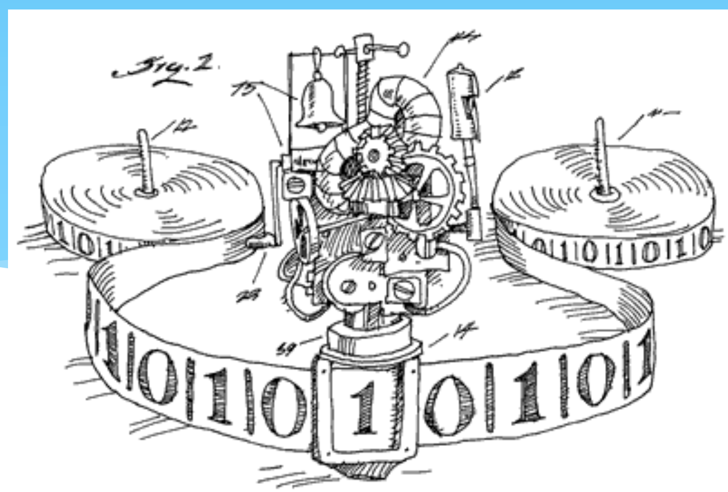
EECS 376: Foundations of Computer Science

Ali Movaghar

Lecture 22



Agenda

* **Last week:** Quicksort, Skip Lists
  * Randomized algorithms that perform well *in expectation*
  * But what's the *probability* of getting a good result?

* **Today:** Concentration Bounds
  * Variance and Chebyshev's inequality
  * Chernoff-Hoeffding bounds
  * Examples: Flipping coins and Polling

several randomized algorithm analyses use Chernoff bound + Union bound
EECS 572: Randomness and Computation for a "real" application

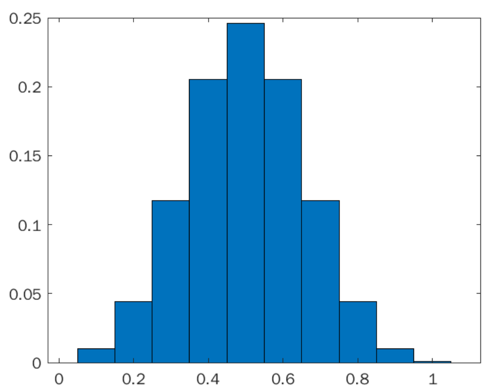
How many heads?

- We want to determine if a coin is fair or not.
- **Q:** How *suspicious* would we be if we flip it n times and see k heads, for the following values of n and k?
  - $n = 100, k = 51$
  - $n = 10{,}000, k = 5{,}100$
  - $n = 1{,}000{,}000, k = 510{,}000$
- Need to estimate $\Pr[X \geq k]$, where X is the number of heads after flipping a fair coin n times!

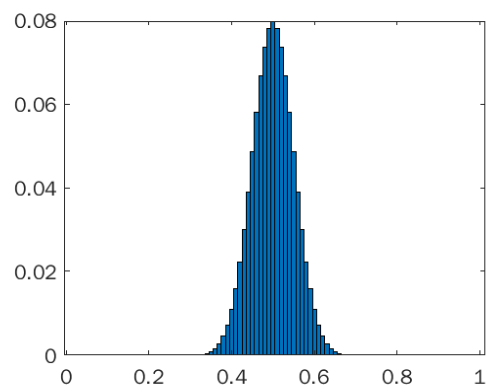
Law of Large Numbers

- **(Informal)** If $X_1, X_2, \ldots$ are ***independent, identically distributed*** (***i.i.d.***) RVs w/ expectation μ, then $\frac{1}{n}\sum_{i=1}^{n} X_i$ converges to μ (a <u>constant</u>) as $n \to \infty$.
- **Example:** The fraction of heads obtained when flipping a fair coin n times converges to $1/2$ as $n \to \infty$.

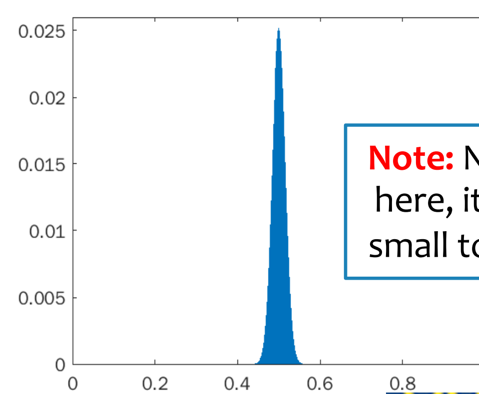
The graphs plot $\Pr\left[\frac{1}{n}\sum_{i=1}^{n} X_i = a\right]$; bars are for possible values of a.



$n = 10$



$n = 100$



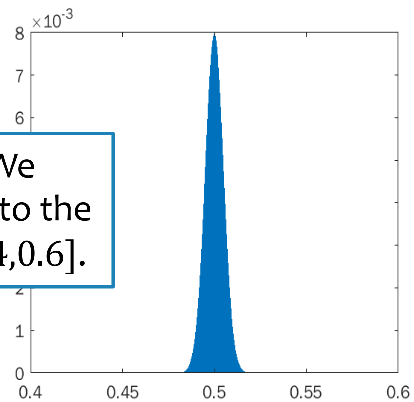
$n = 1000$

Note: Nothing's 0 here, it's just too small to show up.

