Outline

Introduction

Markov’s Inequality and Chebyshev’s Inequality

Weak Law of Large Numbers
Example 1

Let $X_1$ and $X_2$ be two independent random variables, both uniformly distributed on $[0,1]$. How does the probability density of $X_1 + X_2$ look like? What happens for $X_1 + X_2 + X_3$ etc.? Answer
Intro: Sum of Independent (Uniform) Random Variables

Example 1

Let $X_1$ and $X_2$ be two independent random variables, both uniformly distributed on $[0, 1]$. How does the probability density of $X_1 + X_2$ look like? What happens for $X_1 + X_2 + X_3$ etc.?

Let us try to sketch the densities without explicit computations.

Answer

This is also called “convolution”. The detailed calculation for $f_{X_1 + X_2}$ can be found at the end of these slides. The exact distribution is known for any number of random variables under the name Irwin-Hall distribution.
Example 1

Let $X_1$ and $X_2$ be two independent random variables, both uniformly distributed on $[0, 1]$. How does the probability density of $X_1 + X_2$ look like? What happens for $X_1 + X_2 + X_3$ etc.?

---

Let us try to sketch the densities without explicit computations.

---

This is also called “convolution”. The detailed calculation for $f_{X_1 + X_2}$ can be found at the end of these slides. The exact distribution is known for any number of random variables under the name Irwin-Hall distribution.
Let \( X_1 \) and \( X_2 \) be two independent random variables, both uniformly distributed on \([0, 1]\). How does the probability density of \( X_1 + X_2 \) look like? What happens for \( X_1 + X_2 + X_3 \) etc.?

Let us try to sketch the densities without explicit computations\(^a\)

\[ f_{X_1} = f_{X_2} \]

\[ 0 \quad 0.5 \quad 1 \]

\[^a\text{This is also called “convolution”. The detailed calculation for } f_{X_1+X_2} \text{ can be found at the end of these slides. The exact distribution is known for any number of random variables under the name Irwin-Hall distribution.} \]
Let $X_1$ and $X_2$ be two independent random variables, both uniformly distributed on $[0, 1]$. How does the probability density of $X_1 + X_2$ look like? What happens for $X_1 + X_2 + X_3$ etc.?

Let us try to sketch the densities without explicit computations\(^a\)

\(^a\)This is also called “convolution”. The detailed calculation for $f_{X_1+X_2}$ can be found at the end of these slides. The exact distribution is known for any number of random variables under the name Irwin-Hall distribution.
Let $X_1$ and $X_2$ be two independent random variables, both uniformly distributed on $[0, 1]$. How does the probability density of $X_1 + X_2$ look like? What happens for $X_1 + X_2 + X_3$ etc.?

Answer

Let us try to sketch the densities without explicit computations.

---

This is also called “convolution”. The detailed calculation for $f_{X_1 + X_2}$ can be found at the end of these slides. The exact distribution is known for any number of random variables under the name Irwin-Hall distribution.
Intro: Sum of Independent (Uniform) Random Variables

Example 1

Let $X_1$ and $X_2$ be two independent random variables, both uniformly distributed on $[0, 1]$. How does the probability density of $X_1 + X_2$ look like? What happens for $X_1 + X_2 + X_3$ etc.?

Answer

Let us try to sketch the densities without explicit computations$^a$

$^a$This is also called “convolution”. The detailed calculation for $f_{X_1+X_2}$ can be found at the end of these slides. The exact distribution is known for any number of random variables under the name Irwin-Hall distribution.
Intro: Sum of Independent (Uniform) Random Variables

Example 1

Let $X_1$ and $X_2$ be two independent random variables, both uniformly distributed on $[0, 1]$. How does the probability density of $X_1 + X_2$ look like? What happens for $X_1 + X_2 + X_3$ etc.?

Let us try to sketch the densities without explicit computations$^a$

$^a$This is also called “convolution”. The detailed calculation for $f_{X_1 + X_2}$ can be found at the end of these slides. The exact distribution is known for any number of random variables under the name Irwin-Hall distribution.

Answer
Motivation

We will study sums of independent variables. How does their distribution look like, and how well do they concentrate around the expectation?

![Graph showing the distribution of sums of independent variables.]
Motivation

We will study sums of independent variables. How does their distribution look like, and how well do they concentrate around the expectation?

1. Markov’s inequality
2. Chebyshev’s inequality
3. Law of Large Numbers
4. Central Limit Theorem
Motivation

We will study **sums of independent variables**. How does their distribution look like, and how well do they **concentrate** around the expectation?

1. Markov’s inequality
2. Chebyshev’s inequality
3. Law of Large Numbers
4. Central Limit Theorem

Re-use concepts from previous lectures:
1. Independence (Random Var.) (Lec. 1, 7)
2. Expectation and Variance (Lec. 2, 3)
3. Normal Distribution (Lec. 5)
4. Sums of Random Variables (Lec. 6)
Outline

Introduction

Markov's Inequality and Chebyshev's Inequality

Weak Law of Large Numbers
Markov’s Inequality

For any non-negative random variable $X$ with finite $\mathbb{E}[X]$, it holds for any $a > 0$,

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

For any non-negative random variable $X$ with finite $\mathbb{E}[X]$, it holds for any $a > 0$,

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

Markov’s inequality is a so-called tail-bound: it upper bounds the probability that the random variable exceeds its mean.
Markov’s Inequality

For any **non-negative** random variable $X$ with finite $E[X]$, it holds for any $a > 0$,

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

Markov’s inequality is a so-called **tail-bound**: it upper bounds the probability that the random variable **exceeds** its mean.

**Comments:**
Markov’s Inequality

For any non-negative random variable $X$ with finite $\mathbb{E}[X]$, it holds for any $a > 0$,

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

Markov's inequality is a so-called tail-bound: it upper bounds the probability that the random variable exceeds its mean.

Comments:
- Markov’s inequality can be rewritten as: for any $\delta > 0$,

$$\mathbb{P}[X \geq \delta \cdot \mathbb{E}[X]] \leq \frac{1}{\delta}.$$
Markov’s Inequality

For any non-negative random variable \( X \) with finite \( \mathbb{E}[X] \), it holds for any \( a > 0 \),

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

Markov’s inequality is a so-called tail-bound: it upper bounds the probability that the random variable exceeds its mean.

Comments:

- Markov’s inequality can be rewritten as: for any \( \delta > 0 \),

\[
P[X \geq \delta \cdot \mathbb{E}[X]] \leq 1/\delta.
\]

- **Advantage**: Very basic inequality, we only need to know \( \mathbb{E}[X] \)
- **Downside**: For many distributions, the tail bound might be quite loose
Markov’s Inequality

For any non-negative random variable $X$ with finite $\mathbb{E}[X]$, it holds for any $a > 0$,

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

Markov’s inequality is a so-called tail-bound: it upper bounds the probability that the random variable exceeds its mean.

Comments:

- Markov’s inequality can be rewritten as: for any $\delta > 0$,

$$\mathbb{P}[X \geq \delta \cdot \mathbb{E}[X]] \leq 1/\delta.$$

- **Advantage**: Very basic inequality, we only need to know $\mathbb{E}[X]$

- **Downside**: For many distributions, the tail bound might be quite loose

- Proof is similar to the proof of Chebyshev’s inequality (Exercise!)
Applying Markov’s Inequality

Consider throwing an unbiased, six-sided dice 120 times and let $X$ denote the number of times we obtain a six.

1. Derive an upper bound on $P[X \geq 30]$.
2. Can you also derive an upper bound on $P[X \leq 10]$?

Answer

Example 2

Intro to Probability Markov’s Inequality and Chebyshev’s Inequality
Applying Markov’s Inequality

Consider throwing an unbiased, six-sided dice 120 times and let $X$ denote the number of times we obtain a six.

1. Derive an upper bound on $P[X \geq 30]$.
2. Can you also derive an upper bound on $P[X \leq 10]$?

Answer

1. First compute $E[X]$
Applying Markov’s Inequality

Example 2

Consider throwing an unbiased, six-sided dice 120 times and let $X$ denote the number of times we obtain a six.

1. Derive an upper bound on $P[ X \geq 30 ]$.
2. Can you also derive an upper bound on $P[ X \leq 10 ]$?

Answer

1. First compute $E[ X ] = 1/6 \cdot 120 = 20$
Consider throwing an unbiased, six-sided dice 120 times and let \( X \) denote the number of times we obtain a six.

1. Derive an upper bound on \( P \left[ X \geq 30 \right] \).
2. Can you also derive an upper bound on \( P \left[ X \leq 10 \right] \)?

\[
\text{Answer:} \quad 1. \text{First compute } E\left[ X \right] = \frac{1}{6} \cdot 120 = 20. \text{ Then by Markov:}
\]

\[
P \left[ X \geq 30 \right] \leq \frac{20}{30} = \frac{2}{3}.
\]
Applying Markov’s Inequality

Consider throwing an unbiased, six-sided dice 120 times and let $X$ denote the number of times we obtain a six.

1. Derive an upper bound on $P[X \geq 30]$.
2. Can you also derive an upper bound on $P[X \leq 10]$?

---

**Answer**

1. First compute $E[X] = 1/6 \cdot 120 = 20$. Then by Markov:

   $$P[X \geq 30]$$
Consider throwing an unbiased, six-sided dice 120 times and let $X$ denote the number of times we obtain a six.

1. Derive an upper bound on $\Pr[X \geq 30]$.
2. Can you also derive an upper bound on $\Pr[X \leq 10]$?

**Answer**

1. First compute $\mathbb{E}[X] = 1/6 \cdot 120 = 20$. Then by Markov:

$$\Pr[X \geq 30] \leq \frac{20}{30} = \frac{2}{3}.$$
Consider throwing an unbiased, six-sided dice 120 times and let $X$ denote the number of times we obtain a six.

1. Derive an upper bound on $\Pr[X \geq 30]$.
2. Can you also derive an upper bound on $\Pr[X \leq 10]$?

Answer

1. First compute $E[X] = 1/6 \cdot 120 = 20$. Then by Markov:

\[ \Pr[X \geq 30] \leq \frac{20}{30} = \frac{2}{3}. \]

2. Consider now the second bound.
Applying Markov’s Inequality

Example 2

Consider throwing an unbiased, six-sided dice 120 times and let $X$ denote the number of times we obtain a six.

1. Derive an upper bound on $P[ X \geq 30 ]$.
2. Can you also derive an upper bound on $P[ X \leq 10 ]$?

Answer

1. First compute $E[ X ] = 1/6 \cdot 120 = 20$. Then by Markov:

   $P[ X \geq 30 ] \leq \frac{20}{30} = \frac{2}{3}$.

2. Consider now the second bound.
   - Define a new random variable $Y := 120 - X$. 
Applying Markov’s Inequality

Consider throwing an unbiased, six-sided dice 120 times and let $X$ denote the number of times we obtain a six.

1. Derive an upper bound on $P[X \geq 30]$.
2. Can you also derive an upper bound on $P[X \leq 10]$?

---

**Answer**

1. First compute $E[X] = 1/6 \cdot 120 = 20$. Then by Markov:

   $$P[X \geq 30] \leq \frac{20}{30} = \frac{2}{3}.$$

2. Consider now the second bound.
   - Define a new random variable $Y := 120 - X$.
     $\Rightarrow$ This random variable is also non-negative (as $X \leq 120$).
Applying Markov’s Inequality

Example 2

Consider throwing an unbiased, six-sided dice 120 times and let $X$ denote the number of times we obtain a six.

1. Derive an upper bound on $P\left[ X \geq 30 \right]$.
2. Can you also derive an upper bound on $P\left[ X \leq 10 \right]$?

Answer

1. First compute $E\left[ X \right] = \frac{1}{6} \cdot 120 = 20$. Then by Markov:

$$P\left[ X \geq 30 \right] \leq \frac{20}{30} = \frac{2}{3}.$$

2. Consider now the second bound.
   - Define a new random variable $Y := 120 - X$.
   - This random variable is also non-negative (as $X \leq 120$).
   - Applying Markov’s inequality (equivalent version) to $Y$ yields:
Applying Markov’s Inequality

Example 2

Consider throwing an unbiased, six-sided dice 120 times and let $X$ denote the number of times we obtain a six.

1. Derive an upper bound on $P[ X \geq 30 ]$.
2. Can you also derive an upper bound on $P[ X \leq 10 ]$?

Answer

1. First compute $E[ X ] = \frac{1}{6} \cdot 120 = 20$. Then by Markov:

   $$P[ X \geq 30 ] \leq \frac{20}{30} = \frac{2}{3}.$$ 

2. Consider now the second bound.
   - Define a new random variable $Y := 120 - X$.
     ⇒ This random variable is also non-negative (as $X \leq 120$).
   - Applying Markov’s inequality (equivalent version) to $Y$ yields:

     $$P[ X \leq 10 ] = P[ Y \geq 110 ]$$
Applying Markov’s Inequality

**Example 2**

Consider throwing an unbiased, six-sided dice 120 times and let $X$ denote the number of times we obtain a six.

