

1. cvičení z PaSti – 2021-03-03

Na rozehrání

1. Házíme 10 mincemi (neřekneme-li jinak, vždy myslíme poctivými a rozlišitelnými). Jaká je pravděpodobnost, že na aspoň dvou padne vorel? Jak vypadá pravděpodobnostní prostor?

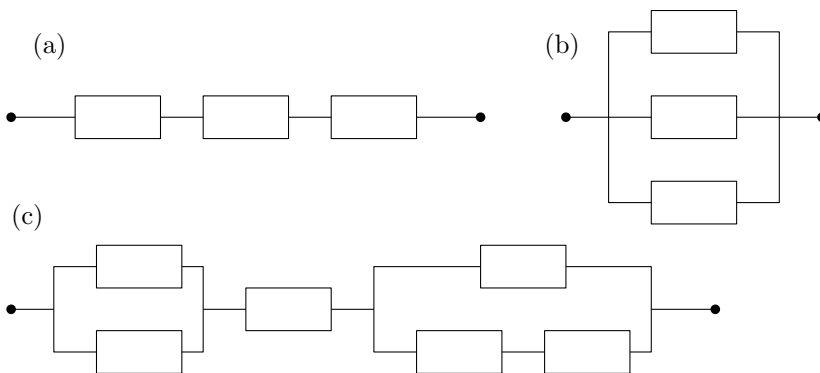
2. Náhodně uspořádáme množinu $\{1, \dots, n\}$. Jaká je pravděpodobnost, že na prvním místě je 1? Jak vypadá pravděpodobnostní prostor?

3. Házíme kostkou, při šestce házíme znovu. Jaká je pravděpodobnost, že hodíme nejvýše třikrát? Jak vypadá pravděpodobnostní prostor?

4. Dokažte, že $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

5. Chceme spravedlivě rozlosovat mezi dvěma lidmi, ale máme jen cinknutou minci, kde padá vorel s pravděpodobností $p \in (0, 1)$. Jak to udělat? Co když p neznáme?

6. Každý obdélník na obrázku je součástka, která se může porouchat s pravděpodobností p . Přesněji řečeno porucha znamená, že skrz součástku neteče proud. Poruchy součástek jsou na sobě nezávislé. Jaká je pravděpodobnost, že stále poteče proud mezi dvěma puntíky?



Podmíněná pravděpodobnost

7. Hodíme korunou a dvoukorunou, na každé z nich padá panna s pravděpodobností p .

(a) Jaká je pravděpodobnost, že na obou padne panna, pokud padla panna na korunu?

(b) Jaká je pravděpodobnost, že na obou padne panna, pokud padla panna aspoň na jedné minci?

Je bez počítání jasné, která pravděpodobnost je větší?

8. Král země má právě jednoho sourozence. Jaká je pravděpodobnost, že král má bratra? (Ujasněte si všechny předpoklady, které používáte! Co jsou elementární jevy?)

9. Král udělí milost dvěma ze tří vězňů (a vybere náhodně). Jednomu z nich dozorce nabídl, že mu řekne jméno jednoho z druhých dvou vězňů, který bude propuštěn. Vězeň ale odmítl: pak bych měl pravděpodobnost propuštění jen $1/2$, teď ji mám $2/3$. Má pravdu?

Bonusy

10. (*Bertrandův paradox*) Rovnostrannému trojúhelníku se stranou 1 opíšeme kružnici, na té vybereme náhodnou tětivu. Jaká je pravděpodobnost, že délka tětivy je větší než 1?

11. Na stole leží dvě obálky, o kterých víme, že v každé z nich je nějaký (nenulový, celočíselný) počet stokorun, v obou jiný. Máme dovoleno jednu obálku otevřít a pak se rozhodnout, zda si necháme tu, nebo tu druhou. Pokud chceme získat obálku s vyšším obnosem, můžeme ji získat s pravděpodobností větší než $1/2$? (Nápověda na poslední stránce.)

12. Máme k nádob, v každé z nich a bílých a b černých míčků. Z první vybereme náhodný míček, vhodíme do druhé. Pak z ní vybereme náhodný míček, vhodíme do třetí, atd. Jaká je pravděpodobnost, že z poslední nádoby vytáhneme bílý míček?

13. V urně je a černých a b bílých míčků. Postupně z ní (bez vracení) taháme míčky. Jaká je pravděpodobnost, že první vytažený míček je černý? Druhý, třetí, ...?

Pro procvičení

14. Házíme mincí, dokud nepadne orel, na každý hod dostaneme novou minci. (Jaká je pravděpodobnost, že získáme k mincí?) Pak všechny získané mince hodíme najednou, pokud na každé z mincí padne orel, můžeme si je všechny nechat. Jaká je pravděpodobnost, že se to stane?

15. Hodíme dvakrát kostkou. Označíme

- SD jev „součet hozených čísel je 10“,
- PS jev „první hod padla šestka“ a
- NS jev „v některém hození padla šestka“ (možná v obou).

Spočítejte pravděpodobnosti daných jevů a všechny podmíněné pravděpodobnosti.

16. V krabici se 100 bateriemi jsou čtyři vybité.

- Jaká je pravděpodobnost, že když vybereme náhodně tři z nich do čelovky, tak bude fungovat? (Musí být všechny v pořádku.)
- Jaká je pravděpodobnost, že budou aspoň dvě ze tří v pořádku?
- Třikrát vytáhneme náhodnou baterii, změříme a (protože je v pořádku) tak ji hodíme zpět. S jakou pravděpodobností se to stane?

Bez počítání: je vyšší pravděpodobnost (a) nebo (c)?

17. Označme Z jev „Kvído si během přístích dvanácti měsíců nastaví automatické zálohování důležitých souborů“ a D jev „Kvído během přístích dvanácti měsíců přijde o nějaký důležitý soubor“.

- (a) Myslíte, že je větší $P(Z \mid D)$ nebo $P(Z \mid D^c)$?
- (b) Myslíte, že je větší $P(D \mid Z)$ nebo $P(D \mid Z^c)$?
- (c) Podle definice ověřte, zda váš tip je matematicky možný.
- (d) Zamyslete se nad tím, proč.

Jak experimentovat v Pythonu

```
# https://docs.python.org/3/library/random.html
# Nepoužívejte pro šifrování!
import random

# https://docs.python.org/3/library/itertools.html
import itertools

# Generuje náhodné celé číslo 1 <= x <= 10:
x = random.randint(1, 10)
print(x)

# Alternativa, pokud máte rádi pythoní range(...):
x = random.randrange(1, 11)
print(x)

# Generuje náhodný prvek dané neprázdné posloupnosti:
drink = random.choice([
    "světlé pivo",
    "tmavé pivo",
    "slivovice",
    "zelený čaj",
    "černý čaj",
    "bílý čaj",
    "mléko",
])
print(drink)

# Výsledek 'B' je třikrát pravděpodobnější než 'A':
print(random.choice(['A', 'B', 'B', 'B']))

# Vrátí seznam k náhodných prvků z dané sekvence (mohou se opakovat):
print(random.choices(['A', 'B', 'C'], k=5))
```

```

# Také můžeme určit váhy jednotlivých prvků:
print(random.choices(['A', 'B', 'C'], weights=[1, 1, 4], k=5))

# Výběr bez opakování:
print(random.sample(['w', 'x', 'y', 'z'], k=2))
print(random.sample(['w', 'x', 'y', 'x'], k=2)) # dvě různá 'x'

# Náhodná permutace:
my_list = ['a', 'b', 'c', 'd']
print(random.sample(my_list, k=len(my_list)))

# Destruktivní náhodná permutace:
random.shuffle(my_list)
print(my_list)

# Inicializace počátečního stavu generátoru
random.seed(42)

# Nebo můžeme chtít pokaždé jiný počáteční stav
random.seed()

# Generování všech možností (vrací iterable):
print(list(itertools.permutations([1, 2, 3])))
print(list(itertools.combinations('ABCD', 2)))
print(list(itertools.combinations_with_replacement('ABCD', 2)))

# Kombinační čísla nejsou v základních knihonách, ale ve scipy.
# import scipy.special
# scipy.special.comb(n, k, exact=True) # vrátí n nad k

```

Řešení baterií hrubou silou (projitím celého PP prostoru)

```

import itertools

N = 12      # počet baterií
V = 4      # počet vybitých
baterky = [0]*V + [1]*(N-V)

celkem = 0  # celkový počet možností
a = 0      # kolik z nich splňuje (a)
b = 0      # kolik z nich splňuje (b)

for vyber in itertools.combinations(baterky, 3):

```

```

celkem += 1
pocet_ok = sum(vyber)
if pocet_ok == 3:
    a += 1
if pocet_ok >= 2:
    b += 1

print(f'(a) P={a/celkem:.3f}')
print(f'(b) P={b/celkem:.3f}')

```

Aproximace řešení baterií náhodnými pokusy

```

import random
random.seed()

N = 12      # počet baterií
V = 4      # počet vybitých
baterky = [0]*V + [1]*(N-V)

pokusu = 10000    # kolik provedeme pokusů
a = 0            # kolik z nich splňuje (a)
b = 0            # kolik z nich splňuje (b)

for _ in range(pokusu):
    vyber = random.sample(baterky, k=3)
    ok = sum(vyber)
    if ok >= 3:
        a += 1
    if ok >= 2:
        b += 1

print(f'(a) P={a/pokusu:.3f}')
print(f'(b) P={b/pokusu:.3f}')

```

Nápověda k problému s obálkami: Vytáhneme minci a házíme, dokud nepadne panna. Označíme X celkový počet hodů (včetně posledního, tj. $X \geq 1$). Pokud v naší obálce je k stokorun, tak si obálku necháme, pokud $X < k$. Jaká je pravděpodobnost, že získáme obálku s vyšší částkou?

2. cvičení z PaSti – 2021-03-10

Podmíněná pravděpodobnost

1. Jaký je vztah tvrzení $P(A | B) > P(A)$ a $P(B | A) > P(B)$?

2. Pro plánování výletu do Krkonoš používáme českou a polskou předpověď počasí. Každá nám dá binární výsledek *hezky/ošklivo* a má pravděpodobnost úspěchu $p \in [0, 1]$; obě předpovědi jsou nezávislé. Používáme je takto: pokud se shodují, věříme jim, pokud ne, hodíme si korunou. Jaká je pravděpodobnost, že se rozhodneme správně?

