(a) $B =$?* *(b) Számoljuk ki a peremsűrűség-függvényeket!* *(c) Független-e $\xi$ és $\eta$?*

**2.10. Feladat.** *A $(\xi, \eta)$ valószínűségi változó-pár egyenletes eloszlású az origó középpontú, egységnyi sugarú körlapon.*
*(a) Határozzuk meg a peremsűrűség-függvényeket!* *(b) Számoljuk ki a korrelációs együtthatót!*

# 3. Nevezetes eloszlások

## 3.1. Elméleti összefoglaló

A természetben és a műszaki életben leginkább előforduló valószínűségi változók sokszor valamilyen ismert, egy-két paraméterrel jellemezhető eloszlást követnek. Az alábbiakban néhány ilyen nevezetes eloszlás tulajdonságait foglaljuk össze.

A diszkrét eloszlások közül csak a két legfontosabbat vizsgáljuk meg. Legyen $n \in \mathbb{Z}^+$ és $p \in (0, 1)$ rögzítve. A $\xi$ diszkrét valószínűségi változó $(n, p)$ paraméterű **binomiális eloszlású**, ha lehetséges értékei a $k = 0, 1 \ldots n$ számok, valószínűségeloszlása pedig:

$$p_k^{(n,p)} = P(\xi = k) = \binom{n}{k} p^k (1-p)^{(n-k)} \qquad (k = 0, 1 \ldots, n).$$

A binomiális eloszlás jelentése: elvégezzük $n$-szer ugyanazt a kísérletet, az egyes próbálkozásaink függetlenek. Minden egyes kísérletet sikeresnek tekintünk, ha az egy bizonyos eredménnyel végződik. Tegyük fel, hogy az egyes kísérletek külön-külön $p$ eséllyel adhatják ezt az eredményt. $\xi$ jelöli a sikerek számát: azt a (véletlentől függő) számot, ahány alkalommal (az $n$ független kísérletből) a kívánt eredmény adódott. Például: feldobunk 10-szer egy szabályos dobókockát, $\xi$ jelölje a hatos dobások számát ($n = 10, p = 1/6$).

A binomiális tétel alapján ellenőrizhető, hogy valóban valószínűség-eloszlást adtunk meg:

$$\sum_{k=0}^{n} p_k^{(n,p)} = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{(n-k)} = (p + 1 - p)^n = 1.$$

A binomiális eloszlás várható értéke:

$$E\xi = \sum_{k=0}^{n} k p_k^{(n,p)} = \sum_{k=1}^{n} \frac{n(n-1)\dots(n-k)}{(k-1)!} p^k (1-p)^{((n-1)-(k-1))} = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{(n-1-j)} = np.$$

(Ez szemléletesen is érthető: a sikerek várható száma a próbálkozások számának és az egyedi kísérletben a siker valószínűségének a szorzata.) Hasonló számolással kaphatjuk meg a szórást: $D\xi = \sqrt{np(1-p)}$.

Legyen most $\lambda > 0$ rögzített: az $\eta$ valószínűségi változó $\lambda$ paraméterű **Poisson eloszlású**, ha értéke tetszőleges $k = 0, 1, 2, \dots$ természetes szám lehet, és valószínűségeloszlása:

$$p_k^{(\lambda)} = P(\eta = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

A Poisson eloszlást a binomiális eloszlás határeseteként is megkaphatjuk, amennyiben $n \to \infty$, $p \to 0$, úgy, hogy $np \to \lambda$. (Belátható, hogy ekkor tetszőleges rögzített $k$-ra $p_k^{(n,p)} \to p_k^{(\lambda)}$.) Szavakban kifejezve a Poisson eloszlású valószínűségi változó jelentése: nagyon sok, egymástól független, külön-külön nagyon kis valószínűségű eseményből ahány bekövetkezik. Ilyen véletlen mennyiségek a természetben és a műszaki életben is gyakran előfordulnak: pl. anyaghibák száma térfogategységnyi mintában, téves kapcsolások száma egy nagy telefonközpont napi forgalmában, egy év alatt ahány baleset előfordul egy forgalmas útkereszteződésnél, stb.

Ezúttal az exponenciális függvény 0 körüli Taylor sorára hivatkozva ellenőrizhető, hogy tényleg valószínűség-eloszlást adtunk meg:

$$\sum_{k=0}^{\infty} p_k^{(\lambda)} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

A várható értékre:

$$E\eta = \sum_{k=0}^{\infty} k p_k^{(\lambda)} = e^{-\lambda} \sum_{k=1}^{\infty} \lambda \frac{\lambda^{(k-1)}}{(k-1)!} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda.$$

Hasonló számolással adódik $D\eta = \sqrt{\lambda}$.

Térjünk át a nevezetes folytonos eloszlásokra. Legyenek $a < b$ valós paraméterek, ekkor $\xi$ az $[a, b]$ **intervallumon egyenletes eloszlású**, ha sűrűségfüggvénye:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{ha } a \le x \le b, \\ 0 & \text{egyébként.} \end{cases}$$

Egyenletes eloszlások leginkább geometriai jellegű problémák során szoktak előkerülni.

**3.1. Megjegyzés.** *Az intervallumon egyenletes eloszlások mellett érdemes megemlíteni az adott tartományon egyenletes (többdimenziós) eloszlásokat, ld. a 2.4 kidolgozott feladatot.*

Könnyű integrálással adódik az egyenletes eloszlás eloszlásfüggvénye:

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \begin{cases} 0 & \text{ha } x < a, \\ \frac{x-a}{b-a} & \text{ha } a \le x \le b, \\ 1 & \text{ha } x > b. \end{cases}$$

A várható érték pedig:

$$E\xi = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{a}^{b} \frac{x}{b-a}\, dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

ami intuíciónknak megfelelően az intervallum felezőpontja. Hasonló számolással: $D\xi = \frac{b-a}{2\sqrt{3}}$.

Legyen paraméterünk ismét $\lambda > 0$; ekkor a $\tau$ valószínűségi változó $\lambda$ paraméterű **exponenciális eloszlású**, ha sűrűségfüggvénye:

$$f_\tau(x) = \begin{cases} 0 & \text{ha } x < 0, \\ \lambda e^{-\lambda x} & \text{ha } x \geq 0; \end{cases}$$

és ebből integrálással eloszlásfüggvénye:

$$F_\tau(x) = \int_{-\infty}^{x} f_\tau(t)\, dt = \begin{cases} 0 & \text{ha } x < 0, \\ 1 - e^{-\lambda x} & \text{ha } x \geq 0. \end{cases}$$

Az exponenciális eloszlás élettartamok, várakozási idők, általában egy esemény bekövetkezéséig eltelő véletlen időtartamok hosszának jellemzésekor szokott előkerülni. Ezzel összefügg, hogy értéke csak pozitív lehet ($F(0) = 0$). Az exponenciális eloszlás legfontosabb tulajdonsága az úgynevezett *örökifjúság*. Legyenek $T_1$ és $T_2$ tetszőleges pozitív számok, és tekintsük az $A = \{\tau \geq T_1\}$; $B = \{\tau \geq T_2\}$; $C = \{\tau \geq T_1 + T_2\}$ eseményeket. Egyrészt $P(A) = P(\tau \geq T_1) = 1 - F_\tau(T_1) = e^{-\lambda T_1}$, és $P(B)$-t és $P(C)$-t számolhatjuk hasonlóan. Másrészt $BC = C$, és így

$$P(\tau \geq T_1 + T_2 | \tau \geq T_2) = \frac{P(BC)}{P(B)} = \frac{e^{-\lambda(T_1+T_2)}}{e^{-\lambda T_2}} = e^{-\lambda T_1} = P(\tau \geq T_1).$$

Szavakban kifejezve: annak valószínűsége, hogy $T_2$ idő várakozás után még további $T_1$ ideig várnunk kell, ugyanannyi, mint a várakozás kezdetében volt annak valószínűsége, hogy $T_1$ időt kell várnunk. Úgy is mondhatnánk, hogy az exponenciális eloszlásnak nincs memóriája, a további várakozási esélyeket nem befolyásolja az, hogy már valamennyi időt vártunk.

Az exponenciális eloszlás várható értékét parciális integrálással kaphatjuk meg:

$$E\tau = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{0}^{\infty} \lambda x e^{-\lambda x}\, dx = \lim_{T\to\infty} \left( \left[-x e^{-\lambda x}\right]_0^T + \int_0^T e^{\lambda x}\, dx \right) = \lim_{T\to\infty} \left[ -\frac{e^{-\lambda x}}{\lambda} \right]_0^T = 1/\lambda.$$

A szóráshoz kétszer kell parciálisan integrálni, $D\tau = 1/\lambda$ adódik.

A nevezetes eloszlások közül kétségkívül a **normális eloszlás**sal lehet leginkább találkozni a természetben és a műszaki életben. A *standard normális eloszlás* sűrűségfüggvénye:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

a jól ismert haranggörbe, eloszlásfüggvénye pedig $\Phi(x) = \int\limits_{-\infty}^{x} \varphi(t)\, dt$. $\Phi(x)$ értékeit táblázatban szokták megadni, fontos tudni, hogy $\Phi(-x) = 1 - \Phi(x)$. Hogy a haranggörbe valóban sűrűségfüggvény, azt a következőképp láthatjuk:

$$\left( \int_{-\infty}^{\infty} \varphi(x)\, dx \right)^2 = \frac{1}{2\pi} \iint_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}}\, dx dy = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} r\, dr d\varphi = 1.$$

A standard normális eloszlás várható értéke 0, szórása pedig 1 (a megfelelő integrálok konvergenciája könnyen látszik, a várható érték esetén egy páratlan függvényt integrálunk a teljes számegyenesre, a szórásnégyzet pedig parciális integrálással számolható.)

