Class Lectures (for Chapter 6)

# Probability Theory framework

# Probability Theory framework

Kolmogorov in the 1930's placed probability theory on a firm mathematical basis;

# Probability Theory framework

Kolmogorov in the 1930's placed probability theory on a firm mathematical basis; namely using measure and integration theory.

# Probability Theory framework

Kolmogorov in the 1930's placed probability theory on a firm mathematical basis; namely using measure and integration theory.

**Definition**

A probability space is a measure space $(\Omega, \mathcal{M}, P)$ with $P(\Omega) = 1$.

# Probability Theory framework

Kolmogorov in the 1930's placed probability theory on a firm mathematical basis; namely using measure and integration theory.

**Definition**

A probability space is a measure space $(\Omega, \mathcal{M}, P)$ with $P(\Omega) = 1$.

$\Omega$ is the set of "outcomes" of some "random" experiment.

# Probability Theory framework

Kolmogorov in the 1930's placed probability theory on a firm mathematical basis; namely using measure and integration theory.

**Definition**

A probability space is a measure space $(\Omega, \mathcal{M}, P)$ with $P(\Omega) = 1$.

$\Omega$ is the set of "outcomes" of some "random" experiment.
$\mathcal{M}$ is the set of "events" to which we will assign a "probability".

# Probability Theory framework

Kolmogorov in the 1930's placed probability theory on a firm mathematical basis; namely using measure and integration theory.

**Definition**

A probability space is a measure space $(\Omega, \mathcal{M}, P)$ with $P(\Omega) = 1$.

$\Omega$ is the set of "outcomes" of some "random" experiment.
$\mathcal{M}$ is the set of "events" to which we will assign a "probability".
For $A \in \mathcal{M}$, $P(A)$ is the the "probability" that our "randomly chosen" $\omega \in \Omega$ falls in $A$.

# Probability Theory framework

Kolmogorov in the 1930's placed probability theory on a firm mathematical basis; namely using measure and integration theory.

**Definition**

A probability space is a measure space $(\Omega, \mathcal{M}, P)$ with $P(\Omega) = 1$.

$\Omega$ is the set of "outcomes" of some "random" experiment.
$\mathcal{M}$ is the set of "events" to which we will assign a "probability".
For $A \in \mathcal{M}$, $P(A)$ is the the "probability" that our "randomly chosen" $\omega \in \Omega$ falls in $A$.

So, $(\Omega, \mathcal{M}, P)$ governs some "random experiment" where $P$ tells us the "likelihood" that $\omega$ (chosen "randomly") falls in different sets.

# Random variables

**Definition**

Given a probability space $(\Omega, \mathcal{M}, P)$, a **random variable** is a measurable function $X$ on $(\Omega, \mathcal{M}, P)$.

# Random variables

## Definition

Given a probability space $(\Omega, \mathcal{M}, P)$, a **random variable** is a measurable function $X$ on $(\Omega, \mathcal{M}, P)$.

So a random variable is not really random as it is just a function.

# Random variables

**Definition**

Given a probability space $(\Omega, \mathcal{M}, P)$, a **random variable** is a measurable function $X$ on $(\Omega, \mathcal{M}, P)$.

So a random variable is not really random as it is just a function. However, if $\omega$ is "random", then $X(\omega)$ is "random".

# Random variables

## Definition

Given a probability space $(\Omega, \mathcal{M}, P)$, a **random variable** is a measurable function $X$ on $(\Omega, \mathcal{M}, P)$.

So a random variable is not really random as it is just a function. However, if $\omega$ is "random", then $X(\omega)$ is "random". Hence we call it a random variable.

# Random variables

## Definition

Given a probability space $(\Omega, \mathcal{M}, P)$, a **random variable** is a measurable function $X$ on $(\Omega, \mathcal{M}, P)$.

So a random variable is not really random as it is just a function. However, if $\omega$ is "random", then $X(\omega)$ is "random". Hence we call it a random variable.