3. (*Paradox Montyho Halla*) V soutěžní hře stojíme na podiu před třemi dveřmi. Za dvojemi je koza (tu nechceme, moc žere), za zbylými auto (to chceme, i když vlastně také moc žere). Vybereme si jedny dveře, ale než je otevřeme, moderátor otevře jedny ze zbylých dveří, ukáže za nimi kozu, a nabídne nám, že můžeme svoji volbu změnit. Máme to udělat? Pomůže to? Uvědomte si, že zadání má (minimálně) následující dvě varianty:

- (a) moderátor ví, kde je auto, a tomu přizpůsobí, které dveře otevře;
- (b) moderátor si hodí korunou, které dveře otevřít. Kdyby odhalil auto, tak bychom asi prohráli, ale to se zrovna nestalo.

Budeme používat strategii, kdy dveře změníme. Spočítejte pravděpodobnost, že vyhrajeme auto, ve variantách (a), (b).

Nezávislé jevy

4. Jsou-li jevy A, B nezávislé, pak jsou nezávislé i A, B^c a také A^c, B^c .

5. Mohou být dva jevy nezávislé a zároveň disjunktní?

6. Najděte jevy A, B, C takové, že $P(A \cap B \cap C) = P(A)P(B)P(C)$, ale jevy nejsou po dvou nezávislé.

Testování vlastností

Motivace: Necht některé elementární jevy mají určitou vlastnost a máme k dispozici test, který ji ne úplně spolehlivě odhaluje – příkladem může být třeba výskyt nemoci v populaci a test na tuto nemoc. Uvažme jevy $N =$ „je nemocný“ a $T =$ „test vyšel pozitivně“.

Představme si, že test danému jedinci vyšel pozitivně. To může znamenat, že nemoc má (*pravý pozitivní výsledek*), nebo že ji nemá, ale test se spletl (*falešný pozitivní výsledek*). Podobně pokud test vyšel negativní, může to být výsledek *pravý negativní* nebo *falešný negativní*. (Anglicky true/false positive/negative.)

Pravděpodobnost pravého pozitivního výsledku je zjevně $P(N | T)$, falešného pozitivního $P(N^c | T)$, podobně pravého negativního $P(N^c | T^c)$ a falešného negativního $P(N | T^c)$. Přitom zjevně $P(N | T) = 1 - P(N^c | T)$ apod.

Při výrobě testu obvykle statisticky testujeme úspěšnost testu na lidech, u kterých víme, zda jsou nemocní. Tím získáme opačně podmíněné pravděpodobnosti $P(T | N)$ a $P(T | N^c)$. Bayseova věta nám je umožňuje obrátit a spočítat pravděpodobnosti z předchozího odstavce.

Ještě k terminologii: U testů se také mluví o *senzitivitě* a *specificitě*. To první je $P(T|N)$ – jaký zlomek nemocných se nám podařilo identifikovat – to druhé pak $P(T^c|N^c)$, tedy totéž pro zlomek zdravých.

Podobnou situaci najdeme i v kontextu vyhledávání informací: představte si třeba internetový vyhledávač, kterému položíte dotaz. Nemocní jsou správné odpovědi na dotaz, pozitivně testovaní jsou výsledky vyhledávače. Tentokrát se parametrům říká *precision* a *recall*. Precision je zlomek správných odpovědí mezi nalezenými, tedy $P(N | T)$, čili pravděpodobnost pravých pozitiv. Recall říká, jaký zlomek správných odpovědí jsme našli – to je $P(T | N)$, čili totéž, co senzitivita. Ať žije jednotnost terminologie ;)

Bayesova věta

7. Petr dostává hodně emailů, ale 80 % z nich jsou spamy. Jeho spamový filtr 90 % spamů správně označí, ale také 5 % řádných emailů označí jako spam.

- (a) Kolik procent emailů bude označeno jako spamy?
- (b) Kolik procent řádných emailů je mezi těmi, co jsou označené jako spamy?
- (c) Kolik procent spamů je mezi emaily, které testem prošly?

8. Kouřovými signály přenášíme binární soubor. Je proto poměrně vysoká pravděpodobnost chyby u každého bitu: 0 se jako 0 přeneso jen s pravděpodobností 0.9, 1 jako 1 jen s pravděpodobností 0.8. Předpokládejme (trochu neseriózně), že jednotlivé znaky se přenášejí nezávisle.

- (a) Pokud jsme dostali signál 0, jaká je pravděpodobnost, že byl opravdu vyslán?
- (b) Dostali jsme zprávu 0010. Jaká je pravděpodobnost, že byla opravdu vyslána?
- (c) Jak se výpočet změní, pokud budeme pro kontrolu vysílat každý symbol třikrát (a pak vezmeme častější z těch tří pokusů)?

Úlohu si můžete zjednodušit předpokladem, že 0 a 1 se mají stejnou pravděpodobnost správného přenosu.

9. Pokud vidíme bílého pudla, zvyšuje to naši důvěru, že je každá vrána černá?

Bonusy

10. (*Simpsonův paradox*) V této úloze budeme mít bonbony dvou druhů: dobré červené a nedobré zelené. Bonbony ale vybíráme z nádoby poslepu (nebo jsme barvoslepi). Rozhodněte, zda se může stát následující podivnost:

- Při vytahování bonbonu z bílé krabice máme vyšší pravděpodobnost, že vytáhneme dobrý bonbon, než z černé krabice.
- Při vytahování bonbonu z bílého sáčku máme vyšší pravděpodobnost, že vytáhneme dobrý bonbon, než z černého sáčku.
- Pokud přesypeme bonbony z bílého sáčku do bílé krabice (a z černého do černé krabice), tak budeme mít lepší pravděpodobnost vytažení dobrého bonbonu v černé krabici.

11. (*Prosecutor's fallacy*) Paní C. umřely dvě děti krátce po narození. Je obžalovaná za dvojnásobnou vraždu. Žalobce argumentuje takto: Pravděpodobnost syndromu náhlého úmrtí kojenců je $1/8500$. Takže pravděpodobnost dvou takových jevů je $1/8500^2$. Tudíž pravděpodobnost, že paní C. je nevinná, je $1/8500^2$, což je hodně málo.

Formulujte argumenty žalobce v řeči pravděpodobnosti a nalezněte v nich dvě chyby.

K procvičení

12. Logická formule $A \implies B$ je ekvivalentní obměně $\neg B \implies \neg A$. Budeme se zabývat analogiemi zahrnujícími pravděpodobnost.

- (a) Ukažte, že pokud $P(B \mid A) = 1$, tak také $P(A^c \mid B^c) = 1$.
- (b) Ukažte, že je však možné, aby $P(B \mid A) \doteq 1$, ale $P(A^c \mid B^c) \doteq 0$.

13. V truhle je sto mincí. Z nich 99 je normálních, ale jedna má na obou stranách orla. Vytáhneme náhodnou minci a šestkrát s ní hodíme, pokaždé padne orel. Jaká je pravděpodobnost, že jsme si vytáhli „dvouorlovou“ minci? (Zkuste napřed odhadnout, pak spočítat.)

14. Na chorobu C máme dva testy, A a B . Test A má sensitivitu i specifitu $p = 0.95$. Test B vždy řekne, že pacient je zdravý. Předpokládejte, že $P(C) = 0.01$.

- (a) Spočítejte pro oba testy pravděpodobnost úspěchu (tj. správné odpovědi), použijeme-li je na náhodného pacienta. Rozmyslete si, co to říká o užitečnosti obou testů.
- (b) Pro jaké p je pravděpodobnost úspěchu obou testů stejná?

15. Ve volbách hlasují lidé pro dva kandidáty, A a B . Při odchodu z volební místnosti jsou voliči náhodně požádáni o účast v exit-poll. Předpokládejme, že kdo odpoví, odpoví popravdě koho volil, ale ne všichni se zúčastní. Označíme-li E množinu voličů, kteří se exit-pollu zúčastní, tak předpokládejme $P(E \mid A) = 0.7$ a $P(E \mid A^c) = 0.4$. Výsledky exit-pollu jsou 60 % pro A . Jaký je skutečný podíl lidí, kteří hlasovali pro A ?

16. Máme tři normální hrací kostky a jednu kostku, kde jsou tři jedničky a tři dvojky. Vybereme uniformně náhodně jednu z kostek, hodíme a padne jednička. Jaká je pravděpodobnost, že jsme vybrali normální kostku?

3. cvičení z PaSti – 2021-03-17

Hrátky s náhodnými veličinami

1. Prokop hází basketbalovým míčem na koš, v každém pokusu má pravděpodobnost zásahu $p = 1/10$, pokusy jsou nezávislé. Skončí po prvním zásahu. Označme X celkový počet hodů.

- (a) Jaká je $P(X > k)$?
- (b) Jaké je rozdělení X ? Tj. určete pravděpodobnostní funkci p_X , tj. pro každé x určete $P(X = x)$.
- (c) Jaká je $P(X \geq 10 \mid X \geq 5)$?
- (d) Jaká je $\mathbb{E}(X)$?

2. Pokračování z minulé úlohy: označme $Y = X \bmod 2$, tj. $Y = 0$, pokud je X sudé, jinak $Y = 1$. Určete rozdělení Y .

3. Quido také hází míčem na koš, má pravděpodobnost q , že se trefí. Označme Z počet zásahů z n pokusů. Určete rozdělení Z .

4. Necht $X \sim \text{Bin}(m, p)$ a $Y \sim \text{Bin}(n, p)$ jsou n.n.v. Pak $X + Y \sim \text{Bin}(m + n, p)$.

5. Necht $X = X_1 + \dots + X_n$, kde pro každé i je $X_i \sim \text{Bern}(p)$. Pokud jsou veličiny $X_1, \dots, X_n$ nezávislé, říkali jsme si na přednášce, že $X \sim \text{Bin}(n, p)$ (zatím bez definice nezávislých veličin...). Ukažte na příkladu, že pokud omezení na nezávislost neuvedeme (tj. chceme jen $X_i \sim \text{Bern}(p)$), tak X může mít i jiné rozdělení.

6. (Kasino v St. Petěrburgu) Házíme opakovaně mincí. Pokud poprvé padla panna v n -tém hoďu, dostaneme odměnu 2^n rublů. Kolik byste byli ochotni zaplatit za účast v této hře?

7. V pytlíku N bonbónů, z nichž K je dobrých. Náhodně vytáhneme dva, označíme X počet dobrých vytažených bonbónů.

- (a) Určete $\mathbb{E}(X)$.
- (b) Můžete i napřed řešit pro tažení jen jednoho bonbónu.
- (c) Jak se jmenuje rozdělení n.v. X ?
- (*) A co když vytáhneme tři, čtyři, ..., n bonbónů?

