Valószínűségszámítás
és
matematikai statisztika

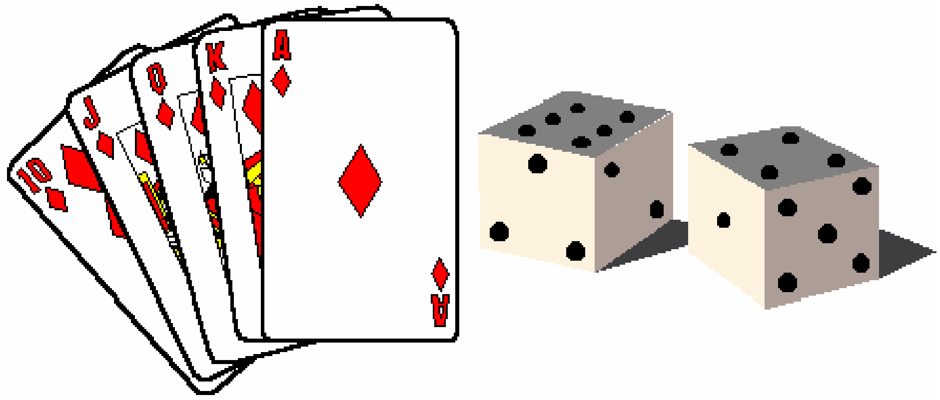
Ketskeméty László

Budapest, 1996

Tartalomjegyzék

I. fejezet

VALÓSZÍNŰSÉGSZÁMÍTÁS



1. Kombinatorikai alapfogalmak

A véges elemszámú halmazok tulajdonságaival foglakozik a kombinatórika. Az alábbiakban egy n elemű halmazból képezhető egyéb halmazok elemszámának meghatározásával fogunk foglalkozni. A képzett halmazok számosságához, azaz elemeinek száma meghatározásához közvetett módszereket fogunk megtanulni. Eredményeinket majd a valószínűség klasszikus kiszámítási módjánál fogjuk felhasználni.

Definíció: n különböző elemből álló halmaz önmagára való kölcsönösen egy-egyértelmű (bijektív) leképezéseit *ismétlés nélküli permutációknak* nevezzük. A permutáció nem más, mint az n különböző elem egy sorrendje. Két permutáció különbözik egymástól, ha valamelyik sorszámú helyeiken más-más elemek állnak.

Tétel: Egy ismétlés nélküli permutációt egyértelműen megadunk, ha az 1,2,...,n természetes számok valamilyen sorrendjét vesszük.

Tétel: Az összes különböző ismétlés nélküli permutációk száma n!=1·2·...·n. (n! n-faktoriális.)

Bizonyítás: Amikor elkészítünk egy sorrendet, az első helyre n elem közül választhatunk, a másodikra (mivel az első helyre egyet már választottunk) n-1 közül. Az első két helyet tehát n·(n-1) féleképpen képezhetjük. A harmadik helyre már csak n-2 lehetőségünk marad: ennyiféleképp folytathatjuk a permutáció felírását, stb. Tehát, ha már i elemet elrendeztem a sorrendbe, n-i féleképpen folytathatom a sort. Ebből már következik az állítás.

Példa: Amikor egy 32 lapos magyar kártyát megkeverünk, a kártyacsomag egy permutációját képezzük. Összesen $32!\sim 2.63\cdot 10^{35}$ sorrend lehetséges.

Definíció: Ha az n elemű halmazban $k_1, k_2, \cdots, k_m$ darab azonosnak tekintett elem van , $\left(k_1 + k_2 + \cdots + k_m = n\right)$ akkor a halmaz önmagára való bijektív *leképezései ismétléses permutációk* lesznek.

Tétel: Az összes különböző ismétléses permutációk száma $\dfrac{n!}{k_1!\,k_2!\cdots k_m!}$.

Bizonyítás: Ha egy adott ismétléses permutációban az azonos elemeket különbözőknek tekintenénk, az azonos elemek egymás közötti sorrendjéből más és más ismétlés nélküli permutációk lennének készíthetők, összesen $k_1!\cdot k_2!\cdots k_m!$ darab.

Példa: Egy 104 darabszámú dupla francia kártyacsomagban minden lapból két példány van. Ezért itt az összes megkülönböztethető permutációk száma: $\dfrac{104!}{(2!)^{52}}$.

Definíció: n különböző elemből álló halmaz egy k elemszámú részhalmazának egy ismétlés nélküli permutációját, az n elem egy k-adosztályú *ismétlés nélküli variációjának* nevezzük.

Tétel: n elem összes különböző k-adosztályú ismétlés nélküli variációinak száma:

$$n\cdot(n-1)\cdots(n-k+1)=\frac{n!}{(n-k)!}=\binom{n}{k}\cdot k!$$

Bizonyítás: Ha egy k-adosztályú variációt elkészítünk, az első helyet n-féleképpen, a másodikat (n-1)-féleképpen , stb. a k-adik helyet (n-k+1)-féleképp választhatjuk.

Példa: A magyar 18 tagú labdarúgó bajnokságból csak három csapat indulhat a nemzetközi kupákért. Elvileg 18·17·16= 4896 variáció képzelhető el.

Definíció: Tekintsünk egy olyan $n{\cdot}k$ elemű halmazt ahol n különböző elemből rendre k darabot azonosnak veszünk. Ezen halmaz összes k elemszámú részhalmazainak ismétléses permutációi az n különböző elem k-adosztályú ismétléses variációi. ($k>n$ is lehet !)

Tétel: n elem összes különböző k-adosztályú ismétléses variációinak száma n^k .

Bizonyítás: Amikor egy ilyen ismétléses variációt elkészítünk, a k hely mindegyikére az n különböző elem bármelyikét tehetjük.

Példa: Amikor egy totó szelvényt kitöltünk, az 1,2,X elemekből álló 3 elemű halmaznak egy k=14 elemű ismétléses variációját képezzük. Összesen tehát 3^{14} = 4 782 969 kitöltési variáció lehetséges.

Definíció: n különböző elemből álló halmaz egy k elemű részhalmaza, az n elem egy k-*adosztályú ismétlés nélküli kombinációja*.

Tétel: Az n elem összes különböző k-adosztályú ismétlés nélküli kombinációinak száma : $\dbinom{n}{k}=\dfrac{n!}{k!\cdot(n-k)!}$.

Bizonyítás: Az n elem ismétlés nélküli variációi, és kombinációi között az a különbség, hogy a kombinációnál a k-elemű részhalmaz elemeinek sorrendjeit nem képezzük. Tehát, egy adott k-adrendű kombinációból, az elemek sorrendjének felcserélésével k! különböző k-adosztályú variáció képezhető, ami már igazolja az állítást.

Példa: Amikor egy hagyományos (ötöt a kilencvenből azaz ötös-) lottószelvényt kitöltünk, a 90 szám egy 5-ödosztályú ismétlés nélküli kombinációját képezzük. Az összes kitöltési kombinációk száma: $\binom{90}{5} = 43\,949\,268$.

Definíció: Tekintsünk egy olyan $n{\cdot}k$ elemű halmazt ahol n különböző elemből rendre k darabot azonosnak veszünk. Ezen halmaz k elemszámú részhalmazait az n különböző elem *k-adosztályú ismétléses kombinációinak nevezzük*. ($k>n$ is lehet !)

Tétel: Az n különböző elem összes különböző k-adosztályú ismétléses kombinációinak száma: $\binom{n+k-1}{k}$.

Bizonyítás: Megmutatjuk, hogy n+k-1 különböző elem k-adosztályú ismétlés nélküli kombinációi és n különböző elem k-adosztályú ismétléses kombinációi között kölcsönösen egy-egyértelmű (bijektív) leképezés adható meg, ami már igazolja az állítást.
Tekintsük a sorszámozott n+k-1 különböző elem egy tetszőleges k-adosztályú ismétlés nélküli kombinációja elemeinek sorszámait természetes sorrendben:

$$i_1 < i_2 < \ldots < i_k \quad (i_\alpha \neq i_\beta \text{ és } i_\alpha \in \{1,2,\ldots,n+k-1\}) .$$

Ha most végrehajtjuk a $j_\alpha = i_\alpha - (\alpha - 1)$ transzformációt, k darab olyan sorszámot kapunk, amellyel egyértelműen azonosíthatjuk n különböző elem egy k-adosztályú ismétléses kombinációját:
$j_1 \leq j_2 \leq \cdots \leq j_k$ ahol bármely indexnél $j_\alpha = j_\beta$ lehet és $j_\alpha \in \{1,2,\ldots n\}$.
Mivel a végrehajtott transzformáció bijektív, az állításunkat bebizonyítottuk.

Példa: Egy analitikus háromváltozós függvénynek elvileg $\binom{3+5-1}{5} = 21$ darab ötödrendű vegyes parciális derivált függvénye lehet.

Ellenőrző kérdések és gyakorló feladatok

1. Hány különböző sorrendje lehet *n* elemnek?
2. Mit nevezünk *n* elem *k*-adosztályú ismétlés nélküli kombinációjának?
3. Mennyi *n* elem *k*-adosztályú ismétléses variációinak száma?
4. Hogyan számoljuk ki az *n!* (*n* faktoriális) számot?

1. Hogyan számoljuk ki az $\binom{n}{k}$ binomiális együtthatót?
2. Mit értünk n elem k-adosztályú ismétléses variációján?

7. Döntse el, az alábbi állítások közül melyik igaz, melyik hamis!
   a. Amikor n elem k-adosztályú ismétléses kombinációjáról beszélünk, $k>n$ is lehet.
   b. Az n elem k-adosztályú ismétléses kombinációinak száma több, mint a k-adosztályú ismétléses variációk száma.
   c. A lottóhúzások számát ismétlés nélküli kombinációval lehet meghatározni.
   d. Ha egy kombinációban két elemet felcserélünk, egy másik kombinációt kapunk.
   e. Ha egy ismétléses variációban két különböző elemet felcserélünk, egy másik ismétléses variációt kapunk.
   f. Az n elem k-adosztályú ismétlés nélküli kombinációinak a száma megegyezik *az* (n-k)-adosztályú ismétlés nélküli kombinációinak a számával. ($k \neq n$-k).
   g. Az n elem k-adosztályú ismétlés nélküli variációinak a száma megegyezik az (n-k)-adosztályú ismétlés nélküli variációinak a számával. ($k \neq n$-k).
   h. Az n elem k-adosztályú ismétléses kombinációinak a száma megegyezik n-k+1 elem k-adosztályú ismétlés nélküli kombinációinak a számával.
   i. Az n elem k-adosztályú ismétlés nélküli kombinációinak a száma megegyezik, az olyan n elemű ismétléses permutációk számával, ahol k illetve n-k elem azonos.
   j. A kenószelvény kitöltésekor egy ismétlés nélküli kombinációt adunk meg.
   k. A totószelvény kitöltésekor egy ismétlés nélküli variációt adunk meg.

8. A VALÓSZÍNŰSÉGSZÁMÍTÁS szó betűiből hány különböző húszkarakteres betűsorozat képezhető?
9. Hány különböző háromtalálatos szelvény képzelhető el elvileg a ötös lottószelvények között?
10. Hány különböző 10 találatos szelvény képzelhető el a 13+1 mérkőzéses totószelvények között?
11. A Morse ABC ti (.) és tá (-) jeleiből mennyi különböző legfeljebb 10 hosszúságú jel kódolható?
12. Hányféleképpen lehet elhelyezni 15 különböző postaládába
    a. két különböző levelet?
    b. két azonos reklámcédulát?

    (Az is lehetséges, hogy mindkét levél illetve reklámcédula ugyanabba a postaládába kerül.)
13. Tíz számozott dobozba hányféleképpen helyezhetek el három különböző színű golyót? (Egy dobozba több golyó is kerülhet, a dobozon belül a sorrendet nem lehet megállapítani.)
14. Tíz egyforma játékkockával dobva, hány különböző eredményt kaphatunk?
15. Öt színből hány trikolór (háromszínű) vízszintes sávos zászló készíthető?
16. Feladatunk, hogy órarendet készítsünk. A hét első öt napjának első hat órájában lehetnek csak tanórák. A heti óraszámok: matematika 5, magyar 4, testnevelés, biológia, földrajz, fizika, történelem 2, ének, rajz, osztályfőnöki 1. Hányféleképpen lehet elvileg elkészíteni az órarendet, ha lyukasóra is elképzelhető?
17. Igazolja, hogy
    a. $\binom{n}{0}+\binom{n}{1}+\cdots+\binom{n}{n}=2^n$

b. $\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - + \cdots + (-1)^n \binom{n}{n} = 0$

c. $\binom{n}{0} - \binom{n}{1} \cdot 2 + \binom{n}{2} \cdot 2^2 \cdots + (-1)^n \binom{n}{n} \cdot 2^n = (-1)^n$

2. A valószínűségszámítás alapfogalmai és axiómarendszere

Az alapfogalmak a szemléletből eredő, magától értetődő fogalmakat jelentenek, amelyeket egyszerűbb fogalmak segítségével nem lehet definiálni, hanem csupán körülírni lehet őket, illetőleg példákat lehet mutatni rájuk.
Hasonlóan, az axiómák bizonyítás nélkül elfogadott tételek, amelyek annyira nyilvánvalóak, hogy csupán a szemléletből vezetjük le őket.

Alapfogalom: *Véletlen kísérleten* (K) olyan folyamatot, jelenséget értünk, amelynek kimenetele előre bizonyosan meg nem mondható, de az igen, hogy elvileg milyen módon fejeződhet be, azaz előre tudható, hogy milyen végállapotok lehetnek. A véletlen kísérletet azonos feltételek mellett, függetlenül meg lehet figyelni, vagy végre lehet hajtani akárhányszor.

Példa: a.) Egy szabályos játékkockát feldobunk. Nem tudjuk előre megmondani az eredményt, de azt állíthatjuk, hogy az 1,2,3,4,5,6 érték közül valamelyiket kapjuk.
b.) Egy csomagból véletlenszerűen kihúzunk 8 lapot. A véletlentől függ, hogy melyik lesz az a 8 lap, de azt tudjuk, hogy a 32 lap összes ismétlés nélküli kombinációja közül lehet csak valamelyik.
c.) Egy telefonkészüléket figyelve mérjük két hívás között eltelt időt. A lehetséges kimenetelek a $[0,\infty)$ intervallum pontjai.
d.) Egy jutalomsorsoláson kihúzott személy kora szintén a véletlentől függ. Előre csak annyi állítható, hogy a kor nyilván pl. 200-nál kisebb szám lesz.
e.) Addig dobálunk egy szabályos játékkockát, amíg 6-ost nem kapunk. Azt persze nem lehet előre biztosan megmondani, hogy a hatoshoz hány dobásra lesz szükség, de azt biztosan tudjuk, hogy a 0,1,2,... (nemnegatív egész) számok valamelyike fog bekövetkezni.

Alapfogalom: A K véletlen kísérlettel kapcsolatos *eseménynek* nevezünk minden olyan logikai állítást, melynek igaz vagy hamis értéke egyértelműen megállapítható a kísérlet befejeződésekor. Az esemény *bekövetkezik* , ha az állítás igaz értéket kap a kísérlet végén, és *nem következik be*, ha a logikai érték hamis. Az eseményeket az abc nagybetűivel fogjuk jelölni: A,B,C,...

Példa: a.)A kockadobás kísérletével kapcsolatos esemény a „párosat dobunk”. Nem tekinthető eseménynek viszont a „Fradi nyeri a bajnokságot” logikai állítás.
b.)A kártyahúzás kísérlethez tartozó esemény pl. az, hogy „van négy piros a lapok között”, de nem esemény a „megnyerhető a piros ulti” állítás.
c.)A telefonhívások közötti időtartamra vonatkozó kísérlethez tartozó esemény az „öt percen belül csengeni fog”, de nem esemény a „Pista fog telefonálni” állítás.
d.)A jutalomsorsoláson „a nyertes fiatalabb mint 20” esemény, „ a nyertes szép ember” pedig nem esemény.
e.)A „nem kell 20 dobásnál több a hatoshoz” állítás esemény, míg a „a kocka nem szabályos” állítás nem esemény.

Definíció: Az A *esemény maga után vonja* a B eseményt, ha az A esemény bekövetkezéséből, már a B esemény bekövetkezése is következik. Jelölés: A⊆B.

Példa: a.)Kockadobásnál a „hatosat dobunk” esemény maga után vonja a „párosat dobunk” eseményt
.b.)„A nyolc pirosat húzunk” esemény maga után vonja a „kihúzott lapok között lesz a piros ász is” eseményt.
c.)„Az öt percen belül megszólal a telefon” esemény maga után vonja a „a tíz percen belül megszólal a telefon” eseményt.
d.)„A kihúzott személy 60 év feletti” esemény maga után vonja „a kisorsolt személy elmúlt 20 éves” eseményt.
e.)„A tíz dobáson belül dobok hatost” esemény maga után vonja a „húsz dobáson belül hatost dobunk” eseményt.

Definíció: Az A és B események *ekvivalensek*, ha A⊆B és B⊆A egyszerre. Ekvivalens események között nem teszünk különbséget.

Definíció: *Lehetetlen esemény*nek nevezzük azt a ∅-val jelölt eseményt, amely a K bármely végrehajtása során soha sem következik be, illetőleg elvileg sem következhet be.(A konstans hamis állítás.)

Definíció: *Biztos esemény*nek nevezzük azt az Ω-val jelölt eseményt, amelyik a K bármely végrehajtása során mindig bekövetkezik, mert elvileg is mindig bekövetkezik. (A konstans igaz állítás).

Példa: a.) A kockadobásnál a „10-nél kisebb értéket dobunk” esemény az Ω-val, a „negatív értéket dobunk” esemény pedig ∅-val ekvivalens.
b.) „A zöld, makk, tök vagy piros színű lapok közül lesz a leosztott nyolc között” esemény biztos esemény, „nyolc piros színű lapom és két ászom is lesz” pedig lehetetlen esemény lesz.
c.) „Negatív szám lesz az eltelt idő” lehetetlen, míg az „eltelt idő nemnegatív lesz” esemény biztos.
d.) „200 év alatti személy nyeri a sorsolást” biztos esemény, a „200-nál öregebb nyer” lehetetlen.
e.) „Egyszer valaha fogunk hatost dobni” biztos esemény, „soha sem fogunk hatost dobni” lehetetlen.

Definíció: A K véletlen kísérlet egy $A \neq \emptyset$ eseményét *elemi esemény*nek nevezzük, ha nincs olyan B esemény, amely A-t maga után vonná. Azaz $\forall B$ ($\neq \emptyset$ és $\neq A$) olyan hogy $B \not\subset A$. Az elemi eseményeket, - a többi ú.n. *összetett esemény*től való megkülönböztetésül - ω-val vagy ω_i-vel fogjuk jelölni.

Definíció: A K véletlen kísérlet összes elemi eseményének halmazát *eseménytér*nek nevezzük.

Megjegyzés: Miután az összetett események elemi események - mint állítások - diszjunkciójából állnak, az összetett eseményeket úgy is felfoghatjuk, mint a megfelelő elemi események halmazát. Ebből a szempontból, az eseménytér éppen az Ω biztos esemény lesz. Pl. kockadobásnál az ω_i =„i értéket dobok" (i=1,2,3,4,5,6) események az elemi események, az A=„3-al osztható számot dobok" esemény az $A = \{\omega_3, \omega_6\}$ halmaz, $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$ pedig a biztos esemény (eseménytér). Tehát, az események az eseménytér részhalmazaiként is elképzelhetőek.

Definíció: Egy A esemény *ellentett esemény*e az az $\overline{A}$-val jelölt esemény, ami pontosan akkor következik be, amikor A nem következik be. $\overline{A}$ az A-nak az Ω-ra vonatkoztatott komplementer halmaza.

Az A és B *események összegén* azt az A+B-vel jelölt eseményt értjük, amely pontosan akkor következik be, ha A és B közül legalább az egyik bekövetkezik. (A+B az A és B események uniója).

Az A és B *események szorzatán* azt az $A \cdot B$-vel jelölt eseményt értjük, amely pontosan akkor következik be, amikor A is és B is egyidejűleg bekövetkezik. ($A \cdot B$ az A és B események metszete).

Az A és B *események különbségén* azt az $A \setminus B$ -vel jelölt eseményt értjük, ami pontosan akkor következik be, amikor A bekövetkezik, de B nem. ($A \setminus B \equiv A \cdot \overline{B}$).

Mivel az események közötti műveletek a logikai állítások közötti diszjunkció és konjukció illetve a negáció segítségével voltak értelmezve, és ott igazak a Boole algebra összefüggései, ezért azok itt is érvényesek. A következő tételben összefoglaljuk az események műveleteinek legfontosabb tulajdonságait.

Tétel: Tetszőleges A,B és C eseményekre igazak az alábbiak:

a.) $A+B=B+A$
b.) $(A+B)+C=A+(B+C)$
c.) $A+A=A$
d.) $A\cdot B=B\cdot C$
e.) $(A\cdot B)\cdot C=A\cdot(B\cdot C)$
f.) $A\cdot A=A$
g.) $A\cdot(B+C)=(A\cdot B)+(A\cdot C)$
h.) $A+(B\cdot C)=(A+B)\cdot(A+C)$
i.) $\overline{\overline{A}} = A$
j.) $\overline{A+B} = \overline{A}\cdot\overline{B}$
k.) $\overline{A\cdot B} = \overline{A}+\overline{B}$
l.) $A\cdot\overline{A} = \emptyset$
m.) $A+\overline{A} = \Omega$
n.) $A\cdot\Omega=A$
o.) $A+\Omega=\Omega$
p.) $A\cdot\emptyset=\emptyset$
r.) $A+\emptyset=A$

Definíció: Az A és B események *egymást kizáróak*, ha $A\cdot B=\emptyset$, azaz szorzatuk a lehetetlen esemény. Egymást kizáró események egyidejűleg nem következhetnek be.

Definíció: Az $A_1, A_2, \ldots, A_n, \ldots$ (nem feltétlenül véges elemszámú) események rendszere *teljes eseményrendszert* alkot, ha $\forall$ i≠j -re $A_i \cdot A_j = \emptyset$ (páronként egymást kizárják) és $\sum_{\forall i} A_i = \Omega$ teljesül.

Megjegyzés: A K véletlen kísérlet egy végrehajtása során a teljes eseményrendszer eseményei közül csak egyikük fog biztosan bekövetkezni.

Példa: A francia kártyacsomagból való húzásnál az A_1=„kört húzok”, A_2=„kárót húzok”, A_3=„pikket húzok” és A_4=„treffet húzok” események teljes eseményrendszert alkotnak.

Axiómák: A K véletlen kísérlettel kapcsolatos összes események $\Im$ rendszere kielégíti az alábbi tulajdonságokat:

1° $\Omega \in \Im$.
2° Ha $A \in \Im \quad \Rightarrow \quad \overline{A} \in \Im$ is.
3° Ha $A_1, A_2, \ldots, A_n, \ldots \in \Im \quad \Rightarrow \quad \sum_{\forall i} A_i \in \Im$ is.

Megjegyzés: a.) $\Im$ nem feltétlenül esik egybe Ω összes részhalmazainak halmazrendszerével. $\Im$-ben csak a kísérlettel kapcsolatba hozható ú.n. megfigyelhető események vannak. Nem zárjuk ki, hogy lehetnek Ω-nak olyan A részhalmazai, amelyeket nem tudunk rendesen megfigyelni, azaz lehet olyan kimenetel, ami végén nem tudjuk megmondani, hogy A bekövetkezett-e vagy sem. Az axiómákkal éppen az ilyen kétes A eseményeket akarjuk kizárni a további vizsgálatainkból.

b.) Az axiómák nyilvánvaló tulajdonságokat fogalmaznak meg. Az 1° pontban azt követeljük meg, hogy a biztos esemény megfigyelhető legyen. A 2°-ben azt állítjuk, hogyha az A eseményt meg tudjuk figyelni, akkor az ellentettjét is meg tudjuk. A 3°-ban pedig az az állítás, hogyha eseményeknek egy rendszerét egyenként meg tudjuk figyelni, akkor azt az eseményt is meg fogjuk tudni figyelni, amely akkor következik be, ha a felsorolt események közül legalább egy bekövetkezik.

Tétel: Az axiómákból levezethetők $\Im$-nek az alábbi tulajdonságai:

a.) $\emptyset \in \Im$, azaz a lehetetlen esemény is megfigyelhető.

b.) Ha A, B $\in \Im \Rightarrow A+B \in \Im$ is, azaz a 3° axióma véges sok esetre is igaz.

c.) Ha A,B $\in \Im \Rightarrow A \cdot B \in \Im$ is, azaz megfigyelhető események szorzata is megfigyelhető.

d.) Ha $A_1, A_2, \dots, A_n, \dots \in \Im \Rightarrow \prod_{\forall i} A_i \in \Im$ is igaz, azaz megfigyelhető események együttes bekövetkezése is megfigyelhető.

e.) Ha A, B $\in \Im \Rightarrow A \setminus B \in \Im$ és $B \setminus A \in \Im$, azaz megfigyelhető események különbségei is megfigyelhetőek.

Axiómák: Adott egy $P: \Im \to [0,1]$ függvény, melyet *valószínűségnek* nevezünk. A P függvény kielégíti az alábbi tulajdonságokat:

1° $P(\Omega) = 1$

2° Ha $A_1, A_2, \dots, A_n, \dots \in \Im$ páronként egymást kizárják, azaz $\forall$ $i \neq j$ -re $A_i \cdot A_j = \emptyset$, akkor $P(\sum_{\forall i} A_i) = \sum_{\forall i} P(A_i)$.

Megjegyzés: a.) A 2° axiómában megfogalmazott tulajdonságot a valószínűség σ-additivitási (szigma additivitási) tulajdonságának nevezzük.

b.) A megfigyelhető események valószínűségeit *ismertnek* tételezzük fel. A P(A) érték az A esemény bekövetkezésének *mértéke*, esélye. Az események valószínűsége az események objektíve, *fizikailag létező* jellemzője, olyan mint pl. a testeknek a tömege vagy térfogata. Attól, hogy egy adott esetben nem tudjuk megmondani egy esemény valószínűségét, nem következik, hogy az eseménynek nincs, vagy nem egyértelmű a valószínűsége. Ha egy test tömegét nem ismerjük, vagy rosszul becsüljük a nagyságát, abból még nem lehet azt a következtetést levonni, hogy a testnek nincs tömege, vagy az nem egyértelmű. Ugyanez igaz a valószínűségre is. Ráadásul a P függvény rendelkezik azokkal a tulajdonságokkal, amikkel minden más mérték is rendelkezik (pl. hossz, terület, térfogat, tömeg stb.)
A 2° axióma azt állítja, hogy egymást át nem fedő események összegének valószínűsége az események valószínűségeinek összege, mint ahogy pl. egymást át nem fedő részekből álló síkidom területe egyenlő a részek területeinek összegével. Az 1° axióma azt posztulálja, hogy legyen a biztos esemény valószínűsége 1, és ehhez képest jellemezzük a többi esemény bekövetkezésének esélyét. A fizikai mennyiségekhez mérőműszerek szerkeszthetők, hogy az

adott test egy fizikai jellemzőjének elméleti értékét nagy pontossággal megbecsülhessük. Ilyen műszer a hosszmérésre a méterrúd, tömegre a karos mérleg. Ugyanúgy, mint más mértéknél, a valószínűség esetén is szerkeszthető „mérőműszer", amivel az elméleti valószínűség számértéke jól becsülhető lesz. Ez a mérőműszer a később értelmezendő *relatív gyakoriság* lesz. (Lásd az 5. pontot !)

Tétel: A valószínűség axiómarendszeréből levezethetőek a valószínűség alábbi tulajdonságai:

a.) $P(\overline{A}) = 1 - P(A)$
b.) $P(\emptyset)=1-P(\Omega)$
c.) Ha $A_1, A_2, \dots, A_n, \dots \in \Im$ események teljes eseményrendszert alkotnak, akkor $\sum_{\forall i} P(A_i) = 1$
d.) Ha $A \subseteq B$ akkor $P(A) \le P(B)$
e.) $P(A \backslash B)=P(B)-P(A \cdot B)$
f.) $P(A+B)=P(A)+P(B)-P(AB)$

A következő nevezetes tétel az előbbi tétel f.) állításának általánosítása kettőnél több esemény esetére.

Tétel: (*Poincare* tétel)

Ha $A_1, A_2, \dots, A_n \in \Im$ tetszőlegesek, akkor $P(\sum_{i=1}^{n} A_i) = \sum_{i=1}^{n} (-1)^{n+1} S_i^n$, ahol

$$S_i^n = \sum_{1 \le j_1 < j_2 < \dots < j_i \le n} P(A_{j_1} \cdot A_{j_2} \cdot \dots \cdot A_{j_i}).$$

Tétel: *(Boole- egyenlőtlenség)*

Legyen $(\Omega, \Im, P)$ Kolmogorov-féle valószínűségi mező. Akkor minden $A_1, A_2, \dots, A_n \in \Im$ esetén

a.) $P(\sum_{i=1}^{n} A_i) \le \sum_{i=1}^{n} P(A_i)$ és

b.) $P(\prod_{i=1}^{n} A_i) \ge 1 - \sum_{i=1}^{n} P(\overline{A}_i)$.

Ellenőrző kérdések és gyakorló feladatok

1. Mit értünk események összegén?
2. Mit értünk események szorzatán?
3. Mik a valószínűség axiómái?
4. Mit állít a Poincare tétel?
5. Mi a teljes eseményrendszer fogalma?
6. Mikor mondjuk azt, hogy az A esemény maga után vonja a B eseményt?
7. Tekintsük azt a véletlen kísérletet, hogy kihúzunk egy kártyalapot a 32 lapos magyar kártyacsomagból. Az alábbiak közül melyik esemény?

A „A kihúzott lap színe makk”
B „Nagy értékű a kihúzott kártya”
C „Nem király a kihúzott lap”
D „Szép figurájú a kihúzott lap”
E „A kihúzott lap a treff kettes”
F „A kihúzott lap nem a treff kettes”

8. Melyik esemény vonja maga után a másikat?
   A „Szabályos kockával párosat dobunk”
   B „Legalább 4-est dobunk”
   C „6-ost dobunk”
   D „Prímszámot dobunk”
9. Mely események zárják ki egymást?
   A „Két szabályos kockával dobva az összeg páros”
   B „A két dobott érték közül legalább az egyik páros”
   C „Az egyik legalább osztható hárommal”
   D „A dobott értékek szorzata páratlan”
   E „A két dobott érték közül az egyik négyszerese a másiknak”
10. Döntse el, az alábbi állítások közül melyik igaz, melyik hamis!
    a. Bármely két esemény közül az egyik maga után vonja a másik bekövetkezését.
    b. Két esemény szorzata olyan esemény, amely a két komponens esemény mindegyikét maga után vonja.
    c. Az események szorzata felcserélhető (kommutatív).
    d. Az események összeadása átzárójelezhető (asszociatív)
    e. Egy esemény az ellentettjével teljes eseményrendszert alkot.
    f. Egy esemény és az ellentettje nem egymást kizáró események.
    g. Az események összege akkor következik be, ha a komponens események valamelyike bekövetkezik.
    h. Az események szorzata akkor következik be, ha a komponens események valamelyike bekövetkezik.
    i. Az események valószínűsége lehet akár 1000 %-os is.
    j. Az események valószínűsége a véletlen kísérlet minden egyes végrehajtásakor más és más.
    k. Az ellentett esemény valószínűsége mindig nagyobb mint az esemény valószínűsége.
    l. Az ellentett esemény valószínűségének és az esemény valószínűségének összege mindig 1.
    m. Az események szorzatának a valószínűsége nem lehet nagyobb bármely komponens esemény valószínűségénél.
    n. Az események összegének a valószínűsége nem lehet nagyobb bármely komponens esemény valószínűségénél.
    o. A független események kizárják egymást.
    p. A független események nem zárják ki egymást.
    q. Két olyan független esemény, melyek közül egyik sem lehetetlen vagy biztos esemény, nem zárhatják egymást ki.
    r. Független események szorzatának valószínűsége egyenlő az események valószínűségeinek szorzatával.
    s. Független események szorzatának valószínűsége egyenlő az események valószínűségeinek összegével.

t. Egymást kizáró események szorzata a lehetetlen esemény.
u. Ha két esemény szorzatának valószínűsége nulla, akkor a két esemény kizárja egymást.
v. Egymást kizáró események összegének valószínűsége a komponens események valószínűségeinek összege.
w. A lehetetlen és a biztos események minden eseménytől függetlenek.
x. Egy esemény nem lehet független a komplementerétől.

11. A próbagyártás során két szempontból vizsgálják a késztermékeket. Az A esemény azt jelenti, hogy egy véletlenszerűen kiválasztott mintadarab anyaghibás, a B pedig az az esemény, hogy a kiválasztott gyártmány mérethibás. Tudjuk, hogy P(A)=0,15 , P(B)=0,3 és P(AB)=0,08. Mennyi annak a valószínűsége, hogy valamelyik termék hibátlan?

12. Mennyi $P(A|\overline{B})$, ha P(A)=0,6 , P(B)=0,5 és P(A+B)=0,8?

13. Egy fekete és fehér golyókat tartalmazó urnából kihúzunk n db golyót. Jelentse A_i azt az eseményt, hogy az i-edeiknek kihúzott golyó fehér ($1 \le i \le n$). Fejezzük ki az A_i események segítségével az alábbi eseményeket:

A „Mindegyik golyó fehér”
B „Legalább egy golyó fehér”
C „Pontosan egy golyó fehér”
D „Mindegyik golyó ugyanolyan színű”

14. Bizonyítsa be, hogy tetszőleges A,B eseményekre

$$\left(P(AB)\right)^2 + \left(P(A\overline{B})\right)^2 + \left(P(\overline{A}B)\right)^2 + \left(P(\overline{A}\,\overline{B})\right)^2 \ge 0{,}25 .$$

15. Ketten sakkoznak. Az A esemény akkor következik be, ha a világossal játszó nyer, a B esemény akkor, ha a sötéttel játszó másik, reminél pedig a C esemény következik be. Fogalmazzuk meg szavakban, mit jelentenek az alábbi események:

a. $AB + \overline{A}\,\overline{B}$
b. $\overline{A}\,\overline{B}$
c. A+C

16. Egy céltábla tíz koncentrikus körből áll és a sugarakra fennáll az $R_1 < R_2 < \cdots < R_{10}$ reláció. A_k azt az eseményt jelenti, hogy egy lövés az R_k sugarú körbe esik. Fogalmazzuk meg szavakban, mit jelentenek az alábbi események:

$B = A_1 + A_3 + A_6$
$C = A_2 A_4 A_6 A_8$
$D = \left(A_1 + A_3\right) A_6$

17. Tegyük fel, hogy A és B olyan események, melyre P(A)=P(B)=0,5. Bizonyítsa be, hogy ekkor $P(AB) = P(\overline{A}\,\overline{B})$!

18. Bizonyítsa be, hogy $P(\overline{A}B + A\overline{B}) = P(A) + P(B) - 2P(AB)$

19. Ha az A és B események közül az egyik feltétlenül bekövetkezik, $P(A|B) = \frac{2}{3}$, $P(B|A) = \frac{1}{3}$, mennyi a P(A) és P(B) valószínűség?

20. Legyen $P(A) = \frac{2}{3}$, $P(A|B) = \frac{2}{3}$, $P(B|A) = \frac{1}{3}$. Határozza meg a P(A+B) és $P(\overline{A}|\overline{B})$ valószínűségeket!

3. A klasszikus valószínűségi mező

Ekkor az eseménytér véges elemszámú elemi esemény halmaza: $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, az $\Im$ eseményosztály Ω összes részhalmazainak rendszere, és mindegyik elemi esemény bekövetkezésének egyforma a valószínűsége: $P(\{\omega_1\}) = P(\{\omega_2\}) = \cdots = P(\{\omega_n\})$. Mivel az összes elemi események rendszere teljes eseményrendszert alkot, ezért

$$1 = P(\Omega) = P(\sum_{i=1}^{n} \{\omega_i\}) = n \cdot P(\{\omega_1\} \Rightarrow p_i = P(\{\omega_i\}) = \frac{1}{n} \quad \forall\, i\text{-re.}$$

Így, ha $A \subseteq \Omega$ tetszőleges esemény, akkor $P(A) = \sum_{\omega \in A} P(\{\omega\}) = \frac{1}{n} \sum_{\omega \in A} 1 = \frac{k_A}{n}$, ahol k_A az A esemény számossága. Vagyis az események valószínűsége ilyenkor úgy számítható, hogy az esemény bekövetkezése szempontjából kedvező elemi események számát osztjuk a kísérlettel kapcsolatos összes elemi események számával.

Klasszikus valószínűségi mezővel modellezhető a kockadobás, a pénzfeldobás, a rulettezés, a kártyahúzás, a lottóhúzás, a totótippelés stb.

Feladat (De Méré lovag feladványa)
Melyik eseménynek nagyobb a valószínűsége: hogy „egy kockával négyszer dobva legalább egyszer hatost dobunk" (A), vagy annak, hogy „két kockával huszonnégyszer dobva legalább egyszer két hatosunk lesz" (B)?

Megoldás: Két különböző valószínűségi mezőről van szó. Az elsőben egy szabályos kockát négyszer feldobunk. Az összes elemi események száma n=6^4. A vizsgált A esemény ellentettje az az esemény, hogy egyszer sem dobunk hatost. Ilyen eset összesen 5^4 lehet, vagyis az ellentett esemény valószínűsége: $P(\overline{A}) = \left(\frac{5}{6}\right)^4$. Így az A esemény valószínűsége: 1-$\left(\frac{5}{6}\right)^4 \approx 0{,}5177472\ldots$ A második vizsgált esemény egy egészen más kísérlethez és eseménytérhez tartozik. Most a véletlen kísérlet az, hogy két szabályos kockát dobunk fel 24-szer. Az összes elemi esemény most sokkal több: 36^{24}. A második esemény ellentettje most az, hogy a dobássorozatban egyszer sem dobunk duplán hatost. Ennek a valószínűsége $P(\overline{B}) = \left(\frac{35}{36}\right)^{24}$. A második esemény valószínűsége így P(B)=1-$\left(\frac{35}{36}\right)^{24} \approx 0{,}4914049\ldots$

Látható, hogy az A esemény valószínűsége a nagyobb.

Megjegyzés: A feladatot De Méré lovag adta fel Blaise Pascal francia matematikusnak, aki ebből kiindulva jutott el a valószínűségszámítás első komoly eredményeihez. A feladatban egyébként első pillantásra az tűnik fel, hogy mindkét esemény esetében a dobások számának és a lehetséges kimenetelek számának aránya azonos: A-nál 4:6, a B-nél 24:36.

Feladat Egy urnából, ahol fehér és fekete golyók vannak, véletlenszerűen kiveszünk visszatevéssel két golyót. Bizonyítsuk be, hogy annak a valószínűsége, hogy a golyók ugyanolyan színűek, nem lehet kisebb mint 0,5.

Megoldás: Legyen a fehér golyók száma n, a feketéké m (n,m $\geq$1). Ekkor a véletlen kísérlet elemi eseményeinek száma $(n+m)^2$, a kedvező eseteké pedig n^2+m^2. A keresett valószínűség: p=$\frac{n^2+m^2}{(n+m)^2}$. Mivel $(n-m)^2 \geq 0$, így $2n^2+2m^2 \geq n^2+2nm+m^2$, azaz p$\geq$0,5.

Feladat (Pólya-féle urnamodell)
Egy urna r darab fekete és s darab fehér golyót tartalmaz. Véletlenszerűen kihúzunk egy golyót. A kihúzott golyót és még plusz c darab ugyanolyan színű golyót visszateszünk az urnába. Mennyi a valószínűsége annak, hogy az n-edik húzás után α-szor húztuk ki a fekete, és β-szor a fehér golyót? (α+β=n).

Megoldás: Pl. annak az eseménynek a valószínűsége, hogy az első α húzáskor mindig fekete és az utolsó β húzáskor pedig csupa fehér golyót fogunk húzni:
$\frac{r(r+c)(r+2c)(r+3c)\cdots\left(r+(\alpha-1)c\right)s(s+c)(s+2c)\cdots\left(s+(\beta-1)c\right)}{(r+s)(r+s+c)(r+s+2c)(r+s+3c)\cdots\left(r+s+(n-1)c\right)}$. De minden más olyan húzássorozatnak, ahol α-szor húztuk ki a fekete, és β-szor a fehér golyót is ugyanekkora a valószínűsége. A különböző kimenetelek száma $\binom{n}{\alpha}$, így a keresett valószínűség:

$$\binom{n}{\alpha}\frac{r(r+c)(r+2c)(r+3c)\cdots\left(r+(\alpha-1)c\right)s(s+c)(s+2c)\cdots\left(s+(\beta-1)c\right)}{(r+s)(r+s+c)(r+s+2c)(r+s+3c)\cdots\left(r+s+(n-1)c\right)}.$$

Feladat Ha egy szabályos pénzérmét n-szer feldobunk, mennyi a valószínűsége, hogy k-val többször fogunk fejet kapni, mint írást? (0$\leq$k$\leq$n).

Megoldás: Ha a fejdobások számát f, az írásokét i jelöli, fenn kell állnia, hogy f+i=n és f-i=k. Innen következik, hogy $2f=n+k$ és $2i=n-k$, vagyis n és k paritásának meg kell egyeznie. Annak valószínűsége, hogy egy n hosszúságú dobássorozatban éppen f fejet dobunk $\binom{n}{f}\left(\frac{1}{2}\right)^n=\binom{n}{\frac{n+k}{2}}\left(\frac{1}{2}\right)^n$. Ugyanis, minden n hosszúságú sorozat egyformán $\left(\frac{1}{2}\right)^n$ valószínűségű, és ezek között $\binom{n}{f}$ olyan különböző dobássorozat lehet, ahol a fejek száma éppen f (kedvező esetek).

Gyakorló feladatok

1. Egy minden oldalán befestett fakockát a lapokkal párhuzamos síkokban 1000 azonos méretű kis kockára fűrészelnek szét. A kapott kis kockákból véletlenszerűen kiválasztunk egyet. Mennyi a valószínűsége, hogy a kockának éppen k oldala festett? (0$\leq$k$\leq$3).
2. Egy kalapban az angol ABC 26 betűje van. Visszatevéssel 11-szer húzva, a kihúzott betűket sorban egy papírra felírva, mennyi a valószínűsége, hogy a kapott szóból legfeljebb két betűt felcserélve éppen a STATISZTIKA szó jön ki?

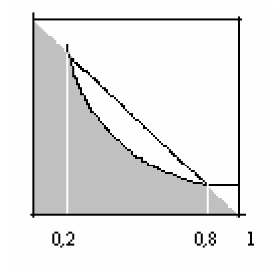
3. Egy szabályos érmével n-szer dobva, mennyi a valószínűsége, hogy a fejdobások száma páratlan lesz?
4. Egy szabályos érmével n-szer dobva, mennyi a valószínűsége, hogy
    a. először az n-edikre jön fej?
    b. ugyanannyi fejet dobunk, mint írást?
    c. pontosan két fejet dobunk?
    d. legalább két fejet dobunk?
5. Egy kalapban három cédula van, amelyekre az 1,2,3 számjegyek vannak felírva. Véletlenszerűen egyesével kihúzzuk a cédulákat. Mennyi a valószínűsége annak, hogy a húzáskor lesz olyan cédula, amelyikre éppen az a szám van felírva, ahányadikként kihúztuk azt?
6. Feldobunk három szabályos pénzérmét. Mennyi a valószínűsége az A,B,C eseményeknek, ahol A: „legalább két érmével fejet dobunk", B: „pontosan két érmével fejet dobunk", C: „legfeljebb két érmével fejet dobunk" ?
7. A ötös lottóhúzás előtt mennyi a valószínűsége, hogy k=1,2,3,4,5 találatunk lesz?
8. Egy urnában fehér és fekete golyók vannak, melyeket egymás után visszatevés nélkül kihúzunk. Az A vagy a B eseménynek nagyobb-e a valószínűsége, ahol A: „az első golyó fehér" , és B: „az utolsó golyó fehér" ?
9. Ha n egyforma ládába elhelyezünk n egyforma golyót úgy, hogy bármely ládába ugyanolyan valószínűséggel tesszük bármelyik golyót, mennyi a valószínűsége annak, hogy mindegyik ládában lesz golyó?
10. Egy 52 lapos francia kártyacsomagból 13 lapot találomra visszatevés nélkül kihúzunk. Mennyi a valószínűsége annak, hogy
    a. a treff király a kihúzott lapok között lesz?
    b. pontosan két treff lesz a leosztott lapok közt?
    c. a treff király és a treff ász a kihúzott lapok közt van?
    d. van treff a leosztott lapok között?

4. Geometriai valószínűségi mező

Alkosson a K véletlen kísérlet elemi eseményeinek halmaza egy véges mértékű geometriai alakzatot, vagy legalábbis, lehessen kölcsönösen egy-egyértelmű leképezést létesíteni Ω pontjai és egy geometriai alakzat pontjai között. Ilyenkor az $\Im$ eseményrendszer a geometriai alakzat mérhető részhalmazait jelenti, és az A esemény valószínűségét a $P(A) = \dfrac{\mu(A)}{\mu(\Omega)}$ módon számítjuk, ahol μ a geometriai térnek megfelelő mértéket jelöli. Ha pl. Ω intervallum, akkor μ hosszmérték, ha Ω síkidom, akkor μ területmérték, ha Ω test, akkor μ térfogatmérték stb.

Feladat Ha x és y két véletlenül választott 0 és 1 közé eső szám, akkor mennyi annak a valószínűsége, hogy $x+y<1$ és $xy < 0{,}16$ lesz ?

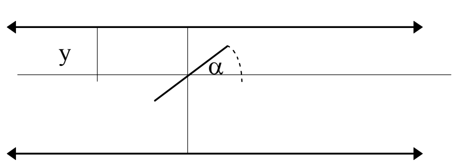
Megoldás: Ω most az egységnégyzet lesz, az kérdéses esemény pedig az ábrán besatírozott területnek felel meg:



A besatírozott terület nagysága: $\int_{0,2}^{0,8} \frac{0{,}16}{x}\,dx + 0{,}2 = 0{,}42.$

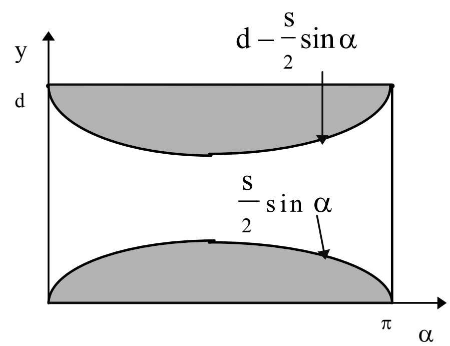
Feladat (A Buffon-tű probléma, 1777)
Egy szobában egymástól d távolságban párhuzamosan padlórések futnak. Leejtve egy s<d hosszúságú tűt, mekkora a valószínűsége, hogy a tű éppen egy padlórést fog metszeni.

Megoldás: A tű helyzetét egyértelműen a felezőpontjának a felső padlóréstől vett y távolságával és a padlórések irányával bezárt α szögével jellemezzük. Azokkal a körülményekkel, hogy melyik két rés által meghatározott sávba esik a középpont, és hogy a párhuzamosokra merőleges faltól milyen messze van a középpont nem foglalkozunk, mert a „tű metszi a padlórést” esemény bekövetkezésére ezek nincsenek hatással.



Nyilván $0\le y \le d$ és $0\le\alpha\le\pi$. A tű leejtése után y és α egyértelműen meghatározható, vagyis a véletlen kísérlet elemi eseményei azon (y,α) pontpárok, melyek elemei a $[0,d]$ és $[0,\pi]$ intervallumok által meghatározott téglalapnak. (Ez a téglalap az Ω eseménytér).

Metszés egyszerre csak egy padlórésnél következhet be, mert s<d. A metszés csak akkor következhet be, ha $0 \le y \le \frac{s}{2}\sin\alpha$, vagy hogyha $(d-y) \le \frac{s}{2}\sin\alpha$ teljesül. A feltételeknek megfelelő (y,α) pontpárok tartományát az alábbi ábrán besatíroztuk:



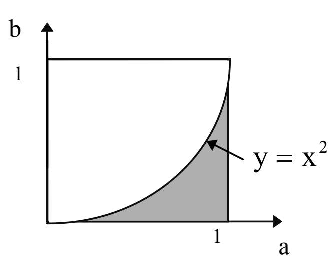
A sötétített terület nagysága $T = 2\int_0^{\pi} \frac{s}{2}\sin\alpha \, d\alpha = s[-\cos\alpha]_0^{\pi} = 2s$, a téglalap területe pedig $d\pi$.

Így a keresett valószínűség: $P(\text{"A tű metszi a padlórést"}) = \frac{2s}{d\pi}$.

Megjegyzés: Mivel a valószínűség kapcsolatos π-vel, lehetőség van statisztikus eszközökkel a π becslésére. Ha nagyon sokszor végrehajtjuk a véletlen kísérletet, és számoljuk a metszések bekövetkezését, azaz a vizsgált esemény gyakoriságát, akkor ezt a kísérletek számával elosztva (relatív gyakoriság) a fenti valószínűséget jól lehet közelíteni. Ebből π-t kifejezve kapjuk a közelítést. 1885-ben Stephan Smith angol matematikus 3200-szer végrehajtva a kísérletet, π-re 3,1553 -at kapott.

Feladat Válasszunk ki egy pontot véletlenszerűen az egységnégyzetben, melynek koordinátáit jelölje (a,b). Tekintve a $p(x) = ax^2 - 2bx + 1$ polinomot, mekkora a valószínűsége annak, hogy a p(x)=0 egyenletnek van valós gyöke?

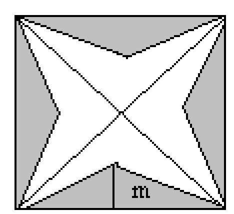
Megoldás: Egy polinomnak akkor van valós gyöke, ha a diszkriminánsa pozitív, azaz $D = 4b^2 - 4a \ge 0$. Innen következik, hogy a véletlenszerűen kiválasztott pont koordinátái között fenn kell állnia a $b^2 > a$ relációnak. Ennek megfelelő tartományt az egységnégyzetben besötétítettük:



A besötétített tartomány területe megegyezik a keresett valószínűséggel, mivel az egységnégyzet területe 1. Így $P(\text{" Van valós gyök" }) = \int_0^1 x^2\, dx = \frac{1}{3}$.

Feladat Válasszunk ki egy pontot véletlenszerűen az egységnégyzetben, melynek koordinátáit jelölje (a,b). Mekkora a valószínűsége annak, hogy a pont közelebb van a négyzet egy oldalához, mint egy átlójához?

Megoldás: Egymást metsző egyenesektől egyenlő távolságra fekvő pontok mértani helye az egyenesek szögének felező egyenese. Az oldalegyenesek és az átló egyeneseinek szögfelezői az oldalegyenesekkel 22,5°-os szöget zárnak be. A vizsgált esemény pontjai ezért az oldalak és a szögfelezők által határolt tartományba esnek:

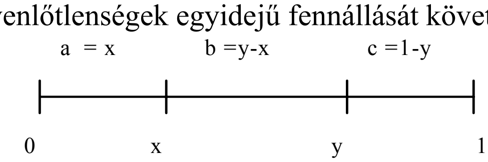


Az ábrán jelölt magasságvonal $m = \frac{1}{2}\operatorname{tg} 22{,}5^{o}$. A besötétített terület most is a keresett valószínűséggel egyezik meg:

$$P(\text{" a pont közelebb van az oldalhoz" }) = T = 4\frac{m \cdot 1}{2} = \operatorname{tg} 22{,}5^{o} = \sqrt{2} - 1 .$$

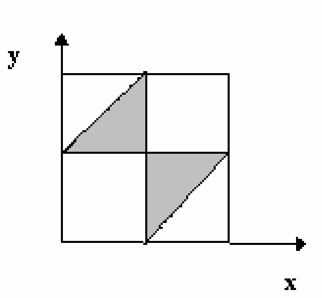
4.5 Példa Az egységintervallumban véletlenszerűen kijelölve két pontot, mekkora a valószínűsége, hogy a keletkező három szakaszból háromszög szerkeszthető?

Megoldás: Jelöljük a két pontnak a 0-tól vett távolságait rendre x-szel és y-nal. Az (x,y) pár ilyenkor egy pontot határoz meg az egységnégyzetben, ami tehát most is a véletlen kísérlethez tartozó Ω eseménytér. A háromszög szerkesztéséhez a keletkező három szakasz a,b,c hosszainak ki kell elégítenie egyidejűleg az a+b≤c, a+c≤b és b+c≤a egyenlőtlenségeket.
Az x<y esetben a három szakasz az a=x,b=y-x és c=1-y. Így a háromszög szerkeszthetősége az alábbi egyenlőtlenségek egyidejű fennállását követeli meg x,y,z-től:



$$x + (y - x) \ge 1 - y \Leftrightarrow y \ge 0{,}5$$
$$x + (1 - y) \ge y - x \Leftrightarrow y \le x + 0{,}5$$
$$(y - x) + (1 - y) \ge x \Leftrightarrow x \le 0{,}5 .$$

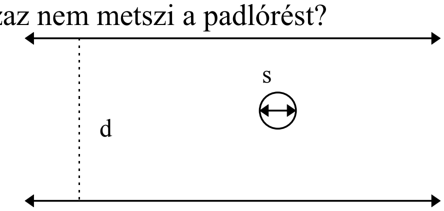
Az y≤x esetben a fenti egyenlőtlenségeknek a x≥0,5, x-0,5≤y és y≤0,5 rendszer fog megfelelni. A két kritériumrendszerhez tartozó tartományt besötétítettük az egységnégyzetben:



Így a keresett valószínűség 0,25 lesz.

Gyakorló feladatok

1. Egy szobában egymástól d távolságban párhuzamosan padlórések futnak. Leejtve egy s<d átmérőjű pénzdarabot, mennyi a valószínűsége, hogy a pénz éppen egy padlódeszka belsejébe esik, azaz nem metszi a padlórést?



2. Egy d=10 cm oldalhosszúságú négyzetrácsos padlózatra leejtünk egy s=3cm átmérőjű pénzdarabot.
    a. Mennyi a valószínűsége, hogy a pénz teljes terjedelmével egy négyzet belsejébe fog esni?
    b. Mennyi a valószínűsége, hogy hússzor végrehajtva a kísérletet, az esemény éppen ötször következik be?
3. Egy d=10 cm oldalhosszúságú négyzetrácsos padlózatra leejtünk egy s=3cm hosszú tűt. Mennyi a valószínűsége, hogy a tű teljes egészében egy négyzet belsejébe kerül?
4. Egy a=1, b=2 oldalhosszúságú téglalapon kiválasztunk egy pontot. Mennyi a valószínűsége, hogy a pont közelebb van egy csúcshoz, mint a középponthoz?
5. Ketten megbeszélik, hogy de. 10 és 11 óra között egy meghatározott helyen találkoznak. Megállapodás szerint, aki korábban érkezik 20 percet vár a másikra, és csak azután távozik. Mennyi a találkozás valószínűsége, ha mindketten véletlenszerűen érkeznek?
6. Egy egységnyi hosszúságú szakaszon találomra választunk két pontot. Mennyi a valószínűsége annak, hogy ezek közelebb vannak egymáshoz, mint bármelyik végponthoz?
7. Egy ötemeletes házban az emeletek között 6 m távolság van, a földszint és az első emelet között 8m. Ha a liftajtó 2m, mennyi a valószínűsége annak, hogy a lift megakadásakor az ajtót teljes egészében fal takarja?
8. Az ABCD egységnégyzeten véletlenszerűen kiválasztva egy pontot, mennyi a valószínűsége, hogy a pont közelebb lesz a négyzet középpontjához, mint az AB oldalhoz?

5. A feltételes valószínűség és az események függetlensége

Definíció: Tekintsünk egy K véletlen kísérletet! Legyen $A \in \Im$ egy esemény. Ha az A esemény bekövetkezéseit figyeljük a K véletlen kísérletet olyan n-szeres azonos körülmények közötti végrehajtása során amikor az egyes megfigyelések eredményei egymást nem befolyásolhatják, egy *n-szeres Bernoulli -féle kísérletsorozatról* van szó.
Ha egy n-szeres Bernoulli-féle kísérletsorozatban az A esemény k_A -szor következett be, akkor k_A az A esemény *gyakorisága*, $r_n(A) = \frac{k_A}{n}$ pedig a *relatív gyakorisága* .

Megjegyzés: Nyilvánvaló, hogy mind a gyakoriság, mind a relatív gyakoriság konkrét értéke függ a véletlentől. Azonban a relatív gyakoriság rendelkezik az alábbi tulajdonságokkal:

Tétel: Egy adott n-szeres Bernoulli kísérletsorozatnál

a.) $r_n : \Im \to [0,1]$

b.) $r_n(\Omega) = 1$

c.) Ha $A_1, A_2, \dots, A_n, \dots$ egymást kizáró események, akkor $r_n(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} r_n(A_i)$.

Megjegyzés: Az előző tétel azt állítja, hogy a relatív gyakoriság rendelkezik a P valószínűség tulajdonságaival. Később látni fogjuk azt is, hogy n növekedtével $r_n(A) \to P(A)$ is fennáll. (Nagy számok Bernoulli féle törvénye). Ezt a törvényszerűséget először tapasztalati úton fedezték fel a XVII. században, mikor megfigyelték, hogy a relatív gyakoriság egyre kisebb mértékben ingadozik egy 0 és 1 közé eső szám körül. A klasszikus matematikusok éppen ez alapján definiálták az események elméleti valószínűségét: az az érték, amely körül a relatív gyakoriság ingadozik. A relatív gyakoriság tehát alkalmas az elméleti valószínűség - mint fizikai mennyiség - mérésére.

