

6. CVIČENÍ Z DATOVÝCH STRUKTUR 1, ZS25/26

Hešujeme: opakování pravděpodobnosti a hešovací funkce univerzální a k -nezávislé

1. *Lehké opakování pravděpodobnosti.* Pat a Mat střílí šípky a snaží se trefit se do terče. Pravděpodobnost, že se trefí, je $p \in (0, 1]$, a je nezávislá na jakýchkoliv jejich předchozích pokusech (a stejná pro oba). Nechť X je náhodná veličina, která označuje počet pokusů do první trefy (včetně). Ukažte, že $\mathbb{E}[X] = \frac{1}{p}$.

2. *Pravděpodobnost kolize.* Ukažte, že v hešovací tabulce velikosti $m = n^2$ s n prvky dojde ke kolizi s pravděpodobností maximálně $\frac{1}{2}$, předpokládáme-li zcela náhodnou hešovací funkci. (Zkuste to dokázat bez použití přímo padnoucího pozorování z přednášky.)

3. *Pevné body permutací.* Mějme uniformně náhodnou permutaci na n prvcích. Určete střední hodnotu počtu pevných bodů této permutace.

4. *Černá krabička.* Dostali jste hešovací funkci $h : \mathcal{U} \rightarrow [m]$. Pokud o této funkci nevíte nic dalšího, kolik vyhodnocení funkce potřebujete, abyste zaručeně našli k -tici prvků, které se všechny zobrazí do téže přihrádky?

5. *Narozeniny.* Na přednášku je přihlášeno 178 lidí. Jaká je pravděpodobnost, že přesně jeden z nich má dnes narozeniny? Ignorujte přestupné roky, uvažujte, že všechny dny jsou stejně pravděpodobné pro narození.

Náhodnou veličinu označující počet studentů, kteří mají dnes narozeniny, označme X . Dále jako Y označme počet studentů, kteří mají narozeniny zítra. Jaký je rozdíl mezi X a Y ?

6. *Nezávislost a univerzalita.* Dokažte následující:

- pokud je systém hešovacích funkcí (k, c) -nezávislý, je také $(k - 1, c)$ -nezávislý (pro $k \geq 2$),
- pokud je systém hešovacích funkcí $(2, c)$ -nezávislý, je též c -univerzální.

7. *Vyloženě praktické systémy.* Uvažme systém funkcí $\mathcal{H}_1 = \{\text{id}\}$, který obsahuje jedinou funkci, jež zobrazí x na x . Je $\mathcal{H}_1$ c -univerzální pro nějaké c ? Je $\mathcal{H}_1$ (k, c) -nezávislý pro nějaká k a c ?

Dále uvažme systém $\mathcal{H}_2 = \{h_a(x) = a : a \in [m]\}$. Dokažte, že tento systém je $(1, 1)$ -nezávislý. Dále ukažte, že $\mathcal{H}_2$ není $(2, c)$ -nezávislý ani c -univerzální pro žádné c .

8. *Modulo univerzálního systému nemusí být univerzální.* Ukažte, že pokud máme univerzální systém hešovacích funkcí $\mathcal{H}$, pak systém $\mathcal{H}'$, kde ke každé fci navíc přidáme modulo m , už nemusí být univerzální. Formálně: Dokažte, že pro každé c a $m > c$ existuje univerzum $\mathcal{U}$ a systém $\mathcal{H}$ z $\mathcal{U}$ do $\mathcal{U}$ tak, že $\mathcal{H}$ je univerzální, ale $\mathcal{H}'$ už není c -univerzální.

9. *Zpět k I/O: Analýza algoritmu dle starověkého „Rozděl a panuj!“* Proveďte analýzu počtu přenesených bloků, tedy I/O složitost, pro algoritmus QuickSelect pro nalezení k -tého nejmenšího prvku s náhodným výběrem pivotu. Jelikož je pravděpodobnostní, určete střední hodnotu.

(Pro jednoduchost můžete nejprve odstranit náhodnost a předpokládat, že v každém kroku vybereme pivotu, který je *pseudomediánem*, tedy leží v prostředních dvou kvartilech.)

Užitečné pojmy

Tvrzení 1 (Union bound) Pro jevy A_1, A_2 platí, že $P[A_1 \cup A_2] \leq P[A_1] + P[A_2]$.

Tvrzení 2 (Linearita střední hodnoty) Pro náhodné veličiny X, Y a $\alpha, \beta \in \mathbb{R}$ platí: $\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y]$.

Definice 1 (Indikátor, nezávislost náhodných veličin) Nechť A je jev v diskrétním pravděpodobnostním prostoru. Potom indikátor A je náhodná veličina I_A definovaná: $I_A(\omega) = 0 \Leftrightarrow \omega \notin A$, jinak $I_A(\omega) = 1$.

Náhodné veličiny X, Y na diskrétním pravděpodobnostním prostoru $(\Omega, 2^\Omega, P)$ jsou nezávislé, pokud $\forall \alpha, \beta \in \mathbb{R}$ jsou jevy $\{\omega \in \Omega : X(\omega) \leq \alpha\}, \{\omega \in \Omega : Y(\omega) \leq \beta\}$ nezávislé.

Definice 2 (c -univerzální systém fci) Systém $\mathcal{H}$ funkcí $h : \mathcal{U} \rightarrow [m]$ je c -univerzální pro $c > 0$, pokud pro všechna $x \neq y$ platí $\Pr_{h \in \mathcal{H}}[h(x) = h(y)] \leq \frac{c}{m}$.

Systém $\mathcal{H}$ je univerzální, pokud je c -univerzální pro nějaké $c > 0$.

Definice 3 (k -nezávislý systém fci) Systém $\mathcal{H}$ funkcí $h : \mathcal{U} \rightarrow [m]$ je (k, c) -nezávislý pro nějaká $k \geq 1, c > 0$, pokud $\Pr_{h \in \mathcal{H}}[h(x_1) = a_1 \wedge \dots \wedge h(x_k) = a_k] \leq \frac{c}{m^k}$ pro libovolná $x_1, \dots, x_k$ různá, $a_1, \dots, a_k$ ne nutně různá.

Systém $\mathcal{H}$ je k -nezávislý, pokud je (k, c) -nezávislý pro nějakou nezávislou konstantu c .