8. Necht X má uniformní¹ rozdělení na množině $\{a, a + 1, a + 2, \dots, b\}$ (pro celá čísla $a < b$). Určete $\mathbb{E}(X)$ a $\text{var}(X)$. Rozptyl je trochu ošklivý na dopočítávání, připomeňte si vzorec z diskrety na součet druhých mocnin – nebo se omezte na konkrétní příklad $a = 1, b = 6$. Připomeňte si vzorec z přednášky:

$$\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

¹Tím se myslí, že všechny hodnoty z dané množiny mají stejnou pravděpodobnost. Někdy též rovnoměrné.

Randomizované algoritmy

9. *Stupidsort* třídí takto: vygeneruje rovnoměrně náhodnou permutaci prvků, otestuje, zda je setříděná, a pokud není, spustí se znovu. Jaká je střední hodnota počtu pokusů, než setříděnou permutaci najdeme? Jaká je střední hodnota časové složitosti algoritmu (to je také náhodná veličina)? A v jakém pravděpodobnostním prostoru se vlastně pohybujeme?

10. *Pseudomedián* množiny $X = \{x_1, \dots, x_n\}$ je jakékoliv $x \in X$, které je větší než aspoň $n/4$ prvků z X a menší než aspoň $n/4$ prvků z X . Jinými slovy kdybychom množinu seřadili, pseudomediány leží v „prostřední polovině“².

Uvažujme následující algoritmus na výpočet pseudomediánu. Rovnoměrně náhodně vybereme jeden prvek, ověříme, je-li to pseudomedián, a pokud není, algoritmus spustíme znovu.

Časová složitost algoritmu je náhodná veličina (v jakém přesně pravděpodobnostním prostoru?). Určete její střední hodnotu. Umíte určit pravděpodobnostní funkci?

11. *k-tý nejmenší prvek* v množině X můžeme hledat následovně. Najdeme *pivota* – nějaký pseudomedián p . Rozdělíme množinu na

$$X_{<} = \{x \in X \mid x < p\} \quad \text{a} \quad X_{>} = \{x \in X \mid x > p\}.$$

- Pokud $k \leq |X_{<}|$, rekurzivně hledáme k -tý nejmenší prvek v $X_{<}$.
- Pokud $k = |X_{<}| + 1$, je k -tým nejmenším prvkem pivot p .
- Jinak rekurzivně hledáme $(k - |X_{<}| - 1)$ -tý nejmenší prvek v $X_{>}$.

Nahlédněte, že tento algoritmus funguje. Pokud budeme pseudomedián hledat podle předchozího cvičení, jaká je střední hodnota časové složitosti tohoto algoritmu?

Jak se střední hodnota změní, pokud jako pivota vybereme rovnoměrně náhodný prvek $p \in X$?

12. *Quicksort* je třídící algoritmus postavený na stejné myšlence rekurzivního dělení. Vybereme pivota p , rozdělíme množinu na $X_{<}$ a $X_{>}$. Rekurzivně setřídíme $X_{<}$ a $X_{>}$. Nakonec slepíme za sebe setříděnou $X_{<}$, pivota a setříděnou $X_{>}$.

Nejprve opět uvažujme verzi, v níž volíme pivota jako pseudomedián spočítaný naším randomizovaným algoritmem. Dokažte, že střední hodnota časové složitosti je $\mathcal{O}(n \log n)$.

Poté uvažujme rovnoměrně náhodnou volbu pivota. Dokažte, že střední časová složitost je nadále $\mathcal{O}(n \log n)$. Může se hodit následující trik: Nejdříve si všimneme, že časová složitost je lineární v celkovém počtu porovnání C . Pak pro každé $x_i \in X$ zavedeme náhodnou veličinu C_i , která bude říkat, kolika porovnání se x_i zúčastní. Všimneme si, že $C = \sum_i C_i$ a použijeme linearity střední hodnoty. A propos, v jakém pravděpodobnostním prostoru pracujeme?

²tedy prostředních dvou čtvrtinách

K procvičení

13. Na dvanáctistěnné kostce je jedna stěna označena jedničkou, dvě dvojkou, čtyři čtyřkou a pět pětkou, všechny stěny padají stejně často. Označíme X výsledek jednoho hodu. Spočítejte $\mathbb{E}(X)$ a $\text{var}(X)$, vyčíslete také směrodatnou odchylku σ_X .³

14. Hugo upravil svoji hrací kostku tak, že ze čtyřky udělal druhou šestku; všechny stěny padají stejně často. Označíme X výsledek jednoho hodu. Spočítejte $\mathbb{E}(X)$, $\text{var}(X)$ a σ_X .

³Směrodatná odchylka σ_X veličiny X se definuje jako $\sqrt{\text{var } X}$.

4. cvičení z PaSti – 2021-03-24

Z každé kapitoly (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 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 je teď střední hodnota?

Zacházení se střední hodnotou a rozptylem

4. Nechť X, Y jsou diskrétní n.v., $\alpha \in \mathbb{R}$.
 - (a) Ukažte, že $\text{var}(X + \alpha) = \text{var}(X)$.
 - (b) Vyjádřete $\text{var}(\alpha X)$ pomocí $\text{var}(X)$.
 - (c) Ukažte, že $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$ pokud X, Y jsou nezávislé.
5. Dokažte:
 - (a) Pokud $\mathbb{E}(X^2) = 0$, tak $P(X = 0) = 1$.
 - (b) Předpokládejme, že $\text{var}(X) = 0$, 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 $\{a, b, c, d\}$, 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 $\text{Im}(X) = \text{Im}(Y) = \{1, 2, \dots, n\}$ 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 psťní 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. $X_1, \dots, X_n$ mají geometrické rozdělení s parametry $p_1, \dots, p_n$. Jaké je rozdělení $\min(X_1, \dots, X_n)$?

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

- (a) Najděte sdruženou pravděpodobnostní funkci pro n.v. $X_1, \dots, X_n$.
- (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) Necht $A = A_1 \cup A_2 \cup \dots \cup A_n$. 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 $M \sim \text{Pois}(\lambda)$ 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 $S \sim \text{Pois}(p\lambda)$.

(c) Odvoďte, že $H \sim \text{Pois}((1 - p)\lambda)$ 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 (Cumulative 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 konkrétními
# parametry, aniž bychom ho explicitně konstruovali.
print(stats.binom.rvs(n, p, size=20))
```

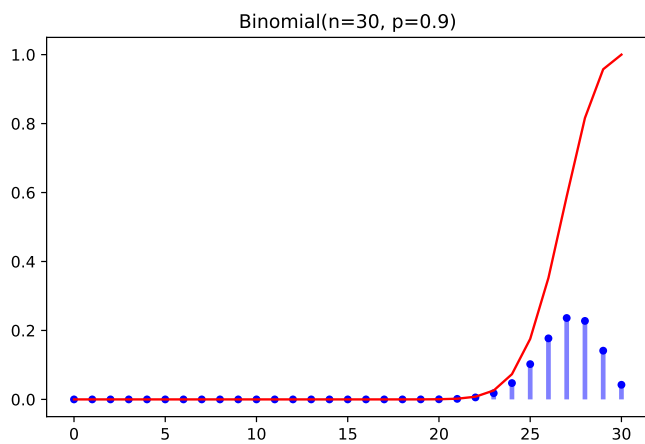
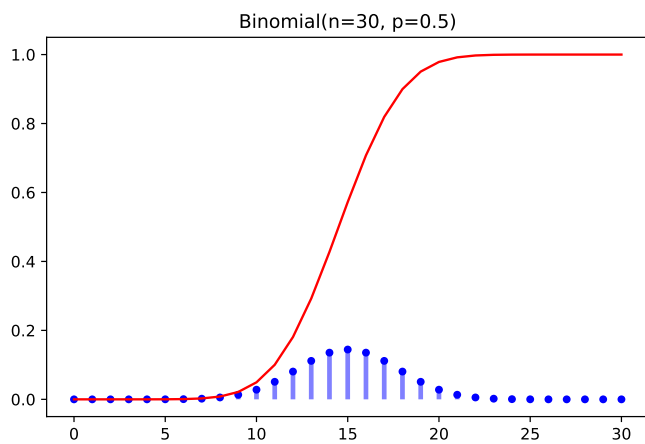
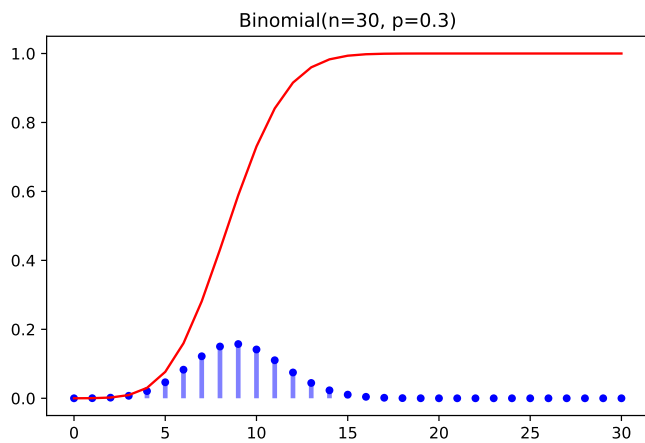

Kreslíme obrázky

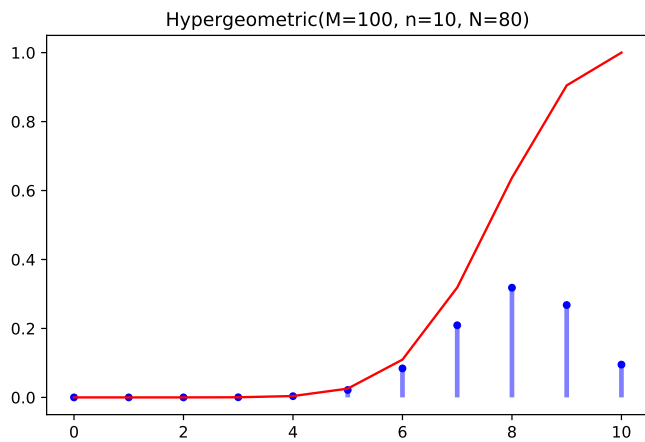
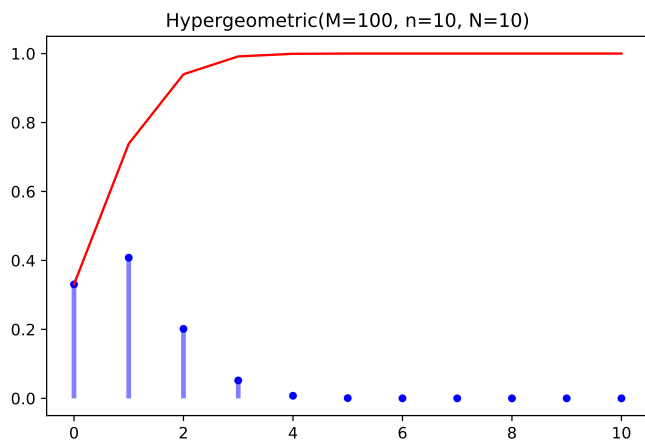
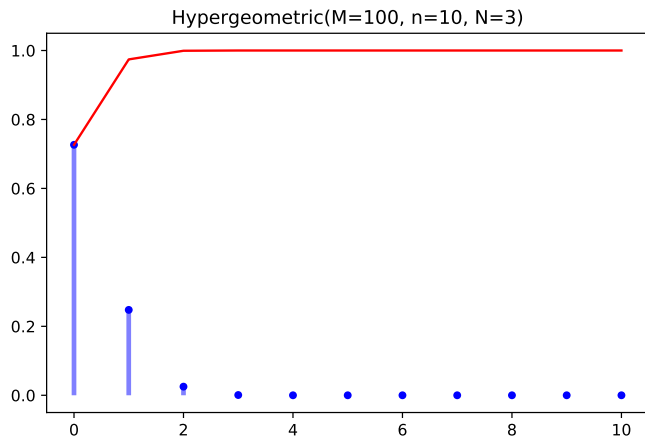
```
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt

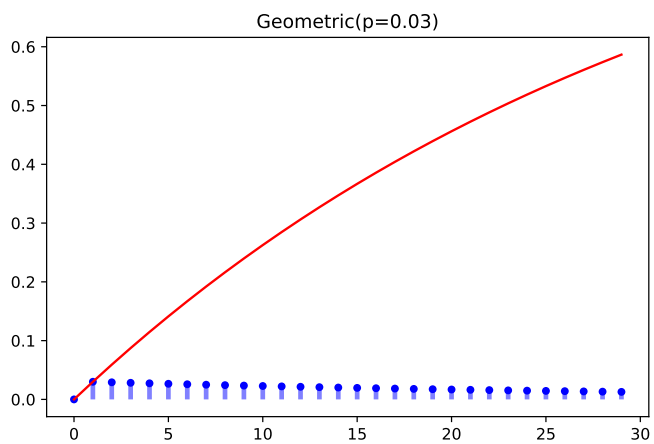
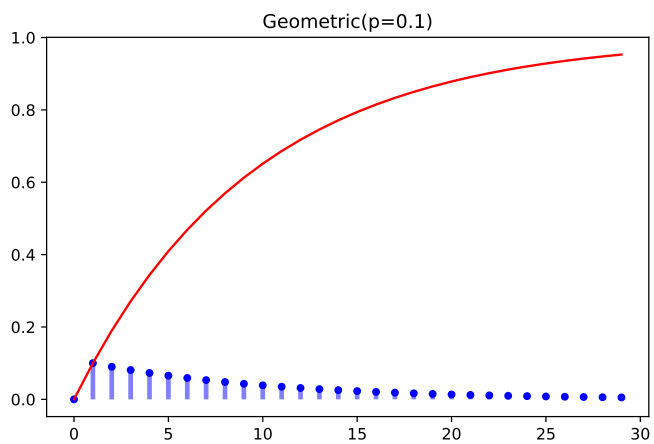
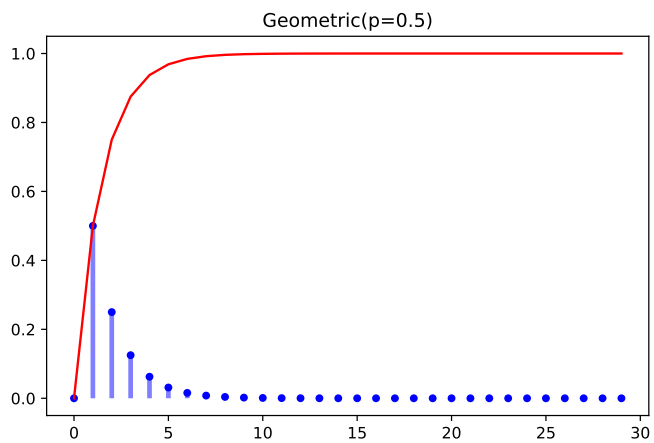
# Rozdělení
n = 20
p = 0.3
b = stats.binom(n, p)

# 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()

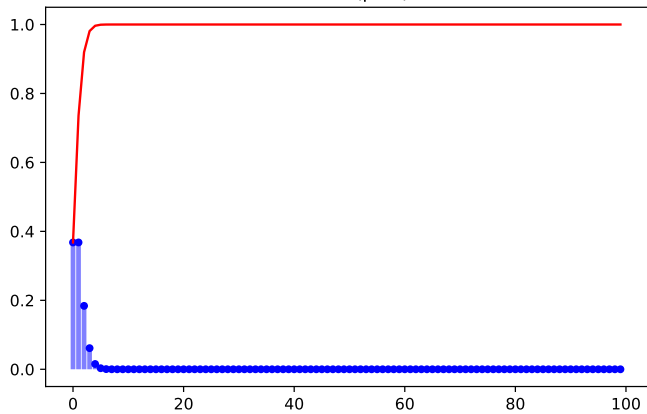
# Nebo více obrázků pohromadě
fig, ax = plt.subplots(1, 2)
x = np.arange(0, n+1)
ax[0].set_ylim(0, 1)
ax[0].plot(x, b.pmf(x), 'bo', ms=8, label='binom pmf')
ax[0].vlines(x, 0, b.pmf(x), colors='b', lw=5, alpha=0.5)
ax[0].legend()
ax[1].plot(x, b.cdf(x), 'r', label='binom cdf')
ax[1].legend()
plt.show()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



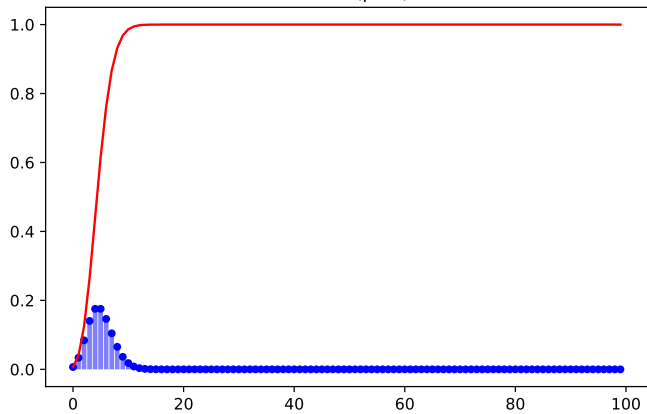




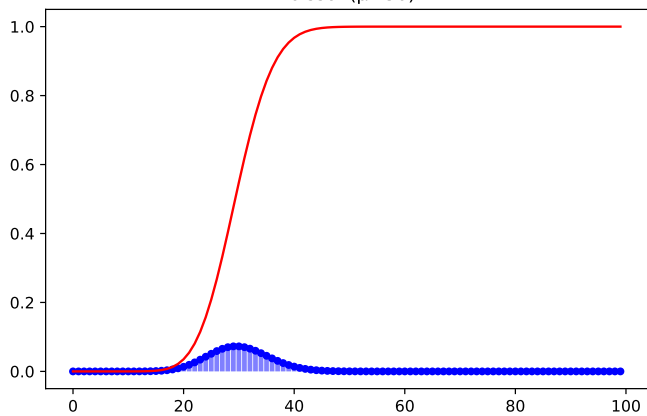
Poisson($\mu=1$)



Poisson($\mu=5$)



Poisson($\mu=30$)



5. cvičení z PaSti – 2021-03-31

Rozdělení náhodných veličin

1. (*Newton-Pepys problem*) Označme A_1 jev „z šesti hodů kostkou padne aspoň jedna šestka“. Dále A_2 jev „z dvanácti hodů kostkou padnou aspoň dvě šestky“. Obecně: A_k jev „z $6k$ hodů kostkou padne aspoň k šestek“.

- (a) Vyjádřete $P(A_i)$ pomocí vzorečku s kombinačními čísly.
- (b) Vyjádřete $P(A_i)$ pomocí distribuční funkce binomického rozdělení.
- (c) Jaká je střední hodnota počtu šestek pro každou z variant?
- (d) * Zapřemýšlejte, která z $P(A_k)$ by měla být největší, bez použití vzorce.

2. (*de Mèreho problém*)

- (a) Jaká je pravděpodobnost, že padne ze čtyř hodů aspoň jedna šestka?
- (b) Jaká je pravděpodobnost, že padne z 24 hodů dvojicí kostek aspoň jedna dvojitá šestka?
- (c) De Mèreho problém spočíval v tom, jestli jsou odpovědi na části a, b stejné.
- (d) Umíte úlohu interpretovat jako otázku o binomickém rozdělení? O geometric-kém rozdělení?

Náhodné vektory

3. Dokažte následující vlastnost Poissonova rozdělení. Necht $X \sim Pois(\lambda)$, $Y \sim Pois(\mu)$ jsou n.n.v. Pak $X + Y \sim Pois(\lambda + \mu)$.

4. Hodíme třikrát mincí. Označíme X počet rubů v prvních dvou hodech a Y počet líců v posledních dvou hodech.

- (a) Určete sdruženou pravděpodobnostní funkci $p_{X,Y}$ a také marginální pravděpodobnostní funkce p_X , p_Y .
- (b) Jsou X a Y nezávislé?
- (c) Určete $P(X < Y)$.
- (d) Určete podmíněnou pravděpodobnostní funkci $p_{X|Y}$.

