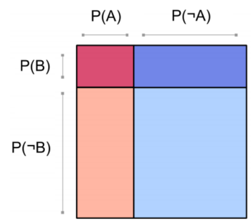
2. A Valószínűség tulajdonságai

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 $A_1, \ldots, A_n$ események **(együttesen) függetlenek**, 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 $\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdot \cdots \cdot \mathbb{P}(A_n)$? 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

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

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 bekövetkezett-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 $B_1, B_2, \ldots$ 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 $A_1, \ldots, A_n \in \mathcal{F}$ páronként kizáró események, amikre $\cup_{i=1}^n A_i = \Omega$ és $\mathbb{P}(A_i) > 0$ minden i-re. Ekkor*

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

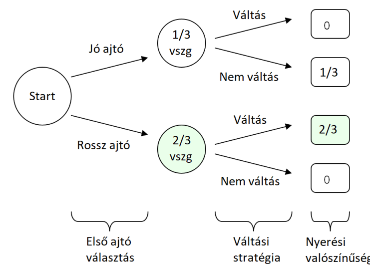
2.2.5. Definíció. Egy $A_1, \ldots, A_n \in \mathcal{F}$ páronként kizáró eseményekből álló sorozatot **teljes eseményrendszer**nek hívunk, ha $\cup_{i=1}^n A_i = \Omega$.

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*

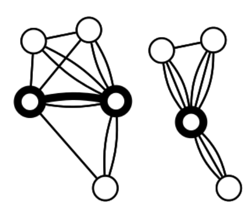
$$\mathbb{P}\Big(\bigcap_{i=1}^{n} A_i\Big) = \mathbb{P}(A_1) \cdot \prod_{i=2}^{n} \mathbb{P}\Big(A_i \Big| \bigcap_{k=1}^{i-1} A_k\Big).$$

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

$$\mathbb{P}(\text{siker}) \geq \left(1-\frac{2}{n}\right)\left(1-\frac{2}{n-1}\right)\dots\left(1-\frac{2}{n-(n-3)}\right) =$$

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

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.