STA 131A: Introduction to Probability Theory

Lecture 23: Markov and Chebyshev Inequalities

Dogyoon Song

Spring 2026, UC Davis

Announcements

Homework 7 is posted

- Start early and post questions on Piazza as needed

Office hours today

- 2:30–3:30 PM at MSB 4220

Agenda

Last time:

- MGFs for fixed sums of independent random variables:

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

- Random sums: sum of a random number of random variables

$$S = \sum_{i=1}^{N} X_i$$

Today: Tail probability bounds

- Markov's inequality
- Chebyshev's inequality
- Using Chebyshev to control sample averages
- Preview: weak law of large numbers

Recap: Random sums

Let

$$S = \sum_{i=1}^{N} X_i,$$

- $X_1, X_2, \dots$ are i.i.d. random variables with
$$\mathbb{E}[X_i] = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2,$$
- $N \in \{0, 1, 2, \dots\}$ is independent of the X_i's.

Then

$$\begin{aligned} \mathbb{E}[S] &= \mathbb{E}[N]\,\mu, \\ \mathrm{Var}(S) &= \mathbb{E}[N]\,\sigma^2 + \mu^2 \mathrm{Var}(N), \\ M_S(t) &= M_N\big(\log M_X(t)\big). \end{aligned}$$

Idea: condition on the random count N, then S becomes an ordinary fixed-length sum.

Recap example: Compound Poisson total cost

Example

Let N be the random number of claims, and let $X_1, X_2, \ldots$ be the claim sizes. Assume all variables are independent, with

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

Let

$$S = \sum_{i=1}^{N} X_i$$

be the total cost. Using the random-sum formulas,

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

This is a compound model: a random number of random-sized contributions.

Interpretation of variance: total cost varies because claim sizes are random and because the number of claims is random.

Recap example: Poisson thinning

Example

Let $N \sim \mathrm{Poisson}(\lambda)$. Given N, suppose each of the N events is marked as a success independently with probability p, independently of all other events. Let S denote the number of successes, i.e.,

$$S = \sum_{i=1}^{N} X_i, \qquad X_i \sim \mathrm{Bernoulli}(p).$$

We have

$$M_N(t) = \exp\left[\lambda(e^t - 1)\right], \qquad M_X(t) = 1 - p + pe^t.$$

Therefore,

$$\begin{aligned} M_S(t) &= M_N\big(\log M_X(t)\big) \\ &= \exp\big[\lambda(M_X(t) - 1)\big] \\ &= \exp\big[\lambda p(e^t - 1)\big]. \end{aligned}$$

This is the MGF of $\mathrm{Poisson}(\lambda p)$. Hence, $S \sim \mathrm{Poisson}(\lambda p)$.

Interpretation: randomly keeping each event with probability p thins the Poisson rate $\lambda \to \lambda p$.

Motivation: Bounding tail probabilities

Often, we want to know how likely it is that a random variable is large:

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

If we know the full distribution of X, we can compute such probabilities exactly.

But sometimes we know only summaries:

$$\mathbb{E}[X], \qquad \mathrm{Var}(X).$$

Question: Can we still say something useful about tail probabilities?

Answer: Yes. Markov and Chebyshev inequalities give distribution-free bounds.

Why tail bounds matter

Tail bounds quantify how often large deviations or outliers can occur

- Typically these tail probabilities tend to be small.
- Tail probabilities often become small for averages of many independent random variables.

However, exact probabilities can be difficult to compute:

- The distribution may be unknown.
- The CDF or density may be complicated.
- We may only have partial information, such as a mean or variance.

A probability bound may be crude, but it is often robust:

$$\text{few assumptions} \quad \Longrightarrow \quad \text{broadly applicable conclusion.}$$

Today's key message:

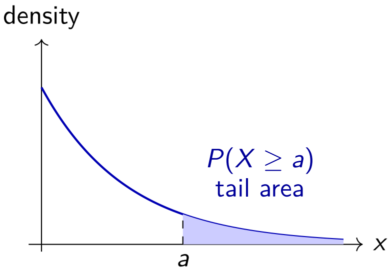
moments can control tail probabilities.

Markov's inequality

Theorem (Markov's inequality)

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

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



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

$$X \geq a\,\mathbf{1}_{\{X \geq a\}}.$$

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 $X \geq 0$ and

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

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.

- 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 gives

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

Follow-up: What would the bound be for $P(X \geq 30)$? Why does the bound get smaller as the threshold increases?

Chebyshev's inequality

Markov's inequality can be applied to any nonnegative random variable. In particular, if $g(X) \geq 0$, 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 $\sigma^2 = \mathrm{Var}(X) < \infty$. Then, for any $c > 0$,

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

Equivalent standardized form: for $k > 0$,

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

Interpretation: Most of the probability mass must lie within a few standard deviations of

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

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

What can we say about

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

Using Chebyshev,

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

Equivalently,

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

Important: This conclusion does not require X to be normal.

Comparison of Markov vs. Chebyshev

Markov's inequality

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

Chebyshev's inequality

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

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.

Pop-up quiz: Chebyshev's inequality

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

Question: Which bound follows 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 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 reduces variance by a factor of n.

Chebyshev bound for sample averages

By Chebyshev's inequality,

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

As n increases,

$$\frac{\sigma^2}{n\epsilon^2} \to 0.$$

This suggests:

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

This is the key idea behind the weak law of large numbers.

Example: How many samples are enough?

Example

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

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

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, Ch. 5.1]

References

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