VALÓSZÍNŰSÉGSZÁMÍTÁS

ELŐADÁSJEGYZET, 2026. ŐSZ

Bevezető

Jelen iromány a BME-VIK Mérnökinformatikus BSc képzésén 2020 őszén elhangzott valószínűségszámítás kurzushoz tartozó előadásjegyzetnek 2026 nyarán a Valószínűségszámítás és statisztika tárgy aktuális tematikájához igazított változata[1]. Előismeretként nem feltételezünk többet, mint a szak Analízis 1 kurzusának tematikájában[2] szereplő fogalmak. A jegyzet főképp a kurzus korábbi évfolyamain használt jegyzetekből építkezik[3]:

[1] Ketskeméty László, Valószínűségszámítás jegyzet, 1998.

[2] Csehi Csongor, Valószínűségszámítás jegyzet 1-4, 2018.

Felhasználtuk még a következő könyveket:

[3] Sheldon Ross, A First Course in Probability Theory, Ed. 8, Pearson Education, 2010.

[4] M. Mitzenmacher, E. Upfal, Probability and Computing, Cambridge Univ. Press, 2005

[5] B. E. Fristedt, L. F. Gray, A Modern Approach to Probability Theory, Springer, 2013.

Az érdeklődő olvasó ezekben találhat a jegyzetben szereplőnél részletesebb kibontást. További magyar nyelvű forrás lehet:

[6] Nándori Péter - Virtuális laboratóriumok a valószínűségszámítás és statisztika oktatásban [4]

Tartalomjegyzék

Bevezető 1

1. Alapfogalmak 3

1.1. Eseménytér 3

1.2. Valószínűségi mező 5

1.3. Szitaformula 6

2. A valószínűség tulajdonságai 8

2.1. Függetlenség 8

2.2. Feltételes valószínűség 9

2.3. Karger algoritmusa (kiegészítő anyag) 10

2.4. Bayes-tétel 12

3. Diszkrét valószínűségi változók 13

3.1. Valószínűségi változó 13

3.2. Várható érték véges esetben 14

3.3. Randomized Quicksort algoritmus (kiegészítő anyag) 17

3.4. Variancia és szórás 18

3.5. Várható érték végtelen diszkrét esetben 19

Csehi Csongor jegyzete alapján készítette Mészáros Szabolcs. Közreműködők: Balázs Barbara, Papp László, Szabó Dániel, Takács Balázs, Tóth Dávid. A tárgy aktuális tematikájához igazította és kibővítette Tóbiás András.
2026. őszi állapot. Utolsó frissítés: 2026. október 8..

[1] A tárgy statisztika részének anyagát Tóbiás András jegyzetkiegészítése tartalmazza.

[2] lásd portal.vik.bme.hu/kepzes/targyak/TE90AX21/

[3] Mindkét jegyzet elérhető a cs.bme.hu/~cscsgy/vsz/ weblapon.

[4] http://math.bme.hu/~nandori/Virtual_lab/stat/index.xhtml, eredetiben http://www.randomservices.org/random/

3.6. Geometriai eloszlás, örökifjú tulajdonság I ... 20
3.7. Poisson-eloszlás ... 22
4. Valószínűségi változók viszonya ... 24
4.1. Függetlenség ... 24
4.2. Diszkrét együttes eloszlás ... 25
4.3. Kovariancia ... 26
4.4. Korreláció ... 28
5. Feltételes várható érték ... 30
5.1. Feltételes várható érték eseményre ... 30
5.2. Feltételes várható érték diszkrét valószínűségi változóra ... 31
5.3. Feltételes várható érték tulajdonságai, teljes várható érték tétele ... 32
6. Folytonos valószínűségi változók ... 34
6.1. Eloszlásfüggvény ... 34
6.2. Sűrűségfüggvény ... 36
6.3. Várható érték és szórás, folytonos eset ... 37
6.4. Exponenciális eloszlás, örökifjú tulajdonság II ... 40
6.5. Várható érték, nemnegatív eset ... 41
6.6. Bertrand-paradoxon ... 43
7. Folytonos együttes eloszlás ... 44
7.1. Valószínűségi vektorváltozók ... 44
7.2. Vektorváltozók függetlensége ... 46
7.3. Kovariancia folytonos esetben ... 48
7.4. Kovarianciamátrix ... 49
8. Normális eloszlás ... 51
8.1. Az eloszlás definíciója ... 51
8.2. Standardizálás ... 52
8.3. De Moivre–Laplace-tétel ... 54
9. Határeloszlás-tételek ... 56
9.1. Nagy számok törvénye ... 56
9.2. Centrális határeloszlás-tétel ... 57
függelék A. ... 59
A.1. Konvolúció ... 59
A.2. Csebisev-egyenlőtlenség és a nagy számok gyenge törvényének bizonyítása ... 60
A.3. Kitekintés: heurisztika a de Moivre–Laplace tételhez ... 62
A.4. A centrális határeloszlás-tétel bizonyításának vázlata ... 63
A.5. Lineáris regresszió ... 65
A.6. Feltételes várható érték folytonos esetben ... 67
A.7. Teljes valószínűség tétele, folytonos eset ... 70
A.8. Multinomiális (polinomiális) eloszlás ... 71
A.9. Marshall–Olkin-féle kétváltozós exponenciális eloszlás ... 71
A.10. Többdimenziós normális eloszlás ... 72

1. Alapfogalmak

A valószínűségszámítás praktikusságát talán nem kell bizonygatni egyetlen olvasónak sem[5]: a legtöbb kísérleti tudomány támaszkodik rá valamilyen formában. Az mégis kérdés, hogy az egyszeri halandónak miért nem elég a „kedvező-per-összes" józan ésszel is kitalálható magasságaiban maradni?

Az egyik ok, hogy néha a naiv megközelítés helytelen vagy ellentmondásos eredményt ad. Ezt jól demonstrálja a számos valószínűségi paradoxon az irodalomban[6], íme az egyik:

Bertrand-féle doboz paradoxon

Adott három egyforma doboz. Az elsőben két arany érme van, a másodikban két ezüst érme, a harmadikban pedig egy arany és egy ezüst. A dobozok tartalmát nem ismerve, (egyenletesen) véletlenszerűen választva kihúzunk egy dobozból egy érmét. Feltéve, hogy a kihúzott érme arany, mi a valószínűsége, hogy a dobozban lévő másik érme is arany?

Első nekifutásra az $\frac{1}{2}$ reális tippnek tűnhet, hiszen két esetben húzhattunk arany érmét: ha az első vagy második dobozból húztunk. Ezek közül pedig csak az egyik esetben lesz a másik érme ezüst. Ugyanakkor a paradoxon helyes megoldása $\frac{2}{3}$, amit kísérlettel is igazolhatunk. Ennek magyarázata, hogy eredetileg 6-féle kimenetele lehet a húzásunknak az alapján, melyik érmét húzzuk (az érméket különbözőnek véve). Ebből a 6 esetből 3-ban húzunk arany érmét, ez tehát az összes eseteink száma. Ebből a 3 esetből 2-ben a dobozban lévő másik érme szintén arany, így a keresett valószínűség $\frac{2}{3}$.

A példából okulva érdemes definiálnunk a vizsgált fogalmainkat.

1.1. Eseménytér

A valószínűség fogalmát a **Kolmogorov-axiómák**[7] segítségével formalizálhatjuk. Kolmogorov a huszadik század nagy hatású matematikusa, aki a fentihez hasonló félreérthetőségek feloldásaként dolgozta ki azt a keretrendszert, aminek a kiindulópontját ma Kolmogorov-axiómáknak nevezünk. Maguk az axiómák a **valószínűségi mező** definíciójában szereplő feltételek (ld 1.2 alfejezet).

1.1.1. Definíció. Legyen Ω egy tetszőleges halmaz. A következő elnevezéseket fogjuk használni:

- **Eseménytér:** Ω,
- **Kimenetel:** az eseménytér egy eleme, $\omega \in \Omega$,
- **Események:** az eseménytér „kitüntetett" $A \subseteq \Omega$ részhalmazai,
- **Valószínűség:** egy eseményhez hozzárendelt $\mathbb{P}(A)$-val jelölt, 0 és 1 közti valós szám.

A fenti paradoxon esetében például 6 kimenetel van, így az eseménytér 6 elemű halmaz. Annak az A eseménynek pedig, hogy „elsőre arany érmét húzunk" a $\mathbb{P}(A)$ valószínűsége $\frac{1}{2}$.

De mi az, hogy az események „kitüntetett" részhalmazok? Honnan fogjuk tudni, egy kérdés esetében mit akarunk eseménynek nevezni, és mit nem? Röviden, azokat a részhalmazokat választjuk eseménynek, amikhez valószínűségeket szeretnénk hozzárendelni. Sok elemi feladat esetében ez nem igazi probléma: minden részhalmazt eseménynek választhatunk, mert feltesszük, hogy mindegyik részhalmaznak van értelme beszélni a valószínűségéről (még ha nem is ismerjük a pontos értékét).

1.1.2. Példa. Egy kockadobás leírásánál az eseménytér így definiálható: $\Omega \stackrel{\text{def}}{=} \{1,2,3,4,5,6\}$. Az Ω elemeit, vagyis a kimeneteleket megfeleltethetjük annak, hogy mikor milyen számot dobunk. Legyen Ω összes részhalmaza esemény. Például $\{2,4,6\}$ egy esemény. Az eseményeket sokszor logikai állításokkal határozzuk meg, így a $\{2,4,6\}$ eseményt röviden írhatjuk úgy is, hogy {párosat dobunk}.

[5] Ha valakinek mégis kellene: robotics.stanford.edu/users/sahami/papers-dir/SIGCSE11-Probability.pdf

[6] lásd még: [youtube] PBS Infinite Series - Making Probability Mathematical

[7] Az axióma – hangzásával ellentétben – nem egy lassú lefolyású megbetegedés, hanem az *alapállítás* másik neve. Olyan kijelentéseket, alapvetéseket nevezünk így, amik globális feltevések az elméletünkben: nem bizonyítjuk, viszont bárhol használhatjuk őket. Kolmogorov eredeti axiómáit lásd Foundations of the Theory of Probability.

Felmerülhet a kérdés: „Miért nem választjuk simán mindig az összes részhalmazt eseménynek, 'oszt csókolom?". Azért, mert vannak olyan helyzetek, amikor szerepe van annak, mi esemény, és mi nem. Ilyen esetekre példa:

(1) **Geometriai valószínűségek** esetén területekkel (vagy azzal analóg fogalommal) definiáljuk a valószínűségeket. Azonban ha minden részhalmazra szeretnénk értelmes területfogalmat definiálni, az nem fog sikerülni, ellentmondásokba futunk[8]. A megoldás, hogy nem minden részhalmaz esemény, így nem kell minden részhalmazra értelmeznünk annak területét.
(2) **Megfigyelhetőség**en is alapulhat, mit nevezünk eseménynek. Például ha a fenti paradoxont szeretnénk modellezni: Ω továbbra is definiálható 6 eleműnek aszerint, hogy mit húzunk. Jelölje Ω elemeit a_1, a_2, b_1, b_2, c_1, c_2 (vegyük észre, hogy Ω elemei nem kell, hogy számok legyenek). Ezen húzások közül a_1, a_2, c_1 jelöl arany érméket, a többi ezüstöt, a_1, a_2 az első láda tartalmát, b_1, b_2 a másodikat és így tovább. A húzás ismeretében $\{a_1, a_2, c_1\}$ illetve $\{b_1, b_2, c_2\}$ részhalmazok megfigyelhetők, míg például $\{c_1, c_2\}$ nem, hiszen nem tudjuk, hogy a harmadik dobozból húztunk-e. Néhány problémánál érdemes pontosan azon részhalmazokat eseménynek nevezni, amik megfigyelhetők. Ilyen probléma például a feltételes várható érték számolása is.
(3) **Folyamatok**, vagyis időben változó véletlen mennyiségek esetében az idő múlásával változhat, hogy mit tudunk megfigyelni és emiatt mit tartunk eseménynek. Lásd még Markov-láncok, martingálok.

A gyakorlatban előforduló legtöbb egyszerű példában azonban megfelelő választás, ha Ω minden részhalmazát eseménynek tekintjük. Leginkább akkor, ha az Ω eseménytér számossága megszámlálható, azaz véges vagy megszámlálhatóan végtelen. (A végtelen halmazok számosságával kapcsolatban lásd a 3.5. fejezetet!)

Nézzük, milyen műveleteket végezhetünk eseményekkel.

1.1.3. Állítás. *Mivel az események halmazok, így értelmezve van események* ***uniója*** *($A \cup B$),* ***metszete*** *($A \cap B$) és Ω-ra vett* ***komplementere*** *($\overline{A}$).*

Két esemény **különbsége** az előbbiekkel leírható: $A \setminus B = A \cap \overline{B}$. Két esemény **kizáró**, ha $A \cap B = \emptyset$. Az Ω-ra használatos még a **biztos esemény** elnevezés. Hasonlóan, az üreshalmaz (jele: $\emptyset$) neve a továbbiakban **lehetetlen esemény**.

1.1.4. Példa. A kockadobálós példánál maradva, a {párosat dobunk} esemény komplementere a {páratlant dobunk}, a {párosat dobunk} és a {3-nál nagyobbat dobunk} események metszete a $\{4, 6\}$, míg uniója a $\{2, 4, 5, 6\}$.

Végiggondolható, hogy ha az események kijelentésekkel vannak megfogalmazva (pl. {párosat dobunk}), akkor az uniójuk megfelel a kijelentések szintjén a „vagy" műveletnek, metszetük az „és"-nek, egy esemény komplementere pedig a logikai tagadásnak.

A halmazoknál megszokott tulajdonságok itt sem vesztik érvényüket: $A \cup B = B \cup A$, $A \cap \Omega = A$ és a többi. Névvel is bíró, megjegyzendő azonosság az alábbi:

1.1.5. Állítás. (de Morgan-azonosságok) *Két halmazra:*

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \qquad \text{és} \qquad \overline{A \cap B} = \overline{A} \cup \overline{B},$$

illetve végtelen sok halmazra:

$$\overline{\bigcup_{i=1}^{\infty} A_i} = \bigcap_{i=1}^{\infty} \overline{A_i} \qquad \text{és} \qquad \overline{\bigcap_{i=1}^{\infty} A_i} = \bigcup_{i=1}^{\infty} \overline{A_i}.$$

Az első állításpár Venn-diagramon könnyen ellenőrizhető.

[8] lásd en.wikipedia.org/wiki/Vitali_set

Feladat. Legyenek A, B és C események. Írjuk fel a következő eseményeket a fenti műveletek segítségével: a) legalább egy esemény teljesül, b) A és B teljesül, de C nem, c) minden esemény teljesül, d) egyik esemény sem teljesül, e) pontosan egy esemény teljesül.

A továbbiakban $\mathcal{F}$-fel fogjuk jelölni az események halmazát. Ω összes részhalmazának halmazát pedig $\mathcal{P}(\Omega)$-val jelölünk és Ω *hatványhalmaz*ának nevezzük. Amint már említettük, a gyakorlatban sokszor megfelelő a $\mathcal{F} = \mathcal{P}(\Omega)$ választás.[9]

1.2. Valószínűségi mező

Volt szó arról, hogy mi az eseménytér, mik az események, de eddig nem kerültek elő valószínűségek. Fentebb említettük, hogy azon $A \subseteq \Omega$ halmazok események, amiknek szeretnénk a $\mathbb{P}(A)$ valószínűségéről beszélni. Vagyis a valószínűség egy $\mathcal{F} \to [0,1]$ függvény kell legyen, ahol $\mathcal{F}$ az események halmaza. De ennél többet is tudnia kell.

1.2.1. Definíció. Legyen $\Omega \neq \emptyset$ egy eseménytér és $\mathcal{F}$ az események halmaza. Ekkor egy $\mathbb{P} : \mathcal{F} \to [0,1]$ függvényt **valószínűségi mérték**nek nevezünk, ha $\mathbb{P}(\Omega) = 1$, és teljesül a következő:

Ha $A_1, A_2, \ldots, A_i, \cdots \in \mathcal{F}$ olyan eseménysorozat, amire minden $i \neq j$ esetén $A_i \cap A_j = \emptyset$, akkor

$$\mathbb{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

A feltétel röviden úgy olvasható: páronként kizáró eseménysorozat uniójának valószínűsége az események valószínűségeinek (végtelen) összege. Röviden ezt a tulajdonságot **σ-additivitás**nak nevezzük.

1.2.2. Definíció. Egy $(\Omega, \mathcal{F}, \mathbb{P})$ hármast (Kolmogorov-féle) **valószínűségi mező**nek hívunk, ha $\mathcal{F}$ az Ω-hoz tartozó események halmaza és $\mathbb{P}$ valószínűségi mérték.

1.2.3. Példa. A mérsékelten kreatív példánknál maradva, egy kockadobás esetén ha $\Omega = \{1,2,3,4,5,6\}$, és $\mathcal{F}$ az összes részhalmaz, akkor egy A esemény $\mathbb{P}(A)$ valószínűsége $|A|/|\Omega|$, például $\mathbb{P}(\{1,2,5,6\}) = \frac{4}{6}$. Az ilyen valószínűségi mezőket (amikor $\mathbb{P}(A) = |A|/|\Omega|$) **klasszikus valószínűségi mező**nek hívjuk. Ennek általánosítása a **geometriai valószínűségi mező**, amikor Ω a sík, tér (vagy $\mathbb{R}^n$) egy részhalmaza, és $\mathbb{P}(A) = \lambda(A)/\lambda(\Omega)$ ahol λ a terület, a térfogat vagy az n-dimenziós térfogat.

1.2.4. Példa. Nézzünk egy 5 kérdéses tesztet, amin minden kérdés eldöntendő (igen-nem típusú), és a kitöltő mindegyikre 60% eséllyel ad helyes választ a többi kérdésre adott választól függetlenül. Ekkor az 5 hosszú 0-1 sorozatok tere, azaz $\Omega = \{0,1\}^5$ modellezheti a feladatot ($\mathcal{F}$ pedig az összes részhalmaz). Ez már nem klasszikus valószínűségi mező, mert a csak a $(0,1,0,1,0)$ kimenetelt tartalmazó esemény valószínűsége $0.4^3 \cdot 0.6^2 \approx 0.023$, nem pedig $\frac{1}{2^5} \approx 0.031$.

Mit várunk el egy jól működő valószínűség fogalomtól? Például a következő tulajdonságot, ami ugyan nem szerepel a definícióban, de könnyen levezethető belőle.

1.2.5. Állítás. *Ha A és B kizáró, akkor* $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Bizonyítás. Használjuk a valószínűségi mérték definíciójában szereplő σ-additivitást azzal a választással, hogy $A_1 = A$, $A_2 = B$, illetve $A_i = \emptyset$ minden $i \geq 3$ esetén. Ekkor

$$\mathbb{P}(A \cup B) = \mathbb{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) = \mathbb{P}(A) + \mathbb{P}(B) + \sum_{i=3}^{\infty} \mathbb{P}(\emptyset).$$

[9] Azokban az itt nem részletezett esetekben, amikor $\mathcal{F} = \mathcal{P}(\Omega)$ nem megfelelő választás, $\mathcal{F}$ akkor is Ω bizonyos részhalmazainak családja lesz, vagyis $\mathcal{F} \subseteq \mathcal{P}(\Omega)$. Ekkor is meg fogjuk például követelni, hogy $\mathcal{F}$ olyan legyen, hogy két esemény uniója (ill. metszete, különbsége) szintén esemény legyen. A pontos követelmény az, hogy $\mathcal{F}$-nek úgynevezett *σ-algebrá*nak (ejtsd: szigma-algebra) kell lennie, ami azt jelenti, hogy eleme kell, hogy legyen a teljes tér (azaz maga Ω), zártnak kell lennie a komplementer-képzésre és a megszámlálható unióra. Ez a három tulajdonság magára a hatványhalmazra természetesen szintén teljesül. Az, hogy a lehetetlen eseményt is mindig eseménynek tekintjük, szintén következik a három tulajdonságból (világos?).

Ez csak úgy történhet, ha $\mathbb{P}(\emptyset) = 0$ (különben a jobb oldal végtelen lenne, míg a bal 0 és 1 közti). Innen az állítás már következik. □

1.2.6. Következmény. *Tetszőleges $A, B \in \mathcal{F}$ eseményekre a következők teljesülnek:*

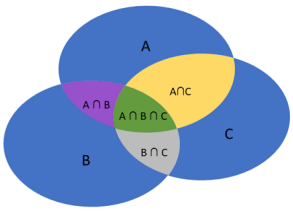
(1) $\mathbb{P}(A) + \mathbb{P}(\overline{A}) = 1$,
(2) $\mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A)$,
(3) ha $B \subseteq A$, *akkor* $\mathbb{P}(B) \leq \mathbb{P}(A)$.

Bizonyítás. Az első állításhoz alkalmazzuk a véges additivitást az (egymást kizáró) A és $\overline{A}$ eseményekre, a második állításhoz pedig az $A \cap B$ és $A \cap \overline{B}$ eseményekre. A harmadik tulajdonság következik a másodikból, hiszen $B \subseteq A$ esetén $A \cap B = B$, és $\mathbb{P}(A \cap \overline{B}) \geq 0$. □

Feladat. Vegyünk egyenletesen véletlenszerűen egy egyszerű irányítatlan gráfot az $\{a, b, c, d\}$ négyelemű csúcshalmazon. (Ekkor az eseménytér egy 64 elemű halmaz.) Melyiknek nagyobb az esélye: hogy a gráf fagráf, vagy hogy legfeljebb három éle van?

1.3. Szitaformula

Hogyan lehet kiszámolni az unió valószínűségét, ha az események nem feltétlenül kizáróak? A válasz a szitaformula (vagy más néven Poincaré-formula), amihez szintén csak a fenti véges additivitást kell használnunk.



1.3.1. Példa. Legyen A, B, C három esemény, például egy céltáblán három részhalmaz eltalálásának eseménye, amelyek közül legalább az egyiket el akarjuk találni.

A $\mathbb{P}(A \cup B \cup C)$ valószínűség kiszámolásához kiindulhatunk a $\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C)$ mennyiségből. Vegyük észre, hogy itt a metszetek valószínűségét duplán számoltuk, vagyis le kell vonjuk a $\mathbb{P}(A \cap B) + \mathbb{P}(A \cap C) + \mathbb{P}(B \cap C)$ összeget, ha jobb közelítést szeretnénk $\mathbb{P}(A \cup B \cup C)$-ra. De ezzel sem vagyunk kész, hiszen az $A \cap B \cap C$ rész valószínűségét háromszor adtuk hozzá, de háromszor is vontuk le, pedig egyszer kellene számoljuk. Tehát végül

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C)$$

adódik. Ezt a formulát általánosítja n eseményre a szitaformula.

A tétel rövid kimondásához további jelölésre van szükségünk. Jelölje $[n]$ az $\{1, 2, \ldots, n\}$ halmazt. Legyenek $A_1, A_2, \ldots, A_n \in \mathcal{F}$ események, továbbá legyen $I = \{i_1, i_2, \ldots, i_k\}$ egy k elemű részhalmaza az $[n]$ halmaznak. Ekkor nézhetjük a $\bigcap_{i \in I} A_i$ esemény valószínűségét, vagyis valamely k darab különböző esemény egyszerre teljesülésének valószínűségét. Végül definiáljuk az

$$S_k \stackrel{\text{def}}{=} \sum_{I \subseteq [n] \colon |I| = k} \mathbb{P}\Big(\bigcap_{i \in I} A_i\Big)$$

számot, azaz az összes lehetséges módon kiválasztunk k darab különböző eseményt, majd ezek metszeteinek valószínűségét mind összegezzük. Például $k = 1$ esetén $S_1 = \sum_{i=1}^{n} \mathbb{P}(A_i)$.

1.3.2. Tétel (Szitaformula)**.** *Legyenek $A_1, A_2, \ldots, A_n \in \mathcal{F}$ események. Ekkor*

$$\mathbb{P}\Big(\bigcup_{j=1}^{n} A_j\Big) = \sum_{k=1}^{n} (-1)^{k+1} S_k.$$

Felmerülhet kérdésként, hogy mi történik, ha a szummának csak az első néhány tagját nézzük, a maradékot elhanyagoljuk (például mert a gyakorlatban annyira kicsik és lassan számolhatók). Az unió valószínűségéről ekkor is kaphatunk információt.

1.3.3. Állítás (Bonferroni-egyenlőtlenségek)**.** *Legyenek $A_1, A_2, \dots, A_n \in \mathcal{F}$ események, és $1 \le m_1, m_2 \le n$ egészek, ahol m_1 páratlan, és m_2 páros. Ekkor*

$$\mathbb{P}\Big(\bigcup_{j=1}^{n} A_j\Big) \le \sum_{k=1}^{m_1} (-1)^{k+1} S_k, \quad \text{és} \quad \mathbb{P}\Big(\bigcup_{j=1}^{n} A_j\Big) \ge \sum_{k=1}^{m_2} (-1)^{k+1} S_k.$$

A fenti két állítást nem bizonyítjuk. Példákat lásd a gyakorlaton.

1.3.4. Következmény (Boole-egyenlőtlenség)**.** *Legyenek $A_1, A_2, \dots, A_n \in \mathcal{F}$ események. Ekkor*

$$\mathbb{P}\Big(\bigcup_{j=1}^{n} A_j\Big) \le \sum_{j=1}^{n} \mathbb{P}(A_j), \quad \text{és} \quad \mathbb{P}\Big(\bigcap_{j=1}^{n} A_j\Big) \ge 1 - \sum_{j=1}^{n} \mathbb{P}(\overline{A_j}).$$

Bizonyítás. Alkalmazzuk az első Bonferroni-egyenlőtlenséget $m_1 = 1$ választással, ebből éppen az első egyenlőtlenségünk adódik, hiszen S_1 a $\mathbb{P}(A_j)$-k összege. A második egyenlőtlenség következik az elsőből és a De Morgan-azonosságból, ha azt az $\overline{A_j}$ eseményekre alkalmazzuk. □

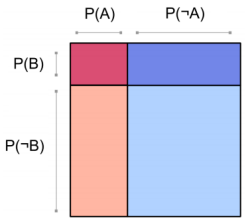
Az állítás a Bonferroni-egyenlőtlenség nélkül, teljes indukcióval is könnyen belátható.

2. A valószínűség tulajdonságai

A továbbiakban mindig feltesszük, hogy adott egy $(\Omega, \mathcal{F}, \mathbb{P})$ valószínűségi mező. Ebben a fejezetben a függetlenség és a feltételes valószínűség fogalmait vesszük sorra.

2.1. Függetlenség

Korábban már foglalkoztunk azzal az esettel, amikor két esemény uniójának valószínűsége összeadódik (azaz $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$). Ehhez arra volt szükség, hogy az események kizáróak legyenek. A feladatokban viszont van olyan eset is, amikor a valószínűségek bizonyos feltételek teljesülése esetén szorzódnak.



2.1.1. Definíció. Az A és B eseményeket **függetlenek**nek nevezzük, ha

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Valójában a függetlenség a kizáró eseményektől nagyban eltérő fogalom. Azt a helyzetet próbálja formalizálni, amikor a két esemény bekövetkezése nem befolyásolja egymást.

2.1.2. Példa. Ha A esemény $\frac{1}{3}$ eséllyel következik be (azaz átlagosan három próbálkozásból egyszer teljesül), B esemény pedig $\frac{1}{4}$ valószínűséggel, és nem tételezünk fel köztük kapcsolatot, akkor a bekövetkezésük esélyét, hétköznapi tapasztalatainkra alapozva, $\frac{1}{3} \cdot \frac{1}{4} = \frac{1}{12}$-nek vesszük.

Vegyük észre, hogy a függetlenség a valószínűségek szintjén van megfogalmazva, így olyan események is lehetnek függetlenek (a fenti definíció értelmében), amikről úgy érezzük „hatásuk van egymásra". Például két kockadobás esetén az {első dobás 1-es} és a {két dobás megegyezik} események függetlenek.

2.1.3. Állítás. *Ha A és B függetlenek, akkor A és $\overline{B}$ is függetlenek.*

Bizonyítás. Használjuk fel a korábban belátott $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B})$ azonosságot. Ebből az A és B függetlenségével következik, hogy

$$\mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\big(1 - \mathbb{P}(B)\big) = \mathbb{P}(A)\mathbb{P}(\overline{B}),$$

ami éppen a belátandó egyenlőség. □

Definiáljuk több esemény függetlenségét is.

2.1.4. Definíció. Az $A_1, \dots, A_n$ események **(együttesen) független**ek, ha minden $I \subseteq [n]$ esetén

$$\mathbb{P}\Big(\bigcap_{i \in I} A_i\Big) = \prod_{i \in I} \mathbb{P}(A_i).$$

Más szavakkal az események közül valahány metszetének valószínűsége a valószínűségek szorzata.

A definíció túlbonyolítottnak tűnhet, de később kiderül, hogy ez a jó fogalom. Felmerülhetne, hogy miért nem csak az összes n eseményre követeljük meg, hogy $\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdot \dots \cdot \mathbb{P}(A_n)$? Hiszen ha az összes esemény független, akkor közülük k is az, nem? Hát nem teljesen.[10] Oké, akkor legyenek páronként függetlenek, abból már biztosan következik az együttes függetlenség? Sajnos ez sem nyert. A következő példa mutatja, mennyire alattomos fogalom az együttes függetlenség.

2.1.5. Példa. Dobjunk fel két szabályos érmét. Legyen $A_1 = \{\text{első érme fej}\}$, $A_2 = \{\text{második érme fej}\}$, $A_3 = \{\text{dobott fejek száma páros}\}$. Ekkor A_i független A_j-től akármilyen $i \neq j$-re, viszont $\{A_1, A_2, A_3\}$ nem együttesen független, hiszen

$$\mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3) = \frac{1}{2^3} = \frac{1}{8}, \qquad \text{míg} \qquad \mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(\text{mindkét érme fej}) = \frac{1}{4}.$$

A példának van lineáris algebrai analógja is: az $(1,0)$, $(0,1)$, $(1,1)$ vektorok közül bármely pár lineárisan független, de együtt már nem azok.

[10] Ha néhány esemény együttesen független, abból valóban következik közülük néhány együttes függetlensége, de ehhez a fenti együttes függetlenség definícióra van szükség.

2.2. Feltételes valószínűség

Hogyan lehet „mérni”, egy esemény mennyire függ egy másiktól?

2.2.1. Definíció. Legyenek $A, B \in \mathcal{F}$ események. Tegyük fel, hogy $\mathbb{P}(A) > 0$. Ekkor a B esemény A-ra vett feltételes valószínűsége

$$\mathbb{P}(B \mid A) \stackrel{\text{def}}{=} \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}.$$

Kiolvasva: „B valószínűsége, feltéve A”.

Vegyük észre, hogy A és B pontosan akkor függetlenek, ha $\mathbb{P}(B \mid A) = \mathbb{P}(B)$. Más szavakkal, B független A-tól, ha B valószínűsége nem függ attól, hogy A bekövekezett-e. Valójában a függetlenséget definiálhatnánk a $\mathbb{P}(B \mid A) = \mathbb{P}(B)$ egyenlettel is, azokban az esetekben, mikor $\mathbb{P}(A) > 0$.[11]

2.2.2. Példa. Nézzünk néhány példát kockadobással. Legyen $A = \{\text{párosat dobunk}\}$. Ekkor $\mathbb{P}(\text{6-ost dobunk} \mid A) = \frac{1}{3}$, $\mathbb{P}(\text{1-est dobunk} \mid A) = 0$, $\mathbb{P}(\text{3-nál nagyobbat dobunk} \mid A) = \frac{2}{3}$ és $\mathbb{P}(\text{párosat dobunk} \mid A) = 1$.

Természetesen a feltételes valószínűség nem csak az események összefüggésének mérésére szolgál. Több problémánál is felmerülhet, hogy feltételes információink vannak, például „ha alaposan felkészülten érkezem vizsgázni, akkor $1 - \varepsilon$ eséllyel átmegyek”.[12]

Nézzük, milyen tulajdonságai vannak a feltételes valószínűségnek.

2.2.3. Állítás. *Legyen $A \in \mathcal{F}$ rögzített esemény, amire $\mathbb{P}(A) > 0$. Ekkor az A-ra vett feltételes valószínűség, vagyis az alábbi $\mathcal{F} \to [0,1]$ függvény:*

$$B \mapsto \mathbb{P}(B \mid A),$$

szintén valószínűségi mérték.

Nagyszerű, de mire megyünk ezzel az állítással? Például arra, hogy az összes korábban $\mathbb{P}$-re elhangzott állításba behelyettesíthetjük $\mathbb{P}(\ \ )$ helyére $\mathbb{P}(\ \mid A)$-t, az állítás akkor is érvényben marad.

Bizonyítás. Egyrészt világos, hogy $\mathbb{P}(\Omega \mid A) = \frac{\mathbb{P}(\Omega \cap A)}{\mathbb{P}(A)} = 1$. Másrészt legyen $B_1, B_2, \dots$ események egy páronként kizáró rendszere. Felhasználva, hogy $\mathbb{P}$ valószínűségi mérték:

$$\mathbb{P}\Big(\bigcup_{i=1}^{\infty} B_i \,\Big|\, A\Big) = \mathbb{P}\Big(\Big(\bigcup_{i=1}^{\infty} B_i\Big) \cap A\Big) \Big/ \mathbb{P}(A) =$$

$$= \mathbb{P}\Big(\bigcup_{i=1}^{\infty} (B_i \cap A)\Big) \Big/ \mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(B_i \cap A)/\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(B_i \mid A),$$

ami épp a bizonyítandó állítás. □

A feltételes valószínűség segítségével lehet kimondani az *esetszétválasztás* valószínűségi megfelelőjét:

2.2.4. Állítás (*Teljes valószínűség tétele*)**.** *Legyenek $A_1, \dots, A_n \in \mathcal{F}$ páronként kizáró események, amikre $\cup_{i=1}^n A_i = \Omega$ és $\mathbb{P}(A_i) > 0$ minden i-re. Ekkor*

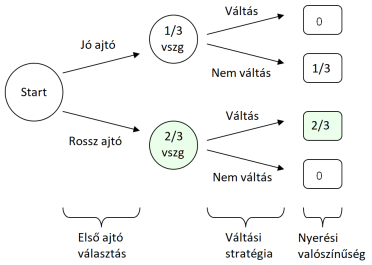
$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i).$$

2.2.5. Definíció. Egy $A_1, \dots, A_n \in \mathcal{F}$ páronként kizáró eseményekből álló sorozatot **teljes eseményrendszer**nek hívunk, ha $\cup_{i=1}^n A_i = \Omega$.

Állítás bizonyítása. A feltételes valószínűség definícióját visszahelyettesítve egyszerűsíthetünk $\mathbb{P}(A_i)$-vel, így kapjuk, hogy a jobb oldal $\sum_{i=1}^n \mathbb{P}(B \cap A_i)$. Mivel a feltételek szerint $\cup_{i=1}^n (B \cap A_i) = B \cap \Omega = B$, így $\mathbb{P}$ additivitásából már következik az állítás. □

[11] Lásd még [youtube] MIT OpenCourseWare - Conditional Probability.

[12] A feltételes valószínűség az első előadáson szerepelt Bertrand doboz paradoxonhoz is kapcsolódik.



2.2.6. Példa (Monty Hall-paradoxon)**.** Adott három ajtó, az egyik mögött egy autó, a másik kettő mögött egy-egy kecske áll. A feladvány, hogy először választanunk kell egy ajtót, majd a játékvezető kinyitja valamelyik másik ajtót, ami mögött kecske van. Ezután lehetőségünk van változtatni a választásunkon. Kérdés: megéri-e, feltéve hogy az autó választását preferáljuk a kecskékkel szemben?

A meglepő válasz a „mindegy" helyett az, hogy igen. Ugyanis ha nem változtatunk a döntésünkön, akkor a nyerési esélyünk nyilván $\frac{1}{3}$. Míg ha változtatunk, akkor

$$\begin{aligned}\mathbb{P}(\text{végül autó}) =& \mathbb{P}(\text{végül autó} \mid \text{elsőre kecske})\mathbb{P}(\text{elsőre kecske}) \\ &+ \mathbb{P}(\text{végül autó} \mid \text{elsőre autó})\mathbb{P}(\text{elsőre autó}) = 1 \cdot \frac{2}{3} + 0 \cdot \frac{1}{3} = \frac{2}{3},\end{aligned}$$

hiszen ha elsőre kecskét választunk, akkor a játékvezető csak a másik kecskés ajtót nyithatja ki.

Előfordul olyan is, amikor egy problémánál több, egymásra épülő feltétel esetén fennálló valószínűségekkel kell dolgozni.

2.2.7. Példa. Három húzást végzünk visszatevés nélkül egy megkevert 52 lapos franciakártya-pakliból. Mekkora a valószínűsége annak, hogy elsőre királyt, másodikra dámát, harmadikra pedig bubit húzunk? Ugyan az első húzás eredménye befolyásolja a második húzás valószínűségeit (egy király kihúzása csökkenti az újbóli király húzásának esélyét), mégis a helyes eredményt a következő számolás adja.

Jelölje K_1, hogy elsőre királyt húzunk, D_2 azt, hogy másodszorra dámát, míg B_3 azt, hogy harmadszorra bubit. Ekkor a keresett esemény valószínűsége:

$$\mathbb{P}(K_1)\mathbb{P}(D_2 \mid K_1)\mathbb{P}(B_3 \mid D_2 \cap K_1) = \frac{4}{52} \cdot \frac{4}{51} \cdot \frac{4}{50} \approx 0{,}0005.$$

Ezt a módszert általánosítja a következő állítás.

2.2.8. Állítás (*Szorzási szabály*)**.** *Legyenek $A_1, \dots, A_n \in \mathcal{F}$ események, amikre $\mathbb{P}(A_i) > 0$ $(\forall i)$. Ekkor*

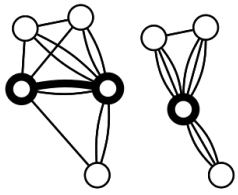
$$\mathbb{P}\Big(\bigcap_{i=1}^{n} A_i\Big) = \mathbb{P}(A_1) \cdot \prod_{i=2}^{n} \mathbb{P}\Big(A_i \,\Big|\, \bigcap_{k=1}^{i-1} A_k\Big).$$

A bizonyításhoz elég kibontani a feltételes valószínűség definícióját és egyszerűsíteni a szorzatot.

2.3. Karger algoritmusa (kiegészítő anyag)

A szorzási szabály és a függetlenség alkalmazásaként nézzünk egy véletlen algoritmust. Legyen $G = (V, E)$ egy irányítatlan (multi)gráf, akár többszörös élekkel együtt, de hurokélek nélkül. Keressük a gráf egy minimális elemszámú vágását, azaz egy $V = A \cup B$ felbontást, ahol A, B diszjunktak, és a lehető legkevesebb él fut A és B közt.

A feladat visszavezethető az irányított gráfok maximális folyam keresésére, aminek megoldását megkereshetjük a Ford-Fulkerson-algoritmus segítségével. Vegyük észre a lényeges különbséget a két kérdés közt: a maximális folyamkeresésénél s és t rögzített, míg a mostani problémában nem.



A fenti úgynevezett globális minimális vágás problémának van egy véletlenített megoldása is, ez a Karger-algoritmus.

Az input: egy összefüggő, irányítatlan gráf (a tárolás módjával most nem foglalkozunk), az output az élek egy részhalmaza. Az algoritmusban két lépést iterálunk felváltva: előbb választunk egyenletesen véletlenszerűen egy élet, majd összehúzzuk/azonosítjuk az él két végpontját, a hurokéleket elhagyjuk, a többi élet megtartjuk.

Ezt addig csináljuk, amíg 2 pontja nem marad a gráfnak. Az eredmény meghatároz egy vágást: az eredeti gráf csúcsai közül az egyik pontra összehúzott csúcsok lesznek az A halmaz, a másikra összehúzottak a B.

Ha az algoritmust egyszer lefuttatjuk, akkor kapunk egy véletlenszerű vágást, de közel sem biztos hogy ez minimális. Futtassuk tehát sokszor, és nézzük meg, melyik eredmény volt a legjobb (vagyis az utolsó lépésben a két pont közt a legkevesebb élet tartalmazó). A következő állítás azt mondja, hogy ez már észszerűen sok próbálkozás után is nagy eséllyel optimális megoldást ad.

2.3.1. Állítás. *A Karger-algoritmus egyszeri futtatása esetén legalább $\frac{2}{n^2}$ eséllyel globális minimális vágást kapunk.*

Bár a $\frac{2}{n^2}$ nagyon kis valószínűségnek tűnik, de ha $\frac{n^2}{2}\ln n$ alkalommal futtatjuk az algoritmust, akkor a sikertelenség esélye a függetlenség miatt már csak

$$\left(1-\frac{2}{n^2}\right)^{\frac{n^2}{2}\ln n} \leq \frac{1}{e^{\ln n}} = \frac{1}{n}$$

felhasználva, hogy az $m \mapsto \left(1-\frac{1}{m}\right)^m$ monoton növő és $\frac{1}{e}$-hez tart. Tehát jó eséllyel globális minimális vágást kapunk.

Bizonyítás. Legyen F egy globális minimális vágás által elvágott élek halmaza. Az algoritmus pontosan akkor találja meg F-et, ha egyetlen élét sem húzza össze. Legyen $|E| = m$, $|F| = k$ és jelölje A_i azt az eseményt, hogy az i-edik lépésben nem F-beli élet húzunk össze. Ekkor a szorzási szabály miatt

$$\mathbb{P}(\text{siker}) = \mathbb{P}\Big(\bigcap_{i=1}^{n-2} A_i\Big) = \prod_{i=1}^{n-2} \mathbb{P}\Big(A_i \,\Big|\, \bigcap_{j=1}^{i-1} A_j\Big)$$

ahol $\mathbb{P}\Big(A_i \,\Big|\, \cap_{j=1}^{i-1} A_j\Big)$ annak a valószínűsége, hogy az i-edik lépésben nem F-beli élet választunk, feltéve, hogy az első $i-1$ lépésben sem választottunk ki egyetlen F-beli élet. Ezt a valószínűséget szeretnénk alulról becsülni, amihez szükségünk van a gráf csúcs- és élszámára.

Az i-edik lépés előtt $n-(i-1)$ csúcsa van a gráfnak. Mivel az F minimális vágás elemszáma k, emiatt minden csúcs foka legalább k, még az összehúzások után is. Hiszen ha valamely (egyesített) csúcs foka kisebb lenne, akkor a csúcsból kiinduló élek megfelelői az eredeti gráfban egy k-nál kisebb elemszámú vágást adnának. Emiatt az i-edik lépés előtt a gráfnak legalább $\frac{k(n-(i-1))}{2}$ éle van. Tehát

$$\mathbb{P}\Big(A_i \,\Big|\, \bigcap_{j=1}^{i-1} A_j\Big) \geq 1 - \frac{k}{\frac{k(n-(i-1))}{2}} = 1 - \frac{2}{n-i+1}$$

Ezt behelyettesítve kapjuk, hogy

$$\mathbb{P}(\text{siker}) \geq \Big(1-\frac{2}{n}\Big)\Big(1-\frac{2}{n-1}\Big)\ldots\Big(1-\frac{2}{n-(n-3)}\Big) =$$

$$= \frac{(n-2)\cdot(n-3)\cdot\ldots\cdot(3-2)}{n\cdot(n-1)\cdot\ldots\cdot 3} = \frac{2}{n(n-1)} \geq \frac{2}{n^2},$$

ami épp a belátandó állítás. □

Megjegyzés. A véletlen algoritmusok két osztályba sorolhatók az alapján, az algoritmus milyen tulajdonsága véletlen: a futásideje vagy a megoldásának helyessége. Ha egy algoritmus biztosan a helyes eredményre jut (avagy jelzi, hogy a feladatnak nincs megoldása), de a futásidő nemcsak a bemenetnek, hanem a véletlennek is függvénye, az algoritmust *Las Vegas algoritmus*nak hívjuk. Míg ha a futásidő csak a bemenettől függ, azaz randomizált választásoktól független, viszont csak bizonyos valószínűséggel kapunk helyes eredményt, akkor egy *Monte Carlo algoritmus*sal állunk szemben.

2.4. Bayes-tétel

A feltételes valószínűséget érintő jelenségek közül kiemelendő a Bayes-tétel és a paradoxon, amit felold. (A paradoxon más néven is ismert, pl. fals pozitív paradoxon, avagy base rate fallacy).

Bayes-paradoxon

Röntgenvizsgálat során 0,95 annak a valószínűsége, hogy tbc-s beteg betegségét felfedezik. Annak valószínűsége, hogy egy egészséges embert betegnek találnak 0,001. A tbc-ben szenvedők aránya a lakosságon belül 0,0001. Mennyi annak a valószínűsége, hogy az ember egészséges, ha átvilágításkor betegnek találták?

A megoldás azon alapul, hogy összefüggést írunk fel a $\mathbb{P}(A \mid B)$ és a $\mathbb{P}(B \mid A)$ feltételes valószínűségek között, ahol $A = \{\text{az alany egészséges}\}$, és $B = \{\text{pozitív a teszt}\}$.[13]

2.4.1. Állítás. (Egyszerű Bayes-tétel) *Legyenek $A, B \in \mathcal{F}$ események, amikre $\mathbb{P}(A) > 0$ és $\mathbb{P}(B) > 0$ teljesül. Ekkor*

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}.$$

A bizonyítás a definíciók behelyettesítésével rögtön következik. Sokszor a tételt a teljes valószínűség tételével kombinálva alkalmazzák:

2.4.2. Állítás. (Bayes-tétel) *Legyenek $B, A_1, A_2, \ldots, A_n \in \mathcal{F}$ események, amikre $\mathbb{P}(B) > 0$, $\mathbb{P}(A_i) > 0$ minden i-re, és $A_1, \ldots, A_n$ teljes eseményrendszer. Ekkor*

$$\mathbb{P}(A_1 \mid B) = \frac{\mathbb{P}(B \mid A_1)\mathbb{P}(A_1)}{\sum_{i=1}^{n} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}.$$

Bizonyítás. Írjuk fel az egyszerű Bayes-tételt A_1-re és B-re, majd bontsuk ki a nevezőt a teljes valószínűség tételével:

$$\mathbb{P}(A_1 \mid B) = \frac{\mathbb{P}(B \mid A_1)\mathbb{P}(A_1)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid A_1)\mathbb{P}(A_1)}{\sum_{i=1}^{n} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)},$$

ami épp a belátandó állítás. □

2.4.3. Példa. Térjünk vissza a fenti példára. Legyen $A_1 = \{\text{az alany egészséges}\}$, $A_2 = \overline{A_1}$ és $B = \{\text{pozitív a teszt}\}$. Ekkor

$$\mathbb{P}(A_1 \mid B) = \frac{\mathbb{P}(B \mid A_1)\mathbb{P}(A_1)}{\mathbb{P}(B \mid A_1)\mathbb{P}(A_1) + \mathbb{P}(B \mid A_2)\mathbb{P}(A_2)} = \frac{0{,}001 \cdot 0{,}9999}{0{,}001 \cdot 0{,}9999 + 0{,}95 \cdot 0{,}0001} \approx 0{,}9132$$

ami nem fest túl jó képet a bizonyos szempontból 95% biztonságúnak tekintett tesztről. Az eredmény csak látszólagos ellentmondás, ami abból fakad, hogy a vizsgált populációban lényegesen több egészséges ember van, így több „lehetőségünk" van fals pozitív eredményt kapni, mint fals negatív eredményt.

Megjegyzés. Bár a Bayes-tétel egy ártatlan állításnak tűnhet a feltételes valószínűségekről, valójában messzemenő következményei vannak. A valószínűségszámítás elsődleges alkalmazási területén, a statisztikában a Bayes-féle modellek külön megközelítést képviselnek; amik közvetve a Bayes-tétel továbbgondolásából alakultak ki, Laplace bábáskodása mellett. A tétel történetével egy könyvet is meg lehetne tölteni, olyannyira, hogy meg is töltöttek:

> S. B. McGrayne, The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy, Yale University Press.

A könyvről összefoglaló: www.lesswrong.com/posts/RTt59BtFLqQbsSiqd/a-history-of-bayes-theorem

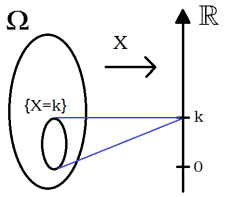
[13] Lásd még [youtube] Crash Course Statistics #24.

3. Diszkrét valószínűségi változók

Az előző két előadásban szereplő definíciók (eseményalgebra, feltételes valószínűség) ugyan alapfogalmai a témának, de nem elégségesek, hogy természetes módon le tudjunk írjunk bizonyos problémákat.

Például hogyan tudnánk megfogalmazni az eddigi eszközökkel olyan egyszerű állításokat, hogy két kockadobás eredménye független? Vagy hogy egy kockadobás átlagos eredménye 3,5, egy 0 és 1 közt egyenletesen kiválasztott véletlen szám átlagos értéke pedig $\frac{1}{2}$? Ezekhez arra van szükségünk, hogy ne csak eseményekről, hanem véletlen mennyiségekről (ún. valószínűségi változókról) is beszélni tudjunk.

3.1. Valószínűségi változó



3.1.1. Definíció. Legyen $X : \Omega \to \mathbb{R}$ függvény. Adott $x \in \mathbb{R}$ valós számra jelölje

$$\{X < x\} \stackrel{\text{def}}{=} \{\omega \in \Omega \mid X(\omega) < x\},$$

vagyis azon kimenetelek halmazát, amikor X kisebb, mint x. Ezeket a halmazokat az X **nívóhalmaz**ainak hívják. Az X függvényt **valószínűségi változó**nak nevezzük, ha minden $x \in \mathbb{R}$-re

$$\{X < x\} \in \mathcal{F},$$

azaz röviden X nívóhalmazai események.

3.1.2. Példa. Az eddigi példáinkban is szerepeltek már valószínűségi változók, csak nem neveztük őket a nevükön. Néhány példa valószínűségi változóra:

(1) Egy kockadobás eredménye. A valószínűségi változó definíciójában szereplő „$\{X < x\} \in \mathcal{F}$ minden valós x-re" feltétel ebben az esetben ekvivalens azzal, hogy minden k-ra azon kimenetelek halmaza, amelyek esetén k-t dobtunk, egy esemény.
(2) Kockadobás eredményének négyzete. Lehetséges értékei 1, 4, 9, 16, 25 és 36, mindegyik lehetőséget $\frac{1}{6}$ eséllyel veszi fel. Formálisan felírva választhatjuk az eseményteret $\Omega = \{1, 2, 3, 4, 5, 6\}$-nak, $\mathcal{F}$ és $\mathbb{P}$ ahogy korábban, a valószínűségi változó pedig $Y(\omega) = \omega^2$.
(3) Egy urnában 2 fehér és 3 piros golyó van. Visszatevés nélkül addig húzunk, amíg fehéret nem húztunk. A fehér kihúzásáig húzott piros golyók száma egy valószínűségi változó.
(4) Valószínűségi változót eseményből is kaphatunk. Legyen

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{ha } \omega \in A, \\ 0 & \text{egyébként.} \end{cases}$$

Ezt hívjuk az A eseményhez tartozó **indikátor valószínűségi változó**nak.

Megjegyzés. Az $\{X < x\} \in \mathcal{F}$ feltétel helyett használhattuk volna az $\{X \leq x\} \in \mathcal{F}$ feltételt is (ahogy néhány más jegyzet teszi is), ahol értelemszerűen $\{X \leq x\} \stackrel{\text{def}}{=} \{\omega \in \Omega \mid X(\omega) \leq x\}$. Ez a módosítás nem változtatna a fenti definíció értelmén, azaz ekvivalens definíciót kapnánk. Hasonlóan definiálhatók az $\{X = x\}$, $\{X > x\}$, de akár az $\{a < X < b\}$ halmazok is, amik azon kimenetelek halmazai, amikre teljesül a zárójeles állítás. Belátható, hogy ezek szintén események.

3.1.3. Definíció. Egy X valószínűségi változó **diszkrét**, ha értékkészlete megszámlálható (azaz véges vagy megszámlálhatóan végtelen). Az értékkészletet $\mathrm{Ran}(X)$-szel jelöljük.

Az X valószínűségi változó **súlyfüggvény**ének a

$$p_X \colon \mathrm{Ran}(X) \to [0, 1], \qquad k \mapsto p_X(k) \stackrel{\text{def}}{=} \mathbb{P}(X = k)$$

függvényt nevezzük.

Vagyis a p_X súlyfüggvény az X valószínűségi változó minden lehetséges k értékére megadja, hogy X milyen valószínűséggel veszi fel a k értéket. Így p_X-et értelmezhetjük $\mathbb{R} \to [0,1]$ függvényként is: minden olyan k-hoz, amely nincs X értékkészletében, nullát rendel.

Kitérő. A végtelen halmazok sem mind ugyanakkorák, azaz nincs bármelyik kettő közt kölcsönösen egyértelmű megfeleltetés. Emiatt megkülönböztetünk megszámlálhatóan végtelen és megszámlálhatatlanul végtelen halmazokat.

Megszámlálhatóan végtelen az, aminek fel tudjuk sorolni az elemeit egy (természetes számokkal indexelt) sorozatként. Ilyen például $\mathbb{Z}$, az egész számok halmaza (ami többek közt felsorolható a következőképp: $0, 1, -1, 2, -2, 3, -3 \dots$), de a racionális számok is[14]. A megszámlálhatóan végtelen halmazok mind ugyanakkorák, vagyis bármely kettő közt van kölcsönösen egyértelmű megfeleltetés.

Megszámlálhatatlanul végtelen az a halmaz, ami végtelen, de nem megszámlálhatóan végtelen. Ilyen például $\mathbb{R}$, a valós számok halmaza, de a $[0,1]$ intervallumon értelmezett Riemann-integrálható függvények halmaza is. A megszámlálhatatlanul végtelen halmazok nem mind ugyanakkorák, például az előző két példa halmaz sem.[15]

3.2. Várható érték véges esetben

Egy véletlen mennyiség esetében az egyik legtermészetesebb kérdés, hogy "Jó-jó, véletlen, de úgy átlagban mennyi?". Ezt a véletlenszerű esetek közti "átlagos" értéket fogja meg a várható érték fogalma.

3.2.1. Definíció. Egy X valószínűségi változó **egyszerű**, ha csak véges sok értéket vehet fel. Egy egyszerű valószínűségi változó **várható értéke**:

$$\mathbb{E}(X) \stackrel{\text{def}}{=} \sum_{k \in \mathrm{Ran}(X)} k \cdot \mathbb{P}(X = k), \tag{3.1}$$

ahol $\mathrm{Ran}(X)$ az X véges értékkészlete, és $\mathbb{P}(X = k)$ jelöli az $\{X = k\}$ esemény valószínűségét.

Mit is jelent ez? Miért lesz ez a fura szumma "átlagos érték"? A képlet azt mondja, hogy a véletlen X értékeinek vegyük a súlyozott közepét, ahol a súlyok az egyes értékek valószínűségei. Az elnevezés némileg szerencsétlen: az érték, amit kapunk nem feltétlenül "várható". Pl. ha csukott szemmel felveszünk egy papucsot, akkor vagy 2 vagy 0 lábunkon lesz a megfelelő papucsfél, mégis a helyesen felvett papucsok számának várható értéke 1 (azonos esélyeket feltételezve).

Fontos, hogy a képletben szerepel a k szorzó, anélkül ugyanis $\sum_{k \in \mathrm{Ran}(X)} \mathbb{P}(X = k) = 1$ bármilyen egyszerű X változó esetén.

3.2.2. Példa. Számoljuk ki a fenti példákban szereplő valószínűségi változók várható értékeit:

(1) A kockadobás esetén $\mathrm{Ran}(X) = \{1,2,3,4,5,6\}$, illetve $\mathbb{P}(X = k) = \frac{1}{6}$ minden $k \in \mathrm{Ran}(X)$-re, ezért
$$\mathbb{E}(X) = \sum_{k=1}^{6} k \cdot \mathbb{P}(X = k) = \sum_{k=1}^{6} k \cdot \frac{1}{6} = \frac{21}{6} = 3{,}5.$$

(2) A kockadobás négyzetére hasonlóan
$$\mathbb{E}(Y) = \sum_{k \in \{1,4,9,16,25,36\}} k \cdot \mathbb{P}(Y = k) = (1+4+9+16+25+36) \cdot \frac{1}{6} = \frac{91}{6} \approx 15{,}1667.$$

[14] lásd BSZ1 jegyzet: cs.bme.hu/bsz1/jegyzet/bsz1_jegyzet.pdf

[15] Formálisan a (3.4) képlet megfelel a (3.1) képletnek, és a kezdeti példában szereplő valószínűségi változó esetén nyugodtan használhattuk volna azt a képletet is, hiszen itt az értékkészlet egy lehetséges felsorolása $k_1 = 1, k_2 = 2$ stb. Lehetnek azonban olyan megszámlálhatóan végtelen értékkészletű valószínűségi változók is, ahol az értékkészlet $\{k_1, k_2, \dots\}$ felsorolása nem lehet monoton növekvő. Ilyen például az az eset, amikor $k_1, k_2, \dots$ a racionális számok egy felsorolása – lásd BSZ1 jegyzet – és X olyan, hogy $\mathbb{P}(X = k_i) = 2^{-i}$. Ez a valószínűségi változó minden racionális számot pozitív valószínűséggel felvesz.

(3) Jelölje Z a fehér golyó kihúzásáig húzott piros golyók számát:

$$\mathbb{E}(Z) = \sum_{k=0}^{3} k \cdot \mathbb{P}(Z=k) \overset{\text{szorzási szabály}}{=}$$
$$= 0 \cdot \frac{2}{5} + 1 \cdot \frac{3}{5}\frac{2}{4} + 2 \cdot \frac{3}{5}\frac{2}{4}\frac{2}{3} + 3 \cdot \frac{3}{5}\frac{2}{4}\frac{1}{3}\frac{2}{2} = \frac{0+3+4+3}{10} = 1.$$

(4) Indikátor valószínűségi változó várható értéke:

$$\mathbb{E}(\mathbf{1}_A) = \sum_{k\in\{0,1\}} k \cdot \mathbb{P}(\mathbf{1}_A = k) = 0 \cdot \mathbb{P}(\mathbf{1}_A = 0) + 1 \cdot \mathbb{P}(\mathbf{1}_A = 1) = \mathbb{P}(A).$$

Ilyen értelemben a várható érték kiterjesztése a valószínűségeknek az indikátor változókról az (egyelőre csak egyszerű) valószínűségi változókra.

Valószínűségi változókra – ahogy valós értékű függvényekre – definiálhatók a szokásos műveletek: ha X és Y valószínűségi változó, akkor $X+Y$ az a függvény, amire $(X+Y)(\omega) = X(\omega) + Y(\omega)$. Belátható, hogy az összeg is valószínűségi változó.[16] Hasonlóan definiálhatjuk valószínűségi változók különbségét, szorzatát, illetve ha a nevező sehol sem nulla, akkor hányadosát.

A várható érték egyik sűrűn használt tulajdonsága, hogy lineáris. Ez alatt egyrészt azt értjük, hogy bármilyen $c \in \mathbb{R}$ esetén $\mathbb{E}(cX) = c \cdot \mathbb{E}(X)$ (ez még egyszerűen látszik). Másrészt, hogy $\mathbb{E}$ additív:

3.2.3. Állítás. *Legyenek X és Y egyszerű valószínűségi változók. Ekkor $\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$.*

Bizonyítás. Jelöle a $\mathrm{Ran}(X+Y)$ halmazt M, $\mathrm{Ran}(X)$-et K és $\mathrm{Ran}(Y)$-t L. Ekkor a definíciókat kibontva

$$\mathbb{E}(X+Y) = \sum_{m\in M} m \cdot \mathbb{P}(X+Y=m) = \sum_{m\in M} m \cdot \mathbb{P}\Big(\bigcup_{\substack{k\in K,\, l\in L \\ k+l=m}} \{X=k,\, Y=l\}\Big) =$$
$$= \sum_{m\in M} \sum_{\substack{k\in K,\, l\in L \\ k+l=m}} (k+l) \cdot \mathbb{P}(X=k,\, Y=l) = \sum_{k\in K}\sum_{l\in L} (k+l) \cdot \mathbb{P}(X=k,\, Y=l)$$
$$= \sum_{k\in K} k \cdot \mathbb{P}\Big(\bigcup_{l\in L} \{X=k,\, Y=l\}\Big) + \sum_{l\in L} l \cdot \mathbb{P}\Big(\bigcup_{k\in K} \{X=k,\, Y=l\}\Big)$$
$$= \sum_{k\in K} k \cdot \mathbb{P}(X=k) + \sum_{l\in L} l \cdot \mathbb{P}(Y=l) = \mathbb{E}X + \mathbb{E}Y,$$

ami épp a bizonyítandó állítás. □

Az additivitás hasznos eszköz olyankor is, amikor alapból nincs szó valószínűségi változók összegéről.

3.2.4. Példa. Bizonyítsuk be a 3 halmazra vonatkozó szitaformulát, azaz hogy

$$\mathbb{P}(A_1 \cup A_2 \cup A_3) = \sum_{i=1}^{3} \mathbb{P}(A_i) - \sum_{1\le i<j\le 3} \mathbb{P}(A_i \cap A_j) + \mathbb{P}(A_1 \cap A_2 \cap A_3).$$

Ekkor a fenti indikátor valószínűségi változós példa miatt:

$$\mathbb{P}(\cup_i A_i) = 1 - \mathbb{P}\big(\cap_i \overline{A_i}\big) = 1 - \mathbb{E}\Big(\mathbf{1}_{\cap_i \overline{A_i}}\Big) = 1 - \mathbb{E}\Big(\prod_i \mathbf{1}_{\overline{A_i}}\Big) = 1 - \mathbb{E}\Big(\prod_i (1 - \mathbf{1}_{A_i})\Big) =$$
$$= 1 - \mathbb{E}\Big(1 - \mathbf{1}_{A_1} - \mathbf{1}_{A_2} - \mathbf{1}_{A_3} + \mathbf{1}_{A_1\cap A_2} + \mathbf{1}_{A_2\cap A_3} + \mathbf{1}_{A_1\cap A_3} - \mathbf{1}_{A_1\cap A_2\cap A_3}\Big)$$

[16] Nem egyszerű valószínűségi változók esetén a bizonyítás nem magától értetődő. Érdemes használni hozzá, hogy a racionális számok sűrűn helyezkednek el, és így $\{X+Y<x\} = \bigcup_{r\in\mathbb{Q}} (\{X<r\} \cap \{Y<x-r\})$.

$$= \sum_{i=1}^{3} \mathbb{P}(A_i) - \sum_{1 \le i < j \le 3} \mathbb{P}(A_i \cap A_j) + \mathbb{P}(A_1 \cap A_2 \cap A_3),$$

ami épp a bizonyítandó állítás.

Vegyük észre, hogy a számolás első sorában nem használtuk, hogy 3 halmazról beszélünk. Valójában ugyanez az érvelés tetszőleges véges sok halmazra elmondható, és így bebizonyítható a szitaformula.

Láttuk, hogy az egyes valószínűségi változók várható értékének meghatározásához elég volt a $\mathbb{P}(X = k)$ értékeket ismernünk. Ezen valószínűségek összességét hívjuk az egyszerű valószínűségi változó **eloszlás**ának. Nézzünk néhány nevezetes eloszlást:

3.2.5. Definíció. Egy X valószínűségi változó **binomiális eloszlás**ú, $n \in \mathbb{N}$ és $p \in [0, 1]$ paraméterekkel, ha

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k \in \{0, 1, 2, \ldots, n\}.$$

Jelölése: $X \sim \mathrm{Bin}(n; p)$.

