8. cvičení

Datové struktury I, 20. 11. 2023 `https://iuuk.mff.cuni.cz/~chmel/2324/ds1/`

Úloha 1 (Lehké opakování pravděpodobnosti)
Matt střílí, a snaží se trefit se do terče. Protože je začátečník, pravděpodobnost, že se trefí, je $p \in (0,1]$, a je nezávislá na jakýchkoliv jeho předchozích pokusech. Nechť X je náhodná veličina, která označuje počet pokusů do Mattovy první trefy (včetně). Ukažte, že $\mathbb{E}[X] = \frac{1}{p}$.

Řešení
Z definice: $\mathbb{E}[X] = p + (1-p)(1+\mathbb{E}[X]) = 1 + (1-p)\mathbb{E}[X] \rightsquigarrow p\mathbb{E}[X] = 1 \rightsquigarrow \mathbb{E}[X] = \frac{1}{p}$.

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

Řešení
$P[\text{kolize}] = P[\exists i \neq j \in [n] : h(i) = h(j)] = P[\bigcup_{i \neq j \in [n]} (h(i) = h(j))] \leq \sum_{i \neq j \in [n]} P[(h(i) = h(j))] = \binom{n}{2} \cdot \frac{1}{m}$

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

Řešení
Použijeme indikátory: je-li F náhodná veličina určující počet pevných bodů náhodné permutace, můžeme psát $F = I_1 + I_2 + \ldots + I_n$, kde I_ℓ je indikátor jevu, že $\pi(\ell) = \ell$. Pak $\mathbb{E}[F] = \mathbb{E}[I_1 + I_2 + \ldots + I_n] = \mathbb{E}[I_1] + \mathbb{E}[I_2] + \ldots + \mathbb{E}[I_n]$ z linearity střední hodnoty, a $\mathbb{E}[I_\ell] = \frac{(n-1)!}{n!} = \frac{1}{n}$, a tedy $\mathbb{E}[F] = 1$.

Úloha 4 (Nezávislost a univerzalita)
Dokažte následující:

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

Řešení
- Chceme ukázat $(k-1,c)$-nezávislost, tedy máme dané $x_1, \ldots, x_{k-1} \in \mathcal{U}, a_1, \ldots, a_{k-1} \in [m]$. Zvolme si dále $x \neq x_i \forall i \in [k-1]$ (takové existuje z k-nezávislosti). Chceme $\Pr_h[h(x_1) = a_1 \wedge \ldots \wedge h(x_{k-1}) = a_{k-1}] = \sum\limits_{a \in [m]} \Pr_h[h(x_1) = a_1 \wedge \ldots \wedge h(x_{k-1}) = a_{k-1} \wedge h(x) = a] \leq \sum\limits_{a \in [m]} \frac{c}{m^k} = \frac{c}{m^{k-1}}$.
- Mějme tedy $x \neq y \in \mathcal{U}$. Chceme omezit shora $\Pr_h[h(x) = h(y)] = \sum\limits_{a \in [m]} \Pr_h[h(x) = a \wedge h(y) = h] \leq \sum\limits_{a \in [m]} \frac{c}{m^2} = \frac{c}{m}$.

Úloha 5 (Černá krabička)
Dostali jste hashovací funkci $h : \mathcal{U} \to [m]$. Pokud o této funkci nevíte nic dalšího, kolik vyhodnocení funkce potřebujete, abyste zaručeně našli k-tici prvků, které se všechny zobrazí do téže přihrádky?

Řešení
Princip holubníku: máme m děr, v každé je $k-1$ holubů, a ještě jeden pokus navíc nám garantuje, že se už opravdu někam musíme trefit, tedy $1 + m(k-1)$.

Úloha 6 (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 $\mathcal{H}_2 = \{h_a(x) = a \colon a \in [m]\}$. Dokažte, že tento systém je (1,1)-nezávislý. Dále ukažte, že $\mathcal{H}_2$ není $(2,c)$-nezávislý ani c-univerzální pro žádné c.

Řešení
$\mathcal{H}_1$ je ε-univerzální pro každé $\varepsilon > 0$. Problém je, že $\Pr[h(x) = x] = 1$, a tedy, pokud $|\mathcal{U}| > 1$, pak nemůže být jakkoliv nezávislá.
U druhého systému: (1,1)-nezávislost plyne z toho, že $\Pr[h_a(x) = b] = \frac{1}{m}$, protože volíme jednu konkrétní volbu a. Na druhou stranu, pro $x \neq y$ máme $\Pr[h_a(x) = b \wedge h_a(y) = b] = \frac{1}{m} > \frac{c}{m^2}$ pro jakoukoliv konstantu, a tedy nezávislost je nemožná. Pro c-univerzalitu, evidentně $\Pr[h_a(x) = h_a(y)] = 1$, a tedy c-univerzalitu také nemáme.

Úloha 7 (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 $m > 1$ 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.

Řešení
Uvažme $\mathcal{H}_1$ z předchozí úlohy - pak $\mathcal{H}_1 \mod m$ nemůže být c-univerzální, protože dokud $m < |\mathcal{U}|$, pak prvky 1 a $m + 1$ se vždycky zobrazí na tentýž prvek 1.

Užitečné definice

Definice (c-univerzální systém fcí)**.** Systém $\mathcal{H}$ funkcí $h : \mathcal{U} \to [m]$ je c-univerzální pro $c > 0$, pokud pro všechna $x \neq y$ platí $\Pr_h[h(x) = h(y)] \leq \frac{c}{m}$.
Systém $\mathcal{H}$ je univerzální, pokud je c-univerzální pro nějaké $c > 0$.

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

Opáčko pravděpodobnosti

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

Tvrzení (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 (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.

Věta (Markovova nerovnost)**.** Buď X nezáporná náhodná veličina. Pak $\forall \varepsilon > 0$ platí $P[X \geq \varepsilon] \leq \frac{\mathbb{E}[X]}{\varepsilon}$.
Ekvivalentně pro jakékoliv $d \geq 1$, $P[X \geq d \cdot \mathbb{E}[X]] \leq \frac{1}{d}$.

Věta (Chernoffova mez)**.** Mějme $X = \sum_{i=1}^{k} X_i$, kde X_i jsou nezávislé Bernoulliovské náhodné veličiny, $\mu = \mathbb{E}[X], c > 1$. Pak $P[X \geq c\mu] \leq \left(\frac{e^{c-1}}{c^c}\right)^\mu$.