STA 130A: Mathematical Statistics: Brief Course

Lecture 18: Markov and Chebyshev Inequalities

Dogyoon Song

Fall 2026, UC Davis

Announcements

Homework 5 is due tomorrow (Thu, Nov 5, 11:59 PM PT)

Extended office hours: Wednesdays, 1–3 PM (MSB 1143)

- The 1–2 PM hour is shared with STA 35C, whose students may receive priority

Agenda

Last time: Moment generating functions

- Compute moments and identify distributions from MGFs.
- For independent variables, multiply MGFs to study their sum.

Today: Tail probability bounds

- Markov's inequality
- Chebyshev's inequality
- Using Chebyshev's inequality to control sample averages

Recap: Moment generating functions

One function describing a distribution:

$$M_X(t) = \mathbb{E}[e^{tX}], \qquad M_X(0) = 1.$$

Recover moments by differentiating at 0:

$$\mathbb{E}[X] = M_X'(0), \qquad \mathrm{Var}(X) = M_X''(0) - \big(M_X'(0)\big)^2.$$

Transform and combine random variables:

$$M_{aX+b}(t) = e^{bt} M_X(at),$$
$$M_{X+Y}(t) = M_X(t) M_Y(t) \qquad \text{when } X, Y \text{ are independent.}$$

Identify distributions: Match the MGF to a known form.

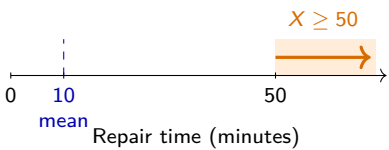
Today: What can we guarantee about probabilities when we know only a mean or variance, not the full distribution?

Motivation: How often can a repair take too long?

A repair shop reserves **50 minutes per job**. Suppose the repair time X satisfies

$$X \geq 0, \qquad \mathbb{E}[X] = 10 \text{ minutes},$$

but its distribution is unknown.



Question: How often does a job take 50 minutes or longer?

$$p = P(X \geq 50).$$

Goal: An upper bound on p, not its exact value.

Key idea: Jobs lasting at least 50 minutes contribute at least $50p$ to the mean; all other durations are nonnegative.

$$10 = \mathbb{E}[X] \geq 50p \quad \Longrightarrow \quad p \leq 0.2.$$

Guarantee: At least 80% of jobs finish in under 50 minutes, whatever the distribution satisfying these assumptions.

Our goal: Distribution-free tail bounds

We often want to know how likely a random variable is to be large or far from its mean:

$$P(X \geq a), \qquad P(|X - \mathbb{E}[X]| \geq a).$$

The full distribution of X determines these probabilities.

Q: When the distribution is unknown or difficult to work with, can we bound these?

Markov and Chebyshev inequalities use **moments** to give **distribution-free** bounds.

Distribution-free does not mean assumption-free.

- The bounds hold for **every distribution satisfying the stated conditions**, without assuming a particular family such as normal or Poisson.

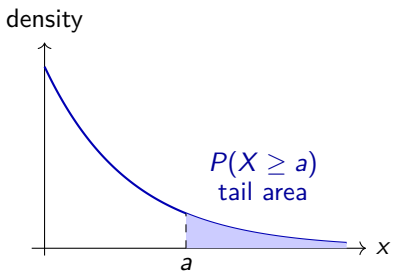
Tradeoff: A guaranteed upper bound, possibly loose—not an exact probability.

Markov's inequality

Theorem (Markov's inequality)

If a random variable X is nonnegative ($X \geq 0$), then

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}, \qquad \text{for all } a > 0.$$



Interpretation:

- If X is nonnegative and has small mean, then X cannot be very large with high probability.
- Markov's inequality uses only nonnegativity and the mean; it does not use the variance or distributional shape.

Proof of Markov's inequality

Assume $X \geq 0$, and fix $a > 0$.