3.2.6. Példa. Dobjunk fel egy olyan pénzérmét n-szer, ami p valószínűséggel mutat fejet egy dobás után. Ekkor a fej dobások száma egy binomiális eloszlású valószínűségi változó.

Általánosan, ha független kísérleteket végzünk, amiknek azonos a sikervalószínűségük, akkor n kísérletből a sikerek száma binomiális eloszlású n és p paraméterekkel. Formálisan ez a következőképp írható fel: legyenek $A_1, \ldots, A_n$ együttesen független események. Tegyük fel, hogy $\mathbb{P}(A_i) = p$ minden i-re. Ekkor

$$\mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_n} \sim \mathrm{Bin}(n; p),$$

vagyis a $\mathrm{Bin}(n; p)$ eloszlású valószínűségi változóra mindig nézhetünk úgy, mint n darab együttesen független esemény indikátorainak összegére.

Ennek a megfigyelésnek a hasznossága rögtön látható, ugyanis ha $X \sim \mathrm{Bin}(n; p)$, akkor

$$\mathbb{E}(X) = \mathbb{E}(\mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_n}) = \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n) = n \cdot p$$

a várható érték additivitása okán.

3.2.7. Definíció. Egy X valószínűségi változó **egyenletes eloszlás**ú egy n elemű $S \subseteq \mathbb{R}$ halmazon, ha

$$\mathbb{P}(X = k) = \frac{1}{n}$$

minden $k \in S$ esetén. Ha $S = \{1, 2, \ldots, n\}$, akkor X várható értéke $\mathbb{E}(X) = \frac{1+2+\cdots+n}{n} = \frac{n+1}{2}$.

A várható érték további hasznos tulajdonsága, hogy monoton:

3.2.8. Állítás. *Legyen X egyszerű valószínűségi változó és a egy valós szám. Ha $\mathbb{P}(X \ge a) = 1$, akkor $\mathbb{E}(X) \ge a$.*

Bizonyítás. Ha $\mathbb{P}(X \ge a) = 1$, akkor

$$\mathbb{E}(X) = \sum_{k \in \mathrm{Ran}(X)} k\mathbb{P}(X = k) \ge \sum_{k \in \mathrm{Ran}(X)} a\mathbb{P}(X = k) = a \sum_{k \in \mathrm{Ran}(X)} \mathbb{P}(X = k) = a.$$

□

3.2.9. Következmény. *Legyenek X, Y egyszerű valószínűségi változók. Ha $\mathbb{P}(X \le Y) = 1$, akkor $\mathbb{E}(X) \le \mathbb{E}(Y)$.*

Bizonyítás. Mivel $\mathbb{P}(X \le Y) = 1$, vagyis $\mathbb{P}(Y - X \ge 0) = 1$, ezért az előbbi állítás szerint $\mathbb{E}(Y - X) \ge 0$. Végül a várható érték additivitása miatt $0 \le \mathbb{E}(Y - X) = \mathbb{E}(Y + (-X)) = \mathbb{E}(Y) - \mathbb{E}(X)$, amiből következik a következmény. □

3.3. Randomized Quicksort algoritmus (kiegészítő anyag)

Az előző előadáson szó volt a Karger-algoritmusról. Nézzünk egy hasonló példát, a várható érték alkalmazhatóságát demonstrálandó.

Input: egy $x_1, \dots, x_n$ különböző elemekből álló tömb ($n \geq 1$). Output: ugyanezen elemek tömbje sorba rendezve. Az algoritmus a következő: Ha a lista egy elemű, visszatérési érték a lista. Egyébként választunk egy p elemet (neve: *pivot elem*), a többieket pedig szétválogatjuk két tömbre: p-nél kisebbek és p-nél nagyobbak (ez $n-1$ összehasonlítást jelent). Alkalmazzuk rekurzívan a quicksort algoritmust a kapott két tömbre, majd adjuk vissza az ebből összekonkatenált eredményt: p-nél kisebbek rendezve, aztán p, végül p-nél nagyobbak rendezve.

Ez egy Las Vegas algoritmus[17], vagyis egyesélyes az eredmény (biztosan jó eredményt kapunk), csak az nem világos, meddig tart eljutni odáig. Legrosszabb esetben mindenkit mindenkivel össze kell hasonlítanunk, így $\binom{n}{2}$ összehasonlítást végzünk: például ha már eleve sorba van rendezve a tömb, és mindig a legelső elemet választjuk pivot elemnek.

Rendben, van amikor lassú, de mégis meddig tart átlagosan? Ez attól is függ, hogyan választjuk a p pivot elemeket. Tegyük fel, hogy a p választása egyenletesen véletlenül történik, és a különböző quicksort hívásokban egymástól függetlenül.

3.3.1. Állítás. *Jelölje X_n a quicksort algoritmusban elvégzett összehasonlítások (véletlen) számát, ha a bemenet hossza n. Ekkor $\mathbb{E}(X_n) \leq 2n \ln n$.*

Bizonyítás. Legyen $y_1, \dots, y_n$ az algoritmus kimenete (vagyis a rendezett tömb). Legyen $X_{i,j}$ az a valószínűségi változó, ami pontosan akkor 1, ha valamikor az eljárás során össze kellett hasonlítanunk az y_i és az y_j számokat, egyébként 0. Mivel minden összehasonlítást legfeljebb egyszer végzünk el, így

$$X_n = \sum_{i<j} X_{i,j}.$$

Vegyük észre, hogy az $X_{i,j}$-k nem függetlenek. De ettől még teljesül, hogy

$$\mathbb{E}X_n = \mathbb{E}\Big(\sum_{i<j} X_{i,j}\Big) = \sum_{i<j} \mathbb{E}X_{i,j}.$$

Tehát elég meghatároznunk az $\mathbb{E}X_{i,j} = \mathbb{P}(X_{i,j} = 1)$ mennyiségeket.

Nézzük az $y_i, y_{i+1}, \dots, y_{j-1}, y_j$ számokat. Az algoritmus definíciója miatt előbb-utóbb mindegyikük lesz pivot elem, de hogy milyen sorrendben, az véletlenszerű. Az $X_{i,j} = 1$ feltétel (azaz hogy y_i-t és y_j-t össze kellett hasonlítanunk valamikor) pontosan akkor teljesül, ha ezen számok közül a legelőször vagy y_i-t vagy y_j-t választjuk pivot elemnek. Ha nem ez történne, y_i és y_j külön résztömbben folytatná karrierjét, és így sosem kerülne összehasonlításra.

Mivel a pivot elemeink egymástól függetlenül egyenletesen választódnak ki, annak az esélye, hogy legelőször y_i-t vagy y_j-t választunk, éppen $\frac{2}{j-i+1}$. Tehát

$$\mathbb{E}X_n = \sum_{i<j} \mathbb{E}X_{i,j} = \sum_{1 \leq i < j \leq n} \frac{2}{j-i+1} = \sum_{k=2}^{n} (n-k+1)\frac{2}{k} =$$

$$= -2(n-1) + (n+1)\sum_{k=2}^{n} \frac{2}{k} = -4n + (2n+2)\sum_{k=1}^{n} \frac{1}{k}.$$

Belátható, hogy $\sum_{k=1}^{n} \frac{1}{k} \leq \ln n + 1$, így $\mathbb{E}X_n \leq -4n + (2n+2)(\ln n + 1) \leq 2n \ln n$. □

[17] Lásd a Karger-algoritmus utáni megjegyzést.

3.4. Variancia és szórás

3.4.1. Definíció. Egy X valószínűségi változó **szórásnégyzet**e, vagy más néven **varianciá**ja:

$$\mathbb{E}\big((X-\mathbb{E}X)^2\big).$$

Jelölés: $\mathbb{D}^2(X)$ (alternatív jelölése: $\mathrm{Var}(X)$). Egy valószínűségi változónak nem mindig létezik szórásnégyzete (hiszen lehet olyan eset, hogy már $\mathbb{E}(X)$ is értelmetlen), de ha létezik, akkor nemnegatív. Sőt, a várható érték definíciójából az is következik, hogy X szórásnégyzete csak akkor lehet 0, ha X értéke 1 valószínűséggel konstans (és vegyük észre, hogy ebben az esetben a konstans értéke megegyezik X várható értékével).

A szórásnégyzet négyzetgyökét **szórás**nak hívjuk, jelölése: $\mathbb{D}(X)$.

Más szavakkal, X szórásnégyzete az X-nek az átlagos értékétől való négyzetes eltérése.

A szórásnégyzetet leggyakrabban az alábbi állítás segítségével a legegyszerűbb kiszámítani.

3.4.2. Állítás. *Ha egy X valószínűségi változó szórásnégyzete létezik, akkor*

$$\mathbb{D}^2(X)=\mathbb{E}(X^2)-\mathbb{E}(X)^2.$$

Bizonyítás. A szórásnégyzet definíciójában szereplő zárójelet felbontva és a várható érték linearitását használva

$$\mathbb{D}^2(X)=\mathbb{E}\big((X-\mathbb{E}X)^2\big)=\mathbb{E}(X^2-2\mathbb{E}(X)X+\mathbb{E}(X)^2)=\mathbb{E}(X^2)-2\mathbb{E}(X)\mathbb{E}(X)+\mathbb{E}(X^2)=\mathbb{E}(X^2)-\mathbb{E}(X^2),$$

ahol az utolsó előtti lépésben azt használtuk, hogy a konstans $\mathbb{E}(X^2)$ értékű valószínűségi változó várható értéke önmaga. □

Hogyan határozhatjuk meg az előző állításban szereplő $\mathbb{E}(X^2)$ várható értéket? Egyrészt meghatározhatjuk X^2 súlyfüggvényének kiszámításával, mint a 3.1.2. Példában. Másrészt pedig érvényes az alábbi állítás.

3.4.3. Állítás. *Ha X egyszerű valószínűségi változó és $g\colon \mathbb{R}\to\mathbb{R}$ függvény, akkor*

$$\mathbb{E}(g(X))=\sum_{k\in\mathrm{Ran}(X)} g(k)\mathbb{P}(X=k). \tag{3.2}$$

Esetünkben konkrétan

$$\mathbb{E}(X^2)=\sum_{k\in\mathrm{Ran}(X)} k^2\mathbb{P}(X=k). \tag{3.3}$$

A 3.4.3. Állítás bizonyítása.

$$\begin{aligned}\mathbb{E}(g(X))&=\sum_{\ell\in\mathrm{Ran}(g(X))}\ell\,\mathbb{P}(g(X)=\ell)=\sum_{\ell\in\mathrm{Ran}(g(X))}\ell\sum_{k\in\mathrm{Ran}(X)\colon\, g(k)=\ell}\mathbb{P}(X=k)\\&=\sum_{\ell\in\mathrm{Ran}(g(X))}\sum_{k\in\mathrm{Ran}(X)\colon\, g(k)=\ell} g(k)\mathbb{P}(X=k)=\sum_{k\in\mathrm{Ran}(X)} g(k)\mathbb{P}(X=k).\end{aligned}$$

□

A szórásnégyzet mennyiség arról szolgáltat információt, hogy X értékei mennyire terülnek szét az átlaga körül. Ilyen „szétterülést” mérő számot többféleképp definiálhatnánk, például nézhetnénk az $\mathbb{E}(|X-\mathbb{E}X|)$-et is. Hogy mégis a szórásnégyzet a népszerű mérőszám erre, annak az egyik oka az alábbi állítás.

3.4.4. Állítás. *Tegyük fel, hogy $\mathbb{D}(X)$ létezik és véges. Ekkor tetszőleges $c\in\mathbb{R}$ esetén*

$$\mathbb{D}^2(X+c)=\mathbb{D}^2(X)\qquad\text{és}\qquad\mathbb{D}^2(cX)=c^2\mathbb{D}^2(X),$$

illetve

$$\mathbb{D}(X+c)=\mathbb{D}(X) \qquad \text{és} \qquad \mathbb{D}(cX)=|c|\mathbb{D}(X),$$

azaz a szórás eltolás-invariáns és abszolút homogén.

Bizonyítás. A szórásnégyzet definícióját kibontva

$$\mathbb{D}^2(X+c)=\mathbb{E}\Big(\big(X+c-\mathbb{E}(X+c)\big)^2\Big)=\mathbb{E}\big((X-\mathbb{E}X)^2\big)=\mathbb{D}^2(X), \qquad \text{illetve}$$

$$\mathbb{D}^2(cX)=\mathbb{E}\Big(\big(cX-\mathbb{E}(cX)\big)^2\Big)=\mathbb{E}\big(c^2(X-\mathbb{E}X)^2\big)=c^2\mathbb{D}^2(X),$$

amely egyenletekből gyökvonással adódnak a szórásra vonatkozó állítások. □

3.4.5. Példa.

(1) Legyen K egyenletes eloszlású valószínűségi változó az $\{1,2,3,4,5,6\}$ halmazon. (Vajon miért jelöljük K-val?) Ekkor a 3. előadás példája szerint $\mathbb{E}(K^2)=\frac{91}{6}$, míg $\mathbb{E}(K)=\frac{7}{2}$, ezért

$$\mathbb{D}^2(K)=\frac{91}{6}-\Big(\frac{7}{2}\Big)^2=\frac{182-147}{12}=\frac{35}{12}\approx 2{,}9167, \quad \text{és} \quad \mathbb{D}(K)=\sqrt{\mathbb{D}^2(K)}\approx 1{,}7078.$$

(2) Vizsgáljuk az $\mathbf{1}_A$ indikátor valószínűségi változót, és jelölje az A esemény valószínűségét p. Ekkor

$$\mathbb{D}^2(\mathbf{1}_A)=\mathbb{E}(\mathbf{1}_A^2)-\mathbb{E}(\mathbf{1}_A)^2=\mathbb{E}(\mathbf{1}_A)-\mathbb{E}(\mathbf{1}_A)^2=\mathbb{P}(A)-\mathbb{P}(A)^2=p(1-p).$$

(3) Legyen $X\sim \mathrm{Bin}(n;p)$. Bár $\mathbb{D}^2(X)$ kibontható a definíció alapján is, célszerűbb felhasználni, hogy felírható $X=\mathbf{1}_{A_1}+\dots+\mathbf{1}_{A_n}$ alakban, ahol $A_1,\dots,A_n$ együttesen független, p valószínűségű események. Így a fenti állítás miatt

$$\mathbb{D}^2(X)=\mathbb{D}^2(\mathbf{1}_{A_1}+\dots+\mathbf{1}_{A_n})=\mathbb{D}^2(\mathbf{1}_{A_1})+\dots+\mathbb{D}^2(\mathbf{1}_{A_n})=np(1-p).$$

(4) A Nagykörúton megfigyeljük a következőként érkező villamost, amely $\frac{1}{2}$ valószínűséggel 4-es és $\frac{1}{2}$ valószínűséggel 6-os. Jelölje X a következő villamos járatának számát (ami 4 vagy 6). Ebben az esetben X szórását a szokásos $\mathbb{E}(X^2)-\mathbb{E}(X^2)$ képlet helyett egyszerűbb az $\mathbb{E}((X-\mathbb{E}(X))^2)$ definíció alapján kiszámolni, ugyanis vegyük észre, hogy $\mathbb{E}(X)=\frac{1}{2}\cdot 4+\frac{1}{2}\cdot 6=5$, és így $(X-\mathbb{E}(X))^2$ értéke 1 valószínűséggel 1, ezért $\mathbb{D}^2(X)=\mathbb{D}(X)=1$.

3.5. Várható érték végtelen diszkrét esetben

Dobáljunk fel egy pénzérmét addig, amíg fejet nem kapunk. Legyen p annak a valószínűsége, hogy az érme a fej oldalát mutatja. Jelölje a dobások számát X. Mi X várható értéke?

Vegyük észre, hogy X nem egyszerű valószínűségi változó, hiszen k akármilyen pozitív egész értéket felvehet. Szerencsére várható értéket nem csak egyszerű valószínűségi változókra számolhatunk.

Intuitívan logikus lenne a 3.1 definíciót használni ebben az esetben is. Vegyük észre azonban, hogy ekkor a $\sum_{k\in\mathrm{Ran}(X)} k\cdot\mathbb{P}(X=k)$ kifejezés egy végtelen sor összegét tartalmazza, hiszen a $\mathrm{Ran}(X)=\{1,2,\dots\}$ értékkészlet a pozitív egész számok halmaza. Mivel azonban az értékkészlet kizárólag nemnegatív számokból áll, a sor összegét mindenképpen tudjuk értelmezni: az vagy egy nemnegatív valós szám, vagy (amennyiben a sor divergens) $+\infty$.

3.5.1. Definíció. Legyen X olyan nemnegatív valószínűségi változó, aminek értékkészlete $\mathrm{Ran}(X)=\{k_1,k_2,\dots\}$ megszámlálhatóan végtelen. Ekkor

$$\mathbb{E}(X)=\sum_{j=1}^{\infty} k_j\cdot\mathbb{P}(X=k_j). \tag{3.4}$$

A kezdeti példára visszatérve: ezzel az állítással már kiszámolható X várható értéke. Határozzuk meg a $\mathbb{P}(X = k)$ mennyiséget. Annak a valószínűsége, hogy éppen k dobásra lesz szükségünk: $(1-p)^{k-1}p$, hiszen $k-1$-szer kell írást dobnunk, majd egyszer fejet. Ezt már behelyettesíthetjük a szummába, és – ahogy azt látni fogjuk a 3.6. alfejezetben, – az eredmény $\frac{1}{p}$.

Mi a helyzet a nem feltétlenül nemnegatív értékkészletű diszkrét valószínűségi változók várható értékével? A (3.4) képlet abban az esetben alkalmazható, ha a sor abszolút konvergens (vagyis ha k_j helyett $|k_j|$-vel is egy véges értéket kapunk). És mit mondjunk a várható értékre, ha a sor nem abszolút konvergens? Már említettük, hogy nemnegatív értékkészlet esetén is előfordulhat, hogy az (3.4) képlet jobb oldalán szereplő sor összege $+\infty$, de ha a $+\infty$ „értéket" megengedjük, akkor legalább értelmezni tudjuk a várható értéket. Hasonlóan járhatunk el nempozitív értékkészletű valószínűségi változókra; itt a $-\infty$ „értéket" kell megengedni. Ha viszont a k_j-k tetszőlegesen nagy abszolút értékű negatív és pozitív számok is lehetnek, akkor lehetséges, hogy még $+\infty$ vagy $-\infty$ értékkel sem tudjuk definiálni a várható értéket.

A 3.2.3. Állításban láttuk, hogy két egyszerű valószínűségi változó összegének várható értéke a várható értékek összege. Ez motiválja azt az ötletet, hogy egy X valószínűségi változót felbontsuk két nemnegatív valószínűségi változó különbségére, majd ezek várható értékének különbségeként definiáljuk az $\mathbb{E}(X) \in [-\infty, \infty]$ várható értéket, kivéve az „$\infty - \infty$" alakú esetet, amikor X várható értékét nem értelmezzük. Erre szolgálnak a következő definíciók.

3.5.2. Definíció. Egy x valós szám **pozitív része**: $x^+ := \max(x, 0)$, x **negatív része** pedig: $x^- = \max(-x, 0)$.

3.5.3. Állítás. *Tetszőleges x valós számra x^+ és x^- nemnegatív számok, $x = x^+ - x^-$ és $|x| = x^+ + x^-$.*

(Ábrázoljuk a fél oldalt konstans, fél oldalt lineárisan növekvő/csökkenő $x \to x^+$ és $x \to x^-$ valós–valós függvényeket!)

A pozitív rész fogalmát nemcsak valós számokra, hanem valószínűségi változókra is értelmezhetjük (ω-nként értelmezve: $X^+(\omega) := (X(\omega))^+$, $X^-(\omega) := (X(\omega))^-$). Ezek segítségével terjeszthetjük ki a várható érték fogalmát a megszámlálhatóan végtelen értékkészletű, esetleg különböző előjelű értékeket is felvevő diszkrét valószínűségi változókra.

3.5.4. Definíció. Legyen X tetszőleges valószínűségi változó. Jelölje $X^+ = \max(X, 0)$ az X pozitív részét, és $X^- = \max(-X, 0)$ az X negatív részét. Ekkor X^+ és X^- nemnegatív valószínűségi változók, továbbá $X = X^+ - X^-$ és $|X| = X^+ + X^-$. Ha $\mathbb{E}(X^+) < \infty$ vagy $\mathbb{E}(X^-) < \infty$, akkor legyen

$$\mathbb{E}(X) \stackrel{\text{def}}{=} \mathbb{E}(X^+) - \mathbb{E}(X^-),$$

ami vagy egy valós szám, vagy $+\infty$, vagy $-\infty$. Ha $\mathbb{E}(X^+) = \mathbb{E}(X^-) = \infty$, akkor a várható értéket nem értelmezzük.[18]

3.6. Geometriai eloszlás, örökifjú tulajdonság I

3.6.1. Definíció. Egy X valószínűségi változó **geometriai eloszlású** p paraméterrel (ahol $p \in (0, 1)$), ha

$$\mathbb{P}(X = k) = (1-p)^{k-1}p$$

minden k pozitív egész esetén. Jelölés: $X \sim \text{Geo}(p)$.

3.6.2. Példa. Cérnát próbálunk befűzni egy tűbe. Tegyük fel, hogy minden egyes próbálkozásnál 0,1 valószínűséggel járunk sikerrel, a korábbi próbálkozásoktól függetlenül. Ekkor a szükséges próbálkozások száma 0,1 paraméterű geometriai eloszlást követ.

[18] Ha $\mathbb{E}(X^+) = \mathbb{E}(X^-) = \infty$, akkor azt szokás mondani, hogy a várható érték *nem létezik*. A hétköznapi szakmai szóhasználatban sokszor csak ebben az esetben mondják ezt; akkor nem feltétlenül mondják, ha a várható érték a fenti definíció szerint $+\infty$ vagy $-\infty$. Ezt az ellentmondást itt nem tudjuk feloldani, csak arra hívjuk fel a figyelmet, hogy a várható érték létezésének fogalmát érdemes mindig pontosítani.

Geometriai eloszlású valószínűségi változó várható értéke:[19]

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k \cdot \mathbb{P}(X = k) = \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} p =$$

$$= p \sum_{k=1}^{\infty} \sum_{i=1}^{k} (1-p)^{k-1} = p \sum_{i=1}^{\infty} \sum_{k=i}^{\infty} (1-p)^{k-1} = p \sum_{i=1}^{\infty} \frac{(1-p)^{i-1}}{1-(1-p)} = \sum_{j=0}^{\infty} (1-p)^{j} = \frac{1}{p}.$$

Geometriai eloszlás szórása: Legyen $X \sim \mathrm{Geo}(p)$. Ekkor

$$\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} p - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2},$$

ahol az első szumma kiszámolható például hatványsorok deriválásával, ettől itt eltekintünk. A 5.3.5. Példában más módszerrel ki fogjuk számítani $\mathbb{E}(X^2)$-et, addig az olvasó türelmét kérjük.

Egy eloszlás nem csak attól lesz nevezetes, hogy sok alkalmazás esetében felbukkan, hanem attól is, hogy speciális tulajdonságokkal bír. A fenti két eloszlás esetében ilyen tulajdonság az örökifjúság (angolul: memorylessness) is.

3.6.3. Definíció. Nevezzünk egy X valószínűségi változót **örökifjú**nak a $G \subseteq \mathbb{R}$ halmazon, ha tetszőleges $s, t \in G$ esetén

$$\mathbb{P}(X > t + s \mid X > s) = \mathbb{P}(X > t), \tag{3.5}$$

illetve $\mathbb{P}(X \in G) = 1$, azaz X értéke 1 valószínűséggel a G halmazba esik.

Ha X-re, mint egy történés bekövetkezésének időpontjára gondolunk, a fenti egyenlet azt jelenti, hogy ha s-ig nem következett be a történés, akkor ugyanakkora eséllyel kell még legalább t időt várnunk, mint ha most kezdenénk el várni. Vagyis az idő nem változtatott semmit a történés bekövetkezési hajlandóságán.

Kérdés: Van-e egyáltalán ilyen valószínűségi változó, és ha igen, mi lehet az eloszlása?

3.6.4. Állítás. *Ha egy X nemkonstans örökifjú valószínűségi változó a $G = \{1, 2, 3, \dots\}$ halmazon, akkor X eloszlása geometriai.*

Bizonyítás. Bontsuk ki a feltételes valószínűség definícióját. Azt kapjuk, hogy a (3.5) egyenlet ekvivalens a következővel:

$$\mathbb{P}(X > t + s) = \mathbb{P}(X > t)\mathbb{P}(X > s) \qquad (\forall s, t \in G).$$

Legyen először $G = \{1, 2, 3, \dots\}$, és jelölje p a $\mathbb{P}(X = 1)$ valószínűséget. Ekkor $s = 1$ esetén az előző egyenlet segítségével tetszőleges t pozitív egészre azt kapjuk, hogy

$$\mathbb{P}(X > t + 1) = \mathbb{P}(X > t)\mathbb{P}(X > 1) = \mathbb{P}(X > t) \cdot (1-p) =$$
$$= \mathbb{P}(X > t - 1) \cdot (1-p)^2 = \dots = (1-p)^{t+1}.$$

Ebből viszont már adódik a változó diszkrét eloszlása:

$$\mathbb{P}(X = t) = \mathbb{P}(X > t - 1) - \mathbb{P}(X > t) = (1-p)^{t-1} - (1-p)^t =$$
$$= \big(1 - (1-p)\big)(1-p)^{t-1} = p(1-p)^{t-1},$$

ami épp a geometriai eloszlás definíciója. Probléma csak azzal lehet, ha $p \in \{0, 1\}$. Ha $p = 0$, akkor $\mathbb{P}(X = t) = 0$ minden $t \in \{1, 2, \dots\}$ esetén, amely egyenletek így nem határozzák meg X eloszlását. Viszont az örökifjúság definíciójában feltettük, hogy X értéke 1 valószínűséggel G-be esik, ami ellentmondás. Hasonlóan, ha $p = 1$, akkor X a konstans 1 valószínűségi változó, de feltettük, hogy X nem konstans, ez ellentmondás. Következésképp X valóban geometriai eloszlású. □

[19] A $\sum_{k=1}^{\infty} k \cdot (1-p)^{k-1}$ sor értéke hatványsor deriválása segítségével is kiszámolható.

3.7. Poisson-eloszlás

A Poisson-eloszlás egy alkalmazásokban sűrűn előkerülő diszkrét eloszlás. Intuitívan, olyankor bukkan fel, ha rengeteg nagyon kis valószínűségű, egymástól független eseményből vizsgáljuk, hogy hány következik be[20]. Ilyen például egy telefonközpontba beérkező hívások száma, vagy az egy adott órában születő gyerekek száma.

3.7.1. Definíció. Az X valószínűségi változó **Poisson-eloszlás**ú λ paraméterrel ($\lambda > 0$), ha

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

minden nemnegatív egész k-ra. Jelölése: $X \sim \mathrm{Pois}(\lambda)$.

Ez tényleg egy lehetséges valószínűségi eloszlás, azaz a valószínűségek összege 1, hiszen

$$\sum_{k=0}^{\infty} \mathbb{P}(X = k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{\lambda - \lambda} = 1.$$

A Poisson-eloszlású valószínűségi változó paramétere szemléletes jelentésű, ugyanis

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} \lambda \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} e^{-\lambda} = \lambda,$$

vagyis a paraméter épp a változó várható értéke (vegyük észre, hogy ez például a geometriai eloszlásra nem teljesül).

Sőt a Poisson-eloszlású valószínűségi változó paramétere megegyezik a változó szórásnégyzetével is. Legyen ugyanis $X \sim \mathrm{Pois}(\lambda)$. Ekkor a (3.2). képlet szerint $g(k) = k^2 - k$ választással

$$\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}(X^2 - X) + \mathbb{E}(X) - \mathbb{E}(X)^2 = \sum_{k=0}^{\infty} (k^2 - k) \frac{\lambda^k}{k!} e^{-\lambda} + \lambda - \lambda^2 =$$

$$= \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} + \lambda - \lambda^2 = \lambda^2 \cdot 1 + \lambda - \lambda^2 = \lambda.$$

3.7.2. Példa. Egy kaszkadőr egy évben átlagosan 2-szer sérül meg. Mi a valószínűsége, hogy idén 4-szer? Mivel egy kaszkadőrnek rengetegszer van alkalma megsérülni, amiket tekinthetünk függetlennek, így a Poisson-eloszlás jó közelítése a sérülések számának, amit jelöljünk Y-nal. A feltétel szerint $\lambda = 2$, így

$$\mathbb{P}(Y = 4) = \frac{2^4}{4!} e^{-2} = \frac{2}{3} e^{-2} \approx 0{,}0902$$

az eredmény.

Azt, hogy rengeteg kis valószínűségű eseménynél a bekövetkezések száma Poisson-eloszlást közelít, be is bizonyíthatjuk. Ha n eseményről van szó, amik mind p valószínűségűek és egymástól függetlenek, akkor a bekövetkezések X száma $\mathrm{Bin}(n; p)$ binomiális eloszlású. A bekövetkezések átlagos száma, ahogy korábban már láttuk, $n \cdot p$, jelölje most ezt λ. A következő állítás épp azt mutatja, hogy nagy n esetén $\mathrm{Bin}(n; p)$ eloszlása körülbelül $\mathrm{Pois}(\lambda)$.

3.7.3. Állítás. *Legyen n pozitív egész, $\lambda \in (0, \infty)$, és jelölje $p_n = \frac{\lambda}{n}$-et. Ekkor*

$$\lim_{n \to \infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$$

tetszőleges $k \in \{0, 1, 2, \dots\}$ esetén.

[20] A hasonlóság az exponenciális eloszlás hasonló leírásával nem véletlen, lásd még Poisson-folyamat, avagy [4], 8.4. fejezet.

Bizonyítás. Fix k és n esetén

$$\binom{n}{k}p_n^k(1-p_n)^{n-k} = \frac{n!}{k!(n-k)!}\Big(\frac{\lambda}{n}\Big)^k\Big(1-\frac{\lambda}{n}\Big)^{n-k} =$$
$$= \frac{n!}{(n-k)!n^k}\cdot\frac{\lambda^k}{k!}\cdot\Big(1-\frac{\lambda}{n}\Big)^n\cdot\Big(1-\frac{\lambda}{n}\Big)^{-k}.$$

Ebből $\frac{\lambda^k}{k!}$ nem fog változni, ha n tart végtelenhez, míg

$$\Big(1-\frac{\lambda}{n}\Big)^n \to e^{-\lambda} \qquad \text{és} \qquad \Big(1-\frac{\lambda}{n}\Big)^{-k} \to 1.$$

Így már csak az első tényezővel kell megküzdeni. Az pedig a következőképp egyszerűsíthető:

$$\frac{n!}{(n-k)!n^k} = \frac{n(n-1)\dots(n-k+1)}{n\cdot n\cdot \ldots \cdot n},$$

ami egy k tényezőjű szorzat, ahol az egyes $\frac{n-i}{n}$ alakú tagok egyenként mind 1-hez tartanak. Mivel csak n tart végtelenhez, de k rögzített, így a szorzat határértéke szintén 1. □

3.7.4. Példa. Egy magyarérettségiben kétszer akkora eséllyel van összesen 3 elírás, mint 1 elírás (a példa nem reprezentatív). Tegyük fel, hogy a hibák egymástól függetlenül, azonos eséllyel következnek be. Mekkora a valószínűsége, hogy egyáltalán nincs elírás a dolgozatban?

Jelölje X egy dolgozatban a hibák számát. Mivel csak korlátos mennyiségű hiba szerepelhet, így a feltételeinkből (függetlenség, azonos esélyek) az következne, hogy X binomiális eloszlású. Ugyanakkor mind a hibák maximális számát, mind egy hiba bekövetkezésének valószínűségét nehéz meghatározni (főképp a rendelkezésre álló információból).

Felhasználhatjuk viszont a fenti állítást, ami szerint X eloszlását Poisson-eloszlással közelíthetjük, azaz

$$\mathbb{P}(X=k) = \frac{\lambda^k}{k!}e^{-\lambda} \qquad (k\in\mathbb{N}).$$

A feltétel szerint

$$2 = \frac{\mathbb{P}(X=3)}{\mathbb{P}(X=1)} = \frac{\lambda^3}{3!}e^{-\lambda}\Big/\frac{\lambda}{1!}e^{-\lambda} = \frac{\lambda^2}{6},$$

vagyis $\lambda = 2\sqrt{3}$. Ez azt jelenti, hogy átlagosan $2\sqrt{3}$ hiba bukkan fel az irományban. Innen a megoldás világos: $\mathbb{P}(X=0) = \frac{(2\sqrt{3})^0}{0!}e^{-2\sqrt{3}} = e^{-2\sqrt{3}} \approx 0{,}0313$.

4. Valószínűségi változók viszonya

Amikor valószínűségi változókról volt szó, mindig egymagukban vizsgáltuk az egyes példákat. Ilyenkor elégséges volt az eloszlásukkal foglalkozni, azaz diszkrét esetben a $\mathbb{P}(X = k)$ alakú valószínűségek sorozatát, vagyis X súlyfüggvényét. A súlyfüggvény minden lényeges tulajdonságot elmondott a valószínűségi változóról.

De ne keverjük össze a súlyfüggvényt magával a valószínűségi változóval: attól, hogy 100 dobásból mind a dobott fejek száma, mind a dobott írások száma $\mathrm{Bin}(100; \frac{1}{2})$ binomiális eloszlású, még nyilván nem mondhatjuk, hogy mindig ugyanannyi fejet dobnánk, mint írást. Ez a megkülönböztetés különösen lényeges, ha több valószínűségi változóról beszélünk egyszerre.

Ebben a fejezetben két valószínűségi változó függetlenségét, illetve lineáris összefüggőségük mértékét vizsgáljuk.

4.1. Függetlenség

Események függetlenségének a fogalmát már bevezettük a 2. előadáson: A és B események függetlenek, ha $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Definiáljuk most ezt felhasználva valószínűségi változók függetlenségét.

4.1.1. Definíció. Legyenek X és Y valószínűségi változók az $(\Omega, \mathcal{F}, \mathbb{P})$ valószínűségi mezőn. Azt mondjuk, hogy X és Y **függetlenek**, ha minden $x, y \in \mathbb{R}$ esetén az $\{X < x\}$ és $\{Y < y\}$ események függetlenek.

A valószínűségi mezőt azért kell emlegessük, mert előfordulhatna, hogy $X : \Omega_1 \to \mathbb{R}$, míg $Y : \Omega_2 \to \mathbb{R}$ alakú függvény, azaz más valószínűségi mezőn vannak értelmezve. Ilyen esetben nem tudunk X és Y függetlenségéről beszélni, mert „más világban élnek”.

4.1.2. Példa. Egy kockadobás eredménye és a ma leeső csapadék mennyisége, amik intuitívan függetlenek, a fenti definíció értelmében is függetlenek, ahogy ezt események függetlenségénél már megjegyeztük.

De a függetlenség nem minden esetben ilyen nyilvánvaló. Például legyen Z egy egyenletes eloszlású valószínűségi változó az $\{1, 2, 3, \ldots, 11, 12\}$ halmazon, jelölje X a Z hármas maradékát, és Y a Z négyes maradékát. Belátható, hogy ekkor X és Y függetlenek.

Hogyan tudjuk ellenőrizni két valószínűségi változó függetlenségét? Ebben az előadásban a diszkrét esetre koncentrálunk. Ekkor a következő állítás szolgáltat módszert a függetlenség ellenőrzésére.

4.1.3. Állítás. *Két diszkrét valószínűségi változó pontosan akkor független, ha minden $x, y \in \mathbb{R}$ esetén az $\{X = x\}$ és $\{Y = y\}$ események függetlenek, azaz*

$$\mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

Megjegyzés. A definícióból az is következik, hogy minden X-szel és Y-nal kifejezhető halmazpár független, például $\{X = x\}$ és $\{1 \leq Y \leq 5\}$ független események.

Fontos különbség az események függetlenségével szemben, hogy eseményekre ugyanannyi fáradtság volt leellenőrizni a függetlenséget és a nem-függetlenséget, hiszen mindkét esetben csak ki kellett számoljuk a metszet, illetve a két esemény külön-külön vett valószínűségét. Ezzel szemben valószínűségi változókra a függetlenséget cáfolni általában jóval egyszerűbb, mint igazolni: ha találunk egy $\{X = x\}$ és $\{Y = y\}$ eseményt, amelyek nem függetlenek, akkor a valószínűségi változók sem azok.

Miért tud hasznos lenni a függetlenség? Például, mert segíthet kiszámolni a várható értéket.

4.1.4. Állítás. *Ha X és Y független valószínűségi változók, és $\mathbb{E}(XY)$, $\mathbb{E}(X)$ és $\mathbb{E}(Y)$ létezik, akkor*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Bizonyítás. Csak arra az esetre bizonyítunk, amikor X és Y egyszerű valószínűségi változók. Az általános eset határátmenet segítségével igazolható, ettől itt eltekintünk.

Először legyen X és Y indikátor valószínűségi változó, azaz $X = \mathbf{1}_A$ és $Y = \mathbf{1}_B$ valamilyen A és B eseményekre. Ekkor

$$\mathbb{E}(\mathbf{1}_A\mathbf{1}_B) = \mathbb{E}(\mathbf{1}_{A\cap B}) = \mathbb{P}(A\cap B) = \mathbb{P}(A)\mathbb{P}(B) = \mathbb{E}(\mathbf{1}_A)\mathbb{E}(\mathbf{1}_B),$$

vagyis az állítás ebben a speciális esetben teljesül.

Nézzük az általánosabb esetet: tegyük fel, hogy X és Y egyszerű valószínűségi változó. Ekkor X és Y felírható indikátor valószínűségi változók lineáris kombinációjaként:

$$X = \sum_{k\in\mathrm{Ran}(X)} k\cdot\mathbf{1}_{\{X=k\}} \qquad \text{és} \qquad Y = \sum_{l\in\mathrm{Ran}(Y)} l\cdot\mathbf{1}_{\{Y=l\}}.$$

Az előző bekezdést és a várható érték additivitását felhasználva kapjuk, hogy

$$\mathbb{E}(XY) = \mathbb{E}\Big(\sum_{k\in\mathrm{Ran}(X)} k\cdot\mathbf{1}_{\{X=k\}}\sum_{l\in\mathrm{Ran}(Y)} l\cdot\mathbf{1}_{\{Y=l\}}\Big) = \sum_{k\in\mathrm{Ran}(X)}\sum_{l\in\mathrm{Ran}(Y)} k\cdot l\cdot\mathbb{E}\big(\mathbf{1}_{\{X=k\}}\mathbf{1}_{\{Y=l\}}\big) =$$

$$= \sum_{k\in\mathrm{Ran}(X)}\sum_{l\in\mathrm{Ran}(Y)} k\cdot l\cdot\mathbb{E}\big(\mathbf{1}_{\{X=k\}}\big)\mathbb{E}\big(\mathbf{1}_{\{Y=l\}}\big) = \mathbb{E}\Big(\sum_{k\in\mathrm{Ran}(X)} k\cdot\mathbf{1}_{\{X=k\}}\Big)\mathbb{E}\Big(\sum_{l\in\mathrm{Ran}(Y)} l\cdot\mathbf{1}_{\{Y=l\}}\Big),$$

ahol a jobb oldal éppen $\mathbb{E}(X)\mathbb{E}(Y)$, ahogy állítottuk. □

Megjegyzés. Felmerülhetne, hogy miért nem az állításban szereplő, kellemesebb egyenlettel definiáltuk valószínűségi változók függetlenségét. Azért, mert az $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ teljesülése gyengébb tulajdonság, nem következik belőle a valószínűségi változók függetlensége. Amit ehelyett felhasználhatnánk, az a következő állítás: ha minden olyan nemnegatív valós f és g függvények esetén $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$, amelyekre $f(X)$ és $g(Y)$ is valószínűségi változó, akkor X és Y függetlenek.[21]

4.2. Diszkrét együttes eloszlás

Diszkrét valószínűségi változók esetén a függetlenségük vizsgálatához a $\mathbb{P}(X = k,\ Y = l)$ (azaz a $\mathbb{P}(\{X = k\}\cap\{Y = l\})$) valószínűségekre, vagyis a változók úgynevezett együttes súlyfüggvényére van szükségünk.

4.2.1. Példa. Legyen X és Y olyan valószínűségi változók, ahol $\mathrm{Ran}(X) = \{2,3,5\}$, $\mathrm{Ran}(Y) = \{0,1,2\}$, és a $\mathbb{P}(X = k,\ Y = l)$ valószínűségeket a következő táblázat foglalja össze. Független-e X és Y, illetve mennyi $\mathbb{E}(XY)$?

Y \ X	2	3	5
0	0,05	0,15	0,1
1	0,1	0,2	0,1
2	0,05	0,2	0,05

Egy fentihez hasonló táblázat pontosan akkor írhatja le két valószínűségi változó együttes súlyfüggvényét, ha a benne szereplő számok nemnegatívak, és **összegük** 1. Leellenőrizhetjük, hogy ez a példában teljesül.

A függetlenség kiszámolásához szükségünk van a $\mathbb{P}(X = k)$ illetve a $\mathbb{P}(Y = l)$ mennyiségekre, vagyis az X és Y súlyfüggvényeire.

[21] Arra a kérdésre itt megfelelő mértékelméleti alapok hiányában nem tudunk teljes választ adni, hogy általában egy X valószínűségi változó és egy f nemnegatív valós függvény esetén $f(X)$ mikor lesz valószínűségi változó. Részleges válaszok: (1) Ha X *diszkrét* valószínűségi változó, akkor $f(X)$ bármely nemnegatív valós f függvény esetén valószínűségi változó. (2) Tetszőleges X esetén ha f *folytonos* nemnegatív valós függvény, akkor $f(X)$ is valószínűségi változó. Később, a 62. lábjegyzetben a folytonosságnál általánosabb feltételt is említünk, amelynek a teljesülése esetén ugyanez igaz.

4.2.2. Definíció. Legyenek X és Y egyszerű valószínűségi változók. Ha adott X és Y **együttes súlyfüggvény**e, vagyis a $\mathbb{P}(X = k,\ Y = l)$ valószínűség minden $k \in \mathrm{Ran}(X)$ és $l \in \mathrm{Ran}(Y)$ esetén, akkor X és Y súlyfüggvényeit az együttes súlyfüggvény **marginális súlyfüggvény**einek (avagy **perem-súlyfüggvényeinek**) nevezzük.

Vegyük észre, hogy az együttes súlyfüggvény egy *kétdimenziós* súlyfüggvény, amely minden (k, l) párhoz $\mathbb{P}(X = k,\ Y = l)$ értékét rendeli hozzá – ezért érdemes kétdimenziós táblázatban (mátrixban) ábrázolni.[22]

A marginális súlyfüggvényeket a valószínűség additivitása miatt a következőképp számolhatjuk ki:

$$\mathbb{P}(X = k) = \sum_{l \in \mathrm{Ran}(Y)} \mathbb{P}(X = k,\ Y = l) \qquad \mathbb{P}(Y = l) = \sum_{k \in \mathrm{Ran}(X)} \mathbb{P}(X = k,\ Y = l),$$

vagyis a táblázat sor- és oszlopösszegei adják az X és Y valószínűségi változók súlyfüggvényeit. A példa esetében így

$$\mathbb{P}(X = 2) = 0{,}2\ , \qquad \mathbb{P}(X = 3) = 0{,}55\ , \qquad \mathbb{P}(X = 5) = 0{,}25\ ,$$
$$\mathbb{P}(Y = 0) = 0{,}3\ , \qquad \mathbb{P}(Y = 1) = 0{,}4\ , \qquad \mathbb{P}(Y = 2) = 0{,}3.$$

Tehát a függetlenség definíciójából adódóan X és Y nem független, hiszen például $\mathbb{P}(X = 5,\ Y = 0) = 0{,}1$, de $\mathbb{P}(X = 5) \cdot \mathbb{P}(Y = 0) = 0{,}075$.

Számoljuk ki a fenti példában szereplő X és Y esetén az $\mathbb{E}(XY)$ mennyiséget is. Ehhez új definícióra nincs szükség, hiszen XY valószínűségi változó, értékkészlete $\{k \cdot l \mid k \in \mathrm{Ran}(X),\ l \in \mathrm{Ran}(Y)\}$. Így

$$\mathbb{E}(XY) = \sum_{m \in \mathrm{Ran}(XY)} m \cdot \mathbb{P}\Big(\bigcup_{\substack{k \in \mathrm{Ran}(X) \\ l \in \mathrm{Ran}(Y) \\ k \cdot l = m}} \{X = k,\ Y = l\} \Big) = \sum_{k \in \mathrm{Ran}(X)} \sum_{l \in \mathrm{Ran}(Y)} k \cdot l \cdot \mathbb{P}(X = k,\ Y = l) =$$

$$= 0 \cdot 0{,}05 + 0 \cdot 0{,}15 + 0 \cdot 0{,}1 + 2 \cdot 0{,}1 + 3 \cdot 0{,}2 + 5 \cdot 0{,}1 + 4 \cdot 0{,}05 + 6 \cdot 0{,}2 + 10 \cdot 0{,}05 = 3{,}2.$$

Vegyük észre, hogy ugyan a változók nem függetlenek, az $\mathbb{E}(XY)$ mennyiség így is kiszámolható.

4.3. Kovariancia

Ahogy az a példa esetében is látható, nem független valószínűségi változók esetében is lehet az összefüggésük mértéke alacsony (azaz intuitívan a $\mathbb{P}(X = k)\mathbb{P}(Y = l)$ szorzatok elég közel vannak az egyes $\mathbb{P}(X = k,\ Y = l)$ valószínűségekhez). Hogyan tudnánk mérni valószínűségi változók összefüggésének fokát? Erre több lehetőség is van,[23] kezdjük a kovariancia fogalmával.

4.3.1. Definíció. Az X és Y valószínűségi változók **kovarianciá**ját a következőképp definiáljuk:

$$\mathrm{cov}(X, Y) \overset{\mathrm{def}}{=} \mathbb{E}\big((X - \mathbb{E}X)(Y - \mathbb{E}Y)\big),$$

feltéve, hogy a várható érték létezik és véges[24].

[22] Az együttes súlyfüggvény táblázatára a szakirodalom egy része gyakran *együttes eloszlás* táblázatként hivatkozik, a marginális súlyfüggvényekre pedig *marginális eloszlásként* avagy *peremeloszlásként*; ennek a tárgynak a korábbi számonkéréseiben is találkozhatunk ezzel a szóhasználattal.
De mit nevezünk általában egy X valószínűségi változó *eloszlásá*nak? Azt a valószínűségi mértéket, amely a valós számok $\mathbb{R}$ halmazának „szép” részhalmazain van értelmezve és egy A részhalmazhoz a $\mathbb{P}(X \in A)$ valószínűséget rendeli, például egy $A = (a; b)$ intervallum esetén $\mathbb{P}(X \in A) = \mathbb{P}(a < X < b)$. Ha X diszkrét, akkor a $\mathbb{P}(X \in \{k\}) = \mathbb{P}(X = k)$, $k \in \mathrm{Ran}(X)$ alakú valószínűségek egyértelműen meghatározzák az eloszlást, ezért az eloszlás megadása lényegében ugyanaz a feladat, mint a súlyfüggvény megadása. Ennek köszönhető, hogy a két fogalmat sok helyen vegyesen használják. Azt itt mértékelméleti előismeretek hiányában nem részletezzük, hogy melyek azok a „szép” A halmazok, amelyekre $\mathbb{P}(X \in A)$ értelmes. Hasonló a helyzet két diszkrét valószínűségi változó, X és Y együttes eloszlásánál, amely egy $A \subseteq \mathbb{R}^2$ halmazhoz a $\mathbb{P}((X, Y) \in A)$ valószínűséget rendeli.

[23] lásd még például: mediántól vett átlagos abszolút eltérés (mean absolute error); távolság-kovariancia.

[24] A 4.4. alfejezetben látni fogjuk, hogy a kovariancia bizonyos praktikus tulajdonságai csak akkor teljesülnek, ha X és Y szórása is véges, vagyis ha $\mathbb{E}(X^2)$ és $\mathbb{E}(Y^2)$ is véges. Ott ki fog derülni az is, hogy ebben az esetben $\mathrm{cov}(X, Y)$ valóban létezik és véges. Egyes források csak ebben az esetben definiálják a kovarianciát.

4.3.2. Állítás. *Ha* $\operatorname{cov}(X,Y)$ *értelmes, akkor* $\operatorname{cov}(X,Y)=\mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)$.

Bizonyítás. A definíciót kibontva kapjuk, hogy

$$\mathbb{E}\big((X-\mathbb{E}X)(Y-\mathbb{E}Y)\big)=\mathbb{E}(XY)-\mathbb{E}\Big(\mathbb{E}(X)Y\Big)-\mathbb{E}\Big(X\mathbb{E}(Y)\Big)+\mathbb{E}\Big(\mathbb{E}(X)\mathbb{E}(Y)\Big)=$$

$$=\mathbb{E}(XY)+(-1-1+1)\mathbb{E}(X)\mathbb{E}(Y)=\mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y),$$

ami épp a belátandó állítás. □

4.3.3. Következmény. *Legyen* X *és* Y *valószínűségi változó, amire* $\operatorname{cov}(X,Y)$ *értelmes.*

(1) Ha Y *konstans, akkor* $\operatorname{cov}(X,Y)=0$.
(2) Ha X *és* Y *függetlenek, akkor* $\operatorname{cov}(X,Y)=0$.
(3) Attól, hogy $\operatorname{cov}(X,Y)=0$*, még nem feltétlenül teljesül, hogy* X *és* Y *független.*
(4) $\operatorname{cov}(X,X)=\mathbb{D}^2(X)$.
(5) A kovariancia bilinearitása: *Ha* $b,c\in\mathbb{R}$ *akkor* $\operatorname{cov}(X,bY+cZ)=b\cdot\operatorname{cov}(X,Y)+c\cdot\operatorname{cov}(X,Z)$ *és hasonlóan* $\operatorname{cov}(bY+cZ,X)=b\cdot\operatorname{cov}(Y,X)+c\cdot\operatorname{cov}(Z,X)$.

Az utolsó előtti tulajdonság miatt nevezik a szórásnégyzetet varianciának is (hiszen ez egy valószínűségi változó önmagával vett kovarianciája). Az utolsó tulajdonságot azért nevezzük bilinearitásnak, mert a kovariancia mindkét változóban lineáris (összeadás esetén mindkét változó szerint tagokra szétbontható, skalár szorzó mindkét változóból kiemelhető).

Bizonyítás. Jelölje Y konstans értékét $c\in\mathbb{R}$. Az előző állítás szerint

$$\operatorname{cov}(X,Y)=\mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)=\mathbb{E}(Xc)-\mathbb{E}(X)c=0.$$

A második ponthoz felhasználhatjuk az előző alfejezet második állítását, így $\operatorname{cov}(X,Y)=\mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)=0$.

A harmadik állításhoz legyen $\operatorname{Ran}(X)=\{-1,0,1\}$, amely értékeket $\frac14$, $\frac12$ és $\frac14$ valószínűségekkel veszi fel X. Legyen $Y=|X|$. Kiszámolható, hogy $\operatorname{cov}(X,Y)=\mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)=0-0\cdot\frac12=0$, pedig a változók nem függetlenek, hiszen $\mathbb{P}(X=0)\mathbb{P}(Y=1)=\frac12\cdot\frac12$, míg $\mathbb{P}(X=0,Y=1)=0$.

A negyedik állítás a definíciókból következik: $\operatorname{cov}(X,X)=\mathbb{E}((X-\mathbb{E}(X))(X-\mathbb{E}(X))^2)$, ami éppen $\mathbb{D}^2(X)$ definíciója volt.

Az ötödik állításnak a szimmetria miatt elég az első felét bizonyítani, az pedig a várható érték linearitásából könnyen adódik:

$$\operatorname{cov}(X,bY+cZ)=\mathbb{E}(X(bY+cZ))-\mathbb{E}(X)\mathbb{E}(bY+cZ)=\mathbb{E}(bXY+cXZ)-b\mathbb{E}(X)\mathbb{E}(Y)-c\mathbb{E}(X)\mathbb{E}(Z)$$
$$=b\mathbb{E}(XY)-b\mathbb{E}(X)\mathbb{E}(Y)+c\mathbb{E}(XZ)-c\mathbb{E}(X)\mathbb{E}(Z)=b\cdot\operatorname{cov}(X,Y)+c\cdot\operatorname{cov}(X,Z).$$

□

4.3.4. *Megjegyzés.* A fenti (1) és (5) tulajdonságot összevonva úgy is megfogalmazhatjuk, hogy ha a,b,c,d valós számok és X,Y valószínűségi változók, akkor

$$\operatorname{cov}(aX+b,cY+d)=a\cdot c\cdot\operatorname{cov}(X,Y).$$

Azaz „konstans szorzó mindkét változóból kiemelhető” és „additív konstans mindkét változóban elhagyható”. Emlékeztetőül: az utóbbi tulajdonságot a szórás(négyzet) esetén *eltolás-invarianciá*nak neveztük, és ugyanúgy hívjuk a kovariancia esetén is.

4.3.5. Példa.

(1) Már láttuk, hogy $\mathbb{E}(XY)=3{,}2$. Kiszámolható, hogy $\mathbb{E}(X)=3{,}8$ és $\mathbb{E}(Y)=1$, így $\operatorname{cov}(X,Y)=\mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)=3{,}2-3{,}8\cdot1=-0{,}6$.

(2) Legyen X egyenletes eloszlású az $\{1, 2, \dots, 10\}$ halmazon, illetve Y egyenletes eloszlású az $\{1, -1\}$ halmazon. Tegyük fel, hogy X és Y függetlenek. Ekkor

$$\begin{aligned}\operatorname{cov}(X,\, 0{,}9 \cdot X + 0{,}1 \cdot Y) &= \mathbb{E}(0{,}9 \cdot X^2 + 0{,}1 \cdot XY) - \mathbb{E}(X)\mathbb{E}(0{,}9 \cdot X + 0{,}1 \cdot Y) = \\ &= 0{,}9 \cdot \mathbb{E}(X^2) + 0{,}1 \cdot \mathbb{E}(XY) - 0{,}9 \cdot \mathbb{E}(X)^2 - 0{,}1 \cdot \mathbb{E}(XY) \\ &= 0{,}9 \sum_{k=1}^{10} k^2 \frac{1}{10} - 0{,}9 \cdot \Big(\frac{11}{2}\Big)^2 \approx 7{,}425.\end{aligned}$$

A példából is látható, hogy a várható érték additivitása könnyíthet a kovariancia kiszámolásán.

Megjegyzés. Adódik a kérdés, hogy ha a kovariancia nulla volta nem is karakterizálja a függetlenséget, akkor miért ezt a definíciót nézzük? Ennek a fő oka, hogy a kovariancia szimmetrikus és bilineáris, azaz

$$\operatorname{cov}(X, Y) = \operatorname{cov}(Y, X) \qquad \text{és} \qquad \operatorname{cov}(X, aY + bZ) = a \cdot \operatorname{cov}(X, Y) + b \cdot \operatorname{cov}(X, Z) \qquad (a, b \in \mathbb{R}),$$

ha a fenti kovarianciák léteznek. Így a kovariancia a vektorok skaláris szorzatának rokonfogalma. A vektoros analógiát felhasználva, ha a kovarianciát a vektorok skaláris szorzatával állítjuk párhuzamba, akkor a szórásnégyzet a vektor hossznégyzetének, míg a szórás a vektor hosszának feleltethető meg. Lényeges különben a következő. hogy míg a vektorok körében csak a nullvektor hossza nulla, addig a valószínűségi változók körében nemcsak a konstans nullának, hanem az összes konstansnak nulla az önmagával vett kovarianciája. Más szavakkal: a kovarianciával (illetve a szórásnégyzettel és szórással) ellentétben a skaláris szorzás nem eltolás-invariáns.

A szórásnégyzet népszerűségének egy további oka az alábbi állítás.

4.3.6. Állítás. *Ha X és Y független, akkor $\mathbb{D}^2(X + Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y)$.*

Bizonyítás. A szórásnégyzet definícióját kibontva:

$$\begin{aligned}\mathbb{D}^2(X + Y) &= \mathbb{E}\big((X + Y)^2\big) - \big(\mathbb{E}(X + Y)\big)^2 = \\ &= \mathbb{E}(X^2) + \mathbb{E}(Y^2) + 2\mathbb{E}(XY) - \mathbb{E}(X)^2 - \mathbb{E}(Y)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) \\ &= \mathbb{D}^2(X) + \mathbb{D}^2(Y) + 2\operatorname{cov}(X, Y).\end{aligned}$$

Mivel X és Y függetlenek, így $\operatorname{cov}(X, Y) = 0$, amiből az állítás már következik. □

4.3.7. *Megjegyzés.* A bizonyításból látható, hogy a fenti állítást függetlenség nélkül is kimondhattuk volna, csak úgy valamivel bonyolultabb eredményt kapunk:

$$\mathbb{D}^2(X + Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y) + 2\operatorname{cov}(X, Y).$$

4.4. Korreláció

Fentebb utaltunk rá, hogy a kovariancia segíthet mérni valószínűségi változók összefüggését. De hogyan kell ezt érteni, ha a kovariancia épp nem 0? Például az első példából adódó $\operatorname{cov}(X, Y) = -0{,}6$ értéknek mi a jelentése, mennyire függnek ettől össze az X és Y változók?

Ezt pusztán a kovariancia alapján nem tudjuk megválaszolni, ahhoz egy ebből származtatott mennyiség lesz a segítségünkre.

4.4.1. Definíció. Legyenek X és Y valószínűségi változók, amelyek értéke nem 1 valószínűséggel konstans. Ha $\operatorname{cov}(X, Y)$, $\mathbb{D}(X)$ és $\mathbb{D}(Y)$ értelmes, akkor X és Y **korreláció**ja:

$$\operatorname{corr}(X, Y) \overset{\text{def}}{=} \frac{\operatorname{cov}(X, Y)}{\mathbb{D}(X)\mathbb{D}(Y)}.$$

Belátható, hogy $-1 \le \operatorname{corr}(X, Y) \le 1$ mindig teljesül. A szélsőséges esetekben X és Y közt tökéletes lineáris összefüggés áll fent, azaz teljesül a következő állítás.

4.4.2. Állítás. *Legyenek X és Y valószínűségi változók. Ha* $\operatorname{corr}(X,Y) \in \{1,-1\}$*, akkor az* $Y = aX + b$ *állítás* 1 *valószínűséggel teljesül valamilyen a és b valós számokra, ahol az a előjele megegyezik* $\operatorname{corr}(X,Y)$ *előjelével.*

A vektoros analógiában (ahol a kovariancia két vektor skaláris szorzatának, a szórás pedig a vektor hosszának felel meg) a korreláció két vektor által bezárt szög koszinuszának felel meg. Itt a $+1$ értékű korreláció annak felel meg, hogy a vektorok párhuzamosak és egyállásúak (bezárt szögük $0°$, így annak koszinusza $+1$), a -1 értékű korreláció annak, hogy a vektorok párhuzamosak és ellentétes állásúak (bezárt szögük $180°$, így annak koszinusza -1), a 0 korreláció pedig annak, hogy a vektorok merőlegesek (bezárt szögük $90°$, így annak koszinusza 0). Ahogy a nullvektornak sem értelmezzük más vektorral bezárt szögét, úgy semmilyen 1 valószínűséggel konstans valószínűségi változónak nem értelmezzük más valószínűségi változóval vett korrelációját.

4.4.3. Példa. A diszkrét együttes eloszlás részben vizsgált példa esetében

$$\operatorname{corr}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\mathbb{D}(X)\mathbb{D}(Y)} = \frac{-0{,}6}{\sqrt{9{,}2}\cdot\sqrt{0{,}6}} \approx -0{,}2554.$$

Ennek a szemléletes jelentése az, hogy X és Y szívesebben tér el ellentétes irányba az átlagától, mint azonos irányba, de a köztük lévő lineáris összefüggés relatíve alacsony (legalábbis amennyire a 0,25 az 1-hez képest alacsony).

Ahogy kovariancia esetében is, a korreláció nulla mivolta nem jelenti, hogy a két valószínűségi változó független volna. Valójában a korreláció a két változó közti lineáris összefüggés fokát méri. Más szavakkal, hiába függ össze két valószínűségi változó, ha az összefüggésük nemlineáris, azt a korreláció nem fogja észrevenni. Például megadható olyan X valószínűségi változó, amire $\operatorname{corr}(X,X^2) = 0$.

Egy nevezetes többdimenziós eloszlás például a *polinomiális*, avagy *multinomiális* eloszlás, amelynek egydimenziós peremeloszlásai binomiálisak. Az ez iránt érdeklődőknek ajánljuk a függelékben szereplő A.8. fejezetet.

5. Feltételes várható érték

Korábban már definiáltuk egy B eseménynek egy (nem 0 valószínűségű) A eseményre vett $\mathbb{P}(B|A) = \frac{\mathbb{P}(B\cap A)}{\mathbb{P}(A)}$ feltételes valószínűségét. Ha X egy valószínűségi változó, akkor a B eseményt megválaszthatjuk például $\{X = x\}$-nek vagy $\{X < x\}$-nek is, ezzel $\mathbb{P}(X = x \mid A)$ vagy $\mathbb{P}(X < x \mid A)$ alakú feltételes valószínűségek is meghatározhatóak.

Ebben a fejezetben azzal foglalkozunk, hogy ez alapján hogyan lehet egy Y *diszkrét* valószínűségi változó *feltételes várható értékét* definiálni. Itt először egy Y valószínűségi változónak egy A eseményre vett feltételes várható értékével fogunk foglalkozni, amit $\mathbb{E}(Y|A)$-val fogunk jelölni. Ebben az esetben a feltételes valószínűség és a várható érték definíciója alapján $\mathbb{E}(Y|A)$ feltételes várható érték definíciója is intuitív lesz, és egy valós számot ad eredményül.

Itt az A esemény is lehet például $\{X = x\}$ alakú, ahol X egy másik diszkrét valószínűségi változó. $\mathbb{E}(Y|A)$ definíciója alapján megtanuljuk azt is, hogy hogyan lehet egy Y diszkrét valószínűségi változó feltételes várható értékét egy X *diszkrét valószínűségi változóra* feltételezve definiálni és példákban meghatározni. Az utóbbi esetben az $\mathbb{E}(Y|X)$-szel jelölt feltételes várható érték már nem egy valós szám, hanem egy *valószínűségi változó* lesz, mégpedig $r(X)$ alakú, azaz X-nek valamilyen transzformáltja. Intuitívan ez az $r(X)$ valószínűségi változó úgy interpretálható, mint „Y legjobb közelítése X ismeretében".

5.1. Feltételes várható érték eseményre

Idézzük fel a feltételes valószínűség definícióját: ha A olyan esemény, amire $\mathbb{P}(A) > 0$, akkor

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$$

a B-nek az A-ra vett feltételes valószínűsége. Szemléletesen, ez a B esemény valószínűsége arra az esetre fókuszálva, amikor A bekövetkezik. A második fejezetben beláttuk, hogy $B \mapsto \mathbb{P}(B \mid A)$ is valószínűségi mérték. Vagyis bármi, amit valószínűségi mértékekre bizonyítottunk, rá is alkalmazható.

5.1.1. Definíció. Legyen Y valószínűségi változó és A olyan esemény, amire $\mathbb{P}(A) > 0$. Ekkor az Y-nak az A-ra vett **feltételes várható értéke** az Y változó $\mathbb{P}(. \mid A)$ valószínűségi mérték szerinti várható értéke. Jelölés: $\mathbb{E}(Y \mid A)$.

A feltételes valószínűséghez hasonlóan, $\mathbb{E}(Y \mid A)$ jelentése az Y átlagos értéke a teljes eseménytér helyett az A eseményre szorítkozva.

