

PaŠ pre informatikov (cv 02)

Beáta Stehlíková, FMFI UK Bratislava

Opakovanie z prednášky: geometrická pravdepodobnosť (bod 18 v texte)

Geometrická pravdepodobnosť

- Ak má množina všetkých možných výsledkov má nekonečne veľa prvkov, nedá sa použiť klasická pravdepodobnosť.
- Pri "rovnomernosti" možných výsledkov na množine Ω sa dá použiť geometrická pravdepodobnosť: **$P(A) = \text{plocha}(A)/\text{plocha}(\Omega)$**

Príklad.

Na Zem dopadne meteorit. Je malý, takže ho v porovnaní s povrchom Zeme budeme považovať za bod. O mieste jeho dopadu nič nevieme, pravdepodobnosť, že padne na určitú územie je teda úmerná jeho ploche. Aká je pravdepodobnosť, že dopadne na súš?

Wikipédia: Voda na Zemi zaberá asi 2/3 celkového povrchu.

Príklad 1: Geometrická pravdepodobnosť I. (bod 19 v texte)

V štvorci so stranou dĺžky 1 zvolíme bod tak, že pravdepodobnosť jeho padnutia do nejakého útvaru je úmerná jeho obsahu (budeme hovoriť “bod zvolíme v tomto štvorci “rovnomerne náhodne”). Aká je pravdepodobnosť, že jeho vzdialenosť od stredu štvorca bude menej ako 0,5?

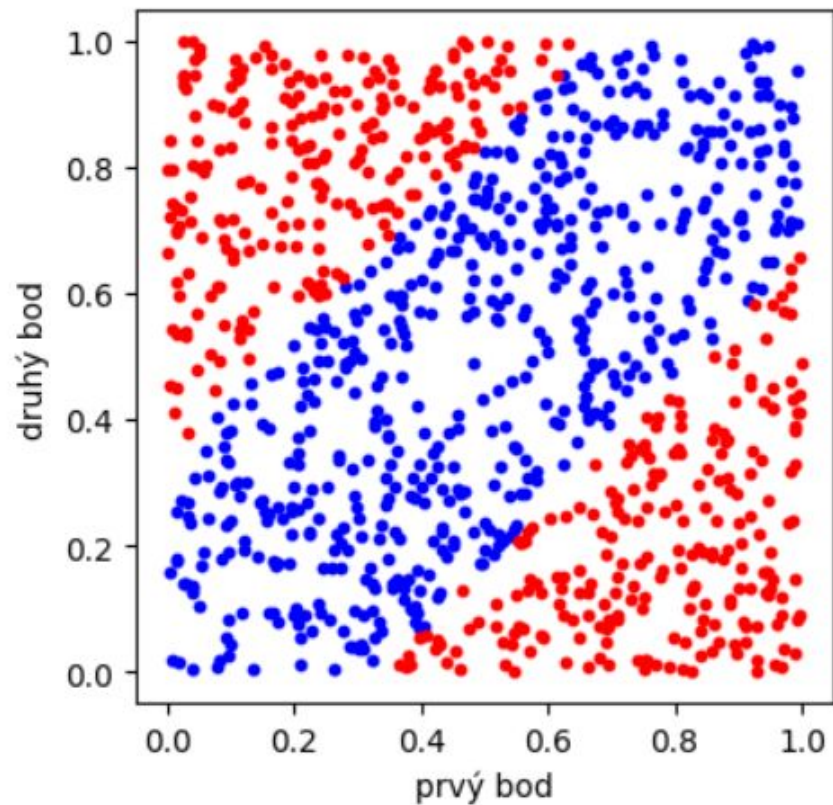
```
import numpy as np

def bod_v_stvorci():
    x = np.random.rand() # rovnomerne medzi 0 a 1
    y = np.random.rand()
    vzdialenost_od_stredu = np.sqrt((x - 0.5)**2 + (y - 0.5)**2)
    return vzdialenost_od_stredu < 0.5

pocet_pokusov = 10**5
simulacia = [bod_v_stvorci() for _ in range(pocet_pokusov)]
print(sum(simulacia)/pocet_pokusov)
```

0.78548

Príklad 2: Geometrická pravdepodobnosť II. (bod 20 v texte)



Na úsečke dĺžky 1 sú náhodne zvolené dva body (každý *rovnomerne náhodne* na celej dĺžke úsečky, nezávisle jeden na druhom). Tieto dva body vytvoria úsečku. Aká je pravdepodobnosť, že jej dĺžka je menej ako $1/3$?

Príklad 3: Geometrická pravdepodobnosť III. (bod 21 v texte, prvé zadanie)

Romeo a Júlia sa chcú stretnúť. Na miesto stretnutia prídu v náhodnom čase, ktorý si každý nezávisle zvolí *rovnomerne náhodne* medzi 18:00 a 20:00. Kto príde prvý, počká na druhého. Aká je pravdepodobnosť, že ten, kto príde prvý, bude čakať viac ako 15 minút?

