Programtervező informatikus BSc, B szakirány
Valószínűségszámítás és statisztika gyakorlat, feladatok megoldása

1. (1-2 hét) Valószínűségek kiszámítása; feltételes valószínűség és Bayes-tétel

1.1. Feladat. Hányféleképpen lehet 8 bástyát letenni egy sakktáblára, hogy ne üssék egymást?

Megoldás

Az első bástya 64 helyre kerülhet. Ekkor a lefedett mező sorába és oszlopába már nem kerülhet újabb bástya, így a következőt már csak 7 sor és 7 oszlop valamelyikébe tehetjük le, ami 49 lehetőség. Minden újabb bástya letételével még egy újabb sor és oszlop kerül lefedésre. Tehát ezután sorra $36, 25, 16, 9, 4$, és 1 lehetőség van a következő bástyák letételére. Tehát összesen $64 \cdot 49 \cdot 36 \cdot 25 \cdot 16 \cdot 9 \cdot 4 \cdot 1 = (8!)^2$ féleképp tehetjük le a bástyákat.Viszont a bástyák letevésének sorrendjét így figyelembe vettük. Amennyiben ezt a sorrendet nem vesszük figyelembe úgy le kell osztanunk a lerakott bástyák permutációinak számával, azaz 8!-sal. Ekkor 8! féleképp tehetjük le a bástyákat. A végeredményt közvetlenül is megkaphatjuk, ha oszloponként (ill. soronként) nézzük a bástyák helyét.

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

Megoldás

Az első számjegyet az $1, 2, \ldots, 9$ számjegyek közül, a többi számjegyet a $0, 1, 2, \ldots, 9$ számjegyek közül választhatjuk. Így az összes esetek száma $9 \cdot 10^5$. Kedvező esetek száma: $9 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5$, mert itt visszatevés nélkül választunk, a sorrend számít, illetve arra figyelünk, hogy az első számjegy ne lehessen 0. Tehát a keresett valószínűség $\frac{9\cdot9\cdot8\cdot7\cdot6\cdot5}{9\cdot10^5} = \frac{136080}{900000} = 0{,}1512$.

1.3. Feladat. Ha egy magyarkártya-csomagból (32 lap: piros, zöld, makk, tök) visszatevéssel húzunk három lapot, akkor mi annak a valószínűsége, hogy

a) pontosan egy piros színű lapot húztunk?

b) legalább egy piros színű lapot húztunk?

Megoldás

a) Az összes lehetőségek száma 32^3. A 3 kihúzott lap közül $\binom{3}{1} = 3$-féleképp dönthetjük el, hogy melyik legyen a piros színű. 8-féle piros és 24 nem piros lap közül választhatunk.
Tehát a keresett valószínűség: $\frac{3\cdot8\cdot24^2}{32^3} = \frac{27}{64} = 0{,}4219$.

b) Kényelmesebb most a komplementer esemény valószínűségét kivonni 1-ből. A komplementer esemény: nincsen piros a húzott lapok között. Ennek valószínűsége $\frac{24^3}{32^3} = \frac{27}{64}$. Tehát a keresett valószínűség $1 - \frac{27}{64} = \frac{37}{64} = 0{,}5781$.

1.4. Feladat. Egy zsákban 10 pár cipő van. 4 db-ot kiválasztva, mi a valószínűsége, hogy van közöttük pár, ha

a) egyformák a párok?

b) különbözőek a párok?

Megoldás

a) 10 balos és 10 jobbos cipő van. Mi a valószínűsége, hogy a 4 kihúzott között van balos és jobbos is? Célszerű most is a komplementer esemény valószínűségét kivonni 1-ből. A komplementer esemény: vagy 4 balosat húztunk, vagy 4 jobbosat. Ennek valószínűsége: $2 \cdot \frac{10\cdot9\cdot8\cdot7}{20\cdot19\cdot18\cdot17} = \frac{28}{323}$ vagy $\frac{\binom{10}{4}}{\binom{20}{4}} + \frac{\binom{10}{4}}{\binom{20}{4}} = \frac{28}{323}$. Tehát a keresett valószínűség $1 - \frac{28}{323} = 0{,}9133$.

b) Most is érdemes a komplementer esemény valószínűségét kiszámítani. Komplementer esemény: nincs pár a 4 cipő között. Ha így akarom a cipőket kiválasztani, akkor az elsőt 20-féleképp választhatom ki, a másodikat 18-féleképp (az első és párja kiesik), a harmadikat 16-féleképp és a negyediket 14-féleképp. Összes eset: $20 \cdot 19 \cdot 18 \cdot 17$. Tehát a komplementer esemény valószínűsége $\frac{20\cdot18\cdot16\cdot14}{20\cdot19\cdot18\cdot17} = \frac{224}{323}$ vagy kiválasztunk 10 párból a 4 párat először, majd ezek balosát ill. jobbosát $\frac{\binom{10}{4}\binom{2}{1}\binom{2}{1}\binom{2}{1}\binom{2}{1}}{\binom{20}{4}} = \frac{224}{323}$. Tehát a keresett valószínűség $1 - \frac{224}{323} = 0{,}3065$.

1.5. Feladat. $\star$ n dobozba véletlenszerűen helyezünk el n golyót úgy, hogy bármennyi golyó kerülhet az egyes dobozokba.

a) Mi a valószínűsége, hogy minden dobozba kerül golyó?

b) Annak mi a valószínűsége, hogy pontosan egy doboz marad üresen?

Megoldás

Vegyük észre hogy a probléma kitűzése nem határozza meg teljesen egyértelműen hogy milyen valószínűségi modellt kell használni, ugyanis nem írja elő hogy milyen módon helyezzük a golyókat a dobozokba, s azt sem rögzíti hogy megkülönböztetett vagy azonos golyókról van szó. Mindenesetre feltesszük hogy a dobozok meg vannak különböztetve (habár a feladat kitűzése ezt sem rögzíti).

a) 1. Értelmezés: A golyókat megkülönböztetjük (ez nem feltétlenül jelenti, hogy a golyók fizikailag különbözőek, már az is megkülönböztetés, hogy ha egymás után rakjuk őket a dobozokba, s így első, második stb., golyóról lehet beszélni). Ilyenkor, hacsak a feladat explicite nem ír elő mást, a "véletlenszerűen" szó értelmezése az, hogy minden golyót egymástól függetlenül, azonos $(1/n)$ valószínűséggel helyezünk a dobozokba.
Tekintsük az $n = 2$ esetet, egyszerűség kedvéért. A valószínűségi tér természetes módon egy szorzattér, $\Omega = \{1, 2\} \times \{1, 2\}$, ahol a Descartes szorzat első komponense azt kódolja el, hogy az első golyó az 1-es vagy a 2-es dobozba kerül, a második komponens ugyanezt teszi a második golyóval. Például $\omega = (2, 1)$ azt jelenti, hogy az első golyó a 2-es, a második golyó az 1-es dobozba került. Összesen $2 \cdot 2 = 4$ kimeneti lehetőség van, és a függetlenségi feltevés miatt mindegyik $1/2 \cdot 1/2 = 1/4$ valószínűségű.
Általánosan: n megkülönböztetett golyót n dobozba n^n féleképpen tudjuk betenni (ismétléses variáció). A kedvező esetek száma $n!$, azaz a lehetséges permutációk száma. Így a keresett valószínűség

$$\mathbb{P}(\text{minden dobozban van egy golyó}) = \frac{n!}{n^n}.$$

2. Értelmezés: Ha a golyók nincsenek megkülönböztetve, és a berakási folyamat sem utal rá, akkor úgy is okoskodhatunk, hogy csupán a végeredményt látjuk és a valószínűségi terünket az összes lehetséges kimenet halmazaként definiáljuk. Vegyük észre, hogy az 1. Értelmezéssel ellentétben most $n = 2-$nél mindössze 3 lehetőségünk van:
(a) az első dobozban két golyó, a másodikban semmi;
(b) mindkét dobozban egy golyó;
(c) első dobozban semmi, a másodikban kettő.
Struktúrájában ez a valószínűségi tér nagyon más mint az előző, nemcsak az elemek száma különbözik, de nincs Descartes szorzat struktúrája sem. A "véletlenszerűen" szó elvileg értelmezhető úgy is, hogy a három lehetséges kimenet egyenlő valószínűségű. Így például $1/3$ annak a valószínűsége hogy mindkét dobozba egy-egy golyó került, míg az első értelmezés szerint ugyanez a valószínűség $1/2$.
Általánosan: n nem megkülönböztetett n dobozba $\binom{2n-1}{n}$ féleképpen tudjuk betenni (ismétléses kombináció). [Rendezzük az n dobozt sorba, ekkor $n - 1$ válaszfal keletkezik közöttük. Az összes esetek száma az n golyó és az $n - 1$ válaszfal sorrendjeinek száma, ami egy ismétléses permutáció: $\frac{\big(n+(n-1)\big)!}{n!\cdot(n-1)!} = \binom{2n-1}{n}$.] A kedvező esetek száma 1, azaz minden dobozba egy golyó kerül. Így a keresett valószínűség

$$\mathbb{P}(\text{minden dobozban van egy golyó}) = \frac{1}{\binom{2n-1}{n}}.$$

A két értelmezés közötti döntés nem matematikai hanem modellezési probléma; sokszor azonban a matematikusnak kell rámutatni a felhasználónál arra, ha esetleg a probléma nincs kellő pontossággal megfogalmazva. Esetünkben az első megközelítés jól jól írja le a gázmolekulák viselkedését, a második pedig a fotonokét. A félév folyamán – ha külön nem jellezzük – az első megközelítést alkalmazzuk a feladatoknál.

b) Ha a golyókat megkülönbözőztetjük, akkor - mint előbb - az n golyót n dobozba n^n féleképpen tudjuk letenni (ismétléses variáció). A kedvező esetek számát a következőképpen kaphatjuk: az üres dobozt n féleképpen, a dobozt melyben 2 golyó lesz pedig $n - 1$ féleképpen választhatjuk ki. Az n golyót $n!$ féleképpen tehetjük le, viszont kétféleképpen is eljuthatunk ugyanahhoz az elrendezéshez, hiszen a 2 golyós dobozban bármelyik jöhetett a most üres dobozból. Így a keresett valószínűség

$$\mathbb{P}(\text{pontosan egy doboz marad üresen}) = \frac{n(n-1)\frac{n!}{2}}{n^n} = \frac{\binom{n}{2}n!}{n^n}.$$

Ha a golyókat nem különbözőztetjük meg, akkor az n golyót n dobozba $\binom{2n-1}{n}$ féleképpen tudjuk betenni (ismétléses kombináció). A kedvező esetek számát a következőképpen kaphatjuk: az üres dobozt n féleképpen, a dobozt melyben 2 golyó lesz pedig $n - 1$ féleképpen választhatjuk ki. Így a keresett valószínűség

$$\mathbb{P}(\text{pontosan egy doboz marad üresen}) = \frac{n(n-1)}{\binom{2n-1}{n}}.$$

1.6. Feladat. Egy boltban 10 látszólag egyforma számítógép közül 3 felújított, a többi új. Mi a valószínűsége, hogy ha veszünk 5 gépet a laborba, akkor pontosan 2 felújított lesz közöttük?

