

(03) Diskrétné náhodné premenné

1. Distribučná funkcia: dve vlastnosti

Na zopakovanie definícia:

Distribučná funkcia náhodnej premennej

- Pre náhodnú premennú X sa jej distribučná funkcia definuje ako

$$F(a) = P(X \leq a)$$

VLASTNOSŤ I.

Monotónnosť F :

- Je zrejmé, že F je **neklesajúca**.
- Napríklad $P(X \leq 2)$ nemôže byť viac ako $P(X \leq 3)$

Je to analógia toho, čo vieme povedať o počtoch prvkov množín: Ak je A podmnožinou B , tak A nemôže mať viac prvkov ako B .

V reči pravdepodobnosti:

Ak nastala udalosť $X \leq 3$, tak nastala aj udalosť $X \leq 2$. Ak doba čakania nebude viac ako 3 minúty, tak nebude ani viac ako 2 minúty. Takže pravdepodobnosť toho, že nebudeme čakať viac ako 2 minúty nemôže byť väčšia ako pravdepodobnosť toho, že nebudeme čakať viac ako 3 minúty.

VLASTNOSŤ II.

Najskôr si pozrime naše dva špeciálne prípady z príkladu:

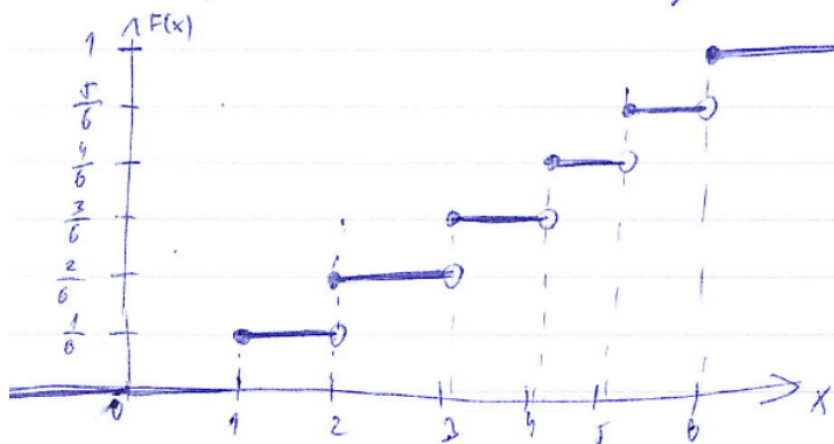
Naše príklady: čakanie Romea a Júlie, hádzanie kockou

- Hodnoty náhodnej premennej boli ohraničené, preto
 - od určitého bodu naľavo bola F nulová
 - od určitého bodu napravo bola F rovná jednej

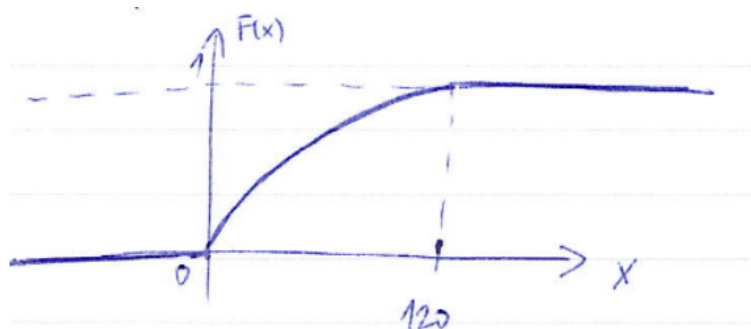
Limitné správanie F vo všeobecnosti:

- **limita v mínus nekonečne je 0**
- **limita v plus nekonečne je 1**

Ak hádzeme kockou a X je počet padnutých bodov, tak napríklad pre každé záporné x je pravdepodobnosť toho, že $X \leq x$, rovná nule, lebo takáto situácia nemôže nastať (platí to aj pre kladné čísla menšie ako 1). Takisto pre x väčšie alebo rovné ako 6 je pravdepodobnosť toho, že $X \leq x$, rovná jednej, lebo táto udalosť určite nastane.



Podobne v prípade čakania Romea a Júlie, keďže čas čakania nemôže byť záporný a je zhora ohraničený na základe informácie o tom, kedy na miesto stretnutia prichádzajú hodnotou 120. Distribučná funkcia je vľavo od tejto dolnej hranice 0 nulová a vpravo od hornej hranice 120 rovná jednej.

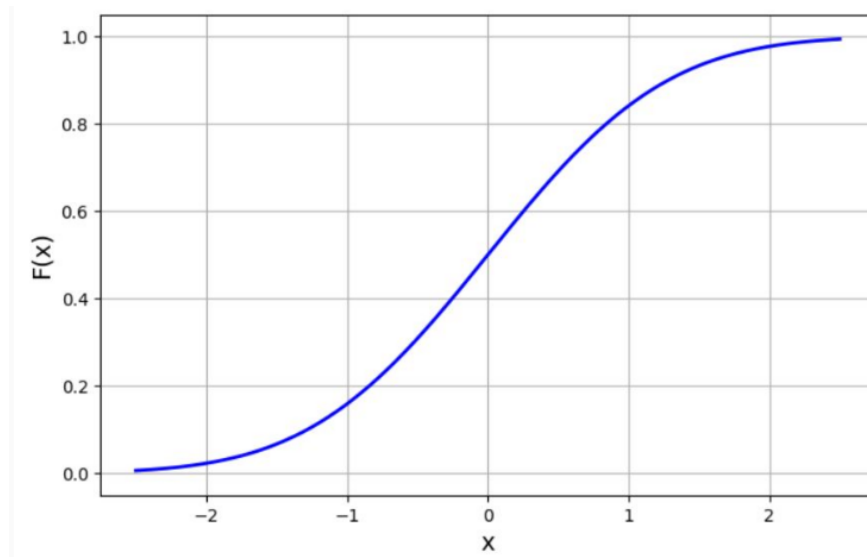


Nemusí to tak byť vždy. Môžeme mať náhodnú premennú, ktorá nie je zdola ohraničená. Hodnota $F(x)$ nemôže byť záporná, lebo je to pravdepodobnosť. Ale ak počítame $P(X \leq x)$ a hodnotu x stále znižujeme, tak sa dostaneme hocijako blízko k nule. Nie je možné, aby sa nedala nájsť hranica, pod ktorou je napríklad len 1 percento hodnôt. To by znamenalo, že pod hocijakým x je viac ako 1 percento hodnôt, čo nie je možné. Toto, samozrejme, nie je matematický dôkaz, ale argumentácia, ktorá by mala presvedčiť, že tá vlastnosť "logicky musí" platiť.

Analogicky, ak náhodná premenná nemá zhora ohraničené hodnoty. Je jasné, že $F(x)$ nemôže byť viac ako 1, lebo je to pravdepodobnosť. Avšak platí, že ak počítame $P(X \leq x)$

a hodnotu x stále zväčšujeme, po čase budeme ľubovoľne blízko jednotky (bez ohľadu na to, ako blízko k jednotke sme si hranicu stanovili). Nemôže sa stať, že hocijako veľké x zvolíme, vždy bude pod ním menej ako 99 percent hodnôt.

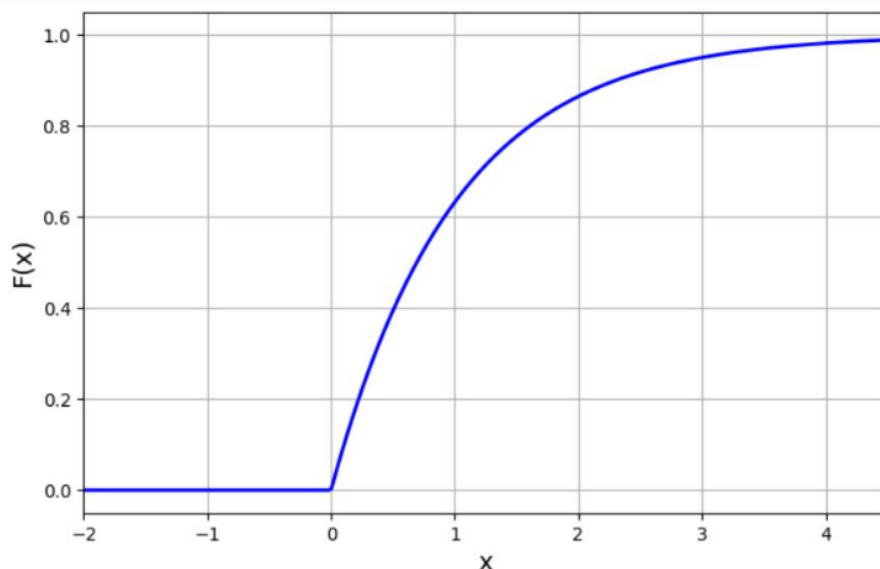
Distribučná funkcia náhodnej premennej, ktorá nadobúda ľubovoľne veľké kladné aj ľubovoľne veľké záporné hodnoty (chápeme to v absolútnej hodnote), môže vyzeráť napríklad takto:



Nikde sa nerovná nule ani jednotke, ale limitne sa k týmto hodnotám približuje – k nule smerom do mínus nekonečna a k jednotke smerom do plus nekonečna.

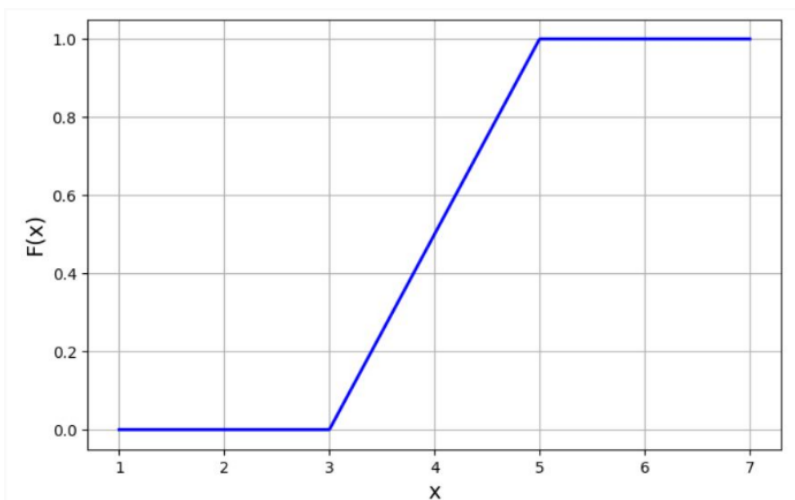
2. Distribučná funkcia: príklady

PRÍKLAD 1:



- Platí $F(0)=0$, teda $P(X \leq 0)=0$. To znamená, že distribučná funkcia zodpovedá náhodnej premennej, ktorá nadobúda iba kladné hodnoty.
- Na druhej strane, hodnoty tejto náhodnej premennej nie sú zhora ohraničené, distribučná funkcia sa nikde nerovná jednej, iba sa k jednotke približuje.

