VALÓSZÍNŰSÉGSZÁMÍTÁS

ELŐADÁSJEGYZET, 2021 ŐSZ

Bevezető

Jelen iromány a BME-VIK Mérnökinformatikus BSc képzésén 2020 őszén elhangzott valószínűségszámítás kurzushoz tartozó előadásjegyzet. Előismeretként nem feltételezünk többet, mint a szak Analízis 1 kurzusának tematikájában[1] szereplő fogalmak. A jegyzet főképp a kurzus korábbi évfolyamain használt jegyzetekből építkezik[2]:

[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 [3]

Tartalomjegyzék

Bevezető 1

1. Alapfogalmak 3

1.1. Eseménytér 3

1.2. Eseményalgebra 5

1.3. Valószínűségi mező 6

1.4. Poincaré-formula 7

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 13

3.3. Randomized Quicksort algoritmus (kiegészítő anyag) 16

3.4. Várható érték végtelen diszkrét esetben 17

4. Folytonos valószínűségi változók 18

4.1. Eloszlásfüggvény 18

4.2. Sűrűségfüggvény 20

Csehi Csongor jegyzete alapján készítette Mészáros Szabolcs. Közreműködők: Balázs Barbara, Papp László, Szabó Dániel, Takács Balázs, Tóth Dávid. Utolsó frissítés: 2021. szeptember 3..

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

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

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

4.3. Várható érték, folytonos eset ... 21
5. Nevezetes eloszlások ... 23
5.1. Bertrand-paradoxon ... 23
5.2. Örökifjú tulajdonság ... 24
5.3. Poisson-eloszlás ... 26
6. Valószínűségi változók viszonya ... 28
6.1. Függetlenség ... 28
6.2. Diszkrét együttes eloszlás ... 29
6.3. Kovariancia ... 30
6.4. Variancia és szórás ... 31
6.5. Korreláció ... 33
7. Folytonos együttes eloszlás és konvolúció ... 34
7.1. Valószínűségi vektorváltozók ... 34
7.2. Vektorváltozók függetlensége ... 36
7.3. Konvolúció ... 37
8. Normális eloszlás ... 39
8.1. Az eloszlás definíciója ... 39
8.2. Standardizálás ... 40
8.3. De Moivre–Laplace-tétel ... 42
8.4. Kitekintés: heurisztika a de Moivre–Laplace tételhez ... 42
9. Határeloszlás-tételek ... 44
9.1. Csebisev-egyenlőtlenség ... 44
9.2. Nagy számok törvénye ... 45
9.3. Centrális határeloszlás-tétel ... 47
10. Lineáris regresszió ... 50
10.1. Szórás és kovariancia folytonos esetben ... 50
10.2. Lineáris regresszió ... 53
11. Feltételes várható érték ... 56
11.1. Feltételes várható érték, diszkrét regresszió ... 56
11.2. Folytonos regresszió ... 58
11.3. Regresszió tulajdonságai, teljes várható érték tétele ... 60
12. Feltételes valószínűség és többdimenziós eloszlások ... 62
12.1. Teljes valószínűség tétele, folytonos eset ... 62
12.2. Többdimenziós eloszlások ... 63
12.3. Többdimenziós normális eloszlás ... 64

1. Alapfogalmak

A valószínűségszámítás praktikusságát talán nem kell bizonygatni egyetlen olvasónak sem[4]: 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[5], 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**[6] 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 1.3 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}.

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

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

[6] 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[7]. 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.

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.

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.

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

1.2. Eseményalgebra

Szeretnénk beszélni az összes eseményt tartalmazó halmazról is: ezt a halmazt **eseményalgebrá**nak hívjuk. Az eseményalgebra már nem Ω részhalmaza, hanem Ω részhalmazainak halmaza, vagyis egy $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ részhalmaz. Itt $\mathcal{P}(\Omega)$ az Ω ún. hatványhalmazát jelöli, azaz $\mathcal{P}(\Omega)$ elemei épp Ω részhalmazai.

1.2.1. Példa. Legyen $\Omega = \{1,2,3,4,5,6\}$ és $\mathcal{F}$ elemei pontosan az $\{1,2,3\}$, a $\{4,5,6\}$ halmaz, a lehetetlen esemény $\emptyset$ és a biztos esemény Ω. Ekkor nem minden kimenetelekből álló részhalmaz esemény. Mégis előfordulhat olyan feladat, ahol ez az $\mathcal{F}$ modellezi jól a problémát (vö. fenti 2. megjegyzés).

Felmerülhet kérdésként, hogy két esemény uniója (ill. metszete, különbsége) szintén esemény-e. A válasz: igen. Egész pontosan azt fogjuk megkövetelni az $\mathcal{F}$ eseményalgebrától, hogy úgynevezett σ-algebra (ejtsd: szigma-algebra) legyen, vagyis teljesítse a következőket.

1.2.2. Definíció. Legyen Ω tetszőleges halmaz, $\mathcal{F}$ pedig az Ω részhalmazainak egy halmaza. Ekkor $\mathcal{F}$-et **σ-algebrá**nak nevezzük az Ω alaphalmazon, ha az alábbi három feltétel mindegyike teljesül:

(1) $\Omega \in \mathcal{F}$,
(2) ha $A \in \mathcal{F}$, akkor $\overline{A} \in \mathcal{F}$,
(3) ha $A_1, A_2, ..., A_i, ... \in \mathcal{F}$, akkor $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Röviden, $\mathcal{F}$ pontosan akkor σ-algebra, ha eleme a teljes tér, zárt a komplementer-képzésre és a megszámlálható unióra.

A definíció valószínűségszámítási szempontból a következőt mondja. Ha $\mathcal{F}$-re úgy gondolunk, mint a megfigyelhető események halmazára, akkor a feltételek szerint meg kell tudjuk figyelni azt, ami biztosan bekövetkezik (első feltétel), és azt is, ha egy A esemény nem történik meg (második feltétel). A harmadik kicsit trükkösebb, azt modellezi, hogy ha események egy sorozatát külön-külön meg tudjuk figyelni, akkor azt is, hogy legalább az egyikük bekövetkezik-e.

A σ-algebra fogalma a mértékelmélet témaköréből származik. A mértékelmélet az analízis azon ága, amely a különböző terület- és térfogatfogalmak általánosításait vizsgálja. Ennek a témának az eredményeit használta fel Kolmogorov a valószínűségszámítás megalapozására.

A σ-algebrák több tulajdonsága adódik a definícióból:

1.2.3. Állítás. *Legyen $\mathcal{F}$ σ-algebra az Ω alaphalmazon. Ekkor teljesülnek a következők:*

- $\emptyset \in \mathcal{F}$,
- *ha* $A, B \in \mathcal{F}$, *akkor* $A \cup B, A \cap B, A \backslash B \in \mathcal{F}$,
- *ha* $A_1, A_2, ..., A_i, ... \in \mathcal{F}$, *akkor* $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

Bizonyítás. Az (1) tulajdonság miatt $\Omega \in \mathcal{F}$, így (2) miatt $\emptyset = \overline{\Omega} \in \mathcal{F}$.

A második pont bizonyításához legyen $A_1 = A$, $A_2 = B$, és minden $i \geq 3$ esetén $A_i = \emptyset$. Ekkor (3) miatt

$$A \cup B = \bigcup_{i=1}^{\infty} A_i \in \mathcal{F},$$

vagyis az unióra valóban zárt $\mathcal{F}$. A metszetre való zártság ebből már levezethető: ha $A, B \in \mathcal{F}$, akkor (2) miatt $\overline{A}, \overline{B} \in \mathcal{F}$. Továbbá, már beláttuk, hogy unió-képzésre zárt az $\mathcal{F}$, tehát $\overline{A} \cup \overline{B} \in \mathcal{F}$. Viszont a két halmazra vonatkozó De Morgan-azonosság okán tudjuk, hogy $\overline{A} \cup \overline{B} = \overline{A \cap B}$, tehát $\overline{A \cap B} \in \mathcal{F}$. Innen ismét a (2) tulajdonságot használva következik, hogy $A \cap B \in \mathcal{F}$.

