Petz Dénes és Lángné Lázi Márta

MATEMATIKA III.

Oktatási segédanyag vegyészmérnök hallgatók számára
a valószínűségelmélet és a matematikai statisztika alapjairól

BME Matematika Intézet

2000. december

Tartalomjegyzék

Előszó

A valószínűségelmélet és a matematikai statisztika viszonylag fiatal területei a matematikának. Napjainkra a valószínűség fogalma a hétköznapi életbe is bevonult, sztochasztikus modellek használata nemcsak a gázok statisztikus leírásában és a kvantumelméletben gyakori, hanem a közgazdaságtantól a piackutatásig számos területen. Statisztikai adatokkal lépten nyomon találkozunk, és a kisérletek eredményének kiértékelése, esetenként pedig a kisérletek megtervezése is, statisztikai gondolkodást kíván.

A számítástechnika fejlődése lehetővé tette igen nagyszámú adat gyors és olcsó kezelését. Így tér nyílt hatalmas adatsorok összefüggéseinek elemzésére. Természetes, hogy az 1960-as évektől kezdve a sztochasztika beépült a műszaki egyetemek tananyagába.

A matematikai értelemben vett statisztika a valószínűségelméleten alapszik, ez a tantárgy ebbe a témába ad bevezetést. Célkitűzése sokkal inkább az alapvető fogalmak tisztázása, mint számolási jártasság kialakítása. (Ma az adatokkal való legfontosabb műveleteket programcsomagok végzik el, némely mérőműszer az ismételt mérések szórását is megadja.) Nem szabad figyelmen kívül hagyni, hogy fogalmak megértése munka- és időigényes folyamat, gyakran keservesebb, mint számolási algoritmusok elsajátítása.

A jegyzet viszonylagos rövidsége ellenére nagy anyagot ölel fel. Az előadás követésének megkönnyítésére íródott, és önálló tanulásra kevésbé alkalmas. Köszönet illeti azokat a hallgatókat, akik megjegyzései elősegítették a korábbi változat esetenkénti hiányosságainak kiküszöbölését.

Az öt fejezet szorosan egymásra épülő témaköröket tárgyal. A véletlen esemény, a valószínűség, a feltételes valószínűség és a függetlenség a sztochasztika sarkalatos fogalmai, az első fejezet tárgyalja azokat. Megvilágításuk szemléletes az érmedobás és a kockadobás mindenki által ismert hétköznapi tapasztalatainak segítségével. A második témakör a valószínűségi változók, amelyek véletlen, azaz sztochasztikus ingadozást mutató mennyiségek matematikai modellezésére szolgálnak. Velük kapcsolatban mindig lényeges kérdés, hogy mi körül és milyen mértékben ingadoznak. A várható érték és a szórás fejezi ki ezeket a jellemzőket. A harmadik fejezet a nevezetes valószínűségeloszlások közül csak keveset tartalmaz, az érdeklődő és a felhasználó sokkal több konkrét eloszlást talál az irodalomban. A negyedik fejezet foglalkozik több véletlen mennyiség belső kapcsolatával, és ízelítőt ad az időben változó valószínűségi változókról, ami a sztochasztikus folyamat nevet viseli. Fel kell hívni a figyelmet arra, hogy ebben a részben jóval több matematikai előismeret használata szükséges, így például kettős integrálok kezelése esetenként. Az ötödik fejezet egy kis bepillantás a szoros értelemben vett matematikai statisztikába. A paraméterbecslés és a statisztikai próbák vannak röviden felvázolva. Az utolsó két fejezetben jelenik meg igazán a sztochasztikus szemléletmód.

A jegyzet minden egyes fejezete tartalmaz megoldott és megoldásra javasolt feladatokat. Előbbiek a fejezetben példa név alatt találhatók, utóbbiak pedig a fejezet végi gyakorlatok.

Ezek némelyikének a megoldását a 6. fejezetben írtuk le. A megoldott gyakorlatokat ,,MM" betű jelöli. A feladatmegoldás szorosan hozzátartozik az anyaghoz.

A csillaggal jelölt feladatok jóval nehezebbek, inkább érdekességként szerepeltetjük őket. Vannak a jegyzetben olyan részek, amelyek nem tartoznak a tárgy törzsanyagához. Ezek apróbb betűkkel jelennek meg.

Budapest, 2000. december 1.

A szerzők

Első fejezet

Elemi valószínűségelmélet

A valószínűségelmélet alapvető fogalma a **véletlen esemény**. A **véletlen kísérlet** végrehajtásakor egy véletlen esemény vagy bekövetkezik, vagy nem következik be. Egyszerű példa véletlen kísérletre az érmedobás, amelynek két kimenetele a FEJ és az ÍRÁS. Valószínűségszámítási szempontból véletlen kísérletnek tekinthető egy mérés végrehajtása is. Ekkor esemény lehet az, hogy az eredmény, amit most valós számnak feltételezünk, egy adott intervallumba esik. Az érmedobással ellentétben ennek a véletlen kísérletnek elvileg végtelen sok kimenetele lehet. A véletlen kísérlet kimeneteleit **elemi eseményeknek** nevezzük. Az összes elemi események halmaza az **eseménytér**, amit a valószínűségelméletben rendszerint Ω jelöl. Az érmedobás esetében az eseménytér két elemű.

A. N. Kolmogorov orosz matematikus 1933-ban publikálta Berlinben a „Grundbegriffe der Wahrscheinlichkeitsrechnung" című könyvét, és ezt az időpontot tekintik sokan a modern valószínűségelmélet kezdetének. Valószínűségszámítás egyszerű formában ennnél sokkal régebben is létezett, többnyire a szerencsejátékokhoz kapcsolódott. Maxwell és Boltzmann a kinetikus gázelméletben és a termodinamikában már a múlt század második felében sztochasztikus meggondolásokat használtak.

A **Kolmogorov-féle** valószínűségelméletben az eseményeket az eseménytér részhalmazaival azonosítjuk. Az A esemény azokból az elemi eseményekből áll, amelyekre az teljesül, hogy a kísérlet ilyen kimenetele mellett az A esemény bekövetkezik. Például, ha arról van szó, hogy egy dobókockával kétszer egymás után dobunk, akkor az az esemény, hogy a két dobás összege legalább 10, a következő részhalmaza az eseménytérnek:

$$\{(6,6),(6,5),(5,6),(5,5)\}$$

Maga a teljes eseménytér egy 36 elemű halmaz.

Az események között van két kitüntetett: a **lehetetlen** és a **biztos** esemény. A lehetetlen esemény sohasem következik be, így nincs olyan elemi esemény, ami megvalósítja, Ω üres részhalmazának felel meg. A másik véglet a biztos esemény, amit minden elemi esemény megvalósít, tehát magának az Ω halmaznak felel meg.

1.1. Műveletek eseményekkel

Legyen A és B két esemény. Az $A \cdot B$ esemény, amelyet A és B **szorzatának** nevezünk, akkor következik be, ha mind A, mind B bekövetkezik. Halmazelméleti nyelven az $A \cdot B$ részhalmaza

Ω-nak az A és B halmazok közös része. A-t és B-t **egymást kizáróknak**, vagy **diszjunktnak** mondjuk, ha szorzatuk a lehetelen esemény, azaz egyszerre nem következhetnek be.

Az $A+B$ esemény akkor következik be, ha A és B közül legalább az egyik bekövetkezik.

Az A esemény **komplementere**, más szóval **kiegészítője**, az az esemény, amely akkor valósul meg, ha A nem következik be. Jelölése: $\bar{A}$, vagy A^c. A komplementer esemény az $\Omega \setminus A$ halmazelméleti különbség. A és $\bar{A}$ mindig egymást kizáró események.

1.1. példa: Ha a véletlen kísérletnek a kétszeri kockadobást tekintjük, akkor az elemi események olyan (i, j) számpárokkal adhatók meg, amelyekre $1 \leq i, j \leq 6$. Az Ω eseménytér az összes lehetséges számpárok 36 elemű halmaza. Ha A jelenti azt az eseményt, hogy a két dobás összege legalább 10, akkor $A = \{(6,6),(6,5),(5,6),(5,5)\}$. Ha B jelenti azt az eseményt, hogy az első dobás páros, akkor $A \cdot B = \{(6,6),(6,5)\}$, $A \cdot \bar{B} = \{(5,6),(5,5)\}$ és

$$\begin{aligned} A+B &= \{(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),(4,1),(4,2),(4,3),(4,4), \\ &\quad (4,5),(4,6),(5,5),(5,6),(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}. \end{aligned}$$ □

1.2. Relatív gyakoriság és valószínűség

Legyen a véletlen kísérletünk a kockadobás, és jelentse A azt az eseményt, hogy a dobás eredménye páros. A kockadobás 20-szori megismétlésével a következő sorozatot kaphatjuk:

$$6,3,2,5,6,6,1,3,3,6,6,2,6,4,5,2,6,5,3,1$$

A 20 dobásból az A esemény 11-szer valósult meg. Azt mondjuk, hogy 11 az A esemény gyakorisága és $11/20 = 0.55$ a **relatív gyakorisága**. Ha a kísérletek számát (azaz most a dobások számát) növeljük, akkor egy eseménynek a relatív gyakorisága stabilitást mutat, és egy bizonyos érték körül ingadozik. Ez az érték az esemény **valószínűsége**. Az A esemény valószínűségét $P(A)$-val jelöljük.

Legyen A tetszőleges esemény és $\bar{A}$ a kiegészítője. Ismételjük meg gondolatban a véletlen kísérletet n-szer. Mondjuk A esemény n_A-szor következett be. Így $\bar{A}$-nak $n - n_A$-szor kellett bekövetkeznie. A valószínűség fenti meghatározása alapján A relatív gyakorisága $= n_A/n$ tart $P(A)$-hoz, és $\bar{A}$ relatív gyakorisága $= (n - n_A)/n$ tart $P(\bar{A})$-hez. Mivel

$$\frac{n_A}{n} + \frac{n - n_A}{n} = 1,$$

eljutottunk a

$$P(A) + P(\bar{A}) = 1$$

összefüggéshez. Ugyanez a gondolatmenet eredményezi a

$$\boxed{P(A+B) = P(A) + P(B), \qquad \text{ahol} \qquad A \cdot B = \emptyset} \tag{1.2.1}$$

összefüggést, ha A és B egymást kizáró események. Ez alapvető tulajdonsága a valószínűségnek, és **additivitásnak** nevezzük. További tulajdonságok:

$$\boxed{P(\emptyset) = 0, \quad 0 \leq P(A) \leq 1, \quad P(\Omega) = 1} \tag{1.2.2}$$

Szavakkal: A lehetetlen esemény valószínűsége 0, a biztos esemény valószínűsége 1, és a valószínűség mindig 0 és 1 közötti szám, (ami egyébként százalékban is kifejezhető).

A valószínűségelmélet Kolmogorov-féle felépítésében (1.2.1) és (1.2.2) **axiómaként** szerepelnek. Matematikai, pontosabban integrálelméleti okokból, az additivitást végtelen sok esemény esetére is meg kell követelni:

$$P\Big(\sum_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i) \qquad (\text{a}hol \quad A_i \cdot A_j = \emptyset,\ \text{ha}\, i \neq j) \tag{1.2.3}$$

Ha A és B tetszőleges,tehát nem szükségképpen egymást kizáróak, akkor (1.2.1) helyett az

$$P(A+B) = P(A) + P(B) - P(A \cdot B) \tag{1.2.4}$$

összefüggés érvényes. Ez az állítás az axiómákból levezethető.

1.2. példa: Igazoljuk, hogy tetszőleges A és B események összege, felírható egymást páronként kizáró események összegeként.

Jelentse I a biztos eseményt! Igazolható, ogy $A \cdot I = A$ és $I = A + \overline{A}$ vagy $I = B + \overline{B}$. Ezeket felhasználva: $A + B = A \cdot I + B \cdot I = A(B + \overline{B}) + B(A + \overline{A})$. Használjuk fel a szorzás és az összeadás tulajdonságait: $A + B = AB + A\overline{B} + BA + B\overline{A}$, azaz $A + B = AB + A\overline{B} + B\overline{A}$, ahol $A\overline{B} \cdot \overline{A}B = \emptyset$, $A\overline{B} \cdot AB = \emptyset$ és $\overline{A}B \cdot AB = \emptyset$. □

1.3. Függetlenség és feltételes valószínűség

Az A és B események **függetlenek**, ha A bekövetkezése nem befolyásolja B bekövetkezésének a valószínűségét. Például, ismételt kockadobás esetében az az esemény, hogy elsőre 3-at dobunk, független attól, hogy másodikra páratlant dobunk. Ugyanakkor, ha piros és fekete golyókat tartalmazó urnából visszatevés nélkül húzunk, akkor az, hogy elsőre pirosat húzunk nem független attól, hogy a másodikra pirosat húzunk. (Tudniillik, az elsőre való piros húzás csökkenti a másodikra való piros húzás esélyét.) Az A és B események függetlenségének matematikai definíciója :

$$\boxed{P(A \cdot B) = P(A)P(B)} \tag{1.3.1}$$

1.3. példa: Legyen A és B két független esemény, és $P(A) = 1/3$, $P(B) = 1/4$. Számítsuk ki a $P(A + B)$ valószínűséget!

$P(A+B) = P(A) + P(B) - P(A \cdot B) = P(A) + P(B) - P(A)P(B) = \frac{1}{2}$. □

Az A_1, $A_2, \ldots, A_n$ eseményeket (teljesen) **függetlennek** nevezzük, ha

$$\boxed{P(A_{i_1} \cdot A_{i_2} \cdot \ldots \cdot A_{i_k}) = P(A_{i_1})P(A_{i_2})\ldots P(A_{i_k})} \tag{1.3.2}$$

valahányszor $1 \leq i_1 < i_2 < \ldots < i_k \leq n$. (Konkrét példával megmutatható, hogy kettőnél több esemény esetén a páronkénti függetlenség nem vonja maga után a teljes függetlenséget.)

1.4. példa: Magyarországon a fiúk születési aránya 51%. Mi a valószínűsége, hogy egy 3 gyermekes magyar családban több a fiú, mint a lány?

Feltételezzük, hogy a gyermekszületések egymástól teljesen függetlenek, jelöljük A-val azt az eseményt, amelynek valószínűségét keressük, és legyen F_i, illetve L_i az az esemény, hogy az i-edik gyermek fiú, illetve lány. Ekkor

$$A = F_1F_2F_3 + L_1F_2F_3 + F_1L_2F_3 + F_1F_2L_3$$

és mivel ezek egymást kizáró események,

$$\begin{aligned} P(A) &= P(F_1F_2F_3) + P(L_1F_2F_3) + P(F_1L_2F_3) + P(F_1F_2L_3) \\ &= 0.51^3 + 3 \cdot 0.51^2 \cdot 0.49 = 0.515 \qquad \square \end{aligned}$$

Természetesen vannak olyan helyzetek, amelyekben egy esemény bekövetkezése igencsak befolyásolja egy másik esemény bekövetkezését. Legyen egy urnában 5 piros és 3 fekete golyó. Kihúzunk egy golyót, majd annak visszatevése nélkül még egyet. Legyen B az az esemény, hogy elsőre pirosat húzunk, A pedig az, hogy a másodikra pirosat húzunk. A feltételes valószínűség $P(A|B)$, annak a valószínűsége, hogy A bekövetkezik, feltételezve, hogy B bekövetkezett. A konkrét példánkban ezt $\frac{4}{7}$-nek gondoljuk. A feltételes valószínűség matematikai definíciója:

$$\boxed{P(A|B) = \frac{P(A \cdot B)}{P(B)}} \tag{1.3.3}$$

A $P(A|B)$ feltételes valószínűség csak akkor értelmes, ha B pozitív valószínűségű esemény.

A feltételes valószínűség segítségével az A és B események függetlensége

$$P(A|B) = P(A) \text{v}agy P(B|A) = P(B)$$

formában is kifejezhető. Szavakkal: A függetlenség azt jelenti, hogy az egyik esemény bekövetkezése semmilyen információt nem ad arról, hogy a másik esemény bekövetkezik-e.

1.5. példa: Tegyük fel, hogy egy urnában 12 piros és 10 fekete golyó van. Az urnából egymás után két golyót húzunk. Mi a valószínűsége, hogy az első piros és a második fekete?

Jelöljük A-val azt az eseményt, hogy az első húzás piros és B-vel azt, hogy a második fekete. Nyilván $P(A) = \frac{12}{22}$. Amit meg kell határoznunk az $P(A \cdot B)$. Feltételes valószínűséget használva: $P(A \cdot B) = P(A)P(B|A)$. Mivel $P(B|A) = \frac{10}{21}$, $P(A \cdot B) = \frac{12}{22} \cdot \frac{10}{21} = \frac{20}{77}$ adódik. $\square$

1.1. tétel: (A teljes valószínűség tétele) *Legyenek $A_1, A_2, \ldots, A_n$ olyan események, hogy*

$$\Omega = A_1 + A_2 + \ldots + A_n \qquad \textit{és} \qquad A_i \cdot A_j = \emptyset \quad \text{ha} \quad i \neq j\,.$$

Ha $P(A_i) > 0$, akkor bármely B eseményre

$$\boxed{P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots + P(B|A_n)P(A_n)\,.}$$

Bizonyítás: A feltevés alapján

$$B \cdot A_1\,, B \cdot A_2\,, \ldots\,, B \cdot A_n$$

egymást kizáró események és összegük B. Így a valószínűség additivitása, az (1.2.1) képlet általánosítása alapján

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

Ha itt $P(B \cdot A_i)$ helyébe $P(A|B_i)P(A_i)$-t írunk, akkor éppen a bizonyítandó tételt kapjuk. □

Az olyan A_1, $A_2, \ldots$, A_n események, amelyekre a teljes valószínűség tétele feltételei teljesülnek **teljes eseményrendszert** alkotnak.

1.6. példa: Tegyük fel, hogy egy urnában 8 piros és 10 fekete golyó van. Az urnából egymás után két golyót húzunk. Mi a valószínűsége, hogy a másodikra húzott golyó fekete?

Legyen A az az esemény, hogy az elsőre húzott golyó piros, B az az esemény, hogy a másodikra húzott golyó fekete. A és $\bar{A}$ teljes eseményrendszert alkotnak, és alkalmazhatjuk a teljes valószínűség tételét $P(B)$ kiszámítására:

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}) = \frac{10}{17}\frac{8}{18} + \frac{9}{17}\frac{10}{18} \approx 56\%\,. \qquad \Box$$

1.7. példa: Az I. urnában k golyó van, a II. urnában n golyó. Egymás után húzunk golyókat, p valószínűséggel az I. urnából, $q = 1 - p$ valószínűséggel a II. urnából. Mi a valószínűsége, hogy az I. urna ürül ki előbb?

Jelölje $A_{k,n}$ azt az eseményt, hogy a k golyót tartalmazó I. urnával és az n golyót tartalmazó II. urnával indulva az I. urna ürül ki előbb. (Tehát éppen $A_{k,n}$ valószínűsége a kérdéses.) Legyen A az az esemény, hogy először az I. urnából húzunk. A teljes valószínűség tétele szerint

$$P(A_{k,n}) = P(A_{k,n}|A) \cdot P(A) + P(A_{k,n}|\bar{A}) \cdot P(\bar{A}).$$

Itt $P(A) = p$ és $P(\bar{A}) = q$ adott értékek. Ugyanakkor $P(A_{k,n}|A)$ annak a valószínűsége, hogy az I. urnában k golyóval, a II. urnában n golyóval indulva, és először az I. urnából húzva az I. urna ürül ki előbb. Ez a valószínűség ugyanaz, mint annak az eseménynek a valószínűsége, hogy az I. urnában $k-1$ golyóval, a II. urnában n golyóval indulva az I. urna ürül ki előbb. A most megfogalmazott esemény éppen $A_{k-1,n}$. Hasonlóan látszik, hogy $P(A_{k,n}|\bar{A}) = P(A_{k,n-1})$. Tehát

$$P(A_{k,n}) = P(A_{k-1,n}) \cdot p + P(A_{k,n-1}) \cdot q \tag{1.3.4}$$

Értelemszerűen

$$P(A_{k,0}) = 0, \qquad P(A_{0,n}) = 1$$

és megállapodhatunk abban, hogy $P(A_{0,0}) = 0$ (bár az $A_{0,0}$ esemény elég értelmetlen). Így $P(A_{1,n})$-et (1.3.4)-ból kiszámíthatjuk.

$$P(A_{1,n}) = P(A_{0,n}) \cdot p + P(A_{1,n-1}) \cdot q$$

és $P(A_{1,1}) = 1 \cdot p + 0 \cdot q = p$, $P(A_{1,2}) = 1 \cdot p + p \cdot q = (1+q)p$, és így tovább. Tetszőleges k-ra és n-re a $p_{k,n} = P(A_{k,n})$ valószínűség az úgynevezett **generátorfüggvény-módszerrel** számolható ki. Adjuk össze az

$$x^n p_{k,n} = x^n p p_{k-1,n} + x^n q p_{k,n-1}$$

egyenleteket az $n = 1, 2 \ldots$ értékekre. Így azt kapjuk, hogy

$$\sum_{n=1}^{\infty} p_{k,n} x^n = p \sum_{n=1}^{\infty} p_{k-1,n} x^n + qx \sum_{n=1}^{\infty} p_{k,n-1} x^{n-1}.$$

Ha a baloldalt egy g_k függvény Taylor-sorának gondoljuk, akkor az egyenlet a

$$g_k(x) = p g_{k-1}(x) + qx g_k(x)$$

alakot ölti. Belőle

$$\begin{aligned} g_k(x) &= \frac{p}{1-qx} g_{k-1}(x) = \frac{p^2}{(1-qx)^2} g_{k-2}(x) = \frac{p^k}{(1-qx)^k} g_0(x) \\ &= \frac{p^k}{(1-qx)^k} \frac{x}{1-x}, \end{aligned}$$

hiszen

$$g_0(x) = \sum_{n=1}^{\infty} P_{0,n} x^n = \sum_{n=1}^{\infty} x^n = \frac{x}{1-x}.$$

Tehát a g_k függvényt sikerült meghatároznunk, és a $p_{k,n}$ valószínűség $g_k(x)$ Taylor-sorában x^n együtthatója, amita binomiális sorfejtés segítségével kiszámolva a

$$p_{k,n} = p^k \Big(1 - \binom{-k}{1} q + \binom{-k}{2} q^2 - \ldots + \binom{-k}{n-1} (-q)^{n-1}\Big)$$

eredmény adódik. □

1.4. Klasszikus valószínűségelmélet

Klasszikus valószínűségelméletről, vagy valószínűségszámításról akkor beszélünk, ha az Ω eseménytér véges sok elemből álló halmaz, és az Ω-t alkotó elemi események mind egyenlő valószínűek. Tipikus példa a szabályos kocka dobása, amikor mind a hat oldalt egyenlő valószínűnek gondoljuk, és természetesen feltételezzük, hogy a feldobott kocka nem gurul úgy el, hogy nem lehet leolvasni, továbbá nem esik élére, stb. Ha Ω elemszáma n, akkor minden egyes elemi esemény $1/n$ valószínűségű kell, hogy legyen. Ha egy A eseményt k elemi esemény valósít meg, akkor $P(A) = k/n$, amit gyakran úgy fogalmazunk, hogy

$$\boxed{P(A) = \frac{\text{kedvező esetek száma}}{\text{összes esetek száma}} = \frac{|A|}{|\Omega|}\,.} \tag{1.4.1}$$

(Itt $|A|$ jelöli A számosságát, vagyis elemei számát.)

Dobókockát ókori egyiptomi sírokban is találtak, és talán a kocka ősrégi használata is szerepet játszott abban, hogy a dobókocka a véletlen szimbólumává vált. A kvantummechanika szerint bizonyos mikrovilágra vonatkozó törvényszerűségek statisztikus jellegűek. Amikor Albert Einstein ebben kételkedett, ellentétes véleményét így fogalmazta meg: „God does not play dice". Fél évszázaddal később Stephen Hawking, a modern fizika másik nagy zsenije, ismét a dobókockába csomagolta véleményét: „God not only plays dice, He also sometimes throws the dice where cannot be seen".

1.8. példa: Legyen A és B az a két esemény, amely az 1.1. példában van leírva. Ekkor $P(A) = 4/36$, mert A-t négy elemi esemény valósítja meg, és minden elemi esemény valószínűsége $1/36$. Hasonlóan, $P(A\bar{B}) = 2/36$, $P(A + B) = 20/36$. Az A és B események nem függetlenek, mert $P(B) = 18/36$, $P(A \cdot B) = 2/36$, $P(A) = 4/36$ és $P(AB) = P(A)P(B)$ nem teljesül. □

A klasszikus valószínűségszámítás körébe tartozó feladatok megoldása többnyire a kedvező és összes esetek számának kombinatorikus összeszámlálásán alapul. Ezért hasznos emlékeztetni a következő képletekre.

Egy n elemű halmazból képezhető k elemű sorozatok száma

$$\boxed{n \cdot (n-1) \cdot \ldots \cdot (n-k+2) \cdot (n-k+1) = \frac{n!}{(n-k)!}} \tag{1.4.2}$$

ha a sorozatok nem tartalmazhatnak ismétlődést, illetve

$$\boxed{n^k} \tag{1.4.3}$$

ha a sorozatok tartalmazhatnak ismétlődést. (Az első esetben ismétlés nélküli, a másodikban ismétléses **variációról** beszélnek a kombinatorikában.) (1.4.2)-nek fontos részesete az $n = k$ eset. Ekkor azt kapjuk, hogy n elem összes lehetséges sorrendjeinek száma $n!$. (A nevezőben felbukkanó $0!$-t 1-nek értelmezzük.)

Az n elemű halmaz k elemszámú részhalmazainak száma

$$\boxed{\frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+2) \cdot (n-k+1)}{1 \cdot 2 \cdot \ldots \cdot (k-1) \cdot k} = \frac{n!}{k! \cdot (n-k)!} = \binom{n}{k}.} \tag{1.4.4}$$

(Ez az **ismétlés nélküli kombinációkra** vonatkozó képlet.)

Az **ismétléses kombináció** alapfeladata az, hogy n különböző fajta tárgyunk van, mindegyikből tetszőlegesen sok és k darabból álló csoportokat képezünk. A lehetőségek száma

$$\boxed{\binom{n+k-1}{k}.} \tag{1.4.5}$$

Ugyanazt a feladatot úgy is megfogalmazhatjuk, hogy n egymástól megkülönböztethetetlen részecskét kell k energiaszintre elhelyezni.

Ha l különböző számunk van, az első fajtából n_1, a másodikból n_2, és így tovább, az l-edik fajtából n_l, akkor az $n_1 + n_2 + \ldots + n_l = n$ darab számot

$$\boxed{\frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_l!}} \tag{1.4.6}$$

féleképpen lehet sorrendbe állítani, ez az **ismétléses permutáció**.

1.9. példa: 8 kártyára felírjuk az F, F, L, O, R, R, U, U betűket, majd urnába téve őket egymás után húzunk. Mi a valószínűsége, hogy olyan sorrendben hú zzuk őket ki, hogy éppen a FURFUROL szó olvasható?

A betűk összes lehetséges sorrendjeinek száma 8!/2!2!2! az (1.4.6) képlet szerint. A keresett valószínűség ennek a reciproka: $\frac{1}{7!}$. □

1.10. példa: Hányszor kell ahhoz dobni egy kockával, hogy 99% valószínűséggel legyen a dobások között 6-os?

Ha n-szer dobunk, akkor az összes lehetőségek száma 6^n, ennyi n tagú sorozat képezhető az $\{1, 2, 3, 4, 5, 6\}$ halmazból, mivel az ismétlődés meg van engedve. Ebből 5^n olyan sorozat van, amely nem tartalmaz 6-ost. Így a keresett n az

$$\frac{5^n}{6^n} < 0.01$$

egyenlőtlenséget elégíti ki. Innen $n > \log 0.01/\log 0.833 \approx 25.2585$. □

Számos, a statisztikus fizikához tartozó feladat megoldható kombinatorikus úton. Erre mutatunk három példát. A feladatok mindegyike visszavezethető n számú golyónak N számú dobozban való elhelyezésére, ahol az elhelyezés feltételei különbözőek.

Gázmolekulák különböző rendszerei alkalmazhatóak az ún. Maxwell– Boltzmann statisztikában.

1.11. példa: Helyezzünk el n megkülönböztethető részecskét N szinten. Mi a valószínűsége annak, hogy egy adott szinten pontosan k darab részecske lesz? Alkalmazzuk a klasszikus valószínűség elméletről tanultakat!

Az összes eset száma: N^n, azaz N elem n-ed osztályú ismétléses variációinak a száma. (A Maxwell–Boltzmann statisztikában az a feltevés, hogy minden elhelyezkedés egyforma valószínűségű.)

A kedvező esetek száma: $\binom{n}{k} \cdot (N-1)^{n-k}$, mivel n különböző részecskéből k darabot $\binom{n}{k}$ féleképpen tudunk kiválaszatni, és a maradék $n-k$ részecskét $(N-1)^{n-k}$ féleképpen lehet a fennmaradt $N-1$ szinten elhelyezni.

Tehaát a keresett valószínűség: $p_k = \binom{n}{k} (N-1)^{n-k}/N^n$. □

Másfajta részecskéket pl. fotonokat tartalmazó rendszerek esetében az ún. Bose–Einstein statisztikát használjuk. Alapvető különbség, hogy ebben a modellben a részecskék megkülönböztethetetlenek.

1.12. példa: Helyezzünk el n egymástól megkülönböztethetetlen részecskét N energia szinten. Mi a valószínűsége annak, hogy egy adott szinten pontosan k darab részecske lesz?

Az összes esetek száma: $\binom{N+n-1}{n}$, azaz N elemből képezhető n-ed osztályú ismétléses kombinációk száma. Az egyes elrendezések ebben a modellben is egyforma valószínűségűek. A kedvező esetek száma megegyezik azon elrendezések számával, ahányféleképpen a fennmaradt

$N-1$ szinten a maradék $n-k$ részecskét el lehet helyezni, ezt megint ismétléses kombinációval tudjuk megadni: $\binom{N+n-k-2}{n-k}$.

Így a keresett valószínűség : $p_k = \dfrac{\binom{N+n-k-2}{n-k}}{\binom{N+n-1}{n}}$. □

A Bose–Einstein modell sem általános érvényű. A részecskék megkülönböztethetetlensége mellett bizonyos rendszerek esetében a Pauli elvet is figyelembe kell venni, azaz minden szintre maximálisan egy részecske kerülhet. Ezt a jelenséget jól leírja a Fermi–Dirac statisztika.

1.13. példa: Helyezzünk el n egymástól megkülönböztethetetlen részecskét N szinten úgy, hogy minden szintre legfeljebb egy részecske kerüljön. Mi annak a valószínűsége, hogy egy tetszőlegesen kiválasztott szinten van részecske?

Az összes esetek száma: $\binom{N}{n}$, azaz ahányféleképpen kitudjuk válaszatni azt az n szintet, ahova részecske kerül. Kedvező esetek száma: $\binom{N-1}{n-1}$, azaz ahányféleképpen a maradék $n-1$ részecskét el tudjuk rendezni a fennmaradt $N-1$ szinten.

Így a keresett valószínűség: $p = \dfrac{\binom{N-1}{n-1}}{\binom{N}{n}}$.

1.5. Geometriai valószínűség

A valószínűségszámítás klasszikus képlete nem mond semmit arra az esetre, ha az eseménytér elemeinek a száma végtelen. Akadnak olyan valószínűségelméleti problémák, amelyekben a keresett valószínűség meghatározását visszavezetjük geometriai alakzatok mértékének kiszámítására.

Az olyan véletlen tömegjelenségek esetében beszélünk geometriai valószínűségről, amelynél a jelenséggel kapcsolatos kísérlet egy geometriai alakzat valamely pontjának véletlenszerű kiválasztásából áll, és az alakzat bármely részhalmazára nézve annak valószínűsége, hogy a kiválasztott pont ebbe a részhalmazba esik, az illető részhalmaz geometriai mértékével arányos.

1.14. példa: Egy üzemben két komponensből állítanak elő egy anyagot. Az egyes komponensek egy órán belül készülnek el, és az elkészülés után 15 percen belül össze kell őket dolgozni, különben tönkremennek. Mi a valószínűsége, hogy sikerül előállítani az anyagot, és nem megy tönkre egyik komponens sem?

A jelenséget felfoghatjuk úgy is, mint a $H = [0.60] \times [0.60]$ négyzet egy pontjának véletlenszerű kiválasztását. Ekkor $A = \{(x,y)|x,y \in [0.60]$ és $|x-y| \leqslant 15\}$ esemény valószínűségét keressük. Az eseményt kielégítő pontok koordinátái kielégítik az $y \geqslant x-15$ és $y \leqslant x+15$ egyenlőtlenségeket.

Az A eseménynek megfelelő halmaz területe: $t_A = 60^2 - 45^2$. Az eseményteret jellemző H halmaz területe $t_H = 60^2$ Így $P(A) = \frac{7}{6}$.

Gyakorlatok

1.1. gyakorlat:$^{\mathrm{M}}$ Legyenek A és B egymást kizáró események, és $P(A) = 0.20$, $P(B) = 0.55$. Számoljuk ki a következő valószínűségeket: **(a)** $P(\bar{A})$, **(b)** $P(\bar{A}\cdot\bar{B})$, **(c)** $P(A+B)$, **(d)** $P(A\cdot B)$!

1.2. gyakorlat: Annak a valószínűsége, hogy egy építkezés időben befejeződik 17/20, annak a valószínűsége, hogy a munkások nem sztrájkolnak 3/4, és annak a valószínűsége, hogy időben elkészül az épület, feltéve, hogy nincsen sztrájk, 14/15. Mi a valószínűsége annak, hogy **(a)** időben elkészül az épület, és nincsen sztrájk, **(b)** nincsen sztrájk, feltéve, hogy időben befejeződik a munka?

1.3. gyakorlat:$^{\mathrm{M}}$ Egy dobozban 7 hibátlan és 3 selejtes termék van. Mi a valószínűsége, hogy a termékeket véletlenszerűen egymás után kivéve a 3 selejtes marad utoljára?

1.4. gyakorlat: 100 alma közül 10 férges. Válogatás nélkül kiválasztunk 5-öt. Mi a valószínűsége, hogy van köztük férges?

1.5. gyakorlat:$^{\mathrm{M}}$ Egy kockát ismételten feldobva mi a valószínűsége, hogy előbb kapunk 1-est, mint párost?

1.6. gyakorlat:$^{\mathrm{M}}$ 10 kockával dobva mi a valószínűsége, hogy a számok összege legalább 58?

1.7. gyakorlat:$^{\mathrm{M}}$ Feldobunk egy kockát. Ha az eredmény 1 vagy 2, akkor a 6 fehér és 3 piros golyót tartalmazó I. urnából húzunk, egyébként a 4 fehér és 4 piros golyót tartalmazó II. urnából. Mi a valószínűsége, hogy a kihúzott golyó fehér?

1.8. gyakorlat: Háromszor dobunk fel egy szabályos érmét. Jelentse A azt az eseményt, hogy a dobások között fej és írás is előfordul, B pedig azt az eseményt, hogy legfeljebb 1 írás fordul elő. Döntsük el, hogy független-e A és B!

1.9. gyakorlat:$^{\mathrm{M}}$ Szabályos érmével $2n$-szer dobunk. Legyen p_{2n} annak a valószínűsége, hogy ugyanannyi fej van mint írás. Bizonyítsuk be, hogy p_{2n} fogyó függvénye n-nek!

1.10. gyakorlat: Igazolja, hogy

$$P(A|\bar{B}) = \frac{P(A) - P(AB)}{1 - P(B)}$$

ha A és B tetszőleges események, és $P(B) \neq 1$!

