Lecture 22

Concentration Inequalities

Motivating Example

Opinion Poll: You want to know how many Americans support increasing taxes.

How to estimate this number?

Motivating Example

<u>Opinion Poll</u>: You want to know how many Americans support increasing taxes.

How to estimate this number?

<u>Method</u>: Pick at random n people from the population and ask for their opinion.

Motivating Example

Opinion Poll: You want to know how many Americans support increasing taxes.

How to estimate this number?

Method: Pick at random n people from the population and ask for their opinion.

Want: The fraction of people supporting taxes increase among your poll is a "good" estimate to the overall fraction in the entire population.

a Mathy Formulation of the Problem

Suppose you have a coin w. heads prob. p

but you don't know p.

Goal: estimate p.

a Mathy Formulation of the Problem

Suppose you have a coin w. heads prob. p

but you don't know p.

Goal: estimate p.

Method: Toss the coin n times

Let $\hat{p}$ = frac. of times the coin

a Mathy Formulation of the Problem

Suppose you have a coin w. heads prob. p

but you don't know p.

Goal: estimate p.

Method: Toss the coin n times

Let $\hat{p}$ = frac. of times the coin turned out heads.

a Mathy Formulation of the Problem

Suppose you have a coin w. heads prob. p

but you don't know p.

Goal: estimate p.

Method: Toss the coin n times

Let $\hat{p}$ = frac. of times the coin turned out heads.

Want: $|p - \hat{p}| \leq \varepsilon$ Can you guarantee it?

a Mathy Formulation of the Problem

Suppose you have a coin w. heads prob. p

but you don't know p.

Goal: estimate p.

Method: Toss the coin n times

Let $\hat{p}$ = frac. of times the coin turned out heads.

Want: $|p - \hat{p}| \leq \varepsilon$ Can you guarantee it? No!

Settle for: $|p - \hat{p}| \leq \varepsilon$ w. confidence $1 - \delta$.

$$\Pr\left[|p - \hat{p}| \leq \varepsilon\right] \geq 1 - \delta$$

a Mathy Formulation of the Problem

Suppose you have a coin w. heads prob. p

but you don't know p.

Goal: estimate p.

Method: Toss the coin n times

Let $\hat{p}$ = frac. of times the coin turned out heads.

Q: How large should n be to guarantee

$$\Pr\left[|p - \hat{p}| \le \varepsilon\right] \geq 1 - \delta$$

A: $n \geq \dfrac{1}{4\varepsilon^2\delta}$ suffices.

For Example:

$\varepsilon = 0.05 \qquad \delta = 0.05$

$n = \dfrac{1}{4 \cdot (0.05)^3} = 2000$ suffices.

Q: Is it possible that everyone is above average (on the midterm)?

Q: Is it possible that everyone is above average (on the midterm)?

Q: How many students can be more than twice the average?

Q: Is it possible that everyone is above average (on the midterm)?

Q: How many students can be more than twice the average?

This is just Markov's Inequality.

Theorem (Markov's Inequality):

If X is a non-negative random variable then for any constant $c \geq 0$

$$\Pr[X \geq c] \leq \frac{\mathbb{E}[X]}{c}$$

Another formulation: for any constant $k \geq 0$

$$\Pr[X \geq k \cdot \mathbb{E}[X]] \leq \boxed{}$$

Theorem (Markov's Inequality):

If X is a non-negative random variable then for any constant $c \geq 0$

$$\Pr[X \geq c] \leq \frac{\mathbb{E}[X]}{c}$$

Proof:

Theorem (Markov's Inequality):

If X is a non-negative random variable then for any constant $c \geq 0$

$$\Pr[X \geq c] \leq \frac{\mathbb{E}[X]}{c}$$

Proof:

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \Pr[\omega]$$

$$\geq \sum_{\omega : X(\omega) \geq c} X(\omega) \cdot \Pr[\omega]$$

$$\geq \sum_{\omega : X(\omega) \geq c} c \cdot \Pr[\omega]$$

$$= c \cdot \sum_{\omega : X(\omega) \geq c} \Pr[\omega] = c \cdot \Pr[X \geq c] \quad \blacksquare$$

Theorem (Markov's Inequality):

If X is a non-negative random variable then for any constant $c \geq 0$

$$\Pr[X \geq c] \leq \frac{\mathbb{E}[X]}{c}$$

Example: Coin Tosses

You toss a fair coin n times. X - number of heads.

