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. $\not\Rightarrow$

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 $\mathbb{P}[H \geq 10]$?

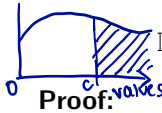
$\mathbb{P}[H \geq 10] \leq \frac{\text{max poss \# of } \geq 10\text{ft trees}}{\text{\# all trees.}}$

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

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

Markov Inequality

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



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

Proof:

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

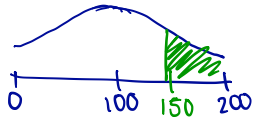
non-negativity allows this $\rightarrow \geq \sum_{a \geq c} a \cdot \mathbb{P}[X = a]$

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

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

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 $|Y|^r$

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

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

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...)

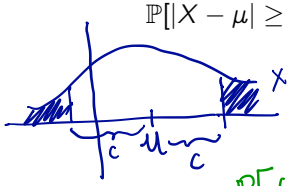
no need for space

Chebyshev Inequality

Let X be an **arbitrary** RV. $\mu = \mathbb{E}[X]$
We have the following bound for two-sided tails:

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

(dist. from X to μ)



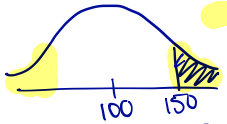
Apply Markov to $(X-\mu)^2, c^2$

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

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

Example: Coin Tosses II

I flip 200 fair coins. $H \sim \text{Bin}(200, \frac{1}{2})$
What is an **upper bound** on the probability that we get more than 150 heads?



what chebyshev gives

$$\mathbb{P}[H \geq 150] = \mathbb{P}[H - 100 \geq 50]$$
$$\leq \mathbb{P}[|H - 100| \geq 50]$$
$$\leq \frac{\text{Var}(H)}{50^2}$$
$$\leq \frac{200(\frac{1}{2})(\frac{1}{2})}{2500} = \frac{1}{50}$$

observe: H's distribution is symmetric about $\mathbb{E}[X]$
can go 1 step more:

$$\mathbb{P}[H \geq 150] = \frac{1}{2}\mathbb{P}[|H-100| \geq 50]$$
$$\leq \frac{1}{2}\left(\frac{1}{50}\right) = \frac{1}{100}$$

(*) will not work if distribution is NOT 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 $\text{Var}(X)$? Chebyshev

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

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: $c = 2$

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

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

$\frac{\text{Var}(X)}{4} \geq \frac{1}{2} \Rightarrow \text{Var}(X) \geq 2$

Break

Whats the best Wi-Fi name you've 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 $1 - \delta$.

↳ $\mathbb{P}[|\hat{p} - p| < \epsilon] > 1 - \delta$ ($\hat{p}$: estimate)

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

Let $\hat{p}$ be my estimate. ($\hat{p}$ is an RV)

Let X_i = indicator for ith 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 $\text{Var}[\hat{p}]$?

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

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

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

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

"unbiased estimator"

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

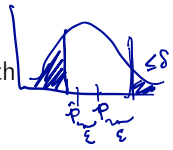
$= \frac{1}{n^2}[\text{Var}(X_1) + \ldots + \text{Var}X_n]$

$= \frac{1}{n^2} \cdot n \cdot \text{Var}(X_1)$ (call this σ^2)

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

Example: Bias of a Coin

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



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

↳ complement

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

using Chebyshev: $\mathbb{P}[|\hat{p} - p| \geq \epsilon] \leq \frac{\text{Var}(\hat{p})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$ ($\mathbb{E}[\hat{p}]$)

want: $\frac{\sigma^2}{n\epsilon^2} \leq \delta \Rightarrow n \geq \frac{\sigma^2}{\epsilon^2 \delta}$

Chebyshev: A Tight Example?

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

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).

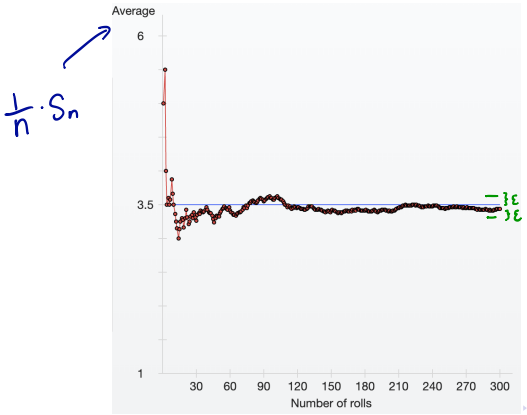
Let $S_n = X_1 + X_2 + \ldots + X_n$. Then:

partial sum.

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

partial avg.

Visualization: Dice Rolls



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

$\pm\epsilon$

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

$\mathbb{P}[\%_{\text{heads}} \geq 60\%] \to 0$

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

$\mathbb{P}[\%_{\text{heads}} \geq 40\%] \to 1$

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?

$\mathbb{P}[40\% \leq \%_{\text{heads}} \leq 60\%] \to 1$ 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% $\Rightarrow$ 4 - 6

100 coins: 40% - 60% 40 - 60

Exactly 50% Heads

Exercise.

Compare: Post Later.

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

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**.
- 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.