Príklad 4 (body na úsečke)

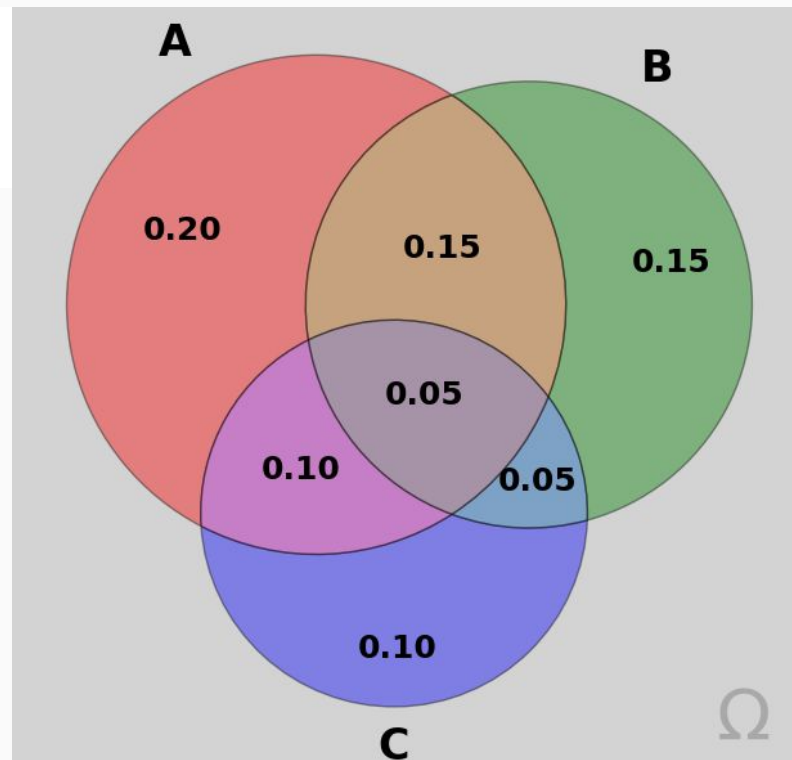
- Na úsečke dĺžky 1 sú nezávisle *rovnomerne náhodne* zvolené dva body.
- Tieto body rozdelia úsečku na tri časti.
- Aká je pravdepodobnosť, že žiadna z týchto častí nebude dlhšia ako $\frac{3}{4}$?

Príklad 5 (križovatky)

A: na prvej križovatke svieti červená
B: na druhej križovatke svieti červená
C: na tretej križovatke svieti červená

- Na prvej križovatke bola červená. Aká je pravdepodobnosť, že červená bude aj na druhej križovatke?
- Sú udalosti *A* a *B* nezávislé?
- Sú nezávislé udalosti *B* a *C*?

Nezávislosť udalostí *M* a *N*:
 $\mathbb{P}(M \cap N) = \mathbb{P}(M)\mathbb{P}(N)$



Príklad 6 (HAPPY HOUR)

- ▶ Dvaja opití kamaráti vychádzajú z krčmy, ktorá má pri vchode nápis HAPPY HOUR.
- ▶ Jeden z nich náhodne odlepí dve písmená (s rovnakou pravdepodobnosťou pre každú dvojicu).
- ▶ Druhý ich prilepí naspäť, ale náhodne, nepozerá sa, či ich vracia správne.
- ▶ Aká je pravdepodobnosť, že nápis HAPPY HOUR zostane zachovaný?



Príklad 7: Naivný Bayesov klasifikátor (republikáni a demokrati, bod 14 v texte)

R: United States Congressional Voting Records 1984 ▾

Find in Topic

HouseVotes84 {mlbench}R Documentation

United States Congressional Voting Records 1984

Description

This data set includes votes for each of the U.S. House of Representatives Congressmen on the 16 key votes

Do Pythonu (po vynechaní NA) pomocou `rpy2.objects` → 281 riadkov

	Class	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
2	republican	n	y	n	y	y	y	n	n	n	n	n	y	y	y	n
6	democrat	n	y	y	n	y	y	n	n	n	n	n	n	y	y	y
9	republican	n	y	n	y	y	y	n	n	n	n	n	y	y	y	n
20	democrat	y	y	y	n	n	n	y	y	y	n	y	n	n	n	y
24	democrat	y	y	y	n	n	n	y	y	y	n	n	n	n	n	y

Na tréningových dátach vytvoríme model, predikujeme **Class** (republikán, demokrat) pomocou hlasovaní 1, 8, 15

```
from sklearn.model_selection import train_test_split

# X - premene ktore pouzijeme na predikovanie
# y - premenna ktoru budeme predikovat
X = votes[["V1", "V8", "V15"]]
y = votes["Class"]

# train/test
X_train, X_test, y_train, y_test = train_test_split(
    X, y,
    test_size=0.2,
    stratify=y,
    random_state=42
)
```

Potom predikujeme hodnoty v testovacej časti, napríklad:

```
print(X_test.head())
```

	V1	V8	V15
266	y	y	n
254	n	n	n
255	y	y	n
341	n	n	n
238	y	y	n

Z tabuliek početností v tréningových dátach spočítame pravdepodobnosti ručne a overíme vo výstupe z Pythonu.

```
=== Class ===  
Class  
democrat      132  
republican    92  
Name: count, dtype: int64
```

```
=== Class vs. V1 ===  
V1           n    y  
Class  
democrat    54   78  
republican  76   16
```

```
=== Class vs. V8 ===  
V8           n    y  
Class  
democrat    26  106  
republican  79   13
```

```
=== Class vs. V15 ===  
V15          n    y  
Class  
democrat    49   83  
republican  82   10
```

Na kontrolu:

	democrat	republican
0	0.920245	0.079755
1	0.063564	0.936436
2	0.920245	0.079755
3	0.063564	0.936436
4	0.920245	0.079755

Zhodnotenie modelu:

