Valószínűségszámítás

2020. szeptember 30.

Mészáros Szabolcs

Tárgyhonlap:
cs.bme.hu/valszam

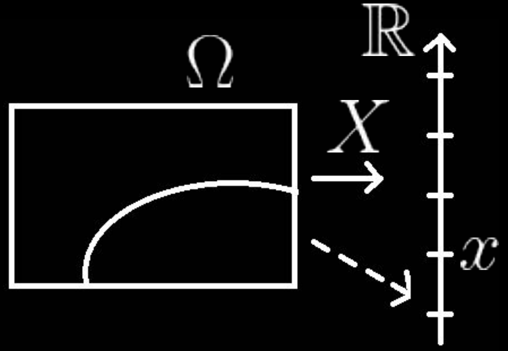
Véletlen mennyiség, példák

Valószínűségi mező	**Véletlen mennyiség**
két kockadobás	• második dobás értéke • két dobás értékének összege • dobott egyesek száma
pakli kártya véletlen fele	• kapott királyok száma • egyszínű lapok maximális száma
múlt heti mail-ek, és feladóik	• összes bejövő • spam mailek száma - Neptuntól jövő mailek száma

Valószínűségi változó

Definíció: Egy $X : \Omega \to \mathbb{R}$ függvényt *valószínűségi változó*nak hívunk, ha minden $x \in \mathbb{R}$ esetén a következő halmaz esemény:

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



Q1: Hogyan kell érteni ezt a feltételt? Miért van rá szükség?

Q2: És miért nem “kisebb-egyenlő”, vagy “nagyobb” van a definícióban?

Val. változó, Q&A

Q1: Hogyan kell érteni ezt a feltételt?

Miért van rá szükség? $\{X < x\}$ esemény

A1: Esemény = aminek értelmes a valószínűsége. Tehát azt tesszük fel, hogy

a $\mathbb{P}(\{X < x\})$ valószínűség értelmes szám.

Q2: És miért nem "kisebb-egyenlő", vagy "nagyobb" van a definícióban?

A2: Ekvivalens definíciót eredményez. $\overline{\{X < x\}} = \{X \geq x\}$

$$\bigcap_{n=1}^{\infty} \left\{X < x + \frac{1}{n}\right\} = \{X \leq x\}$$

Val. változó, példák

1. Egy (szabályos) dobókockával dobunk.
   Jelölje X a dobott szám 4-es maradékát.

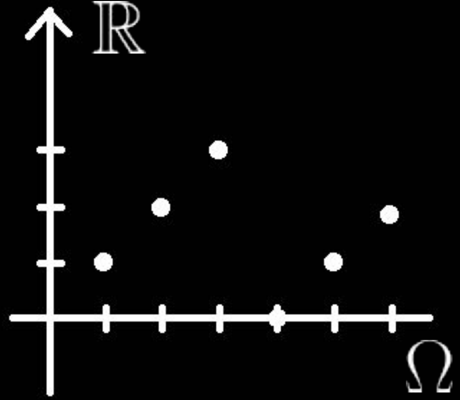
   A valószínűségi mező a szokásos:

$$\Omega = \{1, 2, 3, 4, 5, 6\} \qquad \mathcal{F} = \{\text{összes részhalmaz}\}$$

$$\mathbb{P}(A) = \text{kedvező}/\text{összes}$$

A valószínűségi változó pedig a következő függvény:

ω	1	2	3	4	5	6
$X(\omega)$	1	2	3	0	1	2



Val. változó, példák

2. példa: két pénzérmével dobunk.

X: 1, ha egyformát dobunk, 0 egyébként.

Elnevezés: indikátor valószínűségi változó.

Jel.: $\mathbf{1}_A$ ahol $A = \{\text{egyformát dobunk}\}$

3. példa: megpróbálunk beledobni egy papírgalacsint a szemeteskosárba.

X: a papír távolsága a cél középpontjától

$$\{\text{siker}\} = \{X < 0{,}2\}$$

Val. változó, műveletek

Mit tudunk tenni a valószínűségi változókkal?