A harmadik pont bizonyításához vegyük észre, hogy $A_i \in \mathcal{F}$ miatt $\overline{A_i} \in \mathcal{F}$ is teljesül (2) miatt. Ezen új halmazsorozatra alkalmazhatjuk a (3) tulajdonságot, így

$$\bigcup_{i=1}^{\infty} \overline{A_i} \in \mathcal{F}.$$

Viszont a végtelen halmazokra vonatkozó De Morgan-azonosság miatt ez éppen $\bigcap_{i=1}^{\infty} A_i$ komplementere. Mivel $\mathcal{F}$ zárt a komplementerképzésre, így az állítást ezzel beláttuk. □

1.3. 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 $\mathcal{F}$ egy σ-algebra. De ennél többet is tudnia kell.

1.3.1. Definíció. Legyen $\mathcal{F}$ egy σ-algebra az Ω tetszőleges halmazon. 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.

1.3.2. Definíció. Egy $(\Omega, \mathcal{F}, \mathbb{P})$ hármast (Kolmogorov-féle) **valószínűségi mező**nek hívunk, ha $\mathcal{F}$ σ-algebra az Ω halmazon és $\mathbb{P}$ valószínűségi mérték.

1.3.3. 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.

1.3.4. 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.

1.3.5. 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

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

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. □

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

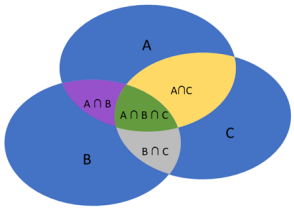
(1) $\mathbb{P}(A) + \mathbb{P}(\overline{A}) = 1$,
(2) $\mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A)$,
(3) ha $B \subseteq A$, akkor $\mathbb{P}(B) \leq \mathbb{P}(A)$.

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?

1.4. Poincaré-formula

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



1.4.1. 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 Poincaré-formula.

A tétel rövid kimondásához további jelölésre van szükségünk. Jelölje $[n]$ az $\{1, 2, \dots, n\}$ halmazt. Legyenek $A_1, A_2, \dots, A_n \in \mathcal{F}$ események, továbbá legyen $I = \{i_1, i_2, \dots, i_k\}$ 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

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

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

1.4.2. Tétel (Poincaré-formula)**.** *Legyenek $A_1, A_2, \dots, A_n \in \mathcal{F}$ 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.

1.4.3. 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.

1.4.4. 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. □

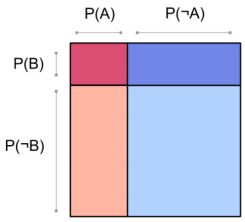
Az állítás a Bonferroni-egyenlőtlenség nélkül, teljes indukcióval is könnyen belátható.

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.[8] 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.

[8] 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 $\mathbb{P}(A) > 0$.[9]

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”.[10]

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*

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

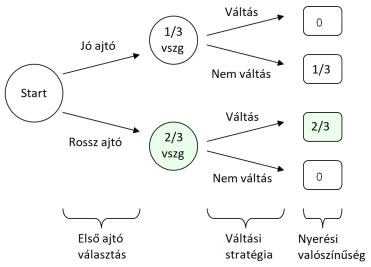
2.2.5. Definíció. Egy $A_1, \dots, A_n \in \mathcal{F}$ páronként kizáró eseményekből álló sorozatot **teljes eseményrendszer**nek hívunk, ha $\cup_{i=1}^n A_i = \Omega$.

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. □

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

[10] 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*

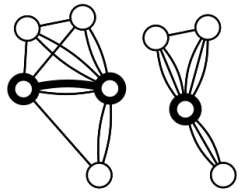
$$\mathbb{P}\Big(\bigcap_{i=1}^{n} A_i\Big) = \mathbb{P}(A_1) \cdot \prod_{i=2}^{n} \mathbb{P}\Big(A_i \,\Big|\, \bigcap_{k=1}^{i-1} A_k\Big).$$

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 $B = \{\text{pozitív a teszt}\}$.[11]

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

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

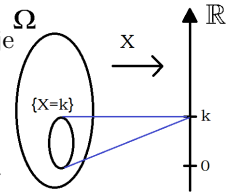
3. Diszkrét valószínűségi változók

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.

3.2. 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**:

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

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

(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.[12] 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
$$\mathbb{E}(X+Y)=\sum_{m\in M}m\cdot\mathbb{P}(X+Y=m)=\sum_{m\in M}m\cdot\mathbb{P}\Big(\bigcup_{\substack{k\in K,\,l\in L\\ k+l=m}}\{X=k,\,Y=l\}\Big)=$$
$$=\sum_{m\in M}\sum_{\substack{k\in K,\,l\in L\\ k+l=m}}(k+l)\cdot\mathbb{P}(X=k,\,Y=l)=\sum_{k\in K}\sum_{l\in L}(k+l)\cdot\mathbb{P}(X=k,\,Y=l)$$

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

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

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 Poincaré-formulát, 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:

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

$$= 1-\mathbb{E}\Big(1-\mathbf{1}_{A_1}-\mathbf{1}_{A_2}-\mathbf{1}_{A_3}+\mathbf{1}_{A_1\cap A_2}+\mathbf{1}_{A_2\cap A_3}+\mathbf{1}_{A_1\cap A_3}-\mathbf{1}_{A_1\cap A_2\cap A_3}\Big)$$

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

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 Poincaré-formula.

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

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

Jelölése: $X\sim B(n;p)$.

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 $A_1,\dots,A_n$ együttesen független események. Tegyük fel, hogy $\mathbb{P}(A_i)=p$ minden i-re. Ekkor

$$\mathbf{1}_{A_1}+\dots+\mathbf{1}_{A_n}\sim B(n;p),$$

vagyis a $B(n;p)$ eloszlású valószínűségi változóra mindig nézhetünk úgy, mint n darab együttesen független esemény indikátorainak összegére.

Ennek a megfigyelésnek a hasznossága rögtön látható, ugyanis ha $X\sim B(n;p)$, akkor

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

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 $S=\{1,2,\dots,n\}$, akkor X várható értéke $\mathbb{E}(X) = \frac{1+2+\dots+n}{n} = \frac{n+1}{2}$.

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[13], 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 $\mathbb{E}X_n \leq -4n + (2n+2)(\ln n + 1) \leq 2n \ln n$. □

[13] Lásd a Karger-algoritmus utáni megjegyzést.

3.4. 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.

3.4.1. Definíció. Legyen X egy kizárólag nemnegatív értékeket felvevő valószínűségi változó. Definiáljuk ekkor a várható értékét:

$$\mathbb{E}(X) \stackrel{\text{def}}{=} \sup_{\substack{Z \text{ egyszerű,} \\ Z \le X}} \mathbb{E}(Z),$$

ami vagy egy nemnegatív valós szám vagy $+\infty$. Vagyis az X-nél (minden $\omega \in \Omega$ ponton) nem nagyobb valószínűségi változók várható értékeinek a „lehető legnagyobb értéke”, a szuprémuma.

Hát ez nem tűnik túl egyszerűen számolhatónak. A kiszámolásban a következő állítás segít úgynevezett diszkrét valószínűségi változók esetében.

3.4.2. Állítás. *Legyen X olyan nemnegatív valószínűségi változó*[14]*, aminek értékkészlete* $\mathrm{Ran}(X) = \{k_1, k_2, \dots\}$ *megszámlálhatóan végtelen. Ekkor*

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

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 az 5. előadásban, – az eredmény $\frac{1}{p}$.

3.4.3. Definíció. Egy valószínűségi változó **diszkrét**, ha értékkészlete megszámlálható (nem feltétlenül végtelen).

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[15]. 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.

[14] Nem feltétlenül nemnegatív valószínűségi változó esetén a várható érték ugyanezzel a formulával definiálható, amennyiben a sor abszolút konvergens.

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

4. 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.

4.1. 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.

4.1.1. 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.

4.1.2. Definíció. Legyen X valószínűségi változó. Ekkor az

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

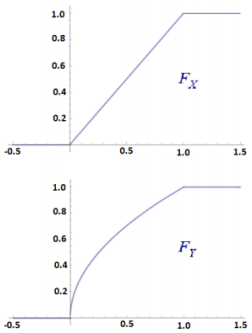
függvényt az X **eloszlásfüggvény**ének hívjuk.

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

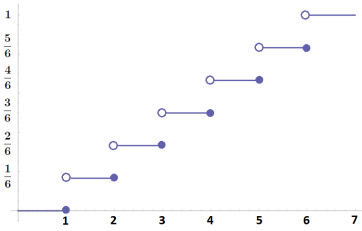
$$F_X(b) - F_X(a) = \mathbb{P}(X < b) - \mathbb{P}(X < a) = \mathbb{P}(a \le X < b)$$

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

4.1.3. 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 1.3 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 $\bigcup_{k=1}^{\infty} \{x_{k-1} \leq X < x_k\} = \{x_0 \leq X < x\}$. Ezt visszahelyettesítve az előző egyenletbe azt kapjuk, hogy $\lim_{n\to\infty} F_X(x_n) = \mathbb{P}(X < x_0) + \mathbb{P}(x_0 \leq X < x) = \mathbb{P}(X < x) = F_X(x)$.[16] Hasonló okból teljesül a harmadik feltétel is.[17]

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.[18] Definiáljuk az X-et, mint

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

Ekkor az X és az infimum definíciója miatt

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

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. □

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

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

[18] 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.

4.2. 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.

4.2.1. 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[19], 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?

4.2.2. 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.

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

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

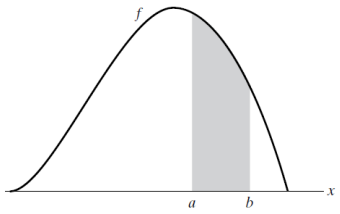
függvények sűrűségfüggvényei X-nek és X^2-nek. Szemléletesen a sűrűségfüggvény értékei annak az esélyét jelölik, hogy az X változó a x érték kis környezetébe esik[20]. Így az alfejezet elején feltett második kérdés válasza $f_{X^2}\left(\frac{1}{4}\right) / f_{X^2}\left(\frac{3}{4}\right) = \frac{1}{1} / \frac{1}{\sqrt{3}} = \sqrt{3}$.

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

[20] 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.



4.2.4. 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.

⚠ **4.2.5. 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?

4.3. Várható érték, folytonos eset

Előző előadáson definiáltuk nemnegatív valószínűségi változók várható értékét:

$$\mathbb{E}(X) \overset{\text{def}}{=} \sup_{\substack{Z \text{ egyszerű,} \\ Z \le X}} \mathbb{E}(Z), \qquad \text{ahol} \qquad \mathbb{E}(Z) = \sum_{k \in \mathrm{Ran}(X)} k \cdot \mathbb{P}(Z = k).$$

Vegyük észre, hogy az első definíció nem csak diszkrét esetre értelmes, folytonos valószínűségi változókra is ad valamit, csak nehezen látszik, hogy mit. Zavaró viszont benne a nemnegativitási feltétel. Ahhoz, hogy megszabaduljunk ettől a feltételtől, felhasználjuk, hogy a várható érték a fenti definícióval is additív, ahogy egyszerű valószínűségi változókra ezt már beláttuk.

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

Ezen azonosság segítségével definiálható nem feltétlenül nemnegatív (és nem is feltétlenül egyszerű) valószínűségi változókra is a várható érték.

4.3.2. 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 $X = X^+ - X^-$. 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.

Bár ezek a definíciók értelmesek folytonos valószínűségi változókra is, nem konstruktívak, konkrét valószínűségi változó várható értékét nehéz így meghatározni. A következő állítás ezt hidalja át.

4.3.3. Állítás. *Legyen X folytonos valószínűségi változó, amire*

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

Ekkor

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

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

4.3.4. Definíció. Egy X valószínűségi változó **egyenletes eloszlású** 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 (1) egyenlet és a folytonos esetre vonatkozó (2) 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.

4.3.5. 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*

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

4.3.6. 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

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

5. Nevezetes eloszlások

A most következő előadásban először a geometriai valószínűségek illetve folytonos valószínűségi változók egy paradoxonjáról lesz szó, majd különböző eloszlásokkal foglalkozunk, vegyesen diszkrét és folytonos kontextusban, ezzel is alkalmazva az előző két előadáson tanultakat.

5.1. Bertrand-paradoxon

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[21] rávilágít, hogy mennyire nem nyilvánvaló feladat az utóbbi esetben meghatározni a valószínűségi változó eloszlását.

Bertrand-paradoxon

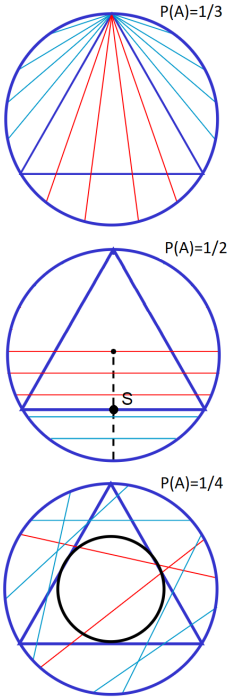
Válasszuk ki egy kör egy húrját véletlenszerűen. Mi a valószínűsége, hogy a húr hosszabb, mint a körbe írható szabályos háromszög egy oldala?

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.

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

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.

5.2. Örökifjú tulajdonság

A példákban többször előforduló, nevezetes eloszlások a (diszkrét) geometriai eloszlás, illetve a (folytonos) exponenciális eloszlás. A két eloszlás közeli rokonságban van, ezért egyszerre tárgyaljuk őket.

5.2.1. 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: $X\sim\mathrm{Geo}(p)$.

5.2.2. 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.

Geometriai eloszlású valószínűségi változó várható értéke:[22]

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

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

5.2.3. 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:

$$\mathbb{E}(Z)=\int_{-\infty}^{\infty}xf_X(x)\mathrm{d}x=\int_{0}^{\infty}\lambda xe^{-\lambda x}\mathrm{d}x=$$

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

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.

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.

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

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

$$\mathbb{P}(X > t + s \mid X > s) = \mathbb{P}(X > t), \tag{3}$$

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?

5.2.5. Állítás. *Legyen X nemkonstans örökifjú valószínűségi változó a G halmazon.*

(1) Ha $G = \{1, 2, 3, \dots\}$, akkor X eloszlása geometriai.
(2) Ha $G = [0, \infty)$, akkor X eloszlása exponenciális.

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) 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.

Legyen most G a nemnegatív valós számok halmaza, és jelöljük $g(t)$-vel az $\ln \mathbb{P}(X > t)$ mennyiséget[23]. Ez a logaritmus értelmes (azaz $\mathbb{P}(X > t) \neq 0$), hacsak X nem konstans 0, amit viszont kizártunk. Ekkor a (3) 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.

5.2.6. 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 $\lambda = -g(1) \geq 0$.*[24]

[23] Itt ln jelöli a természetes alapú logaritmust.

[24] 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.[25]

5.2.7. 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. □

5.3. 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[26]. Ilyen például egy telefonközpontba beérkező hívások száma, vagy az egy adott órában született gyerekek száma.

5.3.1. 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 nem teljesül sem a geometriai, sem az exponenciális eloszlásra).

---

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

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

5.3.2. 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 $B(n;p)$ 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 $B(n;p)$ eloszlása körülbelül Pois(λ).

5.3.3. 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.

