# Valószínűségszámítás
# és
# matematikai statisztika
# gyakorlatok

feladatok és megoldások

2006. június 6.

# Tartalomjegyzék

# 1. fejezet

# Valószínűségszámítási feladatok

## 1.1. Függetlenség, feltételes valószínűség

**1.1. Feladat.** *Egy sakk versenyen $2^N = 32$ versenyző indul. A verseny egyenes kieséses rendszer szerint, $N = 5$ fordulóban zajlik úgy, hogy minden forduló után a győzteseket véletlenszerűen párosítják, és bármelyik versenyző $\frac{1}{2}$ valószínűséggel győzi le ellenfelét (döntetlen kizárva). Mennyi annak valószínűsége, hogy két kijelölt versenyző megmérkőzik egymással?*

**Megoldás:** Vizsgáljuk először annak $p$ valószínűségét, hogy $2K$ számú versenyző párosítása során, mennyi annak valószínűsége, hogy két kijelölt játékos egymás ellen játszik? Az eseteket számoljuk meg úgy, hogy az egyik kiválasztott versenyzőhöz sorsolunk egy másikat:

$$\begin{aligned}
\text{összes eset:} \quad & 2K-1 \\
\text{kedvező eset:} \quad & 1 \\
p \;=\; & \frac{1}{2K-1}
\end{aligned}$$

Jelölje $n = 1, 2, \cdots, N$ esetén

$A_n$ a két versenyző megmérkőzik az $n$-edik fordulóban
$B_n$ mindkét versenyző nyer az $n$-edik fordulóban
$F_n$ mindkét versenyző részt vesz az $n$-edik fordulóban

akkor teljesülnek

$$\begin{aligned}
F_1 &= \Omega \qquad F_{n+1} = F_n \cap \overline{A}_n \cap B_n \quad n = 1, 2, \cdots, N-1 \\
P(A_n|F_n) &= \frac{1}{2^{N+1-n}-1} = \frac{P(A_n)}{P(F_n)} \Longrightarrow P(F_n) = P(A_n)\left(2^{N+1-n}-1\right) \quad n = 1, \cdots, N
\end{aligned}$$

$$\begin{aligned}
P(A_{n+1}) &= P(F_n \cap \overline{A}_n \cap B_n \cap A_{n+1}) = \\
&= P(F_n) \cdot P\left(\overline{A}_n|F_n\right) \cdot P\left(B_n|F_n \cap \overline{A}_n\right) \cdot P\left(A_{n+1}|F_n \cap \overline{A}_n \cap B_n\right) = \\
&= P(A_n)\left(2^{N+1-n}-1\right) \cdot \frac{2^{N+1-n}-2}{2^{N+1-n}-1} \cdot \frac{1}{4} \cdot \frac{1}{2^{N-n}-1} = P(A_n) \cdot \frac{1}{2}
\end{aligned}$$

tehát

$$P(A_n) = P(A_1)\frac{1}{2^{n-1}} \qquad n = 1, 2, \cdots, N \ .$$

A keresett esemény kizáró események uniójaként

$$A = A_1 \cup A_2 \cup \cdots \cup A_N \ ,$$

és valószínűsége

$$P(A) = \frac{1}{31}\left(\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4}\right) = \frac{1}{16} \ .$$

**1.2. Feladat.** *Szinbád, a szultánnak tett szolgálataiért, választhat egyet az $N = 100$ háremhölgy közül úgy, hogy az egyenként előtte elvonuló hölgyek valamelyikére rámutat. Tegyük fel, hogy a háremhölgyek szépségük szerint egyértelműen sorrendbe állíthatók, és Szinbád taktikája a következő: a véletlen sorrendben elvonuló hölgyek közül, az első $n = 10$ szemrevétele után azt választja, aki szebb minden korábban látottnál. Mennyi annak valószínűsége, hogy Szinbád a legszebb háremhölgyet választja?*

**Megoldás:** Vezessük be a következő eseményeket

$$\begin{array}{ccl} A & - & \text{Szinbád a legszebb háremhölgyet választja} \\ B_1 & - & \text{a legszebb hölgy az első helyen áll} \\ B_2 & - & \text{a legszebb hölgy a második helyen áll} \\ \vdots & \vdots & \vdots \\ B_N & - & \text{a legszebb hölgy az } N\text{-edik helyen áll} \end{array}$$

ahol a $B_k \quad k = 1, 2, \ldots, N$ események teljes eseményrendszert alkotnak, és

$$P(B_k) = \frac{1}{N} \quad k = 1, 2, \ldots, N \qquad P(A \mid B_k) = \begin{cases} 0 & \text{ha} \quad k = 1, 2, \ldots, n \\ \frac{n}{k-1} & \text{ha} \quad k = n+1, n+2, \ldots, N \end{cases}$$

amiből a keresett

$$P(A) = \sum_{k=n+1}^{N} \frac{n}{k-1} \cdot \frac{1}{N} = \frac{n}{N} \sum_{k=n+1}^{N} \frac{1}{k-1} = \frac{10}{100} \cdot \sum_{k=11}^{100} \frac{1}{k-1} = 0.2358409$$

valószínűség a teljes valószínűség tétellel számolható.
**Megjegyzés:** Ha az $N = 100$ esetben megkeressük azt az $n$ számot, amire $P(A)$ maximális, kapjuk az

$$\frac{n}{100} \cdot \sum_{k=n+1}^{100} \frac{1}{k-1}$$

kifejezés maximumát $n = 37$ esetén, és ekkor $P(A) = 0.37474278$. Megmutatható, hogy $N \to \infty$ esetén, $\frac{N}{n} = e = 2.718\,3$ arány adja a maximális valószínűséget, ami most $\frac{100}{37} = 2.702\,7$ .

**1.3. Feladat.** *Két testvér, $A$ és $B$, $p$ illetve $q$ valószínűséggel mond igazat. Ha $B$ azt állítja, hogy $A$ hazudik, mennyi annak valószínűsége, hogy $A$ igazat mond?*

**Megoldás:** Vezessük be az alábbi eseményeket

$$
\begin{aligned}
A_1 -&\ A \text{ mond egy igaz állítást}\\
A_2 -&\ A \text{ mond egy hamis állítást}\\
B_1 -&\ B \text{ azt állítja, hogy } A \text{ igaz állítást mond}\\
B_2 -&\ B \text{ azt állítja, hogy } A \text{ hamis állítást mond}
\end{aligned}
$$

akkor feltehetjük:

$$
\begin{aligned}
P(A_1) &= p \quad P(A_2) = 1-p\\
P(B_1 \mid A_1) &= P(B_2 \mid A_2) = q \quad P(B_1 \mid A_2) = P(B_2 \mid A_1) = 1-q
\end{aligned}
$$

amiből a keresett

$$
P(A_1 \mid B_2) = \frac{P(B_2 \mid A_1)\cdot P(A_1)}{P(B_2 \mid A_1)\cdot P(A_1) + P(B_2 \mid A_2)\cdot P(A_2)} = \frac{(1-q)\,p}{(1-q)\,p + q\,(1-p)}
$$

valószínűség a Bayes tétellel számolható.

**1.4. Feladat.** *Egy lőtéren húszan gyakorolnak, köztük* 4 *kiváló,* 7 *jó és* 9 *gyenge felkészültségű lövész van, akik* 0.9, 0.8 *illetve* 0.6 *valószínűséggel találják el a célt. Egy találomra megfigyelt lövés sikeres illetve sikertelen voltából következtessünk a lövész felkészültségére! Adjuk meg a hibás következtetés valószínűségét!*

**Megoldás:** Vezessük be az alábbi eseményeket:

$$
\begin{aligned}
A_1 -&\ \text{a megfigyelt lövés nem talál}\\
A_2 -&\ \text{a megfigyelt lövés talál}\\
B_1 -&\ \text{kiváló lövész adta le a lövést}\\
B_2 -&\ \text{közepes lövész adta le a lövést}\\
B_3 -&\ \text{gyenge lövész adta le a lövést}
\end{aligned}
$$

Foglaljuk össze a két teljes eseményrendszerrel kapcsolatos $P(A_i \cap B_j)$ és $P(B_j \mid A_i)$ valószínűségeket az alábbi táblázatban:

| $A.\setminus B.$ | $B_1$ | $B_2$ | $B_3$ | $P(A.)$ |
|---|---|---|---|---|
| $A_1$ | $0.1\cdot\frac{4}{20}=0.02$<br>$\frac{0.02}{0.27}=0.07407$ | $0.2\cdot\frac{7}{20}=0.07$<br>$\frac{0.07}{0.27}=0.25926$ | $0.4\cdot\frac{9}{20}=0.18$<br>$\frac{0.18}{0.27}=\mathbf{0.66667}$ | 0.27 |
| $A_2$ | $0.9\cdot\frac{4}{20}=0.18$<br>$\frac{0.18}{0.73}=0.24658$ | $0.8\cdot\frac{7}{20}=0.28$<br>$\frac{0.28}{0.73}=\mathbf{0.38356}$ | $0.6\cdot\frac{9}{20}=0.27$<br>$\frac{0.27}{0.73}=0.36986$ | 0.73 |
| $P(B.)$ | $\frac{4}{20}=0.20$ | $\frac{7}{20}=0.35$ | $\frac{9}{20}=0.45$ | 1.00 |

Tehát a Bayes döntés, és a hiba valószínűsége:

| $d^*$ | $P\left(A_i \cap B_{d^*(i)}\right)$ |
|---|---|
| $1 \mapsto 3$ | $0.18$ |
| $2 \mapsto 2$ | $0.28$ |
| | $0.46$ |

$$P(H_{d^*}) = 1 - 0.46 = 0.54$$

**1.5. Feladat (*).** *Egy kosárlabda játékos egymás után végez büntető dobásokat. Az elsőt bedobja, a másodikat nem, és minden további dobása akkora valószínűséggel lesz sikeres, mint amennyi a megelőző dobásokban a kosarak relatív gyakorisága. Mennyi annak valószínűsége, hogy 100 dobásból pontosan 50 kosarat fog dobni?*

**Megoldás:** Jelölje

$$A_k - \text{ a } k\text{-adik dobás sikeres}$$

akkor az az esemény, hogy $3 \leq N$ számú dobásból $n$ számú sikeres:

$$B_{N,n} = \bigcup_{c_3, c_4, \ldots, c_N} A_3^{c_3} \cap A_4^{c_4} \cap \cdots \cap A_N^{c_N} \quad 1 \leq n \leq N-1 \tag{1.1}$$

ahol a $c_3, c_4, \ldots, c_N$ sorozat $n-1$ számú "␣" szóköz karakter és $N-n-1$ számú "$c$" komplementer jel egy permutációja. A (1.1) diszjunkt únió egy tagjának valószínűsége a szorzási szabály segítségével, ha például minden szóköz elöl áll:

$$\begin{aligned} P\left(A_3 \cap A_4 \cap \cdots \cap A_{n+1} \cap A_{n+2}^c \cap A_{n+3}^c \cap \cdots A_N^c\right) = \\ = \frac{1}{2} \cdot \frac{2}{3} \cdot \cdots \cdot \frac{n-1}{n} \cdot \frac{1}{n+1} \cdot \frac{2}{n+2} \cdot \cdots \cdot \frac{N-n-1}{N-1} = \\ = \frac{1}{N-1} \cdot \frac{(n-1)! \cdot (N-n-1)!}{(N-2)!} \end{aligned}$$

Ezt az értéket kapjuk minden $\frac{(N-2)!}{(n-1)! \cdot (N-n-1)!}$ számú permutáció esetén, hiszen az egymást követő tényezők nevezői rendre

$$2, \quad 3, \quad \cdots n, \quad n+1, \quad \cdots N-1$$

és a számlálók két (egyesével) növekvő

$$1, \quad 2, \quad \cdots n-1 \quad \text{illetve} \quad 1, \quad 2, \quad \cdots N-n-1$$

sorozat permutációja lesz. Tehát

$$P\left(B_{N,n}\right) = \frac{1}{N-1}$$

## 1.2. Valószínűségi változók

**1.6. Feladat.** *A következő szerencsejátékot játszuk 10Ft befizetése ellenében: kockát dobunk, és ha az eredmény 1, 2 vagy 3, akkor még fizetünk további 5Ft-ot. Ha a dobás eredménye 4, 5 vagy 6, akkor hatszor ennyi Ft-ot kapunk. Jelölje $\xi$ a játékban elért eredményt (bevétel - kiadás),*

**a)** *adjuk meg a véletlen kísérlet matematikai modelljét!*

**b)** *adjuk meg $\xi$ eloszlását!*

**c)** *mennyi a nyerés ($\xi > 0$) valószínűsége?*

**d)** *mi a legvalószínűbb érték?*

**Megoldás:**

**a)** A véletlen kísérlet matematikai modellje a kombinatórikus v.m. ($n$ – összes eset száma, $k$ – kedvező esetek száma):

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

$$\xi(\omega) = \begin{cases} -15 & \text{ha } \omega = 1,2,3 \\ 14 & \text{ha } \omega = 4 \\ 20 & \text{ha } \omega = 5 \\ 26 & \text{ha } \omega = 6 \end{cases} .$$

Mivel most $\Omega$ minden részhalmaza esemény, a $\xi$ függvény v.v., mellyel kapcsolatos események közül elég vizsgálni a

$$\begin{aligned} \{\xi = -15\} &= \{1,2,3\} & k = 3 \\ \{\xi = 14\} &= \{4\} & k = 1 \\ \{\xi = 20\} &= \{5\} & k = 1 \\ \{\xi = 26\} &= \{6\} & k = 1 \end{aligned}$$

eseményeket.

**b)** $\xi$ értékkészlete véges, ezért diszkrét eloszlása és eloszlásfüggvénye az értékek növekvő sorrendjében megadva:

| $x$ | $P(\xi = x)$ | $F(x)$ |
|---|---|---|
| $-15$ | $\frac{3}{6}$ | 0 |
| 14 | $\frac{1}{6}$ | $\frac{3}{6}$ |
| 20 | $\frac{1}{6}$ | $\frac{4}{6}$ |
| 26 | $\frac{1}{6}$ | $\frac{5}{6}$ |
| $\sum$ | 1 | |

. (1.2)

**c)** Számítsuk ki a

$$P(\xi > 0) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

valószínűséget.

**d)** Mivel

$$\max_x \{P(\xi = x)\} = \frac{1}{2} = P(\xi = -15) \ ,$$

a keresett érték ($\xi$ un. módusza) $-15$.

**1.7. Feladat.** *Egy 2 egység hosszúságú szakaszon találomra választunk egy pontot. Az így kapott két részből, mint oldalakkal, téglalapot készítünk. Jelölje $\xi$ a téglalap területét,*

**a)** *adjuk meg a véletlen kísérlet matematikai modelljét!*

**b)** *adjuk meg $\xi$ eloszlását!*

**c)** *mennyi annak valószínűsége, hogy a terület $\frac{1}{3}$ és $\frac{2}{3}$ közé esik?*

**d)** *milyen értéknél lesz nagyobb illetve kisebb a terület azonos valószínűséggel?*

**Megoldás:**

**a)** A véletlen kísérlet matematikai modellje a geometriai v.m.($H$ – összes eset hossza, $h$ – kedvező esetek hossza):

$$\Omega = [0;2] \quad H = 2$$

$$\xi(\omega) = \omega \cdot (2 - \omega) \quad \omega \in [0;2] \ .$$

Vizsgáljuk a $\xi$-vel kapcsolatos $\{\xi < x\}$ "nívóhalmazokat", ami az

$$\omega \cdot (2 - \omega) < x \quad \omega \in [0;2] \tag{1.3}$$

egyenlőtlenség megoldáshalmaza. Az egyenlőtlenség ekvivalens alakításával kapjuk

$$0 < \omega^2 - 2\omega + x \quad \omega \in [0;2]$$

aminek megoldása, ha

**i)** $4 - 4x < 0 \Leftrightarrow x > 1$

$$\{\xi < x\} = [0;2] \quad h = 2$$

**ii)** $4 - 4x = 0 \Leftrightarrow x = 1$

$$\{\xi < x\} = [0;2] \setminus \{1\} \quad h = 2$$

**iii)** $4 - 4x > 0 \Leftrightarrow x < 1$

Ha $0 < x < 1$

$$\{\xi < x\} = \left[0; 1 - \sqrt{1-x}\right[ \cup \left]1 + \sqrt{1-x}; 2\right[ \quad h = 2\left(1 - \sqrt{1-x}\right) \ ,$$

ha pedig $x \leq 0$

$$\{\xi < x\} = \emptyset \quad h = 0 \ .$$

Tehát $\xi$ valóban valószínűségi változó.

**b)** $\xi$ eloszlása nem lehet diszkrét, ezért adjuk meg eloszlásfüggvényét:

$$F(x) = P(\xi < x) = \begin{cases} 0 & \text{ha } x \leq 0 \\ 1 - \sqrt{1-x} & \text{ha } 0 < x \leq 1 \\ 1 & \text{ha } 1 < x \end{cases}$$

ami szakaszonként folytonosan differenciálható, tehát $\xi$ folytonos eloszlású

$$f(x) = \frac{1}{2\sqrt{1-x}} \quad 0 < x < 1 \tag{1.4}$$

sűrűségfüggvénnyel.

**c)** Számítsuk ki a

$$P\left(\frac{1}{3} < \xi < \frac{2}{3}\right) = \int_{\frac{1}{3}}^{\frac{2}{3}} \frac{1}{2\sqrt{1-x}} dx = F\left(\frac{2}{3}\right) - F\left(\frac{1}{3}\right) = \frac{\sqrt{6}-\sqrt{3}}{3}$$

valószínűséget.

**d)** Mivel $\xi$ folytonos eloszlású, $P(\xi = x) = 0$ minden $x \in \mathbb{R}$ esetén, ezért keressük az

$$P(\xi < x) = P(\xi > x) = F(x) = \frac{1}{2}$$

egyenlet megoldását, amiből

$$1 - \sqrt{1-x} = \frac{1}{2} \Rightarrow x = \frac{3}{4}\ ,$$

tehát a keresett érték ($\xi$ u.n. mediánja) $\frac{3}{4}$.

**1.8. Feladat.** *Két kockát dobunk, és $\xi$ jelölje az eredmények maximumát, $\eta$ pedig a két dobás minimumát.*

**a)** *Adjuk meg $(\xi;\eta)$, $\xi$ és $\eta$ eloszlását!*

**b)** *Függetlenek-e $\xi$ és $\eta$?*

**c)** *Mennyi annak valószínűsége, hogy a maximum legalább kétszer akkora mint a minimum?*

**Megoldás:**

**a)** Mivel $(\xi;\eta)$ véges értékkészletű, a kombinatorikus v.m.-ben számolhatjuk eloszlását, amit az értékek szerint táblázatba foglalva kapjuk az együttes illetve perem eloszlásokat:

| $x \backslash y$ | 1 | 2 | 3 | 4 | 5 | 6 | $P(\xi = x)$ |
|---|---|---|---|---|---|---|---|
| 1 | $\frac{1}{36}$ | 0 | 0 | 0 | 0 | 0 | $\frac{1}{36}$ |
| 2 | $\frac{2}{36}$ | $\frac{1}{36}$ | 0 | 0 | 0 | 0 | $\frac{3}{36}$ |
| 3 | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ | 0 | 0 | 0 | $\frac{5}{36}$ |
| 4 | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ | 0 | 0 | $\frac{7}{36}$ |
| 5 | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ | 0 | $\frac{9}{36}$ |
| 6 | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ | $\frac{11}{36}$ |
| $P(\eta = y)$ | $\frac{11}{36}$ | $\frac{9}{36}$ | $\frac{7}{36}$ | $\frac{5}{36}$ | $\frac{3}{36}$ | $\frac{1}{36}$ | 1 |

(1.5)

**b)** Mivel például

$$P(\xi = 1, \eta = 1) = \frac{1}{36} \neq \frac{1}{36} \cdot \frac{11}{36} = P(\xi = 1) \cdot P(\eta = 1) \ ,$$

ezért $\xi$ és $\eta$ nem függetlenek.

**c)** A keresett valószínűség:

$$P(\xi \geq 2\eta) = \frac{2}{36} + \frac{2}{36} + \frac{2}{36} + \frac{2}{36} + \frac{2}{36} + \frac{2}{36} + \frac{2}{36} + \frac{2}{36} + \frac{2}{36} = \frac{1}{2} \ .$$

**1.9. Feladat.** *Válasszunk véletlen pontot az egység sugarú körben, jelölje a pont koordinátáit $\xi$ és $\eta$, a polárkoordinátákat pedig $\rho$ és $\varphi$.*

**a)** *Adjuk meg a véletlen kísérlet matematikai modelljét!*

**b)** *Keressünk $\xi$-vel és $\eta$-val kapcsolatos eseményeket, melyek nem fügetlenek!*

**c)** *Függetlenek-e, a $\rho$ és $\varphi$ véletlen mennyiségekkel kapcsolatos események?*

**Megoldás:**

**a)** A véletlen kísérlet matematikai modellje a geometriai v.m.($T$ – összes eset területe, $t$ – kedvező esetek területe):

$$\Omega = \left\{(x;y) \mid x^2 + y^2 \leq 1\right\}$$

$$\begin{aligned} \xi(x;y) &= x \quad (x;y) \in \Omega \\ \eta(x;y) &= y \quad (x;y) \in \Omega \end{aligned}$$

$$\begin{aligned} \rho(x;y) &= \sqrt{x^2+y^2} \quad (x;y) \in \Omega \\ \varphi(x;y) &= \arg(x,y) = \begin{cases} 0 & \text{ha } x = y = 0 \\ \frac{\pi}{2} & \text{ha } y > 0 \text{ és } x = 0 \\ \arctan\left(\frac{y}{x}\right) & \text{ha } y \geq 0 \text{ és } x > 0 \\ -\frac{\pi}{2} & \text{ha } y < 0 \text{ és } x = 0 \\ \arctan\left(\frac{y}{x}\right) + \pi & \text{ha } x < 0 \\ \arctan\left(\frac{y}{x}\right) + 2\pi & \text{ha } y < 0 \text{ és } x > 0 \end{cases} \end{aligned}$$

**b)** Mivel a pozitív területű

$$\left\{\xi > \frac{\sqrt{2}}{2}\right\} = \left\{(x;y) \mid \frac{\sqrt{2}}{2} < x \leq 1\right\} \quad \text{és} \quad \left\{\eta > \frac{\sqrt{2}}{2}\right\} = \left\{(x;y) \mid \frac{\sqrt{2}}{2} < y \leq 1\right\}$$

események kizárják egymást, ezért

$$P\left(\xi > \frac{\sqrt{2}}{2}, \eta > \frac{\sqrt{2}}{2}\right) = 0 \neq P\left(\xi > \frac{\sqrt{2}}{2}\right) \cdot P\left(\eta > \frac{\sqrt{2}}{2}\right) > 0$$

tehát a két esemény, és akkor $\xi$ és $\eta$, nem függetlenek

**c)** Adjuk meg a $\rho$-val kapcsolatos eseményeket

$$\{\rho < b\} = \begin{cases} \emptyset & \text{ha } b \leq 0 \quad t = 0 \\ \{(x;y) \mid x^2 + y^2 < b^2\} & \text{ha } 0 < b \leq 1 \quad t = b^2\pi \\ \Omega & \text{ha } 1 < b \quad t = \pi \end{cases}$$

és a $\varphi$-vel kapcsolatos eseményeket

$$\{\varphi < c\} = \begin{cases} \emptyset & \text{ha } c \leq 0 \quad t = 0 \\ c \ \text{ ívmétrékű körcikk} & \text{ha } 0 < c \leq 2\pi \quad t = \frac{c}{2\pi} \cdot \pi \\ \Omega & \text{ha } 2\pi < c \quad t = \pi \end{cases} .$$

Továbbá

$$\{\rho < b\} \cap \{\varphi < c\} = \begin{cases} \emptyset & \begin{array}{l}\text{ha } b \leq 0 \\ \quad \text{vagy } c \leq 0\end{array} & t = 0 \\ \begin{array}{c} c \ \text{ ívmétrékű, } b \text{ sugarú} \\ \text{körgyűrű-cikk}\end{array} & \begin{array}{l}\text{ha } 0 < c \leq 2\pi \\ \quad \text{és } 0 < b \leq 1\end{array} & t = \frac{c}{2\pi} \cdot b^2\pi \\ \begin{array}{c} c \ \text{ ívmétrékű} \\ \text{körcikk}\end{array} & \begin{array}{l}\text{ha } 0 < c \leq 2\pi \\ \quad \text{és } 1 < b\end{array} & t = \frac{c}{2\pi} \cdot \pi \\ \begin{array}{c} b \text{ sugarú} \\ \text{koncentrikus kör}\end{array} & \begin{array}{l}\text{ha } 0 < b \leq 1 \\ \quad \text{és } 2\pi < c\end{array} & t = b^2\pi \\ \Omega & \begin{array}{l}\text{ha } 1 < b \\ \quad \text{és } 2\pi < c\end{array} & t = \pi \end{cases}$$

tehát kapjuk

$$P\left(\rho < b, \varphi < c\right) = P(\rho < b) \cdot P(\varphi < c) \quad b, c \in \mathbb{R} ,$$

ami $\rho$ és $\varphi$ függetlenségét jelenti, mivel a megfelelő eloszlásfüggvényekre kaptuk:

$$F_{\rho,\varphi}(b;c) = F_\rho(b) \cdot F_\varphi(c) \quad b, c \in \mathbb{R} .$$

**Megjegyzés:** Vegyük észre, hogy a $\rho$ és $\varphi$ véletlen mennyiségek ugyanazon $(\xi;\eta)$ pár által meghatározottak, annak függvényei, valószínűségszámítási értelemben mégis függetlenek.

**1.10. Feladat.** *Egy diszkrét eloszlású v.v. eloszlása*

| $x$ | $P(\xi = x)$ |
|---|---|
| $-2$ | $p$ |
| $-1$ | $0.3$ |
| $0$ | $0.1$ |
| $1$ | $q$ |

*és tudjuk, hogy a v.v. negatív értéket* 0.5 *valószínűséggel vehet fel.*

**a)** *Adjuk meg $p$ és $q$ értékét!*

**b)** *Adjuk meg $\eta = \xi^2$ eloszlását!*

**c)** *Adjuk meg a $\left\{\xi^2 + \xi \leq 2\right\}$ esemény valószínűségét!*

**Megoldás:**

**a)** Mivel

$$P(\xi < 0) = p + 0.3 = 0.5 \Rightarrow p = 0.2 \ ,$$

továbbá

$$1 = 0.2 + 0.3 + 0.1 + q \Rightarrow q = 0.4 \ .$$

**b)** Mivel $\eta$ értékkészlete $\{0, 1, 4\}$,

| $y$ | $P(\eta = y)$ |
|---|---|
| $0$ | $0.1 = \sum_{x^2=0} P(\xi = x)$ |
| $1$ | $0.7 = \sum_{x^2=1} P(\xi = x)$ |
| $4$ | $0.2 = \sum_{x^2=4} P(\xi = x)$ |

.

**c)** Az

$$x^2 + x - 2 \leq 0 \quad x \in \Omega$$

egyenlőtlenség megoldáshalmaza $\{-1, 0, 1\}$, tehát

$$P(\xi^2 + \xi < 2) = \sum_{x^2+x<2} P(\xi = x) = 0.3 + 0.1 + 0.4 = 0.8 \ .$$

**1.11. Feladat.** *Legyen a $\xi$ v.v. sűrűségfüggvénye*

$$f(x) = \begin{cases} \frac{c}{\sqrt{|x|}} & \text{ha } -1 < x \leq 1 \\ 0 & \text{egyébként} \end{cases} \ . \tag{1.6}$$

**a)** *Adjuk meg $c$ értékét!*

**b)** *Adjuk meg $\xi$ és $\eta = \xi^2$ eloszlásfüggvényét, $\eta$ sűrűségfüggvényét!*

**c)** *Adjuk meg a* $\left\{\left|\xi^2-\xi\right|<\frac{3}{4}\right\}$ *esemény valószínűségét!*

**Megoldás:**

**a)** Mivel

$$1=\int_{\mathbb{R}} f = 2\int_0^1 \frac{c}{\sqrt{x}}dx = 4c \Rightarrow c=\frac{1}{4}\ .$$

**b)** $\xi$ eloszlásfüggvénye:

$$F_\xi(x)=P(\xi<x)=\begin{cases} 0 & \text{ha } x\le -1 \\ \frac{1}{2}-\frac{1}{2}\sqrt{(-x)} & \text{ha } -1<x\le 0 \\ \frac{1}{2}+\frac{1}{2}\sqrt{x} & \text{ha } 0<x\le 1 \\ 1 & \text{ha } 1<x \end{cases}\ ,$$

mivel $-1<x\le 0$ esetén

$$P(\xi<x)=\int_{-1}^{x}\frac{1}{4\sqrt{-t}}dt=\frac{1}{2}-\frac{1}{2}\sqrt{(-x)}\ ,$$

és $0<x\le 1$ esetén

$$P(\xi<x)=P(\xi<0)+\int_0^x\frac{1}{4\sqrt{t}}dt=\frac{1}{2}+\frac{1}{2}\sqrt{x}\ .$$

$\eta$ eloszlásfüggvénye:

$$F_\eta(x)=P(\xi^2<x)=\begin{cases} 0 & \text{ha } x\le 0 \\ \sqrt[4]{x} & \text{ha } 0<x\le 1 \\ 1 & \text{ha } 1<x \end{cases}$$

mivel $0<x\le 1$ esetén

$$P(\xi^2<x)=P(-\sqrt{x}<\xi<\sqrt{x})=F_\xi(\sqrt{x})-F_\xi(-\sqrt{x})=\sqrt[4]{x}\ .$$

Ebből $\eta$ sűrűségfüggvénye:

$$f_\eta(x)=F'_\eta(x)=\frac{1}{4\sqrt[4]{x^3}}\qquad 0<x\le 1.$$

**c)** Az

$$|x^2-x|<\frac{3}{4}\qquad x\in\mathbb{R}$$

egyenlőtlenséget alakítva, oldjuk meg:

$$0<x^2-x+\frac{3}{4}\quad\text{és}\quad x^2-x-\frac{3}{4}<0\ .$$

A megoldáshalmaz: $]-\frac{1}{2};\frac{3}{2}[$, így

$$\begin{aligned} P\left(\left|\xi^2-\xi\right|<\frac{3}{4}\right) &= P\left(-\frac{1}{2}<\xi<\frac{3}{2}\right)=F_\xi\left(\frac{3}{2}\right)-F_\xi\left(-\frac{1}{2}\right)= \\ &= 1-\left(\frac{1}{2}-\frac{1}{2}\sqrt{\frac{1}{2}}\right)=0.853\,55. \end{aligned}$$

**1.12. Feladat.** *Válasszunk két véletlen pontot a $[0;1]$ intervallumban, jelölje az elsőként választott értéket $\xi$, és $\eta$ legyen a két érték maximuma.*

**a)** *Adjuk meg $(\xi,\eta)$ eloszlását! Független-e $\xi$ és $\eta$?*

**b)** *Adjuk meg a peremek eloszlását!*

**c)** *Folytonos-e $(\xi,\eta)$ eloszlása?*

**Megoldás:**

**a)** Adjuk meg $(\xi,\eta)$ eloszlásfüggvényét:

$$F_{\xi,\eta}(x,y) = \begin{cases} 0 & \text{ha } x \le 0 \text{ vagy } y \le 0 \\ xy & \text{ha } 0 < x \le y \le 1 \\ y^2 & \text{ha } 0 < y \le 1 \text{ és } y < x \\ x & \text{ha } 0 < x \le 1 \text{ és } 1 < y \\ 1 & \text{ha } 1 < x \text{ és } 1 < y \end{cases}$$

Mivel például a

$$\{\xi > 0.5\} \text{ és } \{\eta < 0.5\}$$

pozitív valószínűségű események kizáróak, ezért nem lehetnek függetlenek, és így $\xi$ és $\eta$ sem független.

**b)** $\xi$ eloszlásfüggvénye:

$$F_\xi(x) = \lim_{y\to\infty} F_{\xi,\eta}(x,y) = \begin{cases} 0 & \text{ha } x \le 0 \\ x & \text{ha } 0 < x \le 1 \\ 1 & \text{ha } 1 < x \end{cases}$$

$\xi$ sűrűségfüggvénye:

$$f_\xi(x) = F'_\xi(x) = 1 \quad 0 \le x \le 1 .$$

$\eta$ eloszlásfüggvénye:

$$F_\eta(y) = \lim_{x\to\infty} F_{\xi,\eta}(x,y) = \begin{cases} 0 & \text{ha } y \le 0 \\ y^2 & \text{ha } 0 < y \le 1 \\ 1 & \text{ha } 1 < y \end{cases}$$

$\eta$ sűrűségfüggvénye:

$$f_\eta(y) = F'_\eta(y) = 2y \quad 0 \le y \le 1 .$$

**c)** $(\xi,\eta)$ eloszlása nem lehet folytonos, mert $P(\xi=\eta)=0.5$, de folytonos eloszlás esetén

$$P(\xi=\eta) = \iint_{x=y} f(x,y)dxdy = 0$$

következne, mivel nulla mértékű halmazon kell integrálni.

**Megjegyzés:** Vegyük észre, hogy a peremek ugyan folytonos eloszlásúak, az $F_{\xi,\eta}$ eloszlásfüggvény folytonos, és nulla mértékű halmazon kívül folytonosan differenciálható, mégsem folytonos az együttes eloszlás.

**1.13. Feladat.** *Legyen a* $\xi_1, \xi_2, \dots, \xi_n : \Omega \to \mathbb{R}$ *független, azonos eloszlású valószínűségi változók közös sűrűségfüggvénye* $f : \mathbb{R} \to \mathbb{R}_0^+$, *és jelölje közülük a nagyság szerint* $k$*-adikat* $\xi_k^*$ $k = 1, 2, \dots, n$.