Legyen X és Y valószínűségi változó. Ekkor a következők is val. változók:

$3{,}6 \cdot X$ $\quad$ $c \cdot X$ $\quad$ $(c \in \mathbb{R})$ $\quad$ $X + 42$

$X + Y$ $\quad$ $X - Y$ $\quad$ $\frac{X}{Y}$, ha Y sehol sem nulla

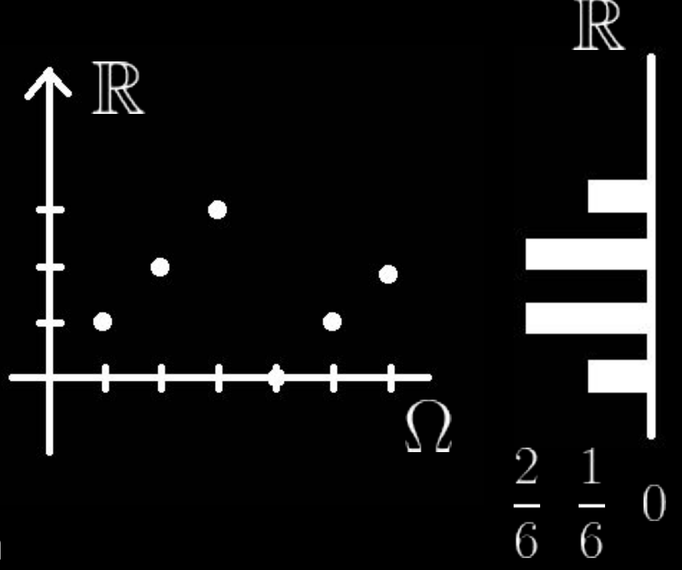
X^2 $\quad$ $X \cdot Y$

Vagyis beszélhetünk a következő eseményekről:

$\{X + 42 < 0\}$ $\quad$ $\{X - Y \geq 10\}$ $\quad$ etc.

Egyszerű val. változó, eloszlás

Definíció: Egy valószínűségi változó *egyszerű*,
ha értékkészlete véges.



Definíció: Egy egyszerű val. változó *eloszlása*:

$$\Big\{\mathbb{P}(X = x) \;\Big|\; x \in \mathrm{Ran}(X)\Big\}$$

Példa: X két kockával a dobott egyesek száma

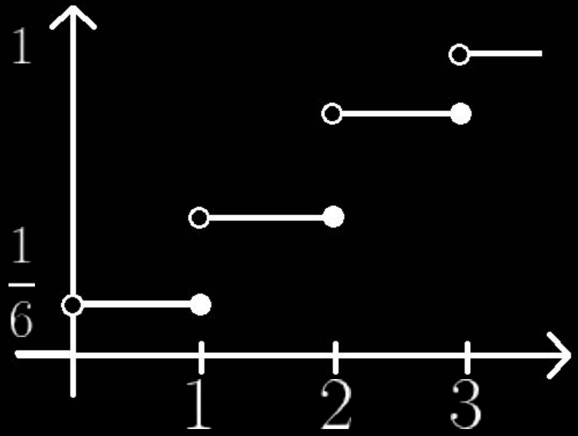
$$\mathbb{P}(X = 0) = \frac{25}{36} \quad \mathbb{P}(X = 1) = \frac{10}{36} \quad \mathbb{P}(X = 2) = \frac{1}{36}$$

Eloszlásfüggvény

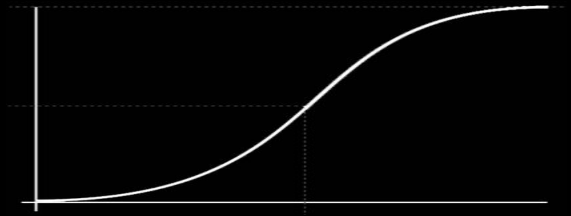
Definíció: Az X val. változó *eloszlásfüggvénye*:

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

Példa: kockadobás négyes maradék, lásd ábra.



(Eloszlásfüggvénye nem csak egyszerű valószínűségi változónak van.)



Várható érték, def.

