6. CVIČENÍ Z ÚVODU DO APROXIMACÍ

Hašování, hashování a hešování. A ještě po dvou nezávislé proměnné.

D: Množina náhodných proměnných $X_1, \dots X_n$ je *po k nezávislá*, jestliže pro každou podmnožinu $I \subseteq \{1, \dots, n\}$ s $|I| \leq k$ a pro každé hodnoty c_i, platí nezávislost pravděpodobností:

$$Pr[\bigwedge_{i \in I}(X_i = c_i)] = \prod_{i \in I} Pr[X_i = c_i].$$

Hašovací funkce budou pro nás funkce $h\colon U \to HT$, kde U je univerzum s $|U| = m$ a HT je hašovací tabulka s $|HT| = n$.

D: Rodina funkcí $\mathcal{H}$ je *(slabě) k-univerzální*, jestliže pro každé navzájem různé $x_1, x_2, \dots x_k \in U$ a hašovací funkci h vybranou uniformně náhodně z $\mathcal{H}$ platí

$$Pr_h[h(x_1) = h(x_2) = \ldots = h(x_k)] \leq \frac{1}{n^{k-1}}.$$

To znamená, že pro náhodnou h ze slabě k-univerzální rodiny je malá pravděpodobnost, že se k prvků zahašuje do stejné buňky HT.

D:Rodina funkcí $\mathcal{H}$ je *silně k-univerzální*, jestliže pro každé navzájem různé $x_1, x_2, \dots x_k \in U$, pro každé (ne nutně různé) hodnoty $y_1, y_2, \dots, y_k \in HT$ a hašovací funkci h vybranou uniformně náhodně z $\mathcal{H}$ platí

$$Pr_h[h(x_1) = y_1 \wedge h(x_2) = y_2 \wedge \ldots \wedge h(x_k) = y_k] = \frac{1}{n^k}.$$

Neboli, rodina je *silně k-univerzální*, jestliže můžeme zahašovat k prvků uniformně náhodnou hašovací funkcí a jejich buňky se budou chovat, jako bychom je vybrali uniformně náhodně.

PŘÍKLAD PRVNÍ Jednou ze zajímavých aplikací této teorie je derandomizace pomocí po dvou nezávislých náhodných proměnných (či jen bitů) místo plně nezávislých proměnných. Ukažte, že můžeme vygenerovat hodně po dvou nezávislých náhodných bitů pomocí jen několika nezávislých náhodných bitů, tedy binárních náhodných proměnných, kde obě hodnoty mají pravděpodobnost 0.5.

Přesněji, máme k plně nezávislých náhodných bitů $x_1, x_2, x_3, \dots, x_k$ a chceme vytvořit $2^k - 1$ po dvou nezávislých náhodných bitů $y_1, y_2, \dots, y_{2^k-1}$.

PŘÍKLAD DRUHÝ Mějte k náhodných bitů. Definujme $X_{i,j}$ pro $1 \leq i < j \leq k$ jako indikátor, jestli i-tý a j-tý náhodný bit jsou stejné. Ukažte, že $X_{i,j}$ jsou po dvou nezávislé, ale ne po třech nezávislé.

PŘÍKLAD TŘETÍ Na přednášce jste viděli, že rodina funkcí $h_{a,b}(x) = ax + b \mod p$ je silně 2-univerzální, když U i HT mají stejnou velikost rovnou prvočíslu p.

To není moc praktické, protože většinou potřebujeme, aby U bylo obrovské a hašovací tabulka rozumně velká. Proto předpokládejme, že $|U| = m$, $|HT| = n$ a $p \geq m$. Dokažte, že skoro stejná rodina $\mathcal{H} = \{h_{a,b} | 1 \leq a \leq p-1, 0 \leq b \leq p-1\}$ s funkcemi

$$h_{a,b}(x) = (ax + b \mod p) \mod n$$

je slabě 2-univerzální.

Důkaz může jít asi takto: pro dané $x_1 \neq x_2$, chceme spočítat počet dvojic (a, b), které způsobí, že x_1 a x_2 se zahašují na stejnou pozici.

1. Použijte silnou 2-univerzalitu nebo přímý argument, že pro danou čtveřici x_1, x_2 a $c, d \in \{0, \ldots p-1\}$, existuje právě jedna dvojice (a, b) taková, že

$$ax_1 + b = c \mod p \qquad \text{a} \qquad ax_2 + b = d \mod p.$$

2. Ukažte, že místo počítání dvojic (a, b) způsobujících kolize můžeme počítat dvojice (c, d) takové, že $c \neq d$ a $c = d \mod n$.
3. No a nakonec tyto dvojice (c, d) spočtěte.

PŘÍKLAD ČTVRTÝ Víme, že k nezávislých náhodných bitů stačí pro vygenerování $2^k - 1$ po dvou nezávislých náhodných bitů. Otázkou tedy je, kolik jich vygeneruje, pokud mají být po *třech* nezávislé.

Překvapivě lze vygenerovat 2^{k-1} po třech nezávislých náhodných bitů. Zbývá ukázat jak.

Tip: Podobná konstrukce jako v prvním příkladu.