CSE 312

Foundations of Computing II

Lecture 20: Tail Bounds Part II
Chebyshev and Chernoff Bounds

Agenda

- Covariance
- Markov's Inequality
- Chebyshev's Inequality

Review Tail Bounds (Idea)

Bounding the probability that a random variable is far from its mean. Usually statements of the form:

$$P(X \geq a) \leq b$$
$$P(|X - \mathbb{E}[X]| \geq a) \leq b$$

Useful tool when

- An approximation that is easy to compute is sufficient
- The process is too complex to analyze exactly

Review Markov's Inequality

Theorem. Let X be a random variable taking only *non-negative* values. Then, for any $t > 0$,

$$P(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

(Alternative form) For any $k \geq 1$,

$$P(X \geq k \cdot \mathbb{E}[X]) \leq \frac{1}{k}$$

Incredibly simplistic – only requires that the random variable is non-negative and only needs you to know expectation. You don't need to know **anything else** about the distribution of X.

Review Example – Geometric Random Variable

Let X be geometric RV with parameter p

$$P(X = i) = (1-p)^{i-1}p \qquad \mathbb{E}[X] = \frac{1}{p}$$

"X is the number of times Alice needs to flip a biased coin until she sees heads, if heads occurs with probability p?

What is the probability that $X \geq 2\mathbb{E}[X] = 2/p$?

Markov's inequality: $P(X \geq 2\mathbb{E}[X]) \leq \frac{1}{2}$

Can we do better?

Review Example

$$P(X \geq k \cdot \mathbb{E}[X]) \leq \frac{1}{k}$$

Suppose that the average number of ads you will see on a website is 25. Compute an upper bound on the probability of seeing a website with 75 or more ads.

Where does that upper bound p lie?

a. $0 \leq p < 0.25$
b. $0.25 \leq p < 0.5$
c. $0.5 \leq p < 0.75$
d. $0.75 \leq p$
e. Unable to compute

X = *RV for number of ads on a website visit*

$\mathbb{E}[X] = 25$

$P(X \geq 75) = P(X \geq 3 \cdot \mathbb{E}[X]) \leq \frac{1}{3} = p$

Note: If this is all you know about X then you can't get a better bound:

Example RV X with $\mathbb{E}[X] = 25$:

$$P(X = 0) = \frac{2}{3}$$
$$P(X = 75) = \frac{1}{3}$$

Example

$$P(X \geq k \cdot \mathbb{E}[X]) \leq \frac{1}{k}$$

Suppose that the average number of ads you will see on a website is 25. Compute an upper bound on the probability of seeing a website with 20 or more ads.

Poll: Where does that upper bound p lie?
pollev.com/paulbeame028

a. $0 \leq p < 0.25$
b. $0.25 \leq p < 0.5$
c. $0.5 \leq p < 0.75$
d. $0.75 \leq p$
e. Unable to compute

Agenda

- Markov's Inequality
- Chebyshev's Inequality
- Chernoff-Hoeffding Bound

Chebyshev's Inequality

Theorem. Let X be a random variable. Then, for any $t > 0$,

$$P(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}.$$

Proof: Define $Z = X - \mathbb{E}[X]$. Then $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[Z^2]$.

$$P(|Z| \geq t) = P(Z^2 \geq t^2) \leq \frac{\mathbb{E}[Z^2]}{t^2} = \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{t^2} = \frac{\mathrm{Var}(X)}{t^2}$$

$|Z| \geq t$ iff $Z^2 \geq t^2$ Markov's inequality ($Z^2 \geq 0$)

Example – Geometric Random Variable

Let X be geometric RV *with* parameter p

$$P(X = i) = (1-p)^{i-1}p \qquad \mathbb{E}[X] = \frac{1}{p} \qquad \mathrm{Var}(X) = \frac{1-p}{p^2}$$

What is the probability that $X \geq 2\mathbb{E}(X) = 2/p$*?*

Markov: $P(X \geq 2\mathbb{E}[X]) \leq \frac{1}{2}$

Chebyshev: $P(X \geq 2\mathbb{E}[X]) \leq P(|X - \mathbb{E}[X]| \geq \mathbb{E}[X]) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} = 1 - p$

Better if $p > 1/2$ ☺

Example

$P(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}.$

Suppose that the average number of ads you will see on a website is 25 and the standard deviation of the number of ads is 4. Give an upper bound on the probability of seeing a website with 30 or more ads.

