4. cvičení z PaSti – 2021-03-24

Z každé kapitolky (mimo Bonusy) vyřešte aspoň jeden příklad.

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 trojciferné 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., $\alpha \in \mathbb{R}$.

(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?

Nezávislost

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

8. 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.

Náhodné vektory

Sdružená pravděpodobnostní funkce je definována vztahem

$$p_{X,Y}(x, y) = P(X = x \wedge Y = y).$$

Připomeňte si, jak z ní zjistit „jednorozměrné funkce" p_X, p_Y. (Těm se říká *marginální rozdělení* – představujte si je na *okrajích* tabulky sdruženého rozdělení.)

9. Ze standardního balíčku s 52 kartami vytáhneme dvě karty. Označíme X počet vytažených es, Y počet králů. Určete sdruženou pravděpodobnostní funkci $p_{X,Y}$ a také marginální pstní funkce p_X, p_Y.

10. Označme X_1, X_2, X_3 výsledky tří nezávislých hodů čtyřstěnnou kostkou (s čísly 1, ..., 4).

(a) Jaká je pravděpodobnostní funkce $X = X_1$?

(b) Jaká je pravděpodobnostní funkce $Y = \max(X_1, X_2)$?

(c) Jaká je pravděpodobnostní funkce $Z = \max(X_1, X_2, X_3)$?

(d) O kolik se zvýší střední hodnota tím, že můžeme házet třikrát? Neboli o kolik je vyšší $\mathbb{E}(Z)$ než $\mathbb{E}(X)$?

Nápověda: Určete napřed $P(Y \leq k)$, $P(Z \leq k)$.

11. Nezávislé n.v. X_1, ..., X_n mají geometrické rozdělení s parametry p_1, ..., p_n. Jaké je rozdělení $\min(X_1, \ldots, X_n)$?

12. Na kostce padne číslo i s pravděpodobností p_i pro $i = 1, \ldots, 6$. Hodíme n-krát a označíme X_i počet hodů, kdy padlo i.

(a) Najděte sdruženou pravděpodobnostní funkci pro n.v. X_1, ..., X_n.

(b) Jaké je marginální rozdělení, tj. rozdělení jednotlivých n.v. X_i?

Bonusy

13. Připomeňte si definici indikátorové náhodné veličiny I_A.

(a) Jaká je $\mathbb{E}(I_A)$?

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

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

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

14. Označme M počet emailů, které dostaneme za den, S počet spamů mezi nimi, H počet „hamů" – těch, co nejsou spamy. Předpokládejme, že $M \sim Pois(\lambda)$ a že každý email má nezávisle na ostatních pravděpodobnost p, že je to spam.

(a) Vyjádřete $P(S = k)$ (jako nekonečnou sumu) pomocí sdruženého rozdělení M a S.

(b) Odvoďte, že $S \sim Pois(p\lambda)$.

(c) Odvoďte, že $H \sim Pois((1-p)\lambda)$ a také, že H, S jsou nezávislé n.v.

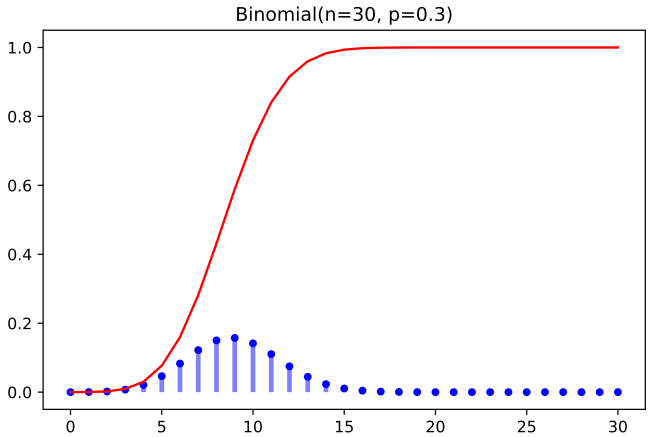
15. V tipovací hře má soutěžící na výběr n otázek, ze kterých si může postupně vybírat. U i-té otázky s pravděpodobností p_i odpoví správně, získá za to h_i korun a právo dalšího výběru. Pokud neodpoví správně, končí. Předpokládejme, že cílem je maximalizovat střední hodnotu zisku. Ukažte, že toho docílí, bude-li vybírat otázky seřazené podle hodnoty $\frac{p_i h_i}{1-p_i}$.