Spojité náhodné veličiny

Připomeňte si, že distribuční funkce F_X je definována vztahem

$$F_X(x) = P(X \leq x).$$

V některých případech je $F_X(x) = \int_{-\infty}^x f_X(t) dt$ pro vhodnou nezápornou funkci f_X (hustotu X); takovým n.v. se říká *spojité*. Pak platí:

- $P(X \in A) = \int_A f_X(t) dt$.
Přitom $\int_A f(t) dt$ definujeme jako $\int_{-\infty}^{+\infty} 1_A(t)f(t) dt$, kde 1_A je charakteristická funkce množiny (1 uvnitř, 0 venku).
- $\mathbb{E}(X) = \int_{-\infty}^{\infty} t f_X(t) dt$ a obecněji

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(t) f_X(t) dt.$$

- Stejně jako pro diskrétní n.v. je $\text{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

5. Pro n.v. X s distribuční funkcí F vyjádřete

- (a) $P(X \in (0, 1])$
- (b) $P(X > 0)$
- (c) $P(X < 0)$
- (d) $P(X \in [0, 1])$

6. Necht X je spojitá náhodná veličina. Vyjádřete pomocí F_X distribuční funkci náhodných veličin

- (a) $-X$
- (b) $X^+ = \max(0, X)$
- (c) $X^- = -\min(X, 0)$
- (d) $|X| = X^+ + X^-$

7. Necht F_X je dána předpisem $F_X(x) = x/3$ pro $x \in [0, 3]$, $F_X(x) = 0$ pro $x < 0$ a $F_X(x) = 1$ pro $x > 3$. Necht $Y = 1/X$ a $Z = X^2$. Spočítejte

- (a) $P(1 \leq X \leq 2)$
- (b) $P(X \leq Y)$
- (c) $P(X \leq Z)$
- (d) hustotní funkci f_X
- (e) distribuční funkce F_Y a F_Z

Bonusy

Viz příklady z minulého cvičení.

6. cvičení z PaSti – 2021-04-07

Z každé kapitoly zkuste aspoň jeden příklad.
Pokud se zaseknete, na poslední straně jsou nápovědy.

Základy pravděpodobnosti

1. Uvažme skupinu 25 lidí. Jaká je pravděpodobnost, že žádní dva z nich nemají narozeniny ve stejný den? (Předpokládejte, že den narozenin je rovnoměrně rozdělená náhodná veličina. Pro jednoduchost ignorujte přestupné roky.) Nemusíte vyčíslovat.

Rozbor možností – věta o úplné psti a úplné střední hodnotě

2. Házíme mincí, dokud nepadne orel, na každý hod dostaneme novou minci. (Jaká je pravděpodobnost, že získáme k mincí?) Pak všechny získané mince hodíme najednou, pokud na každé z mincí padne orel, můžeme si je všechny nechat. Jaká je pravděpodobnost, že se to stane? Jaká je střední hodnota počtu získaných mincí? (Tu můžete nechat ve tvaru nekonečné sumy.)

Diskrétní náhodné veličiny

3. Máme kroužek s pěti klíči, jeden z nich je od vchodových dveří. Nevíme který, tak je budeme zkoušet náhodně. Určete pravděpodobnostní funkci a střední hodnotu počtu vyzkoušených klíčů, pokud

- (a) po vyzkoušení zapomeneme, který klíč jsme zkoušeli, a další vybíráme opět náhodně ze všech.
- (b) vyzkoušené klíče si označíme, příště je už nezkoušíme.
- (c) Části (a), (b) zkuste znovu pro variantu s deseti klíči, z nichž jsou dva správné.

4. Knihovna MFF má 1000 čtenářů – studentů informatiky – a rozhoduje se, kolik kopií nové knihy koupit. Předpokládejme, že o knihu má v daný semestr každý student zájem s pravděpodobností $p = 0.01$, nezávisle na ostatních.

- (a) Určete pravděpodobnostní funkci pro počet studentů, kteří mají o knihu zájem.
- (b) Určete pravděpodobnostní funkci pro Poissonovskou aproximaci tohoto počtu.
- (c) Jaká je pravděpodobnost, že 20 kopií knihy nestačí? Vyjádřete jednak pomocí distribuční funkce, jednak pomocí sumy. A také jednak pomocí přesné formule z části (a), jednak pomocí aproximace z části (b).
- (d) Je popsáný model zájmu studentů o knihy realistický?

Diskrétní náhodné vektory

5. Necht X, Y jsou nezávislé náhodné veličiny s rozdělením $Geom(p)$.

- (a) Napište sdruženou pravděpodobnostní funkci $p_{X,Y}$.
- (b) Spočtete $P(X + Y = n)$.
- (c) Spočtete $P(X = k \mid X + Y = n)$.

6. Označme X náhodnou veličinu, která popisuje počet emailů, které dostaneme za jeden den. Dále $Y \leq X$ necht je počet obdržných emailů, které jsou spamy. Předpokládejme, že $X \sim Pois(\lambda)$. Dále předpokládejme, že každý email, který dostaneme, má (nezávisle na ostatních) pravděpodobnost p , že je spam.

- (a) Napište podmíněnou pravděpodobnostní funkci $p_{Y|X}$.
- (b) Napište sdruženou pravděpodobnostní funkci $p_{X,Y}$.
- (c) Odvoďte z $p_{X,Y}$ marginální pravděpodobnostní funkci p_X a ověřte, že odpovídá $Pois(\lambda)$. (To by mělo být pravda automaticky, ale ověřte to ze vzorce.)
- (d) Odvoďte z $p_{X,Y}$ marginální pravděpodobnostní funkci p_Y a ověřte, že odpovídá $Pois(p\lambda)$. (To je poněkud nečekané.)
- (e) Odsud snadno plyne, že $Y - X \sim Pois((1 - p)\lambda)$.
- (f) * Náhodné veličiny X a $Y - X$ jsou nezávislé.

Spojitě náhodné veličiny

(Definice si můžete připomenout v příkladech z minulého cvičení.)

7. Necht F_X je dána předpisem $F_X(x) = x/3$ pro $x \in [0, 3]$, $F_X(x) = 0$ pro $x < 0$ a $F_X(x) = 1$ pro $x > 3$. Necht $Y = 1/X$ a $Z = X^2$. Spočtete

- (a) $P(1 \leq X \leq 2)$
- (b) $P(X \leq Y)$
- (c) $P(X \leq Z)$
- (d) hustotní funkci f_X
- (e) distribuční funkce F_Y a F_Z

8. *Exponenciální rozdělení* je spojitou analogií rozdělení geometrického. Vyjadřuje dobu čekání na první událost generovanou poissonovským procesem (s daným parametrem λ). Náhodná veličina $X \sim \text{Exp}(\lambda)$ má distribuční funkci

$$F_X(t) = \begin{cases} 1 - e^{-\lambda t} & \text{pro } t \geq 0, \\ 0 & \text{pro } t < 0. \end{cases}$$

Vypočítejte:

- (a) hustotní funkci $f_X(t)$
- (b) střední hodnotu $\mathbb{E}(X)$
- (c) rozptyl $\text{var}(X)$

Nápovědy

- 1. Stačí použít základní formalku počet dobrých děleno počtem všech.
- 2. Použijte rozklad pravděpodobnostního prostoru: B_k budou průběhy experimentu, kde první kolo skončíme s k mincemi.
- 3. (a) Geometrické rozdělení. (b) Představte si, že napřed seřadíme pět klíčů náhodně do řady, pak je bereme v tom pořadí. (c) První část je úplně stejná, druhá je jiná – podobná problému s maximem ze dvou kostek (příklad 4.10.).
- 4. (a) Jedná se o binomické rozdělení. (b) Zde jde o Poissonovo rozdělení s intenzitou $1000 \cdot 0.01$. (c) Které hodnoty je třeba sčítat? (d) Jak je to s tou nezávislostí?
- 5. Postupujte přesně podle definic. Pro kontrolu, část (c) nezávisí na k (pokud je hodnota k možná).
- 6. (a) Víme-li, že pokud $X = n$, tak se Y chová jako $\text{Bin}(n, p)$. (b) Použijte vzorec z přednášky – vztah $p_{X,Y}$ a $p_{X|Y}$.
- 7. Dosadte do definic.
- 8. Dosadte do definic.

7. cvičení z PaSti – 2021-04-14

Z každé kapitoly zkuste aspoň jeden příklad.

Distribuční funkce

Připomeňte si, že distribuční funkce F_X je definována vztahem $F_X(x) = P(X \leq x)$.

1. Pro jistý problém máme k dispozici dva algoritmy, A a B. Algoritmus C spočívá v tom, že si náhodně vybereme, který z algoritmů A, B spustíme – A bude mít pravděpodobnost p , B pravděpodobnost $1 - p$. Dobu běhu A, B, C chápeme jako náhodné veličiny, označíme je X , Y , Z .

- (a) Vyjádřete $\mathbb{E}(Z)$ pomocí $\mathbb{E}(X)$ a $\mathbb{E}(Y)$.
- (b) Určete F_Z pomocí F_X , F_Y .
- (c) Pokud jsou X , Y spojité, určete f_Z pomocí f_X , f_Y .

2. Metrový klacek rozložíme na dva kusy lomem v uniformně náhodném bodě. Bud X délka delší části.

- (a) Jaké je rozdělení X ?
- (b) Určete $\mathbb{E}(X)$.

Hustota

Pro spojitě n.v. je $F_X(x) = \int_{-\infty}^x f_X(t) dt$ pro vhodnou nezápornou funkci f_X (hustotu X). Pak je také $P(X \in A) = \int_A f_X(t) dt$.

3. Házíme na terč – kruh o poloměru 1. Předpokládejme, že každý bod v terči má stejnou pravděpodobnost zásahu, přesněji, každá jeho podmnožina má pravděpodobnost úměrnou své ploše. Označme X vzdálenost od středu.

- (a) Najděte distribuční funkci F_X .
- (b) Najděte hustotní funkci f_X .
- (c) Zjistěte $\mathbb{E}(X)$, $\text{var}(X)$, σ_X .

4. Bublifikem vyfoukneme bublinu o poloměru $R \sim U(1, 5)$. Jaká je střední hodnota povrchu bubliny?

Výpočty momentů

Pro n.v. X definujeme její k -tý moment jako $\mathbb{E}(X^k)$.

5. Buď $X \sim U(a, b)$. Na přednášce jsme si ukazovali výpočet $\mathbb{E}(X)$.

- (a) Spočtete analogicky $\mathbb{E}(X^2)$ a odsud $\text{var}(X)$.
- (b) Alternativně, uvědomte si napřed, jaké je rozdělení veličiny $Y = X - \mathbb{E}(X)$. Pak spočtete $\mathbb{E}(Y^2)$, což je $\text{var}(X)$.

6.