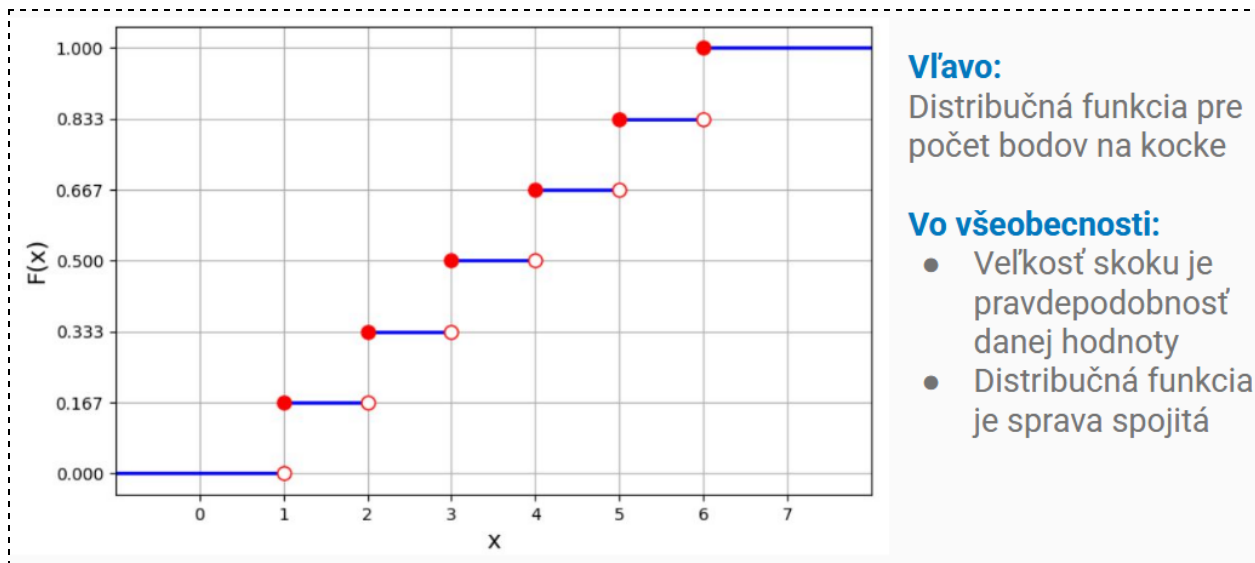
PRÍKLAD 2:



- Analogicky ako v predchádzajúcom príklade platí $F(3)=0$, teda $P(X \leq 3)=0$.
- Na druhej strane, $F(5)=1$, teda $P(X \leq 5)=1$. Nerovnosť $X \leq 5$ je istá udalosť.
- Takže táto náhodná premenná nadobúda iba hodnoty medzi 3 a 5.

3. Distribučná funkcia: ďalšia vlastnosť

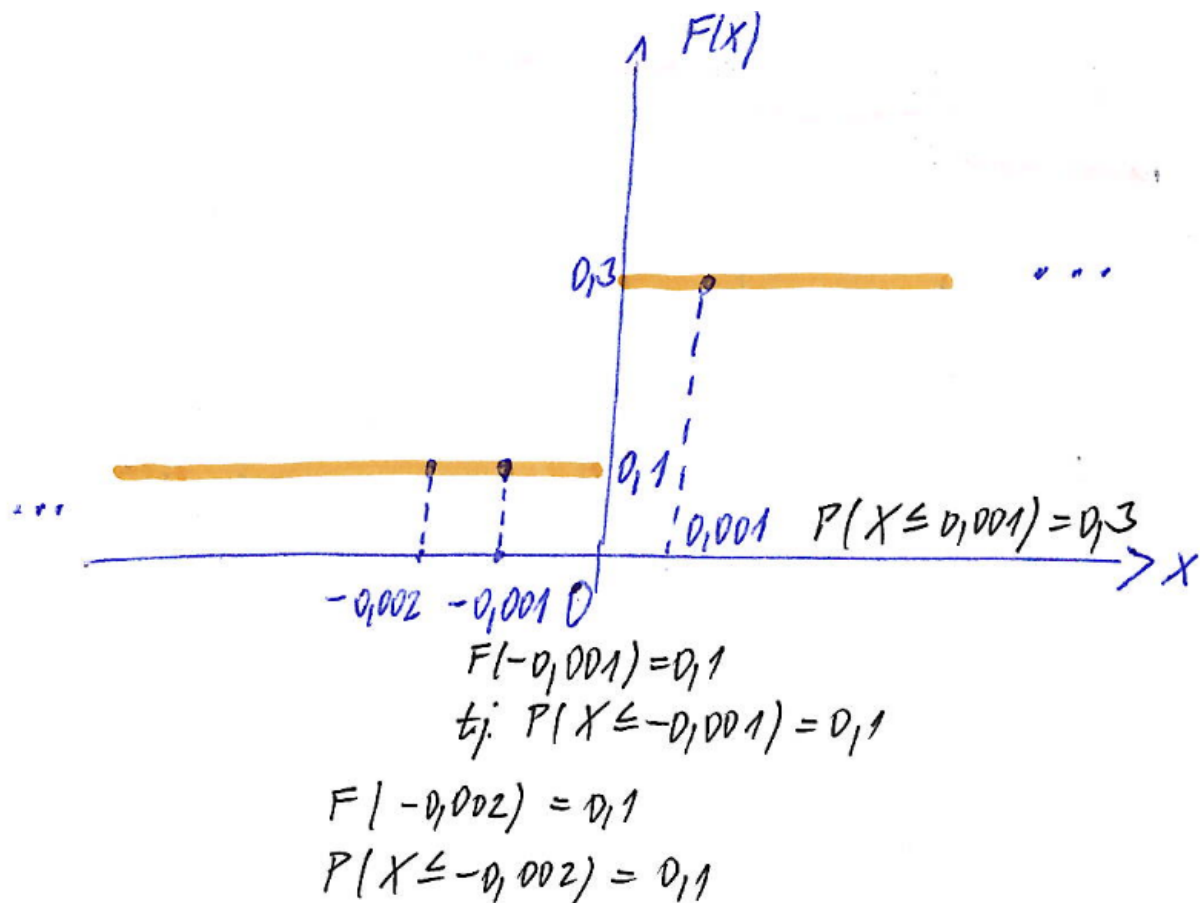
Pripomeňme si distribučnú funkciu počtu bodov pri hode kockou:



“Spojitosť sprava” znamená, že ak sa graf sprava približuje k nejakému bodu, tak ten bod je súčasťou grafu funkcie. Tam, kde je funkcia spojitá, je to jasné. Zaujímavé je to v bodoch, kde spojitá nie je, teda tam, kde sú skoky.

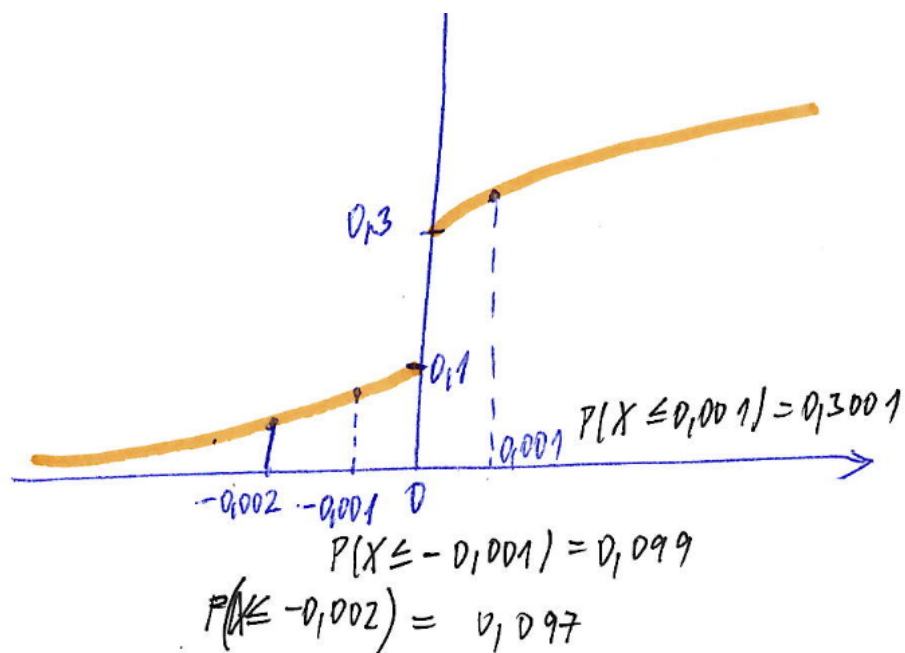
Všimnime si, že v bodoch, v ktorých je skok, sa nadobúda tá vyššia hodnota. Toto platí aj vo všeobecnosti. Ak distribučná funkcia nie je v nejakom bode spojitá, vzhľadom na neklesajúcosť musí byť v tomto bode skok. Pozrime sa, čo to znamená.

Ukážeme si dva príklady. V prvom prípade bude distribučná funkcia podobná tej predchádzajúcej, pred aj po skoku bude na nejakom intervale konštantná. V druhom prípade bude pred aj po skoku rásť. Spoločné majú to, že v nule skočí funkcia z hodnoty 0,1 na 0,3.

