9. cvičení

Datové struktury I, 29. 11. 2022 https://iuuk.mff.cuni.cz/~chmel/2223/ds1/

Úloha 1 (Lineární systém)
Ukažte, že systém funkcí $\{h_{a,b}(x) = (ax + b \mod p) \mod m \colon a, b \in \mathbb{Z}_p\}$ je dokonce 1-univerzální, máme-li $0 < a < p$. Pro jak malé c je tento systém $(2, c)$-nezávislý?

Řešení
Univerzalita: Mějme $x \neq y$ pevné. Pro $a, b \in \mathbb{Z}_p : r = (ax + b) \mod p, s = (ay + b) \mod p$, Máme bijekci $(a, b) \mapsto (r, s)$. Pak (r, s) jsou rovnoměrně rozdělené. Kolize $h_{a,b}(x) = h_{a,b}(y)$ nastane, právě když $r \equiv_m s$, řekneme, že (r, s) je špatná dvojice. Tedy pro dané r máme nejvýše $\lceil p/m \rceil - 1$ špatných s.
Tedy $\Pr[h_{a,b}(x) = h_{a,b}(y)] \leq \frac{p \cdot (\lceil p/m \rceil - 1)}{p(p-1)} \leq \frac{\frac{p-1}{m}}{p-1} = \frac{p-1}{(p-1)m} = \frac{1}{m}$.
Nezávislost: máme $x \neq y$. Pak pro libovolné $z_1, z_2 \in \mathbb{Z}_m$: $\Pr[h_{a,b}(x) = z_1 \wedge h_{a,b}(x) = z_2] \leq \frac{\lceil p/m \rceil}{m^2}$.

Úloha 2 (Lineární systém bez konstanty)
Uvažme systém funkcí $\{h_a(x) = (ax \mod p) \mod m \colon a \in \mathbb{Z}_p \setminus \{0\}\}$. Je k-univerzální pro nějaké k? A co když dovolíme $a = 0$?

Řešení
Je 2-univerzální: $h_a(x) = h_a(y)$ odpovídá $(ax \mod p) \mod m \equiv (ay \mod p) \mod m$, tedy $((ax - ay) \mod p) \mod m \equiv 0$. To můžeme přepsat jako $a(x - y) \mod p = \ell m$ pro $l \in \{-\lfloor p/m \rfloor, \ldots, \lfloor p/m \rfloor\} - \{0\}$ ($\ell \neq 0$, bo $a \neq 0$). Speciálně tedy pro každou volbu x a y máme maximálně $2\lfloor p/m \rfloor$ funkcí, ve kterých můžeme mít kolizi. Tím pádem $|\{h : h(x) = h(y)\}| \leq \frac{2p}{m}$, a pravděpodobnost je tedy $\leq \frac{2}{m}$.
Pro $a = 0$: provedeme to samé, ale ℓ může být až $2\lfloor p/m \rfloor + 1$, a tedy máme odhad $|\{h : h(x) = h(y)\}| \leq \frac{3p}{m}$.

Úloha 3 (Lineární systém není 3-nezávislý)
Ukažte, že systém funkcí $\{h_{a,b}(x) = (ax + b \mod p) \mod m \colon a, b \in \mathbb{Z}_p\}$ není 3-nezávislý.

Řešení
Mějme $x \neq y \neq z \neq x$ a označíme pro konkrétní fixní a, b hodnoty $v_1 = ax + b, v_2 = ay + b, v_3 = az + b$. Pak $\Pr[h(x) = v_1 \wedge h(y) = v_2 \wedge h(z) = v_3] = \Pr[h(z) = v_3 | h(x) = v_1 \wedge h(y) = v_2] \cdot \Pr[h(x) = v_1 \wedge h(y) = v_2] = \Pr[h(x) = v_1 \wedge h(y) = v_2] = \frac{1}{m^2}$.

Úloha 4 (Vyloženě praktické systémy)
Uvažme systém funkcí $\mathcal{H}_1 = \{\mathrm{id}\}$, 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 (resp. zopakujte si z přednášky), 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 5 (Univerzální modulo může rozbít univerzalitu)
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
Až na příštím cvičení :)

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. Sytém je k-nezávislý, pokud je (k, c)-nezávislý pro nějaké c nezávislou konstantu.