CSE 312

Foundations of Computing II

Lecture 18: Chernoff Bounds and Applications

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Stefano Tessaro
tessaro@cs.washington.edu

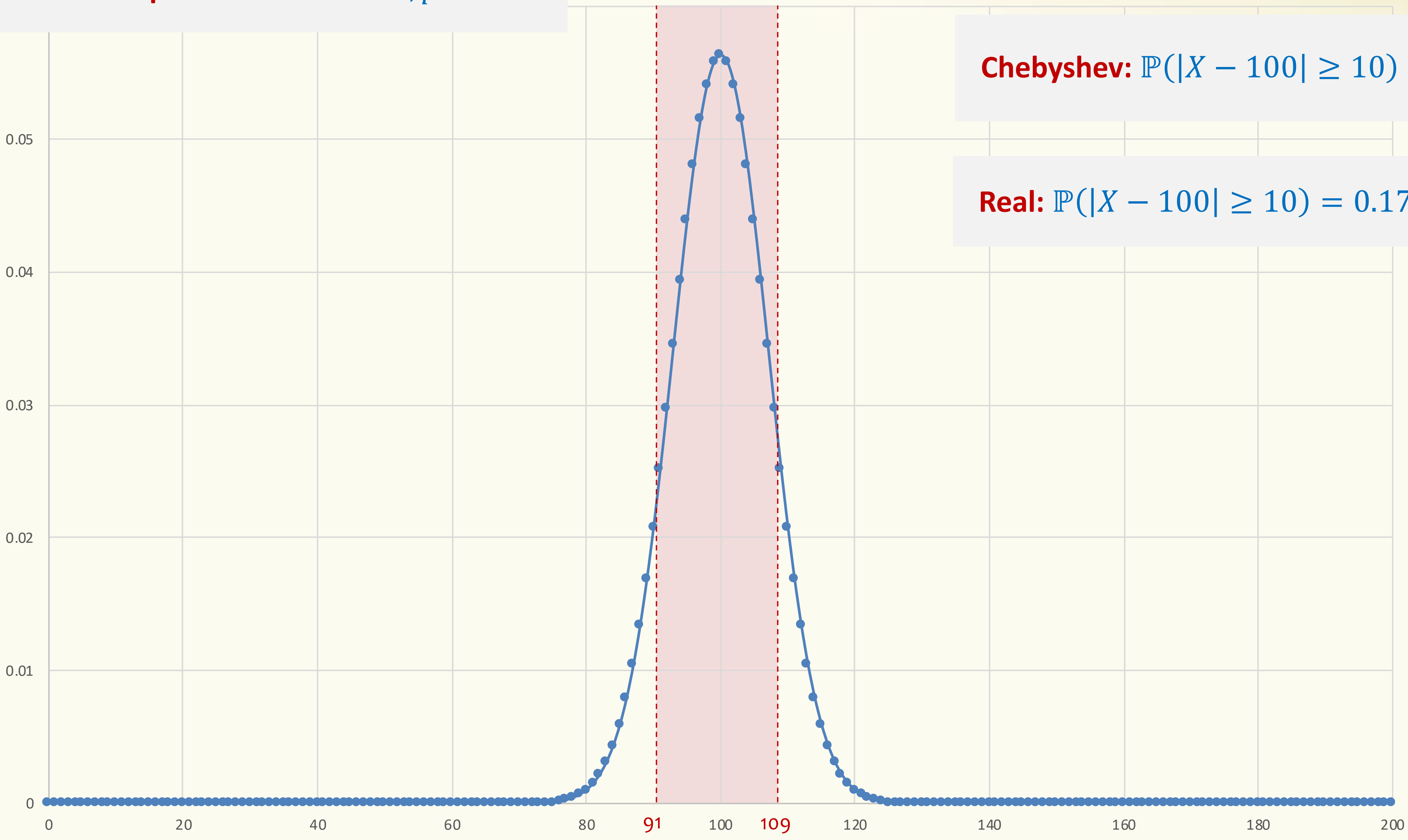
Class survey

- https://forms.gle/wZ2bXyuxdS8EKenc7
- Please elaborate on low scores in the comments portion. (This will help us fix things.)