Poll: Where does that upper bound p lie?
pollev.com/paulbeame028

a. $0 \leq p < 0.25$
b. $0.25 \leq p < 0.5$
c. $0.5 \leq p < 0.75$
d. $0.75 \leq p$
e. Unable to compute

Chebyshev's Inequality – Repeated Experiments

"How many times does Alice need to flip a biased coin <u>until she sees heads</u> n times, if heads occurs with probability p?

X = # of flips until n times "heads"

X_i = # of flips between $(i-1)$-st and i-th "heads"

$$X = \sum_{i=1}^{n} X_i$$

Note: $X_1, \ldots, X_n$ are independent and geometric with parameter p

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{p}$$

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{n(1-p)}{p^2}$$

Chebyshev's Inequality – Coin Flips

"How many times does Alice need to flip a biased coin until she sees heads n times, if heads occurs with probability p?

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{p} \qquad \mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{n(1-p)}{p^2}$$

What is the probability that $X \geq 2\mathbb{E}[X] = 2n/p$?

Markov: $P(X \geq 2\mathbb{E}[X]) \leq \frac{1}{2}$

Chebyshev: $P(X \geq 2\mathbb{E}[X]) \leq P(|X - \mathbb{E}[X]| \geq \mathbb{E}[X]) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} = \frac{1-p}{n}$

Goes to zero as $n \to \infty$ ☺

Tail Bounds

Useful for approximations of complex systems. How good the approximation is depends on the actual distribution and the context you are using it in.

– Very often loose upper-bounds are okay when designing for the worst case

Generally (but not always) making more assumptions about your random variable leads to a more accurate upper-bound.

Brain Break

Agenda

- Markov's Inequality
- Chebyshev's Inequality
- Chernoff-Hoeffding Bound

Chebyshev & Binomial Distribution

$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$.

Reformulated: $P(|X-\mu| \geq \delta\mu) \leq \frac{\sigma^2}{\delta^2\mu^2}$ where $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}(X)$

If $X \sim \mathrm{Bin}(n, p)$, then $\mu = np$ and $\sigma^2 = np(1-p)$

$$P(|X-\mu| \geq \delta\mu) \leq \frac{np(1-p)}{\delta^2 n^2 p^2} = \frac{1}{\delta^2} \cdot \frac{1}{n} \cdot \frac{1-p}{p}$$

E.g., $\delta = 0.1, p = 0.5$: $n = 200$: $P(|X-\mu| \geq \delta\mu) \leq 0.5$

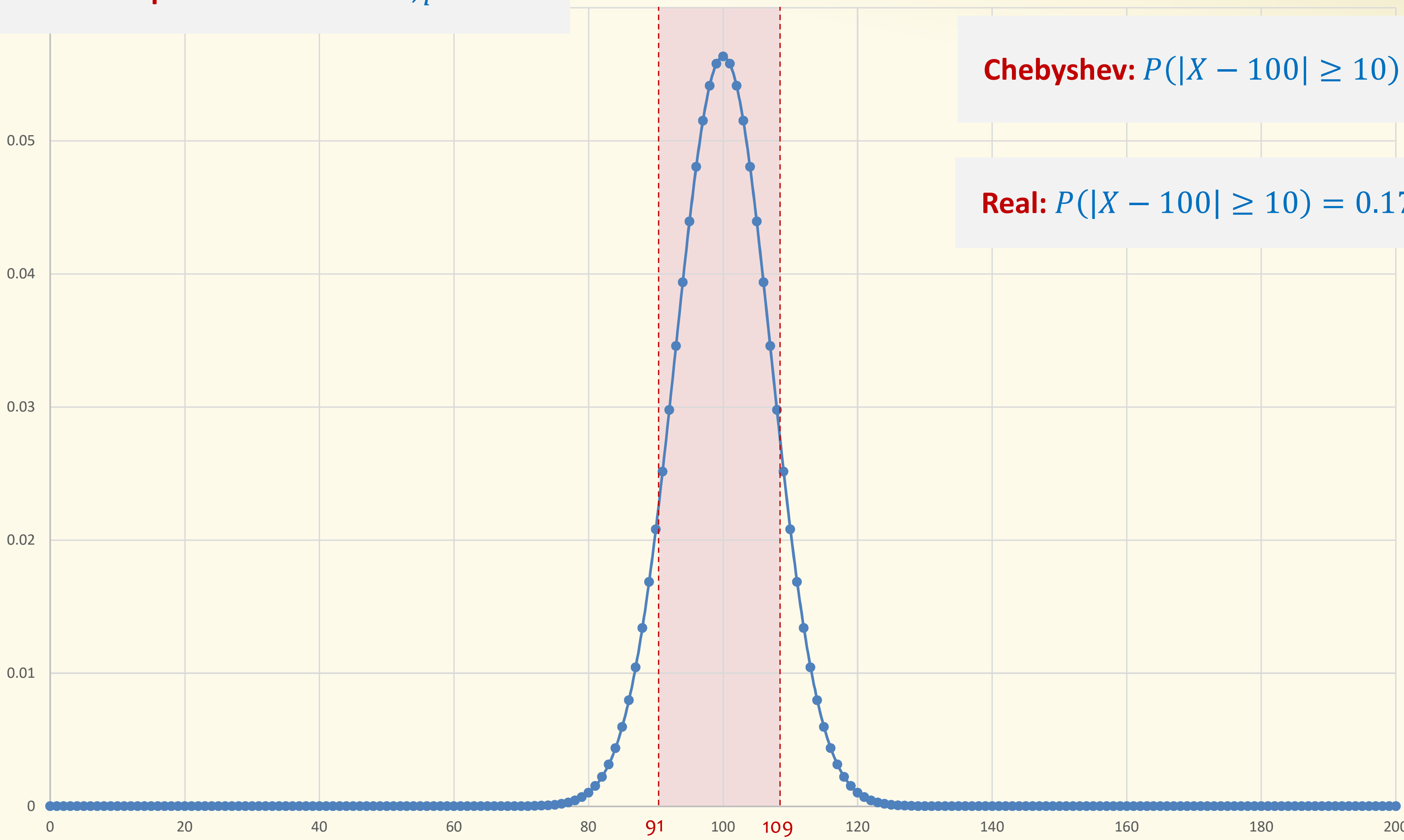
$n = 800$: $P(|X-\mu| \geq \delta\mu) \leq 0.125$

