Lektorálta
Dr. Obádovics Csilla
Dr. Szelényi László
Dr. Szelezsán János

ISBN: 963-9534-00-5

Kiadja a SCOLAR KIADÓ
1114 Budapest, Bartók Béla út 7.
Tel./fax: (06-1) 466-7648
E-mail: scolar@scolar.hu
www.scolar.hu
Felelős kiadó és felelős szerkesztő: Érsek Nándor
A borítót tervezte: Máthé Hanga
A könyv ábráit rajzolta: Érsek-Obádovics Robin, Szabó Béla

Nyomta a Dürer Nyomda Kft.
Felelős vezető: Megyik András

ELŐSZÓ

A valószínűségszámítás és a matematikai statisztika tárgyköre kb. 300 év óta folyamatosan, újabb és újabb eredményekkel gazdagszik. Eredményeit néhány évtized óta számos tudományterületen rendszeresen alkalmazzák különböző problémák megoldására. Folyóiratok és napilapok gazdasági, műszaki, mezőgazdasági, társadalomtudományi cikkeinek szerzői is gyakran alkalmaznak állításaik bizonyításához valószínűségszámítási és statisztikai módszereket. Közvéleménykutatási adatok feldolgozását, választások eredményeinek sok szempontú elemzését az újságolvasók széles köre rendszeresen olvassa. Félő, hogy elemi valószínűségszámítási és statisztikai ismeretek nélkül ezekből a cikkekből hamis következtetéseket vonnak le. Mindezek figyelembevételével szükségesnek tartjuk, hogy már a középiskolák minden típusában a valószínűségszámítás és a statisztika alapjaival megismertessék a tanulókat, mert az a gondolkodásmód, amely az említett cikkek megértéséhez nélkülözhetetlen, csak így alakulhat ki. Mivel a világ jelenségei nem determinisztikusak, hanem véletlenszerűek, ezért ki kell fejlesztenünk azt az érzéket, amely lehetővé teszi számunkra, hogy becsülni tudjuk a jelenségek bekövetkezésének valószínűségét, mert csak így dolgozhatjuk ki azokat a taktikai lépéseket, amelyek a következmények felerősítésére vagy éppen kivédésére, hatásuk csökkentésére alkalmasak.

E könyv a valószínűségszámítás és a matematikai statisztika elemeit tárgyalja a középiskolai ismeretekre támaszkodva. Fejezetei a főiskolai és egyetemi oktatási programokhoz illeszkednek. Kidolgozott példái, feladatai a gyakorlati élet különböző területeinek problémáit ölelik fel. A feladatok megoldásai a könyv III. részében találhatók. Az I. rész a valószínűségszámítást öt fejezetre, a II. rész pedig a matematikai statisztika elemeit négy fejezetre bontva tárgyalja. A fejezetekhez ún. „ellenőrző kérdések" tartoznak, amelyekkel az Olvasó ellenőrizheti tudását, ha a kérdéses fejezet tanulását befejezte. A kérdések természetesen csak a legfontosabb fogalmak definícióira, a további fejezetek megértéséhez nélkülözhetetlen tételekre, és az alkalmazás szempontjából fontos eljárásokra vonatkoznak. Így az az Olvasó, aki korábbi tanulmányai során már szerzett bizonyos ismereteket, az ellenőrző kérdések alapján megállapíthatja mennyire biztos a tudása, és ha azt kielégítőnek találja, akkor a következő fejezet ismereteinek elsajátítására térhet át. Javaslom, hogy az Olvasó a példákat és a feladatokat ilyen esetben is nézze át, ill. oldja meg, mert csak az önállóan jól megoldott feladatok adnak kellő biztonságérzést, a táblázatok használatában való jártasságot, amely nélkül nem lehet jó eredménnyel vizsgázni.

A könyvben való tájékozódást a jól tagolt tartalomjegyzéken kívül a nagyon részletes név- és tárgymutató teszi még könnyebbé. Mindazok, akik a könyvben foglaltakon túlmenő ismereteket kívánnak elsajátítani, az irodalomjegyzékben találnak megfelelő műveket igényeik kielégítésére.

Köszönetet mondok a könyv lektorainak, *Dr. Szelezsán János*nak és *Dr. Obádovics Csillá*nak a lektorálás fárasztó, de számomra igen sok segítséget nyújtó munkájáért és különösen *Érsek Nándor*nak, aki nélkül sem ez a könyv, sem a szép kiállítású *Matematika* könyvem nem jelenhetett volna meg.

Dr. Obádovics J. Gyula

Balatonszárszó, 1994. december havában

ELŐSZÓ A NEGYEDIK KIADÁSHOZ

Köszönetet mondok mindazoknak, akik felhívták figyelmemet a korábbi kiadásokban előforduló elírásokra és hibákra. Ez a negyedik kiadás néhány új témakörrel és számos – az alkalmazást és a tárgyalt módszer megértését elősegítő – példával gazdagodott.

Köszönetet mondok a negyedik kiadás lektorainak, *Dr. Szelényi László*nak és *Dr. Obádovics Csillá*nak igen gondos és precíz munkájukért.

Dr. Obádovics J. Gyula

Balatonszárszó, 2001. június havában

TARTALOMJEGYZÉK

GYAKRABBAN HASZNÁLT JELEK ÉS RÖVIDÍTÉSEK

$=$	egyenlő
$\neq$	nem egyenlő
$\equiv$	azonosan egyenlő
$\approx$	közelítőleg egyenlő
$:=$	definíció szerint egyelő
$<$	kisebb, mint
$>$	nagyobb, mint
$\leq$	kisebb vagy egyenlő
$\geq$	nagyobb vagy egyenlő
$+$	összeadás
$-$	kivonás
$\cdot$	szorzás
$:, /, —$	osztás
$\{a_1, a_2, \ldots\}$	a_1, a_2, ... elemek halmaza
$\{x \mid T(x)\}$	mindazoknak az x elemeknek a halmaza, amelyekre a $T(x)$ tulajdonság érvényes
$\emptyset$	üres halmaz
$\in$	eleme ...-nak (...-nek)
$\notin$	nem eleme ...-nak (...-nek)
$\subseteq$	... részhalmaza ...-nak (...-nek)
$\subset$	... valódi részhalmaza ...-nak (...-nek)
$\not\subset$	... nem részhalmaza ...-nak (...-nek)
$\cup$	halmazok egyesítésének jele
$\cap$	halmazok metszetének jele
$\setminus$	halmazok különbségképzésének jele
$\overline{A}$	A halmaz komplementere
$\times$	direkt vagy *Descartes*-féle szorzat jele
$\forall$	minden ...-ra (...-re)
$\exists$	létezik olyan ..., amelyre érvényes ...; van legalább egy olyan ..., hogy ...
(x_1, x_2)	rendezett pár
$A \times B$	A és B *Descartes*-szorzata

$(x_1, x_2, \ldots, x_n)$	szám n-es
$\rightarrow$	leképezés jele
$\mapsto$	hozzárendelés jele
f, g	függvények jele
$\max(a,b)$	az a, b számok közül a nagyobb
$\min(a,b)$	az a, b számok közül a kisebb
N	természetes számok halmaza
Z	egész számok halmaza
Q	racionális számok halmaza
R	valós számok halmaza
$\lvert a \rvert$	az a szám abszolút értéke
$[a,b]$, $]a,b[$	a-tól b-ig terjedő zárt ill. nyílt intervallum
%	százalék
!	faktoriális: pl. $4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$
e	a természetes logaritmus alapszáma
e^x	e alapú exponenciális függvény
$\binom{n}{k}$	binomiális együttható (n alatt a k) $= \frac{n!}{k!(n-k)!}$
Σ	összegezés jele, pl. $\sum_{i=1}^{n} a_i = a_1 + a_2 + \ldots + a_n$
P_n	n elem ismétlés nélküli permutációinak száma
$P_n^{(k_1, k_2, \ldots, k_r)}$	n elem ismétléses permutációinak száma
$V_{n,k}$	n különböző elem ismétlés nélküli k-ad osztályú variációinak száma

$V_{n,k}^{i}$	n elem k-ad osztályú ismétléses variációinak száma
$C_{n,k}$	n elem k-ad osztályú ismétlés nélküli kombinációinak száma
$C_{n,k}^{i}$	n elem k-ad osztályú ismétléses kombinációinak száma
I	biztos esemény
$\emptyset$	lehetetlen esemény
$P(A)$	az A esemény valószínűsége
$P(A \mid B)$	az A valószínűsége a B feltétel mellett (feltételes valószínűség)
$M(X)$, m	X valószínűségi változó várható értéke
$D(X)$, σ	X valószínűségi változó szórása
Var X, $D^2(X)$, σ^2	variancia, szórásnégyzet
$r_{A\|B}$	feltételes relatív gyakoriság
$v(x_i), v(y_i)$	peremeloszlások
$w(x_i, y_j)$	együttes eloszlás
Cov (X,Y)	kovariancia
R(X,Y), r	korreláció, korrelációs együttható
$f(x)$, $\varphi(x)$	sűrűségfüggvény
$F(x)$, $\Phi(x)$	eloszlásfüggvény
$M(X^k)$	X változó k-adik momentuma
$b(k;n,p)$	binomiális eloszlás
$p(k;\lambda)$	*Poisson*-eloszlás

$N(m,\sigma)$	normális eloszlás
$N(0,1)$	standard normális eloszlás
$\overline{X}$, $\overline{x}$	mintaközép, számtani közép
$F_n^{(r)}(x)$	tapasztalati eloszlásfüggvény
$g_i = \frac{f_i}{n}$	relatív gyakoriság
s, s^*	tapasztalati és korrigált tapasztalati szórás
s^2, s^{*2}	tapasztalati és korrigált tapasztalati szórásnégyzet
u	az u-próba értéke
t	a t-próba értéke
F	az F-próba értéke
χ^2	a χ^2-próba értéke
p	valószínűségi vagy szignifikanciaszint
$[-u_p;u_p]$	konfidencia intervallum
u_{sz}, u_t	számított és táblázatbeli érték jelölése
H_0, H	nullhipotézis és ellenhipotézis jele
μ_k	k-adik centrális momentum
γ_1	ferdeségi együttható
γ_2	lapultsági együttható
m_t	valószínűségi változó várható eltérése
m_e	medián (az eloszlás közepe)
m_t	valószínűségi változó terjedelme
m_d	módusz
x_p	eloszlásfüggvény p-kvantilise

Görög ábécé

Α α	alfa	Ι ι	ióta	Ρ ρ	ró
Β β	béta	Κ κ	kappa	Σ σ	szigma
Γ γ	gamma	Λ λ	lambda	Τ τ	tau
Δ δ	delta	Μ μ	mű	Υ υ	üpszilon
Ε ε	epszilon	Ν ν	nű	Φ ϕ	fí
Ζ ζ	zéta	Ξ ξ	kszí	Χ χ	khí
Η η	éta	Ο ο	omikron	Ψ ψ	pszí
Θ ϑ	théta	Π π	pí	Ω ω	ómega

VALÓSZÍNŰSÉGSZÁMÍTÁS

Esemény, eseménytér
Műveletek eseményekkel
A valószínűség és axiómái
Kombinatórikai összefoglaló
Feltételes valószínűség
A teljes valószínűség tétele
Valószínűségi változók
    Diszkrét és folytonos valószínűségi változó
        Várható érték, szórásnégyzet és szórás
        Együttes- és peremeloszlások
        Eloszlás- és sűrűségfüggvény
        Módusz, medián, *p*-kvantilis
Nevezetes eloszlások
Csebisev-egyenlőtlenség
A nagy számok törvénye

Fejezetenként ellenőrző kérdésekkel és feladatokkal

I. RÉSZ

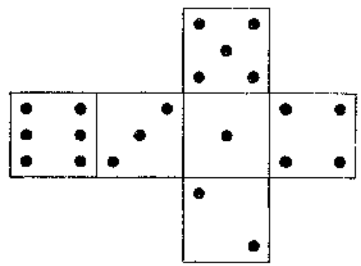
VALÓSZÍNŰSÉGSZÁMÍTÁS

ELSŐ FEJEZET

1.1. Bevezetés

Az emberiség már jóval korábban a valószínűségszámítás matematikai megalapozása előtt, a környezetében ismétlődően lejátszódó jelenségek megfigyelése alapján felkészült bizonyos események bekövetkezésére. Ilyenek voltak pl. az időjárásra, a folyók vízállására, a szerencsejátékokra vonatkozó megfigyelések.

Tekintsünk pl. egy anyagában homogén, szabályos pontozású dobókockát (a továbbiakban röviden kockát, 1. ábra). A kocka sokszori feldobásakor azt tapasztaljuk, hogy mindegyik lapja az összes dobás kb. 1/6-ában lesz felül.



1. ábra. Szabályos kocka lapjai

Ezt a tényt úgy fejezzük ki, hogy a lehetséges hat eset mindegyikének (a kocka egy adott lapja felülre kerülésének a valószínűsége 1/6. Általában egy véletlen esemény valószínűségén azt a számot értjük, amely megadja, hogy a szóban forgó esemény az eseteknek kb. hányadrészében következik be. A szerencsejátékokkal kapcsolatos véletlen események matematikai vizsgálatával elsőként *H. Cardano* (1501–1576), *G. Galilei* (1564–1642), *B. Pascal*, (1623–1662), *P. Fermat* (1601–1665) és *Ch. Huygens* (1629–1695) foglalkozott.

Egy szenvedélyes szerencsejátékos, *Ch. de Méré* a következő két problémával fordult *Pascalhoz* (1654):

1. Mi az oka annak, hogy egy kockával dobva előnyös arra fogadni, hogy az első négy dobás között megjelenik a hatos, de hátrányos arra fogadni, hogy két kockával dobva az első 24 dobás valamelyikében megjelenik a két hatos?

2. Két játékos megállapodik abban, hogy az nyeri a kitűzött pénzösszeget, aki először nyer k számú játszmát. Hogyan kell méltányosan felosztani a kitűzött pénzösszeget, ha valamilyen okból akkor kellett abbahagyniuk a játékot, amikor az első játékosnak még $n < k$, a másodiknak pedig még $m < k$ győzelem hiányzik?

A problémák korábban is ismertek voltak, de helyes általános megoldást *B. Pascal* és *P. Fermat* adtak 1654-ben. Az említett problémákkal *Ch. Huygens* is megismerkedett, és a megoldásokat a *„De ratiociniis in ludo aleae"* c. értekezésében rögzítette, amely 1657-ben jelent meg. Jelentős eredmények fűződnek *J. Bernoulli* (1654–1705) munkásságához, aki felismerte, hogy a valószínűségszámítás jól alkalmazható a természet és társadalom véletlen jelenségeinek vizsgálatára. *Ars conjectandi* (A sejtés művészete) c. művében a nagy számok törvénye is megfogalmazásra került.

A 18. században *A. de Moivre* (1667–1754), *Th. Bayes* (1702–1754), *G. L. L. Buffon* (1707–1788), *J. L. Lagrange* (1736–1813) és mások bővítik a valószínűségszámítást új eredményekkel.

A valószínűségszámítás klasszikus elméletét *P. S. Laplace* foglalta össze. Abban az időben a valószínűséget, a kedvező esetek és az összes esetek számának hányadosaként definiálták, amely csak az ún. „egyenlően valószínű" esetekre teljesül.

Pl. annak valószínűsége, hogy egy kockával páros számot dobunk $\frac{3}{6}=\frac{1}{2}$, mivel a kedvező esetek a 2, 4, 6, vagyis három, az összes eset pedig: 1, 2, 3, 4, 5, 6, azaz hat.

A 19. század elejétől a valószínűségszámítás igen gyors fejlődésnek indult *S. D. Poisson* (1781–1840), *K. F. Gauss* (1777–1855), *P. L. Csebisev* (1821–1894) munkássága következtében. A valószínűségszámítás töretlen fejlődését akadályozta az egzakt matematikai megalapozottságának hiánya. A valószínűségszámítás matematikai megalapozásával foglalkoztak a 20. században pl. *Sz. N. Bernstein* (1880–1969) és *R. Mises* (1883–1953), azonban modern elméletét *A. N. Kolmogorov* dolgozta ki. *Kolmogorov E. Borelnek* (1871–1956), *Jordán Károlynak* (1871–1959) és másoknak az eredményeit továbbfejlesztve, a halmaz- és mértékelméletre alapozva 1933-ban alkotta meg a valószínűségszámítás axiomatikus elméletét. *Kolmogorov* elmélete elősegítette a valószínűségszámítás ugrásszerű fejlődését, mivel megszüntette azt a bizonytalanságot, amit az alapok tisztázatlansága okozott. Munkásságát követően a valószínűségszámítás feladatának megfogalmazása is módosult.

A valószínűségszámítás feladata olyan *mérték* bevezetése, amely a bizonytalanságot numerikusan méri, s erre alapozva olyan matematikai módszerek kidolgozása, amelyekkel bizonyos események (véletlen tömegjelenségek) valószínűsége kiszámítható.

A továbbiakban **kísérletnek** nevezzük az olyan tömegjelenség megfigyelését, amelynek lefolyása a véletlentől függ. A kísérlet különböző *kimeneteleit* a megfigyelés szempontjából kiválasztott mennyiség ill. a kísérlet eredményeit jellemző információ megkülönböztethető értékei alkotják.

Pl. egy kocka n-szeri feldobásánál megfigyeljük hányszor dobunk páros számot, egy célgépen a gyártandó alkatrész méreteinek beállítása után megfigyeljük az elkészült alkatrészek tényleges méreteit.

Gyakran bizonyos feltételek által meghatározott korlátokkal, matematikai eszközökkel leírt jelenség, ún. **modell** megfigyelését végezzük, s a modell megfigyelése alapján következtetünk a modell jóságára. Általában két alapvető modellt különböztetünk meg: a *determinisztikus* és a *nem-determinisztikus* vagy *sztochasztikus* modellt.

Determinisztikus modellnek nevezzük azt, amelynél a megfigyelés szempontjából kiválasztott feltételek egyértelműen meghatározzák az eredményt. Ezzel összhangban egy jelenségről is azt mondjuk, hogy determinisztikus, ha azonos feltételek megléte esetén a jelenség mindig ugyanúgy játszódik le és szükségszerűen bekövetkezik.

Sztochasztikus modellnek pedig azt nevezzük, amelynél a megfigyelt feltételek alapján nem lehet egyértelműen meghatározni a kísérlet kimenetelét. Egy jelenségről is azt mondjuk, hogy *véletlenszerű* vagy sztochasztikus, ha a figyelembe vehető feltételek nem határozzák meg egyértelműen a jelenség kimenetelét. Természetesen ez nem azt jelenti, hogy a jelenségnek nem lennének meg az okai, hanem csak azt, hogy nem ismerjük az okok összességét, vagy egyes okokat nem veszünk figyelembe vizsgálataink során. A valószínűségszámítás témakörében végzett vizsgálatokhoz sztochasztikus modellt alkalmazunk.

Pl. egy szabályos pénzérme feldobásakor a „fej" vagy „írás" eredményt meg tudnánk mondani, ha e jelenséggel kapcsolatos összes befolyásoló tényezőt számításba vennénk, azaz ha figyelembe vennénk a feldobás előtti helyzetét, a dobás magasságát, a pörgés gyorsaságát stb. Ha viszont csak azt vesszük figyelembe, hogy a pénzérmét feldobtuk, és az eredményt befolyásoló tényezőket nem, akkor a dobás eredménye véletlenszerűen lehet „fej" vagy „írás". A pénzérme feldobás kísérletét elegendő sokszor megismételve azt tapasztaljuk, hogy közel fele arányban lesz „fej", ill. „írás". Ezt úgy mondjuk, hogy $\frac{1}{2}$ valószínűséggel lesz a dobás kimenetele „fej" vagy „írás".

1.2. Esemény, eseménytér

Mit nevezünk eseménynek, eseménytérnek?

Ebben a szakaszban felhasználjuk a halmazelmélet elemi ismereteit, a halmazok egyesítésének és a metszet képzésének műveleteit, valamint a kiegészítő halmaz képzésének műveletét.

Kísérletek végrehajtása ill. jelenségek megfigyelése előtt mindig meg kell határoznunk, hogy mit tekintünk lehetséges kimenetelnek. Általában feltesszük, hogy a lehetséges kimenetelek halmaza minden kísérlettel kapcsolatban megadható.

Definíció. Egy kísérlet egyes megkülönböztethető véletlentől függő, lehetséges kimeneteleit (a megfigyelés eredményeit), **elemi eseménynek,** az elemi események összességét, halmazát pedig **eseménytérnek** nevezzük. Az eseménytér bármely részhalmazát **eseménynek** nevezzük. Egy A esemény elemi esemény, ha nem állítható elő tőle különböző események összegeként.

Az eseményteret jelöljük Q-val, az elemi eseményeket pedig az ábécé többi indexes vagy index nélküli dölt nagybetűivel. A Q helyett szokásos jelölés a görög nagy omega, azaz Ω.

Egy kocka feldobásával végzett kísérletnél hat elemi eseményt különböztetünk meg, éspedig a kocka felülre kerülő lapján megszámlált pontok száma alapján. A hat elemi esemény tehát A_1, A_2, A_3, A_4, A_5, A_6, amelyek az 1, 2, 3, 4, 5 és 6-os dobásnak felelnek meg, és így az elemi események Q-val jelölt eseménytere:

$$Q=\{A_1,A_2,A_3,A_4,A_5,A_6\} \text{ ill. } Q=\{1,2,3,4,5,6\}$$

halmazzal adható meg.

A Q eseménytér valamely részhalmazát, azaz *bizonyos kimenetelek összességét,* amelyek bekövetkezése a kérdéses esemény bekövetkezését vonja maga után, az eseménnyel azonosítjuk.

Legyen pl. A a páratlan pontszámok, B a páros pontszámok és C a prímszámok dobásának eseménye, akkor ezen eseményeket rendre $\{1,3,5\}$, $\{2,4,6\}$ és $\{2,3,5\}$ halmazok jelölik, azaz

$$A=\{1,3,5\},\ B=\{2,4,6\},\ C=\{2,3,5\},$$

tehát az A, B és C események a $Q=\{1,2,3,4,5,6\}$ eseménytér részhalmazai, azaz

$$A\subset Q,\ B\subset Q,\ C\subset Q.$$

Ha pl. a kísérlet kimenetele 1, 3 vagy 5, akkor azt mondjuk, hogy az A esemény bekövetkezett. A B esemény bekövetkezik, ha a 2, 4, 6 számok közül bármelyiket dobjuk, és a C esemény bekövetkezik, ha a 2, 3, 5 számok közül bármelyiket dobjuk.

Valamely kísérletnél **biztos esemény** az, amely a kísérlet során minden esetben bekövetkezik, jele: I, (szokás Ω-val is jelölni); **lehetetlen esemény** pedig az, amely a kísérlet során sohasem következhet be, jele: $\varnothing$. A Q teljes eseménytér a biztos eseménynek felel meg, hiszen az összes elemi eseményt tartalmazza, az $\varnothing$ üres halmaz az eseménytér egyetlen elemét sem tartalmazza, ezért akármilyen kimenetele van a kísérletnek a $\varnothing$ esemény nem következhet be.

A 6 elemből álló eseménytérnek a biztos és a lehetetlen eseményt is számbavéve 2^6 részhalmaza van.

Jelölje D a páros vagy prímszámok dobásának eseményét. A D esemény bekövetkezik, ha a 2, 4, 6 vagy a 2, 3, 5 elemi események valamelyike bekövetkezik, azaz ha a B vagy a C esemény bekövetkezik, tehát a D esemény a $B\cup C$ halmazművelettel adható meg:

$$D=B\cup C=\{2,3,4,5,6\}$$

A és B esemény összege tehát az az esemény, amely akkor következik be, ha A és B esemény közül legalább az egyik bekövetkezik.

Jelölje E a páratlan prímszámok dobásának eseményét, akkor az E esemény az A és C halmazok metszete, azaz

$$E = A \cap C = \{3,5\}$$

A és B esemény szorzata tehát az az esemény, amely akkor következik be, ha az A és B események mindegyike bekövetkezik.

Definíció. Egy A esemény **komplementere (ellentéte),** jele: $\overline{A}$ (olv.: á felülvonás), az az esemény, amely akkor és csak akkor következik be, ha A nem következik be. $\overline{A}$ esemény tehát az A ellentett eseménye.

Pl. azt az eseményt, hogy nem dobunk prímszámot a C halmaznak a Q teljes halmazra vonatkozó $\overline{C}$ (olv.: c felülvonás) komplementerével fejezhetjük ki, azaz

$$\overline{C} = \{1,4,6\}$$

A páratlan és a páros számok dobása egyidejűleg nem következhet be, az A és B **diszjunkt halmazok,** ezért az $A \cap B$ üres halmaz, lehetetlen esemény, azaz

$$A \cap B = \varnothing .$$

Definíció. Két eseményt akkor mondunk **egyenlőnek,** ha bármelyik bekövetkezése maga után vonja a másik esemény bekövetkezését is.

Összefoglalva: *eseményeknek* a Q halmaz részhalmazait tekintjük, ill. azzal azonosítjuk. *Lehetetlen esemény* az, amely soha nem következhet be, *biztos esemény* pedig az, amely mindig bekövetkezik. Az A *elemi esemény*, ha nem lehetetlen esemény és nem bontható fel A-tól különböző események összegére. Az A *összetett esemény* ha legalább két, tőle független esemény összegeként állítható elő. Egy kísérlet Q halmazát *eseménytérnek* nevezzük, ha Q minden eleme a kísérlet egy lehetséges kimenetelét jelöli, és a kísérlet bármely végrehajtása során a megfigyelt kimenetel Q pontosan egy elemének felel meg.

Azokat az eseményeket, amelyekről a kísérlet eredményéből egyértelműen eldönthető a bekövetkezésük vagy nem bekövetkezésük, **megfigyelhető eseményeknek** mondjuk.

Pl. a teljesen egyforma kocka feldobásával végzett kísérletsorozatnál a $6+5$ dobás, azaz 11 összeg dobásának eseménye egyértelműen megfigyelhető, de azt nem tudjuk eldönteni, megfigyelni, hogy melyik kockával dobtuk a 6-ost és melyikkel az 5-öst. Így a nem megkülönböztethető két kocka feldobásával végzett kísérlet lehetséges kimeneteleinek száma 36, azaz

$$Q = \{(1,1),(1,2),\dots,(1,6),(2,1),(2,2),\dots,(2,6),\dots,(6,1),(6,2),\dots,(6,6)\}$$

1.2.1. Műveletek eseményekkel

Egy kísérlet eseményeire alkalmazva a halmazműveleteket, miként azt az előző pontban példákkal szemléltettük, új eseményeket hozhatunk létre. Ebben a pontban a halmazműveleteknek megfelelő, – a valószínűségszámításban általánosan elfogadott – az elemi matematikában használt műveleti jelekkel definiáljuk az eseményekre vonatkozó műveleteket.

Legyen A és B két esemény, akkor ezen események

a) $A \cup B$ -vel, ill. $A + B$ -vel jelölt **összegén** azt az eseményt értjük, hogy az A vagy B esemény közül legalább az egyik bekövetkezik;

Pl. $\{1,2\} \cup \{2,3\} = \{1,2,3\}$

b) $A \cap B$ -vel, ill. AB-vel (ill. $A \cdot B$ -vel) jelölt **szorzata** az az esemény, hogy mind az A mind a B esemény, azaz mindegyik bekövetkezik;

Pl. $\{1,2\} \cap \{2,3\} = \{2\}$

c) $B \setminus C$, ill. $B - C$ **különbség** az az esemény, amely akkor következik be, ha a B esemény bekövetkezik, de a C nem;

Pl. $\{1,2\} \setminus \{2,3\} = \{1\}$

E definíció értelmében $B = \{2,4,6\}$ és $C = \{2,3,5\}$ kockadobás eseményeinek $B - C = \{4,6\}$ különbségét, B-nek és C ellentett eseményének szorzataként is definiálhatjuk. Pl.

$$B \setminus C = B \cap \overline{C} = \{2,4,6\} \cap \{1,4,6\} = \{4,6\}.$$

Megjegyzés

Az események összetett kifejezésében előforduló zárójeles eseménykülönbségek általában nem kezelhetők az elemi algebra szabályai szerint.

Pl. $A+(B-C)$, $(A+B)-C$, $(A-C)+B$ általában nem egyenlők, ugyanakkor az $A-(B+C)=(A-B)-C$ egyenlőség mindig teljesül.

Pl. jelentse B a kettővel osztható szám dobását, C pedig a háromnál nagyobb szám dobását. Ekkor $B\setminus C$ a háromnál nem nagyobb páros szám dobását jelenti, azaz a kettes számot.

Annak jelölésére, hogy az A esemény *maga után vonja* a B esemény bekövetkezését $A\subseteq B$ szimbólumot használjuk, amely ekvivalens az $A\cdot B=A$, ill. $A+B=B$ kifejezésekkel.

Definíció. Az A és B eseményeket **egymást kizáróknak** nevezzük, ha egyszerre nem következhetnek be, azaz ha szorzatuk a lehetetlen esemény, vagyis $A\cap B=\varnothing$.

Pl. az A és $\overline{A}$ egymást kizáró események.

Definíció. Legyen $A_1, A_2, \ldots, A_n$, $A_k\neq\varnothing$ $(k=1,2,\ldots,n)$ az események olyan halmaza, amelyre $A_i\cdot A_j=\varnothing$ $(i,j=1,2,\ldots,n;$ $i\neq j)$ és $A_1+A_2+\ldots+A_n=I$ teljesül, azaz összegük (uniójuk) a biztos esemény és páronkénti szorzatuk (metszetük) a lehetetlen esemény. Ekkor a

$$Q=\{A_1, A_2, \ldots, A_n\}$$

eseményhalmazt **teljes eseményrendszernek** nevezzük.

A teljes eseményrendszer eseményei közül – a definíció értelmében, az eseményekre vonatkozó kísérlet során – mindig egy esemény következik be, és nem több.

Pl. Legyen A az az esemény, hogy a kockával 4-nél kisebb számot, B pedig az, hogy 3-nál nagyobb számot dobunk, akkor a $Q=\{A,B\}$ teljes eseményrendszer, közülük legalább az egyik mindig bekövetkezik, azaz $A\cup B=I$ és egymást kizáró események, azaz $A\cap B=\varnothing$.

Megjegyzések

1. Az eseményekre definiált műveletek műveleti jeleit a valószínűségszámítás témakörében az algebrában használt műveleti jelekkel szokás helyettesíteni és azt mondjuk, hogy egy kísérlet eseményhalmazához tartozó események a bevezetett műveletekkel **eseményalgebrát** alkotnak.

Könnyen belátható, hogy az eseményalgebra műveleti szabályai a halmazelmélet műveleti szabályainak felelnek meg, és így a halmazalgebrai azonosságokkal azonosak az eseményalgebrai azonosságok. Legyenek A, B, C tetszőleges események, akkor fennállnak az alábbi azonosságok:

a) 1. $A+A=A$
2. $A+B=B+A$ (az összeadás kommutatív)
3. $A+(B+C)=(A+B)+C$ (az összeadás asszociatív)

b) 1. $AA=A$ (a szorzás idempotens)
2. $AB=BA$ (a szorzás kommutatív)
3. $A(BC)=(AB)C=ABC$ (a szorzás asszociatív)

c) 1. $\overline{A}+A=I$, ill. $\overline{A}+A=Q$
2. $A+\varnothing=A$
3. $A+I=I$, ill. $A+Q=Q$

d) 1. $\overline{A}A=\varnothing$
2. $A\varnothing=\varnothing$
3. $AI=A$, ill. $AQ=A$

e) 1. $A(B+C)=AB+AC$ (A szorzás és összeadás két
2. $A+BC=(A+B)(A+C)$ disztributív törvénye)

f) De Morgan féle szabályok: Ha A és B események komplementerei $\overline{A}$ és $\overline{B}$, akkor

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

azaz *események összegének ellentettje az egyes események ellentettjének szorzatával, és események szorzatának ellentettje az események ellentettjének összegével egyenlő.*

2. Tetszőleges elemekből álló halmazt, amelyben két kétváltozós művelet $A+B$ és AB, valamint egy egyváltozós művelet $\overline{A}$ van értelmezve, továbbá van egy kitüntetett I eleme ($\overline{I}=\varnothing$), és az elemek eleget tesznek az $a)-e)$ axiómáknak *Boole-algebrának* nevezzük (*George Boole* (1815–1864) angol matematikus és filozófus). A halmazok és az események tehát ugyanannak az algebrának, a *Boole*-algebrának tesznek eleget.

Példák

1. Egy gázturbina működik, ha a hozzá csatlakozó három cső bármelyikén áramlik gáz. Az eseményalgebrai azonosságok felhasználásával írjuk le a következő eseményeket, ha A_1, A_2, A_3 esemény jelenti azt, hogy az 1., 2., ill. 3. csövön áramlik gáz:

a) a turbina nem kap gázt,

b) legalább egy cső hibás,

c) pontosan két cső hibátlan.

Megoldás

a) Azt az eseményt, hogy legalább egy csövön áramlik gáz a három esemény összege fejezi ki: $A_1 + A_2 + A_3$. Azt, hogy a turbina nem kap gázt, ennek ellentett eseménye fejezi ki, azaz

$$\overline{A_1 + A_2 + A_3}.$$

b) A legalább egy cső hibás esemény azt jelenti, hogy vagy az 1. vagy a 2. vagy a 3. cső hibás. Ezt az eseményt pl. a *De Morgan*-szabály ismételt alkalmazásával adhatjuk meg:

$$\overline{A_1} + \overline{A_2} + \overline{A_3} = \overline{A_1 A_2 A_3}.$$

c) A pontosan két cső hibátlan esemény azt jelenti, hogy vagy az 1. cső hibás és a másik kettő hibátlan ($\overline{A_1} A_2 A_3$ esemény), vagy a 2. hibás és a másik kettő hibátlan ($A_1 \overline{A_2} A_3$ esemény), vagy a 3. hibás és a másik kettő hibátlan ($A_1 A_2 \overline{A_3}$ esemény) és így a pontosan két cső hibátlan eseményt a három esemény összege adja:

$$\overline{A_1} A_2 A_3 + A_1 \overline{A_2} A_3 + A_1 A_2 \overline{A_3}.$$

2. Egy üzemben a raktárból piros szalagon is és zöld szalagon is beszállítható az alkatrész. Azt az eseményt, hogy egy adott műszakban van beszállítás a piros szalagon jelöljük A-val, azt pedig, hogy a zöld szalagon van beszállítás B-vel. Szavakkal fejezzük ki a következő eseményeket:

a) AB;	*b)* $A+B$;	*c)* $\overline{B}$;
d) $B-A$;	*e)* $\overline{A}B$;	*f)* $\overline{AB}$;
g) $\overline{A+B}$;	*h)* $\overline{A}\overline{B}$;	*i)* $AB+\overline{A}\overline{B}$.

Megoldás

a) Mindkét szalagon van beszállítás;

b) Legalább az egyik szalagon van beszállítás;

c) A zöld szalagon nincs beszállítás, de a piros szalagon lehet, hogy van, de az is lehet, hogy nincs;

d) A zöld szalagon van beszállítás, de a piroson nincs;

e) A piros szalagon nincs, de a zöld szalagon ugyanakkor van beszállítás;

f) Az biztos, hogy mindkét szalagon nem folyik beszállítás, lehet, hogy egyiken igen, vagy egyiken sem;

g) Egyik szalagon sem folyik beszállítás;

h) Egyik szalagon sem folyik beszállítás ($\overline{A+B} = \overline{A}\overline{B}$);

i) Vagy mindkét szalagon van beszállítás vagy egyiken sincs.

3. Egymás után kétszer feldobunk egy kockát. Azt az eseményt, hogy az első dobás pontszáma páros, jelölje A, azt, hogy a második dobás páros, pedig B. Legyen C az az esemény, hogy a két szám szorzata páros, D pedig az az esemény, hogy szorzatuk páratlan. Adjuk meg a C és D eseményt az A és B eseményekkel.

Megoldás

Minthogy két szám szorzata mindig páros, ha legalább az egyik szám páros, ezért $C = A + B$. Két szám szorzata csak akkor páratlan szám, ha mindkét szám páratlan, ami az A-nak is és B-nek is az ellentett eseményeként áll elő, tehát D előállításához alkalmazhatjuk a *De Morgan*-szabályt: $D = \overline{C} = \overline{A+B} = \overline{A} \cdot \overline{B}$.

1.3. A valószínűség és axiómái

Az eseményalgebrában láttuk, hogy egy kísérletet megadhatunk a kísérlet lehetséges kimeneteleinek Ω halmazával, azaz az eseménytérrel. Az eseménytér részhalmazaira előírt műveletekkel eseményalgebrát alkottunk és így azt matematikailag kezelhetővé tettük. Most azt fogjuk megvizsgálni, hogy valamely kísérlet során miként lehet egy A esemény bekövetkezésének esélyét, valószínűségét egy valós számmal megadni.

1.3.1. Gyakoriság, relatív gyakoriság

Egy A esemény bekövetkezésének esélyére akkor nyerhetünk megbízható információt, ha az A-ra vonatkozó kísérletet sokszor megismételjük, és megfigyeljük, hogy az A esemény hányszor következett be. Tegyük fel, hogy egy n-szer elvégzett kísérletnél az A esemény k-szor következett be $(0 \le k \le n)$. A k számot az A esemény **gyakoriságának** nevezzük.

Az n-szer elvégzett kísérletben az A bekövetkezésének arányát a $\frac{k}{n}$ szám fejezi ki. A $\frac{k}{n}$ számot az A esemény **relatív gyakorisá-**

gának nevezzük és g_A-val jelöljük. Mivel az A bekövetkezésére csak $k = 0, 1, 2, ..., n$, jöhet számításba, ezért

$$0 \leq \frac{k}{n} \leq 1 \tag{1}$$

egyenlőtlenség teljesül a relatív gyakoriságra.

A kísérletet ismételten sokszor elvégezve azt tapasztalhatjuk, hogy a relatív gyakoriság egy bizonyos szám körül ingadozik, és ez az ingadozás az n növekedésével egyre kisebb, viszonylagos stabilitása figyelhető meg.

Definíció. Azt a számértéket, amely körül valamely A esemény relatív gyakoriságának ingadozása viszonylagos stabilitást mutat, az esemény **valószínűségének** nevezzük és $P(A)$-val jelöljük.

Minthogy minden Q-ba tartozó A eseményhez hozzárendelhetünk egy az előbbiek szerint értelmezett valós számot, ezért azt mondjuk, hogy definiálható egy P függvény, amelynek $P(A)$ helyettesítési értéke adja az A esemény valószínűségét. A relatív gyakoriság (1) egyenlőtlensége alapján az A esemény valószínűségére is teljesül a

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

egyenlőtlenség. Ha az A a biztos esemény, azaz I, akkor $k = n$, tehát a relatív gyakoriság értéke ebben az esetben 1, így $P(I) = P(Q) = 1$, és a lehetetlen eseményre $k = 0$, tehát a relatív gyakoriság ebben az esetben 0, így a lehetetlen esemény valószínűsége $P(\emptyset) = 0$.

Legyenek $A_1, A_2, ..., A_n$ egymást páronként kizáró események, akkor

$$k_{A_1+A_2+\ldots+A_n} = k_{A_1} + k_{A_2} + \ldots + k_{A_n},$$

ezért

$$g_{A_1+A_2+\ldots+A_n} = \frac{k_{A_1+A_2+\ldots+A_n}}{n} = \frac{k_{A_1}}{n} + \frac{k_{A_2}}{n} + \ldots + \frac{k_{A_n}}{n},$$

amelyből következik, hogy

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

Mivel a Q eseménytér A eseményének valószínűsége az a $P(A)$ valós szám, amely körül a relatív gyakoriság ingadozik, ezért elfogadhatjuk, hogy a relatív gyakoriságra vonatkozó tulajdonságok az esemény valószínűségére is érvényesek, melynek matematikai megfogalmazását a következő pontban adjuk meg.

Példa

Dobjunk fel két szabályos kockát. Mennyi annak a valószínűsége, hogy a dobott pontszámok összege 5?

Megoldás

Az elemi események száma 36, mivel az egyik kocka minden lapjához a másik kocka hat lapja társulhat. Az elemi események közül az 5-ös dobás A eseménye: 1,4; 2,3; 3,2; 4,1 dobásokkal következik be. Mivel ezek az elemi események egyenlően valószínűek, valamint $k = 4$ és $n = 36$, így az A esemény valószínűsége: $P(A) = \frac{4}{36} = \frac{1}{9}$.

1.3.2. A valószínűség matematikai fogalma

A kísérletek eredményeihez, azaz az események mindegyikéhez egy valós számot rendelünk, vagyis a nem üres Q halmazon meghatározunk egy valós értékű P függvényt. A P-t **valószínűségi függvénynek**, $P(A)$-t pedig az $A \subset Q$ esemény **valószínűségének** nevezzük, ha teljesülnek az alábbi axiómák:

I. Minden $A \subset Q$ eseményre

$$0 \leq P(A) \leq 1,$$

azaz $P(A)$ nemnegatív, egynél nem nagyobb valós szám; más szóval minden esemény valószínűsége 0 és 1 között van;

II. $$P(Q) = 1,$$

azaz a biztos esemény valószínűsége 1;

III. Ha A és B egymást kizáró események, vagyis $A \cdot B = \emptyset$, akkor

$$P(A + B) = P(A) + P(B),$$

azaz az egymást kizáró események összegének valószínűsége egyenlő ezen események valószínűségeinek összegével.

IV. Ha az $A_1, A_2, ..., A_n, ...$ egymást páronként kizáró események sorozata, azaz $A_i \cdot A_j = \varnothing$, ha $i \neq j$, akkor

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

Ezeket az axiómákat a valószínűség *Kolmogorov-féle axiómáinak* nevezzük.

Megjegyzés

Ha a Q eseménytér véges, akkor a IV. axióma felesleges.

Az axiómákból a valószínűségszámítás matematikai elmélete tisztán matematikai eszközökkel kidolgozható. Ennek szemléltetésére néhány fontosabb tételt – bizonyítással együtt – ismertetünk.

1. Tétel. Ha $\varnothing$ az üres halmaz, akkor $P(\varnothing) = 0$.

Ui. legyen A egy tetszőleges halmaz, akkor az A és $\varnothing$ diszjunkt halmazok és $A \cup \varnothing = A$. A III. axióma alapján

$$P(A) = P(A + \varnothing) = P(A) + P(\varnothing),$$

s ez csak a $P(\varnothing) = 0$ esetén állhat fenn, vagyis a *lehetetlen esemény valószínűsége* 0.

2. Tétel. Ha az $A_1, A_2, ..., A_n$ események egymást páronként kizárják, azaz $A_i \cdot A_j = \varnothing$, ha $i \neq j$, akkor

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

Ui. a III. axióma e tétel $n = 2$-nek megfelelő esete, így azt felhasználva a tételt teljes indukcióval igazolhatjuk.

3. Tétel. Ha $\overline{A}$ az A esemény komplementere (ellentettje), akkor

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

Ui. az A és $\overline{A}$ egymást kizáró események, és $Q = A + \overline{A}$, tehát a II. és a III. axióma szerint

$$1 = P(Q) = P(A + \overline{A}) = P(A) + P(\overline{A}),$$

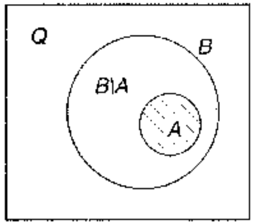
amiből a $P(\overline{A}) = 1 - P(A)$ következik.

4. Tétel. Ha $A \subset B$, akkor $P(A) \leq P(B)$.

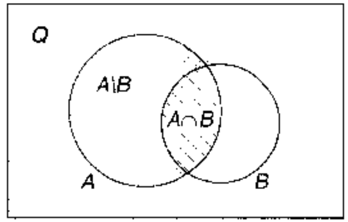
Ui. a Q eseménytér A és $B \setminus A$ (2. ábra) eseményei egymást kölcsönösen kizárják, továbbá $B = A + (B \setminus A)$, így a III. axióma szerint:

$$P(B) = P(A) + P(B \setminus A).$$

Mivel $P(B \setminus A) \geq 0$, ezért a $P(A) \leq P(B)$ egyenlőtlenség teljesül, ha $A \subset B$.



2. ábra. $A \subset B$



3. ábra. $A \setminus B$ és $A \cdot B$

5. Tétel. Ha A és B két tetszőleges esemény, akkor

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

Ui. a Q eseménytér $A \setminus B$ és $A \cdot B$ (3. ábra) eseményei egymást kölcsönösen kizárják, továbbá $A = (A \setminus B) + (A \cdot B)$, így a III. axióma szerint

$$P(A) = P(A \setminus B) + P(A \cdot B),$$

amelyből már állításunk következik.

6. Tétel. Ha A és B két tetszőleges esemény, akkor

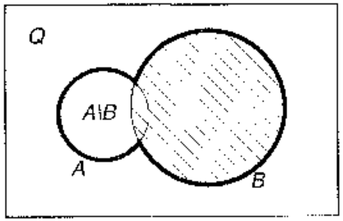
$$P(A + B) = P(A) + P(B) - P(A \cdot B), \qquad (*)$$

azaz két tetszőleges esemény összegének valószínűségét megkapjuk, ha együttes bekövetkezésük valószínűségét kivonjuk a két esemény valószínűségének összegéből.

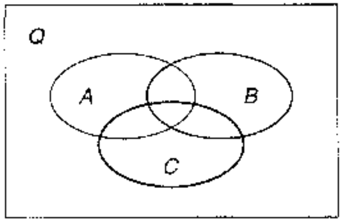
Ui. a Q eseménytér $A \setminus B$ és B (4. ábra) eseményei egymást kölcsönösen kizárják, továbbá $A + B = (A \setminus B) + B$, így a III. axióma és az 5. tétel felhasználásával:

$$P(A + B) = P(A \setminus B) + P(B) = P(A) - P(A \cdot B) + P(B),$$

amelyből állításunk már következik. A (*) formula annak valószínűségét adja, hogy a két tetszőleges esemény közül legalább az egyik bekövetkezik.



4. ábra. $A \setminus B$ és $A \cup B$



5. ábra. Események egyesítése

Következmény. Ha A, B és C tetszőleges események (5. ábra) akkor

$$P(A \cup B \cup C) =$$

$$= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

Ui. legyen $D = B \cup C$, akkor

$A \cap D = A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ és

$$P(A \cap D) = P(A \cap B) + P(A \cap C) - P(A \cap B \cap A \cap C) =$$

$$= P(A \cap B) + P(A \cap C) - P(A \cap B \cap C).$$

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

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

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

Példa

Igazoljuk, hogy $P(A + B) = 1 - P(\overline{A} \cdot \overline{B})$, ha A és B két tetszőleges esemény.

Megoldás

A *De Morgan* szabály szerint $\overline{A} \cdot \overline{B} = \overline{A + B}$, ebből következik, hogy $\overline{A} \cdot \overline{B}$ és $A + B$ ellentett események. Az ellentett események valószínűségeinek összege 1, azaz $P(\overline{A} \cdot \overline{B}) + P(A + B) = 1$, melyből a $P(A + B) = 1 - P(\overline{A} \cdot \overline{B})$ állítás következik.

1.4. Valószínűségi mezők

Hogyan számítjuk ki egy A összetett esemény valószínűségét?

Legyen $Q = \{A_1, A_2, \ldots, A_n\}$ egy teljes eseményrendszer. Tegyük fel, hogy ismerjük mindegyik A_i elemi esemény $P(A_i)$ valószínűségét, amelyet jelöljünk p_i-vel, azaz

$$P(A_i) = p_i.$$

Az A_1, $A_2, \ldots,$ A_n eseményeket a hozzájuk tartozó p_i ($i = 1, 2, \ldots, n$) valószínűségekkel együtt **valószínűségi mezőnek** nevezzük.

A valószínűségi mezőt az előbbiekhez hasonlóan definiáljuk akkor is, ha a teljes eseményrendszer nem véges, de megszámlálhatóan végtelen eseményekből áll. Ekkor *végtelen valószínűségi mezőről* beszélünk.

Egy A **összetett esemény** valószínűségét is könnyen ki tudjuk számítani, ha előállítható a Q-ból vett események összegeként, ugyanis ekkor az A valószínűségét az őt előállító események valószínűségeinek összege adja. Ha pl. $A = A_1 + A_2 + A_3$, akkor

$$P(A) = P(A_1) + P(A_2) + (A_3) = p_1 + p_2 + p_3,$$

mivel A_1, A_2, A_3 a Q teljes rendszer elemei, tehát páronként kizárják egymást, és így alkalmazható a IV. axióma.

Példák

1. Három pénzérme ismételt feldobásakor figyeljük meg a *fej* dobások számát. Mi lesz a Q eseménytér és a valószínűségi mező?

Megoldás

Az eseménytér: $Q = \{0, 1, 2, 3\}$, és a lehetséges esetek száma 8 (F a fej, I az írás). Nulla fej dobásának eseménye (azaz 3 írás dobás eseménye) 1-szer következhet be, egy fej dobásának eseménye 3-szor következhet be, két fej dobásának eseménye 3-szor következhet be stb.:

$$FFF \quad FFI \quad FIF \quad FII \quad IFF \quad IFI \quad IIF \quad III$$

A megfelelő valószínűségi mező. Q és

$$p_0 = P(0) = \frac{1}{8}, \quad p_1 = P(1) = \frac{3}{8}, \quad p_2 = P(2) = \frac{3}{8}, \quad p_3 = P(3) = \frac{1}{8}.$$

(Vegyük észre, hogy a valószínűségek nemnegatív valós számok és összegük: 1, azaz $\frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8}=1$.)

A legalább egy fej dobása legyen A esemény, és a mindhárom érmével fejet dobunk vagy egyetlen fejet sem dobunk esemény legyen B, azaz

$$A=\{1,2,3\} \text{ és } B=\{0,3\}.$$

Az A esemény valószínűsége a definíció szerint:

$$P(A)=P(1)+P(2)+P(3)=\frac{3}{8}+\frac{3}{8}+\frac{1}{8}=\frac{7}{8},$$

$$P(B)=P(0)+P(3)=\frac{1}{8}+\frac{1}{8}=\frac{2}{8}=\frac{1}{4}.$$

2. Robin, Balázs és Miklós futóversenyen vesznek részt. Robin nyerési esélye kétszerese Balázsénak, és Balázs nyerési esélye kétszerese Miklósénak. Mi annak a valószínűsége, hogy Balázs vagy Miklós nyeri a versenyt?

Megoldás

Miklós nyerési valószínűségét jelölje $P(M)=p$, akkor Balázs nyerési valószínűsége $P(B)=2p$, és Robin nyerési valószínűsége:

$$P(R)=2P(B)=2\cdot 2p=4p.$$

Az összes elemi esemény teljes eseményrendszert alkot, a valószínűségek öszszege 1, ezért

$$p+2p+4p=1,$$

amiből

$$p=\frac{1}{7},$$

s így

$$P(R)=4p=\frac{4}{7};\quad P(B)=2p=\frac{2}{7};\quad P(M)=p=\frac{1}{7}.$$

Balázs vagy Miklós nyerési valószínűsége pedig, mivel egymást kizáró események:

$$P(\{B,M\})=P(B)+P(M)=\frac{2}{7}+\frac{1}{7}=\frac{3}{7}.$$

1.4.1. Klasszikus valószínűségi mező

A gyakorlati feladatok megoldása szempontjából fontosak azok a valószínűségi mezők, amelyeknél az egyes események valószínűségei egyenlők. Ha egy A esemény előállítható egyenlő valószínűségű események összegeként, akkor az A valószínűsége egyszerűen kiszámítható.

Definíció. Ha egy véges sok elemi eseményből álló Q eseménytér minden eseményéhez egyenlő valószínűség tartozik, azaz az események egyenlően valószínűek, akkor **egyenlő valószínűségi mezőről**, más szóval **klasszikus valószínűségi mezőről** beszélünk.

Ha Q n elemi eseményből áll, és ezek mindegyike $\frac{1}{n}$ valószínűségű, akkor egy k számú elemi eseményt tartalmazó A esemény valószínűsége $k\cdot\frac{1}{n}=\frac{k}{n}$, azaz

$$P(A)=\frac{A \text{ elemeinek száma}}{Q \text{ elemeinek száma}}.$$

Ez a képlet a $P(A)$ kiszámítására csak **egyenlő valószínűségi mező** esetén használható. Az A esemény kimeneteleit **kedvező eseteknek** szoktuk nevezni és a fenti képletet így írjuk fel:

$$P(A)=\frac{\text{a kedvező esetek száma}}{\text{az összes eset száma}}.$$

Megjegyzés

Ha a $Q=\{A_1,A_2,\ldots,A_n\}$ teljes rendszer eseményei egyenlően valószínűek, akkor nyilvánvaló, hogy $P(A_i)=\frac{1}{n}$, $(i=1,2,\ldots,n)$, hiszen

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

és

$$P(A_1)=P(A_2)=\ldots=P(A_n)=p,$$

tehát

$$np=1 \quad \text{azaz} \quad p=\frac{1}{n}.$$

1.4.2. Kombinatórikai összefoglaló

A klasszikus valószínűségi mező eseményeinek valószínűségei általában **kombinatorikus módszerekkel** számíthatók ki. A kombinatorika alapfogalmaival kapcsolatos számítási formulákat bizonyítások nélkül felsoroljuk, mivel a továbbiakban többször felhasználásra kerülnek. Részletesebb ismertetésük az irodalomjegyzék könyveiben található meg.

A) Permutáció

a) n különböző elem különböző sorrendjének, azaz ***n* elem permutációinak száma:**

$$P_n = n!,$$

ahol $n! = 1\cdot 2\cdot 3\cdot\ldots\cdot(n-1)\cdot n$ (olv. en faktoriális); ($0! = 1$; $1! = 1$).

Példa

Hány hatjegyű szám képezhető a 0, 2, 3, 4, 5, 8 számjegyekből?

Megoldás

Ha a 0-val kezdődő számot nem tekintjük hatjegyűnek, akkor 6 elem $P_6 = 6! = 1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6 = 720$ permutációinak számából le kell vonni a 0-val kezdődő permutációk számát, melynek száma annyi, amennyi az utána álló 5 jegy permutációinak száma: $P_5 = 5! = 120$. Tehát a 0, 2, 3, 4, 5, 8 számjegyekből képezhető hatjegyű számok száma:

$$720 - 120 = 600.$$

b) Ha az adott n elem között $k_1, k_2, \ldots, k_r$ db megegyező van, akkor az összes lehetséges elrendezést ***n* elem ismétléses permutációinak száma** adja:

$$P_n^{(k_1,k_2,\ldots,k_r)} = \frac{n!}{k_1!k_2!\ldots k_r!},$$

ahol $k_1 + k_2 + \ldots + k_r = n$.

Példa

Hány permutáció képezhető a MATEMATIKA szó betűiből?

Megoldás

Az elemek (betűk) száma 10, de ismétlődő elemek is vannak: az M kétszer, az A háromszor, a T kétszer fordul elő, tehát 10 elem ismétléses permutációinak számát kell meghatározni az ismétlődő elemek $k_1 = 2$, $k_2 = 3$, $k_3 = 2$ számának figyelembevételével:

$$P_{10}^{(2,3,2,1,1,1)} = \frac{10!}{2!3!2!1!1!1!} = \frac{3628800}{24} = 151200.$$

B) Variáció

a) Ha n különböző elem közül minden lehetséges módon kiválasztunk k elemet s ezek összes permutációit képezzük, akkor megkapjuk ***n* elem *k*-ad osztályú variációit**, melyeknek száma:

$$V_{n,k} = n(n-1)(n-2)\ldots(n-(k-1)) = \frac{n!}{(n-k)!} \quad (k \le n).$$

Példa

A COMPAQ szó betűiből hány olyan hárombetűs jelsorozat állítható elő, ahol az egyes hárombetűs jelsorozatban a betűk nem ismétlődhetnek?

Megoldás

Hat elem harmadosztályú variációinak számát kell kiszámítani:

$$V_{6,3} = 6\cdot 5\cdot 4 = \frac{6!}{(6-3)!} = \frac{6!}{3!} = \frac{3!4\cdot 5\cdot 6}{3!} = 4\cdot 5\cdot 6 = 120.$$

b) Ha megengedjük, hogy a kiválasztásnál ugyanaz az elem legfeljebb k-szor ismételten kiválasztásra kerüljön, akkor k-ad osztályú ismétléses variációról beszélünk. A ***k*-ad osztályú ismétléses variációk száma:**

$$V_{n,k}^i = n^k \quad (k > n \text{ is felléphet}).$$

Példa

A totóban hány tipposzlopot kell kitölteni a biztos 13, ill. 13+1 találathoz?

Megoldás

Mivel az 1,x,2 elemek bármelyikét elhelyezhetjük a tipposzlop megadott 13, ill. 14 helyére, ezért a különböző módon kitölthető tipposzlopok száma 3 elem 13-ad osztályú, ill. 14-ed osztályú ismétléses variációinak számával egyenlő:

$$V_{3,13}^i = 3^{13} = 1\,594\,323, \quad V_{3,14}^i = 3^{14} = 4782969.$$

C) Kombináció

a) Ha n különböző elem közül minden lehetséges módon kiválasztunk k elemet, de a kiválasztottak sorrendjére nem vagyunk tekintettel, akkor n elem k-ad osztályú kombinációit kpjuk. ***n* elem *k*-ad osztályú kombinációinak száma:**

$$C_{n,k} = \frac{n(n-1)(n-2)\ldots(n-(k-1))}{k!} = \binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (k \le n).$$

ahol $\binom{n}{k}$ a binomiális együtthatók szokásos jelölése. $\left[\binom{n}{0}=1\right]$, $[0!=1]$.

Példa

A lottószelvényen 1-től 90-ig vannak a számok felírva, amelyek közül öt szám „eltalálása" szükséges a legnagyobb összeg elnyeréséhez. Hány szelvényt kellene tervszerűen kitölteni, hogy egy biztos ötös találat legyen?

Megoldás

Mivel az öt különböző szám kijelölésének a sorrendje nem számít, így a kitöltendő szelvények száma 90 elem ötödosztályú kombinációinak számával egyenlő:

$$C_{90,5}=\binom{90}{5}=\frac{90\cdot 89\cdot 88\cdot 87\cdot 86}{1\cdot 2\cdot 3\cdot 4\cdot 5}=43\,949\,268\,.$$

b) Ha n elem k-ad osztályú kombinációiban az elemek ismétlődését megengedjük, akkor n elem k-ad osztályú ismétléses kombinációit kapjuk. **n elem k-ad osztályú ismétléses kombinációinak száma:**

$$C_{n,k}^{i}=\binom{n+k-1}{k}=\frac{(n+k-1)!}{k!(n-1)!}\quad (k>n \text{ is felléphet}).$$

Példa

Hányféle eredményt kaphatunk, ha három egyforma kockát egyszerre dobunk fel s a kockák sorrendje nem számít?

Megoldás

Mivel minden dobás az 1, 2, 3, 4, 5, 6 számok közül három számot határoz meg, melyek között egyenlők is lehetnek, ezért 6 elem harmadosztályú ismétléses kombinációinak számát kell kiszámítani:

$$C_{6,3}^{i}=\binom{6+3-1}{3}=\binom{8}{3}=\frac{8\cdot 7\cdot 6}{1\cdot 2\cdot 3}=56.$$

Példák

1. A 32 lapos magyar kártya csomagból húzzunk egy lapot. A piros lap húzása legyen az A esemény, és a számmal jelölt lap húzása legyen a B esemény. Mi annak a valószínűsége, hogy

a) piros lapot húzunk. $P(A)=?$

b) számmal jelölt lapot húzunk. $P(B)=?$

c) számmal jelölt piros lapot húzunk. $P(A\cdot B)=?$

Megoldás

Mivel az eseménytér egyenlő valószínűségű elemi eseményeket tartalmaz (hiszen mindegyik lap kihúzásának a valószínűsége $\frac{1}{32}$), ezért

$$P(A)=\frac{\text{piros lapok száma}}{\text{kártyacsomag lapszáma}}=\frac{8}{32}=\frac{1}{4};$$

$$P(B)=\frac{\text{számmal jelölt lapok száma}}{\text{kártyacsomag lapszáma}}=\frac{16}{32}=\frac{1}{2};$$

$$P(A\cdot B)=\frac{\text{számmal jelölt piros lapok száma}}{\text{kártyacsomag lapszáma}}=\frac{4}{32}=\frac{1}{8}.$$

2. Egy dobozban 20 csavar van, amelyek közül hatnak hibás a menete. Véletlenszerűen hármat kiválasztva, mi a valószínűsége annak, hogy mindhárom hibás (A esemény), ill. hogy a három közül egyik sem hibás (B esemény)?

Megoldás

A kiemelt 3 csavar egymás közötti sorrendje nem számít, és 20-ból bármelyik csavart ugyanakkora valószínűséggel választhatjuk. Így 20 elem harmadosztályú ismétlés nélküli kombinációja adja a Ω eseménytér elemeinek számát. Tehát 20 csavarból hármat $\binom{20}{3}=\frac{20\cdot 19\cdot 18}{1\cdot 2\cdot 3}=1140$ -féleképpen lehet kiválasztani.

Az A elemeinek száma: 6 hibásból 3-at $\binom{6}{3}=\frac{6\cdot 5\cdot 4}{1\cdot 2\cdot 3}=20$ -féleképpen választhatunk.

A B elemeinek száma: 14 hibátlanból 3-at $\binom{14}{3}=\frac{14\cdot 13\cdot 12}{1\cdot 2\cdot 3}=364$ -féleképpen választhatunk.

Tehát „mindhárom csavar hibás" kiválasztásának valószínűsége:

$$P(A)=\frac{20}{1140}=\frac{1}{57}\approx 0{,}018,$$

„három közül egyik sem hibás" kiválasztás valószínűsége:

$$P(B)=\frac{364}{1140}=\frac{91}{285}\approx 0{,}319.$$

Ha a „legalább egy hibás" C esemény valószínűségét akarjuk meghatározni, akkor a 3. tételt alkalmazhatjuk, mivel $C=\overline{B}$, tehát

$$P(C)=P(\overline{B})=1-P(B)=1-\frac{91}{285}=\frac{194}{285}\approx 0{,}681.$$

3. Egy osztályban 11 lány közül 3 kék szemű és 8 barna szemű. Ha véletlenszerűen kiválasztunk 6 lányt, mi a valószínűsége annak, hogy pontosan egy kék szemű lesz köztük?

Megoldás

Hat lány kiválasztásának összes lehetséges számát 11 elem hatodosztályú kombinációinak a száma adja, azaz

$$C_{11,6}=\binom{11}{6}=\frac{11!}{6!(11-6)!}=\frac{7\cdot 8\cdot 9\cdot 10\cdot 11}{1\cdot 2\cdot 3\cdot 4\cdot 5}=462.$$

A 3 kékszemű közül egyet 3-féleképpen választhatunk, a 8 barnaszemű közül 5-öt $\binom{8}{5}$-féleképpen választhatunk, tehát a kedvező esetek száma:

$$k=C_{3,1}\cdot C_{8,5}=3\cdot\binom{8}{5}=3\cdot 56=168.$$

A kérdéses eseményt A-val jelölve:

$$P(A)=\frac{k}{n}=\frac{168}{462}=0{,}36.$$

Tehát 0,36 annak valószínűsége, hogy a 6 lány között lesz pontosan egy kék szemű.

4. Mi a valószínűsége annak, hogy

a) egy szabályos kocka s-szeri feldobása közül legalább egyszer 6-ost dobunk?

b) két szabályos kocka s-szeri feldobásával legalább egyszer dupla 6-ost dobunk?

Megoldás

a) A 6-os dobás eseményét jelöljük A-val. Egy kocka feldobásakor 6 elemi eseményt különböztetünk meg, ha pedig s-szer feldobjuk, akkor az egyenlően valószínű lehetséges kimenetelek száma 6^s. Mivel a lehetséges esetek közül a nem 6-os dobás eseménye 5^s esetben fordul elő, így a kedvező esetek száma: 6^s-5^s. Annak valószínűsége tehát, hogy s dobás közül legalább az egyik 6-os:

$$P(A)=\frac{6^s-5^s}{6^s}=1-\left(\frac{5}{6}\right)^s. \qquad (*)$$

b) A dupla 6-os dobás eseményét jelöljük B-vel. Két szabályos kocka feldobásakor az egyenlően valószínű kimenetelek száma 36, és így az s-szeri feldobás lehetséges kimeneteleinek száma 36^s. 35^s esetben dobhatunk dupla 6-ostól eltérőt, így a kedvező esetek száma: 36^s-35^s. Annak valószínűsége tehát, hogy két kocka s-szeri feldobása közül legalább egyszer dupla 6-ost dobunk:

$$P(B)=\frac{36^s-35^s}{36^s}=1-\left(\frac{35}{36}\right)^s. \qquad (**)$$

A *de Méré* által felvetett kérdésre (l. 1.1. Bevezetés 1. probléma) a (*) formulából $s=4$ helyettesítéssel:

$$P(A)=1-\left(\frac{5}{6}\right)^4=0{,}5177>\frac{1}{2}$$

adódik, a (**) formulából $s=24$ helyettesítéssel:

$$P(B)=1-\left(\frac{35}{36}\right)^{24}=0{,}4914<\frac{1}{2}$$

adódik. A felvetett problémához tartozóan $\frac{1}{2}$-nél nagyobb valószínűség eléréséhez szükséges dobásszámot nevezzük kritikus értéknek, ami egy kocka dobásánál 4. *De Méré* annak okát kereste, hogy miért nem teljesül az ún. „kritikus érték arányossági szabálya":

$$4:6=24:36,$$

vagyis ha hatodára csökken a valószínűség, akkor miért nem hatszorosára nő a kritikus érték. Ui. $s=25$ érték behelyettesítésével már $P(B)=1-\left(\frac{35}{36}\right)^{25}=0{,}50553>\frac{1}{2}$ adódik. A kérdésre pontos választ *de Moivre* 1718-ban megjelent *Doctrine of Chances* c. művében adott. A $0<p<1$ valószínűséghez az s kritikus szám az a legkisebb egész szám lesz, amely nagyobb az $(1-p)^x=\frac{1}{2}$ egyenlet x megoldásánál, azaz $x=-\frac{\ln 2}{\ln(1-p)}=\frac{\ln 2}{p+\frac{p^2}{2}+\ldots}$ értéknél.

Ebből látható, hogy elegendő kicsi p esetén $\frac{p^2}{2}$ elhanyagolható kicsi érték, s ekkor a kritikus értékek arányossági szabálya megfelelő közelítést ad.

Megjegyzés

De Méré második kérdésére (l. 1.1. Bevezetés 2. probléma) abban az időben számos tudósnak sem sikerült helyes választ adni. Az első játékos győzelmének valószínűségére, *Pascal* és *Fermat* általános megoldást adott. Ha A-val jelöljük azt az eseményt, hogy az első játékos győz, akkor

$$P(A)=\frac{1}{2^{n+m-1}}\sum_{i=n}^{n+m-1}\binom{n+m-1}{i}. \qquad (***)$$

Példa

Péter és Pál abban állapodnak meg, hogy az nyeri a felajánlott T pénzösszeget, aki először nyer 10 játszmát. Tegyük fel, hogy a játék befejezésig Péternek még kettő, Pálnak pedig még négy győzelemre van szüksége. Milyen arányban osztozzanak a megnyerhető összegen, ha nem folytathatják a játékot? Ki jár jobban, ha elfogadják Pál ajánlatát, mely szerint a megnyert játszmák arányában osszák fel a T összeget?

Megoldás

Pál ajánlata szerint a nyert játszmák arányában, azaz 8:6 arányban kellene elosztani a T összeget. Ugyanakkor Péter nyerési esélye a (***) formula szerint:

$$P(\text{Péter})=\frac{1}{2^{2+4-1}}\sum_{i=2}^{2+4-1}\binom{2+4-1}{i}=\frac{1}{2^5}\sum_{i=2}^{5}\binom{5}{i}=$$

$$=\frac{1}{32}\left[\binom{5}{2}+\binom{5}{3}+\binom{5}{4}+\binom{5}{5}\right]=\frac{1}{32}(10+10+5+1)=\frac{26}{32}=\frac{13}{16}.$$

Vagyis, ha a játékot folytatnák, akkor Péter $\frac{13}{16}$ valószínűséggel nyerhet, míg Pál nyerési esélye csak $\frac{3}{16}$, tehát a T összeget igazságosan 13:3 arányban kell felosztani Péter és Pál között. Ez a felosztási arány lényegesen eltér Pál által javasolt aránytól, mely Pálnak kedvezett. Ui. Péter a T összegnek csak $\frac{64}{112}$ részét kapná, az igazságosan neki járó $\frac{91}{112}$ rész helyett.

5. A legyártott termékek közül N db a raktárba került, amelyek között k db hibás van. Találomra kiválasztunk n db terméket $(n \le N)$. Ez azt jelenti, hogy bármelyik n db termék kiválasztásának az esélye azonos. Az ilyen kiválasztást **visszatevés nélküli mintavételnek** nevezzük. Mi a valószínűsége annak, hogy a minta pontosan h db hibását tartalmaz?

Megoldás

N termék közül n db kiválasztásának lehetséges számát, N elem n-ed osztályú ismétlés nélküli kombinációinak számával adhatjuk meg, azaz $\binom{N}{n}$ számú mintát választhatunk. Az egyforma valószínűségű elemi események száma tehát $\binom{N}{n}$. Az

$n-h$ hibátlan elemet $\binom{N-k}{n-h}$-féleképpen, h hibását pedig $\binom{k}{h}$-féleképpen választhatunk ki. A kedvező esetek száma tehát: $\binom{N-k}{n-h}\binom{k}{h}$ és így annak valószínűsége, hogy a találomra kivett n db termék közül pontosan h db hibás, vagyis az $A=\{n-h \text{ hibátlan}, h \text{ hibás}\}$ esemény valószínűsége:

$$P(A)=\frac{\binom{N-k}{n-h}\binom{k}{h}}{\binom{N}{n}}, \quad (k=0,1,2,\dots,n).$$

6. A legyártott termékek közül N db a raktárba került, amelyek közül k db hibás. Most egyenként válasszunk ki n terméket $(n \le N)$, úgy, hogy minden termék kiválasztásánál feljegyezzük annak milyenségét, és visszatesszük a többi közé. Az ilyen kiválasztást **visszatevéses mintavételnek** nevezzük. Mi a valószínűsége annak, hogy a kiválasztott mintában pontosan h db hibás termék van, ha feltesszük, hogy bármely termék kiválasztásának az esélye azonos?

