Concentration

CS 70, Summer 2019

Lecture 23, 8/1/19

Plan:

- Variance measures **deviation from mean**
- How do we **quantify** this? Can we get good **bounds** even if we don't get an exact answer? $\mathbb{P}[X \geq \mathbb{E}[X] + c]$??
- Tools that use **mean** and **variance**:
  - Markov's Inequality
  - Chebyshev's Inequality
  - Law of Large Numbers

Trees. (The Non-Graph Kind.)

I plant 10 trees in my yard.
After a few months, I measure their growth. I tell you that the average height of my trees is 3 ft.

Is it possible that I have 4 trees of 10 ft or taller?

NO.

avg height = 3 $\Rightarrow$ sum of heights = 30

If 4 trees height ≥ 10 $\Rightarrow$ sum ≥ 40. $\not\Leftarrow$

Q: what is max # of trees ≥ 10 ft?

$$3 = \frac{30 \leftarrow \text{sum}}{10 \leftarrow \#}$$

Trees. (The Non-Graph Kind.)

I sample any one of my 10 trees.
Let H be the height of my tree.

What can I say about $\mathbb{P}[H \geq 10]$?

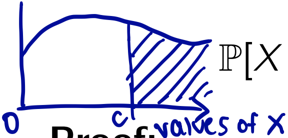
$$\mathbb{P}[H \geq 10] \leq \frac{\text{max poss \# of } \geq 10\text{ft trees}}{\text{\# all trees.}}$$

$$\leq \frac{\frac{\text{sum heights}}{10}}{\text{\# all trees.}}$$

$$\leq \frac{\frac{\text{sum heights}}{\text{\# all trees}}}{10} = \frac{\text{avg height}}{10}$$

Markov Inequality

X is a RV that only takes **non-negative** values.



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

Proof:

$$\mathbb{E}[X] = \sum_{a \in A} a \cdot \mathbb{P}[X=a]$$

non-negativity allows this $\rightarrow$

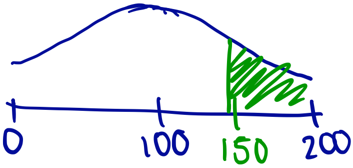
$$\geq \sum_{a \geq c} a \cdot \mathbb{P}[X=a]$$

$$\geq \sum_{a \geq c} c \cdot \mathbb{P}[X=a]$$

$$= c \sum_{a \geq c} \mathbb{P}[X=a] = c \cdot \mathbb{P}[X \geq c]$$

Example: Coin Tosses I

I flip 200 fair coins.
What is an **upper bound** on the probability that we get more than 150 heads?



$H = \#$ heads

$H \sim \text{Bin}(200, \frac{1}{2})$

H non-neg ✓

$\Rightarrow$ use Markov.

$$\mathbb{P}[H \geq 150] \leq \frac{\mathbb{E}[H]}{150}$$

$$= \frac{100}{150} = \frac{2}{3}$$

Example: Generalized Markov

Let Y be a **arbitrary** RV. Let $c > 0$, $r > 0$.
Want to show:

$$\mathbb{P}[|Y| \geq c] \leq \frac{\mathbb{E}[|Y|^r]}{c^r}$$

What should we apply Markov to?

Can't apply to Y.
can apply to $|Y|^r$

$$\Downarrow$$

$$\mathbb{P}[|Y|^r \geq c^r] \leq \frac{\mathbb{E}[|Y|^r]}{c^r}$$

$$\mathbb{P}[|Y| \geq c] \text{ if } r > 0.$$

Markov: A Tight Example?

Is it possible that $\mathbb{P}[X \geq c] = \frac{\mathbb{E}[X]}{c}$?

Strategy: Go through proof of Markov.
For every **inequality** (i.e. $\geq$), determine what would give **equality.**

$$\mathbb{E}[X] = \sum_{a \in A} a \cdot \mathbb{P}[X=a]$$

X can't take values $< c$, except 0

$$= \sum_{a \geq c} a \cdot \mathbb{P}[X=a]$$

X can't take values $> c$.

$$= \sum_{a \geq c} c \cdot \mathbb{P}[X=a]$$

$$= c\,\mathbb{P}[X \geq c]$$

X must be

$$X = \begin{cases} 0 & \text{wp } 1-p. \\ c & \text{wp } p \end{cases}$$

Exercise: Plug X into the Markov, check for equality

Markov: A Tight Example?