## Definition

If $X$ is a random variable on a probability space $(\Omega, \mathcal{M}, P)$, its **expectation**, denoted $E(X)$,

# Random variables

## Definition

Given a probability space $(\Omega, \mathcal{M}, P)$, a **random variable** is a measurable function $X$ on $(\Omega, \mathcal{M}, P)$.

So a random variable is not really random as it is just a function. However, if $\omega$ is "random", then $X(\omega)$ is "random". Hence we call it a random variable.

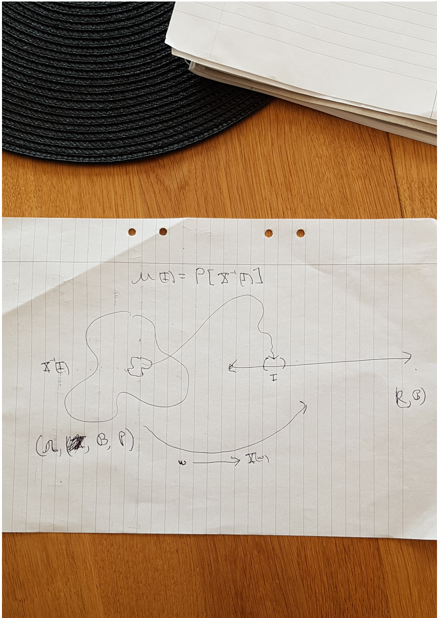
## Definition

If $X$ is a random variable on a probability space $(\Omega, \mathcal{M}, P)$, its **expectation**, denoted $E(X)$, is simply defined by

$$E(X) = \int X dP$$

provided this exists, meaning at least one of $\int X^+ dP$ and $\int X^- dP$ is finite.

# Law or distribution of a random variable



# Law of distribution of a random variable

## Definition

Given a random variable $X$ on a probability space $(\Omega, \mathcal{M}, P)$,

# Law of distribution of a random variable

## Definition

Given a random variable $X$ on a probability space $(\Omega, \mathcal{M}, P)$, the **distribution or law of X** (see picture)

# Law of distribution of a random variable

## Definition

Given a random variable $X$ on a probability space $(\Omega, \mathcal{M}, P)$, the **distribution or law of X** (see picture) is the probability measure $\mu_X$ on $(R, \mathcal{B})$ given by

# Law of distribution of a random variable

## Definition

Given a random variable $X$ on a probability space $(\Omega, \mathcal{M}, P)$, the **distribution or law of X** (see picture) is the probability measure $\mu_X$ on $(R, \mathcal{B})$ given by

$$\mu_X(A) := P(X^{-1}(A)).$$

# Law of distribution of a random variable

## Definition

Given a random variable $X$ on a probability space $(\Omega, \mathcal{M}, P)$, the **distribution or law of X** (see picture) is the probability measure $\mu_X$ on $(R, \mathcal{B})$ given by

$$\mu_X(A) := P(X^{-1}(A)).$$

Remarks: (i) One needs to check that $\mu_X$ is a probability measure.

# Law of distribution of a random variable

## Definition

Given a random variable $X$ on a probability space $(\Omega, \mathcal{M}, P)$, the **distribution or law of X** (see picture) is the probability measure $\mu_X$ on $(R, \mathcal{B})$ given by
$$\mu_X(A) := P(X^{-1}(A)).$$

Remarks: (i) One needs to check that $\mu_X$ is a probability measure.
(ii) The distribution of $X$ contains **all** the essential information of $X$.