Megoldás

A lehetséges esetek számát, mivel bármelyik termék kerülhet az n elemű minta 1., 2., stb. helyére, az ismétléses variációk száma adja: N^n. Az egyforma valószínűségű elemi események száma tehát N^n. A kiválasztottak között akkor van h hibás, ha pontosan $n-h$ a hibátlan termékek száma. A h hibás terméket a visszatevés miatt k^h-szor választhatjuk ki. Az $n-h$ hibátlant pedig $(N-k)^{n-h}$-féleképpen lehet kiválasztani. Így az $A=\{h \text{ hibás}, n-h \text{ hibátlan}\}$ eseményhalmaz elemeinek a száma (a kedvező kimenetelek):

$$\binom{n}{h}(N-k)^{n-h}k^h,$$

és így a keresett valószínűség:

$$P(A)=\frac{\binom{n}{h}(N-k)^{n-h}k^h}{N^n}=\binom{n}{h}\left(1-\frac{k}{N}\right)^{n-h}\left(\frac{k}{N}\right)^h, \quad (k=0,1,2,\dots,n).$$

A formulából látható, hogy a hibás termékek darabszáma helyett elegendő a $p=\frac{k}{N}$ selejtarányt ismerni, ekkor a formulát

$$P(A)=\binom{n}{h}(1-p)^{n-h}p^h \qquad (****)$$

alakban használjuk.

7. Egy dobozban egyenlő számú fehér és piros golyó van. Visszatevéssel 5 elemű mintát választva, mi a valószínűsége annak, hogy a mintában 2 piros golyó lesz (A esemény)?

Megoldás

Mivel a dobozban ugyanannyi fehér és piros golyó van, így bármelyik színű golyó kiválasztásának valószínűsége $p=\frac{1}{2}$. A (****) formulát használva.

$$P(A)=\binom{5}{2}\left(1-\frac{1}{2}\right)^{5-2}\left(\frac{1}{2}\right)^2=10\cdot\frac{1}{8}\cdot\frac{1}{4}=\frac{5}{16}=0{,}31\,,$$

vagyis 0,31 valószínűséggel kerül az 5 elemű mintába 2 db piros golyó.

Megjegyzés

Tegyük fel, hogy a Q eseménytér az A_1, A_2, ..., A_n, ... *megszámlálhatóan végtelen* elemi események halmaza, azaz

$$Q=\{A_1,A_2,\ldots,A_n,\ldots\}.$$

A véges eseménytérhez hasonlóan valószínűségi mezőt kapunk, ha a Q minden A_i eleméhez hozzárendeljük annak p_i valószínűségét. Ekkor a IV. axiómának megfelelően

$$P(A_1+A_2+\ldots+A_n+\ldots)=P(A_1)+P(A_2)+\ldots+P(A_n)+\ldots=$$

$$=p_1+p_2+\ldots+p_n+\ldots=\sum_{i=1}^{\infty}p_i=1$$

Az A esemény $P(A)$ valószínűségét ebben az esetben is az A elemi eseményeinek valószínűségeiből képzett összeg adja. A $P(A_i)=p_i$ számok azt mutatják, hogy miként oszlik meg az egységnyi valószínűség a Q teljes eseményrendszer eseményei között. Ennek értelmében mondhatjuk, hogy minden $p_i\geq 0$ $(i=1,2,\ldots)$, és $\sum_{i=1}^{\infty}p_i=1$ feltételeket kielégítő p_i számok *valószínűségeloszlást* alkotnak.

1.5. Geometriai valószínűség

A Q eseménytér egy geometriai alakzat nem megszámlálható ponthalmazához rendelt elemi események halmaza is lehet. Ilyenkor ha feltesszük, hogy a Q eseménytérben annak valószínűsége, hogy egy véletlen pont az $A\subset Q$ résztartományba essen, arányos az A tartomány mértékével (vonalszakasz hossza, térfogatrész köbtartalma, stb.), akkor *geometriai valószínűségről* beszélünk. A valószínűséget az A és Q mértékének hányadosával adjuk meg, azaz

$$P(A)=\frac{A\text{ részszakasz hossza}}{Q\text{ szakasz hossza}};$$

vagy

$$P(A)=\frac{A\text{ rész területe}}{Q\text{ területe}};$$

vagy

$$P(A)=\frac{A\text{ rész térfogata}}{Q\text{ térfogata}}.$$

A valószínűséget a Q halmazon **egyenletes eloszlásúnak** nevezzük, ha tetszőleges A esemény valószínűsége arányos a halmaz $|A|$ mértékével. Általában az egyenletes eloszlású valószínűségeket **geometriai valószínűségeknek** mondjuk.

Példák

1. Mi annak a p valószínűsége, hogy egy r sugarú körlap véletlenül kiválasztott pontja a vele koncentrikus $\frac{r}{2}$ sugarú körlapba esik?

Megoldás

Jelöljük Q-val az r sugarú, A-val pedig az $\frac{r}{2}$ sugarú körlap ponthalmazát, akkor

$$p=P(A)=\frac{A\text{ területe}}{Q\text{ területe}}=\frac{\pi\left(\frac{r}{2}\right)^2}{\pi r^2}=\frac{1}{4}.$$

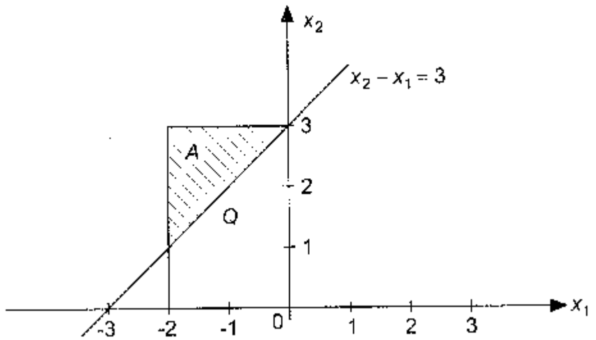
2. Válasszuk ki a valós számegyenesen az a és b pontokat úgy, hogy a

$$-2\leq a\leq 0,\quad 0\leq b\leq 3$$

egyenlőtlenségek teljesüljenek. Mi annak a valószínűsége, hogy a $d=b-a$ egyenes szakasz három egységnél hosszabb lesz?

Megoldás

A Q eseményteret az (a,b) rendezett számpárral adott ponthalmaz alkotja, amelyet derékszögű koordinátarendszerben a 6. ábra téglalapja szemlélteti.



6. ábra. P(A) valószínűsége

Az A halmaz pontjait a téglalaptartományon belül a $d = b - a > 3$ feltételt kielégítő (a,b) pontok alkotják, azaz az $x_2 - x_1 = 3$ egyenes feletti bevonalkázott síkidom. Tehát annak valószínűségét, hogy a $b - a > 3$ lesz, az A és a Q területek mérőszámának aránya adja:

$$p = \frac{A \text{ területe}}{Q \text{ területe}} = \frac{2}{6} = \frac{1}{3}.$$

3. Egy $2r$ alapélű, $3r$ magasságú egyenes állású négyzetes hasáb belsejébe egy r sugarú gömböt képzelünk. Mi a valószínűsége annak az A eseménynek, hogy a hasáb oldallapjai között bolyongó gázmolekula éppen a gömb belsejében van?

Megoldás

A hasáb térfogata: $V_h = 2r \cdot 2r \cdot 3r = 12r^3$. A gömb térfogata: $V_g = \frac{4r^3\pi}{3}$. A geometriai valószínűség definíciója szerint

$$P(A) = \frac{V_g}{V_h} = \frac{\frac{4r^3\pi}{3}}{12r^3} = \frac{\pi}{9} \approx 0{,}35$$

annak a valószínűsége, hogy a gázmolekula az r sugarú gömb belsejében van.

E.1. Ellenőrző kérdések az 1. fejezethez

1. Mit nevezünk kísérletnek, elemi eseménynek, eseménytérnek?

2. Hogyan definiáljuk a valószínűségi függvényt?

3. Melyek a *Kolmogorov*-féle axiómák?

4. Hogyan igazolhatók az axiómák alapján a

$$P(A \setminus B) = P(A) - P(A \cdot B) \text{ és}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cdot B) \text{ állítások.}$$

5. Hogyan definiáljuk a véges valószínűségi mezőt?

6. Hogyan számítjuk ki az egyenlő valószínűségi mező valamely eseményének valószínűségét?

7. Mit nevezünk gyakoriságnak és relatív gyakoriságnak?

8. Mit nevezünk geometriai valószínűségnek?

V.1. Feladatok az 1. fejezethez

V.1.1. Igazoljuk, hogy az AB, $\overline{A}$ és $A - B$ események teljes rendszert alkotnak.

V.1.2. Legyen A, B és C egy Q eseménytér 3 eseménye. Igazoljuk, hogy

a) B és $\overline{B}A$ egymást kizáró események;

b) $A + \overline{A}B = A + B$;

c) $(A - C)(B - C) = AB - C$.

V.1.3. Egy várba egy zöld és egy vörös kapun lehet bejutni. A „zöld kapu nyitva" eseményt jelölje Z, a „vörös kapu nyitva" eseményt jelölje V. Mit jelentenek a következő események:

a) ZV; *b)* $Z + V$; *c)* $\overline{Z}$; *d)* $V - Z$; *e)* $\overline{Z + V}$; *f)* $\overline{Z} + V$,

g) $Z\overline{V}$; *h)* $\overline{Z}\,\overline{V}$; *i)* $\overline{ZV}$; *j)* $\overline{Z} + \overline{V}$; *k)* $Z\overline{V} + \overline{Z}V$.

V.1.4. Legyen a Q eseménytér két tetszőleges eseménye A és B. Igazoljuk, hogy

a) $P(B) = P(AB) + P(\overline{A}B)$;

b) $P(A + B) = 1 - P(\overline{A}\,\overline{B})$.

V.1.5. Dobjunk fel egy pénzérmét hatszor egymásután. Mekkora a következő események valószínűsége:

a) mindegyik fej,

b) mindegyik írás,

c) vagy mindegyik fej, vagy mindegyik írás,

d) pontosan egy fej, a többi írás,

e) pontosan 4 fej,

f) legalább egy fej.

V.1.6. Egy dobozban 20 db hibás, de még egyes esetben használható, és 30 db hibátlan csavar van. Ha 2 db-ot találomra kiveszünk a dobozból, mekkora annak a valószínűsége, hogy

a) mindkettő hibás,

b) egyik sem hibás,

c) csak az egyik hibás,

d) legalább az egyik hibás,

e) legfeljebb az egyik hibás.

A válaszokat visszatevés nélküli és visszatevéses mintavétel esetére is adjuk meg.

V.1.7. Egy 6 hektárnyi terület fölött szálló repülőgépről véletlenszerűen leszakadt egy lemezdarab. Mekkora a valószínűsége annak, hogy egy $100\times 80\ \mathrm{m}^2$ -es sportpályára esik?

V.1.8. Egy kocka és egy pénzérme együttes feldobásával (írás (I), fej (F), 1, 2, 3, 4, 5, 6) adódó elemi eseményekből álló Q eseménytér legyen

$Q=\{F1, F2, F3, F4, F5, F6, I1, I2, I3, I4, I5, I6\}$

a) Adjuk meg Q részhalmazaként a következő eseményeket:

$A=\{\text{fej és páros szám}\}$;

$B=\{\text{fej, valamint írás és prímszámok}\}$;

$C=\{\text{írás és páratlan számok}\}$

b) Adjuk meg az

A vagy B;

B és C;

csak a B eseményeket.

c) Az A, B és C események közül melyek a páronként egymást kizáró események?

V.1.9. Dobjunk fel egy ötforintost, egy tízforintost és egy kockát.

a) Írjuk fel a Q eseményteret az elemi események felsorolásával (pl. $FF1$, stb).

b) Adjuk meg az A, a B és a C eseményt az elemi események felsorolásával, ha

$A=\{\text{két fej és páros szám}\}$,

$B=\{\text{a 2-es szám megjelenése}\}$,

$C=\{\text{pontosan egy fej és prímszám}\}$

c) Adjuk meg az

A és B;

csak a B;

B vagy C eseményeket az elemi események felsorolásával.

V.1.10. Legyen adott a $Q=\{A_1, A_2, A_3, A_4\}$ eseménytér. Az alább adott függvények közül melyek definiálnak a Q-n valószínűségi mezőt?

a) $P(A_1)=\frac{1}{2}$, $P(A_2)=\frac{1}{3}$, $P(A_3)=\frac{1}{4}$, $P(A_4)=\frac{1}{5}$;

b) $P(A_1)=\frac{1}{2}$, $P(A_2)=\frac{1}{4}$, $P(A_3)=-\frac{1}{4}$, $P(A_4)=\frac{1}{2}$;

c) $P(A_1)=\frac{1}{2}$, $P(A_2)=\frac{1}{4}$, $P(A_3)=\frac{1}{8}$, $P(A_4)=\frac{1}{8}$;

d) $P(A_1)=\frac{1}{2}$, $P(A_2)=\frac{1}{4}$, $P(A_3)=\frac{1}{4}$, $P(A_4)=0$.

V.1.11. Legyen P a $Q=\{A_1, A_2, A_3\}$ eseménytéren értelmezett valószínűségi függvény. Határozzuk meg $P(A_1)$ értékét, ha

a) $P(A_2)=\frac{1}{3}$ és $P(A_3)=\frac{1}{4}$;

b) $P(A_1)=2P(A_2)$ és $P(A_3)=\frac{1}{4}$;

c) $P(\{A_2, A_3\})=2P(A_1)$;

d) $P(A_3)=2P(A_2)$ és $P(A_2)=3P(A_1)$.

V.1.12. Nándi (N) és Balázs (B), valamint Merci (M), Anita (A) és Jetta (J) sakkversenyen vesz részt. A nyerés valószínűsége nemenként azonos, de minden fiú kétszer esélyesebb, mint bármelyik lány.

a) Határozzuk meg annak valószínűségét, hogy a sakkversenyt lány nyeri;

b) Határozzuk meg annak valószínűségét, hogy a sakkversenyt N és M (Nándi, Merci) testvérpár valamelyike nyeri.

V.1.13. Egy olyan súlyeloszlású kockánk van, amellyel bármely pontszám dobásának valószínűsége arányos a szóban forgó pontszámmal. Az A, B, C eseményeket rendre a páros, a prím és a páratlan számok alkotják.

a) Adjuk meg a valószínűségi mezőt;

b) Határozzuk meg a $P(A)$, $P(B)$, $P(C)$ valószínűségeket;

c) Határozzuk meg a

c_1) páros vagy prímszámok;

c_2) páratlan prímszámok;

c_3) B-hez nem tartozó páros számok dobásának valószínűségét.

V.1.14. Egy dobozban 12 tranzisztor van, amelyek közül 4 hibás. Véletlenül kiemelve 2 db tranzisztort, mi a valószínűsége annak, hogy mindkettő hibás; egyik sem hibás ill. legalább az egyik hibás?

V.1.15. Egy rekeszbe rendezetlenül lett behelyezve 20 nő és 10 férfi személyi nyilvántartási lapja. A férfiaknak is és a nőknek is a fele mérnök. Mi annak a valószínűsége, hogy egy véletlenül kiemelt lapon férfi (A esemény), mérnök (B esemény), férfi vagy mérnök adatai vannak?

V.1.16. Egy kör kerületén véletlenül kiválasztott három pontot jelölje a, b és c. Mi annak a valószínűsége, hogy a három pont a kör ugyanazon fél kerületén fekszik?

V.1.17. $P(A+B)=\frac{3}{4}$, $P(\overline{A})=\frac{2}{3}$ és $P(A\cdot B)=\frac{1}{4}$ ismeretében, határozzuk meg a $P(A)$, $P(B)$ és $P(A\cdot\overline{B})$ valószínűségeket.

V.1.18. Határozzuk meg a $P(A\cdot B)$, $P(\overline{A}\cdot\overline{B})$, $P(\overline{A}+\overline{B})$ és $P(B\cdot\overline{A})$ valószínűségeket, ha $P(A)=\frac{1}{2}$, $P(A+B)=\frac{3}{4}$ és $P(\overline{B})=\frac{5}{8}$.

V.1.19. Egy szabályos kockát 100-szor feldobtunk, és az egyes pontszámok előfordulását az alábbi ún. gyakorisági táblázatba foglaltuk:

Pontok száma:	1	2	3	4	5	6
Gyakorisága:	14	17	20	18	15	16

Számítsuk ki

a) a 6-os dobás;

b) a 3-as dobás;

c) a páratlan szám dobás;

d) a prímszám dobás relatív gyakoriságát.

MÁSODIK FEJEZET

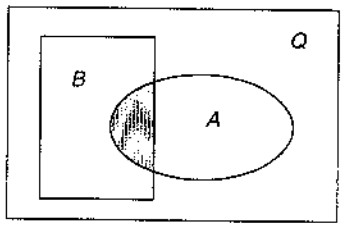
2.1. Feltételes valószínűség

Kísérleteink során sokszor előfordul, hogy valamely esemény valószínűségét egy másik esemény bekövetkezésének figyelembevételével kell meghatározni. Vizsgáljuk meg, hogy mi az A esemény valószínűsége abban az esetben, ha csak azokat a kimeneteleket vesszük figyelembe, amelyek a B eseményhez is hozzátartoznak. Ezt úgy is értelmezhetjük, hogy Q helyett B-t tekintjük az új eseménytérnek, vagy az A valószínűségét a B esemény bekövetkezésének feltétele mellett keressük.

Definíció: Legyenek A, B tetszőleges eseményei egy kísérlet Q eseményterének és B valószínűsége legyen pozitív, azaz $P(B)>0$. Az A esemény B feltételre vonatkozó, $P(A|B)$-vel jelölt (olv. pé A feltéve B) **feltételes valószínűségén** az A és B együttes bekövetkezésének és a B esemény valószínűségének hányadosát, azaz

$$P(A|B)=\frac{P(A\cdot B)}{P(B)} \qquad (1)$$

számot értjük (7. ábra). Ez a szám azt mutatja, hogy A hányadrészben következik be a B bekövetkezéseinek esetei közül, ill. ha B bekövetkezett, akkor annak valószínűsége, hogy A is bekövetkezik, $P(A|B)$.



7. ábra. Venn-*diagram:* $P(A\cdot B)/P(B)$ *szemléltetése*

A képlet alapján a feltételes valószínűséget a feltétel nélküli valószínűségekből kiszámíthatjuk, ha a feltétel valószínűsége nem nulla.

Megjegyzés

Ha A és B a Q-hoz tartozó egyenlő valószínűségi mező eseményei, akkor az A esemény B-re vonatkozó **feltételes relatív gyakoriságát** az

$$r_{A|B} = \frac{m_{A \cdot B}}{m_B},$$

képlettel számíthatjuk, ahol $m_{A \cdot B}$ az $A \cdot B$ esemény bekövetkezéseinek, m_B pedig a B esemény bekövetkezéseinek száma.

Példák

1. Dobjunk fel két szabályos kockát. A B esemény legyen az, hogy a két kockával 6-ot dobunk, azaz amikor a két kocka felső lapján a pontok összege 6:

$$B = \{(1,5), (2,4), (3,3), (4,2), (5,1)\};$$

az A esemény pedig az legyen, hogy legalább az egyik kockán kettest dobunk. Számítsuk ki a $P(A|B)$ feltételes valószínűséget.

Megoldás

A Q eseménytér elemeinek a száma: $6 \cdot 6 = 36$, az A elemeinek a száma pedig:

$$A = \{(2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (1,2), (3,2), (4,2), (5,2), (6,2)\},$$

azaz 11, és $A \cdot B = \{(2,4), (4,2)\}$, ezért

$$P(A|B) = \frac{P(A \cdot B)}{P(B)} = \frac{\frac{2}{36}}{\frac{5}{36}} = \frac{2}{5},$$

ugyanakkor $P(A) = \frac{11}{36}$.

2. Tegyük fel, hogy egy iskola 230 tanulójáról az alábbi kimutatás készült egy járványos betegség időszakában:

	beteg volt	nem volt beteg	összesen	esemény
fiú	50	60	110	B
lány	40	80	120	$\overline{B}$
összesen	90	140	230	
esemény	A	$\overline{A}$		

A nyilvántartó kartonok rendezetlenül vannak letéve. Mi a valószínűsége annak, hogy a véletlenszerűen kiemelt karton

a) fiúé;

b) betegségen átesett tanulóé;

c) betegségen átesett fiúé.

Ha előzetesen a fiúk és lányok kartonját rendezetlenül külön-külön halomba tették, akkor mi a valószínűsége annak, hogy

d) a fiúk kartonjából egyet kiválasztva az betegségen átesett fiú kartonja lesz;

e) lányok kartonja közül egyet kiválasztva az olyan lányé, aki nem volt beteg?

Megoldás

A táblázat jelölései szerint:

a) $P(B) = \frac{110}{230} \approx 0{,}48;$

b) $P(A) = \frac{90}{230} \approx 0{,}39;$

c) $P(A \cdot B) = \frac{50}{230} \approx 0{,}22;$

d) $P(A|B) = \frac{P(A \cdot B)}{P(B)} = \frac{\frac{50}{230}}{\frac{110}{230}} = \frac{50}{110} \approx 0{,}45;$

e) $P(\overline{A}|\overline{B}) = \frac{P(\overline{A} \cdot \overline{B})}{P(\overline{B})} = \frac{\frac{80}{230}}{\frac{120}{230}} = \frac{80}{120} \approx 0{,}67.$

3. Egy üzem A-val és B-vel jelölt két célgépén ugyanazt az alkatrészt gyártja. Az A gépen naponta 500 darabot, melyből 20 db selejt, a B gépen naponta 650 darabot, melyből 30 a selejt. Ha az egyik nap gyártott alkatrészek közül kiveszünk egyet és az selejt, akkor mi annak a valószínűsége, hogy ez a kivett selejtes alkatrész az A gépen készült?

Megoldás

A lehetséges esetek száma a két gépen naponta gyártott alkatrészek száma: 1150 db, a kedvező esetek száma az A gépen gyártott selejt darabszáma: 20 db. Ha E jelenti azt az eseményt, hogy az alkatrész az A gépen készült, S pedig a feltételt jelentő eseményt, azaz egy selejt kihúzását, akkor a $P(E \cdot S)$ annak a valószínűségét jelenti, hogy az alkatrész az A gépen készült és selejt, így

$$P(E|S) = \frac{P(E \cdot S)}{P(S)} = \frac{\frac{20}{1150}}{\frac{50}{1150}} = \frac{20}{50} = \frac{2}{5} = 0{,}4.$$

2.1.1. Szorzási tétel

A feltételes valószínűség fontos következménye az ún. *szorzási tétel (szorzási szabály)*, amelyet, az $A \cdot B = B \cdot A$ felhasználásával a 2.1. pont (1) képletéből

$$P(A \cdot B) = P(B)P(A|B) \tag{2}$$

alakban kapunk. Természetesen A-t is tekinthetjük feltételnek s akkor (2) helyett

$$P(B \cdot A) = P(A)P(B|A) \tag{2*}$$

formulát használjuk.

Amint látható a tétellel két tetszőleges esemény együttes bekövetkezésének valószínűségét megkapjuk, ha az egyik eseményt feltételnek tekintjük, és ennek valószínűségét szorozzuk a másik eseménynek a feltételre vonatkozó valószínűségével.

A szorzási tétel tetszőleges $A_1, A_2, ..., A_n$ eseményekre is kiterjeszthető a teljes indukció alkalmazásával:

$$P(A_1 \cdot A_2 \cdot \ldots \cdot A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cdot A_2)\ldots$$
$$\ldots P(A_n|A_1 \cdot A_2 \cdot \ldots \cdot A_{n-1}). \quad (2^{**})$$

A (2**) formulát általános szorzási szabálynak nevezzük.

Példák

1. Az előző pont 2. példájának adatait felhasználva határozzuk meg annak valószínűségét, hogy a 230 kartonból egymás után kihúzott három karton betegségen átesett fiú kartonja.

Megoldás

A jelölje azt az eseményt, hogy az első karton betegségen átesett fiúé, B azt, hogy a második karton betegségen átesett fiúé, C pedig azt, hogy a harmadik karton betegségen átesett fiúé. Ekkor az $A \cap B \cap C$ esemény valószínűségét kell meghatároznunk a (2**) képlet szerint:

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

A jobb oldali valószínűségek:

$$P(A) = \frac{90}{230}; \quad P(B|A) = \frac{89}{229},$$

(feltettük, hogy az A esemény megvalósult, így mind a betegségen átesettek száma, mind az összes létszám eggyel csökken);

$$P(C|B \cdot A) = \frac{88}{228},$$

így a keresett valószínűség:

$$P(A \cdot B \cdot C) = \frac{90}{230} \cdot \frac{89}{229} \cdot \frac{88}{228} \approx 0{,}06.$$

2. Egy dobozban 16 tranzisztor közül 4 hibás. Mi annak a valószínűsége, hogy

a) visszatevés nélkül három egymás után kivett tranzisztor hibátlan?

b) az első hibátlan, a második hibás, a harmadik ismét hibátlan?

Megoldás

a) 1. módszer: Mivel $16-4=12$ hibátlan van a dobozban, az első hibátlan tranzisztor kivételének valószínűsége $\frac{12}{16} = \frac{3}{4} = 0{,}75$. Ha hibátlant vettünk ki, akkor a dobozban a maradék 15-ből 11 hibátlan, tehát a másodszor is hibátlan tranzisztor kivételének valószínűsége $\frac{11}{15} = 0{,}73$, és a harmadszor is hibátlan tranzisztor kivételének valószínűsége $\frac{10}{14} = \frac{5}{7} = 0{,}71$. A szorzási tétel alkalmazásával

$$p = \frac{12}{16} \cdot \frac{11}{15} \cdot \frac{10}{14} = \frac{11}{28} = 0{,}39,$$

tehát 0,39 valószínűséggel vehetünk ki a dobozból egymás után három hibátlan tranzisztort.

2. módszer: 16 tranzisztorból hármat $\binom{16}{3} = 560$ -féleképpen lehet kiválasztani, 12-ből hármat pedig $\binom{12}{3} = 220$ -féleképpen, tehát

$$p = \frac{220}{560} = \frac{11}{28} = 0{,}39.$$

3. módszer: 16-ból egymás után három tranzisztort kiválasztani $16 \cdot 15 \cdot 14$ -féleképpen lehet, 12-ből hármat pedig $12 \cdot 11 \cdot 10$ -féleképpen, tehát

$$p = \frac{12 \cdot 11 \cdot 10}{16 \cdot 15 \cdot 14} = \frac{11}{28} = 0{,}39.$$

b) Ha T_1, T_2, T_3 jelöli az egymás után, visszatevés nélkül, véletlenszerűen kiválasztott hibátlan tranzisztor választásának eseményét, akkor a $P(T_1\overline{T_2}T_3)$ valószínűséget kell meghatároznunk. Mivel

$$P(T_1)=\frac{16-4}{16}=\frac{3}{4},$$

$$P(\overline{T}_2\,|\,T_1)=\frac{4}{16-1}=\frac{4}{15},$$

$$P(T_3|T_1\overline{T}_2)=\frac{16-4-1}{16-2}=\frac{11}{14},$$

és így az általános szorzási szabály szerint a keresett valószínűség:

$$P(T_1\overline{T}_2T_3)=P(T_1)\cdot P(\overline{T}_2\,|\,T_1)\cdot P(T_3|T_1\overline{T}_2)=\frac{3}{4}\cdot\frac{4}{15}\cdot\frac{11}{14}=0{,}16.$$

Megjegyzés

Adott, véges valószínűségekkel rendelkező véges számú kísérletsorozatokat (véges sztochasztikus folyamatokat) célszerű ún. **fadiagramon** szemléltetni. A fadiagramot úgy készítjük el, hogy egy kezdőpontból (a gyökérből) kiindulva utakat (ágakat) rajzolunk az első kísérletsorozat egymást kizáró eseményeihez, és az utakhoz a megfelelő valószínűségeket írjuk fel. E pontokból a következő kísérletsorozat eseményeihez szintén megrajzoljuk az utakat, a megfelelő valószínűségekkel ellátva, stb. A kezdőponttól a végponthoz vezető utak mindegyike a kísérletsorozat egy kimenetelének felel meg.

A feltételes és a teljes valószínűség tételeiből egyrészt következik, hogy a kísérletsorozat egy-egy útja által meghatározott eseményének valószínűségét az út menti valószínűségek szorzata adja, másrészt, ha az eseményhez több út tartozik, akkor a teljes valószínűséget az egyes utakhoz tartozó valószínűségek összegeként kapjuk.

Példa

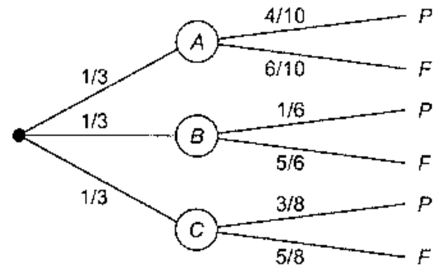
Adott három urna, amelyeket A-val, B-vel és C-vel jelölünk. Az A-ban 6 fehér és 4 piros, a B-ben 5 fehér és 1 piros, a C-ben 5 fehér és 3 piros golyó van. A véletlenszerűen kiválasztott urnából emeljünk ki ugyancsak véletlenszerűen egy golyót. Mi a p valószínűsége annak, hogy piros golyót emeltünk ki az urnából?

Megoldás

A sztochasztikus folyamat most két kísérletből áll: az urna kiválasztásából és a fehér (F) vagy piros (P) golyó kiválasztásából. A fadiagramot a 8. ábra szemlélteti.

Ha az A urnát választottuk, akkor a piros golyó kiemelésének valószínűsége: $\frac{1}{3}\cdot\frac{4}{10}=\frac{2}{15}$, ha a B urnát választottuk, akkor $\frac{1}{3}\cdot\frac{1}{6}=\frac{1}{18}$, ha a C urnát választottuk, akkor $\frac{1}{3}\cdot\frac{3}{8}=\frac{1}{8}$, s minthogy három kölcsönösen független út vezet a piros golyókhoz, ezért egy piros golyó kiemelésének p valószínűségét a három út valószínűségeinek összege adja:

$$p=\frac{1}{3}\cdot\frac{4}{10}+\frac{1}{3}\cdot\frac{1}{6}+\frac{1}{3}\cdot\frac{3}{8}=\frac{2}{15}+\frac{1}{18}+\frac{1}{8}=\frac{113}{360}=0{,}31.$$



8. ábra. Fadiagram

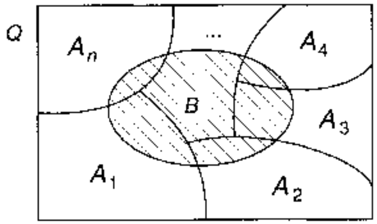
2.2. A teljes valószínűség tétele

Egy Q eseményteret többféle módon is felbonthatunk bizonyos számú esemény összegére. A lehetséges felbontások közül számunkra az egyik legfontosabb az, amelyik a Q eseményteret kölcsönösen kizáró események összegére bontja fel. Legyen $A_1, A_2, \ldots, A_n$ teljes eseményrendszer $\left(\sum_i P(A_i)=1\right)$ és B egy tetszőleges pozitív valószínűségű esemény. Ha $\bigcup_{i=1}^{n}A_i=Q$ (9. ábra), akkor

$$B=Q\cdot B=(A_1+A_2+\ldots+A_n)\cdot B=$$
$$=(A_1\cdot B)+(A_2\cdot B)+\ldots+(A_n\cdot B),$$

ahol az $A_i\cdot B$ $(i=1,2,\ldots,n)$ egymást kölcsönösen kizáró események, így

$$P(B)=P(A_1\cdot B)+P(A_2\cdot B)+\ldots+P(A_n\cdot B).$$



9. ábra. A Q eseménytér felosztása

A szorzási tétel alkalmazásával $P(B)$ a következő alakban is felírható:

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \ldots + P(A_n)P(B|A_n), \quad (3)$$

$$P(A_i) \neq 0 \quad (i = 1, 2, \ldots, n)$$

Ezt a formulát a **teljes valószínűség tételének** nevezzük.

A teljes valószínűség tétele tehát azt fejezi ki, hogy ha A_1, A_2, ..., A_n egymást páronként kizáró, pozitív valószínűségű események és $\sum_{i=1}^{n} P(A_i) = 1$, akkor tetszőleges B esemény valószínűségét a (3) képlete szerint lehet kiszámítani, amelyet

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i) \quad (3^*)$$

alakban szokás megadni.

Példák

1. Az évfolyam matematika szakos hallgatóinak 90%-a, fizika szakos hallgatóinak 70%-a sikeresen vizsgázott. A fizika szakosok az évfolyam 18%-át teszik ki. Mennyi a valószínűsége annak, hogy egy véletlenül kiválasztott hallgató a sikeresen vizsgázottak közül való?

Megoldás

Legyen B a kérdéses esemény. Az F esemény jelentse azt, hogy a kiválasztott hallgató fizikus, az M pedig azt, hogy matematikus. A fizikus hallgató kiválasztásának esélye: $P(F) = \frac{18}{100}$, a matematikus hallgatóé: $P(M) = \frac{82}{100}$. A B esemény

feltételes valószínűsége az F feltétel mellett: $P(B|F) = \frac{70}{100}$, és az M feltétel mellett: $P(B|M) = \frac{90}{100}$.

A teljes valószínűség tételét felhasználva:

$$P(B) = P(B|F)P(F) + P(B|M)P(M) = \frac{70}{100}\cdot\frac{18}{100} + \frac{90}{100}\cdot\frac{82}{100} = \frac{1260+7380}{10000} \approx 0{,}86.$$

Tehát 0,86 vagyis 86% annak a valószínűsége, hogy a véletlenül kiválasztott hallgató a sikeresen vizsgázottak közül való.

2. Egy éremgyűjtő dobozában 19 darab ötforintos van, melyek közül 12 db a 2000. évben, a többi pedig 1997. évben készült. Egy érdeklődő számára véletlenszerűen 3 érmét kiveszünk, és visszatevés előtt mindhármat megjelöljük. A következő érdeklődő számára ismét kiveszünk találomra 3 érmét. Mennyi a valószínűsége annak, hogy a másodszor kivett három érme nincs megjelölve, és 2000. évben készült?

Megoldás

A kérdéses eseményt jelölje A, B_j $(j = 0, 1, 2, 3)$ pedig azt az eseményt, hogy az először választott 3 érme közül j darab 2000-ben vert érme volt. Először meghatározzuk az A esemény feltételes valószínűségét B_j feltétel mellett.

3 érme összes lehetséges kiválasztásának számát 19 elem 3-ad osztályú kombinációi adják: $n = C_{18,3} = \binom{19}{3} = 969$. A 12 db 2000. évben vert érme közül j db kiválasztásának lehetséges száma: $C_{12,j} = \binom{12}{j}$ és a korábban vert 7 érme közül $3-j$ db kiválasztásának lehetséges száma: $C_{7,3-j} = \binom{7}{3-j}$ vagyis a kedvező esetek száma: $k_1 = C_{12,j} \cdot C_{7,3-j} = \binom{12}{j}\binom{7}{3-j}$. A B_j esemény valószínűsége:

$$P(B_j) = \frac{k_1}{n} = \frac{\binom{12}{j}\binom{7}{3-j}}{\binom{15}{3}} \quad (j = 0, 1, 2, 3).$$

Ha először j db 2000-ben vert érmét választottunk, akkor másodszor 3 db nem megjelölt 2000-ben vert érmét csak $12-j$ db nem megjelölt 2000-ben vert érme közül választhatunk. A lehetséges kedvező választások száma:

$$k_2 = C_{12-j,3} = \binom{12-j}{3}.$$

Tehát az A esemény feltételes valószínűsége B_j feltétel mellett:

$$P(A|B_j)=\frac{k_2}{n}=\frac{\binom{12-j}{3}}{\binom{19}{3}},\qquad (j=0,1,2,3).$$

Most alkalmazzuk a teljes valószínűség tételét:

$$P(A)=\sum_{j=0}^{3}P(A\mid B_j)P(B_j)=\sum_{j=0}^{3}\frac{\binom{12-j}{3}}{\binom{19}{3}}\cdot\frac{\binom{12}{j}\binom{7}{3-j}}{\binom{19}{3}}=$$

$$=\frac{1}{969^2}(220\cdot 1\cdot 35+165\cdot 12\cdot 21+120\cdot 66\cdot 7+84\cdot 220\cdot 1)==\frac{123200}{938961}\approx 0{,}13.$$

Tehát 0,13 annak a valószínűsége, hogy másodszorra 3 db nem megjelölt 2000-ben vert ötforintos érmét veszünk ki.

2.2.1. Bayes tétele

A teljes eseményrendszerrel kapcsolatban gyakran nem a B esemény valószínűségét kell meghatároznunk, hanem azt, hogy a B esemény bekövetkezésében az eseményrendszer egyes A_i eseményei milyen szerepet játszanak. A kérdés a következőképpen is megfogalmazható: ha B bekövetkezett, mi annak a valószínűsége, hogy ez pontosan az eseményrendszer A_i (i-edik) eseményének bekövetkezésével együtt valósult meg.

Az A_i események B eseményre vonatkozó feltételes valószínűségeit minden i-re a

$$P(A_i|B)=\frac{P(A_i\cdot B)}{P(B)}\tag{1}$$

képlettel számíthatjuk.

A (1) képletbe $P(B)$ helyett helyettesítsük a 2.2. pont (3) kifejezés jobb oldalát, a $P(A_i\cdot B)$ helyett pedig a neki megfelelő $P(A_i)P(B|A_i)$ kifejezést, akkor a következő összefüggést kapjuk:

$$P(A_i|B)=\frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1)+P(A_2)P(B|A_2)+\ldots+P(A_n)P(B|A_n)},\tag{2}$$

$$(i=1,2,\ldots,n).$$

Ezt a formulát nevezzük ***Bayes*-tételnek**.

Ha tehát az A_1, A_2, ..., A_n egymást páronként kizáró pozitív valószínűségű események, amelyekre $\sum_{i=1}^{n}P(A_i)=1$, akkor tetszőleges pozitív valószínűségű B eseményre és bármely i-re fennáll a (2) összefüggés, amelyet

$$P(A_i|B)=\frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n}P(A_j)P(B|A_j)};\ (i=1,2,\ldots,n)\tag{2*}$$

alakban szokás megadni. Az (2*) formulából látható, hogy a *Bayes*-tételt akkor alkalmazzuk, ha egy teljes eseményrendszer valamelyik eseményének feltételes valószínűségét kell meghatározni, ismerve a feltételi eseménynek az eseményrendszer elemeire vonatkozó feltételes valószínűségeit, valamint a teljes eseményrendszer minden elemének valószínűségeit.

Példák

1. A 2. fejezet 2.1. pontjának 2. példájában szereplő adatok felhasználásával határozzuk meg

a) az A esemény valószínűségét a teljes valószínűség tételével is, és

b) annak valószínűségét, hogy egy betegségen átesett tanulót kiválasztva, az a fiúk közül való; lányok közül való.

Megoldás

a) Az esemény és komplementere mindig teljes eseményrendszert alkot, ezért a B, $\overline{B}$ teljes eseményrendszer. Alkalmazható tehát a teljes valószínűség tétele:

$$P(A)=P(A|B)P(B)+P(A|\overline{B})P(\overline{B})=\frac{50}{100}\cdot\frac{110}{230}+\frac{40}{120}\cdot\frac{120}{230}=\frac{90}{230}\approx 0{,}39,$$

ami megegyezik a közvetlenül számított $P(A)$-val.

b) *Bayes*-tétel felhasználásával számítjuk ki annak valószínűségét, hogy a betegségen átesett tanuló a fiúk közül való:

$$P(B|A)=\frac{P(A|B)P(B)}{P(A)}=\frac{\frac{50}{110}\cdot\frac{100}{230}}{\frac{90}{230}}=\frac{50}{90}\approx 0{,}56;$$

lányok közül való:

$$P(\overline{B}|A)=\frac{P(A|\overline{B})P(\overline{B})}{P(A)}=\frac{\frac{40}{120}\cdot\frac{120}{230}}{\frac{90}{230}}=\frac{40}{90}\approx 0{,}44.$$

2. Egy műhelyben gyártott összes alkatrész 50%-át a G_1 gép 3%-os selejttel, 30%-át a G_2 gép 4%-os selejttel, 20%-át pedig a G_3 gép 5%-os selejttel gyártja.

a) Mi a valószínűsége annak, hogy egy véletlenül kiemelt alkatrész selejtes?

b) Ha a véletlenül kiválasztott alkatrész selejtes, mi a valószínűsége annak, hogy az a G_1 gép terméke?

Megoldás

A selejtes alkatrész választásának eseményét jelölje Z, akkor a teljes valószínűség 2.2. pont (3) képlete szerint

$$P(Z)=P(G_1)P(Z|G_1)+P(G_2)P(Z|G_2)+P(G_3)P(Z|G_3)=$$

$$=0{,}50\cdot 0{,}03+0{,}30\cdot 0{,}04+0{,}20\cdot 0{,}05=0{,}037.$$

Annak valószínűségét, hogy a selejtes alkatrészt a G_1 gép gyártotta a *Bayes*-tétel (2) képletével számíthatjuk ki:

$$P(G_1|Z)=\frac{P(G_1)P(Z|G_1)}{P(G_1)P(Z|G_1)+P(G_2)P(Z|G_2)+P(G_3)P(Z|G_3)}=$$

$$=\frac{0{,}50\cdot 0{,}03}{0{,}50\cdot 0{,}03+0{,}30\cdot 0{,}04+0{,}20\cdot 0{,}05}=\frac{0{,}015}{0{,}037}=\frac{15}{37}\approx 0{,}41.$$

3. Egy kórházban 50 nőt és 10 férfit ápolnak. A férfiak 8%-a, a nők 0,45%-a szívbeteg. A betegségeket leíró kartonok közül véletlenszerűen egyet kiválasztva, azt látjuk, hogy a karton szívbetegé. Mennyi a valószínűsége annak, hogy a karton nő betegé?

Megoldás

Az A esemény jelentse azt, hogy a kiválasztott karton szívbetegé, B_1 jelentse azt, hogy a kiválasztott karton nő betegé, B_2 pedig jelentse azt, hogy a karton férfi betegé.

A B_1 esemény valószínűsége:

$$P(B_1)=\frac{50}{60}=\frac{5}{6},$$

a B_2 esemény valószínűsége:

$$P(B_2)=\frac{10}{60}=\frac{1}{6}.$$

A B_1 esemény A feltétel melletti feltételes valószínűségének kiszámításához a *Bayes*-tételt alkalmazzuk, melyhez szükségünk van az A esemény B_1, és B_2 feltétel melletti feltételes valószínűségére is:

$$P(A|B_1)=\frac{P(AB_1)}{P(B_1)}=\frac{\frac{45}{10000}\cdot\frac{10}{60}}{\frac{10}{60}}=\frac{45}{10000}=\frac{9}{2000},$$

$$P(A|B_2)=\frac{P(AB_2)}{P(B_2)}=\frac{\frac{8}{100}\cdot\frac{50}{60}}{\frac{50}{60}}=\frac{8}{100}=\frac{2}{25}.$$

és így

$$P(B_1|A)=\frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1)+P(A|B_2)P(B_2)}=\frac{\frac{9}{2000}\cdot\frac{5}{6}}{\frac{9}{2000}\cdot\frac{5}{6}+\frac{2}{25}\cdot\frac{1}{6}}\approx 0{,}22.$$

Tehát 0,22 annak a valószínűsége, hogy a szívbetegséget feltüntető kiválasztott karton nő betegé.

2.3. Események függetlensége

Általában függetlennek mondunk két eseményt, ha nincsenek hatással egymásra, vagy más szóval, ha az egyik bekövetkezése nem befolyásolja a másik bekövetkezését. Matematikailag akkor tekintjük függetlennek a két eseményt, ha az egyik bekövetkezése nincs hatással a másik bekövetkezésének valószínűségére.

Definíció. Legyen A és B két tetszőleges véletlen esemény. Az A és B egymástól **független események**, ha

$$P(A\cdot B)=P(A)P(B) \qquad (1)$$

vagyis, ha A és B együttes bekövetkezésének valószínűsége egyenlő A és B valószínűségeinek szorzatával.

Ha A és B független események és $P(B)>0$, akkor $P(A|B)=P(A)$, vagyis az A feltételes valószínűsége a B feltétel mellett egyenlő az A valószínűségével, azaz a B esemény bekövetkezése nem változtatja meg A bekövetkezésének esélyeit. Ha az (1) feltétel nem teljesül, akkor az A és B események *nem függetlenek.*

Általában n eseményt ($n=2,3,\ldots$) függetlennek nevezünk, ha akárhányat kiválasztva közülük, azok együttes bekövetkezésének valószínűsége egyenlő valószínűségeik szorzatával. Az $A_1, A_2, \ldots, A_n$ eseményekről tehát azt mondjuk, hogy *(teljesen) függetlenek,* ha bárhogyan kiválasztva közülük $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ eseményeket, teljesül a

$$P(A_{i_1}\cdot A_{i_2}\cdot\ldots\cdot A_{i_k})=P(A_{i_1})P(A_{i_2})\ldots P(A_{i_k}),\ (k=2,3,\ldots,n)\quad (1^*)$$

egyenlőség, azaz bármely tetszőlegesen kiválasztott k-számú különböző esemény együttes bekövetkezésének valószínűsége az egyes események valószínűségének szorzatával egyenlő. Ez 2^n-n-1 feltétel teljesülését követeli meg.

Ha az (1*) csak $k=2$-re, vagy csak $k=3$-ra, ... teljesül, akkor azt mondjuk, hogy az $A_1, A_2, \ldots, A_n$ események páronként, hármanként, stb.-ként függetlenek.

Példák

1. Egy kocka feldobásának eredménye lehet páros szám és páratlan szám. A páros szám dobásának eseménye legyen A, a páratlan szám dobásáé pedig B. Vizsgáljuk meg az A és B események függetlenségét.

Megoldás

Páros és páratlan szám dobása egyszerre nem következhet be, tehát A és B egymást kizáró események, azaz $P(AB)=0$. Az A és B esemény bekövetkezésének valószínűsége: $P(A)=\frac{3}{6}=\frac{1}{2}$, $P(B)=\frac{3}{6}=\frac{1}{2}$, tehát

$$P(A)\cdot P(B)=\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}.$$

Mivel $0=P(AB)\neq P(A)P(B)=\frac{1}{4}$, az (1) feltétel nem teljesül, ezért az A és B események nem függetlenek.

2. Az A, B és C események függetlenek, ha a három esemény páronként független, azaz

$$P(A\cap B)=P(A)P(B),\ P(A\cap C)=P(A)P(C),\ P(B\cap C)=P(B)P(C)$$

és teljesül még a

$$P(A\cap B\cap C)=P(A)P(B)P(C)\qquad (*)$$

feltétel is. Vegyük észre, hogy a teljes függetlenséghez a páronkénti függetlenség nem elegendő.

3. Két szabályos pénzérme feldobása a $Q=\{FF,FI,IF,II\}$ eseményteret alkotja. Az események egyenlő valószínűségűek. Vizsgáljuk meg az A, B és C események függetlenségét, ha

$$A=\{FF,FI\},\ B=\{FF,IF\},\ C=\{FI,IF\}$$

Megoldás

Az A, B és C események valószínűségei:

$$P(A)=P(B)=P(C)=\frac{2}{4}=\frac{1}{2}$$

és

$$P(AB)=P(\{FF\})=\frac{1}{4},\qquad P(AC)=P(\{FI\})=\frac{1}{4},$$

$$P(BC)=P(\{IF\})=\frac{1}{4}.$$

A páronkénti függetlenség feltételei teljesülnek, ui.

$$P(AB)=P(A)P(B)=\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4},\ \text{stb.}$$

A (*) feltétel azonban nem teljesül, mivel $ABC=\varnothing$ és így $P(ABC)=P(\varnothing)=$ $=0\neq P(A)P(B)P(C)$, vagyis a három esemény nem független.

Megjegyzés

Ha az A és B független események, akkor

a) az $\overline{A}$ és $\overline{B}$ események is függetlenek;

b) az A és $\overline{B}$ valamint az $\overline{A}$ és B események is függetlenek.

Ui. $P(\overline{A}\cap\overline{B})=P(\overline{A\cup B})=1-P(A\cup B)=$

$$=1-P(A)-P(B)+P(A\cap B)=$$
$$=1-P(A)-P(B)+P(A)P(B)=$$
$$=[1-P(A)][1-P(B)]=P(\overline{A})P(\overline{B}).$$

A *b)* állítás hasonlóan látható be.

4. Egy irodában két számítógép egymástól függetlenül működik. Az egyik számítógép műszakonkénti meghibásodásának valószínűsége 0,3, a másik számítógépé pedig 0,2. Mi a valószínűsége annak, hogy legalább egy műszak időtartama alatt egyik számítógép sem hibásodik meg?

Megoldás

Ha A-val jelöljük az egyik számítógép és B-vel a másik számítógép műszakon belüli meghibásodásának eseményét, akkor $P(A) = 0{,}3$, és $P(B) = 0{,}2$. Azt az eseményt, hogy legalább egy műszakon belül egyik számítógép sem hibásodik meg, A és B események ellentétével, azaz $\overline{A}\,\overline{B}$ -vel adhatjuk meg. A függetlenség figyelembevételével

$$P(\overline{A}\,\overline{B}) = P(\overline{A})P(\overline{B}) = (1-0{,}3)\cdot(1-0{,}2) = 0{,}7\cdot 0{,}8 = 0{,}56.$$

Tehát 0,56 annak valószínűsége, hogy egy műszakon belül egyik számítógép sem hibásodik meg.

5. Egy üzem raktárában gumicsizmák és bakancsok vannak. Annak valószínűsége, hogy egy munkás tevékenységéhez csizmát, ill. bakancsot igényel 0,8, ill. 0,5. Ha mindegyik munkás csak egy lábbelit visz el, mennyi a valószínűsége annak az A-val jelölt eseménynek, hogy 6 munkás egymás után csak csizmát vagy 6 munkás egymás után csak bakancsot visz el?

Megoldás

Legyen B esemény az, hogy mind a hat munkás egymás után egy-egy csizmát igényel, C esemény pedig az, hogy mind a hat munkás bakancsot igényel. Nyilvánvaló, hogy a B és C események egymást kizárják, tehát az A valószínűségét a B és C valószínűségének összegeként meghatározhatjuk, azaz

$$P(A) = P(B+C) = P(B) + P(C).$$

Mivel a csizma igénylésének valószínűsége $0{,}8$, így a 6 egymás utáni igénylés valószínűsége:

$$P(B) = 0{,}8^6 = 0{,}262144$$

és hasonlóan:

$$P(C) = 0{,}5^6 = 0{,}015625,$$

vagyis

$$P(A) = P(B) + P(C) = 0{,}262144 + 0{,}015625 \approx 0{,}28.$$

Tehát $0{,}28$ annak valószínűsége, hogy mind a 6 munkás egymás után csizmát vagy bakancsot igényel.

Megjegyzés

1. Ha két esemény, A és B független, vagyis $P(AB)=P(A)P(B)$, és mindkét esemény valószínűsége pozitív, akkor A és B nem lehet egymást kizáró esemény, mivel egyszerre is bekövetkezhetnek, azaz $P(AB) \neq 0$. Ha azonban az egyik esemény valószínűsége 0, akkor $P(AB) = 0$, de AB nem lehetetlen esemény, akkor A és B egymást kizáró események. Hasonlóan, két egymást kizáró esemény csak akkor független, ha legalább az egyik esemény bekövetkezésének valószínűsége 0.

E.2. Ellenőrző kérdések a 2. fejezethez

1. Hogyan definiáljuk a feltételes valószínűséget?

2. Milyen eseménytéren értelmezzük a feltételes relatív gyakoriságot?

3. Hogyan értelmezzük a valószínűségek szorzási tételét?

4. Hogyan alkalmazzuk a fadiagramot valószínűségek kiszámítására?

5. Mit mond ki a teljes valószínűség tétele?

6. Hogyan származtatjuk *Bayes* tételét?

7. Hogyan definiáljuk az események függetlenségét?

V.2. Feladatok a 2. fejezethez

V.2.1. Dobjunk fel két szabályos kockát. Mi annak a valószínűsége, hogy az együttes pontszám 10 vagy 10-nél nagyobb, ha

a) az 1-es kockával 5-öst dobunk.

b) legalább az egyik kockával 5-öst dobunk.

V.2.2. Az $1, 2, \ldots, 9$ számjegyek közül véletlenül válasszunk ki két számjegyet. Ha a két számjegy összege páros, mi annak valószínűsége, hogy mindkét számjegy páratlan?

V.2.3. Bence a jól megkevert 52 lapos bridzs kártyacsomagból négy lapot kapott. Ha a négy lap mindegyike pikk, akkor mi a valószínűsége annak, hogy a következő három lap között is lesz legalább egy pikk? (A pikk, kőr, káró, treff lapok mindegyikéből 13 van.)

V.2.4. Egy dobozban 7 hibátlan és 3 hibás alkatrész van. Ha egymás után három alkatrészt kiveszünk a dobozból, mi a valószínűsége annak, hogy az első kettő hibátlan, a harmadik pedig hibás lesz?

V.2.5. Pongrác a jól megkevert 32 lapos magyar kártyacsomagból 5 lapot kapott. Mi a valószínűsége annak, hogy mindegyik lap zöld?

V.2.6. Számítsuk ki az A eseménynek a B-re vonatkozó, valamint a B eseménynek az A-ra vonatkozó feltételes valószínűségét, ha

$$P(A)=\frac{3}{8},\ P(B)=\frac{5}{8} \text{ és } P(A\cup B)=\frac{3}{4}.$$

V.2.7. Számítsuk ki az

$a)\ P(A|B)$; $b)\ P(B|A)$; $c)\ P(A\cup B)$; $d)\ P(\overline{A}|\overline{B})$; $e)\ P(\overline{B}|\overline{A})$

valószínűségeket, ha $P(A)=\frac{1}{2}$; $P(B)=\frac{1}{3}$ és $P(A\cap B)=\frac{1}{4}$.

V.2.8. Egy évfolyam hallgatóinak 25%-a matematikából, 15%-a fizikából és 10%-a matematikából és fizikából is elégtelenre vizsgázott. Válasszunk ki egy hallgatót az évfolyamból, és állapítsuk meg:

a) Mi a valószínűsége annak, hogy matematikából elégtelen az osztályzata, ha fizikából elégtelen?

b) Mi a valószínűsége annak, hogy fizikából elégtelen az osztályzata, ha matematikából elégtelen?

c) Mi a valószínűsége annak, hogy matematikából vagy fizikából elégtelen az osztályzata?

V.2.9. Egy idős házaspárnál annak valószínűsége, hogy a férfi még 10 évet él $\frac{1}{4}$, annak valószínűsége, hogy a nő még 10 évet él $\frac{1}{3}$. Számítsuk ki annak valószínűségét, hogy

a) mindketten élnek még 10 évet;

b) legalább az egyik él még 10 évet;

c) egyikük sem él még 10 évet;

d) csak a nő él még 10 évet.

V.2.10. Balázs, Robin és Nándor egy céltáblára lőnek. A találat valószínűsége a sorrend szerint $\frac{1}{6};\frac{1}{4};\frac{1}{3}$. Mindegyikük egy lövést ad le. Mi a valószínűsége annak, hogy

a) közülük csak egy talál a céltáblába;

b) ha egy találat van, akkor azt a lövést Balázs adta le?

V.2.11. Egy iskola 2000 diákja közül a fiúk száma 1500, a lányoké 500. A szemorvos minden diákot megvizsgált és azt találta, hogy 60 fiú és 50 lány rövidlátó. Mekkora valószínűséggel

a) rövidlátó egy véletlenszerűen kiválasztott diák;

b) rövidlátó egy véletlenszerűen kiválasztott lány;

c) rövidlátó egy véletlenszerűen kiválasztott fiú;

d) lány egy véletlenszerűen kiválasztott rövidlátó?

V.2.12. Egy cipő nagykereskedő X, Y és Z gyártól vásárol cipőket 25, 35, ill. 40%-os arányban. Az X, Y, ill. Z gyár cipői 5, 4, ill. 2%-ban hibásaknak bizonyultak. Mekkora annak a valószínűsége, hogy egy véletlenszerűen kiválasztott cipő

a) hibás,

b) hibátlan,

c) hibás és azt az Y gyár készítette,

d) hibátlan és azt a X gyár készítette?

V.2.13. Egy gyárban két szalagon ugyanazt a terméktípust szerelik össze. Az első szalag mellett gyakorlott szakmunkások 10%-os selejttel, a második szalag mellett kezdő betanított munkások 20%-os selejttel végzik a szerelést. Véletlenszerűen kiválasztunk mindkét szalagról lekerült termékek közül egyet-egyet. Mekkora annak valószínűsége, hogy

a) mindkettő hibás,

b) egyik sem hibás,

c) legalább az egyik hibás,

d) pontosan egyik hibás?

V.2.14. Egy műhelyben egymástól függetlenül öt forgácsoló gép működik. Az egyes gépek napi üzemszerű működésének valószínűsége 0,9. Mekkora annak valószínűsége, hogy egy véletlenszerűen kiválasztott napon

a) mind az öt gép üzemszerűen működik,

b) egyik gép sem működik,

c) legalább egy gép működik,

d) legfeljebb egy gép működik?

V.2.15. Egy üzem három különböző termelékenységű gépe ugyanazt a terméket gyártja. Az X gép naponta 10 db-ot, az Y gép 15 db-ot, a Z gép pedig 25 db-ot gyárt. Az X, Y, Z gép naponta gyártott termékei között átlagosan szépséghibás 0,3, 0,9, ill. 0,5 db. Az egy nap gyártott összes terméket tartalmazó ládából véletlenszerűen kiveszünk egyet és megállapítjuk, hogy szépséghibás. Mekkora a valószínűsége annak, hogy azt az X gép gyártotta?

HARMADIK FEJEZET

Valószínűségi változók és jellemzőik

Egy kísérlet által meghatározott valószínűségi mező ismeretében rendelkezünk azokkal az információkkal, amelyek lehetővé teszik a kísérlet részletes vizsgálatát. Az eseménytér eseményeinek vizsgálata helyett célszerűbb az eseményekhez rendelt, megfelelően értelmezett valós számokkal végezni vizsgálatainkat. Amennyiben az eseménytéren értelmezünk egy valós értékű függvényt, akkor az analízis jól kidolgozott eszközeivel is elemezhetjük a véletlen jelenségek törvényszerűségeit.

Tekintsük például az első fejezetben tárgyalt kockadobással végzett kísérletet. Láttuk, hogy a Q eseménytér hat elemi eseményt tartalmaz: A_1 az 1-es dobás, A_2 a 2-es dobás,... A_6 a 6-os dobás eseménye. Mindegyik esemény valószínűsége $\frac{1}{6}$, azaz $P(A_i)=\frac{1}{6}$, $(i=1,2,\ldots,6)$. Ezt a jelenséget a következő módon is megfogalmazhatjuk: A kockadobás kimenetelét egy olyan X változónak tekintjük, amelynek értékei, a lehetséges eredmények, az 1, 2, 3, 4, 5 és a 6 számok. Az X minden dobásnál ezek közül csak egy értéket vehet fel, éspedig $\frac{1}{6}$ valószínűséggel. Vizsgálhatjuk pl. a 3-nál nagyobb szám dobásának, vagy a 2 és 5 közötti szám dobásának valószínűségét is stb. Az X változó jelentheti pl. a Balaton vizének véletlenszerűen kiválasztott időpontokban mért hőmérsékletét, egy célgépen gyártott termék véletlenszerűen kiválasztott darabjának valamely méretét, stb. Az ilyen változó értékei véletlentől függő számértékek.

Definíció. Ha egy kísérlet minden lehetséges kimeneteléhez, azaz a kísérlet teljes eseményrendszere mindegyik elemi eseményéhez egyértelműen hozzárendelünk egy-egy valós számot, akkor a Q halmazon egy valós értékű függvényt értelmezünk. Ezt a függvényt **valószínűségi változónak** nevezzük és X-szel (általában az ábécé dőlt nagybetűivel) jelöljük.

A valószínűségi változó jelölésére a X helyett a ξ görög betűt is használjuk.

Ha egy kísérlet során az A_i elemi esemény következik be, és ehhez az x_i értéket rendeljük, akkor ezt az X valószínűségi változó egy *lehetséges értékének* mondjuk és az $X = x_i$ jelölést használjuk. Ekkor az A_i elemi esemény valószínűségét $P(X = x_i)$ jelöli.

A következő pontokban a diszkrét és a folytonos valószínűségi változókkal valamint azok jellemzőivel ismerkedünk meg.

3.1. Diszkrét valószínűségi változó

Definíció. Egy X valószínűségi változót **diszkrétnek** nevezünk, ha csak megszámlálható számú értéket vehet fel, azaz lehetséges értékeinek halmaza véges vagy megszámlálható halmaz. Pl. diszkrét valószínűségi változó a kockadobásnál kapott pontszámokkal meghatározott változó.

Legyen az X valószínűségi változó lehetséges értékeinek halmaza $\{x_1, x_2, \dots, x_n\}$, akkor ezt röviden $X(Q)$-val jelöljük, azaz $X(Q) = \{x_1, x_2, \dots, x_n\}$. Az $X = x_i$ esemény bekövetkezése azt jelenti, hogy a valószínűségi változó értéke az $\{x_1, x_2, \dots, x_n\}$ halmazba esik. Ennek értelmében a *biztos esemény* az, hogy a valószínűségi változó valamelyik lehetséges értékét felveszi. Azt is mondjuk, hogy a biztos esemény valószínűsége, azaz 1, a valószínűségi változó lehetséges értékein oszlik el.

Az X valószínűségi változót adottnak tekintjük, ha ismertek a

$$p_i = P(X = x_i) \quad (i = 1, 2, \dots, n)$$

valószínűségek. Ezt úgy is mondjuk, hogy p_i annak a valószínűsége, hogy az X valószínűségi változó felveszi az x_i $(i = 1, 2, \dots, n)$ értéket.

Ha $X(Q) = \{x_1, x_2, \dots, x_n\}$, akkor az x_i $(i = 1, 2, \dots, n)$ pontokban képzett $p_i = P(X = x_i)$ valószínűségekkel egy **valószínűségi mezőt** kapunk.

Definíció. Az $X(Q)$ halmazon

$$v(x_i) = p_i = P(X = x_i) \quad (i = 1, 2, \dots, n) \tag{1}$$

képlettel adott függvényt, az X változó **valószínűségeloszlásának**, az

$$F(x) = \sum_{x_i < x} p_i \tag{1*}$$

függvényt, ahol az összegezést mindazon i-re végezzük, amelyre $x_i < x$ egyenlőtlenség fennáll, X **diszkrét eloszlásfüggvényének** nevezzük. A valószínűségeloszlást táblázatosan adjuk meg:

x_i	x_1	x_2	....	x_n
$v(x_i)$	$v(x_1)$	$v(x_2)$	....	$v(x_n)$

Az (1) formula azt fejezi ki, hogy az X valószínűségi változó az x_i értéket $p_i = v(x_i)$ valószínűséggel veszi fel. A változó eloszlását tehát egyértelműen jellemzi az, hogy a változó milyen értékeket milyen valószínűséggekkel vesz fel.

Ha az $x_1, x_2, \dots, x_n$ jelölik azokat a különböző értékeket, amelyeket egy diszkrét X valószínűségi változó pozitív valószínűséggel vesz fel, akkor a $\{X = x_k\}$ $(k = 1, 2, \dots, n)$ események *teljes eseményrendszert* alkotnak.

Az értelmezésből következik, hogy az X változó eloszlása kielégíti a következő feltételeket:

$$a)\ v(x_i) \geq 0, \quad \text{és} \quad b)\ \sum_{i=1}^{n} v(x_i) = 1.$$

Példák

1. Egy kockadobás 1, 2, 3, 4, 5, 6 kimenetelei legyenek az X valószínűségi változó értékei. Adjuk meg az X valószínűségeloszlását és eloszlásfüggvényének értékeit.

Megoldás

Mivel a kockadobás kimenetelei közül bármelyik $\frac{1}{6}$ valószínűséggel következhet be, így X valószínűségeloszlása:

x_i	1	2	3	4	5	6
p_i	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Mivel X 1-nél kisebb értéket nem vehet fel, így $x \le 1$ esetén $F(x)=P(X<x)=0$. (Vagyis $\{X<x\}$ lehetetlen esemény). Ha $x>6$, akkor $F(x)=P(X<x)=P(Q)=1$, mivel az $\{X<x\}$ biztos esemény. Ha $1<x\le 2$, akkor X valószínűségi változó az (1*) formulának megfelelően csak x-nél kisebb értéket vehet fel, vagyis az 1 számot, tehát $F(x)=P(X=1)=\frac{1}{6}$. Hasonlóan kapjuk, hogy $2<x\le 3$ esetén X felveheti az 1 és 2 értéket, azaz $F(x)=P(X<x)=\frac{2}{6}$, stb. Tehát az X eloszlásfüggvénye:

$$F(x)=\begin{cases} 0, \text{ ha } x\le 1 \\ p_1=\frac{1}{6}, \text{ ha } 1<x\le 2 \\ p_1+p_2=\frac{2}{6}, \text{ ha } 2<x\le 3 \\ p_1+p_2+p_3=\frac{3}{6}, \text{ ha } 3<x\le 4 \\ p_1+p_2+p_3+p_4=\frac{4}{6}, \text{ ha } 4<x\le 5 \\ p_1+p_2+p_3+p_4+p_5=\frac{5}{6}, \text{ ha } 5<x\le 6 \\ \sum_{i=1}^{6} p_i=1, \text{ ha } x>6 \end{cases}$$

Az X valószínűségi változó eloszlásának ábrázolását l. a 3.1.2. pontban, az eloszlásfüggvény ábrázolását l. a 3.2.1. pontban.

2. Egy dobozban négy jó és három hibás tranzisztor van. Vegyünk ki a dobozból véletlenszerűen négy tranzisztort. Az X valószínűségi változó értéke rendre legyen a jó tranzisztorok száma. Táblázatosan adjuk meg a valószínűségi változó eloszlását.

Megoldás

Az X diszkrét valószínűségi változó, értékei pedig az $x_k=k$ $(k=1, 2, 3, 4)$ számok. Annak az eseménynek a valószínűsége, hogy k darab jó tranzisztort vettünk ki a dobozból a

$$p_k=P(X=k)=\frac{\binom{4}{k}\binom{3}{4-k}}{\binom{7}{4}} \quad (k=1, 2, 3, 4)$$

formulával számítható ki, tehát

$$p_1=P(X=1)=\frac{\binom{4}{1}\binom{3}{3}}{\binom{7}{3}}=\frac{4}{35}=0{,}11; \quad p_2=P(X=2)=\frac{\binom{4}{2}\binom{3}{2}}{\binom{7}{3}}=\frac{18}{35}=0{,}51;$$

$$p_3=P(X=3)=\frac{\binom{4}{3}\binom{3}{1}}{\binom{7}{3}}=\frac{12}{35}=0{,}34; \quad p_4=P(X=4)=\frac{\binom{4}{4}\binom{3}{0}}{\binom{7}{3}}=\frac{1}{35}=0{,}03.$$

Tehát a kérdéses X diszkrét valószínűségi változó táblázatosan adott eloszlása:

x_i	1	2	3	4
p_i	0,11	0,51	0,34	0,03

$\sum p_i = 1$

(A kerekítési hibák miatt tér el az összeg az 1-től.)

3.1.1. A várható érték

Ha egy valószínűségi változóra vonatkozólag független kísérletsorozatot végzünk, akkor a változó által felvett értékek egy meghatározott érték körül ingadoznak. A valószínűségi változót jól tudjuk jellemezni azzal a valós számmal, amely körül a változó tapasztalati értékeinek az átlaga ingadozik. Ezt a valós számot *várható értéknek* nevezzük.

Definíció. Az X diszkrét valószínűségi változó $M(X)$-szel jelölt **várható értékének**, ha eloszlása a

$$\nu(x_i)=p_i=P(X=x_i) \quad (i=1,2,\dots,n)$$

összefüggéssel adott, az

$$M(X)=x_1\nu(x_1)+x_2\nu(x_2)+\dots+x_n\nu(x_n)=\sum_{i=1}^{n} x_i\nu(x_i) \qquad (2)$$

képlettel meghatározott számot nevezzük.

Más megfogalmazásban, az $M(X)$ az X valószínűségi változó lehetséges értékeinek a **súlyozott átlaga**. Szokásos jelölése még $E(X)$, M, μ, m_x vagy m.

A (2) képletet az X diszkrét valószínűségi változó várható értékének kiszámítására, ha az x_i értékeket p_i valószínűséggel veszi fel, az

$$M(X)=\sum_{i=1}^{n} p_i x_i \qquad (2^*)$$

alakban használjuk.

Példák

1. Két szabályos kockával végzett kockadobások esetében a Q eseményteret 36 rendezett számpár alkotja:

$$Q=\{(1,1),(1,2),\ldots,(6,5),(6,6)\}.$$

X értéke legyen a Q elemeit képező $(a,\ b)$ számpárokból mindig a nagyobbik számmal egyenlő, azaz $X(a,b)=\max(a,b)$. Ekkor az X valószínűségi változó lehetséges értékeinek halmaza:

$$X(Q)=\{1,2,3,4,5,6\}.$$

Számítsuk ki az X változó $p_i=\nu(x_i)$ eloszlását és $M(X)$ várható értékét.