Limitations of LLN

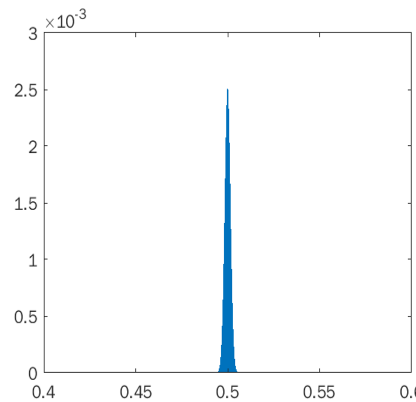
(**Informal**) If $X_1, X_2, \ldots$ are ***independent, identically distributed*** (***i.i.d.***) RVs w/ expectation μ, then $\frac{1}{n}\sum_{i=1}^{n} X_i$ converges to μ (a constant) as $n \to \infty$.

- LLN says distribution of sum is "concentrated" around its expectation as $n \to \infty$. (However, it doesn't say how quickly it happens or what the distribution looks like.)
- **Example:** The probability of seeing at least 0.51-fraction of heads when flipping a *fair* coin n times goes to *zero* as $n \to \infty$. How fast? Not clear.

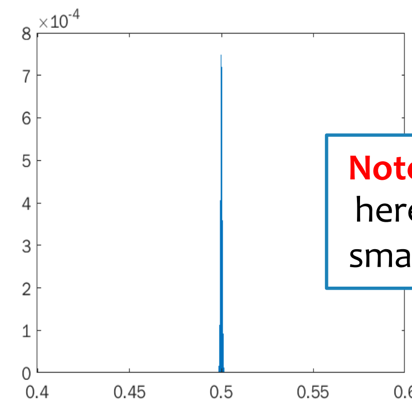


$n = 10{,}000$

Note: We zoomed into the region $[0.4, 0.6]$.



$n = 100{,}000$



$n = 1{,}000{,}000$

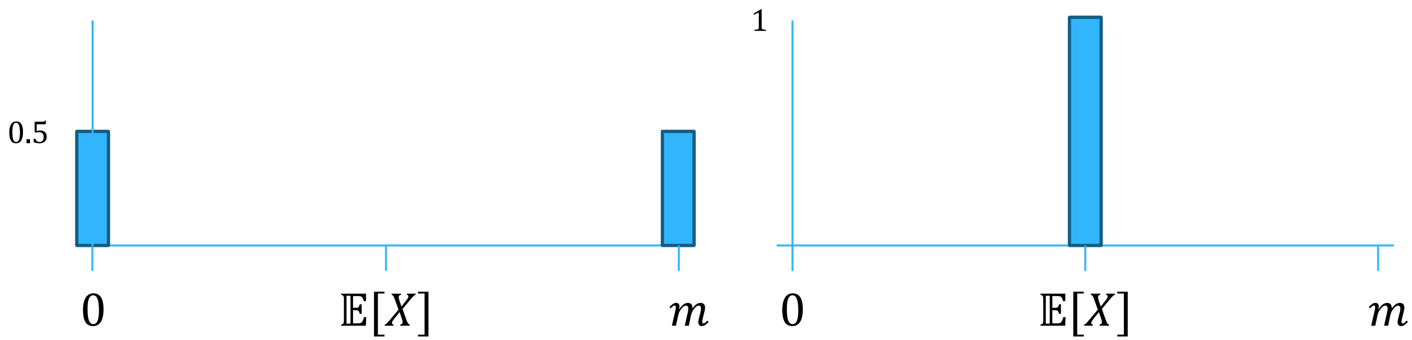
Note: Nothing's 0 here, it's just too small to show up.

Variance

- The ***variance*** of a random variable X is the average <u>*squared-distance*</u> of X from its mean, i.e.,

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

- The ***standard deviation*** is $\mathbf{SD}(X) = \sqrt{\mathbf{Var}(X)}$ (it's an <u>upper bound</u> on the average distance of X from $\mathbb{E}[X]$).



Example

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

* You have a *biased* coin that has probability p of heads, $1 - p$ of tails. For a single flip of the coin:
* Let $X_i = \begin{cases} 1 & \text{if flip } i \text{ is Heads} \\ 0 & \text{if flip } i \text{ is Tails} \end{cases}$
* Then $\mathbb{E}[X_i] = \Pr[X_i = 1] = p$;
* $\mathbb{E}[X_i^2] = 0 \cdot \Pr[X_i^2 = 0] + 1 \cdot \Pr[X_i^2 = 1] = p$
* $\mathbf{Var}(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = p - p^2 = p(1 - p)$
* What about for n flips?

Variance of Sum of Independent RVs

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

* **Fact:** For a sum of *independent* RVs $X = X_1 + \cdots + X_n$, we have $\mathbf{Var}(X) = \mathbf{Var}(\sum X_i) = \sum \mathbf{Var}(X_i)$.
* For one flip of a coin with probability p of heads, we saw that:
  * $\mathbb{E}[X_i] = p$
  * $\mathbf{Var}(X_i) = p(1-p)$
* For n flips, we have:
  * $\mathbb{E}[X] = \sum \mathbb{E}[X_i] = np$
  * $\mathbf{Var}(X) = \sum \mathbf{Var}(X_i) = np(1-p)$

