Heavy Hitters (continued)
Tail Bounds

Anna Karlin

Problem

- Input: sequence of n elements $x_1, x_2, \ldots, x_n$ from a known universe U (e.g., 8-byte integers).

- Goal: perform a computation on the input, in a single left to right pass where
  - Elements processed in real time
  - Can't store the full data. => minimal storage requirement to maintain working "summary"

Heavy Hitters: Keys that occur many times

$\overset{x_1}{32}$, $\overset{x_2}{12}$, $\overset{x_3}{14}$, 32, 7, 12, 32, 7, 32, 12, $\overset{x_{11}}{4}$,

Applications:

- Determining popular products
- Computing frequent search queries
- Identifying heavy TCP

$\forall \quad t \leq n \qquad f_x^t$: # times element x has appeared in $x_1, x_2, \ldots, x_t$

Goal: Find/output all elements with $f_x^n > \frac{n}{k}$

(These elts are "heavy hitters")

Output has size $O(k)$

Provably impossible to solve this problem exactly w/ sublinear space

Modified goal: (solve ε-HH problem)

1. If $f_x^n \geq \frac{n}{k}$, x added to HH list
2. If some elt, say y, added to list, then w.p. $\geq 1-\delta$, $f_y^n \geq \frac{n}{k} - \varepsilon n$

Count-min sketch

- Maintain a short summary of the information that still enables answering queries.
- Cousin of the Bloom filter
    - Bloom Filter solves the “membership problem”.
    - We want to extend it to solve a counting problem.

COUNT-MIN SKETCH

Modified goal: (solve ε-HH problem)

① If $f_x^n \geq \frac{n}{k}$, x added to HH list

② If some elt, say y, added to list, then w.p. $\geq 1-\delta$

$$f_y^n \geq \frac{n}{k} - \varepsilon n$$

designer specifies $k, \varepsilon, \delta \Rightarrow b, \ell$

keep 2D array
ℓ hash tables, each of size b.

initialize tables with all 0s

when elt x shows up.

Update (x): $\forall\ 1 \le j \le \ell$ increment $t_j[h_j(x)]$

Count (x): return $\min\limits_{1 \le j \le \ell} t_j[h_j(x)]$

if Count$(x) \ge \frac{n}{k}$, add x to HH list

		1	2	3	4	5				b
h_1	1									
h_2	2									
h_ℓ	ℓ									

Example $\ell = 2$

x_1	x_2	x_3	x_4
x	y	x	z

	h_1	h_2
x	3	5
y	3	2
z	1	4

Assumptions

① hash functions behave like random maps

$$h_1, \dots, h_\ell : U \longrightarrow \{0, 1, \dots, b-1\}$$

$$\forall x \neq y \quad \Pr(h_j(x) = h_j(y)) = \frac{1}{b}$$

② hash fns $h_1, \dots, h_\ell$ are indep of each other.

$\boxed{\begin{array}{l} \text{initialize tables with all 0s} \\ \text{when elt } x \text{ shows up.} \\ \quad \text{Update}(x)\text{: } \forall\ 1 \le j \le \ell \text{ increment } t_j[h_j(x)] \\ \quad \text{Count}(x)\text{: return } \min_{1 \le j \le \ell} t_j[h_j(x)] \\ \quad \text{if Count}(x) \ge \frac{n}{k}\text{, add } x \text{ to HH list} \end{array}}$

Fix time t. ($\overbrace{x_1, \ldots, x_t}$ have just arrived)

$$Z_j^t \triangleq t_j \left[h_j(x) \right]$$

Count-Min Sketch

- Elegant small space data structure.
- Space used is independent of n.
- Is implemented in several real systems.
  - AT&T used in network switches to analyze network traffic.
  - Google uses a version on top of Map Reduce parallel processing infrastructure and in log analysis.
- Huge literature on sketching and streaming algorithms (algorithms like Distinct Elements, Heavy Hitters and many many other very cool algorithms).

Hash functions

6.1 Tail bounds

Most Slides by Joshua Fan and Alex Tsun

Agenda

- Markov's Inequality
- Chebyshev's Inequality
- The law of large numbers

Markov's Inequality (intuition)

The score distribution of an exam is modelled by a rv X with range $\Omega_X \subseteq [0,110]$ (for extra credit).

If the average was $E[X] = 50$, at most what percentage of the class could have gotten 100 (or higher)?