Megoldás

Az $\{(1,1)\}$ esemény csak egyféleképpen következhet be, a „nagyobbik szám 2" esemény háromféleképpen következhet be, stb. tehát

$$\nu(1)=P(X=1)=P(\{(1,1)\})=\frac{1}{36};$$

$$\nu(2)=P(X=2)=P(\{(2,1),(2,2),(1,2)\})=\frac{3}{36};$$

$$\nu(3)=P(X=3)=P(\{(3,1),(3,2),(3,3),(2,3),(1,3)\})=\frac{5}{36};$$

s hasonlóan kapjuk, hogy

$$\nu(4)=P(X=4)=\frac{7}{36};$$

$$\nu(5)=P(X=5)=\frac{9}{36};$$

$$\nu(6)=P(X=6)=\frac{11}{36}.$$

A kapott eredmények alapján az X valószínűségeloszlását táblázatosan is megadhatjuk:

x_i	1	2	3	4	5	6
$\nu(x_i)$	1/36	3/36	5/36	7/36	9/36	11/36

Az X várható értéke a (2) formula szerint számolva:

$$M(X)=\sum_{i=1}^{n} x_i\nu(x_i)=1\frac{1}{36}+2\frac{3}{36}+3\frac{5}{36}+4\frac{7}{36}+5\frac{9}{36}+6\frac{11}{36}=\frac{161}{36}\approx 4{,}47.$$

Természetesen sosem lesz a két kocka dobásánál megjelenő nagyobbik szám 4,47, pusztán az egymás utáni dobások $\max(a,b)$ pontszámának átlaga fog 4,47 körül ingadozni.

2. Egy kockával végzett kísérletnél kapott valószínűségeloszlásból (l. a 3.1. pont példáját) számított várható érték:

$$M(X)=\sum_{i=1}^{6} p_i x_i=\frac{1}{6}(1+2+3+4+5+6)=\frac{21}{6}=3{,}5$$

is csak azt mutatja, hogy az egymásutáni dobások pontértékének átlaga a 3,5 várható érték körül fog ingadozni. Pl. ha kétszer 100 dobásos kísérletnél az egyes dobások pontjait összeadva 346-ot, ill. 349-et kaptunk eredményül, akkor ezek átlagát képezve $346/100=3{,}46$, ill. $349/100=3{,}49$, azaz mindkét esetben 3,5-hez közeli, de attól eltérő értéket kaptunk.

Megjegyzés

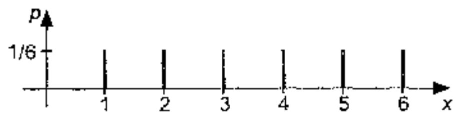
Tegyük fel, hogy az X valószínűségi változó végtelen sok, $x_1,x_2,\ldots,x_n,\ldots$ értéket $\nu(x_1),\nu(x_2),\ldots,\nu(x_n),\ldots$ valószínűséggel vesz fel, akkor abban az esetben, ha $\sum_{i=1}^{\infty}\nu(x_i)x_i$ végtelen sor abszolút konvergens, azaz $\sum_{i=1}^{\infty}\nu(x_i)|x_i|<+\infty$, a várható értéket

$$M(X)=\sum_{i=1}^{\infty}\nu(x_i)x_i$$

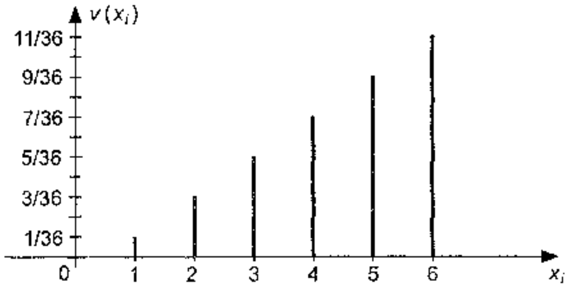
képlettel számítjuk.

3.1.2. Vonaldiagram és hisztogram

A diszkrét valószínűségi eloszlásokat *vonaldiagrammal* vagy *hisztogrammal* szemléltethetjük. A **vonaldiagramot** úgy készítjük, hogy a *Descartes*-féle derékszögű koordinátarendszer x-tengelyén kijelölt pontokhoz az X lehetséges értékeinek megfelelő számokat írjuk egy adott távolságegységgel, és a pontokra olyan – x-tengelyre merőleges – egyeneseket illesztünk, amelyek hossza arányos a szóban forgó pontokhoz tartozó valószínűségekkel. A 3.1. pontban tárgyalt példa valószínűségi eloszlásának vonaldiagramját a 10. ábra, a 3.1.1. pontban tárgyalt 1. példa valószínűségi eloszlásának vonaldiagramját a 11. ábra szemlélteti.



10. ábra. X valószínűségi változó eloszlásának vonaldiagramja



11. ábra. X valószínűségi változó eloszlása

Az ábrán azt rajzoltuk fel, hogy a valószínűségi változó az 1 értéket $\frac{1}{36}$, a 2 értéket $\frac{3}{36}$, és így tovább, a 6 értéket $\frac{11}{36}$ valószínűséggel veszi fel.

Példa

A 3.1.1. pont példájának Q eseményterén most a két kockával dobott pontszámok összege legyen az X értéke, azaz

$$X(a,b) = a + b$$

képlettel képezzük a valószínűségi változó értékeit.

Megoldás

$$X(Q) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$

Az X valószínűségi változó $v(x_i)$ eloszlásának táblázata:

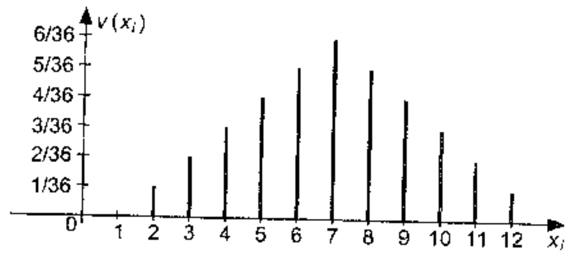
i	1	2	3	4	5	6	7	8	9	10	11
x_i	2	3	4	5	6	7	8	9	10	11	12
$v(x_i)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Pl. $v(5) = P(X = 5) = P(\{(1,4),(2,3),(3,2),(4,1)\}) = \frac{4}{36}$.

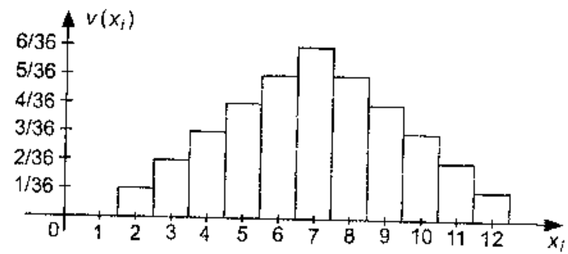
Az X várható értéke:

$$M(X) = \sum_{i=1}^{11} x_i v(x_i) = 2\cdot\frac{1}{36} + 3\cdot\frac{2}{36} + \ldots + 11\cdot\frac{2}{36} + 12\cdot\frac{1}{36} = \frac{252}{36} = 7.$$

Az X valószínűségi változó eloszlását a 12. ábra szemlélteti.



12. ábra. X valószínűségi változó eloszlásának vonaldiagramja



13. ábra. Az X valószínűségi változó eloszlásának hisztogramja

Ha a valószínűségi eloszlást minden x_i-nél azonos szélességű, egymáshoz illeszkedő téglalapokkal ábrázoljuk, és e téglalapok területösszege pontosan 1, akkor **hisztogramot** kapunk. A téglalapok szélességét tehát úgy kell megválasztani, hogy összeadva az egyes téglalapok alapjának $v(x_i)$ -vel (ill. a $v(x_i)$ -vel arányos téglalap magassággal) képzett szorzatát eredményül pontosan 1-et kapjunk. E példánál a téglalapok alapját 1-nek választva a hisztogramot a 13. ábra szemlélteti. Vegyük észre, hogy a két tengelyen az egység megválasztása lényegesen eltér.

Példa

Egy kockajátékos a banktól páros szám dobása esetén a pontszám 18-szorosát kapja forintban számolva, páratlan szám dobása esetén pedig a játékos fizet a banknak, éspedig a pontszám 24-szeresét. Számítsuk ki a várható értéket.

Megoldás

Az eseménytér: $Q = \{1,2,3,4,5,6\}$. Az X valószínűségi változó értékkészlete (a játékos nyereségét pozitív, veszteségét negatív számmal jelölve):

$X(1) = -1\cdot 24 = -24;$ $\quad X(2) = 2\cdot 18 = 36;$ $\quad X(3) = -3\cdot 24 = -72;$

$X(4) = 4\cdot 18 = 72;$ $\quad X(5) = -5\cdot 24 = -120;$ $\quad X(6) = 6\cdot 18 = 108,$

tehát

$$X(Q) = \{36, 72, 108, -24, -72, -120\}$$

Egy szabályos kocka feldobása esetén bármely pontszám bekövetkezése $\frac{1}{6}$ valószínűségű, tehát az X változó $v(x_i)$ eloszlása táblázatosan:

x_i	36	72	108	−24	−72	−120
$v(x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

A várható érték:

$$M(X) = 36\cdot\frac{1}{6} + 72\cdot\frac{1}{6} + 108\cdot\frac{1}{6} - 24\cdot\frac{1}{6} - 72\cdot\frac{1}{6} - 120\cdot\frac{1}{6} = \frac{216-216}{6} = 0.$$

Az ilyen rendszerű játékot, vagyis ahol a várható érték 0, **korrektnek** nevezzük. A játékos szempontjából **kedvező** a játék, ha a várható érték pozitív, és **kedvezőtlen**, ha negatív.

Példa

Tegyük fel, hogy a kockadobás játékszabálya most a következő: prímszám dobásakor a játékos a pontszámmal azonos forintot nyer, nem prímszám dobásakor pedig a pontszámmal azonos forintot veszít. Számítsuk ki a várható értéket.

Megoldás

Az előző példához hasonlóan végezzük a számításokat. Az X valószínűségi változó $v(x_i)$ eloszlása táblázatosan:

x_i	2	3	5	−1	−4	−6
$v(x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

A negatív számok a játékos veszteségeit fejezik ki, ha nem prímszámot dob. A játék várható értéke:

$$M(X) = 2\cdot\frac{1}{6} + 3\cdot\frac{1}{6} + 5\cdot\frac{1}{6} - 1\cdot\frac{1}{6} - 4\cdot\frac{1}{6} - 6\cdot\frac{1}{6} = -\frac{1}{6}.$$

A játékszabály a *játékos számára kedvezőtlen*, mivel a várható érték negatív.

3.1.3. A szórásnégyzet (variancia) és a szórás (diszperzió)

Milyen további számértékekkel jellemezhetjük egy kísérletsorozat eredményeit?

Különböző jelenségek jellemzésére köznapi értelemben is gyakran használjuk az átlagértéket, amely megfeleltethető a várható értéknek. Azonban pusztán az átlagértékből levonható következtetés könnyen megtévesztő, hamis eredményre vezethet. Pl. a 60 éves nagyapa és a 6 éves unoka átlagéletkora 33 év, ami a nagyapára vonatkozólag igen kedvező, de nem úgy az unokára. Az X valószínűségi változó eloszlásának a jellemzésére tehát a várható érték mellett a várható értéktől való átlagos eltérésnek a mértéke is fontos információt szolgáltat.

Egy jelenségre vonatkozó kísérletsorozat mérési adatainak jellemzésére tehát rendszerint megadják az egyes mérési adatok eltérését átlaguktól, az átlagtól való eltérések négyzeteinek átlagát, valamint az utóbbi négyzetgyökét. Ennek megfelelően egy valószínűségi változó várható értékének kiszámítása mellett azt is megvizsgáljuk, hogy a változó értékei mennyit térnek el a várható értéktől, ill. mennyire „szórnak" a várható érték körül.

Legyen adott az X diszkrét valószínűségi változó és eloszlása:

x_i	x_1	x_2	....	x_n
$v(x_i)$	$v(x_1)$	$v(x_2)$	....	$v(x_n)$

Definíció. Az X valószínűségi változó **szórásnégyzetének (varianciájának)** az $(X-m)^2$ valószínűségi változó várható értékét nevezzük, azaz a

$$D^2(X)=\sum_{i=1}^{n}(x_i-m)^2 v(x_i)=M((X-m)^2) \qquad (1)$$

formulával adott számot, ahol m az X várható értéke ($m=M(X)$). A szórásnégyzet szokásos jelölése még s^2, σ^2 és Var(X).

A szórásnégyzet jellemzi tehát az X-re vonatkozó egyes mérési adatok eltérését átlaguktól. Más szóval az adatok négyzetes középhibája $D^2(X)$ körül ingadozik.

Definíció. A $D^2(X)$ szórásnégyzet pozitív négyzetgyökét az X valószínűségi változó **szórásának (standard eltérésének)** nevezzük, azaz

$$D(X)=\sqrt{M((X-m)^2)}. \qquad (2)$$

A szórás szokásos jelölése még s, σ és $\sqrt{\mathrm{Var}(X)}$.

A szórás tehát a valószínűségi változó várható értéke körüli ingadozásainak átlagos nagyságát jellemzi.

Az X valószínűségi változó szórásnégyzetének egyszerűbb kiszámítását teszi lehetővé a következő tétel:

Tétel

$$D^2(X)=\sum x_i^2 v(x_i)-m^2=M(X^2)-m^2. \qquad (3)$$

A tétel bizonyításához a $\sum_{i=1}^{n}x_i v(x_i)=m$ és a $\sum_{i=1}^{n}v(x_i)=1$ összefüggéseket használjuk fel:

$$\sum_{i=1}^{n}(x_i-m)^2 v(x_i)=\sum_{i=1}^{n}(x_i^2-2mx_i+m^2)v(x_i)=$$

$$=\sum_{i=1}^{n}x_i^2 v(x_i)-2m\sum_{i=1}^{n}x_i v(x_i)+m^2\sum_{i=1}^{n}v(x_i)=$$

$$=\sum_{i=1}^{n}x_i^2 v(x_i)-2m^2+m^2=\sum_{i=1}^{n}x_i^2 v(x_i)-m^2.$$

A (3) képletet az X diszkrét valószínűségi változó szórásnégyzetének kiszámítására, ha az x_i értékeket p_i valószínűséggel veszi fel, a

$$D^2(X)=\sum_{i=1}^{n}x_i^2 p_i-\left(\sum_{i=1}^{n}x_i p_i\right)^2 \qquad (3^*)$$

alakban használjuk.

Példák

1. Számítsuk ki az X valószínűségi változó

i	1	2	3	4	5
x_i	−1	0	1	2	3
$v(x_i)$	0,3	0,1	0,1	0,3	0,2

táblázattal adott eloszlása alapján várható értékét, szórásnégyzetét és szórását.

Megoldás

A várható érték:

$$m=M(X)=\sum_{i=1}^{5}x_i v(x_i)=-1\cdot 0{,}3+0\cdot 0{,}1+1\cdot 0{,}1+2\cdot 0{,}3+3\cdot 0{,}2=1.$$

A szórásnégyzet:

$$D^2(X)=M(X^2)-m^2=(-1)^2\cdot 0{,}3+0^2\cdot 0{,}1+1^2\cdot 0{,}1+2^2\cdot 0{,}3+\ 3^2\cdot 0{,}2-1^2=2{,}4.$$

A szórás:

$$D(X)=\sqrt{2{,}4}=1{,}5492.$$

2. Dobjunk fel két kockát, és az X valószínűségi változó értéke legyen a két pontszám közül mindig a kisebb pontszámmal egyenlő. Számítsuk ki X eloszlását, várható értékét és szórását.

Megoldás

A lehetséges esetek száma: 36. A legkisebb pontszám 1, két kockát feldobva valamelyiken az 1 pont 11 esetben fordulhat elő, a 2 mint kisebb pontszám 9 esetben fordulhat elő,..., a 6 pont csak egy esetben, dupla hatos dobása esetén fordulhat elő „kisebb" számként. Így az eloszlás táblázata:

x_i	1	2	3	4	5	6
p_i	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

A várható érték: $m = M(X) = \sum_{i=1}^{6} x_i p_i =$

$$= 1\cdot\frac{11}{36} + 2\cdot\frac{9}{36} + 3\cdot\frac{7}{36} + 4\cdot\frac{5}{36} + 5\cdot\frac{3}{36} + 6\cdot\frac{1}{36} = 2{,}53.$$

A szórásnégyzet: $D^2(X) = \mathrm{Var}(X) = \sum_{i=1}^{6} x_i^2 p_i - m^2 =$

$$= \frac{1}{36}(1^2\cdot 11 + 2^2\cdot 9 + 3^2\cdot 7 + 4^2\cdot 5 + 5^2\cdot 3 + 6^2\cdot 1) - 2{,}53^2 =$$
$$= 8{,}36 - 6{,}40 = 1{,}96.$$

A szórás: $D(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{1{,}96} = 1{,}4.$

Megjegyzések

1. A várható értékre, a szórásnégyzetre és a szórásra érvényesek az alábbi tételek:

a) $M(kX) = kM(X),\ k \in R;$

b) $M(X+k) = M(X) + k;$

c) $M(X_1 + X_2 + \ldots + X_n) = M(X_1) + M(X_2) + \ldots + M(X_n),$

ahol $X_1, X_2, \ldots, X_n$ ugyanazon a Q eseménytéren értelmezett valószínűségi változók;

d) $D^2(X+k) = D^2(X);$

e) $D^2(aX+k) = a^2 D^2(X),\ a \in R;$

f) $D(X+k) = D(X);$

g) $D(aX+k) = |a|\cdot D(X).$

2. Gyakran a számítások egyszerűbbek, ha az X valószínűségi változó helyett az

$$X^* = \frac{X - M(X)}{D(X)}, \qquad (D(X) > 0) \qquad (4)$$

valószínűségi változóra térünk át, mert ennek várható értéke 0 és szórásnégyzete 1, azaz

$$M(X^*) = 0,\ D^2(X^*) = 1. \qquad (*)$$

Az X^*-ot **normált** vagy **standard valószínűségi változónak** nevezzük.

A (*) állításokat az *a*), *b*), *d*), *e*) tételek felhasználásával igazolhatjuk, ui.

$$M(X^*) = M\left(\frac{X - M(X)}{D(X)}\right) = \frac{1}{D(X)} M(X - M(X)) =$$
$$= \frac{1}{D(X)}(M(X) - M(X)) = 0;$$

és

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

3. Egy pozitív X valószínűségi változó szórásának és várható értékének

$$\frac{D(X)}{M(X)}, \qquad (M(X) \neq 0)$$

hányadosát X **relatív szórásának** nevezzük.

Pl. a 2. példa valószínűségi változójának relatív szórása: $\frac{1{,}4}{2{,}53} = 0{,}553.$

3.1.4. Valószínűségi változók együttes- és peremeloszlásai

Miként vizsgálható két valószínűségi változó kapcsolata?

Legyen X és Y a Q eseménytér két valószínűségi változója:

$$X(Q) = \{x_1, x_2, \ldots, x_n\} \text{ és } Y(Q) = \{y_1, y_2, \ldots, y_m\}.$$

Képezzük az

$$X(Q)\times Y(Q)=\{(x_1,y_1),(x_1,y_2),\dots,(x_n,y_m)\}$$

Descartes-szorzatukat és minden (x_i,y_j) rendezett párhoz rendeljük hozzá a $P(X=x_i,Y=y_j)$ valószínűséget, melyet $w(x_i,y_j)$-vel jelölünk.

Definíció. Az $X(Q)\times Y(Q)$ halmazon értelmezett

$$w(x_i,y_j)=P(X=x_i,Y=y_j),$$
$$(i=1,2,\dots,n;j=1,2,\dots,m) \qquad (1)$$

függvényt az X és Y, ill. $(X;Y)$ **valószínűségi vektorváltozó együttes eloszlásának** vagy **együttes diszkrét valószínűségeloszlásnak** nevezzük.

Az X és Y együttes eloszlását az alábbi táblázattal adjuk meg:

$X\setminus Y$	y_1	y_2	…	y_m	$\sum$sor
x_1	$w(x_1,y_1)$	$w(x_1,y_2)$	…	$w(x_1,y_m)$	$v(x_1)$
x_2	$w(x_2,y_1)$	$w(x_2,y_2)$	…	$w(x_2,y_m)$	$v(x_2)$
⋮	⋮	⋮	⋮	⋮	⋮
x_n	$w(x_n,y_1)$	$w(x_n,y_2)$	…	$w(x_n,y_m)$	$v(x_n)$
$\sum$oszlop	$u(y_1)$	$u(y_2)$	…	$u(y_m)$	

A táblázatban a $v(x_i)$ jelöli az i-edik sor elemeinek, $u(y_j)$ pedig a j-edik oszlop elemeinek az összegét:

$$v(x_i)=\sum_{j=1}^{m}w(x_i,y_j);\quad u(y_j)=\sum_{i=1}^{n}w(x_i,y_j) \qquad (2)$$

amelyeket **marginális (határ-)eloszlásoknak** vagy **peremeloszlásoknak** nevezünk az X, ill. az Y változókra vonatkozólag.

Nyilvánvaló, hogy a w együttes eloszlás eleget tesz a

$$w(x_i,y_j)\ge 0 \quad \text{és} \quad \sum_{i=1}^{n}\sum_{j=1}^{m}w(x_i,y_j)=1$$

feltételeknek.

Az X és Y valószínűségi változók **együttes várható értékét**

$$M(XY)=\sum_{i,j}x_iy_jw(x_i,y_j) \qquad (3)$$

formulával számítjuk ki, ahol az összegezést i, és j indexek minden lehetséges értékére végezzük, ugyanúgy, mint az előbbi képletben.

Ha X és Y *független valószínűségi változók*, akkor az együttes eloszlás:

$$w(x_i,y_j)=v(x_i)\cdot u(y_j).$$

Példa

Két szabályos kockával dobásokat végezve – mint azt a 3.1.1. pont példájában láttuk – a Q eseményteret 36 rendezett számpár alkotja:

$$Q=\{(1,1),(1,2),\dots,(6,5),(6,6)\}.$$

Az X valószínűségi változó értéke a Q elemeit képező (a,b) számpárokból mindig a nagyobbik számmal legyen egyenlő, azaz $X(a,b)=\max(a,b)$, az Y változó értéke pedig mindig az (a,b) számpárok összege legyen, azaz $Y(a,b)=a+b$, vagyis

$$X(Q)=\{1,2,3,4,5,6\} \text{ és } Y(Q)=\{2,3,4,5,6,7,8,9,10,11,12\}.$$

Az $X(Q)\times Y(Q)$ *Descartes*-szorzat a Q^* eseményteret állítja elő:

$$X(Q)\times Y(Q)=\{(1,2),(1,3),\dots,(6,11),(6,12)\}.$$

Készítsük el az X és Y változók együttes eloszlásának táblázatát és számítsuk ki

a) az $M(XY)=\sum_{i,j}x_iy_jw(x_i,y_j)$ várható értéket,

b) adjuk meg az $F(x,y)$ eloszlásfüggvény néhány értékét,

c) X és Y változók perem-eloszlásfüggvényét,

d) számítsuk ki az eloszlásfüggvény értékét $x=2{,}5$, $y=3{,}6$ helyen együttesen, ill. külön-külön.

Megoldás

A $\max(a,b)=1$ és $a+b=2$ csak egyszer fordulhat elő a lehetséges kimenetelek közül, tehát $w(1,2)=\frac{1}{36}$. Ha a számpárok közül egyik sem nagyobb 1-nél, akkor 2-nél nagyobb dobás nem jöhet létre, tehát $w(1,y_j)=0$, ha $y_j=3,4,\dots,12$. Hasonlóan kapjuk a $w(2,2)=0$, $w(2,3)=\frac{2}{36}$, $w(2,4)=\frac{1}{36}$, $w(2,5)=0$, stb értékeket. Az együttes eloszlás és a peremeloszlás táblázata, tehát a következő:

$X\setminus Y$	2	3	4	5	6	7	8	9	10	11	12	$\sum$
1	1/36	0	0	0	0	0	0	0	0	0	0	1/36
2	0	2/36	1/36	0	0	0	0	0	0	0	0	3/36
3	0	0	2/36	2/36	1/36	0	0	0	0	0	0	5/36
4	0	0	0	2/36	2/36	2/36	1/36	0	0	0	0	7/36
5	0	0	0	0	2/36	2/36	2/36	2/36	1/36	0	0	9/36
6	0	0	0	0	0	2/36	2/36	2/36	2/36	2/36	1/36	11/36
$\sum$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	

a) Az $M(XY)$ várható értéke a (3) formula szerint számolva:

$$M(X,Y)=\sum_{i,j} x_i y_j w(x_i,y_j)=1\cdot2\frac{1}{36}+2\cdot3\frac{2}{36}+2\cdot4\frac{1}{36}+3\cdot4\frac{2}{36}+\ldots+6\cdot12\frac{1}{36}=34{,}22.$$

$$b)\quad F(x,y)=\begin{cases} 0, & \text{ha } x\le 1 \text{ vagy } y\le 2\\ \frac{1}{36}, & \text{ha } 1<x\le 2 \text{ és } 2<y\le 3\\ \frac{1}{36}+0=\frac{1}{36}, & \text{ha } 2<x\le 3 \text{ és } 2<y\le 3\\ \vdots & \vdots\\ \frac{1}{36}, & \text{ha } 1<x\le 2 \text{ és } 3<y\le 4\\ \frac{1}{36}+\frac{2}{36}=\frac{3}{36}, & \text{ha } 2<x\le 3 \text{ és } 3<y\le 4\\ \frac{1}{36}+\frac{2}{36}=\frac{3}{36} & \text{ha } 3<x\le 4 \text{ és } 3<y\le 4\\ \vdots & \vdots\\ \frac{1}{36}, & \text{ha } 1<x\le 2 \text{ és } 4<y\le 5\\ \frac{4}{36}, & \text{ha } 2<x\le 3 \text{ és } 4<y\le 5\\ \frac{6}{36}, & \text{ha } 3<x\le 4 \text{ és } 4<y\le 5\\ \vdots & \vdots\\ 1, & \text{ha } 6<x \text{ és } 12<y \end{cases}$$

c) Az x változó perem-eloszlásfüggvényét az X marginális eloszlásának összegezésével kapjuk:

$$F_m(x)=\begin{cases} 0, & \text{ha } x\le 1\\ \frac{1}{36}, & \text{ha } 1<x\le 2\\ \frac{4}{36}, & \text{ha } 2<x\le 3\\ \frac{9}{36}, & \text{ha } 3<x\le 4\\ \frac{16}{36}, & \text{ha } 4<x\le 5\\ \frac{25}{36}, & \text{ha } 5<x\le 6\\ 1, & \text{ha } 6<x \end{cases},\quad x\in R \text{ és } F_m(x)=P(X<x).$$

Az y változó perem-eloszlásfüggvényét az Y marginális eloszlásának összegezésével kapjuk:

$$F_m(y)=\begin{cases} 0, & \text{ha } y\le 2\\ \frac{1}{36}, & \text{ha } 2<y\le 3\\ \frac{3}{36}, & \text{ha } 3<y\le 4\\ \frac{6}{36}, & \text{ha } 4<y\le 5\\ \vdots & \vdots\\ 1, & \text{ha } 12<y \end{cases}\quad y\in R \text{ és } F_m(y)=P(Y<y)$$

d) $F(2{,}5;3{,}6)=\frac{3}{36}$; $F_x(2{,}5)=\frac{4}{36}$; $F_y(3{,}6)=\frac{3}{36}$.

3.1.5. A valószínűségi változók közötti kapcsolat szorossága

A valószínűségi változókról fontos tudnunk, hogy függetlenek-e vagy sem. Ha az X és Y független valószínűségi változók, és közülük az egyik ismert, annak alapján a másik változóra nem nyerünk új információt. Pl. abból, hogy ismerjük az X változó eloszlását, a tőle független Y változó eloszlásáról még semmit sem mondhatunk. A gyakorlati feladatokban előforduló valószínűségi változók között azonban gyakran van bizonyos összefüggés. Általában, ha az X és Y változók nem függetlenek, akkor az egyik változóról nyert információkból a másik változóra is szerezhetünk információkat. Pl. ha egy gépkocsi sebessége nő, akkor a fogyasztása is nő, vagy egy árucikk árának ugrásszerű növekedése, az árucikk keresletének csökkenésével jár. Az ilyen típusú változók között általában nem függvénykapcsolatról van szó, hiszen a valószínűségi változó értékeinek létrejöttét számos véletlenszerű hatás befolyásolja. A valószínűségi változók közötti kapcsolatot **sztochasztikus kapcsolatnak** nevezzük.

Az X és Y valószínűségi változók közötti összefüggés számszerű jellemzésére, a valószínűségi változók közötti sztochasztikus kapcsolat szorosságának mérésére bevezetjük a **kovariancia** és a **korrelációs együttható** fogalmát.

X és Y valószínűségi változók összege, szorzata, a változók és várható értékük különbsége és ezek szorzata, valamint várható értékük szintén valószínűségi változók. A valószínűségi változókra vonatkozó vizsgálatok azt mutatják, hogy ha az X és Y változók szoros kapcsolatban állnak, akkor a várható értékeiktől való eltéréseik együtt növekednek ill. csökkennek, és így a várható értékükkel képzett különbségek szorzata pozitív, azaz

$$(X - M(X))(Y - M(Y)) > 0$$

vagy, ha a várható értékeikhez képest ellenkező irányba változnak, akkor negatív, azaz

$$(X - M(X))(Y - M(Y)) < 0.$$

Ezek alapján arra következtethetünk, hogy az

$$(X - M(X))(Y - M(Y))$$

valószínűségi változó várható értéke az X és Y változók függőségének mérésére számszerű információt ad.

Definíció. Legyen X és Y diszkrét, véges valószínűségi változók együttes eloszlása w, a várható értékeket jelölje m_x és m_y, akkor az X és Y változók Cov(X,Y)-nal jelölt **kovarianciáján** a

$$\mathrm{Cov}(X,Y) = \sum_{i,j}(x_i - m_x)(y_j - m_y)w(x_i, y_j) = M\left((X - m_x)(Y - m_y)\right) \quad (4)$$

számot értjük.

Minthogy az X és Y diszkrét, véges valószínűségi változók, ezért a

$$\sum_{i,j} y_j w(x_i, y_j) = \sum_j y_j u(y_j) = m_y;$$

$$\sum_{i,j} x_i w(x_i, y_j) = \sum_i x_i v(x_i) = m_x$$

és a

$$\sum_{i,j} w(x_i, y_j) = 1$$

összefüggések felhasználásával a (4) képletet számításra alkalmasabb alakra hozhatjuk:

$$\mathrm{Cov}(X,Y) = \sum_{i,j}(x_i - m_x)(y_j - m_y)w(x_i, y_j) =$$

$$= \sum_{i,j} x_i y_j w(x_i, y_j) - m_x m_y = M(XY) - m_x m_y. \quad (5)$$

Ha X és Y független valószínűségi változók, akkor

$$w(x_i, y_j) = v(x_i) \cdot u(y_j),$$

és így

$$M(XY) = \sum_{i,j} x_i y_j w(x_i, y_j) = \sum_{i,j} x_i y_j v(x_i) \cdot u(y_j) =$$

$$= \sum_i x_i v(x_i) \cdot \sum_j y_j u(y_j) = M(X)M(Y).$$

Az $M(XY) = M(X)M(Y)$ -ból (5)-re tekintettel következik, hogy $\mathrm{Cov}(X,Y) = 0$, de ez fordítva nem áll fenn, azaz $\mathrm{Cov}(X,Y) = 0$-ból nem következik X és Y változók függetlensége. Mivel a kovariancia nagyságát jelentősen befolyásolja a vizsgálatba bevont (x_i, y_j) adatpárok mértékegysége, ezért ennek kiküszöbölésére az

$$\frac{X - m_x}{D(X)} \quad \text{és az} \quad \frac{Y - m_y}{D(Y)}$$

standardizált változók kovarianciáját képezzük, mely már dimenziómentes:

$$\mathrm{Cov}\left(\frac{X - m_x}{D(X)} \cdot \frac{Y - m_y}{D(Y)}\right) = \frac{\mathrm{Cov}(X,Y)}{D(X)D(Y)}.$$

Az összefüggés szorossága így függetleníthető az (x_i, y_j) adatpárok, a változók mértékegységétől. Ezek figyelembevételével kimondható a következő

Definíció. Ha az X és Y valószínűségi változók szórása és kovarianciája létezik, akkor az $\mathrm{R}(X,Y)$ -nal jelölt **korrelációján** az

$$\mathrm{R}(X,Y)=\frac{\mathrm{Cov}(X,Y)}{D(X)D(Y)}=\frac{M(XY)-m_x m_y}{D(X)D(Y)} \tag{6}$$

képlettel meghatározott számot értjük.

A korrelációt a valószínűségi változók közötti kapcsolat erősségének mérésére használjuk. Az $\mathrm{R}(X,Y)$ korreláció dimenziómentes és az alábbi tulajdonságok jellemzik:

a) $\mathrm{R}(X,Y)=\mathrm{R}(Y,X)$;

b) $-1\le \mathrm{R}(X,Y)\le 1$;

c) $\mathrm{R}(X,X)=1$; $\mathrm{R}(X,-X)=-1$;

d) $\mathrm{R}(aX+b, cY+d)=\mathrm{R}(X,Y)$, ha $a,c\neq 0$.

Összefoglaló megjegyzések

1. A kétváltozós együttes eloszlásokhoz tartozó változónkénti azonos peremeloszlások esetén is különbözhetnek a kovarianciák is és a korrelációk is.

2. A Ω eseménytéren értelmezett X és Y valószínűségi változók **függetlenek**, ha

$$P(X=x_i, Y=y_j)=P(X=x_i)\cdot P(Y=y_j) \tag{7}$$

$$(i=1,2,\dots,n;\ j=1,2,\dots,m)$$

s ebben az esetben a w együttes eloszlás minden tagja a megfelelő peremeloszlások szorzataként állítható elő, azaz

$$w(x_i, y_j)=v(x_i)\cdot u(y_j).$$

Ha X és Y független valószínűségi változók, akkor

a) $M(X+Y)=M(X)+M(Y)$;

b) $D^2(X+Y)=D^2(X)+D^2(Y)$; (8)

Feltesszük, hogy $M(X)$ és $M(Y)$ létezik, akkor (2)-re tekintettel

a) $$M(X+Y)=\sum_i\sum_j (x_i+y_j)w(x_i,y_j)=$$

$$=\sum_i\sum_j x_i\cdot w(x_i,y_j)+\sum_i\sum_j y_j\cdot w(x_i,y_j)=$$

$$=\sum_i x_i v(x_i)+\sum_j y_j u(y_j)=M(X)+M(Y).$$

b) Minthogy

$$D^2(X+Y)=\sum_{i,j}(x_i+y_j)^2 w(x_i,y_j)-\left(M(X+Y)\right)^2,$$

így négyzetre emelés, rendezés és *a)*-ra, valamint (2)-re tekintettel

$$D^2(X+Y)=\sum_i x_i^2 v(x_i)-m_x^2+\sum_j y_j^2 u(y_j)-m_y^2=$$

$$=D^2(X)+D^2(Y).$$

3. Ha az X és Y valószínűségi változók függetlenek, akkor $\mathrm{R}(X,Y)=0$. Ez az állítás fordítva nem áll fenn, vagyis nem következik $\mathrm{R}(X,Y)=0$ -ból, hogy X és Y függetlenek. Ilyenkor azt mondjuk, hogy a valószínűségi változók **korrelálatlanok**. Ha $\mathrm{R}(X,Y)=\pm 1$, akkor a valószínűségi változók között **lineáris függvénykapcsolat** van, azaz $Y=aX+b$, ahol a és b konstansok, és ez fordítva is igaz; ha viszont $-1<\mathrm{R}(X,Y)<1$, akkor az $\mathrm{R}(X,Y)$ értékéből nem lehet a valószínűségi változók közötti függvénykapcsolat milyenségére következtetni, pusztán a kapcsolat lineáris jellegének erősségét lehet becsülni. Ha $|\mathrm{R}(X,Y)|$ közel van az 1-hez, akkor azt mondjuk, hogy X és Y **erősen korreláltak.** A (6) formula alapján mondhatjuk, hogy az X és Y valószínűségi változók korrelálatlanok, ha

$$M(XY)=M(X)M(Y),$$

ill. (8) formulára tekintettel, ha

$$D^2(X+Y)=D^2(X)+D^2(Y).$$

Példa

1. Számítsuk ki az

$X \setminus Y$	4	10	$\sum$ sor
2	0	1/2	1/2
6	1/2	0	1/2
$\sum$ oszlop	1/2	1/2	

$U \setminus V$	4	10	$\sum$ sor
2	1/4	1/4	1/2
6	1/4	1/4	1/2
$\sum$ oszlop	1/2	1/2	

együttes eloszlásokkal adott X és Y, valamint U és V valószínűségi változók kovarianciáit és korrelációjukat.

Megoldás

A peremeloszlások alapján egyrészt látható, hogy az X és Y valószínűségi változók nem függetlenek, az együttes eloszlás tagjai nem egyenlők a megfelelő peremeloszlások szorzataival, azaz $w(x_i, y_j) \neq v(x_i) \cdot u(y_j)$, az U és V valószínűségi változók pedig függetlenek, azaz $w(x_i, y_j) = v(x_i) \cdot u(y_j)$, ($\frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2}$, stb.), másrészt az X és U valamint az Y és V eloszlásai azonosak.

X és U eloszlása:

x_i	2	6
$v(x_i)$	1/2	1/2

Y és V eloszlása:

y_j	4	10
$u(y_j)$	1/2	1/2

Az XY várható értéke:

$$M(XY) = 2 \cdot 4 \cdot 0 + 2 \cdot 10 \cdot \frac{1}{2} + 6 \cdot 4 \cdot \frac{1}{2} + 6 \cdot 10 \cdot 0 = 22.$$

Az UV várható értéke:

$$M(UV) = 2 \cdot 4 \cdot \frac{1}{4} + 2 \cdot 10 \cdot \frac{1}{4} + 6 \cdot 4 \cdot \frac{1}{4} + 6 \cdot 10 \cdot \frac{1}{4} = 28.$$

X és U várható értéke: $m_x = m_u = \frac{2+6}{2} = 4;$

Y és V várható értéke: $m_y = m_v = \frac{4+10}{2} = 7.$

X és Y, valamint U és V kovarianciája:

$$\mathrm{Cov}(X,Y) = M(XY) - m_x m_y = 22 - 4 \cdot 7 = -6;$$

$$\mathrm{Cov}(U,V) = M(UV) - m_u m_v = 28 - 28 = 0.$$

X és Y, valamint U és V szórásnégyzete és szórása:

$$D^2(X) = D^2(U) = M(X^2) - m_x^2 = 2^2 \cdot \frac{1}{2} + 6^2 \cdot \frac{1}{2} - 4^2 = 4.$$

$$D^2(Y) = D^2(V) = M(Y^2) - m_y^2 = 4^2 \cdot \frac{1}{2} + 10^2 \cdot \frac{1}{2} - 7^2 = 9.$$

tehát

$$D(X) = D(U) = \sqrt{4} = 2; \quad D(Y) = D(V) = \sqrt{9} = 3.$$

Az X és Y, valamint U és V korrelációja:

$$\mathrm{R}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{D(X)D(Y)} = \frac{-6}{6} = -1,$$

$$\mathrm{R}(U,V) = \frac{\mathrm{Cov}(U,V)}{D(U)D(V)} = \frac{0}{6} = 0.$$

Vegyük észre, hogy bár az X és U valamint az Y és V eloszlása azonos, de

$$\mathrm{Cov}(X,Y) \neq \mathrm{Cov}(U,V) \quad \text{és} \quad \mathrm{R}(X,Y) \neq \mathrm{R}(U,V).$$

Az X és Y között, mivel $\mathrm{R}(X,Y) = -1$, lineáris függvénykapcsolat van, éspedig $y = \frac{3}{2}x + 1$. Az U és V valószínűségi változók korrelálatlanok, mivel $\mathrm{R}(U,V) = 0$. (ld. a II. rész 4. fejezetét.)

2. Egy szabályos pénzérmét dobjunk fel négyszer. Az X valószínűségi változó értékei legyenek a négy dobásnál jelentkező fejek száma, az Y valószínűségi változó pedig jelölje a fejdobás sorozat hosszát. Számítsuk ki az X és Y változók együttes és peremeloszlását, kovarianciáját és korrelációját.

Megoldás

Négyszeri feldobás fej (F), írás (I) sorozat lehetséges eseteinek száma 16:

$$\begin{array}{cccccccc} FFFF & FFFI & \underline{\underline{FFIF}} & \underline{\underline{FFII}} & \underline{\underline{FIFF}} & \underline{FIFI} & \underline{FIIF} & \underline{FIII} \\ IFFF & \underline{\underline{IFFI}} & \underline{IFIF} & \underline{IFII} & \underline{\underline{IIFF}} & \underline{IIFI} & \underline{IIIF} & IIII \end{array} \qquad (*)$$

Az X valószínűségi változó (fej dobások száma) eloszlása: $v(0)=\frac{1}{16}$, mivel a 0 fejdobás eseménye egyszer következik be, $v(1)=\frac{4}{16}$, mivel 1 fejdobás négyszer következik be, stb.

Az X valószínűségeloszlásának táblázata:

x_i	0	1	2	3	4
$v(x_i)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Az Y valószínűségi változó (fej dobássorozat hossza) eloszlása: $u(0)=\frac{1}{16}$, $u(1)=\frac{7}{16}$, mivel a (*) 16 eseménye közül 1 hosszúságú F sorozat 7 esetben következik be (aláhúzottak), $u(2)=\frac{5}{16}$ (a kétszer aláhúzottak, számuk 5), stb.

Az Y valószínűségeloszlásának táblázata:

y_j	0	1	2	3	4
$u(y_j)$	$\frac{1}{16}$	$\frac{7}{16}$	$\frac{5}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

Az X és Y nem függetlenek, így a peremeloszlások szorzataként nem állíthatók elő az együttes eloszlás tagjai. A

$$w(x_i,y_j)=P(X=x_i,Y=y_j)\quad (i,j=0,1,2,3,4)$$

formulát alkalmazva kapjuk az együttes eloszlás táblázatát:

X \ Y	0	1	2	3	4	$\sum$ sor
0	$\frac{1}{16}$	0	0	0	0	$\frac{1}{16}$
1	0	$\frac{4}{16}$	0	0	0	$\frac{4}{16}$
2	0	$\frac{3}{16}$	$\frac{3}{16}$	0	0	$\frac{6}{16}$
3	0	0	$\frac{2}{16}$	$\frac{2}{16}$	0	$\frac{4}{16}$
4	0	0	0	0	$\frac{1}{16}$	$\frac{1}{16}$
$\sum$ oszlop	$\frac{1}{16}$	$\frac{7}{16}$	$\frac{5}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	

Az XY várható értéke valamint m_x és m_y várható értéke:

$$M(XY)=1\cdot1\cdot\frac{4}{16}+2\cdot1\cdot\frac{3}{16}+2\cdot2\cdot\frac{3}{16}+3\cdot2\cdot\frac{2}{16}+3\cdot3\cdot\frac{2}{16}+4\cdot4\cdot\frac{1}{16}=4{,}25,$$

$$m_x=0\cdot\frac{1}{16}+1\cdot\frac{4}{16}+2\cdot\frac{6}{16}+3\cdot\frac{4}{16}+4\cdot\frac{1}{16}=2$$

$$m_y=0\cdot\frac{1}{16}+1\cdot\frac{7}{16}+2\cdot\frac{5}{16}+3\cdot\frac{2}{16}+4\cdot\frac{1}{16}=1{,}69.$$

Az X és Y változók kovarianciája:

$$\mathrm{Cov}(X,Y)=M(XY)-m_x\cdot m_y=4{,}25-3{,}38=0{,}87.$$

Az X változó szórásnégyzete:

$$D^2(X)=M(X^2)-m_x^2=1^2\cdot\frac{4}{16}+2^2\cdot\frac{6}{16}+3^2\cdot\frac{4}{16}+4^2\cdot\frac{1}{16}-2^2=1.$$

Az X változó szórása:

$$D(X)=1.$$

Az Y változó szórásnégyzete:

$$D^2(Y)=M(Y^2)-m_y^2=1^2\cdot\frac{7}{16}+2^2\cdot\frac{5}{16}+3^2\cdot\frac{2}{16}+4^2\cdot\frac{1}{16}-2{,}86=0{,}96.$$

Az Y változó szórása:

$$D(Y)=0{,}98.$$

Az X és Y változó korrelációja:

$$\mathrm{R}(X,Y)=\frac{\mathrm{Cov}(X,Y)}{D(X)D(Y)}=\frac{0{,}87}{1\cdot0{,}98}=0{,}8878.$$

Az $\mathrm{R}(X,Y)$ közel van az 1-hez, ezért mondhatjuk, hogy X és Y erősen korreláltak.

3.2. Folytonos valószínűségi változó

A diszkrét valószínűségi változók, mint azt az előző pontokban láttuk, értékeiket egy véges vagy egy megszámlálhatóan végtelen számhalmazból vehetik fel. Számos olyan véletlen jelenség is van, amelynél a változó által felvehető értékek egy zárt $[a,b]$ intervallumba esnek. Pl. egy villanyégő 0 és 1000 óra között (a $[0;1000]$ intervallumban) bármelyik pillanatban meghibásodhat (kiéghet), vagy a vízállást jelző mérőléc 0 m és 10 m jelző pontjai között a Duna vízállása bárhol lehet.

A valószínűségszámításban az ilyen problémák tárgyalására bevezették az ún. **folytonos valószínűségi változó** fogalmát. Ez azt jelenti, hogy definiálunk egy olyan X valószínűségi változót, amely „folytonosan" vehet fel értékeket az R valós számhalmaz valamely részhalmazából. A folytonos valószínűségi változó definícióját legegyszerűbben az ún. *eloszlásfüggvény* segítségével adhatjuk meg.

3.2.1. Eloszlásfüggvény

Tegyük fel, hogy minden adott $x \in R$ értéknél ki tudjuk számítani a $P(X < x)$ valószínűséget. Így minden x valós számhoz hozzárendelünk egy $P(X < x)$ valós számot, azaz egy valós-valós függvényt definiálhatunk. Jelöljük ezt a függvényt $F(x)$ -szel, és legyen

$$F(x) = P(X < x), \quad \forall x \in R \qquad (1)$$

Definíció. Az (1) formulával adott $F(x)$ függvényt az X valószínűségi változó **eloszlásfüggvényének** nevezzük.

Az értelmezés szerint tehát az F függvény értéke az x pontban annak a valószínűsége, hogy a X változó értékei kisebbek mint x.

Az eloszlásfüggvény értelmezéséből következik, hogy $F(x)$ -re teljesülnek az alábbi tulajdonságok:

1. $F(x) \geq 0$;

2. monoton növekedő, azaz $\forall a, b$ esetén, ha $a < b$, akkor $F(a) \leq F(b)$;

3. $\lim\limits_{x \to -\infty} F(x) = 0$ és $\lim\limits_{x \to +\infty} F(x) = 1$;

4. $F(x)$ minden pontban balról folytonos, azaz $\forall c \in R$ esetén $\lim\limits_{x \to c-0} F(x) = F(c)$.

Ha egy függvény eleget tesz az 1.–4. pontokba foglaltaknak, akkor valamely folytonos valószínűségi változó eloszlásfüggvényének tekinthető. Figyelembe véve az előzőeket definiálhatjuk a *folytonos valószínűségi változót.*

Definíció. Ha az X valószínűségi változó $F(x)$ eloszlásfüggvénye folytonos, akkor az X *változót és az eloszlását* is **folytonosnak** nevezzük.

Példa

Tegyük fel, hogy egy villanyégő véletlenszerű meghibásodási időpontja 0 és 1000 óra között van (vagyis legfeljebb 1000 óráig jó). Határozzuk meg a jelenség eloszlásfüggvényét.

Megoldás

Ha a $[0;1000]$ időintervallumot egy szakasznak tekintjük, akkor egy geometriai valószínűségi feladatra vezethetjük vissza a problémát, mert annak valószínűsége, hogy a $[0;x]$ szakaszon ég ki az égő, arányos a szakasz hosszával. Jelentse az X valószínűségi változó a meghibásodás időpontját, azaz X a $[0;1000]$ intervallumból veheti fel értékeit. Nyilvánvaló, hogy

$$P(X < 0) = 0,$$

mert $X < 0$ lehetetlen esemény (a meghibásodás időpontja negatív nem lehet) Az is világos, ha $x > 1000$, akkor

$$P(X < x) = 1,$$

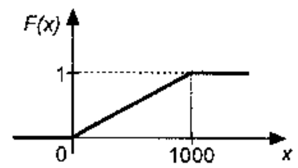
mert kimondtuk, hogy az égő 1000 órán túl biztosan nem üzemelhet. Ha viszont $0 < x \leq 1000$, akkor

$$P(X < x) = \frac{x}{1000},$$

mert feltettük, hogy a meghibásodás valószínűsége arányos a szakasz hosszával. Tehát

$$F(x) = P(X < x) = \begin{cases} 0, & \text{ha } x \leq 0 \\ \frac{x}{1000}, & \text{ha } 0 < x \leq 1000 \\ 1, & \text{ha } x > 1000 \end{cases} \qquad (*)$$

A (*) formulával adott $F(x)$ eloszlásfüggvény gráfját a 14. ábra szemlélteti.



14. ábra. F(x) eloszlásfüggvény

Az eloszlásfüggvény ismeretében ki tudjuk számítani annak valószínűségét is, hogy az X valószínűségi változó értékei egy adott $[a;b]$ intervallumba essenek. Érvényes a következő tétel:

Tétel

Ha F az X valószínűségi változó eloszlásfüggvénye, akkor

$$P(a \le x < b) = F(b) - F(a). \qquad (2)$$

Ui. ha az A, B, C eseményeket a következő módon definiáljuk: A legyen $X < a$; B legyen $X < b$ és C legyen $a \le x < b$, akkor mivel $C = B - A$ és $A \subset B$, ezért

$$P(a \le x < b) = P(C) = P(B - A) = P(B) - P(A) =$$

$$= P(X < b) - P(X < a) = F(b) - F(a).$$

Példa

A tétel felhasználásával számítsuk ki a villanyégőre vonatkozólag annak valószínűségét, hogy a meghibásodás a 30. és a 100. óra között következik be.

Megoldás

$$P(30 \le x < 100) = F(100) - F(30) = \frac{100}{1000} - \frac{30}{1000} = \frac{70}{1000} = 0{,}07.$$

Megjegyzés

A diszkrét valószínűségi változó eloszlásfüggvénye az x-tengellyel párhuzamos szakaszokból áll, ún. **lépcsős függvény**. Az egyes intervallum szakaszokhoz a lépcsős függvény értékeit az alábbi módon számítjuk ki:

$F(x) = 0$, ha $x \le x_0$

$F(x) = p_0$, ha $x_0 < x \le x_1$

$F(x) = p_0 + p_1$, ha $x_1 < x \le x_2$

$F(x) = p_0 + p_1 + p_2$, ha $x_2 < x \le x_3$

$\vdots$

$F(x) = p_0 + p_1 + \ldots + p_k$, ha $x_k < x \le x_{k+1}$

$\vdots$

$F(x) = 1$, ha $x_{n-1} < x$, feltéve, hogy az X változó n számú értéket vesz fel.

Példák

1. Legyen az X diszkrét valószínűségi változó eloszlási táblázata:

x_k	-3	2	6
p_k	1/4	1/2	1/4

Ábrázoljuk az $F(x)$ eloszlásfüggvényt.

Megoldás

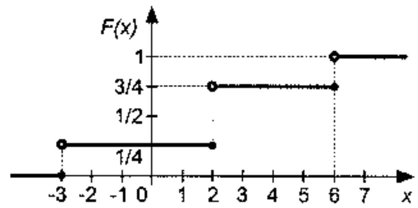
Ha $x \le -3$, akkor $F(x) = 0$,

$-3 < x \le 2$, akkor $F(x) = \frac{1}{4}$,

$2 < x \le 6$, akkor $F(x) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$,

$6 < x$ akkor $F(x) = \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1$.

A megoldást a 15. ábra szemlélteti.



15. ábra. F(x) eloszlásfüggvény (lépcsős függvény)

2. Az adott függvények közül melyik lehet eloszlásfüggvény?

$$a)\ F(x) = \begin{cases} 0, & \text{ha } x < 0 \\ \dfrac{x-4}{x+4}, & \text{ha } 0 \le x; \end{cases} \qquad b)\ F(x) = \begin{cases} e^x, & \text{ha } x \le 0 \\ 1, & \text{ha } 0 < x. \end{cases}$$

Megoldás

a) $F(x)$ nem lehet eloszlásfüggvény, mert $(0,4)$ intervallumban negatív;

b) $F(x)$ valóban lehet eloszlásfüggvény, mivel minden $x \in R$-re értelmezve van, monoton növekvő, minden pontban folytonos és

$$\lim_{x_n \to -\infty} F(x_n) = \lim_{x_n \to -\infty} 0 = 0, \quad \lim_{x_n \to \infty} F(x_n) = \lim_{x_n \to \infty} 1 = 1.$$

3. Dobjunk fel 3 érmét. Az X valószínűségi változó legyen a dobásonkénti fejek száma. Adjuk meg az X változó

a) valószínűségeloszlását és gráfját,

b) eloszlásfüggvényét és gráfját,

c) $F(-1)$, $F(1{,}8)$, $F(2)$ értékeket, valamint az eloszlásfüggvény felhasználásával annak valószínűségét, hogy legalább 1 fejet, ill. legfeljebb 1 fejet dobunk.

Megoldás

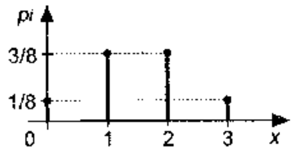
Tegyük fel, hogy az

$$FFF,\ FFI,\ FIF,\ IFF,\ IIF,\ IFI,\ FII,\ III$$

elemi események egyenlő valószínűségűek. Az X diszkrét valószínűségi változó értékei: 0, 1, 2, 3.

a) A fej dobások valószínűsége (16. ábra):

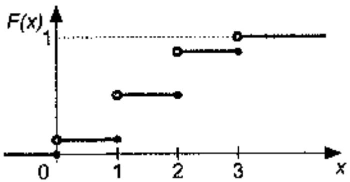
$$P(X=0)=\frac{1}{8},\quad P(X=1)=\frac{3}{8},\quad P(X=2)=\frac{3}{8},\quad P(X=3)=\frac{1}{8}.$$



16. ábra. X valószínűségeloszlása

b) Az eloszlásfüggvény:

$$F(x)=\begin{cases} 0, & \text{ha } x\le 0 \\ \frac{1}{8}, & \text{ha } 0<x\le 1 \\ \frac{4}{8}, & \text{ha } 1<x\le 2; \\ \frac{7}{8}, & \text{ha } 2<x\le 3 \\ 1, & \text{ha } 3<x \end{cases}\qquad x\in R. \quad \text{(17. ábra)}$$



17. ábra. F(x) eloszlásfüggvény

c) $F(-1)=P(X<-1)=0,\quad F(1{,}8)=P(X<1{,}8)=\frac{4}{8}=\frac{1}{2},$

$F(3)=P(X<3)=\frac{7}{8}.$

d) Legalább egy fej dobásának valószínűsége:

$$P(X\ge 1)=1-P(X<1)=1-F(1)=1-\frac{1}{8}=\frac{7}{8}.$$

Legfeljebb 1 fej dobásának valószínűsége:

$$P(X\le 1)=P(X<2)=F(2)=\frac{1}{2}.$$

3.2.2. A sűrűségfüggvény

Definíció. Ha az X valószínűségi változó $F(x)$ eloszlásfüggvénye folytonos, és véges számú pont kivételével létezik az $F'(x)$, akkor a deriváltfüggvényt az X **sűrűségfüggvényének** nevezzük és $f(x)$-szel jelöljük, azaz

$$F'(x)=f(x). \tag{1}$$

Más megfogalmazásban:

Ha az X valószínűségi változó $F(x)$ eloszlásfüggvénye előállítható

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

alakban, ahol $f(t)$ integrálható függvény, akkor az $f(t)$ függvényt az X valószínűségi változó, ill. az $F(x)$ eloszlásfüggvény **sűrűségfüggvényének** nevezzük. Az eloszlásfüggvény tehát a sűrűségfüggvény integrálfüggvénye.

Tétel

Az f sűrűségfüggvény az alábbi tulajdonságokkal rendelkezik:

1. $f(x)\ge 0$, azaz nem lehet negatív;

2. $\int_{-\infty}^{\infty} f(x)\,dx=1$, azaz improprius integrálja 1;

3. $\int_a^b f(x)dx = P(a \le X \le b)$, azaz tetszőleges $[a,b]$-beli integrálja az X változó $[a,b]$-be esésének valószínűségével egyenlő.

Bizonyítás

1. Mivel $F(x)$ monoton nem csökkenő, ezért nem lehet a deriváltja negatív.

2. $\int_{-\infty}^{\infty} f(x)dx = \lim_{R\to\infty} \int_{-R}^{R} f(x)dx = \lim_{R\to\infty} [F(R) - F(-R)] = 1 - 0 = 1,$

az $F(x)$ eloszlásfüggvény 3. tulajdonsága értelmében.

3. A *Newton-Leibniz* tétel értelmében

$$\int_a^b f(x)dx = F(b) - F(a) = P(a \le X < b) =$$

$$= P(a < X \le b) = P(a \le X \le b).$$

Megjegyzés

1. Ha egy f folytonos függvény rendelkezik az 1.–2. tulajdonságokkal, akkor egy folytonos valószínűségi változó sűrűségfüggvényének tekinthető.

2. Annak valószínűsége, hogy az X változó az $[a,b]$ intervallumba esik – a 3. tulajdonság alapján – az $f(x)$ sűrűségfüggvény gráfja alatti $[a,b]$ intervallumhoz tartozó terület mérőszámával egyenlő.

A folytonos valószínűségeloszlásokat gyakran nem az eloszlásfüggvénnyel, hanem a sűrűségfüggvénnyel adjuk meg, és az *eloszlásfüggvényt a sűrűségfüggvényből számítjuk ki* az

$$F(x) = \int_{-\infty}^{x} f(t)dt \qquad (2)$$

integrállal.

Pl. a 3.2.1. pont példájában szereplő X valószínűségi változó sűrűségfüggvénye az

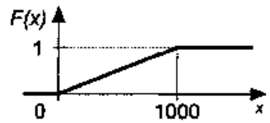
$$F(x) = P(X < x) = \begin{cases} 0, & \text{ha } x \le 0 \\ \dfrac{x}{1000}, & \text{ha } 0 < x \le 1000 \\ 1, & \text{ha } x > 1000 \end{cases} \qquad \text{(18. a) ábra)}$$

eloszlásfüggvényének deriváltja:

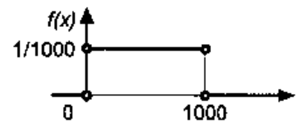
$$F'(x) = f(x) = \begin{cases} 0, & \text{ha } x < 0 \\ \dfrac{1}{1000}, & \text{ha } 0 < x < 1000 \\ 0, & \text{ha } x > 1000 \end{cases}$$

melynek gráfját a 18. b) ábra szemlélteti. $f(x) \ge 0$ az értelmezési tartománybeli minden x-re és

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{1000} \frac{1}{1000}dx + \int_{1000}^{\infty} 0dx = \left[\frac{x}{1000}\right]_0^{1000} = \frac{1000}{1000} = 1.$$



18. a) ábra. F(x) eloszlásfüggvény



18. b) ábra. f(x) sűrűségfüggvény

Példák

1. Az adott függvények közül melyik lehet sűrűségfüggvény?

a) $f(x) = \begin{cases} \dfrac{1}{5}, & \text{ha } 0 \le x \le 10, \\ 0 & \text{egyébként;} \end{cases}$

b) $f(x) = \begin{cases} \dfrac{1}{8}x, & \text{ha } 0 \le x \le 4, \\ 0 & \text{egyébként.} \end{cases}$

Megoldás

a) Mivel $\int_{-\infty}^{+\infty} f(x)dx = \int_0^{10} \frac{1}{5}dx = \left[\frac{x}{5}\right]_0^{10} = \frac{10}{5} - 0 = 2 \ne 1$, ezért ez az f függvény nem lehet egy folytonos valószínűségi változó sűrűségfüggvénye.

b) Mivel $\int_{-\infty}^{+\infty} f(x)dx = \int_0^4 \frac{1}{8}xdx = \left[\frac{x^2}{16}\right]_0^4 = \frac{16}{16} - 0 = 1$, és $f(x) \geq 0$, ezért ez a függvény lehet egy folytonos valószínűségi változó sűrűségfüggvénye, melynek eloszlásfüggvénye:

$$F(x) = \begin{cases} 0, & \text{ha } x < 0, \\ \frac{1}{16}x^2, & \text{ha } 0 \leq x \leq 4, \\ 1, & \text{ha } x > 4. \end{cases}$$

2. Milyen c érték mellett lesz az $f(x) = \begin{cases} cx^2, & \text{ha } 0 \leq x \leq 3, \\ 0 & \text{egyébként,} \end{cases}$ függvény valamely folytonos valószínűségi változó sűrűségfüggvénye? A c ismeretében számítsuk ki $P(1 \leq X \leq 2)$ értékét.

Megoldás

A c értékét

$$\int_{-\infty}^{+\infty} f(x)dx = \int_0^3 cx^2 dx = \left[\frac{cx^3}{3}\right]_0^3 = \frac{27}{3}c - 0 = 1$$

egyenletből számíthatjuk ki, azaz $c = \frac{3}{27} = \frac{1}{9}$, tehát

$$f(x) = \begin{cases} \frac{1}{9}x^2, & \text{ha } 0 \leq x \leq 3, \\ 0 & \text{egyébként,} \end{cases} \quad \text{és} \quad F(x) = \begin{cases} 0, & \text{ha } x < 0, \\ \frac{1}{27}x^3, & \text{ha } 0 \leq x \leq 3, \\ 1, & \text{ha } x > 3. \end{cases}$$

Az f, ill. a F függvény felhasználásával:

$$P(1 \leq X \leq 2) = \int_1^2 \frac{1}{9}x^2 dx = F(2) - F(1) = \frac{8}{27} - \frac{1}{27} = \frac{7}{27}.$$

3.2.3. Várható érték, szórásnégyzet és szórás

Definíció. Ha az X folytonos eloszlású valószínűségi változónak $f(x)$ a sűrűségfüggvénye, akkor **várható értéke**:

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx \tag{1}$$

feltéve, hogy az improprius integrál abszolút konvergens, azaz

$$\int_{-\infty}^{\infty} |x| f(x)dx < +\infty.$$

Tehát az (1) formula csak akkor áll fenn, ha a jobb oldali integrál létezik és véges.

Definíció. Ha az X folytonos eloszlású valószínűségi változónak $f(x)$ a sűrűségfüggvénye, akkor **szórásnégyzetét** és **szórását** a

$$D^2(X) = M((X-m)^2) = \int_{-\infty}^{\infty} (x-m)^2 f(x)dx \tag{2}$$

$$D(X) = \sqrt{\int_{-\infty}^{\infty} (x-m)^2 f(x)dx} \tag{3}$$

képletekkel definiáljuk, ahol $m = M(X)$, feltéve hogy a jobb oldali integrál létezik.

A (2) formula alapján azt is mondhatjuk, hogy a szórásnégyzet az $(X - M(X))^2$ valószínűségi változó várható értéke, ha létezik.

A diszkrét esethez hasonlóan az $m = M(X)$ és $M(X^2)$ létezése szükséges és elegendő feltétel a $D^2(X)$ létezéséhez, és így a *szórásnégyzetet* és a *szórást* a következő képletekkel is számíthatjuk:

$$D^2(X) = M(X^2) - m^2 = \int_R x^2 f(x)dx - m^2 \tag{4}$$

$$D(X) = \sqrt{M(X^2) - m^2} = \sqrt{\int_R x^2 f(x)dx - m^2}. \tag{5}$$

Megjegyzés

Mindazok az állítások, amelyeket a 3.1.3. pont megjegyzésében felsoroltunk a diszkrét valószínűségi változó várható értékére, szórására és szórásnégyzetére, a folytonos valószínűségi változókra is érvényesek.

Pl. Ha $X_1, X_2, \ldots, X_n$ valószínűségi változók mindegyikének létezik várható értéke, akkor *összegüknek* is létezik várható értéke, és az összeg várható értéke az egyes változók várható értékének összegével egyenlő, azaz

$$M(X_1+X_2+\ldots+X_n)=M(X_1)+M(X_2)+\ldots+M(X_n) \text{ stb.}$$

Példák

1. Legyen az X folytonos valószínűségi változó sűrűségfüggvénye:

$$f(x)=\begin{cases}\frac{3}{8}x^2, & \text{ha } 0\le x\le 2\\ 0, & \text{ha } x<0, \text{ ill. ha } x>2\end{cases}$$

Határozzuk meg és ábrázoljuk az $F(x)$ eloszlásfüggvényt.

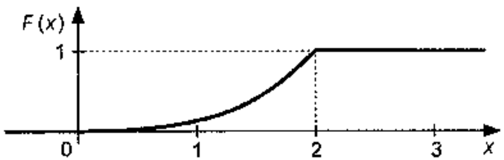
Megoldás

A $0\le x\le 2$ intervallumban:

$$F(x)=\int_0^x \frac{3}{8}t^2dt=\frac{3}{8}\left[\frac{t^3}{3}\right]_0^x=\frac{3}{8}\cdot\frac{1}{3}x^3=\frac{1}{8}x^3,$$

így
$$F(x)=\begin{cases}0, & \text{ha } x<0\\ \frac{1}{8}x^3, & \text{ha } 0\le x\le 2\\ 1, & \text{ha } x>2\end{cases}$$

Az $F(x)$ gráfját a 19. ábra szemlélteti.

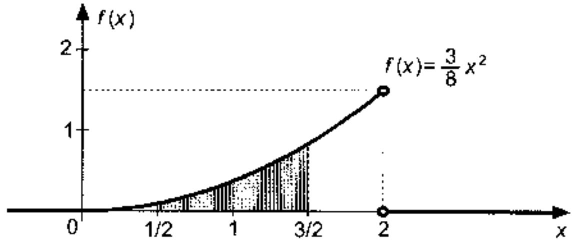


19. ábra. F(x) eloszlásfüggvény

2. Legyen az X folytonos valószínűségi változó sűrűségfüggvénye:

$$f(x)=\begin{cases}\frac{3}{8}x^2, & \text{ha } 0\le x\le 2\\ 0, & \text{ha } x<0 \text{ ill. ha } x>2\end{cases} \qquad \text{(20. ábra)}$$

Számítsuk ki a $P(\frac{1}{2}\le X\le\frac{3}{2})$ valószínűséget, valamint X várható értékét, szórásnégyzetét és szórását.



20. ábra. f(x) sűrűségfüggvény

Megoldás

Annak valószínűsége, hogy X az $\left[\frac{1}{2},\frac{3}{2}\right]$ intervallumba esik:

$$P(\tfrac{1}{2}\le X\le\tfrac{3}{2})=\int_a^b f(x)dx=\int_{\frac{1}{2}}^{\frac{3}{2}}\frac{3}{8}x^2dx=\frac{3}{8}\left[\frac{x^3}{3}\right]_{\frac{1}{2}}^{\frac{3}{2}}=\frac{3}{8}\cdot\frac{1}{3}\left(\frac{27}{8}-\frac{1}{8}\right)=\frac{3}{24}\cdot\frac{26}{28}=\frac{13}{32}.$$

A várható érték:

$$m=M(X)=\int_R xf(x)dx=\int_0^2\frac{3}{8}x^3dx=\frac{3}{8}\left[\frac{x^4}{4}\right]_0^2=\frac{3}{8}\cdot\frac{16}{4}=\frac{3}{2}.$$

$$M(X^2)=\int_R x^2f(x)dx=\int_0^2\frac{3}{8}x^4dx=\frac{3}{8}\left[\frac{x^5}{5}\right]_0^2=\frac{3}{8}\cdot\frac{32}{5}=\frac{12}{5}.$$

Tehát a szórásnégyzet:

$$D^2(X)=M(X^2)-m^2=\frac{12}{5}-\left(\frac{3}{2}\right)^2=\frac{12}{5}-\frac{9}{4}=\frac{3}{20};$$

és a szórás:

$$D(X)=\sqrt{\frac{3}{20}}=\frac{1}{2}\sqrt{\frac{3}{5}}\approx\frac{1}{2}\cdot 0{,}7746=0{,}3873.$$

3.3. A valószínűségi változó egyéb jellemzői

A valószínűségi változó és eloszlása jellemzésére a várható érték, a szórás és szórásnégyzet mellett további információt nyerhetünk az ún. *momentumok* előállításával. A következő pontokban megvizsgáljuk, hogy a momentumokkal miként állíthatók elő az eloszlásokat jellemző várható eltérés, medián, kvantilisek, terjedelem, módusz, ferdeség és lapultság értékei.

3.3.1. A momentumok és alkalmazásuk

Az X valószínűségi változó várható értékét (diszkrét és folytonos esetnek megfelelően)

$$M(X)=\sum_i x_i v(x_i), \text{ ill. } M(X)=\int_{-\infty}^{\infty} xf(x)dx$$

formulákkal, a szórásnégyzetét pedig

$$D^2(X)=\sum_i (x_i-M(X))^2 v(x_i)=M(X^2)-(M(X))^2,$$

ill.

$$D^2(X)=\int_{-\infty}^{\infty}(x-M(X))^2 f(x)dx=\int_{-\infty}^{\infty}x^2 f(x)dx-(M(X))^2$$

formulákkal adtuk meg. A formulákban X és X^2 várható értéke szerepel. A valószínűségi változó magasabb fokú várható értéke is felhasználható eloszlásának és sűrűségfüggvényének jellemzésére.