**3.2. Megjegyzés.** *Általában azt mondjuk, egy $\xi$ valószínűségi változó standard, ha várható értéke* 0*, szórása pedig* 1*.*

Legyenek $m \in \mathbb{R}$ és $\sigma > 0$ paraméterek; azt mondjuk, *X m várható értékű és σ szórású normális eloszlást* követ, ha a megfelelő lineáris átskálázottja, $Y = \frac{X-m}{\sigma}$, standard normális eloszlású. Ennek megfelelően $X$ eloszlás- és sűrűségfüggvénye ilyenkor:

$$\Phi_{m,\sigma}(x) = \Phi\left(\frac{x-m}{\sigma}\right); \qquad \varphi_{m,\sigma}(x) = \Phi'_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Azt, hogy *X m várható értékű és σ szórású normális eloszlást* követ, röviden így is szoktuk jelölni: $X \in \mathcal{N}(m, \sigma)$. A normális eloszlás fontos tulajdonsága az ún. *stabilitás*: ha $X_1$ és $X_2$ független, normális eloszlású valószínűségi változók, akkor összegük, $X_1 + X_2$ is normális eloszlású (és a 2. fejezetnek megfelelően a várható értékek és a szórásnégyzetek összeadódnak). Pontosabban, legyen $X_1 \in \mathcal{N}(m_1, \sigma_1)$ és $X_2 \in \mathcal{N}(m_2, \sigma_2)$, valamint $a \in \mathbb{R}$ és $b \in \mathbb{R}$ tetszőleges, ekkor $aX_1 + bX_2 \in \mathcal{N}(am_1 + bm_2, \sqrt{a^2\sigma_1^2 + b^2\sigma_2^2})$. Fontos, hogy negatív $a$ és/vagy $b$ esetén is ezt a képletet kell alkalmazni.

Mint említettük, a normális eloszlás rendkívül gyakran előfordul a természetben és a műszaki életben: a mérési eredményekben, műszaki adatokban jelentkező ingadozások is jellemzően normális eloszlást követnek. Ennek elsődleges oka a 4. fejezetben tárgyalt centrális határeloszlás-tétel.

További, a műszaki alkalmazásokban előkerülő eloszlások:

- A $\xi$ valószínűségi változó $(m, \sigma)$ paraméterű *lognormális eloszlású*, ha $\ln \xi$ normális eloszlású ugyanezekkel a paraméterekkel. Sűrűségfüggvénye:

$$f(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \cdot e^{-\frac{(\ln x - m)^2}{2\sigma^2}} & \text{ha } x > 0; \\ 0 & \text{egyébként.} \end{cases}$$

  A természetben és műszaki alkalmazásokban a logonormális eloszlás gyakran előkerül, ha egy véletlen mennyiség csak pozitív értékeket vehet fel, és várható értéke viszonylag kicsi a szórásához képest.

- Az $(\eta, \beta)$ paraméterű *Weibull-eloszlás* ($\eta$ és $\beta$ pozitív paraméterek) sűrűségfüggvénye:

$$f(x) = \begin{cases} \frac{\beta}{\eta}\left(\frac{x}{\eta}\right)^{\beta-1} e^{-\left(\frac{x}{\eta}\right)^\beta} & \text{ha } x \geq 0; \\ 0 & \text{egyébként.} \end{cases}$$

  A $\beta = 1$ speciális esetben az $1/\eta$ paraméterű exponenciális eloszlást kapjuk vissza. Ezt az eloszlást is élettartamok leírására használják, abban az esetben, ha az örökifjúsági feltevés nem állja meg a helyét (a $\beta$ paraméter jellemzi, hogy mennyire térünk el az örökifjúságtól).

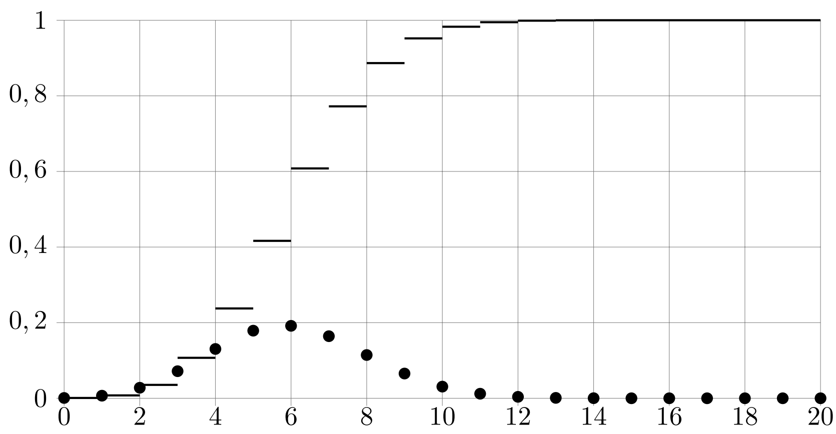
- Az $(n, \lambda)$ paraméterű *Gamma eloszlás* annak a véletlen mennyiségnek az eloszlása, amelyet $n$ független, egyaránt $\lambda > 0$ paraméterű exponenciális eloszlású valószínűségi változó összegeként kapunk meg. Sűrűségfüggvénye:

$$f(x) = \begin{cases} \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} & \text{ha } x \geq 0; \\ 0 & \text{egyébként.} \end{cases}$$

**Nevezetes eloszlások szemléltetése, alkalmazások.** A közismert Excel program a legtöbb nevezetes eloszlást beépített függvényként ismeri, melyek egyszerű függvényutasításokkal kezelhetőek. Például a binomiális eloszlás valószínűség-eloszlásának és eloszlásfüggvényének számítására szolgál a

$$= \text{BINOM.ELOSZLÁS}(k; n; p; \text{IGAZ/HAMIS})$$

utasítás. Itt $k$ a kedvező esetek száma, $n$ az összes esetek száma, és $p$ az esemény valószínűsége. A logikai változó IGAZ értékénél az eloszlásfüggvény, a HAMIS értékénél a valószínűségeloszlás értékét kapjuk vissza. A program használatával könnyen ábrázolhatjuk ezeket a függvényeket (lásd 12. ábra).




12. ábra. A binomiális eloszlás szemléltetése. A monoton, ugró függvény az eloszlásfüggvény. A pontok a konkrét értékek valószínűségei. A paraméterek $n = 20$ és p$=0,3$.

Lássunk két példát a binomiális eloszlás alkalmazására.

Egy üzemben 50 gép dolgozik egymástól függetlenül. Annak a valószínűsége, hogy egy gép dolgozik: $p = 0.6$. Ha az üzemvitelt 100%-os biztonsággal akarjuk biztosítani, akkor az 50 gép együttes teljesítmény-felvételére kell kiépíteni az elektromos hálózatot. Kérdés, hogy ha elegendő 99%-os biztonság, akkor legfeljebb hány gép dolgozhat egyszerre? A binomiális eloszlás eloszlásfüggvényét az Excel programmal számolva azt kapjuk, hogy 38 gép együttes működésének valószínűsége már 99% feletti:

$$= \mathrm{BINOM.ELOSZLÁS}(38; 50; 0.6; \mathrm{IGAZ}) = 0.9943.$$

Vagyis, ha 99%-os biztonság elegendő, akkor az 50 helyett csak 38 gép együttes teljesítmény-felvételére kell méreteznünk az elektromos hálózatot.

Minőségellenőrzés területén is gyakran használják a binomiális eloszlást. Egy sorozatgyártásban legyen a selejtes termék gyártásának valószínűsége $p$. A selejtarányt időről időre úgy ellenőrzik, hogy $n$ elemű mintát vesznek a gyártmányból, és minden mintaelemet megvizsgálnak. Annak a $P$ valószínűsége, hogy az $n$ elemű mintából legfeljebb $k$ selejtet találnak, feltéve, hogy a gyártásban valóban $p$ a selejtszázalék, a binomiális eloszlás segítségével kiszámítható. Ha $P$ értékét elég nagyra választjuk (például $P = 99,9\%$), és a mintában több mint $k$ darab selejtet találunk, akkor joggal feltételezhető, hogy a gyártásban a selejtarány $p$-nél nagyobb. Számokkal bemutatva a fentieket: tegyük fel, hogy a gyártásban $p = 3\%$ a selejtarány. Megvizsgálunk egy $n = 100$ elemű mintát. Annak a valószínűsége, hogy a mintában legfeljebb 9 selejtes darabot találunk 99,91%:

$$= \mathrm{BINOM.ELOSZLÁS}(9; 100; 0.03; \mathrm{IGAZ}) = 0.99912$$

Vagyis, ha a 100 elemű mintában több mint 9 selejtet találunk, akkor feltehetjük, hogy 3%-nál nagyobb a gyártás selejtaránya (a várható selejtszám a mintában $np = 3$ darab).