1. Derive an upper bound on $P[ X \geq 30 ]$.
2. Can you also derive an upper bound on $P[ X \leq 10 ]$?

**Answer**

1. First compute $E[ X ] = \frac{1}{6} \cdot 120 = 20$. Then by Markov:

   $$ P[ X \geq 30 ] \leq \frac{20}{30} = \frac{2}{3}. $$

2. Consider now the second bound.
   - Define a new random variable $Y := 120 - X$.
     $\Rightarrow$ This random variable is also non-negative (as $X \leq 120$).
   - Applying Markov’s inequality (equivalent version) to $Y$ yields:

   $$ P[ X \leq 10 ] = P[ Y \geq 110 ] = P \left[ Y \geq \frac{110}{100} \cdot E[ Y ] \right] $$
Applying Markov’s Inequality

Example 2

Consider throwing an unbiased, six-sided dice 120 times and let $X$ denote the number of times we obtain a six.

1. Derive an upper bound on $P[ X \geq 30 ]$.
2. Can you also derive an upper bound on $P[ X \leq 10 ]$?

Answer

1. First compute $E[ X ] = \frac{1}{6} \cdot 120 = 20$. Then by Markov:

$$P[ X \geq 30 ] \leq \frac{20}{30} = \frac{2}{3}.$$  

2. Consider now the second bound.

- Define a new random variable $Y := 120 - X$.
- This random variable is also non-negative (as $X \leq 120$).
- Applying Markov’s inequality (equivalent version) to $Y$ yields:

$$P[ X \leq 10 ] = P[ Y \geq 110 ] = P \left[ Y \geq \frac{110}{100} \cdot E[ Y] \right] \leq \frac{100}{110} = \frac{10}{11}.$$
Applying Markov’s Inequality

Example 2

Consider throwing an unbiased, six-sided dice 120 times and let $X$ denote the number of times we obtain a six.

1. Derive an upper bound on $P[X \geq 30]$.
2. Can you also derive an upper bound on $P[X \leq 10]$?

---

**Answer**

1. First compute $E[X] = 1/6 \cdot 120 = 20$. Then by Markov:

$$P[X \geq 30] \leq \frac{20}{30} = \frac{2}{3}.$$  

2. Consider now the second bound.
   - Define a new random variable $Y := 120 - X$.
   - This random variable is also non-negative (as $X \leq 120$).
   - Applying Markov’s inequality (equivalent version) to $Y$ yields:

$$P[X \leq 10] = P[Y \geq 110] = P \left[ Y \geq \frac{110}{100} \cdot E[Y] \right]$$

$$\leq \frac{100}{110} = \frac{10}{11}.$$  

Both bounds, especially the second, are quite loose!
Chebyshev’s Inequality

For any random variable $X$ with finite $E[X]$ and $V[X]$, for any $a > 0,$

$$P[|X - E[X]| \geq a] \leq \frac{V[X]}{a^2}.$$
Chebyshev’s Inequality

For any random variable $X$ with finite $E[X]$ and $V[X]$, for any $a > 0$,

$$P[|X - E[X]| \geq a] \leq V[X]/a^2.$$ 

Comments:
- can be rewritten as:

$$P\left[|X - E[X]| \geq \sqrt{\delta \cdot V[X]}\right] \leq 1/\delta.$$ 

The “$\mu \pm \text{a few } \sigma$” rule. Most of the probability mass is within a few standard deviations from $\mu$. 

P. Chebyshev (1821-1894)
Chebyshev’s Inequality

For any random variable $X$ with finite $E[X]$ and $V[X]$, for any $a > 0$, 

$$P[|X - E[X]| \geq a] \leq \frac{V[X]}{a^2}.$$

Comments:

- can be rewritten as:
  $$P\left[|X - E[X]| \geq \sqrt{\delta \cdot V[X]}\right] \leq \frac{1}{\delta}.$$

- Unlike Markov, Chebyshev’s inequality holds is two-sided and also for random variables with negative values

- In most cases, Chebyshev’s inequality yields much stronger bounds than Markov (however, it requires knowledge not only of $E[X]$ but also $V[X]$!)
Chebyshev’s Inequality

For any random variable $X$ with finite $E[X]$ and $V[X]$, for any $a > 0$, 

$$P[|X - E[X]| \geq a] \leq \frac{V[X]}{a^2}.$$ 

Comments:

- can be rewritten as:
  $$P \left[ |X - E[X]| \geq \sqrt{\delta \cdot V[X]} \right] \leq \frac{1}{\delta}.$$

- Unlike Markov, Chebyshev’s inequality holds is two-sided and also for random variables with negative values

- In most cases, Chebyshev’s inequality yields much stronger bounds than Markov (however, it requires knowledge not only of $E[X]$ but also $V[X]$!)

- Chebyshev’s inequality is also known as Second Moment Method
Derivation of Chebychev’s inequality

Proof

We will give a self-contained proof for a continuous random variable $X$ (the case for discrete $X$ is analogous).

Write down the definition of $V[X]$ and then lower bound:

$V[X] = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f_X(x) \, dx \geq \int |x - \mu| \geq a (x - \mu)^2 \cdot f_X(x) \, dx \geq \int a^2 \cdot f_X(x) \, dx = a^2 \cdot P[|X - \mu| \geq a].$

Dividing both sides by $a^2$ yields the result.

Proof

Exercise: Can you find a proof that uses Markov’s inequality?
Derivation of Chebychev’s inequality

Proof

- We will give a self-contained proof for a continuous random variable \( X \) (the case for discrete \( X \) is analogous).

Exercise: Can you find a proof that uses Markov’s inequality?
Derivation of Chebychev’s inequality

Proof

- We will give a self-contained proof for a continuous random variable $X$ (the case for discrete $X$ is analogous).
- Write down the definition of $\mathbb{V}[X]$ and then lower bound:

$$
\mathbb{V}[X] = \mathbb{E}[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f_X(x) \, dx
$$

Exercise: Can you find a proof that uses Markov’s inequality?
Derivation of Chebychev’s inequality

Proof

- We will give a self-contained proof for a continuous random variable $X$ (the case for discrete $X$ is analogous).
- Write down the definition of $V[X]$ and then lower bound:

\[
V[X] = E \left[ (X - \mu)^2 \right] = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f_X(x) \, dx
\]

\[
\geq \int_{|x - \mu| \geq a} (x - \mu)^2 \cdot f_X(x) \, dx
\]

Dividing both sides by $a^2$ yields the result.

