Értékünk az ember

Humánerőforrás-fejlesztési Operatív Program

Kiss Béla – Krebsz Anna

# VALÓSZÍNŰSÉG-SZÁMÍTÁS ÉS MATEMATIKAI STATISZTIKA

SZÉCHENYI ISTVÁN EGYETEM
GYŐR

Magyarország célba ér

NEMZETI FEJLESZTÉSI TERV

Készült a HEFOP 3.3.1-P.-2004-09-0102/1.0 pályázat támogatásával.

Szerzők: dr. Kiss Béla
főiskolai tanár

dr. Krebsz Anna
egyetemi docens

Lektor: dr. Bolla Marianna
egyetemi docens

# A dokumentum használata

## Mozgás a dokumentumban

A dokumentumban való mozgáshoz a Windows és az Adobe Reader megszokott elemeit és módszereit használhatjuk.

Minden lap tetején és alján egy navigációs sor található, itt a megfelelő hivatkozásra kattintva ugorhatunk a használati útmutatóra, a tartalomjegyzékre, valamint a tárgymutatóra. A ◀ és a ▶ nyilakkal az előző és a következő oldalra léphetünk át, míg a Vissza mező az utoljára megnézett oldalra visz vissza bennünket.

## Pozícionálás a könyvjelzőablak segítségével

A bal oldali könyvjelző ablakban tartalomjegyzékfa található, amelynek bejegyzéseire kattintva az adott fejezet/alfejezet első oldalára jutunk. Az aktuális pozíciónkat a tartalomjegyzékfában kiemelt bejegyzés mutatja.

## A tartalomjegyzék és a tárgymutató használata

### Ugrás megadott helyre a tartalomjegyzék segítségével

Kattintsunk a tartalomjegyzék megfelelő pontjára, ezzel az adott fejezet első oldalára jutunk.

### A tárgymutató használata, keresés a szövegben

Keressük meg a tárgyszavak között a bejegyzést, majd kattintsunk a hozzá tartozó oldalszámok közül a megfelelőre. A további előfordulások megtekintéséhez használjuk a Vissza mezőt.

A dokumentumban való kereséshez használjuk megszokott módon a Szerkesztés menü Keresés parancsát. Az Adobe Reader az adott pozíciótól kezdve keres a szövegben.

# Tartalomjegyzék

# Előszó

A jegyzet a Széchenyi István Egyetem mérnöki BSc-szakos hallgatói számára készült, a valószínűség-számítás és a matematikai statisztika klasszikus bevezető fejezeteit tartalmazza. Az anyag tárgyalásánál az elméleti részek rövid, előadásjegyzet formában történő leírására törekedtünk. Ezzel azt kívántuk elérni, hogy egy viszonylag rövid, témakörök szerint jól tagolt anyag álljon a hallgatók rendelkezésére. A példáknál igyekeztünk az elmélet műszaki-gazdasági alkalmazásait is bemutatni. A jegyzet terjedelmi korlátai miatt csak a fontosabb és viszonylag egyszerűen igazolható tételek bizonyítottuk. A statisztika részben is csak a leggyakrabban alkalmazott próbák ismertetésére szorítkoztunk.

A jegyzet újszerű abban az értelemben, hogy a statisztikai feladatoknál a kézi számolás mellett azok Microsoft Excellel történő megoldását is ismertettük, mivel ez a programcsomag minden hallgató számára ingyenesen elérhető. Az anyag hagyományos eloszlási táblázatokkal történő kiegészítését elhagytuk, hiszen a Microsoft Excel eloszlásfüggvényei ezeknél sokkal részletesebb információt szolgáltatnak.

Végül köszönet illeti dr. Bolla Mariannát, a Budapesti Műszaki és Gazdaságtudományi Egyetem docensét a jegyzet gondos lektorálásáért és hasznos megjegyzéseiért, kiegészítéseiért.

A szerzők.

# 1. Kombinatorika

## 1.1. Permutációk, variációk és kombinációk

$n$ különböző elem különböző sorrendjeinek (az $n$ elem **permutációi**nak) a száma:

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

Ha az $n$ elem között $k_1, k_2, \ldots, k_l$ $(k_1 + k_2 + \ldots + k_l = n)$ darab megegyező van, az $n$ elem **ismétléses permutáció**Hiba! A könyvjelző nem létezik.**i**nak a száma:

$$P_n^{k_1, k_2, \ldots, k_l} = \frac{n!}{k_1! \cdot k_2! \cdot \ldots \cdot k_l!} .$$

$n$ különböző elemből $k$ különbözőt kiválasztunk $(k \le n)$ és minden lehetséges sorrendbe állítjuk. Az így keletkező **variációk** száma:

$$V_{n,k} = n \cdot (n-1) \cdot \ldots \cdot (n-k+1).$$

Ha megengedjük a kiválasztásnál az ismétlődést is, az **ismétléses variációk** száma:

$$V_{n,k}^{ism} = n^k .$$

Ha az $n$ különböző elemből $k$-t kiválasztunk $(k \le n)$, de a kiválasztottakat nem rakjuk különböző sorrendbe, a keletkező **kombinációk** száma:

$$C_{n,k} = \frac{V_{n,k}}{P_k} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k!} = \frac{n!}{k! \cdot (n-k)!} = \binom{n}{k}.$$

Ha $n$ különböző elemből kiválasztunk $k$-t, de a kiválasztott elem ismétlődhet, a felírható **ismétléses kombinációk** száma:

$$C_{n,k}^{ism} = \binom{n+k-1}{k}.$$

Megegyezés szerint $0!=1$ és $\binom{n}{n}=\binom{n}{0}=1$.

**1.1. tétel.** Az $n$ elem $k$-ad osztályú ismétléses kombinációinak a száma megegyezik az $(n+k-1)$ elem $k$-ad osztályú ismétlés nélküli kombinációinak számával, azaz $C_{n,k}^{ism}=C_{n+k-1,k}$.

**Bizonyítás.** Jelölje $A$ az $n$ elem $k$-ad osztályú ismétléses kombinációinak és $B$ az $(n+k-1)$ elem $k$-ad osztályú ismétlés nélküli kombinációinak a halmazát. A kombinációknál a kiválasztás sorrendje nem számít. Így az $A$ halmaz elemeit az $\{1,\ldots,n\}$ halmaz elemeiből álló $(a_1,a_2,\ldots,a_k)\in A$ nem csökkenő $a_1\le a_2\le\ldots\le a_k$ sorozatoknak, a $B$ halmaz elemeit pedig az $\{1,\ldots,n+k-1\}$ halmaz elemeiből álló $(b_1,b_2,\ldots,b_k)\in B$ szigorúan növekedő $b_1<b_2<\ldots<b_k$ sorozatoknak tekinthetjük. A két halmaz között a következő $T:A\to B$ leképezést létesítjük. Tetszőleges $(a_1,a_2,\ldots,a_k)\in A$ ismétléses kombinációhoz rendeljük hozzá a

$$T(a_1,a_2,\ldots,a_k)=(a_1,a_2+1,\ldots,a_k+k-1)\in B$$

ismétlés nélküli kombinációt.

A továbbiakban azt igazoljuk, hogy ez a $T$ leképezés kölcsönösen egyértelmű. Ehhez azt kell megmutatnunk, hogy a $T$ leképezés az $A$ halmaz különböző elemeihez a $B$ halmaz különböző elemeit rendeli és a $B$ halmaz minden egyes $(b_1,b_2,\ldots,b_k)$ eleméhez található olyan $A$ halmazbeli $(a_1,a_2,\ldots,a_k)$ elem, amelyre

$$T(a_1,a_2,\ldots,a_k)=(b_1,b_2,\ldots,b_k)$$

teljesül. Legyenek

$$(a_1,a_2,\ldots,a_k),\ (\tilde{a}_1,\tilde{a}_2,\ldots,\tilde{a}_k)\in A$$

tetszőleges különböző ismétléses kombinációk. Ekkor a

$$T(a_1,a_2,\ldots,a_k)=(a_1,a_2+1,\ldots,a_k+k-1)$$

és a

$$T(\tilde{a}_1,\tilde{a}_2,\ldots,\tilde{a}_k)=(\tilde{a}_1,\tilde{a}_2+1,\ldots,\tilde{a}_k+k-1)$$

ismétlés nélküli kombinációk is különböznek, ugyanis

$$(a_1, a_2, \ldots, a_k) - (\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_k) = (a_1 - \tilde{a}_1, a_2 - \tilde{a}_2, \ldots, a_k - \tilde{a}_k) \neq (0, 0, \ldots, 0)$$

esetén

$$T(a_1, a_2, \ldots, a_k) - T(\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_k)$$

sem egyenlő a $(0,0,\ldots,0)$ sorozattal, mivel

$$T(a_1, a_2, \ldots, a_k) - T(\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_k) = (a_1 - \tilde{a}_1, a_2 - \tilde{a}_2, \ldots, a_k - \tilde{a}_k).$$

Legyen $(b_1, b_2, \ldots, b_k) \in B$ egy tetszőleges ismétlés nélküli kombináció. Ekkor $(b_1, b_2 - 1, \ldots, b_k - k + 1) \in A$ egy olyan ismétléses kombináció, amelyre

$$T(b_1, b_2 - 1, \ldots, b_k - k + 1) = (b_1, b_2, \ldots, b_k)$$

teljesül. Ezzel igazoltuk a tételt.

A tétel kimondása és bizonyítása más, szemléletesebb gondolatmenettel is történhet. Vegyük észre, hogy a tétel állítása azzal ekvivalens, hogy $C_{n,k}^{ism} = P_{n+k-1}^{n-1,k}$. Ennek belátását az alábbi példán keresztül szemléltetjük. Cukrászdában *n* fajta sütemény van, ezekből haza akarok vinni *k* darabot ($k > n$ is lehet). Hányféleképpen tehetem ezt meg? A vásárlásokat a következőképp adminisztrálom: felírom az *n* süti nevét, közéjük $(n-1)$ vonalat húzok, és a vásárolt sütiket ugyanannyi ponttal jelölöm. Könnyű látni, hogy pontosan annyi különböző hazavitel létezik, mint ahány különböző sorrendje van az $(n-1)$ vonásból és *k* pontból álló sorozatnak. Ez utóbbi épp $(n-1+k)$ elem ismétléses permutációinak száma, melyek közül $(n-1)$ ill. *k* egymás között megkülönböztethetetlen.

**1. példa.** (Permutációk.) Hány olyan tízjegyű szám van, amelyben minden számjegy csak egyszer fordul elő?

**Megoldás.** Először vegyük az összes olyan számot, amelyek 10 különböző jegyből állnak! Ezek száma a 10 elem összes permutációinak száma, vagyis $P_{10} = 10!$. Mivel a permutációk képzése során egyik számjegynek sincs megkülönböztetett szerepe, ezért nyilván 1/10 részében 0 áll az első helyen. Így a megfelelő permutációk száma az összes permutációk 9/10 része. Tehát $9/10 \cdot 10! = 9 \cdot 9!$ valódi tízjegyű csupa különböző számjegyet tartalmazó szám van.

Más megoldás: az 1–10. helyiértékig a különböző kitöltések száma: $9 \cdot 9 \cdot 8 \cdot \ldots \cdot 2 \cdot 1 = 9 \cdot 9!$.

**2. példa.** (Ismétléses permutációk.) Egy pénzérmét tízszer egymás után feldobunk. Hányféle olyan dobássorozat van, amelyben 6 fej és 4 írás fordul elő?

**Megoldás.** A dobások száma adja az elemek számát, ez tehát 10. Mivel 6 fej és 4 írás megegyező elemeket jelentenek, a sorrendek számát ismétléses permutációval kapjuk meg:

$$P_{10}^{6,4} = \frac{10!}{6! \cdot 4!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6!}{6! \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 3 \cdot 7 \cdot 10 = 210 .$$

**3. példa.** (Variációk.) Egy nyolctagú család egy alkalommal négy színházjegyet kap. Hányféleképpen oszthatók ki a jegyek a családtagok között? (Mivel a jegyek számozottak, a sorrendet is figyelembe kell vennünk!)

**Megoldás.** A család nyolc tagjából négyet kell kiválasztanunk és ezek sorrendjét is meg kell külön adnunk. Képeznünk kell tehát 8 elem negyedosztályú variációit. Ezek száma:

$$V_{8,4} = 8 \cdot 7 \cdot 6 \cdot 5 = 1680 .$$

**4. példa.** (Ismétléses variációk.) A kétféle morzejelből, vagyis pontból és vonalból, 5-öt írunk fel egymás után. Hány ilyen ötös jelsorozat létezik?

**Megoldás.** A kétféle morzejel két elemet jelent, amelyekkel 5 helyet kell betölteni rögzített sorrendben. A lehetséges jelcsoportok száma 2 elem 5-öd osztályú ismétléses variációinak száma:

$$V_{2,5}^{ism} = 2^5 = 32 .$$

Tehát 32-féle jelcsoportot írhatunk fel.

**5. példa.** (Kombinációk.) Hányféleképpen töltheti még ki a lottószelvényt az, aki a 3, 7, 13 számokat már bejelölte a szelvényen?

**Megoldás.** A lottószelvényen 90 számból 5-öt kell megjelölni. Így, aki már hármat beírt, még két számot választhat a fennmaradt 87 szám közül. Tehát 87 elem másodosztályú kombinációit kell képeznünk, hogy az összes lehetséges változatokat megkapjuk. Ezek száma:

$$C_{87,2} = \binom{87}{2} = \frac{87!}{85! \cdot 2!} = \frac{87 \cdot 86}{1 \cdot 2} = 3741 .$$

**6. példa.** (Ismétléses kombinációk.) Egy tisztségre 3 jelölt van, ezekre 20-an szavaznak. Hányféle eredménnyel végződhet a titkos szavazás, ha mindenki egy jelöltre szavaz? (Eredményen annak megadását értjük, hogy a 3 jelölt külön-külön hány szavazatot kap.)

**Megoldás.** A szavazás végén a 20 szavazólap mindegyikén a 3 jelölt valamelyikének a neve áll. A szavazólapok sorrendje nem számít, csupán az, hogy a jelöltek külön-külön hány szavazatot kaptak. A szavazás minden lehetséges eredménye tehát a 3 jelölt egy 20-ad osztályú ismétléses kombinációja. Ennek száma:

$$C_{3,20}^{ism} = \binom{3+20-1}{20} = \binom{22}{20} = \frac{22!}{20! \cdot 2!} = \frac{22 \cdot 21}{2 \cdot 1} = 231 .$$

Tehát 231 különböző eredménnyel zárulhat a szavazás.

**7. példa.** (Ismétléses kombinációk.) Egy gyerek 5 különböző fagylaltból választhat háromgombócos adagot. Hányféle lehetősége van a választásra, ha az adagolás sorrendjére nem vagyunk tekintettel?

**Megoldás.** A fagylaltfajták száma – vagyis az elemek száma – 5. Ezekből hármas csoportokat képezünk, amelyekben a sorrend nem számít, és több gombóc is állhat ugyanabból a fajtából. A csoportok számát 5 elem 3-ad osztályú ismétléses kombinációinak száma adja, és ez

$$C_{5,3}^{ism} = \binom{5+3-1}{3} = \binom{7}{3} = \frac{7!}{3! \cdot 4!} = \frac{7 \cdot 6 \cdot 5}{1 \cdot 2 \cdot 3} = 35.$$

Tehát 35 lehetősége van a választásra.

**8. példa.** (Kombinációk.) Hányféleképp tölthető ki egy ötös lottószelvény?

**Megoldás:** $\binom{90}{5}$-féleképp.

**9. példa.** (Ismétléses variációk.) Hányféleképp tölthető ki totószelvény?

**Megoldás:** A 13-tippes szelvény $3^{13}$-féleképp. A $13+1$-tippes szelvény $3^{14}$-féleképp.

**10. példa.** (Ismétléses kombinációk.) Repülőgépről ledobunk 100 db 1000 forintost. Hányféleképp szedheti össze a lehullott pénzt: *a)* 100 ember *b)* 50 ember *c)* 150 ember?

**Megoldás:** A különböző lehetőségek száma:

*a)* $C_{100,100}^{ism} = \binom{199}{100}$ *b)* $C_{50,100}^{ism} = \binom{149}{100}$ *c)* $C_{150,100}^{ism} = \binom{249}{100}$

# 2. Valószínűség-számítás

A valószínűség-számítás tárgya: véletlen tömegjelenségek vizsgálata.

## 2.1. Események, műveletek eseményekkel

**Kísérlet** alatt egy véletlen tömegjelenség megfigyelését értjük. Egy kísérlet egy lehetséges kimenetele az **elemi esemény.** Az egy kísérlethez tartozó elemi események összessége az **eseménytér,** amit $\Omega$-val jelölünk. Az eseménytér bizonyos részhalmazait **esemény**eknek nevezzük.

Az elemi események egyetlen elemet tartalmazó események.

Ha egy $A$ eseményre vonatkozóan kísérletet végzünk, és a kísérlet során adódó $a$ elemi esemény eleme az $A$-nak ($a \in A$), akkor azt mondjuk, hogy az $A$ esemény **bekövetkezik.**

Az $\Omega$ halmaz is egy eseményt ad. Mivel ez az esemény az összes lehetséges elemi eseményt tartalmazza, ezért biztosan bekövetkezik, és emiatt **biztos esemény**nek nevezzük. Az üres halmazzal megadott esemény sohasem következik be, ezért **lehetetlen esemény**nek nevezzük és a $\varnothing$ szimbólummal jelöljük.

Ha $A$ és $B$ két esemény és $A \subset B$, akkor azt mondjuk, hogy az $A$ esemény bekövetkezése **maga után vonja** a $B$ esemény bekövetkezését.

Egy $A$ egy $B$ eseményt **egyenlő**nek nevezünk, ha bármelyik bekövetkezése egyben a másik bekövetkezését is jelenti.

Az események között az alábbi műveleteket értelmezzük:

Az $A$ és $B$ események $A + B$ **összege** az az esemény, amely akkor következik be, ha az $A$ és $B$ két esemény közül legalább az egyik bekövetkezik.

Az $A$ és $B$ események $A \cdot B$ **szorzata** az az esemény, amely akkor következik be, ha az $A$ és a $B$ is bekövetkezik.

Az $A$ és $B$ események $A - B$ **különbsége** az az esemény, amely akkor következik be, ha az $A$ esemény bekövetkezik, de a $B$ esemény nem.

Az $A$ esemény $\overline{A}$ **ellentettje (komplementere)** az az esemény, amely akkor következik be, ha az $A$ esemény nem következik be.

Ha az $A$ és $B$ esemény egyszerre sohasem következik be, azaz $A \cdot B = \varnothing$, akkor az $A$ és $B$ eseményeket **egymást kizáró eseményeknek** nevezzük.

Hangsúlyozzuk, hogy az eseményekkel végzett *logikai* műveletek a megfelelő részhalmazok közötti *halmazelméleti* műveletekkel ekvivalensek: az összegnek az unió, a szorzatnak a metszet felel meg. A diszjunkt halmazok egymást kizáró eseményeket reprezentálnak.

Az alábbi állítások a definíciók egyszerű következményei:

*Alapvető műveleti tulajdonságok:*

$$A + B = B + A, \qquad A \cdot B = B \cdot A,$$
$$(A + B) + C = A + (B + C), \qquad (A \cdot B) \cdot C = A \cdot (B \cdot C),$$
$$A \cdot (B + C) = A \cdot B + A \cdot C, \qquad A + B \cdot C = (A + B) \cdot (A + C).$$

*Egyéb műveleti tulajdonságok:*

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

*A de Morgan azonosságok:*

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

## 2.2. Az eseményalgebra fogalma

**Definíció.** Legyen $H$ egy tetszőleges alaphalmaz és $G$ a $H$ bizonyos részhalmazaiból készített olyan (nem üres) halmazrendszer, amely tartalmazza

- $H$-t;
- $H$ bármely elemének a komplementerét is;
- $H$ elemei bármely végtelen sorozatának az egyesítését (összegét) is.

Az ilyen tulajdonságú $G$ halmazrendszert **σ-algebrá**nak (**szigma-algebrá**nak) vagy **Boole-féle eseményalgebrá**nak nevezzük.

**Definíció. Eseményalgebrá**n a következőkben mindig egy, az Ω eseménytéren értelmezett σ-algebrát értünk.

## 2.3. A valószínűség-számítás axiómái

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

A tapasztalat azt mutatja, hogy nagyszámú, azonos körülmények között megismételt kísérlet esetén egy esemény relatív gyakorisága bizonyos stabilitást mutat, vagyis egy meghatározott számérték körül ingadozik, és az ingadozások a kísérletek számának növelésével általában egyre kisebbek lesznek. Azt a számot, amely körül az $A$ esemény relatív gyakorisága ingadozik, az **esemény valószínűségé**nek nevezzük és $P(A)$-val jelöljük.

Az egyes események valószínűségeinek létezésére és tulajdonságaira vonatkozóan feltételeket kell tennünk. Erre vonatkoznak a valószínűség-számítás V. N. Kolmogorovtól származó axiómái. Az axiómák ismertetése előtt nézzük meg, hogyan adódnak ezek a tapasztalati adatokból.

1. Nyilvánvaló, hogy egy $A$ esemény relatív gyakorisága 0 és 1 közötti érték, azaz

$$0 \le \frac{k_A}{n} \le 1.$$

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

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

   feltételnek is igaznak kell lennie.

2. A biztos esemény mindig bekövetkezik, ezért relatív gyakorisága mindig 1, azaz
$$\frac{k_\Omega}{n} = 1.$$
Ezért a
$$P(\Omega) = 1$$
egyenlőségnek is teljesülnie kell.
3. Ha az $A$ és $B$ egymást kizáró események, akkor az $A+B$ esemény $(k_A + k_B)$-szer következik be. A relatív gyakoriságokra áttérve innen azt kapjuk, hogy
$$\frac{k_{A+B}}{n} = \frac{k_A + k_B}{n} = \frac{k_A}{n} + \frac{k_B}{n}.$$
Mivel az $A+B$ esemény relatív gyakorisága az $A+B$ esemény valószínűsége körül ingadozik, az $A$ illetve $B$ esemény relatív gyakoriságai pedig az $A$ illetve $B$ esemény valószínűsége körül, ezért egymást kizáró események esetén a
$$P(A+B) = P(A) + P(B)$$
egyenlőségnek is fenn kell állnia.

A **valószínűség-számítás axiómái** ezek után a következők:

1. Az adott $\Omega$ eseménytér minden egyes $A$ eseményéhez tartozik egy 0 és 1 közé eső $P(A)$ szám, azaz
$$0 \le P(A) \le 1,$$
amelyet az $A$ esemény **valószínűségé**nek (**valószínűségi mértékének**) nevezünk.
2. A biztos esemény valószínűsége 1, azaz $P(\Omega) = 1$.
3. Az egymást páronként kizáró események összegének valószínűsége az egyes események valószínűségének összegével egyenlő, azaz ha az $A_1, A_2, \ldots, A_i, \ldots$ események esetén $A_j \cdot A_k = \varnothing$ ha $j \ne k$, akkor

$$P\left(\sum_i A_i\right) = \sum_i P(A_i).$$

(Ezt a tulajdonságot **σ-additivitás**nak nevezzük.)

A valószínűség tehát egy *halmazfüggvény*. Egyik szemléltetése a következő: ha Ω-n egységnyi tömegű festéket képzelünk eloszlatva, akkor $P(A)$ az *A*-ra jutó festékmennyiség.

## 2.4. Valószínűségek meghatározása

**2.4.1. tétel.** Az *A* esemény ellentettjének valószínűsége

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

**Bizonyítás.** Felhasználjuk, hogy $A \cdot \overline{A} = \varnothing$ és $A + \overline{A} = \Omega$. A 3. axiómát alkalmazva így

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

**Definíció.** Események egy összességét **teljes eseményrendszer**nek nevezzük, ha az események páronként kizárják egymást és az összegük a biztos esemény, azaz az $A_1, A_2, \dots, A_n$ események teljes eseményrendszert alkotnak, ha

$$A_i \cdot A_j = \varnothing \text{ ( ha } i \neq j \text{) és } A_1 + A_2 + \dots + A_n = \Omega.$$

**2.4.2. tétel.** Ha az $A_1, A_2, \dots, A_n$ események teljes eseményrendszert alkotnak, akkor a valószínűségük összege 1, vagyis

$$P(A_1) + P(A_2) + \dots + P(A_n) = 1.$$

**Bizonyítás.** A teljes eseményrendszer definíciója szerint $A_i \cdot A_j = \varnothing$ (ha $i \neq j$) és $A_1 + A_2 + \dots + A_n = \Omega$. A 2. és a 3. axiómát alkalmazva így

$$1 = P(\Omega) = P(A_1 + A_2 + \ldots + A_n) = P(A_1) + P(A_2) + \ldots + P(A_n) \Rightarrow$$
$$1 = P(A_1) + P(A_2) + \ldots + P(A_n).$$

**2.4.3. tétel.** Az $A$ és $B$ események $B - A$ különbségének valószínűsége

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

**Bizonyítás.** Felhasználjuk, hogy $B = (B - A) + A \cdot B$, és $(B - A) \cdot (A \cdot B) = \varnothing$. A 3. axiómát alkalmazva így

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

**2.4.4. következmény.** Ha $A \subset B$, akkor $P(B - A) = P(B) - P(A)$.

**2.4.5. tétel.** A valószínűség *monoton halmazfüggvény,* azaz, ha $A \subset B$, akkor $P(A) \le P(B)$.

**Bizonyítás.** Az **2.4.4. következményt** alkalmazva:

$$0 \le P(B - A) = P(B) - P(A), \text{ ahonnan } P(A) \le P(B).$$

**2.4.6. tétel.** Az $A$ és $B$ események összegének valószínűsége:

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

**Bizonyítás.** Felhasználjuk, hogy $A + B = A + (B - A)$ és $A \cdot (B - A) = \varnothing$. A 3. axiómát alkalmazva így

$$P(A + B) = P(A + (B - A)) = P(A) + P(B - A), \text{ ahonnan:}$$

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

Az előző szakasz végén leírt „festékes” szemléltetéssel: az $A \cup B$-re jutó festékmennyiség egyenlő az *A*-ra és *B*-re jutó festékmennyiségek összegével, de ekkor az $A \cap B$-re jutó festékmennyiséget kétszer vettük számításba, így azt le kell vonni az összegből.

A fenti tétel általánosítható több eseményre is *(szitaformulák)*. Említsük még meg a 3 eseményre való általánosítást, amely szemléletesen ugyanúgy látható be, mint a **2.4.6. tétel:**

**2.4.7. tétel.** Az *A*, *B* és *C* események összegének valószínűsége:

$$P(A+B+C)=$$
$$=P(A)+P(B)-P(C)-P(A\cdot B)-P(A\cdot C)-P(B\cdot C)+P(A\cdot BC)$$

## 2.5. A klasszikus valószínűségi mező

A klasszikus valószínűség-számítás olyan eseményekkel foglalkozik, amelyeknél a véges sok $a_1, a_2, \ldots, a_n$ elemi esemény egyenlő valószínűségű, azaz

$$P(a_i)=\frac{1}{n},\ \ (i=1,2,\ldots,n).$$

Ilyen esetekben a $k$-féleképpen bekövetkező $A$ esemény valószínűsége

$$P(A)=\frac{k}{n}=\frac{|A|}{|\Omega|},$$

ahol $k$ a kedvező esetek és $n$ az összes eset száma, $|.|$ pedig a megfelelő eseményt alkotó elemi események száma. Ebben az esetben mondjuk azt, hogy az események és ezek valószínűségei **klasszikus valószínűségi mezőt** alkotnak.

## 2.6. A geometriai valószínűségi mező

Ha egy kísérlettel kapcsolatos események egy véges geometriai alakzat részhalmazainak feleltethetők meg úgy, hogy az egyes események valószínűsége az eseményekhez rendelt részhalmaz geometriai mértékével (terület, térfogat, stb.) arányos, akkor az események valószínűségei **geometriai valószínűségi mezőt** alkotnak. Legyen $A$ egy ilyen kísérlettel kapcsolatos esemény. A kísérlettel kapcsolatban szóba jövő teljes alakzat mértéke legyen $M$, az $A$ eseménynek megfelelő részalakzaté pedig $m$. Az $A$ esemény valószínűsége ekkor tehát $P(A)=\frac{m}{M}=\frac{\mu(A)}{\mu(\Omega)}$ (ahol $\mu$ jelöli a megfelelő alakzat geometriai mértékét).

## 2.7. A feltételes valószínűség

Gyakran előfordul, hogy egy esemény valószínűségét olyan esetben kell megadni, amikor egy másik esemény is bekövetkezik. Ahogy a valószínűség axiómáinak megadása előtt is tettük, itt is visszatérünk a tapasztalati alapokhoz.

Legyen a $B$ esemény relatív gyakorisága $\frac{k_B}{n}$, az $A \cdot B$ esemény relatív gyakorisága pedig $\frac{k_{AB}}{n}$. Ekkor a

$$\frac{k_{AB}}{k_B} \quad \left( = \frac{k_{AB}}{n} \Big/ \frac{k_B}{n} \right)$$

hányadost az $A$ esemény $B$-re vonatkozó **feltételes relatív gyakoriságának** nevezzük.

A tapasztalat szerint ez is egy meghatározott számérték körül ingadozik, ez indokolja az alábbi definíciót:

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

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

kifejezéssel definiáljuk.

A feltételes valószínűség definícióját felhasználva $P(A \cdot B)$ kifejezhető

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

alakban, amit a **valószínűségek szorzási szabályá**nak nevezünk.

A szorzási szabály általánosítható több eseményre: könnyű látni, hogy pl. $P(A_1 \cdot A_2 \cdot A_3) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cdot A_2)$.

A $P(A\,|\,B)$ feltételes valószínűséget úgy is szemléltethetjük, hogy az eseményteret leszűkítjük *B*-re. Figyeljük meg a következő példát: ha tudjuk, hogy egy 2-gyermekes családban van fiú, akkor mi annak a valószínűsége, hogy a másik is fiú? Az eseménytér:

$$\Omega = \{\{F,F\},\{F,L\},\{L,F\},\{L,L\}\}$$

(ahol $F$ értelemszerűen fiút, $L$ lányt jelent), és $A = \{F,F\}$,
$B = \{\{F,F\},\{F,L\},\{L,F\}\}$, így:

$$P(A|B) = \frac{P(A\cdot B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{\frac{1}{4}}{\frac{1}{3}} = \frac{1}{3}.$$

Ha az eseményteret B-re szűkítjük le, akkor abban a keresett (feltételes) valószínűség egyetlen elemi esemény valószínűsége, azaz $\frac{1}{3}$. A kétféle gondolatmenettel tehát ugyanahhoz a feltételes valószínűséghez jutottunk.

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

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

**Bizonyítás.** Mivel a $B_1, B_2, \ldots, B_n$ teljes eseményrendszert alkotnak, ezért

$$B_i \cdot B_j = \varnothing \text{ és } \sum_{i=1}^{n} B_i = \Omega.$$

Ezt felhasználva:

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

alakba írható.

A kapott események páronként kizárják egymást, mivel

$$\left(A \cdot B_i\right) \cdot \left(A \cdot B_j\right) = A \cdot \left(B_i \cdot B_j\right) = \varnothing .$$

A 3. axióma alapján ezért

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

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

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

és ezzel igazoltuk a tételt.

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

$$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{\sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)} \qquad (k = 1,2,\ldots n)$$

**Bizonyítás:**

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

ahol a számlálóban a szorzási szabályt, a nevezőben pedig a teljes valószínűség tételét alkalmaztuk.

## 2.8. Események függetlensége

**Definíció. Két eseményt független**nek nevezünk, ha az együttes bekövetkezésük valószínűsége a két esemény bekövetkezésének szorzata, azaz

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

Az $A_1, A_2, \ldots, A_n$ **eseményeket páronként függetlenek**nek nevezzük, ha bárhogyan is választunk ki közülük két eseményt, ezek szorzatának valószínűsége egyenlő az egyes események valószínűségeinek szorzatával, azaz

$$P(A_i \cdot A_j) = P(A_i) \cdot P(A_j), \quad 1 \le i < j \le n.$$

Az $A_1, A_2, \ldots, A_n$ **eseményeket (teljesen) függetlenek**nek nevezzük, ha bárhogyan is választunk ki közülük $k$ eseményt $(k = 2, 3, \ldots, n)$, ezek szorzatának valószínűsége egyenlő az egyes események valószínűségeinek szorzatával.

Az $A_1, A_2, \ldots, A_n$ események teljes függetlensége esetén tehát a

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

egyenlőségeknek mind teljesülniük kell. Az itt szereplő követelmények száma:

$$\sum_{k=2}^{n} \binom{n}{k} = 2^n - n - 1.$$

**2.8.1. tétel.** Ha az $A$ és a $B$ független események, akkor

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

**Bizonyítás:** $P(A|B) = \dfrac{P(A \cdot B)}{P(B)} = \dfrac{P(A) \cdot P(B)}{P(B)} = P(A)$, és

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

**2.8.2. tétel.** Ha az $A$ és a $B$ független események, akkor az $A$ és a $\overline{B}$, az $\overline{A}$ és a $B$, valamint az $\overline{A}$ és a $\overline{B}$ is független események.

**Bizonyítás.** A valószínűségek meghatározására vonatkozó tételeket felhasználva

$$\begin{aligned} P(A\cdot\overline{B}) &= P(A-B) = P(A) - P(A\cdot B) \\ &= P(A) - P(A)\cdot P(B) = P(A)\cdot(1-P(B)), \\ &= P(A)\cdot P(\overline{B}), \end{aligned}$$

ami igazolja az $A$ és a $\overline{B}$ események függetlenségét. Az $\overline{A}$ és a $B$, valamint az $\overline{A}$ és $\overline{B}$ események függetlensége ezzel megegyező módon igazolható.

Két *pozitív* valószínűségű esemény nem lehet egyszerre független és egymást kizáró – vajon miért?

**1. példa.** (Műveletek eseményekkel.) Két helység között három távbeszélő vonalon folytathatunk beszélgetést. Jelentse $A$ azt, hogy az első vonal hibás, $B$ azt, hogy a második és $C$, hogy a harmadik. Fejezzük ki $A$, $B$, $C$ segítségével a következő eseményeket!

*a)* Csak az első vonal hibás.
*b)* Legalább az egyik hibás.
*c)* Mindhárom hibás.
*d)* Pontosan egy vonal hibás.
*e)* A második nem hibás, de az első és a harmadik közül legalább az egyik hibás.

**Megoldás.**

*a)* $A\cdot\overline{B}\cdot\overline{C}$ ,
*b)* $A+B+C$,
*c)* $A\cdot B\cdot C$,
*d)* $A\cdot\overline{B}\cdot\overline{C}+\overline{A}\cdot B\cdot\overline{C}+\overline{A}\cdot\overline{B}\cdot C$ ,
*e)* $\overline{B}\cdot(A+C)$.

**2. példa.** (Műveletek eseményekkel.) Igazoljuk, hogy tetszőleges $A$ és $B$ eseményekre fennáll az alábbi azonosság:

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

**Megoldás.** A bal oldali eseményből indulunk ki és azonos átalakításokat hajtunk végre:

$$\begin{aligned}(A+B)-A\cdot B &= (A+B)\cdot\overline{(A\cdot B)} = (A+B)\cdot\left(\overline{A}+\overline{B}\right)\\ &= A\cdot\left(\overline{A}+\overline{B}\right)+B\cdot\left(\overline{A}+\overline{B}\right) = A\cdot\overline{A}+A\cdot\overline{B}+B\cdot\overline{A}+B\cdot\overline{B}\\ &= \varnothing + A\cdot\overline{B}+B\cdot\overline{A}+\varnothing = A\cdot\overline{B}+\overline{A}\cdot B.\end{aligned}$$

**3. példa.** (Műveletek eseményekkel.) Igazoljuk, hogy tetszőleges $A$, $B$ és $C$ eseményekre fennáll az alábbi azonosság:

$$(A+B)\cdot\left(\overline{A}+C\right) = A\cdot C+\overline{A}\cdot B$$

**Megoldás.** A bal oldali eseményből indulunk ki és azonos átalakításokat hajtunk végre:

$$\begin{aligned}(A+B)\cdot\left(\overline{A}+C\right) &= A\cdot\left(\overline{A}+C\right)+B\cdot\left(\overline{A}+C\right)\\ &= A\cdot\overline{A}+A\cdot C+B\cdot\overline{A}+B\cdot C = \varnothing + A\cdot C+B\cdot\overline{A}+B\cdot C\\ &= A\cdot C+B\cdot\overline{A}+B\cdot C\cdot\Omega\\ &= A\cdot C+B\cdot\overline{A}+B\cdot C\cdot\left(A+\overline{A}\right)\\ &= A\cdot C+B\cdot\overline{A}+B\cdot C\cdot A+B\cdot C\cdot\overline{A}\\ &= A\cdot C+\overline{A}\cdot B.\end{aligned}$$

Az utolsó azonosságnál felhasználtuk, hogy

$$B\cdot C\cdot A\subseteq A\cdot C \text{ és } B\cdot C\cdot\overline{A}\subseteq B\cdot\overline{A}.$$

**4. példa.** (Műveletek eseményekkel.) Igazoljuk, hogy tetszőleges $A$, $B$, $C$ és $D$ eseményekre fennáll az alábbi azonosság:

$$(A-B)\cdot(C-D) = A\cdot C-(B+D)$$

**Megoldás.** A bal oldali eseményt átalakíthatjuk az alábbi módon:

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

A jobb oldalt is átalakíthatjuk így:

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

Mindkét oldalt ugyanarra az eseményre vezettük vissza, így igazoltuk, hogy az azonosság.

**5. példa.** (Valószínűségek meghatározása.) Igazoljuk, hogy tetszőleges $A$ és $B$ eseményekre fennáll az alábbi azonosság:

$$P(A \cdot B) = 1 - P\left(\overline{A} + \overline{B}\right)$$

**Megoldás.** Az $A \cdot B$ szorzatot

$$A \cdot B = \Omega \cdot (A \cdot B) = \Omega - \overline{(A \cdot B)} = \Omega - \left(\overline{A} + \overline{B}\right)$$

alakra hozhatjuk. Mivel $\left(\overline{A} + \overline{B}\right) \subseteq \Omega$ és $P(\Omega) = 1$, ezért

$$P\left(\Omega - \left(\overline{A} + \overline{B}\right)\right) = P(\Omega) - P\left(\overline{A} + \overline{B}\right) = 1 - P\left(\overline{A} + \overline{B}\right).$$

Így

$$P(A \cdot B) = 1 - P\left(\overline{A} + \overline{B}\right),$$

amit igazolni akartunk.

**6. példa.** (Klasszikus valószínűségi mező.) Tíz golyó van egy dobozban. Közülük kettő fehér, a többi fekete. Kiveszünk találomra öt golyót, (egyszerre, visszatevés nélkül). Mekkora a valószínűsége annak, hogy éppen egy fehér golyó lesz közöttük?

**Megoldás.** A kiválasztás lehetőségeinek a számát megkapjuk, ha a 10 elem ötödosztályú kombinációinak a számát vesszük:

$$n = C_{10,5} = \binom{10}{5}.$$

Ezután nézzük a kedvező eseteket! A két fehérből választunk egyet és a nyolc feketéből négyet:

$$k = C_{2,1} \cdot C_{8,4} = \binom{2}{1} \cdot \binom{8}{4}$$

Ha a vizsgált eseményt $A$-val jelöljük, akkor

$$P(A) = \frac{k}{n} = \frac{C_{2,1} \cdot C_{8,4}}{C_{10,5}} = \frac{5}{9}.$$

Tehát $\frac{5}{9}$ annak a valószínűsége, hogy pontosan egy fehér golyó lesz az öt között.

**7. példa.** (Klasszikus valószínűségi mező.) Adjuk meg annak a valószínségét, hogy egy totószelvényt vaktában kitöltve, az első 13 mérkőzés eredménye közül éppen 11-et találunk el!

**Megoldás.** A szelvény kitöltésére az összes lehetőségek számát – mivel 13 helyre választunk az 1, 2, $x$ elemekből – 3 elem 13-ad osztályú ismétléses variációinak száma adja:

$$n = V_{3,13}^{ism} = 3^{13}.$$

Ezután a következő lehetőségekkel számolunk. A 13 mérkőzés eredményéből 11-et a szelvényen el kell találni, 2-t viszont nem. A 11 eltalált eredmény annyiféleképpen választható ki a 13-ból, mint amennyi 13 elem 11-ed osztályú kombinációinak a száma. A maradék két mérkőzés mindegyikére $2-2$ hibás tippünk van, vagyis ezen tippek számát 2 elem 2-odosztályú ismétléses variációinak száma adja. Így a kedvező esetek száma:

$$k = C_{13,11} \cdot V_{2,2}^{ism} = \binom{13}{11} \cdot 2^2.$$

Ha a vizsgált eseményt $A$-val jelöljük, akkor

$$P(A) = \frac{k}{n} = \frac{C_{13,2} \cdot V_{2,2}^{ism}}{V_{3,13}^{ism}} = \frac{\binom{13}{11} \cdot 2^2}{3^{13}} \approx 0{,}000196.$$

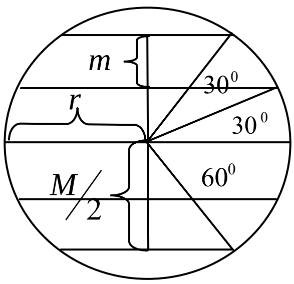
Az MS Excel használata esetén a $P(A)$ valószínűséget az

= KOMBINÁCIÓK(13;11) * (2^2)/(3^13)

kifejezés kiértékelésével határozhatjuk meg.

**8. példa.** (Geometriai valószínűségi mező.) Egy mesterséges égitest kering a Föld körül úgy, hogy a 60°-os északi és a 60°-os déli szelességi kör között bárhol leszállhat. Mennyi a valószínűsége annak, hogy a 60°-os és a 30°-os északi szélességi kör között ér földet, ha a gömbfelület egy darabjára történő leszállás valószínűsége a felületdarab felszínével arányos?

**Megoldás.** A leszállás az $r$ sugarú gömbnek tekintett Föld egy gömbövére történik. Ennek magassága a tengelyen átmenő síkmetszeten látható:




$$M = 2r \cdot \sin\left(60^0\right) = 2r \cdot \frac{\sqrt{3}}{2} = \sqrt{3}r.$$

Így a leszálláskor szóba jövő gömböv felszíne:

$$F = 2\pi r M = 2\sqrt{3}\pi r^2.$$

Az

$$A = \{\text{A 60°-os és 30°-os északi szél. kör között száll le.}\}$$

esemény szempontjából kedvező leszállási pontok ugyancsak egy gömbövön vannak, melynek magassága:

$$m = r \cdot \sin\left(60^0\right) - r \cdot \sin\left(30^0\right) = r \cdot \left(\frac{\sqrt{3}}{2} - \frac{1}{2}\right) = r \cdot \frac{\sqrt{3}-1}{2}.$$

Ennek a kisebb gömbövnek a felszíne:

$$f = 2\pi r m = \left(\sqrt{3}-1\right)\pi r^2.$$

Az $A$ esemény valószínűsége:

$$P(A) = \frac{f}{F} = \frac{\left(\sqrt{3}-1\right)\pi r^2}{2\sqrt{3}\pi r^2} = \frac{\left(\sqrt{3}-1\right)}{2\sqrt{3}} \approx 0.21.$$

Vegyük észre, hogy ez a valószínűség egyezik az $\frac{m}{M}$ hányadossal is.

**9. példa.** (Feltételes valószínűség.) A 32 lapos magyar kártyából 3 lapot húzunk ki egymás után, visszatevés nélkül. Mennyi a valószínűsége annak, hogy az első kihúzott lap hetes, a második kilences és a harmadik szintén hetes?

**Megoldás.** Legyen $A_1$ az az esemény, hogy az első lap hetes. $A_2$ jelentse azt, hogy a második kihúzott lap kilences. Azt az eseményt, hogy a harmadik választott lap hetes jelöljük $A_3$-mal. Az $A_1 A_2 A_3$ szorzat valószínűségét a szorzási szabályt alkalmazva, a

$$P(A_1 A_2 A_3) = P(A_3 | A_1 A_2) \cdot P(A_2 \mid A_1) \cdot P(A_1)$$

összefüggéssel számíthatjuk ki.