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

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

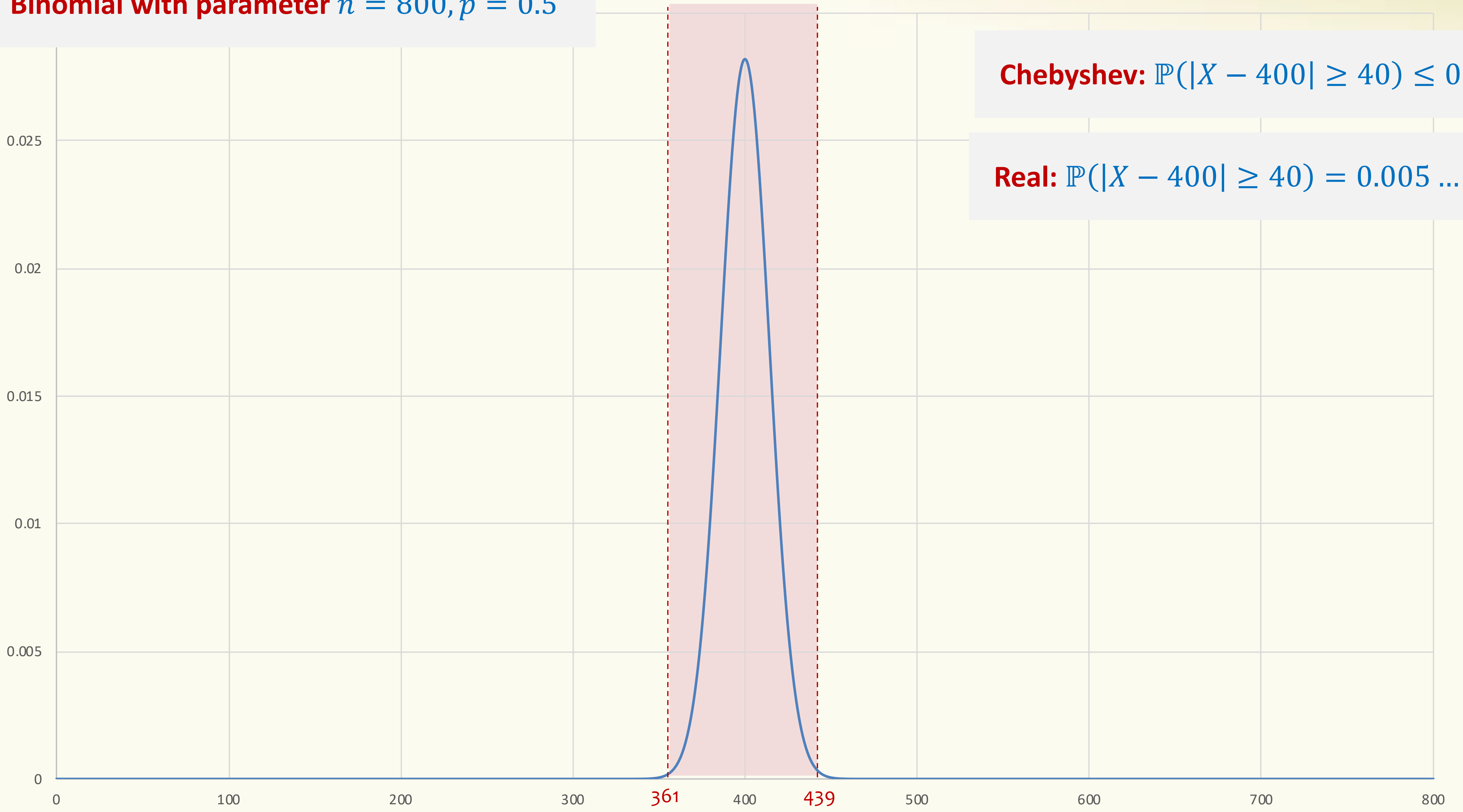
Real: $\mathbb{P}(|X - 100| \geq 10) = 0.179 \ldots$



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

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 $\epsilon > 0$,

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

In particular,

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

[Also holds with $(\leq, \geq) \to (<, >)$]

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.

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 $\epsilon > 0$,

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

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 $\epsilon > 0$,

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

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 $\epsilon > 0$,

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

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

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

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

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

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

Example:

- $n = 10^6, k = 1000$
- $\frac{n}{k} + 3\sqrt{n \ln k / k} \approx 1249$
- *"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

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

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

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

$$= \mathbb{P}\left(X_i > \mu(1+\epsilon)\right)$$

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

What about the maximum load?

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

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

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

Note: $X_1, \dots, X_k$ are <u>not</u> (mutually) independent!

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

Remember: *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

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

Proof.

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

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

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 is a requirement for all NLP/ML classes
- We want to estimate p <u>without</u> asking all M students!

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

Polling (cont'd)

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

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

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

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

Wanted: $\mathbb{P}\left(\left|\hat{P} - p\right| > \theta\right) < \delta$

For which n is this true?!

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

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

$$= \mathbb{P}(|\textstyle\sum_i^n X_i - np| > n\theta)$$

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

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

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

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

We have proved:

$$\mathbb{P}\left(\left|\hat{P} - p\right| > \theta\right) < 2\exp\left(-\frac{\theta^2}{2+\theta}n\right)$$

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

$$n \ge \ln(1/\delta)\frac{2+\theta}{\theta^2}$$

Polling – Summary

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

$$n \geq \ln(1/\delta) \frac{2+\theta}{\theta^2},$$

then

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

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

e.g. $\theta = \delta = 0.1, n \geq 484$

Central question in CS and statistics – can we do better?!

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:

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

- Find t minimizing the right-hand-side.