Valószínűségszámítás és statisztika gyakorlat

Programtervező informatikus szak, esti képzés

1.) Egy érmével dobunk. Ha az eredmény fej, akkor még egyszer dobunk, ha írás, akkor még kétszer.

a.) Mik lesznek a kísérletet leíró eseménytér pontjai?
b.) Határozzuk meg az elemi események valószínűségét!
c.) Mennyi a valószínűsége, hogy összesen 1 fejet dobunk?

Megoldás.

a.) Jelölje I azt, hogy írást dobtunk, F pedig a fejdobást.
$\Omega = \{FI, FF, III, IFI, IIF, IFF\}$
b.) $P(FI) = P(FF) = 1/4$, $P(III) = P(IFI) = P(IIF) = P(IFF) = 1/8$
c.) $P(\{FI, IFI, IIF\}) = \frac{1}{4} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}$

2.) Egy arany és egy ezüst érmével dobunk, majd újra dobunk azzal/azokkal az érmével/érmékkel, amelyikkel/amelyekkel fejet kaptunk. Írjuk fel az eseményteret! Határozd meg az elemi események valószínűségét!

Megoldás.
Jelölje I azt, hogy írást dobtunk, F pedig a fejdobást.
$\Omega = \{II, FII, FIF, IFI, IFF, FFII, FFIF, FFFI, FFFF\}$, ahol például $FFIF$ jelölje azt elemi eseményt, hogy az arany érmével először fejet, az ezüst érmével először fejet, az arany érmével másodszorra írást, az ezüst érmével pedig másodikra fejet dobtunk.
$P(II) = \frac{1}{4}$, $P(FII) = P(FIF) = P(IFI) = P(IFF) = \frac{1}{8}$, $P(FFII) = P(FFIF) = P(FFFI) = P(FFFF) = \frac{1}{16}$

3.) Mi a valószínűsége, hogy egy véletlenszerűen kiválasztott 6 jegyű szám jegyei mind különbözőek?

Megoldás. Összes lehetőség: $9 \cdot 10^5$; jó esetek száma: $9 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5$. Így a keresett valószínűség: $\frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5}{10^5}$ ≈15,12 % .

4.) Aritmethiában az autók rendszámai hatjegyű számok 000000 és 999999 között. Mi a valószínűsége, hogy van 6 a jegyek között?

Megoldás. P(van 6 a jegyek között)=1-P(nincs 6 köztük)=$1 - \frac{9^6}{10^6} = 1 - 0{,}9^6$

5.) Lottóhúzás során (5-ös lottó)

a.) milyen eséllyel lesz telitalálatom?
b.) milyen eséllyel lesz két találatom?
c.) milyen eséllyel lesz legalább két találatom?

Megoldás. A feladat kezelhető mintavételként: N=90, M=5, n=5.

a.) $\frac{\binom{5}{5}\binom{85}{0}}{\binom{90}{5}}$
b.) $\frac{\binom{5}{2}\binom{85}{3}}{\binom{90}{5}}$
c.) $\sum_{k=2}^{5} \frac{\binom{5}{k}\binom{85}{5-k}}{\binom{90}{5}}$

6.) Ha egy magyarkártya-csomagból visszatevéssel húzunk 3 lapot, akkor mi annak a valószínűsége, hogy

a.) pontosan
b.) legalább egy piros színű lapot húzunk?

Megoldás. A feladat kezelhető mintavételként: $N = 32$ (összes lap), $M = 8$ (pirosak), $n = 3$.

<u>visszatevés nélkül</u>

a.) $\frac{\binom{8}{1}\binom{24}{2}}{\binom{32}{3}}$
b.) P(legalább 1 piros)=1-P(0 piros)=$1 - \frac{\binom{8}{0}\binom{24}{3}}{\binom{32}{3}}$

<u>visszatevéssel</u>

a.) a selejtarány=8/32=1/4, így a keresett vsz.: $\binom{3}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^2$
b.) $1 - \left(\frac{3}{4}\right)^3$.

7.) Egy 32 tagú osztályban a diákok angolt, németet vagy franciát tanulhatnak. Tudjuk, hogy angolul 20-an tanulnak, németül 12-en, franciául pedig 9-en. Angolul és németül egyszerre 5-en, németül és franciául egyszerre 3-an, angolul és franciául 2-en, és senki nem tanulja mind a három nyelvet. Mekkora a valószínűsége annak, hogy egy véletlenszerűen választott tanuló legalább az egyik idegen nyelvet tanulja?

Megoldás. Legyen B_1: egy diák angolul tanul; B_2: egy diák németül tanul; B_3: egy diák franciául tanul. Kiszámítandó a $P(B_1 \cup B_2 \cup B_3)$ valószínűség, használjuk a szita formulát.
$P(B_1 \cup B_2 \cup B_3) = P(B_1) + P(B_2) + P(B_3) - P(B_1 \cap B_2) - P(B_1 \cap B_3) - P(B_2 \cap B_3) + P(B_1 \cap B_2 \cap B_3) = \frac{20+12+9-5-3-2+0}{32} = \frac{31}{32}$

8.) Mennyi a valószínűsége, hogy két kockadobásnál mind a két dobás 6-os, azzal a feltétellel, hogy legalább az egyik dobás 6-os?

Megoldás. A feltétel figyelembe vételével oljuk meg: legalább az egyik 6-os
összes eset: 16,61,12,26,36,63,46,64,56,65,66 $\longrightarrow$ 11 darab
jó esetek: 66 $\longrightarrow$ 1 darab

így a keresett valószínűség $\frac{1}{11}$.

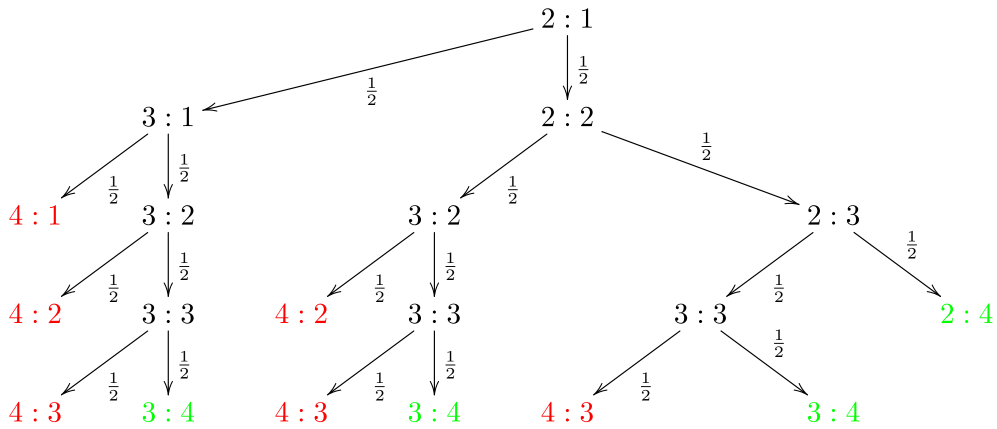
9.) Három különböző kockával dobunk. Mekkora a valószínűsége, hogy az egyik kockával 6-ost dobunk, feltéve, hogy a dobott számok összege 12?
Megoldás. Legyen A: egyikkel 6-ost dobunk; B: az összeg 12.
Írjuk össze az összes lehetséges esetet, amikor 3 kockadobás eredményének az összege 12:

12 felbontása	Esetek száma	Van-e 6-os
6+5+1	3!=6	igen
6+4+2	3!=6	igen
6+3+3	$\frac{3!}{2!}=3$	igen
5+5+2	$\frac{3!}{2!}=3$	nem
5+4+3	3!=6	nem
4+4+4	1	nem
Összesen	25	

Tehát a jó esetek száma: 6+6+3=15, az összes eset száma pedig 25, így a keresett P(A|B) valószínűség 0,6.

10.) 100 érme közül az egyik hamis (ennek mindkét oldalán fej van). Egy érmét kiválasztva és azzal 10-szer dobva, 10 fejet kaptunk. Ezen feltétellel mi a valószínűsége, hogy a hamis érmével dobtunk?
Megoldás. Legyen A: 10 dobásból 10 fej; B_1: jó érmével dobtunk; B_2: hamis érmével dobtunk.
$P(B_1)=\frac{99}{100}$ $\quad P(A|B_1)=\binom{10}{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^{0}=\frac{1}{2^{10}}$
$P(B_2)=\frac{1}{100}$ $\quad P(A|B_2)=1$
Alkalmazzuk a Bayes-tételt: $P(B_2|A)=\frac{P(A|B_2)P(B_2)}{P(A|B_1)P(B_1)+P(A|B_2)P(B_2)}=\frac{1\cdot\frac{1}{100}}{\frac{1}{1024}\cdot\frac{99}{100}+1\cdot\frac{1}{100}}$

11.) Osztozkodási probléma: hogyan osztozzon a téten két játékos, ha 2:1 állásnál félbeszakadt a 4 győzelemig tartó mérkőzésük? (Tfh. az egyes játékok egymástól függetlenek, bármelyikük 1/2 valószínűséggel nyerhet az egyes játékoknál.)
Megoldás. A játék menetét gráffal is lehet ábrázolni. Piros jelöli azt az állást, amikor az első játékos nyer, és zöld, amikor a második. Akkor osztozkodnak "igazságosan", ha a tét annyiad részét kapja az adott játékos, amennyi a nyerési esélye.



Mivel az egyes mérkőzéseket egymástól függetlenül játsszák le, ezért
P(a második játékos nyer)$=1\cdot\frac{1}{8}+3\cdot\frac{1}{16}=\frac{5}{16}$.
Tehát úgy ossza fel a két játékos a tétet, hogy az első játékos kapja a tét $\frac{11}{16}$ részét, a második pedig a tét $\frac{5}{16}$ részét.

12.) Adjuk meg annak a valószínűségi változónak az eloszlását, ami egy hatgyermekes családban a fiúk számát adja meg. Tegyük fel, hogy mindig $\frac{1}{2}$-$\frac{1}{2}$ a fiúk, ill. a lányok születési valószínűsége, és az egyes születések függetlenek egymástól.
Megoldás. Legyen X: fiúk száma. A feladat visszatevéses mintavételként kezelhető: $p=\frac{1}{2}$; a minta mérete 6 $\rightsquigarrow$ n=6.
Így $P(X=k)=\binom{6}{k}\cdot\left(\frac{1}{2}\right)^{k}\cdot\left(\frac{1}{2}\right)^{6-k}=\binom{6}{k}\cdot\left(\frac{1}{2}\right)^{6}$.

