0.1 Learning Objectives

By the end of this lecture, you will be able to:

- State and apply Markov's inequality
- Derive Chebyshev's inequality from Markov's inequality
- Use probability bounds to bound tail probabilities without knowing the full distribution
- Compare the tightness of different bounds

1. Probability Bounds

1.1 Motivation: Why Do We Need Bounds?

Sometimes we don't know the exact distribution of a random variable.

1.1 Motivation: Why Do We Need Bounds?

Sometimes we don't know the exact distribution of a random variable.

But we might know its **mean**, **variance**, or other summary statistics.

1.1 Motivation: Why Do We Need Bounds?

Sometimes we don’t know the exact distribution of a random variable.

But we might know its **mean**, **variance**, or other summary statistics.

Question: Can we still say something useful about probabilities?

1.1 Motivation: Why Do We Need Bounds?

Sometimes we don't know the exact distribution of a random variable.

But we might know its **mean**, **variance**, or other summary statistics.

Question: Can we still say something useful about probabilities?

Example: Server Response Times

A web server has mean response time $\mu = 200$ms. A response over 1 second causes a timeout.

Without knowing the exact distribution, can we bound $P(X \geq 1000)$?

1.2 What Is a Probability Bound?

A **probability bound** is an inequality of the form:

$$P(\text{some event}) \leq \text{expression using summary statistics}$$

1.2 What Is a Probability Bound?

A **probability bound** is an inequality of the form:

$$P(\text{some event}) \leq \text{expression using summary statistics}$$

For example, given only $E(X)$, we might show:

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

1.2 What Is a Probability Bound?

A **probability bound** is an inequality of the form:

$$P(\text{some event}) \leq \text{expression using summary statistics}$$

For example, given only $E(X)$, we might show:

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

We can't compute the *exact* probability

1.2 What Is a Probability Bound?

but we get a **guaranteed upper bound.**

1.2 What Is a Probability Bound?

but we get a **guaranteed upper bound**.

Note: The bound holds for *every* distribution satisfying the assumptions. We trade precision for generality: the result is weaker than an exact answer, but applies much more broadly.

1.3 Markov’s Inequality

Definition: Markov's Inequality

If X is a **nonnegative** random variable and $a > 0$, then:

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

1.3 Markov's Inequality

Definition: Markov's Inequality

If X is a **nonnegative** random variable and $a > 0$, then:

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

Requirements:

- $X \geq 0$ (nonnegative)
- $a > 0$
- We only need to know $E(X)$

1.4 Markov's Inequality: Intuition

Why does this make sense?

1.4 Markov's Inequality: Intuition

Why does this make sense?

If the **average** of a nonnegative quantity is small, then it **can't be large too often**.

1.4 Markov's Inequality: Intuition

Why does this make sense?

If the **average** of a nonnegative quantity is small, then it **can't be large too often**.

Example: Income Analogy

If the average income is \$50,000, then at most $1/5$ of people can earn $\geq$ \$250,000.

$$P(X \geq 250000) \leq \frac{50000}{250000} = \frac{1}{5}$$

1.5 Proof of Markov's Inequality

For a nonnegative random variable X and $a > 0$:

1.5 Proof of Markov's Inequality

For a nonnegative random variable X and $a > 0$:

$$
\begin{aligned}
E(X) &= \int_0^\infty x f(x)\,\mathrm{d}x \\
&= \int_0^a x f(x)\,\mathrm{d}x + \int_a^\infty x f(x)\,\mathrm{d}x
\end{aligned}
$$

1.5 Proof of Markov's Inequality

For a nonnegative random variable X and $a > 0$:

$$\begin{aligned} E(X) &= \int_0^\infty x f(x)\,\mathrm{d}x \\ &= \int_0^a x f(x)\,\mathrm{d}x + \int_a^\infty x f(x)\,\mathrm{d}x \end{aligned}$$

Since both integrals are nonnegative:

$$E(X) \geq \int_a^\infty x f(x)\,\mathrm{d}x \geq \int_a^\infty a f(x)\,\mathrm{d}x = a \cdot P(X \geq a)$$

1.5 Proof of Markov's Inequality

For a nonnegative random variable X and $a > 0$:

$$E(X) = \int_0^\infty x f(x)\, \mathrm{d}x$$

$$= \int_0^a x f(x)\, \mathrm{d}x + \int_a^\infty x f(x)\, \mathrm{d}x$$

Since both integrals are nonnegative:

$$E(X) \geq \int_a^\infty x f(x)\, \mathrm{d}x \geq \int_a^\infty a f(x)\, \mathrm{d}x = a \cdot P(X \geq a)$$

Dividing both sides by a:

$$P(X \geq a) \leq \frac{E(X)}{a} \qquad \blacksquare$$

1.6 Try It: Markov's Inequality

Example: Server Response Times

A web server has mean response time $\mu = 200$ms.