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. □

5.3.4. 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: $\mathbb{P}(X=0) = \frac{(2\sqrt{3})^0}{0!}e^{-2\sqrt{3}} = e^{-2\sqrt{3}} \approx 0{,}0313$.

6. 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, folytonos esetben pedig az F_X eloszlásfüggvényt vagy az f_X sűrűségfüggvényt nézni. Az eloszlás minden lényeges tulajdonságot elmondott a valószínűségi változóról.

De ne keverjük össze az eloszlást magával a valószínűségi változóval: attól, hogy 100 dobásból mind a dobott fejek száma, mind a dobott írások száma $B(100; \frac{1}{2})$ 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.

6.1. 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.

6.1.1. 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”.

6.1.2. 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.

6.1.3. 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. Általánosan: nézhetjük az X által **generált σ-algebrát**, azaz a legkisebb olyan – $\sigma(X)$-el jelölt – halmazt, aminek elemei az $\{X < x\}$ események ($x \in \mathbb{R}$), és teljesíti a σ-algebra definícióját. Ekkor X és Y függetlensége ekvivalens azzal, hogy bármilyen $A \in \sigma(X)$ és $B \in \sigma(Y)$ események függetlenek.

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.

6.1.4. 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

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

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

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 nemnegatív valós f és g függvények esetén $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$, akkor X és Y függetlenek.

6.2. 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 eloszlására van szükségünk. (A folytonos esettel a 8. előadáson fogunk foglalkozni.)

6.2.1. 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 táblázattal megadott együttes eloszlás pontosan akkor lehet két valószínűségi változó együttes eloszlása, 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 eloszlására.

6.2.2. Definíció. Legyenek X és Y egyszerű valószínűségi változók. Ha adott X és Y **együttes eloszlás**a, 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 eloszlásait az együttes eloszlás **marginális eloszlás**ainak nevezzük.

A marginális eloszlásokat 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 eloszlásait. 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.

6.3. 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,[27] kezdjük a kovariancia fogalmával.

6.3.1. 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.

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

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. □

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

(1) Ha Y *konstans, akkor* $\mathrm{cov}(X,Y)=0$.
(2) Ha X *és* Y *függetlenek, akkor* $\mathrm{cov}(X,Y)=0$.
(3) Attól, hogy $\mathrm{cov}(X,Y)=0$*, még nem feltétlenül teljesül, hogy* X *és* Y *független.*

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

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

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

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

A harmadik állításhoz legyen $\mathrm{Ran}(X) = \{-1, 0, 1\}$, amely értékeket $\frac{1}{4}$, $\frac{1}{2}$ és $\frac{1}{4}$ valószínűségekkel veszi fel X. Legyen $Y = |X|$. Kiszámolható, hogy $\mathrm{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0 - 0 \cdot \frac{1}{2} = 0$, pedig a változók nem függetlenek, hiszen $\mathbb{P}(X = 0)\mathbb{P}(Y = 1) = \frac{1}{2} \cdot \frac{1}{2}$, míg $\mathbb{P}(X = 0, Y = 1) = 0$. □

6.3.4. Példa.

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

(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

$$\mathrm{cov}(X,\, 0{,}9 \cdot X + 0{,}1 \cdot Y) = \mathbb{E}(0{,}9 \cdot X^2 + 0{,}1 \cdot XY) - \mathbb{E}(X)\mathbb{E}(0{,}9 \cdot X + 0{,}1 \cdot Y) =$$

$$= 0{,}9 \cdot \mathbb{E}(X^2) + 0{,}1 \cdot \mathbb{E}(XY) - 0{,}9 \cdot \mathbb{E}(X)^2 - 0{,}1 \cdot \mathbb{E}(XY)$$

$$= 0{,}9 \sum_{k=1}^{10} k^2 \frac{1}{10} - 0{,}9 \cdot \Big(\frac{11}{2}\Big)^2 \approx 7{,}425.$$

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

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

ha a fenti kovarianciák léteznek. Így a kovariancia a vektorok skaláris szorzatának rokonfogalma.

6.4. Variancia és szórás

Speciális eset a kovariancia számolásában, amikor $Y = X$.

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

$$\mathrm{cov}(X, X) = \mathbb{E}\big((X - \mathbb{E}X)^2\big) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

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

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

Nyilván ez a mennyiség nem X-nek a saját magával való összefüggőségéről szolgáltat információt, hanem arról, 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.

6.4.2. 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:

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

$$= \mathbb{E}(X^2) + \mathbb{E}(Y^2) + 2\mathbb{E}(XY) - \mathbb{E}(X)^2 - \mathbb{E}(Y)^2 - 2\mathbb{E}(X)\mathbb{E}(Y)$$

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

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

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

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

További elemi tulajdonságai a szórásnégyzetnek:

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

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

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 adódik az állítás. □

6.4.4. 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

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

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

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

(4) Legyen $T \sim \mathrm{Geo}(p)$. Ekkor

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

ahol az első szumma kiszámolható például hatványsorok deriválásával, ettől itt eltekintünk.

(5) Legyen $Y \sim \mathrm{Pois}(\lambda)$. Ekkor

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

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

(6) A szórás definíciója akkor is értelmes, ha a valószínűségi változó például folytonos. Legyen $Z \sim \mathrm{Exp}(\lambda)$ valamilyen λ pozitív valósra. Kiszámolható (és a 10. előadáson ki is számoljuk), hogy ekkor $\mathbb{D}(Z) = \frac{1}{\lambda}$.

6.5. 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.

6.5.1. Definíció. Legyenek X és Y valószínűségi változók. Ha $\operatorname{cov}(X,Y)$, $\mathbb{D}(X)$ és $\mathbb{D}(Y)$ értelmes, akkor X és Y **korreláció**ja:

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

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

6.5.2. 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.*

6.5.3. 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 $\operatorname{corr}(X, X^2) = 0$.

7. Folytonos együttes eloszlás és konvolúció

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.

A fejezet végén 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 harmadik alfejezetben először kiderül, hogy ez miért mégis, majd hogy miért mégsem olyan nehéz.

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.[28] 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 $F_{(X,Y)}(x,y) = 0$.

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

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

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.[29]

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 eloszlásról. 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:

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

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

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

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 megegyzett 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.

7.3. 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.

7.3.1. Példa.

(1) Legyen $X \sim B(n; p)$ 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.[30]

(2) Legyenek X és Y független, 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.

7.3.2. Definíció. Legyenek X és Y független valószínűségi változók. Ekkor $X + Y$ eloszlását az X és Y eloszlásai **konvolúció**jának hívjuk.

[30] Aki úgy gondolná, hogy a binomiális eloszlás nem túl összetett, az megpróbálhatja meggyőzni magát az előző fejezetben szereplő de Moivre–Laplace-tétel segítségével.

7.3.3. 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:

$$\mathbb{P}(X+Y=k)=\mathbb{P}\Big(\bigcup_{i+j=k}\{X=i\}\cap\{Y=j\}\Big)=$$

$$=\sum_{i=0}^{k}\mathbb{P}\Big(\{X=i\}\cap\{Y=k-i\}\Big)=\sum_{i=0}^{k}\mathbb{P}(X=i)\mathbb{P}(Y=k-i).$$

□

7.3.4. Példa. Legyenek X és Y független valószínűségi változók, amire $X\sim\mathrm{Pois}(\lambda)$ és $Y\sim\mathrm{Pois}(\mu)$. Mi $X+Y$ eloszlása? 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.

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.

7.3.5. 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.

7.3.6. Példa. Legyenek X és Y független, $\mathrm{Exp}(\lambda)$ eloszlású valószínűségi változók. Az előző állítás szerint

$$f_{X+Y}(z)=\int_{-\infty}^{\infty}f_X(x)f_Y(z-x)\mathrm{d}x=$$

$$=\int_0^z\lambda e^{-\lambda x}\lambda e^{-\lambda(z-x)}\mathrm{d}x=\lambda^2e^{-\lambda z}\int_0^z1\mathrm{d}x=\lambda^2e^{-\lambda z}z$$

minden $z>0$ esetén. Az így kapott eloszlás neve gamma-eloszlás.