(Continued...)

no need for space

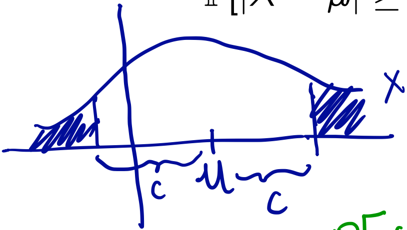
Chebyshev Inequality

Let X be an **arbitrary** RV. $\mu = \mathbb{E}[X]$

We have the following bound for two-sided tails:

dist. from X to μ

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



Apply Markov to $(X-\mu)^2$, c^2

$$\mathbb{P}[(X-\mu)^2 \geq c^2] \leq \frac{\mathbb{E}[(X-\mu)^2]}{c^2}$$

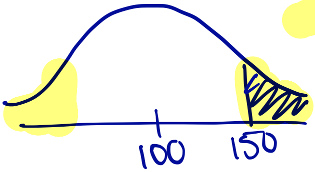
$$\Updownarrow$$

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

Example: Coin Tosses II

I flip 200 fair coins. $H \sim Bin(200, \frac{1}{2})$

What is an **upper bound** on the probability that we get more than 150 heads?



what chebyshev gives

$\mathbb{E}[H]$ ↓

$$\mathbb{P}[H \geq 150] = \mathbb{P}[H - 100 \geq 50]$$

$$\leq \mathbb{P}[|H-100| \geq 50]$$

$$\leq \frac{Var(H)}{50^2}$$

$$\leq \frac{2\cancel{00}(\frac{1}{2})(\frac{1}{2})}{25\cancel{00}} = \frac{1}{50}$$

(*) Observe: H's distribution is symmetric about $\mathbb{E}[X]$

can go 1 step more:

$$\mathbb{P}[H \geq 150] = \frac{1}{2}\mathbb{P}[|H-100| \geq 50]$$

$$\leq \frac{1}{2}\left(\frac{1}{50}\right) = \frac{1}{100}$$

$\Uparrow$

(*) will not work if distribution is <u>NOT</u> symm.

Example: Lower Bound on Variance

Let X be a RV such that $\mathbb{E}[X] = 1$, and

$$\mathbb{P}[-2 < X < 3] = \frac{1}{2}$$

Can I get a lower bound on $\mathrm{Var}(X)$?

Chebyshev

$$\frac{\mathrm{var}(X)}{c^2} \geq \mathbb{P}[|X-1| \geq c]$$

① $X - 1 \geq c$

$x \geq c+1$

② $-(X-1) \geq c$

$X \leq 1-c$

$1+c \leq X$, or $X \leq 1-c$

complement!

$= 1 - \mathbb{P}[1-c < X < 1+c]$

Example: Lower Bound on Variance

(Continued...)

$$1 - \mathbb{P}[1-c < X < 1+c]$$

Know: $\mathbb{P}[-2 < X < 3] = \frac{1}{2}$

choose c carefully:

$c = 2$

$$1 - \mathbb{P}[-1 < X < 3] \geq 1 - \mathbb{P}[-2 < X < 3]$$

$$\geq \frac{1}{2}$$

$$\frac{\mathrm{var}(X)}{4} \geq \frac{1}{2} \Rightarrow \mathrm{var}(X) \geq 2$$

Break

Whats the best Wi-Fi name youve ever seen?

Example: Bias of a Coin

I have a coin with **unknown** head probability p.
I want to **estimate** p within some **error tolerance** ϵ, and I want to be **confident** in my estimate with some probability $1 - \delta$.

estimate

$$\mathbb{P}\left[|\hat{p} - p| < \varepsilon\right] > 1 - \delta$$

To do this, I flip my coin n times, measure the number of heads, and divide by n.

is an RV

Let $\hat{p}$ be my estimate.

Let X_i = indicator for i^{th} coin Heads

$$\hat{p} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Example: Bias of a Coin

What is $\mathbb{E}[\hat{p}]$? What is $\text{Var}[\hat{p}]$?

$$\hat{p} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

$$\mathbb{E}[\hat{p}] = \mathbb{E}\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right]$$

$$= \frac{1}{\cancel{n}}\left(\cancel{n}\,\mathbb{E}[X_1]\right)$$

$$= \mathbb{E}[X_1] = p$$

"unbiased estimator"

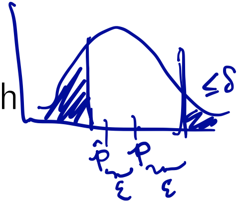
$$\text{var}(\hat{p}) = \text{var}\left(\frac{X_1 + \ldots + X_n}{n}\right)$$

$$= \frac{1}{n^2}\left[\text{var}(X_1) + \ldots + \text{var}(X_n)\right]$$

$$= \frac{1}{n^{\cancel{2}}} \cdot \cancel{n} \cdot \underbrace{\text{var}(X_1)}_{\text{call this } \sigma^2}$$

$$= \frac{\sigma^2}{n}$$

Example: Bias of a Coin



I want $\hat{p}$ to be within ϵ (**error**) of p with probability (**confidence**) $1-\delta$.