Use Markov's inequality to bound $P(X \geq 1000)$.

1.6 Try It: Markov's Inequality

Example: Server Response Times

A web server has mean response time $\mu = 200$ms.

Use Markov's inequality to bound $P(X \geq 1000)$.

Try it yourself

Talk to your neighbor and try to solve this problem.

1.7 Solution: Server Response Times

$$P(X \geq 1000) \leq \frac{E(X)}{1000} = \frac{200}{1000} = \boxed{0.2}$$

1.7 Solution: Server Response Times

$$P(X \geq 1000) \leq \frac{E(X)}{1000} = \frac{200}{1000} = \boxed{0.2}$$

At most 20% of requests take longer than 1 second.

1.7 Solution: Server Response Times

$$P(X \geq 1000) \leq \frac{E(X)}{1000} = \frac{200}{1000} = \boxed{0.2}$$

At most 20% of requests take longer than 1 second.

Note: This bound may be quite loose! If the response times are actually exponential with mean 200ms, the true probability is $e^{-5} \approx 0.0067$. Markov gives us a guarantee even without knowing the distribution.

1.8 How Loose Is Markov's Inequality?

Example: Uniform[0, 4] --- Bertsekas Ex. 7.1

Let $X \sim \mathrm{Uniform}[0, 4]$, so $E(X) = 2$. Compare Markov bounds to exact values:

1.8 How Loose Is Markov's Inequality?

Example: Uniform[0, 4] --- Bertsekas Ex. 7.1

Let $X \sim \mathrm{Uniform}[0, 4]$, so $E(X) = 2$. Compare Markov bounds to exact values:

Event	Markov bound	Exact	Ratio
$P(X \geq 2)$	$\leq 2/2 = 1.00$	0.50	$2\times$
$P(X \geq 3)$	$\leq 2/3 = 0.67$	0.25	$2.7\times$
$P(X \geq 4)$	$\leq 2/4 = 0.50$	0.00	∞

1.8 How Loose Is Markov's Inequality?

Example: Uniform[0, 4] --- Bertsekas Ex. 7.1

Let $X \sim \mathrm{Uniform}[0, 4]$, so $E(X) = 2$. Compare Markov bounds to exact values:

Event	Markov bound	Exact	Ratio
$P(X \geq 2)$	$\leq 2/2 = 1.00$	0.50	$2\times$
$P(X \geq 3)$	$\leq 2/3 = 0.67$	0.25	$2.7\times$
$P(X \geq 4)$	$\leq 2/4 = 0.50$	0.00	∞

The bounds are valid but can be **very** loose

1.8 How Loose Is Markov's Inequality?

especially far in the tail.

1.9 Limitations of Markov's Inequality

Markov's inequality only uses the **mean**

1.9 Limitations of Markov's Inequality

it ignores variance entirely.

1.9 Limitations of Markov's Inequality

it ignores variance entirely.

Low variance: Values cluster near the mean

$\rightarrow$ Tail probabilities should be small

High variance: Values spread out

$\rightarrow$ Tail probabilities could be larger

1.9 Limitations of Markov's Inequality

it ignores variance entirely.

Low variance: Values cluster near the mean

$\to$ Tail probabilities should be small

High variance: Values spread out

$\to$ Tail probabilities could be larger

Can we get a tighter bound by also using the **variance**?

1.9 Limitations of Markov's Inequality

it ignores variance entirely.

Low variance: Values cluster near the mean

$\to$ Tail probabilities should be small

High variance: Values spread out

$\to$ Tail probabilities could be larger

Can we get a tighter bound by also using the **variance**?

Yes! This leads to Chebyshev's inequality.

