Definice 1 (Data Stream model)**.** Uvažujeme algoritmy, které počítají na vstupu $\sigma = a_1, a_2, \ldots, a_m$ kde jednotlivé tokeny jsou čísla $a_i \in \{1, 2, \ldots, n\}$. Stream dostáváme postupně (online) a chceme spočítat nějakou funkci ze streamu v co nejmenším prostoru vzhledem k n, m a případně dalším parametrům. $\mathcal{A}(\sigma)$ nechť je výstup randomizovaného streamovacího algoritmu na vstupním streamu σ, který má počítat funkci $\varphi(\sigma)$. Řekneme, že $\mathcal{A}$ (ε, δ)-aproximuje φ, pokud:

$$\Pr\left[\left|\frac{\mathcal{A}(\sigma)}{\varphi(\sigma)} - 1\right| > \varepsilon\right] < \delta$$

Někdy nás zajímá nějaká statistika σ jako multimnožiny, pak definujeme frekvenční vektor $\vec{f} = f_1, f_2, \ldots, f_n$ takto: $f_j = |\{i \mid a_i = j\}|$ (počet výskytů j v σ).

Definice 2 (PRAS, FPRAS)**.** *Polynomiální randomizované aproximační schéma* (PRAS) pro problém P je randomizovaný algoritmus $\mathcal{A}$, který na vstupu x a $\varepsilon > 0$ běží v čase $|x|^{\mathcal{O}(1)}$ a vydá hodnotu $\mathcal{A}(x)$ splňující $\Pr\left[(1-\varepsilon)\#x \leq \mathcal{A}(x) \leq (1+\varepsilon)\#x\right] \geq \frac{3}{4}$. FPRAS je PRAS, který běží v čase polynomiálním v $|x|$ i $1/\varepsilon$.

Věta 1 (Estimator Theorem)**.** Polož $\rho = |G|/|U|$. Existuje Monte Carlo metoda, která dává ε-aproximaci $|G|$ s pravděpodobností aspoň $1-\delta$, pokud $N \geq \frac{4}{\varepsilon^2 \rho} \ln \frac{2}{\delta}$, kde N je počet nezávislých náhodných vzorků z univerza U.

Značení:

- $G = (U \cup V, E)$ je bipartitní graf, $|U| = |V| = n$
- m_k značí počet párování velikosti k v grafu G (*k-párování*)
- pro hranu $e \in E$ označme m_e počet k-párování obsahujících hranu e a m_{ne} počet k-párování neobsahujících hranu e
- $r_k = m_k / m_{k-1}$

Příklady

1. Streamovací algoritmus na zjištění počtu různých prvků v sekvenci: tedy chceme zjistit $d = |\{j \in [n] \mid f_j > 0\}|$.

Vstup : stream σ
Výstup : odhad počtu různých prvků v σ
Inicializace:
$h\colon [n] \to [n]$ náhodná 2-univerzální hashovací funkce
$z = 0$
Zpracuj : zpracování tokenu j
`//` z `je zatím největší počet nul na konci binárního zápisu` $h(j)$
$z = \max(z, \max\left\{i \mid 2^i \text{ dělí } h(j)\right\})$
return $2^{z+\frac{1}{2}}$

Nechť $X_{r,j}$ je náhodná veličina, která je indikátorem jevu 2^r dělí $h(j)$ (pozor, že i vyšší mocnina může dělit $h(j)$). Nechť $Y_r = \sum_{j : f_j > 0} X_{r,j}$ je také náhodná veličina.

- Spočítejte střední hodnotu Y_r a rozptyl Y_r.
- Pomocí Markovovy nerovnosti ukažte, že $\Pr[Y_r > 0] \leq \frac{d}{2^r}$.
- Pomocí Čebyševovy nerovnosti ukažte, že $\Pr[Y_r = 0] \leq \frac{2^r}{d}$.
- Nechť a je nejmenší celé číslo, že $2^{a+\frac{1}{2}} \geq 3d$. Ukažte, že pravděpodobnost, že náš odhad je větší nebo rovný $3d$ nebo menší nebo rovný $d/3$ je nejvýše $2\frac{\sqrt{2}}{3}$.

- Pokud běžíme k nezávisle náhodných instancí tohoto algoritmu paralelně a odpovíme medián, jak máme zvolit k, aby pravděpodobnost chyby v předchozím bodě byla nejvýš δ.

2. Ukažte, že pokud máme ε-aproximaci $\hat{s}$ čísla s a ε-aproximaci $\hat{t}$ čísla t, potom $\hat{s}/\hat{t}$ je 4ε-aproximace čísla s/t pro dostatečně malé ε.

3. Mějme dáno $\varepsilon > 0$. Nalezněte **vhodnou** volbu $\bar{\varepsilon}$ takovou, že pokud vezmeme $\bar{\varepsilon}$-aproximace $(\hat{a_i})_{i=1}^n$ čísel (a_i) tak $\prod_{i=1}^n \hat{a_i}$ je ε-aproximace čísla $\prod_{i=1}^n a_i$.

4. Buď $\alpha \geq 1$ reálné číslo takové, že $1/\alpha \leq r_k \leq \alpha$. Vyber $N = n^7\alpha$ prvků z $M_k \cup M_{k-1}$ nezávisle náhodně. Položme $\hat{r_k}$ podíl pozorovaných k-párování ku $(k-1)$-párování. Ukažte, že $(1 - 1/n^3)r_k \leq \hat{r_k} \leq (1 + 1/n^3)r_k$ s pravděpodobností alespoň $1 - c^{-n}$ pro nějakou konstantu $c > 1$.

5. Buď $G = (U \cup V, E)$ je bipartitní graf, $|U| = |V| = n$ s $\delta(G) > n/2$. Ukažte, že $r_k \leq n^2$.

6. Buď $G = (U \cup V, E)$ je bipartitní graf, $|U| = |V| = n$ s $\delta(G) > n/2$. Ukažte, že pro libovolné párovámí m velikosti nanejvýš $n - 1$ existuje zlepšující cesta délky nanejvýš 3.

7. Buď $G = (U \cup V, E)$ je bipartitní graf, $|U| = |V| = n$ s $\delta(G) > n/2$. Ukažte, že pro libovolné $2 \leq k \leq n$ a párovámí m velikosti k existuje nanejvýš n^2 párování velikosti $k-1$ takových, že pro každé z nich je možné naleznout zlepšující cestu délky nanejvýš 3, která je zlepší na m.

8. Buď $G = (U \cup V, E)$ je bipartitní graf, $|U| = |V| = n$ s $\delta(G) > n/2$. Ukažte, že $1/n^2 \leq r_k \leq n^2$. (Použijte předchozí 3 cvičení.)

9. Graf G_k vznikne z grafu $G = (U \cup V, E)$ tak, že přidáme $n - k$ vrcholů do každé partity a spojíme každý nový vrchol se všemi starými vrcholy v opačné partitě.

   Ukažte, že pro R podíl perfektních a skoroperfektních párování v G_k platí

$$R = \frac{m_k}{m_{k+1} + 2(n-k)m_k + (n-k+1)^2 m_{k-1}}.$$