1.11. gyakorlat:$^{\mathrm{M}}$ Két ember háromszor feldob egy-egy pénzdarabot. Mi a valószínűsége annak, hogy ugyanannyiszor kapnak fejet?

1.12. gyakorlat:$^{\mathrm{M}}$ Egy csomag 32 lapos magyar kártyából kihúzunk egy lapot, de nem nézzük meg. Ezután kihúzunk még két lapot, és azt találjuk, hogy mindkettő piros. Mi a valószínűsége, hogy az elsőre húzott lap is piros?

1.13. gyakorlat: Az ,,egyszeri statisztikus" soha nem ül repülőgépre, mert olvasta, hogy a gépeken időnként bomba van. Egyszer barátja mégis összetalálkozik vele egy légijáraton. ,,Már nem tartasz a bombától?" - kérdezi tőle. A statisztikus közelebb hajol, és úgy válaszol. ,,Annak a valószínűsége, hogy egy gépen két bomba van rendkívül kicsi, ezért hoztam magammal egy bombát."

Elemezze, hogy hol a hiba a statisztikus elgondolásában!

1.14. gyakorlat: Egy urnában 10 piros és 20 fekete golyó van. Az urnából kihúzunk egy golyót, ha az piros, akkor visszatesszük, ha fekete akkor nem. Ezután még egyszer húzunk. **(a)** Igaz-e, hogy az az esemény, hogy az első húzás fekete független attól az eseménytől, hogy a második húzás fekete? **(b)** Mi a valószínűsége, hogy mindkét húzott golyó fekete? **(c)** Mi a valószínűsége, hogy mindkét húzott golyó piros?

1.15. gyakorlat:* A legrégibb valószínűségi probléma egy 1494-ben Velencében megjelent könyvben olvasható, és osztozkodási probléma néven vált közismertté. Két csapat labdajátékot játszik, és az a csapat nyeri a tétet, amelyik előbb éri el az n pontot. Feltevés szerint annak a valószínűsége, hogy az egyik csapat pontot nyer, ugyanakkora, mint a másik csapat pontnyerésének az esélye. Amikor az egyik csapatnak $t(< n)$ pontja, a másiknak $u(< n)$ pontja van, a játék valamilyen okból félbeszakad. Az a kérdés, hogyan kell elosztani igazságosan, vagyis a nyerési esélyek arányában, a tétet. Az osztozkodási probléma helyes megoldását Pascal és Fermat találta meg 1654-ben egymástól függetlenül.

Oldja meg az osztozkodási problémát az $n = 4$, $t = 2$ és $u = 1$ paraméterválasztás mellett!

1.16. gyakorlat: Egy urnában 10 piros és 30 fekete golyó van. Az urnából egymás után kihúzunk egy-egy golyót. Mi a valószínűsége, hogy mindkét húzott golyó fekete, ha **(a)** az először húzott golyót visszatesszük a második húzás előtt, **(b)** nem tesszük vissza?

1.17. gyakorlat:$^{\mathbf{M}}$ Egy kockával háromszor dobunk. Mi a valószínűsége, hogy három különböző eredményt kapunk?

1.18. gyakorlat: A, B és C egymást páronként kizáró események. Lehetséges-e, hogy $P(A) = 0.4$, $P(B) = 0.4$ és $P(A + C) = 0.2$?

1.19. gyakorlat:$^{\mathbf{M}}$ A és B egymást kizáró események, és $P(A) = 0.2$, $P(B) = 0.55$. Számolja ki a $P(A \cdot \bar{B})$ és a $P(A + B|B)$ valószínűségeket!

1.20. gyakorlat:$^{\mathbf{M}}$ A és B egy adott tárgy hallgatói. A 80 % valószínűséggel, B 60 % valószínűséggel vesz részt egy órán. Hiányzásaik egymástól függetlenek. Mi a valószínűsége, hogy egy órán legalább egyikük jelen van?

1.21. gyakorlat:$^{\mathbf{M}}$ Egy kockával háromszor egymás után dobunk. Mi a valószínűsége, hogy a kapott számok szigorúan növekedő sorozatot alkotnak?

1.22. gyakorlat: Egy urnában 10 piros és 20 fekete golyó van, egy másikban 15 piros és 15 fekete. Először véletlenszerűen kiválasztjuk az egyik urnát, majd abból véletlenszerűen kihúzunk egy golyót. Mi a valószínűsége, hogy feketét húzunk?

1.23. gyakorlat:* A és B olyan játékot játszanak, amelynek minden fordulóját 50-50 % eséllyel nyerik meg. A nyertes minden fordulóban 1 Ft-ot kap a vesztestől, és a játék addig tart, amíg valamelyiküknek elfogy a pénze. A kezdő tőkéjük 10 Ft. Mi a valószínűsége, hogy a játékot B nyeri, ha kezdő tőkéje (a) 1 Ft, (b) 10 Ft, (c) 100 Ft?

1.24. gyakorlat:$^{\mathbf{M}}$ Legyen $\Omega = \{(0,\,0,\,0),(0,\,1,\,1),(1,\,0,\,1),(1,1,0)\}$ eseménytér, amelyben az események egyforma valószínűségűek. Jelentse

A_0: első eleme 0, B_0: második eleme 0, C_0: harmadik eleme 0,
A_1: első eleme egyes, B_1: második eleme egyes, C_1: harmadik eleme egyes
eredményeket. Igazolja, hogy az A_i, B_j, C_k $(i = 0.1,\ j = 0.1,\ k = 0.1)$ események páronként függetlenek, de nem alkotnak teljesen független rendszert.

1.25. gyakorlat: Egy kémiai vegyületben $2K$ számú egyvegyértékű, és N számú kétvegyértékű egység van. Ezek véletlenszerűen egymáshoz kapcsolódva összesen K számú láncmolekulává egyesülnek úgy, hogy mindegyiknek a két végét egy-egy egyvegyértékű egység zárja le. Határozza meg annak a valószínűségét, h0ogy a $0, 1, 2, \ldots N$ darab kétvegyértékű egységet tartalmazó láncmolekulák száma rendre $k_0, k_1, \ldots k_n$, feltéve, hogy minden elrendeződés egyforma valószínűségű!

1.26. gyakorlat:$^{\mathbf{M}}$ Egy állat populációban az egyedek 25 %-a szemmutáns, 50 %-a szárnymutáns, és a szemmutáns állatok 40 %-a szárnymutáns is. Mi annak a valószínűsége, hogy egy véletlenszerűen kiválasztott egyed legalább az egyik mutációtípussal rendelkezik? Mi annak a valószínűsége, hogy az állat nem szárnymutáns, de szemmutáns?

1.27. gyakorlat:$^{\mathbf{M}}$ Tegyük fel, hogy egy radioaktiv sugárzás során egy α részecske valamely adott sejtet 0.75 valószínűséggel talál el, 0.25 valószínűséggel csupán károsítja, 0.5 valószínűséggel pusztulását is okozza. Két részecske által okozott károsodás feltétlenül a sejt pusztulásához vezet. Mennyi a valószínűsége annak, hogy négy találat után a sejt elpusztul?

1.28. gyakorlat:$^{\mathbf{M}}$ N számú sejt besugárzása során n számú elemi részecske mindegyike pontosan egy sejtet talál el, mégpedig minden részecske minden sejtet azonos valószínűséggel. Mennyi a valószínűsége annak, hogy minden sejtet legalább két találat ér.

1.29. gyakorlat:$^{\mathbf{M}}$ Egy alkalmassági vizsgálat tapasztalatai szerint a vizsgált személyeken 0.05 valószínűséggel mozgásszervi, és 0.03 valószínűséggel érzékszervi rendellenesség figyelhető meg. Az együttes előfordulás valószínűsége 0.01. Mennyi a valószínűsége annak, hogy egy találomra kiválasztott személyen egyik rendellenesség sem figyelhető meg?

1.30. gyakorlat:$^{\mathbf{M}}$ Az ábrán látható úgynevezett kontingenciatáblázat hiányos. Kíséreljük meg kiegészíeteni oly módon, hogy a valószínűségeket a megfelelű relatív gyakoriságokkal helyettesítve a táblázat az A és B események függetlenségét tükrözze!

Második fejezet

Valószínűségi változók

A valószínűségelmélet Kolmogorov-féle modelljében a valószínűségi változó az elemi események Ω halmazán értelmezett függvény. Kiindulási példaként gondoljunk arra, hogy véletlen kísérletünk a kétszeri kockadobás. Ekkor a két dobás összege egy véletlen szám, másszóval valószínűségi változó. E valószínűségi változó lehetséges értékei a 2 és 12 közé eső egész számok, a 2-t és a 12-t is beleértve. Másik példa: egy céghez bizonyos időtartam alatt befutó telefonhívások száma. Ennek a valószínűségi változónak is egész számok az értékei. Az ilyen **diszkrét** valószínűségi változók mellett **folytonos** valószínűségi változók is vannak, amelyek értéke elvileg a valós számok valamely intervallumának bármely eleme lehet. Gondoljunk például egy izzó élettartamára, amely pozitív értékeket felvevő valószínűségi változónak tekinthető.

2.1. Diszkrét változó eloszlása

Egy valószínűségi változóval kapcsolatban mindig az a lényeges kérdés, hogy egy adott értéket milyen valószínűséggel vesz fel. (Ha két valószínűségi változó ugyanazokat az értékeket ugyanolyan valószínűséggel veszi fel, akkor valószínűségelméleti szempontból lényegtelen, hogy az egyik testmagasságokat, a másik pedig hőmérsékleteket fejez ki.) Az a ξ (kszi) valószínűségi változó, amelynek értéke két kockadobás esetén a dobások összege, a 3 értéket $2/36$ valószínűséggel veszi fel, mivel a 3-as dobásösszeg $(1,2)$ és $(2,1)$ formában valósulhat meg, és mindkét elemi esemény valószínűsége $1/36$. Valószínűségelméleti szempontból a ξ-re vonatkozó adatok (egy része) a következő táblázatban foglalhatók össze:

n :	2	3	4	5	6	7	8
$P(\xi = n)$:	0.028	0.056	0.083	0.111	0.139	0.167	0.139

A táblázat tartalmazza a

$$P(\xi = n) = P\left(\{\omega \in \Omega : \xi(\omega) = n\}\right)$$

értékeket, amelyek ξ **valószínűség eloszlását** alkotják. ξ eloszlása a

$$P(\xi = n) = \begin{cases} (n-1)/36 & \text{ha } 2 \le n \le 7\,, \\ (13-n)/36 & \text{ha } 8 \le n \le 12\,, \\ 0 & \text{egyébként} \end{cases}$$

képlettel is megadható és hisztogrammal is szemléltethető.

2.1. példa: Legyen az η (éta) valószínűségi változó értéke a fejek száma egy szabályos érme háromszori feldobása során. Mi η valószínűség eloszlása?

Az egyes elemi események egyformán valószínűek, ezért használhatjuk a klasszikus (1.4.1) képletet. Az összes lehetséges esetek száma 8. A kedvező esetek összeszámlálásával

$$P(\eta = 0) = P(\eta = 3) = 1/8, \qquad P(\eta = 2) = P(\eta = 1) = 3/8\,. \qquad \square$$

2.2. példa: Egyszerű alternatívának nevezzük azt a véletlen kísérletet, amelynek két kimenetele van. Jelöljük ezeket **s**-sel (s=siker) és **k**-val (k=kudarc), legyen **s** bekövetkezésének valószínűsége p. Ismételjük meg a kísérletet n-szer. Mi az **s**-ek ξ számának eloszlása?

Meg kell határozni, mi annak a valószínűsége, hogy pontosan m-szer következik be a siker a kisérlet n-szeri ismétlése során. Egy olyan véletlen sorozat, amely m számú **s**-et tartalmaz meghatározott helyen, például az első m helyen, $p^m(1-p)^{n-m}$ valószínűséggel valósul meg. Ugyanakkor az m számú **s**-nek és $n-m$ darab **k**-nak $n!/m!(n-m)!$ sorrendje van (lásd (1.4.4)-et). Ennyiféleképpen választhatjuk ki azt az m számú pozíciót, ahová **s** kerül a többi **k**-val szemben. Így a keresett eloszlás:

$$\boxed{P(\xi = m) = \binom{n}{m} p^m (1-p)^{n-m} \qquad (0 \le m \le n)} \tag{2.1.1.}$$

Ezt az eloszlást (n-ed rendű) **binomiális eloszlásnak** hívják, és p neve **sikervalószínűség**.

$\square$

Ha $p = 1/2$, akkor a binomiális eloszlás szimmetrikus: a $P(\xi = m)$ és $P(\xi = n - m)$ valószínűségek megegyeznek. Ha a sikervalószínűség 1/2-től különbözik, akkor nincsen szimmetria, az eloszlás csúcsa eltolódik.

2.3. példa: Egy érmével egymás után dobunk mindaddig, amíg fejet nem kapunk. A szükséges dobások száma a ξ valószínűségi változó értéke. Mi ξ eloszlása?

Tanulságos pontosan megfogalmazni, hogy ebben a feladatban mi az eseménytér. Ω az összes olyan végtelen sorozatok halmaza, amelyek tagjai F(=fej) vagy I(=írás). Ha

$$\omega = (\omega_1, \omega_2, \ldots)\,,$$

akkor ξ értéke ezen az ω elemi eseményen k, ha $\omega_k = F$, de az $(\omega_1, \omega_2, \ldots, \omega_{k-1})$ véges sorozat nem tartalmaz F-et. (A ξ valószínűségi változót az F-re való **várakozás idejének** is szokás nevezni.) Nekünk a $P(\xi = k)$ valószínűséget kell meghatároznunk. Ez

$$P(\xi = k) = \frac{1}{2^k} \qquad (k \geqslant 1).$$

Megjegyezzük, hogy ha az érmedobás helyett az egyszerű alternatívát ismételjük p sikervalószínűséggel, akkor a sikerre való várakozás valószínűség eloszlása

$$\boxed{P(\xi = k) = \begin{cases} 0 & \text{ha } k = 0, \\ q^{k-1}p & \text{ha } k > 0\,. \end{cases}} \tag{2.1.2}$$

$\square$

A (2.1.2) képlettel adott eloszlást **geometriai eloszlásnak** nevezzük.

2.2. Eloszlásfüggvény és sűrűségfüggvény

Legyen $\eta : \Omega \to \mathbb{R}$ egy valós értékeket felvevő valószínűségi változó. Az η valószínűségi változó F_η **eloszlás függvényét** a

$$\boxed{F_\eta(x) = P(\{\omega \in \Omega : \eta(\omega) < x\})} \tag{2.2.1}$$

képlet értelmezi. Tehát annak a valószínűségét adja meg, hogy $\eta < x$. Ha $x < z$, akkor $\{\omega : \eta(\omega) < x\}$ teljesülése maga után vonja a $\{\omega : \eta(\omega) < z\}$ teljesülését, így $F_\eta(x) \leq F_\eta(z)$, és megállapíthatjuk, hogy az eloszlásfüggvény monoton növekedő. Bármilyen η valószínűségi változó F_η eloszlásfüggvénye rendelkezik a következő tulajdonságokkal:

- F_η monoton növekedő,
- $\lim_{x\to-\infty} F_\eta(x) = 0$,
- $\lim_{x\to\infty} F_\eta(x) = 1$,
- F_η balról folytonos.

Ugyanakkor, ha egy függvény a fenti tulajdonságokkal rendelkezik, akkor alkalmas valószínűségi változó eloszlásfüggvénye lehet.

Mivel tetszőleges η valószínűségi változóra

$$\{\omega : \eta(\omega) < c + d\} = \{\omega : \eta(\omega) < c\} + \{\omega : c \leq \eta(\omega) < c + d\}\,,$$

a valószínűség additivitása alapján megállapíthatjuk, hogy

$$\boxed{P(c \leq \eta < c + d) = F_\eta(c + d) - F_\eta(c)} \tag{2.2.2}$$

Az olyan valószínűségi változót, amely az $[a, b]$ intervallum részintervallumaiban az illető részintervallum hosszával arányos eséllyel veszi fel értékeit, az $[a, b]$-ben **egyenletes eloszlású nak** nevezzük. Az $[a, b]$-ben egyenletes eloszlás eloszlásfüggvénye

$$\boxed{F(x) = \begin{cases} 0 & \text{ha } x \leq a\,, \\ \dfrac{x-a}{b-a} & \text{ha } a < x \leq b\,, \\ 1 & \text{ha } b < x\,. \end{cases}} \tag{2.2.3}$$

2.4. példa: Legyen a ξ valószínűségi változó egyenletes eloszlású a $[-1, 1]$ intervallumban. Írjuk fel ξ^2 eloszlásfüggvényét!

ξ^2 a $[0, 1]$-ból veszi fel értékeit, így az eloszlásfüggvény $F_{\xi^2}(x)$ értéke 0 negatív x-ekre, és $F_{\xi^2}(x) = 1$, ha x 1-nél nagyobb. Ha $0 < x < 1$, akkor (2.2.3) alapján

$$\begin{aligned} F_{\xi^2}(x) &= P\left(\xi^2 < x\right) = P\left(-\sqrt{x} < \xi < \sqrt{x}\right) \\ &= P(\xi < \sqrt{x}) - P(\xi < -\sqrt{x}) = F(\sqrt{x}) - F(-\sqrt{x}) \\ &= \frac{\sqrt{x}+1}{2} - \frac{-\sqrt{x}+1}{2} = \sqrt{x}\,. \end{aligned}$$

Tehát

$$F_{xi^2} = \begin{cases} 0 & \text{ha } x \leqslant 0, \\ \sqrt{x} & \text{ha } 0 < x \leqslant 1\,. \\ 1 & \text{ha } 1 < x \end{cases}$$

□

2.5. példa: Az a és b állandók milyen értékére lesz

$$F(x) := a\,\mathrm{arctg}(x) + b \qquad (x \in \mathbb{R})$$

egy valószínűségi változó eloszlásfüggvénye?
Mivel

$$\lim_{x\to-\infty} F(x) = -\frac{a\pi}{2} + b\,, \qquad \lim_{x\to\infty} F(x) = \frac{a\pi}{2} + b\,,$$

a és b értéke az

$$-\frac{a\pi}{2} + b = 0\,, \quad \frac{a\pi}{2} + b = 1$$

egyenletrendszerből adódik. Tehát $a = 1/\pi$ és $b = 1/2$. F nyilvánvalóan monoton növekedő és balról folytonos. □

Amennyiben a ξ valószínűségi változó F eloszlásfüggvényének $F' = f$ deriváltfüggvénye létezik, úgy azt ξ **sűrűségfüggvényének** nevezzük. Az eloszlás- és sűrűségfüggvények között így az

$$\boxed{F'(x) = f(x)\,, \qquad F(x) = \int_{-\infty}^{x} f(t)\,dt} \tag{2.2.4}$$

összefüggések állnak fenn. Egy sűrűségfüggvényt az

$$f(x) \geqslant 0\,, \qquad \int_{-\infty}^{\infty} f(t)\,dt = 1 \tag{2.2.5}$$

tulajdonságok jellemeznek.

2.6. példa: Egy valószínűségi változó sűrűségfüggvénye

$$f(x) = \begin{cases} 0 & \text{ha } x \le 0, \\ \lambda e^{-\lambda x} & \text{ha } x > 0. \end{cases} \tag{2.2.6}$$

Mi az eloszlásfüggvénye?

A (2.2.4) képlet alapján integrálással

$$F(x) = \begin{cases} 0 & \text{ha } x \le 0, \\ 1 - e^{-\lambda x} & \text{ha } x > 0. \end{cases} \tag{2.2.7}$$

□

A (2.2.6) sűrűségfüggvényű és (2.2.7) eloszlásfüggvényű valószínűségeloszlást **exponenciális eloszlásnak** nevezzük.

2.3. A várható érték és a szórás

A véletlen kísérletet sorozatosan végrehajtva egy véletlen mennyiség értékeinek számtani átlaga konvergál egy számhoz, amit a véletlen mennyiség (azaz valószínűségi változó) **várható értékének** nevezünk. A várható érték tehát átlagérték, és előfordulhat, hogy a valószínűségi változó sohasem veszi fel magát ezt az értéket. A $\xi : \Omega \to \mathbb{R}$ valószínűségi változó várható értéke

$$M(\xi) = \sum_{\omega \in \Omega} \xi(\omega) P(\{\omega\}) \tag{2.3.1}$$

abban az esetben, ha az eseménytér véges sok elemből áll. Mivel $M(\xi)$ a valószínűségekkel súlyozott átlagérték, mindig ξ legnagyobb és legkisebb értéke közé esik. A ξ diszkrét valószínűségi változó várható értékét az

$$\boxed{M(\xi) = \sum_{k} P(\xi = x_k) x_k} \tag{2.3.2}$$

képlet adja meg, ha ξ az x_k értéket veszi fel. Ha az Ω eseménytér véges halmaz, akkor a várható érték a következő alakban is írható:

$$M(\xi) = \sum_{\omega} \xi(\omega) P(\{\omega\}) \, . \tag{2.3.3}$$

Itt különösen jól látszik, hogy a várható érték a valószínűségi változó értékeinek a valószínűségekkel súlyozott átlaga. Az $M(\xi)$ jelölés helyett néha $E(\xi)$-t is írunk.

A várható érték a sztochasztikus mennyiség lehetséges értékeinek egyfajta közepe. Hasonló jelentése van a mediánnak. A **medián** az a szám, amelynél nagyobb és kisebb értéket 50-50% valószínűséggel vesz fel a változó. Ha a sűrűségfüggvény nem szimmetrikus, akkor a medián eltérhet a várható értéktől.

Amennyiben a folytonos ξ valószínűségi változónak van f sűrűségfüggvénye, akkor

$$\boxed{M(\xi) = \int_{-\infty}^{\infty} x\, f(x)\, dx \, .} \tag{2.3.4}$$

Ha (2.3.2)-ben az összeg, vagy (2.3.4)-ban az integrál nem konvergens, akkor várható értékről nem beszélhetünk. Ha a (2.3.2) sor végtelen, akkor a várható érték létezéséhez a sornak abszolút konvergensnek kell lennie, hiszen az összeg értéke nem függhet az összeadás sorrendjétől.

2.7. példa: Egy dobókocka két oldalát pirosra, a többit kékre festjük. Legyen ξ az első piros előfordulásáig a dobások száma, ismételt kockadobás esetén. Mi ξ várható értéke?

Lényegében egyszerű alternatíváról van szó, a piros kimenet 1/3 sikervalószínűségével. A 2.3. példa szerint ξ geometriai eloszlású, eloszlását a (2.1.2) képlet adja. A feladat tehát a geometriai eloszlás várható értékének kiszámítása. (Ehhez használni kell a sorok differenciálásáról tanultakat is a végtelen geometriai sor összegét.)

$$\begin{aligned} \sum_{k=1}^{\infty} k q^{k-1} p &= \sum_{k=1}^{\infty} \frac{d}{dq} q^k p = \frac{d}{dq} \sum_{k=1}^{\infty} q^k p \\ &= \frac{d}{dq} \frac{pq}{1-q} = \frac{p}{(1-q)^2} = \frac{1}{p} \end{aligned}$$

Az eredményt érdemes megjegyezni: p sikervalószínűség mellett a geometriai eloszlás várható értéke $1/p$. □

2.8. példa: Egy nagy mennyiségben gyártott elektroncső órában kifejezett élettartama olyan valószínűségi változónak tekinthető, amelynek sűrűségfüggvénye

$$\alpha^2 x e^{-\alpha x} \qquad (x \in \mathbb{R}^+),\ \text{ahol } \alpha \text{ ismeretlen paraméter.} \tag{2.3.5}$$

Mennyi az elekroncsövek átlagos élettartama? Mi a valószínűsége, hogy egy véletlenszerűen választott cső élettartama $\geqslant m$?

Az átlagos élettartam, azaz a (2.3.5)-ben adott valószínűségi változó várható értéke a (2.3.4) képletből adódik, kétszeres parciális integrálással:

$$E = \int_0^\infty \alpha^2 x^2 e^{-\alpha x}\, dx = 2/\alpha\,.$$

A másik kérdésre az

$$\int_m^\infty \alpha^2 x e^{-\alpha x}\, dx = (1 + m\alpha)e^{-\alpha m}$$

integrál a válasz. □

A várható érték alapvető tulajdonsága a linearitás:

$$\boxed{M(\xi_1 + \xi_2) = M(\xi_1) + M(\xi_2) \quad \text{és} \quad M(a\xi) = aM(\xi)} \tag{2.3.6}$$

ha ξ, ξ_1 és ξ_2 valószínűségi változók és a valós szám. A várható értéknek ezt a tulajdonságát nem bizonyítjuk, de az az egyszerű eset, amikor az eseménytér véges, kiolvasható a (2.3.3) képletből.

2.9. példa: Függetlenül ismételt kockadobásban legyen ξ az ahhoz szükséges dobásszám, hogy mind a hat kimenetel legalább egyszer előforduljon. Mi ξ várható értéke?

A ξ valószínűségi változó egyfajta várakozási idő. Egy csepp furfanggal ξ hat valószínűségi változó összegeként fogható fel. ξ_1 az első számra való várakozás ideje. (ξ_1 triviális, azonosan egy.) ξ_2 a második számra való várakozás ideje, attól számítva, hogy megvan az első szám, és így tovább, egészen ξ_6-ig, ami az utolsó hiányzó kimenetre való várakozási idő, az 5. fajta szám megjelenése után. A jobb érthetőség érdekében nézzük meg, hogy az egyes ξ-k milyen értékeket vesznek fel egy konkrét kísérletben az

$$\omega = (3,3,1,3,2,4,2,1,5,2,2,3,6,1,4,2,5,3,\ldots)$$

elemi eseményen: $\xi_2(\omega) = 2$, $\xi_3(\omega) = 2$, $\xi_4(\omega) = 1$, $\xi_5(\omega) = 3$ és $\xi_6(\omega) = 4$.

$$\begin{aligned} \xi &= \xi_1 + \xi_2 + \xi_3 + \ldots + \xi_6 \\ M(\xi) &= M(\xi_1) + M(\xi_2) + M(\xi_3) + \ldots + M(\xi_6) \end{aligned}$$

A 2.3. és 2.7. példákból tudjuk, hogy ξ_i-k geometriai eloszlásúak, különféle sikervalószínűséggekkel. A ξ_i-hez tartozó sikervalószínűség $(7-i)/6$. Valóban, amikor például a 4. kimenetre várunk, akkor a már megdobott három kimenet kedvezőtlen, és $3/6$ a valószínűsége annak, hogy olyat dobunk, ami még nem volt. A 2.7. példában láttuk, hogy

$$M(\xi_i) = \frac{6}{7-i}\,,$$

és a keresett várható érték ezek összege $1 \le i \le 6$ esetére, tehát $\frac{147}{10}$. □

A ξ valószínűségi változó eloszlásának ismeretében ξ tetszőleges g függvényének várható értékét így számolhatjuk:

$$\boxed{M(g(\xi)) = \int_{-\infty}^{\infty} g(x) f_\xi(x)\, dx\,,} \tag{2.3.7}$$

ahol f_ξ a ξ valószínűségi változó sűrűségfüggvénye. A $g(x) = x^k$ választás a valószínűségi változó **k-adik momentumát** adja. Nevezetesen, az első momentum magának ξ-nek a várható értéke, és a második momentum ξ^2 várható értéke. Minél több momentumát ismerjük egy valószínűségeloszlásnak, annál pontosabb a képünk róla. Az összes momentuma – alkalmas feltételek mellett – egyértelműen meghatározza az eloszlást.

2.10. példa: Legyen ξ exponenciális eloszlású valószínűségi változó. A t pozitív valós paraméterre számoljuk ki $\exp(t\xi)$ várható értékét és ξ első három momentumát!

A (2.3.7) képletre van szükségünk:

$$M(\exp(t\xi)) = \int_0^{\infty} \exp(tx)\lambda \exp(-\lambda x)\, dx$$

Az integrál csak $t < \lambda$ esetén véges. Ekkor

$$M(\exp(t\xi)) = \frac{\lambda}{\lambda - t}\,.$$

A $t \mapsto M(\exp(t\xi))$ függvényt (nemcsak exponenciális eloszlású valószínűségi változók esetén) ξ **momentumgeneráló függvényének** is nevezik, mert belőle ξ momentumai $t = 0$ helyen vett ismételt differenciálással kaphatók meg. Valóban,

$$\frac{d^k}{dt^k} \int_{-\infty}^{\infty} \exp(tx) f_\xi(x)\, dx = \int_{-\infty}^{\infty} x^k f_\xi(x)\, dx \tag{2.3.8}$$

$t = 0$-ban. Például:

$$\frac{d}{dt}\frac{\lambda}{\lambda - t} = \lambda(\lambda - t)^{-2}, \qquad \frac{d^2}{dt^2}\frac{\lambda}{\lambda - t} = 2\lambda(\lambda - t)^{-3}.$$

Tehát: $m_1 = \lambda^{-1}$, $m_2 = 2\lambda^{-2}$ és általában $m_k = k!\lambda^{-k}$. □

A momentumgeneráló függvény meghatározza az eloszlást, vagyis ha két valószínűségi változónak ugyanaz a momentumgeneráló függvénye, akkor azonos eloszlásúak.

A **szórás** egy valószínűségi változó ingadozásának leggyakrabban használt mértéke. A szórás négyzete (vagy a valószínűségi változó **varianciája**) a valószínűségi változó várható értéktől való eltérése négyzetének várható értéke:

$$\boxed{D^2(\xi) := \sigma^2 = M\Big((\xi - M(\xi))^2\Big)} \tag{2.3.9}$$

A kis szórás azt jelenti, hogy a valószínűségi változó nagy valószínűséggel a várható értéke közelében veszi fel az értékeit. A szórásnégyzetet általában előnyösebb a fenti definíció helyett a következő képlettel számolni, amely a várható érték tulajdonságainak felhasználásával egyszerűen megkapható:

$$\boxed{\sigma^2 = M(\xi^2) - M(\xi)^2} \tag{2.3.10}$$

Tehát a szórás az első két momentummal fejezhető ki.

2.11. példa: Legyen ξ egyenletes eloszlású valószínűségi változó a $[0,1]$ intervallumban. Számítsuk ki ξ szórását!

ξ első és második momentumát számoljuk ki:

$$\int_0^1 x \cdot 1\, dx = \frac{1}{2}, \qquad \int_0^1 x^2 \cdot 1\, dx = \frac{1}{3}.$$

(2.3.10) szerint a szórás négyzete $1/3 - 1/4 = 1/12$, és $\sigma = 0.289$. □

2.12. példa: Számítsuk ki a λ paraméterű exponenciális eloszlású valószínűségi változó szórását!

A 2.10. példa tartalmazza a momentumokat, így $\sigma = \lambda^{-1}$. □

2.13. példa: A kinetikus gázelmélet szerint a gázrészecskék sebessége olyan nemnegatív értéket felvevő valószínűségi változó, amelynek sűrűségfüggvénye

$$\frac{4}{\alpha^3\sqrt{\pi}} v^2 e^{-v^2/\alpha^2} \qquad (v > 0)\,. \tag{2.3.11}$$

(2.3.11) az úgynevezett **Maxwell-féle sebességeloszlás**, ahol $\alpha = \sqrt{2kT/m}$, m a részecske tömege, T a hőmérséklet és k a Boltzmann-állandó. Számítsuk ki a részecskék átlagos sebességét!

A várható érték definíciója szerint a

$$\int_0^\infty \frac{4}{\alpha^3\sqrt{\pi}} v^3 e^{-v^2/\alpha^2}\, dv$$

integrált kell kiszámolni, amit parciális integrálással teszünk. A v^3 faktort szétválasztjuk $v \cdot v^2$ alakban, hogy a $-v^2/\alpha^2$ kitevő deriváltja megjelenjen:

$$\begin{aligned}
\int_0^\infty \frac{4}{\alpha^3\sqrt{\pi}} v^3 e^{-v^2/\alpha^2}\, dv &= \frac{-2}{\alpha\sqrt{\pi}} \int_0^\infty v^2 \frac{-2v}{\alpha^2} e^{-v^2/\alpha^2}\, dv \\
&= \frac{2}{\alpha\sqrt{\pi}} \int_0^\infty 2v e^{-v^2/\alpha^2}\, dv \\
&= \left[-\frac{2\alpha}{\sqrt{\pi}} e^{-v^2/\alpha^2} \right]_0^\infty = \frac{2\alpha}{\sqrt{\pi}}
\end{aligned}$$

Érdemes megjegyezni, hogy ebben a példában a második momentumnak nem csak valószínűségelméleti jelentése van. m-mel szorozva a második momentumot a részecske kinetikus energiájának átlagértékét kapjuk. □

2.4. A Csebisev-egyenlőtlenség

Egy üzemben acélgolyókat gyártanak d átlagos átmérővel. Azok a golyók, amelyek átmérője d-től 0.01 mm-rel jobban eltér nem megfelelők. Ha ismerjük az átmérők pontos eloszlását, akkor választ tudunk adni arra a kérdésre, hogy legfeljebb mekkora a valószínűsége annak, hogy egy golyó nem megfelelő (azaz gyártási selejt). A Csebisev-egyenlőtlenség ilyen jellegű kérdések megválaszolására használható. A kérdéses valószínűségre felső korlátot szab a szórás ismeretében, anélkül, hogy azy eloszlás típusát, ill. az eloszlás függvényt ismernénk.

2.1. tétel: (Csebisev-egyenlőtlenség) *Legyen ξ valószínűségi változó m várható értékkel és σ szórással. Ekkor*

$$\boxed{P(|\xi - m| \geqslant t) \leq \frac{\sigma^2}{t^2}}$$

minden pozitív t számra.

Bizonyítás: A bizonyítást csak arra az esetre nézzük, amikor $m = 0$ és ξ diszkrét, vagyis $P(\xi = x_k) = p_k$. Ekkor

$$P(|\xi - m| \geqslant t) = \sum_{k \in I_t} p_k$$

ahol az összegzés azokra a k-kra történik, amelyekre $|x_k| \geqslant t$. (Legyen ezek halmaza I_t.) Nyilván

$$\sum_{k \in I_t} p_k \leq \sum_{k \in I_t} \frac{|x_k|^2}{t^2} p_k \leq \sum_k \frac{|x_k|^2}{t^2} p_k$$

mivel az első esetben az egyes tagokat növeltük, a másodikban pedig újabb (pozitív) tagokat vettünk az összeghez. Az egyenlőtlenséglánc utolsó kifejezése éppen σ^2/t^2, mert $m = 0$. □

2.5. Függetlenség

A ξ és η valószínűségi változók függetlensége azt jelenti, hogy bármely ξ segíségével megadott esemény független bármely η-val megadott eseménytől. Ha $A, B \subset \mathbb{R}$ tetszőleges halmazok, akkor az $\{\omega : \xi(\omega) \in A\}$ esemény független az $\{\omega : \eta(\omega) \in B\}$ eseménytől, képletben

$$P(\xi \in A, \eta \in B) = P(\xi \in A) P(\eta \in B)\,. \tag{2.5.1}$$

Ebből a meghatározásból nyomban adódik, hogy ha ξ és η függetlenek, akkor $g(\xi)$ és $h(\eta)$ is függetlenek akármilyen g és h függvényekre.

