chapter10a_inequalities

Monday, August 17, 2026 12:03 PM

chapter10a_inequalities

$a \le b$

Chapter 10A: Inequalities
MATH/STAT 394: Probability I

Arman Jahangiri

University of Washington Department of Mathematics

Summer 2026

1) $X \le Y$

2) $\mathbb{P}(X \le Y)$

$(f(x) \le g(x))$

$(2 < 3)$

$(\forall x > 2 \quad x < a^x)$

Learning Goals

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

Why Inequalities Matter

What If Exact Calculation Is Impossible?

$\mathbb{P}(X \ge a) = \int_a^\infty f(x)\,dx \le \;?$

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

What Is a Bound?

Suppose we cannot compute

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

exactly.

An inequality may tell us:

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

Even without the exact answer, we gain useful information.

Cauchy–Schwarz Inequality

Cauchy–Schwarz Inequality

$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $\vec{y} = \begin{pmatrix} y_1 \\ \vdots \end{pmatrix}$

$\left|\sum x_i y_i\right| \le \sqrt{\sum x_i^2 \sum y_i^2}$

Theorem

For random variables X, Y with finite second moments,

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

This is one of the most applicable inequalities in mathematics (linear algebra, statistics, functional inequalities, geometry, etc.).

inner product $\langle u, v \rangle \Rightarrow$ $u, v \in V$ $|\langle \vec{u}, \vec{v}\rangle| \le \sqrt{\langle \vec{u}, \vec{u}\rangle \cdot \langle \vec{v}, \vec{v}\rangle}$

induced norm $\|u\| := \sqrt{\langle \vec{u}, \vec{u}\rangle}$ $|\langle \vec{u}, \vec{v}\rangle| \le \sqrt{\|u\|_2^2 \cdot \|v\|_2^2}$

Idea Behind the Proof

For every real number t,

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

Taking expectations:

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

Expanding:

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

This quadratic must always be nonnegative.

$a \cdot t^2 + b \cdot t + c \ge 0$

$a = \mathbb{E}(X^2)$, $b = -2\mathbb{E}(XY)$, $c = \mathbb{E}(Y^2)$

$\Delta \le 0$

$b^2 - 4ac \le 0$

Proof of Cauchy–Schwarz

Since

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

for all t, its discriminant must satisfy:

$$b^2 - 4ac = (-2\mathbb{E}(XY))^2 - 4\mathbb{E}(X^2)\mathbb{E}(Y^2) \le 0.$$

Thus,

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

Taking square roots gives:

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

$X - \mathbb{E}X$ $\quad Y - \mathbb{E}Y$ $\quad X - \mathbb{E}X$

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

$-1 \le Corr(X, Y) = \dfrac{Cov(X, Y)}{\sqrt{Var X} \cdot \sqrt{Var Y}} \le 1$

Apply Cauchy–Schwarz to

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

and

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

Then:

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

Dividing both sides:

$\div \sqrt{Var X\, Var Y}$

$$|Corr(X, Y)| \le 1.$$

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

$|\mathbb{E}XY| \le \sqrt{\mathbb{E}X^2 \mathbb{E}Y^2}$

Apply Cauchy–Schwarz with

$$Y = 1.$$

Then:

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

Therefore,

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

Jensen's Inequality

Convex Functions

Definition

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

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

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

$$0 \le \lambda \le 1.$$

Interpretation: $\lambda x + (1 - \lambda)y$ is a weighted average of x and y, and convexity says $g(\text{average}) \le$ 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:

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

implies that g is convex.

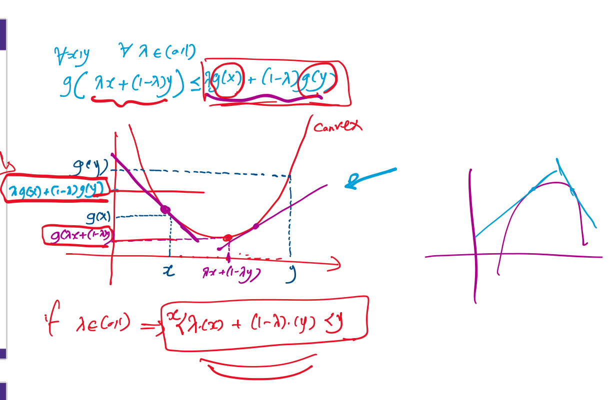
Examples of convex functions:

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

$g''(x) = 2 > 0$ $\quad e^x > 0 \; (\forall x)$

$\forall x < y \quad \forall \lambda \in (0, 1)$

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



$\forall \lambda \in (0, 1) \Rightarrow x \le \lambda(x) + (1-\lambda)(y) \le y$

Jensen's Inequality

Theorem

If g is convex, then

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

If g is concave, the inequality reverses:

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

$Var X = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$ $\quad Var(X) \ge 0$

$g(x) = x^2 \xrightarrow{\text{convex}} g(\mathbb{E}X) \le \mathbb{E}(g(X)) \Rightarrow (\mathbb{E}X)^2 \le \mathbb{E}(X^2)$

Geometric Interpretation of Jensen

For a convex function:
- tangent lines lie below the graph;
- averaging inputs produces smaller values than averaging outputs.