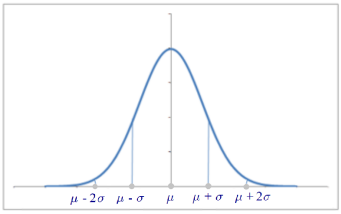
8. Normális eloszlás

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.[31]



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[32]

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

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.[33] □

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

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

[33] 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*

$$f_Y(x) = \frac{1}{\sigma} f_X\Big(\frac{x-\mu}{\sigma}\Big) \tag{4}$$

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 (4) 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 $B(n;p)$ binomiális eloszlású, és emiatt az értékei a $\{0,1,2,\dots,n\}$ 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.[34] 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:

$$\mathbb{P}(Y > 0) = 1 - \mathbb{P}(Y \le 0) = 1 - \mathbb{P}(Y < 0) = 1 - \mathbb{P}\Big(\frac{Y+2}{1{,}3} < \frac{2}{1{,}3}\Big) = 1 - \Phi\Big(\frac{2}{1{,}3}\Big),$$

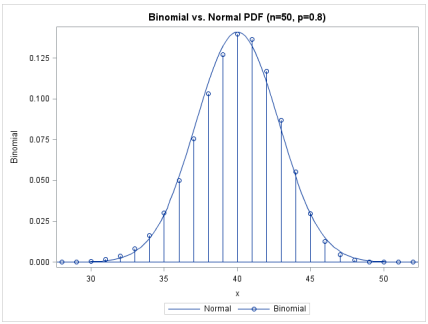
ahol Φ értéke kiszámoltatható megfelelő szoftver segítségével, de egyszerűen ki is kereshető a Φ értékeit tartalmazó táblázatból. Így kapjuk, hogy $\mathbb{P}(Y > 0) \approx 0{,}0620 \approx 6\%$.

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 $x \in (-1,1)$.

[34] Bővebben lásd: math.stanford.edu/∼conrad/papers/elemint.pdf

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 súlya, vagy a napi átlaghőmérséklet az év egy adott hónapjában, stb. A közös ezekben, hogy az eredményét rengeteg apró, egymástól többé-kevésbé független faktor befolyásolja.



A következő tétel azt állítja, hogy nagy n esetén a binomiális eloszlás közelíthető a normális eloszlással.[35]

8.3.1. Tétel. *Legyen $p \in (0,1)$ és $S_n \sim B(n;p)$. 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.)

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.[36]

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 $np \to \lambda$.

8.4. Kitekintés: heurisztika a de Moivre–Laplace tételhez

A tételt nem bizonyítjuk, de körbejárjuk a bal oldalán szereplő 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.)

---

[35] 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

[36] Lásd még: Berry-Esseen-típusú egyenlőtlenségek, pl. arxiv.org/abs/1111.6554

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.[37]

[37] Karakterizációk egy gyűjteményét lásd: [Cross Validated] What is the most surprising characterization of the Gaussian (normal) distribution?

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.[38] 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.

Először 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.

9.1. Csebisev-egyenlőtlenség

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.

9.1.1. 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. 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:

9.1.2. Következmény (Csebisev-egyenlőtlenség[39])**.** *Legyen X valószínűségi változó, amire $\mathbb{D}^2(X)$ véges. Ekkor minden $a > 0$ esetén*

$$\mathbb{P}\big(|X - \mathbb{E}(X)| \geq a\big) \leq \frac{\mathbb{D}^2(X)}{a^2}.$$

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.

---

[38] Szemfüles olvasó felfedezheti, hogy a geometriai eloszlásról is megemlítettük, hogy a pontok sűrítése esetén az exponenciális eloszláshoz tart.

[39] Angol szakirodalomban: Chebyshev's inequality

9.1.3. Példa. Egy adott adatbázis átlagosan 50 lekérést fogad egy időegység alatt. A lekérések számának szórása a tapasztalatok szerint 5.[40] 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.

9.1.4. Következmény (Paraméteres Csernov-egyenlőtlenség[41])**.** *Legyen X valószínűségi változó. Ekkor minden $a, t > 0$ esetén*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(e^{tX})}{e^{ta}}.$$

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. □

9.1.5. Példa. Legyen $X \sim \text{Pois}(5)$. 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.

A fejezet további részében a fenti ötleteket alkalmazva jutunk el a téma két legfontosabb tételéhez.

9.2. 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 $A_1, \ldots, A_n$ 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,

$$\lim_{n \to \infty} \mathbb{P}(A_1 \cup \cdots \cup A_n) \to 1.$$

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

[40] 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.

[41] Angol szakirodalomban: Chernoff bound

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 $\frac{X_1+\ldots+X_n}{n}$ 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.

9.2.1. Tétel. *Legyen $X_1, X_2, \ldots, X_n, \ldots$ 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 $\frac{X_1+\ldots+X_n}{n}$ átlagot.*[42]

- ***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**.

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 $X_1, \ldots, X_n$ függetlenek, így a $\mathbb{D}^2$ tulajdonságai szerint:

$$\mathbb{D}^2(\overline{X_n}) = \mathbb{D}^2\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2}\big(\mathbb{D}^2(X_1) + \cdots + \mathbb{D}^2(X_n)\big) = \frac{n}{n^2}\sigma^2,$$

ahol a jobb oldal 0-hoz tart $n \to \infty$ esetén. Ebből az állítás már következik. □

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.

[42] A tételek a szórás végessége nélkül is teljesülnek, de a bizonyításuk komplikáltabb.

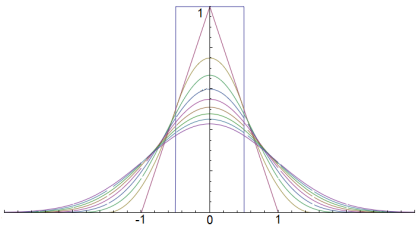
Az erős törvény éppen azt mondja, hogy ez lehetetlen. Vagyis 1 valószínűséggel olyanok az ω-k, hogy egy idő után már nem megyünk μ-től ε-nál távolabb.

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.

9.3. 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 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.



9.3.1. Definíció. Legyen $X_1, X_2, \dots, X_n, \dots$ 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.[43] Jelölése: $X_n \xrightarrow{d} Z$.

9.3.2. 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:

$$\mathbb{P}\left(\frac{X_1 + \dots + X_n - n\mathbb{E}(X_1)}{\sqrt{n}\mathbb{D}(X_1)} < a\right) \to \Phi(a)$$

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.[44]

---

[43] 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.

[44] Lásd Berry-Esseen tétel, [Terence Tao] Central limit theorem, Theorem 23

Megjegyzés. A tételben szereplő szumma más formában is írható:

$$\frac{X_1+\cdots+X_n-n\mathbb{E}(X_1)}{\sqrt{n}\mathbb{D}(X_1)}=\frac{\overline{X_n}-\mathbb{E}(\overline{X_n})}{\mathbb{D}(\overline{X_n})},$$

ami átrendezéssel és a szórás tulajdonságainak felhasználásával kapható. Más szavakkal, a tételben az $\overline{X_n}$ standardizáltjáról van szó.

9.3.3. 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 $X_1, X_2, \ldots, X_{50}$. 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 tétel bizonyításához az első alfejezetből meríthetünk ötletet, egész pontosan a Csernov-egyenlőtlenségben megjelenő mennyiségből: az $\mathbb{E}(e^{tX})$ várható értékből.

9.3.4. Definíció. Az X valószínűségi változó **momentumgeneráló függvény**e:[45]

$$M_X(t)\stackrel{\text{def}}{=}\mathbb{E}(e^{tX}),$$

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.

