Heavy Hitters
Tail Bounds

Anna Karlin

Stream model

- Input elements (e.g. Google queries) enter/arrive one at a time.
- We cannot possibly store the stream.

Question: How do we make critical calculations about the data stream using a limited amount of memory?

Sources of this kind of data

- Sensor data
    - E.g. millions of temperature sensors deployed in the ocean
- Image data from satellites or surveillance camers
    - E.g. London
- Internet and web traffic
    - E.g. millions of streams of IP packets
- Web data
    - E.g. Search queries on Google, clicks on Bing, etc.

Example Applications

- Mining query streams
  - Google wants to know which queries are more frequent today than yesterday.
- Mining click streams
  - Facebook wants to know which of its ads are getting an unusual number of hits in the last hour.
- Mining social network news feeds
  - E.g., looking for trending topics on Twitter and Facebook, trending videos on TikTok

More Applications

- Sensor networks
    - Many sensors feeding into a central controller.
- IP packets
    - Gather congestion information for optimal routing
    - Detect denial-of-service attacks

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

32, 12, 14, 32, 7, 12, 32, 7, 32, 12, 4,

Applications:

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

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

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).

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)?

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)?

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?

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

<u>**Markov's Inequality:**</u> 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 $\alpha > 0$.

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

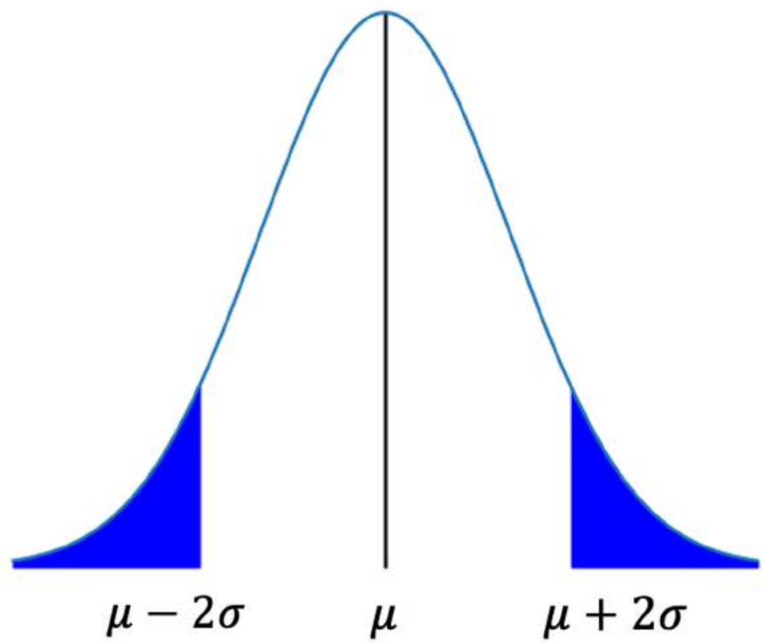
Alternatively, if $k > 0$,

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

Chebyshev's Inequality (picture for Gaussian)

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

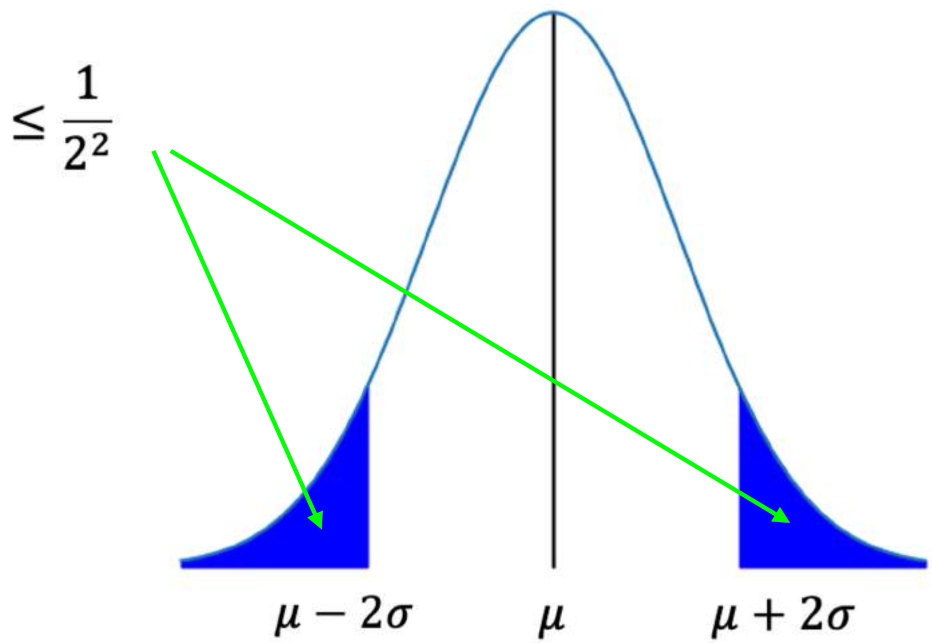
$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$



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

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

Proof of the WLLN

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

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