A normális eloszlás sűrűség- és eloszlásfüggvényének számítására szolgál a

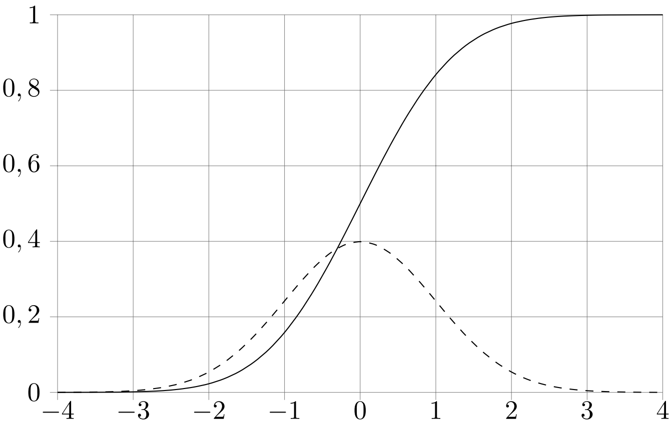
$$= \mathrm{NORM.ELOSZL}(x; m; \sigma; \mathrm{IGAZ/HAMIS})$$

utasítás az Excel programban. Itt $m$ a várható érték, $\sigma$ a szórás, és a logikai változó IGAZ értékénél az eloszlásfüggvény, a HAMIS értékénél a sűrűségfüggvény értékét számolja ki az utasítás. Mondjuk, hogy egy 80 mm névleges átmérőjű darabot sorozatban gyártanak, a gyártott átmérő normális eloszlású valószínűségi változó, szórása 0,5 mm. A mellékelt ábrán bemutatjuk az átmérő eloszlás- és sűrűségfüggvényét. Könnyen kiszámíthatjuk, hogy mekkora a valószínűsége annak, hogy a gyártott darab átmérője az $m \pm 2\sigma$ intervallumba essen:

$$= \text{NORM.ELOSZL}(79; 80; 0.5; \text{IGAZ}) = 0.02275$$
$$= \text{NORM.ELOSZL}(81; 80; 0.5; \text{IGAZ}) = 0.97725$$

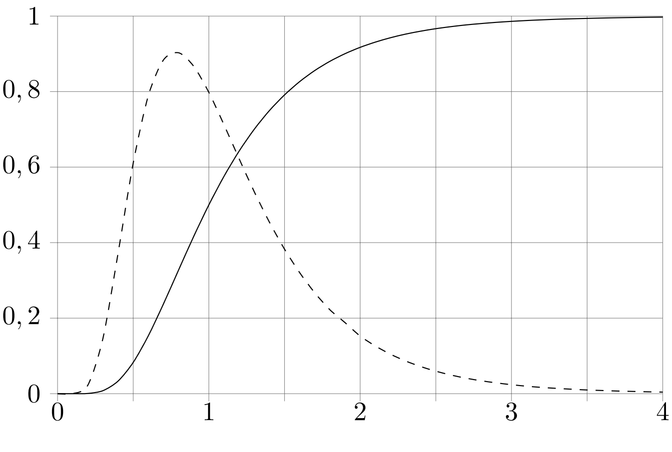
A két érték különbsége $0,9545$, vagyis annak valószínűsége, hogy a gyártott $D$ átmérő a $79 - 81$ mm intervallumba essen:

$$P(79 \leq D \leq 81) = 0.9545.$$




13. ábra. A standard normális eloszlás sűrűség- (szaggatott) és eloszlásfüggvénye

Hasonló beépített függvény van az exponenciális, a lognormális, a normális, a standard normális és a Poisson eloszlásra is.




14. ábra. A lognormális eloszlás sűrűség- (szaggatott) és eloszlásfüggvénye $m = 0$, $\sigma = 0,5$ paraméterekkel

## 3.2. Kidolgozott példák

**3.1. Kidolgozott Feladat.** *Reggelente az Óperenciai Közlekedési Vállalat 13-as buszával járok dolgozni. Ezen a vonalon a buszok harmada légkondicionált. Mi a valószínűsége, hogy a hét öt munkanapjából legfeljebb egyszer kell légkondicionálás nélküli busz miatt bosszankodnom?*

**Megoldás.** Jelölje $\xi$-vel azon reggelek számát, amikor légkondicionált buszhoz van szerencsém. Az egyes napokat tekintsük függetlennek, ezért $\xi$ megmutatja, hogy 5 független próbálkozásból hányszor következik be egy $1/3$ valószínűségű esemény. Így $\xi$ binomiális eloszlású, $p = 1/3$, $n = 5$ paraméterekkel.

A feladat kérdése: $P(\xi \geq 4) =?$ A választ a binomiális eloszlás képlete alapján határozzuk meg:

$$P(\xi \geq 4) = P(\xi = 5) + P(\xi = 4) = \left(\frac{1}{3}\right)^5 + 5\,\frac{2}{3}\left(\frac{1}{3}\right)^4 \approx 0,0453.$$

**3.2. Kidolgozott Feladat.** *A kertünkben álló meggyfát sok kukac támadja meg, de szerencsére bő a termés: tapasztalataim szerint egy kosár meggyben* $\frac{1}{e} \approx 0,37$ *eséllyel egyáltalán nincs kukacos szem.*

*(a) Mi a valószínűsége, hogy egy adott kosár meggyben pontosan két kukacos szemet fogok találni?*

*(b) Mi a valószínűsége, hogy egy adott kosár meggyben legalább két kukacos szemet fogok találni?*

**Megoldás.** Jelölje $\xi$ az egy kosár meggyben található kukacos szemek számát. Észrevételek:

- a fát sok kukac támadja meg,
- az egyes kukacok viselkedése függetlennek tekinthető,
- egy konkrét kukacra (a sok támadó közül) annak valószínűsége, hogy ez a kukac épp az én kosaramban kössön ki, nagyon kicsi, hiszen az én kosaram a teljes meggytermésnek csak töredéke.

A $\xi$ valószínűségi változó megmutatja, hogy sok, független, kis valószínűségű esemény közül – az egyes kukacok épp az én kosaramba jutnak – hány következtik be. Tehát $\xi$ Poisson eloszlásúnak tekinthető. Meg kell határoznunk a $\lambda$ paramétert. A tapasztalatok szerint $P(\xi = 0) = \frac{1}{e}$, másrészt a Poisson eloszlás képlete szerint $P(\xi = k) = e^{-\lambda}\frac{\lambda^k}{k!}$, így

$$\frac{1}{e} = P(\xi = 0) = e^{-\lambda} \qquad \Longrightarrow \qquad \lambda = 1.$$

(a) A feladat kérdése: $P(\xi = 2) =?$ A Poisson eloszlás képlete alapján:

$$P(\xi = 2) = e^{-\lambda}\frac{\lambda^2}{2!} \approx \frac{0,37}{2} = 0,185$$

(b) A feladat kérdése: $P(\xi \geq 2) =?$ A komplementer esemény valószínűségét érdemes számolni:

$$P(\xi \geq 2) = 1 - P(\xi = 1) - P(\xi = 0) = 1 - e^{-\lambda} - \lambda e^{-\lambda} \approx 1 - 0,37 - 0,37 = 0,26.$$

**3.3. Kidolgozott Feladat.** *A HOMÁLY villanykörték élettartama* 1000 *óra – valójában az élettartam véletlen mennyiség, exponenciális eloszlással,* 1000 *óra várható értékkel.*

*(a) Veszek egy HOMÁLY körtét a boltban. Mi a valószínűsége, hogy* 1500 *óra használat alatt nem ég ki?*

*(b) HOMÁLY körtémet már* 1000 *órán át használtam. Mi a valószínűsége, hogy még további* 1500 *óráig fogom tudni használni?*

*(c) Az alagsori mosdó lámpájába, melyet folyamatosan bekapcsolt állapotban tartanak, hétfőn* 0.00*-kor szerelnek be egy HOMÁLY villanykörtét. Mi a valószínűsége, hogy valamikor hétvégén fog kiégni?*

**Megoldás.** Jelölje $\tau$ a HOMÁLY villanykörte (órákban mért) élettartamát. $\tau$ exponenciális eloszlású, tehát sűrűség- és eloszlásfüggvényét ismerjük, csak a $\lambda$ paramétert kell még meghatároznunk. Tudjuk, hogy exponenciális eloszlásra $E\tau = 1/\lambda$. Így:

$$1000 = E\tau = 1/\lambda \qquad \Longrightarrow \qquad \lambda = 0,001.$$

(a) A feladat kérdése: $P(\tau \geq 1500) =?$ Az eloszlásfüggvény képlete alapján:

$$P(\tau > 1500) = 1 - F(1500) = e^{-\lambda 1500} = e^{-1{,}5} \approx 0,0223$$

(b) A feladat kérdése: $P(\tau > 2500|\tau > 1000) =?$ Az exponenciális eloszlás örökifjúsága, valamint az előző részfeladat eredménye alapján:

$$P(\tau > 2500|\tau > 1000) = P(\tau > 1500) \approx 0,0223.$$

(c) Próbáljuk meg $\tau$-ra vonatkozó egyenlőtlenségekre átfogalmazni a feladat kérdését. A villanykörte a (beszerelését követő) első hétvégén ég ki, ha egyrészt $\tau > 5\cdot 24$ (kibírja az első 5 munkanapot), viszont $\tau < 7 \cdot 24$ (a 7. nap végét már nem éli meg). Hasonló okoskodással, a villanykörte a (beszerelését követő) $k$. hétvégén ég ki, ha $(k-1)\cdot 168 + 120 < \tau < k \cdot 168$.

Jelölje $A_k$ azt az eseményt, hogy a villanykörte éppen a (beszerelését követő) $k$. hétvégén ég ki ($k = 1, 2, \dots$). A fenti érvelés, valamint $\tau$ eloszlásfüggvénye alapján:

$$\begin{aligned} P(A_k) &= P((k-1)168 + 120 < \tau < k168) = F(k168) - F((k-1)168 + 120) = \\ &= e^{-\lambda(168(k-1)+120)} - e^{-\lambda 168 k} = e^{-0{,}168(k-1)}(e^{-0{,}12} - e^{-0{,}168}). \end{aligned}$$

Ugyanakkor az $A_k$ események páronként diszjunktak, így:

$$\begin{aligned} P(\text{A körte hétvégén ég ki}) &= \sum_{k=1}^{\infty} P(A_k) = \sum_{k=1}^{\infty} e^{-0{,}168(k-1)}(e^{-0{,}12} - e^{-0{,}168}) = \\ &= \frac{e^{-0{,}12} - e^{-0{,}168}}{1 - e^{-0{,}168}} \approx 0,269. \end{aligned}$$

**3.4. Kidolgozott Feladat.** *Nagymamám egy fazék levest főzött, ennek mennyisége várhatóan* $4,5$ *l*, $4$ *cl szórással. Unokatestvéreim, András, Béla és Cili már vettek belőle egy-egy tányérral, várhatóan fejenként* $0,5$ *litert; a szórás a fiúk esetében* $2$ *cl, Cilinél* $1$ *cl. A szereplő véletlen folyadékmennyiségek függetlennek és normális eloszlásúnak tekinthetők. Mi a valószínűsége, hogy* $2,9$ *liternél kevesebb leves maradt a fazékban?*

**Megoldás.** Vezessük be a következő valószínűségi változókat: $X$ a nagymamám által főzött leves, $Y_1$, $Y_2$ és $Y_3$ rendre az András, Béla és Cili által kivett, $Z$ pedig a fazékban maradt mennyiség, centiliterben kifejezve. Nyílván:

$$Z = X - Y_1 - Y_2 - Y_3$$

és mivel $X$, $Y_1$, $Y_2$ és $Y_3$ függetlenek és normális eloszlásúak, így $Z$ is normális eloszlású. Továbbá:

$$EZ = EX - EY_1 - EY_2 - EY_3 = 450 - 50 - 50 - 50 = 300$$

és

$$D^2Z = D^2X + D^2Y_1 + D^2Y_2 + D^2Y_3 = 16 + 4 + 4 + 1 = 25 \qquad \Longrightarrow \qquad DZ = 5,$$

tehát $Z \in \mathcal{N}(300, 5)$. A feladat kérdése: $P(Z < 290) =?$ $Z$ eloszlásfüggvényét visszavezethetjük a standard normálisra, majd használjuk az arra vonatkozó táblázatot:

$$P(Z < 290) = \Phi_{300,5}(290) = \Phi\left(\frac{290 - 300}{5}\right) = \Phi(-2) = 1 - \Phi(2) = 1 - 0,9772 = 0,0228.$$

**3.5. Kidolgozott Feladat.** *Egy* $X$ *véletlen mennyiség normális eloszlású* $m$ *várható értékkel és* $\sigma$ *szórással. A mennyiség értéke nagy valószínűséggel kevéssé tér el* $m$*-től: a* $\sigma$*-nál lényegesen nagyobb eltéréseknek különösen kicsi az esélye. Határozzuk meg, legalább mennyinek kell* $\alpha$*-t választani, hogy teljesüljön: a várható értéktől való* $\alpha \cdot \sigma$*-nál nagyobb eltérés valószínűsége kisebb, mint egy százalék!*

**Megoldás.** A feladat kérdése: legalább mennyi $\alpha$, ha $P(|X-m| > \alpha \cdot \sigma) < 0,01$? Ehhez először határozzuk meg, mennyinek adódna ez a valószínűség, ha ismernénk $\alpha$-t. A komplementer esemény valószínűségét tudjuk közvetlenül számolni. Mivel $X \in \mathcal{N}(m, \sigma)$:

$$\begin{aligned} P(|X-m| \leq \alpha \cdot \sigma) &= P(m - \alpha \cdot \sigma \leq X \leq m + \alpha \cdot \sigma) = \Phi_{m,\sigma}(m + \alpha \cdot \sigma) - \Phi_{m,\sigma}(m - \alpha \cdot \sigma) = \\ &= \Phi\left(\frac{m + \alpha \cdot \sigma - m}{\sigma}\right) - \Phi\left(\frac{m - \alpha \cdot \sigma - m}{\sigma}\right) = \Phi(\alpha) - \Phi(-\alpha) = 2\Phi(\alpha) - 1. \end{aligned}$$

Keressük, legalább mennyi $\alpha$, ha $P(|X-m| > \alpha \cdot \sigma) < 0,01$, vagy ami ezzel ekvivalens:

$$0,99 < P(|X-m| \leq \alpha \cdot \sigma) = 2\Phi(\alpha) - 1 \qquad \Longrightarrow \qquad \Phi(\alpha) \geq 0,995.$$

Mivel $\Phi$ szigorúan monoton növekvő függvény, és a standard normális eloszlás táblázata szerint $\Phi(2,58) = 0,995$, $\alpha \geq 2,58$ adódik.

**3.3. Megjegyzés.** *Gyakorlati szempontból: normális eloszlásra a szórás háromszorosánál nagyobb kilengéseknek már elenyészően kicsi a valószínűsége.*

**3.6. Kidolgozott Feladat.** *A TRAGACS személyautó csomagtartója szabvány szerint* 70 *cm széles, a RÉMÁLOM gyerekágyakat lapra szerelve, szabványosan 65 cm széles csomagban árulják. Valójában mindkét szélesség normális eloszlású valószínűségi változó, a csomagtató esetében* 3*, a lapra szerelt csomag esetében* 4 *cm szórással (a két szélesség függetlennek tekinthető). Veszek egy RÉMÁLOM gyerekágyat, és szeretném a TRAGACSommal hazavinni. Mi a valószínűsége, hogy bele fog férni a csomagtartóba?*

**Megoldás.** Vezessük be a következő valószínűségi változókat: $X$ a TRAGACSom csomagtartójának, $Y$ a RÉMÁLOM ágy lapra szerelt csomagjának a szélessége. A feladat szövege szerint $X \in \mathcal{N}(70, 3)$, $Y \in \mathcal{N}(65, 4)$ és függetlenek. A feladat kérdése: $P(Y > X) = ?$ ehhez vezessük be még a $Z = X - Y$ valószínűségi változót, ez is normális eloszlású, továbbá:

$$EZ = EX - EY = 70 - 65 = 5$$

és

$$D^2Z = D^2X + D^2Y = 3^2 + 4^2 = 25 \qquad \Longrightarrow \qquad DZ = 5,$$

tehát $Z \in \mathcal{N}(5, 5)$. A kérdést $Z$ eloszlására, azt pedig a standard normális eloszlásfüggvényre visszavezetve:

$$P(Y > X) = P(Z < 0) = \Phi_{5,5}(0) = \Phi\left(\frac{-5}{5}\right) = \Phi(-1) = 1 - \Phi(1) = 1 - 0,8413 = 0,1587.$$

## 3.3. Gyakorló feladatok

**3.7. Feladat.** *Magyarországon nyári éjszakákon rengeteg a hullócsillag, de a teraszunkról ezeket elég rosszul lehet észrevenni: tapasztalataim szerint a nyári éjszakák* $\frac{1}{e^2} \approx 0.135$ *részében egyáltalán nem látok hullócsillagot a teraszról.*

*(a) Mi a valószínűsége, hogy egy adott éjszaka pontosan egy hullócsillagot fogok látni?*

*(b) Mi a valószínűsége, hogy egy adott éjszaka legalább három hullócsillagot fogok látni?*

**3.8. Feladat.** *Az* 100 *km hosszú telefonkábel valahol a* 30 *km-nél és* 65 *km-nél levő ellenőrző pontok között meghibásodott. Tegyük fel, hogy a meghibásodás helye egyenletes eloszlású ezen a* 35 *km hosszú szakaszon.*

*(a) Mi a valószínűsége, hogy a meghibásodás a 100 km hosszú kábel első 50 km-re esik?*

*(b) Számoljuk ki a meghibásodás $\xi$ helyének várható értékét és szórását!*

**3.9. Feladat.** *Ellenőrizzük az $(n,p)$ paraméterű binomiális, a $\lambda$ paraméterű Poisson, a $\lambda$ paraméterű exponenciális, valamint a standard normális eloszlás szórására vonatkozó képletet.*

**3.10. Feladat.** *A hivatalosan 50 grammos RANDOM csokoládé szeletek súlya valójában normális eloszlású, 49 gramm várható értékkel és 3 gramm szórással. A 60 perces valószínűségszámítás vizsgára 6 RANDOM szeletet viszek magammal, és szellemi teljesítőképességem növelésének érdekében 10 percenként ezek közül egyet-egyet elfogyasztok. Átmenni a vizsgán csak akkor van reményem, ha a 6 alkalomból legalább négyszer sikerül legalább 45 gramm csokit juttatnom a szervezetembe. Mi ennek az esélye?*

# 4. Nagy számok törvénye és centrális határeloszlás-tétel

## 4.1. Elméleti összefoglaló

A mérnöki munkában is igen fontos feladat lehet olyan mennyiségek ingadozásainak pontos megértése, amelyek sok, egymásra rakódó, és egymástól függetlennek tekinthető véletlen jelenség hatására alakulnak ki. Ennek a fejezetnek a célja az ilyen ingadozások megértése. Bevezetésül nézzünk két fontos egyenlőtlenséget, amelyek valószínűségi változók ingadozásait általában jellemzik.

**Markov egyenlőtlenség.** Legyen $\xi$ egy nemnegatív valószínűségi változó (azaz $P(\xi < 0) = 0$). Ekkor tetszőleges $a > 0$ számra:

$$P(\xi \geq a) \leq \frac{E\xi}{a}.$$

Az egyenlőtlenség persze semmitmondó, ha $a \leq E\xi$. Másrészt $a \gg E\xi$ esetén érthetjük meg igazán a szemléletes jelentését: igen kicsi annak az esélye, hogy egy pozitív véletlen mennyiség a várható értékénél lényegesen nagyobb legyen.