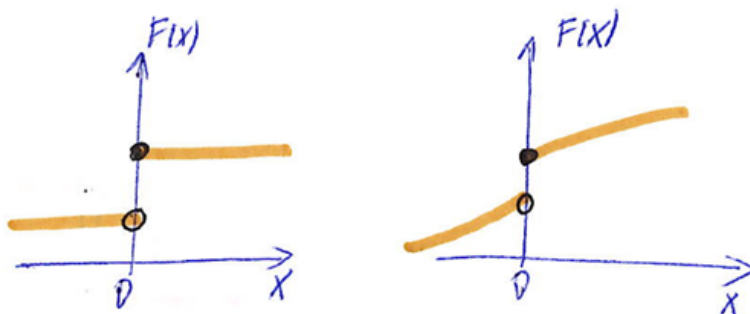


Vidíme, že v tomto prípade sa musí tá nula nadobúdať s pravdepodobnosťou $0.3 - 0.1 = 0.2$. Súčasne to znamená, že funkčná hodnota v nule bude $P(X \leq 0) = 0.3$, lebo udalosť $X \leq 0$ v seba zahŕňa aj $X = 0$ (na rozdiel od napríklad $X \leq -0.00001$, čo nulu nezahŕňa).

Podobne to bude v prípade, ak distribučná funkcia pred skokom aj po skoku rastie:



V oboch prípadoch je teda "plný krúžok", teda nadobúdaná hodnota tá vyššia:

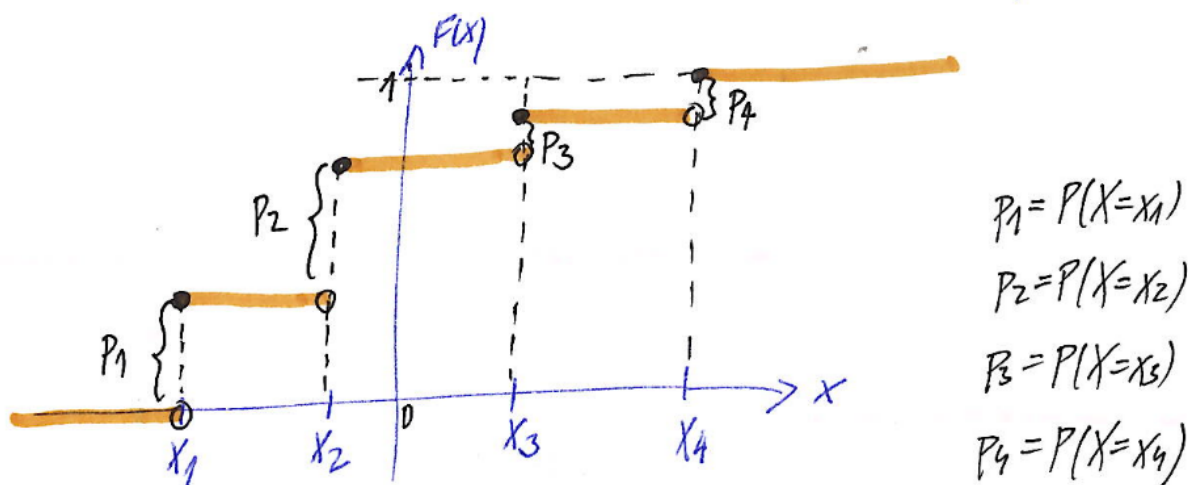


4. Diskrétné náhodné premenné

Diskrétné náhodné premenné

- Nadobúdajú **spočítateľne veľa hodnôt**
 - teda konečný počet
 - alebo nekonečný, ktorý sa dá očíslovať – napríklad všetky prirodzené čísla, všetky celé čísla, ale nie napríklad všetky hodnoty z intervalu $(0, 1)$
- Charakterizujú sa **hodnotami a príslušnými pravdepodobnosťami**.
- Distribučná funkcia je po častiach konštantná

PRÍKLAD:



5. Stredná hodnota a disperzia diskkrétnej náhodnej premennej

Ide o analógiu aritmetického priemeru (**stredná hodnota**) a výberovej disperzie (**disperzia**), ktoré poznáme z popisu dátového súboru.

Súbor dát

hodnota	početnosť	relatívna početnosť
x_1	m_1	$\frac{m_1}{m}$
x_2	m_2	$\frac{m_2}{m}$
$\vdots$	$\vdots$	$\vdots$
x_n	m_n	$\frac{m_n}{m}$

$m = m_1 + \dots + m_n$

priemer: $\frac{1}{m}(x_1 m_1 + \dots + x_n m_n)$
 $= \frac{m_1}{m} x_1 + \dots + \frac{m_n}{m} x_n$

Náhodná premenná

hodnota	pravdepodobnosť
x_1	P_1
x_2	P_2
$\vdots$	$\vdots$
x_n	P_n

stredná hodnota
 $E(X) = x_1 P_1 + \dots + x_n P_n$

Súbor dát

$y_1, y_2, \dots, y_N$

priemer $\bar{y}$

výberová disperzia: $\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$

t.j. priemer hodnôt $(y_i - \bar{y})^2$

Náhodná prem. X

disperzia:

$D(X) = E((X - E(X))^2)$

Matematická poznámka:

V prípade nekonečného počtu hodnôt $x_1, x_2, x_3, \dots$ a príslušných pravdepodobností $p_1, p_2, p_3, \dots$ počítame pri výpočte strednej hodnoty súčet nekonečného radu

$$E(X) = x_1 p_1 + x_2 p_2 + \dots = \sum_{k=1}^{\infty} x_k p_k$$

Aby malo zmysel hovoriť o takejto sume, musí rad $\sum_{k=1}^{\infty} x_k p_k$ konvergovať. Navyše sa však požaduje, aby konvergoval absolútne, teda aby konvergoval rad $\sum_{k=1}^{\infty} |x_k p_k|$.

Dôvod je ten, že z matematickej analýzy známa nasledujúca vlastnosť: Ak by rad konvergoval, ale nie absolútne, tak preusporiadaním jeho členov do iného poradia by sme vedeli dosiahnuť iný súčet (nezávislosť súčtu od poradia sčítovania platí len pre konečný počet sčítancov). To by ale znamenalo, že stredná hodnota náhodnej premennej by závisela od toho, v akom poradí vypíšeme jej hodnoty – čo nedáva zmysel. Aby sme sa vyhli takejto situácii, požaduje sa absolútna konvergencia radu. V tom prípade (to je znovu tvrdenie z matematickej analýzy) súčet nezávisí od poradia členov.

6. Stredná hodnota a disperzia: vlastnosti I. + definícia štandardnej odchýlky

- Ak náhodná premenná nadobúda iba kladné hodnoty, jej stredná hodnota je kladná.
- Ak nadobúda iba nezáporné hodnoty, jej stredná hodnota je nezáporná.
- Analogicky pre hodnoty, ktoré sú záporné a menšie alebo rovné nule.

Keďže v definícii disperzie vystupuje stredná hodnota nejakej druhej mocniny, samotná disperzia je nezáporná. To umožňuje definovať **štandardnú odchýlku** ako odmocninu z disperzie (rovnako, ako to bolo v prípade výberovej disperzie a štandardnej odchýlky).

Dôsledok: **disperzia je vždy nezáporná**

$$\mathbb{D}(X) \geq 0$$

Štandardná odchýlka

- Štandardná odchýlka náhodnej premennej sa definuje ako **odmocnina z disperzie**

7. Stredná hodnota a disperzia: vlastnosti II.

- Pre strednú hodnotu a disperziu konštanty platí:

$$\mathbb{E}(a) = a, \quad \mathbb{D}(a) = 0$$

- Konštantu **a** si môžeme predstaviť ako náhodnú premennú, ktorá nadobúda hodnotu **a** s pravdepodobnosťou 1. Výpočtom strednej hodnoty z definície potom dostaneme, že jej stredná hodnota je **a**.
- Potom je však **a - E(a)** konštantne rovné nule, a teda pri výpočte disperzie počítame strednú hodnotu nuly, čo je nula. Takže disperzia konštantnej náhodnej premennej je nulová.

8. Stredná hodnota a disperzia: vlastnosti III.

Nasledovná vlastnosť nám môže zjednodušiť výpočty.

Ak **X** je náhodná premenná (teraz pracujeme s distrétnymi, ale platí to aj pre ostatné - strednú hodnotu a disperziu pre ne definujeme neskôr):

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

Dá sa jednoducho odvodiť z toho, o čom sme už hovorili:

$$\begin{aligned}
 \boxed{D(X)} &= E((X - E(X))^2) = E(X^2 - 2X \cdot E(X) + (E(X))^2) = \\
 &= E(X^2) - E(\underbrace{2X \cdot E(X)}_{\text{konst.}}) + E(\underbrace{(E(X))^2}_{\text{konst.}}) = \\
 &= E(X^2) - 2E(X) \cdot E(X) + (E(X))^2 = \\
 &= \boxed{E(X^2) - (E(X))^2}
 \end{aligned}$$

9. Stredná hodnota a disperzia: výpočet I.

Príklad 1

Nech X je náhodná premenná označujúca počet bodov, ktoré padli na kocke.

- Aká je stredná hodnota náhodnej premennej X ?
- Aká je disperzia náhodnej premennej X ?

RIEŠENIE:

k	1	2	3	4	5	6
$P(X=k)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = \underline{\underline{3,5}}$$

Pozrime sa, ako vyzerá druhá mocnina tejto náhodnej premennej a jej stredná hodnota:

k	1	4	9	16	25	36
$P(X^2=k)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$E(X^2) = 1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 9 \cdot \frac{1}{6} + 16 \cdot \frac{1}{6} + 25 \cdot \frac{1}{6} + 36 \cdot \frac{1}{6} = \frac{91}{6}$$

To nám umožní vypočítať disperziu:

$$D(X) = E(X^2) - (E(X))^2 = \frac{91}{6} - \left(\frac{21}{6}\right)^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12} \approx 2,9167$$

10. Stredná hodnota a disperzia: výpočet II.

Príklad 2

Hodíme dvakrát kockou. Nech

- X je náhodná premenná označujúca súčet počtu bodov, ktoré padli na kockách,
- Y je maximum z týchto hodnôt,
- Z je rozdiel medzi väčšou a menšou hodnotou (nula, ak sú rovnaké).