**a)** *Adjuk meg* $\xi_n^*$ *eloszlását!*

**b)** *Adjuk meg* $(\xi_1^*, \xi_2^*, \dots, \xi_n^*) : \Omega \to \mathbb{R}^n$ *eloszlását!*

**c)** *Mutassuk meg, hogy az* $A_{kl} = \{\xi_k = \xi_l^*\}$ $l = 1, 2, \dots, n$ *események függetlenek* $(\xi_1^*, \xi_2^*, \dots, \xi_n^*)$*-től minden* $k = 1, 2, \dots, n$ *esetén!*

**Megoldás:**

**a)** Mivel

$$P(\xi_n^* < x) = P(\xi_1 < x, \xi_2 < x, \dots, \xi_n < x) = \left(\int_{-\infty}^{x} f(t)dt\right)^n$$

kapjuk $\xi_n^*$ sűrűségfüggvényét:

$$f_{\xi_n^*}(x) = n \cdot F^{n-1}(x) \cdot f(x) \quad x \in \mathbb{R}$$

ahol

$$F(x) = \int_{-\infty}^{x} f(t)dt \quad x \in \mathbb{R}$$

a közös eloszlásfüggvényt jelöli.

**b)** Mivel $(\xi_1, \xi_2, \dots, \xi_n)$ együttes eloszlása folytonos, 1 valószínűséggel különböző értékeket vesznek fel, ezért elég megadni $(\xi_1^*, \xi_2^*, \dots, \xi_n^*)$ eloszlását a

$$H = \{(x_1, x_2, \dots, x_n) \mid x_1 < x_2 < \dots < x_n \in \mathbb{R}\}$$

halmazon. Legyenek $I_1, I_2, \cdots, I_n \subset \mathbb{R}$ intervallumok olyanok, hogy

$$I_1 \times I_2 \times \cdots \times I_n \subset H \ ,$$

akkor, ha $\Pi$ jelöli az $1, 2, \dots, n$ számok permutációinak halmazát,

$$\begin{aligned} P(\xi_1^* \in I_1, \xi_2^* \in I_2, \cdots \xi_n^* \in I_n) &= \sum_{(k_1, k_2, \cdots, k_n) \in \Pi} P\left(\xi_{k_1} \in I_1, \xi_{k_2} \in I_2, \cdots \xi_{k_n} \in I_n\right) \\ &= n! \cdot \int \cdots \int_{I_1 \times I_2 \times \cdots \times I_n} f(x_1) \cdot f(x_2) \cdot \dots \cdot f(x_n) dx_1 dx_2 \cdots dx_n \ , \end{aligned}$$

tehát $(\xi_1^*, \xi_2^*, \dots, \xi_n^*)$ eloszlása folytonos

$$(x_1, x_2, \dots, x_n) \mapsto n! \cdot f(x_1) \cdot f(x_2) \cdot \dots \cdot f(x_n) \quad (x_1, x_2, \dots, x_n) \in H$$

sűrűségfüggvénnyel.

**c)** Vizsgáljuk a $k = 1$ esetet, és jelölje $A_l = A_{1l}$ $l = 1, 2, \ldots, n$. Mivel az $(A_l)_{l=1}^n$ eseményrendszer 1 valószínűséggel teljes eseményrendszert alkot, és minden tagja azonos valószínűségű, $P(A_l) = \frac{1}{n}$ $l = 1, 2, \ldots, n$. Vizsgáljuk most a

$$P\left(A_l \cap \left(\xi_1^* \in I_1, \xi_2^* \in I_2, \cdots \xi_n^* \in I_n\right)\right) =$$
$$= \int \cdots \int_{I_1 \times I_2 \times \cdots \times I_n} (n-1)! \cdot f(x_1) \cdot f(x_2) \cdot \cdots \cdot f(x_n) dx_1 dx_2 \cdots dx_n$$

valószínűséget, ami éppen $\frac{1}{n} \cdot P\left(\xi_1^* \in I_1, \xi_2^* \in I_2, \cdots \xi_n^* \in I_n\right)$, tehát $I_1 \times I_2 \times \cdots \times I_n \subset H$ esetén

$$A_l \quad \text{és} \quad \{(\xi_1^*, \xi_2^*, \cdots, \xi_n^*) \in I_1 \times I_2 \times \cdots \times I_n\}$$

függetlenek, amiből már következik az állítás.

**1.14. Feladat (*).** *Ha $\xi$ és $\eta$ független skalár valószínűségi változó, és $\xi$ eloszlásfüggvénye folytonos, mutassuk meg, hogy*

$$P(\xi = \eta) = 0.$$

**Bizonyítás.** Mivel

$$P\left(\xi = \eta\right) \leq P\left(F_\xi(\xi) = F_\xi(\eta)\right) ,$$

elég a $\xi \in \mathcal{U}(0; 1)$, és $\mathrm{im}(\eta) \subset [0; 1]$ esetben igazolni az állítést. Ekkor

$$\{\xi = \eta\} = \bigcap_{n=1}^{\infty} \bigcup_{k=0}^{n} \left( \left\{ \frac{k}{n} \leq \xi < \frac{k+1}{n} \right\} \cap \left\{ \frac{k}{n} \leq \eta < \frac{k+1}{n} \right\} \right)$$

amiből

$$\begin{aligned} P\left(\xi = \eta\right) &= \lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{n} \left[ F_\eta\left(\frac{k+1}{n}\right) - F_\eta\left(\frac{k}{n}\right) \right] = \\ &= \lim_{n\to\infty} \frac{1}{n} \left[ F_\eta\left(\frac{n+1}{n}\right) - F_\eta\left(0\right) \right] = \lim_{n\to\infty} \frac{1}{n} = 0 \end{aligned}$$

■

**1.15. Feladat (*).** *Legyenek $\xi, \eta$ független szentpétervári valószínűségi változók, azaz*

$$P(\xi = 2^k, \eta = 2^l) = \frac{1}{2^k} \cdot \frac{1}{2^l} \quad k, l \in \mathbb{N}^+ \ . \tag{1.7}$$

*Mutassuk meg, hogy van olyan $(\Omega, \mathcal{A}, P)$ v.m., melyben $(\xi, \eta)$ értelmezett, és eloszlása 1.7, továbbá megadható $(\xi', \eta')$ ugyanilyen eloszlással, és teljesül*

$$\xi + \eta = 2\xi' + \eta' \cdot \mathbf{1}_{\{\xi' \geq \eta'\}} \ . \tag{1.8}$$

**Bizonyítás.** Legyen

$$\begin{aligned} \Omega &= \left\{(2^k, 2^l, \alpha) \mid k, l \in \mathbb{N}^+, \alpha \in \{0,1\}\right\} \\ \mathcal{A} &= 2^\Omega \\ P\left(\{(2^k, 2^l, \alpha)\}\right) &= \frac{1}{2^k} \cdot \frac{1}{2^l} \cdot \frac{1}{2} \quad k, l \in \mathbb{N}^+, \alpha \in \{0,1\} \end{aligned}$$

vezessük be továbbá a

$$\begin{aligned} \xi' &: \Omega \to \mathbb{R} \quad (2^k, 2^l, \alpha) \mapsto 2^k \quad k, l \in \mathbb{N}^+, \alpha \in \{0,1\} \\ \eta' &: \Omega \to \mathbb{R} \quad (2^k, 2^l, \alpha) \mapsto 2^l \quad k, l \in \mathbb{N}^+, \alpha \in \{0,1\} \\ \delta &: \Omega \to \mathbb{R} \quad (2^k, 2^l, \alpha) \mapsto \alpha \quad k, l \in \mathbb{N}^+, \alpha \in \{0,1\} \end{aligned}$$

független valószínűségi változókat. Ekkor $(\xi', \eta')$ eloszlása (1.7) szerinti, és ha bevezetjük a

$$\begin{aligned} \xi &= \xi' \cdot \mathbf{1}_{\{\xi' < \eta'\}} + [\delta \cdot 2\xi' + (1-\delta) \cdot \eta'] \cdot \mathbf{1}_{\{\xi' \geq \eta'\}} \\ \eta &= \xi' \cdot \mathbf{1}_{\{\xi' < \eta'\}} + [\delta \cdot \eta' + (1-\delta) \cdot 2\xi'] \cdot \mathbf{1}_{\{\xi' \geq \eta'\}} \end{aligned}$$

valószínűségi változókat, akkor teljesül (1.8). Adjuk meg $(\xi, \eta)$ eloszlását:

$$\begin{aligned} &P(\xi = \eta = 2^k) = P(\xi' = 2^k, \xi' < \eta') = \textstyle\sum_{l=k+1}^{\infty} \frac{1}{2^k} \cdot \frac{1}{2^l} = \frac{1}{2^k} \cdot \frac{1}{2^k} \\ &\hspace{20em} k = 1, 2, \ldots \\ &P(\xi = 2^k, \eta = 2^l) = P(\delta = 0, \xi' = 2^{l-1}, \eta' = 2^k) = \tfrac{1}{2} \cdot \tfrac{1}{2^{l-1}} \cdot \tfrac{1}{2^k} = \tfrac{1}{2^k} \cdot \tfrac{1}{2^l} \\ &\hspace{14em} k = 1, 2, \ldots \quad l = k+1, k+2, \cdots \\ &P(\xi = 2^k, \eta = 2^l) = P(\delta = 1, \xi' = 2^{k-1}, \eta' = 2^l) = \tfrac{1}{2} \cdot \tfrac{1}{2^{k-1}} \cdot \tfrac{1}{2^l} = \tfrac{1}{2^k} \cdot \tfrac{1}{2^l} \\ &\hspace{14em} l = 1, 2, \ldots \quad k = l+1, l+2, \cdots \end{aligned}$$

tehát $(\xi, \eta)$ eloszlása is (1.7) szerinti. ■

## 1.3. Nevezetes eloszlású valószínűségi változók

**1.16. Feladat.** *Egy 20 fős tanulócsoportban 12 lány és 8 fiú van. 6 találomra válsztott felelés során, milyen határok között van a lányok száma legalább 0.8 valószínűséggel, ha*

**a)** *minden tanuló csak egyszer felelhet?*

**b)** *minden tanuló tetszőleges számúszor felelhet?*

**Megoldás:** Jelölje $\xi$ a lányok számát a 6 felelő között, akkor feltehetjük, hogy

**a)** $\xi \in \mathcal{H}yp(20; 12; 6)$, és az $n \cdot p = 6 \cdot \frac{12}{20} = 3.6$ értéket közrefogó legvalószínűbb értékekkel kezdve, számoljuk:

| $k$ | $P(\xi = k)$ | $\sum$ |
|---|---|---|
| 4 | $\frac{\binom{12}{4}\binom{8}{2}}{\binom{20}{6}} = 0.357\,59$ | $0.357\,59$ |
| 3 | $\frac{\binom{12}{3}\binom{8}{3}}{\binom{20}{6}} = 0.317\,85$ | $0.675\,44$ |
| 5 | $\frac{\binom{12}{5}\binom{8}{1}}{\binom{20}{6}} = 0.163\,47$ | $0.838\,91$ |

Tehát

$$0.80 < P(3 \le \xi \le 5) = 0.838\,91 \ .$$

**b)** $\xi \in \mathcal{B}in(6; \frac{12}{20} = 0.6)$, és az $n \cdot p = 6 \cdot \frac{12}{20} = 3.6$ értéket közrefogó legvalószínűbb értékekkel kezdve, számoljuk:

| $k$ | $P(\xi = k)$ | $\sum$ |
|---|---|---|
| 4 | $\binom{6}{4} \cdot 0.6^4 \cdot 0.4^2 = 0.311\,04$ | $0.311\,04$ |
| 3 | $\binom{6}{3} \cdot 0.6^3 \cdot 0.4^3 = 0.276\,48$ | $0.587\,52$ |
| 5 | $\binom{6}{5} \cdot 0.6^5 \cdot 0.4^1 = 0.186\,62$ | $0.774\,14$ |
| 2 | $\binom{6}{2} \cdot 0.6^2 \cdot 0.4^4 = 0.138\,24$ | $0.912\,38$ |

Tehát

$$0.80 < P(2 \le \xi \le 5) = 0.912\,38 \ .$$

**Megjegyzés:** A kérdésre más válasz is adható, például a b) esetben

$$P(\xi = 6) = \binom{6}{6} \cdot 0.6^6 \cdot 0.4^0 = 4.665\,6 \times 10^{-2}$$

értékkel számolva kapjuk

$$0.80 < P(3 \le \xi \le 6) = 0.774\,14 + 4.665\,6 \times 10^{-2} = 0.820\,8 \ .$$

**1.17. Feladat.** *Egy háztartási bizosításra átlagosan 5 év alatt egyszer kell kártérítést fizetni.*

**a)** *Mennyi annak valószínűsége, hogy egy biztosított egy adott évben nem jelentkezik kártérítésért?*

**b)** *Mennyi annak valószínűsége, hogy egy biztosításra 5 év alatt egynél több év lesz, amikor kell kártérítést fizetni?*

**Megoldás:**

**a)** Jelölje $\xi$ egy biztosított kártérítéseinek számát egy év alatt, akkor feltehetjük, hogy $\xi \in \mathcal{P}o(\frac{1}{5})$. A keresett valószínűség:

$$P(\xi = 0) = e^{-\frac{1}{5}} = 0.818\,73 \ .$$

**b)** Jelölje $\eta$ az $n = 5$ év alatt bekövetkező $p = 1 - 0.818\,73 = 0.181\,27$ valószínűségű események számát, akkor $\eta \in \mathcal{B}in(5; 0.181\,27)$ . A keresett valószínűség:

$$P(\eta > 1) = 1 - \left(0.818\,73^5 + 5 \cdot 0.181\,27 \cdot 0.818\,73^4\right) = 0.224\,87 \ .$$

**1.18. Feladat.** *Egy bizonyos forrásból származó adatállomány mérete exponenciális eloszlású véletlen mennyiség. Tudjuk, hogy az esetek felében az állomány mérete meghaladja a 120 kB-ot.*

**a)** *Mennyi annak valószínűsége, hogy egy állomány mérete meghladja a 200 kB-ot?*

**b)** *Ha egymás után kapunk ilyen állományokat, mennyi annak valószínűsége, hogy az első 200 kB-ot meghaladó méretű a tizedik után, de még a tizenötödik előtt érkezik?*

**Megoldás:**

**a)** Jelölje $\xi$ egy állomány méretét, akkor $\xi \in \mathcal{E}xp(\lambda)$, és tudjuk, hogy

$$P(\xi > 120) = e^{-\lambda \cdot 120} = 0.5.$$

amiből

$$\lambda = -\frac{\ln(0.5)}{120} = 5.7762 \times 10^{-3}$$

tehát a keresett valószínűség

$$P(\xi > 200) = e^{-5.7762 \times 10^{-3} \cdot 200} = 0.31498$$

**b)** Jelölje $\nu$ annak az állománynak a sorszámát, amely nagyobb mint 200 kB, akkor $\nu \in \mathcal{G}eom(0.31498)$, és a keresett valószínűség

$$P(11 \leq \nu \leq 14) = \sum_{k=11}^{14} (1 - 0.31498)^{k-1} \cdot 0.31498 = 1.7743 \times 10^{-2} \ .$$

**1.19. Feladat.** *Tudjuk, hogy a felnőtt emberek magassága* $\mathcal{N}(175; 10)$ *eloszlású véletlen mennyiség.*

**a)** *Milyen magas legyen egy ajtó, ha azt karjuk, hogy valaki 99%-os biztonsággal gond (lehajlás) nélkül tudja azt használni?*

**b)** *Ha egy lakásban négy felnőtt lakik, mennyi annak valószínűsége, hogy legfeljebb egy fő magasabb az előbb megadott ajtó-méretnél?*

**c)** *Milyen magas legyen az ajtaja egy 200 fős előadó teremnek, ha azt akarjuk, hogy 90%-os valószínűséggel senkinek ne okozzon gondot az ajtó?*

**Megoldás:**

**a)** Jelölje $\xi \in \mathcal{N}(175; 10)$ v.v. egy felnőtt magasságát, és $q$ a keresett értéket, akkor

$$P(\xi < q) = F(q) = \Phi\left(\frac{q - 175}{10}\right) = 0.99$$

amiből a $\Phi(2.3263) = 0.99$ táblázati értékkel kapjuk

$$\frac{q - 175}{10} = 2.3263$$

tehát $q = 198.26$ .

**b)** Jelölje $\nu$ azok számát, akiknek alacsony ez a méret, akkor $\nu \in \mathcal{B}in(4; 0.01)$, tehát a keresett valószínűség:

$$P(\nu \leq 1) = 0.99^4 + 4 \cdot 0.01 \cdot 0.99^3 = 0.999\,41$$

**c)** Jelölje $\nu$ azok számát, akiknek alacsony a keresendő $q$ méret, akkor $\nu \in \mathcal{B}in(200; p) \approxeq \mathcal{P}o(200p)$, tehát

$$P(\nu = 0) = e^{-200p} = 0.9 \Rightarrow p = \frac{\ln(0.9)}{-200} = 5.268 \times 10^{-4}$$

és teljesülni kell

$$P(\xi < q) = F(q) = \Phi\left(\frac{q - 175}{10}\right) = 1 - 5.268 \times 10^{-4}$$

amiből a $\Phi(3.275\,8) = 1 - 5.268 \times 10^{-4}$ táblázati értékkel kapjuk

$$\frac{q - 175}{10} = 3.275\,8$$

tehát $q = 207.76$ .

**1.20. Feladat.** *Egy öt fiúból, és öt lányból álló társaságban a fiúk magassága $\mathcal{N}(180; 8)$, a lányoké $\mathcal{N}(172; 10)$ eloszlású véletlen mennyiség.*

**a)** *Ha választunk egy fiút és egy lányt, mennyi annak valószínűsége, hogy a fiú legalább 5 cm-rel magasabb a lánynál?*

**b)** *Ha öt táncoló párt alkot a társaság, mennyi annak valószínűsége, hogy van köztük legalább egy pár, ahol a fiú nem magasabb legalább 5 cm-rel a lánynál?*

**Megoldás:**

**a)** Jelölje az egymástól független $\xi \in \mathcal{N}(180; 8)$ a fiú, $\eta \in \mathcal{N}(172; 10)$ a lány magasságát, akkor a keresett valószínűség

$$\begin{aligned} P(\xi &> \eta + 5) = P(\xi - \eta > 5) = 1 - F_{\xi-\eta}(5) = 1 - \Phi\left(\frac{5-8}{\sqrt{164}}\right) = \\ &= \Phi\left(\frac{3}{\sqrt{164}}\right) = 0.592\,61 \end{aligned}$$

mivel $\xi - \eta \in \mathcal{N}(8; \sqrt{164})$ .

**b)** $\nu$ jelölje azon párok számát, ahol a fiú nem magasabb 5 cm-rel a lánynál, akkor $\nu \in \mathcal{B}in(5; 0.40739)$, és a keresett valószínűség

$$P(\nu > 0) = 1 - 0.592\,61^5 = 0.926\,91 \ .$$

**1.21. Feladat.** *Legyenek $\xi, \eta \in \mathcal{N}(0;1)$ függetlenek, adjuk meg*

**a)** $\xi^2$ *eloszlását!*

**b)** $\xi^2 + \eta^2$ *eloszlását!*

**c)** $\frac{\xi}{\eta}$ *eloszlását!*

**Megoldás:**

**a)** Adjuk meg az eloszlásfüggvényt:

$$F_{\xi^2}(x) = \begin{cases} 0 & \text{ha } x \le 0 \\ 2\Phi(\sqrt{x}) - 1 & \text{ha } 0 < x \end{cases}$$

mivel $0 < x$ esetén

$$P(\xi^2 < x) = P(-\sqrt{x} < \xi < \sqrt{x}) = 2\Phi(\sqrt{x}) - 1 \ .$$

A sűrűségfüggvény:

$$f_{\xi^2}(x) = F'_{\xi^2}(x) = \frac{1}{\sqrt{x}}\varphi(x) = \frac{1}{\sqrt{2\pi x}} \cdot \exp\left(-\frac{x}{2}\right) \quad 0 < x \tag{1.9}$$

**b)** Vezessük be a

$$\begin{aligned} h(x,y) &= \left(x^2 + y^2; \arg(x,y)\right) \quad x, y \in \mathbb{R} \setminus \{(0;0)\} \\ h^{-1}(R,\phi) &= \left(\sqrt{R} \cdot \cos(\phi); \sqrt{R} \cdot \sin(\phi)\right) \quad R > 0, 0 \le \phi < 2\pi \end{aligned}$$

amivel $\left(\xi^2 + \eta^2; \tau\right) = h(\xi, \eta)$, és $\xi$ és $\eta$ függetlenek, ezért együttes sűrűségfüggvényük

$$f_{\xi,\eta}(u;v) = \frac{1}{2\pi} \cdot \exp\left(-\frac{u^2+v^2}{2}\right) \quad (u;v) \in \mathbb{R}^2 \tag{1.10}$$

és így

$$f_{(\xi^2+\eta^2;\tau)}(R,\phi) = \frac{1}{2} \cdot \frac{1}{2\pi} \cdot \exp\left(-\frac{R}{2}\right) \quad R > 0, 0 \le \phi < 2\pi$$

mivel a $h^{-1}$ függvény derivált mátrixának determinánsa

$$\det\left(\begin{bmatrix} \frac{\cos(\phi)}{2\sqrt{R}} & -\sqrt{R} \cdot \sin(\phi) \\ \frac{\sin(\phi)}{2\sqrt{R}} & \sqrt{R} \cdot \cos(\phi) \end{bmatrix}\right) = \frac{1}{2} \ ,$$

amiből kapjuk

$$f_{\xi^2+\eta^2}(R) = \int_0^{2\pi} \frac{1}{2} \cdot \frac{1}{2\pi} \cdot \exp\left(-\frac{R}{2}\right) d\phi = \frac{1}{2} \cdot \exp\left(-\frac{R}{2}\right) \quad R > 0 \ .$$

**c)** Vezessük be a $\zeta_1 = \frac{\xi}{\eta}$, és $\zeta_2 = \eta$ v.v.-kat, akkor $(\zeta_1, \zeta_2) = h(\xi, \eta)$, ahol

$$h(u,v) = \left(\frac{u}{v}, v\right) \quad u \in \mathbb{R},\ 0 \neq v \in \mathbb{R}$$

a $(\xi, \eta)$ értékkészletének 1-valószínűségű részén értelmezett, invertálható

$$h^{-1}(x,y) = (x \cdot y, y) \quad x \in \mathbb{R},\ 0 \neq y \in \mathbb{R}$$

és differenciálható, az inverz derivált mátrixa:

$$\begin{bmatrix} y & x \\ 0 & 1 \end{bmatrix} \quad x \in \mathbb{R},\ 0 \neq y \in \mathbb{R}$$

és

$$\det \begin{bmatrix} y & x \\ 0 & 1 \end{bmatrix} = y \quad x \in \mathbb{R},\ 0 \neq y \in \mathbb{R}\ .$$

Tehát a (1.10)-ből kapjuk $(\zeta_1, \zeta_2)$ sűrűségfüggvényét

$$f_{\zeta_1,\zeta_2}(x,y) = |y| \frac{1}{2\pi} \exp\left(-\frac{x^2y^2 + y^2}{2}\right) \quad x \in \mathbb{R},\ 0 \neq y \in \mathbb{R}\ ,$$

amiből a keresett sűrűségfüggvény:

$$\begin{aligned} f_{\zeta_1}(x) &= \int_{-\infty}^{\infty} |y| \frac{1}{2\pi} \exp\left(-\frac{x^2y^2+y^2}{2}\right) dy = 2\int_0^{\infty} y \frac{1}{2\pi} \exp\left(-\frac{x^2+1}{2} y^2\right) dy = \\ &= \left[-\frac{1}{\pi} \cdot \frac{1}{x^2+1} \cdot \exp\left(-\frac{x^2+1}{2}y^2\right)\right]_{y=0}^{y\to\infty} = \frac{1}{\pi} \cdot \frac{1}{x^2+1} \quad x \in \mathbb{R}\ . \end{aligned}$$

**Megjegyzés:** Vegyük észre, hogy $\xi^2 + \eta^2$ exponenciális eloszlású $\lambda = \frac{1}{2}$ paraméterrel, $\frac{\xi}{\eta}$ eloszlása pedig az u.n. Cauchy eloszlás.

**1.22. Feladat.** *Legyen $\varepsilon \in \mathcal{N}(0;1)$, adjuk meg olyan $\xi$ valószínűségi változó eloszlását, melyre teljesül*

$$\frac{\lambda(\xi - \mu)^2}{\mu^2 \xi} = \varepsilon^2 \tag{1.11}$$

*ahol $\lambda > 0, \mu > 0$.*

**Megoldás:** Az (1.11) egyenletet alakítva kapjuk a

$$\lambda\xi^2 - (2\lambda\mu + \mu^2\varepsilon^2)\xi + \lambda\mu^2 = 0$$

másodfokú egyenletet $\xi$-re, melynek diszkriminánsa

$$(2\lambda\mu + \mu^2\varepsilon^2)^2 - 4\lambda^2\mu^2 = 4\lambda\mu^3\varepsilon^2 + \mu^4\varepsilon^4 > 0\ ,$$

ezért van két $\xi_1 < \xi_2$ különböző gyök, melyek összege és szorzata

$$\xi_1 + \xi_2 = \frac{2\lambda\mu + \mu^2\varepsilon^2}{2\lambda} > 0 \qquad \xi_1 \cdot \xi_2 = \mu^2 > 0$$

tehát mindkettő pozitív, és $\mu = \sqrt{\xi_1 \cdot \xi_2}$ miatt

$$0 < \xi_1 < \mu < \xi_2 \, .$$

Tehát az (1.11) egyenletnek két megoldása van, és ezek eloszlása:

**a)** $\xi_1$ eloszlása.
Mivel $\xi_1 = h(\varepsilon^2)$, ahol

$$h : \mathbb{R}^+ \to ]0;\mu[ \qquad h^{-1}(x) = \frac{\lambda(x-\mu)^2}{\mu^2 x} \quad h^{-1\prime}(x) = \frac{\lambda}{\mu^2} \cdot \frac{x^2-\mu^2}{x^2} \qquad x \in ]0;\mu[,$$

használjuk $\varepsilon^2$ (1.9) szerinti sűrűségfüggvényét, amivel kapjuk $\xi_1$

$$\begin{aligned} f_1(x) &= \left| \frac{\lambda}{\mu^2} \cdot \frac{x^2-\mu^2}{x^2} \right| \cdot \frac{1}{\sqrt{2\pi \frac{\lambda(x-\mu)^2}{\mu^2 x}}} \cdot \exp\left( -\frac{\frac{\lambda(x-\mu)^2}{\mu^2 x}}{2} \right) = \\ &= \frac{x+\mu}{\mu} \cdot \sqrt{\frac{\lambda}{2\pi x^3}} \cdot \exp\left( -\frac{\lambda(x-\mu)^2}{2\mu^2 x} \right) \qquad x \in ]0;\mu[ \end{aligned}$$

sűrűségfüggvényét.

**b)** $\xi_2$ eloszlása.
Hasonlóan kapjuk $\xi_2$

$$f_2(x) = \frac{x+\mu}{\mu} \cdot \sqrt{\frac{\lambda}{2\pi x^3}} \cdot \exp\left( -\frac{\lambda(x-\mu)^2}{2\mu^2 x} \right) \qquad x \in ]\mu;+\infty[$$

sűrűségfüggvényét.

Az (1.11) egyenlet további megoldásai nyerhetők tetszőleges $A \in \mathcal{A}$ eseménnyel

$$\eta = \xi_1 \cdot \mathbf{1}_A + \xi_2 \cdot \mathbf{1}_{\overline{A}}$$

alakban. Legyen például $\tau \in \mathcal{U}(0;1)$ a $\xi_1$-től (és akkor a $\xi_2 = \frac{\mu^2}{\xi_1}$-től is) független, és $A = \left\{ \tau < \frac{\mu}{\mu+\xi_1} \right\}$. Ekkor $\eta$ eloszlásfüggvénye

$$\begin{aligned} F_\eta(z) &= P\left( \xi_1 < z, \tau < \frac{\mu}{\mu+\xi_1} \right) + P\left( \xi_2 < z, \tau \geq \frac{\mu}{\mu+\xi_1} \right) = \\ &= \int_0^z \sqrt{\frac{\lambda}{2\pi x^3}} \cdot \exp\left( -\frac{\lambda(x-\mu)^2}{2\mu^2 x} \right) dx \qquad 0 < z \end{aligned}$$

mivel

$$P\left(\xi_1 < z, \tau < \frac{\mu}{\mu+\xi_1}\right) = \begin{cases} \int_0^z \int_0^{\frac{\mu}{\mu+x}} f_1(x)dudx & \text{ha } 0 < z < \mu \\ \int_0^\mu \int_0^{\frac{\mu}{\mu+x}} f_1(x)dudx & \text{ha } \mu < z \end{cases} =$$

$$= \begin{cases} \int_0^z \frac{\mu}{\mu+x} f_1(x)dx & \text{ha } 0 < z < \mu \\ \int_0^\mu \frac{\mu}{\mu+x} f_1(x)dx & \text{ha } \mu < z \end{cases}$$

$$P\left(\xi_2 < z, \tau \geq \frac{\mu}{\mu+\xi_1}\right) = P\left(\xi_2 < z, \tau \geq \frac{\xi_2}{\mu+\xi_2}\right) =$$

$$= \begin{cases} 0 & \text{ha } 0 < z < \mu \\ \int_\mu^z \int_{\frac{x}{\mu+x}}^1 f_2(x)dudx & \text{ha } \mu < z \end{cases} =$$

$$= \begin{cases} 0 & \text{ha } 0 < z < \mu \\ \int_\mu^z \frac{\mu}{\mu+x} f_2(x)dx & \text{ha } \mu < z \end{cases}$$

Tehát $\eta$ megoldása az (1.11) egyenletnek, és sűrűségfüggvénye:

$$f_\eta(x) = \sqrt{\frac{\lambda}{2\pi x^3}} \cdot \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right) \qquad 0 < x \ .$$

## 1.4. Várható érték, szórás

**1.23. Feladat.** *Az 1.6. feladatban adjuk meg az eredmény várható értékét szórását!*

**Megoldás:** Használjuk $\xi$ (1.2) eloszlását

| $x$ | $P(\xi = x)$ | $x \cdot P(\xi = x)$ | $x^2 \cdot P(\xi = x)$ |
|---|---|---|---|
| $-15$ | $\frac{3}{6}$ | $-\frac{45}{6}$ | $\frac{675}{6}$ |
| 14 | $\frac{1}{6}$ | $\frac{14}{6}$ | $\frac{196}{6}$ |
| 20 | $\frac{1}{6}$ | $\frac{20}{6}$ | $\frac{400}{6}$ |
| 26 | $\frac{1}{6}$ | $\frac{26}{6}$ | $\frac{676}{6}$ |
| $\sum$ | 1 | $\frac{5}{2}$ | $\frac{649}{2}$ |

Tehát

$$E(\xi) = \frac{5}{2} \quad D(\xi) = \sqrt{\frac{649}{2} - \left(\frac{5}{2}\right)^2} = 17.84$$

**1.24. Feladat.** *Az 1.7. feladatban adjuk meg a terület várható értékét szórását!*

**Megoldás:** Használjuk $\xi$ (1.4) sűrűségfüggvényét:

$$f(x) = \frac{1}{2\sqrt{1-x}} \quad 0 < x < 1 \ .$$

Tehát

$$E(\xi) = \int_0^1 x \cdot \frac{1}{2\sqrt{1-x}} dx = \frac{2}{3} \quad E(\xi^2) = \int_0^1 x^2 \cdot \frac{1}{2\sqrt{1-x}} dx = \frac{8}{15}$$

így kapjuk

$$D(\xi) = \sqrt{\frac{8}{15} - \left(\frac{2}{3}\right)^2} = \frac{2}{15}\sqrt{5}\ .$$

**1.25. Feladat.** *A következő szerencsejátékot játszuk: $T$-összeg befizetése ellenében dobunk egy kockát, és ha az eredmény 1, 2 vagy 3, akkor nem nyerünk semmit, ha 4, 5 vagy 6 a dobás eredménye, akkor 18, 24 illetve 30 a nyereményünk*

**a)** *Milyen $T$-összegig érdemes játszani?*

**b)** *Mennyi a játékban elért eredmény szórása, amikor a várható eredmény 0?*

**Megoldás:**

**a)** Jelölje $\xi$ a játékban elért eredményt, akkor $\xi$ eloszlása

| $x$ | $P(\xi = x)$ |
|---|---|
| $-T$ | $\frac{1}{2}$ |
| $18 - T$ | $\frac{1}{6}$ |
| $24 - T$ | $\frac{1}{6}$ |
| $30 - T$ | $\frac{1}{6}$ |

amiből

$$E(\xi) = -T \cdot \frac{1}{2} + (18 - T) \cdot \frac{1}{6} + (24 - T) \cdot \frac{1}{6} + (30 - T) \cdot \frac{1}{6} = 12 - T\ .$$

Tehát érdemes játszani, ha

$$T < 12\ .$$

**b)** Mivel $E(\xi) = 0 \Leftrightarrow T = 12$, és ekkor

$$\begin{aligned} D^2(\xi) &= E(\xi^2) = (-12)^2 \cdot \frac{1}{2} + 6^2 \cdot \frac{1}{6} + 12^2 \cdot \frac{1}{6} + 18^2 \cdot \frac{1}{6} = 156 \\ D(\xi) &= \sqrt{156} = 2\sqrt{39} = 12.49 \end{aligned}$$