Megoldás

A 10 gépből 3 felújított, 7 új. Tehát a 3 felújított gép közül kell 2-t kiválasztani, illetve a 7 új gép közül kell a maradék 3-mat kiválasztani. A kiválasztás sorrendje nem számít, ez visszatevés nélküli mintavétel. A kedvező esetek száma: $\binom{3}{2} \cdot \binom{7}{3} = 3 \cdot 35 = 105$. Összes esetek száma: $\binom{10}{5} = 252$. Tehát a keresett valószínűség $\frac{105}{252} = 0{,}4167$. (ez megfelel a későbbiekben definiált hipergeometriai eloszlásnak $N = 10, M = 3, n = 5$ paraméterekkel.)

1.7. Feladat. Ha a 6 karakteres jelszavunkat véletlenszerűen választjuk a 10 számjegy és a 26 karakter közül, akkor mi a valószínűsége, hogy pontosan 3 szám lesz benne?

Megoldás

$\binom{6}{3} = 20$-féleképp lehet a 6 karakterből a 3 szám helyét kiválasztani. Ezután feltehető, hogy az első 3 karakter szám, az utolsó 3 karakter betű. Számjegy választásának valószínűsége $\frac{10}{36}$, betűé $\frac{26}{36}$. A keresett valószínűség tehát $\binom{6}{3} \cdot (\frac{10}{36})^3 \cdot (\frac{26}{36})^3 = 0{,}1615$. Itt feltettük, hogy a választások függetlenek. (ez megfelel a későbbiekben definiált binomiális eloszlásnak $n = 60, p = \frac{10}{36}$ paraméterekkel.)

1.8. Feladat. Az ötöslottónál adjuk meg annak a valószínűségét, hogy egy szelvénnyel játszva öttalálatosunk lesz, illetve hogy legalább négyesünk lesz. Mi a valószínűsége, hogy minden kihúzott szám páros? (Hogy viszonylik ez a visszatevéses esethez?)

Megoldás

Annak a valószínűsége, hogy ötösünk lesz: $\frac{\binom{5}{5}}{\binom{90}{5}} = \frac{1}{\binom{90}{5}}$.

Annak a valószínűsége, hogy legalább négyesünk lesz: $\frac{\binom{5}{5}}{\binom{90}{5}} + \frac{\binom{5}{4}\binom{85}{1}}{\binom{90}{5}}$.

Annak a valószínűsége, hogy minden kihúzott szám páros: $\frac{\binom{45}{5}}{\binom{90}{5}} \approx 0,028$.

A visszatevéses esetben (tehát, mikor egy számot többször is kihúzhatunk) annak a valószínűsége, hogy párosakat húzunk: $\left(\frac{45}{90}\right)^5 = \left(\frac{1}{2}\right)^5 \approx 0,031$. Bár a két érték közel van egymáshoz, a visszatevés nélküli esetben kisebb a valószínűség, mert ott fogynak a páros számok a választás során.

1.9. Feladat. Mennyi a valószínűsége, hogy két kockadobásnál mind a két dobás 6-os, feltéve, hogy tudjuk, hogy legalább az egyik dobás 6-os?

Megoldás

Legyen A esemény az, hogy mindkét dobás hatos, B pedig, hogy legalább az egyik hatos. Ekkor

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{\frac{1}{36}}{\frac{11}{36}} = \frac{1}{11}$$

1.10. Feladat. 41 millió ötöslottó-szelvényt töltenek ki egymástól függetlenül. Mennyi a valószínűsége, hogy lesz legalább egy 5-ös találat?

Megoldás

$$P(\text{legalább egy ötös találat lesz a 41M-ból}) = 1 - P(\text{nem lesz ötös találat a 41M-ból}) \overset{\text{függetlenség}}{=}$$

$$= 1 - P(\text{egy embernek nem lesz ötös találata})^{41\cdot 10^6} = 1 - \left(1 - \frac{\binom{5}{5}}{\binom{90}{5}}\right)^{41\cdot 10^6} \approx 0,6066.$$

1.11. Feladat. 100 érme közül az egyik hamis (ennek mindkét oldalán fej található). Egy érmét véletlenszerűen kiválasztva és azzal 10-szer dobva, 10 fejet kaptunk. Ezen feltétellel mi a valószínűsége, hogy a hamis érmével dobtunk?

Megoldás

Jelölje A azt az eseményt, hogy 10 dobásból 10 fej, B_1 azt, hogy jó érmével dobtunk, illetve B_2 azt, hogy hamis érmével dobtunk. Ekkor:

$$P(B_1) = \frac{99}{100}; \qquad P(A|B_1) = \binom{10}{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^{0} = \frac{1}{2^{10}}$$
$$P(B_2) = \frac{1}{100}; \qquad P(A|B_2) = 1$$

Alkalmazva a Bayes-tételt:

$$P(B_2|A) = \frac{P(A|B_2)P(B_2)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} = \frac{1 \cdot \frac{1}{100}}{\frac{1}{1024} \cdot \frac{99}{100} + 1 \cdot \frac{1}{100}} \approx 0.9118.$$

1.12. Feladat. Egy diák a vizsgán p valószínűséggel tudja a helyes választ. Amennyiben nem tudja, akkor tippel (az esélye, hogy eltalálja a helyes választ, ekkor $\frac{1}{3}$). Ha helyesen válaszolt, mennyi a valószínűsége, hogy tudta a helyes választ?

Megoldás

Jelölje A azt az eseményt, hogy helyesen válaszolt, B_1 azt, hogy tudta a választ, illetve B_2, hogy nem tudta a választ. Ekkor:

$$P(B_1) = p; \qquad P(A|B_1) = 1$$
$$P(B_2) = 1 - p; \qquad P(A|B_2) = \frac{1}{3}$$

Alkalmazva a Bayes-tételt:

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} = \frac{1 \cdot p}{1 \cdot p + \frac{1}{3} \cdot (1-p)} = \frac{3p}{2p+1}$$

1.13. Feladat. Egy számítógépes program két független részből áll. Az egyikben $0,2$, a másikban $0,3$ a hiba valószínűsége. Ha a program hibát jelez, akkor mi a valószínűsége, hogy mindkét rész hibás?

Megoldás

Vezessük be a következő jelöléseket:

- A - a program hibát jelez;
- B_1 - egyik rész sem hibás;
- B_2 - pontosan az egyik rész hibás;
- B_3 - mindkét rész hibás.

Ekkor

$$\begin{aligned}
&P(B_1) = P(\text{sem az első, sem a második}) = (1-0,2)(1-0,3) = 0,56 && P(A|B_1) = 0\\
&P(B_2) = P(\text{pontosan az egyik}) = 0,2(1-0,3) + 0,3(1-0,2) = 0,14 + 0,24 = 0,38; && P(A|B_2) = 1\\
&P(B_3) = 0,06; && P(A|B_3) = 1
\end{aligned}$$

Alkalmazva a Bayes-tételt:

$$P(B_3|A) = \frac{P(A|B_3)P(B_3)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)} = \frac{1 \cdot 0,06}{0 \cdot 0,56 + 1 \cdot 0,38 + 1 \cdot 0,06} = \frac{0,06}{0,44} \approx 0,1364.$$

1.14. Feladat. Egy számítógép processzorát 3 üzemben készítik. 20% eséllyel az elsőben, 30% eséllyel a másodikban és 50% eséllyel a harmadikban. A garanciális hibák valószínűsége az egyes üzemekben rendre 10%, 4%, illetve 1%. Ha a gépünk processzora elromlott, akkor mi a valószínűsége, hogy az első üzemben készült?

Megoldás

Vezessük be a következő jelöléseket:

- A - a processzorunk elromlott;
- B_1 - a processzorunk az első üzemben készült;
- B_2 - a processzorunk a második üzemben készült;
- B_3 - a processzorunk a harmadik üzemben készült.

Ekkor

$$P(B_1) = 0,2; \qquad P(A|B_1) = 0,10$$
$$P(B_2) = 0,3; \qquad P(A|B_2) = 0,04$$
$$P(B_3) = 0,5; \qquad P(A|B_3) = 0,01$$

Alkalmazva a Bayes-tételt:

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)} = \frac{0,1 \cdot 0,2}{0,1 \cdot 0,2 + 0,04 \cdot 0,3 + 0,01 \cdot 0,5} \approx 0,5405$$

2. (3-4 hét) Valószínűségi változó, diszkrét eloszlások

Feladatok

2.1. Feladat. 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.)

Megoldás

Jelölje az X valószínűségi változó a fiúk számát. Feltesszük, hogy a gyermekek neme független egymástól (ez a valóságban nem teljesen igaz). Ekkor a feladat visszatevéses mintavételként kezelhetõ, mely paramétereire $p = \frac{1}{2}$ és $n = 6$ teljesülnek $(X\ B(6, \frac{1}{2}))$. Amiből a kívánt eloszlás:

$$P(X = k) = \binom{6}{k} \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{6-k} = \binom{6}{k} \cdot \left(\frac{1}{2}\right)^6.$$

2.2. Feladat. Tegyük fel, hogy az új internet-előfizetők véletlenszerűen választott 20%-a speciális kedvezményt kap. Mi a valószínűsége, hogy 10 ismerősünk közül, akik most fizettek elő, legalább négyen részesülnek a kedvezményben?