Az $A_1$ esemény szempontjából a kedvező esetek száma $k_1 = 4$, mert ennyi hetes van a csomagban. Az összes lehetőségek száma az első húzásra $n_1 = 32$. Az $A_1$ esemény valószínűsége így

$$P(A_1) = \frac{k_1}{n_1} = \frac{4}{32} = \frac{1}{8}.$$

$A_1$ teljesülése esetén az $A_2$ esemény megvalósulására a kedvező esetek száma $k_2 = 4$, mert elsőre nem kilencest húztunk. Mivel az összes esetek száma ekkor $n_2 = 31$, így azt kapjuk, hogy

$$P(A_2 \mid A_1) = \frac{k_2}{n_2} = \frac{4}{31}.$$

Az $A_1 A_2$ esemény bekövetkezését feltételezve, vizsgáljuk az $A_3$ eseményt. A kedvező esetek száma ekkor $k_3 = 3$, mert egy hetest már előzőleg kihúztunk. Az összes lehetőségek száma a harmadik húzásra $n_3 = 30$. Így a feltételes valószínűség

$$P(A_3 \mid A_1 A_2) = \frac{k_3}{n_3} = \frac{3}{30} = \frac{1}{10}.$$

A kapott értékeket behelyettesítve

$$P(A_1 A_2 A_3) = \frac{1}{10} \cdot \frac{4}{31} \cdot \frac{1}{8} = \frac{1}{620}.$$

Tehát $\frac{1}{620}$ a valószínűsége annak, hogy a három lap kihúzására az előírt feltételek teljesülnek.

**10. példa.** (A teljes valószínűség tétele.) Négy termelőtől egy tehergépkocsival almát szállítanak egy üzletbe. Az első termelőtől a mennyiség $\frac{1}{10}$ része származik, amelyből 40% első osztályú. A második termelőtől a tétel $\frac{1}{4}$ részét szállítják, amely 50%-ban első osztályú. A harmadiktól rendelték meg a mennyiség $\frac{2}{5}$ részét, ebből 20% első osztályú. A többi gyümölcs a negyedik termelőtől származik és mind első osztályú. Mekkora a valószínűsége annak, hogy az üzletben e szállítmányból találomra kiválasztva egy almát, az első osztályú?

**Megoldás.** Jelöljük a vizsgált eseményt $A$-val. A $B_i$ $(i = 1,2,3,4)$ esemény jelentse azt, hogy az $i$-edik termelőtől való az alma. A $B_i$ események valószínűségei rendre a következők:

$$P(B_1) = \frac{1}{10}, P(B_2) = \frac{1}{4}, P(B_3) = \frac{2}{5},$$

$$P(B_4) = 1 - \left(\frac{1}{10} + \frac{1}{4} + \frac{2}{5}\right) = \frac{1}{4}.$$

Ezután az $A$ esemény $B_i$ esemény melletti feltételes valószínűségét írjuk fel. Ezek rendre a következők:

$$P(A|B_i) = \frac{4}{10},\ P(A|B_2) = \frac{5}{10},\ P(A|B_3) = \frac{2}{10},\ P(A|B_4) = \frac{10}{10}.$$

A teljes valószínűség tételét alkalmazzuk:

$$P(A)=\sum_{i=1}^{4} P(A|B_i)\cdot P(B_i)=\frac{4}{10}\cdot\frac{1}{10}+\frac{5}{10}\cdot\frac{1}{4}+\frac{2}{10}\cdot\frac{2}{5}+\frac{10}{10}\cdot\frac{1}{4}$$
$$=\frac{99}{200}=0{,}495.$$

Tehát 49,5% a valószínűsége annak, hogy a szállítmányból az üzletben véletlenszerűen választva egy almát az első osztályú.

**11. példa.** (A teljes valószínűség tétele.) Izzólámpákat 100 darabos csomagolásban szállítnak. Az előző megfigyelésekből ismert, hogy a tételek között azonos arányban fordul elő 0, 1, 2, 3, 4 hibás darabot tartalmazó. Mekkora a valószínűsége annak, hogy a tételből véletlenszerűen 3 égőt kiválasztva, mindhárom hibátlan lesz?

**Megoldás.** Jelöljük a vizsgált eseményt $A$-val. A $B_i$ $(i=0,1,2,3,4)$ jelentse azt az eseményt, hogy a tétel $i$ darab hibás égőt tartalmaz. Ezeknek az eseményeknek a valószínűségei azonosak, mégpedig

$$P(B_i)=\frac{1}{5}, \qquad (i=0,1,2,3,4).$$

Most állapítsuk meg az $A$ esemény feltételes valószínűségét $B_i$ bekövetkezése mellett. A 100 darabos tételből a 3 izzólámpa kiválasztására az összes lehetőségek száma:

$$n=C_{100,3}=\binom{100}{3}$$

A kedvező lehetőségek száma 3 jó tétel kiválasztására, ha a tételben $i$ darab hibás van:

$$k=C_{100-i,3}=\binom{100-i}{3}$$

Így a feltételes valószínűség:

$$P(A|B_i)=\frac{k}{n}=\frac{\binom{100-i}{3}}{\binom{100}{3}},\ (i=0,1,2,3,4).$$

A teljes valószínűség tétele szerint:

$$P(A)=\sum_{i=0}^{4}P(A|B_i)\cdot P(B_i)=\sum_{i=0}^{4}\frac{\binom{100-i}{3}}{\binom{100}{3}}\cdot\frac{1}{5}\approx 0{,}94\,.$$

Tehát körülbelül 94% annak a valószínűsége, hogy a tételből 3 hibátlan izzót választunk.
Az MS Excel használata esetén a $P(A)$ valószínűséget az

$$\begin{gathered}=(1/5)*(1/\text{KOMBINÁCIÓK}(100;3))*\\(\text{KOMBINÁCIÓK}(100;3)+\text{KOMBINÁCIÓK}(99;3)\\+\text{KOMBINÁCIÓK}(98;3)+\text{KOMBINÁCIÓK}(97;3)\\+\text{KOMBINÁCIÓK}(96;3))\end{gathered}$$

kifejezés kiértékelésével határozhatjuk meg.

Pontosabb meggondolásokkal megmutatható, hogy az egyes tételekbe kerülő hibás darabok száma inkább *Poisson-eloszlást* követ (ld. a következő szakaszokban, így az, hogy a $B_i$ események mind egyenlő valószínűségűek, csak durva első közelítésnek tekinthető.

**12. példa.** (Bayes-tétel.) Egy gyárban három gép gyárt csavarokat. A termék 25%-a az első, 35%-a a második és 40%-a a harmadik gépen készül. Az első gép 5%, a második gép 4%, míg a harmadik gép 2%-ban termel selejtet. Ha egy találomra kiválasztott csavar selejtes, akkor mennyi a valószínűsége annak, hogy azt az első gépen gyártották?

**Megoldás.** Jelentse $A$ azt az eseményt, hogy a kiválasztott csavar selejtes. Jelöljük $B_i$-vel $(i = 1, 2, 3)$ azt az eseményt, hogy a csavart az $i$-edik gépen gyártották. Ezeknek az eseményeknek a valószínűségei:

$$P(B_1) = 0{,}25; \quad P(B_2) = 0{,}35; \quad P(B_3) = 0{,}40.$$

Az $A$ esemény $B_i$ eseményekre vonatkozó feltételes valószínűségei:

$$P(A \mid B_1) = 0{,}05; \quad P(A \mid B_2) = 0{,}04; \quad P(A \mid B_3) = 0{,}02.$$

A keresett valószínűség a $P(B_1 \mid A)$ feltételes valószínűség. A Bayes-tétel szerint

$$P(B_1 \mid A) = \frac{P(A \mid B_1) \cdot P(B_1)}{\sum_{i=1}^{4} P(A \mid B_i) \cdot P(B_i)}$$

$$= \frac{0{,}05 \cdot 0{,}25}{0{,}05 \cdot 0{,}25 + 0{,}04 \cdot 0{,}35 + 0{,}02 \cdot 0{,}40} = \frac{0{,}125}{0{,}345} \approx 0{,}36.$$

Tehát körülbelül 36% annak a valószínűsége, hogy a kiválasztott selejtes csavart az első gépen gyártották.

**13. példa.** (Független események.) Egy urnában négy egyforma papírlap található. Mindegyikére három számjegy van írva egymás mellé, mégpedig az elsőre $(0; 0; 0)$, a másodikra $(0; 1; 1)$, a harmadikra $(1; 0; 1)$, és a negyedikre $(1; 1; 0)$. Húzzunk ki egy lapot. Feltételezzük, hogy mindegyik lap egyforma valószínűséggel húzható. Jelentse $A_i$ azt az eseményt, hogy olyan lapot húztunk, amelyiknek az $i$-edik jegye 1-es $(i = 1, 2, 3)$. Mutassuk meg, hogy az $A_i$ események páronként függetlenek, együttesen azonban nem!

**Megoldás**. Nyilván

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{2},$$

továbbá

$$P(A_1 \cdot A_2) = \frac{1}{4} = P(A_1) \cdot P(A_2),$$

$$P(A_1 \cdot A_3) = \frac{1}{4} = P(A_1) \cdot P(A_3),$$

$$P(A_2 \cdot A_3) = \frac{1}{4} = P(A_2) \cdot P(A_3),$$

vagyis az $A_1$, $A_2$ és $A_3$ események páronként valóban függetlenek. Összességükben azonban nem, mert

$$P(A_1 \cdot A_2 \cdot A_3) = 0 \neq \frac{1}{8} = P(A_1) \cdot P(A_2) \cdot P(A_3).$$

**14. példa.** (Független események.) Egy gyárban két gép működési idejére végeztek megfigyeléseket és megállapították, hogy az I-es gép átlagban munkaidejének 60%-ában dolgozik, a II-es gép pedig a munkaidő 70%-ában. Határozzuk meg az alábbi események valószínűségeit!

*a)* Mindkét gép dolgozik.
*b)* Legalább az egyik dolgozik.
*c)* Csak az egyik dolgozik.
*d)* Mindkét gép áll.

**Megoldás.**
Legyen $A = \{\text{Az } I. \text{gép dolgozik}\}$ és $B = \{\text{A } II. \text{gép dolgozik}\}$.
Az $A$ és $B$ események függetleneknek tekinthetők.
Mivel most $P(A) = 0{,}6$ és $P(B) = 0{,}7$ ezért:

*a)* $P(A \cdot B) = P(A) \cdot P(B) = 0{,}42$.
*b)* $P(A + B) = P(A) + P(B) - P(A \cdot B) = 0{,}88$.
*c)* $P(A \cdot \overline{B} + \overline{A} \cdot B) = P(A \cdot \overline{B}) + P(\overline{A} \cdot B)$
$= P(A) \cdot P(\overline{B}) + P(\overline{A}) \cdot P(B) = 0{,}46$.
*d)* $P(\overline{A} \cdot \overline{B}) = P(\overline{A}) \cdot P(\overline{B}) = 0{,}4 \cdot 0{,}3 = 0{,}12$.

**15. példa.** (Független események.) Valaki két lottószelvényt tölt ki egymástól függetlenül. Mekkora annak a valószínűsége, hogy nyer?

**Megoldás.** Jelöljük $A_i$-vel azt az eseményt, hogy az $i$-edik $(i=1,2)$ lottószelvény nem nyer. Az $A_i$ esemény valószínűsége

$$P(A_i)=\frac{\binom{5}{0}\cdot\binom{85}{5}+\binom{5}{1}\cdot\binom{85}{4}}{\binom{90}{5}}.$$

Az $A_i$ $(i=1,2)$ események függetleneknek tekinthetők. A keresett valószínűség így

$$P(\Omega-A_1\cdot A_2)=P(\Omega)-P(A_1\cdot A_2)$$

$$=1-P(A_1)\cdot P(A_2)=1-\left(\frac{\binom{5}{0}\cdot\binom{85}{5}+\binom{5}{1}\cdot\binom{85}{4}}{\binom{90}{5}}\right)^2\approx 0{,}046.$$

Az MS Excel használata esetén a $P(\Omega-A_1\cdot A_2)$ valószínűséget az

=1-((1/KOMBINÁCIÓK(90;5))
*(KOMBINÁCIÓK(5;0)*KOMBINÁCIÓK(85;5)
+KOMBINÁCIÓK(5;1)*KOMBINÁCIÓK(85;4)))^2

kifejezés kiértékelésével határozhatjuk meg.

## 2.9. A valószínűségi változó és jellemzői

A gyakorlatban előforduló kísérletek túlnyomó részében a kísérletek egyes kimeneteleivel, az elemi eseményekkel, egyúttal számértékek is adódnak. Például kockadobás esetén a pontok száma. Az említett kísérletnél minden elemi eseményhez egyetlen számérték tartozik, és e számérték megjelenése az elemi esemény bekövetkezésétől függ, vagyis függvényről van szó.

**Definíció.** Ha egy kísérlettel kapcsolatos elemi események mindegyikéhez egyértelműen hozzárendelünk egy-egy valós számot, akkor az elemi események $\Omega$ halmazán egy $\xi:\Omega\to IR$ függvényt értelmezünk. Ezt a függvényt *valószínűségi változónak* nevezzük.

**Definíció.** Ha egy kísérlet során az $a_i$ elemi esemény következik be és a valószínűségi változó megadásakor ehhez az $x_i$ értéket rendeltük, akkor azt mondjuk, hogy a $\xi$ valószínűségi változó az $x_i$ értéket veszi fel, az $x_i$ -t pedig $\xi$ egy lehetséges értékének nevezzük.

A gyakorlatban az a fontos, hogy a $P(\xi \in I)$ valószínűségek meghatározhatók legyenek, azaz $\{\xi \in I\}$ *esemény* legyen, ahol $I$ leggyakrabban valamilyen intervallum, vagy intervallumok egyesítése. A valószínűségi változók bevezetése lényegében kényelmi szempont: a hasonló jelenségek modellezésének egy eszköze.

**Példa**. A $\xi_A : \Omega \to \mathbb{R}$

$$\xi_A := \begin{cases} 1, & \text{ha az } A \text{ bekövetkezik} \\ 0, & \text{ha az } A \text{ nem következik be} \end{cases}$$

valószínűségi változót az $A$ esemény *indikátorváltozójának* nevezzük *(Bernoulli-változó)*.

**Definíció**. Egy valószínűségi változót *diszkrétnek* nevezünk, ha az véges sok értéket vesz fel, vagy értékei sorozatba rendezhetők (a lehetséges értékeinek halmaza megszámlálható).

**Definíció**. Egy valószínűségi változót *folytonosnak* nevezünk, ha értékkészlete nem megszámlálható (jellemzően: korlátos vagy nem korlátos *intervallum,* vagy ilyenek egyesítése), és minden egyes értéket 0 valószínűséggel vesz fel.

Bevezethetők még ún. *kevert eloszlású* valószínűségi változók, melyek diszkrét és folytonos részekből állnak. Ilyen lehet pl. egy sorompó nyitvatartási szöge: ez a 0 és a $\frac{\pi}{2}$ értéket pozitív valószínűséggel, a köztes értékeket 0 valószínűséggel veszi fel. Ebben a jegyzetben a későbbiekben csak diszkrét és folytonos valószínűségi változókról lesz szó.

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

Mivel a $\{\xi = x_i\}$ $(i = 1,2,\ldots)$ események teljes eseményrendszert alkotnak, azért a $\sum_i p_i = 1$ egyenlőség mindig teljesül: az összeg vagy egy véges összeget jelent (ha $\xi$ értékkészlete véges), vagy egy konvergens sor összegét (ha $\xi$ értékkészlete megszámlálhatóan végtelen).

**Definíció.** Egy $\xi$ diszkrét valószínűségi változó lehetséges értékeihez tartozó bekövetkezési valószínűségek összességét $\xi$ *eloszlásának* nevezzük: ha $\xi$ lehetséges értékei az $x_k$ számok $(k = 1,2,\ldots)$, akkor $\xi$ eloszlása a $p_k = P(\xi = x_k)$ $(k = 1,2,\ldots)$ bekövetkezési valószínűségek összessége.

**Definíció.** Egy $\xi$ valószínűségi változó *eloszlásfüggvényének* nevezzük azt az $F$ függvényt, amely minden valós $x$ értékhez hozzárendeli annak valószínűségét, hogy a $\xi$ valószínűségi változó $x$-nél kisebb értéket vesz fel:

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

**2.9.1. tétel.** Tetszőleges $\xi$ valószínűségi változó $F$ eloszlásfüggvénye az alábbi tulajdonságokkal rendelkezik:

1. $R_F = [0,1]$ (értékkészlete a $[0,1]$ intervallum)
2. $F(a) \le F(b)$, ha $a \le b$ (monoton növekedő)
3. $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to+\infty} F(x) = 1$
4. $P(\xi \ge a) = 1 - P(\xi < a) = 1 - F(a)$

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

6. $P(\xi = a) = \lim\limits_{x \to a+0} F(x) - F(a)$.

**Definíció.** Egy $\xi$ valószínűségi változót *folytonos eloszlásúnak* (pontosabban: *abszolút folytonos eloszlásúnak*) mondunk, ha az eloszlásnak van *sűrűségfüggvénye*, azaz van olyan $f : \mathbb{R} \to \mathbb{R}_+$ függvény, hogy minden $I \subset \mathbb{R}$ intervallumra teljesül, hogy

$$\int_I f(x)dx = P(\xi \in I)$$

**2.9.2. tétel.** Tetszőleges folytonos eloszlású $\xi$ valószínűségi változó $f$ sűrűségfüggvénye az alábbi tulajdonságokkal rendelkezik:

1. $F' = f$
2. $f(x) \ge 0, \quad x \in D_f$
3. $\int_{-\infty}^{+\infty} f(x)dt = 1$
4. $P(\xi < a) = F(a) = \int_{-\infty}^{a} f(t)dt$
5. $P(\xi \ge a) = 1 - F(a) = 1 - \int_{-\infty}^{a} f(t)dt = \int_{a}^{+\infty} f(t)dt$
6. $P(a \le \xi < b) = F(b) - F(a) = \int_{a}^{b} f(t)dt$ .

Folytonos eloszlású valószínűségi változók esetén minden valós $c$ számra $P(\xi = c) = 0$, ezért pl. a $P(a \le \xi < b)$, $P(a < \xi < b)$, $P(a \le \xi \le b)$, $P(a < \xi \le b)$ valószínűségek mind megegyeznek.

Azt a számot, amely körül a ξ valószínűségi változó megfigyelt értékeinek átlaga ingadozik a valószínűségi változó *várható értékének* nevezzük és matematikailag a következő módon definiálhatjuk.

**Definíció.** Egy diszkrét ξ valószínűségi változó *várható értékén* a

$$m := M(\xi) := \sum_i x_i \cdot p_i \quad (\text{ahol } p_i = P(\xi = x_i))$$

összeget értjük, amennyiben $\sum_i |x_i| \cdot p_i < \infty$. Egyébként azt mondjuk, hogy *a várható érték nem létezik.*

**Definíció.** Egy folytonos eloszlású ξ valószínűségi változó *várható értékén* az

$$m := M(\xi) := \int_{-\infty}^{\infty} x \cdot f(x)dx$$

integrált értjük, amennyiben $\int_{-\infty}^{\infty} |x| \cdot f(x)dx < \infty$. Egyébként azt mondjuk, hogy *a várható érték nem létezik.*

A várható értéket sokféleképp szemléltethetjük, egy lehetséges szemléltetés a következő. Képzeljünk el egy egyenesen az $x_1, x_2, \ldots$ pontokba koncentrált $p_1, p_2, \ldots$ nagyságú tömegeket, melyek összege épp egységnyi. Akkor ezen (diszkrét tömegpontokból álló) rendszer *s* súlypontjára nézve az előjeles forgatónyomaték-összeg zérus: $\sum_i p_i \cdot (s - x_i) = 0$, ahonnan a súlypont-koordináta épp azon valószínűségi változó várható értéke, mely az $x_1, x_2, \ldots$ értékeket rendre $p_1, p_2, \ldots$ valószínűségekkel veszi fel: $s = \sum_i x_i \cdot p_i$. Hasonló szemléltetés érvényes folytonos valószínűségi változókra is: itt a tömegeloszlás értelemszerűen folytonos lesz, a súlypont koordinátája pedig ismét a várható értékkel egyezik.

**2.9.3. tétel.** Legyen $\xi$ egy tetszőleges valószínűségi változó és jelölje $\eta := a\cdot\xi^2 + b\cdot\xi + c$, ahol $a$, $b$ és $c$ tetszőleges valós számok. Ekkor az $\eta$ valószínűségi változó várható értéke

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

amennyiben $\xi$ és $\xi^2$ várható értéke létezik.

**Bizonyítás**. Ha $\xi$ diszkrét valószínűségi változó, akkor az összefüggést a

$$M(\eta) = \sum_{i=1}^{\infty} y_i\cdot p_i = \sum_{i=1}^{\infty}\left(a\cdot x_i^2 + b\cdot x_i + c\right)\cdot p_i$$

$$= a\cdot\sum_{i=1}^{\infty} x_i^2\cdot p_i + b\cdot\sum_{i=1}^{\infty} x_i\cdot p_i + c\sum_{i=1}^{\infty} p_i = a\cdot M\left(\xi^2\right) + b\cdot M(\xi) + c$$

egyenlőségből kapjuk. Ha $\xi$ folytonos eloszlású valószínűségi változó, akkor az összefüggés a

$$M(\eta) = \int_{-\infty}^{\infty} y(x)\cdot f(x)dx = \int_{-\infty}^{\infty}\left(a\cdot x^2 + b\cdot x + c\right)\cdot f(x)dx$$

$$= a\cdot\int_{-\infty}^{\infty} x^2\cdot f(x)dx + b\cdot\int_{-\infty}^{\infty} x\cdot f(x)dx + c\int_{-\infty}^{\infty} f(x)dx$$

$$= a\cdot M\left(\xi^2\right) + b\cdot M(\xi) + c$$

azonosságból következik.

**Definíció**. Egy $\xi$ valószínűségi változó *szórásnégyzetének* vagy *varianciájának* a $(\xi - M(\xi))^2$ valószínűségi változó várható értékét értjük:

$$\sigma^2 := D^2(\xi) := M\left((\xi - M(\xi))^2\right),$$

feltéve, hogy $\xi$ és $(\xi - M(\xi))^2$ várható értéke létezik. Ennek négyzetgyökét nevezzük a $\xi$ valószínűségi változó *szórásának* (jele: $D(\xi)$).

**2.9.4. tétel.** Ha a $\xi$ valószínűségi változó és annak négyzetének várható értéke is létezik, akkor a szórása is létezik éspedig

$$D^2(\xi) = M\left(\xi^2\right) - M^2(\xi).$$

**Bizonyítás**. A 2.9.3. tétel alapján

$$\begin{aligned} D^2(\xi) &= M\left((\xi - M(\xi))^2\right) = M\left(\xi^2 - 2\cdot M(\xi)\cdot\xi + M^2(\xi)\right) = \\ &= M\left(\xi^2\right) - 2\cdot M(\xi)\cdot M(\xi) + M^2(\xi) = M\left(\xi^2\right) - M^2(\xi), \end{aligned}$$

ami igazolja az állításunkat.

**2.9.5. tétel.** (A szórás meghatározása.)

1. Legyen $\xi$ egy, az $x_1, x_2, \ldots, x_n$ különböző értéket felvevő diszkrét valószínűségi változó. Ekkor

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

illetve a 2.9.4. tétel alapján

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

2. Legyen $\xi$ egy olyan diszkrét valószínűségi változó, amelynek az értékei sorozatba rendezhetők.

Ha $\sum_{i=1}^{\infty} (x_i - M(\xi))^2 \cdot p_i < \infty$, akkor

$$D^2(\xi) = \sum_{i=1}^{\infty} (x_i - M(\xi))^2 \cdot p_i ,$$

illetve a 2.9.4. tétel alapján

$$D^2(\xi) = \sum_{i=1}^{\infty} x_i^2 p_i - M^2(\xi).$$

Egyébként a szórás nem létezik.

3. Legyen $\xi$ egy folytonos eloszlású valószínűségi változó.

Ha $\int\limits_{-\infty}^{\infty}(x-M(\xi))^2 f(x)dx<\infty$, akkor

$$D^2(\xi)=\int\limits_{-\infty}^{\infty}(x-M(\xi))^2 f(x)dx$$

illetve a 2.9.4. tétel alapján

$$D^2(\xi)=\int\limits_{-\infty}^{\infty}x^2 f(x)dx-M^2(\xi).$$

Egyébként a szórás nem létezik.

**2.9.6. tétel.** Legyen $\xi$ egy tetszőleges valószínűségi változó és $\eta=a\cdot\xi+b$. Ekkor $\eta$ szórása

$$D(\eta)=D(a\cdot\xi+b)=|\,a\,|\cdot D(\xi),$$

amennyiben a $\xi$ szórása létezik.

**Bizonyítás**. A tétel állítása a

$$\begin{aligned}D^2(\eta)=D^2(a\cdot\xi+b)=M\left((a\cdot\xi+b)^2\right)-M^2(a\cdot\xi+b)=\\ =M\left(a^2\cdot\xi^2+2ab\cdot\xi+b^2\right)-(a\cdot M(\xi)+b)^2=\\ =a^2\cdot M\left(\xi^2\right)+2ab\cdot M(\xi)+b^2-a^2\cdot M^2(\xi)-2ab\cdot M(\xi)-b^2=\\ =a^2\cdot M\left(\xi^2\right)-a^2\cdot M^2(\xi)=a^2\cdot D^2(\xi)\end{aligned}$$

egyenlőségekből gyökvonással következik.

Korábban már láttuk, hogy a várható érték szoros kapcsolatban áll tömegeloszlások súlypontjával. Ennek alapján nem meglepő, hogy a mechanikában *Steiner-tételként* ismert összefüggés valószínűségi változókra is igaz:

**2.9.8. tétel** (Steiner-egyenlőség ill. egyenlőtlenség): Legyen $\xi$ egy tetszőleges valószínűségi változó, melynek szórása létezik. Akkor minden $c$ valós számra:

$$M((\xi-c)^2)=M((\xi-M(\xi))^2)+(M(\xi)-c)^2\geq D^2(\xi),$$

egyenlőség pedig pontosan akkor érvényes, ha $c = M(\xi)$.

**Bizonyítás:** Jelölje $m := M(\xi)$. Nyilván

$$(\xi - c)^2 = (\xi - m + m - c)^2 = (\xi - m)^2 + 2(\xi - m)(m - c) + (m - c)^2 .$$

Vegyük mindkét oldal várható értékét: az $M(\xi - m) = M(\xi) - m = 0$ egyenlőséget felhasználva, kapjuk, hogy :

$$M((\xi - c)^2) = M((\xi - m)^2) + (m - c)^2$$

A jobboldalról $(m - c)^2$-t elhagyva, az csak csökkenhet:

$$M((\xi - c)^2) = M((\xi - m)^2) + (m - c)^2 \geq M((\xi - m)^2) = D^2(\xi),$$

egyenlőség pedig pontosan akkor érvényes, ha $c = m$.

## 2.10. Nevezetes valószínűség-eloszlások

*1. Karakterisztikus eloszlás (Bernoulli-eloszlás)*

Legyen egy kísérlet valamely $A$ eseményének valószínűsége $P(A) = p$ és az $\overline{A}$ ellentett eseményé $P(\overline{A}) = 1 - p =: q$. Legyen a diszkrét $\xi$ valószínűségi változó az $A$ esemény indikátor változója. Ekkor $\xi$ az $x_k = k$ $(k = 0,1)$ értékeket a következő valószínűségekkel veszi fel:

$$P(\xi = 0) = q, \quad P(\xi = 1) = p .$$

$\xi$ eloszlását *karakterisztikus* (vagy *Bernoulli-*) *eloszlásnak* nevezzük.

$\xi$ várható értéke és szórása: $M(\xi) = p$, $D(\xi) = \sqrt{pq}$.

*2. Binomiális eloszlás*

Legyen egy kísérlet valamely $A$ eseményének valószínűsége $P(A) = p$ és az $\overline{A}$ ellentett eseményé $P(\overline{A}) = 1 - p =: q$. A kísérletet $n$-szer egymástól függetlenül megismételjük. Legyen a diszkrét $\xi$ valószínűségi változó értéke az $A$ esemény bekövetkezéseinek száma a kísérletsorozatban. Ekkor $\xi$ az $x_k = k$ $(k = 0,1,\ldots,n)$ értékeket a következő valószínűségekkel veszi fel:

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

$\xi$ eloszlását *binomiális eloszlásnak* nevezzük.

$\xi$ várható értéke és szórása: $M(\xi) = np, \quad D(\xi) = \sqrt{npq}$ .

*3. Geometriai eloszlás (Pascal-eloszlás)*

Legyen egy kísérlet valamely $A$ eseményének valószínűsége $P(A) = p$ , és folytassunk addig független kísérleteket, míg az *A* esemény be nem következik. Jelölje $\xi$ az ehhez szükséges kísérletek számát: akkor

$$P(\xi = k) = (1-p)^{k-1} p \quad (k = 1,2,\ldots,).$$

$\xi$ eloszlását *geometriai eloszlásnak* (vagy *Pascal-eloszlásnak*) nevezzük.

$\xi$ várható értéke és szórása: $M(\xi) = \frac{1}{p}, \quad D(\xi) = \frac{\sqrt{1-p}}{p}$ .

*4. Hipergeometriai eloszlás*

Legyen $m$ elemünk, amelyből $s$ darabot megkülönböztetünk a többi $m-s$ darabtól (például selejtesnek tekintjük). Ezután találomra kiválasztunk az $m$ elemből $n$ darabot visszatevés nélkül, ahol $n \le s$ és $n \le m-s$ . Legyen a diszkrét $\xi$ valószínűségi változó értéke az $n$ kiválasztott darab között lévő megkülönböztetett elemek száma. Ekkor $\xi$ az $x_k = k \ (k = 0,1,\ldots,n)$ értékeket a következő valószínűségekkel veszi fel:

$$P(\xi = k) = p_k = \frac{\binom{s}{k}\binom{m-s}{n-k}}{\binom{m}{n}} \quad (k = 0,1,\ldots,n)$$

$\xi$ eloszlását *hipergeometriai eloszlásnak* nevezzük.

$\xi$ várható értéke és szórása a $p = \dfrac{s}{m}$ jelöléssel:

$$M(\xi) = np,\ D(\xi) = \sqrt{np(1-p)\left(1 - \frac{n-1}{m-1}\right)}.$$

5. *Poisson-eloszlás*

Egy diszkrét $\xi$ valószínűségi változót $\lambda > 0$ paraméterű *Poisson-eloszlású-nak* nevezünk, ha az $x_k = k\ (k = 0,1,2,\dots)$ értékeket

$$P(\xi = k) = p_k = \frac{\lambda^k}{k!}e^{-\lambda} \quad (k = 0,1,2,\dots)$$

valószínűségekkel veheti fel.

$\xi$ várható értéke és szórása: $M(\xi) = \lambda, \quad D(\xi) = \sqrt{\lambda}$ .

A gyakorlatban, ha egy valószínűségi változó sok, kis valószínűségű esemény bekövetkezéseinek számát méri, akkor az közel Poisson-eloszlású-nak tekinthető (lásd a következő szakaszt is). Ilyen pl. egy forgalmas telefonközpontba 1 sec alatt beérkező hívások száma; ilyen pl. egy májusi erdei séta esetén a sétálóra hullott kullancsok száma stb.

5. *Folytonos egyenletes eloszlás*

Egy folytonos $\xi$ valószínűségi változót az $(a,b)$ intervallumon *egyenletes eloszlásúnak* nevezünk, ha a sűrűségfüggvénye

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

az eloszlásfüggvénye pedig

$$F(x) = P(\xi < x) = \begin{cases} 0, & \text{ha } x \le a \\ \frac{x-a}{b-a}, & \text{ha } a < x \le b, \\ 1, & \text{ha } x > b. \end{cases}$$

$\xi$ várható értéke és szórása: $M(\xi) = \dfrac{a+b}{2}$, $D(\xi) = \dfrac{b-a}{2\sqrt{3}}$.

Szokásos még diszkrét valószínűségi változók közt is bevezetni az egyenletes eloszlást. A $\xi$ diszkrét (éspedig véges sok értéket felvevő) valószínűségi változót egyenletes eloszlásúnak nevezzük, ha az $x_1, x_2, \ldots x_n$ értékeket egyaránt $\dfrac{1}{n}$ valószínűséggel veszi fel. Tipikusan ilyen egy szabályos dobókockával való dobás kimenetele.

*6. Exponenciális eloszlás*

Egy folytonos $\xi$ valószínűségi változót $\lambda > 0$ paraméterű *exponenciális eloszlásúnak* nevezünk, ha a sűrűségfüggvénye

$$f(x) = \begin{cases} 0, & \text{ha } x \le 0, \\ \lambda \cdot e^{-\lambda x}, & \text{ha } x > 0 \end{cases}$$

Az eloszlásfüggvény:

$$F(x) = P(\xi < x) = \begin{cases} 0, & \text{ha } x \le 0, \\ 1 - e^{-\lambda \cdot x}, & \text{ha } x > 0. \end{cases}$$

$\xi$ várható értéke és szórása: $M(\xi) = \dfrac{1}{\lambda}$, $D(\xi) = \dfrac{1}{\lambda}$.

*7. Normális eloszlás*

Egy folytonos $\xi$ valószínűségi változót $(m, \sigma)$-*paraméterű normális eloszlásúnak* nevezünk ($\sigma > 0$), ha a sűrűségfüggvénye

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Az eloszlásfüggvény:

$$F(x) = P(\xi < x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}}\, dt .$$

$\xi$ várható értéke és szórása: $M(\xi) = m, \quad D(\xi) = \sigma$ .

A (0,1)-paraméterű normális eloszlást *standard normális eloszlásnak* nevezzük.

**Jelölés.** Az $(m, \sigma)$ paraméterű normális eloszlást általában $N(m,\sigma)$-val, így a standard normális eloszlást $N(0,1)$-gyel jelöljük.

## 2.11. A nevezetes valószínűség-eloszlások tulajdonságai

**2.11.1. tétel.** Legyen $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ binomiális eloszlású valószínűségi változók olyan sorozata, amelynél $\lim\limits_{n\to\infty} n \cdot p_n = \lambda > 0$. Ekkor tetszőleges rögzített $k$ esetén

$$\lim_{n\to\infty} P(\xi_n = k) = \lim_{n\to\infty} \binom{n}{k} p_n^k \cdot q_n^{n-k} = \frac{\lambda^k}{k!} \cdot e^{-\lambda} \qquad (k = 0,1,2,\ldots)$$

**2.11.2. következmény.** Ha $p$ vagy $q$ elég kicsi és $n$ nagy, akkor a binomiális eloszlást Poisson-eloszlással közelíthetjük:

$$\binom{n}{k} p^k \cdot q^{n-k} \approx \frac{(n\cdot p)^k}{k!} \cdot e^{-n\cdot p}, \text{ ill. } \binom{n}{k} p^k \cdot q^{n-k} \approx \frac{(n\cdot q)^{n-k}}{(n-k)!} \cdot e^{-n\cdot q} .$$

Ha $p$, $q$ egyike sem 0, és $n$ nagy, akkor a binomiális eloszlás a normális eloszlással is közelíthető:

$$\binom{n}{k} p^k \cdot q^{n-k} \approx \frac{1}{\sqrt{npq}} \cdot \varphi\left(\frac{k-np}{\sqrt{npq}}\right), \qquad (k = 0,1,\ldots,n)$$

ahol $\varphi$ a standard normális eloszlás sűrűségfüggvényét jelöli *(Gauss-függvény):*

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

**2.11.3. tétel.** Legyen $\xi_1,\xi_2,\ldots,\xi_m,\ldots$ hipergeometriai eloszlású valószínűségi változók olyan sorozata, amelynél $n$ az $m$-től független állandó és

$$\lim_{m\to\infty}\frac{s_m}{m}=p>0\,.$$

Ekkor tetszőleges rögzített $k$ $(0\le k\le n)$esetén

$$\lim_{m\to\infty}P(\xi_m=k)=\lim_{m\to\infty}\frac{\binom{s_m}{k}\cdot\binom{m-s_m}{n-k}}{\binom{m}{n}}=$$

$$=\binom{n}{k}p^k\cdot(1-p)^{n-k}=\binom{n}{k}p^k\cdot q^{n-k}.$$

**2.11.4. következmény.** Ha $m$ és $s$ nagy a minta $n$ elemszámához képest, akkor a hipergeometriai eloszlás helyett használhatjuk a $p=\frac{s}{m}$ paraméterű binomiális eloszlást.

Következésképp, a fenti feltételek mellett lényegében mindegy, hogy visszatevéssel (binomiális eloszlás) vagy visszatevés nélkül (hipergeometriai eloszlás) vesszük a mintát, ami szemléletesen is nyilvánvaló.

**2.11.5. tétel.** Legyen $A$ egy adott esemény és jelentse $\xi_T$ az $A$ eseménynek a $[0,T]$ időintervallum alatti bekövetkezései számát. Jelöljük $P_i(\Delta t)$-vel annak a valószínűségét, hogy az $A$ esemény $\Delta t$ idő alatt $i$-szer következik be. Az $A$ eseményről feltesszük, hogy

1. Az egymást követő $[0=t_0,t_1),[t_1,t_2),\ldots,[t_{n-2},t_{n-1}),[t_{n-1},t_n=T)$ intervallumokban az $A$ esemény bekövetkezéseinek száma egymástól független.
2. Egyenlő hosszúságú intervallumokban az $A$ esemény bekövetkezéseinek száma egyenlő valószínűségű.
3. Annak valószínűsége, hogy egy $\Delta t$ intervallumban az $A$ esemény egyszer bekövetkezik, az intervallum hosszával arányos, azaz
$$\lim_{\Delta t\to 0}\frac{P_1(\Delta t)}{\Delta t}=\mu,$$
ahol $\mu>0$ egy arányossági tényező.
4. Annak valószínűsége, hogy egy $\Delta t$ intervallumban az $A$ egynél többször következik be, az intervallum hosszához képest elhanyagolható, azaz
$$\lim_{\Delta t\to 0}\frac{\sum_{i=2}^{\infty}P_i(\Delta t)}{\Delta t}=0.$$
Ekkor a $\xi_T$ egy $\lambda=\mu\cdot T$ paraméterű Poisson-eloszlású valószínűségi változó, azaz
$$P(\xi_T=k)=\frac{(\mu\cdot T)^k}{k!}\cdot e^{-\mu\cdot T}.$$

**2.11.6. tétel.** Ha egy nem negatív értékű folytonos eloszlású valószínűségi változóra vonatkozóan
$$P(\xi\geq t+\Delta t\mid \xi\geq t)=P(\xi\geq\Delta t)$$
tetszőleges $\Delta t>0$ esetén, akkor $\xi$ exponenciális eloszlású.

A 2.11.6. tétel meg is fordítható: könnyen belátható, hogy tetszőleges exponenciális eloszlású valószínűségi változóra igaz a tételbeli egyenlőtlenség (bizonyítsuk ezt be!). Ami a tétel szemléletes tartalmát illeti, a tételben szereplő $\xi$ valószínűségi változó jelentse egy berendezés élettartamát: ekkor a tételben szereplő egyenlőség azt fejezi ki, hogy annak a valószínűsége, hogy a berendezés $t$ idő eltelte után még legalább $\Delta t$ ideig működik, nem függ attól, hogy a $t$ mekkora. Ezt a tulajdonságot *örökifjú* tulajdonságnak hívjuk. Ez jellemzi pl. a radioaktív izotópok élettartamát (bomlási idejét), az olyan berendezések élettartamát, ahol a tönkremenetel a folyamatos elhasználódás eredménye (pl. izzólámpák, képcsövek élettartamát).

## 2.12. Valószínűségi változók függvényeinek (transzformáltjainak) eloszlása

**Definíció.** Legyenek a diszkrét $\xi$ valószínűségi változó lehetséges értékei $x_1, x_2, \ldots$ és ezek bekövetkezési valószínűségei $p_1, p_2, \ldots$. Az

$$\eta = h(\xi)$$

*transzformált valószínűségi változó* lehetséges értékei ekkor az

$$y_1 := h(x_1), y_2 := h(x_2), \ldots$$

számok (amelyek között megegyezők is lehetnek) és ezek bekövetkezési valószínűségei a

$$q_k = P(\eta = y_k) = \sum_{h(x_i) = y_k} P(\xi = x_i) = \sum_{h(x_i) = y_k} p_i, \quad (k = 1, 2, \ldots)$$

értékek, ahol az összegzés mindazon $i$-kre vonatkozik, amelyre $h(x_i) = y_k$ fennáll.

Ha az $\eta = h(\xi)$ valószínűségi változó a $\xi$-nek *szigorúan monoton* függvénye, akkor az $\eta$ valószínűségi változó $y_k = h(x_k)$ $(k = 1, 2, \ldots)$ értékeihez tartozó valószínűségeloszlás megegyezik a $\xi$ valószínűségi változó eloszlásával.

**2.12.1. tétel.** (Valószínűségi változó transzformációja.) Legyen $h$ egy szigorúan monoton, differenciálható függvény és a $\xi$ egy folytonos eloszlású valószínűségi változó, amelynek a sűrűségfüggvénye $f$. Ekkor az $\eta = h(\xi)$ transzformált valószínűségi változó sűrűségfüggvénye

$$g(y) = f\left(h^{-1}(y)\right) \cdot \left|\left(h^{-1}\right)'(y)\right| \quad (y \in D_{h^{-1}}),$$

ahol $h^{-1}$ a $h$ függvény inverz függvényét jelöli.

**Bizonyítás.**
*a)* Legyen először $h$ szigorúan monoton növekedő. Ekkor az $\{\eta < y\}$, vagyis a $\{h(\xi) < y\}$ esemény ekvivalens a $\{\xi < h^{-1}(y)\}$ eseménnyel. Így $\eta$ eloszlásfüggvénye:

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

ahol $F$ a $\xi$ eloszlásfüggvénye. Innen az összetett függvény deriváltjára vonatkozó láncszabályt alkalmazva

$$g(y) = G'(y) = \left(F \circ h^{-1}\right)'(y) = f\left(h^{-1}(y)\right) \cdot \left(h^{-1}\right)'(y),$$

ahol a monoton növekedés miatt $\left(h^{-1}\right)'(y) > 0$.

*b)* Ha $h$ szigorúan monoton csökkenő, akkor az $\{\eta < y\}$, vagyis a $\{h(\xi) < y\}$ esemény a $\{\xi > h^{-1}(y)\}$ eseménnyel ekvivalens. $\eta$ eloszlásfüggvénye így:

$$G(y) = P(\eta < y) = P(h(\xi) < y) = P\left(\xi > h^{-1}(y)\right) = 1 - F\left(h^{-1}(y)\right),$$

ahol $F$ a $\xi$ eloszlásfüggvénye. Innen szintén láncszabályt alkalmazva

$$g(y) = G'(y) = -\left(F \circ h^{-1}\right)'(y) = -f\left(h^{-1}(y)\right) \cdot \left(h^{-1}\right)'(y),$$

ahol a monoton csökkenés miatt $\left(h^{-1}\right)'(y) < 0$, a tétel állításával összhangban.

**2.12.2. tétel.** (Speciális változótranszformációk.)

Legyen $\xi$ egy folytonos eloszlású $f$ sűrűségfüggvényű valószínűségi változó. Ekkor:

1. Az $\eta = a\xi + b$ $(a \neq 0)$ valószínűségi változó sűrűségfüggvénye

$$g(y) = \frac{1}{|a|} f\left(\frac{y-b}{a}\right), \quad y \in \mathbb{R}.$$

2. Az $\eta = \xi^2$ valószínűségi változó sűrűségfüggvénye

$$g(y) = \begin{cases} \dfrac{f\left(\sqrt{y}\right) + f\left(-\sqrt{y}\right)}{2\sqrt{y}}, & \text{ha} \quad y > 0, \\ 0, & \text{ha} \quad y \le 0. \end{cases}$$

3. Ha a $\xi$ csak nem negatív értékeket vesz fel, akkor az $\eta = \sqrt{\xi}$ valószínűségi változó sűrűségfüggvénye

$$g(y) = \begin{cases} f\left(y^2\right) \cdot 2y, & \text{ha} \quad y > 0, \\ 0, & \text{ha} \quad y \le 0. \end{cases}$$

4. Az $\eta = |\xi|$ valószínűségi változó sűrűségfüggvénye

$$g(y) = \begin{cases} f(y) + f(-y), & \text{ha} \quad y > 0, \\ 0, & \text{ha} \quad y \le 0. \end{cases}$$

**2.12.3. tétel.** Az $(m,\ \sigma)$ paraméterű $\xi$ normális eloszlású valószínűségi változó $\eta = \dfrac{\xi - m}{\sigma}$ transzformáltja standard normális eloszlású valószínűségi változó.