13.) Jelölje p_k annak a valószínűségét, hogy egy lottóhúzásnál (90/5) a legnagyobb kihúzott szám k. Számítsd ki a p_k értékeket, és mutassuk meg, hogy ez valóban valószínűségi eloszlás!
Megoldás. $p_k=\frac{\binom{k}{5}-\binom{k-1}{5}}{\binom{90}{5}}$, k=5,6,...,90
ugyanis ki kell választanunk 5 számot az első k-ból, viszont nem k lesz a legnagyobb, amennyiben az első $k-1$- ből választottuk ki őket, így ezeket a rossz eseteket le kell vonni.
Ez valószínűségi eloszlás, ugyanis
$$\sum_{k=5}^{90}p_k=\frac{\binom{5}{5}+(\binom{6}{5}-\binom{5}{5})+(\binom{7}{5}-\binom{6}{5})+\ldots+(\binom{90}{5}-\binom{89}{5})}{\binom{90}{5}}=\frac{\binom{90}{5}}{\binom{90}{5}}=1.$$

14.) Legyenek A, B, C, D, egy szabályos tetraéder csúcsai. Egy légy az A csúcsból indulva sétál a tetraéder élein, mégpedig minden csúcsból véletlenszerűen választva a lehetséges három irány közül. Jelölje X azt a valószínűségi változót, hogy A-ból indulva, hányadikra érünk vissza először A-ba. Írjuk fel X eloszlását! Mutassuk meg, hogy ez valóban valószínűségi eloszlás!

Megoldás. Írjuk fel a megoldást a valószínűség klasszikus képlete alapján: $P(X=k)=\frac{3\cdot 2^{k-2}\cdot 1}{3^k}=\frac{1}{3}\left(\frac{2}{3}\right)^{k-2}$ (k=2,3,...) , ugyanis

- legalább 2 lépésre van szükség, hogy visszaérjünk A-ba
- minden lépésben összesen 3 irányba haladhatunk, így az összes eset 3^k
- jó lépések: elsőként 3 helyre mehetünk, utána $(k-2)$ alkalommal 2 helyre, végül vissza kell lépni A-ba

Ez valószínűségi eloszlás, mivel

$\sum\limits_{k=2}^{\infty} P(X=k)=\frac{1}{3}\sum\limits_{k=2}^{\infty}\left(\frac{2}{3}\right)^{k-2}=\frac{1}{3}\sum\limits_{n=0}^{\infty}\left(\frac{2}{3}\right)^{n}=\frac{1}{3}\frac{1}{1-\frac{2}{3}}=1.$

15.) Egy tétova hangya a számegyenesen bolyong. 0-ból indul és minden lépésnél egyforma valószínűséggel vagy jobbra, vagy balra lép. Mennyi a valószínűsége, hogy $2n$ lépés után a hangya k-ban lesz?

Megoldás. Legyen X: hol lesz a hangya 2n lépés után

$P(X=k)=\frac{\binom{2n}{n+k/2}}{2^{2n}} \qquad k=0,\pm 2,\pm 4,...,\pm 2n$

mivel

- páros sok lépés után csak páros helyeken lehet, viszont $\pm 2n$-en túlra nem tud eljutni
- minden lépésben 2 irányba mehet, ezért az összes lépések száma $2n$ lépés után 2^{2n}
- 0-ból úgy tud eljutni k-ba, hogy k alkalommal biztosan jobbra ment, és a maradék $(2n-k)$-ból pedig a felét jobbra, a felét balra tette meg. Tehát összesen $k+\frac{2n-k}{2}=n+\frac{k}{2}$ alkalommal ment jobbra. Ebből adódik, hogy a jó esetek száma $\binom{2n}{n+k/2}$, mivel elég kiválasztani azokat a helyeket, ahol jobbra megy, a többi helyen már csak balra mehet.

16.) Egy sorsjátékon 1 darab 1 000 000Ft-os, 10 db 100 000Ft-os, és 100 db 1 000Ft-os nyeremény van. A játékhoz 10 000 db sorsjegyet adtak ki. Mennyi a sorsjegy ára, ha egy sorsjegyre a nyeremény várható értéke megegyezik a sorsjegy árával?

Megoldás. Legyen X a nyereményünk, ami most 4 értéket vehet fel különböző valószínűségekkel, amit a következő táblázat foglal össze:

x_i (Ft)	Darab	p_i
1.000.000	1	$\frac{1}{10.000}$
100.000	10	$\frac{1}{1000}$
1.000	100	$\frac{1}{100}$
0	9889	$\frac{9889}{100}$

EX=$1.000.000\cdot\frac{1}{10.000}+100.000\cdot\frac{1}{1000}+1.000\cdot\frac{1}{100}$=100+100+10=210 Ft.

17.) Jelölje X az ötöslottón kihúzott lottószámoknál a párosak számát. Adjuk meg X várható értékét.

Megoldás. Ekkor visszatevés nélküli mintavételezésről van szó, 90 számból 45 páros van, és 5 elemű mintát veszünk. Tehát X $\sim$ Hipgeo(90,45,5), így EX=$5\cdot\frac{45}{90}=2,5.$

18.) Két kockával dobunk. Egy ilyen dobást sikeresnek nevezünk, ha van 6-os a kapott számok között. Várhatóan hány sikeres dobásunk lesz n próbálkozásból?

Megoldás. P(sikeres dobás)=$\frac{11}{36}$

Legyen X: n-ből a sikeres dobások száma

Ekkor nyilvánvaló, hogy X $\sim$ Bin(n,$\frac{11}{36}$)

EX=$\frac{11n}{36}$

19.) Háromszor olyan valószínű, hogy egy évben két ember öli magát a Dunába, mint az, hogy öt. Mi a valószínűsége, hogy egy évben legfeljebb egy ember lesz így öngyilkos?

Megoldás. Legyen X: egy év alatt hányan ölik magukat a Dunába. Az öngyilkosságok tipikusan ritka eseménynek tekinthetők, így $X \sim$Poi(λ).

A feladat első mondata alapján $P(X=2)=3\cdot P(X=5) \quad\Rightarrow$

$\Rightarrow \quad \frac{\lambda^2}{2}e^{-\lambda}=3\frac{\lambda^5}{5!}e^{-\lambda} \quad\Rightarrow\quad \lambda^3=\frac{5!}{2\cdot 3} \quad\Rightarrow\quad \lambda=\sqrt[3]{20}$

P(legfeljebb egy ember lesz öngyilkos a Dunába) $=P(X\leq 1)=$

$=P(X=0)+P(X=1)=e^{-\sqrt[3]{20}}(1+\sqrt[3]{20}).$

20.) Egy 200 oldalas könyvben 20 sajtóliba található véletlenszerűen elszórva.

a.) Mennyi a valószínűsége, hogy a 100. oldalon több, mint egy ajtóhiba van?

b.) Hány sajtóhuba a legvalószínűbb a 100. oldalon?

c.) Mennyi a valószínűsége, hogy a 13. és a 14. oldalon együtt több, mint két hajtóhiba van?

Megoldás. Jelölje X_i valószínűségi változó a könyv i. oldalán a sajtóhibák számát. Ekkor X_i-k függetlenek, és mivel egy könyvben a sajtóhibák rendszerint ritkának mondhatók, ezekről feltehető, hogy Poisson eloszlásúak. Ha a teljes könyvben 20 hiba van, akkor egy oldalon átlagosan $\frac{20}{200} = 0,1$ sajtóhiba szerepel, így a Poisson eloszlás várható értékének képlete alapján $\lambda = 0,1$.

a.) $P(X_{100} > 1) = 1 - P(X_{100} = 0) - P(X_{100} = 1) = 1 - e^{-0,1} - 0,1e^{-0,1} = 1 - 1,1e^{-0,1} \approx 0,5\%$

b.) $p_k = P(X_{100} = k)$-t kell maximalizálni k szerint. Nézzük a $\frac{p_{k+1}}{p_k}$ hányadosokat. Egy növekvő sorozatnál ez nagyobb 1-nél, egy csökkenőnél pedig kisebb 1-nél. Keressük meg azt a pontot, ahol ez egyenlő 1-gyel, aztán vizsgáljuk meg, hogy mely egész értékek között van és ezek közül melyiknél veszi fel a maximumot.

$\frac{p_{k+1}}{p_k} = \frac{\frac{0,1^{k+1}}{(k+1)!}e^{-0,1}}{\frac{0,1^k}{k!}e^{-0,1}} = \frac{0,1}{k+1} = 1 \quad \Rightarrow \quad 0,1 = k+1 \quad \Rightarrow \quad k = -0,9$

A maximum vagy -1-ben, vagy 0-ban van, de mivel X_{100} nem vesz fel pozitív valószínűséggel -1 értéket, ezért 0-ban van a maximum.

c.) $P(X_{13} + X_{14} > 2) =$
$1 - P(X_{13} + X_{14} = 0) - P(X_{13} + X_{14} = 1) - P(X_{13} + X_{14} = 2) =$
$= 1 - P(X_{13} = 0)P(X_{14} = 0) - P(X_{13} = 1)P(X_{14} = 0) -$
$P(X_{13} = 0)P(X_{14} = 1) - P(X_{13} = 2)P(X_{14} = 0) -$
$P(X_{13} = 0)P(X_{14} = 2) - P(X_{13} = 1)P(X_{14} = 1) =$
$= 1 - e^{-0,1}e^{-0,1} - 0,1e^{-0,1}e^{-0,1} - e^{-0,1}0,1e^{-0,1} - \frac{0,1^2}{2}e^{-0,1}e^{-0,1} -$
$e^{-0,1}\frac{0,1^2}{2}e^{-0,1} - 0,1e^{-0,1} \cdot 0,1e^{-0,1} =$
$= 1 - e^{-0,2}(1 + 0,1 + 0,1 + 0,005 + 0,005 + 0,01) =$
$= 1 - 1,22e^{-0,2} \approx 0,1\%$

21.) 5-ször dobunk egy szabályos kockával. Legyen X a 6-osok száma. $D^2(X)$=?

Megoldás. Visszatevéses mintavételről van szó, tehát X$\sim$Bin(5,$\frac{1}{6}$). Így $D^2X = 5 \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{25}{36}$.

22.) Az Y és X valószínűségi változók együttes eloszlását a következő táblázat mutatja.

$X \backslash Y$	1	2	3	X peremeloszlása
5	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	
10	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	
Y peremeloszlása				

a.) Határozd meg X és Y eloszlását, várható értékét, szórásnégyzetét!

b.) X és Y függetlenek egymástól? Amennyiben nem, határozd meg a korrelációjukat!

c.) $P(X < 3|Y < 3) =$?

d.) $E(Y|X = 10) =$?

Megoldás.

$X \backslash Y$	1	2	3	X peremeloszlása
5	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{6}{10}$
10	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{4}{10}$
Y peremeloszlása	$\frac{2}{10}$	$\frac{4}{10}$	$\frac{4}{10}$	1

a.) $EX = 5 \cdot \frac{6}{10} + 10 \cdot \frac{4}{10} = 7$
$EX^2 = 25 \cdot \frac{6}{10} + 100 \cdot \frac{4}{10} = 55$
$D^2X = 55 - 49 = 6 \quad \Rightarrow \quad DX = \sqrt{6}$
$EY = 1 \cdot \frac{2}{10} + 2 \cdot \frac{4}{10} + 3 \cdot \frac{4}{10} = \frac{11}{5}$
$EY^2 = 1 \cdot \frac{2}{10} + 4 \cdot \frac{4}{10} + 9 \cdot \frac{4}{10} = \frac{27}{5}$
$D^2Y = \frac{27}{5} - \frac{121}{25} = \frac{14}{25} \quad \Rightarrow \quad DY = \frac{\sqrt{14}}{5}$

b.) Nem függetlenek egymástól, ugyanis például
$P(X = 5, Y = 1) = \frac{1}{10} \neq \frac{6}{10} \cdot \frac{2}{10} = P(X = 5) \cdot P(Y = 1)$
$E(XY) = 5 \cdot 1 \cdot \frac{1}{10} + 5 \cdot 2 \cdot \frac{2}{10} + \ldots + 10 \cdot 3 \cdot \frac{1}{10} = \frac{5+20+45+10+40+30}{10} = 15$
$\mathrm{Cov}(X, Y) = 15 - 7 \cdot \frac{11}{5} = -\frac{2}{5}$
$R(X, Y) = \frac{-\frac{2}{5}}{\sqrt{6} \cdot \frac{\sqrt{14}}{5}} = -\frac{1}{\sqrt{21}} \quad \Rightarrow \quad R^2 = \frac{1}{21} \approx 0,048$
tehát gyenge negatív kapcsolat van X és Y között.

c.) $P(X < 3|Y < 3) = \frac{P(X<3, Y<3)}{P(Y<3)} = \frac{\frac{3}{10}}{\frac{6}{10}} = \frac{3}{6} = \frac{1}{2}$

d.) $P(Y = 1|X = 10) = \frac{\frac{1}{10}}{\frac{4}{10}} = \frac{1}{4}$
$P(Y = 2|X = 10) = \frac{1}{2}$
$P(Y = 3|X = 10) = \frac{1}{4}$
$E(Y|X = 10) = 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} = 2$

23.) Legyen X és Y független, azonos eloszlású. Tegyük fel azt is, hogy véges szórásúak. $R(X, aX + bY) =$?

Megoldás.
$R(X, aX + bY) = \frac{\mathrm{Cov}(X, aX+bY)}{DX D(aX+bY)} = \frac{\mathrm{Cov}(X,aX)+\mathrm{Cov}(X,bY)}{DX\sqrt{D^2(aX)+D^2(bY)+2\mathrm{Cov}(aX,bY)}} =$
$= \frac{aD^2X + bCov(X,Y)}{DX\sqrt{a^2D^2X+b^2D^2Y+2ab\mathrm{Cov}(X,Y)}} = \frac{aD^2X+0}{DX\sqrt{a^2D^2X+b^2D^2X+0}} = \frac{aD^2X}{DX\sqrt{a^2+b^2}DX} =$
$= \frac{a}{a^2+b^2}$.

24.) Írd fel és ábrázold az eloszlásfüggvényt, ha X

a.) indikátorváltozó $p = 1/2$ paraméterrel;

b.) egy olyan kockadobás eredménye, ahol a kockán egy 2-es, két 4-es és három 5-ös van.

Megoldás.

a.) $F(x) = \begin{cases} 0 & x \le 0 \\ \frac{1}{2} & 0 < x \le 1 \\ 1 & x > 1 \end{cases}$

b.) $F(x) = \begin{cases} 0 & x \le 2 \\ \frac{1}{6} & 2 < x \le 4 \\ \frac{1+2}{6} & 4 < x \le 5 \\ 1 & x > 5 \end{cases}$

25.) Mely c-re lesz eloszlásfüggvény $F(x) = \begin{cases} 0 & \text{ha } x \le 0 \\ cx^3 & \text{ha } 0 < x \le 3 \\ 1 & \text{ha } 3 < x \end{cases}$

$P(-1 < X < 1) =?$ Határozd meg a sűrűségfüggvényét!

Megoldás. F(x)-nek monoton növőnek kell lennie, ami csak akkor teljesül, ha $c \ge 0$. További korlátozást jelent c értékére, hogy az eloszlásfüggvény maximum 1 lehet, amit az x=3-ban vesz fel a középső tartományon:

$1 \ge \max\limits_{x \in (0,3]} cx^3 = c \cdot 3^3 = 27c \Rightarrow c \le \frac{1}{27}$

P(-1<X<1)=P(-∞<X<1)=F(1)=c

$c = \frac{1}{27}$ esetén van csak sűrűségfüggvény:

$f(x) = \begin{cases} \frac{1}{9}x^2 & \text{ha } 0 < x < 3 \\ 0 & \text{különben} \end{cases}$

26.) Legyen X sűrűségfüggvénye a következő:

$f(x) = \begin{cases} cx^4 & \text{ha } 0 < x < 1 \\ 0 & \text{különben} \end{cases}$

a.) Határozd meg a c értékét és X eloszlásfüggvényét!

b.) $P(X < -0.5) =?$ $P(X < 0.5) =?$ $P(X < 1.5) =?$

c.) $D^2(X) =?$

Megoldás.

a.) $1 = \int\limits_0^1 cx^4\,dx = c\left[\frac{x^5}{5}\right]_0^1 = c\frac{1}{5} \Rightarrow c = 5$

b.) $F(x) = P(X < x) = \begin{cases} 0 & \text{ha } x \le 0 \\ \int\limits_0^x 5t^4\,dt = x^5 & \text{ha } 0 < x \le 1 \\ 1 & \text{ha } x > 1 \end{cases}$

c.) $P(X < -0.5) = F(-0,5) = 0$
$P(X < 0.5) = F(0,5) = 0,5^5$
$P(X < 1.5) = F(1,5) = 1$

d.) $EX = \int\limits_0^1 5x^5\,dx = 5\frac{1}{6} = \frac{5}{6}$

$EX^2 = \int\limits_0^1 5x^6\,dx = 5\frac{1}{7} = \frac{5}{7}$

$D^2X = \frac{5}{7} - \frac{25}{36} = \frac{5}{252}.$

27.) Legyen X sűrűségfüggvénye a következő: $f(x) = \begin{cases} \frac{c}{x^4} & \text{ha } x > 1 \\ 0 & \text{különben} \end{cases}$

a.) $c =?$, $F(x) =?$

b.) $P(X < 2) =?$, $P(X > 3) =?$

c.) $E(X) =?$

d.) $D^2(X) =?$

Megoldás.

a.) $1 = \int\limits_1^\infty \frac{c}{x^4}\,dx = c\left[\frac{x^{-3}}{-3}\right]_1^\infty = -\frac{c}{3}(0-1) = \frac{c}{3} \Rightarrow c = 3$

$F(x) = P(X < x) = \begin{cases} 0 & \text{ha } x \le 1 \\ \int\limits_1^x \frac{3}{t^4}\,dt = \frac{3}{-3}[t^{-3}]_1^x = -\left(\frac{1}{x^3} - 1\right) = 1 - \frac{1}{x^3} & \text{ha } x > 1 \end{cases}$

b.) $P(X < 2) = F(2) = 1 - \frac{1}{2^3} = \frac{7}{8}$
$P(X > 3) = 1 - F(3) = \frac{1}{3^3} = \frac{1}{27}$

c.) $EX = \int\limits_1^\infty \frac{3}{x^3}\,dx = 3\left[\frac{x^{-2}}{-2}\right]_1^\infty = \frac{3}{2}(1-0) = \frac{3}{2}$

d.) $EX^2 = \int\limits_1^\infty \frac{3}{x^2}\,dx = 3\left[\frac{x^{-1}}{-1}\right]_1^\infty = 3(1-0) = 3$

$D^2X = 3 - \frac{9}{4} = \frac{3}{4}$

28.) Legyen X sűrűségfüggvénye a következő:

$f(x) = \begin{cases} \frac{x}{3} & \text{ha } 0 < x < 2 \\ \frac{1}{6} & \text{ha } 2 < x < c \\ 0 & \text{különben} \end{cases}$

a.) $c=?$ $F(x)=?$
b.) $E(X)=?$ $D(X)=?$
Megoldás.
a.) $1=\int\limits_0^2 \frac{x}{3}\,dx+\frac{c-2}{6}=\frac{2}{3}+\frac{c-2}{6} \quad \Rightarrow \quad c=4$

$$F(x)=\begin{cases} 0 & x\le 0 \\ \int\limits_0^x \frac{t}{3}\,dt=\frac{x^2}{6} & 0<x\le 2 \\ \frac{2}{3}+\frac{x-2}{6}=\frac{x+2}{6} & 2<x\le 4 \\ 1 & x>4 \end{cases}$$

b.) $E(X)=\int\limits_0^2 \frac{t^2}{3}\,dt+\int\limits_2^4 \frac{t}{6}\,dt=\frac{17}{9}$

$E(X^2)=\int\limits_0^2 \frac{t^3}{3}\,dt+\int\limits_2^4 \frac{t^2}{6}\,dt=\frac{40}{9}$

$DX=\frac{\sqrt{71}}{9}$

29.) Tegyük fel, hogy az egyetemisták IQ teszten elért eredménye normális eloszlású 105 várható értékkel és 10 szórással. Mi a valószínűsége, hogy valaki 120-nál több pontot ér el a teszten?
Megoldás. Legyen X: egy egyetemista IQ-pontja $\quad\Rightarrow\quad X\sim N(105,10^2)$
$P(X>120)=1-P(X<120)=1-P\left(\frac{X-105}{10}<\frac{120-105}{10}\right)=$
$=1-\Phi(1,5)=1-0,9332=6,68\%$.