2. Chebyshev's Inequality

2.1 Chebyshev's Inequality

Definition: Chebyshev's Inequality

For **any** random variable X with mean μ and variance σ^2:

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

Equivalently, in terms of standard deviations ($k = c\sigma$):

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

2.1 Chebyshev's Inequality

Definition: Chebyshev's Inequality

For **any** random variable X with mean μ and variance σ^2:

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

Equivalently, in terms of standard deviations ($k = c\sigma$):

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

This bounds how far X can be from its mean, using **both** mean and variance.

2.2 Deriving Chebyshev from Markov

Chebyshev's inequality is a **corollary** of Markov's!

2.2 Deriving Chebyshev from Markov

Chebyshev's inequality is a **corollary** of Markov's!

Let $Y = (X - \mu)^2$. Then $Y \geq 0$.

2.2 Deriving Chebyshev from Markov

Chebyshev's inequality is a **corollary** of Markov's!

Let $Y = (X - \mu)^2$. Then $Y \geq 0$.

Apply Markov's inequality to Y with threshold k^2:

$$P(Y \geq k^2) \leq \frac{E(Y)}{k^2}$$

2.2 Deriving Chebyshev from Markov

Chebyshev's inequality is a **corollary** of Markov's!

Let $Y = (X - \mu)^2$. Then $Y \geq 0$.

Apply Markov's inequality to Y with threshold k^2:

$$P(Y \geq k^2) \leq \frac{E(Y)}{k^2}$$

But $P(Y \geq k^2) = P((X - \mu)^2 \geq k^2) = P(|X - \mu| \geq k)$

2.2 Deriving Chebyshev from Markov

Chebyshev's inequality is a **corollary** of Markov's!

Let $Y = (X - \mu)^2$. Then $Y \geq 0$.

Apply Markov's inequality to Y with threshold k^2:

$$P(Y \geq k^2) \leq \frac{E(Y)}{k^2}$$

But $P(Y \geq k^2) = P((X - \mu)^2 \geq k^2) = P(|X - \mu| \geq k)$

And $E(Y) = E((X - \mu)^2) = \sigma^2$

2.2 Deriving Chebyshev from Markov

Chebyshev's inequality is a **corollary** of Markov's!

Let $Y = (X - \mu)^2$. Then $Y \geq 0$.

Apply Markov's inequality to Y with threshold k^2:

$$P(Y \geq k^2) \leq \frac{E(Y)}{k^2}$$

But $P(Y \geq k^2) = P((X - \mu)^2 \geq k^2) = P(|X - \mu| \geq k)$

And $E(Y) = E((X - \mu)^2) = \sigma^2$

Therefore: $P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$ $\blacksquare$

2.3 Try It: Chebyshev's Inequality

Example: Server Response Times, Revisited

A web server has mean response time $\mu = 200$ms and standard deviation $\sigma = 100$ ms.

Use Chebyshev's inequality to bound $P(X \geq 1000)$.

Hint: First bound $P(|X - 200| \geq 800)$*.*

2.3 Try It: Chebyshev's Inequality

Example: Server Response Times, Revisited

A web server has mean response time $\mu = 200$ms and standard deviation $\sigma = 100$ ms.

Use Chebyshev's inequality to bound $P(X \geq 1000)$.

Hint: First bound $P(|X - 200| \geq 800)$.

Try it yourself

Talk to your neighbor and try to solve this problem.

2.4 Solution: Chebyshev for Server Times

We want $P(X \geq 1000)$.

2.4 Solution: Chebyshev for Server Times

We want $P(X \geq 1000)$.

Note that $X \geq 1000$ implies $|X - 200| \geq 800$, so:

$$P(X \geq 1000) \leq P(|X - 200| \geq 800)$$

2.4 Solution: Chebyshev for Server Times

We want $P(X \geq 1000)$.

Note that $X \geq 1000$ implies $|X - 200| \geq 800$, so:

$$P(X \geq 1000) \leq P(|X - 200| \geq 800)$$

Applying Chebyshev:

$$P(|X - 200| \geq 800) \leq \frac{\sigma^2}{800^2} = \frac{100^2}{800^2} = \frac{10000}{640000} = \boxed{0.0156}$$

2.4 Solution: Chebyshev for Server Times

We want $P(X \geq 1000)$.

Note that $X \geq 1000$ implies $|X - 200| \geq 800$, so:

$$P(X \geq 1000) \leq P(|X - 200| \geq 800)$$

Applying Chebyshev:

$$P(|X - 200| \geq 800) \leq \frac{\sigma^2}{800^2} = \frac{100^2}{800^2} = \frac{10000}{640000} = \boxed{0.0156}$$

Compare:

- Markov: $P(X \geq 1000) \leq 0.2$
- Chebyshev: $P(X \geq 1000) \leq 0.0156$
- Chebyshev is **much tighter** because it uses the variance!