What's the prob. of $X \geq \frac{3}{4}n$?

Theorem (Markov's Inequality):

If X is a non-negative random variable then for any constant $c \geq 0$

$$\Pr[X \geq c] \leq \frac{\mathbb{E}[X]}{c}$$

Example: Coin Tosses

You toss a fair coin n times. X - number of heads.

What's the prob. of $X \geq \frac{3}{4}n$?

using Markov's Ineq. $\Pr[X \geq \frac{3}{4}n] \leq \frac{\mathbb{E}[X]}{\frac{3}{4}n} = \frac{\frac{1}{2}n}{\frac{3}{4}n} = \frac{2}{3}$

Is this a good upper bound?

Theorem (Markov's Inequality):

If X is a non-negative random variable then for any constant $k \geq 0$

$$\Pr[X \geq k \cdot \mathbb{E}[X]] \leq \frac{1}{k}$$

Example 2: Let $\mu \geq 0$, $k \geq 1$.

$$X = \begin{cases} \mu k, & \text{w.p. } \frac{1}{k} \\ 0, & \text{otherwise} \end{cases}$$

$$\mathbb{E}X = \boxed{\phantom{\mu k \cdot \frac{1}{k}}}$$

$$\Pr[X \geq k \cdot \mathbb{E}X] = \boxed{\phantom{\mu k \cdot \frac{1}{k}}}$$

Theorem (Markov's Inequality):

If X is a non-negative random variable then for any constant $c \geq 0$

$$\Pr[X \geq c] \leq \frac{\mathbb{E}[X]}{c}$$

Q: What happens if X can be negative?

Theorem (Markov's Inequality):

If X is a non-negative random variable then for any constant $c \geq 0$

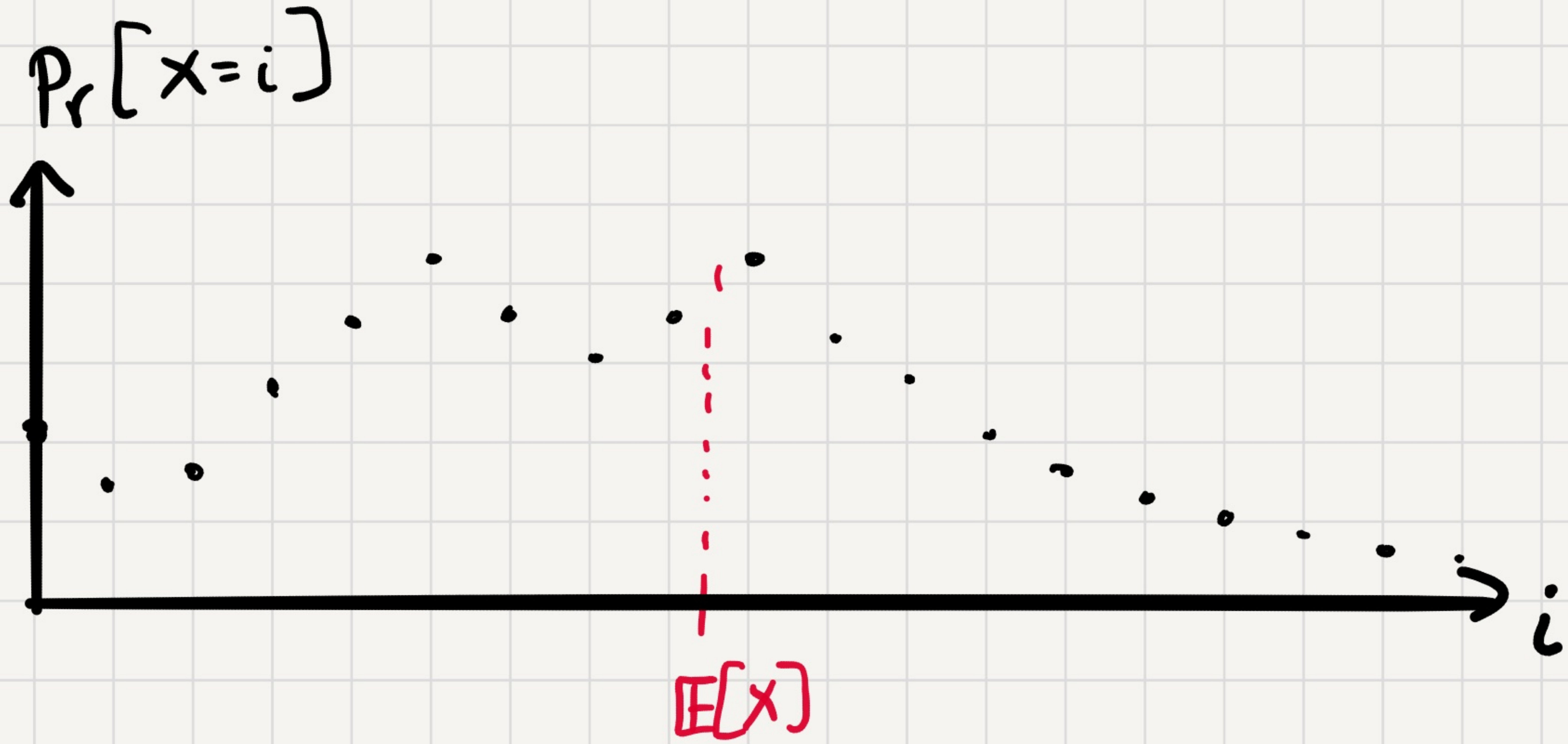
$$\Pr[X \geq c] \leq \frac{\mathbb{E}[X]}{c}$$

