5. cvičení z PaSti – 2022-03-15

Tahák

- *Střední hodnota* veličiny X: $\mathbb{E}(X) = \sum_t t \cdot P(X = T)$.
- *Linearita střední hodnoty:* $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$, $\mathbb{E}(\alpha X) = \alpha \mathbb{E}(X)$.
- *Rozbor případů:* $\mathbb{E}(X) = \mathbb{E}(X \mid A) \cdot P(A) + \mathbb{E}(X \mid A^c) \cdot P(A^c)$.
- *LOTUS (pravidlo naivního statistika):* $\mathbb{E}(g(X)) = \sum_a g(a) \cdot P(X = a)$.
- *Indikátorová náhodná veličina* I_A jevu A nabývá hodnot 0 a 1, přičemž $I_A = 1 \Leftrightarrow A$ nastane. Platí, že $\mathbb{E}(I_A) = P(A)$.
- *Rozptyl* $\mathrm{var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.
- Veličiny X a Y jsou *nezávislé* $\Leftrightarrow P(X = a \wedge Y = b) = P(X = a) \cdot P(Y = b)$.
- Pro X, Y nezávislé platí $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Na zahřátí

1. Mějme neorientovaný graf $G = (V, E)$. Vybereme náhodně podmnožinu vrcholů $V' \subseteq V$ tak, že pro každý vrchol si hodíme mincí, zda ho dáme do V'. Jaká je střední hodnota počtu hran, které vedou mezi V' a $V \setminus V'$?

2. Hodíme třemi kostkami a výsledek přečteme jako tříciferné číslo. Jaká je jeho střední hodnota?

3. Hodíme dvěma kostkami a výsledek přečteme jako dvojciferné číslo, přičemž kostku s větší hodnotou použijeme pro řád desítek. Jaká ja teď střední hodnota?

Zacházení se střední hodnotou a rozptylem

4. Nechť X, Y jsou diskrétní n.v. a $\alpha \in \mathbb{R}$.

(a) Ukažte, že $\mathrm{var}(X + \alpha) = \mathrm{var}(X)$.

(b) Vyjádřete $\mathrm{var}(\alpha X)$ pomocí $\mathrm{var}(X)$.

(c) Ukažte, že $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y)$, pokud X, Y jsou nezávislé.

5. Dokažte:

(a) Pokud $\mathbb{E}(X^2) = 0$, tak $P(X = 0) = 1$.

(b) Předpokládejme, že $\mathrm{var}(X) = 0$, dále že $\mathbb{E}(X)$ existuje a je konečná.
Pak $X = \mathbb{E}(X)$ skoro jistě, neboli $P(X = \mathbb{E}(X)) = 1$.

Podmíněná střední hodnota

6. V testu je 20 otázek s volbami $\{\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}\}$, vždy je správná právě jedna odpověď. Za správnou odpověď dostanete 1 bod, za špatnou $-1/4$ bodu, za nevyplněnou otázku nula. Každá otázka je s pravděpodobností p jednou z těch, co se Kvído naučil, a tedy zná správnou odpověď. Pokud správnou odpověď nezná, ví o tom a může se rozhodnout, zda tipovat.

(a) Jaká je střední hodnota počtu bodů, které Kvído získá, pokud bude odpovídat jenom na otázky, u kterých zná odpověď?

(b) A co když bude tipovat, když nezná správnou odpověď?

(c) Jak by se musela změnit penalizace za chybnou odpověď, aby byly odpovědi v částech a, b stejné?

7. Házíme běžnou kostkou, při šestce házíme znovu, a to i opakovaně. Spočítejte střední hodnotu součtu všech hozených čísel.