Kolmogorov az axiómáiban a relatív gyakoriság a.)-c.) tulajdonságait örökítette át a valószínűségre, minthogy a határátmenet ezeket a tulajdonságokat megtartja.

A K véletlen kísérlet elemi eseményei számunkra véletlenszerűen következnek be, mégpedig azért, mert a végeredményt befolyásoló körülmények bonyolult komplexumát nem ismerjük pontosan. Viszont ismerjük az egyes események, elemi események bekövetkezési esélyeit - a valószínűséget- , vagy legalábbis tetszőleges pontossággal mérhetjük őket. Ha viszont az A esemény bekövetkezési körülményeiről további információkat szerzünk be, vagy bizonyos pontosító feltételezéssel élünk, megváltozhat az A bekövetkezési esélye, az nőhet is, de csökkenhet is. Pl. a kockadobás kísérletnél, a „6-os dobás" esemény valószínűsége 0, ha tudjuk, hogy a dobott érték páratlan szám, és $\frac{1}{3}$, ha tudjuk, hogy a dobott érték páros volt.

Hogyan változik az A esemény valószínűsége, ha az A-val egyidejűleg megfigyelhető B esemény bekövetkezését ismerjük, vagy legalábbis ismernénk ? Tegyük fel, hogy a K kísérlettel végrehajtottunk egy n hosszúságú Bernoulli-féle kísérletsorozatot. Az A eseményt

k_A-szor, a B eseményt k_B-szer, az AB eseményt pedig k_{AB}-szer figyeltük meg. Ekkor a B esemény bekövetkezéséhez képest az A esemény bekövetkezésének relatív gyakorisága nyilván $r_n(A|B) = \frac{k_{AB}}{k_B}$, melyet az *A eseménynek a B eseményre vonatkoztatott relatív gyakoriságának* nevezünk. Ez az arány az A bekövetkezési esélyeit pontosabban tükrözi, ha a B bekövetkezéséről biztos tudomásunk van, mint a $r_n(A) = \frac{k_A}{n}$.

A feltételes relatív gyakoriság tulajdonságai nyilván :

a.) $0 \le r_n(A|B) \le 1$

b.) $r_n(B|B) = 1$

c.) Ha $A_1, A_2, \ldots, A_n, \ldots \in \Im$ egymást kizáró események, akkor $r_n(\sum_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} r_n(A_i|B)$

Az $r_n(A|B) = \frac{k_{AB}}{k_B} = \frac{\frac{k_{AB}}{n}}{\frac{k_B}{n}} = \frac{r_n(AB)}{r_n(B)}$ átírás után, ha $n \to \infty$ kapjuk, hogy $r_n(A|B) \to \frac{P(AB)}{P(B)}$.

Definíció: Legyenek $A, B \in \Im$ olyan események, hogy A tetszőleges és P(B)>0. Akkor az A *eseménynek a* B*-re vonatkoztatott feltételes valószínűségén* a $P(A|B) = \frac{P(AB)}{P(B)}$ számot értjük.

Feladat Számoljuk ki annak feltételes valószínűségét, hogy két kockával dobva mindkét érték páros feltéve, hogy összegük legalább tíz!

Megoldás: Legyen A: „Két szabályos kockával dobva mindkét érték páros lesz” és B: „A dobott értékek összege nem kisebb mint 10”. P(B)=P(„Az összeg 10 vagy 11 vagy 12”)= P(„A dobások eredménye (6,4),(4,6),(5,5) vagy (5,6),(6,5) vagy (6,6)”)=$\frac{1}{6}$. P(A)=$\frac{3 \cdot 3}{36} = \frac{1}{4}$.

P(AB)=P(„A dobások eredménye (6,4),(4,6) vagy (6,6)”)=$\frac{3}{36} = \frac{1}{12}$. A definíciót használva P(A|B)=$\frac{P(AB)}{P(B)} = \frac{1}{2}$. Láthatjuk, hogy a feltételes valószínűség most nagyobb, mint a feltétel nélküli.

Tétel: Tekintsük az $(\Omega,\Im,P)$ Kolmogorov-féle valószínűségi mezőt. $B \in \Im$, P(B)>0 rögzített. Ekkor a $P_B(A) \overset{def}{=} P(A|B)$ feltételes valószínűségre teljesülnek az alábbi tulajdonságok:

a.) $0 \le P_B(A) \le 1 \quad (\forall A \in \Im)$

b.) $P_B(B) = 1$, $P_B(\emptyset) = 0$

c.) $\forall A_1, A_2, \ldots, A_n, \ldots \in \Im : A_i \cdot A_j = \emptyset \ (i \ne j) \Rightarrow P_B(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P_B(A_i)$

Megjegyzés:

a.)Az előző tétel azt állítja, hogyha B-t rögzítjük, $\Im_B \overset{def}{=} \{C | C = A \cdot B, A \in \Im\}$, akkor a $(B, \Im_B, P_B)$ kielégíti a Kolmogorov valószínűségi mező axiómáit, azaz a feltételes valószínűség bevezetésével az eredeti valószínűségi mezőt leszűkítjük.

.b.)Vannak A,B események, amikor $P(A|B) = P(A)$ teljesül, azaz A valószínűsége nem változik meg, ha a B esemény bekövetkezését ismerjük; az A bekövetkezése "független" a B bekövetkezésétől.

Definíció: Legyenek A,B $\in \Im$, P(A)·P(B) > 0. Az A és B események *függetlenek* , ha $P(A|B) = P(A)$ ($\Rightarrow$ $P(B|A) = P(B)$ is) fennáll.

A következő definíció általánosabb, mint a fenti, hiszen nem követeli meg, hogy az események pozitív valószínűségűek legyenek:

Definíció: Legyenek A,B $\in \Im$ tetszőleges események. Az A és B események *függetlenek* , ha P(AB)=P(A)·P(B) fennáll.

Tétel: Ha az A, B $\in \Im$ események függetlenek, akkor

a.) A és $\overline{B}$

b.) $\overline{A}$ és B

c.) $\overline{A}$ és $\overline{B}$

is függetlenek.

Tétel: Az $\emptyset$ és Ω események minden $A \in \Im$ eseménytől függetlenek.

Definíció: Az $A_1, A_2, \ldots, A_n \in \Im$ események *páronként függetlenek*, ha

$$P(A_i \cdot A_j) = P(A_i) \cdot P(A_j) \quad (\forall \ \ i \neq j) \ .$$

Definíció: Az $A_1, A_2, \ldots, A_n \in \Im$ események *teljesen függetlenek*, ha $\forall$ k $\in \{2,3,\ldots,n\}$ és $\forall$ $1 \le i_1 < i_2 < \cdots < i_k \le n$ index kombinációra $P(A_{i_1} A_{i_2} \cdots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_k})$.

Tétel: Ha az $A_1, A_2, \ldots, A_n \in \Im$ események teljesen függetlenek, akkor páronként is függetlenek. Fordítva általában nem igaz.

A teljes függetlenség definíciójában, amikor k=2, éppen a páronkénti függetlenség definícióját kapjuk.

A megfordításra ellenpélda:
K : Dobjunk fel egy szabályos kockát egymás után kétszer.
A : „Elsőre páratlant dobunk”; B : „Másodikra páratlant dobunk”; C : „A két dobott szám összege páratlan”.
P(A)=P(B)=P(C)=0,5 , P(AB)=P(AC)=P(BC)=0,25 $\Rightarrow$ A , B , C páronként függetlenek.
De P(ABC)=0 $\neq$ P(A)P(B)P(C)=0,125 $\Rightarrow$ azaz A,B és C nem teljesen függetlenek.

Tétel: Ha az $A_1, A_2, \dots, A_n \in \mathfrak{I}$ események teljesen függetlenek, akkor közülük bármelyiket az ellentett eseményére felcserélve, újra teljesen független rendszert kapunk.

Tétel: *(szorzási szabály)*
Legyenek az $A_1, A_2, \dots, A_n \in \mathfrak{I}$ tetszőleges események, hogy $P(\prod_{i=1}^{n} A_i) > 0$. Ekkor

$$P\left(\prod_{i=1}^{n} A_i\right) = P\left(A_n \middle| \prod_{i=1}^{n-1} A_i\right) P\left(A_{n-1} \middle| \prod_{i=1}^{n-2} A_i\right) \cdots P\left(A_2 \middle| A_1\right) P\left(A_1\right).$$

A bizonyítás egyszerűen a feltételes valószínűség definíciójának felhasználásával történhet.

$$P(A_n | \prod_{i=1}^{n-1} A_i) = \frac{P(\prod_{i=1}^{n} A_i)}{P(\prod_{i=1}^{n-1} A_i)}, \; P(A_{n-1} | \prod_{i=1}^{n-2} A_i) = \frac{P(\prod_{i=1}^{n-1} A_i)}{P(\prod_{i=1}^{n-2} A_i)}, \dots, \; P(A_2 | A_1) = \frac{P(A_1 A_2)}{P(A_1)}.$$

A baloldalakat $P(A_1)$-gyel összeszorozva, az egyszerűsítés után kapjuk az állítást.

Feladat A 32 lapos magyar kártyából három lapot húzunk egymás után visszatevés nélkül. Mennyi a valószínűsége annak, hogy az első kihúzott lap hetes, a második kilences, a harmadik ismét hetes?

Megoldás: Legyenek $A_7^{(1)}$: „Az elsőnek húzott lap hetes”, $A_9^{(2)}$ „A másodiknak húzott lap kilences” , $A_7^{(3)}$: „A harmadiknak kihúzott lap hetes”. A keresett valószínűség a $P(A_7^{(1)} A_9^{(2)} A_7^{(3)})$. Alkalmazva a szorzási szabályt:
$P(A_7^{(1)} A_9^{(2)} A_7^{(3)}) = P(A_7^{(1)}) P(A_9^{(2)} | A_7^{(1)}) P(A_7^{(3)} | A_7^{(1)} A_9^{(2)})$, ahol az egyes tényezőket egyszerűen meghatározhatjuk:
$P(A_7^{(1)}) = \frac{4}{32} = \frac{1}{8}$, $P(A_9^{(2)} | A_7^{(1)}) = \frac{4}{31}$, $P(A_7^{(3)} | A_7^{(1)} A_9^{(2)}) = \frac{3}{30} = \frac{1}{10}$. Így a keresett valószínűség $\frac{1}{8} \cdot \frac{4}{31} \cdot \frac{1}{10} = \frac{1}{610}$.

Tétel: *(A teljes valószínűség tétele)*

Legyenek $A_1, A_2, \ldots, A_n, \ldots \in \Im$ teljes eseményrendszer, vagyis $A_i \cdot A_j = \varnothing$, $(i \neq j)$ és $\sum_{i=1}^{\infty} A_i = \Omega$. Tegyük fel továbbá, hogy $P(A_i) > 0$ minden i-re. Ekkor tetszőleges $B \in \Im$ eseményre

$$P(B) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i) \ .$$

Bizonyítás:

Mivel $\sum_{i=1}^{\infty} A_i = \Omega$ és $B = B \cdot \Omega = B \cdot \sum_{i=1}^{\infty} A_i = \sum_{i=1}^{\infty} (A_i B)$, valamint $(A_i B) \cdot (A_j B) = \varnothing$, a valószínűség σ-additivitási tulajdonságából következik, hogy

$$P(B) = P(\sum_{i=1}^{\infty} A_i B) = \sum_{i=1}^{\infty} P(A_i B) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i) \ .$$

Feladat Egy rekeszben 15 teniszlabda van, melyek közül 9 még használatlan. Az első játékhoz kiveszünk találomra három labdát, majd a játék után visszarakjuk azokat a rekeszbe. (Nyilván, ha volt közöttük használatlan, az a játék során elveszti ezt a tulajdonságát.) A második játékhoz ismét találomra veszünk ki három labdát. Mennyi a valószínűsége annak, hogy az utóbb kivett labdák mind még használatlanok lesznek?

Megoldás: Vezessük be az alábbi eseményeket:

A_i: „Az első játékhoz éppen i db használatlan labdát vettünk ki” , i=0,1,2,3.

B : „A második játszmához három használatlant vettünk ki”

Látható, hogy az A_i események teljes eseményrendszert alkotnak.

A B eseménynek az A_i eseményekre vonatkozó feltételes valószínűségei : $P(B|A_i) = \frac{\binom{9-i}{3}}{\binom{15}{3}}$, míg az A_i események valószínűségei: $P(A_i) = \frac{\binom{9}{i}\binom{6}{3-i}}{\binom{15}{3}}$ (i=0,1,2,3). A teljes valószínűség tételét alkalmazva:

$$P(B) = \sum_{i=0}^{3} P(B|A_i)P(A_i) = \frac{\binom{9}{3}\binom{6}{3} + \binom{8}{3}\binom{6}{2}\binom{9}{1} + \binom{7}{3}\binom{6}{1}\binom{9}{2} + \binom{6}{3}\binom{9}{3}}{\binom{15}{3}\binom{15}{3}} \approx 0{,}045$$

Tétel: *(Bayes tétele)*
Legyenek $A_1, A_2, \ldots, A_n, \ldots \in \Im$ teljes eseményrendszer, vagyis $A_i \cdot A_j = \emptyset$, ($i \neq j$) és $\sum_{i=1}^{\infty} A_i = \Omega$. Tegyük fel továbbá, hogy $P(A_i) > 0$ minden i-re. Ekkor tetszőleges $B \in \Im$ eseményre, ahol $P(B) > 0$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)} .$$

Bizonyítás:
A feltételes valószínűség definíciójából: $P(A_i|B) = \frac{P(A_i B)}{P(B)}$. A számláló helyébe $P(B|A_i)P(A_i)$-t írva, a nevező helyébe pedig a teljes valószínűség tételéből kapott formulát helyettesítve azonnal adódik az állítás.

Feladat Hat doboz mindegyikében hat-hat darab golyó van, melyek között rendre 1,2,3,4,5,6 darab fehér színű található (a többi fekete). Egy dobozt véletlenszerűen kiválasztunk, majd abból visszatevéssel három golyót kihúzunk. Ha azt tapasztaljuk, hogy mindhárom golyó fehér színű, mennyi annak a valószínűsége, hogy a csupa fehér golyót tartalmazó dobozt választottuk ki előzőleg?

Megoldás: Legyenek A_i-k a következő események: „Azt a dobozt választottuk, amelyikben i db fehér golyó van" , i=1,2,3,4,5,6. Nyilvánvaló, hogy ezek az események teljes eseményrendszert alkotnak, és mindegyikük bekövetkezése egyformán $\frac{1}{6}$ valószínűségű.
Legyen továbbá B az az esemény, hogy „Visszatevéssel húzva mindegyik golyó színe fehér".
$P(B|A_i) = \left(\frac{i}{6}\right)^3$, i=1,2,3,4,5,6. A Bayes-tételt alkalmazva:

$$P(A_6|B) = \frac{P(B|A_6)P(A_6)}{\sum_{i=1}^{6} P(B|A_i)P(A_i)} = \frac{216}{441} \approx 0{,}49 .$$

Ellenőrző kérdések és gyakorló feladatok

1. Mit értünk az A esemény relatív gyakoriságán?
2. Mennyi a lehetetlen és a biztos esemény relatív gyakorisága egy n-szeres kísérletsorozatban?
3. Mi a feltételes valószínűség definíciója?
4. Mikor nevezünk három eseményt teljesen függetlennek?
5. Mit állít a szorzási szabály?
6. Mondja ki a Bayes tételt!
7. Döntse el, az alábbi állítások közül melyik igaz, melyik hamis!

a. Az esemény relatív gyakorisága mindig nagyobb, mint az esemény elméleti valószínűsége.
b. A relatív gyakoriság lehet kisebb is és nagyobb is, mint az elméleti valószínűség.
c. Ha egy esemény relatív gyakorisága 1, akkor az esemény a biztos esemény.
d. A kísérletek számának növekedtével a relatív gyakoriság értéke egyre csökken.
e. Egymást kizáró események relatív gyakoriságainak összege az összegesemény relatív gyakoriságát adja.
f. A teljes eseményrendszer relatív gyakoriságainak összege 1.
g. Egy eseménynek a biztos eseményre vonatkoztatott feltételes valószínűsége nagyobb mint a feltétel nélküli valószínűsége.
h. A feltételes valószínűség lehet 1-nél nagyobb is.
i. Egy esemény rögzítése után a feltételes valószínűség kielégíti a valószínűség axiómáit.
j. A független események kizárják egymást.
k. Ha két esemény ellentettei függetlenek, akkor az események is azok.
l. A teljes eseményrendszer eseményei teljesen függetlenek egymástól.
m. Bármely két esemény vagy független egymástól, vagy pedig kizárják egymást.
n. Bármely pozitív valószínűségű esemény önmagára vonatkoztatott feltételes valószínűsége 1.
o. Bármely pozitív valószínűségű, de nem egy valószínűségű eseménynek az ellentettjére vonatkoztatott feltételes valószínűsége 0.
p. Egymást kizáró eseményeknél az egymásra vonatkoztatott feltételes valószínűség mindig 0.
q. A teljes függetlenségből következik a páronkénti függetlenség.
r. A lehetetlen esemény önmagától is független

8. Mennyi $P(A|\overline{B})$, ha $P(A)=0,6$, $P(B)=0,5$ és $P(A+B)=0,8$?
9. Dobjunk fel két kockát. Mondjunk olyan eseményeket ezzel a kísérlettel kapcsolatban, amelyek függetlenek, és olyanokat amelyek nem függetlenek egymástól!
10. Az A és B események közül legalább az egyik mindig bekövetkezik. Ha $P(A|B)=0,2$ és $P(B|A)=0,5$, mennyi $P(A)$ és $P(B)$?
11. Három szabályos kockát feldobunk. Mennyi a valószínűsége annak, hogy van hatos értékünk, ha tudjuk, hogy mindegyik dobás páros lett?
12. Egy urnában b darab fekete és r darab fehér golyó van. Véletlenszerűen kihúznak egy golyót. A kihúzott golyót és még ugyanolyan színűből c darabot visszatesznek az urnába. A kísérlet eredményét nem ismerve, másodszorra mi húzunk az urnából. Feltéve, hogy a második húzáskor fekete golyót húzunk, mennyi a valószínűsége annak, hogy az első húzáskor is fekete volt az eredmény?
13. Három szabályos kockát feldobunk. Mennyi a valószínűsége annak, hogy a dobások között van hatos, ha mindegyik kockán különböző érték van?
14. Egy ládában 100 darab játékkocka van, melyek közül 99 teljesen szabályos, egy pedig hamis olyan értelemben, hogy vele mindig hatos dobható csak. Ha véletlenszerűen kiveszünk egy kockát a ládából és azt tízszer feldobva mindig hatost kapunk, mennyi a valószínűsége, hogy éppen a hamis kockát vettük ki előzőleg?
15. Két politikus x és y egymástól függetlenül hazudnak illetve mondanak igazat 2/3 illetve 1/3 valószínűséggel. Feltéve, hogy x azt állítja, hogy „y hazudik”, mennyi a valószínűsége, hogy y igazat mond?

16. Két urna közül az egyikben n fekete és m fehér, a másikban N fekete és M fehér golyó van. Az elsőből találomra átrakunk egyet a másodikba, majd onnan találomra vissza veszünk egyet. Megint az elsőből húzva, mennyi a valószínűsége a fehérnek?
17. Két játékos felváltva húz egy-egy golyót visszatevés nélkül egy urnából, amiben egy fehér és három fekete golyó van. Az a játékos nyer, aki először húz fehéret. Mennyi a valószínűsége, hogy az elsőnek húzó játékos fog nyerni?
18. Egy kalapban tíz cédula van, melyekre a 0,1,2,3,4,5,6,7,8,9 számjegyek vannak felírva. Visszatevéssel kiveszünk két cédulát. Jelölje η a számjegyek összegét, ξ pedig a számjegyek szorzatát. Adjuk meg a $P(\eta=i \mid \xi=0)$ valószínűségeket! $(i=0,1,...,18)$.
19. Egy perzsa sah egyszer egy elítéltnek azt mondta, hogy tetszés szerint elhelyezhet 50 fehér és 50 fekete golyót két egyforma vázába. Az egyikből majd a sah kihúz egy golyót, és ha az fehér , megkegyelmez. Ha viszont a kihúzott golyó fekete, vagy kiderül, hogy nem mindegyik golyó volt a vázákba berakva, esetleg a kiválasztott vázában nem volt semmilyen golyó, az ítélet halál. Hogyan kell szétosztania az elitéltnek a golyókat, hogy a megkegyelmezés valószínűsége maximális legyen?

6. A valószínűségi változó és az eloszlásfüggvény fogalma

A gyakorlati alkalmazások jelentős részében a véletlen kísérlet elemi eseményei valós számokkal jellemezhetőek. Gondoljunk csak például a kockadobás kísérletre, a rulett-tárcsa megforgatására, a Duna pillanatnyi vízmagasságára, vagy a legközelebb születendő csecsemő testsúlyára stb. Sokszor, bár az elemi események nem számok, de egy alkalmas függvénnyel (amit majd valószínűségi változónak nevezünk) egy-egyértelmű megfeleltetés létesíthető köztük és a valós számok egy részhalmaza között, és így a valószínűségi változó segítségével átfogalmazható a véletlen jelenség. Pl. a kártyahúzásnál a kártyákat sorszámozzuk, minden addigi esemény ekvivalens módon tárgyalható. A leképező függvények (valószínűségi változók) definiálása az esetek többségében természetes módon adódik.

Felhasználhatók a valószínűségi változók az eredeti kísérlet egyszerűsítésére is. Pl. később látni fogjuk, hogy egy n-szeres hosszúságú Bernoulli kísérletsorozat helyett egyetlen valószínűségi változó megfigyelése is lehetséges.

Definíció: Legyen $(\Omega,\Im,P)$ Kolmogorov- féle valószínűségi mező. A $\xi : \Omega \to \mathrm{lR}$ függvényt *valószínűségi változónak* nevezzük, ha minden $x \in \mathrm{IR}$ esetén a „ξ kisebb értéket fog felvenni mint x” állítás megfigyelhető esemény lesz, azaz $A_x=\{\omega|\ \xi(\omega)<x\}\in\Im$ minden valós x-re.

A valószínűségi változóval kapcsolatos események valószínűségeit az eloszlásfüggvény segítségével fogjuk számolni.

Definíció: Az $F_\xi(x) = P(A_x) = P(\{\omega|\xi(\omega) < x\}) \stackrel{jel}{=} P(\xi < x), x \in \mathrm{lR}$ függvényt a ξ valószínűségi változó *eloszlásfüggvényének* nevezzük.

Mint látható, az eloszlásfüggvény a valós számokat a [0,1] intervallumra leképező valós függvény, azaz $F_\xi : \mathrm{lR} \to [\,0\,,1\,]$. A következő tétel összefoglalja az eloszlásfüggvény legfontosabb tulajdonságait. Bizonyítható, hogy ha egy F(x) valós függvény rendelkezik az alábbi a.),b.),c.) tulajdonsággal, akkor ahhoz mindig található olyan K véletlen kísérlet és azzal kapcsolatos valószínűségi változó, aminek éppen F(x) az eloszlásfüggvénye. Az eloszlásfüggvények, és az a.), b.), c.) tulajdonsággal rendelkező valós függvények halmaza tehát egybeesik!

Tétel: *(Az F_ξ eloszlásfüggvény tulajdonságai)*

a.) F_ξ monoton nemcsökkenő, azaz $F_\xi(x) \le F_\xi(y)$, ha $x < y$.

b.) F_ξ balról folytonos, azaz $\lim\limits_{x\to y+} F_\xi(x) = F_\xi(y)$ minden $y \in \mathrm{lR}$ -re .

c.) $\lim\limits_{x\to+\infty} F_\xi(x) = 1$ és $\lim\limits_{x\to-\infty} F_\xi(x) = 0$.

Feladat Mutassuk meg, hogy az $F(x)=\begin{cases} 0 & \text{, ha } x \le 1 \\ \dfrac{1+2x}{x-0{,}8} & \text{, } x>1 \end{cases}$ függvény nem lehet eloszlásfüggvény!

Megoldás: Mivel $\lim\limits_{x\to\infty} F(x) = 2$, ezért a c.) tulajdonság sérül.

A következő tétel mutat rá arra, hogyan lehet az eloszlásfüggvényt felhasználni a „ξ értékei x és y közé esnek” típusú események valószínűségeinek kiszámításához.

Tétel: Tetszőleges x<y esetén

a.) $P(x \le \xi < y) = F_\xi(y) - F_\xi(x)$

b.) $P(x < \xi < y) = F_\xi(y) - F_\xi(x+0)$

c.) $P(x \le \xi \le y) = F_\xi(y+0) - F_\xi(x)$

d.) $P(x < \xi \le y) = F_\xi(y+0) - F_\xi(x+0)$

e.) $P(\xi = x) = F_\xi(x+0) - F_\xi(x)$

Vegyük észre, hogy ha F_ξ folytonos az x helyen, azaz $F_\xi(x) = F_\xi(x+0)$, akkor az állítás e.) pontjának értelmében $P(\xi = x) = 0$. Tehát, ha egy valószínűségi változóhoz folytonos eloszlásfüggvény tartozik, akkor az azt is jelenti, hogy értékkészletének minden elemét 0 valószínűséggel vesz fel. Pl. a Duna vízmagasságát nyílván egy folytonos valószínűségi változóval jellemezhetjük. Annak valószínűsége, hogy egy tetszőleges pillanatban megfigyelve a vízmagasságot éppen 8 métert kapjunk (mm pontossággal) nulla valószínűségű esemény. (Lehet, hogy a megfigyelt érték közel lesz a 8000 mm-hez, de némi eltérés biztosan fog mutatkozni...) Ez persze nem jelenti azt, hogy a „Duna vízmagassága éppen 8 méter” esemény lehetetlen volna. Ez csupán annyit jelent, hogy az említett esemény bár elvileg bekövetkezhet, de ennek valószínűsége 0. Különbség van tehát a 0 valószínűségű esemény és a lehetetlen ($\emptyset$) esemény között. A lehetetlen esemény speciális nulla valószínűségű esemény.

6.1 Diszkrét valószínűségi változók

Definíció: A ξ valószínűségi változót *diszkrétnek* nevezzük, ha értékkészlete megszámlálható (sorozatba rendezhető) , vagyis $\forall\ \omega \in \Omega$-ra $\xi(\omega) \in X = \{x_1, x_2, \ldots, x_n, \ldots\}$, és X-nek nincsen torlódási pontja. Ez utóbbi azt jelenti, hogy bármely x_i értékhez található olyan pozitív ε szám, hogy az (x_i-ε, x_i+ε) intervallumban egyedül x_i van az X elemei közül.

A diszkrét valószínűségi változóknál a kapcsolatos események valószínűségeit az eloszlással kalkuláljuk, aminek definícióját alant adjuk meg.

Definíció: A $p_i = P(\{\omega |\ \xi(\omega) = x_i\}) \overset{\text{jel}}{=} P(\xi = x_i)$ (i=1,2,...) valószínűségek összességét a ξ diszkrét valószínűségi változó *eloszlásának* nevezzük.

Tétel: A ξ diszkrét valószínűségi változó $p_1, p_2, \ldots, p_n, \ldots$eloszlására teljesül, hogy

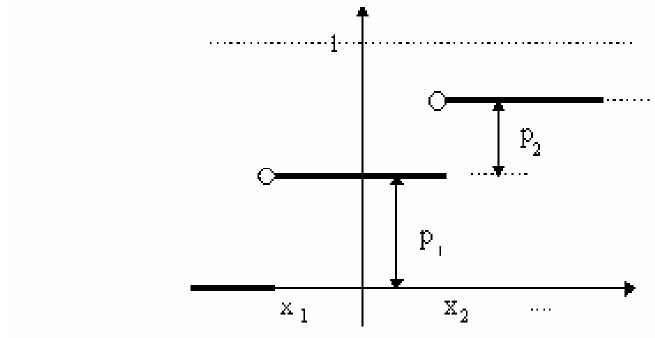
a.) $0 \le p_i \le 1$

b.) $\sum_{i=1}^{\infty} p_i = 1$

Az a.) állítás abból adódik, hogy a p_i számok éppen az $A_i = \{\omega |\ \xi(\omega) = x_i\}$ események valószínűségei.
Mivel a $A_i = \{\omega |\ \xi(\omega) = x_i\}$ (i=1,2,...) események teljes eseményrendszert alkotnak, így a b.) állítás is igaz.

Tétel: A ξ diszkrét valószínűségi változó F_ξ eloszlásfüggvényére igaz, hogy $F_\xi(x) = \sum_{x_i < x} p_i$ másrészt $p_i = F_\xi(x_i + 0) - F_\xi(x_i)$. Azaz a diszkrét valószínűségi változó eloszlásfüggvénye olyan lépcsős függvény, melynek az ugróhelyei az $x_1, x_2, \ldots, x_n, \ldots$ helyeken vannak, és az ugrás nagysága rendre $p_1, p_2, \ldots, p_n, \ldots$.

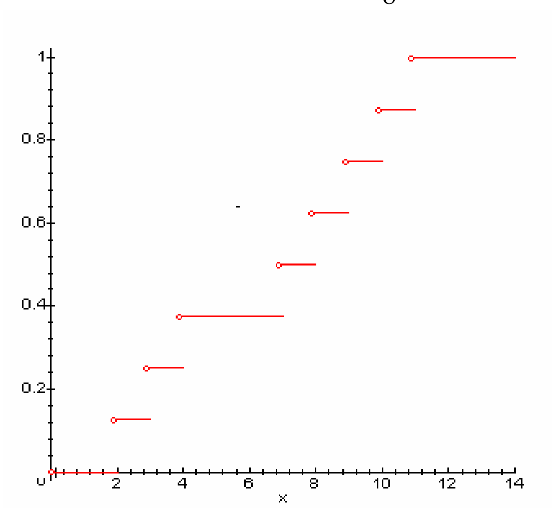
Mivel $A_x = \{\omega |\ \xi(\omega) < x\} = \sum_{x_i < x} A_i = \sum_{x_i < x} \{\omega |\ \xi(\omega) = x_i\}$ és az A_i események egymást páronként kizárják, következik az állítás első része.
Másrészt $p_i = P(\xi = x_i) = P(x_i \le \xi \le x_i) = F_\xi(x_i + 0) - F_\xi(x_i)$.



Diszkrét valószínűségi változó eloszlásfüggvénye

Feladat Egy csomag magyar kártyacsomagból találomra kihúzunk egy lapot. Vegye fel ξ a kártya pontértékét! (alsó:2, felső:3, király:4, ász:11, hetes:7, nyolcas:8, kilences:9, tízes:10). Adjuk meg és ábrázoljuk a ξ eloszlásfüggvényét!

Megoldás: ξ lehetséges értékei, az értékkészlete az $\{2,3,4,7,8,9,10,11\}$ számhalmaz. Mindegyik i értéket $P(\xi = i) = \dfrac{1}{8}$ valószínűséggel veheti fel. Így az eloszlásfüggvény:



$$F_\xi(x) = \begin{cases} 0 \ , \text{ ha } x \le 2 \\ \frac{1}{8}, \text{ ha } 2 < x \le 3 \\ \frac{2}{8}, \text{ ha } 3 < x \le 4 \\ \frac{3}{8}, \text{ ha } 4 < x \le 7 \\ \frac{4}{8}, \text{ ha } 7 < x \le 8 \\ \frac{5}{8}, \text{ ha } 8 < x \le 9 \\ \frac{6}{8}, \text{ ha } 9 < x \le 10 \\ \frac{7}{8}, \text{ ha } 10 < x \le 11 \\ 1 \ , \text{ ha } x > 11 \end{cases}$$

Az alábbiakban a gyakorlati alkalmazásokban leggyakrabban előforduló nevezetes diszkrét valószínűségi változókat fogjuk tárgyalni.

6.1.1 Példa *Karakterisztikus valószínűségi változó*

Legyen $(\Omega,\Im,P)$ Kolmogorov- féle valószínűségi mező, $A \in \Im$ egy pozitív valószínűségű esemény: $p = P(A) > 0$.

A $\xi : \Omega \to \mathrm{lR}$ függvény definíciója a következő : $\xi(\omega) = \begin{cases} 1\,, & \omega \in A \\ 0\,, & \omega \notin A \end{cases}$. (Vagyis ξ az A esemény bekövetkezésekor 1 értéket, különben 0 értéket vesz fel.) Ekkor ξ diszkrét valószínűségi változó, melyet k*arakterisztikus*- vagy *indikátor valószínűségi változónak* nevezünk. Jelölés: $\xi \in \chi(A)$. A ξ eloszlása :

$p_0 = P(\xi = 0) = P(\overline{A}) = 1 - p$, $p_1 = P(\xi = 1) = P(A) = p$.

6.1.2 Példa *Binomiális eloszlású valószínűségi változó*

Legyen $(\Omega,\Im,P)$ Kolmogorov- féle valószínűségi mező, $A \in \Im$ egy pozitív valószínűségű esemény: $p = P(A) > 0$. Hajtsunk végre egy n-szeres Bernoulli-féle kísérletsorozatot. Vegye fel ξ azt az értéket, ahányszor A bekövetkezett a kísérletsorozatban. ξ lehetséges értékei tehát 0,1,2,...,n. Az egyes értékek felvételének valószínűségei, azaz ξ eloszlása :

$$p_k = P(\xi = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = \binom{n}{k} \cdot p^k \cdot q^{n-k} \quad , \ k = 0,1,2,\ldots,n.$$

ξ-t n és p paraméterű *binomiális eloszlású* valószínűségi változónak nevezzük.
Jelölés: $\xi \in B(n,p)$.

A binomiális eloszlás képletét az alábbi felbontás alapján lehet megérteni:

$$\{\omega \mid \xi(\omega) = k\} = \overset{1.}{A} \cdot \overset{2.}{A} \cdots \overset{k.}{A} \cdot \overset{k+1.}{\overline{A}} \cdot \overset{k+2.}{\overline{A}} \cdots \overset{n.}{\overline{A}} + \overset{1.}{A} \cdot \overset{2.}{A} \cdots \overset{k-1.}{A} \cdot \overset{k.}{\overline{A}} \cdot \overset{k+1.}{A} \cdot \overset{k+2.}{\overline{A}} \cdots \overset{n.}{\overline{A}} + \cdots + \overset{1.}{\overline{A}} \cdot \overset{2.}{\overline{A}} \cdots \overset{n-k.}{\overline{A}} \cdot \overset{n-k+1.}{A} \cdot \overset{n-k+2.}{A} \cdots \overset{n.}{A}$$

A jobboldalon álló események egymást kizárják, és mindegyikük valószínűsége a függetlenség miatt $p^k \cdot q^{n-k}$. A tagok száma $\dfrac{n!}{k! \cdot (n-k)!} = \dbinom{n}{k}$, mert n elem olyan ismétléses permutációiról van szó, ahol k illetve n-k elem megegyezik.

A p_k valószínűségek eloszlást alkotnak, hiszen a binomiális tétel szerint:

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

Nyilván $B(1,p) = \chi(A)$, tehát a binomiális eloszlás a karakterisztikus eloszlás kiterjesztése.

Tétel: A binomiális eloszlás p_k elemeire teljesül, hogy

a.) $p_k = \dfrac{n-k+1}{k} \cdot \dfrac{p}{q} \cdot p_{k-1} \quad , (k = 1,2,3,\ldots,n)$, $p_0 = q^n$

b.) Ha $\alpha = [(n+1) \cdot p]$, ahol $[x]$ az egészrészt jelöli, akkor $p_\alpha \geq p_k$, $(k = 0,1,\ldots,n)$

Feladat A véletlen kísérlet az, hogy n-szer feldobunk egy szabályos játékkockát és egy pénzdarabot egyszerre. Jelölje ξ a hatos dobások számát, η pedig a fejdobások számát. Adjuk meg a $P(\xi < \eta)$ valószínűséget!

Megoldás: Mindkét változó binomiális eloszlású: $\xi \in B(n,\frac{1}{6})$ és $\eta \in B(n,\frac{1}{2})$.

$$P(\xi = k) = \binom{n}{k}\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{n-k} \text{ illetve } P(\eta = k) = \binom{n}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{n-k} = \binom{n}{k}\left(\frac{1}{2}\right)^n, \text{ k=0,1,...,n.}$$

$P(\eta < \xi | \xi = 0) = 0$, mert ez lehetetlen,

$P(\eta < \xi | \xi = 1) = P(\eta = 0)$,

$P(\eta < \xi | \xi = 2) = P(\eta = 0) + P(\eta = 1)$,

$\vdots$

$P(\eta < \xi | \xi = n) = P(\eta = 0) + P(\eta = 1) + \cdots P(\eta = n-1)$.

A teljes valószínűség tételét felhasználva kapjuk meg a végeredményt:

$P(\eta < \xi) = P(\eta < \xi | \xi = 0)P(\xi = 0) + \cdots + P(\eta < \xi | \xi = n)P(\xi = n)$.

6.1.3 Példa *Poisson eloszlású valószínűségi változó*

Ha egy ξ valószínűségi változó értékkészlete a természetes számok halmaza: $X = \mathbb{N} = \{0,1,2,\ldots,n,\ldots\}$, eloszlása pedig $p_k = P(\xi = k) = \frac{\lambda^k}{k!}e^{-\lambda}$, $k = 0,1,2,\ldots$, ahol $\lambda > 0$ akkor ξ-t λ *paraméterű Poisson eloszlású valószínűségi változónak* nevezzük. Jelölés: $\xi \in Po(\lambda)$.

A fenti valószínűségek valóban eloszlást alkotnak, mert

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

Poisson eloszlást alkalmazunk a binomiális eloszlás helyett olyankor, amikor n nagy és p kicsi. Erre vonatkozik az alábbi tétel:

Tétel: $\lim\limits_{\substack{n\to\infty \\ p\to 0 \\ np=\lambda}} \binom{n}{k} p^k q^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}$, azaz a Poisson eloszlás a binomiális eloszlás határesete, amikor a kísérletek száma (n) minden határon túl nő, az A esemény valószínűsége pedig 0-hoz tart, miközben az np szorzat állandó.

A Poisson eloszlás tehát jól alkalmazható olyan Bernoulli kísérletsorozat modellezéséhez, ahol a kísérletek száma nagyon nagy, viszont a megfigyelt esemény valószínűsége 0-hoz közeli. Például:

- egy adott térfogatban időegység alatt elbomló atomi részecskék száma;
- a mikroszkóp látóterébe bekerült egysejtűek száma;
- időegység alatt a telefonközpontba beérkező hívások száma;
- egy süteményszeletben található mazsolák száma;
- egy könyvoldalon található sajtóhibák száma; stb.

Az említett esetekben binomiális eloszlás alkalmazása körülményes lenne, mert a binomiális együtthatók számolása a nagy n miatt túlcsorduláshoz, illetve számolási pontatlanságokhoz vezethet.

6.1.4 Példa *Geometriai eloszlású valószínűségi változó*

Legyen K egy véletlen kísérlet, és $(\Omega,\Im,P)$ a hozzátartozó Kolmogorov- féle valószínűségi mező, $A \in \Im$ egy pozitív valószínűségű esemény: $p = P(A) > 0$. A K kísérlet egymástól függetlenül addig hajtsuk végre, amíg az A esemény be nem következik. A ξ valószínűségi változót értelmezzük úgy, mint az A esemény bekövetkezéséhez szükséges ismétlések számát. ξ-t *p paraméterű geometriai eloszlású valószínűségi változónak* nevezzük.
Jelölés: $\xi \in G(p)$.

ξ lehetséges értékei : 1,2,3.4,..., azaz a pozitív egész számok. ξ eloszlása: $p_k = P(\xi = k) = (1-p)^{k-1}p = q^{k-1}p$, hiszen $\left\{\omega \middle| \xi(\omega) = k\right\} = \overset{1.}{\overline{A}} \cdot \overset{2.}{\overline{A}} \cdot \ldots \cdot \overset{k-1.}{\overline{A}} \cdot \overset{k.}{A}$, és a független végrehajtás miatt az esemény valószínűsége: $q \cdot q \cdot \ldots \cdot q \cdot p = q^{k-1}p$.
A geometriai sor összegzőképletét felhasználva láthatjuk be, hogy ezek a valószínűségek valóban eloszlást alkotnak: $\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} q^{k-1}p = p\sum_{k=0}^{\infty} q^k = p\frac{1}{1-q} = p\frac{1}{p} = 1.$

Tétel: A geometriai eloszlás *örökifjú* tulajdonságú:

$$P(\xi = m+k \,|\, \xi > m) = P(\xi = k)\ , \forall m,k\text{-ra.}$$

Annak feltételes valószínűsége, hogy a következő k végrehajtás végén bekövetkezik az A esemény, amennyiben az előző m megfigyelés alatt nem következett be ugyanannyi, mint annak valószínűsége, hogy éppen a k-adik végrehajtás után következik be az A esemény.

Bizonyítás:

$$P(\xi = m+k \,|\, \xi > m) = \frac{P(\xi = m+k, \xi > m)}{P(\xi > m)} = \frac{P(\xi = m+k)}{P(\xi > m)} = \frac{q^{m+k-1}p}{\sum_{\alpha=m+1}^{\infty} q^{\alpha-1}p} =$$

$$= \frac{q^{m+k-1}p}{pq^m \sum_{\alpha=0}^{\infty} q^{\alpha}} = \frac{q^{m+k-1}p}{q^m} = q^{k-1}p = P(\xi = k)$$

A geometriai eloszlás „örökifjú” tulajdonságát a következőképp lehet interpretálni: attól, hogy egy esemény az ismételt végrehajtás során régen fordult elő, még nem fog a bekövetkezési valószínűség megnőni ! Tehát pl. azért, mert régóta lottózom nem lesz nagyobb az ötös találat elérésének esélye.

Feladat A véletlen kísérlet az, hogy n darab dobozba véletlenszerűen golyókat helyezünk el úgy, hogy minden elhelyezésnél bármelyik doboz kiválasztása egyformán valószínű. Akkor állunk meg, ha észrevesszük, hogy az egyes számú dobozba bekerült az első golyó. Jelölje ξ a kísérlet befejeződésekor az elhelyezett golyók számát. Adjuk meg a ξ eloszlását!

Megoldás: Annak valószínűsége, hogy az egyes számú dobozba ejtünk egy golyót p=$\frac{1}{n}$, annak, hogy nem ebbe kerül a golyó $q = \frac{n-1}{n}$. Ha A-val jelöljük a „az egyes dobozba kerül a golyó”, akkor a golyóelhelyezéseket addig kell folytatnunk, amíg A először be nem fog következni, tehát ξ geometriai eloszlású lesz. Az eloszlása:

$P(\xi = k) = q^{k-1}p = \left(\frac{n-1}{n}\right)^{k-1}\frac{1}{n}$, $k = 0,1,2,\ldots$.

6.1.5 Példa *Hipergeometriai eloszlású valószínűségi változó*

Tegyük fel , hogy egy urnában N golyó között F fekete van a többi nem fekete. Kiveszünk egyszerre n db golyót az urnából, ahol $1 \le n \le \min(N - F, F)$. Vegye fel a ξ valószínűségi változó a kivett golyók között található fekete színűek számát! Nyilván, a ξ lehetséges értékei 0,1,...,n. A ξ eloszlása $p_k = P(\xi = k) = \frac{\binom{F}{k}\cdot\binom{N-F}{n-k}}{\binom{N}{n}}$, $k = 0,1,\ldots,n$. ξ -t *n,N,F paraméterű hipergeometriai eloszlású* valószínűségi változónak nevezzük.
Jelölés: $\xi \in HG(n,N,F)$.

A klasszikus valószínűségi képlet alapján következik a p_k-ra fent adott képlet. Az összes lehetséges kiválasztások száma N elem n-edosztályú ismétlés nélküli kombinációi.
A "kedvező" kiválasztások számának meghatározása: az F fekete közül k-t $\binom{F}{k}$ féleképpen, az n-k db nem feketét az N-F közül pedig $\binom{N-F}{n-k}$ féleképpen lehet kiválasztani, így a szorzat megadja a különböző k feketét tartalmazó kiválasztások összes számát.
Azt, hogy a p_k valószínűségek valóban eloszlást alkotnak, úgy tudjuk igazolni, ha az $(1+x)^{N-n}\cdot(1+x)^n = (1+x)^N$ azonosságban összehasonlítjuk mindkét oldalon x^n együtthatóit: $\sum_{k=0}^{n}\binom{F}{k}\cdot\binom{N-F}{n-k} = \binom{N}{n}$. Átosztás után adódik a $\sum_{k=0}^{n} p_k = 1$ összefüggés.
Amikor egy nagyobb széria selejtarányát akarják megbecsülni, mintát vételeznek, és a mintában megfigyelt selejtarányból próbálnak következtetni az egész készlet selejtarányára. A mintát kétféleképpen képezhetjük. *Visszatevés nélküli mintavételezésről* beszélünk, ha a mintaelemeket egyenként vesszük ki és utána végezzük el a selejtességre vonatkozó vizsgálatot. *Visszatevéses a mintavételezés*, ha a mintaelemeket megvizsgálás után visszatesszük, és az újabb húzáskor megint számolunk az összes termékkel, tehát elvileg olyan elemet is kivehetünk, melyet előzőleg már vizsgáltunk. Ha az urnamodellben a golyók helyett termékeket, a fekete golyók helyett selejtes termékeket veszünk, akkor a hipergeometriai eloszlás a teljes készletből való viasztevés nélküli mintavételezést jelenti. Amennyiben az n termék kiválasztását úgy végezzük, hogy minden kiválasztás után a

terméket visszatesszük, $B(n,\frac{F}{N})$ (binomiális) eloszlással írhatjuk le a folyamatot. Az alábbi tétel azt mondja ki, hogy nagy elemszámú sokaság esetén a kétféle mintavételezés között gyakorlatilag nincs különbség.

Tétel: $\lim\limits_{\substack{N\to\infty \\ F\to\infty \\ \frac{F}{N}=p}} \frac{\binom{F}{k}\binom{N-F}{n-k}}{\binom{N}{n}} = \binom{n}{k}p^k q^{n-k}$ .A hipergeometriai eloszlás értékei $B(n,\frac{F}{N})$ eloszlással jól közelíthetőek, ha $N = pF \to \infty$.

Feladat Mennyi a valószínűsége, hogy a hagyományos ötös lottóhúzás során valamennyi kihúzott szám páros lesz?

Megoldás: Ha ξ most a kihúzott páros számok számát jelenti, akkor $\xi \in HG(90,45,5)$, hiszen a páros számok száma 45. A ξ eloszlása $P(\xi = k) = \frac{\binom{45}{k}\binom{45}{5-k}}{\binom{90}{5}}$, k=0,1,2,3,4,5. A kérdés arra vonatkozik, amikor k=5, azaz a keresett valószínűség: $P(\xi = 5) = \frac{\binom{45}{5}\binom{45}{0}}{\binom{90}{5}} \approx 0{,}0278$.

6.2. Folytonos valószínűségi változók

Definíció: Legyen ξ az $(\Omega,\Im,P)$-n értelmezett valószínűségi változó, melynek értékkészlete kontinuum (nem megszámlálhatóan végtelen) számosságú. Jelölje F_ξ az eloszlásfüggvényt. ξ-t *folytonos valószínűségi változónak* nevezzük , ha F_ξ abszolút folytonos, azaz létezik olyan $f_\xi : \mathrm{lR} \to \mathrm{lR}$ függvény, melyre fennáll az $F_\xi(x) = \int\limits_{-\infty}^{x} f_\xi(t)\,dt \quad (x \in \mathrm{lR})$ összefüggés. Az f_ξ függvényt a ξ valószínűségi változó (vagy az F_ξ eloszlásfüggvény) *sűrűségfüggvényének* nevezzük. Ha F_ξ abszolút folytonos, akkor folytonos is és majdnem mindenütt differenciálható, azaz praktikusan véges sok helyen lehet csak töréspontja: $\frac{dF_\xi(x)}{dx} = f_\xi(x)$, ha x folytonossági pontja f_ξ-nek.

Megjegyzés: a.) A diszkrét valószínűségi változók nem folytonosak, már csak azért sem, mert eloszlásfüggvényük nem folytonos.
b.) Léteznek olyan valószínűségi változók, melyek se nem diszkrétek, se nem folytonosak. Ezek az *általános valószínűségi változók*, melyekkel a továbbiakban mi nem foglalkozunk; a gyakorlatban ritkán fordulnak elő. Pl. az a ξ általános valószínűségi változó, melynek eloszlásfüggvénye:

$$F_\xi(x) = \begin{cases} \frac{1}{2} & \text{, ha } x = -1 \\ \frac{1}{\sqrt{2\pi}} \int\limits_{0}^{x} e^{\frac{-t^2}{2}}\,dt & \text{, ha } x > 0 \end{cases}.$$

Tétel: *(A sűrűségfüggvény tulajdonságai)*
Legyen ξ az $(\Omega,\Im,P)$-n értelmezett folytonos valószínűségi változó. Akkor az $f_\xi : \mathrm{lR} \to \mathrm{lR}$ sűrűségfüggvényre teljesül, hogy

a.) $f_\xi(x) \ge 0$, ha x folytonossági pont.

b.) $\int\limits_{-\infty}^{+\infty} f_\xi(t)\,dt = 1$.

Az a.) állítás abból következik, hogy F_ξ monoton nem csökkenő, és $\frac{dF_\xi(x)}{dx} = f_\xi(x)$, ha x folytonossági pontja f_ξ-nek. Ugyanis monoton nem csökkenő függvény deriváltja nemnegatív.
A b.) tulajdonság az eloszlásfüggvény c.) tulajdonságából adódik:
$1 = \lim\limits_{x\to+\infty} F_\xi(x) = \lim\limits_{x\to+\infty} \int\limits_{-\infty}^{x} f_\xi(t)\,dt = \int\limits_{-\infty}^{+\infty} f_\xi(t)\,dt$.

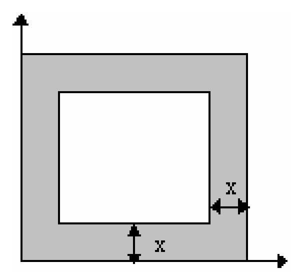
Megjegyzés:
a.) A sűrűségfüggvény a folytonos valószínűségi változóknál ugyanazt a szerepet tölti be, mint diszkrét valószínűségi változóknál az eloszlás. Ugyanis tetszőleges $a \in \mathbb{R}$ és $\Delta x > 0$-ra

$$P(a \le \xi < a + \Delta x) = F_\xi(a + \Delta x) - F_\xi(a) = \int_a^{a+\Delta x} f_\xi(t)\,dt = f_\xi(a^*)\Delta x \text{ , ahol}$$

$a \le a^* < a + \Delta x$. Ha Δx kicsi , akkor $f_\xi(a) \approx f_\xi(a^*)$, így $P(a \le \xi < a + \Delta x) \approx f_\xi(a)\Delta x$.
Tehát a ξ valószínűségi változó az *a* környezetében az $f_\xi(a)$ értékkel arányos valószínűséggel tartózkodik. (Az $f_\xi(a)$ érték lehet 1-nél nagyobb is !)
b.) $f_\xi(x) \overset{\text{def}}{=} 0$, ha $\nexists \dfrac{dF_\xi(x)}{dx}$.

Feladat Az egységnégyzeten kiválasztunk véletlenszerűen egy pontot. Jelölje ξ a pontnak a legközelebbi oldaltól vett távolságát. Adjuk meg a ξ valószínűségi változó sűrűségfüggvényét!

Megoldás: Geometriai módszerrel lehet meghatározni az eloszlásfüggvényt. Az alábbi ábrán sötétítve mutatjuk a ξ<x eseménynek megfelelő tartományt:



A terület nagysága $(1-2x)^2$, így $F_\xi(x) = \begin{cases} 0 \quad , & \text{ha } x \le 0 \\ 1-(1-2x)^2 , & \text{ha } 0 < x \le 0{,}5 \\ 1 \quad , & \text{ha } x > 0{,}5 \end{cases}$.

Deriválás után kapjuk a sűrűségfüggvényt: $f_\xi(x) = \begin{cases} 4-8x , & \text{ha } 0 < x < 0{,}5 \\ 0 , & \text{egyébként} \end{cases}$.

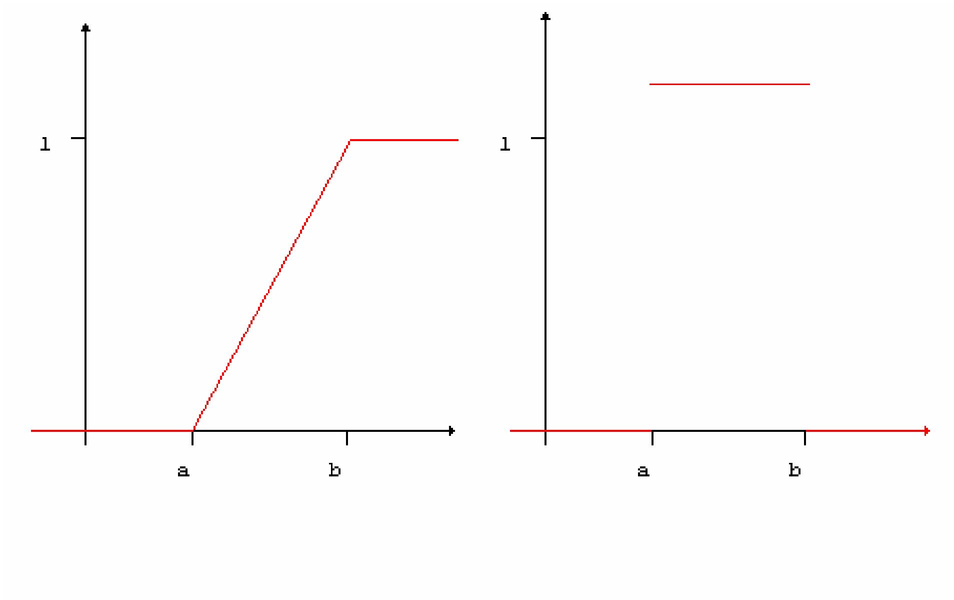
6.2.1 Példa: *Az egyenletes eloszlású valószínűségi változó*

A ξ az $[a\,,b]$ intervallumon egyenletes eloszlású, ha eloszlásfüggvénye:

$$F_\xi(x)=\begin{cases}0 & ,\ x\le a\\ \dfrac{x-a}{b-a}, & a<x\le b\\ 1 & ,\ x>b\end{cases}.$$

Jelölés: $\xi\in U([a\,,b])$.

Ekkor a sűrűségfüggvény: $f_\xi(x)=\begin{cases}\dfrac{1}{b-a} & ,\ x\in(a\,,b)\\ 0 & ,\ x\notin(a\,,b)\end{cases}$.



Az [a,b] intervallumon egyenletes eloszlás eloszlás- és sűrűségfüggvénye

Tétel: Ha ξ a $[0\,,1]$ intervallumon egyenletes eloszlású és $F(y)$ egy szigorúan monoton növekvő eloszlásfüggvény azon az intervallumon, ahol $0<F(y)<1$, akkor az $\eta=F^{-1}(\xi)$ valószínűségi változó eloszlásfüggvénye éppen $F(y)$ lesz.

Ez a tétel könnyen belátható. Először is megjegyezzük, hogy egy szigorúan monoton növekvő függvénynek létezik az inverze.

$P(\eta<y)=P(F^{-1}(\xi)<y)=P(F(F^{-1}(\xi))<F(y))=P(\xi<F(y))=F(y)$, mert $F(y)\in[0\,,1]$.

A tétel lehetőséget ad, hogy a számítógépek egyenletes eloszlású véletlen számokat generáló rutinja segítségével tetszőleges $F(y)$ eloszlásfüggvényhez tartozó véletlen számokat előállítsunk és azokat szimulációs programokhoz felhasználjuk. Például a kockadobás kísérletét úgy szimulálhatjuk, hogy generálunk egy ξ véletlen számot a nulla és egy között. Ha $\xi \in \left[\frac{i-1}{6}, \frac{i}{6}\right]$, akkor az „a kockával i értéket dobtunk” eseménynek fog megfelelni.

Feladat Egy egységnyi hosszúságú szakaszt találomra választott pontjával két részre osztunk. Mi a keletkezett szakaszok közül a kisebbik hosszának sűrűségfüggvénye?

Megoldás: Jelöljük η-val a kiválasztott pont origótól vett távolságát! Ekkor nyílván $\eta \in U[0,1]$, és eloszlásfüggvénye $F_\eta(x) = \begin{cases} 0, \text{ ha } x \le 0 \\ x, \text{ ha } 0 < x < 1 \\ 1, \text{ ha } x \ge 1 \end{cases}$. A keletkező szakaszok közül a rövidebb hosszát jelöljük ξ-vel! A két változó között az alábbi kapcsolat áll fenn: $\xi = \begin{cases} \eta, \text{ ha } \eta \le 0{,}5 \\ 1-\eta, \text{ ha } \eta > 0{,}5 \end{cases}$. Így ξ eloszlásfüggvénye kifejezhető lesz η eloszlásfüggvényével:

$$F_\xi(x) = P(\xi < x) = P(\xi < x, \eta \le 0{,}5) + P(\xi < x, \eta > 0{,}5) = P(\eta < x, \eta \le 0{,}5) + P(1-\eta < x, \eta > 0{,}5) =$$

$$= \begin{cases} 0 & , \text{ha } x \le 0 \\ P(\eta < x) + P(1-x < \eta) = x + (1-(1-x)) = 2x, & \text{ha } x \in (0, 0{,}5) \\ 1 & , \text{ha } x \ge 0{,}5 \end{cases}, \text{ azaz } \xi \in U[0, 0{,}5].$$

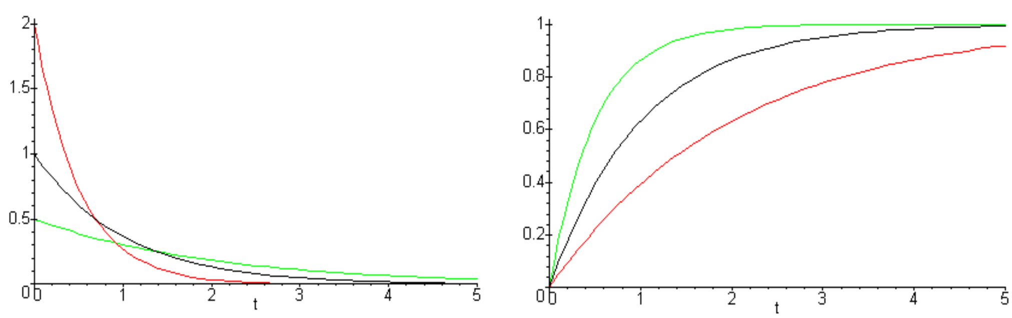
6.2.2 Példa: *Az exponenciális eloszlású valószínűségi változó*

A ξ $\lambda > 0$ paraméterű exponenciális eloszlású valószínűségi változó, ha eloszlásfüggvénye

$$F_\xi(x) = \begin{cases} 1 - e^{-\lambda x} & , x > 0 \\ 0 & , x \le 0 \end{cases}.$$

Jelölés: $\xi \in E(\lambda)$.