- (a) Necht X je diskrétní nebo spojitá náhodná veličina a $X \geq 0$ skoro jistě (tím se myslí, že $P(X \geq 0) = 1$). Pokud $\mathbb{E}(X)$ existuje, tak $\mathbb{E}(X) \geq 0$. Dokažte.
- (b) Necht Y, Z jsou diskrétní nebo spojitě náhodné veličiny a $Y \leq Z$ skoro jistě. Pokud $\mathbb{E}(Y), \mathbb{E}(Z)$ existují, tak $\mathbb{E}(Y) \leq \mathbb{E}(Z)$. Dokažte.

Sampling

7. Necht $X_1, \dots, X_n$ jsou nezávislé náhodné veličiny se stejným rozdělením se střední hodnotou μ a rozptylem σ^2 . Označme $S_n = (X_1 + \dots + X_n)/n$. To můžeme považovat za odhad střední hodnoty μ průměrem z n pokusů.

- (a) Určete $\mathbb{E}(S_n)$ a $\text{var}(S_n)$.
- (b) Ukažte, jak lze počítat S_n z S_{n-1} , X_n a n .
- (c) Použijte vhodné X_i , aby μ obsahovalo číslo π . Sestavte program v libovolném jazyce a spočítejte pomocí něj hodnotu π . (Jak velké n myslíte, že bude potřeba pro pět správných číslic?)

Modelování pomocí náhodných veličin

8. Pan Chen Cheng navštívil Prahu a v uniformně náhodný čas se objeví na Staroměstském náměstí. Každou celou hodinu od 9:00 do 23:00 se na orloji objevuje 12 figur apoštolů. (Pro účely této úlohy předpokládejme, že apoštolové se objeví jen na okamžik a hned zase zmizí.)

- (a) Jaká je pravděpodobnost, že pan Cheng uvidí apoštoly, aniž by čekal déle než 15 minut.
- (b) Co když pan Cheng přijde na Staroměstské náměstí v uniformně náhodném čase po poledni, tj. 12:00–24:00?

9. Předpokládejme, že u poštovní přepážky trvá vyřízení jednoho zákazníka čas, který má exponenciální rozdělení a střední hodnotu 4 minuty.

- (a) Jaký je parametr λ , jaká je distribuční funkce?
- (b) Jaká je pravděpodobnost, že budeme čekat více než 4 minuty?
- (c) Jaká je pravděpodobnost, že budeme čekat něco mezi 3 a 5 minutami?

10. Říkáme, že náhodná veličina X (resp. její rozdělení) *nemá paměť*, pokud

$$P(X > s + t \mid X \geq s) = P(X > t)$$

pro $s, t \geq 0$. Jinými slovy, doba, kterou jsme již čekali, nemá vliv na dobu, kterou budeme ještě čekat. Už jsme viděli, že geometrické rozdělení nemá paměť. Ukažte, že ani exponenciální rozdělení nemá paměť. Platí dokonce, že je to jediné spojité rozdělení na kladných číslech bez paměti (a geometrické je jediné diskrétní bez paměti), ale to dokazovat nemusíte.

11. Necht $X_i \sim \text{Exp}(\lambda_i)$ pro $i = 1, \dots, n$ jsou nezávislé náhodné veličiny. Označme $M = \min(X_1, \dots, X_n)$. Ukažte, že $M \sim \text{Exp}(\lambda_1 + \dots + \lambda_n)$.

12. Budeme modelovat množství sněhu, který bude na Silvestra v lyžarském areálu Ještěd, pomocí normálního rozdělení se střední hodnotou 40 (centimetrů) a směrodatnou odchylkou 10.

- (a) Jaká je pravděpodobnost, že nám model určí zápornou hodnotu sněhové pokrývky?
- (b) Jaká je pravděpodobnost, že sněhu napadne 50–70 cm?

Hodnoty $\Phi(x)$ si spočítejte v Pythonu nebo v R, případně se podívejte do tabulky na https://en.wikipedia.org/wiki/Standard_normal_table (sekce Cumulative).

Bonusy

13. Návod, jak ověřit, že hustota normálního rozdělení se opravdu zintegruje na 1. Chceme vypočítat $I = \int_{-\infty}^{\infty} e^{-x^2/2} dx$, resp. ukázat, že $I = \sqrt{2\pi}$. Ukážeme místo toho, že $I^2 = 2\pi$. Platí totiž

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-x^2/2} \cdot e^{-y^2/2} dx dy$$

a tento dvojný integrál se velmi zjednoduší po převodu do polárních souřadnic.

14.

- (a) Uvažme seznam států světa a jejich aktuálního počtu obyvatel. Odhadněte, kolik z těchto počtů začíná jedničkou! (Lhostejno, na které pozici na první jednička je.)
- (b) Srovnajte s nějakou skutečnou tabulkou, např. na <https://www.worldometers.info/world-population/population-by-country/>.
- (c) Promyslete, proč by to tak mohlo být. Případně si přečtěte o *Benfordově zákonu*.

15. Střední hodnota diskrétní i spojitě náhodné veličiny splňuje

$$\mathbb{E}(X) = \int_0^\infty P(X > t) dt - \int_{-\infty}^0 P(X < t) dt.$$

K procvičení

16. Nechť $U \sim U(0, 1)$ a $p \in [0, 1]$. Uvažme funkci

$$g(t) = \begin{cases} 0 & \text{pro } x > p, \\ 1 & \text{pro } x \leq p. \end{cases}$$

Co můžete říct of n.v. $X = g(U)$? Spočtete její střední hodnotu dvěma způsoby – přímo ze znalosti jejího rozdělení i pomocí pravidla LOTUS.

17. Nechť F_X je dána předpisem $F_X(x) = x/3$ pro $x \in [0, 3]$, $F_X(x) = 0$ pro $x < 0$ a $F_X(x) = 1$ pro $x > 3$. Nechť $Y = 1/X$ a $Z = X^2$. Spočtete

- (a) $P(1 \leq X \leq 2)$
- (b) $P(X \leq Y)$
- (c) $P(X \leq Z)$
- (d) hustotní funkci f_X .
- (e) distribuční funkce F_Y a F_Z .

18. Střední doba života harddisku je 4 roky. Předpokládejme, že tato doba je popsána náhodnou veličinou s exponenciálním rozdělením. (To není realistický předpoklad, viz např. <https://www.backblaze.com/blog/how-long-do-disk-drives-last/>.)

- (a) Jaká je pravděpodobnost, že disk selže během prvních tří let?
- (b) Jaká je pravděpodobnost, že vydrží alespoň 10 let?
- (c) Po jaké době se rozbije 10 % disků?

19. Plutonium-238 má poločas rozpadu 87.7 let. Jeho rozpad budeme modelovat pomocí exponenciálního rozdělení: pro každý atom budeme čas, za který se rozpadne, považovat za nezávislou náhodnou veličinu s rozdělením $\text{Exp}(\lambda)$.

- (a) Jaké je λ ?
- (b) Jaká je střední doba života atomu ^{238}Pu ?
- (c) Po jaké době se rozpadne 90 % atomů?
- (d) Kolik procent atomů se rozpadne po 50 letech? (Mimochodem, některé kosmické sondy a některé kardiostimulátory používají rozpad ^{238}Pu jako zdroj energie.)

20. Pozorujeme meteorický roj. Doba, za kterou uvidíme meteor, je exponenciálně rozdělená se střední hodnotou 1 (minuta).

- (a) Jaká je pravděpodobnost, že budeme muset čekat více než 5 minut?
- (b) Jaká je pravděpodobnost, že se dočkáme za nejvýše jednu minutu?
- (c) * Jaké je rozdělení času, kdy uvidíme druhý meteor? Třetí, ... (Předpokládáme, že jednotlivé meteory jsou navzájem nezávislé.)

21. Nechť $X \sim N(0, 1)$, $Y \sim N(1, 4)$.

- (a) Najděte lineární funkci $f(t) = a \cdot t + b$, aby $f(Y)$ měla stejnou distribuci jako X .
- (b) Spočtete $P(X \leq 1)$, $P(X > 2)$.
- (c) Spočtete $P(Y < 0)$, $P(Y > 2)$.

22. Buď Y minimum z k uniformně náhodných čísel z intervalu $[0, 1]$. Spočtete $\mathbb{E}(Y)$.

23. Frantovi jsme ve skoku do dálky naměřili 9 metrů, což překonává světový rekord o 5 cm. Při měření jsme se ovšem dopustili chyby s rozdělením $N(0, 0.01)$. Jaká je pravděpodobnost, že byl rekord skutečně překonán?

24. Nechť $X \sim N(0, 1)$ a $Y = |X|$. Určete $\mathbb{E}(Y)$ a $\text{var}(Y)$.

9. cvičení z PaSti – 2021-04-28

Z každé kapitoly zkuste aspoň jeden příklad.

Samplování

1. Nechť n.v. X má distribuční funkci F , hustotu f , střední hodnotu μ a rozptyl σ^2 . Nechť dále $Y = aX + b$.

- (a) Jakou střední hodnotu a rozptyl má Y ?
- (b) Jakou distribuční funkci a hustotu má Y ?
- (c) Pokud $X \sim N(0, 1)$, jak nastavit a a b , aby bylo $Y \sim N(\alpha, \beta^2)$?
- (d) Jsou-li Φ a φ distribuční funkce a hustota pro $N(0, 1)$, jak vypadá distribuční funkce a hustota pro $N(\mu, \sigma^2)$?

2. Vzpomeňte si na větu z přednášky. Nechť $U \sim U(0, 1)$. Jak vyrobíte náhodnou veličinu

- (a) s rozdělením $U(a, b)$?
- (b) s rozdělením $Exp(\lambda)$?
- (c) s Cauchyho rozdělením? (připomeňte si, že $(\arctg x)' = 1/(1 + x^2)$)
- (d) s rozdělením $N(0, 1)$?

3. Nechť $X_1, \dots, X_n$ jsou nezávislé náhodné veličiny a mají všechny stejné rozdělení se střední hodnotou μ a rozptylem σ^2 . Označme $S_n = (X_1 + \dots + X_n)/n$. To můžeme považovat za odhad střední hodnoty μ průměrem z n pokusů.

- (a) Určete $\mathbb{E}(S_n)$ a $\text{var}(S_n)$.
- (b) Ukažte, jak lze počítat S_n z S_{n-1} , X_n a n .
- (c) Použijte vhodné X_i , aby μ obsahovalo číslo π . Sestavte program v libovolném jazyce a spočítejte pomocí něj hodnotu π . (Jak velké n myslíte, že bude potřeba pro pět správných číslic?)