**Bizonyítás.** A $\xi$ valószínűségi változó sűrűségfüggvénye

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Mivel $\eta = \dfrac{\xi - m}{\sigma} = \dfrac{1}{\sigma}\xi - \dfrac{m}{\sigma}$ az előző tétel első állítását $a = \dfrac{1}{\sigma}$ és $b = -\dfrac{m}{\sigma}$ választással alkalmazva, $\eta$ sűrűségfüggvénye

$$g(y) = \frac{1}{|a|} f\left(\frac{y - b}{a}\right) = \sigma \cdot f(\sigma \cdot y + m)$$

$$= \sigma \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\sigma \cdot y + m - m)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}},$$

amit igazolni akartunk.

## 2.13. A Markov- és a Csebisev-egyenlőtlenség

**2.13.1. tétel.** (Markov-egyenőtlenség) Ha $\eta$ olyan nem negatív értékeket felvevő valószínűségi változó, amelynek van várható értéke, és az $a$ egy tetszőleges pozitív valós szám, akkor

$$P(\eta \geq a) \leq \frac{M(\eta)}{a}.$$

**Bizonyítás:**

*a)* Legyen $\eta$ először diszkrét valószínűségi változó. Lehetséges értékei legyenek $y_1, y_2, \ldots$, az ezekhez tartozó bekövetkezési valószínűségek pedig $p_1, p_2, \ldots$ . Ha felírjuk az $\eta$ várható értékét, majd az ezt alkotó összeg néhány tagját elhagyjuk, azután a tagokat (legfeljebb egy kivételével) alkalmasan kisebbítjük, az

$$M(\eta) = \sum_{k=1}^{\infty} y_k \cdot p_k \geq \sum_{y_k \geq a} y_k \cdot p_k \geq \sum_{y_k \geq a} a \cdot p_k \geq a \sum_{y_k \geq a} p_k = a \cdot P(\eta \geq a)$$

egyenlőtlenségsorozatot nyerjük, ahonnan

$$M(\eta) \geq a \cdot P(\eta \geq a).$$

A tétel állítását innen mindkét oldalt $a$-val osztva kapjuk.

*b)* Legyen most $\eta$ $f$ sűrűségfüggvényű folytonos eloszlású valószínűségi változó. A bizonyítandó egyenlőtlenség ekkor az

$$M(\eta) = \int_0^{\infty} y \cdot f(y)dy \geq \int_a^{\infty} y \cdot f(y)dy \geq$$

$$\geq \int_a^{\infty} a \cdot f(y)dy = a \cdot \int_a^{\infty} f(y)dy = a \cdot P(\eta \geq a)$$

egyenlőtlenségekből következik.

**2.13.2. tétel.** (Csebisev-egyenlőtlenség) Ha a $\xi$ valószínűségi változónak van várható értéke és szórása, és $\lambda$ tetszőleges pozitív valós szám, akkor

$$P\left(|\xi - M(\xi)| \geq \lambda \cdot D(\xi)\right) \leq \frac{1}{\lambda^2}.$$

**Bizonyítás.** A Markov-egyenlőtlenséget az $\eta := (\xi - M(\xi))^2$ és az $a := \lambda^2 \cdot D^2(\xi)$ választással alkalmazva, a

$$P\left((\xi - M(\xi))^2 \geq \lambda^2 \cdot D^2(\xi)\right) \leq \frac{M\left((\xi - M(\xi))^2\right)}{\lambda^2 \cdot D^2(\xi)} = \frac{D^2(\xi)}{\lambda^2 \cdot D^2(\xi)} = \frac{1}{\lambda^2}$$

egyenlőtlenséget kapjuk. A tétel állítása ebből a

$$\left\{(\xi - M(\xi))^2 \geq \lambda^2 \cdot D^2(\xi)\right\}$$

eseménynek a vele ekvivalens

$$\left\{|\xi - M(\xi)| \geq \lambda \cdot D(\xi)\right\}$$

eseménnyel történő helyettesítésével következik.

A Csebisev-egyenlőtlenséget a

$$P\left(|\xi - M(\xi)| < \lambda \cdot D(\xi)\right) = 1 - P\left(|\xi - M(\xi)| \geq \lambda \cdot D(\xi)\right) \geq 1 - \frac{1}{\lambda^2}$$

formában annak becslésére használjuk, hogy a $\xi$ valószínűségi változó milyen valószínűséggel esik egy adott, a várható érték körüli szimmetrikus intervallumba.

**1. példa.** (Diszkrét valószínűségi változó.) Egy sorsjátékon 1 darab 50000 Ft-os, 10 darab 5000 Ft-os és 50 darab 1000 Ft-os nyeremény van. A játékhoz 1000 darab sorsjegyet adnak ki. A $\xi$ valószínűségi változó vegye fel az egyes nyereményösszegek értékét.

*a)* Adjuk meg $\xi$ eloszlását!

*b)* Határozzuk meg $\xi$ várható értékét!

*c)* Határozzuk meg $\xi$ szórását!

**Megoldás.**

*a)* A $\xi$ valószínűségi változó értékei az

$$x_0 = 0,\ \ x_1 = 1000,\ \ x_2 = 5000,\ \ x_3 = 50000$$

számok lesznek. Ezek bekövetkezési valószínűségei:

$$p_0 = \frac{1000-61}{1000},\ p_1 = \frac{50}{1000},\ p_2 = \frac{10}{1000},\ p_3 = \frac{1}{1000}.$$

*b)*

$$M(\xi) = \sum_{i=0}^{3} x_i \cdot p_i =$$

$$= 0 \cdot \frac{1000-61}{1000} + 1000 \cdot \frac{50}{1000} + 5000 \cdot \frac{10}{1000} + 50000 \cdot \frac{1}{1000} = 150$$

*c)*

$$D(\xi) = \left(M\left(\xi^2\right) - M(\xi)^2\right)^{1/2} = \left(\sum_{i=0}^{3} x_i^2 \cdot p_i - M(\xi)^2\right)^{1/2} =$$

$$= \begin{pmatrix} 0^2 \cdot \frac{1000-61}{1000} + 1000^2 \cdot \frac{50}{1000} \\ +5000^2 \cdot \frac{10}{1000} + 50000^2 \cdot \frac{1}{1000} \\ -(150)^2 \end{pmatrix}^{1/2} =$$

$$= (0 + 50000 + 250000 + 2500000 - 22500)^{1/2} \approx 1666{,}6.$$

**2. példa.** (Folytonos valószínűségi változó.) Egy $r = 18$ cm sugarú céltáblára véletlenszerűen lövéseket adunk le. Feltesszük, hogy minden lövés eltalálja a céltáblát (csak ilyen lövéseket veszünk figyelembe). Tegyük fel, hogy a céltáblára rajzolt bármely síkidom találati valószínűsége arányos a síkidom területével. A $\xi$ valószínűségi változó értéke legyen a találat helyének a kör középpontjától mért távolsága. Számítsuk ki $\xi$ várható értékét és szórásnégyzetét!

**Megoldás.** A folytonos eloszlású $\xi$ valószínűségi változó eloszlásfüggvényét, illetve sűrűségfüggvényét kell először meghatározni. Annak valószínűsége, hogy a $\xi$ valószínűségi változó $x$-nél $(0 < x \le r)$ kisebb értéket vesz fel az $x$ sugarú és az r sugarú körlapok területeinek hányadosával arányos:

$$P(\xi < x) = \frac{x^2 \cdot \pi}{r^2 \cdot \pi} = \frac{x^2}{r^2}.$$

Ennek alapján az eloszlásfüggvény:

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

Ebből a $\xi$ sűrűségfüggvénye szakaszonkénti deriválással adódik:

$$f(x) = \begin{cases} 0, & \text{ha} \quad x \le 0, \\ \dfrac{2 \cdot x}{r^2}, & \text{ha} \quad 0 < x \le r, \\ 0, & \text{ha} \quad r < x. \end{cases}$$

A $\xi$ folytonos eloszlású valószínűségi változó várható értéke így:

$$M(\xi) = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_0^r x \cdot \frac{2 \cdot x}{r^2} dx = \frac{2}{r^2} \int_0^r x^2 dx = \frac{2}{r^2} \left[ \frac{x^3}{3} \right]_0^r$$

$$= \frac{2}{3} \cdot r = \frac{2}{3} \cdot 18 = 12.$$

A szórásnégyzete pedig:

$$D^2(\xi) = M(\xi^2) - M^2(\xi) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx - \left( \int_{-\infty}^{\infty} x \cdot f(x)dx \right)^2 =$$

$$= \int_0^r x^2 \cdot \frac{2 \cdot x}{r^2} dx - \left( \frac{2}{3} r \right)^2 = \frac{2}{r^2} \int_0^r x^3 dx - \left( \frac{2}{3} r \right)^2 =$$

$$= \frac{2}{r^2} \left[ \frac{x^4}{4} \right]_0^r - \frac{4}{9} r^2 = \frac{1}{2} r^2 - \frac{4}{9} r^2 = \frac{1}{18} r^2 = 18.$$

**3. példa.** Határozzuk meg az $n$ és $p$ paraméterű binomiális eloszlású $\xi$ valószínűségi változó várható értékét !

**Megoldás.** $\xi$ várható értéke:

$$M(\xi)=\sum_{k=0}^{n} x_k \cdot p_k = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k \cdot q^{n-k}$$

$$= \sum_{k=0}^{n} k \cdot \frac{n!}{k! \cdot (n-k)!} p^k \cdot q^{n-k}$$

$$= n \cdot p \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)! \cdot ((n-1)-(k-1))!} p^{k-1} \cdot q^{(n-1)-(k-1)}$$

$$= n \cdot p \sum_{i=0}^{n-1} \binom{n-1}{i} p^i \cdot q^{(n-1)-i}$$

$$= n \cdot p \cdot (p+q)^{n-1} = n \cdot p \cdot (p+(1-p))^{n-1} = n \cdot p.$$

Másik – és egyszerűbb – megoldáshoz jutunk, ha észrevesszük, hogy $\xi$ előáll $n$ db független, $p$ paraméterű karakterisztikus eloszlású valószínűségi változó összegeként: $\xi = \eta_1 + \eta_2 + \ldots + \eta_n$. Mindegyik $\eta_i$ várható értéke $1 \cdot p + 0 \cdot (1-p) = p$, innen $M(\xi) = M(\eta_1) + \ldots + M(\eta_n) = np$.

**4. példa.** Határozzuk meg a $\lambda$ paraméterű Poisson-eloszlású $\xi$ valószínűségi változó várható értékét és szórásnégyzetét!

**Megoldás.** $\xi$ várható értéke:

$$M(\xi)=\sum_{k=0}^{\infty} x_k \cdot p_k = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} = \lambda \cdot e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$

$$= \lambda \cdot e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = \lambda \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda.$$

$\xi^2$ várható értéke:

$$M\left(\xi^2\right)=\sum_{k=0}^{\infty} x_k^2 \cdot p_k = \sum_{k=0}^{\infty} k^2 \cdot \frac{\lambda^k}{k!}\cdot e^{-\lambda}$$

$$=\sum_{k=2}^{\infty} k\cdot(k-1)\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda} + \sum_{k=1}^{\infty} k\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}$$

$$=\lambda^2\cdot e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}+\lambda\cdot e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}$$

$$=\lambda^2\cdot e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!}+\lambda\cdot e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!}=\lambda^2+\lambda.$$

$\xi$ szórásnégyzete:

$$D^2(\xi)=M\left(\xi^2\right)-M^2(\xi)=\left(\lambda^2+\lambda\right)-(\lambda)^2=\lambda.$$

**5. példa.** (Binomiális eloszlás.) Egy tétel áru harmadrésze első osztályú. Négy darabot kiválasztunk a tételből találomra. A kiválasztás egyenként megy végbe, és a kiválasztott árut rögtön - a következő kiválasztása előtt - visszatesszük a többi közé. A $\xi$ valószínűségi változó értéke legyen a kiválasztott első osztályú darabok száma. Írjuk fel és ábrázoljuk a valószínűségi változó eloszlását! Számítsuk ki a várható értékét és a szórását!

**Megoldás.** A $\xi$ valószínűségi változó binomiális eloszlású. Jelöljük $A$-val azt az eseményt, hogy egy kiválasztott darab első osztályú. Az $A$ és az $\overline{A}$ események valószínűségei:

$$P(A)=p=\frac{1}{3},\quad P\left(\overline{A}\right)=q=1-p=1-\frac{1}{3}=\frac{2}{3}.$$

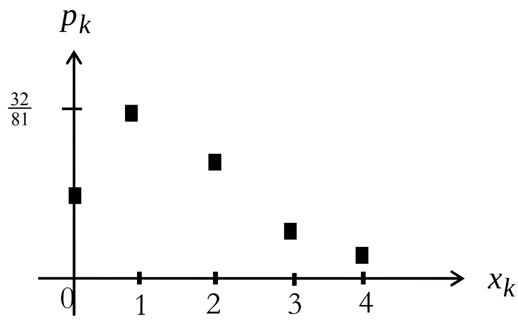
Annak valószínűsége, hogy $\xi$ az $x_k := k$ értéket veszi fel:

$$p_k = P(\xi=k)=\binom{4}{k}\cdot\left(\frac{1}{3}\right)^k\cdot\left(\frac{2}{3}\right)^{4-k}=\frac{4!}{k!\cdot(4-k)!}\cdot\frac{2^{4-k}}{3^4}\quad (k=0,1,2,3,4)$$

Az eloszlás tagjait úgy kapjuk, hogy $k$ lehetséges értékeit behelyettesítjük:

$$p_0=\frac{16}{81},\quad p_1=\frac{32}{81},\quad p_2=\frac{24}{81},\quad p_3=\frac{8}{81},\quad p_4=\frac{1}{81}.$$

Ábrázoljuk az eloszlást:




ξ várható értéke

$$M(\xi) = n \cdot p = 4 \cdot \frac{1}{3} = \frac{4}{3},$$

szórása pedig

$$D(\xi) = \sqrt{n \cdot p \cdot q} = \sqrt{4 \cdot \frac{1}{3} \cdot \frac{2}{3}} = \frac{2}{3} \cdot \sqrt{2}.$$

Az MS Excel használata esetén a $p_0, p_1, \ldots, p_4$ valószínűségeket az

= BINOM.ELOSZLÁS(0;4;1/3;HAMIS),

= BINOM.ELOSZLÁS(1;4;1/3;HAMIS),

…

= BINOM.ELOSZLÁS(4;4;1/3;HAMIS)

kifejezések kiértékelésével határozhatjuk meg.

**6. példa.** (Hipergeometriai eloszlás.) Próbagyártás során 20 gép készült el, amely közül 5 javításra szorul. A teljes mennyiségből 4 találomra kiválasztott gépet felülvizsgálatra küldenek. A gyártás akkor indulhat meg, ha a felülvizsgált gépek közül legfeljebb 1 szorul javításra. Mekkora annak a valószínűsége, hogy megindulhat a gyártás?

**Megoldás.** Jelölje $x_k := k$ $(k = 0,1,2,3,4)$ a felülvizsgálatra küldött gépek között levő hibásak számát. A ξ hipergeometriai eloszlású valószínűségi változó ezeket a $k$ értékeket veheti fel. A teljes mennyiség $m = 20$, a javításra szorulók száma ebből $s = 5$. A felülvizsgálatra küldött gépek száma $n = 4$. Így annak valószínűsége, hogy a ξ valószínűségi változó a $k$ értéket veszi fel:

$$p_k = P(\xi = x_k) = \frac{\binom{5}{k}\binom{15}{4-k}}{\binom{20}{4}} \qquad (k = 0,1,2,3,4)$$

A gyártás akkor indulhat meg, ha legfeljebb egy hibás gép akad a felülvizsgálatra küldött gépek között. Ez két módon állhat elő: ha mind a négy gép hibátlan, vagy pontosan egy hibás van közöttük. Ezek az esetek egymást kizárják, ennélfogva valószínűségeik összege adja a vizsgált esemény valószínűségét:

$$p_0 + p_1 = \frac{\binom{5}{0}\binom{15}{4}}{\binom{20}{4}} + \frac{\binom{5}{1}\binom{15}{3}}{\binom{20}{4}} = \frac{15\cdot 14\cdot 13\cdot 12}{20\cdot 19\cdot 18\cdot 17} + \frac{15\cdot 14\cdot 13}{19\cdot 18\cdot 17}$$

$$= \frac{728}{969} \approx 0{,}75.$$

Tehát kb. 75% annak a valószínűsége, hogy a felülvizsgálat eredménye alapján megindulhat a gyártás.
Az MS Excel használata esetén a $p_0 + p_1$ valószínűséget az

= HIPERGEOM.ELOSZLÁS(0;4;5;20)
+ HIPERGEOM.ELOSZLÁS(1;4;5;20)

kifejezés kiértékelésével határozhatjuk meg.

**7. példa.** (Poisson-eloszlás.) Egy orsózógépen 100 munkaóra alatt átlagban 3 szakadás következik be. Mennyi annak a valószínűsége, hogy egy ilyen időtartam alatt a szakadások száma nem lépi túl az átlagot? Az általános tapasztalat alapján feltehető, hogy a szakadások Poisson-eloszlás szerint következnek be.

**Megoldás.** A $\xi$ valószínűségi változó vegye fel a vizsgált időtartam alatt végbemenő szakadások számát: akkor $\xi$ Poisson-eloszlású. A paramétere a vizsgált időtartam alatti szakadások száma, vagyis $\lambda = M(\xi) = 3$. A $\xi$ valószínűségi változó a $k$ értékeket

$$p_k = P(\xi = k) = \frac{3^k}{k!} \cdot e^{-3} \qquad (k = 0, 1, 2, \ldots)$$

valószínűségekkel veszi fel. Annak valószínűsége, hogy $\xi$ értéke nem lépi túl az $M(\xi) = 3$ várható értéket

$$P(\xi \le 3) = \sum_{k=0}^{3} P(\xi = k) = \sum_{k=0}^{3} p_k = \sum_{k=0}^{3} \frac{3^k}{k!} \cdot e^{-3}$$

$$= e^{-3} \cdot \left(1 + 3 + \frac{3^2}{2!} + \frac{3^3}{3!}\right) = e^{-3} \cdot \left(4 + \frac{9}{2} + \frac{9}{2}\right) = 13 \cdot e^{-3} \approx 0{,}65.$$

Tehát kb. 65% valószínűsége annak, hogy a vizsgált 100 órás időtartam alatt nem következik be szakadás.
Az MS Excel használata esetén a $p_0 + p_1 + p_2 + p_3$ valószínűséget az

$$= \mathrm{POISSON}(0;3;\mathrm{HAMIS}) + \mathrm{POISSON}(1;3;\mathrm{HAMIS})$$
$$+ \mathrm{POISSON}(2;3;\mathrm{HAMIS}) + \mathrm{POISSON}(3;3;\mathrm{HAMIS})$$

vagy az

$$= \mathrm{POISSON}(3;3;\mathrm{IGAZ})$$

kifejezés kiértékelésével határozhatjuk meg.

**8. példa.** Határozzuk meg az $(a, b)$ intervallumon egyenletes eloszlású $\xi$ valószínűségi változó várható értékét és szórásnégyzetét!

**Megoldás.** $\xi$ várható értéke:

$$M(\xi) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \int_a^b x \cdot \frac{1}{b-a}\,dx = \frac{1}{b-a} \cdot \left[\frac{x^2}{2}\right]_a^b =$$

$$= \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2} = \frac{a+b}{2}.$$

$\xi^2$ várható értéke:

$$M\left(\xi^2\right)=\int_{-\infty}^{\infty} x^2\cdot f(x)dx=\int_a^b x^2\cdot\frac{1}{b-a}dx=\frac{1}{b-a}\cdot\left[\frac{x^3}{3}\right]_a^b=$$

$$=\frac{1}{b-a}\cdot\frac{b^3-a^3}{3}=\frac{a^2+a\cdot b+b^2}{3}.$$

$\xi$ szórásnégyzete:

$$D^2(\xi)=M\left(\xi^2\right)-M^2(\xi)=\frac{a^2+a\cdot b+b^2}{3}-\left(\frac{a+b}{2}\right)^2=$$

$$=\frac{4b^2+4a\cdot b+4b^2-3b^2-6a\cdot b-3a^2}{3\cdot 4}=\frac{(b-a)^2}{3\cdot 4}.$$

**9. példa.** Határozzuk meg a $\lambda$ paraméterű exponenciális eloszlású $\xi$ valószínűségi változó várható értékét és szórásnégyzetét!

**Megoldás.** $\xi$ várható értéke (parciálisan integrálva):

$$M(\xi)=\int_{-\infty}^{\infty} x\cdot f(x)dx=\int_0^{\infty} x\cdot\left(\lambda\cdot e^{-\lambda\cdot x}\right)dx$$

$$=\lim_{a\to\infty}\int_0^a x\cdot\left(\lambda\cdot e^{-\lambda\cdot x}\right)dx=\lim_{a\to\infty}\left(\left[-x\cdot e^{-\lambda\cdot x}\right]_0^a-\int_0^a 1\cdot\left(-e^{-\lambda\cdot x}\right)dx\right)$$

$$=\lim_{a\to\infty}\left(\left[-x\cdot e^{-\lambda\cdot x}\right]_0^a+\left[-\frac{e^{-\lambda\cdot x}}{\lambda}\right]_0^a\right)$$

$$=\lim_{a\to\infty}\left(-\frac{a}{e^{\lambda\cdot a}}+0-\frac{1}{\lambda}\cdot\frac{1}{e^{\lambda\cdot a}}+\frac{1}{\lambda}\right)=\frac{1}{\lambda}.$$

$\xi^2$ várható értéke (parciálisan integrálva):

$$M\left(\xi^2\right)=\int_{-\infty}^{\infty} x^2 \cdot f(x)\,dx=\int_0^{\infty} x^2 \cdot \left(\lambda \cdot e^{-\lambda \cdot x}\right)dx$$

$$=\lim_{a\to\infty}\int_0^a x^2 \cdot \left(\lambda \cdot e^{-\lambda \cdot x}\right)dx=\lim_{a\to\infty}\left(\left[-x^2 \cdot e^{-\lambda \cdot x}\right]_0^a-\int_0^a 2x \cdot \left(-e^{-\lambda \cdot x}\right)dx\right)$$

$$=\lim_{a\to\infty}\left(\left[-x^2 \cdot e^{-\lambda \cdot x}\right]_0^a+2 \cdot \int_0^a x \cdot e^{-\lambda \cdot x}dx\right)$$

$$\lim_{a\to\infty}\left(-\frac{a^2}{e^{\lambda \cdot a}}+0\right)+\frac{2}{\lambda}\cdot\lim_{a\to\infty}\int_0^a x \cdot \left(\lambda \cdot e^{-\lambda \cdot x}\right)dx=0+\frac{2}{\lambda}\cdot M(\xi)=\frac{2}{\lambda^2}.$$

.

$\xi$ szórásnégyzete:

$$D^2(\xi)=M\left(\xi^2\right)-M^2(\xi)=\frac{2}{\lambda^2}-\frac{1}{\lambda^2}=\frac{1}{\lambda^2}.$$

**10. példa.** (Egyenletes eloszlás.) Telefonhívás alkalmával a tárcsázás befejezésétől a kapcsolásig eltelt időt tekintsük egy $\xi$ valószínűségi változónak. Tegyük fel, hogy ez egyenletes eloszlású, és a kapcsolás időtartama 5 mp-től 105 mp-ig terjedhet. Adjuk meg $\xi$ sűrűség- és eloszlásfüggvényét! Határozzuk meg $\xi$ várható értékét és szórását! Számítsuk ki annak valószínűségét, hogy legalább egy percig kell várnunk a kapcsolásra!

**Megoldás.** $\xi$ egyenletes eloszlású az $(5,105)$ intervallumban. Sűrűségfüggvénye:

$$f(x)=\begin{cases} 0, & \text{ha} \quad x \le 5, \\ \dfrac{1}{100}, & \text{ha} \quad 5 < x \le 105, \\ 0, & \text{ha} \quad x > 105. \end{cases}$$

Eloszlásfüggvénye:

$$F(x)=\begin{cases} 0, & \text{ha} \quad x \le 5, \\ \dfrac{x-5}{100}, & \text{ha} \quad 5 < x \le 105, \\ 1, & \text{ha} \quad x > 105. \end{cases}$$

$\xi$ várható értéke és szórása:

$$M(\xi)=\frac{5+105}{2}=55,\quad D(\xi)=\frac{105-5}{2\sqrt{3}}=\frac{50}{\sqrt{3}}\approx 28{,}8.$$

Annak valószínűsége, hogy legalább egy percig kell várnunk a kapcsolásra:

$$P(\xi\geq 60)=1-P(\xi<60)=1-F(60)=1-\frac{60-5}{100}=0{,}45.$$

**11. példa.** (Exponenciális eloszlás.) Bizonyos típusú izzólámpák tönkremeneteléig eltelt égési időtartam hosszát tekintsük egy $\xi$ valószínűségi változónak. Megállapították, hogy $\xi$ exponenciális eloszlású és a szórása 1000 óra. Határozzuk meg $\xi$ várható értékét! Írjuk fel $\xi$ sűrűség- és eloszlásfüggvényét! Számítsuk ki annak valószínűségét, hogy egy kiszemelt izzólámpa 3000 órán belül nem megy tönkre!

**Megoldás.** Mivel $\xi$ szórása és várható értéke megegyezik:

$$D(\xi)=M(\xi)=\frac{1}{\lambda}=1000.$$

$\xi$ sűrűségfüggvénye:

$$f(x)=\begin{cases}0, & \text{ha}\quad x\leq 0,\\ \dfrac{1}{1000}\cdot e^{-\frac{x}{1000}}, & \text{ha}\quad x>0.\end{cases}$$

Eloszlásfüggvénye:

$$F(x)=\begin{cases}0, & \text{ha}\quad x\leq 0,\\ 1-e^{-\frac{x}{1000}}, & \text{ha}\quad x>0.\end{cases}$$

Annak valószínűsége, hogy egy kiszemelt izzólámpa 3000 órán belül nem megy tönkre:

$$P(\xi\geq 3000)=1-P(\xi<3000)=1-F(3000)$$
$$=1-\left(1-e^{-\frac{3000}{1000}}\right)=e^{-3}\approx 0{,}05.$$

Az MS Excel használata esetén az $F(3000)$ értékét a

$$=\text{EXP.ELOSZLÁS}(3000;1/1000;\text{IGAZ})$$

kifejezés kiértékelésével határozhatjuk meg.

Az exponenciális eloszlás „örökifjú” tulajdonsága miatt (lásd a 2.11.6. tételt és az utána tett megjegyzést), feltéve, hogy a lámpa számunkra ismeretlen ideig már világított, annak valószínűsége, hogy további 3000 órán belül nem megy tönkre, ugyanez.

**12. példa.** (Normális eloszlás.) Távolságmérést végeznek terepen. A valódi és a mért hosszúság különbségét, vagyis a mérési hibát valószínűségi változónak tekintjük. Ez a $\xi$ valószínűségi változó normális eloszlású, várható értéke $(-20)$ m, szórása 40 m. Számítsuk ki annak valószínűségét, hogy a hiba abszolút értéke 60 m-nél kevesebb!

**Megoldás.** $\xi$ ekkor $m=-20$ várható értékű és $\sigma=40$ szórású normális eloszlású valószínűségi változó, és annak valószínűsége, hogy a hiba abszolút értéke kisebb 60 m-nél:

$$P(|\xi|<60)=P(-60<\xi<60)=P\left(\frac{-60-m}{\sigma}<\frac{\xi-m}{\sigma}<\frac{60-m}{\sigma}\right)=$$
$$=P\left(\frac{-60+20}{40}<\eta<\frac{60+20}{40}\right)=P(-1<\eta<2)=$$
$$=\Phi(2)-\Phi(-1)=\Phi(2)-(1-\Phi(1))=\Phi(2)+\Phi(1)-1\approx$$
$$\approx 0{,}977+0{,}841-1=0{,}818.$$

A fenti levezetésben a standard normális eloszlásra vezető

$$\eta=\frac{\xi-m}{\sigma}$$

transzformációt alkalmaztuk. Felhasználtuk továbbá, hogy a standard normális eloszlás sűrűségfüggvénye szimmetrikus, azaz $\varphi(-x)=\varphi(x)$, és így az eloszlásfüggvényére fennáll a $\Phi(-x)=1-\Phi(x)$ azonosság.

Az MS Excel használata esetén a $\Phi(1)$ és a $\Phi(2)$ értékeket az

$$=\mathrm{STNORMELOSZL}(1) \text{ és } =\mathrm{STNORMELOSZL}(2)$$

kifejezések kiértékelésével határozhatjuk meg.

**13. példa.** (Valószínűségi változó transzformációja.) Legyen $\xi$ egy $(m, \sigma)$ araméterű normális eloszlású valószínűségi változó. Határozzuk meg az $\eta = e^{\xi}$ *lognormális eloszlású* valószínűségi változó sűrűségfüggvényét!

**Megoldás.** $\xi$ sűrűségfüggvénye

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}} \qquad (x \in I\!R)$$

Mivel a $\xi$-ről $\eta$-ra történő transzformáció alakja:

$$h(x) = e^x, \; h^{-1}(y) = \ln(y) \text{ (ezért } \left(h^{-1}\right)'(y) = \frac{1}{y}\text{)},$$

a transzformációs tételt alkalmazva, $\eta$ sűrűségfüggvénye:

$$g(y) = \begin{cases} f\left(h^{-1}(y)\right) \cdot \left|\left(h^{-1}\right)'(y)\right|, & \text{ha} \quad y > 0 \\ 0, & \text{ha} \quad y \le 0 \end{cases} =$$

$$= \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(\ln(y)-m)^2}{2\sigma^2}} \cdot \dfrac{1}{y}, & \text{ha} \quad y > 0 \\ 0, & \text{ha} \quad y \le 0 \end{cases}$$

**14. példa.** (Speciális változótranszformációk.) Legyen $\xi$ egyenletes eloszlású a $(-1, 2)$ intervallumon. Írjuk fel az $\eta := \xi^2$ valószínűségi változó sűrűség- és eloszlásfüggvényét!

**Megoldás.** ξ sűrűségfüggvénye:

$$f(x)=\begin{cases} 0, & \text{ha} \quad x<-1, \\ \frac{1}{3}, & \text{ha} \quad -1\le x<2, \\ 0, & \text{ha} \quad 2\le x. \end{cases}$$

A második speciális esetet alkalmazva, η sűrűségfüggvénye:

$$g(y)=\begin{cases} \dfrac{f\left(\sqrt{y}\right)+f\left(-\sqrt{y}\right)}{2\cdot\sqrt{y}}, & \text{ha} \quad y>0 \\ 0, & \text{ha} \quad y\le 0 \end{cases}$$

$$=\begin{cases} 0, & \text{ha} \quad 4\le y, \\ \dfrac{1}{3}\cdot\dfrac{1}{2\cdot\sqrt{y}}, & \text{ha} \quad 1<y<4, \\ \dfrac{2}{3}\cdot\dfrac{1}{2\cdot\sqrt{y}}, & \text{ha} \quad 0<y\le 1, \\ 0, & \text{ha} \quad y\le 0. \end{cases}$$

Eloszlásfüggvénye pedig:

$$G(y)=\begin{cases} 1, & \text{ha} \quad 4\le y, \\ \dfrac{1}{3}\cdot\sqrt{y}+\dfrac{1}{3}, & \text{ha} \quad 1<y<4, \\ \dfrac{2}{3}\cdot\sqrt{y}, & \text{ha} \quad 0<y\le 1, \\ 0, & \text{ha} \quad y\le 0. \end{cases}$$

ahol felhasználtuk, hogy $G'=g$ és $G(a)\le G(b)$, ha $a<b$.

**15. példa.** (Csebisev-egyenlőtlenség) Egy forgalmas útkereszteződésben egy óra alatt áthaladó gépkocsik száma legyen egy ξ valószínűségi változó. A felmérésekből ismert, hogy ξ várható értéke 500, szórása 25. Legalább mekkora valószínűséggel esik 400 és 600 közé az útkereszteződésben egy óra alatt áthaladó gépkocsik száma?

**Megoldás.** Az, hogy az áthaladó járművek száma 400 és 600 közé esik, azt jelenti, hogy a $\xi$ valószínűségi változó a várható értékétől $\lambda \cdot D(\xi) = 100$-nál nagyobb értékkel nem tér el. Mivel $D(\xi) = 25$, így $\lambda = 4$. Az ellentett események valószínűségei közötti összefüggés és Csebisev-egyenlőtlenség alapján ezért a keresett valószínűség

$$P(|\xi - 500| < 100) = 1 - P(|\xi - 500| \geq 100) \geq 1 - \left(\frac{1}{4}\right)^2 = \frac{15}{16}.$$

## 2.14. Több valószínűségi változó együttes eloszlása

**Definíció.** A $\xi_1, \xi_2, \ldots, \xi_n$ együtt megfigyelt valószínűségi változók összességét $n$-dimenziós *valószínűségi vektorváltozónak* nevezzük, és a $(\xi_1, \xi_2, \ldots, \xi_n)$ szimbólummal jelöljük.

**Definíció.** Egy $(\xi_1, \xi_2, \ldots, \xi_n)$ valószínűségi vektorváltozót *diszkrétnek (folytonosnak)* nevezünk, ha valamennyi komponense diszkrét (folytonos).

**Definíció.** Egy $(\xi_1, \xi_2, \ldots, \xi_n)$ valószínűségi vektorváltozó *eloszlásfüggvényének* nevezzük azt a függvényt, amely minden $(x_1, x_2, \ldots, x_n)$ valós szám-$n$-eshez a

$$\{\xi_1 < x_1\}, \{\xi_2 < x_2\}, \ldots, \{\xi_n < x_n\}$$

események együttes bekövetkezésének valószínűségét rendeli, azaz

$$F(x_1, x_2, \ldots, x_n) = P(\xi_1 < x_1, \xi_2 < x_2, \ldots, \xi_n < x_n)$$
$$((x_1, x_2, \ldots, x_n) \in \mathbb{R}^n)$$

**Definíció.** Egy $(\xi_1, \xi_2, \ldots, \xi_n)$ valószínűségi vektorváltozó $\xi_i$ ($i = 1, 2, \ldots, n$) komponensének az eloszlását a $\xi_i$-hez tartozó *peremeloszlásnak* nevezzük, és az eloszlásfüggvényét $F_i$-vel jelöljük.

A peremeloszlást *vetület-* vagy *marginális eloszlásnak* is szokták nevezni.

**Definíció.** Egy folytonos $(\xi_1, \xi_2, \dots, \xi_n)$ valószínűségi vektorváltozót *folytonos eloszlásúnak* mondunk, ha van olyan $f : \mathbb{R}^n \to \mathbb{R}$ függvény, amelyre

$$\int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f(t_1, t_2, \dots, t_n)\, dt_n \dots dt_2\, dt_1 = F(x_1, x_2, \dots, x_n),$$

$$(x_1, x_2, \dots, x_n) \in \mathbb{R}^n .$$

Ezt az $f$ függvényt a valószínűségi vektorváltozó *sűrűségfüggvényének* nevezzük.

**Definíció.** Egy folytonos eloszlású $(\xi_1, \xi_2, \dots, \xi_n)$ valószínűségi vektorváltozó $\xi_i$ komponensének sűrűségfüggvényét a $\xi_i$-hez tartozó *peremsűrűségfüggvénynek* nevezzük, és $f_i$-vel jelöljük $(i = 1, 2, \dots, n)$.

Az eloszlás- és sűrűségfüggvény tulajdonságaira vonatkozó tételeket a kétdimenziós esetre fogalmazzuk meg, de azok természetes módon általánosíthatók az $n$-dimenziós esetre.

**2.14.1. tétel.** Tetszőleges kétdimenziós $(\xi, \eta)$ valószínűségi vektorváltozó $F$ eloszlásfüggvénye az alábbi tulajdonságokkal rendelkezik:

1. Értékkészlete: $R_F = [0,1]$
2. $F$ mindkét változójában monoton növő, azaz $F(x_1, b) \le F(x_2, b)$, valahányszor $x_1 < x_2$, és $F(a, y_1) \le F(a, y_2)$, valahányszor $y_1 < y_2$
3. $\lim\limits_{x \to -\infty} F(x, b) = \lim\limits_{y \to -\infty} F(a, y) = \lim\limits_{x \to -\infty} \lim\limits_{y \to -\infty} F(x, y) = 0$
4. $\lim\limits_{x \to +\infty} \lim\limits_{y \to +\infty} F(x, y) = 1$

5. $\lim\limits_{y\to+\infty} F(a,y)=F_1(a)$ és $\lim\limits_{x\to+\infty} F(x,b)=F_2(b)$

6. $P(a_1\le\xi\le a_2,\, b_1\le\eta\le b_2)=F(a_2,b_2)+F(a_1,b_1)-F(a_1,b_2)-F(a_2,b_1)$

**Bizonyítás.** Csak az utolsó két állítást igazoljuk.

5. $\lim\limits_{y\to+\infty} F(a,y)=\lim\limits_{y\to+\infty} P(\xi<a,\eta<y)=P(\xi<a,\eta<+\infty)=P(\xi<a)=F_1(a)$

   A $\lim\limits_{x\to+\infty} F(x,b)=F_2(b)$ egyenlőség hasonlóan igazolható.

6. Az $A_i=\{\xi<a_i\},\ B_i=\{\eta<b_i\}\ (i=1,2)$ események bevezetésével az állítás az alábbi egyenlőségekből következik:

$$\begin{gathered}
=P(a_1\le\xi\le a_2,\, b_1\le\eta\le b_2)=\\
=P((A_2-A_1)\cdot(B_2-B_1))=P(A_2\cdot(B_2-B_1)-A_1\cdot(B_2-B_1))=\\
=P(A_2\cdot(B_2-B_1))-P(A_1\cdot(B_2-B_1))=\\
=P(A_2\cdot B_2)-P(A_2\cdot B_1)-(P(A_1\cdot B_2)-P(A_1\cdot B_1))=\\
=P(\xi\le a_2,\,\eta\le b_2)+P(\xi\le a_1,\,\eta\le b_1)-\\
-P(\xi\le a_1,\,\eta\le b_2)-P(\xi\le a_2,\,\eta\le b_1)=\\
=F(a_2,b_2)+F(a_1,b_1)-F(a_1,b_2)-F(a_2,b_1).
\end{gathered}$$

**2.14.2. tétel.** Tetszőleges folytonos eloszlású kétdimenziós $(\xi,\eta)$ valószínűségi vektorváltozó $f$ sűrűségfüggvénye az alábbi tulajdonságokkal rendelkezik:

1. $f(x,y)\ge 0,\quad (x,y)\in I\!R^2$

2. $\partial_x\partial_y F=\partial_y\partial_x F=f$, ahol $\partial_x\partial_y F$ és $\partial_y\partial_x F$ az $F$ eloszlásfüggvény másodrendű vegyes parciális deriváltjait jelölik.

3. $\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(x,y)\,dy\,dx=1$

4. $\int\limits_{-\infty}^{+\infty} f(a,t)\,dt=f_1(a)$ illetve $\int\limits_{-\infty}^{+\infty} f(s,b)\,ds=f_2(b)$.

5. $$P(a_1 \le \xi \le a_2, b_1 \le \eta \le b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f(s,t)\,dt\,ds.$$

**Bizonyítás.** Csak a 4) állítást igazoljuk. Az eloszlásfüggvényre vonatkozó tétel 5) állítása szerint $F_1(x) = \lim_{y \to +\infty} F(x,y)$ és így

$$F_1(x) = \lim_{y \to +\infty} F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f(s,t)\,dt\,ds.$$

A $g(s) := \int_{-\infty}^{+\infty} f(s,t)\,dt$ jelölést bevezetve és felhasználva, hogy $\partial_x \left( \int_{-\infty}^{x} g(s)\,ds \right) = g(x)$, innen azt kapjuk, hogy

$$f_1(x) = \partial_x F_1(x) = \partial_x \left( \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f(s,t)\,dt\,ds \right) = \partial_x \left( \int_{-\infty}^{x} g(s)\,ds \right) =$$

$$= g(x) = \int_{-\infty}^{+\infty} f(x,t)\,dt$$

ami igazolja az állításunkat.

**2.14.3. tétel.** Tetszőleges diszkrét kétdimenziós $(\xi,\eta)$ valószínűségi vektorváltozó valószínűségeloszlásából a peremeloszlásait a

$$p_{i*} = P(\xi = x_i) = \sum_k p_{ik}$$

és a

$$p_{*k} = P(\eta = y_k) = \sum_i p_{ik}$$

formulákkal állíthatjuk elő, ahol

$$p_{ik} := P(\xi = x_i, \eta = y_k).$$

Ha a $p_{ik}$ számokat (véges vagy végtelen) mátrixba rendezzük, a peremeloszlásokat e mátrix sor- és oszlopösszegei adják.

## 2.15. Valószínűségi változók függetlensége

**Definíció.** A $\xi_1, \xi_2, \ldots, \xi_n$ valószínűségi változókat (teljesen) függetleneknek nevezzük, ha tetszőleges $x_1, x_2, \ldots, x_n$ valós számok esetén a $\{\xi_1 < x_1\}, \{\xi_2 < x_2\}, \ldots, \{\xi_n < x_n\}$ események (teljesen) függetlenek.

A definícióból nyomban adódik, hogy ha $\xi_1, \xi_2, \ldots, \xi_n$ (teljesen) függetlenek, akkor minden $c_1, c_2, \ldots, c_n$ szám esetén a $\xi_1 + c_1$, $\xi_2 + c_2$, …, $\xi_n + c_n$ valószínűségi változók is (teljesen) függetlenek.

A függetlenség definíciójából közvetlenül következik az alábbi két tétel:

**2.15.1. tétel.** Ha a $(\xi_1, \xi_2, \ldots, \xi_n)$ valószínűségi vektorváltozó $\xi_i$ komponensei $(i = 1, 2, \ldots, n)$ függetlenek, akkor az eloszlásfüggvény az egyes komponensek eloszlásfüggvényeinek szorzata, azaz:

$$F(x_1, x_2, \ldots, x_n) = F_1(x_1) \cdot F_2(x_2) \cdot \ldots \cdot F_n(x_n) \quad ((x_1, x_2, \ldots, x_n) \in \mathbb{R}^n).$$

**2.15.2. tétel.** Ha a folytonos eloszlású $(\xi_1, \xi_2, \ldots, \xi_n)$ valószínűségi vektorváltozó $\xi_i$ komponensei $(i = 1, 2, \ldots, n)$ függetlenek, akkor a sűrűségfüggvény az egyes komponensek sűrűségfüggvényeinek szorzata, azaz

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1) \cdot f_2(x_2) \cdot \ldots \cdot f_n(x_n) \quad ((x_1, x_2, \ldots, x_n) \in \mathbb{R}^n).$$

A fenti tételek megfordítása is igaz: a szóban forgó valószínűségi változók *pontosan akkor* függetlenek, ha az együttes sűrűségfüggvény (eloszlásfüggvény) a perem-sűrűségfüggvények (perem-eloszlásfüggvények) szorzata.

## 2.16. Több valószínűségi változó transzformáltjának várható értéke és szórása

**2.16.1. tétel.** A $(\xi_1, \xi_2, \dots, \xi_n)$ valószínűségi vektorváltozók $h : \mathbb{R}^n \to \mathbb{R}$ függvényének várható értéke a diszkrét esetben a

$$M(h(\xi_1, \xi_2, \dots, \xi_n)) = \sum_{i_1}\sum_{i_2}\dots\sum_{i_n} h(x_{i_1}, x_{i_2}, \dots, x_{i_n}) \cdot p_{i_1, i_2, \dots, i_n},$$

a folytonos esetben pedig a

$$M(h(\xi_1, \xi_2, \dots, \xi_n)) =$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\dots\int_{-\infty}^{+\infty} h(x_1, x_2, \dots, x_n) \cdot f(x_1, x_2, \dots, x_n)\, dx_n \dots dx_2\, dx_1$$

formulával határozható meg, amennyiben a várható érték létezik.

A $h(\xi_1, \xi_2, \dots, \xi_n)$ valószínűségi változó szórásnégyzete definíció szerint:

$$D^2(h(\xi_1, \xi_2, \dots, \xi_n)) = M\left((h(\xi_1, \xi_2, \dots, \xi_n) - M(h(\xi_1, \xi_2, \dots, \xi_n)))^2\right),$$

feltéve, hogy ez létezik egyáltalán. Ekkor, az egydimenziós esethez hasonlóan, a szórásnégyzet kiszámítása a

$$D^2(h(\xi_1, \xi_2, \dots, \xi_n)) = M(h^2(\xi_1, \xi_2, \dots, \xi_n)) - M^2(h(\xi_1, \xi_2, \dots, \xi_n))$$

formulával is történhet.