Definíció: Legyen X egyszerű val. változó. Ekkor

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

Értelmezések:

- a véletlen mennyiség átlagos értéke (!= tipikus érték)
- értékek súlyozott közepe, ahol a súlyok a valószínűségek

Példa: kockadobás 4-es maradéka

$$\mathbb{E}(X) = 0 \cdot \frac{1}{6} + 1 \cdot \frac{2}{6} + 2 \cdot \frac{2}{6} + 3 \cdot \frac{1}{6} = \frac{3}{2}$$

Várható érték, példák

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

1. Kockadobás

$$\mathbb{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3{,}5$$

2. Az előző val. változó négyzete

$$\mathbb{E}(X^2) = (1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2)\frac{1}{6} = \frac{91}{6}$$

Várható érték, példák

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

3. Egy urnában 2 piros, 3 fehér golyó van. Visszatevés nélkül húzunk, amíg fehéret nem kapunk. Jelölje Z a kihúzott piros golyók számát.

$$\mathbb{E}(Z) = 0 \cdot \frac{2}{5} + 1 \cdot \frac{3}{5} \cdot \frac{2}{4} + 2 \cdot \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{2}{3} + 3 \cdot \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} \cdot \frac{2}{2} = 1$$

4. Indikátor val. változó:

$$\mathbb{E}(\mathbf{1}_A) = 0 \cdot \mathbb{P}(\mathbf{1}_A = 0) + 1 \cdot \mathbb{P}(\mathbf{1}_A = 1) = \mathbb{P}(A)$$

Várható érték additivitása

Legyen X és Y egyszerű valószínűségi változó.

Állítás: Tetszőleges $c \in \mathbb{R}$ esetén $\mathbb{E}(c \cdot X) = c \cdot \mathbb{E}(X)$

Állítás: $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

Ez nem nyilvánvaló, pl. két kockadobás összegének eloszlására:

Megjegyzés: Az additivitás nem csak egyszerű val. változóra lesz igaz.

Poincaré-formula, biz.

Állítás: $\mathbb{P}(A_1 \cup A_2 \cup A_3) =$

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

Bizonyítás:

$$\mathbb{P}(A_1 \cup A_2 \cup A_3) = 1 - \mathbb{P}\Big(\cap_i \overline{A_i} \Big) = 1 - \mathbb{E}\Big(1_{\cap_i \overline{A_i}}\Big)$$

$$= 1 - \mathbb{E}\Big(\prod_i 1_{\overline{A_i}} \Big) = 1 - \mathbb{E}\Big(\prod_i (1 - 1_{A_i}) \Big)$$

Poincaré-formula, biz.

$$= 1 - \mathbb{E}\Big(\prod_i (1 - 1_{A_i})\Big)$$

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

$$\mathbf{1}_{A_1}\mathbf{1}_{A_2} + \mathbf{1}_{A_1}\mathbf{1}_{A_3} + \mathbf{1}_{A_2}\mathbf{1}_{A_3} - \mathbf{1}_{A_1}\mathbf{1}_{A_2}\mathbf{1}_{A_3}\Big)$$

$\Rightarrow$ állítás

Binomiális eloszlás, def.

Definíció: Egy X valószínűségi változó *binomiális eloszlású* az $n \in \mathbb{N}$ és $p \in [0, 1]$ paraméterekkel, ha

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad (\forall k : 0 \leq k \leq n)$$

Jelölés: $X \sim B(n, p)$

Példa: n kosárra dobásból hány megy be?

Általánosan, azonos kísérletekből hány “siker”?

Binomiális eloszlás, v. é.

Állítás: $X \sim B(n,p) \Rightarrow \mathbb{E}(X) = n \cdot p$

Bizonyítás:

Legyenek $A_1, \ldots, A_n$ együttesen független események, amik valószínűsége mind p. Nézzük a következő val. változót:

$$Y = \mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_n}$$

Az additivitás miatt:

$$\begin{aligned}\mathbb{E}(Y) &= \mathbb{E}\left(\mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_n}\right) \\ &= \mathbb{E}\left(\mathbf{1}_{A_1}\right) + \cdots + \mathbb{E}\left(\mathbf{1}_{A_n}\right) = n \cdot p\end{aligned}$$

A várható érték csak az eloszlástól függ, viszont X és Y eloszlása ugyanaz.

