CSE 312

Foundations of Computing II

Lecture 23: Chernoff Bound & Union Bound

PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING

Rachel Lin, Hunter Schafer

Slide Credit: Based on Stefano Tessaro's slides for 312 19au
incorporating ideas from Alex Tsun's and Anna Karlin's slides for 312 20su and 20au

Review Tail Bounds

Putting a limit on the probability that a random variable is in the "tails" of the distribution (e.g., not near the middle).

Usually statements in the form of

$$\Pr(X \geq a) \leq b$$

or

$$\Pr(|X - E[X]| \geq a) \leq b$$

Review Markov's and Chebyshev's Inequalities

Theorem (Markov's Inequality). Let X be a random variable taking only non-negative values. Then, for any $t > 0$,

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

Theorem (Chebyshev's Inequality). Let X be a random variable. Then, for any $t > 0$,

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

Agenda

- Union Bound
- Chernoff Bound
- Application: Polling (again)
- Extra Example: Server Load

Union Bound

Not a tail bound, but a useful formula

Theorem (Union Bound). Let $A_1, \ldots, A_n$ be arbitrary events. Then,

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)$$

Intuition (2 evts.): $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$

Union Bound - Example

Suppose we have $N = 200$ computers, where each one fails with probability 0.001. What is the probability that at least one server fails?

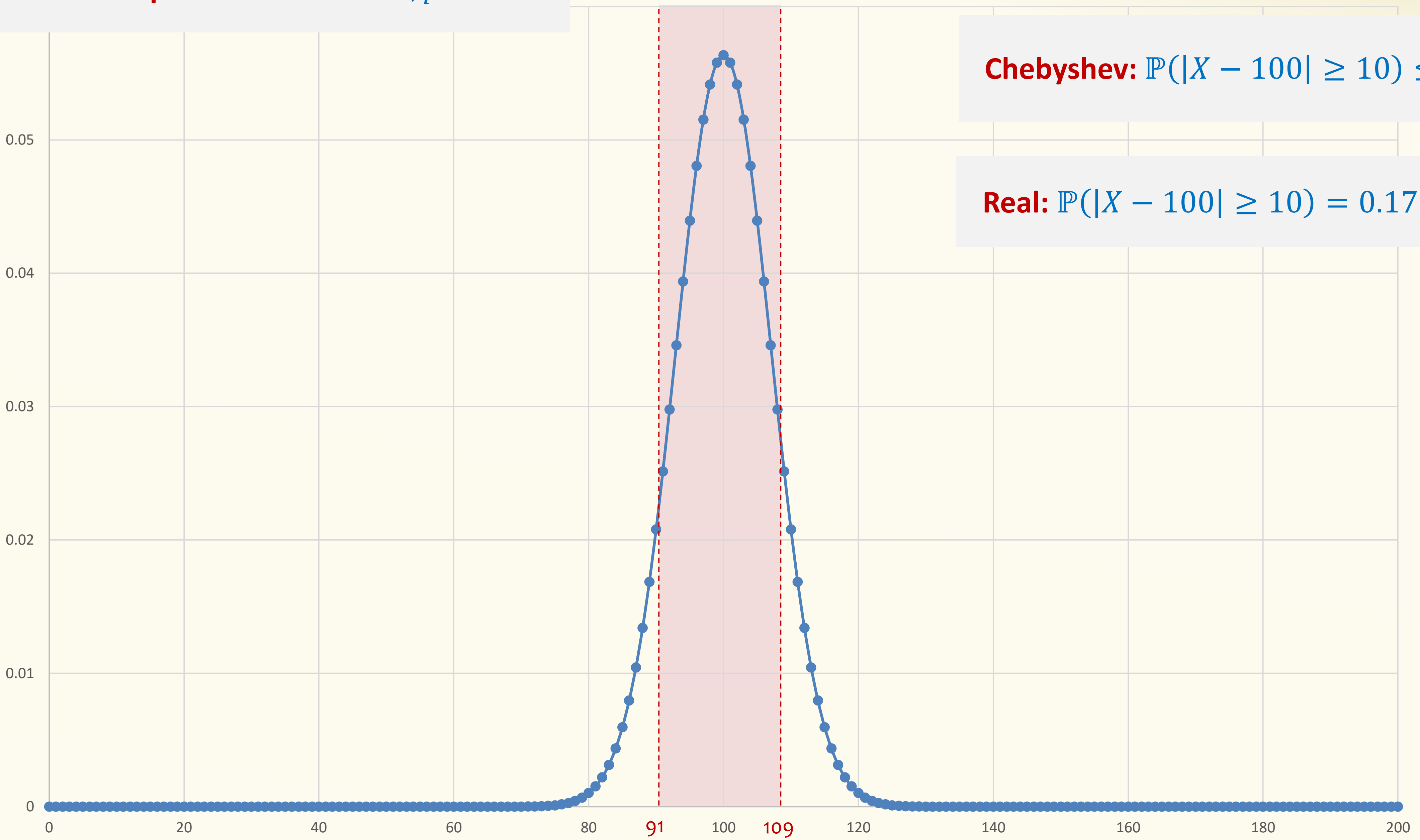
Let A_i be the event that server i fails. Then at least one server fails in the event $\cup_{i=1}^{N} A_i$

