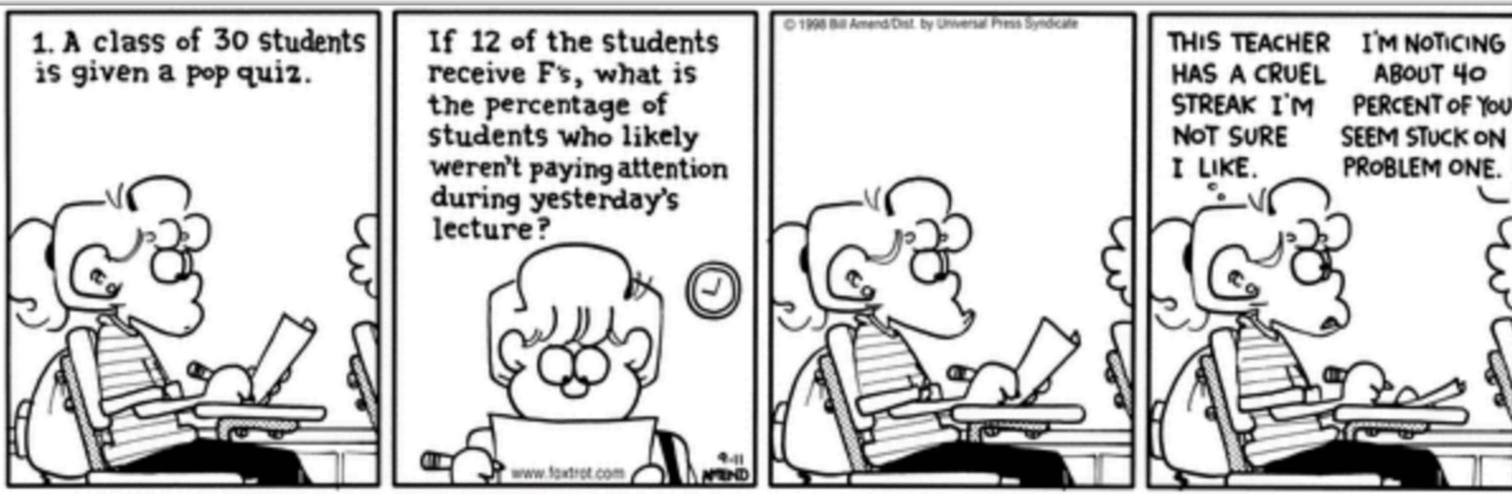
Stat 88: Probability & Mathematical Statistics in Data Science



Lecture 21: 3/10/2021

Sections 6.3, 6.4

Markov and Chebyshev's Inequalities

Recap:

- Variance:

$$Var(X) = E(D^2) = E[(X - E(X))^2]$$

- Standard deviation:

$$SD(X) = \sqrt{Var(X)} = \sqrt{E(D^2)} = \sqrt{E[(X - E(X))^2]}$$

- Shortcut formula for variance:

$$E(X^2) - [E(X)]^2$$

Computational formula.

- Properties of SD (*a*, *b*, *c* are constants):

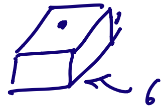
$$SD(aX + b) = |\, a \,|\, SD(X)$$

$$Var(c) = 0 = SD(c)$$

If a is > 0
|a| = a

If a ≤ 0
|a| = -a

Exercise 6.5.5



$|-2| = 2$
$= -(-2)$

- Let $p \in (0,1)$ and let X be the number of spots showing on a flattened die that shows its six faces according to the following chances:
  - $P(X = 1) = P(X = 6)$
  - $P(X = 2) = P(X = 3) = P(X = 4) = P(X = 5) = \frac{1-p}{4}$
  - $P(X = 1 \text{ or } 6) = p \Rightarrow P(X=1) = P(X=6) = p/2 \leftarrow$

Find $SD(X)$ and explain why it is an increasing function of p

$P(X=1) + P(X=6) = p$

$$\mathbb{E}(X) = \sum_x x P(X=x) = \underbrace{\frac{p}{2}(1) + \frac{p}{2}(6)}_{} + \frac{1-p}{4}(2+3+4+5)$$

$$= \frac{7p}{2} + \frac{14}{4}(1-p) = \frac{14p}{4} + \frac{14-14p}{4} = \frac{14}{4} = \boxed{\frac{7}{2}}$$

$$\boxed{Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2}$$

$$\mathbb{E}(X^2) = (1^2+6^2)\frac{p}{2} + (2^2+3^2+4^2+5^2)\frac{1-p}{4}$$

$4+9+16+25$

$$= \frac{37}{2}p + \frac{54}{4}(1-p) = \frac{37p}{2} + \frac{27-27p}{2} = \frac{27+10p}{2} = \frac{27}{2} + 5p$$

$$Var(X) = \frac{27+10p}{2} - \frac{49}{4} = \frac{54}{4} + 5p - \frac{49}{4} = \frac{5}{4} + 5p, \quad SD(X) = \sqrt{\frac{5}{4} + 5p} \leftarrow \text{increasing}$$

as p increases

What can the average tell us?