5.1.2. Lemma. *Legyen Y egyszerű valószínűségi változó, és A esemény, amire $\mathbb{P}(A) > 0$. Ekkor*

$$\mathbb{E}(Y \mid A) = \sum_{k \in \mathrm{Ran}(Y)} k \cdot \mathbb{P}(Y = k \mid A). \tag{5.1}$$

Vagyis elég kicserélnünk a várható érték definíciójában a $\mathbb{P}$-t $\mathbb{P}(. \mid A)$-ra. Általánosabban, az alábbi módon tudjuk a feltételes várható értéket visszavezetni várható érték számolására.

5.1.3. Állítás. *Legyen Y valószínűségi változó, amire $\mathbb{E}(|Y|) < \infty$, és A esemény, amire $\mathbb{P}(A) > 0$. Ekkor $\mathbb{E}(Y \mid A) = \frac{1}{\mathbb{P}(A)}\mathbb{E}(Y\mathbf{1}_A)$.*[25]

Mi ennek a fogalomnak az értelme? Több esetben a feltételes várható érték nem kiszámolandó cél, hanem a feladat megfogalmazásának eszköze, ahogy ezt a feltételes valószínűség esetén is láttuk. Tegyük fel például, hogy egy eszközből kétféle márka is elérhető, az egyik átlagosan 3 évig, míg a másik 4 évig nem romlik el. Válasszunk a kettő közül véletlenszerűen egyet, és jelölje Y az élettartamát. Ekkor a feladatban lévő információink $\mathbb{E}(Y \mid \{\text{elsőt választjuk}\}) = 3$ és $\mathbb{E}(Y \mid \{\text{másodikat választjuk}\}) = 4$.

Feltétel eseményként használhatjuk egy másik valószínűségi változó $\{X < x\}$ nívóhalmazát is, azaz nézhetjük $\mathbb{E}(Y \mid X < x)$-et valamilyen X valószínűségi változóra.

[25] Az állítás következménye, hogy ha $\mathbb{E}(Y)$ létezik és véges, akkor emiatt $\mathbb{E}(Y \mid A)$ is létezik és véges.

5.1.4. Példa. Legyen Y és Z egy-egy szabályos kockadobás eredménye, $X = Y + Z$ és $x = 7$. Ekkor

$$\mathbb{P}(Y = k \mid X < 7) = \frac{\mathbb{P}(Y = k, X < 7)}{\mathbb{P}(X < 7)} = \frac{6-k}{36} \Big/ \frac{15}{36} = \frac{6-k}{15},$$

$$\mathbb{E}(Y \mid X < 7) = \sum_{k=1}^{6} k \cdot \mathbb{P}(Y = k \mid X < 7) = \sum_{k=1}^{6} k \cdot \frac{6-k}{15} = \frac{6 \cdot 21 - 91}{15} = \frac{7}{3}.$$

Továbbá, ha $\mathbb{P}(X = x) > 0$, akkor $\mathbb{E}(Y \mid X = x)$ is értelmes. Az előző példánál maradva

$$\mathbb{E}(Y \mid X = 5) = \sum_{k=1}^{6} k \cdot \mathbb{P}(Y = k \mid X = 5) = \sum_{k=1}^{4} k \cdot \frac{1}{4} = 2{,}5.$$

Szemléletesen, $\mathbb{E}(Y \mid X = 5)$ az Y átlagos értéke abban az esetben, ha tudjuk, hogy X értéke 5.

5.2. Feltételes várható érték diszkrét valószínűségi változóra

Az $\mathbb{E}(Y \mid X = x)$ mennyiségre úgy is tekinthetünk, mint egy függvényre az x változóban. Ez egy determinisztikus függvény, azaz nem valószínűségi változó, hiszen tetszőleges x valós szám esetén $\mathbb{E}(Y \mid X = x)$ valós szám (feltéve, hogy értelmes és véges).

5.2.1. Definíció. Legyenek X és Y valószínűségi változók, ahol X diszkrét és $\mathbb{E}(|Y|) < \infty$. Jelölje S_X az X lényeges értékeinek a halmazát, azaz

$$S_X \stackrel{\text{def}}{=} \{x \in \mathbb{R} \mid \mathbb{P}(X = x) > 0\}.$$

Tekintsük az

$$r \colon S_X \to \mathbb{R}, \qquad r(x) = \mathbb{E}(Y \mid X = x)$$

valós függvényt. Ekkor az Y-nak az X-re vett **feltételes várható érték**e az $r(X)$ valószínűségi változó. Jelölése: $\mathbb{E}(Y \mid X)$. Az r függvény elnevezése **regressziós függvény**.

5.2.2. Példa. Dobunk egy szabályos kockával, majd az eredménynek megfelelő számú szabályos érmével. Legyen X a kockadobás értéke, Y pedig a fejek száma az érmedobások között. Mi Y feltételes várható értéke X-re? Ha ismerjük X értékét, akkor Y binomiális eloszlású, paraméterei: X értéke és $\frac{1}{2}$. Binomiális eloszlású változónak ismerjük a várható értékét (a paraméterek szorzata), ezért $\mathbb{E}(Y \mid X = x) = x\frac{1}{2}$, ahol $x \in \{1, 2, 3, 4, 5, 6\}$.

Hogy mit értünk „legjobb közelítés" alatt? Intuitívan fogalmazva az $\mathbb{E}(Y \mid X)$ valószínűségi változó mindent tud az Y-ról, amit X ismeretében tudni lehet, és ezzel a „lehető legtöbb" információval ad közelítést Y-ra.[26] Ezt precízen úgy tudjuk átfogalmazni, hogy ha további, X-re vonatkozó feltétel esetén nézzük Y feltételes várható értékét, akkor az már kiszámolható az $\mathbb{E}(Y \mid X)$ feltételes várható értékből is, nem kell hozzá az eredeti Y-t ismernünk. Még precízebben:

5.2.3. Állítás. *Legyen X diszkrét valószínűségi változó. Tegyük fel, hogy $\mathbb{P}(X < x) > 0$ és $\mathbb{E}(|Y|)$ véges. Ekkor*

$$\mathbb{E}\big(r(X) \mid X < x\big) = \mathbb{E}(Y \mid X < x),$$

ahol $r(X) = \mathbb{E}(Y \mid X)$.

Bizonyítás. A fenti (5.1) egyenletből, és a feltételes valószínűség definíciójából adódóan

$$\mathbb{E}\big(r(X) \mid X < x\big) \stackrel{(5.1)}{=} \sum_{k \in \mathrm{Ran}(X)} r(k) \cdot \mathbb{P}(X = k \mid X < x) =$$

$$= \sum_{k<x} r(k) \frac{\mathbb{P}(X = k)}{\mathbb{P}(X < x)} = \sum_{k<x} \mathbb{E}(Y \mid X = k) \frac{\mathbb{P}(X = k)}{\mathbb{P}(X < x)}.$$

[26] A „legjobb közelítés" másik megfogalmazását lásd a harmadik alfejezet megfelelő állításában.

Itt felhasználhatjuk a fenti állítást és a várható érték additivitását:

$$= \sum_{k<x} \frac{1}{\mathbb{P}(X=k)} \mathbb{E}(Y \mathbf{1}_{\{X=k\}}) \frac{\mathbb{P}(X=k)}{\mathbb{P}(X<x)} = \frac{1}{\mathbb{P}(X<x)} \mathbb{E}\Big(Y \sum_{k<x} \mathbf{1}_{\{X=k\}}\Big)$$

$$= \frac{1}{\mathbb{P}(X<x)} \mathbb{E}\big(Y \mathbf{1}_{\{X<x\}}\big) = \mathbb{E}(Y \mid X<x).$$

Ez éppen a belátandó állítás. □

5.3. Feltételes várható érték tulajdonságai, teljes várható érték tétele

A teljes várható érték könnyebb meghatározásához érdemes megvizsgálnunk a tulajdonságait, ahogy azt a korábbi fogalmak esetében is tettük.

5.3.1. Állítás. *Legyenek X, Y, Z valószínűségi változók, amelyekre $\mathbb{E}(|Y|) < \infty$ és $\mathbb{E}(|Z|) < \infty$. Ekkor teljesülnek a következők:*

(1) Tetszőleges $a, b \in \mathbb{R}$ esetén $\mathbb{E}(aY + bZ \mid X) = a\mathbb{E}(Y \mid X) + b\mathbb{E}(Z \mid X)$.

(2) Tetszőleges olyan h folytonos[27] függvény esetén, amelyre $\mathbb{E}(|h(X)|) < \infty$ teljesül, igaz, hogy

$$\mathbb{E}\big(h(X)Y \mid X\big) = h(X)\mathbb{E}(Y \mid X).$$

(3) Ha X és Y függetlenek, akkor $\mathbb{E}(Y \mid X) = \mathbb{E}(Y)$.

A linearitás nem meglepő. A második tulajdonság szemléletesen azt állítja, hogy mivel $h(X)$ értéke meghatározható X-ből, ezért ha a $h(X)Y$-ra keressük a legjobb becslést X függvényében, akkor elég az Y becslését megoldanunk, a $h(X)$ szorzóként kiemelhető a várható értékből. A harmadik tulajdonság jelentése, hogy ha X és Y függetlenek, akkor X-ből nem tudunk jobb becslést faragni Y-ra, mint a legjobb konstans becslés, ami a várható értéke.

A feltételes várható érték felírható optimalizálási probléma megoldásaként.[28]

5.3.2. Állítás. *Tegyük fel, hogy $\mathbb{E}(Y^2)$ véges. Ekkor a*

$$\mathbb{E}\Big(\big(Y - g(X)\big)^2\Big)$$

várható érték pontosan akkor minimális a g függvényben, ha $g(X)$ és $r(X) = \mathbb{E}(Y \mid X)$ 1-valószínűséggel megegyeznek.[29]

Hogyan tudjuk hasznosítani a feltételes várható értéket olyan problémában, ahol a probléma kérdésfelvetésében nem szerepel? Úgy, hogy a szokásos várható érték is számolható feltételes várható érték segítségével ugyanúgy, ahogy a második előadáson a szokásos valószínűséget is számoltunk feltételes valószínűségekkel (lásd pl. teljes valószínűség tétele). Ilyen módszer a teljes várható érték tétele is.

5.3.3. Állítás (Teljes várható érték tétele)**.** *Legyen X és Y valószínűségi változó, amikre $\mathbb{E}(X)$ és $\mathbb{E}(Y)$ véges. Ekkor $\mathbb{E}\big(\mathbb{E}(Y \mid X)\big) = \mathbb{E}(Y)$.*

---

[27] A folytonosság itt elégséges feltétel, de valójában csak arra a gyengébb feltételre van szükségünk, hogy $h(X)$ is valószínűségi változó legyen.

[28] Hogy nem egy minimalizálási problémával definiáltuk a feltételes várható értéket, annak az az oka, hogy a karakterizáció feltételezi, hogy $\mathbb{E}(Y^2)$ véges, míg a feltételes várható érték létezéséhez elég, ha $\mathbb{E}(Y)$ véges.

[29] Erre az optimális tulajdonságra utal a *regressziós függvény* elnevezésben a *regresszió* kifejezés. Ez a kifejezés általában a statisztikában valamilyen adathalmaz pontjainak egy másik adathalmaz pontjaival való optimális közelítését jelenti. Esetünkben Y-t próbáljuk X ismeretében X valamilyen $g(X)$ függvényével a lehető legjobban „magyarázni", abbaz a értelemben, hogy $g(X)$-nek az Y-tól való átlagos négyzetes eltérést minimalizáljuk, és ennek a megoldása $g(X) = r(x) = \mathbb{E}(Y|X)$. Hasonló problémát ír le az úgynevezett *lineáris regresszió*, azzal az extra feltétellel, hogy ott g-nek lineáris függvénynek kell lennie (míg r nem feltétlenül lineáris függvény). A részletekért lásd a függelék A.5. fejezetét.

Hát ez mi? Ez nem is úgy néz ki, mint ahogy a teljes valószínűség tétele kinézett. Hol a szumma? Minek két várható érték jel egymás után? Nos, a szumma a külső várható értékben van elrejtve. A várható értékre pedig azért van szükség, mert – ahogy feljebb megvizsgáltuk – $\mathbb{E}(Y \mid X)$ egy valószínűségi változó, így ha számot szeretnénk kapni belőle, ehhez vehetjük a várható értékét.

További érv amellett, hogy a fenti állítás elnevezése találó, hogy ha a feltételes várható értéket kiírjuk a diszkrét esetre, akkor pont olyan formulát kapunk, mint a teljes valószínűség tételében.

5.3.4. Állítás (Teljes várható érték tétele, diszkrét eset)**.** *Legyen X diszkrét valószínűségi változó, és $\{x_1, \dots, x_n, \dots\}$ az X értékkészletének azon pontjai, amire $\mathbb{P}(X = x_i) > 0$. Ha $\mathbb{E}(Y) < \infty$, akkor*

$$\mathbb{E}(Y) = \sum_{i=1}^{\infty} \mathbb{E}(Y \mid X = x_i)\mathbb{P}(X = x_i).$$

5.3.5. Példa. Számoljuk ki a geometriai eloszlás szórását. (Oké, már kiszámoltuk máshogy, de ez rövidebb.) Dobáljunk fel egy cinkelt pénzérmét, amíg fejet nem kapunk, ahol a fej esélye p. Jelölje a szükséges dobások számát Y, és legyen $X = 1$, ha az első dobás fej, különben 0. Ekkor

$$\mathbb{E}(Y^2) = \mathbb{E}(Y^2 \mid X = 1)\mathbb{P}(X = 1) + \mathbb{E}(Y^2 \mid X = 0)\mathbb{P}(X = 0).$$

Ha az első dobás írás, akkor összesen eggyel többet kell majd várnunk, mintha most kezdenénk a dobálást, a geometriai eloszlás örökifjú tulajdonsága miatt. Egyenlettel $\mathbb{P}(Y = k \mid X = 0) = \mathbb{P}(Y + 1 = k)$ minden k pozitív egészre. Vagyis $\mathbb{E}(Y^2 \mid X = 0) = \mathbb{E}\big((Y+1)^2\big)$, hiszen Y-nak a $\mathbb{P}(. \mid X = 0)$ valószínűségi mérték szerinti eloszlása megegyezik az Y szokásos értelemben vett eloszlásával. Következésképp,

$$\begin{aligned}\mathbb{E}(Y^2) &= 1 \cdot \mathbb{P}(X = 1) + \mathbb{E}\big((Y+1)^2\big)\mathbb{P}(X = 0) = \\ &= p + \mathbb{E}(Y^2 + 2Y + 1)(1 - p),\end{aligned}$$

amit átrendezve $\mathbb{E}(Y^2) = \frac{2-p}{p^2}$ adódik, felhasználva hogy $\mathbb{E}(Y) = \frac{1}{p}$. Így $\mathbb{D}^2(Y) = \frac{1-p}{p^2}$.

A tétel teljes eseményrendszerre is kimondható.

5.3.6. Állítás (Teljes várható érték tétele, teljes eseményrendszerrel)**.** *Legyen $A_1, \dots, A_n$ teljes eseményrendszer Ω-n, amire $\mathbb{P}(A_i) > 0$ minden i-re. Ha $\mathbb{E}(Y) < \infty$, akkor*

$$\mathbb{E}(Y) = \sum_{i=1}^{n} \mathbb{E}(Y \mid A_i)\mathbb{P}(A_i).$$

6. Folytonos valószínűségi változók

Eddig olyan valószínűségi változókról beszéltünk, amik értékei egy véges, vagy megszámlálhatóan végtelen halmazt alkotnak, tipikusan az egész számok valamely részhalmazát. Vannak viszont olyan véletlen mennyiségek, amiket célszerű úgy modellezni, hogy bármilyen valós értéket felvehessenek. Ilyen sok fizikai mennyiség, vagy valamely történés bekövetkezéséig eltelt idő. Most ezekről lesz szó.

6.1. Eloszlásfüggvény

A feladatokban már előfordult, hogy választottunk egy számot „egyenletesen véletlenszerűen" egy intervallumból, esetleg egy pontot egy kétdimenziós alakzatból. Ezek éppen olyan véletlen mennyiségek – a múlt heti szóhasználattal valószínűségi változók –, amik nem diszkrétek, azaz nem csak egy megszámlálhatóan végtelen halmazból vesznek fel értékeket.

Ez miért probléma? Amikor X egy diszkrét valószínűségi változó, akkor őt le tudjuk írni az **eloszlás**a segítségével, vagyis a

$$\mathbb{P}(X = k_1),\ \mathbb{P}(X = k_2),\ \dots$$

0 és 1 közti számokból álló sorozattal, ahol $k_1, k_2, \dots$ az X lehetséges értékei. Legyen most X egy egyenletesen véletlen szám a $[0,1]$ intervallumból. Ez alatt azt értjük, hogy X olyan valószínűségi változó, amire $\mathbb{P}(X \le t) = t$ ha $t \in [0,1]$, például $\mathbb{P}(X \le \frac{1}{2}) = \frac{1}{2}$. Ekkor $\mathbb{P}(X = k) = 0$ bármilyen k esetén, vagyis a fenti (diszkrét értelemben vett) eloszlás lényegében semmit nem mond a valószínűségi változóról.

Erre mondhatnánk, hogy „Oké, de a változót akkor is leírja, hogy a $[0,1]$-ből veszi fel az értékeit, így elég, ha ezzel a tulajdonsággal hivatkozunk rá." A következő példa mutatja, hogy miért nem.

6.1.1. Példa. Legyen X egy egyenletesen véletlen szám a $[0,1]$ intervallumból, és nézzük az $Y = X^2$ valószínűségi változót. Ekkor Y értékei szintén a $[0,1]$ intervallumból valók, de Y mégsem ugyanúgy működik, mint X. Hiszen egyrészt $\mathbb{P}(X \le \frac{1}{2}) = \frac{1}{2}$, másrészt

$$\mathbb{P}\Big(Y \le \frac{1}{2}\Big) = \mathbb{P}\Big(X^2 \le \frac{1}{2}\Big) = \mathbb{P}\Big(X \le \frac{1}{\sqrt{2}}\Big) = \frac{1}{\sqrt{2}}.$$

Vagyis Y nagyobb eséllyel vesz fel kisebb értékeket, az „eloszlása jobban koncentrálódik a 0 körül, mint X eloszlása". Mindjárt tisztázzuk, hogy ez mit is jelent.

Általánosan, az X eloszlását az ún. eloszlásfüggvénye segítségével írhatjuk le.

6.1.2. Definíció. Legyen X valószínűségi változó. Ekkor az

$$F_X : \mathbb{R} \to [0,1] \qquad F_X(x) = \mathbb{P}(X < x)$$

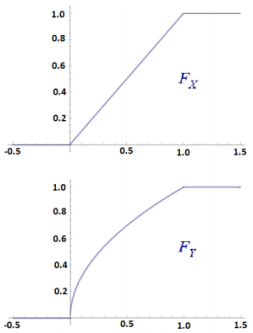
függvényt az X **eloszlásfüggvény**ének hívjuk.

Vegyük észre, hogy $\{X < x\}$ eleme $\mathcal{F}$-nek, azaz esemény (mivel X valószínűségi változó), ezért van értelme beszélni a $\mathbb{P}(X < x)$ valószínűségről.

A fenti példában szereplő X és Y eloszlásfüggvényei:

$$F_X(x) = \mathbb{P}(X < x) = \begin{cases} 0 & \text{ha } x \le 0, \\ x & \text{ha } x \in (0,1], \\ 1 & \text{ha } x > 1, \end{cases}$$

$$F_Y(x) = \mathbb{P}(Y < x) = \mathbb{P}(X^2 < x) = \mathbb{P}(X < \sqrt{x}) = \begin{cases} 0 & \text{ha } x \le 0, \\ \sqrt{x} & \text{ha } x \in (0,1], \\ 1 & \text{ha } x > 1. \end{cases}$$



Nézzük, mit tud általánosan egy eloszlásfüggvény. Egyrészt világos, hogy

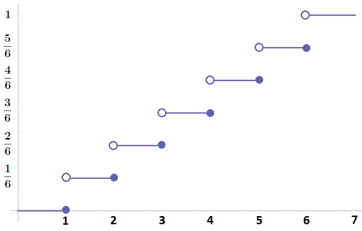
$$F_X(b) - F_X(a) = \mathbb{P}(X < b) - \mathbb{P}(X < a) = \mathbb{P}(a \le X < b)$$

tetszőleges $a < b$ valós számokra, $\mathbb{P}$ additivitása miatt. Az eloszlásfüggvények karakterizálhatók is.

6.1.3. Állítás. *Egy $F : \mathbb{R} \to [0,1]$ függvény akkor és csak akkor eloszlásfüggvénye valamilyen valószínűségi változónak, ha*

(1) F (nem feltétlenül szigorúan) monoton növő,
(2) F balról folytonos, azaz minden x-re az F baloldali határértéke x-ben $F(x)$,
(3) $\lim_{x\mapsto-\infty} F(x) = 0$ és $\lim_{x\mapsto\infty} F(x) = 1$.

A balról folytonosság közel sem jelent folytonosságot. Például diszkrét valószínűségi változók eloszlásfüggvénye sosem folytonos. Igen, eloszlásfüggvény diszkrét esetben is mindig van, és ez is balról folytonos, de jobbról nem. Például egy kockadobás, mint valószínűségi változó, eloszlásfüggvényét lásd jobbra.



Bizonyítás. Legyen X valószínűségi változó, és $x < y$. Belátjuk, hogy F_X-re igaz a fenti három feltétel. Valóban, F_X monoton növő, hiszen $\{X < x\} \subseteq \{X < y\}$, ezért

$$F_X(x) = \mathbb{P}(X < x) \leq \mathbb{P}(X < y) = F_X(y),$$

felhasználva a valószínűségi mezőről szóló 1.2 alfejezet következményét.

Nézzük a második tulajdonságot. Az F_X balról folytonossága ekvivalensen azt jelenti (átviteli elv), hogy bármely monoton növő $(x_n)_{n\in\mathbb{N}}$ sorozatra, amire $x_n \neq x$ és $x_n \to x$, teljesülnie kell, hogy $\lim_{n\to\infty} F_X(x_n) = F_X(x)$. Megmutatjuk, hogy ez valóban teljesül. Egyrészt

$$\lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} \mathbb{P}(X < x_n) = \mathbb{P}(X < x_0) + \lim_{n\to\infty} \mathbb{P}\Big(\bigcup_{k=1}^{n} \{x_{k-1} \leq X < x_k\}\Big),$$

felhasználva a valószínűség additivitását, illetve hogy $\{X < x_n\} = \{X < x_0\} \cup \{x_0 \leq X < x_n\} = \{X < x_0\} \cup \bigcup_{k=1}^{n} \{x_{k-1} \leq X < x_k\}$. A második tag a következőképp alakítható át:

$$\lim_{n\to\infty} \mathbb{P}\Big(\bigcup_{k=1}^{n} \{x_{k-1} \leq X < x_k\}\Big) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{P}(x_{k-1} \leq X < x_k) = \mathbb{P}(x_0 \leq X < x),$$

felhasználva a valószínűség σ-additivitását, illetve hogy $\cup_{k=1}^{\infty} \{x_{k-1} \leq X < x_k\} = \{x_0 \leq X < x\}$. Ezt visszahelyettesítve az előző egyenletbe azt kapjuk, hogy $\lim_{n\to\infty} F_X(x_n) = \mathbb{P}(X < x_0) + \mathbb{P}(x_0 \leq X < x) = \mathbb{P}(X < x) = F_X(x)$.[30] Hasonló okból teljesül a harmadik feltétel is.[31]

Visszafelé, legyen adott F, ehhez keresünk egy megfelelő X valószínűségi változót, amire $F = F_X$. Legyen U egy egyenletesen véletlen szám a $[0,1]$ intervallumból.[32] Definiáljuk az X-et, mint

$$X = \inf\{y \in \mathbb{R} \mid U < F(y)\}.$$

Ekkor az X és az infimum definíciója miatt

$$\mathbb{P}(X < x) = \mathbb{P}\big(\inf\{y \in \mathbb{R} \mid U < F(y)\} < x\big) =$$
$$= \mathbb{P}(\text{van } y \in \mathbb{R}, \text{ amire } y < x \text{ és } U < F(y)\big).$$

Belátható, hogy ilyen y pontosan akkor létezik, ha $U < F(x)$. Valóban, ha van ilyen y, akkor $U < F(y) \leq F(x)$, mivel F monoton növő. Megfordítva, ha $U < F(x)$, akkor F balról folytonossága miatt van x-hez elég közel egy y, amire $U < F(y)$ szintén teljesül. Következésképp:

$$\mathbb{P}(X < x) = \mathbb{P}\big(U < F(x)\big) = F(x),$$

hiszen $0 \leq F(x) \leq 1$, és a példában láttuk, hogy $\mathbb{P}(U < z) = z$ bármilyen $0 < z < 1$ szám esetén. Ezzel az állítást beláttuk. □

[30] Ezzel az érveléssel kaphatjuk, hogy az F_X jobboldali határértéke x-ben $\mathbb{P}(X \leq x)$.

[31] A hasonló ok neve a valószínűség folytonossági tulajdonsága.

[32] Azt illenék bebizonyítani, hogy ilyen egyenletesen véletlen szám, mint valószínűségi változó, valóban létezik. Ennek a precíz leírása támaszkodik a Lebesgue-mérték fogalmára, így ettől most eltekintünk.

6.2. Sűrűségfüggvény

Most a valószínűségi változók másik fontos hozzárendelt függvényét vizsgáljuk: a sűrűségfüggvényét. Ennek többek közt az az oka, hogy az eloszlásfüggvény grafikonjáról nem feltétlenül könnyű leolvasni a valószínűségi változó tulajdonságait. Szó volt már például az X és X^2 változókról, ahol X egyenletesen véletlenszerű $[0,1]$-en. Meg tudjuk-e mondani az eloszlásfüggvény ábrája alapján, melyik szám 0,01 sugarú környezetében lesz a legnagyobb eséllyel X^2? Vagy hogy hányszor akkora eséllyel lesz az X^2 értéke az $\frac{1}{4}$ kis környezetében, mint $\frac{3}{4}$ kis környezetében?

Az első kérdésre hamar rávághatjuk, hogy 0-nál (egész pontosan 0,01-nél), hiszen ott nő a legmeredekebben az eloszlásfüggvény, más szavakkal az ottani x-ek járulnak hozzá leginkább az $F_{X^2}(x) = \mathbb{P}(X^2 < x)$ valószínűség növekedéséhez. A második kérdés valamivel trükkösebb, a válasz $\sqrt{3}$, amihez az eloszlásfüggvény érintőinek meredekségeit kell összehasonlítsuk, lásd lejjebb. Vegyük észre, hogy mindkét esetben a meredekségeket kell vizsgáljuk.

Ha valakinek rossz előérzete lenne az „érintő meredeksége” szókapcsolat hallatán, megnyugtatnám, deriválni fogunk. Hiszen az első példa is éppen azt mutatja, hogy az eloszlásfüggvény meredekségeinek függvénye, azaz deriváltfüggvénye lenne hasznunkra. Már amennyiben létezne, csakhogy 0-ban és 1-ben F_{X^2} nem deriválható. Ezt a problémát megkerülendő, a deriválás egyik általánosítását fogjuk használni, ami a „jó-lesz-az-úgy” filozófiát követve egyszerűen nem foglalkozik azzal, hogy a függvény egy-egy pontban nem deriválható (ha a függvény legalább folytonos). Ez nem fogja elrontani a számításainkat.

6.2.1. Definíció. Egy X valószínűségi változót **folytonos**nak nevezünk, ha létezik olyan nemnegatív, valós $f_X : \mathbb{R} \to \mathbb{R}$ függvény, amire az $\int_{-\infty}^{\infty} f_X(z)\mathrm{d}z$ improprius Riemann-integrál véges[33], és minden $x \in \mathbb{R}$ esetén

$$\int_{-\infty}^{x} f_X(z)\mathrm{d}z = F_X(x),$$

ahol az integrál improprius Riemann-integrál. Az f_X függvényt az X **sűrűségfüggvény**ének hívjuk.

A definíció nem túl konstruktív: a feltételként adott integrálokból nehéz kiszámolni f_X-et. Sőt, valójában azt sem biztosítja, hogy a sűrűségfüggvény egyértelmű legyen, hiszen ha f sűrűségfüggvénye X-nek, akkor az is sűrűségfüggvény, ha f-et megváltoztatjuk egy ponton (hiszen az integrálok nem változnak). Hogyan lehet akkor kiszámolni valamit, ami nem is egyértelmű?

6.2.2. Állítás. *Ha F_X folytonos és végessok pont kivételével mindenhol deriválható, akkor X folytonos valószínűségi változó, és az*

$$f(x) = \begin{cases} F_X'(x) & \text{ha } F_X \text{ deriválható } x\text{-ben,} \\ 0 & \text{egyébként} \end{cases} \qquad (x \in \mathbb{R})$$

függvény sűrűségfüggvénye X-nek.

6.2.3. Példa. Az állítás szerint a fenti $[0,1]$-en egyenletes eloszlású X esetében

$$f_X(x) = \begin{cases} 1 & \text{ha } 0 < x < 1, \\ 0 & \text{egyébként,} \end{cases} \qquad \text{és} \qquad f_{X^2}(x) = \begin{cases} \frac{1}{2\sqrt{x}} & \text{ha } 0 < x < 1, \\ 0 & \text{egyébként} \end{cases} \qquad (x \in \mathbb{R})$$

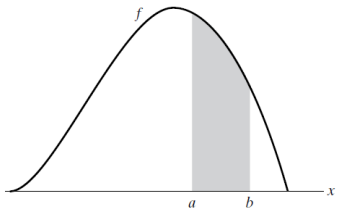
függvények sűrűségfüggvényei X-nek és X^2-nek. Szemléletesen a sűrűségfüggvény értékei annak az esélyét jelölik, hogy az X változó a x érték kis környezetébe esik[34]. Így az alfejezet elején feltett második kérdés válasza $f_{X^2}\left(\frac{1}{4}\right)/f_{X^2}\left(\frac{3}{4}\right) = \frac{1}{1}/\frac{1}{\sqrt{3}} = \sqrt{3}$.

[33] Általánosabban Lebesgue-integrálható sűrűségfüggvényről is beszélhetnénk. Most nem fogunk.

[34] Feltéve, hogy a sűrűségfüggvény épp folytonos. Ha egy-egy pontban megszüntethető szakadása van f_{X^2}-nek, akkor a pontbeli értékének nincs jelentéstartalma.

Megjegyzés. Mi most nem fogunk ilyen esetekkel foglalkozni, de egy valószínűségi változó lehet olyan, ami nem diszkrét, de nem is folytonos (például egy indikátor változó és egy folytonos szorzata). Sőt az sem igaz, hogy ha F_X folytonos, abból következne, hogy X is folytonos (bár a megfordítottja teljesül: egy folytonos változó eloszlásfüggvénye folytonos). Az ilyen kényelmetlen eseteket most tegyük félre.

A sűrűségfüggvény praktikus haszna jóval nagyobb annál, mint hogy a fenti kis környezetekről információval szolgál. Nézzük tehát a tulajdonságait.



6.2.4. Állítás. *Legyen X folytonos valószínűségi változó. Ekkor minden $a < b$ valós szám esetén*

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)\mathrm{d}x.$$

Bizonyítás. A valószínűség additivitása miatt

$$\mathbb{P}(a < X < b) = \mathbb{P}(X < b) - \mathbb{P}(X < a) - \mathbb{P}(X = a) =$$

$$= \int_{-\infty}^b f_X(x)\mathrm{d}x - \int_{-\infty}^a f_X(x)\mathrm{d}x - 0 = \int_a^b f_X(x)\mathrm{d}x,$$

ahol felhasználtuk, hogy az integrálás additív az integrálási intervallumban. □

Az eloszlásfüggvényhez hasonlóan a sűrűségfüggvények is karakterizálhatók.

6.2.5. Állítás. *Egy nemnegatív $f : \mathbb{R} \to \mathbb{R}$ függvényhez akkor és csak akkor létezik X folytonos valószínűségi változó, aminek az f sűrűségfüggvénye, ha f Riemann-integrálható és*

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1.$$

Az világos, hogy ha X folytonos valószínűségi változó, akkor f_X-re teljesül az egyenlet. A visszafelé irányt nem bizonyítjuk.

Feladat. Jelölje egy alkatrész élettartamát Z (órákban számolva). Ha Z eloszlásfüggvénye

$$F_Z(x) = \begin{cases} 0 & \text{ha } x \le 100, \\ 1 - \frac{100}{x} & \text{ha } x > 100, \end{cases}$$

akkor mi a valószínűsége, hogy az alkatrész nem romlik el az első 150 órában?

6.3. Várható érték és szórás, folytonos eset

Diszkrét esetben a valószínűségi várható értékét definíció szerint úgy határoztuk meg (amikor értelmeztük a várható értéket), hogy az értékeket a valószínűségeikkel súlyozva összegeztük. Ennek analógja folytonos esetben, hogy a sűrűségfüggvény $f_X(t)$ értékét t-vel szorozva integráljuk ($-\infty$-től ∞-ig).

6.3.1. Definíció. Legyen X folytonos valószínűségi változó, amire

$$\int_{-\infty}^{\infty} |t| \cdot f_X(t)\mathrm{d}t < \infty.$$

Ekkor

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} t \cdot f_X(t)\mathrm{d}t. \tag{6.1}$$

A definíció feltételére azért van szükség, hogy kizárjuk azt az esetet, amikor $\mathbb{E}(X)$ nem definiált.

6.3.2. Definíció. Egy X valószínűségi változó **egyenletes eloszlás**ú az (a, b) intervallumon, ha sűrűségfüggvénye

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{ha } a < x < b, \\ 0 & \text{egyébként.} \end{cases}$$

Jelölése: $X \sim U(a; b)$.

Ez valóban sűrűségfüggvény, hiszen

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_a^b \frac{1}{b-a}\mathrm{d}x = \Big[\frac{x}{b-a}\Big]_a^b = \frac{b-a}{b-a} = 1.$$

Egy egyenletes eloszlású valószínűségi változó várható értéke

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\mathrm{d}x = \int_a^b \frac{x}{b-a}\mathrm{d}x = \Big[\frac{x^2}{2(b-a)}\Big]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2},$$

ami intuitívan is világos: átlagosan az intervallum közepét kapjuk értékül.

Vegyük észre, mennyire hasonlítanak a diszkrét esetre vonatkozó (3.4) egyenlet és a folytonos esetre vonatkozó (6.1) egyenlet. Ez azért van, mert mindkettő a várható érték általános fogalmának a realizációja. Ez a lenti állításban is megnyilvánul, ahol párhuzamosan tárgyalhatjuk a két esetet.

6.3.3. Állítás (Transzformált várható értéke)**.** *Legyen X valószínűségi változó, és $g : \mathbb{R} \to \mathbb{R}$ függvény. Tegyük fel, hogy $\mathbb{E}(g(X))$ létezik. Ha X diszkrét, akkor*

$$\mathbb{E}(g(X)) = \sum_{j=1}^{\infty} g(k_j) \cdot \mathbb{P}(X = k_j),$$

ahol $\mathrm{Ran}(X) = \{k_1, k_2, \dots\}$*. Ha pedig X folytonos, akkor*

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\mathrm{d}x.$$

Itt a diszkrét eset a 3.2. Állítás általánosítása egyszerű valószínűségi változóról diszkrétre.

6.3.4. Példa. A 6.3.1. Definícióban egy X folytonos valószínűségi változó várható értékét akkor értelmeztük, ha

$$\int_{-\infty}^{\infty} |t| \cdot f_X(t)\mathrm{d}t$$

véges. Vegyük észre, hogy a 6.3.3. Állítás alapján ez a kifejezés nem más, mint $\mathbb{E}(|X|)$.

6.3.5. Példa. Legyen X olyan valószínűségi változó, aminek $f_X : \mathbb{R} \to \mathbb{R}$ sűrűségfüggvényére teljesül, hogy $f(x) = 2e^{-2x}$ ha $x \in [0, \infty)$, és 0 egyébként. (Ellenőrizzük le, hogy ez valóban sűrűségfüggvény.) Ekkor

$$\mathbb{E}(e^X) = \int_{-\infty}^{\infty} e^x \cdot f_X(x)\mathrm{d}x = \int_0^{\infty} e^x 2e^{-2x}\mathrm{d}x = 2\int_0^{\infty} e^{-x}\mathrm{d}x = 2.$$

Legyen X folytonos valószínűségi változó, és jelölje f_X a sűrűségfüggvényét. Hogy tudjuk meghatározni X szórását?

Emiatt az X szórásnégyzetét is ki tudjuk számolni:

$$\mathbb{D}^2(X) \overset{\text{def}}{=} \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)^2\Big) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\mathrm{d}x - \Big(\int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x\Big)^2.$$

Ebből pedig X szórása $\mathbb{D}(X) = \sqrt{\mathbb{D}^2(X)}$.

A szórásnégyzet (illetve szórás) jelentése ilyen esetben is átlagtól való négyzetes eltérés (és annak gyöke). Szemléletesen azt méri, mennyire „terül szét” a sűrűségfüggvény a várható érték körül.[35]

[35] A sűrűségfüggvény alakjáról számos további származtatott mennyiség nyilatkozik, mint a valószínűségi változó átlagos abszolút eltérése, a csúcsossága (más néven lapultsága), vagy a ferdesége.

6.3.6. Példa (Egyenletes eloszlás szórásnégyzete)**.** Legyen $X \sim U(a;b)$. Ekkor a két köb különbségére vonatkozó azonosság alapján

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\mathrm{d}x = \int_a^b x^2 \frac{1}{b-a}\mathrm{d}x$$
$$= \frac{1}{b-a}[x^3/3]_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{(b-a)(b^2+ab+a^2)}{3(b-a)} = \frac{b^2+ab+a^2}{3},$$

és így

$$\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{b^2+ab+a^2}{3} - \frac{(a+b)^2}{4} = \frac{4b^2+4ab+4a^2-3a^2-6ab-3b^2}{12}$$
$$= \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}.$$

Bár a tárgy jelenlegi változatában egy Y valószínűségi változó és egy A esemény esetén az $\mathbb{E}(Y|A)$ feltételes várható érték főleg abban az esetben érdekel minket, amikor Y diszkrét valószínűségi változó, a 5.1.1. Definíció folytonos valószínűségi változókra is értelmes, amennyiben $\mathbb{P}(A) > 0$. Lássunk erre egy példát!

6.3.7. Példa. Legyen Y egyenletes eloszlású a $(0;1)$ intervallumon. Feltéve, hogy $Y > 1/2$, mennyi Y várható értéke?
A definíció alkalmazása előtt próbáljunk meg a kérdésre intuitívan helyes választ adni. Mivel Y egyenletes eloszlású, egy adott intervallumba az intervallum hosszúságával arányos valószínűséggel esik, így a feltételes valószínűség definíciója szerint ez akkor is igaz marad, ha az $\{Y > 1/2\}$ eseményre feltételezünk. Tehát a helyes válasz az $(1/2;1)$ intervallum középpontja, vagyis $\frac{3}{4}$.
Precízebben: Y nemnegatív valószínűségi változó, így a várható értéke $\mathbb{E}(Y) = \int_0^\infty \mathbb{P}(Y \geq t)\mathrm{d}t$. Ezt fogjuk alkalmazni, azonban a 5.1.1. Definíció szerint $\mathbb{P}(\ldots)$ helyett $\mathbb{P}(\ldots|Y > 1/2)$-del. Tehát szükségünk van $\mathbb{P}(Y \geq t|Y > 1/2)$-re minden $t \geq 0$ esetén.
Ha $1/2 \leq t \leq 1$, akkor

$$\mathbb{P}(Y \geq t|Y > 1/2) = \frac{\mathbb{P}(Y \geq t)}{\mathbb{P}(1/2 < Y < 1)} = \frac{1-t}{1/2} = 2(1-t),$$

illetve ha $0 \leq t \leq 1/2$, akkor $\mathbb{P}(Y \geq t|Y > 1/2) = 1$, és ha $t > 1$, akkor $\mathbb{P}(Y \geq t|Y > 1/2) = 0$. Így

$$\mathbb{E}(Y|Y > 1/2) = \int_0^\infty 1 - \mathbb{P}(Y \geq t|Y > 1/2)\mathrm{d}t = \int_0^{1/2} 1\mathrm{d}t + \int_{1/2}^1 2(1-t)\mathrm{d}t = \frac{1}{2} + [(1-t)^2]_{t=1/2}^1 = \frac{3}{4}.$$

Kitekintés. Egy Y valószínűségi változó feltételes várható értéke definiálható egy X folytonos valószínűségi változóra feltételezve is. Itt azonban külön nehézséget jelent a diszkrét esethez képest, hogy ha $\mathbb{E}(Y|X)$-et $g(X)$-ként akarjuk definiálni, ahol $g(x) = \mathbb{E}(Y|X = x)$, akkor azzal a problémával kell szembesülnünk, hogy X folytonossága miatt $\{X = x\}$ minden $x \in \mathbb{R}$ esetén 0 valószínűségű esemény és így nem lehet rá feltételezni, ezért nem világos, hogy $\mathbb{E}(Y|X = x)$ mit is jelent.

A probléma az úgynevezett *feltételes sűrűségfüggvény* bevezetésével orvosolható, amelynek segítségével definiálható $g(x) = \mathbb{E}(Y|X = x)$ (legalábbis abban az esetben, amikor (X, Y) folytonos valószínűségi vektorváltozó), ezután pedig a feltételes várható érték definíciója ismét $g(X) = \mathbb{E}(Y|X)$ lesz. Ezt idén előadáson nem tanuljuk; az érdeklődőknek a függelékben található A.6. fejezetet ajánljuk, ahol megadjuk a feltételes sűrűségfüggvény definícióját is. Az utóbbi fejezetből kiderül az is, hogy a teljes várható érték tétele formálisan igaz marad folytonos esetben is (azzal a különbséggel, hogy $\mathbb{E}(Y|X)$-et máshogy kell kiszámolni), valamint létezik a teljes valószínűség tételének is egy olyan változata, amelyben egy folytonos valószínűségi változó összes lehetséges értékére feltételezve számolhatjuk ki egy esemény valószínűségét. Ezenkívül a feltételes várható érték összes azonossága (lásd a 5.3.1. Állítást) igaz marad folytonos esetben is.

6.4. Exponenciális eloszlás, örökifjú tulajdonság II

6.4.1. Definíció. Egy Z valószínűségi változó **exponenciális eloszlású** λ paraméterrel ($\lambda > 0$ valós), ha

$$f_Z(x) = \begin{cases} \lambda e^{-\lambda x} & \text{ha } x > 0, \\ 0 & \text{egyébként,} \end{cases} \qquad \text{azaz} \qquad F_Z(x) = \begin{cases} 1 - e^{-\lambda x} & \text{ha } x > 0, \\ 0 & \text{egyébként.} \end{cases}$$

Jelölése: $Z \sim \mathrm{Exp}(\lambda)$.

Ez valóban sűrűségfüggvény, hiszen egyrészt nemnegatív, másrészt

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_0^{\infty} \lambda e^{-\lambda x}\mathrm{d}x = \Big[-e^{-\lambda x} \Big]_0^{\infty} = 0 + e^{-0} = 1.$$

Egy exponenciális eloszlású valószínűségi változó várható értéke:

$$\mathbb{E}(Z) = \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x = \int_0^{\infty} \lambda x e^{-\lambda x}\mathrm{d}x =$$

$$= \Big[x(-e^{-\lambda x}) \Big]_0^{\infty} - \int_0^{\infty} (-e^{-\lambda x})\mathrm{d}x = 0 + \int_0^{\infty} e^{-\lambda x}\mathrm{d}x = 0 - \frac{1}{-\lambda} = \frac{1}{\lambda}.$$

Az eloszlás főleg olyan szituációkban kerül elő, amikor sűrű, egymás utáni kísérletek valamelyikének sikerességét várjuk. A λ arányparaméter (angolul: rate) jelentése, hogy egység idő alatt átlagosan hány ilyen „siker” történik.

Az exponenciális eloszlás a geometriai eloszlás folytonos analógja. Mindkét eloszlás fejezik ki, hogy meddig kell várni egy történésre, aminél az eltelt idő nem változtat a bekövetkezés esélyén. A különbség a két eset közt, hogy az első esetben az idő diszkrét, míg a másodikban folytonos módon van modellezve.

6.4.2. Állítás. *Ha X nemkonstans örökifjú valószínűségi változó a $G = [0, \infty)$ halmazon, akkor X eloszlása exponenciális.*

Kicsit szemszögből megfogalmazva ugyanez:

Folytonos és örökifjú eloszlás csak exponenciális lehet (a nemkonstans esetben).

Vagyis ez a két eloszlás egymás analógja. Azt fejezik ki, hogy meddig kell várni egy történésre, aminél az eltelt idő nem változtat a bekövetkezés esélyén. A különbség a két eset közt, hogy az első esetben az idő diszkrét, míg a másodikban folytonos módon van modellezve.

Bizonyítás. Bontsuk ki a feltételes valószínűség definícióját. Azt kapjuk, hogy a (3.5) egyenlet ekvivalens a következővel:

$$\mathbb{P}(X > t + s) = \mathbb{P}(X > t)\mathbb{P}(X > s) \qquad (\forall s, t \in G).$$

Legyen most G a nemnegatív valós számok halmaza, és jelöljük $g(t)$-vel az $\ln \mathbb{P}(X > t)$ mennyiséget[36]. Ez a logaritmus értelmes (azaz $\mathbb{P}(X > t) \neq 0$), hacsak X nem konstans 0, amit viszont kizártunk. Ekkor a (3.5) egyenlet logaritmusát véve kapjuk, hogy

$$g(t + s) = \ln \mathbb{P}(X > t + s) = \ln \mathbb{P}(X > t) + \ln \mathbb{P}(X > s) = g(t) + g(s),$$

vagyis g egy additív függvény a nemnegatív valós számokon. Szerencsére azt is tudjuk, hogy g monoton csökkenő, hiszen az eloszlásfüggvény $t \mapsto F_X(t) = \mathbb{P}(X < t)$ monoton növő, így $t \mapsto \mathbb{P}(X > t)$ monoton csökkenő. De akkor a logaritmusa is monoton csökkenő kell legyen.

6.4.3. Lemma. *Ha $g : [0, \infty) \to \mathbb{R}$ monoton csökkenő függvény, amire $g(t + s) = g(t) + g(s)$ teljesül tetszőleges $s, t \in [0, \infty)$ esetén, akkor $g(t) = -\lambda t$, ahol $\lambda = -g(1) \geq 0$.*[37]

[36] Itt ln jelöli a természetes alapú logaritmust.

[37] A monotonitás nélkül nem igaz az állítás, valami szépségi feltételre szükség van (folytonos, vagy korlátos, Lebesgue-mérhető, ...). Bővebben lásd: Cauchy függvényegyenlete.

A lemmát nem bizonyítjuk. A lemma miatt $\ln \mathbb{P}(X > t) = g(t) = -\lambda t$, ahol $\lambda = -\ln \mathbb{P}(X > 1)$. Emiatt

$$\mathbb{P}(X \leq t) = 1 - \mathbb{P}(X > t) = 1 - e^{-\lambda t},$$

ahol $\lambda \neq 0$, különben $\mathbb{P}(X > t) = 1$ minden t-re teljesülne, ami lehetetlen. Nekünk $\mathbb{P}(X \leq t)$ helyett az eloszlásfüggvényre, azaz $\mathbb{P}(X < t)$-re lenne szükségünk. Vegyük észre, hogy $t \mapsto \mathbb{P}(X \leq t)$ folytonos, ezért $\mathbb{P}(X < t) = \lim_{s \nearrow t} \mathbb{P}(X \leq s) = \mathbb{P}(X \leq t) = 1 - e^{-\lambda t}$, ami épp az exponenciális eloszlás eloszlásfüggvénye. □

A geometriai és az exponenciális eloszlás közti kapcsolat abban is megnyilvánul, hogy a geometriai eloszlás kifejezhető az exponenciálisból.[38]

6.4.4. Állítás. *Ha az X valószínűségi változó exponenciális eloszlású λ paraméterrel, akkor $\lceil X \rceil$ geometriai eloszlású $1 - e^{-\lambda}$ paraméterrel.*

Bizonyítás. Világos, hogy $\mathbb{P}\big(\lceil X \rceil = 0\big) = \mathbb{P}(-1 < X \leq 0) = 0$. Továbbá tetszőleges k pozitív egész esetén

$$\mathbb{P}\big(\lceil X \rceil = k\big) = \mathbb{P}(k-1 < X \leq k) = F_X(k) - F_X(k-1) =$$
$$= (1 - e^{-\lambda k}) - (1 - e^{-\lambda(k-1)}) = e^{-\lambda(k-1)} - e^{-\lambda k} = e^{-\lambda(k-1)}(1 - e^{-\lambda}),$$

ami a $p = 1 - e^{-\lambda}$ jelöléssel éppen $(1-p)^{k-1}p$, vagyis $\lceil X \rceil$ geometriai eloszlású $1 - e^{-\lambda}$ paraméterrel. □

6.4.5. Példa. Legyen $Z \sim \mathrm{Exp}(\lambda)$ valamilyen λ pozitív valósra. Számítsuk ki Z szórásnégyzetét! Két parciális integrálással adódik, hogy

$$\mathbb{E}(Z^2) = \int_0^\infty z^2 \lambda e^{-\lambda z} \mathrm{d}z = \Big[-e^{-\lambda z} z^2 \Big]_0^\infty - \int_0^\infty -e^{-\lambda z} 2z \mathrm{d}z = \int_0^\infty 2z e^{-\lambda z} \mathrm{d}z =$$
$$= \left[2z \Big(-\frac{1}{\lambda} \Big) e^{-\lambda z} \right]_0^\infty - \int_0^\infty 2 \Big(-\frac{1}{\lambda} \Big) e^{-\lambda z} \mathrm{d}z = \frac{2}{\lambda} \int_0^\infty e^{-\lambda z} \mathrm{d}z = \frac{2}{\lambda^2}.$$

Tehát

$$\mathbb{D}^2(Z) = \mathbb{E}(Z^2) - \mathbb{E}(Z)^2 = \frac{2}{\lambda^2} - \Big(\frac{1}{\lambda}\Big)^2 = \frac{1}{\lambda^2} \qquad \Longrightarrow \qquad \mathbb{D}(Z) = \frac{1}{\lambda}.$$

6.5. Várható érték, nemnegatív eset

Diszkrét esetben a súlyfüggvény segítségével összegzéssel, folytonos esetben pedig a sűrűségfüggvény segítségével integrálással számítottuk ki a várható értéket. Felmerülhet a kérdés, hogy létezik-e a várható értéknek olyan általános fogalma, amely mindkettőt magában foglalja. Ez hasznos lehet például olyan valószínűségi változók esetén, amelyek se nem diszkrétek, se nem folytonosak.[39]

Előző előadáson láttuk, hogy ha X valószínűségi változó, akkor X^+ (azaz X pozitív része) és X^- (azaz X negatív része) nemnegatív valószínűségi változók, és $X = X^+ - X^-$. Így a 3.5.4. Definíció alapján ha $\mathbb{E}(X^+)$-t és $\mathbb{E}(X^-)$-t már definiáltuk és ezek közül legalább az egyik véges, akkor $\mathbb{E}(X) := \mathbb{E}(X^+) - \mathbb{E}(X^-)$ definícióval X várható értékét is definiálhatjuk. Így a várható érték egy általános definícióját elég *nemnegatív* valószínűségi változókra megadni. Ez pedig a következő.

6.5.1. Definíció. Legyen X nemnegatív valószínűségi változó és jelölje eloszlásfüggvényét F_X. Ekkor

$$\mathbb{E}(X) = \int_0^\infty (1 - F_X(x)) \mathrm{d}x = \int_0^\infty (1 - \mathbb{P}(X < x)) \mathrm{d}x \in [0, \infty].$$

[38] Megfordítva, az exponenciális eloszlás ún. határeloszlása a geometriainak, ha besűrítjük a geometriai lehetséges bekövetkezési időpontjait az $\frac{1}{n}$ többszöröseivé, $n \to \infty$ és $np \to \lambda$.

[39] Egy ilyen valószínűségi változóra nem túlságosan elvadult példa az az X, amelyet úgy kapunk, hogy egy szabályos pénzérmét feldobunk, majd ha fejet kapunk, akkor X értékét 0-nak választjuk, írás esetén pedig egy számítógéppel generált, a $(0; 1)$ intervallumon egyenletes eloszlású véletlen számnak. Ez nem folytonos valószínűségi változó, hiszen a 0 értéket pozitív valószínűséggel veszi fel, de nem is diszkrét, hiszen értékkészlete nem megszámlálható.

Itt a szemléltetés kedvéért folytonos esetben bizonyítjuk, hogy a várható érték két definíciója ekvivalens egymással. A bizonyítás hasonló diszkrét esetben is, de ezt itt nem fejtjük ki.

6.5.2. Állítás. *Tetszőleges X folytonos és nemnegatív valószínűségi változó esetén teljesül, hogy*

$$\int_{-\infty}^{\infty} t f_X(t)\mathrm{d}t = \int_0^{\infty} 1 - F_X(x)\mathrm{d}x.$$

Bizonyítás. Az eloszlás- és sűrűségfüggvény tulajdonságai alapján

$$\begin{aligned}\int_0^{\infty} 1 - F_X(x)\mathrm{d}x &= \int_0^{\infty} \mathbb{P}(X \geq x)\mathrm{d}x = \int_0^{\infty}\int_x^{\infty} f_X(t)\mathrm{d}t\mathrm{d}x \\ &= \int_0^{\infty}\int_0^{t} 1\ \mathrm{d}x f_X(t)\mathrm{d}t = \int_0^{\infty} [x]_{x=0}^{t} f_X(t)\mathrm{d}t \\ &= \int_0^{\infty} t f_X(t)\mathrm{d}t = \int_{-\infty}^{\infty} t f_X(t)\mathrm{d}t.\end{aligned}$$

A harmadik egyenlőség a Fubini-tétel miatt igaz. Az utolsó lépésben azt használtuk, hogy mivel X nemnegatív, ezért $f_X(t) = 0$, ha $x < 0$. □

Megjegyzés. A bizonyítás harmadik egyenlőségénél szereplő típusú integrálfelcserélést ($x = 0 \dots \infty$, $t = x \dots \infty \Leftrightarrow t = 0 \dots \infty$, $x = 0 \dots t$) az angol nyelvű irodalomban gyakran *Fubini flip* néven említik.

A 6.5.1. Definíció alapján immár minden nemnegatív valószínűségi változó várható értékét tudjuk értelmezni (még ha bizonyos esetekben az integrál kiszámítása nehézséget is jelenthet). Ez alapján értelmes az alábbi állítás is, amely a várható érték additivitását terjeszti ki tetszőleges nemnegatív valószínűségi változókra.

6.5.3. Állítás. *Legyenek X és Y nemnegatív valószínűségi változók. Ekkor $\mathbb{E}(X+Y) = \mathbb{E}(X)+\mathbb{E}(Y) \in [0, \infty]$.*

A bizonyítás a 6.5.1. Definícióból kiindulva az állítás egyszerűsége ellenére sem triviális, ezért itt nem részletezzük. Az állítás igaz a nemnegativitás feltételezése nélkül is abban az esetben, ha X és Y várható értéke is valós számként értelmezett, vagyis ha $\mathbb{E}(X^+)$, $\mathbb{E}(X^-)$, $\mathbb{E}(Y^+)$ és $\mathbb{E}(Y^-)$ mind végesek. Az állítás egyszerű következménye az alábbi.

6.5.4. Következmény. *Tetszőleges X valószínűségi változó esetén $\mathbb{E}(|X|) = \mathbb{E}(X^+)+\mathbb{E}(X^-) \in [0, \infty]$.*

Bizonyítás. Korábban láttuk, hogy tetszőleges X valószínűségi változó esetén $|X| = X^+ + X^-$, továbbá X^+ és X^- nemnegatív valószínűségi változók (lásd 3.5.4. Definíció). Ezért a 6.5.3. Állítás szerint $\mathbb{E}(|X|) = \mathbb{E}(X^+) + \mathbb{E}(X^-) \in [0, \infty]$. □

Végezetül pedig lássunk egy konkrét példát, ahol az 6.5.1. Definíció segítségével könnyebben kiszámíthatjuk egy folytonos valószínűségi változó várható értékét, mint az 6.3.1. Definíció segítségével.

6.5.5. Példa. Az exponenciális eloszlás várható értékét korábban az 6.3.1. Definíció alapján, parciális integrálás segítségével számítottuk ki. Azon olvasók, akik kevésbé elkötelezett hívei a parciális integrálásnak, az 6.5.1. Definíció segítségével elkerülhetik azt. Ugyanis ha $X \sim \mathrm{Exp}(\lambda)$, akkor

$$\mathbb{E}(X) = \int_0^{\infty} (1 - F_X(x))\mathrm{d}x = \int_0^{\infty} e^{-\lambda x}\mathrm{d}x = \Big[-\frac{1}{\lambda}e^{-\lambda x}\Big]_{x=0}^{\infty} = \frac{1}{\lambda}.$$

Ez az integrálás valójában része annak a hosszabb számolásnak, amivel előzőleg számítottuk ki X várható értékét. Azonban annak a számításnak az elejét – és azon belül a parciális integrálást – a „Fubini flipnek” köszönhetően sikerült elkerülnünk.

6.6. Bertrand-paradoxon

Most pedig a geometriai valószínűségek illetve folytonos valószínűségi változók egy paradoxonjáról lesz szó. Folytonos valószínűségi változókat (illetve eloszlásukat) többféleképp megadhatunk. Történhet ez a sűrűségfüggvényük vagy az eloszlásfüggvényük segítségével, de van, hogy ennél közvetettebb információnk van csak a változóról. Bertrand paradoxonja[40] rávilágít, hogy mennyire nem nyilvánvaló feladat az utóbbi esetben meghatározni a valószínűségi változó eloszlását.

Bertrand-paradoxon

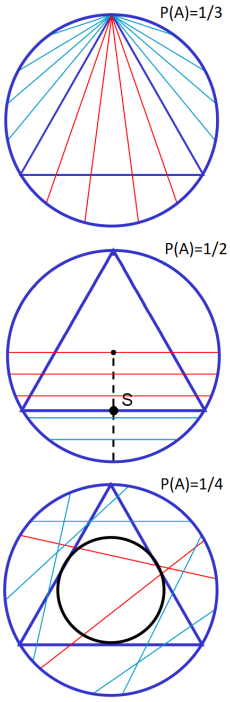
Válasszuk ki egy kör egy húrját véletlenszerűen. Mi a valószínűsége, hogy a húr hosszabb, mint a körbe írható szabályos háromszög egy oldala?

A paradoxon ellentmondása, hogy a feladatra adható olyan megoldás, amiből az $\frac{1}{3}$ válasz adódik, de olyan is, amiből $\frac{1}{2}$ vagy $\frac{1}{4}$. Hogyan történhet ez? Lássuk először a megoldásokat:

(1) Válasszunk egyenletesen véletlenszerűen és egymástól függetlenül egy P és egy Q pontot a körvonalról, és legyen a húr az őket összekötő szakasz. Sorsoljuk ki először a P pontot, majd rajzoljunk a körbe egy olyan szabályos háromszöget, aminek csúcsa ez a P. Világos, hogy pontosan akkor kapunk a háromszög oldalánál hosszabb húrt, ha Q a háromszög másik két csúcsa közé esik. Így a keresett valószínűség $\frac{1}{3}$.

(2) Válasszuk ki egyenletesen véletlenszerűen egy sugarát a körnek, majd rajta egyenletesen véletlenszerűen (a sugár választástól függetlenül) egy P pontot. A húr legyen az, amelyik a sugárra merőleges, és a sugarat P pontban metszi. Sorsoljuk ki először a sugarat, majd rajzoljunk a körbe egy olyan szabályos háromszöget, aminek egyik oldala merőlegesen metszi a sugarat, jelölje a metszéspontot S. Látható, hogy pontosan akkor kapunk a háromszög oldalánál hosszabb húrt, ha a P pont közelebb van a kör középpontjához, mint S. Mivel S felezi a sugarat, így a keresett valószínűség $\frac{1}{2}$.

(3) Válasszunk ki egyenletesen véletlenszerűen egy P pontot a körlapon. Legyen a húr az, aminek éppen P a felezőpontja. Ilyen húr mindig csak egy van (hacsak P nem a középpont, de mivel ez nulla valószínűségű lehetőség, így ezzel az esettel nem kell foglalkoznunk). Vegyünk egy szabályos háromszöget a körben, és nézzük ennek a beírható körét. A húr pontosan akkor lesz hosszabb a háromszög oldalánál, ha P a kisebb körbe esik. Az előző esetben láttuk, hogy ha a nagy kör sugara r, akkor a kis kör sugara $\frac{r}{2}$. A keresett valószínűség pedig a két kör területének aránya, azaz $\left(\frac{r}{2}\right)^2\pi \Big/ r^2\pi = \frac{1}{4}$.



A paradoxon feloldása, hogy a „véletlen húr” fogalma nem pontosan definiált. Függően attól, hogy melyik módszert alkalmazzuk a húr generálására, más-más eloszlást kapunk, ezért különböznek a valószínűségek.

Feladat. Jelölje X_1, X_2 és X_3 a húr hosszát az egyes módszerek esetén, feltéve, hogy a kör sugara 1 egység. Mutassuk meg, hogy a valószínűségi változók sűrűségfüggvényei:

$$a)\ f_{X_1}(x) = \frac{2}{\pi}\frac{1}{\sqrt{4-x^2}}, \qquad b)\ f_{X_2}(x) = \frac{x}{2\sqrt{4-x^2}}, \qquad c)\ f_{X_3}(x) = \frac{x}{2},$$

ha $x \in [0,2]$ és 0 egyébként.

[40] Ez ugyanaz a Joseph Bertrand, mint az első előadásban szereplő doboz paradoxon szerzője.

7. Folytonos együttes eloszlás

A valószínűségi változó fogalmának bevezetésekor először a diszkrét esetet vizsgáltuk meg, majd ezután következtek a folytonos valószínűségi változók. Az együttes eloszlás esetében hasonló utat járunk be: korábban megnéztük, hogy lehet leírni két diszkrét valószínűségi változó együttes viselkedését, függetlenségét; most ugyanezt tesszük folytonos valószínűségi változókkal.

7.1. Valószínűségi vektorváltozók

Sokszor találkozhatunk olyan esettel, ahol több valószínűségi változóról kell egyszerre beszélnünk. Vagy azért, mert a két véletlen mennyiség viszonyát vizsgáljuk (pl. véletlen program futásideje és memóriahasználata), vagy azért mert a változónk nem írható le egy paraméterrel (pl. egy drón helyzete a térben). Mindkét esetben természetes ötlet vektorként kezelni a valószínűségi változóinkat.

7.1.1. Definíció. Legyenek $X_1, \dots, X_n$ valószínűségi változók valamilyen n pozitív egész esetén. Ekkor az $\underline{X} \stackrel{\text{def}}{=} (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ függvényt **valószínűségi vektorváltozó**nak hívjuk. Az $\underline{X}$ **(együttes) eloszlásfüggvény**e az

$$F_{\underline{X}} : \mathbb{R}^n \to [0,1] \qquad F_{\underline{X}}(x_1, \dots, x_n) = \mathbb{P}(X_1 < x_1, \dots, X_n < x_n)$$

vektorváltozós, skalárértékű függvény.

7.1.2. Példa. Két ismerősünkkel, Aladárral és Bélával beszélgetünk külön-külön valamilyen csevegőprogrammal. Ha épp meccs van, akkor Aladár dupla annyi idő alatt reagál (hiszen nézi a meccset), míg Béla feleannyi idő alatt (mondjuk mert arról panaszkodik, hogy a saját gondolatát sem hallja a szomszédból áthallatszó üvöltés miatt). Tegyük fel, hogy alapesetben, azaz ha nincs meccs, egymástól függetlenül, exponenciális eloszlás szerint válaszolnak, $\lambda = 6$ paraméterrel.[41] Annak az esélye, hogy meccs van $\frac{1}{5}$. Mi a válaszadásuk együttes eloszlása?