Chebyshev's Inequality

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

* **(Recall) Markov's Inequality:** For a *non-negative* RV X and $a > 0$:

$$\Pr[X \geq a] \leq \mathbb{E}[X]/a$$

* ***Chebyshev's Inequality:*** For *any* RV X and $a > 0$:

$$\Pr[|X - \mathbb{E}[X]| \geq a] \leq \mathbf{Var}(X)/a^2$$

* **Proof:** square both sides and apply Markov's ineq.
  * $\Pr[|X - \mathbb{E}[X]| \geq a] = \Pr[(X - \mathbb{E}[X])^2 \geq a^2]$ (sq. both sides)

$$\leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{a^2}$$ (apply Markov)

$$\leq \frac{\mathbf{Var}(X)}{a^2}$$ (defn. of Variance)

Chebyshev's Inequality

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

* ***Chebyshev's Inequality:*** For *any* RV X and $a > 0$:
$$\Pr[|X - \mathbb{E}[X]| \geq a] \leq \mathbf{Var}(X)/a^2$$
* **Example:** What's the probability of getting $\leq$ 49% or $\geq$ 51% heads in n tosses of a *fair* coin?
* $\Pr\left[\left|X - \frac{n}{2}\right| \geq 0.01n\right] \leq \frac{\mathbf{Var}(X)}{(0.01n)^2} = 10{,}000 \cdot \frac{n \cdot 1/4}{n^2} = \frac{2{,}500}{n}$
  * $n = 10{,}000 \Longrightarrow \Pr[\text{deviating by } 1\%] \leq 1/4$
  * $n = 1{,}000{,}000 \Longrightarrow \Pr[\text{deviating by } 1\%] \leq 1/400$

Chebyshev's Inequality

For any RV X and $a > 0$:
$\Pr[|X - \mathbb{E}[X]| \geq a] \leq \mathbf{Var}(X)/a^2$

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

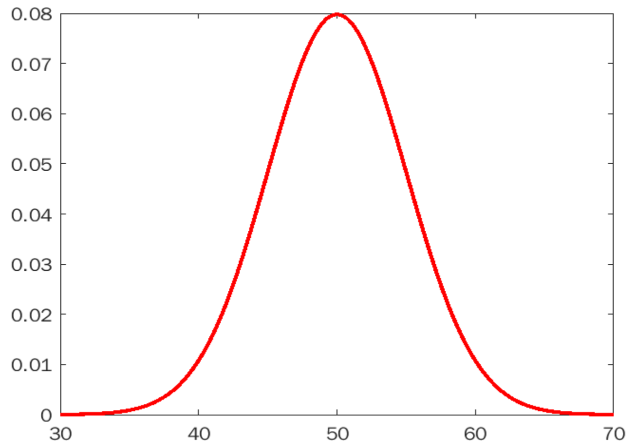
* For a sum of ***i.i.d.*** RVs $X = X_1 + \cdots + X_n$, we are often interested in $\frac{1}{n}X$ rather than X itself, since $\mathbb{E}\left[\frac{1}{n}X\right] = \mathbb{E}[X_i]$ does not depend on n (unlike $\mathbb{E}[X] = n\mathbb{E}[X_i]$).
* **Fact:** For a constant c, $\mathbf{Var}(cX) = c^2\mathbf{Var}(X)$.
* ***Chebyshev (Alternative):*** For a sum of ***i.i.d.*** $X = X_1 + \cdots + X_n$:

$$\Pr\left[\left|\frac{1}{n}X - \mathbb{E}[X_i]\right| \geq \varepsilon\right] \leq \frac{\mathbf{Var}(X_i)}{\varepsilon^2 n}$$

* **Example:** What's the probability of getting $\leq 49\%$ or $\geq 51\%$ heads in n tosses of a fair coin?
  * $\Pr\left[\left|\frac{1}{n}X - \frac{1}{2}\right| \geq 0.01\right] \leq \frac{\mathbf{Var}(X_i)}{0.01^2 n} = 10{,}000\frac{1/4}{n} = \frac{2{,}500}{n}$

Normal Distribution

- A ***normal distribution*** has a *bell-curve* shape and is characterized by two parameters, *mean* and *standard deviation*.
  - **Examples:** Height, exam scores, measurement error, are "normal-like"…
- ***66-95-99.7 rule:*** ≈66 / 95 / 99.7% of the area under the curve (i.e., probability) is within 1 / 2 / 3 SD from the mean, respectively

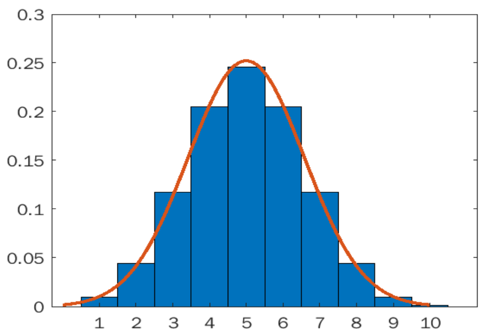


mean 50, SD 5

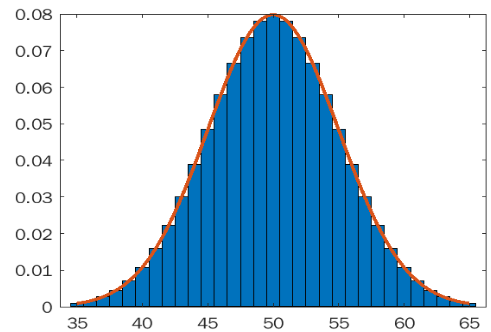
Note: Distribution is from $-\infty$ to ∞ and nothing's 0 here.