2.2. tétel: *Ha a ξ és η valószínűségi változók függetlenek, akkor*

$$\boxed{M(\xi\eta) = M(\xi) M(\eta).}$$

Bizonyítás: A bizonyítást csak abban az esetben tárgyaljuk, amikor ξ és η diszkrétek. Jelöljük x_i-vel ξ, y_j-vel η és z_k-val $\xi\eta$ lehetséges értékeit. Ekkor

$$\begin{aligned} M(\xi) &= \sum_i x_i P(\xi = x_i) \\ M(\eta) &= \sum_i y_j P(\eta = y_j) \\ M(\xi\eta) &= \sum_i z_k P(\xi\eta = z_k) \end{aligned}$$

$\xi\eta$ a z_k értéket úgy veheti fel, ha $z_k = x_i y_j$, ami esetleg többféleképpen is megvalósulhat, de mindig igaz, hogy

$$\begin{aligned} P(\xi\eta = z_k) &= \sum_{x_i y_j = z_k} P(\xi = x_i, \eta = y_j) \\ &= \sum_{x_i y_j = z_k} P(\xi = x_i) P(\eta = y_j)\,. \end{aligned}$$

Ezt figyelembe véve adódik, hogy $M(\xi)$-t és $M(\eta)$-t összeszorozva $M(\xi\eta)$-t kapjuk. □

2.3. tétel: *Ha a ξ és η valószínűségi változók függetlenek, akkor*

$$\boxed{\sigma^2(\xi + \eta) = \sigma^2(\xi) + \sigma^2(\eta)}\,.$$

Bizonyítás: A (2.3.10) képletből indulunk ki, amely azt adja, hogy

$$\begin{aligned} \sigma^2(\xi + \eta) &= M((\xi^2 + 2\xi\eta + \eta^2) - (M(\xi^2) + 2M(\xi)M(\eta) + M(\eta^2) \\ &= (M(\xi^2) - M(\xi)^2) + (M(\eta^2) - M(\eta)^2) \\ &\quad +2(M(\xi\eta) - M(\xi)M(\eta))\,, \end{aligned}$$

ahol az utolsó tag zérus a függetlenség miatt. □

2.14. példa: Egy dohányos minden nap p valószínűséggel szív el 10 és $1-p$ valószínűséggel 11 cigarettát. Mennyi az 1 hónap (= 30) nap alatt elszívott cigaretták számának várható értéke és szórása?

Jelölje a ξ_n valószínűségi változó az n-edik napon szívott cigaretták számát. Feltételezzük, hogy ezek a valószínűségi változók teljesen függetlenek. Kiszámítandó

$$E\Big(\sum_{n=1}^{30} \xi_n\Big) = \sum_{n=1}^{30} E(\xi_n) \qquad \text{és} \qquad \sigma^2\Big(\sum_{n=1}^{30} \xi_n\Big) = \sum_{n=1}^{30} \sigma^2(\xi_n)\,.$$

Mivel bármilyen n-re $E(\xi_n) = 10p + 11(1-p) = 11 - p$ és $\sigma^2(\xi_n) = 10^2 p + 11^2(1-p) - (11-p)^2 = p(1-p)$ a keresett várható érték $30(11-p)$ és a keresett szórás $\sqrt{30p(1-p)}$. □

2.4. tétel: *Ha a ξ és η valószínűségi változók függetlenek és momentumgeneráló függvényük m_ξ és m_η, akkor $\xi + \eta$ momentumgeneráló függvénye $m_\xi m_\eta$.*

Bizonyítás: A momentumgeneráló függvény definíciója szerint az

$$m_{\xi+\eta}(t) = M(\exp t(\xi + \eta)) = M(\exp(t\xi)\exp(t\eta))$$

várható értéket kell kiszámolni. Mivel $\exp(t\xi)$ és $\exp(t\eta)$ független valószínűségi változók, a 2.2. Tétel alkalmazható, és

$$M(\exp(t\xi)\exp(t\eta)) = M(\exp(t\xi))M(\exp(t\eta))\,.$$

Ez a tétel bizonyítását teljessé teszi. □

2.6. A nagy számok törvénye

A nagy számok törvényére a hétköznapi életben is gyakran hivatkozunk. A törvény (azaz a tétel) azt fejezi ki, hogy függetlenül ismételt kísérletsorozat esetén a mérési eredmények átlagértéke a várható értéket közelíti, elvileg tetszőleges pontossággal. Például a kockadobásra vonatkoztatva a nagy számok törvénye azt mondja, hogy átlagosan a dobások egy hatoda 5-ös. (Arról nem szól a törvény, hogy legalább hányat kell dobni ahhoz, hogy 5-öst kapjunk.)

2.5. tétel: *Legyen $\xi_1, \xi_2, \ldots \xi_n$ azonos eloszlású, m várható értékű, σ szórású, és teljesen független valószínűségi változók sorozata, és legyen $\eta_n = (\xi_1 + \xi_2 + \ldots + \xi_n)/n$ az első n változó átlaga. Ekkor*

$$\lim_{n\to\infty} P(|\eta_n - m| < r) = 1$$

bármilyen $r > 0$ számra.

Bizonyítás: Azt fogjuk belátni, hogy $\lim_{n\to\infty} P(|\eta_n - m| \geqslant r) = 0$, ami ekvivalens magával a tétellel. A várható érték tulajdonságaiból következik, hogy η_n várható értéke is m, legyen szórása σ_n. A Csebisev-egyenlőtlenséget használjuk:

$$P(|\eta_n - m| \geqslant r) \leq \frac{\sigma_n^2}{r^2}.$$

Mivel $\xi_1, \xi_2, \ldots \xi_n$ szórása σ, ezért a 2.3. tétel szerint $\sigma_n^2 = \sigma^2/n$, ami r^2-tel osztva is 0-hoz tart, ha n tart a végtelenhez. □

A fenti tételt a nagy számok gyenge törvényének is szokás nevezni, de itt más változatokkal nem foglalkozunk. Megjegyezzük azonban, hogy a $P(|\eta_n - m| \geqslant r)$ valószínűségek 0-hoz tartása sokkal gyorsabb, mint ahogyan az a bizonyításból látszik. Ott $1/n$ gyorsaságú konvergencia szerepelt, de valójában exponenciálisan tartanak a $P(|\eta_n - m| \geqslant r)$ valószínűségek a 0-hoz. Ez azt jelenti, hogy vannak olyan $C_r > 0$ és $0 < D_r < 1$ állandók, hogy

$$P(|\eta_n - m| \geqslant r) \leq C_r D_r^n.$$

Minnél nagyob az r, annál gyorsabb az exponenciális konvergencia, tehát nagyobb r-re D_r kisebb. (Ez a tartalma az ún. **nagyeltérés-tételeknek**.)

Gyakorlatok

2.1. gyakorlat: N ember vérvizsgálata a következő két módon szervezhető: **(a)** Minden egyes embertől származó vérmintát külön-külön megvizsgálnak, és így N próbára van szükség. **(b)** k darab mintát összeöntenek, és azokat egyszerre vizsgálják meg. Ha a próba negatív, akkor ez az egyetlen próba elegendő a k személy vérvizsgálatához. Amennyiben a próba pozitív, mind a k személy vérmintáját újabb próbának vetik alá. Ebben az esetben tehát $k + 1$ próba szükséges.

Tételezzük fel, hogy az egyes személyek vizsgálati eredményei statisztikailag egymástól függetlenek, és annak a valószínűsége, hogy bármely ember vérmintája pozitív eredményt ad ugyanaz a p szám. Legyen η az a valószínűségi változó, amely megadja az N ember vizsgálatához szükséges próbák számát a (b) szerinti szervezésben. Mi η várható értéke?

2.2. gyakorlat:$^{\mathrm{M}}$ Legyen ξ valószínűségi változó és a, b valós számok. Fejezzük ki $a\xi + b$ varianciáját ξ varianciája segítségével!

2.3. gyakorlat: Mi a valószínűsége annak, hogy az $x^2 + \xi x + \eta = 0$ egyenlet gyökei valósak, ha ξ és η függetlenek és [-1, 1]-ben egyenletes eloszlásúak?

2.4. gyakorlat: Egy érmével dobunk. Ha az eredmény fej, akkor még kétszer dobunk, ha írás, akkor még egyszer. Mennyi az összes fejdobások számának várható értéke?

2.5. gyakorlat:$^{\mathrm{M}}$ Érmével addig dobunk, amíg először fordul elő két egymás utáni azonos dobás. Mennyi a szükséges dobások számának várható értéke?

2.6. gyakorlat:$^{\mathrm{M}}$ Legyen ξ egy λ paraméterű exponenciális eloszlású valószínűségi változó. Határozzuk meg $2\xi + 3$ sűrűségfüggvényét!

2.7. gyakorlat:$^{\mathrm{M}}$ Legyen ξ és η két olyan valószínűségi változó, hogy (ξ, η) az $(1, 0)$, $(0, 1)$, $(-1, 0), (0, -1)$ értékeket egyenlő valószínűséggekkel veszi fel. Döntsük el, hogy független-e ξ és η!

2.8. gyakorlat: Egy vegyipari termék pH-ja olyan valószínűségi változónak tekinthető, amelynek sűrűségfüggvénye

$$f(x) = \begin{cases} 25(x - 3.8) & \text{ha } 3.8 \leq x \leq 4, \\ -25(x - 4.2) & \text{ha } 4 < x \leq 4.2, \\ 0 & \text{különben.} \end{cases}$$

Mi a valószínűsége, hogy a pH 3.90 és 4.05 közé esik?

2.9. gyakorlat:$^{\mathrm{M}}$ A ξ valószínűségi változó egyenletes eloszlású a $[-2\pi, 2\pi]$ intervallumban. Számítsa ki a $4\xi + 1$, ξ^2, e^{ξ}, $\sin \xi$ valószínűségi változók várható értékét!

2.10. gyakorlat:$^{\mathrm{M}}$ Egy ékszíj (években kifejezett) élettartama olyan valószínűségi változó, amelynek sűrűségfüggvénye

$$f(x) = \begin{cases} 3e^{-3x} & \text{ha } x \geqslant 0, \\ 0 & \text{ha } x < 0. \end{cases}$$

Mi a valószínűsége, hogy egy ékszíj legalább 4 évig használható?

2.11. gyakorlat:$^{\mathrm{M}}$ Szabályos érmével 16-szor dobunk. A Csebisev-egyenlőtlenség segítségével becsülje meg annak a valószínűségét, hogy a fejek száma 2 és 14 közé esik, a határokat is megengedve! Számolja ki a valószínűség pontos értékét is!

2.12. gyakorlat: Egy szakaszt véletlenszerűen két részre osztunk. Számolja ki a kisebbik rész hosszának eloszlásfüggvényét és várható értékét!

2.13. gyakorlat: A (0,0) és (2,0) pontokat összekötő szakaszon véletlenszerűen választunk egy pontot. Mi a választott pont (1,1) ponttól való távolságának sűrűségfüggvénye?

2.14. gyakorlat: Legyen ξ eloszlása n-ed rendű binomiális eloszlás. Számítsa ki ξ várható értékét **(a)** a várható érték definíciója alapján, **(b)** annak felhasználásával, hogy ξ n független egyszerű alternatíva összegére bontható!

2.15. gyakorlat: Számítsa ki az n-ed rendű binomiális eloszlás szórását!

2.16. gyakorlat: $^{\mathbf{M}}$ 400 kg megolvasztott üvegben, amelyből 1 kg-os palackok készülnek, 100 db, hibát okozó szilárd részecske van. Mi a valószínűsége, hogy egy palackba legalább két, hibát okozó részecske kerül?

2.17. gyakorlat: $^{\mathbf{M}}$ Egy ötvözet magnézium tartalma olyan valószínűségi változó, amelynek sűrűségfüggvénye

$$f(x) = \begin{cases} \frac{x}{18} & \text{ha } 0 \leq x \leq 6 \\ 0 & \text{különben} \end{cases}$$

Az ötvözet gyártásával elérhető haszon a $k(\xi) = 10 + 2\xi$ függvénnyel írható le. Adja meg a k függvény valószínűség eloszlását! Mi lesz a haszon varható értéke?

2.18. gyakorlat: $^{\mathbf{M}}$ A ξ valószínűségi változó sűrűségfüggvénye:

$$f_\xi(x) = \begin{cases} 2\beta^{-2}xe^{-x^2/\beta^2} & \text{ha } x > 0, \\ 0 & \text{ha } x \leq 0. \end{cases}$$

Mi ξ eloszlásfüggvénye?

2.19. gyakorlat: $^{\mathbf{M}}$ ξ és η független valószínűségi változók, mindkettő varianciája 3. Számítsa ki $2\xi - 3\eta + 1$ szórását!

2.20. gyakorlat: Egy szórakozott titkárnő legépel n levelet és megcímezi hozzájuk az n borítékot. Ezután véletlenszerűen helyezi a leveleket borítékba (minden borítékba csak egyet). Mi a várható értéke a korrektül borítékolt levelek számának?

2.21. gyakorlat: $^{\mathbf{M}}$ Addig ismételjük az érmedobást, amíg pontosan k-szor jelenik meg a fej. Mi a dobások számának várható értéke?

2.22. gyakorlat: $^{\mathbf{M}}$ Számítógépünk véletlenszám-generátora a $[0, 1]$ intervallumban egyenletes eloszlással hozza létre a ξ véletlen számot. Nekünk egy olyan η véletlen számra van szükségünk, amelynek sűrűségfüggvénye

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

Milyen r függvényt kell használnunk ahhoz, hogy η-t $r(\xi)$ alakban kaphassuk meg?

2.23. gyakorlat: $^{\mathbf{M}}$ Az exponenciális eloszlású ξ valószínűségi változó $1/2$ valószínűséggel vesz fel 3-nál nagyobb értéket. **(a)** Mi ξ várható értéke? **(b)** Mi a valószínűsége, hogy ξ 1-nél kisebb értéket vesz fel?

2.24. gyakorlat: $^{\mathbf{M}}$ A ξ valószínűségi változó egyenletes eloszlású a $[-1, 1]$ intervallumban. Mi a valószínűsége, hogy első tizedesjegye 4-es?

2.25. gyakorlat: Szabályos dobókockával addig dobunk, amíg másodszor kapunk páratlan számot. Mennyi a szükséges dobások számának várható értéke?

Harmadik fejezet

Nevezetes valószínűségi változók

Valamely valószínűségi változóhoz kapcsolódó kérdésekre akkor tudunk pontos választ adni, ha a változó eloszlása ismert, vagy megközelítőleg ismert. Ebben a fejezetben a leggyakrabban előforduló valószínűségeloszlásokat tekintjük át. Megjegyezzük, hogy a matematikai statisztika irodalma jóval több valószínűségeloszlást tart nyilván. Természetesen minden eloszlás nem adható meg, de elegendően sok eloszlás ismeretében jó közelítések kaphatók szinte minden eloszlásra, illetve van olyan eloszlás (a gamma eloszlás), amelynek paramétereit alkalmasan megválasztva a legtöbb eloszlás jól közelíthetõ.

3.1. A Poisson-eloszlás

A Poisson-eloszlás egyike a mérnöki gyakorlatban és az informatikában leggyakrabban megjelenő valószínűség-eloszlásoknak. Ennek ellenére az előző fejezetben tárgyalt eloszlásokhoz képest csak jóval összetettebb matematikai modellekben mutatható be. Legegyszerűbben a binomiális eloszlásból kiindulva juthatunk el a Poisson-eloszláshoz.

Bizonyos mennyiségű nyersanyagból m számú terméket készítenek. Az összes nyersanyagban n számú szennyeződésszemcse van. Mi a valószínűsége, hogy egy termékbe pontosan k szemcse kerül? (Ez a kérdés például akkor válhat lényegessé, amikor bizonyos számú szennyeződésszemcsét tartalmazó termék selejtnek tekintendő.) Leegyszerűsítve így fogalmazhatunk: n számú szemcsét kell elhelyezni m dobozban. Mi a valószínűsége, hogy egy dobozba k számú szemcse jut? Annak a valószínűsége, hogy egy adott szemcse egy adott dobozba kerül $p = 1/m$. Így a binomiális eloszlás képletét használva

$$\binom{n}{k} p^k (1-p)^{n-k}$$

annak a valószínűsége, hogy k szemcse jut egy dobozba. Ez a valószínűség egy hosszabb gyártási periódusban érdekes, amikor m nagyon nagy, de az egy termékre jutó szennyeződésszemcsék n/m száma nem változik. Legyen $\lambda = n/m$. A

$$\lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

határérték kiszámításához kiírjuk a binomiális együtthatót:

$$\frac{n(n-1)(n-2)\dots(n-k+1)}{n^k}\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}$$

Itt az első tört és az utolsó tényező egyhez tart, a középső (n-től függő) tényező pedig $e^{-\lambda}$-hoz egy nevezetes határértéktétel szerint. Tehát:

$$\boxed{\binom{n}{k} p^k(1-p)^{n-k} \to \frac{\lambda^k}{k!}e^{-\lambda}.} \tag{3.1.1}$$

λ **paraméterű Poisson-eloszlásnak** nevezzük a

$$\boxed{P(\xi = k) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad \text{ahol} k = 0, 1, \dots n \text{és} \lambda > 0} \tag{3.1.2}$$

valószínűségeloszlást. Kiszámolható, hogy λ az eloszlás várható értéke.

Amit így kaptunk, az a Poisson-féle határértéktétel:

3.1. tétel: (Poisson-féle határértéktétel) *A p sikervalószínűségű binomiális eloszlás Poisson-eloszláshoz tart, ha az alternatíva ismétlésének n száma tart a végtelenhez és az np várható érték eközben állandó marad.*

Más szóval, ha n nagy p-hez képest, akkor a binomiális eloszlás Poisson-eloszlással közelíthető.

A Poisson-eloszlásra vonatkozó több kérdésre a **generátorfüggvény** módszere hatékonyan használható. Ha egy diszkrét valószínűségi változó csak pozitív egész értékeket vesz fel, akkor generátorfüggvénye

$$G(z) = \sum_{n=1}^{\infty} P(\xi = n)z^n \tag{3.1.3}$$

hatványsorral adott, feltételezve, hogy a G összegfüggvény létezik. A Poisson-eloszlás generátorfüggvénye

$$G(z) = e^{\lambda(z-1)} \tag{3.1.4}$$

minden $z \in \mathbb{R}$ esetén értelmezve van.

3.1. példa: Használjuk a generátorfüggvényt a Poisson-eloszlás várható értékének és szórásnégyzetének kiszámítására!

(3.1.3)-ból differenciálással, és $z = 1$ helyettesítéssel kapjuk, hogy

$$G'(1) = \sum_{n=1}^{\infty} P(\xi = n)n = M(\xi)\,, \tag{3.1.5}$$

és ismételt differenciálással

$$G''(1) = \sum_{n=1}^{\infty} P(\xi = n)n(n-1) = M(\xi^2) - M(\xi).$$

Ebből

$$\boxed{\sigma(\xi)^2 = G''(1) + G'(1) - (G'(1))^2\,.} \tag{3.1.6}$$

Most alkalmazzuk a (3.1.5) és (3.1.6) általános képleteket a Poisson-eloszlásra:

$$G'(z) = \lambda e^{\lambda(z-1)}, \qquad G''(z) = \lambda^2 e^{\lambda(z-1)}.$$

□

A λ szám a Poisson-eloszlás várható értéke és egyben szórásnégyzete.

A következő, a gyakorlatban nagy jelentõségû tétel bizonyítása ismét a generátorfüeggvény–módszeren alapul. Azt fogjuk felhasználni, hogy független, egész értékeket felvevő valószínűségi változók összegének generátorfüggvénye az összeadandók generátorfüggvényének a szorzata.

3.2. tétel: *Ha ξ_1 és ξ_2 λ_1 ill. λ_2 paraméterű Poisson-eloszlású független valószínűségi változók, akkor $\xi_1 + \xi_2$ ugyancsak Poisson-eloszlású, és paramétere $\lambda_1 + \lambda_2$.*

Bizonyítás: ξ_i generátorfüggvénye

$$G_i(z) = e^{\lambda_i(z-1)} \qquad (i = 1, 2)$$

(3.1.4) szerint. $\xi_1 + \xi_2$ generátorfüggvénye $G = G_1 \cdot G_2$, azaz

$$G(z) = e^{\lambda_1(z-1)} e^{\lambda_2(z-1)} = e^{(\lambda_1+\lambda_2)(z-1)}.$$

Ez pedig éppen egy $\lambda_1 + \lambda_2$ paraméterű Poisson-eloszlás generátorfüggvénye. □

3.2. Az exponenciális és a gamma-eloszlás

Az exponenciális eloszlás már az előző fejezetben is felbukkant, sűrűségfüggvénye $\lambda \exp(-\lambda x)$, ahol λ pozitív paraméter. Az exponenciális eloszlás várható értéke λ^{-1}.

A tipikus exponenciális eloszlású valószínűségi változó egy olyan véletlen időtartam, amely ha egy x időpontig nem ért véget, akkor úgy tekinthető, mintha az egész folyamat csak az x időpontban kezdődött volna:

$$P(\xi \leqslant x + y | \xi \geqslant x) = P(\xi \leqslant y) \tag{3.2.1}$$

Ez mindig teljesül egy exponenciális eloszlású ξ valószínűségi változóra. Szavakkal úgy fejezzük ki, hogy az **exponenciális eloszlásnak nincsen emlékezete**. (3.2.1) bizonyításához a valószínűségeket ki kell fejezni az eloszlásfüggvény segítségével:

$$\begin{aligned} P(\xi \leqslant x + y | \xi \geqslant x) &= \frac{P(x \leqslant \xi \leqslant x + y)}{P(x \geqslant \xi)} = \frac{F_\xi(x+y) - F_\xi(x)}{1 - F_\xi(x)} \\ &= \frac{e^{-\lambda x} - e^{-\lambda(x+y)}}{e^{-\lambda x}} = 1 - e^{-\lambda y} = F_\xi(y) = P(\xi \leqslant y). \end{aligned}$$

Bizonyítható, hogy csak az exponenciális eloszlásnak nincs emlékezete a sűrűségfüggvénnyel rendelkező folytonos eloszlások között.

Az exponenciális eloszlás befoglalható a **gamma-eloszlások** családjába. Legyen α és β pozitív paraméter. A gamma-eloszlás a gamma-függvényről kapta a nevét, mivel definíciójában a gamma függvény szerepel. A gamma függvényre vonatkozó legfontosabb ismeretek a függelékben megtalálhatók.

Az α-drendű gamma-eloszlás sűrűségfüggvénye

$$\boxed{f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & \text{ha } x > 0, \\ 0 & \text{különben.} \end{cases}} \tag{3.2.2}$$

Kiszámolható, hogy a gamma-eloszlás várható értéke $\alpha\beta$ és varianciája $\alpha\beta^2$. Az $\alpha = 1$ választás vezet az exponenciális eloszláshoz. A gamma-eloszlás momentumgeneráló függvénye $t > \beta$ esetén van értelmezve és

$$m(t) = (1-\beta t)^{-\alpha} \qquad (t > \beta). \tag{3.2.3}$$

3.3. tétel: *Ha a $\xi_1, \xi_2, \ldots, \xi_n$ valószínűségi változók függetlenek és ξ_i gamma-eloszlású α_i és β paraméterrel $(1 \leqslant i \leqslant n)$, akkor $\xi_1+\xi_2+\ldots+\xi_n$ ugyancsak gamma-eloszlású, $\alpha_1+\alpha_2+\ldots+\alpha_n$ és β paraméterrel.*

Bizonyítás: A momentumgeneráló függvény módszerét használjuk. (3.2.3) szerint ξ_i momentum generáló függvénye

$$m_i(t) = (1-\beta t)^{-\alpha_i}$$

és a 2.4. tétel szerint $\xi_1 + \ldots + \xi_n$ momentumgeneráló függvénye

$$m(t) = \prod_{i=1}^{n} m_i(t) = (1-\beta t)^{-(\alpha_1+\alpha_2+\ldots+\alpha_n)}$$

amiből látszik, hogy egy gamma-eloszláshoz tartozik, és a paraméterek kiolvashatók. □

3.3. A normális eloszlás

A ξ valószínűségi változót **normális**, avagy $N(m,\sigma)$-eloszlásúnak nevezzük, ha sűrűségfüggvénye

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}} \tag{3.3.1}$$

alakú, ahol $-\infty < m < \infty$ és $\sigma > 0$. Mivel $f(m-x) = f(m+x)$, az $N(m,\sigma)$ sűrűségfüggvény szimmetrikus m-re, és m a várható értéke. Parciális integrálással számolható ki, hogy $N(m,\sigma)$ szórása éppen σ.

Ha $m = 0$ és $\sigma = 1$, akkor **standard normális eloszlásról** beszélünk. Ennek eloszlásfüggvénye

$$\boxed{\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt} \tag{3.3.2}$$

A $\Phi(x)$ értékeket táblázatból vehetjük, mivel az integrál nem adható meg könnyen kezelhető képlettel. A táblázat csak pozitív x-ekre tartalmazza $\Phi(x)$-et, negatív x értékekre a $\Phi(-x) = 1-\Phi(x)$ összefüggést használjuk. A táblázatból látható, de fejben tartani is érdemes, hogy a standard normális eloszlás 99% valószínűséggel -3 és 3 között veszi fel értékeit.

3.4. tétel: *Ha a ξ valószínűségi változó $N(m,\sigma)$ eloszlású, akkor*

$$\eta = \frac{\xi - m}{\sigma}$$

standard normális eloszlású. Nevezetesen

$$P(a < \xi < b) = P\Big(\frac{a-m}{\sigma} < \eta < \frac{b-m}{\sigma}\Big) = \Phi\Big(\frac{b-m}{\sigma}\Big) - \Phi\Big(\frac{a-m}{\sigma}\Big)$$

3.2. példa: Legyen a ξ valószínűségi változó $N(3,2)$ eloszlású. Mekkora legyen az A szám ahhoz, hogy a $(2,A)$ intervallumba $1/2$ valószínűséggel essenek ξ értékei?

Ha F a ξ valószínűségi változó eloszlásfüggvénye, akkor

$$\begin{aligned} P(2 \leqslant \xi \leqslant A) &= F(A) - F(2) \\ &= \Phi\Big(\frac{A-3}{2}\Big) - \Phi\Big(\frac{2-3}{2}\Big) = \Phi\Big(\frac{A-3}{2}\Big) - 1 + \Phi\Big(\frac{1}{2}\Big), \end{aligned}$$

és a

$$\Phi\Big(\frac{A-3}{2}\Big) = \frac{3}{2} - \Phi\Big(\frac{1}{2}\Big)$$

egyenlethez jutunk, amit Φ táblázata segítségével oldunk meg: $(A-3)/2 = \Phi^{-1}(0.8085) = 0.8725$, amiből $A = 4.745$. $\square$

3.3. példa: Egy gyártó 1000 Ft-os egységáron árulja termékeit. Ha egy termék 80 g-nál kisebb, akkor eladhatatlan, és teljes veszteséget jelent. A termékek tömege normális eloszlást mutat w_0 várható értékkel és 10 g szórással. Egy termék előállítási költsége $c = 5w + 30$, ahol w a termék súlya. Milyen átlagos w_0 súly maximalizálja a profitot?

Annak a valószínűsége, hogy egy termék eladhatatlan

$$p = \Phi\Big(\frac{w-80}{10}\Big),$$

és $(1-p)1000 - (5w+30)$ a bevétel, amelynek várható értéke a maximalizálandó. A $970 - 1000p - 5w_0$ várható érték maximumát differenciálással határozzuk meg:

$$1000\frac{d}{dw_0}\Phi\Big(\frac{w_0-80}{10}\Big) = -100 \quad \Phi'\Big(\frac{w_0-80}{10}\Big) = -5$$

(összetett függvényt kellett differenciálni). Az egyenlet két gyöke közül $w_0 = 109.6$ ad csak maximumot, a feladat természetéből adódóan. $\square$

3.5. tétel: (Moivre–Laplace–tétel) *Legyen ξ_n $0,1,2,\ldots,n$ értékeket felvevő valószínűségi változó, amelynek eloszlása binomiális p sikervalószínűséggel. Ekkor a*

$$\xi_n^* = \frac{\xi_n - np}{\sqrt{npq}}$$

valószínűségi változók eloszlása a standard normális eloszláshoz tart, amikor $n \to \infty$.

A tételt a gyakorlatban a binomiális eloszlás közelítésére használjuk a következő alakban:

$$\boxed{P(a \leqslant \xi_n \leqslant b) \approx \Phi\Big(\frac{b-np}{\sqrt{npq}}\Big) - \Phi\Big(\frac{a-np}{\sqrt{npq}}\Big)} \tag{3.3.3}$$

Ez a közelítés akkor alkalmazható, ha np és nq 5-nél nagyobb.

3.4. példa: Egy hallgatónak 20 tesztkérdésre kell igennel vagy nemmel válaszolnia, és $p = 50\,\%$-os valószínűséggel ad helyes választ. Mi a valószínűsége, hogy legalább 15 kérdésre ad helyes választ?

$np = nq = 10 > 5$, így alkalmazhatjuk a (3.3.3) közelítést. $P(15 \leqslant \xi) = 1 - 0.989 = 1.1\%$ a meglepően alacsony esély. $\square$

3.6. tétel: *Az $N(m,\sigma)$ normális eloszlású valószínűségi változó momentumgeneráló függvénye*

$$\boxed{m(t) = \exp(mt + \sigma^2 t^2/2)\,.} \tag{3.3.4}$$

Bizonyítás: A momentumgeneráló függvény definíciója szerint az

$$\begin{aligned} m(t) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{tx} e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \exp\left[(-\sigma^{-2}/2)(x^2 - 2mx + m^2 - 2\sigma^2 tx)\right] dx. \end{aligned}$$

A kitevőben teljes négyzetté való kiegészítés után azt kapjuk, hogy

$$m(t) = e^{mt+t^2\sigma^2/2}\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \exp\Big(-\frac{(x-(m+t\sigma^2))^2}{2\sigma^2}\Big)\,dx\,.$$

Az integrál értéke $\sqrt{2\pi}\sigma$, és így a bizonyítandó képlethez jutottunk. □

3.7. tétel: *Legyen ξ_i egy $N(m_i,\sigma_i)$ normális eloszlású valószínűségi változó $(i = 1,2)$. Ha ξ_1 és ξ_2 függetlenek, akkor $\xi_1 + \xi_2$ is normális eloszlású, $m = m_1 + m_2$ várható értékkel és $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$ szórással.*

Bizonyítás: Valóban, az $m_i(t) = \exp(m_i t + \sigma_i^2 t^2/2)$ momentumgeneráló függvények szorzata

$$\exp\Big(m_1 t + m_2 t + \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}\Big)$$

a 2.4. tétel alapján $\xi_1 + \xi_2$ momentumgeneráló függvénye. Belőle a normális eloszlás paraméterei kiolvashatók. □

A szórás a valószínűségi változók ingadozásának leggyakrabban használt mértéke. Egy valószínűségi változó bizonytalanságának – főként az információelméletben használt – másik mértéke az **entrópia**. ξ entrópiáját az

$$\boxed{-\int_{-\infty}^{\infty} f_\xi(x)\ln f_\xi(x)\,dx = H(\xi)} \tag{3.3.5.}$$

integrállal lehet megadni. Könnyű kiszámolni az $N(m,\sigma)$ eloszlás entrópiáját, ami az

$$\ln\sqrt{2\pi e} + \ln\sigma \tag{3.3.6}$$

értéknek adódik. (A képletből látszik, hogy normális eloszlásra az entrópia bármilyen valós értéket felvehet.) A nagy szórás nagy entrópiát, azaz nagy bizonytalanságot jelent. Az $N(m,\sigma)$ normális eloszlás nevezetes tulajdonsága, hogy a σ szórású és m várható értékű eloszlások között a legnagyobb az entrópiája, azaz minden más σ szórású m várható értékű eloszlás entrópiája (3.3.6)-nál kisebb.

3.4. A logaritmikus normális eloszlás

Egy pozitív értékeket felvevő ξ valószínűségi változót **logaritmikus normális eloszlásúnak** (vagy lognormális eloszlásúnak) mondunk, ha a $\eta = \ln\xi$ valószínűségi változó normális eloszlású. Tételezzük fel, hogy η eloszlása $N(m, \sigma)$. Ekkor az eloszlásfüggvény definíciója szerint

$$F_\xi(x) = P(\xi < x) = P(\eta < \ln x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\ln x} e^{-\frac{(t-m)^2}{2\sigma^2}}\, dt.$$

Ebből differenciálással adódik, hogy

$$\boxed{f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\Big(- \frac{(\ln x - m)^2}{2\sigma^2}\Big) \qquad (x > 0).} \tag{3.4.1}$$

Az irodalomban m helyett α és σ helyett β paraméter is szokott szerepelni. Az m és σ betűk használata azonban azért célszerű, mert utalnak a megfelelő normális eloszlásra.

Integrálással kiszámolható ξ várható értéke és szórása:

$$M(\xi) = e^{m+\sigma^2/2}, \qquad \sigma(\xi)^2 = e^{2m+\sigma^2}(e^{\sigma^2} - 1) \tag{3.4.2}$$

A lognormális eloszlás nevezetes tulajdonsága, hogy ha ξ lognormális eloszlású, akkor a $a\xi^b$ valószínűségi változó is lognormális eloszlású tetszőleges pozitív a és b számokra.

3.5. A khi-négyzet és a khi-eloszlás

n számú független standard normális eloszlású valószínűségi változó négyzetösszegének eloszlását n**-szabadságfokú** χ^2**-eloszlásnak** nevezik. A χ^2-eloszlás sűrűségfüggvénye

$$f_n(x) = \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma(n/2)} \qquad (x > 0) \tag{3.5.1}$$

(A nevezőben előforduló gamma függvény értelmezése és tulajdonságai a függelékben találhatók meg.)

3.8. tétel: *Az n-szabadságfokú χ^2-eloszlás momentumgeneráló függvényét, várható értékét és szórását a következő képletekkel adják meg:*

$$m(t) = (1 - 2t)^{-n/2}, \qquad M = n, \qquad \sigma = \sqrt{2n}\,. \tag{3.5.2}$$

Az η valószínűségi változó eloszlását χ**-eloszlásnak** nevezzük, ha az η^2 valószínűségi változó χ^2-eloszlású. Ez az eloszlás a matematikai statisztikában játszik fontos szerepet. Az n-szabadságfokú χ-eloszlás sűrűségfüggvénye

$$f_n(x) = \frac{2x^{n-1} e^{-x^2/2}}{2^{n/2}\Gamma(n/2)} \qquad (x > 0) \tag{3.5.3}$$

és várható értéke

$$M = \sqrt{2}\frac{\Gamma((n+1)/2)}{\Gamma(n/2)}\,. \tag{3.5.4}$$

Megjegyezzük, hogy a 4.6. példa tartalmazza a két szabadságfokú χ-eloszlás sűrűségfüggvényének levezetését. A következő példában rámutatunk a Maxwell-eloszlás (lásd a 2.13. példát) és a három szabadságfokú χ-eloszlás kapcsolatára.

3.5. példa: A kinetikus gázelmélet szerint egy nyugalomban lévő gáz egy molekulájának egymásra merőleges x, y és z irányú sebességkomponensei függetlenek, és $N(0,\sigma)$ eloszlású valószínűségi változónak tekinthetők. Legyen ξ_x, ξ_y és ξ_z a három sebességkomponens. A sebességvektor

$$v = \sqrt{\xi_x^2 + \xi_y^2 + \xi_z^2}$$

hosszának sűrűségfüggvényét a Maxwell-féle sebességeloszlási törvény adja meg, amit most levezetünk.

A ξ_x/σ, ξ_y/σ és ξ_z/σ valószínűségi változók standard normális eloszlásúak, így v/σ három szabadságfokú χ-eloszlású valószínűségi változó. A χ-eloszlás sűrűségfüggvényéből

$$f_{v/\sigma}(x) = \frac{2x^2 e^{-x^2/2}}{2^{3/2}\Gamma(3/2)} = \frac{2x^2 e^{-x^2/2}}{\sqrt{2\pi}}$$

és

$$f_v(x) = \sigma f_{v/\sigma}(\sigma x) = \frac{2\sigma^3 x^2 e^{-x^2\sigma^2/2}}{\sqrt{2\pi}}$$

(2.3.11) α paramétere és az utóbbi képlet σ-ja között az $\alpha = \sqrt{2}/\sigma$ összefüggés áll fenn. Ezért $\sigma^2 = m/(kT)$.