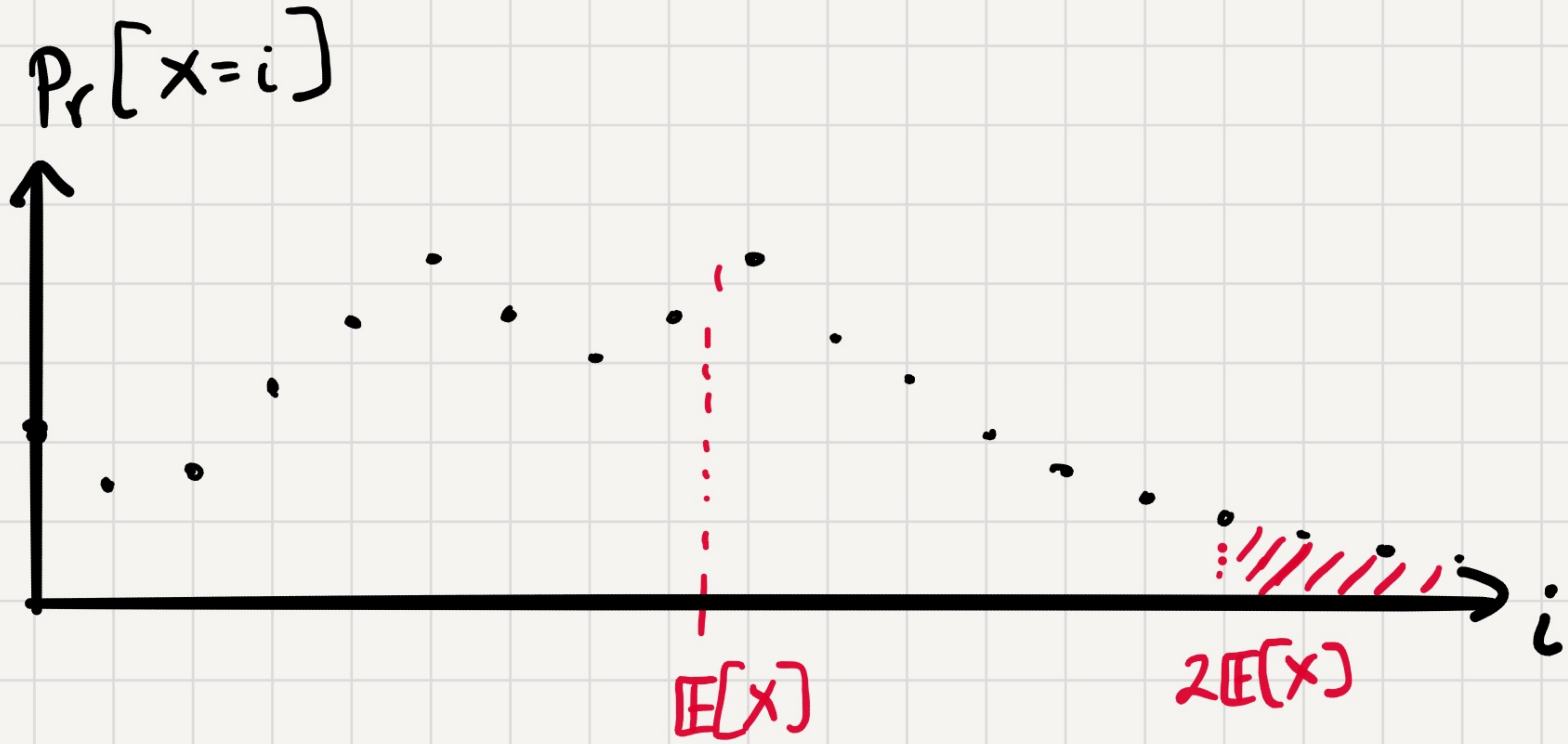
Q: What happens if X can be negative?