Normální rozdělení

4. Nechť $Z \sim N(0, 1)$. Pomocí tabulky funkce Φ ověřte pravidlo 3σ , neboli spočítejte

- (a) $P(|Z| \leq 1)$
- (b) $P(|Z| \leq 2)$
- (c) $P(|Z| \leq 3)$
- (d) Přepište, co to znamená pro n.v. $X \sim N(\mu, \sigma^2)$.

Modelování pomocí náhodných veličin

Viz úlohy z minulého cvičení.

Spojité vektory

Připomeňte si:

- dvojné integrály jde prohazovat (Fubiniho věta)

$$\int_X \int_Y f(x, y) \, dy \, dx = \int_Y \int_X f(x, y) \, dx \, dy.$$

Potřeba je, aby se nejednalo o „integrály typu $\infty - \infty$ “, neboli $\int_X \int_Y |f(x, y)|$ musí být konečný.

- pro „rozumnou“ množinu A

$$P((X, Y) \in A) = \int_A f_{X,Y}(x, y) \, dx \, dy.$$

- $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dy$
- $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx$
- nezávislost $X, Y \Leftrightarrow F_{X,Y}(x, y) = F_X(x)F_Y(y) \Leftrightarrow f_{X,Y}(x, y) = f_X(x)f_Y(y)$

5. Nechť X, Y mají sdruženou hustotu $f_{X,Y}(x, y) = e^{-x-y}$ pro $x, y > 0$ (a 0 jinak).

- Určete marginální hustoty f_X, f_Y .
- Určete také distribuční funkce $F_X, F_Y, F_{X,Y}$.
- Jsou X, Y nezávislé?
- Najděte $P(X + Y \leq 1)$ a $P(X > Y)$.

6. Volme uniformně náhodně bod z půlkruhu o poloměru 1, se středem v počátku a ležícím v horní polovině. (Uniformně znamená, že pravděpodobnost každé podmnožiny je úměrná jejímu obsahu.) Označme X, Y souřadnice zvoleného bodu.

- Najděte sdruženou hustotu $f_{X,Y}$.
- Najděte marginální hustotu f_Y a spočítejte pomocí ní $\mathbb{E}(Y)$.
- Pro kontrolu spočítejte $\mathbb{E}(Y)$ přímo (pomocí pravidla LOTUS).

7. (Buffonova jehla) Na nekonečnou podlahu hodíme náhodně jehlu délky ℓ . Podlaha je z prken, jejich okraje tvoří rovnoběžné přímky ve vzdálenosti d . Určete pravděpodobnost, že jehla bude přesahovat okraj některého prkna.

Nápověda: Nakreslete obrázek a popište polohu jehly pomocí dvou náhodných proměnných (posun a úhel).

K procvičení

8. Buď Y minimum z k uniformně náhodných čísel z intervalu $[0, 1]$. Spočtete $\mathbb{E}(Y)$.

9. Najděte analogii „pravidla 3σ “, neboli spočtete $P(|X - \mathbb{E}(X)| < c \cdot \sigma_X)$ pro $c = 1, 2, 3$, pokud:

- (a) X má uniformní rozdělení,
- (b) $X \sim \text{Exp}(1)$,
- (c) $X \sim \text{Exp}(2)$.

10. cvičení z PaSti – 2021-05-05

Z každé kapitoly zkuste aspoň jeden příklad.

Konvoluce

1. Buďte $X, Y, Z \sim U(0, 1)$ nezávislé náhodně veličiny.
 - (a) Jaké je rozdělení $X + Y$? Určete hustotu – jak podle konvolučního vzorce, tak „podle obrázku“.
 - (b) Jaké je rozdělení $X + Y + Z$? Pro jednoduchost určete hustotní funkci jen na intervalu $[0, 1]$.
 - (c) Jak výsledek ověřit smplováním? (Proveďte rychlý experiment, např. v Pythonu, nebo jen popište, co byste dělali.)
2. Buďte $X, Y, Z \sim \text{Exp}(\lambda)$ nezávislé náhodně veličiny.
 - (a) Jaké je rozdělení $X + Y$?
 - (b) Jaké je rozdělení $X + Y + Z$?

Aplikace nerovností a Centrální Limitní Věty

3. Statistik chce odhadnout průměrnou výšku h (v metrech) lidí v nějaké populaci, pomocí n nezávislých vzorků $X_1, \dots, X_n$, které vybíráme uniformně náhodně se všech možných lidí. Pro odhad použije výběrový průměr $S_n = (X_1 + \dots + X_n)/n$. Odhaduje, že směrodatná odchylka jednoho výběru je nejvýše 1 metr.
 - (a) Jak velké n má volit, aby směrodatná odchylka S_n byla nejvýše 1 cm?
 - (b) Pro jaké n zajistí Čebyševova nerovnost, že pravděpodobnost, že M_n se liší od h nejvýše o 5 cm s pravděpodobností alespoň 99 %?
4. Označme $S = \sum_{k=0}^{30} \binom{100}{k}$. Označme dále $X = \sum_{i=1}^{100} X_i$, kde X_i je ± 1 s pravděpodobností $1/2$ a veličiny $X_1, \dots, X_n$ jsou nezávislé.
 - (a) Vyjádřete S pomocí vhodné pravděpodobnosti výroku o X .
 - (b) Použijte CLV na odhad této pravděpodobnosti.
 - (c) Vyčíslete S vhodným softwarem a srovnejte.
5. Odhadněte $\binom{100}{30}$ pomocí CLV.

Podmíněná střední hodnota

6. Nechť X je n.v. s hustotou

$$f_X(x) = \begin{cases} x/4 & \text{pro } 1 < x \leq 3, \\ 0 & \text{jinak.} \end{cases}$$

Označme A jev $\{X \geq 2\}$.

- (a) Spočítejte $\mathbb{E}(X)$, $P(A)$, $f_{X|A}$ a $\mathbb{E}(X | A)$.
- (b) Označme $Y = X^2$. Spočítejte $\mathbb{E}(Y)$ a $\text{var}(Y)$.

Podmíněná hustota

7. Nechť X, Y mají sdruženou hustotu

$$f(x, y) = \begin{cases} e^{-y} & \text{pro } 0 < x < y < \infty, \\ 0 & \text{jinak.} \end{cases}$$

- (a) Určete podmíněnou hustotu $f_{X|Y}$.
- (b) Určete podmíněnou hustotu $f_{Y|X}$.

8. Metrový klacek zlomíme v uniformně náhodném bodě a ponecháme si levý kus. Jeho délku označíme Y . V něm opět vybereme uniformně náhodný bod, kde klacek zlomíme, a délku levého kusu označíme X .

- (a) Najděte sdruženou hustotu $f_{X,Y}$. Může vám pomoci podmíněná hustota $f_{X|Y}$.
- (b) Najděte marginální hustotu f_X .
- (c) Pomocí f_X spočítejte $\mathbb{E}(X)$.
- (d) Spočítejte $\mathbb{E}(X)$ pomocí vztahu $X = Y \cdot (X/Y)$.

9. Metrový klacek rozlomíme na tři kusy jedním z níže popsaných způsobů. Pro každý z nich spočítejte, jaká je pravděpodobnost, že ze získaných tří kusů jde sestavit trojúhelník. (Nápověda: napřed si rozmyslete, kdy jsou tři kladná čísla se součtem 1 stranami nějakého trojúhelníku.)

- (a) Vybereme uniformně náhodně dva body zlomu.
- (b) Vybereme uniformně náhodně první bod zlomu. Pak totéž uděláme s kusem klacku v pravé ruce.
- (c) Vybereme uniformně náhodně první bod zlomu. Pak totéž uděláme s větším kusem klacku.

10. Volme uniformně náhodně bod z trojúhelníku s vrcholy v bodech $[0, 0]$, $[0, 1]$ a $[1, 0]$, tj. pravděpodobnost každé podmnožiny je úměrná jejímu obsahu. Označme X, Y souřadnice zvoleného bodu.

- (a) Najděte sdruženou hustotu $f_{X,Y}$.
- (b) Najděte marginální hustotu f_Y .
- (c) Najděte podmíněnou hustotu $f_{X|Y}$.
- (d) Spočtěte $\mathbb{E}(X \mid Y = y)$ a podle věty o rozboru možností spočtěte $\mathbb{E}(X)$ (pomocí $\mathbb{E}(Y)$).
- (e) Spočtěte $\mathbb{E}(X)$ pomocí předchozí části a symetrie.

11. cvičení z PaSti – 2021-05-19

Bodové odhady

1. Máme náhodný výběr $X_1, \dots, X_n \sim U(0, \vartheta)$.

- (a) Navrhněte bodový odhad ϑ momentovou metodou.
- (b) Navrhněte bodový odhad ϑ metodou maximální věrohodnosti.
- (c) Pro každý z nich zjistěte, zda je nestranný a konzistentní.
- (d) Pro každý z nich spočítejte střední kvadratickou odchylku (MSE).
- (e) Který odhad je lepší? Napadá vás nějaký ještě lepší?

2. Máme náhodný výběr $X_1, \dots, X_n \sim \text{Exp}(\lambda)$. Označme $\vartheta = 1/\lambda$.

- (a) Navrhněte bodový odhad ϑ momentovou metodou.
- (b) Navrhněte bodový odhad ϑ metodou maximální věrohodnosti.
- (c) Pro každý z nich zjistěte, zda je nestranný a konzistentní.
- (d) Spočítejte střední kvadratickou odchylku (MSE).

3. Máme náhodný výběr $X_1, \dots, X_n \sim \text{Geom}(p)$.

- (a) Navrhněte bodový odhad p momentovou metodou.
- (b) Navrhněte bodový odhad p metodou maximální věrohodnosti.
- (c) Pro každý z nich zjistěte, zda je nestranný a konzistentní.

4. Máme náhodný výběr $X_1, \dots, X_n \sim \text{Exp}(\lambda)$. Zajímá nás pravděpodobnost p , že $X > 1$ pro $X \sim \text{Exp}(\lambda)$. (Připomeňme, že $p = e^{-\lambda \cdot 1}$.)

- (a) Navrhněte bodový odhad p (libovolnou metodou), případně několik odhadů.
- (b) Prozkoumejte jeho vlastnosti.

5. Nechť $X \sim \text{Exp}(\lambda)$ popisuje dráhu, kterou uletí radioaktivní částice, než se rozpadne. Náš přístroj její rozpad (a polohu rozpadu, tj. hodnotu X) zachytí, ale jen pokud $1 \leq X \leq 2$. Formálně, budeme zkoumat náhodný výběr $X_1, \dots, X_n \sim F_{X|B}$ pro jev $B = 1 \leq X \leq 2$.