$$\mathbb{P}[|\hat{p}-p|<\varepsilon] > 1-\delta$$

↳ complement

$$\mathbb{P}[|\hat{p}-p|\geq\varepsilon]\leq\delta$$

using Chebyshev: $\mathbb{E}[\hat{p}]$

$$\mathbb{P}[|\hat{p}-p|\geq\varepsilon]\leq\frac{\mathrm{Var}(\hat{p})}{\varepsilon^2}=\frac{\sigma^2}{n\varepsilon^2}$$

want: $\dfrac{\sigma^2}{n\varepsilon^2}\leq\delta \implies n\geq\dfrac{\sigma^2}{\varepsilon^2\delta}$

Chebyshev: A Tight Example?

Is it possible that $\mathbb{P}[|X - \mathbb{E}[X]| \geq c] = \frac{\mathrm{Var}(X)}{c^2}$?

Strategy: Go through proof of Chebyshev.
Where do we use inequality (i.e. $\leq$ or $\geq$) instead of equality?

Exercise.

Chebyshev: A Tight Example?

(Continued...)

The Law of Large Numbers

Intuition: If we observe a RV X **many times**, and **average** the observations, the average **converges** to $\mathbb{E}[X]$.

Formally: Let $X_1, X_2, \ldots$ be a sequence of i.i.d. RVs with expectation μ (where μ is finite).

Let $S_n = X_1 + X_2 + \ldots + X_n$. Then:

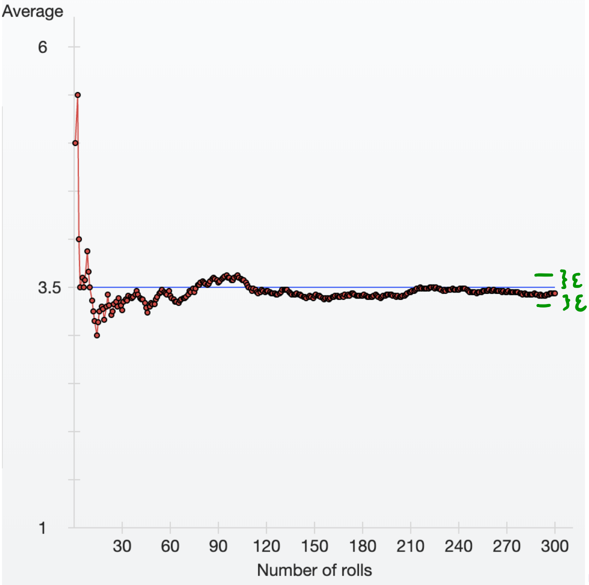
partial sum.

$$\mathbb{P}[|\frac{1}{n} \cdot S_n - \mu| < \epsilon] \to 1 \text{ as } n \to \infty$$

partial avg.

Visualization: Dice Rolls

$\frac{1}{n} \cdot S_n$



Example: Coin Game

A fair coin is tossed.
1) You win if there are more than 60% heads.
Which is better, 10 or 100 tosses?

For larger n

$$\mathbb{P}\left[\%\,\text{heads} \geq 60\%\right] \to 0$$

2) You win if there are more than 40% heads.
Which is better, 10 or 100 tosses?

$$\mathbb{P}\left[\%\,\text{heads} \geq 40\%\right] \to 1$$

Example: Coin Game

A fair coin is tossed.

1) You win if there are between than 40% and 60% heads. Which is better, 10 or 100 tosses?

$\mathbb{P}[40\% \le \%\text{ heads} \le 60\%] \to 1$ as $n \to \infty$

2) You win if there are **exactly** 50% heads. Which is better, 10 or 100 tosses?

LLN deals w/ an interval

10 coins: 40% - 60% ⇒ 4 - 6

100 coins: 40% - 60% 40 - 60

Exactly 50% Heads

Exercise. Post Later.

Compare:

$$\mathbb{P}[n \text{ heads out of } 2n \text{ tosses}]$$

and

$$\mathbb{P}[(n+1) \text{ heads out of } (2n+2) \text{ tosses}]$$

Exactly 50% Heads

(Continued...)

Summary

- Markov helps find **one-sided** tail probability for **non-negative** RVs, given the **mean**.

(*) Not always symmetric!

- Chebyshev helps find **two-sided** tails for **any** RV, given the mean **and variance**.
- LLN tells us that if we observe a RV many times, the **probability that we are "close" to the mean** nears 1.