# Streaming space complexity of nearly all functions of one variable

Vladimir Braverman, Stephen Chestnut, David P. Woodruff, Lin F. Yang

January 7, 2016

A stream of $m = 7$ items from $[n] = [4]$

$$4, \ 2, \ 3, \ 2, \ 4, \ 2, \ 2$$

$$f = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\sum f_i^2 = 0$$

A stream of $m = 7$ items from $[n] = [4]$

$$4, \ 2, \ 3, \ 2, \ 4, \ 2, \ 2$$

$$f = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\sum f_i^2 = 1$$

A stream of $m = 7$ items from $[n] = [4]$

$$2,\ \ 3,\ \ 2,\ \ 4,\ \ 2,\ \ 2$$

$$f = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

$$\sum f_i^2 = 2$$

A stream of $m = 7$ items from $[n] = [4]$

$$3, \quad 2, \quad 4, \quad 2, \quad 2$$

$$f = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\sum f_i^2 = 3$$

A stream of $m = 7$ items from $[n] = [4]$

$2, \quad 4, \quad 2, \quad 2$

$$f = \begin{bmatrix} 0 \\ 2 \\ 1 \\ 1 \end{bmatrix}$$

$$\sum f_i^2 = 6$$

A stream of $m = 7$ items from $[n] = [4]$

$$4, \quad 2, \quad 2$$

$$f = \begin{bmatrix} 0 \\ 2 \\ 1 \\ 2 \end{bmatrix}$$

$$\sum f_i^2 = 9$$

A stream of $m = 7$ items from $[n] = [4]$

$$2, \quad 2$$

$$f = \begin{bmatrix} 0 \\ 3 \\ 1 \\ 2 \end{bmatrix}$$

$$\sum f_i^2 = 14$$

A stream of $m = 7$ items from $[n] = [4]$

$$2$$

$$f = \begin{bmatrix} 0 \\ 4 \\ 1 \\ 2 \end{bmatrix}$$

$$\sum f_i^2 = 21$$

A stream of $m = 7$ items from $[n] = [4]$

$$f = \begin{bmatrix} 0 \\ 4 \\ 1 \\ 2 \end{bmatrix}$$

$$\sum f_i^2 = 21$$

**How much storage for a streaming $(1 \pm \epsilon)$-approximation to $\sum_i f_i^2$?**

## Classify $g : \mathbb{Z}_{\geq 0} \to \mathbb{R}$

Is there a streaming $(1 \pm \epsilon)$-approximation for $\sum_i g(f_i)$ using only $\text{poly}(\frac{1}{\epsilon} \log nm)$ bits?

**Previous works**

- $g(x) = \mathbf{1}(x \neq 0)$: [FM85],[KNW10]
- $g(x) = x^p$: [F85],[AMS96],[IW05],[I06]
- $g(x) = x \log x$: [CDM06],[CCM07],[HNO08]
- monotonic $g$: [BO10],[BC15]

$$\epsilon = \Omega(\frac{1}{\text{polylog}(n)})$$
$$m = \text{poly}(n)$$
$$g(0) = 0$$
$$g(x) > 0, \ \forall x > 0$$

# Recursive Subsampling [Indyk & Woodruff 2005]

An $\alpha$-heavy hitter is any item $i^*$ such that $g(f_{i^*}) \geq \alpha \sum_i g(f_i)$.

**Theorem (Braverman & Ostrovsky 2010)**

$\frac{\epsilon^2}{\log^3 n}$*-heavy hitters* $\Rightarrow$ $(1 \pm \epsilon)$*-approximation to* $\sum_i g(f_i)$.

# Recursive Subsampling [Indyk & Woodruff 2005]

An $\alpha$-heavy hitter is any item $i^*$ such that $g(f_{i^*}) \geq \alpha \sum_i g(f_i)$.