Jelölje X Aladár válaszidejét, Y pedig Béláét, továbbá legyen M az az esemény, hogy meccs van. (Itt most X skalár valószínűségi változó, nem vektor, mint korábban.) Legyen Z az alábbi valószínűségi változó:

$$Z = \begin{cases} 2 & \text{ha meccs van,} \\ 1 & \text{egyébként.} \end{cases}$$

A feltételek szerint létezik U és V, amire

$$X = U \cdot Z \quad \text{és} \quad Y = V/Z, \qquad \text{ahol } U, V \sim \text{Exp}(6) \text{ függetlenek egymástól és } M\text{-től.}$$

Ha $x, y > 0$, akkor a teljes valószínűség tétele miatt

$$F_{(X,Y)}(x,y) = \mathbb{P}(X < x, Y < y) = \mathbb{P}(U \cdot Z < x,\ V/Z < y) =$$

$$= \mathbb{P}(U \cdot Z < x, V/Z < y \mid M)\mathbb{P}(M) + \mathbb{P}(U \cdot Z < x, V/Z < y \mid \overline{M})\mathbb{P}(\overline{M}).$$

Itt felhasználhatjuk, hogy M, illetve $\overline{M}$ meghatározza Z értékét, továbbá U és V függetlenek, emiatt

$$= \mathbb{P}(U \cdot 2 < x, V/2 < y)\frac{1}{5} + \mathbb{P}(U < x, V < y)\frac{4}{5} = \mathbb{P}(U < \frac{x}{2})\mathbb{P}(V < 2y)\frac{1}{5} +$$

$$\mathbb{P}(U < x)\mathbb{P}(V < y)\frac{4}{5} = (1 - e^{-6\frac{x}{2}})(1 - e^{-6 \cdot 2y})\frac{1}{5} + (1 - e^{-6x})(1 - e^{-6y})\frac{4}{5}.$$

Ha pedig $x \le 0$ vagy $y \le 0$, akkor $F_{(X,Y)}(x,y) = 0$.

Folytonos esetben – az egyváltozós esethez hasonlóan – definiáljuk egy valószínűségi vektorváltozó együttes sűrűségfüggvényét is.

[41] Hogy reális-e feltenni az exponenciális eloszlást, az a körülményektől is függ. Lásd például [article] Estimating response time in Twitter

7.1.3. Definíció. Legyen $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó. Egy $f_{\underline{X}} : \mathbb{R}^n \to [0, \infty)$ függvény az $\underline{X}$ **(együttes) sűrűségfüggvény**e, ha $f_{\underline{X}}$-nek létezik az improprius Riemann-integrálja $\mathbb{R}^n$-en, és

$$\int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{\underline{X}}(z_1, \dots, z_n) \mathrm{d}z_1 \dots \mathrm{d}z_n = F_{\underline{X}}(x_1, \dots, x_n)$$

tetszőleges $x_1, \dots, x_n \in \mathbb{R}$ esetén. $\underline{X}$-et **folytonos**nak hívjuk, ha létezik együttes sűrűségfüggvénye.

7.1.4. Állítás. *Legyen $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó. Ha $\underline{X}$ folytonos, akkor az alábbi függvény az $\underline{X}$ sűrűségfüggvénye:*

$$f_{\underline{X}}(x_1, \dots, x_n) = \begin{cases} \partial_{x_1} \dots \partial_{x_n} F_{\underline{X}}(x_1, \dots, x_n) & \text{ha ez létezik,} \\ 0 & \text{egyébként.} \end{cases}$$

(A deriválások tetszőleges sorrendben elvégezhetők.)

Megjegyzés. Az egyváltozós állítással ellentétben az állításnak feltétele, nem pedig következménye, hogy a sűrűségfüggvény létezzen (azaz $\underline{X}$ folytonos legyen). Például, ha $X_1 = X_2$, ahol X_1 egyenletes eloszlású a $[0, 1]$ intervallumon, akkor $F_{(X_1, X_2)}$ folytonos, és néhány egyenes kivételével 2-szer folytonosan differenciálható, mégis $\partial_{x_1} \partial_{x_2} F_{(X_1, X_2)}(x_1, x_2) = 0$, ami nyilván nem lehet sűrűségfüggvény. Vagyis az (X_1, X_2) vektorváltozónak $X_1 = X_2$ esetén nincs együttes sűrűségfüggvénye.

Az egyváltozós esethez hasonlóan a többváltozós sűrűségfüggvények is karakterizálhatók.

7.1.5. Állítás. *Legyen $f : \mathbb{R}^n \to [0, \infty)$. Ekkor*

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) \mathrm{d}x_1 \dots \mathrm{d}x_n = 1.$$

pontosan akkor teljesül, ha létezik $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó, aminek f sűrűségfüggvénye.[42]

Hogy nézett ki az együttes eloszlás diszkrét esetben? Ott nem együttes sűrűségfüggvényről beszéltünk, simán csak a $\mathbb{P}(X = x, Y = y)$ értékekből álló táblázatról, mint együttes súlyfüggvényről. Az előző állítás megfelelője, hogy a táblázatban szereplő számok összege 1.

Mire jó az együttes sűrűségfüggvény? Amikor Z egyváltozós, akkor többek közt a $\mathbb{P}(a < Z < b)$ alakú valószínűségeket lehetett kiszámolni vele. Ez a szerepe többváltozós esetben is megmarad:

7.1.6. Állítás. *Legyen $H \subseteq \mathbb{R}^n$ Jordan-mérhető halmaz és $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó. Ekkor*

$$\mathbb{P}(\underline{X} \in H) = \int_H f_{\underline{X}}(\underline{x}) \mathrm{d}\underline{x}.$$

7.1.7. Példa. A fenti példára visszatérve, mi az esélye, hogy Aladár előbb ír vissza, mint Béla. Az első állítás szerint x illetve y szerinti deriválással kaphatjuk a sűrűségfüggvényt, ha az létezik:

$$f_{(X,Y)}(x, y) = \begin{cases} 3e^{-3x} \cdot 12e^{-12y} \frac{1}{5} + 6e^{-6x} \cdot 6e^{-6y} \frac{4}{5} & \text{ha } x, y > 0, \\ 0 & \text{egyébként.} \end{cases}$$

Integrálással leellenőrizhető, hogy ez valóban (X, Y) sűrűségfüggvénye, ettől most eltekintünk. Innen a keresett valószínűség:

$$\mathbb{P}(X < Y) = \int_{\{x<y\}} f_{(X,Y)}(x, y) \mathrm{d}x\mathrm{d}y = \int_0^{\infty} \int_0^{y} \Big(3e^{-3x} \cdot 12e^{-12y} \frac{1}{5} + 6e^{-6x} \cdot 6e^{-6y} \frac{4}{5}\Big) \mathrm{d}x\mathrm{d}y,$$

amiről némi számolás után kiderül, hogy 0,44 (nem csak kerekítve).

[42] Az együttes eloszlásfüggvények is karakterizálhatók, de a leírásuk komplikáltabb szerkezetű, mint az egyváltozós esetben.

Hiányzó fogalom még a marginális eloszlás.

7.1.8. Definíció. Ha $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó, akkor az X_i valószínűségi változó eloszlását az $\underline{X}$ i-edik **marginális eloszlás**ának (vagy peremeloszlásának) hívjuk.

7.1.9. Állítás. *Ha $\underline{X} = (X_1, \dots, X_n)$ folytonos valószínűségi vektorváltozó, akkor X_i is folytonos, sűrűségfüggvénye:*

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(x_1, \dots, x_n) \mathrm{d}x_1 \dots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \dots \mathrm{d}x_n \qquad (\forall x_i \in \mathbb{R}).$$

Vagyis az együttes sűrűségfüggvény x_i-től eltérő változói szerinti integrálja.

7.1.10. Példa. A fenti példában (X, Y) együttes eloszlának marginálisa például az Y eloszlása, azaz

$$f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) \mathrm{d}x = \int_0^{\infty} \Big(3e^{-3x} \cdot 12e^{-12y} \frac{1}{5} + 6e^{-6x} \cdot 6e^{-6y} \frac{4}{5} \Big) \mathrm{d}x =$$

$$= \frac{12}{5} e^{-12y} \big[-e^{-3x} \big]_0^{\infty} + \frac{24}{5} e^{-6y} \big[-e^{-6x} \big]_0^{\infty} = \frac{12}{5} e^{-12y} + \frac{24}{5} e^{-6y},$$

ha $y > 0$, és 0 egyébként. Y eloszlása egy úgynevezett kevert exponenciális eloszlás.

Megjegyzés. Valószínűségi vektorváltozókról és azok együttes eloszlásfüggvényéről akkor is van értelme beszélni, ha a $X_1, \dots, X_n$ diszkrét. A sűrűségfüggvény esetében nem ez a helyzet. Sőt, ha X_1 és X_2 külön-külön folytonos, akkor sem feltétlenül van együttes sűrűségfüggvényük. Lásd az előző megjegyzésben szereplő példát.

7.2. Vektorváltozók függetlensége

Valószínűségi változók viszonyának szempontjából speciális eset, amikor a változók függetlenek. Két valószínűségi változó függetlenségét a hatodik fejezetben már definiáltuk. Bár ott elsősorban diszkrét eloszlásokról volt szó, a definíció nem feltétlenül diszkrét változókra is értelmes. Ezt általánosítja a következő definíció.

7.2.1. Definíció. Az $X_1, \dots, X_n$ valószínűségi változók **(együttesen) függetlenek**, ha az

$$\{X_1 < x_1\}, \dots, \{X_n < x_n\}$$

események függetlenek minden $x_1, \dots, x_n \in \mathbb{R}$ esetén.

A továbbiakban az „együttesen” kitételt hanyagoljuk. Ha nem együttes függetlenségről beszélünk (hanem mondjuk páronkénti függetlenségről), azt külön jelezzük.

Hogy lehet n valószínűségi változó együttes függetlenségét ellenőrizni? Diszkrét esetben ehhez az egyes $\{X_i = x_i\}$ események valószínűségét vizsgáltuk. Amikor az $\mathbb{P}(X_i = x_i)$ valószínűségek szorzata minden paraméterválasztásra megegyezett az események metszetének valószínűségével, az bizonyította a változók függetlenségét. Folytonos esetben ez nem járható út. Helyette a következőt mondhatjuk:

7.2.2. Állítás. *Az $X_1, \dots, X_n$ valószínűségi változók pontosan akkor függetlenek, ha*

$$F_{(X_1,\dots,X_n)}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n)$$

minden $x_1, \dots, x_n \in \mathbb{R}$ esetén.

A fenti állítás, bár hasznos folytonos valószínűségi változókra, de nem csak rájuk érvényes. Ugyanúgy, ahogy az eloszlásfüggvény sem csak folytonos valószínűségi változókra definiált, a fenti tétel is teljesül tetszőleges $X_1, \dots, X_n$ valószínűségi változók esetén.

A folytonos esetre visszatérve: vannak feladatok, amikor az eloszlásfüggvény nem áll rögtön rendelkezésünkre, viszont az együttes sűrűségfüggvény igen. Ilyenkor a következő állítást tudjuk használni.

7.2.3. Állítás. *Legyenek $X_1, \dots, X_n$ folytonos valószínűségi változók. Ezek pontosan akkor függetlenek, ha $(X_1, \dots, X_n)$ folytonos valószínűségi vektorváltozó, és*

$$f_{(X_1,\dots,X_n)}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n)$$

minden $x_1, \dots, x_n \in \mathbb{R}$ esetén.

Röviden, független valószínűségi változók együttes sűrűségfüggvénye szorzattá bomlik. Speciálisan, ilyen esetben létezik az együttes sűrűségfüggvény.

7.2.4. Példa. A korábbi példa feltételei mellett igaz-e, hogy Aladár válaszideje független Béláétól? Ezt az előző állítás segítségével ellenőrizhetjük. Legyen $x, y > 0$. A korábbiak szerint

$$f_{(X,Y)}(x, y) = 3e^{-3x} \cdot 12e^{-12y}\frac{1}{5} + 6e^{-6x} \cdot 6e^{-6y}\frac{4}{5} \quad \text{és} \quad f_Y(y) = \frac{12}{5}e^{-12y} + \frac{24}{5}e^{-6y},$$

illetve az f_Y-hoz hasonlóan kiszámolható, hogy

$$f_X(x) = \frac{3}{5}e^{-3x} + \frac{24}{5}e^{-6x}.$$

Tehát X és Y nem függetlenek, hiszen

$$f_X(x) \cdot f_Y(y) = \Big(\frac{3}{5}e^{-3x} + \frac{24}{5}e^{-6x}\Big)\Big(\frac{12}{5}e^{-12y} + \frac{24}{5}e^{-6y}\Big)$$

nem ugyanaz, mint

$$f_{(X,Y)}(x, y) = 3e^{-3x} \cdot 12e^{-12y}\frac{1}{5} + 6e^{-6x} \cdot 6e^{-6y}\frac{4}{5},$$

csak speciális x, y esetén, pedig a függetlenséghez az $f_X(x) \cdot f_Y(y) = f_{(X,Y)}(x, y)$ egyenlőségnek minden x, y esetén teljesülnie kellene.

Végül lássunk egy nevezetes többdimenziós folytonos eloszlást, mégpedig azt, amelynek együttes sűrűségfüggvénye a lehető legegyszerűbb alakú.

7.2.5. Példa (Többdimenziós egyenletes eloszlás)**.** Az $(a; b)$ intervallumon egyenletes eloszlású valószínűségi változót az a tulajdonság karakterizálta, hogy a valószínűségi változó egy tetszőleges $I \subseteq (a; b)$ intervallumba az intervallum hosszával arányos valószínűséggel esik. Ennek általánosítása a **kétdimenziós egyenletes eloszlás**, amely szorosan kötődik a geometriai valószínűségi mező fogalmához.

Egy (X, Y) valószínűségi vektorváltozó definíció szerint akkor kétdimenziós egyenletes eloszlású egy $H \subseteq \mathbb{R}^2$ Jordan-mérhető tartományon, ha együttes sűrűségfüggvénye H-n konstans, mindenhol máshol pedig 0. Mivel az együttes sűrűségfüggvény kettős integráljának 1-et kell adnia, ezért ez a konstans $1/T_H$-val kell, hogy megegyezzen, ahol T_H jelöli H területét. Tehát az együttes sűrűségfüggvény

$$f_{(X,Y)}(x, y) = \begin{cases} \frac{1}{T_H}, & \text{ha } (x, y) \in H, \\ 0, & \text{egyébként.} \end{cases}$$

Ha H egy koordinátatengelyekkel párhuzamos oldalakkal rendelkező téglalap, akkor könnyű belátni, hogy X és Y peremeloszlása is egyenletes eloszlású lesz (egy dimenzióban), továbbá X és Y függetlenek. Ha viszont például H egy háromszög vagy kör, akkor általában X és Y se nem függetlenek, se nem egyenletes eloszlásúak.

Egy másik ismert példa többdimenziós eloszlásokra az exponenciális peremeloszlásokkal rendelkező úgynevezett **Marshall–Olkin-féle kétváltozós exponenciális eloszlás**, amely idén nem képezi a tárgy anyagát, de az érdeklődő olvasó a függelékben szereplő A.9 fejezetben találhat róla információt.

7.3. Kovariancia folytonos esetben

Észrevehetjük, hogy folytonos valószínűségi változók kovarianciája is értelmes a kovariancia eredeti definíciójával, feltéve, hogy az ott szereplő várható értékek léteznek. Sőt, az alábbi állítás is érvényben marad:

$$\operatorname{cov}(X,Y) \overset{\text{def}}{=} \mathbb{E}\big((X-\mathbb{E}X)(Y-\mathbb{E}Y)\big) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Konkrét esetben számolási nehézséget tipikusan az $\mathbb{E}(XY)$ tag jelent, hiszen az XY valószínűségi változó eloszlása az (X,Y) valószínűségi vektorváltozó együttes eloszlásától függ, nem csak X és Y peremeloszlásaitól. A következő állítás segítségével XY eloszlásának kiszámolása nélkül is meghatározható $\mathbb{E}(XY)$.

7.3.1. Állítás. *Legyen $\underline{X} = (X_1,\dots,X_n)$ folytonos valószínűségi vektorváltozó, és legyen $g:\mathbb{R}^n\to\mathbb{R}$ olyan függvény, amire $\mathbb{E}(g(X_1,\dots,X_n))$ létezik. Ekkor*

$$\mathbb{E}(g(X_1,\dots,X_n)) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1,\dots,x_n)\cdot f_{\underline{X}}(x_1,\dots x_n)\mathrm{d}x_1\dots\mathrm{d}x_n.$$

Ha g folytonos és nemnegatív, akkor $\mathbb{E}(g(X_1,\dots,X_n))$ létezik, beleértve, hogy értéke esetleg $+\infty$.

Az állításnak speciális esete, hogy ha (X,Y) folytonos valószínűségi vektorváltozó, akkor

$$\mathbb{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\cdot f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y,$$

feltéve, hogy a várható érték létezik (ugyebár a $g:(x,y)\mapsto x\cdot y$ függvény nem nemnegatív).

7.3.2. Példa. Jelölje X az éves összes csapadékmennyiséget (1000 mm-ben számolva), Y pedig az évben eladott esernyők számát (1000 db-ban számolva). Tegyük fel, hogy az együttes sűrűségfüggvényük:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{5}(4-2x^2+xy-y^2) & \text{ha } 0<x<1 \text{ és } 0<y<2,\\ 0 & \text{egyébként.}\end{cases}$$

Számoljuk ki X és Y kovarianciáját. Az előző állítás szerint

$$\mathbb{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\cdot f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y = \int_0^2\int_0^1 xy\cdot\frac{1}{5}(4-2x^2+xy-y^2)\mathrm{d}x\mathrm{d}y =$$

$$= \frac{1}{5}\int_0^2\int_0^1 (4xy-2x^3y+x^2y^2-xy^3)\mathrm{d}x\mathrm{d}y = \frac{1}{5}\int_0^2\Big[2x^2y-\frac{1}{2}x^4y+\frac{1}{3}x^3y^2-\frac{1}{2}x^2y^3\Big]_{x=0}^{1}\mathrm{d}y =$$

$$= \frac{1}{5}\int_0^2\Big(\frac{3}{2}y+\frac{1}{3}y^2-\frac{1}{2}y^3\Big)\mathrm{d}y = \frac{1}{5}\Big[\frac{3}{4}y^2+\frac{1}{9}y^3-\frac{1}{8}y^4\Big]_0^2 = \frac{1}{5}\Big(3+\frac{8}{9}-2\Big) = \frac{17}{45}.$$

A kovarianciához szükségünk van még a várható értékekre. Annyi csak a probléma, hogy ehhez az f_X sűrűségfüggvény még nem áll rendelkezésünkre. Szerencsére azt tudjuk, hogy a peremeloszlás sűrűségfüggvénye hogyan számolható az együttes sűrűségfüggvényből:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\cdot f_X(x)\mathrm{d}x = \int_{-\infty}^{\infty} x\cdot\int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}y\mathrm{d}x = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\cdot f_{X,Y}(x,y)\mathrm{d}y\mathrm{d}x.$$

Ezen a ponton észre is vehetjük, hogy $\mathbb{E}(X)$ igazából a $g(x,y)=x$ függvény szerinti transzformált várható értéke, így hamarabb eljutunk ugyanehhez a formulához. Némi integrálással kapjuk, hogy

$$\mathbb{E}(X) = \int_0^2\int_0^1 x\cdot\frac{1}{5}(4-2x^2+xy-y^2)\mathrm{d}x\mathrm{d}y = \frac{7}{15}$$

$$\mathbb{E}(Y) = \int_0^2\int_0^1 y\cdot\frac{1}{5}(4-2x^2+xy-y^2)\mathrm{d}x\mathrm{d}y = \frac{4}{5}$$

$$\operatorname{cov}(X,Y) = \mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y) = \frac{17}{45}-\frac{7}{15}\cdot\frac{4}{5} = \frac{1}{225}\approx 0{,}0044.$$

A kovariancia illetve szórás korábban tárgyalt tulajdonságai (azonosságai) szintén teljesülnek, függetlenül attól, hogy folytonos esetről beszélünk-e vagy sem.

7.3.3. Lemma. *Legyen (X, Y, Z) valószínűségi vektorváltozó. Ekkor teljesülnek az alábbiak, feltéve, hogy a bennük szereplő mennyiségek értelmezettek:*

(1) Ha $c \in \mathbb{R}$, akkor $\mathbb{D}(X + c) = \mathbb{D}(X)$ és $\mathbb{D}(cX) = |c|\mathbb{D}(X)$.
(2) $\mathbb{D}^2(X + Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y) + 2\mathrm{cov}(X, Y)$.
(3) $\mathbb{D}^2(X) = 0$ pontosan akkor teljesül, ha $\mathbb{P}(X = c) = 1$ valamilyen $c \in \mathbb{R}$-re.
(4) Ha X és Y függetlenek, akkor $\mathrm{cov}(X, Y) = 0$, speciálisan $\mathbb{D}^2(X + Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y)$.
(5) (bilineáris) Ha $b, c \in \mathbb{R}$ akkor $\mathrm{cov}(X, bY + cZ) = b \cdot \mathrm{cov}(X, Y) + c \cdot \mathrm{cov}(X, Z)$.

Megjegyzés. A lemma 4. pontja általánosabban alkalmazható, ha felhasználjuk az alábbi lemmát.

7.3.4. Lemma. *Ha X és Y független valószínűségi változók, g és h folytonos, valós függvények, akkor $g(X)$ és $h(Y)$ is függetlenek.*

A lemma nem nyilvánvaló, de itt nem bizonyítjuk.

7.4. Kovarianciamátrix

Valószínűségi vektorváltozó esetén a szórásnégyzeteket és kovarianciákat mátrixba rendezve szokás kezelni. Ennek a motivációja nem a kompakt leírhatóság, hanem az, hogy a valószínűségi vektorváltozókkal való számolásokban természetes módon előkerül a kovarianciamátrix vektorokkal vett szorzata, a mátrix determinánsa, illetve nyoma is, lásd például a többváltozós normális eloszlást a 12. előadáson.

7.4.1. Definíció. Az $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó **kovarianciamátrix**a az alábbi $n \times n$-es valós mátrix:

$$\mathrm{cov}(\underline{X}) = \begin{pmatrix} \mathrm{cov}(X_1, X_1) & \mathrm{cov}(X_1, X_2) & \dots & \mathrm{cov}(X_1, X_n) \\ \mathrm{cov}(X_2, X_1) & \mathrm{cov}(X_2, X_2) & & \vdots \\ \vdots & & \ddots & \\ \mathrm{cov}(X_n, X_1) & \dots & & \mathrm{cov}(X_n, X_n) \end{pmatrix},$$

azaz $\mathrm{cov}(\underline{X})_{i,j} = \mathrm{cov}(X_i, X_j)$ minden $1 \le i, j \le n$ esetén.

De hol van ebben szórásnégyzet? Mivel $\mathbb{D}^2(X_1) = \mathrm{cov}(X_1, X_1)$, így a mátrix főátlójában lévő elemek a vektorváltozó koordinátáinak szórásnégyzetei.

7.4.2. Állítás. *Legyen $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó. Ekkor*

(1) $\mathrm{cov}(\underline{X})$ szimmetrikus, azaz $\mathrm{cov}(X_i, X_j) = \mathrm{cov}(X_j, X_i)$ minden $1 \le i, j \le n$ esetén.
(2) $\mathrm{cov}(\underline{X})$ pozitív szemidefinit mátrix, azaz $\sum_{i=1}^n \sum_{j=1}^n a_i \mathrm{cov}(X_i, X_j) a_j \ge 0$ minden $(a_1, \dots, a_n) \in \mathbb{R}^n$ esetén, és pontosan akkor 0, ha $\sum_{i=1}^n a_i X_i$ 1-valószínűséggel konstans.

Bizonyítás. A kovariancia szimmetrikussága a definíciója szimmetrikusságából adódik, ezt nem ragozzuk. A pozitív szemidefinitség belátását kezdjük az extrém esettel: tegyük fel, hogy $\sum_{i=1}^n a_i X_i$ 1-valószínűséggel konstans valószínűségi változó, azaz $\mathbb{P}\big(\sum_{i=1}^n a_i X_i = c\big) = 1$ valamilyen $c \in \mathbb{R}$ esetén. Az előző lemma 3-as pontja szerint ez ekvivalens azzal, hogy a valószínűségi változó szórásnégyzete 0. Továbbá, a lemma 5-ös pontja miatt

$$\mathbb{D}^2\Big(\sum_{i=1}^n a_i X_i\Big) = \mathrm{cov}\Big(\sum_{i=1}^n a_i X_i, \sum_{j=1}^n a_j X_j\Big) = \sum_{i=1}^n \sum_{j=1}^n a_i \mathrm{cov}(X_i, X_j) a_j. \tag{7.1}$$

Tehát ha a valószínűségi változó 1-valószínűséggel konstans, akkor a jobb oldalon lévő összeg is 0. Az érvelés fordított irányba ugyanígy elmondható, így az állítás „pontosan akkor” része teljesül. Az egyenlőtlenség belátásához már csak azt kell észrevennünk, hogy a szórásnégyzet nemnegatív, ezért a (7.1) egyenlet jobb oldala is mindig nemnegatív. □

7.4.3. Példa. Írjuk fel az előző példában szereplő (X,Y) valószínűségi vektorváltozó kovarianciamátrixát. Ehhez szükségünk van $\mathbb{D}^2(X)$-re és $\mathbb{D}^2(Y)$-ra is. A korábbiakkal analóg átalakításokkal, illetve polinomok integrálásával kapjuk, hogy

$$\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f_X(x,y)\mathrm{d}x\mathrm{d}y - \Big(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_X(x,y)\mathrm{d}x\mathrm{d}y\Big)^2 =$$

$$= \int_0^2\int_0^1 x^2\cdot\frac{1}{5}(4-2x^2+xy-y^2)\mathrm{d}x\mathrm{d}y - \Big(\int_0^2\int_0^1 x\cdot\frac{1}{5}(4-2x^2+xy-y^2)\mathrm{d}x\mathrm{d}y\Big)^2 = \frac{7}{90}.$$

És hasonlóan $\mathbb{D}^2(Y) = \frac{58}{225}$. Tehát ha $\underline{Z}$ jelöli az (X,Y) valószínűségi vektorváltozót, akkor

$$\mathrm{cov}(\underline{Z}) = \begin{pmatrix} \frac{7}{90} & \frac{1}{225} \\ \frac{1}{225} & \frac{58}{225} \end{pmatrix}.$$

7.4.4. Definíció. Egy $\underline{Y} = (Y_1,\dots,Y_n)$ valószínűségi vektorváltozó **várhatóérték-vektora** az

$$(\mathbb{E}Y_1,\dots,\mathbb{E}Y_n)$$

$\mathbb{R}^n$-beli vektor (amennyiben a szóban forgó várható értékek léteznek). Jelölés: $\mathbb{E}\underline{Y}$.

A kovarianciamátrix szintén kifejezhető a várhatóérték-vektor segítségével. Ha oszlopvektorokként kezeljük az $\underline{Y}$ és $\mathbb{E}\underline{Y}$ vektorokat, akkor

$$\mathrm{cov}(\underline{Y}) = \mathbb{E}\big((\underline{Y}-\mathbb{E}\underline{Y})\cdot(\underline{Y}-\mathbb{E}\underline{Y})^T\big) \in \mathbb{R}^{n\times n},$$

ahol a szorzás az $n\times 1$ és $1\times n$ alakú mátrixok mátrixszorzatát jelöli, illetve a kapott mátrix várható értékét koordinátánként értelmezzük.

Az A.1 függelékben még egy módszerrel bővítjük eszköztárunkat, amivel meghatározhatjuk két független valószínűségi változó összegének eloszlását. Ezt az eloszlást hívják az eredeti két eloszlás konvolúciójának. Első hangzásra talán nem érződik a feladat komplikáltsága, hiszen összeadni nem lehet olyan nehéz. A függelékben először kiderül, hogy ez miért mégis, majd hogy miért mégsem olyan nehéz.

A kovariancia és szórás fogalmak egyik alkalmazása a *lineáris regresszió*. Lineáris regresszió alatt elsősorban egy statisztikai modellt értünk, ami a változók közötti lineáris kapcsolatra alapozva vezet le összefüggéseket a változók viselkedésére. A modellt használják prediktív, illetve magyarázó célzattal is. Az előbbi alkalmazás a becsléselmélet, míg utóbbi a hipotézisvizsgálat témaköréhez sorolható, amik a statisztika részterületei. A témát idén előadáson nem tárgyaljuk; az érdeklődőknek a függelékben szereplő A.5 fejezetet ajánljuk.

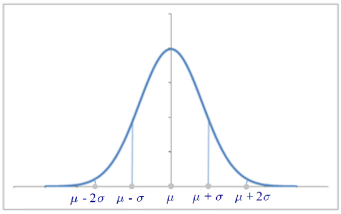
8. Normális eloszlás

Volt már szó nevezetes eloszlásokról, de a legnevezetesebbről még nem. Ez a Gauss-féle normális eloszlás, avagy Gauss-eloszlás, amely elméleti és gyakorlati szempontból is központi jelentőséggel bír.

Alkalmazásokban tipikusan akkor találkozunk vele, ha nagyszámú, független, de egymagában elhanyagolható méretű hatás eredményeként létrejövő valószínűségi változó eloszlásáról van szó, mint amilyen egy fizikai mérés eredménye. Ennek elméleti hátterét, a centrális határeloszlás tételt, a 9. előadásban tárgyaljuk.

8.1. Az eloszlás definíciója

Kezdjük azzal, miről is beszélünk, utána ráérünk megmagyarázni, miért nevezetes ez az eloszlás.[43]



8.1.1. Definíció. Egy Y valószínűségi változó **normális eloszlású** μ és σ^2 valós paraméterekkel ($\sigma^2 > 0$), ha Y folytonos valószínűségi változó, aminek sűrűségfüggvénye

$$f_Y(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (x \in \mathbb{R}).$$

Jelölése: $Y \sim N(\mu; \sigma^2)$.

Ha $\mu = 0$ és $\sigma^2 = 1$, akkor az eloszlás neve **standard normális eloszlás** (néha: sztenderd normális eloszlás). Sűrűségfüggvényét φ jelöli, azaz

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

tetszőleges $x \in \mathbb{R}$ esetén.

Érdemes rögtön ellenőrizni, hogy a fenti f_Y valóban sűrűségfüggvény. A 4. előadás szerint ehhez elég belátnunk, hogy nemnegatív, és a teljes valós egyenesen az integrálja 1. A nemnegativitás világos (hiszen e^x mindig pozitív), de az integrált ki kell számolni. Kezdjük a standard esettel.

8.1.2. Állítás. $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \mathrm{d}x = \sqrt{2\pi}$.

Bizonyítás. Az eredeti integrál helyett vizsgáljuk először a négyzetét. Ez a következőképp írható át:

$$\left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \mathrm{d}x\right)\left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} \mathrm{d}y\right) = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} \mathrm{d}y\right) \mathrm{d}x = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} \mathrm{d}y\mathrm{d}x.$$

Ez egy kettős improprius integrál. Mivel $e^{-\frac{x^2+y^2}{2}}$ sehol sem negatív, így a kettős integrálja éppen[44]

$$\lim_{R\to\infty} \iint_{B_R} e^{-\frac{x^2+y^2}{2}} \mathrm{d}x\mathrm{d}y,$$

ahol B_R jelöli az R sugarú, origó körüli zárt körlapot. Ezt pedig polárkoordinátákra való áttéréssel számolhatjuk ki: legyen $x = r\cos(\alpha)$ és $y = r\sin(\alpha)$. Ekkor

$$\iint_{B_R} e^{-\frac{x^2+y^2}{2}} \mathrm{d}x\mathrm{d}y = \int_0^{2\pi}\int_0^R e^{-\frac{r^2}{2}} r\mathrm{d}r\mathrm{d}\alpha = \int_0^{2\pi} \left[-e^{-\frac{r^2}{2}}\right]_0^R \mathrm{d}\alpha = \int_0^{2\pi} (-e^{-\frac{R^2}{2}}+1)\mathrm{d}\alpha = 2\pi(1-e^{-\frac{R^2}{2}}).$$

Ha R tart végtelenhez, akkor az $e^{-\frac{R^2}{2}}$ tart 0-hoz, így a keresett mennyiségünk négyzete 2π. Mivel az e^{-x^2} pozitív, így az integrálja is pozitív, ezért innen gyököt vonva kapjuk az állítást.[45] □

[43] A türelmetlen olvasó addig rákereshet a Galton-deszka fogalmára, lásd pl. [youtube] The Galton Board

[44] Ez a kettősintegrál alaptulajdonságaiból belátható, feltéve, hogy a függvény minden véges téglalapon integrálható.

[45] Igazából arra is szükségünk van, hogy az integrál értelmes. Ez szintén abból adódik, hogy $e^{-\frac{x^2}{2}}$ pozitív és folytonos, tehát az integrál létezik, esetleg végtelen. A fenti számolás pedig meghatározza az értékét.

8.2. Standardizálás

Az állítás szerint a standard normális eloszlás tényleg egy valószínűségi eloszlás. Mi a helyzet a többi normális eloszlással? A hozzájuk tartozó sűrűségfüggvényeket is ki lehetne integrálni, de ennél most koncepciózusabb megközelítést alkalmazunk.

8.2.1. Lemma. *Legyen $\mu, \sigma \in \mathbb{R}$, $\sigma > 0$, és legyen X folytonos valószínűségi változó, aminek sűrűségfüggvényét f_X jelöli. Ekkor $Y = \sigma X + \mu$ sűrűségfüggvénye*

$$f_Y(x) = \frac{1}{\sigma} f_X\Big(\frac{x-\mu}{\sigma}\Big) \tag{8.1}$$

tetszőleges $x \in \mathbb{R}$ esetén.

Bizonyítás. A sűrűségfüggvény definíciója szerint azt kell belátnunk, hogy $\frac{1}{\sigma} f_X\big(\frac{x-\mu}{\sigma}\big)$ nemnegatív és

$$\int_{-\infty}^{a} \frac{1}{\sigma} f_X\Big(\frac{x-\mu}{\sigma}\Big)\mathrm{d}x = F_Y(a)$$

minden $a \in \mathbb{R}$ esetén, ahol F_Y az Y eloszlásfüggvénye. A nemnegativitás világos, hiszen f_X nemnegatív és $\sigma > 0$. Az integrált pedig a következőképp számolhatjuk ki:

$$\int_{-\infty}^{a} \frac{1}{\sigma} f_X\Big(\frac{x-\mu}{\sigma}\Big)\mathrm{d}x \overset{z=\frac{x-\mu}{\sigma}}{=} \int_{-\infty}^{\frac{a-\mu}{\sigma}} \frac{1}{\sigma} f_X(z)\sigma \mathrm{d}z = F_X\Big(\frac{a-\mu}{\sigma}\Big) =$$

$$= \mathbb{P}\Big(X < \frac{a-\mu}{\sigma}\Big) = \mathbb{P}(\sigma X + \mu < a) = \mathbb{P}(Y < a) = F_Y(a),$$

felhasználva, hogy $\int_{-\infty}^{b} f_X(z)\mathrm{d}z = F_X(b) = \mathbb{P}(X < b)$ tetszőleges valós b esetén. □

8.2.2. Következmény. *Legyen $\mu, \sigma \in \mathbb{R}$ és $\sigma > 0$. Egy Y valószínűségi változó pontosan akkor $N(\mu; \sigma^2)$ eloszlású, ha létezik $X \sim N(0;1)$ amire $Y = \sigma X + \mu$.*

Más szavakal, a többi normális eloszlás nem más, mint a standard normális eloszlás lineáris függvénye (transzformáltja). Emiatt a legtöbb esetben elegendő megértenünk a standard eset működését.

Bizonyítás. Először tegyük fel, hogy $X \sim N(0;1)$ és $Y = \sigma X + \mu$. A lemma miatt f_Y kiszámolható f_X felhasználásával, ahol $f_X = \varphi$. Emiatt

$$f_Y(x) = \frac{1}{\sigma} f_X\Big(\frac{x-\mu}{\sigma}\Big) = \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} e^{-\frac{\left(\frac{x-\mu}{\sigma}\right)^2}{2}} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

vagyis Y valóban $N(\mu; \sigma^2)$ eloszlású.

Visszafelé, ha $Y \sim N(\mu; \sigma^2)$, akkor legyen $X = \frac{1}{\sigma}(Y - \mu)$. Átrendezve, $Y = \sigma X + \mu$. Ekkor ismét használhatjuk a lemmát, ami miatt tetszőleges $x \in \mathbb{R}$ esetén (8.1) egyenlet teljesül. Alkalmazva az $x = \sigma z + \mu$ helyettesítést, azt kapjuk, hogy

$$f_X(z) = \sigma f_Y(\sigma z + \mu) = \sigma \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\sigma z + \mu - \mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

tetszőleges $z \in \mathbb{R}$ esetén. Tehát X eloszlása valóban standard normális. □

Megjegyzés. A legtöbb eddig tárgyalt eloszlásra nem teljesül, hogy a különböző paraméterű verziók egymás egyszerű transzformáltjai. Például, ha egy X valószínűségi változó $\mathrm{Bin}(n; p)$ binomiális eloszlású, és emiatt az értékei a $\{0, 1, 2, \ldots, n\}$ halmazból adódnak (mindegyik pozitív eséllyel), akkor $3X$-nek nem lehet az értéke 2, így már csak emiatt sem lehet $3X$ binomiális eloszlású. Ugyanígy $\sigma X + \mu$ csak akkor lehet binomiális, ha $\sigma = 1$ és $\mu = 0$.

Az egyetlen kivételünk ez alól az exponenciális eloszlás, ugyanis ha $X \sim \mathrm{Exp}(\lambda)$, akkor $\sigma X \sim \mathrm{Exp}(\sigma^{-1}\lambda)$ tetszőleges λ, σ pozitív valósakra.

Rendben, láttuk a normális eloszlás sűrűségfüggvényét, de mi van az eloszlásfüggvényével?

Jelölés: A standard normális eloszlás eloszlásfüggvényét Φ jelöli, azaz

$$\Phi(x) = \int_{-\infty}^{x} \varphi(z)\mathrm{d}z = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\mathrm{d}z.$$

A titkolózást – mármint hogy csak integrálalakban írjuk fel az eloszlásfüggvényt – az indokolja, hogy Φ nem írható fel zárt alakban.[46] Ettől függetlenül az integrál létezik és véges minden valós x-re, sőt numerikusan is jól közelíthető. Pusztán nem tudjuk kifejezni elemi módon, olyan alakban, ami nem használ határértéket. Fontos azonosság Φ-re:

$$\Phi(-x) = 1 - \Phi(x) \qquad (\forall x \in \mathbb{R}),$$

amely tulajdonság abból adódik, hogy φ szimmetrikus a 0-ra, azaz $\varphi(x) = \varphi(-x)$.

Mit jelent a normális eloszlás definíciójában a μ és a σ?

8.2.3. Állítás. *Legyen $Y \sim N(\mu;\sigma^2)$. Ekkor $\mathbb{E}(Y) = \mu$ és $\mathbb{D}^2(Y) = \sigma^2$.*

Más szavakkal, a normális eloszlások éppen a várható értékükkel és– a szórásnégyzetükkel vannak paraméterezve. A fenti összefüggésből adódik az ún. **standardizálás** módszere is. Ez azt jelenti, hogy ha Y eloszlása normális, akkor

$$\frac{Y - \mathbb{E}Y}{\mathbb{D}(Y)} \sim N(0;1).$$

Ez a gyakorlatban is jól alkalmazható: általában egyszerűbb a valószínűségi változót transzformálni, és a standard normális sűrűségfüggvényével vagy eloszlásfüggvényével számolni, lásd még a lenti példát.

Bizonyítás. A következmény miatt létezik X, amire $Y = \sigma X + \mu$. Emiatt

$$\mathbb{E}(Y) = \mathbb{E}(\sigma X + \mu) = \sigma\mathbb{E}(X) + \mu, \quad \text{és} \quad \mathbb{D}^2(Y) = \mathbb{D}^2(\sigma X + \mu) = \mathbb{D}^2(\sigma X) = \sigma^2\mathbb{D}^2(X).$$

Tehát az állításhoz elég a standard normális eloszlású X esetére kiszámolni, hogy $\mathbb{E}(X) = 0$, illetve $\mathbb{D}^2(X) = 1$. Ez pedig valóban teljesül, hiszen a transzformált várható értékére vonatkozó állítás miatt:

$$\mathbb{E}(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}}\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \Big[-e^{-\frac{x^2}{2}} \Big]_{-\infty}^{\infty} = \frac{1}{\sqrt{2\pi}}(0-0) = 0,$$

$$\mathbb{E}(X^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}}\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \Big[x\big(-e^{-\frac{x^2}{2}}\big) \Big]_{-\infty}^{\infty} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \big(-e^{-\frac{x^2}{2}}\big)\mathrm{d}x = 0 + 1 = 1.$$

Következésképp, $\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 1 - 0^2 = 1$. □

8.2.4. Példa. Egy tartályból a gyártási folyamat végeztével mintát veszünk. Tegyük fel, hogy a minta hőmérséklete Celsius-fokban mérve $N(-2;1{,}69)$ eloszlású. Mi a valószínűsége, hogy a minta hőmérséklete nagyobb, mint $0°C$?

Jelölje a minta hőmérsékletét Y. A kérdésünk a $\mathbb{P}(Y > 0)$ mennyiség. Nézzük az Y standardizáltját:

$$X \overset{\text{def}}{=} \frac{Y+2}{\sqrt{1{,}69}} \sim N(0;1) \qquad \text{azaz} \qquad \mathbb{P}\Big(\frac{Y+2}{1{,}3} < x\Big) = \mathbb{P}(X < x) = \Phi(x) \qquad (x \in \mathbb{R}).$$

Így $\mathbb{P}(Y > 0)$-ra a következőt kapjuk:

$$\mathbb{P}(Y > 0) = 1 - \mathbb{P}(Y \le 0) = 1 - \mathbb{P}(Y < 0) = 1 - \mathbb{P}\Big(\frac{Y+2}{1{,}3} < \frac{2}{1{,}3}\Big) = 1 - \Phi\Big(\frac{2}{1{,}3}\Big),$$

ahol Φ értéke kiszámoltatható megfelelő szoftver segítségével, de egyszerűen ki is kereshető a Φ értékeit tartalmazó táblázatból. Így kapjuk, hogy $\mathbb{P}(Y > 0) \approx 0{,}0620 \approx 6\%$.

Megjegyzés. A szórást vizuálisabban is azonosíthatjuk: a sűrűségfüggvény inflexiós pontjai, vagyis azon pontok, ahol a grafikon konvexitást vált, éppen $\mu-\sigma$ és $\mu+\sigma$. Speciálisan, a standard normális eloszlás φ sűrűségfüggvénye a $[-1,1]$ intervallumon konkáv, azon kívül konvex. Ez következik abból, hogy $\varphi''(x) = (x^2-1)\varphi(x)$, így $\varphi''(x) < 0$ pontosan akkor teljesül, ha $x \in (-1,1)$.

[46] Bővebben lásd: math.stanford.edu/∼conrad/papers/elemint.pdf

A normális eloszlás további szép tulajdonsága, hogy független normális eloszlású valószínűségi változók összege normális, és a paramétereik összeadódnak. Hasonló tulajdonság igaz még például a Poisson-eloszlásra is (lásd a függelékben szereplő A.1. fejezetet). Ezt az állítást a matematikai statisztikában lépten-nyomon használjuk (lásd például u-próba).

8.2.5. Állítás. *Legyenek $X \sim N(\mu_1; \sigma_1^2)$ és $Y \sim N(\mu_2; \sigma_2^2)$ függetlenek (ahol $\mu_1, \mu_2 \in \mathbb{R}$ és $\sigma_1, \sigma_2 > 0$). Ekkor $X + Y \sim N(\mu_1 + \mu_2; \sigma_1^2 + \sigma_2^2)$.*

Ebben az állításban az nem meglepő, hogy $X+Y$ várható értéke $\mu_1+\mu_2$ és szórásnégyzete $\sigma_1^2+\sigma_2^2$, hiszen a várható érték és szórásnégyzet azonosságai alapján

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y) = \mu_1 + \mu_2$$

és

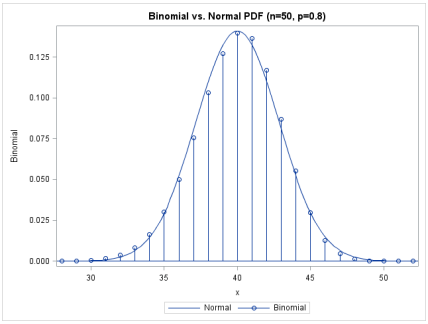
$$\mathbb{D}^2(X+Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y) = \sigma_1^2 + \sigma_2^2$$

(ahol az első egyenlőségnél felhasználtuk, hogy X és Y függetlenek). Ebből azonban még nem következik, hogy $X+Y$ normális eloszlású. Az utóbbi állítást olyan módszerekkel bizonyíthatjuk, amelyek jelenleg nem képezik a tárgy anyagát és ezért ebben a jegyzetben csak függelékben szerepelnek. A bizonyítás lehetséges konvolúcióval (lásd a A.1. fejezetet) vagy a momentumgeneráló függvény (lásd a A.4. fejezetet) segítségével.

A normális eloszlásnak létezik többdimenziós analógja is, amelyről az érdeklődő olvasó a függelék A.10. fejezetében találhat rövid összefoglalót.

8.3. De Moivre–Laplace-tétel

Miért akarna valaki épp egy ilyen eloszlást használni? Empirikus tapasztalat, hogy a normális eloszlás sok esetben jó közelítést jelent mérési eredmények eloszlására. Ilyen például egy ország lakosságának magassága vagy születési súlya, vagy a napi átlaghőmérséklet az év egy adott hónapjában, stb. A közös ezekben, hogy az eredményét rengeteg apró, egymástól többékevésbé független faktor befolyásolja.



A következő tétel azt állítja, hogy nagy n esetén a binomiális eloszlás közelíthető a normális eloszlással.[47]

8.3.1. Tétel. *Legyen $p \in (0,1)$ és $S_n \sim \mathrm{Bin}(n;p)$. Ekkor minden $a < b$ valós számokra*

$$\lim_{n\to\infty} \mathbb{P}\left(a < \frac{S_n - \mathbb{E}(S_n)}{\mathbb{D}(S_n)} < b\right) = \int_a^b \varphi(x)\mathrm{d}x = \Phi(b) - \Phi(a),$$

ahol a binomiális eloszlásról tanultak szerint $\mathbb{E}(S_n) = np$ és $\mathbb{D}(S_n) = \sqrt{np(1-p)}$.

A tétel egyik érdekessége, hogy hiába $p \neq \frac{1}{2}$, és így a binomiális eloszlás nem szimmetrikus, a megfelelően lenormált változó határeloszlása mégis 0-ra szimmetrikus sűrűségfüggvényű.

8.3.2. Példa. A matematikusok 31,4% százaléka szandált hord. Száz találomra választott matematikust nézve, közelítőleg mi az esélye, hogy kevesebb, mint 25 pár szandált találunk rajtuk. (A szituációt némileg idealizálva, tegyük fel, hogy egy matematikuson vagy egy teljes pár szandál van, vagy nincs rajta szandál.)

[47] Ettől nem független tény, hogy a normális eloszlás a statisztika szempontjából is központi jelentőségű, lásd pl. [youtube] Crash course statistics - The shape of the data: distributions

Jelölje S_n a szandálok számát, ahol $n = 100$ és $p = 0{,}314$. Kiszámolható, hogy $\mathbb{E}S_n = 31{,}4$ és $\mathbb{D}(S_n) = \sqrt{100p(1-p)} = 4{,}6412$. A fenti tétel szerint $X = (S_n - 31{,}4)/4{,}6412$ közelítőleg standard normális eloszlású. Emiatt

$$\mathbb{P}(S_n < 25) = \mathbb{P}\left(\frac{S_n - 31{,}4}{4{,}6412} < \frac{25 - 31{,}4}{4{,}6412}\right) = \mathbb{P}(X < -1{,}3790) \approx$$

$$\approx \Phi(-1{,}3790) = 1 - \Phi(1{,}3790) \approx 0{,}0839 \approx 8\%.$$

A tétel segítségével így megspórolhattuk a binomiális eloszlás tagjaival történő fáradtságos számolást. Hogy ez a közelítés mennyire jó, arról további tételek rendelkeznek.[48]

Megjegyzés. Lehet némi déjà vu érzésünk: a Poisson-eloszlás esetében is elmondtuk, hogy a binomiális eloszlás a Poisson-eloszláshoz tart, ha $n \to \infty$. Most meg azt mondjuk, hogy a normálishoz. Nincs itt ellentmondás? Nincs, ugyanis másféle paraméterezés esetén kapjuk egyik, illetve másik határértéket.

A mostani helyzetben p rögzített, $n \to \infty$, és az eloszlás standardizálva van (S_n helyett $\frac{S_n - \mathbb{E}S_n}{\mathbb{D}(S_n)}$ határeloszlását nézzük), míg a Poisson-eloszlás esetében azzal a feltétellel dolgoztunk, hogy $p \to 0$ és $n \to \infty$, de olyan ütemben, hogy $np \to \lambda$.

A tételt nem bizonyítjuk. Az érdeklődőknek a függelékben szereplő A.3 fejezetet ajánljuk, ahol körbejárjuk a tételben szereplő egyenlet bal oldalán álló kifejezést, és hogy miért épp a normális eloszlás jelenik meg a jobb oldalon. (A gyakorlat szempontjából a fenti tétel és példa a lényeges, nem a most következő vázlatos érvelés.)

[48] Lásd még: Berry-Esseen-típusú egyenlőtlenségek, pl. arxiv.org/abs/1111.6554

9. Határeloszlás-tételek

A korábbi fejezetek során már többször előkerült az ún. határeloszlások témaköre: megnéztük, hogy a Poisson-eloszlás valószínűségeit közelítik a binomiális eloszlás valószínűségei, megfelelő paraméterezés esetén, illetve a normális eloszlásnál szó esett a de Moivre–Laplace-tételről.[49] De egyáltalán mit jelent az, hogy eloszlások „tartanak" egy másik eloszláshoz? És csak speciális alakú eloszlások esetén lehet határeloszlást bizonyítani? A mostani előadásban erről is szó lesz.

9.1. Nagy számok törvénye

A köznyelvben a nagy számok törvényét sokszor abban az értelemben használjuk, hogy ha valamit nagyon sokszor próbálgatunk, akkor előbb-utóbb sikerülni fog. Ez az állítás (illetve a precíz megfogalmazása) speciális esete annak, amit a valószínűségszámításban nagy számok törvényének hívunk. Valójában a „törvény" ennél általánosabb, ugyanis nem csak valószínűségekről beszél, hanem várható értékről is.

Fogalmazzuk meg először a fenti köznyelvi igazságot precízen. Legyenek $A_1, \dots, A_n$ együttesen független események, amelyek valószínűsége egységesen p, ahol $0 < p < 1$. Ezek az események felelnek meg annak, hogy n-szer próbálgatjuk ugyanazt a kísérletet. Az állítás olyasmi, hogy elég nagy n esetén biztosan sikerül valamelyik, még akkor is, ha p nagyon kicsi. Precízen megfogalmazva,

$$\lim_{n\to\infty} \mathbb{P}(A_1 \cup \dots \cup A_n) \to 1.$$

Vegyük észre, hogy valójában a valószínűség fogalmának bevezetésénél már ennél többet is elfogadtunk: azt az intuitív igazságot, hogy független kísérletek esetén a sikerek aránya épp a siker valószínűségéhez közelít (és emiatt elég sok kísérlet esetén nem is lesz nulla). Ez szintén a nagy számok törvényének speciális esete.

Hogy az utóbbit formulával is felírhassuk, vezessük be a következő jelölést: legyen $X_i = 1$ ha A_i teljesül, egyébként 0 (azaz $X_i = \mathbf{1}_{A_i}$). Jelölje $\overline{X_n}$ az $\frac{X_1+\dots+X_n}{n}$ átlagot. Ekkor a „korábban elfogadott intuitív igazság" azt állítja, hogy ha $n \to \infty$, akkor

$$\overline{X_n} \to p.$$

A probléma az, hogy ez az állítás nem precíz: milyen értelemben tud valószínűségi változók egy sorozata konvergálni? Ahogy az a következő tételből kiderül, többféleképp.

9.1.1. Tétel. *Legyen $X_1, X_2, \dots, X_n, \dots$ független, azonos eloszlású valószínűségi változók egy sorozata. Tegyük fel, hogy $\mathbb{E}X_n = \mu$ és $\mathbb{D}(X_n) = \sigma$ minden n-re, ahol $\mu, \sigma \in \mathbb{R}$ rögzített. Jelölje $\overline{X_n}$ az $\frac{X_1+\dots+X_n}{n}$ átlagot.*[50]

- ***Nagy számok gyenge törvénye:*** *Tetszőleges $\varepsilon > 0$ esetén*

$$\lim_{n\to\infty} \mathbb{P}\Big(\Big|\overline{X_n} - \mu\Big| \geq \varepsilon\Big) = 0.$$

- ***Nagy számok erős törvénye:***

$$\mathbb{P}\Big(\lim_{n\to\infty} \overline{X_n} = \mu\Big) = 1,$$

ahol a határérték úgy értendő, mint az $\omega \mapsto \overline{X_n}(\omega)$ függvények pontonkénti határértéke.

Ahogy az a tételből is kiderül, a nagy számok törvénye nem egyetlen tétel, hanem egy ún. tételkör, adott témájú állítások egy halmaza, különböző feltételekkel.

Az állítás annyiban általánosabb az előtte szereplő diszkussziónál, hogy X_i nem feltétlenül $\{0,1\}$-értékű, hanem bármilyen egyéb eloszlású is lehet. Ha X_i $\{0,1\}$-értékű, akkor a gyenge törvény neve: **Bernoulli-féle nagy számok gyenge törvénye**.

[49] Szemfüles olvasó felfedezheti, hogy a geometriai eloszlásról is megemlítettük, hogy a pontok sűrítése esetén az exponenciális eloszláshoz tart.

[50] A tételek a szórás végessége nélkül is teljesülnek, amennyiben $\mathbb{E}(|X_n|) < \infty$, de a bizonyításuk komplikáltabb.

A gyenge törvényt az úgynevezett *Csebisev-egyenlőség* segítségével könnyű bizonyítani, azonban ezt az egyenlőtlenséget idén előadáson nem tanuljuk, így a bizonyítással együtt csak függelékben adjuk meg, lásd a A.2 fejezetet.[51]

Mitől gyenge törvény az első, és erős a második? És egyáltalán, mi a különbség a kettő közt? Egyrészt, az erős törvényből következik a gyenge (ezt nem bizonyítjuk). Másrészt, a két állítás abban különbözik, hogy milyen értelemben nézzük az $\overline{X_n}$ konvergenciáját. A gyenge törvényben szereplő konvergencia neve *sztochasztikus konvergencia*, míg az erős törvényben *1-valószínűségű konvergencia*ról beszélünk.

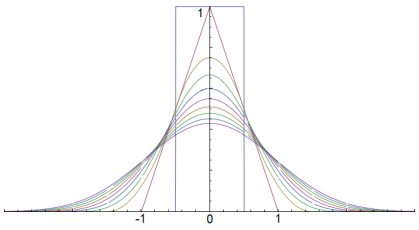
Rendben, de mit jelent az, hogy „más a konvergencia"? Van itt valami kézzelfogható eltérés, vagy ez csak valami elméleti-technikai kötözködés? És egyébként is, ha igaz az erős törvény, minek egyáltalán beszélni a gyengéről? Ezek jogos kérdések, haladjunk sorban.

A valószínűségszámítást néha azért nehéz összevetni a valósággal, mert míg a valószínűségek az összes kimenetelről beszélnek egyszerre, addig a valóságban mi (egyszerre) csak egyetlen kimenetelt látunk, egy ω esetén felvett értéket tapasztalunk. A gyenge törvény (és általában a sztochasztikus konvergencia), csak arról beszél, hogy azon ω-k, ahol az átlagtól való eltérés lényeges (azaz ε-nál nagyobb), egyre kevesebben vannak (sőt, részarányuk nullához tart). De ettől még megtörténhetne, hogy akármilyen ω-t is választunk, egyre-másra felbukkan egy olyan lépés, ahol $\overline{X_n}$ nagyon eltér μ-től. Az erős törvény éppen azt mondja, hogy ez lehetetlen. Vagyis 1 valószínűséggel olyanok az ω-k, hogy egy idő után már nem megyünk μ-től ε-nál távolabb.

Ennek fényében még jogosabb a kérdés: minek beszélünk egyáltalán az „elavult" gyenge törvényről? Azért, mert gyengébb feltételek esetén megeshet, hogy a gyenge tétel teljesül, de az erős már nem. Részletesebben: mind a függetlenség, mind az azonos eloszlásúság nagyon erős feltételek, amik a gyakorlatban sem mindig teljesülnek. (Például, ha elég ideig méregetjük a napi középhőmérsékleteket, azok nem lesznek függetlenek, továbbá előbb-utóbb nyár után tél lesz, legalábbis reméljük.) Szerencsére ezeket a feltételeket gyengítve is bizonyíthatók nagy számok törvényei. Viszont van olyan feltétel-gyengítés, ahol a gyenge törvény még igaz, de az erős már nem.

9.2. Centrális határeloszlás-tétel

A nagy számok törvénye csak arról beszél, hogy mihez konvergálunk (a várható értékhez, ugyebár), de arról nem, hogy ezt milyen gyorsan tesszük. Más szavakkal, arról nem mond semmit, hogy a várható értéktől való eltérés milyen ütemben csökken, n növekedésének függvényében.



Ahhoz, hogy ezt megválaszoljuk, szükségünk van az eloszlásbeli konvergencia fogalmára, ugyanis az átlagtól való (egyre kisebb) eltérés már nem egy konkrét szám, hanem egy (egyre koncentráltabb) eloszlás a várható érték körül.

9.2.1. Definíció. Legyen $X_1, X_2, \ldots, X_n, \ldots$ valószínűségi változók egy sorozata, jelölje F_{X_n} az X_n eloszlásfüggvényét, és hasonlóan F_Z a Z eloszlásfüggvényét. Az $(X_n)_{n\in\mathbb{N}}$ sorozat **eloszlásban konvergál** egy Z valószínűségi változóhoz, ha

$$F_{X_n}(x) \to F_Z(x) \qquad (n \to \infty)$$

minden olyan $x \in \mathbb{R}$ esetén, ahol F_Z folytonos.[52] Jelölése: $X_n \xrightarrow{d} Z$.

[51] Az erős törvény bizonyítása némi mértékelméleti alapozást és több mint egy egész előadást igényel. Az elszánt érdeklődőknek a BME-TTK Valószínűségszámítás 2 nevű tárgyát ajánljuk.

[52] Bebizonyítható, hogy az eloszlásbeli konvergencia fenti definíciójával ekvivalens a következő feltétel: $\lim_{n\to\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X)$ minden $f : \mathbb{R} \to \mathbb{R}$ folytonos, korlátos függvény esetén.

9.2.2. Tétel (Centrális határeloszlás-tétel)**.** *Legyen $X_1, X_2, \dots, X_n, \dots$ független, azonos eloszlású valószínűségi változók egy sorozata. Tegyük fel, hogy $0 < \mathbb{D}^2 X_1 < \infty$. Legyen $Z \sim N(0;1)$. Ekkor*

$$\frac{X_1 + \dots + X_n - n\mathbb{E}(X_1)}{\sqrt{n}\mathbb{D}(X_1)} \xrightarrow{d} Z \qquad (n \to \infty).$$

Az eloszlásbeli konvergencia definícióját kibontva ez azt jelenti, hogy a bal oldal eloszlásfüggvénye tart Z eloszlásfüggvényéhez, azaz Φ-hez, a Φ minden folytonossági pontjában. Mivel Φ mindenhol folytonos, így a tétel következtetése a következő:

9.2.3. Következmény (Centrális határeloszlás-tétel, valószínűségekkel megfogalmazva)**.** *Legyen X_1, $X_2, \dots, X_n, \dots$ független, azonos eloszlású valószínűségi változók egy sorozata. Tegyük fel, hogy $0 < \mathbb{D}^2 X_1 < \infty$. Ekkor*

$$\mathbb{P}\left(\frac{X_1 + \dots + X_n - n\mathbb{E}(X_1)}{\sqrt{n}\mathbb{D}(X_1)} < a\right) \to \Phi(a)$$

minden $a \in \mathbb{R}$ esetén, ha $n \to \infty$.

Az utóbbi alakból már jobban látható, hogy a centrális határeloszlás-tétel a de Moivre–Laplace-tétel általánosítása. Valóban, hiszen itt olyan független, azonos eloszlású valószínűségi változókról beszélünk, amik véges szórásúak, míg a de Moivre–Laplace-tétel csak független indikátor valószínűségi változók összegéről (azaz egy binomiális eloszlású valószínűségi változóról) állít konvergenciát. De az állítás mindkét esetben az, hogy az eloszlásfüggvény Φ-vel közelíthető. Hogy ez a közelítés mennyire jó, arról külön tételek rendelkeznek.[53]

Megjegyzés. A tételben szereplő szumma más formában is írható:

$$\frac{X_1 + \dots + X_n - n\mathbb{E}(X_1)}{\sqrt{n}\mathbb{D}(X_1)} = \frac{\overline{X_n} - \mathbb{E}(\overline{X_n})}{\mathbb{D}(\overline{X_n})},$$

ami átrendezéssel és a szórás tulajdonságainak felhasználásával kapható. Más szavakkal, a tételben az $\overline{X_n}$ standardizáltjáról van szó.

9.2.4. Példa. Egy pincészetben munkanapokon átlagosan 100 liter bort mérnek ki, 20 szórással. Tegyük fel, hogy az egyes napok mérései függetlenek és azonos eloszlásúak. Az évből hátralévő 50 munkanap alatt 4750 liter bort kellene eladniuk ahhoz, hogy felülmúlják a tavalyi teljesítményt. Mi annak a valószínűsége, hogy ez sikerül?

Jelölje az egyes napok eredményeit $X_1, X_2, \dots, X_{50}$. A feltételek szerint $\mathbb{E}(X_1) = 100$ és $\mathbb{D}(X_1) = 20$. A centrális határeloszlás-tétel szerint a megfelelően standardizált összeg közelítőleg normális, azaz

$$\mathbb{P}\Big(\sum_{i=1}^{50} X_i \geq 4750\Big) = 1 - \mathbb{P}\left(\frac{\sum_{i=1}^{50} X_i - 50 \cdot 100}{\sqrt{50} \cdot 20} < \frac{4750 - 5000}{\sqrt{50} \cdot 20}\right) \approx 1 - \Phi\left(\frac{4750 - 5000}{\sqrt{50} \cdot 20}\right)$$
$$= 1 - \Phi(-1{,}7678) = \Phi(1{,}7678) \approx 0{,}9615.$$

Tehát a siker valószínűsége nagyjából 96%.

A bizonyítás ötletéről idén előadáson nem lesz szó, az érdeklődőknek a A.4 fejezetet ajánljuk a függelékben.

[53] Lásd Berry-Esseen tétel, [Terence Tao] Central limit theorem, Theorem 23

függelék A.

A.1. Konvolúció

Amióta bevezettük a várható érték fogalmát, a legsűrűbben használt tulajdonsága, hogy additív, azaz $\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$. Ez teljesült akkor is, ha X és Y függetlenek, de akkor is, ha nem. Ez alapján azt is gondolhatnánk, hogy akkor az $X+Y$ valószínűségi változó eloszlásának meghatározása is magától értetődő feladat. Ha mást nem, legalább akkor, ha X és Y függetlenek. Nézzünk néhány példát, hogy ez mennyire nincs így.