Exercise: Can you find a proof that uses Markov’s inequality?
Proof

- We will give a self-contained proof for a continuous random variable $X$ (the case for discrete $X$ is analogous).
- Write down the definition of $V[X]$ and then lower bound:

$$V[X] = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f_X(x) \, dx$$

$$\geq \int_{|x-\mu| \geq a} (x - \mu)^2 \cdot f_X(x) \, dx$$

$$\geq \int_{|x-\mu| \geq a} a^2 \cdot f_X(x) \, dx$$

Exercise: Can you find a proof that uses Markov’s inequality?
Derivation of Chebychev’s inequality

Proof

- We will give a self-contained proof for a continuous random variable \( X \) (the case for discrete \( X \) is analogous).
- Write down the definition of \( \mathbb{V}[X] \) and then lower bound:

\[
\mathbb{V}[X] = \mathbb{E} \left[ (X - \mu)^2 \right] = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f_X(x) \, dx \\
\geq \int_{|x-\mu| \geq a} (x - \mu)^2 \cdot f_X(x) \, dx \\
\geq \int_{|x-\mu| \geq a} a^2 \cdot f_X(x) \, dx \\
= a^2 \cdot \int_{|x-\mu| \geq a} f_X(x) \, dx
\]

Dividing both sides by \( a^2 \) yields the result.

Exercise: Can you find a proof that uses Markov’s inequality?
Derivation of Chebychev’s inequality

Proof

- We will give a self-contained proof for a continuous random variable $X$ (the case for discrete $X$ is analogous).
- Write down the definition of $\text{Var}[X]$ and then lower bound:

\[
\text{Var}[X] = \mathbb{E}[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f_X(x) \, dx
\]

\[
\geq \int_{|x-\mu| \geq a} (x - \mu)^2 \cdot f_X(x) \, dx
\]

\[
\geq \int_{|x-\mu| \geq a} a^2 \cdot f_X(x) \, dx
\]

\[
= a^2 \cdot \int_{|x-\mu| \geq a} f_X(x) \, dx
\]

\[
= a^2 \cdot \mathbb{P}[|X - \mu| \geq a].
\]

Exercise: Can you find a proof that uses Markov's inequality?
Proof

- We will give a self-contained proof for a continuous random variable $X$ (the case for discrete $X$ is analogous).

- Write down the definition of $V[X]$ and then lower bound:

\[
V[X] = E \left[ (X - \mu)^2 \right] = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f_X(x) \, dx
\]

\[
\geq \int_{|x-\mu| \geq a} (x - \mu)^2 \cdot f_X(x) \, dx
\]

\[
\geq \int_{|x-\mu| \geq a} a^2 \cdot f_X(x) \, dx
\]

\[
= a^2 \cdot \int_{|x-\mu| \geq a} f_X(x) \, dx
\]

\[
= a^2 \cdot P[|X - \mu| \geq a].
\]

- Dividing both sides by $a^2$ yields the result.

Exercise: Can you find a proof that uses Markov’s inequality?
Example 3

Throw an unbiased coin $n$ times and let $X$ be the total number of heads. In an experiment, with $n$ large, we would usually expect a number of heads that is close to the expectation. Can we justify that?

Answer

$X \sim \text{Bin}(n, \frac{1}{2})$ so $E[X] = n \cdot \frac{1}{2}$.

Markov's inequality:

For any $\delta > 0$, $P[X \geq (1 + \delta) \cdot E[X]] \leq \frac{1}{1 + \delta}$.

Chebychev's inequality:

We have $V[X] = np(1-p) = n \cdot \frac{1}{2} \cdot \frac{1}{2}$. For any $\delta > 0$, $P[|X - E[X]| \geq \delta \cdot E[X]] \leq \frac{1}{\delta^2} \cdot \frac{1}{4} \cdot \frac{n}{2}$.

Not good! Independent of $n$.

Much better! (Inversely) Linear in $n$.
Example: Chebychev is (usually) much stronger than Markov

Example 3

Throw an unbiased coin $n$ times and let $X$ be the total number of heads. In an experiment, with $n$ large, we would usually expect a number of heads that is close to the expectation. Can we justify that?

Answer

$$X \sim Bin(n, 1/2) \text{ so } \mathbb{E}[X] = n \cdot \frac{1}{2}.$$
Example: Chebychev is (usually) much stronger than Markov

Example 3

Throw an unbiased coin $n$ times and let $X$ be the total number of heads. In an experiment, with $n$ large, we would usually expect a number of heads that is close to the expectation. Can we justify that?

Answer

$X \sim Bin(n, 1/2)$ so $E[X] = n \cdot \frac{1}{2}$.  
- Markov's inequality: For any $\delta > 0$,

\[
P[X \geq (1 + \delta) \cdot E[X]] \leq \frac{1}{1 + \delta}
\]
Example: Chebychev is (usually) much stronger than Markov

Example 3

Throw an unbiased coin $n$ times and let $X$ be the total number of heads. In an experiment, with $n$ large, we would usually expect a number of heads that is close to the expectation. Can we justify that?

$X \sim Bin(n, 1/2)$ so $E[X] = n \cdot \frac{1}{2}$.

- **Markov’s inequality:** For any $\delta > 0$,

$$P[X \geq (1 + \delta) \cdot E[X]] \leq \frac{1}{1 + \delta}$$

Not good! **Independent of $n$**
Example: Chebychev is (usually) much stronger than Markov

Example 3

Throw an unbiased coin $n$ times and let $X$ be the total number of heads. In an experiment, with $n$ large, we would usually expect a number of heads that is close to the expectation. Can we justify that?

Answer

$X \sim Bin(n, 1/2)$ so $E[X] = n \cdot \frac{1}{2}$.

- **Markov’s inequality:** For any $\delta > 0$,

$$P[X \geq (1 + \delta) \cdot E[X]] \leq \frac{1}{1 + \delta}$$

- **Chebychev’s inequality:**

Not good! Independent of $n$
Example: Chebychev is (usually) much stronger than Markov

Example 3

Throw an unbiased coin $n$ times and let $X$ be the total number of heads. In an experiment, with $n$ large, we would usually expect a number of heads that is close to the expectation. Can we justify that?

Answer

$X \sim Bin(n, 1/2)$ so $\mathbb{E}[X] = n \cdot \frac{1}{2}$.

- **Markov’s inequality:** For any $\delta > 0$,
  
  $$
  \mathbb{P}[X \geq (1 + \delta) \cdot \mathbb{E}[X]] \leq \frac{1}{1 + \delta}
  $$

- **Chebychev’s inequality:**
  
  $\Rightarrow$ We have $\mathbb{V}[X] = np(1 - p) = n \cdot 1/2 \cdot 1/2$. For any $\delta > 0$,
Example: Chebychev is (usually) much stronger than Markov

Example 3

Throw an unbiased coin \( n \) times and let \( X \) be the total number of heads. In an experiment, with \( n \) large, we would usually expect a number of heads that is close to the expectation. Can we justify that?

\[
\begin{align*}
X &\sim Bin(n, 1/2) \text{ so } \mathbb{E}[X] = n \cdot \frac{1}{2}. \\
\text{Markov’s inequality: For any } \delta > 0, \\
\mathbb{P}[X \geq (1 + \delta) \cdot \mathbb{E}[X]] &\leq \frac{1}{1 + \delta} \\
\text{Chebychev’s inequality:} \\
\Rightarrow \text{ We have } \mathbb{V}[X] = np(1 - p) = n \cdot \frac{1}{2} \cdot \frac{1}{2}. \text{ For any } \delta > 0, \\
\mathbb{P}[X \geq (1 + \delta) \cdot \mathbb{E}[X]] &= \mathbb{P}[X - \mathbb{E}[X] \geq \delta \cdot \mathbb{E}[X]] \\
&\leq \mathbb{P}[|X - n/2| \geq \delta \cdot (n/2)] \\
&\leq \frac{n \cdot 1/4}{\delta^2 (n/2)^2} = \frac{1}{\delta^2 n}
\end{align*}
\]
Example: Chebychev is (usually) much stronger than Markov

Throw an unbiased coin $n$ times and let $X$ be the total number of heads. In an experiment, with $n$ large, we would usually expect a number of heads that is close to the expectation. Can we justify that?