Definíció. Legyen X valószínűségi változó. Az X^k változó várható értékét ***k*-adik vagy *k*-adrendű momentumának** nevezzük és M_k-val jelöljük, azaz

$$M_k=M(X^k)=\sum_i x_i^k v(x_i) \quad (k=1,2,\ldots) \tag{1}$$

ill.

$$M_k=M(X^k)=\int_{-\infty}^{\infty}x^k f(x) \quad (k=1,2,\ldots) \tag{2}$$

ha a várható értékek léteznek. Az $M(X)$ várható értéket, tehát elsőrendű momentumnak is nevezhetjük.

Az X változó k-adik momentuma mellett azt is vizsgálhatjuk, hogy X egy adott r értéktől (ponttól) való eltérésének k-adik momentuma mivel egyenlő. Ha ez az r érték éppen $M(X)$ értékével egyenlő, akkor a várható értékre vonatkozó ún. ***k*-adik centrális momentumról** beszélünk, melyet μ_k-val jelölünk:

$$\mu_k=M\left((X-M(X))^k\right) \quad (k=0,1,2,\ldots) \tag{3}$$

Az első centrális momentum 0, ui.

$$M(X-M(X))=M(X)-M(M(X))=M(X)-M(X)=0,$$

mivel $M(X)$ konstans, így várható értéke önmagával egyenlő.

A második centrális momentum pedig a szórásnégyzettel egyenlő, azaz

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

Ebből látható, hogy az első és második momentum ismeretében a második centrális momentum, azaz a szórásnégyzet is előállítható.

Az $M(X)$ várható értéket az **eloszlás centrumának** is nevezzük.

Definíció. Az

$$M\left(|X|^k\right) \tag{4}$$

várható értéket, ha létezik ***k*-adik abszolút momentumnak**, az

$$M\left(|X-M(X)|^k\right) \tag{5}$$

várható értéket, ha létezik ***k*-adik centrális abszolút momentumnak** nevezzük.

Ha a diszkrét valószínűségi változó x_i-hez tartozó p_i valószínűségei, ill. folytonos valószínűségi változó sűrűségfüggvénye szimmetrikusak a várható értékre, akkor a páratlanrendű centrális momentumok 0-val egyenlők.

Ha az eloszlás, ill. a sűrűségfüggvény nem szimmetrikus, akkor annak mérésére egy ún. *ferdeségi együtthatót* számíthatunk ki, ha ismert a változó szórása, és harmadik centrális momentuma.

Definíció. Az X valószínűségi változó γ_1-gyel jelölt harmadik centrális momentumának és szórása harmadik hatványának hányadosát **ferdeségi együtthatónak** nevezzük, azaz

$$\gamma_1 = \frac{M\left((X - M(X))^3\right)}{D^3(X)} = \frac{\mu_3}{\sigma^3}. \tag{6}$$

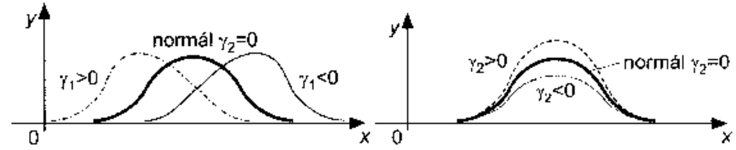
Egycsúcsú eloszlás esetén a valószínűségi változónak csak egy módusza van (l. 3.3.3. pontot), ekkor a folytonos valószínűségi változó sűrűségfüggvényének gráfja negatív γ_1 esetén a módusztól balra, pozitív γ_1 esetén pedig a módusztól jobbra elnyúlik. Ha $\gamma_1 = 0$, akkor a gráf szimmetrikus.

Az X valószínűségi változó sűrűségfüggvényének jellemzésére a vele azonos várható értékű és szórású normális eloszlású valószínűségi változó sűrűségfüggvényével való összehasonlításból is következtetést vonhatunk le.

Definíció. Az X valószínűségi változó γ_2-vel jelölt negyedik centrális momentuma és szórása negyedik hatványa hányadosának 3-mal csökkentett számértékét **lapultsági együtthatónak** nevezzük, azaz

$$\gamma_2 = \frac{M\left((X - M(X))^4\right)}{D^4(X)} - 3 = \frac{\mu_4}{\sigma^4} - 3. \tag{7}$$

A leggyakrabban előforduló normális eloszlás (l. a 4.2.3. pontot) lapultsági együtthatója 0, így mondhatjuk, hogy pozitív γ_2 esetén az X változó sűrűségfüggvénye rendszerint magasabb és csúcsosabb, mint az összehasonlításul választott, megegyező várható értékű és szórású normális eloszlású változó sűrűségfüggvényéé (21. a) és b) ábra).



21. a) és b) ábra. Lapultság és ferdeség a normáleloszláshoz képest

Az (5) formulát $k = 1$ esetén a valószínűségi változó várható értéke körüli véletlen ingadozásának mérőszámaként használhatjuk a szórásnégyzet helyett, pontosabb értéket ad, de használata az abszolút érték miatt nehézkesebb.

Definíció. Az X valószínűségi változó m_l-lel jelölt első centrális abszolút momentumát az X **várható eltérésének** nevezzük, ha várható értéke létezik, azaz

$$m_l = M\left(|X - M(X)|\right). \tag{8}$$

Példa

Számítsuk ki a ferdeségi és a lapultsági együtthatót, ha

a) $f(x) = \begin{cases} \frac{1}{2}\cos x, & \text{ha } -\frac{\pi}{2} \le x \le \frac{\pi}{2} \\ 0 & \text{egyébként;} \end{cases}$

b) $f(x) = \begin{cases} \frac{4}{\pi}\frac{1}{1+x^2}, & \text{ha } 0 \le x \le 1 \\ 0 & \text{egyébként.} \end{cases}$

Megoldás

a) Mivel az $f(x)$ sűrűségfüggvény szimmetrikus az y tengelyre, ezért a páratlanrendű centrális momentumok 0-val egyenlők, azaz $M(X) = 0$, és $\mu_3 = 0$. A szórásnégyzet:

$$\mu_2 = D^2(X) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} x^2 \cdot \frac{1}{2}\cos x\,dx = \frac{1}{4}\pi^2 - 2 = 0{,}467401101,$$

amelyből a szórás: $\sigma = D(X) = \sqrt{0{,}467401101} = 0{,}6836673906,$

$$\sigma^3 = 0{,}3195468911, \quad \sigma^4 = 0{,}2184637892,$$

$$\mu_4 = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} x^4 \cdot \frac{1}{2}\cos x dx = \frac{1}{16}\pi^4 + 24 - 3\pi^2 = 0{,}47925498,$$

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{0}{0{,}3195468911} = 0,$$ minthogy a sűrűségfüggvény szimmetrikus.

$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{0{,}47925498}{0{,}2184637892} - 3 = -0{,}806249806.$ A negatív γ_2-ből arra következtethetünk, hogy a megegyező várható értékű és szórású normális eloszlású változó sűrűségfüggvényéhez képest a vizsgált változó sűrűségfüggvénye laposabb.

b) A várható érték:

$$m = M(X) = \int_0^1 x \cdot \frac{4}{\pi}\frac{1}{1+x^2}dx = 2\frac{\ln 2}{\pi} = 0{,}4412712002,$$

szórásnégyzet:

$$D^2(X) = \int_0^1 (x-m)^2 \cdot \frac{4}{\pi}\frac{1}{1+x^2}dx = 0{,}07851926940,$$

szórás:

$$\sigma = \sqrt{0{,}07851926940} = 0{,}2802129001,$$

és így

$$\sigma^3 = 0{,}02200211219, \quad \sigma^4 = 0{,}006165275664.$$

$$\mu_3 = \int_0^1 (x-m)^3 \frac{4}{\pi}\frac{1}{1+x^2}dx = 0{,}02200211219,$$

$$\mu_4 = \int_0^1 (x-m)^4 \frac{4}{\pi}\frac{1}{1+x^2}dx = 0{,}01185056671.$$

A kiszámított értékekkel:

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = 0{,}2490325103 \text{ és } \gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = -1{,}077852905.$$

A γ_1 ferdeségi együttható pozitív, γ_2 lapultsági együttható negatív, tehát összehasonlítva a megfelelő normális eloszlású változó sűrűségfüggvényével, arra következtethetünk, hogy a vizsgált változó sűrűségfüggvénye a módusztól jobbra elnyúlik, és laposabb.

3.3.2. A medián

Definíció. Azt az r számot, amelyre az X valószínűségi változó $M(|X-r|)$ elsőrendű abszolút momentuma felveszi minimumát, ha ez létezik, **mediánnak** (az *eloszlás közepének*) nevezzük és m_e-vel jelöljük. Ez az m_e szám a valószínűségi változó értékeit két olyan részre osztja, amelyek bekövetkezési valószínűsége ugyanaz.

A definícióból és a várható érték fogalmából következik, hogy a medián bizonyos esetekben egyenlő is lehet a várható értékkel, de általában attól különbözik.

A mediánt szokás az X valószínűségi változó $F(x)$ eloszlásfüggvényével is definiálni. Igazolható, hogy az

$$F(x) = \frac{1}{2} \tag{1}$$

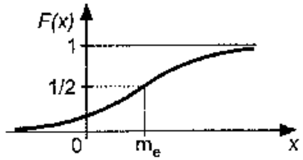
egyenletnek, ha egyetlen x_0 megoldása létezik, akkor az az eloszlás mediánjával egyenlő, azaz $x_0 = m_e$. Ha nincs egyetlen megoldása az (1) egyenletnek, de valamely $]a,b]$ intervellumbeli értékekre $F(x) = \frac{1}{2}$, akkor ennek az intervallumnak a középpontját nevezzük mediánnak, azaz

$$m_e = \frac{a+b}{2}. \tag{2}$$

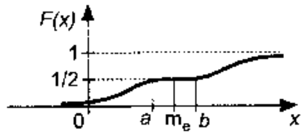
Ha az egyenletnek nincs megoldása, akkor azoknak az x értékeknek a felső határát nevezzük mediánnak, amelyekre $F(x) < \frac{1}{2}$.

A *medián geometriai szemléltetésére* ábrázoljuk az $F(x)$ eloszlásfüggvényt és az x tengellyel párhuzamosan, $F(x)$ ordinátájának $\frac{1}{2}$ magasságában vegyünk fel egy egyenest. Ha az $F(x)$ gráfjának az egyenessel való metszéspontját levetítjük az x tengelyre, megkapjuk a mediánt. A 22. ábra az egyetlen megoldás esetét, a 23.

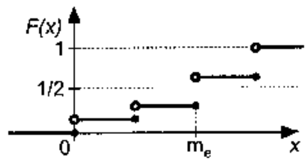
ábra a $\forall x \in]a,b]$-re $F(x)=\frac{1}{2}$ esetét, és a 24. ábra pedig „felső-határ $F(x)<\frac{1}{2}$" esetét szemlélteti.



22. ábra. A medián grafikus meghatározása



23. ábra. A medián grafikus meghatározása



24. ábra. A medián grafikus meghatározása

Az X folytonos valószínűségi változó $f(x)$ sűrűségfüggvényének grafikonja alatti területet az $x=m_e$ egyenes felezi, melyből már következik, hogy szimmetrikus eloszlás esetén a medián egyenlő a változó várható értékével, azaz $m_e=M(X)$.

3.3.3. A p-kvantilis, a terjedelem és a módusz

A medián, m_e olyan szám, amely az eloszlást $\frac{1}{2}:\frac{1}{2}$ arányban osztja ketté. A gyakorlatban előfordul, hogy olyan $0<p<1$ számot kell megadnunk, amely az eloszlást $p:(1-p)$ arányban osztja ketté.

Definíció. Azt az x_p valós számot, amelyre a

$$P(X \le x_p)=p, \quad P(X>x_p)=1-p, \quad 0<p<1 \tag{1}$$

egyenlőtlenségek teljesülnek, az $F(x)$ **eloszlásfüggvény p-kvantilisének** nevezzük, ha csak egyetlen ilyen x_p létezik.

A mediánhoz hasonlóan itt is még két esetet különböztethetünk meg. Előfordulhat, hogy $\forall x \in]a,b]$-re p-vel egyenlő, akkor $F(x)$ p-kvantilise ezen intervallum középpontja, azaz

$$p=\frac{a+b}{2}.$$

Ha a p értékkel nem lesz egyenlő az $F(x)$ eloszlásfüggvény egyetlen x_p helyen sem, akkor p-kvantilisnek az $F(x)<p$ feltételt kielégítő x értékek felső határát nevezzük.

Abban az esetben, ha $F(x)$ szigorúan monoton növekedő, akkor az

$$F(x)=p \tag{2}$$

egyenletnek egyetlen megoldása van, és az x_p-vel egyenlő. A medián definiciójának megfelelően, a $p=\frac{1}{2}$ értékhez tartozó $x_{\frac{1}{2}}$ kvantilis a medián. Gyakrabban használjuk a $p=\frac{1}{4}$, és a $p=\frac{3}{4}$ értékekhez tartozó kvantiliseket, az előbbit **alsó,** az utóbbit pedig **felső kvantilisnek** nevezzük.

A valószínűségi változó jellemzését kiegészíthetjük azzal is, hogy megadjuk azt az m_t-vel jelölt legszűkebb intervallumot, amelybe 1 valószínűséggel beletartozik az X.

Definíció. Ha valamely X valószínűségi változó korlátos és $]a,b]$ az a legszűkebb intervallum, amelyre

$$P(a \le X \le b) = 1 \tag{3}$$

akkor az intervallum $m_t = b - a$ hosszát az X **valószínűségi változó terjedelmének** nevezzük.

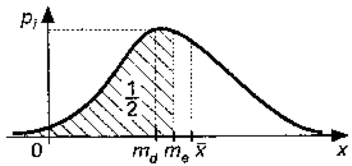
Igazolható, hogy *korlátos eloszlásokra* a variancia legfeljebb a terjedelem fél hosszával egyenlő, azaz

$$\mathrm{Var}(X) = D^2(X) \le \frac{m_t}{2}. \tag{4}$$

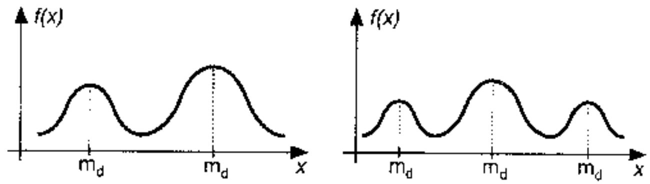
Mind a diszkrét, mind a folytonos valószínűségi változó jellemzésére azt is célszerű megadni, hogy melyik x_i értéket veszi fel a legnagyobb valószínűséggel, ill. mely értékeket vesz fel relatíve nagyobb valószínűséggel.

Definíció. *Diszkrét valószínűségi változó* $x_1, x_2, \ldots$ lehetséges értékei közül azt az x_i számot nevezzük az X változó **móduszának**, amelyre a $P(X = x_i) = p_i$ szám a legnagyobb. Ez az x_i a lehetséges értékek közül a legvalószínűbb. Ha több ilyen érték van, akkor mindegyik módusza a valószínűségi változónak.

Ha *X folytonos eloszlású valószínűségi változó*, akkor a sűrűségfüggvénye lokális maximumhelyeit nevezzük az eloszlás móduszainak, jelölése: m_d. Az eloszlásnak tehát több módusza is lehet. Az egy, két, három stb. móduszú eloszlást unimodális, bimodális, trimodális stb. eloszlásnak mondjuk (25. ábra. és 26. ábra).



25. ábra. Unimodális eloszlás módusz (m_d), medián (m_e) és középérték ($\bar{x}$)



26. ábra. Bimodális és trimodális eloszlás

Példa

Számítsuk ki az

$$F(x) = \begin{cases} 0, \text{ ha } x < 0 \\ \frac{1}{8}x^3, \text{ ha } 0 \le x \le 2 \\ 1, \text{ ha } x > 2 \end{cases}$$

eloszlásfüggvény mediánját, alsó- és felső kvantilisét.

Megoldás

Az $\frac{1}{8}x^3 = \frac{1}{2}$, $\frac{1}{8}x^3 = \frac{1}{4}$, $\frac{1}{8}x^3 = \frac{3}{4}$ egyenletek megoldásai adják a medián, az alsó és felső kvantilis x_p értékeit:

Medián: $m_e = x = \sqrt[3]{4} \approx 1{,}5874$

Alsó kvantilis: $x_{\frac{1}{4}} = \sqrt[3]{2} \approx 1{,}2599$

Felső kvantilis: $x_{\frac{3}{4}} = \sqrt[3]{6} \approx 1{,}8171$.

E.3. Ellenőrző kérdések a 3. fejezethez

1. Mit nevezünk egy X valószínűségi változó eloszlásának és várható értékének?

2. Egy játékot mikor nevezünk korrektnek ill. a játékos szempontjából kedvezőnek vagy kedvezőtlennek?

3. Hogyan definiáljuk az X valószínűségi változó szórásnégyzetét és szórását?

4. Hogyan értelmezzük az X és Y valószínűségi változók együttes eloszlását és peremeloszlásukat?

5. Hogyan definiáljuk az X és Y valószínűségi változók kovarianciáját és korrelációját?

6. Milyen következtetést vonhatunk le a korreláció értékéből az X és Y valószínűségi változók közötti kapcsolatra vonatkozólag?

7. Hogyan értelmezzük az X folytonos valószínűségi változó esetén az eloszlást, a várható értéket, a szórásnégyzetet és a szórást?

8. Hogyan értelmezzük az X valószínűségi változó eloszlásfüggvényét és sűrűségfüggvényét?

V.3. Feladatok a 3. fejezethez

V.3.1. Számítsuk ki az m várható értéket, a szórásnégyzetet és a szórást az alábbi táblázatokkal adott eloszlásokhoz:

a)

x_i	2	3	11
$v(x_i)$	1/3	1/2	1/6

b)

x_i	−5	−4	1	2
$v(x_i)$	1/4	1/8	1/2	1/8

V.3.2. Egy szabályos pénzérmét dobjunk fel négyszer egymás után. Határozzuk meg az X és Y eloszlását, várható értékét, szórásnégyzetét és szórását, ha

a) X jelenti a négy dobásonkénti fejek számát;

b) Y jelenti a négy dobásonkénti fej dobások sorozatának hosszát.

V.3.3. Egy játék három szabályos pénzérme feldobásából áll. A játékos 3 fej dobásakor 5 forintot, 2 fej dobásakor 3 forintot, 1 fej dobásakor 1 forintot kap, egyébként pedig azaz 3 írás dobásakor a játékos fizet a banknak 15 forintot. Állapítsuk meg, hogy a játékos szempontjából kedvező vagy kedvezőtlen a játék?

V.3.4. Tegyük fel, hogy az X és Y valószínűségi változók együttes eloszlása az alábbi táblázattal adott:

$X \setminus Y$	−3	2	4	$\sum$ sor
1	0,1	0,2	0,2	0,5
3	0,3	0,1	0,1	0,5
$\sum$ oszlop	0,4	0,3	0,3	

Határozzuk meg az X és Y

a) eloszlását;

b) kovarianciáját;

c) korrelációját és

d) vizsgáljuk meg X és Y függetlenségét.

V.3.5. Határozzuk meg az X és Y valószínűségi változókhoz az

a) $M(X)$ és $M(Y)$;

b) $\mathrm{Cov}(X,Y)$;

c) $D(X)$, $D(Y)$ és $R(X,Y)$ értékeket, ha az együttes eloszlások táblázata:

A)

$X \setminus Y$	−4	2	7	$\sum$ sor
1	1/8	1/4	1/8	1/2
5	1/4	1/8	1/8	1/2
$\sum$ oszlop	3/8	3/8	1/4	

B)

$X \setminus Y$	−2	−1	4	5	$\sum$ sor
1	0,1	0,2	0	0,3	0,6
2	0,2	0,1	0,1	0	0,4
$\sum$ oszlop	0,3	0,3	0,1	0,3	

V.3.6. Legyenek X és Y független valószínűségi változók az

x_i	1	2
$v(x_i)$	0,7	0,3

y_i	−2	5	8
$u(y_i)$	0,3	0,5	0,2

eloszlásokkal adottak. Állítsuk elő az X és Y együttes eloszlástáblázatát, és igazoljuk, hogy $\mathrm{Cov}(X,Y)=0$.

V.3.7. Legyen az X folytonos valószínűségi változó:

$$f(x)=\begin{cases}\dfrac{1}{6}x+b, & \text{ha } 0\le x\le 3\\ 0, & \text{ha } x<0 \text{ és ha } x>3\end{cases}$$

sűrűségfüggvénnyel adott.

a) Számítsuk ki a b értékét;

b) Számítsuk ki a $P(1 \le X \le 2)$ értékét.

V.3.8. Az X diszkrét valószínűségi változó az

x_i	-2	1	2	4
$v(x_i)$	1/4	1/8	1/2	1/8

eloszlási táblázattal adott. Ábrázoljuk az $F(x)$ eloszlásfüggvényt.

V.3.9. Legyen az X folytonos valószínűségi változó sűrűségfüggvénye:

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

Határozzuk meg és ábrázoljuk az $F(x)$ eloszlásfüggvényt.

V.3.10. Számítsuk ki az X valószínűségi változó első négy momentumát, ha eloszlása:

x_i	-2	1	3
$v(x_i)$	1/2	1/4	1/4

V.3.11. Egy érme olyan súlyozású, hogy a fejdobás valószínűsége $P(F) = \frac{2}{3}$, az írásdobásé pedig $P(I) = \frac{1}{3}$. Dobjuk fel az érmét háromszor egymásután. Jelentse az X valószínűségi változó a dobásonkénti fejdobások sorozatának hosszát. Határozzuk meg az X változó értékeit, eloszlását, várható értékét.

V.3.12. Egy X valószínűségi változó a -2, 0 és 4 értékeket veheti fel, melyekhez tartozó valószínűségeloszlás:

$$P(X = -2) = \frac{3}{10}; \quad P(X = 0) = \frac{1}{2}; \quad P(X = 4) = \frac{1}{5}.$$

a) Írjuk fel és ábrázoljuk az X változó eloszlásfüggvényét;

b) Számítsuk ki az eloszlásfüggvény értékét az $X = -3$, $X = -2$, $X = 1$, $X = 5$ helyeken;

c) $P(X \ge 0) = ?$; $P(X \le 0) = ?$.

V.3.13. Egy dobozban 3 friss és 2 záptojás van. Visszatevés nélkül véletlenszerűen kiválasztunk hármat. Az X valószínűségi változó jelentse a kivett friss tojások számát. Adjuk meg az X változó

a) valószínűségeloszlását és gráfját;

b) eloszlásfüggvényét és gráfját;

valamint annak valószínűségét, hogy legalább egy friss, ill. legfeljebb egy friss tojás van a kivettek között.

V.3.14. Egy csokoládégyár 110 egységnyi és 220 egységnyi titkosított adalékanyaggal gyárt 20, 40, 60 és 100 g-os csokoládé szeleteket. Az ALFA üzletháznak a következő táblázat szerinti összetételben teljesítették a megrendelést:

$X \backslash Y$	20	40	60	100
110	50	100	100	250
220	250	300	400	550

Válasszunk ki véletlenszerűen egyet. Az X valószínűségi változó jelentse a csokoládé szelet adalékanyag mennyiségét, Y pedig a tömegét. Adjuk meg az (X, Y) valószínűségi vektorváltozó

a) együttes valószínűségeloszlását és peremeloszlását,

b) együttes eloszlásfüggvényét és a perem-eloszlásfüggvényeket,

c) X és Y változóinak kovarianciáját és korrelációs együtthatóját.

NEGYEDIK FEJEZET

Nevezetes eloszlások

A valószínűségszámítás, mint minden más tudomány, a különböző jelenségek vizsgálata során arra törekszik, hogy feltárja azokat az elvonatkoztatható jegyeket, amelyek lehetővé teszik a különböző jelenségek bizonyos szempontok szerinti csoportba sorolását, mert ezzel azonos módszerek alkalmazására nyílik lehetőségünk. Számos jelenségről megállapították, hogy a valószínűség szempontjából hasonlóan viselkednek, pl. ugyanazt az eloszlást követik, vagyis az X valószínűségi változó x_i értékeihez a p_i valószínűségek azonos eljárással (képlettel) számíthatók ki, függetlenül x_i jelentésétől, vagy csak paraméterértékekben különbözőek, de ugyanazzal az eloszlásfüggvénnyel írhatók le. A következő pontokban azt vizsgáljuk, hogy bizonyos feltételeknek eleget tevő valószínűségi változóknak milyen jellegzetes eloszlásfüggvényei vannak.

4.1. Diszkrét valószínűségeloszlások

A gyakorlati feladatok kísérleteinél legtöbbször csak az érdekel bennünket, hogy a kísérlet eredményes volt-e vagy eredménytelen. Olyan kísérletsorozatot vizsgálunk tehát, amelynek pontosan két kimenetele van, vagyis a kísérlet során azt figyeljük meg, hogy valamely A esemény bekövetkezett-e vagy sem. Ilyen kérdésre ad választ a következő eloszlás.

4.1.1. Binomiális eloszlás

Mi annak a valószínűsége, hogy egy kísérletben az A esemény pontosan k-szor következik be?

Tekintsünk egy független kísérletekből álló kísérletsorozatot. Legyen az A esemény bekövetkezésének valószínűsége minden kísérletben $P(A)=p$ és az ellentett esemény $\overline{A}$ bekövetkezésének valószínűsége:

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

Ismételjük meg a kísérletet n-szer egymástól függetlenül, és az X diszkrét valószínűségi változó értéke legyen az A esemény bekövetkezéseinek számával egyenlő.

Tétel

Annak valószínűsége, hogy n $(n=1,2,\ldots)$ független kísérletsorozatban az A esemény pontosan k-szor következik be, azaz az A esemény bekövetkezéseinek számát megadó X valószínűségi változó x_k értéke éppen k (tehát $x_k=k$) a

$$p_k=P(X=k)=\binom{n}{k}p^kq^{n-k}, \tag{1}$$

$$(k=0,1,2,\ldots,n)\ (0\le p\le 1)$$

képlettel adható meg, ahol $q=1-p$, és $\binom{n}{k}=\frac{n!}{k!(n-k)!}$ binomiális együttható.

Az (1) eloszlást ***n*-edrendű, *p* paraméterű binomiális eloszlásnak** nevezzük, mivel a $p_k=P_n(k)$ valószínűségek a $(p+q)^n$ hatványmennyiség binomiális tétel szerinti kifejtésének tagjai, azaz az eloszlást a következő táblázattal adhatjuk meg:

k	0	1	2	....	n
$P_n(k)$	q^n	$\binom{n}{1}q^{n-1}p$	$\binom{n}{2}q^{n-2}p^2$	....	p^n

Az 1. tétel bizonyítását az A és B elemekből álló olyan n elemhosszúságú sorozatok vizsgálata alapján végezhetjük, amelyekben az A elem pontosan k-szor, B elem pedig $(n-k)$-szor fordul elő.

Az ilyen elem-n-es sorozatok száma: $\binom{n}{k}$. Mivel feltettük, hogy a kísérletek egymástól függetlenek, ezért minden elem-n-es valószínűsége: p^kq^{n-k}, ahol $q=1-p$, s így az A előfordulásának valószínűsége:

$$P(A)=\binom{n}{k}p^kq^{n-k};\ (k=0,1,2,\ldots,n).$$

A binomiális eloszlást általánosabban is definiálhatjuk.

Legyen X diszkrét valószínűségi változó, amelynek értékei a természetes számok, azaz $X=\{0,1,2,\ldots,n,\ldots\}$. Ha annak valószínűsége, hogy X éppen a k értéket veszi fel

$$p_k=\binom{n}{k}p^kq^{n-k};\quad (k=0,1,2,\ldots,n)$$

akkor az X valószínűségi változó binomiális eloszlású. A binomiális eloszlás tagjai kezdetben a k-val együtt növekednek, majd a maximum elérése után csökkennek. Ha np egész szám, akkor az ehhez tartozó valószínűség a legnagyobb, ha pedig nem egész szám, akkor az np-hez legközelebbi egész számhoz tartozó valószínűség a legnagyobb.

Az (1) binomiális eloszlást $b(k;n,p)$-vel is szokás jelölni, azaz

$$b(k;n,p)=\binom{n}{k}p^kq^{n-k} \tag{1*}$$

amelynek $k<x$-re vonatkozó összegezésével kapjuk az **eloszlásfüggvényt:**

$$F_n(x)=\sum_{k<x}\binom{n}{k}p^kq^{n-k}. \tag{2}$$

Megjegyzések

1. Ilyen kétkimenetelű független kísérletek vizsgálatával először *J. Bernoulli* foglalkozott, ezért a fentiekkel kapcsolatban *Bernoulli*-kísérlet és *Bernoulli*-eloszlás elnevezések is használatosak.

2. A $P_n(k)$ valószínűségek kényelmesebb kiszámítását a

$$P_n(k+1)=\frac{n-k}{k+1}\cdot\frac{p}{1-p}\cdot P_n(k) \qquad (1^{**})$$

rekurziós képlettel végezhetjük.

A binomiális eloszlású X valószínűségi változó

várható értéke: $m=M(X)=np$ (3)

szórásnégyzete: $s^2=\mathrm{Var}(X)=D^2(X)=npq$ (4)

szórása: $s=\sqrt{\mathrm{Var}(X)}=D(X)=\sqrt{npq}$ (5)

A (3), (4) és (5) formulák a megfelelő definíciók és a binomiális sorbafejtés képleteinek felhasználása alapján igazolhatók. Ui. az (1*) jelölést használva

$$M(X)=\sum_{k=0}^{n}k\cdot b(k;n,p)=\sum_{k=0}^{n}k\frac{n!}{k!(n-k)!}p^kq^{n-k}.$$

$k=0$-ra az első tag kiesik, egyszerűsítsünk k-val és emeljük ki np tényezőt:

$$M(X)=np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}q^{n-k}.$$

$k-1=r$ helyettesítéssel az összegezést $r=0$ -tól $n-1$-ig végezzük:

$$M(X)=np\sum_{r=0}^{n-1}\frac{(n-1)!}{r!(n-1-r)!}p^rq^{n-1-r}=np\sum_{r=0}^{n-1}b(r;n-1,p),$$

de a binomiális tétel értelmében:

$$\sum_{r=0}^{n-1}b(r;n-1,p)=(p+q)^{n-1}=1^{n-1}=1,$$

tehát

$$M(X)=np.$$

A fenti lépéseket alkalmazva $M(X^2)=\sum_{k=0}^{n}k^2b(k;n,p)$ formulára, kapjuk:

$$M(X^2)=np(np+q),$$

és felhasználva a $D^2(X)=M(X^2)-(M(X))^2$ összefüggést, kapjuk, hogy

$$D^2(X)=np(np+q)-(np)^2=npq.$$

Példák

1. Egy szabályos pénzérmét dobjunk fel hatszor és figyeljük meg a fej dobások számát. Állítsuk elő a valószínűségi változó eloszlását, eloszlásfüggvényét és grafikonjait.

Megoldás

Bernoulli-kísérletről van szó, tehát az (1*) ill. (1**) képlettel előállítjuk az eloszlás táblázatát $n=6$, $p=q=\frac{1}{2}$, $k=0,1,2,3,4,5,6$ értékekre. A

$$P_6(k)=\binom{6}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{6-k}$$

ill.

$$P_n(k+1)=\frac{6-k}{k+1}\cdot\frac{\frac{1}{2}}{1-\frac{1}{2}}\cdot P_6(k)=\frac{6-k}{k+1}P_6(k)$$

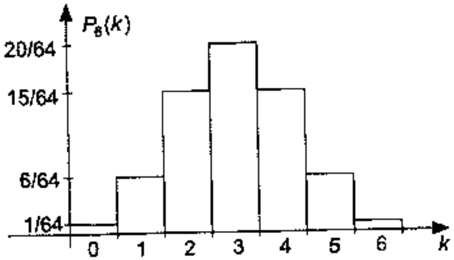
képleteket használva:

k	0	1	2	3	4	5	6
$P_6(k)$	$\frac{1}{64}$	$\frac{6}{64}$	$\frac{15}{64}$	$\frac{20}{64}$	$\frac{15}{64}$	$\frac{6}{64}$	$\frac{1}{64}$

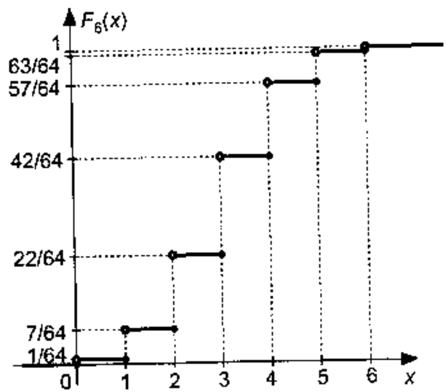
Az eloszlásfüggvényt $F_6(x)=\sum_{k<x}P(k)$ képlettel számítjuk:

k	0	1	2	3	4	5	6	6 <
$F_6(k)$	0	$\frac{1}{64}$	$\frac{7}{64}$	$\frac{22}{64}$	$\frac{42}{64}$	$\frac{57}{64}$	$\frac{63}{64}$	1

A binomiális eloszlású valószínűségi változó eloszlását a 27. ábra, a binomiális eloszlásfüggvényt a 28. ábra szemlélteti.



27. ábra. $P_6(k)$ valószínűségi változó eloszlása (hisztogramja)



28. ábra. Az $F_6(x)$ binomiális eloszlásfüggvény

2. Mennyi a valószínűsége annak, hogy egy házaspár 6 születendő gyermeke közül

a) 3 fiú lesz;

b) a fiúk száma kevesebb lesz a lányok számánál, ha a fiúk születésének valószínűségét $\frac{1}{2}$-nek vesszük?

Megoldás

Mivel most $n=6$ és $p=q=\frac{1}{2}$, ezért az (1) képlet szerint

a) $P(3\text{ fiú})=\binom{6}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^3=\frac{6\cdot5\cdot4}{1\cdot2\cdot3}\cdot\frac{1}{8}\cdot\frac{1}{8}=\frac{5}{16}$;

b) Ha a fiúk száma 0, 1, 2, akkor kevesebb a lányok számánál, tehát

$$P(0\text{ fiú})+P(1\text{ fiú})+P(2\text{ fiú})=\left(\frac{1}{2}\right)^6+\binom{6}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^5+\binom{6}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^4=\frac{11}{32}.$$

3. Egy számítógéphez 4 egymástól függetlenül működő lemezegység tartozik. Az egyes lemezegységek $6\cdot10^{-4}$ valószínűséggel hibásodhatnak meg. Ha az adatfeldolgozás elvégzéséhez legalább 2 hibátlan lemezegység kell, mi a valószínűsége annak, hogy az adatfeldolgozás sikeresen befejezhető?

Megoldás

Annak, hogy egy lemezegység nem hibásodik meg $p=1-6\cdot10^{-4}=0{,}9994$ a valószínűsége. Ha a hibátlanul működő lemezegységek számát X jelöli, akkor az X valószínűségi változó eloszlása egy $n=4$, $p=0{,}9994$ paraméterű binomiális eloszlás. Tehát a sikeres adatfeldolgozás valószínűsége $P(X\geq2)$, azaz

$$P(\text{sikeres adatfeldolgozás})=P(X\geq2)=\sum_{k=2}^{4}b(k;4,0{,}9994)=$$

$$=\binom{4}{2}0{,}9994^2\cdot0{,}0006^2+\binom{4}{3}0{,}9994^3\cdot0{,}0006+$$

$$+\binom{4}{4}0{,}9994^4\cdot0{,}0006^0=0{,}9999998.$$

4.1.2. Poisson-eloszlás

Mi a valószínűsége annak, hogy egy sajtóhibákat tartalmazó könyv véletlenül kinyitott oldalán van sajtóhiba?

Legyen λ pozitív állandó és X egy diszkrét valószínűségi változó, amelynek értékei 0, 1, 2,..., k,... Ha X a k értékeket

$$P(X=k)=p(k;\lambda)=\frac{\lambda^k e^{-\lambda}}{k!},\ (\lambda>0),\ (k=0,1,2,\ldots)\qquad(1)$$

valószínűséggel veszi fel, akkor az X eloszlását λ paraméterű ***Poisson*-eloszlásnak** nevezzük. A $p(k,\lambda)$ számok valószínűségeloszlást alkotnak, ui.

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

A *Poisson*-eloszlású valószínűségi változóhoz tartozó **eloszlásfüggvényt** a (1) -gyel adott valószínűségek összegezésével kapjuk minden $k<x$ -re kiterjesztve:

$$F_n(x)=\sum_{k<x}\frac{\lambda^k e^{-\lambda}}{k!}. \tag{2}$$

A *Poisson*-eloszlású X valószínűségi változó

várható értéke: $m=M(X)=\lambda;$

szórásnégyzete: $s^2=\mathrm{Var}(X)=D^2(X)=\lambda;$

szórása: $s=\sqrt{\mathrm{Var}(X)}=D(X)=\sqrt{\lambda},$

amelyeket a megfelelő definíciók alapján állíthatunk elő. A bizonyításhoz felhasználjuk, hogy $e^{\lambda}=\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}$. Az (1) formula jelölését használva:

$$M(X)=\sum_{k=0}^{\infty}k\cdot p(k;\lambda)=\sum_{k=0}^{\infty}k\frac{\lambda^k e^{-\lambda}}{k!}.$$

Az első tag $k=0$-ra 0-val egyenlő, így λ kiemelésével és $k-1=r$ helyettesítéssel:

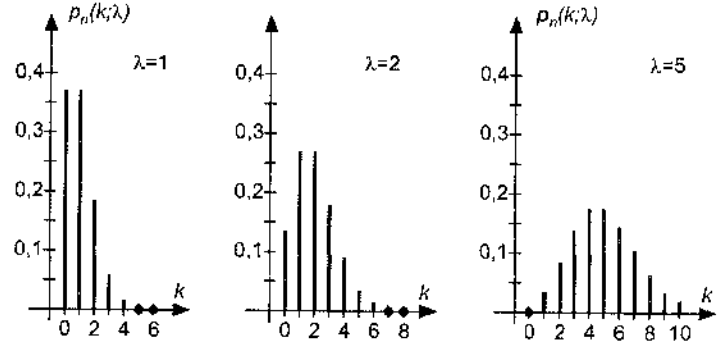
$$M(X)=\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}e^{-\lambda}}{(k-1)!}=$$

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

Hasonló meggondolással $M(X^2)=\lambda(\lambda+1)$ adódik, és így

$$D^2(X)=M(X^2)-\left(M(X)\right)^2=\lambda(\lambda+1)-\lambda^2=\lambda.$$

Poisson-eloszlást szemléltet a 29. ábra $\lambda=1$, $\lambda=2$ és $\lambda=5$ -re:



29. ábra. Poisson-*eloszlások*

Megjegyzések

1. Az egyes valószínűségek kiszámítását megkönnyíti a következő **rekurziós formula:**

$$p_n(k+1;\lambda)=\frac{\lambda}{k+1}p_n(k;\lambda). \tag{1*}$$

2. A *Poisson*-eloszlás a binomiális eloszlásból is származtatható, annak n szerinti határértékeként, ha feltesszük, hogy λ állandó és $p=\frac{\lambda}{n}$, azaz $\lambda=pn$; ekkor ugyanis

$$\lim_{n\to\infty}b(k;n,p)=\lim_{n\to\infty}\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}=\frac{\lambda^k e^{-\lambda}}{k!}.$$

A binomiális eloszlást elég nagy n és kicsi p valószínűség esetén az $np=\lambda$ paraméterű *Poisson*-eloszlás jól közelíti. Ennek belátásához felhasználjuk az

$$\ln(1-z)=-z-\frac{z^2}{2}-\frac{z^3}{3}-\dots \tag{*}$$

Taylor-sort. Ui. ha $\lambda=np$, akkor

$$b(0;n,p)=(1-p)^n=(1-\frac{\lambda}{n})^n.$$

Mindkét oldal természetes logaritmusát képezve:

$$\ln b(0;n,p)=\ln(1-\frac{\lambda}{n})^n=n\ln(1-\frac{\lambda}{n}).$$

(*) figyelembe vételével, $z=\frac{\lambda}{n}$ helyettesítéssel:

$$\ln(1-\frac{\lambda}{n})=-\frac{\lambda}{n}-\frac{\lambda^2}{2n^2}-\frac{\lambda^3}{3n^3}-\ldots$$

Ha n igen nagy, akkor

$$\ln b(0;n,p)=n\ln(1-\frac{\lambda}{n})=-\lambda-\frac{\lambda^2}{2n}-\frac{\lambda^3}{3n^2}\approx-\lambda,$$

és így

$$b(0;n,p)\approx e^{-\lambda}. \qquad (**)$$

Ha a p elegendő kicsi, akkor $q\approx 1$, és

$$\frac{b(k;n,p)}{b(k-1;n,p)}=\frac{(n-k+1)p}{kq}=\frac{\lambda-(k-1)p}{kq}\approx\frac{\lambda}{k}.$$

Ekkor $b(k;n,p)\approx\frac{\lambda}{k}b(k-1;n,p)$ összefüggésből (**) felhasználásával $b(1;n,p)\approx\lambda e^{-\lambda}$, $b(2;n,p)\approx\frac{\lambda^2 e^{-\lambda}}{2}$, és teljes indukcióval kapjuk, hogy

$$b(k;n,p)\approx\frac{\lambda^k e^{-\lambda}}{k!}=p(k;\lambda).$$

3. Az e^x és e^{-x} értékeket a [0;3] intervallumban a 6. táblázat tartalmazza.

Példák

1. Számítsuk ki az X valószínűségi változó *Poisson*-eloszlását $\lambda=1$, $\lambda=2$, $\lambda=5$ paraméterekkel, ha $k=0,1,2,3,4,5,6,7,8,9,10$.

Megoldás

A $k=0$ értékhez a (1) képletet, a k további értékeihez a (1*) képletet használjuk. Az eredményt táblázatba foglaltuk. (L. 29. ábra.) Pl.

$$P(X=0)=p(0;1)=\frac{1^0\cdot e^{-1}}{0!}=\frac{1}{e}=\frac{1}{2,718282}=0,3679\approx 0,368 \text{ stb.}$$

k	0	1	2	3	4	5
$p(k;1)$	0,368	0,368	0,184	0,0613	0,0153	0,00307
$p(k;2)$	0,1353	0,2707	0,2707	0,1805	0,0902	0,0361
$p(k;5)$	0,00674	0,0337	0,0842	0,1404	0,1755	0,1755

k	6	7	8	9	10
$p(k;1)$	0,00051	0,00007	0,0000		
$p(k;2)$	0,012	0,0034	0,0009	0,0002	0,0000
$p(k;5)$	0,1462	0,1045	0,0653	0,0363	0,0181

2. Tegyük fel, hogy az X valószínűségi változó *Poisson*-eloszlású. Számítsuk ki a $p(3;\frac{1}{2})$, $p(2;0,7)$ valószínűségeket.

Megoldás

$$P(X=3)=p(3;\frac{1}{2})=\frac{\left(\frac{1}{2}\right)^3\cdot e^{-\frac{1}{2}}}{3!}=\frac{0,607}{48}=0,013;$$

$$P(X=2)=p(2;0,7)=\frac{(0,7)^2\cdot e^{-0,7}}{2!}=\frac{0,49\cdot 0,497}{2}=0,12.$$

3. Tegyük fel, hogy egy 500 oldalas könyvben véletlen eloszlásban 300 sajtóhiba van. Számítsuk ki annak valószínűségét, hogy egy adott oldalon

a) pontosan 2 sajtóhiba van;

b) 2 vagy több (legalább 2) sajtóhiba van.

Megoldás

Ha azt vizsgáljuk, hogy oldalanként hány sajtóhiba van, akkor $n=300$ és $p=\frac{1}{500}$ paraméterű binomiális eloszlást kapunk. Mivel p elég kicsi és n elég nagy, a binomiális eloszlást *Poisson*-eloszlással közelítjük $\lambda=np=0,6$ paraméterérték mellett.

a) $p(2;0,6)=\frac{(0,6)^2\cdot e^{-0,6}}{2!}=\frac{0,36\cdot 0,549}{2}=0,0988.$

b) Annak az eseménynek a valószínűségét, hogy egy oldalon 2 vagy több, azaz legalább 2 hiba van, az ellentett események valószínűségei közötti összefüggés alapján számítjuk ki. Kettőnél kevesebb hiba egy oldalon úgy következhet be, hogy az X a 0 vagy 1 értéket veszi fel, s minthogy ezek az események egymást kizárják, így összegük valószínűsége p_0+p_1. A vizsgált esemény valószínűsége:

$$P(X\geq 2)=1-(p_0+p_1)=1-\left(\frac{(0,6)^0\cdot e^{-0,6}}{0!}+\frac{0,6\cdot e^{-0,6}}{1!}\right)=$$

$$=1-(e^{-0,6}+0,6\cdot 0,549)=1-(0,549+0,329)=0,122.$$

Megjegyzés

A gyakorlati feladatokban gyakran találkozunk *Poisson*-eloszlással. Pl. egy telefonközpontban egy adott időintervallumban jelentkező hívások átlagos száma, vérsejtek száma egy térrészben, egy adott tartományba hulló esőcseppek száma, a radioaktív bomlások száma egy adott időtartam alatt, a csillagok a térben, közelítőleg *Poisson*-eloszlásúak stb.

4.1.3. Hipergeometrikus eloszlás

Mi annak a valószínűsége, hogy N számú alkatrészből visszatevés nélkül kiválasztott n db-ból álló mintában a selejtesek száma k db, ha a N számú alkatrész között M db selejtes van?

Definíció. Azt az X valószínűségi változót, amely az $x_k=k$ $(k=0,1,2,\ldots,n)$ értékeket

$$p_k=P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\quad (k=0,1,2,\ldots,n) \qquad (1)$$

valószínűséggel veszi fel, **hipergeometrikus eloszlású** valószínűségi változónak nevezzük, ahol N, M, n nemnegatív egészek és $M<N$, $0<n\leq \min(M,N-M)$.

A hipergeometrikus eloszlás tagjai kezdetben k-val együtt növekednek, majd a maximum elérése után csökkennek. Az eloszlás tagjait jól közelíthetjük az n-edrendű $\frac{M}{N}=p$ paraméterű binomiális eloszlás megfelelő tagjaival, ha a k-hoz képest M és N elég nagy.

Egy hipergeometrikus eloszlású X valószínűségi változó

várható értéke: $m=M(X)=n\frac{M}{N}$;

szórásnégyzete: $s^2=\mathrm{Var}(X)=D^2(X)=np(1-p)\left(1-\frac{n-1}{N-1}\right)$;

szórása: $s=\sqrt{\mathrm{Var}(X)}=D(X)=\sqrt{np(1-p)\left(1-\frac{n-1}{N-1}\right)}$.

Példa

Egy 32 lapos magyar kártyacsomagot négy játékos között egyenlően osztunk el. Az X változó értéke legyen az egyik kijelölt játékoshoz kerülő piros lapok száma. (Feltesszük, hogy jól megkevert kártyacsomagból mindig véletlenszerű az elosztás.) Készítsük el az X valószínűségi változó eloszlásának táblázatát, továbbá számítsuk ki várható értékét és szórását.

Megoldás

Az X valószínűségi változó $x_k=k$ $(k=0,1,\ldots,8)$ értékek felvételével hipergeometrikus eloszlású. A lapok száma $N=32$, amelyből $M=8$ a piros lapok száma, és $n=8$ lapot kap mindegyik játékos. Az adatok alapján annak valószínűsége, hogy az X a k értéket veszi fel az (1) képlet szerint:

$$p_k=P(X=k)=\frac{\binom{8}{k}\binom{24}{8-k}}{\binom{32}{8}}\qquad (k=0,1,2,\ldots,8)$$

Az X valószínűségi változó eloszlástáblázata:

k	0	1	2	3	4	5	6	7	8
p_k	0,07	0,26	0,36	0,23	0,07	0,01	0,0	0,0	0,0

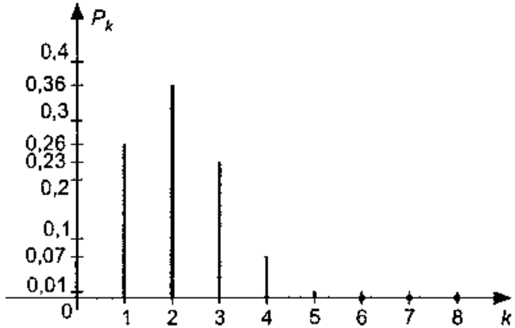
várható értéke:

$$M(X)=n\frac{M}{N}=8\cdot\frac{8}{32}=2;$$

szórása:

$$D(X)=\sqrt{np(1-p)\left(1-\frac{n-1}{N-1}\right)}=\sqrt{8\frac{8}{32}\left(1-\frac{8}{32}\right)\left(1-\frac{8-1}{32-1}\right)}=1{,}16.$$

Az X hipergeometrikus változó eloszlását a 30. ábra szemlélteti.



30. ábra. Hipergeometrikus eloszlás

4.2. Folytonos eloszlások

Az előzőekben három nevezetes diszkrét valószínűségi eloszlással ismerkedtünk meg: a *binomiális*, a *Poisson-* és a *hipergeometrikus* eloszlással. A következő pontokban folytonos eloszlásokat tárgyalunk. A folytonos eloszlásokat a sűrűségfüggvényeikkel definiáljuk, de magadjuk az eloszlásfüggvényeket is. Mivel az X folytonos valószínűségi változó egyes konkrét érékeinek valószínűsége 0-val egyenlő, így a változó jellemzésére a sűrűségfüggvényt vagy az eloszlásfüggvényt használhatjuk. Megkülönböztetés és rövidítés céljából a folytonos eloszlások várható értékét μ-vel, szórásnégyzetét σ^2-tel, szórását pedig σ-val fogjuk jelölni.

Példa

Legyen

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

a) Vizsgáljuk meg, hogy teljesül-e az $f(x)\ge 0$ és az $\int_{-\infty}^{\infty} f(x)dx=1$ feltétel, ha igen, határozzuk meg a sűrűségfüggvény eloszlásfüggvényét, és mindkét függvényt ábrázoljuk.

b) Számítsuk ki az X változó várható értékét, szórásnégyzetét és szórását.

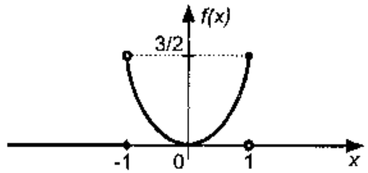
c) Mekkora valószínűséggel tér el az X változó a várható értékétől legfeljebb 0,25-dal?

Megoldás

a) Mivel $x^2\ge 0$ minden valós számra, így nyilvánvaló, hogy $f(x)\ge 0$, ha $x\in R$.

$$\int_{-\infty}^{\infty} f(x)dx=\int_{-\infty}^{-1}0dx+\int_{-1}^{1}\frac{3}{2}x^2dx+\int_{1}^{\infty}0dx=0+\frac{3}{2}\left[\frac{x^3}{3}\right]_{-1}^{1}+0=$$
$$=\frac{3}{2}\cdot\left(\frac{1}{3}-(-\frac{1}{3})\right)=\frac{3}{2}\cdot\frac{2}{3}=1.$$

Tehát az adott f függvény valóban sűrűségfüggvény, melynek gráfja (31. ábra.):



31. ábra. Az f(x) sűrűségfüggvény

Az X valószínűségi változó sűrűségfüggvényének $F(x)$ eloszlásfüggvénye:

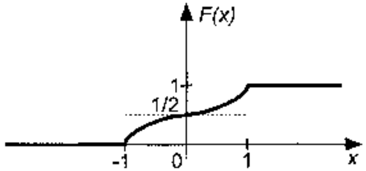
Ha $x\le -1$, $F(x)=\int_{-\infty}^{x}0dt=0$;

Ha $-1\le x\le 1$, $F(x)=\int_{-\infty}^{-1}0dt+\int_{-1}^{x}\frac{3}{2}t^2dt=\frac{3}{2}\left[\frac{t^3}{3}\right]_{-1}^{x}=\frac{x^3}{2}+\frac{1}{2}$;

ha $x>1$, $F(x)=\int_{-\infty}^{x}f(t)dt=\int_{-\infty}^{-1}0dt+\int_{-1}^{1}\frac{3}{2}t^2dt+\int_{1}^{x}0dt=\frac{3}{2}\cdot\frac{2}{3}=1.$

Tehát az eloszlásfüggvény:

$$F(x)=\begin{cases}0, \text{ ha } x\le -1\\ \dfrac{x^3}{2}+\dfrac{1}{2}, \text{ ha } -1\le x\le 1,\\ 1, \text{ ha } x>1\end{cases}$$ melynek gráfja (32. ábra):



32. ábra. Az $F(x)$ eloszlásfüggvény

b) Az X valószínűségi változó várható értéke:

$$\mu=M(X)=\int_{-\infty}^{\infty}xf(x)=\int_{-1}^{1}x\cdot\frac{3}{2}x^2dx=\frac{3}{2}\left[\frac{x^4}{4}\right]_{-1}^{1}=\frac{3}{2}\left(\frac{1}{4}-\frac{1}{4}\right)=0,$$

szórásnégyzete:

$$\sigma^2=D^2(X)=\int_{-\infty}^{\infty}x^2f(x)dx-\mu^2=\int_{-1}^{1}x^2\cdot\frac{3}{2}x^2dx-0^2=\frac{3}{2}\left[\frac{x^5}{5}\right]_{-1}^{1}=$$

$$=\frac{3}{2}\left(\frac{1}{5}+\frac{1}{5}\right)=\frac{3}{5}=0{,}6;$$

és szórása:

$$\sigma=D(X)=\sqrt{0{,}6}\approx 0{,}774597.$$

c) Ki kell számítani a $P(|X-\mu|<0{,}25)$ valószínűséget:

$$P(|X-\mu|<0{,}25)=P(\mu-0{,}25<X<\mu+0{,}25)=$$

$$=P(-0{,}25<X<0{,}25)=F(0{,}25)-F(-0{,}25)\approx$$

$$\approx 0{,}5078-0{,}4921=0{,}0157.$$

Tehát az X a várható értékétől 0,25-dal legfeljebb 0,0157 valószínűséggel tér el.

A folytonos valószínűségi változók eloszlásainak vizsgálatát általában a bemutatott példához hasonlóan végezhetjük. A nevezetesebb folytonos eloszlások közül a következő pontokban az *egyenletes eloszlást*, az *exponenciális eloszlást* és a *normális eloszlást* tárgyaljuk részletesebben.

4.2.1. Egyenletes eloszlás

Definíció. Az X valószínűségi változót **egyenletes eloszlásúnak** nevezzük az $]a;b[$ intervallumon, ha sűrűségfüggvénye (33. ábra):

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & \text{ha } a<x<b,\\ 0 & \text{egyébként}\end{cases}\quad x\in R \qquad (1)$$

Azonnal látható, hogy $f(x)>0$, mivel $b-a>0$, és

$$\int_{-\infty}^{\infty}f(x)dx=\int_{a}^{b}\frac{dx}{b-a}=\frac{1}{b-a}[x]_a^b=\frac{b-a}{b-a}=1.$$

A sűrűségfüggvény definíciójából következik, hogy az egyenletes eloszlású X változó **eloszlásfüggvénye** (34. ábra):

$$F(x)=P(X<x)=\int_{-\infty}^{x}f(t)\mathrm{d}t=\int_{a}^{x}\frac{dt}{b-a}=\begin{cases}\dfrac{x-a}{b-a}, & \text{ha } a<x<b;\\ 0, & \text{ha } x\le a;\\ 1, & \text{ha } x>b.\end{cases} \qquad (2)$$

és a várható érték, szórás, szórásnégyzet definíciói alapján:

várható értéke: $\mu=M(X)=\dfrac{a+b}{2}$;

szórása: $\sigma=D(X)=\dfrac{b-a}{\sqrt{12}}$;

szórásnégyzete: $\sigma^2=D^2(X)=\dfrac{(b-a)^2}{12}$.

Ui. a várható érték:

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

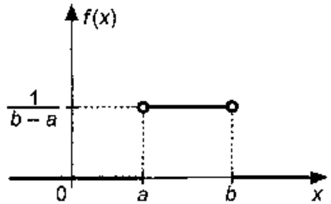
$$=\frac{1}{b-a}\cdot\frac{1}{2}\cdot(b^2-a^2)=\frac{b+a}{2}.$$

A szórásnégyzet és szórás:

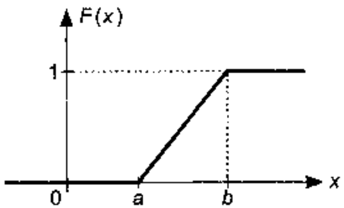
$$\sigma^2=D^2(X)=M(X^2)-(M(X))^2=\int_a^b x^2\cdot\frac{1}{b-a}dx-\left(\frac{b+a}{2}\right)^2=$$

$$=\frac{1}{b-a}\cdot\frac{1}{3}\cdot(b^3-a^3)-\left(\frac{b+a}{2}\right)^2=\frac{(b-a)^2}{12},$$

$$\sigma=\frac{b-a}{\sqrt{12}}=\frac{b-a}{2\sqrt{3}}.$$



33. ábra. Egyenletes eloszlás sűrűségfüggvénye



34. ábra. Egyenletes eloszlásfüggvény

Ha feltesszük, hogy X egyenletes eloszlású az $]a,b]$ intervallumban, akkor annak valószínűsége, hogy X az $]r,s]\subset]a,b]$ részintervallumba esik

$$P(r\le X<s)=\int_r^s f(x)dx=\frac{1}{b-a}\int_r^s dx=\frac{s-r}{b-a},$$

vagyis arányos az intervallum hosszával. (L. a geometriai valószínűséget az 1.5. pontban.)

Példa

Egy üzemi telefonközpont telefonhívásainál azt tapasztaljuk, hogy a tárcsázást követő kapcsolásig terjedő időtartam 1 mp-től 100 mp-ig terjedhet. Az eltelt idő legyen az X egyenletes eloszlású valószínűségi változó. Határozzuk meg az X valószínűségi változó sűrűségfüggvényét, eloszlásfüggvényét, várható értékét, szórását, valamint annak valószínűségét, hogy legalább 50 mp-ig kell várnunk a kapcsolásra.

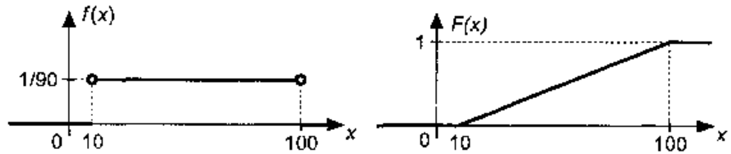
Megoldás

Feltettük, hogy X egyenletes eloszlású a]10;100[intervallumban, ezért sűrűségfüggvénye (35. ábra):

$$f(x)=\begin{cases}0, & \text{ha } x<10,\\ \dfrac{1}{90}, & \text{ha } 10<x<100,\\ 0, & \text{ha } x>100;\end{cases}$$

és eloszlásfüggvénye (35. ábra):

$$F(x)=P(X<x)=\begin{cases}0, & \text{ha } x\le 10,\\ \dfrac{x-10}{90}, & \text{ha } 10<x<100\\ 1, & \text{ha } x\ge 100.\end{cases}$$



35. ábra. f(x) sűrűségfüggvény és F(x) eloszlásfüggvény

Az X egyenletes eloszlású valószínűségi változó várható értéke:

$$\mu = \frac{a+b}{2} = \frac{10+100}{2} = 55\text{mp};$$

szórása:

$$\sigma = \frac{b-a}{\sqrt{12}} = \frac{100-10}{\sqrt{12}} = \frac{90}{2\sqrt{3}} = \frac{45\cdot\sqrt{3}}{3} \approx 25{,}98 \approx 26\text{mp};$$

szórásnégyzete:

$$\sigma^2 = \frac{(b-a)^2}{12} = \frac{90^2}{12} = 675.$$

Annak valószínűsége, hogy legalább 50 mp-ig kell várnunk a kapcsolásra:

$$P(X \ge 50) = \frac{100-50}{100-10} = \frac{50}{90} \approx 0{,}56.$$

4.2.2. Exponenciális eloszlás

Mi a valószínűsége annak, hogy egy alkatrész pl. 2000 órán belül nem hibásodik meg?

Definíció. Az X folytonos valószínűségi változót λ **paraméterű exponenciális eloszlásúnak** nevezzük, ha **sűrűségfüggvénye**

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{ha } x > 0; \\ 0, & \text{ha } x \le 0 \end{cases} \quad x \in R \tag{1}$$

ahol a λ állandó tetszőleges pozitív szám, az eloszlás paramétere (36. ábra).

A feltételeknek eleget tesz az f, mivel $f \ge 0$ és

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^{\infty} \lambda e^{-\lambda x}dx =$$

$$= \lim_{C\to\infty} \lambda \int_0^C e^{-\lambda x}dx = \lim_{C\to\infty} \left[-e^{-\lambda x}\right]_0^C = 0+1 = 1.$$

Az (1) sűrűségfüggvényű X valószínűségi változó **eloszlásfüggvénye** (37. ábra):

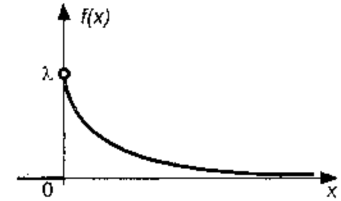
$$F(X) = P(X < x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 1-e^{-\lambda x}, & \text{ha } x > 0; \\ 0, & \text{ha } x \le 0. \end{cases} \quad x \in R \tag{2}$$

és a megfelelő definíciók alapján, egyszerű integrálás és határértékszámítás alapján az exponenciális eloszlású X változó

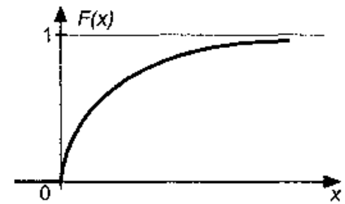
várható értéke: $\mu = M(X) = \dfrac{1}{\lambda}$;

szórásnégyzete: $\sigma^2 = \text{Var}X = D^2(X) = \dfrac{1}{\lambda^2}$;

szórása: $\sigma = \sqrt{\text{Var}X} = D(X) = \dfrac{1}{\lambda}$.



36. ábra. X exponenciális eloszlású változó sűrűségfüggvénye



37. ábra. X eloszlásfüggvénye

Mint látható az exponenciális eloszlású valószínűségi változó várható értéke és szórása egyenlő egymással.

A *medián* definíciója értelmében egyszerű számítással kapjuk a medián értékét:

$$m_e = M(X)\cdot \ln 2 = \frac{\ln 2}{\lambda},$$

amiből következik, hogy az X változó a várható értékénél nagyobb értékeit kisebb valószínűséggel veszi fel, mint a várható értékénél kisebb értékeket, minthogy $m_e < M(X)$.

Megjegyzés

Exponenciális eloszlású pl. a radioaktív atomok élettartama, azaz keletkezésüktől az elbomlásig terjedő időszakasz hossza, hosszú öregedési időtartamú berendezések, alkatrészek meghibásodásai nagy időintervallum alatt stb.

Példa

Legyen az X valószínűségi változó bizonyos típusú alkatrészek meghibásodásáig eltelt használati időtartam hossza. Legyen X exponenciális eloszlású, amelynek szórása 500 óra. Határozzuk meg az X

a) várható értékét;

b) sűrűség- és eloszlásfüggvényét és

c) annak valószínűségét, hogy egy kiszemelt alkatrész 2000 órán belül még nem hibásodik meg.

Megoldás

a) Mivel X szórása 500, ezért várható értéke

$$\sigma = D(X) = M(X) = 500 \text{ óra},$$

és így az eloszlás paramétere: $\lambda = \frac{1}{500}$.

b) A sűrűségfüggvény: $f(x) = \begin{cases} \frac{1}{500} e^{-\frac{x}{500}}, & \text{ha } x > 0; \\ 0, & \text{ha } x < 0, \end{cases}$

az eloszlásfüggvény: $F(x) = \begin{cases} 1 - e^{-\frac{x}{500}}, & \text{ha } x > 0; \\ 0, & \text{ha } x \le 0. \end{cases}$

c) Annak valószínűsége, hogy egy alkatrész 2000 óráig nem hibásodott meg:

$$P(X \ge 2000) = 1 - P(X < 2000) = 1 - F(2000) =$$

$$= 1 - \left(1 - e^{-\frac{2000}{500}}\right) = e^{-4} \approx 0{,}0183 \approx 0{,}02.$$

Tehát annak valószínűsége, hogy 2000 órán belül bekövetkezik az alkatrész meghibásodása 98%.

4.2.3. Normális eloszlás

A leggyakrabban előforduló folytonos eloszlás az ún. *normális* vagy *Gauss-eloszlás*, amely sok jelenség leírásában jelentős szerepet játszik. (*C. F. Gauss* 1777–1855, német matematikus.)

Definíció. Egy X folytonos valószínűségi változót m **és** σ **paraméterű normális eloszlásúnak** nevezünk, ha **sűrűségfüggvénye** (38. ábra):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad (-\infty < x < +\infty) \tag{1}$$

ahol m tetszőleges valós szám, σ pedig tetszőleges pozitív szám lehet.

Az (1) sűrűségfüggvény $-\infty$-től x-ig vett határozott integrálja adja az X változó m, σ **paraméterű normális eloszlásfüggvényét** (39. ábra), azaz

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

Az X normális eloszlású valószínűségi változó

várható értéke: $\mu = M(X) = m$;

szórásnégyzete: $D^2(X) = \sigma^2$;

szórása: $D(X) = \sigma$,

melyek a definíciók felhasználásával igazolhatók. A feltételekből következik, hogy $f(x) > 0$, és $-\infty$-től $+\infty$-ig integrálva 1-et ad, azaz

$$\int_{-\infty}^{+\infty} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = 1,$$

mely $t = \frac{x-m}{\sigma}$ helyettesítéssel

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt$$

integrálba megy át.

Az $M(X)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} x\cdot e^{-\frac{(x-m)^2}{2\sigma^2}}dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(\sigma t+m)e^{-\frac{t^2}{2}}dt=$

$$=\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} te^{-\frac{t^2}{2}}dt+m\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}dt.$$

Az első tag integrálja 0-val egyenlő, mivel az integrandusz páratlan függvény, a második tag pedig $m\cdot 1$, tehát

$$M(X)=\frac{\sigma}{\sqrt{2\pi}}\cdot 0+m\cdot 1=m.$$

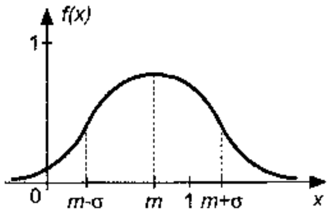
A szórásnégyzethez az

$$M(X^2)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^2\cdot e^{-\frac{(x-m)^2}{2\sigma^2}}dx$$

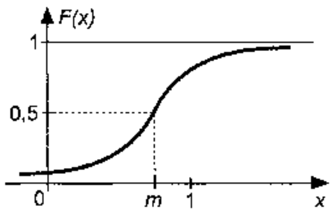
kiszámítását a fentiekhez hasonlóan végezve, kapjuk, hogy

$$M(X^2)=\sigma^2+m^2.$$

Tehát $D^2(X)=M(X^2)-(M(X))^2=\sigma^2+m^2-m^2=\sigma^2.$



38. ábra. Normális eloszlás sűrűségfüggvénye



39. ábra. Normális eloszlásfüggvény

A normális eloszlást követő valószínűségi változók az m és σ paraméterben térnek el egymástól. Az *m várható értékű, σ szórású normális eloszlás szokásos jelölése*: $N(m,\sigma)$, sűrűségfüggvényének gráfja az m várható értékkel adott $x=m$ egyenesre, eloszlásfüggvényének gráfja pedig a $(m,\frac{1}{2})$ pontra szimmetrikus.

Vezessük be az X valószínűségi változó

$$X^*=\frac{X-M(X)}{D(X)}$$

standardizáltját, melynek várható értéke $M(X^*)=0$, és szórása $D(X^*)=1$. Ha az X valószínűségi változó $N(m,\sigma)$ eloszlású, akkor standardizáltját

$$X=\frac{X-m}{\sigma}$$

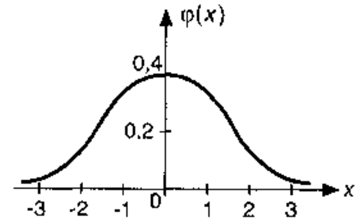
képlettel számítjuk, ugyanis $M(X)=m$, és $D(X)=\sigma$.

Definíció. Az $m=0$, $\sigma=1$ paraméterű normális eloszlást, melyet $N(0,1)$-gyel jelölünk, **standard normális eloszlásnak** nevezzük, *sűrűségfüggvénye* (gráfja szimmetrikus az y tengelyre, 40. ábra):

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\quad (x\in R);\qquad (3)$$

eloszlásfüggvénye (gráfja szimmetrikus a $(0,\frac{1}{2})$ pontra, 41. ábra):

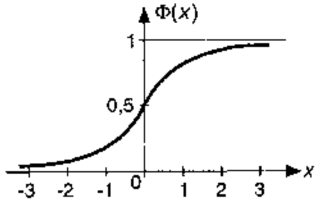
$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt.\qquad (4)$$



40. ábra. Standard normális eloszlás sűrűségfüggvénye

Megjegyzés

$-1\le x\le 1$ intervallumokhoz tartozó terület a görbe alatti terület 68,2 %-a; $-2\le x\le 2$-höz tartozó pedig 95,4 %-a.



41. ábra. Standard normális eloszlás eloszlásfüggvénye

A (4) jobb oldalán álló integrál nem fejezhető ki elemi függvények segítségével, azonban a $\Phi(x)$ értékc a gyakorlati alkalmazások által megkívánt pontossággal táblázatokból kiszámítható. (1. és 2. táblázat).

A standard normális eloszlás (3) és (4) függvényének segítségével kifejezhetők az (1) és (2) függvények:

$$f(x)=\frac{1}{\sigma}\varphi\left(\frac{x-m}{\sigma}\right), \tag{5}$$

$$F(x)=\Phi\left(\frac{x-m}{\sigma}\right), \quad (x\in R). \tag{6}$$

Mivel a normális eloszlás sűrűségfüggvénye szimmetrikus a várható értékre, ezért

$$\varphi(-x)=\varphi(x); \quad \Phi(-x)=1-\Phi(x) \tag{7}$$

továbbá a (7)-ből a standard normális eloszlásra a következő összefüggést kapjuk:

$$P(-x<X\le x)=\Phi(x)-\Phi(-x)=\Phi(x)-(1-\Phi(x))=2\Phi(x)-1 \tag{8}$$

A (7) és (8) összefüggések az 1. és a 2. táblázatból kiolvasható értékek alapján a $\Phi(x)$ és a $\varphi(x)$ függvény értéktáblázatának bővítésére jól felhasználhatók.