Central Limit Theorem

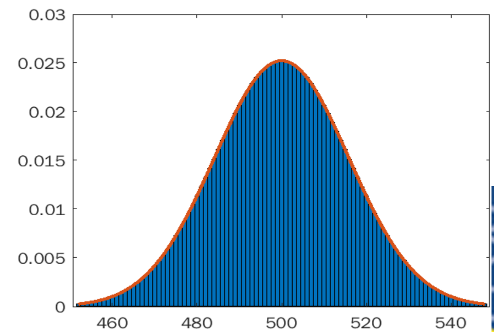
- **(Informal)** For large n, the sum $X = X_1 + \cdots + X_n$ of n ***i.i.d.*** RVs is "*close*" to a normal distribution with mean = $\mathbb{E}[X]$and SD = $\mathbf{SD}(X)$
- **Example:** The number of heads after flipping a fair coin n times, is "close" to a normal distribution with mean $n/2$ and standard deviation $\sqrt{n}/2$.
- **Q:** How suspicious are we if see $510{,}000$ heads after $1{,}000{,}000$ tosses?



$n = 10$
mean 5, SD 1.58



$n = 100$
mean 50, SD 5



$n = 1000$
mean 500, SD 15.8

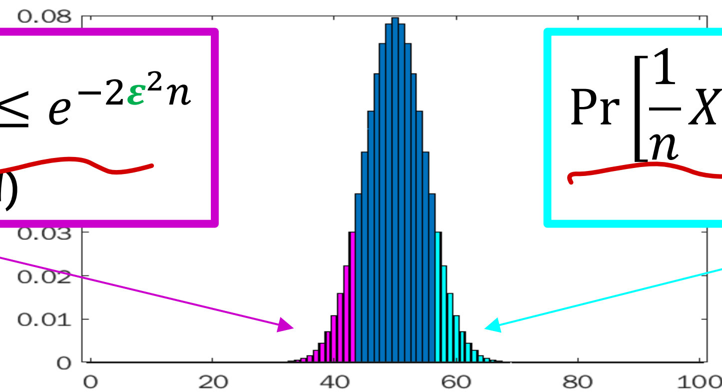
Chernoff-Hoeffding Bounds

(Usually tighter than Chebyshev)

* If $X = X_1 + X_2 + \cdots + X_n$ is the sum of n ***i.i.d.*** RVs with **each $X_i \in [0, 1]$**, then, for any $\varepsilon > 0$:

$$\Pr\left[\frac{1}{n}X \le \mathbb{E}[X_i] - \varepsilon\right] \le e^{-2\varepsilon^2 n}$$

(lower tail bound)



$$\Pr\left[\frac{1}{n}X \ge \mathbb{E}[X_i] + \varepsilon\right] \le e^{-2\varepsilon^2 n}$$

(upper tail bound)

Example: If we flip a fair coin $n = 1{,}000{,}000$ times, the probability we see $\ge (50 + \mathbf{1})\%$ heads is at most $e^{-2(\mathbf{0.01})^2 \cdot 1{,}000{,}000} = e^{-200}$!

Chernoff-Hoeffding Bounds
Proof Sketch

* $X = \sum_{i=1}^{n} X_i$ is the sum of n independent indicators with $E[X_i] = p$.
* $\Pr(X - \mathbb{E}(X) \geq t) = \Pr\left(e^{s(X-\mathbb{E}(X))} \geq e^{st}\right)$ for any $s > 0$ we wish.

$$\leq \frac{\mathbb{E}\left(e^{s(X-\mathbb{E}(X))}\right)}{e^{st}} \qquad \text{(Markov)}$$

$$= \frac{\mathbb{E}\left(e^{s(X_1+\cdots+X_n-\mathbb{E}(X_1)-\cdots-\mathbb{E}(X_n))}\right)}{e^{st}} \qquad \text{(Defn of } X\text{)}$$

$$= \frac{\mathbb{E}\left(\prod_{i=1}^{n} e^{s(X_i-\mathbb{E}(X_i))}\right)}{e^{st}}$$

"Hoeffding's Lemma"

$$= \frac{\prod_{i=1}^{n} \mathbb{E}\left(e^{s(X_i-\mathbb{E}(X_i))}\right)}{e^{st}} \leq \frac{\left(e^{\frac{s^2}{8}}\right)^n}{e^{st}} = \exp(-st + s^2 n/8)$$

$$= \exp(-2t^2/n). \qquad \text{(Choose } s = 4t/n.\text{)}$$

6/17/24

Hoeffding's Lemma

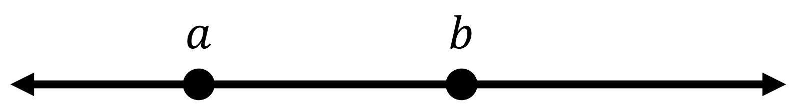
Let X be a real random variable such that X $\in [0, 1]$ almost surely. Then for any real number of s, we have:

$$\boxed{\mathrm{E}[e^{s(X-E[X])}] \leq \exp(\frac{1}{8}s^2)}$$

The above inequality is proved using the convexity of exponential functions and the arithmetic and geometric means (AM-GM) inequality.