Answer

$X \sim Bin(n, 1/2)$ so $E[X] = n \cdot \frac{1}{2}$.

- **Markov’s inequality**: For any $\delta > 0$,

  $$P[X \geq (1 + \delta) \cdot E[X]] \leq \frac{1}{1 + \delta}$$

- **Chebychev’s inequality**: Not good! Independent of $n$

  $$P[X \geq (1 + \delta) \cdot E[X]] = P[X - E[X] \geq \delta \cdot E[X]]$$

  $$\leq P[|X - n/2| \geq \delta \cdot (n/2)]$$

  $$\leq \frac{n \cdot 1/4}{\delta^2 (n/2)^2} = \frac{1}{\delta^2 n}$$

Much better! (Inversely) Linear in $n$
Outline

Introduction

Markov's Inequality and Chebyshev's Inequality

Weak Law of Large Numbers
Let $X_n := 1/n \cdot \sum_{i=1}^{n} X_i$, where the $X_i$’s are i.i.d. with finite expectation $\mu$ and finite variance $\sigma^2$.
Law of Large Numbers

The Weak Law of Large Numbers

Let \( \bar{X}_n := \frac{1}{n} \cdot \sum_{i=1}^{n} X_i \), where the \( X_i \)'s are i.i.d. with finite expectation \( \mu \) and finite variance \( \sigma^2 \).

"For even the most stupid of men, by some instinct of nature, by himself and without any instruction (which is a remarkable thing), is convinced that the more observations have been made, the less danger there is of wandering from one's goal."

J. Bernoulli (1655-1705)
The Weak Law of Large Numbers

Let $\overline{X}_n := \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$, where the $X_i$'s are i.i.d. with finite expectation $\mu$ and finite variance $\sigma^2$. Then, for any $\epsilon > 0$,

$$
\lim_{n \to \infty} \mathbb{P} \left[ |\overline{X}_n - \mu| > \epsilon \right] = 0
$$
Law of Large Numbers

The Weak Law of Large Numbers

Let $\bar{X}_n := \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$, where the $X_i$’s are i.i.d. with finite expectation $\mu$ and finite variance $\sigma^2$. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} P \left[ |\bar{X}_n - \mu| > \epsilon \right] = 0$$

"Power of Averaging": repeated samples allow us to estimate $\mu$. A similar statement holds even if the $X_i$’s are not identically distributed.

There is also a strong law of large numbers:

$$P \left[ \lim_{n \to \infty} X_n = \mu \right] = 1.$$
The Weak Law of Large Numbers

Let \( \bar{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i \), where the \( X_i \)'s are i.i.d. with finite expectation \( \mu \) and finite variance \( \sigma^2 \). Then, for any \( \epsilon > 0 \),

\[
\lim_{n \to \infty} P \left[ |\bar{X}_n - \mu| > \epsilon \right] = 0
\]

\( \forall \epsilon > 0 : \)
Law of Large Numbers

Let $X_n := \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$, where the $X_i$'s are i.i.d. with finite expectation $\mu$ and finite variance $\sigma^2$. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} P \left[ |X_n - \mu| > \epsilon \right] = 0$$

The Weak Law of Large Numbers

∀ε > 0: ∀δ > 0: independent and identically distributed
Law of Large Numbers

The Weak Law of Large Numbers

Let \( \overline{X}_n := \frac{1}{n} \cdot \sum_{i=1}^{n} X_i \), where the \( X_i \)'s are i.i.d. with finite expectation \( \mu \) and finite variance \( \sigma^2 \). Then, for any \( \epsilon > 0 \),

\[
\lim_{n \to \infty} \mathbb{P} \left[ |\overline{X}_n - \mu| > \epsilon \right] = 0
\]

\( \forall \epsilon > 0 : \forall \delta > 0 : \exists N > 0 : \)

"For even the most stupid of men, by some instinct of nature, by himself and without any instruction (which is a remarkable thing), is convinced that the more observations have been made, the less danger there is of wandering from one's goal."

J. Bernoulli (1655-1705)
Law of Large Numbers

The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i$, where the $X_i$’s are i.i.d. with finite expectation $\mu$ and finite variance $\sigma^2$. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P} \left[ |\overline{X}_n - \mu| > \epsilon \right] = 0$$

∀$\epsilon > 0$: ∀$\delta > 0$: ∃$N > 0$: ∀$n \geq N$: 
Law of Large Numbers

The Weak Law of Large Numbers

Let \( \overline{X}_n := \frac{1}{n} \cdot \sum_{i=1}^{n} X_i \), where the \( X_i \)'s are i.i.d. with finite expectation \( \mu \) and finite variance \( \sigma^2 \). Then, for any \( \epsilon > 0 \),

\[
\lim_{n \to \infty} P \left[ |\overline{X}_n - \mu| > \epsilon \right] = 0
\]

\( \forall \epsilon > 0: \forall \delta > 0: \exists N > 0: \forall n \geq N: P \left[ |\overline{X}_n - \mu| > \epsilon \right] \leq \delta \)
Law of Large Numbers

The Weak Law of Large Numbers

Let \( \overline{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i \), where the \( X_i \)'s are i.i.d. with finite expectation \( \mu \) and finite variance \( \sigma^2 \). Then, for any \( \epsilon > 0 \),

\[
\lim_{n \to \infty} P \left[ |\overline{X}_n - \mu| > \epsilon \right] = 0
\]

\( \forall \epsilon > 0: \forall \delta > 0: \exists N > 0: \forall n \geq N: P \left[ |\overline{X}_n - \mu| > \epsilon \right] \leq \delta \)

- “Power of Averaging”: repeated samples allow us to estimate \( \mu \)
Law of Large Numbers

The Weak Law of Large Numbers

Let $\overline{X}_n := \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$, where the $X_i$’s are i.i.d. with finite expectation $\mu$ and finite variance $\sigma^2$. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}\left[|\overline{X}_n - \mu| > \epsilon\right] = 0$$

$\forall \epsilon > 0: \forall \delta > 0: \exists N > 0: \forall n \geq N: \mathbb{P}\left[|\overline{X}_n - \mu| > \epsilon\right] \leq \delta$

“Power of Averaging”: repeated samples allow us to estimate $\mu$

“For even the most stupid of men, by some instinct of nature, by himself and without any instruction (which is a remarkable thing), is convinced that the more observations have been made, the less danger there is of wandering from one’s goal.”