Example:

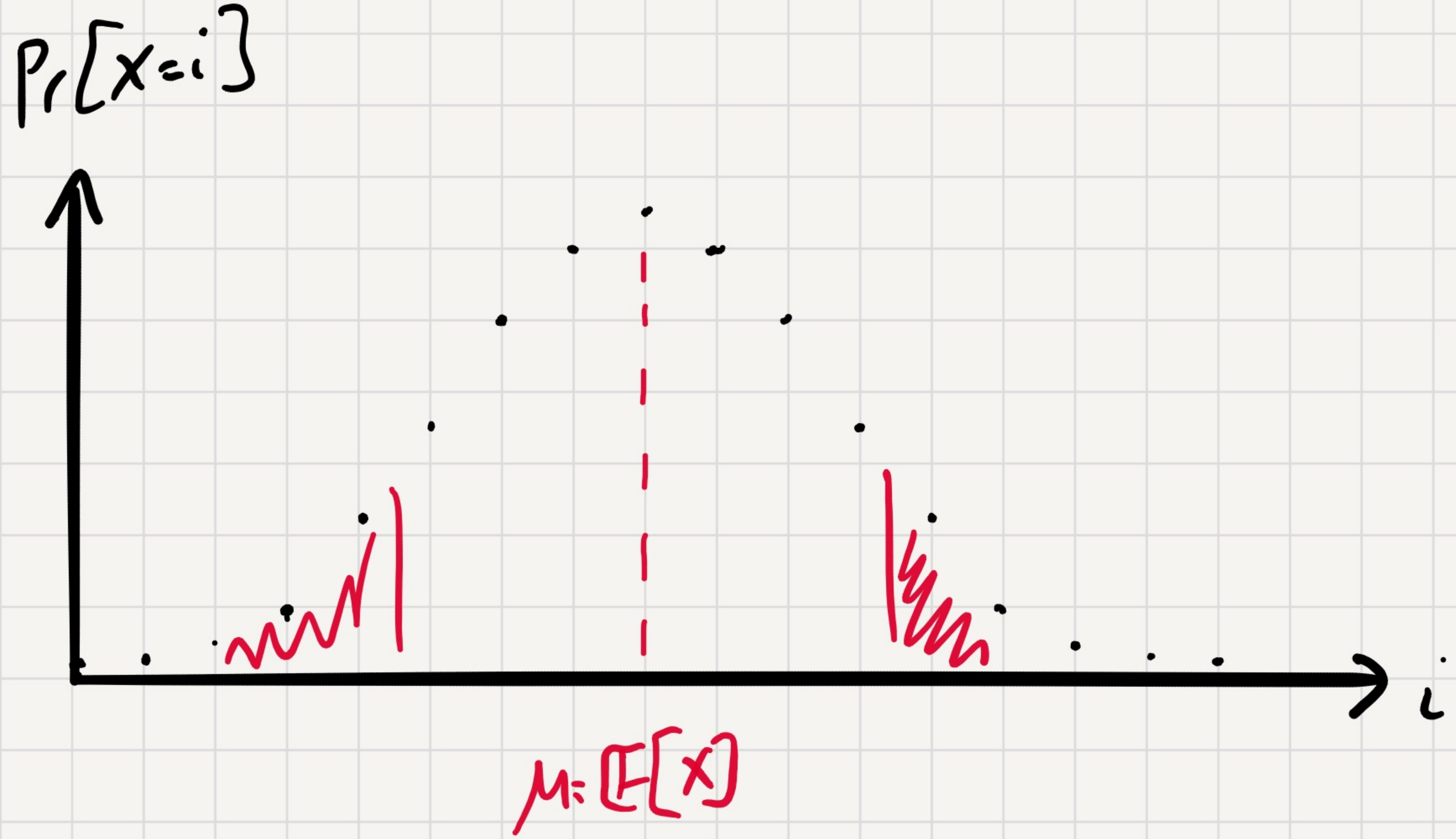
$$X = \begin{cases} +1, & \text{w.p. } \frac{1}{2} \\ -1, & \text{w.p } \frac{1}{2} \end{cases} \qquad \mathbb{E}X = 0$$