**2.16.2. tétel.** Ha a $\xi_1, \xi_2, \dots, \xi_n$ valószínűségi változók várható értéke létezik, akkor létezik az összegük várható értéke is és

$$M(\xi_1 + \xi_2 + \dots + \xi_n) = M(\xi_1) + M(\xi_2) + \dots + M(\xi_n).$$

**Bizonyítás.** Az állítást csak az $n = 2$ esetre igazoljuk. Ez teljes indukcióval általánosítható az $n > 2$ esetre.

*a)* A diszkrét esetben az állítás az

$$M(\xi+\eta)=\sum_i\sum_k(x_i+y_k)\cdot p_{ik}=$$
$$=\sum_i\sum_k x_i\cdot p_{ik}+\sum_i\sum_k y_k\cdot p_{ik}=\sum_i x_i\sum_k p_{ik}+\sum_k y_k\sum_i p_{ik}=$$
$$=\sum_i x_i\cdot p_{i*}+\sum_k y_k\cdot p_{*k}=M(\xi)+M(\eta)$$

egyenlőségekből következik.
*b)* A folytonos esetben pedig az állítást az

$$M(\xi+\eta)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x+y)\cdot f(x,y)\,dy\,dx=$$
$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x\cdot f(x,y)\,dy\,dx+\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}y\cdot f(x,y)\,dy\,dx=$$
$$=\int_{-\infty}^{+\infty}x\int_{-\infty}^{+\infty}f(x,y)\,dy\,dx+\int_{-\infty}^{+\infty}y\int_{-\infty}^{+\infty}f(x,y)\,dx\,dy=$$
$$=\int_{-\infty}^{+\infty}x\cdot f_1(x)\,dx+\int_{-\infty}^{+\infty}y\cdot f_2(y)\,dy=M(\xi)+M(\eta)$$

egyenlőségekből kapjuk.

**2.16.3. tétel.** Ha a $\xi_1,\xi_2,\ldots,\xi_n$ független valószínűségi változók várható értéke létezik, akkor létezik a szorzatuk várható értéke is és

$$M(\xi_1\cdot\xi_2\cdot\ldots\cdot\xi_n)=M(\xi_1)\cdot M(\xi_2)\cdot\ldots\cdot M(\xi_n).$$

**Bizonyítás.** Az állítást csak az $n=2$ esetre igazoljuk. Ez teljes indukcióval általánosítható az $n>2$ esetre.

*a)* A diszkrét esetben az állítás az

$$M(\xi\cdot\eta)=\sum_i\sum_k(x_i\cdot y_k)\cdot p_{ik}=\sum_i\sum_k(x_i\cdot y_k)\cdot(p_{i*}\cdot p_{*k})=$$
$$=\left(\sum_i x_i\cdot p_{i*}\right)\cdot\left(\sum_k y_k\cdot p_{*k}\right)=M(\xi)\cdot M(\eta)$$

egyenlőségekből következik.

*b)* A folytonos esetben pedig az állítást az

$$M(\xi\cdot\eta)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x\cdot y)\cdot f(x,y)\,dy\,dx$$
$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x\cdot y)\cdot(f_1(x)\cdot f_2(y))\,dy\,dx$$
$$=\left(\int_{-\infty}^{+\infty}x\cdot f_1(x)dx\right)\cdot\left(\int_{-\infty}^{+\infty}y\cdot f_2(y)dy\right)=M(\xi)\cdot M(\eta)$$

egyenlőségekből kapjuk.

**2.16.4. tétel.** Ha a $\xi_1,\xi_2,\ldots,\xi_n$ független valószínűségi változók szórása létezik, akkor létezik az összegük szórása is, éspedig:

$$D^2(\xi_1+\xi_2+\ldots+\xi_n)=D^2(\xi_1)+D^2(\xi_2)+\ldots+D^2(\xi_n)$$

**Bizonyítás.** Az állítást csak az $n=2$ esetre igazoljuk. Ez teljes indukcióval általánosítható az $n>2$ esetre.

$$D^2(\xi+\eta)=M((\xi+\eta-M(\xi+\eta))^2)=M((\xi-M(\xi)+\eta-M(\eta))^2)=$$
$$=M((\xi-M(\xi))^2)+2M\big((\xi-M(\xi))\cdot(\eta-M(\eta))\big)+M((\eta-M(\eta))^2)=$$
$$=D^2(\xi)+D^2(\eta)$$

mivel $\xi,\eta$ függetlensége miatt $M\big((\xi-M(\xi))\cdot(\eta-M(\eta))\big)=0$.

A bizonyításban előforduló $\operatorname{cov}(\xi,\eta):=M\big((\xi-M(\xi))\cdot(\eta-M(\eta))\big)$ számot a $\xi$ és $\eta$ valószínűségi változók *kovarianciájának* nevezzük. A 2.16.3. tétel szerint tehát *független valószínűségi változók kovarianciája mindig zérus*. Ez megfordítva általában nincs így: ha a kovariancia zérus, ebből a függetlenség még nem következik.

**2.16.5. következmény.** Ha a $\xi_1,\xi_2,\ldots,\xi_n$ független valószínűségi változók szórása megegyezik, akkor

$$D(\xi_1+\xi_2+\ldots+\xi_n)=\sqrt{n}\cdot\sigma,$$

ahol $\sigma$ a valószínűségi változók közös szórását jelöli.

## 2.17. A nagy számok törvényei

**2.17.1. tétel.** (A nagy számok törvényének Csebisev-féle alakja.) Ha a $\xi_1, \xi_2, \ldots, \xi_n$ azonos várható értékű és szórású független valószínűségi változók $M(\xi_i) = m$ várható értékkel és $D(\xi_i) = \sigma$ szórással $(i = 1, 2, \ldots, n)$, akkor

$$P\left(\left|\frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n} - m\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2 \cdot n}.$$

tetszőleges $\varepsilon > 0$ szám esetén.

**Bizonyítás.** A Csebisev-egyenlőtlenséget a $\xi_1, \xi_2, \ldots, \xi_n$ független valószínűségi változók

$$\xi = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n}$$

számtani közepére alkalmazzuk. $\xi$ várható értéke a 2.16.3. tétel szerint:

$$M(\xi) = M\left(\frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n}\right) = \frac{1}{n} \cdot (m + m + \ldots + m) = m$$

$\xi$ szórása (a 2.16.6.Következményt felhasználva):

$$D(\xi) = D\left(\frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n}\right) = \frac{1}{n} \cdot \left(\sqrt{n} \cdot \sigma\right) = \frac{\sigma}{\sqrt{n}}.$$

A Csebisev-egyenlőtlenség szerint így minden $\lambda > 0$ mellett:

$$P\left(|\xi - m| \geq \lambda \cdot \frac{\sigma}{\sqrt{n}}\right) \leq \frac{1}{\lambda^2}.$$

Speciálisan, $\lambda := \dfrac{\varepsilon \cdot \sqrt{n}}{\sigma}$ választással:

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

mint állítottuk.

A nagy számok törvényének Csebisev-féle alakja független, azonos várható értékű és szórású valószínűségi változók számtani közepének a közös várható értéktől való eltérésére ad becslést. A tétel lényegében azon alapul, hogy ilyen valószínűségi változók átlagának várható értéke változatlan marad, míg szórása *n* növekedésével csökken.

**2.17.2. tétel.** (A nagy számok Bernoulli-törvénye.) Ha $n$ független kísérletet végzünk egy $p = P(A)$ valószínűségű $A$ esemény megfigyelésére és a kísérletek során az $A$ esemény $k_n$ -szer következett be, akkor

$$P\left(\left|\frac{k_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{p\cdot(1-p)}{\varepsilon^2\cdot n},$$

tetszőleges $\varepsilon > 0$ szám esetén.

**Bizonyítás.** Legyen $\xi_i$ az $A$ esemény $i$-edik kísérletében történő megfigyeléséhez rendelt karakterisztikus valószínűségi változó $(i = 1, 2, \ldots, n)$. Ekkor

$$M(\xi_i) = p \text{, és } D(\xi_i) = \sqrt{p\cdot(1-p)} \quad (i = 1, 2, \ldots, n).$$

Ezt felhasználva, az előző tételt a

$$\frac{k_n}{n} = \xi = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n}$$

választással alkalmazva, az igazolni kívánt

$$P\left(\left|\frac{k_n}{n} - p\right| \geq \varepsilon\right) = P(|\xi - p| \geq \varepsilon) \leq \frac{p\cdot(1-p)}{\varepsilon^2\cdot n}$$

egyenlőtlenséget kapjuk.

A nagy számok Bernoulli-törvényét általában

$$P\left(\left|\frac{k_n}{n} - p\right| < \varepsilon\right) = 1 - P\left(\left|\frac{k_n}{n} - p\right| \geq \varepsilon\right) \geq 1 - \frac{p\cdot(1-p)}{\varepsilon^2\cdot n}$$

formában annak becslésére használjuk, hogy a relatív gyakoriság milyen valószínűséggel közelíti meg az adott $A$ esemény $p$ valószínűségét.

Mivel $\lim\limits_{n\to\infty} \frac{p\cdot(1-p)}{\varepsilon^2\cdot n}=0$, ezért a relatív gyakoriságok határértéke az esemény $p$ valószínűségével egyezik meg.

**2.17.3. következmény.** Mivel $p\cdot(1-p)\le \frac{1}{4}$, a nagy számok Bernoulli-törvényéből a

$$P\left(\left|\frac{k_n}{n}-p\right|\ge\varepsilon\right)\le\frac{1}{4\cdot\varepsilon^2\cdot n}$$

egyenlőtlenség teljesülése is következik, ami ismeretlen $p$ esetén is alkalmazható becslést tesz lehetővé.

**1. példa.** (Diszkrét valószínűségi változók együttes eloszlása.) Egy dobozban 22 darab 1-től 22-ig megszámozott cédulát helyeztünk el. Véletlenszerűen kihúzunk egyet. A rajta levő számot két szempontból vizsgáljuk. A $\xi$ valószínűségi változó értéke legyen 0, ha páratlan számot húztunk és 1, ha párosat. Az $\eta$ valószínűségi változó értéke pedig legyen 0, ha nem osztható hárommal a kihúzott szám és 1, ha igen.

*a)* Írjuk fel a $(\xi,\eta)$ kétdimenziós valószínűségi vektorváltozó valószínűség-eloszlását!

*b)* Számítsuk ki a peremeloszlásokat!

**Megoldás.**

*a)* $(\xi,\eta)$ lehetséges értékei:

$$(x_i,y_k)=(0,0),(0,1),(1,0),(1,1).$$

Ezek bekövetkezési valószínségei:

$$p_{00}=P(\xi=0,\eta=0)=\frac{7}{22},\quad p_{01}=P(\xi=0,\eta=1)=\frac{4}{22},$$
$$p_{10}=P(\xi=1,\eta=0)=\frac{8}{22},\quad p_{11}=P(\xi=1,\eta=1)=\frac{3}{22}.$$

*b)* $\xi$ peremeloszlása:

$$p_{0*} = P(\xi = 0) = \frac{11}{22}, \quad p_{1*} = P(\xi = 1) = \frac{11}{22}.$$

$\eta$ peremeloszlása:

$$p_{*0} = P(\eta = 0) = \frac{15}{22}, \quad p_{*1} = P(\eta = 1) = \frac{7}{22}.$$

Táblázatos alakban:

| $\xi / \eta$ | $y_0 = 0$ | $y_1 = 1$ | $\xi$ |
|---|---|---|---|
| $x_0 = 0$ | $p_{00} = \frac{7}{22}$ | $p_{01} = \frac{4}{22}$ | $p_{0*} = \frac{11}{22}$ |
| $x_1 = 1$ | $p_{10} = \frac{8}{22}$ | $p_{11} = \frac{3}{22}$ | $p_{1*} = \frac{11}{22}$ |
| $\eta$ | $p_{*0} = \frac{15}{22}$ | $p_{*1} = \frac{7}{22}$ | 1 |

A peremeloszlás tagjai egy-egy sor (oszlop) adatainak összege.

**2. példa.** (Diszkrét valószínűségi változók együttes eloszlása.) Egy dobozban 30 darab 40 wattos, 20 darab 60 wattos és 40 darab 100 wattos villanyégő van. Kiveszünk véletlenszerűen, visszatevés nélkül, 20 villanyégőt. Jelentse a $\xi$ valószínűségi változó a mintában szereplő 40 wattos, $\eta$ pedig a 60 wattos égők számát. (A 100 wattos égőkből a mintában ekkor nyilvánvalóan $20 - \xi - \eta$ darab van.)

*a)* Írjuk fel a kétdimenziós $(\xi, \eta)$ valószínűségi vektorváltozó valószínűség-eloszlását!

*b)* Számítsuk ki $\xi$ peremeloszlását!

**Megoldás.**

*a)* A $(\xi, \eta)$ valószínűség-eloszlását a klasszikus képlet alapján számítjuk ki.

Az összes lehetséges elemi esemény száma $C_{90,20} = \binom{90}{20}$, mivel 90 darab villanyégő közül 20 darabot ennyiféleképpen választhatunk ki. Ennyi az összes esetek száma.

A 30 darab 40 wattos égő közül $i$ számút, a 20 darab 60 wattos égő közül $j$ számút, és végül a 40 darab 100 wattos égő közül $20-i-j$ számút összesen

$$C_{30,i}\cdot C_{20,j}\cdot C_{40,20-i-j}=\binom{30}{i}\cdot\binom{20}{j}\cdot\binom{40}{20-i-j},\quad 0\le i+j\le 20$$

féleképpen húzhatunk ki. Ennyi a kedvező esetek száma. Ezért

$$p_{ij}=P(\xi=i,\eta=j)=$$
$$=\begin{cases}\dfrac{\binom{30}{i}\cdot\binom{20}{j}\cdot\binom{40}{20-i-j}}{\binom{90}{20}}, & \text{ha } 0\le i+j\le 20,\\ 0, & \text{egyebként.}\end{cases}$$

Ez $(\xi,\eta)$ valószínűség-eloszlása. Az ilyen jellegű eloszlásokat *polihipergeometriai eloszlásnak* szokás nevezni.

*b)* $\xi$ peremeloszlását úgy kapjuk meg, hogy rögzített $i$ érték mellett a $p_{i,j}$ valószínűségeket $j$ szerint összegezzük. A $0\le i+j\le 20$ feltétel miatt az összegzés csak 0-tól $(20-i)$-ig futhat. A peremeloszlás tagjai így:

$$p_{i*}=P(\xi=i)=\sum_{j=0}^{20-i}\frac{\binom{30}{i}\cdot\binom{20}{j}\cdot\binom{40}{20-i-j}}{\binom{90}{20}}$$

$$=\frac{\binom{30}{i}}{\binom{90}{20}}\sum_{j=0}^{20-i}\binom{20}{j}\cdot\binom{40}{20-i-j}=\frac{\binom{30}{i}\cdot\binom{60}{20-i}}{\binom{90}{20}}$$

$$(i=0,1,2,\dots,20).$$

Az összegzésnél felhasználtuk, hogy a hipergeometriai eloszlásból tudjuk, hogy

$$\sum_{j=0}^{n} \frac{\binom{s}{j} \cdot \binom{m-s}{n-j}}{\binom{m}{n}} = 1 \text{ azaz } \sum_{j=0}^{n} \binom{s}{j} \cdot \binom{m-s}{n-j} = \binom{m}{n}.$$

Innen $m = 60$, $s = 20$ és $n = 20 - i$ választással kapjuk a

$$\sum_{j=0}^{20-i} \binom{20}{j} \cdot \binom{40}{20-i-j} = \binom{60}{20-i}$$

egyenlőséget.

Vegyük észre, hogy a peremeloszlások hipergeometriaiak.

**3. példa.** (Folytonos valószínűségi változók együttes eloszlása.) Legyen a $(\xi, \eta)$ valószínűségi vektorváltozó egyenletes eloszlású a $(-a, a) \times (-b, b)$ téglalap alakú tartományon.

*a)* Írjuk fel $(\xi, \eta)$ sűrűség- és eloszlásfüggvényét!

*b)* Az eloszlásfüggvény alapján számítsuk ki a

$$P(\xi < 0, \eta < 0), \quad P\left(\xi < 0, \eta \geq \frac{b}{2}\right), \quad P\left(0 \leq \xi < \frac{a}{2}, 0 \leq \eta < \frac{b}{2}\right), \quad P(\xi \geq 0)$$

valószínűségeket!

**Megoldás.**
*a)* A sűrűségfüggvény:

$$f(x, y) = \begin{cases} \dfrac{1}{2a \cdot 2b}, & \text{ha } (x, y) \in (-a, a) \times (-b, b), \\ 0, & \text{egyébként.} \end{cases}$$

Az eloszlásfüggvény:

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)\,dt\,ds$$

$$= \int_{-a}^{\min\{a,\max\{-a,x\}\}} \int_{-b}^{\min\{b,\max\{-b,y\}\}} \frac{1}{2a\cdot 2b}\,dt\,ds$$

$$= \frac{1}{2a\cdot 2b}\left(\int_{-a}^{\min\{a,\max\{-a,x\}\}} 1\,ds\right)\cdot\left(\int_{-b}^{\min\{b,\max\{-b,y\}\}} 1\,dt\right)$$

$$= \frac{1}{2a\cdot 2b}[s]_{-a}^{\min\{a,\max\{-a,x\}\}}\cdot[t]_{-b}^{\min\{b,\max\{-b,y\}\}}$$

$$= \frac{1}{2a\cdot 2b}\cdot\left(\min\{a,\max\{-a,x\}\}+a\right)\cdot\left(\min\{b,\max\{-b,y\}\}+b\right),$$

vagy más alakban:

$$F(x,y) = \begin{cases} 0, & \text{ha } x \le -a, \text{ vagy } y \le -b, \\ \dfrac{(x+a)}{2a}\cdot\dfrac{(y+b)}{2b}, & \text{ha } -a < x \le a \text{ és } -b < y \le b, \\ \dfrac{y+b}{2b}, & \text{ha } a < x \text{ és } -b < y \le b, \\ \dfrac{x+a}{2a}, & \text{ha } -a < x \le a \text{ és } b < y, \\ 1, & \text{ha } a < x \text{ és } b < y. \end{cases}$$

*b)* A keresett valószínűségek:

$$P(\xi < 0, \eta < 0) = P(\xi < 0)\cdot P(\eta < 0) = F(0,0) = \frac{a\cdot b}{2a\cdot 2b} = \frac{1}{4};$$

$$P\left(\xi<0,\ \frac{b}{2}\le\eta\right)=P\left(-\infty\le\xi<0,\ \frac{b}{2}\le\eta<+\infty\right)=$$
$$=F(0,+\infty)+F\left(-\infty,\frac{b}{2}\right)-F(-\infty,+\infty)-F\left(0,\frac{b}{2}\right)=$$
$$=\frac{a}{2a}+0-0-\frac{a}{2a}\cdot\frac{\frac{b}{2}+b}{2b}=\frac{1}{2}-\frac{3}{8}=\frac{1}{8};$$
$$P\left(0\le\xi<\frac{a}{2},\ 0\le\eta<\frac{b}{2}\right)=$$
$$=F\left(\frac{a}{2},\frac{b}{2}\right)+F(0,0)-F\left(0,\frac{b}{2}\right)-F\left(\frac{a}{2},0\right)=$$
$$=\frac{\frac{a}{2}+a}{2a}\cdot\frac{\frac{b}{2}+b}{2b}+\frac{a}{2a}\cdot\frac{b}{2b}-\frac{a}{2a}\cdot\frac{\frac{b}{2}+b}{2b}-\frac{\frac{a}{2}+a}{2a}\cdot\frac{b}{2b}=$$
$$=\frac{9}{16}+\frac{4}{16}-\frac{6}{16}-\frac{6}{26}=\frac{1}{16};$$
$$P(\xi\ge0)=P(0\le\xi<+\infty,-\infty\le\eta<+\infty)=$$
$$=F(+\infty,+\infty)+F(0,-\infty)-F(0,+\infty)-F(+\infty,-\infty)=$$
$$=1+0-\frac{a}{2a}-0=\frac{1}{2}.$$

A *b)* eredményhez egyszerűbben is eljuthatunk, ha észrevesszük, hogy a komponensek *független* valószínűségi változók.

**4. példa.** (Folytonos valószínűségi változók együttes eloszlása.) Legyen a folytonos eloszlású $(\xi,\eta)$ valószínűségi vektorváltozó sűrűségfüggvénye

$$f(x,y)=\begin{cases} e^{-x-y}, & \text{ha } x>0 \text{ és } y>0,\\ 0, & \text{egyébként.}\end{cases}$$

*a)* Írjuk fel ξ és η perem-sűrűségfüggvényét!

*b)* Számítsuk ki a $P(\xi < 1, \eta < 1)$ és a $P\left(\xi < 1, \eta > \frac{2}{3}\right)$ valószínűségeket!

**Megoldás.** Felhasználva, hogy

$$\int_{-\infty}^{\infty} f(x,y)dy = \int_{0}^{\infty} e^{-x-y}\,dy = e^{-x}\cdot\int_{0}^{\infty} e^{-y}\,dy =$$

$$= e^{-x}\cdot\lim_{a\to\infty}\int_{0}^{a} e^{-y}\,dy = e^{-x}\cdot\lim_{a\to\infty}\left[-e^{-y}\right]_{0}^{a} =$$

$$= e^{-x}\cdot\lim_{a\to\infty}\left[-e^{-a}+1\right] = e^{-x}\cdot\lim_{a\to\infty}\left[-\frac{1}{e^{a}}+1\right] = e^{-x}, \quad (x\in(0,\infty)),$$

a kérdéses perem-sűrűségfüggvények

$$f_1(x) = \begin{cases} e^{-x} & \text{ha } x > 0, \\ 0, & \text{egyébként,} \end{cases}$$

és

$$f_2(y) = \begin{cases} e^{-y} & \text{ha } y > 0, \\ 0, & \text{egyébként.} \end{cases}$$

*b)*

$$P(\xi < 1, \eta < 1) = \int_0^1\int_0^1 e^{-x-y}\,dydx = \int_0^1\int_0^1 e^{-x}\cdot e^{-y}\,dydx =$$

$$= \int_0^1 e^{-x}\left(\int_0^1 e^{-y}\,dy\right)dx = \left(\int_0^1 e^{-x}dx\right)\cdot\left(\int_0^1 e^{-y}\,dy\right) =$$

$$= \left[-e^{-x}\right]_0^1\cdot\left[-e^{-y}\right]_0^1 = \left(-e^{-1}+1\right)^2 \approx 0{,}397.$$

$$P\left(\xi < 1, \eta \geq \frac{3}{2}\right) = \int_0^1 \int_{\frac{3}{2}}^{+\infty} e^{-x-y}\, dy\, dx = \int_0^1 \int_{\frac{3}{2}}^{+\infty} e^{-x} \cdot e^{-y}\, dy\, dx =$$

$$= \int_0^1 e^{-x} \left( \int_{\frac{3}{2}}^{+\infty} e^{-y}\, dy \right) dx = \left( \int_0^1 e^{-x} dx \right) \cdot \left( \int_{\frac{3}{2}}^{+\infty} e^{-y}\, dy \right) =$$

$$= \left[ -e^{-x} \right]_0^1 \cdot \lim_{a \to \infty} \left[ -e^{-y} \right]_{\frac{3}{2}}^{a} = \left( -e^{-1} + 1 \right) \cdot \lim_{a \to \infty} \left( -e^{-a} + e^{-\frac{3}{2}} \right) =$$

$$= \left( -e^{-1} + 1 \right) \cdot e^{-\frac{3}{2}} \approx 0{,}14.$$

A b) eredményhez egyszerűbben is eljuthatunk, ha észrevesszük, hogy a komponensek *független* valószínűségi változók.

**5. példa.** (Valószínűségi változók összegének várható értéke.) Egy löveg addig tüzel egy célpontra, amíg három találatot el nem ér. Az egyes találatokat függetleneknek tekintjük. A célba találás valószínűsége minden egyes lövésnél 0,05. Határozzuk meg a szükséges lövedékek számának várható értékét!

**Megoldás.** A $\xi_i$ valószínűségi változó jelölje az $(i-1)$-edik és az $i$-edik találat közötti lőszerszükségletet $(i = 1, 2, 3)$. Ekkor az elhasznált lőszer

$$\eta = \xi_1 + \xi_2 + \xi_3 .$$

A várható lőszerszükséglet az $\eta$ várható értéke, azaz

$$M(\eta) = M(\xi_1) + M(\xi_2) + M(\xi_3).$$

Feltehető, hogy az itt szereplő valószínűségi változók egyenlő várható értékűek.

A $\xi_i$ valószínűségi változó $(i=1,2,3)$ lehetséges értékei

$$x_k = k\,,\ (k=1,2,\ldots).$$

Mivel minden egyes lövésnél 0,05 a találat valószínűsége, és feltehető, hogy a találatok egymástól függetlenek, $\xi_i$ eloszlása $(i=1,2,3)$:

$$p_k = P(\xi_i = k) = 0{,}95^{k-1}\cdot 0{,}05 \quad (k=1,2,\ldots).$$

$\xi_i$ várható értéke $(i=1,2,3)$ ezért

$$M(\xi_i) = \sum_{k=1}^{\infty} k\cdot 0{,}95^{k-1}\cdot 0{,}05 = 0{,}05\cdot\sum_{k=1}^{\infty} k\cdot 0{,}95^{k-1}$$
$$= 0{,}05\cdot\frac{1}{(1-0{,}95)^2} = \frac{1}{0{,}05} = 20 \quad (i=1,2,3).$$

Itt felhasználtuk, hogy

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}\,,\ \text{ha } |x|<1 \text{ és így } \left(\sum_{k=0}^{\infty} x^k\right)' = \left(\frac{1}{1-x}\right)',$$

azaz

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

Így

$$M(\eta) = M(\xi_1)+M(\xi_2)+M(\xi_3) = 3\cdot 20 = 60\,,$$

tehát a várható lőszerszükséglet a 3 találat eléréséig 60 darab lőszer.

**6. példa.** (Független valószínűségi változók összegének várható értéke.) Mutassuk meg, hogyha a $\xi$ és $\eta$ valószínűségi változók függetlenek, akkor

$$D^2(\xi-\eta) = D^2(\xi)+D^2(\eta).$$

**Megoldás:** Felhasználva, hogy független valószínűségi változók kovarianciája zérus:

$$D^2(\xi-\eta)=M((\xi-\eta-M(\xi-\eta))^2)=M((\xi-M(\xi)-\eta+M(\eta))^2)=$$
$$=M((\xi-M(\xi))^2)-2M\left((\xi-M(\xi))\cdot(\eta-M(\eta))\right)+M((\eta-M(\eta))^2)=$$
$$=D^2(\xi)+2\operatorname{cov}(\xi,\eta)+D^2(\eta)=D^2(\xi)+D^2(\eta)$$

**7. példa.** (Független valószínűségi változók függvényeinek várható értéke.) Legyen $\xi$ és $\eta$ együttes sűrűségfüggvénye

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

Számítsuk ki $\xi-\eta$ és a $\xi\cdot\eta$ várható értékét!

**Megoldás.**

$$M(\xi-\eta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-y)\cdot f(x,y)\,dy\,dx=\int_0^1\int_0^1(x-y)\cdot(x+y)\,dy\,dx=$$

$$=\int_0^1\int_0^1\left(x^2-y^2\right)dy\,dx=\int_0^1\left[x^2\cdot y-\frac{y^3}{3}\right]_0^1 dx=$$

$$=\int_0^1\left(x^2-\frac{1}{3}\right)dx=\left[\frac{x^3}{3}-\frac{1}{3}\cdot x\right]_0^1=\frac{1}{3}-\frac{1}{3}=0,$$

$$M(\xi\cdot\eta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x\cdot y)\cdot f(x,y)\,dy\,dx=\int_0^1\int_0^1(x\cdot y)\cdot(x+y)\,dy\,dx=$$

$$=\int_0^1\int_0^1\left(x^2\cdot y+x\cdot y^2\right)dy\,dx=\int_0^1\left[x^2\cdot\frac{y^2}{2}+x\cdot\frac{y^3}{3}\right]_0^1 dx=$$

$$=\int_0^1\left(\frac{x^2}{2}+\frac{x}{3}\right)dx=\left[\frac{x^3}{6}+\frac{x^2}{6}\right]_0^1=\frac{1}{6}+\frac{1}{6}=\frac{1}{3}.$$

**8. példa.** (A nagy számok Bernoulli törvénye.) Hányszor kell egy szabályos érmét feldobnunk ahhoz, hogy a fejek számának relatív gyakorisága legalább 0,9 valószínűséggel 0,1-nél kevesebb hibával térjen el az esemény valószínűségétől?

**Megoldás.** Az $A$ esemény jelentse azt, hogy fejet dobtunk. Ennek valószínűsége $p = P(A) = \frac{1}{2}$. A hibakorlát $\varepsilon = 0{,}1$. A nagy számok törvényét alkalmazva

$$P\left(\left|\frac{k_n}{n} - p\right| < \varepsilon\right) = 1 - P\left(\left|\frac{k_n}{n} - p\right| \geq \varepsilon\right) \geq 1 - \frac{p\cdot(1-p)}{n\varepsilon^2}.$$

Ezt felhasználva a dobások $n$ számát a

$$1 - \frac{\frac{1}{2}\cdot\frac{1}{2}}{n\cdot\frac{1}{10}^{2}} \geq \frac{9}{10} \quad \Leftrightarrow \quad n \geq 250$$

egyenlőtlenség legkisebb megoldásának választhatjuk. Tehát legalább 250 dobást kell végeznünk ahhoz, hogy a fejek számának relatív gyakorisága legalább 0,9 valószínűséggel 0,1-nél kevesebb hibával térjen el az esemény valószínűségétől.

**9. példa.** (A nagy számok Bernoulli törvénye.) Egy csavargyártó automata esetében kívánjuk meghatározni a selejtgyártás valószínűségét. E célból megvizsgálunk 5000 csavart. Összesen 80 selejtest találunk közöttük. Határozzuk meg, hogy az ebből számított relatív gyakoriság az ismeretlen $p$ valószínűséget 90%-os valószínűséggel mekkora hibával közelíti meg!

**Megoldás.**

A kísérletek száma itt $n = 5000$, $k_n = 80$ és $\frac{k_n}{n} = \frac{80}{5000} = 0{,}016$ a relatív gyakoriság. A feladat megoldásához a nagy számok törvényének

$$P\left(\left|\frac{k_n}{n}-p\right|<\varepsilon\right)=1-P\left(\left|\frac{k_n}{n}-p\right|\geq\varepsilon\right)\geq 1-\frac{1}{4n\varepsilon^2}$$

alakját alkalmazzuk, mivel a $p$ valószínűség nem ismert. A legkisebb $\varepsilon$ hibakorlátot így az

$$1-\frac{1}{4\cdot 5000\cdot\varepsilon^2}\geq 0{,}9 \quad\Leftrightarrow\quad \varepsilon^2\geq\frac{1}{2000} \quad\Leftrightarrow\quad \varepsilon\geq 0{,}022$$

egyenlőtlenségek legkisebb megoldásának választva kapjuk. A nagy számok törvényét alkalmazva tehát $\varepsilon = 0{,}022$ az a legkisebb hibakorlát, amelyre

$$P\left(\left|\frac{k_n}{n}-p\right|<\varepsilon\right)=P\left(\left|0{,}016-p\right|<0{,}022\right)\geq 0{,}9$$

teljesül. Ezért a 0,016 relatív gyakoriság az ismeretlen $p$ valószínűséget 90%-os valószínűséggel 0,022-nél kisebb hibával közelíti meg.

## 2.18. Feltételes eloszlások

**Definíció.** Legyen $(\xi,\eta)$ egy diszkrét valószínűségi vektorváltozó és a lehetséges értékei az $(x_i, y_k)$ $(i,k=1,2,\ldots)$ számpárok összessége. A $\{\xi = x_i\}$ *esemény* $\{\eta = y_k\}$ *feltétel melletti valószínűségén* a

$$P(\xi = x_i \mid \eta = y_k)=\frac{P(\xi = x_i, \eta = y_k)}{P(\eta = y_k)}=:\frac{p_{ik}}{p_{*k}}$$

számot értjük, ha $p_{*k} > 0$.

**2.18.1. tétel.** A $P(\xi = x_i \mid \eta = y_k)=\dfrac{p_{ik}}{p_{*k}}$ $(i=1,2,\ldots)$ számok összessége valószínűség-eloszlást határoz meg, amelyet a $\xi$ valószínűségi változó $\{\eta = y_k\}$ feltétel melletti eloszlásának nevezünk.

**Bizonyítás.** $\sum\limits_{i=1}^{\infty}\dfrac{p_{ik}}{p_{*k}}=\dfrac{1}{p_{*k}}\sum\limits_{i=1}^{\infty}p_{ik}=\dfrac{p_{*k}}{p_{*k}}=1,$ ami igazolja az állítást.

**Definíció.** Legyen $(\xi, \eta)$ egy tetszőleges valószínűségi vektorváltozó. A $\xi$ valószínűségi változó $\{y_1 \leq \eta < y_2\}$ *feltétel melletti eloszlásfüggvényén* az

$$F(x | y_1 \leq \eta < y_2) := P(\xi < x | y_1 \leq \eta < y_2) \quad (x \in \mathbb{R})$$

függvényt értjük.

**2.18.2. tétel.** Legyen a $(\xi, \eta)$ valószínűségi vektorváltozó eloszlásfüggvénye $F$ és $\eta$ peremeloszlás-függvénye $F_2$. Ekkor a $\xi$ valószínűségi változó $\{y_1 \leq \eta < y_2\}$ feltétel melletti eloszlásfüggvénye

$$F(x | y_1 \leq \eta < y_2) = \frac{F(x, y_2) - F(x, y_1)}{F_2(y_2) - F_2(y_1)} \quad (x \in \mathbb{R})$$

alakba írható, ha $F_2(y_2) \neq F_2(y_1)$.

**Bizonyítás.** Az $A = \{\xi < x\}$ és $B_i = \{\eta < y_i\}$ $(i = 1, 2)$ események bevezetésével az állítás az alábbi egyenlőségekből következik:

$$\begin{aligned} F(x | y_1 \leq \eta < y_2) &= P(\xi < x | y_1 \leq \eta < y_2) = P(A | B_2 - B_1) = \\ &= \frac{P(A \cdot (B_2 - B_1))}{P(B_2 - B_1)} = \frac{P(A \cdot B_2 - A \cdot B_1)}{P(B_2 - B_1)} = \\ &= \frac{P(A \cdot B_2) - P(A \cdot B_1)}{P(B_2) - P(B_1)} = \frac{P(\xi < x, \eta < y_2) - P(\xi < x, \eta < y_1)}{P(\eta < y_2) - P(\eta < y_1)} = \\ &= \frac{F(x, y_2) - F(x, y_1)}{F_2(y_2) - F_2(y_1)}. \end{aligned}$$

**Definíció.** Legyen $(\xi, \eta)$ egy tetszőleges valószínűségi vektorváltozó. A $\xi$ valószínűségi változó $\{\eta = y\}$ *feltétel melletti eloszlásfüggvényét* az

$$F(x | y) = \lim_{h \to 0} F(x | y \leq \eta < y + h) \qquad (x \in \mathbb{R}),$$

határértékkel értelmezzük, amennyiben ez a határérték létezik.

**2.18.3. tétel.** Legyen $(\xi, \eta)$ egy tetszőleges folytonos eloszlású valószínűségi vektorváltozó, melynek eloszlásfüggvénye $F$, és $\eta$ perem-sűrűségfüggvénye $f_2$. Ekkor a $\xi$ valószínűségi változó $\{\eta = y\}$ feltétel melletti eloszlásfüggvénye létezik és az

$$F(x|y) = \frac{\partial_y F(x,y)}{f_2(y)} \qquad (x \in \mathbb{R})$$

alakba írható, ha $f_2(y) \neq 0$.

**Bizonyítás**. A 2.18.2. tétel eredményeit felhasználva

$$F(x|y) = \lim_{h\to 0} F(x|y \le \eta < y+h) =$$

$$= \lim_{h\to 0} \frac{F(x,y+h)-F(x,y)}{F_2(y+h)-F_2(y)} = \lim_{h\to 0} \frac{\frac{F(x,y+h)-F(x,y)}{h}}{\frac{F_2(y+h)-F_2(y)}{h}} =$$

$$= \frac{\lim\limits_{h\to 0} \frac{F(x,y+h)-F(x,y)}{h}}{\lim\limits_{h\to 0} \frac{F_2(y+h)-F_2(y)}{h}} = \frac{\partial_y F(x,y)}{f_2(y)},$$

mint állítottuk.

**Definíció.** Legyen $(\xi, \eta)$ egy folytonos eloszlású valószínűségi vektorváltozó. A $\xi$ valószínűségi változó $\{\eta = y\}$ *feltétel melletti sűrűségfüggvényét* az $F(x|y)$ feltétel melletti eloszlásfüggvényének $x$ változó szerinti parciális deriváltjaként értelmezzük:

$$f(x|y) = \partial_x F(x|y) \qquad (x \in \mathbb{R}).$$

**2.18.4. tétel.** Legyen $(\xi, \eta)$ egy tetszőleges folytonos eloszlású valószínűségi vektorváltozó, amelynek a sűrűségfüggvénye $f$ és $\eta$ perem-sűrűségfüggvénye $f_2$. Ekkor a $\xi$ valószínűségi változó $\{\eta = y\}$ feltétel melletti sűrűségfüggvénye létezik és

$$f(x|y)=\frac{f(x,y)}{f_2(y)} \qquad (x\in I\!R)$$

alakba írható, ha $f_2(y)\neq 0$.

**Bizonyítás**. A 2.18.3. tétel eredményeit felhasználva

$$f(x|y)=\partial_x F(x|y)=\partial_x\left(\frac{\partial_y F(x,y)}{f_2(y)}\right)=$$
$$=\frac{1}{f_2(y)}\partial_x\left(\partial_y F(x,y)\right)=\frac{1}{f_2(y)}\cdot f(x,y),$$

mint állítottuk.

## 2.19. A feltételes várható érték

**Definíció.** Legyen $(\xi,\eta)$ egy diszkrét valószínűségi vektorváltozó és a lehetséges értékei az $(x_i, y_k)$ $(i,k=1,2,\ldots)$ számpárok összessége. Ha $\sum_{i=1}^{\infty}|x_i|\cdot P(\xi=x_i\mid\eta=y_k)<\infty$, akkor a $\xi$ valószínűségi változó $\{\eta=y_k\}$ *feltétel melletti várható értékén* az

$$M(\xi|\eta=y_k):=\sum_{i=1}^{\infty}x_i\cdot P(\xi=x_i\mid\eta=y_k)=$$
$$=\sum_{i=1}^{\infty}x_i\cdot\frac{P(\xi=x_i,\,\eta=y_k)}{P(\eta=y_k)}=\sum_{i=1}^{\infty}x_i\cdot\frac{p_{ik}}{p_{*k}}$$

összeget értjük. Egyébként azt mondjuk, hogy ez a várható érték nem létezik.

Megjegyezzük, hogy $M(\xi|\eta=y_k)$ nem függ $y_k$ konkrét értékétől.

**Definíció.** Legyen $(\xi,\eta)$ egy folytonos eloszlású valószínűségi vektorváltozó és legyen $\xi$-nek az $\{\eta=y\}$ feltétel melletti sűrűségfüggvénye

$f(x|y)$. Ha $\int_{-\infty}^{+\infty}|x|\cdot f(x|y)dx<\infty$, akkor a $\xi$ valószínűségi változó $\{\eta=y\}$ *feltétel melletti várható értékén* az

$$M(\xi|\eta=y):=\int_{-\infty}^{+\infty}x\cdot f(x|y)dx=\int_{-\infty}^{+\infty}x\cdot\frac{f(x,y)}{f_2(y)}dx$$
$$=\frac{1}{f_2(y)}\cdot\int_{-\infty}^{+\infty}x\cdot f(x,y)dx$$

határozott integrált értjük. Egyébként azt mondjuk, hogy ez a várható érték nem létezik.

Amennyiben az $M(\xi|\eta=y_k)$ ill. az $M(\xi|\eta=y)$ feltételes várható értéket $\eta$-nak a véletlentől függő $y_k$ ill. $y$ értékeivel képezzük, valószínűségi változót kapunk (melyet röviden $M(\xi|\eta)$-val jelölünk). A fenti kifejezésekből könnyen látható, hogy az $M(\xi|\eta)$ feltételes várható érték valójában az $\eta$ valószínűségi változó függvénye.

**2.19.1. tétel.** Egy $\xi$ valószínűségi változó $\eta$-ra vonatkoztatott feltételes várható értéke, mint az $\eta$ függvényének várható értéke, megegyezik a $\xi$ feltétel nélküli várható értékével, azaz

$$M(M(\xi|\eta))=M(\xi).$$

**Bizonyítás.**
*a)* A diszkrét esetben az állítást az

$$M(M(\xi|\eta))=\sum_{k=1}^{\infty}M(\xi|\eta=y_k)\cdot P(\eta=y_k)=$$
$$=\sum_{k=1}^{\infty}\left(\sum_{i=1}^{\infty}x_i\cdot\frac{p_{ik}}{p_{*k}}\right)\cdot p_{*k}=\sum_{k=1}^{\infty}\left(\sum_{i=1}^{\infty}x_i\cdot p_{ik}\right)=$$
$$=\sum_{i=1}^{\infty}x_i\cdot\left(\sum_{k=1}^{\infty}p_{ik}\right)=\sum_{k=i}^{\infty}x_i\cdot p_{i*}=M(\xi)$$

egyenlőség bizonyítja.

*b)* A folytonos esetben pedig az állítást az

$$M\left(M(\xi\,|\,\eta)\right)=\int_{-\infty}^{+\infty} M(\xi|\eta=y)\cdot f_2(y)dy=$$

$$=\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} x\cdot\frac{f(x,y)}{f_2(y)}dx\right)\cdot f_2(y)dy=\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} x\cdot f(x,y)dx\right)dy=$$

$$=\int_{-\infty}^{+\infty} x\cdot\left(\int_{-\infty}^{+\infty} f(x,y)dy\right)dx=\int_{-\infty}^{+\infty} x\cdot f_1(x)dx=M(\xi)$$

egyenlőségekből kapjuk.

**2.19.2. tétel.** Ha $\xi$ és $\eta$ független valószínűségi változók, akkor a feltételes várható értékük megegyezik a feltétel nélküli várható értékükkel, vagyis

$$M(\xi|\eta=y)=M(\xi) \text{ és } M(\eta|\xi=x)=M(\eta).$$

(Szemléletesen: Ha $\eta$ változása nincs hatással $\xi$ eloszlására, akkor a várható értékét sem befolyásolja.)

**Bizonyítás.** Csak az $M(\xi|\eta=y)=M(\xi)$ esetet igazoljuk.
*a)* A diszkrét esetben az állítást az

$$M(\xi|\eta=y_k)=\sum_{i=1}^{\infty} x_i\cdot\frac{p_{ik}}{p_{*k}}$$

$$=\sum_{i=1}^{\infty} x_i\cdot\frac{p_i\cdot p_{*k}}{p_{*k}}=\sum_{i=1}^{\infty} x_i\cdot p_{i*}=M(\xi)$$

egyenlőség bizonyítja, ahol felhasználtuk, hogy független valószínűségi változók esetén

$$p_{ik}=p_{i*}\cdot p_{*k}\,.$$

*b)* A folytonos esetben az állítást az

$$M(\xi|\eta=y)=\frac{1}{f_2(y)}\cdot\int_{-\infty}^{+\infty}x\cdot f(x,y)dx=$$
$$=\frac{1}{f_2(y)}\cdot\int_{-\infty}^{+\infty}x\cdot f_1(x)\cdot f_2(y)dx=\int_{-\infty}^{+\infty}x\cdot f_1(x)dx=M(\xi)$$

egyenlőségekből kapjuk, ahol felhasználtuk, hogy független valószínűségi változók esetén
$$f(x,y)=f_1(x)\cdot f_2(y).$$

## 2.20. A korrelációs együttható

Az alábbiakban olyan mérőszámot vezetünk be, amely alkalmas lesz arra, hogy valószínűségi változók közötti esetleges lineáris jellegű kapcsolat erősségét jól jellemezze.