Observe that

$$X \geq a\,\mathbf{1}_{\{X \geq a\}} = \begin{cases} a, & \text{if } X \geq a, \\ 0, & \text{if } X < a. \end{cases}$$

Taking expectations,

$$\mathbb{E}[X] \geq \mathbb{E}\Big[a\,\mathbf{1}_{\{X \geq a\}}\Big] = a\,\mathbb{E}\Big[\mathbf{1}_{\{X \geq a\}}\Big] = a\,P(X \geq a).$$

Therefore,

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Key intuition: if X exceeds a, then it contributes at least a to the average.

Example: Markov's inequality

Example

Suppose $X \geq 0$ (e.g., waiting time in minutes) and

$$\mathbb{E}[X] = 10.$$

Question: What can we say about $P(X \geq 50)$?

By Markov's inequality,

$$P(X \geq 50) \leq \frac{\mathbb{E}[X]}{50} = \frac{10}{50} = 0.2.$$

Important: This bound may be loose, but universally applicable for nonnegative X.

- It uses only nonnegativity and the mean; it ignores variance and distributional shape.
- The same bound holds for every nonnegative random variable with mean 10.

Pop-up quiz: Markov's inequality

Suppose $X \geq 0$ and $\mathbb{E}[X] = 6$.

Question: Which statement is guaranteed by Markov's inequality?

A) $P(X \geq 12) \leq 1/4$

B) $P(X \geq 12) \leq 1/2$

C) $P(X \geq 12) = 1/2$

D) $P(X \leq 12) \leq 1/2$

Answer: B. Markov's inequality gives

$$P(X \geq 12) \leq \frac{\mathbb{E}[X]}{12} = \frac{6}{12} = \frac{1}{2}.$$

Follow-up: Does Markov's inequality hold without assuming $X \geq 0$? Which step of the proof would fail?

Chebyshev's inequality

Markov's inequality can be applied to any nonnegative random variable.

In particular, if $g(X) \geq 0$ and $a > 0$, then

$$P(g(X) \geq a) \leq \frac{\mathbb{E}[g(X)]}{a}.$$

Theorem (Chebyshev's inequality)

Let X have mean $\mu = \mathbb{E}[X]$ and variance $\sigma^2 = \mathrm{Var}(X)$. Then, for any $c > 0$,

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

If $\sigma > 0$, the standardized form is, for $k > 0$,

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

Interpretation: Most of the probability must lie in a few standard deviations of the mean.

Proof of Chebyshev's inequality

Let $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}(X)$.

Apply Markov's inequality to the nonnegative random variable

$$(X - \mu)^2.$$

Then, for $c > 0$,

$$\begin{aligned} P(|X - \mu| \geq c) &= P\left((X - \mu)^2 \geq c^2\right) \\ &\leq \frac{\mathbb{E}[(X - \mu)^2]}{c^2} \\ &= \frac{\mathrm{Var}(X)}{c^2} = \frac{\sigma^2}{c^2}. \end{aligned}$$

Key idea: Large deviation from the mean implies large squared deviation.

Example: Chebyshev's inequality

Example

Suppose

$$\mathbb{E}[X] = 100, \qquad \mathrm{SD}(X) = 10.$$

Question: What can we say about

$$P(|X - 100| \geq 30)?$$

Using Chebyshev's inequality,

$$P(|X - 100| \geq 30) \leq \frac{10^2}{30^2} = \frac{1}{9}.$$

Equivalently,

$$P(|X - 100| < 30) \geq 1 - \frac{1}{9} = \frac{8}{9}.$$

Important: No distributional shape is assumed beyond the stated mean and variance.

Comparison of Markov vs. Chebyshev

Markov's inequality

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a} \qquad (X \geq 0,\ a > 0).$$

Chebyshev's inequality

$$P(|X - \mu| \geq c) \leq \frac{\mathrm{Var}(X)}{c^2} \qquad (c > 0).$$

Comparison:

- Markov controls a one-sided tail for a nonnegative random variable.
- Chebyshev controls deviation from the mean.
- Chebyshev uses more information: variance as well as mean.
- Both are distribution-free: they hold for all distributions satisfying the stated moment assumptions, but they can be loose.