How good is it?

Binomial with parameter $n = 200, p = 0.5$

Chebyshev: $P(|X - 100| \geq 10) \leq \frac{1}{2}$

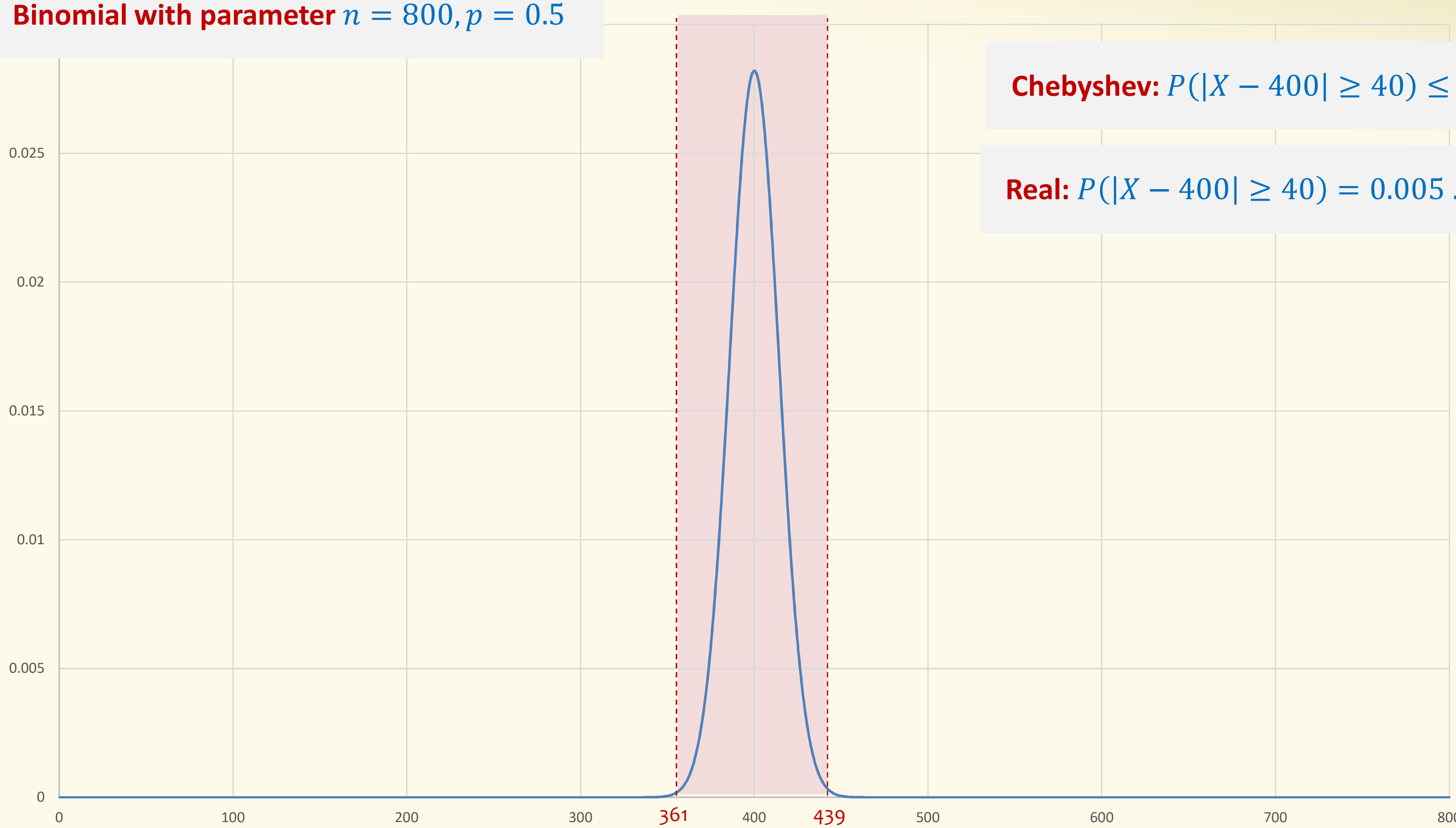
Real: $P(|X - 100| \geq 10) = 0.179 \ldots$