Nezávislost

8. Ukažte, že jevy A, B jsou nezávislé, právě když jsou nezávislé jejich indikátorové veličiny.

9. Ukažte, že pro diskrétní nezávislé n.v. X, Y platí

$$P(X \leq x \wedge Y \leq y) = P(X \leq x) \cdot P(Y \leq y).$$

Pro jednoduchost můžete předpokládat, že $\mathrm{Im}(X) = \mathrm{Im}(Y) = \{1, 2, \ldots, n\}$ pro nějaké n.

Bonusy

10. Velký hon: n myslivců se snaží ulovit n lišek. Míří náhodně: každý myslivec střefí rovnoměrně náhodně vybranou lišku. Spočítejte střední počet lišek, které hon přežijí (nikdo je netrefil).

11. Mějme náhodnou permutaci π na množině $\{1, \ldots, n\}$. Prvek i je *levým maximem*, pokud $\pi(j) < \pi(i)$ pro všechna $j < i$. Jaká je střední hodnota počtu levých maxim? Analogie: různě vysocí lidé stojí ve frontě, kolik z nich vidí na začátek fronty?

12. Připomeňte si definici indikátorové náhodné veličiny I_A.

(a) Nechť $A = A_1 \cup A_2 \cup \cdots \cup A_n$. Ověřte rovnost

$$1 - I_A = \prod_{i=1}^{n} (1 - I_{A_i}).$$

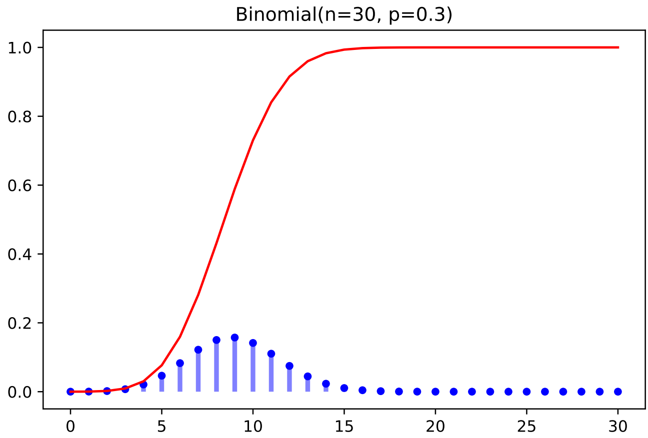
(b) Roznásobte a použijte větu o linearitě střední hodnoty, abyste získali princip inkluze a exkluze.