A bizonyítás igen egyszerű, tekintsük pl. az abszolút folytonos $\xi$ esetét. Ekkor $\xi$ nemnegativitása miatt $f(x) = 0$ tetszőleges $x < 0$ esetén, így

$$E\xi = \int_0^\infty x f(x)\,dx \geq \int_a^\infty x f(x)\,dx \geq a \int_a^\infty f(x)\,dx = aP(\xi \geq a);$$

ahonnan átrendezéssel adódik a Markov egyenlőtlenség.

**Csebisev egyenlőtlenség.** Legyen $\xi$ egy tetszőleges (véges szórású) valószínűségi változó, és a jelölés egyszerűsítésének céljából vezessük be az $m := E\xi$, $\sigma := D\xi$ jelöléseket. Ekkor tetszőleges $\varepsilon > 0$ esetén:

$$P(|\xi - m| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

A Csebisev egyenlőtlenség szemléletes jelentése ($\varepsilon \gg \sigma$ választás esetén): a véletlen mennyiségek a várható értékük körül ingadoznak, a szórásnál lényegesen nagyobb kilengések azonban csak kis valószínűséggel fordulnak elő. Például a szórás tízszeresét meghaladó ingadozásoknak legfeljebb $0,01$ lehet a valószínűsége.

A Csebisev egyenlőtlenség könnyen adódik, ha bevezetjük az $\eta := (\xi - m)^2$ valószínűségi változót, hiszen ekkor $E\eta = D^2\xi = \sigma^2$. Ha $\eta$-ra felírjuk a Markov egyenlőtlenséget $a = \varepsilon^2$ választással, éppen a $\xi$-re vonatkozó Csebisevet kapjuk vissza.

**Nagy számok törvénye.** Az alábbiakban valószínűségi változóknak egy fontos típusát fogjuk vizsgálni: végezzük el $n$-szer ugyanazt a kísérletet, próbálkozásaink függetlenek, és jelöljük $X$-szel, hogy az $n$ próbálkozásból összesen hányszor adódott egy bizonyos eredmény, amelynek esélye minden egyes kísérletben $p$. Erre az eredményre úgy is gondolhatunk, mint „sikerre” az egyes kísérletekben, így $X$ azt méri, hány próbálkozásunk volt sikeres. $n$ jellemzően nagyon nagy szám (sokszor próbálkozunk), $(0 <)p(< 1)$ pedig valamilyen rögzített paraméter. Néhány konkrét feladat: feldobunk sokszor egy szabályos pénzérmét, és megszámoljuk, hányszor esett a Fej oldalára ($p = 1/2$) vagy egy szabályos dobókockát, és $X$ a hatos dobások száma ($p = 1/6$). Egy a műszaki alkalmazásokhoz közelebb álló példa: egy gépsor gyárt valamilyen alkatrészeket, a legyártott termékekből minőségellenőrzés céljából nagy elemszámú mintát veszünk, és összeszámoljuk, hány alkatrész selejtes. Itt $p$-t nem ismerjük, sőt, éppen $p$-t – a gépsor hibaszázalékát – szeretnénk minél jobban megbecsülni mintavételünkkel.

Matematikai szempontból $X$ valószínűségi változó, hiszen értéke függ a véletlentől. $X$ eloszlását ismerjük – emlékezzünk vissza a 3 fejezetre, ezen belül a binomiális eloszlásra, $X$ ezt az eloszlást követi – így várható értékét és szórását is meg tudjuk állapítani: $EX = np$, $DX = \sqrt{np(1-p)}$. Mindig tartsuk szem előtt, hogy $X$-re most abban a határesetben gondolunk, amikor $n$ nagyon nagy. Fontos észrevétel, hogy $n$ növelésével a várható érték és a szórás is nő, de *$EX$ növekedésének üteme nagyságrendben nagyobb, mint $DX$ növekedésének üteme.*

Tekintsük $X/n$-t, vagyis a *sikerek arányát*, ez a mennyiség szintén valószínűségi változó, intuíciónk alapján azt várjuk, hogy ez $p$-t, az egyedi kísérletben a siker esélyét fogja megközelíteni $n$ növelésével. Alapvetően ezt az intuíciót formalizálja a *Nagy számok törvénye*: legyen $\delta > 0$ tetszőleges(en pici) rögzített szám, ekkor:

$$P\left(\left|\frac{X}{n} - p\right| \geq \delta\right) \to 0, \text{ ha } n \to \infty.$$

A bizonyítás egyben a konvergencia gyorsaságáról is információt ad. Egyszerű átalakítások mellett a Csebisev egyenlőtlenséget használjuk az $X$ valószínűségi változóra, $\varepsilon := n\delta$ választással (emlékeztetésképp $EX = np$, $DX = \sqrt{np(1-p)}$):

$$P\left(\left|\frac{X}{n} - p\right| \geq \delta\right) = P(|X - np| \geq n\delta) \leq \frac{D^2(X)}{n^2\delta^2} = \frac{np(1-p)}{n^2\delta^2} = \frac{p(1-p)}{n\delta^2} \qquad (3)$$

és az utolsó kifejezés $n$ növelésével mindenképp 0-hoz tart.

A nagy számok törvénye szavakban kifejezve: a *kísérletek számának növelésével a sikerek aránya egyre pontosabban, egyre nagyobb valószínűséggel megközelíti $p$-t.* Persze akármilyen sokszor dobok fel egy dobókockát, elképzelhető, hogy a hatosok aránya lényegesen el fog térni 1/6-tól (mondjuk 1/2-nél is nagyobb lesz), de ennek rendkívül kicsi a valószínűsége. A (3) becslés három mennyiség között teremt kapcsolatot: a kísérletek száma ($n$), a pontosság ($\delta$, tehát a sikerarány és $p$ eltérése), valamint a biztonság (legalább $\frac{p(1-p)}{n\delta^2}$ valószínűséggel lesz $\delta$-nál kisebb az eltérés) között. A feladatokban (és a mérnöki gyakorlatban is!) jellemzően a három mennyiség közül kettő adott, és a harmadikat kell kiszámolni.

**4.1. Megjegyzés.** *A konvergenciának ezt a valószínűségi változókra felírt alakját a matematikában sztochasztikus konvergenciának hívják, a nagy számok törvényének ezt az alakját pedig a nagy számok gyenge törvényének, vagy a nagy számok Bernoulli-féle törvényének is szokták mondani. A nagy számok törvényének további változatai is léteznek, a mérnöki alkalmazások szempontjából azonban az itt tárgyalt változat a legfontosabb.*

**Egy fontos általánosítás.** A fenti $X$ valószínűségi változóra vonatkozó becsléseinket annak binomiális eloszlása alapján kaptuk. $X$-re azonban egy kicsit másképp is gondolhatunk. Legyen $i = 1, 2, \ldots, n$ esetén

$$\xi_i = \begin{cases} 1 & \text{ha az } i. \text{ kísérlet sikeres,} \\ 0 & \text{ha az } i. \text{ kísérlet nem sikeres.} \end{cases}$$

Tehát például kockadobásra $\xi_2 = 1$, ha a második kockadobás hatos, és $\xi_2 = 0$, ha a második kockadobás eredménye hatostól különböző. Ezeket a véletlen mennyiségeket indikátor valószínűségi változóknak is szokták hívni. Könnyen kiszámolható, hogy $E\xi_i = p$ és $D^2\xi_i = p(1-p)$, minden $i$-re. másrészt a kísérletek függetlensége miatt a $\xi_i$ valószínűségi változók függetlenek, továbbá $X = \xi_1 + \xi_2 + \cdots + \xi_n$, a sikerek száma a teljes kísérletsorozatban. Felidézve a 2 fejezetben mondottakat:

$$EX = E\xi_1 + E\xi_2 + \cdots + E\xi_n = np, \quad D^2X = D^2\xi_1 + D^2\xi_2 + \cdots + D^2\xi_n = np(1-p) \implies DX = \sqrt{np(1-p)},$$

ahol a szórás számolásánál kihasználtuk a $\xi_i$-k függetlenségét. Az észrevétel rávilágít arra, hogy a nagy számok törvénye nem $X$ binomiális eloszlásán múlik (a törvény levezetésénél a várható érték és a szórás növekedésének üteme volt meghatározó). Egyben megkapjuk a következő általánosítást:

Legyenek $\eta_1, \eta_2, \ldots, \eta_n$ független és azonos eloszlású valószínűségi változók, $m = E\eta_i$ várható értékkel és $\sigma = D\eta_i$ szórással. Ez a mérnök számára a következőt jelenti: mérjük meg ugyanazt a (véletlentől is függő) mennyiséget $n$-szer egymástól független mérésekkel, azonos körülmények között. Képezzük a mérési eredmények $Y = \eta_1 + \eta_2 + \cdots + \eta_n$ összegét, illetve $Y/n$ átlagát. Ekkor tetszőleges(en kicsi) $\delta > 0$ számra:

$$P\left(\left|\frac{Y}{n} - m\right| \geq \delta\right) \to 0, \text{ ha } n \to \infty,$$

amit a nagy számok törvényével azonos módon vezethetünk le, kihasználva, hogy

$$EY = E\eta_1 + E\eta_2 + \cdots + E\eta_n = nm, \quad D^2Y = D^2\eta_1 + D^2\eta_2 + \cdots + D^2\eta_n = n\sigma \implies DY = \sqrt{n\sigma}.$$

A kapott általánosítás jelentőségét átláthatjuk a következő, műszaki jellegű alkalmazásnál: kiváncsiak vagyunk valamilyen mennyiség tényleges értékére (pl. egy anyagfajta szakítószilárdságára), amelyre azonban a gyakorlatban számos, véletlennek tekinthető fluktuáció rakódik rá (pl. a hőmérsékletből, a páratartalomból vagy más hasonló tényezőkből adódóan). Ha a mennyiséget sokszor megmérjük, egymástól független mérésekkel, akkor a mért értékek átlaga egyre pontosabban, egyre nagyobb biztonsággal meg fogja közelíteni a mennyiség tényleges értékét.

**A centrális határeloszlás-tétel.** A nagy számok törvényének fenti alakjai úgy is értelmezhetőek: ha az $n$ független mennyiség összegeként kapott $Y$ (illetve $X$) valószínűségi változóból kivonjuk a várható értékét (így egy 0 várható értékű valószínűségi változó adódik), majd leosztjuk *az* $Y$ *szórásánál lényegesen nagyobb* $n$-nel, akkor a kapott mennyiségből (nagy $n$-re) kiskálázódnak a véletlen fluktuációk.

Természetes módon vetődik fel a kérdés, hogy mi történik, ha $n$ helyett éppen $Y$ szórásával, $\sqrt{n}\sigma$-val osztunk. Ezzel a lineáris átskálázással

$$Z = \frac{Y - nm}{\sqrt{n}\sigma},$$

egy 0 várható értékű, 1 szórású valószínűségi változó, $Y$ úgynevezett *standardizálja* (ld. a 3.2 megjegyzést) adódik. *A centrális határeloszlás-tétel szerint asszimptotikusan (nagy $n$-re) a független, azonos eloszlású valószínűségi változók összegének standardizálásával kapott $Z$ valószínűségi változó standard normális eloszlásúnak tekinthető.* A centrális határeloszlás-tétel bizonyítása meghaladja ennek az összefoglalónak a kereteit (az érdeklődő olvasó megtalálhatja a valószínűségszámítást részletesen tárgyaló tankönyvekben), azonban valóban centrális jelentőségű eredményről van szó, az alkalmazások szempontjából is. A centrális határeloszlás-tétel értelmében ugyanis a $Z$-re (és ezáltal az igazán fontos átlagra, $Y/n$-re) vonatkozó becsléseinket számolhatjuk a standard normális eloszlásfüggvény segítségével. Így a nagy számok törvényéből adódónál lényegesen pontosabb válaszokat kapunk ugyanazokra a kérdésekre. Tekintsük például a mintavétel esetét, azaz $\eta_i := \xi_i$, $Y := X$ – persze így $X$ is független, azonos eloszlású valószínűségi változók összegeként áll elő, és standardizáltja $Z = \frac{X-np}{\sqrt{np(1-p)}}$. Legyen $\delta > 0$ rögzített, és vizsgáljuk meg, ezúttal a centrális határeloszlás-tétel segítségével, a sikerek arányának $p$-hez képesti eltéréseire vonatkozó kérdésünket (most a komplementer esemény valószínűségét könnyebb számolni):

$$\begin{aligned} P\left(\left|\frac{X}{n}-p\right|\le\delta\right) &= P(|X-np|\le n\delta)=P\left(\left|\frac{X-np}{\sqrt{np(1-p)}}\right|\le\frac{n\delta}{\sqrt{np(1-p)}}\right)= \\ &= P\left(|Z|\le\frac{\sqrt{n}\delta}{\sqrt{p(1-p)}}\right)=P\left(-\frac{\sqrt{n}\delta}{\sqrt{p(1-p)}}\le Z\le\frac{\sqrt{n}\delta}{\sqrt{p(1-p)}}\right)\approx \\ &\approx \Phi\left(\frac{\sqrt{n}\delta}{\sqrt{p(1-p)}}\right)-\Phi\left(-\frac{\sqrt{n}\delta}{\sqrt{p(1-p)}}\right)=2\Phi\left(\frac{\sqrt{n}\delta}{\sqrt{p(1-p)}}\right)-1. \end{aligned} \tag{4}$$

A kidolgozott példák és a gyakorló feladatok megoldása során látni fogjuk, hogy (4) lényegesen jobb becsléseket ad, mint (3). Például a centrális határeloszlás-tétel rávilágít, hogy a nagy számok törvényéből adódónál lényegesen kisebb mintavétellel is elérhetjük ugyanazt a pontosságot, ugyanolyan biztonsággal. Hasonlóképp a centrális határeloszlás-tétel alapján érdemes számolni általános $Y$ esetén (ha tetszőleges mennyiség független mérésekből adódó átlagát tekintjük). Éppen ezért alapulnak a matematikai statisztika módszerei is a centrális határeloszlás-tételen.

**A centrális határeloszlás-tétel szemléltetése.** A centrális határeloszlás tétel állítását a 15. ábrasor szemlélteti: legyen $\xi$ egyenletes eloszlású valószínűségi változó $[0,1]$-ben. Ennek $f_1(x)$ sűrűségfüggvénye $[0,1]$-ben egységnyi magasságú, egyébként nulla (legfelső ábra). Ha összeadunk két független, ilyen eloszlású változót, akkor az $f_2(x)$ sűrűségfüggvény kiszámítható (lásd [4]; p.195), grafikonját a felülről második ábrán mutatjuk be, "háromszög" alakú. Ha három illetve négy független ilyen eloszlású változót adunk össze, akkor az $f_3(x)$ és $f_4(x)$ sűrűségfüggvények grafikonját az alsó két ábra mutatja. Az ábrasorozat alapján vizuálisan érzékelhetjük, hogy az összeg sűrűségfüggvénye az összeadott változók számának növekedésével egyre jobban hasonlít a normális eloszlás sűrűségfüggvényéhez.

Az ábrasorozaton tehát azt látjuk, hogy az összeadott független változók $n$ számának növekedtével a sűrűségfüggvény alakja egyre jobban hasonlít a haranggörbére. Ugyanakkor az is megfigyelhető, hogy $n$ növekedtével a sűrűségfüggvény egyre szélesebb tartományra húzódik szét. Ahhoz, hogy a szokásos haranggörbe alakot kapjuk meg, a kiszámolt véletlen mennyiséget $\sqrt{n}$-nel kell osztanunk.

Ezt a jelenséget szemlélteti a 16. ábra. A kék, piros illetve zöld hisztogramok rendre azokhoz a változókhoz tartoznak, amelyeket a szövegben $(Y-nm)$-mel, $(Y/n-m)$-mel, illetve $Z$-vel jelöltünk. Normálás nélkül (kék hisztogram) az eloszlás egyre inkább szétterül, $n$-nel osztva (piros hisztogram) a nagy számok törvényének megfelelően ráhúzódik egyetlen pontra, $\sqrt{n}$-nel osztva (zöld hisztogram) pedig a centrális határeloszlás-tételnek megfelelően egyre pontosabban kirajzolódik a standard normális eloszlás sűrűségfüggvénye.

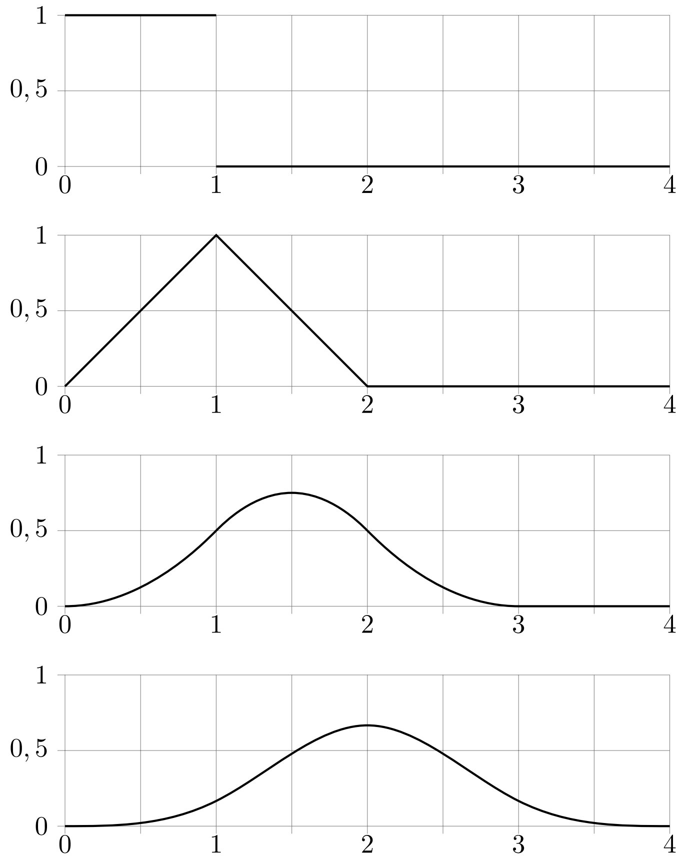
## 4.2. Kidolgozott példák

**4.1. Kidolgozott Feladat.** *A kristálycukrot* 1*kg-os zacskókban árulják: egy zacskó cukor tömege* $m=1000$ *g,* $\sigma=2$ *g szórással. Legalább mekkora valószínűséggel (a forgalomba kerülő zacskók legalább mekkora hányadára) esik a tömeg* 995 *g és* 1005 *g közé? Válaszoljunk a feladat kérdésére ha a tömeg eloszlása (a) ismeretlen; (b) normális.*

**Megoldás.** Jelöljük $\xi$-vel egy zacskó cukor tömegét, mint valószínűségi változót, gramm egységekben. Ekkor $E\xi=m=1000$, $D\xi=\sigma=2$. A feladat kérdése: legalább mennyi $P(995<\xi<1005)$ ?