# Law of distribution of a random variable

**Definition**

Given a random variable $X$ on a probability space $(\Omega, \mathcal{M}, P)$, the **distribution or law of X** (see picture) is the probability measure $\mu_X$ on $(R, \mathcal{B})$ given by

$$\mu_X(A) := P(X^{-1}(A)).$$

Remarks: (i) One needs to check that $\mu_X$ is a probability measure.
(ii) The distribution of $X$ contains **all** the essential information of $X$.
(iii) If someone says

# Law of distribution of a random variable

## Definition

Given a random variable $X$ on a probability space $(\Omega, \mathcal{M}, P)$, the **distribution or law of X** (see picture) is the probability measure $\mu_X$ on $(R, \mathcal{B})$ given by
$$\mu_X(A) := P(X^{-1}(A)).$$

Remarks: (i) One needs to check that $\mu_X$ is a probability measure.
(ii) The distribution of $X$ contains **all** the essential information of $X$.
(iii) If someone says
"Let $X$ be a Poisson random variable with parameter $\lambda$",

## Law of distribution of a random variable

### Definition

Given a random variable $X$ on a probability space $(\Omega, \mathcal{M}, P)$, the **distribution or law of X** (see picture) is the probability measure $\mu_X$ on $(R, \mathcal{B})$ given by

$$\mu_X(A) := P(X^{-1}(A)).$$

Remarks: (i) One needs to check that $\mu_X$ is a probability measure.
(ii) The distribution of $X$ contains **all** the essential information of $X$.
(iii) If someone says
"Let $X$ be a Poisson random variable with parameter $\lambda$",
what they mean is

# Law of distribution of a random variable

## Definition

Given a random variable $X$ on a probability space $(\Omega, \mathcal{M}, P)$, the **distribution or law of X** (see picture) is the probability measure $\mu_X$ on $(R, \mathcal{B})$ given by

$$\mu_X(A) := P(X^{-1}(A)).$$

Remarks: (i) One needs to check that $\mu_X$ is a probability measure.
(ii) The distribution of $X$ contains **all** the essential information of $X$.
(iii) If someone says
"Let $X$ be a Poisson random variable with parameter $\lambda$",
what they mean is $X$ is a random variable on some probability space

# Law of distribution of a random variable

**Definition**

Given a random variable $X$ on a probability space $(\Omega, \mathcal{M}, P)$, the **distribution or law of X** (see picture) is the probability measure $\mu_X$ on $(R, \mathcal{B})$ given by

$$\mu_X(A) := P(X^{-1}(A)).$$

Remarks: (i) One needs to check that $\mu_X$ is a probability measure.
(ii) The distribution of $X$ contains **all** the essential information of $X$.
(iii) If someone says
"Let $X$ be a Poisson random variable with parameter $\lambda$",
what they mean is $X$ is a random variable on some probability space (which we often don't care about)

# Law of distribution of a random variable

## Definition

Given a random variable $X$ on a probability space $(\Omega, \mathcal{M}, P)$, the **distribution or law of X** (see picture) is the probability measure $\mu_X$ on $(R, \mathcal{B})$ given by

$$\mu_X(A) := P(X^{-1}(A)).$$

Remarks: (i) One needs to check that $\mu_X$ is a probability measure.
(ii) The distribution of $X$ contains **all** the essential information of $X$.
(iii) If someone says
"Let $X$ be a Poisson random variable with parameter $\lambda$",
what they mean is $X$ is a random variable on some probability space (which we often don't care about) and the law of $X$, $\mu_X$, satisfies

# Law of distribution of a random variable

## Definition

Given a random variable $X$ on a probability space $(\Omega, \mathcal{M}, P)$, the **distribution or law of X** (see picture) is the probability measure $\mu_X$ on $(R, \mathcal{B})$ given by

$$\mu_X(A) := P(X^{-1}(A)).$$

Remarks: (i) One needs to check that $\mu_X$ is a probability measure.
(ii) The distribution of $X$ contains **all** the essential information of $X$.
(iii) If someone says
"Let $X$ be a Poisson random variable with parameter $\lambda$",
what they mean is $X$ is a random variable on some probability space (which we often don't care about) and the law of $X$, $\mu_X$, satisfies

$$\mu_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

# Independence

# Independence

## Definition

$n$ random variables $X_1, X_2 \ldots, X_n$ on a probability space $(\Omega, \mathcal{M}, P)$ are called **independent**