$$\Pr\left(\bigcup_{i=1}^{N} A_i\right) \le \sum_{i=1}^{N} \Pr(A_i) = 0.001N = 0.2$$

Agenda

- Union Bound
- Chernoff Bound
- Application: Polling (again)
- Extra Example: Server Load

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



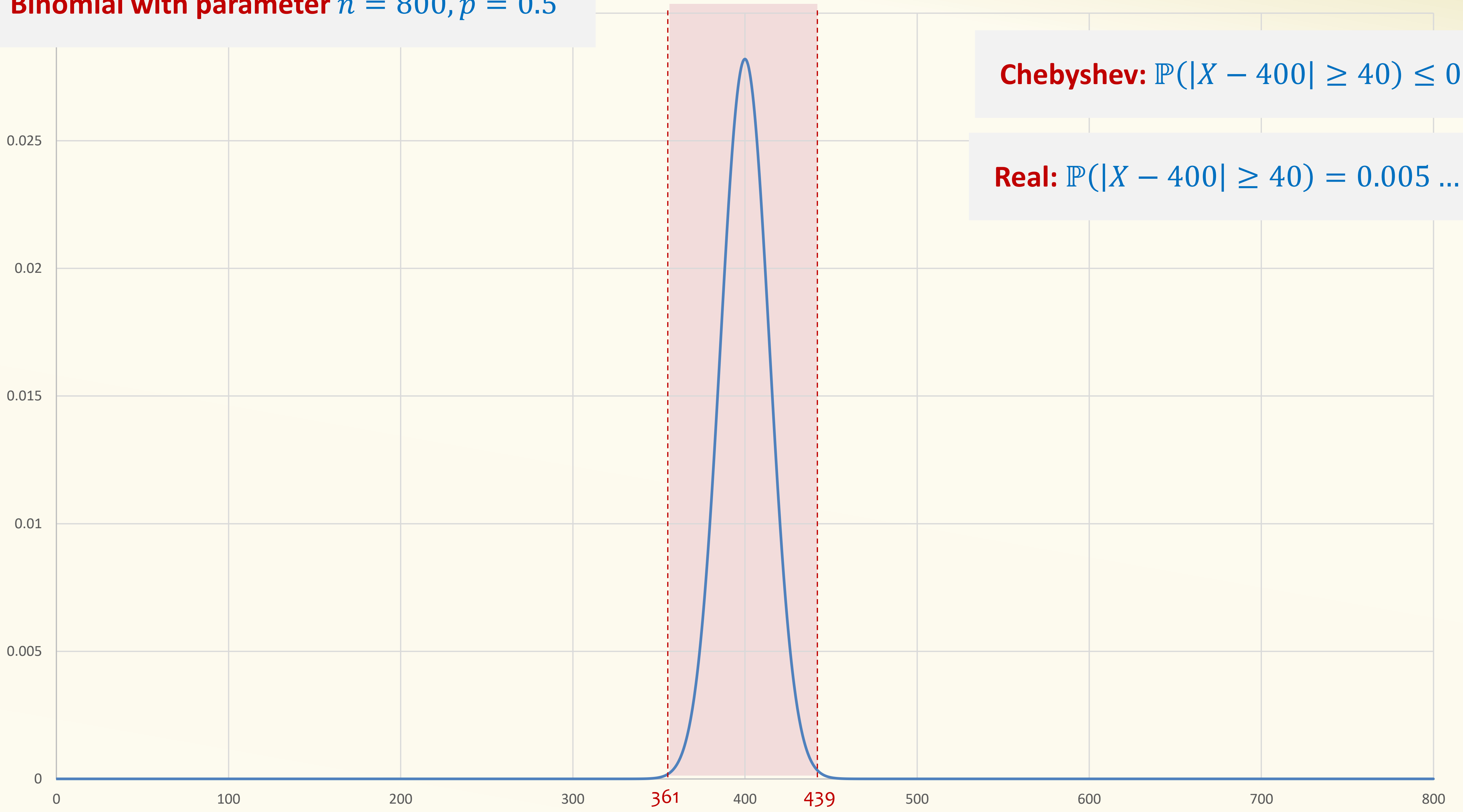
Chebyshev: $\mathbb{P}(|X - 100| \geq 10) \leq \frac{1}{2}$