A $\varphi(x)$ sűrűségfüggvény $y=\varphi(x)$ gráfja az y tengelyre szimmetrikus, mivel páros függvény (40. ábra). A $\varphi(x)$ *mediánja* 0, mivel az $N(0,1)$ eloszlás szimmetrikus a 0-ra, a páratlan rendű momentumai 0-val egyenlők, így a *ferdeségi együtthatója* is 0, és

mivel 0 a maximum helye ($\varphi(0)=\frac{1}{\sqrt{2\pi}}\approx 0{,}3989$), ezért a *módusza* is 0. Az $y=\varphi(x)$ gráfjának inflexiós pontjai vannak az $x=1$ és $x=-1$ helyeken.

Tekintettel az (5) formulára, az $y=f(x)$ gráfnak $x=m$ egyenes a szimmetriatengelye, és így a maximuma is m-nél van ($f(m)=\frac{\varphi(0)}{\sigma}\approx\frac{1}{\sigma}\cdot 0{,}3989$). Ha tehát az X valószínűségi változó $N(m,\sigma)$ eloszlású, akkor eloszlása szimmetrikus az m-re, *mediánja* is és *módusza* is m-mel egyenlő, minden páratlan rendű centrális momentuma és így *ferdeségi együtthatója* is 0-val egyenlő, inflexiós pontjai pedig az $x=m-\sigma$ és $x=m+\sigma$ helyeken vannak. A normális eloszlás gráfjának *lapultsági együtthatója* 0, mely a 3.3.1. pont (7) lapultsági együttható formulája alapján számítva igazolható.

A φ és f függvényekre elmondottakat felhasználva, valamint Φ és F eloszlásfüggvények definícióira tekintettel, Φ és F gráfjait jellemezhetjük. Az $y=\Phi(x)$ gráfnak $x=0$ helyen inflexiós pontja van, $\Phi(0)=0{,}5$, és így *mediánja* 0. A (6) és (7) formulákra tekintettel, $y=\Phi(x)$ a $(0,\frac{1}{2})$ pontra szimmetrikus, az $y=F(x)$ gráfnak pedig az $x=m$ helyen van inflexiós pontja, $F(m)=\frac{1}{2}$, és az $(m,\frac{1}{2})$ pontra szimmetrikus.

Ha az X valószínűségi változó $N(m,\sigma)$ eloszlású, akkor annak valószínűsége, hogy az m várható értékre szimmetrikus σ szórásával arányos hosszúságú intervallumba esik, csak attól függ, hogy az intervallum hossza a 2σ -nak hányszorosa. Felhasználva a (8) és a (6) formulákat:

$$P(m-k\sigma\le X\le m+k\sigma)=2\Phi(k)-1, \quad (k>0) \tag{9}$$

amiből látható, hogy $N(m,\sigma)$ eloszlású valószínűségi változó eltérése a várható értékétől valóban csak a k-tól függ.

Példák

1. Számítsuk ki, hogy az X valószínűségi változó a várható értékétől legfeljebb mekkora valószínűséggel tér el a szórásának 1-szeresével, 2-szeresével, ill. 3-szorosával.

Megoldás

Felhasználva a (9) formulát és az 1. táblázatban adott $\Phi(t)$ függvény $t=1{,}00$, $t=2{,}00$, $t=3{,}00$ helyekhez tartozó értékeit, kapjuk:

$$P(m-\sigma \le X \le m+\sigma) = 2\cdot\Phi(1)-1 = 2\cdot 0{,}8413-1 = 0{,}6826,$$

$$P(m-2\sigma \le X \le m+2\sigma) = 2\cdot\Phi(2)-1 = 2\cdot 0{,}9772-1 = 0{,}9544,$$

$$P(m-3\sigma \le X \le m+3\sigma) = 2\cdot\Phi(3)-1 = 2\cdot 0{,}9987-1 = 0{,}9974, \qquad (*)$$

Tehát az X változó a várható értékétől *egy szórásnyira* közelítőleg 0,683, *két szórásnyira* 0,954, *három szórásnyira* 0,997 valószínűséggel tér el. A (*) formula az ún. *háromszigma* szabály, mely a gyakorlatban jól használható, hiszen azt fejezi ki pl., hogy az esemény 10000 kísérletből 9974 esetben bekövetkezik. A 3σ-t a *megengedhető legnagyobb hibának* nevezzük.

A műszaki- és természettudományok számos problémájának megoldásához igen gyakran olyan valószínűségi változót rendelhetünk, amelynek eloszlása normális vagy majdnem normális. (Az eltérésből adódó hiba a vizsgált jelenség szempontjából elhanyagolható.) Ilyenek pl. az alkatrészgyártásban jelentkező méretingadozások, távolság és területmérés hibaeloszlása, stb.

2. Egy célgép 0,75 cm várható átmérőjű korongokat készít. Tegyük fel, hogy az X átmérő normális valószínűségi változó, melynek szórása 0,06 cm. Hány százalékos hibával dolgozik a célgép, ha 0,60 cm-nél kisebb, és 0,84 cm-nél nagyobb korongokat tekintünk hibásnak?

Megoldás

A 0,60 cm standard értéke: $\dfrac{0{,}60-0{,}75}{0{,}06} = -2{,}5,$

A 0,84 cm standard értéke: $\dfrac{0{,}84-0{,}75}{0{,}06} = 1{,}5.$

Annak valószínűsége, hogy hibátlan korongok készülnek (az 1. táblázatból vett értékekkel):

$$P(-2{,}5 \le X^* \le 1{,}5) = \Phi(1{,}5)-(1-\Phi(2{,}5)) = 0{,}9332-(1-0{,}9938) =$$

$$= 0{,}9332-0{,}0062 = 0{,}927$$

tehát a hibás korongok gyártásának valószínűsége: $1-0{,}927 = 0{,}073$, vagyis a korongok 7,3%-a a feltételek szerint hibás.

Megjegyzések

1. A binomiális eloszlású változók esetén a

$$P(k) = b(k;n,p) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} p^k q^{n-k}$$

valószínűségek kiszámítása nagy n értékekre igen fáradságos munkát jelent. Ha n nagy, valamint ha sem a p sem a q nem esik közel a 0-hoz, akkor a binomiális eloszlás jól közelíthető az np várható értékű, $\sqrt{npq}$ szórású normális eloszlással. Így pl. ha n számú kísérlet esetén a p eleget tesz a következő egyenlőtlenségnek:

$$\frac{0{,}637}{\sqrt[3]{n}} < p < 1-\frac{0{,}637}{\sqrt[3]{n}},$$

akkor a $b(k;n,p)$ kiszámítását célszerűbb az

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{npq}}\,\varphi\left(\frac{k-np}{\sqrt{npq}}\right) \qquad (10)$$

közelítő képlet alapján végezni, ahol φ a standard normális sűrűségfüggvény, s ennek megfelelően

$$\sum_{k<x} \binom{n}{k} p^k q^{n-k} \approx \Phi\left(\frac{x-np}{\sqrt{npq}}\right) \qquad (11)$$

2. Centrális határeloszlás-tétel

Mint említettük, a természeti jelenségek vizsgálatánál gyakran találkozunk normális eloszlással. Egy véletlen esemény kialakulása általában nagy számú független véletlen hatás eredménye, ezért célszerű annak vizsgálata, hogy miként viselkedik nagy számú független valószínűségi változó összege. Erre a kérdésre a választ az ún. **centrális (központi) határeloszlás-tétel** adja:

Ha az $X_1, X_2, \ldots, X_n, \ldots$ azonos eloszlású, független, véges várható értékű és szórású valószínűségi változók, akkor

$$\lim_{n\to\infty} P\left(\frac{X_1+X_2+\ldots+X_n-nm}{\sigma\sqrt{n}}<x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,\mathrm{d}t=\Phi(x) \quad (12)$$

ahol $m=M(X_k)$, $\sigma=D(X_k)$ $(k=1,2,3,\ldots)$, és $\Phi(x)$ a standard normális eloszlásfüggvény, nm az $X_1+X_2+\ldots+X_n$ összeg várható értéke, $\sigma\sqrt{n}$ pedig az összeg szórása, melynek következtében

$$\frac{X_1+X_2+\ldots+X_n-nm}{\sigma\sqrt{n}}$$

valószínűségi változó várható értéke 0, szórása pedig 1. A centrális határeloszlás-tétel tehát azt mondja ki, hogy *sok független valószínűségi változó összege normális eloszlású*, vagyis (12)

$$\lim_{n\to\infty} P(X_n<x)=\Phi\left(\frac{x-nm}{\sqrt{n}\,\sigma}\right)=F(x) \quad (12^*)$$

alakban is írható.

Példa

Tegyük fel, hogy júliusban a H hőmérséklet normális eloszlású, átlaga 26°C, szórása 4°C. Számítsuk ki annak a valószínűségét, hogy a hőmérséklet 28°C és 34°C közé esik.

Megoldás

A 28°C standard egysége a $t=\dfrac{x-m}{\sigma}$ összefüggés felhasználásával:

$$\frac{28-26}{4}=\frac{2}{4}=\frac{1}{2},$$

és 34°C standard egysége:

$$\frac{34-26}{4}=\frac{8}{4}=2,$$

tehát

$$P(28\le H\le 34)=P\left(\frac{1}{2}\le H^*\le 2\right)=\Phi(2)-\Phi\left(\frac{1}{2}\right)=0{,}9772-0{,}6915=0{,}2857.$$

A hőmérséklet kb. 0,29 valószínűséggel esik 28°C és 34°C közé.

A H^* a H-nak megfelelő standard valószínűségi változót jelöli. A számításhoz felhasználtuk az 1. sz. táblázatot, amelynek 2.00 sorában a $\Phi(x)$ értéke 0,9772, és a 0,50 sorában $\Phi(x)$ értéke 0,6915.

E.4. Ellenőrző kérdések a 4. fejezethez

1. Mikor mondjuk az X valószínűségi változóról, hogy binomiális eloszlású?

2. Mikor nevezzük az X valószínűségi változót *Poisson*-eloszlásúnak?

3. Mikor nevezzük az X valószínűségi változót hipergeometrikus eloszlásúnak?

4. Mi a jellemzője az egyenletes eloszlásnak?

5. Hogyan definiáljuk az exponenciális eloszlású valószínűségi változó sűrűségfüggvényét és eloszlásfüggvényét?

6. Hogyan definiáljuk a normális eloszlású valószínűségi változó sűrűségfüggvényét és eloszlásfüggvényét?

7. Hogyan értelmezzük a standard normális eloszlás sűrűségfüggvényét és eloszlásfüggvényét?

8. Milyen feltételek teljesülése esetén közelíthető a binomiális eloszlás normális eloszlással?

V.4. Feladatok a 4. fejezethez

V.4.1. Számítsuk ki az

a) $b\left(3;6,\frac{1}{2}\right)$ *b)* $b\left(3;4,\frac{1}{4}\right)$ értékeket $(p+q=1)$.

V.4.2. Az A csapat nyerési esélye minden játszmánál a $p=\frac{2}{3}$. Négy játszma lejátszása esetén mi a valószínűsége annak, hogy az A csapat

a) pontosan 2 játszmát nyer;

b) legalább 1 játszmát nyer;

c) a játszmák felénél többet nyer?

V.4.3. Egy üzem által gyártott alkatrészek 2%-a selejtes. A megrendelő 10000 db-os csomagban kapja az alkatrészeket.

a) Mi a selejtes darabok várható értéke és szórása?

b) Hány db-ot kell véletlenszerűen kivenni és megvizsgálni ahhoz, hogy legalább 0,96 valószínűséggel legyen köztük selejtes is? (A kiválasztott darabokat vizsgálat után azonnal visszatesszük.)

V.4.4. Egy lövész $\frac{1}{4}$ valószínűséggel találja el a célpontot.

a) Mi a valószínűsége annak, hogy 7 lövés közül legalább kétszer célba talál?

b) Hány lövést kell leadni ahhoz, hogy a célt $\frac{2}{3}$-nál nagyobb valószínűséggel eltalálja?

V.4.5. Legyen X valószínűségi változó *Poisson*-eloszlású $\lambda = 1{,}8$ paraméterrel.

a) Határozzuk meg $p(k;\lambda)$ értékeit $k = 0,1,2,3,4$ értékekre.

b) Milyen valószínűséggel vesz fel az X a várható értékénél kisebb értéket?

V.4.6. Egy üzem termékei között 2% a selejtesek száma. Mi a valószínűsége annak, hogy egy 100-as mintában 3 db selejtes?

V.4.7. A tv-képcsövek működőképességének időtartam-hossza legyen egy X valószínűségi változó. Legyen X exponenciális eloszlású, amelynek szórása 800 óra. Határozzuk meg az X

a) várható értékét;

b) sűrűség- és eloszlásfüggvényét és

c) annak valószínűségét, hogy egy véletlenül kijelölt képcső 3200 órán belül még nem hibásodik meg.

V.4.8. Számítsuk ki a standard normális eloszlású X valószínűségi változó sűrűségfüggvényének $\varphi(x)$ helyettesítési értékeit, ha

a) $x = 1{,}63$; *b*) $x = -0{,}75$; *c*) $x = -2{,}08$.

V.4.9. Számítsuk ki a standard normális eloszlású X valószínűségi változó

a) $P(0 \le X \le 1{,}42)$; *b*) $P(-1{,}37 \le X \le 2{,}02)$;

c) $P(X \ge 1{,}13)$; *d*) $P(|X| \le \frac{1}{2})$

valószínűségeit az 1. táblázatban adott $\Phi(x)$ függvényértékek segítségével.

V.4.10. Egy évfolyam 400 hallgatójának L magassága legyen normális eloszlású 170 cm-es átlaggal és 16 cm szórással. Hány hallgató

a) tartozik a 166 cm és 182 cm közötti magassági intervallumba;

b) nagyobb vagy éppen 190 cm?

V.4.11. Igazolja, hogy a λ paraméterű exponenciális eloszlású valószínűségi változó várható értéke $\frac{1}{\lambda}$, szórásnégyzete $\frac{1}{\lambda^2}$.

V.4.12. Alma osztályozásánál azt tapasztalták, hogy 100 db közül 95% elsőosztályúnak minősíthető. Visszatevés nélkül is és visszatevéssel is vegyünk 4 elemű mintát. Az X valószínűségi változó jelentse a mintában lévő minőségi előírásoknak nem megfelelők számát. Mindkét mintavételhez adjuk meg X valószínűségeloszlását, várható értékét és szórását.

V.4.13. Egy automata gép által gyártott alkatrészek átmérője normális eloszlású valószínűségi változó, várható értéke 8 mm, szórása 0,1 mm. Az alkatrész nem alkalmazható, ha átmérője a várható értékétől 4%-kal eltér. Mekkora a hibás alkatrészgyártás valószínűsége?

ÖTÖDIK FEJEZET

Milyen következtetést vonhatunk le a nagy számok törvénye alapján a kísérletsorozat eredményére vonatkozólag?

A gyakorlat számos feladata olyan, hogy az ismeretlen eloszlásfüggvényű, tetszőleges eloszlású X valószínűségi változó várható értékét és szórását jó közelítésben meg tudjuk adni, és ezen értékek alapján kívánjuk becsülni a változó megfigyelhető értékei és várható értékük eltérését. Ebben a fejezetben a valószínűség és a relatív gyakoriság közötti összefüggéseket vizsgáljuk. Ilyen problémákkal kapcsolatos első eredmények már *J. Bernoulli* (1654–1705) halála után, 1713-ban megjelent könyvében olvashatók.

5.1. A Csebisev-egyenlőtlenség

1. Tétel. Ha X olyan tetszőleges nemnegatív valószínűségi változó, amelynek van várható értéke és c egy tetszőleges pozitív szám, akkor

$$P(X \geq c) \leq \frac{M(X)}{c}. \tag{1}$$

Ui. ha X diszkrét valószínűségi változó megszámlálhatóan végtelen $x_1, x_2, \ldots$ értékeihez $p_1, p_2, \ldots$ valószínűségek tartoznak, akkor várható értékére a következő becslést kapjuk:

$$M(X) = \sum_{i=1}^{\infty} x_i p_i \geq \sum_{x_i \geq c} x_i p_i \geq \sum_{x_i \geq c} c p_i = c \sum_{x_i \geq c} p_i = cP(X \geq c),$$

amelyből már (1) következik.

Ha pedig X folytonos és sűrűségfüggvénye $f(x)$, akkor az alábbi integrálegyenlőtlenségekből kapható (1):

$$M(X) = \int_0^{+\infty} xf(x)dx \geq \int_c^{+\infty} xf(x)dx \geq \int_c^{+\infty} cf(x)dx = c\int_c^{+\infty} f(x)dx = cP(X \geq c)$$

Az (1) ún. *Markov-egyenlőtlenségből* (*Markov, A. A.* 1856–1922 orosz matematikus) már könnyen előállíthatjuk a *Csebisev*-egyenlőtlenséget (*Csebisev, P. L.* 1821–1894 orosz matematikus).

2. Tétel. Ha az X tetszőleges valószínűségi változónak van várható értéke és szórása, valamint $k>1$ tetszőleges valós szám, akkor

$$P(|X-M(X)|\geq kD(X))\leq \frac{1}{k^2} \tag{2}$$

vagyis annak valószínűsége, hogy az X a várható értékétől abszolút értékben a szórás k-szorosánál többel térjen el, legfeljebb $\frac{1}{k^2}$.

Ui. ha az (1) egyenlőtlenségben az 1. tétel feltételeinek eleget tevő

$$X=(X-M(X))^2$$

és

$$c=k^2D^2(X)$$

helyettesítést elvégezzük, akkor a

$$P\left((X-M(X))^2\geq k^2D^2(X)\right)\leq \frac{M\left((X-M(X))^2\right)}{k^2D^2(X)}=\frac{1}{k^2}$$

egyenlőtlenséget kapjuk. Mivel

$$(X-M(X))^2\geq k^2D^2(X)$$

egyenlőtlenséggel ekvivalens az

$$|X-M(X)|\geq kD(X) \tag{3}$$

egyenlőtlenséggel, így igazoltuk a (2) *Csebisev*-egyenlőtlenséget.

Megjegyzések

1. Ha arra akarunk választ kapni, hogy az X mekkora valószínűséggel esik egy adott intervallum belsejébe, akkor a (2)-ből (3) komplementerének behelyettesítésével kapható

$$P\left(M(X)-kD(X)<X<M(X)+kD(X)\right)=$$
$$=1-P\left(|X-M(X)|\geq kD(X)\right)\geq 1-\frac{1}{k^2} \tag{4}$$

egyenlőtlenséget használjuk. Ennek jelentése: $1-\frac{1}{k^2}$-nél nem kisebb annak valószínűsége, hogy az X a várható értékének $kD(X)$ sugarú környezetébe esik. Pl. a (2) formula értelmében annak valószínűsége, hogy nagyszámú független kísérletet végezve a megfigyelt érték a várható értékétől abszolút értékben szórásának legfeljebb kétszeresével tér el 0,25, vagyis 25%, (4) szerint pedig 0,75 valószínűséggel az $(M(X)-2D(X), M(X)+2D(X))$ intervallumba esik.

2. A *Csebisev*-egyenlőtlenséget néha előnyösebb $kD(X)=\varepsilon>0$ helyettesítéssel

$$P\left(|X-M(X)|\geq \varepsilon\right)\leq \frac{D^2(X)}{\varepsilon^2} \tag{5}$$

alakban használni.

Példa

Legyen egy X pozitív valószínűségi változó várható értéke: $M(X)=8$; szórása: $D(X)=8$. Számítsuk ki, hogy legfeljebb mekkora valószínűséggel vesz fel a változó 52-t vagy annál nagyobb értéket. Mennyi a valószínűség pontos értéke, ha feltesszük, hogy az eloszlás exponenciális?

Megoldás

Az X a várható értéktől $52-8=44$ értékkel tér el, ha eléri az 52 értéket. Negatív értéket nem vesz fel, tehát annak valószínűsége 0, így a Csebisev-egyenlőtlenség alkalmazható az $\varepsilon=44$ értékre:

$$P\left(|\overline{X}_n-8|\geq 44\right)\leq \frac{8^2}{44^2}=0{,}0331.$$

Tehát legfeljebb 0,0331 valószínűséggel veszi fel a változó legalább az 52 értéket.

Feltesszük, hogy X exponenciális eloszlású, így az eloszlásfüggvény értéke:

$$F(X)=1-e^{-\frac{x}{8}}$$

tehát a keresett valószínűség

$$P(X \geq 52) = 1 - F(52) = e^{-\frac{52}{8}} = e^{-6,5} = \frac{1}{e^{6,5}} \approx 0{,}0015.$$

Amint látható, az exponenciális eloszlás feltételezésével számított 0,0015 pontos érték lényegesen kisebb, mint a *Csebisev*-egyenlőtlenséggel kapott 0,0331 felső korlát.

5.2. A nagy számok törvénye

A valószínűségszámításban **nagy számok törvényeinek** nevezzük a tételek olyan csoportját, amelyek a valószínűségi változók és várható értékei közötti sztochasztikus konvergenciára vonatkoznak.

Definíció. Az $X_1, X_2, \ldots$ valószínűségi változók sorozatáról akkor mondjuk, hogy **sztochasztikusan konvergál** az X valószínűségi változóhoz, ha bárhogy választva az $\varepsilon > 0$ számot, az $|X_n - X| > \varepsilon$ egyenlőtlenség teljesülésének valószínűsége 0-hoz tart, ha n minden határon túl nő, azaz

$$\lim_{n \to \infty} P(|X_n - X| > \varepsilon) = 0.$$

3. Tétel. (*Nagy számok gyenge törvénye*):

Legyenek az $X_1, X_2, \ldots, X_n, \ldots$ valószínűségi változók függetlenek; eloszlásuk, m-mel jelölt várható értékük, s^2-tel jelölt szórásnégyzetük azonos, akkor az

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

számtani közép sztochasztikusan konvergál az m várható értékhez, ha n minden határon túl nő, azaz

$$\lim_{n \to \infty} P(|\overline{X}_n - m| \geq \varepsilon) = 0 \quad \text{vagy} \quad \lim_{n \to \infty} P(|\overline{X}_n - m| < \varepsilon) = 1. \tag{1}$$

Ui. $M(\overline{X}_n) = \dfrac{M(X_1) + M(X_2) + \ldots + M(X_n)}{n} = \dfrac{n \cdot m}{n} = m,$

s mivel $X_1, X_2, \ldots, X_n$ független valószínűségi változók, ezért

$$D^2(X_1 + X_2 + \ldots + X_n) =$$
$$= D^2(X_1) + D^2(X_2) + \ldots + D^2(X_n) = n \cdot s^2.$$

Tekintettel arra, hogy

$$D^2(\overline{X}_n) = D^2\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) =$$
$$= \frac{1}{n^2} D^2(X_1 + X_2 + \ldots + X_n) =$$
$$= \frac{1}{n^2}(n \cdot s^2) = \frac{s^2}{n}$$

így a *Csebisev*-egyenlőtlenség (5) alakját használva:

$$P(|\overline{X}_n - m| \geq \varepsilon) \leq \frac{s^2}{n\varepsilon^2}, \tag{2}$$

amelyből $n \to \infty$ esetén, a jobb oldal tart a 0-hoz, így következik a tétel állítása.

Megjegyzések

1. Ha egy kísérletsorozat valamely A eseménye tetszőleges sokszor megismétlődhet, akkor megmutatható, hogy van olyan N szám, amelynél $n \geq N$ számú kísérlet független megismétlésénél 1-hez tetszőleges közel lesz annak a valószínűsége, hogy az esemény $\frac{k}{n}$ relatív gyakorisága és p valószínűségének abszolútértékben vett eltérése kisebb egy előre megadott tetszőleges pozitív ε-nál.

Alkalmazzuk a *Csebisev*-egyenlőtlenséget a binomiális eloszlásra, akkor a nagy számok ún. **Bernoulli-féle törvényét** kapjuk:

$$P\left(\left|\frac{k}{n} - p\right| \geq \varepsilon\right) \leq \frac{pq}{\varepsilon^2 n} \tag{3}$$

ahol $p = P(A)$, $q = P(\overline{A}) = 1 - p$ és $\frac{k}{n}$ az A esemény relatív gyakorisága.

2. A 3. tétel akkor is igaz, ha csak azt a feltételt kötjük ki, hogy létezik az X_i valószínűségi változók $M(X_i)$ $(i = 1,2,\ldots)$ várható értéke.

Példa

Hányszor kell egy kockát feldobni, hogy a 6-os dobás $\frac{1}{6}$ valószínűségét annak relatív gyakorisága legalább 0,75 valószínűséggel 0,15-nál kisebb hibával megközelítse, ha a szórásnégyzet 0,14?

Megoldás

Jelentse A a 6-os dobás eseményét. A (3) képletbe $p = \frac{1}{6}$, $q = 1 - \frac{1}{6} = \frac{5}{6}$ és $pq = \frac{1}{6} \cdot \frac{5}{6} \approx 0{,}14$ helyettesítéssel

$$P\left(\left|\frac{k}{n} - \frac{1}{6}\right| \geq 0{,}15\right) \leq \frac{0{,}14}{0{,}15^2 n}.$$

Az egyenlőtlenség bal oldali valószínűségének 0,15-nál kisebbnek kell lennie, hogy a kérdéses eseménnyel ellentétes esemény legalább 0,85 valószínűséggel bekövetkezzék. Ennek feltétele, hogy

$$\frac{0{,}14}{0{,}15^2 n} < 0{,}25,$$

amelyből kiszámítható az n értéke:

$$n > \frac{0{,}14}{0{,}15^2 \cdot 0{,}25} = 24{,}89 \approx 25.$$

Tehát 25-nél többször kell feldobni a kockát az adott feltételek mellett.

E.5. Ellenőrző kérdések az 5. fejezethez

1. Hogyan becsülhetjük a változó értékei és várható értéke közötti eltérést?

2. Hogyan becsüljük egy esemény relatív gyakorisága és valószínűsége közötti eltérést?

3. Miként értelmezhető a *Csebisev*-egyenlőtlenség és a nagy számok törvénye?

V.5. Feladatok az 5. fejezethez

V.5.1. Számítsuk ki, hogy az X pozitív valószínűségi változó legfeljebb mekkora valószínűséggel vesz fel legalább 70-es értéket, ha $M(X) = 20$ és $D(X) = 20$?

V.5.2. Egy kötélgyártó 35 m hosszú köteleket gyárt 0,3 m szórással. Legfeljebb mennyi annak valószínűsége, hogy a kötél hossza legalább 1 m-rel eltér a várható 35 m-es értéktől?

V.5.3. Egy gumikesztyűgyár termékeinek 10%-a hibás. A vevő csak akkor hajlandó a leszállított tételt átvenni, ha abban legfeljebb 12% hibás. Hány darabos tétel szállítására kössön szerződést a gyár, hogy a hibás kesztyűk relatív gyakorisága a megfelelő valószínűségtől legalább 0,95 valószínűséggel ne térjen el 0,02-nál nagyobb értékkel?

A MATEMATIKAI STATISZTIKA ELEMEI

Statisztikai mintavétel
- Mintaátlag, mintaterjedelem
- Tapasztalati eloszlás- és sűrűségfüggvény

Statisztikai becslések
- A pontbecslés módszere
- Konfidencia-intervallum

Statisztikai hipotézisek vizsgálata
- Egy- és kétmintás *u*- és *t* próba
- A Welch-próba és *F*-próba
- Illeszkedés- és homogenitásvizsgálat
- Függetlenségvizsgálat khi-négyzet-próba

Empirikus képletek előállítása
- Kiválasztott pontok módszere
- Közepek módszere
- Legkisebb négyzetek módszere

Korreláció- és regressziószámítás

Fejezetenként ellenőrző kérdésekkel és feladatokkal

II. RÉSZ

A MATEMATIKAI STATISZTIKA ELEMEI

ELSŐ FEJEZET

1.1. Bevezetés

Milyen tárgykörök tartoznak a matematikai statisztikához?

A statisztika a tömegjelenségeknél észlelhető tapasztalati törvények empirikus mérések általi feltárásával foglalkozó tudomány. Számos természettudománynak (fizika, biológia, stb.) és társadalomtudománynak (gazdaságtudományok, demográfia, stb.) nélkülözhetetlen segédeszköze, amely elsősorban a valószínűségszámítás és a matematikai statisztika eredményeire és módszereire támaszkodik.

A matematikai statisztika a valószínűségszámítás egy önálló fejezete, amely a megfigyelések és mérések eredményeiből az ún. **statisztikai adatokból** következtet események ismeretlen valószínűségeire vagy valószínűségi változók ismeretlen eloszlásfüggvényeire és ezek paramétereire. Következtetései ún. **valószínűségi ítéletek,** amelyeknek a bizonytalanságaiból fakadó hatásokat is számításba tudjuk venni. A matematikai statisztika feladata egyrészt – az előbbiekben említett problémák kezeléséhez – olyan módszerek kidolgozása, amelyekkel a jelenségek megfigyeléséből, mérések útján előállított tapasztalati adatokból a keresett elméleti értékekre, az eloszlásfüggvények paramétereire (várható értékére, szórására) a lehető legtöbb információt nyerhetjük, másrészt az adatokat szolgáltató kísérletek optimális tervezése.

Például egy darabológépet adott hosszúságú pálcikák előállítására állítottunk be. A pálcikák, a gép fizikai állapotától, a levegő hőmérsékletétől, stb.-től függően, a pontos mérettől eltérő – hosszabb és rövidebb – méretűek is lehetnek. Tapasztalatból tudjuk, hogy a pálcikák X hossza normális eloszlású valószínűségi változó. Véletlenszerűen kiválasztunk n pálcikát, melyeket lemérve $x_1, x_2, \ldots, x_n$ méreteket kapunk. A kapott mérési eredményekből statisztikai módszerekkel kiszámítjuk közelítően a várható értéket és a szórást, és a továbbiakban a tervezéshez az így meghatározott normális eloszlású valószínűségi változóval végezhetjük számításainkat. Arra a kérdésre, hogy egy ismert eloszlású X valószínűségi változó milyen valószínűséggel esik egy adott intervallumba, a valószínűségszámítás témakörében olyan formulával adhattunk választ, amelyben a várható érték és a szórás is ismert volt. A matematikai statisztikában az ilyen típusú kérdésre csak akkor adhatunk választ, ha előbb az adatok alapján a várható értéket és a szórást közelítően meghatározzuk, értékeiket megbecsüljük.

A matematikai statisztika modern elmélete, bár egyes módszerei régebbi keletűek, csak a valószínűségszámítás *Kolmogorov*-féle megalapozása óta alakult ki. Alapjait a 18. és a 19. században rakták le. Így pl. *T. Bayes* (1702–1761) módszert dolgozott ki az eloszlások meghatározására, *P. S. Laplace* (1749–1827), *K. F. Gauss* (1777–1855) és *A. Legendre* (1752–1833) a becsléselmélet megalapozását végezték el azzal, hogy kidolgozták a hibaszámításhoz a legkisebb négyzetek módszerét. A demográfiában és az ipari minőségellenőrzésben *M. V. Osztrogradszkij* (1801–1862) alkalmazta a matematikai statisztikát, melynek igen gyors fejlődését *P. L. Csebisev* (1821–1894), *A. A. Markov* (1856–1922), *A. M. Ljapunov* (1857–1918) és *A. Quetelet* (1796–1874) munkássága segítette elő.

A 20. században pedig *K. Pearson* (1857–1936), *Jordán Károly* (1871–1959), *R. A. Fisher* (1890–1962), *Student* (eredeti neve: *W. S. Gosset* (1876–1937), *J. Neyman, E. S. Pearson, M. D. Kendall, V. I. Glivenko, A. N. Kolmogorov, A. J. Hincsin, N. V. Szmirnov* és *B. V. Gnyegyenko* kutatásai, valamint az alkalmazásban elért eredményei hozták létre a matematikai statisztika modern elméletét, amelyet a magyar származású *Wald Ábrahám* a döntésfüggvények elméletében egyesít és általánosít.

A matematikai statisztika elmélete és alkalmazása szempontjából jól körülhatárolható főbb fejezetei: a mintavétel elmélete, a becsléselmélet, a hipotézisvizsgálat, a korreláció- és regressziószámítás, a szóráselemzés, a faktoranalízis, a kísérletek tervezése és a hibaszámítás. Ebben a II. részben az első négy tárgykörrel foglalkozunk. Azok számára, akik a többi tárgykörrel is meg akarnak ismerkedni, az irodalomjegyzékben felsorolt műveket ajánljuk.

1.2. Statisztikai mintavétel

Hogyan válasszunk mintát egy alapsokaságból?

A továbbiakban **statisztikai sokaságnak** nevezzük az elemek (egyedek) olyan halmazát, amelyeknek tulajdonságait a matematikai statisztika fogalmaival és módszereivel jól jellemezhetjük. Statisztikai sokaságot alkotnak pl. népcsoportok egyedeinek összessége, különböző ágazatok által előállított termékek, a mezőgazdaság állatállománya.

A statisztikai sokaság egészének vizsgálata gyakran kivihetetlen vagy csak igen nagy fáradsággal és költséggel valósítható meg, ezért a vizsgálat céljára kiválasztjuk egy részét, amelyet **statisztikai mintának** nevezünk. A **mintavétel** azt jelenti, hogy a statisztikai sokaságból véletlenszerűen többször kiválasztunk bizonyos számú elemet. Az elemek független kísérlet vagy megfigyelés eredményei, azonos eloszlású független valószínűségi változók. Pl. egy $F(x)$ eloszlásfüggvényű változóra vonatkozóan n mérést végzünk, melynek eredményeként $x_1, x_2, \ldots, x_n$ értékeket kapjuk. Többször megismételve a mérést, az $x_{11}, x_{21}, \ldots, x_{n1}$, $x_{12}, x_{22}, \ldots, x_{n2}, \ldots$ értékek általában különböznek egymástól. Jelöljük a sokaságból kiválasztott n elemű mintát $X_1, X_2, \ldots, X_n$-nel. Az egyes változók, az egyes megfigyelések eredményei, X-szel azonos eloszlásúak és egymástól függetlenek. Így pl. a folyamatosan gyártott termékek összességét, a gyártmánysokaságot (az alapsokaságot) minősítjük a bizonyos számú véletlenszerűen kiválasztott termék, a minta minősége alapján. Általában **alapsokaságnak** nevezzük az egyedeknek azt a halmazát, amelyből a mintavétel során a mintát vesszük.

A statisztikai mintavétellel szemben támasztott alapvető követelmény, hogy az **reprezentatív mintavétel** legyen. Általános értelemben reprezentatív a **véletlen mintavétel,** ha minden lehetséges mintának egyenlő valószínűsége van a kiválasztásra. Az alapsokaságból kiválasztott $X_1, X_2, ..., X_n$ mintaelemekkel szemben fontos követelmény, hogy hűen tükrözze azt a sokaságot, amelyből való és a lehető legtöbb információt nyújtsa az ismeretlen eloszlásról. Ez elérhető, ha a mintaelemek eloszlása azonos és az alapsokaságéval is megegyező, továbbá, ha a mintaelemek összességükben független valószínűségi változók. Az első követelmény azt jelenti, hogy

$$P(X_i < x) = F(x)$$

feltételnek i-től függetlenül kell teljesülni. Ekkor mondhatjuk azt, hogy a véletlenszerűen kiválasztott mintaelemek **reprezentálják** az alapsokaságot. A második követelmény teljesülése biztosítja, hogy elegendő információt kapjunk a vizsgált valószínűségi változó eloszlására és egyes paramétereire. A mintaelemek függetlenségének és az azonos eloszlásának biztosítására a kísérleteket, megfigyeléseket egymástól függetlenül kell végezni.

A matematikai statisztikában *mintának* nevezzük azoknak a számértékeknek a sorozatát is, amelyet pl. a minta elemein végzett mérések útján nyerünk. Például, ha valamely E esemény valószínűségét akarjuk meghatározni, akkor az eseményhez rendelt X_i változó értéke 1 vagy 0 lehet, attól függően, hogy az i-edik kísérletben az E vagy $\overline{E}$ következik be. Az $X_1, X_2, ..., X_n$ mintaelemek a véletlen eseményre vonatkozó megfigyelési értékek, amelyek a kísérlet befejezése után n számú rögzített adatot jelentenek. Így pl. a minta elemei lehetnek termésbecslés esetén a véletlenszerűen kiválasztott n számú adott nagyságú terület terméseredményei; n db véletlenszerűen kiválasztott gépalkatrész hosszára, súlyára, stb.-re kapott számértékek; fizikai, kémiai kísérletek anyagaira vonatkozó független mérések eredményei, stb.

Pl. a Tisza ólomtartalmának megállapításához a szennyezési időszakban naponta többször bizonyos mennyiségű vizet kimernek és laboratóriumban meghatározzák ólomtartalmát. Jelölje az X valószínűségi változó a vízadagonként mért ólomtartalmat, ekkor az n-szer kimert vízadag méréssel kapott ólomtartalom értékei legyenek $x_1, x_2, \ldots, x_n$, és ezzel egy statisztikai mintát állítottunk elő. Többször megmérve n vízadagot, azt tapasztaljuk, hogy a minta elemei más-más értékeket vesznek fel. Az előállított statisztikai mintából következtetünk az X valószínűségi változó eloszlására, várható értékére, szórására.

Gyakran az a követelmény, hogy a statisztikai mintavétel csak bizonyos jellemzők szempontjából legyen reprezentatív, ilyenkor ún. **rétegezett mintavételt** alkalmazunk. Ez olyan mintavétel, amelynél az alapsokaságot a mintavételt megelőzően bizonyos rétegekre (részekre) bontjuk, és az egyes rétegekből vesszük ki a minta meghatározott (általában arányos) részeit. Pl. az ország lakosságát nemek szerint két rétegre, vagy kor szerint több rétegre osztjuk és az egyes rétegekből a részmintákat véletlen mintavétellel választjuk ki.

A véletlen mintavétel legmegbízhatóbb módja a *véletlen számtáblázat* alapján való kiválasztás. A mintavétel alkalmával az alapsokaságnak azokat az elemeit választjuk ki, amelyek sorszámait a véletlen számtáblázatból kiolvastuk. Előfordulhat, hogy az alapsokaság elemeinek elhelyezkedése véletlenszerűnek tekinthető, ilyenkor a **mechanikus (szisztematikus) mintavétel** is alkalmazható, pl. az elemek közül minden 30-adikat választjuk ki.

A matematikai tárgyalás szempontjából a statisztikai mintavételből származó minta elemeit **független, egyenlő eloszlású valószínűségi változóknak** tekintjük. (Ha az alapsokaság elég nagy és a mintavétel reprezentatív vagy az alapsokaság elemei is független, egyenlő eloszlású valószínűségi változóknak tekinthetők, akkor az előbbi feltevés a valóságot jól megközelíti.)

1.2.1. A statisztikai minta jellemzői

Legyen az X valószínűségi változó eloszlásfüggvénye $F(x)$. Tekintsük az X-re vonatkozó

$$X_1,\ X_2,\ ...,\ X_n \tag{1}$$

n elemű mintát. Mivel a minta elemeinek kiválasztása véletlenszerűen történik, tehát – mint említettük – azok is valószínűségi változók. Pl. az X valószínűségi változóra vonatkozó n-elemű méréssorozat n-szeri elvégzésével kapott

$$X_{i1}, X_{i2}, \ldots, X_{in} \quad (i = 1, 2, \ldots, n)$$

eredménysorozat általában nem azonos. Így nyilvánvaló, hogy az

$X_1, X_2, \ldots, X_n$ mintaelemek valószínűségi változók, X-szel azonos eloszlásúak és egymástól függetlenek. Mint a bevezetőben említettük a matematikai statisztika feladata, hogy a mintából következtessen az alapsokaság (elméleti) eloszlására, sűrűségfüggvényére és azok paramétereire.

Az X változó (1) mintaelemei meghatározzák a *tapasztalati-(empirikus)* vagy *mintaeloszlást*, amely a *tapasztalati eloszlásfüggvénnyel* vagy a *gyakorisági eloszlással* jellemezhető. A minta elemeiből kiszámított értékek a *statisztikai függvények*, röviden *statisztikák*, amelyek közül a legfontosabbak a *mintaközép* vagy *számtani közép*, a *medián*, a *mintaterjedelem* és a *tapasztalati szórás*. A mintaelemekből alkotott statisztikák alapján tudunk jó információkat szerezni az eloszlás elméleti jellemzőire.

Definíció. Az (1) n elemű minta $\overline{X}$ **mintaközepének, mintaátlagának** vagy **számtani közepének (empirikus várható értéknek)** az

$$\overline{X}=\frac{X_1+X_2+\ldots+X_n}{n}=\frac{\sum_{i=1}^{n}X_i}{n} \tag{2}$$

képlettel meghatározott számértéket nevezzük.

A mintaátlag várható értéke az egész sokaság várható értéke, azaz

$$M(\overline{X})=M(X).$$

Az egyes mintaelemek $X_i-\overline{X}$ *eltérése* $(i=1,2,\ldots,n)$ összegezve mindig zérus, azaz

$$\sum_{i=1}^{n}\left(X_i-\overline{X}\right)=0. \tag{3}$$

Ha az $F(x)$ eloszlásfüggvényről feltesszük, hogy folytonos, akkor 1 a valószínűsége annak, hogy a mintaelemek között két egyenlő érték nem fordul elő. Ekkor a minta elemeit *nagyság szerint rendezhetjük* (jelölje ezeket X_i^*):

$$X_1^*<X_2^*<\ldots<X_n^*.$$

Az X_i^* az i-edik rendezett mintaelem szintén statisztika, mint-hogy az n számú véletlen mintaelem függvénye.

Definíció. Az

$$X_1^*, X_2^*, \ldots, X_n^* \tag{4}$$

rendezett mintaelemek közül a **medián**:

$$X_{m+1}^*,$$

ha $n=2m+1$ páratlan szám (a (4) középső eleme) és

$$\frac{X_m^*+X_{m+1}^*}{2}, \tag{5}$$

ha $n=2m$ páros szám.

Definíció. A (4) rendezett minta legnagyobb és legkisebb elemének R különbsége a **mintaterjedelem (empirikus terjedelem)**

$$\mathrm{R}=X_n^*-X_1^*. \tag{6}$$

Definíció. Az $X_1^*, X_2^*, \ldots, X_n^*$ véletlen minta $F_n(x)$ **tapasztalati (empirikus) eloszlásfüggvényét** az

$$F_n(x)=\begin{cases}0, \text{ ha } x\le X_1^*\\ \dfrac{k}{n}, \text{ ha } X_k^*<x\le X_{k+1}^*, \ (k=1,2,\ldots,n-1)\\ 1, \text{ ha } X_n^*<x\end{cases} \tag{7}$$

képlettel határozzuk meg, ahol k jelenti az x-néi kisebb mintaelemek számát, $\frac{k}{n}$ pedig a relatív gyakoriságot.

Az $0\le F_n(x)\le 1$ tapasztalati eloszlásfüggvény monoton, nem csökkenő, balról folytonos lépcsős függvény, a mintaelemek által megszabott ugráspontokkal. Azt is mondjuk, hogy az $F_n(x)$ lépcsős függvény minden x helyen felvett értéke az $X<x$ eseménynek a mintából számított relatív gyakoriságával egyenlő.

Megjegyzés

Igazolható a következő állítás: ha a minta elemei függetlenek és azonos eloszlásúak, akkor a minta tapasztalati eloszlásfüggvénye 1 valószínűséggel egyenletesen konvergál a megfelelő elméleti eloszlásfüggvényhez.

Ha a mintaelemek n száma elég nagy, pl. $n=100$, akkor olyan eloszlásfüggvényt kapunk, amelynek ugrásai kicsik $\left(\frac{1}{100}\right)$ és így ez a gyakorlat számára jól használható eloszlásfüggvény.

Különösen nagy minták esetén a (7) tapasztalati eloszlásfüggvény helyett célszerűbb az

$$F_n^{(r)}(x)=\begin{cases} 0, & \text{ha } x \le x_0 \\ \sum_{i=1}^{k} g_i, & \text{ha } x_{k-1} < x \le x_k, \quad (k=1,2,\dots,r) \\ 1, & \text{ha } x_r < x \end{cases} \tag{8}$$

ún. **közelítő tapasztalati eloszlásfüggvényt** használni, amelynél az ugrás nagysága $\frac{1}{n}$ helyett a $g_i=\frac{f_i}{n}$ relatív gyakoriság, ahol f_i az $(x_{i-1};x_i]$ intervallumba eső mintaelemek száma. A (8) előállítása a következő lépésekből áll:

a) Az $X_1, X_2, \dots, X_n$ mintából kikeressük a legkisebb $X_1^*=a$, és a legnagyobb $X_n^*=b$ elemet, és $x_0=a,\ x_1,\ x_2,\dots,x_r=b$ osztópontokkal az (a,b) intervallumot r egyenlő (vagy nem egyenlő) részre osztjuk, ún. **osztályközöket** képezünk. Pl. az osztályközök r számát $\sqrt{n}$-hez legközelebb eső egész számmal adjuk meg, de szükség szerint változtathatjuk az osztályközök számát, hosszát. Fontos, hogy minden osztályközhöz tartozzék mintaelem.

b) Megállapítjuk az egyes intervallumokba eső mintaelemek számát, azaz az egyes intervallumok gyakoriságát, az $f_1, f_2, \dots, f_r$ értékeket, amelyek összege $f_1+f_2+\dots+f_r=n$;

c) Kiszámítjuk az intervallumok

$$g_i=\frac{f_i}{n}, \quad (i=1,2,\dots,r;\ g_1+g_2+\dots+g_r=1)$$

relatív gyakoriságait;

d) Az $F_n^{(r)}(x)$ közelítő tapasztalati eloszlásfüggvényt koordinátarendszerben ábrázoljuk. (L. a 42. és a 46. ábrát.)

A valószínűségi eloszlást is szemléltethetjük, ha a koordinátarendszerben minden $(x_{i-1};x_i)$ intervallum fölé f_i-vel arányos magasságú paralelogrammát rajzolunk. Ha a magasságot éppen g_i-nek választjuk, akkor az ún. **relatív gyakoriság hisztogramját** kapjuk.

Ha a koordinátarendszerben minden $(x_{i-1};x_i)$ részintervallum fölé

$$\frac{g_i}{x_i-x_{i-1}}, \quad (i=1,2,\dots,r) \tag{9}$$

magasságú téglalapot szerkesztünk, akkor az ún. **sűrűséghisztogramot** kapunk. A téglalapok területösszege 1:

$$\sum_{i=1}^{r}\frac{g_i}{x_i-x_{i-1}}(x_i-x_{i-1})=\sum_{i=1}^{r} g_i=1.$$

Ha téglalap helyett csak az x tengellyel párhuzamos egyenes szakaszokkal készítjük el a gráfot, akkor az $f(x)$ elméleti sűrűségfüggvényt közelítő $f_r(x)$ **tapasztalati sűrűségfüggvényt** kapjuk. Az így kapott ábra jellegzetes pontjait egy folytonos vonallal, ún. *burkoló* görbével összekötve, az $f(x)$ elméleti sűrűségfüggvény jellegére következtethetünk (44. ábra).

Példa

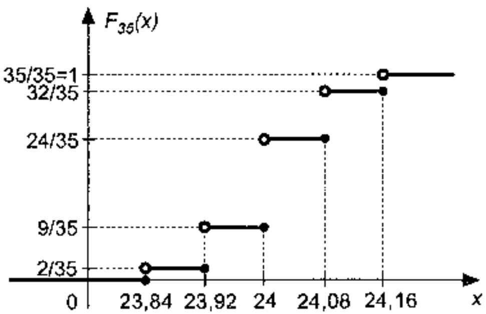
Egy présgépen $d=24$ mm átmérőjű korongokat kell gyártani. A korongok véletlenszerű fizikai hatások következtében d értékétől eltérő átmérőjűek is lehetnek. Az elkészült korongok közül 35 db-os mintát lemérve 23,80 mm és 24,20 mm közötti értékeket kaptunk. 0,08 mm-es osztályközöket képezve megállapítottuk a gyakoriságot, majd kiszámítottuk a relatív gyakoriságokat, azok összegét és a sűrűséghisztogramhoz a téglalapok magasságát. Ábrázoljuk az empirikus eloszlásfüggvényt, a sűrűséghisztogramot és a közelítő tapasztalati sűrűségfüggvényt a táblázatba foglalt adatok alapján.

i	Osztályközök: 0,08 mm	Gyakoriság: f_i	Relatív gyakoriság: g_i	$\sum_i g_i$	$\frac{g_i}{0,08}$
1	23,80-23,88	2	2/35	2/35	0,7
2	23,88-23,96	7	7/35	9/35	2,5
3	23,96-24,04	15	15/35	24/35	5,36
4	24,04-24,12	8	8/35	32/35	2,86
5	24,12-24,20	3	3/35	35/35	1,07
		$\sum_i f_i = 35 = n$	$\sum_{i=1}^{5} g_i = \frac{35}{35} = 1$		

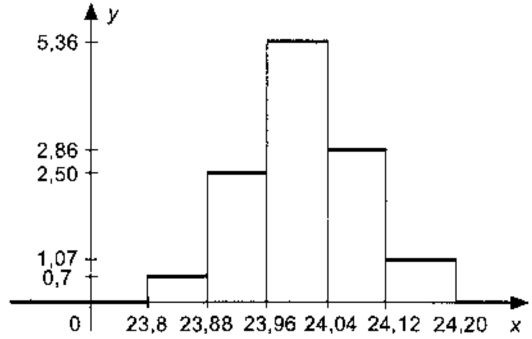
Megoldás

Az $F_{35}(x)$ empirikus eloszlásfüggvény 23,80-nál kisebb x-ekre 0-val egyenlő, majd 0,08-os intervallumonként x tengelytől $\sum_i g_i$ távolságú, az x tengellyel párhuzamos egyenes szakaszokkal alkotott, balról folytonos lépcsősfüggvény (intervallum felosztásnak az osztályközepeket vettük, 42. ábra).

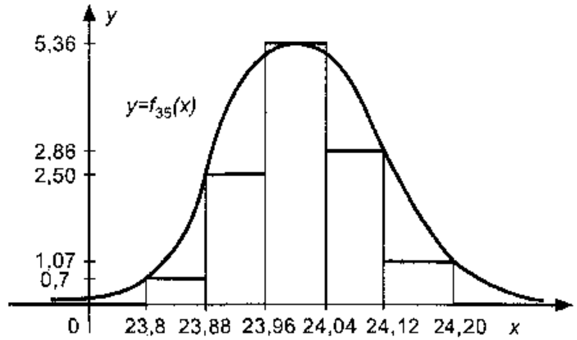
A sűrűséghisztogramot intervallumonként az x tengelyre merőlegesen álló $\frac{g_i}{0,08}$ magasságú téglalapok alkotják (43. ábra). Az $f_{35}(x)$ tapasztalati sűrűségfüggvény és a közelítő burkológörbe normális eloszlást mutat (44. ábra).



42. *ábra. Az $F_{35}(x)$ tapasztalati eloszlásfüggvény*



43. *ábra. Sűrűséghisztogram*



44. *ábra. Az $f_{35}(x)$ tapasztalati sűrűségfüggvény és burkolója*

A minta jellemzésére fontos mutató a tapasztalati (empirikus) vagy mintabeli szórásnégyzet és szórás, valamint a tapasztalati momentum.

A **tapasztalati szórásnégyzet** s^2, a mintaelemek mintaközéptől való eltérései négyzetének számtani közepe, azaz

$$s^2 = \frac{(X_1-\overline{X})^2+(X_2-\overline{X})^2+\ldots+(X_n-\overline{X})^2}{n} = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n} \qquad (10)$$

A (10) helyett kis n számú minta esetén $(n \le 10)$, az ún. **korrigált tapasztalati szórásnégyzetet** használjuk:

$$s^{*2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}. \tag{11}$$

A tapasztalati szórásnégyzet is, és a korrigált tapasztalati szórásnégyzet is véletlentől függő változó. Bizonyítható, hogy a korrigált tapasztalati szórásnégyzet várható értéke az X valószínűségi változó szórásnégyzete, azaz

$$M(s^{*2}) = \sigma^2.$$

Nagy mintaelemszám esetén az s^2 és s^{*2} közötti eltérés elhanyagolható. A tapasztalati szórást az

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}}, \tag{12}$$

a korrigált tapasztalati szórást pedig

$$s^* = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}} \tag{13}$$

képlettel számítjuk ki.

Megjegyzés

1. Osztályközök képzésekor, ha a minta elemei f_i gyakorisággal adottak az osztályközökben, akkor a **mintaátlagot**

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i \cdot x_i}{n}, \tag{14}$$

a tapasztalati szórást, ill. a korrigált tapasztalati szórást

$$s = \sqrt{\frac{\sum_{i=1}^{k} f_i \cdot (x_i - \overline{X})^2}{n}}, \quad \text{ill.} \quad s^* = \sqrt{\frac{\sum_{i=1}^{k} f_i \cdot (x_i - \overline{X})^2}{n-1}} \tag{15}$$

képlettel számítjuk, ahol x_i az egyes osztályközök osztályközepe, k az osztályközepek (ill. a ténylegesen különböző x_i értékek) száma és

$$n = \sum_{i=1}^{k} f_i.$$

2. A tapasztalati szórás és a mintaközép hányadosát, az $r_{sz} = \frac{s}{\overline{X}}$, $\overline{X} > 0$ statisztikát, **relatív szórásnak** *(variációs tényezőnek)* nevezzük.

3. Az (1) mintához tartozó ***k*-adik tapasztalati momentum,** m_k kiszámítását az

$$m_k = \frac{1}{n}(X_1^k + X_2^k + \ldots + X_n^k), \quad (k = 2, 3, \ldots) \tag{16}$$

ill. a μ_k *centrális momentum* kiszámítását

$$\mu_k = \frac{\sum_{i=1}^{n} f_i (x_i - \overline{X})^k}{n}$$

képlettel végezzük. Ha a *d osztályköz nem egységnyi,* a (16) helyett a

$$\mu_{k_d} = \frac{\sum_{i=1}^{n} f_i \cdot \left(\frac{x_i - \overline{X}}{d}\right)^k}{n} \tag{17}$$

formulát használjuk a k-adik centrális momentum kiszámítására. A

$$\gamma_1 = \frac{\mu_3}{s^3} = \frac{\sum_{i=1}^{n} f_i (x_i - \overline{X})^3}{n \cdot s^3}$$

hányadost, azaz a harmadik centrális momentum és az empirikus szórás harmadik hatványának hányadosát **empirikus ferdeségi együtthatónak,** a

$$\gamma_2 = \frac{\mu_4}{s^4} - 3 = \frac{\sum_{i=1}^{n} f_i (x_i - \overline{X})^4}{n \cdot s^4} - 3$$

formulát pedig a **lapultság empirikus mérőszámának** nevezzük (l. a 3.3.1. pontot).

A rendezett minta *p-edrendű kvantilise* az a mintaelem, amelynél kisebb a mintaelemek $100p\%$-a $(0 < p < 1)$, vagyis ez az elem a rendezett minta $[np]+1$-edik eleme, ahol $[np]$ az np szám egész része.

Ha pl. a rendezett minta elemei

$$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16$$

ill.

$$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17$$

akkor a $p = \frac{1}{4}$ -hez tartozó alsó (negyedrendű) kvantilis az első esetben

$$[np]+1 = \left[16 \cdot \frac{1}{4}\right]+1 = [4]+1 = 4+1 = 5,$$

a második esetben:

$$[np]+1 = \left[17 \cdot \frac{1}{4}\right]+1 = [4{,}25]+1 = 4+1 = 5.$$

Az alsó kvantilisnek mindkét esetben az

$$x_{\frac{1}{4}} = X_5^* = 5$$

elemet tekintjük. Tehát a mintaelemek $\frac{1}{4}$ része, azaz $100 \cdot \frac{1}{4} = 25\%$ -a kisebb, mint $X_5^* = 5$.

Példa

Egy automata darabológép adott hosszúságú pálcikákat készít. Az elkészített pálcika hossza legyen az X valószínűségi változó. A pálcikák hossza az adott hosszúságnál nagyobb is, kisebb is lehet. Tegyük fel, hogy $n = 26$ mérést végezve a méreteltérések számát mm-es intervallumonként rögzítettük: a ± 4 és ± 3 mm közé 1-1, a ± 3 és ± 2 mm közé 3-3, a ± 2 és ± 1 mm közé 4-4, a ± 1 és 0 mm közé pedig 5-5 esett.

Szerkesszük meg

a) a hosszúság eltérés sűrűséghisztogramját;

b) a közelítő tapasztalati eloszlásfüggvényt,

és számítsuk ki

c) a mintaátlagot;

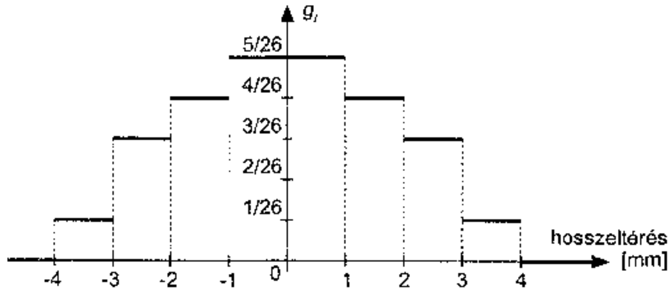
d) a korrigált tapasztalati szórást.

Megoldás

Először kiszámítjuk és táblázatba foglaljuk a szükséges részeredményeket.

Osztályközök: mm	Osztályközép x_i mm	Gyakoriság: f_i	Relatív gyakoriság: g_i	$\sum_i g_i$
(−4;−3)	−3,5	1	1/26	1/26
(−3;−2)	−2,5	3	3/26	4/26
(−2;−1)	−1,5	4	4/26	8/26
(−1;−0)	−0,5	5	5/26	13/26
(0;1)	0,5	5	5/26	18/26
(1;2)	1,5	4	4/26	22/26
(2;3)	2,5	3	3/26	25/26
(3;4)	3,5	1	1/26	26/26
		$\sum f_i = 26$	$\sum g_i = 1$	

a) Mivel $x_i - x_{i-1} = 1$, ezért intervallumonként a g_i magasságú téglalapok adják a hosszúságeltérés sűrűséghisztogramját (45. ábra):

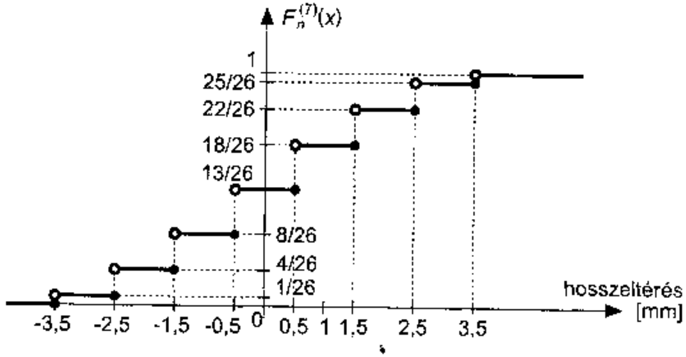


45. ábra. Sűrűséghisztogram

b) A közelítő tapasztalati eloszlásfüggvényt az

$$F_n^{(7)}(x) = \begin{cases} 0, & \text{ha } x \le -3{,}5; \\ \sum_{i=1}^{k} g_i, & \text{ha } x_{k-1} < x \le x_k, (k = 1, 2, \ldots, 7); \\ 1, & \text{ha } x > 3{,}5 \end{cases}$$

képlet definiálja. (Intervallum felosztásnak az osztályközepeket vettük, 46. ábra.)



46. ábra. Az $F_n^{(7)}(x)$ tapasztalati eloszlásfüggvény

c) A mintaátlag, mivel az egyes osztályközepet f_i -szer veszi fel a változó, ezért

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i \cdot x_i}{n} =$$

$$= \frac{1\cdot(-3{,}5)+3\cdot(-2{,}5)+4\cdot(-1{,}5)+5\cdot(-0{,}5)+5\cdot(0{,}5)+4\cdot 1{,}5+3\cdot 2{,}5+1\cdot 3{,}5}{26} = 0$$

amit a hosszúságeltérés szimmetrikus volta miatt közvetlenül is beláthatunk.

d) A korrigált tapasztalati szórás:

$$s^* = \sqrt{\frac{\sum_{i=1}^{k} f_i \cdot (x_i - \overline{X})^2}{n-1}} =$$

$$= \sqrt{\frac{(-3{,}5)^2+3\cdot(-2{,}5)^2+4\cdot(-1{,}5)^2+5\cdot(-0{,}5)^2+5\cdot(0{,}5)^2+4\cdot(1{,}5)^2+3\cdot(2{,}5)^2+(3{,}5)^2}{25}} =$$

$$= \sqrt{\frac{82{,}5}{25}} = \sqrt{3{,}3} \approx 1{,}82 \text{ mm.}$$

E.1. Ellenőrző kérdések az 1. fejezethez

1. Hogyan határozható meg a matematikai statisztika, valamint feladata és tárgyköre?

2. Mit nevezünk sokaságnak és mintának?

3. Milyen követelményeket támasztunk a statisztikai mintavétellel szemben?

4. Melyek a statisztikai minta jellemzői?

5. Mit nevezünk korrigált tapasztalati szórásnégyzetnek és szórásnak?

6. Hogyan definiáljuk a tapasztalati (empirikus) eloszlásfüggvényt?

7. Hogyan szerkesztjük meg a sűrűséghisztogramot?

S.1. Feladatok az 1. fejezethez

S.1.1. A szénosztályozóba érkező vagonokból véletlenszerűen kiválasztanak egyet-egyet, és megmérik a kiválasztott vagonban levő meddő mennyiségét. A vizsgált 142 vagonnál tapasztaltakat a következő táblázatba foglalták:

Meddőmennyiség: kg (osztályközök)	Gyakoriság (vagonok száma): f_i
163–243	8
244–323	19
324–403	23
404–483	27
484–563	18
564–643	15
644–723	12
724–803	12
804–884	8

Szerkesszük meg

a) a mérési eredményekhez tartozó sűrűséghisztogramot;

b) a közelítő tapasztalati eloszlásfüggvényt.

S.1.2. Egy automata palacktöltőgép cm^3-ben megadott mennyiségű folyadékot tölt a palackokba. A palackokba – véletlenszerűen – az adott folyadékmennyiségtől eltérően hol több, hol kevesebb kerül. Az X valószínűségi változó jelölje a palackba töltött folyadékmennyiséget. A palackok tartalmát 12 elemű véletlen minta alapján ellenőrizték és a vizsgálat eredményét osztályba sorolással a következő táblázatba foglalták:

Folyadékmennyiség eltérés, osztályhatárok cm^3	Osztályközép cm^3	Gyakoriság f_i
(–3;–2)	–2,5	1
(–2;–1)	–1,5	2
(–1;0)	–0,5	3
(0;1)	0,5	3
(1;2)	1,5	2
(2;3)	2,5	1

Szerkesszük meg

a) a folyadékmennyiség-eltérés sűrűséghisztogramját;

b) a tapasztalati eloszlásfüggvényt;

és számítsuk ki

c) a mintaátlagot;

d) a korrigált tapasztalati szórást.

Mennyi a valószínűsége annak, hogy a folyadékmennyiség-eltérés 1 cm^3-nél nagyobb legyen?

S.1.3. Egy homokbánya markolója által kiemelt homok kg súlyát véletlen kiválasztással 31 esetben lemérték. A 31 elemű véletlen minta alapján végzett vizsgálat eredményét a következő táblázatba foglalták:

Osztályhatárok: kg	Gyakoriság: f_i
470–480	1
480–490	4
490–500	10
500–510	10
510–520	4
520–530	2

Szerkesszük meg

a) a súlyeltérés sűrűséghisztogramját;

b) a tapasztalati eloszlásfüggvényt;

és számítsuk ki

c) a mintaátlagot;

d) a korrigált tapasztalati szórást.

Mi a valószínűsége annak, hogy 485 kg-nál kevesebb homokot markol a gép?

S.1.4. 12,8 mm szélességű alkatrészek gyártására egy olyan daraboló gép került beállításra, amely $12{,}8 \pm 1$ mm pontossággal készíti az elemeket. Véletlenszerűen 25 db-ot kiválasztva, az alábbi mérési eredményt kaptuk:

Osztályközök	Gyakoriság
10–11	1
11–12	5
12–13	13
13–14	4
14–15	2

a) Számítsuk ki az adatok átlagát, korrigált szórását, relatív szórását;

b) Az elemek hány százaléka felel meg az előírt pontosságnak?

MÁSODIK FEJEZET

Statisztikai becslések

Melyek az ún. jó becslés feltételei?

Ebben a fejezetben a valószínűségeloszlások ismeretlen adatainak, paramétereinek a mintából való közelítő meghatározásával, más szóval becslésével foglalkozunk. Ha pl. egy fizikai mennyiséget olyan mérésekből kell meghatároznunk amelyek véletlen hibát tartalmaznak, akkor általában a feladat megoldásához egy valószínűségeloszlás egyik jellemző paraméterének becslését kell elvégeznünk. A statisztika alkalmazása során tudjuk, – elméleti megfontolások vagy mérések alapján – hogy egy X valószínűségi változó pl. normális eloszlású, vagy eloszlása azzal közelíthető. A sűrűségfüggvény ekkor, mint ismert,

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

ahol m és σ ismeretlen paraméterek, amelyeket a statisztikai mintából kell becsülnünk. $f(x)$ tehát függ az m és σ értékektől, melyet

$$f(x;m,\sigma)$$

alakban jelölünk.

Ha valamelyik ismeretlen paramétert egyetlen számértékkel becsüljük, akkor **pontbecslésről** beszélünk, ha pedig egy olyan intervallummal, amely nagy valószínűséggel tartalmazza az ismeretlen paramétert, akkor **intervallumbecslésről** beszélünk.

2.1. A pontbecslés módszere

A pontbecslés módszereit részletesen ismertetik az irodalomjegyzék művei. Mi ebben a pontban két, paraméter becslésére használható képletet adunk meg.

Ha az X valószínűségi változó eloszlásfüggvénye k számú ismeretlen paramétertől függ:

$$F(x; a_1, a_2, \ldots, a_k) \tag{1}$$

és az X-re vonatkozó n számú mérés eredménye $X_1, X_2, \ldots, X_n$, akkor az ismeretlen a_i állandók becslését a mintaelemek

$$b_i = b_i(X_1, X_2, \ldots, X_n) \tag{2}$$

függvényei, statisztikái segítségével végezzük. A (2) függvényeket az a_i *paraméterek statisztikai becsléseinek* nevezzük.

Tegyük fel pl. hogy a_1 az X változó ismeretlen várható értéke és a_2 pedig az ismeretlen szórásnégyzete, akkor ezek becslését általában az

$$a_1 = M(X) \approx \overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \tag{3}$$

mintaközéppel, ill. az

$$a_2 = D^2(X) \approx s^{*2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} \tag{4}$$

korrigált tapasztalati szórásnégyzettel végezzük. A (3) és (4) becslések helyessége a nagy számok törvényei alapján igazolhatók.

Hogyan juthatunk jó becslésekhez? Erre a becsléselméletben kidolgozott, alábbi követelmények adnak választ:

1. Torzítatlanság: Egy $b_i = b_i(X_1, X_2, \ldots, X_n)$ statisztikát valamely a paraméter torzítatlan becslésének nevezzük, ha várható értéke a-val egyenlő, azaz

$$M\big(b_i(X_1, X_2, \ldots, X_n)\big) = a.$$

Ez azt követeli meg, hogy a méréssorozatonként ingadozó

$$b_i = b_i(X_1, X_2, \ldots, X_n)$$

értékek átlaga a paraméter valódi értéke körül ingadozzék. Ha $M\big(b_i(X_1, X_2, \ldots, X_n)\big) \neq a$, akkor a becslés torzított.

2. Hatásosság (efficiencia): A becslést hatásosság szempontjából akkor mondjuk jónak, ha a $b_i = b_i(X_1, X_2, \ldots X_n)$ statisztika ingadozása megfelelően kicsiny. Az a becslés a **leghatékonyabb (legefficiensebb)**, amelynek szórásnégyzete minden más torzítatlan becslés szórásnégyzeténél kisebb. Pl. ha a $D_a^2(b_{i1}) \leq D_a^2(b_{i2})$, akkor a b_{i1} becslést hatásosabbnak (efficiens becslésnek) nevezzük. A becslés H_f **hatásfokán (efficienciáján)** az összes lehetséges becslések szórásnégyzetei alsó határának és a kérdéses becslés szórásnégyzetének hányadosát értjük, azaz a $\frac{D_a^2}{D^2(b_i)}$ hányadost, melyre $0 < H_f \leq 1$ egyenlőtlenség teljesül.

3. Konzisztencia (megegyezés): A b_i statisztika sorozatot valamely a paraméter konzisztens becslésének nevezzük, ha elegendő nagy n esetén $b_i = b_i(X_1, X_2, \ldots, X_n)$ nagy valószínűséggel jól megközelíti a paraméter értékét, vagyis minden pozitív ε-hoz és δ-hoz található olyan N szám, amelyre teljesül hogy

$$P\big(|b_i(X_1, X_2, \ldots, X_n) - a| \geq \varepsilon\big) \leq \delta, \qquad \text{ha } n > N.$$

Ha a $b_i = b_i(X_1, X_2, \ldots, X_n)$ torzítatlan becslés és szórásnégyzete n növekedésével 0-hoz tart, akkor azt mondjuk, hogy b_i az a-nak *erősen konzisztens becslése*.