A.1.1. Példa.

(1) Legyen $X \sim \mathrm{Bin}(n;p)$ binomiális eloszlású. Tudjuk, hogy ekkor X független indikátor valószínűségi változók összege, azaz olyan valószínűségi változóké, amik csak 1 vagy 0 értéket vehetnek fel. Látható, hogy hiába a lehető legegyszerűbb a kiindulási változók eloszlása, amik még függetlenek is, X eloszlása náluk lényegesen összetettebb.[54]

(2) Legyenek X és Y független, $\mathrm{Geo}(p)$ eloszlású valószínűségi változók, és $Z = X+Y$ (ami például egy központba beérkező második telefonhívásig eltelt másodpercek számát jelöli). Ekkor $X+Y$ nem geometriai, hanem úgynevezett negatív binomiális eloszlású, $r=2$ rend-paraméterrel.

(3) Legyenek U és V független, egyenletes eloszlású valószínűségi változók a $[0,1]$ intervallumon. Mi $U+V$ eloszlása? Elsőre talán rámondanánk, hogy egyenletes a $[0,2]$ intervallumon. Ez viszont nem igaz, mégpedig ugyanazon okból nem, amiért két kockadobás összege sokkal gyakrabban lesz 7, mint 12. Az eredmény neve Irwin–Hall-eloszlás, $n=2$ paraméterrel.

A.1.2. Definíció. Legyenek X és Y független valószínűségi változók. Ekkor $X+Y$ eloszlását az X és Y eloszlásai **konvolúció**jának hívjuk.

A.1.3. Állítás. *Legyenek X és Y független, diszkrét valószínűségi változók, amik értékei nemnegatív egész számok. Ekkor*

$$\mathbb{P}(X+Y=k) = \sum_{i=0}^{k} \mathbb{P}(X=i)\mathbb{P}(Y=k-i)$$

minden $k \in \mathbb{N}$ esetén.

Bizonyítás. A bizonyításhoz csak a valószínűség additivitását kell felhasználjuk:

$$\mathbb{P}(X+Y=k) = \mathbb{P}\Big(\bigcup_{i+j=k} \{X=i\} \cap \{Y=j\}\Big) =$$

$$= \sum_{i=0}^{k} \mathbb{P}\Big(\{X=i\} \cap \{Y=k-i\}\Big) = \sum_{i=0}^{k} \mathbb{P}(X=i)\mathbb{P}(Y=k-i).$$

□

A.1.4. Példa. Legyenek X és Y független valószínűségi változók, amire $X \sim \mathrm{Pois}(\lambda)$ és $Y \sim \mathrm{Pois}(\mu)$. Mi $X+Y$ súlyfüggvénye? A fenti állítás szerint

$$\mathbb{P}(X+Y=k) = \sum_{i=0}^{k} \mathbb{P}(X=i)\mathbb{P}(Y=k-i) = \sum_{i=0}^{k} \frac{\lambda^i}{i!} e^{-\lambda} \frac{\mu^{k-i}}{(k-i)!} e^{-\mu} =$$

$$= e^{-(\lambda+\mu)} \sum_{i=0}^{k} \frac{k!}{k! \cdot i! \cdot (k-i)!} \lambda^i \mu^{k-i} = e^{-(\lambda+\mu)} \frac{1}{k!} (\lambda+\mu)^k,$$

felhasználva a binomiális tételt. Vegyük észre, hogy az eredmény $\mathrm{Pois}(\lambda+\mu)$ eloszlású.

[54] Aki úgy gondolná, hogy a binomiális eloszlás nem túl összetett, az megpróbálhatja meggyőzni magát a de Moivre–Laplace-tétel segítségével.

A konvolúció kiszámolása folytonos esetben valamivel komolyabb feladat. Itt a fentebb összegzett i és $k-i$ párok kontinuum sokan vannak, így a szummát integrál, a valószínűségi változó eloszlását pedig a sűrűségfüggvény helyettesíti.

A.1.5. Állítás. *Legyenek X és Y független, folytonos valószínűségi változók. Ekkor a*

$$z \mapsto \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) \mathrm{d}x$$

függvény az $X+Y$ sűrűségfüggvénye.[55]

A.1.6. Példa. Legyenek X és Y független, $\mathrm{Exp}(\lambda)$ eloszlású valószínűségi változók. Az előző állítás szerint

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) \mathrm{d}x =$$

$$= \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)} \mathrm{d}x = \lambda^2 e^{-\lambda z} \int_0^z 1 \mathrm{d}x = \lambda^2 e^{-\lambda z} z$$

minden $z > 0$ esetén. Az így kapott eloszlás neve gamma-eloszlás.

A.2. Csebisev-egyenlőtlenség és a nagy számok gyenge törvényének bizonyítása

Ebben a függelékben olyan egyenlőtlenségeket nézünk, amelyekkel egy valószínűségi változó eloszlásának „széleinek” valószínűségei becsülhetők. Ezek az egyenlőtlenségek hasznos segédeszközök határeloszlás-tételek bizonyításában is. Ezután a valószínűségszámítás két alaptételéről fogunk beszélni: a nagy számok törvényéről, és a centrális határeloszlás-tételről.

Egy valószínűségi változó eloszlása nem feltétlenül ismert egy probléma esetében, különösen igaz ez a gyakorlatra. Igény viszont lenne rá, hogy ilyenkor is meg tudjuk becsülni (például felülről), mekkora valószínűséggel lesz az X értéke szélsőséges.

A.2.1. Állítás (Markov-egyenlőtlenség)**.** *Legyen X nemnegatív értékű valószínűségi változó. Ekkor minden $a > 0$ esetén*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Bizonyítás. Definiáljuk a következő valószínűségi változót: legyen $Y = a$ ha $X \geq a$, és 0 egyébként. Más szavakkal, $Y = a\mathbf{1}_{\{X \geq a\}}$. Mivel X nemnegatív, így $Y \leq X$ a teljes eseménytéren. Emiatt $\mathbb{E}(Y) \leq \mathbb{E}(X)$ is teljesül, a várható érték monotonitásából adódóan (lásd 3.2.9. Következmény). Tehát

$$\mathbb{E}(X) \geq \mathbb{E}(Y) = 0 \cdot \mathbb{P}(Y = 0) + a \cdot \mathbb{P}(Y = a) = a \cdot \mathbb{P}(X \geq a).$$

Az egyenlőtlenség átrendezéséből következik az állítás. □

A Markov-egyenlőtlenség önmagában nem feltétlenül erős becslés. Például, ha $a < \mathbb{E}(X)$, akkor annyit állít, hogy egy valószínűség kisebb, mint egy 1-nél nagyobb szám, ami azért nem egy mély tény. Tipikus alkalmazása ehelyett, amikor a jóval nagyobb $\mathbb{E}(X)$-nél. Ekkor intuitívan azt állítja, hogy annak a valószínűsége, hogy az X változó a-nál szélsőségesebb értéket vesz fel, legalább olyan ütemben csökken, mint az $x \mapsto \frac{c}{x}$ függvény, ahol $c = \mathbb{E}(X)$.

A Markov-egyenlőtlenség helyett gyakrabban alkalmazzák az alábbi következményét:

A.2.2. Következmény (Csebisev-egyenlőtlenség[56])**.** *Legyen X valószínűségi változó, amire $\mathbb{D}^2(X)$ véges. Ekkor minden $a > 0$ esetén*

$$\mathbb{P}\big(|X - \mathbb{E}(X)| \geq a\big) \leq \frac{\mathbb{D}^2(X)}{a^2}.$$

[55] Mivel ez a teljes alfejezet már csak kiegészítő anyag, az ennél az állításnál szereplő felkiáltójelet is töröltük.

[56] Angol szakirodalomban: Chebyshev's inequality

Bizonyítás. A bal oldal átrendezése után alkalmazzuk a Markov-egyenlőtlenséget az X helyett az $\big(X-\mathbb{E}(X)\big)^2$ valószínűségi változóra (és a helyett a^2-re):

$$\mathbb{P}\big(|X-\mathbb{E}(X)|\geq a\big)=\mathbb{P}\Big(\big(X-\mathbb{E}(X)\big)^2\geq a^2\Big)\leq\frac{\mathbb{E}\Big(\big(X-\mathbb{E}(X)\big)^2\Big)}{a^2}=\frac{\mathbb{D}^2(X)}{a^2}.$$

felhasználva, hogy $\big(X-\mathbb{E}(X)\big)^2$ mindig nemnegatív. Ez épp a bizonyítandó állítás. □

Vegyük észre, hogy míg a Markov-egyenlőtlenség csak nemnegatív valószínűségi változókra érvényes, addig a Csebisev-egyenlőtlenség tetszőleges valós esetben is igaz.

A.2.3. Példa. Egy adott adatbázis átlagosan 50 lekérést fogad egy időegység alatt. A lekérések számának szórása a tapasztalatok szerint 5.[57] Adjunk alsó becslést annak a valószínűségére, hogy 40-nél több, de 60-nál kevesebb lesz a lekérések száma egy időegység alatt.

Jelölje X az egy időegység alatti lekérések számát. Ekkor a Csebisev-egyenlőtlenség szerint

$$\mathbb{P}(40<X<60)=\mathbb{P}(|X-50|<10)=1-\mathbb{P}(|X-50|\geq 10)\geq 1-\frac{\mathbb{D}^2(X)}{a^2}=1-\frac{5^2}{10^2}=\frac{3}{4}.$$

Vegyük észre, hogy ez a korlát nem igényelt semmilyen feltételt az eloszlásra, a várható értékének és szórásának ismeretén felül.

A Markov-egyenlőtlenség további, erősebb becslések alapjául is szolgálhat, amennyiben a négyzetre emelés helyett egyéb függvényt alkalmazunk, illetve X-re erősebb feltételeket teszünk.

A.2.4. Következmény (Paraméteres Csernov-egyenlőtlenség[58])**.** *Legyen X valószínűségi változó. Ekkor minden $a,t>0$ esetén*[59]

$$\mathbb{P}(X\geq a)\leq\frac{\mathbb{E}(e^{tX})}{e^{ta}}.$$

Bizonyítás. Tetszőleges $t>0$ esetén az $x\mapsto e^{tx}$ függvény monoton növő. Így a Markov-egyenlőtlenség miatt

$$\mathbb{P}\big(X\geq a\big)=\mathbb{P}\big(e^{tX}\geq e^{ta}\big)\leq\frac{\mathbb{E}(e^{tX})}{e^{ta}},$$

felhasználva, hogy e^{tX} és e^{ta} mindig pozitív. □

A.2.5. Példa. Legyen $X\sim\mathrm{Pois}(5)$. Adjunk felső becslést $\mathbb{P}(X\geq 10)$-re. Ekkor a Csernov-egyenlőtlenség szerint

$$\mathbb{P}(X\geq 10)\leq\frac{\mathbb{E}(e^{tX})}{e^{10t}}=e^{-10t}\sum_{k=0}^{\infty}e^{tk}\frac{5^k}{k!}e^{-5}=e^{-10t}\sum_{k=0}^{\infty}\frac{(5e^t)^k}{k!}e^{-5}=e^{-10t+5e^t-5}.$$

Ez a felső becslés bármilyen pozitív t választása esetén igaz, így akkor is, ha a kitevő a lehető legkisebb. Deriválással megkapható, hogy a függvény minimuma $t=\ln(2)$ helyen van, értéke: $e^5/1024\approx 0{,}1449$. Vegyük észre, hogy ezzel szemben a Markov-egyenlőtlenség csak $\mathbb{E}(X)/10=\frac{1}{2}$ értékkel becsülne felülről.

Végezetül a Csebisev-egyenlőtlenség segítségével bebizonyítjuk a nagy számok gyenge tételét.

---

[57] Hogy hogyan lehet a tapasztalat alapján, például ismételt kísérletekkel információt szerezni az eloszlás szórásáról, az a statisztika témakörébe eső nem nyilvánvaló kérdés, amire most nem térünk ki.

[58] Angol szakirodalomban: Chernoff bound

[59] Előfordulhat, hogy $\mathbb{E}(e^{tX})=\infty$, ekkor az egyenlőtlenség az „∞/pozitív konstans $=\infty$” konvencióval triviálisan teljesül, de haszna nincsen.

Nagy számok gyenge törvényének bizonyítása. A várható érték linearitása miatt az $\overline{X_n}$ átlag várható értéke is μ. Így alkalmazhatjuk a Csebisev-egyenlőtlenséget az állításban szereplő valószínűségre:

$$\mathbb{P}\big(\big|\overline{X_n}-\mu\big|\geq\varepsilon\big)=\mathbb{P}\Big(\big|\overline{X_n}-\mathbb{E}(\overline{X_n})\big|\geq\varepsilon\Big)\leq\frac{\mathbb{D}^2(\overline{X_n})}{\varepsilon^2}.$$

Mivel $X_1,\dots,X_n$ függetlenek, így a $\mathbb{D}^2$ tulajdonságai szerint:

$$\mathbb{D}^2(\overline{X_n})=\mathbb{D}^2\left(\frac{X_1+\dots+X_n}{n}\right)=\frac{1}{n^2}\big(\mathbb{D}^2(X_1)+\dots+\mathbb{D}^2(X_n)\big)=\frac{n}{n^2}\sigma^2,$$

ahol a jobb oldal 0-hoz tart $n\to\infty$ esetén. Ebből az állítás már következik. □

A.3. Kitekintés: heurisztika a de Moivre–Laplace tételhez

Jelölje h az $\frac{1}{\sqrt{np(1-p)}}$ számot. Vizsgáljuk meg a következő valószínűségi változó eloszlását:

$$\frac{S_n-\mathbb{E}(S_n)}{\mathbb{D}(S_n)}=h(S_n-np),\qquad\text{ahol}$$

$$\operatorname{Ran}\Big(h(S_n-np)\Big)=\big\{h(k-np)\mid k=0,1,2,\dots,n\big\},$$

hiszen S_n binomiális. Vegyük észre, hogy minél nagyobb n, annál kisebb h, vagyis a fenti valószínűségi változó annál sűrűbben veszi fel az értékeit. Annak a valószínűsége, hogy épp $h(k-np)$ az értéke, $\binom{n}{k}p^k(1-p)^{n-k}$.

Készítsünk $h(S_n-np)$ imént végiggondolt eloszlásából egy szakaszonként lineáris, sűrűségfüggvényre emlékeztető függvényt a következőképp. Definiáljuk az $f_n:\mathbb{R}\to\mathbb{R}$ függvényt úgy, hogy minden $k\in\{0,1,2,\dots,n\}$ esetén

$$f_n\big(h(k-np)\big)=\frac{1}{h}\binom{n}{k}p^k(1-p)^{n-k},$$

továbbá két szomszédos $h(k-np)$ alakú pont közt legyen f_n lineáris, a pontokat tartalmazó intervallumon kívül pedig azonosan nulla.

Mi köze f_n-nek a tétel állításához? Az, hogy a bal oldali kifejezés közelítőleg $\int_a^b f_n(x)\mathrm{d}x$, precízebben

- $\lim_{n\to\infty}\int_a^b f_n(x)\mathrm{d}x-\mathbb{P}\big(a<h(S_n-np)<b\big)=0$

Ha ezt a konvergenciát elhisszük, akkor elég bizonyítani, hogy $\int_a^b f_n\to\int_a^b\varphi$ minden $a<b$ esetén. Az ebben szereplő $n\mapsto f_n$ függvénysorozatról elemi módon belátható a következő:

- Ha $x\in\mathbb{R}$, ami nem $h(k-np)$ alakú semmilyen k és n esetén, akkor f_n deriválható x-ben minden n-re, és
$$\lim_{n\to\infty}\frac{f_n'(x)}{f_n(x)}=-x.$$

Ez az állítás lényeges információt tartalmaz az $f_n(x)$ határértékekről: azt, hogy az $f'(x)=-xf(x)$ differenciálegyenlet aszimptotikus értelemben teljesül f_n-re, néhány speciális alakú x kivételével. Ezen a ponton viszont számos, nem elhanyagolható technikai kérdésbe ütközünk:

- Létezik-e egyáltalán az $f_n(x)$ határértéke minden $x\in\mathbb{R}$ esetén?
- Ha létezik az $x\mapsto f(x)=\lim_{n\to\infty}f_n(x)$ függvény, akkor igaz-e, hogy minden pontban pozitív, folytonos, illetve folytonosan differenciálható? Továbbá sűrűségfüggvény-e?
- Teljesül-e $\lim_{n\to\infty}f_n'(x)=f'(x)$ minden $x\in\mathbb{R}$ esetén?
- Ha f_n-ről belátjuk, hogy $\lim_{n\to\infty}f_n(x)=\varphi(x)$ minden $x\in\mathbb{R}$ esetén, abból következik-e, hogy $\lim_{n\to\infty}\int_a^b f_n=\int_a^b\varphi$ minden $a<b$ esetén?

A fenti vázlatpontok tárgyalása meghaladja a jegyzet kereteit. Ha feltételezzük, hogy ezek a tulajdonságok teljesülnek, akkor a következőt kapjuk:

$$-x = \lim_{n\to\infty} \frac{f_n'(x)}{f_n(x)} = \frac{\lim_{n\to\infty} f_n'(x)}{\lim_{n\to\infty} f_n(x)} = \frac{f'(x)}{f(x)} = \big(\ln f(x)\big)'$$

majdnem minden $x \in \mathbb{R}$ esetén (azaz ha x nem $h(k-np)$ alakú). Ezt integrálva adódik, hogy $\ln f(x) = -\frac{x^2}{2} + c$ valamilyen $c \in \mathbb{R}$ esetén, azaz

$$f(x) = e^{-\frac{x^2}{2}} e^c \qquad (x \in \mathbb{R}).$$

Mivel elfogadtuk, hogy f sűrűségfüggvény, így e^c szükségképpen $\frac{1}{\sqrt{2\pi}}$, vagyis f_n valóban a standard normális sűrűségfüggvényhez tart.

A normális eloszlásnak léteznek további karakterizációi, de ezeket itt idő hiányában megint csak nem tárgyaljuk.[60]

A.4. A centrális határeloszlás-tétel bizonyításának vázlata

A centrális határeloszlás-tétel bizonyításához a A.2 fejezetből meríthetünk ötletet, egész pontosan a Csernov-egyenlőtlenségben megjelenő mennyiségből: az $\mathbb{E}(e^{tX})$ várható értékből.

A.4.1. Definíció. Az X valószínűségi változó **momentumgeneráló függvény**e:[61]

$$M_X(t) \stackrel{\text{def}}{=} \mathbb{E}(e^{tX}),$$

azon t helyeken értelmezve, ahol $\mathbb{E}(e^{tX})$ véges.

A momentumgeneráló függvényt az alábbi hasznos tulajdonságai teszik alkalmassá, hogy felhasználjuk a centrális határeloszlás-tétel bebizonyításában.

A.4.2. Állítás. *Legyenek Y, Z illetve $X_1, X_2, \dots$ valószínűségi változók. Tegyük fel, hogy $M_Y(t)$ és $M_Z(t)$ minden $t \in \mathbb{R}$ esetén értelmezett.*

(1) Ha $M_Y(t) = M_Z(t)$ minden $t \in \mathbb{R}$ esetén, akkor Y és Z azonos eloszlásúak.

(2) Ha $\lim_{n\to\infty} M_{X_n}(t) = M_Z(t)$ minden $t \in \mathbb{R}$ esetén, akkor $X_n \stackrel{d}{\to} Z$.

A bizonyításban feltesszük, hogy a momentumgeneráló függvények mindenhol értelmezettek. Az állítás enélkül is teljesülne, de az általános esetet itt nem bizonyítjuk. Továbbá a bizonyításban megjelenő lemmát szintén érvelés nélkül elfogadjuk.

Centrális határeloszlás-tétel bizonyításvázlata. A számolást egyszerűsítendő, csak azt az esetet vizsgáljuk, amikor $\mathbb{E}(X_1) = 0$ és $\mathbb{D}(X_1) = 1$. Ez valóban elegendő, ugyanis ha tudjuk a tételt a valószínűségi változók $\frac{X_i - \mathbb{E}(X_1)}{\mathbb{D}(X_1)}$ standardizáltjaira (amikre már teljesül, hogy a várható értékük 0, szórásuk 1), akkor ebből a tétel az alábbi átrendezéssel következik:

$$\frac{X_1 + \dots + X_n - n\mathbb{E}(X_1)}{\sqrt{n}\mathbb{D}(X_1)} = \sqrt{n}\frac{\sum_{i=1}^n X_i - \mathbb{E}(X_1)}{n\mathbb{D}(X_1)} = \sqrt{n}\frac{\sum_{i=1}^n \frac{X_i - \mathbb{E}(X_1)}{\mathbb{D}(X_1)}}{n}.$$

Tehát azt kell belátnunk, hogy $\sqrt{n} \cdot \overline{X_n} \stackrel{d}{\to} Z$. A momentumgeneráló függvényre vonatkozó állítás szerint ehhez elég belátni, hogy

$$\lim_{n\to\infty} M_{\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}(t) = M_Z(t) \tag{A.1}$$

[60] Karakterizációk egy gyűjteményét lásd: [Cross Validated] What is the most surprising characterization of the Gaussian (normal) distribution?

[61] A momentumgeneráló függvény a Laplace-transzformálttal rokon fogalom. A másik sűrűn előkerülő fogalom a témában a karakterisztikus függvény, ami pedig a Fourier-transzformáció valószínűségszámításbeli megfelelője. Fourier-transzformációról nagyszerű magyarázó videó: [youtube - 3blue1brown, Fourier Transform]

minden $t \in \mathbb{R}$ esetén.

Számoljuk ki először az (A.1) egyenlet jobb oldalát:

$$M_Z(t) = \mathbb{E}(e^{tZ}) = \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \mathrm{d}z = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(z-t)^2}{2} + \frac{t^2}{2}} \mathrm{d}z = e^{\frac{t^2}{2}}.$$

Más szavakkal, a standard normális eloszlás momentumgeneráló függvénye $t \mapsto e^{\frac{t^2}{2}}$.

Az (A.1) egyenlet bal oldalának vizsgálata előtt számoljuk ki az X_1 momentumgeneráló függvényét is:

$$\text{(A.2)} \quad M_{X_1}(t) = \mathbb{E}\Big(\sum_{k=0}^{\infty} \frac{1}{k!} t^k X_1^k\Big) = 1 + t\mathbb{E}X_1 + \frac{t^2}{2}\mathbb{E}(X_1^2) + \frac{t^2}{2} \cdot \mathbb{E}\Big(\sum_{k=3}^{\infty} \frac{2}{k!} t^{k-2} X_1^k\Big) = 1 + \frac{t^2}{2}\big(1 + r(t)\big),$$

ahol $r(t)$ jelöli az utolsó szumma várható értékét. Egyáltalán nem nyilvánvaló, de belátható a következő:

A.4.3. Lemma. $\lim_{t \to 0} r(t) = 0$.

A lemma miatt az (A.1) egyenlet bal oldaláról a következő mondható:

$$\text{(A.3)} \quad M_{\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}(t) = \mathbb{E}\left(e^{t \frac{\sum_{i=1}^n X_i}{\sqrt{n}}}\right) = \mathbb{E}\left(\prod_{i=1}^n e^{\frac{t}{\sqrt{n}} X_i}\right) = \\ = \prod_{i=1}^n \mathbb{E}\left(e^{\frac{t}{\sqrt{n}} X_i}\right) = \prod_{i=1}^n M_{X_i}\left(\frac{t}{\sqrt{n}}\right) = M_{X_1}\left(\frac{t}{\sqrt{n}}\right)^n,$$

ahol felhasználtuk a következő két tulajdonságot: Egyrészt az $e^{\frac{t}{\sqrt{n}} X_i}$ valószínűségi változók függetlenek, másrészt $M_{X_i}\left(\frac{t}{\sqrt{n}}\right)$ minden i esetén megegyezik $M_{X_1}\left(\frac{t}{\sqrt{n}}\right)$-nel, hiszen a momentumgeneráló függvény csak X_i eloszlásától függ, ami minden i esetén ugyanaz, mint X_1 eloszlása.

Helyettesítsük be az (A.3) egyenlet logaritmusába az $M_{X_1}\left(\frac{t}{\sqrt{n}}\right)$ (A.2) egyenlet szerinti értékét:

$$\ln M_{\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}(t) = n \ln M_{X_1}\left(\frac{t}{\sqrt{n}}\right) = n \ln\left(1 + \frac{\left(\frac{t}{\sqrt{n}}\right)^2}{2}\left(1 + r\left(\frac{t}{\sqrt{n}}\right)\right)\right).$$

Jelölje $r\left(\frac{t}{\sqrt{n}}\right)$-et r_n. A lemma szerint $r_n \to 0$, ha $n \to \infty$. Ezért az egyenletet a következőképp folytathatjuk:

$$\ln M_{\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}(t) = n \ln\left(1 + \frac{t^2}{2n}(1 + r_n)\right) = \frac{\ln\left(1 + \frac{t^2}{2n}(1 + r_n)\right)}{\frac{t^2}{2n}(1 + r_n)} \cdot \frac{t^2}{2} \cdot (1 + r_n).$$

A lemma miatt $\frac{t^2}{2n}(1 + r_n) \to 0$, ha $n \to \infty$ (rögzített t esetén). Emiatt a hármas szorzat első tényezője 1-hez tart, felhasználva, hogy $\frac{\ln(1+y)}{y} \to 1$, ha $y \to 0$. Összességében:

$$\lim_{n \to \infty} \ln M_{\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}(t) = 1 \cdot \frac{t^2}{2} \cdot (1 + 0) = \frac{t^2}{2}.$$

Erre alkalmazhatjuk az exponenciális függvényt (ami folytonos, így felcserélhető a határértékképzéssel). Tehát $\lim_{n \to \infty} M_{\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}(t) = e^{\frac{t^2}{2}} = M_Z(t)$ minden $t \in \mathbb{R}$ esetén. □

A.5. Lineáris regresszió

Tegyük fel, hogy egy esernyőket áruló bolt tulajdonosai vagyunk, és kapunk egy hosszútávú előrejelzést a jövő évi csapadékmennyiségről. Jobb híján ezen előrejelzés alapján próbáljuk megtippelni, mekkora készletet rendeljünk, azaz körülbelül hány esernyőt fogunk eladni. Hogyan kellene tippeljünk, ha a korábbi évek alapján van némi elképzelésünk a csapadékmennyiség és az eladott esernyők száma közti összefüggésről? Ilyen becslésre az egyik lehetséges módszerünk a lineáris regresszió.

Jelölje X az éves csapadékmennyiséget, Y pedig az eladott esernyők számát, ahogy a második példában. Tegyük fel, hogy (X, Y) együttes sűrűségfüggvénye a példában szereplő $f_{X,Y}$. A lineáris regresszió alapötlete, hogy próbáljuk meg Y-t az X-nek egy lineáris függvényével, azaz $\beta \cdot X + \alpha$ alakban, a lehető legjobban közelíteni.

Vegyük észre, hogy a „legjobb közelítés" nem egy jóldefiniált fogalom: azt még meg kéne mondanunk, mi alapján tekintünk egy közelítést jónak vagy rossznak. Erre többféle megközelítés is bevethető,[62] de a legalapvetőbb, az ún. **legkisebb négyzetek módszere**.

A.5.1. Definíció. Legyenek X és Y valószínűségi változók. Ekkor Y-nak az X-re vett **lineáris regresszió**ján azt a $\beta X + \alpha$ valószínűségi változót értjük, ahol $\alpha, \beta \in \mathbb{R}$, és az

$$\mathbb{E}\Big(\big(Y - (\beta X + \alpha)\big)^2\Big) \tag{A.4}$$

mennyiség minimális.

Ennek az optimalizálási problémának a megoldása lényegében mindig létezik és egyértelmű:

A.5.2. Állítás. *Legyenek X és Y olyan valószínűségi változók, amire $\mathbb{D}^2(X)$, $\mathbb{D}^2(Y)$ és $\operatorname{cov}(X, Y)$ véges, továbbá $\mathbb{D}^2(X) \neq 0$. Ekkor a (A.4) egyenletben szereplő várható érték pontosan akkor minimális, ha*

$$\beta = \frac{\operatorname{cov}(X, Y)}{\mathbb{D}^2(X)} \qquad \text{és} \qquad \alpha = \mathbb{E}(Y) - \frac{\operatorname{cov}(X, Y)}{\mathbb{D}^2(X)}\mathbb{E}(X).$$

A.5.3. Definíció. Az Y valószínűségi változó X-re vett **regressziós egyenes**e az

$$\{(x, y) \in \mathbb{R}^2 \mid y = \beta x + \alpha\}$$

egyenes a síkon, ahol β és α értéke a fenti állításban szerepel.

Vizuálisabban, az (X, Y) valószínűségi vektorváltozó lehetséges értékeinek a síkján az eloszlást „legjobban közelítő" egyenes a regressziós egyenes. A lineáris regresszió akkor lesz jól használható modell, ha az (X, Y) együttes eloszlása ezen egyenes környékén koncentrálódik.

Megjegyzés. A β-ra és az α-ra vonatkozó egyenleteket nem feltétlenül egyszerű sem megjegyezni, sem megindokolni. Egy heurisztika (de nem bizonyítás) a helyes α és β megtalálására, hogy olyannak válasszuk őket, amire Y-nak és $\beta X + \alpha$-nak ugyanaz a várható értéke és az X-el vett kovarianciája. Emiatt

$$\mathbb{E}(Y) = \mathbb{E}(\beta X + \alpha) = \beta\mathbb{E}(X) + \alpha \qquad \text{és} \qquad \operatorname{cov}(X, Y) = \operatorname{cov}(X, \beta X + \alpha) = \beta\mathbb{D}^2(X) + 0,$$

amely egyenletekből adódik is β és α értéke.

Egy hasonló, kompaktabb megközelítés a korreláció fogalmán keresztül vezet. Idézzük fel, X és Y korrelációja:

$$\operatorname{corr}(X, Y) \stackrel{\text{def}}{=} \frac{\operatorname{cov}(X, Y)}{\mathbb{D}(X)\mathbb{D}(Y)}$$

[62] A lineáris regresszió alternatív változatai, amelyek máshogy definiálják a „legjobb közelítés" fogalmát: súlyozott lineáris regresszió, ridge regresszió, avagy az ℓ_1 regresszió.

egy -1 és 1 közti valós szám, ami X és Y lineáris összefüggését méri. Azt állítjuk, hogy ha $\beta X+\alpha$ az Y lineáris regressziója X-re, akkor teljesül, hogy

$$\frac{(\beta X+\alpha)-\mathbb{E}(Y)}{\mathbb{D}(Y)}=\frac{X-\mathbb{E}(X)}{\mathbb{D}(X)}\cdot\operatorname{corr}(X,Y).$$

Más szavakkal, ha Y standardizáltjába az első Y helyére a $\beta X+\alpha$ regressziót helyettesítjük, akkor az eredmény X standardizáltjának korreláció-szorosa. Ez az azonosság egyszerű átrendezéssel belátható.

Bizonyítás. A következő függvényt kellene minimalizálnunk:

$$h(\alpha,\beta)=\mathbb{E}\Big(\big(Y-(\beta X+\alpha)\big)^2\Big)=\mathbb{E}\Big(Y^2+\beta^2X^2+\alpha^2-2\beta XY-2\alpha Y+2\alpha\beta X\Big)=$$

$$=\mathbb{E}(Y^2)+\beta^2\mathbb{E}(X^2)+\alpha^2-2\beta\mathbb{E}(XY)-2\alpha\mathbb{E}(Y)+2\alpha\beta\mathbb{E}(X).$$

Az eredeti formából látszik, hogy h nemnegatív (hiszen valószínűségi változó négyzetének várható értéke), továbbá az átalakított formából világos, hogy α-ban és β-ban h másodfokú polinom. Egy ilyen polinomnak csak ott lehet globális minimuma, ahol mind az α-ban, mind a β-ban vett parciális derivált eltűnik.

Bár egy (α_0,β_0) pontban a parciális deriváltak eltűnése nem elégséges feltétele annak, hogy ez a pont a h függvény globális minimuma legyen, jelen esetben a nemnegativitás miatt mégis ez a helyzet. Valóban, indirekt tegyük fel, hogy az (α_0,β_0) pontban eltűnik mindkét parciális derivált, de $h(\alpha_1,\beta_1)<h(\alpha_0,\beta_0)$. Nézzük a függvényt a két pontot összekötő egyenesen, vagyis tekintsük az $f(t)=h(t\alpha_0+(1-t)\alpha_1,t\beta_0+(1-t)\beta_1)$ egyváltozós függvényt. Mivel ezt h-ból lineáris behelyettesítéssel kaptuk, így polinom kell legyen t-ben, ami legfeljebb másodfokú. Sőt, a 0-ban vett deriváltját is ki tudjuk fejezni h parciális deriváltjaival az (α_0,β_0) pontban, ezért $f'(0)=0$, hiszen a parciális deriváltak eltűnnek. Összefoglalva, f egy olyan legfeljebb másodfokú polinom, amire $f'(0)=0$, és f mindenhol nemnegatív (ebből már látjuk, hogy f vagy egy felfelé álló parabola, vagy konstans), de mégis $h(\alpha_1,\beta_1)=f(1)<f(0)=h(\alpha_0,\beta_0)$. Ez ellentmondás, ilyen polinom nincs.

Visszatérve a globális minimum pontos értékére, h parciális deriváltjai a következők:

$$\beta\text{ szerint: }2\beta\mathbb{E}(X^2)-2\mathbb{E}(XY)+2\alpha\mathbb{E}(X)\qquad\text{és}\qquad\alpha\text{ szerint: }2\alpha-2\mathbb{E}(Y)+2\beta\mathbb{E}(X).$$

Vagyis a parciális deriváltak közös nullhelyeit megadó egyenletek:

$$\alpha\mathbb{E}(X)+\beta\mathbb{E}(X^2)=\mathbb{E}(XY)\qquad\text{és}\qquad\alpha+\beta\mathbb{E}(X)=\mathbb{E}(Y).$$

Ez egy 2×2-es lineáris egyenletrendszer α-ban és β-ban. Megoldása:

$$\beta=\frac{\mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)}{\mathbb{E}(X^2)-\mathbb{E}(X)^2}=\frac{\operatorname{cov}(X,Y)}{\mathbb{D}^2(X)}\qquad\text{és}\qquad\alpha=\mathbb{E}(Y)-\beta\mathbb{E}(X)=\mathbb{E}(Y)-\frac{\operatorname{cov}(X,Y)}{\mathbb{D}^2(X)}\mathbb{E}(X),$$

amik éppen a kívánt egyenletek. □

A.5.4. Példa. Mit kapunk a felvezető példa esetében, ahol X a csapadékmennyiség, Y az eladott esernyők száma? A már kiszámolt kovarianciamátrix koordinátáiból rögtön felírhatók az Y-nak az X-re vett lineáris regressziójának együtthatói:

$$\beta=\frac{\operatorname{cov}(X,Y)}{\mathbb{D}^2(X)}=\frac{1/225}{7/90}=\frac{2}{35},\qquad\alpha=\mathbb{E}(Y)-\frac{\operatorname{cov}(X,Y)}{\mathbb{D}^2(X)}\mathbb{E}(X)=\frac{4}{5}-\frac{2}{35}\frac{7}{15}=\frac{58}{75}.$$

Tehát ha X-re kapunk egy előrejelzést, akkor ezen együtthatókkal közelíthetjük Y-t. Némi értelmezést hozzáadva: az eső mennyiségének emelkedése csak kismértékben fogja növelni a már alapból magas készletszükségletet.

Mivel a lineáris regresszió csak közelítés, így fontos információ lehet, hogy mekkora hibával találja el Y értékét. (Hiba alatt itt átlagos négyzetes hibát, vagyis szórásnégyzetet értünk.)

A.5.5. Állítás. *Legyen az Y valószínűségi változó X-re vett lineáris regressziója $\beta X + \alpha$. Ekkor az eltérés szórásnégyzete:*

$$\mathbb{D}^2\Big(Y - (\beta X + \alpha)\Big) = \mathbb{D}^2(Y) - \frac{\mathrm{cov}(X,Y)^2}{\mathbb{D}^2(X)}.$$

Bizonyítás. A szórásnégyzet fentebb felsorolt tulajdonságai és $\beta = \frac{\mathrm{cov}(X,Y)}{\mathbb{D}^2(X)}$ miatt:

$$\mathbb{D}^2\Big(Y - (\beta X + \alpha)\Big) = \mathbb{D}^2(Y - \beta X) = \mathbb{D}^2(Y) + \beta^2\mathbb{D}^2(X) - 2\mathrm{cov}(Y, \beta X) =$$

$$= \mathbb{D}^2(Y) + \frac{\mathrm{cov}(X,Y)^2}{\big(\mathbb{D}^2(X)\big)^2}\mathbb{D}^2(X) - 2\frac{\mathrm{cov}(X,Y)}{\mathbb{D}^2(X)}\mathrm{cov}(Y,X) = \mathbb{D}^2(Y) - \frac{\mathrm{cov}(X,Y)^2}{\mathbb{D}^2(X)}.$$

Éppen ez volt az állítás. □

Megjegyzés. Másképpen felírva:

$$\mathbb{D}^2\Big(Y - (\beta X + \alpha)\Big) = \mathbb{D}^2(Y) \cdot \big(1 - \mathrm{corr}(X,Y)^2\big).$$

Speciálisan, minél nagyobb a korreláció X és Y közt, annál kisebb rész járul hozzá a hiba szórásnégyzetéhez $\mathbb{D}^2(Y)$-ból. Továbbá, ez az átfogalmazás azt is mutatja, hogy a fenti állításból következik a 4.4 alfejezet állítása.

A.5.6. Példa. Az előző példa esetében

$$\mathbb{D}^2\Big(Y - (\beta X + \alpha)\Big) = \frac{58}{225} - \frac{(1/225)^2}{(7/90)^2} \approx 0{,}2545.$$

Vagyis az eladások jócskán eltérhetnek a lineáris regresszió által becsült értéktől.

Hasznos észben tartani, hogy statisztikai témakörben nem ugyanezt értik lineáris regresszió alatt. A különbség, hogy ott nem feltételezik, hogy a valószínűségi változók eloszlása ismert, de általában azt sem, hogy (az esetlegesen ebből levezethető) kovariancia és szórásnégyzet értékeit ismernénk. Így a statisztikai értelemben vett lineáris regresszióba beleértik azt is, hogy a β és α értékek maguk is becsült mennyiségek, egy véges nagyságú minta alapján. Ez lényegesen eltérő egyenleteket és értelmezést jelent, de ettől még a lineáris regresszió ötlete ugyanaz marad: keressünk közelítőleg lineáris összefüggést a vizsgált változók között.

A.6. Feltételes várható érték folytonos esetben

A feltételes várható érték fogalmát abban az esetben is szeretnénk bevezetni, ha X folytonos valószínűségi változó. Ezzel az a lényeges probléma, hogy $\mathbb{P}(X = x) = 0$ minden x érték esetén, ezért $\mathbb{E}(Y \mid X = x)$ a fenti definícióval értelmetlen. De ne adjuk fel rögtön, ugyanis az „Y legjobb közelítése X függvényében” fogalom viszont nem tűnik értelmetlennek.

A feltételes várható értéket általánosan az előző alfejezet utolsó állításával definiálhatjuk.

A.6.1. Definíció. Legyenek X és Y valószínűségi változók és tegyük fel, hogy $\mathbb{E}(|Y|) < \infty$. Az Y-nak az X-re vett **feltételes várható értéke** az az $r(X)$ alakú[63] valószínűségi változó, amire

$$\mathbb{E}\big(r(X) \mid X < x\big) = \mathbb{E}(Y \mid X < x) \tag{A.5}$$

minden olyan $x \in \mathbb{R}$ esetén, amire $\mathbb{P}(X < x) > 0$ (ami miatt $\mathbb{E}(. \mid X < x)$ definiált az előző alfejezet feltételes várható érték definíciójával). Az $r(X)$ valószínűségi változó jelölése: $\mathbb{E}(Y \mid X)$.

[63] Az r nem akármilyen függvény, hiszen $r(X)$ valószínűségi változó kell legyen. Ehhez általában azt követelik meg, hogy r úgynevezett Borel-mérhető függvény legyen; ekvivalensen, folytonos függvények pontonkénti határértéke.

Láttuk, hogy a diszkrét feltételes várható érték is teljesíti ezt a tulajdonságot (lásd az erre vonatkozó állítást). Emellett itt jegyezzük meg, hogy az $\mathbb{E}(Y \mid X)$ feltételes várható érték, mint valószínűségi változó, nem egyértelmű (de azért majdnem). Ugyanúgy, ahogy a sűrűségfüggvény sem egyértelmű, ugyanis r értékét néhány olyan ponton büntetlenül megváltoztathatjuk, amiket 0 valószínűséggel vesz fel X. Egy ilyen változtatás $r(X)$ értékét is megváltoztatja, de a definícióban szereplő egyenlet érvényben marad.

Az r függvény[64] jelölése: $x \mapsto \mathbb{E}(Y \mid X = x)$, elnevezése **regressziós függvény**. A jelölés több értelemben sem pontos. Egyrészt a fejezet legelején felírt feltételes várható érték értelemben az $\mathbb{E}(Y \mid X = x)$ nem feltétlenül van értelmezve, hiszen $\mathbb{P}(X = x)$ lehet nulla is. Ettől függetlenül a jelölés szemléletes, hiszen informálisan $r(x)$ az Y átlagos legjobb közelítése az $\{X = x\}$ feltétel esetén.

Másrészt a jelölés abban az értelemben sem precíz, hogy r egyáltalán nem egyértelmű, így rögzített x-re a függvénynek nincs jóldefiniált értéke. Például ha X nemnegatív, akkor r a negatív félegyenesen akárhogy megválasztható. Mivel r kiszámolása tipikusan egy köztes lépés az $\mathbb{E}(Y \mid X)$ valószínűségi változó kiszámolásához, amely valószínűségi változó már lényegében egyértelmű (lásd előző bekezdés), így r nem egyértelmű voltával nem fogunk a továbbiakban foglalkozni.

Megjegyzés. Ha $\mathbb{E}(|Y|) < \infty$, akkor $\mathbb{E}(Y \mid X)$ is létezik. Ezt nem bizonyítjuk.

Rendben, most már definiálva van a feltételes várható érték. De hogyan lehet kiszámolni akár az r regressziós függvényt, akár a $r(X) = \mathbb{E}(Y \mid X)$ feltételes várható értéket? Ha X egyszerű, akkor ez az előző alfejezet (5.1) egyenletéből világos:

$$\mathbb{E}(Y \mid X = x) = \sum_{y \in \mathrm{Ran}(Y)} y \cdot \mathbb{P}(Y = y \mid X = x)$$

minden $x \in S_X$ esetén, hiszen ez a $\mathbb{P}(. \mid X = x)$ valószínűségi mérték szerinti várható érték. Folytonos esetben ehelyett a következőt tudjuk mondani.

A.6.2. Definíció. Legyen (X, Y) folytonos valószínűségi vektorváltozó, és jelölje együttes sűrűségfüggvényét $f_{X,Y}$. Ekkor Y-nak az X-re vett **feltételes sűrűségfüggvénye**:

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X,Y}(x,y)}{\int_{-\infty}^{\infty} f_{X,Y}(x,u)\mathrm{d}u},$$

olyan $x, y \in \mathbb{R}$ számokra értelmezve, amire $f_X(x) \neq 0$, és $f_{Y|X}(y \mid x) = 0$ ha $f_X(x) = 0$.

A definíció megjegyzésében segíthet, ha észrevesszük a hasonlóságát a $\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$ egyenlőséggel.

A.6.3. Állítás. *Legyen (X, Y) folytonos valószínűségi vektorváltozó, és tegyük fel, hogy $\mathbb{E}(|Y|) < \infty$. Ekkor*

$$\mathbb{E}(Y \mid X = x) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid x)\mathrm{d}y$$

az Y-nak az X-re vett regressziós függvénye.

Bizonyítás. Jelölje r a $z \mapsto \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid z)\mathrm{d}y$ függvényt ($z \in \mathbb{R}$). Azt kell leellenőriznünk, hogy r-re teljesül a (A.5) egyenlet.

Legyen $x \in \mathbb{R}$ olyan, amire $\mathbb{P}(X < x) > 0$. A fejezet első állításа miatt

$$\mathbb{E}\big(r(X) \mid X < x\big) = \frac{1}{\mathbb{P}(X < x)}\mathbb{E}\big(r(X)\mathbf{1}_{\{X<x\}}\big),$$

[64] Az r értelmezési tartományát nem specifikáltuk, nem is igazán tudjuk, hiszen r nem egyértelmű. Ami lényeges feltétel a r értelmezési tartományára, hogy tartalmazza azon $x \in \mathbb{R}$ pontokat, aminek bármilyen kis környezetébe pozitív eséllyel esik X, legfeljebb egy 0 mértékű halmaz kivételével.

ahol

$$\mathbb{E}\big(r(X)\mathbf{1}_{\{X<x\}}\big) = \int_{-\infty}^{\infty} r(z)\mathbf{1}_{\{z<x\}} f_X(z)\mathrm{d}z = \int_{-\infty}^{x} r(z) f_X(z)\mathrm{d}z.$$

Ha $f_X(z) > 0$, akkor r definíciója szerint

$$r(z)f_X(z) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid z)\mathrm{d}y \cdot f_X(z) = \int_{-\infty}^{\infty} y \cdot \frac{f_{X,Y}(z,y)}{f_X(z)} f_X(z)\mathrm{d}y = \int_{-\infty}^{\infty} y \cdot f_{X,Y}(z,y)\mathrm{d}y.$$

Ha $f_X(z) = 0$, akkor $r(z)f_X(z) = 0$. Leellenőrizhető, hogy a jobb oldali integrál szintén 0, ha $f_X(z) = 0$, felhasználva, hogy $\int_{-\infty}^{\infty} f_{X,Y}(z,y)\mathrm{d}y = f_X(z)$. Összességében, az előző három egyenletből az adódik,

$$\mathbb{E}\big(r(X) \mid X < x\big) = \frac{1}{\mathbb{P}(X<x)} \int_{-\infty}^{x}\int_{-\infty}^{\infty} y \cdot f_{X,Y}(z,y)\mathrm{d}y\mathrm{d}z.$$

Itt felhasználhatjuk a valószínűségi vektorváltozó transzformáltjának várható értékére vonatkozó állítást, és a fejezet első állítását, így a fentit folytatva

$$= \frac{1}{\mathbb{P}(X<x)}\mathbb{E}(Y\mathbf{1}_{\{X<x\}}) = \mathbb{E}(Y \mid X < x),$$

ami éppen a belátandó állítás. □

A.6.4. Példa. Legyen az (X,Y) valószínűségi vektorváltozó együttes sűrűségfüggvénye $15x^2y$, ha $0 < x < y < 1$, és 0 egyébként. Határozzuk meg az $\mathbb{E}(Y \mid X)$ feltételes várható értéket.
Először számoljuk ki X sűrűségfüggvényét a $0 < x < 1$ esetekben:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}y = \int_{x}^{1} 15x^2y\mathrm{d}y = \frac{15}{2}(x^2 - x^4).$$

Tehát a feltételes sűrűségfüggvény értéke $0 < x < y < 1$ esetén

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{15x^2y}{\frac{15}{2}(x^2-x^4)} = \frac{2y}{1-x^2},$$

és $f_{Y|X}(y \mid x) = 0$ egyébként. Tehát $0 < x < 1$ esetén

$$\mathbb{E}(Y \mid X = x) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid x)\mathrm{d}y = \int_{x}^{1} y \cdot \frac{2y}{1-x^2}\mathrm{d}y = \frac{2}{3} \cdot \frac{1-x^3}{1-x^2}.$$

Behelyettesítve, az $\mathbb{E}(Y \mid X)$ feltételes várható érték a $\frac{2}{3} \cdot \frac{1-X^3}{1-X^2}$ valószínűségi változó.[65]

A.6.5. Példa. Vegyük észre, hogy a 5.3.1. Állításban nem szerepel, hogy az X, Y, Z valószínűségi változóknak diszkrétnek kell lenniük. Az állítás szerint például ha X egyenletes eloszlású a $(0;1)$ intervallumon és Y tőle független és exponenciális eloszlású $\lambda = 2$ paraméterrel, akkor $\mathbb{E}(4Ye^X + X + 1|X) = 4e^X\mathbb{E}(Y) + X + 1 = 2e^X + X + 1$.

Ugyanúgy igaz marad az állítás, ha a benne szereplő valószínűségi változók egy része diszkrét, másik része pedig folytonos. Például ha az előbbi, $(0;1)$-en egyenletes eloszlású X helyett bármilyen véges várható értékű diszkrét X-et írunk, Y-t pedig változatlanul X-től függetlennek és Exp(2) eloszlásúnak hagyjuk, $\mathbb{E}(4Ye^X + X + 1|X) = 4e^X\mathbb{E}(Y) + X + 1 = 2e^X + X + 1$ akkor is teljesülni fog.

A teljes várható érték tételét általános valószínűségi változókra mondtuk ki, bár eddig csak diszkrét esetben tudtuk kiszámolni a benne szereplő mennyiségeket. Az együttesen folytonos esetben a tétel a következőképpen írható.

[65] Érdekes utánaszámolni, hogy a feltételes várható érték egyáltalán nem szimmetrikus X-ben és Y-ban, azaz $\mathbb{E}(X \mid Y)$ egyáltalán nem biztos, hogy hasonlít $\mathbb{E}(Y \mid X)$-re.

A.6.6. Állítás (Teljes várható érték tétele, folytonos eset)**.** *Legyen X folytonos valószínűségi változó, $\mathbb{E}(Y) < \infty$ és jelölje X sűrűségfüggvényét f_X. Ekkor*

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} \mathbb{E}(Y \mid X = x) f_X(x) \mathrm{d}x,$$

ahol $\mathbb{E}(Y \mid X = x)$ az Y-nak az X-re vett regressziós függvénye.

A.7. Teljes valószínűség tétele, folytonos eset

Az előző fejezet után maradhatott némi hiányérzet az olvasóban: míg a teljes várható érték tételét többféle formában is kimondtuk (teljes eseményrendszerrel és valószínűségi változó $\{X = x\}$ szinthalmazaival is diszkrét és folytonos esetben), addig a teljes valószínűség tételét csak teljes eseményrendszerre fogalmaztuk meg. Mi a helyzet az utóbbi tétellel akkor, ha a feltétel egy valószínűségi változó (szinthalmaza)?

Ha X diszkrét valószínűségi változó, akkor a teljes valószínűség tétele nem újdonság:

$$\mathbb{P}(A) = \sum_{k \in S_X} \mathbb{P}(A \mid X = k) \cdot \mathbb{P}(X = k),$$

ahol S_X azon k értékek halmaza, ahol $\mathbb{P}(X = k) > 0$, és így van értelme a $\mathbb{P}(A \mid X = k)$ feltételes valószínűségről beszélni. Ez az eredeti teljes valószínűség tételének speciális esete. Viszont ha X folytonos, akkor $\mathbb{P}(A \mid X = k)$ értelmetlen. A probléma feloldása ugyanaz, mint a regressziós függvény esetében.

A.7.1. Definíció. Legyen X valószínűségi változó, és A esemény. Ekkor A-nak az **X-re vett feltételes valószínűsége** az

$$x \mapsto \mathbb{E}(\mathbf{1}_A \mid X = x)$$

regressziós függvény. Jelölése: $\mathbb{P}(A \mid X = x)$.

Itt a regressziós függvényt a A.6 alfejezet definíciója szerint értjük, vagyis ez az az r függvény, amire $r(X) = \mathbb{E}(\mathbf{1}_A \mid X)$. Az $\mathbb{E}(\mathbf{1}_A \mid X)$ feltételes várható érték pedig a (A.5) egyenlettel definiált. (Mivel $\mathbb{E}(\mathbf{1}_A)$ azaz $\mathbb{P}(A)$ véges, így r létezik a A.6 alfejezet megjegyzése okán.) Szemléletesen, $\mathbb{E}(\mathbf{1}_A \mid X = x)$ jelentése az A esemény valószínűsége (avagy precízebben, annak legjobb átlagos közelítése), tudván az X értékét.

Innen a teljes valószínűség tétele már megtippelhető:

A.7.2. Tétel (Teljes valószínűség tétele)**.** *Legyen X folytonos valószínűségi változó, és A esemény. Ekkor*

$$\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid X = x) f_X(x) \mathrm{d}x,$$

ahol f_X az X sűrűségfüggvénye.

A.7.3. Példa. Jelölje X egy hallgatónak a valószínűségszámítás vizsgára szánt felkészülési idejét. Tegyük fel, hogy X egyenletes eloszlású az $[\varepsilon, 20]$ intervallumon (napban számolva, ahol ε 20-nál kisebb, és őszintén remélem, hogy pozitív valós szám). Feltételezve, hogy x időt szán felkészülésre a hallgató, $\left(\frac{x}{21}\right)^2$ a valószínűsége, hogy ötös érdemjegyet kap. Mi a valószínűsége az ötös vizsgának?

Az előző tétel jelölésével: tudjuk, hogy $f_X(x) = \frac{1}{20-\varepsilon}$ ha $\varepsilon \le x \le 20$, illetve 0 egyébként. Továbbá $\mathbb{P}(A \mid X = x) = \left(\frac{x}{21}\right)^2$. Tehát

$$\mathbb{P}(A) = \int_{\varepsilon}^{20} \left(\frac{x}{21}\right)^2 \frac{1}{20-\varepsilon} \mathrm{d}x = \left[\frac{x^3}{3 \cdot 21^2 (20-\varepsilon)}\right]_{\varepsilon}^{20} = \frac{\varepsilon^2 + 20\varepsilon + 20^2}{3 \cdot 21^2}$$

Ha $\varepsilon = 1$, akkor ez kerekítve 0,3182.

Megjegyzés. A feltételes valószínűség speciális esete a feltételes eloszlásfüggvény:

$$F_{Y|X}(y \mid x) = \mathbb{P}(Y < y \mid X = x).$$

A.8. Multinomiális (polinomiális) eloszlás

Nevezetes diszkrét eloszlás a binomiális. Hogyan általánosítható ez több változóra? Erre van egy kézenfekvő módszer: legyenek $X_1, \dots, X_m$ együttesen függetlenek, és legyen $X_i \sim \mathrm{Bin}(n; p_i)$ valamilyen $n \in \mathbb{N}$ és $0 < p_i < 1$ számokra $(i = 1, \dots, m)$. Így értelmes többdimenziós eloszlást kapunk, de a binomiális eloszlás általánosításának nem ez az egyetlen módja.

A.8.1. Példa. Átcímkéztünk egy szabályos dobókockát: egy 1-es, két 2-es és három 3-mas számjegyet írtunk rá. Dobjunk 13-szor a kockával. Jelölje X_i a dobott i számjegyek számát. Mi a valószínűsége, hogy $X_1 = 3$, $X_2 = 4$ és $X_3 = 6$?

A valószínűség kombinatorikus módon meghatározható:

$$\mathbb{P}(X_1 = 3, X_2 = 4, X_3 = 6) = \frac{13!}{3!4!6!}\left(\frac{1}{6}\right)^3\left(\frac{1}{3}\right)^4\left(\frac{1}{2}\right)^6 \approx 0{,}05364,$$

hiszen a 3 db 1-es, 4 db 2-es és 6 db 3-mas lehetséges elhelyezéseinek száma $\frac{13!}{3!4!6!}$ (ismétléses permutáció), és az ilyen esetek valószínűsége $p_1^3 p_2^4 p_3^6$, ahol az i dobás valószínűsége p_i.

A.8.2. Definíció. Az $\underline{X} = (X_1, \dots, X_m)$ valószínűségi vektorváltozó **polinomiális** (más néven: multinomiális) eloszlású, $n \in \mathbb{N}$ és $(p_1, p_2, \dots, p_m) \in [0,1]^m$ paraméterekkel, ha $p_1 + \dots + p_m = 1$ és

$$\mathbb{P}(X_1 = k_1, \dots, X_m = k_m) = \frac{n!}{k_1! k_2! \dots k_m!} p_1^{k_1} \dots p_m^{k_m}$$

minden $0 \le k_i \le n$ $(i = 1, \dots, m)$, $\sum_{i=1}^m k_i = n$ értékek esetén.

Ha $m = 2$ és $(p_1, p_2) = (p, 1-p)$ valamilyen $p \in [0,1]$ esetén, akkor X_1 eloszlása $\mathrm{Bin}(n; p)$ (az X_2 pedig nem hordoz extra információt, hiszen $X_2 = n - X_1$).

Világos, hogy az X_i változók nem függetlenek (hiszen például $X_1, \dots, X_{m-1}$ egyértelműen meghatározza X_m-et), ugyanakkor az $\underline{X}$ peremeloszlásai mind $\mathrm{Bin}(n; p_i)$ binomiális eloszlások. Ez a példa is mutatja, hogy a peremeloszlások nem határozzák meg az együttes eloszlást, továbbá, hogy nem mindig az együttesen független koordináták adják egy eloszlás természetes többváltozós általánosítását.

A.9. Marshall–Olkin-féle kétváltozós exponenciális eloszlás

A.9.1. Definíció. Legyenek Y_1, Y_2, Y_3 együttesen független valószínűségi változók, ahol $Y_i \sim \mathrm{Exp}(\lambda_i)$, $(i = 1, 2, 3)$. Definiáljuk az $\underline{X} = (X_1, X_2)$ vektorváltozót: $X_1 = \min(Y_1, Y_3)$ és $X_2 = \min(Y_2, Y_3)$. Az $\underline{X}$ eloszlását **Marshall–Olkin-féle kétváltozós exponenciális eloszlás**nak (röviden Marshall–Olkin-eloszlásnak) hívják.[66]

A motiváció a következő: ha X exponenciális eloszlású, akkor teljesíti az örökifjúság feltételét, azaz $\mathbb{P}(X > t + s \mid X > s) = \mathbb{P}(X > t)$ minden $s, t > 0$ esetén. Ennek lehetséges általánosítása a

$$\mathbb{P}(\underline{X} > \underline{t} + \underline{s} \mid \underline{X} > \underline{s}) = \mathbb{P}(\underline{X} > \underline{t})$$

feltétel, ahol $\underline{t}, \underline{s} \in [0, \infty)^2$, és a vektorok közti $>$ reláció akkor teljesül, ha mindkét koordinátában külön-külön teljesül. Ez a fajta örökifjúság meghatároz egy értelmes kétdimenziós eloszlást: azt, aminek a koordinátái független, exponenciális eloszlású valószínűségi változók (vagyis ez nem a fenti Marshall–Olkin-eloszlás). Alternatív általánosítás viszont a következő feltétel:

$$\mathbb{P}(\underline{X} > t \cdot \underline{1} + \underline{s} \mid \underline{X} > \underline{s}) = \mathbb{P}(\underline{X} > t \cdot \underline{1}), \tag{A.6}$$

ahol $\underline{s} \in [0, \infty)^2$, $t \ge 0$ és $\underline{1} = (1, 1)$. Ezt a tulajdonságot a független, exponenciális eloszlású koordinátákkal bíró valószínűségi vektorváltozón túl a fenti Marshall–Olkin-eloszlás is teljesíti.

[66] Érdekesség, hogy a Marshall–Olkin-eloszlás nem folytonos, azaz nincs együttes sűrűségfüggvénye. Ennek az az oka, hogy a két koordináta pozitív eséllyel megegyezhet. Lásd A.W. Marshall, I. Olkin, A generalized bivariate exponential distribution, J. Appl. Probab. 4 (1967) 291–302.

A.9.2. Példa. Egy gépben két fontos alkatrész van. Jelölje X_1 és X_2 a két alkatrész (véletlen) élettartamát. Tegyük fel, hogy az alkatrészek kora nem befolyásolja, hogy elromlanak-e t idő alatt, vagyis ha az első alkatrész s_1 idős, a második s_2 idős, akkor annak a valószínűsége, hogy t ideig nem romlik el egyik alkatrész sem, ugyanaz, mintha mindkét alkatrész új lenne. Egyenlettel:

$$\mathbb{P}\big((X_1, X_2) > (t+s_1, t+s_2) \mid (X_1, X_2) > (s_1, s_2)\big) = \mathbb{P}\big((X_1, X_2) > (t, t)\big)$$

tetszőleges $s_1, s_2, t \in [0, \infty)$ esetén. Ez éppen az előző (A.6) egyenlet, azaz (X_1, X_2) Marshall–Olkin-eloszlású is lehet, valamilyen $\lambda_1, \lambda_2, \lambda_3$ paraméterekkel. (Szemléletesen, az Y_3 azt a közös hatást reprezentálja, ami mindkét alkatrészt egyszerre elronthatja.)

A.10. Többdimenziós normális eloszlás

Bár számos többdimenziós eloszlásról lehetne beszélni, a legnevezetesebbet nem hagyhatjuk ki, ez a többváltozós normális eloszlás.

Hogyan tudnánk általánosítani a normális eloszlást kétdimenziós eloszlásként? Az egydimenziós normális eloszlás tipikusan egy fizikai mérés eredményének a tényleges érték körüli szóródását (hibáját) írja le. A kétdimenziós általánosítás meghatározásához tekintsünk egy kétdimenziós mérési eredményt, például egy olyan jeladó X szélességi és Y hosszúsági koordinátáit, aminek helyzetét nem ismerjük pontosan, de a jel alapján bemérjük. Idealizált esetben milyen tulajdonságot várnánk ettől az eloszlástól?

Egyrészt feltesszük, hogy az eloszlás *folytonos*, azaz létezik az $f_{X,Y}$ együttes sűrűségfüggvény. Az egyszerűség kedvéért legyen a jeladó tényleges helye az origó. Természetes feltételezés, hogy az eloszlás *forgásszimmetrikus*, azaz $f_{X,Y}$ értéke csak (x, y) hosszától függ. Egyenlettel:

$$f_{X,Y}(x, y) = h(x^2 + y^2) \tag{A.7}$$

valamilyen h valós függvényre. Másrészt, nem irreális feltétel az sem, hogy X és Y *függetlenek*, vagyis hogy az x és y koordinátában mért hibák nem befolyásolják egymást. Az X és Y függetlensége ekvivalensen:

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y) \qquad (\forall x, y \in \mathbb{R}). \tag{A.8}$$

Megmutatjuk, hogy ezek a feltételek meghatározzák az eloszlást.

A.10.1. Állítás. *Ha (X, Y) folytonos valószínűségi vektorváltozó, ami forgásszimmetrikus, és az X, Y koordináták függetlenek, akkor $f_{X,Y}(x, y) = e^{a(x^2+y^2)-c}$ valamilyen $a, c \in \mathbb{R}$ esetén, ahol $a < 0$.*

Bizonyítás. Helyettesítsünk $y = 0$-t a (A.7) és (A.8) egyenletekbe:

$$h(x^2 + 0^2) = f_{X,Y}(x, 0) = f_X(x) \cdot f_Y(0),$$

tehát $f_X(x) = \frac{1}{f_Y(0)} h(x^2)$ $(x \in \mathbb{R})$. Közben felhasználtuk, hogy ha $f_Y(0) = 0$ lenne, akkor h azonosan nulla, ami lehetetlen. Hasonlóan, $f_Y(y) = \frac{1}{f_X(0)} h(y^2)$ $(y \in \mathbb{R})$. Visszahelyettesítve,

$$h(x^2 + y^2) = f_X(x) \cdot f_Y(y) = \frac{1}{f_Y(0)} h(x^2) \cdot \frac{1}{f_X(0)} h(y^2).$$

Jelöljük ezt át a következőképp: $u = x^2$, $v = y^2$ és $c = \ln\big(f_X(0) f_Y(0)\big)$. Ekkor a fenti egyenlet logaritmusa:

$$\ln h(u + v) = \ln h(u) + \ln h(v) - c.$$

Legyen $G(u) = \ln h(u) - c$. Az utolsó egyenletből c-t levonva mindkét oldalról $G(u+v) = G(u) + G(v)$ adódik. Ez ugyanaz a Cauchy-egyenlet, amiről korábban már beszéltünk. Integrálható megoldása ennek csak a $G(u) = a \cdot u$ függvény, valamilyen $a \in \mathbb{R}$ esetén. Tehát $h(u) = e^{au-c}$, vagyis $f_{X,Y}(x, y) = e^{a(x^2+y^2)-c}$ valamilyen $a, c \in \mathbb{R}$ esetén. Ha $a \geq 0$ lenne, akkor nem lehetne $f_{X,Y}$ integrálja 1. □

A paraméterek alkalmas megválasztásával adódik a standard normális eloszlás. Általánosan, n-dimenziós esetben ez a következő.