$\leq$

a) 100%

b) 50%

c) 25%

d) no bound

Markov's Inequality (intuition)

The score distribution of an exam is modelled by a rv X with range $\Omega_X \subseteq [0,110]$ (for extra credit).

If the average was $E[X] = 50$, at most what percentage of the class could have gotten 100 (or higher)?

$$\frac{1}{2}$$

Markov's Inequality (intuition)

The score distribution of an exam is modelled by a rv X with range $\Omega_X \subseteq [0,110]$ (for extra credit).

If the average was $E[X] = 50$, at most what percentage of the class could have gotten 100 (or higher)?

$$\frac{1}{2}$$

If the average was $E[X] = 25$, at most what percentage of the class could have gotten 100 (or higher)?

$\leq$

a) 100%

b) 50%

c) 25%

d) no bound

Markov's Inequality (intuition)

The score distribution of an exam is modelled by a rv X with range $\Omega_X \subseteq [0,110]$ (for extra credit).

If the average was $E[X] = 50$, at most what percentage of the class could have gotten 100 (or higher)?

$$\frac{1}{2}$$

If the average was $E[X] = 25$, at most what percentage of the class could have gotten 100 (or higher)?

$$\frac{1}{4}$$

What if you could get a negative score?

$\leq$

a) 100%
b) 50%
c) 25%
d) no bound

Markov's Inequality

Markov's Inequality: Let $X \geq 0$ be a **nonnegative** random variable (discrete or continuous), and let $k > 0$. Then,

$$P(X \geq k) \leq \frac{E[X]}{k}$$

Markov's Inequality

Markov's Inequality: Let $X \geq 0$ be a **nonnegative** random variable (discrete or continuous), and let $k > 0$. Then,

$$P(X \geq k) \leq \frac{E[X]}{k}$$

Alternatively,

$$P(X \geq kE[X]) \leq \frac{1}{k}$$

Markov's Inequality (Proof)

Markov's Inequality: Let $X \geq 0$ be a **nonnegative** rv and let $k > 0$. Then,

$$P(X \geq k) \leq \frac{E[X]}{k}$$

Proof (Markov):

$X \geq 0$

$$E[X] = \int_0^\infty x f_X(x)dx = \int_0^k x f_X(x)dx + \int_k^\infty x f_X(x)dx$$

$$\geq \int_k^\infty x f_X(x)dx \geq \int_k^\infty k f_X(x)dx = k\int_k^\infty f_X(x)dx = kP(X \geq k)$$

Rearranging gives

$$P(X \geq k) \leq \frac{E[X]}{k}$$

Chebyshev's Inequality

Chebyshev's Inequality: Let X be any random variable, with mean $\mu = E[X]$ and (finite) variance. Let $\alpha > 0$.

$$P(|X - \mu| \geq \alpha) \leq \frac{Var(X)}{\alpha^2}$$

Chebyshev's Inequality

Chebyshev's Inequality: Let X be any random variable, with mean $\mu = E[X]$ and (finite) variance. Let $\alpha > 0$.

$$P(|X - \mu| \geq \alpha) \leq \frac{Var(X)}{\alpha^2}$$

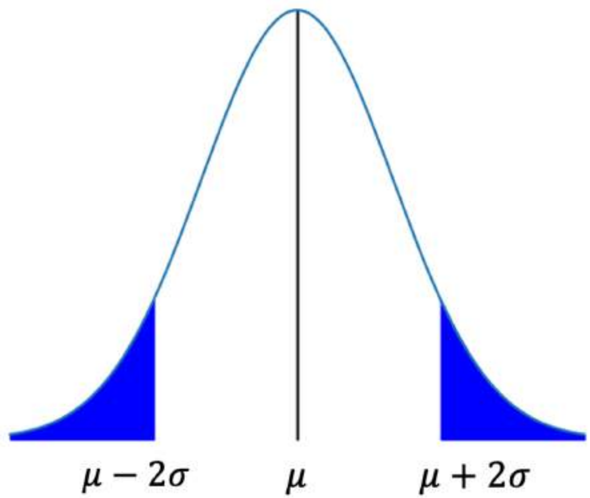
Alternatively, if $k > 0$,

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Chebyshev's Inequality (picture for Gaussian)

Chebyshev's Inequality: Let X be any random variable, with mean $\mu = E[X]$ and (finite) variance. Let $k > 0$.