$$\Pr[X \geq 1] = \frac{1}{2} \not\leq \frac{\mathbb{E}X}{1}$$


Pr[X=i]
i
E[X]



Pr[X=i]
i
E[X]
2E[X]



What's the prob. to be "far" from the mean?

That is captured by Chebyshev's Inequality.

Theorem (Chebyshev's Inequality)

Let X be a r.v. with expectation μ

Then $$\Pr\left[\,|X-\mu| \geq c\,\right] \leq \frac{Var(X)}{c^2}$$

for any constant $c \geq 0$.

Theorem (Chebyshev's Inequality)

Let X be a r.v. with expectation μ

Then $$\Pr[\,|X-\mu| \geq c\,] \leq \frac{Var(X)}{c^2}$$

for any constant $c \geq 0$.

Equivalent Formulation:

If X has mean μ and stdev σ then

for any $k \geq 0$ $\quad \Pr[\,|X-\mu| \geq k\sigma\,] \leq \boxed{}$

Theorem (Chebyshev's Inequality)

Let X be a r.v. with expectation μ

Then $$\Pr[\,|X-\mu| \geq c\,] \leq \frac{Var(X)}{c^2}$$

for any constant $c \geq 0$.

Proof: $\Pr[\,|X-\mu| \geq c\,] =$

Theorem (Chebyshev's Inequality)

Let X be a r.v. with expectation μ

Then $$\Pr[\,|X-\mu| \geq c\,] \leq \frac{Var(X)}{c^2}$$

for any constant $c \geq 0$.

Proof: $\Pr[\,|X-\mu| \geq c\,] = \Pr[(X-\mu)^2 \geq c^2]$

Let $Y = (X-\mu)^2$. Y is a non-negative r.v.

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

$$\Pr[\,|X-\mu| \geq c\,] = \Pr[Y \geq c^2] \leq \frac{\mathbb{E}[Y]}{c^2} = \frac{Var(X)}{c^2} \quad \blacksquare$$

Theorem (Chebyshev's Inequality)

Let X be a r.v. with expectation μ

Then $$\Pr[\,|X-\mu| \geq c\,] \leq \frac{Var(X)}{c^2}$$

for any constant $c \geq 0$.

Example: Coin Tosses

You toss a fair coin n times. X - number of heads.

What's the prob. of $X \geq \frac{3}{4}n$?

$X \sim Bin(n, \frac{1}{2})$. $\mathbb{E}[X] =$ $\qquad$ $Var(X) =$

Theorem (Chebyshev's Inequality)

Let X be a r.v. with expectation μ

Then $$\Pr[\,|X-\mu| \geq c\,] \leq \frac{Var(X)}{c^2}$$

for any constant $c \geq 0$.

Example: Coin Tosses

You toss a fair coin n times. X - number of heads.

What's the prob. of $X \geq \frac{3}{4}n$?