□

Gyakorlatok

3.1. gyakorlat: Számolja ki az $N(0,\sigma)$ eloszlás összes momentumát!

3.2. gyakorlat:$^{\mathrm{M}}$ Egy mosógép élettartama normális eloszlást követ 6.4 év várható értékkel és 2.3 év szórással. Tudva azt, hogy egy gép 5 évig nem hibásodott meg, mi a valószínűsége annak, hogy még további 3 évig használható?

3.3. gyakorlat:$^{\mathrm{M}}$ Legyen ξ egy λ paraméterű Poisson- eloszlású valószínűségi változó. Számolja ki $1/(\xi+1)$ várható értékét.

3.4. gyakorlat: Igazolja, hogy (3.2.2) valóban egy sűrűségfüggvényt értelmez!

3.5. gyakorlat: A 2.13. példában közölt Maxwell-féle sebességeloszlásból számolja ki egy olyan gáz molekuláinak átlagos kinetikus energiáját, amelynek abszolút hőmérséklete T, és amely m tömegű molekulákból áll!

3.6. gyakorlat:$^{\mathrm{M}}$ Hasonlítsa össze az 1 szórású egyenletes és normális eloszlások entrópiáját!

3.7. gyakorlat:$^{\mathrm{M}}$ Legyen η gamma-eloszlású valószínűségi változó. Mi $c\eta$ sűrűségfüggvénye, ha $c > 0$?

3.8. gyakorlat:$^{\mathbf{M}}$ A ξ valószínűségi változó sűrűségfüggvénye a következő:

$$f(x) = \begin{cases} \frac{a}{b^a} x^{a-1} e^{-(x/b)^a} & \text{ha } x > 0, \\ 0 & \text{ha } 0 < 0 \end{cases}.$$

(Itt a és b pozitív paraméterek és ξ-t ilyenkor Weibull-féle eloszlásúnak mondjuk.) Milyen eloszlású a ξ^a valószínűségi változó?

3.9. gyakorlat:$^{\mathbf{M}}$ Egy valószínűségi változó olyan lognormális eloszlású, amelynek paraméterei $m = -2$ és $\sigma = 2$. Mi a valószínűsége, hogy a valószínűségi változó az értékét 4.7 és 9.8 között veszi fel?

3.10. gyakorlat: A Poisson-eloszlás diszkrét eloszlás, tehát nincs sűrűségfüggvénye. Lehetne-e valahogy mégis folytonos görbével szemlélteni?

3.11. gyakorlat:$^{\mathbf{M}}$ Igazolja az exponenciális eloszlás példáját követve, hogy a geometriai eloszlásnak nincsen ,,emlékezete" !

3.12. gyakorlat:$^{\mathbf{M}}$ Egy nagy népességben az emberek 20 %-a balkezes. Ha a népességből nagy számú mintát ($n = 10000$) vizsgálunk, mi annak a valószínűsége, hogy

a.) legalább 2100 ember balkezes,

b.) legalább 1960, de nem több, mint 2040 ember balkezes

3.13. gyakorlat: Állatokkal végzett kísérletben valamely állat egy adott feladatot p valószínűséggel tud végrehajtani, $g = 1 - p$ valószínűséggel a probálkozas sikertelen. Átlagosan hány próbálkozás szükséges ahhoz, hogy k esetben sikert érjünk el?

3.14. gyakorlat:$^{\mathbf{M}}$ Egy adott mohafaj spóráinak megszámlálásához N számú talajfúrást végeznek. A megfigyelések eredményét a következő táblázat foglalja össze, ahol k a spóraszám, f_k a k-hoz tartozó fúrások gyakorisága.

$k=$	0,	1,	2,	3,	> 3
$f_k=$	59,	27,	9,	1,	0

Jogosan feltételezhetjuk-e, hogy a spóraeloszlás a talajban Poisson-eloszlással jellemezhető?

3.15. gyakorlat:$^{\mathbf{M}}$ Kísérletek alapján azt állíthatjuk, hogy egy adott tóban a baktériumok eloszlása Poisson-típusú és 2 baktérium/cm^3 sűrűségű. Mi annak a valószínűsége, hogy egy 2 cm^3 nagyságú minta

a.) baktériummentes

b.) legalább két baktériumot tartalmaz?

3.16. gyakorlat:$^{\mathbf{M}}$ Egy rovarfaj térbeli diffúziójának vizsgálatára kísérletet végeznek. Egy adott A pontban nagyszámú egyedet bocsátanak szabadon, és azt tapasztalják, hogy az egyedek e^{-2r} hányada jut el adott idő alatt A-tól legalább r távolságra.

a.) Határozza meg, hogy az egyedek hányad része vándorolt el A-tól 1 m-nél nagyobb távolságra!

b.) A vizsgált egyedek átlagosan milyen messze kerültek a kiindulási ponttól?

3.17. gyakorlat:[M] Két különböző forrásból származó inger erőssége egymástól független, normális eloszlású valószínűségi változó m_1 és σ_1, illetve m_2 és σ_2 paraméterekkel. A két inger azonos időben éri a receptort és hatásuk összegződik. Mi a valószínűsége annak, hogy a két inger együttes hatása eléri a receptor x_0 ingerküszöbét.

Negyedik fejezet

Többdimenziós eloszlások

Több valószínűségi változó együttes vizsgálatához nem elegendő az egyes változók eloszlásának ismerete. Ez a tény jól érzékelhető a következő hétköznapi életből vett pédából. Tegyük fel, hogy ismerjük a magyar személygépkocsik kor szerinti eloszlását, és szín szerinti eloszlását. E két adat alapján nem tudjuk megmondani, hogy mi a valószínűsége, hogy egy 5 éves autó X színű. Lehetséges, hogy 5 évvel ezelőtt X a legnagyobb divatszín volt, így ez a valószínűség meglehetősen nagy, de az is lehet, hogy az X szín 5 évvel ezelőtt elő sem fordult. Tehát két véletlen mennyiségre vonatkozó kérdések nem válaszolhatók meg az egyes változók eloszlása ismeretében, a sztochasztikus kapcsolatot maga a két eloszlás nem tartalmazza. Egy elvontabb, de ugyanilyen következtetésre vezető példa: Ha ξ és η jelöli a két valószínűségi változót, akkor ξ és η eloszlásfüggvénye semmit nem mond például a

$$P(a \le \xi \le b,\, c \le \eta \le d)$$

valószínűségről, ami az ún. együttes eloszlás segítségével adható meg. Az együttes eloszlás az a fogalom, amely különböző valószínűségi változók egymással való sztochasztikus kapcsolatát kifejezi.

4.1. A többdimenziós eloszlás- és sűrűségfüggvény

A ξ és η valószínűségi változók **együttes eloszlása** az a kétváltozós $F : \mathbb{R}^2 \to \mathbb{R}_0^+$ függvény, amelyre

$$F(x,y) = P(\xi < x, \eta < y) \qquad (x, y \in \mathbb{R}) \tag{4.1.1}$$

Az F eloszlásfüggvény mindkét változójában növekedő és balról folytonos. Amennyiben ξ és η független, akkor $P(\xi < x, \eta < y) = P(\xi < x)P(\eta < y)$, és $F(x,y) = F_\xi(x)F_\eta(y)$, tehát két független valószínűségi változó együttes eloszlása egyszerűen a két eloszlásfüggvény szorzata.

Ha létezik olyan $f : \mathbb{R}^2 \to \mathbb{R}_0^+$ függvény, amelyre

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,dv\,du, \tag{4.1.2}$$

akkor ezt ξ és az η **együttes sűrűségfüggvényének** nevezzük. (Szokás azt is mondani, hogy f a (ξ,η) valószínűségi vektorváltozó sűrűségfüggvénye.) (4.1.2) megfordítása a

$$\frac{\partial^2}{\partial x \partial y} F(x,y) = f(x,y) \tag{4.1.3}$$

formula.

Ha a ξ és η valószínűségi változóknak van együttes sűrűségfüggvénye, akkor ξ-nek és η-nak külön-külön is van f_ξ és f_η sűrűségfüggvénye.

$$\boxed{f_\xi(x) = \int_{-\infty}^{\infty} f(x,y)\,dy, \qquad f_\eta(y) = \int_{-\infty}^{\infty} f(x,y)\,dx.} \tag{4.1.4}$$

Az f_ξ és f_η sűrűségfüggvényeket a (ξ,η) valószínűségi vektorváltozó **peremsűrűségeinek** nevezzük. Az f_ξ-re vonatkozó képletet a következőképpen láthatjuk be. Ha (4.1.2)-ben végrehajtjuk az $y \to \infty$ határátmenetet, akkor a

$$F_\xi(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u,v)\,dv\,du$$

formulát kapjuk. Ezt x szerint differenciálva a bal oldalon ξ sűrűségfüggvénye, a jobb oldalon pedig az együttes sűrűségfüggvény második változó szerinti integrálja adódik, (4.1.4)-nek megfelelően.

A (4.1.1) és (4.1.2) képletek kombinációja a

$$P(\xi < x, \eta < y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\,dv\,du$$

formula, ami nagy mértékben általánosítható. A bal oldalon annak a valószínűsége áll, hogy a (ξ,η) véletlen pontja a síknak az $A = \{(u,v) \in \mathbb{R}^2 : u < x, v < y\}$ részhalmazába esik, a jobb oldal pedig az együttes sűrűségfüggvénynek ezen halmazon vett kettős integrálja. Ez az összefüggés nem csak erre a speciális negyed-síkra, hanem a síknak bármilyen A részhalmazára teljesül:

$$\boxed{P((\xi,\eta) \in A) = \iint_A f(u,v)\,du\,dv} \tag{4.1.5}$$

Az eddigiekben az egyszerűség kedvéért csak két valószínűségi változó együttes eloszlás- és sűrűségfüggvényével foglalkoztunk. A definíciók és formulák természetes módon akárhány valószínűségi változóra is kiterjeszthetők. Ha $\xi_1, \xi_2, \ldots, \xi_n$ n darab valószínűségi változó, akkor az együttes sűrűségfüggvényük n-változós függvény, $x_1, x_2, \ldots, x_n$ változókkal, és például (4.1.3)-ban n-szeres deriválás, (4.1.4)-ben egy $(n-1)$-szeres, (4.1.5)-ben egy n-szeres integrálás lesz. Kettőnél több valószínűségi változó főleg a normális eloszlással kapcsolatban fog a továbbiakban szerepelni.

4.1. példa: Tételezzük fel, hogy két véletlen történés következik be egy időtartam, például egy óra alatt. A ξ valószínűségi változó jelenti egy bizonyos szállítmány beérkezésének az idejét, η pedig annak a szállító vállalathoz intézett telefonhívásnak az idejét, amelyben a szállítmányról érdeklődnek. Tételezzük fel, hogy mind ξ mind η egyenletes eloszlású. Ennek alapján az együttes eloszlásuk nem meghatározott, tehát további feltevésre van szükség. Legyen ξ és η egymástól független. Ekkor az együttes sűrűségfüggvényük:

$$f(x,y) = \begin{cases} 1 & \text{ha } 0 \le x, y \le 1, \\ 0 & \text{egyébként.} \end{cases}$$

A sűrűségfüggvény tartója – az a halmaz, ahol nullától különböző értékeket vesz fel – tehát az egységnégyzet. Az együttes sűrűség ismeretében számos ξ-re és η-ra vontatkozó kérdésre válaszolhatunk. Például: mi a valószínűsége, hogy $\eta < \xi$?

Annak a valószínűségét számoljuk tehát ki, hogy azelőtt telefonálnak, hogy a szállítmány beérkezett.

$$P(\eta < \xi) = \int_{-\infty}^{\infty} \int_{-\infty}^{x} f(x,y)\, dy\, dx = \int_0^1 \int_0^x 1\, dy\, dx = \frac{1}{2}. \qquad \square$$

Az előző példában az együttes sűrűségfüggvény állandó a tartóján. Általában azt mondjuk, hogy a ξ és η valószínűségi változók **együttesen egyenletes eloszlásúak** az $A \subset \mathbb{R}^2$ halmazon, ha együttes sűrűségfüggvényük

$$f(x,y) = \begin{cases} c & \text{ha } (x,y) \in A, \\ 0 & \text{különben.} \end{cases}$$

Mivel egy sűrűségfüggvény integrálja 1, a c állandó az A halmaz területének reciproka. Figyelmeztetés: Két együttesen egyenletes eloszlású valószínűségi változó nem mindig független!

4.2. példa: Egy berendezés A és B részegységének meghibásodási idejét a ξ és η valószínűségi változó fejezi ki. Együttes eloszlásuk

$$f(x,y) = \begin{cases} 2e^{-(x+2y)} & \text{ha } x, y \geq 0, \\ 0 & \text{különben.} \end{cases}$$

Mi annak a valószínűsége, hogy

- (i) A és B legalább egy évig nem mondja fel a szolgálatot,
- (ii) B előbb megy tönkre mint A?

Az (i) esetben a $P(\xi \geq 1, \eta \geq 1)$, (ii)-ben $P(\eta < \xi)$ valószínűség a kérdés. Egyszerű számolással:

$$\begin{aligned} P(\xi \geq 1, \eta \geq 1) &= \int_1^{\infty} \int_1^{\infty} 2e^{-(x+2y)}\, dx\, dy \\ &= \int_1^{\infty} e^{-x}\, dx \int_1^{\infty} 2e^{-2y}\, dy = e^{-1}e^{-2} = e^{-3} = 0.05, \\ P(\eta \leq \xi) &= \int_0^{\infty} \int_0^{x} 2e^{-(x+2y)}\, dy\, dx = \int_0^{\infty} e^{-x} dx \int_0^x 2e^{-2y}\, dy \\ &= \int_0^{\infty} e^{-x}(1 - e^{-2x}) dx = \frac{2}{3}. \qquad \square \end{aligned}$$

4.3. példa: A ξ és η valószínűségi változók együttes sűrűségfüggvénye a következő:

$$h(x,y) = \begin{cases} 2 & \text{ha } 0 \leq x \leq 1 \text{ és } 0 \leq y \leq x, \\ 0 & \text{egyébként.} \end{cases}$$

Határozzuk meg ξ és η sűrűségfüggvényét!

Az együttes sűrűség tartója az a háromszög, amelynek csúcsai a (0, 0), (1, 0) és (1, 1) pont. Így ξ és η egyaránt a [0, 1] intervallumban veszi fel értékeit.

$$f_\xi(x) = \int_0^x 2dy = 2x, \qquad f_\eta(y) = \int_y^1 2dx = 2(1-y)\,.$$

Tehát

$$f_\xi(x) = \begin{cases} 2x & \text{ha } 0 \le x \le 1 \\ 0 & \text{egyébként,} \end{cases} \qquad f_\eta(y) = \begin{cases} 2(1-y) & \text{ha } 0 \le y \le 1 \\ 0 & \text{egyébként.} \end{cases}$$

Érdemes hangsúlyozni, hogy annak ellenére, hogy (ξ, η) együttesen egyenletes eloszlású, a peremeloszlások nem egyenletesek. □

4.2. Függetlenség

Tételezzük fel, hogy a folytonos ξ és η valószínűségi változóknak létezik együttes sűrűségfüggvénye. Ha ξ és η függetlenek, akkor együttes eloszlásfüggvényükre

$$F(x, y) = F_\xi(x)F_\eta(y),$$

amiből differenciálással adódik

$$f(x, y) = f_\xi(x)f_\eta(y). \tag{4.2.1}$$

Azt is szoktuk mondani, hogy függetlenség esetén az együttes sűrűségfüggvény tényezőkre bomlik (faktorizál).

4.1. tétel: *Tételezzük fel, hogy a ξ és az η valószínűségi változóknak létezik együttes sűrűségfüggvénye. Ekkor ξ és η pontosan akkor függetlenek, ha (4.2.1) fennáll.*

4.4. példa: A 4.2. példában szereplő ξ és η valószínűségi változók függetlenek, mert – amint megmutatható – együttes sűrűségfüggvényük szorzattá bomlik.

Most két független valószínűségi változó összegének eloszlás- ill. sűrűségfüggvényét fogjuk kiszámolni. Tehát legyenek ξ és η függetlenek, $\zeta = \xi + \eta$. A $\zeta < z$ esemény azt jelenti, hogy a (ξ, η) pont az $y = z - x$ egyenes által határolt (alsó) félsíkba esik. Jelöljük A-val ezt a félsíkot.

$$F_\zeta(z) = P(\zeta < z) = \iint_A f_\xi(x)f_\eta(y)\,dx\,dy$$

A kettős integrálban helyettesítést hajtunk végre: $u = x + y$, $v = y$. Ez a transzformáció az A félsíkot az $u - v$ sík $u = z$ függőleges egyenese által határolt baloldali félsíkba viszi. A transzformáció függvénydeterminánsa 1, ezért

$$F_\zeta(z) = \iint_B f_\xi(u-v)f_\eta(v)\,du\,dv = \int_{-\infty}^{z}\int_{-\infty}^{\infty} f_\xi(u-v)f_\eta(v)\,dv\,du\,.$$

Differenciálással kapjuk ζ sűrűségfüggvényét

$$\boxed{f_\zeta(z) = \int_{-\infty}^{\infty} f_\xi(z-v)f_\eta(v)\,dv,} \tag{4.2.2}$$

és ezt az f_ξ és f_η sűrűség **konvolúciójának** nevezzük. □

4.5. példa: Legyenek ξ és η független azonos eloszlású valószínűségi változók. Közös sűrűségfüggvényük

$$f(x) = \begin{cases} e^{-x} & \text{ha } x \geq 0, \\ 0 & \text{egyébként.} \end{cases}$$

Számítsuk ki $\xi + \eta$ sűrűségfüggvényét!

A konvolúció (4.2.2) képletét kell használni. Mivel ξ és η pozitív értékeket vesz fel, a $\xi + \eta$ valószínűségi változó g sűrűségfüggvénye 0 negatív z értékekre. Ha $z > 0$, akkor

$$g(z) = \int_{-\infty}^{\infty} f_\xi(z-x) f_\eta(x)\, dx = \int_0^z e^{-(z-x)} e^{-x}\, dx = z e^{-z}.$$

Megemlítjük, hogy ez a g függvény egy másodrendű **gamma eloszlás** sűrűségfüggvénye, lásd (3.2.2)-t. □

4.3. Több valószínűségi változó függvényei

Ha a $\xi_1, \xi_2, \ldots, \xi_n$ valószínűségi változók együttes eloszlása ismeretes, akkor egy n változós h függvényre a $h(\xi_1, \xi_2, \ldots, \xi_n)$ valószínűségi változó eloszlása meghatározott. Leggyakrabban, például a statisztikában,

$$a_1\xi_1 + a_2\xi_2 + \ldots + a_n\xi_n$$

alakú lineáris kifejezések fordulnak elő, és $\xi_1, \xi_2, \ldots, \xi_n$ teljesen független változók.

4.6. példa: ξ_1, ξ_2 független $N(0, \sigma)$ eloszlású valószínűségi változók. Határozzuk meg az

$$\eta = \sqrt{\xi_1^2 + \xi_2^2}$$

valószínűségi változó sűrűségfüggvényét!

A függetlenség miatt ξ_1, ξ_2 együttes sűrűségfüggvénye

$$\begin{aligned} f(x,y) &= \frac{1}{2\pi\sigma^2} \exp\left(-x^2/2\sigma^2\right) \exp\left(-y^2/2\sigma^2\right) \\ &= \frac{1}{2\pi\sigma^2} \exp\left(-(x^2+y^2)/2\sigma^2\right). \end{aligned}$$

Először η eloszlásfüggvényét számoljuk ki.

$$\begin{aligned} F_\eta(z) &= P\left(\sqrt{\xi_1^2 + \xi_2^2} < z\right) = P(\xi_1^2 + \xi_2^2 < z^2) \\ &= \iint_{x^2+y^2<z^2} \frac{1}{2\pi\sigma^2} \exp\left(-(x^2+y^2)/2\sigma^2\right) dx\, dy. \end{aligned}$$

Itt a kettős integrál a z sugarú körlemezen van véve. Célszerű áttérni polárkoordinátákra.

$$\begin{aligned} F_\eta(z) &= \frac{1}{2\pi\sigma^2} \int_0^{2\pi} \left(\int_0^z r \exp\left(-r^2/2\sigma^2\right) dr \right) d\theta \\ &= \frac{1}{\sigma^2} \int_0^z r \exp\left(-r^2/2\sigma^2\right) dr = 1 - \exp\left(-z^2/2\sigma^2\right). \end{aligned}$$

A sűrűségfüggvényt differenciálással kapjuk:

$$f_\eta(z) = \frac{d}{dz} F_\eta(z) = \sigma^{-2} z \exp\left(- z^2/2\sigma^2 \right). \tag{4.3.1}$$

(A $\sigma = 1$ esetben ez egy 2-szabadságfokú χ-eloszlás.) □

A következő tételt valószínűségi változók függvényének várható értékéről bizonyítás nélkül közöljük.

4.2. tétel: *Ha a ξ_1 és ξ_2 valószínűségi változók együttes sűrűségfüggvénye f és h egy kétváltozós függvény, akkor*

$$M(h(\xi_1, \xi_2)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) f(x,y)\, dx\, dy.$$

4.7. példa: A ξ és η valószínűségi változók együttes sűrűségfüggvénye a következő:

$$h(x,y) = \begin{cases} 2 & \text{ha } 0 \le x \le 1 \text{ és } 0 \le y \le x, \\ 0 & \text{egyébként.} \end{cases}$$

Határozzuk meg $\xi\eta$ várható értékét!

Az előző tétel szerint az $xyh(x,y)$ kifejezést kell integrálni a (0,0), (1,0) és (1,1) pontokkal meghatározott háromszögön, azaz

$$M(\xi\eta) = 2 \iint_H xy\, dx\, dy = 2 \int_0^1 \Big(\int_0^x xy\, dy \Big)\, dx = \int_0^1 x^3\, dx = 1/4\,.$$ □

4.4. Kovariancia és korrelációs együttható

Legyen ξ és η két tetszőleges valószínűségi változó. Egymástól való függőségüket az

$$M\Big((\xi - M(\xi))(\eta - M(\eta)) \Big) \tag{4.4.1}$$

várható értékkel mérjük, amit **kovarianciának** nevezünk. (A kovariancia jelölésére a $\sigma_{\xi,\eta}$, $\mathrm{Cov}(\sigma, \eta)$ jelölések használatosak.) Ha ξ és η függetlenek, akkor kovarianciájuk 0. A 2.2. tételből adódik, hogy a kovariancia a

$$M(\xi\eta) - M(\xi)M(\eta) \tag{4.4.2}$$

formában is írható. A kovariancia segítségével van értelmezve a **korrelációs együttható**:

$$\boxed{r = \frac{M(\xi - M(\xi))(\eta - M(\eta))}{\sigma(\xi)\sigma(\eta)}} \tag{4.4.3}$$

A korrelációs együttható -1 és 1 közé esik, és abszolút értékben az 1 értéket csak akkor veszi fel, ha ξ és η egymásnak lineáris függvénye, azaz vannak olyan a és b számok, hogy $\xi = a\eta + b$. A 2.7. és 4.3. feladat két nem független, de 0 korrelációs együtthatójú (azaz korrelálatlan) valószínűségi változóra ad példát. A korrelálatlanság tehát még nem jelent függetlenséget.

A korrelációs együttható egy szög koszinuszaként értelmezhető, ha a ξ és η valószínűségi változókat olyan vektorként képzeljük el, amelyek hossza a szórásuk. A

$$\sigma^2(\xi + \eta) = \sigma^2(\xi) + \sigma^2(\eta) + 2r\sigma(\xi)\sigma(\eta) \tag{4.4.4}$$

összefüggésben a háromszög-geometria koszinusztételére ismerhetünk.

4.8. példa: A ξ és η valószínűségi változók együttes sűrűségfüggvénye a következő:

$$h(x,y) = \begin{cases} 2 & \text{ha } 0 \le x \le 1 \text{ és } 0 \le y \le x, \\ 0 & \text{egyébként.} \end{cases}$$

Számoljuk ki ξ és η korrelációs együtthatóját!

A 4.3. példa tartalmazza a peremsűrűségeket, felhasználásukkal kiszámoljuk ξ és η várható értékét:

$$\begin{aligned} M(\xi) &= \int_0^1 x \cdot 2x\,dx = \frac{2}{3}, & \sigma^2(\xi) = \int_0^1 (x-2/3)^2 \cdot 2x\,dx = \frac{1}{18}, \\ M(\eta) &= \int_0^1 y \cdot 2(1-y)\,dy = \frac{1}{3}, & \sigma^2(\eta) = \int_0^1 (y-1/3)^2 2(1-y) dy = \frac{1}{18}. \end{aligned}$$

Végül a 4.7. példából tudjuk, hogy $M(\xi\eta) = 1/4$ és

$$r = \frac{M(\xi\eta) - M(\xi)M(\eta)}{\sigma(\xi)\sigma(\eta)} = \frac{1}{2}$$ □

Több valószínűségi változó esetén a páronkénti kovarianciákat és korrelációs együtthatókat egy-egy mátrixba foglalhatjuk össze. Legyen $(\xi_1, \xi_2, \ldots, \xi_n)$ n valószínűségi változó. Azt a $\mathbf{C}$ mátrixot, amelynek i-edik sorának j-edik eleme a $\mathrm{Cov}(\xi_i, \xi_j)$ kovariancia, a valószínűségi változók **kovarianciamátrixának** nevezzük. Hasonlóan van definiálva az **R korrelációmátrix** a páronkénti korrelációs együtthatókkal. Mivel bármely valószínűségi változónak önmagával vett korrelácios együtthatója 1, az **R** mátrix diagonálisa csupa egyesből áll.

4.3. tétel: **C** *és* **R** *pozitív szemidefinit mátrixok, és közöttük a*

$$\mathbf{C} = \mathbf{S} \cdot \mathbf{R} \cdot \mathbf{S}$$

kapcsolat áll fent, ahol $\cdot$ a közönséges mátrixszorzást jelöli, és

$$\mathbf{S} = \mathrm{Diag}(\sigma(\xi_1), \sigma(\xi_2), \ldots, \sigma(\xi_n))$$

egy olyan diagonális mátrix, amely a szórásokból áll.

Megjegyezzük, hogy egy mátrix pozitiv szemidefinit volta azt jelenti, hogy a mátrix X^tX alakú valamilyen X mátrixra, amelynek transzponáltja X^t. Pozitív szemidefinit mátrix sajátértékei nemnegatívak.

4.5. A többdimenziós normális eloszlás

A $\xi_1, \xi_2, \ldots, \xi_n$ valószínűségi változók együttes eloszlását normálisnak nevezzük, ha együttes sűrűségfüggvényük

$$\boxed{f(x_1, x_2, \ldots, x_n) = \sqrt{\frac{|\mathbf{B}|}{(2\pi)^n}} \exp\Big(-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}(x_i - m_i)(x_j - m_j)\Big)} \qquad (4.5.1)$$

alakú, ahol $m_1, m_2, \ldots, m_n$ valós számok és $\mathbf{B} = (b_{ij})$ egy $n \times n$-es szimmetrikus pozitív definit mátrix. Mind az m_i paraméterek, mind pedig a $\mathbf{B}$ mátrix valószínűségelméleti jelentéssel bír. Az m_i szám nem más mint ξ_i várhatóértéke, a $\mathbf{B}$ mátrix pedig a $\mathbf{C}$ kovarianciamátrix inverze.

A vektor-mátrix írásmóddal (4.5.1) jóval tömörebbé válik:

$$\begin{aligned} f(x) &= \sqrt{\frac{|\mathbf{B}|}{(2\pi)^n}} \exp\left(-\tfrac{1}{2} < \mathbf{B}(x-m),\, (x-m) >\right) \\ &= \frac{1}{\sqrt{|\mathbf{C}|(2\pi)^n}} \exp\left(-\tfrac{1}{2} < \mathbf{C}^{-1}(x-m), (x-m) >\right) \end{aligned}$$

4.9. példa: Ellenőrizzük, hogy az alábbi kétváltozós függvény kétdimenziós normális eloszlás sűrűségfüggvénye-e!

$$f(x,y) = \frac{1}{\pi\sqrt{3}} \exp\Big(-\frac{2}{3}(x^2 - xy + y^2)\Big)$$

Keressük meg az eloszlás kovarianciamátrixát!

Azt kell megvizsgálni, hogy az adott függvény (4.5.1) alakú-e. Először a kitevőt nézzük meg. (4.5.1)-ben az áll, hogy

$$-\frac{1}{2}\Big(b_{11}(x-m_1)^2 + (b_{12}+b_{21})(x-m_1)(y-m_2) + b_{22}(y-m_2)^2\Big),$$

ami akkor lesz

$$-\frac{2}{3}(x^2 - xy + y^2),$$

ha

$$\mathbf{B} = \begin{pmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{pmatrix} = \frac{2}{3}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix},$$

és $m_1 = m_2 = 0$. $\mathbf{B}$ determinánsa $4/3$, és a kis számolás

$$\sqrt{\frac{4/3}{(2\pi)^2}} = \frac{1}{\pi\sqrt{3}}$$

azt mutatja, hogy f (4.5.1) alakú. A kovarianciamátrix $\mathbf{B}$ inverze, ami a lineáris algebrában tanultak alapján számolható ki, és

$$\frac{1}{2}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} -\text{nek}$$

adódik. □

Ha a $(\xi_1, \xi_2, \ldots, \xi_n)$ valószínűségi változók együttes eloszlása normális (azaz (4.5.1) alakú), akkor $\xi_1, \xi_2, \ldots,$ és ξ_n külön-külön is normális eloszlásúak. Az egyszerűség kedvéért korlátozódjunk az $n = 2$ esetre, és legyen ξ_i $N(m_i, \sigma_i)$ eloszlású. Ha r a ξ_1 és ξ_2 valószínűségi változók korrelációs együtthatója, akkor együttes sűrűségük

$$\boxed{\begin{aligned} f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\Big(-\frac{1}{2(1-r^2)}\times \\ \times\Big[\frac{(x-m_1)^2}{\sigma_1^2} - 2r\frac{x-m_1}{\sigma_1}\frac{y-m_2}{\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\Big]\Big) \end{aligned}} \tag{4.5.2}$$

alakú. Megfigyelhetjük, hogy $r=0$ esetén az együttes sűrűségfüggvény faktorizál egy x-től és egy y-től függő sűrűségfüggvény szorzatára, tehát ilyenkor ξ_1 és ξ_2 függetlenek.

4.4. tétel: *Ha két valószínűségi változó együttes eloszlása normális, és a változók korrelálatlanok, akkor függetlenek is.*

Mint azt már több ízben hangsúlyoztuk, teljes általánosságban a korrelálatlanság nem vonja maga után a függetlenséget. A tételben az együttes eloszlás normalitásának a feltételezése igen lényeges.

4.10. példa: Mi a 4.9. példában szereplő együttes eloszlással adott valószínűségi változók korrelációs együtthatója?

A korrelációs együtthatót az együttes sűrűségfüggvény (4.5.2)-vel való összevetésével kaphatjuk meg:

$$\frac{1}{\pi\sqrt{3}}=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}},$$

amiből $3=4\sigma_1^2\sigma_2^2(1-r^2)$. A szórásnégyzetek a 4.9. példából ismert kovarianciamátrixból kipotyognak, $\sigma_1^2=\sigma_2^2=1$, és $r=1/2$. □

4.5. tétel: *Ha a ξ_1 és ξ_2 valószínűségi változók együttes eloszlása normális, akkor bármilyen $\lambda,\mu\in R$ esetén $\lambda\xi_1+\mu\xi_2$ normális eloszlású. Nevezetesen ξ_1 és ξ_2 is normális eloszlású.*

Ez a tétel azt is jelenti, hogy ha az együttes eloszlás normális, akkor a peremeloszlások is normálisak.

4.11. példa: Határozzuk meg a 4.9. példában szereplő együttes eloszlással adott valószínűségi változók peremeloszlását!

A peremeloszlások normálisak, így elegendő tudnunk várható értéküket és szórásukat. Ezek kiderülnek a 4.9. példából. Tehát mindkét peremeloszlás standard normális. □

4.12. példa: Legyen ξ_1 egy véletlenszerűen kiválasztott házaspárból a nő, ξ_2 pedig a férfi testmagassága. Tételezzük fel, hogy ξ_1 és ξ_2 együttesen normális eloszlású, ξ_1 várható értéke 169 cm, szórása 5 cm, ξ_2 várható értéke 177 cm, szórása 5 cm, továbbá ξ_1 és ξ_2 korrelációs együtthatója 0.68. Mi a valószínűsége, hogy egy feleség magasabb a férjénél?

A $P(\xi_1-\xi_2>0)$ valószínűséget kell kiszámolni. A 4.4. tétel szerint $\xi_1-\xi_2$ normális eloszlású.

$$\begin{aligned} M(\xi_1-\xi_2) &= 169-177=-8 \\ \sigma^2(\xi_1-\xi_2) &= \sigma^2(\xi_1)+\sigma^2(\xi_2)-2\mathrm{Cov}(\xi_1,\xi_2)=25+25-2\cdot 0.68\cdot 5\cdot 5=16 \end{aligned}$$

Így $\xi_1-\xi_2$ szórása 4 és $\eta=(\xi_1-\xi_2+8)/4$ standard normális eloszlású. $\xi_1-\xi_2>0$ ekvivalens $\eta>2$ egyenlőtlenséggel. Ezért

$$P(\xi_1-\xi_2>0)=P(\eta>2)=1-\Phi(2)=2.3\%.$$

Tehát 2.3 % annak a valószínűsége, hogy egy feleség magasabb a férjénél. □

4.6. A feltételes sűrűségfüggvény és a regresszió

Legyen ξ és η két olyan folytonos valószínűségi változó, amelyek f együttes sűrűségfüggvénye létezik. Tudjuk, hogy ekkor η-nak van sűrűségfüggvénye, és tegyük fel, hogy $f_\eta(y) > 0$. Rögzített y_0 mellett

$$f(x|y_0) = \frac{f(x, y_0)}{f_\eta(y_0)} \tag{4.6.1}$$

egy sűrűségfüggvény, hiszen

$$\int_{-\infty}^{\infty} f(x|y_0)\, dx = \frac{1}{f_\eta(y_0)} \int_{-\infty}^{\infty} f(x, y_0)\, dx = \frac{1}{f_\eta(y_0)} f_\eta(y_0) = 1.$$

$f(x|y_0)$-t ξ-nek az $\eta = y_0$ **feltételre vonatkozó feltételes sűrűségfüggvényének** nevezzük. Az $x \mapsto f(x|y_0)$ sűrűségfüggvényhez tartozó valószínűségi változóra a „ξ feltéve $\eta = y_0$” elnevezést, és a $\xi|\eta = y_0$ jelölést használjuk. A $\xi|\eta = y_0$ feltételes valószínűségi változó arról ad felvilágosítást, hogy ξ milyen valószínűséggel veszi fel értékeit feltéve, hogy η-ról tudunk valamit, nevezetesen, hogy $\eta = y_0$. Ha ξ és η függetlenek, akkor

$$f(x|y_0) = \frac{f(x, y_0)}{f_\eta(y_0)} = \frac{f_\xi(x) f_\eta(y_0)}{f_\eta(y_0)} = f_\xi(x)\,.$$

Tehát ilyenkor a ξ és a $\xi|\eta = y_0$ valószínűségi változók azonosak.

4.13. példa: ξ és η együttes sűrűségfüggvénye

$$f(x, y) = \begin{cases} \frac{1+xy}{3} & \text{ha } 0 < x < 2 \text{ és } 0 < y < 1, \\ 0 & \text{különben.} \end{cases}$$

Írjuk fel az $f(x|y_0)$ feltételes sűrűségfüggvényt!