(a) Mivel $\xi$ eloszlása ismeretlen, a komplementer valószínűséget számoljuk a Csebisev egyenlőtlenséggel ($\varepsilon=5$):

$$P(|\xi-1000|\ge 5)\le\frac{2^2}{5^2}=0,16$$




15. ábra. $[0,1]$-n egyenletes, független változók összegeinek sűrűségfüggvényei.

így

$$P(995 \leq \xi \leq 1005) = 1 - P(995 < \xi < 1005) \geq 0,84.$$

(b) Felidézve az $m$ várható értékű, $\sigma$ szórású normális eloszlásról tanultakat a 3. fejezetből:

$$\begin{aligned} P(995 < \xi < 1005) &= \Phi_{1000,2}(1005) - \Phi_{1000,2}(995) = \Phi(2,5) - \Phi(-2,5) \\ &= 2 \cdot \Phi(2,5) - 1 = 2 \cdot 0,9938 - 1 = 0,9876. \end{aligned}$$

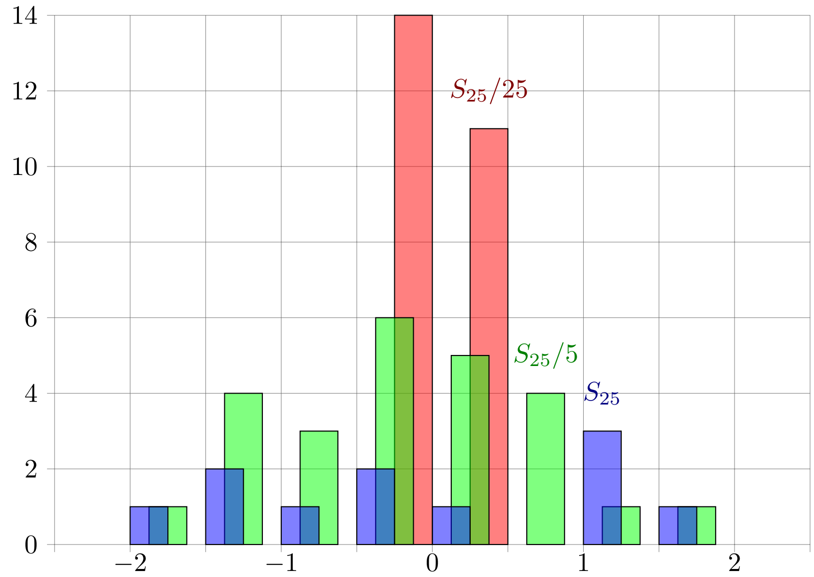
**4.2. Kidolgozott Feladat.** *Feldobunk* 10000*-szer egy szabályos pénzémét, (legalább) mekkora a valószínűsége, hogy a Fejek aránya* 0,49 *és* 0,51 *közé esik? Válaszoljunk a feladat kérdésére (a) a Nagy számok törvénye, illetve (b) a Centrális határeloszlás-tétel segítségével.*

**Megoldás.** Az elméleti összefoglaló jelöléseit használva: $p = 1/2$ a fej valószínűsége egy érmedobásra, $n = 10000$ az érmedobások száma, $X$ jelöli ebből a fej eredmények számát, $\delta = 0,01$ pedig a relatív gyakoriság és $p$ eltérése.

(a) A (3) formulával a komplementer esemény valószínűsége becsülhető:

$$P\left(\left|\frac{X}{n} - p\right| \geq \delta\right) \leq \frac{p(1-p)}{n\delta^2} = \frac{1}{4 \cdot 10000 \cdot 0,01^2} = 0,25$$

tehát legalább 0,75 a keresett valószínűség.




16. ábra. $S_{25}$ hisztogramja különböző normálásokkal 25 kísérletből. A hisztogram téglalapjait a jobb láthatóság kedvéért húztuk szét, valójában egymás mellett kellene álljanak, mint a 6. ábrán. A függőleges tengelyen az adott 1/2 hosszú intervallumba eső kísérletek számát tüntettük fel. A kék téglalapok folytatódnak az ábrázolt tartományon túl is.

(b) Most a (4) formulát használjuk:

$$P\left(\left|\frac{X}{n}-p\right|\geq\delta\right)\approx 2\Phi\left(\frac{\sqrt{n}\delta}{\sqrt{p(1-p)}}\right)-1=2\Phi(2)-1=2\cdot 0,9772-1=0,9544.$$

**4.3. Kidolgozott Feladat.** *Hányszor kell feldobni egy szabályos dobókockát, hogy 99% valószínűséggel a hatosok aránya 0,1 és 0,2 közé essen? Válaszoljunk a feladat kérdésére (a) a Nagy számok törvénye, illetve (b) a Centrális határeloszlás-tétel segítségével.*

**Megoldás.** Most tudjuk, hogy $p = 1/6 \approx 0,1667$, és az a kérdés, hogy mennyinek kell választanunk $n$-t, hogy $P(0,1 < X/n < 0,2)$ legalább $0,99$ legyen.

(a) Egy kicsit el kell gondolkodnunk, mert a Csebisev egyenlőtlenséggel (és ennek megfelelően a nagy számok törvényével) a várható értékre szimmetrikus intervallumok valószínűségei becsülhetők, és a $[0,1; 0,2]$ intervallum nem ilyen. Ezért a következőképp járunk el:

$$P(0,1 < X/n < 0,2) \geq P(0,1333 < X/n < 0,2) = P(|X/n - 1/6| < 1/30),$$

így ha $P(|X/n - 1/6| > 1/30) \geq 0,99$, akkor nyertünk. Másrészt az utóbbi esemény komplementerének valószínűségére a (3) formulából:

$$P(|X/n - 1/6| \geq 1/30) \leq \frac{1/6 \cdot 5/6}{n \cdot (1/30)^2} = \frac{125}{n},$$

tehát $\frac{125}{n} \leq 0,01$-t kell biztosítani, vagyis $n \geq 12500$.

(b) Most is ugyanaz a gondolatmenet, mint az előző részfeladatnál, csak a (4) formulát használjuk. Így:

$$P(0,1 < X/n < 0,2) \geq P(|X/n - 1/6| > 1/30) \approx 2\Phi\left(\frac{\sqrt{n}/30}{\sqrt{1/6 \cdot 5/6}}\right) - 1 = 2\Phi\left(\frac{\sqrt{n}}{5\sqrt{5}}\right) - 1,$$

tehát elegendő $2\Phi(\frac{\sqrt{n}}{5\sqrt{5}}) - 1 \geq 0,99$-t biztosítani. Mivel $\Phi(2,58) \approx 0,995$, és $\Phi(x)$ monoton növő függvény, ez $\sqrt{n} \geq 2,58 \cdot 5\sqrt{5}$-tel ekvivalens, tehát $n \geq 833$.

**4.4. Kidolgozott Feladat.** *Magyarország lakosságának számunkra ismeretlen $p$ hányada dohányzik. Felmérést készítünk $p$ meghatározására: megkérdezünk $n$ magyar állampolgárt, közülük $X$ mondja magát dohányosnak: ez alapján $p$-t $X/n$-nel becsüljük. Persze $X$ valószínűségi változó, hiszen függ a véletlentől, hogy milyen állampolgárokat sikerült megszólítanunk. Feltételezve, hogy a felmérés résztvevőit egymástól független körülmények között választjuk, határozzuk meg, legalább mennyi legyen $n$, ha azt szeretnénk, hogy $X/n$ legfeljebb $0,01$-gyel térjen el $p$-től, 98% valószínűséggel. Válaszoljunk a feladat kérdésére (a) a Nagy számok törvénye, illetve (b) a Centrális határeloszlás-tétel segítségével.*

**Megoldás.** Ezt a feladatot pontosan ugyanúgy kell megoldani, mint az előzőt – adott biztonság és pontosság eléréséhez keressük a megfelelő $n$-t – egy különbség van csupán, nem ismerjük $p$-t (hiszen épp $p$ felmérése a célunk!). Így olyan becslésre van szükség, amely $0,98$ valószínűséggel biztosítja a $\delta = 0,01$ pontosságot, akármennyi is legyen $p$. Ehhez mindkét részfeladatnál a számtani és mértani közép közti egyenlőtlenségből következő

$$\sqrt{p(1-p)} \leq \frac{p + (1-p)}{2} = 1/2 \qquad \Longleftrightarrow \qquad p(1-p) \leq 1/4 \tag{5}$$

elemi becslést fogjuk használni.

(a) Célunk $n$ meghatározása, ha

$$0,02 \geq P(|X/n - p| \geq 0,01).$$

A (3) és (5) formulák alapján:

$$P(|X/n - p| \geq 0,01) \leq \frac{p(1-p)}{n\delta^2} \leq \frac{1}{4 \cdot n \cdot 10^{-4}},$$

tehát $n \geq (0,02 \cdot 4 \cdot 10^{-4})^{-1} = 125000$.

(b) Célunk $0,98 \leq P(|X/n - p| \leq 0,01)$. Kihasználva a (4) és (5) formulákat, valamint $\Phi(x)$ monoton növekedését:

$$P(|X/n - p| \leq 0,01) \approx 2\Phi\left(\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 = 2 \cdot \Phi(0,02\sqrt{n}) - 1.$$

Ugyanakkor $\Phi(2,34) \approx 0,99$, tehát $0,02\sqrt{n} \geq 2,34$ szükséges a cél eléréséhez, vagyis $n \geq 13689$.