Some Useful Inequalities

* Let X be a random variable. Then:
  * $\Pr[X \geq a] \geq \Pr[X > a]$ and $\Pr[X \leq a] \geq \Pr[X < a]$
    * Why? $\Pr[X \geq a] = \Pr[X > a] + \Pr[X = a]$
  * $\Pr[X \geq a] = 1 - \Pr[X < a] \geq 1 - \Pr[X \leq a]$
  * $\Pr[X \leq a] = 1 - \Pr[X > a] \geq 1 - \Pr[X \geq a]$



* Let $a \leq b$. Then:
  * $\Pr[X \geq a] \geq \Pr[X \geq b]$ and $\Pr[X \leq a] \leq \Pr[X \leq b]$
    * Why? $\Pr[X \geq a] = \Pr[X \geq b] + \Pr[b > X \geq a]$

Union Bound

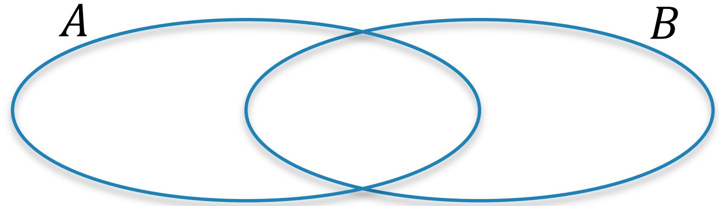
$$\Pr\left[\frac{1}{n}X \le \mathbb{E}[X_i] - \boldsymbol{\varepsilon}\right] \le e^{-2\boldsymbol{\varepsilon}^2 n}$$

(lower tail bound)

$$\Pr\left[\frac{1}{n}X \ge \mathbb{E}[X_i] + \boldsymbol{\varepsilon}\right] \le e^{-2\boldsymbol{\varepsilon}^2 n}$$

(upper tail bound)

* For any two events A, B:
  * $\Pr[A \cup B] \le \Pr[A] + \Pr[B]$



* This also implies:
  * $\Pr[A \cap B] \ge 1 - (\Pr[\overline{A}] + \Pr[\overline{B}])$

* ***Example:*** Combined Chernoff-Hoefding bound:
  * $\Pr\left[\left|\frac{1}{n}X - \mathbb{E}[X_i]\right| \ge \varepsilon\right] \le 2e^{-2\varepsilon^2 n}$

Since $\Pr\left[\left|\frac{1}{n}X - \mathbb{E}[X_i]\right| \ge \varepsilon\right] = \Pr\left[\left(\frac{1}{n}X \le \mathbb{E}[X_i] - \varepsilon\right) \cup \left(\frac{1}{n}X \ge \mathbb{E}[X_i] + \varepsilon\right)\right]$

Polling

* There are m candidates for president. How can we estimate their relative support *without* asking the entire population?
* **A:** Sample people at random and compute the relative frequencies.
* Two types of "accuracy":
  1. The probability that we indeed obtain a "good" estimate
  2. The extent to which our estimate approximates reality
* **Fine print:** "This poll has been conducted with a ***confidence level*** of **95%** and ***statistical error*** of $\pm$**2%**"

Polling

- **Algorithm for one candidate (approval rating):**
  - Sample at random n people (ask: "Do you support?" Yes/No)
  - Let X be the number of supporters
  - Return X/n as an estimate
- Let $0 \le p \le 1$ be the true level of support. How large does n have to be so that we get good "accuracy" with high "confidence"?
- **Fine print:** "This poll has been conducted with a ***confidence level*** of **95%** and ***statistical error*** of $\pm$**2%**"
- Thus, we want $\Pr\left[\left|\frac{1}{n}X - p\right| \le 0.02\right] \ge 0.95.$

Combined Chernoff-Hoeffding Bound

Combined Chernoff-Hoeffding:
$$\Pr\left[\left|\frac{1}{n}X - \mathbb{E}[X_i]\right| \geq \varepsilon\right] \leq 2e^{-2\varepsilon^2 n}$$

Compare to Chebyshev:
$$\Pr\left[\left|\frac{1}{n}X - \mathbb{E}[X_i]\right| \geq \varepsilon\right] \leq \frac{\mathbf{Var}(X_i)}{\varepsilon^2 n}$$

* **Goal:** Find n such that $\Pr\left[\left|\frac{1}{n}X - p\right| \leq 0.02\right] \geq 0.95.$
* Define indicators for $i = 1..n$:
$$X_i = \begin{cases} 1, & \text{person } i \text{ supports the candidate} \\ 0, & \text{otherwise} \end{cases}$$
* Then $\mathbb{E}[X_i] = \Pr[X_i = 1] = p$ and $X = X_1 + X_2 + \cdots + X_n$.
* **Q:** What should the value of n be to satisfy the fine print?
* **Equivalently:** We want $\Pr\left[\left|\frac{1}{n}X - p\right| > 0.02\right] \leq 0.05.$
* By the combined CH bound:
$$\Pr\left[\left|\frac{1}{n}X - p\right| > 0.02\right] \leq \Pr\left[\left|\frac{1}{n}X - p\right| \geq 0.02\right] \leq 2e^{-2\cdot 0.02^2 n}$$

Polling Analysis

Combined Chernoff-Hoeffding:
$$\Pr\left[\left|\frac{1}{n}X - \mathbb{E}[X_i]\right| \geq \varepsilon\right] \leq 2e^{-2\varepsilon^2 n}$$

- By the combined CH bound:

$$\Pr\left[\left|\frac{1}{n}X - p\right| > 0.02\right] \leq \Pr\left[\left|\frac{1}{n}X - p\right| \geq 0.02\right] \leq 2e^{-2\cdot 0.02^2 n}$$

- Therefore, we need $2e^{-2\cdot 0.02^2 n} \leq 0.05$

$$\Longrightarrow 40 \leq e^{2\cdot 0.02^2 n} \Longrightarrow \ln 40 \leq 2 \cdot 0.02^2\, n \Longrightarrow n \geq 4612$$

- **Remark:** n does not depend on the population size!
- **Q:** How large should n be if we want error ε with probability $\leq \delta$?
  - $2e^{-2\varepsilon^2 n} \leq \delta \Leftrightarrow n \geq \dfrac{\ln(2/\delta)}{2\varepsilon^2}$.