- We want to say something about the accuracy of estimates. If we don't know much about a list beyond its mean, what can we say about the values of the random variable?

- True or false? Half the data are always above the average. F

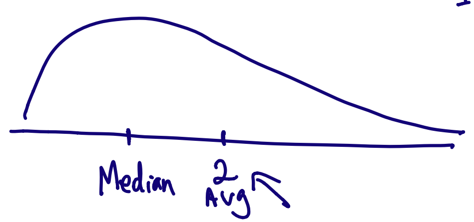
- Suppose we have 100 *non-negative* numbers and the average of the list is 2. True or false? At most 25 of the numbers could be greater than or equal to 8. Why?

If avg is 2, $n=100$

Total = 200 $\left(\frac{\text{Total}}{n} = \text{avg}\right)$

$25 \times 8 = 200$

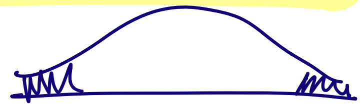
If 25 of the list values are ≥ 8, the total is already 200, no more #s can be >0



- The average gives us an upper bound on proportion. That is, it puts a firm wall that prevents too many numbers from being big compared to the average.

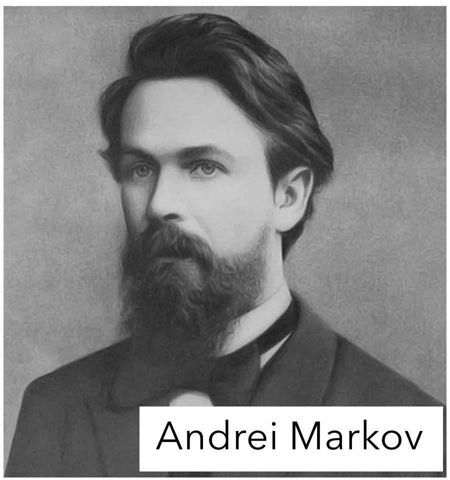
What can the average tell us?

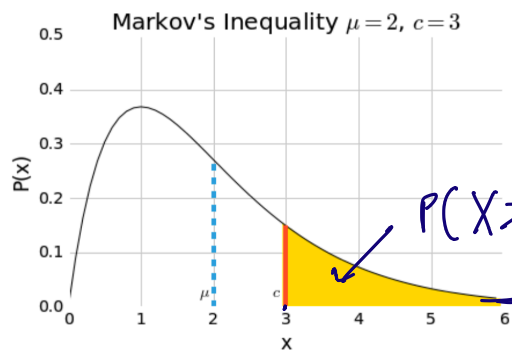
- We will begin with a simple bound, that works for non-negative random variables when the only information we have is the mean of the distribution.



Tail probabilities

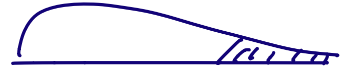
- We are going to consider tail probabilities, or probabilities of the type $P(X \geq c)$, for some $c > 0$





$P(X \geq 3)$ given $\mathbb{E}(X) = 2$.

https://mathshistory.st-andrews.ac.uk/Biographies/Markov/



Markov's inequality: Bounding tail probabilities

- "Bounding tail probabilities" means that we put bounds on what fraction of points can fall far away from the mean.

$X \geq 0$

- Let X be a non-negative random variable (all possible values taken by X are at least 0). Fix $c > 0$. We want to find an upper bound for $P(X \geq c)$in terms of $E(X)$.

- Start with the definition of $E(X) = \sum_{x \geq 0} xP(X = x)$

Split into those $x < c$, $x \geq c$

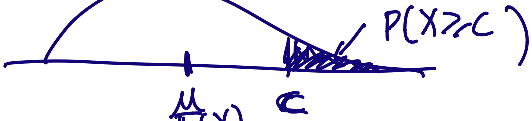
$$E(X) = \cancel{\sum_{x<c} x \cdot P(X=x)} + \sum_{x \geq c} x \cdot P(X=x)$$

$$E(X) \geq \sum_{x \geq c} x \cdot P(X=x) \geq \sum_{x \geq c} c \cdot P(X=x) = c \sum_{x \geq c} P(X=x)$$

$$E(X) \geq c \cdot P(X \geq c)$$

$$\boxed{1+2+3+4+5 = (1+2+3)+(4+5) \quad \text{If } a \geq \cancel{(1+2+3)}+(4+5) \quad a \geq 4+5}$$

$$\boxed{P(X \geq c) \leq \frac{E(X)}{c}}$$



Examples

$$P(X \geq c) \leq \frac{E(X)}{c}$$

1. The mean weight of students in a certain class of students is 140 lbs. What is the largest possible fraction that could weigh over 210 lbs.?