Aká je stredná hodnota a disperzia týchto náhodnej premenných?

Na kontrolu našich výsledkov si spravíme simulácie, výsledky by im mali byť podobné (ak sa príliš líšia, treba si hľadať chybu):

```
import numpy as np
```

```
pocet_pokusov = 10**6
```

```
# Simulácia hodov kockou
```

```
kocka1 = np.random.randint(1, 7, size=pocet_pokusov)
```

```
kocka2 = np.random.randint(1, 7, size=pocet_pokusov)
```

```
# Definícia náhodných premenných
```

```
X = kocka1 + kocka2
```

```
# sucet
```

```
Y = np.maximum(kocka1, kocka2)
```

```
# maximum
```

```
Z = np.abs(kocka1 - kocka2)
```

```
# vacsie - mensie
```

```
X (súčet): stredná hodnota = 6.997, disperzia = 5.843
```

```
Y (maximum): stredná hodnota = 4.470, disperzia = 1.976
```

```
Z (väčšie - menšie): stredná hodnota = 1.944, disperzia = 2.055
```

("Stredná hodnota" a "disperzia" tu označuje ich odhad pomocou aritmetického priemeru a výberovej disperzie.)

Pomôcka k určovaniu pravdepodobností:

1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

1	2	3	4	5	6
1	2	3	4	5	6
1	2	3	4	5	6
1	2	3	4	5	6
1	2	3	4	5	6
1	2	3	4	5	6

RIEŠENIE PRE X:

X:	k	2	3	4	5	6	7	8	9	10	11	12
	$P(X=k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	

$$E(X) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} + \dots =$$

$$= \frac{2+6+12+20+30+42+40+36+30+22+12}{36} = \frac{252}{36} = 7$$

$$E(X^2) = 4 \cdot \frac{1}{36} + 9 \cdot \frac{2}{36} + 16 \cdot \frac{3}{36} + 25 \cdot \frac{4}{36} + \dots =$$

$$= \frac{1974}{36} = \frac{987}{18} = \frac{329}{6}$$

$$D(X) = E(X^2) - (E(X))^2 = \frac{329}{6} - 7^2 = \frac{35}{6} \approx 5,8333$$

RIEŠENIE PRE Y:

Y:	k	1	2	3	4	5	6
	$P(Y=k)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

$$E(Y) = 1 \cdot \frac{1}{36} + 2 \cdot \frac{3}{36} + 3 \cdot \frac{5}{36} + 4 \cdot \frac{7}{36} + 5 \cdot \frac{9}{36} + 6 \cdot \frac{11}{36} =$$

$$= \frac{161}{36} \doteq 4,4722$$

$$E(Y^2) = 1 \cdot \frac{1}{36} + 4 \cdot \frac{3}{36} + 9 \cdot \frac{5}{36} + 16 \cdot \frac{7}{36} + 25 \cdot \frac{9}{36} + 36 \cdot \frac{11}{36} =$$

$$= \frac{791}{36}$$

$$DM) = E(Y^2) - (E(Y))^2 = \frac{791}{36} - \left(\frac{161}{36}\right)^2 = \frac{2555}{1296} \doteq 1,9715$$

RIEŠENIE PRE Z:

Z:	k	0	1	2	3	4	5
	$P(Z=k)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

$$E(Z) = 0 \cdot \frac{6}{36} + 1 \cdot \frac{10}{36} + 2 \cdot \frac{8}{36} + 3 \cdot \frac{6}{36} + 4 \cdot \frac{4}{36} + 5 \cdot \frac{2}{36} =$$

$$= \frac{70}{36} = \frac{35}{18} \doteq 1,9444$$

$$E(Z^2) = 0 \cdot \frac{1}{36} + 1 \cdot \frac{10}{36} + 4 \cdot \frac{8}{36} + 9 \cdot \frac{6}{36} + 16 \cdot \frac{4}{36} + 25 \cdot \frac{2}{36} =$$

$$= \frac{210}{36} = \frac{35}{6}$$

$$D(Z) = E(Z^2) - (E(Z))^2 = \frac{35}{6} - \left(\frac{35}{18}\right)^2 = \frac{665}{324} \approx 2,0525$$

11. Stredná hodnota a disperzia: výpočet II.

Príklad 3

- Hádzeme kockou. Potom, čo uvidíme padnuté číslo, môžeme si zobrať výhru rovnú padnutému číslu alebo hádzať znovu. V druhom prípade sa naša výhra rovná číslu, ktoré padne v druhom hode.
- Zvolili sme si stratégiu, že v hre budeme pokračovať, ak by sme po prvom kole dostali menej, ako je očakávaná výhra v prípade, že budeme pokračovať.
- Aká je očakávaná výška výhry (teda stredná hodnota výhry, ktorá je náhodnou premennou) v hre pri tejto stratégii?

RIEŠENIE:

Ak budeme pokračovať v druhom kole, naša očakávaná výhra je 3,5 (stredná hodnota počtu bodov, ktoré padnú pri hode kockou). Stratégia teda spočíva v tom, že:

- ak padne 1, 2 alebo 3, tak hážeme ďalej,
- ak padne 4, 5 alebo 6, tak si zoberieme príslušnú výhru.

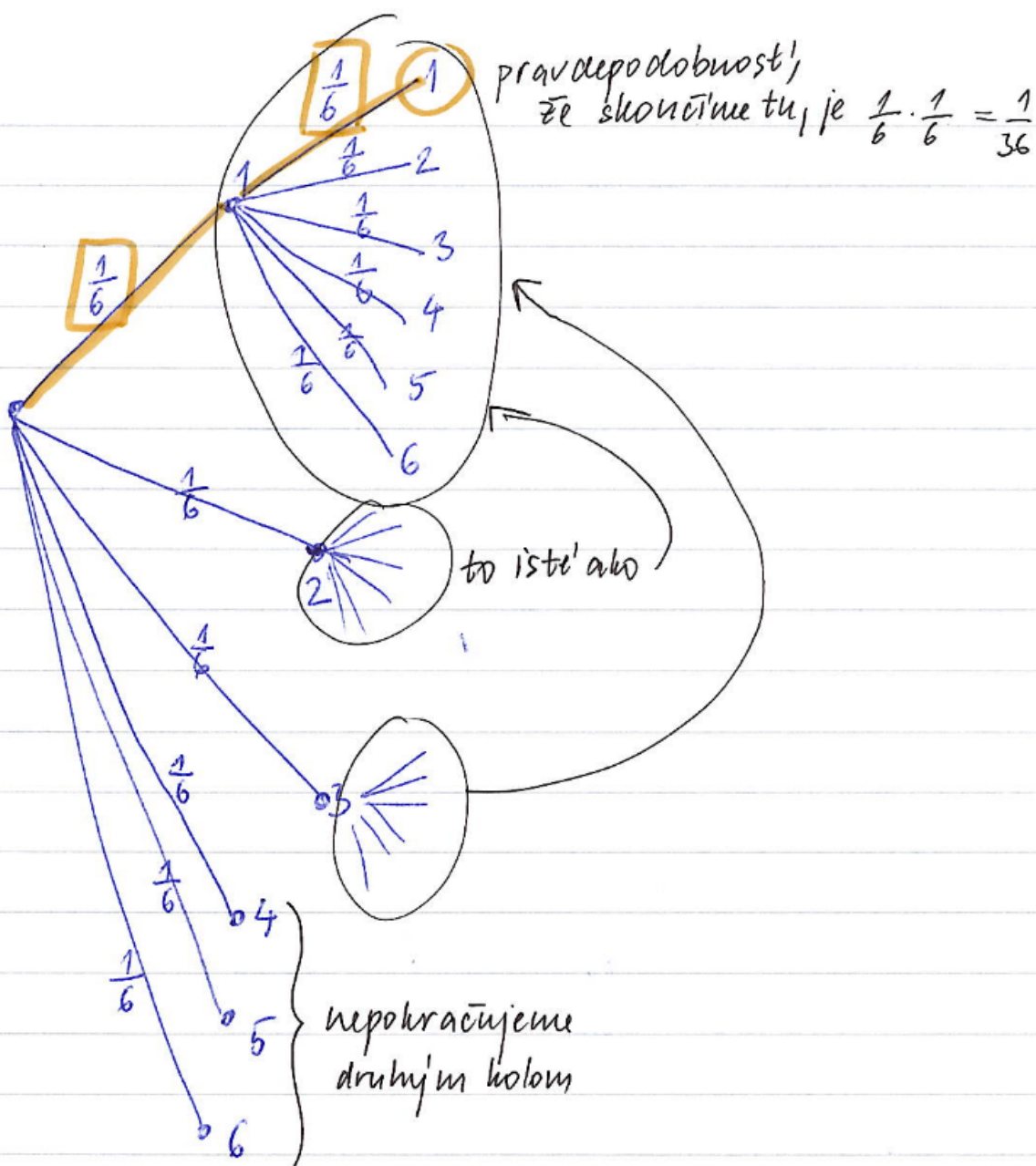
Pravdepodobnostný strom je na nasledujúcej strane. Z neho spočítame, s akou pravdepodobnosťou nastanú jednotlivé možnosti výhry:

výhra	pravdepodobnosť
1	$\frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{3}{36} = \frac{1}{12}$
2	$\frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{3}{36} = \frac{1}{12}$
3	$\frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{3}{36} = \frac{1}{12}$
4	$\frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{6} = \frac{9}{36} = \frac{3}{12}$
5	$\frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{6} = \frac{9}{36} = \frac{3}{12}$
6	$\frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{6} = \frac{9}{36} = \frac{3}{12}$