**1.26. Feladat.** *Adjuk meg a 1.8. feladat $\xi$ és $\eta$ véletlen mennyiségeivel kapcsolatban az*

**a)** *$E(\xi), E(\eta), E(\xi\eta)$ várható értékeket és a $D(\xi), D(\eta)$ szórásokat!*

**b)** *$\eta \approx a\xi + b$ közelítést!*

**Megoldás:**

**a)** Használjuk az (1.5)

| $x \diagdown y$ | 1 | 2 | 3 | 4 | 5 | 6 | $P(\xi = x)$ |
|---|---|---|---|---|---|---|---|
| 1 | $\frac{1}{36}$ | 0 | 0 | 0 | 0 | 0 | $\frac{1}{36}$ |
| 2 | $\frac{2}{36}$ | $\frac{1}{36}$ | 0 | 0 | 0 | 0 | $\frac{3}{36}$ |
| 3 | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ | 0 | 0 | 0 | $\frac{5}{36}$ |
| 4 | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ | 0 | 0 | $\frac{7}{36}$ |
| 5 | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ | 0 | $\frac{9}{36}$ |
| 6 | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ | $\frac{11}{36}$ |
| $P(\eta = y)$ | $\frac{11}{36}$ | $\frac{9}{36}$ | $\frac{7}{36}$ | $\frac{5}{36}$ | $\frac{3}{36}$ | $\frac{1}{36}$ | 1 |

eloszlást, amiből

$$\begin{aligned}
E(\xi) &= 1\cdot\frac{1}{36}+2\cdot\frac{3}{36}+3\cdot\frac{5}{36}+4\cdot\frac{7}{36}+5\cdot\frac{9}{36}+6\cdot\frac{11}{36}=\frac{161}{36}\\
E(\eta) &= 1\cdot\frac{11}{36}+2\cdot\frac{9}{36}+3\cdot\frac{7}{36}+4\cdot\frac{5}{36}+5\cdot\frac{3}{36}+6\cdot\frac{1}{36}=\frac{91}{36}
\end{aligned}$$

$$\begin{aligned}
E(\xi\eta) = 1\cdot 1\cdot\frac{1}{36}+2\cdot 1\cdot\frac{2}{36}+2\cdot 2\cdot\frac{1}{36}+(3\cdot 1+3\cdot 2)\cdot\frac{2}{36}+3\cdot 3\frac{1}{36}+\\
+(4\cdot 1+4\cdot 2+4\cdot 3)\cdot\frac{2}{36}+4\cdot 4\cdot\frac{1}{36}+(5\cdot 1+5\cdot 2+5\cdot 3+5\cdot 4)\cdot\frac{2}{36}+\\
+5\cdot 5\cdot\frac{1}{36}+(6\cdot 1+6\cdot 2+6\cdot 3+6\cdot 4+6\cdot 5)\cdot\frac{2}{36}+6\cdot 6\cdot\frac{1}{36}=\frac{49}{4}
\end{aligned}$$

továbbá

$$\begin{aligned}
E(\xi^2) &= 1^2\cdot\frac{1}{36}+2^2\cdot\frac{3}{36}+3^2\cdot\frac{5}{36}+4^2\cdot\frac{7}{36}+5^2\cdot\frac{9}{36}+6^2\cdot\frac{11}{36}=\frac{791}{36}\\
E(\eta^2) &= 1^2\cdot\frac{11}{36}+2^2\cdot\frac{9}{36}+3^2\cdot\frac{7}{36}+4^2\cdot\frac{5}{36}+5^2\cdot\frac{3}{36}+6^2\cdot\frac{1}{36}=\frac{301}{36}
\end{aligned}$$

tehát

$$\begin{aligned}
D(\xi) &= \sqrt{\frac{791}{36}-\left(\frac{161}{36}\right)^2}=\frac{1}{36}\sqrt{2555}=1.404\,1\\
D(\eta) &= \sqrt{\frac{301}{36}-\left(\frac{91}{36}\right)^2}=\frac{1}{36}\sqrt{2555}=1.404\,1
\end{aligned}$$

**b)** Mivel

$$\begin{aligned}
\operatorname{cov}(\xi,\eta) &= \frac{49}{4}-\frac{161}{36}\cdot\frac{91}{36}=\frac{1225}{1296}\\
r &= \frac{\frac{1225}{1296}}{\frac{1}{36}\sqrt{2555}\cdot\frac{1}{36}\sqrt{2555}}=\frac{35}{73}
\end{aligned}$$

kapjuk a regressziós együtthatók értékét

$$
\begin{aligned}
a &= \frac{\frac{1225}{1296}}{\left(\frac{1}{36}\sqrt{2555}\right)^2} = \frac{35}{73} = 0.479\,45 \\
b &= \frac{91}{36} - \frac{35}{73}\cdot\frac{161}{36} = \frac{28}{73} = 0.383\,56
\end{aligned}
$$

és a maradék szórásnégyzetet

$$
\sigma_R^2 = \left(\frac{1}{36}\sqrt{2555}\right)^2\left(1-\left(\frac{35}{73}\right)^2\right) = \frac{665}{438} = 1.518\,3\ .
$$

**1.27. Feladat.** *Adjuk meg az 1.11. feladatban szereplő $\xi$ és $\eta$ valószínűségi változók várható értékét és szórását, valamint a két véletlen mennyiség korrelációs együthatóját!*

**Megoldás:** Mivel $\xi$ sűrűségfüggvénye

$$
f(x) = \begin{cases} \frac{1}{4\sqrt{|x|}} & \text{ha } -1 < x \le 1 \\ 0 & \text{egyébként} \end{cases}\ .
$$

kapjuk

$$
\begin{aligned}
E(\xi) &= \int_{-1}^{0} x\cdot\frac{1}{4\sqrt{-x}}dx + \int_0^1 x\cdot\frac{1}{4\sqrt{x}}dx = 0 \\
E(\eta) &= E(\xi^2) = \int_{-1}^{0} x^2\cdot\frac{1}{4\sqrt{-x}}dx + \int_0^1 x^2\cdot\frac{1}{4\sqrt{x}}dx = \frac{1}{5} \\
E(\eta^2) &= E(\xi^4) = \int_{-1}^{0} x^4\cdot\frac{1}{4\sqrt{-x}}dx + \int_0^1 x^4\cdot\frac{1}{4\sqrt{x}}dx = \frac{1}{9} \\
E(\xi\eta) &= E(\xi^3) = \int_{-1}^{0} x^3\cdot\frac{1}{4\sqrt{-x}}dx + \int_0^1 x^3\cdot\frac{1}{4\sqrt{x}}dx = 0
\end{aligned}
$$

kapjuk

$$
D(\xi) = \sqrt{\frac{1}{5}} \quad D(\eta) = \sqrt{\frac{1}{9}-\left(\frac{1}{5}\right)^2} = \frac{4}{15} \quad \operatorname{cov}(\xi,\eta) = 0 \quad r = 0\ .
$$

**Megjegyzés:** Vegyük észre, hogy $\xi$ és $\eta$ korrelálatlan, de nem független, mert például

$$
\begin{aligned}
P\left(\xi < -\frac{1}{4}\right) &= \int_{-1}^{-\frac{1}{4}} \frac{1}{4\sqrt{-x}}dx = \frac{1}{4} \\
P\left(\eta > \frac{1}{81}\right) &= P\left(\xi < -\frac{1}{9}\right) + P\left(\xi > \frac{1}{9}\right) = 2\int_{-1}^{-\frac{1}{9}} \frac{1}{4\sqrt{-x}}dx = \frac{2}{3}
\end{aligned}
$$

és

$$
P\left(\xi < -\frac{1}{4}, \eta > \frac{1}{81}\right) = P\left(\xi < -\frac{1}{4}\right) = \frac{1}{4} \ne \frac{1}{4}\cdot\frac{2}{3}\ .
$$

**1.28. Feladat.** *Egy ujságárusnál egy adott lapot átlagosan 100 vásárló keres egy napon. Ha naponta 120 lapot rendel, mennyi a várható haszna, ha az 50Ft-ért beszerzett lapot 60Ft-ért adja a vásárlónak, és a megmaradt példányokat 30Ft-ért veszi vissza a terjesztő?*

**Megoldás:** Felthetjük, hogy a napi kereslet $\nu \in \mathcal{P}o(100)$, és ha $N = 120$ jelöli a rendelt mennyiséget, akkor a haszon véletlen mennyisége:

$$\begin{aligned}
\xi &= 10 \cdot \nu \cdot \mathbf{1}_{\{\nu \leq N\}} - 20 \cdot (N - \nu) \cdot \mathbf{1}_{\{\nu \leq N\}} + (10 \cdot N) \cdot \mathbf{1}_{\{\nu > N\}} = \\
&= 30 \cdot \nu \cdot \mathbf{1}_{\{\nu \leq N\}} - 30 \cdot N \cdot \mathbf{1}_{\{\nu \leq N\}} + 10 \cdot N = \\
&= 30 \, (\nu - N) \cdot \mathbf{1}_{\{\nu \leq N\}} + 10N = 30 \sum_{k=0}^{N} (k - N) \mathbf{1}_{\{\nu = k\}} + 10N \ .
\end{aligned}$$

Adjuk meg ennek várható értékét:

$$E(\xi) = 30 \sum_{k=0}^{120} (k - 120) \, \frac{100^k}{k!} e^{-100} + 10 \cdot 120 = 596.92 \ .$$

**1.29. Feladat.** *Legyenek $\xi_1 \in \mathcal{N}(10; 2)$, $\xi_2 \in \mathcal{E}xp(2)$ függetlenek, és jelölje $\xi = \xi_1$ és $\eta = \xi_1 \cdot \xi_2$. Adjuk meg az $\eta \approx a\xi + b$ regressziós közelítést!*

**Megoldás:** Számítsuk ki a szükséges várható értékeket:

$$\begin{aligned}
E(\xi) &= 10 \\
E(\eta) &= E\left(\xi_1 \cdot \xi_2\right) = 10 \cdot \frac{1}{2} = 5 \\
D^2(\xi) &= 4 \\
E(\xi\eta) &= E\left(\xi_1^2 \cdot \xi_2\right) = \left(2^2 + 10^2\right) \cdot \frac{1}{2} = 52 \\
\operatorname{cov}(\xi, \eta) &= 52 - 10 \cdot 5 = 2 \\
E(\eta^2) &= E\left(\xi_1^2 \cdot \xi_2^2\right) = \left(2^2 + 10^2\right) \cdot \left( \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 \right) = 52 \\
D^2(\eta) &= 52 - 5^2 = 27
\end{aligned}$$

amiből kapjuk

$$\begin{aligned}
a &= \frac{2}{4} = \frac{1}{2} \quad b = 5 - \frac{1}{2} \cdot 10 = 0 \\
r &= \frac{2}{\sqrt{4} \cdot \sqrt{27}} = \frac{\sqrt{3}}{9} \quad \sigma_R^2 = 27 \left( 1 - \left( \frac{\sqrt{3}}{9} \right)^2 \right) = 26 \ .
\end{aligned}$$

**1.30. Feladat (*).** *Legyen $\xi \in \Omega \to \mathbb{R}$ v.v., mutassuk meg, hogy $E(\xi)$ akkor és csak akkor létezik, ha a*

$$\sum_{n=1}^{\infty} P\left(|\xi| > n\right) \tag{1.12}$$

*sor konvergens!*

**Bizonyítás.** Elég $\xi \geq 0$ esetet vizsgálni, amikor teljesül

$$\sum_{n=1}^{N} \mathbf{1}_{\{\xi > n\}} \leq \xi \leq 1 + \sum_{n=1}^{N} \mathbf{1}_{\{\xi > n\}} \ ,$$

és ebből már következik az állítás. ■

**1.31. Feladat (*).** *Bizonyítsuk az un. "parciális integrálás" szabályát: Legyen $\xi \geq 0$ v.v. eloszlásfüggvénye $F$, akkor*

$$\int_0^x t dP_\xi = \int_0^x (1 - F(t))\, dt - x \cdot (1 - F(x+0)) \qquad x > 0. \tag{1.13}$$

**Bizonyítás.** Mivel

$$\sum_{k=0}^{n-1} x\frac{k}{n} \mathbf{1}_{\left\{x\frac{k}{n} \leq \xi < x\frac{k+1}{n}\right\}} + x \cdot \mathbf{1}_{\{\xi = x\}} \nearrow \xi \cdot \mathbf{1}_{\{\xi \leq x\}} \ ,$$

kapjuk

$$\begin{aligned}
\int_0^x t dP_\xi = E\left(\xi \cdot \mathbf{1}_{\{\xi \leq x\}}\right) = x \cdot P(\xi = x) + \lim_{n\to\infty} \sum_{k=0}^{n-1} x\frac{k}{n}\left[F\left(x\frac{k+1}{n}\right) - F\left(x\frac{k}{n}\right)\right] = \\
x \cdot (F(x+0) - F(x)) + \lim_{n\to\infty} \left\{ x\frac{2}{n} F\left(x\frac{2}{n}\right) - x\frac{1}{n} F\left(x\frac{1}{n}\right) - x\frac{1}{n} F\left(x\frac{2}{n}\right) + \right. \\
x\frac{3}{n} F\left(x\frac{3}{n}\right) - x\frac{2}{n} F\left(x\frac{1}{n}\right) - x\frac{1}{n} F\left(x\frac{3}{n}\right) + \\
\vdots \\
\left. x\frac{n}{n} F\left(x\frac{n}{n}\right) - x\frac{n-1}{n} F\left(x\frac{n-1}{n}\right) - x\frac{1}{n} F\left(x\frac{n}{n}\right) \right\} = \\
= xF(x+0) - \lim_{n\to\infty} \sum_{k=0}^{n-1} x\frac{1}{n} F\left(x\frac{k}{n}\right) = xF(x+0) - \int_0^x F(t)dt = \\
= \int_0^x (1 - F(t))\, dt - x \cdot (1 - F(x+0)) \ .
\end{aligned}$$

■

**Következmények:**

1. Vegyük észre, hogy (1.13) jobb oldalán szereplő integrál $F$ monotonitása miatt mindíg létezik, és véges.

2. Mivel az eloszlásfüggvény $x \in \mathbb{R}$ folytonossági helyein (tehát egy legfeljebb megszámlálható halmaz kivételével)

$$\xi \cdot \mathbf{1}_{\{\xi \leq x\}} \leq x \cdot \mathbf{1}_{\{\xi > x\}} + \xi \cdot \mathbf{1}_{\{\xi \leq x\}} \leq \xi$$

kapjuk

$$\int_0^x t dP_\xi \le x \cdot (1 - F(x)) + \int_0^x t dP_\xi = \int_0^x (1 - F(t))\, dt \le E(\xi) = \lim_{x\to\infty} \int_0^x t dP_\xi \ .$$

Tehát $E(\xi)$ pontosan akkor létezik, ha az $\int_0^\infty (1 - F(t))\, dt$ impropríus integrál véges értékű, és ekkor

$$E(\xi) = \int_0^\infty (1 - F(t))\, dt \ .$$

3. Ha $\mathrm{im}(\xi) \subset \mathbb{N}$, akkor a (1.12) sort kiegészítve, a $\sum_{n=0}^\infty P(\xi > n)$ összege pontosan akkor véges, ha létezik a várható érték, és ekkor

$$E(\xi) = \int_0^\infty (1 - F(t))\, dt = \int_0^1 (1 - F(t))\, dt + \int_1^2 (1 - F(t))\, dt + \cdots = \\ = \sum_{n=0}^\infty P(\xi > n) \ .$$

**1.32. Feladat (*).** *Ha $\xi$ v.v.-nak van várható értéke és szórása, akkor*

$$P\big(\, |\xi - E(\xi)| > 2D(\xi)\big) + P\big(\, |\xi - E(\xi)| > 3D(\xi)\big) < \frac{1}{4}$$

*és nincs $\frac{1}{4}$-nél kisebb felső korlát.*

**Bizonyítás.** Elég az $E(\xi) = 0$, $D(\xi) = 1$ eset vizsgálata, és jelölje $\eta = |\xi|$. Mivel

$$\eta > 2 \cdot 1_{\{\eta > 2\}} + 1_{\{\eta > 3\}} \Rightarrow \eta^2 > 4 \cdot 1_{\{\eta > 2\}} + 5 \cdot 1_{\{\eta > 3\}} \ ,$$

ezért

$$1 > 4 \cdot (P(|\xi - E(\xi)| > 2D(\xi)) + P(|\xi - E(\xi)| > 3D(\xi)))\,.$$

Legyen továbbá $\xi_n$ v.v., melyre

$$P\left(\xi_n = \pm\left(2 + \frac{1}{n}\right)\right) = \frac{1}{2} \cdot \frac{1 - \frac{1}{n^2}}{4 + \frac{4}{n}} \quad P\left(\xi_n = \pm\frac{1}{n}\right) = \frac{1}{2} \cdot \frac{3 + \frac{4}{n} + \frac{1}{n^2}}{4 + \frac{4}{n}} \quad n = 2, 3, \ldots$$

amiből következik $E(\xi_n) = 0$, $D(\xi_n) = 1$, és

$$\begin{aligned} P(|\xi_n| &> 2) = \frac{1 - \frac{1}{n^2}}{4 + \frac{4}{n}} \to \frac{1}{4} \\ P(|\xi_n| &> 3) = 0 \end{aligned}$$

tehát $\frac{1}{4}$ a legkisebb felső korlát. ■

**1.33. Feladat (*).** *Legyen $0 \le \xi$ v.v., és $m_n = E(\xi^n)$ $\quad n = 0, 1, 2, \ldots, N + 2$, akkor teljesül*

$$m_{n+1}^2 \le m_n \cdot m_{n+2} \quad n = 0, 1, 2, \ldots, N.$$

**Bizonyítás.** Tegyük fel először, hogy van olyan $0 < \varepsilon$, melyre $\varepsilon \leq \xi$. Megmutatjuk, hogy a

$$\phi(x) = \ln\left(E(\xi^x)\right) \quad x \in [0; N]$$

függvény konvex. $\phi$ nyilván értelmezett, és mivel elég kis $h$ esetén

$$\left|\frac{\xi^{x+h} - \xi^x}{h}\right| = \xi^x \cdot \left|\frac{\xi^h - \xi^0}{h}\right| \leq \xi^x \cdot (\xi + 1 + |\ln \varepsilon|) \leq \xi^{N+1} + \xi + |\ln \varepsilon|$$

$$|\ln \xi| \left|\frac{\xi^{x+h} - \xi^x}{h}\right| \leq \xi^{N+2} + \xi^2 + \xi \cdot |\ln \varepsilon| + \xi^{N+1} + \xi + |\ln \varepsilon| + \xi^{N+1} \cdot |\ln \varepsilon| + + \xi \cdot |\ln \varepsilon| + |\ln \varepsilon|^2,$$

ezért a várható érték és a deriválás sorrendje felcserélhető, így

$$\phi'(x) = \frac{E(\ln \xi \cdot \xi^x)}{E(\xi^x)} \quad x \in [0; N]$$

amiből

$$\phi''(x) = \frac{E(\ln^2 \xi \cdot \xi^x) \cdot E(\xi^x) - E^2(\ln \xi \cdot \xi^x)}{E^2(\xi^x)} \quad x \in [0; N]$$

ahol a számláló nem lehet negatív, ugyanis

$$E^2\left(\left(\ln \xi \cdot \sqrt{\xi^x}\right) \cdot \sqrt{\xi^x}\right) \leq E(\ln^2 \xi \cdot \xi^x) \cdot E(\xi^x).$$

Tehát $\phi$ konvex, és így

$$\ln m_{n+1} \leq \frac{1}{2}\left(\ln m_n + \ln m_{n+2}\right)$$

vagyis amit meg kellett mutatni.

Legyen most $0 \leq \xi$ és $\xi_\varepsilon = \varepsilon + \xi$, és jelölje $m_n^\varepsilon = E(\xi_\varepsilon^n)$, akkor a fentiekből, és

$$\lim_{\varepsilon \to 0+} m_n^\varepsilon = m_n$$

miatt következik az állítás. ■

## 1.5. Generátorfüggvény, karakterisztikus függvény

**1.34. Feladat.** *Legyen a $\xi$ és $\eta$ független v.v.-k generátorfüggvénye:*

$$\begin{aligned} G_\xi(z) &= \frac{1}{2} + \frac{1}{4}z + \frac{1}{4}z^2 \\ G_\eta(z) &= \frac{1}{6}z + \frac{1}{3}z^2 + \frac{1}{2}z^3 \end{aligned}$$

*adjuk meg $\xi + \eta$ generátorfüggvényét, várható értékét és szórását!*

**Megoldás:** Mivel

$$\begin{aligned} G_{\xi+\eta}(z) &= \left(\frac{1}{2}+\frac{1}{4}z+\frac{1}{4}z^2\right)\cdot\left(\frac{1}{6}z+\frac{1}{3}z^2+\frac{1}{2}z^3\right) = \\ &= \frac{1}{8}z^5+\frac{5}{24}z^4+\frac{3}{8}z^3+\frac{5}{24}z^2+\frac{1}{12}z \quad z\in\mathbb{C} \end{aligned}$$

kapjuk

$$\begin{aligned} G'_{\xi+\eta}(z) &= \frac{5}{8}z^4+\frac{5}{6}z^3+\frac{9}{8}z^2+\frac{5}{12}z+\frac{1}{12} \\ G'_{\xi+\eta}(1) &= \frac{5}{8}+\frac{5}{6}+\frac{9}{8}+\frac{5}{12}+\frac{1}{12}=\frac{37}{12}=E(\xi+\eta) \end{aligned}$$

$$\begin{aligned} G''_{\xi+\eta}(z) &= \frac{5}{2}z^3+\frac{5}{2}z^2+\frac{9}{4}z+\frac{5}{12} \\ G''_{\xi+\eta}(1) &= \frac{5}{2}+\frac{5}{2}+\frac{9}{4}+\frac{5}{12}=\frac{23}{3}=E\left((\xi+\eta)^2\right)-E(\xi+\eta) \end{aligned}$$

tehát

$$\begin{aligned} E(\xi+\eta) &= \frac{37}{12} \\ D(\xi+\eta) &= \sqrt{\frac{23}{3}+\frac{37}{12}-\left(\frac{37}{12}\right)^2}=\frac{1}{12}\sqrt{179} \end{aligned}$$

**1.35. Feladat.** *Legyen a $\xi$ v.v. sűrűségfüggvénye*

$$f_\xi(x)=\frac{1}{2}e^{-|x|} \quad x\in\mathbb{R}$$

*$a,b\in\mathbb{R}$ adjuk meg $\eta=a\xi+b$ karakterisztikus függvényét!*

**Megoldás:** Adjuk meg először $\xi$

$$\begin{aligned} \phi_\xi(t) &= E(e^{it\xi})=\int_{-\infty}^{\infty}e^{itx}\cdot\frac{1}{2}e^{-|x|}dx= \\ &= \int_{-\infty}^{0}e^{itx}\cdot\frac{1}{2}e^{x}dx+\int_{0}^{\infty}e^{itx}\cdot\frac{1}{2}e^{-x}dx= \\ &= \frac{1}{2\left(1+t^2\right)}-\frac{1}{2}i\frac{t}{1+t^2}+\frac{1}{2\left(1+t^2\right)}+\frac{1}{2}i\frac{t}{1+t^2}=\frac{1}{1+t^2} \quad t\in\mathbb{R} \end{aligned}$$

karakterisztikus függvényét, amiből

$$\phi_\eta(t)=e^{itb}\varphi_\xi(at)=\frac{e^{itb}}{1+a^2t^2} \quad t\in\mathbb{R}\ .$$

**Megjegyzés:** Vegyük észre, hogy az inverziós formula következménye szerint, a folytonos eloszlás sűrűségfüggvényét megkaphatjuk

$$\frac{1}{2}e^{-|x|} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx} \cdot \frac{1}{1+t^2}dt$$

amiből a $t = -u$ helyettesítéssel, és rendezve

$$e^{-|x|} = \int_{-\infty}^{\infty} e^{iux} \cdot \frac{1}{\pi}\frac{1}{1+u^2}du \quad x \in \mathbb{R} \ ,$$

kapjuk a Cauchy eloszlás karakterisztikus függvényét.

**1.36. Feladat.** *Legyen a $\xi$ v.v. sűrűségfüggvénye*

$$f(x) = \begin{cases} \frac{3}{8}x^2 & ha\ 0 < x < 2 \\ 0 & egyébként \end{cases} ,$$

*adjuk meg $\xi$ karakterisztikus függvényét!*

**Megoldás:** Adjuk meg a karakterisztikus függvényt:

$$\begin{aligned}
\phi_\xi(t) &= E(e^{it\xi}) = \int_0^2 e^{itx} \cdot \frac{3}{8}x^2 dx = \left[\frac{e^{itx}}{it} \cdot \frac{3}{8}x^2\right]_0^2 - \int_0^2 \frac{e^{itx}}{it} \cdot \frac{3}{4}x dx = \\
&= -\frac{3i}{2t}e^{2it} - \left[\frac{e^{itx}}{i^2t^2} \cdot \frac{3}{4}x\right]_0^2 + \int_0^2 \frac{e^{itx}}{i^2t^2} \cdot \frac{3}{4}dx = \\
&= -\frac{3i}{2t}e^{2it} + \frac{3}{2t^2}e^{2it} - \frac{3}{4}\left[\frac{e^{itx}}{it^3}\right]_0^2 = \left(-\frac{3i}{2t} + \frac{3}{2t^2} + \frac{3i}{4t^3}\right)e^{2it} - \frac{3i}{4t^3} \quad t \in \mathbb{R} \setminus \{0\} \ .
\end{aligned}$$

**Megjegyzés:** Vegyük észre, hogy a fenti eredmény csak $t \neq 0$ esetben értelmezett, de minden karakterisztikus függvény folytonos a $t = 1$ helyen, és értéke 1, amit esetünkben ellenőrizhetünk:

$$\lim_{t\to 0}\left[\left(-\frac{3i}{2t} + \frac{3}{2t^2} + \frac{3i}{4t^3}\right)e^{2it} - \frac{3i}{4t^3}\right] = 1$$

**1.37. Feladat.** *Legyen a $\xi$ v.v. exponenciális eloszlású, és $E(\xi) = 10$, és a $\nu$ v.v. eloszlása:*

| $k$ | $P(\nu = k)$ |
|---|---|
| 1 | $\frac{1}{2}$ |
| 2 | $\frac{1}{4}$ |
| 3 | $\frac{1}{4}$ |

*adjuk meg $\nu$ generátorfüggvényét! Adjuk meg továbbá az*

$$\eta = \sum_{k=1}^{\nu} \xi_k$$

*v.v. karakterisztikus függvényét és várható értékét, ha $\xi_k$ $k = 1, 2, \ldots$ függetlenek, és közös eloszásuk azonos a $\xi$ valószínűségi változóéval!*

**Megoldás:** Adjuk meg $\nu$

$$G(z) = \frac{1}{2}z + \frac{1}{4}z^2 + \frac{1}{4}z^3 \quad z \in \mathbb{C}$$

generátorfüggvényét, és $\xi_k$

$$\phi(t) = \frac{1}{1 - it10} \quad t \in \mathbb{R}$$

karaktrisztikus függvényét, amivel kapjuk

$$\phi_\eta(t) = \frac{1}{2} \cdot \frac{1}{1 - it10} + \frac{1}{4} \cdot \left(\frac{1}{1 - it10}\right)^2 + \frac{1}{4} \cdot \left(\frac{1}{1 - it10}\right)^3 \quad t \in \mathbb{R} ,$$

továbbá

$$\begin{aligned} \phi'_\eta(t) &= \frac{1}{2} \cdot i10\,(1 - it10)^{-2} + \frac{1}{4} \cdot 2i10\,(1 - it10)^{-3} + \frac{1}{4} \cdot 3i10\,(1 - it10)^{-4} \\ \phi'_\eta(0) &= i\frac{35}{2} \quad \Rightarrow \quad E(\eta) = \frac{35}{2} \end{aligned}$$

## 1.6. Központi határeloszlás tétel

**1.38. Feladat.** *Milyen határok között lehet az a véletlen mennyiség, amit 12 darab,* 0 *és* 1 *közötti véletlen szám összegéből* 6*-ot levonva kapunk, 90%-os valószínűséggel?*

**Megoldás:** Jelölje $\xi_k \in \mathcal{U}(0;1) \quad k = 1, 2, \cdots, 12$ a független véletlen számokat, akkor az

$$\eta = \sum_{k=1}^{12} \xi_k - 6 \quad E(\eta) = 0 \quad D(\eta) = \sqrt{12 \cdot \frac{1}{12}} = 1$$

v.v. közelítően $\eta \in \mathcal{N}\,(0;1)$, tehát

$$P(-x \le \eta \le x) = 0.9 \Leftrightarrow 2\Phi\,(x) - 1 = 0.9 \Leftrightarrow \Phi\,(x) = 0.95$$

aminek $x = 1.645$ a megoldása, mivel $\Phi\,(1.645) = 0.95$ .

**1.39. Feladat.** *Egy üzletben egy nap átlagosan 120 vevő vásárol. Mennyi annak valószínűsége, hogy 90-nél kevesebben vásárolnak egy napon?*

**Megoldás:** Jelölje $\nu$ a vásárlók számát, akkor $\nu \in \mathcal{P}\,(120)$, ami már közelíthető az $\mathcal{N}\left(120; \sqrt{120}\right)$ eloszlással, tehát

$$P(\nu < 70) = \Phi\left(\frac{90 - 120}{\sqrt{120}}\right) = 1 - \Phi(2.74) = 1 - 0.\,996\,93 = 0.00\,307$$

**1.40. Feladat.** *Hányszor kell egy szabályos érmét feldobnunk ahhoz, hogy a fejek számának relatív gyakorisága legalább 0.9 valószínűséggel 0.01-nél kevesebbel térjen el a valószínűségtől*

**Megoldás:** Jelölje $\nu$ v.v. a fejek számát $n$ dobásból, akkor szabályos érme esetén $\nu \in \mathcal{B}in(n, \frac{1}{2})$, és keressük azt az $n$ számot, melyre teljesül

$$P\left(\left|\frac{\nu}{n} - \frac{1}{2}\right| < 0.01\right) \geq 0.9 \ .$$

Mivel közelítően $\nu \in \mathcal{N}\left(n\frac{1}{2}, \sqrt{n\frac{1}{4}}\right)$, ezért $\frac{\nu}{n} - \frac{1}{2} \in \mathcal{N}\left(0, \frac{1}{2\sqrt{n}}\right)$, amivel kapjuk

$$2\Phi\left(2\sqrt{n} \cdot 0.01\right) - 1 \geq 0.9 \Leftrightarrow \Phi\left(2\sqrt{n} \cdot 0.01\right) \geq 0.95 \Leftrightarrow 2\sqrt{n} \cdot 0.01 \geq 1.645 \Leftrightarrow \sqrt{n} \geq \frac{164.5}{2}$$

mivel $\Phi(1.645) = 0.95$ , tehát $n \geq 6765$ esetén teljesül a feltétel.
**Megjegyzés:** Ha az érme nem szabályos, akkor

$$P\left(\left|\frac{\nu}{n} - p\right| < 0.01\right) \geq 0.9$$

kell, hogy teljesüljön az ismeretlen $p \in ]0;1[$ értékkel, és ekkor $\frac{\nu}{n} - p \in \mathcal{N}\left(0, \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right)$, amiből

$$2\Phi\left(\frac{0.01\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \geq 0.9 \Leftrightarrow \Phi\left(\frac{0.01\sqrt{n}}{\sqrt{p(1-p)}}\right) \geq 0.95 \Leftrightarrow \sqrt{n} \geq 164.5\sqrt{p(1-p)}$$

ahol a kívánt egyenlőtlenség következik az

$$\sqrt{n} \geq 164.5 \cdot \frac{1}{2}$$

egyenlőtlenségből, mivel $\sqrt{p(1-p)} \leq \frac{1}{2}$ . Tehát az előző eredmény minden $p$ esetén megoldása a feladatnak.