A sűrűségfüggvény $F_\xi{}'(x) = f_\xi(x) = \begin{cases} \lambda e^{-\lambda x} & , x > 0 \\ 0, \text{egyébként} \end{cases}$.



Az exponenciális eloszlás sűrűség – és eloszlásfüggvénye a λ = 0.5 , 1 , 2 esetekben

Tétel: *(Az exponenciális eloszlás örökifjú tulajdonsága)*
Ha $\xi \in E(\lambda)$, akkor $P(\xi < y | \xi \geq x) = P(\xi < y - x) \quad \forall x < y$, vagyis annak feltételes valószínűsége, hogy ξ legfeljebb y-ig „él″, ha már x-et „megélt″ egyenlő annak valószínűségével, hogy ξ legfeljebb y-x ideig „él″, azaz a túlélési kondíciók az idő múlásával nem csökkennek, hisz 0 és y-x között ugyanaz a túlélési esély mint x és x+y között.

Legyen ugyanis $x < y$ tetszőleges, ekkor

$$P(\xi < y | \xi \geq x) = \frac{P(x \leq \xi < y)}{P(\xi \geq x)} = \frac{F_\xi(y) - F_\xi(x)}{1 - F_\xi(x)} = \frac{1 - e^{-\lambda y} - 1 + e^{-\lambda x}}{1 - 1 + e^{-\lambda x}} =$$
$$= 1 - e^{-\lambda(y-x)} = P(\xi < y - x)$$

A tétel megfordítása is igaz, vagyis csak az exponenciális eloszlás örökifjú a folytonos valószínűségi változók között. Az exponenciális eloszlást véletlen időtartamok modellezésére használják. Például exponenciális eloszlású két telefonhívás között eltelt idő, a fodrásznál eltöltött várakozási idő, egy berendezés hibamentes üzemelési ideje, stb.

Feladat Egy szobában öt telefon van, melyek közül bármelyik megszólalhat a többiektől teljesen függetlenül ξ időn belül, ahol ξ $\lambda=1$ paraméterű exponenciális eloszlású valószínűségi változó. Mennyi az esélye annak, hogy egységnyi időn belül pontosan két telefonkészülék fog csörögni?

Megoldás: Az „egy telefon megcsörren egységnyi időn belül” esemény valószínűsége: $p = P(\xi < 1) = F_\xi(1) = 1 - e^{-1}$. Mivel öt függetlenül üzemelő készülékünk van, a feladat átfogalmazható úgy, mintha az A eseményre vonatkozó ötszörös Bernoulli kísérletsorozatról volna szó. Így a binomiális eloszlást figyelembevéve, annak valószínűsége, hogy az A esemény pontosan kétszer következik be: $\binom{5}{2} p^2 (1-p)^3 = 10(1 - \frac{1}{e})^2 (\frac{1}{e})^3 \approx 0{,}1989$.

6.2.3 Példa *A normális eloszlású valószínűségi változó*

A ξ valószínűségi változó $\mu \in \mathbb{R}$ és $\sigma > 0$ *paraméterű normális eloszlású*, ha eloszlásfüggvénye

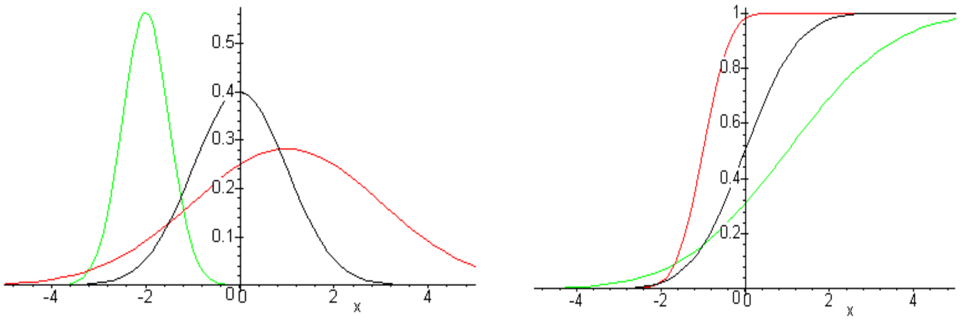
$$F_\xi(x) = \Phi_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad , x \in \mathbb{R}.$$

Jelölés: $\xi \in N(\mu, \sigma)$.

A ξ sűrűségfüggvénye: $f_\xi(x) = \varphi_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad , x \in \mathbb{R}$.

Ha $\xi \in N(0,1)$, akkor *standard normális eloszlásról* beszélünk. Ilyenkor

$$\varphi_{0,1}(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \text{ és } \Phi_{0,1}(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt .$$



. A normális eloszlás sűrűség - és eloszlásfüggvénye $(-1,0.5)$, $(0,1)$ és $(1,2)$ paraméterekkel.

Tétel: *(Transzformációs tulajdonságok)*

a.) $\Phi_{\mu,\sigma}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$,

b.) $\varphi_{\mu,\sigma}(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$,

vagyis a standard normális eloszlás sűrűségfüggvényével és eloszlásfüggvényével tetszőleges $\mu \in \mathrm{lR}$ és $\sigma > 0$ paraméterű normális eloszlású sűrűségfüggvény és eloszlásfüggvény előállítható.

Az előző tétel bizonyítása:

a.) $\Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{t^2}{2}}\,dt = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-\frac{(u-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sigma}\,du = \Phi_{\mu,\sigma}(x)$

$t = \frac{u-\mu}{\sigma}$, $\sigma t + \mu = u$, $\frac{dt}{du} = \frac{1}{\sigma}$.

b.) az a.) mindkét oldalát deriváljuk.

Tétel: *(A φ Gauss-függvény tulajdonságai)*

a.) $\varphi(-x) = \varphi(x)$, vagyis φ páros függvény,

b.) $\lim\limits_{x\to+\infty}\varphi(x) = \lim\limits_{x\to-\infty}\varphi(x) = 0$,

c.) $\frac{1}{\sqrt{2\pi}} = \varphi(0) \geq \varphi(x) > 0$, $\forall x \in \mathrm{lR}$,

d.) φ inflexiós helyei a +1 és -1, azaz $\varphi''(-1) = \varphi''(+1) = 0$,

e.) φ analitikus,

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

Tétel: *(A Φ eloszlásfüggvény tulajdonságai)*

a.) $\Phi(x) = 1 - \Phi(-x)$, $\forall x > 0$, azaz Φ grafikonja szimmetrikus a (0, 0,5)-ra,

b.) Φ szigorúan monoton növekedő,

c.) Φ analitikus, és

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\left(x - \frac{x^3}{1!2\cdot 3} + \cdots + \frac{(-1)^k x^{2k+1}}{k!2^k\cdot(2k+1)} + \cdots\right),\ \forall x > 0.,$$

d.) $\lim\limits_{x\to\infty}\Phi(x) = 1$, $\lim\limits_{x\to-\infty}\Phi(x) = 0$.

Feladat Egy automata zacskókba cukorkát adagol. A zacskók ξ súlyát μ=100 (gramm), σ=2 (gramm) paraméterű normális eloszlásúnak tekinthetjük. Mennyi a valószínűsége annak, hogy három véletlenszerűen kiválasztott zacskó között legalább egy olyan van, aminek a súlya 99 és 101 gramm közé fog esni?

Megoldás: Legyen A a „zacskó súlya 99 és 101 gramm közé esik" esemény. Az A bekövetkezésének valószínűségét a ξ eloszlásfüggvénye segítségével határozhatjuk meg:

$$P(A) = P(99 \le \xi < 101) = F_\xi(101) - F_\xi(99) = \Phi\left(\frac{101-100}{2}\right) - \Phi\left(\frac{99-100}{2}\right) = \Phi(0{,}5) - \Phi(-0{,}5) =$$

$= 2\Phi(0{,}5) - 1 \approx 0{,}383$. A három zacskó kiválasztása n=3 és P(A) paraméterű binomiális eloszlással modellezhető, ami alapján a keresett valószínűség:

$$1 - \binom{3}{0}(P(A))^0(1-P(A))^3 \approx 1-(1-0{,}383)^3 = 0{,}765114887.$$

Ellenőrző kérdések és gyakorló feladatok

1. Mi a valószínűségi változó definíciója?
2. Milyen tulajdonságok jellemzik egyértelműen az eloszlásfüggvényt?
3. Definiálja a binomiális eloszlást!
4. Mi a sűrűségfüggvénye a μ,σ paraméterű normális eloszlásnak?
5. Mit jelent az, hogy az exponenciális eloszlás örökifjú?
6. Döntse el, hogy az alábbi állítások közül melyik igaz, és melyik hamis!
   a.) A valószínűségi változó olyan valós értékű függvény, amelynek értelmezési tartománya a K véletlen kísérlet Ω eseménytere.
   b.) Minden Ω-t IR-be leképező függvény valószínűségi változó.
   c.) Egy véletlen kísérlethez több valószínűségi változót is értelmezni lehet.
   d.) Az eloszlásfüggvény szigorúan monoton növekedő.
   e.) Az eloszlásfüggvény értékei nemnegatívak.
   f.) Az eloszlásfüggvény folytonos.
   g.) Az eloszlásfüggvény balról folytonos függvény, aminek jobbról lehet elsőfajú szakadása.
   h.) Az eloszlásfüggvény nem veheti fel a 0 és az 1 értékeket csak határértékben.
   i.) Annak valószínűségét, hogy egy valószínűségi változó az értékeit egy intervallumban veszi fel, az eloszlásfüggvény segítségével meg lehet határozni.

j.) Ha egy pontban az eloszlásfüggvény folytonos, akkor az azt is jelenti, hogy azt a pontot a valószínűségi változó nulla valószínűséggel veszi fel.

k.) Ha a $\xi: \Omega \to IR$ függvény értékkészlete az irracionális számok halmaza, ξ nem lehet diszkrét valószínűségi változó.

l.) Ha a $\xi: \Omega \to IR$ függvény értékkészlete véges, akkor ξ csak diszkrét valószínűségi változó lehet.

m.) A folytonos valószínűségi változók eloszlásfüggvénye lépcsős.

n.) Egy n-szeres Bernoulli kísérletsorozatban a p valószínűségű A esemény gyakorisága binomiális eloszlású valószínűségi változó.

o.) Az egyenletes eloszlású valószínűségi változó sűrűségfüggvénye lépcsős.

p.) Az exponenciális eloszlás sűrűségfüggvénye páros.

q.) A μ,σ paraméterű normális eloszlás sűrűségfüggvénye szimmetrikus az $x=\mu$ függőleges tengelyre.

r.) A μ,σ paraméterű normális eloszlás eloszlásfüggvénye szimmetrikus a $(\mu, 0{,}5)$ pontra.

s.) A hipergeometriai eloszlással modellezhető a visszatevés nélküli mintavételezés.

t.) A binomiális eloszlással modellezhető a visszatevés nélküli mintavételezés.

u.) A normális eloszlás örökifjú tulajdonságú.

v.) Az exponenciális eloszlás örökifjú tulajdonságú.

w.) A geometriai eloszlású valószínűségű változó értékkészlete véges.

x.) A karakterisztikus eloszlás speciális binomiális eloszlás.

7. Legyen a ξ valószínűségi változó folytonos eloszlásfüggvénye olyan, hogy $1>F(x)>0$ esetben szigorúan monoton növekedő is. Bizonyítsa be, hogy ekkor az $\eta=F(\xi)$ valószínűségi változó egyenletes eloszlású a [0,1] intervallumon!
8. Legyen a ξ valószínűségi változó folytonos eloszlásfüggvénye olyan, hogy $1>F(x)>0$ esetben szigorúan monoton növekedő is. Bizonyítsa be, hogy ekkor az $\eta=\ln\dfrac{1}{F(\xi)}$ eloszlása $\lambda=1$ paraméterű exponenciális lesz!
9. Ha a ξ standard normális eloszlású valószínűségi változó, mi a sűrűségfüggvénye az $\eta=\xi^2$ valószínűségi változónak?
10. Ha a ξ valószínűségi változó sűrűségfüggvénye f(x), akkor mi a sűrűségfüggvénye az $\eta=|\xi|$ valószínűségi változónak?
11. Ha ξ λ-paraméterű Poisson eloszlású valószínűségi változó, akkor mi az eloszlása az $\eta=2\xi+1$ valószínűségi változónak?
12. Ha a ξ a [0,1] intervallumon egyenletes eloszlású valószínűségi változó, mi a sűrűségfüggvénye az $\eta=\dfrac{1}{\xi}$ és $\zeta=\dfrac{\xi}{1+\xi}$ valószínűségi változóknak?
13. Ha a ξ μ,σ paraméterű normális eloszlású valószínűségi változó, mi a sűrűségfüggvénye az $\eta=e^{\xi}$ valószínűségi változónak? (η az ú.n. lognormális eloszlású valószínűségi változó).
14. Ha ξ a [0,2] intervallumon egyenletes eloszlású, akkor mi a sűrűségfüggvénye az $\eta=|\xi-1|$ valószínűségi változónak?
15. Ha ξ λ-paraméterű exponenciális eloszlású valószínűségi változó, akkor mi a sűrűségfüggvénye az $\eta=3\xi+3$ valószínűségi változónak?

16. Ha ξ λ-paraméterű exponenciális eloszlású valószínűségi változó, akkor mi a sűrűségfüggvénye az $\eta=\sqrt{\xi}$ valószínűségi változónak?
17. Ha ξ λ-paraméterű exponenciális eloszlású valószínűségi változó, akkor mi a sűrűségfüggvénye az $\eta=\dfrac{1}{\xi^2}$ valószínűségi változónak?
18. Egy szabályos pénzdarabbal végzünk dobásokat. A pénzfeldobást addig folytatjuk, amíg a dobások sorozatában mind a fej, mind az írások száma eléri a k számot. Jelölje ξ az ehhez szükséges dobások száma. Adja meg a ξ eloszlását!
19. A [0,1] szakaszon véletlenszerűen kiválasztunk két pontot. Legyen ξ a két pont távolsága. Adja meg ξ sűrűségfüggvényét!

7. Vektor valószínűségi változók, valószínűségi változók együttes eloszlása

Nagyon gyakran nem lehet a véletlen jelenséget egyetlen számadattal jellemezni. Pl. amikor az időjárási helyzetet próbálják előrejelezni, megadják a várható hőmérséklet, csapadékösszeg, légnyomás, szélerősség stb. adatokat, azaz a prognosztizált helyzetet egy vektorral jellemzik. A vektor komponensei valószínűségi változók, értékeik a véletlentől függnek. Felmerülhet az egyes komponensek között fennálló kapcsolatok kérdése is.

Definíció: Legyen $(\Omega,\Im,P)$ Kolmogorov- féle valószínűségi mező. Tekintsük a $\underline{\xi}:\Omega \to \mathrm{lR}^p$ függvényt! A $\underline{\xi} = (\xi_1,\xi_2,\ldots,\xi_p)^T$ vektor valószínűségi változó, ha minden $\underline{x} \in \mathrm{IR}^p$ p-dimenziós vektor esetén $\left\{\omega | \xi_1(\omega) < x_1, \xi_2(\omega) < x_2, \cdots, \xi_p(\omega) < x_p\right\} \in \Im$ teljesül.

Tétel: A $\underline{\xi}$ vektor valószínűségi változó $\Leftrightarrow$ Mindegyik komponense valószínűségi változó.

Definíció: Legyen $(x_1,x_2,\ldots,x_p)^T = \underline{x} \in \mathrm{lR}^p$. Ekkor az $F_{\underline{\xi}}(\underline{x}) = F_{\xi_1,\xi_2,\ldots,\xi_p}(x_1,x_2,\ldots,x_p) = P(\xi_1 < x_1, \xi_2 < x_2, \ldots, \xi_p < x_p)$ p-változós skalár-vektor függvényt a $\underline{\xi}$ vektor valószínűségi változó *eloszlásfüggvényének*, illetve a $\xi_1,\xi_2,\ldots,\xi_p$ komponens valószínűségi változók *együttes eloszlásfüggvényének* nevezzük.

Tétel: *(Az együttes eloszlásfüggvény tulajdonságai)*

a.) $F_{\underline{\xi}}$ minden változójában monoton nem csökkenő függvény, azaz $\forall$ i-re ha $x_i^* < x_i^{**}$, akkor
$F_{\underline{\xi}}(x_1,\ldots,x_i^*,\ldots,x_p) \le F_{\underline{\xi}}(x_1,\ldots,x_i^{**},\ldots,x_p)$.

b.) $F_{\underline{\xi}}$ minden változójában balról folytonos függvény, azaz
$$\lim_{y\to x_i^0-0} F_{\underline{\xi}}(x_1,\ldots,y,\ldots,x_p) = F_{\underline{\xi}}(x_1,\ldots,x_i^0,\ldots,x_p) .$$

c.) $\lim\limits_{\forall x_i\to+\infty} F_{\underline{\xi}}(x_1,\ldots,x_i,\ldots,x_p) = 1$ és $\lim\limits_{\exists x_i\to-\infty} F_{\underline{\xi}}(x_1,\ldots,x_i,\ldots,x_p) = 0$.

d.) Legyen $T = \left[a_1,b_1\right) \times \left[a_2,b_2\right) \times \cdots \times \left[a_p,b_p\right)$ tetszőleges p-dimenziós tégla, és $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_p \in \{0,1\}$ tetszőlegesek (0 vagy 1 - diadikus -számok) . Akkor
$$\sum_{\forall(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_p)} (-1)^j F_{\underline{\xi}}(\varepsilon_1 a_1 + (1-\varepsilon_1)b_1, \varepsilon_2 a_2 + (1-\varepsilon_2)b_2, \ldots, \varepsilon_p a_p + (1-\varepsilon_p)b_p) \ge 0, \text{ahol } j = \sum_{i=1}^{p} \varepsilon_i .$$

A d.) állítás nem szerepelt az egydimenziós esetben, akkor a neki megfelelő alak a tetszőleges $\left[a_1,b_1\right)$ intervallum esetén $F_\xi(b_1) - F_\xi(a_1) \ge 0$, ami a monotonitási tulajdonsággal esik egybe.

Többdimenziós esetben szükség van d.)-re, mert pl. p=2 esetben a

$$F(x_1, x_2) = \begin{cases} 0 & \text{, ha } x_1 + x_2 \le 0 \\ 1 & \text{, ha } x_1 + x_2 > 0 \end{cases}$$

függvény kielégíti a.),b.) és c.)-t , de d.) nem teljesül rá. A bizonyítás azon múlik, hogy megmutathatjuk, hogy d.) jobboldalán a $P(\underline{\xi} \in T)$ valószínűség áll, ami nyilvánvalóan nemnegatív.

Tétel: Ha $F(\underline{x})$ tetszőleges, az előző tétel a.) - d.) tulajdonságaival rendelkező skalár-vektor függvény, akkor megadható olyan $(\Omega,\Im,P)$ Kolmogorov- féle valószínűségi mező és hozzá olyan $\underline{\xi}$ vektor valószínűségi változó melynek eloszlásfüggvénye éppen $F(\underline{x})$.

Definíció: Ha $\underline{\xi} = (\xi_1, \xi_2, \ldots, \xi_p)^T$ vektor valószínűségi változó eloszlásfüggvénye $F_{\underline{\xi}}$ és $1 \le j_1 < j_2 < \cdots < j_k \le p$ egy tetszőleges k elemű index kombináció, akkor az indexekhez tartozó $\xi_{j_1}, \xi_{j_2}, \ldots, \xi_{j_k}$ komponens valószínűségi változók együttes eloszlásfüggvénye az $F_{\underline{\xi}}$ egy k-dimenziós *perem* vagy *vetületi eloszlásfüggvénye.*

Tétel: Ha a $\xi_1, \xi_2, \ldots, \xi_p$ valószínűségi változók együttes eloszlásfüggvénye $F_{\underline{\xi}}$ ismert, akkor bármely vetületi eloszlásfüggvénye meghatározható. Fordítva általában nem igaz: ha ismerjük az összes alacsonyabb dimenziós vetületi eloszlásfüggvényt, az együttes eloszlásfüggvény nem állítható elő.

Hasonló az eset, mint a geometriában a testeknél. A test vetülete bármely síkra vonatkozóan egyértelműen képezhető, de a vetületek ismeretében nem feltétlenül állítható vissza a térbeli alakzat. A megfelelő vetületi eloszlásfüggvényt az együttes eloszlásfüggvényből határátmenettel kaphatjuk:

$$F_{\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_k}}(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) = \lim_{\substack{\forall x_j \to \infty \\ j \notin \{i_1, i_2, \ldots, i_k\}}} F_{\xi_1, \xi_2, \ldots, \xi_p}(x_1, x_2, \ldots, x_p) .$$

Arra, hogy a fordított állítás nem igaz, p=2 esetben adunk ellenpéldát:
Legyenek ξ_1 és ξ_2 olyan valószínűségi változók, melyek csak a -1,0 és +1 értékeket vehetik fel az alábbi eloszlástáblázat szerint

$\xi_1 \setminus \xi_2$	-1	0	+1	ξ_1 perem
-1	$0.125+\varepsilon$	0	$0.125-\varepsilon$	0.25
0	0	0.5	0	0.5
+1	$0.125-\varepsilon$	0	$0.125+\varepsilon$	0.25
ξ_2 perem	0.25	0.5	0.25	1

ahol $0 < \varepsilon < 0.125$ tetszőleges.

Ekkor

$$F_{\xi_i}(x) = \begin{cases} 0 & , \quad x \le -1 \\ 0.25 & , -1 < x \le 0 \\ 0.75 & , \quad 0 < x \le 1 \\ 1 & , \quad x > 1 \end{cases}$$

a két vetületi eloszlásfüggvény, ami nyilván nem határozza meg az együttes eloszlásfüggvényt, mely az ε paramétert is tartalmazza.

Az együttes eloszlásfüggvény segítségével értelmezhetjük a függetlenség fogalmakat valószínűségi változók között.

Definíció: Legyenek $\xi_1, \xi_2, \dots, \xi_p$ valószínűségi változók.

a.) $\xi_1, \xi_2, \dots, \xi_p$ *páronként függetlenek*, ha $\forall \;\; 1 \le i < j \le n$-re $F_{\xi_i, \xi_j}(x, y) = F_{\xi_i}(x) \cdot F_{\xi_j}(y)$ teljesül $\forall \;\; x, y \in \mathrm{lR}$-re.

b.) $\xi_1, \xi_2, \dots, \xi_p$ *teljesen függetlenek*, ha $\forall \;\; 2 \le k \le p$ és $\forall \; 1 \le i_1 < i_2 < \dots < i_k \le p$ index kombinációra $F_{\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_k}}(x_{i_1}, x_{i_2}, \dots, x_{i_k}) = \prod_{j=1}^{k} F_{\xi_{i_j}}(x_{i_j})$, $\forall x_{i_1}, x_{i_2}, \dots, x_{i_k} \in \mathrm{lR}$-re.

A valószínűségi változók páronként (illetve teljesen) függetlenek, ha velük kapcsolatos bármely nívó-eseményrendszer páronként (illetve teljesen) független eseményekből áll.

Tétel: Ha $\xi_1, \xi_2, \dots, \xi_p$ teljesen függetlenek , akkor páronként is függetlenek. A megfordítás általában nem igaz.

A tétel első fele nyilvánvaló, hisz a teljesen függetlenség feltételrendszere a páronkénti függetlenség feltételrendszerét is tartalmazza. Az ellenpélda ugyanazon a kockadobásos példán alapulhat, amikor megmutattuk, hogy a páronként független események rendszere nem feltétlenül alkot teljesen független eseményrendszert.

Definíció: a.) Ha ξ és η diszkrét valószínűségi változók $X = \{x_1, x_2, \dots, x_n, \dots\}$ illetve $Y = \{y_1, y_2, \dots, y_n, \dots\}$ értékkészletekkel, akkor az $r_{ij} = P\left(\{\omega \mid \xi(\omega) = x_i\} \cap \{\omega \mid \eta(\omega) = y_j\}\right) \overset{\text{jel}}{=} P(\xi = x_i, \eta = y_j)$ (i,j=1,2,...) valószínűségek összességét a két diszkrét valószínűségi változó *együttes eloszlásának* nevezzük.

b.) A $\xi_1, \xi_2, \dots, \xi_p$ diszkrét valószínűségi változók értékkészleteit jelölje rendre $X^{(i)} = \{x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}, \dots\}$ (i=1,2,...,p). Ekkor a $r_{i_1, i_2, \dots, i_p} = P(\xi_1 = x_{i_1}^{(1)}, \xi_2 = x_{i_2}^{(2)}, \dots, \xi_p = x_{i_p}^{(p)})$ valószínűségek összessége a $\xi_1, \xi_2, \dots, \xi_p$ diszkrét valószínűségi változók *együttes eloszlása* .

Definíció: Ha adott a $\xi_1, \xi_2, \ldots, \xi_p$ diszkrét valószínűségi változók $\left\{ r_{i_1, i_2, \ldots, i_p} = P(\xi_1 = x_{i_1}^{(1)}, \xi_2 = x_{i_2}^{(2)}, \ldots, \xi_p = x_{i_p}^{(p)}) \ , \forall i_k \right\}$ együttes eloszlása, és $1 \le j_1 < j_2 < \ldots < j_k \le p$, akkor a $\xi_{j_1}, \xi_{j_2}, \ldots, \xi_{j_k}$ diszkrét valószínűségi változók együttes eloszlását *k-dimenziós vetületi-* vagy *peremeloszlásnak* nevezzük.

Tétel: A diszkrét valószínűségi változók együttes eloszlása kielégíti az alábbi tulajdonságokat:

a.) $0 \le r_{i_1, i_2, \ldots, i_p} \le 1$

b.) $\sum\limits_{\forall i_1, i_2, \ldots, i_p} r_{i_1, i_2, \ldots, i_p} = 1$

c.) $P(\xi_{j_1} = x_{i_{j_1}}^{(j_1)}, \xi_{j_2} = x_{i_{j_2}}^{(j_2)}, \ldots, \xi_{j_k} = x_{i_{j_k}}^{(j_k)}) = \sum\limits_{\forall i_\alpha \notin \left\{ i_{j_1}, i_{j_2}, \ldots, i_{j_k} \right\}} r_{i_1, i_2, \ldots, i_p}$

Az a.) állítás nyilvánvaló, hiszen az $A_{i_1, i_2, \ldots, i_p} = \left\{ \omega \middle| \ \xi_1(\omega) = x_{i_1}^{(1)} \right\} + \left\{ \omega \middle| \ \xi_2(\omega) = x_{i_2}^{(2)} \right\} + \cdots + \left\{ \omega \middle| \ \xi_p(\omega) = x_{i_p}^{(p)} \right\}$ esemény valószínűségéről van szó. Mivel az $A_{i_1, i_2, \ldots, i_p}$ események teljes eseményrendszert alkotnak, igaz a b.) állítás. A c.) állítás speciálisan a p=2 esetben: $P(\xi = x_i) = \sum\limits_{j=1}^{\infty} P(\xi = x_i, \eta = y_j)$ és $P(\eta = y_j) = \sum\limits_{i=1}^{\infty} P(\xi = x_i, \eta = y_j)$.

Tétel:

a.) A ξ és η diszkrét valószínűségi változók függetlenek, ha $\forall$ i, j-re $P(\xi = x_i, \eta = y_j) = P(\xi = x_i) \cdot P(\eta = y_j)$.

b.) A $\xi_1, \xi_2, \ldots, \xi_p$ diszkrét valószínűségi változók teljesen függetlenek, ha $\forall\ 2 \le k \le p$-re és $\forall\ 1 \le j_1 < j_2 < \ldots < j_k \le p$ esetén

$P(\xi_{j_1} = x_{i_{j_1}}^{(j_1)}, \xi_{j_2} = x_{i_{j_2}}^{(j_2)}, \ldots, \xi_{j_k} = x_{i_{j_k}}^{(j_k)}) = \prod\limits_{\alpha=1}^{k} P(\xi_{j_\alpha} = x_{i_{j_\alpha}}^{(j_\alpha)})$.

Látható, hogy a valószínűségi változók páronkénti (illetve teljesen) függetlensége ekvivalens a kapcsolatos $A_i = \left\{ \omega \middle| \ \xi(\omega) = x_i \right\}$ nívóesemények páronkénti (illetve teljesen) függetlenségével.

Feladat Két szabályos kockát feldobunk. ξ jelentse a hatos dobások számát, η pedig a dobott számok összegét. Adjuk meg ξ és η együttes eloszlását!

Megoldás: Az alábbi táblázatban az oszlopok tetején szerepelnek a ξ lehetséges értékei, a sorok elején pedig az η értékkészletének megfelelő számok állnak. Az (i,j) koordinátáknak megfelelő cellában a P(ξ=i,η=j) valószínűségek találhatók.

$\eta \backslash \xi$	0	1	2	η peremeloszlása
2	$\frac{1}{36}$	0	0	$\frac{1}{36}$
3	$\frac{2}{36}$	0	0	$\frac{2}{36}$
4	$\frac{3}{36}$	0	0	$\frac{3}{36}$
5	$\frac{4}{36}$	0	0	$\frac{4}{36}$
6	$\frac{5}{36}$	0	0	$\frac{5}{36}$
7	$\frac{4}{36}$	$\frac{2}{36}$	0	$\frac{6}{36}$
8	$\frac{3}{36}$	$\frac{2}{36}$	0	$\frac{5}{36}$
9	$\frac{2}{36}$	$\frac{2}{36}$	0	$\frac{4}{36}$
10	$\frac{1}{36}$	$\frac{2}{36}$	0	$\frac{3}{36}$
11	0	$\frac{2}{36}$	0	$\frac{2}{36}$
12	0	0	$\frac{1}{36}$	$\frac{1}{36}$
ξ peremeloszlása	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$	1

Például a táblázat nyolcadik sorának és második oszlopának kereszteződésében azért áll $\frac{2}{36}$, mert a 36 dobási lehetőségből csak kettő felel meg a $\xi=1,\eta=8$ feltételeknek, a (6,2) és a (2,6). Az η eloszlását a sorokban álló valószínűségek összeadásával, a ξ eloszlását pedig az oszlopokban álló valószínűségek összeadásával kapjuk meg. Látható az is, hogy ξ nem független η-tól, hiszen pl. $P(\xi = 2, \eta = 2) = 0 \neq \frac{25}{36^2} = P(\xi = 2)P(\eta = 2)$.

7.1 Példa *Polinomiális eloszlás*

Legyen $(\Omega,\Im,P)$ Kolmogorov- féle valószínűségi mező, $A_1,A_2,\ldots,A_r \in \Im$ egy r eseményből álló teljes eseményrendszer, azaz $A_i \cdot A_j = \emptyset$, $\sum_{i=1}^{r} A_i = \Omega$. Ekkor $0 < P(A_i) = p_i$ esetben $\sum_{i=1}^{r} p_i = 1$. Hajtsunk végre egy n-szeres Bernoulli-féle kísérletsorozatot! Vegye fel ξ_i azt az értéket, ahányszor A_i bekövetkezett a kísérletsorozatban. A $\xi_1,\xi_2,\ldots,\xi_r$ valószínűségi változók együttes eloszlását $n,p_1,p_2,\ldots,p_r$ paraméterű *polinomiális eloszlásnak* nevezzük. A ξ_i valószínűségi változók értékei a 0,1,2,...,n számok közé esnek. A $\xi_1,\xi_2,\ldots,\xi_r$ valószínűségi változók értékei között szoros összefüggés van: $\sum_{i=1}^{r} \xi_i = n$. A $\xi_1,\xi_2,\ldots,\xi_r$ valószínűségi változók együttes eloszlása:

$$P(\xi_1 = k_1, \xi_2 = k_2, \ldots, \xi_r = k_r) = \frac{n!}{k_1!k_2!\cdots k_r!} p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}.$$

A fenti valószínűségek valóban eloszlást alkotnak, hiszen:

$$\sum_{\substack{\forall k_i=0 \\ k_1+k_2+\ldots+k_r=n}}^{n} \frac{n!}{k_1!k_2!\cdots k_r!} p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r} = (p_1 + p_2 + \cdots + p_r)^n = 1^n = 1.$$

A polinomiális eloszlás a binomiális eloszlás többdimenziós kiterjesztése. A polinomiális eloszlás ξ_i komponensei egyenként $B(n,p_i)$ eloszlásúak, azaz a polinomiális eloszlás egydimenziós peremeloszlásai binomiálisak.

Definíció: A $\xi_1,\xi_2,\ldots,\xi_p$ folytonos valószínűségi változók együttes sűrűségfüggvényén azt az $f_{\xi_1,\xi_2,\ldots,\xi_p}(x_1,x_2,\ldots,x_p)$ függvényt értjük, melyre

$$F_{\xi_1,\xi_2,\ldots,\xi_p}(x_1,x_2,\ldots,x_p) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_p} f_{\xi_1,\xi_2,\ldots,\xi_p}(t_1,t_2,\ldots,t_p)\, dt_p \ldots dt_2\, dt_1 \text{ , azaz}$$

$$\frac{\partial^p F_{\xi_1,\xi_2,\ldots,\xi_p}(x_1,x_2,\ldots,x_p)}{\partial x_1 \partial x_2 \cdots \partial x_p} = f_{\xi_1,\xi_2,\ldots,\xi_p}(x_1,x_2,\ldots,x_p)$$, ha $\underline{x} = (x_1,x_2,\ldots,x_p)^T$ folytonossági pontja $f_{\xi_1,\xi_2,\ldots,\xi_p}(x_1,x_2,\ldots,x_p)$-nek.

Definíció: Az $f_{\xi_1,\xi_2,\ldots,\xi_p}(x_1,x_2,\ldots,x_p)$ együttes sűrűségfüggvény egy *k-dimenziós vetületi sűrűségfüggvényén* $(2 \le k \le p-1)$ valamely $1 \le i_1 < i_2 < \cdots < i_k \le p$ indexkombinációra a $\xi_{i_1},\xi_{i_2},\ldots,\xi_{i_k}$ valószínűségi változók együttes sűrűségfüggvényét értjük.

Tétel: $f_{\xi_{i_1},\xi_{i_2},\dots,\xi_{i_k}}(x_{i_1},x_{i_2},\dots,x_{i_k}) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f_{\xi_1,\xi_2,\dots,\xi_p}(t_1,t_2,\dots,t_p)\,dt_{j_1}\dots dt_{j_{p-k-1}}\,dt_{j_{p-k}}$,
azaz az együttes sűrűségfüggvényt az összes többi, a kiválasztott indexkombinációban nem szereplő indexhez tartozó változóra kell kiintegrálni a teljes számegyenesen, hogy előállítsuk a k-dimenziós vetület sűrűségfüggvényt ($j_1,j_2,\dots,j_{p-k} \notin \{i_1,i_2,\dots,i_k\}$).

Tétel: Legyenek $\xi_1,\xi_2,\dots,\xi_p$ folytonos valószínűségi változók az $(\Omega,\Im,P)$ Kolmogorov- féle valószínűségi mezőn.

a.) $\xi_1,\xi_2,\dots,\xi_p$ *páronként függetlenek* $\Leftrightarrow \forall\ 1 \le i < j \le n$-re
$f_{\xi_i,\xi_j}(x,y) = f_{\xi_i}(x)\cdot f_{\xi_j}(y)$ teljesül $\forall\ x,y \in \mathrm{lR}$-re.

b.) $\xi_1,\xi_2,\dots,\xi_p$ *teljesen függetlenek* $\Leftrightarrow \forall\ 2 \le k \le p$ és $\forall\ 1 \le i_1 < i_2 < \dots < i_k \le p$
indexkombinációra $f_{\xi_{i_1},\xi_{i_2},\dots,\xi_{i_k}}(x_{i_1},x_{i_2},\dots,x_{i_k}) = \prod_{j=1}^{k} f_{\xi_{i_j}}(x_{i_j}), \forall x_{i_1},x_{i_2},\dots,x_{i_k} \in \mathrm{lR}$.

Az függetlenség definícióból egyszerűen deriválással következik az állítás.

p=2 esetben az előző tételek speciális alakjai:

$$\frac{\partial^2 F_{\xi,\eta}(x,y)}{\partial x \partial y} = f_{\xi,\eta}(x,y)\ ,\ f_\xi(x) = \int_{-\infty}^{+\infty} f_{\xi,\eta}(x,y)\,dy\ ,\ f_\eta(y) = \int_{-\infty}^{+\infty} f_{\xi,\eta}(x,y)\,dx\ ,$$

A ξ és η függetlenek $\Leftrightarrow f_{\xi,\eta}(x,y) = f_\xi(x) f_\eta(y)\ (\forall\, x,y \in \mathrm{lR})$.

Tétel: *(Az együttes sűrűségfüggvény tulajdonságai)*

a.) $f_{\xi_1,\xi_2,\dots,\xi_p}(x_1,x_2,\dots,x_p) \ge 0$

b.) $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f_{\xi_1,\xi_2,\dots,\xi_p}(t_1,t_2,\dots,t_p)\,dt_p\dots dt_2\,dt_1 = 1$.

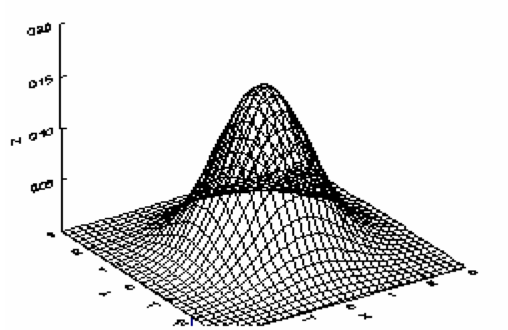
7.2 Példa *A kétdimenziós normális eloszlás*

Amennyiben a (ξ,η) pár együttes eloszlását az

$$f_{\xi,\eta}(x,y) = \frac{1}{2\cdot\pi\cdot\sigma_1\cdot\sigma_2\cdot\sqrt{1-\rho^2}}\cdot e^{-\frac{1}{2\cdot(1-\rho^2)}\cdot\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]} \qquad (x,y \in \mathrm{lR})$$

együttes sűrűségfüggvénnyel lehet leírni azt mondjuk, hogy a két valószínűségi változó együttes eloszlása kétdimenziós normális, ahol a peremeloszlásokra $\xi \in N(\mu_1,\sigma_1), \eta \in N(\mu_2,\sigma_2)$ teljesül. (A képletben $-1 \le \rho \le 1$).

A kétdimenziós normális eloszlás sűrűségfüggvénye egy olyan felületet ír le, melynek minden, az x-y síkra merőleges, a (μ_1,μ_2) pontot tartalmazó síkkal való metszete Gauss-féle haranggörbe, míg az x-y síkkal párhuzamos nemüres síkmetszetei ellipszisek.



A kétdimenziós normális eloszlás sűrűségfüggvénye

Ellenőrző kérdések és gyakorló feladatok

1. Mi az együttes eloszlásfüggvény és a peremeloszlásfüggvény fogalma, és mi a kapcsolat közöttük?
2. Mikor teljesen függetlenek a ξ_1, ξ_2, ξ_3 valószínűségi változók?
3. Hogyan nevezzük a binomiális eloszlás többdimenziós megfelelőjét?
4. Mik az együttes sűrűségfüggvény tulajdonságai?
5. Az alábbiak közül melyik állítás helyes és melyik hamis?
    a. Az együttes eloszlásfüggvény többváltozós valós függvény.
    b. Az együttes eloszlásfüggvény értékei az 1-hez tartanak, ha valamelyik változójával $+\infty$-hez tartunk.
    c. Az együttes eloszlásfüggvény értékei az 0-hez tartanak, ha valamelyik változójával $-\infty$-hez tartunk.
    d. Diszkrét valószínűségi változók együttes eloszlása a peremeloszlások összegeként áll elő.
    e. Diszkrét valószínűségi változók peremeloszlásait az együttes eloszlásból összegzéssel számolhatjuk ki.
    f. Független folytonos valószínűségi változók együttes sűrűségfüggvénye a peremsűrűségfüggvények szorzata.
    g. Független folytonos valószínűségi változók együttes eloszlásfüggvénye a peremeloszlásfüggvények szorzata.
    h. Az együttes eloszlás elemeinek összege 1.
    i. A vektor valószínűségi változó minden komponense valószínűségi változó.
    j. Az együttes sűrűségfüggvény minden változójában folytonos.
6. A ξ és η valószínűségi változók együttes eloszlását tartalmazza az alábbi táblázat:

$\eta \backslash \xi$	-1	0	1
-1	p	3p	6p
1	5p	15p	30p

Mekkora a p paraméter értéke? Függetlenek-e ξ és η ?

7. Először egy szabályos kockával dobunk, majd a dobott értéknek megfelelően kihúzunk lapokat egy 32 lapos kártyatömegből. Jelölje ξ a kihúzott lapok között található figurás

lapok számát, η pedig legyen a kihúzott királyok száma. Adja meg a $P(\xi=4,\eta=2)$ valószínűséget!

8. Legyen a ξ és η együttes sűrűségfüggvénye $f(x,y)=2e^{-2x-y}$, $0<x,y<\infty$ (egyébként $f(x,y)=0$). Határozza meg a peremsűrűségfüggvényeket! Függetlenek-e ξ és η?
9. Legyen a ξ és η együttes sűrűségfüggvénye

$$f(x,y)=\begin{cases}\frac{4}{5}\left(x+xy+y\right), & \text{ha } 0<x<1 \text{ és } 0<y<1\\ 0, & \text{egyébként}\end{cases}$$

Határozza meg a peremsűrűségfüggvényeket! Függetlenek-e ξ és η?

8. Várható érték, szórás, szórásnégyzet, magasabb momentumok, kovariancia és a korrelációs együttható

Definíció: a.) A ξ diszkrét valószínűségi változónak akkor létezzék *várható értéke,* ha a $\sum_{i=1}^{\infty}|x_i|\cdot P(\xi = x_i)$ sor konvergens. Ekkor a ξ várható értékén az $\mathbf{M}\xi=\sum_{i=1}^{\infty} x_i P(\xi = x_i)$ sorösszeget értjük.
b.) A ξ folytonos valószínűségi változónak akkor létezzék *várható értéke,* ha az $\int_{-\infty}^{+\infty}|x|\cdot f_\xi(x)\,dx$ improprius integrál konvergens. Ekkor a ξ várható értékén az $\mathbf{M}\xi=\int_{-\infty}^{+\infty} x\cdot f_\xi(x)\,dx$ számot értjük.

Egy valószínűségi változónak nem feltétlenül létezik várható értéke. (Ld. A 10 számú gyakorló feladatot!)

Tétel: Legyen $g:\mathrm{lR}\to\mathrm{lR}$ tetszőleges valós függvény. Ekkor, ha az $\eta = g(\xi)$ valószínűségi változó, és létezik a várható értéke, akkor

a.) ha ξ diszkrét : $\mathbf{M}\eta=\sum_{i=1}^{\infty} g(x_i)\cdot P(\xi = x_i)$

b.) ha ξ folytonos: $\mathbf{M}\eta=\int_{-\infty}^{+\infty} g(x)\cdot f_\xi(x)\,dx$.

Tétel: Legyen a ξ valószínűségi változó várható értéke $\mathbf{M}\xi$. Ekkor az $\eta = a\cdot\xi + b$ valószínűségi változónak is létezik várható értéke, és $\mathbf{M}\eta=a\cdot\mathbf{M}\xi+b$.

Alkalmazzuk a megelőző tételt a $g(x)=a\cdot x+b$ lineáris függvényre !
a.) diszkrét eset:

$$\mathbf{M}\eta=\sum_{i=1}^{\infty}(a\cdot x_i + b)p_i = a\cdot\sum_{i=1}^{\infty} x_i\cdot p_i + b\cdot\sum_{i=1}^{\infty} p_i = a\cdot\mathbf{M}\xi+b\cdot 1.$$

b.) folytonos eset:

$$\mathbf{M}\eta=\int_{-\infty}^{+\infty}(ax + b)f_\xi(x)\,dx = a\cdot\int_{-\infty}^{+\infty} xf_\xi(x)\,dx + b\cdot\int_{-\infty}^{+\infty} f_\xi(x)\,dx = a\cdot\mathbf{M}\xi+b\cdot 1.$$

Következmény: A konstans valószínűségi változó várható értéke önmaga.

Tétel: a.) A $\xi_1,\xi_2,...,\xi_p$ diszkrét valószínűségi változók értékkészleteit jelölje rendre $X^{(i)}=\{x_1^{(i)},x_2^{(i)},\dots,x_n^{(i)},\dots\}$ (i=1,2,...,p) , együttes eloszlásukat pedig

$\{\ r_{i_1,i_2,...,i_p}=P(\xi_1=x_{i_1}^{(1)},\xi_2=x_{i_2}^{(2)},\dots,\xi_p=x_{i_p}^{(p)})\ \}$. Legyen $g:\mathrm{lR}^p\to\mathrm{lR}$ tetszőleges p-változós valós függvény.

Akkor ha az $\eta=g(\xi_1,\xi_2,...,\xi_p)$ valószínűségi változó és létezik a várható értéke,

$$\mathbf{M}\eta=\sum_{\forall(i_1,i_2,...,i_p)} g(x_{i_1}^{(1)},x_{i_2}^{(2)},\dots,x_{i_p}^{(p)})P(\xi_1=x_{i_1}^{(1)},\xi_2=x_{i_2}^{(2)},\dots,\xi_p=x_{i_p}^{(p)})\ .$$

b.) A $\xi_1,\xi_2,...,\xi_p$ folytonos valószínűségi változók együttes sűrűségfüggvényét jelölje $f_{\xi_1,\xi_2,...,\xi_p}(x_1,x_2,...,x_p)$.Legyen $g:\mathrm{lR}^p\to\mathrm{lR}$ tetszőleges p-változós valós függvény.

Akkor ha az $\eta=g(\xi_1,\xi_2,...,\xi_p)$ valószínűségi változó és létezik a várható értéke,

$$\mathbf{M}\eta=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} g(x_1,x_2,\dots,x_p)\cdot f_{\xi_1,\xi_2,...,\xi_p}(x_1,x_2,\dots,x_p)\,dx_p\cdots dx_2dx_1\ .$$

Tétel: Az $\underline{\eta=\xi_1+\xi_2+\dots+\xi_p}$ valószínűségi változó várható értéke létezik, amennyiben a ξ_i tagok várható értéke létezik, és $\mathbf{M}\eta=\mathbf{M}\xi_1+\mathbf{M}\xi_2+...+\mathbf{M}\xi_p$.

Az előző tétel következménye, amikor $g(x_1,x_2,...,x_p)=x_1+x_2+...+x_p$.

Tétel: Legyenek a ξ és η valószínűségi változók függetlenek, létezzék a várható értékük. Akkor a $\zeta=\xi\cdot\eta$ valószínűségi változónak is létezik a várható értéke, és $\mathbf{M}\zeta=\mathbf{M}\xi\cdot\mathbf{M}\eta$.

Legyen most $g(x,y)=x\cdot y$!

a.) diszkrét eset:

$$\mathbf{M}\zeta=\sum_{\forall i}\sum_{\forall j}x_iy_jP(\xi=x_i,\eta=y_j)=\sum\sum x_iy_jP(\xi=x_i)P(\eta=y_j)=$$

$$=\left(\sum_{\forall i}x_iP(\xi=x_i)\right)\cdot\left(\sum_{\forall j}y_jP(\eta=y_j)\right)=\mathbf{M}\xi\cdot\mathbf{M}\eta\ .$$

b.) folytonos eset:

$$\mathbf{M}\zeta=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf_{\xi,\eta}(x,y)\,dy\,dx=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf_\xi(x)f_\eta(y)\,dy\,dx=\left(\int_{-\infty}^{+\infty}xf_\xi(x)\,dx\right)\cdot\left(\int_{-\infty}^{+\infty}yf_\eta(y)\,dy\right)=$$

$$=\mathbf{M}\xi\cdot\mathbf{M}\eta\ .$$

Definíció: A ξ valószínűségi változó n-edik *momentumán* a ξ^n valószínűségi változó várható értékét értjük, ha az létezik. Jelölés: $\mu_n=\mathbf{M}\xi^n$.

Diszkrét esetben: $\mu_n = \sum_{i=1}^{\infty} x_i^{\,n} P(\xi = x_i)$

Folytonos esetben $\mu_n = \int_{-\infty}^{+\infty} x^n \cdot f_\xi(x)\,dx$.

Definíció: A ξ valószínűségi változó *szórásnégyzetén* vagy *varianciáján* az $\eta = (\xi - \mathbf{M}\xi)^2$ valószínűségi változó várható értékét értjük (amennyiben az létezik). Jelölés: $\mathbf{D}^2\,\xi = \mathbf{M}(\xi - \mathbf{M}\xi)^2$.
A ξ valószínűségi változó *szórása* a szórásnégyzet pozitív négyzetgyöke: $\mathbf{D}\xi = +\sqrt{\mathbf{M}(\xi - \mathbf{M}\xi)^2}$.

Diszkrét esetben: $\mathbf{D}^2\,\xi = \sum_{i=1}^{\infty} (x_i - \mathbf{M}\xi)^2 \cdot P(\xi = x_i)$.

Folytonos esetben : $\mathbf{D}^2\,\xi = \int_{-\infty}^{+\infty} (x - \mathbf{M}\xi)^2 \cdot f_\xi(x)\,dx$.

Tétel: Legyen ξ olyan valószínűségi változó , melynek létezik szórásnégyzete. Akkor minden valós x esetén:
$\mathbf{D}^2\xi = \mathbf{M}(\xi - \mathbf{M}\xi)^2 \le \mathbf{M}(\xi - x)^2$.

Legyen $g(x) = \mathbf{M}(\xi - x)^2 = \mathbf{M}(\xi^2 - 2\xi \cdot x + x^2) = \mathbf{M}\xi^2 - 2x \cdot \mathbf{M}\xi + x^2$.
Mivel $g'(x) = 2x - 2 \cdot \mathbf{M}\xi = 0$, $\Rightarrow$ $x = \mathbf{M}\xi$ és $g''(x) = 2 > 0$, ezért az $x = \mathbf{M}\xi$ hely minimumhely, ami már igazolja az állítást.

A ξ valószínűségi változó értékei a várható érték körül ingadoznak a legkisebb mértékben az összes valós szám közül, és ezt a minimális ingadozást, bizonytalanságot jellemzi a szórásnégyzet. Ha tehát egy valószínűségi változónak nagy a szórása, értékeit bizonytalanul tudjuk csak megbecsülni. Ha a szórásnégyzet egyre kisebb, a bizonytalanságunk a változó értékeit illetően csökken. Ad abszurdum, a konstans szórásnégyzete 0. A tétel megfordítása is igaz, azaz a 0 szórású „valószínűségi változó” a konstans.

Tétel: $\mathbf{D}^2\xi = 0 \iff P(\xi = \mathbf{M}\xi) = 1$.

Tétel: *(Steiner formula)*
$$\mathbf{D}^2\xi = \mathbf{M}(\xi - A)^2 - [\mathbf{M}(\xi - A)]^2 \text{ , minden } A \in \mathrm{lR}\text{-re.}$$
Speciálisan az A = 0-ra
$$\mathbf{D}^2\xi = \mathbf{M}\xi^2 - [\mathbf{M}\xi]^2 \ .$$

Bizonyítás: Legyen $A \in \mathrm{lR}$ tetszőleges !
$\mathbf{M}(\xi - A)^2 = \mathbf{M}(\xi^2 - 2A\xi + A^2) = \mathbf{M}\xi^2 - 2A \cdot \mathbf{M}\xi + A^2$,
$[\mathbf{M}(\xi - A)]^2 = [\mathbf{M}\xi - A]^2 = (\mathbf{M}\xi)^2 - 2A \cdot \mathbf{M}\xi + A^2$.

Így $\mathbf{M}(\xi-A)^2-[\mathbf{M}(\xi-A)]^2=\mathbf{M}\xi^2-[\mathbf{M}\xi]^2$.
Viszont .
$\mathbf{D}^2\xi=\mathbf{M}(\xi-\mathbf{M}\xi)^2=\mathbf{M}(\xi^2-2\xi\cdot\mathbf{M}\xi+\left[\mathbf{M}\xi\right]^2)=\mathbf{M}\xi^2-2\mathbf{M}\xi\cdot\mathbf{M}\xi+\left[\mathbf{M}\xi\right]^2=\mathbf{M}\xi^2-\left[\mathbf{M}\xi\right]^2$,
amiből már következik az állítás.

Következmény: Mivel $\mathbf{D}^2\xi=\mathbf{M}(\xi-\mathbf{M}\xi)^2\geq 0\Rightarrow\mathbf{M}\xi^2\geq\left[\mathbf{M}\xi\right]^2$, tehát, ha ξ második momentuma (így a szórásnégyzete is) létezik, akkor a várható értéknek is kell léteznie !

Tétel: $\mathbf{D}^2(a\xi+b)=a^2\cdot\mathbf{D}^2\xi$, minden a,b∈lR-re. Azaz a szórásnégyzet *eltolás invariáns*.

Bizonyítás: $\mathbf{D}^2(a\xi+b)=\mathbf{M}(a\xi+b)^2-[\mathbf{M}(a\xi+b)]^2=a^2\cdot\mathbf{M}\xi^2+2ab\mathbf{M}\xi+b^2-a^2\cdot[\mathbf{M}\xi]^2-$
$-2ab\cdot\mathbf{M}\xi-b^2=a^2\cdot[\mathbf{M}\xi^2-[\mathbf{M}\xi]^2]=a^2\cdot\mathbf{D}^2\xi$.

Tétel: Legyenek a ξ és η valószínűségi változók függetlenek, létezzék a szórásnégyzetük. Akkor

$$\mathbf{D}^2(\xi\pm\eta)=\mathbf{D}^2\xi+\mathbf{D}^2\eta\ .$$

Bizonyítás: $\mathbf{D}^2(\xi\pm\eta)=\mathbf{M}(\xi\pm\eta)^2-[\mathbf{M}(\xi\pm\eta)]^2=$
$=\mathbf{M}\left[\xi^2\pm 2\xi\cdot\eta+\eta^2\right]-\left[(\mathbf{M}\xi)^2\pm 2\mathbf{M}\xi\cdot\mathbf{M}\eta+(\mathbf{M}\eta)^2\right]=\mathbf{M}\xi^2\pm 2\mathbf{M}(\xi\cdot\eta)+\mathbf{M}\eta^2-$
$-(\mathbf{M}\xi)^2\mp 2\mathbf{M}\xi\cdot\mathbf{M}\eta-(\mathbf{M}\eta)^2=\mathbf{M}\xi^2-(\mathbf{M}\xi)^2+\mathbf{M}\eta^2-(\mathbf{M}\eta)^2=\mathbf{D}^2\xi+\mathbf{D}^2\eta$.
Felhasználtuk, hogy függetlenség esetén : $\mathbf{M}(\xi\cdot\eta)=(\mathbf{M}\xi)\cdot(\mathbf{M}\eta)$.

8.1 Nevezetes eloszlások várható értéke és szórásnégyzete

Diszkrét eloszlások

8.1.1. Példa *Karakterisztikus eloszlás*

Az eloszlás : $P(\xi=1)=p$, $P(\xi=0)=1-p=q$.
$\mathbf{M}\xi=1\cdot p+0\cdot q=p$, $\mathbf{D}^2\xi=(1-p)^2\cdot p+(0-p)^2\cdot q=q^2\cdot p+p^2\cdot q=p\cdot q(q+p)=pq$.