Diszkrét egyenletes eloszlás

Definíció: Egy X valószínűségi változó *diszkrét egyenletes eloszlás*ú az n elemű S halmazon, ha

$$\mathbb{P}(X = k) = \frac{1}{n} \qquad (\forall k \in S)$$

Megjegyzés: Ha $S = \{1, 2, \ldots, n\}$

akkor

$$\mathbb{E}(X) = \frac{1 + 2 + \cdots + n}{n} = \frac{n+1}{2}$$

Randomized quicksort

Input: n elemű tömb, elemei összehasonlíthatók (pl. számok)

Output: ugyanez sorba rendezve

Algoritmus:

0. Ha $n \leq 1$ akkor csak visszaadjuk.
1. Egyébként veszünk egy P pivot elemet, és ketté osztjuk a tömböt:
   - P-nél kisebbekre, és
   - P-nél nagyobbakra.
2. Hívjuk meg az eredeti függvényt erre a két kisebb tömbre.
3. Konkatentáljuk az eredményeket az alábbi sorrendben:

   rendezett(P-nél kisebbek), [P], rendezett(P-nél nagyobbak)

Kérdés: Hogyan válasszuk P-t?

Randomized quicksort, áll.

Megjegyzés: Ha mindig a legnagyobbat választjuk, akkor $\binom{n}{2} \approx n^2$ összehasonlítást kell tegyünk.

Válasszuk a P-t egyenletesen véletlenszerűen. Jelölje X_n egy n elemű tömb sorba rendezése közben az elvégzett összehasonlítások számát.

Állítás: $\mathbb{E}(X_n) \leq 2n\ln(n)$

Randomized quicksort, biz.

Bizonyítás: Legyen $X_{i,j}$ az a val. változó, ami 1, ha az i-edik és j-edik elemet valamikor össze kellett hasonlítanunk, és 0 egyébként. Ekkor

$$X_n = \sum_{1 \le i < j \le n} X_{i,j}$$

(Ezek nem függetlenek.)

Az additivitás miatt

$$\mathbb{E}(X_n) = \sum_{1 \le i < j \le n} \mathbb{E}(X_{i,j})$$

Randomized quicksort, biz.

Meggondolható: $\mathbb{E}(X_{i,j}) = \dfrac{2}{j-i+1}$

És tényleg: Legyenek a tömb elemei $y_1, \dots, y_n$

Nézzük csak az $y_i, y_{i+1}, \dots, y_j$ elemeket. (Ez $j - i + 1$ db elem.)

Melyiket választjuk ki legelőször pivot elemnek?

- Ha az y_i vagy az y_j elemet, akkor $X_{i,j} = 1$
- Ha bármi mást, akkor ők sose lesznek összehasonlítva, ezért $X_{i,j} = 0$

Az eloszlás egyenletes, ezért $\mathbb{E}(X_{i,j}) = \mathbb{P}(X_{i,j} = 1) = \dfrac{2}{j-i+1}$

Randomized quicksort, biz.

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

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

$$\Rightarrow \begin{matrix} \text{calc} \ldots \\ \quad \text{calc} \ldots \\ \qquad \text{calc} \ldots \end{matrix} \quad \le 2n \ln(n)$$

Várható érték, végtelen eset

Kérdés: mi van ha $\mathrm{Ran}(X)$ nem véges? Akkor is van várható érték?

Ilyenkor az eredeti képlet nem feltétlenül működik. Mi legyen a def?

Definíció: Ha X nemnegatív valószínűségi változó, akkor

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

Ez általánosítja a korábbi definíciót.

Várható érték, diszkrét eset

Állítás: Legyen X olyan nemnegatív val. változó, amire $\mathrm{Ran}(X)$ megszámlálhatóan végtelen. Jelölje az elemeit $k_1, k_2, \ldots$. Ekkor

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

Példa: Egy tű fokán próbálunk áthúzni egy szál cérnát. Minden alkalommal p eséllyel sikerül. Legyen X a szükséges próbálkozások száma. Ekkor

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k \cdot \mathbb{P}(X = k) =$$ (folyt. köv.)

Folytonos eset, dióhéjban

Geometriai valószínűség kontextusában is vannak valószínűségi változók.

Probléma: $\mathbb{P}(X = k) = 0$ minden k -ra. Vagyis az "eloszlás" nem mond semmit.

Megoldás(ok):

- Az eloszlásfüggvény itt is értelmes.
- Az eloszlás helyett az eloszlásfüggvény deriváltját fogjuk nézni (ez lesz a sűrűségfüggvény).

Köszönöm a figyelmet!