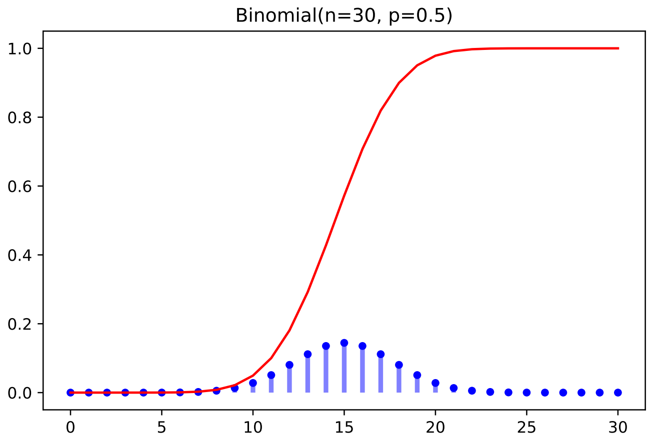
Pravděpodobnostní rozdělení v Pythonu

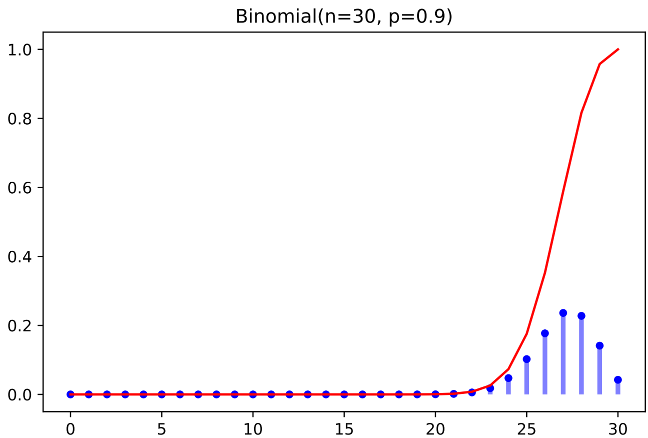
```
import scipy.stats as stats

# Pořídíme si binomiální rozdělení s danými parametry
n = 20
p = 0.3
b = stats.binom(n, p)

# Necháme si vygenerovat náhodný vzorek
print('RVS:', b.rvs(size=20))

# Pravděpodobnostní funkce (Probability Mass Function)
print('PMF:', [b.pmf(x) for x in range(0, n+1)])

# Distribuční funkce (Cummulative Distribution Function)
print('CDF:', [b.cdf(x) for x in range(0, n+1)])

# Kvantily ("inverze" k CDF, Percent Point Function)
print('Quantile 0.1:', b.ppf(0.1))
print('Quantile 0.9:', b.ppf(0.9))

# Střední hodnota, rozptyl a směrodatná odchylka
print('Mean:', b.mean())
print('Variance:', b.var())
print('Std deviation:', b.std())

# Také se jde přímo zeptat na vlastnosti rozdělení s konrétními
# parametry, aniž bychom ho explicitně konstruovali.
print(stats.binom.rvs(n, p, size=20))
```

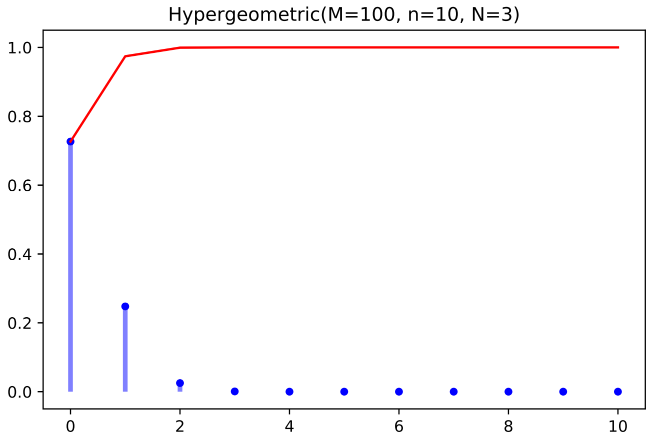
Kreslíme obrázky

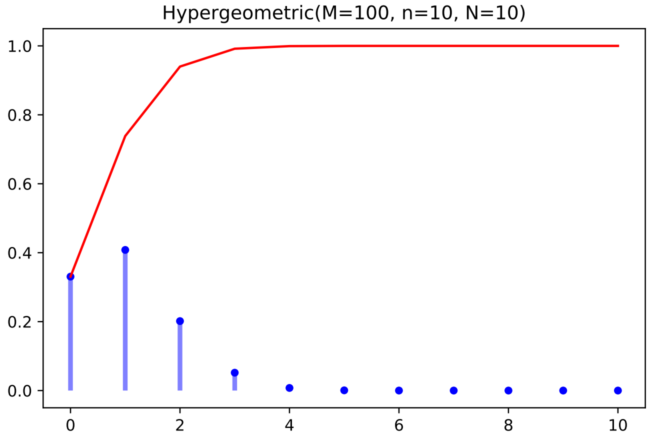
```
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt

# Rozdělení
n = 20
p = 0.3
b = stats.binom(n, p)

# Jeden obrázek
fig, ax = plt.subplots(1, 1)
x = np.arange(0, n+1)
ax.plot(x, b.pmf(x), 'bo', ms=8, label='binom pmf')
ax.vlines(x, 0, b.pmf(x), colors='b', lw=5, alpha=0.5)
ax.legend()
plt.show()

# Nebo více obrázků pohromadě
fig, ax = plt.subplots(1, 2)
x = np.arange(0, n+1)
ax[0].set_ylim(0, 1)
ax[0].plot(x, b.pmf(x), 'bo', ms=8, label='binom pmf')
ax[0].vlines(x, 0, b.pmf(x), colors='b', lw=5, alpha=0.5)
ax[0].legend()
ax[1].plot(x, b.cdf(x), 'r', label='binom cdf')
ax[1].legend()
plt.show()
```