30.) Mennyi garanciát adjunk, ha azt szeretnénk, hogy termékeink legfeljebb 10%-át kelljen garanciaidőn belül javítani, ha a készülék élettartama 10 év várható értékű és 2 év szórású normális eloszlással közelíthető?
Megoldás. Legyen X: a termékek élettartama $\quad\Rightarrow\quad X\sim N(10,2^2)$
Jelölje a garanciaidőt t
A feladat szövege alapján $0,1\ge P(X<t)$
$P(X<t)=P\left(\frac{X-10}{2}<\frac{t-10}{2}\right)=\Phi\left(\frac{t-10}{2}\right)$
Átrendezve t-re: $t\le 2\cdot\Phi^{-1}(0,1)+10=2\cdot(-\Phi^{-1}(0,9))+10=$
$=-2\cdot 1,28+10=7,44$
Tehát legfeljebb 7 év garanciát kell adnunk (ha a garanciaidő csak egész szám lehet).

31.) U és V valószínűségi változókról a következőket tudjuk: $R(U,V)=-0,75$; $EU=4$; $EV=6$; $D(U)=D(V)=\frac{1}{\sqrt{2}}$. Becsüld alulról a $P(8<U+V<12)$ valószínűséget!
Megoldás. $X:=U+V$
Ekkor $EX=EU+EV=4+6=10$
$\mathrm{Cov}(U,V)=R(U,V)\cdot DU\cdot DV=-\frac{3}{4}\cdot\frac{1}{2}=-\frac{3}{8}$
$D^2X=D^2U+D^2V+2\mathrm{Cov}(U,V)=\frac{1}{2}+\frac{1}{2}+2\cdot-\frac{3}{8}=\frac{1}{4}$
$P(8<X<12)=P(-2<X-10<2)=P(|X-10|<2)=$
$=1-P(|X-10|\ge \underbrace{2}_{\varepsilon=})\ge 1-\frac{D^2X}{\varepsilon^2}=1-\frac{\frac{1}{4}}{2^2}=1-\frac{1}{16}=\frac{15}{16}$.

32.) Hamis érmével dobunk, a fej valószínűsége 0,51.
a.) Becsüljük meg a Csebisev-egyenlőtlenséggel, majd a centrális határértéktétel segítségével is annak a valószínűségét, hogy 10 ezer dobásból legalább 5150 fej!
b.) Hányszor kell dobni, hogy a fejek relatív gyakorisága legalább 97,5 %-os valószínűséggel több legyen, mint 0,505?

Megoldás.
a.) Legyen X_i valószínűségi változó, ami 1 értéket vesz fel, ha fejet dobunk, és 0-t, amikor írást. Ekkor nyilvánvalóan $X_i\sim\mathrm{Ind}(0,51)$
Legyen Y: mennyi fejet kaptunk 10000 dobásból, azaz
$Y=\sum\limits_{i=1}^{10000} X_i\sim\mathrm{Bin}(10000;0,51)$.
Ekkor $EY=5100$ és $D^2Y=2499$.
A becsülendő valószínűség: $P(Y\ge 5150)$
Csebisev-egyenlőtlenséggel:
$P(Y-5100\ge 50)\le P(|Y-5100|\ge 50)\le\frac{D^2Y}{50^2}=\frac{2499}{2500}$
Használjuk a centrális határeloszlás-tételt:
$P\left(\frac{Y-5100}{\sqrt{2499}}\ge\frac{50}{\sqrt{2499}}\right)\approx 1-\Phi\left(\frac{50}{\sqrt{2499}}\right)\approx 1-\Phi(1)=0,1587$

b.) Legyen $\overline{X}_n=\frac{\sum\limits_{i=1}^{n} X_i}{n}$ a fejek relatív gyakorisága, ekkor $E\overline{X}_n=0,51$ és $D^2\overline{X}_n=\frac{0,51\cdot 0,49}{n}$
A becsülendő valószínűség: $P(\overline{X}_n>0,505)$

Használjuk a Csebisev-egyenlőtlenséget:
$P(\overline{X}_n>0,505)=P(\overline{X}_n-0,51>-0,005)\ge$
$\ge P(|\overline{X}_n-0,51|\le 0,005)\ge 1-\frac{D^2\overline{X}_n}{0,005^2}=1-\frac{0,51\cdot 0,49}{n}\cdot 200^2=1-\frac{9996}{n}$
A feladat szövege alapján ennek legalább 0,975-nek kell lennie:
$1-\frac{9996}{n}\ge 0,975=\frac{39}{40}$
ezt megoldva, $n\ge 399840$ jön ki: legalább 399840-szer kell dobnunk.

Centrális határérték-tétellel:

$P(\overline{X}_n > 0,505) = P\left(\frac{\overline{X}_n - 0,51}{\sqrt{\frac{0,2499}{n}}} > \frac{-0,005}{\sqrt{\frac{0,2499}{n}}}\right) \approx 1 - \Phi\left(-\frac{0,005}{\sqrt{\frac{0,2499}{n}}}\right) =$

$= \Phi\left(\frac{\sqrt{n}}{2\sqrt{2499}}\right)$

A feladat szövege alapján ennek legalább 0,975-nek kell lennie:

$\Phi\left(\frac{\sqrt{n}}{2\sqrt{2499}}\right) \geq 0,975 \quad \Rightarrow \quad \sqrt{n} \geq \Phi^{-1}(0,975) \cdot 2\sqrt{2499} = 195,96$

tehát $n \geq 38399,2$ adódik, ezáltal legalább 38400-szor kell dobnunk.

33.)

a.) Legyenek X_i ~Ind(p) $(i = 1, 2, \ldots)$ val. változók. Mihez konvergál $\frac{X_1^5+\ldots+X_n^5}{n}$?

b.) X_i jelölje az i-edik kockadobás eredményét. Mihez konvergál $\frac{X_1^2+\ldots+X_n^2}{n}$?

Megoldás.

a.) A nagy számok erős törvénye szerint $n \to \infty$ esetén

$\frac{X_1^5+\ldots+X_n^5}{n} \overset{m.m.}{\to} EX_1^5 = 1^5 \cdot p + 0^5 \cdot (1-p) = p$

b.) X_i-k közös eloszlása: $P(X_i = k) = \frac{1}{6}$, $k = 1, 2, ..., 6$

A nagy számok erős törvénye szerint $n \to \infty$ esetén

$\frac{X_1^2+\ldots+X_n^2}{n} \overset{m.m.}{\to} EX_1^2 = \sum_{k=1}^{6} k^2 \cdot \frac{1}{6} = \frac{6\cdot7\cdot13}{6}\frac{1}{6} = \frac{91}{6}$

34.) Tegyük fel, hogy egy tábla csokoládé tömege normális eloszlású 100 g várható értékkel és 3 g szórással, valamint, hogy az egyes táblák tömege egymástól független. Legalább hány csokoládét csomagoljunk egy dobozba, hogy a dobozban levő táblák átlagos tömege legalább 0,9 valószínűséggel nagyobb legyen 99,5 g-nál?

Megoldás. Legyen X_i: az i. tábla tömege $\quad \Rightarrow \quad X_i \sim N(100, 3^2)$

Átlagos tömeg: $\overline{X} = \frac{X_1+\ldots+X_n}{n} \sim N\left(100, (\frac{3}{\sqrt{n}})^2\right)$, ugyanis

- $E\overline{X} = \frac{E\sum_{i=1}^{n} X_i}{n} = \frac{nEX_1}{n} = EX_1 = 100$
- $D^2\overline{X} = D^2\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} D^2X_i = \frac{D^2X_1}{n} = \frac{9}{n}$

A feladat szövege alapján $0,9 < P(\overline{X} > 99,5)$

$P(\overline{X} > 99,5) = 1 - P(\overline{X} < 99,5) =$

$= 1 - P\left(\frac{\overline{X}-100}{\frac{3}{\sqrt{n}}} < \frac{99,5-100}{\frac{3}{\sqrt{n}}} = -\frac{\sqrt{n}}{6}\right) = 1 - \Phi\left(-\frac{\sqrt{n}}{6}\right) =$

$= 1 - \left[1 - \Phi\left(\frac{\sqrt{n}}{6}\right)\right] = \Phi\left(\frac{\sqrt{n}}{6}\right)$

Átrendezve n-re: $n > [6\Phi^{-1}(0,9)]^2 = [6 \cdot 1,28]^2 = 58,9$

Tehát legalább 59 csokit kell becsomagolni a dobozba.

35.) Legyen X ~Exp(λ). Határozd meg X móduszát és tetszőleges kvantilisét! Hasonlítsd össze a mediánt és a várható értéket!

Megoldás. $Mo = 0$, mivel a sűrűségfüggvény csökkenő.

$y = F(x) = 1 - e^{-\lambda x}$ -et x-re rendezve, megkapjuk a kvantilisfüggvényt:

$F^{-1}(y) = -\frac{1}{\lambda}\log(1-y)$.

$Me = F^{-1}\left(\frac{1}{2}\right) = -\frac{1}{\lambda}\log\left(\frac{1}{2}\right) = \frac{1}{\lambda}\cdot\log 2$, ez pedig nagyobb a várható értéknél, ami $\frac{1}{\lambda}$.

36.) Legyen X ~Ind(p). Határozd meg X móduszát és kvantilisfüggvényét!

Megoldás.

$$Mo = \begin{cases} 1 & \text{ha } p > \frac{1}{2} \\ 0 & \text{ha } p < \frac{1}{2} \\ \{0, 1\} & \text{ha } p = \frac{1}{2} \end{cases}$$

$$F^{-1}(y) = \begin{cases} 0 & \text{ha } y \leq p \\ 1 & \text{ha } y > p \end{cases}$$

37.) Legyen $X_1, ..., X_{20}$ i.i.d. minta $N(m, 1^2)$ eloszlásból. Célunk az ismeretlen m paraméter becslése. Tekintsük az alábbi három statisztikát:

- $T_1(\mathbf{X}) = X_8$,
- $T_2(\mathbf{X}) = \frac{X_3+X_7}{2}$,
- $T_3(\mathbf{X}) = \frac{X_9+X_{19}}{8}$.

a.) A fenti statisztikák közül melyek torzítatlanok? Amelyik nem torzítatlan, hogyan tudnánk torzítatlanná tenni?

b.) Vizsgáljuk meg a fenti statisztikák közül a torzítatlanokat hatásosság szempontjából!