8.1.2. Példa *Binomiális eloszlás*

Az eloszlás : $p_k = P(\xi = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = \binom{n}{k} \cdot p^k \cdot q^{n-k}$, $k = 0,1,2,\ldots,n.$

$$\mathbf{M}\xi = \sum_{k=0}^{n} k \cdot p_k = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^{n} k \cdot \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k} =$$

$$= np \cdot \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-1-(k-1))!} p^{k-1} q^{n-1-(k-1)} = np \sum_{\alpha=0}^{n-1} \frac{(n-1)!}{\alpha!(n-1-\alpha)!} p^{k-1} q^{n-1-\alpha} =$$

$$= np \sum_{\alpha=0}^{n-1} \binom{n-1}{\alpha} p^{k-1} q^{n-1-\alpha} = np \cdot (p+q)^{n-1} = np \text{, azaz } \mathbf{M}\xi = np.$$

$$\mathbf{D}^2\xi = \mathbf{M}\xi^2 - [\mathbf{M}\xi]^2 \text{ , } \mathbf{M}\xi^2 = \sum_{k=0}^{n} k^2 \cdot p_k = \sum_{k=0}^{n} k^2 \cdot \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^{n} k^2 \cdot \binom{n}{k} p^k q^{n-k} =$$

$$= \sum_{k=1}^{n} k \cdot (k-1) \cdot \binom{n}{k} p^k q^{n-k} + \sum_{k=1}^{n} k \cdot \binom{n}{k} p^k q^{n-k} = \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} p^k q^{n-k} + \mathbf{M}\xi =$$

$$= n \cdot (n-1) \cdot p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-2-(k-2))!} \cdot p^{k-2} \cdot q^{n-2-(k-2)} + np =$$

$$= n \cdot (n-1) p^2 \sum_{\alpha=0}^{n-2} \binom{n-2}{\alpha} \cdot p^{\alpha} \cdot q^{n-2-\alpha} + np = n \cdot (n-1) \cdot p^2 \cdot (p+q)^{n-2} + np = n^2 p^2 - np^2 + np$$

Így $\mathbf{D}^2\xi = n^2p^2 - np^2 + np - (np)^2 = np(1-p) = npq.$

8.1.3 Példa *Poisson eloszlás*

Az eloszlás : $p_k = P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0,1,2,\ldots$

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

$$\mathbf{M}\xi^2 = \sum_{k=0}^{\infty} k^2 \cdot p_k = \sum_{k=1}^{\infty} k \cdot (k-1) \cdot p_k + \sum_{k=0}^{\infty} k \cdot p_k = \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-2)!} e^{-\lambda} + \lambda =$$

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

Így $\mathbf{D}^2\xi = \mathbf{M}\xi^2 - [\mathbf{M}\xi]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

8.1.4 Példa *Geometriai eloszlás*

Az eloszlás : $p_k = P(\xi = k) = (1-p)^{k-1}p = q^{k-1}p$, k=1,2,3,...

$$\mathbf{M}\xi = \sum_{k=1}^{\infty} k\cdot p_k = \sum_{k=1}^{\infty} k\cdot q^{k-1}\cdot p = p\sum_{k=1}^{\infty} k\cdot q^{k-1} = p\cdot\frac{1}{(1-q)^2} = \frac{1}{p}, \text{ hiszen } \sum_{k=0}^{\infty} q^k = \frac{1}{1-q} \text{ , és}$$

$$\frac{d}{dq}\sum_{k=0}^{\infty} q^k = \sum_{k=1}^{\infty} k\cdot q^{k-1} = \frac{d}{dq}\left(\frac{1}{1-q}\right) = \frac{1}{(1-q)^2}.$$

$$\mathbf{M}\xi^2 = \sum_{k=1}^{\infty} k^2\cdot p_k = \sum_{k=1}^{\infty} k\cdot(k-1)\cdot p_k + \sum_{k=1}^{\infty} k\cdot p_k = pq\sum_{k=2}^{\infty} k\cdot(k-1)q^{k-2} + \frac{1}{p} =$$

$$= pq\frac{2}{(1-q)^3} + \frac{1}{p} = \frac{2q}{p^2} + \frac{1}{p}.$$

Így $\mathbf{D}^2\xi = \mathbf{M}\xi^2 - [\mathbf{M}\xi]^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{p+2q-1}{p^2} = \frac{(p+q)+q-1}{p^2} = \frac{q}{p^2}.$

8.1.5 Példa *Hipergeometriai eloszlás*

Az eloszlás : $p_k = P(\xi = k) = \dfrac{\binom{F}{k}\cdot\binom{N-F}{n-k}}{\binom{N}{n}}$, $k = 0,1,\ldots,n.$

$$\mathbf{M}\xi = \sum_{k=0}^{n} k\cdot p_k = \sum_{k=1}^{n} k\cdot\frac{\binom{F}{k}\cdot\binom{N-F}{n-k}}{\binom{N}{n}} = \sum_{k=1}^{n}\frac{F\cdot\binom{F-1}{k-1}\cdot\binom{N-1-(F-1)}{n-1-(k-1)}}{\frac{N}{n}\binom{N-1}{n-1}} =$$

$$= n\frac{F}{N}\sum_{k=1}^{n}\frac{\binom{F-1}{k-1}\cdot\binom{N-1-(F-1)}{n-1-(k-1)}}{\binom{N-1}{n-1}} = n\frac{F}{N}, \text{ hiszen a szumma mögött a HG(n-1,F-1,N-1)}$$

eloszlás valószínűségei állnak, melyek összege 1.

$$\mathbf{M}\xi^2 = \sum_{k=1}^{n} k^2\cdot p_k = \sum_{k=2}^{n} k\cdot(k-1)\cdot p_k + \sum_{k=1}^{n} k\cdot p_k = \sum_{k=2}^{n} k\cdot(k-1)\frac{\binom{F}{k}\cdot\binom{N-F}{n-k}}{\binom{N}{n}} + n\frac{F}{N} =$$

$$= \sum_{k=2}^{n}\frac{F\cdot(F-1)\binom{F-2}{k-2}\cdot\binom{N-2-(F-2)}{n-2-(k-2)}}{\frac{N\cdot(N-1)}{n\cdot(n-1)}\binom{N-2}{n-2}} + n\frac{F}{N} = \frac{F\cdot(F-1)}{N\cdot(N-1)}n\cdot(n-1) + n\frac{F}{N} =$$

$$= n\frac{F}{N}\left(\frac{(F-1)\cdot(n-1)+N-1}{N-1}\right).$$

Így $\mathbf{D}^2\xi = \mathbf{M}\xi^2 - [\mathbf{M}\xi]^2 = n\frac{F}{N}\left(\frac{(F-1)\cdot(n-1)+N-1}{N-1}\right) - n^2\frac{F^2}{N^2} = n\frac{F}{N}\left(\frac{N-F}{N}\right)\left(\frac{N-n}{N-1}\right).$

Folytonos eloszlások

8.1.6 Példa *Egyenletes eloszlás*

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

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

$$\mathbf{D}^2\xi=\mathbf{M}\xi^2-[\mathbf{M}\xi]^2=\frac{a^2+a\cdot b+b^2}{3}-\frac{(a+b)^2}{4}=\frac{a^2-2ab+b^2}{12}=\frac{(b-a)^2}{12}\quad .$$

8.1.7 Példa *Exponenciális eloszlás*

$$\mathbf{M}\xi=\int_{-\infty}^{+\infty} x\cdot f_\xi(x)\,dx=\int_0^{\infty} x\cdot\lambda e^{-\lambda x}\,dx=\left[-xe^{-\lambda x}\right]_0^{\infty}+\int_0^{\infty} e^{-\lambda x}\,dx=0+\left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^{\infty}=\frac{1}{\lambda}\quad .$$

$$\mathbf{M}\xi^2=\int_{-\infty}^{+\infty} x^2\cdot f_\xi(x)\,dx=\int_0^{\infty} x^2\cdot\lambda e^{-\lambda x}\,dx=\left[-x^2e^{-\lambda x}\right]_0^{\infty}+\int_0^{\infty} 2xe^{-\lambda x}\,dx=0+\frac{2}{\lambda}\int_0^{\infty} x\lambda e^{-\lambda x}\,dx=$$

$=\frac{2}{\lambda}\cdot\frac{1}{\lambda}=\frac{2}{\lambda^2}$, így $\mathbf{D}^2\xi=\mathbf{M}\xi^2-[\mathbf{M}\xi]^2=\frac{2}{\lambda^2}-\left(\frac{1}{\lambda}\right)^2=\frac{1}{\lambda^2}$.

8.1.8 Normális eloszlás

a.) Standard normális eloszlás

$$\mathbf{M}\xi=\int_{-\infty}^{+\infty} x\cdot f_\xi(x)\,dx=\int_{-\infty}^{\infty} x\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx=\left[-\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right]_{-\infty}^{+\infty}=0\quad .$$

$$\mathbf{M}\xi^2=\int_{-\infty}^{+\infty} x^2\cdot f_\xi(x)\,dx=\int_{-\infty}^{\infty} x^2\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx=\int_{-\infty}^{\infty} x\cdot x\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx=$$

$=\left[x\cdot(-\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}})\right]_{-\infty}^{+\infty}+\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx=0+1=1$. Így $\mathbf{D}^2\xi=\mathbf{M}\xi^2-[\mathbf{M}\xi]^2=1-0=1$.

b.) Az általános eset, $\xi\in N(\mu,\sigma)$.

$$\mathbf{M}\xi=\int_{-\infty}^{+\infty} x\cdot f_\xi(x)\,dx=\int_{-\infty}^{\infty} x\cdot\varphi_{\mu,\sigma}(x)\,dx=\int_{-\infty}^{\infty} x\cdot\frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma})\,dx=\int_{-\infty}^{\infty}(\sigma y+\mu)\cdot\varphi(y)\,dy=$$

$$.\sigma\int_{-\infty}^{+\infty} y\varphi(y)\,dy+\mu\cdot\int_{-\infty}^{+\infty}\varphi(y)\,dy=\sigma\cdot 0+\mu\cdot 1=\mu\quad .$$

$$\mathbf{M}\xi^2=\int_{-\infty}^{+\infty} x^2\cdot f_\xi(x)\,dx=\int_{-\infty}^{\infty} x^2\cdot\varphi_{\mu,\sigma}(x)\,dx=\int_{-\infty}^{\infty} x^2\cdot\frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma})\,dx=$$

$$=\int_{-\infty}^{\infty}(\sigma y+\mu)^2\cdot\varphi(y)\,dy=\sigma^2\int_{-\infty}^{+\infty} y^2\varphi(y)\,dy+2\mu\cdot\sigma\int_{-\infty}^{+\infty} y\varphi(y)\,dy+\mu^2\int_{-\infty}^{+\infty}\varphi(y)\,dy=$$

$=\sigma^2\cdot 1+2\mu\cdot\sigma\cdot 0+\mu^2\cdot 1=\sigma^2+\mu^2$ .Innen $\mathbf{D}^2\xi=\mathbf{M}\xi^2-[\mathbf{M}\xi]^2=\sigma^2+\mu^2-\mu^2=\sigma^2$.

Tehát a normális eloszlás μ paramétere a várható értéket, a σ paraméter pedig a szórást jelenti.

Definíció: Legyenek ξ és η valószínűségi. Tegyük fel, hogy létezik a szórásnégyzetük. Akkor a ξ és η *kovarianciáján* a $\zeta = (\xi - \mathbf{M}\xi)\cdot(\eta - \mathbf{M}\eta)$ valószínűségi változó várható értékét értjük. Jelölés: $\operatorname{cov}(\xi,\eta) = \mathbf{M}\left[(\xi - \mathbf{M}\xi)\cdot(\eta - \mathbf{M}\eta)\right]$.

Megjegyzés: $\operatorname{cov}(\xi,\xi) = \mathbf{D}^2\xi$.

Definíció: Egy ξ valószínűségi változó standardizáltján a $\tilde{\xi} = \dfrac{\xi - M\xi}{D\xi}$ valószínűségi változót értjük.

Definíció: A ξ és η valószínűségi változók *korrelációs együtthatóján* standardizáltjaik kovarianciáját értjük. Jelölés: $R(\xi,\eta) = \operatorname{cov}(\tilde{\xi},\tilde{\eta}) = \dfrac{\operatorname{cov}(\xi,\eta)}{D\xi\cdot D\eta}$.

Megjegyzés: $M\tilde{\xi} = 0$, $D^2\tilde{\xi} = 1$.

Tétel: $\operatorname{cov}(\xi,\eta) = \mathbf{M}(\xi\cdot\eta) - (\mathbf{M}\xi)\cdot(\mathbf{M}\eta)$.

Bizonyítás:
$$\operatorname{cov}(\xi,\eta) = \mathbf{M}((\xi - \mathbf{M}\xi)\cdot(\eta - \mathbf{M}\eta)) = \mathbf{M}(\xi\cdot\eta - \xi\cdot\mathbf{M}\eta - \eta\cdot\mathbf{M}\xi + (\mathbf{M}\xi)\cdot(\mathbf{M}\eta)) =$$
$$= \mathbf{M}(\xi\cdot\eta) - (\mathbf{M}\xi)\cdot(\mathbf{M}\eta) - (\mathbf{M}\eta)\cdot(\mathbf{M}\xi) + (\mathbf{M}\xi)\cdot(\mathbf{M}\eta) = \mathbf{M}(\xi\cdot\eta) - (\mathbf{M}\xi)\cdot(\mathbf{M}\eta).$$

Tétel: Ha ξ és η függetlenek, akkor $\operatorname{cov}(\xi,\eta) = 0$ és $\mathbf{R}(\xi,\eta) = 0$. A tétel megfordítása általában nem igaz.

A megfordításra ellenpélda:
Legyenek a ξ és η diszkrét valószínűségi változók, {-1,0,1} értékkészletekkel. Az együttes eloszlásukat az alábbi táblázatban láthatjuk:

$\eta \backslash \xi$	-1	0	+1	η perem
-1	0	0,25	0	0,25
0	0,25	0	0,25	0,5
+1	0	0,25	0	0,25

ξ perem	0,25	0,5	0,25	1

$$\mathbf{M}\xi = \mathbf{M}\eta = \frac{1}{4}\cdot(-1) + \frac{1}{2}\cdot 0 + \frac{1}{4}\cdot 1 = 0\ ,\ \mathbf{M}(\xi\cdot\eta) = (-1)\cdot(-1)\cdot 0 + 1\cdot 1\cdot 0 = 0$$

$\Rightarrow$ $\operatorname{cov}(\xi,\eta) = \mathbf{M}(\xi\cdot\eta) - (\mathbf{M}\xi)\cdot(\mathbf{M}\eta) = 0.$ A ξ és η nem függetlenek, mert pl. $P(\xi = 0, \eta = 1) = \frac{1}{4} \neq P(\xi = 0)\cdot P(\eta = 1) = \frac{1}{2}\cdot\frac{1}{4}$.

Definíció: A ξ és η valószínűségi változók *korrelálatlanok* , ha
$\operatorname{cov}(\xi,\eta) = \mathbf{M}(\xi\cdot\eta) - (\mathbf{M}\xi)\cdot(\mathbf{M}\eta) = 0.$

A korrelálatlanság a függetlenség szükséges, de nem feltétlenül elégséges feltétele.

Diszkrét esetben a kovariancia számítása:

$$\operatorname{cov}(\xi,\eta) = \sum_{\forall i}\sum_{\forall j} x_i y_j P(\xi = x_i, \eta = y_j) - (\sum_{\forall i} x_i P(\xi = x_i))\cdot(\sum_{\forall j} y_j P(\eta = y_j)).$$

Folytonos esetben a kovariancia számítása:

$$\operatorname{cov}(\xi,\eta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy\cdot f_{\xi,\eta}(x,y)\,dx\,dy - \left(\int_{-\infty}^{+\infty} x\cdot f_\xi(x)\,dx\right)\cdot\left(\int_{-\infty}^{+\infty} y\cdot f_\eta(y)\,dy\right) .$$

Tétel: Ha a ξ és η valószínűségi változók szórásnégyzetei léteznek, úgy

$$\mathbf{D}^2(\xi\pm\eta) = \mathbf{D}^2\xi + \mathbf{D}^2\eta \pm 2\operatorname{cov}(\xi,\eta) .$$

Bizonyítás:

$$\mathbf{D}^2(\xi\pm\eta) = \mathbf{M}\left[(\xi\pm\eta)^2\right] - \left[\mathbf{M}(\xi\pm\eta)\right]^2 = \mathbf{M}\left[\xi^2 \pm 2\xi\eta + \eta^2\right] - \left[(\mathbf{M}\xi)^2 \pm 2(\mathbf{M}\xi)(\mathbf{M}\eta) + (\mathbf{M}\eta)^2\right] =$$

$$= M\xi^2 \pm 2M\xi\eta + M\eta^2 - (M\xi)^2 \mp 2(M\xi)(M\eta) - (M\eta)^2 = D^2\xi + D^2\eta \pm 2\operatorname{cov}(\xi,\eta) .$$

Tétel:

$$\mathbf{D}^2\left(\sum_{\mathbf{i}=1}^{\mathbf{p}} \xi_\mathbf{i}\right) = \sum_{\mathbf{i}=1}^{\mathbf{p}} \mathbf{D}^2\xi_\mathbf{i} + 2\sum_{\mathbf{i}<\mathbf{j}} \operatorname{cov}(\xi_\mathbf{i},\xi_\mathbf{j}) .$$

Bizonyítás:

p=2 -re éppen az előző tételt kapjuk. Tegyük fel, hogy az állítás igaz valamely p≥2 -re.

$$\mathbf{D}^2\left(\sum_{\mathbf{i}=1}^{\mathbf{p}+1} \xi_\mathbf{i}\right) = \mathbf{D}^2\left(\sum_{\mathbf{i}=1}^{\mathbf{p}} \xi_\mathbf{i}\right) + \mathbf{D}^2\xi_{\mathbf{p}+1} + 2\operatorname{cov}\left(\sum_{\mathbf{i}=1}^{\mathbf{p}} \xi_\mathbf{i}, \xi_{\mathbf{p}+1}\right) =$$

$$= \sum_{i=1}^{p+1} D^2\xi_i + 2\sum_{\substack{i<j\\ i,j=1,2,\dots,p}} \operatorname{cov}(\xi_i,\xi_j) + 2\sum_{i=1}^{p} \operatorname{cov}(\xi_i,\xi_{p+1}) \quad \Rightarrow \text{ állítás} .$$

Tétel: Ha a ξ és η valószínűségi változók szórásnégyzetei léteznek, úgy

$$-1 \le \mathbf{R}(\xi,\eta) \le 1 .$$

Bizonyítás:

Legyenek $\tilde{\xi} = \dfrac{\xi - \mathbf{M}\xi}{\mathbf{D}\xi}$ és $\tilde{\eta} = \dfrac{\eta - \mathbf{M}\eta}{\mathbf{D}\eta}$ a standardizált valószínűségi változók.

$$\operatorname{cov}(\tilde{\xi}, \tilde{\eta}) = \mathbf{M}\left(\frac{\xi - \mathbf{M}\xi}{\mathbf{D}\xi} \cdot \frac{\eta - \mathbf{M}\eta}{\mathbf{D}\eta} \right) = \frac{\operatorname{cov}(\xi, \eta)}{\mathbf{D}\xi \cdot \mathbf{D}\eta} = \mathrm{R}(\xi, \eta) \ .$$

$$0 \le \mathbf{D}^2(\tilde{\xi} \pm \tilde{\eta}) = \mathbf{D}^2\tilde{\xi} + \mathbf{D}^2\tilde{\eta} \pm 2\operatorname{cov}(\tilde{\xi}, \tilde{\eta}) = 1 + 1 \pm 2\mathbf{R}(\xi, \eta) \ \Rightarrow -1 \le \mathbf{R}(\xi, \eta) \le 1 \ .$$

Következmény: $\left|\operatorname{cov}(\xi, \eta)\right| \le \mathbf{D}\xi \cdot \mathbf{D}\eta$.

Tétel: Ha a ξ és η valószínűségi változók szórásnégyzetei léteznek, úgy

$$\mathbf{R}(\xi, \eta) = \pm 1 \quad \Leftrightarrow \quad \exists\, a, b \in \mathbb{R} \quad : \mathrm{P}(\xi = a \cdot \eta + b) = 1 \ .$$

Bizonyítás:

Legyenek $\tilde{\xi} = \dfrac{\xi - \mathrm{M}\xi}{\mathrm{D}\xi}$ és $\tilde{\eta} = \dfrac{\eta - \mathrm{M}\eta}{\mathrm{D}\eta}$ a standardizált valószínűségi változók.

$\exists\, a, b \in \mathbb{R} \quad : \mathrm{P}(\xi = a \cdot \eta + b) = 1 \Leftrightarrow \mathrm{P}(\tilde{\xi} = \pm\tilde{\eta}) = 1 \Leftrightarrow$

$\Leftrightarrow \ 0 = \mathbf{D}^2(\tilde{\xi} \mp \tilde{\eta}) = 2(1 \mp \mathbf{R}(\xi, \eta)) \Leftrightarrow \mathbf{R}(\xi, \eta) = \pm 1$ ráadásul $\mathbf{R}(\xi, \eta) = \operatorname{sign}(a)$.

(A bizonyításkor felhasználtuk, hogy csak az egy valószínűséggel konstans valószínűségi változónak lehet 0 a szórása.)

Ellenőrző kérdések és gyakorló feladatok

1. Mi a várható érték definíciója?
2. Milyen tulajdonságai vannak a várható értéknek és a szórásnégyzetnek?
3. Mi a kovariancia és a korrelációs együttható?
4. Mi a kapcsolat a függetlenség és a korrelálatlanság között?
5. Mikor 1 a korrelációs együttható abszolút értéke?
6. Melyik állítás igaz, melyik hamis?
   a. Minden valószínűségi változónak van várható értéke.
   b. Minden valószínűségi változónak van szórása.
   c. A binomiális valószínűségi változónak van várható értéke és szórása is.
   d. A szórásnégyzet lineáris.
   e. A várható érték lineáris.
   f. Két valószínűségi változó szorzatának várható értéke egyenlő a várható értékek szorzatával.
   g. Két független valószínűségi változó szorzatának várható értéke egyenlő a várható értékek szorzatával.
   h. Két független valószínűségi változó szorzatának szórásnégyzete egyenlő a szórásnégyzetek szorzatával.
   i. Két független valószínűségi változó összegének szórásnégyzete egyenlő a szórásnégyzetek összegével.
   j. Két független valószínűségi változó összegének várható értéke egyenlő a várható értékek összegével.
   k. Egy valószínűségi változó önmagával vett kovarianciája éppen a szórásnégyzet.
   l. Egy valószínűségi változó önmagával vett korrelációs együtthatója mindig 1.

m. Két független valószínűségi változó korrelációja 0.
n. Ha két valószínűségi változó kovarianciája 0, akkor függetlenek.
o. Egy valószínűségi változó négyzetének várható értéke nem kisebb mint a várható értékének négyzete.
p. Egy valószínűségi változó négyzetének várható értéke egyenlő a várható értékének négyzetével.
q. A standardizált valószínűségi változó szórása 1.
r. A standardizált valószínűségi változó várható értéke 1.
s. A pozitív konstans szórása pozitív.
t. Független valószínűségi változók különbségének szórásnégyzete a szórásnégyzetek különbsége.

7. Legyen ξ Poisson eloszlású $\lambda>0$ paraméterrel, $\eta=2\xi+1$. Adjuk meg η várható értékét és szórását!
8. Legyen ξ n és p paraméterű binomiális eloszlású valószínűségi változó, $\eta = \frac{1}{1+\xi}$. Adjuk meg η várható értékét és szórását!
9. Legyenek ξ és η azonos eloszlású valószínűségi változók. Igaz-e, hogy $\mathbf{M}\frac{\xi}{\xi+\eta} = \mathbf{M}\frac{\eta}{\xi+\eta}$?
10. Ha a ξ sűrűségfüggvénye $f_\xi(x) = \frac{1}{\pi(1+x^2)}$, $(x \in \mathrm{IR})$, akkor létezik-e a várható értéke?
11. Egy szabályos kockával dobunk ismételten. ξ az első dobás, η a második dobás eredménye. Számoljuk ki $R(\xi,\xi+\eta)$-t!
12. Legyenek ξ és η n=1 és p=0,5 paraméterű független binomiális eloszlású valószínűségű változók. Mutassuk meg, hogy $\xi+\eta$ és $|\xi-\eta|$ bár korrelálatlanok, de nem függetlenek!
13. Legyen $\xi\in U(0,2)$, azaz a (0,2) intervallumon egyenletes eloszlású valószínűségi változó. $\eta=\cos\xi$ és $\zeta=\sin\xi$. Határozzuk meg $cov(\eta,\zeta)$-t! Függetlenek-e η és ζ?
14. Legyen $\xi\in N(\mu,\sigma)$, azaz μ,σ paraméterű normális eloszlású valószínűségi változó! Adjunk képletet $\mathbf{M}\xi^n$-re!

9. A nagy számok törvényei és a centrális határeloszlás tételek

A valószínűségi változók várható értékének és szórásnégyzetének fontos tulajdonságát fogalmazza meg a Markov és a Csebisev egyenlőtlenség.

Tétel: *(A Markov egyenlőtlenség)*
Legyen $\xi \geq 0$ olyan valószínűségi változó, melynek létezik a várható értéke: $\mathbf{M}\xi \geq 0$.
Akkor $\forall \varepsilon > 0$ esetén $P(\xi > \varepsilon) \leq \dfrac{\mathbf{M}\xi}{\varepsilon}$.

Bizonyítás:
Diszkrét valószínűségi változó esetében:
$\mathbf{M}\xi = \sum_{\forall i} x_i P(\xi = x_i) \geq \sum_{x_i \geq \varepsilon} x_i P(\xi = x_i) \geq \varepsilon \cdot \sum_{x_i \geq \varepsilon} P(\xi = x_i) = \varepsilon \cdot P(\xi \geq \varepsilon) \Rightarrow$állítás.
Folytonos valószínűségi változó esetében:
$M\xi = \int_0^\infty x \cdot f_\xi(x)\,dx \geq \int_\varepsilon^\infty x \cdot f_\xi(x)\,dx \geq \varepsilon \cdot \int_\varepsilon^\infty f_\xi(x)\,dx = \varepsilon \cdot \left(1 - F_\xi(\varepsilon)\right) = \varepsilon \cdot P(\xi \geq \varepsilon) \Rightarrow$ állítás.
Megjegyzés:
1.) Az egyenlőtlenségben most akkor kapunk nem semmitmondó állítást, ha $\varepsilon \geq \mathbf{M}\xi$. Különben a Markov egyenlőtlenség csak annyit jelentene, hogy egy valószínűség nem nagyobb, mint egy 1-nél nagyobb szám...Tehát most $\varepsilon > 0$ nem azt sugallja - mint általában a matematikai tételekben-, hogy ε tetszőlegesen kicsiny pozitív szám, hanem éppen ellenkezőleg, most ε nagy.
2.) A Markov egyenlőtlenséget átfogalmazhatjuk a következő módon, ha végrehajtjuk az $\varepsilon = \delta \cdot \mathbf{M}\xi$ helyettesítést: $\forall \delta > 0$ esetén $P(\xi > \delta \cdot \mathbf{M}\xi) \leq \dfrac{1}{\delta}$.
Innen viszont az olvasható le, hogy ξ kicsi valószínűséggel vehet csak fel a saját várható értékénél sokkal nagyobb értékeket, vagyis ξ hajlamos a várható értéke közelében értéket felvenni. Pl. annak valószínűsége, hogy egy nemnegatív valószínűségi változó a várható értékének ötszörösénél nagyobb értéket felvegyen, 20%-nál kisebb.

Tétel: *(A Csebisev-egyenlőtlenség)*
Legyen ξ olyan valószínűségi változó, amelynek véges a szórásnégyzete :$\mathbf{D}^2\xi < \infty$.
Akkor minden $\varepsilon > 0$ esetén $P(|\xi - \mathbf{M}\xi| \geq \varepsilon) \leq \dfrac{\mathbf{D}^2\xi}{\varepsilon^2}$.

Bizonyítás: Alkalmazzuk a Markov egyenlőtlenséget a $\xi = \left(\xi - M\xi\right)^2$, $\varepsilon = \varepsilon^2$ helyettesítéssel:
$$P((\xi - \mathbf{M}\xi)^2 \geq \varepsilon^2) = P(|\xi - \mathbf{M}\xi| \geq \varepsilon) \leq \frac{\mathbf{M}(\xi - \mathbf{M}\xi)^2}{\varepsilon^2} = \frac{\mathbf{D}^2\xi}{\varepsilon^2}.$$
Megjegyzés:
1.) Az ε-ról ugyanaz elmondható, mint a Markov egyenlőtlenség esetén: $\varepsilon \geq \mathbf{D}\xi$ esetben lesz csak nem-triviális az egyenlőtlenség.
2.) A Csebisev egyenlőtlenség is átfogalmazható, ha $\varepsilon = \delta \cdot \mathbf{D}\xi$:

Minden $\delta > 0$ esetén $P(|\xi - \mathbf{M}\xi| \geq \delta \cdot \mathbf{D}\xi) \leq \frac{1}{\delta^2}$. Vagyis a valószínűségi változó a várható értéke körül ingadozik, és annál kisebb mértékben, minél kisebb a szórása. Pl. egy valószínűségi változó nem térhet el jobban a várható értékétől , mint a szórása háromszorosa, csak legfeljebb $\frac{1}{9} \approx 0.11$ valószínűséggel.

Feladat Egy célpontra 200 lövést adnak le. A találat valószínűsége minden lövésnél 0,4. Milyen határok közé fog esni 90%-os valószínűséggel a találatok száma?

Megoldás: Jelöljük ξ-vel a találatok számát! A lövéssorozat felfogható egy n=200 hosszúságú Bernoulli kísérletsorozatnak, ahol a megfigyelt esemény a célpont eltalálása. Ezért ξ binomiális eloszlású n=200 és p=0,4 paraméterekkel. Így

$\mathbf{M}\xi = np = 200 \cdot 0{,}4 = 80$, és $\mathbf{D}^2\xi = npq = 200 \cdot 0{,}4 \cdot 0{,}6 = 48$. A Csebisev egyenlőtlenséget alkalmazzuk erre az esetre ε=$\sqrt{10}$ választással:

$P\left(|\xi - \mathbf{M}\xi| > \sqrt{10}\mathbf{D}\xi\right) = P\left(|\xi - 80| > \sqrt{480}\right) \leq 0{,}1$, ahonnan

$P\left(|\xi - 80| \leq \sqrt{480}\right) = P\left(80 - \sqrt{480} \leq \xi \leq 80 + \sqrt{480}\right) = P\left(58 \leq \xi \leq 102\right) \geq 0{,}9$ adódik, azaz a lövések 58 és 101 közé fognak esni legalább 90%-os valószínűséggel.

Feladat Automata minőségvizsgáló n=100 000 elemű mintát ellenőriz le egy gyártósoron előállított számítógépes alkatrésztömegből. A vizsgálat után milyen valószínűséggel állíthatjuk, hogy a mintából meghatározott selejtarány a készlet elméleti p selejtvalószínűségétől legfeljebb 0,01-dal tér el?

Megoldás: ξ most a selejtes termékek számát jelölje a mintában! Ekkor a selejtarány a mintában $\frac{\xi}{10^5}$ lesz. Nyílván $\xi \in B(100\ 000, p)$, ahol a p ismeretlen.

$\mathbf{M}\xi = np = 10^5 p$, és $\mathbf{D}^2\xi = npq = 10^5 \cdot pq$. A Csebisev egyenlőséget most ε=1000-rel alkalmazzuk: $P\left(|\xi - 10^5 p| \leq 1000\right) = P\left(\left|\frac{\xi}{10^5} - p\right| \leq 0{,}01\right) \geq 1 - \frac{10^5 pq}{10^6} \geq \frac{39}{40}$≈0,975.

A levezetésben felhasználtuk, hogy $pq = p - p^2 \leq 0{,}25$.

A nagy számok törvényei azt a megfigyelést támasztják alá elméletileg is, hogy egy valószínűségi változót sokszor megfigyelve, az átlagérték mindig közel van az elméleti várható értékhez. Az is igaz, hogy a megfigyelések növekedtével az eltérés csökken, azaz az átlagértékek konvergálnak is a várható értékhez. A köznapi életben a tételt úgy fogalmazzák meg, hogy a véletlen jelenségek is kiszámíthatóak hosszútávon.

Tétel: *(A nagy számok tételének Csebisev-féle gyenge alakja)*
Legyenek a $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ valószínűségi változók páronként függetlenek és azonos eloszlásúak (azonos eloszlásfüggvénnyel rendelkezők). Létezzék a közös $\mu = \mathbf{M}\xi_i$ várható értékük és a közös $\sigma^2 = \mathbf{D}^2\xi_i$ szórásnégyzetük.
Akkor a $\zeta_n = \dfrac{\xi_1 + \xi_2 + \cdots + \xi_n}{n}$ valószínűségi változó sorozatra $\forall \varepsilon > 0$ esetén $P(|\zeta_n - \mu| \geq \varepsilon) \to 0 \quad (n \to \infty)$ teljesül.

Bizonyítás: $\mathrm{M}\zeta_n = \mathrm{M}\left(\dfrac{\xi_1 + \xi_2 + \cdots + \xi_n}{n} \right) = \dfrac{1}{n} \sum_{i=1}^{n} \mathrm{M}\xi_i = \dfrac{1}{n} \cdot n \cdot \mu = \mu$.

A páronkénti függetlenség miatt: $\mathbf{D}^2\zeta_n = \mathbf{D}^2\left(\dfrac{1}{n} \sum_{i=1}^{n} \xi_i \right) = \dfrac{1}{n^2} \sum_{i=1}^{n} \mathbf{D}^2\xi_i = \dfrac{1}{n^2} \cdot n \cdot \sigma^2 = \dfrac{\sigma^2}{n}$.

Látható tehát, hogy ζ_n minden indexre teljesíti a Csebisev egyenlőtlenség feltételét, így:

$P(|\zeta_n - \mathbf{M}\zeta_n| \geq \varepsilon) = P(|\zeta_n - \mu| \geq \varepsilon) \leq \dfrac{\mathbf{D}^2\zeta_n}{\varepsilon^2} = \dfrac{\sigma^2}{n \cdot \varepsilon^2} \to 0 \quad (n \to \infty)$, ami már igazolja az állítást.

Megjegyzés: A tétel azt állítja, hogy azon valószínűségek számsorozata, hogy az átlag az elméleti várható értéktől akármilyen kis ε-nál is jobban eltérjen, nullához konvergál.

Tétel: *(A nagy számok tételének Bernoulli-féle gyenge alakja)*

Legyen $(\Omega, \Im, P)$ Kolmogorov- féle valószínűségi mező, $A \in \Im$ egy pozitív valószínűségű esemény: $p = P(A) > 0$. Hajtsunk végre egy végtelen Bernoulli-féle kísérletsorozatot, vagyis figyeljük meg az A bekövetkezéseit! Legyen $\xi_i = \begin{cases} 1 & , \omega \in A \\ 0 & , \omega \notin A \end{cases}$, vagyis az i-edik végrehajtáskor az esemény karakterisztikus valószínűségi változója. ξ_i-k teljesen függetlenek és azonos eloszlásúak:
$p_0 = P(\xi_i = 0) = P(\overline{A}) = q$, $p_1 = P(\xi_i = 1) = P(A) = p$, $\mathbf{M}\xi_i = p$, $\mathbf{D}^2\xi_i = pq$.
Legyen $\zeta_n = \dfrac{\xi_1 + \xi_2 + \cdots + \xi_n}{n} = r_n(A)$ a relatív gyakoriság.
Akkor $\forall \varepsilon > 0$ esetén $P(|r_n(A) - p| \geq \varepsilon) \to 0 \quad (n \to \infty)$.

Bizonyítás: A tétel feltételei speciális esetben a nagy számok Csebisev-féle alakjának felelnek meg. Ekkor a Csebisev egyenlőtlenségnek a

$P(|\zeta_n - \mathbf{M}\zeta_n| \geq \varepsilon) = P(|r_n(A) - p| \geq \varepsilon) \leq \dfrac{\mathbf{D}^2\zeta_n}{\varepsilon^2} = \dfrac{pq}{n \cdot \varepsilon^2} \leq \dfrac{1}{4 \cdot n \cdot \varepsilon^2} \to 0 \quad (n \to \infty)$

felel meg, mert $p \cdot q \leq \dfrac{1}{4}$ mindig teljesül.

Megjegyzés: A tétel azt mondja ki, hogy a relatív gyakoriság jól közelíti az esemény elméleti valószínűségét, ahogyan azt már a valószínűség axiómái után tett megjegyzésünkben előre jeleztük.

Tétel: *(A centrális határeloszlás tétel)*
Legyenek a $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ valószínűségi változók páronként függetlenek és azonos eloszlásúak (azonos eloszlásfüggvénnyel rendelkezők) az $(\Omega,\Im,P)$ Kolmogorov-féle valószínűségi mezőn. Létezzék a közös $\mu = \mathbf{M}\xi_i$ várható értékük és a közös $\sigma^2 = \mathbf{D}^2\xi_i$ szórásnégyzetük.
Akkor a $\zeta_n = \dfrac{\xi_1 + \xi_2 + \cdots + \xi_n - n\cdot\mu}{\sqrt{n}\cdot\sigma}$ valószínűségi változó sorozatra teljesül, hogy az eloszlásfüggvényeik függvénysorozata minden pontban a standard normális eloszlás eloszlásfüggvényéhez konvergálnak, azaz
$$F_{\zeta_n}(x) = P(\zeta_n < x) = P\left(\frac{\xi_1 + \xi_2 + \cdots + \xi_n - n\cdot\mu}{\sqrt{n}\cdot\sigma} < x\right) \to \Phi(x) \ \ (n \to \infty) \ \ \forall x \in \mathrm{lR}.$$

Megjegyzés: A tétel rámutat a normális eloszlásnak az elméletben játszott fontos szerepének okára: tetszőleges eloszlású valószínűségi változók átlaga normális eloszlást követ. Tehát, ha egy véletlen jelenséget sok egyenként nem jelentős, független hatás összegeként kapunk, akkor az jól közelíthető a normális eloszlással. Tipikusan ilyenek a mérésekből származó adatok: a Duna közepes vízállása, a napi középhőmérséklet stb. Az elektromos elosztó központban is normális eloszlásúnak tekinthető a lakossági fogyasztás, hiszen nagyon sok kisfogyasztó eredőjeként áll elő. És bár lehet, hogy az egyes fogyasztók külön-külön nem a normális elosztás szerint fogyasztanak, de az átlagos fogyasztást a nagy számok törvénye értelmében biztosan tekinthetjük normálisnak modelljeinkben.

Tétel: *(A Moivre-Laplace tétel ,1733.)*
Legyen $(\Omega,\Im,P)$ Kolmogorov-féle valószínűségi mező, $A \in \Im$ egy pozitív valószínűségű esemény: $p = P(A) > 0$. Hajtsunk végre egy végtelen Bernoulli-féle kísérletsorozatot, vagyis figyeljük meg az A bekövetkezéseit az 1,2,...,n,...-edik kísérletnél! Legyen $\xi_i = \begin{cases} 1 & ,\omega \in A \\ 0 & ,\omega \notin A \end{cases}$, vagyis az i-edik végrehajtáskor az esemény karakterisztikus valószínűségi változója. ξ_i-k teljesen függetlenek és azonos eloszlásúak:
$p_0 = P(\xi_i = 0) = P(\overline{A}) = q$, $p_1 = P(\xi_i = 1) = P(A) = p$, $\mathbf{M}\xi_i = p$, $\mathbf{D}^2\xi_i = pq$.
Akkor a $\zeta_n = \dfrac{\xi_1 + \xi_2 + \cdots + \xi_n - n\cdot p}{\sqrt{n\cdot p\cdot(1-p)}}$ valószínűségi változó sorozatra teljesül, hogy az eloszlásfüggvényeik függvénysorozata minden pontban a standard normális eloszlás eloszlásfüggvényéhez konvergálnak, azaz
$F_{\zeta_n}(x) = P(\zeta_n < x) \to \Phi(x) \ \ (n \to \infty) \ \ \forall x \in \mathrm{lR}$.

Bizonyítás: A Moivre-Laplace tétel a centrális határeloszlástétel azon speciális este, amikor a $\xi_i \in \chi(A)$, azaz karakterisztikus eloszlásúak.
Ráadásul $F_{\zeta_n}(x) = P(\zeta_n < x) = P(n\cdot S_n < \sqrt{n\cdot p\cdot q}\,.\cdot x + n\cdot p)$, mivel $r_n(A) = S_n = \dfrac{\xi_1 + \xi_2 + \cdots + \xi_n}{n}$ a relatív gyakoriság és $n\cdot S_n \in B(n,p)$.

Így $P(n\cdot S_n = k) = \binom{n}{k} p^k \cdot q^{n-k}$, amiből az eloszlásfüggvényre:

$$P(n\cdot S_n < \sqrt{npq}\cdot x + np) = \sum_{k<\sqrt{npq}\cdot x+np} \binom{n}{k} p^k \cdot q^{n-k} = \sum_{\frac{k-np}{\sqrt{npq\cdot}}<x} \binom{n}{k} p^k \cdot q^{n-k}.$$

Tehát a tétel azt állítja, hogy a binomiális eloszlás standardizáltja határeloszlásban standard normális eloszlás lesz, azaz $\lim\limits_{n\to\infty} \sum\limits_{\frac{k-np}{\sqrt{npq}}<x} \binom{n}{k} p^k \cdot q^{n-k} = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt \quad (\forall x \in \mathrm{lR})$.

Másképpen fogalmazva, hosszú Bernuolli kísérletsorozatok esetén az esemény gyakorisága közelítőleg normális eloszlást fog követni.

Feladat Közelítőleg határozzuk meg a A= $\sum\limits_{k=220}^{260} \binom{500}{k}$ összeget!

Megoldás: Legyen $\xi \in B(500,\, 0{,}5)$! Ekkor a kiszámítandó A összeget felírhatjuk:

$A = 2^{500} \sum\limits_{k=200}^{260} P(\xi = k)$ alakban. A Moivre-Laplace tétel szerint: $\sum\limits_{y\le\frac{k-250}{\sqrt{125}}<x} P(\xi = k) \approx \Phi(x) - \Phi(y)$.

Most úgy kell x-et és y-t megválasztani, hogy 220=250+$\sqrt{125}y$ és 261=250+$\sqrt{125}x$ legyen.

Tehát y=-2,683281573 és x= 0,9838699100999, amivel

$2^{-500} A \approx \Phi(0{,}9839) - \Phi(-2{,}6833) = \Phi(0{,}9839) + \Phi(+2{,}6833) - 1 \approx 0{,}8365 + 0{,}9963 - 1 =$ =0,8328, azaz A≈ 2,726079698256e+150. A Φ függvény értékeit a standard normális eloszlás táblázatából olvastuk ki.(Ld. A függelékben!)

Megjegyzés: Az előbbi összeg kiszámítása még számítógépre írt program segítségével sem triviális a binomiális együtthatókban szereplő nagy faktoriálisok miatt.

Ellenőrző kérdések és gyakorló feladatok

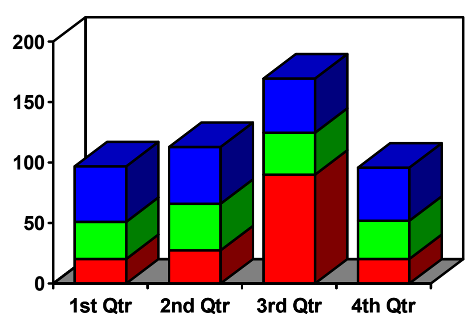
1. Mit állít a Csebisev egyenlőtlenség?
2. Mondja ki a nagy számok törvényének Csebisev féle alakját!
3. Mit állít a nagy számok törvényének Bernoulli féle alakja?
4. Mit állít a centrális határeloszlás tétel?
5. Mondja ki a Moivre-Laplace tételt!
6. Melyik állítás igaz, melyik hamis?
   a. A Markov egyenlőtlenség csak a várható értéknél nagyobb ε valós számok esetén érvényes.
   b. A Csebisev egyenlőtlenség csak diszkrét valószínűségi változókra igaz.
   c. A Moivre-Laplace tétel egy speciális esete a nagy számok törvényének.
   d. A Csebisev egyenlőtlenség a Markov egyenlőtlenség speciális esete.
   e. A centrális határeloszlás tétel azt állítja, hogy a binomiális eloszlás nagy n paraméter esetén közelíthető normális eloszlással.
7. Egy üzemben csavarokat csomagolnak. Egy-egy dobozba átlagosan 5000 csavar kerül. A csavarok számának szórása a tapasztalat szerint 20 darab. Mit mondhatunk annak

valószínűségéről, hogy egy dobozban a csavarok száma 4900 és 5100 közé esik, ha az eloszlást nem ismerjük.

8. Egy szövőgép 500 szállal dolgozik. Annak a valószínűsége, hogy egy szál időegység alatt elszakad 0,008 minden szálra. Határozzuk meg, hogy 0,95 valószínűséggel milyen határok között várható a szálszakadások száma egy időegység alatt?
9. Legyenek $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ független azonos eloszlású valószínűségi változók véges szórással. Bizonyítsuk be, hogy tetszőleges $x \in \mathrm{IR}$ valós szám esetén
$\lim\limits_{n\to\infty} P(\xi_1 + \xi_2 + \cdots \xi_n < x) \in \left\{0, \frac{1}{2}, 1\right\}$, vagyis a határérték csak 0 vagy 0,5 vagy 1 lehet!
10. Legyen ξ standard normális eloszlású valószínűségi változó! A standard normális eloszlás táblázatának használata nélkül bizonyítsa be, hogy ekkor fennáll a
$P(-3 < \xi < 3) \geq 1 - \frac{2}{\sqrt{18\pi}}$ egyenlőtlenség!
11. Ha egy gyár egyforma energiaigényű gépe közül átlagosan 70% működik és 30% vár javításra, vagy éppen javítják, akkor átlagosan 210 gép energiaigényét kell kielégíteni. Mennyi energiát kell biztosítani akkor, ha 99,9%-os biztonsággal szeretnénk elérni azt, hogy minden működőképes gép valóban működni tudjon? (Feltesszük, hogy a gépek meghibásodása egymástól független.)

II. fejezet

MATEMATIKAI STATISZTIKA



1. A matematikai statisztika alapfogalmai

A valószínűségszámítás elméletében az $(\Omega, \Im, P)$ Kolmogorov valószínűségi mezőn fogalmaztuk meg a tételeinket, azaz a P valószínűségi mértéket végig adottnak tételeztük fel. A gyakorlati problémáknál azonban a valószínűség nem ismert, legfeljebb logikus előfeltételezéseink vannak róla. A matematikai statisztika alapfeladata éppen az, hogy a véletlen kísérletre, vagy a véletlen tömegjelenségre vonatkozó megfigyeléssorozat segítségével következtetni tudjunk a jelenséghez tartozó adekvát valószínűségi mértékre, azt megfelelő pontossággal közelíteni tudjuk. Ilyen értelemben a véletlen jelenségek matematikai modellezésénél a matematikai statisztika módszerei megelőzik a valószínűségszámítás módszereit. A matematikai statisztika fogalomköre, módszertana viszont a valószínűségszámítás fogalmaira és módszereire alapul, és ilyen szempontból a matematikai statisztika követi a valószínűségszámítást. A matematikai jelző arra utal, hogy az állításokat matematikai formulákkal fogalmazzuk meg, és az ott szokott egzaktsággal bizonyítunk, legtöbbször a valószínűségszámításnál igazolt tételekre hivatkozva. Ezzel a jelzővel is szeretnénk hangsúlyozni a témakör különbségét a főleg társadalomtudományoknál (közgazdaságtan, szociológia, politológia) alkalmazott statisztikai módszerektől, ahol a rendelkezésekre álló adatokat leíró statisztikákkal jellemeznek valahogy, és a következtetéseket heurisztikusan hozzák meg.
Ugyanúgy, mint a valószínűségszámításnál, a *véletlen kísérlet* ($\mathtt{k}$) alapfogalmából indulunk ki. Azt is feltesszük, hogy ismert az elemi események Ω halmaza és az események $\Im$ halmazrendszere. A P valószínűség pontosan nem ismert, csak azt tudjuk, hogy a $\mathtt{k}$ véletlen kísérlethez tartozó valószínűség eleme egy $\mathtt{P}$ halmaznak. Tehát $\forall\ P \in \mathtt{P}$ esetén Kolmogorov-féle valószínűségi mezőt kapunk. A matematikai statisztika alapfeladata ezen $\mathtt{P}$ halmazból kiválasztani azt a valószínűségi mértéket, amely ténylegesen a kísérlethez tartozik.
Ezért a $\mathtt{k}$ véletlen kísérlethez megfigyeléssorozatot szervezünk, azaz adatokat gyűjtünk.

Definíció: Legyen $(\Omega, \Im)$ egy $\mathtt{k}$ véletlen kísérlethez tartozó eseménytér és eseményalgebra, $\mathtt{P}$ valószínűségi mértékek egy halmaza, ahol $\forall\ P \in \mathtt{P}$ szóbajöhető valószínűség. Az $X = (\xi_1, \xi_2, \dots, \xi_n)^T$ statisztikai megfigyelést *statisztikai mintának* nevezzük, ha ξ_i-k teljesen független, azonos eloszlású valószínűségi változók $\forall\ P \in \mathtt{P}$ esetén $(\Omega, \Im, P)$-n, azaz $\forall\ P \in \mathtt{P}$-re $P(\xi_i < x) = F_P(x) \quad (i = 1, 2, 3, \dots, n)$ és

$$P(\xi_{i_1} < x_{i_1}, \xi_{i_2} < x_{i_2}, \dots, \xi_{i_k} < x_{i_k}) = \prod_{\alpha=1}^{k} F_P(x_{i_\alpha}) \quad (\forall 2 \le k \le n).$$

n a *minta elemszáma*, $F_P(x)$ a *minta eloszlásfüggvénye*, ξ_i az i-edik *mintaelem*. Egy $\omega \in \Omega$ esetén a $\xi_1(\omega) = x_1, \xi_2(\omega) = x_2, \dots, \xi_n(\omega) = x_n$ szám n-es a *minta egy realizációja*.

Megjegyzés: Amikor egy statisztikai módszert alkalmazunk, mindig egy méréssorozat eredménye, azaz a statisztikai minta realizáltja áll a rendelkezésünkre. Ez a szám n-es azonban lehetett volna egy teljesen más szám n-es is, hiszen ha megismételnénk a mintavételezést, egészen biztos, hogy más adatokhoz jutnánk, azaz a minta függ a véletlentől. A módszerek elméletének tárgyalásakor ezért a mintát független, azonos eloszlású valószínűségi változók

sorozatának tekintjük. A függetlenség és az azonos eloszlásból való származás feltételeit a mintavételezés megtervezésekor kell figyelembe venni.
A statisztikai következtetéseinket a statisztikai mintából számolt mennyiségek segítségével hozzuk. A statisztika a minta egy függvénye, tehát maga is valószínűségi változó. A matematikai definíciója a következő:

Definíció: A $T_n : \mathrm{IR}^n \to \mathrm{IR}$ skalár-vektor függvényt statisztikai függvénynek nevezzük, ha adott $X = (\xi_1, \xi_2, \ldots, \xi_n)^T$ statisztikai minta esetén $T_n(\xi_1, \xi_2, \ldots, \xi_n)$ $\forall$ $P \in \mathcal{P}$ esetén valószínűségi változó lesz az $(\Omega, \Im, P)$ valószínűségi mezőn.

A gyakorlati alkalmazásokban az alábbi statisztikák fordulnak legtöbbször elő:

Definíció: A $\overline{\xi}_n \overset{\mathrm{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} \xi_i$ statisztikát a $\xi_1, \xi_2, \ldots, \xi_n$ statisztikai minta *átlag-* vagy *empirikus közép* statisztikájának nevezzük.

Definíció: Az $s_n^2 \overset{\mathrm{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} (\xi_i - \overline{\xi}_n)^2$ statisztikát a $\xi_1, \xi_2, \ldots, \xi_n$ statisztikai minta *empirikus szórásnégyzet* statisztikájának nevezzük. $s_n \overset{\mathrm{def}}{=} +\sqrt{s_n^2}$ az *empirikus szórás* statisztika.

Definíció: Az $s^{*2}_n \overset{\mathrm{def}}{=} \frac{1}{n-1} \cdot \sum_{i=1}^{n} (\xi_i - \overline{\xi}_n)^2$ statisztikát a $\xi_1, \xi_2, \ldots, \xi_n$ statisztikai minta *korrigált empirikus szórásnégyzet* statisztikájának nevezzük. $s^*_n \overset{\mathrm{def}}{=} +\sqrt{s^{*2}_n}$ a *korrigált empirikus szórás* statisztika.

Definíció: Tekintsük azokat az $r_k(x_1, x_2, \ldots, x_n)$ skalár-vektor függvényeket, melyek definíciója:
$x^*_k \overset{\mathrm{def}}{=} r_k(x_1, x_2, \ldots, x_n) = x_j$, ha x_j $k-$adik legnagyobb elem $x_1, x_2, \ldots, x_n$ között. A
$\xi^*_k \overset{\mathrm{def}}{=} r_k(\xi_1, \xi_2, \ldots, \xi_n)$ $(k = 1,2,\ldots, n)$ statisztikák a *rendezett mintaelem statisztikák*.

Megjegyzés: A rendezett mintaelem-statisztikák között $\forall$ $P \in \mathcal{P}$ esetén 1 valószínűséggel fennáll, hogy $\xi^*_1 \le \xi^*_2 \le \cdots \le \xi^*_n$. Speciálisan $\xi^*_1 = \min\{\xi_1, \xi_2, \ldots, \xi_n\}$,
és $\xi^*_n = \max\{\xi_1, \xi_2, \ldots, \xi_n\}$.

Definíció: Az $F_n(x) \overset{def}{=} \begin{cases} 0 & , x \le \xi^*_1 \\ \dfrac{k}{n} & , \xi^*_k < x \le \xi^*_{k+1} \quad (k = 1,2,\dots,n-1) \\ 1 & , x > \xi^*_n \end{cases}$ véletlen függvényt a $\xi_1, \xi_2, \dots, \xi_n$ statisztikai minta *empirikus eloszlásfüggvényének* nevezzük.

Megjegyzés: Az empirikus eloszlásfüggvény minden rögzített $x \in \mathbb{R}$ esetén statisztika, azaz valószínűségi változó! $F_n(x)$ minden realizációja diszkrét eloszlásfüggvény, azaz olyan lépcsős függvény, melynek ugráshelyei a véletlen mintától függenek, és az ugrások magassága
1 valószínűséggel $\dfrac{1}{n}$.

Definíció: A $T_n = \xi^*_n - \xi^*_1$ statisztika a $\xi_1, \xi_2, \dots, \xi_n$ statisztikai minta *terjedelme.*

Definíció: Legyen most az $X = \left((\xi_1, \eta_1)^T, (\xi_2, \eta_2)^T, \dots, (\xi_n, \eta_n)^T \right)^T$ statisztikai megfigyelés kétdimenziós statisztikai minta, ahol az $(\xi_i, \eta_i)^T$ párok azonos eloszlású, teljesen független valószínűségi vektor változók. Akkor a $c_n = \dfrac{1}{n-1} \cdot \sum_{i=1}^{n} \left(\xi_i - \overline{\xi}_n \right) \cdot \left(\eta_i - \overline{\eta}_i \right)$ statisztika a $(\xi_1, \eta_1)^T, (\xi_2, \eta_2)^T, \dots, (\xi_n, \eta_n)^T$ minta *empirikus kovarianciája,*

$$\rho_n = \frac{\sum_{i=1}^{n} \left(\xi_i - \overline{\xi}_n \right) \cdot \left(\eta_i - \overline{\eta}_n \right)}{\sqrt{\sum_{i=1}^{n} \left(\xi_i - \overline{\xi}_n \right)^2 \cdot \sum_{i=1}^{n} \left(\eta_i - \overline{\eta}_n \right)^2}} = \frac{c_n}{s^*_\xi \cdot s^*_\eta}$$ pedig az empirikus korrelációs együtthatója,

ahol pl. $s^*_\xi = \sqrt{\dfrac{1}{n-1} \sum_{i=1}^{n} \left(\xi_i - \overline{\xi}_n \right)^2}$ a $\xi_1, \xi_2, \dots, \xi_n$ statisztikai minta korrigált empirikus szórását jelöli.

2. Becsléselmélet

2.1 Pontbecslések

Tekintsük az $(\Omega,\Im)$ mérhető teret, ahol Ω egy véletlen kísérlet elemi eseményeinek halmaza, $\Im$ pedig Ω részhalmazainak σ-algebrája, azaz a megfigyelhető események halmazosztálya. Legyen továbbá adott a $\mathbb{P}=\{P\}$ valószínűségi mérték család, ahol $\forall\ P\in\mathbb{P}$ olyan valószínűség, ami lehetséges. Adott ezenkívül egy $\vartheta:\mathbb{P}\dashrightarrow IR$ függvény, melyet a $\mathbb{P}$-hez tartozó *paraméternek* nevezünk. Jelölje $\xi_1,\xi_2,\ldots,\xi_n,\ldots$a (nem feltétlenül véges elemszámú) statisztikai mintát. Feladat olyan $T_n(\xi_1,\xi_2,\ldots,\xi_n)\in IR\ (n=1,2,\ldots)$ statisztika-sorozat megadása, amely segítségével "jól" tudjuk becsülni a ϑ paramétervektort. Ha a paramétert "pontosan" meg tudjuk becsülni, akkor ez egyben azt is jelenti, hogy az adekvát P valószínűséget is közelítőleg megkapjuk. Az alábbiakban az elvárandó "jó", "pontos" becslési tulajdonságokat definiáljuk.

Definíció: A $T_n(\xi_1,\xi_2,\ldots,\xi_n)\in IR$ statisztika a $\vartheta\in IR$ paraméter *torzítatlan becslése*, ha $\forall\ P\in\mathbb{P}$ esetén a T_n -nek, mint vektor valószínűségi változónak létezik várható érték vektora és $\mathbf{M}_P T_n=\vartheta$.

Megjegyzés: 1.) Az $\mathbf{M}_P T_n$ azt jelöli, hogy a várható érték vektor függ attól, hogy melyik P valószínűségi mérték alapján számoljuk az $F_{T_n}(x)=P\big(T_n<x\big)$ eloszlásfüggvényt, majd abból a várható értéket.
2.) A Csebisev-egyenlőtlenségből tudjuk, hogy egy valószínűségi változó értékei a várható értéke körül ingadoznak. Az tehát, hogy egy statisztika a paraméter torzítatlan becslése, azt az elvárható tulajdonságot fejezi ki, hogy a becslési statisztika realizáltjai az ismeretlen paraméter körül ingadoznak a paramétertérben.

Definíció: A $T_n(\xi_1,\xi_2,\ldots,\xi_n)\in IR$ statisztika-sorozat a $\vartheta\in IR$ paraméter *aszimptotikusan torzítatlan becslése*, ha $\forall\ P\in\mathbb{P}$ esetén a T_n -nek, mint valószínűségi változónak létezik várható értéke és $\lim\limits_{n\to\infty}\mathbf{M}_P T_n=\vartheta$.