Mindenekelőtt idézzük fel két valószínűségi változó kovarianciájának fogalmát (2.17. szakasz):

**Definíció.** A $\xi$ és $\eta$ valószínűségi változókból képzett $\xi-M(\xi)$ és $\eta-M(\eta)$ valószínűségi változók szorzatának várható értékét a $\xi$ és $\eta$ valószínűségi változók *kovarianciájának* nevezzük és
$$\operatorname{cov}(\xi,\eta)=M((\xi-M(\xi))\cdot(\eta-M(\eta)))$$
módon jelöljük, ha az itt szereplő várható értékek léteznek.

**2.20.1. tétel.** Ha a $\xi$ és $\eta$ valószínűségi változók kovarianciája létezik, akkor az a
$$\operatorname{cov}(\xi,\eta)=M(\xi\cdot\eta)-M(\xi)\cdot M(\eta)$$
alakba is írható.

**Bizonyítás.** A várható értékre vonatkozó tételek felhasználásával
$$\operatorname{cov}(\xi,\eta)=M((\xi-M(\xi))\cdot(\eta-M(\eta)))=$$
$$=M(\xi\cdot\eta-M(\eta)\cdot\xi-M(\xi)\cdot\eta+M(\xi)\cdot M(\eta))=$$
$$=M(\xi\cdot\eta)-M(\eta)\cdot M(\xi)-M(\eta)\cdot M(\xi)+M(\eta)\cdot M(\xi)=$$
$$=M(\xi\cdot\eta)-M(\eta)\cdot M(\xi),$$
amit igazolni akartunk.

E tétel segítségével különösen egyszerűen igazolható a korábban már szerepelt megállapítás:

**2.20.2. tétel.** Ha a független $\xi$ és $\eta$ valószínűségi változók kovarianciája létezik, akkor szükségképp zérus: $\operatorname{cov}(\xi,\eta)=0$.

**Bizonyítás.** Az előző tétel szerint

$$\operatorname{cov}(\xi,\eta)=M(\xi\cdot\eta)-M(\eta)\cdot M(\xi).$$

Ha a $\xi$ és $\eta$ függetlenek, akkor viszont

$$M(\xi\cdot\eta)=M(\eta)\cdot M(\xi),$$

és így

$$\operatorname{cov}(\xi,\eta)=M(\eta)\cdot M(\xi)-M(\eta)\cdot M(\xi)=0,$$

amit igazolni akartunk.

**2.20.3. tétel.** Ha a $\xi$ és $\eta$ valószínűségi változók kovarianciája létezik és közöttük lineáris kapcsolat áll fenn, azaz $\eta=a\cdot\xi+b$, ahol $a\neq 0$, akkor

$$|\operatorname{cov}(\xi,\eta)|=D(\xi)\cdot D(\eta).$$

**Bizonyítás.**

$$\begin{aligned}
|\operatorname{cov}(\xi,\eta)|&=|M(\xi\cdot\eta)-M(\xi)\cdot M(\eta)|=\\
&=|M(\xi\cdot(a\cdot\xi+b))-M(\xi)\cdot M(a\cdot\xi+b)|=\\
&=\left|M\left(a\cdot\xi^2+b\cdot\xi\right)-M(\xi)\cdot M(a\cdot\xi+b)\right|=\\
&=\left|a\cdot M\left(\xi^2\right)+b\cdot M(\xi)-a\cdot M(\xi)\cdot M(\xi)-b\cdot M(\xi)\right|=\\
&=\left|a\cdot M\left(\xi^2\right)-a\cdot M^2(\xi)\right|=|a|\cdot D^2(\xi)=\\
&=D(\xi)\cdot|a|\cdot D(\xi)=D(\xi)\cdot D(\eta),
\end{aligned}$$

ahol felhasználtuk azt, hogy $D(\eta)=|a|\cdot D(\xi)$.

A tétel meg is fordítható, sőt a kovariancia abszolút értékére éles becslés is adható:

**2.20.4. tétel.** Ha a $\xi$ és $\eta$ valószínűségi változók kovarianciája létezik, akkor mindig érvényes az alábbi becslés:

$$|\mathrm{cov}(\xi,\eta)| \le D(\xi)\cdot D(\eta),$$

A becslésben az egyenlőség pontosan akkor teljesül, ha $\xi$ és $\eta$ között lineáris kapcsolat áll fenn, azaz $\eta = a\cdot\xi + b$ vagy $\xi = a\cdot\eta + b$ alakú.

**Definíció.** A $\xi$ és $\eta$ valószínűségi változók *korrelációs együtthatóján* az

$$R(\xi,\eta) = \frac{\mathrm{cov}(\xi,\eta)}{D(\xi)\cdot D(\eta)} = \frac{M(\xi\cdot\eta) - M(\xi)\cdot M(\eta)}{D(\xi)\cdot D(\eta)}$$

hányadost értjük, ha az itt szereplő várható értékek és szórások léteznek, és a szórások egyike sem 0.

$D(\xi)$ (ill. $D(\eta)$) csak úgy lehet 0, hogy $\xi$ (ill. $\eta$) 1 valószínűséggel konstans. Ekkor pedig $\mathrm{cov}(\xi,\eta) = 0$, ezért ekkor a korrelációs együtthatót 0-nak definiálhatjuk.

Két valószínűségi változó kovarianciájának tulajdonságai alapján a korrelációs együtthatóról a következőket mondhatjuk:

**2.20.5. tétel.** Ha a $\xi$ és $\eta$ valószínűségi változók korrelációs együtthatója létezik, akkor

1. A korrelációs együttható értéke mindig $-1$ és 1 közé esik, azaz
$$-1 \le R(\xi,\eta) \le 1.$$
2. Ha $\xi$ és $\eta$ függetlenek, akkor a korrelációs együttható értéke 0, azaz
$$R(\xi,\eta) = 0.$$
3. A korrelációs együttható abszolút értéke akkor és csak akkor 1, ha a két valószínűségi változó között lineáris kapcsolat áll fenn, azaz
$$\eta = a\cdot\xi + b \text{ vagy } \xi = a\cdot\eta + b \text{ (ahol } a \ne 0\text{)}.$$
Ez esetben $R(\xi,\eta) = 1$, ha $a > 0$, és $R(\xi,\eta) = -1$, ha $a < 0$.

Ha $R(\xi,\eta)=0$, akkor azt mondjuk, hogy a $\xi$ és $\eta$ valószínűségi változók *korrelálatlanok*. Mint a fenti tétel mutatja, a korrelálatlanság valamivel gyengébb a függetlenségnél: megmutatható azonban, hogy bizonyos esetekben, így pl. ha ($\xi$, $\eta$) kétdimenziós normális eloszlású, akkor a korrelálatlanságból már következik a függetlenség is.

A korrelációs együttható négyzetét *meghatározottsági együtthatónak* is szokás nevezni.

## 2.21. A regresszió

**Definíció.** Az $r(y)=M(\xi|\eta=y)$ függvényt a $\xi$ valószínűségi változó $\eta$-ra vonatkozó *regressziós függvényének* nevezzük (ha az $\{\eta=y\}$ feltétel melletti várható értékékek léteznek).

**2.21.1. tétel.** Legyen $(\xi,\eta)$ tetszőleges valószínűségi vektorváltozó és tegyük fel, hogy az ehhez tartozó $r(y)$ regressziós függvény létezik. Ekkor bármilyen $u(y)$ egyváltozós függvény esetében

$$M\left((\xi-u(\eta))^2\right)\geq M\left((\xi-r(\eta))^2\right),$$

feltéve, hogy az itt szereplő várható értékek léteznek.

**Bizonyítás.** A tételt csak a folytonos esetben igazoljuk. Induljunk ki a bal oldal várható értékéből és alakítsuk át az alábbiak szerint:

$$\begin{aligned} M\left((\xi-u(\eta))^2\right) &= M\left((\xi-r(\eta)+r(\eta)-u(\eta))^2\right)= \\ &= M\left((\xi-r(\eta))^2\right)+2\cdot M\left((\xi-r(\eta))\cdot(r(\eta)-u(\eta))\right)= \\ &\quad + M\left((r(\eta)-u(\eta))^2\right) \end{aligned}$$

A jobb oldal második tagját az

$$f(x,y)=f(x\mid y)\cdot f_2(y)$$

azonosság alkalmazásával és az integrálás sorrendjének felcserélésével

$$M\left((\xi - r(\eta))\cdot(r(\eta)-u(\eta))\right) =$$
$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x - M(\xi|\eta = y))\cdot(M(\xi|\eta = y) - u(y)) f(x,y)\,dy\,dx =$$
$$= \int_{-\infty}^{+\infty}(M(\xi|\eta = y) - u(y))\cdot f_2(y)\cdot\left(\int_{-\infty}^{+\infty}(x - M(\xi|\eta = y))\cdot f(x|y)\,dx\right)dy$$

alakba írhatjuk. A belső integrál és így a második tag 0, ugyanis

$$\int_{-\infty}^{+\infty}(x - M(\xi|\eta = y))\cdot f(x|y)\,dx =$$
$$= \int_{-\infty}^{+\infty} x\cdot f(x|y)\,dx - M(\xi|\eta = y)\int_{-\infty}^{+\infty} f(x|y)\,dx =$$
$$= M(\xi|\eta = y) - M(\xi|\eta = y)\int_{-\infty}^{+\infty}\frac{f(x,y)}{f_2(y)}\,dx =$$
$$= M(\xi|\eta = y) - M(\xi|\eta = y)\frac{1}{f_2(y)}\cdot f_2(y) =$$
$$= M(\xi|\eta = y) - M(\xi|\eta = y) = 0.$$

Mivel $(r(\eta)-u(\eta))^2 \geq 0$ ezért a várható értéke biztosan nem negatív, azaz $M\left((r(\eta)-u(\eta))^2\right) \geq 0,$ és így a jobb oldal harmadik tagja sem az. A második és a harmadik tag elhagyásával tehát nem negatív értékeket hagyunk el, és ez igazolja a tételben szereplő egyenlőtlenséget.

Az imént bizonyított tétel szerint nem független $\xi$ és $\eta$ valószínűségi változók esetén, ha az egyikre vonatkozó kísérleti eredményekből a másikra akarunk következtetni, a legjobb közelítést – a tételben szereplő legkisebb négyzetek elve alapján – a regressziós függvény segítségével nyerhetjük. Független valószínűségi változók esetén a bizonyított $M(\xi|\eta = y) = M(\xi)$ azonosság miatt a regressziós függvény konstans, és ez azt mutatja, hogy ez esetben az egyik változóra kapott adatokból a másikra következtetni nem lehet.

A tétel a *Steiner-egyenlőtlenség* (lásd 2.9.8. tétel) segítségével egyszerűbben is bizonyítható:

$$M((\xi-u(\eta))^2)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-u(y))^2 f(x,y)\,dxdy=$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-u(y))^2 f(x\mid y)f_2(y)\,dxdy=$$

$$=\int_{-\infty}^{+\infty} f_2(y)\left(\int_{-\infty}^{+\infty}(x-u(y))^2 f(x\mid y)\,dx\right)dy$$

és a belső integrál a Steiner-tétel értelmében akkor minimális, ha $u(y)=M(\xi\mid\eta=y)$.

## 2.22. A másodfajú regresszió

Az előző szakaszban láttuk, hogy a $\xi$ és az $\eta$ valószínűségi változók regressziós függvénye egy minimumfeladat megoldása, aminek meghatározása általában nem könnyű. Ezért természetes az a törekvés, hogy egyszerűbb függvénytípust keressünk, és azon belül azt a függvényt, amelynek hasonló minimum tulajdonsága van, mint a regressziós függvénynek. Így jutunk el a regressziós egyenes, regressziós parabola, stb. fogalmához. Az ilyen módon készített függvényeket *másodfajú regressziós függvényeknek* nevezzük.

Itt részletesebben csak a regressziós egyenessel foglalkozunk. Az ismertetett eljárás természetesen másfajta függvények meghatározására is alkalmazható.

Feladatunk a következő: Keressük azt az

$$x=a\cdot y+b$$

egyenest, amelyre az

$$S(a,b)=M\left((\xi-(a\cdot\eta+b))^2\right)$$

függvény minimális lesz.

A feladatot a kétdimenziós függvényekre vonatkozó szélsőérték számítási módszerekkel oldjuk meg. Ehhez végezzük el az

$$\begin{aligned} S(a,b) &= M\left((\xi - a\cdot\eta - b)^2\right) \\ &= M\left(\xi^2 - 2a\cdot\xi\cdot\eta - 2b\cdot\xi + a^2\cdot\eta^2 + 2ab\cdot\eta + b^2\right) \\ &= M\left(\xi^2\right) - 2a\cdot M(\xi\cdot\eta) - 2b\cdot M(\xi) + a^2\cdot M\left(\eta^2\right) + 2ab\cdot M(\eta) + b^2 \end{aligned}$$

átalakítást.

Írjuk fel $S(a,b)$-nek az $a$ és a $b$ szerinti parciális deriváltjait, és állapítsuk meg, hogy mely értékekre lesznek ezek zérusok:

$$\partial_a S(a,b) = -2M(\xi\cdot\eta) + 2a\cdot M\left(\eta^2\right) + 2b\cdot M(\eta) = 0,$$
$$\partial_b S(a,b) = -2\cdot M(\xi) + 2a\cdot M(\eta) + 2b = 0.$$

Ebből átrendezéssel adódnak az úgynevezett *normálegyenletek*:

$$a\cdot M\left(\eta^2\right) + b\cdot M(\eta) = M(\xi\cdot\eta), \quad a\cdot M(\eta) + b = M(\xi).$$

Ezek minden nehézség nélkül megoldhatók és a megoldás

$$a = R(\xi,\eta)\cdot\frac{D(\xi)}{D(\eta)}, \quad b = M(\xi) - R(\xi,\eta)\cdot\frac{D(\xi)}{D(\eta)}\cdot M(\eta)$$

alakba írható. Így a keresett egyenes egyenlete

$$\begin{aligned} x &= a\cdot y + b \\ &= R(\xi,\eta)\cdot\frac{D(\xi)}{D(\eta)}\cdot y + \left(M(\xi) - R(\xi,\eta)\cdot\frac{D(\xi)}{D(\eta)}\cdot M(\eta)\right), \end{aligned}$$

amit $\xi$-nek az $\eta$-ra vonatkoztatott *regressziós egyenesének* nevezünk. Hogy ez a függvény valóban a kívánt minimum tulajdonsággal rendelkezik, az a

$$\begin{pmatrix} \partial_a^2 S(a,b) & \partial_b\partial_a S(a,b) \\ \partial_b\partial_a S(a,b) & \partial_b^2 S(a,b) \end{pmatrix}$$

Hesse-mátrix pozitív definitségéből következik, amit a

$$\partial_b^2 S(a,b) = 2 > 0$$

és a

$$\partial_a^2 S(a,b)\cdot\partial_b^2 S(a,b) - \left(\partial_b\partial_a S(a,b)\right)^2 = 4\cdot D^2(\eta) > 0$$

egyenlőtlenségek igazolnak.

Az $\eta$-nak a $\xi$-re vonatkoztatott regressziós egyenese ezzel analóg módon határozható meg és alakja a következő:

$$y = a \cdot x + b$$
$$= R(\xi,\eta) \cdot \frac{D(\eta)}{D(\xi)} \cdot x + \left( M(\eta) - R(\xi,\eta) \cdot \frac{D(\eta)}{D(\xi)} \cdot M(\xi) \right).$$

**1. példa.** (Feltételes eloszlások.) Egy előadás látogatóinak a száma 0-tól $N$-ig akárhány személy lehet. Tegyük fel, hogy ezek közül minden szám egyenlő valószínűséggel fordul elő. Előző tapasztalatokból tudjuk, hogy a látogatók 80%-a nő. Állapítsuk meg, mekkora lesz egy adott előadáson a megjelent nők számának várható értéke!

**Megoldás.** Jelentse $\eta$ a megjelent látogatók számát, $\xi$ pedig az előadáson megjelent nők számát. Az $M(\xi)$-t kell meghatároznunk. Ezt most az

$$M(\xi) = M(M(\xi|\eta))$$

képlet felhasználásával fogjuk elvégezni.
Az $\{\eta = n\}$ feltétel mellett $\xi$ binomiális eloszlású, így

$$M(\xi|\eta = n) = n \cdot p = n \cdot 0{,}8.$$

Ezt felhasználva a keresett várható érték

$$M(\xi) = M(M(\xi|\eta)) = \sum_{n=0}^{N} M(\xi|\eta = n) \cdot P(\eta = n)$$
$$= \sum_{n=0}^{N} (n \cdot 0{,}8) \cdot \left( \frac{1}{N+1} \right) = \frac{0{,}8}{N+1} \cdot \frac{(N+1)\cdot(N+2)}{2} = 0{,}4 \cdot (N+2).$$

**2. példa.** (Feltételes eloszlások.) Számítsuk ki $\xi$-nek az $\{\eta = y\}$ feltételre vonatkozó feltételes várható értékét, ha a $\xi$ és $\eta$ együttes sűrűségfüggvénye

$$f(x,y) = \begin{cases} \frac{4}{5}(x+3y) \cdot e^{-x-2y}, & \text{ha } x > 0,\, y > 0, \\ 0, & \text{egyébként.} \end{cases}$$

**Megoldás.** η perem-sűrűségfüggvénye:

$$f_2(y)=\int_{-\infty}^{\infty} f(x,y)dx=\begin{cases}\int_0^{\infty}\frac{4}{5}(x+3y)\cdot e^{-x-2y}dx, & \text{ha } y>0,\\ 0, & \text{egyébként.}\end{cases}$$

$$=\begin{cases}\frac{4}{5}(1+3y)\cdot e^{-2y}, & \text{ha } y>0,\\ 0, & \text{egyébként.}\end{cases}$$

ξ-nek az $\{\eta=y\}$ feltételre vonatkozó feltételes sűrűségfüggvénye:

$$f(x|y)=\frac{f(x,y)}{f_2(y)}=\begin{cases}\frac{x+3y}{1+3y}\cdot e^{-x}, & \text{ha } x>0,\ y>0,\\ 0, & \text{egyébként.}\end{cases}$$

Így a keresett várható érték:

$$M(\xi|\eta=y)=\int_0^{\infty} x\cdot f(x|y)dx=\int_0^{\infty} x\cdot\left(\frac{x+3y}{1+3y}\cdot e^{-x}\right)dx$$

$$=\frac{2+3y}{1+3y},\quad \text{ha } y>0.$$

**3. példa.** (Korrelációs együttható.) Egy üzem $A$ és $B$ jelű termékeket gyárt. Ezeket I., II. és III. osztályú minősítéssel látták el. Egy alkalommal a késztáruraktárban e termékek a következő megoszlásban szerepeltek:

| | $I.o.$ | $II.o.$ | $III.o.$ |
|---|---|---|---|
| $A$ | 500 | 200 | 100 |
| $B$ | 650 | 300 | 50 |

E halmazból egy terméket veszünk ki véletlenszerűen. $\xi=0$ jelentse azt, hogy $A$ jelű, $\xi=1$ pedig, hogy $B$ jelű a termék. $\eta=1$, $\eta=2$, illetve $\eta=3$ jelentse azt, hogy a kivett termék I., II., illetve III. osztályú.

*a)* Készítsük el az eloszlás táblázatát!

*b)* Számítsuk ki ξ és η korrelációs együtthatóját!

**Megoldás.**

*a)* Az eloszlás táblázatát egyszerű számolással kapjuk, és a peremeloszlásokat is felírtuk:

| $\xi \setminus \eta$ | $y_1 = 1$ | $y_2 = 2$ | $y_3 = 3$ | $\xi$ |
|---|---|---|---|---|
| $x_1 = 0$ | $p_{11} = \frac{50}{180}$ | $p_{12} = \frac{20}{180}$ | $p_{13} = \frac{10}{180}$ | $p_{1*} = \frac{80}{180}$ |
| $x_2 = 1$ | $p_{21} = \frac{65}{180}$ | $p_{22} = \frac{30}{180}$ | $p_{23} = \frac{5}{180}$ | $p_{2*} = \frac{100}{180}$ |
| $\eta$ | $p_{*1} = \frac{115}{180}$ | $p_{*2} = \frac{50}{180}$ | $p_{*3} = \frac{15}{180}$ | 1 |

*b)* A korrelációs együttható kiszámításához szükséges adatok:

$$M(\xi \cdot \eta) = \sum_{i=1}^{2} \sum_{k=1}^{3} (x_i \cdot y_k) \cdot p_{ik} = \frac{7}{9},$$

$$M(\xi) = \sum_{i=1}^{2} x_i \cdot p_{i*} = \frac{5}{9},$$

$$M(\eta) = \sum_{k=1}^{3} y_k \cdot p_{*k} = \frac{13}{9},$$

$$D(\xi) = \sqrt{\sum_{i=1}^{2} x_i^2 \cdot p_{i*} - M^2(\xi)} = \frac{2\sqrt{5}}{9},$$

$$D(\eta) = \sqrt{\sum_{k=1}^{3} y_k^2 \cdot p_{*k} - M^2(\eta)} = \frac{1}{9} \cdot \sqrt{\frac{67}{2}}.$$

A korrelációs együttható ezek után:

$$R(\xi, \eta) = \frac{M(\xi \cdot \eta) - M(\xi) \cdot M(\eta)}{D(\xi) \cdot D(\eta)} = -\sqrt{\frac{2}{335}}.$$

**4. példa.** (Korrelációs együttható.) Az alábbi táblázat a ξ és η valószínűségi változók együttes valószínűség-eloszlását és peremeloszlását tartalmazza. Mutassuk meg, hogy ξ és η nem függetlenek, bár korrelálatlanok!

| $\xi \setminus \eta$ | $y_1 = -1$ | $y_2 = 0$ | $\xi$ |
|---|---|---|---|
| $x_1 = 0$ | $p_{11} = 0$ | $p_{12} = \frac{1}{3}$ | $p_{1*} = \frac{1}{3}$ |
| $x_2 = 1$ | $p_{21} = \frac{1}{3}$ | $p_{22} = 0$ | $p_{2*} = \frac{1}{3}$ |
| $x_3 = 2$ | $p_{31} = 0$ | $p_{32} = \frac{1}{3}$ | $p_{3*} = \frac{1}{3}$ |
| $\eta$ | $p_{*1} = \frac{1}{3}$ | $p_{*2} = \frac{2}{3}$ | 1 |

**Megoldás.**
ξ és η nem függetlenek, ugyanis például

$$P(\xi = 0, \eta = -1) = p_{11} = 0 \neq$$
$$\frac{1}{3} \cdot \frac{1}{3} = p_{1*} \cdot p_{*1} = P(\xi = 0) \cdot P(\eta = -1).$$

A korrelációs együttható kiszámításához szükséges adatok:

$$M(\xi \cdot \eta) = \sum_{i=1}^{3} \sum_{k=1}^{2} (x_i \cdot y_k) \cdot p_{ik} = -\frac{1}{3},$$

$$M(\xi) = \sum_{i=1}^{3} x_i \cdot p_i = 1, \; M(\eta) = \sum_{k=1}^{2} y_k \cdot q_k = -\frac{1}{3},$$

A kovariancia ezek után:

$$\operatorname{cov}(\xi, \eta) = M(\xi \cdot \eta) - M(\xi) \cdot M(\eta) = -\frac{1}{3} - 1 \cdot \left(-\frac{1}{3}\right) = 0,$$

tehát ξ és η valóban korrelálatlanok.
Megmutatható azonban, hogy $\eta = \xi^2 - 2 \cdot \xi$.

**5. példa.** (Regressziós függvény és egyenes.) Az alábbi táblázat a $\xi$ és $\eta$ valószínűségi változók együttes valószínűség-eloszlását és peremeloszlását tartalmazza.

| $\xi \setminus \eta$ | $y_1 = 0$ | $y_2 = 1$ | $\xi$ |
|---|---|---|---|
| $x_1 = 0$ | $p_{11} = \frac{1}{12}$ | $p_{12} = \frac{1}{12}$ | $p_{1*} = \frac{2}{12}$ |
| $x_2 = 1$ | $p_{21} = \frac{1}{12}$ | $p_{22} = \frac{3}{12}$ | $p_{2*} = \frac{4}{12}$ |
| $x_3 = 2$ | $p_{31} = \frac{2}{12}$ | $p_{32} = \frac{4}{12}$ | $p_{3*} = \frac{6}{12}$ |
| $\eta$ | $p_{*1} = \frac{4}{12}$ | $p_{*2} = \frac{8}{12}$ | 1 |

*a)* Határozzuk meg $\xi$-nek az $\eta$-ra, illetve $\eta$-nak a $\xi$-re vonatkozó regressziós függvényét!

*b)* Számítsuk ki $\xi$-nek az $\eta$-ra, illetve $\eta$-nak a $\xi$-re vonatkoztatott regressziós egyenesét!

**Megoldás.**

*a)* A táblázatból $\xi$ feltételes eloszlásaira a következő értékeket kapjuk:

$$P(\xi = 0 | \eta = 0) = \frac{p_{11}}{p_{*1}} = \frac{1}{4}, \quad P(\xi = 0 | \eta = 1) = \frac{p_{12}}{p_{*2}} = \frac{1}{8},$$
$$P(\xi = 1 | \eta = 0) = \frac{p_{21}}{p_{*1}} = \frac{1}{4}, \quad P(\xi = 1 | \eta = 1) = \frac{p_{22}}{p_{*2}} = \frac{3}{8},$$
$$P(\xi = 2 | \eta = 0) = \frac{p_{31}}{p_{*1}} = \frac{1}{2}, \quad P(\xi = 2 | \eta = 1) = \frac{p_{32}}{p_{*2}} = \frac{1}{2}.$$

A feltételes várható értékek:

$$M(\xi | \eta = 0) = \sum_{i=1}^{3} x_i \cdot \frac{p_{i1}}{p_{*1}} = \frac{5}{4}, \quad M(\xi | \eta = 1) = \sum_{i=1}^{3} x_i \cdot \frac{p_{i2}}{p_{*2}} = \frac{11}{8}.$$

$\xi$-nek az $\eta$-ra vonatkozó regressziós függvénye így az alábbi táblázat alakban adott függvény:

| $y_k$ | 0 | 1 |
|---|---|---|
| $r(y_k) = M(\xi \mid \eta = y_k)$ | $\frac{5}{4}$ | $\frac{11}{8}$ |

A táblázatból $\eta$ feltételes eloszlásaira a következő értékeket kapjuk:

$$P(\eta=0|\xi=0)=\frac{p_{11}}{p_{1*}}=\frac{1}{2},\quad P(\eta=1|\xi=0)=\frac{p_{12}}{p_{1*}}=\frac{1}{2},$$
$$P(\eta=0|\xi=1)=\frac{p_{21}}{p_{2*}}=\frac{1}{4},\quad P(\eta=1|\xi=1)=\frac{p_{22}}{p_{2*}}=\frac{3}{4},$$
$$P(\eta=0|\xi=2)=\frac{p_{31}}{p_{3*}}=\frac{1}{3},\quad P(\eta=1|\xi=2)=\frac{p_{32}}{p_{3*}}=\frac{2}{3}.$$

A feltételes várható értékek:

$$M(\eta|\xi=0)=\sum_{k=1}^{2} y_k\cdot\frac{p_{1k}}{p_{1*}}=\frac{1}{2},$$
$$M(\eta|\xi=1)=\sum_{k=1}^{2} y_k\cdot\frac{p_{2k}}{p_{2*}}=\frac{3}{4},$$
$$M(\eta|\xi=2)=\sum_{k=1}^{2} y_k\cdot\frac{p_{3k}}{p_{3*}}=\frac{2}{3}.$$

$\eta$-nak a $\xi$-re vonatkozó regressziós függvénye így az alábbi táblázat alakban adott függvény:

| $x_i$ | 0 | 1 | 2 |
|---|---|---|---|
| $r(x_i)=M(\eta\|\xi=x_i)$ | $\frac{1}{2}$ | $\frac{3}{4}$ | $\frac{2}{3}$ |

*b)* A táblázat alapján:

$$M(\xi\cdot\eta)=\frac{11}{12},\ M(\xi)=\frac{4}{3},\ M\left(\xi^2\right)=\frac{14}{16},\ D(\xi)=\frac{\sqrt{5}}{3},$$
$$M(\eta)=\frac{2}{3},\ M\left(\eta^2\right)=\frac{2}{3},\ D(\eta)=\frac{\sqrt{2}}{3}.$$

$\xi$ és $\eta$ korrelációs együtthatója ezért

$$R(\xi,\eta)=\frac{M(\xi\cdot\eta)-M(\xi)\cdot M(\eta)}{D(\xi)\cdot D(\eta)}=\frac{1}{4\sqrt{10}}.$$

$\xi$-nek $\eta$-ra vonatkoztatott regressziós egyenese így

$$x = a \cdot y + b$$

$$= R(\xi,\eta) \cdot \frac{D(\xi)}{D(\eta)} \cdot y + \left( M(\xi) - R(\xi,\eta) \cdot \frac{D(\xi)}{D(\eta)} \cdot M(\eta) \right) = \frac{1}{8} y + \frac{15}{12},$$

$\eta$-nak $\xi$-re vonatkoztatott regressziós egyenese pedig

$$y = c \cdot x + d$$

$$= R(\xi,\eta) \cdot \frac{D(\eta)}{D(\xi)} \cdot x + \left( M(\eta) - R(\xi,\eta) \cdot \frac{D(\eta)}{D(\xi)} \cdot M(\xi) \right) = \frac{1}{20} x + \frac{3}{5}.$$

## 2.23. Folytonos valószínűségi változók egyszerűbb függvényeinek eloszlása

**2.23.1. tétel.** (Folytonos valószínűségi változók egyszerűbb függvényeinek eloszlása.) $\xi_1$ és a $\xi_2$ legyenek független folytonos eloszlású valószínűségi változók. A sűrűségfüggvényeiket jelöljük $f_i$-vel $(i = 1, 2)$. Ekkor

1. az $\eta = \xi_1 + \xi_2$ valószínűségi változó sűrűségfüggvénye

$$g(x) = \int_{-\infty}^{\infty} f_1(x - t) \cdot f_2(t) dt,$$

2. az $\eta = \xi_1 - \xi_2$ valószínűségi változó sűrűségfüggvénye

$$g(x) = \int_{-\infty}^{\infty} f_1(x + t) \cdot f_2(t) dt,$$

3. az $\eta = \xi_1 \cdot \xi_2$ valószínűségi változó sűrűségfüggvénye

$$g(x) = \int_{-\infty}^{\infty} \frac{1}{|t|} \cdot f_1\left(\frac{x}{t}\right) \cdot f_2(t) dt,$$

4. az $\eta = \frac{\xi_1}{\xi_2}$ valószínűségi változó sűrűségfüggvénye pedig

$$g(x) = \int_{-\infty}^{\infty} |t| \cdot f_1(x \cdot t) \cdot f_2(t) dt.$$

A következő szakaszban ismertetésre kerülő eloszlások sűrűségfüggvényei ennek a tételnek a valószínűségi változó transzformáltjára vonatkozó korábbi tétellel való együttes alkalmazásával vezethetők le.

## 2.24. Nevezetes többváltozós eloszlások

Ebben a fejezetben a matematikai statisztikában leggyakrabban használt többváltozós eloszlásokat ismertetjük. Az eloszlások sűrűségfüggvényeinek ismertetésénél szükségünk van az Euler-féle gamma függvényre, amelynek definíciója és fontosabb tulajdonságai a következők:

**Definíció.** Az *Euler-féle gamma-függvényt* a

$$\Gamma(x) = \int_0^\infty e^{-t} \cdot t^{x-1}\,dt \qquad (x \in (0,\infty))$$

képlettel definiáljuk.

A gamma-függvény főbb tulajdonságai:

$a)\ \Gamma(x+1) = x \cdot \Gamma(x) \quad (x \in (0,\infty))$
$b)\ \Gamma(n+1) = n!\ (n \in \mathbf{N})$
$c)\ \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

**Definíció.** Ha $\xi_1, \xi_2, \ldots, \xi_n$ független $N(0,1)$ standard normális eloszlású valószínűségi változók, akkor a

$$\chi^2 := \xi_1^2 + \xi_2^2 + \ldots + \xi_n^2$$

valószínűségi változó eloszlását *n-szabadságfokú* $\chi^2$*-eloszlásúnak* (olv.: khi-négyzet eloszlásúnak) nevezzük.

**2.24.1. tétel.** Az $n$-szabadságfokú $\chi^2$-eloszlás sűrűségfüggvénye, várható értéke és a szórása:

$$f_{\chi^2}(x)=\begin{cases}\dfrac{x^{\frac{n}{2}-1}\cdot e^{-\frac{x}{2}}}{\left(\frac{1}{2}\right)^{\frac{n}{2}}\cdot\Gamma\left(\frac{n}{2}\right)}, & \text{ha} \quad x>0\\ 0, & \text{ha} \quad x\le 0\end{cases},$$

$$M\left(\chi^2\right)=n,\ \ D\left(\chi^2\right)=\sqrt{2n}.$$

A tétel bizonyítását ld. pl. Bolla Marianna és Krámli András könyvében.

**Definíció.** Ha $\xi_1,\xi_2,\ldots,\xi_n$ és $\eta$ független $N(0,1)$ standard normális eloszlású valószínűségi változók, akkor a

$$t:=\frac{\eta}{\sqrt{\dfrac{\xi_1^2+\xi_2^2+\ldots+\xi_n^2}{n}}}$$

valószínűségi változó *eloszlását n-szabadságfokú t-eloszlásúnak (Student-eloszlásúnak)* nevezzük.

Másképp megfogalmazva, ha $\xi$ *n*-szabadságfokú $\chi^2$-eloszlású, $\eta$ pedig $N(0,1)$ standard normális eloszlású független valószínűségi változók, akkor $t:=\dfrac{\eta}{\sqrt{\xi}}\cdot\sqrt{n}$ *n*-szabadságfokú Student-eloszlású valószínűségi változó.

**2.24.2. tétel.** Az $n$-szabadságfokú $t$-eloszlás sűrűségfüggvénye, a várható értéke és a szórása:

$$f_t(x)=\frac{1}{\sqrt{n\pi}}\cdot\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\cdot\frac{1}{\left(\frac{x^2}{n}+1\right)^{\frac{n+1}{2}}},$$

$$M(t) = \begin{cases} \text{nem létezik}, & \text{ha } n = 1, \\ 0, & \text{ha } n \geq 2, \end{cases}$$

$$D(t) = \begin{cases} \text{nem létezik}, & \text{ha } n = 1, 2, \\ \sqrt{\dfrac{n}{n-2}}, & \text{ha } n \geq 3. \end{cases}.$$

**Definíció.** Ha $\xi_1, \xi_2, \ldots, \xi_m$ és $\eta_1, \eta_2, \ldots, \eta_n$ független, $N(0,1)$ standard normális eloszlású valószínűségi változók, akkor az

$$F = \frac{\xi_1^2 + \xi_2^2 + \ldots + \xi_m^2}{m} \bigg/ \frac{\eta_1^2 + \eta_2^2 + \ldots + \eta_n^2}{n}$$

valószínűségi változó eloszlását $(m,n)$*-szabadságfokú F-eloszlásúnak* nevezzük.

Másképp megfogalmazva, ha $\xi$ és $\eta$ független, *m*- ill. *n*-szabadságfokú $\chi^2$-eloszlású valószínűségi változók, akkor $F := \dfrac{\xi}{\eta} \cdot \dfrac{n}{m}$ (*m*,*n*)-szabadságfokú *F*-eloszlású valószínűségi változó.

**2.24.3. tétel.** Az $(m,n)$-szabadságfokú *F*-eloszlás sűrűségfüggvénye, a várható értéke és a szórása:

$$f_F(x) = \begin{cases} \left(\dfrac{m}{n}\right)^{\frac{m}{2}} \cdot \dfrac{\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right) \cdot \Gamma\left(\dfrac{n}{2}\right)} \cdot \dfrac{x^{\frac{m}{2}-1}}{\left(1 + \dfrac{m}{n} \cdot x\right)^{\frac{m+n}{2}}}, & \text{ha} \quad x \geq 0, \\ 0, & \text{ha} \quad x < 0, \end{cases}$$

$$M(F) = \begin{cases} \text{nem létezik}, & \text{ha } n = 1, 2, \\ \dfrac{n}{n-2}, & \text{ha } n \geq 3, \end{cases}$$

$$D(F)=\begin{cases}\text{nem létezik,} & \text{ha } n=1,2,3,4\\ \sqrt{\dfrac{2\cdot n^2\cdot(m+n-2)}{m\cdot(n-2)^2\cdot(n-4)}}, & \text{ha } n\geq 5.\end{cases}$$

A várható értékek különbségére vonatkozó statisztikai próbáknál alapvető fontosságú a normális eloszlású valószínűségi változókra vonatkozó alábbi tétel.

**2.24.4. tétel.** Ha $\xi_1,\xi_2,\ldots,\xi_m$ $N(m_1,\sigma_1)$, $\eta_1,\eta_2,\ldots,\eta_n$ pedig $N(m_2,\sigma_2)$ normális eloszlású független valószínűségi változók, akkor

$$\frac{\xi_1+\xi_2+\ldots+\xi_m}{m}$$

$N\left(m_1,\dfrac{\sigma_1}{\sqrt{n}}\right)$ normális eloszlású,

$$\frac{\xi_1+\xi_2+\ldots+\xi_m}{m}-\frac{\eta_1+\eta_2+\ldots+\eta_n}{n}$$

pedig $N\left(m_1-m_2,\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}\right)$ normális eloszlású valószínűségi változó.

## 2.25. A centrális (központi) határeloszlás tétel

A normális eloszlás a gyakorlat egyik legfontosabb eloszlása. Ezt az állítást támasztja alá a következő, ún. *centrális határeloszlás tétel.*

**2.25.1. tétel.** (Centrális határeloszlás tétel.) Ha $\xi_1,\xi_2,\ldots,\xi_n,\ldots$ azonos eloszlású és véges szórású független valószínűségi változók egy sorozata, $M(\xi_i)=m$, $D(\xi_i)=\sigma$ $(i=1,2,\ldots)$, akkor a 0 várható értékű és 1 szórású

$$\eta_n=\frac{\xi_1+\xi_2+\ldots+\xi_n-n\cdot m}{\sqrt{n}\cdot\sigma}\qquad(n=1,2,\ldots)$$

valószínűségi változók sorozata *aszimptotikusan standard normális eloszlású,* azaz tetszőleges $x\in(-\infty,+\infty)$ szám esetén

$$\lim_{n\to\infty} P(\eta_n < x) = \Phi(x),$$

ahol $\Phi(x)$ a standard normális eloszlás eloszlásfüggvényét jelöli.

**1. példa.** (Folytonos valószínűségi. változók egyszerűbb függvényeinek eloszlása) $\xi_1$ és a $\xi_2$ legyenek független $N(0,1)$ standard normális eloszlású valószínűségi változók. Határozzuk meg az $\eta = \xi_1 + \xi_2$ valószínűségi változó sűrűségfüggvényét!

**Megoldás.** Az $\eta = \xi_1 + \xi_2$ valószínűségi változó sűrűségfüggvényét a

$$g(x) = \int_{-\infty}^{\infty} f_1(x-t) \cdot f_2(t) dt$$

képlet alkalmazásával határozhatjuk meg. Mivel $\xi_1$ és $\xi_2$ sűrűségfüggvényei egyaránt

$$f_i(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad (i = 1,2),$$

azért $g$-t az alábbi alakba írhatjuk:

$$g(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2+t^2}{2}} dt =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2+2\cdot t^2-2\cdot x\cdot t}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{4}-\left(t-\frac{x}{2}\right)^2} dt =$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{4}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\left(t-\frac{x}{2}\right)^2} dt =$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{4}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} \cdot \frac{1}{\sqrt{2}} ds =$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{x^2}{2\cdot\left(\sqrt{2}\right)^2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{x^2}{2\cdot\left(\sqrt{2}\right)^2}}$$

ahol helyettesítéses integrálást alkalmaztunk ($s := \sqrt{2} \cdot \left(t - \frac{x}{2}\right)$ helyettesítéssel), és felhasználtuk, hogy

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds = \int_{-\infty}^{\infty} f(s) ds = \lim_{+\infty} F = 1.$$

A kapott alakból az következik, hogy $\eta$ egy $N\left(0, \sqrt{2}\right)$ normális eloszlású valószínűségi változó lesz.

# 3. Matematikai statisztika

**A statisztika tárgya:** Véletlen tömegjelenségek, az úgynevezett **alapsokaság**ok viselkedésének leírása, esetenként döntési eljárások adása véges sok kísérlet eredményének felhasználásával.

## 3.1. Statisztikai minta, statisztikai függvények

A statisztikában nagyszámú (gyakorlatilag végtelen sok) egyedből álló alapsokaság viselkedésére annak viszonylag kevés egyedének vizsgálata alapján kívánunk következtetni. Ez a kevés egyed alkotja a **mintát.**

A **matematikai statisztiká**ban a valószínűség-számítás eszközeit alkalmazzuk. Az alapsokaságot egy **valószínűségi változó**val azonosítjuk és az alapsokaság viselkedése helyett e valószínűségi változó **eloszlás**áról beszélünk.

**Definíció. Statisztikai mintá**n $n$ számú független $\xi_1, \xi_2, \ldots, \xi_n$, a megfigyelt $\xi$ valószínűségi változóval megegyező eloszlású, **valószínűségi (mintavételi) változó** összességét értjük.

**Definíció.** A $\xi_1, \xi_2, \ldots, \xi_n$ mintavételi változók tetszőleges $\hat{\alpha}_n : \mathbb{R}^n \to \mathbb{R}$ függvényét **statisztikai függvény**nek, vagy röviden **statisztiká**nak nevezzük.

A statisztikák maguk is valószínűségi változók.

Az alábbiakban a legfontosabb statisztikai függvényeket soroljuk fel.

**Definíció.** A mintaelemek számtani középértékét **mintaátlag**nak, vagy **empirikus közép**nek nevezzük és $\hat{m}_n$ -pal jelöljük, azaz

$$\hat{m}_n := \hat{m}_n(\xi) := \hat{m}_n(\xi_1, \xi_2, \ldots, \xi_n) := \frac{1}{n} \cdot \sum_{i=1}^{n} \xi_i .$$

**Definíció.** A mintaelemek közepüktől való eltérésnégyzeteinek átlagát a **minta szórásnégyzeté**nek, vagy **empirikus szórásnégyzeté**nek nevezzük és $\hat{\sigma}_n^2$-tel jelöljük, azaz

$$\hat{\sigma}_n^2 := \hat{\sigma}_n^2(\xi) := \hat{\sigma}_n^2(\xi_1, \xi_2, \ldots, \xi_n) := \frac{1}{n} \cdot \sum_{i=1}^{n} (\xi_i - \hat{m}_n)^2.$$

A **korrigált tapasztalati szórásnégyzet**et az

$$\hat{s}_n^2 := \frac{n}{n-1} \hat{\sigma}_n^2$$

formulával definiáljuk.

Könnyen igazolható, hogy az empirikus szórásnégyzet a

$$\hat{\sigma}_n^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} \xi_i^2 - \left( \frac{1}{n} \cdot \sum_{i=1}^{n} \xi_i \right)^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} \xi_i^2 - \hat{m}_n^2$$

ekvivalens alakban írható. Ez a Steiner-tétel (2.9.8. tétel) következménye, $c := 0$ választással.

**Definíció.** A $\xi_1, \xi_2, \ldots, \xi_n$ mintavételi változók **nagyság szerint növekvően rendezett értékei** közül az $i$-ediket $\xi_i^*$-vel $(i = 1, \ldots, n)$ jelöljük.

**Definíció.** A minta legkisebb és legnagyobb elemének számtani közepét, azaz $\dfrac{\xi_1^* + \xi_n^*}{2}$-t a **minta középpontjá**nak nevezzük.

**Definíció.** A minta legnagyobb és legkisebb elemének különbségét, azaz $\xi_n^* - \xi_1^*$-t a **minta terjedelmé**nek nevezzük.

**Definíció.** A minta (empirikus) **mediánja**

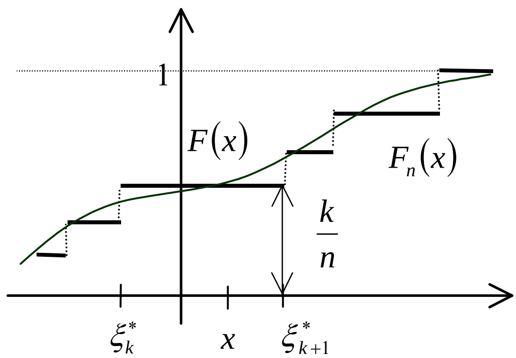
$$\xi_m^*, \text{ ha } n = 2m-1, \text{ és } \frac{\xi_m^* + \xi_{m+1}^*}{2}, \text{ ha } n = 2m,$$

azaz páratlan elemszám esetén a középső érték, páros elemszám esetén pedig a két középső érték átlaga.