Megoldás

Legyen X az a valószínűségi változó, mely megadja a speciális kedvezményt kapó ismerőseink számát. Ekkor ez egy olyan visszatevéses mintavételként kezelhető feladat, mely paramétereire $p = \frac{1}{5}$ és $n = 10$, azaz $X\ B(10, \frac{1}{5})$. Így pedig

$$\begin{aligned} P(X \geq 4) &= 1 - P(X < 4) = \\ &= 1 - \left[\binom{10}{0}\left(\frac{1}{5}\right)^0\left(\frac{4}{5}\right)^{10} + \binom{10}{1}\left(\frac{1}{5}\right)^1\left(\frac{4}{5}\right)^9 + \binom{10}{2}\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^8 + \binom{10}{3}\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^7\right] \\ &= 1 - \left[\binom{10}{0}4^{10} + \binom{10}{1}4^9 + \binom{10}{2}4^8 + \binom{10}{3}4^7\right]\left(\frac{1}{5}\right)^{10} \approx 0,1209. \end{aligned}$$

2.3. Feladat. Egy tétel áru 1% selejtet tartalmaz. Hány darabot kell találomra kivennünk és megvizsgálnunk, hogy a megvizsgált darabok között legalább 0,95 valószínűséggel selejtes is legyen, ha az egyes kiválasztott darabokat vizsgálatuk után visszatesszük?

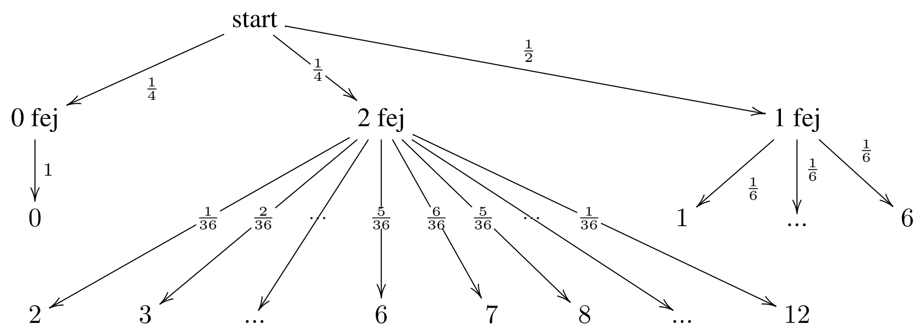
Megoldás

Legyen X = a selejtes áruk száma a vizsgált darabok közt. Ekkor mivel X binomiális eloszlású n és $p = 0,01$ paraméterekkel $P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{n}{0} \cdot 0,01^0 \cdot 0,99^n > 0,95 \Rightarrow 0,05 > 0,99^n \Rightarrow n > \frac{\ln 0,05}{\ln 0,99} \approx 298,07 \Rightarrow n \geq 299.$

2.4. Feladat. Dobjunk egy kockával annyiszor, ahány fejet dobtunk két szabályos érmével. Jelölje X a kapott számok összegét. Adjuk meg X eloszlását!

Megoldás

Esetszétbontással érdemes próbálkozni. Annak a valószínűsége, hogy 0,1,2 fejet dobunk rendre 1/4, 1/2, 1/4. Az összegek 0 és 12 közé eshetnek, attól függően, hogy hány fejet dobtunk.



$P(X = 0) = \frac{1}{4} \cdot 1$
$P(X = 1) = \frac{1}{2} \cdot \frac{1}{6}$
$P(X = 2) = \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{1}{36}$
$\vdots$
$P(X = 6) = \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{5}{36}$
$P(X = 7) = \frac{1}{4} \cdot \frac{6}{36}$
$\vdots$
$P(X = 12) = \frac{1}{4} \cdot \frac{1}{36}$

2.5. Feladat. Jelölje X az ötöslottón kihúzott lottószámok legkisebbikét. Adjuk meg X eloszlását!

Megoldás

Jelentse $X = k$ azt, hogy a legkisebb kihúzott szám k. Ez $1 - 86$-ig bármelyik szám lehet. Ezek alapján, ha tudjuk, hogy k a legkisebb:

$$P(X = k) = \frac{\binom{90-k}{4}}{\binom{90}{5}},$$

mert a maradék kihúzott szám $k + 1$ és 90 közé eshet.

2.6. Feladat. Egy érmével dobva (tfh. p a fej valószínűsége), jelölje X az első azonosakból álló sorozat hosszát. (Azaz pl., ha a sorozat FFI..., akkor $X = 2$.) Adjuk meg X eloszlását!

Megoldás

Tegyük fel, hogy k-szor dobtunk egymás után fejet. Ez akkor lesz pontosan k hosszú sorozat, ha a k fej után közvetlenül írást dobtunk. Ugyanez fordítva is kell, hogy teljesüljön, azaz k írás után 1 fej kell. Ezek alapján az eloszlás:

$$P(X = k) = p^k(1 - p) + (1 - p)^k p$$

2.7. Feladat. Legyenek az X diszkrét valószínűségi változó értékei $-2, 1, 3$, a következő valószínűségekkel:

$$P(-2) = 1/2, \quad P(1) = 1/3, \quad P(3) = 1/6.$$

Rajzolja fel az $F(x)$ eloszlásfüggvényt!

Megoldás

$$F(x) = P(X < x) = \begin{cases} 0, & \text{ha } x \leq -2 \\ \frac{1}{2}, & \text{ha } -2 < x \leq 1 \\ \frac{1}{2} + \frac{1}{3} = \frac{5}{6}, & \text{ha } 1 < x \leq 3 \\ 1, & \text{ha } x > 3 \end{cases}$$

2.8. Feladat. Tegyük fel, hogy a 3 valószínűségszámítás gyakorlatra rendre 15, 20, illetve 25 diák jár. Várhatóan mekkora egy véletlenszerűen kiválasztott diák csoportja?

Megoldás

Legyen X a valószínűségszámítás gyakorlatra járó diákok száma. Ekkor
$P(X = 15) = 15/60 = 1/4$
$P(X = 20) = 20/60 = 1/3$
$P(X = 25) = 25/60 = 5/12$
Így a várható érték $EX = 15 \cdot 1/4 + 20 \cdot 1/3 + 25 \cdot 5/12 = (45 + 80 + 125)/12 = 250/12 = 20,83$. Itt azt feltételeztük, hogy minden diákot ugyanakkora $\frac{1}{60}$ valószínűséggel választunk ki. Természetesen lehetnek más választási módok is.

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

Megoldás

Legyen X a sikeres dobások száma az n dobásból. Ekkor X egy p paraméterű binomiális eloszlást követ, melyre $p = \frac{11}{36}$ a sikeres dobás valószínűsége. Így X várható értéke $EX = np$, azaz várhatóan $\frac{11}{36}n$ sikeres dobásunk lesz.

2.10. Feladat. Tegyük fel, hogy egy dobozban van $2N$ kártyalap, melyek közül kettőn 1-es, kettőn 2-es szám van és így tovább. Válasszunk ki véletlenszerűen m lapot. Várhatóan hány pár marad a dobozban?

Megoldás

Legyen X_i annak az indikátora, hogy mindkét i feliratú lap bent marad az m lap kivétele után, azaz

$$X_i = \begin{cases} 1, & \text{ha mindkét } i \text{ feliratú lap bent marad} \\ 0, & \text{különben.} \end{cases}$$

Ekkor

$$p = P(X_i = 1) = \frac{\binom{2N-2}{m}}{\binom{2N}{m}}. \qquad \left(\text{Legyen} \binom{n}{k} := 0, \text{ ha } n < k.\right)$$

Legyen X a dobozban maradt párok száma az m lap kivétele után. Ekkor $X = X_1 + X_2 + \cdots + X_N$, melynek várható értéke

$$EX = EX_1 + EX_2 + \cdots + EX_N = Np = N\frac{\binom{2N-2}{m}}{\binom{2N}{m}} = \frac{(2N - m)(2N - 1 - m)}{2(2N - 1)}.$$

2.11. Feladat. Mennyi az ötöslottón kihúzott

a) számok összegének várható értéke?

b) páros számok számának várható értéke?

Megoldás

a) Egy húzásnál a várható érték $1 \cdot \frac{1}{90} + 2 \cdot \frac{1}{90} + \cdots + 90 \cdot \frac{1}{90} = \frac{1+2+\ldots+90}{90} = 45,5$. Öt szám kihúzása esetén pedig az összeg várható értéke a várható értékek összege: $5 \cdot 45,5 = 227,5$.

b) A lottón kihúzott (páros és páratlan) számok számának várható értéke 5, azaz $E(\text{párosak száma}) + E(\text{páratlanok száma}) = 5$. Mivel ugyanannyi páros és páratlan szám közül választhatunk, így $E(\text{párosak száma}) = E(\text{páratlanok száma})$. Ez viszont csak akkor teljesülhet, ha a $E(\text{párosak száma}) = 2,5$.

Más megoldás: Jelölje X a kihúzott páros számok darabszámát. Ekkor X hipergeometrikus eloszlást követ $N = 90$, $K = 45$ és $m = 5$ paraméterekkel, így $EX = m\dfrac{K}{N} = 5\dfrac{45}{90} = 2.5$.

2.12. Feladat. Egy bükkösben a bükkmagoncok négyzetméterenkénti száma Poisson-eloszlású, $\lambda = 2,5$ db / m^2 paraméterrel. Mi a valószínűsége annak, hogy egy 1 m^2-es mintában

a) legfeljebb egy, ill.

b) több, mint három magoncot találunk?

c) Adja meg a magoncok számanak várható értékét és szórását!

Megoldás

Legyen X a bükkmagoncok négyzetméterenkénti száma. Ekkor $X \sim Poisson(\lambda)$, ahol $\lambda = 2,5$.

a) $P(X \leq 1) = P(X = 0) + P(X = 1) = 1 \cdot e^{-2,5} + 2,5 \cdot e^{-2,5} = (1 + 2,5)e^{-2,5} \approx 0,287$.

b) $P(X > 3) = 1 - P(X \leq 3) = 1 - (P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)) = 1 - (1 \cdot e^{-2,5} + 2,5 \cdot e^{-2,5} + \frac{2,5^2}{2} \cdot e^{-2,5} + \frac{2,5^3}{6} \cdot e^{-2,5}) = 1 - \left(1 + 2,5 + \frac{2,5^2}{2} + \frac{2,5^3}{6}\right) e^{-2,5} \approx 0,242$.

c) $EX = \lambda = 2,5,\ DX = \sqrt{\lambda} = \sqrt{2,5} \approx 1,58$.

2.13. Feladat. Egy adott területről származó talajmintákban a spórák száma Poisson-eloszlású. A minták harmadában egyáltalán nincs spóra. Mi a valószínűsége annak, hogy egy mintában a spórák száma egynél több? Mekkora a spórák számának várható értéke és szórása?

