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. Nechť $X \sim Bin(m, p)$ a $Y \sim Bin(n, p)$ jsou n.n.v. Pak $X + Y \sim Bin(m + n, p)$.

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

6. (Kasino v St. Petěrburgu) Házíme opakovaně mincí. Pokud poprvé padla panna v n-tém hodu, dostaneme odměnu 2^n rublů. Kolik byste byli ochotní 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. Nechť X má uniformní[1] rozdělení na množině $\{a, a + 1, a + 2, \ldots, b\}$ (pro celá čísla $a < b$). Určete $\mathbb{E}(X)$ a $\mathrm{var}(X)$. Rozptyl je trochu ošklivý na dopočítávání, připomeňte si vzorec z diskrétky na součet druhých mocnin – nebo se omezte na konkrétní příklad $a = 1, b = 6$. Připomeňte si vzorec z přednášky:

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

---

[1] Tím 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 $X = \{x_1, \ldots, x_n\}$ 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 $k \le |X_<|$, 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 třídicí 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 linearitu střední hodnoty. A propos, v jakém pravděpodobnostním prostoru pracujeme?

---

[2] tedy 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 $\operatorname{var}(X)$, 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 $\mathbb{E}(X)$, $\operatorname{var}(X)$ a σ_X.

[3] *Směrodatná odchylka* σ_X veličiny X se definuje jako $\sqrt{\operatorname{var} X}$.