Polling General Case

* **Algorithm for m candidates:**
  * Sample at random n people (ask: “Who do you support?”)
  * Let $X^{(j)}$ be the number of supporters of candidate j
  * For each j: Return $X^{(j)}/n$
* **Fine print:** “This poll has been conducted with a ***confidence level*** of $1-\delta$ and ***statistical error*** of $\pm\varepsilon$”
* **Formally:** Let $p_1, \ldots, p_m$ be the support levels of the candidates.
* **We want:** $\Pr\left[\text{for every } j = 1..m\text{: } \left|\frac{1}{n}X^{(j)} - p_j\right| \le \varepsilon\right] \ge 1 - \delta.$

Polling General Case

* **We want:** $\Pr\left[\text{for every } j = 1..m\text{: } \left|\frac{1}{n}X^{(j)} - p_j\right| \leq \varepsilon\right] \geq 1 - \delta.$
* How many samples do we need now?
* When $m = 1$, we need $n \geq \frac{1}{2\varepsilon^2} \ln\left(\frac{2}{\delta}\right)$
* **Wrong answer:** for m candidates we need $n \geq m \cdot \frac{1}{2\varepsilon^2} \ln\left(\frac{2}{\delta}\right)$
* **Sampling Theorem:** If $n \geq \frac{1}{2\varepsilon^2} \ln\left(\frac{2m}{\delta}\right)$ then we can assert that our estimates satisfy the fine print. (proof via union bound)
* **Conclusion:** The dependence on m is logarithmic!

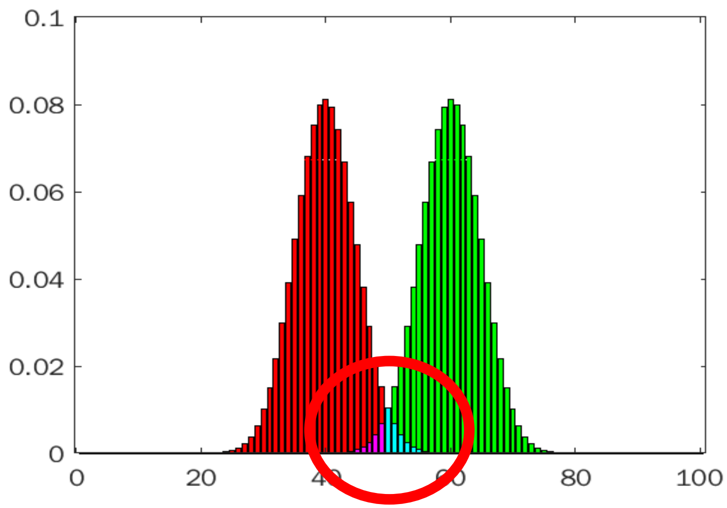
EECS M ECE CSE

Distinguishing Biased Coins

* You're given a coin that is ε-biased to either heads or tails.
  * i.e., $\Pr[H] = \frac{1}{2} + \varepsilon$ and $\Pr[T] = \frac{1}{2} - \varepsilon$ **or** $\Pr[H] = \frac{1}{2} - \varepsilon$ and $\Pr[T] = \frac{1}{2} + \varepsilon$
* To determine if it's biased towards heads, you flip the coin n times.
  * If you see at least $\frac{1}{2}n$ heads, you guess "yes"
  * Otherwise, you guess "no".

Note: We have ***two-sided*** error; ***false positives*** and ***false negatives*** are possible!

Q: How large should n be to guarantee an error probability of δ?



Probability of False Negatives

If $X = X_1 + X_2 + \cdots + X_n$ is the sum of n ***i.i.d.*** RVs with **each $\boldsymbol{X_i} \in [\mathbf{0}, \mathbf{1}]$**, then, for any $\boldsymbol{\varepsilon} > 0$:

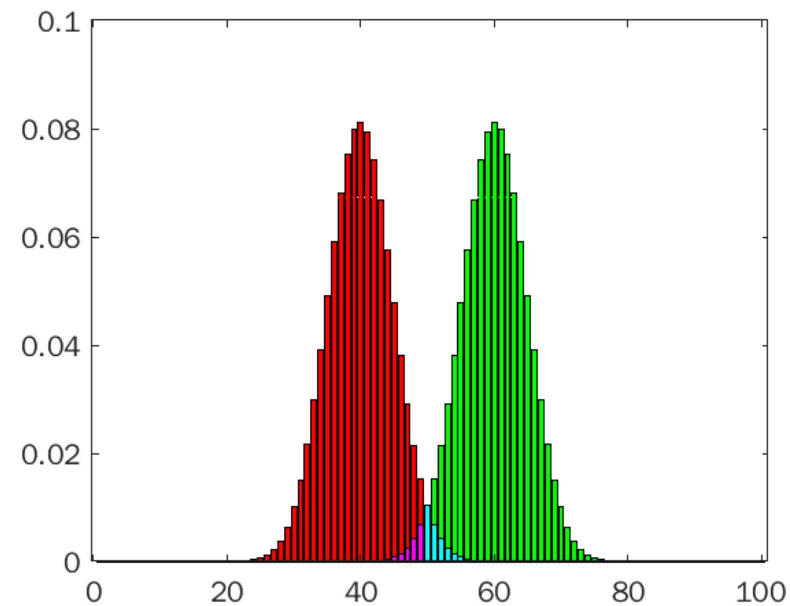
$$\Pr\left[\frac{1}{n}X \leq \mathbb{E}[X_i] - \boldsymbol{\varepsilon}\right] \leq e^{-2\boldsymbol{\varepsilon}^2 n}$$