Megjegyzés: A torzítatlanságból nyilvánvalóan következik az aszimptotikusan torzítatlanság, tehát ez utóbbi a gyengébb tulajdonság.

Definíció: A $T_n(\xi_1,\xi_2,\ldots,\xi_n)\in IR$ statisztika-sorozat a $\vartheta\in IR$ paraméter *konzisztens becslése*, ha $\forall\ P\in\mathbb{P}$ és $\forall\ \varepsilon>0$ esetén $\lim\limits_{n\to\infty}P\Big(\big|T_n-\vartheta\big|>\varepsilon\Big)=0.$

Megjegyzés: A konzisztencia más követelményt fejez ki, mint a torzítatlanság. A konzisztencia tulajdonsága azt a jogos elvárást fejezi ki, hogy a megfigyelések számának növekedtével javuljon a becslés pontossága. Ezt a tulajdonságot még szemléletesebben fejezi ki az erősen konzisztencia, hiszen a megfigyelések számának növekedtével a becslés szórása, így a várható értéktől (a paramétertől) való eltérése is nullához tart.

Definíció: A $T_n(\xi_1, \xi_2, \ldots, \xi_n) \in \mathrm{IR}$ statisztika-sorozat a $\vartheta \in \mathrm{IR}$ paraméter *erősen konzisztens becslése*, ha $\forall$ n-re T_n torzítatlan becslése a paraméternek , és $\lim\limits_{n\to\infty} \mathbf{D}^2{}_P T_n = 0$.

Tétel: Ha a T_n (n=1,2,...) statisztika sorozat erősen konzisztens becslése ϑ-nak, akkor konzisztens becslése is.

Bizonyítás: Most $M_P T_n = \vartheta$. A Csebisev egyenlőtlenségből:

$$P\left(\left|T_n - \vartheta_i\right| > \varepsilon\right) \le \frac{\mathbf{D}_\mathbf{P}{}^2 T_n}{\varepsilon^2} \to 0 \ \ (n \to \infty) \ \Rightarrow \text{állítás.}$$

Definíció: Legyenek T^* és T^{**} a $\vartheta \in \mathrm{lR}$ paraméter torzítatlan becslései, ahol $\exists\, \mathbf{D}_\mathbf{P}{}^2 T^*$ és $\mathbf{D}_\mathbf{P}{}^2 T^{**}$ $(\forall P \in \mathbb{P})$. Azt mondjuk, hogy T^* *hatásosabb* becslése ϑ-nak mint, T^{**} ha $\mathbf{D}_\mathbf{P}{}^2 T^* \le \mathbf{D}_\mathbf{P}{}^2 T^{**}$ $(\forall P \in \mathbb{P})$-re.

Megjegyzés: Adott becslési problémánál lehetőleg a legkisebb szórásnégyzetű torzítatlan becslést kell alkalmazni.

Definíció: Ha a t^* torzítatlan statisztikára igaz, hogy $\mathbf{D}_\mathbf{P}{}^2 t^* = \min\limits_{\substack{Mt=\vartheta \\ \exists D^2 t}} \mathbf{D}_\mathbf{P}{}^2 t$ $(P \in \mathbb{P})$, akkor t^*-ot *hatásos* becslésnek nevezzük.

Megjegyzés: A Csebisev egyenlőtlenségből tudjuk, hogy a valószínűségi változó annál kisebb mértékben ingadozik a várható értéke körül, minél kisebb a szórása. Ez az oka, hogy a torzítatlan becslések között a hatásos becslés megkeresése a cél, hisz várhatólag ez pontosabb, mint bármely más torzítatlan becslés.

Példa: Legyen $(\Omega,\Im)$, $\mathbb{P}$ és rajta $\xi:\Omega\to\mathrm{IR}$ valószínűségi változó adott. Tegyük fel, hogy $\forall\ P\in\mathbb{P}$-re ξ egyenletes eloszlású valószínűségi változó a $[0,\vartheta]$ intervallumon, ahol $\vartheta>0$ ismeretlen paraméter.

Tehát most $\forall P\in\mathbb{P}$-re $F_{\xi,\vartheta}(x)=\dfrac{x}{\vartheta}$, $F'_{\xi,\vartheta}(x)=f_{\xi,\vartheta}(x)=\dfrac{1}{\vartheta}$, $x\in(0,\vartheta)$, $\mathbf{M}_\vartheta\xi=\dfrac{\vartheta}{2}$, $\mathbf{D}^2_\vartheta\xi=\dfrac{\vartheta^2}{12}$. Legyen továbbá $\xi_1,\xi_2,\ldots,\xi_n,\ldots$ statisztikai minta, amelynek eloszlásfüggvénye ξ-ével azonos $\forall\ P\in\mathbb{P}$-re.

Tekintsük a $T_1\overset{\text{def}}{=}\dfrac{n+1}{n}\xi^*_n$, $T_2\overset{\text{def}}{=}2\,\overline{\xi}_n$ statisztikákat! Megmutatjuk, hogy mindegyikük torzítatlan, de különböző szórású becslés, tehát eltér a hatásosságuk.

$$\mathbf{M}_\vartheta T_2=\mathbf{M}_\vartheta 2\,\overline{\xi}_n=2\,\mathbf{M}_\vartheta\overline{\xi}_n=2\,\mathbf{M}_\vartheta\xi=2\frac{\vartheta}{2}=\vartheta\Rightarrow T_2 \text{ torzítatlan.}$$

$$\mathbf{D}^2_\vartheta T_2=4\,\mathbf{D}^2_\vartheta\overline{\xi}_n=4\frac{\mathbf{D}^2_\vartheta\xi}{n}=4\frac{\vartheta^2}{12\cdot n}=\frac{\vartheta^2}{3\cdot n}.$$

A ξ^*_n eloszlásfüggvénye:

$$P\left(\xi^*_n<x\right)=\left[F_{\xi,\vartheta}(x)\right]^n=\left(\frac{x}{\vartheta}\right)^n,\quad x\in[0,\vartheta]\Rightarrow \text{sűrűségfüggvénye}\quad f_{n,\vartheta}(x)\overset{\text{def}}{=}n\frac{x^{n-1}}{\vartheta^n}\quad x\in(0,\vartheta).$$

$$\mathbf{M}_\vartheta\xi^*_n=\int_0^\vartheta x\,f_{n,\vartheta}(x)\,dx=\int_0^\vartheta n\frac{x^n}{\vartheta^n}\,dx=n\frac{1}{\vartheta^n}\left[\frac{x^{n+1}}{n+1}\right]_0^\vartheta=\frac{n}{n+1}\vartheta\Rightarrow\mathbf{M}_\vartheta T_1=\vartheta\text{, torzítatlan.}$$

$$\mathbf{D}^2_\vartheta T_1=\mathbf{M}_\vartheta T_1^2-\left(\mathbf{M}_\vartheta T_1\right)^2=\frac{(n+1)^2}{n^2}\mathbf{M}_\vartheta(\xi^*_n)^2-\vartheta^2=\frac{(n+1)^2}{n^2}\int_0^\vartheta x^2\,n\frac{x^{n-1}}{\vartheta^n}\,dx-\vartheta^2=$$

$$=\frac{(n+1)^2}{n}\frac{1}{\vartheta^n}\left[\frac{x^{n+2}}{n+2}\right]_0^\vartheta-\vartheta^2=\frac{(n^2+2n+1-n^2-2n)\,\vartheta^2}{n(n+2)}=\frac{\vartheta^2}{n(n+2)}.$$

Az eredményt összegezve: $\mathbf{D}^2_\vartheta T_1<\mathbf{D}^2_\vartheta T_2$, azaz a T_1.hatásosabb torzítatlan becslés T_2-nél.

Tétel: Legyen $(\Omega,\Im)$, $\mathbb{P}$ és rajta $\xi : \Omega \to \mathbb{R}$ valószínűségi változó adott. Tegyük fel, hogy $\forall\ P\in\mathbb{P}$ -re $\exists\, \mathbf{M}_P\xi$. Legyen most a paraméter $\vartheta = \vartheta(P) = \mathbf{M}_P\xi$. Legyen továbbá $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ statisztikai minta, amelynek eloszlásfüggvénye ξ-ével azonos $\forall\ P\in\mathbb{P}$ -re.
Akkor

a.) A $\overline{\xi}_n = \frac{1}{n}\cdot\sum_{i=1}^{n}\xi_i$ empirikus közép statisztika a ϑ elméleti várható érték paraméter torzítatlan becslése.

b.) Ha a feltételekhez azt is hozzávesszük, hogy $\forall\ P\in\mathbb{P}$-re $\exists\mathbf{D}_P^2\xi$ is, úgy $\overline{\xi}_n$ erősen konzisztens becslés is.

c.) A lineáris statisztikák között a $\overline{\xi}_n$ statisztika a leghatásosabb, azaz ha tetszőleges $c_1, c_2, \ldots, c_n$, $\sum_{i=1}^{n} c_i = 1$ valós súlyokkal tekintjük a $t = \sum_{i=}^{n} c_i\cdot\xi_i$ lineáris becslést, akkor t torzítatlan, és $\mathbf{D}_P^2\overline{\xi}_n \le \mathbf{D}_P^2 t$.

Bizonyítás:

a.) $M_P\overline{\xi}_n = M_P\left(\frac{1}{n}\cdot\sum_{i=1}^{n}\xi_i\right) = \frac{1}{n}\cdot\sum_{i=1}^{n} M_P\xi_i = \frac{1}{n}\cdot n\cdot\vartheta = \vartheta$.

b.) $\mathbf{D}_P^2\overline{\xi}_n = \mathbf{D}_P^2\left(\frac{1}{n}\cdot\sum_{i=1}^{n}\xi_i\right) = \frac{1}{n^2}\cdot\sum_{i=1}^{n}\mathbf{D}_P^2\xi_i = \frac{1}{n^2}\cdot n\cdot\mathbf{D}_P^2\xi = \frac{\mathbf{D}_P^2\xi}{n} \to 0$.

c.) Először is megjegyezzük, hogy a $c_i = \frac{1}{n}$ $(i = 1, 2, \ldots, n)$ súlyválasztásnál az átlagstatisztika is lineáris becslés. Legyen $\varepsilon_i = c_i - \frac{1}{n}$ $(i = 1, 2, \ldots, n)$. Ekkor $\sum_{i=1}^{n}\varepsilon_i = \sum_{i=1}^{n} c_i - 1 = 0$. Így

$$\mathbf{D}_P^2\left(\sum_{i=1}^{n} c_i\cdot\xi_i\right) = \sum_{i=1}^{n} c_i^2\cdot\mathbf{D}_P^2\xi_i = \mathbf{D}_P^2\xi\cdot\sum_{i=1}^{n} c_i^2 = \mathbf{D}_P^2\xi\cdot\sum_{i=1}^{n}(\varepsilon_i + \frac{1}{n})^2 =$$

$$= \mathbf{D}_P^2\xi\cdot\left[\sum_{i=1}^{n}\varepsilon_i^2 + \frac{2}{n}\cdot\sum_{i=1}^{n}\varepsilon_i + \frac{1}{n}\right] \ge \frac{\mathbf{D}_P^2\xi}{n} = \mathbf{D}_P^2\overline{\xi}_n \quad .$$

Tétel: Legyen $(\Omega,\Im)$, $\mathbb{P}$ és rajta $\xi:\Omega\to \mathrm{IR}$ valószínűségi változó adott. Tegyük fel, hogy $\forall\ P\in\mathbb{P}$-re $\exists\ \mathbf{D}_{\mathbf{P}}^{\mathbf{2}}\xi$. Legyen most a paraméter $\vartheta=\vartheta(P)=\mathbf{D}_{\mathbf{P}}^{\mathbf{2}}\xi$. Legyen továbbá $\xi_1,\xi_2,\dots,\xi_n,\dots$ statisztikai minta, amelynek eloszlásfüggvénye ξ-ével azonos $\forall\ P\in\mathbb{P}$-re.
Akkor

a.) Az $s_n^{\ 2}=\frac{1}{n}\cdot\sum_{i=1}^{n}(\xi_i-\overline{\xi}_n)^2$ empirikus szórásnégyzet statisztika a ϑ elméleti szórásnégyzet paraméter aszimptotikusan torzítatlan becslése, az $s_n^{*\,2}=\frac{1}{n-1}\cdot\sum_{i=1}^{n}(\xi_i-\overline{\xi}_n)^2$ korrigált empirikus szórásnégyzet statisztika pedig a ϑ elméleti szórásnégyzet paraméter torzítatlan becslése.

b.) Ha a feltételekhez azt is hozzávesszük, hogy $\forall\ P\in\mathbb{P}$-re $\exists\mathbf{M}_P\xi^4$ is, úgy $s_n^{\ 2}$ konzisztens, $s_n^{*\,2}$ erősen konzisztens becslés is.

Bizonyítás:
Fel fogjuk használni a Steiner tételt:

Segédtétel: *(Steiner)*
Tetszőleges $a,x_1,x_2,\dots,x_n$ valós számokra

$$\frac{1}{n}\cdot\sum_{i=1}^{n}(a-x_i)^2=(a-\overline{x}_n)^2+\frac{1}{n}\cdot\sum_{i=1}^{n}(\overline{x}_n-x_i)^2\ge\frac{1}{n}\cdot\sum_{i=1}^{n}(\overline{x}_n-x_i)^2.$$

Másrészt a=0 választással, átrendezés után: $\frac{1}{n}\cdot\sum_{i=1}^{n}(\overline{x}_n-x_i)^2=\frac{1}{n}\cdot\sum_{i=1}^{n}x_i^{\ 2}-\overline{x}_n^{\ 2}$.

A segédtétel bizonyítása:

$$\frac{1}{n}\cdot\sum_{i=1}^{n}(a-x_i)^2=\frac{1}{n}\cdot\sum_{i=1}^{n}(a-\overline{x}_n+\overline{x}_n-x_i)^2=(a-\overline{x}_n)^2+2\cdot(a-\overline{x}_n)\cdot\frac{1}{n}\cdot\sum_{i=1}^{n}(\overline{x}_n-x_i)+\frac{1}{n}\cdot\sum_{i=1}^{n}(\overline{x}_n-x_i)^2.$$

A tétel bizonyítása:

$$\mathbf{M}_P s_n^{\ 2}=\mathbf{M}_P\left(\frac{1}{n}\cdot\sum_{i=1}^{n}(\xi_i-\overline{\xi}_n)^2\right)=\mathbf{M}_P\left[\frac{1}{n}\cdot\sum_{i=1}^{n}\xi_i^{\ 2}-(\overline{\xi}_n)^2\right]=\frac{1}{n}\cdot\sum_{i=1}^{n}\mathbf{M}_P\xi_i^{\ 2}-\mathbf{M}_P(\overline{\xi}_n)^2=$$

a.)

$$=\frac{1}{n}\cdot n\cdot(\vartheta+(\mathbf{M}_P\xi)^2)-(\frac{\vartheta}{n}+(\mathbf{M}_P\xi)^2)=\frac{n-1}{n}\vartheta\to\vartheta\ \ (n\to\infty).$$

Mivel $s_n^{*\,2}=\frac{n}{n-1}s_n^{\ 2}\Rightarrow\mathbf{M}_P s_n^{*\,2}=\frac{n}{n-1}\mathbf{M}_P s_n^{\ 2}=\vartheta$.

b.)Belátható, hogy $\mathbf{D}_{\mathbf{P}}^{\mathbf{2}}s_n^2=\frac{n^2}{(n-1)^2}\left(\frac{\mathbf{M}_P\xi^4}{n}-\frac{n-3}{n(n-1)}\vartheta^2\right)\to 0$, $\mathbf{D}_{\mathbf{P}}^{\mathbf{2}}s_n^{*2}\to 0$ amiből, már $s_n^{\ 2}$ konzisztenciája következik.

Tétel: Legyen $(\Omega,\Im)$, $\mathbb{P}$ és rajta $\xi:\Omega\to \mathrm{IR}$ valószínűségi változó adott. Legyen továbbá $\xi_1,\xi_2,\ldots,\xi_n,\ldots$ statisztikai minta, amelynek eloszlásfüggvénye ξ-ével azonos $\forall$ $P\in\mathbb{P}$-re. Rögzítsük most az $x\in\mathrm{IR}$ valós számot. Ekkor a $\vartheta=\vartheta(P)=F_P(x)$ valós paraméter $\mathbb{P}$-n. Akkor az $F_n(x)$ empirikus eloszlásfüggvény a ϑ elméleti eloszlásfüggvény torzítatlan, erősen konzisztens becslése.

Bizonyítás: Az empirikus eloszlásfüggvény definíciójából nyilvánvaló, hogy $n\cdot F_n(x)\in\{0,1,2,\ldots,n\}$ és

$P(n\cdot F_n(x)=k)=P(k{:}\text{db} \quad i \quad \text{indexre} \quad \xi_i<x\,,(n-k) \quad \text{db} \quad j \quad \text{indexre } j\neq i \ \ \xi_j\geq x)=$

$$=\binom{n}{k}\cdot\left[F_P(x)\right]^k\cdot\left[1-F_P(x)\right]^{n-k} \Rightarrow n\cdot F_n(x)\in B(n,\vartheta).$$

Azaz $n\cdot F_n(x)$ binomiális eloszlású n és $F_p(x)=\vartheta$ paraméterekkel. Viszont ekkor $\mathbf{M}_P(n\cdot F_n(x))=n\cdot\vartheta$ és $\mathbf{D}^2_\mathbf{P}(n\cdot F_n(x))=n\cdot\vartheta\cdot(1-\vartheta)$. Innét pedig $\mathbf{M}_P(F_n(x))=\vartheta$ és $\mathbf{D}^2_\mathbf{P}(F_n(x))=\dfrac{\vartheta\cdot(1-\vartheta)}{n}\leq\dfrac{1}{4n}\to 0 \ \ (n\to\infty)$ következik, ami az állítást igazolja.

Megjegyzés: 1.) $\vartheta-\vartheta^2\leq 0.25.$

2.) Mivel az erősen konzisztenciából következik a konzisztencia, ezért $\forall\varepsilon>0$, $\forall x\in\mathrm{IR}$, $\forall P\in\mathbb{P}$-re $P(|F_n(x)-F_P(x)|>\varepsilon)\to 0 \ \ (n\to\infty)$. Ennél az állításnál lényegesen erősebbet fogalmaz meg a következő Glivenko-tól származó tétel: az empirikus eloszlásfüggvény 1 valószínűséggel, egyenletesen konvergál az elméleti eloszlásfüggvényhez. Elméleti jelentősége miatt a tételt a *matematikai statisztika alaptételének* is hívják.

Tétel: a.) A c_n empirikus kovariancia az $\mathbf{M}_P(\xi-\mathbf{M}_P\xi)(\eta-\mathbf{M}_P\eta)$ elméleti kovariancia torzítatlan becslése. Ha még azt is feltehetjük, hogy $\exists\,\mathbf{M}_P\xi^4$, $\mathbf{M}_P\eta^4$ is, akkor c_n erősen konzisztens becslés is.
b.) Az ρ_n empirikus korrelációs együttható az elméleti korrelációs együttható aszimptotikusan torzítatlan becslése. Ha még azt is feltehetjük, hogy $\exists\,\mathbf{M}_P\xi^4$, $\mathbf{M}_P\eta^4$ is, akkor ρ_n konzisztens becslés is.

Eddig csak arról volt szó, hogy milyen jó tulajdonságai lehetnek egy statisztikának, de még nem tudjuk, milyen módszerekkel lehet egy adott becslési problémához alkalmas statisztikát előállítani. A következőkben két általános becslési módszert fogunk ismertetni.

Definíció: Legyen adva az $(\Omega,\Im)$ mérhető téren a $\mathbb{P}$ valószínűségi mértékek tere, amelyhez adott a $\xi_1,\xi_2,\ldots,\xi_n$ statisztikai minta , amelyek eloszlásfüggvénye abszolút folytonos $\forall\ P\in\mathbb{P}$-re, azaz létezik a minta sűrűségfüggvénye. Legyen továbbá adott a $\vartheta:\mathbb{P}\to \mathrm{IR}$ paraméterfüggvény. Jelölje most $L(\underline{x},\vartheta)=\prod_{i=1}^{n} f_\vartheta(x_i)$ a minta együttes sűrűségfüggvényét. A ϑ paraméter *maximum likelihood* becslésén azt a $T_n(\xi_1,\xi_2,\ldots,\xi_n)$ statisztikát értjük, melyre $L(\underline{x},T_n(\underline{x}))=\max\limits_{\vartheta\in \mathrm{IR}^k} L(\underline{x},\vartheta)$ teljesül $(\forall\ \underline{x}\in \mathrm{IR}^n)$.

Definíció: Legyen adva az $(\Omega,\Im)$ mérhető téren a $\mathbb{P}$ valószínűségi mértékek tere, amelyhez adott a $\xi_1,\xi_2,\ldots,\xi_n$ diszkrét eloszlású statisztikai minta $\forall\ P\in\mathbb{P}$-re. Legyen továbbá adott a $\vartheta:\ \mathbb{P}\to \mathrm{IR}$ paraméterfüggvény.

Jelölje most $L(\underline{x},\vartheta)=P_\vartheta(\xi_1=x_1,\xi_2=x_2,\ldots,\xi_n=x_n)=\prod_{i=1}^{n} P_\vartheta(\xi_i=x_i)$ a minta együttes eloszlását. A ϑ paraméter *maximum likelihood* becslésén azt a $T_n(\xi_1,\xi_2,\ldots,\xi_n)$ statisztikát értjük, melyre $L(\underline{x},T_n(\underline{x}))=\max\limits_{\vartheta\in \mathrm{IR}} L(\underline{x},\vartheta)$ teljesül $(\forall\ \underline{x}\in \mathrm{IR}^n)$.

Megjegyzések: 1.) $L(\underline{x},\vartheta)$-t *likelihood függvénynek* is nevezik. Az elnevezés jogos, mert most az együttes sűrűségfüggvényben nem $\underline{x}$-et, hanem ϑ-t tekintjük változónak.

2.) A módszer alapgondolata a következő: mintavételezés során az $\underline{x}=(x_1,x_2,\ldots,x_n)^T$ realizációt mértük. Feltételezzük, hogy azért éppen ezt a realizációt kaptuk, és nem mást, mert az összes realizációk közül ennek a legnagyobb a bekövetkezési valószínűsége. Vegyük tehát, az összes ϑ paraméter közül azt, amelynél éppen az $\underline{x}$ realizáció bekövetkezése a maximális. A választ mind a folytonos, mind a diszkrét esetben a $L(\underline{x},\vartheta)\to\max\limits_{\vartheta\in \mathrm{IR}}$ szélsőérték-feladat megoldásából kapjuk meg.

3.) Mivel a természetes alapú logaritmusfüggvény szigorúan monoton növekvő, a $L(\underline{x},\vartheta)\to\max\limits_{\vartheta\in \mathrm{IR}}$ feladat helyett sokszor célszerű az $\ln L(\underline{x},\vartheta)\to\max\limits_{\vartheta\in \mathrm{IR}}$ szélsőérték-feladatot megoldani, ugyanis ugyanott lépnek fel a maximumhelyek. Az $l(\underline{x},\vartheta)\overset{\text{def}}{=}\ln L(\underline{x},\vartheta)$ függvényt *loglikelihood függvénynek* nevezzük.

4.) A maximumhelyet az $\dfrac{\partial l(\underline{x},\vartheta)}{\partial\vartheta}=0$ egyenlet megoldásai között kereshetjük.

Példa: A várható érték maximum-likelihood becslése normális esetben, amikor ismert a szórás.

Legyen $f_\vartheta(x) = \frac{1}{\sqrt{2\pi}\,\sigma_0} e^{-\frac{1}{2\sigma_0^2}(x-\vartheta)^2}$, ahol $\sigma_0 > 0$ rögzített, $\vartheta \in \mathrm{lR}$ az ismeretlen paraméter.

Most a likelihood függvény: $L(\underline{x},\vartheta) = \left(\frac{1}{\sqrt{2\pi}\,\sigma_0}\right)^n \cdot e^{-\frac{1}{2\sigma_0^2}\cdot\sum_{i=1}^{n}(x_i-\vartheta)^2}$, a loglikelihood függvény pedig $l(\underline{x},\vartheta) = n\cdot\ln\left(\frac{1}{\sqrt{2\pi}\,\sigma_0}\right) - \frac{1}{2\sigma_0^2}\cdot\sum_{i=1}^{n}(x_i-\vartheta)^2$. A maximumhely keresése:

$\frac{dl(\underline{x},\vartheta)}{d\vartheta} = \frac{1}{\sigma_0^2}\cdot\sum_{i=1}^{n}(x_i-\vartheta) = 0 \Rightarrow \vartheta = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}_n$. Mivel $\frac{d^2 l(\underline{x},\vartheta)}{d\vartheta^2} = -\frac{1}{\sigma_0^2} < 0$, a kapott stacionárius hely maximumhely $\Rightarrow$ az átlagstatisztika normális esetben a várható érték maximum-likelihood becslése.

A maximum likelihood módszerrel akkor is dolgozhatunk, ha a paraméterek száma egynél több. Ilyenkor többváltozós szélsőérték számítás útján lehet a statisztikákat előállítani, ahogy azt a következő példában látjuk.

Példa: A várható érték és a szórásnégyzet maximum-likelihood becslései normális esetben.

Legyen $f_{\vartheta_1,\vartheta_2}(x) = \frac{1}{\sqrt{2\pi\,\vartheta_2}} e^{-\frac{1}{2\vartheta_2}(x-\vartheta_1)^2}$, ahol $\vartheta_2 > 0$ rögzített, $\vartheta_1 \in \mathrm{lR}$ az ismeretlen paraméterek.

Most a likelihood függvény: $L(\underline{x},\vartheta_1,\vartheta_2) = \left(\frac{1}{\sqrt{2\pi\,\vartheta_2}}\right)^n \cdot e^{-\frac{1}{2\vartheta_2}\cdot\sum_{i=1}^{n}(x_i-\vartheta_1)^2}$, a loglikelihood függvény pedig $l(\underline{x},\vartheta_1,\vartheta_2) = -n\cdot\ln\sqrt{2\pi} - \frac{n}{2}\ln\vartheta_2 - \frac{1}{2\vartheta_2}\cdot\sum_{i=1}^{n}(x_i-\vartheta_1)^2$. A maximumhely keresése:

$$\frac{\partial l(\underline{x},\vartheta_1,\vartheta_2)}{\partial\vartheta_1} = \frac{1}{\vartheta_2}\cdot\sum_{i=1}^{n}(x_i-\vartheta_1) = 0 \Rightarrow \vartheta_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}_n$$

$$\frac{\partial l(\underline{x},\vartheta_1,\vartheta_2)}{\partial\vartheta_2} = -\frac{n}{2\vartheta_2} + \frac{1}{2\vartheta_2^2}\cdot\sum_{i=1}^{n}(x_i-\vartheta_1)^2 = 0 \Rightarrow \vartheta_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\vartheta_1)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x}_n)^2 = s_n^2$$

Mivel $\frac{d^2 l(\underline{x},\vartheta_1,\vartheta_2)}{d\vartheta_1^2} = -\frac{n}{\vartheta_2}$, $\frac{d^2 l(\underline{x},\vartheta_1,\vartheta_2)}{d\vartheta_2^2} = \frac{n}{2\vartheta_2^2} - \frac{1}{\vartheta_2^3}\sum_{i=1}^{n}\left(x_i-\vartheta_1\right)^2$,

$\frac{d^2 l(\underline{x},\vartheta_1,\vartheta_2)}{d\vartheta_1 d\vartheta_2} = -\frac{1}{\vartheta_2^2}\sum_{i=1}^{n}\left(x_i-\vartheta_1\right)$ a kapott stacionárius hely Hesse mátrixa: $\begin{pmatrix} \frac{-n}{s_n^2} & 0 \\ 0 & \frac{-n}{2\left(s_n^2\right)^2} \end{pmatrix}$

amiből látszik, hogy a hely maximumhely $\Rightarrow$ az átlagstatisztika és az empirikus szórásnégyzet statisztikák normális esetben az elméleti várható érték és szórásnégyzet maximum-likelihood becslései.

Példa: *A várható érték maximum-likelihood becslése Poisson eloszlás esetében.*

Most a minta eloszlása: $p_i(\vartheta)=\dfrac{\vartheta^i}{i!}e^{-\vartheta}$ $i=0,1,2,\ldots$. A likelihood függvény, a minta együttes eloszlásából számolható: $L(\underline{x},\vartheta)=\prod_{i=1}^{n}\dfrac{\vartheta^{x_i}}{x_i!}e^{-\vartheta}=\dfrac{\vartheta^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!}\cdot e^{-n\vartheta}$, a loglikelihood függvény pedig: $l(\underline{x},\vartheta)=\ln\vartheta\cdot\sum_{i=1}^{n}x_i-n\cdot\vartheta-\ln\left(\prod_{i=1}^{n}x_i!\right)$. A stacionárius helyek megkeresése: $\dfrac{\partial l(\underline{x},\vartheta)}{\partial\vartheta}=\dfrac{1}{\vartheta}\cdot\sum_{i=1}^{n}x_i-n=0\ \Rightarrow\ \vartheta=\dfrac{1}{n}\sum_{i=1}^{n}x_i=\overline{x}_n$. Mivel $\dfrac{\partial^2 l(\underline{x},\vartheta)}{\partial\vartheta^2}=-\dfrac{1}{\vartheta^2}\cdot\sum_{i=1}^{n}x_i<0$, a kapott stacionáriushely maximum. Tehát a Poisson eloszlás esetén is a paraméternek maximum likelihood becslése az átlagstatisztika.

Megjegyzés: A maximum likelihood módszer alapvető fontosságú a becsléselméletben. Ahol lehet, célszerű alkalmazni. Vannak azonban esetek, amikor a likelihood egyenlet a paraméterre transzcendens egyenletet ad, azaz a paraméterre a kifejtés lehetetlen. Ilyen esetekben sokszor hasznos a momentumok módszere. A módszer lényege az, hogy a minta elméleti momentumai függvénykapcsolatban vannak az eloszlás paramétereivel, és ebbe az ismert függvénybe a mintából becsült empirikus momentumokat beírva kapjuk a becslési statisztikákat.

Definíció: Legyen adva az $(\Omega,\Im)$ mérhető téren a $\mathcal{P}$ valószínűségi mértékek tere, amelyhez adott a $\xi_1,\xi_2,\ldots,\xi_n$ statisztikai minta. Legyen továbbá adott a $\underline{\vartheta}:\mathcal{P}\to \mathrm{IR}^k$ paraméterfüggvény. Tegyük fel, hogy léteznek a $\mu_\alpha=\mathbf{M}_{\underline{\vartheta}}\xi_i^{\alpha}\overset{def}{=}g_\alpha(\underline{\vartheta})$ $(\alpha=1,2,\ldots,k)$ momentumok, és $\exists\, g_\alpha^{-1}(\mu_1,\mu_2,\ldots,\mu_k)=\vartheta_\alpha$ $(\alpha=1,2,\ldots,k)$. Tekintsük a $\hat{\mu}_\alpha\overset{def}{=}\dfrac{1}{n}\cdot\sum_{i=}^{n}\xi_i^{\alpha}$ $(\alpha=1,2,\ldots,k)$ empirikus momentum statisztikákat. Akkor a $m_\alpha\overset{def}{=}g_\alpha^{-1}(\hat{\mu}_1,\hat{\mu}_2,\ldots,\hat{\mu}_k)$ $(\alpha=1,2,\ldots,k)$ statisztikák a ϑ_α paraméterek *momentumos becslései.*

Megjegyzés: A momentumok módszere nem rendelkezik olyan optimális tulajdonságokkal, mint a maximum likelihood módszer, de azért az általános feltételek mellet belátható, hogy a becslései konzisztensek. A konzisztencia azon múlik, hogy az empirikus momentumok is konzisztens becslései az elméleti momentumoknak.

Példa: *(A normális eloszlás paramétereinek becslése a momentumok módszerével)*

A minta sűrűségfüggvénye $f_{\mu,D}(x) = \frac{1}{\sqrt{2\pi D}} e^{-\frac{1}{2D}(x-\mu)^2}$. A normális eloszlás esetén tudjuk, hogy $\mu = g(\mu_1, \mu_2) = \mu_1$, $D = g(\mu_1, \mu_2) = \mu_2 - \mu_1^2$.

Az empirikus momentumok: $\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} \xi_i = \overline{\xi}_n$, $\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} \xi_i^2$. Így a momentumbecslések egyből adódnak: $\mu \approx m_1 = g_1(\overline{\mu}_1, \overline{\mu}_2) = \overline{\xi}_n$, $D \approx g_2(\overline{\mu}_1, \overline{\mu}_2) = \frac{1}{n}\sum_{i=1}^{n} \xi_i^2 - \frac{1}{n}\sum_{i=1}^{n} \xi_i = s_n^2$.

Látható, hogy ugyanazok a statisztikák adódtak, mint a maximum likelihood módszernél.

Példa: *(A Poisson eloszlás paraméterének becslése a momentumok módszerével)*

A minta eloszlása most $P_\vartheta(\xi_i = k) = \frac{\vartheta^k}{k!} e^{-\vartheta}$ $(k = 0,1,2,\ldots)$ .A $\vartheta > 0$ paraméter éppen az elméleti várható érték, az első momentum, így a momentumbecslés egyből adódik: $\vartheta \approx \overline{\mu}_1 = \overline{\xi}_n$. Ezúttal is ugyanazt a statisztikát kaptuk, mint a maximum likelihood módszernél.

2.2 Intervallumbecslések

A 2.1 pontban az ismeretlen paramétervektort a minta egy függvényével, azaz statisztikával próbáltuk meg közelíteni. Konkrét realizációnál tehát, a paramétertér egy pontját egy másik ponttal becsüljük. Ezért beszélünk pontbecslésről. De tudjuk azt is, hogy folytonos eloszlásoknál, annak valószínűsége, hogy a valószínűségi változó az értékkészletének éppen egy tetszőlegesen kiválasztott pontját fogja felvenni, nulla. Az intervallumbecsléseknél a mintából készített tartományokat definiálunk, amely tartományok nagy valószínűséggel lefedik a kérdéses paraméter-pontot.

Definíció: Legyen adva az $(\Omega,\Im)$ mérhető téren a $\mathcal{P}$ valószínűségi mértékek tere, amelyhez adott a $\xi_1, \xi_2, \ldots, \xi_n$ statisztikai minta. Legyen továbbá adott a $\vartheta : \mathcal{P} \to IR$ paraméterfüggvény. Legyen $\varepsilon > 0$ rögzített. Azt mondjuk, hogy a ϑ paraméterhez megadtunk egy *legalább* $1-\varepsilon$ *szignifikancia szintű konfidencia intervallumot*, ha $t_1(\xi_1, \xi_2, \ldots, \xi_n)$ és $t_2(\xi_1, \xi_2, \ldots, \xi_n)$ olyan statisztikák, hogy $P_\vartheta(t_1(\xi_1, \xi_2, \ldots, \xi_n) \le t_2(\xi_1, \xi_2, \ldots, \xi_n)) = 1$, és $P_\vartheta(t_1(\xi_1, \xi_2, \ldots, \xi_n) \le \vartheta \le t_2(\xi_1, \xi_2, \ldots, \xi_n)) \ge 1-\varepsilon$ fennáll minden $\forall\, P_\vartheta \in \mathcal{P}$-re.

Megjegyzés: Ahhoz, hogy példákat mutassunk konfidencia intervallumra, ki kell mondanunk az alábbi tételt.

Tétel: *(Lukács)*
Legyen $\xi_1, \xi_2, \ldots, \xi_n \in N(\mu, \sigma)$ eloszlásból származó statisztikai minta.
Akkor

1.) $\overline{\xi}_n \in N(\mu, \frac{\sigma}{\sqrt{n}})$,

2.) $\frac{n \cdot s_n^2}{\sigma^2} \in \chi^2_{n-1}$,

3.) $\overline{\xi}_n$ és s_n^2 függetlenek ($\Rightarrow \overline{\xi}_n$ és s_n^{*2} is függetlenek).

Megjegyzés: Belátható, hogy ha $\xi_1, \xi_2, \ldots, \xi_n \in N(\mu, \sigma)$ eloszlásból származó statisztikai minta, akkor a

$\frac{\overline{\xi}_n - \mu}{\sigma}\sqrt{n} \in N(0,1)$, $\frac{(n-1)\, s_n^{*2}}{\sigma^2} \in \chi^2_{n-1}$ statisztikák függetlenek, így –

$$\frac{\frac{\overline{\xi}_n - \mu}{\sigma}\sqrt{n}}{\sqrt{\frac{\frac{(n-1)\, s_n^{*2}}{\sigma^2}}{n-1}}} = \frac{\overline{\xi}_n - \mu}{s_n^{*}}\sqrt{n} \in t_{n-1} \text{ (n-1 szabadságfokú Student eloszlású).}$$

Példa: *(Konfidencia intervallum szerkesztése az ismeretlen várható értékre ismert szórású normális eloszlás esetében)*
Legyenek $\xi_1, \xi_2, \ldots, \xi_n \in N(\mu, \sigma_0)$ eloszlásból származó statisztikai minta, ahol $\sigma_0 > 0$ ismert, $\mu \in \mathbb{R}$ ismeretlen. Szerkesszünk μ-re adott $0 < \varepsilon < 1$ mellett $1-\varepsilon$ szintű konfidencia intervallumot! A Lukács tételből tudjuk, hogy $u \overset{def}{=} \frac{\overline{\xi}_n - \mu}{\sigma_0}\sqrt{n} \in N(0,1)$, azaz a statisztika sűrűségfüggvénye: $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. $\varphi(x)$ segítségével megadható olyan $u_\varepsilon > 0$ szám, hogy $\int\limits_{-u_\varepsilon}^{+u_\varepsilon} \varphi(t)\, dt = P(-u_\varepsilon < u < u_\varepsilon) = \Phi(u_\varepsilon) - \Phi(-u_\varepsilon) = 2\,\Phi(u_\varepsilon) - 1 = 1 - \varepsilon$ teljesüljön. Az $u_\varepsilon > 0$ szám meghatározását a $\Phi(u_\varepsilon) = 1 - \frac{\varepsilon}{2}$ egyenletből, standard normális eloszlás-táblázat segítségével határozhatjuk meg. Mivel az $\left\{\omega \middle| -u_\varepsilon < u(\omega) < u_\varepsilon\right\}$ esemény ekvivalens az $\left\{\omega \middle| \overline{\xi}_n(\omega) - \frac{u_\varepsilon \sigma_0}{\sqrt{n}} < \mu < \overline{\xi}_n(\omega) + \frac{u_\varepsilon \sigma_0}{\sqrt{n}}\right\}$ eseménnyel, ezért

$P(\overline{\xi}_n - \frac{u_\varepsilon \sigma_0}{\sqrt{n}} < \mu < \overline{\xi}_n + \frac{u_\varepsilon \sigma_0}{\sqrt{n}}) = 1 - \varepsilon$, azaz a $T_1 = T_1 \overset{def}{=} \overline{\xi}_n - \frac{u_\varepsilon \sigma_0}{\sqrt{n}}$, $T_2 \overset{def}{=} \overline{\xi}_n + \frac{u_\varepsilon \sigma_0}{\sqrt{n}}$ $1-\varepsilon$ szintű konfidencia intervallum μ-re.

Példa: *(Konfidencia intervallum szerkesztése az ismeretlen várható értékre ismeretlen szórású normális eloszlás esetében)*
Legyenek $\xi_1, \xi_2, ..., \xi_n \in N(\mu, \sigma_0)$ eloszlásból származó statisztikai minta, ahol $\sigma > 0$ is és , $\mu \in \mathrm{IR}$ is ismeretlen. Szerkesszünk μ-re adott $0 < \varepsilon < 1$ mellett $1 - \varepsilon$ szintű konfidencia intervallumot! A Lukács tétel után tett megjegyzés alapján: $\frac{\overline{\xi}_n - \mu}{s_n^*}\sqrt{n} \in t_{n-1}$, azaz az n-1 szabadságfokú Student eloszláshoz tartozó táblázatból kiolvasható olyan $t_\varepsilon > 0$ szám, amellyel $1 - \varepsilon = P(-t_\varepsilon < \frac{\overline{\xi}_n - \mu}{s_n^*}\sqrt{n} < t_\varepsilon) = P(\overline{\xi}_n - \frac{t_\varepsilon s_n^*}{\sqrt{n}} < \mu < \overline{\xi}_n + \frac{t_\varepsilon s_n^*}{\sqrt{n}}) \Rightarrow$ azaz most a $T_1 \overset{def}{=} \overline{\xi}_n - \frac{u_\varepsilon s_n^*}{\sqrt{n}}$, $T_2 \overset{def}{=} \overline{\xi}_n + \frac{u_\varepsilon s_n^*}{\sqrt{n}}$ statisztika pár lesz $1 - \varepsilon$ szintű konfidencia intervallum μ-re.

Példa: *(Konfidencia intervallum szerkesztése az ismeretlen szórásra normális eloszlás esetében)*
Legyenek $\xi_1, \xi_2, ..., \xi_n \in N(\mu, \sigma_0)$ eloszlásból származó statisztikai minta, ahol $\sigma > 0$ is és $\mu \in \mathrm{IR}$ is ismeretlen. Szerkesszünk σ-ra adott $0 < \varepsilon < 1$ mellett $1 - \varepsilon$ szintű konfidencia intervallumot! A Lukács tételre hivatkozva megint: $\frac{(n-1)s_n^{*2}}{\sigma^2} \in \chi^2_{n-1}$. Az n-1 szabadságfokú χ^2-eloszlás táblázatból (ld. függeléket) megadhatók olyan $0 < c_1 < c_2$ számok, hogy $1 - \varepsilon = P(c_1 < \frac{(n-1)s_n^*}{\sigma^2} < c_2)$ teljesüljön . (A c_1, c_2 értékek nyilván kielégítik a $P(\chi^2_{n-1} > c_1) = 1 - \frac{\varepsilon}{2}$ és $P(\chi^2_{n-1} > c_2) = \frac{\varepsilon}{2}$ feltételeket.) Egyszerű átrendezéssel kapjuk, hogy $1 - \varepsilon = P(\sqrt{\frac{(n-1)}{c_2}}s_n^* < \sigma < \sqrt{\frac{(n-1)}{c_1}}s_n^*)$, azaz $T_1 \overset{def}{=} \sqrt{\frac{(n-1)}{c_2}}s_n^*$, $T_2 \overset{def}{=} \sqrt{\frac{(n-1)}{c_1}}s_n^*$ statisztika pár lesz $1 - \varepsilon$ szintű konfidencia intervallum σ-ra.

Példa: *(Konfidencia intervallum szerkesztése az ismeretlen paraméterre exponenciális eloszlás esetében)*
Legyen $\xi_1, \xi_2, ..., \xi_n \in E(\lambda)$ eloszlásból származó statisztikai minta, ahol $\lambda > 0$ ismeretlen. Szerkesszünk λ-ra adott $0 < \varepsilon < 1$ mellett $1 - \varepsilon$ szintű konfidencia intervallumot!
A probléma megoldásához felhasználjuk az alábbi segédtételt:

Segédtétel: Legyen $\xi_1, \xi_2, ..., \xi_n \in E(\lambda)$ eloszlásból származó statisztikai minta.
Akkor

a.) $\lambda \cdot \xi_i \in E(1)$,

b.) $\sum_{\alpha=1}^{n} \lambda\xi_\alpha = \lambda \cdot n \cdot \overline{\xi}_n \in \Gamma(n,1)$, azaz n,1 paraméterű gamma eloszlású,

$f_\Gamma(x) = \frac{x^{n-1}e^{-x}}{(n-1)!}$ $(x > 0)$ sűrűségfüggvénnyel.

A konfidencia intervallum szerkesztése:

Az n,1 paraméterű gamma eloszláshoz tartozó táblázatból kiolvashatóak olyan $0 < \gamma_1 < \gamma_2$ számok, melyekkel $1-\varepsilon = P(\gamma_1 < \lambda n \overline{\xi}_n < \gamma_2) = P(\frac{\gamma_1}{n \overline{\xi}} < \lambda < \frac{\gamma_2}{n \overline{\xi}}) \Rightarrow$

azaz $T_1 \stackrel{\text{def}}{=} \frac{\gamma_1}{n \overline{\xi}_n}$, $T_2 \stackrel{\text{def}}{=} \frac{\gamma_2}{n \overline{\xi}_n}$ statisztika pár lesz $1-\varepsilon$ szintű konfidencia intervallum λ-ra.

A γ_1, γ_2 számokat úgy kell meghatározni, hogy $P(0 < \Gamma(n,1) < \gamma_1) = P(\Gamma(n,1) > \gamma_2) = \frac{\varepsilon}{2}$ legyen.

Ellenőrző kérdések és gyakorló feladatok

1. Mi a statisztikai minta és mi a mintarealizáció fogalma? Mi a különbség?
2. Mi a statisztika?
3. Mikor beszélünk paraméteres problémáról?
4. Mit nevezünk a paraméter torzítatlan, és mit aszimptotikus torzítatlan becslésének?
5. Mikor erősen konzisztens, és mikor konzisztens egy statisztika sorozat?
6. Mikor hatásos egy torzítatlan becslés?
7. Hogyan definiálunk egy konfidencia intervallumot? Mi a megbízhatósági szint?
8. Melyik állítás igaz, melyik hamis?
    a. Az átlagstatisztika a minta elméleti várható értékének aszimptotikusan torzítatlan becslése.
    b. Az átlagstatisztika a minta elméleti várható értékének mindig erősen konzisztens becslése.
    c. Az átlagstatisztika, ha létezik a statisztikai minta elméleti szórásnégyzete, a minta elméleti várható értékének konzisztens becslése.
    d. A minta elméleti szórásnégyzet statisztikája az elméleti szórásnégyzet aszimptotikusan torzítatlan becslése.
    e. Normális eloszlású alapsokaság esetén a mintaátlag lesz az elméleti várható érték paraméterének maximum likelihood becslése.
    f. Normális eloszlású alapsokaság esetén a korrigált empirikus szórásnégyzet statisztika lesz az elméleti szórásnégyzet paraméterének maximum likelihood becslése.
    g. A paraméterhez tartozó 90%-os konfidencia intervallum 90%-os valószínűséggel lefedi a becsülendő paramétert.
    h. A 90%-os konfidencia intervallum tartalmazza a 95%-os konfidencia intervallumot.
    i. Az empirikus eloszlásfüggvény az argmentumának minden rögzített értékénél statisztika.
    j. Az erősen konzisztens tulajdonságból következik a konzisztens tulajdonság.
    k. A maximum likelihood függvény folytonos esetben a minta együttes sűrűségfüggvényének helyettesítési értéke a mintarealizáció helyén.
    l. A maximum likelihood módszer mindig torzítatlan becslést ad.
    m. Normális eloszlású minta esetén az elméleti várható értékre, ha a szórás nem ismert, a normális eloszlás táblázata segítségével lehet konfidencia intervallumot szerkeszteni.
    n. Normális eloszlású minta esetén a mintaátlag és empirikus szórásnégyzet statisztikák függetlenek.

o. Normális eloszlású minta esetén a mintaátlag és empirikus szórásnégyzet statisztikák, mivel mindketten ugyanannak a mintának a függvényei, nem lehetnek függetlenek.
p. Az n szabadságfokú χ^2-eloszlás előáll n elemű standard normális eloszlású statisztikai minta elemeinek négyzetösszegeként.
q. Az n szabadságfokú χ^2-eloszlás előáll n elemű standard normális eloszlású statisztikai minta elemeinek összegeként (konvolúciójaként).

9. Legyen $\xi_1, \xi_2, \ldots, \xi_n$ exponenciális eloszlásból származó statisztikai minta, azaz a minta sűrűségfüggvényéről tudjuk, hogy $f(x) = \frac{1}{\vartheta} e^{-\frac{x}{\vartheta}}$, $x > 0$. Igazolja, hogy a $T_1 = n\xi_1^*$ és $T_2 = \overline{\xi}_n$ statisztikák a ϑ paraméter torzítatlan becslései. Melyikük a hatásosabb?
10. Legyen $\xi_1, \xi_2, \ldots, \xi_n$ statisztikai minta $f(x) = e^{\vartheta - x}$, $x > \vartheta$ sűrűségfüggvénnyel, ahol ϑ az eloszlás ismeretlen paramétere. Igazolja, a $T = \xi_1^* - \frac{1}{n}$ statisztika a ϑ paraméter torzítatlan és erősen konzisztens becslése!
11. Legyen $\xi_1, \xi_2, \ldots, \xi_n$ Laplace eloszlásból származó statisztikai minta, azaz a minta elméleti sűrűségfüggvénye $f(x) = \frac{\vartheta}{2} e^{-\vartheta|x|}$, $x \in IR$. Adja meg a ϑ paraméter maximum likelihood becslését!
12. Előző statisztikai vizsgálatokból tudjuk, hogy egy csapágy átmérőjének mérőszáma normális eloszlást követ σ=1,2 (mm) szórással. Legalább hány elemű mintát kell képeznünk ahhoz, hogy az elméleti várható értéket 90%-os megbízhatósági szinten lefedő konfidencia intervallum hossza legfeljebb 1 mm legyen?
13. Az alábbi ötelemű minta realizáció ismeretlen szórású normális eloszlásból származik. Adja meg a várható érték 90%-os konfidencia intervallumát!
$x_1 = 1{,}1$, $x_2 = 1$, $x_3 = 1{,}3$, $x_4 = 1{,}1$, $x_5 = 0{,}5$.

2. Hipotéziselmélet

Tekintsük a K véletlen kísérletet és a hozzátartozó $(\Omega,\Im)$ mérhető teret, és a $\mathcal{P}$ valószínűségi mértékek osztályát, ahol $(\Omega,\Im,P)$ Kolmogorov-féle valószínűségi mező $\forall\ P\in\mathcal{P}$-re. Tegyük fel, hogy $\mathcal{P}$ két diszjunkt részhalmazra bontható: $\mathcal{P}=\mathcal{P}_0 \cup \mathcal{P}_1$, és $\mathcal{P}_0 \cap \mathcal{P}_1=\emptyset$. Statisztikai módszert (ú.n. *próbát*) akarunk konstruálni annak eldöntésére, hogy a véletlen kísérlethez tartozó tényleges P valószínűségi mérték melyik halmazhoz tartozik $\mathcal{P}_0$ és $\mathcal{P}_1$ közül. Ehhez felállítunk egy $H_0: P\in\mathcal{P}_0$ *nullhipotézist*, és egy $H_1: P\in\mathcal{P}_1$ *alternatív hipotézist*. A nullhipotézis azt a feltevésünket fogalmazza meg, hogy az elméleti $P\in\mathcal{P}$ valószínűség a $\mathcal{P}_0$ részhez tartozik, az alternatív hipotézisünk szerint pedig azt, hogy ellenkezőleg, pont a $\mathcal{P}_1$ részhez. A kettő feltevés közül az eljárás végén egyértelműen kiválasztjuk és elfogadjuk majd az egyiket. A döntést a $\xi_1,\xi_2,\ldots,\xi_n$ statisztikai minta segítségével fogjuk meghozni. Először is, el fogjuk készíteni a $t_n(\xi_1,\xi_2,\ldots,\xi_n)$ ú.n. *próbastatisztikát*, amely rendelkezni fog az alábbi tulajdonsággal: adott $0<\varepsilon<1$ számhoz megadhatók olyan $K_1(\varepsilon)<K_2(\varepsilon)$ számok, hogy $P(K_1(\varepsilon)\le t_n \le K_2(\varepsilon))\ge 1-\varepsilon$, $\forall P\in\mathcal{P}_0$.

A $K_1(\varepsilon), K_2(\varepsilon)$ értékeket *kritikus értékeknek*, a segítségükkel definiált $X_e \stackrel{\text{def}}{=} \left\{ \underline{x} \mid \underline{x}\in \mathrm{lR}^n ,\ K_1(\varepsilon)\le t_n(\underline{x})\le K_2(\varepsilon)\right\}$ n-dimenziós vektorhalmazt *elfogadási tartománynak*, az $X_k \stackrel{\text{def}}{=} \mathrm{lR}^n \setminus X_e$ komplemens halmazát pedig *kritikus tartománynak* nevezzük.

Az $1-\varepsilon$ szám a próba *szignifikancia szintje*. A döntést úgy hajtjuk végre, hogy ellenőrizzük, hogy a $\xi_1,\xi_2,\ldots,\xi_n$ minta $x_1,x_2,\ldots,x_n$ realizáltja beleesik-e az X_e elfogadási tartományba. Ha beleesik, akkor a H_0 hipotézist, ellenkező esetben a H_1 alternatív hipotézist fogjuk elfogadni. A hipotézis eldöntése másképpen alakulhat az egyes ε szignifikancia szinteken, ezért mindig jelezni kell, hogy milyen mellett, azaz milyen szignifikancia szint mellett fogadjuk el (vagy vetjük el) a nullhipotézist. Természetesen számolunk azzal is, hogy a döntésünk hibás. Azt mondjuk, hogy *elsőfajú hibát* követünk el, ha elvetjük a nullhipotézist, holott valójában az igaz. *Másodfajú hibát* akkor követünk el, ha elfogadjuk a nullhipotézist, holott az nem igaz. Minden más esetben helyesen döntünk. A döntési hibafajtákat az alábbi táblázatban mutatjuk:

Valóság / Döntés	H_0 igaz	H_1 igaz
H_0 mellett	jó döntés	másodfajú hiba
H_1 mellett	elsőfajú hiba	jó döntés

A hibavétések elméleti valószínűségeit az alábbiakban definiáljuk:

Definíció: A $p_1(\varepsilon, n, P) \overset{def}{=} P((\xi_1, \xi_2, \ldots, \xi_n)^T \in X_k)$, $P \in \mathcal{P}_0$ függvényt *elsőfajú hibavalószínűségnek* nevezzük.

Definíció: A $p_2(\varepsilon, n, P) \overset{def}{=} P((\xi_1, \xi_2, \ldots, \xi_n)^T \in X_e)$, $P \in \mathcal{P}_1$ függvényt *másodfajú hibavalószínűségnek* nevezzük.

2.1 Paraméteres próbák

Ha adott egy $\underline{\vartheta} : \mathcal{P} \to \mathrm{IR}^k$ bijektív paraméter-leképezés, akkor a $\mathcal{P}=\mathcal{P}_0 \cup \mathcal{P}_1$, és $\mathcal{P}_0 \cap \mathcal{P}_1 = \emptyset$ felbontás helyett a Θ paramétertér $\Theta = \Theta_0 \bigcup \Theta_1$, $\Theta_0 \bigcap \Theta_1 = \emptyset$ diszjunkt felbontása segítségével is megfogalmazhatjuk a hipotéziseinket: $H_0 : \underline{\vartheta} \in \Theta_0$, $H_1 : \underline{\vartheta} \in \Theta_1$.

2.1.1 Egymintás u-próba

Most csak olyan $\mathcal{P}$ valószínűségi mértékeket tekintünk, ahol a $\xi_1, \xi_2, \ldots, \xi_n$ minta adott $\sigma_0 > 0$ szórású, ismeretlen μ várható értékű normális eloszlású lesz, a ϑ paraméter a várható érték ($\vartheta = \mu$).

$\Theta_0 \overset{def}{=} \{\mu_0\}$, $\Theta_1 \overset{def}{=} \{\mu \neq \mu_0\}$. Azaz most a nullhipotézis $H_0 : \mathbf{M}_P\xi = \mu_0$, az alternatív hipotézis pedig $H_1 : \mathbf{M}_P\xi \neq \mu_0$. Azt akarjuk tehát eldönteni, hogy lehet-e a minta elméleti várható értéke egy adott μ_0 érték, vagy attól szignifikánsan különböző. Ha a H_0 hipotézis igaz, akkor a mintaelemek $N(\mu_0, \sigma_0)$ eloszlásúak, amiből következik, hogy a mintaátlag statisztika szintén normális eloszlású: $\overline{\xi}_n \in N(\mu_0, \frac{\sigma_0}{\sqrt{n}})$. Standardizálás után : $u \overset{def}{=} \frac{\overline{\xi}_n - \mu_0}{\sigma_0}\sqrt{n} \in N(0,1)$. A standard normális eloszláshoz a $\Phi(u_\varepsilon) = 1 - \frac{\varepsilon}{2}$ összefüggés alapján megadhatók olyan $K_1(\varepsilon) \overset{def}{=} -u_\varepsilon$, $K_2(\varepsilon) \overset{def}{=} u_\varepsilon$ kritikus értékek, melyekre, ha a H_0 hipotézis igaz, akkor fenn kell állnia, hogy $P(-u_\varepsilon < u < u_\varepsilon) = 1 - \varepsilon$. Adjuk meg tehát az u-próba kritikus tartományát az $X_k \overset{def}{=} \left\{ \underline{x} \;\middle|\; \left| \frac{\overline{x}_n - \mu_0}{\sigma_0}\sqrt{n} \right| \geq u_\varepsilon \right\}$ definícióval.

A nullhipotézist az adott mintarealizáció felhasználásával az $|u| = \left| \frac{\overline{x}_n - \mu_0}{\sigma_0}\sqrt{n} \right| < u_\varepsilon$ reláció ellenőrzése alapján döntjük el. Ha az előbbi egyenlőtlenség fennáll, akkor az adott szignifikancia szinten elfogadjuk a nullhipotézist, ellenkező esetben a minta várható értéke szignifikánsan különbözik a hipotetikus μ_0 értéktől.

Megjegyzés: A nullhipotézis annál megbízhatóbban fogadható el, minél nagyobb az ε értéke. A gyakorlatban, ha ε közel van 1-hez a nullhipotézis erősen igaznak mutatkozik, $\varepsilon \le 0.01$ esetben viszont csak nagyon nagy elemszámú minta esetén célszerű elfogadni azt.