$X \sim Bin(n, \frac{1}{2})$. $\mathbb{E}[X] = n/2$ $\quad Var(X) = n \cdot p \cdot (1-p) = \frac{n}{4}$

$$\Pr[\,X \geq \tfrac{3}{4}n\,] \leq \Pr\left[|X - \tfrac{n}{2}| \geq \tfrac{1}{4}n\right] \leq \frac{Var(X)}{(\frac{1}{4}n)^2} = \frac{n/4}{(n/4)^2} = \frac{4}{n}$$

goes to 0 as $n \to \infty$.

Back to Motivating Example

You have a coin w. unknown heads prob. p.

Want to estimate p.

<u>Method:</u> toss the coin n times

$$X = \text{number of heads}$$

$$\hat{p} = \frac{X}{n} \qquad \text{"emperical mean"}$$

$$\Pr[\,|p-\hat{p}| \geq \varepsilon\,] \leq \;?$$

Back to Motivating Example

You have a coin w. unknown heads prob. p.

Toss the coin n times.

$X =$ number of heads. $X = X_1 + \dots + X_n$

X_i – indicator that i^{th} coin is heads

$$\hat{p} = \frac{X}{n} = \frac{X_1 + \dots + X_n}{n}$$

$$\mathbb{E}[\hat{p}] =$$

$$Var[\hat{p}] =$$

$$\Pr\left[|\hat{p} - p| \geq \varepsilon \right] \leq \frac{Var[\hat{p}]}{\varepsilon^2} \leq$$

Back to Motivating Example

You have a coin w. unknown heads prob. p.

Toss the coin n times.

$X =$ number of heads. $X = X_1 + \dots + X_n$

X_i – indicator that i^{th} coin is heads

$$\hat{p} = \frac{X}{n} = \frac{X_1 + \dots + X_n}{n}$$

$$\mathbb{E}[\hat{p}] = \mathbb{E}\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{np}{n} = p.$$

$$\mathrm{Var}[\hat{p}] = \frac{\mathrm{Var}(X)}{n^2} = \frac{\mathrm{Var}(X_1) + \dots + \mathrm{Var}(X_n)}{n^2} = \frac{np \cdot (1-p)}{n^2} \leq \frac{1/4}{n}$$

$$\Pr\left[|\hat{p} - p| \geqslant \varepsilon\right] \leqslant \frac{\mathrm{Var}[\hat{p}]}{\varepsilon^2} = \frac{1}{4n\varepsilon^2}$$

Back to Motivating Example

You have a coin w. unknown heads prob. p.

Toss the coin n times.

X = number of heads. $X = X_1 + \dots + X_n$

X_i – indicator that ith coin is heads

$\hat{p} = \frac{X}{n} = \frac{X_1 + \dots + X_n}{n}$ $\quad \mathbb{E}[\hat{p}] = p \quad \mathrm{Var}[\hat{p}] = \frac{p(1-p)}{n} \leq \frac{1}{4n}$

$$\Pr[|\hat{p} - p| \geq \varepsilon] \leq \frac{\mathrm{Var}[\hat{p}]}{\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

Theorem: If $n \geq \frac{1}{4\varepsilon^2\delta}$ then

$$\Pr[|\hat{p} - p| \geq \varepsilon] \leq \delta.$$

Weak Law of Large Numbers

If $X_1, X_2, X_3, \ldots$ are independent identically distributed (i.i.d.)

with $\mathbb{E} X_i = \mu$, $\operatorname{Var}(X_i) < \infty$, then for every $\varepsilon > 0$,

$$\Pr\left[\left| \frac{X_1 + \cdots + X_n}{n} - \mu \right| \geq \varepsilon \right] \xrightarrow[n \to \infty]{} 0.$$

Weak Law of Large Numbers

If $X_1, X_2, X_3, \ldots$ are independent identically distributed (i.i.d.)

with $\mathbb{E} X_i = \mu$, $\mathrm{Var}(X_i) < \infty$, then for every $\varepsilon > 0$,

$$\Pr\left[\left| \frac{X_1 + \ldots + X_n}{n} - \mu \right| \geq \varepsilon \right] \xrightarrow[n \to \infty]{} 0.$$

Proof: Let $\sigma^2 = \mathrm{Var}(X_i)$

$S_n = X_1 + \ldots + X_n$. $\quad Y_n = \frac{S_n}{n}$

$\mathbb{E}[Y_n] = \boxed{}$

$\mathrm{Var}[Y_n] = \boxed{}$

$\Pr[\, |Y_n - \mu| \geq \varepsilon \,] \leq$

Q: Is Chebyshev's Inequality best possible?