4. Elégségesség: Ha a $b_i = b_i(X_1, X_2, \ldots, X_n)$ becslés a mintaelemekből nyerhető minden információt megad a kérdéses paraméterre vonatkozólag, akkor a $b_i = b_i(X_1, X_2, \ldots, X_n)$ statisztikai függvényt **elégséges becslésnek vagy elégséges statisztikának** nevezzük.

Megjegyzések

1. Minden eloszlás esetén a mintaelemek mintaközepe (3) torzítatlan becslés a várható értékekre, továbbá a mintaelemek korrigált tapasztalati szórásnégyzete (4) torzítatlan becslés az elméleti szórásnégyzetre.

2. A normális-, az exponenciális- valamint a *Poisson*-eloszlás várható értékére a mintaközép (3) elégséges becslés.

3. A mintaelemekből a teljes eloszlásfüggvényt az 1.2.1. pont (7) ill. (8) képletével, a becslését pedig a tapasztalati sűrűségfüggvénynyel ill. sűrűséghisztogrammal végezzük.

2.1.1. A maximum-likelihood módszer

A maximum-likelihood (legnagyobb valószínűség) módszerét jól alkalmazhatjuk annak a statisztikának a meghatározására, amely az ismeretlen paraméter legjobb becslését adja az adott információk mellett. A módszert *R. A. Fisher* 1922-ben megjelent cikke tárgyalja részletesen, bár már korábban is alkalmazták.

Tegyük fel, hogy $X_1, X_2, \ldots, X_n$ az adott minta, amelynek segítségével az ismeretlen a paramétert becsülni akarjuk. Ha az ismeretlen a paramétertől függő eloszlás mintaelemeinek közös sűrűségfüggvénye $f(x;a)$, akkor a független mintaelemek együttes sűrűségfüggvénye

$$f(x_1;a)\cdot f(x_2;a)\cdot\ldots\cdot f(x_n;a)=\prod_{i=1}^{n} f(x_i;a) \qquad (1)$$

ahol $x_1, x_2, \ldots, x_n$ számok a mintaelemeknek a kísérlet során mért értékei. Ekkor a maximum-likelihood módszer szerint az a paraméter becslésének az $X_1, X_2, \ldots, X_n$ mintaelemeknek azt az $\hat{a}=\hat{a}(X_1, X_2, \ldots X_n)$ függvényét nevezzük, melyre az $\prod_{i=1}^{n} f(x_i,\hat{a})$ szorzat a lehető legnagyobb értéket veszi fel, feltéve, hogy a maximum létezik és egyértelmű. Feladat, tehát az

$$L=\prod_{i=1}^{n} f(x_i,a) \qquad (2)$$

egyparaméteres függvény szélsőértékének meghatározása. Mivel egy függvénynek és logaritmusának a szélsőérték helyei megegyeznek, ezért a (2) szorzatfüggvénynél célszerűbb a függvény logaritmusából számítani a szélsőértékhelyet, azaz

$$\frac{d\ln L}{da}=0 \qquad (3)$$

egyenletből meghatározni az $\hat{a}$ értékét. Ha a (3) egyenletnek több gyöke van, akkor a lokális szélsőértékek közül választjuk ki az abszolút maximum helyet. Ha az a paramétert $\hat{a}$-val becsüljük, akkor e paraméter mellett van a legnagyobb valószínűsége annak, hogy a vizsgált minta $x_1, x_2, \ldots, x_n$ értékeit figyeljük meg.

Diszkrét valószínűségeloszlás esetén is alkalmazható a maximum-likelihood módszer. Ekkor a likelihood-függvény:

$$L=\prod_{i=1}^{m} p(x_i,\hat{a})^{r_i},$$

ahol $P(X=x_i)=p(x_i,a)$ $(i=1,2,\ldots,m)$, és r_i az x_i mintabeli értékek gyakorisága.

Megjegyzés

1. Ha az a paramétert a maximum-likelihood módszerrel számított $\hat{a}$-val becsüljük, akkor az a valamely $g(a)$ függvényének maximum-likelihood becslése $g(\hat{a})$ (invariancia tulajdonság).

2. A gyakorlat szempontjából elegendően általános feltételek mellett kimutatható, hogy maximum-likelihood becslés konzisztens és nagy n értékekre közelítőleg minimális szórású, valamint ha van az a paraméternek elégséges becslése, akkor maximum-likelihood módszerrel ennek valamely függvényét kapjuk.

3. A többváltozós függvény szélsőérték-számítási eljárásának alkalmazásával több paramétert tartalmazó valószínűségeloszlás esetén is alkalmazható a maximum-likelihood módszer.

Példa

a) Vizsgáljuk meg, hogy a $0\le p\le 1$ intervallumban folytonos

$$L(p)=\binom{n}{k}p^k(1-p)^{n-k}$$

függvénynek hol van maximuma.

b) Tegyük fel, hogy egy binomiális eloszlású statisztikai sokaságból vett 25 elemű mintának 17 eleme az előírt tulajdonságú. A statisztikai sokaságban az előírt tulajdonságú elemek arányát akarjuk meghatározni. Számítsuk ki a valószínűségértékeket 0,4-től 0,9-ig 0,5 lépéstávolsággal és keressük ki a legnagyobb értékhez tartozó p értéket.

Megoldás

a) Képezzük az $L(p)$ függvény első p-szerinti deriváltját, tegyük egyenlővé nullával, és az így kapott egyenletet oldjuk meg p-re:

$$\frac{dL(p)}{dp}=\binom{n}{k}\cdot k\cdot p^{k-1}(1-p)^{n-k}+$$

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

$$=\binom{n}{k}p^k(1-p)^{n-k}\left(\frac{k}{p}-\frac{n-k}{1-p}\right),$$

és az

$$\binom{n}{k}p^k(1-p)^{n-k}\left(\frac{k}{p}-\frac{n-k}{1-p}\right)=0$$

egyenletből a megoldás:

$$\hat{p}=\frac{k}{n}$$

a p maximum-likelihood becslése, mivel ebben az intervallumban az $L(p)$ függvény alulról konkáv.

b) A $v(p)=\binom{25}{17}p^{17}(1-p)^8$ értékeket kiszámítva a táblázatból közelítőleg kiolvashatjuk az ismeretlen p értéket közelítő $\hat{p}$ -t.

p	$v(p)$
0,4	0,003121
0,45	0,011523
0,50	0,032233
0,55	0,070133
0,60	0,119980
0,65	0,160742
0,70	0,165080
0,75	0,124056
0,80	0,062349
0,85	0,017495
0,90	0,001804

A maximum közelítőleg a $p=0{,}70\approx\frac{k}{n}=\frac{17}{25}=0{,}68$ relatív gyakorisággal kapott értéknél található, azaz $\hat{p}=0{,}68\approx0{,}70$.

2.2. Konfidencia-intervallum

Gyakran az eloszlásfüggvény valamely paramétere pontos értékének becslése helyett megelégszünk azzal, hogy a mintaelemek két statisztikai függvényével megadunk egy intervallumot, amely előírt valószínűséggel tartalmazza az ismeretlen paramétert.

Definíció. Legyen az X valószínűségi változó eloszlásfüggvénye $F(x;a)$, és tegyük fel, hogy az a rögzített ismeretlen paraméter értéke valamely $(a_1;a_2)$ intervallumba esik, és tekintsük az X változó n elemű X_1, X_2, ..., X_n mintáját. Ha egy adott $1-p$ valószínűséghez található olyan

$$b_1=b_1(X_1,X_2,\dots,X_n;p) \text{ és } b_2=b_2(X_1,X_2,\dots,X_n;p)$$

statisztikai függvény, hogy a $(b_1;b_2)$ intervallum $1-p$ valószínűséggel tartalmazza az a ismeretlen állandót, akkor (b_1,b_2) az a paraméter **konfidencia-intervalluma (megbízhatósági intervalluma)** $1-p$ **megbízhatósággal,** ahol p egy 0-hoz közel eső kis valószínűség.

Ezt úgy is mondhatjuk, hogy az a paramétert egy olyan intervallummal becsüljük, amely $1-p$ valószínűséggel lefedi az a paraméter értékét. Ezt az intervallumot az a paraméter $100(1-p)\%$ -os konfidencia-intervallumának, a $100(1-p)\%$ -ot a megbízhatóság szintjének, az intervallum kezdő és végpontját pedig konfidencia határoknak is mondjuk. Az $1-p$ megbízhatóság tehát azt jelenti, hogy nagyszámú mintavétel esetén a paraméter pontos értékét az esetek $100(1-p)\%$ -ában tartalmazza az intervallum, és $100p\%$ -ban nem. Ha változatlan mintanagyságnál csökkentjük a p értékét, akkor nagyobb megbízhatósági szintet kapunk, de általában az intervallum is hosszabb lesz. Rögzített megbízhatósági szinten az intervallum szűkítéséhez a mintaelemszám növelése szükséges. Az intervallum hossza n növelésével $\sqrt{n}$ arányban csökken. A fejezetben a várható értékre és a szórásra vonatkozó $100(1-p)\%$ -os konfidencia-intervallum meghatározásával foglalkozunk még.

2.2.1. A várható érték becslése

Tegyük fel, hogy az X valószínűségi változó eloszlása (eloszlásfüggvénye) az a paramétertől függ, azaz

$$P(X < x) = F(x;a).$$

Határozzuk meg az a paraméter konfidencia-intervallumát az X-re vonatkozó $X_1, X_2, \ldots, X_n$ minta alapján.

a) Legyen X egy normális eloszlású valószínűségi változó ismert σ szórással és ismeretlen a várható értékkel, ekkor

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

Az a várható értéket $\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$ mintaközéppel becsüljük. Kimutatható, hogy az $\overline{X}$ szintén normális eloszlású valószínűségi változó, amelynek várható értéke: $M(\overline{X}) = a$ és szórása az $n \cdot D^2(\overline{X}) = \sigma^2$ összefüggésből $D(\overline{X}) = \frac{\sigma}{\sqrt{n}}$, tehát az $u = \frac{\overline{X} - a}{\sigma}\sqrt{n}$ egy az y-tengelyre szimmetrikus standard normális eloszlású valószínűségi függvény. A konfidencia-intervallumot tehát célszerű 0-ra szimmetrikusan $(-u_p; u_p)$ intervallumként előállítani úgy, hogy

$$P(-u_p \leq \frac{\overline{X} - a}{\sigma}\sqrt{n} < u_p) = 1 - p \tag{1}$$

legyen. Az u_p értékét az

$$1 - p = 2\Phi(u_p) - 1 \tag{2}$$

egyenletből az 1. sz. táblázat alapján meghatározhatjuk, ui. (2)-ből

$$\Phi(u_p) = 1 - \frac{p}{2}$$

és így az $1 - \frac{p}{2}$ ismeretében az u_p értéke a táblázatból visszakereshető.

Mivel $1 - p$ valószínűséggel érvényesek a következő egyenlőtlenségek:

$$-u_p \leq \frac{\overline{X} - a}{\sigma}\sqrt{n} < u_p \tag{3}$$

$$\overline{X} - u_p\frac{\sigma}{\sqrt{n}} < a \leq \overline{X} + u_p\frac{\sigma}{\sqrt{n}} \tag{4}$$

ezért az m várható értékre (az a paraméterre) az

$$\left]X - u_p\frac{\sigma}{\sqrt{n}}; X - u_p\frac{\sigma}{\sqrt{n}}\right] \tag{5}$$

intervallum egy $100(1 - p)\%$ szintű konfidencia-intervallum, ami azt jelenti, hogy az m várható érték $1 - p$ valószínűséggel esik ebbe az intervallumba, ha az intervallumbecslést elég sok mintán elvégezzük. Tévedésünk pedig $100p\%$ értékkel becsülhető. (5)-ből következik, hogy az $\overline{X}$ és az m eltérésének abszolút értéke legfeljebb az intervallum hosszának a fele, azaz

$$u_p\frac{\sigma}{\sqrt{n}},$$

amely n növekedésével nullához tart.

Ha előírt megbízhatósági szinthez megadjuk, hogy a konfidencia-intervallum félhossza legfeljebb d legyen, akkor az

$$u_p\frac{\sigma}{\sqrt{n}} \leq d$$

egyenlőtlenségből **a minta szükséges elemszámának nagyságára** az

$$n \geq u_p^2\frac{\sigma^2}{d^2} \tag{6}$$

egyenlőtlenséget kapjuk.

Ha a konfidencia-intervallum a számunkra elfogadhatónál nagyobb, akkor arra következtethetünk, hogy a statisztikai minta elemszáma nem elegendő a paraméter kielégítő pontosságú becsléséhez, bővíteni kell az adathalmazt.

Példa

Határozzuk meg u_p értékét 95%-os valószínűségi szinthez. Számítsuk ki az m várható értéket lefedő, u_p értéknek megfelelő konfidencia-intervallumot, ha 25 mérés mintaközepe 8,4 cm, és szórása 0,2 cm.

Megoldás

A 95%-os valószínűségi szintnek $p = 0{,}05$ felel meg, tehát $1 - \frac{p}{2} = 1 - \frac{0{,}05}{2} = 0{,}9750$. Az 1. sz. táblázatban a 0,9750 érték az 1,9 sor 6-os fejlécű oszlopában van, tehát $u_p = 1{,}96$.

A (4) formula értelmében az X valószínűségi változó m várható értékét lefedő intervallum:

$$8{,}4 - 1{,}96 \cdot \frac{0{,}2}{\sqrt{25}} < m < 8{,}4 + 1{,}96 \cdot \frac{0{,}2}{\sqrt{25}},$$

vagyis,

$$8{,}3216 < m < 8{,}4784.$$

Az m tehát 95%-os valószínűséggel esik a $(8{,}3216; 8{,}4784)$ konfidencia-intervallumba.

Megjegyzés

A továbbiakban a gyakorlati feladatok megoldásához felhasználjuk a **Student-eloszlás** 4. táblázatát, valamint a χ^2 **(khi-négyzet)-eloszlás** 3. táblázatát.

b) Ha $Y, X_1, X_2, \ldots, X_n$ független, $N(0;1)$ eloszlású valószínűségi változók, akkor a

$$t = \frac{\sqrt{n}Y}{\sqrt{X_1^2 + X_2^2 + \ldots + X_n^2}} \tag{7}$$

valószínűségi változó eloszlását **n szabadságfokú, Student-** vagy **t-eloszlásnak** nevezzük, melynek **sűrűségfüggvénye:**

$$f(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \cdot \frac{1}{\left(\frac{x^2}{n}+1\right)^{\frac{n+1}{2}}} \tag{8}$$

várható értéke: $M(X) = 0$, csak $n \geq 2$-re létezik,

szórásnégyzete: $D^2(X) = \frac{n}{n-2}$, csak $n \geq 3$-ra létezik.

A (8) sűrűségfüggvény n növekedésével az $N(0;1)$ eloszlás sűrűségfüggvényéhez konvergál.

A 4. táblázat a t-eloszlás értékeit tartalmazza adott %-ú statisztikai biztonság és adott n-számú mérés esetére.

Pl. ha a normális eloszlású sokaságnak nem ismerjük a szórását, akkor a σ helyett az s^* korrigált tapasztalati szórást helyettesítve az u kifejezésébe a

$$t = \sqrt{n}\,\frac{\overline{X} - m}{s^*} \tag{9}$$

valószínűségi változót kapjuk, amely egy $n-1$ szabadságfokú *Student*-eloszlású valószínűségi változó. A *Student*-eloszlás táblázatából (4. sz. táblázat) adott p-hez meghatározható a t_p szám, amelyre teljesül a

$$P\left(-t_p \leq t \leq t_p\right) = P\left(|t| \leq t_p\right) = 1 - p \tag{10}$$

egyenlőség, azaz t értéke $1-p$ valószínűséggel esik a $\left[-t_p; t_p\right]$ intervallumba. Ez viszont (9) figyelembevételével azt jelenti, hogy m értéke $1-p$ valószínűséggel esik az

$$\left[\overline{X} - t_p \frac{s^*}{\sqrt{n}}; \overline{X} + t_p \frac{s^*}{\sqrt{n}}\right] \tag{11}$$

intervallumba. Tehát ismeretlen szórás esetén a várható értékre ez lesz a $100(1-p)\%$ szintű konfidencia-intervallum.

Példák

1. Egy szerves vegyület oxigéntartalmának vizsgálatához 16 mérést végeztünk, melyek alapján $\overline{X} = 2{,}75\%$ átlagot kaptunk. Tegyük fel, hogy a szórás ismert, pl. $\sigma = 0{,}28\%$. Számítsuk ki a várható értékre vonatkozó 96%-os szintnek megfelelő konfidencia-intervallumot, ha az $\overline{X}$ mintaközép $N(m, \frac{\sigma}{\sqrt{n}})$ eloszlású.

Megoldás

Mivel a 96%-os szinthez $p = 0{,}04$ érték tartozik, így

$$\Phi(u_p) = 1 - \frac{p}{2} = 1 - 0{,}02 = 0{,}98,$$

és így

$$u_p \approx 2{,}06.$$

A konfidencia határok:

$$\overline{X} - u_p \frac{\sigma}{\sqrt{n}} = 2{,}75 - 2{,}06 \cdot \frac{0{,}28}{\sqrt{16}} \approx 2{,}61;$$

$$\overline{X} + u_p \frac{\sigma}{\sqrt{n}} = 2{,}75 + 2{,}06 \frac{0{,}28}{\sqrt{16}} \approx 2{,}89,$$

tehát a szerves vegyület oxigéntartalma 96%-os valószínűséggel esik a $[2{,}61; 2{,}89]$ intervallumba.

2. A tv-képcsövek vizsgálatánál 19 db-nak mérték meg az élettartamát, amely közelítőleg normális eloszlású volt 99%-os biztonsági szinten. Milyen konfidencia-intervallumba esik az egész sokaság várható értéke, ha a tapasztalati várható értéke $\overline{X} = 2200$ óra, korrigált tapasztalati szórása pedig $s^* = 190$ óra.

Megoldás

Mivel a minta korrigált tapasztalati szórása ismert, ezért az

$$\overline{X} - t_p \frac{s^*}{\sqrt{n}} \le m \le \overline{X} + t_p \frac{s^*}{\sqrt{n}}$$

konfidencia-intervallum határait kell kiszámítanunk. $19 - 1 = 18$ a szabadságfokok száma, így a 99%-nak megfelelő $t_p = 2{,}898$.

A konfidencia határok:

$$\overline{X} - t_p \frac{s^*}{\sqrt{n}} = 2200 - 2{,}898 \cdot \frac{190}{\sqrt{19}} \approx 2073{,}68 \text{ óra};$$

$$\overline{X} + t_p \frac{s^*}{\sqrt{n}} = 2200 + 2{,}898 \cdot \frac{190}{\sqrt{19}} \approx 2326{,}32 \text{ óra},$$

tehát az egész sokaság várható értékét a $[2073{,}68; 2326{,}32]$ intervallum lefedi.

2.2.2. A szórás becslése

Az n számú $X_1, X_2, \ldots, X_n$ független, standard normális eloszlású valószínűségi változó

$$\chi^2 = X_1^2 + X_2^2 + \ldots + X_n^2$$

négyzetösszegének eloszlását **n szabadságfokú χ^2 (khi-négyzet)-eloszlásnak** nevezzük.

Igazolható, hogy χ^2 **sűrűségfüggvénye**

$$f_n(x) = \begin{cases} \dfrac{x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}, & \text{ha } x \ge 0 \\ 0, & \text{ha } x < 0 \end{cases} \tag{1}$$

ahol a Γ-val jelölt ún. *gamma függvényt* a következő képlettel értelmezzük:

$$\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} \mathrm{d}x; \; p > 0.$$

A χ^2 *(khi-négyzet)-eloszlás*

várható értéke: $M(\chi^2) = n;$

szórásnégyzete: $D^2(\chi^2) = 2n.$

A 3. sz. táblázat a χ^2-eloszlású változó eloszlásfüggvényének értékeit tartalmazza a valószínűség százalékában a szabadságfokok szerint.

Pl. $n = 12$ szabadságfokhoz $p = 0{,}1$ esetén $100(1 - p) = 100(1 - 0{,}1) = 90{,}0$ -hez 18,5 érték tartozik.

Vegyünk most egy n elemű mintát az $N(m; \sigma)$ normális eloszlású sokaságból és tegyük fel, hogy ismeretlen az eloszlás szórása.

Igazolható, hogy az

$$\frac{ns^{*2}}{\sigma^2} \qquad (2)$$

valószínűségi változó $n-1$ szabadságfokú χ^2-eloszlású, ahol s^{*2} az n elemű minta alapján számított tapasztalati szórásnégyzet. Kimutatható, hogy a (2) -vel adott érték $1-p$ valószínűséggel esik a

$$\left[\chi^2_{1-\frac{p}{2}};\chi^2_{\frac{p}{2}}\right]$$

intervallumba, azaz

$$P\left(\chi^2_{1-\frac{p}{2}} \le \frac{ns^{*2}}{\sigma^2} \le \chi^2_{\frac{p}{2}}\right) = 1-p. \qquad (3)$$

A (3) -ból következik, hogy a sokaság σ^2 szórásnégyzete $1-p$ valószínűséggel esik a

$$\left[\frac{ns^{*2}}{\chi^2_{\frac{p}{2}}};\frac{ns^{*2}}{\chi^2_{1-\frac{p}{2}}}\right] \qquad (4)$$

intervallumba.

Ez az intervallum az ismeretlen σ^2 paraméter $100(1-p)\%$ -os konfidencia-intervalluma.

A (4) -ből pedig felírható a σ szórás $100(1-p)\%$ -os konfidencia-intervalluma:

$$\left[\frac{\sqrt{n}s^*}{\sqrt{\chi^2_{\frac{p}{2}}}};\frac{\sqrt{n}s^*}{\sqrt{\chi^2_{1-\frac{p}{2}}}}\right]. \qquad (5)$$

Ha a p értéke adott, akkor a $\chi^2_{\frac{p}{2}}$ és $\chi^2_{1-\frac{p}{2}}$ értékeket a 3. sz. táblázatból olvassuk ki a megfelelő szabadságfok figyelembevételével.

Példa

18 doboz mérlegelése alapján a töltési súly tapasztalati szórására $s^* = 12$g értéket kaptunk. Határozzunk meg a szórásra 95%-os konfidencia-intervallumot.

Megoldás

Mivel a 95%-os szinthez $p = 0{,}05$ érték tartozik, és a szabadságfokok száma 17, így a 3. sz. táblázatból

$$\chi^2_{\frac{p}{2}} = \chi^2_{0{,}025} = 30{,}191; \quad \chi^2_{1-\frac{p}{2}} = \chi^2_{0{,}975} = 7{,}564.$$

Az intervallum határai (5)-nek megfelelően:

$$\frac{\sqrt{n}\cdot s^*}{\sqrt{\chi^2_{\frac{p}{2}}}} = \frac{4{,}24264\cdot 12}{5{,}4946337} \approx 9{,}27\text{g};;$$

$$\frac{\sqrt{n}\cdot s^*}{\sqrt{\chi^2_{1-\frac{p}{2}}}} = \frac{4{,}24264\cdot 12}{2{,}7502727} \approx 18{,}51\text{g},$$

tehát a szórás 95%-os konfidencia-intervalluma $[9{,}27;18{,}51]$.

E.2. Ellenőrző kérdések a 2. fejezethez

1. Mit nevezünk statisztikai becslésnek?

2. Melyek az ún. jó becslés követelményei?

3. Hogyan definiáljuk a konfidencia-intervallumot?

4. Hogyan számítjuk ki az m várható érték konfidencia-intervallumát?

5. Mit nevezünk n szabadságfokú χ^2-eloszlásnak?

6. Mit nevezünk n szabadságfokú *Student-* vagy *t*-eloszlásnak?

7. Hogyan számítjuk ki a szórásnégyzet és a szórás $100(1-p)\%$ -os konfidencia-intervallumát, ha a sokaság $N(m;\sigma)$ eloszlású?

S.2. Feladatok a 2. fejezethez

S.2.1. Egy hajóra szerelt kotrógép egy adott időszakban 5000 kanál kavicsot emel ki. Egy kanalának (puttonyának) átlagos töltési súlya, 100 mérlegelés alapján, 705 kg. Legyen a töltési súly normális eloszlású $\sigma = 50$ kg szórással. Határozzuk meg a 90, 95 és 99%-os megbízhatósági szinteknek megfelelő konfidencia-intervallumokat a töltési súly várható értékére vonatkozóan.

S.2.2. Az 1. feladatban említett 5000 puttony közül hányat kell lemérni ahhoz, hogy a konfidencia-intervallum félhossza 95%-os szint mellett 7 legyen?

S.2.3. Oldjuk meg az 1. feladatot abban az esetben, ha az 5000 puttony közül csak 30-at mértünk le és nem ismerjük a σ szórást.

S.2.4. Oldjuk meg az 1. feladatot, ha az ott közölt adatok közül a σ szórást a 100 mérlegelés alapján számított $s^* = 51$ korrigált tapasztalati szórással helyettesítjük.

S.2.5. Egy vegyület hidrogén tartalmának meghatározására végeztünk méréseket. 12 mérésből az $\overline{X} = 3{,}25\%$ átlagot kaptuk. Tegyük fel, hogy ismert a szórás: $\sigma = 0{,}30\%$. Számítsuk ki a várható értékre vonatkozóan a 95%-os szintnek megfelelő konfidencia-intervallumot.

S.2.6. Egy markológépnél véletlenszerűen 16-szor lemértük a kimarkolt homok mennyiségét. A mérlegelés alapján a kimarkolt homok súlyának tapasztalati szórására $s^* = 15$ kg értéket kaptunk. Határozzunk meg a szórásra 90%-os konfidencia-intervallumot.

HARMADIK FEJEZET

Statisztikai hipotézisek vizsgálata

Hogyan ellenőrizhetjük feltevéseink helyességét?

A **hipotézis (feltevés)** általában jelenségek feltételezett magyarázata, ill. feltételezett törvényszerű összefüggések elképzelése, amelyeknek bizonyítására még nem került sor. A statisztikai vizsgálatok során feltevéseket teszünk pl. az esemény valószínűségére, a valószínűségi változók típusára, várható értékére, varianciájára, két valószínűségi változó függetlenségére, azonos eloszlására, stb. Ezeket a feltevéseket nevezzük **statisztikai hipotéziseknek**.

A **hipotézisvizsgálat (feltevésvizsgálat)** a matematikai statisztika azon fejezete, amely a statisztikai hipotézisek közötti döntésekre vonatkozó módszereket és azok elvi kérdéseit vizsgálja. A hipotézisvizsgálat a kísérleti tudományok és a műszaki gyakorlat igen hatékony eszköze, pl. fontos alkalmazási területe a minőségellenőrzés. A hipotéziseket statisztikai módszerekkel ellenőrizzük. Ezek az ún. **statisztikai próbák**, amelyek a szóban forgó változókra vett mintának, ill. a mintaelemek valamely függvényének vizsgálatán alapszanak. A következő pontokban néhány alapvető hipotézist, hipotézisvizsgálati módszert, statisztikai próbát tárgyalunk.

Állapodjunk meg néhány jelölésben. Azt a feltevést, amelyet igaznak tételezünk fel **alaphipotézisnek** vagy **nullhipotézisnek** nevezzük és H_0-val jelöljük. Pl. ha feltesszük, hogy az X valószínűségi változó várható értéke m_0, akkor ezt a nullhipotézist

$$H_0 \colon M(X) = m_0$$

szimbólummal fejezzük ki. Ezzel szembeállítjuk az ún. **ellenhipotézist**, melyet H-val jelölünk és

$$H \colon M(X) = m \neq m_0$$

kifejezéssel adunk meg, melyet akkor fogadunk el, ha H_0 nem teljesül. A két hipotézist mindig úgy kell megalkotni, hogy ha az egyik teljesül, akkor a másik nem teljesülhet, azaz H_0-nak és H-

nak egymást kizáró hipotéziseknek kell lenniük a vizsgált valószínűségi változóra vonatkozólag.

A **hipotézisvizsgálat célja** tehát a felállított hipotézis *helyességének ellenőrzése* a szóban forgó statisztikai minta alapján, és **döntéshozatal** a nullhipotézis elfogadásáról vagy elvetéséről vagyis az ellenhipotézis elfogadásáról.

A hipotézisvizsgálat menete. Legyen az X valószínűségi változó n elemű statisztikai mintája: $X_1, X_2, ..., X_n$, az összes n elemű statisztikai minta halmaza pedig legyen S. Képezzünk egy $h(X_1, X_2, \ldots, X_n)$ statisztikai függvényt ($h: S \to R$).

a) Adjunk meg egy $0 < p < 1$ számot és egy $T \subset R$ **elfogadási tartományt,** amelyre **igaz,** hogy $h(X_1, X_2, \ldots, X_n) \in T$ nagy valószínűséggel teljesül H_0 fennállása esetén, vagyis fennáll a

$$P(h(X_1, X_2, \ldots, X_n) \in T | H_0) \geq 1 - p,$$

egyenlőtlenség; azaz $1 - p$ -nél nagyobb valószínűséggel lesz igaz az a hipotézis, hogy a $h(X_1, X_2, \ldots, X_n)$ függvény értékei a T tartományba esnek. Az $1 - p$ számot a próba **szignifikancia szintjének** nevezzük.

b) Adjunk meg egy $K \subset R \setminus T$ **kritikus tartományt,** amelyre H_0 fennállása esetén kicsiny a valószínűsége annak, hogy

$$h(X_1, X_2, \ldots, X_n) \in K,$$

vagyis hogy h a kritikus halmazba esik, azaz

$$P(h(X_1, X_2, \ldots, X_n) \in K | H_0) \leq p;$$

c) **Döntés:**

ha $h(X_1, X_2, \ldots, X_n) \in T$, akkor elfogadjuk a H_0 hipotézist;

ha $h(X_1, X_2, \ldots, X_n) \in K$, akkor elutasítjuk a H_0 hipotézist, vagyis a H hipotézist fogadjuk el.

A döntés eredménye, mivel valószínűségi értékek alapján hozzuk, jó is, hibás is lehet.

Elsőfajú hibát követünk el, ha elvetjük az igaz H_0 hipotézist, mert a h a K tartományba esik. Ennek valószínűsége

$$P(\text{elsőfajú hiba}) = P(h(X_1, X_2, \ldots, X_n) \in K | H_0) \leq p,$$

és ennek alapján azt mondjuk, hogy a H_0 hipotézist $100(1-p)\%$ -os szinten elfogadjuk ill. elutasítjuk.

Másodfajú hibát követünk el, ha elfogadjuk a hibás H_0 hipotézist, mert h a T tartományba esik. Ennek valószínűsége:

$$P(\text{másodfajú hiba}) =$$
$$= P(h(X_1, X_2, \ldots, X_n) \in T | H) = 1 - P(h(X_1, X_2, \ldots, X_n) \in K | H)$$

A *hibás döntés valószínűségét* tehát az első- és másodfajú hiba valószínűségeinek összege adja, azaz

$$P(\text{hibás döntés}) = P(\text{elsőfajú hiba}) + P(\text{másodfajú hiba})$$

A hibás döntés valószínűségét tehát úgy tudjuk csökkenteni, hogy az elsőfajú hiba valószínűségének csökkentésével egyidejűleg csökkentjük a másodfajú hiba valószínűségét is. Ezt pl. a minta elemszámának növelésével tudjuk elérni.

A döntések és az elkövetett hibák logikai kapcsolatát szemlélteti a következő táblázat:

	A H_0 hipotézis	
	igaz	nem igaz
H_0 elfogadása: $h \in T$	helyes döntés	másodfajú hiba
H_0 elvetése: $h \in K$	elsőfajú hiba	helyes döntés

3.1. Az egy- és kétmintás *u*-próba

a) **Az egymintás *u*-próba.** Tegyük fel, hogy X egy $N(m;\sigma)$ normális eloszlású valószínűségi változó, ismeretlen m és ismert σ értékkel. Ellenőrizni akarjuk, hogy az X valószínűségi változó várható értéke m egyenlő-e egy adott m_0 számmal. Egy n elemű minta $\overline{X}$ átlaga általában nem lesz pontosan m_0. Kérdés: a mintaátlag mekkora eltérése esetén feltételezhetjük, hogy a várható érték m_0?

Ha a

$$H_0 : M(X) = m_0$$

nullhipotézis teljesül a

$$H : M(X) = m \neq m_0 \qquad (*)$$

ellenhipotézissel szemben, akkor az

$$u = \sqrt{n}\,\frac{\overline{X} - m_0}{\sigma} \qquad (1)$$

formulával képzett u valószínűségi változó (próbafüggvény) $N(0;1)$ standard normális eloszlású. Ekkor a II. rész 2. fejezet 2.2. pontjának (1) relációja alapján az u konfidencia-intervalluma $1-p$ valószínűséggel megadható, azaz

$$P\left(-u_p \leq \sqrt{n}\,\frac{\overline{X} - m_0}{\sigma} \leq u_p\right) = 1 - p = 2\Phi(u_p) - 1$$

és így a $\Phi(u_p) = 1 - \frac{p}{2}$ összefüggésből adott p-hez u_p értéke a 1. sz. táblázatból visszakereshető. A (*) ún. *kétoldali ellenhipotézis*, mely azt fejezi ki, hogy vagy $m < m_0$ vagy $m > m_0$.

Az (1) képlettel **számított** $u_{sz} = u$ **érték** nagy valószínűséggel (pl. $1 - p = 1 - 0{,}05 = 0{,}95$) az $\left[-u_p; u_p\right]$ elfogadási tartományba esik, és csak kis valószínűséggel (pl. 0,05) esik a kritikus tartományba.

Azt mondjuk, hogy $100(1-p)\%$ biztonsági szinten elfogadjuk a nullhipotézist, ha

$$|u_{sz}| = \left|\sqrt{n}\,\frac{\overline{X} - m_0}{\sigma}\right| \leq u_p = u_t$$

ahol u_t a táblázatból kiolvasott érték. Ekkor elfogadjuk, hogy az alapsokaság várható értéke és a feltételezett m_0 érték között az

eltérés véletlenszerű, statisztikailag nem igazolható, más szóval az **eltérés nem szignifikáns**.

Azt mondjuk, hogy a H_0 hipotézist $100(1-p)\%$ szinten elutasítjuk, vagyis a H ellentett hipotézist fogadjuk el, ha

$$|u_{sz}| > u_t,$$

azaz ha u_{sz} a kritikus tartományba esik. Ez utóbbi esetben statisztikailag igazoltnak fogadjuk el, hogy az alapsokaság várható értéke és a feltételezett m_0 érték között **szignifikáns különbség** van.

Mint az előző pontban említettük, mindkét döntésünk véletlen folytán téves lehet. Elsőfajú hibát követünk el, ha H_0 hipotézis igaz, de elvetjük az $M(X) = m_0$ feltevést, mert pl. az u_{sz} érték a kritikus tartományra esett. Másodfajú hibát követünk el, ha H_0 hipotézis nem igaz, de elfogadjuk az $M(X) = m_0$ feltevést pl. mert az u_{sz} érték az elfogadási tartományra esett.

Példák

1. Egy automata gép 200 mm hosszúságú pálcikákat készít. Vajon a gép által gyártott pálcikák hossza megfelel-e az előírt méretnek? Előzetes adatfelvételből tudjuk, hogy a gép által gyártott pálcikák hossza normális eloszlású valószínűségi változó, $\sigma = 3$ mm szórással. Az $n = 16$ elemű minta elemeinek hosszmérete:

193	198	203	191
195	196	199	191
201	196	193	198
204	196	198	200

Elfogadható-e, hogy a pálcikák hosszának eltérése az előírt mérettől 99,9%-os szinten nem szignifikáns, vagyis az egész sokaságban a várható érték $m_0 = 200$ mm.

Megoldás

Mivel a változó normális eloszlású, a minta átlaga

$$\overline{X} = \frac{\sum_{i=1}^{16} x_i}{16} = 197 \text{ mm; és } m_0 = 200 \text{ mm; } \sigma = 3 \text{ mm; } \sqrt{n} = \sqrt{16} = 4,$$

ezért az u-próba alkalmazásával dönthetünk:

$$|u_{sz}| = \left|\sqrt{n}\,\frac{\overline{X} - m_0}{\sigma}\right| = \left|4 \cdot \frac{197 - 200}{3}\right| = 4.$$

A $p = 0{,}001$ értékhez tartozó u_p táblázatbeli érték $u_t = 3{,}29$ (az 1. sz. táblázatból, mivel $1-\frac{p}{2} = 0{,}9995$), melynek 3,27 és 3,32 között bármely érték megfelel, ezért az $u_t = 3{,}29$ közbülső értéket választjuk, tehát

$$|u_{sz}| = 4 > u_t = 3{,}29$$

ezért az $m_0 = 200$ feltevést 99,9%-os szignifikancia szinten el kell vetni, a mintaátlag nem véletlenül tér el a beállított 200 mm-es hosszúságtól. A gyártást le kell állítani és a gépet újra be kell szabályozni.

2. Egy csokoládégyár speciális 14 dekagrammos csokoládészeletek gyártását kívánta megvalósítani egyik gépén, melynek szórása $\sigma = 2{,}0$ dkg volt. A véletlenszerűen kiválasztott 25 darabos minta átlaga $\overline{X} = 14{,}8$ dkg-nak adódott. Feltételezhető-e az eltérés véletlenszerűsége?

Megoldás

A nullhipotézis: $H_0 : M(X) = m_0 = 14{,}0$ dkg, és

$$|u_{sz}| = \left|\sqrt{n}\,\frac{\overline{X}-m_0}{\sigma}\right| = \left|\sqrt{25}\,\frac{14{,}8-14{,}0}{2}\right| = 2{,}0.$$

Az általában szokásos $p = 5\%$ -nak megfelelő

$$u_{p=0{,}05} = u_t = 1{,}96 \quad \left(\text{ui. } \Phi(u_t) = 1-\frac{p}{2} = 1-0{,}025 = 0{,}975\right)$$

Mivel

$$|u_{sz}| = 2{,}0 > u_t = 1{,}96,$$

ezért a nullhipotézist elvetjük, ugyanis a csokoládé szeletek várható súlyértéke 95%-os szinten szignifikánsan eltér az előírt súlyértéktől.

b) **A kétmintás *u*-próba.** Legyen X és Y két normális eloszlású valószínűségi változó, melyekhez $X_1, X_2, \ldots, X_k$ és $Y_1, Y_2, \ldots, Y_m$ független minták tartoznak, valamint szórásuk, σ_x és σ_y, ismert.

Ha

$$H_0 : M(X) = M(Y)$$

nullhipotézis teljesül a

$$H : M(X) \neq M(Y)$$

ellenhipotézissel szemben, akkor az

$$u = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_x^2}{k}+\frac{\sigma_y^2}{m}}} \tag{2}$$

valószínűségi változó (próbastatisztika) $N(0;1)$ eloszlású, ahol $\overline{X}$, $\overline{Y}$ a két mintából számított mintaközép. Ekkor a kétmintás u-statisztika u konfidencia-intervalluma $1-p$ valószínűséggel megadható, azaz

$$P(-u_p \le u \le u_p) = 1-p = 2\Phi(u_p)-1.$$

Az u_p értékét adott p-hez az 1. sz. táblázatból $\Phi(u_p) = 1-\frac{p}{2}$ összefüggés alapján kiszámíthatjuk. Ha a (2) formulával kiszámított u-statisztika nem esik a $[-u_p; u_p]$ intervallumba, akkor $1-p$ szignifikanciaszinten a nullhipotézist elvetjük, mivel ekkor a kiszámított u értéke a $[-u_p; u_p]$ intervallumon kívüli kritikus tartományba esik.

Példa

Tegyük fel, hogy két célgép, A és B, azonos henger alakú alkatrészeket gyárt. A hengerek átmérőjének mérete normális eloszlású. Az A gépen gyártottak átmérőjét jelölje X, a B gépen gyártottakét pedig Y. A szórásuk különbözők, de ismertek, azaz

$$D(X) = \sigma_x = 0{,}5 \text{ cm és } D(Y) = \sigma_y = 0{,}7 \text{ cm}.$$

Vizsgáljuk meg a $H_0 : M(X) = M(Y)$ nullhipotézist a $H : M(X) \neq M(Y)$ ellenhipotézissel szemben $1-p = 0{,}95$ szignifikanciaszinten. $\Phi(u_{p=0{,}05}) = 1-\frac{0{,}05}{2} = 0{,}975$ mivel $p = 0{,}05$, és az 1. sz. táblázatból $u_{p=0{,}05} = 1{,}96$.

Mind az A mind a B gép által gyártott hengerek közül 30-30 db-ot lemérve mintaközepekre $\overline{X} = 16{,}4$ cm és $\overline{Y} = 15{,}7$ cm adódott. A (2) formula szerint

$$u = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_x^2}{k}+\frac{\sigma_y^2}{m}}} = \frac{16{,}4-15{,}7}{\sqrt{\frac{(0{,}5)^2}{30}+\frac{(0{,}7)^2}{30}}} = 4{,}457.$$

Mivel a kiszámított u érték a táblázatból kiolvasott $u_{p=0{,}05} = 1{,}96$ értéknél nagyobb, vagyis kívül esik a $[-1{,}96; 1{,}96]$ intervallumon, ezért a H_0 nullhipotézist 95%-os szinten el kell vetnünk.

3.2. Egy- és kétmintás *t*-próba

a) **Egymintás *t*-próba.** A gyakorlati feladatok többségében a normális eloszlású valószínűségi változóra a várható érték is és a szórás is ismeretlen. Ilyen esetben az

$$s^{*2} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$$

korrigált szórásnégyzet kiszámítása után a

$$H_0 : M(X) = m_0$$

nullhipotézis ellenőrzésére a

$$t = \sqrt{n}\,\frac{\overline{X} - m_0}{s^{*2}} \qquad (1)$$

próbastatisztikát képezzük. Ez $n-1$ **szabadságfokú** (1-gyel csökkentett mintaelemszámú) **Student-eloszlású változó.** A *Student*-eloszlású táblázatból adott p-hez meghatározható az a $t_p = t_t$ (táblabeli) érték, amelyre

$$P\left(-t_p \le t \le t_p\right) = 1 - p \qquad (2)$$

ahol a p a valószínűségi szintet jellemző kis pozitív szám pl. $p = 0{,}05$ (5%).

Az u-próbához hasonlóan azt mondjuk, hogy $1-p$ szignifikanciaszinten elfogadjuk a H_0 hipotézist, ha

$$\left|t_{sz}\right| = \left|\sqrt{n}\,\frac{\overline{X} - m_0}{s^{*2}}\right| \le t_t = t_p \qquad (3)$$

és a H_0 hipotézist $1-p$ szinten elutasítjuk, ha

$$\left|t_{sz}\right| > t_t = t_p\,,$$

azaz a

$$H : M(X) = m \ne m_0$$

ellenhipotézist fogadjuk el. Az (1) próbastatisztikát *t-statisztikának* nevezzük.

Adott n mintaelemszám és adott p értékhez a 4. sz. táblázatból a t_p értéket az $n-1$ -edik sor p % oszlopának megfelelő t_t táblabeli értékkel vesszük fel.

Pl. ha mintaelemszám $n = 21$, $p = 0{,}02$, akkor a szabadságfok 20, tehát t_p értékét a 4. sz. táblázat 20. sorának 5% fejlécű oszlopából olvassuk ki, amely $t_{p=0{,}02} = t_t = 2{,}528$.

Példa

Egy konzervdoboztöltő adagolóautomata 1000 g anyag betöltésére van beállítva. Mintavétel során az alábbi értékeket kaptuk:

985	987	1003	993	996
991	994	1004	1002	985

Vizsgáljuk meg, hogy 95%-os biztonsági szinten teljesül-e a várható értékre az $m_0 = 1000$ g előírás, azaz $H_0 : M(X) = m_0 = 1000$ hipotézis.

Megoldás

Normális eloszlást feltételezve, *t*-próbával ellenőrizzük, hogy az eltérések csak a véletlennek tulajdoníthatók-e vagy szisztematikusak.

A minta adatai alapján $(n = 10)$:

$$\overline{X} = \frac{\sum_{i=1}^{10} x_i}{10} = 994{,}0; \quad s^{*2} = \frac{\sum_{i=1}^{10}\left(x_i - \overline{X}\right)^2}{n-1} = 52{,}222;$$

$$s^* = 7{,}226; \quad \sqrt{n} = \sqrt{10} = 3{,}162,$$

tehát

$$t_{sz} = \sqrt{n}\,\frac{\overline{X} - m_0}{s^*} = 3{,}162 \cdot \frac{994{,}0 - 1000}{7{,}226} = -2{,}626.$$

A *Student*-eloszlás 4. sz. táblázata alapján a $10-1 = 9$ szabadságfokú sorban a $p = 0{,}05$ értékhez $t_{p=0{,}05} = t_t = 2{,}267$ tartozik, és mivel

$$\left|t_{sz}\right| = 2{,}626 > t_t = t_{p=0{,}05} = 2{,}267,$$

ezért az automata nagy valószínűséggel nem működik jól, szignifikáns eltérés van 95%-os szinten, tehát a $H_0 : M(X) = m_0 = 1000$ nullhipotézist elvetjük.

A 99%-os szinten (szignifikanciaszinthez) a 9 szabadságfok mellett a $p = 0{,}01$ értékhez $t_p = t_t = 3{,}250 > \left|t_{sz}\right|$ adódik, tehát 99%-os statisztikai biztonsági szinten az eltérések véletlennek tulajdoníthatók, de a mintaelemszám nagyon kicsi ahhoz, hogy ezt elfogadhassuk.

b) **A kétmintás *t*-próba.** Gyakran két normális eloszlású, azonos szórású alapsokaság várható értékeinek egyenlőségét kell vizsgálnunk, vagy arra a kérdésre kell választ adnunk, hogy két azonos szórású minta azonos alapsokaságból származik-e?

Legyen $X \in N(m_1;\sigma)$ és $Y \in N(m_2;\sigma)$, ahol m_1 és m_2 a várható értékek, σ a szórás, továbbá $X_1, X_2, ..., X_n$ az X valószínűségi változóhoz, és $Y_1, Y_2, ..., Y_k$ pedig az Y valószínűségi változóhoz tartozó egymástól független minták.

A nullhipotézis:

$$H_0 : M(X) = M(Y), \quad \text{azaz} \quad m_1 = m_2;$$

az ellenhipotézis pedig:

$$H : M(X) \neq M(Y), \quad \text{azaz} \quad m_1 \neq m_2.$$

Jelölje s_n^2 az X, s_k^2 pedig az Y valószínűségi változó mintaelemeiből képzett empirikus szórásnégyzetet. Képezzük a

$$t = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{(n-1)s_n^2 + (k-1)s_k^2}{n+k-2}}\sqrt{\frac{1}{n} + \frac{1}{k}}} \tag{4}$$

statisztikát, amely H_0 fennállása esetén $n+k-2$ szabadságfokú *Student*-eloszlású változó. Jelölje $|t_{sz}|$ a (4) képlettel számított abszolút értéket, azaz $|t_{sz}| = |t|$, a *Student*-eloszlás $n+k-2$ szabadságfokú táblázatbeli értéket pedig $t_t = t_p$, akkor

$$P(-t_p \leq t \leq t_p) = 1 - p.$$

Ha

a) $|t_{sz}| \leq t_t$ esetén (a számított érték a $[-t_p, t_p]$ intervallumba esik) a H_0 hipotézist $100(1-p)\%$ -os szinten elfogadjuk ($m_1 = m_2$);

b) $|t_{sz}| > t_t$ esetén pedig (a számított érték nem esik a $[-t_p, t_p]$ intervallumba) a H_0 hipotézist $100(1-p)\%$ -os szinten elvetjük ($m_1 \neq m_2$).

Példa

Legyenek az X és Y normális valószínűségi változókra vonatkozó egymástól független minták mintaszáma 21, ill. 41, mintaközepei: $\overline{X} = 25{,}4$, $\overline{Y} = 25{,}2$, szórásuk: $D(X) = D(Y) = 0{,}6$. Vizsgáljuk meg, hogy 95%-os biztonsági szinten egyenlőknek tekinthetők-e a várható értékek.

Megoldás

A nullhipotézis: $H_0 : M(X) = m_x = M(Y) = m_y$.

Az ellenhipotézis: $H : M(X) \neq M(Y)$, azaz $m_x \neq m_y$.

A (4) formula alapján kiszámítjuk t értékét:

$$t = \frac{25{,}4 - 25{,}2}{\sqrt{\frac{20 \cdot 0{,}36 + 40 \cdot 0{,}36}{60}} \cdot \sqrt{\frac{1}{21} + \frac{1}{41}}} = 1{,}242.$$

A *Student*-eloszlás 4. sz. táblázatában a 60 szabadságfokú sorban a $p = 0{,}05$ (5% oszlopfejű) értékhez $t_{p=0{,}05} = t_t = 2{,}000$ tartozik, és mivel

$$|t_{sz}| = 1{,}242 < t_t = t_{p=0{,}05} = 2{,}000,$$

ezért a két valószínűségi változó várható értékének egyenlőségét 95%-os szinten elfogadjuk, az ellenhipotézist pedig elvetjük.

3.3. A Welch-próba

Legyen X és Y két független normális eloszlású valószínűségi változó. Ha ismeretlenek a szórások, de feltehető, hogy azonosak, akkor a várható értékük összehasonlítására alkalmazhatjuk. Ha szórásuk ismeretlen, de különböző akkor a kétmintás t-próba a várható értékek összehasonlítására nem alkalmazható. Szükségünk van olyan próbára, amely akkor is alkalmas

$$H_0 : M(X) = M(Y)$$

hipotézis eldöntésére, ha σ_X, σ_Y ismeretlen és $\sigma_X \neq \sigma_Y$. *B. L. Welch* a kétmintás u-próba formulájához hasonló próbastatisztikát javasolt, azzal a különbséggel, hogy az ottani elméleti szórások helyére az empirikus szórásokat helyettesítette:

$$w = \frac{\overline{X}_n - \overline{Y}_m}{\sqrt{\frac{s_{xn}^2}{n} + \frac{s_{ym}^2}{m}}}, \tag{1}$$

ahol $\overline{X}_n$, $\overline{Y}_m$ a mintákból számított mintaközepek, n, m a minták elemszámai, s_{xn}^2, s_{ym}^2 pedig a mintákból számított empirikus szórásnégyzetek.

Igazolható, hogy a w változó eloszlása jól közelíthető f szabadságfokú t-eloszlással. Az f szabadságfokra teljesül a

$$\min(n,m) \le f \le n+m$$

egyenlőtlenség.

Az f kiszámítására javasolt közelítő képletek:

$$c = \frac{s_{xn}^2}{\sqrt{\frac{x_{xn}^2}{n} + \frac{x_{ym}^2}{m}}} \text{ felhasználásával } f = \frac{1}{\frac{c^2}{m-1} + \frac{(1-c)^2}{n-1}};$$

ill.

$$f = \frac{\left(\frac{s_{xn}^2}{n} + \frac{s_{ym}^2}{m}\right)^2}{\frac{s_{xn}^4}{n^2-1} + \frac{s_{ym}^4}{m^2-1}}.$$

A *Student*-eloszlás táblázatából adott f szabadságfokhoz, az adott p értéknek megfelelő táblabeli értéket kiolvassuk. Ha a 4. sz. táblázat megfelelő szabadságfokú sorának p oszlopából kiolvasott táblázatbeli értéket t_p -vel jelöljük, és

$$|w| \le t_p,$$

akkor a hipotézist p szinten elfogadjuk, ha

$$|w| \ge t_p,$$

akkor p szinten elvetjük.

Példa

Egy üzem férfi és nő dolgozóinak testsúlyát lemérve, a súlycsoportonkénti eredményt a következő oldali táblázat tartalmazza. Vizsgáljuk meg, hogy a két sokaság várható értéke megegyezik-e.

Osztályköz	Osztályközép	Gyakoriság: ffi f_i	Gyakoriság: nő $\tilde{f}_i$
40–44	42	57	62
45–49	47	63	71
50–54	52	78	80
55–65	60	90	85
66–70	68	84	72
71–80	75,5	72	66
81–90	85,5	30	13
90–	95	10	10
		$n = 484$	$m = 459$

Megoldás

A hipotézis: $H_0: M(X) = M(Y)$.

A mintaközepek:

$$\overline{X}_n = \frac{\sum_{i=1}^{n} f_i x_i}{n} = 60{,}8967, \quad \overline{Y}_m = \frac{\sum_{i=1}^{n} \tilde{f}_i x_i}{m} = 59{,}1318.$$

Az empirikus szórásnégyzetek:

$$s_{xn}^2 = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \overline{X}_n)^2}{n-1} = 182{,}5038,$$

$$s_{ym}^2 = \frac{\sum_{i=1}^{k} \tilde{f}_i \cdot (x_i - \overline{Y}_m)^2}{m-1} = 170{,}3685.$$

Welch-próba: $$w = \frac{\overline{X}_n - \overline{Y}_m}{\sqrt{\frac{s_{xn}^2}{n} + \frac{s_{ym}^2}{m}}} = 2{,}0403.$$

A w közelítőleg $459 \le f \le 459 + 484$ egyenlőtlenségnek eleget tevő f szabadságfokú t eloszlást követ. A p értékét 0,02-nak választva, a 4. sz. táblázat 0,02 oszlopának utolsó sorában a táblabeli érték: $t_{0,02} = 2{,}326$. Mivel $w = 2{,}0403 < t_{0,02} = 2{,}326$, így a hipotézist 98%-os szinten elfogadjuk. Ugyanakkor, ha p értékét 0,10-nek választjuk, akkor $t_{0,10} = 1{,}645$, vagyis $w > t_{0,10} = 1{,}645$, tehát 90%-os szinten a hipotézist elvetjük.

3.4. Az F-próba

A kétmintás t-próba alkalmazásakor feltettük, hogy a minták ismeretlen szórásai azonosak. Az **F-próba** alkalmazásával dönthetjük el, hogy két normális eloszlású, ismeretlen várható értékű statisztikai sokaság szórásnégyzete ill. szórása azonos-e vagy sem.

Legyen $X \in N(m_1;\sigma_1)$ és $Y \in N(m_2;\sigma_2)$, ahol m_1 és m_2 a várható értékek, σ_1 és σ_2 a szórások, továbbá $X_1, X_2, ..., X_n$ az X-hez és $Y_1, Y_2, ..., Y_k$ az Y-hoz tartozó egymástól független minták. A nullhipotézis:

$$H_0 : D^2(X) = D^2(Y), \quad \text{azaz} \quad \sigma_1^2 = \sigma_2^2 \quad \text{ill.} \quad \sigma_1 = \sigma_2;$$

az ellenhipotézis pedig:

$$H : D^2(X) \neq D^2(Y), \quad \text{azaz} \quad \sigma_1^2 \neq \sigma_2^2 \quad \text{ill.} \quad \sigma_1 \neq \sigma_2.$$

Jelölje s_n^2 az X, s_k^2 pedig az Y minta empirikus szórásnégyzetét. Jelölje $s_{\max}^2$ a két szórásnégyzet közül a nagyobbat és $s_{\min}^2$ a kisebbet, azaz

$$s_{\max}^2 = \max(s_n^2, s_k^2); \qquad s_{\min}^2 = \min(s_n^2, s_k^2).$$

Az $s_{\max}^2$ szabadságfoka legyen f_1, az $s_{\min}^2$ szabadságfoka legyen f_2. Képezzük az

$$F = \frac{s_{\max}^2}{s_{\min}^2} \tag{5}$$

statisztikát, amely (f_1 ; f_2) szabadságfokú F-eloszlás. Az (5) formulába helyettesített értékekkel képezzük az F_{sz} számított értéket és összehasonlítjuk az F-táblázat $100(1-p)\%$ szignifikancia szintű f_2-edik sorának f_1-edik oszlopából vett F_t értékkel. (Az 5. táblázat értékei csak $p = 0{,}05$-ra, azaz 95%-os biztonsági szinthez vannak kiszámítva. f_1 a nagyobb empirikus szórású, f_2 pedig a kisebb empirikus szórású változó mintaelemszámának 1-gyel csökkentett értéke. Pl. ha $s_n^2 > s_k^2$, akkor $f_1 = n-1$, és $f_2 = k-1$.)

Döntés, ha

a) $F_{sz} \le F_t$, akkor a H_0 hipotézist $100(1-p)\%$-os szinten elfogadjuk ($\sigma_1 = \sigma_2$); az eltérést véletlennek tulajdonítjuk,

b) $F_{sz} > F_t$, akkor pedig a H_0 hipotézist $100(1-p)\%$-os szinten elvetjük ($\sigma_1 \neq \sigma_2$), az eltérést szignifikánsnak tekintjük.

Példa

Egy konzervgyár mindkét próbaüzembe állított gépén 300 grammos konzervek gyártását kezdte meg. Legyen a gépenként véletlenszerűen kiválasztott 6-6 doboz töltősúlya az X-re ill. az Y-ra vonatkozó 6 elemű minta:

X elemei:	300	301	303	288	294	296
Y elemei:	305	317	308	300	314	316.

Ellenőrizzük, hogy teljesül-e a $H_0: D(X) = D(Y)$ nullhipotézis.

Megoldás

Mintaátlagok: $\overline{X} = 297$; $\overline{Y} = 310$;

Az empirikus szórásnégyzetek: $s_1^2 = 38{,}4$; $s_2^2 = 46$.

Alkalmazzuk az F-próbát, $n-1 = 6-1 = 5$; $k-1 = 6-1 = 5$ szabadságfokkal:

$$F_{sz} = F = \frac{s_2^2}{s_1^2} = 1{,}2.$$

Az 5. sz. táblázatból $p = 0{,}05$ értékhez, 5%-os szinthez tartozó táblázatbeli (kritikus) érték (5. sor 5-ös fejlécű oszlop): $F_t = 5{,}05$. Mivel $F_{sz} < F_t$ $(1{,}2 < 5{,}05)$, ezért a nullhipotézist 95%-os szinten elfogadjuk, ami azt jelenti, hogy a szórások (szórásnégyzetek) közötti különbség véletlennek tulajdonítható.

3.5. Illeszkedés- és homogenitásvizsgálat

A matematikai statisztikában gyakran olyan próbát kell alkalmaznunk, amellyel megvizsgálhatjuk, hogy valamely minta származhat-e egy bizonyos elméleti valószínűségeloszlásból, melyet paramétereivel együtt megadunk. Az ilyen statisztikai próbát **tiszta illeszkedésvizsgálatnak** nevezzük. Ha a vizsgálat csak arra vonatkozik, hogy valamilyen típusú eloszlás feltételezhető-e, az eloszlás paramétereit a mintából becsülve, akkor **becsléses illeszkedésvizsgálatokat** alkalmazunk. Legtöbbször ez utóbbiakat használjuk, ezek közül is a leggyakoribb a **normalitásvizsgálat,** a normális eloszláshoz való illeszkedés vizsgálata.

A **homogenitásvizsgálat** is statisztikai próba, éspedig annak megvizsgálására vonatkozik, hogy bizonyos minták azonos alapsokaságból származnak-e, két vagy több valószínűségi változó eloszlása megegyezik-e? Általánosabban homogenitásvizsgálatnak nevezzük az olyan statisztikai próbát is, amellyel csak azt vizsgáljuk, hogy a megfelelő alapsokaságok bizonyos paramétereikben megegyeznek-e.

A fenti vizsgálatokhoz számos eljárást dolgoztak ki, mi itt az ún. χ^2-*(khi-négyzet)-próba* alkalmazását mutatjuk be.

3.5.1. Illeszkedésvizsgálat

Tegyük fel, hogy egy X valószínűségi változó eloszlása ismert. Azt akarjuk eldönteni, hogy egy n elemű minta eloszlása megfelel-e az ismert eloszlásnak. A próba az ismert eloszlás alapján várható gyakoriságok és a minta gyakorisága közötti eltérések vizsgálatából áll.

Diszkrét valószínűségi változó esetén
a nullhipotézis:

$$H_0 : P(X = x_i) = p_i, \quad (i = 1, 2, \ldots, r);$$

az ellenhipotézis:

$$H : \exists i, \text{amelyre } P(X = x_i) \neq p_i.$$

Folytonos eloszlás esetén
a nullhipotézis:

$$H_0 : P(X \le x) = F(x);$$

az ellenhipotézis:

$$H : P(X \le x) \neq F(x),$$

ahol $F(x)$ ismert eloszlás.

Folytonos eloszlásnál a valószínűségi változó értékeit soroljuk be nagyság szerint r osztályba, miként azt a hisztogram készítésénél is tettük. Az n adatból az i-edik osztályba eső mintaelemek számát, a gyakoriságot jelöljük f_i-vel, akkor $g_i = \frac{f_i}{n}$ jelenti relatív gyakoriságát. Az ismert (feltételezett) eloszlás alapján az elméleti gyakoriság np_i. A minta alapján számított f_i gyakoriságok és az elméleti gyakoriságok általában különböznek egymástól. Ezt az eltérést a

$$\chi^2 = \sum_{i=1}^{r} \frac{(f_i - np_i)^2}{np_i} = n\sum_{i=1}^{r} \frac{(g_i - p_i)^2}{p_i} \qquad (1)$$

próbastatisztikával mérjük. Bizonyítható, hogy ha n minden határon túl nő, akkor az (1) statisztika eloszlása tiszta illeszkedésvizsgálat esetén az $r-1$ szabadságfokú χ^2-eloszláshoz, és becsléses illeszkedésvizsgálat esetén, ha a becsült paraméterek száma k, az $r-1-k$ szabadságfokú χ^2-eloszláshoz tart. Az (1) statisztika tehát közelítőleg χ^2-eloszlású. A közelítés elég jó, ha $np_i \ge 10$ minden i-re.

Az elfogadási tartományt úgy adjuk meg, hogy tiszta illeszkedés esetén

$$P(\chi^2_{sz} \le \chi^2_{p,r-1}) = 1 - p,$$

becsléses illeszkedés esetén pedig

$$P(\chi^2_{sz} \le \chi^2_{p,r-1-k}) = 1 - p$$

legyen, ahol a χ^2_{sz} az (1) alapján számított, $\chi^2_{p,r-1}$ és $\chi^2_{p,r-1-k}$ a 3. sz. táblázatból kiolvasott érték, p előre megadott 1-nél nem nagyobb pozitív szám. Tehát,

a) ha $\chi^2_{sz} \le \chi^2_{p,r-1}$ ill. $\chi^2_{sz} \le \chi^2_{p,r-1-k}$, akkor a nullhipotézist elfogadjuk, a minta eloszlása az adott biztonsági szinten megfelel az elméleti eloszlásnak, az eltéréseket véletlenszerűnek ítéljük;

b) ha $\chi^2_{sz} > \chi^2_{p,r-1}$ ill. $\chi^2_{sz} > \chi^2_{p,r-1-k}$, akkor a nullhipotézist az adott biztonsági szinten elvetjük, a minta eloszlása szignifikánsan eltér az elméleti eloszlástól.

Megjegyzés

Ha azt akarjuk eldönteni, hogy a mintasokaság normális eloszlású-e vagy sem, akkor a vizsgálatot **normalitásvizsgálatnak** nevezzük és az előbbiekkel azonos módon járunk el.

3.5.2. Homogenitásvizsgálat

Legyen *X* és *Y* két tetszőleges eloszlású valószínűségi változó. Azt akarjuk megvizsgálni, hogy *X* és *Y* azonos eloszlásúnak tekinthető-e.

A nullhipotézis, tehát

$$H_0 : P(X \le x) = P(Y \le x);$$

az ellenhipotézis pedig

$$H : P(X \le x) \ne P(Y \le x).$$

Az eljárás alkalmazása a következő lépésekből áll:

Osszuk fel a számegyenest *r* számú osztályra:

$$-\infty = b_0 < b_1 < b_2 < \ldots < b_r = \infty .$$

Legyen az *i*-edik intervallumba eső mintaelemek száma *X*-re f_i, *Y*-ra v_i $(i = 1, 2, \ldots, r)$.

Legyen az *X* változóhoz tartozó mintaelemszám: $\sum_{i=1}^{r} f_i = n$, és az *Y* változóhoz tartozó mintaelemszám: $\sum_{i=1}^{r} v_i = m$. Igazolható, hogy a

$$\chi^2 = mn \sum_{i=1}^{r} \frac{\left(\frac{f_i}{n} - \frac{v_i}{m}\right)^2}{f_i + v_i} \qquad (2)$$

próbastatisztika $n \to \infty$ és $m \to \infty$ esetén $r-1$ szabadságfokú χ^2-eloszlású valószínűségi változó.

Ebből következik, hogy ha az *n* és *m* elég nagy, akkor χ^2-próba alkalmazható a két eloszlás megegyezőségének vizsgálatára.

A nullhipotézist elfogadjuk, ha $\chi^2_{sz} \le \chi^2_{p,r-1}$, és elvetjük, ha $\chi^2_{sz} > \chi^2_{p,r-1}$, ahol χ^2_{sz} a (2) képlettel számított érték, $\chi^2_{p,r-1}$ pedig a χ^2 3. sz. táblázatában *p* szinten található érték (szabadságfok: $r-1$).

Példák

1. Egy kockát 1400-szor feldobva az egyes számok gyakoriságára a következő táblázatot kaptuk:

Dobott számok	gyakoriság (k_i)
1–es	228
2–es	240
3–as	224
4–es	237
5–ös	235
6–os	236

Kérdés: szabályosnak tekinthető-e a kocka?

Megoldás

Mivel szabályos kockánál bármely szám dobása egyenlő valószínűségű, ezért a nullhipotézis:

$$H_0 : \ P(X = x_i) = p_i = \frac{1}{6} \quad (i = 1, 2, \ldots, 6).$$

Az *X* egyenletes eloszlású, becsülendő paraméter nincs. A feladatot χ^2-próbával oldjuk meg. A χ^2 kiszámítását

$$\chi^2 = \sum_{i=1}^{r} \frac{(k_i - np_i)^2}{np_i} = \frac{1}{n} \sum_{i=1}^{r} \frac{k_i}{p_i} - 2\sum_{i=1}^{r} k_i + n \sum_{i=1}^{r} p_i =$$

$$= \frac{1}{n} \sum_{i=1}^{r} \frac{k_i^2}{p_i} - 2n + n = \frac{1}{n} \sum_{i=1}^{r} \frac{k_i^2}{p_i} - n$$

képlettel végezzük.

$$\chi_{sz}^2 = \frac{1}{n}\sum_{i=r}^{r}\frac{k_i^2}{p_i} - n = \frac{1}{1400}\cdot\sum_{i=1}^{6}\frac{k_i^2}{\frac{1}{6}} - 1400 =$$

$$= \frac{6}{1400}(228^2 + 240^2 + 224^2 + 237^2 + 235^2 + 236^2) - 1400 =$$

$$= \frac{6}{1400}\cdot 326850 - 1400 = 0{,}7857.$$

A szabadságfokok száma: $r-1 = 6-1 = 5$. A 3. sz. táblázatból, ha $p = 0.05$ értékkel adott, akkor $\chi_{95}^2 = 11{,}070$ tehát

$$\chi_{sz}^2 = 0{,}7857 < \chi_{95}^2 = 11{,}070$$

vagyis a nullhipotézis elfogadható, a kocka szabályosnak tekinthető.

A χ^2-próba alkalmazhatóságának a feltétele teljesült, hiszen az np_i érték minden i-re lényegesen nagyobb 10-nél $(np_i = 1400\frac{1}{6} \approx 233)$.

2. Egy adagolóautomata működésének vizsgálatához 174 db doboz mérlegelését végezték el. A grammokban adott névleges tömegértéktől való eltérés értékeit osztályba sorolva az alábbi táblázat tartalmazza:

Osztályok $x_{i-1};-x_i$	Gyakoriság f_i
……–1,5	12
1,51–3,5	25
1,51–4,5	22
4,51–5,5	24
5,51–7,5	35
7,51–9,5	26
9,51–11,5	19
11,51–……	11

Feltehető-e, hogy az előírt tömegértéktől való eltérések normális eloszlásúak?

Megoldás

A feladathoz becsléses illeszkedésvizsgálatot kell végeznünk, és a χ^2-próbát alkalmazzuk.

A nullhipotézis: $H_0: P(X < x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}}\,dt = F(x);$

az ellenhipotézis: $H: P(X < x) \neq F(x).$

A várható értéket

$$m \approx \overline{X} = \frac{\sum_{i=1}^{n} f_i k_i}{n};$$

a szórásnégyzetet

$$s^{*2} = \frac{1}{n-1}\sum_{i=1}^{n} f_i k_i^2 - \frac{n}{n-1}\overline{X}^2$$

formulákkal becsüljük.

A p_i értékek kiszámításához a

$$p_i = P(x_{i-1} \le X < x_i) = F(x_i) - F(x_{i-1}) = \Phi\left(\frac{x_i - m}{\sigma}\right) - \Phi\left(\frac{x_{i-1} - m}{\sigma}\right)$$

és

$$\Phi(-u) = 1 - \Phi(u)$$

képleteket használjuk. Amennyiben $np_i \ge 10$ $(i = 1, 2, \dots, r)$, akkor kiszámítjuk a

$$\chi^2 = \sum_{i=1}^{r}\frac{(f_i - np_i)^2}{np_i}$$

értéket. A szabadságfokok száma: $r - 1 - k = 8 - 1 - 2 = 5$.