Először η peremsűrűségét számoljuk ki.

$$f_\eta(y) = \int_0^2 \frac{1 + xy}{3}\, dx = \frac{2 + 2y}{3}$$

ha $0 < y < 1$, egyébként $f_\eta(y) = 0$.

A feltételes sűrűségfüggvény definíciója alapján

$$f(x|y_0) = \begin{cases} \dfrac{1 + xy_0}{2 + 2y_0} & \text{ha } 0 < x < 1, \quad 0 < y_0 < 1, \\ 0 & \text{különben.} \end{cases}$$

□

Azt a függvényt, ami y_0-hoz az $M(\xi|\eta = y_0)$ várható értéket rendeli, a ξ valószínűségi változó η-ra vonatkozó *regressziójának*, vagy elméleti regressziójának nevezzük. Az $y \mapsto M(\xi|\eta = y)$ függvény gráfja a **regressziós görbe**.

Ha ξ és η függetlenek, akkor – mint fent láttuk – a ξ és a $\xi|\eta = y$ változók azonosak, ezért ξ-nek az η-ra vonatkozó regressziója az $M(\xi)$ állandó függvény. A regressziós görbe meredeksége ξ és η sztochasztikus összefüggését mutatja. Legjobban látszik ez abban az esetben, amikor ξ és η együttes eloszlása normális. (Az egyszerűség érdekében tételezzük fel, hogy ξ és η várható értéke 0.) Ekkor az $u = M(\xi|\eta = v)$ regressziós görbe az $u = -r\sigma(\eta)v/\sigma(\xi)$ egyenes. (r jelöli a korrelációs együtthatót.)

4.14. példa: Egy hőkezelési folyamattal kapcsolatban ξ a kezelés időtartama és η a kezelés által kiváltott keményedés mélysége. (ξ-t másodpercben, η-t mm-ben mérjük.) Egy modell szerint ξ és η együttes eloszlása normális, $M(\xi) = 18$, $M(\eta) = 9$ $\sigma(\xi) = 4.8$, $\sigma(\eta) = 2$ és a korrelációs együttható $r = 0,87$.

(a) Adja meg az $x \mapsto M(\eta|\xi = x)$ regressziót!

(b) Számolja ki a feltételes sűrűségfüggvényt, a „kezelés hossza 15 másodperc"-re vonatkozóan!

(c) Mi a valószínűsége, hogy a keményedés mélysége 9 és 12 mm közé esik, feltéve, hogy a hőkezelés időtartama 15 másodperc?

Az $m_1 = 18$, $m_2 = 9$, $\sigma_1 = 4.8$, $\sigma_2 = 2$, $r = 0.87$ adatok felhasználásával (4.5.2)-be helyettesítve felírjuk az f együttes sűrűségfüggvényt, és azt elosztjuk ξ-nek az f_ξ sűrűségfüggvényével. Így kapjuk, az

$$f(y|x) = C \exp\Big(-\frac{1}{2}\frac{(y - 0.36x - 2.52)^2}{2}\Big) \tag{4.6.2}$$

feltételes sűrűségfüggvényt, amely normális. Kiolvassuk, hogy 2 a szórásnégyzete és $0.36x+2.52$ a várhatóértéke. A (b) kérdés válaszához úgy jutunk, hogy (4.6.2)-ben $x = 15$-öt helyettesítünk. Így egy normális ζ valószínűségi változót kapunk, amelynek 7.92 a várható értéke és $\sqrt{2}$ a szórása. A (c) kérdés megválaszolásához a $P(9 \le \zeta \le 12)$ valószínűséget kell megadni, ami standardizálással $\Phi(3.05) - \Phi(0.92) = 0.1777$. □

A mérnöki-statisztikai problémákban gyakran ismerjük két sztochasztikusan összefüggő valószínűségi változó néhány össszetartozó értékpárját, és keressük a változók közötti kapcsolat legjobb lineáris közelítését. Adott ξ és η. Ha η-t $A\xi + B$ alakban akarjuk közelíteni, A és B értékét úgy kell meghatározni, hogy a $\eta - (A\xi + B)$ valószínűségi változó szórása a lehető legkisebb legyen. A

$$\begin{aligned} g(A,B) &= M((\eta - (A\xi + B))^2) \\ &= M(\eta^2) + A^2M(\xi^2) + B^2 - 2AM(\xi\eta) - 2BM(\eta) + 2ABM(\xi) \end{aligned}$$

függvény minimumát az ismert módon parciális deriválással határozzuk meg. Az adódik, hogy

$$\frac{M(\eta) + \mathrm{Cov}(\xi,\eta)}{\sigma^2(\xi)(\xi - M(\xi))} \tag{4.6.3}$$

η legjobb megközelítése a legkisebb négyzetek értelmében. (Az $u = M(\eta) + c(v - M(\xi))/\sigma^2(\xi)$ egyenest **regressziós egyenesnek** is hívjuk. Ez átmegy a $v_0 = M(\xi)$, $u_0 = M(\eta)$ ponton.) Abban az esetben, ha az elméleti regresszió egy lineáris függvény, akkor gráfja a regressziós egyenes. Amennyiben az elméleti regresszió közel lineáris, helyettesíthetjük a lineáris regresszióval.

4.7. A sztochasztikus folyamat

Egy időtől függő véletlen mennyiséget **sztochasztikus folyamatnak** nevezünk. Itt nem célunk a sztochasztikus folyamatok kimerítő ismertetése, elsősorban a **Poisson-folyamat** példájára szorítkozunk.

A Poisson-folyamat jól szemléltethető a polimerizáció egyszerű valószínűségelméleti modelljén. A polimerizáció láncmolekulák növekedési folyamata, amelynek során a már kialakult lánchoz újabb és újabb monomer csoport kapcsolódik. A láncok végének összekapcsolódását az egyszerű modellben kizárjuk, és azt gondoljuk, hogy a láncmolekula és a monomer csoport véletlen találkozásakor a csoport hozzákötődik a lánchoz, és annak hosszát eggyel megnöveli. η_1 legyen az a véletlen időtartam, ami az első monomer csoport kapcsolódásáig eltelik, η_2 időtartam múlva kötődik a második csoport az első kapcsolódása után, és így tovább. Tehát $\eta_1, \eta_2, \ldots$ valószínűségi változók egy sorozata. A valóságban ezek a véletlen időtartamok nagyon rövidek, ez azonban elvi megfontolásainkat nem zavarja. Mivel η_i egy „emlékezet nélküli várakozási idő”, η_i exponenciális eloszlású kell, hogy legyen, mondjuk α_i paraméterrel, és feltételezhető, hogy az η_i valószínűségi változók teljesen függetlenek. Amennyiben a polimerizáció során bizonyos körülmények változatlanok (például hőmérséklet, a nagy feleslegben jelenlévő szabad monomerek koncentrációja, stb.), akkor feltételezhető, hogy $\eta_1, \eta_2 \ldots$ azonos eloszlásúak, azaz $\alpha_i = \alpha$, i-től függetlenül.

Vizsgáljuk meg, hogy t idő eltelte után milyen hosszúak a láncmolekulák! Jelölje N_t azt a valószínűségi változót, amely a t időpontban véletlenül választott lánc hosszát adja meg. N_t értékei pozitív egész számok és

$$P(N_t \geq n) = P(\eta_1 + \eta_2 + \ldots + \eta_n \leq t), \tag{4.7.1}$$

azaz a láncmolekula hossza pontosan akkor nagyobb vagy egyenlő mint n a t időpontban, ha az $\eta_1, \eta_2, \ldots, \eta_n$ várakozási idők összege legfeljebb t. (4.7.1)-ből kiindulva számoljuk ki N_t eloszlását. Nyilván

$$P(N_t = n) = P(N_t \leq n) - P(N_t \leq n-1). \tag{4.7.2}$$

Érdemes bevezetni a $\xi_n = \eta_1 + \eta_2 + \ldots + \eta_n$ valószínűségi változót. Mivel független változókat adunk össze, ξ_n momentum generáló függvénye

$$M(e^{t\xi_n}) = \left(\frac{\alpha}{\alpha - t}\right)^n \tag{4.7.3}$$

(a 2.10. példa és 2.4. tétel alapján), amiből a Laplace-transzformáció táblázatát használva kapjuk, hogy ξ_n sűrűségfüggvénye

$$f_n(x) = \frac{\alpha^n}{(n-1)!} x^{n-1} e^{-\alpha x} \qquad (x > 0). \tag{4.7.4}$$

(Különben (3.2.2)-vel összevetve megállapítható, hogy ez gamma eloszlás.) Némi parciális integrálással látszik, hogy

$$P(N_t = n) = \int_0^t f_n(x)\,dx - \int_0^t f_{n-1}(x)\,dx = \frac{(\alpha t)^n e^{-\alpha t}}{n!}. \tag{4.7.5}$$

Tehát N_t Poisson-eloszlású, αt paraméterrel. Amit így megkonstruáltunk, az a Poisson-folyamat.

Tételezzük fel, hogy minden $t \in \mathbb{R}^+$ számra adott egy X_t valószínűségi változó úgy, hogy

(a) $X_0 \equiv 0$,

(b) ha $t_1 < s_1 < t_2 < \ldots < s_n$, akkor az $X_{s_1} - X_{t_1}, X_{s_2} - X_{t_2}, \ldots, X_{s_n} - X_{t_n}$ valószínűségi változók függetlenek,

(c) $X_s - X_t$ Poisson-eloszlású $\alpha(s-t)$ paraméterrel, ha $s > t$. Az ilyen (X_t) valószínűségi változó

családot α intenzitású Poisson-folyamatnak nevezzük. A (b) és (c) feltételek úgy fogalmazhatók meg, hogy a folyamat növekményei (diszjunkt időintervallumokban) függetlenek, és Poisson-eloszlásúak az időintervallumok hosszával arányos paraméterrel.

Ellenőrizhető, hogy a fent konstruált N_t folyamat eleget tesz az (a)–(c) feltételeknek, tehát egy konkrét példa az absztrakt Poisson-folyamatra. A példában a folyamat α intenzitása a reakciósebességgel függ össze, α az időegység alatt bekövetkezett átlagos láncnövekedés, másként a vizsgált nulladrendű reakció sebességi együtthatója.

4.15. példa: Egy polimerizációs folyamatot Poisson-féle sztochasztikus folyamat ír le. Miután a polimerizáció 10 percig folyt, a láncmolekulák átlagosan 280 monomercsoportból álltak. Még hány percig kell folytatni az eljárást ahhoz, hogy a molekulaláncok 95 %-a tartalmazzon legalább 1200 monomer csoportot?

Legyen λ a folyamat intenzitása, ekkor 10 perc után a láncok eloszlása 10λ paraméterű Poisson-eloszlás. A Poisson-eloszlás várható értéke maga a paraméterérték, így $10\lambda = 280$ és $\lambda = 28$.

Kiszámoljuk, hogy milyen Λ paraméterű Poisson-eloszlásra igaz, hogy a 1200-nál nagyobb értékeket 95 % valószínűséggel vesz fel. Ehhez a

$$\sum_{n=1200}^{\infty} \frac{\Lambda^n}{n!} e^{-\Lambda} = 0.95$$

egyenletet kell megoldani Λ-ra. Ez numerikusan, számítógép segítségével lehetséges, $\Lambda = 1144,575$ (lásd függelék).

A $(t+10)\lambda = \Lambda$ egyenletekből kapjuk a megoldást, $t = \Lambda/\lambda - 10 = 30.87$. □

4.16. példa: Mi a magyarázata annak a tapasztalati ténynek, hogy a polimerek lánchosszának eloszlása akkor is Poisson-eloszlás, ha a reakciósebesség a polimerizáció során időben változik?

Gondoljunk arra, hogy a polimerizáció 3 percig α intenzitással folyik, és 3 perc után megváltozik az intenzitása β-ra. Ekkor a lánchossz $t > 3$-ra

$$N_3^{(\alpha)} + N_{t-3}^{(\beta)}.$$

Itt mindkét változó Poisson-eloszlású, és egymástól függetlenek. Ezért összegük is Poisson-eloszlású a 3.2. tétel következtében. □

Poisson-folyamattal írható le egy üzletbe betérő vásárlók száma, vagy egy telefonközpontba beérkező hívások száma is.

Gyakorlatok

4.1. gyakorlat:$^{\mathrm{M}}$ Legyen ξ_1 és ξ_2 független valószínűségi változó és tételezzük fel, hogy ξ_i exponenciális eloszlású β_i paraméterrel $(i = 1, 2)$. Bizonyítsuk be, hogy

$$P(\xi_1 > \lambda \xi_2) = \frac{\beta_2}{\lambda \beta_1 + \beta_2}$$

bármilyen $\lambda > 0$ számra!

4.2. gyakorlat: Legyenek ξ_1 és ξ_2 független valószínűségi változók, ξ_i n_i-ed rendű, p sikervalószínűségű binomiális eloszlású $(i = 1, 2)$. Mi a $\xi_1 + \xi_2$ eloszlása?

4.3. gyakorlat:$^{\mathrm{M}}$ Legyen ξ egyenletes eloszlású a $(0, 2\pi)$ intervallumban. Bizonyítsuk be, hogy $\sin \xi$ és $\cos \xi$ korrelálatlan valószínűségi változók.

4.4. gyakorlat: ξ és η együttesen egyenletes eloszlású az origó középpontú egységnyi sugarú körlemezen. Számítsa ki a $P(\xi < \eta)$, $P(\xi^2 + \eta^2 < 1/4)$, $P(\max(\xi, \eta) < 1/2)$ valószínűségeket!

4.5. gyakorlat:$^{\mathrm{M}}$ A síkon úgy választunk ki véletlenszerűen egy pontot, hogy x és y koordinátáit egymástól függetlenül $N(0,1)$ eloszlással adjuk meg. Milyen origó középpontú R sugarú körre igaz, hogy a véletlen pont 99% valószínűséggel beleesik? (A 4.6. példa sok segítséget nyújt a megoldáshoz.)

4.6. gyakorlat: A folyadékban (vagy gázban) levő részecskék az ütközések következtében szabálytalan, ún. Brown-féle mozgást végeznek, amely úgy modellezhető, hogy t idő alatt a részecske x, y és z koordinátájának megváltozása olyan $\xi_x^{(t)}, \xi_y^{(t)}, \xi_z^{(t)}$ valószínűségi változók, amelyek függetlenek és $N(0, \sigma\sqrt{t})$ eloszlásúak. Mi a valószínűsége, hogy a $t = 0$ időpontban az origóban levő részecske a $t = 2$ időpontban az origó középpontú 4σ sugarú gömbben van?

4.7. gyakorlat:$^{\mathrm{M}}$ Milyen eloszlásúak a 4.2. példában szereplő ξ és η valószínűségi változók?

4.8. gyakorlat: Legyen ξ_1 és ξ_2 két olyan nem független valószínűségi változó, amelynek együttes eloszlása normális. Bizonyítsuk be, hogy mindig találhatók olyan α és β valós számok, hogy a $\xi_1 + \alpha\xi_2$ és a $\xi_1 + \beta\xi_2$ valószínűségi változók függetlenek.

4.9. gyakorlat: Legyen $A = (a_{ij})$ egy nem szinguláris $n \times n$-es mátrix és $\xi_1, \xi_2, \dots, \xi_n$ olyan valószínűségi változók, amelyeknek kovarianciamátrixa C. Ha

$$\eta_i = \sum_{k=1}^{n} a_{ik}\xi_k,$$

akkor bizonyítsuk be, hogy az $\eta_1, \eta_2, \dots, \eta_n$ valószínűségi változók kovarianciamátrixa $A \cdot C \cdot A^t$, ahol A^t az A mátrix transzponáltja.

4.10. gyakorlat: A ξ és η valószínűségi változók egyenletes eloszlásúak az origó középpontú egységnyi sugarú körön. Számítsa ki a korrelációs együtthatójukat!

4.11. gyakorlat: ξ és η együttesen egyenletes eloszlású valószínűségi változók a $0 < x < 1$, $0 < y < x$ háromszögön. η-t a ξ^2 alakban akarjuk közelíteni. Milyen $v = an^2$ függvényt kell ahhoz választani, hogy $(\eta - a\xi^2)^2$ várható értéke a legkisebb legyen?

4.12. gyakorlat: A ξ valószínűségi változó egyenletes eloszlású a $(-2, 2)$ intervallumban és $\eta = \xi^6$. Számolja ki a ξ és η valószínűségi változók korrelációs együtthatóját!

4.13. gyakorlat:$^{\mathrm{M}}$ A ξ_1, ξ_2, ξ_3 valószínűségi változók kovarianciamátrixa a következő:

$$\begin{pmatrix} 2.44 & -2.4 & 0.84 \\ -2.4 & 5 & -1.4 \\ 0.84 & -1.4 & 1.49 \end{pmatrix}$$

Írja fel a ξ_1 és ξ_3 korrelációmátrixát!

4.14. gyakorlat:$^{\mathrm{M}}$ A ξ és η valószínűségi változók együttes sűrűségfüggvénye

$$f(x,y) = \begin{cases} \frac{1}{3}(x+y) & \text{ha } 0 \le x \le 1, 0 \le y \le 2 \\ 0 & \text{egyébként.} \end{cases}$$

Írja fel η eloszlásfüggvényét!

4.15. gyakorlat:$^{\mathbf{M}}$ A ξ és η valószínűségi változók együttes sűrűségfüggvénye

$$f(x,y)=\begin{cases}\frac{1}{3}(x+y) & \text{ha } 0\le x\le 1, 0\le y\le 2\\ 0 & \text{különben.}\end{cases}$$

Számolja ki $2\xi-3\eta+8$ szórását!

4.16. gyakorlat: ξ és η független valószínűségi változók, mindkettő egyenletes eloszlású $[-1,1]$-en mi a valószínűsége, hogy ξ^2 kisebb mint η?

4.17. gyakorlat:$^{\mathbf{M}}$ Legyen ξ egy véletlenszerűen választott házaspárból a férfi, η pedig a nő testmagassága. Tételezzük fel, hogy ξ és η együttesen normális eloszlású 0.68 korrelációs együtthatóval ξ várható értéke 177 cm, szórása 5 cm, η várható értéke 170 cm és szórása 5 cm. Mi a valószínűsége, hogy egy férj legalább 6 cm-rel magasabb a feleségénél? (Használja a Φ függvényt!)

4.18. gyakorlat:$^{\mathbf{M}}$ Egy kis üzletbe délután 1 és 3 óra között bármely percben átlagosan ugyanannyi vevő érkezik, és az egy perc alatt érkező vevők számának szórása 2.3. Feltételezve, hogy bármely időtartam alatt érkező vevők száma Poisson eloszlást mutat, számítsa ki annak a valószínűségét, hogy 2 percig nem jön be vevő!

4.19. gyakorlat:$^{\mathbf{M}}$ ξ és η független, azonos várható értékű, normális eloszlású valószínűségi változók. $\xi+\eta$ eloszlása N(2,4). **(a)** Milyen eloszlású $\xi-\eta$? **(b)** Számolja ki $\xi+\eta$ és $\xi-\eta$ korrelációs együtthatóját!

4.20. gyakorlat:$^{\mathbf{M}}$ ξ és η együttesen egyenletes eloszlású az origó középpontú 2 sugarú körön. Írja fel ξ sűrűségfüggvényét!

4.21. gyakorlat:$^{\mathbf{M}}$ A ξ_1, ξ_2 és ξ_3 valószínűségi változók kovarianciamátrixa a következő:

$$\begin{pmatrix}2.44 & -2.4 & 0.8\\ -2.4 & 5 & -1.4\\ 0.8 & -1.4 & 1.5\end{pmatrix}$$

Számolja ki $\xi_1+\xi_2$ és $\xi_1-\xi_2$ kovarianciáját!

Ötödik fejezet

Statisztikai következtetések

Általánosságban fogalmazva a statisztikai következtetések a tanulmányozott jelenség megfigyelési adataiból mondanak valamit a jelenséget modellező véletlen változókra vonatkozó törvényszerűségekről. A statisztikai vizsgálat tárgyát képező objektumok a hozzájuk rendelt jellemzőkkel együtt alkotják a **statisztikai sokaságot**. A jellemzők lehetnek minőségiek vagy numerikusak, de az egyszerűség kedvéért itt csak egyetlen numerikus jellemzővel foglalkozunk. Az a kísérlet, amely a sokaság egy elemének véletlen kiválasztásából áll, az egyes egyedeken értelmezett számértékű függvényt valószínűségi változóvá teszi, és ennek eloszlását a **sokaság eloszlásának** nevezzük. Legtöbbet normális eloszlású sokasággal fogunk foglalkozni.

Az egyik tipikus statisztikai probléma abban áll, hogy korábbi tapasztalatokból vagy elméleti ill. heurisztikus meggondolásból ismerjük a statisztikus sokaság eloszlásának jellegét, de nem ismerjük pontosan az eloszlástípusban szereplő valamely paraméter aktuális értékét. Ilyenkor megfigyelési adatokból becsüljük a paraméter értékét.

5.1. A statisztikai minta

A statisztika lényegéhez tartozik, hogy a valódi statisztikus sokaságnak csak egy kis részével számol. Gondoljunk arra, hogy egy termék élettartamának vizsgálata a terméket általában tönkreteszi, ezért az össztermeléshez képest viszonylag kis mintán hajtják végre. A minta kiválasztásakor ügyelni kell arra, hogy az hűen tükrözze vissza a statisztikus sokaságot, azaz a minta reprezentatív legyen. Valószínűségelméleti értelemben az n**-elemű minta** egymástól teljesen független valószínűségi változókat jelent, amelyek eloszlása megegyezik a sokaság eloszlásával. Ha $(\xi_1, \xi_2, \ldots, \xi_n)$ az n-elemű statisztikai minta, akkor a $\xi_1, \xi_2, \ldots, \xi_n$ valószínűségi változók bármely függvényét statisztikának nevezzük. A

$$\bar{x}_n = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n} \tag{5.1.1}$$

statisztikát **tapasztalati várható értéknek** hívják, és az egyik leggyakrabban használt statisztika. Ne feledkezzünk meg arról, hogy a tapasztalati várható érték sztochasztikus mennyiség, azaz egy valószínűségi változó! Értéke más és más lehet a véletlen kísérlet kimenetelétől függően.

A $(\xi_1, \xi_2, \ldots, \xi_n)$ n-elemű minta elemeit növekvően elrendezve kapjuk a **rendezett mintát**. Jelölése: $(\xi_1^*, \xi_2^*, \ldots, \xi_n^*)$. Más szóval ξ_1^* a $\xi_1, \xi_2, \ldots, \xi_n$ számok közül a legkisebb, ξ_2^* a második legkisebb, és így tovább. A rendezett mintából értelmezett

$$Y = \frac{\xi_1^* + \xi_n^*}{2} \tag{5.1.2}$$

valószínűségi változó további példa statisztikára.

5.1. példa: Egymástól függetlenül 20-szor végrehajtott mérés a következő eredményeket adta.

0.094	0.108	0.114	0.123	0.128
0.117	0.099	0.093	0.105	0.119
0.126	0.122	0.125	0.115	0.097
0.113	0.109	0.111	0.132	0.120

Tekintsük ezt 20 elemű statisztikai mintának és számoljuk ki az (5.1.1)-gyel adott tapasztalati várható értéket és az (5.1.2)-vel adott Y statisztikát!

A 20 mérési eredmény összege 2.27, amit 20-szal osztva kapjuk, hogy $\bar{x} = 0.1135$. A legkisebb mintaelem 0.093, a legnagyobb 0.132, így $Y = 0.1125$. □

Legyen $\xi_1, \xi_2, \ldots, \xi_n$ egy n-elemű statisztikai minta. Ekkor a ξ_i változók azonos eloszlásúak, ezért várható értékük ugyanaz. Ez a statisztikus sokaság várható értéke, amit elméleti várható értéknek is nevezünk, ha hangsúlyosan kívánjuk megkülönböztetni a tapasztalati várható értéktől. Nagy mintaméret esetén a standardizált tapasztalati várható érték közelítőleg normális eloszlású. Ez a tartalma a következő tételnek.

5.1. tétel: (Központi határeloszlás-tétel) *Legyen $(\xi_1, \xi_2, \ldots, \xi_n)$ az n-elemű statisztikai minta az m várható értékű és σ szórású statisztikai sokaságból. Ekkor*

$$P\Big(a < \frac{(\bar{x}_n - m)\sqrt{n}}{\sigma} < b\Big) \to \Phi(b) - \Phi(a),$$

ha $n \to \infty$.

A tételt nem bizonyítjuk. $(\bar{x}_n - m)\sqrt{n}/\sigma$ valóban $\bar{x}_n$ standardizálása, mert $M(\bar{x}_n) = M(\xi_i) = m$ és $\sigma^2(\bar{x}_n) = n\sigma^2/n^2 = \sigma^2/n$.

Más megfogalmazásban a központi határeloszlás tétel azt mondja, hogy a tapasztalati várható érték nagy mintaméretre közelítőleg normális eloszlású, függetlenül attól, hogy mi volt az eloszlása ξ_i változóknak $M(\bar{x}_n) = m$ várható értékkel és $\sigma^2(\bar{x}_n) = \sigma^2/n$ szórásnégyzettel.

5.2. Paraméterbecslés

A becslés fogalmát először egy példával világítjuk meg. Tételezzük fel, hogy egy bizonyos várakozási (vagy kiszolgálási) idő exponenciális eloszlású valószínűségi változóval modellezhető. A sűrűségfüggvény

$$f(x;b) = \begin{cases} be^{-bx} & \text{ha } x \geqslant 0, \\ 0 & \text{ha } x < 0. \end{cases}$$

A b paraméter értékét nem ismerjük, és statisztikai úton kívánjuk megbecsülni. Az exponenciális eloszlás várható értéke a b paraméter reciproka, így heurisztikusan b becslése a várható érték reciproka lehet. A $(\xi_1, \xi_2, \ldots, \xi_n)$ véletlen mintából a várható értéket az (5.1.1) tapasztalati várható értékkel becsülhetjük, és a b paramétert az

$$\frac{n}{\xi_1 + \xi_2 + \ldots + \xi_n}$$

statisztikával becsülhetjük. Ha például a 6 elemű véletlen minta a 11, 29, 14, 13, 16, 13 értékeket adja (másodpercben mérve), akkor a várható érték becslése 16, és b becslése 1/16.

A **becslés** a matematikai statisztikának azt az alapfeladatát jelenti, amikor a statisztikus sokaság eloszlásának jellege ismert, de az abban szereplő valamely paraméter értékét mintavétel útján kívánjuk megbecsülni. A becsléshez egy statisztikát konstruálunk. Természetesen nem használhatunk bármilyen statisztikát. Egy statisztikát **torzítatlan becslésnek** nevezünk, ha várható értéke a paraméter valódi értéke. Két becslés közül a kisebb szórásút tekintjük jobbnak, vagy **hatásosabbnak**.

5.2. példa: Adott statisztikus sokaság (elméleti) várható értéke m. Az n-elemű $(\xi_1, \xi_2, \dots, \xi_n)$ véletlen mintából az

$$X = \sum_{i=1}^{n} \lambda_i \xi_i$$

statisztikát képezzük olyan nemnegatív λ_i számokkal, amelyek összege 1. (X-et **súlyozott tapasztalati várható értéknek** nevezhetjük, a $\lambda_i = 1/n$ választás a szokásos tapasztalati várható értéket adja.)

A várható érték linearitása alapján

$$M(X) = \sum_{i=1}^{n} \lambda_i M(\xi_i) = \sum_{i=1}^{n} \lambda_i m = m,$$

ezért X az (elméleti) várhatóérték torzítatlan becslése. Kiszámítjuk X szórását.

$$\begin{aligned} M(X^2) &= M((\sum_i \lambda_i \xi_i)(\sum_j \lambda_j \xi_j)) = \sum_{i \neq j} \lambda_i \lambda_j M(\xi_i \xi_j) + \sum_i \lambda_i^2 M(\xi_i^2) \\ &= \sum_{i \neq j} \lambda_i \lambda_j m^2 + \sum_i \lambda_i^2 M(\xi^2) \end{aligned}$$

Mivel $\sum_{i \neq j} \lambda_i \lambda_j = 1 - \sum_i \lambda_i^2$, a $\sum \lambda_i = 1$ feltételből azt kapjuk, hogy

$$\sigma^2(X) = \sum_i \lambda_i^2 (M(\xi^2) - m^2).$$

X szórása akkor a lehető legkisebb, ha $\lambda_i = 1/n$. Valóban,

$$\sum_i \lambda_i^2 = \sum_i (\lambda_i - 1/n)^2 + \frac{1}{n}.$$

Tehát **a súlyozott tapasztalati várható értékek közül a közönséges tapasztalati várható érték a leghatékonyabb becslése az elméleti várható értéknek**. Utóbbi szórása $\sigma(\xi)/\sqrt{n}$, és az is látható, hogy ha nagyobb a mintaméret, hatékonyabb a becslés. □

Legyen $(\xi_1, \xi_2, \dots, \xi_n)$ egy statisztikus sokaságból vett n-elemű minta és

$$\bar{x}_n = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n} \tag{5.2.1}$$

a tapasztalati várható érték. A tapasztalati szórásnégyzet

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \bar{x}_n)^2 \tag{5.2.2}$$

olyan statisztika, amely a sokaság σ^2 (elméleti) szórásnégyzetének becslésére használható. Belátjuk, hogy

$$\boxed{M(s_n^2) = \frac{n-1}{n}\sigma^2.} \tag{5.2.3}$$

A tapasztalati szórásnégyzet definíciója alapján

$$M(ns_n^2) = \sum_{i=1}^{n}(M(\xi_i^2) - 2M(\xi_i\bar{x}_n) + M(\bar{x}_n^2)). \tag{5.2.4}$$

Most minden tagot külön kiszámolunk. $M(\xi_i^2) = M(\xi^2)$, mert a mintaelemek azonos eloszlásúak.

$$\begin{aligned} M(\xi_i\bar{x}_n) &= \frac{1}{n}M(\xi_i\sum_{j=1}^{n}\xi_j) = \frac{1}{n}M(\xi_i^2) + \frac{1}{n}\sum_{j\neq i}M(\xi_i\xi_j) \\ &= \frac{1}{n}M(\xi^2) + \frac{n-1}{n}m^2, \end{aligned}$$

mert a mintaelemek függetlensége miatt $M(\xi_i\xi_j) = M(\xi_i)M(\xi_j) = m^2$. Végül

$$M(\bar{x}_n^2) = \frac{1}{n^2}\sum_{j=1}^{n}M(\xi_j)^2 + \frac{1}{n^2}\sum_{k\neq j}M(\xi_k\xi_j) = \frac{1}{n}M(\xi^2) + \frac{n^2-n}{n^2}m^2.$$

Visszahelyettesítve (5.2.4)-be:

$$\begin{aligned} M(ns_n^2) &= \sum_{i=1}^{n}\Big(M(\xi^2) - \frac{2}{n}M(\xi^2) - \frac{2(n-1)}{n}m^2 + \frac{1}{n}M(\xi^2) + \frac{n^2-n}{n^2}m^2\Big) \\ &= (n-1)(M(\xi^2) - m^2) = (n-1)\sigma^2, \end{aligned}$$

amiből n-nel való osztással (5.2.3) adódik.

(5.2.3)-ból látható, hogy a tapasztalati szórásnégyzet torzított becslése az elméleti szórásnégyzetnek, de a torzítás mértéke n növekedésével csökken, és aszimptotikusan torzítatlan becslést kapunk. A példából a következő definíció fogalmazható meg. Legyen

$$X_n = F_n(\xi_1, \xi_2, \ldots, \xi_n) \tag{5.2.5}$$

az n-elemű mintán alapuló statisztika. Azt mondjuk, hogy az X_n becsléssorozat **aszimptotikusan torzítatlan becslése** a θ paraméternek, ha

$$\lim_{n\to\infty} M(X_n) = \theta.$$

Tehát a tapasztalati szórásnégyzetek sorozata példa aszimptotikusan torzítatlan becsléssorozatra.

5.3. példa: Legyen $(\xi_1, \xi_2, \ldots, \xi_n)$ a $[0,t]$ intervallumban egyenletes eloszlású sokaságból vett minta. Használhatjuk-e az $\xi_n^* = \max(\xi_1, \xi_2, \ldots, \xi_n)$ statisztikát az ismeretlen t paraméter becslésére?

A matematika nyelvére átfogalmazva az a kérdés, hogy torzítatlan becslése-e ξ_n^* a t-nek. Tehát ki kell számolnunk ξ_n^* várható értékét, és ehhez a sűrűségfüggvényét. Az eloszlásfüggvény definíciója szerint

$$F(x) = P(\xi_n^* < x) = P(\xi_1 < x, \xi_2 < x, \ldots, \xi_n < x).$$

Ez a valószínűség faktorizál, mert a mintaelemek függetlenek. Azonos eloszlásúak is, ezért

$$F(x) = P(\xi_1 < x)^n = \begin{cases} 0 & \text{ha } x \le 0, \\ (x/t)^n & \text{ha } 0 < x \le t, \\ 1 & \text{ha } 1 < x. \end{cases}$$

Differenciálva kapjuk ξ_n^* f sűrűségfüggvényét:

$$f(x) = \begin{cases} nx^{n-1}t^{-n} & \text{ha } 0 < x \le t \\ 0 & \text{egyébként.} \end{cases}$$

Most ki tudjuk számolni ξ_n^* várható értékét,

$$M(\xi_n^*) = \int_0^t nx^n t^{-n}\,dx = \frac{n}{n+1}t\,.$$

Tehát ξ_n^* várható értéke nem t, így ξ_n^* torzított becslése a t paraméternek. Ha n nő, akkor a torzítás csökken. Mivel $n/(n+1)$ határértéke 1, a becslés aszimptotikusan torzítatlan.

Azt mondhatjuk, hogy nagy n-re az ξ_n^* statisztika használható t becslésére, de legjobb, ha a korrigált

$$Y_n = \frac{n+1}{n}\max(\xi_1, \xi_2, \ldots, \xi_n) \tag{5.2.6}$$

statisztikát használjuk, ez ugyanis torzítatlan becslés lévén bármilyen n-re jól használható.

A t paraméternek a $2\bar{x}_n$ statisztika is torzítatlan becslése. Melyiket érdemesebb használni? Y_n-t vagy a tapasztalati várható érték kétszeresét? Az a hatékonyabb becslés, amelynek kisebb a szórása. Számoljuk ki tehát a szórásokat!

A $[0,t]$ intervallumban egyenletes eloszlás szórásnégyzete a 2.11. példa alapján $t^2/12$. Ezért $2\bar{x}_n$ szórásnégyzete $t^2/3n$. Ezt kell összehasonlítani Y_n szórásnégyzetével.

$$M(\xi_n)^2 = \int_0^t nx^{n+1}t^{-n}\,dx = \frac{n}{n+2}t^2\,,$$

és

$$\sigma^2(Y_n) = \Big(\frac{n+1}{n}\Big)^2 \frac{n}{n+2}t^2 - t^2 = \Big(\frac{(n+1)^2}{n(n+2)} - 1\Big)t^2 = \frac{t^2}{n(n+2)}\,.$$

Nagy mintaméretre, azaz nagy n-re, $t^2/n(n+2)$ sokkal kisebb mint $t^2/3n$, tehát az Y_n statisztika jóval hatékonyabb. □

Legyen η_1 és η_2 két olyan valószínűségi változó, hogy $\eta_1 < \eta_2$. Ekkor (η_1, η_2) egy véletlen elhelyezkedésű intervallum, ami tartalmazhatja a becsülendő statisztikai paramétert. Azt mondjuk, hogy (η_1, η_2) **megbízhatósági intervallum** az ismeretlen paraméter számára, az x %-os szinten, ha annak a valószínűsége, hogy a paraméter η_1 és η_2 közé esik legalább $x/100\%$.