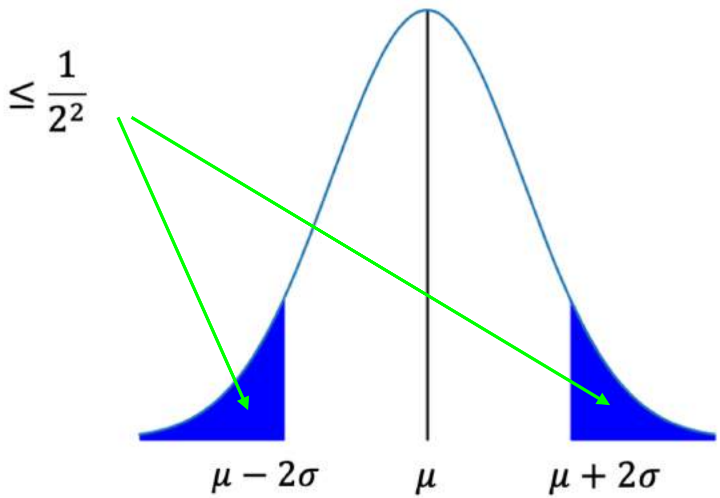
$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$



Chebyshev's Inequality (picture for Gaussian)

Chebyshev's Inequality: Let X be any random variable, with mean $\mu = E[X]$ and (finite) variance. Let $k > 0$.

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$



Chebyshev's Inequality (Proof)

Markov's Inequality: Let $X \geq 0$ be a **nonnegative** rv and let $k > 0$. Then,

$$P(X \geq k) \leq \frac{E[X]}{k}$$

Chebyshev's Inequality: Let X be any random variable, with mean $\mu = E[X]$ and (finite) variance. Let $\alpha > 0$.

$$P(|X - \mu| \geq \alpha) \leq \frac{Var(X)}{\alpha^2}$$

Chebyshev's Inequality (Proof)

Markov's Inequality: Let $X \geq 0$ be a **nonnegative** rv and let $k > 0$. Then,

$$P(X \geq k) \leq \frac{E[X]}{k}$$

Chebyshev's Inequality: Let X be any random variable, with mean $\mu = E[X]$ and (finite) variance. Let $\alpha > 0$.

$$P(|X - \mu| \geq \alpha) \leq \frac{Var(X)}{\alpha^2}$$

Proof (Chebyshev): $(X - \mu)^2$ is a nonnegative random variable.

Chebyshev's Inequality (Proof)

Markov's Inequality: Let $X \geq 0$ be a **nonnegative** rv and let $k > 0$. Then,

$$P(X \geq k) \leq \frac{E[X]}{k}$$

Chebyshev's Inequality: Let X be any random variable, with mean $\mu = E[X]$ and (finite) variance. Let $\alpha > 0$.

$$P(|X - \mu| \geq \alpha) \leq \frac{Var(X)}{\alpha^2}$$

Proof (Chebyshev): $(X - \mu)^2$ is a nonnegative random variable.

$$\begin{aligned} P(|X - \mu| \geq \alpha) &= P((X - \mu)^2 \geq \alpha^2) \\ &\leq \frac{E[(X - \mu)^2]}{\alpha^2} \text{ [Markov]} \\ &= \frac{Var(X)}{\alpha^2} \end{aligned}$$

The Law of Large Numbers

Weak Law of Large Numbers (WLLN): Let $X_1, X_2, \ldots X_n$ be a sequence of iid random variables with mean μ. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean. Then, $\bar{X}_n$ **converges in probability** to μ. That is, for any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0$$

Proof of the WLLN

<u>**Weak Law of Large Numbers (WLLN):**</u> Let $X_1, X_2, \ldots X_n$ be a sequence of iid random variables with mean μ. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean. Then, $\bar{X}_n$ **converges in probability** to μ. That is, for any $\epsilon > 0$,

$$\lim_{n \to \infty} P(|\bar{X}_n - \mu| > \epsilon) = 0$$

<u>**Proof**</u>: Recall $E[\bar{X}_n] = \mu$ and $Var(\bar{X}_n) = \sigma^2/n$.

Proof of the WLLN

<u>Weak Law of Large Numbers (WLLN):</u> Let $X_1, X_2, \ldots X_n$ be a sequence of iid random variables with mean μ. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean. Then, $\bar{X}_n$ **converges in probability** to μ. That is, for any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0$$

<u>Proof</u>: Recall $E[\bar{X}_n] = \mu$ and $Var(\bar{X}_n) = \sigma^2/n$. By Chebyshev's inequality,

$$P(|\bar{X}_n - \mu| > \epsilon) \leq \frac{Var(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0 \text{ (as } n \to \infty)$$