2.5 How Loose Is Chebyshev? (1/2)

Example: Uniform[0, 4] --- Bertsekas Ex. 7.2

Let $X \sim \mathrm{Uniform}[0, 4]$. We know $\mu = 2$ and $\sigma^2 = 16/12 = 4/3$.

Question: What is $P(|X - 2| \geq 1)$? That is, how likely is X to be more than 1 unit from its mean?

2.5 How Loose Is Chebyshev? (1/2)

Example: Uniform[0, 4] --- Bertsekas Ex. 7.2

Let $X \sim \mathrm{Uniform}[0, 4]$. We know $\mu = 2$ and $\sigma^2 = 16/12 = 4/3$.

Question: What is $P(|X - 2| \geq 1)$? That is, how likely is X to be more than 1 unit from its mean?

Chebyshev gives:

$$P(|X - 2| \geq 1) \leq \frac{4/3}{1^2} = \boxed{4/3 \approx 1.33}$$

This exceeds 1

2.5 How Loose Is Chebyshev? (1/2)

the bound is **trivially true** and tells us nothing!

2.5 How Loose Is Chebyshev? (1/2)

the bound is **trivially true** and tells us nothing!

Exact: $P(|X-2| \geq 1) = P(X \leq 1) + P(X \geq 3) = 1/4 + 1/4 = 1/2$

Chebyshev is uninformative here because the deviation $k = 1$ is small relative to $\sigma \approx 1.15$.

2.6 How Loose Is Chebyshev? (2/2)

Example: Exponential --- Bertsekas Ex. 7.2

Let $X \sim \mathrm{Exp}(\lambda = 1)$, so $\mu = 1$ and $\sigma^2 = 1$.

Question: Bound $P(X \geq c)$ for $c > 1$ using Chebyshev, and compare to the exact tail probability.

2.6 How Loose Is Chebyshev? (2/2)

2.6 How Loose Is Chebyshev? (2/2)

Example: Exponential --- Bertsekas Ex. 7.2

Let $X \sim \mathrm{Exp}(\lambda = 1)$, so $\mu = 1$ and $\sigma^2 = 1$.

Question: Bound $P(X \geq c)$ for $c > 1$ using Chebyshev, and compare to the exact tail probability.

$X \geq c$ implies $|X - 1| \geq c - 1$, so:

$$P(X \geq c) \leq P(|X - 1| \geq c - 1) \leq \frac{1}{(c-1)^2}$$

2.6 How Loose Is Chebyshev? (2/2)

2.6 How Loose Is Chebyshev? (2/2)

Example: Exponential --- Bertsekas Ex. 7.2

Let $X \sim \mathrm{Exp}(\lambda = 1)$, so $\mu = 1$ and $\sigma^2 = 1$.

Question: Bound $P(X \geq c)$ for $c > 1$ using Chebyshev, and compare to the exact tail probability.

$X \geq c$ implies $|X - 1| \geq c - 1$, so:

$$P(X \geq c) \leq P(|X - 1| \geq c - 1) \leq \frac{1}{(c-1)^2}$$

c	Chebyshev	Exact (e^{-c})	Ratio
3	≤ 0.25	0.050	$5\times$
5	≤ 0.0625	0.0067	$9\times$
10	≤ 0.012	0.000045	$270\times$

2.6 How Loose Is Chebyshev? (2/2)

Chebyshev gets **looser** further into the tail

2.6 How Loose Is Chebyshev? (2/2)

it can't capture exponential decay.

2.7 Chebyshev in Standard Deviations

A useful form: what fraction of values can lie more than c standard deviations from the mean?

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

2.7 Chebyshev in Standard Deviations

A useful form: what fraction of values can lie more than c standard deviations from the mean?

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

c (std devs)	Chebyshev bound	Normal (for comparison)
1	≤ 1.00 (trivial)	0.3173
2	≤ 0.25	0.0455
3	≤ 0.111	0.0027
4	≤ 0.0625	0.0001

2.7 Chebyshev in Standard Deviations

A useful form: what fraction of values can lie more than c standard deviations from the mean?