9.3.5. Állítás. *Legyenek Y, Z illetve $X_1, X_2, \ldots$ 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:

$$\frac{X_1+\cdots+X_n-n\mathbb{E}(X_1)}{\sqrt{n}\mathbb{D}(X_1)}=\sqrt{n}\frac{\sum_{i=1}^n X_i-\mathbb{E}(X_1)}{n\mathbb{D}(X_1)}=\sqrt{n}\frac{\sum_{i=1}^n\frac{X_i-\mathbb{E}(X_1)}{\mathbb{D}(X_1)}}{n}.$$

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

$$\lim_{n\to\infty}M_{\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}(t)=M_Z(t) \tag{5}$$

[45] 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 (5) 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 (5) egyenlet bal oldalának vizsgálata előtt számoljuk ki az X_1 momentumgeneráló függvényét is:

$$M_{X_1}(t) = \mathbb{E}\Big(\sum_{k=0}^{\infty} \frac{1}{k!} t^k X_1^k\Big) = 1 + t\mathbb{E}X_1 + \frac{t^2}{2}\mathbb{E}(X_1^2) + \frac{t^2}{2} \cdot \mathbb{E}\Big(\sum_{k=3}^{\infty} \frac{2}{k!} t^{k-2} X_1^k\Big) = 1 + \frac{t^2}{2}\big(1 + r(t)\big), \tag{6}$$

ahol $r(t)$ jelöli az utolsó szumma várható értékét. Egyáltalán nem nyilvánvaló, de belátható a következő:

9.3.6. Lemma. $\lim_{t\to 0} r(t) = 0$.

A lemma miatt az (5) egyenlet bal oldaláról a következő mondható:

$$\begin{aligned} M_{\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}(t) &= \mathbb{E}\Big(e^{t\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}\Big) = \mathbb{E}\Big(\prod_{i=1}^n e^{\frac{t}{\sqrt{n}} X_i}\Big) = \\ &= \prod_{i=1}^n \mathbb{E}\Big(e^{\frac{t}{\sqrt{n}} X_i}\Big) = \prod_{i=1}^n M_{X_i}\Big(\frac{t}{\sqrt{n}}\Big) = M_{X_1}\Big(\frac{t}{\sqrt{n}}\Big)^n, \end{aligned} \tag{7}$$

ahol felhasználtuk a következő két tulajdonságot: Egyrészt az $e^{\frac{t}{\sqrt{n}} X_i}$ valószínűségi változók függetlenek, másrészt $M_{X_i}\Big(\frac{t}{\sqrt{n}}\Big)$ minden i esetén megegyezik $M_{X_1}\Big(\frac{t}{\sqrt{n}}\Big)$-nel, hiszen a momentumgeneráló függvény csak X_i eloszlásától függ, ami minden i esetén ugyanaz, mint X_1 eloszlása.

Helyettesítsük be az (7) egyenlet logaritmusába az $M_{X_1}\Big(\frac{t}{\sqrt{n}}\Big)$ (6) egyenlet szerinti értékét:

$$\ln M_{\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}(t) = n \ln M_{X_1}\Big(\frac{t}{\sqrt{n}}\Big) = n \ln\left(1 + \frac{\Big(\frac{t}{\sqrt{n}}\Big)^2}{2}\Big(1 + r\Big(\frac{t}{\sqrt{n}}\Big)\Big)\right).$$

Jelölje $r\Big(\frac{t}{\sqrt{n}}\Big)$-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. □

10. Lineáris regresszió

Valószínűségi változók kovarianciáját eddig csak diszkrét esetben vizsgáltuk, annak ellenére, hogy ugyanaz a definíció alkalmas folytonos valószínűségi változók kovarianciájának definiálására is. Amiért ezt a témát mégis eddig halogattuk, az az együttes sűrűségfüggvény fogalmának hiánya volt, amely fogalom lehetővé teszi a kovariancia kiszámolását folytonos esetben is.

A kovariancia és szórás fogalmak alkalmazásaként a lineáris regressziót is itt tárgyaljuk. 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.

Ugyanakkor a lineáris regressziónak van egy tisztán valószínűségszámítási vonatkozása is, amihez nincs szükség a statisztika alapfogalmaira, mint a minta vagy a becslés. Ez annak a kérdésnek a környékre, hogy hogyan lehet adott X és Y valószínűségi változók esetén olyan α, β számokat választani, hogy $\beta X + \alpha$ a lehető legközelebb legyen Y-hoz.

10.1. Szórás és kovariancia folytonos esetben

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

Korábban vizsgáltuk már az X várható értékét, sőt $g(X)$ transzformált várható értékét is, ahol $g : \mathbb{R} \to \mathbb{R}$ tetszőleges folytonos függvény. Emiatt az X szórásnégyzetét is ki tudjuk számolni (ahogy azt a normális eloszlás esetében már számoltuk is):

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

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.[46]

10.1.1. Példa. Legyen $Z \sim \mathrm{Exp}(\lambda)$ valamilyen λ pozitív valósra. Ekkor két parciális integrálással adódik, hogy

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

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

Tehát

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

Ezen gondolatmeneten továbbhaladva 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:

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

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 $\mathbb{E}(XY)$.

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

10.1.2. 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).

10.1.3. 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

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

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

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

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

10.1.4. 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.

10.1.5. 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.

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.

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

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

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

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

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

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

(2) $\operatorname{cov}(\underline{X})$ *pozitív szemidefinit mátrix, azaz* $\sum_{i=1}^n \sum_{j=1}^n a_i \operatorname{cov}(X_i, X_j) a_j \ge 0$ *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

$$\mathbb{D}^2\Big(\sum_{i=1}^n a_i X_i\Big) = \operatorname{cov}\Big(\sum_{i=1}^n a_i X_i, \sum_{j=1}^n a_j X_j\Big) = \sum_{i=1}^n \sum_{j=1}^n a_i \operatorname{cov}(X_i, X_j) a_j. \tag{8}$$

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 (8) egyenlet jobb oldala is mindig nemnegatív. □

10.1.8. 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

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

10.2. 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,[47] de a legalapvetőbb, az ún. **legkisebb négyzetek módszere**.

10.2.1. 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

$$\mathbb{E}\Big(\big(Y - (\beta X + \alpha)\big)^2\Big) \tag{9}$$

mennyiség minimális.

Ennek az optimalizálási problémának a megoldása lényegében mindig létezik és egyértelmű:

10.2.2. 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 (9) egyenletben szereplő várható érték pontosan akkor minimális, ha*

$$\beta = \frac{\operatorname{cov}(X, Y)}{\mathbb{D}^2(X)} \qquad \text{és} \qquad \alpha = \mathbb{E}(Y) - \frac{\operatorname{cov}(X, Y)}{\mathbb{D}^2(X)}\mathbb{E}(X).$$

10.2.3. 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:

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

[47] 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. □

10.2.4. 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.)

10.2.5. Állítás. *Legyen az Y valószínűségi változó X-re vett lineáris regressziója $\beta X + \alpha$. Ekkor az eltérés szórásnégyzete:*

$$\mathbb{D}^2\Big(Y - (\beta X + \alpha)\Big) = \mathbb{D}^2(Y) - \frac{\operatorname{cov}(X,Y)^2}{\mathbb{D}^2(X)}.$$

Bizonyítás. A szórásnégyzet fentebb felsorolt tulajdonságai és $\beta = \frac{\operatorname{cov}(X,Y)}{\mathbb{D}^2(X)}$ miatt:

$$\mathbb{D}^2\Big(Y - (\beta X + \alpha)\Big) = \mathbb{D}^2(Y - \beta X) = \mathbb{D}^2(Y) + \beta^2\mathbb{D}^2(X) - 2\operatorname{cov}(Y, \beta X) =$$

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

Éppen ez volt az állítás. □

Megjegyzés. Másképpen felírva:

$$\mathbb{D}^2\Big(Y - (\beta X + \alpha)\Big) = \mathbb{D}^2(Y) \cdot \big(1 - \operatorname{corr}(X,Y)^2\big).$$

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 6.5 alfejezet állítása.