Example: Compare bounds for the same event

Example

Return to $\mathbb{E}[X] = 100$, $\mathrm{SD}(X) = 10$, and now additionally assume $X \geq 0$.

Question: Bound $P(X \geq 130)$.

$$\text{Markov:} \qquad P(X \geq 130) \leq \frac{100}{130} = \frac{10}{13}.$$

Since $\{X \geq 130\} \subseteq \{|X - 100| \geq 30\}$,

$$\text{Chebyshev:} \qquad P(X \geq 130) \leq \frac{10^2}{30^2} = \frac{1}{9}.$$

Takeaway: Compare bounds for the same event and check each set of assumptions. The smaller bound is stronger (in this example, Chebyshev is stronger).

Pop-up quiz: Chebyshev's inequality

Suppose $\mathbb{E}[X] = 20$ and $\mathrm{Var}(X) = 16$.

Question: Which option gives the smallest valid upper bound on $P(|X - 20| \geq 8)$ from Chebyshev's inequality?

A) $P(|X - 20| \geq 8) \leq 1/4$

B) $P(|X - 20| \geq 8) \leq 1/2$

C) $P(|X - 20| \geq 8) \leq 2$

D) $P(X \geq 8) \leq 1/4$

Answer: A. Chebyshev's inequality gives

$$P(|X - 20| \geq 8) \leq \frac{\mathrm{Var}(X)}{8^2} = \frac{16}{64} = \frac{1}{4}.$$

Follow-up: What bound do we get for $P(|X - 20| \geq 12)$, and why is it smaller than the bound for threshold 8?

Application: Sample averages

Let $X_1, \ldots, X_n$ be independent and identically distributed with

$$\mathbb{E}[X_i] = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2.$$

Define the sample average

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

This is itself a random variable with

$$\mathbb{E}[\bar{X}_n] = \mu,$$

and, by independence,

$$\mathrm{Var}(\bar{X}_n) = \mathrm{Var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{\sigma^2}{n}.$$

Message: averaging i.i.d. random variables reduces variance by a factor of n.

Chebyshev bound for sample averages

For i.i.d. observations with mean μ and variance σ^2, Chebyshev's inequality yields

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\mathrm{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}, \qquad \epsilon > 0.$$

To guarantee $P(|\bar{X}_n - \mu| \geq \epsilon) \leq \delta$, with $0 < \delta < 1$, it is sufficient that

$$n \geq \frac{\sigma^2}{\delta\epsilon^2}.$$

In particular, as n increases, $\frac{\sigma^2}{n\epsilon^2} \to 0$, and hence,

$$\bar{X}_n \approx \mu \quad \text{with high probability for large } n.$$

This is the key idea behind the **weak law of large numbers** (Lecture 20).

Example: How many samples are enough?

Example

Suppose $X_1, \ldots, X_n$ are i.i.d. with

$$\mathbb{E}[X_i] = \mu, \qquad \mathrm{Var}(X_i) = 4.$$

Question: How large should the sample size n be to guarantee

$$P(|\bar{X}_n - \mu| \geq 0.5) \leq 0.01$$

using Chebyshev's inequality?

Chebyshev's inequality gives

$$P(|\bar{X}_n - \mu| \geq 0.5) \leq \frac{4}{n(0.5)^2} = \frac{16}{n}.$$

It is sufficient to require

$$\frac{16}{n} \leq 0.01 \qquad \Longrightarrow \qquad n \geq 1600.$$

Remark: Chebyshev can be conservative, but it works without knowing the distribution.

Wrap-up

Markov's inequality

$$X \geq 0 \quad \Longrightarrow \quad P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

- Uses only nonnegativity and the mean.

Chebyshev's inequality

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

- Uses mean and variance, but no distributional shape.

Sample averages

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}.$$

- This is the bridge to the weak law of large numbers.

Suggested reading: [BT08, Sec. 5.1]

Next lecture: Review for Midterm 2.

References


Dimitri Bertsekas and John N Tsitsiklis.
Introduction to probability, volume 1.
Athena Scientific, 2nd edition, 2008.