A számítást célszerű a következő táblázatos formában végezni (a $\Phi(u)$ értékek kiszámításánál lineáris interpolációt használtunk):

osztályok $x_{i-1};x_i$	gyak. f_i	középp k_i	$f_i k_i$	$f_i k_i^2$	u	$\Phi(u)$	p_i	np_i	$\frac{(f_i - np_i)^2}{np_i}$
–1,5	12	1	12,0	12,00	–1,420	0,0690	0,0690	12,0060	0,00003
1,5–3,5	25	2,5	62,5	156,25	–0,796	0,2071	0,1381	24,0294	0,0392
3,5–4,5	22	4	88,0	352,00	–0,480	0,3156	0,1085	18,8790	0,5159
4,5–5,5	24	5	120,0	600,00	–0,166	0,4341	0,1185	20,6190	0,5544
5,5–7,5	35	6,5	227,5	1478,75	0,461	0,6776	0,2435	42,3690	1,2816
7,5–9,5	26	8,5	221,0	1878,50	1,088	0,8617	0,1841	32,0334	1,1364
9,5–11,5	19	10,5	199,5	2094,75	1,715	0,9569	0,0952	16,3908	0,4154
11,5–	11	12	132,0	1584,00	∞	1,0000	0,0431	7,6734	1,4422
Σ	174		1062,5	8156,25			1,0000	174,00	$\chi^2 = 5{,}3851$

ahol

$$u = \frac{x_i - \overline{X}}{s^*},\quad m \approx \overline{X} = \frac{\sum_{i=1}^{8} f_i k_i}{174} = \frac{1062{,}5}{174} = 6{,}106.$$

$$s^{*2} = \frac{1}{n-1}\sum_{i=1}^{8} f_i k_i^2 - \frac{n}{n-1}\overline{X}^2 = \frac{1}{173}\cdot 8156{,}25 - \frac{174}{173}\cdot 6{,}106^2 = 9{,}6433;$$

és így

$$s^* = \sqrt{9{,}6433} \approx 3{,}105.$$

Pl.

$$p_1 = \Phi\left(\frac{1{,}5 - 6{,}106}{3{,}105}\right) - \Phi(-\infty) = 1 - \Phi(1{,}4833).$$

Az 1. sz. táblázatból

$$\Phi(1{,}4833) = 0{,}9310,$$

tehát

$$p_1 = 1 - 0{,}9310 = 0{,}0690.$$

Mivel a 95%-os szinthez az 5 szabadságfokú χ^2-táblázatban a 11,07 érték tartozik, melynél a számított $\chi^2_{sz} = 5{,}3851$ lényegesen kisebb, ezért a nullhipotézist elfogadjuk, vagyis az eltérésértékek normális eloszlásúnak tekinthetők. (Az np_i értékek egy kivételével nagyobbak 10-nél, ettől az egy kivételtől eltekintünk.)

3.6. Függetlenségvizsgálat χ^2-próbával

Egy férficsoport vizsgálatát pl. magasság és testsúly szerint végezzük. Jelölje X valószínűségi változó aktuális értéke egyedenként a magasságot, Y pedig ugyanarra az egyedre a testsúlyt. Arra szeretnénk választ kapni, hogy a két tulajdonság, az adott statisztikai sokaságra vonatkozóan, függetlennek tekinthető-e. Általában, ha egyugyanazon statisztikai sokaság két valószínűségi változó által adott két tulajdonságának függetlenségvizsgálatát végezzük, mindig ilyen összetartozó értékpárok az alapadatok.

Legyen az X valószínűségi változóhoz tartozó teljes eseményrendszer: $X_1, X_2, \ldots, X_k$, az Y valószínűségi változóhoz tartozó pedig $Y_1, Y_2, \ldots, Y_m$. A szorzási tétel értelmében felírható a nullhipotézis:

$$H_0 : P(X_i \cap Y_j) = P(X_i)\cdot P(Y_j) \ \forall (i, j)\text{-re}$$

$$(i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, m).$$

A mintaelemekből mind az X mind az Y valószínűségi változóhoz a vizsgált két tulajdonságnak megfelelően egy-egy érték rendelhető. Mindegyik mintaelem a két tulajdonság szerint osztályozható. Ha az X_i és az Y_j esemény együttesen következett be, akkor azt mondjuk, hogy az $(X_i \cap Y_j)$ esemény következett be. Ha a mintaelemek számát n-nel, a $(X_i \cap Y_j)$ esemény bekövetkezésének gyakoriságát f_{ij}-vel jelöljük, akkor ezeket az értékeket ún. *kontingenciatáblázatba* rendezzük.

Az f_{ij} gyakorisági értékek soronkénti összege:

$$v_{ij} = \sum_{j=1}^{m} f_{ij},\ (i = 1, 2, \ldots, k),$$

oszloponkénti összeg pedig:

$$u_{ij} = \sum_{i=1}^{k} f_{ij},\ \ (j = 1, 2, \ldots, k).$$

A sor- és oszlopösszegeket *marginális értékeknek* nevezzük.

A definiált értékekből következik, hogy

$$\sum_{i=1}^{k}\sum_{j=1}^{m} f_{ij} = n, \text{ ill. } \sum_{i=1}^{k} v_{ij} = n, \ \ \sum_{j=1}^{m} u_{ij} = n.$$

A kontingenciatáblázatot tehát a következő elrendezésben készítjük el:

X\Y	Y_1	Y_2	...	Y_m	$\sum$ sor
X_1	f_{11}	f_{12}	...	f_{1m}	v_{1j}
X_2	f_{21}	f_{22}	...	f_{2m}	v_{2j}
⋮	⋮	⋮	⋮	⋮	⋮
X_k	f_{k1}	f_{k2}	...	f_{km}	v_{kj}
$\sum$ oszlop	u_{i1}	u_{i2}	...	u_{im}	n

A χ^2-próbához ki kell számítani az elméleti gyakoriságok értékeit az

$$\tilde{f}_{ij} = \frac{u_{i1} \cdot v_{1j}}{n} \qquad (1)$$

formulával minden i-re és j-re. Célszerű ezeket az értékeket a táblázatba f_{ij} értékei mellé írni.

A *becsléses függetlenségvizsgálat* valószínűségi változója:

$$\chi^2_{sz} = \sum_{i=1}^{k}\sum_{j=1}^{m} \frac{(f_{ij} - \tilde{f}_{ij})^2}{f_{ij}}. \qquad (2)$$

Az (2) formulával számított valószínűségi változó $n \to \infty$ esetén $(k-1)\cdot(m-1)$ szabadságfokú χ^2 eloszlású. Ha a mintából meghatároztuk a χ^2_{sz} értékét, majd kikeressük a 3. sz. táblázatból adott $100(1-p)\%$-os valószínűségi szint és $(k-1)\cdot(m-1)$ szabadságfok mellett a táblabeli χ^2_t értéket, akkor a függetlenségre vonatkozó H_0 hipotézist $\chi^2_{sz} \le \chi^2_t$ esetben elfogadjuk, $\chi^2_{sz} > \chi^2_t$ esetben pedig $100(1-p)\%$-os szinten elvetjük.

Példa

A PB gázpalack elosztóhoz négy töltőállomás (Alfa, Béta, Gamma, Delta) szállítja a 11 kg névleges töltő súlyú palackokat. Az elosztó súlyvizsgálatot tartott véletlenszerűen kiválasztott 182 palack lemérésével. Az osztályozást töltőállomásonként és adott súlyhatárok szerint végezve az eredményt, a következő táblázat tartalmazza:

Súly kg	Alfa	Béta	Gamma	Delta	$\sum$sor
10,8–11,2	25	30	22	23	100
10,3–10,7	20	10	16	18	64
x<10,3	5	3	6	4	18
$\sum$oszlop	50	43	44	45	182

Állapítsuk meg χ^2 próbával, hogy a töltőállomásoktól függetlenek-e a súlyeltérések.

Megoldás

A nullhipotézis: az osztályozás szerinti két valószínűségi változó független.

Számítsuk ki az (1) formulával az elméleti gyakoriság $\tilde{f}_{ij}$ értékeit és írjuk be a táblázatba: $\tilde{f}_{11} = \frac{100\cdot 50}{182} = 27{,}47$; $\tilde{f}_{12} = \frac{100\cdot 43}{182} = 23{,}63$; stb.

Súly kg	Alfa		Béta		Gamma		Delta		$\sum$sor	
10,8-11,2	27,5	25	23,6	30	24,2	22	24,7	23	100	100
10,3-10,7	17,6	20	15,1	10	15,4	16	15,8	18	64	64
x<10,3	4,9	5	4,3	3	4,4	6	4,5	4	18	18
$\sum$oszlop	50	50	43	43	44	44	45	45	182	182

Képezzük a (2) formula szerint χ^2_{sz} értékét:

$$\chi^2_{sz} = \sum_{i=1}^{k}\sum_{j=1}^{m} \frac{(f_{ij} - \tilde{f}_{ij})^2}{f_{ij}} = \frac{(25-27{,}5)^2}{27{,}5} + \frac{(30-23{,}6)^2}{23{,}6} + \ldots +$$

$$+ \frac{(6-4{,}4)^2}{4{,}4} + \frac{(4-4{,}5)^2}{4{,}5} = 6{,}656.$$

A szabadságfokok száma: $(k-1)(m-1) = (3-1)(4-1) = 2\cdot 3 = 6$. a 3. sz. táblázat 6. sorában a $p_1 = 0{,}10$-hez $\chi^2_t = 10{,}645$. a $p_2 = 0{,}95$ értékhez $\chi^2_t = 1{,}635$. A nullhipotézist, vagyis a két tulajdonság függetlenségét $p_1 = 0{,}10$ választás mellett elfogadjuk, mivel

$$\chi^2_{sz} = 6{,}656 < \chi^2_{0,10} = 10{,}645,$$

míg

$$p_2 = 0{,}95$$

választás esetén elvetjük, mivel

$$\chi^2_{sz} = 6{,}656 > \chi^2_{0,95} = 1{,}635.$$

E.3. Ellenőrző kérdések a 3. fejezethez

1. Mit nevezünk nullhipotézisnek és ellenhipotézisnek?

2. Hogyan értelmezzük a statisztikai próbák elfogadási és kritikus tartományát?

3. Mit nevezünk elsőfajú és másodfajú hibának?

4. Hogyan alkalmazzuk az u-próbát?

5. Hogyan alkalmazzuk az egy- és kétmintás t-próbát?

6. Milyen döntéshez használjuk az F-próbát?

7. Mit nevezünk tiszta- és becsléses illeszkedésvizsgálatnak?

8. Mire vonatkozik a homogenitásvizsgálat?

S.3. Feladatok a 3. fejezethez

S.3.1. Egy adagolóautomata 1200 g tömegű massza töltésére van beállítva. A dobozokba 1200 g-tól eltérő tömegű massza is betöltésre kerülhet. A massza tömegének értéke véletlentől függő változó, az előzetes méréssorozatok azt igazolták, hogy normális eloszlású $\sigma = 3\,\text{g}$ szórással. A mintát véletlenszerűen kiválasztott 16 db doboz alkotja, amelyeknek a mért értéke:

1193	1198	1203	1191
1195	1196	1199	1191
1201	1196	1193	1198
1204	1196	1198	1200

Van-e szignifikáns eltérés 99,9%-os megbízhatósági szinten az előírt tömegértéktől?

S.3.2. Egy csomagológép által készített bálákat vizsgálunk. 100 bála mérlegelése alapján azt kaptuk, hogy a bálák átlagos tömege 705 kg. Legyen a bálák tömege normális eloszlású, $\sigma = 50\,\text{kg}$ szórással. Van-e szignifikáns eltérés 95%-os megbízhatósági szinten az előírt 700 kg-hoz képest?

S.3.3. Egy adagológép által dobozba töltendő anyag tömegének várható értékére az előírás 500 g. Mintavétel során az alábbi értékeket kaptuk:

483	498	502	494	505
502	496	483	491	486

Teljesül-e a várható értékre az $m_0 = 500\,\text{g}$ előírás 95%-os biztonsági szinten?

S.3.4. A szénosztályozóba érkező vagonok meddőeloszlásáról feltesszük, hogy normális eloszlású. 142 vagont megvizsgálva, osztályok szerint rendezve a következő adatokat kaptuk:

i	osztályok $x_{i-1}; x_i$	gyakoriság f_i
1	– 243	9
2	244 – 323	18
3	324 – 403	23
4	404 – 483	27
5	484 – 563	18
6	564 – 643	15
7	644 – 723	12
8	724 –	20
Σ		142

Vizsgáljuk meg, hogy a meddő mennyisége normális eloszlású-e.

NEGYEDIK FEJEZET

Empirikus képletek előállítása, korreláció- és regressziószámítás

A matematikának a természettudományok, a műszaki tudományok, a közgazdaságtudományok területein történő alkalmazásai során gyakran közvetlen megfigyelés, mérés útján nyert tapasztalati adatok között, a vizsgált folyamatra jellemző mennyiségi összefüggések elemzését kell elvégezni. Rendszerint ezeket az összefüggéseket bizonyos feltételek fennállása esetén is csak közelítő pontossággal lehet kifejezni. Az ilyen típusú mennyiségi összefüggéseket számításra alkalmas ún. **empirikus képletekkel** adjuk meg. A tudományok fejlődésének kezdeti időszakában pl. a legtöbb természeti törvényt is egy, a kísérleti adatoknak megfelelő empirikus képletbe való foglalásával alkották meg. Célunk, hogy lehetőleg egyszerű képlettel kifejezzük az összetartozó (x_k, y_k) $(k = 1, 2, \ldots, n)$ tapasztalati eredmények közötti kapcsolatot. A probléma megoldása általában két lépésből áll:

1. Megállapítjuk az empirikus képlet típusát,

2. Meghatározzuk a választott típusú képletben szereplő paramétereket a megfigyelés adatai alapján.

4.1. Az empirikus képlet kiválasztása

Az empirikus képlet típusának meghatározásakor arra törekszünk, hogy annak grafikonja jól illeszkedjék a tapasztalati adatoknak a koordinátarendszerben ábrázolt pontjai közé és a lehetséges legegyszerűbb függvényosztályba tartozzék. Így az empirikus képlet típusának meghatározása négy lépésre bontható:

1. A tapasztalati úton kapott értékpárokat koordinátarendszerben ábrázoljuk;

2. A pontok közé jól illeszkedő görbét (egyenest) rajzolunk;

3. A görbét összehasonlítjuk a legfontosabb elemi függvényekkel és kiválasztjuk az empirikus képlet típusát;

4. Kiszámítjuk a paramétereket, és a kiválasztott empirikus képlet alkalmasságát a linearizáló módszerrel ellenőrizzük.

Linearizáló módszernek nevezzük azt az eljárást, amely révén a síknak valamely

$$\begin{cases} \overline{x} = f(x, y) \\ \overline{y} = g(x, y) \end{cases}$$

leképezésével a függvénykapcsolatban álló (x_k, y_k) pontokat olyan helyzetbe hozzuk, hogy az $(\overline{x}, \overline{y})$ síkon egy egyenesre illeszkednek, vagyis lineárisan függnek össze.

Könnyebb kezelhetőség kedvéért feltesszük, hogy az x_k, y_k értékek pozitívak. Ez mindig elérhető

$$\overline{x}_k = x_k + c_1, \quad \overline{y}_k = y_k + c_2$$

transzformációval.

Pl. A megfigyelt adatpárokat derékszögű koordinátarendszerben ábrázoljuk. A pontok közé közelítőleg egy hiperbola jellegű görbe illeszthető. Feltesszük, hogy az $y = \dfrac{x}{c + dx}$ függvény a megfelelő empirikus képlettípus. Ebben az esetben az $\overline{x} = x$, $\overline{y} = \dfrac{x}{y}$ leképezés a $c + d\,\overline{x} = \overline{y}$ lineáris kapcsolatot állítja elő.

A gyakorlatban a kiválasztott empirikus képlet alkalmasságának eldöntéséhez az adott x_k, y_k értékekkel kiszámítjuk az $\overline{x}_k, \overline{y}_k$ értékeket, és megvizsgáljuk, hogy az $(\overline{x}_k, \overline{y}_k)$ értékpárok által felvett pontok közelítőleg egy egyenesre esnek-e. Ha igen, akkor az empirikus képlet típusa megfelelő. A pontok az x értékekkel skálázott abszcisszatengelyű és $\dfrac{x}{y}$ értékekkel skálázott ordinátatengelyű koordinátarendszerben ábrázolva esnek közelítőleg egy egyenesre.

Néhány egyszerű empirikus képlettípus és a lineáris kapcsolatot előállító leképezés:

Empirikus képlet: $y = ax^b$;

linearizálás: $\overline{x} = \lg x$, $\overline{y} = \lg y$;

lineáris függvénykapcsolat: $\overline{y} = \lg a + b\overline{x} \equiv A + b\overline{x}$.

A pontok kettős logaritmikus papíron ábrázolva közelítőleg egy egyenesre esnek.

Empirikus képlet: $y = ae^{bx}$;

linearizálás: $\overline{x} = x$, $\overline{y} = \lg y$;

lineáris függvénykapcsolat: $\overline{y} = \lg a + b \lg e \cdot x \equiv A + B\overline{x}$.

A pontok féllogaritmikus papíron ábrázolva közelítőleg egy egyenesre esnek.

Empirikus képlet: $y = ax^n + c$;

linearizálás: $\overline{x} = x^n$, $\overline{y} = y$;

lineáris függvénykapcsolat: $\overline{y} = a\overline{x} + c$.

A pontok x^n értékekkel skálázott abszcisszatengelyű és y értékekkel skálázott ordinátatengelyű koordinátarendszerben ábrázolva közelítőleg egy egyenesre esnek.

4.2. A paraméterek meghatározása

Az empirikus képlet megfelelő típusának kiválasztása után meghatározzuk a képletben szereplő paraméterek numerikus értékét úgy, hogy a közelítés valamely feltétel szerint optimális legyen. Leggyakrabban az *a) kiválasztott pontok módszere, b) közepek módszere,* és *c) legkisebb négyzetek módszere* használatos.

***a)* A kiválasztott pontok módszere.** A tapasztalati úton kapott n-számú (x_k, y_k) $(k = 1, 2, \ldots, n)$ értékpárhoz az előbb leírt lépések szerint megalkotjuk a megfelelő

$$y = \varphi(x; a_1, a_2, \ldots, a_m)\ (m < n), \quad x \in R \tag{1}$$

empirikus képlettípust, $a_1, a_2, \ldots, a_m$ paraméterekkel. A koordinátarendszerben ábrázolt (x_k, y_k) $(k = 1, 2, \ldots, n)$ pontok közé illesztett görbén kiválasztunk m számú (x_i, y_i) $(i = 1, 2, \ldots, m)$ pontot és ezeket behelyettesítve az (1) képletbe, az $a_1, a_2, \ldots, a_m$ paraméterekre m egyenletből álló

$$y_i = \varphi(x_i; a_1, a_2, \ldots a_m) \quad (i = 1, 2, \ldots, m) \tag{2}$$

egyenletrendszert kapunk. A (2) egyenletrendszer megoldása szolgáltatja az a_i paraméterek numerikus értékeit. Ha a (2) egyenletrendszer ellentmondó, akkor az ellentmondást a görbe (x_i, y_i) $(i = 1, 2, \ldots, m)$ pontjainak alkalmasabb választásával szüntethetjük meg. Az egyenletrendszer megoldása után a paramétereket behelyettesítjük az (1) képletbe. Ellenőrzésül az így meghatározott empirikus képlet felhasználásával kiszámítjuk a $\varphi(x_k, a_1, a_2, \ldots, a_m)$ értékeket és képezzük az y_k tapasztalati értékek és a $\varphi(x_k, a_1, a_2, \ldots, a_m)$ értékek

$$d_k = y_k - \varphi(x_k; a_1, a_2, \ldots, a_m) \quad (k = 1, 2, \ldots, n) \tag{3}$$

különbségét, a különbségek

$$d = \frac{\sum_{k=1}^{n} |d_k|}{n} \tag{4}$$

átlagát, és a különbségek

$$d_q = \sum_{k=1}^{n} d_k^2 \tag{5}$$

négyzetösszegét.

***b*) A középek módszere.** A még ismeretlen paramétert tartalmazó (1) empirikus képlettípusba helyettesítsük be a kísérlettel kapott (x_k, y_k) $(k = 1, 2, \ldots, n)$ értékeket és képezzük az

$$e_k = y_k - \varphi(x_k; a_1, a_2, \ldots, a_m) \quad (k = 1, 2, \ldots, n) \tag{6}$$

eltéréseket. Az $a_1, a_2, \ldots, a_m$ paraméterek numerikus értékét abból a feltételből határozzuk meg, hogy a (6) eltérések algebrai összege zérus legyen, azaz

$$\sum_{k=1}^{n} e_k = 0 \tag{7}$$

feltétel teljesüljön. Mivel így csak egy feltételünk van az m számú paraméter meghatározásához, ezért az adott n számú (x_k, y_k) pontot m csoportba osztjuk, és mindegyik csoportban az eltérések összegét zérussal tesszük egyenlővé. Ezzel az eljárással előállítjuk a szükséges m számú feltételi egyenletet. Természetesen a

$$\sum_{k=1}^{n_1} e_k = 0, \quad \sum_{k=n_1+1}^{n_2} e_k = 0, \quad \sum_{k=n_{m-1}+1}^{n_m} e_k = 0, \tag{8}$$

részfeltételekkel együtt a (7) is teljesül. Ellenőrzésül, a kiszámított $a_1, a_2, \ldots, a_m$ paramétereknek (1) képletbe való behelyettesítése után, képezzük a (3), (4), (5) értékeket.

Megjegyzés

A középek módszerével leggyakrabban az $\begin{cases} \overline{x} = f(x, y) \\ \overline{y} = g(x, y) \end{cases}$ transzformáció után kapott közötti lineáris kapcsolatot határozzuk meg, melynek általános alakja: $\overline{y} = a\overline{x} + b$. Így a kísérlet eredményeként kapott (x_k, y_k) értékpárokkal x_k vagy y_k változó növekedésének sorrendjében felírjuk a $y_k = ax_k + b$ $(k = 1, 2, \ldots, n)$ feltételi egyenleteket és a rendszert kb. megfelezve, két csoportra osztva különkülön összeadjuk. Az így kapott két egyenletből álló egyenletrendszerből kiszámítjuk a és b értékét. Az $\overline{x}$ és $\overline{y}$ helyére az eredeti változóknak megfelelő kifejezések behelyettesítésével kapjuk az x és y közötti kapcsolat empirikus képletét. Amennyiben még nem ismert az összes paraméter, akkor ugyanezt a módszert alkalmazzuk még egyszer, de előbb egy újabb, $\overline{x}$ és $\overline{y}$ közötti lineáris kapcsolatot hozunk létre.

***c*) A legkisebb négyzetek módszere.** Az ún. legkisebb négyzetek módszerével a paraméterek numerikus értékét az eltérések négyzetösszegének minimum feltételéből határozzuk meg.

A tapasztalati úton kapott n-számú (x_k, y_k) $(k = 1, 2, \ldots, n)$ értékpárhoz az előbb leírt lépések szerint megalkotjuk a megfelelő (1) empirikus képlettípust. Mivel az adatokat pl. véletlenszerű mérési hibák terhelik, ezért az y_k mért értékek és a φ függvény x_k helyen vett helyettesítési értékei általában ε_k értékkel eltérnek egymástól, azaz

$$y_k - \varphi(x_k; a_1, a_2, \dots, a_m) = \varepsilon_k, \quad (k = 1, 2, \dots, n). \tag{9}$$

Az ε_k mérési hibákról (a tapasztalatok szerint) általában feltehető a függetlenség és az $N(0,\sigma)$ normális eloszlás. A feladat tehát olyan $y = \varphi(x; a_1, a_2, \dots, a_m)$ $(m < n)$ függvény keresése, amely minimalizálja a mérési hibák hatását, azaz az eltérések négyzetösszegét, $\sum_{k=1}^{n} \varepsilon_k^2$ -et. Ez azt jelenti, hogy képeznünk kell az

$$F(a_1, a_2, \dots, a_m) = \sum_{k=1}^{n} \left[y_k - \varphi(x_k; a_1, a_2, \dots, a_m)\right]^2 \tag{10}$$

függvény minimumát amely az a_1, $a_2, \dots, a_m$ változók nemnegatív függvénye. Ha φ a paraméterek szerint differenciálható, akkor a feladat visszavezethető a

$$\frac{\partial F}{\partial a_1} = 0; \ \frac{\partial F}{\partial a_2} = 0; \ \dots; \ \frac{\partial F}{\partial a_m} = 0 \tag{11}$$

ún. **feltételi egyenletrendszer** megoldására. Kiszámítva az m számú (11) egyenletrendszerből az m számú paramétert és visszahelyettesítve az (1) egyenletbe, megkapjuk az empirikus képletet. A véletlen hibákkal rendelkező mérési adatok alapján meghatározott

$$y = \varphi(x; a_1, a_2, \dots, a_m) \ (m < n) \quad (x, a_1, a_2, \dots, a_n \in R) \tag{12}$$

függvényt **regressziós függvénynek** vagy **regressziós görbének** szokás nevezni. Ellenőrzésül képezzük (3), (4) és (5) értékeket.

A legkisebb négyzetek módszerének előnye, hogy ha a különbségek négyzetösszege, d_q, kicsi, akkor a különbségek d_k értéke $(k = 1, 2, \dots, n)$ is kicsiny, ami a közepek módszerénél nem mindig áll fenn. Hátránya, hogy az előbbi módszerekhez viszonyítva több számítási munkát igényel. Használata akkor célszerű, ha a tapasztalati adatok is elég pontosak.

Példa

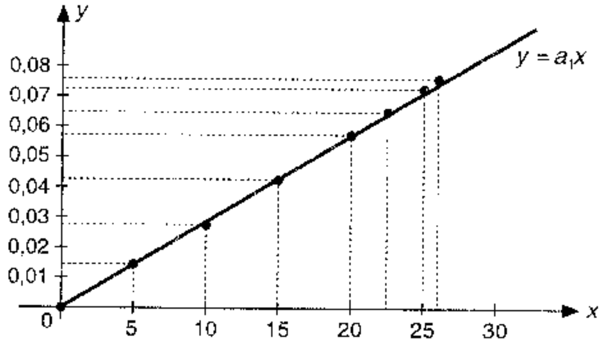
Tegyük fel, hogy méréseink táblázatba foglalt adatai:

x	0	5	10	15	20	22,5	25	26
y	0	0,0143	0,0276	0,0426	0,0572	0,0648	0,0725	0,0758

Határozzuk meg az empirikus képletet az *a*), *b*) és *c*) módszer szerint, és hasonlítsuk össze a d_k különbségek értékét, a különbségek d átlagát és d_q négyzetösszegét.

Megoldás

A pontokat koordinátarendszerben ábrázolva (47. ábra) látjuk, hogy azok jó közelítéssel egy egyenesre esnek.



47. ábra. Pontok közé illeszkedő egyenes

A pontok közé illeszkedő egyenes empirikus képlettípusa $y = ax$.

a) A *kiválasztott pontok módszerével* a_0 értékének meghatározásához válasszunk $x_7 = 25$ és $y_0 = 0{,}0715$ értéket, amely a pontok közé illesztett egyenes ordináta értéke. Így $0{,}0715 = a_0 \cdot 25$, vagyis $a_0 = 0.00286$. Tehát az a) empirikus képlet:

$$y_a = 0{,}00286\,x.$$

b) A közepek módszeréhez

$$\sum_{k=1}^{8} e_k = \sum (y_k - a_0 x_k) = 0 \text{, és így } a_0 = \frac{\sum_{k=1}^{8} y_k}{\sum_{k=1}^{8} x_k} \approx 0{,}00287,$$

tehát

$$y_b = 0{,}00287\,x.$$

c) A *legkisebb négyzetek módszerét* alkalmazva a (10)-nek megfelelő függvény

$$F(x;a_0)=\sum_{k=1}^{8}(y_k-a_0x_k)^2.$$

A a_0 -t a szélsőérték létezésének feltételi egyenletéből kapjuk:

$$\frac{dF}{da_0}=\frac{d}{da_0}\sum_{k=1}^{8}(y_k-a_0x_k)^2=0.$$

Deriválás után 2-vel egyszerűsítve

$$\sum_{k=1}^{8}x_k(y_k-a_0x_k)=0,$$

és így

$$a_0=\frac{\sum_{k=1}^{8}y_kx_k}{\sum_{k=1}^{8}x_k^2}\approx 0{,}00288,$$

tehát az egyenes (regressziós egyenes) egyenlete:

$$y_c=\varphi(x)=0{,}00288.$$

Az adatokat, a valamint az egyes módszerekkel kapott empirikus függvényhez tartozó kiszámított ordináta és d_k értékeket táblázatba foglaltuk, melyből látható, hogy legjobb eredményt a legkisebb négyzetek módszerével kaptuk.

x_k	y_k	$(y_a)_k$	$(y_b)_k$	$(y_c)_k$	$(d_a)_k$ (10^{-4})	$(d_b)_k$ (10^{-4})	$(d_c)_k$ (10^{-4})
0	0	0	0	0	0	0	0
5	0,0143	0,0143	0,0144	0.0144	0	-1	-1
10	0,0276	0,0286	0,0287	0,0288	-10	-11	-12
15	0,0426	0,0429	0,0430	0,0432	-3	-4	-6
20	0,0572	0,0572	0,0547	0,0576	0	-2	-4
22,5	0,0648	0,0644	0,0646	0,0648	+4	+2	0
25	0,0725	0,0715	0,0718	0,0720	+10	+7	+5
26	0,0758	0,0744	0,0746	0,0749	+14	+12	+9
123,5	0.3548			$\sum_{k=1}^{8}\lvert d_k\rvert$	41	39	37
				$\frac{\sum_{k=1}^{8}\lvert d_k\rvert}{8}$	5,125	4,875	4,625
				$\sum_{k=1}^{8}d_k^2$	421	339	287

4.3. A statisztikai modell

Az I. rész 3. fejezet 3.1.4. és 3.1.5. pontjában definiáltuk két valószínűségi változó, *X* és *Y*

$$w(x_i,y_j)=P(X=x_i,Y=y_j),\quad (i=1,2,\dots,n;\ j=1,2,\dots,m)$$

együttes eloszlását,

$$v(x_i)=\sum_{j=1}^{m}w(x_i,y_j);\quad u(y_i)=\sum_{i=1}^{n}w(x_i,y_j);$$

peremeloszlásait, a

$$\mathrm{Cov}(X,Y)=M\left[(X-M(X))(Y-M(Y))\right]\equiv\sum_{i,j}x_iy_jw(x_i,y_j)-m_xm_y\equiv$$
$$\equiv M(XY)-m_xm_y$$

kovarianciáját és az

$$\mathrm{R}(X,Y)=\frac{\mathrm{Cov}(X,Y)}{D(X)D(Y)}\equiv\frac{M(XY)-m_xm_y}{D(X)D(Y)}$$

korrelációját.

A **korrelációanalízis** a valószínűségi változók közötti sztochasztikus kapcsolatok felderítésével és e kapcsolatok szorosságának, erősségének mérésével, a **regresszióanalízis** pedig az összefüggő változókra vonatkozó, adott tulajdonságú függvénykapcsolatok megadásával, képlettel való leírásával foglalkozik.

A kovariancia a két változó együttes változását, a korreláció a két változó kapcsolatának szorosságát jellemzi.

Ha egy *Y* valószínűségi változó sztochasztikus kapcsolatban áll egy *X* valószínűségi változóval, akkor az *Y* értékét célszerű az $M(Y|X=x)$ feltételes várható értékkel becsülni. A becslést minden *x*-re elvégezve, kapjuk a

$$\varphi(x)=M(Y|X=x)\qquad (x\in R)\qquad (6)$$

ún. regressziós függvényt.

A $\varphi(x)$ függvényt úgy célszerű megválasztani, hogy a $\varphi(x)$ értékek a lehető legközelebb legyenek az *Y* értékeihez. A közelséget mérhetjük például az $(Y-\varphi(x))^2$ várható értékével. A $\varphi(x)$-et úgy választjuk meg, hogy a várható érték minimális legyen, tehát az

$$M\left((Y-\varphi(x))^2\right)=\min$$

feladatot kell megoldanunk. Ez azt jelenti, hogy a $\varphi(x)$ regressziós görbétől az Y értékeinek négyzetes átlageltérése a legkisebb.

Ha az X és Y együttes eloszlása normális, akkor a regressziós függvény lineáris, grafikonja egyenes. A legkisebb négyzetek módszerével kapott eredmény alapján az Y értékeit legjobban közelítő függvény a regressziós függvény. Az együttes sűrűségfüggvényből a **regressziós függvényt**

$$y=\varphi(x)=\mathrm{R}(X,Y)\frac{D(Y)}{D(X)}(X-M(X))+M(Y),\ (x\in R) \qquad (7)$$

alakban állíthatjuk elő.

Ha az X, Y változókra ismert egy (X_1,Y_1), (X_2,Y_2), ..., (X_n,Y_n) statisztikai minta, akkor lineáris korrelációt feltételezve, a paramétereket a mintából kapott (x_i,y_i) egyenletes eloszlású diszkrét értékekkel becsülve:

$$M(X)=m_1\approx\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n};\ D(X)=\sigma_1\approx\frac{1}{n}\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$M(Y)=m_2\approx\bar{y}=\frac{\sum_{i=1}^{n}y_i}{n};\ D(Y)=\sigma_2\approx\frac{1}{n}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}\ ;$$

és így

$$\mathrm{R}(X,Y)\approx r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\cdot\sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad (8)$$

A korrelációanalízis szóhasználatában az $\mathrm{R}(X,Y)$ a lineáris kapcsolat mérőszáma, elnevezése: **korrelációs együttható (koefficiens)**, és leggyakrabban r-rel jelölik. Mint már láttuk $-1\le r\le 1$ és $r=\pm1$ esetén a valószínűségi változók között lineáris függvénykapcsolat van, ha viszont $|r|<1$, akkor a függvénykapcsolat típusára nem, de a 0 és 1 között elfoglalt helyétől függően a lineáris kapcsolat erősségére következtetést vonhatunk le. Ha $r=0$, akkor az X és Y valószínűségi változókat **korrelálatlanoknak** mondjuk. A független valószínűségi változókra $r=0$, de abból, hogy $r=0$, általában még nem következik a valószínűségi változók függetlensége, de normális eloszlásúak esetén igen.

K. Pearson a változók szorosságának mérésére az Y-nak X-re vonatkozó várható értéke varianciájának és az Y varianciájának hányadosát javasolja:

$$Q=\frac{\mathrm{Var}(M(Y|X))}{\mathrm{Var}(Y)}=\frac{D^2(M(Y|X))}{D^2(Y)} \qquad (*)$$

A (*) formulával kiszámított Q számra teljesül az

$$0\le Q\le 1$$

egyenlőtlenség.

Q pozitív négyzetgyöke $\sqrt{Q}$, az ún. **korrelációhányados** csak akkor 0, ha X és Y valószínűségi változók függetlenek, és csak akkor 1, ha X és Y között függvénykapcsolat van.

A becslő formulák felhasználásával a (7)-ből a regressziós egyenes mintabeli becslésére

$$y=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}\cdot\frac{\frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}{\frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}(x-\bar{x})+\bar{y}$$

egyenletet kapjuk, melyből az egyszerűsítéssel előálló:

$$y=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}(x-\overline{x})+\overline{y} \qquad (9)$$

alakot használjuk.

Megjegyzés

Ha a legkisebb négyzetek módszerét alkalmazzuk a pontok közé illeszkedő egyenes $y=ax+b$ képlettípusára, akkor az

$$F(x;a,b)=\sum_{i=1}^{n}(y_i-ax_i-b)^2 \qquad (10)$$

kétváltozós függvény minimumát kell kiszámítani, amely az

$$M\left[(Y-aX-b)^2\right]\to \min$$

követelmény mintabeli becslése. A (10) kétváltozós függvény minimumának kiszámításához képezzük a függvény a és b szerinti parciális deriváltját. A deriváltakat 0-val egyenlővé téve, a-ra és b-re a következő egyenletrendszert kapjuk:

$$F'_a=\sum_{i=1}^{n}2(y_i-ax_i-b)(-x_i)=0,$$

$$F'_b=\sum_{i=1}^{n}2(y_i-ax_i-b)(-1)=0.$$

Az egyenletrendszer egyetlen (a,b) megoldása az F nemnegatív volta miatt, a (10) függvény minimumát adja. Az egyenletrendszer rendezésével kapjuk az ún. **normálegyenletrendszert**:

$$\sum_{i=1}^{n}x_iy_i=a\sum_{i=1}^{n}x_i^2+b\sum_{i=1}^{n}x_i\,,$$

$$\sum_{i=1}^{n}y_i=a\sum_{i=1}^{n}x_i+b\cdot n \qquad (11)$$

melynek megoldása:

$$a=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2} \quad \text{és} \quad b=\overline{y}-\overline{x}a \qquad (12)$$

Példa

A táblázat egy finom fémszál átlagterheléseknél mért átlag megnyúlásait tartalmazza:

x kg	12	14	16	18	20	22
y m	0,7516	0,7628	0,7728	0,7836	0, 7956	0,8065

A két változó között lineáris regressziót feltételezve számítsuk ki a korrelációs együtthatót és a korrelációs egyenes egyenletét.

Megoldás

A számítás menetét az alábbi táblázat szemlélteti:

k	x_k	y_k	$x_k-\overline{x}$	$y_k-\overline{y}$	$(x_k-\overline{x})(y_k-\overline{y})$	$(x_k-\overline{x})^2$	$(y_k-\overline{y})^2$
1	12	0,7516	–5	–0,0272	0,1360	25	0,0007
2	14	0,7628	–3	–0,0160	0,0480	9	0,0003
3	16	0,7728	–1	–0,0060	0,0060	1	0,0000
4	18	0,7836	1	0,0048	0,0048	1	0,0000
5	20	0,7956	3	0,0168	0,0504	9	0,0003
6	22	0,8065	5	0,0277	0,1385	25	0,0008
$\sum$	102	4,6729			0,3837	70	0,0021
$\sum/6$	17	0,7788					

A lineáris kapcsolat feltételezése jogos volt, tehát a regressziós egyenes egyenletének paraméterei:

$$a=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\approx\frac{0{,}3837}{70}=0{,}0055,$$

$$b=\overline{y}-\overline{x}a\approx 0{,}7788-17\ \ 0{,}0055=0{,}6853.$$

A regressziós egyenes egyenlete:

$$y=ax+b=0{,}0055x+0{,}6853.$$

E.4. Ellenőrző kérdések a 4. fejezethez

1. Hogyan választunk empirikus képletet?

2. Mit nevezünk linearizáló módszernek?

3. Hogyan állítható elő néhány nevezetes függvénytípus linearizáló képlete?

4. Hogyan definiáljuk a legkisebb négyzetek módszerét?

5. Mit mér a korrelációs együttható?

6. Hogyan állítjuk elő a regressziós függvényt?

7. Hogyan számítjuk ki a lineáris regressziós egyenes együtthatóit?

S.4. Feladatok a 4. fejezethez

S.4.1. Egy darabológép 100 cm-es rudak vágására van beállítva. Véletlenszerű kiválasztással hat rúd hosszát és súlyát lemérik. Az eredményt a következő táblázat tartalmazza:

x_k cm	101,3	103,7	98,6	99,9	97,2	100,1
y_k dkg	609	626	586	594	579	605

Számítsuk ki a korrelációs együttható értékét és ha a két változó között feltételezhető a lineáris kapcsolat, akkor határozzuk meg a regressziós egyenes egyenletét.

S.4.2. Egy változó lélekszámú faluban a születések számát öt egymást követő évben az alábbi táblázat tartalmazza:

x_k évek	1970	1971	1972	1973	1974
y_k születések	13	24	32	46	57

Állapítsuk meg a születések száma és az évek közötti kapcsolatot. Ha az adatok alapján lineáris kapcsolat feltételezhető, akkor határozzuk meg a regressziós egyenes egyenletét.

MEGOLDÁSOK

III. RÉSZ

MEGOLDÁSOK

AZ I. RÉSZ FELADATAINAK MEGOLDÁSAI

V.1. Az 1. fejezet feladatainak megoldásai

V.1.1. Az események összege I, és páronként kizárják egymást:

$$\overline{A}+(A-B)+AB=\overline{A}+A\overline{B}+AB=\overline{A}+A(\overline{B}+B)=I,$$

$$\overline{A}(A-B)=\overline{A}(A\overline{B})=(\overline{A}A)\overline{B}=\varnothing\overline{B}=\varnothing,$$

$$\overline{A}(AB)=(\overline{A}A)B=\varnothing B=\varnothing.$$

$$(A-B)AB=(A\overline{B})(AB)=A\overline{B}AB=AA\overline{B}B=A\varnothing=\varnothing.$$

V.1.2. *a)* $B\left(\overline{B}\,A\right)=(B\,\overline{B})A=\varnothing A=\varnothing$;

A műveleti szabályok *e*, 2. pontja szerint:

b) $A+\overline{A}\,B=(A+\overline{A})(A+B)=I\cdot(A+B)=A+B.$

c) $(A-C)(B-C)=(A\overline{C})(B\overline{C})=AB\overline{C}\,\overline{C}=(AB)\overline{C}=AB-C.$

V.1.3. *a)* Mindkét kapu nyitva;

b) legalább az egyik kapu nyitva;

c) a zöld kapu nincs nyitva (a vörös lehet, hogy nyitva van, de az is lehet hogy zárva van);

d) a vörös kapu nyitva, de a zöld zárva;

e) egyik kapu sincs nyitva;

f) a „zöld zárva" és a „vörös nyitva" események közül legalább az egyik teljesül, vagyis, ha a vörös nyitva, akkor a zöld lehet nyitva is meg nem is. Ha pedig a vörös zárva van, akkor a zöld sincs nyitva;

g) a zöld nyitva, de a vörös zárva;

h) mindkét kapu zárva;

i) legfeljebb egyik lehet nyitva;

j) legfeljebb egyik lehet nyitva;

k) csak az egyik kapu lehet nyitva.

V.1.4. *a)* minthogy $AB+\overline{A}B=(A+\overline{A})B=I\cdot B=B$, ezért $P(B)=P(AB+\overline{A}B)$, de AB és $\overline{A}B$ egymást kizáró események, ezért

$$P(B)=P(AB+\overline{A}B)=P(AB)+P(\overline{A}B).$$

b) $P(A+B)=1-P(\overline{A+B})=1-P(\overline{A}\,\overline{B})$, (*de Morgan*).

V.1.5. A fej és írás dobás elemi események lehetséges számát 2 elem hatod osztályú ismétléses variációi adják:

$$V_{2,6}^{i}=2^6=64.$$

Ha A-val jelöljük a kérdezett eseményt, akkor a $P(A)=\frac{k}{n}$ képletet használva:

a) $k=1$, $P(A)=\frac{1}{64}$;

b) $k=1$, $P(A)=\frac{1}{64}$;

c) $P(A)=\frac{1}{64}+\frac{1}{64}=\frac{2}{64}$;

d) Azt a helyet, amelyen fej áll $\binom{6}{1}=6$ módon választhatjuk ki, így

$$P(A)=\frac{6}{64}=\frac{3}{32};$$

e) 4 fej dobása $k=\binom{6}{4}=15$-féleképpen állhat elő, tehát

$$P(A)=\frac{15}{64};$$

f) Ha $\overline{A}$ jelöli a „mind írás" eseményét, akkor legalább egy fej (nem mind írás) A eseményének valószínűsége:

$$P(A)=1-P(\overline{A})=1-\frac{1}{64}=\frac{63}{64}=0{,}98.$$

V.1.6. *a)* Visszatevés nélkül: $P(A)=\dfrac{\binom{20}{2}\binom{30}{0}}{\binom{50}{2}}=0{,}1551,$

Visszatevéssel: $P(A)=\binom{2}{2}\left(\frac{20}{50}\right)^2\left(\frac{30}{50}\right)^0=0{,}16,$

b) Visszatevés nélkül: $P(A)=\dfrac{\binom{20}{0}\binom{30}{2}}{\binom{50}{2}}=0{,}3551,$

Visszatevéssel: $P(A)=\binom{2}{0}\left(\frac{20}{50}\right)^0\left(\frac{30}{50}\right)^2=0{,}36,$

c) Visszatevés nélkül: $P(A)=\dfrac{\binom{20}{1}\binom{30}{1}}{\binom{50}{2}}=0{,}48979,$

Visszatevéssel: $P(A)=\binom{2}{1}\left(\frac{20}{50}\right)^1\left(\frac{30}{50}\right)^1=0{,}48.$

d) Legyen A legalább az egyik hibás, B csak az egyik hibás, C mindkettő hibás esemény, akkor $A=B+C$, $B\cdot C=\varnothing$, és így $P(A)=P(B)+P(C)$.

Visszatevés nélkül: *c)* és *a)* összege: $P(A)=0{,}6449,$

Visszatevéssel: *c)* és *a)* összege: $P(A)=0{,}64.$

e) Legyen A legfeljebb az egyik hibás, B egyik sem hibás, C csak az egyik hibás esemény, akkor $A=B+C$, $B\cdot C=\varnothing$, és így $P(A)=P(B)+P(C)$.

Visszatevés nélkül: *b)* és *c)* összege: $P(A)=0{,}8449,$

Visszatevéssel: *b)* és *c)* összege: $P(A)=0{,}84.$

V.1.7. A 6 hektár $6\cdot 10000\ \mathrm{m}^2$, a sportpálya területe $100\cdot 80=8000\,\mathrm{m}^2$, tehát

$$P(A)=\frac{8000}{60000}=\frac{2}{15}=0{,}13.$$

V.1.8. *a)* A Q elemeiből azok az A elemei, amelyek az F betűből és a páros számokból állnak (hasonló meggondolással adjuk meg B és C elemeit):

$A=\{F2,F4,F6\}$

$B=\{F2,F3,F5,I2,I3,I5\}$

$C=\{I1,I3,I5\}$.

b) A vagy B azaz $A\cup B=\{F2,F4,F6,F3,F5,I2,I3,I5\}$;

B és C azaz $B\cap C=\{I3,I5\}$;

a B azon elemei, amelyek sem az A-hoz sem a C-hez nem tartoznak:

$$B\cap\overline{A}\cap\overline{C}=\{F3,F5,I2\}$$

c) A páronként egymást kizáró események A és C, $A \cap C = \varnothing$.

$$(A \cap B = \{F2\},\ B \cap C = \{I3, I5\})$$

V.1.9. *a)* $\Omega = \{FF1, FF2, FF3, FF4, FF5, FF6, FI1, FI2, FI3, FI4, FI5, FI6,$
$IF1, IF2, IF3, IF4, IF5, IF6, II1, II2, II3, II4, II5, II6\}$

b) $A = \{FF2, FF4, FF6\}$, $B = \{FF2, FI2, IF2, II2\}$,

$C = \{FI2, IF2, FI3, IF3, FI5, IF5\}$;

c) $A \cap B = \{FF2\}$, $B \setminus (A \cup C) = \{II2\}$;

$B \cup C = \{FF2, FI2, IF2, II2, FI3, IF3, FI5, IF5\}$.

V.1.10. *a)* Nem definiál valószínűségi mezőt, mivel a valószínűségek összege:

$$\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}=\frac{77}{60}>1.$$

b) Mivel $P(A_3) = -\frac{1}{4} < 0$, ezért a függvény nem definiál valószínűségi mezőt.

c) A függvény valószínűségi mezőt definiál a Ω-n, mivel minden valószínűség 1-nél kisebb pozitív szám és összegük

$$\frac{1}{2}+\frac{1}{2}+\frac{1}{8}+\frac{1}{8}=1.$$

d) A függvény valószínűségi mezőt definiál a Ω-n, mivel a valószínűségek nemnegatív értékek és összegük 1.

V.1.11. *a)* Legyen $P(A_1) = p$, akkor $p + \frac{1}{3} + \frac{1}{4} = 1$, amelyből $p = 1 - \frac{7}{12} = \frac{5}{12}$, vagyis $P(A_1) = \frac{5}{12}$.

b) Legyen $P(A_2) = p$, akkor $2p + p + \frac{1}{4} = 1$; $3p = \frac{3}{4}$, $p = \frac{1}{4}$, tehát

$$P(A_1) = 2p = 2 \cdot \frac{1}{4} = \frac{1}{2}.$$

c) A III. axióma szerint: $P(\{A_2, A_3\}) = P(A_2) + P(A_3)$, így a feladat szerint $P(A_2) + P(A_3) = 2P(A_1)$. Ha $P(A_1) = p$, akkor $p + 2p = 1$, és $p = \frac{1}{3}$, vagyis $P(A_1) = \frac{1}{3}$.

d) Legyen $P(A_2) = p$, akkor $P(A_3) = 2p$ és $p = 3P(A_1)$, amiből $\frac{p}{3} + p + 2p = 1$, $\frac{10}{3}p = 1$, $p = \frac{3}{10}$ és így $P(A_1) = \frac{p}{3} = \frac{1}{10}$.

V.1.12. Ha $P(M) = p$, akkor $P(M) = P(A) = P(J) = p$, és $P(N) = P(B) = 2p$. Az öt elemi esemény valószínűségének összege 1, tehát $p+p+p+2p+2p=1$, amelyből $p = \frac{1}{7}$.

a) $P(\{M, A, J\}) = P(M) + P(A) + P(J) = \frac{1}{7} + \frac{1}{7} + \frac{1}{7} = \frac{3}{7}$;

b) $P(\{N, M\}) = P(N) + P(M) = \frac{2}{7} + \frac{1}{7} = \frac{3}{7}$.

V.1.13. *a)* Ha $P(1) = p$, akkor $P(2) = 2p$, $P(3) = 3p$, $P(4) = 4p$, $P(5) = 5p$, $P(6) = 6p$. Mivel a valószínűségek összege 1, ezért $p+2p+3p+4p+5p+6p=1$, amelyből $p = \frac{1}{21}$. Tehát

$$P(1) = \frac{1}{21},\quad P(2) = \frac{2}{21},\quad P(3) = \frac{3}{21},\quad P(4) = \frac{4}{21},\quad P(5) = \frac{5}{21},\quad P(6) = \frac{6}{21}.$$

b) $P(A) = P(\{2,4,6\}) = \frac{2}{21} + \frac{4}{21} + \frac{6}{21} = \frac{12}{21} = \frac{4}{7}$; $P(B) = P(\{2,3,5\}) = \frac{10}{21}$;

$P(C) = P(\{1,3,5\}) = \frac{3}{7}$.

c)

c_1) a páros vagy prímszámok eseménye: $A \cup B = \{2,4,6,3,5\}$, tehát csak az 1 nem tartozik hozzá, ezért $P(A \cup B) = 1 - P(1) = \frac{20}{21}$.

c_2) a páratlan prímszámok eseménye: $B \cap C = \{3,5\}$, tehát

$$P(B \cap C) = P(\{3,5\}) = \frac{8}{21}.$$

c_3) a B-hez nem tartozó páros számok eseménye: $A \cap \overline{B} = \{4,6\}$, tehát

$$P(A \cap \overline{B}) = P(\{4,6\}) = \frac{10}{21}.$$

V.1.14. Az $A = \{\text{mindkettő hibás}\}$ ill. $B = \{\text{egyik sem hibás}\}$ események valószínűségét $P(A)$-t és $P(B)$-t kell meghatározni. Mivel 12-ből 2-t $\binom{12}{2} = \frac{12 \cdot 11}{1 \cdot 2} = 66$-féleképpen lehet kiválasztani, ezért a Ω eseménytér elemszáma 66. A négy hibás tranzisztorból kettőt $\binom{4}{2} = \frac{4 \cdot 3}{1 \cdot 2} = 6$-féleképpen lehet kiválasztani, tehát $P(A) = \frac{6}{66} = \frac{1}{11}$.

A nyolc hibátlan tranzisztorból kettőt $\binom{8}{2} = \frac{8 \cdot 7}{1 \cdot 2} = 28$-féleképpen lehet kiválasztani, tehát $P(B) = \frac{28}{66} = \frac{14}{33} = 0{,}42$.

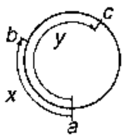
A $C = \{\text{legalább egy hibás}\}$ esemény valószínűségét, mivel C a B komplemense, azaz $C = \overline{B}$, a $P(\overline{B}) = 1 - P(B)$ összefüggéssel számítjuk:

$$P(C) = P(\overline{B}) = 1 - P(B) = 1 - \frac{14}{33} = \frac{19}{33}.$$

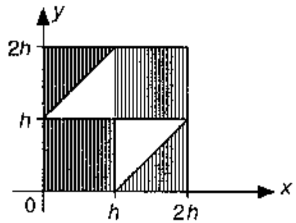
V.1.15. $P(A) = \frac{10}{30} = \frac{1}{3}$; $P(B) = \frac{10+5}{30} = \frac{1}{2}$; $P(A \cap B) = \frac{5}{30} = \frac{1}{6}$, tehát a férfi vagy mérnök személyi lap kiemelésének valószínűsége, a 6. tétel felhasználásával:

$$p = P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{3} + \frac{1}{2} - \frac{1}{6} = \frac{2}{3}.$$

V.1.16. A kör kerülete legyen $2h$ hosszúságú. Az óramutató járásával megegyezően haladva a kör kerületének a és b közötti ívhosszát jelölje x, az a és c közötti ívhosszát pedig y (48. ábra).



48. ábra. 2h kerületű kör



49. ábra. 2h oldalú négyzet

A $0 < x < 2h$ és $0 < y < 2h$ egyenlőtlenségeket kielégítő (x, y) pontok alkotják a Q halmazt (49. ábra). Az A halmazt a Q azon részhalmaza alkotja, amelynek pontjai kielégítik az alábbi egyenlőtlenségeket:

a) $x, y < h$;

b) $x, y > h$;

c) $x < h$ és $y - x > h$;

d) $y < h$ és $x - y > h$.

(A 49. ábra négyzetének bevonalkázott részei.) Ekkor az A azon pontokból áll, amelyekkel a, b és c a félköríven fekszik, tehát

$$p = \frac{A \text{ területe}}{Q \text{ területe}} = \frac{3h^2}{4h^2} = \frac{3}{4}.$$

V.1.17. $P(A) = 1 - P(\overline{A}) = 1 - \frac{2}{3} = \frac{1}{3}$; a $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ egyenletbe helyettesítve: $\frac{3}{4} = \frac{1}{3} + P(B) - \frac{1}{4}$, amiből $P(B) = \frac{2}{3}$.

$$P(A \cap \overline{B}) = P(A) - P(A \cap B) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

V.1.18. $P(A \cap B) = \frac{1}{8}$; $P(\overline{A} \cap \overline{B}) = \frac{1}{4}$; $P(\overline{A} \cup \overline{B}) = \frac{7}{8}$ és $P(B \cap \overline{A}) = \frac{1}{4}$.

V.1.19. A relatív gyakoriságot az

$$r = \frac{\text{gyakoriság (az esemény bekövetkezésének száma)}}{\text{kísérletek száma}}$$

képlettel számítjuk ki.

a) $r = \left(\frac{\text{6-os dobás száma}}{\text{kísérletek száma}}\right) = \frac{16}{100} = 0{,}16;$

b) $r = \left(\frac{\text{3-as dobás száma}}{100}\right) = \frac{20}{100} = 0{,}20;$

c) $r = \left(\frac{\text{páratlan szám dobás száma}}{100}\right) = \frac{14+20+15}{100} = 0{,}49;$

d) $r = \left(\frac{\text{prímszám dobás száma}}{100}\right) = \frac{17+20+15}{100} = 0{,}52.$

V.2. A 2. fejezet feladatainak megoldásai

V.2.1. *a)* Jelöljük A-val azt az eseményt, ha az 1-es kockán 5-ös van felül, akkor

$$A=\{(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)\}$$

A pontszámok összege 10 vagy 10-nél nagyobb: $(5,5)$, $(5,6)$ dobás esetén, tehát

$$p=\frac{2}{6}=\frac{1}{3}.$$

b) $B=\{(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),(1,5),(2,5),(3,5),(4,5),(5,5),(6,5)\}$.

A 10 vagy 10-nél nagyobb pontszám dobása háromféleképpen jöhet létre (5,5), (5,6), (6,5), tehát $p=\frac{3}{11}$, hiszen B-nek csak 11 különböző eleme van.

V.2.2. Két szám összege páros szám, ha mindkét összeadandó páros ill. mindkettő páratlan. A 2, 4, 6, 8 páros számok közül kettőt $\binom{4}{2}=6$-féleképpen választhatunk ki, az 1, 3, 5, 7, 9 páratlan számok közül kettőt pedig $\binom{5}{2}=10$-féleképpen, így $6+10=16$ esetben kaphatunk páros összeget. Mivel 10 esetben következhet be két páratlan szám összegeként páros szám, így $p=\frac{10}{16}=\frac{5}{8}$.

V.2.3. Négy pikk lap kiosztása után az $52-4=48$ lap között még $13-4=9$ pikk van. 48 lapból hármat $\binom{48}{3}=17296$-féleképpen lehet kiválasztani. Mivel $48-9=39$ lap egyike sem pikk, ezért $\binom{39}{3}=9139$ módon lehet hármat kiosztani úgy, hogy egyik lap sem pikk, vagyis ennek valószínűsége $p_1=\frac{9139}{17296}$. Tehát annak valószínűsége, hogy legalább egy pikk lesz a három lap között:

$$p=1-p_1=\frac{8157}{17296}\approx 0{,}47.$$

V.2.4. A szorzástétel felhasználásával: $p=\frac{7}{10}\cdot\frac{6}{9}\cdot\frac{3}{8}=\frac{7}{40}$.

V.2.5. A szorzástétel alkalmazásával:

$$p=\frac{8}{32}\cdot\frac{7}{31}\cdot\frac{6}{30}\cdot\frac{5}{29}\cdot\frac{4}{28}=\frac{21}{75516}\approx 0{,}000278.$$

V.2.6. A $P(A\cup B)=P(A)+P(B)-P(A\cap B)$ összefüggésből:

$$P(A\cap B)=-P(A\cup B)+P(A)+P(B)=-\frac{3}{4}+\frac{3}{8}+\frac{5}{8}=\frac{2}{8}=\frac{1}{4},$$

és így

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{\frac{1}{4}}{\frac{5}{8}}=\frac{2}{5},$$

és

$$P(B|A)=\frac{P(B\cap A)}{P(A)}=\frac{\frac{1}{4}}{\frac{3}{8}}=\frac{2}{3}.$$

V.2.7. *a)* $P(A|B)=\frac{3}{4}$;

b) $P(B|A)=\frac{1}{2}$;

c) $P(A\cup B)=\frac{7}{12}$;

d) Mivel $P(\overline{B})=1-P(B)=1-\frac{1}{3}=\frac{2}{3}$,

és az 1. megjegyzés szerint

$$P(\overline{A}\cap\overline{B})=P(\overline{A\cup B})=1-P(A\cup B)=1-\frac{7}{12}=\frac{5}{12},$$

ezért

$$P(\overline{A}|\overline{B})=\frac{P(\overline{A}\cap\overline{B})}{P(\overline{B})}=\frac{\frac{5}{12}}{\frac{2}{3}}=\frac{5}{8}.$$

$$P(\overline{A})=1-P(A)=\frac{1}{2}$$

és így

$$P(\overline{B}|\overline{A})=\frac{P(\overline{B}\cap\overline{A})}{P(\overline{A})}=\frac{\frac{5}{12}}{\frac{1}{2}}=\frac{5}{6}.$$

V.2.8. Ha $M=\{$matematikából elégtelen hallgatók$\}$,
$F=\{$fizikából elégtelen hallgatók$\}$, akkor
$P(M)=0{,}25$, $P(F)=0{,}15$ és $P(M\cap F)=0{,}10$.

a) $P(M|F)=\dfrac{P(M\cap F)}{P(F)}=\dfrac{0{,}10}{0{,}15}=\dfrac{2}{3}$.

b) $P(F|M)=\dfrac{P(F\cap M)}{P(M)}=\dfrac{0{,}10}{0{,}25}=\dfrac{2}{5}$.

c) $P(M\cup F)=P(M)+P(F)-P(M\cap F)=0{,}25+0{,}15-0{,}10=0{,}30=\dfrac{3}{10}$.

V.2.9. Legyen $A=\{\text{férfi él még 10 évet}\}$, $B=\{\text{nő él még 10 évet}\}$, akkor $P(A)=\dfrac{1}{4}$, $P(B)=\dfrac{1}{3}$.

a) Az A és B független események, tehát

$$P(A\cap B)=P(A)P(B)=\frac{1}{4}\cdot\frac{1}{3}=\frac{1}{12}.$$

b) $P(A\cup B)=P(A)+P(B)-P(A\cap B)=\dfrac{1}{4}+\dfrac{1}{3}-\dfrac{1}{12}=\dfrac{1}{12}$.

c) A $P(\overline{A}\cap\overline{B})$ valószínűséget kell meghatározni.

$$P(\overline{A})=1-P(A)=1-\frac{1}{4}=\frac{3}{4}\ ;\ P(\overline{B})=1-P(B)=\frac{2}{3},$$

s mivel $\overline{A}$ és $\overline{B}$ esemény is független, ezért

$$P(\overline{A}\cap\overline{B})=P(\overline{A})P(\overline{B})=\frac{3}{4}\cdot\frac{2}{3}=\frac{1}{2}.$$

A megoldást az $\overline{A\cup B}=\overline{A}\cap\overline{B}$ összefüggés felhasználásával is előállíthatjuk:

$$P(\overline{A}\cap\overline{B})=P(\overline{A\cup B})=1-P(A\cup B)=1-\frac{1}{2}=\frac{1}{2}.$$

d) A $P(\overline{A}\cap B)$ valószínűséget kell meghatároznunk: $P(\overline{A})=1-P(A)=\dfrac{3}{4}$, s mivel $\overline{A}$ és B független események, ezért $P(\overline{A}\cap B)=P(\overline{A})P(B)=\dfrac{1}{4}$.

V.2.10. Események: $A=\{\text{Balázs talál}\}$, $B=\{\text{Robin talál}\}$, $C=\{\text{Nándor talál}\}$, tehát $P(A)=\dfrac{1}{6}$, $P(B)=\dfrac{1}{4}$, $P(C)=\dfrac{1}{3}$, a három esemény független, és

$$P(\overline{A})=\frac{5}{6},\quad P(\overline{B})=\frac{3}{4},\quad P(\overline{C})=\frac{2}{3};$$

a) A „pontosan egy személy talál a céltáblába" T eseményét

$$T=(A\cap\overline{B}\cap\overline{C})\cup(\overline{A}\cap B\cap\overline{C})\cup(\overline{A}\cap\overline{B}\cap C)$$

képlettel adhatjuk meg, s mivel ezek egymást kizáró események, ezért

$$p=p(T)=P(A\cap\overline{B}\cap\overline{C})+P(\overline{A}\cap B\cap\overline{C})+P(\overline{A}\cap\overline{B}\cap\overline{C})=$$

$$=P(A)P(\overline{B})P(\overline{C})+P(\overline{A})P(B)P(\overline{C})+P(\overline{A})P(\overline{B})P(C)=$$

$$=\frac{1}{6}\cdot\frac{3}{4}\cdot\frac{2}{3}+\frac{5}{6}\cdot\frac{1}{4}\cdot\frac{2}{3}+\frac{5}{6}\cdot\frac{3}{4}\cdot\frac{1}{3}=\frac{31}{72}.$$

b) Annak valószínűsége, hogy ha van találat azt Balázs érte el: $P(A|T)$. Mivel $A\cap T=A\cap\overline{B}\cap\overline{C}$ esemény jelenti, hogy csak Balázs ért el találatot, tehát

$$P(A\cap T)=P(A\cap\overline{B}\cap\overline{C})=\frac{1}{6}\cdot\frac{3}{4}\cdot\frac{2}{3}=\frac{1}{12},$$

másrészt

$$P(T)=\frac{31}{72},$$

így

$$P(A|T)=\frac{P(A\cap T)}{P(T)}=\frac{\frac{1}{12}}{\frac{31}{72}}=\frac{6}{31}.$$

V.2.11. Ha rövidlátó A, nem rövidlátó $\overline{A}$, lány B, fiú $\overline{B}$ jelölést használjuk az egyes eseményekre, akkor az adatokat a következő táblázatba rendezhetjük:

	B	$\overline{B}$	$\sum$sor
A	50	60	110
$\overline{A}$	450	1440	1890
$\sum$oszlop	500	1500	2000

a) $P(A)=\dfrac{110}{2000}=0{,}06$;

b) $P(A|B)=\dfrac{50}{500}=0{,}10$;

c) $P(A|\overline{B})=\dfrac{60}{1500}=0{,}04$;

d) $P(B|A)=\dfrac{50}{110}=0{,}45$.

V.2.12. Legyen A hibátlan, $\overline{A}$ hibás esemény, B_x, B_y, B_z az X, Y, Z üzem termékének eseménye. B_x, B_y, B_z egymást kizáró események, teljes rendszert alkotnak, teljesülnek a teljes valószínűség és a *Bayes*-tétel feltételei. Az adatok alapján:

$$P(B_x)=0{,}25,\quad P(A|B_x)=0{,}95,\quad P(\overline{A}|B_x)=0{,}05,$$

$$P(B_y)=0{,}35,\quad P(A|B_y)=0{,}96,\quad P(\overline{A}|B_y)=0{,}04,$$

$$P(B_z)=0{,}4,\quad P(A|B_z)=0{,}98,\quad P(\overline{A}|B_z)=0{,}02.$$

a) Az $\overline{A}$ esemény teljes valószínűsége:

$$P(\overline{A})=P(B_x)P(\overline{A}|B_x)+P(B_y)P(\overline{A}|B_y)+P(B_z)P(\overline{A}|B_z)=$$

$$=0{,}25\cdot 0{,}05+0{,}35\cdot 0{,}04+0{,}4\cdot 0{,}02=0{,}0345,$$

azaz a cipők $3{,}45\%$ hibás.

b) Az A esemény teljes valószínűsége:

$$P(A)=P(B_x)P(A|B_x)+P(B_y)P(A|B_y)+P(B_z)P(A|B_z)=$$

$$=0{,}25\cdot 0{,}95+0{,}35\cdot 0{,}96+0{,}4\cdot 0{,}98=0{,}9655,$$

amelyet az *a*) eredményéből is képezhetjük:

$$P(A)=1-P(\overline{A})=1-0{,}0345=0{,}9655,$$

vagyis a cipők $96{,}55\%$ hibátlan.

c) *Bayes* tétellel számítjuk, hogy az Y üzem a hibás cipők hány százalékát szállította:

$$P(B_y|\overline{A})=\frac{P(B_y)P(\overline{A}|B_y)}{P(\overline{A})}=\frac{0{,}014}{0{,}0346}=0{,}4058,$$

azaz a hibás cipők $40{,}58\%$ az Y gyárból származik.

d) A *c*)-hez hasonlóan kapjuk:

$$P(B_x|A)=\frac{P(B_x)P(A|B_x)}{P(A)}=\frac{0{,}2375}{0{,}9655}=0{,}246,$$

azaz a hibátlan cipőknek csak $24{,}6\%$ -a származik az X gyárból.

V.2.13. Az első szalagon gyártott hibás termék választás eseménye A, a második szalagon gyártotté pedig B. Mivel az A és B események függetlenek, így

$$P(AB)=P(A)P(B)=0{,}1\cdot 0{,}2=0{,}02,$$

$$P(\overline{A}\,\overline{B})=P(\overline{A})P(\overline{B})=0{,}9\cdot 0{,}8=0{,}72,$$

Az előző ellentettjeként számítható: $1-P(\overline{A}\,\overline{B})=0{,}28,$

$$P(A\overline{B}+\overline{A}B)=P(A\overline{B})+P(\overline{A}B)=0{,}1\cdot 0{,}8+0{,}9\cdot 0{,}2=0{,}26.$$

V.2.14. Az A_k jelentse azt, hogy a k-adik gép üzemszerűen működik, $(k=1,2,3,4,5)$. Az A_k -k függetlenek és $P(A_k)=0{,}9$.

a) $P(A_1A_2A_3A_4A_5)=P(A_1)P(A_2)P(A_3)P(A_4)P(A_5)=0{,}9^5=0{,}59049;$

b) $P(\overline{A_1}\,\overline{A_2}\,\overline{A_3}\,\overline{A_4}\,\overline{A_5})=P(\overline{A_1})P(\overline{A_2})P(\overline{A_3})P(\overline{A_4})P(\overline{A_5})=0{,}1^5=0{,}00001;$

Az előző esemény ellentettje: $p=1-0{,}00001=0{,}99999.$

„Egy gép sem működik", ill. „pontosan egy gép működik" egymást kizáró események, tehát $p=0{,}1^5+5\cdot 0{,}9\cdot 0{,}1^4=0{,}00045.$

V.2.15. Jelölje B_k $(k=1,2,3)$ azt, hogy a terméket a k-adik gép készítette. Ekkor

$$P(B_1)=\frac{10}{50},\quad P(B_2)=\frac{15}{50},\quad P(B_3)=\frac{25}{50} \text{ és}$$

$$P(A|B_1)=\frac{0{,}3}{10},\quad P(A|B_2)=\frac{0{,}9}{15},\quad P(A|B_3)=\frac{0{,}5}{25}.$$

Alkalmazzuk a *Bayes*-tételt:

$$P(B_1|A)=\frac{P(A|B_1)P(B_1)}{\sum_{k=1}^{3}P(A|B_k)P(B_k)}=\frac{\frac{0{,}3}{10}\cdot\frac{10}{50}}{\frac{0{,}3}{10}\cdot\frac{10}{50}+\frac{0{,}9}{15}\cdot\frac{15}{50}+\frac{0{,}5}{25}\cdot\frac{25}{50}}=\frac{3}{17}=0{,}17647.$$

Tehát a szépséghibás darabot kb. 0,18 valószínűséggel az X gép készítette.

V.3. A 3. fejezet feladatainak megoldásai

V.3.1. *a)* $m=\sum_i x_i v(x_i)=2\cdot\frac{1}{3}+3\cdot\frac{1}{2}+11\cdot\frac{1}{6}=4.$

$$s^2=\sum_i x_i^2 v(x_i)-m^2=26-16=10.$$

$$s=\sqrt{10}\approx 3{,}2.$$

b) $m=-1;\ s^2=9{,}25-1=8{,}25;\ s=\sqrt{8{,}25}\approx 2{,}87.$

V.3.2. A lehetséges esetek száma 16:

FFFF FFFI FFIF FFII FIFF FIFI FIIF FIII

IFFF IFFI IFIF IFII IIFF IIFI IIIF IIII

a) Az X eloszlása:

x_i	0	1	2	3	4
$v(x_i)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

(pl. 2 fej 6-szor fordul elő, stb.)

$$M(X)=2;\ s^2=\left(\mathrm{Var}(X)=D^2(X)\right)=1;\ D(X)=1.$$

b) Az Y eloszlása:

y_i	0	1	2	3	4
$v(y_i)$	$\frac{1}{16}$	$\frac{7}{16}$	$\frac{5}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

(pl. 2 fej egymás után 5 esetben következik be, stb.)

$$M(Y)=1{,}7;\ s^2=\left(\mathrm{Var}(Y)=D^2(Y)\right)=0{,}9;\ D(Y)\approx 0{,}95.$$

V.3.3. Mivel 3 fej dobásának valószínűsége $\frac{1}{8}$, 2 és 1 fej dobásának valószínűsége $\frac{3}{8}$, 3 írás dobás valószínűsége $\frac{1}{8}$, ezért a várható érték:

$$M=5\cdot\frac{1}{8}+3\cdot\frac{3}{8}+1\cdot\frac{3}{8}-15\cdot\frac{1}{8}=\frac{2}{8}=\frac{1}{4}=0{,}25\ \text{Ft},$$

tehát a játékos szempontjából kedvező a játék.