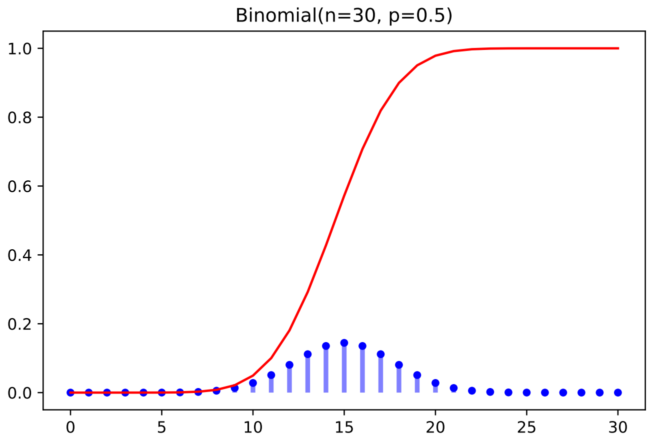
Pravděpodobnostní rozdělení v Pythonu

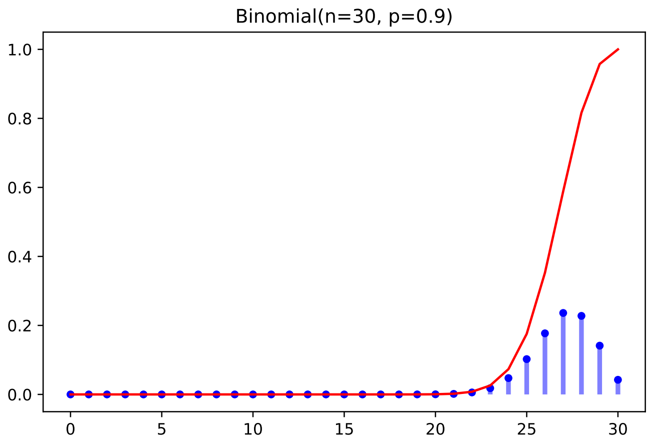
```
import scipy.stats as stats

# 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))
```

