Chapter 10A: Inequalities

MATH/STAT 394: Probability I

Arman Jahangiri

University of Washington Department of Mathematics

Summer 2026

Learning Goals

By the end of this lecture, students should be able to:

- understand why inequalities are important in probability;
- apply the Cauchy–Schwarz inequality;
- use Jensen's inequality with convex and concave functions;
- apply Markov's inequality;
- apply Chebyshev's inequality;
- apply Chernoff bounds;
- interpret inequalities geometrically and probabilistically.

Why Inequalities Matter

What If Exact Calculation Is Impossible?

In many real problems:

- exact probabilities are difficult;
- integrals may be impossible;
- distributions may be unknown;
- computations may be too expensive.

Three major strategies:

1. simulation;
2. inequalities (bounds);
3. asymptotic approximations.

Today's focus

Probability inequalities

Suppose we cannot compute

$$\mathbb{P}(A)$$

exactly.

An inequality may tell us:

$$0.01 \leq \mathbb{P}(A) \leq 0.03.$$

Even without the exact answer, we gain useful information.

Cauchy–Schwarz Inequality

Cauchy–Schwarz Inequality

Theorem

For random variables X, Y with finite second moments,

$$|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$

This is one of the most important inequalities in mathematics.

For every real number t,

$$(Y - tX)^2 \geq 0.$$

Taking expectations:

$$\mathbb{E}(Y - tX)^2 \geq 0.$$

Expanding:

$$\mathbb{E}(Y^2) - 2t\mathbb{E}(XY) + t^2\mathbb{E}(X^2) \geq 0.$$

This quadratic must always be nonnegative.

Proof of Cauchy–Schwarz

Since

$$\mathbb{E}(Y^2) - 2t\mathbb{E}(XY) + t^2\mathbb{E}(X^2) \geq 0$$

for all t, its discriminant must satisfy:

$$(-2\mathbb{E}(XY))^2 - 4\mathbb{E}(X^2)\mathbb{E}(Y^2) \leq 0.$$

Thus,

$$\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2).$$

Taking square roots gives:

$$|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$

Application of Cauchy-Schwarz: $-1 \leq Corr(X, Y) \leq 1$

Apply Cauchy–Schwarz to

$$X - \mathbb{E}(X)$$

and

$$Y - \mathbb{E}(Y).$$

Then:

$$|\operatorname{Cov}(X, Y)| \leq \sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}.$$

Dividing both sides:

$$|\operatorname{Corr}(X, Y)| \leq 1.$$

Application of Cauchy-Schwarz: $Var(X) \geq 0$

Apply Cauchy–Schwarz with

$$Y = 1.$$

Then:

$$|\mathbb{E}(X)|^2 \leq \mathbb{E}(X^2).$$

Therefore,

$$\boxed{\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \geq 0.}$$

Jensen's Inequality

Convex Functions

Definition

A function $g : I \to \mathbb{R}$ is called convex if

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)$$

for all $x, y \in I$ and all

$$0 \leq \lambda \leq 1.$$

Interpretation: $\lambda x + (1-\lambda)y$ is a weighted average of x and y, and convexity says $g(\text{average}) \leq \text{average of } g$.
Geometrically: the graph of a convex function always lies below its secant lines.

Second Derivative Test

If g is twice differentiable, then:

$$g''(x) \geq 0 \quad \text{for all } x$$

implies that g is convex.

Examples of convex functions:

$$g(x) = x^2, \qquad g(x) = e^x.$$

Jensen's Inequality

Theorem

If g is convex, then
$$\mathbb{E}(g(X)) \geq g(\mathbb{E}(X)).$$

If g is concave, the inequality reverses:
$$\mathbb{E}(g(X)) \leq g(\mathbb{E}(X)).$$

For a convex function:

- tangent lines lie below the graph;
- averaging inputs produces smaller values than averaging outputs.

Mnemonic

$$g(\text{average}) \leq \text{average of } g.$$

for convex functions.