A z toho strednú hodnotu:

$$\begin{aligned} E(\text{výhra}) &= 1 \cdot \frac{1}{12} + 2 \cdot \frac{1}{12} + 3 \cdot \frac{1}{12} + 4 \cdot \frac{3}{12} + 5 \cdot \frac{3}{12} + 6 \cdot \frac{3}{12} = \\ &= \frac{1+2+3+12+15+18}{12} = \frac{51}{12} = \underline{\underline{4,25}} \end{aligned}$$

Pravdepodobnostný strom:



12. Stredná hodnota a disperzia: vlastnosti IV.

Ak X, Y sú náhodné premenné (teraz pracujeme s distrétnymi, ale platí to aj pre ostatné - strednú hodnotu a disperziu definujeme neskôr) a a je konštanta, tak:

- pre strednú hodnotu platí:

$$\mathbb{E}(aX) = a\mathbb{E}(X)$$

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

- pre disperziu platí:

$$\mathbb{D}(aX) = a^2\mathbb{D}(X)$$

- ale vo všeobecnosti disperzia súčtu náhodných premenných nie je súčet ich disperzií

Ukážeme, prečo platí prvé a tretie tvrdenie:

Prvé tvrdenie:

Ak náhodná premenná X nadobúda hodnoty x_i s pravdepodobnosťami p_i , tak $E(X) = \sum_i x_i p_i$. V tejto situácii

nadobúda náhodná premenná aX hodnoty ax_i s tými istými pravdepodobnosťami p_i . Takže

$$E(aX) = \sum_i (ax_i) p_i = a \cdot \underbrace{\sum_i x_i p_i}_{=E(X)} = a \cdot E(X)$$

Tretie tvrdenie:

$$D(aX) = E((aX)^2) - (E(aX))^2 = E(a^2 X^2) - (E(aX))^2 =$$

$$= a^2 E(X^2) - (a \cdot E(X))^2 =$$

$$= a^2 E(X^2) - a^2 (E(X))^2 =$$

$$= a^2 \cdot \underbrace{(E(X^2) - (E(X))^2)}_{D(X)} = a^2 D(X)$$

POZNÁMKA: Indukciou dostaneme, že druhé tvrdenie platí aj pre súčet n premenných.

PRÍKLAD: V časti 10 sme pre náhodnú premennú X (súčet bodov na dvoch kockách) našli strednú hodnotu tak, že sme našli jej pravdepodobnostné rozdelenie – možné hodnoty a ich pravdepodobnosti. Pomocou tohto tvrdenia však vieme nájsť strednú hodnotu jednoduchšie – spočítať stredné hodnoty pre jednotlivé kocky.

13. Nezávislé náhodné premenné

Nezávislé náhodné premenné

- Niekedy je to intuitívne jasné, napríklad hádzeme niekoľkokrát kockou
- Matematické vyjadrenie nezávislosti diskretných náhodných premenných X, Y :

$$\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i) \mathbb{P}(Y = y_j) \quad \forall x_i, y_j$$

- na overenie nezávislosti
- na výpočet pravdepodobnosti, ak je nezávislosť známa

14. Stredná hodnota a disperzia: vlastnosti V.

Stredná hodnota a disperzia pre nezávislé náhodné premenné:

$$\mathbb{D}(X + Y) = \mathbb{D}(X) + \mathbb{D}(Y)$$

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

POZNÁMKA: Indukciou dostaneme, že tvrdenia platia aj pre n náhodných premenných.

PRÍKLAD: V časti 10 sme pre náhodnú premennú X (súčet bodov na dvoch kockách) našli disperziu tak, že sme našli jej pravdepodobnostné rozdelenie – možné hodnoty a ich pravdepodobnosti, a následne aj pre jej druhú mocninu. Pomocou tohto tvrdenia však vieme nájsť disperziu jednoduchšie – spočítať disperzie pre jednotlivé kocky, keďže hody kockami sú nezávislé.

V nasledujúcich častiach si ukážeme niektoré špeciálne typy distrétnych náhodných premenných a ich použitie.

15. Bernoulliho rozdelenie (iný názov: alternatívne)

Bernoulliho rozdelenie (tiež alternatívne)

- Náhodná premenná s Bernoulliho rozdelením s parametrom p :
 - hodnoty **1, 0**
 - pravdepodobnosti **$p, 1 - p$**

Vlastnosti:

- Stredná hodnota: **p**
- Disperzia: **$p(1-p)$**

Použitie:

- Sledujeme, či určitý jav nastal alebo nenastal.

Jacob Bernoulli



Odvodenie strednej hodnoty a disperzie:

X_i	x_i	0	1
P_i		$1-p$	p

$$E(X) = 0 \cdot (1-p) + 1 \cdot p = p$$
$$E(X^2) = 0^2 \cdot (1-p) + 1^2 p = p$$
$$D(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1-p)$$

16. Bernoulliho rozdelenie: príklad

**marilynvossavant.com**
[> home](#) [> ask a question](#) [> discussions](#) [> about marilyn](#) [> ideas](#)

Príklad 4
Vrátime sa k príkladu o písomke z dejepisu.

Marilyn would love to hear from you!
To send her a question or comment, fill out the form on the right, then click "Submit."

Ask a Question
Name:*

A high school student who hadn't opened his American history book in weeks was dismayed to walk into class and be greeted with a pop quiz. It was in the form of two lists, one naming the 24 presidents in office during the 19th century in alphabetical order and another list noting their terms in office, but scrambled. The object was to match the presidents with their terms. The completely clueless student had to guess every time. On average, how many did he guess correctly?

RIEŠENIE:

$$X_i = \begin{cases} 1 & \text{ak je } i\text{-ty prezident priradený správne} \\ 0 & \text{inak} \end{cases}$$

$$P(X_i = 1) = \frac{1}{24}$$

$$E(X_i) = 1 \cdot P(X_i=1) + 0 \cdot P(X_i=0) = P(X_i=1) = \frac{1}{24}$$

počet správnych odpovedí ozn. X

platí: $X = X_1 + \dots + X_{24}$

$$\Rightarrow E(X) = E(X_1) + \dots + E(X_{24}) = \underbrace{\frac{1}{24} + \dots + \frac{1}{24}}_{24 \text{ identical}} = \underline{\underline{1}}$$

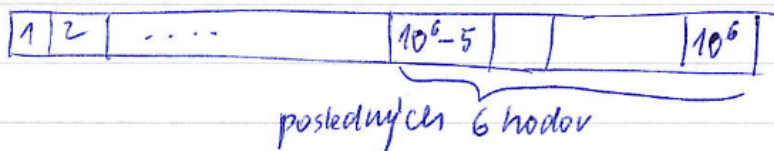
17. Bermoulliho rozdelenie: ďalší príklad

Príklad 5

Hádzeme 1 000 000 krát mincou (výsledky označíma H - hlava, Z - znak). Aká je stredná hodnota počtu výskytov reťazca HHHHHHZZZZZZ (6 hláv, potom 6 znakov)?

RIEŠENIE:

$$X_i = \begin{cases} 1 & \text{ak na } i\text{-tom mieste začína } 6 \times H, 6 \times Z \\ 0 & \end{cases}$$



počet výskytů $X = X_1 + X_2 + \dots + X_N$, kde $N = 10^6 - 5$

$$P(X_i=1) = \underbrace{\frac{1}{2} \cdot \frac{1}{2} \cdots \frac{1}{2}}_{12} = \left(\frac{1}{2}\right)^{12}$$

$$\Rightarrow E(X_i) = P(X_i=1) = \frac{1}{2^{12}} \quad \Rightarrow E(X) = E(X_1) + \dots + E(X_N) =$$

ako v predch.
prípade

$$= \underbrace{\frac{1}{2^{12}} + \dots + \frac{1}{2^{12}}}_{N \text{ sčítancov}} =$$

$$= \frac{N}{2^{12}} = \boxed{\frac{10^6 - 5}{2^{12}}}$$

18. Binomické rozdelenie

Binomické rozdelenie: $\text{Bin}(n, p)$

- n = počet nezávislých pokusov (prirodzené číslo)
- p = pravdepodobnosť úspechu v jednom pokuse (z intervalu $(0,1)$)
- náhodná premenná X vyjadruje počet úspechov v týchto pokusoch

Hovoríme, že " X má binomické rozdelenie s parametrami n a p ", označujeme $X \sim \text{Bin}(n, p)$

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

$$\mathbb{E}(X) = np, \quad \mathbb{D}(X) = np(1-p)$$

Odvozenie strednej hodnoty a disperzie:

Počet úspechov X môžeme napísať ako $X = X_1 + \dots + X_n$, kde $X_i = \begin{cases} 1 & \text{ak v } i\text{-tom pokuse nastal úspech} \\ 0 & \text{inak} \end{cases}$

Vieme, že $E(X_i) = p$, $D(X_i) = p(1-p)$ pre všetky i

Potom: $E(X) = E(X_1) + \dots + E(X_n) = p + \dots + p = np$

Z nezávislosti pokusov tiež vyplýva, že $D(X) = D(X_1) + \dots + D(X_n) =$
 $= p(1-p) + \dots + p(1-p) = np(1-p)$

19. Binomické rozdelenie: príklad, výpočet v Pythone

Príklad 6

Basketbalista hodí loptu do koša s pravdepodobnosťou 0,4. Spravil 10 pokusov. Aké sú pravdepodobnosti pre možné počty úspešných pokusov?

Počet úspechov má binomické rozdelenie $\text{Bin}(10; 0,4)$.