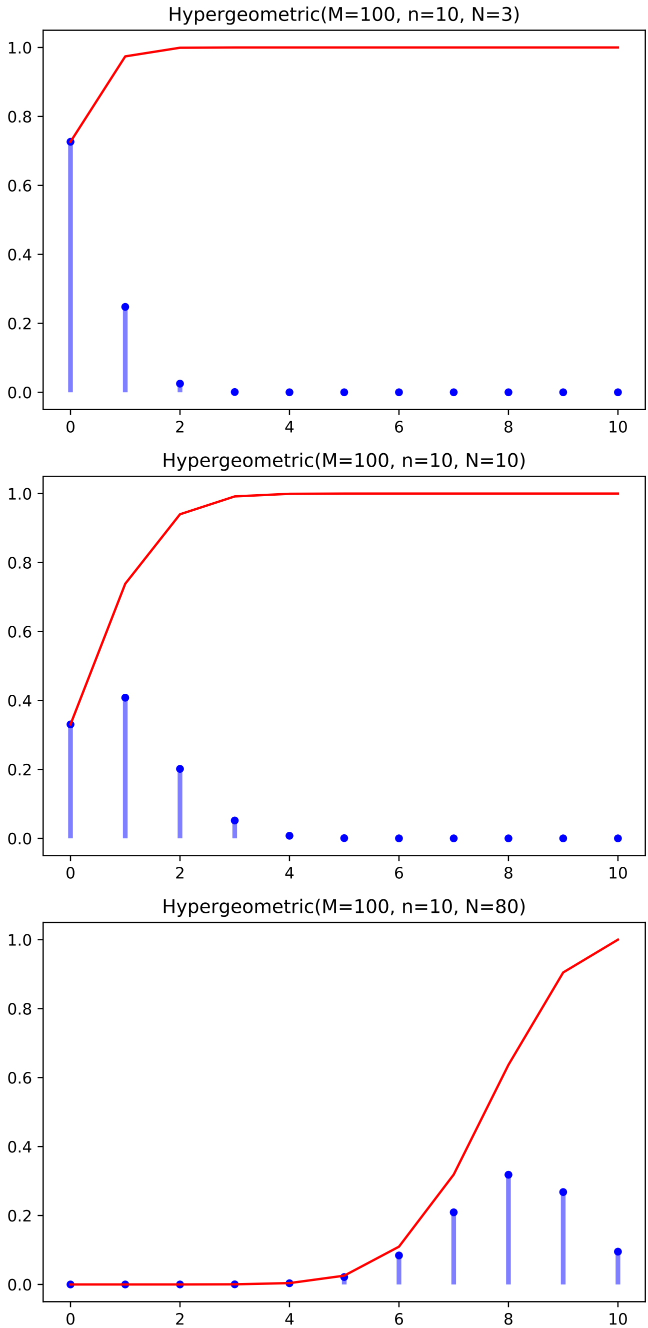
Kreslíme obrázky

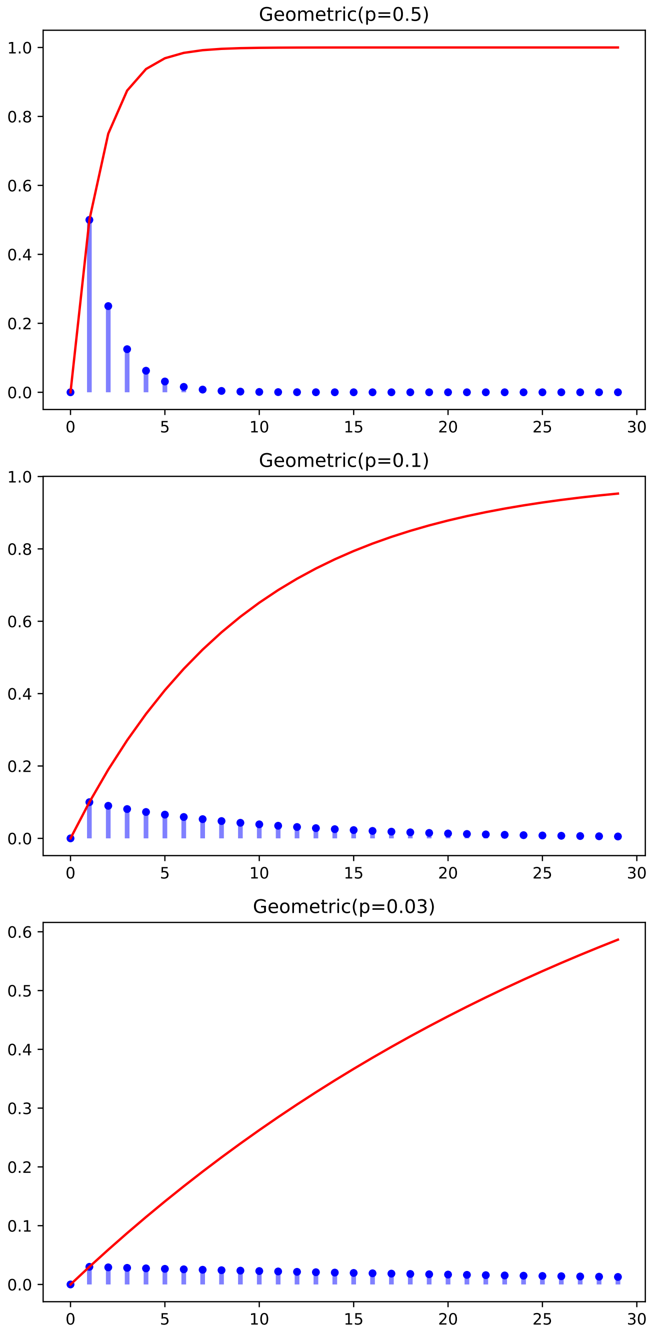
```python
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt

# 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()
```