**Definíció.** A $\xi_1, \xi_2, \ldots, \xi_n$ minta **empirikus eloszlásfüggvénye:**

$$F_n(x) := \begin{cases} 0, & \text{ha } x < \xi_1^*, \\ \dfrac{k}{n}, & \text{ha } \xi_k^* \le x < \xi_{k+1}^* \quad (k = 1,2,\ldots,n-1), \\ 1, & \text{ha } x \ge \xi_n^*, \end{cases}$$

ahol $\xi_i^*$ a minta növekvő nagyság szerint rendezett elemei közül az $i$-edik.




**Megjegyzések.**

1. Az empirikus eloszlásfüggvény egy lépcsősfüggvény, minden $\xi_i^*$ helyen $\frac{1}{n}$ nagyságú ugrással.

2. Az $F_n$ empirikus eloszlásfüggvény minden egyes $x$ helyen maga is valószínűségi változó, az értéke pedig a $\{\xi < x\}$ esemény mintabeli relatív gyakorisága, azaz

$$F_n(x) = \frac{1}{n} \sum_{\xi_i < x} 1.$$

A nagy számok Bernoulli-törvénye szerint elég nagy $n$ elemszámú minta esetén az esemény relatív gyakorisága az esemény valószínűséget tetszőleges $\varepsilon > 0$ esetén

$$P\left(\left| F_n(x) - P(\xi < x)\right| < \varepsilon\right) =$$

$$= 1 - P\left(\left| F_n(x) - P(\xi < x)\right| \geq \varepsilon\right) > 1 - \frac{1}{4n\varepsilon^2}$$

valószínűséggel közelíti meg. Mivel $F(x) = P(\xi < x)$, ezért

$$\lim_{n\to\infty} P\left(\left| F_n(x) - F(x)\right| < \varepsilon\right) = 1,$$

azaz tetszőleges $\varepsilon > 0$ hibakorlát esetén, az empirikus eloszlásfüggvények $x$ helyhez tartozó értékeinek $F_n(x)$ $(n = 1,2,\ldots)$ sorozatának határértéke a vizsgált $\xi$ valószínűségi változó elméleti eloszlásfüggvényének $x$ helyen felvett értékét 1 valószínűséggel $\varepsilon$-nál kisebb hibával közelíti meg. De ennél lényegesen több is igaz, mint azt az alábbi tétel mutatja.

**3.1.1. tétel.** (Glivenko és Cantelli tétele) Az $F_n$ empirikus eloszlásfüggvény az egész számegyenesen 1 valószínűséggel, egyenletesen konvergál az $F$ elméleti eloszlásfüggvényhez, azaz

$$P\left(\lim_{n\to\infty}\left(\sup_{-\infty<x<\infty}\left| F_n(x) - F(x)\right|\right) = 0\right) = 1.$$

Glivenko és Cantelli tétele teszi jogossá a mintavételen alapuló statisztikai következtetéseket. Például azt, hogy a minta várható értékét és szórását az elméleti várható érték és szórás közelítő értékének, statisztikai becslésének tekintsük.

**Definíció.** Legyen $\xi_1, \xi_2, \ldots, \xi_n$ egy adott $n$ elemű minta, az $a$, $b$ számokra pedig teljesüljön az $a \le \xi_1^*$ és a $\xi_n^* < b$ feltétel. Osszuk fel az $[a,b]$ intervallumot $m$ részintervallumra (osztályra) az

$$a = x_0 < x_1 < \ldots < x_{m-1} < x_m = b$$

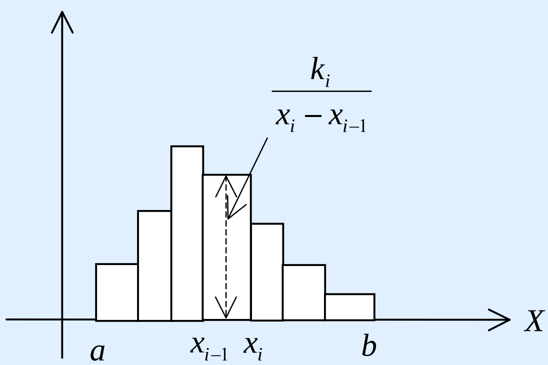
osztópontok segítségével. Az egyes $[x_{i-1}, x_i)$ részintervallumba eső mintaelemek számát jelöljük $k_i$-vel $(i = 1, 2, \ldots, m)$.

A **gyakorisági hisztogram**ot úgy kapjuk, hogy az $[x_{i-1}, x_i)$ intervallumra

$$\frac{k_i}{x_i - x_{i-1}}$$

magasságú téglalapot rajzolunk $(i = 1,2, \ldots, m)$.
(Ekkor a téglalapok által lefedett terület:

$$\sum_{i=1}^{m} \frac{k_i}{x_i - x_{i-1}} \cdot (x_i - x_{i-1}) = \sum_{i=1}^{m} k_i = n.)$$



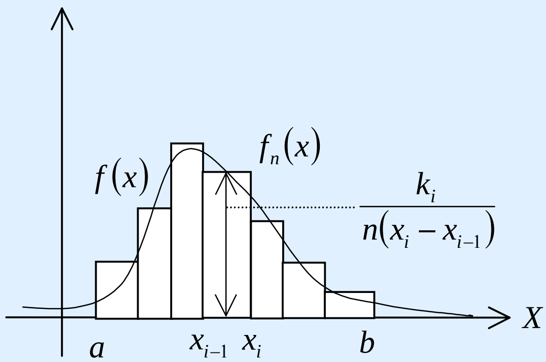

A **sűrűséghisztogram**ot úgy kapjuk, hogy az $[x_{i-1}, x_i)$ intervallumra

$$\frac{k_i}{n \cdot (x_i - x_{i-1})}$$

magasságú téglalapot rajzolunk $(i = 1,2, \ldots, m)$.

(Ekkor a téglalapok által lefedett terület:

$$\sum_{i=1}^{m} \frac{k_i}{n\cdot(x_i - x_{i-1})}\cdot(x_i - x_{i-1}) = \frac{1}{n}\sum_{i=1}^{m} k_i = 1.)$$




Szokás a sűrűséghisztogramot az **empirikus sűrűségfüggvény grafikonjá**nak is nevezni és $f_n$ -nel jelölni.

Az $[x_{i-1}, x_i)$ intervallumon

$$f_n(x) = \frac{k_i}{n\cdot(x_i - x_{i-1})} = \frac{F_n(x_i) - F_n(x_{i-1})}{x_i - x_{i-1}} \quad (x_{i-1} \le x < x_i,\ i = 1,\dots,m)$$

ahol $F_n(x_i) - F_n(x_{i-1})$ az $F_n$ empirikus eloszlásfüggvény növekménye az $[x_{i-1}, x_i)$ intervallumon. A két függvény közti kapcsolat analóg az elméleti sűrűségfüggvényre ismert $f = F'$ relációval, csak itt a derivált helyett a differenciahányadost kell venni.
A fenti módon $m$ osztályba gyűjtött minta esetén a mintaátlagot

$$\hat{m}_n \approx \frac{1}{n}\sum_{i=1}^{m} \frac{x_i + x_{i-1}}{2}\cdot k_i,$$

a **minta szórásnégyzeté**t pedig

$$\hat{\sigma}_n^2 \approx \frac{1}{n}\cdot\sum_{i=1}^{n}\left(\frac{x_i + x_{i-1}}{2}\right)^2\cdot k_i - \left(\frac{1}{n}\sum_{i=1}^{m}\frac{x_i + x_{i-1}}{2}\cdot k_i\right)^2$$

módon becsüljük.

**1. példa.** A Tisza árhullámaira vonatkozólag Tokajnál az 1903–1971 időszakban a tetőzési értékek az alábbi gyakorisággal estek a feltüntetett intervallumokba (mindegyik év első félévében bekövetkezett árvizeket vizsgálva).

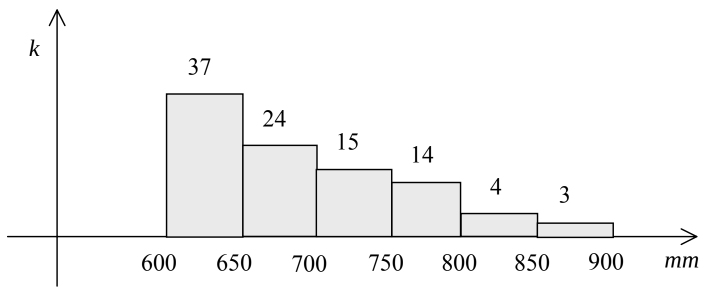
| | |
|---|---|
| 600-650 mm | 37 árhullám |
| 650-700 mm | 24 árhullám |
| 700-750 mm | 15 árhullám |
| 750-800 mm | 14 árhullám |
| 800-850 mm | 4 árhullám |
| 850-900 mm | 3 árhullám |
| összesen: | 97 árhullám |

Készítsük el a tetőzési értékek gyakorisági hisztogramját, a tapasztalati eloszlásfüggvényét, valamint a mintaátlag és minta szórásnégyzetének becslését!

**Megoldás.** Minthogy egyenletes felosztást használunk, az egyes részintervallumok hossza $\Delta_i = x_i - x_{i-1} = 5$ cm. A gyakorisági hisztogram alatti terület most

$$\sum_{i=1}^{m} \frac{k_i}{x_i - x_{i-1}} \cdot (x_i - x_{i-1}) = (74 + 48 + 30 + 28 + 8 + 6) = 97\,,$$

azaz a mintaelemek számával egyenlő.




A gyakorisági hisztogram

Ha a gyakorisági hisztogramban minden ordinátát 97-tel osztjuk, akkor a sűrűség hisztogramot kapjuk.

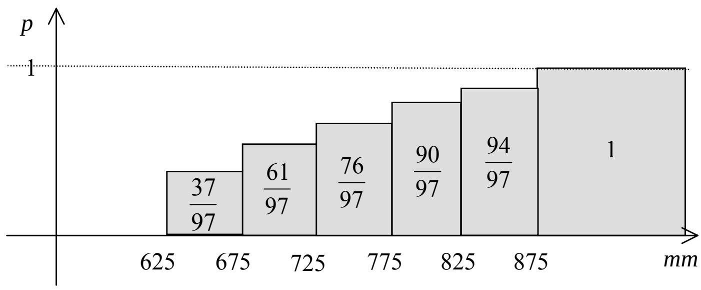
Osztályokban adott gyakoriságok esetén a tapasztalati eloszlásfüggvénynél az ugrásokat az osztályközepeknél jelöljük. Példánkban az osztályok a

$$[600;\ 650),\ [650;700),\ [700;750),\ [750;800),\ [800;850),\ [850;\ 900)$$

intervallumok, az osztályközepek pedig a

$$625,\ 675,\ 725,\ 775,\ 825,\ 875$$

számok.




A tapasztalati eloszlásfüggvény

A mintaátlag becslése

$$\hat{m}_n \approx \frac{1}{n}\sum_{i=1}^{m}\frac{x_i + x_{i-1}}{2}\cdot k_i =$$

$$= \frac{1}{97}\cdot\left(625\cdot 37 + 675\cdot 24 + 725\cdot 15 + 775\cdot 14 + 825\cdot 4 + 875\cdot 3\right) \approx$$

$$\approx 523{,}45.$$

A minta szórásnégyzetének becslése pedig

$$\hat{\sigma}_n^2 \approx \frac{1}{n}\cdot\sum_{i=1}^{n}\left(\frac{x_i + x_{i-1}}{2}\right)^2\cdot k_i - \left(\frac{1}{n}\sum_{i=1}^{m}\frac{x_i + x_{i-1}}{2}\cdot k_i\right)^2 =$$

$$= \frac{1}{97}\cdot\left(\begin{array}{l}(625)^2\cdot 37 + (675)^2\cdot 24 + (725)^2\cdot 15\\ + (775)^2\cdot 14 + (825)^2\cdot 4 + (875)^2\cdot 3\end{array}\right) - (523{,}45)^2 \approx$$

$$\approx 94714{,}1\,.$$

## 3.2. Becsléselméleti alapfogalmak

A becsléselmélet ismert eloszlástípusú valószínűségi változók ismeretlen paramétereire vonatkozó becsléseket és ezen becslések tulajdonságait vizsgálja.

Legyen $\xi$ a megfigyelt valószínűségi változó, és $\alpha$ az eloszlás ismeretlen paramétere. Legyen továbbá $\xi_1, \xi_2, \ldots, \xi_n$ a $\xi$-ből vett $n$-elemű minta. Készítsünk el egy olyan statisztikát, amelyből következtetni lehetet $\alpha$-ra. Legyen egy ilyen $\alpha$-ra vonatkozó becslés az

$$\hat{\alpha}_n = \hat{\alpha}_n(\xi_1, \xi_2, \ldots, \xi_n)$$

statisztika.

**Definíció.** A **becslés torzítatlan,** ha a becslés eloszlásának várható értéke a becsült paraméter, azaz

$$M(\hat{\alpha}_n) = \alpha.$$

**Definíció.** Egy $\hat{\alpha}_{1,n}$ **becslés legalább olyan hatásos,** (illetve **hatásosabb**) mint egy $\hat{\alpha}_{2,n}$ becslés, ha azonos $n$ elemszámú minta esetén, az $\hat{\alpha}_{1,n}$ becslés eloszlásának szórása nem nagyobb, mint az $\hat{\alpha}_{2,n}$-é, azaz

$$D(\hat{\alpha}_{1,n}) \leq D(\hat{\alpha}_{2,n})$$

az $\alpha$ paraméter bármely értékére (és legalább egy értékre a szigorú egyenlőtlenség teljesül).

Itt és a későbbiekben: a statisztika szórását az $\alpha$ valódi paraméterérték melletti eloszlás alapján számítjuk.

**Definíció.** Az $\alpha$ paraméter, adott minta elemszámhoz tartozó legkisebb szórású torzítatlan becslését **hatásos becslésnek** nevezzük, amennyiben létezik ilyen becslés.

Ha létezik egyáltalán hatásos becslés, akkor az egyértelmű is.

**Definíció.** Amennyiben az $\alpha$ paraméternek létezik $\hat{\alpha}_{H,n}$ hatásos becslése, egy $\hat{\alpha}_n$ **becslés** $e(\hat{\alpha}_n)$ **hatásfokán** az

$$e(\hat{\alpha}_n) = \frac{D^2(\hat{\alpha}_{H,n})}{D^2(\hat{\alpha}_n)}$$

hányadost értjük, amelynek értéke mindig 0 és 1 közé esik.

A gyakorlatban szokás kis mértékben torzított becslést is alkalmazni, ha a szórása kicsiny.

**Definíció.** Az $\alpha$ paraméter becsléseinek $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_n, \ldots$ sorozatát **aszimptotikusan torzítatlan**nak nevezzük, ha

$$\lim_{n \to \infty} M(\hat{\alpha}_n) = \alpha.$$

**3.2.1. tétel.** Az $\hat{m}_n$ mintaátlag a megfigyelt $\xi$ valószínűségi változó $M(\xi)$ várható értékének torzítatlan becslése.

**Bizonyítás.** A valószínűségi változók összegének várható értékére vonatkozó tétel felhasználásával azt kapjuk, hogy

$$M(\hat{m}_n) = M\left(\frac{1}{n} \cdot \sum_{i=1}^{n} \xi_i\right) = \frac{1}{n} \sum_{i=1}^{n} M(\xi_i) = \frac{n}{n} \cdot M(\xi) = M(\xi),$$

ami igazolja az állítást.

**3.2.2. tétel.** Egy adott $A$ esemény esetében a $\frac{k}{n}$ relatív gyakoriság az esemény $p = P(A)$ valószínűségének torzítatlan becslése, ahol $n$ a kísérletek és $k$ az $A$ esemény bekövetkezéseinek a száma.

**Bizonyítás.** Legyen $\xi$ a vizsgált $p$ valószínűségű $A$ esemény karakterisztikus valószínűségi változója. Felhasználva, hogy a karakterisztikus eloszlás esetén $M(\xi) = p$ és az $\hat{m}_n$ mintaközép éppen az esemény relatív gyakorisága, azaz $\hat{m}_n = \dfrac{k}{n}$, a tétel állítása az előző tétel közvetlen következménye.

**3.2.3. tétel.** A megfigyelt $\xi$ valószínűségi változó $D^2(\xi)$ szórásnégyzetének a $\hat{\sigma}_n^2$ tapasztalati szórásnégyzet nem torzítatlan becslése, az $\hat{s}_n^2$ korrigált tapasztalati szórásnégyzet viszont igen.

**Bizonyítás.** A valószínűségi változók összegének várható értékére és a független valószínűségi változók szórásnégyzetére vonatkozó tételek felhasználásával azt kapjuk, hogy

$$M\left(\hat{\sigma}_n^2\right) = M\left(\frac{1}{n}\cdot\sum_{i=1}^{n}(\xi_i - \hat{m}_n)^2\right) = M\left(\frac{1}{n}\cdot\sum_{i=1}^{n}\left(\xi_i - \frac{1}{n}\sum_{j=1}^{n}\xi_j\right)^2 = \right)$$

$$= M\left(\frac{1}{n}\cdot\sum_{i=1}^{n}\left((\xi_i - M(\xi)) + \left(M(\xi) - \frac{1}{n}\sum_{j=1}^{n}\xi_j\right)\right)^2\right) =$$

$$= M\left(\frac{1}{n}\cdot\sum_{i=1}^{n}\left((\xi_i - M(\xi)) - \frac{1}{n}\cdot\left(\sum_{j=1}^{n}\left(\xi_j - M(\xi)\right)\right)\right)^2\right) =$$

$$= M\left(\begin{array}{c} \dfrac{1}{n}\cdot\sum_{i=1}^{n}(\xi_i - M(\xi))^2 \\ -2\cdot\dfrac{1}{n^2}\sum_{i=1}^{n}(\xi_i - M(\xi))\cdot\sum_{j=1}^{n}\left(\xi_j - M(\xi)\right) \\ +\dfrac{1}{n^2}\left(\sum_{j=1}^{n}\left(\xi_j - M(\xi)\right)\right)^2 \end{array}\right) =$$

$$= M\left( \begin{array}{c} \frac{1}{n} \cdot \sum_{i=1}^{n} (\xi_i - M(\xi))^2 - 2 \cdot \frac{1}{n^2} \left( \sum_{j=1}^{n} (\xi_i - M(\xi)) \right)^2 \\ + \frac{1}{n^2} \left( \sum_{j=1}^{n} (\xi_i - M(\xi)) \right)^2 \end{array} \right) =$$

$$= M\left( \frac{1}{n} \cdot \sum_{i=1}^{n} (\xi_i - M(\xi))^2 - \frac{1}{n^2} \left( \sum_{i=1}^{n} (\xi_i - M(\xi)) \right)^2 \right) =$$

$$= M\left( \frac{n-1}{n^2} \cdot \sum_{i=1}^{n} (\xi_i - M(\xi))^2 \right) =$$

$$= \frac{n-1}{n^2} \cdot n \cdot M\left( (\xi - M(\xi))^2 \right) = \frac{n-1}{n} \cdot D^2(\xi),$$

ami igazolja, hogy a tapasztalati szórásnégyzet nem torzítatlan becslés. A $\hat{s}_n^2$ becslés torzítatlansága az

$$M\left(\hat{s}_n^2\right) = M\left( \frac{n}{n-1} \cdot \hat{\sigma}_n^2 \right) = \frac{n}{n-1} \cdot M\left(\hat{\sigma}_n^2\right) = D^2(\xi)$$

egyenlőségből következik.

A bizonyításból ennél több is kiderült:

**3.2.4. tétel.** A megfigyelt $\xi$ valószínűségi változó $D^2(\xi)$ szórásnégyzetének a $\hat{\sigma}_n^2$ tapasztalati szórásnégyzet aszimptotikusan torzítatlan becslése.

**Bizonyítás.** Az előző tétel bizonyításából adódóan:

$$\lim_{n \to \infty} M\left(\hat{\sigma}_n^2\right) = \lim_{n \to \infty} \frac{n-1}{n} \cdot D^2(\xi) = D^2(\xi)$$

**3.2.5. tétel.** Az $\hat{m}_n$ mintaátlag a megfigyelt $\xi$ valószínűségi változó $M(\xi)$ várható értékének

$$\hat{m}_{C,n} = \sum_{i=1}^{n} c_i \cdot \xi_i$$

alakú (lineáris) torzítatlan becslései közül a leghatékonyabb.

**Bizonyítás.** Először azt kell megmutatnunk, hogy az $\hat{m}_{C,n} = \sum_{i=1}^{n} c_i \cdot \xi_i$ alakú becslések csak akkor lehetnek torzítatlanok, ha $\sum_{i=1}^{n} c_i = 1$. Valóban:

$$M\left(\hat{m}_{C,n}\right) = M\left(\sum_{i=1}^{n} c_i \cdot \xi_i\right) = \sum_{i=1}^{n} c_i \cdot M(\xi_i) = \sum_{i=1}^{n} c_i \cdot M(\xi) =$$

$$= M(\xi) \cdot \sum_{i=1}^{n} c_i = M(\xi),$$

ahonnan következik, hogy $\sum_{i=1}^{n} c_i = 1$ (amennyiben $M(\xi) \neq 0$).

Hátra van még annak belátása, hogy $D^2(\hat{m}_n) \leq D^2\left(\hat{m}_{C,n}\right)$. Ehhez írjuk $c_i$-ket $c_i = \frac{1}{n} + \varepsilon_i$ alakba, ahol $\sum_{i=1}^{n} c_i = 1$ miatt $\sum_{i=1}^{n} \varepsilon_i = 0$. A független valószínűségi változók összegének szórására vonatkozó tétel felhasználásával azt kapjuk, hogy

$$D^2\left(\hat{m}_{C,n}\right) = D^2\left(\sum_{i=1}^{n} c_i \cdot \xi_i\right) = \sum_{i=1}^{n} D^2(c_i \cdot \xi_i) =$$

$$= \sum_{i=1}^{n} c_i^2 \cdot D^2(\xi_i) = D^2(\xi) \cdot \sum_{i=1}^{n} c_i^2 =$$

$$= D^2(\xi) \cdot \sum_{i=1}^{n} \left(\frac{1}{n} + \varepsilon_i\right)^2 = D^2(\xi) \cdot \sum_{i=1}^{n} \left(\frac{1}{n^2} + 2 \cdot \frac{\varepsilon_i}{n} + \varepsilon_i^2\right) =$$

$$= D^2(\xi) \cdot \left(\frac{1}{n} + \frac{2}{n} \sum_{i=1}^{n} \varepsilon_i + \sum_{i=1}^{n} \varepsilon_i^2\right) =$$

$$= D^2(\xi) \cdot \left(\frac{1}{n} + \sum_{i=1}^{n} \varepsilon_i^2\right) \geq \frac{1}{n} D^2(\xi).$$

Egyenlőség csak akkor áll fenn, ha $\varepsilon_i = 0$ $(i = 1, 2, \ldots, n)$, vagyis, ha $c_i = \frac{1}{n}$ minden $i$-re, ami igazolja a $D^2(\hat{m}_n) \le D^2(\hat{m}_{C,n})$ becslést.

Ha a megfigyelt valószínűségi változó normális eloszlású, ennél több is igaz:

**3.2.6. tétel.** Ha a megfigyelt $\xi$ valószínűségi változó normális eloszlású, akkor az $\hat{m}_n$ mintaátlag a várható érték *valamennyi* torzítatlan becslése közül a leghatásosabb.

**Definíció.** Az $\alpha$ paraméter becsléseinek $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_n, \ldots$ sorozatát **konzisztens becslés**ének nevezzük, ha becsléssorozat sztochasztikusan konvergál az $\alpha$ paraméterhez, azaz bármely tetszőlegesen kicsi $\varepsilon > 0$ szám esetén

$$\lim_{n \to \infty} P\left(\left|\hat{\alpha}_n - \alpha\right| > \varepsilon\right) = 0.$$

**3.2.7. tétel.** Ha $\hat{\alpha}_n$ az $\alpha$ torzítatlan becslése és $\lim\limits_{n \to \infty} D^2(\hat{\alpha}_n) = 0$, akkor az $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_n, \ldots$ becsléssorozat az $\alpha$ konzisztens becslése.

**Bizonyítás.** A Csebisev-egyenlőtlenség alapján

$$P\left(\left|\hat{\alpha}_n - M(\hat{\alpha}_n)\right| \ge \varepsilon\right) = P\left(\left|\hat{\alpha}_n - \alpha\right| \ge \varepsilon\right) \le \frac{D^2(\hat{\alpha}_n)}{\varepsilon^2}$$

tetszőleges $\varepsilon > 0$-ra. Mivel $D^2(\hat{\alpha}_n) \to 0$ esetén a bal oldalon álló valószínűség 0-hoz tart, ezért

$$\lim_{n \to \infty} P\left(\left|\hat{\alpha}_n - \alpha\right| \ge \varepsilon\right) = 0.$$

**3.2.8. tétel.** A mintaátlagok $\hat{m}_1, \ldots, \hat{m}_n, \ldots$ sorozata a megfigyelt $\xi$ valószínűségi változó $M(\xi)$ várható értékének konzisztens becslése, amennyiben a $\xi$ szórása létezik.

**Bizonyítás.** Az előző tétel alapján ehhez elég azt megmutatni, hogy $\lim\limits_{n\to\infty} D^2(\hat{m}_n)=0$. A független valószínűségi változók összegének szórására vonatkozó tétel felhasználásával azt kapjuk, hogy

$$\lim_{n\to\infty} D^2(\hat{m}_n)=\lim_{n\to\infty} D^2\left(\frac{1}{n}\cdot\sum_{i=1}^{n}\xi_i\right)=\lim_{n\to\infty}\frac{1}{n^2}D^2\left(\sum_{i=1}^{n}\xi_i\right)=$$

$$=\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}D^2(\xi_i)=\lim_{n\to\infty}\frac{1}{n}\cdot D^2(\xi)=0,$$

ami igazolja az állítást.

**Definíció.** Az $\hat{\alpha}_n=\hat{\alpha}_n(\xi_1,\xi_2,\ldots,\xi_n)$ statisztika **elégséges becslés**e a megfigyelt $\xi$ valószínűségi változó $\alpha$ paraméterének, ha a bármilyen módon megvalósuló $\{\hat{\alpha}_n=y\}$ feltétel esetén, a mintavételi változók e feltételre vonatkozó feltételes valószínűsége nem tartalmazza a becsült $\alpha$ paramétert. Diszkrét valószínűségi változó esetén ez a

$$P(\xi_1=x_1,\xi_2=x_2,\ldots,\xi_n=x_n\mid\hat{\alpha}(\xi_1,\xi_2,\ldots,\xi_n)=y)$$

feltételes valószínűség $\alpha$ paramétertől való függetlenségét, folytonos eloszlású valószínűségi változó esetén pedig az

$$f(x_1,x_2,\ldots,x_n\mid\hat{\alpha}_n(x_1,x_2,\ldots,x_n)=y)$$

feltételes sűrűségfüggvény $\alpha$ paramétertől való függetlenségét jelenti.

A mintavételi változók együttes feltételes eloszlásának ismerete a legtöbb, amire a megfigyelt $\xi$ valószínűségi változó ismeretlen $\alpha$ paraméterére vonatkozó információ-szerzésünk során számíthatunk. Ha ez az eloszlás a becsült $\alpha$ paraméterétől független, akkor az $\hat{\alpha}$ becslés arról minden információt tartalmaz.

**1. példa.** Mutassuk meg, hogy a Poisson-eloszlású $\xi$ valószínűségi változó $\lambda$ paraméterének az $\hat{m}_n$ mintaátlag elégséges becslése!

**Megoldás.** Az elégséges becslés definíciója értelmében elő kell állítani a $\xi_1, \xi_2, \ldots, \xi_n$ mintavételi változók $\{\hat{m}_n(\xi_1, \xi_2, \ldots, \xi_n) = y\}$ feltételhez tartozó

$$P(\xi_1 = k_1, \xi_2 = k_2, \ldots, \xi_n = k_n \mid \hat{m}_n(\xi_1, \xi_2, \ldots, \xi_n) = y)$$

feltételes valószínűséget, ahol $\hat{m}_n(\xi_1, \xi_2, \ldots, \xi_n) = \frac{1}{n} \cdot \sum_{i=1}^{n} k_i = y$.

A feltételes valószínűség definícióját és a mintavételi változók függetlenségét felhasználva ez a feltételes valószínűség

$$\begin{aligned} &P(\xi_1 = k_1, \xi_2 = k_2, \ldots, \xi_n = k_n \mid \hat{m}_n = y) = \\ &= \frac{P(\xi_1 = k_1) \cdot P(\xi_2 = k_2) \cdot \ldots \cdot P(\xi_n = k_n)}{P(\hat{m}_n = y)} = \\ &= \frac{P(\xi_1 = k_1) \cdot P(\xi_2 = k_2) \cdot \ldots \cdot P(\xi_n = k_n)}{P(n \cdot \hat{m}_n = n \cdot y)} \end{aligned}$$

alakba írható. Mivel a $\xi_i$ $(i = 1, 2, \ldots, n)$ mintavételi változók valamennyien a megfigyelt $\xi$ valószínűségi változóval megegyező, $\lambda$ paraméterű Poisson-eloszlású valószínűségi változók, ezért

$$P(\xi_i = k_i) = \frac{\lambda^{k_i}}{k_i!} e^{-\lambda} \qquad (i = 1, 2, \ldots, n).$$

Bizonyítható továbbá, hogy $n \cdot \hat{m}_n$ eloszlása $n \cdot \lambda$ paraméterű Poisson-eloszlású, azaz

$$P(n \cdot \hat{m}_n = n \cdot y) = \frac{(n \cdot \lambda)^{n \cdot y}}{(n \cdot y)!} \cdot e^{-n \cdot \lambda}.$$

A fentiek figyelembevételével azt kapjuk, hogy

$$P(\xi_1 = k_1, \xi_2 = k_2, \ldots, \xi_n = k_n \mid \hat{m}_n = y) =$$
$$= \frac{P(\xi_1 = k_1) \cdot P(\xi_2 = k_2) \cdot \ldots \cdot P(\xi_n = k_n)}{P(n \cdot \hat{m}_n = n \cdot y)} =$$
$$= \frac{\frac{\lambda^{k_1}}{k_1!} e^{-\lambda} \cdot \frac{\lambda^{k_2}}{k_2!} e^{-\lambda} \cdot \ldots \cdot \frac{\lambda^{k_n}}{k_n!} e^{-\lambda}}{\frac{(n \cdot \lambda)^{n \cdot y}}{(n \cdot y)!} \cdot e^{-n \cdot \lambda}} = \frac{\frac{1}{k_1!} \cdot \frac{1}{k_2!} \cdot \ldots \cdot \frac{1}{k_n!}}{\frac{(n)^{n \cdot y}}{(n \cdot y)!}} =$$
$$= \frac{(n \cdot y)!}{(n)^{n \cdot y}} \cdot \frac{1}{k_1!} \cdot \frac{1}{k_2!} \cdot \ldots \cdot \frac{1}{k_n!},$$

ami független a $\lambda$ paramétertől, tehát a becslés valóban elégséges.

## 3.3. A legnagyobb valószínűség (maximum likelihood) elve

Eddig nem vizsgáltuk azt a kérdést, hogy valamely eloszlás paraméterére vonatkozólag hogyan konstruálható jó becslés? Van-e olyan általános matematikai módszer, amely megadja, hogy a mintaelemekből milyen $\hat{\alpha}_n = \hat{\alpha}_n(\xi_1, \xi_2, \ldots, \xi_n)$ statisztikát kell ahhoz számítanunk, hogy az adott $\alpha$ paraméterre jó becslést kapjunk?

A legfontosabb általános módszer a becslésre az ún. **maximum likelihood módszer,** amit magyarra fordítva a **legnagyobb valószínűség módszeré**nek nevezhetünk. A módszer R. A. Fishertől származik. Először a módszer gondolatmenetét ismertetjük.

**1. példa.** Egy alapsokaság ismeretlen $p$ selejtarányát szeretnénk mintavétel útján megbecsülni. Ezért az alapsokaságból $n$ elemet választunk ki egymás után, visszatevéssel. (A selejtarány így minden egyes húzás után változatlan marad.) Tegyük fel, hogy a vizsgálatunk során $k$ selejtest találtunk. Ennek alapján milyen becslést adhatunk az alapsokaság selejtarányára?

**Megoldás.** Jelölje $A$ azt az eseményt, hogy egy kiválasztásnál selejtes elemet választottunk. Az alapsokaságot így az $A$ esemény $\xi_A$ indikátor vál-

tozójával azonosítjuk, és a feladatunk az $A$ esemény ismeretlen $P(A)=p$ valószínűségének a becslése.
Legyen $\xi_i$ az *i*-edik kísérletnél az $A$ esemény indikátor változója, azaz

$$\xi_i=\begin{cases}1, & \text{ha } A \text{ bekövetkezett,}\\ 0, & \text{ha } A \text{ nem következett be.}\end{cases}$$

A $\xi_1,\xi_2,\ldots,\xi_n$ minta tehát most indikátor változók sorozata. Ha az $A$ esemény $k$-szor következett be, akkor a mintában $k$ darab 1-es és $(n-k)$ darab 0-ás kísérlet szerepel. Mivel független kísérleteket végeztünk, ezen minta létrejöttének valószínűsége az ismeretlen $p$ paraméter függvényében:

$$f(p)=p^k\cdot(1-p)^{n-k}.$$

Bármely két lehetséges $p$ érték közül azt fogadjuk el szívesebben, amely mellett a kapott minta létrejöttének $f(p)$ valószínűsége a nagyobb. Így egy valószínűbb esemény bekövetkezésének adunk nagyobb esélyt egy valószínűtlenebbel szemben. Ezt az elvet követve azt a $p\in(0,1)$ értéket keressük, amely mellett az $f(p)$, vagyis a minta létrejöttének valószínűsége a legnagyobb. Ez egy egyváltozós szélsőértékfeladat. Felhasználva, hogy

$$f'(p)=k\cdot p^{k-1}\cdot(1-p)^{n-k}-p^k\cdot(n-k)\cdot(1-p)^{n-k-1},$$

szélsőérték csak ott lehet, ahol

$$\begin{aligned}f'(p)&=p^{k-1}\cdot(1-p)^{n-k-1}\cdot[k\cdot(1-p)-p\cdot(n-k)]=\\&=p^{k-1}\cdot(1-p)^{n-k-1}\cdot[k-n\cdot p]=0.\end{aligned}$$

Mivel $p\in(0,1)$, az $f$-nek a

$$p=\frac{k}{n}$$

esetben lehet szélsőértéke. Azonnal látható, hogy $f(p)$-nek itt valóban maximuma van. Mintánk elemzése során így arra a következtetésre jutottunk, hogy az alapsokaság selejtarányának legvalószínűbb értéke a selejt mintabeli gyakorisága, így a

$$\hat{p}_n = \hat{p}_n(\xi_1, \xi_2, \ldots, \xi_n) = \hat{m}_n(\xi_1, \xi_2, \ldots, \xi_n) = \frac{1}{n} \cdot \sum_{i=1}^{n} \xi_i = \frac{k}{n}$$

statisztika adja $p$ legjobb becslését.

### 3.3.1. A legnagyobb valószínűség elvének matematikai megfogalmazása

**Diszkrét eset:** Jelölje az alapsokaság valószínűség-eloszlását

$$p(x, \alpha) := P(\xi = x),$$

ahol $\alpha$ ismeretlen paraméter. Legyen továbbá $x_1, x_2, \ldots, x_n$ egy, az alapsokaságból vett konkrét $n$ elemű minta. A minta létrejöttének valószínűsége a mintaelemek függetlensége folytán az egyes valószínűségek szorzata. Ezt a szorzatot **likelihood-függvény**nek nevezzük és

$$L(x_1, x_2, \ldots, x_n, \alpha) = p(x_1, \alpha) \cdot p(x_2, \alpha) \cdot \ldots \cdot p(x_n, \alpha)$$

módon jelöljük. Azt az $\hat{\alpha} = \hat{\alpha}(x_1, x_2, \ldots, x_n)$ értéket fogjuk az ismeretlen paraméter valódi értéke becslésének tekinteni, amely mellett a likelihood-függvény a legnagyobb értékét veszi fel.

**Folytonos eset:** Itt $P(\xi = x) = 0$ minden $x$-re, de

$$P(x \le \xi < x + \Delta x) \approx f(x, \alpha) \cdot \Delta x,$$

ahol $f$ az eloszlás sűrűségfüggvénye és $\alpha$ az eloszlás ismeretlen paramétere. Így annak a valószínűsége, hogy valamely mintaelem egy kicsiny, rögzített hosszúságú intervallumba esik, közelíthető a sűrűségfüggvény segítségével. A szélsőérték meghatározása szempontjából a rögzített részintervallum hosszának nincs jelentősége, csak egy konstans szorzóként szerepel, ezért a likelihood-függvény ebben az esetben az

$$L(x_1, x_2, \ldots, x_n, \alpha) = f(x_1, \alpha) \cdot f(x_2, \alpha) \cdot \ldots \cdot f(x_n, \alpha)$$

függvény.

**Megjegyzések:**

1. A gyakorlatban az $n$-tényezős szorzat deriválása elkerülhető, ha a likelihood-függvény helyett annak logaritmusával dolgozunk, mivel a logaritmus függvény szigorúan monoton növekedő volta miatt a szélsőértékhely a transzformációval nem változik.
2. Több paraméter együttes becslése nem jelent elvi változtatást.
3. A legnagyobb valószínűség elvének alkalmazásakor konzisztens, de nem feltétlenül torzítatlan becslést kapunk. Be lehet azonban látni, hogy az így nyert becslés legalábbis aszimptotikusan torzítatlan.
4. Ha a kérdéses paraméternek létezik hatásos becslése, akkor a likelihood-függvénynek csak egyetlen szélsőértéke van, és ez éppen a hatásos becslés.
5. A maximum likelihood becslés az invertálható leképezéssel végrehajtott transzformációra *invariáns*, azaz, ha $\hat{\alpha}$ maximum likelihood becslés $\alpha$-ra, és $g$ egy kölcsönösen egyértelmű függvény, akkor $g(\hat{\alpha})$ maximum likelihood becslése $g(\alpha)$-nak.

**2. példa.** Legyen $\xi$ egy $N(m,\sigma)$ normális eloszlású valószínűségi változó. Határozzuk meg az $m$ és $\sigma$ paraméterek maximum likelihood becslését!

**Megoldás.** A likelihood-függvény és annak logaritmusa

$$L(x_1,x_2,\ldots,x_n,m,\sigma)=\frac{1}{\left(\sigma\sqrt{2}\right)^n}\cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-m)^2},$$

$$\ln\left(L(x_1,x_2,\ldots,x_n,m,\sigma)\right)=-n\cdot\ln\left(\sigma\sqrt{2}\right)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-m)^2.$$

Egy függvénynek és a logaritmusának ugyanazokon a helyeken van maximuma. Ezeket a helyeket a függvény logaritmusa esetén határozzuk meg, mivel így egyszerűbb. Egy függvénynek ott lehet maximuma, ahol a parciális deriváltjai 0-val egyenlők. A lehetséges maximum helyeket ezért a

$$\partial_m \ln(L(x_1, x_2, \ldots, x_n, m, \sigma)) = \frac{1}{\sigma^2} \sum_{i=1}^{n} x_i - \frac{n \cdot m}{\sigma^2} = 0$$

$$\partial_\sigma \ln(L(x_1, x_2, \ldots, x_n, m, \sigma)) = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i - m)^2 = 0$$

egyenletrendszer megoldásai adják. Az egyenletrendszernek egyetlenegy $(m_0, \sigma_0)$ megoldása van,

$$m_0 = \frac{\sum_{i=1}^{n} x_i}{n} = \hat{m}_n, \quad \sigma_0^2 = \frac{\sum_{i=1}^{n} \left( x_i - \frac{1}{n} \sum_{i=1}^{n} x_i \right)^2}{n} = \hat{\sigma}_n^2$$

Megmutatható (a részleteket itt mellőzzük), hogy ez a hely valóban maximumhely.

A maximum likelihood módszer tehát normális eloszlás esetén az $m$ várható érték becslésére a mintaközepet, a $\sigma^2$ szórásnégyzet becslésére pedig az empirikus szórást szolgáltatja. Ez utóbbi becslés, mint már megmutattuk, csak aszimptotikusan torzítatlan. Megjegyezzük, hogy az előző Megjegyzés 5) pontja értelmében $\hat{\sigma}_n$ maximum likelihood becslése a $\sigma$ szórásnak.

## 3.4. A konfidencia (megbízhatósági) intervallum

Egy alapsokaságból vett konkrét $(x_1, x_2, \ldots, x_n)$ minta esetén kiszámított $\hat{\alpha}_n = \hat{\alpha}_n(x_1, x_2, \ldots, x_n)$ becslés az $\hat{\alpha}_n(\xi_1, \xi_2, \ldots, \xi_n)$ statisztikai függvény helyettesítési értéke az $(x_1, x_2, \ldots, x_n)$ helyen, és így mintánként változik. A becsült $\alpha$ paraméter és az $\hat{\alpha}_n$ becslés valóságos eltérésének nagyságára ezért biztos kijelentést nem tudunk adni. Azonban, ha ismerjük az $\alpha - \hat{\alpha}_n$ eloszlását, akkor meg tudunk adni olyan $C^1_{n,\varepsilon} \cdot \hat{\beta}_n$ és $C^2_{n,\varepsilon} \cdot \hat{\beta}_n$ úgynevezett **konfidencia (megbízhatósági) határok**at, hogy a becsült $\alpha$ paraméter és az $\hat{\alpha}_n$ becslés eltérése nagy $(1-\varepsilon)$ valószínűséggel e határok között legyen, azaz teljesüljön a

$$P\left(C^1_{n,\varepsilon} \cdot \hat{\beta}_n < \alpha - \hat{\alpha}_n < C^2_{n,\varepsilon} \cdot \hat{\beta}_n\right) = 1 - \varepsilon$$

feltétel, amit

$$P\left(C_{n,\varepsilon}^{1}\cdot\hat{\beta}_n+\hat{\alpha}_n<\alpha<C_{n,\varepsilon}^{2}\cdot\hat{\beta}_n+\hat{\alpha}_n\right)=1-\varepsilon$$

ekvivalens alakba is írhatunk, ahol $\hat{\beta}_n$ az $\alpha-\hat{\alpha}_n$ eloszlásától függő statisztika, $C_{n,\varepsilon}^{i}$ $(i=1,2)$ pedig az $n$-től és az $\varepsilon$-tól függő valós számok.

A $P\left(C_{n,\varepsilon}^{1}\cdot\hat{\beta}_n+\hat{\alpha}_n<\alpha<C_{n,\varepsilon}^{2}\cdot\hat{\beta}_n+\hat{\alpha}_n\right)=1-\varepsilon$ alak azt mutatja,hogy az $\alpha$ paraméter valódi értéke $1-\varepsilon$ valószínűséggel esik a

$$\left(C_{n,\varepsilon}^{1}\cdot\hat{\beta}_n+\hat{\alpha}_n, C_{n,\varepsilon}^{2}\cdot\hat{\beta}_n+\hat{\alpha}_n\right)$$

úgynevezett **konfidencia (megbízhatósági) intervallum**ba.

**Definíció.** Egy $\hat{\alpha}_n$ becsléshez tartozó

$$\left(C_{n,\varepsilon}^{1}\cdot\hat{\beta}_n+\hat{\alpha}_n, C_{n,\varepsilon}^{2}\cdot\hat{\beta}_n+\hat{\alpha}_n\right)$$

**konfidencia intervallum**ot $(1-\varepsilon)$ **megbízhatósági szinthez tartozó**nak nevezzük, ha az esetek $100\cdot(1-\varepsilon)\,\%$-ában tartalmazza a becsült $\alpha$ paraméter valódi értékét, azaz

$$P\left(C_{n,\varepsilon}^{1}\cdot\hat{\beta}_n+\hat{\alpha}_n<\alpha<C_{n,\varepsilon}^{2}\cdot\hat{\beta}_n+\hat{\alpha}_n\right)=1-\varepsilon\,.$$

Diszkrét eloszlások esetén pontos egyenlőség nem mindig érhető el: ekkor *legalább* $(1-\varepsilon)$ megbízhatósági szintről szokás beszélni.