J. Bernoulli (1655-1705)
The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^n X_i$, where the $X_i$’s are i.i.d. with finite expectation $\mu$ and finite variance $\sigma^2$. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} P \left[ |\overline{X}_n - \mu| > \epsilon \right] = 0$$

$\forall \epsilon > 0: \forall \delta > 0: \exists N > 0: \forall n \geq N: P \left[ |\overline{X}_n - \mu| > \epsilon \right] \leq \delta$

- “Power of Averaging”: repeated samples allow us to estimate $\mu$
- A similar statement holds even if the $X_i$’s are not identically distributed.

“For even the most stupid of men, by some instinct of nature, by himself and without any instruction (which is a remarkable thing), is convinced that the more observations have been made, the less danger there is of wandering from one’s goal.”

J. Bernoulli (1655-1705)
Law of Large Numbers

The Weak Law of Large Numbers

Let $\bar{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i$, where the $X_i$’s are i.i.d. with finite expectation $\mu$ and finite variance $\sigma^2$. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} P \left[ |\bar{X}_n - \mu| > \epsilon \right] = 0$$

For all $\epsilon > 0$: For all $\delta > 0$: There exists $N > 0$: For all $n \geq N$: $P \left[ |\bar{X}_n - \mu| > \epsilon \right] \leq \delta$

- “Power of Averaging”: repeated samples allow us to estimate $\mu$
- A similar statement holds even if the $X_i$’s are not identically distributed.
- There is also a strong law of large numbers:

$$P \left[ \lim_{n \to \infty} \bar{X}_n = \mu \right] = 1.$$  

“For even the most stupid of men, by some instinct of nature, by himself and without any instruction (which is a remarkable thing), is convinced that the more observations have been made, the less danger there is of wandering from one’s goal.”

J. Bernoulli (1655-1705)
Let $X_i$ be independent random variables taking values in $\{-1, +1\}$ with probability $1/2$ each. Consider $\tilde{X}_n := \sum_{i=1}^{n} X_i$ for any $n = 0, 1, ..., 200$.

How does a "typical" realisation look like?
Let $X_i$ be independent random variables taking values $\in \{-1, +1\}$ with probability $1/2$ each.
Let $X_i$ be independent random variables taking values $\in \{-1, +1\}$ with probability $1/2$ each.

Consider $\tilde{X}_n := \sum_{i=1}^{n} X_i$ for any $n = 0, 1, \ldots, 200$.
Let $X_i$ be independent random variables taking values $\in \{-1, +1\}$ with probability $1/2$ each.

Consider $\tilde{X}_n := \sum_{i=1}^{n} X_i$ for any $n = 0, 1, \ldots, 200$.

How does a “typical” realisation look like?
Illustration of Weak Law of Large Numbers (2/4)
Illustration of Weak Law of Large Numbers (2/4)
Illustration of Weak Law of Large Numbers (2/4)
Illustration of Weak Law of Large Numbers (2/4)
Illustration of Weak Law of Large Numbers (2/4)
Illustration of Weak Law of Large Numbers (2/4)
Illustration of Weak Law of Large Numbers (2/4)
Plot of the Distributions for $n = 0, 1, \ldots, 20$
Plot of the Distributions for $n = 0, 1, \ldots, 50$
Plot of the Distributions for $n = 0, 1, \ldots, 80$
Plot of the Distributions for $n = 0, 1, \ldots, 80$

$P[\tilde{X}_n = x]$
Interlude: Approximation of $P[\tilde{X}_n = 0]$

Try to find an expression for $P[\tilde{X}_n = 0]$. Using Stirling's approximation for $n!$, conclude that $P[\tilde{X}_n = 0] = \Theta(1/\sqrt{n})$ for even integers $n$. 

Exercise: Intro to Probability Weak Law of Large Numbers 18
Try to find an expression for $P[\tilde{X}_n = 0]$. Using Stirling’s approximation for $n!$, conclude that $P[\tilde{X}_n = 0] = \Theta(1/\sqrt{n})$ for even integers $n$. 

Exercise
• Let $X_i$ be independent random variables taking values $\in \{-1, +1\}$ with probability $1/2$ each.

• Consider $\tilde{X}_n := \sum_{i=1}^{n} X_i$ for any $n = 0, 1, \ldots, 200$. This does not converge!
Illustration of Weak Law of Large Numbers (3/4)

- Let $X_i$ be independent random variables taking values $\in \{-1, +1\}$ with probability $1/2$ each.
- Consider $\tilde{X}_n := \sum_{i=1}^{n} X_i$ for any $n = 0, 1, \ldots, 200$.

This does not converge!
Let \( X_i \) be independent random variables taking values \( \in \{-1, +1\} \) with probability 1/2 each.

Consider \( \tilde{X}_n := \sum_{i=1}^{n} X_i \) for any \( n = 0, 1, \ldots, 200 \).

This does not converge!

Consider now the average (sample mean): \( \bar{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i \).
Illustration of Weak Law of Large Numbers (4/4)

\[ \frac{1}{\sqrt{n}} - \frac{1}{\sqrt{n}} \]

\( X_n \)

Intro to Probability

Weak Law of Large Numbers
Illustration of Weak Law of Large Numbers (4/4)

\[ \frac{1}{\sqrt{n}} - \frac{1}{\sqrt{n}} \]

Intro to Probability

Weak Law of Large Numbers
Illustration of Weak Law of Large Numbers (4/4)

\[ \frac{1}{\sqrt{n}} \]
Illustration of Weak Law of Large Numbers (4/4)

\[ \frac{1}{\sqrt{n}} - \frac{1}{\sqrt{n}} \]

\[ 0 \quad 20 \quad 40 \quad 60 \quad 80 \quad 100 \quad 120 \quad 140 \quad 160 \quad 180 \quad 200 \]

\[ -1 \quad -0.8 \quad -0.6 \quad -0.4 \quad -0.2 \quad 0 \quad 0.2 \quad 0.4 \quad 0.6 \quad 0.8 \quad 1 \]
Illustration of Weak Law of Large Numbers (4/4)

\[ \frac{1}{\sqrt{n}} - \frac{1}{\sqrt{n}} \]
Illustration of Weak Law of Large Numbers (4/4)

\[ \frac{1}{\sqrt{n}} - \frac{1}{\sqrt{n}} \]
Proof of the Weak Law of Large Numbers