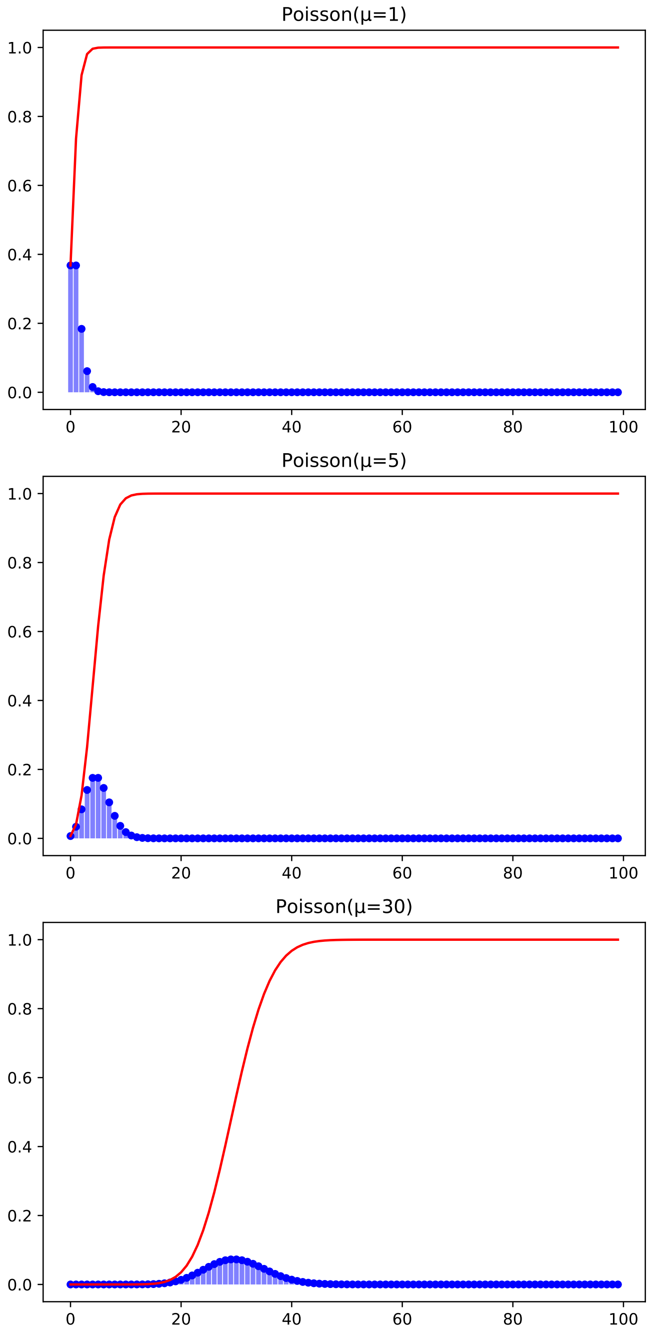

Binomial(n=30, p=0.3)
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30


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


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


Hypergeometric(M=100, n=10, N=3)
Hypergeometric(M=100, n=10, N=10)
Hypergeometric(M=100, n=10, N=80)
0.0
0.2
0.4
0.6
0.8
1.0
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)
Poisson(μ=5)
Poisson(μ=30)
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100