Egy konfidencia intervallum megadásakor először a $\hat{\beta}_n$ statisztikát határozzuk meg. Feltéve, hogy $\hat{\beta}_n>0$, a $C_{n,\varepsilon}^{i}$ értékeket ezt követően a

$$P\left(C_{n,\varepsilon}^{1}\cdot\hat{\beta}_n+\hat{\alpha}_n<\alpha<C_{n,\varepsilon}^{2}\cdot\hat{\beta}_n+\hat{\alpha}_n\right)=1-\varepsilon$$

kifejezés

$$P\left(C_{n,\varepsilon}^{1} < \frac{\alpha - \hat{\alpha}_n}{\hat{\beta}_n} < C_{n,\varepsilon}^{2}\right) = 1 - \varepsilon$$

ekvivalens alakjának felhasználásával a

$$\hat{\gamma}_n = \frac{\alpha - \hat{\alpha}_n}{\hat{\beta}_n}$$

valószínűségi változó (statisztika) $F$ eloszlásfüggvényének segítségével az

$$F\left(C_{n,\varepsilon}^{1}\right) = P\left(\hat{\gamma}_n < C_{n,\varepsilon}^{1}\right) = \frac{\varepsilon}{2}, \quad F\left(C_{n,\varepsilon}^{2}\right) = P\left(\hat{\gamma}_n < C_{n,\varepsilon}^{2}\right) = 1 - \frac{\varepsilon}{2}$$

összefüggésekből határozzuk meg.
Ha $\hat{\beta}_n < 0$, akkor

$$P\left(C_{n,\varepsilon}^{1} > \frac{\alpha - \hat{\alpha}_n}{\hat{\beta}_n} > C_{n,\varepsilon}^{2}\right) = 1 - \varepsilon$$

az ekvivalens alak és a $C_{n,\varepsilon}^{i}$ értékeket az

$$F\left(C_{n,\varepsilon}^{2}\right) = P\left(\hat{\gamma}_n < C_{n,\varepsilon}^{2}\right) = \frac{\varepsilon}{2}, \quad F\left(C_{n,\varepsilon}^{1}\right) = P\left(\hat{\gamma}_n < C_{n,\varepsilon}^{1}\right) = 1 - \frac{\varepsilon}{2}$$

összefüggésekből határozzuk meg.

A gyakorlatban a

$$P\left(\alpha < C_{n,\varepsilon}^{2} \cdot \hat{\beta}_n + \hat{\alpha}_n\right) = 1 - \varepsilon \text{ és } P\left(C_{n,\varepsilon}^{1} \cdot \hat{\beta}_n + \hat{\alpha}_n < \alpha\right) = 1 - \varepsilon$$

feltételekkel megfogalmazott

$$\left(-\infty, C_{n,\varepsilon}^{2} \cdot \hat{\beta}_n + \hat{\alpha}_n\right) \text{ és } \left(C_{n,\varepsilon}^{1} \cdot \hat{\beta}_n + \hat{\alpha}_n, +\infty\right)$$

egyoldali konfidencia intervallumok is használatosak.

**1. példa.** Legyen $\xi$ egy ismeretlen $m$ várható érékű és ismert $\sigma$ szórású valószínűségi változó. A $m$ várható értéket egy nagy $n$ elemszámú $\xi_1, \xi_2, \ldots, \xi_n$ minta felhasználásával az $\hat{m}_n$ mintaátlaggal becsüljük. Határozzuk meg a várható érték $\hat{m}_n$ mintaátlaggal történő becslésének $100 \cdot (1-\varepsilon)\%$-os megbízhatósági szinthez tartozó konfidencia intervallumát a $\sigma = 5$, $n = 36$, $\hat{m}_n = 12$ és $\varepsilon = 0{,}05$ esetben!

**Megoldás.** A konfidencia intervallumot annak felhasználásával konstruáljuk meg, hogy a centrális határeloszlás tétel szerint nagy $n$ esetén a

$$\hat{\gamma}_n = \frac{\alpha - \hat{\alpha}_n}{\hat{\beta}_n} = \frac{m - \hat{m}_n}{-\frac{\sigma}{\sqrt{n}}} = \frac{\sum_{i=1}^{n} \xi_i - n \cdot m}{\sqrt{n} \cdot \sigma}$$

valószínűségi változó közelítőleg standard normális, azaz $N(0,1)$ eloszlásúnak tekinthető. Mivel $\hat{\beta}_n < 0$ a $C_{n,\varepsilon}^{i}$ értékeket a

$$\Phi\left(C_{n,\varepsilon}^{2}\right) = \frac{\varepsilon}{2}, \quad \Phi\left(C_{n,\varepsilon}^{1}\right) = 1 - \frac{\varepsilon}{2}$$

összefüggésből határozzuk meg, ahol $\Phi$ a standard normális eloszlás eloszlásfüggvényét jelöli. Felhasználva, hogy a standard normális eloszlás eloszlásfüggvénye szimmetrikus, azaz $\Phi(-x) = \Phi(x)$, ezért $C_{n,\varepsilon}^{1} = C_{\varepsilon}$ és $C_{n,\varepsilon}^{2} = -C_{\varepsilon}$ alakú, ahol $\Phi(C_{\varepsilon}) = 1 - \frac{\varepsilon}{2}$. Az $\varepsilon = 0{,}05$ esetben a $C_{0,05}$ értékét a $\Phi(C_{0,05}) = 0{,}975$ feltételből határozzuk meg, azaz $C_{0,05} = \Phi^{-1}(0{,}975) = 1{,}96$ lesz.

Az MS Excel használata estén a $C_{0,05}$ értéket az

= INVERZ.STNORM(0,975)

kifejezés kiértékelésével kaphatjuk meg.

Így az ismeretlen $m$ értékére vonatkozó konfidencia intervallum

$$\left(-C_{n,\varepsilon}^{1}\cdot\hat{\beta}_n+\hat{\alpha}_n;-C_{n,\varepsilon}^{2}\cdot\hat{\beta}_n+\hat{\alpha}_n\right)=\left(-1{,}96\cdot\frac{5}{6}+12;\ 1{,}96\cdot\frac{5}{6}+12\right)\approx$$
$$\approx(10{,}366,\,13{,}633)$$

lesz.

Vegyük észre, hogy a konfidencia intervallum szélessége mintaelemek számának növekedésével és a szórás csökkenésével csökken, viszont a megbízhatósági szint emelésével nő.

## 3.5. Statisztikai hipotézisek (feltevések) vizsgálata

A gyakorlatban igen sok esetben már eleve van valamilyen feltevésünk egy valószínűségi változó eloszlásáról, illetve az eloszlás valamely paraméteréről, és azt szeretnénk ellenőrizni, hogy ez a feltevésünk helyes-e?

Ilyenek például a következő esetek:

*a)* Adott technológiával gyártott vasbeton gerendák mérete normális eloszlást követ, ismert szórással. Az elkészült gerendák mérete az előírt várható méret érték körül ingadozik-e?
*b)* Két különböző keverőgéppel készült aszfaltkeverékben a bitumentartalom szórása megegyezik-e?

**Statisztikai hipotézis**en (**feltevés**en) egy vagy több, valószínűség-eloszlásra vonatkozó feltevést értünk.

**Statisztikai próbá**nak nevezzük azt az eljárást, amely alapján egy statisztikai hipotézisről döntünk.

Egy statisztikai próba lépései következők.

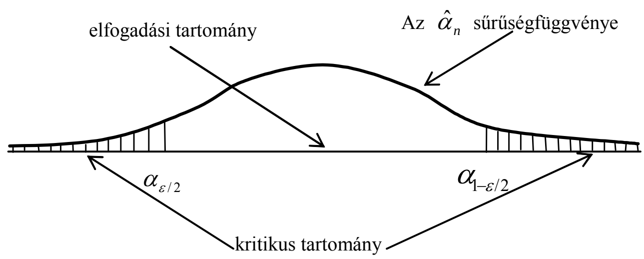
**Elméleti lépések:**
1. Az ismeretlen eloszlásra vagy az eloszlás ismeretlen paraméterére (például az előző tapasztalatok alapján) egy $H_0$ **null-hipotézis**t állítunk fel. Az eloszlás vagy paraméter számára a null-hipotézistől eltérő más lehetőségek bizonyos halmazát, esetleg az összes más lehetőséget együttesen **ellenhipotézis**nek, vagy **alternatív hipotézis**nek nevezzük és $H_1$-gyel jelöljük.

2. A feltevésünk ellenőrzésére mintavételen alapuló **konzisztens** $\hat{\alpha}_n = \hat{\alpha}_n(\xi_1, \xi_2, \ldots, \xi_n)$ statisztikai függvényt konstruálunk.

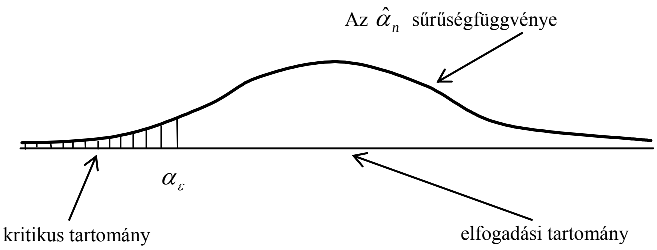
3. Meghatározzuk a $\hat{\alpha}_n$ statisztikai függvény eloszlását.

4. Kijelöljük az eloszlás **kritikus tartományát**, ahova a statisztikai $\hat{\alpha}_n$ függvény, mint valószínűségi változó, értéke csak kicsiny $\varepsilon$ (pl.: $5\%, 1\%, 0{,}1\%$) valószínűséggel esik. A kritikus tartományt **elutasítási tartomány**nak is nevezzük. A kritikus tartomány komplementerét **elfogadási tartomány**nak hívjuk. A döntés szintjét jellemző $100 \cdot (1-\varepsilon)\%$ számot a próba **szignifikancia szintjének** vagy röviden a **próba szintjének** nevezzük.

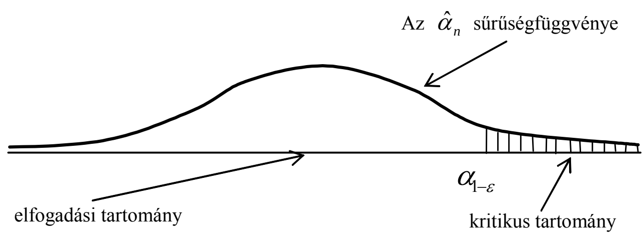
*A kritikus tartomány kijelölése:*
4.1. A $P(\alpha_{\varepsilon/2} < \hat{\alpha}_n < \alpha_{1-\varepsilon/2}) = 1-\varepsilon$ feltétellel megadott kétoldali próba esetén a kritikus tartomány egy $(-\infty, \alpha_{\varepsilon/2})$ és egy $(\alpha_{1-\alpha/2}, +\infty)$ intervallum uniójából áll. Az intervallumok $\alpha_{\varepsilon/2}$ és a $\alpha_{1-\varepsilon/2}$ határait az $\hat{\alpha}_n$ $F_\alpha$ eloszlásfüggvényének felhasználásával az $F_\alpha(\alpha_{\varepsilon/2}) = \frac{\varepsilon}{2}$ és az $F_\alpha(1-\alpha_{\varepsilon/2}) = 1-\frac{\varepsilon}{2}$ feltételekből határozzuk meg. Az elfogadási tartomány ekkor az $(\alpha_{\varepsilon/2}, \alpha_{1-\varepsilon/2})$ intervallum. (Egyenlőséggel megfogalmazott $H_0$ hipotézis esetén alkalmazzuk, azaz, ha $H_0 = \{\alpha = \alpha_0\}$, $H_1 = \{\alpha \neq \alpha_0\}$.)




4.2. A $P(\alpha_\varepsilon < \hat{\alpha}_n) = 1-\varepsilon$ feltétellel megadott egyoldali próba esetén a kritikus tartomány egy $(-\infty, \alpha_\varepsilon)$ intervallum. Az intervallum $\alpha_\varepsilon$ határát az $\hat{\alpha}_n$ $F_\alpha$ eloszlásfüggvényének felhasználásával a $F_\alpha(\alpha_\varepsilon) = \varepsilon$ feltételből határozzuk meg. Az elfogadási tartomány ekkor az $(\alpha_\varepsilon, +\infty)$ intervallum. (Egyenlőtlenséggel megfogalmazott $H_0$ hipotézis esetén alkalmazzuk, azaz, ha $H_0 = \{\alpha \geq \alpha_0\}$, $H_1 = \{\alpha < \alpha_0\}$.)




4.3. A $P(\hat{\alpha}_n < \alpha_{1-\varepsilon}) = 1 - \varepsilon$ feltétellel megadott egyoldali próba esetén a kritikus tartomány egy $(\alpha_{1-\varepsilon}, +\infty)$ intervallum. Az intervallum $\alpha_{1-\varepsilon}$ határát az $\hat{\alpha}_n$ $F_\alpha$ eloszlásfüggvényének felhasználásával a $F_\alpha(\alpha_{1-\varepsilon}) = 1 - \varepsilon$ feltételből határozzuk meg. Az elfogadási tartomány ekkor a $(-\infty, \alpha_{1-\varepsilon})$ intervallum. (Egyenlőtlenséggel megfogalmazott $H_0$ hipotézis esetén alkalmazzuk, azaz, ha $H_0 = \{\alpha \leq \alpha_0\}$, $H_1 = \{\alpha > \alpha_0\}$.)




**Gyakorlati lépések:**

1. Mintát veszünk az alapsokaságból.
2. A minta adatokból kiszámítjuk a statisztikai függvény helyettesítési értékét.
3. A kapott számérték alapján döntünk:

Megtartjuk a null-hipotézist, ha a kiszámított helyettesítési érték az elfogadási tartományba esik.

Elvetjük a null-hipotézist, ha a kiszámított helyettesítési érték a kritikus (elutasítási) tartományba esik. Ekkor a feltételezett eloszláshoz képest egy valószínűtlen esemény következett be. Ekkor **szignifikáns (jelentős)** az adatokban tükröződő eltérés a feltétevésünkhöz képest.

A statisztikai próbák alkalmazásakor hibás döntéseket is hozhatunk. Kétféle hibát követhetünk el:

*Elvetjük a* $H_0$ *hipotézist, noha az helyes.* Ez az úgynevezett **elsőfajú hiba** akkor fordul elő, ha a statisztikánk helyettesítési értéke a kicsiny valószínűségű kritikus tartományba esik.

*Megtartjuk a* $H_0$ *hipotézist, noha az hibás.* Ez az úgynevezett **másodfajú hiba** akkor fordul elő, ha a statisztikai függvény helyettesítési értéke a véletlen folytán az elfogadási tartományba esik.

Egy **próbá**t **konzisztens**nek nevezünk, ha a minta elemszám minden határon túl történő növelésével a másodfajú hiba minden rögzített alternatív hipotézis esetén 0-hoz tart.

Az $\alpha = \alpha_0$ ún. *egyszerű null-hipotézis* esetén az elsőfajú hiba $\varepsilon$, mely uralható, hiszen mi állítjuk be az $(1-\varepsilon)$ szintet. A másodfajú hiba azonban függ a paraméter (ismeretlen) értékétől, és így nem uralható. Mindenesetre konzisztens próba és elég nagy mintaelemszám esetén a másodfajú hiba „kicsi”, így a szintet választhatjuk „nagyra”. Az első- és másodfajú hiba mozgása ellentétes, megfelelő kompromisszumra és az alternatívák közti szereposztás olyan beállítására van szükség, hogy az uralhatatlan másodfajú hiba elkövetése legyen a kisebb vétség.

A további fejezetekben csak ismert eloszlású konzisztens statisztikai próbákat ismertetünk.

Az első- és a másodfajú hiba egyensúlyba tartása szempontjából kialakult szokás a $100 \cdot (1-\varepsilon)\,\% = 95\%$-os szignifikancia szint választása, de ez alól számos kivétel is létezik. A legtöbb statisztikai programcsomag (így az MS Excel is) kiírja azt a legkisebb $\varepsilon$-t ill. azt a legnagyobb szintet, amely mellett el tudjuk utasítani a null-hipotézist (kisebb $\varepsilon$-ra elutasítjuk, nagyobbra elfogadjuk), ez következik abból, hogy magasabb szinthez (azaz kisebb $\varepsilon$-hoz) tágabb konfidenciaintervallum tartozik.

## 3.6. Paraméteres próbák

Egy statisztikai próbát **paraméteres próbának** nevezünk, ha az alapsokaság valamely paraméterére vagy paramétereire vonatkozik.

Az alábbiakban várható értékre és szórásra vonatkozó egy- és kétmintás statisztikai próbákat ismertetünk.

A továbbiakban egy statisztikai mintát *nagy elemszámúnak* mondunk, ha a mintaelemek $n$ száma legalább 30, mivel ekkor a standard normális és a Student-eloszlással számított valószínűségértékek már alig különböznek.

### 3.6.1. Az egymintás u-próba

*Alkalmazása:* Egy ismert $D(\xi) = \sigma$ szórású, de ismeretlen $M(\xi) = m$ várható értékű normális eloszlású $\xi$ valószínűségi változó várható értékére vonatkozó $m_0$ hipotézis helyességének ellenőrzése.

A próba statisztika:

$$\hat{u} = \frac{\hat{m}_n - m_0}{\frac{\sigma}{\sqrt{n}}},$$

ahol $n$ a minta elemszáma.

*Eloszlása*: $\hat{u}$ standard normális eloszlású valószínűségi változó.

Nagy elemszámú minta esetén, a centrális határérték tétel miatt a próba $\xi$ eloszlásától függetlenül közelítőleg normális eloszlású, ezért ebben az esetben a próba nem normális eloszlású valószínűségi változó esetén is alkalmazható.

Nagy elemszámú minta esetén a próba akkor is alkalmazható, ha a $\sigma$ szórás nem ismert. Ekkor a szórás a mintából történő becslésével helyettesíthető, és a próba statisztika alakja a következő lesz:

$$\hat{u} = \frac{\hat{m}_n - m_0}{\frac{\hat{s}_n}{\sqrt{n}}}$$

Nagy elemszámú minta esetén, ha $\xi$ egy $A$ esemény karakterisztikus valószínűségi változója, akkor a próbával az $A$ esemény $p = P(A)$ valószínűségére vonatkozó $p_0$ hipotézis ellenőrzésére is alkalmazható, mivel ekkor $M(\xi) = p$.

### 3.6.2. A kétmintás u-próba

*Alkalmazása:* Ismert $D(\xi) = \sigma_1$ és $D(\eta) = \sigma_2$ szórású, de ismeretlen $M(\xi) = m_1$ és $M(\eta) = m_2$ várható értékű normális eloszlású $\xi$ és $\eta$ valószínűségi változók várható értékének különbségére vonatkozó $m_0 = m_1 - m_2$ hipotézis helyességének ellenőrzése.
A próba statisztika:

$$\hat{u} = \frac{\hat{m}_1 - \hat{m}_1 - m_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}},$$

ahol $n_1$ a $\xi$-re, $n_2$ pedig az $\eta$-ra vonatkozó független minta elemszáma.

*Eloszlása:* $\hat{u}$ standard normális eloszlású valószínűségi változó (a nullhipotézis fennállása esetén).

Nagy elemszámú minta esetén a centrális határérték tétel miatt a próba a $\xi$ és $\eta$ eloszlásától függetlenül közelítőleg normális eloszlású. Ezért ebben az esetben a próba nem normális eloszlású valószínűségi változók esetében is alkalmazható.

Ha $\sigma = \sigma_1 = \sigma_2$, azaz a szórások megegyeznek, akkor a próba statisztika alakja egyszerűsödik:

$$\hat{u} = \frac{\hat{m}_1 - \hat{m}_2 - m_0}{\sigma \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Nagy elemszámú minta esetén ha $\sigma = \sigma_1 = \sigma_2$, de a közös $\sigma$ szórás nem ismert, akkor az a mintából történő becslésével helyettesíthető és a próba statisztika alakja a következő lesz:

$$\hat{u} = \frac{\hat{m}_2 - \hat{m}_2 - m_0}{\sqrt{\frac{(n_1-1)\cdot \hat{s}_1^2 + (n_2-1)\cdot \hat{s}_2^2}{n_1+n_2-2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Nagy elemszámú minta esetén a próba akkor is alkalmazható, ha a $\sigma_1$ és a $\sigma_2$ szórások nem ismertek. Ekkor ezeket a szórásokat a mintából történő becslésükkel helyettesítjük és a Welch-féle

$$\hat{w} = \frac{\hat{m}_1 - \hat{m}_2 - m_0}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}}$$

próba statisztikát alkalmazzuk, ami ebben az esetben közelítőleg szintén standard normális eloszlású.

### 3.6.3. Az egymintás t-próba

*Alkalmazása:* Ismeretlen $D(\xi) = \sigma$ szórású és ismeretlen $M(\xi) = m$ várható értékű normális eloszlású $\xi$ valószínűségi változó várható értékére vonatkozó $m_0$ hipotézis helyességének ellenőrzése.
A próba statisztika:

$$\hat{t}_{(n-1)} = \frac{\hat{m}_n - m_0}{\frac{\hat{s}_n}{\sqrt{n}}},$$

ahol $n$ a minta elemszáma.

*Eloszlása:* $\hat{t}_{(n-1)}$ $(n-1)$ szabadságfokú Student-eloszlású valószínűségi változó.

Az egymintás *t*-próbát általában csak kis minta elemszám esetén használjuk, mivel nagy elemszámú minta esetén az eloszlása közelítőleg standard normális eloszlású és az $u$ próba megfelelő alakjával helyettesíthető.

Érdekességképp megjegyezzük, hogy a Student-eloszlást V. Gosset vezette be, aki eredményeit Student álnéven publikálta.

### 3.6.4. A kétmintás t-próba

*Alkalmazása:* Ismeretlen $D(\xi)=\sigma_1$ és $D(\eta)=\sigma_2$, de megegyező $\sigma=\sigma_1=\sigma_2$ szórású és ismeretlen $M(\xi)=m_1$ és $M(\eta)=m_2$ várható értékű normális eloszlású $\xi$ és $\eta$ valószínűségi változók várható értékének különbségére vonatkozó $m_0=m_1-m_2$ hipotézis helyességének ellenőrzése.
A próba statisztika:

$$\hat{t}_{(n_1+n_2-2)}=\frac{\hat{m}_1-\hat{m}_2-m_0}{\sqrt{\frac{(n_1-1)\cdot\hat{s}_1^2+(n_2-1)\cdot\hat{s}_2^2}{n_1+n_2-2}}\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}},$$

ahol $n_1$ a $\xi$-re, $n_2$ pedig az $\eta$-ra vonatkozó független minta elemszáma.

*Eloszlása:* $\hat{t}_{(n_1+n_2-2)}$ $(n_1+n_2-2)$ szabadságfokú Student-eloszlású valószínűségi változó.

Általában csak kis minta elemszám esetén használjuk, mivel nagy elemszámú minta esetén az eloszlása közelítőleg standard normális eloszlású és az $u$ próba megfelelő alakjával helyettesíthető.

Ha a $\sigma_1$ és a $\sigma_2$ szórások nem egyenlők, akkor az $u$ próbánál már ismertetett Welch-féle

$$\hat{w}_f=\frac{\hat{m}_1-\hat{m}_2-m_0}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1}+\frac{\hat{\sigma}_2^2}{n_2}}}$$

próba statisztikát alkalmazhatjuk. A $\hat{w}_f$ próba statisztika

$$f = \frac{1}{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}} \cdot \left( \frac{1}{n_1 - 1} \cdot \left( \frac{\hat{\sigma}_1^2}{n_1} \right)^2 + \frac{1}{n_2 - 1} \cdot \left( \frac{\hat{\sigma}_2^2}{n_2} \right)^2 \right) \text{-}$$

szabadságfokú Student-eloszlású valószínűségi változó. (*f* általában nem egész, ekkor kerekítjük.)

### 3.6.5. Az F-próba

*Alkalmazása:* Ismeretlen $D(\xi) = \sigma_1$ és $D(\eta) = \sigma_2$ szórású normális eloszlású $\xi$ és $\eta$ valószínűségi változók szórásainak megegyezőségére vonatkozó $\frac{\sigma_1}{\sigma_2} = 1$ hipotézis helyességének ellenőrzése.

A próba statisztika:

$$\hat{F}_{(n_1 - 1, n_2 - 1)} = \frac{\hat{s}_1^2}{\hat{s}_2^2},$$

ahol $n_1$ a $\xi$-re, $n_2$ pedig az $\eta$-ra vonatkozó független minta elemszáma.

*Eloszlása*: $\hat{F}_{(n_1 - 1, n_2 - 1)}$ $(n_1 - 1, n_2 - 2)$ szabadságfokú $F$-eloszlású valószínűségi változó.

Az $F$-próbával pl. a kétmintás $t$-próba alkalmazása előtt a $\xi$ és $\eta$ valószínűségi változók szórásának egyezőségét vizsgálhatjuk.

A gyakorlatban az $\hat{F}_{(n_1 - 1, n_2 - 1)}$ értékét mindig 1-nél nagyobbnak választjuk, azaz az

$$F = \max\left( \frac{\hat{s}_1^2}{\hat{s}_2^2}, \frac{\hat{s}_2^2}{\hat{s}_1^2} \right) \geq 1$$

statisztikát alkalmazzuk.

**1. példa.** (Egymintás $u$-próba.) Egy elektronikus készülék a hirdetések szerint az információt az áram kikapcsolása után még 70-től 90 óráig képes tárolni. A gyártó állításának ellenőrzésére $n = 250$ mérést végeztek és következő eredményeket kapták: $\hat{m}_n = 78{,}73$ és $\hat{s}_n = 10{,}22$. Egyoldali $u$-próbát alkalmazva 95%-os szignifikancia szinten igaz-e, hogy a tárolás várható értéke legalább 80 óra?

**Megoldás.** A null-hipotézis: $H_0 = \{m \geq m_0\}$, azaz az tárolási időt leíró ismeretlen $\xi$ valószínűségi változó $m$ várható értéke legalább $m_0 = 80$.

$\xi$ szórását nem ismerjük, de a minta elemszáma nagy, így azt a korrigált tapasztalati szórással becsüljük. A próba statisztika így

$$\hat{u} = \frac{\hat{m}_n - m_0}{\frac{\hat{s}_n}{\sqrt{n}}},$$

helyettesítési értéke pedig

$$\hat{u} = \frac{78{,}73 - 80}{\frac{10{,}22}{\sqrt{250}}} \approx -1{,}965.$$

Egyoldali próbát alkalmazunk, $\varepsilon = 1 - 0{,}95 = 0{,}05$ és az elfogadási tartomány $(u_\varepsilon, +\infty) = (u_{0,05}, +\infty)$ alakú. Az $u_{0,05}$ értékét a $\Phi(u_{0,05}) = 0{,}05$ feltételből határozzuk meg, azaz $u_{0,05} = \Phi^{-1}(0{,}05) = -1{,}6448$ lesz.
Az MS Excel használata esetén $u_{0,05}$ értékét az

= INVERZ.STNORM(0,05)

kifejezés kiértékelésével kaphatjuk meg.
Mivel $\hat{u} = -1{,}965 \notin (-1{,}6448, +\infty) = (u_{0,05}, +\infty)$ ezért a null-hipotézist elvetjük, azaz az elektronikus készülék információ tárolásának várható értéke nem haladja meg a 80 órát.

**2. példa.** (Kétmintás $u$-próba.) Egy vizsgálatot végeztek abból a célból, hogy megbecsüljék bizonyos vitaminok étrendhez adásának hatását. Ebből a célból 64 kéthetes patkány kapott az étrendjéhez vitamin kiegészítést négyhetes időtartam alatt. Ezt követően a tömegüket lemérték. Egy 36 patkányból álló, ugyanolyan korú kontroll csoport ugyanezen idő alatt a szokásos étrendet kapta és a tömegüket négy hét után szintén feljegyezték.

| minta típusa | elemszáma | empirikus közép | korrigált empirikus szórás |
|---|---|---|---|
| Patkányok vitamin kiegészítővel | 64 | 89,6 g | 12,96 g |
| Patkányok szokásos étrenddel | 36 | 83,5 g | 11,41 g |

A mintákat megegyező szórásúnak tekintve, vizsgáljuk meg 95%-os szignifikancia szinten, hogy a vitamin kiegészítést kapott patkányok hathetes korukban nagyobb tömegűek-e, mint azok, amelyek csak szokásos étrendet kaptak?

**Megoldás.** Jelöljük $\xi$-vel a vitaminokkal kiegészített étrendű és $\eta$-val a szokásos étrendű patkányok súlygyarapodását jellemző valószínűségi változót. A null-hipotézis: $H_0 = \{m_1 - m_2 > m_0\}$, ahol $m_0 = 0$, azaz a $\xi$ valószínűségi változó várható értéke nagyobb, mint az $\eta$-é.

$\xi - \eta$ szórását nem ismerjük, de a minta elemszáma nagy, így azt a mintából számított szórással becsüljük. A próba statisztika így

$$\hat{u} = \frac{\hat{m}_1 - \hat{m}_2 - m_0}{\sqrt{\dfrac{(n_1 - 1)\cdot \hat{s}_1^2 + (n_2 - 1)\cdot \hat{s}_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}},$$

helyettesítési értéke pedig

$$\hat{u} = \frac{89{,}6 - 83{,}5 - 0}{\sqrt{\dfrac{(64-1)\cdot(12{,}96)^2 + (36-1)\cdot(11{,}41)^2}{64 + 36 - 2}} \cdot \sqrt{\dfrac{1}{64} + \dfrac{1}{36}}} \approx 2{,}3558.$$

Egyoldali próbát alkalmazunk, $\varepsilon = 1 - 0{,}95 = 0{,}05$ és az elfogadási tartomány $(u_\varepsilon, +\infty) = (u_{0,05}, +\infty)$ alakú. $u_{0,05}$ értékét a $\Phi(u_{0,05}) = 0{,}05$ feltételből határozzuk meg, azaz $u_{0,05} = \Phi^{-1}(0{,}05) = -1{,}6448$ lesz.

Az MS Excel használata esetén $u_{0,05}$ értékét az

= INVERZ.STNORM(0,05)

kifejezés kiértékelésével kaphatjuk meg.

Mivel $\hat{u} = 2{,}3558 \in (-1{,}6448; +\infty) = (u_{0,05}, +\infty)$ ezért a null-hipotézist elfogadjuk, azaz a vitaminokkal kiegészített étrendű patkányok súlygyarapodása átlagosan nagyobb, mint a szokásos étrendűeké.

Ha a kétféle táplálást ugyanazokon a patkányokon végezték volna el (más, azonos hosszúságú időszakokban), akkor ún. *önkontrollos* vizsgálatot végezhetnénk. A null-hipotézis az lenne, hogy a kétféle diéta hatására történt súlygyarapodás különbsége 0, melyet egymintás *u*-próbával ellenőrizhetnénk.

**3. példa.** (Egymintás $t$-próba.) Egy véletlenszerűen vett, nyolc nőtől származó minta koleszterin szintjei mmol/l-ben a következők:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 3,1 | 2,8 | 1,5 | 1,7 | 2,4 | 1,9 | 3,3 | 1,6 |

Feltéve, hogy a koleszterin szint normális eloszlást követ, 98%-os szignifikancia szinten igaz-e, hogy a minta 3,1 mmol/l koleszterinátlagú népességtől származik?

**Megoldás**. A null-hipotézis: $H_0 = \{m = m_0\}$, azaz a koleszterin szintet leíró ismeretlen $\xi$ valószínűségi változó $m$ várható értéke $m_0 = 3{,}1$.

Az alkalmazott Student-eloszlás szabadságfoka $n - 1 = 7$. A próba statisztika:

$$\hat{t}_{(n-1)} = \frac{\hat{m}_n - m_0}{\frac{\hat{s}_n}{\sqrt{n}}},$$

helyettesítési értéke pedig

$$\hat{t}_{(7)} = \frac{2{,}2875 - 3{,}1}{\frac{0{,}712014}{\sqrt{8}}} \approx -3{,}2276.$$

Az MS Excel használata esetén $\hat{m}_8$ értékét az

$$= \text{ÁTLAG}(3{,}1;2{,}8;1{,}5;1{,}7;2{,}4;1{,}9;3{,}3;1{,}6),$$

$\hat{s}_8$ értékét a

$$= \text{SZÓRÁS}(3{,}1;2{,}8;1{,}5;1{,}7;2{,}4;1{,}9;3{,}3;1{,}6)$$

kifejezés kiértékelésével számíthatjuk ki.

Kétoldali próbát alkalmazunk, $\varepsilon = 1 - 0{,}98 = 0{,}02$ és az elfogadási tartomány $\left(t_{\varepsilon/2}, t_{1-\varepsilon/2}\right) = \left(t_{0,01}, t_{0,99}\right)$ alakú.

A $k$ szabadságfokú Student-eloszlás esetén az eloszlásfüggvény helyett általában az

$$\widetilde{T}_k(x) = P\left(|\xi| > x\right) = 2 - 2 \cdot T_k(x) \qquad \left(x \in (0, +\infty)\right)$$

függvényt adják meg, ahol $T_k$-val a Student-eloszlás eloszlásfüggvényét jelöltük. Az MS Excel-ben a T.ELOSZLÁS név alatt is ezt a függvényt találjuk és az INVERZ.T függvény ennek inverzét jelöli. A $\widetilde{T}_k$ függvény felhasználásával $T_k$ értékeit a

$$T_k(x) = \begin{cases} \dfrac{2 - \widetilde{T}_k(x)}{2}, & \text{ha} \quad x \ge 0, \\ 1 - T_k(-x), & \text{ha} \quad x < 0 \end{cases}$$

képlettel határozhatjuk meg.

$\widetilde{T}_k$ inverzének felhasználásával $T_k$ inverzét a

$$(T_k)^{-1}(p) = \begin{cases} \left(\widetilde{T}_k\right)^{-1}(2 - 2 \cdot p), & \text{ha} \quad p \ge \dfrac{1}{2}, \\ -(T_k)^{-1}(1 - p), & \text{ha} \quad p < \dfrac{1}{2} \end{cases}$$

formulával számíthatjuk ki.

Az MS Excel használata esetén $t_{0,99} = 2{,}9979$, amit az

$$= \text{INVERZ.T}(2 - 2 * 0{,}99;7)$$

kifejezés kiértékelésével kapunk. A Student-eloszlás szimmetriája miatt $t_{\varepsilon/2} = 1 - t_{1-\varepsilon/2}$ és így $t_{0,01} = -2{,}9979$.
Mivel $\hat{t}_{(7)} = -3{,}22769 \notin (-2{,}9979;\ 2{,}9979) = (t_{0,01}, t_{0,99})$ ezért a null-hipotézisünket elvetjük, azaz a minta nem 3,1 koleszterin átlagú népességtől származik.

**4. példa.** (Kétmintás *t*-próba.) A füstgáz por tartalmát vizsgálták két szilárd fűtőanyaggal működő kazán típus esetén. E célból tizenhárom *A* és kilenc *B* típusú kazánt vizsgáltak meg megegyező tüzelési feltételek mellett. A vizsgálati periódus után a kéményekbe helyezett porcsapdákra a 22 kazán esetén az alábbi mennyiségű por rakódott le (grammban mérve).

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *A* típus | 73,1 | 56,4 | 82,1 | 67,2 | 78,7 | 75,1 | 48,0 | 53,3 | 55,5 |
| | 61,5 | 60,6 | 55,2 | 63,1 | | | | | |
| *B* típus | 53,0 | 39,3 | 55,8 | 58,8 | 41,2 | 66,6 | 46,0 | 56,4 | 58,9 |

Feltéve, hogy a por lerakódása normális eloszlást követ és a minták szórása megegyezik, 95%-os szignifikancia szinten vizsgáljuk meg, hogy van-e lényeges különbség a porlerakódás mértékében a két kazán típus esetében. Ha van, akkor melyiknek nagyobb a porkibocsátása?

**Megoldás.** Jelöljük $\xi$-vel az *A* típusú, és $\eta$-val a *B* típusú kazánok porkibocsátását jellemző valószínűségi változót.
*a)* A null-hipotézis: $H_0 = \{m_1 - m_2 = m_0\}$, ahol $m_0 = 0$, azaz a $\xi$ és az $\eta$ valószínűségi változó várható értéke megegyezik.

Az alkalmazott Student-eloszlás szabadságfoka $n_1 + n_2 - 2 = 20$.
A próba statisztika

$$\hat{t}_{(n_1+n_2-2)} = \frac{\hat{m}_1 - \hat{m}_2 - m_0}{\sqrt{\dfrac{(n_1-1)\cdot\hat{s}_1^2 + (n_2-1)\cdot\hat{s}_2^2}{n_1+n_2-2}}\cdot\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}},$$

helyettesítési értéke pedig

$$\hat{t}_{(20)} = \frac{63,8308 - 52,8889 - 0}{\sqrt{\dfrac{12 \cdot 113,0123 + 8 \cdot 81,07861}{20}} \cdot \sqrt{\dfrac{1}{13} + \dfrac{1}{9}}} \approx 2,5203\,.$$

Kétoldali próbát alkalmazunk, $\varepsilon = 1 - 0,95 = 0,05$ és az elfogadási tartomány $\left(t_{\varepsilon/2}, t_{1-\varepsilon/2}\right) = \left(t_{0,025}, t_{0,975}\right)$ alakú.

Az MS Excel használata esetén $t_{0,975} = 2,0860$, amit az

$$= \text{INVERZ.T}(2 - 2 * 0,975;20)$$

kifejezés kiértékelésével kapunk. A Student-eloszlás szimmetriája miatt $t_{\varepsilon/2} = 1 - t_{1-\varepsilon/2}$ és így $t_{0,025} = -2,0860$.

Mivel $\hat{t}_{(20)} = 2,5203 \notin (-2,0860, 2,0860) = \left(t_{0,025}, t_{0,975}\right)$ ezért a null-hipotézisünket elvetjük, azaz lényeges különbség van a két kazán típus porkibocsátása között!

A próba függvény értéke pozitív volt és az előző esetben azt kaptuk, hogy jelentős különbség van a két kazántípus porkibocsátása között. Emiatt most ugyanezen a szignifikancia szinten a következő null-hipotézist vizsgáljuk meg:

*b)* A null-hipotézis: $H_0 = \{m_1 - m_2 > m_0\}$, ahol $m_0 = 0$, azaz a $\xi$ valószínűségi változó várható értéke nagyobb, mint az $\eta$ valószínűségi változóé.

Egyoldali próbát alkalmazunk, $\varepsilon = 1 - 0,95 = 0,05$ és az elfogadási tartomány $\left(t_{\varepsilon}, +\infty\right) = \left(t_{0,05}, +\infty\right)$ alakú.

Az MS Excel használata esetén $t_{0,95} = 1,7248$, amit az

$$= \text{INVERZ.T}(2 - 2 * 0,95;20)$$

kifejezés kiértékelésével kapunk. A Student-eloszlás szimmetriája miatt $t_{0,05} = -t_{0,95} = -1,7248$ lesz.

Mivel $\hat{t}_{(20)} = 2,5203 \in (-1,7248, +\infty) = \left(t_{0,5}, +\infty\right)$ ezért a null-hipotézisünket elfogadjuk, azaz az $A$ típusú kazán lényegesen több port bocsát ki.

A kétmintás $t$-próba az MS Excel

$$\text{T.PRÓBA}(\mathit{tömb}1;\mathit{tömb}2;\mathit{oldal};\mathit{típus})$$

függvényének felhasználásával egyszerűbben is elvégezhető. A *tömb1* és a *tömb2* paraméterek a mintákat tartalmazó cella-tartományok. Az *oldal* paraméter értéke 1, ha egyoldali és 2, ha kétoldali a próbát végzünk. A *típus* paraméter értéke 2, ha feltesszük a szórások megegyezését és 3 ha nem (Welch-féle statisztika). A függvény valószínűség értéket ad eredményül.

Ha az $A$ típusú kazánhoz tartozó mintaértékeket az A1:M1 cella-tartományba, a $B$ típusú kazánhoz tartozó mintaértékeket pedig az A2:I2 cella-tartományba írjuk, akkor az előző példa a) esetében a függvényt

$$=\text{T.PRÓBA}(\text{A1}:\text{M1};\text{A2}:\text{I2};2;2)$$

módon kiértékelve a 20 szabadságfokú Student – eloszláshoz tartozó

$$P\left(|\xi|>\left|\hat{t}_{(20)}\right|\right)=0{,}0203$$

valószínűséget kapjuk eredményül, míg a b) esetben

$$=\text{T.PRÓBA}(\text{A1}:\text{M1};\text{A2}:\text{I2};1;2)$$

módon kiértékelve a 20 szabadságfokú Student – eloszláshoz tartozó

$$P\left(\xi>\left|\hat{t}_{(20)}\right|\right)=0{,}01016$$

valószínűség lesz az eredmény.

Kétoldali próba esetén a null-hipotézis elfogadásához elegendő azt ellenőriznünk, hogy a kapott valószínűség érték nagyobb-e, mint az adott szignifikancia szinthez tartozó $\frac{\varepsilon}{2}$. Az előző példa a) esetében $\frac{\varepsilon}{2}=0{,}025$ és mivel $0{,}0203<0{,}025$ ezért elutasítjuk a null-hipotézist, mint azt a kézi számolás esetén is tettük.

Egyoldali próba esetében a null-hipotézis elfogadása egy kicsit bonyolultabb. A $(t_\varepsilon,+\infty)$ esetben a próbát csak akkor kell elvégezni, ha a próba statisztika értéke negatív, a $(-\infty,t_{1-\varepsilon})$ esetben pedig amikor pozitív, mivel az ellenkező esetekben a null-hipotézis automatikusan teljesül. A próba statisztika előjelének meghatározásakor csak a számláló előjelet kell meghatároznunk, amit az MS Excel már ismertetett ÁTLAG függvényének alkalmazásával számíthatunk ki a legegyszerűbben. Amennyiben a próbát alkalmaznunk kell, akkor ha a kapott valószínűség érték nagyobb mint az adott ε, akkor a null-hipotézist elfogadjuk, ellenkező esetben pedig elvetjük.

**5. példa.** ($F$-próba.) Vizsgáljuk meg 95%-os szignifikancia szinten, hogy az előző példa esetében jogos volt-e a szórások egyezésére vonatkozó feltevésünk!

**Megoldás.** A null-hipotézis: $H_0 = \left\{ \frac{\sigma_1}{\sigma_2} = 1 \right\}$, ahol, azaz a ξ és η valószínűségi változó szórása megegyezik.

A próbát általában csak ezzel a null-hipotézissel alkalmazzuk!

Az alkalmazott $F$-eloszlás szabadságfoka $(n_1 - 1, n_2 - 2) = (12, 8)$.

A próba statisztika

$$\hat{F}_{(n_1 - 1, n_2 - 1)} = \frac{\hat{s}_1^2}{\hat{s}_2^2},$$

helyettesítési értéke pedig

$$\hat{F}_{(12,8)} = \frac{113{,}0123}{81{,}07861} \approx 1{,}3939.$$

A próba leírásánál már említettük, hogy ha a próba statisztika értéke 1-nél kisebb, azaz $\hat{F}_{(n_1 - 1, n_2 - 1)} = \frac{\hat{s}_1^2}{\hat{s}_2^2} < 1$, akkor helyette az

$\hat{F}_{(n_2 - 1, n_1 - 1)} = \frac{\hat{s}_2^2}{\hat{s}_1^2} > 1$ statisztikát kell alkalmaznunk, mert általában csak olyan táblázatokat adnak meg, amelyekben csak az $F$ eloszlás 1-nél nagyobb értékeihez tartozó valószínűségek szerepelnek.

Az $F$-próba esetén mindig csak $\left(-\infty; F_{1-\varepsilon}\right)$ elfogadási tartományú egyoldali próbát használunk. Mivel az $F$-eloszlás csak a $(0,+\infty)$ intervallumon vesz fel pozitív értékeket, ezért ez itt $\left(0, F_{1-\varepsilon}\right)$ alakúra egyszerűsödik. A jelen $\varepsilon = 1-0{,}95 = 0{,}05$ esetben az elfogadási tartomány így $\left(0, F_{1-\varepsilon}\right) = \left(0, F_{0{,}95}\right)$ alakú lesz.

A $(k,l)$ szabadságfokú $F$-eloszlás estén az eloszlásfüggvény helyett általában az

$$\widetilde{F}_{(k,l)}(x) = P(\xi > x) = 1 - F_{(k,l)}(x),\ x \in (0,+\infty)$$

függvényt adják meg, ahol $F_{(k,l)}$-lel az $F$-eloszlás eloszlásfüggvényét jelöltük. Az MS Excelben az F.ELOSZLÁS név alatt is ezt a függvényt találjuk és az INVERZ.F függvény ennek inverzét jelöli.

Az MS Excel használata esetén $F_{0{,}95} = 3{,}2839$, amit az

$$= \text{INVERZ.F}(1\text{-}\ 0{,}95;12;8)$$

kifejezés kiértékelésével kapunk.