(***lower tail bound***)

- Let X_i be the indicator RV for whether i'th coin flip was H.
- Suppose the coin we had was ε-biased towards **heads**.
  - Then $\mathbb{E}[X_i] = \frac{1}{2} + \varepsilon$.
- **Q:** When do we get an error (**false negative**) in this case?
- **A:** When $\frac{1}{n}X < \frac{1}{2} = \mathbb{E}[X_i] - \varepsilon$
- Therefore:

$$\Pr[\text{error}|H \text{ bias}]$$

$$= \Pr\left[\frac{1}{n}X < \mathbb{E}[X_i] - \varepsilon\right] \leq e^{-2\varepsilon^2 n}$$



To determine if it's biased towards heads, you flip the coin n times.

If you see at least $\frac{1}{2}n$ heads, you guess "yes"

Otherwise, you guess "no".

Probability of False Positives

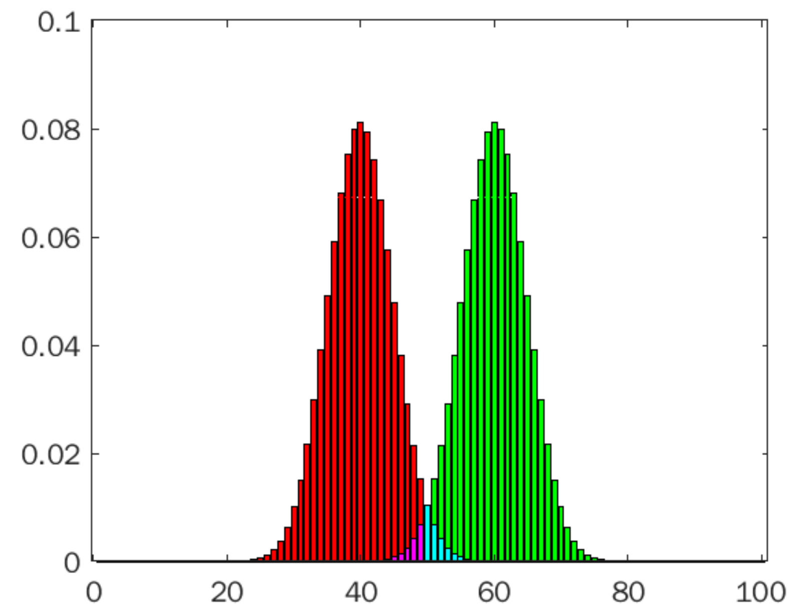
If $X = X_1 + X_2 + \cdots + X_n$ is the sum of n ***i.i.d.*** RVs with **each $\boldsymbol{X_i} \in [\mathbf{0}, \mathbf{1}]$**, then, for any $\boldsymbol{\varepsilon} > 0$:

$$\Pr\left[\frac{1}{n}X \geq \mathbb{E}[X_i] + \boldsymbol{\varepsilon}\right] \leq e^{-2\boldsymbol{\varepsilon}^2 n}$$

(***lower tail bound***)

* Let X_i be the indicator RV for whether i'th coin flip was H.
* Suppose the coin we had was ε-biased towards **tails**.
  * Then $\mathbb{E}[X_i] = \frac{1}{2} - \varepsilon$.
* **Q:** When do we get an error (**false positive**) in this case?
* **A:** When $\frac{1}{n}X \geq \frac{1}{2} = \mathbb{E}[X_i] + \varepsilon$
* Therefore:

$$\Pr[\text{error}|T\text{ bias}]$$
$$= \Pr\left[\frac{1}{n}X \geq \mathbb{E}[X_i] + \varepsilon\right] \leq e^{-2\varepsilon^2 n}$$



To determine if it's biased towards heads, you flip the coin n times.
If you see at least $\frac{1}{2}n$ heads, you guess "yes"
Otherwise, you guess "no".