Megoldás

Legyen X a spórák száma a vizsgált mintában. Ekkor $X \sim Poisson(\lambda)$.
$P(X = 0) = e^{-\lambda} = \frac{1}{3}$, így $\lambda = -\ln\dfrac{1}{3} = \ln 3 \approx 1,099$.
$P(X > 1) = 1 - P(X \leq 1) = 1 - (P(X = 0) + P(X = 1)) = 1 - (1 \cdot e^{-\ln 3} + \ln 3 \cdot e^{-\ln 3}) \approx 0,3$.
$EX = \lambda = \ln 3$ és $DX = \sqrt{\ln 3} \approx 1,048$.

3. (5-6 hét) Abszolút folytonos eloszlások, függetlenség, egyenlőtlenségek, aszimptotikus tulajdonságok)

Feladatok

3.1. Feladat. Tegyük fel, hogy egy számítógép meghibásodási időpontja 0 és 10 év között van és itt geometriai modellel írható le. Határozzuk meg a jelenség eloszlásfüggvényét!

Megoldás

Legyen a ξ valószínűségi változó a meghibásodás időpontja, azaz ξ a [0,10] intervallumból veheti fel értékeit. Ekkor $P(\xi < 0) = 0$, mivel a meghibásodás időpontja nem lehet negatív. Hasonlóan $P(\xi < 10) = 1$, mivel a számítógép 10 éven túl nem üzemelhet. Ha viszont $0 < x < 10$, akkor $P(\xi < x) = \frac{x}{10}$, mivel a meghibásodás valószínűsége arányos a szakasz hosszával.

Ekkor az eloszlásfüggvény a következő alakú: $F(x) = P(\xi < x) = \begin{cases} 0 & \text{ha } x \leq 0 \\ \frac{x}{10} & \text{ha } 0 < x \leq 10 \\ 1 & \text{ha } 10 < x \end{cases}$

Az ilyen eloszlásfüggvényű valószínűségi változót egyenletes eloszlásúnak nevezzük a $[0, 10]$ intervallumon.

3.2. Feladat. Legyen $0 < Y < 3$ valószínűségi változó. Eloszlásfüggvénye ezen az intervallumon $F(x) = cx^3$. Mennyi c és $P(-1 < Y < 1)$?

Megoldás

Mivel $Y < 3$, így $P(Y = 3) = 0$, tehát $F(x)$ folytonos az $x = 3$ pontban. Az eloszlásfüggvénynek monoton növekedőnek kell lennie és legfeljebb 1 lehet, vagyis c pozitív lehet csak és $x = 3$-ban már 1, vagyis

$$1 = \max_{x \in (0,3]} cx^3 = c \cdot 3^3 = 27c \quad \Rightarrow \quad c = \frac{1}{27}.$$

Tudjuk, hogy -1-ben az eloszlásfüggvény 0-át vesz fel, emiatt $P(-1 < Y < 1) = F(1) - F(-1) = \dfrac{1}{27} - 0.$

3.3. Feladat. Legyen X egy abszolút folytonos valószínűségi változó a $[0, c]$ intervallumon, sűrűségfüggvénye:

$$f(x) = \begin{cases} \dfrac{1}{9}x^2, & \text{ha } 0 \leq x < c \\ 0, & \text{ha } x < 0 \text{ vagy } x \geq c. \end{cases}$$

Határozza meg c-t és X eloszlásfüggvényét!

Megoldás

Mivel a sűrűségfüggvény integrálja $= 1$ a $[0, c]$ intervallumon, így $1 = \displaystyle\int_0^c \frac{1}{9}t^2\,dt = \frac{1}{9}\left[\frac{t^3}{3}\right]_0^c = \frac{1}{9}\frac{c^3}{3}$, amiből $c = 3$.

Felhasználva, hogy az eloszlásfüggvény a sűrűségfüggvény integrálja:

$$F(x) = \int_0^x \frac{1}{9}t^2\,dt = \left[\frac{1}{9}\frac{t^3}{3}\right]_0^x = \frac{x^3}{27} \quad 0 < x \leq 3, \text{ így } F(x) = P(X < x) = \begin{cases} 0, & \text{ha } x \leq 0 \\ \dfrac{x^3}{27}, & \text{ha } 0 < x \leq 3 \\ 1, & \text{ha } x > 3 \end{cases}$$

3.4. Feladat. Az X valószínűségi változó a $[0, c]$ intervallumon veszi fel értékeit és ott sűrűségfüggvénye $4e^{-2x}$. Határozzuk meg c értékét és annak valószínűségét, hogy $\frac{1}{4} < X < \frac{1}{2}$!

Megoldás

Mivel az eloszlásfüggvény a sűrűségfüggvény integrálja, így

$$F(x) = \int_0^x 4e^{-2t}\,dt = \left[\frac{e^{-2t}}{-2}4\right]_0^x = -2e^{-2x} + 2 \qquad 0 < x \leq c,$$

és $F(c) = 1$-ből következik, hogy

$$\begin{aligned} -2e^{-2c} + 2 &= 1 \\ e^{-2c} &= \frac{1}{2} \\ -2c &= \ln\frac{1}{2} \\ -2c &= -\ln 2 \end{aligned}$$

azaz $c = \dfrac{\ln(2)}{2} \approx 0,35$.

$$P\left(\frac{1}{4} < X < \frac{1}{2}\right) = F\left(\frac{1}{2}\right) - F\left(\frac{1}{4}\right) = 1 - (-2e^{-2\frac{1}{4}} + 2) = \frac{2}{\sqrt{e}} - 1 \approx 0,21.$$

3.5. Feladat. Véletlenszerűen választunk egy pontot az $x^2 + y^2 < 1$ kör belsejében. Jelölje Z a távolságát a középponttól. Adjuk meg Z eloszlás- és sűrűségfüggvényét valamint várható értékét!

Megoldás

Legyen Z a középponttól való távolság. Ekkor $0 \leq Z \leq 1$, így itt $F(r) = P(Z < r) = \frac{r^2\pi}{1^2\pi} = r^2$, így

$$F(r) = \begin{cases} 0, & \text{ha } r \leq 0 \\ r^2, & \text{ha } 0 < r \leq 1 \\ 1, & \text{ha } 1 < r \end{cases}$$

Ebből deriválással adódik, hogy $f(r) = F'(r) = 2r$ a $[0,1]$-en, így

$$f(r) = \begin{cases} 0, & \text{ha } r \leq 0 \text{ és } r > 1 \\ 2r, & \text{ha } 0 < r \leq 1 \end{cases}$$

$$EZ = \int_0^1 r \cdot 2r \text{ dr} = \left[\frac{2r^3}{3}\right]_0^1 = \frac{2}{3}.$$

3.6. Feladat. Legyen X sűrűségfüggvénye $\dfrac{c}{x^4}$ ha $x > 1$, és 0 különben.

a) $c = ?$

b) $EX = ?$

Megoldás

a) Mivel $1 = \displaystyle\int_1^\infty \frac{c}{x^4} dx = \lim_{t\to\infty} \int_1^t \frac{c}{x^4} dx = \lim_{t\to\infty} \left[\frac{c}{-3 \cdot x^3}\right]_1^t = \lim_{t\to\infty} \left[\frac{c}{-3 \cdot t^3} - \frac{c}{-3 \cdot 1^3}\right] = 0 + \frac{c}{3} = \frac{c}{3}$

$\left(\text{egyszerűbb jelöléssel: } 1 = \displaystyle\int_1^\infty \frac{c}{x^4} dx = \left[\frac{c}{-3 \cdot x^3}\right]_1^\infty = \frac{c}{3}\right)$, így következik, hogy $c = 3$.

b) $EX = \displaystyle\int_1^\infty x \frac{3}{x^4}\, dx = \left[\frac{-3}{2 \cdot x^2}\right]_1^\infty = 1,5$

3.7. Feladat. Tapasztalatok szerint az út hossza, amit egy bizonyos típusú robogó megtesz az első meghibásodásáig exponenciális eloszlású valószínűségi változó. Ez a távolság átlagosan 6000 km. Mi a valószínűsége annak, hogy egy véletlenszerűen kiválasztott robogó

a) kevesebb, mint 4000 km megtétele után meghibásodik?

b) több, mint 6500 km megtétele után hibásodik meg?

c) 4000 km-nél több, de 6000 km-nél kevesebb út megtétele után hibásodik meg?

d) Legfeljebb mekkora utat tesz meg az első meghibásodásig a robogók leghamarabb meghibásodó 20%-a?

Megoldás

Legyen X az első meghibásodásig megtett út. Ekkor $X \sim Exp(\lambda)$, ahol $\lambda = \dfrac{1}{6000}$.

a) $P(X < 4000) = 1 - e^{-\frac{1}{6000}4000} \approx 0,4866$

b) $P(X > 6500) = 1 - P(X < 6500) = e^{-\frac{1}{6000}6500} \approx 0,3385$

c) $P(4000 < X < 6000) = P(X < 6000) - P(X < 4000) = 1 - e^{-\frac{1}{6000}6000} - (1 - e^{-\frac{1}{6000}4000}) \approx 0,1455$

d) $0,2 = P(X < c) = 1 - e^{-\frac{1}{6000}c}$, azaz $0,8 = e^{-\frac{1}{6000}c}$, amiből $c = -6000\ln(0,8) \approx 1338,86$.

3.8. Feladat. Egy tehén napi tejhozamát normális eloszású valószínűségi változóval, $m = 22,1$ liter várható értekkel és $\sigma = 1,5$ liter szórással, modellezzük.

a) Mi annak a valószínűsége, hogy egy adott napon a tejhozam 23 és 25 liter közé esik?

b) Mekkora valószínűséggel esik a napi tejhozam $m - \sigma$ es $m + \sigma$ közé?

$(\Phi(0,6) = 0,7257, \Phi(1,93) = 0,9732, \Phi(1) = 0,8413)$

Megoldás

Legyen X a napi tejhozam. Ekkor $X \sim N(22,1; 1,5^2)$.

a) $P(23 < X < 25) = P(X < 25) - P(X < 23) = P\left(\frac{X-m}{\sigma} < \frac{25-m}{\sigma}\right) - P(X < 23) = P\left(\frac{X-22,1}{1,5} < \frac{25-22,1}{1,5}\right) - P(X < 23) = \Phi\left(\frac{25-22,1}{1,5}\right) - \Phi\left(\frac{23-22,1}{1,5}\right) = \Phi(1,93) - \Phi(0,6) = 0,9732 - 0,7257 = 0,2475$.

b) $P(m-\sigma < X < m+\sigma) = P(X < m+\sigma) - P(X < m-\sigma) = \Phi\left(\frac{(m+\sigma)-m}{\sigma}\right) - \Phi\left(\frac{(m-\sigma)-m}{\sigma}\right) = \Phi(1) - \Phi(-1) = \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1 = 2 \cdot 0,8413 - 1 = 0,6826$.