Q: Is Chebyshev's Inequality best possible?

A: Yes, if we only know μ and σ.

<u>**For example:**</u> For any μ, σ and any $k \geq 1$

$$X = \begin{cases} \mu + \sigma k & \text{w.p. } \frac{1}{2k^2} \\ \mu - \sigma k & \text{w.p. } \frac{1}{2k^2} \\ \mu & \text{w.p. } 1 - \frac{1}{k^2} \end{cases}$$

$\mathbb{E}[X] =$

$\mathrm{Var}[X] =$

$\Pr[|X - \mathbb{E}X| \geq \sigma k] =$

Theorem: If $n \geq \frac{1}{4\varepsilon^2\delta}$

$$\Pr[|\hat{p} - p| \geq \varepsilon] \leq \delta.$$

Q: Is $\frac{1}{4\varepsilon^2\delta}$ best possible?

Theorem: If $n \geq \frac{1}{4\varepsilon^2\delta}$

$$\Pr\left[\,|\hat{p}-p| \geq \varepsilon\,\right] \leq \delta.$$

Q: Is $\frac{1}{4\varepsilon^2\delta}$ best possible?

A: No $n = \frac{1}{2\varepsilon^2}\ln\left(\frac{2}{\delta}\right)$ suffices

using Chernoff bounds – out of scope.

Theorem: If $n \geq \frac{1}{4\varepsilon^2\delta}$

$$\Pr[|\hat{p} - p| \geq \varepsilon] \leq \delta.$$

Q: Is $\frac{1}{4\varepsilon^2\delta}$ best possible?

A: No. $n = \frac{1}{2\varepsilon^2}\ln\left(\frac{2}{\delta}\right)$ suffices

using Chernoff bounds - out of scope.

$$\boxed{\begin{array}{l}\text{Confidence is "cheap"} \\ \text{Accuracy is "expensive."}\end{array}}$$

Confidence Interval

You don't know the prob. p of heads.

You tossed it a 1,000,000 times and get H 35% of the times.

How confident are you that $p = 0.35$?

" " " " that $p \in [0.3, 0.4]$?

Confidence Interval

You don't know the prob. p of heads.

You tossed it a 1,000,000 times and get H 35% of the times.

How confident are you that $p = 0.35$?

" " " " that $p \in [0.3, 0.4]$?

Def'n: An interval $[a,b]$ is a 95%-confidence interval for an unknown quantity p if

$$\Pr[\ p \in [a,b]\ ] \geq 0.95$$

The interval $[a,b]$ is calculated based on an experiment.

Confidence Interval

You don't know the prob. p of heads.

Theorem: $\hat{p}$ = frac. of heads in n tosses

$$\Pr\left[\,|\hat{p}-p| \leq \varepsilon\,\right] \geq 1 - \frac{1}{4n\varepsilon^2}$$

We can pick $n = \frac{1}{4\varepsilon^2 \cdot 0.05} = \frac{5}{\varepsilon^2}$ to get confidence 95%.

What's the confidence interval?

Confidence Interval

You don't know the prob. p of heads.

Theorem: $\hat{p}$ = frac. of heads in n tosses

$$\Pr\left[\,|\hat{p}-p| \le \varepsilon\,\right] \geq 1 - \frac{1}{4n\varepsilon^2}$$

We can pick $n = \frac{1}{4\varepsilon^2 \cdot 0.05} = \frac{5}{\varepsilon^2}$ to get confidence 95%.

What's the confidence interval?

$a = \hat{p} - \varepsilon \qquad b = \hat{p} + \varepsilon$

$$\Pr\left[\, p \in [a,b] \,\right] = \Pr\left[\, |p - \hat{p}| \le \varepsilon \,\right] \geq 0.95.$$