**1.41. Feladat.** *Egy 1000 darabos tételben 200 hibás darab van. Ha 100 darabot kiválasztunk, legalább hány hibás darab lesz a kivettek között 90%-os valószínűséggel?*

**Megoldás:** Jelölje $\nu$ a hibásak számát a mintában, akkor $\nu \in \mathcal{H}yp(1000; 200; 100)$, amit közelíthetünk $\mathcal{N}(20; 3.796\,6)$ eloszlással, mivel

$$E(\nu) = 100 \cdot \frac{200}{1000} = 20 \quad D(\nu) = \sqrt{20 \cdot 0.8 \cdot \left[1 - \frac{99}{999}\right]} = 3.796\,6 \ .$$

Keressük tehát $k \in \mathbb{R}$, melyre

$$P(\nu \geq k) = 1 - \Phi\left(\frac{k-20}{3.796\,6}\right) = 0.9 \Rightarrow -\frac{k-20}{3.796\,6} = 1.282 \Rightarrow k = 15.133 \approx 15 \ .$$

**1.42. Feladat.** *Az 1.28 feladatban keressük meg azt az $N$ rendelés számot, mellyel a várható haszon maximális!*

**Megoldás:** Jelölje $\xi$ a hasznot, $\nu \in \mathcal{P}o(\lambda) \cong \mathcal{N}(\lambda, \sqrt{\lambda})$ a vásárlók számát, ahol $\lambda = 100$, $b = 10$ a haszon egy eladott újságon és $c = 20$ a veszteség minden megmaradt példányon, akkor

$$\begin{aligned} m_N &= E(\xi) = E\left((b+c)\sum_{k=0}^{N}(k-N)\mathbf{1}_{\{\nu=k\}} + bN\right) = \\ &= (b+c)\sum_{k=0}^{N} k\cdot\mathbf{P}(\nu=k) - (b+c)N\sum_{k=0}^{N}\mathbf{P}(\nu=k) + bN\ . \end{aligned}$$

Vizsgáljuk az $m_N \quad N = 1, 2, \ldots$ sorozat monotonitását:

$$m_{N+1} - m_N = -(b+c)\sum_{k=0}^{N}\mathbf{P}(\nu=k) + b \geq 0\ ,$$

ami pontosan akkor teljesül, ha

$$\frac{b}{b+c} \geq \sum_{k=0}^{N}\mathbf{P}(\nu=k) = P(\nu \leq N) \approx \Phi\left(\frac{N-\lambda}{\sqrt{\lambda}}\right)\ .$$

$m_N$ maximális, ha

$$\frac{b}{b+c} = \Phi\left(\frac{N-\lambda}{\sqrt{\lambda}}\right)\ ,$$

adatainkat behelyettesítve:

$$\frac{1}{3} = \Phi\left(\frac{N-100}{\sqrt{100}}\right) \Rightarrow \frac{N-100}{\sqrt{100}} = -0.430\,73 \Rightarrow N = 95.693.$$

Tehát $N \approx 96$ darabot kell rendelni a maximális várható haszonhoz, és ekkor

$$E(\xi) = 30\sum_{k=0}^{96}(k-96)\,\frac{100^k}{k!}e^{-100} + 10\cdot 96 = 891.71\ .$$

## 1.7. Vektor valószínűségi változók

**1.43. Feladat.** *A $\xi = (\xi_1, \xi_2)$ v.v.v. kovariancia mátrixa és várható értéke:*

$$\mathrm{cov}(\xi,\xi) = \begin{bmatrix} 7 & \sqrt{3} \\ \sqrt{3} & 5 \end{bmatrix} \qquad E(\xi) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

***a)*** *Adjuk meg az első (a nagyobbik sajátértékhez tartozó) főkomponenssel és főfaktorral való közelítést, és a közelítés hibáját, és a közelítés egyenesének egyenletét!*

***b)*** *Legyenek*

$$\begin{aligned} \eta_1 &= \xi_1 - \xi_2 + 1 \\ \eta_2 &= \xi_1 + 2\xi_2 - 5 \end{aligned}$$

*ajuk meg az $\eta = (\eta_1, \eta_2)$ v.v.v. kovariancia mátrixát és várható érték vektorát!*

***c)*** *Ha* $(\xi_1, \xi_2)$ *egy megfigyelt értéke* $(2; 1)$, *adjuk meg a főkomponensek és főfaktorok megfelelő értékeit!*

**Megoldás:**

**a)** $\text{cov}(\xi, \xi)$ sajátértékei és normált sajátvektorai:

$$\lambda_1 = 8 \quad v_1 = \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix} \quad \lambda_2 = 4 \quad v_2 = \begin{bmatrix} -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix}$$

tehát

$$\xi \sim \tau_1 \cdot \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \tau_1^* \cdot \begin{bmatrix} \sqrt{6} \\ \sqrt{2} \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \sigma_R^2 = \lambda_2 = 4$$

és az első főkomponens illetve főfaktor egyenesének egyenlete

$$-x + \sqrt{3}y = -1 + 2\sqrt{3}$$

**b)**

$$\text{cov}(\eta, \eta) = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 7 & \sqrt{3} \\ \sqrt{3} & 5 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}^T = \begin{bmatrix} 12 - 2\sqrt{3} & -3 + \sqrt{3} \\ -3 + \sqrt{3} & 27 + 4\sqrt{3} \end{bmatrix}$$

$$E(\eta) = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ -5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

***c)***

| $\xi_1$ | $\xi_2$ | $\xi_1^*$ | $\xi_2^*$ | $\tau_1$ | $\tau_2$ | $\tau_1^*$ | $\tau_2^*$ |
|---|---|---|---|---|---|---|---|
| 2 | 1 | 1 | $-1$ | $\frac{1}{2}\sqrt{3} - \frac{1}{2}$ | $-\frac{1}{2} - \frac{1}{2}\sqrt{3}$ | $\frac{\frac{1}{2}\sqrt{3} - \frac{1}{2}}{\sqrt{8}} = 0.129$ | $\frac{-\frac{1}{2} - \frac{1}{2}\sqrt{3}}{2} = -0.683$ |

mivel

$$\begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{2}\sqrt{3} - \frac{1}{2}$$

$$\begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -\frac{1}{2} - \frac{1}{2}\sqrt{3}$$

**1.44. Feladat.** *Legyenek egy téglalap oldalai a* $\xi$ *és* $\eta$ *független valószínűségi változók, és*

$$E(\xi) = 100 \quad D(\xi) = 1 \quad E(\eta) = 150 \quad D(\eta) = 2$$

*jelölje továbbá a terület logaritmusát* $\tau$, *a kerület logaritmusát pedig* $\kappa$. *Adjuk meg a* $(\tau, \kappa)$ *v.v.v. várható értékének és kovariancia mátrixának közelítését!*

**Megoldás:** Mivel

$$\left[\begin{array}{c}\tau\\ k\end{array}\right]=\left[\begin{array}{c}\ln(\xi)+\ln(\eta)\\ \ln(2)+\ln(\xi+\eta)\end{array}\right]$$

használjuk az elsőfokú Taylor formulát a várható érték középponttal:

$$\left[\begin{array}{c}\tau\\ k\end{array}\right]\approx\left[\begin{array}{c}\ln(15000)\\ \ln(500)\end{array}\right]+\begin{array}{cc}\frac{1}{100} & \frac{1}{150}\\ \frac{1}{250} & \frac{1}{250}\end{array}\left[\begin{array}{c}\xi-100\\ \eta-150\end{array}\right]$$

amiből

$$E\left(\left[\begin{array}{c}\tau\\ k\end{array}\right]\right)\approx\left[\begin{array}{c}\ln(15000)\\ \ln(500)\end{array}\right]=\left[\begin{array}{c}9.6158\\ 6.2146\end{array}\right]$$

$$\operatorname{cov}\left(\left[\begin{array}{c}\tau\\ k\end{array}\right],\left[\begin{array}{c}\tau\\ k\end{array}\right]\right)\approx\left[\begin{array}{cc}\frac{1}{100} & \frac{1}{150}\\ \frac{1}{250} & \frac{1}{250}\end{array}\right]\left[\begin{array}{cc}1 & 0\\ 0 & 4\end{array}\right]\left[\begin{array}{cc}\frac{1}{100} & \frac{1}{150}\\ \frac{1}{250} & \frac{1}{250}\end{array}\right]^T=\left[\begin{array}{cc}\frac{1}{3600} & \frac{11}{75\,000}\\ \frac{11}{75\,000} & \frac{1}{12\,500}\end{array}\right]$$

**1.45. Feladat.** *Legyen a $(\xi,\eta)$ v.v.v. sűrűségfüggvénye:*

$$f(x,y)=c\cdot\exp\left(-x^2+\sqrt{2}xy-y^2-x+y+3\right)\quad (x,y)\in\mathbb{R}^2\ .$$

*Adjuk meg c értékét, $(\xi,\eta)$ kovariancia mátrixát és várható érték vektorát!*

**Megoldás:** Mivel a normális eloszlás sűrűségfüggvénye az exponensben egy kvadratikus formát tartalmaz, vizsgáljuk meg a normális eloszlás lehetőségét, amikor is a sűrűségfüggvény az alábbi alakban írható:

$$f(x,y)=\frac{1}{2\pi\cdot\sigma_1\cdot\sigma_2\cdot\sqrt{1-r^2}}\exp\left\{-\frac{1}{2}\left[\begin{array}{c}x-m_1\\ y-m_2\end{array}\right]^T V^{-1}\left[\begin{array}{c}x-m_1\\ y-m_2\end{array}\right]\right\}$$

Tehát a "tiszta" másodfokú tagokból

$$-\frac{1}{2}\left(2x^2-2\sqrt{2}xy+2y^2\right)=-\frac{1}{2}\left[\begin{array}{cc}x & y\end{array}\right]V^{-1}\left[\begin{array}{c}x\\ y\end{array}\right]$$

kell hogy teljesüljön, amiből

$$V^{-1}=\left[\begin{array}{cc}2 & \sqrt{2}\\ \sqrt{2} & 2\end{array}\right]\Rightarrow\operatorname{cov}\left(\left[\begin{array}{c}\xi\\ \eta\end{array}\right],\left[\begin{array}{c}\xi\\ \eta\end{array}\right]\right)=V=\left[\begin{array}{cc}1 & -\frac{\sqrt{2}}{2}\\ -\frac{\sqrt{2}}{2} & 1\end{array}\right]$$

ami valóban egy pozitív definit (szimmetrikus) mátrix. Az elsőfokú tagokra teljesül

$$-\frac{1}{2}(2x+2y)=-\frac{1}{2}(-2)\left[\begin{array}{c}m_1\\ m_2\end{array}\right]^T V^{-1}\left[\begin{array}{c}x\\ y\end{array}\right]=\left(2m_1+m_2\sqrt{2}\right)x+\left(m_1\sqrt{2}+2m_2\right)y\ ,$$

ezért az

$$\begin{aligned}-1&=2m_1+m_2\sqrt{2}\\ -1&=m_1\sqrt{2}+2m_2\end{aligned}$$

egyenlet megoldásával kapjuk:

$$m_1 = m_2 = -1 + \frac{\sqrt{2}}{2} \Rightarrow E\left(\begin{bmatrix} \xi \\ \eta \end{bmatrix}\right) = \begin{bmatrix} -1+\frac{\sqrt{2}}{2} \\ -1+\frac{\sqrt{2}}{2} \end{bmatrix}.$$

A konstans tag az exponensben így

$$-\frac{1}{2}\begin{bmatrix} -1+\frac{\sqrt{2}}{2} \\ -1+\frac{\sqrt{2}}{2} \end{bmatrix}^T \begin{bmatrix} 2 & \sqrt{2} \\ \sqrt{2} & 2 \end{bmatrix} \begin{bmatrix} -1+\frac{\sqrt{2}}{2} \\ -1+\frac{\sqrt{2}}{2} \end{bmatrix} = \frac{\sqrt{2}}{2} - 1$$

és a sűrűségfüggvény

$$f(x,y) = c \cdot \exp\left\{-\frac{1}{2}\begin{bmatrix} x+1-\frac{\sqrt{2}}{2} \\ y+1-\frac{\sqrt{2}}{2} \end{bmatrix}^T V^{-1} \begin{bmatrix} x+1-\frac{\sqrt{2}}{2} \\ y+1-\frac{\sqrt{2}}{2} \end{bmatrix}\right\} \cdot \exp\left(3 - \frac{\sqrt{2}}{2} + 1\right),$$

tehát teljesülni kell az

$$c \cdot \exp\left(3 - \frac{\sqrt{2}}{2} + 1\right) = \frac{1}{2\pi \cdot \sigma_1 \cdot \sigma_2 \cdot \sqrt{1-r^2}}$$

egyenletnek, ahol $\sigma_1 = \sigma_2 = 1 \quad r = -\frac{\sqrt{2}}{2}$, amiből

$$c = \frac{1}{\exp\left(4 - \frac{\sqrt{2}}{2}\right) \cdot 2\pi \cdot \sqrt{1 - \frac{1}{2}}} = \frac{\sqrt{2}}{2}\frac{e^{-4+\frac{\sqrt{2}}{2}}}{\pi} = 8.360\,8 \times 10^{-3}.$$

Tehát összefoglalva,

$$c = 8.360\,8 \times 10^{-3} \qquad \begin{bmatrix} \xi \\ \eta \end{bmatrix} \in \mathcal{N}\left(\begin{bmatrix} -1+\frac{\sqrt{2}}{2} \\ -1+\frac{\sqrt{2}}{2} \end{bmatrix}, \begin{bmatrix} 1 & -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & 1 \end{bmatrix}\right).$$

## 1.8. $\chi^2$, $\mathcal{T}$ és $\mathcal{F}$ eloszlás

**1.46. Feladat.** *Legyen* $\xi = (\xi_1, \xi_2) \in \mathcal{N}(m, V)$ *2-dimenziós normális eloszlású, ahol*

$$m = \begin{bmatrix} 10 \\ 20 \end{bmatrix} \qquad V = \begin{bmatrix} 2 & \sqrt{3} \\ \sqrt{3} & 4 \end{bmatrix}$$

*és keressünk olyan* $H \subset \mathbb{R}^2$ *tartományt, hogy*

$$P(\xi \in H) = 0.95$$

*teljesüljön!*

**Megoldás:** Mivel a $V$ kovariancia mátrix invertálható, keressük a $H$ tartományt az

$$(x-m)^T V^{-1}(x-m) \le k \quad x \in \mathbb{R}^2$$

másodfokú egyenlőtlenség megoldáshalmazaként egy alkalmas $k > 0$ számmal. Ekkor

$$\{\xi \in H\} = \left\{(W(\xi - m))^T WV^{-1}W^T (W(x-m)) \le k\right\}$$

ahol

$$W = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} v_1^T \\ \frac{1}{\sqrt{\lambda_2}} v_2^T \end{bmatrix}$$

a főfaktorok előállításának transzformációs mátrixa a $v_1, v_2$ sajátvektorokkal és $\lambda_1, \lambda_2$ sajátértékekkel, ezért

$$\{\xi \in H\} = \left\{\tau^{*T}\tau^* \le k\right\} \ ,$$

ahol $\tau^{*T}\tau^* \in \chi_2^2$, tehát táblázatból válasszuk a $k = \chi_{0.05}^2 = 5.99$ értéket, amivel teljesül

$$P\left(\left(\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} - \begin{bmatrix} 10 \\ 20 \end{bmatrix}\right)^T \begin{bmatrix} 2 & \sqrt{3} \\ \sqrt{3} & 4 \end{bmatrix}^{-1} \left(\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} - \begin{bmatrix} 10 \\ 20 \end{bmatrix}\right) \le 5.99\right) = 0.95 \ ,$$

vagy kifejtve a kvadratikus formát:

$$P\left(\frac{4}{5}\xi_1^2 - 16\xi_1 + 240 - \frac{2}{5}\sqrt{3}\xi_2\xi_1 + 4\sqrt{3}\xi_2 + 8\sqrt{3}\xi_1 - 80\sqrt{3} + \frac{2}{5}\xi_2^2 - 16\xi_2 \le 5.99\right) = 0.95$$

**Megjegyzés:** Vegyük észre, hogy az ilyen módon keresett $H$ tartományba esés valószínűsége csak $k$ értékétől (és a dimenziótól) függ, ami a skalár esetben jól ismert "$k \cdot \sigma$ szabály" megfelelője. A $H$ tartománynak szemléletes jelentése is van, ugyanis a főkomponensekkel írva

$$\{\xi \in H\} = \left\{\frac{\tau_1^2}{\lambda_1} + \frac{\tau_2^2}{\lambda_2} \le k\right\} \ ,$$

tehát $H$ egy olyan ellipszissel határolt tartomány, melynek középpontja a várható érték, a tengelyek irányvektorai a sajátvektorok, a fél-tengelyhosszak pedig

$$a = \sqrt{\lambda_1 k} \quad b = \sqrt{\lambda_2 k}$$

a sajátértékek gyökével arányos értékek.

**1.47. Feladat.** *Egy berendezés működtetéséhez szükséges alkatrész élettartama exponenciális eloszlású véletlen mennyiség, és az átlagos élettartam 100 óra*

***a)*** *Milyen határok között van egy alkatrész élettartama 90%-os valószínűséggel?*

***b)*** *Legalább hány óra működési időre számíthatunk 90%-os biztonsággal, ha 3 ilyen alkatrészünk van?*

**Megoldás:**

**a)** Jelölje $\xi \in \mathcal{E}xp\left(\frac{1}{100}\right)$ egy ilyen alkatrész élettartamát, akkor

$$\frac{2}{100} \cdot \xi \in \mathcal{E}xp\left(\frac{1}{2}\right) \equiv \chi_2^2 \ ,$$

és táblázatból (szabadsági fok: 2)

$$\chi_{0.95}^2 = 0.10 \qquad \chi_{0.05}^2 = 5.99 \ ,$$

amivel

$$P\left(0.10 \le \frac{2}{100} \cdot \xi \le 5.99\right) = 0.90 \ ,$$

tehát

$$P\left(5 \le \xi \le 300\right) = 0.90 \ .$$

**b)** Jelölje a három független élettartamot $\xi_i \in \mathcal{E}xp\left(\frac{1}{100}\right) \quad i = 1, 2, 3$, akkor $\eta = \xi_1 + \xi_2 + \xi_3$ összegükre teljesül

$$\frac{2}{100} \cdot \eta \in \chi_6^2 \ ,$$

és táblázatból (szabadsági fok: 6)

$$\chi_{0.90}^2 = 2.20 \ ,$$

amivel

$$P\left(2.20 \le \frac{2}{100} \cdot \eta\right) = 0.90 \ ,$$

tehát

$$P\left(110 \le \eta\right) = 0.90 \ .$$

**1.48. Feladat.** *Egy 100 óra várható élettartamú alkatrész élettartamánál hányszor több négy egyenként 200 órás várható élettartamú alkatrész össz-élettartama 0.95 valószínűséggel? Tegyük fel, hogy az alkatrészek élettartama független exponenciális eloszlású!*

**Megoldás:** Jelölje $\xi \in \mathcal{E}xp\left(\frac{1}{100}\right)$ és $\eta_i \in \mathcal{E}xp\left(\frac{1}{200}\right) \quad i = 1, 2, 3, 4$ a kétfle élettartamot, keressük azt a $k > 0$ számot, melyre

$$P(k\xi < \eta_1 + \eta_2 + \eta_3 + \eta_4) = 0.95$$

vagy rendezve

$$P\left(k \cdot \frac{50}{100} \cdot \frac{2}{8} < \frac{\frac{\frac{2}{200}(\eta_1+\eta_2+\eta_3+\eta_4)}{8}}{\frac{\frac{2}{100}\xi}{2}}\right) = 0.95$$

ahol

$$\frac{\frac{\frac{2}{200}(\eta_1+\eta_2+\eta_3+\eta_4)}{8}}{\frac{\frac{2}{100}\xi}{2}} \in F_{8;2} \ ,$$

ezért táblázatból (szabadsági fok: $(2;8)$) kapjuk $f_{0.05} = 4.46$, és a $(8;2)$ szabadsági fokhoz

$$f_{0.95} = \frac{1}{4.46} = k \cdot \frac{50}{100} \cdot \frac{2}{8} \ ,$$

amiből $k = 1.7937$, tehát

$$P(1.7937 \cdot \xi < \eta_1 + \eta_2 + \eta_3 + \eta_4) = 0.95 \ .$$

**1.49. Feladat (*).** *Legyen $\xi \in \mathcal{N}(\mathbf{m}; \mathbf{I}_n)$ $n$-dimenziós v.v., ahol $\mathbf{0}_n \neq \mathbf{m} \in \mathbb{R}^n$ és $\mathbf{I}_n \in \mathbb{R}^{n\times n}$ egység mátrix.*

***a)*** *Mutassuk meg, hogy a $\xi^T\xi$ skalár v.v. eloszlása csak az $m = \mathbf{m}^T\mathbf{m}$ értéktől függ. Ezt az eloszlást nem-centrált $\chi^2$-eloszlásnak nevezzük $n$ szabadsági fokkal, jelölése $\xi^T\xi \in \chi^2_n(m)$.*

***b)*** *Ajuk meg az $\eta \in \chi^2_n(m)$ v.v. karakterisztikus függvényét!*

***c)*** *Mutassuk meg, hogy a $\chi^2_n(m)$ eloszlás megadható, mint a $\chi^2_{n+2k}$ $\quad k = 0, 1, 2, \ldots$ eloszlások keveréke a $\mathcal{P}o\left(\frac{m}{2}\right)$ eloszlással, azaz*

$$\chi^2_n(m) = e^{-\frac{m}{2}} \sum_{k=0}^{\infty} \frac{\left(\frac{m}{2}\right)^k}{k!} \chi^2_{n+2k} \ .$$

**Bizonyítás.**

**a)** Ha $n = 1$, az állítás nyilvánvaló, ha $n > 1$ legyen $Q = \mathbf{I}_n - \mathbf{m}_0\mathbf{m}_0^T$ szimmetrikus mátrix, ahol $\mathbf{m}_0 = \frac{1}{\sqrt{m}}\mathbf{m} \in \mathbb{R}^n$, akkor

$$\mathbf{I}_n = Q + \mathbf{m}_0\mathbf{m}_0^T$$

és teljesül $Q \cdot \mathbf{m}_0\mathbf{m}_0^T = \mathbf{0}_{n\times n}$, ezért $QQ = Q$ és $\mathrm{rang}(Q) = n - 1$, továbbá

$$\xi^T\xi = \xi^TQ\xi + \left(\mathbf{m}_0^T\xi\right)^2$$

ahol az összeadandók függetlenek, $Q\mathbf{m}_0 = \mathbf{0}_n$, ezért $\xi^TQ\xi \in \chi^2_{n-1}$. Tehát $\xi^T\xi$ eloszlása olyan, mint egy $\chi^2_{n-1}$ eloszlású v.v. és egy tőle független $\mathbf{m}_0^T\xi \in \mathcal{N}(\sqrt{m}; 1)$ eloszlású v.v. négyzete összegének az eloszlása.

**b)** Legyen $\eta \in \chi^2_1(m)$, akkor $\eta$ karakterisztikus függvénye

$$\begin{aligned} \phi(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx^2 - \frac{(x-\sqrt{m})^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\left(it - \frac{1}{2}\right)x^2 + x\sqrt{m} - \frac{m}{2}} dx = \\ &= e^{-\frac{m}{2}} \int_{-\infty}^{\infty} \sum_{k=0}^{\infty} \frac{\left(x\sqrt{m}\right)^k}{k!} \cdot \frac{1}{\sqrt{2\pi}} e^{\left(it-\frac{1}{2}\right)x^2} dx \ . \end{aligned} \tag{1.14}$$

Mivel

$$\left|\sum_{k=0}^{l} \frac{(x\sqrt{m})^k}{k!} \cdot \frac{1}{\sqrt{2\pi}} e^{\left(it-\frac{1}{2}\right)x^2}\right| \le \sum_{k=0}^{l} \frac{|x\sqrt{m}|^k}{k!} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \le e^{\left|x\sqrt{m}\right|-\frac{1}{2}x^2} \quad l = 1, 2, \ldots$$

és a majoráló függvény integrálható, (1.14)-ben az integrálás és összegzés felcserélésével kapjuk:

$$\phi(t) = e^{-\frac{m}{2}} \sum_{k=0}^{\infty} \int_{-\infty}^{\infty} \frac{(x\sqrt{m})^k}{k!} \cdot \frac{1}{\sqrt{2\pi}} e^{\left(it-\frac{1}{2}\right)x^2} dx$$

Számoljuk ki a sor tagjait:

$k = 0$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\left(it-\frac{1}{2}\right)x^2} dx = \phi_1(t)$$

$k = 1$

$$\int_{-\infty}^{\infty} x\sqrt{m} \frac{1}{\sqrt{2\pi}} e^{\left(it-\frac{1}{2}\right)x^2} dx = \left[\frac{\sqrt{m}}{2it-1} \frac{1}{\sqrt{2\pi}} e^{\left(it-\frac{1}{2}\right)x^2}\right]_{-\infty}^{\infty} = 0$$

$k = 2$

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{(x\sqrt{m})^2}{2!} \cdot \frac{1}{\sqrt{2\pi}} e^{\left(it-\frac{1}{2}\right)x^2} dx &= \\ &= \left[\frac{x\sqrt{m}}{2!} \cdot \frac{\sqrt{m}}{2it-1} \frac{1}{\sqrt{2\pi}} e^{\left(it-\frac{1}{2}\right)x^2}\right]_{-\infty}^{\infty} + \\ &\quad + \frac{\frac{m}{1-2it}}{2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\left(it-\frac{1}{2}\right)x^2} dx = \frac{\frac{m}{1-2it}}{2} \cdot \phi_1(t)\end{aligned}$$

$\vdots$

$k = 2l - 1$

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{(x\sqrt{m})^{2l-1}}{(2l-1)!} \cdot \frac{1}{\sqrt{2\pi}} e^{\left(it-\frac{1}{2}\right)x^2} dx &= \\ &= \left[\frac{(x\sqrt{m})^{2l-2}}{(2l-1)!} \cdot \frac{\sqrt{m}}{2it-1} \frac{1}{\sqrt{2\pi}} e^{\left(it-\frac{1}{2}\right)x^2}\right]_{-\infty}^{\infty} + \\ &\quad + \frac{\frac{m}{1-2it}}{2l-1} \int_{-\infty}^{\infty} \frac{(x\sqrt{m})^{2l-3}}{(2l-3)!} \cdot \frac{1}{\sqrt{2\pi}} e^{\left(it-\frac{1}{2}\right)x^2} dx = 0\end{aligned}$$

$k = 2l$

$$\int_{-\infty}^{\infty} \frac{(x\sqrt{m})^{2l}}{(2l)!} \cdot \frac{1}{\sqrt{2\pi}} e^{\left(it-\frac{1}{2}\right)x^2} dx =$$

$$= \left[ \frac{(x\sqrt{m})^{2l-1}}{(2l)!} \cdot \frac{\sqrt{m}}{2it-1} \frac{1}{\sqrt{2\pi}} e^{\left(it-\frac{1}{2}\right)x^2} \right]_{-\infty}^{\infty} +$$

$$+ \frac{\frac{m}{1-2it}}{2l} \int_{-\infty}^{\infty} \frac{(x\sqrt{m})^{2l-2}}{(2l-2)!} \cdot \frac{1}{\sqrt{2\pi}} e^{\left(it-\frac{1}{2}\right)x^2} dx =$$

$$= \frac{\left(\frac{m}{1-2it}\right)^l}{2l \cdot (2l-2) \cdot \ldots \cdot 2} \cdot \phi_1(t) = \frac{\left(\frac{\frac{m}{2}}{1-2it}\right)^l}{l!} \cdot \phi_1(t) \quad t \in \mathbb{R}$$

$\vdots$

Tehát

$$\phi(t) = e^{-\frac{m}{2}} \sum_{n=0}^{\infty} \frac{\left(\frac{\frac{m}{2}}{1-2it}\right)^l}{l!} \cdot \phi_1(t) = e^{-\frac{m}{2}} \sum_{l=0}^{\infty} \frac{\left(\frac{m}{2}\right)^l}{l!} \cdot \left(\frac{1}{1-2it}\right)^{\frac{2l+1}{2}} =$$

$$= \left(\frac{1}{1-2it}\right)^{\frac{1}{2}} \cdot e^{-\frac{m}{2}+\frac{\frac{m}{2}}{1-2it}} = \left(\frac{1}{1-2it}\right)^{\frac{1}{2}} \cdot \exp\left(\frac{itm}{1-2it}\right) \quad t \in \mathbb{R}$$

ahol $\phi_1(t) = \left(\frac{1}{1-2it}\right)^{\frac{1}{2}} \quad t \in \mathbb{R}$ a $\chi_1^2$ eloszlás karakterisztikus függvénye, ezért az $\eta \in \chi_n^2(m)$ v.v. karakterisztikus függvénye:

$$\phi(t) = e^{-\frac{m}{2}} \sum_{k=0}^{\infty} \frac{\left(\frac{m}{2}\right)^k}{k!} \cdot \left(\frac{1}{1-2it}\right)^{\frac{2k+n}{2}} = \left(\frac{1}{1-2it}\right)^{\frac{n}{2}} \cdot \exp\left(\frac{itm}{1-2it}\right) \quad t \in \mathbb{R}\ . \tag{1.15}$$

**c)** Az állítás következik a karakterisztikus függvény (1.15) szerinti alakjából.

■

**Következmény:** Ha $\eta_1 \in \chi_{n_1}^2(m)$ és $\eta_2 \in \chi_{n_2}^2$ függetlenek, akkor az $f = \frac{\eta_1}{\eta_2} \cdot \frac{n_2}{n_1}$ v.v. eloszlását nemcentrált $F$-eloszlásnak nevezzük $(n_1, n_2)$ szabadsági fokkal, jelölése $f \in F_{(n_1,n_2)}(m)$, és

$$F_{(n_1,n_2)}(m) = e^{-\frac{m}{2}} \sum_{k=0}^{\infty} \frac{\left(\frac{m}{2}\right)^k}{k!} \cdot F_{(n_1+2k,n_2)}\ .$$

## 1.9. Regresszió analízis

**1.50. Feladat.** *A $(\xi_1, \xi_2, \xi_3)$ v.v.v. kovariancia mátrixa és várható érték vektora:*

$$\operatorname{cov}(\xi,\xi) = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 7 & 3 \\ 1 & 3 & 3 \end{bmatrix} \qquad E(\xi) = \begin{bmatrix} 10 \\ 20 \\ 30 \end{bmatrix}$$

***a)*** *Ajuk meg a $\xi_1 \sim a\xi_2 + b\xi_3 + c$ regressziós közelítést, a $\rho_{\xi_1|\xi_2,\xi_3}$ többszörös korrelációs együtthatót és a maradék szórást!*

***b)*** *Ajuk meg a $\begin{bmatrix} \xi_1 \\ \xi_3 \end{bmatrix} \sim \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \xi_3 + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ regressziós közelítést, a maradék szórásnégyzetet, és a $\rho_{\xi_1\xi_3|\xi_2}$ parciális korrelációs együtthatót*

**Megoldás:**

**a)** Jelölje

$$\begin{aligned} m_1 &= E(\xi_1) = 10 \quad m_2 = E\left(\begin{bmatrix} \xi_2 \\ \xi_3 \end{bmatrix}\right) = \begin{bmatrix} 20 \\ 30 \end{bmatrix} \\ V_{11} &= D^2(\xi_1) = 4 \quad V_{12} = \operatorname{cov}\left(\xi_1, \begin{bmatrix} \xi_2 \\ \xi_3 \end{bmatrix}\right) = \begin{bmatrix} 2 & 1 \end{bmatrix} \\ V_{22} &= \operatorname{cov}\left(\begin{bmatrix} \xi_2 \\ \xi_3 \end{bmatrix}, \begin{bmatrix} \xi_2 \\ \xi_3 \end{bmatrix}\right) = \begin{bmatrix} 7 & 3 \\ 3 & 3 \end{bmatrix} \end{aligned}$$

amivel a regressziós együtthatók "normál"-egyenlete

$$\begin{aligned} 7a + 3b &= 2 \\ 3a + 3b &= 1 \end{aligned}$$

és megoldása:

$$\begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{12} \end{bmatrix} \quad c = 10 - \begin{bmatrix} \frac{1}{4} & \frac{1}{12} \end{bmatrix} \begin{bmatrix} 20 \\ 30 \end{bmatrix} = \frac{5}{2}$$

A maradék szórásnégyzet

$$\sigma_R^2 = 4 - \begin{bmatrix} \frac{1}{4} & \frac{1}{12} \end{bmatrix} \begin{bmatrix} 2 & 1 \end{bmatrix}^T = \frac{41}{12}$$

és a meghatározottság mértéke:

$$R^2 = 1 - \frac{\frac{41}{12}}{4} = \frac{7}{48}$$

Tehát a regressziós közelítés

$$\xi_1 \sim \frac{1}{4}\xi_1 + \frac{1}{12}\xi_2 + \frac{5}{2} \ ,$$

a többszörös korrelációs együttható és a maradék szórás

$$\rho_{\xi_1|\xi_2\xi_3} = \sqrt{\frac{7}{48}} = 0.381\,88 \quad \sigma_R = \sqrt{\frac{41}{12}} = 1.848\,4 \ .$$

**b)** Jelölje

$$\begin{aligned} m_1 &= E\left(\begin{bmatrix} \xi_1 \\ \xi_3 \end{bmatrix}\right) = \begin{bmatrix} 10 \\ 30 \end{bmatrix} \quad m_2 = E(\xi_2) = 20 \\ V_{11} &= \operatorname{cov}\left(\begin{bmatrix} \xi_1 \\ \xi_3 \end{bmatrix}, \begin{bmatrix} \xi_1 \\ \xi_3 \end{bmatrix}\right) = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} \quad V_{12} = \operatorname{cov}\left(\begin{bmatrix} \xi_1 \\ \xi_3 \end{bmatrix}, \xi_2\right) = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \\ V_{22} &= D^2(\xi_2) = 7 \end{aligned}$$

amivel a regressziós együtthatókat komponensenként számolva:

$$\begin{aligned} a_1 &= \frac{2}{7} \quad b_1 = 10 - \frac{2}{7} \cdot 20 = \frac{30}{7} \\ a_2 &= \frac{3}{7} \quad b_2 = 30 - \frac{3}{7} \cdot 20 = \frac{150}{7} \end{aligned}$$

a maradék kovariancia mátrix

$$V_R = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \end{bmatrix} \frac{1}{7} \begin{bmatrix} 2 \\ 3 \end{bmatrix}^T = \begin{bmatrix} \frac{24}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{12}{7} \end{bmatrix} .$$

Tehát a regressziós közelítés

$$\begin{bmatrix} \xi_1 \\ \xi_3 \end{bmatrix} \sim \begin{bmatrix} \frac{2}{7} \\ \frac{3}{7} \end{bmatrix} \xi_3 + \begin{bmatrix} \frac{30}{7} \\ \frac{150}{7} \end{bmatrix}$$

a maradék szórásnégyzet

$$\sigma_R^2 = \frac{24}{7} + \frac{12}{7} = \frac{36}{7}$$

a parciális korrelációs együttható

$$\rho_{\xi_1\xi_3|\xi_2} = \frac{\frac{1}{7}}{\sqrt{\frac{24}{7}} \cdot \sqrt{\frac{12}{7}}} = 0.005893 .$$

**Megjegyzés:** A $\xi_1$ és $\xi_3$ kapcsolatának szorosságát "mérő"

$$\rho_{\xi_1\xi_3} = \frac{1}{\sqrt{4} \cdot \sqrt{3}} = 0.288\,68$$

korrelációs együttható lényegesen nagyobb az előző értéknél, aminek oka $\xi_2$, ugyanis ezen komponens hatására figyelhetünk meg azonos "irányú" változást $\xi_1$ és $\xi_3$ értékében. Ha pedig $\xi_2$ értéke nem változik, a cél-változó komponensei közti korreláció mértéke elhanyagolható.