Megoldás.

a.) $E_mT_1(\mathbf{X}) = E_mX_8 = m$

$E_mT_2(\mathbf{X}) = E_m\frac{X_3+X_7}{2} = m$

$E_mT_3(\mathbf{X}) = E_m\frac{X_9+X_{19}}{8} = \frac{m}{4}$

Tehát T_1 és T_2 torzítatlanul becsüli m-et, míg T_3 nem, azonban torzítatlanná tehető: $T_3' = 4T_3$ már torzítatlan becslés.

b.) Csak T_1 és T_2 vizsgálható hatásosság szempontjából.

$D_m^2T_1(\mathbf{X}) = D_m^2X_8 = 1$

$D^2_m T_2(\mathbf{X}) = D^2_m \frac{X_3+X_7}{2} = \frac{1}{4}\cdot[D^2_m X_3 + D^2_m X_7 + 2\underbrace{cov_m(X_3,X_7)}_{0}] =$

$= \frac{1}{4}\cdot(m+m) = \frac{1}{2}$

Tehát mivel minden m paraméterérték esetén $D^2_m T_2(\mathbf{X}) \le D^2_m T_1(\mathbf{X})$ teljesül, ezért T_2 hatásosabb becslés T_1 becslésnél.

38.) n elemű λ-paraméterű exponenciális minta esetén adjunk torzítatlan becslést $e^{-3\lambda}$-ra és $\frac{1}{\lambda}$-ra!

Megoldás. Ha $X_i \sim \text{Exp}(\lambda)$, akkor $F(x) = (1-e^{-\lambda x})\mathrm{I}(x\ge 0)$ és $EX_i = \frac{1}{\lambda}$. Nézzük az eloszlásfüggvényt az 3 helyen: $F(3) = 1-e^{-3\lambda} \quad\Rightarrow\quad 1-F(3) = e^{-3\lambda}$, és épp ezt akarjuk torzítatlanul becsülni, tehát $T_1(\mathbf{X}) := 1-F_n(3)$ torzítatlan becslése $e^{-3\lambda}$-nak. $F_n(3)$ pedig éppen a 3-nál kisebb megfigyelések relatív gyakorisága.

Most térjünk rá $\frac{1}{\lambda}$-ra.

$EX_i = \frac{1}{\lambda}$, és épp ezt akarjuk torzítatlanul becsülni. Tudjuk, hogy a várható értéket a mintaátlag torzítatlanul becslüli, tehát $T_2(\mathbf{X}) := \overline{X}$ torzítatlan becslése $\frac{1}{\lambda}$-nak.

39.) Adjunk meg torzítatlan becslést a $[0,\theta]$ intervallumon egyenletes eloszlás paraméterére

a.) a mintaátlag

b.) a maximum

segítségével. Hasonlítsuk őket össze hatásosság szempontjából! Melyik becslés konzisztens?

Megoldás. $E(0,\theta)$ eloszlás- és sűrűségfüggvénye:

$$F_\theta(x) = \begin{cases} 0 & \text{ha } x\le 0 \\ \frac{x}{\theta} & \text{ha } 0<x\le\theta \\ 1 & \text{ha } \theta< x\end{cases} \qquad f_\theta(x) = \begin{cases} \frac{1}{\theta} & \text{ha } 0<x<\theta \\ 0 & \text{különben}\end{cases}$$

a.) $E\overline{X} = EX_1 = \frac{\theta}{2} \quad\Rightarrow\quad T_1(\mathbf{X}) := 2\overline{X}$ torzítatlan becslése θ-nak

$D^2(T_1(\mathbf{X})) = 4D^2(\overline{X}) = \frac{4D^2X_1}{n} = \frac{4\frac{\theta^2}{12}}{n} = \frac{\theta^2}{3n}$. Látható, hogy a szórásnégyzet a 0-hoz tart, így T_1 konzisztens becslés.

b.) $EX^*_n = \int\limits_0^\theta xnf(x)(F(x))^{n-1}\,dx = \int\limits_0^\theta xn\frac{1}{\theta}\left(\frac{x}{\theta}\right)^{n-1}dx =$

$= \frac{n}{\theta^n}\int\limits_0^\theta x^n\,dx = \frac{n}{\theta^n}\left[\frac{x^{n+1}}{n+1}\right]_0^\theta dx = \frac{n}{\theta^n}\frac{\theta^{n+1}}{n+1} = \frac{\theta n}{n+1}$

$\Rightarrow T_2(\mathbf{X}) := \frac{n+1}{n}X^*_n$ torzítatlan becslése θ-nak

Szükség van a második momentumra is, hogy ki tudjuk számítani a szórásnégyzetet.

$E(X^*_n)^2 = \int\limits_0^\theta x^2nf(x)(F(x))^{n-1}\,dx = \int\limits_0^\theta x^2n\frac{1}{\theta}\left(\frac{x}{\theta}\right)^{n-1}dx =$

$= \frac{n}{\theta^n}\int\limits_0^\theta x^{n+1}\,dx = \frac{n}{\theta^n}\left[\frac{x^{n+2}}{n+2}\right]_0^\theta dx = \frac{n}{\theta^n}\frac{\theta^{n+2}}{n+2} = \frac{\theta^2 n}{n+2}$

$D^2(X^*_n) = \frac{\theta^2 n}{n+2} - \frac{\theta^2n^2}{(n+1)^2} = \theta^2 n\frac{(n+1)^2-n(n+2)}{(n+1)^2(n+2)} =$

$= \theta^2 n\frac{n^2+2n+1-n^2-2n}{(n+1)^2(n+2)} = \frac{\theta^2 n}{(n+1)^2(n+2)}$

Így $D^2(T_2(\mathbf{X})) = \frac{(n+1)^2}{n^2}D^2(X^*_n) = \frac{(n+1)^2}{n^2}\frac{\theta^2 n}{(n+1)^2(n+2)} = \frac{\theta^2}{n(n+2)}$. Látható, hogy a szórásnégyzet a 0-hoz tart, így T_2 is konzisztens becslés.

Mivel $D^2(T_2) \le D^2(T_1)$ minden n esetén, ezért T_2 hatásosabb becslés T_1-nél.

40.) Mutassuk meg, hogy exponenciális eloszlású minta esetén $T(\mathbf{X}) = n\cdot\min(X_1,...,X_n)$ statisztika torzítatlan, de nem konzisztens becslése a várható értéknek.

Megoldás. Először nézzük meg, hogy exponenciálisok minimumának mi az eloszlásfüggvénye:

$F_{X^*_1}(x) = n\cdot\lambda e^{-\lambda x}\mathrm{I}_{(x\ge 0)}\cdot(e^{-\lambda x})^{n-1} = (n\lambda)e^{-(n\lambda)x}\mathrm{I}_{(x\ge0)}$

ebből pedig látható, hogy $X^*_1 \sim \text{Exp}(n\lambda)$

Legyen $\varepsilon>0$ tetszőleges.

$P\left(\left|nX^*_1-\frac{1}{\lambda}\right|\ge\varepsilon\right) = P\left(nX^*_1-\frac{1}{\lambda}\ge\varepsilon\right) + P\left(nX^*_1-\frac{1}{\lambda}\le-\varepsilon\right) =$

$= P\left(X^*_1\ge\frac{1}{n}\left(\frac{1}{\lambda}+\varepsilon\right)\right) + P\left(X^*_1\le\frac{1}{n}\left(\frac{1}{\lambda}-\varepsilon\right)\right) =$

$= e^{-n\lambda\cdot\frac{\frac{1}{\lambda}+\varepsilon}{n}} + 1 - e^{-n\lambda\cdot\frac{\frac{1}{\lambda}-\varepsilon}{n}} = e^{-1-\varepsilon\lambda} + 1 - e^{-1+\varepsilon\lambda}$

ez pedig nem tart 0-hoz, ha $n\to\infty$, így a becslés nem konzisztens.

41.) Egy osztályban a diákok magassága: (cm)

180	163	150
157	165	165
174	191	172
165	168	186

Elemezd a diákok testmagasságát az átlag, a korrigált tapasztalati szórás, szórási együttható és boxplot ábra (kvartilisek) segítségével! Értelmezd is az eredményeket!

Megoldás. $n=12$

$\overline{x} = \frac{180+157+...+186}{12} = 169,7$ cm

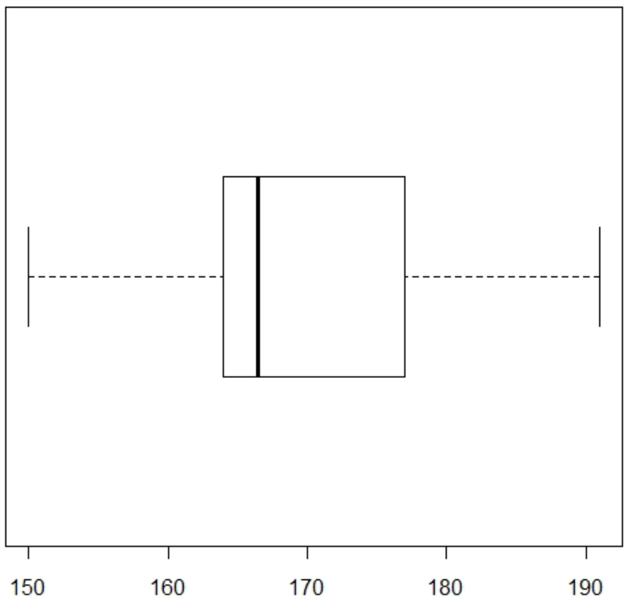
$s_n^* = \sqrt{\frac{(180-169{,}6)^2+(157-169{,}6)^2+...+(186-169{,}6)^2}{11}} = 11,7$ cm
$V = \frac{11{,}7}{169{,}7} = 6,9$ %

Rendezett minta:

150	165	174
157	165	180
163	168	186
165	172	191

Q_1-hez Sorszám: $\frac{13}{4}$=3+0,25
$Q_1 = X_3^* + 0,25(X_4^* - X_3^*) = 163 + 0,25 \cdot (165 - 163) = 163,5$ cm
Me-hoz Sorszám: $\frac{13}{2}$=6+0,5
$Me = X_6^* + 0,5(X_7^* - X_6^*) = 165 + 0,5 \cdot 3 = 166,5$ cm
Q_3-hoz Sorszám: $\frac{13\cdot 3}{4}$=9+0,75
$Q_3 = X_9^* + 0,75(X_{10}^* - X_9^*) = 174 + 0,75 \cdot 6 = 178,5$ cm

Boxplot ábra:



Értelmezések: a diákok átlagos testmagassága 169,6 cm, az egyes testmagasságok az átlagos testmagasságtól átlagosan 11,56 cm-rel, azaz 6,8 %-kal térnek el. A hallgatók negyede 163,5 cm-nél alacsonyabb, míg háromnegyede ennél magasabb. A hallgatók fele 167,5 cm-nél alacsonyabb, másik fele ennél magasabb. A hallgatók negyede 178,5 cm-nél magasabb.