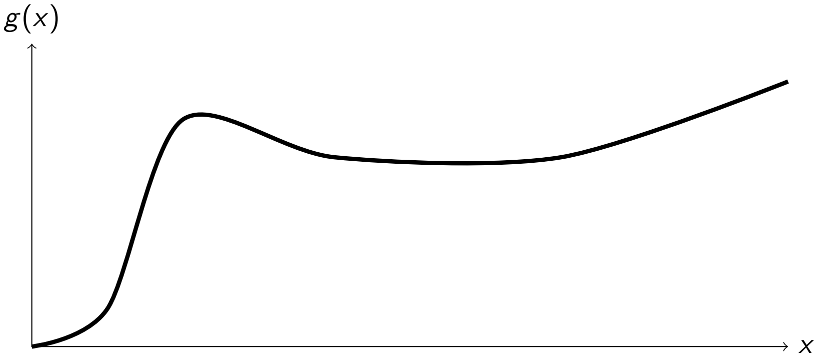
**Theorem (Braverman & Ostrovsky 2010)**

$$\frac{\epsilon^2}{\log^3 n}\text{-heavy hitters} \quad \Rightarrow \quad (1 \pm \epsilon)\text{-approximation to } \sum_i g(f_i).$$

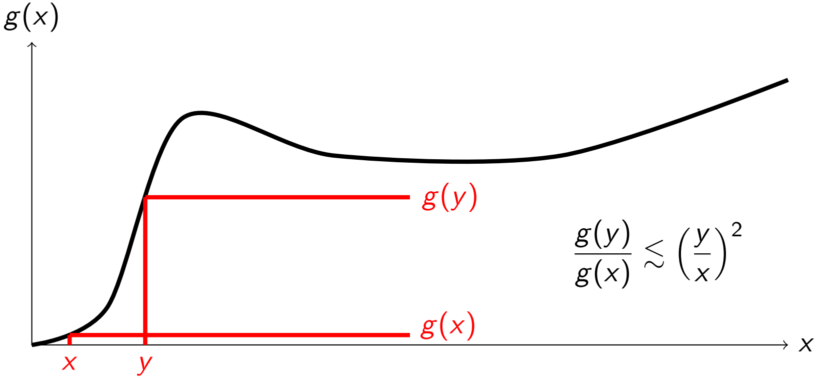
Heavy hitters by **CountSketch**[Charikar, Chen & Farach-Colton 2002]

- Find $i^*$ such that $f_{i^*}^2 \geq \alpha \sum_i f_i^2$
- Estimate $f_{i^*}$
- $O(\alpha^{-1} \log^2 n)$ bits.

Three properties are **sufficient** and **almost necessary** for $\tilde{O}(1)$ bits



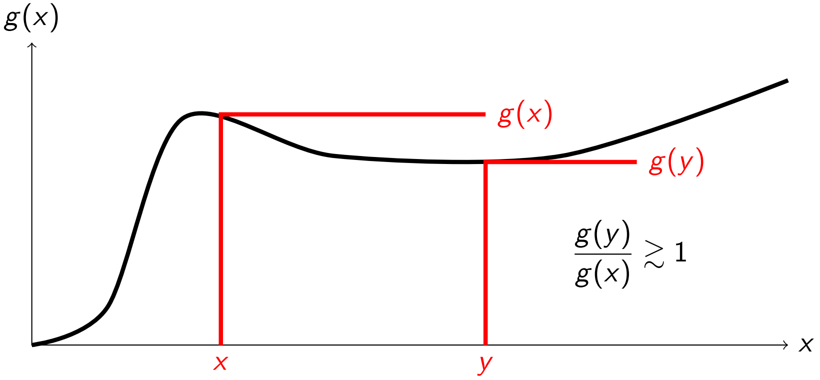
# Slow-jumping




$$\frac{g(y)}{g(x)} \lesssim \left(\frac{y}{x}\right)^2$$

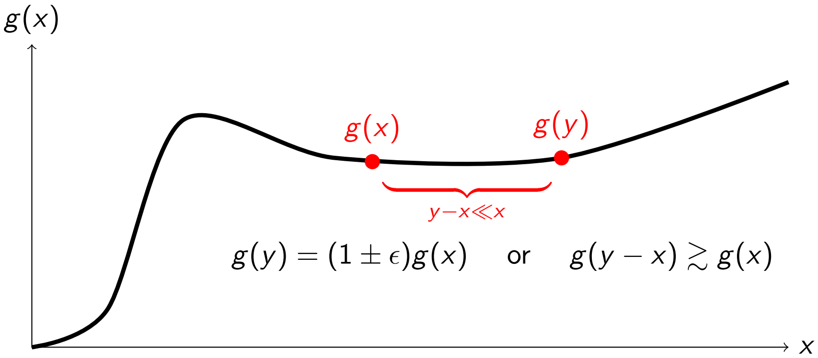
YES: $g(x) = x^2 \log x$ NO: $g(x) = x^3$