- (a) Navrhněte bodový odhad λ momentovou metodou.
- (b) Navrhněte bodový odhad λ metodou maximální věrohodnosti.
- (c) Pro každý z nich zjistěte, zda je nestranný a konzistentní.

K procvičení

6. Máme náhodný výběr $X_1, \dots, X_n \sim \text{Pois}(\lambda)$.

- (a) Navrhněte bodový odhad λ momentovou metodou.
- (b) Navrhněte bodový odhad λ metodou maximální věrohodnosti.
- (c) Spočtěte střední kvadratickou odchylku (MSE).

7. Máme náhodný výběr $X_1, \dots, X_n \sim U(\vartheta, \vartheta + 1)$.

- (a) Navrhněte bodový odhad ϑ momentovou metodou.
- (b) Navrhněte bodový odhad ϑ metodou maximální věrohodnosti.
- (c) Pro každý z nich zjistěte, zda je nestranný a konzistentní.
- (d) Pro každý z nich spočtěte střední kvadratickou odchylku (MSE).
- (e) Který odhad je lepší? Napadá vás nějaký ještě lepší?

12. cvičení z PaSti – 2021-05-26

Z každé kapitoly zkuste aspoň jeden příklad.

Intervalové odhady

1. Máme jedno měření $X \sim N(\mu, 1)$. (Parametrem je tedy $\vartheta = \mu$.)
 - (a) Najděte intervalový odhad pro μ se spolehlivostí 95 %.
 - (b) Místo jednoho měření jich provedeme n (pochopitelně nezávislých). Jaký bude teď intervalový odhad pro μ ?
 - (c) Nechť X má stále střední hodnotu μ a rozptyl 1, ale není už nutně normální. Co se změní?
2. Tentokrát vybíráme z rozdělení $N(\mu, \sigma^2)$: μ ani σ neznáme, parametr $\vartheta = (\mu, \sigma)$. Naměřili jsme hodnoty 8.47, 10.91, 10.87, 9.46, 10.40.
 - (a) Spočtete výběrový průměr a výběrový rozptyl.
 - (b) Kdybychom věřili, že spočtený výběrový rozptyl je skutečná hodnota σ^2 , najděte intervalový odhad pro μ .
 - (c) Najděte intervalový odhad pro μ použitím Studentova t -rozdělení.
3. Počet emailů za den modelujeme pomocí Poissonova rozdělení $Pois(\lambda)$. První týden v prosinci jsme dostali postupně 34, 35, 29, 31, 30 emailů. Najděte pro λ intervalový odhad se spolehlivostí 95 %.

Použijte k tomu poslední metodu z přednášky – tu využívající Studentova rozdělení. (Poissonovo rozdělení sice není normální, ale pro dostatečně vysokou hodnotu λ je normálnímu dost podobné, metoda bude mít spolehlivost blízkou 95 %.)

Testování hypotéz

4. (Všimněte si podobnosti a rozdílu oproti prvnímu příkladu.) Máme jedno měření $X \sim N(\mu, 1)$. Chceme ověřit hypotézu $H_0: \mu = 5$ s hladinou významnosti $\alpha = 5\%$.
 - (a) Jaký zvolíme kritický obor – množinu měření, ve které hypotézu zamítneme?
 - (b) Místo jednoho měření jich provedeme n (pochopitelně nezávislých). Jaký bude kritický obor pro $\bar{X}_n$?
 - (c) Pokud je ve skutečnosti $\mu = 4$ a máme $n = 10$ měření, jaká je pravděpodobnost, že hypotézu nezamítneme?
 - (d) Nechť X má stále střední hodnotu μ a rozptyl 1, ale není už nutně normální. Co se změní?

5. Podle slibu výrobce bude jeho stroj dělat chyby nejvýše ve 3 % případů. Z 600 pokusů došlo k chybě v 28 případech. Posuďte slib výrobce (coby nulovou hypotézu) na hladině významnosti 5 %.

- (a) Počet chyb modelujte přesně, tj. pomocí binomického rozdělení.
- (b) Počet chyb modelujte přibližně pomocí normálního rozdělení (s vhodným μ , σ^2).

Test dobré shody

6. Otestujte náhodnost kostky. Můžete házet ručně, máte-li kostku po ruce, nebo využít vhodný generátor (na webu např. <https://www.random.org/dice/?num=60> – výsledek jde kopírovat a vložit jako posloupnost čísel – nebo softwarový generátor na vašem stroji). Použijte Pearsonův χ^2 test dobré shody.

7. Pojdme zkontrolovat, zda počty emailů jsou opravdu dobře modelovány Poissonovým rozdělením. Můžete použít svoje data, nebo použít data Roberta Šamala, za loňský listopad. (Nejsou to všechny emaily, jenom „ty důležité“ podle klasifikace Gmailu, ale to by na statistických vlastnostech nemělo nic měnit.)

0, 6, 14, 8, 8, 9, 3, 3, 12, 12, 15, 7, 15, 2, 5, 13, 5, 17, 15, 11, 9, 2, 16, 8, 9, 11, 6, 2, 2, 9.

Rozmyslete, jak použít test dobré shody, zjistěte velikost λ a test proveďte. Zamyslete se nad výsledkem.

13. cvičení z PaSti – 2021-06-02

Bayesovská statistika

1. Kvído zase píše test, tentokrát s deseti otázkami, každá má tři volby. Pro každou z otázek nastane (nezávisle na ostatních) jedna ze dvou stejně pravděpodobných možností: buď se Kvído tuto látku učil, otázku „umí“ a odpoví určitě správně, nebo se neučil a tipne uniformně náhodnou volbu.

- (a) Pokud první otázku odpověděl Kvído správně, jaká je pravděpodobnost, že tuto otázku uměl?
- (b) Jaká je apriorní distribuce (pravděpodobnostní funkce) počtu otázek, které Kvído umí?
- (c) Kvído odpověděl správně na šest otázek z deseti. Jaká je posteriorní distribuce počtu otázek, které Kvído uměl?

2. Stejný test píše několik studentů. Každý z nich patří (uniformně náhodně) do jedné ze tří skupin – všichni v první skupině umí otázku s pravděpodobností $\vartheta_1 = 0.3$, ve druhé s pravděpodobností $\vartheta_2 = 0.7$, a ve třetí s pravděpodobností $\vartheta_3 = 0.95$. Náhodně vybraný student odpoví správně k otázek (z deseti).

- (a) Metodou MAP rozhodněte, do které skupiny student patří.
- (b) U budiž počet otázek, které student umí. Student odpoví správně na 5 otázek. Odvoďte posteriorní distribuci, MAP odhad U a odhad U pomocí střední hodnoty.

3. Máme dvě krabičky, v první jsou dva bílé a jeden černý míček, ve druhé dva černé a jeden bílý. Vybereme náhodně jednu z krabiček (pravděpodobnost volby první krabičky je p) a pak vytáhneme míček.

- (a) Popište, jak metodou MAP rozhodnout podle barvy taženého míčku, zda jsme vybrali první nebo druhou krabičku.
- (b) Pro $p = 1/2$ spočtete pravděpodobnost chybného rozhodnutí. Srovnajte s chybou chybného rozhodnutí bez tahání míčku.

4. Házíme cinknutou mincí s pravděpodobností líce Θ . Naše apriorní rozdělení je dáno funkcí $f = f_\Theta$ takovou, že $f(0) = f(1) = 0$, $f(0.5) = 2$, f je lineární na intervalech $[0, 0.5]$ a $[0.5, 1]$ a f je nulová mimo $[0, 1]$.

Najděte MAP odhad Θ , pokud z n nezávislých hodů padl líc k -krát.

5. Předpokládejme, že policejní radar naměří rychlost auta vyšší o $U \sim U(0, 5)$. Předpokládejme, že skutečná rychlost auta je $V \sim U(45, 65)$. Metodou podmíněné pravděpodobnosti odhadněte rychlost auta na základě naměřené rychlosti.

6. Počet nákupních vozíků v obchodě je Θ – uniformně rozdělená náhodná veličina s hodnotami $\{1, 2, \dots, 100\}$. Každý vozík má číslo $(1, \dots, \Theta)$. Na našem vozíku (předpokládáme, že je uniformně náhodný) je číslo X . Odhadněte Θ

- (a) metodou MAP.
- (b) pomocí podmíněné pravděpodobnosti.

(Úloha je podobná příkladu o Romeovi a Julii z přednášky, ale tam šlo o spojité náhodné veličiny.)

7. Počet minut mezi příjezdy autobusů na zastávku u kolejí má exponenciální rozdělení s parametrem Θ , použijeme apriorní distribuci s hustotou $f_{\Theta}(\vartheta) = 10\vartheta$ pro $\vartheta \in [0, 1/\sqrt{5}]$ (a nula jinde – ověřte, že se jedná o hustotu).

- (a) Jdeme na zastávku a musíme čekat 30 minut. Jaká je posteriorní hustota a odhad Θ (metodou MAP nebo podmíněnou střední hodnotou)?
- (b) Provedeme pět měření (jdeme na autobus pětkrát a zapíšeme si dobu čekání). Musíme čekat 30, 25, 15, 40 a 20 minut, předpokládáme, že jednotlivé dny jsou nezávislá měření se stejným rozdělením. Jaká je posteriorní hustota a odhad Θ (metodou MAP nebo podmíněnou střední hodnotou)?

8. Varianta problému z přednášky: Máme nezávislé náhodné veličiny $X_1, \dots, X_n$. Všechny jsou uniformní na intervalu $[0, \vartheta]$, kde ϑ je hodnota n.v. Θ . Apriorní distribuce Θ je uniformní na $[0, 1]$. Na přednášce jsme měli $n = 1$, teď budeme předpokládat $n > 3$.

- (a) Najděte odhad Θ pomocí podmíněné pravděpodobnosti (při daných hodnotách $x_1, \dots, x_n$ veličiny $X_1, \dots, X_n$).
- (b) Nakreslete graf podmíněné střední kvadratické chyby (MSE) pro odhady MAP a podmíněné střední hodnoty, jako funkce $\bar{x} = \max\{x_1, \dots, x_n\}$, pro hodnotu $n = 5$.
- (c) Pokud $\bar{x} = 0.5$, jak se vyvíjejí odhad MAP a podmíněná střední hodnota, a odpovídající střední kvadratická chyba, v závislosti na $n \rightarrow \infty$?

The first rule of the Bayesian Conspiracy is that you talk about it exactly as much as a non-member would.