

HOMEWORK SECTION 4: DATA STREAM

SV Nguyễn Văn Quyết¹

¹ MSSV: 20122310

²CNTT 2.04 - Bộ môn Khoa Học Máy Tính - Đại học Bách Khoa Hà Nội

1. Solution problem 4

Trong chứng minh ta sử dụng 2 tính chất sau đây:

1. Cho một họ hàm băm $H = \{h : U \rightarrow [m]\}$ là k-independent nếu với k phần tử khóa phân biệt $(x_1, x_2, \dots, x_k) \in U^k$ và k giá trị hàm băm (không nhất thiết phân biệt) $(y_1, y_2, y_3 \dots y_k) \in [m]^k$, ta có:

$$Pr[h(x_1) = y_1 \wedge h(x_2) = y_2 \wedge \dots \wedge h(x_k) = y_k] = \frac{1}{m^k} \forall h \in H$$

2. Với X là một biến ngẫu nhiên và với $a \geq 0$, ta có bất đẳng thức Markov:

$$Pr(|X| \geq a) \leq \frac{E(|X|)}{a}$$

Áp dụng ta có:

$$Pr(\tilde{F}[i] \leq F[i] + \epsilon t) = 1 - Pr(\tilde{F}[i] \geq F[i] + \epsilon t)$$

$$= 1 - Pr(c_{j, h_j(i)} \geq F[i] + \epsilon t) \forall 1 \leq j \leq \lceil \log(\frac{1}{\delta}) \rceil \text{ bởi vì } \tilde{F}[i] = \min(c_{j, h_j(i)})$$

$$= 1 - Pr(\bigwedge_{1 \leq j \leq \lceil \log(\frac{1}{\delta}) \rceil} c_{j, h_j(i)} \geq F[i] + \epsilon t) = 1 - \prod_1^{\lceil \log(\frac{1}{\delta}) \rceil} Pr(c_{j, h_j(i)} \geq F[i] + \epsilon t) \text{ (áp dụng 1)}$$

$$= 1 - \prod_1^{\lceil \log(\frac{1}{\delta}) \rceil} Pr(c_{j, h_j(i)} - F[i] \geq \epsilon t)$$

Áp dụng bất đẳng thức Markov ta có:

$$Pr(c_{j, h_j(i)} - F[i] \geq \epsilon t) \leq E\left(\frac{c_{j, h_j(i)} - F[i]}{\epsilon t}\right) \leq \frac{1}{e}$$

Từ đó suy ra

$$Pr(\tilde{F}[i] \leq F[i] + \epsilon t) \geq 1 - \frac{1}{e^{\lceil \log(\frac{1}{\delta}) \rceil}} = 1 - \delta \text{ (đpcm)}$$