STAT 88: Lecture 21

Contents

Warm up: A study on college students found that the men had an average weight of about 66 kg and an SD of about 9 kg. The women had an average weight of about 55 kg and SD of 9 kg. If you took the men and women together, would the SD of their weights be:

(a) smaller than 9 kg.

(b) just about 9 kg.

(c) bigger than 9 kg.

(d) you need more information.

Last time

SD(X) is the average deviation of X from the mean $E(X)$.

$$\mathrm{SD}(X) = \sqrt{E((X-\mu_X)^2)} \text{ where } \mu_X = E(X),$$

or

$$\mathrm{SD}(X) = \sqrt{E(X^2) - \mu_X^2}.$$

$$\mathrm{Var}(X) = (\mathrm{SD}(X))^2.$$

You should be able to tell which of two distributions has a larger SD.

<u>Ex:</u> (Exercise 6.5.4)

4. Let X have distribution

x	1	2	3	4
$P(X=x)$	0.4	0.1	0.1	0.4

Let Y have distribution

y	1	2	3	4
$P(Y=Y)$	0.1	0.4	0.4	0.1

Which of these distributions has a larger SD?

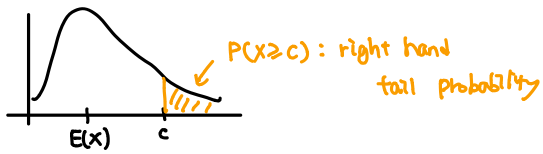
6.3. Markov's inequality

We study what we can say about how far a non-negative random variable can be from its mean, using only the mean and not the SD.

all the possible values of $X \geq 0$

Tail Probabilities Let X be a a non-negative random variables. Fix $c > 0$. We want to find $P(X \geq c)$ in terms of $E(X)$.

Intuition :



We know

$$
\begin{aligned}
E(X) &= \sum_{\text{all } x \geq 0} xP(X = x) \\
&= \sum_{\text{all } x < c} xP(X = x) + \sum_{\text{all } x \geq c} xP(X = x).
\end{aligned}
$$

Then

$$
\begin{aligned}
E(X) &\geq \sum_{\text{all } x \geq c} xP(X = x) \\
&\geq \sum_{\text{all } x \geq c} cP(X = x) \\
&= c \sum_{\text{all } x \geq c} P(X = x) \\
&= cP(X \geq c).
\end{aligned}
$$

Therefore we obtain **Markov's inequality**: for a non-negative random variable X and a positive constant $c > 0$,

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

Markov's inequality is a *tail bound*.

→ Tail probability is bounded by expectation (= 1st moment)

Ex: Give an upper bound for the probability that a Stat 88 student takes 4 or more math classes ($E(X) = 1.1$).

Ex: Let X be a non-negative RV and let k be any positive constant. Find an upper bound for $P(X \geq kE(X))$.

What does Markov say if $k = 0.5$?

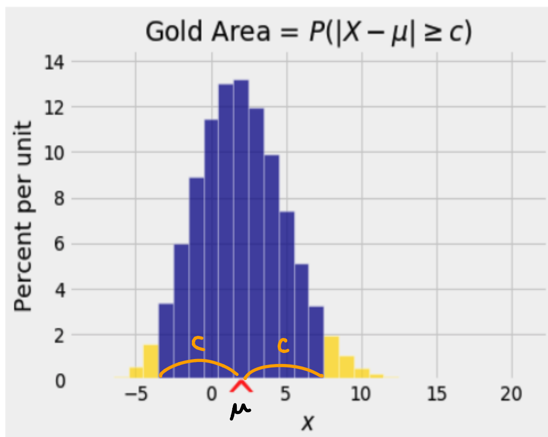
Ex: Let $X \sim$ Binomial$(100, 1/2)$. What is an upper bound for $P(X \geq 4E(X))$? What is $P(X \geq 4E(X))$ exactly?