3.9. Feladat. 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 egy termék meghibásodásának ideje. Ekkor $X \sim N(10, 2^2)$

$$0,1 = P(X < c) = P\left(\frac{X-10}{2} < \frac{c-10}{2}\right) = \Phi\left(\frac{c-10}{2}\right)$$

$$c = 2 \cdot \Phi^{-1}(0,1) + 10 = 2 \cdot (-\Phi^{-1}(0,9)) + 10 = -2 \cdot 1,28 + 10 = 7,44.$$

Standard normáslis eloszás eloszlásfüggvényének értékei: `http://www.cs.elte.hu/~kovacsa/stdnormelo.pdf`

3.10. Feladat. Tegyük fel, hogy egy populációban az intelligenciahányados (IQ) normális eloszlású 110 várható értékkel és 10 szórással. Mi a valószínűsége, hogy egy véletlenszerűen kiválasztott ember IQ-ja 120 feletti?
/ $\Phi(1) = 0,8413$ /

Megoldás

Legyen X egy véletlenszerűen kiválasztott ember IQ-ja. Ekkor $X \sim N(110, 10^2)$.

$$P(X > 120) = 1 - P(X < 120) = 1 - P\left(\frac{X-110}{10} < 1\right) = 1 - \Phi(1) = 1 - 0,8413 \approx 16\%$$

3.11. Feladat. Legyen X sűrűségfüggvénye $\frac{c}{x^4}$ ha $1 < x$, és 0 különben. Mi a c konstans értéke és mennyi D^2X?

Megoldás

$$1 = \int_1^\infty \frac{c}{x^4} dx = \left[\frac{cx^{-3}}{-3}\right]_1^\infty = 0 - \left(-\frac{c}{3}\right) = \frac{c}{3}, \text{ így } c = 3$$

$$EX = \int_1^\infty x\frac{3}{x^4} dx = \int_1^\infty 3x^{-3} dx = \left[-\frac{3}{2}x^{-2}\right]_1^\infty = 0 - \left(-\frac{3}{2}\right) = \frac{3}{2}$$

Mivel $E(g(x)) = \int_{-\infty}^{\infty} g(x)f(x)\,dx$, így

$$EX^2 = \int_1^\infty x^2\frac{3}{x^4} dx = \int_1^\infty 3x^{-2} dx = \left[-3x^{-1}\right]_1^\infty = 0 - (-3) = 3$$

$$D^2X = EX^2 - E^2X = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4}$$

3.12. Feladat. Legyen X egyenletes eloszlású az $[1,4]$ intervallumon Számítsuk ki $(X-1)^2$ várható értékét!

Megoldás

Ha $X \sim Egyenletes[1,4]$, akkor $Y = X - 1 \sim Egyenletes[0,3]$. Ekkor

$$E(X-1)^2 = EY^2 = \int_0^3 y^2 \frac{1}{3}\, dy = \frac{1}{3}\left[\frac{y^3}{3}\right]_0^3 = 3$$

Más megoldás:

$$E(X-1)^2 = D^2(X-1) + E^2(X-1) = \frac{(3-0)^2}{12} + \left(\frac{0+3}{2}\right)^2 = \frac{3}{4} + \frac{9}{4} = 3$$

3.13. Feladat. Legyen X és Y független valószínűségi változók mindkettő 0 várható értékkel és 1 szórással. Legyen $W = X - Y$. Számítsa ki W várható értékét és szórását!

Megoldás

$EW = EX - EY = 0$ és $DW = \sqrt{D^2X + D^2Y} = \sqrt{2}$

3.14. Feladat. Adjon meg véges sok értéket felvehető (X) ill. végtelen sok értéket felvehető (Y) diszkrét valószínűségi változókat melyeknek szórása 1!

Megoldás

Például: Legyen $P(X=-1) = \frac{1}{2}, P(X=1) = \frac{1}{2}$ ill. $Y \sim Poisson(1)$.

3.15. Feladat. Legyen $X \sim N(2, \sqrt{5}^2)$ és $Y \sim N(5, 3^2)$ függetlenek és legyen $W = 3X - 2Y + 1$. Számítsa ki
a) EW-t és D^2W-t, ill.
b) $P(W \leq 6)$-ot!
$(\Phi(1) = 0,8413)$

Megoldás

a) $EW = 3EX - 2EY + 1 = 6 - 10 + 1 = -3$ és
$D^2W = D^2(3X - 2Y) = D^2(3X) + D^2(-2Y) = 3^2D^2X + (-2)^2D^2Y = 9D^2X + 4D^2Y = 45 + 36 = 81$
b) Mivel független normális eloszlású valószínűségű változók összege is normális eloszlású, és $3X \sim N(6,\, 3^2 \cdot \sqrt{5}^2)$ továbbá $-2Y \sim N(-10,\, (-2)^2 \cdot 3^2)$, így $W \sim N(-3, 9^2)$.

$$P(W \leq 6) = P\left(\frac{W-(-3)}{9} < \frac{6-(-3)}{9}\right) = \Phi(1) = 0,8413$$

3.16. Feladat. Legyen X egy véges szórású valószínűségi változó és legyen $a, b \in \mathbb{R}$.
a) Mutassa meg, hogy $aX + b$ és X kovarianciája egyenlő a-szor X szórásnégyzetével!
b) Számolja ki $aX + b$ és X korrelációját $(a \neq 0)$!

Megoldás

a)

$$cov(aX+b, X) = cov(aX, X) + cov(b, X) = acov(X, X) = aD^2(X)$$

b)

$$corr(aX+b, X) = \frac{cov(aX+b, X)}{D(aX+b)DX} = \frac{aD^2X}{\sqrt{a^2}DXDX} = \begin{cases} 1, & \text{ha } a > 0 \\ -1, & \text{ha } a < 0 \end{cases}$$

3.17. Feladat. Legyen X és Y független valószínűségi változók, melyre $D^2X < \infty$ és $D^2Y < \infty$.
a) Mutassa meg, hogy $X + Y$ és X kovarianciája egyenlő X szórásnégyzetével!
b) Számolja ki $X + Y$ és X korrelációját!

Megoldás

a)

$$cov(X+Y, X) = E(X+Y)X - E(X+Y)EX = EX^2 + E(YX) - E^2X - EYEX =$$
$$= EX^2 - E^2X + E(YX) - EYEX = cov(X, X) + cov(Y, X) = D^2(X)$$

b)

$$corr(X+Y, X) = \frac{cov(X+Y, X)}{D(X+Y)DX} = \frac{D^2X}{\sqrt{D^2X + D^2Y}DX} = \frac{DX}{\sqrt{D^2X + D^2Y}}$$

3.18. Feladat. Tegyük fel, hogy egy tábla csokoládé tömege normális eloszlású 100g várható értékkel és 3g szórással. Legalább hány csokoládét csomagoljunk egy dobozba, hogy a dobozban levő táblák átlagos tömege legalább 0,9 valószínűséggel nagyobb legyen 99,5 g-nál, ha feltételezzük, hogy az egyes táblák tömege egymástól független? ($\Phi(1{,}28) = 0{,}8997$)

Megoldás

Legyen X egy tábla csokoládé tömege, $X \sim N(100, 3^2)$. Ekkor n tábla csokoládé átlagos tömege $\overline{X} \sim N(100, \frac{9}{n})$, mivel $D^2(\overline{X}) = D^2\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} D^2(X_i) = \frac{n \cdot 9}{n^2} = \frac{9}{n}$.

$$0{,}9 = P(\overline{X} > 99{,}5) = 1 - P(\overline{X} < 99{,}5) = 1 - P\left(\frac{\overline{X}-100}{\frac{3}{\sqrt{n}}} < \frac{-0{,}5\cdot\sqrt{n}}{3}\right) = 1 - \Phi\left(\frac{-\sqrt{n}}{6}\right) = \Phi\left(\frac{\sqrt{n}}{6}\right)$$

Mivel tudjuk, hogy $\Phi(1{,}28) = 0{,}8997 \approx 0{,}9$, így $1{,}28 = \frac{\sqrt{n}}{6}$. Ebből következik, hogy $n = (6 \cdot 1{,}28)^2 = 58{,}9$, azaz legalább 59 csokit kell egy dobozba csomagolni.

3.19. Feladat. Egy scannelt kép átlagos mérete 600KB, 100KB szórással. Mi a valószínűsége, hogy 80 ilyen kép együttesen 47 és 48MB közötti tárhelyet foglal el, ha feltételezzük, hogy a képek mérete egymástól független?
($\Phi(1{,}12) = 0{,}8686$)

Megoldás

Jelölje X egy kép eloszlását $\mu = 600$KB várható értékkel és $\sigma = 100$KB szórással. Legyen S_n n db ilyen valószínűségi változó összege ($n = 80$). A centrális határeloszlás tétel szerint

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \to Z \text{ ha } n \to \infty, \text{ ahol } Z \sim N(0,1).$$

Tehát

$$P(47000 \le S_n \le 48000) = P\left(\frac{47000 - 80 \cdot 600}{\sqrt{80} \cdot 100} \le \frac{S_n - n\mu}{\sqrt{n}\sigma} \le \frac{48000 - 80 \cdot 600}{\sqrt{80} \cdot 100}\right) \approx$$

$$\approx P(-1,12 \le Z \le 0) = \Phi(0) - \Phi(-1,12) = 0,5 - (1 - \Phi(1,12)) = 0,5 - (1 - 0,8686) = 0,3686 = 36,9\%$$

3.20. Feladat. Egy szoftver frissítéséhez 68 file-t kell installálni, amik egymástól függetlenül 10mp várható értékű és 2mp szórású ideig töltődnek.
a) Mi a valószínűsége, hogy a teljes frissítés lezajlik 12 percen belül?
b) A cég a következő frissítésnél azt ígéri, hogy az már 95% valószínűséggel 10 percen belül betöltődik. Hány file-ból állhat ez a frissítés?
($\Phi(2,42) = 0,992, \Phi(1,645) = 0,95$)

Megoldás

Legyen X egy fájl telepítési ideje $\mu = 10$ mp várható értékkel és $\sigma = 2$ mp szórással. Jelölje S_n n db fájl telepítési idejének az összegét ($n = 68$).
a) Használva a Centrális Határeloszlás Tételt,

$$P(\text{teljes frissítés lezajlik 12 percen belül}) = P(S_n < 720) = P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} < \frac{720 - 680}{2\sqrt{68}}\right) \approx \Phi(2,42) = 99,2\%$$

b)

$$0,95 = P(S_n < 600) = P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} < \frac{600 - 10n}{2\sqrt{n}}\right) \approx \Phi\left(\frac{600 - 10n}{2\sqrt{n}}\right)$$

Mivel tudjuk, hogy $\Phi(1,645) = 0,95$, így $1,645 = \dfrac{600 - 10n}{2\sqrt{n}}$, vagyis

$$\begin{aligned}
3{,}29\sqrt{n} &= 600 - 10n \ \ / \ y := \sqrt{n} \\
3{,}29y &= 600 - 10y^2 \\
10y^2 + 3{,}29y - 600 &= 0 \\
&\to y_1 = 7{,}58, y_2 = -7{,}91 \\
y = \sqrt{n} &\ge 0 \Rightarrow \sqrt{n} = 7{,}58
\end{aligned}$$

Így következik, hogy $n = 57{,}51$, azaz legfeljebb 57 fájlból állhat a frissítés.