Az elsőfajú hiba valószínűségére:
$p_1(\varepsilon, n, \mu_0) = P_{\mu_0}(|u| \ge u_\varepsilon) = 1 - P_{\mu_0}(-u_\varepsilon < u < u_\varepsilon) = 1 - (\Phi(u_\varepsilon) - \Phi(-u_\varepsilon)) = 2 - 2\Phi(u_\varepsilon) = \varepsilon.$ A másodfajú hiba valószínűsége pedig:

$$p_2(\varepsilon, n, \mu) = P_\mu(-u_\varepsilon < u < u_\varepsilon) = P_\mu(-u_\varepsilon < \frac{\overline{\xi}_n - \mu_0}{\sigma_0}\sqrt{n} < u_\varepsilon) =$$

$$= P_\mu(-u_\varepsilon - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma_0} < \frac{\overline{\xi}_n - \mu}{\sigma_0}\sqrt{n} < u_\varepsilon - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma_0}) =$$

$= \Phi\left(u_\varepsilon - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma_0}\right) - \Phi\left(-u_\varepsilon - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma_0}\right)$, ugyanis az alternatív hipotézis fennállása esetén lesz $\frac{\overline{\xi}_n - \mu}{\sigma_0}\sqrt{n} \in N(0,1)$.

Feladat Egy automata darabolónak 1200 mm hosszúságú acélszalagokat kell levágnia. Előzetes adatfelvételből ellenőriztük, hogy a gép által készített darabok hossza normális eloszlású valószínűségi változónak tekinthető σ=3 mm szórással. Ellenőrizni akarjuk a gép beállításának helyes voltát. Ezért a gyártmányokból 16 db szalagot véletlenszerűen kiválasztunk, és lemérünk. Az adatok az alábbiak voltak mm-ben: 1193 , 1196 , 1198 , 1195 , 1198 , 1199 , 1204 , 1193 , 1203 , 1201 , 1196 , 1200 , 1191 , 1196 , 1198 , 1191 . Vizsgáljuk meg, hogy van-e szignifikáns eltérés az előírt mérettől!

Megoldás: Egymintás u-próbával kell a $H_o : m = 1200$ nullhipotézisről dönteni. A mintaátlag $\overline{x}_{16} = 1197$, így a próbastatisztika számított értéke: $u = \frac{|1197 - 1200|}{3}\sqrt{16} = 4$. Mivel csak ε=0,0001 elsőfajú hibánál lehetne elfogadni a nullhipotézist, így akkor járunk el helyesen, ha elvetjük azt, azaz a méretek az előírt 1200 mm-es hossztól szignifikánsan eltérnek.

2.1.2 A kétmintás u-próba

Adott a $\xi_1, \xi_2, \ldots, \xi_n$ és a $\eta_1, \eta_2, \ldots, \eta_m$ egymástól független statisztikai minták. Most csak olyan $\mathcal{P}$ valószínűségi mértékeket tekintünk, ahol a minták peremeloszlásai $\sigma_1 > 0$ illetve $\sigma_2 > 0$ ismert szórású, de ismeretlen μ_1 illetve μ_2 várható értékű normális eloszlásúak, azaz a két mintához tartozó együttes sűrűségfüggvény: $f_{\mu_1,\mu_2}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(y-\mu_2)^2}{2\sigma_2^2}}$. Feltett hipotézisek:
$H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \ne \mu_2$ .A feltételek miatt a két minta átlagstatisztikájára: $\overline{\xi}_n \in N(\mu_1, \frac{\sigma_1}{\sqrt{n}})$, $\overline{\eta}_m \in N(\mu_2, \frac{\sigma_2}{\sqrt{m}})$. Mivel a két minta független volt, így a különbségükre:

$\overline{\xi}_n - \overline{\eta}_m \in N(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}})$. Ha feltesszük, hogy a nullhipotézis igaz, akkor $\overline{\xi}_n - \overline{\eta}_m \in N(0, \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}})$ is fennáll. Standardizálás után: $\frac{\overline{\xi}_n - \overline{\eta}_m}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \in N(0,1)$.

Adott $0 < \varepsilon < 1$ esetén, tehát most az elfogadási tartomány:

$$X_e = \left\{ (\underline{x}^T, \underline{y}^T)^T \;\middle|\; \left| \frac{\overline{x}_n - \overline{y}_m}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \right| < u_\varepsilon \right\}$$, ahol az $u_\varepsilon > 0$ kritikus értékre: $\Phi(u_\varepsilon) = 1 - \frac{\varepsilon}{2}$.

A hipotézis eldöntése tehát úgy történik, hogyha az adott mintarealizációknál teljesedik a $\left| \frac{\overline{x}_n - \overline{y}_m}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \right| < u_\varepsilon$ reláció, akkor a nullhipotézist az adott ε szignifikancia szinten elfogadjuk, ellenkező esetben pedig elvetjük. Ha a H_0 hipotézist fogadjuk el, úgy is fogalmazhatunk, hogy a két minta elméleti várható értéke között "nincsen szignifikáns különbség".
A kétmintás u-próba elsőfajú hibája is ε.

Feladat A textiliparban minőségellenőrzésnél fontos annak a vizsgálata, hogy két, minta alapján minősített tétel azonos tulajdonságúnak tekinthető-e vagy sem. Az első $n_1 = 100$ elemű mintatételben az egységnyi fonalhossz tömegének átlagára $\overline{x}_{100} = 195\,(\text{gr})$, a második $n_2 = 70$ elemű mintatételben ugyanakkor az átlagra $\overline{y}_{70} = 185\,(\text{gr})$ adódott. Feltételezve, hogy a tömegmérés pontossága (a szórás) σ=18 gr, döntsünk 99%-os szignifikancia szinten arról, hogy a két tétel fonaltömegei azonosaknak tekinthetők-e!

Megoldás: A mintákat független normális eloszlásúaknak tekintve, mivel a szórás ismert, kétmintás t-próbával dönthetünk a várható értékek egyezésére vonatkozó nullhipotézisről. A próbastatisztika számított értéke most: $u = \frac{|195 - 185|}{18 \cdot \sqrt{\frac{1}{100} + \frac{1}{70}}} \approx 3{,}56$. Az ε=0,01 elsőfajú hiba valószínűséghez tartozó kritikus érték $u_{0,001} = 2{,}58$. Mivel a számított érték nagyobb, mint a kritikus érték, így az adott szignifikancia szinten a két tétel szignifikánsan eltérőnek mutatkozik, azaz a nullhipotézist elvetjük.

2.1.3 Az egymintás t-próba

Most csak olyan P valószínűségi mértékeket tekintünk, ahol a $\xi_1, \xi_2, \ldots, \xi_n$ minta ismeretlen $\sigma > 0$ szórású és ismeretlen μ várható értékű normális eloszlású lesz, a ϑ paraméter a várható érték ($\vartheta = \mu$).

$\Theta_0 \stackrel{\text{def}}{=} \{\mu_0, \sigma > 0\}$, $\Theta_1 \stackrel{\text{def}}{=} \{\mu \neq \mu_0, \sigma > 0\} \subseteq IR^2$. Azaz most a nullhipotézis $H_0 : \mathbf{M}_P\xi = \mu_0$, az alternatív hipotézis pedig $H_1 : \mathbf{M}_P\xi \neq \mu_0$. Azt akarjuk tehát eldönteni, hogy lehet-e a minta elméleti várható értéke egy adott μ_0 érték, vagy attól szignifikánsan különböző. Ha a H_0

hipotézis igaz, akkor a mintaelemek $N(\mu_0,\sigma)$ eloszlásúak, amiből következik, hogy a mintaátlag statisztika szintén normális eloszlású: $\overline{\xi}_n \in N(\mu_0,\frac{\sigma}{\sqrt{n}})$. Standardizálás után : $\frac{\overline{\xi}_n-\mu_0}{\sigma}\sqrt{n} \in N(0,1)$. Az ismeretlen σ szórás kiküszöbölését a Lukács tétel segítségével végezzük. Tudjuk, hogy $\frac{(n-1)s_n^{*2}}{\sigma^2} \in \chi^2_{n-1}$, akár igaz a nullhipotézis, akár nem. Felhasználva a Lukács tétel utáni megjegyzést: $t \stackrel{def}{=} \frac{\overline{\xi}_n-\mu_0}{s_n^*}\sqrt{n} \in t_{n-1}$.

Az n-1 szabadságfokú Student eloszlás táblázatából adott $0<\varepsilon<1$-hoz kiolvasható olyan $t_\varepsilon>0$ kritikus érték, mellyel a H_0 fennállása esetén $P(|t|<t_\varepsilon)=1-\varepsilon$ kell, hogy teljesüljön. Így a nullhipotézist aszerint fogadjuk vagy vetjük el, hogy $\left|\frac{\overline{\xi}_n-\mu_0}{s_n^*}\sqrt{n}\right|<t_\varepsilon$ fennáll-e vagy sem az adott mintarealizációnál. Mivel $p_1(\varepsilon,n,\mu_0)=P(|t|\geq t_\varepsilon)=\varepsilon$, így a t-próba esetében is ε az elsőfajú hiba nagysága.

Feladat Egy konzervgyárban adagolóautomata tölti a dobozokat. Az egy dobozba töltendő anyag tömegének várható értékére az előírás 500 gr. Mintavétel során az alábbi értékeket kapták grammokban: 483, 502, 498, 496, 502, 483, 494, 491, 505, 486. Döntsünk 95%-os szignifikancia szinten, hogy teljesül-e a várható értékre az m=500 gr előírás.

Megoldás: A dobozok súlyát normális eloszlásúnak tekintjük. Mivel a szórás ismeretlen, egymintás t-próbával dönthetünk a nullhipotézisről. A mintából számolt statisztikák: átlag $\overline{x}_{10}=494$, a korrigált empirikus szórásnégyzet $s_{10}^{*\,2}=64{,}9$. A számított próbastatisztika $t=\frac{|494-500|}{\sqrt{64{,}9}}\sqrt{10}\approx 2{,}36$. A 95%-hoz tartozó kritikus értéket a Student-eloszlás táblázatából olvashatjuk ki figyelembevéve, hogy a szabadságfok 9: $t_{0,05}=2{,}262$. Látható, hogy a számított érték a nagyobb, így a tétel átlaga szignifikánsan nagyobb, mint az előírt érték. A műszakvezetőnek intézkednie kell, hogy állítsák be pontosabban az adagoló automatát.

2.1.4 A kétmintás t-próba

Adott a $\xi_1,\xi_2,\ldots,\xi_n$ és a $\eta_1,\eta_2,\ldots,\eta_m$ egymástól független statisztikai minták. Most csak olyan P valószínűségi mértékeket tekintünk, ahol a minták peremeloszlásai $\sigma_1>0$ illetve $\sigma_2>0$ ismeretlen, de egyenlő nagyságú ($\sigma_1=\sigma_2=\sigma$) szórású és ismeretlen μ_1 illetve μ_2 várható értékű normális eloszlásúak. A két mintához tartozó együttes sűrűségfüggvény: $f_{\mu_1,\mu_2}(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{(x-\mu_1)^2}{2\sigma^2}-\frac{(y-\mu_2)^2}{2\sigma^2}}$. Feltett hipotézisek:
$H_0:\mu_1=\mu_2$, $H_1:\mu_1\neq\mu_2$ .A feltételek miatt a két minta átlagstatisztikájára: $\overline{\xi}_n\in N(\mu_1,\frac{\sigma}{\sqrt{n}})$, $\overline{\eta}_m\in N(\mu_2,\frac{\sigma}{\sqrt{m}})$. Mivel a két minta független volt, így a különbségükre:

$\overline{\xi}_n - \overline{\eta}_m \in N(\mu_1 - \mu_2, \sigma\sqrt{\frac{1}{n} + \frac{1}{m}})$. Ha feltesszük, hogy a nullhipotézis igaz, akkor $\overline{\xi}_n - \overline{\eta}_m \in N(0, \sigma\sqrt{\frac{1}{n} + \frac{1}{m}})$ is fennáll. Standardizálás után: $\dfrac{\overline{\xi}_n - \overline{\eta}_m}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \in N(0,1)$.

Ahhoz, hogy az ismeretlen σ értéket kiküszöbölhessük, felhasználjuk, hogy $\dfrac{(n-1){s^*_{\xi,n}}^2}{\sigma^2} \in \chi^2_{n-1}$, $\dfrac{(m-1){s^*_{\eta,m}}^2}{\sigma^2} \in \chi^2_{m-1}$, valamint azt, hogy a ${s^*_{\xi,n}}^2$, ${s^*_{\eta,m}}^2$, $\overline{\xi}_n$, $\overline{\eta}_m$ statisztikák a feltételek és a Lukács tétel miatt függetlenek egymástól. Először is $\dfrac{(n-1){s^*_{\xi,n}}^2}{\sigma^2} + \dfrac{(m-1){s^*_{\eta,m}}^2}{\sigma^2} \in \chi^2_{n+m-2}$, akár igaz a nullhipotézis, akár nem. Másrészt, a Lukács tétel után tett megjegyzés értelmében, ha a H_0 hipotézis igaz,

$$tt \overset{def}{=} \frac{\dfrac{\overline{\xi}_n - \overline{\eta}_m}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}}{\sqrt{\dfrac{\frac{(n-1){s^*_{\xi,n}}^2}{\sigma^2} + \frac{(m-1){s^*_{\eta,m}}^2}{\sigma^2}}{n+m-2}}} = \frac{\overline{\xi}_n - \overline{\eta}_m}{\sqrt{(n-1){s^*_{\xi,n}}^2 + (m-1){s^*_{\eta,m}}^2}}\sqrt{\frac{nm(n+m-2)}{n+m}} \in t_{n+m-2} \text{A}$$

fentiek alapján, az n+m-2 szabadságfokú Student eloszlás táblázatból adott $0 < \varepsilon < 1$ szignifikancia szinthez kiolvasható olyan $t_\varepsilon > 0$ kritikus érték, mellyel a H_0 fennállása esetén $P(|tt| < t_\varepsilon) = 1 - \varepsilon$ kell, hogy teljesüljön. Így a nullhipotézist aszerint fogadjuk vagy vetjük el, hogy $|tt| < t_\varepsilon$ fennáll-e vagy sem az adott mintarealizációnál. Mivel $p_1(\varepsilon, n, \mu_0) = P(|tt| \geq t_\varepsilon) = \varepsilon$, így a kétmintás t-próba esetében is ε az elsőfajú hiba nagysága.

Megjegyzés: Hangsúlyozzuk, hogy a kétmintás t-próba csak akkor alkalmazható, ha a két minta ismeretlen szórásait egyenlőnek tételezzük fel. (Különben nem tudtuk volna kiküszöbölni a tt próbastatisztikából σ-t !) A két minta szórásai egyezésének ellenőrzését az F-próbával végezhetjük, tehát ennek meg kell előznie a kétmintás t-próbát.

2.1.5 Az F-próba

Adott a $\xi_1, \xi_2, \ldots, \xi_n$ és a $\eta_1, \eta_2, \ldots, \eta_m$ egymástól független statisztikai minták. Most csak olyan P valószínűségi mértékeket tekintünk, ahol a minták peremeloszlásai $\sigma_1 > 0$ illetve $\sigma_2 > 0$ ismeretlen szórású és ismeretlen μ_1 illetve μ_2 várható értékű normális eloszlásúak. A két mintához tartozó együttes sűrűségfüggvény: $f_{\mu_1,\mu_2,\sigma_1,\sigma_2}(x,y) = \dfrac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(y-\mu_2)^2}{2\sigma_2^2}}$.

Felállított hipotézisek most a szórások egyezésére, illetve szignifikáns különbségére vonatkoznak: $H_0 : \sigma_1 = \sigma_2$, $H_1 : \sigma_1 \neq \sigma_2$. Ha feltesszük, hogy a nullhipotézis igaz, akkor a Lukács tétel szerint igaz lesz, hogy $\dfrac{(n-1){s^*_{\xi,n}}^2}{\sigma^2} \in \chi^2_{n-1}$, $\dfrac{(m-1){s^*_{\eta,m}}^2}{\sigma^2} \in \chi^2_{m-1}$, ahol $\sigma_1 = \sigma_2 = \sigma$. A minták függetlensége miatt a két statisztika is független lesz.

Belátható, hogy: $\dfrac{\dfrac{\dfrac{(n-1){s^*_{\xi,n}}^2}{\sigma^2}}{n-1}}{\dfrac{\dfrac{(m-1){s^*_{\eta,m}}^2}{\sigma^2}}{m-1}} = \dfrac{{s^*_{\xi,n}}^2}{{s^*_{\eta,m}}^2} \in F_{n-1,m-1}$, azaz a minták korrigált empirikus szórásnégyzeteinek hányadosa n-1,m-1 szabadságfokú Fisher-eloszlást fog követni, ha a nullhipotézis igaz. Ezek alapján a nullhipotézis eldöntésére a kritikus tartományt úgy szerkeszthetjük meg, hogy adott $0 < \varepsilon < 1$ szignifikancia szinthez az n-1,m-1 szabadságfokú F-eloszlás táblázatból kiolvasunk olyan $0 < K_1 < K_2$ kritikus értékeket, melyekre $P(K_1 < F_{n-1,m-1}) = 1 - \dfrac{\varepsilon}{2}$, $P(K_2 < F_{n-1,m-1}) = \dfrac{\varepsilon}{2}$. Ha az adott mintarealizációnál $K_1 < \dfrac{{s^*_{\xi,n}}^2}{{s^*_{\eta,m}}^2} < K_2$ reláció teljesül, a nullhipotézist elfogadjuk, ellenkező esetben pedig elvetjük.

A próba elsőfajú hibájának a valószínűsége most is ε, a másodfajú hiba valószínűsége az n és m mintaelemszámoktól, ε-tól és a $\sigma_1 - \sigma_2$ különbségtől függ.

Megjegyzések: 1.) Ha $\varepsilon < 0.33$, n és m kettőnél nagyobb mintaelemszámok (ez gyakorlatilag mindig fennáll), akkor a $0 < K_1 < K_2$ kritikus értékekre mindig teljesül a $K_1 < 1 < K_2$ reláció. Így, ha ${s^*_{\xi,n}}^2$, ${s^*_{\eta,m}}^2$ közül a nagyobbikat írjuk a számlálóba, a próba eldöntéséhez elég a próbastatisztika értékét csupán K_2-vel összehasonlítani. Ha a számított érték kisebb, mint K_2, a nullhipotézist elfogadjuk. Az F-eloszlás táblázatból csak egyetlen kritikus érték meghatározása elégséges ilyenkor, de ügyeljünk arra, hogy az első szabadságfok mindig abból a mintaelemszámból képződik, amelyhez tartozó korrigált empirikus szórásnégyzet statisztika a számlálóban van !

2.) Statisztikai elemzéseket napjainkban valamilyen statisztikai programrendszer segítségével szokás elvégezni. A programok egy próba esetén mindig azt a $0 < \varepsilon < 1$ elsőfajú hibavalószínűséget adják meg eredményül, amelynél már elfogadható a nullhipotézis. Ha tehát túl közel van 0-hoz, akkor az azt jelenti, hogy a nullhipotézist el kell vetni. 0.01-nél kisebb elsőfajú hibavalószínűség mellett „nem illik” elfogadni H_0-t, míg 0.1 felett a nullhipotézis fennállása erősnek mutatkozik. A két szélső érték közötti szignifikancia szintek esetén a felhasználó felelőssége, hogy elfogadja, vagy elveti H_0-t, vagy esetleg újabb mintavételezéssel bővíti a mintát (mintákat), majd megismétli a próbát.

Feladat Ki akarjuk mutatni, hogy egy kezelés amelyet állatokon végeztek, hatást gyakorol az állat testsúlyának növekedésére. Annál az n=10 állatnál, ahol nem volt speciális kezelés, az időegység alatt mért súlynövekedés 61,52,47,51,58,64,60,55,49,53 (kg) volt. Az állatok azon m=12 létszámú csoportjánál, ahol a kezelést elvégezték ugyanezen idő alatt a súlygyarapodások értékei 53,59,63,67,60,57,73,65,58,68,62,71 (kg) voltak. Igaz-e az a nullhipotézis, hogy a kezelés nem növeli az állatok testsúlyát?

Megoldás: A mintákat tekintsük normális eloszlásúaknak! Ekkor a nullhipotézisről kétmintás t-próbával dönthetünk, ha a minták ismeretlen szórásai egyenlők. Ezt előzetesen F-próbával ellenőrizni kell. A kezeletlen állatcsoport adatainak átlaga $\overline{x}_{10} = 55$, a korrigált empirikus szórásnégyzet pedig ${s^*_{10}}^{*} = 31$, míg a kezelt állatcsoportnál ugyanezekre a statisztikákra: $\overline{y}_{12} = 63$ és ${\sigma^*_{12}}^2 = 36$ számolható a mintákból. A F-próba statisztikájának számított értéke:

$\frac{\sigma_{10}^{*\,2}}{s_{12}^{*\,2}} = 1{,}16$, míg a 95%-os szignifikancia szinthez tartozó kritikus érték $K_{0,05} = 3{,}1$. (A szabadsági fokok $f_1 = 9$ és $f_2 = 11$). Tehát a két minta szórásai 95%-os szignifikancia szinten azonosnak tekinthetők, így alkalmazható a kétmintás t-próba a minták elméleti várható értékeinek egyezésére. A kétmintás t-próba statisztikájának számított értéke: $t = \frac{|\overline{x}_{12} - \overline{y}_{10}|}{\sqrt{(n-1)\sigma_{10}^{*\,2} + (m-1)s_{12}^{*\,2}}}\sqrt{\frac{nm(n+m-2)}{n+m}} \approx 3{,}214$, míg a 95%-hoz és a n+m-2=20 szabadságfokhoz tartozó kritikusértéket a Student eloszlás táblázatából kiolvasva $K_{0,05} = 2{,}528$-t kapunk. Mivel a számított érték nagyobb mint a kritikus érték, el kell vetnünk a két csoport azonos súlyára vonatkozó nullhipotézisünket. Mivel $\overline{x}_{12} > \overline{y}_{10}$, ezért csak az a feltevés állja meg a helyét, hogy a kezelés szignifikánsan megnöveli az állatok testsúlyát.

2.1.6 A Welch-próba

Ha az F-próbát el kell vetnünk, nem alkalmazható a kétmintás t-próba a két minta várható értékeinek egyezésének ellenőrzésére. Erre az esetre dolgozta ki Welch a most ismertetendő próbát. Adott a $\xi_1, \xi_2, \ldots, \xi_n$ és az $\eta_1, \eta_2, \ldots, \eta_m$ egymástól független statisztikai minták. Most is csak olyan P valószínűségi mértékeket tekintünk, ahol a minták peremeloszlásai $\sigma_1 > 0$ illetve $\sigma_2 > 0$ ismeretlen szórású és ismeretlen μ_1 illetve μ_2 várható értékű normális eloszlásúak. A két mintához tartozó együttes sűrűségfüggvény: $f_{\mu_1,\mu_2,\sigma_1,\sigma_2}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(y-\mu_2)^2}{2\sigma_2^2}}$. Feltett hipotézisek ugyanazok mint a kétmintás t-próbánál voltak: $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$. Megmutatható, hogy a nullhipotézis fennállása esetén a $W_{n,m} = \frac{\overline{\xi}_n - \overline{\eta}_m}{\sqrt{\frac{s_{\xi,n}^2}{n} + \frac{s_{\eta,m}^2}{m}}}$ próbastatisztika közelítőleg Student eloszlású [f] (egészrész f) szabadságfokkal, ahol $\frac{1}{f} = \frac{c^2}{m-1} + \frac{(1-c^2)}{n-1}$, $c \stackrel{def}{=} \frac{\frac{s_{\eta,m}^2}{m}}{\frac{s_{\eta,m}^2}{m} + \frac{s_{\xi,n}^2}{n}}$. A kritikus értéket a Student eloszlás táblázatából kiolvasva dönthetünk a szokásos módon a nullhipotézisről: elfogadjuk, ha az adott realizációknál a $|W_{n,m}|$ számított érték kisebb lesz. Ha $n, m \geq 40$, akkor $W_{n,m} \approx N(0, \frac{[f]}{[f]-2})$, azaz akkor a normális eloszlás táblázatából is kiolvashatjuk a kritikus értéket.

2.2 Nemparaméteres próbák

Ha az alapsokaság (a statisztikai minta) eloszlását nem tekintjük eleve ismertnek, akkor nemparaméteres próbákról beszélünk. Ilyenkor tehát az előzetes feltevéseink nagyon általánosak, de természetesek; pl. feltesszük, hogy a minta eloszlása folytonos, vagy feltesszük, hogy a szórás véges, stb. Nyilvánvaló, mivel kevesebb feltételt követelünk meg kiinduláskor (a'priori feltevések), a következtetéseink levonásához nagyobb elemszámú mintákra lesz szükségünk, mint a paraméteres próbák esetén.

2.2.1 χ^2-próbák

Tiszta illeszkedésvizsgálat

Adott a $\xi_1, \xi_2, \ldots, \xi_n$ statisztikai minta. Ellenőrizni akarjuk azt a feltevést, hogy a minta elméleti eloszlásfüggvénye éppen az $F_0(x)$, az összes szóbajöhető eloszlásfüggvény között. $F_0(x)$-nek nincsenek ismeretlen paraméterei, egy bizonyos, konkrét eloszlásfüggvény. A nullhipotézisünk most $H_0 : P(\xi < x) \equiv F_0(x)$, míg az alternatív hipotézis $H_1 : P(\xi < x) \not\equiv F_0(x)$. Adjuk meg a számegyenesnek egy tetszőleges r diszjunkt intervallumból álló felosztását! Legyen
$-\infty < x_1 < x_2 < \cdots < x_{r-1} < \infty$, $I_k \overset{def}{=} \left[x_{k-1}, x_k\right), (k = 1,2,\ldots,r)$, $x_0 \overset{def}{=} -\infty$, $x_r \overset{def}{=} +\infty$. Ha H_0 igaz, akkor $p_k \overset{def}{=} P(\xi \in I_k) = F_0(x_k) - F_0(x_{k-1})$.
Az $A_k \overset{def}{=} \left\{\omega \,\middle|\, \xi(\omega) \in I_k\right\}$ $(k = 1,2,\ldots,r)$ teljes eseményrendszer, így $\sum_{k=1}^{r} p_k = 1$. Jelölje υ_k azt a gyakoriságot, ahány mintaelemre teljesült a $\xi_\alpha \in I_k$ reláció, azaz υ_k nem más, mint az A_k esemény bekövetkezéseinek a száma egy n-szeres Bernoulli féle kísérletsorozatban. Ha a nullhipotézis igaz, akkor belátható, hogy $\sum_{i=1}^{r} \frac{\left(\upsilon_i - n \cdot p_i\right)^2}{n \cdot p_i} \to \chi^2_{r-1}$ $(n \to \infty)$. Vagyis, ha nagy a mintaelemszám, a $T_n = \sum_{i=1}^{r} \frac{\left(\upsilon_i - n \cdot p_i\right)^2}{n \cdot p_i} = \sum_{i=1}^{r} \frac{\upsilon_i^{\,2}}{np_i} - n$ statisztika a nullhipotézis fennállása esetén közelítőleg r-1 szabadságfokú χ^2-eloszlást követ. Erre alapozhatjuk a döntési eljárásunkat. Adott $0 < \varepsilon < 1$ szignifikancia szinthez meghatározunk olyan K_ε kritikus értéket, mellyel $P(\chi^2_{r-1} < K_\varepsilon) = 1 - \varepsilon$. Ezek után, ha az adott statisztikai minta realizációjánál teljesül a $T_n < K_\varepsilon$ reláció, a nullhipotézist elfogadjuk, ellenkező esetben pedig elvetjük. Az elsőfajú hibavalószínűség most csak aszimptotikusan lesz ε.

Megjegyzések: 1.) $x_1 < x_2 < \cdots < x_{r-1}$ osztópontokat úgy célszerű megválasztani, hogy a realizálódott mintánál $\upsilon_i \geq 10$ és $p_i \approx \frac{1}{r}$ legyen minden i-re.
2.) Ha $r \geq 30$, akkor a χ^2-eloszlás táblázat helyett a normális eloszlás táblázatát is használhatjuk, mert ilyenkor már $T_n \approx \chi^2_{r-1} \approx N(r-1, \sqrt{2r-2})$.
3.) Ha a statisztikai minta diszkrét, akkor az intervallumok helyett a minta értékkészletének diszjunkt felbontását vesszük. Pl., ha a k-adik partíciót az $I_k = \left\{z_1, z_2, \ldots, z_{n_k}\right\}$ számhalmaz jelenti, akkor $p_k \overset{def}{=} \sum_{i=1}^{n_k} P(\xi = z_i)$.

Feladat Egy textilüzemben korábbi tapasztalatok azt mutatták, hogy a fonalszakadások száma egy bizonyos géptípus és fonal esetén Poisson eloszlású $\lambda=8$ paraméterrel. Vizsgáljuk meg újabb adatokkal, hogy fennáll-e a Poisson eloszlás!

A fonalszakadások száma	Gyakoriság	Az elméleti Poisson eloszlásérték
0	-	0,00033
1	-	0,00268
2	-	0,01073
3	-	0,02863
4	1	0,05725
5	4	0,09160
6	9	0,12214
7	7	0,13959
8	8	0,13959
9	10	0,12408
10	8	0,09926
11	10	0,07219
12	5	0,04813
13	7	0,02962
14	2	0,01692
15	3	0,00903
16	1	0,00451
17	-	0,00212
18	-	0,00160
Összesen	75=n	1

(A táblázat harmadik oszlopának 5. sorában a $p_5 = \frac{8^5}{5!}e^{-8} \approx 0{,}09160$ valószínűség áll, a második oszlop ötödik eleme pedig azt a gyakoriságot mutatja az n=75 elemű mintában hányszor volt ötszörös fonalszakadás.)

Megoldás: A statisztikai módszer most a tiszta illeszkedésvizsgálat, hiszen az illesztendő eloszlás paramétere ismert. A mintaelemeket eleve összesen 18 csoportba osztottuk. Mivel a csoportok gyakorisága kicsi, célszerű először másik csoportokat kialakítani. Egy lehetséges átcsoportosítás lehet például:

Csoport	Gyakoriság (f_i)	Csoportvalószínűség (p_i)	$n\,p_i$	$\frac{(f_i - np_i)^2}{np_i}$
0-tól 6-ig	14	0,31336	23,6	3,905
7-től 8-ig	15	0,27918	20,9	1,666
9-től 10-ig	18	0,22334	16,7	0,101
11-től ∞-ig	28	0,18412	13,8	14,612
Összesen	75	1	75	20,284

A táblázat jobb alsó cellájában lehet a próbastatisztika számított értékét kiolvasni. Mivel az r-1=3 szabadságfokú χ^2-eloszlás táblázatában még az ε=0,05 elsőfajú hibavalószínűséghez is csak $K_{0,05} = 17,7$ kritikus érték tartozik, ezért a az illeszkedésre vonatkozó feltevést el kell vetni, esetleg becsléses illeszkedésvizsgálattal meg kell ismételni az eljárást. Látható ugyanis, hogy a mintaátlag most 9,48, ami jelentősen eltér az elméleti 8-as értéktől.

Becsléses illeszkedésvizsgálat

Adott a $\xi_1, \xi_2, \ldots, \xi_n$ statisztikai minta. Ellenőrizni akarjuk azt a feltevést, hogy a minta elméleti eloszlásfüggvénye $F_{\underline{\vartheta}}(x)$ alakú, az összes szóbajöhető eloszlásfüggvény között. $F_{\underline{\vartheta}}(x)$ egy eloszláscsalád paraméterektől függő általános eleme. A nullhipotézisünk most

$$H_0 : \exists \underline{\vartheta} \in \mathrm{lR}^k : P(\xi < x) \equiv F_{\underline{\vartheta}}(x),$$

míg az alternatív hipotézis

$$H_1 : \nexists \underline{\vartheta} \in \mathrm{lR}^k : P(\xi < x) \equiv F_{\underline{\vartheta}}(x).$$

A próba végrehajtása nagyon hasonlít az előző esetre, csak először venni kell a $\underline{\vartheta}$ paramétervektor $\underline{t}_n$ konzisztens becslését, majd az adott mintarealizációnál kapott $\hat{\underline{\vartheta}} = \underline{t}_n$ becsléssel képezzük az $F_0(x) \overset{\text{def}}{=} F_{\hat{\underline{\vartheta}}}(x)$ eloszlásfüggvényt, ami már konkrét, hiszen ismeretlen paramétereket már nem tartalmaz. Ezután végrehajtva mindazt, amit a tiszta illeszkedésvizsgálatnál leírtunk, kiszámoljuk a T_n próbastatisztikát. A különbség csak ott jelentkezik, hogy most az mutatható meg, hogy $T_n \overset{e}{\rightarrow} \chi^2_{r-1-k}$, ahol k a becsült paraméterek száma. Ezek alapján a döntési algoritmus az előzőekhez hasonlóan történik.

Feladat A megadott százéves minta alapján döntsünk arról a nullhipotézisről, hogy Budapest levegőjének januári középhőmérséklete normális eloszlást követ vagy sem!

A levegő január havi közepes hőmérséklete
Budapest, [°C]

1861	-3,6	1894	-2,7	1927	2,7
1862	-2,7	1895	-1,5	1928	-,3
1863	1,9	1896	-6,4	1929	-3,8
1864	-7,7	1897	0,1	1930	0,3
1865	0,3	1898	0,2	1931	0,4
1866	0,1	1899	2,7	1932	-1,2
1867	0,7	1900	0,9	1933	-2,1
1868	-0,6	1901	-5,3	1934	-1,5
1869	-2,7	1902	2,8	1935	-2,0
1870	-0,6	1903	-1,1	1936	4,2
1871	-2,0	1904	-2,1	1937	-2,3
1872	-0,4	1905	-3,8	1938	0,0
1873	1,7	1906	-0,7	1939	1,7
1874	-1,1	1907	-2,3	1940	-6,8
1875	-1,1	1908	-2,4	1941	-2,1
1876	-4,7	1909	-3,0	1942	-8,3

1877	1,7	1910	0,9	1943	-3,9
1878	-2,6	1911	1,0	1944	3,2
1879	-2,1	1912	-3,3	1945	-3,3
1880	-3,0	1913	-1,9	1946	-2,8
1881	-4,1	1914	-4,0	1947	-5,7
1882	0,8	1915	2,2	1948	4,2
1883	-1,5	1916	3,8	1949	1,5
1884	1,2	1917	0,5	1950	-2,2
1885	-0,8	1918	1,1	1951	2,8
1886	0,0	1919	2,5	1952	1,2
1887	-1,6	1920	3,1	1953	1,3
1888	-4,3	1921	4,6	1954	-4,8
1889	-1,9	1922	-1,5	1955	-1,0
1890	0,3	1923	1,8	1956	1,5
1891	-6,2	1924	-3,1	1957	-1,3
1892	-1,3	1925	0,2	1958	-0,3
1893	-9,0	1926	0,2	1959	0,1
				1960	1,2

Megoldás: Mivel a minta elméleti várható értéke és szórásnégyzete nem ismert, ezért azokat a mintából számolt statisztikákkal fogjuk becsülni: $\mu \approx \overline{x}_{100} = -1{,}02$ és $\sigma^2 \approx {s^*_{100}}^2 = 2{,}78$. Az átlagstatisztika és a korrigált empirikus szórásnégyzet normális esetben torzítatlan, erősen konzisztens becsléseket adnak, így a becsléses illeszkedésvizsgálatnál felhasználhatók. A nullhipotézisünk most az, hogy a minta eloszlása N(-1,02, $\sqrt{2{,}78}$). A mintaelemeket r=8 csoportba osztva elkészíthető az alábbi táblázat:

index (i)	csoport (intervallum)	gyakoriság (υ_i)	csoport valószínűség (p_i)	$\frac{(\upsilon_i - np_i)^2}{np_i}$
1	[-10°C, -4°C]	13	0,1466	0,1879672578445
2	(-4°C,-3°C]	9	0,0966	0,04509316770186
3	(-3°C,-2°C]	15	0,1243	0,5313676588898
4	(-2°C,-1°C]	14	0,1458	0,02307270233196
5	(-1°C, 0°C]	9	0,1453	2,104673090158
6	(0°C,+1°C]	16	0,1237	1,06523039612
7	(+1°C,+2°C]	12	0,0948	0,6698734177215
8	(+2°C,+5°]	12	0,1229	0,006842961757526
összesen		100	1	4,634120652525

A szabadságfok most 8-1-2=5, mert az illesztett eloszlás két paraméterét a mintából becsültük. A χ^2-eloszlás táblázat ból a 90%-os szignifikanciaszinhez tartozó kritikus érték $K_{0,1} = 9{,}236$, ami jóval nagyobb mint a táblázat jobb alsó cellájában olvasható számított érték, azaz a minta eloszlása normálisnak tekinthető. (Még ε=0,3 esetén is elfogadhatnánk a nullhipotézist.)

Függetlenségvizsgálat

a.) Először teljes eseményrendszerekre fogalmazzuk meg a problémát. Legyenek $A_1, A_2, \ldots, A_r$ és $B_1, B_2, \ldots, B_s$ a K véletlen kísérlettel kapcsolatos teljes eseményrendszerek $A_i \cdot A_j = \varnothing$ $(i \neq j)$, $\sum_{i=1}^{r} A_i = \Omega$ és $B_i \cdot B_j = \varnothing$ $(i \neq j)$, $\sum_{i=1}^{s} B_i = \Omega$. Hajtsuk végre a K kísérletet n-szer egymástól függetlenül. Jelölje υ_{ij} azon esetek számát, ahányszor az $A_i \cdot B_j$ esemény előfordult, vagyis υ_{ij} az $A_i \cdot B_j$ esemény gyakorisága az n-szeres Bernoulli kísérletsorozatban. Ekkor az $\upsilon_{i\cdot} \overset{def}{=} \sum_{j=1}^{s} \upsilon_{ij}$ az A_i esemény gyakorisága, $\upsilon_{\cdot j} \overset{def}{=} \sum_{i=1}^{r} \upsilon_{ij}$ pedig a B_j esemény gyakorisága.
Használni fogjuk továbbá az alábbi jelöléseket:
$p_{ij} \overset{def}{=} P(A_i \cdot B_j)$, $p_{i\cdot} \overset{def}{=} P(A_i) = \sum_{j=1}^{s} p_{ij}$, $p_{\cdot j} \overset{def}{=} P(B_j) = \sum_{i=1}^{r} p_{ij}$. A nullhipotézisünk az, hogy $A_1, A_2, \ldots, A_r$ és $B_1, B_2, \ldots, B_s$ függetlenek egymástól, azaz $H_0 : P(A_i \cdot B_j) = P(A_i)P(B_j)$ minden $i \neq j$-re. Evvel ellentétes állítás az alternatív hipotézis:
$H_1 : \exists i, j \colon P(A_i \cdot B_j) \neq P(A_i)P(B_j)$. A nullhipotézis fennállása esetén belátható, hogy :

$$T = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(\upsilon_{ij} - n \cdot p_{i\cdot} \cdot p_{\cdot j}\right)^2}{n \cdot p_{i\cdot} \cdot p_{\cdot j}} \overset{e}{\rightarrow} \chi^2_{r \cdot s - 1},$$

vagyis a T statisztika nagy mintaelemszám esetén közel χ^2 eloszlású. Ezért az r·s-1 szabadságfokú χ^2-eloszlás táblázatának segítségével szerkeszthetünk próbát a szokásos módon. A nullhipotézis átfogalmazható úgy is, hogy az $A_i \cdot B_j$ $(i = 1,2,\ldots,r$, $j = 1,2,\ldots,s)$ teljes eseményrendszerhez tartozó diszkrét eloszlás ellenőrzésére diszkrét tiszta illeszkedésvizsgálatot hajtunk végre.
b.) A két valószínűségi változó függetlenségére vonatkozó hipotézisvizsgálati módszer felhasználja az előző, a.) pontban kidolgozott eljárást. Legyen $(\xi_1, \eta_1)^T, (\xi_2, \eta_2)^T, \ldots, (\xi_n, \eta_n)^T$ n mintaelemszámú kétdimenziós statisztikai minta. Ellenőrizni akarjuk, hogy a minta komponensei függetlenek-e egymástól, vagy pedig szignifikáns sztochasztikus összefüggés tapasztalható közöttük:

$$H_0 : P(\xi_i < x, \eta_i < y) = P(\xi_i < x) \cdot P(\eta_i < y) \ \forall x, y;$$
$$H_1 : P(\xi_i < x, \eta_i < y) \not\equiv P(\xi_i < x) \cdot P(\eta_i < y).$$

Legyen $-\infty < x_1 < x_2 < \cdots < x_{r-1} < \infty, I_k \overset{def}{=} \left[x_{k-1}, x_k\right), (k = 1,2,\ldots,r)$, $x_0 \overset{def}{=} -\infty$, $x_r \overset{def}{=} +\infty$ és $-\infty < y_1 < y_2 < \cdots < y_{s-1} < \infty$, $J_k \overset{def}{=} \left[y_{k-1}, y_k\right), (k = 1,2,\ldots,s)$, $y_0 \overset{def}{=} -\infty$, $y_s \overset{def}{=} +\infty$ két különböző partícióra bontása lR-nek ! Azért kell két különböző felosztást tekintenünk, mert a két minta értékei másképpen oszolhatnak el a számegyenesen; az első beosztás az első komponens értékkészletét, a második partíció a második komponens értékkészletét fedi le.

A két felosztás segítségével értelmezhetjük az $A_k \overset{def}{=} \left\{\omega \middle| \xi(\omega) \in I_k\right\}$, $k = 1,2,\ldots,r$ és $B_\alpha \overset{def}{=} \left\{\omega \middle| \eta(\omega) \in J_\alpha\right\}, \alpha = 1,2,\ldots,s$ teljes eseményrendszereket. Nyilvánvaló, hogyha a nullhipotézis igaz, akkor a két teljes eseményrendszer is független egymástól, abban az értelemben, ahogy azt az a.) pontban leírtuk. A megfordítás nem igaz, azaz ez csak szükséges de nem elegendő feltétele a H_0 teljesülésének. A próbánkat ennek ellenére az eseményrendszerek függetlenségére alapozzuk: ha ugyanis az eseményrendszerek nem függetlenek egymástól, akkor a két valószínűségi változó sem lehet független, míg az eseményrendszerek függetlenségének elfogadása nem mond ellent a változók függetlensége elfogadásának.

Tehát a H_0 hipotézist átfogalmazzuk, és a $H_0^*: P(A_i \cdot B_j) = P(A_i)P(B_j)$ hipotézissel foglalkozunk a továbbiakban.

Az a.) pontban használt jelölésekkel: υ_{ij} az $A_i \cdot B_j$ esemény gyakorisága az n-szeres Bernoulli kísérletsorozatban, azaz azon mintaelemek száma, ahol $(\xi_\alpha, \eta_\alpha)^T \in I_i \times J_j$ teljesül. A $p_{ij} \overset{def}{=} P(A_i \cdot B_j)$, $p_{i\cdot} \overset{def}{=} P(A_i) = \sum_{j=1}^{s} p_{ij}$, $p_{\cdot j} \overset{def}{=} P(B_j) = \sum_{i=1}^{r} p_{ij}$ valószínűségek most nem ismertek, de azokat a relatív gyakoriságok segítségével becsülni lehet: $p_{i\cdot} \approx \hat{p}_{i\cdot} \overset{def}{=} \frac{1}{n}\upsilon_{i\cdot} = \frac{1}{n}\sum_{j=1}^{s}\upsilon_{ij}$, $p_{\cdot j} \approx \hat{p}_{\cdot j} \overset{def}{=} \frac{1}{n}\upsilon_{\cdot j} = \frac{1}{n}\sum_{i=1}^{r}\upsilon_{ij}$. A becslések száma r-1 illetve s-1, mivel eloszlásokról van szó, és így $\sum_{i=1}^{r} p_{i\cdot} = \sum_{j=1}^{s} p_{\cdot j} = 1$, vagyis $p_{r\cdot} = 1 - \sum_{i=1}^{r-1} p_{i\cdot}$ és $p_{\cdot s} = 1 - \sum_{j=1}^{s-1} p_{\cdot j}$, azaz az eloszlás utolsó elemei már a többi becslésből számolhatók.

Most tehát az $A_i \cdot B_j$ $(i = 1,2,\ldots,r\,,\ j = 1,2,\ldots,s)$ teljes eseményrendszerhez tartozó diszkrét eloszlás ellenőrzésére becsléses illeszkedésvizsgálatot kell végrehajtani, ahol a becsült paraméterek száma: r-1+s-1=r+s-2 . Az a.) pontban elmondottak szerint, a

$$T_n \overset{def}{=} \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(\upsilon_{ij} - n\cdot\hat{p}_{i\cdot}\cdot\hat{p}_{\cdot j}\right)^2}{n\cdot\hat{p}_{i\cdot}\cdot\hat{p}_{\cdot j}} = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(\upsilon_{ij} - \frac{\upsilon_{i\cdot}\cdot\upsilon_{\cdot j}}{n}\right)^2}{\frac{\upsilon_{i\cdot}\cdot\upsilon_{\cdot j}}{n}} = n\cdot\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\upsilon_{ij}^2}{\upsilon_{i\cdot}\cdot\upsilon_{\cdot j}} - n \text{ próbastatisztika}$$

eloszlása aszimptotikusan r·s-1-(r+s-2)=(r-1)·(s-1) szabadságfokú χ^2-eloszlású lesz!

A nullhipotézis eldöntéséhez táblázatból meg kell határoznunk olyan K_ε kritikus értéket, melyre $P(\chi^2_{(r-1)(s-1)} < K_\varepsilon) = 1 - \varepsilon$ teljesül. Ha a T_n számított értéke kisebb mint a K_ε kritikus érték, a nullhipotézist az 1-ε szignifikancia szinten elfogadjuk, ellenkező esetben az alternatív hipotézist tartjuk igaznak, azaz a komponensek között szignifikáns összefüggést tapasztalunk.

Feladat Független-e egymástól a szem és a haj színe? A döntést a mellékelt n=467 ember vizsgálatának eredményét tartalmazó táblázat adatai alapján hozzuk meg!

Megoldás:

	Hajszín		
Szemszín	Világos	Sötét	Összesen
Világos	307	32	339
Sötét	33	95	128
Összesen	340	127	467

A próbastatisztika a függetlenségvizsgálat során most $T = 467\cdot\frac{(307\cdot 95 - 32\cdot 33)^2}{339\cdot 128\cdot 340\cdot 127} = 196{,}93158$ jóval nagyobb, mint az 1 szabadságfokú χ^2-táblázatból kiolvasott kritikus érték. Például a táblázatban található leggyengébb szignifikancia szintnél (ε=0,001) a kritikus érték $K_{0,001} = 10{,}827$, vagyis a semmiképpen sem fogadható el az a nullhipotézis, hogy az embereknél a szemszín független lenne a hajszíntől.

Homogenitásvizsgálat

A homogenitásvizsgálat annak a kérdésnek az eldöntésére szolgál, hogy két valószínűségi változó azonos eloszlású-e, azaz ugyanaz a függvény-e az eloszlásfüggvényük, vagy sem. Adottak a $\xi_1, \xi_2, \ldots, \xi_n$ és az $\eta_1, \eta_2, \ldots, \eta_m$ statisztikai minták, amelyek egymástól is függetlenek. Eldöntendő, hogy: $H_0 : P(\xi < x) \equiv P(\eta < x)$ vagy $H_1 : P(\xi < x) \not\equiv P(\eta < x)$.

Tekintsük most a $-\infty < x_1 < x_2 < \cdots < x_{r-1} < \infty$, $I_k \stackrel{\text{def}}{=} [x_{k-1}, x_k)$, $(k = 1,2,\ldots,r)$, $x_0 \stackrel{\text{def}}{=} -\infty$, $x_r \stackrel{\text{def}}{=} +\infty$ felosztást. A két minta ellenére elég most egyetlen intervallumrendszer, hiszen a homogenitás fennállása esetén ugyanaz a két változó értékkészlete. A minták és a felosztás segítségével definiáljuk az $A_k \stackrel{\text{def}}{=} \{\omega \mid \xi(\omega) \in I_k\}$, $B_k \stackrel{\text{def}}{=} \{\omega \mid \eta(\omega) \in I_k\}$ $(k = 1,2,\ldots,r)$. Mindkét eseményrendszer teljes. A nullhipotézis fennállása esetén a két minta egyesítése is statisztikai minta. Tekintsük most a υ_i és λ_i gyakoriságokat, ahol $\upsilon_i \stackrel{\text{def}}{=} k$, ha k db α indexre teljesül a mintában, hogy $\xi_\alpha \in I_i$, és $\lambda_i \stackrel{\text{def}}{=} j$, ha j db α indexre teljesül a mintában, hogy $\eta_\alpha \in I_i$.

Nyilvánvalóan: $\sum_{i=1}^{r} \upsilon_i = n$, $\sum_{i=1}^{r} \lambda_i = m$. Ha a nullhipotézis igaz, akkor fenn kell állnia a $H_0^* : P(A_i) = P(B_i) = p_i$ feltételezésnek is, ami tehát csak szükséges, de nem elegendő feltétele H_0-nak. A döntési eljárást úgy szerkesztjük meg, hogy H_0^*-ra vonatkozzék, de annak eredményét H_0-ra is átörökítjük. Ha ugyanis H_0^* nem igaz, akkor H_0 sem lehet igaz.
H_0^* átfogalmazható úgy, hogy az egy az 1,2,...,r értékeket felvevő, a p_i diszkrét eloszláshoz tartozó valószínűségi változó illeszkedésére vonatkozik, melyhez n+m elemszámú megfigyelés-sorozat tartozik. A p_i értékeket nem ismerjük, de a mintákból a relatív gyakoriságokkal becsülni tudjuk: $p_i \approx \hat{p}_i = \dfrac{\upsilon_i + \lambda_i}{n+m}$. Összesen r-1 becslést alkalmazunk, mivel az r-edik eloszláselem a többiből számolható. Tehát megint becsléses illeszkedésvizsgálatról van szó. A tiszta illeszkedésvizsgálatnál elmondottak szerint a $T_n^* \stackrel{\text{def}}{=} \sum_{i=1}^{r} \dfrac{(\upsilon_i - n \cdot p_i)^2}{n \cdot p_i}$ és a $T_m^{**} \stackrel{\text{def}}{=} \sum_{i=1}^{r} \dfrac{(\lambda_i - m \cdot p_i)^2}{m \cdot p_i}$ statisztikák aszimptotikusan r-1 szabadságfokú χ^2-eloszlást követnek, ha H_0^* igaz. Az összegük viszont akkor 2r-2 szabadságfokú χ^2-eloszlású lesz: $T_n^* + T_m^{**} \xrightarrow{e} \chi^2_{2r-1}$. Az összesen r-1 db paraméterbecslés miatt azonban, ahogy arra a becsléses illeszkedésvizsgálatnál utaltunk, a szabadságfokot r-1 -gyel csökkenteni kell:

$$\sum_{i=1}^{r} \frac{(\upsilon_i - n \cdot \hat{p}_i)^2}{n \cdot \hat{p}_i} + \sum_{i=1}^{r} \frac{(\lambda_i - m \cdot \hat{p}_i)^2}{m \cdot \hat{p}_i} = \sum_{i=1}^{r} \frac{\left(\upsilon_i - n \cdot \dfrac{\upsilon_i + \lambda_i}{n+m}\right)^2}{n \cdot \dfrac{\upsilon_i + \lambda_i}{n+m}} + \sum_{i=1}^{r} \frac{\left(\lambda_i - m \cdot \dfrac{\upsilon_i + \lambda_i}{n+m}\right)^2}{m \cdot \dfrac{\upsilon_i + \lambda_i}{n+m}} =$$

$$= n \cdot m \sum_{i=1}^{r} \frac{\left(\frac{\upsilon_i}{n} - \frac{\lambda_i}{m}\right)^2}{\upsilon_i + \lambda_i} \xrightarrow{e} \chi^2_{r-1}.$$

A H_0^* hipotézis eldöntéséhez, tehát az r-1 szabadságfokú χ^2-eloszlás táblázatból meghatározzuk azt a K_ε kritikus értéket, melyre $1-\varepsilon = P\left(\chi^2_{r-1} < K_\varepsilon\right)$ teljesül. Ezek után a H_0^*-ot, így H_0-t is elfogadjuk, ha az adott realizálódott mintánál $n \cdot m \sum_{i=1}^{r} \frac{\left(\frac{\upsilon_i}{n} - \frac{\lambda_i}{m}\right)^2}{\upsilon_i + \lambda_i} < K_\varepsilon$ teljesül.

Feladat El akarjuk dönteni, hogy a Tisza Szegedenél mért évi maximális vízállásai ugyanazt az eloszlást követték-e 1876-1925 között, mint 1926-1975 között, vagy pedig megváltozott a Tiszának ez a tulajdonsága. Adatainkat a mellékelt táblázatban közöljük.

Megoldás:

A Tisza maximális vízállása (V)	Gyakoriság az első 50 évben (λ_i)	Gyakoriság a második 50 évben (μ_i)	$\frac{\left(\lambda_i - \mu_i\right)^2}{\lambda_i + \mu_i}$
V<5 m	5	10	1,6667
5 m ≤V<6 m	11	11	0
6 m ≤V<7 m	13	13	0
7 m ≤V<8 m	13	10	0,3913
8 m ≤V	8	6	0,2857
összesen	50	50	2,3437

A feladatot homogenitásvizsgálattal oldjuk meg. A táblázatban a vízmagasság adatokat r=5 csoportba adtuk, tehát a szabadságfok r-1=4. A 90%-hoz tartozó kritikus érték most $K_{0,1} = 7{,}779$, ami jóval nagyobb, mint a táblázat jobb alsó cellájában leolvasható számított érték. A két terminusban tehát azonosnak tekinthető a vízállásmaximumok eloszlása.

Ellenőrző kérdések és gyakorló feladatok

1. Mi az elsőfajú és a másodfajú hiba definíciója?
2. Hogyan fogalmazható meg a nullhipotézis az egymintás u-próbánál?
3. Mi a próbastatisztika a kétmintás u-próbánál?
4. Mik a feltételei a t-próba alkalmazásának?
5. Milyem próbát kell elvégezni a kétmintás t-próba elvégzése előtt?
6. Hogyan osztjuk fel a minta értékkészletét csoportokba tiszta illeszkedésvizsgálatnál a diszkrét és folytonos esetekben?
7. Milyen becsléssel kell közelítenünk az eloszlás ismeretlen paramétereit becsléses illeszkedésvizsgálatnál?
8. Hogyan osztjuk fel a minták közös értékkészletét homogenitásvizsgálatnál?

9. Hogyan számoljuk a szabadságfokot függetlenségvizsgálatnál?
10. Melyik állítás igaz, melyik hamis?
a. Az elsőfajú hiba akkor keletkezik, amikor elfogadom a nullhipotézist, holott nem igaz.
b. Az elsőfajú hiba akkor keletkezik, amikor elvetem a nullhipotézist, holott igaz.
c. Az elsőfajú hiba valószínűségének nagyságát a próba végrehajtója szabályozhatja.
d. Az egymintás u-próba alkalmazásakor fel kell tenni, hogy ismert a normális eloszlású minta szórása.
e. Az egymintás t-próba alkalmazásakor fel kell tenni, hogy ismert a normális eloszlású minta szórása.
f. Az F-próbával ellenőrizhetjük, hogy teljesülnek-e a kétmintás t-próba feltételei.
g. Ha az F-próbával elvetjük a minták szórásainak egyezésére vonatkozó feltevést, akkor t-próba helyett Welch-próbát kell alkalmazni.
h. A kétmintás t-próba, mivel a normális eloszlás paraméterei most nem ismertek, nemparaméteres próba.
i. A kétmintás t-próbához a szabadságfokot úgy számolhatjuk ki, hogy a két minta elemszámai összegéből kettőt levonunk.
j. A kétmintás t-próbához a szabadságfokot úgy számolhatjuk ki, hogy a két minta elemszámai összegéből egyet levonunk.
k. A kétmintás t-próbát akkor alkalmazhatjuk csak, ha a két normális eloszlásból származó statisztikai minta ismeretlen szórásai egyenlőeknek tekinthetők.
l. A kétmintás t-próba a két normális eloszlású minta ismeretlen szórásainak egyezését ellenőrzi.
m. Az F-próba szabadsági fokai n-1 és m-1, ahol n a nevezőben álló korrigált empirikus szórásnégyzethez tartozó minta elemszáma, m pedig a számlálóhoz tartozó minta elemszáma.
n. Egy statisztikai döntés érvényessége annál erősebb, minél nagyobb elsőfajú hibavalószínűség mellett lehetett a nullhipotézist elfogadni.
o. Egy statisztikai döntés érvényessége annál erősebb, minél kisebb elsőfajú hibavalószínűség mellett lehetett a nullhipotézist elfogadni.
p. Diszkrét eloszlás illeszkedését nem lehet χ^2-próbával ellenőrizni.
q. A becsléses illeszkedésvizsgálatnál a szabadságfok r-1-k, ahol r a kialakított csoportok száma, k pedig a becsült paraméterek száma.
r. A becsléses illeszkedésvizsgálatnál a szabadságfok r-1-k, ahol k a kialakított csoportok száma, r pedig a becsült paraméterek száma.
s. Homogenitásvizsgálatnál két minta eloszlásának azonosságát ellenőrizzük.
t. χ^2-próbáknál a próbastatisztikák eloszlása elméletileg χ^2-eloszlású.
u. χ^2-próbáknál a próbastatisztikák eloszlásfüggvénye a mintaelemszám növekedtével pontonként χ^2-eloszlású eloszlásfüggvényhez konvergál, ha igaz a nullhipotézis.
v. χ^2-próbáknál a szabadságifok függ a mintaelemszámtól.
w. A χ^2-próbák paraméteres próbák.
x. Függetlenségvizsgálathoz a két minta elemeit egyidejű (szinkron) megfigyelésekből kell beszerezni.
11. Meg akarjuk vizsgálni, hogy egy új készítési eljárás javítja-e a beton minőségét, nevezetesen növeli-e a törőszilárdságát. E célból ugyanabból az alapanyagból 12 egyenlő mennyiségű mintát vesznek. Ezeket véletlenszerűen kettéosztva mind a régi, mind az új technológiával 6-6 próba kockát készítenek. A mellékelt táblázat az egyes próbakockák

törésszilárdságait tartalmazza kg/cm^2-ben. Döntésünket 95%-os szinten hozzuk! (A minták normális eloszlásúaknak tekinthetők.