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

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

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

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



Chernoff-Hoeffding Bound – Binomial Distribution

Theorem. (CH bound, binomial case) Let X be a binomial RV with parameters p and n. Let $\mu = np = \mathbb{E}(X)$. Then, for any $\epsilon > 0$,

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

Binomial: $n = 800, p = 0.5 \rightarrow \mu = np = 400$

Chebyshev: $\mathbb{P}(|X - \mu| \geq 0.1\mu) \leq 0.125$

CH: $\mathbb{P}(|X - \mu| \geq 0.1\mu) \leq 2e^{-\frac{4}{2.1}} = 0.296 \ldots$

Chernoff-Hoeffding Bound – Binomial Distribution

Theorem. (CH bound, binomial case) Let X be a binomial RV with parameters p and n. Let $\mu = np = \mathbb{E}(X)$. Then, for any $\epsilon > 0$,

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

Binomial: $n = 8000, p = 0.5 \rightarrow \mu = np = 4000$

Chebyshev: $\mathbb{P}(|X-\mu| \geq 0.1\mu) \leq 0.0125$

CH: $\mathbb{P}(|X-\mu| \geq 0.1\mu) \leq 2e^{-\frac{40}{2.1}} \approx 1.7 \times 10^{-8}$

Chernoff-Hoeffding Bound, beyond Binomial RV

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 $\epsilon > 0$,

$$\mathbb{P}(X \geq (1+\epsilon)\cdot\mu) \leq e^{-\frac{\epsilon^2\mu}{2+\epsilon}}, \qquad \mathbb{P}(X \leq (1-\epsilon)\cdot\mu) \leq e^{-\frac{\epsilon^2\mu}{2}}$$

In particular,

$$\mathbb{P}(|X-\mu| \geq \epsilon\cdot\mu) \leq 2e^{-\frac{\epsilon^2\mu}{2+\epsilon}}$$

Herman Chernoff, Herman Rubin, Wassily Hoeffding

Example: If X binomial w/ parameters n, p, then $X = X_1 + \cdots + X_n$ is a sum of independent $\{0,1\}$-Bernoulli variables, and $\mu = np$

Agenda

- Union Bound
- Chernoff Bound
- Application: Polling (again)
- Extra Example: Server Load

Application – Polling

We have a (large) population of M CS students.

- A fraction $p \in [0,1]$ supports the introduction of **CSE 313**
  - a harder, follow-up class to CSE 312, with even more math
  - CSE 313 would be a hard requirement for all NLP/ML classes
- We want to estimate p without asking all M students!

How can we do this with enough accuracy?
[Say, estimate within absolute error ϵ]

Polling (cont'd)

Solution: For $i = 1, \ldots, n$ do:

- Pick a random student P_i (out of the M students) and ask them whether they want **CSE 313**
- Let $X_i = 1$ if student P_i wants **CSE 313**, and $X_i = 0$ else.