Binomial with parameter $n = 800, p = 0.5$

Chebyshev: $P(|X - 400| \geq 40) \leq 0.125$

Real: $P(|X - 400| \geq 40) = 0.005 \ldots$



Chernoff-Hoeffding Bound

Theorem. Let $X = X_1 + \cdots + X_n$ be a sum of independent RVs, each taking values in $[0,1]$, such that $\mathbb{E}[X] = \mu$. Then, for every $\delta \in [0,1]$,

$$P(|X - \mu| \geq \delta \cdot \mu) \leq e^{-\frac{\delta^2 \mu}{4}}.$$

Herman Chernoff, Herman Rubin, Wassily Hoeffding

Example: If $X \sim \mathrm{Bin}(n, p)$, then $X = X_1 + \cdots + X_n$ is a sum of independent $\{0,1\}$-Bernoulli variables, and $\mu = np$

Note: More accurate versions are possible, but with more cumbersome right-hand side (e.g., see textbook)

Chernoff-Hoeffding Bound – Binomial Distribution

Theorem. (CH bound, binomial case) Let $X \sim \text{Bin}(n, p)$. Let $\mu = np = \mathbb{E}[X]$. Then, for any $\delta \in [0,1]$,

$$P(|X - \mu| \geq \delta \cdot \mu) \leq e^{-\frac{\delta^2 np}{4}}.$$

Example:

$p = 0.5$
$\delta = 0.1$

n	Chebyshev $\frac{1}{\delta^2} \cdot \frac{1}{n} \cdot \frac{1-p}{p}$	Chernoff $e^{-\frac{\delta^2 np}{4}}$
800	0.125	0.3679
2600	0.03846	0.03877
8000	0.0125	0.00005
80000	0.00125	3.72×10^{-44}

Chernoff Bound – Example

$$\mathbb{P}(|X-\mu| \geq \delta \cdot \mu) \leq e^{-\frac{\delta^2 np}{4}}.$$

Alice tosses a fair coin n times, what is an upper bound for the probability that she sees heads at least $0.75 \times n$ times?

Poll: pollev.com/paulbeame028

a. $e^{-n/64}$
b. $e^{-n/32}$
c. $e^{-n/16}$
d. $e^{-n/8}$

Chernoff vs Chebyshev – Summary

$$\frac{1}{\delta^2} \cdot \frac{1}{n} \cdot \frac{1-p}{p}$$

Chebyshev, <u>linear</u> decrease in n

vs

Chernoff, <u>exponential</u> decrease in n

$$e^{-\frac{\delta^2 np}{4}}$$

Why is the Chernoff Bound True?

Proof strategy (upper tail): For any $t > 0$:

- $P(X \geq (1+\delta) \cdot \mu) = P\left(e^{tX} \geq e^{t(1+\delta)\cdot\mu}\right)$
- Then, apply Markov + independence:

$$P\left(e^{tX} \geq e^{t(1+\delta)\cdot\mu}\right) \leq \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\delta)\mu}} = \frac{\mathbb{E}[e^{tX_1}] \cdots \mathbb{E}[e^{tX_n}]}{e^{t(1+\delta)\mu}}$$

- Find t minimizing the right-hand-side.

Next time

Examples using the Chernoff bound together with the simple "Union Bound"

Beginning "estimation"