The Weak Law of Large Numbers

Let $\overline{X}_n := \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$, where the $X_i$'s are i.i.d. with finite expectation $\mu$ and finite variance $\sigma^2$. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} P \left[ |\overline{X}_n - \mu| > \epsilon \right] = 0$$

Proof
Proof of the Weak Law of Large Numbers

The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i$, where the $X_i$’s are i.i.d. with finite expectation $\mu$ and finite variance $\sigma^2$. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} P \left[ |\overline{X}_n - \mu| > \epsilon \right] = 0$$

Proof

- Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i$
Proof of the Weak Law of Large Numbers

The Weak Law of Large Numbers

Let \( \bar{X}_n := \frac{1}{n} \cdot \sum_{i=1}^{n} X_i \), where the \( X_i \)'s are i.i.d. with finite expectation \( \mu \) and finite variance \( \sigma^2 \). Then, for any \( \epsilon > 0 \),

\[
\lim_{n \to \infty} P \left[ |\bar{X}_n - \mu| > \epsilon \right] = 0
\]

Proof

- Let \( \bar{X}_n := \frac{1}{n} \cdot \sum_{i=1}^{n} X_i \)
- Then \( E \left[ \bar{X}_n \right] = \mu \) and
  \[
  V \left[ \bar{X}_n \right] = \frac{1}{n^2} \cdot V \left[ \sum_{i=1}^{n} X_i \right] = \frac{1}{n^2} \cdot \sum_{i=1}^{n} V \left[ X_i \right] = \frac{1}{n} \cdot \sigma^2.
  \]
Proof of the Weak Law of Large Numbers

Let \( \overline{X}_n := \frac{1}{n} \cdot \sum_{i=1}^{n} X_i \), where the \( X_i \)'s are i.i.d. with finite expectation \( \mu \) and finite variance \( \sigma^2 \). Then, for any \( \epsilon > 0 \),

\[
\lim_{n \to \infty} P \left[ | \overline{X}_n - \mu | > \epsilon \right] = 0
\]

The Weak Law of Large Numbers

Proof

- Let \( \overline{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i \)
- Then \( E \left[ \overline{X}_n \right] = \mu \) and
  \[
  V \left[ \overline{X}_n \right] = \frac{1}{n^2} \cdot V \left[ \sum_{i=1}^{n} X_i \right] = \frac{1}{n^2} \cdot \sum_{i=1}^{n} V \left[ X_i \right] = \frac{1}{n} \cdot \sigma^2.
  \]
- Applying Chebyshev's inequality yields:
  \[
P \left[ \left| \overline{X}_n - E \left[ \overline{X}_n \right] \right| > \epsilon \right] \leq \frac{1}{\epsilon^2} \cdot V \left[ \overline{X}_n \right]
  \]
Proof of the Weak Law of Large Numbers

Let $X_n := 1/n \cdot \sum_{i=1}^{n} X_i$, where the $X_i$'s are i.i.d. with finite expectation $\mu$ and finite variance $\sigma^2$. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} P \left[ |X_n - \mu| > \epsilon \right] = 0$$

Proof

- Let $X_n := 1/n \cdot \sum_{i=1}^{n} X_i$
- Then $E \left[ X_n \right] = \mu$ and
  $$V \left[ X_n \right] = 1/n^2 \cdot V \left[ \sum_{i=1}^{n} X_i \right] = 1/n^2 \cdot \sum_{i=1}^{n} V \left[ X_i \right] = 1/n \cdot \sigma^2$$. 
- Applying Chebyshev's inequality yields:
  $$P \left[ \left| X_n - E \left[ X_n \right] \right| > \epsilon \right] \leq \frac{1}{\epsilon^2} \cdot V \left[ X_n \right] = \frac{\sigma^2}{n\epsilon^2}$$. 

Proof of the Weak Law of Large Numbers

Let \( \overline{X}_n := \frac{1}{n} \cdot \sum_{i=1}^{n} X_i \), where the \( X_i \)'s are i.i.d. with finite expectation \( \mu \) and finite variance \( \sigma^2 \). Then, for any \( \epsilon > 0 \),

\[
\lim_{n \to \infty} P \left[ |\overline{X}_n - \mu| > \epsilon \right] = 0
\]

Proof:

- Let \( \overline{X}_n := \frac{1}{n} \cdot \sum_{i=1}^{n} X_i \)
- Then \( E \left[ \overline{X}_n \right] = \mu \) and
  \[
  V \left[ \overline{X}_n \right] = \frac{1}{n^2} \cdot V \left[ \sum_{i=1}^{n} X_i \right] = \frac{1}{n^2} \cdot \sum_{i=1}^{n} V \left[ X_i \right] = \frac{1}{n} \cdot \sigma^2.
  \]
- Applying Chebyshev's inequality yields:
  \[
  P \left[ \left| \overline{X}_n - E \left[ \overline{X}_n \right] \right| > \epsilon \right] \leq \frac{1}{\epsilon^2} \cdot V \left[ \overline{X}_n \right] = \frac{\sigma^2}{n\epsilon^2}.
  \]
- For any (fixed) \( \epsilon > 0 \), the right hand side vanishes as \( n \to \infty \).
Proof of the Weak Law of Large Numbers

The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^n X_i$, where the $X_i$'s are i.i.d. with finite expectation $\mu$ and finite variance $\sigma^2$. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} P \left[ |\overline{X}_n - \mu| > \epsilon \right] = 0$$

Proof

- Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^n X_i$

- Then $E \left[ \overline{X}_n \right] = \mu$ and
  $$V \left[ \overline{X}_n \right] = 1/n^2 \cdot V \left[ \sum_{i=1}^n X_i \right] = 1/n^2 \cdot \sum_{i=1}^n V [X_i] = 1/n \cdot \sigma^2.$$

- Applying Chebyshev’s inequality yields:
  $$P \left[ \left| \overline{X}_n - E \left[ \overline{X}_n \right] \right| > \epsilon \right] \leq \frac{1}{\epsilon^2} \cdot V \left[ \overline{X}_n \right] = \frac{\sigma^2}{n\epsilon^2}.$$

- For any (fixed) $\epsilon > 0$, the right hand side vanishes as $n \to \infty$.

  (Let $\epsilon > 0$, $\delta > 0$. Pick $N = \frac{\sigma^2}{\epsilon^2 \cdot \delta}$. Then for any $n \geq N$, the probability above is smaller than $\delta$.)
Suppose that, instead of the expectation $\mu$, we want to estimate the probability of an event, e.g.,

$$p := \mathbb{P} \left[ X \in (a, b) \right], \text{ where } a < b.$$ 

How can we use the Law of Large Numbers?

---

**Example 4**

Let $X_1, X_2, \ldots, X_n \sim X$. For each $1 \leq i \leq n$, define:

$$Y_i = \begin{cases} 
1 & \text{if } X_i \in (a, b), \\
0 & \text{otherwise}.
\end{cases}$$

We have:

$$E[Y_i] = \mathbb{P}[X_i \in (a, b)] \cdot 1 + \mathbb{P}[X_i \notin (a, b)] \cdot 0 = p.$$ 

Similarly,

$$\text{Var}[Y_i] = p \left(1 - p\right).$$

The random variables $Y_1, Y_2, \ldots, Y_n$ are i.i.d., so we can apply the Law of Large Numbers to $Y_n$.

Can use similar argument to recover the probability mass or density!
Inferring Probabilities of an Event

Suppose that, instead of the expectation $\mu$, we want to estimate the probability of an event, e.g.,

$$p := P \left[ X \in (a, b) \right], \text{ where } a < b.$$

How can we use the Law of Large Numbers?

- Let $X_1, X_2, \ldots, X_n \sim X$. For each $1 \leq i \leq n$, define:

$$Y_i = \begin{cases} 
1 & \text{if } X_i \in (a, b), \\
0 & \text{otherwise.}
\end{cases}$$
Suppose that, instead of the expectation \( \mu \), we want to estimate the probability of an event, e.g.,

\[
p := P[X \in (a, b)], \text{ where } a < b.
\]

How can we use the Law of Large Numbers?

**Answer**

- Let \( X_1, X_2, \ldots, X_n \sim X \). For each \( 1 \leq i \leq n \), define:

\[
Y_i = \begin{cases} 
1 & \text{if } X_i \in (a, b], \\
0 & \text{otherwise}.
\end{cases}
\]

- We have:

\[
E[Y_i] = P[X_i \in (a, b)] \cdot 1 + P[X_i \notin (a, b)] \cdot 0 = p.
\]
Inferring Probabilities of an Event

Example 4

Suppose that, instead of the expectation $\mu$, we want to estimate the probability of an event, e.g.,

$$p := P[ X \in (a, b) ], \text{ where } a < b.$$ 

How can we use the Law of Large Numbers?

Let $X_1, X_2, \ldots, X_n \sim X$. For each $1 \leq i \leq n$, define:

$$Y_i = \begin{cases} 1 & \text{if } X_i \in (a, b], \\ 0 & \text{otherwise}. \end{cases}$$

- We have:

$$E[ Y_i ] = P[ X_i \in (a, b) ] \cdot 1 + P[ X_i \notin (a, b) ] \cdot 0 = p.$$ 

- Similarly, $V[ Y_i ] = p(1 - p)$. 

Can use similar argument to recover the probability mass or density!
Suppose that, instead of the expectation $\mu$, we want to estimate the probability of an event, e.g.,

$$p := P \left[ X \in (a, b) \right], \text{ where } a < b.$$

How can we use the Law of Large Numbers?

- Let $X_1, X_2, \ldots, X_n \sim X$. For each $1 \leq i \leq n$, define:

$$Y_i = \begin{cases} 1 & \text{if } X_i \in (a, b], \\ 0 & \text{otherwise}. \end{cases}$$

- We have:

$$E \left[ Y_i \right] = P \left[ X_i \in (a, b) \right] \cdot 1 + P \left[ X_i \notin (a, b) \right] \cdot 0 = p.$$

- Similarly, $V \left[ Y_i \right] = p(1 - p)$

- The random variables $Y_1, Y_2, \ldots, Y_n$ are i.i.d., so we can apply the Law of Large Numbers to $\overline{Y}_n$. 

Can use similar argument to recover the probability mass or density!
Inferring Probabilities of an Event

**Example 4**

Suppose that, instead of the expectation $\mu$, we want to estimate the probability of an event, e.g.,

$$p := P[X \in (a, b)], \text{ where } a < b.$$ 

How can we use the Law of Large Numbers?

**Answer**

- Let $X_1, X_2, \ldots, X_n \sim X$. For each $1 \leq i \leq n$, define:

$$Y_i = \begin{cases} 
1 & \text{if } X_i \in (a, b], \\
0 & \text{otherwise}.
\end{cases}$$

- We have:

$$E[Y_i] = P[X_i \in (a, b)] \cdot 1 + P[X_i \notin (a, b)] \cdot 0 = p.$$ 

- Similarly, $V[Y_i] = p(1 - p)$

- The random variables $Y_1, Y_2, \ldots, Y_n$ are i.i.d., so we can apply the Law of Large Numbers to $\bar{Y}_n$.

Can use similar argument to recover the probability mass or density!
Appendix: Sum of Two Uniform R.V. (non-examinable)

Let $X$ and $Y$ be two independent random variables, both uniformly distributed on $[0, 1]$. How does the probability density of $X + Y$ look like?

Answer

Further, for $0 \leq a \leq 1$ we have $f_X(a - y) = 1$ and $f_X(a - y) = 0$ otherwise, and thus $f_{X+Y}(a) = \int_0^a dy = a$.

Similarly, for $1 < a < 2$, $f_{X+Y}(a) = \int_{2-a}^2 dy = 2 - a$.

Therefore, $f_{X+Y}(a) = \begin{cases} a & \text{if } 0 \leq a \leq 1, \\ 2 - a & \text{if } 1 \leq a \leq 2, \\ 0 & \text{otherwise}. \end{cases}$
Appendix: Sum of Two Uniform R.V. (non-examinable)

Example

Let $X$ and $Y$ be two independent random variables, both uniformly distributed on $[0, 1]$. How does the probability density of $X + Y$ look like?

We have

$$f_{X+Y}(a) \overset{(*)}{=} \int_{-\infty}^{+\infty} f_X(a - y)f_Y(y)dy,$$

where for $(*)$, see Chapter 6.3 in Ross (Chapter 11.2 in Dekking et al.). Since $f_Y(y) = 1$ if $0 \leq y \leq 1$ and $f_Y(y) = 0$ otherwise, we have

$$f_{X+Y}(a) = \int_{0}^{1} f_X(a - y)dy.$$

Further, for $0 \leq a \leq 1$ we have $f_X(a - y) = 1$ and $f_X(a - y) = 0$ otherwise, and thus

$$f_{X+Y}(a) = \int_{0}^{a} dy = a.$$

Similarly, for $1 < a < 2$, $f_{X+Y}(a) = \int_{a}^{2} dy = 2 - a$. Therefore,

$$f_{X+Y}(a) = \begin{cases} a & \text{if } 0 \leq a \leq 1, \\ 2 - a & \text{if } 1 \leq a \leq 2, \\ 0 & \text{otherwise.} \end{cases}$$