**1.51. Feladat.** *Keressük meg az*

$$(-1; 2) \quad (0; 1) \quad (1; 2) \quad (2; 4) \in \mathbb{R}^2$$

*pontokra "legjobban" illeszkedő másodfokú polinomot!*

**Megoldás:** Használjuk a legkisebb négyzetek módszerének regressziós feladatként történő megfogalmazását, amihez számoljuk ($n = 4$):

| $x$ | $y$ | $x^2$ | $x^3$ | $x^4$ | $xy$ | $x^2y$ |
|---|---|---|---|---|---|---|
| $-1$ | 2 | 1 | $-1$ | 1 | $-2$ | 2 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 1 | 2 | 1 | 1 | 1 | 2 | 2 |
| 2 | 4 | 4 | 8 | 16 | 8 | 16 |
| 2 | 9 | 6 | 8 | 18 | 8 | 20 |

Amiből

$$\begin{aligned}
m_1 &= \frac{9}{4} \quad m_2 = \begin{bmatrix} \frac{6}{4} \\ \frac{2}{4} \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ \frac{1}{2} \end{bmatrix} \\
V_{12} &= \begin{bmatrix} \frac{20}{4} - \frac{9}{4} \cdot \frac{3}{2} & \frac{8}{4} - \frac{9}{4} \cdot \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{13}{8} & \frac{7}{8} \end{bmatrix} \\
V_{22} &= \begin{bmatrix} \frac{18}{4} - \left(\frac{3}{2}\right)^2 & \frac{8}{4} - \frac{3}{2} \cdot \frac{1}{2} \\ \frac{8}{4} - \frac{3}{2} \cdot \frac{1}{2} & \frac{6}{4} - \left(\frac{1}{2}\right)^2 \end{bmatrix} = \begin{bmatrix} \frac{9}{4} & \frac{5}{4} \\ \frac{5}{4} & \frac{5}{4} \end{bmatrix}
\end{aligned}$$

amiből a regressziós együtthatók normál egyenlete:

$$\begin{aligned}
\frac{9}{4}a + \frac{5}{4}b &= \frac{13}{8} \\
\frac{5}{4}a + \frac{5}{4}b &= \frac{7}{8}
\end{aligned}$$

megoldása $a = \frac{3}{4}$, $\quad b = -\frac{1}{20}$, és $\quad c = \frac{9}{4} - \frac{3}{4} \cdot \frac{3}{2} - \left(-\frac{1}{20}\right) \cdot \frac{1}{2} = \frac{23}{20}$. Tehát a regressziós közelítés eredménye az

$$y = \frac{3}{4}x^2 - \frac{1}{20}x + \frac{23}{20}$$

másodfokú függvény.

**1.52. Feladat.** *Egy férfi magasságát három szemtanú 177cm-, 181cm- és 179cm-nek becsülte. Ha tudjuk, hogy a férfiak magassága $\mathcal{N}(178; 10)$ eloszlású véletlen mennyiség, és a szemtanúk becslése ettől, és egymástól is független $\mathcal{N}(0; 2)$ eloszlású véletlen hibát tartalmaz, adjunk közelítést a látott férfi magasságára!*

**Megoldás:** Jelölje a független valószínűségi változókat

$$\begin{aligned}
\eta &\in \mathcal{N}(178; 10) \\
\varepsilon_k &\in \mathcal{N}(0; 2) \quad k = 1, 2, 3
\end{aligned}$$

akkor a három becsült érték véletlen mennyisége

$$\begin{aligned}
\xi_1 &= \eta + \varepsilon_1 \\
\xi_2 &= \eta + \varepsilon_2 \\
\xi_3 &= \eta + \varepsilon_3
\end{aligned}$$

és keressük a legjobb

$$\eta \approx E(\eta \mid \xi)$$

közelítést. Mivel

$$\begin{bmatrix} \eta \\ \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta \\ \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

ezért $(\eta, \xi_1, \xi_2, \xi_3)$ normális eloszlású vektor valószínűségi változó, melynek várható érték vektora

$$m = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 178 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 178 \\ 178 \\ 178 \\ 178 \end{bmatrix}$$

és kovariancia mátrixa

$$V = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 100 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}^T = \begin{bmatrix} 100 & 100 & 100 & 100 \\ 100 & 104 & 100 & 100 \\ 100 & 100 & 104 & 100 \\ 100 & 100 & 100 & 104 \end{bmatrix} .$$

Együttes normális eloszlás esetén a legjobb $\eta \approx E(\eta \mid \xi)$ közelítés a lineáris regressziós feladat

$$\eta \approx a_1\xi_1 + a_2\xi_2 + a_3\xi_3 + b$$

megoldása. A regressziós együtthatók

$$\begin{bmatrix} 104 & 100 & 100 \\ 100 & 104 & 100 \\ 100 & 100 & 104 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 100 \\ 100 \\ 100 \end{bmatrix}$$

normál egyenletéből kapjuk

$$\begin{aligned} a_1 &= a_2 = a_3 = \frac{25}{76} , \\ b &= 178 - 3 \cdot \frac{25}{76} \cdot 178 = \frac{89}{38} . \end{aligned}$$

Tehát a legjobb közelítés

$$\eta \approx \frac{25}{76} (\xi_1 + \xi_2 + \xi_3) + \frac{89}{38} .$$

Ezt felhasználva, a becsült értékekből számolhatjuk a látott férfi magasságának közelítését:

$$\frac{25}{76} (177 + 181 + 179) + \frac{89}{38} = 178.99\text{cm} .$$

**1.53. Feladat.** *Legyen a $(\xi,\eta)$ v.v.v. diszkrét eloszlása:*

| $\xi \diagdown \eta$ | 1 | 2 | 3 |
|---|---|---|---|
| 10 | 0.01 | 0.25 | 0.04 |
| 20 | 0.12 | 0.08 | 0.50 |

*Adjuk meg az $E(\eta \mid \xi)$ feltételes várható értéket! Mennyi a maradék szórásnégyzet?*

**Megoldás:** Adjuk meg $\eta$-nak $\xi$-re vonatkozó $\left(P(\eta = y \mid \xi = x)\right)_y$ feltételes eloszlását

| $\xi \diagdown \eta$ | 1 | 2 | 3 |
|---|---|---|---|
| 10 | $\frac{0.01}{0.01+0.25+0.04}$ | $\frac{0.25}{0.01+0.25+0.04}$ | $\frac{0.04}{0.01+0.25+0.04}$ |
| 20 | $\frac{0.12}{0.12+0.08+0.50}$ | $\frac{0.08}{0.12+0.08+0.50}$ | $\frac{0.50}{0.12+0.08+0.50}$ |

amiből kapjuk:

$$E(\eta \mid \xi = x) = \begin{cases} 1 \cdot \frac{0.01}{0.30} + 2 \cdot \frac{0.25}{0.30} + 3 \cdot \frac{0.04}{0.30} = 2.1 & x = 10 \\ 1 \cdot \frac{0.12}{0.70} + 2 \cdot \frac{0.08}{0.70} + 3 \cdot \frac{0.50}{0.70} = 2.542\,9 & x = 20 \end{cases}$$

A maradék szórásnégyzethez számoljuk ki:

$$\begin{aligned} E(\eta^2) &= 1^2 \cdot 0.13 + 2^2 \cdot 0.33 + 3^2 \cdot 0.54 = 6.31 \\ E\left(E^2(\eta \mid \xi)\right) &= 2.1^2 \cdot 0.30 + 2.542\,9^2 \cdot 0.70 = 5.849\,4 \end{aligned}$$

amiből

$$\sigma_R^2 = E(\eta^2) - E\left(E^2(\eta \mid \xi)\right) = 6.31 - 5.849\,4 = 0.460\,6 \ .$$

**1.54. Feladat.** *Legyen a $(\xi,\eta)$ v.v.v. sűrűségfüggvénye:*

$$f(x,y) = x \cdot y^{x-1} \quad \text{ha } 0 < x < 1 \ \text{ és } \ 0 < y < 1 \ .$$

*Adjuk meg az $E(\eta \mid \xi)$ feltételes várható értéket, és a közelítés maradék szórásnégyzetét.*

**Megoldás:** Adjuk meg $\xi$ sűrűségfüggvényét

$$f_\xi(x) = \int_0^1 x \cdot y^{x-1} dy = 1 \quad \text{ha } 0 < x < 1 \ (\Rightarrow \xi \in \mathcal{U}(0;1)) \ ,$$

amivel a feltételes sűségfüggvény

$$f_{\eta|\xi}(y|x) = \frac{f(x,y)}{f_\xi(x)} = x \cdot y^{x-1} \quad \text{ha } 0 < x < 1 \ \text{ és } \ 0 < y < 1 \ .$$

Tehát

$$E(\eta \mid \xi = x) = \int_0^1 y \cdot x \cdot y^{x-1} dy = \frac{x}{1+x} \quad \text{ha } 0 < x < 1$$

vagyis

$$E(\eta \mid \xi) = \frac{\xi}{1+\xi} \ .$$

A maradék szórásnégyzethez számoljuk ki :

$$\begin{aligned} E(\eta^2) &= \int_0^1 \int_0^1 y^2 x \cdot y^{x-1} dx dy = 1 + 2\ln 2 - 2\ln 3 \\ E\left(E^2(\eta \mid \xi)\right) &= \int_0^1 \left(\frac{x}{1+x}\right)^2 dx = \frac{3}{2} - 2\ln 2 \end{aligned}$$

amivel kapjuk:

$$\sigma_R^2 = 1 + 2\ln 2 - 2\ln 3 - \left(\frac{3}{2} - 2\ln 2\right) = 7.5364 \times 10^{-2} \ .$$

**1.55. Feladat.** *Tudjuk, hogy a férfiak magassága $\mathcal{N}(180; 10)$, a nők magassága pedig $\mathcal{N}(172; 10)$ eloszlású véletlen mennyiség. Ha egy 60 férfiból és 40 nőből álló társaságból találomra választott személy 175cm magas, mennyi annak valószínűsége, hogy nőt, illetve férfit választottunk? Adjunk döntési szabályt a magasság alapján egy találomra választott személy nemére!*

**Megoldás:** Jelölje $\xi$ a választott személy magasságát, továbbá

$F$ a választott személy férfi
$N$ a válsztott személy nő

akkor feltehetjük

$$P(F) = 0.6 \quad P(N) = 0.4 \ ,$$

és a magasság feltételes sűrűségfüggvényei

$$f_F(x) = \frac{1}{\sqrt{2\pi} \cdot 10} \exp\left(-\frac{(x-180)^2}{200}\right) \quad f_N(x) = \frac{1}{\sqrt{2\pi} \cdot 10} \exp\left(-\frac{(x-172)^2}{200}\right) \quad x \in \mathbb{R} \ ,$$

amivel

$$P(F \mid \xi = 175) = \frac{\frac{0.6}{\sqrt{2\pi}\cdot 10} \exp\left(-\frac{1}{200}(175-180)^2\right)}{\frac{0.6}{\sqrt{2\pi}\cdot 10} \exp\left(-\frac{1}{200}(175-180)^2\right) + \frac{0.4}{\sqrt{2\pi}\cdot 10} \exp\left(-\frac{1}{200}(175-172)^2\right)} = 0.58066$$

$$P(N \mid \xi = 175) = 1 - 0.58066 = 0.41934$$

A Bayes döntéshez

$$d^*(x) = 1 \Leftrightarrow \frac{0.6}{\sqrt{2\pi} \cdot 10} \exp\left(-\frac{1}{200}(x-180)^2\right) > \frac{0.4}{\sqrt{2\pi} \cdot 10} \exp\left(-\frac{1}{200}(x-172)^2\right) \Leftrightarrow 170.93 < x \ .$$

Tehát

$$d^*(x) = \begin{cases} 1 & 170.93 < x \\ 2 & 170.93 \geq x \end{cases}$$

és a hibás döntés valószínűsége:

$$\begin{aligned} P(H_{d^*}) &= 1 - (P(\xi > 170.93|F) \cdot P(F) + P(\xi \leq 170.93|N) \cdot P(N)) = \\ &= 1 - \left(1 - \Phi\left(\frac{170.93 - 180}{10}\right)\right) \cdot 0.6 - \Phi\left(\frac{170.93 - 172}{10}\right) = \\ &= 1 - 0.8178 \cdot 0.6 - 0.45739 \cdot 0.4 = 0.32636 \end{aligned}$$

**1.56. Feladat.** *Legyenek* $\eta \in \mathcal{N}(0;\sigma)$, $\varepsilon_k \in \mathcal{N}(0;\sigma_0)$ $\quad k = 1,2,\ldots,n$ *független v.v.-k, és*

$$\xi_k = \eta + \varepsilon_k \quad k = 1,2,\ldots,n\ ,$$

*adjuk meg az* $E(\eta \mid \xi_1,\xi_2,\ldots,\xi_n)$ *feltételes várható értéket!*

**Megoldás:** Mivel $(\eta,\xi_1,\xi_2,\ldots,\xi_n) \in \mathcal{N}\left(\mathbf{0}_{n+1};V\right)$, ahol

$$V = \begin{bmatrix} \sigma^2 & \sigma^2 & \sigma^2 & \cdots & \sigma^2 \\ \sigma^2 & \sigma_0^2+\sigma^2 & \sigma^2 & \cdots & \sigma^2 \\ \sigma^2 & \sigma^2 & \sigma_0^2+\sigma^2 & \cdots & \sigma^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma^2 & \sigma^2 & \sigma^2 & \cdots & \sigma_0^2+\sigma^2 \end{bmatrix},$$

a feltételes várható érték a lineáris regressziós függvény, melynek paraméterei $a = a_1 = a_2 = \cdots a_n$ azonosak a független változók szimmetrikus eloszlása miatt. Tehát

$$\left(\sigma_0^2+\sigma^2\right)\cdot a + \sigma^2\cdot a + \cdots + \sigma^2 \cdot a = \sigma^2 \Longrightarrow a = \frac{\sigma^2}{n\sigma^2+\sigma_0^2}$$

és így kapjuk

$$E(\eta \mid \xi_1,\xi_2,\ldots,\xi_n) = \frac{\sigma^2}{n\sigma^2+\sigma_0^2}\sum_{i=1}^{n}\xi_i = \frac{1}{1+\frac{\sigma_0^2}{n\sigma^2}}\cdot\overline{\xi}$$

és a maradék szórásnégyzet, vagyis az $\eta \approx \frac{1}{1+\frac{\sigma_0^2}{n\sigma^2}}\cdot\overline{\xi}$ közelítés hibája

$$\sigma_R^2 = \sigma^2 - \frac{(n-1)\,\sigma^4}{n\sigma^2+\sigma_0^2} = \sigma^2\cdot\frac{\sigma_0^2+\sigma^2}{n\sigma^2+\sigma_0^2}\ .$$

**1.57. Feladat.** *Legyenek a* $\xi_1,\xi_2,\ldots,\xi_n \in \mathcal{U}(0;d)$ *valószínűségi változók függetlenek, és jelölje a legnagyobbat*

$$\xi_n^* = \max\left(\xi_1,\xi_2,\ldots,\xi_n\right)\ .$$

*Adjuk meg az* $E(\xi_k|\xi_n^*)$ $\quad k = 1,2,\ldots,n \quad$ *feltételes várható értéket!*

**Megoldás:** Mivel $(\xi_k, \xi_n^*)$ eloszlása nem lehet folytonos, ugyanis $P(\xi_k = \xi_n^*) = \frac{1}{n}$, a feladatot átfogalmazzuk egy folytonos eloszlással kezelhető esetre. $(\xi_k, \xi_n^*)$ eloszlása azonos minden $k = 1, 2, \ldots, n$ esetben, ezért a keresett feltételes várható érték

$$E(\xi_k|\xi_n^*) = E\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i|\xi_n^*\right) = E\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i^*|\xi_n^*\right) = \frac{1}{n}\left(E\left(\sum_{i=1}^{n-1}\xi_i^*|\xi_n^*\right) + \xi_n^*\right)$$

ahol

$$\begin{aligned}
\xi_1^* &= \min\{\xi_1, \xi_2, \ldots, \xi_n\} \\
\xi_2^* &= \min\{\xi_1, \xi_2, \ldots, \xi_{k-1}, \xi_{k+1}, \ldots \xi_n\} \quad \text{ahol } \xi_k = \xi_1^* \\
&\vdots \\
\xi_n^* &= \max(\xi_1, \xi_2, \ldots, \xi_n)
\end{aligned}$$

Az 1.13. feladat szerint kaphatjuk $(\xi_1^*, \xi_2^*, \ldots, \xi_{n-1}^*)$ feltételes sűrűségfüggvényét $\xi_n^* = y$ esetén

$$f(x_1, \ldots, x_{n-1}|y) = \frac{n!\frac{1}{d^n}}{n\frac{y^{n-1}}{d^n}} = (n-1)!\frac{1}{y^{n-1}} \quad 0 < x_1 < x_2 < \ldots < x_{n-1} \quad 0 < y < d\ ,$$

ezért $\left(\xi_1^*, \xi_2^*, \ldots, \xi_{n-1}^*\right)$ feltételes eloszlása olyan, mint egy $\mathcal{U}(0; y)$ eloszlású $n$ számú független v.v. rendezéséből származó vektoré, tehát

$$E\left(\sum_{i=1}^{n-1}\xi_i^*|\xi_n^* = y\right) = E\left(\sum_{i=1}^{n-1}\xi_i|\xi_n^* = y\right) = \frac{n-1}{2}y \quad 0 < y < d$$

és így kapjuk

$$E(\xi_k|\xi_n^*) = \frac{n-1}{2n}\xi_n^* + \frac{1}{n}\xi_n^* = \frac{n+1}{2}\xi_n^*\ .$$

**1.58. Feladat.** *Használjuk az előző feladat jelöléseit, és mutassuk meg, hogy*

$$E(\xi_i^*) = \frac{i}{n+1}d \quad \text{és} \quad E(\xi_i^* \mid \xi_n^*) = \frac{i}{n}\xi_n^* \quad i = 1, 2, \ldots, n\ .$$

**Bizonyítás.** Az állítás nyilván teljesül az $n = 1$ esetben. Tegyük most fel, hogy $n \in \mathbb{N}$ esetén teljesül az állítás minden $d > 0$ esetben. Mivel $\xi_{n+1}^*$ sűrűségfüggvénye (lásd: 1.13. feladat)

$$f_{\xi_{n+1}}^*(y) = (n+1)\frac{y^n}{d^{n+1}} \quad 0 < y < d\ ,$$

a $(\xi_1^*, \xi_2^*, \ldots, \xi_n^*)$ v.v.v. $\xi_{n+1}^* = y$ feltétel melletti sűrűségfüggvénye

$$f(x_1, \ldots, x_n|y) = \frac{(n+1)!\frac{1}{d^{n+1}}}{(n+1)\frac{y^n}{d^{n+1}}} = n!\frac{1}{y^n} \quad 0 < x_1 < x_2 < \ldots < x_n \quad 0 < y < d$$

ezért $(\xi_1^*, \xi_2^*, \ldots, \xi_n^*)$ feltételes eloszlása olyan, mint egy $\mathcal{U}(0; y)$ eloszlású $n$ számú független v.v. rendezéséből származó vektoré. Feltételezésünk szerint tehát teljesül

$$E(\xi_i^* | \xi_{n+1}^* = y) = \frac{i}{n+1} y \quad 0 < y < d \ , \quad i = 1, 2, \ldots, n+1$$

továbbá

$$E(\xi_{n+1}^*) = \int_0^d y \cdot (n+1) \cdot \frac{y^n}{d^{n+1}} dy = \frac{(n+1)}{d^{n+1}} \cdot \left[ \frac{x^{n+2}}{n+2} \right]_0^d = \frac{n+1}{(n+1)+1} d$$

és

$$E(\xi_i^*) = E\left( \frac{i}{n+1} \xi_{n+1}^* \right) = \frac{i}{n+1} \cdot \frac{n+1}{(n+1)+1} d = \frac{i}{(n+1)+1} d$$

amit bizonyítani kellett. ■

**1.59. Feladat (*).** *Két pózna között kifeszített drótra $n \geq 2$ számú madár száll le véletlenszerűen válsztott helyekre. Minden madártól, a hozzá legközelebbi madárig tartó drótszakaszt fessük be. Ha $n \to \infty$, hova tart a befestett részek várható össz hosszának és a drót teljes hosszának aránya?*

**Megoldás:** Feltehetjük, hogy a drót teljes hossza egységnyi, és a madarak helye az egyik végpottól mérve, a

$$\xi_1, \xi_2, \cdots, \xi_n \in \mathcal{U}(0; 1)$$

független valószínűségi változókkal adott. Rendezzük ezeket nagyság szerint, akkor kapjuk a

$$\xi^* = (\xi_1^*, \xi_2^*, \cdots, \xi_n^*)$$

v.v.v.-t, melynek sűrűségfüggvénye:

$$f_{\xi^*}(x_1, x_2, \cdots, x_n) = n! \qquad 0 < x_1 < x_2 < \cdots < x_n < 1 \ .$$

Vezessük be a

$$\delta_1 = \xi_1^* \quad \delta_2 = \xi_2^* - \xi_1^* \quad \cdots \quad \delta_n = \xi_n^* - \xi_{n-1}^*$$

valószínűségi változókat, akkor a $\delta = (\delta_1, \delta_2, \cdots, \delta_n)$ v.v.v. sűrűségfüggvénye:

$$f_\delta(x_1, x_2, \cdots, x_n) = n! \qquad 0 < x_i \quad i = 1, 2, \ldots, n \quad \sum_{i=1}^n x_i < 1 \tag{1.16}$$

ami a változók szimmetrikus függvénye az $n$-dimenziós egység élű szimplexen, ezért például minden perem azonos eloszlású. A vizsgálandó arány, vagyis a befestett részek hossza

$$\begin{aligned} \eta \ = \ & \delta_2 + \delta_3 \cdot \mathbf{1}_{\{\delta_2 > \delta_3 \vee \delta_4 > \delta_3\}} + \delta_4 \cdot \mathbf{1}_{\{\delta_3 > \delta_4 \vee \delta_5 > \delta_4\}} + \cdots \\ & + \delta_{n-1} \cdot \mathbf{1}_{\{\delta_{n-2} > \delta_{n-1} \vee \delta_n > \delta_{n-1}\}} + \delta_n \end{aligned}$$

aminek várható értéke

$$E(\eta) = 2 \cdot E(\delta_2) + (n-3) \cdot E\left(\delta_3 \cdot \mathbf{1}_{\{\delta_2 > \delta_3 \vee \delta_4 > \delta_3\}}\right) .$$

Mivel

$$\sum_{i=1}^{n} \delta_i = \xi_n^*$$

kapjuk

$$nE(\delta_2) = E(\xi_n^*) = \frac{n}{n+1} \Rightarrow E(\delta_2) = \frac{1}{n+1} .$$

Vizsgáljuk most az

$$E\left(\delta_3 \cdot \mathbf{1}_{\{\delta_2 > \delta_3 \vee \delta_4 > \delta_3\}}\right) = \frac{1}{n} E\left(n\delta_3 \cdot \mathbf{1}_{\{n\delta_2 > n\delta_3 \vee n\delta_4 > n\delta_3\}}\right)$$

várható értéket, ami az $(n\delta_2, n\delta_3, n\delta_4)$ v.v.v. eloszlásától függ. A $\delta$ vektor $(\delta_1, \delta_2, \delta_3)$ peremének sűrűségfüggvénye a (1.16) sűrűségfüggvényből:

$$f_{(\delta_2,\delta_3,\delta_3)}(x,y,z) = n(n-1)(n-2) \cdot (1-x-y-z)^{n-3} \qquad 0 < x,y,z \quad x+y+z < 1$$

amiből a $(n\delta_2, n\delta_3, n\delta_4)$ vektor sűrűségfüggvénye

$$f_{(n\delta_2,n\delta_3,n\delta_3)}(x,y,z) = \frac{n(n-1)(n-2)}{n^3} \cdot \left(1 - \frac{x+y+z}{n}\right)^{n-3} \qquad 0 < x,y,z \quad x+y+z < n .$$

Mivel a

$$\lim_{n\to\infty} f_{(n\delta_2,n\delta_3,n\delta_3)}(x,y,z) = e^{-x} \cdot e^{-y} \cdot e^{-z} \qquad 0 < x,y,z$$

határérték három független $\mathcal{E}xp(1)$ eloszlású v.v. együttes sűrűségfüggvénye, megmutatjuk, hogy

$$\begin{aligned} E\left(\tau_3 \cdot \mathbf{1}_{\{\tau_1 > \tau_2 \vee \tau_3 > \tau_2\}}\right) &= \lim_{n\to\infty} E\left(n\delta_3 \cdot \mathbf{1}_{\{n\delta_2 > n\delta_3 \vee n\delta_4 > n\delta_3\}}\right) = \\ &= \lim_{n\to\infty} \iiint_{\mathbb{R}^3} \mathbf{1}_{\{n>x>0 \wedge n-x>y>0 \wedge n-x-y>z>0\}} \cdot \mathbf{1}_{\{x>y \vee z>y\}} \cdot y \cdot \\ &\quad \cdot \frac{n(n-1)(n-2)}{n^3} \cdot \left(1 - \frac{x+y+z}{n}\right)^{n-3} dxdydz \end{aligned}$$

ahol $\tau_1, \tau_2\tau_3 \in \mathcal{E}xp(1)$ független v.v.-k. Ehhez elég ez utóbbi integrandus sorozathoz integrálható majoránst keresni, mint pl.:

$$\begin{aligned} 0 \leq \mathbf{1}_{\{n>x>0 \wedge n-x>y>0 \wedge n-x-y>z>0\}} \cdot \mathbf{1}_{\{x>y \vee z>y\}} \cdot y \cdot \frac{n(n-1)(n-2)}{n^3} \cdot \left(1 - \frac{x+y+z}{n}\right)^{n-3} \leq \\ \leq y \cdot \exp\left(-\frac{1}{2}(x+y+z)\right) \qquad x,y,z \in \mathbb{R}^+ \end{aligned}$$

Vizsgáljuk most a következő feltételes várható értéket:

$$E\left(\tau_3 \cdot \mathbf{1}_{\{\tau_1>\tau_2 \vee \tau_3>\tau_2\}} | \tau_3 = y\right) = yP\left(\tau_1 > y \vee \tau_3 > y\right) = y \cdot \left(2e^{-y} - e^{-2y}\right) \qquad 0 < y \ ,$$

amiből

$$E\left(\tau_3 \cdot \mathbf{1}_{\{\tau_1>\tau_2 \vee \tau_3>\tau_2\}}\right) = \int_0^\infty y \cdot \left(2e^{-y} - e^{-2y}\right) e^{-y} dy = \frac{7}{18} \ .$$

Továbbá

$$\lim_{n\to\infty} E(\eta) = \lim_{n\to\infty} \left[2 \cdot E\left(\delta_2\right) + (n-3) \cdot E\left(\delta_3 \cdot \mathbf{1}_{\{\delta_2>\delta_3 \vee \delta_4>\delta_3\}}\right)\right] =$$
$$= \lim_{n\to\infty} \left[\frac{2}{n+1} + \frac{n-3}{n} E\left(n\delta_3 \cdot \mathbf{1}_{\{n\delta_2>n\delta_3 \vee n\delta_4>n\delta_3\}}\right)\right] = E\left(\tau_3 \cdot \mathbf{1}_{\{\tau_1>\tau_2 \vee \tau_3>\tau_2\}}\right) = \frac{7}{18} \ .$$

**Megjegyzések:**

1. Ha $n = 2$, akkor $\eta = \delta_2$ és $E(\eta) = \frac{1}{3}$, ha pedig $n = 3$, akkor $\eta = \delta_2 + \delta_3$ és $E(\eta) = 2 \cdot \frac{1}{4} = \frac{1}{2}$.

2. Hasonlóan számolhatjuk a $\lim_{n\to\infty} D(\eta)$ határértéket is, ugyanis

$$\begin{aligned}\eta^2 \ = \ & \delta_2^2 + \delta_3^2 \mathbf{1}_{\{\delta_2>\delta_3 \vee \delta_4>\delta_3\}} + \delta_4^2 \mathbf{1}_{\{\delta_3>\delta_4 \vee \delta_5>\delta_4\}} + \cdots + \delta_{n-1}^2 \mathbf{1}_{\{\delta_{n-2}>\delta_{n-1} \vee \delta_n>\delta_{n-1}\}} + \\ & + \delta_n^2 + 2\delta_2\delta_n + 2\sum_{i=3}^{n-1} \delta_2\delta_i \mathbf{1}_{\{\delta_{i-1}>\delta_i \vee \delta_{i+1}>\delta_i\}} + 2\sum_{i=3}^{n-1} \delta_n\delta_i \mathbf{1}_{\{\delta_{i-1}>\delta_i \vee \delta_{i+1}>\delta_i\}} + \\ & + 2 \sum_{i\neq j=3,4,\ldots,n-1} \delta_i\delta_j \mathbf{1}_{\{\delta_{i-1}>\delta_i \vee \delta_{i+1}>\delta_i\}} \mathbf{1}_{\{\delta_{j-1}>\delta_j \vee \delta_{j+1}>\delta_j\}}\end{aligned}$$

amiből

$$\lim_{n\to\infty} E(\eta^2) = \left(\frac{7}{18}\right)^2$$

felhasználva

$$\lim_{n\to\infty} E(\delta_2^2) = \lim_{n\to\infty} E(\delta_n^2) = \lim_{n\to\infty} E\left(2\delta_2\delta_n\right) = 0$$

$$0 \le E\left(2\sum_{i=3}^{n-1}\delta_2\delta_i\mathbf{1}_{\{\delta_{i-1}>\delta_i\vee\delta_{i+1}>\delta_i\}}\right) \le \frac{2(n-3)}{n^2}\sqrt{E((n\delta_2)^2)E((n\delta_3)^2)} \to 0$$

$$0 \le E\left(2\sum_{i=3}^{n-1}\delta_n\delta_i\mathbf{1}_{\{\delta_{i-1}>\delta_i\vee\delta_{i+1}>\delta_i\}}\right) \le \frac{2(n-3)}{n^2}\sqrt{E((n\delta_n)^2)E((n\delta_3)^2)} \to 0$$

$$0 \le E\left(2\sum_{i=3}^{n-2}\delta_{i+1}\delta_i\mathbf{1}_{\{\delta_{i-1}>\delta_i\vee\delta_{i+1}>\delta_i\}}\cdot\mathbf{1}_{\{\delta_i>\delta_{i+1}\vee\delta_{i+2}>\delta_{i+1}\}}\right) \le$$

$$\le \frac{n-4}{n^2}\sqrt{E((n\delta_4)^2)E((n\delta_3)^2)} \to 0$$

$$0 \le E\left(2\sum_{i=3}^{n-3}\delta_{i+2}\delta_i\cdot\mathbf{1}_{\{\delta_{i-1}>\delta_i\vee\delta_{i+1}>\delta_i\}}\mathbf{1}_{\{\delta_{i+1}>\delta_{i+2}\vee\delta_{i+3}>\delta_{i+2}\}}\right) \le$$

$$\le \frac{n-5}{n^2}\sqrt{E((n\delta_5)^2)E((n\delta_3)^2)} \to 0$$

$$\lim_{n\to\infty} E\left(2\sum_{i\neq j=3,4,\dots,n-1}\delta_i\delta_j\mathbf{1}_{\{\delta_{i-1}>\delta_i\vee\delta_{i+1}>\delta_i\}}\mathbf{1}_{\{\delta_{ji-1}>\delta_j\vee\delta_{j+1}>\delta_j\}}\right) =$$

$$= \lim_{n\to\infty}\frac{(n-3)(n-4)-(n-4)-(n-5)}{n^2}\cdot$$

$$\cdot E\left((n\delta_3)(n\delta_6)\mathbf{1}_{\{n\delta_2>n\delta_3\vee n\delta_4>n\delta_3\}}\mathbf{1}_{\{n\delta_4>n\delta_6\vee n\delta_7>n\delta_6\}}\right) = \left(\frac{7}{18}\right)^2$$

$$E\left(\tau_3\tau_6\mathbf{1}_{\{\tau_2>\tau_3\vee\tau_4>\tau_3\}}\mathbf{1}_{\{\tau_5>\tau_6\vee\tau_7>\tau_6\}}\right) = \left(E\left(\tau_3\cdot\mathbf{1}_{\{\tau_1>\tau_2\vee\tau_3>\tau_2\}}\right)\right)^2 = \left(\frac{7}{18}\right)^2$$

Tehát kaptuk $\lim_{n\to\infty} D(\eta) = 0$, vagyis $\eta$ sztochasztikusan konvergál a $\frac{7}{18}$ arányhoz.