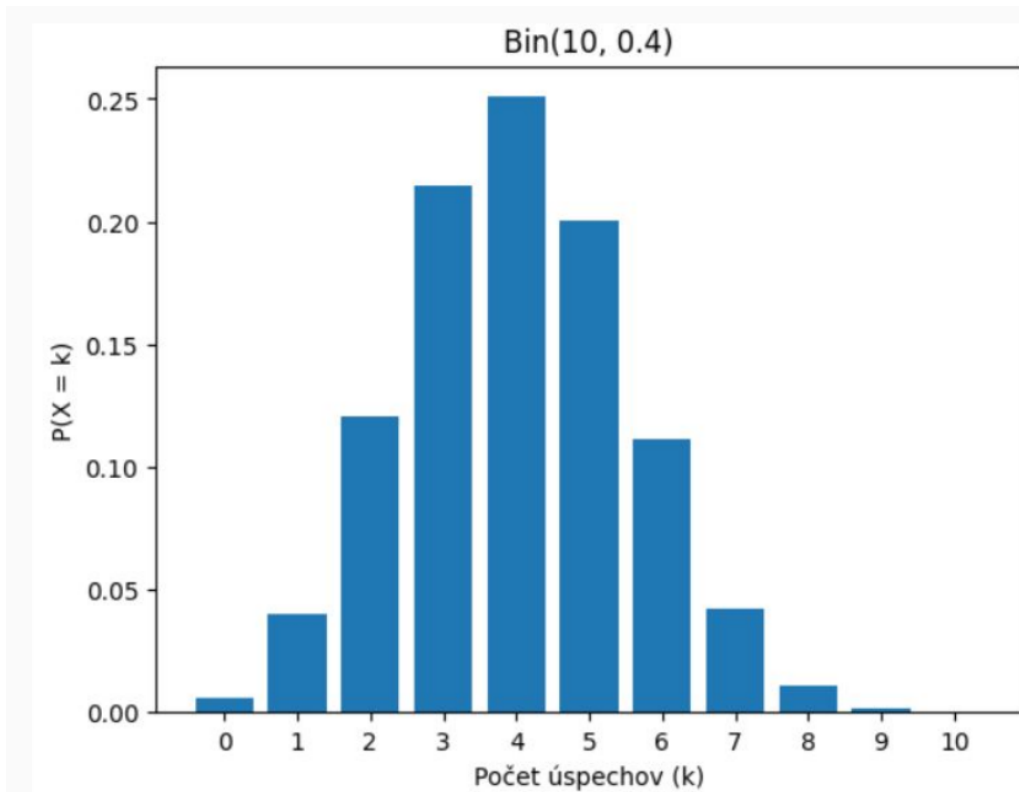
`binom.pmf(k, n, p)`

pmf - probability mass function

```
from scipy.stats import binom
# parametre
n, p = 10, 0.4
# mozne hodnoty nahodnej premennej x
k_hodnoty = range(n + 1)
# pravdepodobnosti P(X = k)
pravdepodobnosti = [binom.pmf(k, n, p) for k in k_hodnoty]
# vypisanie hodnot a pravdepodobnosti
for k, pr in zip(k_hodnoty, pravdepodobnosti):
    print(f"P(X = {k}) = {pr:.4f}")
```

```
P(X = 0) = 0.0060
P(X = 1) = 0.0403
P(X = 2) = 0.1209
P(X = 3) = 0.2150
P(X = 4) = 0.2508
P(X = 5) = 0.2007
P(X = 6) = 0.1115
P(X = 7) = 0.0425
P(X = 8) = 0.0106
P(X = 9) = 0.0016
P(X = 10) = 0.0001
```

Grafické znázornenie:



Príklad 6 - pokračovanie

Aká je pravdepodobnosť, že úspechov nebude viac ako 3?

```
p_vyhovujuce = [binom.pmf(k, n, p) for k in range(0, 4)]  
p_spolu = sum(p_vyhovujuce)  
print(f"P(nie viac ako 3 úspechy) = P(X ∈ {{0, 1, 2, 3}}) = {p_spolu:.4f}")
```

$P(\text{nie viac ako 3 úspechy}) = P(X \in \{0, 1, 2, 3\}) = 0.3633$

```
# P(X ≤ 3) = F(3)  
p_cdf = binom.cdf(3, n, p)  
print(f"P(nie viac ako 3 úspechy) = P(X ≤ 3) = F(3) = {p_cdf:.4f}")
```

$P(\text{nie viac ako 3 úspechy}) = P(X \leq 3) = F(3) = 0.3633$

`binom.cdf(k, n, p)`

`pmf` - cumulative distribution function

20. Binomické rozdelenie: príklad

Príklad 7

Hráme tenis s rovnako silným súperom (teda pravdepodobnosť výhry aj prehry je rovnaká, remíza tu nie je možná). V ktorom z nasledujúcich prípadov máme väčšiu pravdepodobnosť, že vyhráme viac ako polovicu zápasov?

- Hráme 4 zápasy.
- Hráme 8 zápasov.

RIEŠENIE:

$X \sim \text{Bin}(4, \frac{1}{2})$ počet výhier pri prvej možnosti

$P(\text{vyhráme viac ako } \underbrace{\text{polovicu}}_2) = P(X=3) + P(X=4) =$

$$= \binom{4}{3} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^1 + \binom{4}{4} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^0 =$$

$$= \boxed{\frac{5}{16}}$$

$Y \sim \text{Bin}(8, \frac{1}{2})$ počet výhier pri druhej možnosti

$$P(\text{vyhráme viac ako polovicu}) = P(Y=5) + P(Y=6) + P(Y=7) + P(Y=8) =$$

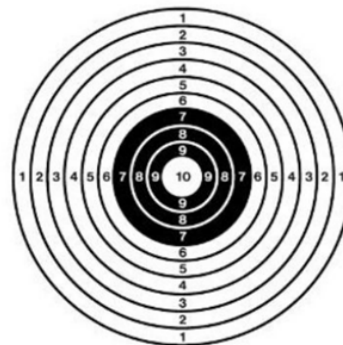
$$= \binom{8}{5} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^3 + \binom{8}{6} \cdot \frac{1}{2^8} + \binom{8}{7} \cdot \frac{1}{2^8} + \binom{8}{8} \frac{1}{2^8} =$$

$$= \frac{56+28+8+1}{2^8} = \boxed{\frac{93}{256}}$$

21. Binomické rozdelenie: ďalší príklad

Príklad 8

Strelec strieľa na terč, jednotlivé pokusy sú nezávislé. Desiatku v strede zasiahne s pravdepodobnosťou 0,2. Koľkokrát musí strieľať, aby bola desiatka zasiahnutá s pravdepodobnosťou aspoň 0,9?



RIEŠENIE:

$$X \sim \text{Bin}(n, \frac{4}{5})$$

$$\begin{aligned} P(\text{zasiahnutá desiatka}) &= 1 - P(\text{nezasiahnutá}) = \\ &= 1 - P(X=0) = \\ &= 1 - \left(\frac{4}{5}\right)^n \geq 0,9 \end{aligned}$$

$$0,1 \geq \left(\frac{4}{5}\right)^n \quad / \ln$$

$$\ln 0,1 \geq \ln \left(\frac{4}{5}\right)^n$$

$$\ln 0,1 \geq n \cdot \ln \left(\frac{4}{5}\right) \quad /: \ln \left(\frac{4}{5}\right) < 0$$

$$\frac{\ln 0,1}{\ln \left(\frac{4}{5}\right)} \leq n$$

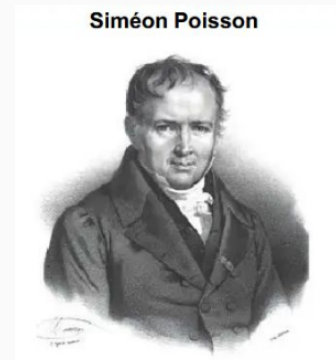
$$\hat{=} 10,32$$

Musí teda strieľať aspoň 11 krát.

22. Poissonovo rozdelenie: motivácia, modelovanie zriedkavých udalostí

Poissonovo rozdelenie: motivácia

- Predpokladajme, že v binomickom rozdelení je
 - **n** veľké (**veľa pokusov**)
 - **p** malé (**udalosť je zriedkavá**)
 - kvôli porovnateľnosti nech je očakávaný počet úspechov **np** rovnaký (označme ho **λ**)
- Pre pravdepodobnosti jednotlivých možností potom v limite platí:



$$\lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}$$

Základná myšlienka pri výpočte limity:

Prečo to platí:

- Upravíme pravdepodobnosti:

$$\begin{aligned} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} &= \frac{n(n-1) \cdots (n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \prod_{i=0}^{k-1} \left(1 - \frac{i}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{n-k}. \end{aligned}$$

- Využijeme limitu z matematickej analýzy:

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

Podrobnejšie:

$$\begin{aligned}
 \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} &= \frac{n(n-1)\dots(n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} = \\
 &= \frac{\lambda^k}{k!} \frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \dots \frac{n-(k-1)}{n} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} = \\
 &= \boxed{\frac{\lambda^k}{k!}} \underbrace{1 \cdot \left(1 - \frac{1}{n}\right)}_{\substack{\downarrow \\ 1}} \underbrace{\left(1 - \frac{2}{n}\right)}_{\substack{\downarrow \\ 1}} \dots \underbrace{\left(1 - \frac{k-1}{n}\right)}_{\substack{\downarrow \\ 1}} \cdot \boxed{\left(1 - \frac{\lambda}{n}\right)^n} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\substack{\downarrow \\ 1}} \\
 &\quad \text{lebo } \frac{1}{n} \rightarrow 0 \quad \quad \quad \boxed{e^{-\lambda}} \\
 &\rightarrow \frac{\lambda^k}{k!} e^{-\lambda} \text{ pre } n \rightarrow \infty
 \end{aligned}$$

23. Poissonovo rozdelenie

Poissonovo rozdelenie: $Po(\lambda)$

- λ je kladný parameter
- Náhodná premenná nadobúda ľubovoľný nezáporné celočíselné hodnoty s pravdepodobnosťami:

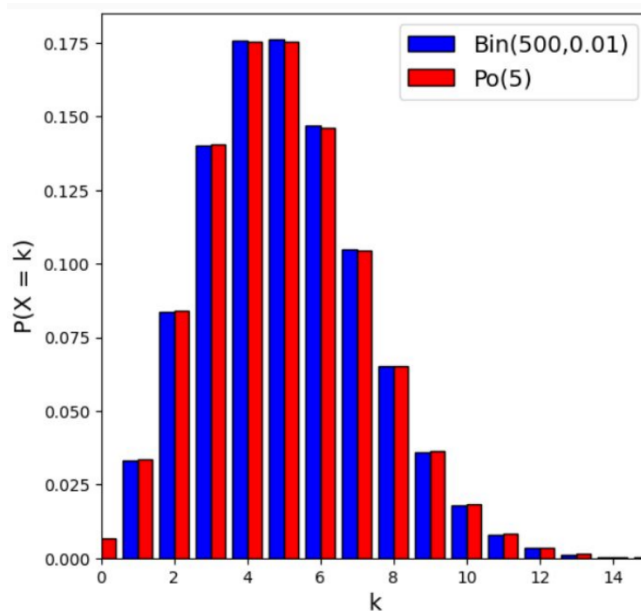
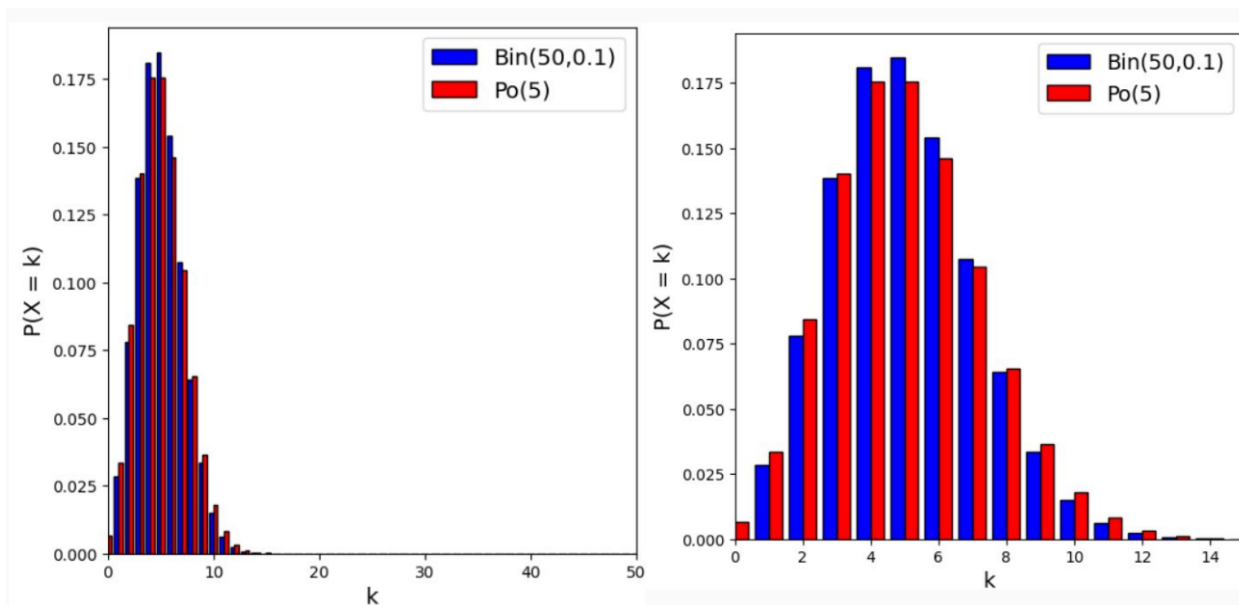
$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \in \{0, 1, 2, \dots\}$$

- Stredná hodnota a disperzia:

$$\mathbb{E}(X) = \lambda, \quad \mathbb{D}(X) = \lambda$$

- Stredná hodnota je rovnaká ako pre každé binomické rozdelenie, z ktorých sme počítali limitu pravdepodobností.
- Disperzia je limita disperzií týchto binomických rozdelení.
- Dajú sa spočítať aj priamo z predpisu pre $P(X=k)$ Poissonovho rozdelenia.

24. Binomické a Poissonovo rozdelenie - porovnanie kvality aproximácie



25. Poissonovo rozdelenie: príklad

Príklad 9

- Z dlhodobých štatistík vieme, že za jednu hodinu počas bežného pracovného dňa zaznamenaná server priemerne dve požiadavky o prístup.
- Predpokladajme teda, že počet požiadaviek o prístup má počas jednej hodiny Poissonovo rozdelenie so strednou hodnotou 2.
- Vypočítajte pravdepodobnosť, že sa počas hodiny o prístup
 - nepokúsi ani jeden používateľ
 - pokúsia aspoň dvaja používatelia
 - pokúsia aspoň traja používatelia

RIEŠENIE:

$$1) X \sim Po(2)$$

$$P(\text{ani jeden}) = P(X=0) = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \quad \text{pre } \lambda=2, k=0$$

$$= e^{-2} \cdot \frac{2^0}{0!} = e^{-2} \approx 0,1353$$

$$P(\text{aspoň 2}) = P(X \geq 2) = 1 - P(X=0) - P(X=1) =$$

$$= 1 - e^{-2} \cdot \frac{2^0}{0!} - e^{-2} \cdot \frac{2^1}{1!} =$$

$$= 1 - e^{-2} - 2e^{-2} = 1 - 3e^{-2} \approx 0,5940$$

$$P(\text{aspoň 3}) = P(X \geq 3) = 1 - P(X=0) - P(X=1) - P(X=2) =$$

$$= 1 - e^{-2} - 2e^{-2} - e^{-2} \cdot \frac{2^2}{2!} = 1 - 5e^{-2}$$

$$\approx 0,3233$$

26. Poissonovo rozdelenie: porovnanie dát s Poissonovým rozdelením

▪ How random was the 1998 World Cup?

Getting the Best from Teaching Statistics

Príklad 10

The average number of goals per game scored by the 32 teams in the group matches was 1.3125 (i.e. 48 matches yielded 126 goals: $126/96 = 1.3125$).

The actual distribution was as follows:

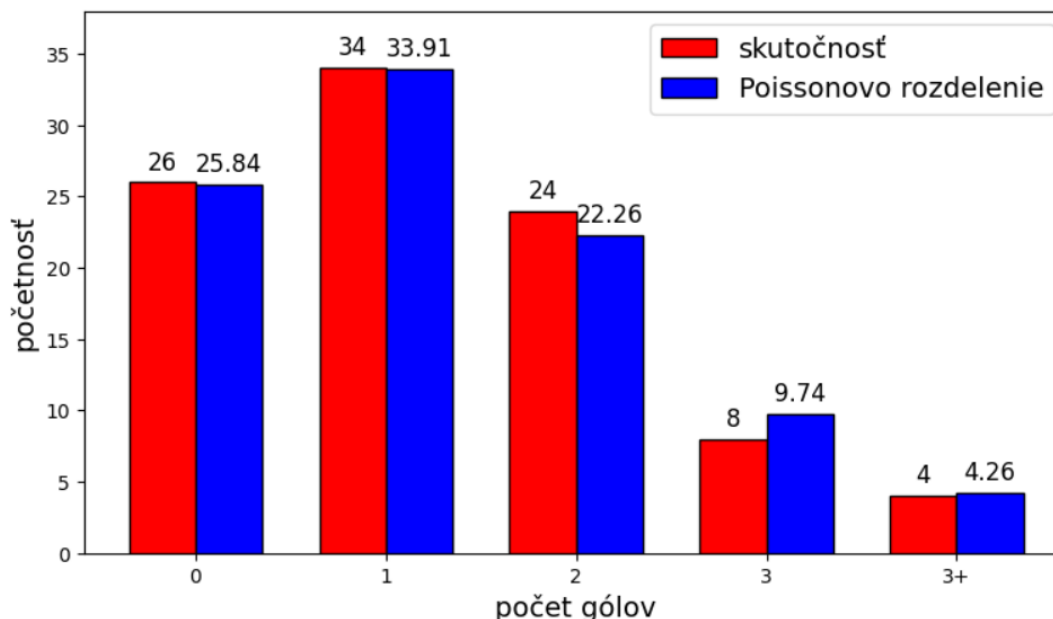
Goals	Frequency
0	26
1	34
2	24
3	8
4	1
5	2
6	1

A Poisson distribution with $m = 1.3125$ gives expected frequencies:

Goals	Frequency
0	25.8
1	33.9
2	22.3
3	9.7
>3	4.3

NOT A BAD FIT!!!

Parameter Poissonovho rozdelenia sa tu určil tak, aby sa stredná hodnota Poissonovho rozdelenia zhodovala s aritmetickým priemerom z dát. Takémuto spôsobu odhadovania parametrov rozdelenia sa hovorí **momentová metóda** (lebo stredná hodnota je tzv. prvý moment rozdelenia, disperzia druhý moment, dajú sa definovať aj momenty vyššieho rádu).



27. Hypergeometrické rozdelenie: motivácia

Hypergeometrické rozdelenie: motivácia

Čo už vieme - **binomické rozdelenie**:

- V triede je prítomných 12 študentov, na hodinu literatúry sa z pripravilo 7.
- Profesor sa pýta 5 otázok a študenta, ktorý má odpovedať, vyberá pre každú otázku hádzaním 12-stennou kockou.
- Koľkokrát bude vyvolaný študent, ktorý sa na hodinu pripravil?

Modifikácia:

- Profesor nevyberá študentov samosť pre každú otázku, ale náhodne vyberie **päťicu študentov**
- Koľko študentov z nich je pripravených?