V.3.4. *a)* Peremeloszlások adják az X és Y eloszlását:

x_i	1	3
$v(x_i)$	0,5	0,5

y_i	−3	2	4
$v(y_i)$	0,4	0,3	0,3

b) A várható értékek:

$$m_x=\sum_i x_i v(x_i)=1\cdot 0{,}5+3\cdot 0{,}5=2;$$

$$m_y=\sum_j y_j u(y_j)=-3\cdot 0{,}4+2\cdot 0{,}3+4\cdot 0{,}3=0{,}6;$$

$$M(XY)=\sum_{i,j} x_i y_j w(x_i,y_j)=1\cdot(-3)\cdot 0{,}1+1\cdot 2\cdot 0{,}2+\ldots+3\cdot 4\cdot 0{,}1=0,$$

tehát

$$\mathrm{Cov}(X,Y)=M(XY)-m_x m_y=0-2\cdot 0{,}6=-1{,}2.$$

c) $D^2(X)=M(X^2)-m_x^2=\sum_i x_i^2 v(x_i)-m_x^2=1\cdot 0{,}5+9\cdot 0{,}5-2^2=1$

$$D(X)=\sqrt{1}=1,$$

és

$$D^2(Y)=M(Y^2)-m_y^2=\sum_j y_j^2 v(y_j)-m_y^2=9\cdot 0{,}4+4\cdot 0{,}3+16\cdot 0{,}3-0{,}6^2=9{,}24$$

$$D(Y)=\sqrt{9{,}24}\approx 3{,}04$$

tehát a korrelációs együttható:

$$R(X,Y)=\frac{\mathrm{Cov}(X,Y)}{D(X)D(Y)}=\frac{-1{,}2}{1\cdot 3{,}04}\approx -0{,}39.$$

d) Az X és Y nem független, mivel pl.

$$P(X=1,Y=-3)\neq P(X=1)\cdot P(Y=-3),$$

ui. $w(1,-3)=0{,}1$ ez pedig nem egyenlő a megfelelő peremeloszlások szorzatával:

$$v(1)\cdot u(-3)=0{,}5\cdot 0{,}4=0{,}2.$$

V.3.5. *A)*

a) $M(X)=3;\ M(Y)=1;$

b) $\mathrm{Cov}(X,Y)=1{,}5;$

c) $D(X)=2;\ D(Y)=4{,}3;\ \mathrm{R}(X,Y)=0{,}17.$

B)

a) $M(X)=1{,}4$; $M(Y)=1$;

b) $\mathrm{Cov}(X,Y)=-0{,}5$;

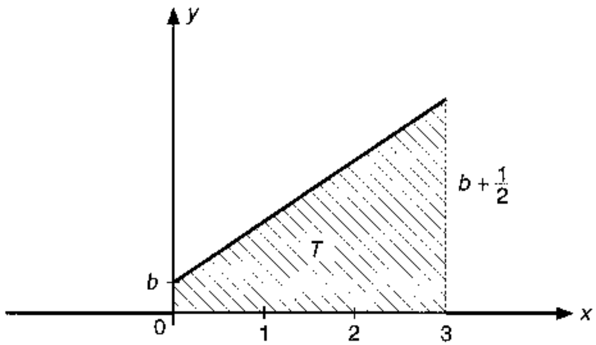
c) $D(X)=0{,}49$; $D(Y)=3{,}1$; $R(X,Y)=-0{,}3$.

V.3.6. Az eloszlás táblázata $w(x_i,y_j)=P(x_i)P(y_j)$ képlet alapján állítható elő, mivel X és Y független valószínűségi változók.

$X \setminus Y$	-2	5	8	$\sum$
1	0,21	0,35	0,14	0,7
2	0,09	0,15	0,06	0,3
$\sum$	0,3	0,5	0,2	

V.3.7. *a)* Mivel f folytonos valószínűségi változó sűrűségfüggvénye, ezért, feltéve hogy $b>0$, az 50. ábrán a $[0;3]$ intervallum feletti és az $y=\frac{1}{6}x+b$ egyenes alatti bevonalkázott terület 1 területegységgel egyenlő és így

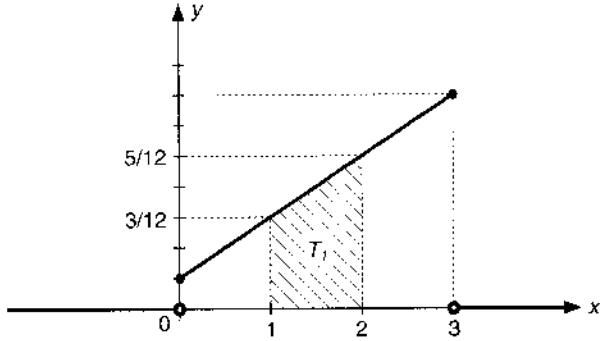
$$T=\frac{1}{2}\left(b+b+\frac{1}{2}\right)\cdot 3=1,\ \text{amelyből}\ b=\frac{1}{12}.$$



50. ábra

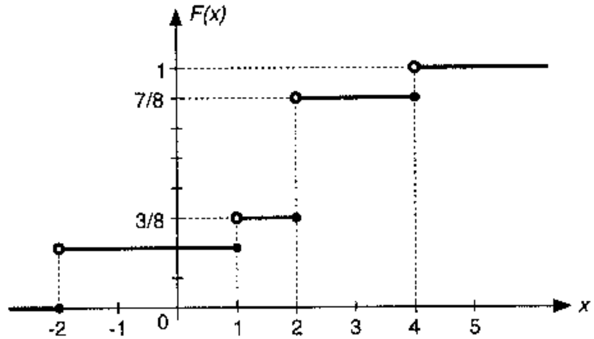
b) A $P(1\le X\le 2)$ értéke egyenlő a 51. ábra bevonalkázott T_1 területével.

Mivel $f(1)=\frac{1}{6}+\frac{1}{12}=\frac{3}{12}$, $f(2)=\frac{1}{3}+\frac{1}{12}=\frac{5}{12}$, ezért $\frac{1}{2}\left(\frac{3}{12}+\frac{5}{12}\right)\cdot 1=\frac{1}{3}$ területegységgel egyenlő a T_1 terület, tehát $P(1\le X\le 2)=\frac{1}{3}$.



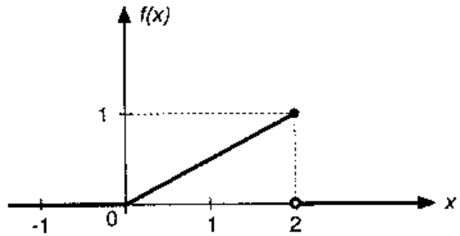
51. ábra

V.3.8.



52. ábra

V.3.9. Az f gráfja:

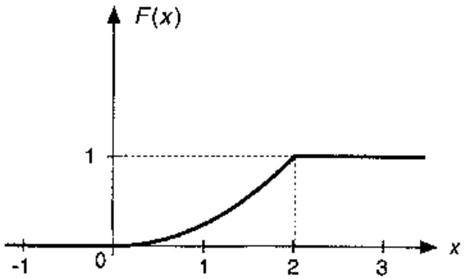


53. ábra

Az eloszlásfüggvény a $0 \le x \le 2$ intervallumon:

$$F(x)=\int_0^x \frac{1}{2}t\,dt=\frac{1}{4}x^2,\ \text{és így}\ F(x)=\begin{cases}0, & \text{ha } x<0\\ \frac{1}{4}x^2, & \text{ha } 0\le x\le 2\\ 1, & \text{ha } x>2.\end{cases}$$

Az $F(x)$ gráfja:



54. ábra

V.3.10. Elsőrendű momentum: $M_1=\sum_i x_i v(x_i)=-2\cdot\frac{1}{2}+1\cdot\frac{1}{4}+3\cdot\frac{1}{4}=0;$

Másodrendű momentum: $M_2=\sum_i x_i^2 v(x_i)=4\cdot\frac{1}{2}+1\cdot\frac{1}{4}+9\cdot\frac{1}{4}=4{,}5;$

Harmadrendű momentum: $M_3=\sum_i x_i^3 v(x_i)=-8\cdot\frac{1}{2}+1\cdot\frac{1}{4}+27\cdot\frac{1}{4}=3;$

Negyedrendű momentum: $M_4=\sum_i x_i^4 v(x_i)=16\cdot\frac{1}{2}+1\cdot\frac{1}{4}+81\cdot\frac{1}{4}=28{,}5.$

V.3.11. $Q=\{FFF, FFI, FIF, FII, IFF, IFI, IIF, III\}$.

F és *F* és *F* dobás valószínűsége: $P(FFF)=\frac{2}{3}\cdot\frac{2}{3}\cdot\frac{2}{3}=\frac{8}{27}$, és így tovább

$$P(FFI)=\frac{2}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}=\frac{4}{27},\quad P(FIF)=\frac{4}{27},\quad P(FII)=\frac{2}{27},$$

$$P(IFF)=\frac{4}{27},\quad P(IFI)=\frac{2}{27},\quad P(IIF)=\frac{2}{27},\quad P(III)=\frac{1}{27},$$

és

$$X(III)=0,\qquad X(FIF=X(FII)=F(IFF)=X(IFF)=1,$$

$$X(FFI)=X(IFF)=2,\qquad X(FFF)=3.$$

Tehát $X(Q)=\{0,1,2,3\}$, és

$$v(0)=P(III)=\frac{1}{27},\quad v(1)=P(FIF,FII,IFI,IIF)=\frac{4}{27}+\frac{2}{27}+\frac{2}{27}+\frac{2}{27}=\frac{10}{27},$$

$$v(2)=P(FFI,IFF)=\frac{4}{27}+\frac{4}{27}=\frac{8}{27},\quad v(3)=P(FFF)=\frac{8}{27}.$$

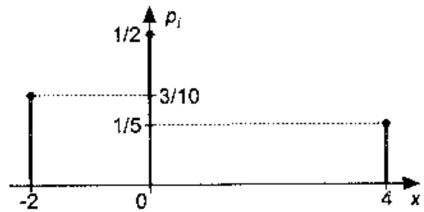
Az eloszlás táblázat:

x_i	0	1	2	3
$v(x_i)$	$\frac{1}{27}$	$\frac{10}{27}$	$\frac{8}{27}$	$\frac{8}{27}$

Várható érték:

$$M(X)=\sum_{i=1}^{4}x_i v(x_i)=0\cdot\frac{1}{27}+1\cdot\frac{10}{27}+2\cdot\frac{8}{27}+3\cdot\frac{8}{27}=\frac{50}{27}=1{,}85.$$

V.3.12. *a)* Az *X* valószínűségi változó eloszlásának gráfja:



55. ábra. X valószínűségeloszlása

Minden $x\in R$-re meg kell határozni az $F(x)=P(X<x)$ valószínűséget:

$x\le -2:,\quad F(x)=P(X<x)=0,$

$-2<x\le 0:\quad F(x)=P(X<x)=P(X=-2)=\frac{3}{10},$

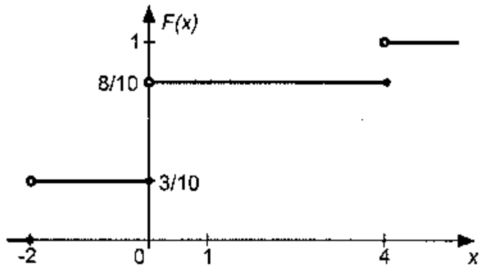
$0<x\le 4:\quad F(x)=F(X<x)=P(X=-2)+P(X=0)=\frac{3}{10}+\frac{1}{2}=\frac{8}{10},$

$4<x:\quad F(x)=P(X<x)=P(X=-2)+P(X=0)+P(4)=$

$$=\frac{3}{10}+\frac{1}{2}+\frac{1}{5}=\frac{10}{10}=1.$$

Az X valószínűségi változó eloszlásfüggvénye:

$$F(x)=\begin{cases}0, \text{ ha } x\le -2\\ \frac{3}{10}, \text{ ha } -2<x\le 0\\ \frac{4}{5}, \text{ ha } 0<x\le 4\\ 1, \text{ ha } 4<x.\end{cases} \quad x\in R;$$



56. ábra. Az X eloszlásfüggvénye

b) $F(-3)=P(X<-3)=0 \qquad F(-2)=P(X<-2)=0$

$F(1)=P(X<1)=\frac{8}{10} \qquad F(5)=P(X<5)=1$

c) $P(X\ge 0)=1-P(<0)=1-F(0)=1-\frac{3}{10}=\frac{3}{10};$

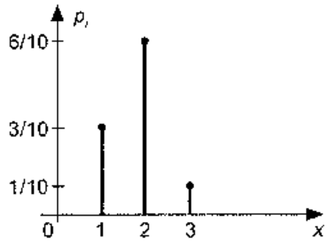
$$P(X\le 0)=P(X<4)=F(4)=\frac{8}{10}.$$

V.3.13. Az X változó diszkrét értékei: 0, 1, 2, 3. Az 5 elemből 3-at 10-féleképpen $\left[\binom{5}{3}=\frac{5\cdot 4\cdot 3}{1\cdot 2\cdot 3}=10\right]$ választhatunk.

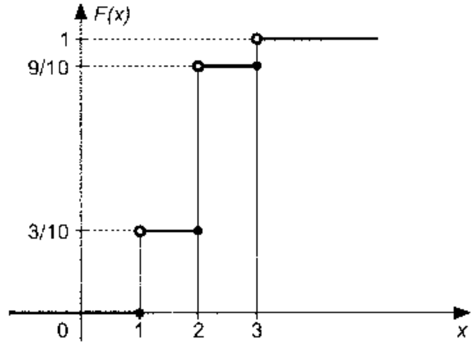
a) $P(X=0)=0; \qquad P(X=1)=\frac{3}{10};$

$P(X=2)=\frac{6}{10}; \qquad P(X=3)=\frac{1}{10};$ (57. ábra)

b) $$F(x)=\begin{cases}0, \text{ ha } x\le 1\\ \frac{3}{10}, \text{ ha } 1<x\le 2\\ \frac{9}{10}, \text{ ha } 2<x\le 3\\ 1, \text{ ha } 3<x\end{cases} \quad x\in R \text{ (58.ábra).}$$



57. ábra. X valószínűségeloszlása



58. ábra. F(x) eloszlásfüggvény

c) Legalább egy friss tojás van a kiválasztott 3 tojás között:

$$P(X\ge 1)=1-P(X<1)=1-F(1)=1-0=1;$$

Legfeljebb egy friss tojás van a kiválasztott 3 tojás között:

$$P(X\le 1)=P(X<2)=F(2)=\frac{3}{10}.$$

V.3.14. Az (X,Y) valószínűségi vektorváltozó együttes valószínűségei és peremvalószínűségei:

a)

$X\setminus Y$	20	40	60	100	$\sum_{sor}$
110	$\frac{50}{2000}$	$\frac{100}{2000}$	$\frac{100}{2000}$	$\frac{250}{2000}$	$\frac{500}{2000}$
220	$\frac{250}{2000}$	$\frac{300}{2000}$	$\frac{400}{2000}$	$\frac{550}{2000}$	$\frac{1500}{2000}$
$\sum_{oszlop}$	$\frac{300}{2000}$	$\frac{400}{2000}$	$\frac{500}{2000}$	$\frac{800}{2000}$	1

b) $$F(x,y)=\begin{cases}0, & \text{ha } x\le 110 \text{ vagy } y\le 20\\ \frac{50}{2000}, & \text{ha } 110<x\le 220 \text{ és } 20<y\le 40\\ \frac{300}{2000}, & \text{ha } 220<x \text{ és } 20<y\le 40\\ \frac{150}{2000}, & \text{ha } 110<x \text{ és } 40<y\le 60\\ \frac{700}{2000}, & \text{ha } 220<x \text{ és } 40<y\le 60\\ \frac{250}{2000}, & \text{ha } 110<x\le 220 \text{ és } 60<y\le 100\\ \frac{1200}{2000}, & \text{ha } 220<x \text{ és } 60<y\le 100\\ \frac{500}{2000}, & \text{ha } 110<x\le 220 \text{ és } 100<y\\ 1, & \text{ha } 220<x \text{ és } 100<y\end{cases}\quad (x,y)\in R^2.$$

$$F_m(x)=\begin{cases}0, & \text{ha } x\le 110\\ \frac{500}{2000}, & \text{ha } 110<x\le 220\\ 1, & \text{ha } 220<x\end{cases}$$

$$F_m(y)=\begin{cases}0, & \text{ha } y\le 20\\ \frac{300}{2000}, & \text{ha } 20<y\le 40\\ \frac{700}{2000}, & \text{ha } 40<y\le 60\\ \frac{1200}{2000}, & \text{ha } 60<y\le 100\\ 1, & \text{ha } 100<y\end{cases}\quad x\in R,\ y\in R.$$

c) $M(X)=0{,}6266$; $D(X)=0{,}3454$; $M(Y)=2{,}0633$; $D(Y)=0{,}5338$;

$M(XY)=\sum_{i,j}x_i y_j w(x_i,y_j)=1{,}2556$;

$\mathrm{Cov}(X,Y)=M(XY)-M(X)M(Y)=-0{,}0373$;

$R(X,Y)=\dfrac{\mathrm{Cov}(X,Y)}{D(X)D(Y)}=-0{,}2023.$

V.4. A 4. fejezet feladatainak megoldásai

V.4.1. *a)* $b\left(3;6,\frac{1}{2}\right)=\binom{6}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^3=\frac{6\cdot5\cdot4}{1\cdot2\cdot3}\cdot\frac{1}{8}\cdot\frac{1}{8}=\frac{5}{16}$;

b) $b\left(3;4,\frac{1}{4}\right)=\binom{4}{3}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)=\frac{3}{64}$.

V.4.2. Mivel $n=4$, $p=\frac{2}{3}$ és $q=1-p=\frac{1}{3}$, ezért

a) $P(2 \text{ játszma nyerése})=b\left(2;4,\frac{2}{3}\right)=\binom{4}{2}\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^2=\frac{8}{27}$.

b) Mind a 4 játszma elvesztésének valószínűsége: $q^4=\left(\frac{1}{3}\right)^4=\frac{1}{81}$, tehát annak, hogy az A csapat legalább egy játszmát nyer $1-q^4=\frac{80}{81}$ a valószínűsége.

c) 3 vagy 4 játszma megnyerése jelenti azt, hogy a 4 játszmából felénél többet nyer az A csapat, tehát

$P(3 \text{ játszma nyerése})+P(4 \text{ játszma nyerése})=$

$$=\binom{4}{3}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)+\binom{4}{4}\left(\frac{2}{3}\right)^4=\frac{32}{81}+\frac{16}{81}=\frac{16}{27}.$$

V.4.3. *a)* Mivel $p=0{,}02$, ezért $m=np=10000\cdot0{,}02=200$, és így

$$s=\sqrt{npq}=\sqrt{10000\cdot0{,}02\cdot0{,}98}=\sqrt{196}=14.$$

b) Selejtes darab választásának valószínűsége: $p=0{,}02$ és így

$$q=1-p=0{,}98.$$

Az X valószínűségi változó értéke legyen n db-ból a selejtesek száma, így X binomiális eloszlású. Annak valószínűsége, hogy n db-ból egy sem selejtes, azaz

$$X=0: P(X=0)=\binom{n}{0}p^0q^n=0{,}98^n.$$

Tehát n értékét az $1-0{,}98^n\ge0{,}96$ egyenlőtlenségből számíthatjuk ki, amelyből

$$0{,}98^n\le0{,}04$$

$$n\lg0{,}98\le\lg0{,}04$$

s mivel

$$\lg 0{,}98 < 0$$

ezért

$$n \geq \frac{\lg 0{,}04}{\lg 0{,}98} \approx 159{,}33.$$

Tehát legalább 160 darabot kell visszatevéssel megvizsgálni, hogy a kiválasztottak között 0,96-nál nagyobb valószínűséggel legyen selejtes darab.

V.4.4. *a)* A $k = 2,3,4,5,6,7$ értékekre kiszámított valószínűségek összegezése helyett, vonjuk ki 1-ből a 0 és az 1 találat valószínűségének összegét:

$$P(0\ \text{találat}) = \left(\frac{3}{4}\right)^7 = \frac{2187}{16384};$$

$$P(1\ \text{találat}) = \binom{7}{1}\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^6 = \frac{5103}{16384},$$

tehát

$$P = 1 - \left(\frac{2187}{16384} + \frac{5103}{16384}\right) = \frac{4547}{8192}.$$

b) Annak valószínűsége, hogy a lövés nem talál célba: q^n. Mivel $q = 1 - p = = 1 - \frac{1}{4} = \frac{3}{4}$ és q^n-nek $1 - \frac{2}{3} = \frac{1}{3}$-nál kel kisebbnek lenni, ezért a $q^n < \frac{1}{3}$, azaz $\left(\frac{3}{4}\right)^n < \frac{1}{3}$ egyenlőtlenségből meghatározhatjuk n értékét:

$$n \lg\left(\frac{3}{4}\right) < \lg\left(\frac{1}{3}\right) \text{ és így } n \geq \frac{\lg\left(\frac{1}{3}\right)}{\lg\left(\frac{3}{4}\right)} \approx 3{,}8184 \approx 4,$$

tehát a lövész $\frac{2}{3}$ valószínűséggel célba talál, ha legalább 4 lövést ad le.

V.4.5. *a)* A számítást, ha $\lambda = 1{,}8$ a

$$P(X = k) = p_k = \frac{1{,}8^k}{k!} e^{-1{,}8} \quad (k = 0,1,2,3,4)$$

formulával végezzük.

$$p_0 = e^{-1{,}8} = 0{,}1653;$$

$$p_1 = 1{,}8 e^{-1{,}8} = 0{,}2976;$$

$$p_2 = \frac{1{,}8^2}{2!} e^{-1{,}8} = \frac{3{,}24}{2} e^{-1{,}8} = 0{,}2678;$$

$$p_3 = \frac{1{,}8^3}{3!} e^{-1{,}8} = \frac{5{,}832}{6} e^{-1{,}8} = 0{,}1607;$$

$$p_4 = \frac{1{,}8^4}{4!} e^{-1{,}8} = \frac{10{,}4976}{24} e^{-1{,}8} = 0{,}0723.$$

b) Várható érték: $M(X) = \lambda = 1{,}8$;

szórás: $D(X) = \sqrt{\lambda} = \sqrt{1{,}8} = 1{,}3416.$

Annak valószínűsége, hogy X várható értékénél kisebb értéket vesz fel $P(X < 1{,}8)$ kiszámításával adható meg

$$P(X < 1{,}8) = p_0 + p_1 = 0{,}1653 + 0{,}2976 = 0{,}4629.$$

V.4.6. Bár $n = 100$, $p = 0{,}02$ paraméterű binomiális eloszlású valószínűségi változó értékét kellene kiszámítani, de n elég nagy és p elég kicsi, így $\lambda = np = 100 \cdot 0{,}02 = 2$ paraméterű *Poisson*-eloszlás $p(3;2)$ értékével közelítjük $b(3;100,0{,}02)$ értékét:

$$P = p(3;2) = 2^3 \cdot \frac{e^{-2}}{3!} = 8 \cdot \frac{0{,}1354}{6} = 0{,}1805.$$

V.4.7. Az X szórása és várható értéke azonos.

a) $M(X) = D(X) = 800$ óra, ezért az eloszlás paramétere: $\lambda = \frac{1}{800}$.

b) A sűrűségfüggvény: $f(x) = \begin{cases} 0, \text{ ha } x \leq 0 \\ \frac{1}{800} e^{-\frac{x}{800}}, \text{ ha } x > 0 \end{cases}$

és az eloszlásfüggvény: $F(x) = P(X < x) = \begin{cases} 0, \text{ ha } x \leq 0 \\ 1 - e^{-\frac{x}{800}}, \text{ ha } x > 0. \end{cases}$

c) Annak valószínűsége, hogy a kijelölt képcső 3200 órán belül nem hibásodott meg:

$$P(X \geq 3200) = 1 - P(X < 3200) = 1 - F(3200) = 1 - \left(1 - e^{-\frac{3200}{800}}\right) = \\ = e^{-4} \approx 0{,}0183.$$

V.4.8. A 2. táblázatból olvashatók ki a sűrűségfüggvény $\varphi(x)$ értékei:

a) $\varphi(1{,}63) = 0{,}1057$;

b) $\varphi(-0{,}75) = \varphi(0{,}75) = 0{,}3011$;

c) $\varphi(-2{,}08) = \varphi(2{,}08) = 0{,}0459.$

V.4.9. Az 1. táblázatból olvashatók ki a $\Phi(x)$ értékei:

a) $P(0 \le X \le 1{,}42) = \Phi(1{,}42) - \Phi(0) = 0{,}9222 - 0{,}5 = 0{,}4222;$

b) $P(-1{,}37 \le X \le 2{,}01) = \Phi(2{,}01) - \Phi(-1{,}37) = \Phi(2{,}01) - (1 - \Phi(-1{,}37)) =$

$$= 0{,}9778 - 1 + 0{,}9147 = 0{,}8925.$$

c) $P(X \ge 1{,}13) = 1 - P(X < 1{,}13) = 1 - 0{,}8708 = 0{,}1292.$

d) $P(|X| \le 0{,}5) = P(-0{,}5 \le X \le 0{,}5) = -(1 - \Phi(0{,}5)) + \Phi(0{,}5) =$

$$= -1 + 2\Phi(0{,}5) = -1 + 2 \cdot 0{,}6915 = 0{,}3830.$$

V.4.10. *a)* 166 cm standard egysége: $\dfrac{166-170}{16} = -0{,}25;$

182 cm standard egysége: $\dfrac{182-170}{16} = 0{,}75.$

Tehát $P(166 \le L \le 182) = P(-0{,}25 \le L^* \le 0{,}75) =$

$$= \left(\Phi(0{,}75) - (1 - \Phi(0{,}25))\right) = 0{,}7734 - 1 + 0{,}5987 = 0{,}3721,$$

és így $L = 400 \cdot 0{,}3721 \approx 149$ hallgató esik a 166 cm-rel és a 182 cm-rel adott magassági határok közé.

b) A 190 cm standard egység: $\dfrac{190-170}{16} = 1{,}25.$

Tehát $P(L \ge 190) = P(L^* \ge 1{,}25) = 1 - \Phi(1{,}25) = 1 - 0{,}8944 = 0{,}1056$ és így

$$L = 400 \cdot 0{,}1056 \approx 42$$

hallgató nagyobb vagy éppen 190 cm.

V.4.11. Az exponenciális eloszlású valószínűségi változó várható értéke, mivel $f(x) = 0$, ha $x \le 0$, így

$$M(X) = \int_0^\infty x\lambda e^{-\lambda x}dx = \lim_{R\to\infty}\int_0^R x\lambda e^{-\lambda x}dx = \lim_{R\to\infty}\left(-Re^{-\lambda R} - \frac{1}{\lambda}\left[e^{-\lambda x}\right]_0^R\right) = \frac{1}{\lambda},$$

(a határérték kiszámításához alkalmaztuk a *L'Hospitál* szabályt). A szórásnégyzet kiszámításához előbb az X^2 várható értékét számítjuk ki, az előzőhöz hasonlóan parciális integrálást alkalmazva: $M(X^2) = \int_0^\infty x^2\lambda e^{-\lambda x}dx = 2\dfrac{1}{\lambda^2}.$

Tehát $D^2(X) = M(X^2) - M^2(X) = \dfrac{2}{\lambda^2} - \dfrac{1}{\lambda^2} = \dfrac{1}{\lambda^2}$, melyből négyzetgyökvonással kapjuk a szórást: $P(X = 4) = \binom{4}{4}\left(\dfrac{5}{100}\right)^4\left(\dfrac{95}{100}\right)^0 = 0{,}0000062.$

V.4.12. Visszatevés nélküli minta választásánál X hipergeometrikus eloszlású, értékei: 0, 1, 2, 3, 4, visszatevéses minta választása esetén pedig X binomiális eloszlású, értékei ugyancsak: 0, 1, 2, 3, 4.

Visszatevés nélküli választás esetén:

$P(X = 0) = 0{,}8119,\quad P(X = 1) = 0{,}1765,\quad P(X = 2) = 0{,}0114,$

$P(X = 3) = 0{,}0002,\quad P(X = 4) = 0{,}000001.$

$M(X) = np = 4 \cdot \dfrac{5}{100} = 0{,}2,$

$D(X) = 0{,}43.$

Visszatevéses választás esetén:

$$P(X = 0) = \binom{4}{0}\left(\frac{5}{100}\right)^0\left(\frac{95}{100}\right)^4 = 0{,}8145,$$

$$P(X = 1) = \binom{4}{1}\left(\frac{5}{100}\right)^1\left(\frac{95}{100}\right)^3 = 0{,}1715,$$

$$P(X = 2) = \binom{4}{2}\left(\frac{5}{100}\right)^2\left(\frac{95}{100}\right)^2 = 0{,}01354,$$

$$P(X = 3) = \binom{4}{3}\left(\frac{5}{100}\right)^3\left(\frac{95}{100}\right)^1 = 0{,}000475,$$

$$P(X = 4) = \binom{4}{4}\left(\frac{5}{100}\right)^4\left(\frac{95}{100}\right)^0 = 0{,}0000062,$$

$M(X) = np = 4\dfrac{5}{100} = 0{,}2,$

$D(X) = \sqrt{npq} = 0{,}4359.$

V.4.13. Az X normális eloszlású valószínűségi változó jelentse az átmérőt, akkor $m = 8$, $\sigma = 0{,}1$, és a 4%-os eltérés: $0{,}04 \cdot 8 = 0{,}32$ adataival: $P(|X - 8| \ge 0{,}32) =$

$= -P(7{,}68 < X < 8{,}32) = 1 - (F(8{,}32) - F(7{,}68)) =$

$$= 1 - (\Phi(3{,}2) - \Phi(-3{,}2)) = 2 - 2\Phi(3{,}2) = 0{,}00138.$$

V.5. Az 5. fejezet feladatainak megoldásai

V.5.1. A *Csebisev*-egyenlőtlenség felhasználásával:

$$P(|X-20|\geq 50)\leq \frac{20^2}{50^2}=\frac{4}{25}=0{,}16.$$

V.5.2. A *Csebisev*-egyenlőtlenséget használva $\varepsilon=1$ értékre:

$$P(|X-35|\geq 1)=\frac{0{,}3^2}{1^2}=0{,}09$$

tehát legfeljebb 0,09 annak valószínűsége, hogy a kötél hossza *nem esik* a 34 m és 36 m közé.

V.5.3. A hibás áru választásának eseménye legyen A.

$$p=0{,}1;\quad q=1-p=0{,}9.$$

$$P\left(\left|\frac{k}{n}-0{,}1\right|\geq 0{,}02\right)\leq\frac{0{,}1\cdot 0{,}9}{0{,}02^2\cdot n},$$

és

$$\frac{0{,}09}{0{,}0004\cdot n}<0{,}05$$

feltételből

$$n>\frac{0{,}09}{0{,}0004\cdot 0{,}05}=\frac{0{,}09}{0{,}00002}=4500,$$

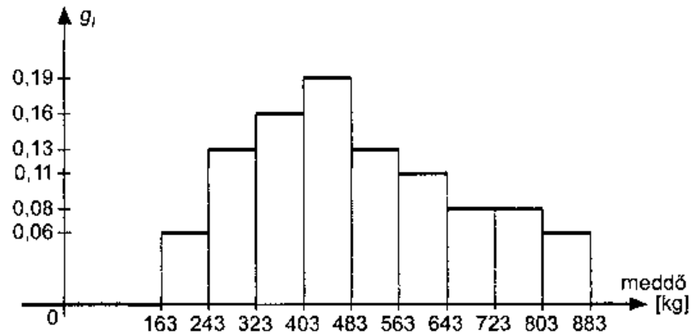
vagyis, ha a tételben 4500-nál több kesztyű van, akkor 0,95 valószínűséggel a vevő átveszi azt.

A II. RÉSZ FELADATAINAK MEGOLDÁSAI

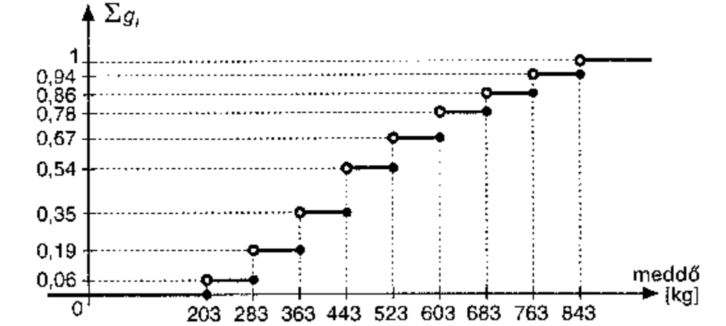
S.1. Az 1. fejezet feladatainak megoldásai

S.1.1. Kiszámítjuk a g_i relatív gyakoriságokat és a $\sum_i g_i$ értékeket:

Meddő mennyiség kg (osztályközök)	f_i	g_i	$\sum_i g_i$
.....–243	8	0,06	0,06
244–323	19	0,13	0,19
324–403	23	0,16	0,35
404–483	27	0,19	0,54
484–563	18	0,13	0,67
564–643	15	0,11	0,78
644–723	12	0,08	0,86
724–803	12	0,08	0,94
804 –	8	0,06	1,00
	$\sum_i f_i=142$	$\sum_i g_i=1$	



59. ábra. Sűrűséghisztogram

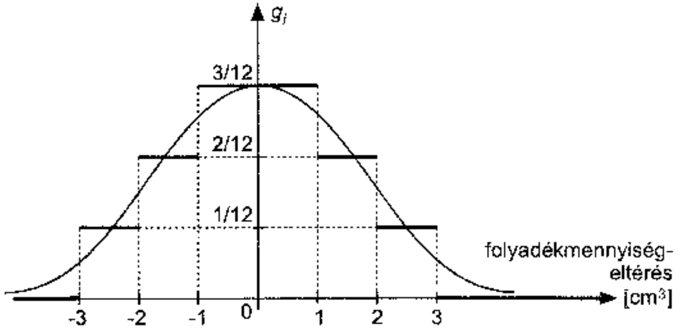


60. ábra. Tapasztalati eloszlásfüggvény

S.1.2. *a*) Kiszámítjuk a relatív gyakoriságot és a g_i részösszegeket, majd megrajzoljuk a sűrűséghisztogramot (tapasztalati sűrűségfüggvényt) intervallumonként

$$f(x) \approx g_i = \frac{\text{intervallumra eső adatok száma}}{(\text{összes adat száma}) \cdot (\text{intervallum hossza})}$$

képlettel számított értékek alapján. $]-\infty; -3[$ és $]3; +\infty[$ intervallumokba nem esett adat, ezért ezen intervallumokban a sűrűségfüggvényt 0-nak tekintjük.

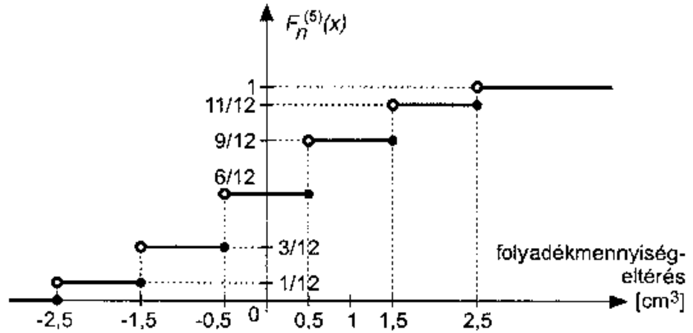


61. ábra. Sűrűséghisztogram

Megszerkesztjük az

$$F_n^{(5)}(x) = P(X < x) = \begin{cases} 0, & \text{ha } x \le -2{,}5 \\ \sum_{i=1}^{k} g_i, & \text{ha } x_{k-1} < x \le x_k \ (k = 1,2,3,4,5) \\ 1, & \text{ha } x > 2{,}5 \end{cases}$$

tapasztalati eloszlásfüggvény ábráját. A $]-\infty; -2{,}5]$ intervallumba nem esik folyadékmennyiség-eltérés, a $]-2{,}5; -1{,}5]$ intervallumba 1 esik a 12 közül, $P(X < -1{,}5) = \frac{1}{12}$, a $]-1{,}5; -0{,}5]$ intervallumba 2 esik, tehát $P(X < -0{,}5) = \frac{1+2}{12} = \frac{3}{12}$, stb. A balról nyílt jobbról zárt intervallumok mindegyikében a függvény állandó. Az intervallum végpontjaiban $f_i \cdot \frac{1}{12}$ az ugrás mértéke.



62. ábra. Tapasztalati eloszlásfüggvény

c) Mintaátlag: $\bar{x} = 0$.

d) Annak valószínűsége, hogy a folyadékmennyiség-eltérés 1 cm^3 -nél nagyobb:

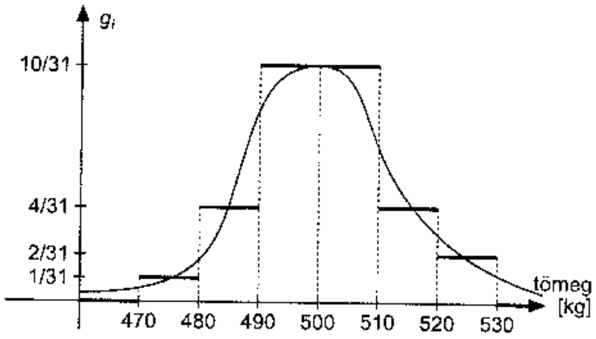
$$P(X \ge 1) = 1 - P(X < 1) = 1 - F(1) = 1 - \Phi\left(\frac{1-0}{1{,}45}\right) = 1 - \Phi(0{,}69) = 1 - 0{,}7549 = 0{,}2451$$

vagyis 24,51%-ban kapunk olyan palackot, amelyben az előírthoz képest 1 cm^3 -nél több a folyadékmennyiség eltérése.

S.1.3. Kiszámítjuk és táblázatba foglaljuk a szükséges részeredményeket:

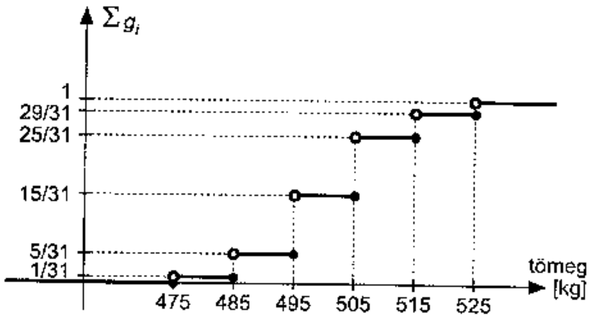
osztályhatár kg	osztály-közép x_i	gyak. f_i	g_i	$\sum_i g_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
470–480	475	1	1/31	1/31	–25,8	665,64
480–490	485	4	4/31	5/31	–15,8	249,64
490–500	495	10	10/31	15/31	–5,8	33,64
500–510	505	10	10/31	25/31	4,2	17,64
510–520	515	4	4/31	24/31	14,2	201,64
520–530	525	2	2/31	31/31	24,2	585,64
		$\sum_i f_i = 31$	$\sum_i g_i = 1$			

a) A 63. ábra a sűrűségfüggvényt (hisztogramot) a normális eloszlás sűrűségfüggvényével együtt szemlélteti (a gyakoriságértékeket $10 \cdot 31 = 310$ -zel osztottuk).



63. ábra. Hisztogram és $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$ *gráfja*

b) A tapasztalati eloszlásfüggvény a táblázatból vett $\sum_i g_i$ értékekkel (64. ábra):



64. ábra. Tapasztalati eloszlásfüggvény

c) Mintaátlag:

$$\bar{x} = \frac{1 \cdot 475 + 4 \cdot 485 + 10 \cdot 495 + 10 \cdot 505 + 4 \cdot 515 + 2 \cdot 525}{31} = 500{,}81.$$

d) Korrigált tapasztalati szórás:

$$s^* = \sqrt{\frac{1 \cdot 665{,}64 + 2 \cdot 249{,}64 + \ldots + 2 \cdot 585{,}64}{30}} = 11{,}77.$$

Annak valószínűsége, hogy a gép 485 kg-nál kevesebbet markol:

$$P(X < 485) \approx F_n^{(6)}(485) = \frac{1}{31} \approx 0{,}032,$$

vagyis a markolások számának 3,2%-ában 485 kg-nál kevesebb a kimarkolt homok súlya.

S.1.4. *a)* $\overline{X} = 12{,}54$; $s^* = 0{,}93$; $\frac{s^*}{\overline{X}} = 0{,}07$;

b) $F(13{,}8) - F(11{,}8) = \Phi(1{,}35) - \Phi(-0{,}8) = 0{,}6996$, vagyis az elemeknek kb. 70%-a felel meg az előírt pontosságnak.

S.2. A 2. fejezet feladatainak megoldásai

S.2.1. A 90, 95 és 99%-nak a $p_1 = 0{,}1$; $p_2 = 0{,}05$ és $p_3 = 0{,}01$ értékek felelnek meg. Az 1. sz. táblázat felhasználásával rendre $\left(\Phi(u_{p_1}) = 1 - \frac{p_1}{2} = 1 - \frac{0{,}1}{2} = 0{,}95 \text{ stb.}\right)$:

$$u_{p_1} = 1{,}645;\ u_{p_2} = 1{,}960 \text{ és } u_{p_3} = 2{,}5785.$$

Mivel $n = 100$, $\sigma = 50$ és $\overline{X} = 705$, ezért a 90%-os szintnek megfelelő konfidencia-határok:

$$\overline{X} - u_{p_1} \cdot \frac{\sigma}{\sqrt{n}} = 705 - 1{,}645 \cdot \frac{50}{\sqrt{100}} = 696{,}8\,\text{kg},$$

$$\overline{X} + u_{p_1} \cdot \frac{\sigma}{\sqrt{n}} = 705 + 1{,}645 \cdot \frac{50}{\sqrt{100}} = 713{,}2\,\text{kg}.$$

A töltési súly 90% valószínűséggel esik a $[696{,}8;713{,}2]$ intervallumba.

Hasonlóan kapjuk a 95 ill. a 99%-nak megfelelő konfidencia-intervallumokat:

$$[695{,}2;714{,}8],\ \text{ill.}\ [692{,}1;719{,}9]$$

S.2.2. A (6) formula alapján, mivel $\sigma = 50$, $u_p = 1{,}960$ és $d = 7$:

$$n > u_p^2 \cdot \frac{\sigma^2}{d^2} = 1{,}96^2 \cdot \frac{2500}{49} \approx 196.$$

Tehát 196 puttony leméréséből kapott mérési eredményekből számított számtani középtől ± 7 kg eltérésű intervallumba 95% valószínűséggel esik az 5000 puttony átlagos töltési súlya.

S.2.3. A 30 puttony mérlegelése alapján kapott átlagos töltési súly $\overline{X} = 708$ kg, a korrigált tapasztalati szórás $s^* = 52$ kg. A szabadságfokok száma $n - 1 = 29$. A 90, 95 és 99%-os szinteknek megfelelő p értékek rendre $p_1 = 0{,}1$; $p_2 = 0{,}05$ és $p_3 = 0{,}01$. A 4. táblázatból az ezekhez tartozó t_p értékek:

$$t_{p_1} = 1{,}699;\ t_{p_2} = 2{,}045 \text{ és } t_{p_3} = 2{,}756.$$

Így a (12) formula szerint a 90%-os szint esetén

$$\overline{X} - t_{p_1} \frac{s^*}{\sqrt{n}} = 708 - 1{,}669 \cdot \frac{52}{\sqrt{30}} = 692{,}15\,\text{kg};$$

$$\overline{X} + t_{p_1} \frac{s^*}{\sqrt{n}} = 708 + 1{,}669 \cdot \frac{52}{\sqrt{30}} = 723{,}84\,\text{kg};$$

tehát a 90%-os szinthez tartozó konfidencia-intervallum: $[692{,}15;723{,}84]$.

Hasonlóan kapjuk a 95 és 99%-nak megfelelő intervallumokat:

$$[688{,}6;722{,}4] \text{ és } [681{,}9;724{,}1].$$

S.2.4. A szabadságfokok száma: 99. A 90%-nak megfelelő konfidencia-határok (a 4. táblázatból interpolált $t_{p_1} = 1{,}662$; $t_{p_2} = ,986$, $t_{p_3} = 2{,}632$ értékekkel):

$$\overline{X} - t_{p_1} \frac{s^*}{\sqrt{n}} = 705 - 1{,}662 \cdot \frac{51}{\sqrt{100}} = 696{,}5\,\text{kg};$$

$$\overline{X} + t_{p_1} \frac{s^*}{\sqrt{n}} = 705 + 1{,}662 \cdot \frac{51}{\sqrt{100}} = 713{,}5\,\text{kg}.$$

Hasonlóan kapjuk a 95 és 99%-nak megfelelő intervallumokat:

$$[694{,}9;715{,}1] \text{ és } [691{,}9;718{,}4].$$

S.2.5. $u_p = 1{,}960$; a konfidencia-intervallum: $[3{,}08;3{,}42]$

S.2.6. A 90%-os szinthez $p = 0{,}1$ érték tartozik. A szabadságfokok száma: 15. A 3. sz. táblázat alapján $\chi^2_{p/2} = \chi^2_{0{,}05} = 24{,}996$ és $\chi^2_{1-p/2} = \chi^2_{0{,}95} = 7{,}261$.

A (16) szerint számított konfidencia-intervallum: $[12{,}0;22{,}3]$.

S.3. A 3. fejezet feladatainak megoldásai

S.3.1. $m_0 = 1200$ g; $\sigma = 3$ g; $\overline{X} = 1197$ g, $\sqrt{n} = \sqrt{100} = 10$, így

$$u = \sqrt{n}\,\frac{\overline{X} - m_0}{\sigma} = -4{,}0.$$

$|u| = 4 > 3{,}3$ a $p = 0{,}001$ értékhez tartozó értéknél, az előírt tömegtől való eltérés szignifikáns, a gépet be kell állítani.

S.3.2. Mivel $m_0 = 700$ kg; $\sigma = 50$ kg; $\overline{X} = 705$ kg és $\sqrt{n} = \sqrt{100} = 10$, így

$$u = \sqrt{n}\,\frac{\overline{X} - m_0}{\sigma} = 1{,}0.$$

A $p = 0{,}05$ értékhez $u_p = 1{,}960$, tehát $u = 1 < 1{,}96$, ezért 95%-os szinten nincs szignifikáns eltérés a 700 kg-hoz képest.

S.3.3. Az $\overline{X} = 494$; $s^{*2} = 64{,}9$ kiszámítása után meghatározzuk a $t_{sz} = -2{,}36$ értéket. Mivel $|t_{sz}| = 2{,}36 > t_{95} = 2{,}262$, ezért az automata nagy valószínűséggel nem működik jól, be kell állítani.

S.3.4. A $\chi^2 = \sum_{i=1}^{r} \frac{(f_i - np_i)^2}{np_i} = n\sum_{i=1}^{r} \frac{(g_i - p_i)^2}{p_i}$ képletet használva, a számítást a következő táblázat szemlélteti

x_i	$u = \frac{x_i - m}{\sigma}$	$\Phi(u)$	p_i	$g_i = \frac{f_i}{n}$	$\frac{(g_i - p_i)^2}{p_i}$
243	−1,340	0,0901	0,0901	0,0634	0,00791
323	−0,916	0,1798	0,0897	0,1268	0,01534
403	−0,493	0,3110	0,1312	0,1620	0,00723
483	−0,070	0,4721	0,1611	0,1901	0,00522
563	0,353	0,6379	0,1658	0.1268	0,00917
643	0,776	0,7811	0,1432	0.1056	0,00987
723	1,199	0,8847	0,1036	0,0845	0,00352
∞	∞	1,0000	0,1153	0,1408	0,00565
Σ			1,0000		0,06391

Tehát $\chi^2 = 142 \cdot 0{,}06391 = 9{,}075$.

A szabadságfokok száma: $8 - 3 = 5$.

A 3. sz. táblázatban a 9,07 értékhez 9,24 esik legközelebb, amelynek $p = 0{,}10$ érték felel meg, tehát kb. 10%-os szinten állítható, hogy a minta normális eloszlású sokaságból származott, ezért hipotézisünket elvetjük.

S.4. A 4. fejezet feladatainak megoldásai

S.4.1. A korrelációs együttható: $r = 0{,}9815$.

A regressziós egyenes egyenlete: $y = 7{,}47x - 148{,}51$.

S.4.2. A korrelációs együttható: $r = 0{,}997$.

A regressziós egyenes egyenlete: $y = 11x - 21657{,}6$.

IRODALOMJEGYZÉK

[1] Bognár–Mogyoródi–Prékopa–Rényi–Szász: *Valószínűségszámítás*. Feladatgyűjtemény. Tankönyvkiadó, Budapest, 1982.

[2] Denkinger Géza: *Valószínűségszámítás*. Tankönyvkiadó, Budapest, 1978.

[3] Denkinger Géza: *Valószínűségszámítási gyakorlatok*. Tankönyvkiadó, Budapest, 1977.

[4] W. Feller: *Bevezetés a valószínűségszámításba és alkalmazásaiba*. Műszaki Könyvkiadó, 1978.

[5] Lukács O.: *Matematikai statisztika*. Példatár. Műszaki Könyvkiadó, Budapest, 1987.

[6] Meszéna Gy.–Ziermann M.: *Valószínűségelmélet és matematikai statisztika*. Közgazdasági és Jogi Könyvkiadó, Budapest, 1981.

[7] Mundruczó Gy.: *Alkalmazott regressziószámítás*. Akadémia Kiadó, Budapest, 1981.

[8] dr. Medgyessy P. – dr. Takács L.: *Műszaki Matematikai Gyakorlatok C. V. Valószínűségszámítás*. (Szerk.: dr Fazekas F.). Tankönyvkiadó, Budapest, 1966.

[9] Obádovics J. Gy.: *Matematika*. 16. kiadás. SCOLAR Kiadó, Budapest, 2000.

[10] Obádovics J. Gy.: *Gyakorlati számítási eljárások*. Gondolat Kiadó, Budapest, 1972.

[11] Prékopa A.: *Valószínűségelmélet műszaki alkalmazásokkal*. Műszaki Könyvkiadó, Budapest, 1961.

[12] dr. Reimann J. – Tóth Julianna: *Valószínűségszámítás és matematikai statisztika*. Nemzeti Tankönyvkiadó, Budapest, 1985.

[13] Rényi A.: *Valószínűségszámítás*. 2. kiadás. Tankönyvkiadó, Budapest, 1966.

[14] Sachs, L.: *Angewandte Statistik*. 4. kiadás. Springer Verlag, Berlin–Heidelberg–New York, 1974.

[15] Solt Gy.: *Valószínűségszámítás*. Példatár. 6. kiadás. Műszaki Könyvkiadó, Budapest, 1993.

[16] *Statisztikai módszerek alkalmazása a mezőgazd.-ban*. (Szerk.: dr. Manczel J.) Mezőgazdasági Kiadó, Budapest, 1983.

[17] Székely J. G.: *Paradoxonok a véletlen matematikájában*. Műszaki Könyvkiadó, Budapest, 1982.

[18] Vincze I.: *Matematikai statisztika ipari alkalmazásokkal*. Műszaki Könyvkiadó, Budapest, 1968.

1. sz. táblázat

t	0	1	2	3	4	5	6	7	8	9
0,0	,5000	,5040	,5080	,5120	,5160	,5199	,5239	,5279	,5319	,5359
0,1	,5398	,5438	,5478	,5517	,5557	,5596	,5636	,5675	,5714	,5754
0,2	,5793	,5832	,5871	,5910	,5948	,5987	,6026	,6064	,6103	,6141
0,3	,6179	,6217	,6255	,6293	,6331	,6368	,6406	,6443	,6480	,6517
0,4	,6554	,6591	,6628	,6664	,6700	,6736	,6772	,6808	,6844	,6879
0,5	,6915	,6950	,6985	,7019	,7054	,7088	,7123	,7157	,7190	,7224
0,6	,7258	,7291	,7324	,7357	,7389	,7422	,7454	,7486	,7518	,7549
0,7	,7580	,7612	,7642	,7673	,7704	,7734	,7764	,7794	,7823	,7852
0,8	,7881	,7910	,7939	,7967	,7996	,8023	,8051	,8078	,8106	,8133
0,9	,8159	,8186	,8212	,8238	,8264	,8289	,8315	,8340	,8365	,8389
1,0	,8413	,8438	,8461	,8485	,8508	,8531	,8554	,8577	,8599	,8621
1,1	,8643	,8665	,8686	,8708	,8729	,8749	,8770	,8790	,8810	,8830
1,2	,8849	,8869	,8888	,8907	,8925	,8944	,8962	,8980	,8997	,9015
1,3	,9032	,9049	,9066	,9082	,9099	,9115	,9131	,9147	,9162	,9177
1,4	,9192	,9207	,9222	,9236	,9251	,9265	,9279	,9292	,9306	,9319
1,5	,9332	,9345	,9357	,9370	,9382	,9394	,9406	,9418	,9429	,9441
1,6	,9452	,9463	,9474	,9484	,9495	,9505	,9515	,9525	,9535	,9545
1,7	,9554	,9564	,9573	,9582	,9591	,9599	,9608	,9616	,9625	,9633
1,8	,9641	,9649	,9656	,9664	,9671	,9678	,9686	,9693	,9699	,9706
1,9	,9713	,9719	,9726	,9732	,9738	,9744	,9750	,9756	,9761	,9767
2,0	,9772	,9778	,9783	,9788	,9793	,9798	,9803	,9808	,9812	,9817
2,1	,9821	,9826	,9830	,9834	,9838	,9872	,9846	,9850	,9854	,9857
2,2	,9861	,9864	,9868	,9871	,9875	,9878	,9881	,9884	,9887	,9890
2,3	,9893	,9896	,9898	,9901	,9904	,9906	,9909	,9911	,9913	,9916
2,4	,9918	,9920	,9922	,9925	,9927	,9929	,9931	,9932	,9934	,9936
2,5	,9938	,9940	,9941	,9943	,9945	,9946	,9948	,9949	,9951	,9952
2,6	,9953	,9955	,9956	,9957	,9959	,9960	,9961	,9962	,9963	,9964
2,7	,9965	,9966	,9967	,9968	,9969	,9970	,9971	,9972	,9973	,9974
2,8	,9974	,9975	,9976	,9977	,9977	,9978	,9979	,9979	,9980	,9981
2,9	,9981	,9982	,9982	,9983	,9984	,9984	,9985	,9985	,9986	,9986
3,0	,9987	,9987	,9987	,9988	,9988	,9989	,9989	,9989	,9990	,9990
3,1	,9990	,9991	,9991	,9991	,9992	,9992	,9992	,9992	,9993	,9993
3,2	,9993	,9993	,9994	,9994	,9994	,9994	,9994	,9995	,9995	,9995
3,3	,9995	,9995	,9995	,9996	,9996	,9996	,9996	,9996	,9996	,9997
3,4	,9997	,9997	,9997	,9997	,9997	,9997	,9997	,9997	,9997	,9998
3,5	,9998	,9998	,9998	,9998	,9998	,9998	,9998	,9998	,9998	,9998
3,6	,9998	,9998	,9999	,9999	,9999	,9999	,9999	,9999	,9999	,9999
3,7	,9999	,9999	,9999	,9999	,9999	,9999	,9999	,9999	,9999	,9999
3,8	,9999	,9999	,9999	,9999	,9999	,9999	,9999	,9999	,9999	,9999
3,9	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0

Megjegyzés: $\Phi(-t) = 1 - \Phi(t)$.

2. sz. táblázat

t	0	1	2	3	4	5	6	7	8	9
0,0	,3989	,3989	,3989	,3988	,3986	,3984	,3982	,3980	,3977	,3973
0,1	,3970	,3965	,3961	,3956	,3951	,3945	,3939	,3932	,3925	,3918
0,2	,3910	,3902	,3894	,3885	,3876	,3867	,3857	,3847	,3836	,3825
0,3	,3814	,3802	,3790	,3778	,3765	,3752	,3739	,3725	,3712	,3697
0,4	,3683	,3668	,3653	,3637	,3621	,3605	,3589	,3572	,3555	,3538
0,5	,3521	,3503	,3485	,3467	,3448	,3429	,3410	,3391	,3372	,3352
0,6	,3332	,3312	,3292	,3271	,3251	,3230	,3209	,3187	,3166	,3144
0,7	,3123	,3101	,3079	,3056	,3034	,3011	,2989	,2966	,2943	,2920
0,8	,2897	,2874	,2850	,2827	,2803	,2780	,2756	,2732	,2709	,2685
0,9	,2661	,2637	,2613	,2589	,2565	,2541	,2516	,2492	,2468	,2444
1,0	,2420	,2396	,2371	,2347	,2323	,2299	,2275	,2251	,2227	,2203
1,1	,2179	,2155	,2131	,2107	,2083	,2059	,2036	,2012	,1989	,1965
1,2	,1942	,1919	,1895	,1872	,1849	,1826	,1804	,1781	,1758	,1736
1,3	,1714	,1691	,1669	,1647	,1626	,1604	,1582	,1561	,1539	,1518
1,4	,1497	,1476	,1456	,1435	,1415	,1394	,1374	,1354	,1334	,1315
1,5	,1295	,1276	,1257	,1238	,1219	,1200	,1182	,1163	,1145	,1127
1,6	,1109	,1092	,1074	,1057	,1040	,1023	,1006	,0989	,0973	,0957
1,7	,0940	,0925	,0909	,0893	,0878	,0863	,0848	,0833	,0818	,0804
1,8	,0790	,0775	,0761	,0748	,0734	,0721	,0707	,0694	,0681	,0669
1,9	,0656	,0644	,0632	,0620	,0608	,0596	,0584	,0573	,0562	,0551
2,0	,0540	,0529	,0519	,0508	,0498	,0488	,0478	,0468	,0459	,0449
2,1	,0440	,0431	,0422	,0413	,0404	,0396	,0387	,0379	,0371	,0363
2,2	,0355	,0347	,0339	,0332	,0325	,0317	,0310	,0303	,0297	,0290
2,3	,0283	,0277	,0270	,0264	,0258	,0252	,0246	,0241	,0235	,0229
2,4	,0224	,0219	,0213	,0208	,0203	,0198	,0194	,0189	,0184	,0180
2,5	,0175	,0171	,0167	,0163	,0158	,0154	,0151	,0147	,0143	,0139
2,6	,0136	,0132	,0129	,0126	,0122	,0119	,0116	,0113	,0110	,0107
2,7	,0104	,0101	,0099	,0096	,0093	,0091	,0088	,0086	,0084	,0081
2,8	,0079	,0077	,0075	,0073	,0071	,0069	,0067	,0065	,0063	,0061
2,9	,0060	,0058	,0056	,0055	,0053	,0051	,0050	,0048	,0047	,0046
3,0	,0044	,0048	,0042	,0040	,0039	,0038	,0037	,0036	,0035	,0034
3,1	,0033	,0032	,0031	,0030	,0029	,0028	,0027	,0026	,0025	,0025
3,2	,0024	,0023	,0022	,0022	,0021	,0020	,0020	,0019	,0018	,0018
3,3	,0017	,0017	,0016	,0016	,0015	,0015	,0014	,0014	,0013	,0013
3,4	,0012	,0012	,0012	,0011	,0011	,0010	,0010	,0010	,0009	,0009
3,5	,0009	,0008	,0008	,0008	,0008	,0007	,0007	,0007	,0007	,0006
3,6	,0006	,0006	,0006	,0005	,0005	,0005	,0005	,0005	,0005	,0004
3,7	,0004	,0004	,0004	,0004	,0004	,0004	,0003	,0003	,0003	,0003
3,8	,0003	,0003	,0003	,0003	,0003	,0002	,0002	,0002	,0002	,0002
3,9	,0002	,0002	,0002	,0002	,0002	,0002	,0002	,0002	,0001	,0001

Megjegyzés: $\varphi(-t)=\varphi(t)$.

3. sz. táblázat

	p =						
Szabadság-fok	0,10	0,05	0,01	0,001	0,025	0,975	0,95
1	2,706	3,841	6,635	10,827	5,024	9,82E-4	3,93E-3
2	4,605	5,991	9,210	13,815	7,378	5,06E-2	0,103
3	6,251	7,815	11,435	16,268	9,348	0,0216	0,352
4	7,779	9,488	13,277	18,465	11,143	0,484	0,711
5	9,236	11,070	15,086	20,517	12,833	0,831	1,145
6	10,645	12,592	16,812	22,457	14,449	1,237	1,635
7	12,017	14,067	18,475	24,322	16,013	1,690	2,167
8	13,362	15,507	20,090	26,125	17,535	2,180	2,733
9	14,684	16,919	21,666	27,877	19,023	2,700	3,325
10	15,987	18,307	23,209	29,588	20,483	3,247	3,940
11	17,275	19,675	24,725	31,264	21,920	3,816	4,575
12	18,549	21,026	26,217	32,909	23,337	4,404	5,226
13	19,812	22,362	27,688	34,528	24,736	5,009	5,892
14	21,064	23,685	29,141	36,123	26,119	5,629	6,571
15	22,307	24,996	30,578	37,697	27,488	6,262	7,261
16	23,542	26,296	32,000	39,252	28,845	6,908	7,962
17	24,769	27,587	33,409	40,790	30,191	7,564	8,672
18	25,989	28,869	34,805	42,312	31,526	8,231	9,390
19	27,204	30,144	36,191	43,820	32,852	8,907	10,117
20	28,412	31,410	37,566	45,315	34,170	9,591	10,851
21	29,615	32,671	38,932	46,797	35,479	10,283	11,591
22	30,813	33,924	40,289	48,268	36,781	10,982	12,338
23	32,007	35,172	41,638	49,728	38,076	11,689	13,091
24	33,196	36,415	42,980	51,179	39,364	12,401	13,484
25	34,382	37,652	44,314	52,620	40,646	13,120	14,611
26	35,563	38,885	45,642	54,052	41,923	13,844	15,379
27	36,741	40,113	46,963	55,476	43,194	14,573	16,151
28	37,916	41,337	48,278	56,893	44,461	15,308	16,928
29	39,087	42,557	49,558	58,302	45,772	16,047	17,708
30	40,256	43,773	50,892	59,703	46,979	16,791	18,493

4. sz. táblázat

Szabadság-fok	$p =$ 0,20	0,10	0,05	0,02	0,01	0,001
1	3,078	6,314	12,706	31,821	63,657	636.619
2	1,886	2,920	4,303	6,965	9,925	31,598
3	1,638	2,353	3,182	4,541	5,841	12,941
4	1,530	2,132	2,776	3,747	4,604	8,610
5	1,476	2,015	2,571	3,365	4,032	6,859
6	1,440	1,943	2,447	3,143	3,707	5,959
7	1,415	1,895	2,365	2,998	3,499	5,405
8	1,397	1,860	2,306	2,896	3,355	5,041
9	1,383	1,833	2,262	2,821	3,250	4,781
10	1,372	1,812	2,228	2,764	3,169	4,587
11	1,363	1,796	2,201	2,718	3,106	4,437
12	1,356	1,782	2,179	2,681	3,055	4,318
13	1,350	1,771	2,160	2,650	3,012	4,221
14	1,345	1,761	2,145	2,624	2,977	4,140
15	1,341	1,753	2,131	2,602	2,947	4,073
16	1,337	1,746	2,120	2,583	2,921	4,015
17	1,333	1,740	2,110	2,567	2,898	3,965
18	1,330	1,734	2,101	2,552	2,878	3,922
19	1,328	1,729	2,093	2,539	2,861	3,883
20	1,325	1,725	2,086	2,528	2,845	3,850
21	1,323	1,721	2,080	2,518	2,831	3,819
22	1,321	1,717	2,074	2,508	2,819	3,792
23	1,319	1,714	2,069	2,500	2,807	3,767
24	1,318	1,711	2,064	2,492	2,797	3,745
25	1,316	1,708	2,060	2,485	2,787	3,725
26	1,315	1,706	2,056	2,479	2,779	3,707
27	1,314	1,703	2,052	2,473	2,771	3,690
28	1,313	1,701	2,048	2,467	2,763	3,647
29	1,311	1,699	2,045	2,462	2,756	3,659
30	1,310	1,697	2,042	2,457	2,750	3,646
40	1,303	1,684	2,021	2,423	2,704	3,551
60	1,296	1,671	2,000	2,390	2,660	3,460
120	1,289	1,658	1,980	2,358	2,617	3,373
∞	1,282	1,645	1,960	2,326	2,576	3,291

5. sz. táblázat

$f_2 \setminus f_1$	1	2	3	4	5	6	7	8	9	12
1	161,4	199,5	215,7	224,6	230,0	234,0	236,8	238,9	240,5	243,9
2	18,51	19,0	19,16	19,25	19,3	19,33	19,35	19,37	19,38	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,84	8,81	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,83	2,77	2,71	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,50	2,54	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,38
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,34
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	242	2,31
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,28
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,25
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,23
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,38	2,32	2,20
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,18
25	4,24	3,38	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,16
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,15
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,30	2,25	2,13
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,12
29	4,18	3,33	2,93	2,70	2,54	2,43	2,35	2,28	2,22	2,10
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,09
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,00
60	4,00	3,15	2,76	2,52	2,37	2,25	2,17	2,10	2,04	1,92
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,83
∞	3,84	2,99	2,60	2,37	2,21	2,09	2,01	1,94	1,88	1,75

f_1 a nagyobb empirikus szórásnégyzetű minta elemszámának, f_2 pedig a másik minta elemszámának eggyel kisebbített értéke.

6. sz. táblázat

x	e^x	e^{-x}	x	e^x	e^{-x}	x	e^x	e^{-x}
0,00	1,0000	1,0000	0,40	1,4918	0,6703	0.80	2,2255	0,4493
.01	1,0101	0.9900	,41	1,5068	,6637	.81	2,2479	,4449
,02	1,0202	,9802	,42	1,5220	,6570	,82	2,2705	,4404
,03	1,0305	,9704	,43	1,5373	,6505	.83	2,2933	,4360
,04	1,0408	,9608	,44	1,5527	,6440	.84	2,3164	,4317
,05	1,0513	,9512	,45	1,5683	,6376	,85	2,3397	,4274
,06	1,0618	,9418	,46	1,5841	,6313	,86	2,3632	,4232
,07	1,0725	,9324	,47	1,6000	,6250	,87	2,3869	,4190
,08	1,0833	,9231	,48	1,6161	,6188	,88	2,4109	,4148
,09	1,0942	.9139	,49	1,6323	,6126	,89	2.4351	,4107
,10	1,1052	,9048	,50	1,6487	,6065	,90	2,4596	,4066
,11	1,1163	,8958	,51	1,6653	,6005	,91	2,4843	,4025
,12	1,1275	,8869	,52	1,6820	,5945	,92	2,5093	,3985
,13	1,1388	,8781	,53	1,6989	,5886	,93	2,5345	,3946
,14	1,1503	,8694	,54	1,7160	,5827	,94	2,5600	,3906
,15	1,1618	,8607	,55	1,7333	,5769	,95	2,5857	,3867
,16	1,1735	,8521	,56	1,7507	,5712	,96	2,6117	,3829
,17	1,1853	,8437	,57	1,7683	,5655	,97	2,6380	,3791
,18	1,1972	,8353	,58	1,7860	,5599	,98	2,6645	,3753
,19	1,2096	,8270	,59	1,8040	,5543	,99	2,6912	,3716
,20	1,2214	,8187	,60	1,8221	,5488	1,0	2,7183	,3679
,21	1,2337	,8106	,61	1,8404	,5434	1,1	3,004	,3329
,22	1,2461	,8025	,62	1,8589	,5379	1,2	3,320	,3012
,23	1,2586	,7945	,63	1,8776	,5326	1,3	3,669	,2725
,24	1,2713	,7866	,64	1,8965	,5273	1,4	4,055	,2466
,25	1,2840	,7788	,65	1,9155	,5220	1,5	4,482	,2231
,26	1,2969	,7711	,66	1,9348	,5169	1,6	4,953	,2019
,27	1,3100	,7634	,67	1,9542	,5117	1,7	5,474	,1827
,28	1,3231	,7558	,68	1,9739	,5066	1,8	6,050	,1653
,29	1,3364	,7483	,69	1,9937	,5016	1,9	6,686	,1496
,30	1,3497	,7408	,70	2,0138	,4966	2,0	7,389	,1353
,31	1,3634	,7334	,71	2,0340	,4916	2,1	8,166	,1225
,32	1,3771	,7261	,72	2,0544	,4868	2,2	9,025	,1108
,33	1,3910	,7189	,73	2,0751	,4819	2,3	9,974	,1003
,34	1,4050	,7118	,74	2,0960	,4771	2,4	11,023	,0907
,35	1,4191	,7047	,75	2,1170	,4724	2,5	12,182	,0821
,36	1,4333	,6977	,76	2,1383	,4677	2,6	13,464	,0743
,37	1,4477	,6907	,77	2,1598	,4630	2,7	14,880	,0672
,38	1,4623	,6839	,78	2,1815	,4584	2,8	16,445	,0608
,39	1,4770	,6771	,79	2,2034	,4538	2,9	18,174	,0550
,40	1,4918	,6703	,80	2,2255	,4493	3,0	20,086	,0498

NÉV- ÉS TÁRGYMUTATÓ

A

Á

B

C

D

E