Probability and Mathematical Statistics in Data Science

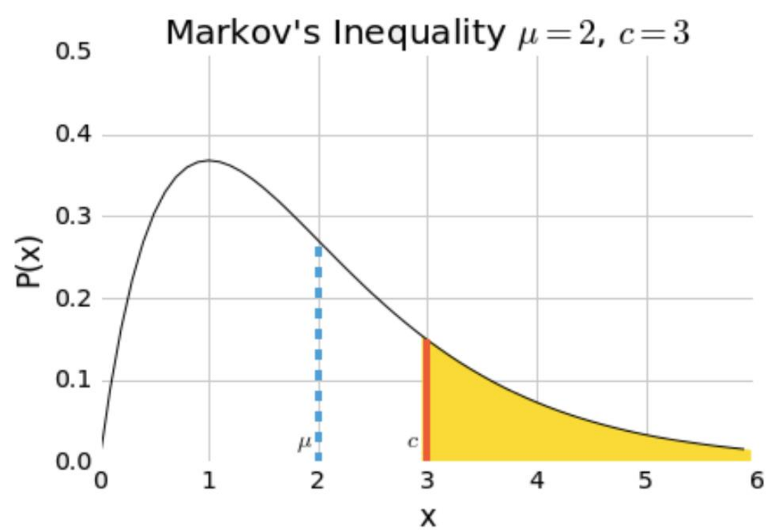
Lecture 17: Section 6.3: Markov's Inequality

Section 6.4: Chebyshev's Inequality

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.

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



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.

- 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 $E(X)$.

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

Exmples

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.?

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?

Chebyshev's Inequality

- What about if we have more information? Say we also know $SD(X)$.

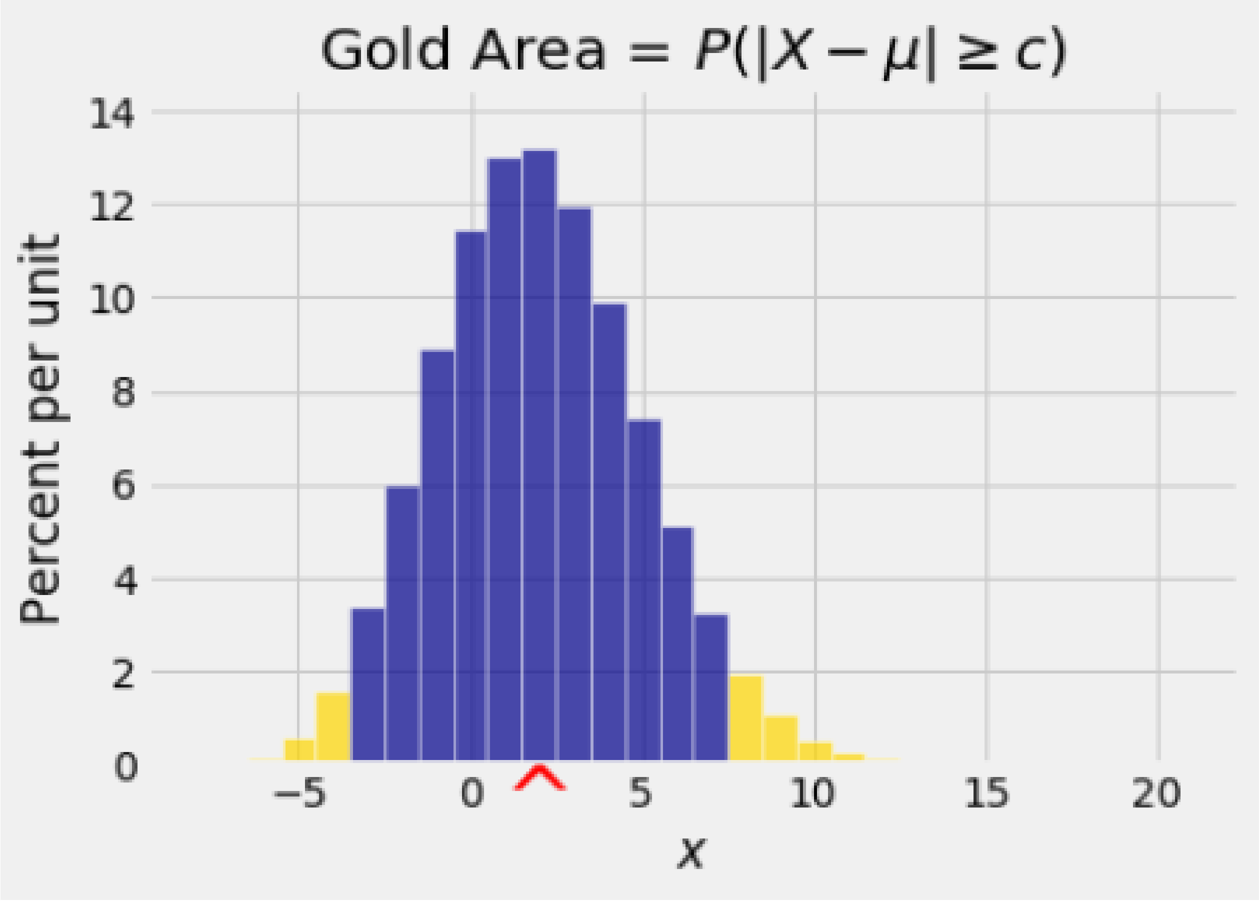
- The first improvement is that we don't need X to be non-negative.

- 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.

$$P(|X - \mu| \geq c)$$

Chebyshev's Inequality



Chebyshev's Inequality

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

$$
\begin{aligned}
P(|X-\mu| \geq c) &= P((X-\mu)^2 \geq c^2) \\
&\leq \frac{E((X-\mu)^2)}{c^2} \\
&= \frac{\sigma^2}{c^2}
\end{aligned}
$$

Chebyshev's Inequality

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

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

- 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:

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

Chebyshev's inequality interpreted as distances

- 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?

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

Chebyshev or Markov? $P(X \geq c) \leq \frac{E(X)}{c}$

- Suppose X is a non-negative random variable with expectation 60 and SD 5.

(a) What can we say about $P(X \geq 70)$?

(b) What is the chance that X is outside the interval $(50, 70)$?

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

Exercise 6.5.6

Ages in a population have a mean of 40 years. Let X be the age of a person picked at random from the population.

a) If possible, find $P(X \geq 80)$. If it's not possible, explain why, and find the best upper bound you can based on the information given.

b) Suppose you are told in addition that the SD of the ages is 15 years. What can you say about $P(10 < X < 70)$?

c) With the information as in Part b, what can you say about $P(10 \leq X \leq 70)$?