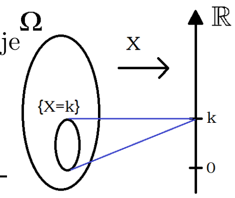
3. Diszkrét valószínűségi változók

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**:

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

ahol $\operatorname{Ran}(X)$ 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 $x_1, \ldots, x_n$ 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 $y_1, \ldots, y_n$ 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

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

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

Nézzük az $y_i, y_{i+1}, \ldots, y_{j-1}, y_j$ 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:

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

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