## 1.10. Sztochasztikus folyamatok

**1.60. Feladat.** *Egy urnában 2 piros és 1 fehér golyó van. Egymás után véletlenszerűen kiveszünk egy golyót és minden húzás után egy fehéret beteszünk a kivett golyó helyett. Jelölje $\xi_n$ az n-edik húzás után az urnában lévő piros golyók számát, $n = 1, 2, \ldots$. Markov láncot alkot-e a $(\xi_n)_{n\in\mathbb{N}}$ idősor? Ha igen, adjuk meg az átmenetvalószínűségek mátrixát, van-e stacionárius eloszlás, mennyi a $P(\xi_2 \ge 1)$ valószínűség?*

**Megoldás:** A kezdő $\xi_0$ állapot eloszlása

$$P(\xi_0 = 2) = 1\ ,$$

és teljesül

$$P(\xi_{n+1}=l|\xi_n=k,\xi_{n-1}=k_1,\cdots)=\begin{cases}\frac{3-k}{3} & \text{ha } l=k\\ \frac{k}{3} & \text{ha } l=k-1\\ 0 & \text{egyébként}\end{cases}\qquad k=0,1,2$$

tehát $(\xi_n)_{n\in\mathbb{N}}$ homogén Markov lánc $\{0,1,2\}$ állapotérrel és

$$\Pi=\begin{bmatrix}1&0&0\\ \frac{1}{3}&\frac{2}{3}&0\\ 0&\frac{2}{3}&\frac{1}{3}\end{bmatrix}$$

az átmenetvalószínűségek mátrixa. Keressük a stacionárius elsoszlást, mint az

$$\begin{aligned}\begin{bmatrix}x&y&z\end{bmatrix}\Pi &= \begin{bmatrix}x&y&z\end{bmatrix}\\ x+y+z &= 1 \quad x,y,z\geq 0\end{aligned}$$

egyenlet megoldását, azaz

$$\begin{array}{ccccccc}x&+&\frac{1}{3}y&&&=&x\\ &&\frac{2}{3}y&+&\frac{2}{3}z&=&y\\ &&&&\frac{1}{3}z&=&z\\ x&+&y&+&z&=&1\end{array}$$

amiből kapjuk az egyetlen stacionárius eloszlást:

$$\mathbf{p}=\begin{bmatrix}1&0&0\end{bmatrix}^T .$$

A $\xi_2$ állapot eloszlása

| $k$ | $P(\xi_2=k)$ |
|---|---|
| 0 | $\frac{2}{9}$ |
| 1 | $\frac{2}{3}$ |
| 2 | $\frac{1}{9}$ |

$$\Rightarrow P(\xi_2\geq 1)=\frac{2}{3}+\frac{1}{9}=\frac{7}{9}$$

mivel

$$\begin{bmatrix}0&0&1\end{bmatrix}\begin{bmatrix}1&0&0\\ \frac{1}{3}&\frac{2}{3}&0\\ 0&\frac{2}{3}&\frac{1}{3}\end{bmatrix}^2=\begin{bmatrix}\frac{2}{9}&\frac{2}{3}&\frac{1}{9}\end{bmatrix} .$$

**1.61. Feladat.** *Egy mozgó pont helyzetét a $(\xi_t)_{t\in\mathbb{R}_0^+}$ Wiener folyamt adja meg, és $D(\xi_2)=2$. Mennyi annak valószínűsége, hogy a pont egy adott időpillantbeli tartózkodási helyétől 4 időegység alatt $2\sqrt{2}$ távolság-egységnél jobban távolodik el.*

**Megoldás:** Mivel

$$D(\xi_2)=2=\sigma\cdot\sqrt{2}\Rightarrow\sigma=\sqrt{2}\Rightarrow\delta=\xi_{t+4}-\xi_t\in\mathcal{N}(0;\sqrt{2}\cdot 2) ,$$

a keresett valószínűség:

$$P\left(|\delta|>2\sqrt{2}\right)=1-\Phi\left(\frac{2\sqrt{2}}{\sqrt{2}\cdot 2}\right)+\Phi\left(-\frac{2\sqrt{2}}{\sqrt{2}\cdot 2}\right)=2-2\Phi(1)=0.317\,31 .$$

**1.62. Feladat.** *Legyen*

$$\xi_t = t \cdot \tau \quad t \in \mathbb{R}_0^+$$

*ahol* $\tau$ *egyenletes eloszlású valószínűségi változó a* $[0;4]$ *intervallumon.*

**a.** *Adjuk meg a várható érték és kovariancia függvényeket!*

**b.** *Stacionárius a* $(\xi_t)_{t\in\mathbb{R}}$ *folyamat?*

**c.** *Stacionárius illetve független növekményű a folyamat?*

**d.** *Markov tulajdonságú a folyamat?*

**e.** *Martingál a folyamat?*

**Megoldás:**

**a.**

$$m(t) = E(t\cdot\tau) = t\cdot 2 \quad \Gamma(t,s) = \mathrm{cov}(t\cdot\tau, s\cdot\tau) = t\cdot s\cdot D^2(\tau) = t\cdot s\cdot\frac{4}{3} \quad t,s\in\mathbb{R}_0^+$$

**b.** Nem, mert pl. $m(.)$ nem állandó.

**c.** A növekmények stacionáriusak, mert

$$\xi_{t_2} - \xi_{t_1} = \xi_{t_2+h} - \xi_{t_1+h} = (t_2 - t_1)\,\tau \quad \xi_{t_3} - \xi_{t_2} = \xi_{t_3+h} - \xi_{t_2+h} = (t_3 - t_2)\tau \quad \ldots$$

tehát a növekmények nem csak azonos eloszlásúak, hanem most még azonosak is. A folyamat nem független növekményű, mert a

$$(t_2 - t_1)\,\tau \quad (t_3 - t_2)\tau \quad \ldots$$

növekmények nem függetlenek.

**d.** Igen, mert $0 < t_1 < t_2 < \cdots < t_n < t$ esetén

$$P(\xi_t < x \mid \xi_{t_n}, \ldots \xi_{t_1}) = \begin{cases} 1 & \text{ha } t\cdot\frac{\xi_{t_n}}{t_n} < x \\ 0 & \text{egyébként} \end{cases} = P(\xi_t < x \mid \xi_{t_n}) \quad x\in\mathbb{R}\,.$$

**e.** Nem, mert ha $0 < t_1 < t_2 < \cdots < t_n < t$

$$E\left(\xi_t \mid \xi_{t_n}, \ldots \xi_{t_1}\right) = t\cdot\frac{\xi_{t_n}}{t_n} \neq \xi_{t_n}\,.$$

**1.63. Feladat.** *Legyen*

$$\begin{aligned} \xi_0 &= 0 \\ \xi_{n+1} &= \frac{\sqrt{2}}{2}\xi_n + \varepsilon_{n+1} \quad n = 1,2,3\ldots \\ \textit{ahol } \varepsilon_n &\in \mathcal{N}(0,1) \quad n = 1,2,3\ldots \textit{ függetlenek.} \end{aligned}$$

**a.** *Adjuk meg a $(\xi_n)_{n\in\mathbb{N}}$ idősor véges dimenziós peremeloszlásait!*

**b.** *Mennyi annak valószínűsége, hogy $P(\xi_{1000} > 2)$ ?*

**Megoldás:**

**a.** Mivel egy véges dimenziós perem a diszkrét fehér zaj idősor megfelelllő peremének lineáris transzformáltja, pl.

$$\begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \vdots \\ \xi_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \frac{\sqrt{2}}{2} & 1 & 0 & \cdots & 0 \\ \left(\frac{\sqrt{2}}{2}\right)^2 & \frac{\sqrt{2}}{2} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \left(\frac{\sqrt{2}}{2}\right)^{n-1} & \left(\frac{\sqrt{2}}{2}\right)^{n-2} & \left(\frac{\sqrt{2}}{2}\right)^{n-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \xi_3 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

ezért az idősor peremei normális eloszlásúak, amit meghatároz a várható érték és kovariancia függvény. Tehát $n = 1, 2, 3, \ldots$ esetén

$$\begin{aligned} m_n &= E(\xi_n) = \left(\frac{\sqrt{2}}{2}\right)^{n-1} E(\varepsilon_1) + \cdots + E(\varepsilon_n) = 0 \\ \Gamma(n,n) &= D^2(\xi_n) = \left(\frac{\sqrt{2}}{2}\right)^{2(n-1)} D^2(\varepsilon_1) + \cdots + D^2(\varepsilon_n) = \frac{\frac{1}{2^n} - 1}{\frac{1}{2} - 1} = 2 - \frac{1}{2^{n-1}} \end{aligned}$$

továbbá $s < t \in \mathbb{N}^+$ esetén

$$\begin{aligned} \Gamma(s,t) &= E(\xi_s \cdot \xi_t) = \left(\frac{\sqrt{2}}{2}\right)^{s+t-2} E(\varepsilon_1^2) + \cdots + \left(\frac{\sqrt{2}}{2}\right)^{t-s} E(\varepsilon_s^2) = \\ &= \left(\frac{\sqrt{2}}{2}\right)^{t-s} \frac{\frac{1}{2^s} - 1}{\frac{1}{2} - 1} = 2\left(\frac{\sqrt{2}}{2}\right)^{t-s} \left(1 - \frac{1}{2^{s-1}}\right) \end{aligned}$$

és tetszőleges $s, t \in \mathbb{N}^+$ esetén

$$\Gamma(s,t) = 2\left(\frac{\sqrt{2}}{2}\right)^{|t-s|} \left(1 - \frac{1}{2 \cdot 2^{\min\{s,t\}}}\right) .$$

Ez egyben azt is jelenti, hogy a $(\xi_n)_{n=N,N+1,\ldots}$ idősor stacionárius Gauss idősornak tekinthető elég nagy $N$ esetén.

**b.** Mivel közelítően $\xi_{1000} \in \mathcal{N}(0; 2)$, a keresett valószínűség

$$P(\xi_{1000} > 2) = 1 - \Phi(1) = 0.158\,66 .$$

**1.64. Feladat.** *Adjuk meg az előző feladat* $(\xi_n)_{n=N,N+1,\dots}$ *idősorának (elég nagy* $N$ *esetén) spektrális eloszlását!*

**Megoldás:** Írjuk a kovariancia függvényt az időpontok különbségének függvényeként

$$R(n) = \Gamma(n_1, n_1+n) = 2\left(\frac{\sqrt{2}}{2}\right)^{|n|} \qquad n \in \mathbb{Z}\ ,$$

és keressük azt az $f : [-\pi;\pi] \to \mathbb{R}$ páros függvényt mellyel

$$R(n) = \int_{-\pi}^{\pi} f(\omega)\cos(n\omega)d\omega \quad n = 0,1,2,\dots$$

teljesül. Tehát $R(n)$ az $f$ Fourier sorának megfelelő együtthatója, vagyis

$$\begin{aligned}
f(\omega) &= \frac{1}{2\pi}R(0) + \frac{1}{\pi}\sum_{n=1}^{\infty} R(n)\cos(n\omega) = \frac{1}{\pi} + \frac{2}{\pi}\sum_{n=1}^{\infty}\left(\frac{\sqrt{2}}{2}\right)^n \cos(n\omega) = \\
&= \frac{1}{\pi} + \frac{2}{\pi}\operatorname{Re}\left[\sum_{n=1}^{\infty}\left(\frac{\sqrt{2}}{2}\exp(i\omega)\right)^n\right] = \frac{1}{\pi} + \frac{2}{\pi}\operatorname{Re}\left[\frac{\frac{\sqrt{2}}{2}\left(\cos(\omega)+i\sin(\omega)\right)}{1-\frac{\sqrt{2}}{2}\left(\cos(\omega)+i\sin(\omega)\right)}\right] = \\
&= \frac{1}{\pi} + \frac{2}{\pi}\operatorname{Re}\left[-\frac{\sqrt{2}\cos\omega - 1 + i\sqrt{2}\sin\omega}{-3+2\sqrt{2}\cos\omega}\right] = \frac{1}{\pi} - \frac{2}{\pi}\frac{1-\sqrt{2}\cos(\omega)}{3-2\sqrt{2}\cos(\omega)} = \\
&= \frac{1}{\pi\left(3-2\sqrt{2}\cos(\omega)\right)} \qquad \omega \in [-\pi;\pi]
\end{aligned}$$

**1.65. Feladat (*).** *Egy pont a számegyenes "1" pontjából indulva, az alábbi véletlen módon változtatja helyzetét: ha az* $n \in \mathbb{N}$ *helyen van, akkor* $n > 0$ *esetben a következő hely a* $0, 1, \dots, n, n+1$ *helyek valamelyike lesz azonos valószínűséggel, és marad* $0$*, ha* $n = 0$*. Adjuk meg a 0-állapot elérésének várható lépésszámát, és annak valószínűségét, hogy a 0-ra az 1-helyről lép a mozgó pont!*

**Megoldás:** Jelölje $\xi_n$ $n = 0,1,2,\dots$, v.v. a pont helyzetét, $\xi_0 = 1$. Legyen továbbá $\tau$ v.v. a 0-ba érés lépészáma. Ekkor

$$\begin{aligned}
P(\xi_n &= k \mid \xi_{n-1} = l) = \begin{cases} 1 & \text{ha } k = l = 0 \\ \frac{1}{l+2} & \text{ha } k = 0,1,\dots,l+1, l = 1,2,\dots,n \\ 0 & \text{egyébként} \end{cases} \qquad n = 1,2,\dots \\
P(\tau &> n) = \sum_{k=1}^{n+1} P(\xi_n = k) \qquad n = 0,1,2,\dots\ , \\
P(\tau &= n) = P(\xi_n = 0 \wedge \xi_{n-1} > 0) = P(\xi_n = 1) = P(\xi_n = 2) = \\
&= \sum_{k=1}^{n}\frac{1}{k+2}P(\xi_{n-1} = k) \qquad n = 1,2,\dots\ , \\
P(\xi_n &= l) = \sum_{k=l-1}^{n}\frac{1}{k+2}P(\xi_{n-1} = k) > P(\xi_n = l+1) \qquad n = 1,2,\dots \qquad l = 3,4,\dots
\end{aligned}$$

amivel kapjuk:

$$\begin{aligned}
P(\tau &> n) = \sum_{k=1}^{n} \frac{k+1}{k+2} P(\xi_{n-1} = k) \leq \frac{n+1}{n+2} P(\tau > n-1) \leq \\
&\leq \frac{n+1}{n+2}\frac{n}{n+1} P(\tau > n-2) \leq \ldots \leq \frac{1}{n+2} \qquad n = 1, 2, \ldots
\end{aligned}$$

tehát $\lim_{n\to\infty} P(\tau > n) = 0$, továbbá

$$\begin{aligned}
P(\tau &> 0) = 1 \\
P(\tau &> 1) = P\left(\tau > 0\right) - P(\tau = 1) = P\left(\tau > 0\right) - P(\xi_1 = 1) \\
P(\tau &> 2) = P\left(\tau > 1\right) - P(\tau = 2) = P\left(\tau > 1\right) - P(\xi_2 = 2) \\
&\vdots \\
P(\tau &> n) = P\left(\tau > n-1\right) - P(\tau = n) = P\left(\tau > n-1\right) - P(\xi_n = n)
\end{aligned}$$

amit összegezve kapjuk

$$0 = \lim_{n\to\infty} P(\tau > n) = 1 - \sum_{k=1}^{\infty} P(\xi_k = 1) \Rightarrow \sum_{k=0}^{\infty} P(\xi_k = 1) = 2.$$

Mivel $P(\xi_n = l) \geq P(\xi_n = l+1)$, az alábbi sorok konvergensek, és összegüket jelölje:

$$q_l = \sum_{k=l-1}^{\infty} P(\xi_k = l) \qquad l = 1, 2, 3, 4, \ldots \ ,$$

amivel kapjuk:

$$\begin{aligned}
q_l &= \sum_{k=l-1}^{\infty} \sum_{s=k-1}^{k} \frac{1}{s+2} P(\xi_{k-1} = s) = \sum_{s=l-1}^{\infty} \frac{1}{s+2} \sum_{k=s}^{\infty} P(\xi_{k-1} = s) = \\
&= \sum_{s=l-1}^{\infty} \frac{1}{s+2} q_s \qquad l = 2, 3, 4, \ldots \ .
\end{aligned}$$

Tehát

$$\begin{aligned}
q_1 &= 2 \\
q_2 &= 1 \\
q_l &= \frac{1}{l+1} q_{l-1} + \frac{1}{l+2} q_l + \frac{1}{l+3} q_{l+1} \ldots \\
q_{l+1} &= \qquad\qquad + \frac{1}{l+2} q_l + \frac{1}{l+3} q_{l+1} \ldots
\end{aligned}$$

amiből kapjuk a

$$q_l - q_{l+1} = \frac{1}{l+1} q_{l-1} \qquad l = 2, 3, 4, \ldots$$

rekurzív egyenletet, melynek megoldása, figyelembe véve, hogy az első két sorozattag egyértelműen meghatározza a sorozatot:

$$q_l = \frac{2}{l!} \qquad l = 1, 2, 3, 4, \dots \ .$$

$\tau$ várható értékét megkapjuk az alábbi sor összegeként:

$$\begin{aligned} E(\tau) &= \sum_{n=0}^{\infty} P(\tau > n) = \sum_{n=0}^{\infty} \sum_{k=1}^{n+1} P(\xi_n = k) = \\ &= \sum_{k=1}^{\infty} \sum_{n=k-1}^{\infty} P(\xi_n = k) = 2 \sum_{k=0}^{\infty} \frac{1}{l!} - 2 = 2(e-1) \ . \end{aligned}$$

Számítsuk ki most az $A = \cup_{n=1}^{\infty} \left\{ \tau = n \wedge \xi_{n-1} = 1 \right\}$ esemény valószínűségét!

$$\begin{aligned} P(A) &= \sum_{n=1}^{\infty} P(\tau = n \wedge \xi_{n-1} = 1) = \sum_{n=1}^{\infty} P(\xi_n = 0 \wedge \xi_{n-1} = 1) = \\ &= \sum_{n=1}^{\infty} P(\xi_n = 0 \mid \xi_{n-1} = 1) \cdot P(\xi_{n-1} = 1) = \sum_{n=1}^{\infty} \frac{1}{3} P(\xi_{n-1} = 1) = \frac{2}{3}. \end{aligned}$$

**1.66. Feladat (*).** *Legyenek $\Pi_k \quad k = 1, 2, \dots, r \quad d \times d$-típusú sztochasztikus mátrixok, az $A_{nk} \quad k = 1, 2, \dots, r$ teljes eseményrendszerek $n = 1, 2, \dots$ teljesen függetlenek, pozitív valószínűségeik:*

$$0 < p_k = P(A_{nk}) \quad k = 1, 2, \dots, r \quad n = 1, 2, \dots \ .$$

*Ha tetszőleges $v_0 \in \left(\mathbb{R}_0^+\right)^d$ valószínűségeloszlás esetén a*

$$v_n^T = v_{n-1}^T \left( \sum_{k=1}^{r} \mathbf{1}_{A_{nk}} \cdot \Pi_k \right) \qquad n = 1, 2, \dots$$

*d-dimenziós idősor majdnem biztosan konvergál egy $m \in \mathbb{R}^d$ (valószínűségeloszlás-) vektorhoz, akkor*

$$m^T = m^T \cdot \left( \sum_{k=1}^{r} p_k \cdot \Pi_k \right) \tag{1.17}$$

*és teljesül*

$$m^T = m^T \cdot \Pi_k \quad k = 1, 2, \dots, r \ . \tag{1.18}$$

**Bizonyítás.** A feltételes várható érték tulajdonságait, továbbá $v_{n-1}$ és az $A_{nk} \quad k = 1, 2, \dots, r$ eseményrendszer függetlenségét kihasználva kapjuk

$$E(v_n \mid v_{n-1}) = E\left( \sum_{k=1}^{r} \mathbf{1}_{A_{nk}} \cdot \Pi_k^T \right) v_{n-1} = \overline{\Pi}^T v_{n-1},$$

ahol

$$\overline{\Pi} = \sum_{k=1}^{r} p_k \cdot \Pi_k \ ,$$

és ebből

$$E(v_n) = \overline{\Pi}^T E(v_{n-1}) \quad n = 1, 2, \dots \ .$$

Mivel $v_n$ minden komponense nemnegatív, és 1-nél kisebb, a várható érték és a határétérték felcserélésével kapjuk (1.17)-et. Hasonlóan számoljuk továbbá

$$\begin{aligned}
E(v_n v_n^T \mid v_{n-1}) = E\left(\left(\sum_{k=1}^{r} \mathbf{1}_{A_{nk}} \cdot \Pi_k^T\right) v_{n-1} v_{n-1}^T \left(\sum_{l=1}^{r} \mathbf{1}_{A_{nl}} \cdot \Pi_l\right) \mid v_{n-1}\right) = \\
= E\left(\left(\sum_{k=1}^{r} \mathbf{1}_{A_{nk}} \cdot \Pi_k^T \cdot v_{n-1} v_{n-1}^T \cdot \Pi_k\right) \mid v_{n-1}\right) \\
= \sum p_k \cdot \Pi_k^T \cdot v_{n-1} v_{n-1}^T \cdot \Pi_k \ ,
\end{aligned}$$

amiből

$$E(v_n v_n^T) = \sum p_k \cdot \Pi_k^T \cdot E(v_{n-1} v_{n-1}^T) \cdot \Pi_k \qquad n = 1, 2, \dots .$$

Ismét határértéket véve, átrendezés után kapjuk

$$\sum p_k \cdot \Pi_k^T \cdot m \cdot m^T \cdot \Pi_k - m \cdot m^T = \mathbf{0}_{r \times r}$$

Jelölje $q_k^T = m^T \cdot \Pi_k \quad k = 1, 2, \dots, r$, akkor felhasználva, hogy $m = \sum_{k=1}^{r} p_k \cdot q_k$, kapjuk

$$\sum p_k \cdot q_k \cdot q_k^T - m \cdot m^T = \mathbf{0}_{r \times r} \ ,$$

ami azt jelent, hogy a $(q_k; p_k)_{k=1,2,\dots,r}$ diszkrét valószínűségeloszlás várható érték vektora $m$, kovariancia mátrixa a zérusmátrix, tehát az ilyen eloszlású vektor valószínűségi változó 1-valószínűséggel állandó, azaz ha $p_k > 0$ akkor $m = q_k$ teljesül, ami (1.18) teljesülését jelenti. ■

# 2. fejezet

# Matematikai statisztika feladatok

## 2.1. Paraméter becslések

**2.1. Feladat.** *Egy $N = 100$ elemű termék halmazban $M > 0$ ismeretlen számú hibás termék van. Visszatevés nélkül, egyesével válsztva amíg hibásat nem találtunk, $k = 20$ terméket kellett megvizsgálni. Becsüljük $M$ értékét!*

**Megoldás:** Jelölje $\nu$ a minta statisztikát (megfigyelt értéke $k = 20$), $M$ az ismeretlen (meghatározó) paramétert, melynek függvényében a likelihood függvény:

$$L(k; M) = \begin{cases} \frac{M}{N} & k = 1 \\ \frac{(N-M)(N-M-1)\cdots(N-M-k+2)M}{N(N-1)\cdots(N-k+1)} & k = 2, \ldots N - M + 1 \end{cases}$$

vagy

$$L(k; M) = \frac{\binom{N-M}{k-1}}{\binom{N-1}{k-1}} \frac{M}{N} \quad k = 1, 2, \ldots N - M + 1 \ .$$

Keressük ennek maximumát az

$$\begin{aligned} 1 \quad \leq \quad \frac{L(k; M+1)}{L(k; M)} &= \frac{\frac{(N-M-1)(N-M-2)\cdots(N-M-k+1)(M+1)}{N(N-1)\cdots(N-k+1)}}{\frac{(N-M)(N-M-1)\cdots(N-M-k+2)M}{N(N-1)\cdots(N-k+1)}} = \\ = \quad \frac{(N-M-k+1)(M+1)}{(N-M)M} & \quad M = 1, 2, \ldots, N - k + 1 \end{aligned}$$

egyenlőtlenség vizsgálatával, ami ekvivalens a

$$kM \leq N - k + 1 \Leftrightarrow M \leq \frac{N - k + 1}{k}$$

egyenlőtlenséggel, tehát a maximum hely

$$\hat{M}(k) = \begin{cases} N & \text{ha } k = 1 \\ \mathrm{Int}\left(\frac{N+1}{k}\right) & \text{ha } k > 1 \end{cases}$$

ahol $\mathrm{Int}(x)$ jelöli az $x$ szám egész részét. Tehát kaptuk az

$$M \sim \hat{M} = \begin{cases} N & \text{ha } \nu = 1 \\ \mathrm{Int}\left(\frac{N+1}{\nu}\right) & \text{ha } \nu > 1 \end{cases}$$

maximum likelihood becslést, amivel az ismeretlen paraméter becsült értéke:

$$M \approx \hat{M}(20) = \mathrm{Int}\left(\frac{101}{20}\right) = 5\ .$$

**Megjegyzés:** Könnyen ellenőrízhető, hogy ez a becslés torzított, mert pl. $M = 5$ esetén

$$\mathbb{E}_{|M=5}\left(\hat{M}\right) = 100 \cdot \frac{\binom{100-5}{0}}{\binom{99}{0}} \cdot \frac{5}{100} + \sum_{k=2}^{94} \mathrm{Int}\left(\frac{101}{k}\right) \cdot \frac{\binom{100-5}{k-1}}{\binom{99}{k-1}} \cdot \frac{5}{100} = 16.167 \neq 5\ .$$

A torzítás relatív mértéke $M$ nagyobb értékeire csökken, pl.

$$\mathbb{E}_{|M=50}\left(\hat{M}\right) = 69.429 \qquad \mathbb{E}_{|M=80}\left(\hat{M}\right) = 89.295$$

és $M = 100$ esetben nincs torzítás, mert $\mathbb{E}_{|M=100}\left(\hat{M}\right) = 100$.

**2.2. Feladat.** *Egy autóbusz járat egy megállóba ismeretlen, de állandó időközönként érkezik. Véletlen időpontokban érkezve ebbe a megállóba, az alábbi várakozási időtartamokat figyeltük meg.*

1.1 0.6 3.5 2.2 4.1 0.2 4.8 1.8 *perc*

*Becsüljük a járatsűrűséget! Készítsünk maximum likelihood becslést!*

**Megoldás:** Jelölje $\xi = (\xi_1, \ldots, \xi_n)$ a minta statisztikát, $n = 8$, és

$$x = \begin{pmatrix} 1.1 & 0.6 & 3.5 & 2.2 & 4.1 & 0.2 & 4.8 & 1.8 \end{pmatrix}$$

a megfigyelt mintát. Feltételezhetjük, hogy $\xi_k \in \mathcal{U}(0; d) \quad k = 1, 2, \ldots, n$, ahol $d$ az ismeretlen járatsűrűség paraméter. Mivel $\mathbb{E}(\xi_k) = \frac{d}{2}$, használjuk először a várható érték szokásos becslését

$$\frac{d}{2} \approx \bar{x} = 2.2875 \quad \mathbb{D}(\bar{\xi}) = \frac{d}{2\sqrt{3n}} \approx \frac{2.2875}{\sqrt{3 \cdot 8}} = 0.46693$$

amiből a becsült érték és standard hibája

$$d \approx 2\bar{x} = 2 \cdot 2.2875 = 4.575 \quad \mathbb{D}(2\bar{\xi}) = \frac{d}{\sqrt{3n}} \approx \frac{4.575}{\sqrt{3 \cdot 8}} = 0.93387\ .$$

Írjuk a likelihood függvényt a $d$ paraméter függvényeként

$$L(x; d) = \begin{cases} \frac{1}{d^n} & \text{ha } 0 < x_k < d \quad k = 1, 2, \ldots, n \\ 0 & \text{egyébként} \end{cases},$$

aminek maximum helye

$$\hat{d}(x) = \max\{x_1, x_2, \dots, x_n\} = x_n^* \ .$$

Tehát kaptuk a $d \sim \xi_n^*$ maximum likelihood becslést. Vizsgáljuk a becslés torzítatlanságát, amihez szükséges $\xi_n^*$ eloszlásának megadása.

$$F_{\xi_n^*}(x) = \mathbb{P}(\xi_1 < x, \xi_2 < x, \cdots, \xi_n < x) = \begin{cases} \frac{x^n}{d^n} & \text{ha } 0 < x < d \\ 0 & \text{egyébként} \end{cases}$$

$$f_{\xi_n^*}(x) = \frac{n}{d^n} x^{n-1} \quad 0 < x < d \ .$$

Így

$$\mathbb{E}(\xi_n^*) = \int_0^d x \frac{n}{d^n} x^{n-1} dx = \frac{n}{1+n} \cdot d$$

vagyis a maximum likelihood becslés torzított, de korrigálva kapjuk a $d \sim \frac{n+1}{n}\xi_n^*$ torzítatlan becslést. Ennek standard hibája

$$\mathbb{D}\left(\frac{n+1}{n}\xi_n^*\right) = \frac{n+1}{n} \cdot \mathbb{D}(\xi_n^*) = d \cdot \frac{n+1}{n} \cdot \sqrt{\frac{n}{2+n} - \frac{n^2}{(1+n)^2}} = \frac{d}{\sqrt{n(2+n)}}$$

mivel

$$\mathbb{E}(\xi_n^{*2}) = \int_0^d x^2 \frac{n}{d^n} x^{n-1} dx = \frac{n}{2+n} \cdot d^2 \quad \mathbb{D}(\xi_n^*) = d\sqrt{\frac{n}{2+n} - \frac{n^2}{(1+n)^2}} \ .$$

Tehát az így korrigált becsléssel kapott becsült érték és standard hiba:

$$d \approx \frac{9}{8} \cdot 4.8 = 5.4 \quad \mathbb{D}\left(\frac{n+1}{n}\xi_n^*\right) = \frac{d}{\sqrt{n(2+n)}} \approx \frac{5.4}{\sqrt{8 \cdot 10}} = 0.603\,74 \ .$$

**Megjegyzés:** Vegyük észre, hogy a maximum likelihood becslésből nyert utóbbi eredmény standard hibája lényegesen kisebb ($\sqrt{n}$ nagyságrenddel jobb). A Fisher féle információs mennyiséggel most nem adható alsó korlát torzítatlan becslés standard hibájára, mert a likelihood függvény a $d$ paramétertől függő $[0; d]$ intervallumon pozitív.

**2.3. Feladat.** *Legyen a s.m. egy*

$$f(x) = \alpha \cdot x^{\alpha-1} \qquad 0 < x < 1$$

*sűrűségfüggvényű v.v. $n$-ismételt megfigyelésével kapcsolatos, ahol $\alpha > 0$ ismeretlen paraméter. Keressünk elégséges statisztikát, adjunk maximum likelihood becslést a $\vartheta = \frac{1}{\alpha}$ paraméterre, vizsgáljuk meg a becslés tulajdonságait!*

**Megoldás:** Jelölje $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ a minta statisztikát, $\alpha$ az eloszlást meghatározó paramétert. Írjuk a likelihodd függvényt az $\alpha$ paraméter függvényeként:

$$L(x;\alpha) = \alpha^n \cdot \prod_{k=1}^{n} x_k^{\alpha-1} \quad x = (x_1, x_2, \ldots, x_n) \in ]0;1[^n$$

Ekkor a

$$x \mapsto \prod_{k=1}^{n} x_k \quad x = (x_1, x_2, \ldots, x_n) \in ]0;1[^n$$

statisztika elégséges, és kölcsönösen egyértelmű függvénye

$$T(x) = \sum_{k=1}^{n} \ln(x_k) \quad x = (x_1, x_2, \ldots, x_n) \in ]0;1[^n$$

is elégséges statisztika. Keressük először $\alpha$ maximum likelihood becslését, mint

$$\ln \circ L(x;\alpha) = n\ln(\alpha) + (\alpha - 1)T(x)$$

maximum helyét egy $x = (x_1, x_2, \ldots, x_n) \in ]0;1[^n$ megfigyelt minta esetén, ami az

$$S(x;\alpha) = \frac{\partial}{\partial\alpha} \ln \circ L(x;\alpha) = \frac{n}{\alpha} + T(x)$$

zérushelyéből

$$\hat{\alpha}(x) = -\frac{n}{T(x)} \ ,$$

tehát kaptuk az $\alpha \sim \hat{\alpha} = \frac{n}{T}$ becslést, és így a

$$\vartheta \sim \frac{T}{n}$$

maximum likelihood becslést.

Mivel

$$\begin{aligned}
\mathbb{E}\left(-\ln(\xi_k)\right) &= \int_0^1 -\ln(x) \cdot \alpha x^{\alpha-1} dx = \frac{1}{\alpha} = \vartheta \\
\mathbb{E}\left(\ln^2(\xi_k)\right) &= \int_0^1 \ln^2(x) \cdot \alpha x^{\alpha-1} dx = \frac{2}{\alpha^2} \quad \mathbb{D}\left(-\ln(\xi_k)\right) = \sqrt{\frac{2}{\alpha^2} - \left(\frac{1}{\alpha}\right)^2} = \frac{1}{\alpha}
\end{aligned}$$

kapjuk

$$\mathbb{E}\left(\frac{T}{n}\right) = \frac{1}{\alpha} = \vartheta \ ,$$

tehát a becslés torzítatlan, és standard hibája

$$\mathbb{D}\left(\frac{T}{n}\right) = \frac{1}{\alpha\sqrt{n}} \to 0$$

ezért a becslés konzisztens.

Adjuk meg a $\vartheta$ paraméter torzítatlan becslése hatékonyságának alsó határát az

$$I(\alpha) = \mathbb{D}^2\left(S(\xi;\alpha)\right) = -\mathbb{E}\left(\frac{\partial}{\partial\alpha}S(\xi;\alpha)\right) = \mathbb{E}\left(\frac{n}{\alpha^2}\right) = \frac{n}{\alpha^2}$$

Fisher féle információs mennyiséggel. Ha $\vartheta = \frac{1}{\alpha} \sim t$ torzítatlan becslés, akkor

$$\mathbb{D}^2(t) \geq \frac{\left(-\frac{1}{\alpha^2}\right)^2}{\frac{n}{\alpha^2}} = \frac{1}{\alpha^2 n} = \mathbb{D}^2\left(\frac{T}{n}\right) ,$$

tehát a kapott maximum likelihood becslés a leghatékonyabb. Ez egyébként a

$$\frac{T}{n} = S(\xi;\alpha) - \frac{n}{\alpha}$$

összefüggésből is következik.