5.4. példa: Normális eloszlású statisztikai sokaság szórása σ és várható értéke ismeretlen. Keressünk megbízhatósági intervallumot a tapasztalati várható értékhez 95%-os megbízhatósági szinten!

Olyan $c > 0$ számot keresünk, amelyre

$$P(|\bar{x}_n - m| < c) \leq 0.95$$

teljesül. Az $\bar{x}_n - m$ valószínűségi változó szórása $\sigma/\sqrt{n}$, várható értéke 0 és normális eloszlású. Standardizálunk, hogy a standard normális eloszlás eloszlásfüggvényét használhassuk.

$$\begin{aligned} P(|\bar{x}_n - m| < c) &= P\Big(\frac{|\bar{x}_n - m|\sqrt{n}}{\sigma} < \frac{c\sqrt{n}}{\sigma}\Big) = \Phi\Big(\frac{c\sqrt{n}}{\sigma}\Big) - \Big(1 - \Phi\Big(\frac{c\sqrt{n}}{\sigma}\Big)\Big) \\ &= 2\Phi\Big(\frac{c\sqrt{n}}{\sigma}\Big) - 1 \end{aligned}$$

Ez akkor lesz 0.95, ha

$$\Phi\Big(\frac{c\sqrt{n}}{\sigma}\Big) = 0.975 = \Phi(1.96)$$

azaz $c\sqrt{n}/\sigma = 1.96$ és $c = 1.96\sigma/\sqrt{n}$.

Tekintsük most az $n = 20$ mintaméretet az 5.1. példa számszerű adataival és tételezzük fel, hogy $\sigma = 0.013$. Ekkor $c = 0.025/4.472 = 0.006$. Az 5.1. példában $\bar{x}_n = 0.114$. Analízisünk eredményét így foglalhatjuk össze: 95%-os megbízhatósággal állíthatjuk, hogy a mért paraméter értéke $0.114 - 0.006 = 0.108$ és $0.114 + 0.006 = 0.120$ közé esik. □

A gyakorlatban használt megbízhatósági szintek 90%, 95% és 99%. Túlságosan nagy megbízhatósági szintet általában azért nem érdemes megkövetelni, mert akkor a megbízhatósági (vagy konfidencia) intervallum hossza erősen megnőhet. Az előző példában a 99.9% megbízhatósági szinthez már a $c = 0.010$ érték tartozik, ami csak a nagyon kis szórás és a nagy mintaméret miatt nem ad sokkal hoszabb megbízhatósági intervallumot.

5.5. példa: Legyen ξ és η két valószínűségi változó. Gondoljunk arra, hogy ξ valamely, Budapesten eladott lakás alapterülete, η pedig ugyannek a lakásnak az eladási ára. η értékére becslést kaphatunk ξ ismeretében a lineáris regressziót használva:

$$\eta \approx M(\eta) + \frac{\mathrm{Cov}(\xi, \eta)}{\sigma^2(\xi)}\Big(\xi - M(\xi)\Big)$$

Ennek a képletnek a segítségével megbecsülhetjük egy lakás eladási árát az alapterülete ismeretében, feltéve persze, hogy a képletben szereplő $M(\eta)$ várható értéket és a

$$\frac{\mathrm{Cov}(\xi, \eta)}{\sigma^2(\xi)}$$

regressziós együtthatót ismerjük. Ha adott a $(\eta_1, \eta_2, \ldots, \eta_n)$ statisztikai minta, akkor $M(\eta)$ becslésére használhatjuk az empirikus várható értéket. Ahhoz, hogy a regressziós együtthatót becsüljük, természetesen a ξ-re vonatkozó mintát is ismernünk kell. Legyen ez $(\xi_1, \xi_2, \ldots, \xi_n)$. (Tehát ξ_i annak a lakásnak az eladási ára, amelynek alapterülete η_i.) A regressziós együttható szokásos becslése

$$\frac{\sum_{i=1}^{n}(\xi_i - \bar{\xi})(\eta_i - \bar{\eta})}{\sum_{i=1}^{n}(\xi_i - \bar{\xi})^2}. \tag{5.2.7}$$

Így
$$\eta \approx \bar{\eta} + \frac{\sum_{i=1}^{n}(\xi_i - \bar{\xi})(\eta_i - \bar{\eta})}{\sum_{i=1}^{n}(\xi_i - \bar{\xi})^2}\Big(\xi - \bar{\xi}\Big). \tag{5.2.8}$$
Itt a jobb oldalon csak a mintaelemek szerepelnek. Ez a képlet a legkisebb négyzetek elvén alapuló **egyenesillesztés kapcsán** is ismert. Annak az egyenesnek az egyenlete a ξ-η síkon, amely „leginkább" a (ξ_1, η_1), (ξ_2, η_2), $\ldots, (\xi_n, \eta_n)$ pontok közelében halad. Természetesn az (5.2.8) képletnek csak abban az esetben van haszna, ha ξ és η között a kapcsolat megközelítőleg lineáris. □

5.3. Normális eloszlású sokaságok

Tegyük fel, hogy a statisztikai sokaság $N(m, \sigma)$ eloszlású. Ekkor az (5.2.1)-gyel értelmezett $\bar{x}_n$ tapasztalati várható érték is normális eloszlású a 3.7 tétel általánosítása alapján. $\bar{x}_n$ eloszlása $N(m, \sigma/\sqrt{n})$. A tapasztalati szórásnégyzet eloszlásáról szól a következő tétel.

5.2. tétel: *$N(m, \sigma)$ eloszlású sokaság esetén ns_n^2/σ^2 $(n-1)$-szabadságfokú χ^2-eloszlást követ.*

Az egyszerűség kedvéért csak az $n = 2$ esetet tekintjük.
$$2s_2^2 = \Big(\xi_1 - \frac{\xi_1 + \xi_2}{2}\Big)^2 + \Big(\xi_2 - \frac{\xi_1 + \xi_2}{2}\Big)^2 = \Big(\frac{\xi_1 - \xi_2}{\sqrt{2}}\Big)^2,$$
és ezért
$$\frac{2s_2^2}{\sigma^2} = \Big(\frac{\xi_1 - \xi_2}{\sqrt{2}\sigma}\Big)^2.$$
Azt kell látnunk, hogy $(\xi_1 - \xi_2)/\sqrt{2}\sigma$ standard nomális eloszlású. Ez a tény a 3.7. tételből következik.

5.3. tétel: *Normális eloszlású sokaság esetén a tapasztalati várhatóérték és a tapasztalati szórásnégyzet egymástól független statisztikák.*

A tételt nem bizonyítjuk. $\bar{x}_n$ és s_n^2 függetlensége a normális eloszlású statisztikus sokaságok jellemzője. (Más eloszlások esetén a függetlenség általában nem áll fenn.) Az $n = 2$ mintaméret esetén a tétel a
$$\frac{\xi_1 + \xi_2}{\sqrt{2}} \quad \text{és} \quad \frac{\xi_1 - \xi_2}{\sqrt{2}}$$
valószínűségi változók függetlenségét állítja. A kinetikus gázelméletből vett példával azt gondolhatjuk, hogy ξ_1 egy molekula x irányú sebessége, ξ_2 pedig az y irányú sebessége. A sebességkomponensek függetlensége bármilyen más derékszögű koordinátarendszerben fenn kell, hogy álljon. Nézzük, most azt az (x', y', z) koordináta rendszert, amiben az x' irányt az $(1/\sqrt{2}, 1/\sqrt{2}, z)$ vektor, az y' irányt pedig az $(1/\sqrt{2}, -1/\sqrt{2}, z)$ vektor adja meg. Jól látható, hogy x' és y' merőlegesek lesznek. Ezek után vegyük észre, hogy $(\xi_1 + \xi_2)/\sqrt{2}$ a részecske x' irányú sebessége, $(\xi_1 - \xi_2)/\sqrt{2}$ pedig az y' irányú sebessége. Merőleges irányokról lévén szó a két sebességösszetevő független. Ez az okoskodás természetesen nem matematikai bizonyítás, de pontosan rámutat arra, hogy a tétel miért igaz.

Ha ξ és η olyan független valószínűségi változók, hogy ξ standard normális eloszlású és η n-szabadságfokú χ eloszlást követ, akkor $\sqrt{n}\xi/\eta$ eloszlását n-**szabadságfokú t-eloszlásnak** nevezzük.

A t-eloszlást W.S. Gosset vizsgálta először 1908-ban. Egy ír sörfőzde alkalmazottjaként nem publikálhatta statisztikai vizsgálatai eredményét, és „Student” álnéven írt. Ezért nevezik a t-eloszlást gyakran Student-eloszlásnak.

A legfontosabb példát a t-eloszlásra a normális eloszlású sokaság szolgáltatja. Az $N(m, \sigma)$ normális sokaság esetén a

$$t = \sqrt{n-1}\frac{\bar{x}_n - m}{s_n} \tag{5.3.1}$$

statisztika Student-eloszlású, $(n-1)$-szabadságfokkal. Valóban, mivel az **??** tétel szerint $\sqrt{n}s/\sigma$ $(n-1)$-szabadságfokú χ-eloszlást követ, $(\bar{x}_n - m)\sqrt{n}/\sigma$ standard normális eloszlású, és az 5.3. tétel szerint s_n és $\bar{x}_n$ függetlenek.

A t-eloszlás sűrűségfüggvénye

$$g(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \tag{5.3.2}$$

és ennek grafikonja harang alakú, emlékeztet a normális eloszlásra. Ennél sokkal több is igaz. Ha a szabadságfok tart a végtelenhez, akkor az n-szabadságfokú t-eloszlás a standard normális eloszláshoz konvergál. Megemlítjük, hogy az n-szabadságfokú t-eloszlás varianciája $n/(n-2)$.

5.4. Hipotézisvizsgálat és statisztikai próbák

A statisztikaihipotézis vizsgálatban nem az ismeretlen paraméter értékét akarjuk megtudni (mint a becslések alkalmazásakor), hanem a paraméter értékére vonatkozó hipotézisünket akarjuk ellenőrizni. Az ellenőrzendő hipotézis lehet például az, hogy egy darabológép által levágott csődarabok átlagos hossza 115 mm és 123 mm közé esik. Az ellenőrzendő hipotézist a statisztikusok rendszerint H_0-lal jelölik, és a neve nullahipotézis. H_1 jelöli az alternatív hipotézist, ami a fenti példában az lehet, hogy a csődarabok átlagos hossza 123 mm-nél nagyobb. A statisztikai hipotézisvizsgálat abban áll, hogy a véletlen minta alapján eldöntjük, hogy elfogadjuk a H_0 hipotézist, vagy elvetjük. A döntéskor kétféle hibát követhetünk el. A hipotézis valójában igaz, de mi tévesen elvetjük, ez az **első típusú hiba**. A másik tévedési lehetőség az, hogy elfogadjuk a hipotézist, de az valójában hamis. Ez a **második típusú hiba**. Természetesen mindkét hiba valószínűségét kicsinek kívánjuk. A hipotézisvizsgálat a **statisztikai próbán** keresztül valósul meg. Példaként tekintsünk egy palackozó gépet, amely 200 ml vegyi anyagot tölt egy-egy palackba. Statisztikai próbával kívánjuk eldönteni, hogy az egy palackba jutó átlagos mennyiség valóban 200 ml-e. Ezt választjuk nullahipotézisnek, szemben azzal az alternatívával, hogy az átlag nem 200 ml. Feltételezzük, hogy az egy palackba töltött anyag mennyisége normális eloszlást mutat. A próbához az (5.3.1)-gyel adott t-statisztikát használjuk. Matematikailag t egy valószínűségi változó, amelynek eloszlása ismert. Ha n a minta elemszáma, akkor t $(n-1)$ szabadságfokú Student-eloszlást mutat. Kell választanunk egy $0 < p < 1$ számot úgy, hogy a p valószínűségű esemény gyakorlatilag elhanyagolható legyen. $p = 0,005$ elfogadható, és az $(1-p)100\% = 99,5\%$-ot a **szignifikancia szintjének** nevezzük. Ez adja meg, hogy milyen bizonyossággal akarjuk a próbát alkalmazni. A Student-eloszlás táblázatából kikeressük azt a t_p értéket, amelyre

$$P(|t| > t_p) = p \tag{5.4.1}$$

Példaként gondoljunk arra, hogy 25 palackot vizsgáltak meg, és azt találták, hogy azok átlagosan $\bar{x} = 200,3$ ml anyagot tartalmaztak, és a tapasztalati szórás $s = 0,04$ ml volt. Ekkor

$$t = \sqrt{24}\frac{200,3 - 200}{0,4} = 3.67$$

A 24 szabadságfokú Student-eloszlás táblázatából $t_{0,005} = 2,797$, ami kevesebb, mint 3.67. A próba alkalmazásakor tehát olyan esemény következett be, amelynek valószínűsége $p = 0,005$-nél kisebb. Mivel ilyen kis valószínűségű eseményt gyakorlatilag lehetetlennek ítélünk, el kell vetni azt a hipotézist, hogy az egy palackba jutó mennyiség átlagosan 200 ml. A fent leírt próba a **t-próba**. Akkor alkalmazható, ha a sokaság normális eloszlású, szórása ismeretlen, és a várható értékre vonatkozó hipotézist kell ellenőrizni.

5.6. példa: Egy vegyész a titán meghatározására olcsó, gyors és pontos eljárást fejleszt ki. Módszere pontosságát bemutatandó egy mintán 50 egymástól független vizsgálatot végez, amelyek átlaga 0.0095 ppm és varianciája $81.0 \cdot 10^{-8}$. Az új módszerrel vizsgált anyagot a régi, drága de beváltan nagyon pontos vizsgálatnak alávetve azt kapják, hogy 0.0093 ppm titánt tartalmaz. Döntsük el, hogy a szignifikancia 95%-os szintjén van-e okunk kételkedni az új módszer pontosságában!

Ismert a tapasztalati szórás, a tapasztalati várható érték, és a várható értékre vonatkozó hipotézist kell ellenőrizni. Ha normális eloszlást tételezünk fel, akkor a t-próba alkalmazható. (5.3.1)-be behelyettesítve

$$t = \sqrt{49}\frac{0.0095 - 0.0093}{9 \cdot 10^{-4}} = 1.55556\,.$$

Az (5.4.1) egyenlőséghez tartozó t_p értéket a 49 szabadságfokú t-eloszlás táblázatából vesszük: $p = 0.05$ és $t_p = 2.01$.

$$t = 1.62 < t_p = 2.01\,,$$

ezért az eltérés nem szignifikáns a 95%-os szinten, nem kell elvetnünk a 0-hipotézist. □

Gyakorlatok

5.1. gyakorlat: Legyen $(\xi_1, \xi_2, \ldots, \xi_6)$ egy $N(0,1)$ sokaságból vett minta. A c állandó milyen értékére lesz az

$$x = c(\xi_1 + \xi_2 + \xi_3)^2 + c(\xi_4 + \xi_5 + \xi_6)^2$$

statisztika χ^2-eloszlású?

5.2. gyakorlat: Legyen $(\xi_1, \xi_2, \ldots, \xi_n)$ egy $N(m,\sigma)$ sokaságból vett minta. Mi lesy az

$$n(\bar{x}_{(n)} - m)^2/\sigma^2$$

eloszlása?

5.3. gyakorlat: Bizonyítsa be, hogy a

$$Y = \frac{\xi_1^* + \xi_n^*}{2}$$

statisztika egyenletes eloszlású sokaság esetén torzítatlan becslése az elméleti várható értéknek!

5.4. gyakorlat: Mi a valószínűsége, hogy a normális eloszlású $\sigma = 10$ szórású sokaságból vett 5 elemű minta varianciája több, mint 120? Használja fel, hogy a 4 szabadságfokú χ^2-eloszlás sűrűségfüggvénye

$$f(x) = \begin{cases} \frac{1}{4}xe^{-x/2} & \text{ha } x > 0, \\ 0 & \text{egyébként.} \end{cases}$$

5.5. gyakorlat:$^{\mathbf{M}}$ $\xi_1, \xi_2, \ldots, \xi_5$ független standard normális eloszlású valószínűségi változók. Milyen c állandóra lesz

$$\frac{c(\xi_1 + \xi_2)}{(\xi_3^2 + \xi_4^2 + \xi_5^2)^{1/2}}$$

t-eloszlású valószínűségi változó?

5.6. gyakorlat: Adjon részletes bizonyítást arra, hogy az n-szabadságfokú t-eloszlás a standard normális eloszláshoz tart, ha $n \to \infty$!

5.7. gyakorlat:$^{\mathbf{M}}$ ξ és η független normális eloszlású valószínűségi változók 0 várható értékkel és 2 varianciával. Határozza meg a

$$P\left(\frac{(\xi + \eta)^2}{(\xi - \eta)^2} < 4\right)$$

értékét!

5.8. gyakorlat: 50 hallgató megmérte az alumínium fajhőjét. A méréseikből számolt (tapasztalati) átlagérték 0.2210 (cal/g°C) és a tapasztalati szórás 0.0240 (cal/g°C). Mit mondhatunk a megbízhatóság 95%-os szintjén a hibáról, ha az átlagértéket tekintjük az alumínium igazi fajhőjének?

5.9. gyakorlat: Egy statisztikai adat azt mondja, hogy az írógépek 30%-át egy A cég gyártja. Ugyanakkor egy felmérés azt mutatja, hogy 500 megnézett írógépből 118-at gyártott az A cég. Kételkedünk-e a 30%-os adatban a szignifikancia 99%-os szintjén?

5.10. gyakorlat: 0.013 szórású normálsi eloszlású statisztikai sokaság tapasztalati várható értékeihez keresünk megbízhatósági intervallumot a 99%-os szinten. Mekkora legyen a mintaméret, ha a 0.002 hosszúságú intervallumot akarunk biztosítani?

5.11. gyakorlat:$^{\mathbf{M}}$ $\eta_1, \eta_2, \ldots, \eta_5$ standard normális eloszlású független valószínűségi változók. Milyen eloszlása van a

$$\frac{2\eta_5}{\sqrt{\eta_1^2 + \eta_2^2 + \eta_3^2 + \eta_4^2}}$$

valószínűségi változónak?

5.12. gyakorlat: Mikor nevezünk egy becsléssorozatot aszimptotikusan torzítatlannak? Mondjon példát! Mit értünk azon, hogy egy becslés hatékonyabb a másiknál? Mondjon példát!

5.13. gyakorlat:$^{\mathbf{M}}$ $\xi_1, \xi_2, \ldots, \xi_6$ 0 várható értékű és 2 szórású statisztikai sokaságból vett 6 elemű véletlen minta. Milyen c számra lesz a

$$c(\xi_1 + \xi_2 + \xi_3)^2 + c(\xi_4 + \xi_5 + \xi_6)^2$$

statisztika χ^2 eloszlású?

5.14. gyakorlat: Mit nevezünk torzítatlan statisztikai becslésnek? Egy statisztikus sokaság eloszlásáról tudjuk, hogy egyenletes a $[-t, t]$ intervallumon, de a t paramétert nem ismerjük. Javasoljon statisztikát a t paraméter becslésére! Indokolja javaslatát!

5.15. gyakorlat: M Normális eloszlású statisztikus sokaság szórását ismerjük, 0.13. Mekkora legyen a mintaméret, hogy az ismeretlen elméleti várható értékhez a tapasztalati várható értékből a 95%-os szinten 0.1 hosszúságú megbízhatósági intervallumot kapjunk?

5.16. gyakorlat: $(\xi_1, \xi_2, \xi_3, \xi_4)$ egy $[0, t]$ intervallumban egyenletes eloszlású statisztikus sokaságból vett véletlen minta. Az

$$\eta_1 = \frac{\xi_1 + \xi_2 + \xi_3 + \xi_4}{4}, \quad \eta_2 = \frac{\xi_1 + 2\xi_2 + \xi_3 + 2\xi_4}{3}, \quad \eta_3 = \xi_1 + \xi_2$$

statisztikák közül melyiket használná az ismeretlen t paraméter becslésére? Részletesen indokolja válaszát!

5.17. gyakorlat: M Exponenciális eloszlású sokaságból vett 3 elemű minta (ξ_1, ξ_2, ξ_3). Használható-e az

$$\eta = \mathrm{MAX}(\xi_1, \xi_2, \xi_3)$$

statisztika a sokaság várható értékének becslésére? (Az η valószínűségi változó a mintaelemek legnagyobbikát jelöli.)

5.18. gyakorlat: M A $[0, t]$ intervallumban egyenletes eloszlású statisztikus sokaság t paramétere ismeretlen. A $(\xi_1, \xi_2, \xi_3, \xi_4)$ véletlen mintán alapuló

$$\eta_1 = \xi_1 + \xi_2 + \xi_3 - \xi_4, \qquad \eta_2 = \xi_1 + \xi_4$$

statisztikák közül melyik előnyösebb t becslésére? (Miért?)

5.19. gyakorlat: Magyarázza meg, hogy (5.2.7) miért becslése a regressziós együtthatónak!

Hatodik fejezet

Megoldások

6.1. Az első fejezet gyakorlatai

1.1. gyakorlat megoldása:

$A \cdot B = \emptyset$, mivel egymást kizáró események, tehát $P(A \cdot B) = 0$.

(a) $P(\overline{A}) = 1 - P(A)$, $P(\overline{A}) = 0,8$.

(b) $P(\overline{A} \cdot \overline{B}) = P\left(\overline{(A+B)}\right) = 1 - P(A+B) = 1 - (P(A) + P(B) - P(A \cdot B))$ és $P(\overline{A} \cdot \overline{B}) = 0,25$ (De Morgan azonosság).

(c) $P(A + B) = P(A) + B(B)$ (Kolmogorov III. axióma), $P(A + B) = 0,75$.

(d) $P(A \cdot B) = 0$ (a feltételekből következően). □

1.3. gyakorlat megoldása:

Visszatevés nélkül egymás után kivesszük a dobozból a H illetve S jelű elemeket. $P(HHHHHHHSSS) =?$

(a) Megoldás a klasszikus valószínűségelmélet alapján: A H jelű ill. S jelű elemek nem különböztethetők meg egymás közt. Az elemi események a 10 db (7 db H és 3 db S) elem ismétléses permutációi.

$$P = \frac{\text{kedvező esetek száma}}{\text{összes esetek száma}} = \frac{1}{\frac{10!}{7!3!}} = \frac{1}{120}$$

(b) Ha a különböző jelű elemek egymás közt megkülönböztethetők, akkor a feladat ekvivalens a 10 db különböző elemmel kapcsolatos alábbi feladattal:

Mennyi annak a valószínűsége, hogy egy urnából, melyben 7 db piros az 1, 2, ..., 7 számokkal sorszámozott és 3 db az 1, 2, 3 számokkal sorszámozott zöld golyó van, egymás után kivéve mind a 10 db golyót, az első 7 db golyó piros lesz?

Az elemi események ebben az esetben úgy származtathatók, hogy 10 db lehetséges helyből kiválasztunk hármat (az összes lehetséges módon) a zöld golyó számára, majd a kiválasztott helyekre az összes lehetséges módon elhelyezzük a zöld golyókat.

10 elemből $\binom{10}{3}$ féleképpen választhatunk ki hármat (kombináció), az 1, 2, 3 sorszámú zöld golyókat pedig 3! féleképpen helyezhetjük el a három ,,éppen kiválasztott" helyre.

A 7 db sorszámozott piros golyót pedig 7! féleképpen helyezhetjük el a fennmaradó 7 helyre. Az összes lehetséges esetek száma: $\binom{10}{3} \cdot 3! \cdot 7!$
A kedvező esetek száma: $7! \cdot 3!$

$$P = \frac{7! \cdot 3!}{\binom{10}{3} \cdot 3!7!} = \frac{1}{\frac{10\cdot 9\cdot 8}{1\cdot 2\cdot 3}} = \frac{1}{120}$$

(c) A feladat modellezhető az alábbi módon is. Adott 10 db urna.

$7H3S$	$6H3S$	$5H3S$	$4H3S$	$3H3S$	$2H3S$	$1H3S$	$3S$	$2S$	$1S$
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.

Mindegyik urnából az 1. sorszámúval kezdve és az elhelyezési sorrendet betartva – véletlenszerűen – választunk egy elemet. Kérdés: mennyi annak a valószínűsége, hogy a húzás során az első 7 elem H, az utolsó 3 pedig S? A megoldás során az egymás utáni húzásokat egy független kísérletsorozat elemeinek tekinthetjük, azaz, ha $E_1, E_2, \ldots, E_{10}$ rendre az egyes húzások során bekövetkező események, akkor $P(E_1 \cdot E_2 \cdot E_3 \cdot \ldots \cdot E_{10}) = P(E_1)P(E_2)\ldots P(E_{10})$. Ezzel a keresett valószínűség:

$$P = \frac{7}{10} \cdot \frac{6}{9} \cdot \frac{5}{8} \cdot \frac{4}{7} \cdot \frac{3}{6} \cdot \frac{2}{5} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{3}{3} \cdot \frac{2}{2} \cdot \frac{1}{1} = \frac{1}{120}.$$

(d) A feladat az ún. szorzási szabály segítségével is megoldható.

$$\begin{aligned} & P(H \cdot H \cdot H \cdot H \cdot H \cdot H \cdot H \cdot S \cdot S \cdot S) \\ = & \, P(H_1) \cdot P(H_2|H_1) \cdot P(H_3|H_1 \cdot H_2) \cdot P(H_7|H_1 \cdots H_6) \\ & \cdot P(S_1|H_1 \cdots H_7) \cdot P(S_2|H_1 \cdots H_7 S_1) \cdot P(S_3|H_1 \cdots H_7 S_1 S_2). \end{aligned}$$

□

1.5. gyakorlat megoldása:

Legyen A az az esemény, amelynek a valószínűségét keressük. Ha A megvalósul, akkor az 1-est megdobhatjuk az első, a második, a harmadik, …, az n-edik és így tovább dobásra. Legyen A_n az az esemény, hogy n-edikre 1-est dobunk és előtte nem dobunk sem párosat, sem 1-est. Ekkor

$$A = A_1 + A_2 + \ldots + A_n + \ldots.$$

Mivel ezek egymást kizáró események

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

Ki kell számolnunk a $P(A_n)$ valószínűséget. A_n független események szorzata: az első, a második, a harmadik, …, az $(n-1)$-edik dobásra a $\{3, 5\}$ esemény valósul meg, az n-edik dobásra pedig az $\{1\}$ esemény. Ezek valószínűsége $2/6$ és $1/6$. Függetlenség esetén a valószínűségek szorzódnak, tehát

$$P(A_n) = \left(\frac{2}{6}\right)^{n-1} \frac{1}{6},$$

és

$$P(A) = \frac{1}{6}\left(1 + \frac{2}{6} + \left(\frac{2}{6}\right)^2 + \ldots + \left(\frac{2}{6}\right)^k + \ldots\right) = \frac{1}{6} \cdot \frac{1}{1-\frac{1}{3}} = \frac{1}{6} \cdot \frac{3}{2} = \frac{1}{4}$$

a végtelen mértani sor összegzésével. (A $\sum\limits_{n=0}^{\infty} a_1 q^n$ végtelen mértani sor összege: $S = a_1 \cdot \frac{1}{1-q}$.)

Megjegyzés: A „klasszikus valószínűségelmélet” nem használható a feladat megoldására, mivel az elemi események száma nem véges. □

1.6. gyakorlat megoldása:

Azt az eseményt, hogy a számok összege legalább 58 három, egymást páronként kizáró esemény összegére bonthatjuk: $A =$ (a számok összege 58), $B =$ (a számok összege 59), $C =$ (a számok összege 60). Így

$$P(\text{a számok összege legalább 58}) = P(A) + P(B) + P(C).$$

A C esemény csak úgy valósulhat meg, ha mind a 10 kocka 6-ost mutat, így a kedvező lehetőségek száma 1. Az összes lehetőségek száma 6 elem 10-ed osztályú ismétléses kombinációi száma, tehát

$$\binom{6+10-1}{10}$$

(Megjegyezzük, hogy kombinációkat kell számolni, mert a kockák egymástól nem megkülönböztethetők, tehát sorrendnek nincs is értelme.)

$$P(A) = \frac{1}{\binom{15}{10}}$$

A B esemény úgy valósulhat meg, ha 9 db. 6-os és 1db. 5-ös adódik. Mivel sorrendnek nincs értelme, most is csak 1 kedvező lehetőség van:

$$P(B) = \frac{1}{\binom{15}{10}}$$

A C esemény már kétféleképpen jöhet létre: 1. 8 db. 6-os és 2 db. 5-ös 2. 9 db. 6-os és 1db 4-es. Tehát

$$P(C) = \frac{2}{\binom{15}{10}}.$$

1.7. gyakorlat megoldása:

Megoldás a teljes valószínűség tétele alapján. Jelölje A azt az eseményt, hogy „1” vagy „2” a kockadobás eredménye, B pedig a „3”, „4”, „5”, „6” dobások bekövetkezését! Akkor

$$P(A) = \frac{2}{6} = \frac{1}{3} \qquad \text{és} \quad P(B) = \frac{4}{6} = \frac{2}{3}.$$

Mivel $A \cdot B = \emptyset$ és $A + B = \Omega$, A és B teljes eseményrendszert alkot.

A teljes valószínűség tétele szerint

$$P(F) = P(F|A) \cdot P(A) + P(F|B) \cdot P(B).$$

Mivel

$$P(A) = \frac{1}{3}, \quad P(B) = \frac{2}{3}, \quad P(F|A) = \frac{6}{9}, \quad P(F|B) = \frac{1}{2},$$

így

$$P(F) = \frac{6}{9} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{2}{3} = \frac{5}{9}.$$

Megjegyzés: Az eredeti feladat átfogalmazható a következő módon. Adott 2 urna, 6 fehér és 3 piros illetve 4 fehér és 4 piros golyóval. Ezek közül véletlenszerűen választunk egyet. Az I. urna választásának valósznűsége $\frac{1}{3}$, a II. urnáé $\frac{2}{3}$. □

1.9. gyakorlat megoldása:

A feladat állítása azt jelenti, hogy $p_{2n} \geqslant p_{2(n+1)}$, vagyis ha $2(n+1)$-szer dobunk, akkor annak a valószínűsége, hogy a dobott fejek és írások száma továbbra is egyenlő, kisebb lesz, mint $2n$ dobás esetén.

p_{2n} értéke adott n esetére, a klasszikus valószínűségelmélet alapján számítható, mivel az eseményteret az F és I elemekből álló $2n$ hosszúságú sorozatok alkotják. Másrészt F és I bekövetkezése – szabályos érmét feltételezve – egyenlő valószínűségű. Tehát minden elemi esemény azonos valószínűségű, ezért

$$\frac{\text{kedvező esetek száma}}{\text{összes esetek száma}} = p_{2n}$$

A kedvező esetek száma annyi, ahány féleképpen ki tudunk választani a $2n$ szamú lehetséges helyből n számút (pl.) a fejek (F) számára. Ezek száma: $\binom{2n}{n}$. ($2n$ elem n-ed osztályú, ismétlés nélküli kombinációinak száma.)

Az összes lehetséges esetek száma (az eseménytér elemszáma): 2^{2n}.

Ezzel

$$p_{2n} = \frac{\binom{2n}{n}}{2^{2n}}.$$

Bizonyítani kell, hogy

$$\frac{\binom{2n}{n}}{2^{2n}} \geqslant \frac{\binom{2(n+1)}{n+1}}{2^{2(n+1)}}.$$

Ha a fenti egyenlőtlenséget megoldjuk n-re, akkor eldönthető, hogy p_{2n} monoton növekvő sorozat-e. Használva $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ definíciót és az egyenlőtlenségbe behelyettesítve: $2(n+1) > 2n+1$ adódik, és ez minden n-re igaz. □

1.11. gyakorlat megoldása:

Azt az eseményt, hogy a fejek száma ugyanannyi, 4 egymást páronként kizáró eseményre bonthatjuk. Jelölje A_i azt az eseményt, hogy mindkét ember i-szer dob fejet. i lehetséges értékei 0,1,2,3. A keresett valószínűség

$$P(A_0) + P(A_1) + P(A_2) + P(A_3)$$

összeg, amelynek tagjait ay alábbi módon kiszámoljuk.

Annak a valószínűsége, hogy az egyik ember nem dob fejet 1/8, annak a valószínűsége, hogy a másik nem dob fejet szintén 1/8. Mivel a két ember egymástól függetlenül dob, annak a valószínűsége, hogy egyikük sem dob fejet $1/8 \cdot 1/8$. Tehát

$$P(A_0) = \frac{1}{64}.$$

Annak a valószínűsége, hogy az egyik ember a három dobásból egyszer dob fejet 3/8, ezért

$$P(A_1) = \frac{9}{64}.$$

Hasonlóan

$$P(A_2) = \frac{9}{64}, \qquad P(A_3) = \frac{1}{64},$$

és így a keresett valószínűség 20/64.

1.12. gyakorlat megoldása:

Jelölje P_1, P_2, P_3 azt az eseményt, hogy elsőre, másodikra, harmadikra pirosat húztunk! A $P(P_1|P_2 \cdot P_3)$ feltételes valószínűséget keressük. Definíció szerint:

$$P(P_1|P_2 \cdot P_3) = \frac{P(P_1 \cdot P_2 \cdot P_3)}{P(P_2 \cdot P_3)}.$$

$P(P_1P_2P_3) = P(P_3|P_1 \cdot P_2) \cdot P(P_1 \cdot P_2)$, ahol $P(P_1P_2) = P(P_2|P_1) \cdot P(P_1)$ és $\Omega = P_1 + \overline{P}_1$
$P(P_2P_3) = P(P_3P_2(P_1 + \overline{P}_1)) = P(P_1 \cdot P_2 \cdot P_3 + \overline{P}_1P_2P_3) = P(P_1P_2P_3) + P(\overline{P}_1P_2P_3)$

A teljes valószínűség tétele szerint $P(P_2) = P(P_2 \cdot \Omega) = P(P_2(P_1 + \overline{P}_1)) = P(P_2 \cdot P_1) + P(P_2 \cdot \overline{P}_1)$
$P(P_2P_1) = P(P_2|P_1) \cdot P(P_1)$
$P(\overline{P}_1P_2P_3) = P(P_3|\overline{P}_1P_2) \cdot P(P_2|\overline{P}_1) \cdot P(\overline{P}_1)$

A szükséges adatok: $P(P_1) = \frac{8}{32}$; $P(\overline{P}_1) = \frac{3}{4}$; $P(P_2|P_1) = \frac{7}{31}$; $P(P_2|\overline{P}_1) = \frac{8}{31}$; $P(P_3|P_1 \cdot P_2) = \frac{6}{30}$ és $P(P_3|P_2 \cdot \overline{P}_1) = \frac{7}{30}$. Behelyettesítés után: $P(P_1|P_2P_3) = \frac{1}{5}$

Megjegyzés:

(1) A megoldás alapja lehet a Bayes-tétel.