3.21. Feladat. Legyen egy X pozitív valószínűségi változó várható értéke $EX = 3$ és szórása $DX = 3$. Számítsuk ki, hogy legfeljebb mekkora valószínűséggel vesz fel a változó 13-at vagy annál nagyobb értéket! Mennyi a valószínűség pontos értéke, ha feltesszük, hogy az eloszlás exponenciális?

Megoldás

Markov-egyenlőtlenséggel: $P(X \geq 13) \leq \dfrac{EX}{13} = \dfrac{3}{13} \approx 0{,}23$
A Csebisev-egyenlőtlenséget $\varepsilon = 10$ értékre használva

$$P(X \geq 13) = P(X - 3 \geq 13 - 3) = P(X - 3 \geq 10) \leq P(|X - 3| \geq 10) \leq \frac{D^2X}{10^2} = \frac{9}{100} = 0,09$$

Ha X exponenciális eloszlású, akkor eloszlásfüggvénye $F(x) = 1 - e^{-\frac{1}{3}x}$, így

$$P(X \geq 13) = 1 - P(X < 13) = 1 - (1 - e^{-\frac{13}{3}}) = e^{-\frac{13}{3}} = 0,013$$

3.22. Feladat. Egy elektromos vezetékgyártó cég 40 m-es vezetékeket gyárt $0,2$ m szórással. Legfeljebb mennyi annak a valószínűsége, hogy a vezeték hossza legalább 1 m-rel eltér a várható 40 m-es értéktől?

Megoldás

A Csebisev-egyenlőtlenséget $\varepsilon = 1$ értékre használva

$$P(|X - 40| \geq 1) \leq \frac{D^2X}{1^2} = \frac{0,2^2}{1^2} = 0,04$$

Vagyis legfeljebb 0,04 annak a valószínűsége, hogy a vezeték rövidebb, mint 39 m ill. hosszabb, mint 41 m.

4. (7-8 hét) Leíró statisztikák, statisztikai alapfogalmak: becslések (maximum likelihood, momentum)

Feladatok

4.1. Feladat. Legyen $X_1, \ldots, X_n$ független, azonos eloszlású valószínűségi változók m várható értékkel. Célunk az ismeretlen m paraméter becslése. Tekintsük az alábbi statisztikákat és állapítsuk meg, hogy melyek torzítatlanok! Amelyik nem torzítatlan, hogyan tudnánk torzítatlanná tenni?
$T_1(\mathbf{X}) = X_8, \qquad T_2(\mathbf{X}) = \dfrac{X_9 + X_{19}}{9}, \qquad T_3(\mathbf{X}) = \overline{X}$

Megoldás

$E(T_1(\mathbf{X})) = E(X_8) = m$, így T_1 torzítatlan
$E(T_2(\mathbf{X})) = E\left(\dfrac{X_9 + X_{19}}{9}\right) = \dfrac{E(X_9) + E(X_{19})}{9} = \dfrac{2m}{9}$, így T_2 nem torzítatlan, viszont $\dfrac{9}{2}T_2$ már igen
$E(T_3(\mathbf{X})) = E(\overline{X}) = E\left(\dfrac{1}{n}\sum_{1=1}^{n} E(X_i)\right) = m$, így T_3 torzítatlan

4.2. Feladat. Adjon torzítatlan becslést a független, azonos $E[0, \vartheta]$ eloszlású $X_1, \ldots, X_n$ minta ϑ paraméterére a mintaátlag segítségével!

Megoldás

Mivel $X_1, \ldots, X_n \sim E[0, \vartheta]$, így $E(X_i) = \frac{\vartheta}{2}$. Ekkor $E(\overline{X}) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{n}{n}E(X_1) = \frac{\vartheta}{2}$, tehát $E(2\overline{X}) = \vartheta$, vagyis $2\overline{X}$ torzítatlan becslése ϑ-nak.

4.3. Feladat. Legyen az alábbi gyakorisági tábla egy 20 elemű minta, a következő diszkrét eloszlásból:
$P(X_i = -1) = c, P(X_i = 1) = 3c, P(X_i = 2) = 1 - 4c$ $\left(i = 1, \ldots 20\right.$ és c az ismeretlen paraméter, $\left.0 < c < \frac{1}{4}\right)$.

érték	-1	1	2
gyakoriság	4	10	6

Határozza meg c ML-becslését és c becslését a momentum módszerrel!

Megoldás

c ML-becslése:

$$L(c, \mathbf{x}) = P(X_1 = x_1, \ldots, X_{20} = x_{20}) = c^4(3c)^{10}(1 - 4c)^6$$

$$lnL(c, \mathbf{x}) = 4\ln(c) + 10\ln(3c) + 6\ln(1 - 4c)$$

$$(lnL(c, \mathbf{x}))'_c = \frac{4}{c} + \frac{10}{c} - \frac{6 \cdot 4}{1 - 4c}$$

Átrendezve a $(lnL(c, \mathbf{x}))'_c = 0$ egyenletet, kapjuk, hogy

$$\frac{14}{c} - \frac{24}{1 - 4c} = 0$$
$$14(1 - 4c) - 24c = 0$$
$$14 = 80c$$

így $\hat{c} = \dfrac{7}{40} = \dfrac{21}{120}$. Ez valóban maximum, mivel $(lnL(c, \mathbf{x}))''_c$-t kiértékelve a $\hat{c}$ helyen $(lnL(c; \mathbf{x}))''_c = -\dfrac{14}{c^2} - \dfrac{96}{(1 - 4c)^2} < 0$.
c becslése momentum-módszerrel:

$$M_1(c) = EX = -1 \cdot c + 1 \cdot 3c + 2 \cdot (1 - 4c) = 2 - 6c, \qquad m_1 = \frac{1}{20}(-1 \cdot 4 + 1 \cdot 10 + 2 \cdot 6) = 0,9$$

így az $M_1(c) = m_1$ egyenletet c-re megoldva kapjuk, hogy $\hat{c} = \dfrac{2 - 0,9}{6} = \dfrac{11}{60} = \dfrac{22}{120}$

4.4. Feladat. Legyenek $X_1, X_2, \ldots, X_n$ független azonos eloszlású valószínűségi változók az alábbi eloszlásokból. Számolja ki az ismeretlen paraméter ML-becslését!
a) $Bin(m, p)$ binomiális eloszlás, ahol $m \in \mathbb{N}$ adott és p a paraméter
b) $Exp(\lambda)$ exponenciális eloszlás
c) $N(\mu, \sigma^2)$ normális eloszlás, ahol $\sigma \in \mathbb{N}$ adott és μ a paraméter

Megoldás

(Továbbá lehet, hogy érdemes megjegyezni, hogy az $\overline{x} = m$ eset külön meggondolást igényel.)

a)

$$L(m,p;\mathbf{x}) = \prod_{k=1}^{n} \binom{m}{x_k} p^{x_k}(1-p)^{m-x_k} \qquad (x_k = 0, 1, \ldots, m)$$

$$lnL(m,p;\mathbf{x}) = \sum_{k=1}^{n} \ln \binom{m}{x_k} + \ln p \sum_{k=1}^{n} x_k + \ln(1-p) \sum_{k=1}^{n}(m - x_k)$$

$$(lnL(m,p;\mathbf{x}))'_p = \frac{1}{p}\sum_{k=1}^{n} x_k + \frac{-1}{1-p}\sum_{k=1}^{n}(m-x_k) = \frac{1}{p}\sum_{k=1}^{n} x_k + \frac{-1}{1-p}\left(nm - \sum_{k=1}^{n} x_k\right) = \frac{1}{p}n\overline{x} + \frac{-1}{1-p}\left(nm - n\overline{x}\right)$$

Átrendezve a $(lnL(m,p;\mathbf{x}))'_p = 0$ egyenletet, kapjuk, hogy

$$\begin{aligned}\frac{1}{p}n\overline{x} + \frac{-1}{1-p}(nm - n\overline{x}) &= 0\\ \frac{\overline{x}}{p} - \frac{m-\overline{x}}{1-p} &= 0\\ \overline{x} - p\overline{x} - pm + p\overline{x} &= 0\end{aligned}$$

így $\hat{p} = \frac{\overline{X}}{m}$. Ez valóban maximum, mivel $(lnL(m,p))''_p$-t kiértékelve a $\hat{p}$ helyen $(lnL(m,p;\mathbf{x}))''_p = \frac{-n\overline{x}}{p^2} + \frac{-n(m-\overline{x})}{(1-p)^2} = -n\left(\frac{\overline{x}}{p^2} + \frac{m-\overline{x}}{(1-p)^2}\right) < 0$.

b)

$$L(\lambda;\mathbf{x}) = \prod_{k=1}^{n} \lambda e^{-\lambda x_k} \qquad (x_k > 0)$$

$$lnL(\lambda;\mathbf{x}) = \sum_{k=1}^{n} \ln \lambda e^{-\lambda x_k} = \sum_{k=1}^{n} \ln \lambda + \sum_{k=1}^{n} \ln e^{-\lambda x_k} = n\ln\lambda - \lambda\sum_{k=1}^{n} x_k = n\ln\lambda - \lambda n\overline{x}$$

$$(lnL(\lambda;\mathbf{x}))'_\lambda = \frac{n}{\lambda} - \sum_{k=1}^{n} x_k = \frac{n}{\lambda} - n\overline{x}$$

Átrendezve a $(lnL(\lambda;\mathbf{x}))'_\lambda = 0$ egyenletet, kapjuk, hogy $\hat{\lambda} = \frac{1}{\overline{X}}$. Ez valóban maximum, mivel $(lnL(\lambda))''_\lambda = -\frac{n}{\lambda^2} < 0$.

c)

$$L(\mu,\sigma^2;\mathbf{x}) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum\limits_{k=1}^{n}(x_i-\mu)^2}$$

$$lnL(\mu,\sigma^2;\mathbf{x}) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_i-\mu)^2$$

$$\left(lnL(\mu,\sigma^2;\mathbf{x})\right)'_\mu = -\frac{1}{2\sigma^2}(-2)\sum_{k=1}^{n}(x_i - \mu)$$

Átrendezve a $\left(lnL(\mu,\sigma^2;\mathbf{x})\right)'_\mu = 0$ egyenletet, kapjuk, hogy $\hat{\mu} = \frac{\sum\limits_{k=1}^{n} X_i}{n} = \overline{X}$. Ez valóban maximum, mivel $\left(lnL(\mu,\sigma^2;\mathbf{x})\right)''_\mu = -\frac{n}{\sigma^2} < 0$.