**2.4. Feladat.** *Egy folytonos eloszlású v.v. sűrűségfüggvénye:*

$$f(x) = 2\alpha x \cdot e^{-\alpha x^2} \quad \text{ha } 0 < x ,$$

*ahol $\alpha > 0$ ismeretlen paraméter. $n$-ismételt megfigyelés esetén adjuk meg a likelihood függvényt, keressünk elégséges statisztikát, adjunk maximum likelihood becslést az $\alpha$ paraméterre és a várható értékre! Mit mondhatunk a becslésekről? Ha egy megfigyelt minta*

$$x = (16.24, 4.19, 4.77, 6.38, 4.14, 0.54, 12.87, 7.35, 9.28, 8.11),$$

*adjuk meg a becsült értékeket, és becsüljük a standard hibákat!*

**Megoldás:** Jelölje $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ a minta statisztikát, $\alpha$ az eloszlást meghatározó paramétert. Írjuk a likelihood függvényt az $\alpha$ paraméter függvényeként:

$$L(x;\alpha) = 2^n\alpha^n \prod_{k=1}^{n} x_k \cdot e^{-\alpha\sum_{k=1}^{n} x_k^2} \quad x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^{+n} ,$$

tehát

$$T : \mathbb{R}^n \to \mathbb{R} \quad x \mapsto \sum_{k=1}^{n} x_k^2$$

elégséges statisztika. Keressük

$$\ln \circ L(x;\alpha) = n\ln(2) + n\ln(\alpha) + \sum_{k=1}^{n}\ln(x_k) - \alpha T(x)$$

maximum helyét, ami az

$$S(x;\alpha) = \frac{\partial}{\partial\alpha}\ln \circ L(x;\alpha) = \frac{n}{\alpha} - T(x) = 0$$

egyenlet

$$\hat{\alpha} = \frac{n}{T(x)}$$

megoldása. Tehát kaptuk az

$$\alpha \sim \frac{n}{T}$$

maximum likelihood becslést. A várható értéket adjuk meg először az $\alpha$ paraméter függvényeként:

$$\begin{aligned} m &= \int_0^\infty x \cdot 2\alpha x \cdot e^{-\alpha x^2} dx = \left[x \cdot \left(-e^{-\alpha x^2}\right)\right]_0^\infty + \int_0^\infty e^{-\alpha x^2} dx = \\ &= \sqrt{\frac{\pi}{\alpha}} \int_0^\infty \frac{1}{\sqrt{2\pi} \cdot \frac{1}{\sqrt{2\alpha}}} \exp\left(-\frac{x^2}{2\left(\frac{1}{\sqrt{2\alpha}}\right)^2}\right) dx = \frac{1}{2}\sqrt{\frac{\pi}{\alpha}} \end{aligned}$$

Tehát a várható érték maximum likelihood becslése

$$m \sim \frac{1}{2}\sqrt{\frac{\pi}{n} T} \; .$$

A standard hibák közelítéséhez vizsgáljuk a $T$ elégséges statisztika várható értékét és szórását. Mivel

$$\mathbb{P}(\xi_k^2 < t) = \mathbb{P}(\xi_k < \sqrt{t}) = \int_0^{\sqrt{t}} 2\alpha x \cdot e^{-\alpha x^2} dx = 1 - e^{-\alpha t} \quad 0 < t$$

a minta elem statisztikák négyzetei független exponenciális eloszlásúak $\alpha$ paraméterrel, tehát

$$\mathbb{E}(T) = n\frac{1}{\alpha} \quad \mathbb{D}(T) = \sqrt{n}\frac{1}{\alpha} \; .$$

Használjuk ezt a

$$\begin{aligned} \frac{n}{T} &\approx \frac{n}{n\frac{1}{\alpha}} - \frac{n}{\left(n\frac{1}{\alpha}\right)^2}\left(T - n\frac{1}{\alpha}\right) = \alpha - \frac{\alpha^2}{n}\left(T - n\frac{1}{\alpha}\right) \\ \frac{1}{2}\sqrt{\frac{\pi}{n}T} &\approx \frac{1}{2}\sqrt{\frac{\pi}{n} n \frac{1}{\alpha}} + \frac{1}{2}\sqrt{\frac{\pi}{n}} \frac{1}{2\sqrt{n\frac{1}{\alpha}}}\left(T - n\frac{1}{\alpha}\right) = m + \frac{\sqrt{\pi\alpha}}{4n}\left(T - n\frac{1}{\alpha}\right) \end{aligned}$$

elsőfokú Taylor formulákban, és így a becsléseink közelítően torzítatlanok

$$\mathbb{D}\left(\frac{n}{T}\right) \approx \frac{\alpha^2}{n}\sqrt{n}\frac{1}{\alpha} = \frac{\alpha}{\sqrt{n}} \qquad \mathbb{D}\left(\frac{1}{2}\sqrt{\frac{\pi}{n}T}\right) \approx \frac{\sqrt{\pi\alpha}}{4n}\sqrt{n}\frac{1}{\alpha} = \frac{1}{4}\sqrt{\frac{\pi}{\alpha n}} = \frac{m}{2\sqrt{n}}$$

standard hibákkal, melyek $n \to \infty$ esetben nullához tartanak, tehát a becslések konzisztensek.

Adjuk meg a paraméterek torzítatlan becslései hatékonyságának alsó határát az

$$I(\alpha) = \mathbb{D}^2\left(S(\xi;\alpha)\right) = -\mathbb{E}\left(\frac{\partial}{\partial\alpha}S(\xi;\alpha)\right) = \mathbb{E}\left(\frac{n}{\alpha^2}\right) = \frac{n}{\alpha^2}$$

Fisher féle információs mennyiséggel. Ha $\alpha \sim t_1$ $\quad m = \frac{1}{2}\sqrt{\frac{\pi}{\alpha}} \sim t_2$ torzítatlan becslések, akkor

$$\mathbb{D}^2(t_1) \geq \frac{1}{\frac{n}{\alpha^2}} = \frac{\alpha^2}{n} \approx \mathbb{D}^2\left(\frac{n}{T}\right) \qquad \mathbb{D}^2(t_2) \geq \frac{\left(-\frac{1}{4}\frac{\sqrt{\pi}}{\sqrt{\alpha^3}}\right)^2}{\frac{n}{\alpha^2}} = \frac{1}{16}\frac{\pi}{\alpha n} \approx \mathbb{D}^2\left(\frac{1}{2}\sqrt{\frac{\pi}{n}T}\right)$$

mivel $\frac{\partial}{\partial\alpha}\frac{1}{2}\sqrt{\frac{\pi}{\alpha}} = -\frac{1}{4}\frac{\sqrt{\pi}}{\sqrt{\alpha^3}}$. Tehát a kapott becslések közelítően (elég nagy $n$ esetén) a leghatékonyabbak.

Az $n = 10$ elemű mintából

$$T(x) = 16.24^2 + 4.19^2 + 4.77^2 + 6.38^2 + 4.14^2 + 0.54^2 + 12.87^2 + \\ + 7.35^2 + 9.28^2 + 8.11^2 = 733.73$$

amivel kapjuk:

$$\begin{aligned} \alpha &\approx \frac{n}{T(x)} = \frac{10}{733.73} = 0.013629 \quad \text{s.h.: } \frac{\alpha}{\sqrt{n}} \approx \frac{0.013629}{\sqrt{10}} = 4.3099 \times 10^{-3} \\ m &\approx \frac{1}{2}\sqrt{\frac{\pi}{n}T(x)} = \frac{1}{2}\sqrt{\frac{\pi}{10}733.73} = 7.5912 \quad \text{s.h.: } \frac{m}{2\sqrt{n}} \approx \frac{7.5912}{2\sqrt{10}} = 1.2003 \end{aligned}$$

**2.5. Feladat.** *Egy 500 darabos termékhalmazban ismeretlen számú hibás termék van. Becsüljük a hibásak számát, ha 100 elemű mintát véve, 25 hibásat találtunk! Adjunk 90% biztonsággal felső korlátot a hibásak számára!*

**Megoldás:** Jelölje $\nu$ a minta statisztikát, akkor $\nu \in \mathcal{H}yp(500; M; 100)$, ahol $M$ az ismeretlen paraméter. A valószínűség becsült értéke, és a becslés standard hibája:

$$\frac{M}{500} \approx \frac{25}{100} = 0.25 \quad \mathbb{D}\left(\frac{\nu}{100}\right) \approx \frac{1}{\sqrt{100}} \cdot \sqrt{0.25 \cdot 0.75 \cdot \left[1 - \frac{99}{499}\right]} = 3.8769 \times 10^{-2} \ ,$$

tehát $M$ becsült értéke, és a becslés standard hibája

$$M \approx 125 \quad \mathbb{D}\left(500 \cdot \frac{\nu}{100}\right) \approx 500 \cdot 3.8769 \times 10^{-2} = 19.385 \ .$$

Táblázatból $u_{0.2} = 1.282$, amivel a valószínűség 80%-os szintű kétoldali határai

$$\frac{M}{500} \approx \frac{25}{100} \pm 1.282 \cdot 3.8769 \times 10^{-2}$$

tehát 90%-os szint mellett

$$M \leq 125 + 1.282 \cdot 19.385 = 149.852 \approx 150 \ .$$

**2.6. Feladat.** *Hány embert kell egy közvélemény kutató cégnek megkérdeznie, hogy egy ismeretlen arányt 0.01 pontossággal tudjanak megadni 95%-os szint mellett?*

**Megoldás:** Jelölje $\nu$ a minta statisztikát, akkor (a feltehetően nagy létszámú alapsokaság miatt) $\nu \in \mathcal{B}in(n;p)$, ahol $p$ az ismeretlen arány (valószínűség) paraméter. A 95%-os kétoldali intervallum becslés pontossága

$$u_{0.05} \cdot \frac{1}{\sqrt{n}} \cdot \sqrt{\frac{\nu}{n}\left(1-\frac{\nu}{n}\right)} \leq u_{0.05} \cdot \frac{1}{2\sqrt{n}} \; .$$

Mivel $u_{0.05} = 1.96$, kapjuk

$$1.96 \cdot \frac{1}{2\sqrt{n}} \leq 0.01 \Longrightarrow 9604 \leq n \; .$$

**2.7. Feladat.** *Egy mérési eljárás legnagyobb véletlen hibája $\Delta = \pm 0.03$, $n = 3$ mérés eredményéből $\bar{x} = 12.13$.*

**a)** *Becsüljük 90%-os biztonsággal a mért mennyiség értékét!*

**b)** *Legfeljebb mennyi lehet (ugyanezen mennyiség esetén) egy mérés eredménye 95%-os biztonsággal?*

**c)** *Hány méresre van szükség, hogy a 95%-os szintű intervallum becslés pontossága* 0.005 *legyen?*

**Megoldás:** Feltehetjük, hogy a mérések eredménye $\mathcal{N}(m, \sigma_0)$ eloszlású, ahol $m$ az ismeretlen várható érték paraméter, $\sigma_0 = \frac{|\Delta|}{3} = 0.01$ az ismert szórás, tehát

$$m \approx \bar{x} = 12.13 \qquad \mathbb{D}(\bar{\xi}) = \frac{\sigma_0}{\sqrt{n}} = \frac{0.01}{\sqrt{3}} = 5.7735 \times 10^{-3} \; .$$

**a)** Táblázatból $u_{0.1} = 1.645$, amivel kapjuk a 90%-os határokat:

$$m \approx 12.13 \pm 1.645 \cdot 5.7735 \times 10^{-3} = \begin{cases} 12.139 \\ 12.121 \end{cases} .$$

**b)** Jelölje $\eta_1$ egy további mérés eredményét, akkor $u_{0.1} = 1.645$ táblázati értékkel kapjuk a 90%-os

$$\bar{\eta} \approx 12.13 \pm 1.645 \cdot 0.01 \cdot \sqrt{\frac{1}{3} + \frac{1}{1}}$$

kétoldali határokat, amiből

$$\eta_1 \leq 12.13 + 1.645 \cdot 0.01 \cdot \sqrt{\frac{1}{3} + \frac{1}{1}} = 12.149$$

a keresett 95%-os felső határ.

**c)** $u_{0.05} = 1.96$ amivel a kétoldali intervallum becslés pontosságára

$$1.96 \cdot \frac{0.01}{\sqrt{n}} \leq 0.005 \Rightarrow 15.366 \leq n \ ,$$

tehát $n = 16$ mérés szükséges.

**2.8. Feladat.** *$n = 40$ felnőtt férfi súlyát megmérve, kaptuk $\bar{x} = 78.25$ kg $\quad s^*(x) = 12.01$ kg.*

***a)*** *Adjuk meg az átlagos (várható) súly 90%-os határait!*

***b)*** *Legalább hány kg lesz egy gépkocsi terhelése, ha 5 férfi foglal benne helyet, 90%-os szint mellett?*

***c)*** *Milyen határok között van a súly szórása 90%-os szint mellett?*

**Megoldás:** Feltehetjük a súly adatok $\mathcal{N}(m, \sigma)$ eloszlását (már a minta mérete miatt is), ahol $m$ a várható érték, $\sigma$ a szórás paraméter.

**a)** Az átlagsúly, azaz várható érték határai a $t_{0.1} = 1.6849$ táblázati értékkel (szabadsági fok: 39)

$$m \approx 78.25 \pm 1.6849 \frac{12.01}{\sqrt{40}} \begin{cases} 81.45 \\ 75.05 \end{cases} .$$

**b)** Jelölje $\eta_1, \eta_2, \eta_3, \eta_4$ a további (gépkocsiba szálló) férfiak súlyát, akkor $t_{0.2} = 1.3036$ táblázati értékkel (szabadsági fok: 39) a 80%-os kétoldali határok

$$\bar{\eta} \approx 78.25 \pm 1.3036 \cdot 12.01 \sqrt{\frac{1}{40} + \frac{1}{4}}$$

amiből kapjuk a 90%-os alsó határt az összegre:

$$\eta_1 + \eta_2 + \eta_3 + \eta_4 = 4 \cdot \bar{\eta} \geq 4 \cdot \left( 78.25 - 1.3036 \cdot 12.01 \sqrt{\frac{1}{40} + \frac{1}{4}} \right) = 280.16 \text{ kg.}$$

**c)** A $\chi^2_{39}$ eloszlás táblázatából $\chi^2_{0.95} = 25.695 \quad \chi^2_{0.05} = 54.572$ amivel kapjuk a szórás 90%-os konfidencia határait:

$$\sigma \approx \left( 12.01 \cdot \sqrt{\frac{39}{54.572}}; 12.01 \cdot \sqrt{\frac{39}{25.695}} \right) = (10.153; 14.796) \ .$$

**2.9. Feladat (*).** *Egy exponenciális eloszlású véletlen mennyiség ismeretlen $n$-számú megfigyeléséből csak egy adott $T$ időtartamnál kisebb értékekből becsüljük az exponenciális eloszlás paraméterét, és az ismeretlen $n$ értékét!*

**Megoldás:** Legyenek $\xi_1, \xi_2, \ldots, \xi_n$ független, azonos folytonos eloszlású v.v.-k! Jelölje a rendezett minta elemeit

$$\tau_1 < \tau_2 < \ldots < \tau_n \; .$$

Ha a megfigyelés $[0;T]$ időintervallumába $\nu = r$ számú megfigyelt érték esik, a likelihood függvény az ismeretlen $f$ sűrűségfüggvénnyel:

$$\begin{aligned} L(r, x_1, x_2, \ldots; n, f) &= \frac{n!}{(n-r)!} \cdot f(x_1) \cdot f(x_2) \cdot \ldots \cdot f(x_r) \cdot (1 - F(T))^{n-r} \qquad (*) \\ \text{ha} \quad x_1 &< x_2 < \ldots < x_r \in \mathbb{R}, \quad r = 0, 1, \ldots, n \end{aligned}$$

ahol $F$ az eloszlásfüggvényt jelöli.
Legyen a továbbiakban

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

keressük (*) maximumát, azaz

$$L(r, x_1, x_2, \ldots; n, \lambda) = \frac{n!}{(n-r)!} \cdot \lambda^r \cdot \exp\left(-\lambda \sum_{k=1}^{r} x_k - \lambda(n-r)T\right)$$

az ismeretlen $\lambda$, vagy még inkább $m = \frac{1}{\lambda}$ illetve $n$ szerint! Mivel minden $n$-re (*) pontosan akkor maximális, ha

$$\hat{\lambda} = \frac{r}{\sum_{k=1}^{r} x_k + (n-r)T} = \frac{r}{r \cdot \overline{x} + (n-r)T}$$

illetve

$$\hat{m} = \bar{x} + (\frac{n}{r} - 1)T$$

ahol

$$\bar{x} = \frac{1}{r} \sum_{k=1}^{r} x_k .$$

Ekkor a maximum értéke

$$\hat{L}_n = \left(\hat{\lambda}\right)^r \cdot \frac{n!}{(n-r)!} \cdot e^{-r} \quad n = r, r+1, \ldots \quad .$$

Ennek a pozitív sorozatnak $n$ szerinti maximumát a

$$q(n) = \frac{\hat{L}_{n+1}}{\hat{L}_n} = \left(\frac{r \cdot \bar{x} + (n-r)T}{r \cdot \bar{x} + (n+1-r)T}\right)^r \frac{n+1}{n+1-r} \quad n = r, r+1, \ldots$$

hányados vizsgálatával keressük. A $c = \frac{\bar{x}}{T}$ és $N = n+1$ jelöléssel vizsgáljuk a

$$Q(N) = \ln(q(n)) = r \ln\left(1 - \frac{1}{r \cdot c + N - r}\right) - \ln\left(1 - \frac{r}{N}\right) \quad N \geq r+1$$

függvény előjelét. Mivel $\lim_{N\to\infty} Q(N) = 0$, és az

$$\frac{\partial}{\partial N}Q(N) = r\left[\frac{1}{(N-r(1-c))^2-(N-r(1-c))} - \frac{1}{N^2-rN}\right] > 0$$

egyenlőtlenség $0 < 1-c < 1$ miatt ekvivalens módon alakítható

$$\begin{aligned} N^2 - rN &> (N-r(1-c))^2 - (N-r(1-c)) \\ N\left(2r(1-c)-r+1\right) &> r^2(1-c)^2 + r(1-c) \end{aligned}$$

tehát a derivált legfeljebb egy helyen válthat előjelet, ezért az alábbi esetek lehetségesek:

1. $2r(1-c)-r+1 \le 0 \Leftrightarrow 1-c \le \frac{r-1}{2r} \Leftrightarrow \frac{r+1}{2r} \le \frac{\overline{x}}{T} < 1$
   Tehát $\frac{\partial}{\partial N}Q(N) \le 0$, vagyis $Q$ monoton csökkenve tart 0-hoz, azaz $Q > 0$, vagyis a likelihood függvény maximuma

$$\widehat{L}_{\max} = \lim_{n\to\infty} \widehat{L}_n = \left(\frac{r}{e\cdot T}\right)^r$$

2. $2r(1-c)-r+1 > 0 \Leftrightarrow \frac{r+1}{2r} > \frac{\overline{x}}{T} > 0$
   Tehát elég nagy $N$ esetén $\frac{\partial}{\partial N}Q(N) > 0$, vagyis $Q$ monoton növekedve tart 0-hoz, azaz $Q < 0$, vagyis a likelihood függvény maximum helye, illetve m.l. becslések

   (a) ha $Q(r+1) < 0$ akkor

$$\widehat{n} = r \qquad \widehat{\lambda} = \frac{1}{\overline{x}}$$

   és ekkor

$$\widehat{L}_{\max} = \left(\frac{1}{e\cdot\overline{x}}\right)^r$$

   (b) ha $Q(r+1) \ge 0$ akkor

$$\widehat{n} > r \qquad \widehat{\lambda} = \frac{r}{r\cdot\overline{x} + (n-r)T}$$

   amire $Q(\widehat{n}-1) \ge 0$ és $Q(\widehat{n}) < 0$, és ekkor

$$\widehat{L}_{\max} = \left(\frac{\widehat{\lambda}}{e}\right)^r \cdot \frac{\widehat{n}!}{(\widehat{n}-r)!}$$

**Megjegyzések:**

1. Az 1. esetben a maximum likelihood becslések a formális

$$\widehat{n} = \infty \qquad \widehat{\lambda} = 0$$

   értékeket adják.

2. A korlátos maximum hely léte (2. eset) azon múlik, hogy teljesül-e, a $r+1 > \frac{2}{T}\sum_{i=1}^{r} x_i$ feltétel. Erre akkor "számíthatunk", ha a várható értékekre teljesül

$$\begin{aligned} E(\nu+1) &= n(1-e^{-\lambda T})+1 > E\left(\frac{2}{T}\sum_{i=1}^{n}\varepsilon_i \cdot \mathbf{1}_{\{\varepsilon_i<T\}}\right) = \\ &= n\frac{2}{T}\int_0^T t\lambda e^{-\lambda t}dt = 2n\cdot\left(\frac{1-e^{-\lambda T}}{\lambda T} - e^{-\lambda T}\right) \end{aligned} \tag{2.1}$$

ahol

$$\begin{aligned} \nu &= \sum_{i=1}^{n} \mathbf{1}_{\{\varepsilon_i<T\}} \\ \varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n &\in \mathcal{E}xp(\lambda) \text{ függetlenek.} \end{aligned}$$

Alakítsuk az (2.1) egyenlőtlenséget,

$$1-e^{-\lambda T} + \frac{1}{n} > 2\left(\frac{1-e^{-\lambda T}}{\lambda T} - e^{-\lambda T}\right)$$

ami teljesül, ha

$$1-e^{-\lambda T} \geq 2\left(\frac{1-e^{-\lambda T}}{\lambda T} - e^{-\lambda T}\right) \Leftrightarrow \lambda T e^{\lambda T} + \lambda T - 2e^{\lambda T} + 2 \geq 0$$

Ez pedig mindíg igaz, mivel $0 \leq x$ esetén

$$\begin{aligned} xe^x + x - 2(e^x-1) &= x + x^2 + \frac{x^3}{2!} + \ldots + x - 2\left(x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots\right) = \\ &= \left(\frac{x^3}{2!} - 2\frac{x^3}{3!}\right) + \left(\frac{x^4}{3!} - 2\frac{x^4}{4!}\right) + \ldots \geq 0 \end{aligned}$$

Tehát elég nagy $n$ esetén a 2. eset bekövetkezésére számíthatunk.

3. Legyen a továbbiakban az egyszerűség kedvéért $T=1$. Feltételezve tehát a 2. eset bekövetkezését, a kapott $\widehat{n}$, és $\widehat{\lambda}$ becsült értékek (közelítő) megoldásai az

$$\begin{aligned} S(r,x;n,\lambda) &= \left[\ \frac{\partial}{\partial n}\ln\circ L \quad \frac{\partial}{\partial\lambda}\ln\circ L\ \right]^T = \\ &= \left[\ -\lambda + \sum_{i=0}^{r-1}\frac{1}{n-i} \quad \frac{r}{\lambda} - (n - r + \sum_{i=1}^{r} x_i)\ \right]^T = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{aligned}$$

egyenletnek. Továbbá

$$B(r,x;n,\lambda) = \begin{bmatrix} \frac{\partial^2}{\partial n^2}\ln\circ L & \frac{\partial^2}{\partial n\partial\lambda}\ln\circ L \\ \frac{\partial^2}{\partial n\partial\lambda}\ln\circ L & \frac{\partial^2}{\partial\lambda^2}\ln\circ L \end{bmatrix} = -\begin{bmatrix} \sum_{i=0}^{r-1}\frac{1}{(n-i)^2} & 1 \\ 1 & \frac{r}{\lambda^2} \end{bmatrix},$$

amiből

$$-B^{-1}(r,x;n,\lambda) = \frac{1}{\frac{r}{\lambda^2}\sum_{i=0}^{r-1}\frac{1}{(n-i)^2} - 1}\begin{bmatrix} \frac{r}{\lambda^2} & -1 \\ -1 & \sum_{i=0}^{r-1}\frac{1}{(n-i)^2}\end{bmatrix}$$

és $-\widehat{\lambda} + \sum_{i=0}^{r-1}\frac{1}{\widehat{n}-i} = 0$ miatt

$$\frac{r}{\widehat{\lambda}^2}\sum_{i=0}^{r-1}\frac{1}{(\widehat{n}-i)^2} - 1 = \frac{r^2}{\widehat{\lambda}^2}\left[\frac{1}{r}\sum_{i=0}^{r-1}\frac{1}{(\widehat{n}-i)^2} - \left(\frac{1}{r}\sum_{i=0}^{r-1}\frac{1}{\widehat{n}-i}\right)^2\right] > 0 \ .$$

Tehát a $-B^{-1}(r,x;\widehat{n},\widehat{\lambda})$ pozitív definit mátrix a $\left(\widehat{n},\widehat{\lambda}\right)$ becslő statisztika pár kovariancia mátrixa becslésére használható, vagyis az $n \approx \widehat{n}$ becslés standard hibájának becsült értéke

$$\mathbb{D}(\widehat{n}) \approx \frac{1}{\sqrt{\sum_{i=0}^{r-1}\frac{1}{(\widehat{n}-i)^2} - \frac{\widehat{\lambda}^2}{r}}}$$

és a $\lambda \approx \widehat{\lambda}$ becslés standard hibájának becslése

$$\mathbb{D}(\widehat{\lambda}) \approx \frac{\widehat{\lambda}}{\sqrt{r}} \cdot \sqrt{\frac{\sum_{i=0}^{r-1}\frac{1}{(\widehat{n}-i)^2}}{\sum_{i=0}^{r-1}\frac{1}{(\widehat{n}-i)^2} - \frac{\widehat{\lambda}^2}{r}}} \ .$$

## 2.2. Paraméteres próbák

**2.10. Feladat.** *Egy mérési eljárás hibája $\mathcal{N}(0;0.02)$ eloszlású véletlen mennyiség. Egy termék egyik fontos jellemzőjét mérjük ezzel az eljárással, és azt jónak minősítjük, ha értéke $m_0 = 20.00$. Tervezzünk próbát ennek ellenőrzésére úgy, hogy a jó terméket csak 0.01 valószínűséggel minősítsük hibásnak, és a 20.05 vagy annál nagyobb, illetve 19.97 vagy annál kisebb jellemzőjű terméket legfeljebb 0.02 valószínűséggel minősítsük jónak!*

**Megoldás:** $u$-próbával ellenőrízzük a $H_0 : m = 20.00$ hipotézist a $H_1 : m \neq 20.00$ kétoldali alternativ hipotézissel szemben, és keressük $n$ értékét, amivel teljesülnek:

$$\alpha = 0.01 \quad \beta\Big|_{m\geq 20.05} \leq 0.02 \quad \beta\Big|_{m\leq 19.97} \leq 0.02 \ .$$

Mivel a másodfajú hiba szimmetrikus az $m = 20.00$ értékre, és az alternatíva monoton csökkenő függvénye mindkét irányban, elég az

$$\alpha = 0.01 \quad \beta\Big|_{m=19.97} \leq 0.02$$

feltételeket teljesíteni. Tehát $u_{0.01} = 2.575\,8$ és így a másodfajú hiba

$$\beta\Big|_{m=19.97} = \Phi\left(2.5758 + \frac{20.00-19.97}{0.02}\sqrt{n}\right) - \Phi\left(-2.5758 + \frac{20.00-19.97}{0.02}\sqrt{n}\right) \leq 0.02 \ ,$$

amit alakítva kpajuk

$$\Phi\left(2.5758+1.5\sqrt{n}\right)-\Phi\left(-2.5758+1.5\sqrt{n}\right)\leq 0.02$$

ahol a bal oldali első tag értéke közelítően már 1-nek vehető, amit rendezve:

$$\begin{aligned}
0.98 &\leq \Phi\left(-2.5758+1.5\sqrt{n}\right)\\
2.0537 &\leq -2.5758+1.5\sqrt{n}\\
9.5255 &= \left(\frac{2.0537+2.5758}{1.5}\right)^2 \leq n
\end{aligned}$$

Tehát $n=10$ mérésre van szükség.

**2.11. Feladat.** *Egy tantárgy (maximum 100 pontos) dolgozatainak sok éves átlageredménye 58.26 pont. Egy 20 fős tanulócsoport tagjai a szokásos felkészítés mellett kidolgozott feladatokat is kaptak a sikeres felkészüléshez. Az így megírt dolgozatok átlageredménye 62.15 pont, és az eredmények korrigált empírikus szórása 4.56. Egy másik, 15 fős csoport tagjai pedig képletgyüjteményt használhattak a dolgozat írásakor, és eredményeik átlaga 61.01, korrigált empírikus szórása 6.69 pont lett. Van-e kimutatható javulás valamelyik esetben? Ha következtetésünk nemleges, mennyi annak kockázata, hogy az adott módon segített csoport átlageredménye mégis 62 pontra nőtt? Tételezzük fel az eredmények normális eloszlását!*

**Megoldás:** $t$-próbával vizsgáljuk a $H_0 : m = 58.26$ hipotézist a $H_1^+ : m > 58.26$ alternatívával szemben, mivel most a másik oldali alternatíva kizárható.

A próba statisztika értéke a 20 fős csoport esetén:

$$t=\frac{62.15-58.26}{4.56}\sqrt{20}=3.815$$

ami a $t_{2\alpha=0.01}=2.861$ (szabadsági fok: 19) táblázati értékkel

$$3.815>2.861 \Rightarrow H_0 \text{ -t elutasítjuk,}$$

tehát van javulás az eredményekben, és az így vállalt elsőfajú hiba értéke kevesebb mint $\alpha=0.005$, pontosan

$$1-F_{T_{19}}(3.815)=5.8453\times 10^{-4}\ ,$$

ahol $F_{T_{19}}$ jelöli a $T_{19}$ eloszlás eloszlásfüggvényét (értékének számításához használjunk MATLAB, MAPLE, SPSS, R, stb. programcsomagot).

A próba statisztika értéke a 15 fős csoport esetén

$$t=\frac{61.01-58.26}{6.69}\sqrt{15}=1.592$$

ami a $t_{2\alpha=0.1}=1.761$ (szabadsági fok: 14) táblázati értékkel

$$1.592 \not> 1.761 \Rightarrow H_0 \text{ -t elfogadjuk,}$$

és ekkor a vállalt másodfajú hiba értéke $m = 62$ alternatíva esetén az $u$-próba erőfüggvényével közelítve:

$$\beta\Big|_{m=62} \approx \Phi\left(1.592 + \frac{58.26-62}{6.69}\sqrt{15}\right) = 0.28327\ .$$

**Megjegyzés:** Pontosabb értéket kaphatunk a

$$\beta\Big|_{m=62} = P\left(\frac{\frac{\bar{\xi}-62}{\sigma}\sqrt{15} + \frac{62-58.26}{\sigma}\sqrt{15}}{\sqrt{\frac{s^*}{\sigma^2}}} < 1.592\right) = F_{T_{19}}(1.592) = 0.2820342$$

másodfajú hibára, ahol $F_{T_{19}}$ most az un. nem centrált $T$ eloszlás eloszlásfüggvénye, a nem-centráltsági paraméter becsült értéke

$$\frac{62-58.26}{\sigma}\sqrt{15} \approx \frac{62-58.26}{6.69}\sqrt{15} = 2.1652$$

amivel az R programcsomag

```
>pt(1.592,df=14,ncp=2.1652)
[1]  0.2820342
```

függvény-hívásával kapjuk az eredményt. Az így kapott érték lényegében azonos a fenti közelítéssel, és ez az eredmény sem tekinthető pontos értéknek, mivel használjuk hozzá az ismeretlen szórás becsült értékét.

**2.12. Feladat.** *Ugyanazt a mennyiséget kétféle eljárással mérték meg, és kaptuk:*

$$\begin{array}{lll} n_1 = 11 & \bar{x} = 123.01 & s_1^*(x) = 0.12 \\ n_2 = 21 & \bar{y} = 123.21 & s_2^*(y) = 0.13 \end{array}$$

*Van-e kimutatható különbség a két eljárás pontossága, és várható eredménye között? Feltételezhetjük a mérési eredmények normális eloszlását.*

**Megoldás:** Jelölje $m_1$ illeteve $m_2$ a két várható érték paramétert, $\sigma_1$ és $\sigma_2$ a szórás paramétereket. Vizsgáljuk először a $H_0 : \sigma_1 = \sigma_2$ hipotézist. A próba statisztika értéke (a nagyobb becsült értéket osztva a kisebbel):

$$f = \frac{s_2^{*2}(y)}{s_1^{*2}(x)} = \frac{0.13^2}{0.12^2} = 1.1736$$

ami az $f_{0.1} = 2.20$ (szabadsági fok: 20;10) táblázati értékkel

$$f = 1.1736 \not> f_{0.1} = 2.20\ ,$$

tehát $\alpha = 2 \cdot 0.1 = 0.2$ terjedelmű próbával elfogadható a szórások egyenlősége.

A várható értékek $H_0 : m_1 = m_2$ azonosságát a szórások egyenlőségének feltételezése mellett, a kétmintás $T$-próbával ellenőrízzük. A szórás becsült értéke a két mintából

$$S(x, y) = \sqrt{\frac{10 \cdot 0.12^2 + 20 \cdot 0.13^2}{30}} = 0.\,126\,75 \ ,$$

amivel a próba statisztika értéke

$$t = \frac{123.01 - 123.21}{0.\,126\,75} \cdot \sqrt{\frac{11 \cdot 21}{32}} = -4.\,239\,5 \ ,$$

és a $t_{0.001} = 3.646$ (szabadsági fok: 30) táblázati értékkel

$$t = |-4.\,239\,5| > t_{0.001} = 3.646$$

tehát $\alpha = 0.001$ elsőfajú hiba mellett elvetjük a várható értékek azonosságát.

Összefoglalva, a mérések pontossága (szórása) azonosnak tekinthető, a várható eredmények azonban szignifikánsan különböznek.

## 2.3. Nem paraméteres próbák

**2.13. Feladat.** *Szabályosnak tekinthető-e az a dobókocka, melyet 120-szor dobva, az alábbi gyakoriságokat figyeltük meg:*

| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 24 | 16 | 25 | 10 | 30 | 15 |

**Megoldás:** Vizsgáljuk a $H_0 : p_i = \frac{1}{6} \quad i = 1, 2, \ldots, 6$ hipotézist $\chi^2$-próbával.

| | $f_i$ | $n \cdot p_{i0}$ | $\frac{(f_i - n \cdot p_{i0})^2}{n \cdot p_{i0}}$ |
|---|---|---|---|
| 1 | 24 | 20 | $\frac{16}{20}$ |
| 2 | 16 | 20 | $\frac{16}{20}$ |
| 3 | 25 | 20 | $\frac{25}{20}$ |
| 4 | 10 | 20 | $\frac{100}{20}$ |
| 5 | 30 | 20 | $\frac{100}{20}$ |
| 6 | 15 | 20 | $\frac{25}{20}$ |
| $\sum$ | 120 | | $\frac{282}{20}$ |

tehát a próba statisztika értéke

$$\chi^2 = \frac{282}{20} = 14.\,1 \ ,$$

és a $\chi^2_{0.025} = 12.83$ (szabadsági fok: 5) táblázati értékkel

$$\chi^2_{0.025} = 12.83 < \chi^2 = 14.\,1 \ ,$$

tehát a $H_0$ hipotézist elutasítjuk 0.025 elsőfajú hibával, vagyis a kocka nem tekinthető szabályosnak.