# Independence

## Definition

$n$ random variables $X_1, X_2 \ldots, X_n$ on a probability space $(\Omega, \mathcal{M}, P)$ are called **independent** if for all Borel sets $B_1, B_2 \ldots, B_n$

$$P(\bigcap_{i=1}^{n} X_i^{-1}(B_i)) = \prod_{i=1}^{n} P(X_i^{-1}(B_i)).$$

# Independence

## Definition

$n$ random variables $X_1, X_2 \ldots, X_n$ on a probability space $(\Omega, \mathcal{M}, P)$ are called **independent** if for all Borel sets $B_1, B_2 \ldots, B_n$

$$P(\bigcap_{i=1}^{n} X_i^{-1}(B_i)) = \prod_{i=1}^{n} P(X_i^{-1}(B_i)).$$

## Definition

An infinite collection of random variables on a probability space $(\Omega, \mathcal{M}, P)$ is called **independent** if each finite collection is independent as above.

# A simple example of convergence in probability but not a.s.

# A simple example of convergence in probability but not a.s.

Step 1: (will use $1 - x \leq e^{-x}$ for all $x$)

## A simple example of convergence in probability but not a.s.

Step 1: (will use $1 - x \leq e^{-x}$ for all $x$)

### Theorem

*(Second Borel Cantelli Lemma)*

# A simple example of convergence in probability but not a.s.

Step 1: (will use $1 - x \leq e^{-x}$ for all $x$)

**Theorem**

*(Second Borel Cantelli Lemma)*
*Let $E_1, E_2, \ldots$ be a sequence of independent events in the probability space $(\Omega, \mathcal{M}, P)$.*

# A simple example of convergence in probability but not a.s.

Step 1: (will use $1 - x \leq e^{-x}$ for all $x$)

**Theorem**

*(Second Borel Cantelli Lemma)*
*Let $E_1, E_2, \ldots$ be a sequence of independent events in the probability space $(\Omega, \mathcal{M}, P)$. If $\sum_i P(E_i) = \infty$,*

## A simple example of convergence in probability but not a.s.

Step 1: (will use $1 - x \leq e^{-x}$ for all $x$)

### Theorem

*(Second Borel Cantelli Lemma)*
*Let $E_1, E_2, \ldots$ be a sequence of independent events in the probability space $(\Omega, \mathcal{M}, P)$. If $\sum_i P(E_i) = \infty$, then*

$$P(\limsup E_i) = 1.$$

## A simple example of convergence in probability but not a.s.

Step 1: (will use $1 - x \leq e^{-x}$ for all $x$)

### Theorem

*(Second Borel Cantelli Lemma)*
*Let $E_1, E_2, \ldots$ be a sequence of independent events in the probability space $(\Omega, \mathcal{M}, P)$. If $\sum_i P(E_i) = \infty$, then*

$$P(\limsup E_i) = 1.$$

Proof: $(\limsup E_i)^c := \bigcup_{n=1}^{\infty} (\bigcap_{k=n}^{\infty} E_k^c)$.

## A simple example of convergence in probability but not a.s.

Step 1: (will use $1 - x \le e^{-x}$ for all $x$)

### Theorem

*(Second Borel Cantelli Lemma)*
*Let $E_1, E_2, \ldots$ be a sequence of independent events in the probability space $(\Omega, \mathcal{M}, P)$. If $\sum_i P(E_i) = \infty$, then*

$$P(\limsup E_i) = 1.$$

Proof: $(\limsup E_i)^c := \bigcup_{n=1}^{\infty} (\bigcap_{k=n}^{\infty} E_k^c)$. For each $n$,

# A simple example of convergence in probability but not a.s.

Step 1: (will use $1 - x \leq e^{-x}$ for all $x$)