Output estimate $\hat{P} = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\mathbb{P}(X_i = 1) = p$ $\mathbb{E}(\hat{P}) = p$

What's the chance $|\hat{P} - p| \geq \epsilon$

Want: $\mathbb{P}(|\hat{P} - p| \geq \epsilon) \leq \delta$

For which n is this true?!

Polling (cont'd) $\mathbb{P}(X_i = 1) = p$

Theorem. Let X be a binomial RV with parameters p and n. Let $\mu = np = \mathbb{E}(X)$. Then, for any $\epsilon > 0$,

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

$$\mathbb{P}\left(\left|\hat{P} - p\right| \geq \epsilon\right) = \mathbb{P}\left(\left|n\hat{P} - np\right| \geq n\epsilon\right)$$

$$= \mathbb{P}\left(\left|\sum_i^n X_i - np\right| \geq n\epsilon\right)$$

$$= \mathbb{P}\left(\left|\sum_i^n X_i - np\right| \geq np\frac{\epsilon}{p}\right)$$

$$\leq 2\exp\left(-\frac{\epsilon^2/p^2}{2 + \epsilon/p}pn\right)$$

$$= 2\exp\left(-\frac{\epsilon^2}{2p + \epsilon}n\right) \leq 2\exp\left(-\frac{\epsilon^2}{2 + \epsilon}n\right)$$

Reminder: $\exp(x) = e^x$

Polling (cont'd) $\mathbb{P}(X_i = 1) = p$

We have proved:

$$\mathbb{P}\left(\left|\hat{P} - p\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{\epsilon^2}{2+\epsilon}n\right)$$

We have $2\exp\left(-\frac{\epsilon^2}{2+\epsilon}n\right) \leq \delta$ if (and only if)

$$n \geq \ln(2/\delta)\frac{2+\epsilon}{\epsilon^2}$$

Polling – Summary

Theorem. (Sampling Theorem) Assume we use independent uniformly random samples to produce an estimate $\hat{P}$ of $p \in [0,1]$. If

$$n \geq \ln(2/\delta)\frac{2+\epsilon}{\epsilon^2},$$

then

$$\mathbb{P}\left(\left|\hat{P} - p\right| \leq \epsilon\right) \geq 1 - \delta.$$

Important: "Sample size" n is <u>independent</u> of the population size, M. <u>Only</u> depends on desired accuracy.

e.g. $\epsilon = 0.02, \delta = 0.05, n \geq 15{,}128$

Central question in CS and statistics – can we do better?!
Central question in polling – how can we sample n iid samples?

Agenda

- Union Bound
- Chernoff Bound
- Application: Polling (again)
- Extra Example: Server Load

Why is the Chernoff Bound True?

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

$$\mathbb{P}(X \geq (1+\epsilon)\cdot\mu) \leq e^{-\frac{\epsilon^2\mu}{2+\epsilon}}, \qquad \mathbb{P}(X \leq (1-\epsilon)\cdot\mu) \leq e^{-\frac{\epsilon^2\mu}{2}}$$

Proof strategy: For any $t > 0$:

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

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

- Find t minimizing the right-hand-side.

Application – Distributed Load Balancing

We have k processors, and $n \gg k$ jobs. We want to distribute jobs evenly across processors.

Strategy: Each job assigned to a randomly chosen processor!

X_i = load of processor i $\qquad X_i \sim \text{Binomial}(n, 1/k) \qquad \mathbb{E}(X_i) = n/k$

$X = \max\{X_1, \dots, X_k\}$ = max load of a processor

Question: How close is X to n/k?