4.5. Feladat. Határozza meg az ismeretlen paraméter ML-becslését, ha a minta $E[a,1]$ eloszlású!

Megoldás

A paraméter függvényében nem deriválható a likelihood függvény (ugrik):

$$L(a;\mathbf{x}) = \prod_{i=1}^{n} \frac{1}{1-a} I(a \leq x_i \leq 1) = \frac{1}{(1-a)^n} I(a \leq x_1, x_2, ..., x_n \leq 1) =$$

$$= \frac{1}{(1-a)^n} I(a \le x_1^* \le ... \le x_n^* \le 1) = \frac{1}{(1-a)^n} I(a \le x_1^*) I(x_n^* \le 1)$$

Az $I(a \le x_1^*)I(x_n^* \le 1)$ rész 0 vagy 1 lehet, tehát úgy kell megválasztani a paramétereket, hogy 1 legyen: $a \le x_1^*$ és $x_n^* \le 1$ teljesüljön. Mivel a $(-\infty, x_1^*]$ intervallumon az $\dfrac{1}{(1-a)^n}$ függvény maximuma az $a = x_1^*$ pontban van, így $\hat{a} = X_1^*$.

4.6. Feladat. Legyenek $X_1, X_2, \ldots, X_n$ független azonos $E[a,b]$ eloszlású valószínűségi változók. Számolja ki az ismeretlen paraméterek becslését a momentum módszerrel!

Megoldás

$M_1(a,b) = E(X) = \frac{a+b}{2}, \qquad m_1 = \overline{x}$

$M_2(a,b) = E(X^2) = D^2(X) + E(X)^2 = \frac{(b-a)^2}{12} + \left(\frac{a+b}{2}\right)^2, \qquad m_2 = \frac{1}{n}\sum\limits_{k=1}^{n} x_k^2$

Így $M_1(a,b) = m_1$ és $M_2(a,b) = m_2$-ből kapjuk, hogy

$\frac{a+b}{2} = m_1$

$\frac{(b-a)^2}{12} + \left(\frac{a+b}{2}\right)^2 = m_2$

és ezt oldjuk meg a, b-re, először m_1 és m_2-vel kifejezve. Átrendezve a fenti adja, hogy $\frac{(b-a)^2}{12} = m_2 - m_1^2$, így

$b - a = \sqrt{12(m_2 - m_1^2)}$

$b + a = 2m_1.$

Ezeket összeadva kapjuk, hogy $b = m_1 + \sqrt{3(m_2 - m_1^2)}$ és $a = m_1 - \sqrt{3(m_2 - m_1^2)}$. Azaz a paraméterek becslése a momentum módszerrel:

$$\hat{a} = \overline{X} - \sqrt{3\left(\frac{1}{n}\sum_{k=1}^{n} X_k^2 - \overline{X}^2\right)} = \overline{X} - \sqrt{3}S_n \quad \text{és} \quad \hat{b} = \overline{X} + \sqrt{3\left(\frac{1}{n}\sum_{k=1}^{n} X_k^2 - \overline{X}^2\right)} = \overline{X} + \sqrt{3}S_n$$

5. (9-10 hét) Konfidenciaintervalluok, paraméteres próbák

Feladatok

5.1. Feladat. Legyen X_1, X_2, X_3, X_4 független azonos $N(\mu, 2^2)$ eloszlású minta. A megfigyelt értékek a következők:
14,8; 12,2; 16,8; 11,1
a) Adjon 95%-os megbízhatóságú konfidenciaintervallumot μ-re!
b) Hány elemű mintára van szükség, ha azt szeretnénk, hogy a konfidenciaintervallum legfeljebb 1,6 hosszúságú legyen?
($u_{0,975} = 1{,}96$)

Megoldás

a) Adatok: $n = 4$
$\overline{x} = \frac{14{,}8+12{,}2+16{,}8+11{,}1}{4} = 13,725$
$\sigma = 2$
$\alpha = 0{,}05$
Ekkor $u_{0,975} = 1{,}96$, így az $(1-\alpha)100\%$ megbízhatóságú konfidenciaintervallum μ-re:

$$\left(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{x} + u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = \left(13{,}725 - u_{0,975}\frac{2}{\sqrt{4}}, 13{,}725 + u_{0,975}\frac{2}{\sqrt{4}}\right) = (11{,}765; 15{,}685)$$

R-kód:
```
minta<-c(14.8, 12.2, 16.8, 11.1)
n<-length(minta)
sigma<-2
alpha<-0.05

mean(minta)-qnorm(1-alpha/2)*sigma/sqrt(n)
mean(minta)+qnorm(1-alpha/2)*sigma/sqrt(n)
```

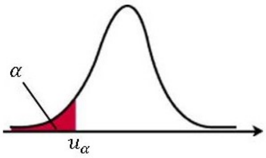
b) A konfidenciaintervallum hossza: $2u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = 2u_{0,975}\frac{2}{\sqrt{n}} < 1{,}6$, így $n > \left(\frac{4u_{0,975}}{1{,}6}\right)^2 = \left(\frac{4 \cdot 1,96}{1{,}6}\right)^2 \approx 24{,}01$
tehát legalább 25 elemű mintára van szükség.

R-kód (folytatás):
```
hossz<-1.6
(2*qnorm(1-alpha/2)*sigma/hossz)^2
```

5.2. Feladat. Azt szeretnénk vizsgálni, hogy a napi középhőmérséklet október 18-án Budapesten 15°C alatt volt-e? Az elmúlt 4 év napi középhőmérsékletei a következők voltak: $14,8; 12,2; 16,8; 11,1\ °C$, valamint tegyük fel, hogy az adatok normális eloszlásból származnak.
a) Írjuk fel a null- és ellenhipotézist!
b) Tegyük fel, hogy a napi középhőmérséklet szórása $\sigma = 2$. Tesztelje a fenti hipotézist $\alpha = 0.05$ terjedelem mellett! Adja meg a kritikus tartományt és p-értéket! Mi a döntés?
c) Tesztelje a hipotézist úgy is, hogy nem használja a szórásra vonatkozó előzetes információt!
d) Milyen hipotézist írjunk fel, ha azt szeretnénk vizsgálni, hogy a napi középhőmérséklet október 18-án Budapesten 15°C-tól különböző volt? Teszteljük a fenti adatok segítségével!
($u_{0,05} = -1{,}645,\ \Phi(1{,}275) = 0{,}899,\ t_{3;0,05} = -2{,}353\ u_{0,975} = 1,96$)

Megoldás

a) Legyen m a napi középhőmérséklet október 18-án Budapesten. Ekkor
$H_0\colon m \geq 15$
$H_1\colon m < 15$
b) Adatok: $n = 4$
$\overline{x} = \frac{14{,}8+12{,}2+16{,}8+11{,}1}{4} = 13,725$
$\sigma = 2$
$\alpha = 0,05$
Milyen próbát használjunk? Egymintás, egyoldali u-próbát.
Próbastatisztika: $U = \dfrac{\overline{X} - 15}{\frac{\sigma}{\sqrt{n}}} \overset{H_0 \text{esetén}}{\sim} N(0,1)$, melynek értéke: $u = \dfrac{13,725 - 15}{\frac{2}{\sqrt{4}}} = -1,275$
U mely értékeire utasítjuk el H_0-t?
H_0 esetén $P(U < u_\alpha) = \Phi(u_\alpha) = \alpha$, azaz $\Phi(u_{0,05}) = 0,05$ tehát $u_{0,05} = -1,645$ így a



kritikus tartomány $= \{\mathbf{x} \in \chi : U < u_\alpha\} = \{\mathbf{x} \in \chi : U < -1,645\}$.
Mivel most $u = -1,275 > -1,645$, nem utasítjuk el H_0-t. Azaz nincs elég bizonyítékunk, hogy a napi középhőmérséklet október 18-án Budapesten 15°C alatt lenne.

A hipotézist a p-érték α-val való összehasonlításával is tesztelhetjük:
p-érték $= \Phi(-1,275) = 1 - \Phi(1,275) = 1 - 0,899 = 0,101 > \alpha = 0,05$, így nem vetjük el H_0-t.

c) Adatok: $n = 4$

$\overline{x} = \frac{14,8+12,2+16,8+11,1}{4} = 13,725$

$s_n^* = \sqrt{\frac{(14,8-13,725)^2+(12,2-13,725)^2+(16,8-13,725)^2+(11,1-13,725)^2}{3}} = \sqrt{6,6092} = 2,57$

$\alpha = 0,05$

Milyen próbát használjunk? Egymintás, egyoldali t-próbát.

Próbastatisztika: $T = \frac{\overline{X} - 15}{\frac{s_n^*}{\sqrt{n}}} \overset{H_0 \text{esetén}}{\sim} t_{n-1}$, melynek értéke: $t = \frac{13,725 - 15}{\frac{2,57}{\sqrt{4}}} = -0,99$

T mely értékeire utasítjuk el H_0-t?
H_0 esetén $P(T < t_{n-1;\alpha}) = \alpha$, azaz $P(T < t_{3;0,05}) = 0,05$ tehát $t_{3;0,05} = -2,353$ így a kritikus tartomány $= \{\mathbf{x} \in \chi : T < t_{n-1;\alpha}\} = \{\mathbf{x} \in \chi : T < -2,353\}$.
Mivel most $t = -0,99 > -2,353$, nem utasítjuk el H_0-t. Azaz nincs elég bizonyítékunk, hogy a napi középhőmérséklet október 18-án Budapesten 15°C alatt lenne.