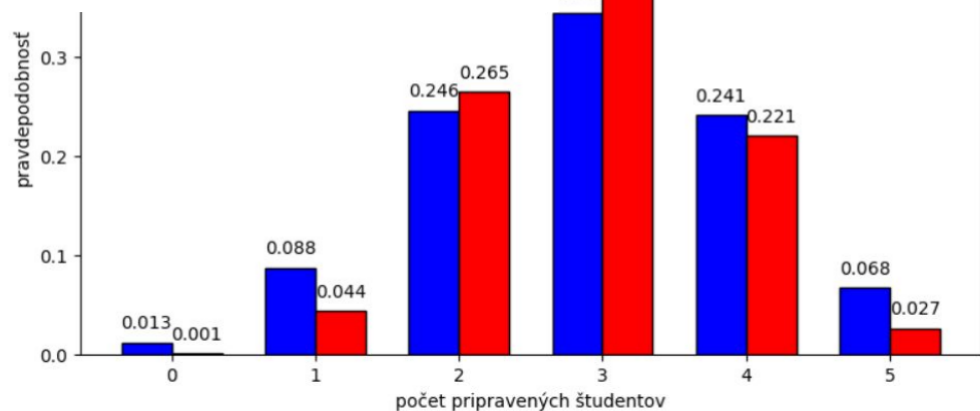


<https://www.ihrysko.sk/kocka-koplow-12k-elemental-p104043>

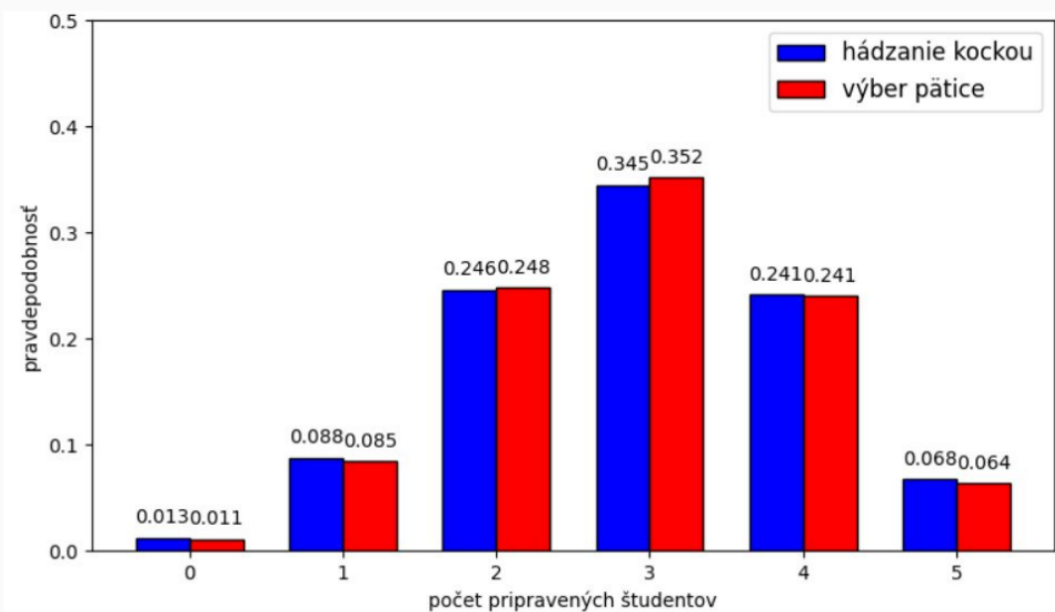
Nech **X** je počet pripravených študentov pri prvom prístupe a **Y** pri druhom:

$$\mathbb{P}(X = k) = \binom{5}{k} \left(\frac{7}{12}\right)^k \left(\frac{5}{12}\right)^{5-k}, \quad k \in \{0, 1, 2, 3, 4, 5\}$$

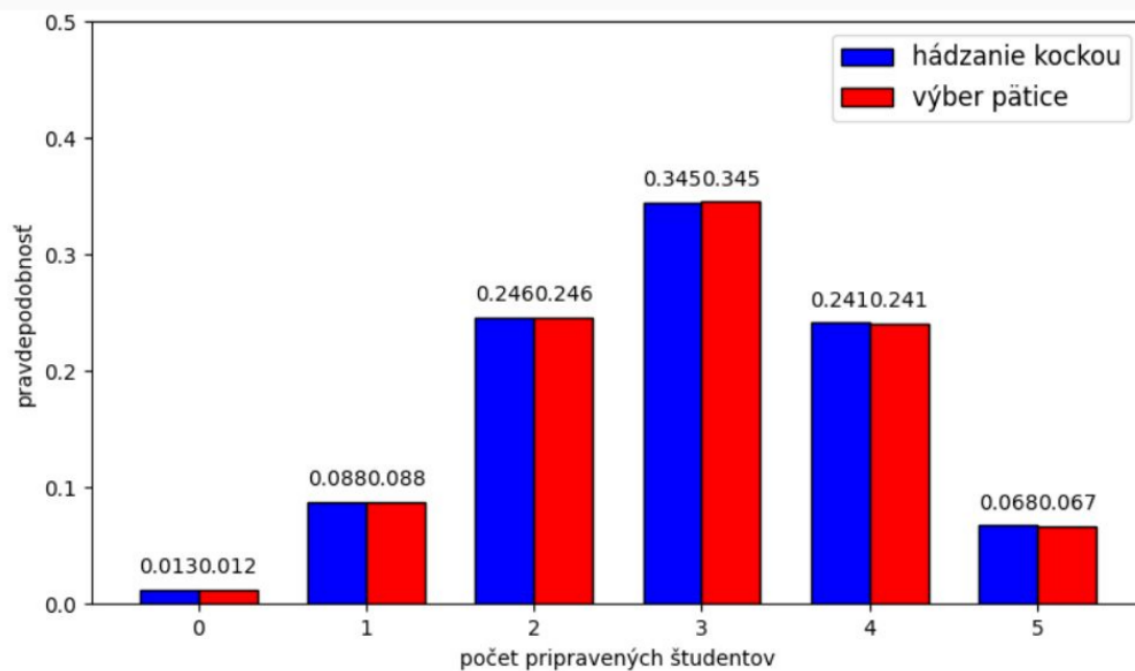
$$\mathbb{P}(Y = k) = \frac{\binom{7}{k} \binom{5}{5-k}}{\binom{12}{5}}, \quad k \in \{0, 1, 2, 3, 4, 5\}$$



Ak by študentov bolo 120 a z nich by bolo pripravených 70:



Ak by študentov bolo 1200 a z nich by bolo pripravených 700:



28. Hypergeometrické rozdelenie (výber bez návratu)

Hypergeometrické rozdelenie

- **N** - počet všetkých objektov
- **K** - počet objektov so sledovanou vlastnosťou
- **n** - počet vybraných objektov

$$\mathbb{E}(X) = n \cdot \frac{K}{N}$$

Rozdelenie náhodnej premennej **X**, vyjadrujúcej počet vybraných objektov so sledovanou vlastnosťou, sa nazýva hypergeometrické.

$$\mathbb{P}(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}, \quad k \in \{\max(0, n + K - N), \dots, \min(n, K)\}$$

- Binomické rozdelenie zodpovedá výberu s návratom, hypergeometrické výberu bez návratu, pri veľkom počte všetkých objektov **N** je rozdiel malý

29. Geometrické rozdelenie (čakanie na úspech)

Geometrické rozdelenie

- Robíme nezávislé pokusy, pravdepodobnosť úspechu je **p**.
- Rozdelenie náhodnej premennej **X**, vyjadrujúcej **počet pokusov, ktoré musíme spraviť, kým nastane prvý úspech** (vrátane toho posledného úspešného), sa nazýva geometrické.

$$\mathbb{P}(X = k) = (1 - p)^{k-1} p, \quad k \in \{1, 2, 3, \dots\}$$

- Stredná hodnota a disperzia:

$$\mathbb{E}(X) = \frac{1}{p}, \quad \mathbb{D}(X) = \frac{1 - p}{p^2}$$

Vzťah pre strednú hodnotu je intuitívne logický: Ak hádzeme kockou, pravdepodobnosť šestky je $\frac{1}{6}$ a zdá sa byť logické, že na padnutie šestky by sme mali v priemere potrebovať šesť pokusov.

30. Geometrické rozdelenie: príklad

Príklad 11

- Basketbalisti A a B sa striedajú pri hádzaní na kôš, až kým nenastane situácia, že jeden z nich v danom kole trafi a druhý netrafi. Ten, ktorý trafi, sa potom stáva víťazom súťaže.
- Basketbalista A zasiahne kôš s pravdepodobnosťou p_A a basketbalista B s pravdepodobnosťou p_B , obe pravdepodobnosti sú z otvoreného intervalu $(0, 1)$. Pokusy sú nezávislé.
- Náhodná premenná X označuje počet kôl v tejto súťaži. Nájdite pravdepodobnostné rozdelenie tejto náhodnej premennej a jej strednú hodnotu.

RIEŠENIE:

> úspech, na ktorý očakávame - jeden trafi, druhý nie

$$p = \underbrace{p_A}_{\substack{\uparrow \\ \text{A trafi}}} \underbrace{(1-p_B)}_{\substack{\uparrow \\ \text{B netrafi}}} + \underbrace{(1-p_A)}_{\substack{\uparrow \\ \text{A netrafi}}} \underbrace{p_B}_{\substack{\uparrow \\ \text{B trafi}}}$$

násobníme kvôli
nezávislosti

$X=k$ znamená: neúspech, neúspech, ..., neúspech, úspech
($k-1$) krát

$$P(X=k) = \underbrace{(1-p) \cdot (1-p) \cdots (1-p)}_{(k-1) \text{ krát}} \cdot p$$

$$= (1-p)^{k-1} p \quad \text{pre } k=1, 2, 3, \dots$$

$$E(X) = \frac{1}{p} \quad \text{lebo ide o geometrické rozdelenie}$$

Pri výpočte pravdepodobností vychádzame zo základných princípov, je zbytočné pamätať si vzorec. Pri strednej hodnote si spomenieme na logické zdôvodnenie z geometrického rozdelenia.

31. Geometrické rozdelenie (iná parametrizácia)

Poznámka

- Niekedy sa geometrické rozdelenie definuje ako **počet neúspechov pred prvým úspechom**.
- V našom značení túto hodnotu vyjadruje **$X - 1$** .
- V Pythone:
 - Geometrické rozdelenie sa defaultne uvažuje v zmysle našej pôvodnej definície
 - Túto modifikáciu môžeme získať parametrom **`loc=-1`**

```
from scipy.stats import geom
import numpy as np
```

```
k_hodnoty = np.arange(0, 5)
print(k_hodnoty)
```

```
[0 1 2 3 4]
```

```
p1 = geom.pmf(k_hodnoty, 0.5)
print(p1)
```

```
[0.         0.5       0.25      0.125     0.0625]
```

```
p2 = geom.pmf(k_hodnoty, 0.5, loc=-1)
print(p2)
```

```
[0.5       0.25      0.125     0.0625   0.03125]
```