**2.14. Feladat.** *Egy bizonyos típusú file hossz adataiból kaptuk az alábbi gyakorisági adatokat:*

| | $f_i$ |
|---|---|
| $-1\ kB$ | 14 |
| $1-2\ kB$ | 16 |
| $2-3\ kB$ | 25 |
| $3-4\ kB$ | 10 |
| $4-\quad kB$ | 10 |

*Tekinthetjük-e exponenciális eloszlású véletlen értéknek egy ilyen típusú file méretét?*

**Megoldás:** Vizsgáljuk a $H_0 : p_1 = p_{10}(\lambda) \quad p_2 = p_{20}(\lambda) \quad p_3 = p_{30}(\lambda) \quad p_4 = p_{40}(\lambda) \quad p_5 = p_{50}(\lambda)$ becsléses illeszkedési hipotézist, ahol $\lambda$ jelöli a feltételezett exponenciális eloszlás ismeretlen paraméterét, aminek maximum likelihood becsléssel kapott becsült értéke

$$\lambda \approx \frac{1}{\bar{x}} \approx \frac{75}{14 \cdot 0.5 + 16 \cdot 1.5 + 25 \cdot 2.5 + 15 \cdot 3.5 + 10 \cdot 4.5} = 0.392\,67$$

amit az osztályközök megadásával kaphatunk. Egészítsük ki továbbá táblázatunkat a becsült értékből nyert várható gyakoriságokkal, és a próba statisztika számításával

| | $f_i$ | $n \cdot \hat{p}_{i0}$ | $\frac{(f_i - n\cdot\hat{p}_{i0})^2}{n\cdot\hat{p}_{i0}}$ |
|---|---|---|---|
| 0.5 | 14 | $75 \cdot 0.324\,75 = 24.356$ | $\frac{10.356^2}{24.356} = 4.403\,3$ |
| 1.5 | 16 | $75 \cdot 0.219\,29 = 16.447$ | $\frac{0.447^2}{16.447} = 1.214\,9 \times 10^{-2}$ |
| 2.5 | 25 | $75 \cdot 0.148\,07 = 11.105$ | $\frac{(25-11.105)^2}{11.105} = 17.386$ |
| 3.5 | 10 | $75 \cdot 9.998\,7 \times 10^{-2} = 7.499$ | $\frac{(10-7.499)^2}{7.499} = 0.834\,11$ |
| 4.5 | 10 | $75 \cdot 0.207\,9 = 15.593$ | $\frac{(10-15.593)^2}{15.593} = 2.006\,1$ |
| $\sum$ | 75 | 75.000 | 24.642 |

ahol

$$\begin{aligned}
\hat{p}_{10} &= 1 - e^{-0.392\,67\cdot 1} = 0.324\,75 \\
\hat{p}_{20} &= e^{-0.392\,67\cdot 1} - e^{-0.392\,67\cdot 2} = 0.219\,29 \\
\hat{p}_{30} &= e^{-0.392\,67\cdot 2} - e^{-0.392\,67\cdot 3} = 0.148\,07 \\
\hat{p}_{40} &= e^{-0.392\,67\cdot 3} - e^{-0.392\,67\cdot 4} = 9.998\,7 \times 10^{-2} \\
\hat{p}_{50} &= e^{-0.392\,67\cdot 4} = 0.207\,9
\end{aligned}$$

A próba statisztika értéke 24.642, és $\chi^2_{0.001} = 16.27$ (szabadsági fok: $5 - 1 - 1 = 3$), amivel

$$24.642 > 16.27$$

ezért a $H_0$ hipotézist elutasítjuk $\alpha = 0.001$ elsőfajú hiba mellett. Tehát a véletlen mennyiség nem tekinthető exponenciális eloszlásúnak.

**2.15. Feladat.** *Egy termék három különböző technológiával készülhet I., II. és III. osztályú minősgben. Egy felmérésből kaptuk az alábbi gyakorisági táblázatot:*

| minőség \ technológia | 1. | 2. | 3. |
|---|---|---|---|
| I. | 50 | 4 | 1 |
| II. | 44 | 41 | 33 |
| III. | 6 | 15 | 26 |

*Van-e kimutatható kapcsolat a minőség és az alkalmazott technológia között?*

**Megoldás:** Vizsgáljuk a $H_0 : p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$ függetlenségi hipotézist, amihez egészítsük ki a táblázatot az összegekkel, és a várható gyakoriságokkal:

| minőség \ technológia | 1. | 2. | 3. | $\sum$ |
|---|---|---|---|---|
| I. | 50 $\frac{55\cdot100}{220} = 25$ | 4 $\frac{55\cdot60}{220} = 15$ | 1 $\frac{55\cdot60}{220} = 15$ | 55 |
| II | 44 $\frac{118\cdot100}{220} = \frac{590}{11}$ | 41 $\frac{118\cdot60}{220} = \frac{354}{11}$ | 33 $\frac{118\cdot60}{220} = \frac{354}{11}$ | 118 |
| III. | 6 $\frac{47\cdot100}{220} = \frac{235}{11}$ | 15 $\frac{47\cdot60}{220} = \frac{141}{11}$ | 26 $\frac{47\cdot60}{220} = \frac{141}{11}$ | 47 |
| $\sum$ | 100 | 60 | 60 | 220 |

A próba statisztika értéke:

$$\frac{(50-25)^2}{25} + \frac{(4-15)^2}{15} + \frac{(1-15)^2}{15} +$$
$$+ \frac{\left(44-\frac{590}{11}\right)^2}{\frac{590}{11}} + \frac{\left(41-\frac{354}{11}\right)^2}{\frac{354}{11}} + \frac{\left(33-\frac{354}{11}\right)^2}{\frac{354}{11}} +$$
$$+ \frac{\left(6-\frac{235}{11}\right)^2}{\frac{235}{11}} + \frac{\left(15-\frac{141}{11}\right)^2}{\frac{141}{11}} + \frac{\left(26-\frac{141}{11}\right)^2}{\frac{141}{11}} = 79.446$$

továbbá $\alpha = 0.001$ $\chi^2_\alpha = 18.47$ (szabadsági fok: $(3-1)\cdot(3-1) = 4$), amivel

$$18.47 < 79.446$$

tehát $H_0$-t elutasítjuk $\alpha = 0.001$ elsőfajú hibával, vagyis van kapcsolat a minőség és a technológia között.

**2.16. Feladat.** *Egy véletlen-szám generátor hívásának $n = 20$ eredménye:*

| | | | | |
|---|---|---|---|---|
| 0.69278073 | 0.89428280 | 0.72684370 | 0.94735859 | 0.25751298 |
| 0.32268032 | 0.21392526 | 0.33039386 | 0.06091866 | 0.08225982 |
| 0.54508760 | 0.90420617 | 0.87403826 | 0.28252200 | 0.84477326 |
| 0.54796052 | 0.94966769 | 0.33603055 | 0.19376286 | 0.15896395 |

*Elfogadható-e, a [0;1] intervallumon egyenletes eloszlás feltételezése?*

**Megoldás:** Vizsgáljuk a $H_0 : F(x) = x \quad 0 \leq x \leq 1$ hipotézist, amihez keressük meg az empirikus eloszlásfüggvény és az egyenletes eloszlás eloszlásfüggvényének legnagyobb eltérését:

| $x_k^*$ | $F_x^n(x_k^*)$ | $F_x^n(x_k^*+0)$ | $F_0(x_k^*)$ | $\lvert\Delta\rvert$ |
|---|---|---|---|---|
| 0.06091866 | 0.00 | 0.05 | 0.06091866 | 0.06091866 |
| 0.08225982 | 0.05 | 0.10 | 0.08225982 | |
| 0.15896395 | 0.10 | 0.15 | 0.15896395 | |
| 0.19376286 | 0.15 | 0.20 | 0.19376286 | |
| 0.21392526 | 0.20 | 0.25 | 0.21392526 | |
| 0.25751298 | 0.25 | 0.30 | 0.25751298 | |
| 0.28252200 | 0.30 | 0.35 | 0.28252200 | 0.067478 |
| 0.32268032 | 0.35 | 0.40 | 0.32268032 | 0.0 773 2 |
| 0.33039386 | 0.40 | 0.45 | 0.33039386 | 0. 119 61 |
| 0.33603055 | 0.45 | 0.50 | 0.33603055 | 0. 163 97 |
| 0.54508760 | 0.50 | 0.55 | 0.54508760 | |
| 0.54796052 | 0.55 | 0.60 | 0.54796052 | |
| 0.69278073 | 0.60 | 0.65 | 0.69278073 | |
| 0.72684370 | 0.65 | 0.70 | 0.72684370 | |
| 0.84477326 | 0.70 | 0.75 | 0.84477326 | |
| 0.87403826 | 0.75 | 0.80 | 0.87403826 | |
| 0.89428280 | 0.80 | 0.85 | 0.89428280 | |
| 0.90420617 | 0.85 | 0.90 | 0.90420617 | |
| 0.94735859 | 0.90 | 0.95 | 0.94735859 | |
| 0.94966769 | 0.95 | 1.00 | 0.94966769 | |

A próba statisztika értéke

$$z = 0.163\,97 \cdot \sqrt{20} = 0.\,733\,3\ ,$$

és a Kolmogorov-féle $K$ függvény értéke táblázatból:

$$K(0.\,73) = 0.339$$

tehát $\alpha = 1 - 0.339 = 0.\,661$ terjedelmű próbával elfogadjuk az egyenletes eloszlás feltételezését.

## 2.4. Függőségi kapcsolatok

**2.17. Feladat.** *Egy $\mathcal{N}(0;\sigma)$ eloszlású véletlen hibával mérhető "függő" mennyiség egy "független" változó*

$$y = ax + b \qquad x \in \mathbb{R}$$

*lineáris függvénye. A független változó*

$$x = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \end{bmatrix}^T$$

*értékeihez tartozó függő változó mért (véletlen hibát tartalmazó) értékei*

$$y = \left[\begin{array}{ccccc} 2.0235 & 2.8557 & 3.8921 & 4.9978 & 5.7415 \end{array}\right]^T .$$

**a)** *Becsüljük az $a, b$ regressziós együtthatókat és a hiba $\sigma$ szórását!*

**b)** *Adjunk 90%-os szintű intervallum becslést a függő változó (várható-) értékére $x = 6$ esetén!*

**c)** *Legfeljebb mennyi lehet a függő változó három mért (véletlen-) értékének átlaga 90%-os biztonsággal $x = 4$ esetén?*

**d)** *95%-s biztonsággal milyen határok között van a független változó értéke, amelyhez tartozó három mérési eredmény átlaga 3.52?*

**e)** *Lehet-e a mérési hiba szórása 0.1?*

**Megoldás:** A függőségi kapcsolat modellje:

$$\eta(x) = ax + b + \varepsilon$$

ahol $\varepsilon \in \mathcal{N}(0; \sigma)$. Ekkor a terv-mátrix

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \end{bmatrix}$$

amivel kapjuk

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 55 & 15 \\ 15 & 5 \end{bmatrix} \qquad \mathbf{X}^T y = \begin{bmatrix} 68.11 \\ 19.511 \end{bmatrix} \qquad \left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \begin{bmatrix} .1 & -.3 \\ -.3 & 1.1 \end{bmatrix} .$$

**a)** A paraméterek maximum likelihood becslése

$$\begin{bmatrix} a \\ b \end{bmatrix} \approx \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \mathbf{X}^T y = \begin{bmatrix} 0.95781 \\ 1.0287 \end{bmatrix}$$

és

$$\sigma^2 \approx \frac{1}{5}\left(y^T y - y^T \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T y\right) = 6.8284 \times 10^{-3} ,$$

illetve a torzítatlanság miatt korrigált becslés

$$\sigma \approx s_R(y) = \sqrt{\frac{5}{3} \cdot 6.8284 \times 10^{-3}} = 0.10668$$

amivel az együtthatók becslésének kovariancia mátrixa az alábbi módon becsülhető:

$$(0.10668)^2 \begin{bmatrix} .1 & -.3 \\ -.3 & 1.1 \end{bmatrix} = \begin{bmatrix} 1.1381 \times 10^{-3} & -3.4142 \times 10^{-3} \\ -3.4142 \times 10^{-3} & 1.2519 \times 10^{-2} \end{bmatrix} .$$

**b)** Mivel az együtthatók $a \cdot 6 + b$ lineáris függvénye szórásának becslése

$$0.10668 \cdot \sqrt{\begin{bmatrix} 6 & 1 \end{bmatrix} (\mathbf{X}^T\mathbf{X})^{-1} \begin{bmatrix} 6 & 1 \end{bmatrix}^T} = 0.11189 \ ,$$

és

$$\alpha = 0.1 \qquad t_{0.1} = 2.353 \qquad \text{sz.f.: } 3$$

kapjuk

$$y(5) \approx 0.95781 \cdot 6 + 1.0287 \pm 2.353 \cdot 0.11189 \begin{array}{l} \nearrow 7.0388 \\ \searrow 6.5123 \end{array} .$$

**c)** Három további mért érték $\overline{\eta}(4)$ átlagának szórás becslése

$$0.10668 \cdot \sqrt{\frac{1}{3} + \begin{bmatrix} 4 & 1 \end{bmatrix} (\mathbf{X}^T\mathbf{X})^{-1} \begin{bmatrix} 4 & 1 \end{bmatrix}^T} = 8.4898 \times 10^{-2} \ ,$$

és

$$\alpha = 0.2 \qquad t_{0.2} = 1.638 \qquad \text{sz.f.: } 3$$

amivel

$$\overline{\eta}(4) \approx 0.95781 \cdot 4 + 1.0287 \pm 1.638 \cdot 8.4898 \times 10^{-2} \begin{array}{l} \nearrow 4.999 \\ \searrow \end{array} ,$$

tehát $\overline{\eta}(4) \leq 4.999$ 90%-os biztonsággal.

**d)** Mivel az ismeretlen $x_0$ értékhez tartozó $k = 3$ független megfigyelés átlagára teljesül

$$\overline{\eta}(x_0) - \left(\widehat{a}x_0 + \widehat{b}\right) \in \mathcal{N}\left(0; \sigma\sqrt{\frac{1}{k} + \begin{bmatrix} x_0 & 1 \end{bmatrix} (\mathbf{X}^T\mathbf{X})^{-1} \begin{bmatrix} x_0 & 1 \end{bmatrix}^T}\right)$$

és így

$$\mathbb{P}\left(\left|\frac{\overline{\eta}(x_0) - \left(\widehat{a}x_0 + \widehat{b}\right)}{s_R\sqrt{\frac{1}{k} + \begin{bmatrix} x_0 & 1 \end{bmatrix} (\mathbf{X}^T\mathbf{X})^{-1} \begin{bmatrix} x_0 & 1 \end{bmatrix}^T}}\right| \leq t_a\right) = 1 - \alpha \ .$$

Tehát keressük a megfigyelt mintából számolt értékekkel és $t_{0.05} = 3.182$ táblázati értékkel az

$$(3.52 - 0.95781x - 1.0287)^2 \leq 3.182^2 \cdot 0.10668^2 \cdot \left(\frac{1}{3} + \frac{1}{10}x^2 - \frac{3}{5}x + \frac{11}{10}\right)$$

egyenlőtlenség megoldását, ami

$$\{2.3316 \leq x \leq 2.8603\}$$

tehát a keresett érték 90%-os határai:

$$x_0 \in (2.3316; 2.8603) \ .$$

**e)** Vizsgáljuk a $H_0 : \sigma = 0.1$ hipotézist! Mivel $H_0$ esetén

$$\frac{3}{0.1^2} s_R^2 \in \chi_3^2 \ ,$$

a próba statisztika értéke:

$$\frac{3}{0.1^2} s_R^2(y) = \frac{3}{0.1^2} 0.106\,68^2 = 3.414\,2 \ .$$

Válasszuk $\alpha = 0.2$, akkor $\chi^2_{0.9} = 0.58 \quad \chi^2_{0.1} = 6.25$ sz.f.: 3, amivel

$$0.58 < 3.414\,2 < 6.25$$

tehát elfogadjuk a $H_0 : \sigma = 0.1$ hipotézist.

**2.18. Feladat.** *Egy függőleges irányban, felfelé elhajított kődarab magasságát a*

$$t = (1; 1.5; 2; 2.5; 3) \quad [s]$$

*időpontokban megmérve, kaptuk az*

$$y = \left[\ 9.117\,6 \quad 12.694 \quad 16.024 \quad 18.606 \quad 20.892\ \right]^T \quad [m]$$

*értékeket.*

**a)** *Becsüljük az*

$$y = at^2 + bt + c$$

*mozgás-egyenlet paramétereit 90%-os szintű határokkal, és a hiba szórását!*

**b)** *Milyen határok között lesz a magasság várható értéke, illetve a megfigyelt érték a $t = 5$ időpontban 95%-os szint mellett?*

**c)** *A magasságot még a "felszálló ágon" három független kísérletben, ugyanazon ismeretlen $t_0$ időpontban megmérve, kaptuk*

$$y_{t_0} = [15.118, 14.944, 15.024]$$

*Milyen határok között van a $t_0$ időpont 90%-os biztonsággal?*

**d)** *Milyen határok között van 90%-os biztonsággl az a $t_m$ időpont, amikor a legnagyobb magasságot éri el a kődarab?*

**Megoldás:** A függőségi kapcsolat modellje:

$$\eta(t) = at^2 + bt + c + \varepsilon \quad t \in \mathbb{R}$$

ahol $\varepsilon \in \mathcal{N}(0;\sigma)$. Ekkor a terv-mátrix

$$\mathbf{X} = \begin{bmatrix} 1^2 & 1 & 1 \\ 1.5^2 & 1.5 & 1 \\ 2^2 & 2 & 1 \\ 2.5^2 & 2.5 & 1 \\ 3^2 & 3 & 1 \end{bmatrix},$$

amivel kapjuk

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 142.13 & 55.0 & 22.5 \\ 55.0 & 22.5 & 10.0 \\ 22.5 & 10.0 & 5 \end{bmatrix} \qquad \mathbf{X}^T y = \begin{bmatrix} 406.09 \\ 169.4 \\ 77.334 \end{bmatrix}$$

és

$$\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \begin{bmatrix} 1.1364 & -4.5455 & 3.9773 \\ -4.5455 & 18.582 & -16.709 \\ 3.9773 & -16.709 & 15.72 \end{bmatrix}.$$

**a)** A paraméterek maximum likelihood becslése

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} \approx \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T y = \begin{bmatrix} -0.95109 \\ 9.6965 \\ 0.3536 \end{bmatrix}$$

és

$$\sigma^2 \approx \frac{1}{5}\left(y^T y - y^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T y\right) = 2.5994 \times 10^{-3},$$

illetve a torzítatlanság miatt korrigált becslés

$$\sigma \approx s_R(y) = \sqrt{\frac{5}{2} \cdot 2.5994 \times 10^{-3}} = 8.0613 \times 10^{-2}$$

amivel az együtthatók becslésének kovariancia mátrixa az alábbi módon becsülhető:

$$\left(8.0613 \times 10^{-2}\right)^2 \begin{bmatrix} 1.1364 & -4.5455 & 3.9773 \\ -4.5455 & 18.582 & -16.709 \\ 3.9773 & -16.709 & 15.72 \end{bmatrix} =$$
$$= \begin{bmatrix} 7.3848 \times 10^{-3} & -2.9539 \times 10^{-2} & 2.5846 \times 10^{-2} \\ -2.9539 \times 10^{-2} & .12075 & -.10858 \\ 2.5846 \times 10^{-2} & -.10858 & .10216 \end{bmatrix}.$$

Az együtthatók intervallum becsléséhez

$$\alpha = 0.1 \qquad t_{0.1} = 2.920 \qquad \text{sz.f.: } 2$$

és így

$$\begin{aligned} a &\approx -0.95109 \pm 2.920 \cdot \sqrt{7.3848 \times 10^{-3}} \begin{array}{l} \nearrow -0.700 \\ \searrow -1.202 \end{array} \\ b &\approx 9.6965 \pm 2.920 \cdot \sqrt{0.12075} \begin{array}{l} \nearrow 10.711 \\ \searrow 8.682 \end{array} \\ c &\approx 0.3536 + 2.920 \cdot \sqrt{0.10216} \begin{array}{l} \nearrow 1.287 \\ \searrow -0.580 \end{array} \end{aligned}$$

**b)** Mivel az együtthatók lineáris függvénye szórásának becslése

$$8.0613 \times 10^{-2} \cdot \sqrt{\begin{bmatrix} 25 & 5 & 1 \end{bmatrix} (\mathbf{X}^T\mathbf{X})^{-1} \begin{bmatrix} 25 & 5 & 1 \end{bmatrix}^T} = 0.74918 \ ,$$

és

$$\alpha = 0.05 \qquad t_{0.05} = 4.303 \qquad \text{sz.f.: } 2$$

kapjuk

$$m(5) \approx -0.95109 \cdot 5^2 + 9.6965 \cdot 5 + 0.3536 \pm 4.303 \cdot 0.74918 \begin{array}{l} \nearrow 28.283 \\ \searrow 21.835 \end{array} .$$

A magasság értékének becsléséhez a szórás becslése

$$8.0613 \times 10^{-2} \cdot \sqrt{\frac{1}{1} + \begin{bmatrix} 25 & 5 & 1 \end{bmatrix} (\mathbf{X}^T\mathbf{X})^{-1} \begin{bmatrix} 25 & 5 & 1 \end{bmatrix}^T} = 0.75351 \ ,$$

amivel

$$\eta(5) \approx -0.95109 \cdot 5^2 + 9.6965 \cdot 5 + 0.3536 \pm 4.303 \cdot 0.75351 \begin{array}{l} \nearrow 28.301 \\ \searrow 21.816 \end{array} .$$

**c)** Mivel a $t_0$ időponthoz tartozó $k = 3$ független megfigyelés átlagára teljesül

$$\overline{\eta}(t_0) - \left(\widehat{a}t_0^2 + \widehat{b}t_0 + \widehat{c}\right) \in \mathcal{N}\left(0; \sigma\sqrt{\frac{1}{k} + \begin{bmatrix} t_0^2 & t_0 & 1 \end{bmatrix} (\mathbf{X}^T\mathbf{X})^{-1} \begin{bmatrix} t_0^2 & t_0 & 1 \end{bmatrix}^T}\right)$$

és így

$$\mathbb{P}\left(\left|\frac{\overline{\eta}(t_0) - \left(\widehat{a}t_0^2 + \widehat{b}t_0 + \widehat{c}\right)}{s_R\sqrt{\frac{1}{k} + \begin{bmatrix} t_0^2 & t_0 & 1 \end{bmatrix} (\mathbf{X}^T\mathbf{X})^{-1} \begin{bmatrix} t_0^2 & t_0 & 1 \end{bmatrix}^T}}\right| \leq t_a\right) = 1 - \alpha \ .$$

Tehát keressük a megfigyelt mintából számolt értékekkel és $t_{0.1} = 2.920$ táblázati értékkel az

$$\begin{aligned} &\left(15.029 + 0.95109t^2 - 9.6965t - 0.3536\right)^2 \leq \\ &\qquad 6.3327 \times 10^{-2}t^4 - .5066t^3 + 1.4786t^2 - 1.8617t + .89391 \end{aligned}$$

egyenlőtlenség megoldását, ami

$$\{1.8209 \le t \le 1.8766\} \qquad \text{vagy} \qquad \{7.0656 \le t \le 10.559\}$$

tehát a (felszálló ágban) keresett időpont 90%-os határai:

$$t_0 \in (1.8209; 1.8766)$$

**d)** Mivel a keresett időpont $t_m = -\frac{b}{2a}$, ezért $2t_m a + b = 0$, tehát

$$2t_m\widehat{a} + \widehat{b} \in \mathcal{N}\left(0; \sigma\sqrt{\begin{bmatrix} 2t_m & 1 & 0 \end{bmatrix} (\mathbf{X}^T\mathbf{X})^{-1} \begin{bmatrix} 2t_m & 1 & 0 \end{bmatrix}^T}\right)$$

és így

$$\mathbb{P}\left(\left|\frac{2t_m\widehat{a} + \widehat{b}}{s_R\sqrt{\begin{bmatrix} 2t_m & 1 & 0 \end{bmatrix} (\mathbf{X}^T\mathbf{X})^{-1} \begin{bmatrix} 2t_m & 1 & 0 \end{bmatrix}^T}}\right| \le t_a\right) = 1 - \alpha \ .$$

Tehát keressük a megfigyelt mintából számolt értékekkel és $t_{0.1} = 2.920$ táblázati értékkel az

$$\begin{aligned}(2t \cdot (-0.95109) + 9.6965)^2 &\le \\ &\le 2.920^2 \cdot \left(8.0613 \times 10^{-2}\right)^2 \cdot \left(4t^2 \cdot 1.1364 + 4t \cdot (-4.5455) + 18.582\right)\end{aligned}$$

egyenlőtlenség megoldását, ami

$$\{4.4472 \le t \le 6.2115\}$$

tehát a 90%-os határok

$$t_m \in (4.4472; 6.2115) \ .$$

**2.19. Feladat.** *Egy tantárgy vizsgáján a bukások száma a félév során gyüjtött pontok száma szerint az alábbi volt:*

| *pontszám*$(x)$ | *hallgatók száma*$(n)$ | *közülük megbukott*$(f)$ |
|---|---|---|
| 25 | 10 | 2 |
| 15 | 20 | 12 |
| 10 | 30 | 24 |

**a)** *Keressük a bukás valószínűségének*

$$p(x) = \frac{e^{ax+b}}{1 + e^{ax+b}} \tag{2.2}$$

*alakú függvényét!*

**b)** *Adjunk* 90% *szintű intervallum becslést az* $\frac{1}{2}$ *valószínűséget adó pontszámra!*

**Megoldás:** Keressük az odds-hányados logaritmusának

$$\ln \frac{p(x)}{1-p(x)} = ax + b$$

lineáris közelítését, ahol az

$$y = \left[ \begin{array}{ccc} \ln \frac{\frac{2}{10}}{1-\frac{2}{10}} & \ln \frac{\frac{12}{20}}{1-\frac{12}{20}} & \ln \frac{\frac{24}{30}}{1-\frac{24}{30}} \end{array} \right]^T = \left[ \begin{array}{ccc} -1.3863 & .40547 & 1.3863 \end{array} \right]^T$$

megfigyelt mintát

$$\widehat{y_i} = ax_i + b \quad i = 1, 2, 3$$

várható értékű, ahol

$$x = \left[ \begin{array}{ccc} 25 & 15 & 10 \end{array} \right]^T$$

és

$$\sigma_i = \frac{1}{\sqrt{n_i}} \cdot \frac{1}{\sqrt{p(x_i)\,(1-p(x_i))}} = \frac{1}{\sqrt{n_i}} \frac{1+e^{ax_i+b}}{e^{\frac{1}{2}(ax_i+b)}} \quad i = 1, 2, 3$$

szórású független, normális eloszlásra vonatkozónak tekintjük, mivel

$$\frac{\partial}{\partial p} \ln \frac{p}{1-p} = \frac{1}{p\,(1-p)}$$

és a $p$ valószínűséget becslő relatív gyakoriság szórása

$$\sqrt{\frac{p(1-p)}{n}}\,.$$

**a)** Használjunk iterációs lépésenkénti súlyozást a lineáris kapcsolat paramétereinek becslésére!

1. Indulásként adjuk meg a becsléseket súlyozás nélkül:

$$\mathbf{X} = \left[ \begin{array}{cc} x & \mathbf{1} \end{array} \right] = \left[ \begin{array}{cc} 25 & 1 \\ 15 & 1 \\ 10 & 1 \end{array} \right] \quad \Rightarrow \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \mathbf{X}^T y = \left[ \begin{array}{c} -.18403 \\ 3.2023 \end{array} \right]$$

2. Használjuk a

$$W = \left[ \begin{array}{ccc} \frac{10e^{a\cdot 25+b}}{\left(1+e^{a\cdot 25+b}\right)^2} & 0 & 0 \\ 0 & \frac{20e^{a\cdot 15+b}}{\left(1+e^{a\cdot 15+b}\right)^2} & 0 \\ 0 & 0 & \frac{30e^{a\cdot 10+b}}{\left(1+e^{a\cdot 10+b}\right)^2} \end{array} \right] = \left[ \begin{array}{ccc} 1.5883 & 0 & 0 \\ 0 & 4.7637 & 0 \\ 0 & 0 & 4.87 \end{array} \right]$$

súlymátrixot, amivel

$$\left(\mathbf{X}^T W \mathbf{X}\right)^{-1} \mathbf{X}^T W y = \left[ \begin{array}{c} -.18558 \\ 3.2212 \end{array} \right]$$

3. Mint előbb, kapjuk

$$W = \begin{bmatrix} 1.5693 & 0 & 0 \\ 0 & 4.7682 & 0 \\ 0 & 0 & 4.8602 \end{bmatrix}$$

amivel

$$\left(\mathbf{X}^T W \mathbf{X}\right)^{-1} \mathbf{X}^T W y = \begin{bmatrix} a = -.1856 \\ b = 3.2215 \end{bmatrix}$$

4.

$$W = \begin{bmatrix} 1.5691 & 0 & 0 \\ 0 & 4.7682 & 0 \\ 0 & 0 & 4.8599 \end{bmatrix}$$

$$\left(\mathbf{X}^T W \mathbf{X}\right)^{-1} \mathbf{X}^T W y = \begin{bmatrix} -.1856 \\ 3.2215 \end{bmatrix}$$

ami már ($10^{-4}$ pontossággal) azonos az előző lépésben kapott eredménnyel.

Tehát a paraméterek becsült értéke

$$a \approx \widehat{a}(y) = -0.1856 \qquad b \approx \widehat{b}(y) = 3.2215$$

amivel

$$p(x) \approx \widehat{p}(x) = \frac{e^{-0.1856x+3.2215}}{1+e^{-0.1856x+3.2215}} .$$

A maradék négyzetösszeg és a hiba szórásnégyzetének becslése

$$\begin{aligned} SS(y) &= y^T W y - y^T W \mathbf{X} \left(\mathbf{X}^T W \mathbf{X}\right)^{-1} \mathbf{X}^T W y = 8.6212 \times 10^{-3} \\ \sigma^2 &\approx s_R^2(y) = \frac{1}{3-2} SS = 8.6212 \times 10^{-3} \end{aligned}$$

ahol

$$\left(\mathbf{X}^T W \mathbf{X}\right)^{-1} = \begin{bmatrix} 3.6793 \times 10^{-3} & -5.2361 \times 10^{-2} \\ -5.2361 \times 10^{-2} & .83447 \end{bmatrix}$$

$$\begin{bmatrix} 3.6793 \times 10^{-3} \cdot 8.6212 \times 10^{-3} & -5.2361 \times 10^{-2} \cdot 8.6212 \times 10^{-3} \\ -5.2361 \times 10^{-2} \cdot 8.6212 \times 10^{-3} & .83447 \cdot 8.6212 \times 10^{-3} \end{bmatrix}$$

tehát az együtthatók becslésének kovariancia mátrixa (pontosabban annak becsült értéke):

$$s_R^2(y) \left(\mathbf{X}^T W \mathbf{X}\right)^{-1} = \begin{bmatrix} 3.172 \times 10^{-5} & -4.5141 \times 10^{-4} \\ -4.5141 \times 10^{-4} & 7.1941 \times 10^{-3} \end{bmatrix} .$$

**b)** Mivel azt az $x_0$ értéket keressük, amivel $ax_0 + b = 0$, tehát a becsült érték

$$x_0 \approx -\frac{3.\,221\,5}{-0.\,185\,6} = 17.\,357\ .$$

Az intervallum becsléshez írhatjuk

$$\widehat{a}x_0 + \widehat{b} \in \mathcal{N}\left(0; \sigma\sqrt{\left[\begin{array}{cc} x_0 & 1 \end{array}\right] (\mathbf{X}^T W \mathbf{X})^{-1} \left[\begin{array}{cc} x_0 & 1 \end{array}\right]^T}\right)$$

és

$$\mathbb{P}\left(\left|\frac{\widehat{a}x_0 + \widehat{b}}{s_R\sqrt{\left[\begin{array}{cc} x_0 & 1 \end{array}\right] (\mathbf{X}^T W \mathbf{X})^{-1} \left[\begin{array}{cc} x_0 & 1 \end{array}\right]^T}}\right| \leq t_\alpha\right) = 1 - \alpha\ .$$

Válasszuk tehát $\alpha = 0.1$ esetén $t_{0.1} = 6.314$ (sz.f.: 1) kritikus értéket, és keressük az

$$\frac{(-0.\,185\,6x + 3.\,221\,5)^2}{3.\,172 \times 10^{-5}x^2 - 2 \cdot 4.\,514\,1 \times 10^{-4}x + 7.\,194\,1 \times 10^{-3}} \leq 6.314^2$$

egyenlőtlenség megoldását, amiből a 90%-os határok:

$$x_0 \in (16.\,331; 18.\,622)$$

**Megjegyzés:** A regressziós függvény 2.2 alakú keresését az indokolja, hogy ha egy $\xi$ mennyiség feltételes eloszlása két alternatíva (egy esemény és komplementere) esetén normális, azonos szórásokkal, akkor az események feltételes valószínűsége ilyen függvénye a $\xi = x$ értéknek. Az ilyen problémát *LOGIT* regressziónak nevezik. Ha még a szórások is különböznek, az odds-hányados logaritmusa az $x$ érték másodfokú függvénye lesz. A maradék négyzetösszegből kapott $s_R^2(y) = 8.\,621\,2 \times 10^{-3}$ a súlyozás miatt tudottan $\sigma^2 = 1$ érték becslése, ami az igazi értéknél jóval kisebb ugyan, de az 1 szabadsági fokú $\chi^2$ eloszlás eloszlásfüggvényével (lásd pl. az R programcsomagban a `pchisq(0.0086212,df=1)` parancs hívásával)

$$F(8.\,621\,2 \times 10^{-3}) = 0.0739776$$

tehát a becsült érték még így is nagyobb mint a 90%-os szintű konfidencia intervallum alsó határa.