Mnemonic

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

for convex functions.

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

Take

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

Since

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

the function is convex.

Jensen gives:

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

Thus:

$$Var(X) \ge 0.$$

Example (Jensen's Inequality): Logarithms

Take

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

Since

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

the logarithm is concave.

Thus:

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

This inequality is fundamental in statistics and information theory.

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

(skip)

Recall:

$$\mathbb{E}[S_n^2] = \sigma^2,$$

i.e., S^2 is an unbiased estimator for σ^2.
Now, since

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

is concave, Jensen's inequality gives:

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

Conclusion

Sample standard deviation tends to underestimate the true standard deviation.

Tail Bounds

Tail Probabilities

We often want bounds on:

$$\int_a^\infty f(x)\,dx = \mathbb{P}(X \ge a) = \;?$$

or

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

These are called tail probabilities.

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

Markov's Inequality

Theorem (Markov's Inequality)

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

$$\mathbb{P}(X \ge a) \le \frac{\mathbb{E}(X)}{a}.$$

Proof
First, note that $X \ge aI(X \ge a)$.
Taking expectations from both sides,

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

Using the fundamental bridge:

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

Therefore:

$$\mathbb{P}(X \ge a) \le \frac{\mathbb{E}(X)}{a}.$$

$X \ge a \cdot 1$

$a \cdot 1\{X \ge a\} = \begin{cases} a & X \ge a \\ 0 & \text{o.w.} \end{cases} \le X$

$X < a \Rightarrow 0 \le X$

Chebyshev's Inequality

Markov: $P(X \ge a) \le \dfrac{\mathbb{E}X}{a}$

$X := |X - \mu| \ge 0$

Theorem (Chebyshev's Inequality)

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

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

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

$$P(|X - \mu| \ge a) = \mathbb{P}((X - \mu)^2 \ge a^2) \le \frac{\mathbb{E}(X - \mu)^2}{a^2} = \frac{\sigma^2}{a^2}.$$

$P(|X - \mu| \ge a) = P(|X - \mu|^2 \ge a^2) \le \dfrac{\mathbb{E}(X - \mu)^2}{a^2} = \dfrac{var(X)}{a^2} = \dfrac{\sigma^2}{a^2}$

Standardized Form of Chebyshev

Let

$$a = c\sigma.$$

$\sigma^2 / a^2 = \dfrac{\sigma^2}{c^2 \sigma^2} = \dfrac{1}{c^2}$

Then:

$$\mathbb{P}(|X - \mu| \ge c\sigma) \le \frac{1}{c^2}.$$

Examples:

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

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

Chernoff Bounds

Markov $\to$ Chebyshev $\to$ Chernoff

Markov becomes stronger if we apply it to:

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

$P(X \ge a) \le \dfrac{\mathbb{E}X}{a}$

Chernoff Bound

For every

$$t > 0,$$

$$\mathbb{P}(X \ge a) \le \frac{\mathbb{E}(e^{tX})}{e^{ta}}.$$

The numerator is the MGF of X.

$\to X = e^{tX} \ge a$

$P(e^{tX} \ge e^{ta}) \le \dfrac{\mathbb{E}(e^{tX})}{a} = \dfrac{M_X(t)}{a} \quad \forall t > 0$

$\mathbb{P}(X \ge a)$

Comparing Bounds for Normal Tails

Suppose

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

True probability:

$$\mathbb{P}(|Z| > 3) \approx 0.003. \quad < 0.27$$

Bounds:

$(X > 0)$ Markov: 0.27,

Chebyshev: 0.11,

Chernoff: 0.022.

Observation

Chernoff bounds are often dramatically better.

Summary

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

- Cauchy–Schwarz bounds expectations.
- Jensen compares

$$\mathbb{E}(g(X)) \;\; \overset{\ge}{\le} \;\; g(\mathbb{E}(X)).$$

and

concave/convex

- 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.

LLN CLT