Régi technológia (x_i)	Új technológia (y_i)
300	305
301	317
303	308
288	300
294	314
296	316

12. Döntsünk 90%-os szignifikancia szinten arról, hogy szabályosnak tekinthető-e az a dobókocka, melyet n=1200-szor feldobva az alábbi gyakoriság táblázatot produkálja.

A kockával dobott érték (i)	Az előfordulás gyakorisága (υ_i)
1	184
2	212
3	190
4	208
5	212
6	194
Összesen	n=1200

13. A légi közlekedésben fontos figyelemmel kísérni az utasok átlagos testsúlyának alakulását. Egyrészt, hogy ne terheljék túl a gépet, másrészt ne utazzon a gép fölös kapacitással. Ezért időről időre ellenőrzik, hogy a felnőtt utasok testsúlya nem tér-e el a feltételezettől. A légitársaság a terhelést 78 kg-os átlagos testsúlyra és 11 kg-os szórásra tervezi. A feltételezés ellenőrzése céljából megmérték n=100 véletlenszerűen kiválasztott utas súlyát. A mérések eredményét a mellékelt táblázatban foglaltuk össze. Végezze el az eloszlás normalitására vonatkozó feltételezés ellenőrzését!

Testsúly (kg)	Utasok száma (fő)
- 60	7
61 - 70	16
71 - 80	32
81 - 90	28
91 - 100	13
101-	4
Összesen	100

14. Megvizsgáltak összesen n=460 db csavart, amelyek közül méretre 439 db volt megfelelő. Ez utóbbiak közül szakítószilárdság szempontjából is megfelelt 416 db. A maradék közül 18 db bizonyult selejtesnek szakítószilárdság szempontjából. Vizsgálja meg, hogy a méretre nézve és a szakítószilárdságra nézve megfelelőség (illetve selejtesség) független tulajdonságok-e! A döntést 95%-os szignifikancia szinten hozza!

3. Regresszióanalízis

A feladat két, erősen összefüggő ξ és η valószínűségi változó közötti függvénykapcsolat jellegének, és paramétereinek feltárása. η fogja jelölni a célváltozót, és a ξ a független változót., vagyis feladat olyan f függvény megadása, ahol $\eta \approx f(\xi)$. A függvénykapcsolatot a két változóra vonatkozó $(\xi_1,\eta_1)^T,(\xi_2,\eta_2)^T,\ldots,(\xi_n,\eta_n)^T$ statisztikai minta alapján kell meghatározni. A regresszióanalízis végrehajtásának csak akkor van értelme, ha kimutatható ξ és η között a sztochasztikus összefüggés (pl. el kellett vetni a nullhipotézist függetlenségvizsgálatnál, vagy a minta empirikus korrelációs együtthatója közel van 1-hez). A regresszióanalízis tipikus módszere az, hogy egy jól körülírt többparaméteres függvényhalmazból határozunk meg egy bizonyos függvényt úgy, hogy annak paramétereit a minta segítségével megbecsüljük. Legyen adott tehát az $\mathrm{F}=\{f\}$ függvényosztály. Meghatározandó az az $f^* \in \mathrm{F}$ függvény, ahol $\mathbf{M}(\eta - f^*(\xi))^2 = \min\limits_{\forall f \in F} \mathbf{M}(\eta - f(\xi))^2$. F-et legtöbbször a mintarealizációnak a koordinátarendszerben való ábrázolásával kapott szóródás grafikon alapján lehet meghatározni, de az a változók fizikai tartalmából fakadó "elvárt" típusú függvények halmaza is lehet.

Definíció: Adott a $(\xi_1,\eta_1)^T,(\xi_2,\eta_2)^T,\ldots,(\xi_n,\eta_n)^T$ statisztikai minta és az $\mathrm{F}=\{f(x;a_1,a_2,\ldots,a_k)\}$ k-paraméteres függvényosztály.
A $\sum\limits_{i=1}^{n}(\eta_i - f(\xi_i;a_1,a_2,\ldots,a_k))^2 \to \min\limits_{\forall a_1,a_2,\ldots,a_k}$ szélsőérték feladat megoldásából kapott $a_i^* = a_i^*(\underline{\xi},\underline{\eta})$ $(i=1,2,\ldots,n)$ statisztikákat, az $M(\eta - f^*(\xi))^2 = \min\limits_{\forall f \in F} M(\eta - f(\xi))^2$ regressziós probléma paramétereinek *legkisebb négyzetek módszerével* kapott becsléseinek nevezzük.

Definíció: Tekintsük a ξ és η valószínűségi változókat, és tegyük fel, hogy η-t $\hat{\eta} \stackrel{def}{=} f(\xi)$-vel közelítjük: $\hat{\eta} \approx \eta$. A közelítés jóságának mérésére az $I^2 \stackrel{def}{=} 1 - \dfrac{\mathbf{D}^2(\eta - \hat{\eta})}{\mathbf{D}^2\eta}$ *meghatározottsági együtthatót* használjuk. Adott $(\xi_1,\eta_1)^T,(\xi_2,\eta_2)^T,\ldots,(\xi_n,\eta_n)^T$ statisztikai minta esetén a meghatározottsági együtthatót az $1 - \dfrac{\sum\limits_{i=1}^{n}(\eta_i - f(\xi_i))^2}{\sum\limits_{i=1}^{n}(\eta_i - \overline{\eta}_n)^2}$ ($\approx I^2$) statisztikával közelítjük.

Megjegyzés: I^2 minél közelebb van az 1-hez, annál jobb a regressziós közelítés. Ha I^2 0 közeli érték, vagy negatív, a regressziós illesztés elfogadhatatlan.

Alapvető fontosságú a regressziószámításnak az a speciális esete, amikor F a lineáris függvények halmaza.

3.1 Lineáris regresszió két változó között

Definíció: Legyen ξ és η két adott valószínűségi változó. Az $a^*\xi + b^*$ valószínűségi változó az η-nak a ξ-re vonatkozó *lineáris regressziója,* ha $\mathbf{M}(\eta - a^*\xi - b^*)^2 = \min\limits_{\forall a,b \in R} \mathbf{M}(\eta - a\xi - b))^2$.

Tétel: $a^* = \mathbf{R}(\xi,\eta)\dfrac{\mathbf{D}\eta}{\mathbf{D}\xi}$, $b^* = \mathrm{M}\eta - \mathbf{R}(\xi,\eta)\dfrac{\mathbf{D}\eta}{\mathbf{D}\xi}\mathrm{M}\xi$.

Bizonyítás: Legyen $h(a,b) \overset{\text{def}}{=} \mathbf{M}(\eta - a\xi - b))^2$. A lineáris regresszió meghatározásához ezt a kétváltozós függvényt kell minimalizálni. A minimumhely létezésének szükséges feltétele, hogy:
$\dfrac{\partial h}{\partial a} = -2\mathbf{M}\left[(\eta - a\xi - b)\xi\right] = 0$, $\dfrac{\partial h}{\partial b} = -2\mathbf{M}\left[\eta - a\xi - b\right] = 0$. Innen:
$a\mathbf{M}\xi^2 + b\mathbf{M}\xi = \mathbf{M}\xi\eta$, $a\mathbf{M}\xi + b = \mathbf{M}\eta \Rightarrow b = \mathbf{M}\eta - a\mathbf{M}\xi \Rightarrow a\mathbf{M}\xi^2 + (\mathbf{M}\eta - a\mathbf{M}\xi)\mathbf{M}\xi = \mathbf{M}\xi\eta \Rightarrow$
$\Rightarrow \quad a = \mathbf{R}(\xi,\eta)\dfrac{\mathbf{D}\eta}{\mathbf{D}\xi}$, $b = \mathrm{M}\eta - \mathbf{R}(\xi,\eta)\dfrac{\mathbf{D}\eta}{\mathbf{D}\xi}\mathrm{M}\xi$, $\begin{pmatrix} \mathrm{M}\xi^2 & \mathrm{M}\xi \\ \mathrm{M}\xi & 1 \end{pmatrix}$ pozitív definit , és ez volt az állítás.

A gyakorlatban általában nem ismertek a ξ és η változók momentumai, ezért az elméleti lineáris regressziós összefüggés nem határozható meg. A $(\xi_1,\eta_1)^T, (\xi_2,\eta_2)^T, \ldots, (\xi_n,\eta_n)^T$ statisztikai minta alapján a legkisebb négyzetek módszerével lehet az egyenes paramétereit megadni.

Tétel: Lineáris regresszió esetén az egyenes paramétereinek becslései a legkisebb négyzetek módszerével: $a^* = \hat{R}_n \dfrac{s_\eta}{s_\xi}$, $b^* = \overline{\eta}_n - \hat{R}_n \dfrac{s_\eta}{s_\xi}\overline{\xi}_n$, ahol $\hat{R}_n = \dfrac{\sum\limits_{i=1}^{n}(\eta_i - \overline{\eta}_n)(\xi_i - \overline{\xi}_n)}{\sqrt{\sum\limits_{i=1}^{n}(\eta_i - \overline{\eta}_n)^2 \sum\limits_{i=1}^{n}(\xi_i - \overline{\xi}_n)^2}}$ az empirikus korrelációs együttható, $s_\eta = \sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}(\eta_i - \overline{\eta}_n)^2}$, $s_\xi = \sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}(\xi_i - \overline{\xi}_n)^2}$ az empirikus szórások, és $\overline{\xi}_n$, $\overline{\eta}_n$ az átlagstatisztikák.

Bizonyítás: A tétel állítása könnyen belátható, ha az előző tétel bizonyítását megismételjük a $h(a,b)=\sum_{i=1}^{n}(\eta_i-a\xi_i-b)^2$ kétváltozós függvénnyel.

Megjegyzések: Látható, hogy az empirikus lineáris regresszió együtthatói az elméleti regressziós egyenes együtthatóitól annyiban különböznek, hogy a képletekben az elméleti momentumok helyett a mintából számolt megfelelő empirikus momentumok állnak.

Tétel: *(Gauss-Markov tétel)*
Ha $\eta_i = a\cdot x_i + b + \varepsilon_i\ (i=1,2,\dots,n)$, ahol az ε_i teljesen független valószínűségi változók, és $\mathbf{M}\varepsilon_i = 0$, $\mathbf{D}^2\varepsilon_i = \sigma^2$, akkor az a , b együtthatók legkisebb négyzetek módszerével kapott becslései torzítatlanok, és az összes lineáris becslés közül minimális szórással rendelkeznek.

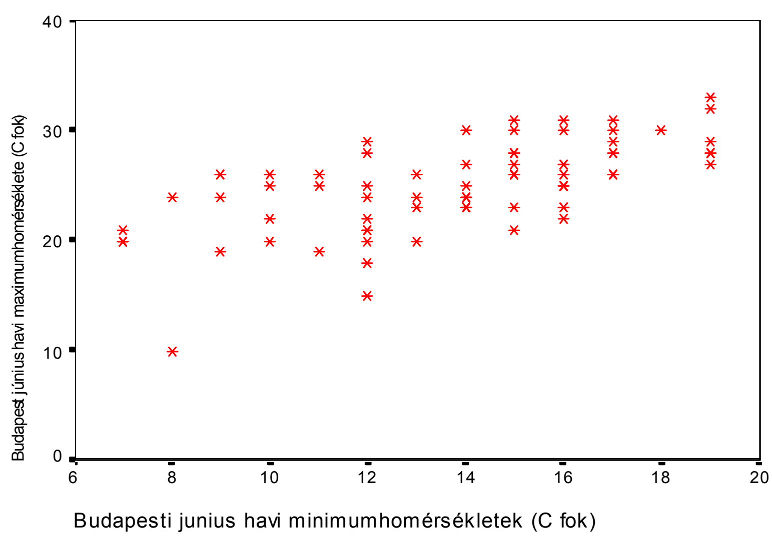
Megjegyzés: A legkisebb négyzetek módszere a legjobb torzítatlan becslést adja, ami angolul: *b*est *l*inear *u*nbaised *e*stimation = BLUE.

Feladat Jelölje ξ a június 8-i budapesti minimum hőmérsékleteket, és legyen a feladat az η június 8-i budapesti maximum hőmérséklet becslése, a minimum hőmérséklet alapján. Tegyük fel, hogy a két valószínűségi változó kapcsolata lineáris regresszióval leírható. Becsüljük meg az adott minta alapján a $\eta\approx a\xi+b$ lineáris összefüggés együtthatóit!

év (i)	max (x_i)	min (y_i)	év (i)	max (x_i)	min (y_i)
1901	28°C	17°C	1941	24°C	14°C
1902	24°C	13°C	1942	31°C	15°C
1903	21°C	12°C	1943	23°C	13°C
1904	25°C	16°C	1944	15°C	12°C
1905	26°C	16°C	1945	33°C	19°C
1906	20°C	10°C	1946	31°C	17°C
1907	18°C	12°C	1947	20°C	12°C
1908	19°C	11°C	1948	22°C	16°C
1909	23°C	16°C	1949	30°C	14°C
1910	27°C	16°C	1950	30°C	18°C
1911	23°C	14°C	1951	25°C	16°C
1912	31°C	16°C	1952	23°C	15°C
1913	27°C	15°C	1953	28°C	19°C
1914	25°C	10°C	1954	22°C	12°C
1915	30°C	17°C	1955	28°C	19°C
1916	25°C	11°C	1956	28°C	17°C
1917	28°C	15°C	1957	28°C	15°C
1918	26°C	11°C	1958	26°C	9°C
1919	25°C	12°C	1959	28°C	15°C
1920	20°C	7°C	1960	27°C	16°C
1921	24°C	9°C	1961	26°C	15°C
1922	26°C	16°C	1962	10°C	8°C
1923	21°C	7°C	1963	26°C	17°C

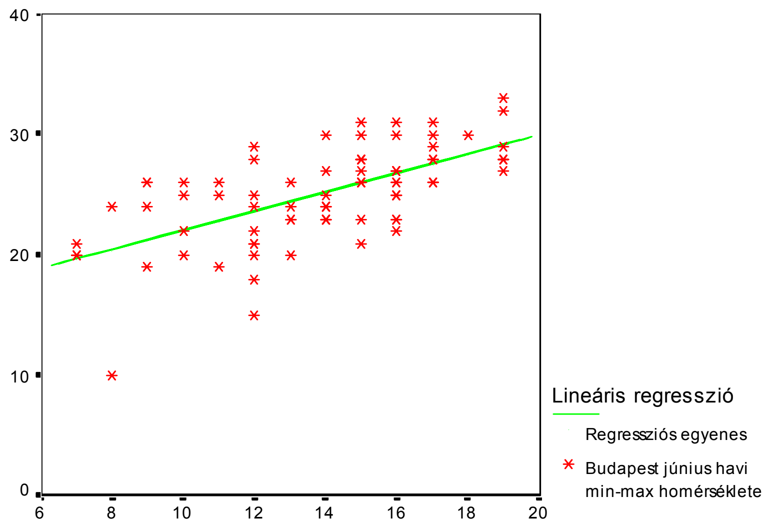
1924	24°C	12°C	1964	27°C	19°C
1925	24°C	8°C	1965	24°C	14°C
1926	23°C	14°C	1966	28°C	15°C
1927	26°C	10°C	1967	28°C	17°C
1928	26°C	15°C	1968	30°C	15°C
1929	29°C	17°C	1969	21°C	15°C
1930	26°C	9°C	1970	23°C	16°C
1931	25°C	16°C	1971	26°C	17°C
1932	22°C	10°C	1972	29°C	19°C
1933	20°C	7°C	1973	24°C	14°C
1934	23°C	14°C	1974	19°C	9°C
1935	29°C	12°C	1975	21°C	12°C
1936	20°C	13°C	1976	25°C	14°C
1937	30°C	16°C	1977	26°C	15°C
1938	27°C	14°C	1978	28°C	19°C
1939	28°C	12°C	1979	32°C	19°C
1940	26°C	13°C	1980	23°C	13°C

A két változó közötti kapcsolatot az alábbi pontgrafikonon szemlélhetjük meg:



A minták átlagai $\overline{x}_{80} = 25{,}1°\,C$ és $\overline{y}_{80} = 13{,}95°\,C$, a korrigált empirikus szórás statisztikák pedig $s^*_{80} = 3{,}94°\,C$ illetve $\sigma^*_{80} = 3{,}14°\,C$. A két minta között számolt empirikus korrelációs együttható: $\hat{r} = \dfrac{\sum_{i=1}^{80}\left(x_i - \overline{x}_{80}\right)\cdot\left(y_i - \overline{y}_{80}\right)}{\sqrt{\sum_{i=1}^{80}\left(x_i - \overline{x}_{80}\right)^2 \sum_{i=1}^{80}\left(y_i - \overline{y}_{80}\right)^2}} \approx 0{,}64$. Ezek alapján a regressziós egyenes meredekségére $\hat{a} = r\dfrac{\sigma^*_{80}}{s^*_{80}} \approx 0{,}8$, a konstans tagra pedig $\hat{b} = \overline{y}_{80} - \hat{a}\cdot\overline{x}_{80} \approx 13{,}94$ adódik.

A pontgrafikonon ábrázolva az egyenest, szemléltethetjük az összefüggést:



A regressziós illeszkedés jósága I^2 =0,397, ami nem mutat erős lineáris összefüggést.

3.2 Polinomiális regresszió

Definíció: Amikor az $F=\{p_n(x)=a_0+a_1x+\cdots+a_nx^n\}$ függvényosztály a legfeljebb n-edrendű polinomosztály, az $\mathbf{M}(\eta-f^*(\xi))^2=\min\limits_{\forall f\in F}\mathbf{M}(\eta-f(\xi))^2$ minimumfeladat megoldását *polinomiális regressziós illesztésnek* nevezzük.

Tétel: Az elméleti polinomiális regressziós görbe együtthatóit az

$$\begin{pmatrix} 1 & M\xi & \cdots & M\xi^n \\ M\xi & M\xi^2 & \cdots & M\xi^{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ M\xi^i & M\xi^{i+1} & \cdots & M\xi^{i+n} \\ \vdots & \vdots & \ddots & \vdots \\ M\xi^n & M\xi^{n+1} & \cdots & M\xi^{2n} \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} M\eta \\ M\eta\xi \\ \vdots \\ M\eta\xi^i \\ \vdots \\ M\eta\xi^n \end{pmatrix}$$

lineáris egyenletrendszer megoldásával kaphatjuk meg. Ennek mindig van megoldása, hiszen az együtthatómátrix szimmetrikus és pozitív definit.

Bizonyítás: A feladatot a $h(a_0,a_1,\ldots,a_n)\overset{\text{def}}{=}\mathbf{M}(\eta-(a_0+a_1\xi+\ldots+a_n\xi^n))^2$ n+1 változós függvény minimumhelyének megkeresésével oldhatjuk meg:

$$\frac{\partial h(a_0,a_1,\ldots,a_n)}{\partial a_i}=-2M([\eta-(a_0+a_1\xi+\cdots+a_n\xi^n)]\cdot\xi^i)=0\ (i=0,1,2,\ldots,n)\Rightarrow$$

$\Rightarrow \sum\limits_{j=0}^{n} a_j\cdot M\xi^{i+j}=M\eta\xi^i \Rightarrow$következik az állítás.

A tapasztalati polinomiális görbe együtthatóinak meghatározását a $(\xi_1,\eta_1)^T,(\xi_2,\eta_2)^T,\ldots,(\xi_N,\eta_N)^T$ statisztikai minta segítségével az

$$\begin{pmatrix} 1 & \frac{1}{N}\sum_{j=1}^{N}\xi_j & \cdots & \frac{1}{N}\sum_{j=1}^{N}\xi_j^{\,n} \\ \frac{1}{N}\sum_{j=1}^{N}\xi_j & \frac{1}{N}\sum_{j=1}^{N}\xi_j^{\,2} & \cdots & \frac{1}{N}\sum_{j=1}^{N}\xi_j^{\,n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{N}\sum_{j=1}^{N}\xi_j^{\,i} & \frac{1}{N}\sum_{j=1}^{N}\xi_j^{\,i+1} & \cdots & \frac{1}{N}\sum_{j=1}^{N}\xi_j^{\,i+n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{N}\sum_{j=1}^{N}\xi_j^{\,n} & \frac{1}{N}\sum_{j=1}^{N}\xi_j^{\,n+1} & \cdots & \frac{1}{N}\sum_{j=1}^{N}\xi_j^{\,2n} \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \frac{1}{N}\sum_{j=1}^{N}\eta_j \\ \frac{1}{N}\sum_{j=1}^{N}\eta_j\xi_j \\ \vdots \\ \frac{1}{N}\sum_{j=1}^{N}\eta_j\xi_j^{\,i} \\ \vdots \\ \frac{1}{N}\sum_{j=1}^{N}\eta_j\xi_j^{\,n} \end{pmatrix}$$

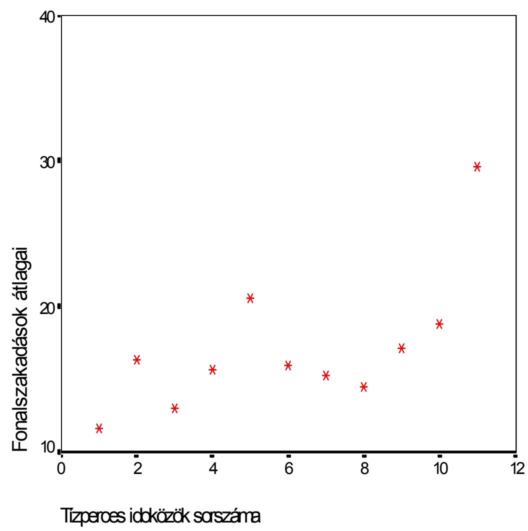
lineáris egyenletrendszer megoldásából kapjuk. Ehhez úgy jutunk el, hogy a $\hat{h}(a_0,a_1,\ldots,a_n)\stackrel{\text{def}}{=}\frac{1}{N}\sum_{j=1}^{N}(\eta_j-(a_0+a_1\xi_j+\cdots+a_n\xi_j^{\,n}))^2$ függvény minimumhelyét meghatározzuk, hasonlóan, mint ahogy azt az előző tételben tettük.

Megjegyzés: Nyilvánvalóan az N mintaelemszámnak jóval nagyobbnak kell lennie, mint az n-nek, az illesztendő polinom rendjének.

Feladat Fonodai felvételi lapok összesítésével akarják megállapítani az eltelt tíz perces időközök sorszáma (x_i) és a fonalszakadások átlagai (y_i) közötti összefüggést. Keressünk parabolikus összefüggést a változók között!

x_i	y_i
1	11,7
2	16,4
3	13,1
4	15,7
5	20,6
6	16,0
7	15,3
8	14,6
9	17,2
10	18,9
11	29,7

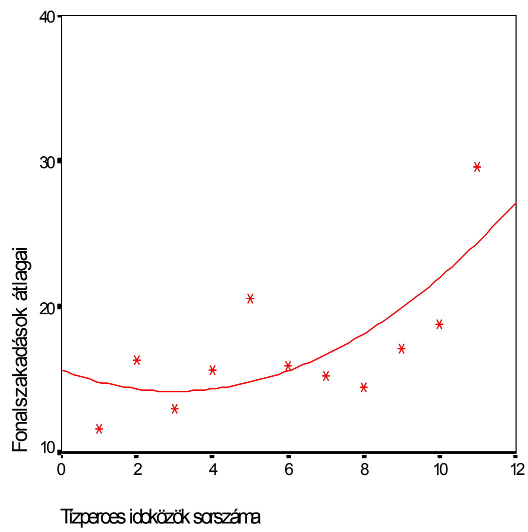
A változók közötti pontszóródás grafikon:



Az $a + bx + cx^2 = y$ összefüggés együtthatóit a

$$\begin{array}{lllllll} 189{,}2 & = & 11a & + & 66b & + & 506c \\ 1239{,}7 & = & 66a & + & 506b & + & 4356c \\ 10098 & = & 506a & + & 4356b & + & 39974c \end{array}$$

egyenletrendszer megoldásából kapjuk: a=15,76, b=-1,0175, c=0,164. A regressziós parabolát együtt ábrázolva az adatokkal szemléltethetjük az illeszkedést:



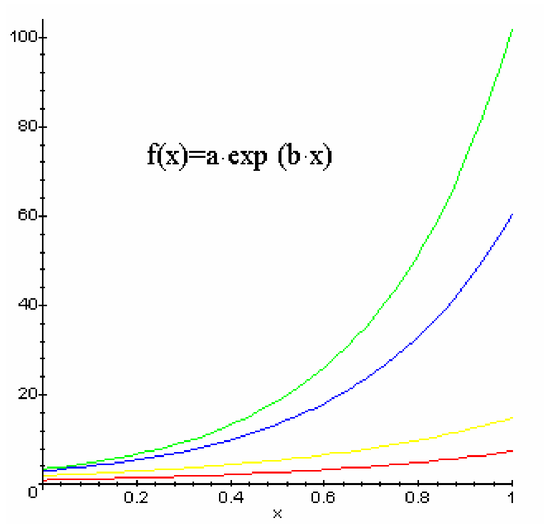
Az illeszkedés jósága: $I^2 = 0{,}543$.

3.3 Lineárisra visszavezethető kétparaméteres regressziós összefüggések keresése

Ha a lineáris regresszió feltételei valahol sérülnek, vagy rossz illesztést kapunk, a függő és a független változók transzformációjával kell megpróbálkozni. A transzformált input adatokon azután már lineáris regressziós elemzést hajtunk végre, de ez az eredeti adatoknál már nem lineáris összefüggést fog magyarázni. Az inverz leképezés és a regressziós együtthatók segítségével képezhetők azok a paraméterek, amelyekkel a kapcsolatot leíró függvény felírható. Tehát, ha az $F=\{f(x;a,b)\}$ függvényosztály kétparaméteres, és találhatók olyan g,h,k_1,k_2 függvények, hogy $y=f(x;a,b)\Leftrightarrow g(y)=k_1(a,b)\cdot h(x)+k_2(a,b)$ teljesül.

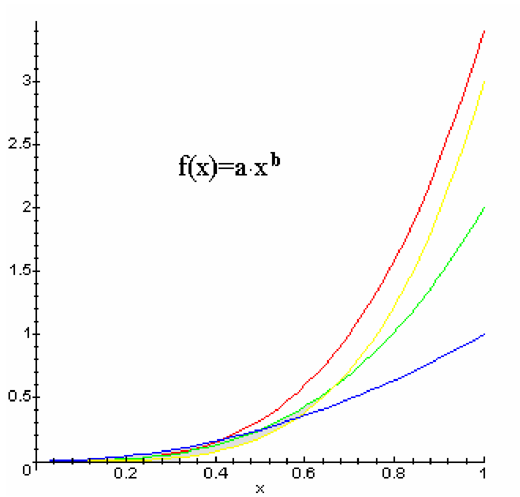
Ezután az $\mathbf{M}(\eta-f(\xi;a^*,b^*))^2=\min\limits_{\forall f\in F}\mathbf{M}(\eta-f(\xi;a,b))^2$ feladat helyett az $\mathbf{M}(g(\eta)-k_1^*\cdot h(\xi)-k_2^*)^2=\min\limits_{\forall k_1,k_2}\mathbf{M}(g(\eta)-k_1\cdot h(\xi)-k_2)^2$ lineáris regressziós feladatot oldjuk meg. Végül $a^*\approx k_1^{-1}(k_1^*,k_2^*)$, $b^*\approx k_2^{-1}(k_1^*,k_2^*)$. Általában más eredményeket kapunk, mintha az eredeti függvényen hajtottuk volna végre a legkisebb négyzetek módszerével a paraméterbecslést. Viszont az eredeti problémánál, nem biztos, hogy a stacionárius helyekre kapott (sokszor transzcendens) egyenletet meg tudnánk oldani. A továbbiakban megadunk néhány lehetőséget nemlineáris kapcsolatnak a lineáris regresszió segítségével való megadására.

- $y=a\cdot e^{b\cdot x}$ **exponenciális függvénykapcsolat:**



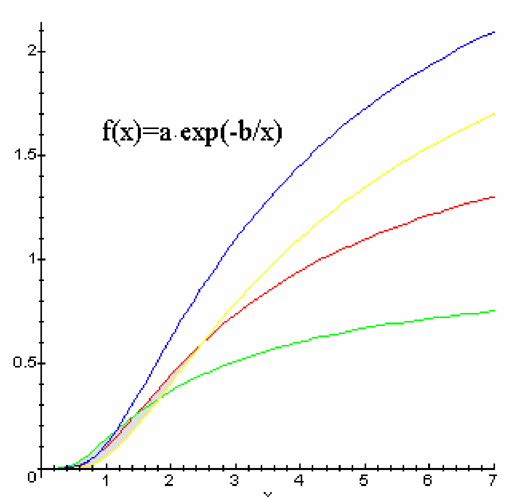
Az egyenlet két oldalát logaritmizálva már lineáris összefüggést kapunk ln y és x között: $y^*=\ln y=b\cdot x+\ln a=k_1\cdot x+k_2$. Ilyenkor a $(\xi_1,\ln\eta_1),(\xi_2,\ln\eta_2),\ldots,(\xi_n,\ln\eta_n)$ transzformált mintára illesztünk egyenest. A kapott k_1^* és k_2^* együtthatókból az $a=e^{k_2^*}$ és $b=k_1^*$ transzformációval kapjuk meg az eredeti összefüggés paramétereit.

- $y = a \cdot x^b$ **hatványfüggvény kapcsolat:**



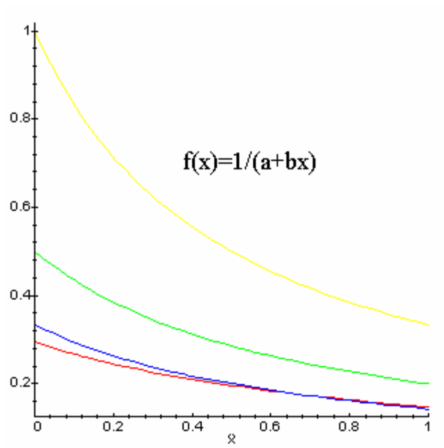
A lineáris kapcsolatot a logaritmizálás után most ln y és ln x között kell megadni: $y^* = \ln y = b \cdot \ln x + \ln a = k_1 \cdot x^* + k_2 \Rightarrow b = k_1$, $a = e^{k_2}$.

- $y = a \cdot e^{\frac{-b}{x}}$ **Arrhenius függvénykapcsolat:**



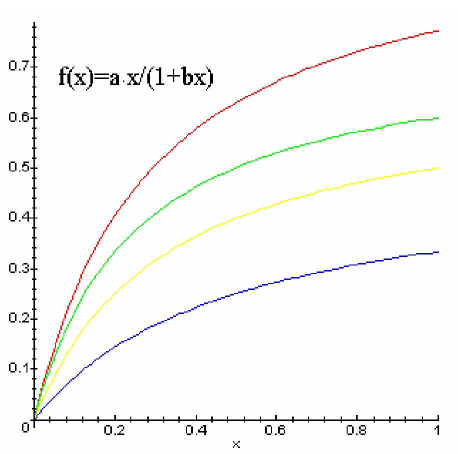
Logaritmizálás után: $y^* = \ln y = -b \cdot \frac{1}{x} + \ln a = k_1 \cdot x^* + k_2$ az ln y és x reciproka között lép fel a lineáris kapcsolat. ($b = -k_1$, $a = e^{k_2}$).

- $y = \dfrac{1}{a + b \cdot x}$ **reciprok függvénykapcsolat:**



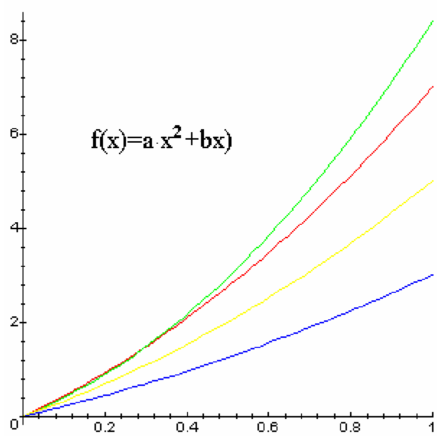
Itt most y reciproka és x között kell a lineáris regressziót kiszámolni.

- $y = \dfrac{a \cdot x}{1 + b \cdot x}$ **racionális törtfüggvény kapcsolat:**



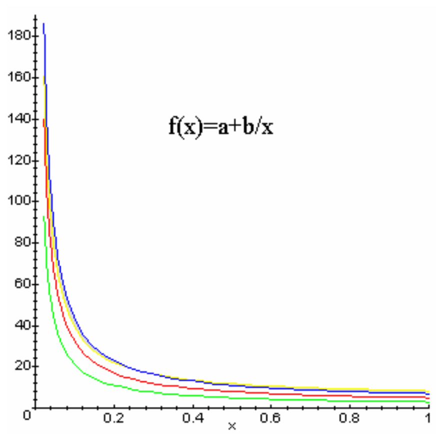
Most az egyenlet két oldalának reciprokát képezzük : $y^* = \dfrac{1}{y} = \dfrac{1}{a} \cdot \dfrac{1}{x} + \dfrac{b}{a} = k_1 \cdot x^* + k_2 \Rightarrow a = \dfrac{1}{k_1}, b = \dfrac{k_2}{k_1}$. , és a reciprokértékek között keresünk lineáris regressziót.

- $y = a \cdot x^2 + b \cdot x$ **kvadratikus függvénykapcsolat:**



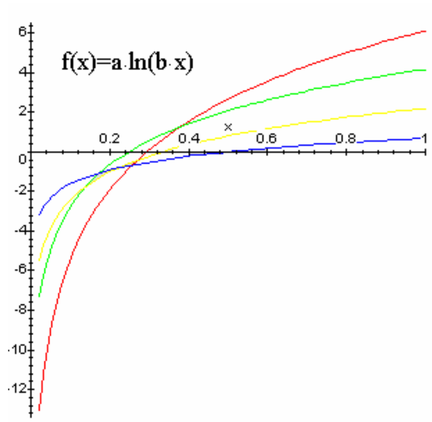
Ekkor ha x-szel átosztunk : $y^* = \frac{y}{x} = a \cdot x + b$, máris lineáris az összefüggés y/x és x között.

- $y = a + \frac{b}{x}$ **hiperbolikus függvénykapcsolat:**



Ez eleve lineáris összefüggés y és 1/x között.

- $y = a \cdot \ln(b \cdot x) = a \cdot \ln b + a \cdot \ln x$ **a logaritmikus függvénykapcsolat:**



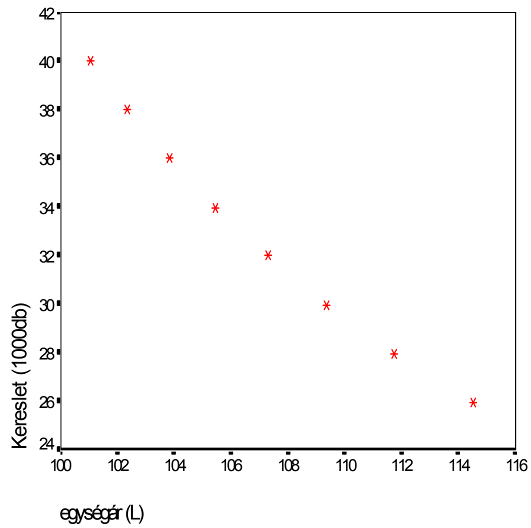
Ez lineáris kapcsolat y és ln x között.

Feladat Egy árérzékeny terméknél a piackutató megfigyelte a kereslet (y_i) és az egységár (x_i) összefüggését. Adjuk meg az $y = a + \frac{b}{x}$ hiperbolikus függvénykapcsolatot a legkisebb négyzetek módszerével!

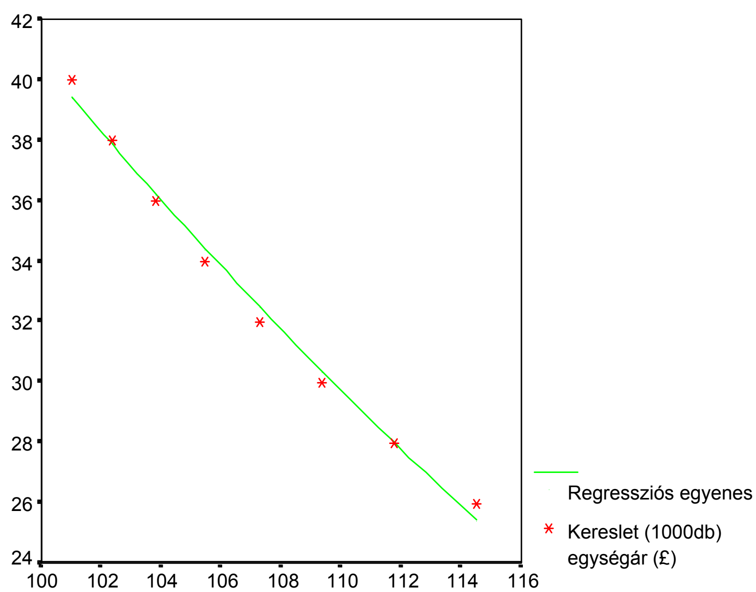
Az adatokat az alábbi táblázatba foglaltuk bele:

Kereslet (y_i)	Egységár (x_i)
26	114,51
28	111,76
30	109,38
32	107,30
34	105,46
36	103,83
38	102,37
40	101,05

A változok közötti pontszóródási grafikon:



Most az $\left(\frac{1}{x_i}, y_i\right)$ pontok között kell lineáris kapcsolatot keresni. A keresett $y = a + \frac{b}{x}$ összefüggés együtthatóinak becsléseire a legkisebb négyzetek módszerével az $\hat{a} = -79{,}529$, $\widehat{b} = 12015{,}8$ becslések adódnak. Az illeszkedés most nagyon jó, mert $I^2 = 0{,}992$. A pontszóródást együtt ábrázolva a regressziós hiperbolával:



Ellenőrző kérdések és gyakorló feladatok

1. Mit nevezünk a regressziós összefüggés paraméterei legkisebb négyzetek módszerével kapott becsléseinek?
2. Mikor beszélünk lineáris regresszióról?
3. Milyen függvénykapcsolatok regressziós vizsgálatát lehet visszavezetni lineáris regresszióra?

4. Mivel mérjük a regressziós illeszkedés jóságát?
5. Melyik állítás igaz, és melyik hamis?
    a. Kétdimenziós összetartozó minták között a legjobb lineáris összefüggés mindig megadható.
    b. Az elméleti regressziós egyenes együtthatója arányos a két változó korrelációs együtthatójával.
    c. A legkisebb négyzetek módszerével kapjuk a regressziós együtthatókra a legkisebb szórású torzítatlan becsléseket.
    d. Ha két valószínűségi változó független, akkor jó lineáris regresszió várható közöttük.
    e. A változók között minél erősebb a korreláció, annál jobb lineáris összefüggés várható.
6. Adjuk meg a lineáris regressziót a csapadék (x_i) és a talajvízállás (y_i) között a mellékelt minta pár feldolgozásával!

Talajvízszint (y_i) mm	Csapadék (x_i) cm
1,25	10,36
1,40	8,94
2,13	13,21
1,19	15,80
1,65	11,18
1,89	13,64
1,68	19,53
1,77	24, 56
1,28	11,48
1,16	7,77
0,94	11,30
3,69	28,13
3,51	30,18
3,14	23,14
1,22	15,88
2,29	19,76
4,42	35,36
2,90	25,40
átlag 2,1	átlag 18,1

7. Az alábbi táblázatban a magyarországi szén-(x) és villamosenergia-termelés (y) előző évihez mért százalékos változásai (az ún. láncindexek) láthatók. Számolja ki és értékelje a lineáris regressziót!

év	széntermelési láncindex (x)	villamos energia láncindex (y)
1950	9,4	19,0
1951	13,3	16,8
1952	14,5	19,9
1953	9,3	10,0
1954	1,4	4,4
1955	3,8	12,4

1956	-8,4	-4,2
1957	0,3	4,8
1958	12,3	18,9
1959	7,2	9,1

8. Egy vezetőn ismeretlen erősségű áram folyik. Az áramerősséget (x) 1,2,3,5,7,10 amperrel megnövelve, a felmelegedés (y) rendre 11,12,15,22,34,61 °C-nak adódott. Adjuk meg a legjobb parabolikus összefüggést az áramerősség és a felmelegedés között!

FÜGGELLÉK

Válaszok és megoldások

I. fejezet. VALÓSZÍNŰSÉGSZÁMÍTÁS

1. Kombinatorika

1. n! **2.** Egy k elemű részhalmaza elemeinek egy permutációját. **3.** n^k **4.** n!=1·2·3·...·n

5. $\binom{n}{k}=\frac{n!}{k!(n-k)!}$ **6.** Az n elem olyan k hosszúságú sorrendjét, ahol ismétlődések is előfordulhatnak. **7.** a.-I, b-H, c-I, d-H, e-I, f-I, g-H, h-I, i-I, j-I,k-H. **8.** Ismétléses permutációval: $\frac{20!}{2!2!4!}$ **9.** Kombinációval: $\binom{5}{3}\binom{85}{2}$ **10.** $\binom{13}{10}2^3\cdot 3$ **11.** Ismétléses variációval: $2+2^2+2^3+\ldots+2^{10}=2^{11}-2$ **12.** a. $15+\binom{15}{2}2$ b. $15+\binom{15}{2}$ **13.** Ismétléses kombinációval. A golyókhoz választjuk a dobozt ismétléssel. $\binom{3+10-1}{10}=66$ **14.** Ismétléses kombinációval. $\binom{6+10-1}{10}=\binom{15}{10}$ **15.** Ismétlés nélküli variáció: 5·4·3=6

16. Kombinációval: $\binom{30}{5}\binom{25}{4}\binom{21}{2}\binom{19}{2}\binom{17}{2}\binom{15}{2}\binom{13}{2}\binom{11}{1}\binom{10}{1}\binom{9}{1}$ **17.** A biominális tétellel: $(a+b)^n=\sum_{k=1}^{n}\binom{n}{k}a^k b^{n-k}$, a. a=b=1 , b. a=-1, b=1, c. a=-2, b=1 .

2. A valószínűségszámítás alapfogalmai és axiómarendszere

7. A , C , E , F események. Az E esemény a lehetetlen, F pedig a biztos esemény. **8.** $C\subset A, C\subset B$ **9.** $AE=CE=DE=\varnothing$ **10.** a-H,b-I,c-I,d-I,e-I,f-H,g-I,h-H,i-H,j-H,k-H,l-I,m-I,n-H,o-H,p-H,q-I,r-I,s-H,t-I,u-H,v-I,w-I,x-H

11. $P(\overline{A}\,\overline{B})=1-P(A+B)=1-P(A)-P(B)+P(AB)=0{,}63$ **12.** $P(A|B)=\frac{P(A\,\overline{B})}{P(\overline{B})}=$ $=\frac{P(A)-P(AB)}{1-P(B)}=\frac{P(A+B)-P(B)}{1-P(B)}=0{,}6$ **13.** $A=A_1A_2\cdots A_n$, $B=A_1+A_2+\cdots+A_n$, $C=\sum_{i=1}^{n}A_i\prod_{i\neq j}A_j$, D=$\prod_{i=1}^{n}A_i+\prod_{i=1}^{n}\overline{A}_i$ **14.** $P(AB)+P(\overline{A}B)+P(A\overline{B})+P(\overline{A}\,\overline{B})=1$, legyen P(AB)=x+0,25 , $P(\overline{A}B)=y+0{,}25$, $P(A\overline{B})=z+0{,}25$ és $P(\overline{A}\,\overline{B})=v+0{,}25$. Mivel x+y+z+u=0,

$$(x+0{,}25)^2+(y+0{,}25)^2+(z+0{,}25)^2+(u+0{,}25)^2=x^2+y^2+z^2+u^2+\frac{x+y+z+u}{4}+0{,}25\geq 0{,}25.$$

15. a. , b. a C eseményt jelent, c. a $\overline{B}$, azaz nem a sötéttel játszó játékos nyer. **16.** $B=A_6$, vagyis a találat belül lesz az R_6 sugarú körön. C=A_2, D=A_3.

17. $P(\overline{A}\,\overline{B})=1-P(A+B)=1-P(A)-P(B)+P(AB)=P(AB)$ **18.** $P(A\,\overline{B}+\overline{A}B)=$ $=P(A\overline{B})+P(\overline{A}B)=P(A)-P(AB)+P(B)-P(AB)$. **19.** P(A+B)=1, P(AB)=0,2P(B)=0,5P(A), így

1=P(A+B)=P(A)+P(B)-P(AB)=P(A)+2,5P(A)-0,5P(A)=3P(A), azaz P(A)=1/3 és P(B)=0,5.

20. P(AB)=P(A)P(B|A)=2/9, P(B)= P(AB)/P(A|B)=1/3, így P(A+B)=2/3+1/3-2/9=7/9
$P(\overline{A}|\overline{B}) = P(\overline{A}\,\overline{B})/P(\overline{B}) = (1-P(A+B))/(1-P(B)) = 1/3$

3. A klasszikus valószínűségi mező

1. Összes eset n=1000. Kedvező esetek k=0-nál 8^3 (a belső 8x8x8 kisnégyzetben lévő mindegyik részkocka jó), k=1-nél $6 \cdot 64$ (mindegyik lapon a belső 8x8-as négyzethez tartozóan), k=2-nél $12 \cdot 8$ (minden élen van 8 ilyen kocka) és végül k=3-nál 8 (a csúcsok nál lehet ilyen eset). **2.** Az összes eset n= 26^{11}, kedvező esetek száma k=1+$\binom{11}{2}-\binom{3}{2}-3\binom{2}{2}=50$ (az azonos betűk egymás közti cseréit le kell vonni). **3.** A valószínűség éppen 0,5. Ugyanis, ha tekintünk egy olyan sorozatot, amelyben a fejek száma páratlan, akkor ha az első dobást kicserélnénk az ellenkezőjére, olyan sorozatot kapunk, melyben a fejek száma már páros lesz. Azaz a páros és a páratlan fejdobásos sorozatok között kölcsönösen egy-egyértelmű leképezés hozható létre, vagyis mindegyikük ugyanolyan valószínű.**4.** a.- $\left(\frac{1}{2}\right)^n$, b.- 0, ha n páros és $\binom{n}{\frac{n}{2}}\left(\frac{1}{2}\right)^n$, ha n páros , c.- $\binom{n}{2}\left(\frac{1}{2}\right)^n$,

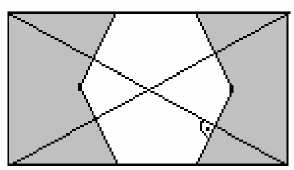
d.- $1-\left(\frac{1}{2}\right)^n - n\left(\frac{1}{2}\right)^n$ **5.** Az összes eset n=3!=6. Ezek között a nem kedvező eset csak kettő van: 2,3,1 és 3,1,2. A keresett valószínűség: 2/3. **6.** P(A)=P(„vagy kettő, vagy három fejet dobunk")=
$=\left(\binom{3}{2}+\binom{3}{3}\right)\left(\frac{1}{2}\right)^3 = 0{,}5$, P(B)=$\binom{3}{2}\left(\frac{1}{2}\right)^3 = \frac{3}{8}$, P(C)=P(„nem három fejet dobunk")=

=1-P(„három fejet dobunk")=1-$\left(\frac{1}{2}\right)^3 = \frac{7}{8}$. **7.** Az összes lehetsége lottóhúzások száma $n=\binom{90}{5}=43949268$, a kedvező esetek száma k=1 találatnál: $\binom{5}{1}\binom{85}{4}$, k=2-nél, $\binom{5}{2}\binom{85}{3}$, k=3-nál $\binom{5}{3}\binom{85}{2}$, k=4-nél $\binom{5}{4}\binom{85}{1}$ és végül k=5-nél $\binom{5}{5}\binom{85}{0}=1$. **8.** Ha N a golyók száma, ebből K a fehéreké, akkor $P(A)=\frac{K(N-1)(N-2)\cdots 1}{N!}=\frac{K}{N}$, és $P(B)=\frac{(N-1)(N-2)\cdots 1\cdot K}{N!}=\frac{K}{N}$, azaz a két esemény ugyanolyan valószínűségű. **9.** Összes eset n= n^n, a kedvező esetek száma pedig: n! . **10.** a.- $\frac{\binom{51}{12}}{\binom{52}{13}}$, b.- $\frac{\binom{13}{2}\binom{39}{11}}{\binom{52}{13}}$

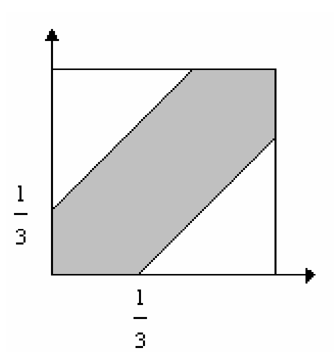
c.- $\frac{\binom{50}{11}}{\binom{52}{13}}$ d.- $1-\frac{\binom{39}{13}}{\binom{52}{13}}$

4. Geometriai valószínűségi mező

1. A pénz középpontjának s/2-nél nagyobb távolságra kell lennie egy padlóréstől, így a valószínűség p=1-s/d. **2.** a) Ahhoz, hogy a pénzdarab benne legyen a négyzetben, a pénz középpontjának a belső 7cm oldalhosszúságú négyzetben, igy a valószínűség p=0,49. b.) Az előző p valószínűséggel: $\binom{20}{5}p^5(1-p)^{15}$. **3.** A keresett valószínűség p= $1-\frac{4sd-s^2}{d^2\pi}$. Ha A azt az eseményt jelenti, hogy a tű a vízszintes oldalt metszi, B pedig az, hogy a tű függőleges oldalt keresztez, akkor meghatározandó a P(A+B) valószínűség. A Poincare tételéből: P(A+B)=P(A)+P(B)-P(AB). A Buffon-tű problémánál láttuk, hogy P(A)=P(B)=$\frac{2s}{d\pi}$. Az AB szorzatesemény valószínűségét a P(AB)=$\frac{4}{d^2\pi}\int_0^{\pi}\int_0^{\frac{s}{2}\sin\alpha}\int_0^{\frac{s}{2}|\cos\alpha|}dxdyd\alpha=\frac{s^2}{d^2\pi}$ képlettel számolhatjuk ki. A képletben x és y a tű középpontjának koordinátái, α pedig a tű egyenesének a vízszintessel bezárt szöge. A P(AB) valószínűség a két oldalt egyszerre metsző tűelhelyezkedésekhez tartozó (x,y,α) pontok alkotta térrész térfogatának és a dxdxπ hasáb térfogatának aránya. **4.** Két pont között egyenlő távolságra lévő pontok mértani helye a pontokat összekötő szakasz felezőmerőlegese. Így a keresett eseménynek megfelelő tartományt az alábbi ábrán besötétítéssel szemléltethetjük:

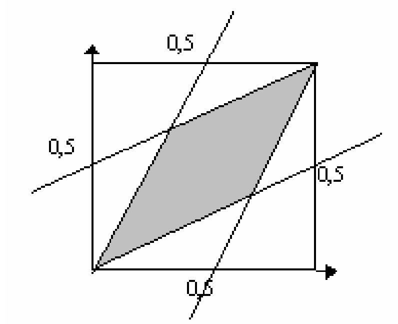


A középső (fehér) alakzat két szimmetrikus trapézból van összetéve. Mivel a trapézok középvonalai az átlók meghatározta háromszög középvonalával egyeznek meg, a hosszuk 1. A trapéz magasság 0,5. Így a fehér alakzat területe éppen 1 lesz. Ezért a besötétített alakzat területe is 1, így a keresett valószínűség 0,5. **5.** Jelölje x az egyik, y a másik ember véletlenmegérkezésének idejét. Az (x,y) pár egy véletlen pontot határoz meg az egységnégyzetben. A találkozáshoz fenn kell állnia a $|x-y|<\frac{1}{3}$ relációnak, melyet kielégítő pontok besötétítve láthatók az alábbi ábrán:

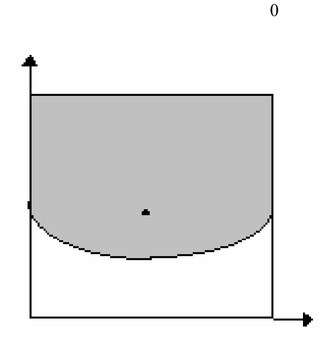


Az ábráról közvetlenül leolvasható, hogy a keresett valószínűség: $1-\frac{4}{9}=\frac{5}{9}$. **6.** A vizsgált eseményhez tartozó pontok (x,y) koordinátáira fennáll x<y esetben, hogy y-x<1-y és y-x<x.

(Az y<x esetben ezek a kritériumok x-y<1-x és x-y<y lennének.) Az egységnégyzeten bejelölve a relációknak eleget tevő pontok alkotta tartományt:



Ezek alapján a keresett valószínűség: $\frac{1}{3}$. **7.** A lift teljesen a fal mögötti takarásbanvan a földszinten 4 m-en keresztül, az 1.,2.,3. és 4. Emeleten 2-2 m-en át. A lift összútja 8+4x6+2=34 m. Így a keresett valószínűség: $p = \frac{12}{34}$. **8.** Egy ponttól és egy egyenestől azonos távolságban fekvő pontok mértani helye a síkban a parabóla. Így a négyzet pontjai közül azok lesznek a középponthoz közelebb, mint az alapon fekvő AB oldalhoz, amelyek felette vannak azon parabóla vonalának, melynek a középpont a fókusza, és az AB vonala a direktrisze. Ha AB az x tengelyre esik, és az A pont éppen az origó, akkor a parabóla egyenlete: $y = (x-0{,}5)^2 + 0{,}25$. A keresett terület: $1 - \int_0^1 (x-0{,}5)^2 + 0{,}25\,dx = \frac{2}{3}$.



5. A feltételes valószínűség és az események függetlensége

1. Egy n-szeres Bernoulli kísérletsorozatban a megfigyelt A esemény bekövetkezései gyakoriságának és az n-nek a hányadosa. **2.** A lehetetlen eseménye 0, a biztos eseményé 1. **3.** Az AB valószínűségének és a B eseményének hányadosa adja meg az A eseménynek a B eseményre vonatkozó feltételes valószínűségét. **4.** Az A,B,C események teljesen függetlenek, ha P(AB)=P(A)P(B), P(AC)=P(A)P(C), P(BC)=P(B)P(C), P(ABC)=P(A)P(B)P(C) . **5.** Legyenek az $A_1, A_2, \ldots, A_n \in \Im$ tetszőleges események, hogy $P(\prod_{i=1}^{n} A_i) > 0$. Ekkor $P\left(\prod_{i=1}^{n} A_i\right) = P\left(A_n \middle| \prod_{i=1}^{n-1} A_i\right) P\left(A_{n-1} \middle| \prod_{i=1}^{n-2} A_i\right) \cdots P(A_2|A_1)P(A_1)$. **6.** Legyenek $A_1, A_2, \ldots, A_n, \ldots \in \Im$ teljes eseményrendszer, vagyis $A_i \cdot A_j = \varnothing$, ($i \neq j$) és $\sum_{i=1}^{\infty} A_i = \Omega$.