A hipotézist a p-érték α-val való összehasonlításával is tesztelhetjük:
p-érték $= P(t_3 < -0,99) = 0.198 > \alpha = 0,05$, így nem vetjük el H_0-t.

d) Legyen m a napi középhőmérséklet október 18-án Budapesten. Ekkor
$H_0 : m = 15$
$H_1 : m \neq 15$

Adatok: $n = 4$

$\overline{x} = \frac{14,8+12,2+16,8+11,1}{4} = 13,725$

$\sigma = 2$

$\alpha = 0,05$

Milyen próbát használjunk? Egymintás, kétoldali u-próbát.

Próbastatisztika: $U = \frac{\overline{X} - 15}{\frac{\sigma}{\sqrt{n}}} \overset{H_0 \text{esetén}}{\sim} N(0,1)$, melynek értéke: $u = \frac{13,725 - 15}{\frac{2}{\sqrt{4}}} = -1,275$

U mely értékeire utasítjuk el H_0-t?
H_0 esetén $P(|U| > u_{1-\frac{\alpha}{2}}) = \alpha$, azaz $P(U < u_{0,975}) = 0,975$ tehát $u_{0,975} = 1,96$ így a kritikus tartomány $= \{\mathbf{x} \in \chi : |U| > u_{1-\frac{\alpha}{2}}\} = \{\mathbf{x} \in \chi : |U| > 1,96\}$.
Mivel most $|u| = 1,275 < 1,96$, nem utasítjuk el H_0-t. Azaz nincs elég bizonyítékunk, hogy a napi középhőmérséklet október 18-án Budapesten 15°C-tól különböző lenne.

A hipotézist a p-érték α-val való összehasonlításával is tesztelhetjük:
p-érték $= 2 \cdot (1 - \Phi(1,275)) = 0.202 > \alpha = 0,05$, így nem utasítjuk el H_0-t.

A hipotézist a várható értékre vonatkozó 95%-os konfidenciaintervallum segítségével is tesztelhetjük:
A konfidenciaintervallum $(11,765; 15,685)$ (Feladat 1.) tartalmazza a 15-öt, így nem utasítjuk el H_0-t.

5.3. Feladat. Az alábbi két minta két különböző gyáregységben tapasztalt selejtarányra vonatkozik (ezrelékben). Állítható-e, hogy az „A" gyáregység jobban dolgozott? (Feltételezhetjük, hogy a minták normális eloszlásúak, függetlenek.)

A	11,9	12,1	12,8	12,2	12,5	11,9	12,5	11,8	12,4	12,9
B	12,1	12,0	12,9	12,2	12,7	12,6	12,6	12,8	12,0	13,1

$(F_{9,9;0,975} = 4{,}026,\ \ t_{18;0,05} = -1{,}734)$

Megoldás

Jelölje m_A az „A" és m_B az „B" gyáregység selejtarányát. Ekkor

$H_0\colon m_A \geq m_B$
$H_1\colon m_A < m_B$

Adatok: $n_A = 10, n_B = 10$
$\overline{x}_A = \frac{11{,}9+\dots+12{,}9}{10} = 12,3$
$\overline{x}_B = \frac{12{,}1+\dots+13{,}1}{10} = 12,5$
$s_A^{*2} = \frac{(11{,}9-12{,}3)^2+\dots+(12{,}9-12{,}3)^2}{9} = \frac{132}{900} = 0,147$
$s_B^{*2} = \frac{(11{,}9-12{,}5)^2+\dots+(12{,}9-12{,}5)^2}{9} = \frac{142}{900} = 0,158$
$\alpha = 0,05$

Van különbség a szórások közt? Előzetes F-próbával tesztelünk.

$H_0\colon \sigma_A^2 = \sigma_B^2$
$H_1\colon \sigma_A^2 \neq \sigma_B^2$

$F = \frac{s_B^{*2}}{s_A^{*2}} = \frac{142}{132} = 1,076 <$ kritikus érték $= F_{9,9;0,975} = 4,026$, tehát nem utasítjuk el H_0-t, így nincs rá okunk, hogy a két szórást különbözőnek tekintsük.

Milyen próbát használjunk? Kétmintás, egyoldali t-próbát.
$T = \sqrt{\frac{n_A n_B}{n_A+n_B}} \frac{\overline{x}_A-\overline{x}_B}{\sqrt{\frac{(n_A-1)s_A^{*2}+(n_B-1)s_B^{*2}}{n_A+n_B-2}}} \overset{H_0 \text{ esetén}}{\sim} \mathrm{t}_{n+m-2}$, melynek értéke: $t = \sqrt{\frac{10\cdot 10}{10+10}} \frac{12{,}3-12{,}5}{\sqrt{\frac{9\cdot 0{,}147+9\cdot 0{,}158}{10+10-2}}} = -1,13$

T mely értékeire utasítjuk el H_0-t?
H_0 esetén $P(T < t_{n_A+n_B-2;\alpha}) = \alpha$, azaz $P(T < t_{18;0{,}05}) = 0,05$ tehát $t_{18;0{,}05} = -1,734$ így a kritikus tartomány $= \{\mathbf{x} \in \chi : T < t_{n_A+n_B-2;\alpha}\} = \{\mathbf{x} \in \chi : T < -1,734\}$.
Mivel most $t = -1,13 > -1,734$, nem utasítjuk el H_0-t, azaz nincs elég bizonyítékunk arra, hogy az „A" gyáregység jobban dolgozott.

5.4. Feladat. Két szervert hasonlítottunk össze. Az elsőn 30 futás átlagos ideje 6,7 mp volt, míg ettől függetlenül a másodikon 20 futásé 7,2 mp. Vizsgáljuk meg, hogy van-e szignifikáns különbség a két szerver sebessége közt, ha a futási idők szórása mindkét gépen 0,5 volt?
($u_{0{,}975} = 1{,}96$)

Megoldás

Jelölje m_1 és m_2 az első illetve a második szerveren való futás idejét. Ekkor

$H_0\colon m_1 = m_2$
$H_1\colon m_1 \neq m_2$

Adatok: $n_1 = 30, n_2 = 20$
$\overline{x}_1 = 6.7$
$\overline{x}_2 = 7.2$
$\sigma_1 = \sigma_2 = 0.5$

Milyen próbát használjunk? Kétmintás, kétoldali u-próbát (szórások ismertek).
$U = \frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \overset{H_0 \text{esetén}}{\sim} N(0,1)$, melynek értéke: $u = \frac{6.7-7.2}{\sqrt{\frac{0.5^2}{30}+\frac{0.5^2}{20}}} = -3.464$

U mely értékeire utasítjuk el H_0-t?
H_0 esetén $P(|U| > u_{1-\frac{\alpha}{2}}) = \alpha$, azaz $P(U < u_{0{,}975}) = 0,975$ tehát $u_{0{,}975} = 1,96$ így a kritikus tartomány $= \{\mathbf{x} \in \chi : |U| > u_{1-\frac{\alpha}{2}}\} = \{\mathbf{x} \in \chi : |U| > 1,96\}$.
Mivel most $|u| = 3,464 > 1,96$, elutasítjuk H_0-t, azaz a két szerver futási ideje közt szignifikáns különbség van.

5.5. Feladat. Az alábbi két minta 10 forgalmas csomópont levegőjében található szennyezőanyag koncentrációra vonatkozó két adatsort tartalmaz. Az első sorban a november 15-i, a másodikban a november 29-i számok szerepelnek. Szignifikánsan változott-e a légszennyezettség?

november 15.	20,9	17,1	15,8	18,8	20,1	15,6	14,8	24,1	18,9	12,5
november 29.	21,4	16,7	16,4	19,2	19,9	16,6	15,0	24,0	19,2	13,2

($t_{9;0{,}975} = 2{,}262$)

Megoldás

Jelölje m_1 és m_2 a november 15-i illetve a november 29-i légszennyeződés várható értékét. Ekkor

$H_0\colon m_1 = m_2$

$H_1 : m_1 \neq m_2$

Mivel ugyanazokon a helyeken mérték a légszennyezettséget, a minták páronként összetartozóak (egymástól nem független megfigyeléseink vannak). A légszennyeződés változására vonatkozó információ a két mérési eredmény különbségében rejlik.

november 29. - november 15.	0,5	-0,4	0,6	0,4	-0,2	1,0	0,2	-0,1	0,3	0,7

Jelöljük m-mel a november 29-én és a november 15-én mért légszennyeződés várható értékének különbségét, azaz $m = m_2 - m_1$. Ekkor a fenti hipotézis a következőképpen fogalmazható meg:

$H_0 : m = 0$
$H_1 : m \neq 0$

Adatok: $n = 10$

$\overline{x} = \frac{0{,}5+\cdots+0{,}7}{10} = 0{,}3$

$s_n^* = \sqrt{\frac{(0{,}5-0{,}3)^2+\cdots+(0{,}07-0{,}3)^2}{10-1}} = 0{,}435$

$\alpha = 0{,}05$

Milyen próbát használjunk? Egymintás, kétoldali t-próbát.

$$T = \frac{\overline{X} - 0}{\frac{s_n^*}{\sqrt{n}}} \overset{H_0 \text{ esetén}}{\sim} t_{n-1}, \text{ melynek értéke: } t = \frac{0{,}3 - 0}{\frac{0{,}435}{\sqrt{10}}} = 2{,}18$$

T mely értékeire utasítjuk el H_0-t?

H_0 esetén $P(|T| > t_{n-1;1-\frac{\alpha}{2}}) = \alpha$, azaz $P(T < t_{9;0,975}) = 0,975$ tehát $t_{9;0,975} = 2{,}262$ így a kritikus tartomány $= \{\mathbf{x} \in \chi : |T| > t_{n-1;1-\frac{\alpha}{2}}\} = \{\mathbf{x} \in \chi : |T| > 2{,}262\}$.

Mivel most $t = 2{,}18 < 2{,}262$, nem utasítjuk el H_0-t, azaz nincs elég bizonyítékunk, hogy különbség lenne a november 15-i és 29-i légszennyeződés mértéke közt.

Viszont vegyük észre, hogy a próbastatisztika értéke közel van a kritikus értékhez. Ezt a p-érték α-hoz közeli értékéből is látjuk: p-érték $= P(|t_9| > 2{,}18) = 0{,}057$. Ez utóbbi azt mutatja, hogy az $\alpha = 0{,}05$ szignifikanciaszinten nem utasítjuk el a nullhipotézist, viszont egy $0{,}057$-nél magasabb szinten már igen.