10.2.6. 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.

11. Feltételes várható érték

Az előző fejezetben vizsgált lineáris regresszió egyik hátulütője, hogy csak a változók közötti lineáris összefüggést fogja meg, mélyebb relációkat nem. Példán érzékeltetve, hiába lehet X értékéből tökéletesen meghatározni X^2 értékét, az X^2 lineáris regressziója X-re nem feltétlenül lesz jó közelítése X^2-nek. Természetes tehát a kérdés, nem lehetne ezt jobban csinálni valahogy? Persze a válasz igenlő, amiben a feltételes várható érték fogalma, illetve az ebből adódó regressziós függvény lesz segítségünkre.

11.1. Feltételes várható érték, diszkrét regresszió

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.

11.1.1. 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.

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

$$\mathbb{E}(Y \mid A) = \sum_{k \in \mathrm{Ran}(Y)} k \cdot \mathbb{P}(Y = k \mid A). \tag{10}$$

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.

11.1.3. Állítás. *Legyen Y valószínűségi változó, és $\mathbb{P}(A) > 0$. Ekkor $\mathbb{E}(Y \mid A) = \frac{1}{\mathbb{P}(A)}\mathbb{E}(Y\mathbf{1}_A)$.*[48]

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.

11.1.4. 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.

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

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).

11.1.5. Definíció. Legyenek X és Y valószínűségi változók, ahol X diszkrét. 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 a

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

valós függvényt. Ekkor az Y-nak az X-re vett (diszkrét) **regressziój**a a $g(X)$ valószínűségi változó. Jelölése: $\mathbb{E}(Y \mid X)$.

11.1.6. 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 regressziója 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 $x \in \{1, 2, 3, 4, 5, 6\}$.

Az elnevezés motivációja a lineáris regresszió elnevezéséből kiindulva érthető meg: Y lineáris regressziója X-re egy olyan $\beta X + \alpha$ alakú lineáris függvénye az X valószínűségi változónak, ami a „legjobb lineáris közelítést" adja Y-ra. Ezt általánosítva, az Y változó X-re vett regressziója nem szorítkozik lineáris függvényekre, hanem azt a $g(X)$ függvényét adja vissza az a X valószínűségi változónak, amely a „legjobb közelítést" adja Y-ra. Ez a „legjobb" g függvény történetesen a fenti $x \mapsto \mathbb{E}(Y \mid X = x)$, hiszen mi lehetne jobb közelítés, mint az Y átlagos értéke abban az esetben, amikor X értékéről tudjuk, hogy x.

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.[49] Ezt precízen úgy tudjuk átfogalmazni, hogy ha további, X-re vonatkozó feltétel esetén nézzük Y feltételes várható értékét, akkor az már kiszámolható $\mathbb{E}(Y \mid X)$ regresszióból is, nem kell hozzá az eredeti Y-t ismernünk. Még precízebben:

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

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

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

Bizonyítás. A fenti (10) egyenletből, és a feltételes valószínűség definíciójából adódóan

$$\mathbb{E}\big(g(X) \mid X < x\big) \stackrel{(10)}{=} \sum_{k \in \mathrm{Ran}(X)} g(k) \cdot \mathbb{P}(X = k \mid X < x) =$$

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

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. □

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

11.2. Folytonos regresszió

A regresszió 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 regressziót általánosan az előző alfejezet utolsó állításával definiálhatjuk.

11.2.1. Definíció. Legyenek X és Y valószínűségi változók. Az Y-nak az X-re vett **regresszió**ja az a $g(X)$ alakú[50] valószínűségi változó, amire

$$\mathbb{E}\big(g(X) \mid X < x\big) = \mathbb{E}(Y \mid X < x) \tag{11}$$

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). A $g(X)$ valószínűségi változó jelölése: $\mathbb{E}(Y \mid X)$.

Mivel a diszkrét regresszió is regresszió (lásd előző alfejezet utolsó állítása), így a diszkrét jelzőt elhagyjuk, és a továbbiakban minden regresszió alatt ezt a definíciót értjük. Emellett itt jegyezzük meg, hogy az $\mathbb{E}(Y \mid X)$ regresszió, mint valószínűségi változó, nem egyértelmű (de azért majdnem). Ugyanúgy, ahogy a sűrűségfüggvény sem egyértelmű, ugyanis g értékét néhány olyan ponton büntetlenül megváltoztathatjuk, amiket 0 valószínűséggel vesz fel X. Egy ilyen változtatás $g(X)$ értékét is megváltoztatja, de a definícióban szereplő egyenlet érvényben marad.

A g függvény[51] 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 $g(x)$ az Y átlagos legjobb közelítése az $\{X = x\}$ feltétel esetén.

Másrészt a jelölés abban az értelemben sem precíz, hogy g egyáltalán nem egyértelmű, így rögzített x-re a függvénynek nincs jóldefiniált értéke. Például ha X nemnegatív, akkor g a negatív félegyenesen akárhogy megválasztható. Mivel g kiszámolása tipikusan egy köztes lépés az $\mathbb{E}(Y \mid X)$ valószínűségi változó kiszámolásához, amely valószínűségi változó már lényegében egyértelmű (lásd előző bekezdés), így g 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 regresszió. De hogyan lehet kiszámolni akár a g regressziós függvényt, akár a $g(X) = \mathbb{E}(Y \mid X)$ regressziót? Ha X egyszerű, akkor ez az előző alfejezet (10) 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.

11.2.2. 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 $f_X(x) = 0$.

[50] A g nem akármilyen függvény, hiszen $g(X)$ valószínűségi változó kell legyen. Ehhez általában azt követelik meg, hogy g úgynevezett Borel-mérhető függvény legyen; ekvivalensen, folytonos függvények pontonkénti határértéke.

[51] A g értelmezési tartományát nem specifikáltuk, nem is igazán tudjuk, hiszen g nem egyértelmű. Ami lényeges feltétel a g értelmezési tartományára, hogy tartalmazza azon $x \in \mathbb{R}$ pontokat, aminek bármilyen kis környezetébe pozitív eséllyel esik X, legfeljebb egy 0 mértékű halmaz kivételével.

A definíció megjegyzésében segíthet, ha észrevesszük a hasonlóságát a $\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$ egyenlőséggel.

11.2.3. Állítás. *Legyen (X, Y) folytonos valószínűségi vektorváltozó. 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 g az $z \mapsto \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid z)\mathrm{d}y$ függvényt ($z \in \mathbb{R}$). Azt kell leellenőriznünk, hogy g-re teljesül a (11) egyenlet.

Legyen $x \in \mathbb{R}$ olyan, amire $\mathbb{P}(X < x) > 0$. A fejezet első állításа miatt

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

ahol

$$\mathbb{E}\big(g(X)\mathbf{1}_{\{X<x\}}\big) = \int_{-\infty}^{\infty} g(z)\mathbf{1}_{\{z<x\}} f_X(z)\mathrm{d}z = \int_{-\infty}^{x} g(z) f_X(z)\mathrm{d}z.$$

Ha $f_X(z) > 0$, akkor g definíciója szerint

$$g(z)f_X(z) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid z)\mathrm{d}y \cdot f_X(z) = \int_{-\infty}^{\infty} y \cdot \frac{f_{X,Y}(z,y)}{f_X(z)} f_X(z)\mathrm{d}y = \int_{-\infty}^{\infty} y \cdot f_{X,Y}(z,y)\mathrm{d}y.$$

Ha $f_X(z) = 0$, akkor $g(z)f_X(z) = 0$. 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,

$$\mathbb{E}\big(g(X) \mid X < x\big) = \frac{1}{\mathbb{P}(X < x)} \int_{-\infty}^{x} \int_{-\infty}^{\infty} y \cdot f_{X,Y}(z,y)\mathrm{d}y\mathrm{d}z.$$

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. □

11.2.4. Példa. Legyen az (X, Y) valószínűségi vektorváltozó együttes sűrűségfüggvénye $15x^2y$, ha $0 < x < y < 1$, és 0 egyébként. Határozzuk meg az $\mathbb{E}(Y \mid X)$ regressziót.
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 $\mathbb{E}(Y \mid X)$ regresszió a $\frac{2}{3} \cdot \frac{1-X^3}{1-X^2}$ valószínűségi változó.[52]

[52] Érdekes utánaszámolni, hogy a regresszió 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.