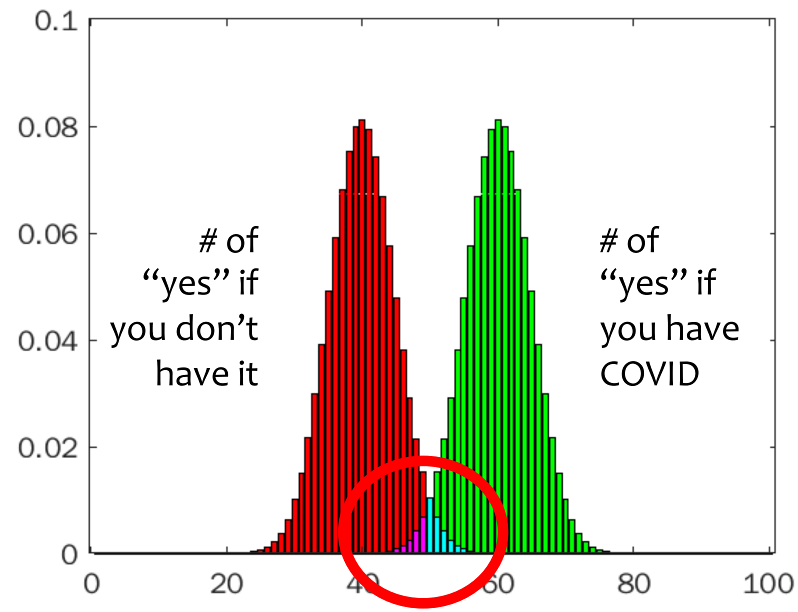
How large should n be?

- By previous analysis, $\Pr[\text{error}] \leq e^{-2\varepsilon^2 n}$.
  - We saw the error in either case is at most this.
- How large should n be if we want error to be $\leq \delta$?
  - $e^{-2\varepsilon^2 n} \leq \delta \Leftrightarrow n \geq \frac{\ln(1/\delta)}{2\varepsilon^2}$.
- **Example:** If $\varepsilon = 0.01$ and $\delta = 0.0001$ (correct 99.99%), then we need $n \geq \frac{\ln\left(0.0001^{-1}\right)}{2 \cdot 0.01^2} \approx 46{,}052$ flips.

Extra Practice

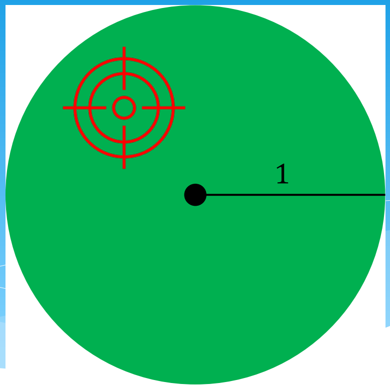
Decreasing Error

- A low-quality COVID test has *two-sided* error:
  - If a person has COVID, it says "yes" w.p. 2/3.
  - Otherwise, it says "no" w.p. 2/3.
  - Different runs are independent.
- You decide to buy and run the test n times and take the *majority* answer you get.
- **Q:** How large should n be to guarantee that the answer is correct w.p. $1-\delta$?
  - Same as distinguishing ε-biased coins with $\varepsilon = 1/6$!



Note: ***false positives*** and ***false negatives*** are possible!

Estimating π



- Suppose there is a 2x2 square with a unit circle inside
- **Q:** If we toss a dart *uniformly at random* towards the square, what's the probability that we hit the circle?
  - (area of circle)/(area of board) = $\pi/4$
- We toss n darts uniformly at random towards the square
  - X_i = indicator 0/1 RVs for whether we hit circle on i'th toss
  - **Q:** What is $\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$?
- **Q:** How might we estimate π by tossing darts?
  - It's *roughly* 4*fraction of times we hit circle; CH to bound error.

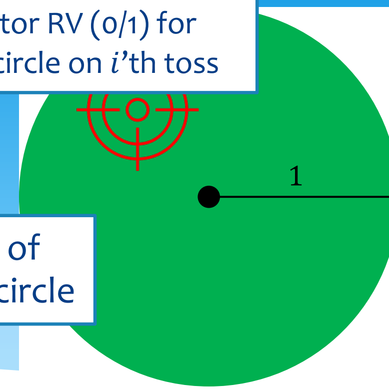
$$\Pr\left[\left|\frac{1}{n}X - \mathbb{E}[X_i]\right| \geq \varepsilon\right] \leq 2e^{-2\varepsilon^2 n}$$

(combined bound)

Math

(to show that this is a bad idea)

We toss n darts uniformly at random towards the square
Let X_i = indicator RV (0/1) for whether we hit circle on i'th toss



$\pi \approx$ 4*fraction of times we hit the circle

* Let $X = X_1 + \cdots + X_n$ be the number of times we hit the circle.
* $\mathbb{E}[X_i] = \frac{\pi}{4}$ so $\left|\frac{1}{n}X - \frac{\pi}{4}\right| < \varepsilon$ with probability $\geq 1 - 2e^{-2\varepsilon^2 n}$
* To estimate π within γ, i.e. $\left|\frac{4}{n}X - \pi\right| < \gamma$, set $\varepsilon = \frac{\gamma}{4}$.
    * $\left|\frac{4}{n}X - \pi\right| < \gamma \Leftrightarrow \left|\frac{1}{n}X - \frac{\pi}{4}\right| < \frac{\gamma}{4} = \varepsilon$
      (with probability $\geq 1 - 2e^{-2\varepsilon^2 n} = 1 - 2e^{-\gamma^2 n/8}$)
* For probability $\geq 1 - \delta$, set $n = 8\ln(2/\delta)/\gamma^2$.
    * $1 - 2e^{-\gamma^2 n/8} \geq 1 - \delta \Leftrightarrow \delta \geq 2e^{-\gamma^2 n/8} \Leftrightarrow \ln\delta/2 \geq -\gamma^2 n/8 \Leftrightarrow n \geq 8\ln(2/\delta)/\gamma^2$

Example: To get our estimate between 3.140 and 3.142 ($\gamma = 0.001$) 99.99% of the time ($\delta = 0.0001$), we should toss $n \approx$ 79,227,901 darts