Preliminary: Probability and tail bound for sketches algorithms

Lecturer: Jiaheng Lu

Fall 2016

Basic Probability Theory

A **probability space** is a triple (Ω, E, P) with

- a set Ω of elementary events (sample space),
- a family E of subsets of Ω with $\Omega \in E$ which is closed under $\cap$, $\cup$, and $-$ with a countable number of operands (with finite Ω usually $E=2^{\Omega}$), and
- a **probability measure $P: E \rightarrow [0,1]$** with $P[\Omega]=1$ and $P[\cup_i A_i] = \sum_i P[A_i]$ for countably many, pairwise disjoint A_i

Probabilities of events

Properties of P:

$P[A] + P[\neg A] = 1$

$P[A \cup B] = P[A] + P[B] - P[A \cap B]$

$P[\varnothing] = 0$ (null/impossible event)

$P[\Omega] = 1$ (true/certain event)

Independence of events

- Two events A and B are independent if

$$p(A \cap B) = p(A)p(B)$$

- Conditional probability: For two events A and B with $p(B) > 0$, the *probability of A conditioned on B* is $p(A|B) = \frac{p(A \cap B)}{p(B)}$.

Random variable

- A random variable X is a function $X: \Omega \to R$.
- $\Pr[X = a] = \sum_{s \in \Omega : X(s) = a} p(s)$.
- Two random variables X and Y are independent if

$$\Pr[(X = a) \wedge (Y = b)] = \Pr[X = a] \Pr[Y = b] .$$

Expectation

- Expectation:

$$\mathbf{E}[X] = \sum_{s\in\Omega} p(s)X(s)$$
$$= \sum_{i\in\mathrm{Range}(X)} i \cdot \Pr[X = i]$$

- Linearity of expectation:

$$\mathbf{E}[\sum_i X_i] = \sum_i \mathbf{E}[X_i]$$

no matter whether X_i's are independent or not.

variance

- The variance of X is

$$\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$$

- The standard deviation of X is

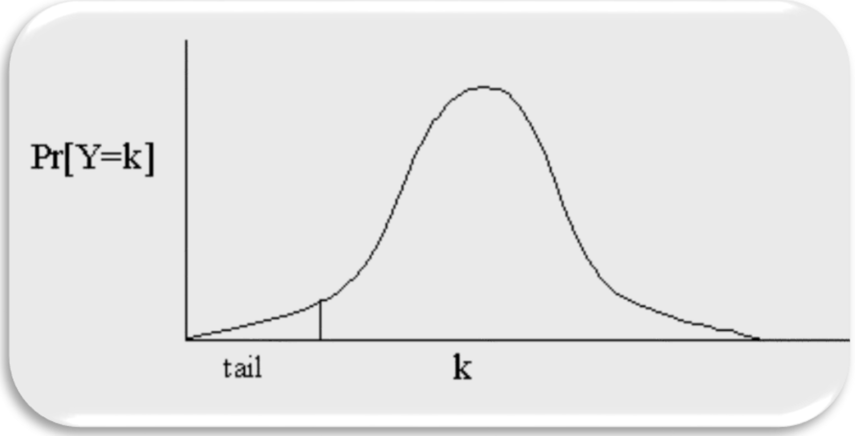
$$\sigma = \sqrt{\mathbf{Var}[X]}$$

Concentration and tail bounds

- In many analysis of randomized algorithms, we need to study how concentrated a random variable X is close to its mean $E[X]$.
  - Many times $X = X_1 + \cdots + X_n$.
- Upper bounds of

$$\Pr[X \text{ deviates from } E[X] \text{ a lot}]$$

is called *tail bounds*.



Markov's Inequality: when you only know expectation

- [Thm] If $X \geq 0$, then

$$\mathbf{Pr}[X \geq a] \leq \frac{\mathbf{E}[X]}{a}.$$

In other words, if $E[X] = \mu$, then

$$\mathbf{Pr}[X \geq k\mu] \leq \frac{1}{k}.$$

Chebyshev's Inequality: when you also know variance

- [Thm] $\mathbf{Pr}[|X - \mathbf{E}[X]| \geq a] \leq \frac{\mathbf{Var}[X]}{a^2}$.

  In other words,

  $$\mathbf{Pr}\left[|X - \mathbf{E}[X]| \geq k \cdot \sqrt{\mathbf{Var}[X]}\right] \leq \frac{1}{k^2}.$$

Chernoff's Bound

- [Thm] Suppose $X_i = \begin{cases} 1 & \text{with prob. } p \\ 0 & \text{with prob. } 1-p \end{cases}$

  and let

  $$X = X_1 + \cdots + X_n.$$

  Then

  $$\mathbf{Pr}[|X - \mu| \geq \delta\mu] \leq 2\, e^{-\delta^2 \mu/3},$$

  where $\mu = np = \mathbf{E}[X]$.

Some basic applications: coin tossing

- A fair coin
  - $f(x) = 0$: with probability 1/2
  - $f(x) = 1$: with probability 1/2

- Repeatedly toss the coin, Let S_n be the number of heads from the first n tosses.