42.) Határozzuk meg az ismeretlen paraméter(ek) maximum likelihood becslését, ha a minta
a.) Pascal (=Geom(p));
b.) Exp(λ);
c.) Poi(λ).

Megoldás.

a.) $P(X_i = x_i) = p(1-p)^{x_i - 1}$

$$L(p;\mathbf{x}) = \prod_{i=1}^{n} p(1-p)^{x_i-1} = p^n(1-p)^{\sum_{i=1}^{n} x_i - n}$$

$$l(p;\mathbf{x}) = n\log p + \left(\sum_{i=1}^{n} x_i - n\right)\log(1-p)$$

Elsőrendű feltétel: $0 = \partial_p l = \frac{n}{p} + \left(\sum_{i=1}^{n} x_i - n\right)\frac{-1}{1-p} \quad \Rightarrow$

$$\Rightarrow \quad n(1-p) = p(\sum_{i=1}^{n} x_i - n) \quad \Rightarrow \quad n = p\sum_{i=1}^{n} x_i \quad \Rightarrow \quad \hat{p} = \frac{1}{\overline{X}}.$$

$$\partial_p^2 l(p;\mathbf{x}) = -\frac{n}{p^2} - \frac{\sum_{i=1}^{n} x_i - n}{(1-p)^2} = -\frac{n}{p^2(1-p)^2}\left((1-p)^2 + p^2(\overline{x}-1)\right) =$$
$$= -\frac{n}{p^2(1-p)^2}\left(1 - 2p + p\overline{x}\right)$$

Másodrendű feltétel: $0 > \partial_p^2 l(\hat{p};\mathbf{x}) = -\frac{n}{\hat{p}^2(1-\hat{p})^2}\left(1 - 2\hat{p} + \hat{p}^2\overline{x}\right) \quad \Leftrightarrow$
$\Leftrightarrow 0 < 1 - 2\frac{1}{\overline{x}} + \frac{1}{\overline{x}^2}\overline{x} = 1 - \frac{1}{\overline{x}} \Leftrightarrow \overline{x} > 1$, és ez az egyenlőtlenség egy kivétellel teljesül: amikor $\overline{x} = 1$, ilyenkor pedig $x_1 = ... = x_n = 1 \quad \Rightarrow$ $L = p^n$, amit a $p = 1$ maximalizál.

b.) $f_{X_i}(x_i) = \lambda e^{-\lambda x_i}\mathrm{I}(x_i \geq 0)$

$$L(\lambda;\mathbf{x}) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}\mathrm{I}(x_i \geq 0) = \lambda^n e^{-\lambda\sum_{i=1}^{n} x_i}\mathrm{I}(x_i \geq 0)$$

$$l(\lambda;\mathbf{x}) = n\log\lambda - \lambda\sum_{i=1}^{n} x_i$$

Elsőrendű feltétel: $0 = \partial_\lambda l = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \quad \Rightarrow \quad \lambda = \frac{n}{\sum_{i=1}^{n} x_i} \quad \Rightarrow \quad \hat{\lambda} = \frac{1}{\overline{X}}$

$\partial_\lambda^2 l = -\frac{n}{\lambda^2}$, ami < 0 minden λ-ra, így teljesül a másodrendű feltétel, azaz $\hat{\lambda}$ maximumhely.

c.) $P(X_i = x_i) = \frac{\lambda^{x_i}}{x_i!}e^{-\lambda}$

$$L(\lambda;\mathbf{x}) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!}e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}e^{-n\lambda}$$

$l(\lambda;\mathbf{x}) = -\sum\limits_{i=1}^{n}\log x_i! + \sum\limits_{i=1}^{n} x_i\cdot\log\lambda - n\lambda$

Elsőrendű feltétel: $0 = \partial_\lambda l = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n \quad \Rightarrow \quad \lambda = \frac{\sum_{i=1}^{n} x_i}{n} \quad \Rightarrow \quad \hat{\lambda} = \overline{X}$

$\partial_\lambda^2 l = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2}$, ami < 0 minden λ-ra, így teljesül a másodrendű feltétel, azaz $\hat{\lambda}$ maximumhely.

43.) Becsüld a paramétert momentum-módszerrel az alábbi esetekben:

a.) $\mathrm{Exp}(\lambda)$;

b.) $\mathrm{Poi}(\lambda)$;

c.) $\mathrm{E}(a,b)$;

d.) $\mathrm{E}(-a,a)$.

Megoldás.

a.) $EX = m_1 \quad \Rightarrow \quad \frac{1}{\lambda} = \overline{x} \quad \Rightarrow \quad \hat{\lambda} = \frac{1}{\overline{x}}$;

b.) $EX = m_1 \quad \Rightarrow \quad \lambda = \overline{x} \quad \Rightarrow \quad \hat{\lambda} = \overline{x}$;

c.) Használjuk az előző feladat eredményét

$$\begin{cases} E_{a,b}X &= m_1 \\ D^2_{a,b}X &= s_n^2 \end{cases} \Rightarrow \begin{cases} \frac{a+b}{2} &= \overline{x} \\ \frac{(b-a)^2}{12} &= s_n^2 \end{cases} \Rightarrow \begin{cases} b+a &= 2\overline{x} \\ b-a &= 2\sqrt{3s_n^2} \end{cases}$$

$\Rightarrow \hat{b} = \overline{X} + \sqrt{3}S_n$ (összeadva egymáshoz a két egyenletet)

$\Rightarrow \hat{a} = \overline{X} - \sqrt{3}S_n$ (kivonva egymásból a két egyenletet);

d.) $EX = m_1 \quad \Rightarrow \quad 0 = \overline{x} \quad \rightsquigarrow$ nem kaptunk semmit a paraméterre, ezért nézzük a következő momentumra

$EX^2 = m_2 \quad \Rightarrow \quad \frac{(2a)^2}{12} + 0 = s_n^2 \quad \Rightarrow \quad \hat{a} = \sqrt{3}s_n$.

44.) Legyen az $X_1, \ldots, X_n$ minta a következő diszkrét eloszlásból: $\mathrm{P}(X_1{=}1){=}c$, $\mathrm{P}(X_1{=}2){=}3c$, $\mathrm{P}(X_1{=}3){=}1\text{-}4c$ (c az ismeretlen paraméter). Tegyük fel, hogy az n mintaelemből y_i darab veszi fel az i értéket (i=1,2,3).

a.) Határozzuk meg c momentum-becslését!

b.) Határozzuk meg c ML-becslését!

Megoldás. Hogy érthető legyen, az ML-módszernél honnan jön a képlet, definiálunk egy új, többváltozós eloszlást, ami a binomiális eloszlás általánosításaként fogható fel.

<u>Def.:</u> $\mathbf{Y} = (Y_1, \ldots, Y_k)$ **polinomiális eloszlású** n renddel és $p_1, \ldots, p_k$ paraméterekkel, ha

- $\sum\limits_{i=1}^{k} p_i = 1$ és
- $P(Y_1 = y_1, \ldots, Y_k = y_k) = \frac{n!}{y_1!\ldots y_k!} p_1^{y_1} \ldots p_k^{y_k}$, amennyiben $0 \leq y_i \in \mathbb{Z}$ és $\sum\limits_{i=1}^{k} y_i = n$.

Jelölés ekkor: $\mathbf{Y} \sim \mathrm{Polin}(n; p_1, \ldots, p_k)$

a.) $\overline{x} = m_1 = \mathrm{E}X = c + 6c + 3(1-4c) = 3 - 5c$.

Ezt átrendezve, $c = \frac{3-\overline{x}}{5}$, azaz $\hat{c} = \frac{3-\overline{X}}{5}$.

b.) Legyen $\underline{Y} = (Y_1, Y_2, Y_3)$, ahol Y_1 az a val. változó, aminél $Y_1 = y_1$ azt jelenti, hogy az X_i-kből y_1 alkalommal kaptuk az 1 értéket. Y_2 és Y_3 hasonlóan értendő. Ekkor $\underline{Y} \sim \mathrm{Polin}(n; c, 3c, 1-4c)$.

A likelihood-függvényt az X_i-k együttes eloszlásából lehet kiszámítani, ami persze megegyezik Y_i-k együttes eloszlásával a feladat szövege alapján, azaz $\underline{X} = \underline{Y}$.

Így $\mathrm{L}(c, \underline{x}) = P(X_1 = x_1, \ldots, X_n = x_n) =$

$= P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) = \frac{n!}{y_1!y_2!y_3!} c^{y_1} (3c)^{y_2} (1-4c)^{y_3}$

$\mathrm{l}(c, \underline{x}) = \log(\frac{n!}{y_1!y_2!y_3!}) + y_1\log(c) + y_2\log(3c) + y_3\log(1-4c)$

Deriválásnál figyelni kell a belső függvények deriváltjaira is:

$\partial_c \mathrm{l}(c, \underline{x}) = y_1\frac{1}{c} + y_2\frac{1}{3c}3 + y_3\frac{1}{1-4c}(-4) = \frac{y_1+y_2}{c} + \frac{-4y_3}{1-4c}$

Ezt 0-val tesszük egyenlővé és átrendezgetünk:

$y_1 + y_2 = 4c(y_1 + y_2 + y_3) = 4cn \longrightarrow c = \frac{y_1+y_2}{4n}$

Hogy a becslést fel tudjuk írni, szükség van y_1 meghatározására az X_i-k segítségével. Nade y_1 az a szám, ahány alkalommal az X_i-ink értéke 1 lett, tehát $y_1 = \sum_{i=1}^{n} I(X_i = 1)$. Ugyanígy az y_2.

Tehát a becslés így írható: $\hat{c} = \frac{\sum_{i=1}^{n}(I(X_i=1)+I(X_i=2))}{4n}$.