**4.5. Kidolgozott Feladat.** *Egy fémötvözet fajhőjét 100 független, azonos körülmények között elvégzett mérés segítségével szeretnénk megállapítani, a mérések átlaga 293 $\frac{J}{kg\,^oC}$. Határozzuk meg azt az intervallumot, amelybe az ötvözet fajhője 95% biztonsággal beleesik, ha a fajhő szórása 15 $\frac{J}{kg\,^oC}$.*

***4.2. Megjegyzés.*** *A statisztikában ezt úgy is mondják, hogy határozzuk meg a fajhő 95%-os megbízhatósági (konfidencia) intervallumát.*

**Megoldás.** Jelöljük az $n = 100$ mérési eredmény összegét $Y$-nal, átlagát $Y/n$-nel (ez utóbbiról tudjuk, hogy 293). A minden egyes mérésre azonos $m$ várható értéket nem ismerjük, tudjuk viszont, hogy a szórás $\sigma = 15$. A centrális határeloszlás-tétel alapján $Z = \frac{Y-nm}{\sqrt{n}\sigma}$ eloszlását tekinthetjük standard normálisnak. Lemásolva a (4) formulánál látott levezetést, tetszőleges $\delta > 0$-ra:

$$\begin{aligned} P\left(\left|\frac{Y}{n} - m\right| \le \delta\right) &= P(|Y - nm| \le n\delta) = P\left(\left|\frac{Y - nm}{\sqrt{n}\sigma}\right| \le \frac{n\delta}{\sqrt{n}\sigma}\right) = \\ &= P\left(|Z| \le \frac{\sqrt{n}\delta}{\sigma}\right) = P\left(-\frac{\sqrt{n}\delta}{\sigma} \le Z \le \frac{\sqrt{n}\delta}{\sigma}\right) \approx \\ &\approx \Phi\left(\frac{\sqrt{n}\delta}{\sigma}\right) - \Phi\left(-\frac{\sqrt{n}\delta}{\sigma}\right) = 2\Phi\left(\frac{\sqrt{n}\delta}{\sigma}\right) - 1 = 2\Phi\left(\frac{10\delta}{15}\right) - 1. \end{aligned}$$

Célunk annak a (minimális) $\delta$-nak a meghatározása, amelyre ez a valószínűség legalább $0,95$. Ekkor ugyanis 95% valószínűséggel: $Y/n$ beleesik az $m$ körüli $\delta$ sugarú intervallumba, tehát a kimért érték és $m$ eltérése legfeljebb $\delta$ lehet. Ehhez:

$$0,95 \ge 2\Phi\left(\frac{10\delta}{15}\right) - 1 \qquad \Longleftrightarrow \qquad 0,975 \ge \Phi(2\delta/3),$$

továbbá $\Phi(1,96) = 0,975$, és $\Phi(x)$ monoton nő, tehát $\delta = 1,4625$. Összefoglalva, 95% biztonsággal állíthatjuk, hogy $\frac{J}{kg\,^oC}$ egységekben az ötvözet fajhője a $[293 - 1,4625; 293 + 1,4625] = [291,5375; 294,4625]$ intervallumba esik.

**4.3. Megjegyzés.** *A feladat fenti megfogalmazásában az a valószerűtlen, hogy a szórást eleve ismerjük: jellemzően az átlaghoz hasonlóan a szórást is a mérésekből becsüljük meg a tapasztalati szórásnégyzet segítségével (lásd a 2.1.2. fejezetben a mérnöki bevezetést.) Ilyenkor egy kicsit másképp kell számolni (a standard normális eloszlás helyett az ún. Student eloszlást kell használni, erről bővebben olvashatunk bármilyen bevezető statisztika könyvben). Kellően nagy minta esetén azonban (és $n = 100$ már feltétlen elég nagy) számolhatunk a fent leírt módon akkor is, ha a szórást a mért adatokból becsüljük meg.*

## 4.3. Gyakorló feladatok

**4.6. Feladat.** *A TV-t véletlenszerűen bekapcsolva az esetek $\frac{1}{4}$-ében látunk éppen reklámot. Százszor bekapcsolva a TV-t, mi a valószínűsége, hogy legalább 22 és legfeljebb 28 alkalommal megy éppen reklám? Becsüljük meg ezt a valószínűséget a centrális határeloszlás-tétel alapján.*

**4.7. Feladat.** *Egy oltóanyag hatékonyságát állatkísérletekkel teszteljük: egy emlős szervezete ismeretlen $p$ valószínűséggel válik a betegséggel szemben rezisztenssé az oltóanyag hatására. $n$ patkányon elvégezve a kísérletet, $p$-t a rezisztenssé váló patkányok $n$-hez képesti arányával becsüljük. Legalább hány patkányon kell a kísérletet végrehajtani, hogy 97% valószínűséggel becslésünk $p$ tényleges értékétől legfeljebb $0,02$-vel térjen el? Válaszoljunk a kérdésre (a) a Nagy számok törvénye, illetve (b) a Centrális határeloszlás-tétel alapján.*

**4.8. Feladat.** *Egy automata gépsor fogaskerekeket gyárt, melyek átmérőjének várható értéke $m = 20$ mm, szórása $\sigma = 0,5$ mm. A várhatóhoz képest legalább $\delta$ mm eltérést mutató fogaskerekeket selejtesnek minősítik. Mennyi lehet $\delta$, ha tudjuk, hogy a selejtarány 4%? Válaszoljunk a feladat kérdésére, ha az átmérő eloszlása (a) ismeretlen; (b) normális.*

**4.9. Feladat.** *Tekintsük a 4.8. feladat gépsorát: felmerült, hogy a gépsort újra kell kalibrálni. Erre akkor van szükség, ha a legyártott fogaskerekek átmérője már nem 20 mm, hanem valamilyen attól lényegesen, legalább $0,1$ mm-rel eltérő érték. Ennek eldöntésére megmérjük $n$ legyártott fogaskerék átmérőjét, és* *megnézzük, a mérések mm-ben vett átlaga beleesik-e a* $[19,9; 20,1]$ *intervallumba. Persze egy ilyen eljárás csak akkor megbízható, ha a mérések átlagának és a fogaskerék-átmérő várható értékének az eltérése nagy valószínűséggel nem haladja meg a* $0,1$ *mm-t. Hány mérést kell elvégezni, ha azt szeretnénk, hogy ez a valószínűség legalább* $94\%$ *legyen? Számoljunk a 4.8. feladat* $\sigma = 0,5$ *mm szórásával, és becsüljünk a centrális határeloszlás-tétel alapján.*

# Függelék: gyakorló feladatok numerikus eredményei

| Feladat száma | Megoldás |
|---|---|
| **1.8** | (a) $H\overline{M}$, (b) $\overline{H} \cdot \overline{M}$, (c) $H\overline{M} + \overline{H}M$, (d) $\overline{HM}$. |
| **1.9** | 0,9856. |
| **1.10** | (a) $P(A) = 1/2$, (b) $P(A\|B) = 1/3$, (c) igen. |
| **1.11** | (a) 1/12, (b) nem. |
| **1.12** | 1/8. |
| **1.13** | 1/3. |
| **2.6** | $E\xi = 7$, $D\xi = \sqrt{\frac{35}{6}} \approx 2,415$. |
| **2.7** | (a) $B = 1$, (b) 1/2, (c) $E\xi = 1$, (d) $D\xi = \sqrt{\pi - 3} \approx 0,3763$. |
| **2.8** | (a) $p = 0,1$, (b) nem, (c) $R(\xi, \eta) \approx -0,4138$. |
| **2.9** | (a)$B = \frac{35}{48}$, (b) $f_\xi(x) = \frac{35}{48}(2x^4 + \frac{2}{7})$ ha $\|x\| \leq 1$, egyébként 0, $f_\eta(y) = \frac{35}{48}(2y^6 + \frac{2}{5})$ ha $\|y\| \leq 1$, egyébként 0, (c) nem. |
| **2.10** | (a) $f_\xi(x) = \frac{2\sqrt{1-x^2}}{\pi}$ ha $\|x\| \leq 1$, egyébként 0, $f_\eta(y) = \frac{2\sqrt{1-y^2}}{\pi}$ ha $\|y\| \leq 1$, egyébként 0, (b) $R(\xi, \eta) = 0$. |
| **3.7** | (a) $2e^{-2} \approx 0,27$, (b) $1 - 5e^{-2} \approx 0,325$. |
| **3.8** | (a) 4/7, (b) $E\xi = 47,5$, $D\xi \approx 10,104$. |
| **3.9** | Ld. az elméleti összefoglalót. |
| **3.10** | $\approx 0,9875$. |
| **4.6** | $\approx 0,51$. |
| **4.7** | (a) $n \geq 20834$, (b) $n \geq 2944$. |
| **4.8** | (a) $\delta \leq 2,5$, (b) $\delta \approx 1,03$. |
| **4.9** | $n \geq 85$. |

# Hivatkozások

[1] Balázs Márton, Tóth Bálint: *Valószínűségszámítás 1. jegyzet matematikusoknak és fizikusoknak*, elérhető: `http://www.math.bme.hu/~balazs/vsz1jzetb-t.pdf`

[2] Jay Devore, Nicholas Farnum: *Applied Statistics for Engineers and Scientists*; Thomson Brooks/Cole, Belmont CA, 2005

[3] Monostory Iván: *Valószínűségelmélet és Matematikai statisztika Pédatár*; Műegyetemi Kiadó, Budapest, 2001

[4] Prékopa András: *Valószínűségelmélet*; Műszaki Könyvkiadó, Budapest, 1980

[5] Rényi Alfréd: *Valószínűségszámítás*; Tankönyvkiadó, Budapest, 1972

[6] Szelezsán János: *Valószínűségszámítás és Matematikai Statisztika*; LSI Oktatóközpont, Budapest, 1999