$E(X) = \mu = 140$

$$P(X \geq 210) \leq \frac{140}{210} = \frac{14}{21} = \frac{2}{3}$$

Say $E(X) = 220$

$$P(X \geq 210) \leq \frac{220}{210} > 1$$

$$P(X \geq 210) \leq 1$$

2. A student has a GPA (grade point average) of 2.8. In each course she takes, she gets a grade between 0 (failing) and 4.0 (A+). What is the largest decimal fraction of her grades that could be 4 or higher?

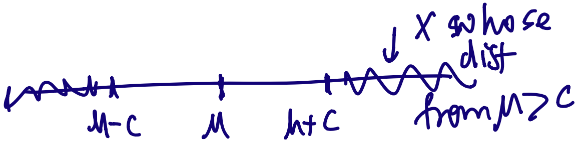
$E(X) = 2.8, \quad c = 4$

$$P(X = 4) \leq \frac{E(X)}{c} = \frac{2.8}{4}$$

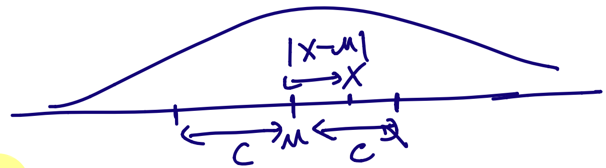
c

Chebyshev's inequality

- What about if we have more information? Say we also know $SD(X)$.
- The first improvement is that we don't need X to be non-negative.
- Let $E(X) = \mu, SD(X) = \sigma$, where X is *any* random variable. Fix $c > 0$.
- We are now interested in the chance of being in either tail, so the chance of the random variable being extreme in either direction. We sum the two tail probabilities.



$$P(\underbrace{| X - \mu |}_{\text{distance b/w } X \,\&\, \mu} \geq c)$$



- Now, we don't know much about $E(| X - \mu |)$ but know about the squared deviation, whose expectation is the variance of X. So we consider that random variable instead (D^2):

$P(|X-\mu|)$

$c > 0$ $\quad P(|X - \mu|^2 \geq c^2) = P\left((X - \mu)^2 \geq c^2\right)$ What can Markov's inequality tell us about this?

Let $Y = (X-\mu)^2$, then $Y \geq 0$

$E(Y) = Var(X) = E((X-\mu)^2)$

$|X-\mu|^2 = (X-\mu)^2$

Markov's ineq $\quad P(Y \geq k) \leq \dfrac{E(Y)}{k}$

$k > 0$

$$P(|X-\mu| \geq c) = P\left((X-\mu)^2 \geq c^2\right) \leq \frac{E\left((X-\mu)^2\right)}{c^2} = \frac{Var(X)}{c^2} = \frac{\sigma^2}{c^2}$$

Chebyshev's inequality

- For a random variable X, with mean μ and standard deviation σ, for any positive constant $c > 0$, we have:

$$\boxed{P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2} = \frac{Var(X)}{c^2}}$$

- Suppose a random variable X has $\mu = 60$, and $\sigma = 5$. what is the chance that it is outside the interval $(50, 70)$?

- What about $P\left(X \in (50, 70)\right)$?

Chebyshev's inequality interpreted as distances

- Say that $E(X)$ is the origin, and we are measuring distances in terms of $SD(X)$.
- We want to know the chance that the rv X is at least $k\ SD$'s away from its mean:

$$P(|X-\mu| \geq k\cdot\sigma) \leq \frac{\cancel{\sigma^2}}{(k\cancel{\sigma})^2} = \frac{1}{k^2}$$

$k > 0$

k not nec. integer.

By Chebyshev's ineq $P(|X-\mu| \geq c) \leq \frac{var(X)}{c^2}$ $\rightarrow$ $c = k\sigma,\ var(X) = \sigma^2$

$P((X-\mu)^2 \geq c^2)$

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

$$P(|X-\mu| < k\sigma) \geq 1 - 1/k^2$$

- What if we are only interested in one tail? A certain type of light bulb has an average lifetime of 10,000 hours. The SD of bulb lifetimes is 550 hours. What decimal fraction of bulbs could last more than 11,980 hours?

$k\cdot 550$

p

10000 11980

these X are within $k\sigma$ of μ

X farther then $k\cdot\sigma$ distance from μ.

$\mu - k\sigma$ μ $\mu + k\sigma$