Distributed Load Balancing

Claim. (Load of single server) If $n > 9k \ln k$, then

$$\mathbb{P}\left(X_i > \frac{n}{k} + 3\sqrt{\frac{n \ln k}{k}}\right) = \mathbb{P}\left(X_i > \frac{n}{k}\left(1 + 3\sqrt{\frac{k \ln k}{n}}\right)\right) \leq 1/k^3.$$

Example:

- $n = 10^6 \gg k = 1000$
- $\frac{n}{k} + 3\sqrt{n \ln k / k} \approx 1249$
- *"The probability that server* i *processes more than* 1249 *jobs is at most 1-over-one-billion!"*

Distributed Load Balancing

Claim. (Load of single server) If $n > 9k \ln k$, then

$$\mathbb{P}\left(X_i > \frac{n}{k} + 3\sqrt{\frac{n \ln k}{k}}\right) = \mathbb{P}\left(X_i > \frac{n}{k}\left(1 + 3\sqrt{\frac{k \ln k}{n}}\right)\right) \leq 1/k^3.$$

Proof. Set $\mu = \mathbb{E}(X_i) = \frac{n}{k}$ and $\epsilon = 3\sqrt{\frac{k}{n}\ln k} < 3\sqrt{\frac{k}{9k \ln k}\ln k} = 1$

$n > 9k \ln k$

$$\mathbb{P}\left(X_i > \mu\left(1 + 3\sqrt{\frac{k \ln k}{n}}\right)\right) = \mathbb{P}(X_i > \mu(1+\epsilon))$$

$$\leq e^{-\frac{\epsilon^2 \mu}{2+\epsilon}} < e^{-\frac{\epsilon^2 \mu}{3}} = e^{-3 \ln k} = \frac{1}{k^3}$$

What about the maximum load?

Claim. (Load of single server) If $n > 9k \ln k$, then

$$\mathbb{P}\left(X_i > \frac{n}{k} + 3\sqrt{\frac{n \ln k}{k}}\right) \leq 1/k^3.$$

What about $X = \max\{X_1, \dots, X_k\}$?

Note: $X_1, \dots, X_k$ are not (mutually) independent!

In particular: $X_1 + \cdots + X_k = n$

When non-trivial outcome of one RV can be derived from other RVs, they are non-independent.

Distributed Load Balancing

Claim. (Load of single server) If $n > 9k \ln k$, then

$$\mathbb{P}\left(X_i > \frac{n}{k} + 3\sqrt{n \ln k \,/k}\right) \le 1/k^3.$$

Claim. (Max load) Let $X = \max\{X_1, \dots, X_k\}$. If $n > 9k \ln k$, then

$$\mathbb{P}\left(X > \frac{n}{k} + 3\sqrt{n \ln k \,/k}\right) \le 1/k^2.$$

Union Bound: $\mathbb{P}(\boldsymbol{A_1} \cup \boldsymbol{A_2} \cdots \cup \boldsymbol{A_n}) \le \Sigma_i\, \mathbb{P}(\boldsymbol{A_i})$

Always holds. No assumption on A_i's

Distributed Load Balancing

Claim. (Load of single server) If $n > 9k \ln k$, then

$$\mathbb{P}\left(X_i > \frac{n}{k} + 3\sqrt{n \ln k \,/k}\right) \le 1/k^3.$$

Claim. (Max load) Let $X = \max\{X_1, \dots, X_k\}$. If $n > 9k \ln k$, then

$$\mathbb{P}\left(X > \frac{n}{k} + 3\sqrt{n \ln k \,/k}\right) \le 1/k^2.$$

Union Bound: $\mathbb{P}(\boldsymbol{A_1} \cup \boldsymbol{A_2} \cdots \cup \boldsymbol{A_n}) \le \Sigma_i\, \mathbb{P}(\boldsymbol{A_i})$

Proof.

$$\mathbb{P}\left(X > \frac{n}{k} + 3\sqrt{n \ln k \,/k}\right) = \mathbb{P}\left(\left\{X_1 > \frac{n}{k} + 3\sqrt{n \ln k \,/k}\right\} \cup \cdots \cup \left\{X_k > \frac{n}{k} + 3\sqrt{n \ln k \,/k}\right\}\right)$$

$$\le \mathbb{P}\left(X_1 > \frac{n}{k} + 3\sqrt{\frac{n \ln k}{k}}\right) + \cdots + \mathbb{P}\left(X_k > \frac{n}{k} + 3\sqrt{n \ln k \,/k}\right) \le k \cdot \frac{1}{k^3} = 1/k^2$$