11.3. Regresszió tulajdonságai, teljes várható érték tétele

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

11.3.1. Állítás. *Legyenek X, Y, Z valószínűségi változók. 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 h folytonos[53] függvény esetén $\mathbb{E}\big(h(X)Y \mid X\big) = h(X)\mathbb{E}(Y \mid X)$.
(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.

Az első alfejezetben párhuzamot vontunk a regresszió és a lineáris regresszió közt: mindkét módszer jó közelítést keres X függvényében Y-ra. A lineáris regresszió esetében pontosan meg is fogalmaztuk, hogy mi az az optimalizálási probléma, amit a lineáris regresszió megold (sőt, ez volt a definíció). Hasonlóan, a regresszió is felírható optimalizálási probléma megoldásaként.[54]

11.3.2. 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 $\mathbb{E}(Y \mid X)$ 1-valószínűséggel megegyeznek.

Röviden, itt nem X lineáris függvényével próbáljuk minimalizálni az Y-tól való átlagos négyzetes eltérést, hanem ennél általánosabb függvények segítségével. Ilyen értelemben a regresszió a lineáris regresszió javítása.

Azt gondolhatnánk, hogy a regresszió szinte sosem egyezik meg a lineáris regresszió esetével, csak teljesen elfajuló esetekben. A következő állítás mutatja, hogy ez nem igaz.

11.3.3. Állítás. *Legyenek $Z_1, \dots, Z_n$ független, normális eloszlású valószínűségi változók, $X = \sum_{i=1}^n a_i Z_i$ és $Y = \sum_{i=1}^n b_i Z_i$ valamilyen $a_1, \dots, a_n, b_1, \dots, b_n$ valós számokra. Ekkor $\mathbb{E}(Y \mid X)$ megegyezik az Y változó X-re vett lineáris regressziójával.*

Hogyan tudjuk hasznosítani a regressziót olyan problémában, ahol a probléma kérdésfelvetésében nem szerepel feltételes várható érték? Úgy, hogy a szokásos várható érték is számolható regresszió 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.

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

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.

---

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

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

11.3.5. 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*

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

11.3.6. 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.

11.3.7. 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*

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

Folytonos esetben pedig a következőképp írható.

11.3.8. 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.

12. Feltételes valószínűség és többdimenziós eloszlások

A mostani fejezetben kimondjuk a teljes valószínűség tételének azon verzióját, ahol a feltételben valószínűségi változó szerepel teljes eseményrendszer helyett. Ettől független témaként közelebbről megvizsgálunk néhány valószínűségi vektorváltozóhoz tartozó (ún. többdimenziós) nevezetes eloszlást, különös tekintettel a többdimenziós normális eloszlásra.

12.1. 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.

12.1.1. 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 11.2 alfejezet definíciója szerint értjük, vagyis ez az a g függvény, amire $g(X) = \mathbb{E}(\mathbf{1}_A \mid X)$. Az $\mathbb{E}(\mathbf{1}_A \mid X)$ regresszió pedig a (11) egyenlettel definiált. (Mivel $\mathbb{E}(\mathbf{1}_A)$ azaz $\mathbb{P}(A)$ véges, így g létezik a 11.2 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:

12.1.2. 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.

12.1.3. 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:

$$F_{Y|X}(y \mid x) = \mathbb{P}(Y < y \mid X = x).$$

12.2. Többdimenziós eloszlások

Legyen $\underline{X} = (X_1, \dots, X_m)$ valószínűségi vektorváltozó. Az egydimenziós esethez hasonlóan beszélhetünk az $\underline{X}$ eloszlásáról (amit például az együttes eloszlásfüggvény ír le), ahogy ezt tettük is már az együttes eloszlás témakörénél. Nézzünk most néhány gyakrabban előkerülő többdimenziós eloszlást.

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 $X_i \sim B(n; p_i)$ 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.

12.2.1. 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.

12.2.2. 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 $B(n; p)$ (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 $B(n; p_i)$ 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.

Egy másik érdekes többdimenziós eloszlás:

12.2.3. 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.[55]

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:

$$\mathbb{P}(\underline{X} > t \cdot \underline{1} + \underline{s} \mid \underline{X} > \underline{s}) = \mathbb{P}(\underline{X} > t \cdot \underline{1}), \tag{12}$$

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.

[55] 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.

12.2.4. 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 (12) 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.)

12.3. 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:

$$f_{X,Y}(x,y)=h(x^2+y^2) \tag{13}$$

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:

$$f_{X,Y}(x,y)=f_X(x)\cdot f_Y(y)\qquad(\forall x,y\in\mathbb{R}). \tag{14}$$

Megmutatjuk, hogy ezek a feltételek meghatározzák az eloszlást.

12.3.1. 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 (13) és (14) 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.

12.3.2. 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?

12.3.3. 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. Vizsgáljuk meg a többdimenziós normális eloszlás hasonló paramétereit.

12.3.4. Definíció. Egy $\underline{Y} = (Y_1, \dots, Y_n)$ valószínűségi vektorváltozó **várható érték vektora** az $(\mathbb{E}Y_1, \dots, \mathbb{E}Y_n)$ $\mathbb{R}^n$-beli vektor. Jelölés $\mathbb{E}\underline{Y}$.

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

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

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

12.3.5. 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.

12.3.6. Állítás. *Legyen $\underline{Y}$ nemelfajuló n-dimenziós normális eloszlású vektorváltozó. Jelölje a várható érték 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 $\underline{\mu}$ várható érték 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:

12.3.7. 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 $Y_i \sim N\big(\mu_i, \Sigma_{i,i}\big)$.*

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.

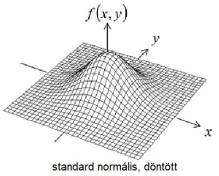
A normális eloszlás több egyéb tulajdonsága okán is a „túl szép, hogy igaz legyen” díjas eloszlás első számú jelöltje; ezeket a tulajdonságokat a következő állításban foglaljuk össze:

12.3.8. 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

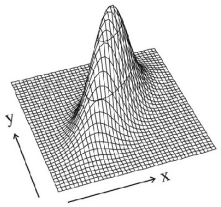
$$\mathbb{E}(Y_2 \mid Y_1) = \frac{b}{a}Y_1 + \Big(\mu_2 - \frac{b}{a}\mu_1\Big), \qquad \text{ahol} \quad \underline{\mu} = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \quad \underline{\underline{\Sigma}} = \begin{pmatrix}a & b\\ b & c\end{pmatrix}.$$

Az eloszlás vizualizációjáról még érdemes szót ejteni: hogyan is néz ki egy normális eloszlás sűrűségfüggvénye, például kétdimenziós esetben?



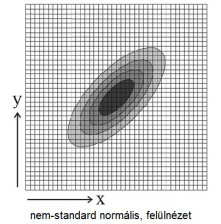
standard normális, döntött

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 $\det\big(\underline{\underline{\Sigma}}\big)^{\frac{1}{2}}$ 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 $\operatorname{Tr}\big(\underline{\underline{\Sigma}}\big)$ nyoma is használatos mennyiség. A diagonalizált változó szórásaival kifejezve $\operatorname{Tr}\big(\underline{\underline{\Sigma}}\big) = \sigma_1^2 + \sigma_2^2$. 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.