45.) Legyen a $Z_1, \ldots, Z_5$ minta $\mathrm{N}(m, 2^2)$ eloszlású. A megfigyelt értékek a következők: 6; 4,5; 2,5; 2; 1.

a.) Határozzunk meg 95%-os (99%-os) megbízhatóságú konfidenciaintervallumot m-re!

b.) Hány elemű mintára van szükségünk 95%-os megbízhatósági szinten, ha azt szeretnénk, hogy a konfidenciaintervallum legfeljebb 0,01 hosszúságú legyen?

c.) Mi változik az a.) esetben, ha a szórást nem ismerjük?

d.) Adjunk a szórásra 98%-os megbízhatóságú konfidenciaintervallumot.

$\chi^2_{4;0{,}01} = 0{,}3 \quad \chi^2_{4;0{,}99} = 13{,}28$

Megoldás. $n = 5 \qquad \overline{x} = 3{,}2 \qquad s_5^* = 2{,}019$

a.) Először $\alpha = 0{,}05$, most ismert a szórás: $\sigma = 2$

$u_{\frac{0{,}05}{2}} = \Phi^{-1}(1 - 0{,}025) = \Phi^{-1}(0{,}975) = 1{,}96$

Konfidencia intervallum: $3,2 \pm 1,96 \cdot \frac{2}{\sqrt{5}} = 3,2 \pm 1,753 = [1,447; 4,953]$
Amennyiben $\alpha = 0,01$, akkor $u_{\frac{0,01}{2}} = \Phi^{-1}(0,995) = 2,58$
Konfidencia intervallum: $3,2 \pm 2,58 \cdot \frac{2}{\sqrt{5}} = 3,2 \pm 2,308 = [0,892; 5,508]$

b.) $\alpha = 0,05$, a konfidencia intervallum hossza $2 \cdot u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$, így a megoldandó egyenlőtlenség a következő:
$0,01 \geq 2 \cdot 1,96 \cdot \frac{2}{\sqrt{n}} \quad \Rightarrow \quad \sqrt{n} \geq 100 \cdot 4 \cdot 1,96 = 784 \quad \Rightarrow \quad n \geq 614656.$

c.) $t_{5-1;\frac{0,05}{2}} = t_{4;0,025} = 2,776$
Konfidencia intervallum: $3,2 \pm 2,776 \frac{2,019}{\sqrt{5}} = 3,2 \pm 2,507 = [0,693; 5,707]$

d.) $\alpha = 0,05$, a szóráshoz kellenek a 4 szabadságfokú χ^2-eloszlás $\frac{0,05}{2}$- és $1 - \frac{0,05}{2}$-kvantilisei.
Konfidencia intervallum σ^2-re:
$\left[\frac{4 \cdot 2,019^2}{\chi^2_{4;0,99}}; \frac{4 \cdot 2,019^2}{\chi^2_{4;0,01}}\right] = \left[\frac{4 \cdot 2,019^2}{13,28}; \frac{4 \cdot 2,019^2}{0,3}\right] = [1,2278; 54,35]$
Konfidencia intervallum σ-ra: $[1,1; 7,37]$.

46.) Valaki azt állítja, hogy a klíma változik, és ezt azzal véli bizonyítottnak, hogy az elmúlt 10 évben 2-szer is volt jégeső, pedig korábban az egyes évekre a jégeső valószínűsége a hivatalos adatok alapján csupán p=0.1 volt. Írjuk fel a hipotéziseket, a próbát és állapítsuk meg az elsőfajú hiba valószínűségét, valamint az erőfüggvényt a p=0.2 pontban!
Megoldás.
$H_0 : p = 0,1$
$\underline{H_1 : p \neq 0,1}$
$\Theta_0 = \{0,1\}$ és $\Theta_1 = \{p \in [0;1], p \neq 0,1\}$
Próba: H_0-t elfogadjuk, ha 10-ből pontosan 1 jégeső volt
Legyen X valószínűségi változó: 10-ből hány évben volt jégeső. Ekkor nyilvánvalóan $X \sim$Bin(10,p)
$\alpha(\theta) = P_{\theta \in \Theta_0}(\mathcal{X}_k) = P_{p=0,1}(\mathcal{X}_k) = 1 - P_{p=0,1}(\mathcal{X}_e) =$
$= 1 - P_{p=0,1}(\text{10-ből 1 jégeső}) = 1 - \binom{10}{1} p(1-p)^9\Big|_{p=0,1} =$
$= 1 - 10 \cdot 0,1 \cdot 0,9^9 = 1 - 0,9^9 = 61,25\%$
$\psi(p) = P_p(\mathcal{X}_k) = 1 - 10p(1-p)^9$
$\psi(0,2) = 1 - 10 \cdot 0,2 \cdot 0,8^9 = 0,7316.$

47.) Az alábbi minta 4 év október 18-án Budapesten mért napi középhőmérséklet adatait tartalmazza. Ellenőrizzük a H_0: m =15 hipotézist α =0.05 elsőfajú hibavalószínűség mellett *értelmes* alternatív hipotézissel szemben.
Középhőm. (C fok) adatok:

14,8	12,2	16,8	11,1

a.) A korábbi tapasztalatok alapján tekintsük az értékek szórását 2-nek. Adjuk meg a p-értéket is.
b.) Ne használjunk a szórásra vonatkozóan előzetes információt.

Megoldás. $\overline{x} = 13,725$, $n = 5$, $\alpha = 0,05$

a.) $\sigma = 2$, $u = \sqrt{4} \cdot \frac{13,725-15}{2} = -1,275.$
Két értélmes H_1-et lehet választani:

- $\underline{H_1 : m \neq 15}$
  Kritikus értékhez: $u_{\frac{\alpha}{2}} = \Phi^{-1}\left(1 - \frac{0,05}{2}\right) = \Phi^{-1}(0,975) = 1,96 \quad \Rightarrow$ a kritikus tartomány azokat a mintákat tartalmazza, amikből számolt próbastatisztikára $u \in K := ((-\infty; -1,96) \cup (1,96; \infty))$
  A próbastatisztika nem esik K halmazba $\Rightarrow$ a minta az elfogadási tartományban van $\Rightarrow$ elfogadjuk H_0-t (nem tudjuk elvetni), tehát állíthatjuk, hogy a hőmérséklet a vizsgált 4 éven keresztül 15 fok volt.

  p-érték: azon α, amire $1,275 = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \quad \Rightarrow$
  $\Rightarrow \quad 1 - \frac{\alpha}{2} = \Phi(1,275) = 0,899 \quad \Rightarrow \quad$ p-érték$= 20,2\%$
  Természetesen ugyanazt a döntést hozzuk p-értékkel is:
  $\alpha = 5\% <$ p-érték $\Rightarrow$ elfogadjuk H_0-t
- $\underline{H_1 : m < 15}$
  Kritikus értékhez: $u_\alpha = \Phi^{-1}(1 - 0,05) = \Phi^{-1}(0,95) = 1,645 \quad \Rightarrow$ a kritikus tartomány azokat a mintákat tartalmazza, amikből számolt próbastatisztikára $u \in L := (-\infty; -1,645)$
  A próbastatisztika nem esik L halmazba $\Rightarrow$ a minta az elfogadási tartományban van $\Rightarrow$ elfogadjuk H_0-t (nem tudjuk elvetni), tehát állíthatjuk, hogy a hőmérséklet a vizsgált 4 éven keresztül 15 fok volt.

  p-érték: azon α, amire $1,275 = \Phi^{-1}(1 - \alpha) \quad \Rightarrow$
  $\Rightarrow \quad 1 - \alpha = \Phi(1,275) = 0,899 \quad \Rightarrow \quad$ p-érték$= 10,1\%$
  Döntés p-értékkel: $\alpha = 5\% <$ p-érték $\Rightarrow$ elfogadjuk H_0-t

b.) $s_n^* = 2,57; t = \sqrt{4} \cdot \frac{13,725-15}{2,57} = -0,99$; szabadságfok $= 4 - 1 = 3$
Két értélmes H_1-et lehet választani:

- $\underline{H_1 : m \neq 15}$
  Kritikus értékhez: $t_{n-1;\frac{\alpha}{2}} = t_{3;\frac{0,05}{2}} = 3,182 \quad \Rightarrow$ a kritikus tartomány azokat a mintákat tartalmazza, amikből számolt próbastatisztikára $t \in K := ((-\infty; -3,182) \cup (3,182; \infty))$

A próbastatisztika nem esik K halmazba $\Rightarrow$ a minta az elfogadási tartományban van $\Rightarrow$ elfogadjuk H_0-t (nem tudjuk elvetni), tehát állíthatjuk, hogy a hőmérséklet a vizsgált 4 éven keresztül 15 fok volt.

- $H_1 : m < 15$
  Kritikus értékhez: $t_{n-1;\alpha} = t_{3;0,05} = 2,353$ $\Rightarrow$ a kritikus tartomány azokat a mintákat tartalmazza, amikből számolt próbastatisztikára $t \in L := (-\infty; -2,353)$
  A próbastatisztika nem esik L halmazba $\Rightarrow$ a minta az elfogadási tartományban van $\Rightarrow$ elfogadjuk H_0-t (nem tudjuk elvetni), tehát állíthatjuk, hogy a hőmérséklet a vizsgált 4 éven keresztül 15 fok volt.

48.) Az Informatikai Kar III. évfolyamán 300-an tanulnak. Megszámolták, hogy a legutóbbi vizsgaidőszakban hányszor buktak az egyes hallgatók. Az eredményeket tartalmazza az alábbi táblázat.

Bukások száma	0	1	2	3	4
Hallgatók száma	80	113	77	27	3

a.) Elfogadhatjuk-e azt a hipotézist, hogy egy hallgató bukásszáma Bin(4; 0,25) eloszlású?

b.) és azt, hogy Bin(4;p) eloszlású?