(2) A feladathoz tartozó eseménytér felépíthető a P és $\overline{P}$ elemekből:

$$\Omega = \{PPP,\ PP\overline{P},\ P\overline{P}P,\ \overline{P}PP,\ P\overline{P}\,\overline{P},\ \overline{P}P\overline{P},\ \overline{P}\,\overline{P}P,\ \overline{P}\,\overline{P}\,\overline{P}\} = \{A,\ B,\ C,\ D\}.$$

Ha az egyes eseményeket realizáló összes elemi eseményt felírjuk, akkor az eredeti feladat modellezhető úgy, hogy tekintjük a PPP-t és a $\overline{P}PP$-t realizáló elemi események halmazát, és azt kérdezzük, hogy mivel egyenlő $\frac{|\{PPP\}|}{|\{PPP\}|+|\{\overline{P}PP\}|}$ hányados. Ezzel a feladat megoldását visszavezettük egyrészt a klasszikus valószínűségelméletre, másrészt – technikailag – a visszatevés nélküli mintavételnél alkalmazott hipergeometrikus eloszlásra, amiből:

$$\frac{\binom{8}{3} \cdot \binom{24}{0}}{\binom{8}{3}\binom{24}{0} + \binom{8}{2}\binom{24}{1} \cdot \frac{1}{3}} = \frac{1}{5}$$

1.17. gyakorlat megoldása:

Megoldás a klasszikus valószínűségelmélet alapján. Elemi események: az 1, 2, 3, 4, 5, 6 számok harmadosztályú ismétléses variációi.
Az összes esetek száma: 6^3. A kedvező esetek száma: $\binom{6}{3} \cdot 3!$, mivel 6 különböző elem közül kell kiválasztani hármat az összes lehetséges módon. A sorrendjük ugyan nem számít, egy adott számhármas – mint három elemű halmaz – 3! féleképpen valósulhat meg, ezért a választott modell (klasszikus valószínűségelmélet) használhatósága érdekében a három egymás utáni dobás során előforduló különböző sorrendű bekövetkezéseket meg kell különböztetni.
Így a keresett valószínűség:

$$p = \frac{\binom{6}{3} \cdot 3!}{6^3} = \frac{\frac{6 \cdot 5 \cdot 4}{3 \cdot 2 \cdot 1} \cdot 1 \cdot 2 \cdot 3}{6 \cdot 6 \cdot 6} = \frac{5}{9}$$

1.19. gyakorlat megoldása:

A feltételekből adódóan $P(A \cdot B) = 0$. A B és $\overline{B}$ események teljes eseményrendszert alkotnak. Így $P(A) = P(A(B+\overline{B})) = P(AB+A\overline{B}) = P(AB)+P(A\overline{B}) = 0,2$, mivel AB és $A\overline{B}$ események is egymást kizáró események. A definíciót felhasználva: $P(A+B|B) = \frac{P\big((A+B) \cdot B\big)}{P(B)} = \frac{P(A \cdot B+B)}{P(B)} = \frac{P(B)}{P(B)} = 1$. □

1.20. gyakorlat megoldása:

Feladat a $P(A+B)$ valószínűség kiszámítása. A függetlenség felhasználásával (1.2.4.) alapján

$$P(A+B) \quad = 0,8 + 0,6 - 0,8 \cdot 0,6 = 1,4 - 0,48 = 0,92$$

□

1.21. gyakorlat megoldása:

Elemi események az $1, 2, 3, 4, 5, 6$ számok harmad osztályú ismétléses variációi. Az összes elemi események száma: $|\Omega| = 6^3$. A feladat szempontjából kedvező eseteket megvalósító elemi események:
$(1,5,6), (1,4,5), (1,4,6), (1,3,4), (1,3,5), (1,3,6), (1,2,3), (1,2,4), (1,2,5), (1,2,6), (2,5,6), (2,4,5), (2,4,6), (2,3,4), (2,3,5), (2,3,6), (3,5,6), (3,4,5), (3,4,6), (4,5,6)$.
Ezek száma 20, és mindegyik azonos valószínűségű (ha három egymás utáni kockadobást független kísérletsorozatnak feltételezünk). Tehát

$$P = \frac{20}{6^3} = \frac{2^2 \cdot 5}{2^3 \cdot 3^3} = \frac{5}{54}.$$

□

1.24. gyakorlat megoldása:

A feltételekbő adott, hogy

$$P(A_0) = P(A_1) = P(B_0) = P(B_1) = P(C_0) = P(C_1) = \frac{1}{2}$$

és

$$\begin{aligned} P(A_1 \cdot B_j) &= \tfrac{1}{4} = P(A_i) \cdot P(B_j) \quad i = 0,1, \quad j = 0,1 \\ P(A_i \cdot C_k) &= \tfrac{1}{4} = P(A_i) \cdot P(C_k) \quad i = 0,1, \quad k = 0,1 \\ P(B_j \cdot C_k) &= \tfrac{1}{4} = P(B_j) \cdot P(C_k) \quad j = 0,1, \quad k = 0,1\,, \end{aligned}$$

de

$$P(A_0 \cdot B_0 \cdot C_0) = \frac{1}{4} \neq P(A_0) \cdot P(B_0) \cdot P(C_0)\,.$$

1.26. gyakorlat megoldása:

Jelentse A : az állat szemmutáns, B : a szárnymutáns eseményeket.

a.) $P(A+B) = P(A) + P(B) - P(A \cdot B) = 0.25 + 0.50 - 0.10 = 0.65$

b.) $P(A-B) = P(A) - P(A \cdot B) = 0.25 - 0.10 = 0.15$

1.27. gyakorlat megoldása:

Jelöljük az ,,elpusztít", ,,károsít" és ,,nem talája" eseményeket rendre A, B, C-vel. Ekkor $P(A) = 0.5$, $P(B) = 0.25$ és $P(C) = 1 - 0.75 = 0.25$. A négy részecske hatása X_1, X_2, X_3, X_4; itt az X_i az i-edik részecske hatásától függően A, B vagy C. A sejt pusztulása akkor következik be, ha a sorozatban van A, vagy nincs ugyan A, de legalább két B előfordul benne. Az előbbi események ellentettjének valószínvalószínűsége 0.5^4. Ezért annakvalószínűsége, hogy a részecskék egyike elpusztítja a sejtet, $1 - 0.5^4$. A másik összetett esemény valószínűsége $0.25^4 \left[\binom{4}{2} + \binom{4}{3} + \binom{4}{4}\right] = 11 \cdot 0.25^4$. Innen a végeredmény 0.98.

1.28. gyakorlat megoldása:

A sejthalmaz besugárzás utáni állapotát egy N elemű számsorozattal írhatjuk le, melynek i-edik eleme ($i = 1, 2, \ldots, N$) az i-edik sejtet ért találatok száma. Az állapotok száma annyi, ahányféleképpen az N sejt közül visszatevéses eljárással n számút ki lehet választani (a kiválasztás sorrendje nem számít): $\binom{N+ \quad n-1}{n}$ Azon esetek száma, melyekben minden sejtet legalább 2 találat ér, a fentiekhez hasonló megoldással $\binom{n-N-1}{n-2N}$. A keresett valószínûség tehát $\dfrac{\binom{n-N-1}{n-2n}}{\binom{N+n-1}{n}}$

1.29. gyakorlat megoldása:

Jelöljük a mozgásszervi és érzékszervi rendellenesség elõfordulásának eseményét A_1-gyel, illetve A_2-vel. Ekkor a $P(A_1 \cup A_2) = P(A_1) + P(A_2) - PA_1 \cap A_2)$ összefüggés szerint $P(\overline{A_1 \cup A_2}) = 1 - P(A_1 \cup A_2) = 0.93$

1.30. gyakorlat megoldása:

Legyen az AB eseményhez tarozó esetszám N_{AB}. Ekkor AB relatív gyakorisága (melyet a feladat szerint AB valószínûségének tekintünk) $\frac{N_{AB}}{N_{AB}+4+9+10}$. Az A esemény relatív gyakorisága $\frac{N_{AB}+9}{N_{AB}+4+9+10}$, a B eseményé $\frac{N_{AB}+4}{N_{AB}+4+9+10}$. Ezeket is valószínûségi értéknek tekintve, függetlenség esetén $\frac{N_{AB}}{N_{AB}+23} = \frac{N_{AB}+9}{N_{AB}+23} \cdot \frac{N_{AB}+4}{N_{AB}+23}$. Innen N_{AB}-re 3.6 adódik, ami esetszám nyilván nem lehet. Az eredményt úgy interpretálhatjuk, hogyha A és B független, akkor a megadott cellagyakoriságok mellett AB gyakorisága átlagosan 3.6.

6.2. A második fejezet gyakorlatai

2.2. gyakorlat megoldása:

A megoldás alapja a várható érték (2.3.6)-ban leírt linearitása és a szórás (2.3.10)-ben megadott definíciója. Tegyük fel, hogy ξ varianciája σ^2. Ugyanekkor $a\xi$ varianciája $a^2\sigma^2$, és ugyanekkora $a\xi + b$ varianciája. (Konstans hozzáadása az ingadozást nem befolyásolja.) □

2.5. gyakorlat megoldása:

Elemi események típusai:

$$\begin{array}{c} FIFI\ \ldots\ IFF \\ FIFI\ \ldots\ FII \\ IFIF\ \ldots\ IFF \\ IFIF\ \ldots\ FII \end{array}$$

k számú I és F váltakozva.

Legyen az η valószínűségi változó értéke a szükséges dobások száma.

$$\begin{array}{lll} k=0 & FF \text{ vagy } II & p_0 = 2\cdot\left(\frac{1}{2}\right)^2 = P(\eta=2) \\ k=1 & IFF \text{ vagy } FII & p_1 = 2\cdot\left(\frac{1}{2}\right)^3 = P(\eta=3) \\ k=2 & IFII \text{ vagy } FIFF & p_2 = 2\cdot\left(\frac{1}{2}\right)^4 = P(\eta=4) \\ k=3 & FIFII \text{ vagy } FIFFI & p_3 = 2\cdot\left(\frac{1}{2}\right)^5 = P(\eta=5), \end{array}$$

és így tovább.

A keresett várható érték:

$$M(\eta) = 2P(\eta=2) + 3P(\eta=3) + \ldots + nP(\eta=n) + \ldots$$

Behelyettesítve a megfelelő valószínűségeket:

$$\begin{array}{rcl} M(\eta) = & 2\cdot\frac{1}{2}\cdot\left(2\cdot\frac{1}{2} + 3\cdot\left(\frac{1}{2}\right)^2 + \ldots + n\cdot\left(\frac{1}{2}\right)^{n-1} + \ldots\right) \\ M(\eta) = & (2\cdot x + 3x^2 + \ldots + nx^{n-1} + \ldots)_{x=\frac{1}{2}} \\ = & (x^2 + x^3 + \ldots + x^n + \ldots)'_{x=\frac{1}{2}} = \\ = & \left(x^2\cdot\frac{1}{1-x}\right)'\Big|_{x=\frac{1}{2}} & \text{(végtelen mértani sor összege)} \\ = & \left(\frac{x^2}{1-x}\right)' = \frac{2x-x^2}{(1-x)^2}. \end{array}$$

Ezzel

$$M(\eta) = \frac{2x-x^2}{(1-x)^2}\Big|_{x=\frac{1}{2}} \qquad M(\eta) = \frac{1-\frac{1}{4}}{\left(\frac{1}{2}\right)^2} = \frac{\frac{3}{4}}{\frac{1}{4}} = 3.$$

Tehát a szükséges dobások számának várható értéke 3.

Megjegyzés: A feladat egyszerűen megoldható a 2.9. példához hasonlóan. A ξ valószínűségi változó itt is jelentsen várakozási időt! Így ξ k számú valószínűségi változó összege lesz.

□

2.6. gyakorlat megoldása:

Megoldás a (2.2.7) képlet alapján.

$$F_\xi(x) = \begin{cases} 1 - e^{-\lambda x} & \text{ha} \quad x > 0 \\ 0 & \text{ha} \quad x \leqslant 0 \end{cases} \quad \text{és} \quad \eta = 2\xi + 3 \, .$$

$$F_\eta(x) = P(\eta < x) = P(2\xi + 3 < x) = P\left(\xi < \tfrac{x-3}{2}\right) = F_\xi\left(\tfrac{x-3}{2}\right), \text{ tehát}$$

$$F_\eta(x) = \begin{cases} 1 - e^{-\lambda\left(\frac{x-3}{2}\right)} & \text{ha} \quad \frac{x-3}{2} > 0 \quad \text{ha} \quad x > 3 \\ 0 & \text{ha} \quad x \leqslant 3 \end{cases}$$

$$f_\eta(x) = F'_\eta(x) = \begin{cases} \frac{\lambda}{2} e^{-\lambda\left(\frac{x-3}{2}\right)} & \text{ha} \quad \text{x¿3} \\ 0 & \text{különben} \end{cases}$$

□

2.7. gyakorlat megoldása:

A feladat egyszerűen megoldható. Pl.: a $(\xi, \eta) = (1,\ 0)$ esemény bekövetkezésének valószínűsége $\frac{1}{4}$. A $P(\xi = 1) = \frac{1}{4}$, $P(\eta = 0) = \frac{2}{4} = \frac{1}{2}$ és $P(\xi = 1) \cdot P(\eta = 0) \neq P(\xi = 1,\ \eta = 0)$. Tehát ξ és η nem függetlenek. □

2.9. gyakorlat megoldása:

Megoldás a (2.3.6) vagy a (2.3.7) alapján:

$$f_\xi = \begin{cases} \frac{1}{4\pi} & \text{ha} \quad -2\pi \leq x \leq \frac{\pi}{2} \quad \text{és} \quad M(\xi) = 0 \\ 0 & \text{különben} \end{cases}$$

(a)

$$\eta = 4\xi + 1 \quad M(\eta) = 1$$

(b)

$$\eta = \xi^2 \quad M(\eta) = ?$$

$$M(\eta) = \int_{-2\pi}^{2\pi} x^2 \cdot \frac{1}{4\pi} dx = \frac{1}{4\pi} \left[\frac{x^3}{3}\right]_{-2\pi}^{2\pi} = \frac{4\pi^2}{3}$$

(c)

$$\eta = e^\xi$$

$$M(\eta) = \int_{-2\pi}^{2\pi} e^x \frac{1}{4\pi} dx = \frac{1}{4\pi} \left[e^x\right]_{-2\pi}^{2\pi} = \frac{1}{4\pi}\left(e^{2\pi} - e^{-2\pi}\right) = \frac{s\,h(2\pi)}{2\pi}$$

(d)

$$\eta = \sin \xi$$

$$M(\eta) = \int_{-2\pi}^{2\pi} \sin x \cdot \frac{1}{4\pi} dx = \frac{1}{4\pi} \left[\cos x\right]_{-2\pi}^{2\pi} = 0 \qquad \square$$

2.10. gyakorlat megoldása:

Megoldás a sűrűségfüggvény jelentése alapján: $P(\xi \geqslant 4) = \int\limits_4^\infty 3e^{-3t} dt = e^{-12} = 0,61 \cdot 10^{-5}$, mivel a keresett valószínűség $\lim\limits_{b\to\infty} \left[-e^{-3b}\right] = 0$. $\square$

2.11. gyakorlat megoldása:

Jelölje ξ a 16 egymás utáni független éremdobás „fej" (F) eredményeinek számát! ξ lehetséges értékei: $k = 0, 1, 2, \dots, 16$ és ξ binomiális eloszlású valószínűségi változó, $p = \frac{1}{2}$ paraméterrel, $n = 16$, mivel $m = n \cdot p$ és $\sigma = \sqrt{n \cdot p(1-p)}$, tehát $m = 16 \cdot \frac{1}{2} = 8$ $\sigma = \sqrt{16 \cdot \frac{1}{2} \cdot 12} = 20$

Meg kell határozni a $P(2 \leqslant \xi \leqslant 14)$ valószínűséget: 2.1. alapján:

$$P(2 \leqslant \xi \leqslant 4) = 1 - P(|\xi - 8| \geqslant 7) \geqslant 1 - \frac{4}{49} = \frac{45}{49}.$$

A pontos értéke:

$$P(\xi = 2) + P(\xi = 3) + \ldots + P(\xi = 14) = \sum_{k=2}^{14} P(\xi = k)$$

$$= \sum_{k=2}^{14} \binom{16}{k} \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{16-k},$$

mivel az esemény összeg tagjai egymást kizáró események, és ξ binomiális eloszlású valószínűségi változó.

$$P(2 \leqslant \xi \leqslant 14) = \left(\frac{1}{2}\right)^{16} \cdot \sum_{k=2}^{14} \binom{16}{k} = \left(\frac{1}{2}\right)^{16}$$

$$\cdot \left(\sum_{k=0}^{16} \binom{16}{k} - \binom{16}{0} - \binom{16}{1} - \binom{16}{15} - \binom{16}{16} \right)$$

$$= 1 - \frac{1}{2^{16}}(1 + 16 + 16 + 1) = 1 - \frac{17}{2^{15}}.$$

Mivel $\sum\limits_{k=0}^{16} \binom{16}{k} = 2^{16}$, ha alkalmazzuk a binomiális tételt az $(1+1)^{16}$ kiszámítására.

Megjegyzés: A $P(2 \leqslant \xi \leqslant 14) = 1 - (P(\xi = 0) + P(\xi = 1) + P(\xi = 15) + P(\xi = 16))$ összefüggés – az ellentett esemény valószínűségére vonatkozó tétel alapján – közvetlenül szolgáltatja a pontos valószínűséget.

2.16. gyakorlat megoldása:

A Poisson-eloszlást jól közelítő valószínűségi változók között az egyik legtipikusabb a pontszerű hibák száma valamely anyagban. Az egy oldalon előforduló hibák várható értéke $\frac{1}{4}$. Számoljuk ki az ellentett esemény valószínűségét!

$$P(\xi=0)+P(\xi=1)+P(\xi=2)=\frac{\left(\frac{1}{4}\right)^0}{0!}e^{-\frac{1}{4}}+\frac{\frac{1}{4}}{1!}e^{-\frac{1}{4}}+\frac{\left(\frac{1}{4}\right)^2}{2!}e^{-\frac{1}{4}}=\sum_{k=0}^{2}\frac{\left(\frac{1}{4}\right)^k}{k!}e^{-\frac{1}{4}}\approx 0.997839.$$

Annak a valószínűsége, hogy egy oldalon több, mint két hiba található: $P(\xi>2)=1-\sum_{k=0}^{2}\frac{\left(\frac{1}{4}\right)^k}{k!}e^{-\frac{1}{4}}\approx 0,002161.$

2.17. gyakorlat megoldása:

A második momentum az

$$\int_0^\infty x^2 f_\xi(x)\,dx$$

integrál. Helyettesítéssel

$$\int_0^\infty x^2\left(\alpha\beta^{-\alpha}x^{\alpha-1}e^{-(x/\beta)^\alpha}\right)dx=\beta^2\int_0^\infty x^{2/\alpha}e^{-x}\,dx.$$

Az integrál értéke a gamma függvény definiciója szerint $\Gamma(1+2/\alpha)$.

Az $\alpha=1$ eset exponenciális eloszlást ad. □

2.18. gyakorlat megoldása:

(2.2.4) szerint $F(x)=\int_{-\infty}^{x}f(t)dt$, így

$$F_\xi(x)=\begin{cases}\int_0^x 2\cdot\beta^{-2}\cdot x\cdot e^{-\frac{x^2}{\beta^2}}dx & \text{ha}\quad 0<x\\ 0 & \text{ha}\quad x\leqslant 0\end{cases}$$

$$\int_0^x\frac{2x}{\beta^2}\cdot e^{-\frac{x^2}{\beta^2}}dx=-\int_{u_1}^{u_2}e^u du=\left[-e^{-\frac{x^2}{\beta^2}}\right]_0^x$$

tehát

$$F_\xi(x)=\begin{cases}1-e^{-\frac{x^2}{\beta^2}} & \text{ha}\quad 0<x\\ 0 & \text{ha}\quad x\leqslant 0\end{cases}$$

□

2.19. gyakorlat megoldása:

(a) Megoldás 2.3. alapján.

$$\begin{aligned}\sigma^2(2\xi-3\eta+1) &= M\left((2\xi-3\eta+1)^2\right)-\left(M(2\xi-3\eta+1)\right)^2\\ (2\xi-3\eta+1)^2 &= 4\xi^2-12\xi\eta+9\eta^2+4\xi-6\eta+1\\ M\left((2\xi-3\eta+1)^2\right) &= 4M(\xi^2)-12M(\xi)\cdot M(\eta)+9M(\eta^2)+4M(\xi)-6M(\eta)+1\\ \left(M(2\xi-3\eta+1)\right)^2 &= (2M(\xi)-3M(\eta)+1)^2=4(M(\xi))^2-12M(\xi)M(\eta)\\ &\quad +9\left(M(\eta)\right)^2+4M(\xi)-6M(\eta)+1.\end{aligned}$$

Ezzel

$$\begin{aligned}
\sigma^2(2\xi-3\eta+1) &= 4\cdot\left(M(\xi^2)-(M(\xi))^2\right)+9\left(M(\xi^2)-(M(\eta))^2\right)\\
\sigma^2(2\xi-3\eta+1) &= \underbrace{4\sigma^2(\xi)}_{3}+\underbrace{9\sigma^2(\eta)}_{3}\\
\sigma(2\xi-3\eta+1) &= \sqrt{39}
\end{aligned}$$

(b) Megoldás

(A) Belátjuk, ha ξ és η függetlenek, akkor $a\xi$ és $b\eta+1$ is független ($a\neq 0$) és $b\neq 0$).

Bizonyítás 2.5.1./28. utáni megjegyzés alapján. Legyen $g(\xi)=a\cdot\xi$ és $k(\eta)=b\eta+1$. Tehát igaz.

Ekkor 2.3. alapján

$$\sigma^2(2\xi-3\eta+1)=\sigma^2(2\xi)+\sigma^2(-3\eta+1).$$

(B) Belátjuk, ha $\sigma^2(a\cdot\xi)=a^2\sigma^2(\xi)$. A 2.3.6. összefüggés alapján

$$\begin{aligned}
&\sigma^2(a\cdot\xi)=M(a^2\cdot\xi^2)-(M(a\xi))^2\\
=\;& a^2M(\xi^2)-a^2M^2(\xi)=a^2\left(M(\xi^2)-(M(\xi))^2\right)=a^2\sigma^2(\xi).
\end{aligned}$$

(C) Belátjuk, $\sigma^2(\xi+c)=\sigma^2(\xi)$, c tetszőleges állandó.

$$\begin{aligned}
&\sigma^2(\xi+c)\\
=\;& M(\xi^2+2\cdot c\cdot\xi+c^2)-((M(\xi))^2+2\cdot c\cdot M(\xi)+c^2)\\
=\;& M(\xi^2)+2\cdot M(\xi)+c^2-(M(\xi))^2-2\cdot c\cdot M(\xi)-c^2\\
=\;& M(\xi^2)-(M(\xi))^2=\sigma^2(\xi)
\end{aligned}$$

B. és C. alapján

$$\sigma^2(2\xi)+\sigma^2(-3\eta+1)=4\sigma^2(\xi)+9\sigma^2(\eta)$$

Ezzel az a) megoldáshoz hasonlóan:

$$\sigma^2(2\xi-3\eta+1)=4\cdot 3+9\cdot 3\rightarrow\sigma(2\xi-3\eta+1)=\sqrt{39}\qquad\square$$

2.21. gyakorlat megoldása:

Legyen ξ a k-dik fejre való várakozás ideje, és ξ_i az a véletlen dobásszám, ami a $(i-1)$-ik fej dobás után következik a újabb fejig. Tehát $\xi=\xi_1+\xi_2+\ldots+\xi_k$ és $M(\xi)=M(\xi_1)+M(\xi_2)+\ldots+M(\xi_k)$. Mindegyik ξ_i valószínűségi változó geometriai eloszlású $p=1/2$ sikervalószínűséggel és 2 várható értékkel. Ezért $M(\xi)=2k$.

2.22. gyakorlat megoldása:

$$F_\xi(y)=\begin{cases} y & \text{ha}\quad 0<y\leqslant 1\\ 1 & \text{ha}\quad 1<y\end{cases}$$

Az η eloszlás függvénye a megadott sűrűségfüggvényből:

$$F_\eta(y) = \begin{cases} \frac{y^3}{8} & \text{ha} \quad 0 < y \leqslant 2 \\ 1 & \text{ha} \quad 2 < y \end{cases}$$

Így $\eta = r(\xi)$, azaz $\eta = \sqrt[3]{8\xi}$ kell, hogy legyen. $F_\eta(y) = P(\eta < y) = P\left(\sqrt[3]{8\xi} < y\right) = P\left(\xi < \frac{y^3}{8}\right) = F_\xi\left(\frac{y^3}{8}\right)$. □

2.23. gyakorlat megoldása:

$$F(x) = \begin{cases} 0 & \text{ha} \quad x \leqslant 0 \\ 1 - e^{-\lambda x} & \text{ha} \quad 0 < x \end{cases} \qquad \begin{matrix} M(\xi) = ? \\ P(\xi < 1) = ? \end{matrix}$$

$$P\left(\{\omega \in \Omega : \xi(\omega) < 3\}\right) = 1 - P\left(\{\omega \in \Omega : 3 \leqslant \xi\}\right) = \frac{1}{2}.$$

(2.2.1) alapján:

$$F_\xi(3) = 1 - e^{-3\lambda} = \frac{1}{2}; \quad e^{-3\lambda} = \frac{1}{2}; \quad -3\lambda = -\ln 2; \quad \lambda = \frac{\ln 2}{3} = \ln\sqrt[3]{2}$$

(a)

$$M(\xi) = \int_{-\infty}^{\infty} x f(x)(dx) = \frac{1}{3}\int_0^\infty x \ln 2 \cdot e^{-\frac{\ln 2}{3}x} dx = \frac{3}{\ln 2} = \frac{1}{\lambda}$$

(b)

$$P\left(\{\omega \in \Omega : \xi(\omega) < 1\}\right) = F_\xi(1) = 1 - e^{-\ln\sqrt[3]{2}} = 1 - \frac{1}{e^{\ln\sqrt[3]{2}}} = 1 - \frac{1}{\sqrt[3]{2}}$$

Megjegyzés: Tetszőleges valós c esetén:

$$p_c = P\left(\{\omega \in \Omega : \xi(\omega) = c\cdot\}\right) = 0, \quad \text{2.2.2. alapján} \quad p_c = \lim_{d \to 0}\left(F_\xi(c+d) - F_\xi(c)\right)$$

kifejezéssel definiálható az exponenciális eloszlás esetén.

Ebből pedig: $p_c = \lim\limits_{d \to 0}\left(1 - e^{-\lambda(c+d)} - \left(1 - e^{-\lambda c}\right)\right) = 0$ adódik az exponenciális függvény folytonossága miatt. Ezért $P\left(\{w \in \Omega : \xi > 3\}\right) = P\left(\{w \in \Omega : \xi \geqslant 3\}\right)$. □

2.24. gyakorlat megoldása:

A feladat megoldása a (-0,5, -0,4] és a [0,4, 0,5) intervallumokba esés valószínűségét jelenti. Felhasználva a sűrűségfüggvény jelentését – egy adott intervallumba esés valószínűsége egyenlő az adott intervallum feletti sűrűségfüggvény alatti terület számértékével – valamint $f(x)$ „egyszerű” alakját:

$$P\left((-0,5 < \xi \leqslant -0,4) + (0,4 \leqslant \xi < 0,5)\right) = 2 \cdot 0,1 \cdot \frac{1}{2} = 0,1. \qquad \square$$

6.3. Harmadik fejezet gyakorlatai

3.2. gyakorlat megoldása: qnoindent

ξ jelölje az első meghibásodásig eltelt időt. ξ normális eloszlású $N(6,4;\ 2,3)$ paraméterekkel. $P(\xi \geqslant 8|\xi \geqslant 5)$ feltételes valószínűségét az (1.3.2) alapján számoljuk. $P(\xi \geqslant 8|\xi \geqslant 5) = \frac{P(\xi \geqslant 8)}{P(\xi \geqslant 5)}$, mivel annak az eseménynek a bekövetkezése, hogy $\xi \geqslant 8$ maga után vonja azt, hogy $\xi \geqslant 5$. A keresett valószínűséget az ellentett esemény valószínűségének a meghatározásával kapjuk meg. Tehát

$$\begin{aligned} & P(\xi \geqslant 8|\xi \geqslant 5) \\ = & \frac{1-P(\xi \leqslant 8)}{1-P(\xi \leqslant 5)} = \frac{1-\Phi\left(\frac{8-6,4}{2,3}\right)}{1-\Phi\left(\frac{5-6,4}{2,3}\right)} \\ = & \frac{1-\Phi\left(\frac{1,6}{2,3}\right)}{1-\Phi\left(\frac{-1,4}{2,3}\right)} = \frac{1-\Phi(0,69)}{1-(1-\Phi(0,60))} = \frac{1-0,7549}{0,7257} \approx 0,333943 \end{aligned}$$

□

3.3. gyakorlat megoldása:
Az $1/(\xi+1)$ valószínűségi változó az $1/(k+1)$ értékeket veszi fel

$$\frac{\lambda^k}{k!}e^{-\lambda}$$

valószínűséggel, $k = 0, 1, 2, \ldots$. Ezért a várható érték meghatározása alapján

$$M\left(\frac{1}{\xi+1}\right) = \sum_{k=0}^{\infty} \frac{1}{k+1}\frac{\lambda^k}{k!}e^{-\lambda},$$

amit összegeznünk kell. Az exponenciális függvény Taylor-sorát használjuk fel:

$$\begin{aligned} M &= \sum_{k=0}^{\infty} \frac{1}{k+1}\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda} + \sum_{k=1}^{\infty} \frac{1}{k+1}\frac{\lambda^k}{k!}e^{-\lambda} \\ &= e^{-\lambda} + e^{-\lambda}\sum_{k=1}^{\infty} \frac{\lambda^k}{(k+1)!} = e^{-\lambda} + e^{-\lambda}\sum_{j=2}^{\infty} \frac{\lambda^{j-1}}{j!} \\ &= e^{-\lambda} + e^{-\lambda}\lambda^{-1}\sum_{j=2}^{\infty} \frac{\lambda^j}{j!} = e^{-\lambda} + e^{-\lambda}\lambda^{-1}(e^{-\lambda}-1-\lambda) = \lambda^{-1}(1-e^{-\lambda}). \end{aligned}$$

□

3.6. gyakorlat megoldása:

Megoldás a (3.3.5) egyenletes eloszlás definíciója alapján.

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{ha} \quad a < x < b \\ 0 & \text{különben} \end{cases} \tag{3.2.2}$$

$\sigma^2(\xi) = \frac{(b-a)^2}{12} = 1$, amiből $b - a = \sqrt{12}$. ξ entrópiája:

$$-\int_{-\infty}^{\infty} f_\xi(x)\ln f_\xi(x)dx = -\int_a^b \left(\frac{1}{b-a}\ln\frac{1}{b-a}\right)dx = (a-b)\left(\frac{1}{b-a}\ln\frac{1}{b-a}\right)$$
$$= \quad -\ln\frac{1}{b-a} = \ln(b-a) = \ln\sqrt{12} \approx 1,24.$$

Az $N(m,1)$ eloszlás entrópiája:

$$\ln\sqrt{2\pi e} \approx 1,42.$$

□

3.7. gyakorlat megoldása:

Ha η gamma eloszlású, akkor sűrűségfüggvénye:

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}\cdot e^{-\frac{x}{\beta}}, & \text{ha} \quad x > 0 \\ 0 & \text{különben} \end{cases} \qquad (3.2.2)$$

A sűrűségfüggvényt az eloszlásfüggvény deriváltjaként értelmezzük.

$$f(x) = F'_\xi(x) \qquad (2.2.4)$$

Az eloszlásfüggvény definíciója:

$$F_\xi(x) = P\left(\{\omega \in \Omega : \xi(\omega) < x\}\right) \qquad (2.2.1)$$

Ha

$$\xi(\omega) = c\cdot\eta(\omega),$$

akkor

$$F_\xi(x) = P\left(\left\{\omega \in \Omega : \eta(\omega) < \frac{x}{c}\right\}\right) = F_\eta\left(\frac{x}{c}\right).$$

Ebből az összetett függvény deriválási szabálya alapján

$$f_\xi(x) = F'_\xi(x) = F'_\eta\left(\frac{x}{c}\right)\cdot\frac{1}{c} = f_\eta\left(\frac{x}{c}\right)\cdot\frac{1}{c}.$$

Ezzel

$$f_{c\eta}(x) = \begin{cases} \frac{1}{\beta^\alpha\Gamma(\alpha)}\cdot\left(\frac{x}{c}\right)^{\alpha-1}\cdot e^{-\frac{x}{c\cdot\beta}}\cdot\frac{1}{c} & \text{ha} \quad 0 < x \\ 0 & \text{különben} \end{cases}. \quad \square$$

3.8. gyakorlat megoldása:

Legyen ξ^a eloszlásfüggvénye F_a, sűrűségfüggvénye f_a. Ekkor $F_a(x) = P(\xi^a < x) = P(\xi < x^{1/a}) = F(x^{1/a})$. A közvetett függvény differenciálási szabályával

$$f_a(x) = \frac{d}{dx}F_a(x) = \frac{d}{dx}F(x^{1/a}) = F'(x^{1/a})\frac{d}{dx}x^{1/a} = f(x^{1/a})\frac{1}{a}x^{\frac{1}{a}-1}$$

Behelyettesítve kapjuk

$$f_a(x) = b^{-a}e^{-b^{-a}x},$$

ami exponenciális eloszlás b^{-a} paraméterrel. □

3.9. gyakorlat megoldása:

A lognormális eloszlás sűrűségfüggvénye

$$f_\eta(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma\cdot x}\cdot e^{-\frac{(\ln(x)-m)^2}{2\sigma^2}} \qquad (3.4.1)$$

A keresett valószínűség

$$P(4,7 < \eta < 9,8) = F_\eta(9,8) - F_\eta(4,7) \qquad (2.2.2.)$$

$$F_\eta(9,8) = \int_{-\infty}^{9,8} f_\eta(x)dx \qquad F_\eta(4,7) = \int_{-\infty}^{4,7} f_\eta(x)dx$$

$$F_\eta(9,8) - F_\eta(4,7) = \int_{x_1=4,7}^{x_2=9,8} \frac{1}{\sqrt{2\pi}\cdot 2}\frac{1}{x}e^{-\frac{(\ln(x)+2)^2}{8}}dx$$

$$u = \frac{\ln(x)+2}{2} \longrightarrow du = \frac{1}{2}\cdot\frac{1}{x}dx$$

$$F_\eta(9,8) - F_\eta(4,7) = \frac{1}{\sqrt{2\pi}\int_{\frac{\ln 4,7+2}{2}}^{\frac{\ln 9,8+2}{2}}}e^{-\frac{n^2}{2}}dn$$

$$= \Phi\left(\frac{\ln(9,8)+2}{2}\right) - \Phi\left(\frac{\ln(4,7)+2}{2}\right)$$

A standard normális eloszlás táblázatából a $\Phi(u_1)$ és $\Phi(u_2)$ értékek kikereshetők, és különbségük, $\Phi(u_2) - \Phi(u_1)$ szolgáltatja a keresett valószínűséget.

$$\begin{aligned} \Phi(2,1412) - \Phi(1,7738) &= 0,9838 - 0,9616 \\ P(4,7 < \eta < 9,8) &= 0,0222. \end{aligned}$$

□

3.11. gyakorlat megoldása:

Megoldás (3.2.1) szerint. Igazolni kell, hogy

$$P(\xi \leqslant x + y | \xi \geqslant x) = P(\xi \leqslant y).$$

A geometriai eloszlás definíciója

$$P(\xi = k) = \begin{cases} 0, & \text{ha} \quad k = 0 \\ q^{k-1}\cdot p, & \text{ha} \quad k > 0 \end{cases} \qquad (2.1.2.)$$

$$F(x) = \begin{cases} 0 & \text{ha} \quad x \leqslant 0 \\ p\cdot\frac{q^k-1}{q-1}, & \text{ha} \quad k < x \leqslant k+1 \ \text{ ahol } \quad 1 \leqslant k \end{cases}$$

Ezt felhasználva

$$P(\xi \leqslant x + y | \xi \geqslant x) = \frac{P(x \leqslant \xi \leqslant x + y)}{P(x \leqslant \xi)} = \frac{F_\xi(x+y) - F_\xi(x)}{1 - F_\xi(x)}.$$

Az állítás triviálisan igaz $y \leqslant 0$ esetén, ezért elég a bizonyítást az $1 \leqslant y$ esetre elvégezni. Tegyük fel, hogy $k < x \leqslant k+1$ és $l < y \leqslant l+1$, ahol $1 \leqslant k$ és $0 \leqslant l$.