Tegyük fel továbbá, hogy $P(A_i) > 0$ minden i-re. Ekkor tetszőleges $B \in \Im$ eseményre, ahol $P(B) > 0$ Akkor $P(A_i|B) = \dfrac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)}$ **7.** a.-H, b.-I , c.- H , d.-H, e.- I, f.-I, g.-H, h.-H, i.-I, j.-H, k.-I, l.-H, m.-H, n.-I, o.-I, p.-I, q.-I, r.-I **8.** $P(A|\overline{B}) = \dfrac{P(A\overline{B})}{P(\overline{B})} = \dfrac{P(A) - P(AB)}{1 - P(B)} = \dfrac{P(A+B) - P(B)}{1 - P(B)} = \dfrac{0{,}8 - 0{,}5}{0{,}5} = 0{,}6$ **9.** Pl. A: „Az egyik kockán kettest dobunk", B: „A másik kockán hármast dobunk", C:"Van hatos a két dobott érték között", D:"A dobott értékek nem egyenlőek". Az A és B függetlenek, C és D nem, hiszen P(CD)=$\dfrac{10}{36} \neq P(C)P(D) = \dfrac{55}{216}$ **10.** A feltétel szerint P(A+B)=1. 1=P(A+B)=P(A)+P(B)-P(AB), P(A|B)=P(AB)/P(B) és P(B|A)=P(AB)/P(A), azaz P(AB)=0,2P(B)=0,5P(A), amiből P(B)=2,5P(A) és így 1=3P(A), azaz P(A)=1/3 és P(B)=5/6

11. Ha B: „Mindegyik dobás páros", A: „Van hatos dobás". P(B)=$\dfrac{3^3}{6^3} = \dfrac{1}{8}$,

P(AB)=P(B)-P($\overline{A}$B)=$\dfrac{1}{8} - \dfrac{2^3}{6^3} = \dfrac{19}{216}$. Így P(A|B)=$\dfrac{P(AB)}{P(B)} = \dfrac{19}{27}$. **12.** A „Az első húzás fekete volt", B: „A második golyó fekete". A Bayes tételt alkalmazva: $P(A|B) = \dfrac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})}$, ahol P(B|A)=$\dfrac{b+c}{b+r+c}$, $P(B|\overline{A}) = \dfrac{b}{b+r+c}$, P(A)=$\dfrac{b}{b+r}$ és $P(\overline{A}) = \dfrac{r}{b+r}$. Így P(A|B)=$\dfrac{b+c}{b+r+c}$.**13.** Ha B: „Mindhárom kockán más-más eredmény van", A: „Az egyik kockán hatos van", akkor P(AB)=$\dfrac{3 \cdot 5 \cdot 4}{6^3} = \dfrac{10}{36}$, P(B)=$\dfrac{6 \cdot 5 \cdot 4}{6^3} = \dfrac{20}{36}$, így P(A|B)=0,5. **14.** Legyen A: „A hamis kockát választottuk ki",

B: „Tízszer dobva mindig hatost kapunk". A Bayes tételt alkalmazva: $P(A|B) = \dfrac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})}$, ahol P(A)=0,01 , P($\overline{A}$)=0,99 , P(B|A)=1 , P(B|$\overline{A}$)=$\dfrac{1}{6^{10}}$. Behelyettesítve: P(A|B)≈0,99999983 **15.** A: „x azt állítja, hogy y hazudik" , B: „y igazat mond". P(A|B)=P(„x hazudik")=2/3 , P(B)=1/3 , P(A|$\overline{B}$)=P(„x igazat mond")=1/3 , $P(\overline{B}) = 2/3$. A Bayes tételt alkalmazva $P(B|A) = \dfrac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\overline{B})P(\overline{B})} = \dfrac{1}{2}$. **16.**

A_1: „Az első urnából fehéret rakunk a másodikba, a másodikból fehéret rakunk vissza",
A_2: „Az első urnából fehéret rakunk a másodikba, a másodikból feketét rakunk vissza"
A_3: „Az első urnából feketét rakunk a másodikba, a másodikból fehéret rakunk vissza"
A_4: „Az első urnából feketét rakunk a másodikba, a másodikból feketét rakunk vissza"
B: „Harmadszorra az első urnából fehéret húzunk". A_1, A_2, A_3, A_4 teljes eseményrendszer.

$P(A_1) = \dfrac{m}{m+n}\dfrac{M+1}{N+M+1}$, $P(B|A_1) = \dfrac{m}{m+n}$, $P(A_2) = \dfrac{m}{m+n}\dfrac{N}{N+M+1}$, $P(B|A_2) = \dfrac{m-1}{m+n}$

$P(A_3) = \dfrac{n}{m+n}\dfrac{N+1}{N+M+1}$, $P(B|A_3) = \dfrac{m}{m+n}$, $P(A_4) = \dfrac{n}{m+n}\dfrac{M}{N+M+1}$, $P(B|A_4) = \dfrac{m+1}{m+n}$

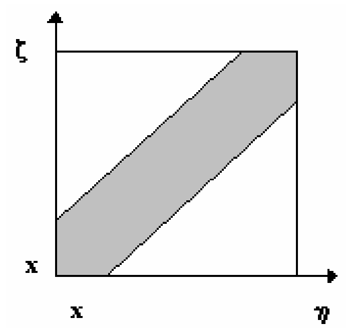
A teljes valószínűség tételéből: $P(B)=\sum_{i=1}^{4}P(B|A_i)P(A_i)=\ldots$. **17.** A: „Az első húzás után nyer a kezdő játékos”, B: „A harmadik húzás után nyer a kezdő játékos”, C: „Nyer a kezdő játékos”. Nyilván: P(C)=P(A)+P(B)=$\frac{1}{4}+\frac{3}{4}\cdot\frac{2}{3}\cdot\frac{1}{2}=\frac{1}{2}$. **18.** P(ξ=0)=$\frac{19}{100}$, P(η=i|ξ=0)=0, ha i>9. P(η=i|ξ=0)=$\frac{P(\text{"0 – át és i – t húztunk" , "i – t és 0 – át húztunk"})}{P(\xi=0)}=\frac{2}{19}$, ha i=1,2,...,9.**19.** Az optimális stratégia az, ha az egyik vázába egy fehér golyót teszünk, a másikba az összes többit. Ekkor a teljes valószínűség tételét alkalmazva: P(„A sah fehéret húz”)=$\frac{1}{2}\left(1+\frac{49}{99}\right)\approx 0{,}747$. Minden más szétosztásnaál csökken ez a valószínűség.

6. A valószínűségi változó és az eloszlásfüggvény fogalma

6. a-I,b-H,c-I,d-H,e-I,f-H,g-I,h-H,i-I,j-I,k-I,l-I,m-H,n-I,o-I,p-H,q-I,r-I,s-I,t-H,u-H,v-I,w-H,x-I.
7. P(η<x)=P(F(ξ)<x)=P(ξ<$F^{-1}(x))=F(F^{-1}(x))=x$ **8.**P(η<x)=$P(\ln\frac{1}{F(\xi)}<x)=$
$=P(F(\xi)>e^{-x})=1-P(F(\xi)\le e^{-x})=1-P(\xi<F^{-1}(e^{-x}))=1-e^{-x}\Rightarrow \xi\in E(1)$. **9.** Ha x>0:
$P(\eta<x)=P(\xi^2<x)=P(|x|<\sqrt{x})=P(-\sqrt{x}<\xi<\sqrt{x})=\Phi(\sqrt{x})-\Phi(-\sqrt{x})=2\Phi(\sqrt{x})-1\Rightarrow$ deriválás után kapjuk a sűrűségfüggvényt: $f_\eta(x)=2\varphi(\sqrt{x})\frac{1}{2\sqrt{x}}=\frac{1}{\sqrt{2\pi x}}e^{-\frac{x}{2}}$, ha x>0. **10.** P(η<x)=P(|ξ|<x)=P(-x<ξ<x)=F(x)-F(-x) , deriválás után kapjuk a sűrűségfüggvényt: $f_\eta(x)=f(x)+f(-x)$, x>0. **11.**Mivel ξ∈{0,1,2,…}⇒η∈{1,3,5,…,2n+1,…} és $P(\xi=k)=P(\eta=2k+1)=\frac{\lambda^k}{k!}e^{-\lambda}$ **12.** Ha x≤0, akkor P(η<x)=$P(\frac{1}{\xi}<x)=0$, mert ez lehetetlen. Ha x>0, akkor P(η<x)=$P(\frac{1}{\xi}<x)=P(\xi>\frac{1}{x})=1-P(\xi\le\frac{1}{x})=1-F_\xi(\frac{1}{x})=1-\frac{1}{x}$, ha még az is fennáll, hogy $\frac{1}{x}\le 1$, azaz x≥1. Így $F_\eta(x)=\begin{cases}0\text{ , ha } x\le 1\\ 1-\frac{1}{x}\text{ , ha } x>1\end{cases}$. A sűrűségfüggvényt deriválással határozhatjuk meg: $f_\eta(x)=x^{-2}$, ha x > 1 (különben =0). Másrészt P(ζ<x)=
$=P(\frac{\xi}{1+\xi}<x)=P(\xi<\frac{x}{1-x})=\begin{cases}\frac{x}{1-x}\text{ , ha } x\le 0{,}5\\ 1\text{ , ha } x>0{,}5\end{cases}$. (A $\frac{x}{1-x}\le 0$ sohasem teljesül.) Deriválás után: $f_\zeta(x)=(1-x)^{-2}$, ha x<0,5 (különben =0). **13.** P(η<x)=$P(e^\xi<x)=P(\xi<\ln x)=F_\xi(\ln x)=\Phi\left(\frac{\ln x-\mu}{\sigma}\right)$, így $f_\eta(x)=\frac{1}{\sqrt{2\pi}\sigma x}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$. **14.** P(η<x)=P(|ξ|<x)=P(-x<<x)=P(1-x<ξ<1+x)=$F_\xi(x+1)-F_\xi(1-x)=\begin{cases}\frac{x+1}{2}-\frac{1-x}{2}=x\text{ , ha } x\le 1\\ 1\text{ , ha } x>1\end{cases}$

, vagyis $\eta \in U[0,1]$. **15.** $P(\eta<x)=P(\xi < \frac{x-3}{2}) = 1 - e^{-\lambda\frac{x-3}{2}}$, ha $x \geq 3$. $f_\eta(x) = \frac{\lambda}{2} e^{-\lambda\frac{x-3}{2}}$, $x>3$. **16.** $P(\eta<x)=P(\xi < x^2) = 1 - e^{-\lambda x^2}$, $x>0$. $f_\eta(x) = 2\lambda x e^{-\lambda x^2}$, $x>0$. **17.** $P(\eta<x)=P(0<\frac{1}{\xi^2} < x) =$

$= P(\xi^2 > \frac{1}{x}) = P(\xi > \frac{1}{\sqrt{x}}) = 1 - P(\xi \leq \frac{1}{\sqrt{x}}) = 1 - F_\xi(\frac{1}{\sqrt{x}})$. Deriválás után $f_\eta(x) = \frac{\lambda}{2\sqrt{x^3}} e^{-\frac{\lambda}{\sqrt{x}}}$, $x>0$. **18.** Jelölje η a fejek száma, ζ az írások száma az n dobás közben. Így $P(\xi=n)=P(\eta=k,\zeta=n-k)+P((\eta=n-k,\zeta=k)=\binom{n-1}{k-1}\frac{1}{2^n}+\binom{n-1}{k-1}\frac{1}{2^n}=\binom{n-1}{k-1}\frac{1}{2^{n-1}}$. **19.** Jelölje η illetve ζ a két pont origótol vett távolságát! Ekkor $\xi=|\eta-\zeta|$. $P(\xi < x) = P(\eta - x < \zeta < \eta + x)$. Geometriai valószínűségszámítási módszerrel: (η,ζ) egy véletlen pont az egységnégyzetben, így a $\eta - x < \zeta < \eta + x$ feltételnek megfelelő tartomány:



A keresett eloszlásfüggvény: $F_\xi(x) = \begin{cases} 0, & \text{ha } x \leq 0 \\ 1-(1-x)^2, & \text{ha } x \in (0,1) \\ 1 & \text{ha } x \geq 1 \end{cases}$.

7. Vektor valószínűségi változók, valószínűségi változók együttes eloszlása

5. a-I,b-H,c-I,d-H,e-I,f-I,g-I,h-I,i-I,j-H **6.** Mivel az együttes eloszlás elemeinek összege 1, így 60p=1, azaz p=1/60. ξ és η függetlenek, mert minden lehetséges értékpárnál teljesül a függetlenség feltétele pl. $P(\xi=-1)=1/6$, $P(\eta=-1)=1/10$, és $P(\xi=-1, \eta=-1)=1/60$ stb. **7.** Ha a kockával 1,2,3-t dobunk, $P(\xi=4,\eta=2)=0$ nyilván, mert négynél kevesebb lapból nem lehet négy figurást kihúzni. Ha a kockával 4-et dobunk akkor a keresett esemény : „2 király és 2 figurás nem király”. $p_1 = P(\xi = 4, \eta = 2 | \text{"négyet dobtunk a kockával"}) = \frac{\binom{4}{2}\binom{8}{2}}{\binom{32}{4}}$. Ha a kockán ötöst kapunk, az esemény: „2 király és 2 figurás nem király és 1 egyéb”. $p_2 = P(\xi = 4, \eta = 2 | \text{"ötöt dobtunk a kockával"}) = \frac{\binom{4}{2}\binom{8}{2}\binom{20}{1}}{\binom{32}{4}}$ Végül, ha a dobás hatos volt, a keresett esemény: „2 király, 2 figurás nem király, 2 egyéb”.

$p_3 = P(\xi = 4, \eta = 2 | \text{"hatot dobtunk a kockával"}) = \dfrac{\binom{4}{2}\binom{8}{2}\binom{20}{2}}{\binom{32}{4}}$. A teljes valószínűség tételéből: $P(\xi = 4, \eta = 2) = \frac{1}{6}(p_1 + p_2 + p_3)$. **8.** $f_\xi(x) = \int_0^\infty 2e^{-2x-y}\,dy = 2e^{-2x}\int_0^\infty e^{-y}\,dy = 2e^{-2x}$ $x > 0$. $f_\eta(x) = \int_0^\infty 2e^{-2x-y}\,dx = 2e^{-y}\int_0^\infty e^{-2x}\,dx = e^{-y}$ $y > 0$. **9.** $f_\xi(x) = \int_0^1 0{,}8(x + xy + y)\,dy = 0{,}8\left[xy + x\frac{y^2}{2} + \frac{y^2}{2}\right]_0^1 = 1{,}2x + 0{,}4$, $f_\eta(y) = \int_0^1 0{,}8(x + xy + y)\,dy = 0{,}8\left[\frac{x^2}{2} + \frac{x^2}{2}y + xy\right]_0^1 = 1{,}2y + 0{,}4$. ξ és η nem függetlenek, mert $f_{\xi,\eta}(x, y) \neq f_\xi(x) f_\eta(y)$.

8. Várható érték, szórás, szórásnégyzet, magasabb momentumok, kovariancia és a korrelációs együttható

7. a-H,b-H,c-H,d-H,e-I,f-H,g-I,h-H,i-I,j-I,k-I,l-I,m-I,n-H,o-I,p-H,q-I,r-H,s-H,t-H **8.** $\mathbf{M}\eta = 2\mathbf{M}\xi + 1 = 2\lambda + 1$, $\mathbf{D}^2\eta = 4\mathbf{D}^2\xi = 4\lambda$ **9.** $\mathbf{M}\eta = \sum_{k=0}^{n} \frac{1}{1+k}\binom{n}{k} p^k (1-p)^{n-k} = \ldots$, $\mathbf{D}^2\eta = \sum_{k=0}^{n} \frac{1}{(1+k)^2}\binom{n}{k} p^k (1-p)^{n-k} - (\mathbf{M}\eta)^2 = \ldots$ **10.** Nem létezik, mert $\int_{-\infty}^{+\infty} x\frac{1}{\pi(1+x^2)}\,dx = \left[\frac{1}{2\pi}\ln(1+x^2)\right]_{-\infty}^{\infty}$ divergens. **11.** Egyrészt, a függetlenség miatt $\mathrm{cov}(\xi, \xi+\eta) = \mathrm{cov}(\xi, \xi) + \mathrm{cov}(\xi, \eta) = \mathbf{D}^2\xi$, másrészt $\mathbf{D}^2(\xi+\eta) = \mathbf{D}^2\xi + \mathbf{D}^2\eta = 2\mathbf{D}^2\xi$. Így $R(\xi, \xi+\eta) = \dfrac{\mathrm{cov}(\xi, \xi+\eta)}{\mathbf{D}\xi\mathbf{D}(\xi+\eta)} = \dfrac{\mathbf{D}^2\xi}{\sqrt{2}\mathbf{D}\xi\mathbf{D}\xi} = \dfrac{\sqrt{2}}{2}$ **12.** $P(\xi = 1) = P(\xi = 0) = P(\eta = 1) = P(\eta = 0) = 0{,}5$. $P(\xi+\eta = 0) = P(\xi = 0)P(\eta = 0) = 0{,}25$, $P(\xi+\eta = 1) = P(\xi = 1)P(\eta = 0) + P(\xi = 0)P(\eta = 1) = 0{,}5$, $P(\xi+\eta = 2) = P(\xi = 1)P(\eta = 1) = 0{,}25$. $P(|\xi-\eta| = 0) = P(\xi = 0)P(\eta = 0) + P(\xi = 1)P(\eta = 1) = 0{,}5$, $P(|\xi-\eta| = 1) = P(\xi = 1)P(\eta = 0) + P(\xi = 0)P(\eta = 1) = 0{,}5$. $\mathbf{M}(\xi+\eta) = \mathbf{M}\xi + \mathbf{M}\eta = 1$, $\mathbf{M}(|\xi-\eta|) = 0{,}5$. $\mathbf{M}((\xi+\eta)|\xi-\eta|) = \mathbf{M}(|\xi^2 - \eta^2|) = 0{,}5$. Így $\mathrm{cov}(\xi+\eta, |\xi-\eta|) = 0{,}5 - 1\cdot 0{,}5 = 0$. $\xi+\eta$ és $|\xi-\eta|$ nem lehetnek függetlenek, mert pl. $P(\xi+\eta = 0, |\xi-\eta| = 1) = 0$ de $P(\xi+\eta = 0)P(|\xi-\eta| = 1) = 0{,}25\cdot 0{,}5 = 0{,}125 \neq 0$. **13.** $\mathbf{M}\zeta = \int_0^2 \sin x \cdot 0{,}5\,dx = \frac{1-\cos 2}{2}$, $\mathbf{M}\eta = \int_0^2 \cos x \cdot 0{,}5\,dx = \frac{\sin 2}{2}$,

$\mathbf{M}\zeta\eta = \mathbf{M}\left(\sin\xi\cos\xi\right) = 0{,}5\mathbf{M}\sin 2\xi = 0{,}5\int_0^2 \sin\ 2x\cdot 0{,}5\,dx = \frac{1-\cos 4}{8}$.

$\mathrm{cov}(\eta,\zeta) = \frac{1-\cos 4}{8} - \frac{1-\cos 2}{2}\frac{\sin 2}{2} \approx 0{,}216 \Rightarrow$ nem függetlenek! (Megjegyzés: $P(\eta^2+\zeta^2=1)=1$) **14.** Ha $\tilde{\xi} = \frac{\xi-\mu}{\sigma}$ jelöli a standardizáltat, akkor $\tilde{\xi}\in N(0,1)$. $\mathbf{M}\tilde{\xi}^n = \int_{-\infty}^{\infty} x^n\varphi(x)\,dx = (n-1)\int_{-\infty}^{\infty} x^{n-2}\varphi(x)\,dx = (n-1)\mathbf{M}\tilde{\xi}^{n-2}$. Mivel $\mathbf{M}\tilde{\xi}=0$, így a standardizált minden páratlan hatványának várható értéke 0. $\mathbf{M}\tilde{\xi}^{2n} = (n-1)(n-3)\cdots 1$=(n-1)!!, mivel $\mathbf{M}\tilde{\xi}^{\mathbf{2}} = 1$. Másrészt $\mathbf{M}\xi^n = \mathbf{M}\left(\sigma\tilde{\xi}+\mu\right)^n = \sum_{k=0}^{n}\binom{n}{k}\sigma^k\mu^{n-k}\mathbf{M}\tilde{\xi}^k$ is fennáll. Behelyettesítve kaphatjuk a végeredményt.

9. A nagy számok törvényei és a centrális határeloszlás tételek

6. a.-H,b.-H,c,-H,d-I,e.-H **7.** Jelölje ξ a csavarok számát! Ekkor a Csebisev egyenlőtlenségből: $P\left(4900\le\xi\le 5100\right) = P\left(\left|\xi-5000\right|\le 100\right) \ge 1-\frac{400}{10000} = 0{,}96$. **8.** Jelölje ξ a szálszakadások számát! Ekkor a Moivre-Laplace törvényből: $P\left(\frac{\xi-500\cdot 0{,}008}{\sqrt{500\cdot 0{,}008\cdot 0{,}992}} < x\right) = P\left(\xi < 1{,}99\cdot x+4\right) \approx \Phi(x)$. Másrészt $\Phi\left(1{,}65\right) = 0{,}95$, azaz x=1,65-nél: $P\left(\xi < 1{,}99\cdot 1{,}65+4\right) = P\left(\xi < 7{,}28\right) = 0{,}95$, vagyis a szálszakadások száma 8-nál kisebb lesz legalább 95%-os valószínűséggel. **9.** A centrális határeloszlás tételt használva: $\lim_{n\to\infty} P\left(\xi_1+\xi_2+\dots+\xi_n < x\right) = \lim_{n\to\infty} P\left(\frac{\xi_1+\xi_2+\dots+\xi_n-nm}{\sqrt{n}\sigma} < x^*\right) = \Phi\left(\lim_{n\to\infty} x^*\right)$, ahol $m=\mathbf{M}\xi_i$ $\sigma=\mathbf{D}^2\xi_i$, $x^* = \frac{x-nm}{\sqrt{n}\sigma}$. De $\lim_{n\to\infty}\frac{x-nm}{\sqrt{n}\sigma} = \begin{cases} -\infty, \text{ ha } m>0 \\ 0, \text{ ha } m=0 \\ \infty, m<0 \end{cases}$, amiből már következik az állítás. **10.** A Markov egyenlőtlenségből: $P\left(|\xi|>3\right) \le \frac{\mathbf{M}|\xi|}{3} = \frac{2}{3}\frac{1}{\sqrt{2\pi}}$, amiből már következik $P\left(-3<\xi<3\right) \ge 1-\frac{2}{3}\frac{1}{\sqrt{2\pi}}$. **11.** Jelölje ξ a működő gépek számát! Nyilván $\xi\in B(300,\ 0{,}7)$. A Moivre-Laplace tételből $P\left(\xi < np+x\sqrt{npq}\right) = P\left(\xi < 210+7{,}93\cdot x\right) \approx \Phi(x)$. Mivel $\Phi(3)\approx 0{,}999$, így $P\left(\xi<234\right)\approx 0{,}999$, vagyis az üzemelő gépek száma kevesebb mint 234 99,9%-kal.

II. fejezet. MATEMATIKAI STATISZTIKA

1. A matematikai statisztika alapfogalmai

8.a.-I,b.-H,c.-I,d.-I,e.-I,f.-H,g.-I,h.-H,i.-I,j.-I,k..-I,l.-H,m.-H,n.-I,o.-H,p.-I,q.-H **9.** Az átlagstatisztika tulajdonságait kihasználva $\mathbf{M}T_2 = \mathbf{M}\overline{\xi}_n = \frac{1}{\lambda} = \vartheta$, $\mathbf{D}^2 T_2 = \mathbf{D}^2\overline{\xi}_n = \frac{1}{\lambda n} = \frac{\vartheta}{n}$. Másrészt a rendezett minta eloszlására vonatkozó tétel miatt: $P(T_1 < x) = P\left(\xi_1^* < \frac{x}{n}\right) = 1 - \left(1 - \left(1 - e^{-\lambda\frac{x}{n}}\right)\right)^n = 1 - e^{-\lambda x}$, azaz $T_1 \in E(\lambda)$. Innen már közvetlenül következik, hogy $\mathbf{M}T_1 = \frac{1}{\lambda} = \vartheta$, és $\mathbf{D}^2 T_1 = \frac{1}{\lambda} = \vartheta$. Látható, hogy $\mathbf{D}^2 T_2 < \mathbf{D}^2 T_1$, azaz T_2 a hatásosabb torzítatlan becslés. **10.** Mivel $P(\xi_1^* < x) = 1 - (1 - F(x,\vartheta))^n$, ahol $F(x,\vartheta) = \int_\vartheta^x f(t,\vartheta)\,dt = \int_\vartheta^x e^{\vartheta - t}\,dt = e^\vartheta \int_\vartheta^x e^{-t}\,dt = 1 - e^{\vartheta - x}$, $x > \vartheta$. Így $P(\xi_1^* < x) = 1 - e^{n\vartheta - nx} \Rightarrow f_{\xi_1^*}(x) = ne^{n\vartheta - nx}$. $\mathbf{M}\xi_1^* = \int_\vartheta^\infty x f_{\xi_1^*}(x)\,dx = \int_\vartheta^\infty x n e^{n\vartheta - nx}\,dx = \vartheta + \frac{1}{n}$, vagyis $\mathbf{M}T = \vartheta$, azaz T torzítatlan becslés. Továbbá $\mathbf{M}(\xi_1^*)^2 = \int_{-\infty}^\infty x^2 f_{\xi_1^*}(x)\,dx = \int_\vartheta^\infty x^2 n e^{n\vartheta - nx}\,dx = \vartheta^2 + \frac{2}{n}\vartheta + \frac{2}{n^2}$, ahonnan $\mathbf{D}^2 T = \mathbf{D}^2\xi_1^* = \frac{1}{n^2}$, azaz T erősen konzisztens, vagyis konzisztens is. **11.** a.) $\mathbf{M}\xi_i = \int_{-\infty}^\infty x \frac{\vartheta}{2} e^{-\vartheta|x|}\,dx = 0$, mert páratlan az inegranduszfüggvény. b.) A likelihood függvény: $L(x_1, x_2, x_3, x_4, x_5, \vartheta) = \frac{\vartheta^5}{32} e^{-\vartheta \sum_{i=1}^5 |x_1|}$, a loglikelihood függvény: $l(x_1, x_2, x_3, x_4, x_5, \vartheta) = \ln L(x_1, x_2, x_3, x_4, x_5, \vartheta) = 5 \ln \frac{\vartheta}{2} - \vartheta \sum_{i=1}^5 |x_1|$. Innen deriválás után: $\frac{\partial l}{\partial \vartheta} = \frac{5}{\vartheta} - \sum_{i=1}^5 |x_i| = 0 \Rightarrow \vartheta = \frac{1}{\frac{1}{5}\sum_{i=1}^5 |x_i|} = 5$. Az adott mintarealizációnál a ϑ paraméter maximum likelihood becslése 5. **12.** Ismert szórású normális eloszlású minta esetén az ismeretlen várható értékre az a $[T_1, T_2]$ konfidenciaintervallum szerkeszthető, ahol $T_1 = \overline{\xi}_n - \frac{\sigma}{\sqrt{n}} u_e$, $T_2 = \overline{\xi}_n + \frac{\sigma}{\sqrt{n}} u_e$, $\Phi(u_\varepsilon) = 1 - \frac{\varepsilon}{2}$. Most ε=0,1, így $u_\varepsilon = 1{,}65$. A konfidencia intervallum hossza $1 = \frac{2\sigma}{\sqrt{n}} u_\varepsilon \Rightarrow \sqrt{n} = 2{,}4 \cdot 1{,}65 = 3{,}96$. A mintaelemszámnak tehát legalább 16 nak kell lennie. **13.** A számított statisztikák most: $\overline{x}_5 = 1$ és $s_5^{*2} = 0{,}09$. Ismeretlen szórású normális minta esetén a $[T_1, T_2]$ konfidencia intervallum $T_1 = \overline{x}_5 - \frac{s_5^*}{\sqrt{5}} t_e$, $T_2 = \overline{x}_5 + \frac{s_5^*}{\sqrt{5}} t_e$, ahol a t_ε számot a 4 szabadságfokú Student eloszlás táblázatából kell kiolvasni. Most ε=0,1 így $t_{0,1} = 2{,}132$. Azaz a 90%-os konfidenciaintervallum: [0,713 , 1,286].

2. Hipotéziselmélet

10.a.-H,b.-I,c.-I,d.-I,e.-H,f.-I,g.-I,h.-H,i.-I,j.-H,k.-I,l.-H,m.-H,n.-I,o.-H,p.-H,q.-I,r.-H,s.-I,t.-H,u.-I,v.-H,w.-H,x.-I **11.** Kétmintás t-próbával döntünk. A mintákból számolt korrigált empirikus szórásnégyzet statisztikák rendre $s_6^{*2} = 25{,}33$ és $\sigma_6^{*2} = 38{,}33$. Az F-próba statisztikájának számított értéke 1,513225424398, amihez kiolvasható kritikus érték (a szabadságfokok most $f_1 = f_2 = 5$) $F_{0,05} = 5{,}05$, ami azt is jelenti, hogy a minták szórásai egyenlőeknek tekinthetők, vagyis alkalmazhatjuk a kétmintás t-próbát. A mintaátlagok $\overline{x}_6 = 297$ és $\overline{y}_6 = 310$, próbastatisztika számított értéke:

$\frac{|297-310|}{\sqrt{5(25{,}33+38{,}33)}}\sqrt{\frac{36(6+6-2)}{6+6}} \approx 3{,}64$, az előírt szignifikancia szinthez tartozó kritikus érték a 10 szabadságfokot figyelembevéve: $K_{0,05} = 2{,}228$. Mivel a számított érték a nagyobb, így a két minta várható értékei nem lehetnek azonosak, vagyis az új technológiával előállított próbakockák törőszilárdsága szignifikánsan nagyobb, hiszen $\overline{y}_6 > \overline{x}_6$. **12.** Tiszta illeszkedésvizsgálatot kell végeznünk. A kocka szabályossága azt jelenti, hogy mindegyik értéket ugyanakkora valószínűséggel dobhatunk vele, azaz $p_1 = p_2 = \ldots = p_6 = \frac{1}{6}$. A számított próbastatisztika: $\sum_{i=1}^{6}\frac{(\upsilon_i - 200)^2}{200} \approx 5{,}72$. A 90%-os szignifikancia szinthez és a 5 szabadságfokhoz tartozó kritikus érték: $K_{0,1} = 9{,}236$. Vagyis, az adott szinten a kocka szabályosnak tekinthető. **13.** Tiszta illeszkedésvizsgálattal kell dolgozni. Az egyes súlycsoportokhoz tartozó elméleti valószínűségek, ha az N(78, 11) eloszlással számolunk:

$$p_1 = P(\xi < 60) = \Phi\left(\frac{60-78}{11}\right) = 1-\Phi\left(\frac{18}{11}\right) = 0{,}05089,$$

$$p_2 = P(60 \le \xi < 70) = \Phi\left(\frac{-8}{11}\right) - \Phi\left(\frac{-18}{11}\right) = \Phi\left(\frac{18}{11}\right) - \Phi\left(\frac{8}{11}\right) = 0{,}18266,$$

$$p_3 = P(70 \le \xi < 80) = \Phi\left(\frac{2}{11}\right) + \Phi\left(\frac{8}{11}\right) - 1 = 0{,}33858,$$

$$p_4 = P(80 \le \xi < 90) = \Phi\left(\frac{12}{11}\right) - \Phi\left(\frac{2}{11}\right) = 0{,}29021,$$

$$p_5 = P(90 \le \xi < 100) = \Phi(2) - \Phi\left(\frac{12}{11}\right) = 0{,}11491,$$

$p_6 = P(\xi \ge 101) = 1-\Phi(2) = 0{,}02275$. A próbastatisztika számított értéke

$$T = \frac{(7-5{,}089)^2}{5{,}089} + \frac{(16-18{,}266)^2}{18{,}266} + \frac{(32-33{,}858)^2}{33{,}858} + \frac{(28-29{,}021)^2}{29{,}021} +$$

$+\frac{(13-11{,}491)^2}{11{,}491} + \frac{(4-2{,}275)^2}{2{,}275} \approx 2{,}6427$. A szabadságfok 5, a kritikus értékε=0,05-nél $K_{0,05} = 11{,}070$. A megvizsgált mintán az utasok súlyának eloszlása nem tér el a feltételezettől. **14.** A feladatot függetlenségvizsgálattal kell megoldani. A mérési adatokat az alábbi kontingenciatáblázatba foglalhatjuk:

méretre	szakítószilárdságra		
	megfelelő	selejtes	összesen
megfelelő	416	23	439

selejtes	16	5	21
összesen	432	28	460

A próbastatisztika számított értéke: $T = 460\dfrac{(416\cdot 5-16\cdot 23)^2}{432\cdot 28\cdot 21\cdot 439} \approx 1{,}209$. A szabadságifok 1, a kritikus érték $K_{0,05} = 3{,}841$. Mivel $T < K_{0,05}$, ezért a függetlenségre vonatkozó nullhipotézist 95%-os szignifikancia szinten elfogadjuk.

3. Regresszióanalízis

5. a.-I,b.-I,c.-I,d.-H,e.-I **6.** $\hat{a} = \dfrac{\sum\limits_{i=1}^{18} x_i y_i - 18\cdot \overline{x}_{18}\overline{y}_{18}}{\sum\limits_{i=1}^{18} x_i^2 - 18(\overline{x}_{18})^2} \approx 0{,}113$ és $\hat{b} = \overline{y}_{18} - \hat{a}\cdot \overline{x}_{18} \approx 0{,}040$.

$I^2 = 0{,}79$. **7.** Lineáris regresszióval. A számított statisztikák: $\overline{x}_{10} = 6{,}31$ és $\overline{y}_{10} = 11{,}11$, $s_{10}^{*\,2} = 50{,}445$ és $\sigma_{10}^{*\,2} = 62{,}083$, a minták empirikus korrelációs együtthatója: $\hat{r} = 0{,}931$, $\hat{a} = \hat{r}\dfrac{\sigma_{10}^*}{s_{10}^*} \approx 10{,}36$ és $\hat{b} = \overline{y}_{10} - \hat{a}\cdot \overline{x}_{10} \approx 4{,}57$. $I^2 = 0{,}873$. **8.** Az $y = a\cdot x^2 + b\cdot x + c$ parabóla együtthatóit a

$$\begin{pmatrix} \sum\limits_{i=1}^{6} x_i^4 & \sum\limits_{i=1}^{6} x_i^3 & \sum\limits_{i=1}^{6} x_i^2 \\ \sum\limits_{i=1}^{6} x_i^3 & \sum\limits_{i=1}^{6} x_i^2 & \sum\limits_{i=1}^{6} x_i^1 \\ \sum\limits_{i=1}^{6} x_i^2 & \sum\limits_{i=1}^{6} x_i^1 & 6 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \sum\limits_{i=1}^{6} x_i^2 y_i \\ \sum\limits_{i=1}^{6} x_i y_i \\ \sum\limits_{i=1}^{6} y_i \end{pmatrix}$$

egyenletrendszer megoldásából kapjuk. Az összegeket a mintából kiszámolva ez konkrétan a

$$\begin{pmatrix} 13124 & 1504 & 188 \\ 1504 & 188 & 28 \\ 188 & 28 & 6 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 8510 \\ 1038 \\ 155 \end{pmatrix}$$

háromismeretlenes lineáris egyenletrendszer megoldását jelenti. A megoldások: $\hat{a} = 0{,}56$, $\hat{b} = -0{,}63$, $\hat{c} = 11{,}191$. Az illeszkedés jósága $I^2 = 1$, ami igen erős regressziós összefüggést sejtet.

TÁBLÁZATOK

A standard normális eloszlás sűrűségfüggvény és eloszlásfüggvény táblázata

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,,\ \Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}\,dt$$

1. Ha $\xi\in N(\mu,\sigma)$, akkor $P(\xi<x)=\Phi\left(\frac{x-\mu}{\sigma}\right)$ és $f_\xi(x)=\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$. (Ezen tulajdonságok miatt van csak standard normális eloszlás-táblázat).
2. Ha x>0, akkor $\Phi(-x)=1-\Phi(x)$. (Ezen tulajdonság miatt van a táblázatban csak nemnegatív x argumentum)
3. Ha $\varepsilon\in(0,1)$, akkor $P(-u_\varepsilon<\frac{\xi-\mu}{\sigma}<u_\varepsilon)=2\Phi(u_\varepsilon)-1=1-\varepsilon$, azaz $\Phi(u_\varepsilon)=1-\frac{\varepsilon}{2}$.

x	Φ(x)	φ(x)
,00	,500000	,398942
,01	,503989	,398922
,02	,507978	,398862
,03	,511966	,398763
,04	,515953	,398623
,05	,519939	,398444
,06	,523922	,398225
,07	,527903	,397966
,08	,531881	,397668
,09	,535856	,397330
,10	,539828	,396953
,11	,543795	,396536
,12	,547758	,396080
,13	,551717	,395585
,14	,555670	,395052
,15	,559618	,394479
,16	,563559	,393868
,17	,567495	,393219
,18	,571424	,392531
,19	,575345	,391806
,20	,579260	,391043
,21	,583166	,390242
,22	,587064	,389404
,23	,590954	,388529
,24	,594835	,387617
,25	,598706	,386668
,26	,602568	,385683
,27	,606420	,384663
,28	,610261	,383606
,29	,614092	,382515
,30	,617911	,381388
,31	,621720	,380226
,32	,625516	,379031
,33	,629300	,377801
,34	,633072	,376537
,35	,636831	,375240
,36	,640576	,373911
,37	,644309	,372548
,38	,648027	,371154
,39	,651732	,369728
,40	,655422	,368270
,41	,659097	,366782
,42	,662757	,365263
,43	,666402	,363714
,44	,670031	,362135
,45	,673645	,360527
,46	,677242	,358890
,47	,680822	,357225
,48	,684386	,355533
,49	,687933	,353812
,50	,691462	,352065

x	Φ(x)	φ(x)
,51	,694974	,350292
,52	,698468	,348493
,53	,701944	,346668
,54	,705401	,344818
,55	,708840	,342944
,56	,712260	,341046
,57	,715661	,339124
,58	,719043	,337180
,59	,722405	,335213
,60	,725747	,333225
,61	,729069	,331215
,62	,732371	,329184
,63	,735653	,327133
,64	,738914	,325062
,65	,742154	,322972
,66	,745373	,320864
,67	,748571	,318737
,68	,751748	,316593
,69	,754903	,314432
,70	,758036	,312254
,71	,761148	,310060
,72	,764238	,307851
,73	,767305	,305627
,74	,770350	,303389
,75	,773373	,301137
,76	,776373	,298872
,77	,779350	,296595
,78	,782305	,294305
,79	,785236	,292004
,80	,788145	,289692
,81	,791030	,287369
,82	,793892	,285036
,83	,796731	,282694
,84	,799546	,280344
,85	,802337	,277985
,86	,805105	,275618
,87	,807850	,273244
,88	,810570	,270864
,89	,813267	,268477
,90	,815940	,266085
,91	,818589	,263688
,92	,821214	,261286
,93	,823814	,258881
,94	,826391	,256471
,95	,828944	,254059
,96	,831472	,251644
,97	,833977	,249228
,98	,836457	,246809
,99	,838913	,244390
1,00	,841345	,241971
1,01	,843752	,239551

x	Φ(x)	φ(x)
1,02	,846136	,237132
1,03	,848495	,234714
1,04	,850830	,232297
1,05	,853141	,229882
1,06	,855428	,227470
1,07	,857690	,225060
1,08	,859929	,222653
1,09	,862143	,220251
1,10	,864334	,217852
1,11	,866500	,215458
1,12	,868643	,213069
1,13	,870762	,210686
1,14	,872857	,208308
1,15	,874928	,205936
1,16	,876976	,203571
1,17	,879000	,201214
1,18	,881000	,198863
1,19	,882977	,196520
1,20	,884930	,194186
1,21	,886861	,191860
1,22	,888768	,189543
1,23	,890651	,187235
1,24	,892512	,184937
1,25	,894350	,182649
1,26	,896165	,180371
1,27	,897958	,178104
1,28	,899727	,175847
1,29	,901475	,173602
1,30	,903200	,171369
1,31	,904902	,169147
1,32	,906582	,166937
1,33	,908241	,164740
1,34	,909877	,162555
1,35	,911492	,160383
1,36	,913085	,158225
1,37	,914657	,156080
1,38	,916207	,153948
1,39	,917736	,151831
1,40	,919243	,149727
1,41	,920730	,147639
1,42	,922196	,145564
1,43	,923641	,143505
1,44	,925066	,141460
1,45	,926471	,139431
1,46	,927855	,137417
1,47	,929219	,135418
1,48	,930563	,133435
1,49	,931888	,131468
1,50	,933193	,129518
1,51	,934478	,127583
1,52	,935745	,125665
1,53	,936992	,123763

x	Φ(x)	φ(x)
1,54	,938220	,121878
1,55	,939429	,120009
1,56	,940620	,118157
1,57	,941792	,116323
1,58	,942947	,114505
1,59	,944083	,112704
1,60	,945201	,110921
1,61	,946301	,109155
1,62	,947384	,107406
1,63	,948449	,105675
1,64	,949497	,103961
1,65	,950529	,102265
1,66	,951543	,100586
1,67	,952540	,098925
1,68	,953521	,097282
1,69	,954486	,095657
1,70	,955435	,094049
1,71	,956367	,092459
1,72	,957284	,090887
1,73	,958185	,089333
1,74	,959070	,087796
1,75	,959941	,086277
1,76	,960796	,084776
1,77	,961636	,083293
1,78	,962462	,081828
1,79	,963273	,080380
1,80	,964070	,078950
1,81	,964852	,077538
1,82	,965620	,076143
1,83	,966375	,074766
1,84	,967116	,073407
1,85	,967843	,072065
1,86	,968557	,070740
1,87	,969258	,069433
1,88	,969946	,068144
1,89	,970621	,066871
1,90	,971283	,065616
1,91	,971933	,064378
1,92	,972571	,063157
1,93	,973197	,061952
1,94	,973810	,060765
1,95	,974412	,059595
1,96	,975002	,058441
1,97	,975581	,057304
1,98	,976148	,056183
1,99	,976705	,055079
2,00	,977250	,053991
2,01	,977784	,052919
2,02	,978308	,051864
2,03	,978822	,050824
2,04	,979325	,049800
2,05	,979818	,048792

x	Φ(x)	φ(x)
2,06	,980301	,047800
2,07	,980774	,046823
2,08	,981237	,045861
2,09	,981691	,044915
2,10	,982136	,043984
2,11	,982571	,043067
2,12	,982997	,042166
2,13	,983414	,041280
2,14	,983823	,040408
2,15	,984222	,039550
2,16	,984614	,038707
2,17	,984997	,037878
2,18	,985371	,037063
2,19	,985738	,036262
2,20	,986097	,035475
2,21	,986447	,034701
2,22	,986791	,033941
2,23	,987126	,033194
2,24	,987455	,032460
2,25	,987776	,031740
2,26	,988089	,031032
2,27	,988396	,030337
2,28	,988696	,029655
2,29	,988989	,028985
2,30	,989276	,028327
2,31	,989556	,027682
2,32	,989830	,027048
2,33	,990097	,026426
2,34	,990358	,025817
2,35	,990613	,025218
2,36	,990863	,024631
2,37	,991106	,024056
2,38	,991344	,023491
2,39	,991576	,022937
2,40	,991802	,022395
2,41	,992024	,021862
2,42	,992240	,021341
2,43	,992451	,020829
2,44	,992656	,020328
2,45	,992857	,019837
2,46	,993053	,019356
2,47	,993244	,018885
2,48	,993431	,018423
2,49	,993613	,017971
2,50	,993790	,017528
2,51	,993963	,017095
2,52	,994132	,016670
2,53	,994297	,016254
2,54	,994457	,015848
2,55	,994614	,015449
2,56	,994766	,015060
2,57	,994915	,014678

x	Φ(x)	φ(x)
2,58	,995060	,014305
2,59	,995201	,013940
2,60	,995339	,013583
2,61	,995473	,013234
2,62	,995604	,012892
2,63	,995731	,012558
2,64	,995855	,012232
2,65	,995975	,011912
2,66	,996093	,011600
2,67	,996207	,011295
2,68	,996319	,010997
2,69	,996427	,010706
2,70	,996533	,010421
2,71	,996636	,010143
2,72	,996736	,009871
2,73	,996833	,009606
2,74	,996928	,009347
2,75	,997020	,009094
2,76	,997110	,008846
2,77	,997197	,008605
2,78	,997282	,008370
2,79	,997365	,008140
2,80	,997445	,007915
2,81	,997523	,007697
2,82	,997599	,007483
2,83	,997673	,007274
2,84	,997744	,007071
2,85	,997814	,006873
2,86	,997882	,006679
2,87	,997948	,006491
2,88	,998012	,006307
2,89	,998074	,006127
2,90	,998134	,005953
2,91	,998193	,005782
2,92	,998250	,005616
2,93	,998305	,005454
2,94	,998359	,005296
2,95	,998411	,005143
2,96	,998462	,004993
2,97	,998511	,004847
2,98	,998559	,004705
2,99	,998605	,004567

A Student eloszlás táblázata

$$P(|\xi| > t_\varepsilon) = \varepsilon,\ P(-t_e < \xi < t_e) = 1-\varepsilon$$

Sűrűségfüggvény

$$f_\xi(x) = c(n)\cdot\left(\frac{n}{n+x^2}\right)^{\frac{n+1}{2}},\ \text{ahol } c(n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}},\ \Gamma(x) = \int_0^\infty e^{-t}t^{x-1}\,dt.$$

n a szabadságfok

ε	0,90	0,80	0,70	0,60	0,50	0,40	0,30
Szabadsági fok							
1,00	**,158384**	**,324920**	**,509525**	**,726543**	**1,000000**	**1,376382**	**1,962611**
2,00	**,142134**	**,288675**	**,444750**	**,617213**	**,816497**	**1,060660**	**1,386207**
3,00	**,136598**	**,276671**	**,424202**	**,584390**	**,764892**	**,978472**	**1,249778**
4,00	**,133830**	**,270722**	**,414163**	**,568649**	**,740697**	**,940965**	**1,189567**
5,00	**,132175**	**,267181**	**,408229**	**,559430**	**,726687**	**,919544**	**1,155767**
6,00	**,131076**	**,264835**	**,404313**	**,553381**	**,717558**	**,905703**	**1,134157**
7,00	**,130293**	**,263167**	**,401538**	**,549110**	**,711142**	**,896030**	**1,119159**
8,00	**,129707**	**,261921**	**,399469**	**,545934**	**,706387**	**,888890**	**1,108145**
9,00	**,129253**	**,260955**	**,397868**	**,543480**	**,702722**	**,883404**	**1,099716**
10,00	**,128890**	**,260185**	**,396591**	**,541528**	**,699812**	**,879058**	**1,093058**
11,00	**,128594**	**,259556**	**,395551**	**,539938**	**,697445**	**,875530**	**1,087666**
12,00	**,128347**	**,259033**	**,394686**	**,538618**	**,695483**	**,872609**	**1,083211**
13,00	**,128139**	**,258591**	**,393955**	**,537504**	**,693829**	**,870152**	**1,079469**
14,00	**,127961**	**,258213**	**,393331**	**,536552**	**,692417**	**,868055**	**1,076280**
15,00	**,127806**	**,257885**	**,392790**	**,535729**	**,691197**	**,866245**	**1,073531**
16,00	**,127671**	**,257599**	**,392318**	**,535010**	**,690132**	**,864667**	**1,071137**
17,00	**,127552**	**,257347**	**,391902**	**,534377**	**,689195**	**,863279**	**1,069033**
18,00	**,127447**	**,257123**	**,391533**	**,533816**	**,688364**	**,862049**	**1,067170**
19,00	**,127352**	**,256923**	**,391202**	**,533314**	**,687621**	**,860951**	**1,065507**
20,00	**,127267**	**,256743**	**,390906**	**,532863**	**,686954**	**,859964**	**1,064016**
21,00	**,127190**	**,256580**	**,390637**	**,532455**	**,686352**	**,859074**	**1,062670**
22,00	**,127120**	**,256432**	**,390394**	**,532085**	**,685805**	**,858266**	**1,061449**
23,00	**,127056**	**,256297**	**,390171**	**,531747**	**,685306**	**,857530**	**1,060337**
24,00	**,126998**	**,256173**	**,389967**	**,531438**	**,684850**	**,856855**	**1,059319**
25,00	**,126944**	**,256060**	**,389780**	**,531154**	**,684430**	**,856236**	**1,058384**
26,00	**,126895**	**,255955**	**,389607**	**,530892**	**,684043**	**,855665**	**1,057523**
27,00	**,126849**	**,255858**	**,389448**	**,530649**	**,683685**	**,855137**	**1,056727**
28,00	**,126806**	**,255768**	**,389299**	**,530424**	**,683353**	**,854647**	**1,055989**
29,00	**,126767**	**,255684**	**,389161**	**,530214**	**,683044**	**,854192**	**1,055302**
30,00	**,126730**	**,255605**	**,389032**	**,530019**	**,682756**	**,853767**	**1,054662**
40,00	**,126462**	**,255039**	**,388100**	**,528606**	**,680673**	**,850700**	**1,050046**
60,00	**,126194**	**,254473**	**,387170**	**,527198**	**,678601**	**,847653**	**1,045469**
120,00	**,125928**	**,253910**	**,386244**	**,525796**	**,676540**	**,844627**	**1,040932**

ε	0,20	0,10	0,05	0,02	0,01	0,001
szabadsági fok						
1,00	**3,077684**	**6,313752**	**12,70620**	**31,82052**	**63,65674**	**636,6205**
2,00	**1,885618**	**2,919986**	**4,302653**	**6,964557**	**9,924843**	**31,59905**
3,00	**1,637744**	**2,353363**	**3,182446**	**4,540703**	**5,840909**	**12,92398**
4,00	**1,533206**	**2,131847**	**2,776445**	**3,746947**	**4,604095**	**8,610302**
5,00	**1,475884**	**2,015048**	**2,570582**	**3,364930**	**4,032143**	**6,868827**
6,00	**1,439756**	**1,943180**	**2,446912**	**3,142668**	**3,707428**	**5,958816**
7,00	**1,414924**	**1,894579**	**2,364624**	**2,997952**	**3,499483**	**5,407883**
8,00	**1,396815**	**1,859548**	**2,306004**	**2,896459**	**3,355387**	**5,041305**
9,00	**1,383029**	**1,833113**	**2,262157**	**2,821438**	**3,249836**	**4,780913**
10,00	**1,372184**	**1,812461**	**2,228139**	**2,763769**	**3,169273**	**4,586894**
11,00	**1,363430**	**1,795885**	**2,200985**	**2,718079**	**3,105807**	**4,436979**
12,00	**1,356217**	**1,782288**	**2,178813**	**2,680998**	**3,054540**	**4,317791**
13,00	**1,350171**	**1,770933**	**2,160369**	**2,650309**	**3,012276**	**4,220832**
14,00	**1,345030**	**1,761310**	**2,144787**	**2,624494**	**2,976843**	**4,140454**
15,00	**1,340606**	**1,753050**	**2,131450**	**2,602480**	**2,946713**	**4,072765**
16,00	**1,336757**	**1,745884**	**2,119905**	**2,583487**	**2,920782**	**4,014996**
17,00	**1,333379**	**1,739607**	**2,109816**	**2,566934**	**2,898231**	**3,965126**
18,00	**1,330391**	**1,734064**	**2,100922**	**2,552380**	**2,878440**	**3,921646**
19,00	**1,327728**	**1,729133**	**2,093024**	**2,539483**	**2,860935**	**3,883406**
20,00	**1,325341**	**1,724718**	**2,085963**	**2,527977**	**2,845340**	**3,849516**
21,00	**1,323188**	**1,720743**	**2,079614**	**2,517648**	**2,831360**	**3,819277**
22,00	**1,321237**	**1,717144**	**2,073873**	**2,508325**	**2,818756**	**3,792131**
23,00	**1,319460**	**1,713872**	**2,068658**	**2,499867**	**2,807336**	**3,767627**
24,00	**1,317836**	**1,710882**	**2,063899**	**2,492159**	**2,796940**	**3,745399**
25,00	**1,316345**	**1,708141**	**2,059539**	**2,485107**	**2,787436**	**3,725144**
26,00	**1,314972**	**1,705618**	**2,055529**	**2,478630**	**2,778715**	**3,706612**
27,00	**1,313703**	**1,703288**	**2,051831**	**2,472660**	**2,770683**	**3,689592**
28,00	**1,312527**	**1,701131**	**2,048407**	**2,467140**	**2,763262**	**3,673906**
29,00	**1,311434**	**1,699127**	**2,045230**	**2,462021**	**2,756386**	**3,659405**
30,00	**1,310415**	**1,697261**	**2,042272**	**2,457262**	**2,749996**	**3,645959**
40,00	**1,303077**	**1,683851**	**2,021075**	**2,423257**	**2,704459**	**3,550966**
60,00	**1,295821**	**1,670649**	**2,000298**	**2,390119**	**2,660283**	**3,460200**
120,00	**1,288646**	**1,657651**	**1,979930**	**2,357825**	**2,617421**	**3,373454**

Az F (Fisher) eloszlás táblázata

Szignifikancia szint 95%-os

$P(\xi > K_{0,05}) = 0{,}975$

f_1+1 a számláló korrigált empirikus szórásnégyzetéhez tartozó mintaelemszám

f_2 +1 a nevező korrigált empirikus szórásnégyzetéhez tartozó mintaelemszám

f_1 az oszlopok tetején, f_2 a sorok elején áll

A sűrűségfüggvény

$$f_\xi(x) = \frac{\Gamma\left(\frac{f_1+f_2}{2}\right)}{\Gamma\left(\frac{f_1}{2}\right)\Gamma\left(\frac{f_2}{2}\right)} x^{\frac{f_1}{2}-1}\left(f_2 + f_1 \cdot x\right)^{-\frac{f_1+f_2}{2}}, \quad x > 0, \text{ ahol } \Gamma(x) = \int_0^\infty e^{-t} t^{x-1}\, dt .$$

f1	1	2	3	4	5	6	7	8	9	10	12	16	20	24	30	50	100
f2																	
1	161,44	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	243,91	246,46	248,01	249,05	250,10	251,77	253,04
2	18,512	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,45	19,46	19,48	19,49
3	10,127	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,69	8,66	8,64	8,62	8,58	8,55
4	7,7086	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,84	5,80	5,77	5,75	5,70	5,66
5	6,6078	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,60	4,56	4,53	4,50	4,44	4,41
6	5,9873	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,92	3,87	3,84	3,81	3,75	3,71
7	5,5914	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,49	3,44	3,41	3,38	3,32	3,27
8	5,3176	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,20	3,15	3,12	3,08	3,02	2,97
9	5,1173	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	2,99	2,94	2,90	2,86	2,80	2,76
10	4,9646	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,83	2,77	2,74	2,70	2,64	2,59
11	4,8443	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,70	2,65	2,61	2,57	2,51	2,46
12	4,7472	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,60	2,54	2,51	2,47	2,40	2,35
13	4,6671	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,51	2,46	2,42	2,38	2,31	2,26
14	4,6001	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,44	2,39	2,35	2,31	2,24	2,19
15	4,5430	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,38	2,33	2,29	2,25	2,18	2,12
16	4,4939	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,33	2,28	2,24	2,19	2,12	2,07
17	4,4513	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,29	2,23	2,19	2,15	2,08	2,02
18	4,4138	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,25	2,19	2,15	2,11	2,04	1,98
19	4,3807	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,21	2,16	2,11	2,07	2,00	1,94
20	4,3512	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,18	2,12	2,08	2,04	1,97	1,91
21	4,3247	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,16	2,10	2,05	2,01	1,94	1,88
24	4,2596	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,09	2,03	1,98	1,94	1,86	1,80
26	4,2252	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,05	1,99	1,95	1,90	1,82	1,76
28	4,1959	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,02	1,96	1,91	1,87	1,79	1,73
32	4,1490	3,29	2,90	2,67	2,51	2,40	2,31	2,24	2,19	2,14	2,07	1,97	1,91	1,86	1,82	1,74	1,67
36	4,1131	3,26	2,87	2,63	2,48	2,36	2,28	2,21	2,15	2,11	2,03	1,93	1,87	1,82	1,78	1,69	1,62
40	4,0847	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,90	1,84	1,79	1,74	1,66	1,59
60	4,0011	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,82	1,75	1,70	1,65	1,56	1,48
100	3,9361	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,85	1,75	1,68	1,63	1,57	1,48	1,39
200	3,8883	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,80	1,69	1,62	1,57	1,52	1,41	1,32

A χ^2-eloszlás táblázata

$$P\left(\xi > K_\varepsilon\right) = \varepsilon$$

Az f szabadságfok a sorok elején olvasható, az ε szint az oszlopok tetején áll
A K_ε kritikus érték a megfelelő sor-oszlop kereszteződés cellájában áll

Sűrűségfüggvény

$$f_\xi(x) = \frac{1}{2^{\frac{f}{2}}\Gamma\left(\frac{f}{2}\right)} e^{-\frac{x}{2}} \cdot x^{\frac{f}{2}-1}\ , x > 0, \text{ ahol } \Gamma(x) = \int_0^\infty e^{-t} t^{x-1}\, dt$$

ε	0,99	0,98	0,90	0,80	0,70	0,50
szab.fok						
1	,0002	,0006	,0039	,0158	,1485	,4549
2	,0201	,0404	,1026	,2107	,7133	1,3863
3	,1148	,1848	,3518	,5844	1,4237	2,3660
4	,2971	,4294	,7107	1,0636	2,1947	3,3567
5	,5543	,7519	1,1455	1,6103	2,9999	4,3515
6	,8721	1,1344	1,6354	2,2041	3,8276	5,3481
7	1,2390	1,5643	2,1673	2,8331	4,6713	6,3458
8	1,6465	2,0325	2,7326	3,4895	5,5274	7,3441
9	2,0879	2,5324	3,3251	4,1682	6,3933	8,3428
10	2,5582	3,0591	3,9403	4,8652	7,2672	9,3418
11	3,0535	3,6087	4,5748	5,5778	8,1479	10,3410
12	3,5706	4,1783	5,2260	6,3038	9,0343	11,3403
13	4,1069	4,7654	5,8919	7,0415	9,9257	12,3398
14	4,6604	5,3682	6,5706	7,7895	10,8215	13,3393
15	5,2293	5,9849	7,2609	8,5468	11,7212	14,3389
16	5,8122	6,6142	7,9616	9,3122	12,6243	15,3385
17	6,4078	7,2550	8,6718	10,0852	13,5307	16,3382
18	7,0149	7,9062	9,3905	10,8649	14,4399	17,3379
19	7,6327	8,5670	10,1170	11,6509	15,3517	18,3377
20	8,2604	9,2367	10,8508	12,4426	16,2659	19,3374
21	8,8972	9,9146	11,5913	13,2396	17,1823	20,3372
22	9,5425	10,6000	12,3380	14,0415	18,1007	21,3370
23	10,1957	11,2926	13,0905	14,8480	19,0211	22,3369
24	10,8564	11,9918	13,8484	15,6587	19,9432	23,3367
25	11,5240	12,6973	14,6114	16,4734	20,8670	24,3366
26	12,1981	13,4086	15,3792	17,2919	21,7924	25,3365
27	12,8785	14,1254	16,1514	18,1139	22,7192	26,3363
28	13,5647	14,8475	16,9279	18,9392	23,6475	27,3362
29	14,2565	15,5745	17,7084	19,7677	24,5770	28,3361
30	14,9535	16,3062	18,4927	20,5992	25,5078	29,3360