- $E(S_n) = n/2$, $Var(S_n) = n/4$
- $E(S_n/n) = 1/2$, $Var(S_n/n) = 1/(4n)$

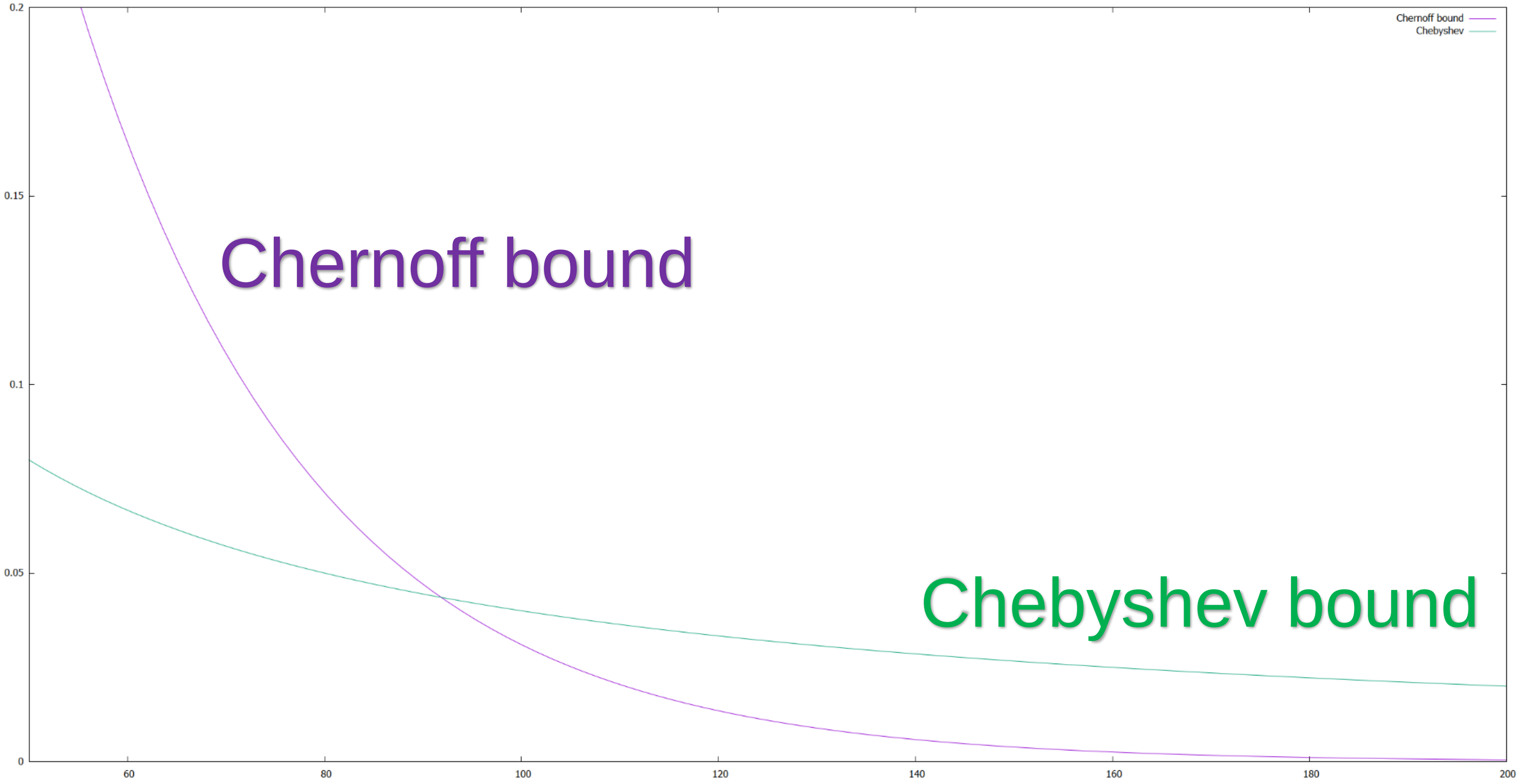
Some basic applications: coin tossing

- In terms of Chebyshev's inequality
- $P(|\frac{S_n}{\mathbf{n}} - 1/2| \geq \varepsilon) \leq 1 / (4n\, \varepsilon^2)$

- For example $P(|\frac{S_n}{\mathbf{n}} - 1/2| \geq 1/4) \leq 4/n$

- But if we use Chernoff bound, $E(S_n) = n/2$
- $\mathbf{Pr}[|S_n - n/2| \geq \delta n/2] \leq 2\, e^{-\delta^2 n/6}$,

- Taking $\delta = ½$, $\mathbf{Pr}[|S_n/n - 1/2| \geq 1/4] \leq 2\, e^{-\delta^2 n/6} = 2e^{-n/24}$.
-


Chernoff bound
Chebyshev
0.2
0.15
0.1
0.05
0
60
80
100
120
140
160
180
200
Chernoff bound
Chebyshev bound