Ekkor

$$F_\xi(x+y) = p\cdot\frac{q^{k+l}-1}{q-1} \qquad F_\xi(x) = p\cdot\frac{q^k-1}{q-1}$$

$$\frac{F_\xi(x+y)-F_\xi(x)}{1-F_\xi(x)} = \frac{p}{q-1}\cdot\frac{q^k(q^l-1)}{1-p\cdot\frac{q^k-1}{q-1}} = pq^k\cdot\frac{q^l-1}{q-1-pq^k+p}$$

rendezve és felhasználva, hogy $p+q=1$

$$P(\xi\leqslant x+y|\xi\geqslant x) = \;-\left(q^l-1\right) = \tfrac{p}{-p}\cdot(q^l-1) = \tfrac{p}{-(1-q)}(q^l-1)$$
$$P(\xi\leqslant x+y|\xi\geqslant x) = \; p\cdot\tfrac{q^l-1}{q-1} = F_\xi(y).$$

3.12. gyakorlat megoldása:

A mintában szereplõ balkezesek ξ száma binomiális eloszlású valószínûségi változó $p=0.2$ és $n=10.000$ paraméterekkel, amelyre $M(\xi)=2.000$ és $D^n(\xi)=40$. Így

(a) $P(\xi\geqslant 2.100 = \sum_{k=2.100}^{10.000}\binom{10.000}{k}0.2^k\cdot(1-0.2)^{10.000-k}$

(b) $P = (1.960\leqslant\xi\leqslant 2.040) = \sum^{2}.040_{k=1.960}\binom{10.000}{k}\cdot 0.2^k(1-0.2)^{10.000-k}$

Ennek a kiszámítása elég nehézkes, de nem szükséges, mivel $n\cdot p=2.000$ és $n\cdot g=8.000$, így a 3.5 tétel alkalmazható. Tehát $\eta=\frac{\xi-2.000}{40}$ standard normális eloszlású valószínûségi változó.

(a) $P(\xi\geqslant 2.100) = P(\xi\geqslant 2.5) = 1-P(\xi<2.5) = 0.0062$

(b) $P(1.960\leqslant\xi\leqslant 2.040) = P(-1\leqslant\eta\leqslant 1)\approx 0.68$.

3.15. gyakorlat megoldása:

2 cm^3 vízben átlagosan 4 baktérium van, így $\lambda=4$ paraméterû Poisson eloszlásról van szó. Tehát

(a) $P(k=0)=e^{-4}=0.0183$

(b) Az elemzett események valószínûségét felhasználva $1-P(k=0)-P(k=1)=1-5e^{-4}=0.908$.

3.16. gyakorlat megoldása:

Legyen ξ olyan valószínûségi változó, amely az A ponttól idõegység alatt megtett távolságot jelenti. Ekkor ξ eloszlás függvénye $F(r)=\int_0^r e^{-2t}dt = 1-e^{-2r}$.

(a) $P(\xi\geqslant 1) = 1-F(1) = e^{-2} = 0.315.$

(b) A jelenség exponenciális eloszlással jellemezhetõ $x=2$ paraméterrel. Így a várható érték $\frac{1}{\lambda}=\frac{1}{2}$ méter.

3.17. gyakorlat megoldása:

Független normális eloszlású valószínûségi változók összege szintén normális eloszlás m_1+m_2 és $\sqrt{\sigma_1^2+\sigma_2^2}$ paraméterekkel. A keresett valószínûséget a

$$\int_{-\infty}^{x_0} \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2}} \cdot e^{\frac{-[x-(m_1+m_2)]^2}{2(\sigma_1^2+\sigma_2^2)}}$$

integrál értéke adja meg.

6.4. Negyedik fejezet gyakorlatai

4.1. gyakorlat megoldása:

A (ξ_1, ξ_2) pár az $\{(x,y) : x \geqslant 0, y \geqslant 0\}$ első síknegyedben veszi fel értékeit, és a $P(\xi_1 > \lambda\xi_2)$ valószínűség a

$$\int\int_H f(x,y)\,dx\,dy$$

kettős integrál a $H = \{(x,y) : 0 \leq \lambda y \leq x\}$ tartományon. A függetlenség miatt

$$f(x,y) = \beta_1\beta_2 e^{-\beta_1 x} e^{-\beta_2 y}$$

az együttes sűrűségfüggvény. Ezért

$$\int\int_H f(x,y)\,dx\,dy = \int_{x=0}^{\infty} \Big(\int_{y=0}^{x/\lambda} \beta_1\beta_2 e^{-\beta_1 x} e^{-\beta_2 y}\,dy \Big) dx\,,$$

ami kiintegrálva azt az eredményt adja, ami az igazolandó állításban van. □

4.3. gyakorlat megoldása: Megoldás a 4.2. tétel (4.4.2) alapján

$$f_\xi(x) = \begin{cases} \frac{1}{2\pi} & \text{ha} \quad 0 \leqslant x \leqslant 2\pi \\ \\ 0 & \text{különben} \quad M(\xi) = \pi \end{cases} .$$

$M(\cos(\xi)) = 0$, $M(\sin(\xi))$ és $M(\sin\xi\cos\xi) = 0 \Rightarrow r = 0$, azaz korrelálatlanok. □

4.5. gyakorlat megoldása:

Legyen ξ_1 a véletlen pont x-koordinátája és ξ_2 az y-koordinátája. Ekkor a pont origótól való távolsága $\eta = \sqrt{\xi_1^2+\xi_2^2}$, ami kettő szabadságfokú χ-eloszlást követ. η sűrűségfüggvényét a (3.5.3) képlet tartalmazza $f(x) = xe^{-x^2/2}$, $x > 0$. Az eloszlásfüggvény

$$F(x) = \begin{cases} 1 - e^{-x^2/2} & \text{ha } x > 0 \\ 0 & \text{ha } x < 0. \end{cases}$$

Nekünk olyan r érték kell, hogy $F(r) \geqslant 0.99$ teljesüljön. Az

$$1 - e^{-x^2/2} = 0.99$$

egyenlet megoldása

$$x = \sqrt{-2\log 0.01} = 2\sqrt{\log 10}.$$

□

4.7. gyakorlat megoldása:

$$f(x,y) = \begin{cases} 1 \cdot e^{-x} \cdot 2 \cdot e^{-2y} & \text{ha} \quad 0 \leqslant x, y \\ 0 & \text{különben} \end{cases},$$

ebből $f(x) = e^{-x}$ $\lambda_\xi = 1$ exponenciális eloszlású valószínűségi változó $(0 \leqslant x)$,
$g(y) = 2 \cdot e^{-2y}$ $\lambda_\eta = 2$ exponenciális eloszlású valószínűségi változó $0 \leqslant y)$.

Összefoglalva: ξ és η független valószínűségi változók, mivel együttes sűrűségfüggvényük $f(x) \cdot g(y)$ alakú szorzat formájában írható fel, ahol

$$f(x) = \begin{cases} e^{-x} & \text{ha} \quad 0 \leqslant x \\ 0 & \text{különben} \end{cases}$$

$$g(y) = \begin{cases} 2e^{-2y} & 0 \leqslant y \\ 0 & \text{különben} \end{cases}$$

□

4.13. gyakorlat megoldása:

Megoldás definíció alapján. A mátrix $c_{i\,j}$ eleme az i-edik változónak a j-edik változóval vett kovarianciáját jelenti $\Rightarrow$ a főátlóban az egyes változók szórásnégyzetei állnak. Így

$$\sigma(\xi_1) = \sqrt{2 \cdot 44} = 1,56, \quad \sigma(\xi_2) = \sqrt{5} = 2,23, \quad \sigma(\xi) = 1,22.$$

Ezt felhasználva, ξ_1 és ξ_3 korreláció mátrixa:

$$\begin{pmatrix} 1 & \frac{0,84}{1,56 \cdot 1,22} \\ \frac{0,84}{1,56 \cdot 1,22} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0,44 \\ 0,44 & 1 \end{pmatrix}$$

□

4.14. gyakorlat megoldása:

Megoldás (4.1.4) alapján:

$$\begin{aligned} F_\eta(y) &= \int_0^y \int_0^1 \frac{1}{3}(u+v)dudv = \int_0^y \frac{1}{3}\left[\frac{u^2}{2} + uv\right]_0^1 dv \\ &= \int_0^y \left(\frac{1}{6} + \frac{v}{3}\right) dv = \frac{y}{6} + \frac{y^2}{2}. \end{aligned}$$

$$F_\eta(y) = \begin{cases} 0 & \text{ha} & y \leqslant 0 \\ \frac{y^2}{6} + \frac{y}{6} & \text{ha} & 0 < y \leqslant 2 \end{cases}$$

□

4.15. gyakorlat megoldása:

Megoldás (4.3.2) alapján:

$$M(2\xi - 3\eta + 8) = \int_0^1 \int_0^2 (2x - 3y + 8)\,\frac{1}{3}(x+y)dydx = \frac{7}{9} \qquad l \qquad \square$$

4.17. gyakorlat megoldása:

Megoldás 4.12. példa alapján. A ?? tétel szerint a $\xi - \eta$ valószínűségi változó is normális eloszlású lesz.
$M((\xi - \eta) = 177 - 170 = 7$ a várható értéke,
$\sigma^2(\xi - \eta) = 5^2 + 5^2 - 2 \cdot 0,68 \cdot 5 \cdot 5 = 16 \quad \Rightarrow \sigma = 4$ a szórása,
$P(\xi - \eta > 6)$ a kérdés.

$$P(\xi - \eta > 6) = P\left(\frac{\xi - \eta - 7}{4} > -\frac{1}{4}\right) = \Phi\left(-\frac{1}{4}\right) = 1 - \Phi\left(\frac{1}{4}\right) \approx 0,40.$$

$\square$

4.18. gyakorlat megoldása:

Megoldás (3.1.2) alapján. Mivel $\sigma = \sqrt{2,3} = 1,52$

$$P(\xi = 0) = \frac{(1,52)^0}{0!}e^{-1,53} = e^{-1,53}$$

$\square$

4.19. gyakorlat megoldása:

A megoldás alapja a 4.5. tétel és az, hogy a várható érték lineáris funkcionál, azaz

$$M(c_1X_1 + c_2X_2) = c_1M(X_1) + c_2M(X_2).$$

$$M(\xi) = M(\eta) \text{ és } M(\xi) + M(\eta) = 2 \Rightarrow M(\xi) = M(\eta) = 1$$

i.) 4.5. tételből következik $\xi - \eta$ normális eloszlású.

ii.)

$$\mathrm{cov}((\xi + \eta)(\xi - \eta)) = M((\xi + \eta)) - M(\xi + \eta)M(\xi - \eta)$$

ahol

$$M(\xi + \eta) = 2 \qquad M(\xi - \eta) = 0$$

$\square$

4.20. gyakorlat megoldása:

$$f(x, y) = \begin{cases} \frac{1}{4\pi} & \text{ha} \quad x^2 + y^2 \leqslant 4 \\ 0 & \text{egyébként} \end{cases}$$

A (4.1.4) szerint $f_\xi(x) = \int_{-\infty}^{\infty} f(x,y)dy$

$$\begin{aligned} f_\xi(x) &= \int_{-\infty}^{-\sqrt{4-x^2}} 0\ dy + \int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} \frac{1}{4\pi} dy + \int_{\sqrt{4-x^2}}^{\infty} 0\ dy \\ f_\xi(x) &= \frac{1}{4\pi}[y]_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} = \frac{1}{4\pi} \cdot 2 \cdot \sqrt{4-x^2} \\ f_\xi &= \frac{1}{2\pi}\sqrt{4-x^2} \end{aligned}$$ □

4.21. gyakorlat megoldása:

$$\sigma_{\xi_1}^2 = 2,44 \quad \sigma_{\xi_2}^2 = 5 \quad \sigma_{\xi_3}^2 = 1,5$$

Definíció szerint $\xi_1 + \xi_2$ és $\xi_1 - \xi_2$ valószínűségi változók kovarianciája:

$$\begin{aligned} & M((\xi_1+\xi_2)\cdot(\xi_1-\xi_2)) - M(\xi_1+\xi_2)\cdot M(\xi_1-\xi_2) \\ = \ & M(\xi_1^2 - \xi_2^2) - (M(\xi_1)+M(\xi_2))(M(\xi_1) - (\xi_2)) \\ = \ & M(\xi_1^2) - M(\xi_2^2) - M^2(\xi_1) + M^2(\xi_2) \\ = \ & M(\xi_1^2) - M^2(\xi_1) - (M(\xi_2^2) - M^2(\xi_2)) = \sigma_{xi_1}^2 - \sigma_{\xi_2}^2 = -2.56. \end{aligned}$$ □

6.5. Ötödik fejezet gyakorlatai

5.3. gyakorlat megoldása:

Azt kell belátni, hogy az $M(Y)$ várható érték a sokaság várható értékével azonos. A sokaság egyenletes eloszlású egy intervallumon, és feltehetjük, hogy ez az intervallum $[0,t]$ alakú. A feladat megoldásában az 5.3. példát követjük.

$$M(Y) = \frac{M(\xi_1^*) + M(\xi_n^*)}{2}$$

Az 5.3. példában kiszámoltuk az $M(\xi_n^*)$ várható értéket, ez $nt/(n+1)$ volt. Most ξ_1^* sűrűségfüggvényére van szükségünk a várható érték kiszámításához.

Legyen ξ_1^* eloszlásfüggvénye F_* sűrűségfüggvénye f_*. Ekkor

$$1 - F_*(x) = P(\xi_1^* \geqslant x) = P(\xi_1 \geqslant x, \xi_2 \geqslant x \ldots \xi_n \geqslant x) = P(\xi_1 \geqslant x)^n = \Big(1 - \frac{x}{t}\Big)^n.$$

A közvetett függvény differenciálási szabályával deriválunk:

$$\frac{d}{dx}(1 - F_*(x)) = -f_*(x) = -n\Big(1 - \frac{x}{t}\Big)^{n-1},$$

így

$$f_*(x) = n\frac{(t-x)^{n-1}}{t^{n-1}}.$$

Most jön a várható érték:

$$M(\xi_1^*) = \int_0^t x f_*(x)\, dx,$$

amit például parciális integrálással számolhatunk ki. Eredményül $t/(n+1)$ adódik. Ezért $M(Y) = t/2$. □

5.5. gyakorlat megoldása:

Az n szabadság fokú t-eloszlás definíciója: ha ξ és η olyan független valószínűségi változók,hogy ξ $N(0;1)$ eloszlású és η n-szabadságfokú χ eloszlást követ, akkor a $\frac{\sqrt{n}\xi}{\eta}$ valószínűségi változó eloszlását „n-szabadságfokú t-eloszlás"-nak nevezzük. (n számú független standard normális eloszlású valószínűségi változó négyzetösszegének eloszlását „n-szabadságfokú χ-eloszlás"-nak nevezzük.)

A nevezőben álló kifejezés χ eloszlású, mivel négyzete három darab független standard normális eloszlású valószínűségi változó négyzetösszege. A számlálóban álló kifejezés a 3.7. tétel szerint normális eloszlású ($N(0;\sqrt{2})$). A $c \cdot (\xi_1 + \xi_2)$ kifejezés akkor lesz $N(0;1)$ eloszlású, ha standardizáljuk, azaz osztjuk a szórásával.

Ebből következik, hogy $c = \sqrt{\frac{3}{2}}$.

Megjegyzés: Az, hogy $M(\xi) = 0$ és $D(\xi) = 1$, könnyen belátható a (2.3.6) azonosság alapján:

$$M\left(\frac{1}{\sqrt{2}}(\xi_1+\xi_2)\right) = \frac{1}{\sqrt{2}}\left(M(\xi_1)+M(\xi_2)\right) = 0$$

$$D^2(\xi) = D^2\left(\frac{1}{\sqrt{2}}(\xi_1+\xi_2)\right) = M\left(\frac{1}{2}\left(\xi_1^2+2\xi_1\xi_2+\xi_2^2\right)\right) = 1$$

□

5.7. gyakorlat megoldása:

A keresett valószínűség a ξ és η valószínűségi változók együttes sűrűségfüggvényének ($h(x,y)$) az $\frac{(x+y)^2}{(x-y)^2} < 4$ feltételek által meghatározott xy síkbeli tartomány feletti integráljával egyenlő (4.1.5. összefüggés).

A feltételbeli esemény pontosan akkor következik be, ha a (ξ,η) vektor-valószínűségi változó olyan tartományból veszi fel értékét, amely pontjainak koordinátái kielégítik a fenti egyenlőtlenséget. Így utóbbi bekövetkezésének valószínűsége egyenlő (ξ,η) ezen tartományokba, esésének valószínűségével, amit az együttes sűrűségfüggvény adott tartomány feletti integrálja ad meg.

A feladat megoldása három részből áll:
1.) $h(x,y)$ felírása.
2.) Az integrálási tartomány meghatározása.
3.) Az integrál kiszámítása

1.) Mivel ξ és η függetlenek, ezért együttes sűrűségfüggvényük ξ és η sűrűségfüggvényének szorzataként kapható meg.

$$h(x,y) = \frac{1}{4\pi} \cdot e^{-\frac{x^2+y^2}{4}}$$

2.) A feltételi egyenlőtlenséget egy vele ekvivalens, de könnyebben kezelhető alakra hozzuk:

$$\frac{(\xi+\eta)^2}{(\xi-\eta)^2} < 4 \Leftrightarrow \left(\frac{\xi+\eta}{\xi-\eta}\right)^2 < 4 \Leftrightarrow -2 < \frac{\xi+\eta}{\xi-\eta} < 2 \Leftrightarrow \left|\frac{\xi+\eta}{\xi-\eta}\right| < 2$$

Gyakorlásként ábrázolja!

3.) A keresett valószínűséget a $P(A) = 1 - P(\overline{A})$ összefüggés alapján számítjuk.

$$P(\overline{A}) = 2 \cdot \int_0^{\infty} \left(\int_{\frac{1}{3}x}^{3x} h(x,y)dy \right) dx$$

(Felhasználjuk, hogy $h(x,y)$ szimmetrikus az xyz koordinátarendszer z tengelyére, mivel $M(\xi)$ és $M(\eta)$ értéke nulla.)

Az integrálást polár koordinátákra való áttéréssel kell elvégezni, és az $u = \frac{-r^2}{4}$ helyettesítést kell alkalmazni.

$$P(\overline{A}) = -\frac{1}{\pi} \int_{\varphi_1}^{\varphi} \left(\int_0^{-\infty} e^u du \right) d\Phi$$

$$= \frac{1}{\pi} \left(arc\text{tg}(3) - arc\text{tg}\left(\frac{1}{3}\right)\right).$$

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

□

5.11. gyakorlat megoldása:

Az $\eta_1^2 + \eta_2^2 + \eta_3^2 + \eta_4^2$ valószínűségi változó 4 db független $N(0;1)$ eloszlású valószínűségi változó négyzetének összege, ezért eloszlása „4-szabadságfokú χ^2-eloszlás" (lásd 3.5.). Legyen

$$\eta = \sqrt{\eta_1^2 + \eta_2^2 + \eta_3 +^2 + \eta_4^2}$$

Mivel η^2 4-szabadságfokú χ^2-eloszlás, ezért η „4-szabadságfokú χ eloszlás". η felhasználásával w az alábbi alakban írható:

$$w = \frac{2\eta_5}{\eta}$$

Mivel $\eta_1, \eta_2, \ldots, \eta_5$ független valószínűségi változók, ezért η_5 és η is függetlenek. így w megfelel a „(4-szabadságfokú) t eloszlás" definíciójának.

$$w = \frac{\sqrt{4} \cdot \eta_5}{\eta} \qquad \text{lásd 5.3.}$$

□

5.13. gyakorlat megoldása:

A c paraméter értékét úgy szeretnénk megválasztani, hogy a megadott összeg a 3.5.-nek megfelelően $(*)\eta_1^2 + \eta_2^2$ alakú legyen, ahol η_1 és η_2 $N(0;1)$ eloszlású független valószínűségi változók. A függetlenség 5.1. szerint – a statisztikai mintára vonatkozó feltevés alapján – teljesül.

$(*)$-ból következik, hogy $\eta_1 = \sqrt{c}(\xi_2 + \xi_3 + \xi_3)$ illetve $\eta_2 = \sqrt{c}(\xi_4 + \xi_5 + \xi_6)$ és $c > 0$, ez $(*)$-ból következik. A várható érték linearitása miatt (2.3.6.) következik, hogy $M(\eta_1) = M(\eta_2) = 0$ / $M(\xi_i) = 0 \;\; i = 1, \ldots, 6$ / η_1 és η_2 varianciája, felhasználva a 2.5.2. tételt, és a várható érték képzés linearitását:

$$\sigma^2(\eta_1) = M(\eta_1^2) - M^2(\eta_1) = c \cdot M\left((\xi_1 + \xi_2 + \xi_3)^2\right) - (\sqrt{c})^2 M^2(\xi_1 + \xi_2 + \xi_3),$$

ahol a második tag értéke nulla. Tehát

$$\sigma^2(\eta_1) = c \cdot M\left((\xi_1+\xi_2+\xi_3)^2\right)$$

és

$$M\left((\xi_1+\xi_2+\xi_3)^2\right) = M(\xi_1^2 + 2\xi_1\cdot\xi_2 + \xi_2^2 + 2\xi_1\cdot\xi_3 + 2\xi_2\cdot\xi_3 + \xi_3^2)$$

M linearitása és a függetlenség alapján:

$$M\left((\xi_1+\xi_2+\xi_3)^2\right) = M(\xi_1^2) + M(\xi_2^2) + M(\xi_3)^2)$$

Utóbbi viszont – felhasználva, hogy $M(\eta_i) = 0$ – a közös szórásnégyzet (variancia) háromszorosa, azaz:

$$\sigma^2(\eta_1) = c\cdot 3\cdot 2^2 \quad \text{és} \quad \sigma^2(\eta_2) = 12c$$

$\eta_1^2+\eta_2^2$, „akkor 2-szabadságfokú χ^2 eloszlású”, ha $\sigma^2(\eta_1)=\sigma^2(\eta_2)=1$ teljesül. Ebből

$$c = \frac{1}{12} \quad \square$$

5.15. gyakorlat megoldása:

Megoldás az 5.4. péda jelöléseivel: $c = 0,05$, $\sigma = 0,13$, $n = ?$ Olyan n mintaméretet keresünk, amelyre

$$\begin{aligned} P\left(|\overline{x}-m|<c\right) &= 2\Phi\left(\tfrac{c\cdot\sqrt{n}}{\sigma}\right)-1 = 0,95 \\ \Phi\left(\frac{c\cdot\sqrt{n}}{\sigma}\right) &= \tfrac{0,95+1}{2} = 0,975 \\ \Phi\left(\frac{0,05\cdot\sqrt{n}}{0,13}\right) &= 0,975 \end{aligned}$$

A 6.6. táblázatból kikereshető az az x érték, amelyre $\Phi(x) = 0,975$. Azt kapjuk, hogy $x = 1,96$, így $\frac{0,05\cdot\sqrt{n}}{0,13} = 1,96 \Rightarrow n \approx 26$. $\square$

5.16. gyakorlat megoldása:

Először azt nézzük meg, hogy a három statisztika közül melyik ad torzítatlan becslést. $M(\eta_1) = t/2$, $M(\eta_2) = t$ és $M(\eta_3) = t$. Az η_1 statisztika erősen torzít, ezért teljesen elvetendő. η_2 és η_3 torzítatlanok, közülük azt kell választani, aminek kisebb a szórása. Mivel független valószínűségi változókat összeadva a szórás négyzet adódik össze, azzal érdemes dolgozni.

$$\sigma^2(\eta_2) = \sigma^2\Big(\frac{\xi_1}{3}\Big) + \sigma^2\Big(\frac{2\xi_2}{3}\Big) + \sigma^2\Big(\frac{\xi_3}{3}\Big) + \sigma^2\Big(\frac{2\xi_4}{3}\Big) = \frac{10}{9}\sigma^2,$$

hasonlóan

$$\sigma^2(\eta_3) = 2\sigma^2,$$

ahol σ jelenti a sokaság szórását. η_2 és η_3 közül ezért η_2 a hatékonyabb. $\square$

5.17. gyakorlat megoldása:

Az 5.3. példa mintájára felírjuk az eloszlásfüggvényt:

$$F_\eta(x) = P(\eta < x) = P\left(\xi_1 < x, \xi_2 < x, \xi_3 < x\right) = \left(P\left(\xi_1 < x\right)\right)^3.$$

Exponenciális eloszlás esetén:

$$F_\xi(x) = \int_{-\infty}^{x} f(x)dx = \begin{cases} 0 & \text{ha} \quad x \leqslant 0 \\ 1 - e^{-\lambda x} & \text{ha} \quad 0 < x \end{cases}$$

és

$$F_\xi(x) = \begin{cases} 0 & \text{ha} \quad x \leqslant 0 \\ \lambda e^{\lambda x} & \text{ha} \quad 0 < x \end{cases}$$

Ezt felhasználva

$$F_\eta(x) = \begin{cases} 0 & \text{ha} \quad x \leqslant 0 \\ (1 - e^{-\lambda x})^3 & \text{ha} \quad 0 < x \end{cases}$$

Deriválással ebből megkapjuk η sűrűségfüggvényét, amellyel várható értékét ki tudjuk számolni. Ha $M(\eta) = \frac{1}{\lambda}$ vagy esetleg $\frac{1}{\lambda}$ valamely aszimptotikusan torzítatlan becslésével, akkor használható az exponenciális eloszlás várható értékének becslésére.

$$F'_\eta(x) = f_\eta = \begin{cases} 0 & \text{ha} \quad x \leqslant 0 \\ 3(1 - e^{-\lambda x})^2 \cdot \left(-e^{-\lambda x}(-\lambda)\right) & \text{ha} \quad 0 < x \end{cases}$$

ezzel

$$\begin{aligned} M(\xi) = {} & \int_0^\infty x3\lambda(1 - e^{-\lambda x})^2 \cdot e^{-\lambda x} dx \\ = {} & 3\lambda \int_0^\infty xe^{-\lambda}\left(1 - 2e^{-\lambda x} + e^{-2\lambda x}\right) dx = \tfrac{11}{6} \cdot \tfrac{1}{\lambda}. \end{aligned}$$

$\frac{11}{6} \cdot \frac{1}{\lambda}$ nem egyenlő $\frac{1}{\lambda}$-val (az exponenciális eloszlás várható értékével), de az $\eta = \frac{6}{11} \cdot \max(\xi_1, \xi_2, \xi_3)$ torzítatlan becslés $\frac{1}{\lambda}$-ra! □

5.18. gyakorlat megoldása:

Egy statisztikától megköveteljük, hogy torzítatlan becslése legyen a becsülni kívánt paraméternek, azaz várható értéke legyen egyenlő a paraméter valódi értékével. Két statisztika közül azt tekintjük jobbnak vagy hatásosabbnak, amelyiknek kisebb a szórása.

Első lépésként kiszámítjuk η_1 és η_2 várható értékét

$$M(\eta_1) = M(\xi_1 + \xi_2 + \xi_3 - \xi_4) = M(\xi_1) + M(\xi_2) + M(\xi_3) - M(\xi_4)$$

Mivel $\xi_1, \xi_2, \xi_3, \xi_4$ eloszlása megegyezik a statisztikus sokaság eloszlásával, ezért a várható érték linearitása miatt:

$$M(\eta_1) = 2m \quad \text{és} \quad M(\eta_2) = M(\xi_1 + \xi_2) = 2m,$$

ahol m a $[0,\ t]$ intervallumon egyenletes eloszlású valószínűségi változó várható értékét jelöli, sűrűségfüggvénye:

$$f(x) = \begin{cases} \frac{1}{t} & \text{ha} \quad 0 < x < t \\ 0 & \text{különben} \end{cases}$$

így

$$m = \int_0^t x\frac{1}{t}dx = \frac{1}{t}\left[\frac{x^2}{2}\right]_0^t = \frac{t}{2}.$$

$M(\eta_1) = M(\eta_2) = 2\cdot\frac{t}{2} = t$, tehát mind a két statisztika torzítatlan becslése a paraméternek. Hasonlítsa össze a szórásokat!

$$\sigma^2(\eta_1) = M(\eta_1^2) - (M(\eta_1))^2 = M\left((\xi_1 + \xi_2 + \xi_3 - \xi_4)^2\right) - t^2$$

Felhasználva, hogy a tekintett valószínűségi változók függetlenek, és az $M(\xi_i^2)_{i=1,2,3,4} = \frac{t^2}{3}$ $\sigma^2(\eta_1) = \frac{t^2}{3}$ és $\sigma^2(\eta_2) = \frac{t^2}{6}$ adódik. Ennek alapján η_2 statisztikát hatásosabbnak tekintjük t becslésére. □

Függelék

1.. A Stirling-formula

A valószínűségszámításban a faktoriálisok közelítő kiszámítására gyakran használják a **Stirling-formulát**, amely legegyszerűbb alakjában az

$$\boxed{n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n}} \tag{7.1.1}$$

formában írható. Pontosabban

$$\left(\frac{n}{e}\right)^n \sqrt{2\pi n} \le n! \le \left(\frac{n}{e}\right)^n \sqrt{2\pi n} e^{\frac{1}{12n}}. \tag{6.1.2}$$

egyenlőtlenség érvényes minden n egész számra, azaz

$$n! = \theta_n \left(\frac{n}{e}\right)^2 \sqrt{2\pi n}, \qquad \text{ahol} \quad 1 \le \theta_n \le \exp(1/12n)\,. \tag{7.1.3}$$

1. példa: Alkalmazzuk a Stirling-formulát a

$$p_{2n} = \binom{2n}{n} 2^{-2n}$$

valószínűségek határértékének kiszámítására, ha $n \to \infty$! (p_{2n} az 1.9. gyakorlatban fordul elő.)

Mivel

$$p_{2n} = \frac{(2n)!}{n!n!} 2^{-2n}$$

a (7.1.3) alakú Stirling-formulát alkalmazva kapjuk, hogy

$$p_{2n} = \frac{\sqrt{4\pi n}}{2\pi n} \cdot \frac{\theta_{2n}}{\theta_n^2}\,,$$

ami 0-hoz tart, hiszen oda tart az első tört, a második pedig 1-hez.

□

1. gyakorlat: Legyen $(p_1, p_2, p_3, \ldots, p_k)$ egy valószínűségeloszlás. Igazolja, hogy

$$\lim_{n\to\infty} \frac{1}{n} \ln \frac{n!}{(p_1 n)!(p_2 n)! \ldots (p_k n)!} = -\sum_{k=1}^{k} p_i \ln p_i$$

a Stirling-formula segítségével!

Ezt a „gyakorlatot" elsőként *Ludwig Boltzmann* oldotta meg a statisztikus fizika megalapozása közben.

2. gyakorlat: Vezesse le az

$$n \ln n - n + 1 < \ln(n!) < (n+1)\ln(n+1) - (n+1) + 1$$

becslését $n!$-nak úgy, hogy az $\ln x$ függvény gráfja alatti területet integrálással számolja ki!

2.. A gamma-függvény

A **gamma-függvényt** a pozitív számokon a

$$\boxed{\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt} \tag{7.2.1}$$

integrállal értelmezzük. $\Gamma(x)$ tetszőlegesen sokszor differenciálható függvény, a differenciálás az integráljel mögött végezhető el. A gamma-függvény legfontosabb tulajdonsága

$$\Gamma(x+1) = x\Gamma(x)\,, \tag{7.2.2}$$

és ez parciális integrálással igazolható. Mivel $\Gamma(1) = 1$, teljes indukcióval adódik, hogy

$$\Gamma(n+1) = n! \tag{7.2.3}$$

minden pozitív egész n-re. A gamma-függvény a természetes számokon értelmezett faktoriális függvényt terjeszti ki tetszőlegesen sokszor differenciálható függvénnyé.

A gamma-függvény (7.2.3) tulajdonságát a Laplace-transzformáció táblázata segítségével is beláthatjuk. Ha $f(x) = x^n$, akkor a Laplace-transzformáció táblázata szerint

$$\int_0^\infty t^n e^{-tx}\, dt = \frac{n!}{x^{n+1}}\,.$$

Ez $x = 1$-et véve éppen (7.2.3)-at adja.

2. példa: Bebizonyítjuk a

$$\Gamma(1/2) = \sqrt{\pi}$$

összefüggést!

A $\Gamma(1/2)$ definíciójában a $t = u^2$ helyettesítést alkalmazzuk:

$$\Gamma(1/2) = \int_0^\infty x^{-1/2} e^{-x}\, dx = 2\int_0^\infty e^{-u^2} = 2\frac{\sqrt{\pi}}{2} = \sqrt{\pi}\,.$$

□

3. gyakorlat: Mennyi $\Gamma(3/2)$?

3.. A maximum-entrópia elve

Egy $f(x)$ sűrűségfüggvényű eloszlás entrópiáját az

$$-\int_{-\infty}^{\infty} f(x) \ln f(x)\, dx \tag{7.3.1.}$$

integrál értelmezi, és egy diszkrét $(p_1, p_2, \ldots, p_n)$ eloszlás entrópiája

$$-\sum_{i=1}^{n} p_i \ln p_i\,. \tag{7.3.2}$$

A **maximum-entrópia elve** azt mondja, hogy ha egy eloszlásról bizonyos tulajdonságokat ismerünk, de ezek nem határozzák meg az eloszlást egyértelműen, akkor legcélszerűbb azt az eloszlást feltételeznünk, amely az adott tulajdonságú eloszlások között a legnagyobb entrópiájú. Itt van néhány példa:

(a) Ha a $(p_1, p_2, \ldots, p_n)$ valószínűségeloszlásról semmit sem tudunk, de n értékét ismerjük, akkor a $p_i = 1/n$ feltételezés a legjobb, ennek a legnagyobb az entrópiája, $\ln n$.

(b) Azon $(p_1, p_2, \ldots, p_n)$ valószínűségeloszlások között, amelyekre a $\sum_i p_i E_i = E$ feltétel teljesül adott E_i és E értékekkel, a

$$p_i = \frac{E_i}{E} \frac{\exp(-E_i)}{\sum_i \exp(-E_i)} \qquad (1 \leq i \leq n) \tag{7.3.3}$$

a legnagyobb entrópiájú. Ha p_i annak valószínűsée, hogy egy részecske az i-edik állapotban van, ahol energiája E_i, akkor a $\sum_i p_i E_i = E$ éppen az energia átlagértéke. Tehát rögzitett energia átlag mellett a (7.3.3) eloszlás a legnagyobb entrópiájú. A statisztikus fizikában (7.3.3)-at **Boltzmann féle eloszlásnak** nevezik.

(c) Egy adott véges intervallumra koncentrálódó eloszlások közül az egyenletes eloszlás entrópiája a legnagyobb.

(d) Az adott várható értékű $\mathbb{R}^+$-on levő eloszlások közül a megfelelő exponenciális eloszlas entrópiája a legnagyobb.

(e) Az adott várható értékű és adott szórású eloszlások közül a megfelelő normális eloszlás entrópiája a legnagyobb.

A maximum-entrópia elve érthetővé teszi, hogy miért használjuk olyan gyakran a normális, exponenciális és egyenletes eloszlásokat.

1949-ben Claude Shannon villamosmérnök a (7.3.2) mennyiséget a bizonytalanság mértékéül válsztotta. Ha egy kisérlet n különféle kimenetele $p_1, p_2, \ldots, p_n$ valószínűségekkel következik be, akkor az igazi kimenetelt annál nagyobb eséllyel lehet megjósólni, minél kisebb az eloszlás entrópiája. A Shannon-féle entrópia az információelméletben játszik fontos szerepet.