## Slow-dropping




YES: $g(x) = \Theta(\frac{1}{\log x})$ NO: $g(x) = \Theta(\frac{1}{x})$

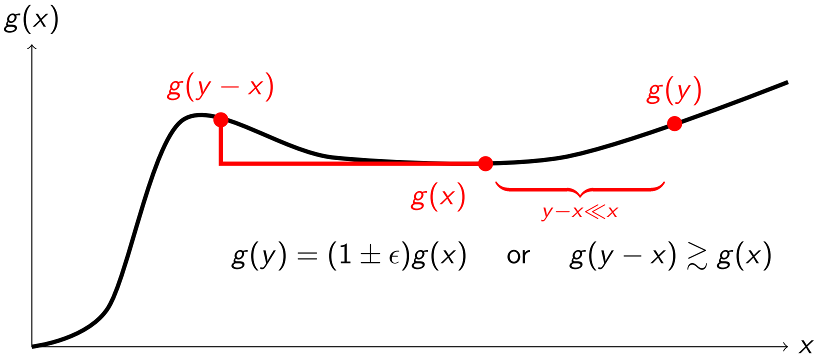
## Predictable




$$g(y) = (1 \pm \epsilon) g(x) \quad \text{or} \quad g(y - x) \gtrsim g(x)$$

YES: $g(x) = (2 + \sin x)\mathbf{1}(x > 0)$ NO: $g(x) = (2 + \sin x)x^2$

## Predictable




YES: $g(x) = (2 + \sin x)\mathbf{1}(x > 0)$ NO: $g(x) = (2 + \sin x)x^2$

Three properties are **sufficient** and **almost necessary** for $\widetilde{O}(1)$ bits

**slow-jumping** $\frac{g(y)}{g(x)} \lesssim \left(\frac{y}{x}\right)^2$,

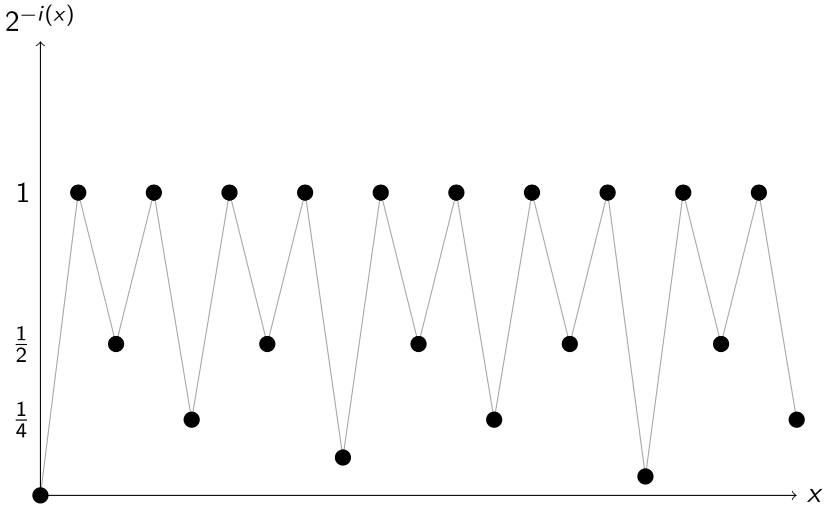
**slow-dropping** $g(y) \gtrsim g(x)$, and

**predictable** whenever $0 < y - x \ll x$

$$g(y) = (1 \pm \epsilon)g(x) \text{ or } g(y-x) \gtrsim g(x).$$

| **g(x)** | **lower bound** | **fails** |
|---|---|---|
| $x^3$ | $\Omega(n^{1/3})$ | slow-jumping |
| $1/x$ | $\Omega(n)$ | slow-dropping |
| $g(x) = (2 + \sin x)x^2$ | $\Omega(n)$ | predictability |

## **Almost necessary**?




$$i(x) = \max\{j \in \mathbb{N} : 2^j \text{ divides } x\}$$