

# **Data Streams**

**Làm bài tập Data Streams**

**Nguyễn Văn Đạt 20121499**

Ngày 28 tháng 11 năm 2016

# Mục lục

<b>1</b>	<b>Lời giải</b>	<b>2</b>
1.1	Câu (a) . . . . .	2
1.2	Câu (b) . . . . .	2

# Chương 1

## Lời giải

### 1.1 Câu (a)

$$\begin{aligned} Pr(\tilde{F}[i] \leq F[i] + \varepsilon t) &= 1 - Pr(\tilde{F}[i] \geq F[i] + \varepsilon t) \\ &= 1 - Pr(c_{j, h_j(i)} \geq F[i] + \varepsilon t \ \forall 1 \leq j \leq \lceil \log(\frac{1}{\delta}) \rceil) \\ &= 1 - \prod_{j=1}^{\lceil \log(\frac{1}{\delta}) \rceil} Pr(c_{j, h_j(i)} \geq F[i] + \varepsilon t) \end{aligned}$$

Sử dụng bất đẳng thức Markov và các tính chất mà đề bài đã cho.. Ta có :

$$Pr(c_{j, h_j(i)} \geq F[i] + \varepsilon t) \leq \frac{E[c_{j, h_j(i)} - F[i]]}{\varepsilon t} \leq \frac{\varepsilon t/e}{\varepsilon t} = \frac{1}{e}$$

Từ đó:

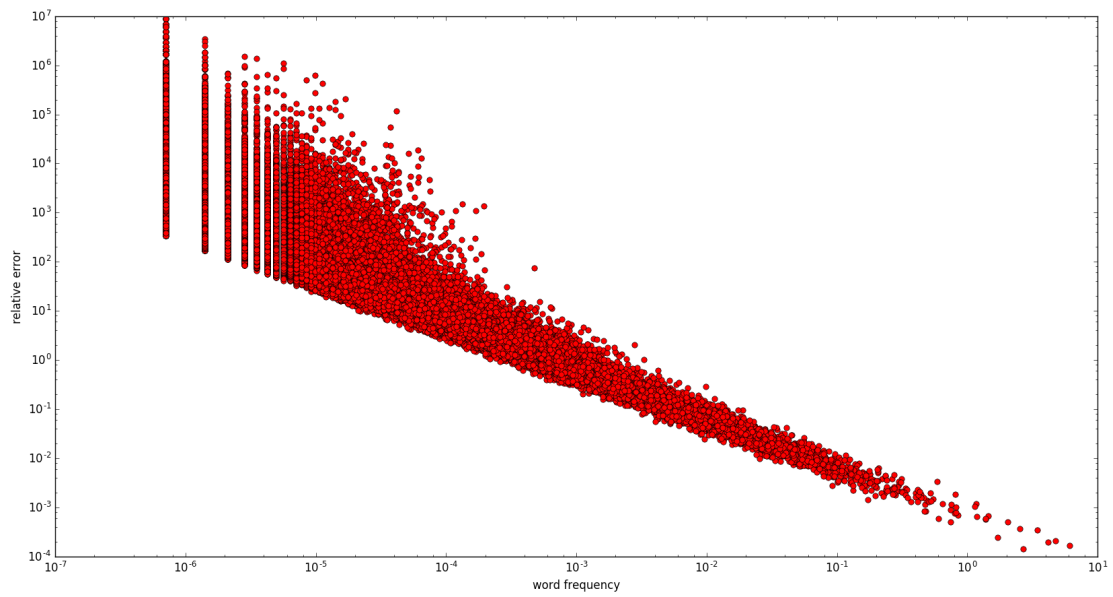
$$Pr(\tilde{F}[i] \leq F[i] + \varepsilon t) \geq 1 - (1/e)^{\lceil \log(\frac{1}{\delta}) \rceil}$$

Suy ra:

$$Pr(\tilde{F}[i] \leq F[i] + \varepsilon t) \geq 1 - \delta \ (\square)$$

### 1.2 Câu (b)

Ta có đồ thị biểu diễn quan hệ giữa tần số (frequency) của từ và sai số tương đối (relative error) ứng với từ đó:



Nhìn vào đồ thị ta thấy thuật toán làm việc tốt với các từ có tần số (frequency) lớn. Cụ thể những từ có tần số lớn hơn  $10^{-5}$  thì có sai số tương đối nhỏ hơn 1.