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

c (std devs)	Chebyshev bound	Normal (for comparison)
1	≤ 1.00 (trivial)	0.3173
2	≤ 0.25	0.0455
3	≤ 0.111	0.0027
4	≤ 0.0625	0.0001

Chebyshev is always valid; the Normal values require assuming normality.

2.8 When to Use Which Bound?

Markov's Inequality

- Only need: $E(X)$
- Requires: $X \geq 0$
- Typically loose
- Best for: quick upper bounds when you know very little

Chebyshev's Inequality

- Need: $E(X)$ and $\mathrm{Var}(X)$
- Works for any X
- Tighter than Markov
- Best for: bounding deviations from the mean

2.8 When to Use Which Bound?

Markov's Inequality

- Only need: $E(X)$
- Requires: $X \geq 0$
- Typically loose
- Best for: quick upper bounds when you know very little

Chebyshev's Inequality

- Need: $E(X)$ and $\mathrm{Var}(X)$
- Works for any X
- Tighter than Markov
- Best for: bounding deviations from the mean

Note: Both bounds hold for **any** distribution satisfying their conditions. This generality is powerful but means the bounds can be loose for specific distributions.

2.9 Application: Hash Table Load Balancing

Example: Hashing

We hash n items into n bins uniformly at random.

Let X = number of items in a particular bin.

Then $E(X) = 1$ (by linearity of expectation).

What's the probability that a bin has ≥ 10 items?

2.9 Application: Hash Table Load Balancing

Example: Hashing

We hash n items into n bins uniformly at random.

Let X = number of items in a particular bin.

Then $E(X) = 1$ (by linearity of expectation).

What's the probability that a bin has ≥ 10 items?

By Markov: $P(X \geq 10) \leq \frac{1}{10} = \boxed{0.1}$

2.9 Application: Hash Table Load Balancing

Example: Hashing

We hash n items into n bins uniformly at random.

Let X = number of items in a particular bin.

Then $E(X) = 1$ (by linearity of expectation).

What's the probability that a bin has ≥ 10 items?

By Markov: $P(X \geq 10) \leq \frac{1}{10} = \boxed{0.1}$

Since $\mathrm{Var}(X) = 1 - \frac{1}{n} \approx 1$ for large n:

By Chebyshev: $P(X \geq 10) \leq P(|X - 1| \geq 9) \leq \frac{1}{81} \approx \boxed{0.012}$

2.10 Application: Algorithm Analysis

Example: Randomized QuickSort

The expected number of comparisons in QuickSort on n elements is $E(C) = 2n \ln n$.

Using Markov's inequality, bound the probability that QuickSort uses more than $10n \ln n$ comparisons.

2.10 Application: Algorithm Analysis

Example: Randomized QuickSort

The expected number of comparisons in QuickSort on n elements is $E(C) = 2n \ln n$.

Using Markov's inequality, bound the probability that QuickSort uses more than $10n \ln n$ comparisons.

$$P(C \geq 10n \ln n) \leq \frac{2n \ln n}{10n \ln n} = \boxed{\frac{1}{5}}$$

2.10 Application: Algorithm Analysis

Example: Randomized QuickSort

The expected number of comparisons in QuickSort on n elements is $E(C) = 2n \ln n$.

Using Markov's inequality, bound the probability that QuickSort uses more than $10n \ln n$ comparisons.

$$P(C \geq 10n \ln n) \leq \frac{2n \ln n}{10n \ln n} = \boxed{\frac{1}{5}}$$

At most a 20% chance of using $5 \times$ the expected comparisons.

2.11 Application: Polling and Sample Size

Example: How Many Voters Do We Need? --- Bertsekas Ex. 7.4

Let p = true fraction supporting a candidate. We poll n random voters.

Sample mean M_n estimates p. Since $\mathrm{Var}(X_i) = p(1-p) \leq 1/4$:

$$P(|M_n - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$$

2.11 Application: Polling and Sample Size

Example: How Many Voters Do We Need? --- Bertsekas Ex. 7.4

Let p = true fraction supporting a candidate. We poll n random voters.

Sample mean M_n estimates p. Since $\mathrm{Var}(X_i) = p(1-p) \leq 1/4$:

$$P(|M_n - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$$

With $n = 100$, $\varepsilon = 0.1$: $P(|M_{100} - p| \geq 0.1) \leq \frac{1}{4 \cdot 100 \cdot 0.01} = \boxed{0.25}$

2.11 Application: Polling and Sample Size

Example: How Many Voters Do We Need? --- Bertsekas Ex. 7.4

Let p = true fraction supporting a candidate. We poll n random voters.