6.4. Chebyshev's inequality

Can we get a better upper bound for the chance a stat 88 Student takes 4 or more math classes knowing the average is 1.1 classes **AND** the SD is 1.5 classes?

This is answered by Chebyshev's inequality.

Let $\mu = E(X)$ and $\sigma = \text{SD}(X)$. Let X be any random variable (possibly negative) and fix $c > 0$. We are interested in the chance of being in both tails, $P(|X - \mu| \geq c)$.



We have

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

By Markov's inequality since $(X-\mu)^2 \geq 0$

$$\leq \frac{E((X-\mu)^2)}{c^2}$$

$$= \frac{\sigma^2}{c^2}.$$

This proves **Chebyshev's Inequality**: for a random variable X with mean μ and SD σ and a positive constant $c > 0$,

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

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

What is $P(X \in (50, 70))$?

Chebyshev's inequality revisited

Chebyshev inequality can give an upper bound for the chance your data is $k > 0$ or more SD away from the mean, e.g. $k = 2$.

Let X be a random variable with mean μ and SD σ. Then for all $k > 0$,

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

The most important point about Chebyshev's inequality is that it makes no assumption about the shape of the distribution. **No matter what the shape of the distribution of X:**

- the chance that X is at least 2 SDs away from its mean is at most?
- the chance that X is at least 3 SDs away from its mean is at most?
- the chance that X is at least 4 SDs away from its mean is at most?
- the chance that X is at least 5 SDs away from its mean is at most?

This holds for ANY DISTRIBUTION.

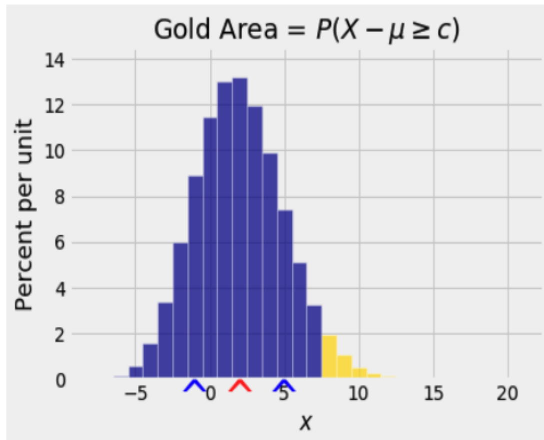
Example: (Exercise 6.5.6) Ages in a population have a mean of 40 years. Let X be the age of a person picked at random from the population.

(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.

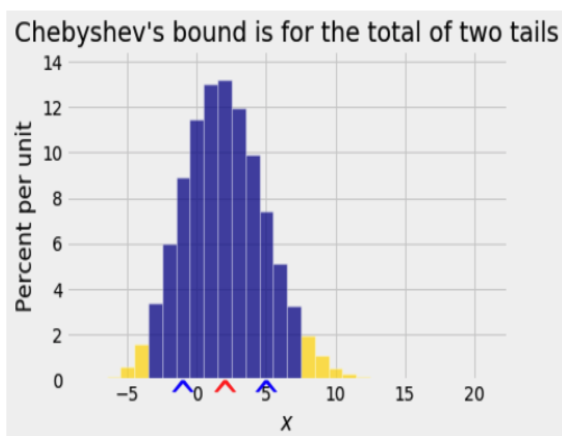
(b) Suppose you are told in addition that the SD of the ages is 15 years. What can you say about $P(10 < X < 70)$?

Bound on One Tail

Suppose we want an upper bound on just one tail, as in the figure below. The right hand tail probability is $P(X - \mu \geq c)$.



Chebyshev's inequality gives an upper bound on the total of two tails starting at equal distances on either side of the mean: $P(|X - \mu| \geq c)$.



You cant just use half of Chebyshev upper bound. Note each single tail is no bigger than the total of two tails.

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

Ex: What is chance that a Stat 88 student takes 2 or more math classes given $\mu = 1.1$ and $\sigma = 1.5$?