Megoldás. Mindkét esetben diszkrét illeszkedésvizsgálatot kell elvégeznünk.

a.) H_0: a bukások száma Bin(4; 0,25) eloszlású
H_1: nem ilyen eloszlású

Foglaljuk táblázatba a számításokhoz szükséges értékeket:

Bukások sz.	0	1	2	3	4	Összesen
N_i	80	113	77	27	3	300
p_i	0,3164	0,4219	0,2109	0,0469	0,0039	1
$n \cdot p_i$	94,92	126,57	63,27	14,07	1,17	300

$n = 300$, $r = 5$, a p_i értékeket a Bin(4;0,25) eloszlásból kapjuk:
$p_i = \binom{4}{i}\left(\frac{1}{4}\right)^i\left(\frac{3}{4}\right)^{4-i} \quad i = 0, 1, ..., 4$
$T_{300} = \frac{(80-94,92)^2}{94,92} + ... + \frac{(3-1,17)^2}{1,17} = 21,97$
Szabadságfok: 5-1=4
Kritikus érték: $\chi^2_{4;1-0,05} = 9,49$
A próbastatisztika nagyobb a kritikus értéknél $\Rightarrow$ a minta a kritikus tartományban van $\Rightarrow$ elvetjük H_0-t, tehát azt állíthatjuk, hogy a bukások száma NEM Bin(4; 0,25) eloszlású.

b.) H_0: a bukások száma 4 rendű binomiális eloszlású
H_1: nem ilyen eloszlású

Mindenekelőtt meg kell becsülni az ismeretlen p paramétert ML-módszerrel. Korábban levezettük, hogy Bin(m, p) minta esetén, ha m ismert, akkor p ML-becslése a $\hat{p} = \frac{\overline{X}}{m}$.
Így p becslése: $\frac{\frac{80\cdot0+113\cdot1+77\cdot2+27\cdot3+4\cdot3}{300}}{4} = 0,3 \quad \Rightarrow \quad p_i = \binom{4}{i}0,3^i0,7^{4-i}$

Foglaljuk táblázatba a számításokhoz szükséges értékeket:

Bukások sz.	0	1	2	3	4	Összesen
N_i	80	113	77	27	3	300
$n \cdot p_i$	72	123,5	79,4	22,7	2,4	300

$T_{300} = \frac{(80-72)^2}{72} + ... + \frac{(3-2,4)^2}{2,4} = 2,82$
Szabadságfok: $5 - 1 - 1 = 3$ (mert 1 darab paramétert becsültünk)
Kritikus érték: $\chi^2_{3;1-0,05} = 7,81$
A próbastatisztika kisebb a kritikus értéknél $\Rightarrow$ a minta az elfogadási tartományban van $\Rightarrow$ elfogadjuk H_0-t, tehát azt állíthatjuk, hogy a bukások száma 4 rendű binomiális eloszlású.

49.) Az alábbi kontingencia-táblázat mutatja, hogy 100 évben a csapadék mennyisége és az átlaghőmérséklet hogyan alakult.

Csapadék / Hőmérséklet	Kevés	Átlagos	Sok
Hűvös	15	10	5
Átlagos	10	10	20
Meleg	5	20	5

(A cellákban az egyes esetek gyakoriságai találhatóak.) Tekinthető-e a csapadékmennyiség és a hőmérséklet függetlennek?

Megoldás.
Függetlenségvizsgálatot kell elvégeznünk.

Csapadék / Hőmérséklet	Kevés	Átlagos	Sok	Összesen
Hűvös	15	10	5	30
Átlagos	10	10	20	40
Meleg	5	20	5	30
Összesen	30	40	30	100

$n = 100$, α legyen 5%

H_0: a csapadék éa a hőmérséklet függetlenek

H_1: nem függetlenek

$T_{100} = 100\left(\frac{15^2}{30\cdot 30} + ... + \frac{5^2}{30\cdot 30} - 1\right) = 22,9$

A szabadságfok $(3-1)(3-1) = 4$, így $\chi^2_{4;1-\alpha} = 9,49$

A próbastatisztika nagyobb a kritikus értéknél $\Rightarrow$ a minta a kritikus tartományban van $\Rightarrow$ elvetjük H_0-t, tehát azt állíthatjuk, hogy a csapadék és a hőmérséklet nem függetlenek.

50.) Legyen X a hatosok száma 6 kockadobásból, Y pedig $X+Z$, ahol Z további 6 kockadobásból a hatosok száma. Mi lesz Y legkisebb négyzetes közelítése X segítségével, ha

a.) X lineáris függvényével közelítünk;

b.) X tetszőleges függvényével közelítünk?

Megoldás. X, Z $\mathrm{Bin}(6,\frac{1}{6})$ eloszlásúak és függetlenek egymástól $\Rightarrow$ $Y \sim \mathrm{Bin}(12,\frac{1}{6})$.

a.) $a_{\text{opt}} = \frac{\mathrm{Cov}(X,Y)}{D^2X} = \frac{\mathrm{Cov}(X,X+Z)}{D^2X} = \frac{D^2X+\mathrm{Cov}(X,Y)}{D^2X} = 1$

$b_{\text{opt}} = EY - a_{\text{opt}}EX = 12\cdot\frac{1}{6} - 1\cdot 6\cdot\frac{1}{6} = 2-1 = 1.$

Tehát a legjobb lineáris becslés: $X+1$.

b.) $f_{\text{opt}}(X) = E(Y|X) = E(X+Z|X) = E(X|X) + E(Z|X) =$

$= X + EZ = X + 1.$

Tehát a legjobb becslés: $X+1$.

51.) Legyenek adottak a következő (x,y) párok:

x_i	0	1	6	5	3
y_i	4	3	0	1	2

a.) Határozzuk meg és ábrázoljuk is az $aX+b$ alakú regessziós egyenest.

b.) Számoljuk ki a reziduálisokat és becsüljük meg a hiba-szórásnégyzetet.

c.) Adjunk előrejelzést x=10-re a regressziós egyenes alapján.

Megoldás. Számítsuk ki a szükséges értékeket, ehhez célszerű táblázatot készíteni:

	x_i	y_i	$x_i-\overline{x}$	$y_i-\overline{y}$	$\hat{a}x_i+\hat{b}$	$\hat{\varepsilon}_i$
	0	4	-3	2	$\frac{50}{13}$	$\frac{2}{13}$
	1	6	-2	1	$\frac{42}{13}$	$-\frac{3}{13}$
	6	0	3	-2	$\frac{2}{13}$	$-\frac{2}{13}$
	5	1	2	-1	$\frac{10}{13}$	$\frac{3}{13}$
	3	2	0	0	$\frac{26}{13}$	0
$\sum$	15	10	0	0	-	0

$\overline{x} = 3$; $\overline{y} = 2$

a.) $\hat{a} = \frac{-6-2-6-2}{2\cdot 9+2\cdot 4} = -\frac{8}{13}$

$\hat{b} = 2 - \left(-\frac{8}{13}\right)\cdot 3 = \frac{50}{13}$

Tehát a regressziós egyenes: $-\frac{8}{13}x + \frac{50}{13}$.

b.) RNÖ$= \left(\frac{2}{13}\right)^2 + ... + \left(\frac{3}{13}\right)^2 = \frac{2}{13}$

$\hat{\sigma}^2 = \frac{\frac{2}{13}}{3} = \frac{2}{39}$.

c.) $\frac{-80+50}{13} = -\frac{30}{13}$.

52.) Véletlenszerűen választunk egy szót az alábbi mondatból: EGY TEVE LEGEL A KERTBEN. A feladatunk az, hogy kitaláljuk a szó hosszát úgy, hogy a tényleges és a tippelt szóhossz közötti eltérés négyzetének várható értéke minimális legyen.

a.) Mit tippelünk, ha semmi információ nem áll rendelkezésünkre?

b.) Hogyan tippelünk, ha valaki megsúgta a szóban szereplő "e"-betűk számát?

c.) Hogyan tippeljünk, ha az "e" betűk számának lineáris függvényét használhatjuk?

Megoldás. Legyen X: tippelt szóhossz; Y: tényleges szóhossz; Z: "e" betűk száma egy szóban

Ekkor X és Y függetlenek és azonos eloszlásúak, az 1,3,4,5 és 7 értékeket egyaránt $\frac{1}{5}$ valószínűséggel veszik fel.

Z eloszlása: $P(Z=0) = \frac{1}{5}, P(Z=1) = \frac{1}{5}, P(Z=2) = \frac{3}{5}$

a.) Y-t akarjuk közelíteni X-szel $\Rightarrow$ $f_{\text{opt}}(X) = E(Y|X) = EY =$

$= \frac{1+3+4+5+7}{5} = 4$

b.) Y-t akarjuk közelíteni Z tetszőleges függvényével $\Rightarrow$ $f_{\text{opt}}(Z) = E(Y|Z)$

$E(Y|Z=0) = 1$, mert az egyetlen "e" betűt nem tartalmazó szó az "a", ami 1 betűből áll

$E(Y|Z=1) = 3$, mert az egyetlen 1 darab "e" betűt tartalmazó szó az "egy", ami 3 betűből áll

$E(Y|Z=2) = \frac{1}{3}(4+5+7) = \frac{16}{3}$, mert a 3 darab "e" betűt tartalmazó szavak a "teve", "legel", "kertben, amik 4,5 és 7 betűből állnak

Tehát $E(Y|Z) = \mathrm{I}(Z=0) + 3\mathrm{I}(Z=1) + \frac{16}{3}\mathrm{I}(Z=2)$.

c.) Y-t akarjuk közelíteni Z lineáris függvényével $\Rightarrow$ $a_{\text{opt}} = \frac{\mathrm{Cov}(Z,Y)}{D^2Z}$

$b_{\text{opt}} = EY - a_{\text{opt}}EZ$

Írjuk fel Z és Y együttes eloszlását:

$Z\backslash Y$	1	3	4	5	7	Z peremeloszlása
0	$\frac{1}{5}$	0	0	0	0	$\frac{1}{5}$
1	0	$\frac{1}{5}$	0	0	0	$\frac{1}{5}$
2	0	0	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{3}{5}$
Y peremeloszlása	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	1

$EZ = \frac{1+6}{5} = \frac{7}{5}$
$EZ^2 = \frac{1+12}{5} = \frac{13}{5}$
$D^2Z = \frac{13}{5} - \frac{49}{25} = \frac{16}{25}$
$E(ZY) = \frac{3+8+10+14}{5} = 7$
$\mathrm{Cov}(Z,Y) = 7 - \frac{7}{4} \cdot 4 = \frac{7}{5}$
$a_{\mathrm{opt}} = \frac{\frac{7}{5}}{\frac{16}{25}} = \frac{35}{16}$
$b_{\mathrm{opt}} = 4 - \frac{35}{16} \cdot \frac{7}{5} = \frac{15}{16}$
Tehát a legjobb lineáris becslés: $\frac{35}{16}Z + \frac{15}{16}$.