Sample mean M_n estimates p. Since $\mathrm{Var}(X_i) = p(1-p) \leq 1/4$:

$$P(|M_n - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$$

With $n = 100$, $\varepsilon = 0.1$: $P(|M_{100} - p| \geq 0.1) \leq \frac{1}{4 \cdot 100 \cdot 0.01} = \boxed{0.25}$

For 95% confidence within 0.01 of p: need $\frac{1}{4n \cdot 0.0001} \leq 0.05$

$$\to n \geq \boxed{50,000}$$

Conservative! (CLT gives tighter results

2.11 Application: Polling and Sample Size

stay tuned.)

2.12 Chebyshev Proves the Weak Law of Large Numbers

Let $X_1, ..., X_n$ be i.i.d. with mean μ, variance σ^2. Define $M_n = \frac{X_1 + ... + X_n}{n}$.

2.12 Chebyshev Proves the Weak Law of Large Numbers

Let $X_1, ..., X_n$ be i.i.d. with mean μ, variance σ^2. Define $M_n = \frac{X_1 + ... + X_n}{n}$.

Then $E(M_n) = \mu$ and $\mathrm{Var}(M_n) = \frac{\sigma^2}{n}$.

2.12 Chebyshev Proves the Weak Law of Large Numbers

Let $X_1, ..., X_n$ be i.i.d. with mean μ, variance σ^2. Define $M_n = \frac{X_1 + ... + X_n}{n}$.

Then $E(M_n) = \mu$ and $\mathrm{Var}(M_n) = \frac{\sigma^2}{n}$.

Applying Chebyshev:

$$P(|M_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty$$

2.12 Chebyshev Proves the Weak Law of Large Numbers

Let $X_1, ..., X_n$ be i.i.d. with mean μ, variance σ^2. Define $M_n = \frac{X_1+...+X_n}{n}$.

Then $E(M_n) = \mu$ and $\text{Var}(M_n) = \frac{\sigma^2}{n}$.

Applying Chebyshev:

$$P(|M_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty$$

This is the **Weak Law of Large Numbers** (WLLN): the sample mean converges in probability to the true mean.

2.12 Chebyshev Proves the Weak Law of Large Numbers

Let $X_1, ..., X_n$ be i.i.d. with mean μ, variance σ^2. Define $M_n = \frac{X_1+...+X_n}{n}$.

Then $E(M_n) = \mu$ and $\mathrm{Var}(M_n) = \frac{\sigma^2}{n}$.

Applying Chebyshev:

$$P(|M_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty$$

This is the **Weak Law of Large Numbers** (WLLN): the sample mean converges in probability to the true mean.

Note: We've now seen Chebyshev used to *prove a theorem*, not just compute a bound. This is a common pattern in probability theory.

2.13 Summary: The Hierarchy of Bounds

As we use **more information**, we get **tighter bounds**:

2.13 Summary: The Hierarchy of Bounds

As we use **more information**, we get **tighter bounds**:

Bound	**Information Used**	**Tightness**
Markov	$E(X)$ only	Loosest
Chebyshev	$E(X)$ and $\text{Var}(X)$	Tighter
Exact distribution	Full PMF/PDF	Exact

2.13 Summary: The Hierarchy of Bounds

As we use **more information**, we get **tighter bounds**:

Bound	**Information Used**	**Tightness**
Markov	$E(X)$ only	Loosest
Chebyshev	$E(X)$ and $\mathrm{Var}(X)$	Tighter
Exact distribution	Full PMF/PDF	Exact

Note: There are even stronger bounds (Chernoff, Hoeffding) that use moment-generating functions. These are common in algorithms and machine learning but beyond the scope of this course.

2.14 Recap

Today we covered:

- **Markov's inequality**: $P(X \geq a) \leq \frac{E(X)}{a}$ for nonnegative X
- **Chebyshev's inequality**: $P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$ for any X
- Chebyshev follows from applying Markov to $(X - \mu)^2$
- These bounds are **distribution-free** — valid without knowing the exact distribution
- Chebyshev proves the **Weak Law of Large Numbers**: $M_n \to \mu$ in probability
- Applications: algorithm analysis, hashing, polling, sample size determination