Mivel $\hat{F}_{(12,8)} = 1{,}3939 \in (0;3{,}2839) = \left(0;\ F_{\ 0{,}95}\right)$ ezért a null-hipotézisünket elfogadjuk, azaz a valószínűségi változók szórása megegyezik.

---

Az $F$-próba az MS Excel

$$\text{F.PRÓBA}(tömb1;tömb2)$$

függvényének felhasználásával egyszerűbben is elvégezhető. A *tömb1* és a *tömb2* paraméterek a mintákat tartalmazó cella-tartományok. A függvény egy valószínűségértéket ad eredményül.

---

Ha az $A$ típusú kazánhoz tartozó mintaértéket az A1:M1 cella-tartományba, a $B$ típusú kazánhoz tartozó mintaértékeket pedig az A2:I2 cella-tartományba írjuk, akkor az előző példa esetében a függvényt

$$= \text{F.PRÓBA}(\text{A1}:\text{M1};\text{A2}:\text{I2})$$

módon kiértékelve a $(12,8)$ szabadságfokú $F$-eloszláshoz tartozó

$$2 \cdot P\left(\xi > \hat{F}_{(12,8)}\right) = 0{,}6516$$

valószínűséget kapjuk eredményül.

Az $\hat{F}_{(n_1-1,n_2-1)} = \dfrac{\hat{s}^2_{n_1}}{\hat{s}^2_{n_2}} < 1$ esetet az F.PRÓBA automatikusan lekezeli.

Az F.PRÓBA alkalmazása esetén a null-hipotézis elfogadásához csak azt kell ellenőriznünk, hogy a kapott valószínűség érték nagyobb-e mint az adott szignifikancia szinthez tartozó ε érték kétszerese. Az előző példa esetén $\varepsilon = 0{,}05$ és mivel $0{,}6516 > 2 \cdot 0{,}05 = 0{,}1$ ezért elfogadjuk a null-hipotézist, mint azt a kézi számolás esetén is tettük.

## 3.7. Nemparaméteres próbák

Ezekben az esetekben a hipotéziseink nem a paraméterekre vonatkoznak. Az alábbiakban a három fő nemparaméteres próba típus (illeszkedésvizsgálat, homogenitásvizsgálat, függetlenségvizsgálat) $\chi^2$-eloszláson alapuló próbáit ismertetjük. Mivel a próbák aszimptotikus közelítéseken alapulnak, ezért csak nagy elemszámú minták esetén vezetnek megbízható következtetéshez.

**Illeszkedésvizsgálatnak** nevezzük azokat a statisztikai próbákat, amelyek alapján arról döntünk, hogy valamely ξ valószínűségi változó eloszlása lehet-e egy adott $F_0$ eloszlásfüggvénnyel jellemzett eloszlás.

### 3.7.1. Illeszkedésvizsgálat $\chi^2$-próbával

*Alkalmazása*: A

$$-\infty \le x_0 < x_1 < \ldots < x_{r-1} < x_r \le +\infty$$

osztópontokkal alkalmas $r$ csoport (intervallum) kijelölésével annak eldöntése, hogy egy ismeretlen eloszlású ξ valószínűségi változó eloszlása lehet-e egy adott $F_0$ eloszlásfüggvénnyel jellemzett eloszlás.

A próbastatisztika:

$$\hat{\chi}^2_{(r-1)} = \sum_{i=1}^{r} \frac{(\mu_i - n \cdot p_i)^2}{n \cdot p_i},$$

ahol $r$ a csoportok száma, $n$ a minta elemszáma, $p_i := F_0(x_i) - F_0(x_{i-1})$ a $P(x_{i-1} \le \xi < x_i)$ valószínűség feltételezett értéke, és $\mu_i$ az $[x_{i-1}, x_i)$ intervallumba eső mintaelemek száma, tehát $\sum_{i=1}^{r} \mu_i = n$.

*Eloszlása*: $\hat{\chi}^2_{(r-1)}$ közel $(r-1)$ szabadságfokú $\chi^2$-eloszlású valószínűségi változó.

A próba esetén a feltételezett és a mintából számolt relatív gyakoriságokkal becsült tényleges valószínűségek eltérését az osztópontok által meghatározott

$$A_i = \{x_{i-1} \le \xi < x_i\} \quad (i = 1, \ldots, r)$$

teljes eseményrendszer esetén vizsgáljuk.

Definíciójukból következően $\mu_i$-k binomiális eloszlású (de nem független) valószínűségi változók, és megmutatható, hogy ha a mintaelemek $n$ száma nagy, akkor a $\hat{\chi}^2_{(r-1)}$ próbastatisztika közel $(r-1)$ szabadságfokú $\chi^2$-eloszlású. A próba alkalmazásakor az $n$ elemszámának legalább olyan nagynak kell lennie, hogy az $n \cdot p_i > 10$ feltétel teljesüljön. Megjegyezzük, hogy ez a feltétel olykor a cellák $r$ számának csökkentésével is elérhető.

Ha a $p_i$ valószínűségek meghatározásakor mintából becsült ismeretlen paramétereket is felhasználunk, akkor becsléses illeszkedésvizsgálatról beszélünk. A próba szabadságfoka ekkor $(r-1)$ és a becsült paraméterek számának különbsége lesz.

**Homogenitásvizsgálatnak** nevezzük azokat a statisztikai próbákat, amelyek alapján arról döntünk, hogy két valószínűségi változó, $\xi$ és $\eta$ egyforma eloszlású-e vagy sem.

### 3.7.2. Homogenitásvizsgálat $\chi^2$-próbával

*Alkalmazása:* A

$$-\infty \le x_0 < x_1 < \ldots < x_{r-1} < x_r \le +\infty$$

osztópontokkal alkalmas $r$ csoport (intervallum) kijelölésével annak eldöntése, hogy két ismeretlen eloszlású $\xi$ és $\eta$ valószínűségi változó egyforma eloszlású-e vagy sem.
A próbastatisztika:

$$\hat{\chi}^2_{(r-1)} = \sum_{i=1}^{r} \frac{\left(\frac{\mu_i}{m} - \frac{\nu_i}{n}\right)^2}{\frac{\mu_i + \nu_i}{m \cdot n}} = m \cdot n \cdot \sum_{i=1}^{r} \frac{(n \cdot \mu_i - m \cdot \nu_i)^2}{\mu_i + \nu_i},$$

ahol $r$ a csoportok száma, $m$ a $\xi$-re és $n$ az $\eta$-ra vonatkozó minta elemszáma, $\mu_i$ a $\xi$-re és $\nu_i$ az $\eta$-ra vonatkozó minta $[x_{i-1}, x_i)$ intervallumba eső mintaelemeinek száma: nyilván $\sum_{i=1}^{r} \mu_i = m$, és $\sum_{i=1}^{r} \nu_i = n$.

*Eloszlása:* $\hat{\chi}^2_{(r-1)}$ közel $(r-1)$ szabadságfokú $\chi^2$-eloszlású valószínűségi változó.

A próba estén a mintából számolt relatív gyakoriságokkal becsült tényleges valószínűségek eltérését, az osztópontok által meghatározott

$$A_i = \{x_{i-1} \le \xi < x_i\} \ (i = 1, \ldots, r) \text{ és } B_i = \{x_{i-1} \le \eta < x_i\} \ (i = 1, \ldots, r)$$

teljes eseményrendszerek esetén vizsgáljuk.

A statisztikai függvényben szereplő $\frac{\mu_i}{m}$ relatív gyakoriság a $P(A_i)$, a $\frac{\nu_i}{n}$ pedig a $P(B_i)$ $(i = 1, \ldots, r)$ valószínűség mintából történő becslése.

**Függetlenségvizsgálatnak** nevezzük azokat a statisztikai próbákat, amelyek alapján arról döntünk, hogy két valószínűségi változó, $\xi$ és $\eta$ független-e vagy sem.

### 3.7.3. Függetlenségvizsgálat $\chi^2$-próbával

*Alkalmazása:* Az *X* tengelyen felvett

$$-\infty \le x_0 < x_1 < \ldots < x_{r-1} < x_r \le +\infty$$

és az *Y* tengely felvett

$$-\infty \le y_0 < y_1 < \ldots < y_{s-1} < y_s \le +\infty$$

osztópontokkal az *XY* síkon alkalmas $r \cdot s$ csoport (téglalap) kijelölésével annak eldöntése, hogy két ismeretlen eloszlású $\xi$ és $\eta$ valószínűségi változó független-e vagy sem.
A próbastatisztika:

$$\hat{\chi}^2_{(r-1)\cdot(s-1)} = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(\mu_{ij} - \dfrac{\mu_{i*} \cdot \mu_{*j}}{n}\right)^2}{\dfrac{\mu_{i*} \cdot \mu_{*j}}{n}},$$

ahol $r \cdot s$ a csoportok száma, $n$ a $(\xi, \eta)$-ra vonatkozó minta elemszáma, $\mu_{ij}$ az $[x_{i-1}, x_i) \times [y_{j-1}, y_j)$ téglalapba, $\mu_{i*} = \sum_{j=1}^{s} \mu_{ij}$ az $[x_{i-1}, x_i) \times [y_0, y_s)$ sávba, $\mu_{*j} = \sum_{i=1}^{r} \mu_{ij}$ pedig az $[x_0, x_r) \times [y_{j-1}, y_j)$ sávba eső mintaelemek száma, azaz $\sum_{i=1}^{r} \sum_{j=1}^{s} \mu_{ij} = n$.

*Eloszlása:* $\hat{\chi}^2_{(r-1)\cdot(s-1)}$ közel $(r-1)\cdot(s-1)$ szabadságfokú $\chi^2$-eloszlású valószínűségi változó.

Ha $\xi$ és $\eta$ független valószínűségi változók, akkor

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

tetszőleges

$$A=\{x\le\xi<x+\Delta x\} \text{ és } B=\{y\le\eta<y+\Delta y\}$$

események esetén. A próba alkalmazásakor ezen egyenlőségek teljesülését csak a megadott osztópontok által meghatározott

$$A_i=\{x_{i-1}\le\xi<x_i\}\quad(i=1,\ldots,r),\ B_j=\{y_{j-1}\le\eta<y_j\}\quad(j=1,\ldots,s)$$

teljes eseményrendszerek esetén vizsgáljuk, az egyes valószínűségek relatív gyakoriságokkal való közelítését alkalmazva. Tehát csak a

$$P(A_i\cdot B_j)=P(A_i)\cdot P(B_j)$$

egyenlőségek teljesülését ellenőrizzük. Az egyes valószínűségek helyett pedig azok

$$P(A_i\cdot B_j)\approx\frac{\mu_{ij}}{n},\quad P(A_i)\approx\frac{\mu_{i*}}{n},\quad P(B_j)\approx\frac{\mu_{*j}}{n}$$

a relatív gyakoriságokkal való közelítésével számolunk.

Ha a próbát alkalmazva a valószínűségi változók függetlenségére vonatkozó null-hipotézisünket el kell vetni, akkor a függőségük mértékét a következő fejezetben ismertetésre kerülő korreláció- és regresszió analízissel vizsgálhatjuk.

**1. példa.** (Illeszkedésvizsgálat.) Egy adott virág színenkénti eloszlására vonatkozó genetika elmélet szerint, a virág rózsaszín, fehér és kék színű előfordulásainak aránya 3:2:5. 100 palánta esetén az alábbi eredményeket kapták:

| Szín: | rózsaszín | fehér | kék |
|---|---|---|---|
| Palánták száma: | 24 | 14 | 62 |

Vizsgáljuk meg 98%-os szignifikancia szinten, hogy az elméleti és mért gyakoriságok lényegesen különböznek-e!

**Megoldás**. A null-hipotézis: $H_0 = \{$a megfigyelt és az elméleti gyakoriságok nem különböznek lényegesen$\}$. A rózsaszint 1-gyel, a fehéret 2-vel, és a kéket 3-mal számozva a virág színenkénti eloszlását jellemző diszkrét $\xi$-val valószínűségi változó lehetséges értékei:

$$\begin{aligned} \xi &= 1, \quad \text{ha a virág rószaszín,} \\ \xi &= 2, \quad \text{ha a virág kék,} \\ \xi &= 3, \quad \text{ha a virág fehér.} \end{aligned}$$

Az elméleti (feltételezett) $F_0$ eloszlásfüggvénye pedig a megadott 3:2:5 színeloszlási arány alapján

$$F_0(x) = P_0(\xi < x) = \begin{cases} 0, & \text{ha } x < 1, \\ \dfrac{3}{10}, & \text{ha } 1 \le x < 2, \\ \dfrac{5}{10}, & \text{ha } 2 \le x < 3, \\ 1, & \text{ha } 3 \le x. \end{cases}$$

Ennek megfelelően az

$$x_0 = 1,\ x_1 = 2,\ x_2 = 3,\ x_3 = 4$$

osztópontokkal az

$$[1,2),\ [2,\ 3),\ [3,4)$$

$r = 3$ csoport (intervallum) kijelölésével az

$$A_1 = \{1 \le \xi < 2\},\ A_2 = \{2 \le \xi < 3\}, A_3 = \{3 \le \xi < 4\}$$

teljes eseményrendszert kapjuk.
Az $A_i$ $(i = 1, 2, 3)$ események feltételezett $p_i$ valószínűségei

$$p_1 = F_0(2) - F_0(-\infty) = \frac{3}{10},$$

$$p_2 = F_0(3) - F_0(2) = \frac{2}{10},$$

$$p_3 = F_0(+\infty) - F_0(3) = \frac{5}{10},$$

az $n=100$ elemű mintából kapott gyakoriságértékek pedig

$$\mu_1=24,\ \mu_2=14,\ \mu_3=62.$$

A színeket önkényesen számoztuk meg, bármilyen más számozásuk is jó lett volna. Arra azonban figyelni kell, hogy az elméleti eloszlásfüggvényt és az $x_i$ osztópontokat a számozással összhangban írjuk fel!

Az alkalmazott $\chi^2$-eloszlás szabadságfoka $r-1=2$.
A próbastatisztika

$$\hat{\chi}^2_{(r-1)}=\sum_{i=1}^{r}\frac{(\mu_i-n\cdot p_i)^2}{n\cdot p_i},$$

helyettesítési értéke pedig

$$\hat{\chi}^2_{(2)}=\frac{\left(24-100\cdot\frac{3}{10}\right)^2}{100\cdot\frac{3}{10}}+\frac{\left(14-100\cdot\frac{2}{10}\right)^2}{100\cdot\frac{2}{10}}+\frac{\left(62-100\cdot\frac{5}{10}\right)^2}{100\cdot\frac{5}{10}}$$
$$=5{,}88.$$

$\chi^2$-próba esetében mindig csak $\left(-\infty,\chi^2_{1-\varepsilon}\right)$ elfogadási tartományú egyoldali próbát alkalmazunk: mivel a $\chi^2$-eloszlás csak a $(0,+\infty)$ intervallumon vesz fel pozitív értékeket ezért ez itt $\left(0,\chi^2_{1-\varepsilon}\right)$ alakúra egyszerűsödik. A jelen $\varepsilon=1-0{,}98=0{,}02$ esetben az elfogadási tartomány így $\left(0,\chi^2_{1-\varepsilon}\right)=\left(0,\chi^2_{0,98}\right)$ alakú.

A $k$ szabadságfokú $\chi^2$-eloszlás esetén az eloszlásfüggvény helyett általában a

$$\tilde{\chi}^2_k(x)=P(\xi>x)=1-\chi^2_k(x)\qquad (x\in(0,+\infty))$$

függvényt adják meg, ahol $\chi^2_k$-val a $\chi^2$-eloszlás eloszlásfüggvényét jelöltük. Az MS Excel-ben KHI.ELOSZLÁS név alatt is ezt a függvényt találjuk és az INVERZ.KHI függvény ennek inverzét jelöli.

Az MS Excel használata esetén $\chi^2_{0,98} = 7{,}824$, amit az

$$= \text{INVERZ.KHI}(1 - 0{,}98;2)$$

kifejezés kiértékelésével kapunk.

Mivel $\hat{\chi}^2_{(2)} = 5{,}88 \in (0,\ 7{,}824) = \left(0,\ \chi^2_{0,98}\right)$ ezért a null-hipotézisünket elfogadjuk, azaz az elméleti és a megfigyelt gyakoriságok nem különböznek jelentősen.

**2. példa.** (Homogenitásvizsgálat.) A Tisza Szegednél mért évi maximális vízállásaira 1876–1925 és 1926–1975 között a következő eredményeket kapták.

| Maximális vízállás méterben (V) | Gyakoriság 1876–1925 között $(\mu_i)$ | Gyakoriság 1926–1975 között $(\nu_i)$ |
|---|---|---|
| $V < 5$ | 5 | 10 |
| $5 \le V < 6$ | 11 | 11 |
| $6 \le V < 7$ | 13 | 13 |
| $7 \le V < 8$ | 13 | 10 |
| $8 \le V$ | 8 | 6 |
| Összesen | $m = 50$ | $n = 50$ |

Vizsgáljuk meg 95%-os szignifikancia szinten, hogy a Tisza Szegednél mért évi maximális vízállásai ugyanazt az eloszlást követték-e 1876–1925, mint 1926–1975 között!

**Megoldás.** A null-hipotézis: $H_0 = \{$a Tisza Szegednél mért évi vízállás maximumai ugyanazt az eloszlást követték 1876–1925, mint 1926–1975 között.$\}$

Az 1876–1925 közötti években mért maximális vízállásokat jellemző valószínűségi változót jelöljük $\xi$-vel, az 1926–1975 közöttit pedig $\eta$-val. Az adott gyakoriság táblázatnak megfelelően az

$$x_0 = -\infty,\ x_1 = 5,\ x_2 = 6,\ x_3 = 7,\ x_4 = 8, x_5 = \infty$$

osztópontokkal a

$$(-\infty, 5), [5, 6), [6, 7), [7, 8), [8, \infty)$$

$r = 5$ csoport (intervallum) kijelölésével az

$$A_1 = \{-\infty \le \xi < 5\},\ A_2 = \{5 \le \xi < 6\},\ A_3 = \{6 \le \xi < 7\},$$
$$A_4 = \{7 \le \xi < 8\},\ A_5 = \{8 \le \xi < \infty\}$$

és a

$$B_1 = \{-\infty \le \eta < 5\},\ B_2 = \{5 \le \eta < 6\},\ B_3 = \{6 \le \eta < 7\},$$
$$B_4 = \{7 \le \eta < 8\},\ B_5 = \{8 \le \eta < \infty\}$$

teljes eseményrendszereket kapjuk.

A magadott $\frac{\mu_i}{m}$ értékek a $P(A_i)$, a $\frac{\nu_i}{n}$ értékek a $P(B_i)$ valószínűségek közelítései $(i = 1, 2, 3, 4, 5)$.

Az alkalmazott $\chi^2$-eloszlás szabadságfoka $r - 1 = 4$.
A próbastatisztika

$$\hat{\chi}^2_{(r-1)} = \sum_{i=1}^{r} \frac{\left(\frac{\mu_i}{m} - \frac{\nu_i}{n}\right)^2}{\frac{\mu_i + \nu_i}{m \cdot n}} = \frac{1}{m \cdot n} \sum_{i=1}^{r} \frac{(n \cdot \mu_i - m \cdot \nu_i)^2}{\mu_i + \nu_i},$$

helyettesítési értéke pedig

$$\hat{\chi}^2_{(4)} = \frac{1}{50 \cdot 50} \cdot \left( \frac{(50 \cdot 5 - 50 \cdot 10)^2}{15} + \frac{(50 \cdot 11 - 50 \cdot 11)^2}{22} + \frac{(50 \cdot 13 - 50 \cdot 13)^2}{26} + \frac{(50 \cdot 13 - 50 \cdot 10)^2}{23} + \frac{(50 \cdot 8 - 50 \cdot 6)^2}{14} \right) \approx$$
$$\approx 2{,}3437.$$

$\chi^2$-próba esetében mindig csak $\left(-\infty, \chi^2_{1-\varepsilon}\right)$ elfogadási tartományú egyoldali próbát alkalmazunk: mivel a $\chi^2$-eloszlás csak a $(0,+\infty)$ intervallumon vesz fel pozitív értékeket ezért ez itt $\left(0, \chi^2_{1-\varepsilon}\right)$ alakúra egyszerűsödik. A jelen $\varepsilon = 1-0{,}95 = 0{,}05$ esetben az elfogadási tartomány így $\left(0, \chi^2_{1-\varepsilon}\right) = \left(0, \chi^2_{0{,}95}\right)$ alakú.

Az MS Excel használata esetén $\chi^2_{0{,}95} = 9{,}4877$, amit a

$$= \text{INVERZ.KHI}(1-0{,}95;4)$$

kifejezés kiértékelésével kapunk.

Mivel $\hat{\chi}^2_{(4)} = 2{,}3437 \in (0,\ 9{,}4877) = \left(0,\ \chi^2_{0{,}95}\right)$ ezért a null-hipotézisünket elfogadjuk, azaz az vízállás maximumai ugyanazt az eloszlást követték 1876–1925 között, mint 1926–1975 között.

**3. példa.** (Függetlenségvizsgálat.) Vizsgálatot végeztek annak megállapítására, hogy van-e összefüggés a szemszín és a bőr napon való leégése között. Erre vonatkozóan egy 180 fős véletlenszerű mintát véve az alábbi eredményeket kapták.

| Szemszín | Bőrérzékenység | | | Összesen |
|---|---|---|---|---|
| | magas | közepes | alacsony | |
| kék | 19 | 27 | 4 | 50 |
| barna | 1 | 13 | 16 | 30 |
| szürkés zöld | 27 | 48 | 25 | 100 |
| Összesen | 47 | 88 | 45 | |

Vizsgáljuk meg 95%-os szignifikancia szinten, hogy a szemszín és a bőr leégésre való érzékenysége között van-e összefüggés!

**Megoldás.** A null-hipotézis: $H_0 = \{$a szemszín és a bőr napon való leégésre való érzékenysége között nincs összefüggés$\}$.

A kéket 1-gyel, a barnát 2-vel, és a szürkés zöldet 3-mal számozva, a szemszínt jellemző diszkrét, $\xi$-val jelölt valószínűségi változó lehetséges értékei:

$$\begin{aligned} \xi &= 1, \quad \text{ha a szemszín kék,} \\ \xi &= 2, \quad \text{ha a szemszin barna,} \\ \xi &= 3, \quad \text{ha a szemszín szürkés zöld.} \end{aligned}$$

Ennek megfelelően az

$$x_0 = 1,\ x_1 = 2,\ x_2 = 3,\ x_3 = 4$$

osztópontokkal az

$$[1,2),\ [2,\ 3),\ [3,4)$$

$r = 3$ csoport (intervallum) kijelölésével az

$$A_1 = \{1 \le \xi < 2\},\ A_2 = \{2 \le \xi < 3\}, A_3 = \{3 \le \xi < 4\}$$

teljes eseményrendszert kapjuk.

A nagyot 1-gyel, a közepest 2-vel, és az alacsonyt 3-mal számozva, a bőrérzékenységet mértékét jellemző diszkrét, $\eta$-val jelölt valószínűségi változó lehetséges értékei:

$$\begin{aligned} \eta &= 1, \quad \text{ha magas,} \\ \eta &= 2, \quad \text{ha közepes,} \\ \eta &= 3, \quad \text{ha alacsony.} \end{aligned}$$

Ennek megfelelően az

$$y_0 = 1,\ y_1 = 2,\ y_2 = 3,\ y_3 = 4$$

osztópontokkal az

$$[1,2),\ [2,\ 3),\ [3,4)$$

$s = 3$ csoport (intervallum) kijelölésével a

$$B_1 = \{1 \le \eta < 2\},\ B_2 = \{2 \le \eta < 3\}, B_3 = \{3 \le \eta < 4\}$$

teljes eseményrendszert kapjuk.

A $P(A_i \cdot B_j)$, $P(A_i)$ és a $P(B_j)$ valószínűségekkel kapcsolatos gyakoriság közelítések a minta alapján $(i = 1,2,3;\ j = 1,2,3)$:

| | $B_1$ | $B_2$ | $B_3$ | |
|---|---|---|---|---|
| $A_1$ | $\nu_{11} = 19$ | $\nu_{,2} = 27$ | $\nu_{13} = 4$ | $\nu_{1*} = 50$ |
| $A_2$ | $\nu_{21} = 1$ | $\nu_{22} = 13$ | $\nu_{23} = 16$ | $\nu_{2*} = 30$ |
| $A_3$ | $\nu_{31} = 27$ | $\nu_{32} = 48$ | $\nu_{33} = 25$ | $\nu_{3*} = 100$ |
| | $\nu_{*1} = 47$ | $\nu_{*2} = 88$ | $\nu_{*3} = 45$ | |

Az $P(A_i) \cdot P(B_j)$ valószínűségekkel kapcsolatos gyakoriság közelítések a $P(A_i)$ -vel és a $P(B_j)$-vel kapcsolatos gyakoriság közelítések felhasználásával $(i = 1,2,3;\ j = 1,2,3)$:

| | $B_1$ | $B_2$ | $B_3$ | |
|---|---|---|---|---|
| $A_1$ | $\frac{\nu_{1*} \cdot \nu_{*1}}{n} = \frac{50 \cdot 47}{180}$ | $\frac{\nu_{1*} \cdot \nu_{*2}}{n} = \frac{50 \cdot 88}{180}$ | $\frac{\nu_{1*} \cdot \nu_{*3}}{n} = \frac{50 \cdot 45}{180}$ | $\nu_{1*} = 50$ |
| $A_2$ | $\frac{\nu_{2*} \cdot \nu_{*1}}{n} = \frac{30 \cdot 47}{180}$ | $\frac{\nu_{2*} \cdot \nu_{*2}}{n} = \frac{30 \cdot 88}{180}$ | $\frac{\nu_{2*} \cdot \nu_{*3}}{n} = \frac{30 \cdot 45}{180}$ | $\nu_{2*} = 30$ |
| $A_3$ | $\frac{\nu_{3*} \cdot \nu_{*1}}{n} = \frac{100 \cdot 47}{180}$ | $\frac{\nu_{3*} \cdot \nu_{*2}}{n} = \frac{100 \cdot 88}{180}$ | $\frac{\nu_{3*} \cdot \nu_{*3}}{n} = \frac{100 \cdot 45}{180}$ | $\nu_{3*} = 100$ |
| | $\nu_{*1} = 47$ | $\nu_{*2} = 88$ | $\nu_{*3} = 45$ | |

Az alkalmazott $\chi^2$ -eloszlás szabadságfoka $(r-1)\cdot(s-1) = 4$.

A próbastatisztika

$$\hat{\chi}^2_{(r-1)\cdot(s-1)}=\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(\mu_{ij}-\frac{\mu_{i*}\cdot\mu_{*j}}{n}\right)^2}{\frac{\mu_{i*}\cdot\mu_{*j}}{n}},$$

helyettesítési értéke pedig

$$\hat{\chi}^2_{(4)}=\frac{\left(19-\frac{50\cdot 47}{180}\right)^2}{\frac{50\cdot 47}{180}}+\frac{\left(27-\frac{50\cdot 88}{180}\right)^2}{\frac{50\cdot 88}{180}}+$$

$$+\frac{\left(4-\frac{50\cdot 45}{180}\right)^2}{\frac{50\cdot 45}{180}}+\frac{\left(1-\frac{30\cdot 47}{180}\right)^2}{\frac{30}{180}\cdot\frac{47}{180}}+\frac{\left(13-\frac{30\cdot 88}{180}\right)^2}{\frac{30\cdot 88}{180}}+$$

$$+\frac{\left(16-\frac{30\cdot 45}{180}\right)^2}{\frac{30\cdot 45}{180}}+\frac{\left(27-\frac{100\cdot 47}{180}\right)^2}{\frac{100\cdot 47}{180}}+\frac{\left(48-\frac{100\cdot 88}{180}\right)^2}{\frac{100\cdot 88}{180}}+$$

$$+\frac{\left(25-\frac{100\cdot 45}{180}\right)^2}{\frac{100\cdot 45}{180}}\approx 24{,}6082$$

.

$\chi^2$-próba esetén mindig csak $\left(-\infty,\chi^2_{1-\varepsilon}\right)$ elfogadási tartományú egyoldali próbát alkalmazunk. Mivel a $\chi^2$-eloszlás csak a $(0,+\infty)$ intervallumon vesz fel pozitív értékeket, ezért ez itt $\left(0,\chi^2_{1-\varepsilon}\right)$ alakúra egyszerűsödik. A jelen $\varepsilon = 1 - 0{,}95 = 0{,}05$ esetben az elfogadási tartomány így $\left(0, \chi^2_{1-\varepsilon}\right) = \left(0, \chi^2_{0,95}\right)$ alakú.

Az MS Excel használata esetén $\chi^2_{0,95} = 9{,}4877$, amit az

$$= \text{INVERZ.KHI}(1 - 0{,}95;4)$$

kifejezés kiértékelésével kapunk.

Mivel $\hat{\chi}^2_{(4)} = 24{,}6082 \notin (0,\ 9{,}4877) = \left(0,\ \chi^2_{0,95}\right)$ ezért a null-hipotézisünket elutasítjuk, azaz a szemszín és a bőr napon való leégésre való érzékenysége között jelentős (szignifikáns) összefüggés van

---

A függetlenségvizsgálat az MS Excel

$$\text{KHI.PRÓBA}(tömb1;tömb2)$$

függvényének felhasználásával egyszerűbben is elvégezhető, ahol a *tömb1* a $\mu_{ij}$, a *tömb2* pedig a $\frac{\mu_{i,*} \cdot \mu_{*,j}}{n}$ értékeket tartalmazó cella-tartományok. A függvény egy valószínűségértéket ad eredményül.

---

Ha az előző feladat $\mu_{ij}$ értékeit az $A1:C3$ cella-tartományba, a $\frac{\mu_{i,*} \cdot \mu_{*,j}}{n}$ értékeit pedig az $D1:F3$ cella-tartományba írjuk, akkor a függvényt

$$= \text{KHI.PRÓBA}(\text{A}1:\text{C}3;\text{D}1:\text{F}3)$$

módon kiértékelve a 4 szabadságfokú $\chi^2$-eloszláshoz tartozó

$$P\left(\xi > \hat{\chi}^2_{(4)}\right) \approx 0{,}00006$$

valószínűséget kapjuk eredményül.

Az KHI.PRÓBAalkalmazása esetén a null-hipotézis elfogadásához csak azt kell ellenőriznünk, hogy a kapott valószínűség érték nagyobb-e, mint az adott szignifikancia szinthez tartozó ε érték. Az előző példa esetén

$\varepsilon = 0{,}05$ és mivel $0{,}00006 < 0{,}05$, ezért null-hipotézist elutasítjuk, mint azt a kézi számolás esetén is tettük.

## 3.8. Korreláció- és regresszió elemzés

A gyakorlatban egy $(\xi, \eta)$ valószínűségi változó pár együttes eloszlását rendszerint nem ismerjük, a korrelációs együtthatót és a másodfajú regressziós függvényeket ezért nem tudjuk kiszámítani. Egy, a $(\xi, \eta)$ párra vonatkozó, $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ két-dimenziós mintából (ponthalmazból) kiindulva csak e mennyiségek közelítő értékeit (becsléseit) tudjuk meghatározni.

A két valószínűségi változó közötti esetleges lineáris jellegű kapcsolat erősségét jól jellemező korrelációs együtthatót értékét az alábbi tapasztalati (empirikus) korrelációs együtthatóval közelítjük.

**Definíció.** Legyen $(\xi_1, \eta_1), (\xi_2, \eta_2), \ldots, (\xi_n, \eta_n)$ a $(\xi, \eta)$ valószínűségi vektorváltozóból vett $n$-elemű minta. A $\xi$ és $\eta$ valószínűségi változók **tapasztalati** (**empirikus**) **korrelációs együtthatóján** a

$$\hat{r}_n = \hat{r}_n(\xi, \eta) = \frac{\hat{m}_n(\xi \cdot \eta) - \hat{m}_n(\xi) \cdot \hat{m}_n(\eta)}{\hat{\sigma}_n(\xi) \cdot \hat{\sigma}_n(\eta)}$$

$$= \frac{\left(\frac{1}{n} \cdot \sum_{i=1}^{n} \xi_i \cdot \eta_i\right) - \left(\frac{1}{n} \cdot \sum_{i=1}^{n} \xi_i\right) \cdot \left(\frac{1}{n} \cdot \sum_{i=1}^{n} \eta_i\right)}{\left(\frac{1}{n} \cdot \sum_{i=1}^{n} \xi_i^2 - \left(\frac{1}{n} \cdot \sum_{i=1}^{n} \xi_i\right)^2\right)^{1/2} \cdot \left(\frac{1}{n} \cdot \sum_{i=1}^{n} \eta_i^2 - \left(\frac{1}{n} \cdot \sum_{i=1}^{n} \eta_i\right)^2\right)^{1/2}}$$

hányadost értjük.

Az empirikus korrelációs együttható a korrelációs együtthatóhoz hasonló tulajdonságokkal rendelkezik:

**3.8.1. tétel.** A $\xi$ és $\eta$ valószínűségi változók $\hat{r}_n$ empirikus korrelációs együtthatójának tulajdonságai:

1. Értéke mindig $-1$ és $1$ közé esik, azaz

$$-1 \le \hat{r}_n(\xi,\eta) \le 1.$$

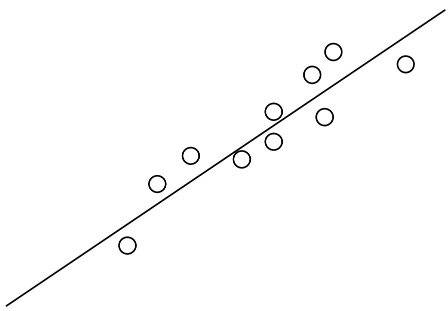
2. Ha $\xi$ és $\eta$ függetlenek, akkor

$$\hat{r}_n(\xi,\eta) = 0.$$

3. Ha a két valószínűségi változó között lineáris kapcsolat áll fenn, azaz, $\eta = a \cdot \xi + b$, ahol $a \neq 0$, akkor

$$\hat{r}_n(\xi,\eta) = 1, \text{ ha } a > 0, \text{ és } \hat{r}_n(\xi,\eta) = -1, \text{ ha } a < 0.$$

Ha a $(\xi,\eta)$ párra vonatkozó konkrét $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ kétdimenziós minta (ponthalmaz) a $\xi$ és az $\eta$ valószínűségi változók közötti lineáris összefüggésre utal, akkor azt az egyenest keressük, ami a legkisebb négyzetek elve alapján a legjobban illeszkedik a megadott $(x_i,y_i)$ $(i = 1,2,\ldots,n)$ pontokhoz. Így a valószínűségszámítási részben ismertetett, $\eta$-nak a $\xi$-re vonatkozó regressziós egyenesének egy közelítését (becslését) kapjuk.



$\xi$ és $\eta$ közötti lineáris jellegű kapcsolat

Feladatunk a következő: Keressük azt az

$$y = a \cdot x + b$$

egyenest, amelyre az

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

függvény minimális lesz.

A feladatot a kétdimenziós függvényekre vonatkozó szélsőérték-számítási módszerekkel oldjuk meg.

Írjuk fel $S(a,b)$-nek az $a$ és a $b$ szerinti parciális deriváltjait, és állapítsuk meg, hogy mely értékekre lesznek ezek zérusok.

$$\partial_a S(a,b) = -2 \cdot \sum_{i=1}^{n} (y_i - (a \cdot x_i + b)) \cdot x_i = 0,$$

$$\partial_b S(a,b) = -2 \cdot \sum_{i=1}^{n} (y_i - (a \cdot x_i + b))^2 = 0.$$

Ebből átrendezéssel adódnak az úgynevezett normálegyenletek:

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

Ezek megoldása

$$a = \hat{r}_n(\xi,\eta) \cdot \frac{\hat{\sigma}_n(\eta)}{\hat{\sigma}_n(\xi)}, \quad b = \hat{m}_n(\eta) - \hat{r}_n(\xi,\eta) \cdot \frac{\hat{\sigma}_n(\eta)}{\hat{\sigma}_n(\xi)} \cdot \hat{m}_n(\xi)$$

alakba írható.
Így a keresett egyenes egyenlete

$$y = a \cdot x + b =$$

$$= \hat{r}_n(\xi,\eta) \cdot \frac{\hat{\sigma}_n(\eta)}{\hat{\sigma}_n(\xi)} \cdot x + \left( \hat{m}_n(\eta) - \hat{r}_n(\xi,\eta) \cdot \frac{\hat{\sigma}_n(\eta)}{\hat{\sigma}_n(\xi)} \cdot \hat{m}_n(\xi) \right).$$

Hogy ez a függvény valóban a kívánt minimum tulajdonsággal rendelkezik az a

$$\begin{pmatrix} \partial_a^2 S(a,b) & \partial_b \partial_a S(a,b) \\ \partial_b \partial_a S(a,b) & \partial_b^2 S(a,b) \end{pmatrix}$$

Hesse-mátrix pozitív definitségéből következik, amit az $\partial_b^2 S(a,b) = 2 > 0$ és $\partial_a^2 S(a,b) \cdot \partial_b^2 S(a,b) - (\partial_b \partial_a S(a,b))^2 = 4 \cdot \hat{\sigma}_n^2(\xi) > 0$ egyenlőtlenségek igazolnak.

A gyakorlatban az $(x_i, y_i)$ $(i = 1, 2, \ldots, n)$ ponthalmaz alakjától függően, az $y = a \cdot x + b$ lineáris közelítésen kívül a $\xi$ és az $\eta$ valószínűségi változók közötti összefüggés közelítésére az alábbi nemlineáris közelítések is szokásosak:

| | |
|---|---|
| $y = a \cdot x^2 + b \cdot x + c$ | (másodfokú kapcsolat) |
| $y = \dfrac{a}{x + b}$ | (hiperbolikus kapcsolat) |
| $y = a \cdot e^{b \cdot x}$ | (exponenciális kapcsolat) |

Ezeken kívül más, bonyolultabb függvények is használatosak.

Némelyik kapcsolatot könnyen visszavezethető lineárisra. Exponenciális kapcsolat esetén például

$$\ln(y) = b \cdot x + \ln(a),$$

ami azt mutatja, hogy $x$ és $\ln(y)$ között lineáris a kapcsolat, és az $(x_i, \ln(y_i))$ $(i = 1, 2, \ldots, n)$ transzformált ponthalmazra így lineáris regressziós közelítést alkalmazhatunk.

**1. példa.** Azt tapasztalták, hogy az import értékének növekedésével az importanyagok fuvarköltsége lineárisan növekszik. Hét egymást követő év adatait a következő táblázat mutatja. Határozzuk meg az adatokat négyzetesen legjobban közelítő $y = a \cdot x + b$ egyenest!

| Az import értéke milliárd forintban $(y_i)$ | Az import fuvarköltsége milliárd forintban $(x_i)$ |
|---|---|
| 94,8 | 2,64 |
| 112,8 | 2,83 |
| 122,7 | 3,12 |
| 134,1 | 3,31 |
| 148,2 | 3,69 |
| 164,7 | 3,86 |
| 175,0 | 3,95 |

**Megoldás.** A fuvarköltségek alakulását a $\xi$, az import értékek alakulást pedig az $\eta$ valószínűségi változóval jelölve, a megadott táblázat a $(\xi,\eta)$-ra vonatkozó konkrét $n=7$ elemű mintát tartalmazza.
A keresett $a$ és $b$ együtthatókat a

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

egyenletrendszer megoldásával kapjuk.
Az egyenletrendszer megoldása

$$a=\hat{r}_n(\xi,\eta)\cdot\frac{\hat{\sigma}_n(\eta)}{\hat{\sigma}_n(\xi)}=$$
$$=\frac{\left(\frac{1}{n}\cdot\sum_{i=1}^{n}x_i\cdot y_i\right)-\left(\frac{1}{n}\cdot\sum_{i=1}^{n}x_i\right)\cdot\left(\frac{1}{n}\cdot\sum_{i=1}^{n}y_i\right)}{\left(\frac{1}{n}\cdot\sum_{i=1}^{n}x_i^2-\left(\frac{1}{n}\cdot\sum_{i=1}^{n}x_i\right)^2\right)^{1/2}\cdot\left(\frac{1}{n}\cdot\sum_{i=1}^{n}y_i^2-\left(\frac{1}{n}\cdot\sum_{i=1}^{n}y_i\right)^2\right)^{1/2}}\times$$
$$\times\frac{\left(\frac{1}{n}\cdot\sum_{i=1}^{n}y_i^2-\left(\frac{1}{n}\cdot\sum_{i=1}^{n}y_i\right)^2\right)^{1/2}}{\left(\frac{1}{n}\cdot\sum_{i=1}^{n}x_i^2-\left(\frac{1}{n}\cdot\sum_{i=1}^{n}x_i\right)^2\right)^{1/2}}\approx 56{,}37,$$
$$b=\left(\frac{1}{n}\cdot\sum_{i=1}^{n}y_i\right)-a\cdot\left(\frac{1}{n}\cdot\sum_{i=1}^{n}x_i\right)\approx -51{,}97.$$

Így a négyzetesen legjobban közelítő egyenes:

$$y=a\cdot x+b=56{,}37\cdot x-51{,}97$$

Az MS Excel alkalmazásával a keresett $a$ együtthatót például a

$$= \mathrm{KORREL}(\{2{,}64;2{,}83;3{,}12;3{,}31;3{,}69;3{,}86;3{,}95\};$$
$$\{94{,}8;112{,}8;122{,}7;134{,}1;148{,}2;167{,}7;175\})*$$
$$\mathrm{SZÓRÁSP}(\{94{,}8;112{,}8;122{,}7;134{,}1;148{,};167{,}7;175\})/$$
$$\mathrm{SZÓRÁSP}(\{2{,}64;2{,}83;3{,}12;3{,}31;3{,}69;3{,}86;3{,}95\}),$$

$b$ együtthatót pedig a

$$= \mathrm{ÁTLAG}(\{94{,}8;112{,}8;122{,}7;134{,}1;148{,}2;167{,}7;175\})$$
$$-56{,}37 * \mathrm{ÁTLAG}(\{2{,}64;2{,}83;3{,}12;3{,}31;3{,}69;3{,}86;3{,}95\})$$

kifejezések kiértékelésével határozhatjuk meg.

A négyzetesen legjobban közelítő egyenes az MS Excel

$$\mathrm{LIN.ILL}(tömb1;tömb2)$$

függvényének felhasználásával egyszerűbben is meghatározható, ahol a *tömb1* az $y_i$, a *tömb2* pedig az $x_i$ értékeket tartalmazó cella-tartományok. A függvény az $a$ és a $b$ értékeket tartalmazó tömböt adja eredményül.

Ha az előző feladat $y_i$ értékeit az A1:G1 cellatartományba, az $x_i$ értékeit pedig az $A2:G2$ cella-tartományba írjuk, akkor a függvényt

$$= \mathrm{INDEX}(\mathrm{LIN.ILL}(\mathrm{A1:G1;A2:G2});1;1)$$

módon kiértékelve az $a$, és

$$= \mathrm{INDEX}(\mathrm{LIN.ILL}(\mathrm{A1:G1;A2:G2});1;2)$$

módon kiértékelve a $b$ paraméter értékét kapjuk.

Az MS Excelben a LIN.ILL függvénynek a most mondottaknál sokkal általánosabb alakja is létezik.

A regresszió elemzéssel kapcsolatosan felhívjuk az Olvasó figyelmét az MS Excel TREND , LOG.ILL és NÖV függvényeire.

Önmagában is érdekes a korrelációs együttható szignifikanciájának vizsgálata. Ha $(\xi,\eta)$ együttes eloszlása normális, akkor annak a null-hipotézisnek a tesztelése, hogy a köztük levő korrelációs együttható 0, egyben $\xi$ és $\eta$ függetlenségének vizsgálatát is jelenti.

# Ajánlott irodalom

Bolla Marianna – Krámli Sándor: *Statisztikai következtetések elmélete*. 2005, Typotex.

Reimann József: *Valószínűség-elmélet és matematikai statisztika mérnököknek*. Budapest, 1992, Tankönyvkiadó.

Reimann József – Tóth Julianna: *Valószínűség-számítás és matematikai statisztika*. Budapest, 2000, Nemzeti Tankönyvkiadó Zrt.

Vetier András: *Szemléletes mérték- és valószínűség-elmélet*. Budapest, 1991, Tankönyvkiadó.

Vincze István: *Matematikai statisztika ipari alkalmazásokkal*. Budapest, 1968, Műszaki Könyvkiadó.

# Név- és tárgymutató