Example (Jensen's Inequality): $Var(X) \geq 0$ (again)

Take

$$g(x) = x^2.$$

Since

$$g''(x) = 2 > 0,$$

the function is convex.

Jensen gives:

$$\mathbb{E}(X^2) \geq (\mathbb{E}(X))^2.$$

Thus:

$$\text{Var}(X) \geq 0.$$

Example (Jensen's Inequality): Logarithms

Take

$$g(x) = \log x.$$

Since

$$g''(x) = -\frac{1}{x^2} < 0,$$

the logarithm is concave.

Thus:

$$\mathbb{E}(\log X) \leq \log(\mathbb{E}(X)).$$

This inequality is fundamental in statistics and information theory.

Example (Jensen's Inequality): Sample Standard Deviation Is Biased

Recall:

$$S_n^2$$

is unbiased for

$$\sigma^2.$$

But:

$$S_n = \sqrt{S_n^2}.$$

Since

$$g(x) = \sqrt{x}$$

is concave, Jensen gives:

$$\mathbb{E}(S_n) = \mathbb{E}(\sqrt{S_n^2}) \leq \sqrt{\mathbb{E}(S_n^2)} = \sigma.$$

Conclusion

Sample standard deviation tends to underestimate the true standard deviation.

Tail Bounds

Tail Probabilities

We often want bounds on:

$$\mathbb{P}(X \geq a),$$

or

$$\mathbb{P}(|X - \mu| \geq a).$$

These are called tail probabilities.

Applications

- rare events;
- risk analysis;
- statistics;
- machine learning;
- concentration inequalities.

Markov's Inequality

Theorem (Markov's Inequality)

If $X \geq 0$ and $a > 0$, then

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

Proof

First, note that $X \geq a\, I(X \geq a)$.
Taking expectations from both sides,

$$\mathbb{E}(X) \geq a\mathbb{E}(I(X \geq a)).$$

Using the fundamental bridge:

$$\mathbb{E}(I(X \geq a)) = \mathbb{P}(X \geq a).$$

Therefore:

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

Theorem (Chebyshev's Inequality)

If X has mean μ and variance σ^2, then for every $a > 0$,

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

Proof

Apply Markov to $(X - \mu)^2 \geq 0$. Then,

$$\mathbb{P}((X - \mu)^2 \geq a^2) \leq \frac{\mathbb{E}(X - \mu)^2}{a^2} = \frac{\sigma^2}{a^2}.$$

Standardized Form of Chebyshev

Let

$$a = c\sigma.$$

Then:

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

Examples:

$$\mathbb{P}(|X - \mu| \geq 2\sigma) \leq \frac{1}{4},$$

$$\mathbb{P}(|X - \mu| \geq 3\sigma) \leq \frac{1}{9}.$$

Markov becomes stronger if we apply it to:

$$e^{tX}, \qquad t > 0.$$

Chernoff Bound

For every

$$t > 0,$$

$$\boxed{\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(e^{tX})}{e^{ta}}.}$$

The numerator is the MGF of X.

Why Chernoff Is Powerful

Advantages:

- often much sharper than Chebyshev;
- uses MGFs;
- can optimize over

$$t.$$

Chernoff bounds are central in:

- machine learning;
- randomized algorithms;
- statistical learning theory;
- concentration inequalities.

Comparing Bounds for Normal Tails

Suppose

$$Z \sim N(0,1).$$

True probability:

$$\mathbb{P}(|Z| > 3) \approx 0.003.$$

Bounds:

$$\text{Markov: } 0.27,$$

$$\text{Chebyshev: } 0.11,$$

$$\text{Chernoff: } 0.022.$$

Observation

Chernoff bounds are often dramatically better.

Summary

- Cauchy–Schwarz bounds expectations.
- Jensen compares

$$\mathbb{E}(g(X))$$

  and

$$g(\mathbb{E}(X)).$$

- Markov bounds nonnegative tails.
- Chebyshev controls deviations from the mean.
- Chernoff bounds are often much sharper.
- Inequalities are fundamental tools throughout probability and statistics.

Next lecture

Law of Large Numbers and Central Limit Theorem.