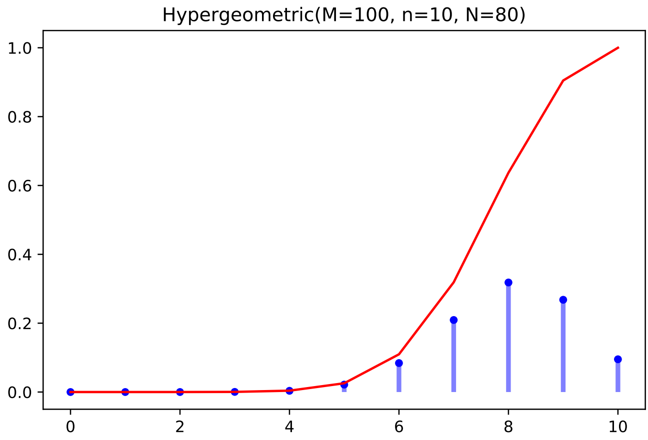

Binomial(n=30, p=0.3)
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30

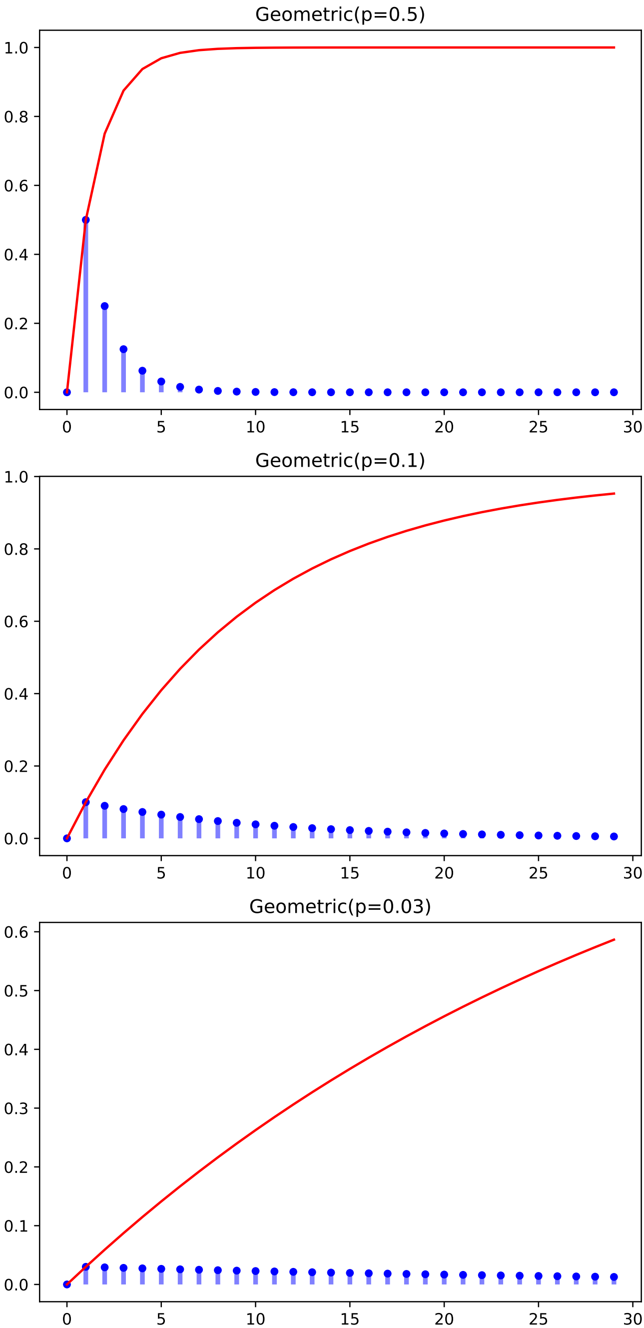

Binomial(n=30, p=0.5)
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30

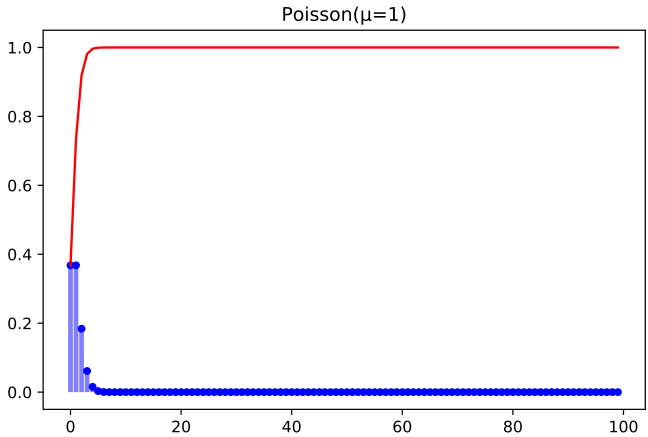

Binomial(n=30, p=0.9)
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30

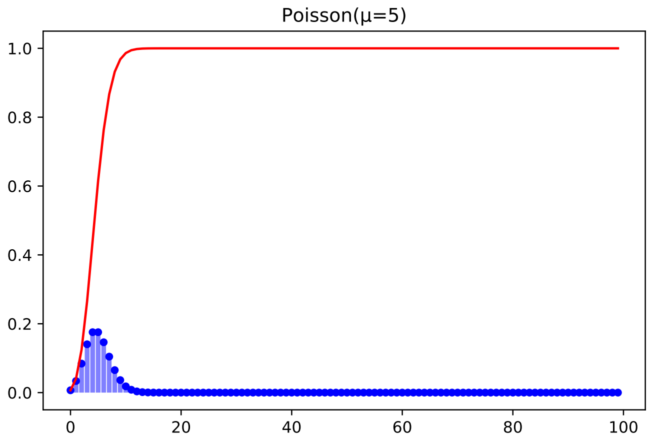

Hypergeometric(M=100, n=10, N=3)
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10

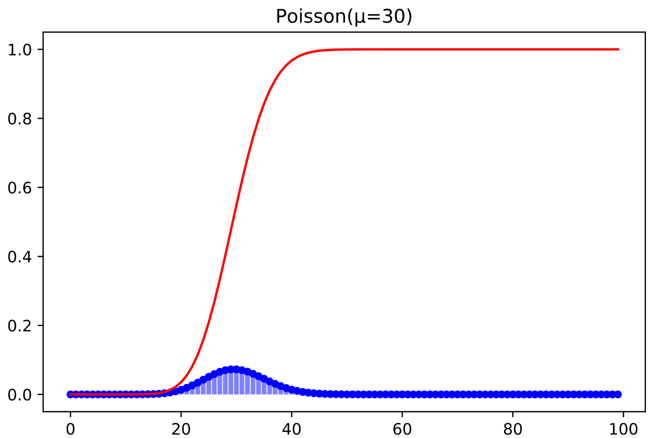

Hypergeometric(M=100, n=10, N=10)
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10


Hypergeometric(M=100, n=10, N=80)
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10


Geometric(p=0.5)
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
Geometric(p=0.1)
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
Geometric(p=0.03)
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
5
10
15
20
25
30


Poisson(μ=1)
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100


Poisson(μ=5)
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100


Poisson(μ=30)
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100