A.10.2. Definíció. Az $\underline{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó **n-dimenziós standard normális eloszlás**ú, ha folytonos, és együttes sűrűségfüggvénye:

$$f_{\underline{X}}(x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2} \qquad (x_1, \dots, x_n \in \mathbb{R}).$$

Hogyan kapjuk a nem feltétlenül standard, többdimenziós normális eloszlásokat?

A.10.3. Definíció. Az $\underline{Y} = (Y_1, \dots, Y_n)$ valószínűségi vektorváltozó **többdimenziós normális eloszlás**ú, ha létezik $\underline{\underline{A}} \in \mathbb{R}^{n\times n}$, $\underline{\mu} \in \mathbb{R}^n$ és $\underline{X}$ n-dimenziós standard normális eloszlású valószínűségi vektorváltozó, amire

$$\underline{Y} = \underline{\underline{A}} \cdot \underline{X} + \underline{\mu},$$

$\underline{X}$-et oszlopvektorként kezelve. Az $\underline{Y}$ eloszlása **nemelfajuló**, ha $\underline{\underline{A}}$ választható nemelfajuló mátrixnak (azaz $\det(\underline{\underline{A}}) \neq 0$).

Ez a leírásmód eltér az egydimenziós esetben alkalmazott paraméterezéstől, ahol egy (nem feltétlenül standard) normális eloszlást a várható értékével és a szórásnégyzetével adtunk meg. Ezek analógja a többdimenziós esetben a várhatóérték-vektor, illetve a kovarianciamátrix.

A.10.4. Állítás. *Legyen $\underline{X} = (X_1, \dots, X_n)$ standard normális eloszlású valószínűségi vektorváltozó, és $\underline{Y} = \underline{\underline{A}} \cdot \underline{X} + \underline{\mu}$. Ekkor $\mathbb{E}\underline{Y} = \underline{\mu}$ és $\operatorname{cov}(\underline{Y}) = \underline{\underline{A}} \cdot \underline{\underline{A}}^T$.*

Ezekkel a paraméterekkel felírható a többdimenziós normális eloszlás sűrűségfüggvénye is.

A.10.5. Állítás. *Legyen $\underline{Y}$ nemelfajuló n-dimenziós normális eloszlású vektorváltozó. Jelölje a várhatóérték-vektorát $\underline{\mu}$, a kovarianciamátrixát $\underline{\underline{\Sigma}}$. Ekkor $\underline{Y}$ sűrűségfüggvénye*

$$f_{\underline{Y}}(x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(\underline{\underline{\Sigma}})^{\frac{1}{2}}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^T \underline{\underline{\Sigma}}^{-1} (\underline{x}-\underline{\mu})},$$

ahol $\det(\underline{\underline{\Sigma}})$ a $\underline{\underline{\Sigma}}$ determinánsa, $\underline{\underline{\Sigma}}^{-1}$ pedig az inverz mátrixa.

A kitevőben a szorzat egy hármas mátrixszorzat (vektor, mátrix és megint vektor tagokkal), ami valós számot eredményez. A mátrix tag kétdimenziós esetben:

$$\underline{\underline{\Sigma}} = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \qquad \Rightarrow \qquad \underline{\underline{\Sigma}}^{-1} = \frac{1}{\det\left(\underline{\underline{\Sigma}}\right)} \begin{pmatrix} c & -b \\ -b & a \end{pmatrix} \qquad \det\left(\underline{\underline{\Sigma}}\right) = ac - b^2,$$

ahol $a = \mathbb{D}^2(Y_1)$, $b = \operatorname{cov}(Y_1, Y_2)$ és $c = \mathbb{D}^2(Y_2)$.

Az állítás fontos következménye, hogy egy nemelfajuló normális eloszlást meghatároz a $\underline{\mu}$ várhatóérték-vektora és a $\underline{\underline{\Sigma}}$ kovarianciamátrixa. (Vegyük észre, hogy adott $\underline{\underline{\Sigma}}$ többféle $\underline{\underline{A}}$ mátrixból is előállhat, ezért ez nem nyilvánvaló állítás.) Valójában az elfajuló esettel is ez a helyzet, de ekkor nincs sűrűségfüggvényünk, de ezzel itt részletesebben nem foglalkozunk.

A fentiek miatt értelmes a következő jelölés:

Jelölés. Az n-dimenziós normális eloszlást $N(\underline{\mu}, \underline{\underline{\Sigma}})$ jelöli, ahol $\underline{Y} = \underline{\underline{A}} \cdot \underline{X} + \underline{\mu}$, $\underline{X}$ n-dimenziós standard normális, és $\underline{\underline{\Sigma}} = \underline{\underline{A}} \cdot \underline{\underline{A}}^T$. Speciálisan, a standard normális eloszlás jelölése $N(\underline{0}, \underline{\underline{I}})$, ahol $\underline{0}$ az n-dimenziós nullvektor, és $\underline{\underline{I}}$ az n-dimenziós egységmátrix.

Vegyük észre, hogy sem a standard, sem az általános esetben nem beszéltünk még az Y_i koordináták eloszlásáról, sőt szóba sem került az egydimenziós normális eloszlás. Kérdés tehát, hogy mik a normális eloszlás marginálisai? A válasz mérsékelten meglepő:

A.10.6. Állítás. *Legyen $\underline{Y} \sim N(\underline{\mu}, \underline{\underline{\Sigma}})$, ahol $\underline{\mu} \in \mathbb{R}^n$ és $\underline{\underline{\Sigma}} \in \mathbb{R}^{n\times n}$. Ekkor $Y_i \sim N(\mu_i, \Sigma_{i,i})$.*

A standard esetben ennél többet is tudunk: mivel a sűrűségfüggvény szorzattá bomlik (hiszen $\frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{i=1}^n x_i^2} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_i^2}$), így az X_i koordináták együttesen független, egydimenziós standard normális eloszlásúak. Vagyis a normális eloszlásnál teljesül az a szép tulajdonság, ami a polinomiálisnál vagy a Marshall–Olkin-eloszlásnál nem: a természetes többdimenziós általánosítás az egydimenziós eloszlások együttesen független példányai, vektorba rendezve.

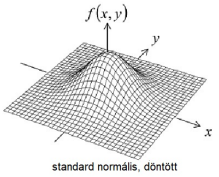
A normális eloszlás több egyéb tulajdonsága okán is a „túl szép, hogy igaz legyen” díjas eloszlás első számú jelöltje; ezeket a tulajdonságokat a következő állításban foglaljuk össze:

A.10.7. Következmény. *Legyen* $(Y_1, Y_2) \sim N(\underline{\mu}, \underline{\underline{\Sigma}})$ *kétdimenziós normális eloszlású valószínűségi vektorváltozó. Ekkor*

(1) tetszőleges $c_1, c_2 \in \mathbb{R}$ *esetén,* $c_1Y_1 + c_2Y_2$ *egydimenziós normális eloszlású, vagy konstans,*
(2) ha $\operatorname{corr}(Y_1, Y_2) = 0$*, akkor* Y_1 *és* Y_2 *függetlenek,*
(3) az $\mathbb{E}(Y_2 \mid Y_1)$ *regresszió megegyezik az* Y_2*-nek az* Y_1*-re vett lineáris regressziójával, azaz*

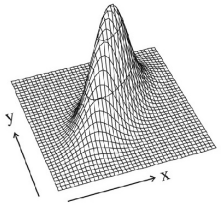
$$\mathbb{E}(Y_2 \mid Y_1) = \frac{b}{a} Y_1 + \left(\mu_2 - \frac{b}{a}\mu_1\right), \qquad \text{ahol} \quad \underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \underline{\underline{\Sigma}} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

Az eloszlás vizualizációjáról még érdemes szót ejteni: hogyan is néz ki egy normális eloszlás sűrűségfüggvénye, például kétdimenziós esetben?



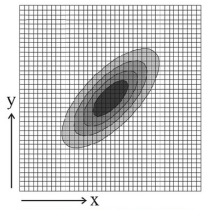
standard normális, döntött

A standard esetben egy „domb” az origó körül (ahogy egydimenziós esetben is), ami forgásszimmetrikus, azaz a szintvonalai körök. Nem standard esetben a szintvonalak ellipszisek lesznek. Tehát a nem standard normális eloszlás nem feltétlenül forgásszimmetrikus, de továbbra is tengelyesen szimmetrikus az ellipszis(ek) főtengelyeire. Tekintsük az egyik ilyen ellipszist.



nem-standard normális, döntött

Az egyszerűség kedvéért tegyük fel, hogy $\underline{\mu} = \underline{0}$, vagyis az ellipszis középpontja az origó. Az ellipszis főtengelyei egymásra merőlegesek, így létezik olyan $\underline{\underline{U}} \in \mathbb{R}^{2\times 2}$ ortogonális transzformáció, ami a főtengelyeket átviszi a koordinátatengelyekbe. Kiszámolható, hogy ekkor $\underline{\underline{U}} \cdot \underline{Y} \sim N(\underline{0}, \underline{\underline{D}})$, ahol $\underline{\underline{D}}$ diagonális mátrix. A következmény második pontja szerint ekkor $\underline{\underline{U}} \cdot \underline{Y}$ két koordinátája független. Összefoglalva, megfelelő koordináta-rendszert választva minden normális eloszlás független, egydimenziós, normális eloszlású valószínűségi változókból áll.



nem-standard normális, felülnézet

A diagonalizálással kapott független valószínűségi változók szórásai implicit módon korábban is megjelentek a normális eloszlás felírásában: ha $\underline{\underline{D}} = \operatorname{diag}(\sigma_1^2, \sigma_2^2)$, akkor a sűrűségfüggvényben megjelenő $\det\left(\underline{\underline{\Sigma}}\right)^{\frac{1}{2}}$ éppen $\sigma_1 \cdot \sigma_2$, azaz a szórások szorzata. A kovarianciamátrix determinánsa nem változik ortogonális transzformáció alkalmazása esetén, így mindegy, hogy az eredeti $\underline{Y}$ vagy a transzformált $\underline{\underline{U}} \cdot \underline{Y}$ kovarianciamátrixáról beszélünk. Vizuálisabban, ez méri az ellipszis területének az egységkör területéhez viszonyított arányát.

Többdimenziós eloszlások esetén a (teljes) variancia mérésére a kovarianciamátrix determinánsa mellett a $\operatorname{Tr}\left(\underline{\underline{\Sigma}}\right)$ nyoma is használatos mennyiség. A diagonalizált változó szórásaival kifejezve $\operatorname{Tr}\left(\underline{\underline{\Sigma}}\right) = \sigma_1^2 + \sigma_2^2$. Szemléletesen, ez az $\underline{Y}$-nak a $\underline{\mu}$-től való eltérésének az átlagos hossznégyzetét méri.

A többdimenziós normális eloszlás további mélységeiért lásd:

- J.K. Patel, C.B. Read, Handbook of the Normal Distribution, CRC Press, 1982.
- Y.L. Tong, The Multivariate Normal Distribution, Springer, 1990.

Statisztika

Tóbiás András

2026. június 26-i állapot

Ezen dokumentum a Mészáros Szabolcs-féle Valószínűségszámítás előadásjegyzetet egészíti ki az abban nem szereplő statisztika anyagrésszel. Aki a szövegben hibát talál, kérem, jelezze nekem.

Tartalomjegyzék

10. A matematikai statisztika alapfogalmai **1**
10.1. Minta és realizáció . . . 3
10.2. Alapstatisztikák . . . 4
10.3. Empirikus eloszlásfüggvény . . . 7

11. Pontbecslések tulajdonságai, átlag és szórás becslései **8**

12. Konfidenciaintervallumok **11**
12.1. Alapfogalmak: konfidenciaintervallum, folytonos eloszlás kvantilisei . . . 11
12.2. Konfidenciaintervallum normális eloszlás várható értékére ismert szórás esetén . . . 12

13. Hipotézisvizsgálati módszerek **14**
13.1. Bevezetés . . . 14
13.2. u-próba . . . 16

A. Példák becsléselméleti módszerekre **22**
A.1. Maximum likelihood-becslés . . . 22
A.2. Momentumbecslés . . . 28

B. További konfidenciaintervallumok, t- és χ^2-eloszlás, t-próba **30**
B.1. χ^2-eloszlás, Student-eloszlás és tulajdonságaik . . . 30
B.2. Konfidenciaintervallum normális eloszlás szórásnégyzetére ismert várható érték esetén . . . 31
B.3. Konfidenciaintervallum normális eloszlás várható értékére ismeretlen szórás esetén . . . 33
B.4. t-próba . . . 33

C. Konfidenciaintervallumokhoz és hipotézisvizsgálatokhoz képletek **38**

10. A matematikai statisztika alapfogalmai

Az előadás eddigi részében a valószínűségi változók eloszlása ismert volt. A matematikai statisztikára viszont az a jellemző, hogy a valószínűségi változók valamilyen mérési eredményeknek felelnek meg, és ezért az eloszlásukat nem ismerjük pontosan. Gyakran előfordul, hogy elméleti megfontolások alapján tudjuk, hogy jó közelítéssel milyen típusú eloszlásúak a mérési eredményeink, de az eloszlás valamilyen ismeretlen

ϑ *paramétertől* függ, amelynek lehetséges értékei a $\Theta \subseteq \mathbb{R}^d$ paramétertartomány elemei lehetnek valamely $d \geq 1$-re.[1] Például:

- Ha az utcán találunk egy pénzérmét, amiről nem tudjuk, hogy szabályos-e, az egy ismeretlen $\vartheta \in \Theta = [0,1] \subset \mathbb{R}^1$ valószínűséggel mutat fejet; a pénzérmét feldobva próbálhatjuk ϑ értékét megbecsülni, illetve igazolni vagy cáfolni, hogy a pénzérme szabályos. Az

$$\mathbb{1}_{\{\text{a pénzérme egy adott feldobásának eredménye fej}\}}$$

  indikátorváltozó paramétere tehát az ismeretlen ϑ.

- A Magyarországon egy adott hónapban vasúti útátjárókban bekövetkező balesetek számáról feltehetjük, hogy jó közelítéssel Poisson-eloszlású, mert a közúton közlekedők száma nagy, egy-egy közlekedő kis valószínűséggel szenved ilyen balesetet, és a többi közlekedőtől többé-kevésbé függetlenül. A Poisson-eloszlás paramétere egy *ismeretlen* $\vartheta \in \Theta = (0, \infty) \subset \mathbb{R}^1$ paraméter.

- A BME egy véletlenszerűen kiválasztott női hallgatójának testmagasságát modellezhetjük normális eloszlással[2]. Itt alapesetben a $\mu \in \mathbb{R}$ várható érték és a $\sigma^2 > 0$ szórásnégyzet is ismeretlen, így a paramétertartomány

$$\Theta = \{(\mu, \sigma^2) \in \mathbb{R}^2 \colon \sigma^2 > 0\} \subset \mathbb{R}^2.$$

  (Persze nyilvánvaló, hogy μ-nek a gyakorlatban pozitívnak kell lennie, de ettől most az általánosság kedvéért eltekintünk.)

A matematikai statisztikában jellemző, hogy az ismeretlen paraméter tanulmányozása céljából *mintát veszünk*, azaz „generálunk" olyan $X_1, \ldots, X_n$ független, azonos eloszlású valószínűségi változókat, amelyek az ismeretlen paraméterű eloszlást követik. Például az előbbi példákban:

- A pénzérmét n-szer feldobjuk, ekkor $i = 1, \ldots, n$ esetén X_i értéke legyen 1, ha az i-edik dobás eredménye fej, és 0, ha írás. Ekkor az $X_1, \ldots, X_n$ *mintaelemek* független és azonos eloszlású indikátorváltozók, amelyek paramétere az ismeretlen ϑ.

- A Magyarországon vasúti útátjárókban bekövetkező balesetek havi számát n hónapon át megfigyeljük, legyen X_i az i-edik hónapban bekövetkező balesetek száma. Ekkor $X_1, \ldots, X_n$ (többé-kevésbé) függetlenek és (nagyjából) Poisson(ϑ) eloszlásúak.

- A BME női hallgatóinak névsorából egyenletesen és egymástól függetlenül választva kisorsolunk n hallgatót, legyen X_i az i-edik hallgató testmagassága. Ekkor $X_1, \ldots, X_n$ (hozzávetőleg) függetlenek és (nagyjából) $\mathrm{N}(\mu; \sigma^2)$ eloszlásúak, ahol mindkét paraméter ismeretlen.

A matematikai statisztika két legfontosabb ága a *becsléselmélet* és a *hipotézisvizsgálat*. A becsléselmélet célkitűzése az ismeretlen paraméter értékének a minta alapján történő, minél pontosabban meghatározása. (A „minél pontosabban" persze matematikailag egyáltalán nem precíz kifejezés, később legalábbis konkrét becsléselméleti módszereknél tisztázni fogjuk, hogy mit is értünk ez alatt.) A hipotézisvizsgálat célja valamilyen előzetesen rögzített hipotézis (feltevés) igazolása vagy megcáfolása a minta segítségével. Ilyen hipotézis lehet például:

- A pénzérme szabályos.

- Az egy hónapban bekövetkező balesetek átlagos száma 2.

- A BME női hallgatóinak átlagos testmagassága legfeljebb 166 cm.

- A BME női hallgatói testmagasságának szórása 3 cm.

[1] A ϑ és a Θ a görög „teta" betű kétféle kisbetűs írásmódja.

[2] Ugyanígy egy véletlenszerűen kiválasztott férfi hallgatóét is (más várható értékkel), azonban a BME egy tetszőleges hallgatójáét nem, mert a nők és a férfiak átlagos testmagassága különbözik, ami 2 db lokális maximummal rendelkező – tehát közelítőleg sem normális – sűrűségfüggvényhez vezet.

10.1. Minta és realizáció

A minta fogalmát úgy definiáljuk, hogy abban nem utalunk a ϑ ismeretlen paraméterre. Ez azért lesz hasznos, mert a későbbiekben a matematikai statisztika olyan módszereivel is fogunk találkozni, amelyek nemcsak egy ismeretlen paramétertől függő eloszlásra, hanem teljesen ismeretlen eloszlásokra is alkalmazhatóak. (Például a 10.2. és a 10.3. fejezetekben nem fog szerepet játszani ismeretlen paraméter.)

10.1.1. Definíció. Legyenek $X_1, \dots, X_n$ független, azonos eloszlású változók, amelyek peremeloszlása nem feltétlenül ismert. Ekkor az

$$\mathbf{X} = (X_1, \dots, X_n)$$

valószínűségi vektorváltozót n **elemű független, azonos eloszlású mintának** (röviden n **elemű fae. mintának**) nevezzük.

Most pedig bevezetünk néhány hasznos jelölést a paraméteres esetre:

10.1.2. Jelölés. Legyen $\mathbf{X} = (X_1, \dots, X_n)$ egy n elemű fae. minta, ahol a mintaelemek eloszlása valamilyen ϑ paramétertől függ (azaz az eloszlásuk a ϑ ismeretében meghatározható), ahol a ϑ lehetséges értékeinek halmaza $\Theta \subseteq \mathbb{R}^d$ valamely $d \geq 1$-re. Ekkor egy $\vartheta \in \Theta$ paraméter esetén:

(1) abban az esetben, ha ϑ a valódi paraméter, bármely A esemény valószínűségét $\mathbb{P}_\vartheta(A)$-val jelöljük,

(2) az X_1 valószínűségi változó ϑ paraméterhez tartozó eloszlásfüggvényét F_ϑ-val jelöljük. (Az eloszlásfüggvény definíciója alapján az előbbi jelöléssel tehát $F_\vartheta(x) = \mathbb{P}_\vartheta(X_1 < x)$, ha $x \in \mathbb{R}$.)

(3) ha X_1-nek ezen paraméter esetén létezik sűrűségfüggvénye, akkor ezt f_ϑ-val jelöljük. Vagyis:

$$f_\vartheta(x) = \begin{cases} F'_\vartheta(x), & \text{ha } F_\vartheta \text{ differenciálható } x\text{-ben,} \\ 0, & \text{különben,} \end{cases}$$

(4) illetve ha X_1-nek ezen paraméter esetén diszkrét az eloszlása, akkor p_ϑ-val jelöljük az ehhez az eloszláshoz tartozó súlyfüggvényt. Vagyis $x \in \mathbb{R}$ esetén $p_\vartheta(x)$ jelöli annak a valószínűségét, hogy $X_1 = x$, feltéve, hogy ϑ a valódi paraméter. Az előbbi jelöléssel: $p_\vartheta(x) = \mathbb{P}_\vartheta(X_1 = x)$.

10.1.3. *Példa.*

- Az utcán talált pénzérme esetén $p_\vartheta(0) = 1 - \vartheta$ és $p_\vartheta(1) = \vartheta$, $\vartheta \in [0, 1]$.
- Az útátjárókban bekövetkező balesetek száma esetén

$$p_\vartheta(k) = \frac{\vartheta^k}{k!} \mathrm{e}^{-\vartheta}, \qquad \vartheta > 0, k = 0, 1, \dots.$$

- A BME női hallgatóinak magassága esetén az ismeretlen paraméterünk (μ, σ^2), így tehát

$$f_{(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}, \qquad \mu \in \mathbb{R}, \sigma^2 > 0.$$

Az eloszlásfüggvények felírását az olvasóra bízzuk.

Sokszor a mintaelemek értékkészlete is függ a ϑ paramétertől, amint ezt a következő példa is mutatja (ellentétben az eddigi példákkal).

10.1.4. *Példa.* Egy barátunk számítógéppel 5 db $\mathrm{U}(0; 1)$ eloszlású véletlen számot generált, majd a kapott számokat mind megszorozta ugyanazzal a szigorúan 1 és 2 közé eső ϑ számmal, amit ő választott és nekünk nem árult el. Az így kapott $\mathbf{x} = (x_1, \dots, x_5) \in \mathbb{R}^5$ vektort (vagyis az eredetileg generált véletlen számok ϑ-szorosainak vektorát) átadta nekünk. Ez a vektor tehát egy 5 elemű $\mathbf{X} = (X_1, \dots, X_5)$ fae. mintából származik, annak egy *realizációját* adja. A minta elemei függetlenek és $\mathrm{U}(0; \vartheta)$ eloszlásúak (világos? Ha nem, határozzuk meg az egyes mintaelemek sűrűségfüggvényét!).

A ϑ paraméter ismeretében az $X_1, \dots, X_n$ valószínűségi változók értékkészlete (ahol a sűrűségfüggvényük pozitív) a $(0, \vartheta)$ intervallum, ami tehát függ ϑ-tól.

Ahhoz, hogy a *realizáció* szó precíz definícióját megadhassuk, először bevezetjük a valószínűségi változó (ϑ paramétertől függő) *lényeges értékei* halmazának fogalmát. Diszkrét esetben ezzel a definícióval a regresszió témakörénél már találkoztunk (lásd a 11.1.5. Definíciót Mészáros Szabolcs jegyzetében), csak akkor nem volt szó ismeretlen paraméterről.

10.1.5. Definíció (és jelölés)**.** Legyen $\mathbf{X} = (X_1, \ldots, X_n)$ egy n elemű fae. minta, ahol a mintaelemek eloszlása egy $\vartheta \in \Theta \subseteq \mathbb{R}^d$ paramétertől függ. Legyen $\vartheta \in \Theta$ és $i = 1, \ldots, n$. Ekkor ha létezik az f_ϑ sűrűségfüggvény, akkor definiáljuk az $S_{X_i}^{(\vartheta)}$ halmazt a következőképpen:

$$S_{X_i}^{(\vartheta)} = \{x \in \mathbb{R} |\ f_\vartheta(x) > 0\} \subseteq \mathbb{R}.$$

Ha pedig létezik a p_ϑ súlyfüggvény, akkor definiáljuk az $S_{X_i}^{(\vartheta)}$ halmazt az alábbi képlettel:

$$S_{X_i}^{(\vartheta)} = \{x \in \mathbb{R} |\ p_\vartheta(x) > 0\} \subseteq \mathbb{R}.$$

$S_{X_i}^{(\vartheta)}$-t mindkét esetben az X_i valószínűségi változó ϑ paraméterhez tartozó **lényeges értékei halmazá**nak nevezzük.[3]

10.1.6. *Példa.* Az ismeretlen ϑ valószínűséggel fejet mutató pénzérme n-szeri feldobásából kapott $(X_1, \ldots, X_n)$ fae. mintánál, ahol X_i indikátorváltozó ϑ paraméterrel, $S_{X_i}^{(\vartheta)} = \{0, 1\}$ minden $\vartheta \in \Theta = (0, 1)$-re, $S_{X_i}^{(0)} = \{0\}$ és $S_{X_i}^{(1)} = \{1\}$. Továbbá $p_\vartheta(1) = \vartheta$, $p_\vartheta(0) = 1 - \vartheta$ és $p_\vartheta(x) = 0$ minden $x \in \mathbb{R} \setminus \{0, 1\}$ esetén.
A 10.1.4 példában szereplő $(X_1, \ldots, X_n)$ fae. mintánál $(n = 5)$, ahol X_i egyenletes eloszlású a $(0, \vartheta)$ intervallumon, $S_{X_i}^{(\vartheta)} = (0, \vartheta)$ és

$$f_\vartheta(x) = \begin{cases} \frac{1}{\vartheta}, & \text{ha } x \in (0, \vartheta), \\ 0, & \text{különben} \end{cases}$$

teljesül minden $\vartheta \in \Theta = (1, 2)$ paraméterre.

10.1.7. Definíció. Legyen $\mathbf{X} = (X_1, \ldots, X_n)$ egy n elemű fae. minta, ahol a mintaelemek eloszlása egy $\vartheta \in \Theta \subseteq \mathbb{R}^d$ paramétertől függ. Az $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ az $\mathbf{X} = (X_1, \ldots, X_n)$ minta egy (lehetséges) **realizációja** a $\vartheta \in \Theta$ paraméter esetén, ha $x_i \in S_{X_i}^{(\vartheta)}$ teljesül minden $i \in \{1, \ldots, n\}$-re.

10.1.8. *Példa.* Demonstráljuk a realizáció fogalmát a 10.1.6 példával!
Az utcán talált pénzérme esetén az $(x_1, \ldots, x_7) = (1, 0, 0, 0, 1, 1, 0)$ az $n = 7$ elemű $\mathbf{X} = (X_1, \ldots, X_7)$ fae. minta egy realizációja minden $0 < \vartheta < 1$ paraméter esetén (de $\vartheta = 0$ és $\vartheta = 1$ esetén nem).
A $(0, \vartheta)$ intervallumon egyenletes eloszlású $\mathbf{X} = (X_1, \ldots, X_5)$ fae. mintának

$$(x_1, \ldots, x_5) = (0.14, 0.79, 1.13, 1, 1.2),$$

egy lehetséges realizációja, ha $1.2 < \vartheta < 2$, de ha $1 < \vartheta \leq 1.2$, akkor nem.

10.2. Alapstatisztikák

Egy n elemű fae. minta egy statisztikájának a mintaelemek egy olyan függvényét nevezzük, amely szimmetrikus, azaz „minden mintaelemtől ugyanúgy függ". Az utóbbi tulajdonságot formalizálja a következő definíció.

[3] X_i lényeges értékeinek halmaza (adott ϑ esetén) csaknem ugyanaz, mint X_i (mint $\Omega \to \mathbb{R}$ függvény) értékkészlete. Az értékkészlet lehet bővebb, mint a lényeges értékek halmaza, azonban X_i a lényeges értékek halmazán kívüli összes értéket együttvéve 0 valószínűséggel veszi fel.

10.2.1. Definíció. Legyen $\mathbf{X} = (X_1, \ldots, X_n)$ egy n elemű fae. minta. Ha $T\colon \mathbb{R}^n \to \mathbb{R}$ szimmetrikus függvény, azaz

$$T(x_1, \ldots, x_n) = T(x_{\pi(1)}, \ldots, x_{\pi(n)})$$

minden $x_1, \ldots, x_n \in \mathbb{R}$ és $\pi\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$ permutáció[4] esetén teljesül, akkor a $T(\mathbf{X}) = T(X_1, \ldots, X_n)$ valószínűségi vektorváltozót az $X_1, \ldots, X_n$ egy **statisztikájának** nevezzük.

Lássunk most néhány klasszikus, részben már az előadás korábbi részeiből is ismert statisztikát! Az első a mintaátlag, amit már a középiskolából is ismerünk.

10.2.2. *Példa.* Legyen $\mathbf{X} = (X_1, \ldots, X_n)$ egy n elemű fae. minta. Ekkor a 9.2. fejezetben bevezetett

$$\overline{X_n} = \frac{X_1 + \ldots + X_n}{n}$$

mintaátlag az $\mathbf{X}$ egy statisztikája. Ha $\mathbf{x} = (x_1, \ldots, x_n)$ az $\mathbf{X} = (X_1, \ldots, X_n)$ egy realizációja, akkor realizáció átlagát $\overline{x_n}$-nel fogjuk jelölni: $\overline{x_n} = \frac{x_1 + \ldots + x_n}{n}$.

Az előadás eddigi anyaga alapján nem meglepő, hogy a mintaátlag a várható érték egyfajta becslése a minta alapján. Az világos, hogy ha $\mathbb{E}(X_i)$ létezik (vagyis ha $\mathbb{E}[|X_i|] < \infty$), akkor a mintaátlag várható értéke megegyezik az egyes mintaelemek várható értékével:

$$\mathbb{E}\big[\overline{X_n}\big] = \frac{1}{n}(\mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n)) = \mathbb{E}(X_1).$$

Ezen becslés további tulajdonságait a becsléselmélet nyelvezetével a 11. fejezetben fogjuk elemezni.

A következő klasszikus statisztika a korrigált empirikus szórásnégyzet, amely a szórásnégyzet egyfajta közelítése a mintaelemek segítségével.

10.2.3. Definíció. Legyen $\mathbf{X} = (X_1, \ldots, X_n)$ egy n elemű fae. minta. Ekkor az

$$(S_n^*)^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X_n})^2 \tag{10.1}$$

mennyiséget az **X korrigált empirikus szórásnégyzetének**[5] nevezzük, $S_n^* = \sqrt{(S_n^*)^2}$-et pedig az **X korrigált empirikus szórásának**.

Világos, hogy $(S_n^*)^2$ az $\mathbf{X}$ minta egy statisztikája. Az olvasóban felmerülhet a kérdés, hogy miért éppen $n-1$-gyel (és nem például n-nel) osztunk a 10.1 formulában. Ennek oka, hogy így kapunk olyan statisztikát, amelynek várható értéke megegyezik a mintaelemek szórásnégyzetével:

10.2.4. Állítás. *Legyen $\mathbf{X} = (X_1, \ldots, X_n)$ egy n elemű fae. minta, amelyre teljesül, hogy $\mathbb{E}(X_i^2) < \infty$. Ekkor*

$$\mathbb{E}((S_n^*)^2) = \mathbb{D}^2(X_1).$$

Bizonyítás. A bizonyításban szereplő számítások kissé hosszadalmasak, de a várható érték linearitásán kívül kizárólag azt a tulajdonságot használják, hogy független valószínűségi változók szorzatának várható értéke megegyezik a várható értékek szorzatával (6.1.4. Állítás). Először is,

$$\begin{aligned}\mathbb{E}[(S_n^*)^2] &= \frac{1}{n-1} \sum_{i=1}^{n} \mathbb{E}\big(X_i - \overline{X_n}\big)^2 = \frac{1}{n-1} \sum_{i=1}^{n} \Big(\mathbb{E}(X_i^2) - 2\mathbb{E}(X_i \overline{X_n}) + \mathbb{E}(\overline{X_n}^2)\Big) \\ &= \frac{1}{n-1} \sum_{i=1}^{n} \mathbb{E}(X_i^2) - \frac{2}{n-1} \sum_{i=1}^{n} \mathbb{E}(X_i \overline{X_n}) + \frac{n}{n-1} \mathbb{E}(\overline{X_n}^2).\end{aligned}$$

[4] Emlékeztetünk arra, hogy permutációnak az $\{1, \ldots, n\}$ halmaz önmagára való kölcsönösen egyértelmű leképezéseit (vagyis bijekcióit) nevezzük. Ez ekvivalens például a permutációnak a Szeszlér Dávid-féle BSZ1 jegyzetben (`http://cs.bme.hu/bsz1/jegyzet/bsz1_jegyzet.pdf`) a determináns témakörénél szereplő definíciójával, lásd az idei Valszám diasort is a kombinatorika témakörénél.

[5] Az „empirikus" szó jelentése: „tapasztalati". Néha használatos a „korrigált tapasztalati szórásnégyzet" kifejezés is.

Számoljuk ki külön-külön a jobb oldalon szereplő három tag várható értékét! Az első tag:

$$\frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}(X_i^2)=\frac{n}{n-1}\mathbb{E}(X_1^2),$$

mivel $X_1,\dots,X_n$ azonos eloszlásúak. A második tag -1-szerese:

$$\begin{aligned}\frac{2}{n-1}\sum_{i=1}^{n}\mathbb{E}(X_i\overline{X_n})&=\frac{2}{n-1}\Big(\sum_{i=1}^{n}\frac{1}{n}\mathbb{E}(X_i^2)+\frac{1}{n}\sum_{j\neq i}\mathbb{E}(X_iX_j)\Big)=\frac{2}{n-1}\Big(\mathbb{E}(X_1^2)+(n-1)\mathbb{E}(X_1)^2\Big)\\&=\frac{2}{n-1}\mathbb{E}(X_1^2)+2\mathbb{E}(X_1)^2,\end{aligned}$$

mivel $X_1,\dots,X_n$ azonos eloszlásúak és függetlenek. A harmadik tag:

$$\begin{aligned}\frac{n}{n-1}\mathbb{E}(\overline{X_n}^2)&=\frac{n}{(n-1)n^2}\mathbb{E}\Big[\Big(\sum_{i=1}^{n}X_i\Big)^2\Big]=\frac{1}{n(n-1)}\mathbb{E}\Big(\sum_{i=1}^{n}X_i^2+2\sum_{1\le i<j\le n}X_iX_j\Big)\\&=\frac{1}{n(n-1)}\Big(\sum_{i=1}^{n}(\mathbb{E}(X_1^2)+(n-1)\mathbb{E}(X_1)^2)\Big)\\&=\frac{1}{n-1}\mathbb{E}(X_1^2)+\mathbb{E}(X_1)^2.\end{aligned}$$

mivel $X_1,\dots,X_n$ függetlenek és azonos eloszlásúak. Összességében tehát

$$\mathbb{E}[(S_n^*)^2]=\mathbb{E}(X_1^2)\big(\frac{n}{n-1}-\frac{2}{n-1}+\frac{1}{n-1}\big)+\mathbb{E}(X_1)^2(-2+1)=\mathbb{E}(X_1^2)-\mathbb{E}(X_1)^2=\mathbb{D}^2(X_1).$$

□

10.2.5. *Megjegyzés.* Láthatjuk, hogy ha az (10.1) formulában $\frac{1}{n-1}$ helyett $\frac{1}{n}$-nel osztanánk, akkor az így kapott statisztika szórásnégyzete (a várható érték linearitása miatt) nem $\mathbb{D}^2(X_1)$, hanem annak $\frac{n-1}{n}$-szerese lenne. Az $S_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X_n})^2$ statisztikát *empirikus szórásnégyzet*nek (a négyzetgyökét pedig *empirikus szórás*nak) nevezik, $(S_n^*)^2$-et pedig azért hívják *korrigált*nak, mert szórásnégyzete megegyezik $\mathbb{D}^2(X_1)$-gyel. A magyar nyelvű szakirodalomban ez a bevett szóhasználat. Idegen nyelvű művekben előfordulhat az is, hogy $(S_n^*)^2$-et hívják empirikus szórásnégyzetnek, de ettől mi eltekintünk.

Szintén alapstatisztika a minta **módusza**, azaz a leggyakoribb érték a mintában. Ha több ilyen érték van, akkor mindegyiket módusznak tekintjük.

További alapstatisztikák definiálása érdekében először bevezetjük a rendezett minta fogalmát.

10.2.6. Definíció. Legyen $\mathbf{X}=(X_1,\dots,X_n)$ egy n elemű fae. minta. Legyen $(X_1^*,X_2^*,\dots,X_n^*)$ az $X_1,\dots,X_n$ mintaelemek egy olyan felsorolása, hogy

$$X_1^*\le X_2^*\le\ldots\le X_n^*.$$

Ekkor $(X_1^*,X_2^*,\dots,X_n^*)$-t **rendezett mintá**nak nevezzük.

10.2.7. *Példa.* Legyen $\mathbf{x}=(x_1,\dots,x_8)=(1,2,1,3,4,6,5,2)$ egy szabályos dobókocka nyolcszori feldobásából származó $\mathbf{X}=(X_1,\dots,X_8)$ minta egy realizációja. Ekkor ehhez a realizációhoz az $(X_1^*,\dots,X_8^*)$ rendezett minta $(x_1^*,\dots,x_8^*)=(1,1,2,2,3,4,5,6)$ realizációja tartozik. Tehát a rendezett minta realizációja az eredeti (nem rendezett) minta realizációjából egyértelműen megkapható, akkor is, ha vannak ismétlődő elemek.

Fontos kiemelni, hogy $X_1^*,\dots,X_n^*$ nem függetlenek (ezt formálisan nem bizonyítjuk, de érezhető, hogy pl. a legkisebb mintaelem eloszlása befolyásolja az összes többi rendezett mintaelem eloszlását, hiszen azoknak a legkisebbnél nagyobbnak kell lenniük) és jellemzően nem is azonos eloszlásúak. A rendezett mintaelemek segítségével definiálható az empirikus medián, ami páratlan elemszámú mintánál a rendezett minta középső elemét, páros elemszámúnál pedig a két középső számtani közepét jelenti:

10.2.8. Definíció. Az $\mathbf{X} = (X_1, \dots, X_n)$ n elemű fae. minta **empirikus mediánj**án az $m_{n,\mathbf{X}} = X^*_{k+1}$ rendezett mintaelemet értjük, ha $n = 2k+1$, és az $m_{n,\mathbf{X}} = \frac{X^*_k + X^*_{k+1}}{2}$ kifejezést, ha $n = 2k$.

10.2.9. *Példa.* Ha egy n elemű minta realizációja $(x_1, \dots, x_4) = (\sqrt{2}, \pi, -1, \mathrm{e})$, akkor a rendezett minta realizációja $(x^*_1, x^*_2, x^*_3, x^*_4) = (-1, \sqrt{2}, \mathrm{e}, \pi)$ és empirikus mediánja $\frac{\sqrt{2}+\mathrm{e}}{2}$. Ha az n elemű minta realizációja $(x_1, x_2, x_3) = (-1, 3, 2)$, akkor a rendezett minta realizációja $(x^*_1, x^*_2, x^*_3) = (-1, 2, 3)$ és az empirikus medián 2.

10.3. Empirikus eloszlásfüggvény

Ebben az alfejezetben azt tesszük fel, hogy az $X_1, \dots, X_n$ valószínűségi változók függetlenek és azonos eloszlásúak valamilyen ismeretlen eloszlással, ennek eloszlásfüggvénye legyen $x \mapsto F(x) = \mathbb{P}(X_i < x)$. (A ϑ paraméter itt sem fog szerepet játszani.) Az X_i-k eloszlásáról semmilyen ismeretünk nincs, például azt sem tudjuk, hogy az eloszlásuk diszkrét-e, folytonos vagy egyik sem[6]. Hogyan tudjuk az $\mathbf{X} = (X_1, \dots, X_n)$ egy $\mathbf{x} = (x_1, \dots, x_n)$ realizációja alapján az F eloszlásfüggvényt „megbecsülni", úgy, hogy $n \to \infty$ esetén a valódi eloszlásfüggvényhez tartó eredményt kapjunk? Ezt a célt szolgálja az empirikus (tapasztalati) eloszlásfüggvény.

10.3.1. Definíció. Legyen $n \in \mathbb{N}$ és legyenek $X_1, \dots, X_n$ fae. valószínűségi változók. Az $\mathbb{R} \to \mathbb{R}$,

$$x \mapsto F^*_n(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i < x\}} = \frac{1}{n} \big| \{ i \in \{1, \dots, n\} |\ X_i < x \} \big| \tag{10.2}$$

függvényt az $\underline{X} = (X_1, \dots, X_n)$ n elemű fae. mintához tartozó **empirikus eloszlásfüggvény**nek nevezzük.

Fontos az elején rögzíteni, hogy F^*_n egy *véletlen* függvény: a függvény értékei az $X_1, \dots, X_n$ valószínűségi változók értékeitől függenek.

A (10.2) egyenletben szereplő definíciót az $(X^*_1, \dots, X^*_n)$ rendezett minta (lásd 10.2.6. Definíció) segítségével a következő egyszerű, explicit formulával is megfogalmazhatjuk:

$$F^*_n(x) = \begin{cases} 0, & \text{ha } x \le X^*_1, \\ \frac{k}{n}, & \text{ha } X^*_k < x \le X^*_{k+1}, \quad k = 1, \dots, n-1, \\ 1, & \text{ha } x > X^*_n. \end{cases} \tag{10.3}$$

Ez az átírás azt is mutatja, hogy az $\mathbf{X} = (X_1, \dots, X_n)$ bármely fix $\mathbf{x} = (x_1, \dots, x_n)$ realizációja esetén valóban eloszlásfüggvényt kapunk: a $-\infty$-ben vett határérték valóban 0, az ∞-ben vett határérték 1, és a függvény monoton növő és (a (10.2) definícióban szereplő szigorú egyenlőtlenségeknek köszönhetően) balról folytonos.

A következő állítás megmutatja, hogy az empirikus eloszlásfüggvény „jó becslése" a valódi eloszlásfüggvénynek: várható értékben minden pontban megegyezik vele, a szórása pedig n függvényében csökken. Továbbá az empirikus eloszlásfüggvény minden pontban (1 valószínűséggel) a valódi eloszlásfüggvényhez konvergál $n \to \infty$ esetén, vagyis ha kellően nagy elemszámú mintát veszünk, akkor egy tetszőleges $x \in \mathbb{R}$ pontban (1 valószínűséggel) tetszőlegesen megközelíthetjük az eloszlásfüggvény $F(x)$ értékét az empirikus eloszlásfüggvény $F^*_n(x)$ értékével.

10.3.2. Állítás. *Legyen $n \in \mathbb{N}$ és legyenek $X_1, \dots, X_n$ független, azonos eloszlású valószínűségi változók F eloszlásfüggvénnyel. Ekkor minden $x \in \mathbb{R}$ esetén teljesül, hogy*

(1) $\mathbb{E}(F^*_n(x)) = F(x)$,

[6] Utóbbira példa a következőképpen definiált X valószínűségi változó: dobjunk fel egy pénzérmét, és ha fejet kapunk, dobjunk fel egy kockát és legyen X a dobás eredménye, ha pedig írást, akkor legyen X egy $\mathrm{U}(0;1)$ eloszlású, a kockadobástól független véletlen szám.

(2) $\mathbb{D}^2(F_n^*(x)) = \frac{F(x)(1-F(x))}{n}$, *és*

(3) $\mathbb{P}(\lim_{n\to\infty} F_n^*(x) = F(x)) = 1$.

Bizonyítás. Legyen $x \in \mathbb{R}$. Az (10.2) definíció alapján észrevehetjük, hogy $n \cdot F_n^*(x)$ a független, azonos eloszlású $\mathbb{1}_{\{X_i<x\}}$ indikátorváltozók összege, amelyek paramétere (és egyben várható értéke) $\mathbb{P}(X_i < x) = F(x)$. Emiatt $nF_n^*(x) \sim B(n; F(x))$. A binomiális eloszlás, illetve a várható érték és a szórásnégyzet ismert tulajdonságai miatt tehát

$$\mathbb{E}\big(F_n^*(x)\big) = \frac{1}{n}\mathbb{E}\big(nF_n^*(x)\big) = \frac{1}{n}nF(x) = F(x)$$

és

$$\mathbb{D}^2\big(F_n^*(x)\big) = \frac{1}{n^2}\mathbb{D}^2\big(nF_n^*(x)\big) = \frac{1}{n^2}nF(x)(1-F(x)) = \frac{F(x)(1-F(x))}{n},$$

amint állítottuk. A (3) állítás igazolása érdekében alkalmazzuk a független, azonos eloszlású $\mathbb{1}_{\{X_i<x\}}$ indikátorváltozókra a nagy számok erős törvényét (azaz a 9.2.1. Tételt). Ezek várható értéke

$$\mathbb{E}\big(\mathbb{1}_{\{X_1<x\}}\big) = \mathbb{P}(X_1 < x) = F(x)$$

(és a szórásnégyzetük véges, pontosabban $F(x)(1-F(x))$-szel egyenlő), ezért a 9.2.1. Tétel szerint az átlaguk 1 valószínűséggel a várható értékükhöz tart:

$$\mathbb{P}\Big(\lim_{n\to\infty} F_n^*(x) = F(x)\Big) = \mathbb{P}\Big(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i<x\}} = \mathbb{E}\big(\mathbb{1}_{\{X_1<x\}}\big)\Big) = 1.$$

Ezzel a (3) állítást is beláttuk. □

A 10.3.2 Állítás (3) részénél több is igaz: az empirikus eloszlásfüggvény nemcsak hogy minden pontban (1 valószínűséggel) konvergál a valódihoz, hanem (1 valószínűséggel) *az egész számegyenesen egyenletesen konvergál* hozzá, azaz a két függvény közötti távolság nullához tart. Ezt fogalmazza meg a következő tétel:

10.3.3. Tétel (Glivenko–Cantelli-tétel)**.** *Legyen $n \in \mathbb{N}$ és legyenek $X_1, \dots, X_n$ független, azonos eloszlású valószínűségi változók F eloszlásfüggvénnyel. Ekkor az $x \mapsto F_n^*(x)$ függvény 1 valószínűséggel egyenletesen konvergál $x \mapsto F(x)$-hez, azaz*

$$\mathbb{P}\Big(\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \big|F_n^*(x) - F(x)\big| = 0\Big) = 1.$$

Miért erősebb ez a tétel a 10.3.2 Állítás (3) részénél? A 10.3.2. Állításban az a 0 valószínűségi esemény, ahol a (3)-beli konvergencia nem teljesül, elvileg x-től is függ. A Glivenko–Cantelli-tétel alapján azonban ha vesszük ezen események (megszámlálhatatlanul végtelenül sok nulla valószínűségű esemény!) unióját, akkor még mindig 0 valószínűségű eseményt kapunk. A Glivenko–Cantelli tételt ezen az előadáson nem bizonyítjuk, a bizonyítás megtalálható szinte bármelyik matematikai statisztika tankönyben.

11. Pontbecslések tulajdonságai, átlag és szórás becslései

Ebben a fejezetben ismét azt a helyzetet tekintjük, amikor az $X_1, \dots, X_n$ mintaelemek eloszlása egy $\vartheta \in \Theta$ paramétertől függ, és az ismeretlen ϑ paraméterre vagy annak valamilyen $\psi(\vartheta)$ függvényére[7] kívánunk egy n elemű fae. minta valamilyen statisztikája alapján minél jobb becslést adni. Az itt részletezett becslési módszerek idő hiányában nem lesznek kimerítőek; az érdeklődőknek a mérnökinformatikus és gazdaságinformatikus MSc képzés Matematikai statisztika tárgyát ajánljuk.

11.0.1. Definíció. Legyen adott egy $\mathbf{X} = (X_1, \dots, X_n)$ egy n elemű fae. minta, ahol a mintaelemek eloszlása egy ismeretlen $\vartheta \in \Theta \subseteq \mathbb{R}^d$ paramétertől függ, és legyen $\psi\colon \mathbb{R}^d \to \mathbb{R}$ egy függvény. Legyen továbbá $T(\mathbf{X}) = T(X_1, \dots, X_n)$ a minta egy statisztikája.

Azt mondjuk, hogy a $T(\mathbf{X})$ statisztika

[7] A ψ is görög betű, pszí-nek ejtjük.

(1) a $\psi(\vartheta)$ paraméterfüggvény egy **torzítatlan** becslése, ha minden $\vartheta \in \Theta$ esetén

$$\mathbb{E}_\vartheta(T(X_1, \ldots, X_n)) = \psi(\vartheta),$$

ahol $\mathbb{E}_\vartheta$ jelöli a $\mathbb{P}_\vartheta$ valószínűség szerinti várható értéket,

(2) a $\psi(\vartheta)$ paraméterfüggvény egy **aszimptotikusan torzítatlan** becslése, ha minden $\vartheta \in \Theta$ esetén

$$\lim_{n\to\infty} \mathbb{E}_\vartheta(T(X_1, \ldots, X_n)) = \psi(\vartheta),$$

(3) a $\psi(\vartheta)$ paraméterfüggvény egy **erősen konzisztens** becslése[8], ha minden $\vartheta \in \Theta$ esetén

$$\mathbb{P}_\vartheta\big(\lim_{n\to\infty} T(X_1, \ldots, X_n) = \psi(\vartheta)\big) = 1. \tag{11.1}$$

11.0.2. Definíció. A 11.0.1 Definícióban szereplő szituációban ha $T(\mathbf{X})$ és a minta egy másik $T'(\mathbf{X}) = T'(X_1, \ldots, X_n)$ statisztikája is torzítatlan becslése $\psi(\vartheta)$-nak, akkor azt mondjuk, hogy $T(\mathbf{X})$ **legalább annyira hatásos, mint** $T'(\mathbf{X})$, ha

$$\mathbb{D}^2_\vartheta(T(\mathbf{X})) \le \mathbb{D}^2_\vartheta(T'(\mathbf{X})),$$

ahol $\mathbb{D}^2_\vartheta$ jelöli a $\mathbb{P}_\vartheta$ szerinti szórásnégyzetet.[9] Ha $T(\mathbf{X})$ legalább annyira hatásos, mint $\psi(\vartheta)$ bármely torzítatlan becslése, akkor azt mondjuk, hogy $T(\mathbf{X})$ a $\psi(\vartheta)$ paraméterfüggvény egy **hatásos** becslése.

Nyilvánvaló, hogy ha egy statisztika torzítatlan becslést ad egy adott paraméterfüggvényre, akkor aszimptotikusan torzítatlan becslést is ad. Az erős konzisztencia definíciójában szereplő (11.1) egyenlőséget úgy interpretálhatjuk, hogy bármely $\vartheta \in \Theta$-ra igaz, hogy ha ϑ a valódi paraméter, akkor $T(X_1, \ldots, X_n)$ egy valószínűséggel tart $\psi(\vartheta)$-hoz.

Lássunk példát torzítatlan, aszimptotikusan torzítatlan, illetve erősen konzisztens becslésekre!

11.0.3. *Példa.* A 11.0.1. Definícióban szereplő szituációban egy lehetséges ψ függvény a

$$\Theta \ni \vartheta \mapsto \psi(\vartheta) = \mathbb{E}_\vartheta(X_1),$$

ha az $\mathbb{E}_\vartheta(X_1)$ minden $\vartheta \in \Theta$ esetén létezik (azaz ha $\mathbb{E}_\vartheta(|X_1|) < \infty$ minden $\vartheta \in \Theta$ esetén teljesül). Ebben az esetben a

$$T(X_1, \ldots, X_n) = \overline{X}_n = \frac{X_1 + \ldots + X_n}{n}$$

mintaátlag a $\psi(\vartheta) = \mathbb{E}_\vartheta(X_1)$ paraméterfüggvény egy torzítatlan becslése, mivel minden $\vartheta \in \Theta$ esetén

$$\mathbb{E}_\vartheta(\overline{X}_n) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_\vartheta(X_i) = \mathbb{E}_\vartheta(X_1)$$

(lásd a 10.2.2. Példát). Továbbá a nagy számok erős törvénye miatt ha feltesszük, hogy ϑ a valódi paraméter, akkor a mintaátlag 1 valószínűséggel $\psi(\vartheta) = \mathbb{E}_\vartheta(X_1)$-hez konvergál. Formálisan: minden $\vartheta \in \Theta$ esetén

$$\mathbb{P}_\vartheta\Big(\lim_{n\to\infty} \overline{X_n} = \mathbb{E}_\vartheta(X_1)\Big) = 1.$$

Emiatt $T(X_1, \ldots, X_n) = \overline{X_n}$ erősen konzisztens becslése is a $\psi(\vartheta) = \mathbb{E}_\vartheta(X_1)$ paraméternek.

[8] Léteznek más fogalmak is becslések konzisztenciájára, például a *gyenge konzisztencia*, ami ugyanazt fejezi ki, mint az erős konzisztencia, csak az (11.1) egyenletben szereplő 1 valószínűségű konvergencia helyett valószínűségben való konvergenciával. Ha azt mondjuk, hogy egy becslés *konzisztens*, az alatt azt értjük, hogy gyengén konzisztens. Amint a nagy számok törvényeinek témájánál már láttuk, az 1 valószínűségű konvergenciából következik a valószínűségben való konvergencia, így minden erősen konzisztens becslés konzisztens is. Vannak más konzisztenciafogalmak is, amiket itt nem részletezünk, például a *négyzetes középben való konzisztencia*, lásd a mérnök- és gazdaságinformatikus MSc képzés Matematikai statisztika tárgyát.

[9] Hasonlóan és értelemszerűen vezethetők be a „(szigorúan) hatásosabb", a „legfeljebb annyira hatásos" és a „(szigorúan) kevésbé hatásos" relációk két torzítatlan becslés között.

11.0.4. *Példa.* Egy másik lehetséges ψ függvény a

$$\Theta \ni \vartheta \mapsto \psi(\vartheta) = \mathbb{D}^2_\vartheta(X_1),$$

ha $\mathbb{E}_\vartheta(X_1^2)$ minden $\vartheta \in \Theta$ esetén véges. Ebben az esetben a

$$T(X_1, \dots, X_n) = (S_n^*)^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X_n})^2$$

korrigált empirikus szórásnégyzet torzítatlan becslése a

$$\psi(\vartheta) = \mathbb{D}^2_\vartheta(X_1)$$

paraméterfüggvénynek. Valóban, a 10.2.4. Állítás szerint minden $\vartheta \in \Theta$ esetén teljesül, hogy

$$\mathbb{E}_\vartheta((S_n^*)^2) = \mathbb{D}^2_\vartheta(X_1).$$

Így azt is láthatjuk, hogy a

$$T'(X_1, \dots, X_n) = S_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X_n})^2 \tag{11.2}$$

statisztika nem torzítatlan becslése $\mathbb{D}^2_\vartheta(X_1)$-nek, hiszen (a várható érték linearitása miatt, lásd a 10.2.5. Megjegyzést is)

$$\mathbb{E}_\vartheta[S_n^2] = \frac{n-1}{n}\mathbb{E}_\vartheta[(S_n^*)^2] = \frac{n-1}{n}\mathbb{D}^2_\Theta(X_1).$$

Ugyanakkor $T'(X_1, \dots, X_n) = S_n^2$ is aszimptotikusan torzítatlan becslése $\mathbb{D}^2_\vartheta(X_1)$-nek, mert

$$\lim_{n\to\infty} \mathbb{E}_\vartheta[S_n^2] = \lim_{n\to\infty} \frac{n-1}{n}\mathbb{D}^2_\Theta(X_1) = \mathbb{D}^2_\Theta(X_1).$$

11.0.5. *Példa.* Egy becslés torzítatlansága önmagában nem garantálja a becslés jóságát, például a hatásosságát; előfordulhat, hogy egy másik torzítatlan becslés szigorúan hatásosabb nála. Legyen pl. (X_1, X_2, X_3) fae. minta $\mathrm{U}(\vartheta; 1+\vartheta)$ eloszlással, ahol $\vartheta \geq 0$ ismeretlen paraméter, és tekintsük a $T(X_1, X_2, X_3) = X_2^*$ rendezett mintaelemet mint (X_1, X_2, X_3) egy statisztikáját. Gyakorlaton korábban beláttuk, hogy $\vartheta = 0$ esetén X_2^* sűrűségfüggvénye

$$f_{X_2^*}(x) = \begin{cases} 6x - 6x^2, & \text{ha } 0 < x < 1, \\ 0, & \text{különben.} \end{cases}$$

Azt is kiszámoltuk már, hogy $\mathbb{E}(X_2^*) = 1/2$, illetve a sűrűségfüggvény alapján könnyen kiszámolható, hogy $\mathbb{D}^2(X_2^*) = 0{,}05$. Általános $\vartheta > 0$ esetén X_1, X_2, X_3-mat úgy is megkaphatjuk, hogy a $(0, 1)$ intervallumon egyenletes fae. Y_1, Y_2, Y_3 val. változók mindegyikéhez ϑ-t adunk (közülük a 2. rendezett mintaelemet jelölje Y_2^*). Ezért általános ϑ esetén $\mathbb{E}(X_2^*) = \mathbb{E}(Y_2^* + \vartheta) = 1/2 + \vartheta = \mathbb{E}_\vartheta(X_1)$ és $\mathbb{D}^2(X_2^*) = \mathbb{D}^2(Y_2^* + \vartheta) = \mathbb{D}^2(Y_2^*) = 0{,}05$. Ezért tehát a $T(X_1, X_2, X_3) = X_2^*$ statisztika torzítatlan becslése a $\psi(\vartheta) = \mathbb{E}_\vartheta(X_1) = \vartheta + 1/2$ paraméterfüggvénynek. Ugyanakkor ez a statisztika nem hatásos: pl. $S(X_1, X_2, X_3) = \overline{X_3}$ is torzítatlan becslése $\psi(\vartheta)$-nak, a szórásnégyzete viszont

$$\mathbb{D}^2(\overline{X_3}) = \mathbb{D}^2\left(\frac{X_1 + X_2 + X_3}{3}\right) = \frac{1}{3}\mathbb{D}^2(X_1) = \frac{1}{12 \cdot 3} = \frac{1}{36} < 0{,}05,$$

vagyis kisebb, mint $T(X_1, X_2, X_3)$-é. Tehát $S(X_1, X_2, X_3)$ hatásosabb, mint $T(X_1, X_2, X_3)$.

A becsléselmélet két általánosabban használható módszere a *maximum likelihood becslés* és a *momentum-becslés*. Ezek idén nem képezik a tárgy anyagának részét, de az érdeklődő olvasó találhat róluk információkat A. és A.2. fejezetekben, részletesebben pedig az MSc-s Matematikai statisztika tárgyban tanulhat róluk.

12. Konfidenciaintervallumok

12.1. Alapfogalmak: konfidenciaintervallum, folytonos eloszlás kvantilisei

Az eddig ismertetett maximum likelihood becslés és momentumbecslés a *pontbecslések* kategóriájába tartoznak, mivel az ismeretlen ϑ paraméterre (vagy annak valamilyen függvényére) vonatkozó becslésünk eredménye egy, a mintától függő „pont", vagyis egy valós szám vagy vektor. Bár a pontbecslés bizonyos értelemben a véletlen mintaeredmény ismeretében a „legjobb tippünket" adja meg a paraméterre, a gyakorlatban ritkán fordul elő, hogy a valódi paraméter valóban közel van a becslésünk eredményéhez. Ezért ebben a fejezetben bevezetjük a konfidencia-intervallumokat, amelyek az *intervallumbecslésekhez* tartoznak. Itt a becslés eredménye egy olyan, a realizációtól függő intervallum, amelybe a valódi ϑ paraméter egy előre megadott $1-\varepsilon$ valószínűséggel beletartozik. Ezt a valószínűséget az intervallum hosszának növelésével 1-hez tetszőlegesen közelivé tehetjük, például gyakran választjuk 0.99-nek (itt a hibavalószínűség $\varepsilon = 0.01$) vagy 0.95-nek ($\varepsilon = 0.05$). Ezt írja le formálisan a következő definíció.

12.1.1. Definíció. Legyen $\mathbf{X} = (X_1, \dots, X_n)$ egy n elemű fae. minta, ahol a mintaelemek eloszlása egy $\vartheta \in \Theta \subseteq \mathbb{R}^d$ paramétertől függ, legyen $\psi\colon \mathbb{R}^d \to \mathbb{R}$ egy függvény, és legyen $0 < \varepsilon < 1$. Azt mondjuk, hogy a $[T_1(\mathbf{X}), T_2(\mathbf{X})]$ intervallum (pontosan) $1-\varepsilon$ **szintű konfidenciaintervallum** a $\psi(\vartheta)$ paraméterfüggvényre, ha

$$\mathbb{P}_\vartheta\big(T_1(\mathbf{X}) \le \psi(\vartheta) \le T_2(\mathbf{X})\big) = 1-\varepsilon,$$

minden $\vartheta \in \Theta$ esetén.

12.1.2. *Megjegyzés.*

1. A definícióban szereplő $T_1(\mathbf{X})$ és $T_2(\mathbf{X})$ intervallum-végpontok az $\mathbf{X}$ minta statisztikái, tehát a véletlen mintától függenek, de a realizáció ismeretében már nem véletlenek.

2. Pontosan $1-\varepsilon$ szintű konfidenciaintervallum általában csak akkor adható meg, ha $X_1, \dots, X_n$ peremeloszlása folytonos. Diszkrét esetben általában *legalább* $1-\varepsilon$ *szintű konfidenciaintervallum*okról szokás beszélni, ahol a definícióban „$= 1-\varepsilon$" helyett „$\ge 1-\varepsilon$" szerepel. Ennek ebben a kurzusban nem lesz nagy jelentősége.

Mielőtt példákat mutathatnánk a konfidenciaintervallumra, szükségünk van még a *kvantilis* fogalmára is. Ezt is csak folytonos valószínűségi változók esetén vezetjük be.

Legyen X folytonos valószínűségi változó. A 4.2.1. Definíció értelmében ez azt jelenti, hogy X-nek létezik az f_X sűrűségfüggvénye. Tudjuk, hogy ebben az esetben az $x \mapsto F_X(x)$ eloszlásfüggvény folytonos. Az egyszerűség kedvéért tegyük fel, hogy a sűrűségfüggvény pontosan egy I nyílt intervallumon nem 0 (azaz I minden pontjában pozitív, $\mathbb{R} \setminus I$ minden pontjában pedig 0). Erre példák:

1. az egyenletes eloszlás, ahol $I = (a, b)$, $-\infty < a < b < \infty$,
2. az exponenciális eloszlás, ahol $I = (0, \infty)$,
3. és a normális eloszlás, ahol $I = (-\infty, \infty) = \mathbb{R}$.

Ekkor az I intervallumon az F_X eloszlásfüggvény szigorúan monoton nő. Továbbá ha az I intervallum (a, b) vagy (a, ∞) alakú, akkor $F_X(x) = 0$ minden $x \le a$-ra teljesül, és ha az I intervallum (a, b) vagy $(-\infty, b)$ alakú, akkor $F_X(x) = 1$ minden $x \ge b$-re teljesül. Ebből az következik, hogy létezik az

$$F_X^{-1}\colon (0,1) \to I, \quad y \mapsto F_X^{-1}(y)$$

inverzfüggvény (ahol $x \in I$ és $0 < y < 1$ esetén $x = F_X^{-1}(y)$ pontosan akkor teljesül, ha $y = F_X(x)$). Ez az inverzfüggvény is folytonos és szigorúan monoton növő. Fontos azonban hangsúlyozni, hogy F_X^{-1} értékkészlete általában nem az egész $\mathbb{R}$, hanem csak I. Valóban, az előbbi példák azt mutatják, hogy a 0 értéket az F_X vagy végtelen sok helyen felveszi (pl. egyenletes és exponenciális eloszlás) vagy sehol nem veszi fel (pl. normális eloszlás), hasonlóan az 1 értékhez.

12.1.3. Definíció. Legyen X folytonos valószínűségi változó F_X eloszlásfüggvénnyel és f_X sűrűségfüggvénnyel, és tegyük fel, hogy létezik egy olyan I nyílt intervallum, hogy minden $x \in \mathbb{R}$ esetén $f_X(x) \neq 0$ pontosan akkor teljesül, ha $x \in I$.

Legyen $0 < y < 1$. Ekkor az $F_X^{-1}(y) \in I$ pontot az X **y-kvantilis**ének nevezzük. Az $\frac{1}{2}$-kvantilist ($y = 1/2$) **medián**nak nevezzük.

Ha x az X y-kvantilise, akkor teljesül, hogy

$$F_X(x) = \mathbb{P}(X < x) = y,$$

vagyis X pontosan y valószínűséggel vesz fel x-nél kisebb értékeket. A medián esetén $y = 1/2$, vagyis annak a valószínűsége, hogy $X < x$, megegyezik annak a valószínűségével, hogy $X \geq x$ (sőt a folytonosság miatt az $X > x$ valószínűséggel is). Ilyen értelemben a medián valóban az empirikus medián, vagyis a „rendezett minta középső eleme" (páros elemszámnál a két középső átlaga) egyfajta analógiája.

Például diszkrét valószínűségi változó esetén elképzelhető, hogy a $0 < y < 1$ értéket az eloszlásfüggvény átugorja. Például a szabályos kockadobás eloszlásfüggvénye az $1/12$ értéket sehol nem veszi fel, mert 0-ról egyből $1/6$-ra ugrik (az $x = 1$ pontban). Emiatt az y-kvantilis fenti definíciója csak akkor helyes, ha a valószínűségi változó eloszlásfüggvénye folytonos, de a definíció – némi technikai nehézség árán – általánosítható diszkrét valószínűségi változókra is.

12.2. Konfidenciaintervallum normális eloszlás várható értékére ismert szórás esetén

Következzék most első példánk a konfidenciaintervallumra, amely valószínűleg a legismertebb példa is a témakörben:

12.2.1. *Példa.* Legyenek $X_1, \ldots, X_n \sim \mathrm{N}(\mu; \sigma^2)$ fae. valószínűségi változók, ahol a normális eloszlás $\mu \in \mathbb{R}$ várható értéke ismeretlen, de $\sigma^2 > 0$ szórásnégyzete adott. $\varepsilon \in (0, 1)$ esetén szerkesszünk $1 - \varepsilon$ szintű konfidenciaintervallumot a normális eloszlás várható értékére úgy, hogy annak hossza a lehető legkisebb legyen!

Mivel az

$$f_\mu(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

sűrűségfüggvény az x pont μ-től való távolsága függvényében szigorúan monoton csökken (bármi is legyen a μ paraméter értéke), ezért adott ε esetén a legrövidebb $1 - \varepsilon$ szintű konfidenciaintervallumot akkor fogjuk kapni, ha azt az $\overline{X_n}$ mintaátlag körül szimmetrikusan választjuk, vagyis

$$[\overline{X_n} - r_\varepsilon, \overline{X_n} + r_\varepsilon]$$

alakban, úgy, hogy

$$\mathbb{P}_\mu(\mu \in [\overline{X_n} - r_\varepsilon, \overline{X_n} + r_\varepsilon]) = 1 - \varepsilon \tag{12.1}$$

minden $\mu \in \mathbb{R}$ esetén teljesüljön. A feladatunk tehát az r_ε (ε-tól függő) értékének meghatározása.

A mintaátlag eloszlása a normális eloszlás ismert tulajdonságai alapján:

$$\overline{X_n} = \frac{1}{n} \underbrace{\sum_{i=1}^{n} \underbrace{X_i}_{\sim \mathrm{N}(\mu;\sigma^2) \text{ fae.}}}_{\sim \mathrm{N}(n\mu; n\sigma^2)} \sim \mathrm{N}(\mu; \sigma^2/n).$$

Ebből következik, hogy

$$\frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\overline{X_n} - \mu)}{\sigma} \sim \mathrm{N}(0; 1). \tag{12.2}$$

Fejezzük ki most a (12.1) egyenletet az $\frac{\overline{X_n}-\mu}{\sigma}\sqrt{n}$ valószínűségi változó segítségével:

$$\mathbb{P}_\mu(\mu \in [\overline{X_n} - r_\varepsilon, \overline{X_n} + r_\varepsilon]) = 1-\varepsilon \qquad \Leftrightarrow \qquad \mathbb{P}_\mu(-r_\varepsilon \le \overline{X_n} - \mu \le r_\varepsilon) = 1-\varepsilon$$

$$\Leftrightarrow \qquad \mathbb{P}_\mu\Big(-\frac{r_\varepsilon}{\sigma}\sqrt{n} \le \frac{\overline{X_n}-\mu}{\sigma}\sqrt{n} \le \frac{r_\varepsilon}{\sigma}\sqrt{n}\Big) = 1-\varepsilon.$$

Az utóbbi pedig (szokás szerint Φ-vel jelölve a standard normális eloszlás eloszlásfüggvényét) pontosan akkor teljesül, ha

$$\Phi\big(\frac{r_\varepsilon}{\sigma}\sqrt{n}\big) - \Phi\big(-\frac{r_\varepsilon}{\sigma}\sqrt{n}\big) = 2\Phi\big(\frac{r_\varepsilon}{\sigma}\sqrt{n}\big) - 1 = 1-\varepsilon,$$

vagyis átrendezve

$$\Phi\big(\frac{r_\varepsilon}{\sigma}\sqrt{n}\big) = 1 - \frac{\varepsilon}{2}. \tag{12.3}$$

A $\Phi\colon \mathbb{R} \to (0,1)$ eloszlásfüggvény egész $\mathbb{R}$-en szigorúan monoton nő, tehát létezik a $\Phi^{-1}\colon (0,1) \to \mathbb{R}$ inverzfüggvény, amely szintén szigorúan monoton nő. Így tehát (12.3) mindkét oldalára Φ^{-1}-t alkalmazva azt kapjuk, hogy

$$\frac{r_\varepsilon}{\sigma}\sqrt{n} = \Phi^{-1}(1-\varepsilon/2),$$

vagyis

$$r_\varepsilon = \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\varepsilon/2).$$

A végeredmény megadásához vezessük be $\delta \in (0,1)$ esetén az

$$u_\delta = \Phi^{-1}(1-\delta) \tag{12.4}$$

jelölést! Vegyük észre, hogy ez éppen a standard normális eloszlás $1-\delta$-kvantilise, hiszen ha $X \sim \mathrm{N}(0;1)$, akkor

$$\mathbb{P}(X < u_\delta) = \Phi(u_\delta) = \Phi(\Phi^{-1}(1-\delta)) = 1-\delta.$$

Így tehát

$$r_\varepsilon = \frac{\sigma}{\sqrt{n}}u_{\varepsilon/2},$$

és a keresett konfidenciaintervallum

$$\big[\overline{X_n} - \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}, \overline{X_n} + \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}\big].$$

Eredményeinket a következő állításban foglaljuk össze.

12.2.2. Állítás. *Legyenek $X_1, \dots, X_n$ fae. $N(\mu;\sigma^2)$ eloszlású valószínűségi változók, ahol a μ várható érték ismeretlen és a $\sigma^2 > 0$ szórásnégyzet ismert, és jelölje $\overline{X_n}$ a mintához tartozó mintaátlagot. Ekkor $\varepsilon \in (0,1)$ esetén a μ várható értékre egy $1-\varepsilon$ szintű konfidenciaintervallum a következő:*

$$\big[T_1(\mathbf{X}), T_2(\mathbf{X})\big] = \big[\overline{X_n} - \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}, \overline{X_n} + \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}\big],$$

ahol $u_{\varepsilon/2}$ a standard normális eloszlás $1-\varepsilon/2$-kvantilise.

12.2.3. *Megjegyzés.* Figyeljük meg, hogy

1. minél nagyobb a minta elemszáma, annál kevésbé hosszú konfidenciaintervallum szükséges (vö. nagy számok törvénye),
2. minél nagyobb a szórás, annál hosszabb konfidenciaintervallum szükséges (nagy szórás esetén „laposabb”, jobban szétterülő a sűrűségfüggvény),

3. minél kisebb az ε hibaszint, annál hosszabb konfidenciaintervallum szükséges (mert $1-\varepsilon$ valószínűséggel kell μ-nek a konfidenciaintervallumba esnie, ha μ a valódi paraméter).

Az olvasóban joggal merülhet fel a kérdés: hogyan kell a normális eloszlás szórásnégyzetére konfidenciaintervallumot szerkeszteni, ha a várható érték ismert? Illetve hogyan lehet a normális eloszlás várható értékére akkor is konfidenciaintervallumot megadni, ha a szórás ismeretlen? Ezek idén nem képezik a tárgy anyagának részét, de az érdeklődő olvasó a függelék B fejezetében olvashat és az MSc-s Matematikai statisztika tárgyban is tanulhat róluk.

13. Hipotézisvizsgálati módszerek

13.1. Bevezetés

A becsléselmélet mellett matematikai statisztika másik nagy ága a *hipotézisvizsgálat*, ahol valamilyen előzetes hipotézisünket (azaz feltevésünket) akarjuk valamilyen mérés, illetve adatelemzés alapján igazolni vagy cáfolni. Ilyen hipotézis lehet például:[10]

- A mintaelemek várható értéke 2.
- A mintaelemek várható értéke legfeljebb 2.
- A mintaelemek várható értéke megegyezik egy adott másik minta elemeinek várható értékével.
- A mintaelemek normális eloszlásúak.
- A realizációk függetlenek valamely más, kapcsolódó mérés eredményeitől.

Célunk ebben a fejezetben az első három típusba tartozó hipotézisek vizsgálata. A továbbiak iránt érdeklődőknek megint csak a mérnök- és gazdaságinformatikus MSc képzés Matematikai statisztika tárgyát ajánljuk.

A mintavétel után számítások segítségével megvizsgáljuk, hogy a mérési eredmények a hipotézisünknek ellentmondanak-e, azaz hogy a hipotézisünket feltéve a mérési eredmények nagyon valószínűtlenek-e. Ha igen, akkor a hipotézist elutasítjuk. Általános esetben a hipotézisvizsgálat (röviden *próba*) az alábbi lépéssorozat végrehajtását jelenti:

1. A H_0-lal jelölt *nullhipotézis* formalizálása.
2. A mintára vonatkozó *elfogadási tartomány* és *kritikus tartomány* megalkotása.
3. A kísérlet végrehajtása, amelyben az $x_1, \dots, x_n$ realizációt (adatpontokat) kapjuk.
4. Annak megvizsgálása, hogy az adatpontok az elfogadási tartományba esnek-e. Ha igen, H_0-t elfogadjuk. Ellenkező esetben H_0-t elutasítjuk.

Az alábbiakban megadjuk az imént említett fogalmaknak – és néhány továbbinak – a definícióját.

13.1.1. Definíció. Legyen $\mathbf{X} = (X_1, \dots, X_n)$ egy fae. minta az $(\Omega, \mathcal{F}, \mathbb{P})$ valószínűségi mezőn, ahol a mintaelemek peremeloszlása ismeretlen. Legyen $\mathbf{x} = (x_1, \dots, x_n)$ az $\mathbf{X} = (X_1, \dots, X_n)$ egy realizációja.

(1) Egy H_0 **nullhipotézis** egy olyan kijelentés, amelynek igazsága csak az $X_1, \dots, X_n$ együttes eloszlásától függ. A H_1 **ellenhipotézis** a H_0 tagadása: $H_1 = \neg H_0$. Az adott próbához tartozó **alternatíva**[11] a

$$H_0 \text{ vs. } H_1.$$

[10] Példa gyanánt az itt felsorolt hipotéziseken kívül természetesen gondolhatunk a 10. fejezet elején ismertetett példahipotézisekre is.

[11] vs.: ejtsd „versus” (latinul „ellen”).

(2) Egy $\mathfrak{X}_\mathrm{e}$ **elfogadási tartomány** az $\mathbb{R}^n$ egy részhalmaza. Az $\mathfrak{X}_\mathrm{e}$ elfogadási tartományhoz tartozó $\mathfrak{X}_\mathrm{k}$ **kritikus tartomány** az $\mathfrak{X}_\mathrm{e}$ komplementere, azaz $\overline{\mathfrak{X}_\mathrm{e}} = \mathbb{R}^n \setminus \mathfrak{X}_\mathrm{e}$. A $\mathfrak{X}_\mathrm{k}$ kritikus tartománnyal definiált **(statisztikai) próba** az az eljárás, amely során H_0-t elfogadjuk, ha $\mathbf{X} \in \mathfrak{X}_\mathrm{e}$ és H_0-t elutasítjuk (azaz H_1-et fogadjuk el), ha $\mathbf{X} \in \mathfrak{X}_\mathrm{k}$.

(3) A $\mathfrak{X}_\mathrm{k}$ kritikus tartománnyal definiált próba **terjedelme** a $0 < \alpha < 1$ szám, ha

$$\mathbb{P}\big((X_1, \dots, X_n) \in \mathfrak{X}_\mathrm{k} \big| H_0 \text{ igaz}\big) = \alpha.$$

(4) **Elsőfajú hibá**nak nevezzük azt, amikor egy próba eljárása során H_0-t elutasítjuk, pedig H_0 igaz. **Másodfajú hibá**nak nevezzük azt, amikor az eljárás során H_0-t elfogadjuk, pedig H_1 igaz (azaz H_0 hamis).

13.1.2. *Megjegyzés.* Egy α terjedelmű próba elsőfajú hibájának valószínűsége éppen a terjedelem:

$$\mathbb{P}(H_0\text{-t elutasítjuk} | H_0 \text{ igaz}) = \mathbb{P}\big((X_1, \dots, X_n) \in \mathfrak{X}_\mathrm{k} \big| H_0 \text{ igaz}\big) = \alpha.$$

A másodfajú hiba valószínűségének kiszámítására viszont nem mindig áll rendelkezésre explicit módszer, amint azt konkrét példákban látni fogjuk.

Konkrét próbák működési elvének megértésében segíthetnek a következő általános, matematikailag nem precíz meggondolások és példák[12].

13.1.3. *Megjegyzés.* A H_0 nullhipotézist gyakran úgy fogalmazzuk meg, hogy annak teljesülése esetén ismert legyen az $X_1, \dots, X_n$ mintaelemek eloszlása. Ilyen értelemben például az utcán talált pénzérme esetén jó nullhipotézis a H_0: „a pénzérme szabályos" kijelentés, mert ekkor annak ismeretében, hogy H_0 igaz, hogy a mintaelemek fae. 1/2 paraméterű indikátorváltozók. Ezzel szemben nem jó nullhipotézis például a H_0': „a pénzérme nem szabályos" kijelentés, mert ekkor H_0' ismeretében csak azt tudjuk, hogy a mintaelemek valamilyen, 1/2-től különböző paraméterrel rendelkező fae. indikátorváltozók, ami nem ad elegendő információt a terjedelem kiszámításához.

13.1.4. *Megjegyzés.* Sokszor helyes az az interpretáció, hogy a nullhipotézist „az ártatlanság vélelmében" fogalmazzuk meg. Például a pénzérme vizsgálatára tipikusan azért kerül sor, mert arra gyanakodunk, hogy a pénzérme cinkelt, de mégis ezen állítás tagadását tesszük meg H_0-nak. Ebben az értelemben

- az elsőfajú hiba úgy interpretálható, mint az ártatlan elítélése,
- a másodfajú hiba úgy interpretálható, mint a bűnös felmentése.

(Még egyszer hangsúlyozzuk, hogy ezek matematikailag nem precíz, sőt nem is igazán formalizálható állítások.) A legtöbb ismert próbában csak az elsőfajú hiba valószínűségét, vagyis a terjedelmet tudjuk kontrollálni: előre megadunk egy α értéket, és a próba konstrukciója alapján a terjedelem éppen α lesz. Viszont érezhető, – anélkül is, hogy bármit tudnánk konkrét próbákról – hogy ha az előirányzott terjedelmet nagyon kicsinek választjuk, akkor a másodfajú hiba valószínűsége megnő. („Ha biztosra akarunk menni, hogy ártatlanokat ne ítéljünk el, akkor bűnösöket is gyakrabban fogunk felmenteni.") A gyakorlatban $\varepsilon = 0.05$ és $\varepsilon = 0.01$ a leggyakrabban választott terjedelemértékek.

13.1.5. *Példa.* Ha például egy új gyógyszert szeretnénk forgalomba hozni, akkor H_0: „a gyógyszer hatástalan vagy káros" vs. H_1: „a gyógyszer hatásos" jó alternatíva, mert ekkor az elsőfajú hiba azt a szcenáriót jelenti, hogy a mintavétel alapján egy hatástalan vagy káros gyógyszert hatásosnak mondunk (és ez alapján vélhetően forgalomba hozzuk). Mivel ez az elsőfajú hiba, ennek valószínűségét tudjuk kontrollálni, tetszőlegesen alacsony $1 - \alpha$ szint alatt tarthatjuk, a másodfajú hiba valószínűségének megnövekedésének árán. A másodfajú hiba itt az az eshetőség, hogy a gyógyszer hatásos, de a mintavétel alapján mégis hatástalannak vagy károsnak nyilvánítjuk (és ezért bizonyára nem vezetjük be). Természetesen a másodfajú hiba bekövetkezése sem kedvező (pl. gazdasági szempontból), de nem jár olyan fatális következményekkel, mint az elsőfajú hiba. Ezért ha olyan hipotézisvizsgálati módszert használunk a gyógyszer tesztelésére, amelyben csak az elsőfajú hiba valószínűsége kontrollálható, akkor H_0-t a fenti módon kell megválasztani.

[12] Ezek Bolla Marianna 2012-es *Matematikai statisztika* előadásáról származnak.

13.2. u-próba

A fejezet hátralévő részében ismertetjük az u-próba több alesetét. Ezek a normális eloszlás várható értékével kapcsolatos próbák ismert szórás esetén. Látni fogjuk, hogy ezek a próbák számos ponton kapcsolódnak a konfidenciaintervallumok témájához (ami már az u jelölés alapján is sejthető). A fejezet ezen részével kapcsolatban érdemes mindjárt az elején tisztázni, hogy az u-próbák csak normális eloszlású minták esetén működnek. A centrális határeloszlás-tétel miatt sokféle mintáról feltehető, hogy (jó közelítéssel) normális eloszlású, de azért számos olyan alkalmazás is van, amikor ez nem teljesül. Például a mérnök- és gazdaságinformatikus MSc képzés Matematikai statisztika tárgyának anyagában szerepelnek ún. *nemparaméteres próbák* is, amelyek a normális eloszlást nem tételezik fel és nagyon sokfajta eloszlásra alkalmazhatóak. A nemparaméteres próbák legismertebb példái közé tartozik a χ^2*-próba* és a *Kolmogorov–Szmirnov-próba.*

Az u-próba kétoldali, egymintás változatát egy példán vezetjük be, majd áttekinthető formában összefoglaljuk az eljárást.

13.2.1. *Példa* (Kétoldali, egymintás u-próba: kenyérvásárlás). A sarki péknél mindennap egy kilós kenyeret veszünk, és a kenyerek az utóbbi napokban kisebbnek tűnnek a korábban megszokottnál, ezért felmerül a gyanúnk, hogy újabban várható értékben is 1 kg-nál kisebb tömegű kenyereket ad el nekünk a pék. Feltesszük, hogy a vásárolt kenyerek kilogrammban mért tömegei egymástól függetlenek és normális eloszlásúak ismeretlen μ várható értékkel és ismert $\sigma = 0.02$ kilogramm szórással (tehát $\sigma^2 = 0.0004$ kg^2). Ezért felállítjuk a

$$H_0\colon \mu = \mu_0 \qquad \text{vs.} \qquad H_1\colon \mu \neq \mu_0$$

alternatívát, ahol esetünkben $\mu_0 = 1$. Terjedelemként $\varepsilon = 0.05$-öt irányzunk elő, vagyis azt szeretnénk, hogy a próba során legfeljebb 0.05 valószínűséggel utasítsuk el H_0-t, amennyiben H_0 igaz.

Ezután mintát veszünk: $n = 25$ napon keresztül az aznap vásárolt kenyeret mindig mérlegre tesszük. Egy olyan $\mathbf{x} = (x_1, \dots, x_{25})$ realizációt kapunk, amelyre az átlagos tömeg $\overline{x_n} = \overline{x_{25}} = 0.98$ kg. A minta alapján szeretnénk egy olyan próbát konstruálni, amelynek terjedelme pontosan $\varepsilon = 0.05$, azaz az $\mathfrak{X}_\mathrm{e}$ elfogadási tartományt és az $\mathfrak{X}_\mathrm{k}$ kritikus tartományt úgy akarjuk megválasztani, hogy

$$\mathbb{P}(H_0\text{-t elutasítjuk}|H_0) = \varepsilon = 0.05$$

teljesüljön.

Az elfogadási tartományt keressük

$$\mathfrak{X}_\mathrm{e} = \{\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n\colon \mu_0 \in [\overline{x_n} - h, \overline{x_n} + h]\}$$

alakban, ahol $h > 0$ és $\overline{x_n} = \frac{1}{n}\sum_{i=1}^n x_i$ a bal oldalon álló $(x_1, \dots, x_n)$ realizáció átlaga. Ekkor h-t úgy kívánjuk megválasztani, hogy $\mu = \mu_0 = 1$ esetén (az ennek megfelelő valószínűséget $\mathbb{P}_{\mu_0}$-lal jelöljük) éppen ε valószínűséggel essen $\mu = \mu_0$ a kritikus tartományba, azaz $1 - \varepsilon$ valószínűséggel essen az elfogadási tartományba, vagyis

$$\mathbb{P}_{\mu_0}\big(\mu_0 \in [\overline{X_n} - h, \overline{X_n} + h]\big) = 1 - \varepsilon$$

teljesüljön. Azaz $[\overline{X_n} - h, \overline{X_n} + h]$ egy $\overline{X_n}$ körül szimmetrikus, $1 - \varepsilon$ szintű konfidenciaintervallum kell, hogy legyen a μ_0 paraméterre. Azt viszont már a 12.2. alfejezetből tudjuk, hogy egy ilyen konfidenciaintervallum

$$\big[\overline{X_n} - \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}, \overline{X_n} + \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}\big],$$

amiből tehát azt kapjuk, hogy

$$h = \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}.$$

Vagyis H_0-t pontosan akkor fogadjuk el, ha

$$\overline{X_n} - \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}} \leq \mu_0 \leq \overline{X_n} + \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}} \qquad \Leftrightarrow \qquad \mu_0 - \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}} \leq \overline{X_n} \leq \mu_0 + \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}$$

$$\Leftrightarrow \qquad -u_{\varepsilon/2} \leq \frac{\overline{X_n} - \mu_0}{\sigma}\sqrt{n} \leq u_{\varepsilon/2}.$$

Az elfogadási tartományunk emiatt

$$\mathfrak{X}_{\mathrm{e}} = \big\{\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n \colon -u_{\varepsilon/2} \le \frac{\overline{x_n} - \mu_0}{\sigma}\sqrt{n} \le u_{\varepsilon/2}\big\}.$$

Ha tehát az $u(\mathbf{X}) = \frac{\overline{X_n}-\mu_0}{\sigma}\sqrt{n}$ *próbastatisztika* $u(\mathbf{x}) = \frac{\overline{x_n}-\mu_0}{\sigma}\sqrt{n}$ realizációja $-u_{\varepsilon/2}$ és $u_{\varepsilon/2}$ közé esik, akkor H_0-t elfogadjuk, különben pedig elutasítjuk (azaz H_1-et fogadjuk el).

A konkrét esetünkben $\varepsilon = 0.05$, $n = 25$, $\overline{x_n} = 0.98$, $\mu_0 = 1$ és $\sigma = 0.03$. Ez alapján tehát $u(\mathbf{x}) = \frac{0.98-1}{0.02}\sqrt{25} = -5$. Ugyanakkor a normális eloszlás táblázatába pillantva látjuk, hogy $u_{\varepsilon/2} = \Phi^{-1}(0.975) \approx 1.9600$. Tehát H_0-t akkor fogadnánk el, ha az $u(\mathbf{X})$ próbastatisztika értéke -1.96 és 1.96 közé esne. Mivel ez nem áll fent, H_0-t elutasítjuk, vagyis kellő bizonyítékot találtunk arra, hogy a kenyerek várható értékben nem egykilósak.

Az alábbiakban formalizáljuk az egymintás u-próba általános eljárását, amelynek egyes részletei a próba altípusától (egy- vagy kétoldali) függnek:

Eljárás (Egymintás u-próba)**.** *Feltevések: az $x_1, \dots, x_n$ adatpontok fae. normális eloszlású $X_1, \dots, X_n$ valószínűségi változók realizációi, amelyek μ várható értéke ismeretlen, de $\sigma > 0$ szórásuk ismert. Általános eljárás:*

1. *A H_0 nullhipotézis felállítása.*
2. *Az előirányzott $\varepsilon \in (0, 1)$ terjedelem kiválasztása.*
3. *Az $x_1, \dots, x_n$ adatpontok kinyerése.*
4. *Az elfogadási tartomány meghatározása. Ez függ a próba altípusától, de mindig magába foglalja a normális eloszlás egy megfelelő kvantilisének kiszámítását (táblázatból való kiolvasását).*
5. *A próbastatisztika $u(\mathbf{x}) = \frac{\overline{x_n}-\mu_0}{\sigma}\sqrt{n}$ értékének meghatározása.*
6. *Annak eldöntése, hogy a próbastatisztika értéke az elfogadási tartományba esik-e. Ha igen, H_0-t elfogadjuk. Ha nem, H_0-t elutasítjuk (tehát H_1-et fogadjuk el).*

Példánk alapján a kétoldali, egymintás u-próba esetén az eljárást így pontosíthatjuk:

Kétoldali, egymintás u-próba.

- *Nullhipotézis:* $H_0 \colon \mu = \mu_0$.
- *Releváns kvantilis:* $u_{\varepsilon/2} = \Phi^{-1}(1 - \varepsilon/2)$.
- *Elfogadási tartomány:* $\mathfrak{X}_{\mathrm{e}} = \{(x_1, \dots, x_n) \colon |u(\mathbf{x})| \le u_{\varepsilon/2}\}$, *tehát:*
  - *ha* $|u(\mathbf{x})| > u_{\varepsilon/2}$, H_0*-t elutasítjuk,*
  - *különben H_0-t elfogadjuk.mi*

Most előző példánknál maradva bevezetjük az egyoldali, egymintás u-próbát is.

13.2.2. *Példa* (Egyoldali, egymintás u-próba: kenyérvásárlás). Az előző példánál felmerülhet a kérdés, hogy miért azt tettük meg nullhipotézisnek, hogy a kenyér várható értékben egykilós. Ugyanis a 0.98 kg-os mintaátlagot józan ésszel senki nem gondolná bizonyítéknak arra, hogy a kenyér tömege várható értékben *nagyobb*, mint 1 kg, hanem csak arra, hogy kisebb. Ennek megfelelően felállíthatunk egy másik alternatívát (megint csak „az ártatlanság vélelmében”):

$$H_0 \colon \mu \ge \mu_0 \qquad \text{vs.} \qquad H_1 \colon \mu < \mu_0, \tag{13.1}$$

ahol μ_0 változatlanul 1-gyel egyenlő.

Vajon teljesíti-e ez az alternatíva azt a feltételt, hogy ha tudjuk, hogy teljesül, akkor a mintaelemek eloszlása is ismert? Szigorú értelemben nem: H_0 teljesülése esetén lehetnek a mintaelemek például $N(\mu_0;\sigma^2)$ és $N(\mu_0+113;\sigma^2)$ eloszlásúak is. Azonban a H_0 teljesülése esetén egy μ_0-nál kisebb átlagú mintát akkor kapunk a legnagyobb valószínűséggel, ha μ megegyezik μ_0-lal (nem pedig nagyobb nála). Ezért ha egy tetszőleges A eseményre – a jelöléssel enyhén visszaélve – bevezetjük a

$$\mathbb{P}(A|H_0 \text{ igaz}) := \mathbb{P}(A|\text{a mintaelemek várható értéke } \mu_0)$$

jelölést, vagyis úgy fogjuk fel, hogy H_0 teljesülése esetén a várható érték egyenlő μ_0-lal, akkor a kétoldali, egymintás u-próbához hasonlóan járhatunk el.

Az $u(\mathbf{x})$ próbastatisztika-realizácónk most is ugyanaz lesz, mint a kétoldali esetben, a különbség csak az lesz, hogy most csak akkor utasítjuk el H_0-t, ha az $u=\frac{\overline{x_n}-\mu_0}{\sigma}\sqrt{n}$ értéke nagyon kicsi. Pontosabban olyan elfogadási tartományt választunk, amely jobbról végtelen, és $\mu=\mu_0$ esetén $1-\varepsilon$ valószínűséggel esik μ_0 az $\mathfrak{X}_\mathrm{e}$ elfogadási tartományba, tehát ε valószínűséggel a tőle balra eső $\mathfrak{X}_\mathrm{k}$ kritikus tartományba:

$$\mathfrak{X}_\mathrm{k}=\{\mathbf{x}=(x_1,\dots,x_n)\in\mathbb{R}^n\colon u(\mathbf{x})<-u_\varepsilon\},$$

ahol $u_\varepsilon=\Phi^{-1}(1-\varepsilon)$ továbbra is a standard normális eloszlás $1-\varepsilon$-kvantilisét jelöli.

Példánkban $u(\mathbf{x})=-5$-öt kaptunk. A standard normális eloszlás $1-\varepsilon=0.95$-kvantilise $\Phi^{-1}(1-\varepsilon)=\Phi^{-1}(0.95)\approx 1.6449$. Mivel $u(\mathbf{x})=-5<-u_\varepsilon$, H_0-t ismét elutasítjuk, vagyis azt kapjuk, hogy a kenyerek tömege várható értékben kisebb, mint 1 kg.

Az egyoldali, egymintás u-próba tehát μ_0-nál nagyobb átlagú minta esetén soha nem utasítja el H_0-t, viszont μ_0-nál kisebb átlagú minták esetén szigorúbb, mint a kétoldali, egymintás próba: a példánkban ha $u(\mathbf{x})$ a -1.9600 és a -1.6449 számok közé esik, akkor az egyoldali próba már elutasítja H_0-t, de a kétoldali még elfogadja. Ez az egyoldali próba fő előnye.

Ha viszont a

$$H_0\colon \mu\le\mu_0 \qquad \text{vs.} \qquad H_1\colon \mu>\mu_0 \tag{13.2}$$

alternatívára kívánunk egyoldali, egymintás u-próbát szerkeszteni, akkor a szimmetria miatt az

$$\mathfrak{X}_\mathrm{k}=\{\mathbf{x}=(x_1,\dots,x_n)\in\mathbb{R}^n\colon u(\mathbf{x})>u_\varepsilon\}$$

kritikus tartományt kapjuk. Most tehát csak akkor utasítjuk el H_0-t, ha a próbastatisztika értéke nagyon nagy. Az eddigiek alapján összefoglaljuk az egyoldali, egymintás u-próba jellemzőit mindkét lehetséges nullhipotézis esetén.

Egyoldali, egymintás u-próba ($H_0\colon \mu\ge\mu_0$ változat)**.**

- *Nullhipotézis:* $H_0\colon \mu\ge\mu_0$.
- *Releváns kvantilis:* $-u_\varepsilon=-\Phi^{-1}(1-\varepsilon)$.
- *Elfogadási tartomány:* $\mathfrak{X}_\mathrm{e}=\{(x_1,\dots,x_n)\colon u(\mathbf{x})\ge -u_\varepsilon\}$*, tehát:*
  - *ha* $u(\mathbf{x})<-u_\varepsilon$*,* H_0*-t elutasítjuk,*
  - *különben* H_0*-t elfogadjuk.*

Egyoldali, egymintás u-próba ($H_0\colon \mu\le\mu_0$ változat)**.**

- *Nullhipotézis:* $H_0\colon \mu\le\mu_0$.
- *Releváns kvantilis:* $u_\varepsilon=\Phi^{-1}(1-\varepsilon)$.
- *Elfogadási tartomány:* $\mathfrak{X}_\mathrm{e}=\{(x_1,\dots,x_n)\colon u(\mathbf{x})\le u_\varepsilon\}$*, tehát:*

– *ha* $u(\mathbf{x}) > u_\varepsilon$*,* H_0*-t elutasítjuk,*
– *különben* H_0*-t elfogadjuk.*

13.2.3. *Megjegyzés* (Próba ereje, az u-próba optimalitása). Igazoltuk (legalább vázlatosan), hogy a kétoldali, egymintás u-próba terjedelme, vagyis elsőfajú hibájának valószínűsége legfeljebb az előzetesen választott ε paraméter lesz. Mit mondhatunk a másodfajú hibáról? Ha H_1 teljesül, akkor nem tudjuk, milyen eloszlásúak a mintaelemeink, mert normális eloszlásuk várható értéke ismeretlen. Vegyünk most egy $\mu_1 \neq \mu_0$ paramétert, ahol μ_0 a nullhipotézis szerinti várható érték. Tehát ha μ_1 a valódi paraméter, akkor H_1 igaz. Ha tudnánk, hogy μ_1 a valódi paraméter, akkor a másodfajú hibát a

$$\mathbb{P}(H_0\text{-t elfogadjuk}|\ \mu_1 \text{ a valódi paraméter}) = \mathbb{P}\big(u(\mathbf{X}) \in [-u_{\varepsilon/2}, u_{\varepsilon/2}]\big|\ X_1 \sim \mathrm{N}(\mu_1; \sigma^2)\big)$$

képlettel írhatnánk le, ami egy kiszámítható, konkrét érték. Ebben az esetben a próba **erejé**nek 1 mínusz a másodfajú hiba valószínűségét nevezzük. Az erő tehát ebben az esetben csak konkrét H_1-beli paraméter esetén definiálható. Látható, hogy minél távolabb van μ_1 a μ_0 értéktől, annál kisebb lesz a másodfajú hiba valószínűsége, vagyis annál nagyobb lesz a próba ereje.

Miért a (kétoldali, egymintás) u-próbát alkalmazzuk a gyakorlatban és nem valamilyen más, szintén ε terjedelmű próbát ugyanerre az alternatívára (feltételezve, hogy a mintaelemek eloszlása normális és a σ szórásuk adott)? Azért, mert bizonyítható, hogy az u-próba az **egyenletesen legerősebb** próba valamennyi ilyen próba közül. Ez azt jelenti, hogy bárhogy is választunk egy H_1-beli μ_1 paramétert, $\mu = \mu_1$ esetén a másodfajú hiba akkor lesz a legkisebb, ha az u-próbát alkalmazzuk. Ezt az állítást itt nem bizonyítjuk, sőt nem is mondjuk ki formálisan, csak megjegyezzük, hogy ez a **Neyman–Pearson-lemma** következménye. A Neyman–Pearson-lemma az egyenletesen legerősebb próba létezését mondja ki és konstrukcióját adja meg általános esetben, és többek között matematikus hallgatóknak szóló matematikai statisztikai tárgyaknál találkozhatunk vele (pl. a BME-TTK-n is).

Ilyen értelemben egyenletesen legerősebb próba az egyoldali, egymintás és a kétoldali, kétmintás u-próba is, mindegyik a maga alternatívájára vonatkozóan, a saját feltevései mellett.

Lássunk most egy olyan példát, ahol eredendően nem normális eloszlással van dolgunk, de a CHT miatt az u-próba mégis alkalmazható!

13.2.4. *Példa.* Az utcán talált, ismeretlen $\vartheta \in (0,1)$ valószínűséggel fejet mutató pénzérmét n =100-szor feldobjuk, a dobás eredménye 60-szor lesz fej és 40-szer írás. Döntsük el az egyoldali, egymintás u-próba segítségével, hogy a pénzérme szabályosnak tekinthető-e, ha a próba terjedelme $\varepsilon = 0.05$, illetve $\varepsilon = 0.01$!

Az $X_i = \mathbb{1}_{\{\text{az } i.\text{ dobás eredménye fej}\}}$ fae. mintaelemeink ebben az esetben nem normális eloszlásúak. Ezek ϑ paraméterű indikátorváltozók, tehát $\mathbb{E}_\vartheta[X_i] = \vartheta$ valószínűséggel mutatnak fejet, a szórásnégyzetük pedig szintén ismeretlen: $\mathbb{D}^2_\vartheta[X_i] = \vartheta(1-\vartheta)$.

A minta mérete $n \geq 30$ esetén elég nagy ahhoz, hogy a centrális határeloszlás-tételt alkalmazzuk. Tekintsük a

$$H_0\colon \vartheta \leq \frac{1}{2} \qquad \text{vs.} \qquad H_1\colon \vartheta > \frac{1}{2}$$

alternatívát. Ismét igaz az, hogy egy $n\vartheta = 50$-nél nagyobb átlagú mintát H_0 teljesülése esetén akkor kapunk a legnagyobb valószínűséggel, ha $\vartheta = \frac{1}{2}$. Ezért (a jelöléssel ismét némiképpen visszaélve) feltehetjük, hogy H_0 teljesülése esetén $\vartheta = \frac{1}{2}$. Ebben az esetben az X_i mintaelemek várható értéke $\frac{1}{2}$, szórásnégyzete $\frac{1}{4}$. Így a centrális határeloszlás-tétel miatt

$$u(X_1, \ldots, X_n) = \frac{\sum_{i=1}^n X_i - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}}$$

valószínűségi változó közelítőleg $\mathrm{N}(0;1)$ eloszlású. Az egyoldali, egymintás u-próba szellemében ha ezen statisztika értéke a $(-\infty, u_\varepsilon]$ intervallumba esik, akkor H_0-t elfogadjuk, különben pedig H_0-t elutasítjuk.

Konkrét példánkban

$$u(x_1, \ldots, x_n) = \frac{60 - 50}{10/2} = 2.$$

Ezt kell $\varepsilon = 0.05$ esetén $u_\varepsilon = \Phi^{-1}(1-\varepsilon) = \Phi^{-1}(0.95) = 1.6449$-cel, $\varepsilon = 0.01$ esetén pedig $u_\varepsilon = \Phi^{-1}(0.99) = 2.3263$-mal összehasonlítani. Ezért $\varepsilon = 0.05$ esetén H_0-t elutasítjuk (tehát arra következtetünk, hogy a pénzérme nem szabályos), viszont $\varepsilon = 0.01$ esetén elfogadjuk (vagyis arra a konklúzióra jutunk, hogy az érme szabályos).

13.2.5. *Megjegyzés* (p-érték). Láthatjuk, hogy ha ε-t csökkentjük, akkor H_0-t gyakrabban fogjuk elfogadni; ez nem meglepő, hiszen ε a próba terjedelmével, vagyis az elsőfajú hiba valószínűségével egyezik meg. A fenti pénzérmés példában $\varepsilon = 0.01$ és $\varepsilon = 0.05$ között található az a legnagyobb terjedelem, amelyre H_0-t az adott mintának megfelelő $u(\mathbf{x})$ próbastatiszika-érték esetén még elfogadjuk. Ezt szokás a próba ***p*-értéké**nek (az angol nyelvű irodalomban *p-value*) nevezni. Minél kisebb a p-érték, annál jogosabb a H_0 nullhipotézis elutasítása.

A pénzérmés példában (egyoldali, egymintás u-próba) a p-érték éppen 1 mínusz a standard normális eloszlásfüggvénynek a próbastatisztikában vett értéke:

$$u(\mathbf{x}) = u_\varepsilon \;\Leftrightarrow\; \mathbb{P}(u(\mathbf{X}) > u(\mathbf{x})) = \varepsilon \;\Leftrightarrow\; \varepsilon = 1 - \Phi(u(\mathbf{x})) = 1 - \Phi(2) \approx 1 - 0.9772 = 0.0228.$$

Vagyis a próba p-értéke kb. 2,28%. Ennél kisebb (vagy ezzel egyenlő) terjedelem esetén H_0-t elfogadjuk, ennél nagyobb terjedelem esetén elutasítjuk.

Kétoldali u-próba esetén, például a kenyeres példában:

$$|u(\mathbf{x})| = u_{\varepsilon/2} \;\Leftrightarrow\; \varepsilon = 2(1 - \Phi(|u(\mathbf{x})|)) = 2(1 - \Phi(5)) \approx 5.733 \times 10^{-7}.$$

Vagyis a próba p-értéke rendkívül alacsony; H_0-t csak akkor fogadjuk el, ha $\varepsilon \leq 5.733 \times 10^{-7}$.

A p-érték használata egy alternatív szemléletet jelent az előirányzott terjedelemhez képest: itt nem a mintavétel előtt határozzuk meg a próba ε terjedelmét, ez és a minta alapján döntést hozva H_0 elfogadásáról vagy elutasításáról, hanem a minta alapján határozzuk meg azt az ε terjedelmet, amely pont a H_0 elfogadása és elutasítása határán áll, azaz a p-értéket.

Gyakori szóhasználat, hogy valamilyen mintára vonatkozó állítás „*szignifikáns*”, ha a p-érték legfeljebb 0.05, és „*nagyon szignifikáns*”, ha legfeljebb 0.01. Ez a kifejezésmód szinte bármilyen hipotézisvizsgálati módszer esetén használható, és azt jelenti, hogy ha elfogadjuk a minta alapján a nullhipotézist, akkor legfeljebb 0.05, illetve 0.01 lehet az elsőfajú hiba valószínűsége (azaz a terjedelem).

Alfejezetünk végén térjünk rá a **kétmintás** u-próbára, amellyel két normális eloszlású minta várható értékének egyenlőségét tesztelhetjük. Ennek itt csak a kétoldali változatát közöljük, ebből az egyoldali változat ugyanúgy vezethető le, mint az egymintás esetben. Kezdjük ismét a kenyeres példával, illetve annak egy folytatásával.

13.2.6. *Példa* (Kétoldali, kétmintás u-próba: kenyérvásárlás). Miután megtudtuk, hogy $\varepsilon = 0.05$ terjedelem mellett el kell utasítanunk azt a nullhipotézist, hogy a péknél vásárolt kenyerek várható értékben egykilósak, alternatív kenyérforrás után nézünk. Lakóhelyünk közelében van egy másik pék, aki nagyon hasonló kinézetű, ízű és árú egykilós kenyereket gyárt, mint az első. Ez a pék azzal reklámozza magát, hogy precíziós eszközök segítségével képes olyan kenyeret gyártani, amelyek tömegének szórása csak 1 dkg (azaz $\sigma_1 = 0.01$ kg, tehát $\sigma_1^2 = 0.0001$ kg^2 a szórásnégyzet). Most $n_1 = 10$ napon keresztül ennél a péknél vásárolunk napi egy kenyeret, és megmérjük a tömegüket. Egy olyan $\mathbf{x} = (x_1, \ldots, x_{n_1})$ realizációt kapunk, amelynek átlaga $\overline{x_{n_1}} = 0.9825$ kg. Vezessünk be egy új, a későbbiekben hasznos jelölést a régi péknél vásárolt kenyerek tömegeire: $\mathbf{y} = (y_1, \ldots, y_{n_2})$, ahol a mintaelemszám $n_2 = 25$, a mintaátlag $\overline{y_{n_2}} = 0.98$, a szórás $\sigma_2 = 0.02$ volt. (Feltételezzük, hogy bármely, a két pék egyike által gyártott kenyér tömege független bármely, ugyanezen pék vagy a másik pék által gyártott kenyér tömegétől.)

Ez alapján kellene eldöntenünk, hogy ugyanúgy járunk-e, ha ehhez a pékhez járunk, mint a másikhoz, vagyis hogy az új péknél vásárolt kenyerek tömegének ismeretlen μ_1 várható értéke ugyanannyi-e, mint a régi péknél vásároltak tömegének μ_2 várható értéke. Ezért a

$$H_0 \colon \mu_1 = \mu_2 \qquad \text{vs.} \qquad H_1 \colon \mu_1 \neq \mu_2$$

alternatívára kívánunk egy $\varepsilon = 0.95$ terjedelmű próbát konstruálni. Feltételezve, hogy az $(X_1, \ldots, X_{n_1})$ minta független az $(Y_1, \ldots, Y_{n_2})$ mintától, az $\overline{X_{n_1}}$ mintaátlag is független az $\overline{Y_{n_2}}$ mintaátlagtól. Ezért $\overline{X_{n_1}}$ és $\overline{Y_{n_2}}$ lineáris kombinációi is normális eloszlásúak. Elemi számításokkal (amiket itt nem részletezünk) megmutatható, hogy

$$u(\mathbf{X}, \mathbf{Y}) = \frac{\overline{X_{n_1}} - \overline{Y_{n_2}}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

standard normális eloszlású. Ennek ismeretében hasonlóan járhatunk el, mint a kétoldali, egymintás u-próba esetén. Ha

$$u(\mathbf{X},\mathbf{Y}) \in [-u_{\varepsilon/2}, u_{\varepsilon/2}] = \big[-\Phi^{-1}(1-\varepsilon/2), \Phi^{-1}(1-\varepsilon/2)\big],$$

akkor H_0-t elfogadjuk, különben pedig elutasítjuk. Vagyis a kritikus tartományunk

$$\mathfrak{X}_{\mathrm{k}} = \big\{(\mathbf{x},\mathbf{y}) = (x_1,\dots,x_{n_1},y_1,\dots,y_{n_2}) \in \mathbb{R}^{n_1+n_2} \colon |u(\mathbf{x},\mathbf{y})| > u_{\varepsilon/2}\big\}.$$

Konkrét példánkban

$$u(\mathbf{x},\mathbf{y}) = \frac{0.9825 - 0.98}{\sqrt{\frac{0.0001}{10} + \frac{0.0004}{25}}} \approx 0{,}4903.$$

Mivel $|u(\mathbf{x},\mathbf{y})| < u_{\varepsilon/2} = \Phi^{-1}(0.975) = 1.9600$, H_0-t elfogadjuk. Vagyis várható értékben az új péknél is ugyanolyan tömegűek a kenyerek, mint a réginél, tehát a pékváltás nem változtat a helyzetünkön...

Ha a próba itt nem részletezett egymintás változatát alkalmaznánk $H_0 \colon \mu_1 \geq \mu_2$ vs. $\mu_1 < \mu_2$ alternatívával, akkor a „változik-e a helyzetünk, ha az új péket választjuk?" helyett a „javul-e a helyzetünk, ha az új péket választjuk?" kérdésre kapnánk választ.

Foglaljuk össze tehát a kétoldali, kétmintás u-próba eljárását:

Eljárás (Kétoldali, kétmintás u-próba)**.** *Feltevések: az $x_1,\dots,x_{n_1}$ adatpontok fae. normális eloszlású $X_1,\dots,X_{n_1}$ valószínűségi változók realizációi, amelyek $\sigma_1 > 0$ szórása ismert, μ_1 várható értéke viszont ismeretlen, az $y_1,\dots,y_{n_2}$ adatpontok pedig az $X_1,\dots,X_{n_1}$-től (együttesen) független, egymás között fae. normális eloszlású $Y_1,\dots,Y_{n_2}$ valószínűségi változók realizációi, ezek $\sigma_2 > 0$ szórása ismert, μ_2 várható értéke viszont ismeretlen. Eljárás:*

1. *Nullhipotézis: $H_0 \colon \mu_1 = \mu_2$ vs. $H_1 \colon \mu_1 \neq \mu_2$.*
2. *Az előirányzott $\varepsilon \in (0,1)$ terjedelem kiválasztása.*
3. *Az $x_1,\dots,x_{n_1},y_1,\dots,y_{n_2}$ adatpontok kinyerése.*
4. *Próbastatisztika:*
$$u(\mathbf{x},\mathbf{y}) = \frac{\overline{x_{n_1}} - \overline{y_{n_2}}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$
5. *Elfogadási tartomány: $\mathfrak{X}_{\mathrm{e}} = \big\{(\mathbf{x},\mathbf{y}) = (x_1,\dots,x_{n_1},y_1,\dots,y_{n_2}) \in \mathbb{R}^{n_1+n_2} \colon |u(\mathbf{x},\mathbf{y})| \leq u_{\varepsilon/2}\big\}$, ahol $u_{\varepsilon/2} = \Phi^{-1}(1-\varepsilon/2)$.*
6. *Döntés: ha $|u(\mathbf{x},\mathbf{y})| > u_{\varepsilon/2}$, akkor H_0-t elutasítjuk, különben elfogadjuk.*

Ismeretlen szórás esetén az u-próba analógja a *t-próba*. Ezt idén előadáson nem tárgyaljuk, de az elve igen hasonló az u-próbáéhoz. Az érdeklődő olvasó a függelékben szereplő B.4 fejezetben vagy az MSc-s Matematikai statisztika tárgyban találkozhat vele.

Függelék

A. Példák becsléselméleti módszerekre

A.1. Maximum likelihood-becslés

A paraméterbecslés egyik módszere a *maximum likelihood módszer*, amelyet ebben az alfejezetben ismertetünk. A módszer lényege, hogy az $\mathbf{X} = (X_1, \dots, X_n)$ n elemű fae. minta $\mathbf{x} = (x_1, \dots, x_n)$ realizációja alapján meghatározzuk azt a $\vartheta_* = \vartheta_*(x_1, \dots, x_n)$-gal jelölt paramétert, amely esetén a legvalószínűbb, hogy (közelítőleg) ezt a realizációt kapjuk. Miután meghatároztuk a $T(x_1, \dots, x_n) = \vartheta_*(x_1, \dots, x_n)$ függvényt, a becslés eredménye a

$$T(X_1, \dots, X_n) = \vartheta_*(X_1, \dots, X_n)$$

statisztika lesz. (A becslés eredménye tehát a véletlen minta egy függvénye, nem pedig a realizációé – ezt érdemes mindig észben tartani!)

Annak a matematikai leírására, hogy egy adott paraméter esetén milyen valószínűséggel kapunk (közelítőleg) egy adott $x_1, \dots, x_n$ realizációt, bevezetjük a *likelihood-függvény* fogalmát. A következő definíció használja a 10.1.1. Definícióban bevezetett p_ϑ súlyfüggvényt és f_ϑ sűrűségfüggvényt.

A.1.1. Definíció. Legyenek $X_1, \dots, X_n$ fae. valószínűségi változók, amelyek eloszlása egy $\vartheta \in \Theta \subseteq \mathbb{R}^d$ paramétertől függ.

(1) Ha X_1 eloszlása bármely $\vartheta \in \Theta$ esetén diszkrét és p_ϑ jelöli a ϑ paraméterhez tartozó súlyfüggvényt, akkor az $\mathbf{X} = (X_1, \dots, X_n)$ n elemű fae. minta bármely $(x_1, \dots, x_n)$ realizációja esetén legyen

$$L_\vartheta(x_1, \dots, x_n) = \prod_{i=1}^{n} p_\vartheta(x_i).$$

(2) Ha X_1 eloszlása bármely $\vartheta \in \Theta$ esetén folytonos és $x \mapsto f_\vartheta(x)$ jelöli a ϑ paraméterhez tartozó sűrűségfüggvényt, akkor az $\mathbf{X} = (X_1, \dots, X_n)$ n elemű fae. minta bármely $(x_1, \dots, x_n)$ realizációja esetén legyen

$$L_\vartheta(x_1, \dots, x_n) = \prod_{i=1}^{n} f_\vartheta(x_i).$$

Az $\mathbb{R}^n \to \mathbb{R}$, $(x_1, \dots, x_n) \mapsto L_\vartheta(x_1, \dots, x_n)$ függvényt mindkét esetben (a ϑ paraméterhez tartozó) **likelihood-függvény**nek nevezzük.

A.1.2. *Megjegyzés.* Diszkrét esetben az L_ϑ likelihood-függvény a mintaelemek együttes eloszlását írja le abban az esetben, ha ϑ a valódi paraméter. Idézzük fel a 10.1.2. Jelölésből, hogy ha A egy esemény, akkor $\mathbb{P}_\vartheta(A)$ jelöli az A valószínűségét, feltéve, hogy ϑ a valódi paraméter. Ezzel a jelöléssel az $X_1, \dots, X_n$ függetlensége miatt azt kapjuk, hogy

$$L_\vartheta(x_1, \dots, x_n) = \prod_{i=1}^{n} p_\vartheta(x_i) = \prod_{i=1}^{n} \mathbb{P}_\vartheta(X_i = x_i) = \mathbb{P}_\vartheta(X_1 = x_1, \dots, X_n = x_n).$$

Folytonos esetben pedig, ha ϑ a valódi paraméter, akkor mivel $f_\vartheta(x_i)$ az X_i valószínűségi változó sűrűségfüggvényének értéke az x_i helyen,

$$L_\vartheta(x_1, \dots, x_n) = \prod_{i=1}^{n} f_\vartheta(x_i)$$

a függetlenség miatt megegyezik az $X_1, \dots, X_n$ valószínűségi változók együttes sűrűségfüggvényével. A likelihood-függvény szemléletes jelentése tehát ebben az esetben

$$L_\vartheta(x_1, \dots, x_n) = \prod_{i=1}^{n} f_\vartheta(x_i) = \prod_{i=1}^{n} \lim_{\Delta x_i \downarrow 0} \frac{1}{\Delta x_i} \mathbb{P}_\vartheta(X_i \in [x_i, x_i + \Delta x_i))$$
$$= \lim_{\Delta x_1 \downarrow 0, \dots, \Delta x_n \downarrow 0} \frac{1}{\Delta x_1 \dots \Delta x_n} \mathbb{P}_\vartheta\big(X_1 \in [x_1, x_1 + \Delta x_1), \dots, X_n \in [x_n, x_n + \Delta x_n)\big).$$

Ezáltal nyer értelmet az a pongyola megfogalmazás, hogy $L_\vartheta(x_1, \dots, x_n)$ azt adja meg, hogy milyen valószínűséggel kapjuk (közelítőleg) $(x_1, \dots, x_n)$-t az $(X_1, \dots, X_n)$ realizációjaként, ha ϑ a valódi paraméter. Diszkrét esetben közelítésre nincs is szükség: $L_\vartheta(x_1, \dots, x_n)$ pontosan annak a valószínűsége, hogy az $(x_1, \dots, x_n)$ realizációt kapjuk. Folytonos esetben ugyanez a valószínűség azonban 0, ezért $L_\vartheta(x_1, \dots, x_n)$ valóban csak azt adja meg, hogy milyen valószínűséggel kapunk egy $(x_1, \dots, x_n)$-hez nagyon közeli realizációt.

A.1.3. Definíció. Legyen $\mathbf{X} = (X_1, \dots, X_n)$ egy n elemű fae. minta, ahol a mintaelemek egy $\vartheta \in \Theta \subseteq \mathbb{R}^d$ paramétertől függő eloszlással rendelkeznek, és tegyük fel, hogy (minden $\vartheta \in \Theta$ esetén) létezik az L_ϑ likelihood-függvény. Legyen $\mathbf{x} = (x_1, \dots, x_n)$ az $\mathbf{X} = (X_1, \dots, X_n)$ egy realizációja.
A **maximum likelihood becslés** ϑ-ra **az** $(x_1, \dots, x_n)$ **realizáció alapján** az a $\vartheta_*(x_1, \dots, x_n) \in \Theta$ paraméter, amely esetén a likelihood-függvény $L_\vartheta(x_1, \dots, x_n)$ értéke maximális. Formálisan kifejezve,

$$\vartheta_*(x_1, \dots, x_n) = \arg\max_{\vartheta \in \Theta} L_\vartheta(x_1, \dots, x_n). \tag{A.1}$$

Más szavakkal, $\vartheta_*(x_1, \dots, x_n)$ az a $\vartheta \in \Theta$ paraméter, amelyre teljesül, hogy bármely $\vartheta' \in \Theta$ esetén

$$L_{\vartheta_*(x_1,\dots,x_n)}(x_1, \dots, x_n) \geq L_{\vartheta'}(x_1, \dots, x_n). \tag{A.2}$$

Elvileg elképzelhető, hogy a (A.2) egyenlet $\vartheta_*(x_1, \dots, x_n)$ több értéke esetén is teljesül, vagyis a maximum nem szigorú és a maximumhely nem egyértelmű, de a gyakorlatban ilyen probléma ritkán fordul elő. Mint ahogy az sem jellemző probléma, hogy a likelihood-függvénynek egyáltalán nincs maximuma.

Amint már utaltunk rá, a maximum likelihood becslés végeredményeként egy statisztikát kapunk, vagyis nem az $(x_1, \dots, x_n)$ realizáció, hanem az $\mathbf{X} = (X_1, \dots, X_n)$ valószínűségi vektorváltozó egy függvényét. Erről szól a következő definíció:

A.1.4. Definíció. Az A.1.3. Definícióban leírt helyzetben vezessük be a

$$T\colon \mathbb{R}^n \to \mathbb{R}, \quad (x_1, \dots, x_n) \mapsto \vartheta_*(x_1, \dots, x_n)$$

függvényt. Ekkor a

$$T(X_1, \dots, X_n) = \vartheta_*(X_1, \dots, X_n)$$

statisztika neve: **maximum likelihood becslés** ϑ-ra.

Hogyan lehet a maximum likelihood-becslést a gyakorlatban kiszámítani? Az A.1 egyenletben szereplő, ϑ szerinti maximalizálási probléma egy egydimenziósszélsőérték-számítási feladat, ami elvileg az első féléves analízistudásunkkal megoldható, ha a likelihood-függvény differenciálható. Azonban a likelihood-függvény még viszonylag egyszerű súly- vagy sűrűségfüggvénnyel rendelkező eloszlások esetén is egy n tényezős szorzat, amelynek a szorzat- és láncszabály alkalmazásával való deriválása hosszadalmas és kellemetlen művelet. Gyakran segít, ha a likelihood-függvény helyett annak a logaritmusát próbáljuk ϑ szerint maximalizálni, ugyanis a logaritmus a szorzatokat összegekké alakítja, amit lényegesen könnyebb deriválni. Ezért vezetjük be a következő fogalmat.

A.1.5. Definíció. Ha $(x_1, \dots, x_n) \mapsto L_\vartheta(x_1, \dots, x_n)$ egy likelihood-függvény, akkor az

$$\mathbb{R}^n \to \mathbb{R}, \quad (x_1, \dots, x_n) \mapsto l_\vartheta(x_1, \dots, x_n) = \ln L_\vartheta(x_1, \dots, x_n)$$

függvényt **log-likelihood-függvénynek** nevezzük.

Jegyezzük meg, hogy a log-likelihood-függvény minden likelihood-függvény esetén jóldefiniált, mert a likelihood-függvény definíciójában szereplő súly- és sűrűségfüggvények minden realizációban vett értéke pozitív. (Ezért is volt szükségünk a lényeges értékek halmaza és realizáció – viszonylag bonyolult – definíciójára!) A log-likelihood-függvény pedig ugyanott veszi fel a maximumát, mint a likelihood-függvény:

A.1.6. Állítás. *Ha $(x_1, \dots, x_n) \mapsto L_\vartheta(x_1, \dots, x_n)$ egy likelihood-függvény és $(x_1, \dots, x_n) \mapsto l_\vartheta(x_1, \dots, x_n) = \ln L_\vartheta(x_1, \dots, x_n)$ az ehhez tartozó log-likelihood-függvény, akkor $\vartheta \in \Theta$ esetén $L_\vartheta(x_1, \dots, x_n)$ pontosan akkor maximális ϑ-ban, ha $l_\vartheta(x_1, \dots, x_n)$ maximális ϑ-ban.*

Bizonyítás. Az állítás a (természetes alapú) logaritmusfüggvény szigorú monotonitásából következik. □

Ha a likelihood-függvényt log-likelihood-függvénnyé alakítjuk, akkor a maximum likelihood-becslésre az alábbi eljárást adhatjuk (ez az eljárás csak akkor működik, ha a log-likelihood-függvény differenciálható):

Eljárás (Maximum likelihood-becslés log-likelihood-függvénnyel)**.**

1. *Rögzített $x_1, \dots, x_n$ realizáció és $\vartheta \in \Theta$ paraméter esetén meghatározzuk a likelihood-függvény $L_\vartheta(x_1, \dots, x_n)$ értékét.*
2. *Vesszük ennek (*e *alapú) logaritmusát, így megkapjuk $l_\vartheta(x_1, \dots, x_n)$ értékét.*
3. *Az a $\vartheta \mapsto l_\vartheta(x_1, \dots, x_n)$ függvényt deriváljuk (ϑ szerint).*
4. *Megvizsgáljuk, hogy a derivált hol egyenlő 0-val.*
5. *Megvizsgáljuk, hogy a derivált zérushelyei lokális maximumhelyek-e. (Emlékeztetünk arra, hogy lokális maximumhellyel állunk szemben, ha a második derivált az adott helyen negatív, vagy ha az első derivált pozitívból negatívba vált előjelet az adott pontban.) Ha esetleg több lokális maximumhely is van, akkor ezek közül az lesz a globális maximumhely, ahol a likelihood-függvény értéke a legnagyobb.*[13]
6. *Tegyük fel, hogy egy egyértelmű $\vartheta_*(x_1, \dots, x_n)$ globális maximumhelyet kaptunk. Ekkor $\vartheta_*(X_1, \dots, X_n)$ lesz a keresett maximum likelihood becslés.*

Az alábbiakban bemutatjuk a módszert egy diszkrét és egy folytonos példán, egyéb példákkal pedig a gyakorlatokon (és a vizsgákon) találkozhatunk.

A.1.7. *Példa* (Maximum likelihood módszer, diszkrét eset: Poisson-eloszlás). A fenti eljárást alkalmazzuk a Poisson-eloszlás esetére: tegyük fel, hogy az $X_1, \dots, X_n$ mintaelemek peremeloszlása Poisson-eloszlás ismeretlen $\vartheta > 0$ paraméterrel. Legyen $(x_1, \dots, x_n)$ az $(X_1, \dots, X_n)$ egy realizációja, vagyis egy n darab nemnegatív egész számból álló vektor. A továbbiakban először tegyük fel, hogy az $x_1, \dots, x_n$ számok közül legalább az egyik nem nulla.

1. Mivel
$$p_\vartheta(x_i) = \mathbb{P}_\vartheta(X_i = x_i) = \frac{\vartheta^{x_i}}{x_i!}\mathrm{e}^{-\vartheta}, \qquad i = 1, \dots, k,$$
ezért
$$L_\vartheta(x_1, \dots, x_n) = \prod_{i=1}^n p_\vartheta(x_i) = \prod_{i=1}^n \frac{\vartheta^{x_i}}{x_i!}\mathrm{e}^{-\vartheta} = \frac{\vartheta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}\mathrm{e}^{-n\vartheta}.$$
2. A log-likelihood függvényre tehát a logaritmus azonosságait használva azt kapjuk, hogy
$$l_\vartheta(x_1, \dots, x_n) = \ln\Big(\frac{\vartheta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}\mathrm{e}^{-n\vartheta}\Big) = \ln\vartheta\sum_{i=1}^n x_i - \sum_{i=1}^n \ln(x_i!) - n\vartheta.$$

[13] Amennyiben a Θ paramétertartomány nem az egész $\mathbb{R}^d$, akkor vizsgálni kell azt is, hogy a tartomány határán nincs-e maximumhely.

3. A log-likelihood függvényt ϑ szerint deriváljuk:

$$\frac{\partial}{\partial\vartheta} l_\vartheta(x_1,\dots,x_n) = \frac{1}{\vartheta}\sum_{i=1}^{n} x_i - n.$$

4. Láthatjuk, hogy a derivált alakja meglehetősen egyszerű, vagyis a logaritmus alkalmazása valóban hasznos volt. A derivált zérushelyét is könnyű meghatározni:

$$\frac{\partial}{\partial\vartheta} l_\vartheta(x_1,\dots,x_n) = 0 \qquad \Leftrightarrow \qquad \frac{1}{\vartheta}\sum_{i=1}^{n} x_i - n = 0 \qquad \Leftrightarrow \qquad \vartheta = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x_n}.$$

Vagyis a log-likelihood-függvény (és a likelihood-függvény) egyetlen zérushelye a mintaátlag realizációja, vagyis a realizáció elemeinek számtani közepe.

5. Ebben a pontban a log-likelihood-függvénynek szigorú lokális maximuma van, mivel a második derivált értéke bármely $\vartheta \in \Theta$ pontban negatív:

$$\frac{\partial^2}{\partial\vartheta^2} l_\vartheta(x_1,\dots,x_n) = -\frac{1}{\vartheta^2}\sum_{i=1}^{n} x_i < 0.$$

Mivel ez az egyetlen lokális szélsőérték van és a $\Theta = (0,\infty)$ paramétertartomány egy nyílt intervallum, a kapott maximum globális is.

6. A log-likelihood-függvény tehát a

$$\vartheta_*(x_1,\dots,x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x_n}$$

pontban lesz maximális. Így a maximum likelihood becslésünk:

$$\vartheta_*(X_1,\dots,X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X_n}. \tag{A.3}$$

Ha $x_1 = \ldots = x_n = 0$, akkor a likelihood-függvénynek nincs a $(0,\infty)$ intervallumon lokális maximuma. A globális maximum ekkor a Θ határán, a $\vartheta = 0$ pontban vétetik fel. (Interpretáció: minél kisebb $\vartheta > 0$, annál valószínűbb, hogy minden mintaelem értéke 0.) Ebben az esetben tehát $\vartheta_*(x_1,\dots,x_n) = 0$, vagyis ekkor is igaz, hogy $\vartheta_*(x_1,\dots,x_n) = \overline{x_n}$. Az A.3 egyenlet tehát mindig teljesül.

A.1.8. *Példa* (Maximum likelihood módszer, folytonos eset: exponenciális eloszlás). A fenti eljárást alkalmazzuk az exponenciális eloszlás esetére: tegyük fel, hogy az $X_1,\dots,X_n$ mintaelemek peremeloszlása exponenciális eloszlás ismeretlen $\vartheta > 0$ paraméterrel. Legyen $(x_1,\dots,x_n)$ az $(X_1,\dots,X_n)$ egy realizációja, vagyis egy n darab pozitív valós számból álló vektor.

1. A mintaelemek sűrűségfüggvénye

$$f_\vartheta(x_i) = \vartheta \mathrm{e}^{-\vartheta x_i}, \qquad i = 1,\dots,n,$$

így a likelihood-függvényre a következő kifejezés adódik:

$$L_\vartheta(x_1,\dots,x_n) = \prod_{i=1}^{n} f_\vartheta(x_i) = \prod_{i=1}^{n} \vartheta \mathrm{e}^{-\vartheta x_i} = \vartheta^n \mathrm{e}^{-\vartheta\sum_{i=1}^{n} x_i}.$$

2. A log-likelihood-függvény tehát

$$l_\vartheta(x_1,\dots,x_n) = \ln\left(\vartheta^n \mathrm{e}^{-\vartheta\sum_{i=1}^n x_i}\right) = n\ln\vartheta - \vartheta\sum_{i=1}^n x_i.$$

3. A log-likelihood függvényt ϑ szerint deriváljuk:

$$\frac{\partial}{\partial\vartheta}\ln l_\vartheta(x_1,\dots,x_n) = \frac{n}{\Theta} - \sum_{i=1}^n x_i.$$

4. A derivált zérushelyét ismét egyszerű meghatározni:

$$\frac{\partial}{\partial\vartheta}\ln l_\vartheta(x_1,\dots,x_n) = 0 \qquad \Leftrightarrow \qquad \frac{n}{\Theta} - \sum_{i=1}^n x_i = 0 \qquad \Leftrightarrow \qquad \Theta = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\overline{x_n}}.$$

5. Mivel

$$\frac{\partial^2}{\partial\vartheta^2}\ln l_\vartheta(x_1,\dots,x_n) = -\frac{n}{\vartheta^2}$$

minden $\vartheta > 0$ esetén negatív, a kapott zérushely lokális maximumhelye a log-likelihood-függvénynek. Mivel ez az egyetlen lokális maximumhely és $\Theta = (0,\infty)$ egy nyílt intervallum, ez egyben a (log-)-likelihood-függvény egyetlen globális maximumhelye.

6. A maximum-likelihood becslés az $(x_1,\dots,x_n)$ realizáció alapján tehát

$$\vartheta_*(x_1,\dots,x_n) = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\overline{x_n}}.$$

Így tehát a keresett maximum likelihood becslés

$$\vartheta_*(X_1,\dots,X_n) = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\overline{X_n}}.$$

A.1.9. *Megjegyzés.* Az előző két példa azt is megmutatja, hogy a maximum likelihood becslés lehet torzítatlan becslése ϑ-nak, de nem mindig az. Ugyanis az ismeretlen ϑ paraméterű Poisson-eloszlás esetén ϑ éppen megegyezik $\mathbb{E}_\vartheta(X_1)$-gyel, a mintaelemek várható értékére pedig az $\vartheta_*(X_1,\dots,X_n) = \overline{X_n}$ mintaátlag mindig torzítatlan becslés (vö. 11.0.3. Példa). Ezért ebben az esetben a maximum likelihood becslés torzítatlan becslése a ϑ paraméternek.

Ugyanakkor az exponenciális eloszlás esetén

$$\mathbb{E}_\vartheta\big(\vartheta_*(X_1,\dots,X_n)\big) = \mathbb{E}_\vartheta\big(\frac{n}{\sum_{i=1}^n X_i}\big) \neq \frac{n}{\sum_{i=1}^n \mathbb{E}_\vartheta(X_i)} = \frac{1}{\frac{1}{\vartheta}} = \vartheta.$$

Tehát ha ϑ a valódi paraméter, akkor $\mathbb{E}_\vartheta(\vartheta_*(X_1,\dots,X_n))$ nem egyezik meg ϑ-val, tehát a maximum likelihood becslés nem torzítatlan becslése ϑ-nak.[14]

[14] Az előbbi „$\neq$” állítást itt részletesen nem bizonyítjuk, de általában igaz az, hogy bármilyen szigorúan konvex f függvény és X valószínűségi változó esetén $\mathbb{E}(f(X)) \geq f(\mathbb{E}(X))$, és egyenlőség akkor és csak akkor áll fenn, ha X 1 valószínűséggel konstans. (Ennek egy speciális esete az az általunk már ismert állítás, hogy $\mathbb{E}(X^2) \geq \mathbb{E}(X)^2$, – tehát $\mathbb{D}^2(X) \geq 0$ – és egyenlőség – vagyis $\mathbb{D}^2(X) = 0$ – akkor és csak akkor áll fenn, ha X 1 valószínűséggel konstans.) Itt $f(x) = 1/x$ (ami $(0,\infty)$-n szigorúan konvex) és $X = \sum_{i=1}^n X_i$ (és $\mathbb{P}$ helyett $\mathbb{P}_\vartheta$-ról van szó).

A likelihood-függvény log-likelihood-függvénnyé alakítása különösen abban az esetben segíthet, ha a mintát alkotó független, azonos eloszlású X_i valószínűségi változók súly- illetve sűrűségfüggvényében exponenciális tényezők is szerepelnek. Erre példa a Poisson-eloszlás, az exponenciális eloszlás vagy független exponenciálisok összegének eloszlása, illetve a normális eloszlás – amelynek a szituációtól függően lehet egy vagy két ismeretlen paramétere is. Vannak azonban olyan eloszlások is, ahol érdemes közvetlenül a likelihood-függvényt maximalizálni, anélkül, hogy a logaritmusát vennénk, sőt a maximalizálást nem feltétlenül deriválás segítségével érdemes végezni. Lássunk erre is egy példát!

A.1.10. *Példa* (Maximum likelihood módszer: egyenletes eloszlás ismeretlen alsó végpontú intervallumon). Legyenek $X_1, \dots, X_n$ független és $U(-\vartheta; 0)$ eloszlású, tehát a $(-\vartheta, 0)$ intervallumon egyenletes eloszlású valószínűségi változók, ahol $\vartheta > 0$ egy (ismeretlen) paraméter. Adjuk meg a $\vartheta_*(X_1, \dots, X_n)$ maximum likelihood becslést!

Az egyenletes eloszlás tulajdonságai miatt az f_ϑ sűrűségfüggvény

$$f_\vartheta(x) = \begin{cases} \frac{1}{\vartheta}, & \text{ha } x \in (-\vartheta, 0), \\ 0, & \text{különben.} \end{cases}.$$

Így a likelihood-függvény $(X_1, \dots, X_n)$ bármely $(x_1, \dots, x_n)$ realizációja esetén

$$L_\vartheta(x_1, \dots, x_n) = \begin{cases} \frac{1}{\vartheta^n}, & \text{ha } x_1, x_2, \dots, x_n \in (-\vartheta, 0), \\ 0, & \text{különben.} \end{cases}$$

Vegyük észre, hogy ebben az esetben X_1 lényeges értékeinek $S_{X_1}^{(\vartheta)} = (-\vartheta, 0)$ halmaza függ a ϑ paramétertől, az intervallum hosszának csökkentésével pedig a likelihood-függvény értéke az $(x_1, \dots, x_n)$ pontban szigorúan monoton csökken. Az intervallum hosszát mindaddig csökkenthetjük, amíg az $(x_1, \dots, x_n)$ az $(X_1, \dots, X_n)$ egy realizációja marad, azaz amíg

$$\min\{x_1, \dots, x_n\} > -\vartheta,$$

azaz $\vartheta > -\min\{x_1, \dots, x_n\}$. Így a likelihood-függvény éppen abban a határesetben lesz maximális, amikor $-\min\{x_1, \dots, x_n\} = \vartheta$. (Ebben az esetben $(x_1, \dots, x_n)$ már éppen nem realizáció, de bármely $\vartheta > -\min\{x_1 \dots, x_n\}$ esetén igen.) Ezért az $(x_1, \dots, x_n)$ realizációhoz tartozó maximum likelihood becslésünk

$$\vartheta_*(x_1, \dots, x_n) = -\min\{x_1, \dots, x_n\},$$

a maximum likelihood becslésünk pedig

$$\vartheta_*(X_1, \dots, X_n) = -\min\{X_1, \dots, X_n\} = -X_1^*.$$

Zárásul lássunk egy olyan nevezetes példát a maximum likelihood módszerre, ahol az ismeretlen paraméter lehetséges értékeinek száma megszámlálható, ezért a maximalizálás során a (log-)likelihood-függvény deriválása szóba sem jöhet.

A.1.11. *Példa* (Hány hal van a tóban? – avagy a hipergeometriai eloszlás maximum likelihood becslése). Egy tóban ismeretlen számú hal él (a halak száma egy pozitív egész szám). A környékbeli halászok szeretnék humánus módon megbecsülni, hogy hány hal él a tóban, ezért kifognak $j \geq 1$ darab különböző halat, ezeket megjelölik egy-egy olyan chippel, ami nem ártalmas a halakra, majd visszaengedik őket a tóba. Ezután pedig kifognak $n \geq 1$ darab különböző halat a tóból és ezek között $0 \leq x \leq n$ darab jelöltet találnak. Mi a maximum likelihood becslésünk a jelöletlen halak k számára?

Az egyetlen ismeretlen paraméterünk k, amely egy nemnegatív egész szám. Mivel a különböző halak választása egy urnából való visszatevés nélküli húzásnak felel meg[15], a likelihood-függvény

$$L_k(x) = \mathbb{P}_k(n \text{ kifogott halból pontosan } x \text{ jelölt}) = \frac{\binom{j}{x}\binom{k}{n-x}}{\binom{j+k}{n}}.$$

[15]Egy olyan urnamodell esetén, ahol az urnában lévő minden golyó vagy kedvező, vagy kedvezőtlen, és az urnából rögzített számú golyót húzunk visszatevés nélkül, a kihúzott kedvező golyók számának eloszlását *hipergeometriai* eloszlásnak nevezzük.

(Ezt $L_k(x)$ helyett jelölhettük volna $p_k(x)$-szel is, hiszen az x szám egy egyetlen elemű minta realizációja, tehát a likelihood-függvény tulajdonképpen az ismeretlen k paramétertől függő súlyfüggvény.) A feladatunk tehát az, hogy ezt a likelihood-függvényt a k paraméter szerint maximalizáljuk, amely csak nemnegatív egész értékeket vehet fel (a paramétertartomány $\Theta = \mathbb{N}$). Ebben segítene, ha tudnánk, hogy $k = 0$-tól indulva és k-t egyesével növelve $L_k(x)$ egy darabig szigorúan monoton növekszik, majd legfeljebb egyszer nem változik, ami után szigorúan monoton csökkenésbe kezd. Igazoljuk először ezt az állítást. Ha $k \geq 1$, akkor

$$\frac{L_k(x)}{L_{k-1}(x)} = \frac{\binom{k}{n-x}\binom{j+k-1}{n}}{\binom{k-1}{n-x}\binom{j+k}{n}} = \frac{k(j+k-n)}{(k-(n-x))(j+k)} \geq 1$$

pontosan akkor teljesül, ha

$$k \leq \frac{jn}{x} - j,$$

illetve ugyanez igaz „$\geq$” helyett „$>$”-vel és „$\leq$” helyett „$<$”-vel az utóbbi két egyenlőtlenségnél. Így ha $\frac{jn}{x} \notin \mathbb{N}$, akkor $k = \lfloor \frac{jn}{x} \rfloor - j$ a maximum likelihood becslés k-ra, mert k ezen értékéig a likelihood-függvény szigorúan monoton nő, innentől kezdve pedig szigorúan monoton csökken. Ha viszont $\frac{jn}{x} \in \mathbb{N}$, akkor két maximum likelihood becslés is van (azonos maximumértékkel): $k = \frac{jn}{x} - j$ és $k = \frac{jn}{x} - j - 1$. Az összes hal számára, $j + k$-ra a maximum likelihood becslés értelemszerűen $\lfloor \frac{jn}{x} \rfloor$, ha $\frac{jn}{x} \notin \mathbb{N}$, illetve $\frac{jn}{x}$ és $\frac{jn}{x} - 1$, ha $\frac{jn}{x} \in \mathbb{N}$.

A.2. Momentumbecslés

Egy másik fontos paraméterbecslési módszer a *momentumok módszere*. Ehhez először vezessük be a momentumok és az empirikus momentumok fogalmát:

A.2.1. Definíció. Legyen k egy pozitív egész szám és X egy valószínűségi változó. Ha $\mathbb{E}[|X|^k] < \infty$, akkor az X k**. momentumán** az $\mathbb{E}[X^k] \in \mathbb{R}$ mennyiséget értjük.

A.2.2. Definíció. Legyen k egy pozitív egész szám és legyenek $X_1, \ldots, X_n$ független, azonos eloszlású valószínűségi változók. Az $(X_1, \ldots, X_n)$ n-elemű fae. minta k**-adik empirikus momentumán** az

$$\widehat{m}_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

statisztikát értjük.

A momentumok módszere különösen akkor lesz hasznos, ha egynél több paraméterünk van. Ilyen például a normális eloszlás ismeretlen μ várható értékkel és ismeretlen σ^2 szórásnégyzettel, ahol az ismeretlen paraméterek vektora $\underline{\vartheta} = (\mu, \sigma^2) \in \mathbb{R}^2$, ahol $\mu \in \mathbb{R}$ és $\sigma^2 > 0$. Vagy ilyen például az egyenletes eloszlás egy ismeretlen kezdő- és végpontú (a, b) intervallumon, ahol a paramétervektor $\underline{\vartheta} = (a, b) \in \mathbb{R}^2$ valamely $-\infty < a < b < \infty$ valós számokra.

Általában tegyük fel, hogy az $\mathbf{X} = (X_1, \ldots, X_n)$ fae. minta elemeinek eloszlását meghatározó ismeretlen paraméterek vektora $\underline{\vartheta} = (\vartheta_1, \ldots, \vartheta_k)$, amely egy $\Theta \subseteq \mathbb{R}^k$ halmaz egy eleme. Tegyük fel azt is, hogy az $X_1, \ldots, X_n$ mintaelemeknek a paraméter bármely választása esetén létezik az első k momentuma. Ezek a momentumok természetesen a $\vartheta_1, \ldots, \vartheta_k$ paraméterek függvényei:

$$\begin{aligned} m_1 &= g_1(\vartheta_1, \ldots, \vartheta_k), \\ m_2 &= g_2(\vartheta_1, \ldots, \vartheta_k), \\ &\vdots \\ m_k &= g_k(\vartheta_1, \ldots, \vartheta_k) \end{aligned}$$

valamely $g_1, g_2, \dots, g_k \colon \mathbb{R}^k \to \mathbb{R}$ függvényre. A

$$(\vartheta_1, \dots, \vartheta_k) \mapsto \underline{g}(\vartheta_1, \dots, \vartheta_k) := (g_1(\vartheta_1, \dots, \vartheta_k), \dots, g_k(\vartheta_1, \dots, \vartheta_k))$$

leképezés tehát egy $\underline{g} \colon \mathbb{R}^k \to \mathbb{R}^k$ függvény, vagyis egy vektormező. Ha ennek a vektormezőnek létezik inverze, akkor jelöljük ezt $\underline{h} = (h_1, \dots, h_k)$-val. Ezen inverz segítségével a momentumok ismeretében megkaphatjuk a $\underline{\vartheta} = (\vartheta_1, \dots, \vartheta_k)$ paramétert:

$$\begin{aligned} \vartheta_1 &= h_1(m_1, \dots, m_k), \\ \vartheta_2 &= h_2(m_1, \dots, m_k), \\ &\vdots \\ \vartheta_k &= h_k(m_1, \dots, m_k) \end{aligned}$$

A.2.3. Definíció. A fenti szituációban az $\underline{\vartheta} = (\vartheta_1, \dots, \vartheta_n)$ paramétervektor **momentumbecslésé**n az $\widehat{\underline{\vartheta}} = (\widehat{\vartheta}_1, \dots, \widehat{\vartheta}_k)$ vektort értjük, ahol

$$\widehat{\vartheta}_i = h_i(\widehat{m}_1, \dots, \widehat{m}_k), \qquad i = 1, \dots, k,$$

és $\widehat{m}_i$ az i-edik empirikus momentumot jelöli.

Differenciálható $\underline{g}$ leképezés esetén az $\underline{h} = (h_1, \dots, h_k)$ inverz létezésének szükséges feltétele, hogy $\underline{g}$ Jacobi-determinánsa a Θ halmazon sehol se legyen 0. Lássunk példákat a momentumbecslésre:

A.2.4. *Példa* (Momentumbecslés: normális eloszlás). Legyenek $X_1, \dots, X_n$ független és egyenként $\mathrm{N}(\mu; \sigma^2)$ eloszlású valószínűségi változók, ahol $\mu \in \mathbb{R}$ és $\sigma^2 > 0$. Az első $k = 2$ momentumot $\underline{\vartheta} = (\mu, \sigma^2)$ esetében ekkor a következőképpen kaphatjuk meg:

$$\begin{aligned} m_1 &= \mathbb{E}_{\underline{\vartheta}}(X_1) = \mu, \\ m_2 &= \mathbb{E}_{\underline{\vartheta}}(X_1^2) = \mathbb{D}^2_{\underline{\vartheta}}(X_1) + \mathbb{E}_{\underline{\vartheta}}(X_1)^2 = \sigma^2 + \mu^2. \end{aligned}$$

Ebben az esetben a $\underline{g} \colon (\mu, \sigma^2) \mapsto (\mu, \sigma^2 + \mu^2)$ leképezés természetesen invertálható[16]:

$$\begin{aligned} \mu &= m_1, \\ \sigma^2 &= m_2 - m_1^2, \end{aligned}$$

azaz az inverz függvény képlete $\underline{h} \colon (m_1, m_2) \mapsto (m_1, m_2 - m_1^2)$. Így a momentumbecslésünk

$$\begin{aligned} \widehat{\mu} &= \widehat{m}_1 = \overline{X_n}, \\ \widehat{\sigma^2} &= \widehat{m}_2 - \widehat{m}_1^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \overline{X_n}^2. \end{aligned}$$

Az $\widehat{m}_2 - \widehat{m}_1^2$ pedig éppen a (11.2) egyenletben definiált S_n^2 statisztikával (azaz az empirikus szórásnégyzettel, vagyis a korrigált empirikus szórásnégyzet $(n-1)/n$-szeresével) egyezik meg, ugyanis

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X_n})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{2}{n} \sum_{i=1}^n X_i \overline{X_n} + \frac{1}{n} \sum_{i=1}^n \overline{X}_n^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - 2\overline{X_n}^2 + \overline{X_n}^2 = \widehat{m}_2 - \widehat{m}_1^2. \quad \text{(A.4)}$$

Ez az egyenlőség pedig bármilyen eloszlású minta esetén igaz.

[16] Ebben az esetben a Jacobi-mátrixot úgy kell kiszámítani, hogy változóként σ^2-et (és nem σ-t) tekintjük, így a Jacobi-mátrix:

$$\begin{bmatrix} \frac{\partial g_1}{\partial \mu}(\mu, \sigma^2) & \frac{\partial g_1}{\partial (\sigma^2)}(\mu, \sigma^2) \\ \frac{\partial g_2}{\partial \mu}(\mu, \sigma^2) & \frac{\partial g_2}{\partial (\sigma^2)}(\mu, \sigma^2) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2\mu & \sigma^2 \end{bmatrix}.$$

Ennek determinánsa σ^2, ami a Θ paramétertartományon sehol sem 0.

A.2.5. *Példa* (Momentumbecslés: egyenletes eloszlás). Tegyük fel, hogy az $X_1, \dots, X_n$ független, azonos eloszlású valószínűségi változók az (a, b) intervallumon egyenletes eloszlásúak, ahol most az ismeretlen paraméterek vektora $\underline{\vartheta} = (a, b)$, $-\infty < a < b < \infty$. Ekkor az egyenletes eloszlás tulajdonságai alapján az első $k = 2$ momentum a következő:

$$m_1 = \mathbb{E}_{\underline{\vartheta}}(X_1) = \frac{a+b}{2},$$
$$m_2 = \mathbb{E}_{\underline{\vartheta}}(X_1^2) = \mathbb{D}^2_{\underline{\vartheta}}(X_1) + \mathbb{E}_{\underline{\vartheta}}(X_1)^2 = \frac{(b-a)^2}{12} + \Big(\frac{a+b}{2}\Big)^2 = \frac{(b-a)^2 + 3(a+b)^2}{12}.$$

Tehát azt kapjuk, hogy

$$\frac{(b-a)^2}{12} = m_2 - m_1^2 \qquad \Rightarrow \qquad b - a = 2\sqrt{3(m_2 - m_1^2)}$$

és

$$a + b = 2m_1.$$

(Vagyis itt is invertálható leképezéssel állunk szemben; ellenőrizhetjük, hogy az a, b szerinti deriválással kapott Jacobi-mátrix determinánsa $-\infty < a < b < \infty$ esetén sohasem 0.) Ezzel tehát egy lineáris egyenletrendszert kaptunk a-ra és b-re, ami $-\infty < a < b < \infty$ esetén mindig egyértelműen megoldható. A megoldás

$$a = m_1 - \sqrt{3(m_2 - m_1^2)},$$
$$b = m_1 + \sqrt{3(m_2 - m_1^2)}.$$

Felhasználva az $\widehat{m}_2 - \widehat{m}_1^2 = S_n^2$ egyenletet (vö. (A.4)), azt kapjuk, hogy a momentumbecslés

$$\widehat{a} = \widehat{m}_1 - \sqrt{3(\widehat{m}_2 - \widehat{m}_1^2)} = \overline{X_n} - \sqrt{3S_n^2},$$
$$\widehat{b} = \widehat{m}_1 + \sqrt{3(\widehat{m}_2 - \widehat{m}_1^2)} = \overline{X_n} + \sqrt{3S_n^2}.$$

B. További konfidenciaintervallumok, t- és χ^2-eloszlás, t-próba

Ebben a fejezetben először a B.1. fejezetben bevezetünk két újabb nevezetes eloszláscsaládot, a χ^2- és a Student-eloszlásokat, és megismerjük ezek kapcsolatát a normális eloszlású fae. mintákkal. A Student-eloszlás a 13. fejezetben, a hipotézisvizsgálatnál is fontos szerephez fog jutni. Az itt ismertetett állításokat nem bizonyítjuk, a bizonyítások iránt érdeklődő olvasó számára a mérnök- és gazdaságinformatikus MSc képzés Matematika statisztika tárgyát (vagy bármelyik matematikus BSc képzés megfelelő tárgyát) ajánljuk. Ezek után a B.2. fejezetben az ismert várható értékű normális eloszlás szórására, a B.3. fejezetben pedig az ismeretlen szórású normális eloszlás várható értékére vonatkozó konfidenciaintervallum konstrukcióját ismertetjük.

B.1. χ^2-eloszlás, Student-eloszlás és tulajdonságaik

B.1.1. Definíció. Legyenek $X_1, \dots, X_n$ fae. $\mathrm{N}(0;1)$ eloszlású valószínűségi változók. Ekkor $\sum_{i=1}^n X_i^2$ eloszlását n **szabadságfokú χ^2-eloszlásnak** (ejtsd: khí-négyzet-eloszlásnak) nevezzük. Jelölés: $X \sim \chi^2(n)$.

B.1.2. Állítás. *Legyen egy n pozitív egész szám. Ha az X valószínűségi változó n szabadságfokú χ^2-eloszlású, akkor a sűrűségfüggvénye*

$$f_n(x) = \begin{cases} \frac{x^{n/2-1}\mathrm{e}^{-x/2}}{2^{n/2}\Gamma(n/2)}, & \text{ha } x > 0, \\ 0, & \text{ha } x \le 0. \end{cases}.$$

*Itt Γ az úgynevezett **Euler-féle Gamma-függvény**, amelyre $\Gamma(n/2)$ rekurzívan a $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$ és $\Gamma(r+1) = r\Gamma(r)$, $r \in [0, \infty)$ egyenletekből számítható ki.*

B.1.3. *Megjegyzés.* 1. Vegyük észre, hogy $n \geq 1$ esetén

$$\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)(n-2)\Gamma(n-2) = \ldots = (n-1)(n-2)\ldots\cdot 1\Gamma(1) = (n-1)!$$

A Γ függvény folytonos (sőt analitikus) módon kiterjeszthető az $\mathbb{R} \setminus \{0, -1, -2, \ldots\}$ halmazra, ez tehát a faktoriális egyfajta kiterjesztése majdnem minden valós számra.

2. Megfigyelhetjük azt is, hogy $n = 2$ esetén

$$f_n(x) = \begin{cases} \frac{\mathrm{e}^{-x/2}}{2}, & \text{ha } x > 0, \\ 0, & \text{ha } x \leq 0, \end{cases}$$

vagyis a 2 szabadságfokú χ^2-eloszlás megegyezik a $\lambda = 1/2$ paraméterű exponenciális eloszlással. Általában igaz (de itt nem bizonyítjuk), hogy ha n páros, akkor az n szabadságfokú χ^2-eloszlás megegyezik $n/2$ db független Exp(1/2) eloszlású valószínűségi változó összegének eloszlásával.

B.1.4. Definíció. Legyen n egy pozitív egész szám. Az X valószínűségi változó eloszlását n **szabadságfokú t-eloszlásnak**, más szóval n **szabadságfokú Student-eloszlásnak** nevezzük, ha a sűrűségfüggvénye

$$f_n(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \Big(1 + \frac{x^2}{n}\Big)^{-(n+1)/2}, \qquad x \in \mathbb{R},$$

ahol Γ ismét az Euler-féle Gamma-függvény. Jelölés: $X \sim t(n)$.

B.1.5. *Megjegyzés.* 1. A ‚Student" név az eloszlást felfedező William Sealy Gosset angol statisztikus álneve volt.

2. Az 1 szabadságfokú,

$$f_1(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

sűrűségfüggvényű t-eloszlást ismertebb nevén *Cauchy-eloszlásnak* is nevezik, ez a valószínűségszámítás számos területén fontos szerepet játszik. A Cauchy-eloszlású X valószínűségi változó érdekes tulajdonsága, hogy $\mathbb{E}[|X|] = \infty$, és emiatt az $\mathbb{E}[X]$ várható érték nem létezik. (A várható érték nem definiálható még $+\infty$-ként vagy $-\infty$-ként sem, mert a sűrűségfüggvény a 0-ra szimmetrikus.) Általában az n szabadságfokú t-eloszlásnak az első $n-1$ momentuma létezik, az n-edik már nem.

B.1.6. Állítás. *Legyenek $X_1, \ldots, X_n$ független és $\mathrm{N}(\mu; \sigma^2)$ eloszlású valószínségi változók. Jelölje $\overline{X_n}$ a hozzájuk tartozó mintaátlagot és $S_n^* = \sqrt{(S_n^*)^2}$ a korrigált empirikus szórásukat. Ekkor a*

$$\frac{\sqrt{n}(\overline{X_n} - \mu)}{S_n^*}$$

valószínűségi változó t-eloszlású $n-1$ szabadságfokkal.

B.2. Konfidenciaintervallum normális eloszlás szórásnégyzetére ismert várható érték esetén

Az ismert μ várható értékű és ismeretlen σ^2 szórásnégyzetű normális eloszlás szórásnégyzetére vonatkozó konfidenciaintervallum meghatározása az alábbi tételre épül, amelyet itt nem bizonyítunk:

B.2.1. Tétel (Lukács Jenő tétele)**.** *Legyenek $X_1, \ldots, X_n \sim \mathrm{N}(\mu; \sigma^2)$ fae. valószínűségi változók. Ekkor a*

$$\frac{(n-1)(S_n^*)^2}{\sigma^2}$$

valószínűségi változó χ^2-eloszlású $n-1$ szabadságfokkal, továbbá $\overline{X_n}$ és $(S_n^*)^2$ függetlenek.

Ez alapján most levezetjük a szórásnégyzetre vonatkozó konfidenciaintervallum képletét ismeretlen várható érték esetén, majd ennek felhasználásával tárgyaljuk az ismert várható érték esetét.

B.2.2. *Példa.* A várható értékre való konfidenciaintervallumhoz képest most az a jelentős különbség, hogy a szórásnégyzet csak nemnegatív lehet. Az $1-\varepsilon$ szintű konfidenciaintervallumok közül a lehető legkisebb hosszúságú $[a_\varepsilon, b_\varepsilon]$ $(0 < a_\varepsilon < b_\varepsilon)$ konfidenciaintervallumot keressük. Bizonyítás nélkül felhasználjuk, hogy a legrövidebb konfidenciaintervallumot akkor kapjuk, ha annak a valószínűsége, hogy a valódi σ^2 paraméter kisebb, mint az alsó a_ε határ, ugyanakkora, mint azé, hogy σ^2 nagyobb, mint a felső b_ε határ, vagyis pontosan $\varepsilon/2$:

$$\mathbb{P}_{\sigma^2}(0 < \sigma^2 < a_\varepsilon) = \mathbb{P}_{\sigma^2}(\sigma^2 > b_\varepsilon) = \frac{\varepsilon}{2}.$$

Ezt átírhatjuk úgy is, hogy

$$\mathbb{P}_{\sigma^2}\big(\frac{(n-1)(S_n^*)^2}{\sigma^2} > \frac{(n-1)(S_n^*)^2}{a_\varepsilon}\big) = \mathbb{P}_{\sigma^2}\big(\frac{(n-1)(S_n^*)^2}{\sigma^2} < \frac{(n-1)(S_n^*)^2}{b_\varepsilon}\big) = \frac{\varepsilon}{2}. \tag{B.1}$$

Jelölje $\alpha \in (0,1)$ esetén $\chi^2_{n-1,\alpha}$ az $n-1$ szabadságfokú χ^2-eloszlás $1-\alpha$-kvantilisét. Lukács tételének fenti részlete alapján $\frac{(n-1)(S_n^*)^2}{\sigma^2} \sim \chi^2(n-1)$. Az a szám, amelynél ez a valószínűségi változó pontosan $\varepsilon/2$ valószínűséggel nagyobb, éppen $\chi^2_{n-1,\varepsilon/2}$, az a szám pedig, amelynél ez a valószínűségi változó pont $\varepsilon/2$ valószínűséggel kisebb, vagyis pont $1-\varepsilon/2$ valószínűséggel nagyobb, éppen $\chi^2_{n-1,1-\varepsilon/2}$. Ezt felhasználva a (B.1) egyenletben szereplő egyenlőségek éppen akkor teljesülnek, ha

$$\frac{(n-1)(S_n^*)^2}{a_\varepsilon} = \chi^2_{n-1,\varepsilon/2} \qquad \text{és} \qquad \frac{(n-1)(S_n^*)^2}{b_\varepsilon} = \chi^2_{n-1,1-\varepsilon/2}.$$

Nagy örömünkre a kapott egyenletek már nem függnek az ismeretlen σ^2 paramétertől, sőt az ismeretlen μ várható értéktől sem. Átrendezés után azt kapjuk, hogy $(a_\varepsilon, b_\varepsilon)$ csak a mintától függ, így valóban egy konfidenciaintervallum:

$$[a_\varepsilon, b_\varepsilon] = \big[\frac{(n-1)(S_n^*)^2}{\chi^2_{n-1,\varepsilon/2}}, \frac{(n-1)(S_n^*)^2}{\chi^2_{n-1,1-\varepsilon/2}}\big].$$

A példa tanulságait az alábbi állítás foglalja össze.

B.2.3. Állítás. *Legyenek $X_1, \dots, X_n$ fae. $\mathrm{N}(\mu;\sigma^2)$ eloszlású valószínűségi változók, ahol a μ várható érték és a $\sigma^2 > 0$ szórásnégyzet is ismeretlen, és jelölje $(S_n^*)^2$ a mintához tartozó korrigált empirikus szórásnégyzetet. Ekkor $\varepsilon \in (0,1)$ esetén a σ^2 szórásnégyzetre egy $1-\varepsilon$ szintű konfidenciaintervallum a következő:*

$$[T_1(\mathbf{X}), T_2(\mathbf{X})] = \big[\frac{(n-1)(S_n^*)^2}{\chi^2_{n-1,\varepsilon/2}}, \frac{(n-1)(S_n^*)^2}{\chi^2_{n-1,1-\varepsilon/2}}\big], \tag{B.2}$$

ahol $\alpha \in (0,1)$ esetén $\chi^2_{n-1,\alpha}$ az $n-1$ szabadságfokú χ^2-eloszlás $1-\alpha$-kvantilise.

Mi a helyzet az ismert várható érték esetével? A fenti érvelés alapján a (B.2) képlet ismert μ esetén is egy $1-\varepsilon$ szintű konfidenciaintervallumot ad σ^2-re, azonban ez a konfidenciaintervallum nem optimális, mivel tipikusan nem ez a legrövidebb minden ilyen szintű konfidenciaintervallum közül. Ugyanis a B.1.1 Definíció és a normális eloszlás standardizálásának ismert módja szerint ha $X_1, \dots, X_n \sim \mathrm{N}(\mu;\sigma^2)$ fae. valószínűségi változók, akkor $\sum_{i=1}^n (\frac{X_i-\mu}{\sigma})^2$ pontosan n darab független standard normális eloszlású valószínűségi változó négyzetének összege, ezért $\chi^2(n)$-eloszlású. Így ismert μ esetén az előző érveléssel analóg módon új konfidenciaintervallumot kapunk σ^2-re az alábbi állítás szerint.[17]

[17] Mivel ezen konfidenciaintervallum kiszámítása a mintaelemek és a várható érték pontos értékére épül, ellentétben az (B.2) képlet szerintivel, valóban ez adja a pontosabb intervallumbecslést, ezt az állítást azonban nem bizonyítjuk, sőt nem is mondjuk ki formálisan.

B.2.4. Állítás. *Legyenek $X_1, \dots, X_n$ fae. $\mathrm{N}(\mu;\sigma^2)$ eloszlású valószínűségi változók, ahol a μ várható érték ismert és a $\sigma^2 > 0$ szórásnégyzet ismeretlen. Ekkor $\varepsilon \in (0,1)$ esetén a σ^2 szórásnégyzetre egy $1-\varepsilon$ szintű konfidenciaintervallum a következő:*

$$[T_1(\mathbf{X}), T_2(\mathbf{X})] = \Big[\frac{\sum_{i=1}^n (X_i-\mu)^2}{\chi^2_{n,\varepsilon/2}}, \frac{\sum_{i=1}^n (X_i-\mu)^2}{\chi^2_{n,1-\varepsilon/2}}\Big], \tag{B.3}$$

ahol $\alpha \in (0,1)$ esetén $\chi^2_{n,\alpha}$ az n szabadságfokú χ^2-eloszlás $1-\alpha$-kvantilise.

Az ismert és az ismeretlen várható érték esete között tehát nem csupán a szabadsági fokban van különbség, hanem abban is, hogy míg ismeretlen várható érték esetén $(n-1)(S_n^*)^2 = \sum_{i=1}^n (X_i - \overline{X_n})^2$ állt az intervallumvégpontok számlálóiban, ezek helyett itt $\sum_{i=1}^n (X_i-\mu)^2$ szerepel. Vagyis a várható érték becslését a valódi várható értékre cserélve pontosabb becslést (rövidebb konfidenciaintervallumot) kapunk.

B.3. Konfidenciaintervallum normális eloszlás várható értékére ismeretlen szórás esetén

A B.1.6. Állításban érdemes egy dolgot megfigyelni. Az ismert σ szórású normális eloszlás μ várható értékére vett konfidenciaintervallum meghatározásakor láttuk, hogy

$$\frac{\sqrt{n}(\overline{X_n}-\mu)}{\sigma} \sim \mathrm{N}(0;1)$$

(lásd (12.2)) ami elengedhetetlen volt a konfidenciaintervallum megadásához. Az állításban azt látjuk, hogy ha a σ szórást a S_n^* korrigált empirikus szórásra cseréljük, akkor $n-1$ szabadságfokú t-eloszlást kapunk:

$$\frac{\sqrt{n}(\overline{X_n}-\mu)}{S_n^*} \sim t(n-1).$$

Miért nagyszerű ez? Azért, mert az utóbbi egyenletben egyáltalán nem szerepel a σ szórás, vagyis az egyenlet *bármilyen* μ várható értékű normális eloszlás esetén igaz, a szórás értékétől függetlenül! Ezt felhasználva ha ismeretlen szórású normális eloszlás várható értékére kell $1-\varepsilon$ szintű konfidenciaintervallumot szerkesztenünk, azt a 12.2.1. Példában megadott eljárás szerint tehetjük meg, azzal a különbséggel, hogy a szórást a korrigált empirikus szórásra, a standard normális eloszlás $1-\varepsilon/2$-kvantilisét pedig az $n-1$ szabadságfokú t-eloszlás $1-\varepsilon/2$-kvantilisére kell cserélnünk. Ezzel megkapjuk a 12.2.2. Állítás analógiáját az ismeretlen szórás esetére:

B.3.1. Állítás. *Legyenek $X_1, \dots, X_n$ fae. $\mathrm{N}(\mu;\sigma^2)$ eloszlású valószínűségi változók, ahol a μ várható érték és a $\sigma^2 > 0$ szórásnégyzet ismeretlen, és jelölje $\overline{X_n}$ a mintához tartozó mintaátlagot, S_n^* pedig a mintához tartozó korrigált empirikus szórást. Ekkor $\varepsilon \in (0,1)$ esetén a μ várható értékre egy $1-\varepsilon$ szintű konfidenciaintervallum a következő:*

$$\big[T_1(\mathbf{X}), T_2(\mathbf{X})\big] = \big[\overline{X_n} - \frac{S_n^* t_{n-1,\varepsilon/2}}{\sqrt{n}}, \overline{X_n} + \frac{S_n^* t_{n-1,\varepsilon/2}}{\sqrt{n}}\big],$$

ahol $\alpha \in (0,1)$ esetén $t_{n-1,\alpha}$ az $n-1$ szabadságfokú t-eloszlás $1-\alpha$-kvantilisét jelöli.[18]

B.4. t-próba

Amint már említettük, a t-próba ismeretlen szórású normális eloszlással rendelkező fae. mintaelemek esetén használható. Ebből a próbatípusból is az egyoldali, egy- illetve kétmintás, valamint a kétoldali, kétmintás

[18] A t-eloszlás nagy szabadságfok esetén „közel van" a standard normális eloszláshoz. Precízebben fogalmazva: bármely $\alpha \in (0,1)$ esetén a t-eloszlás α-kvantilise a standard normális eloszlás α-kvantiliséhez tart, ha a szabadsági fok a végtelenhez tart. Ezért nagy mintaelemszám (kb. $n \geq 30$), azaz nagy szabadsági fok esetén szabad a t-eloszlás α-kvantilise helyett a standard normális eloszlás α-kvantilisével dolgozni. Nagy szabadsági fok esetén sem szabad viszont a korrigált empirikus szórás és a valódi szórás szerepét a konfidenciaintervallumokra, illetve a hipotézisvizsgálatok témakörénél a próbastatisztikákra vonatkozó képletekben összekeverni.

próbát ismertetjük ebben a jegyzetben. Ezek többé-kevésbé analóg módon viselkednek, mint az u-próba megfelelő változataival, azzal a különbséggel, a próbastatisztikákban a szórás helyett a korrigált empirikus szórás fog szerepelni, és a próbastatisztikát a standard normális eloszlás helyett egy bizonyos szabadságfokú t-eloszlás megfelelő kvantilisével kell összehasonlítani. Kétmintás esetben a próbastatisztika formája jelentősebb mértékben változik az u-próbához képest. Kezdjük most az egymintás esettel.

B.4.1. *Példa* (Kétoldali, egymintás t-próba). Legyen $\mathbf{X} = (X_1, \dots, X_n)$ fae. minta $\mathrm{N}(\mu; \sigma^2)$ eloszlású mintaelemekkel, ahol ezúttal mind a μ várható érték, mind a σ^2 szórásnégyzet ismeretlen. Kétoldali, egymintás esetben az alternatívánk továbbra is

$$H_0 \colon \mu = \mu_0 \qquad \text{vs.} \qquad H_1 \colon \mu \neq \mu_0.$$

A (szintén kétoldali, egymintás) ε terjedelmű u-próba alapja az volt, hogy

$$u(\mathbf{X}) = \frac{\sqrt{n}(\overline{X_n} - \mu)}{\sigma^2} \sim \mathrm{N}(0; 1),$$

és ez alapján ha $|u(\mathbf{X})| \leq u_{\varepsilon/2}$, akkor H_0-t elfogadjuk, különben elutasítjuk. Ugyanakkor a B.3. alfejezetben említettük és használtuk, hogy

$$t(\mathbf{X}) = \frac{\sqrt{n}(\overline{X_n} - \mu)}{S_n^*} \sim t(n-1),$$

ahol $S_n^* = \sqrt{(S_n^*)^2}$ a mintához tartozó korrigált empirikus szórást jelöli. Emiatt az u-próba levezetésével (lásd 13.2.1. Példa) analóg módon azt kapjuk, hogy akkor kapunk ε terjedelmű próbát, ha H_0-t akkor fogadjuk el, ha $|t(\mathbf{X})| \leq t_{n-1,\varepsilon/2}$, ahol $t_{n-1,\varepsilon/2}$ az $n-1$ szabadságfokú t-eloszlás $1-\varepsilon/2$-kvantilisét jelöli, és különben H_0-t elutasítjuk.

Hasonlóan az egyoldali, kétmintás t-próbát is az u-próba megfelelő változatával definiálhatjuk, a próbastatisztikában (amit most $t(\mathbf{X})$-szel jelölünk) a szórást a korrigált empirikus szórásra cserélve és a normális eloszlás releváns kvantilise helyett az $n-1$ szabadságfokú t-eloszlás ugyanezen kvantilisét tekintve. Írjuk le az egymintás t-próbák eljárását az u-próbákéhoz hasonló formában:

Eljárás (Egymintás t-próba)**.** *Feltevések: az $x_1, \dots, x_n$ adatpontok fae. normális eloszlású $X_1, \dots, X_n$ valószínűségi változók realizációi, amelyek szórása és μ várható értéke ismeretlen. Általános eljárás:*

1. *A H_0 nullhipotézis felállítása.*
2. *Az előirányzott $\varepsilon \in (0,1)$ terjedelem kiválasztása.*
3. *Az $x_1, \dots, x_n$ adatpontok kinyerése.*
4. *Az $f = n-1$ szabadsági fok meghatározása.*
5. *Az elfogadási tartomány meghatározása. Ez függ a próba altípusától, de mindig magába foglalja az $n-1$ szabadságfokú t-eloszlás egy megfelelő kvantilisének kiszámítását (táblázatból való kiolvasását). Ez a táblázat csak nemnegatív helyeken adja meg az $n-1$ szabadságfokú t-eloszlás eloszlásfüggvényének értékeit, a negatív helyeken vett értékek kiszámítása a normális eloszlással analóg módon történik (mivel a t-eloszlások sűrűségfüggvénye is a 0-ra szimmetrikus).*
6. *A próbastatisztika $t(\mathbf{x}) = \frac{\overline{x_n} - \mu_0}{s_n^*}\sqrt{n}$ értékének meghatározása.*
7. *Annak eldöntése, hogy a próbastatisztika értéke az elfogadási tartományba esik-e. Ha igen, H_0-t elfogadjuk. Ha nem, H_0-t elutasítjuk.*

Kétoldali, egymintás t-próba.

- *Nullhipotézis:* $H_0 \colon \mu = \mu_0$.

- *Releváns kvantilis:* $t_{n-1,\varepsilon/2}$ *az* $f = n-1$ *szabadságfokú* t*-eloszlás* $1-\varepsilon/2$*-kvantilise.*
- *Elfogadási tartomány:* $\mathfrak{X}_\mathrm{e} = \{(x_1,\dots,x_n)\colon |t(\mathbf{x})| \le t_{n-1,\varepsilon/2}\}$*, tehát:*
  - *ha* $|t(\mathbf{x})| > t_{n-1,\varepsilon/2}$*,* H_0*-t elutasítjuk,*
  - *különben* H_0*-t elfogadjuk.*

Egyoldali, egymintás t-próba ($H_0\colon \mu \ge \mu_0$ változat)**.**

- *Nullhipotézis:* $H_0\colon \mu \ge \mu_0$*.*
- *Releváns kvantilis:* $-t_{n-1,\varepsilon}$*, ahol* $t_{n-1,\varepsilon}$ *az* $n-1$ *szabadságfokú* t*-eloszlás* $1-\varepsilon$*-kvantilise.*
- *Elfogadási tartomány:* $\mathfrak{X}_\mathrm{e} = \{(x_1,\dots,x_n)\colon t(\mathbf{x}) \ge -t_{n-1,\varepsilon}\}$*, tehát:*
  - *ha* $t(\mathbf{x}) < -t_{n-1,\varepsilon}$*,* H_0*-t elutasítjuk,*
  - *különben* H_0*-t elfogadjuk.*

Egyoldali, egymintás t-próba ($H_0\colon \mu \le \mu_0$ változat)**.**

- *Nullhipotézis:* $H_0\colon \mu \le \mu_0$*.*
- *Releváns kvantilis:* $t_{n-1,\varepsilon}$ *az* $n-1$ *szabadságfokú* t*-eloszlás* $1-\varepsilon$*-kvantilise.*
- *Elfogadási tartomány:* $\mathfrak{X}_\mathrm{e} = \{(x_1,\dots,x_n)\colon t(\mathbf{x}) \le t_{n-1,\varepsilon}\}$*, tehát:*
  - *ha* $t(\mathbf{x}) > t_{n-1,\varepsilon}$*,* H_0*-t elutasítjuk,*
  - *különben* H_0*-t elfogadjuk.*

B.4.2. *Példa* (Egymintás t-próba: kenyérvásárlás.). Miután elszomorodtunk azon, hogy az u-próbáink eredménye alapján a pékünk valóban várható értékben 1 kg-nál kisebb kenyereket ad nekünk, viszont érdemben a másik péknél sem jobb a helyzet, eszünkbe jut, hogy talán csak azért kaptuk ezeket az eredményeket, mert a szórás értékéről rossz feltételezéseink voltak. Ezért elhatározzuk, hogy az adott alternatívákra elvégezzük a t-próba megfelelő változatait is. Kezdjük ebben a példában az egymintás t-próbákkal!

Szerencsére mindennap felírtuk, hogy aznap milyen nehéz kenyeret vettünk, ezért ki tudjuk számolni a korrigált empirikus szórásokat is. Az első péknél vett kenyerek tömegéből álló, n =25 elemű $(x_1,\dots,x_n)$ mintarealizáció esetén, amelyre $\overline{x_n} = 0.98$ kg teljesült, azt kapjuk, hogy a korrigált empirikus szórás $s^*_{25} = 0.05$ kg (vagyis lényegesen nagyobb a valódi szórás korábban feltételezett 0.02 értékénél).

Kétoldali, egymintás t-próba esetén a

$$H_0\colon \mu = 1 \qquad \text{vs.} \qquad H_1\colon \mu \ne 1$$

alternatívára felállított $\varepsilon = 0.05$ terjedelmű próbánál a próbastatisztika értéke

$$t(\mathbf{x}) = \frac{\sqrt{25}(0.98-1)}{0.05} = -\frac{10}{5} = -2.$$

Ennek abszolút értékét kell az $n-1 = 24$ szabadságfokú t-eloszlás $1-\varepsilon/2 = 0.975$-kvantilisével összehasonlítani. A kvantilis értéke

$$t_{24,0.025} \approx 2.0639.$$

Ez nagyobb, mint $u_{0.975} \approx 1.9600$, és valamivel nagyobb $|t(\mathbf{x})|$-nél is. Ezért H_0-t elfogadjuk, vagyis azzal a következtetéssel vigasztalhatjuk magunkat, hogy a kenyerek várható értékben mégis 1 kilogramm tömegűek.

Egyoldali, egymintás t-próba esetén a

$$H_0\colon \mu \ge 1 \qquad \text{vs.} \qquad H_1\colon \mu < 1$$

alternatívára felállított $\varepsilon = 0.05$ terjedelmű próbánál a próbastatisztika értéke megint csak $t(\mathbf{x}) = -2$. Ezt kell $-t_{24,0.05} \approx -1.7109$-cel összehasonlítani. Mivel $t(\mathbf{x}) < -t_{24,0.05}$, H_0-t a „szigorúbb" egyoldali próba alapján még mindig elutasítjuk, vagyis arra a következtetésre jutunk, hogy a kenyerek várható értékben 1 kg-nál kisebb tömegűek.

Végül térjünk rá a kétoldali, kétmintás t-próba esetére.

B.4.3. *Példa.* Legyenek $X_1, \dots, X_{n_1} \sim \mathrm{N}(\mu_1; \sigma_1^2)$ és $Y_1, \dots, Y_{n_2} \sim \mathrm{N}(\mu_2; \sigma_2^2)$ függetlenek, ahol a μ_1 és μ_2 várható értékek ismeretlenek, és tekintsük a

$$H_0 \colon \mu_1 = \mu_2 \qquad \text{vs.} \qquad H_1 \colon \mu_1 \neq \mu_2$$

alternatívát. Láttuk (bár nem bizonyítottuk), hogy

$$u(\mathbf{X}, \mathbf{Y}) = \frac{\overline{X_{n_1}} - \overline{Y_{n_2}}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0; 1).$$

Ezért ha a σ_1^2 és σ_2^2 szórásnégyzetek ismertek, akkor az $u(\mathbf{X}, \mathbf{Y})$ statisztikát használva konstruálhatunk egy ε terjedelmű próbát erre az alternatívára, ami nem más, mint a kétoldali, kétmintás u-próba.

Hogyan alkossunk ε terjedelmű próbát ugyanerre az alternatívára, ha σ_1^2 és σ_2^2 is ismeretlenek? Ebben az esetben belátható, hogy amennyiben $\sigma_1^2 = \sigma_2^2$,

$$t(\mathbf{X}, \mathbf{Y}) = \frac{\overline{X_{n_1}} - \overline{Y_{n_2}}}{\sqrt{\frac{(n_1-1)(s_X^*)^2 + (n_2-1)(s_Y^*)^2}{n_1+n_2-2}}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

t-eloszlású $n_1 + n_2 - 2$ szabadsági fokkal, ahol $(s_X^*)^2$ jelöli az $\mathbf{X}$ mintához tartozó korrigált empirikus szórásnégyzetet és $(s_Y^*)^2$ az $\mathbf{Y}$-hoz tartozót (ezt sem bizonyítjuk).[19] A kritikus tartományunk ebben az esetben tehát

$$\mathfrak{X}_\mathrm{k} = \{(\mathbf{x}, \mathbf{y}) = (x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}) \in \mathbb{R}^{n_1+n_2} \colon |t(\mathbf{x}, \mathbf{y})| > t_{n_1+n_2-2, \varepsilon/2}\},$$

ahol $t_{n_1+n_2-2,\varepsilon/2}$ az n_1+n_2-2 szabadsági fokú t-eloszlás $1-\varepsilon/2$-kvantilise (az eddigiekkel konzisztens jelölést használva). Vagyis ha $|t(\mathbf{x}, \mathbf{y})| \leq t_{n_1+n_2-2,\varepsilon/2}$, akkor H_0-t elfogadjuk, különben pedig H_0-t elutasítjuk.

Ez alapján formalizálhatjuk a kétoldali, kétmintás t-próba eljárását, amely természetesen sokban hasonlít a kétoldali, kétmintás u-próbáéra. Vigyázat: a próbastatisztika változása itt nem csupán annyi, hogy a szórásnégyzeteket a korrigált empirikus megfelelőjükre cseréljük, és a próba csak akkor alkalmazható, ha a két minta ismeretlen szórásai megegyeznek!

Eljárás (Kétoldali, kétmintás t-próba)**.** *Feltevések: az $x_1, \dots, x_{n_1}$ adatpontok fae. normális eloszású $X_1, \dots, X_{n_1}$ valószínűségi változók realizációi, amelyek szórása ismeretlen, ahogy μ_1 várható értékük is, az $y_1, \dots, y_{n_2}$ adatpontok pedig az $X_1, \dots, X_{n_1}$-től (együttesen) független, egymás között fae. normális eloszlású $Y_1, \dots, Y_{n_2}$ valószínűségi változók realizációi, ezek μ_2 várható értékük sem ismert, szórásuk pedig megegyezik az X_i valószínűségi változók (ismeretlen) szórásával.*[20] *Eljárás:*

[19] Általában igaz az a (matematikailag nem mindig egyértelmű) ökölszabály, hogy a szabadsági fok az összes mintaelem száma mínusz a becsült paraméterek száma. Esetünkben az egymintás t-próbánál egyetlen becsült paraméterünk volt: a valódi szórás helyett a minta korrigált empirikus szórásával dolgoztunk, így a szabadsági fok $n-1$ volt. A kétmintás t-próbánál már két szórást helyettesítünk a becsült értékével, ezért kapunk $n_1 + n_2 - 2$ szabadsági fokot. A normális eloszlás szórásnégyzetére vonatkozó konfidenciaintervallumok esetén is látható, hogy ismeretlen várható érték esetén eggyel kevesebb $(n-1)$ a szabadsági fok, mint ismert várható érték esetén (n), mivel az előbbi esetben a konfidenciaintervallum képletében a várható érték helyett a becslése, azaz a mintaátlag szerepel.

[20] Ebből a tárgyból nem foglalkozunk azzal az esettel, amikor $\sigma_1^2 = \sigma_2^2$ nem teljesül. A szórásnégyzetek egyenlőségét is lehet statisztikai próbával tesztelni, mégpedig a $H_0 \colon \sigma_1^2 = \sigma_2^2$ vs. $H_1 \colon \sigma_1^2 \neq \sigma_2^2$ alternatívájú *F-próba* segítségével. Ha az F-próba azt az eredményt adja, hogy $\sigma_1^2 = \sigma_2^2$, akkor alkalmazható a kétoldali, kétmintás t-próba. Ellenkező esetben ehelyett a *Welch-próbát* kell alkalmazni, amelynek próbastatisztikáját más módon kell kiszámolni. A próbastatisztika szintén t-eloszlású lesz, de a t-eloszlás szabadsági foka általában nem egész szám lesz. Ezek a próbák a mérnök- és gazdaságinformatikus MSc képzés Matematikai statisztika tárgyának anyagához tartoznak.

1. *Nullhipotézis:* $H_0\colon \mu_1 = \mu_2$ *vs.* $H_1\colon \mu_1 \neq \mu_2$.
2. *Az előirányzott* $\varepsilon \in (0,1)$ *terjedelem kiválasztása.*
3. *Az* $x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}$ *adatpontok kinyerése.*
4. *Az* $f = n_1 + n_2 - 2$ *szabadsági fok meghatározása.*
5. *Próbastatisztika:*
$$t(\mathbf{x}, \mathbf{y}) = \frac{\overline{x_{n_1}} - \overline{y_{n_2}}}{\sqrt{\frac{(n_1-1)(s_x^*)^2+(n_2-1)(s_y^*)^2}{n_1+n_2-2}}} \sqrt{\frac{n_1 n_2}{n_1+n_2}}.$$
6. *Elfogadási tartomány:* $\mathfrak{X}_\mathrm{e} = \{(\mathbf{x}, \mathbf{y}) = (x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}) \in \mathbb{R}^{n_1+n_2}\colon |t(\mathbf{x}, \mathbf{y})| \leq t_{n_1+n_2-2,\varepsilon/2}\}$, *ahol* $t_{n_1+n_2-2,\varepsilon/2}$ *az* $f = n_1 + n_2 - 2$ *szabadságfokú* t*-eloszlás* $1 - \varepsilon/2$*-kvantilise.*
7. *Döntés: ha* $|t(\mathbf{x}, \mathbf{y})| > t_{\varepsilon/2}$*, akkor* H_0*-t elutasítjuk, különben elfogadjuk.*

Zárásként alkalmazzuk a kétoldali, kétmintás t-próbát a kenyeres példára.

B.4.4. *Példa.* A kétmintás próbának megfelelő jelöléssel az eddigiek szerint az új péknél vásárolt 10 db kenyér adatai: $n_1 = 10$, $\overline{x_{n_1}} = 0.9825$ kg. Szerencsére gondosan felírtuk az új péknél vásárolt kenyerek tömegadatait is, ezek korrigált empirikus szórására pedig $s_x^* = 0.015$ kg adódott. A régi péknél vásárolt 25 kenyérhez tartozó adatok az eddigiek szerint: $n_2 = 25$, $\overline{y_{n_2}} = 0.98$ kg, $s_y^* = 0.05$ kg. Így a szabadsági fok $f = n_1 + n_2 - 2 = 33$. A
$$H_0\colon \mu_1 = \mu_2 \text{ vs. } H_1\colon \mu_1 \neq \mu_2$$
alternatívára – ismeretlen szórások mellett – a kétoldali, kétmintás t-próbát alkalmazzuk. A próbastatisztika értéke:
$$t(\mathbf{x}, \mathbf{y}) = \frac{\overline{x_{n_1}} - \overline{y_{n_2}}}{\sqrt{\frac{(n_1-1)(s_x^*)^2+(n_2-1)(s_y^*)^2}{n_1+n_2-2}}} \sqrt{\frac{n_1 n_2}{n_1+n_2}} = \frac{0.9825 - 0.98}{\sqrt{\frac{9\cdot 0.015^2+24\cdot 0.05^2}{33}}} \sqrt{\frac{10\cdot 25}{10+25}} \approx 0.1541.$$
Ez (abszolút értékben) jóval kisebb, mint az $n_1 + n_2 - 2 = 33$ szabadságfokú t-eloszlás $1 - \varepsilon/2 = 0.975$-kvantilise, $t_{33,0.025} \approx 2.0345$. Ezért H_0-t elfogadjuk, azaz (ismét) arra a következtetésre jutunk, hogy a régi és az új péknél vásárolt kenyér várható értéke megegyezik.[21]

Felhasznált irodalom

A jegyzetkiegészítés anyaga (a Mészáros Szabolcs-féle jegyzeten túl) Noemi Kurt *Stochastik für Informatiker* című tankönyvére (Springer, 2020) és Bolla Marianna (BME-TTK) 2013. tavaszi félévi *Matematikai statisztika* című előadásának anyagára épül. A (B.3) képletben szereplő konfidenciaintervallum Zakhar Kabluchko *Skript zur Vorlesung zur mathematischen Statistik* című előadásjegyzetében[22] található meg.

[21] A kapott próba p-értéke rendkívül alacsony, mert a próbastatisztika $t(\mathbf{x}, \mathbf{y})$ értéke közel van a 0-hoz.
[22] `https://www.uni-muenster.de/imperia/md/content/Stochastik/skript_statistik.pdf`, utolsó frissítés: 2018.10.23.

C. Konfidenciaintervallumokhoz és hipotézisvizsgálatokhoz képletek

$s_n^2 = \frac{1}{n}\sum_{i=1}^n x_i^2 - \overline{x_n}^2$, $\quad s_n = \sqrt{s_n^2}$, $\quad (s_n^*)^2 = \frac{n}{n-1}s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x_n})^2$, $\quad s_n^* = \sqrt{(s_n^*)^2}$.

u-próba

[23]

1. Kétoldali, egymintás: $u = \frac{\overline{x_n}-\mu_0}{\sigma}\sqrt{n}$, $u_{\varepsilon/2} = \Phi^{-1}(1-\varepsilon/2)$.
   Alternatíva: $H_0\colon \mu = \mu_0$ vs. $H_1\colon \mu \neq \mu_0$, kritikus tartománya:

   $$\mathfrak{X}_\mathrm{k} = \{(x_1,\dots,x_n) \in \mathbb{R}^n \colon |u(x_1,\dots,x_n)| > u_{\varepsilon/2}\}.$$

   Konfidenciaintervallum μ-re ismert σ esetén:

   $$\big[\overline{x_n} - \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}, \overline{x_n} + \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}\big].$$

2. Egyoldali, egymintás: $u = \frac{\overline{x_n}-\mu_0}{\sigma}\sqrt{n}$, $u_\varepsilon = \Phi^{-1}(1-\varepsilon)$.
   (a) Alternatíva: $H_0\colon \mu \leq \mu_0$ vs. $H_1\colon \mu > \mu_0$, kritikus tartomány:

   $$\mathfrak{X}_\mathrm{k} = \{(x_1,\dots,x_n) \in \mathbb{R}^n \colon u(x_1,\dots,x_n) > u_\varepsilon\}.$$

   (b) Alternatíva: $H_0\colon \mu \geq \mu_0$ vs. $H_1\colon \mu < \mu_0$, kritikus tartomány:

   $$\mathfrak{X}_\mathrm{k} = \{(x_1,\dots,x_n) \in \mathbb{R}^n \colon u(x_1,\dots,x_n) < -u_\varepsilon\}.$$

3. Kétoldali, kétmintás: $u = \frac{\overline{x_{n_1}}-\overline{y_{n_2}}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$, $u_{\varepsilon/2} = \Phi^{-1}(1-\varepsilon/2)$.
   Alternatíva: $H_0\colon \mu_1 = \mu_2$ vs. $H_1\colon \mu_1 \neq \mu_2$, kritikus tartománya:

   $$\mathfrak{X}_\mathrm{k} = \{(x_1,\dots,x_{n_1},y_1,\dots,y_{n_2}) \in \mathbb{R}^{n_1+n_2} \colon |u(x_1,\dots,x_{n_1},,y_1,\dots,y_{n_2})| > u_{\varepsilon/2}\}.$$

[23] Itt ε a próba terjedelmét, illetve konfidenciaintervallum esetén $1-\varepsilon$ a konfidenciaintervallum szintjét jelöli. (Ez nem lesz megadva a vizsgán.)