## Theorem

*(Second Borel Cantelli Lemma)*
*Let* $E_1, E_2, \ldots$ *be a sequence of independent events in the probability space* $(\Omega, \mathcal{M}, P)$*. If* $\sum_i P(E_i) = \infty$*, then*

$$P(\limsup E_i) = 1.$$

Proof: $(\limsup E_i)^c := \bigcup_{n=1}^{\infty}(\bigcap_{k=n}^{\infty} E_k^c)$. For each $n$,

$$P(\bigcap_{k=n}^{\infty} E_k^c)$$

# A simple example of convergence in probability but not a.s.

Step 1: (will use $1 - x \leq e^{-x}$ for all $x$)

## Theorem

*(Second Borel Cantelli Lemma)*
*Let $E_1, E_2, \ldots$ be a sequence of independent events in the probability space $(\Omega, \mathcal{M}, P)$. If $\sum_i P(E_i) = \infty$, then*

$$P(\limsup E_i) = 1.$$

Proof: $(\limsup E_i)^c := \bigcup_{n=1}^{\infty}(\bigcap_{k=n}^{\infty} E_k^c)$. For each $n$,

$$P(\bigcap_{k=n}^{\infty} E_k^c) = \prod_{k=n}^{\infty} P(E_k^c) =$$

## A simple example of convergence in probability but not a.s.

Step 1: (will use $1 - x \leq e^{-x}$ for all $x$)

### Theorem

*(Second Borel Cantelli Lemma)*
*Let $E_1, E_2, \ldots$ be a sequence of independent events in the probability space $(\Omega, \mathcal{M}, P)$. If $\sum_i P(E_i) = \infty$, then*

$$P(\limsup E_i) = 1.$$

Proof: $(\limsup E_i)^c := \bigcup_{n=1}^{\infty}(\bigcap_{k=n}^{\infty} E_k^c)$. For each $n$,

$$P(\bigcap_{k=n}^{\infty} E_k^c) = \prod_{k=n}^{\infty} P(E_k^c) = \prod_{k=n}^{\infty}(1 - P(E_k))$$

# A simple example of convergence in probability but not a.s.

Step 1: (will use $1 - x \le e^{-x}$ for all $x$)

## Theorem

*(Second Borel Cantelli Lemma)*
*Let $E_1, E_2, \dots$ be a sequence of independent events in the probability space $(\Omega, \mathcal{M}, P)$. If $\sum_i P(E_i) = \infty$, then*

$$P(\limsup E_i) = 1.$$

Proof: $(\limsup E_i)^c := \bigcup_{n=1}^{\infty}(\bigcap_{k=n}^{\infty} E_k^c)$. For each $n$,

$$P(\bigcap_{k=n}^{\infty} E_k^c) = \prod_{k=n}^{\infty} P(E_k^c) = \prod_{k=n}^{\infty}(1 - P(E_k)) \le \prod_{k=n}^{\infty} e^{-P(E_k)}$$

# A simple example of convergence in probability but not a.s.

Step 1: (will use $1 - x \leq e^{-x}$ for all $x$)

**Theorem**

*(Second Borel Cantelli Lemma)*
*Let $E_1, E_2, \ldots$ be a sequence of independent events in the probability space $(\Omega, \mathcal{M}, P)$. If $\sum_i P(E_i) = \infty$, then*

$$P(\limsup E_i) = 1.$$

Proof: $(\limsup E_i)^c := \bigcup_{n=1}^{\infty}(\bigcap_{k=n}^{\infty} E_k^c)$. For each $n$,

$$P(\bigcap_{k=n}^{\infty} E_k^c) = \prod_{k=n}^{\infty} P(E_k^c) = \prod_{k=n}^{\infty}(1-P(E_k)) \leq \prod_{k=n}^{\infty} e^{-P(E_k)} = e^{-\sum_{k=n}^{\infty} P(E_k)} =$$

# A simple example of convergence in probability but not a.s.

Step 1: (will use $1 - x \leq e^{-x}$ for all $x$)

## Theorem

*(Second Borel Cantelli Lemma)*
*Let $E_1, E_2, \ldots$ be a sequence of independent events in the probability space $(\Omega, \mathcal{M}, P)$. If $\sum_i P(E_i) = \infty$, then*

$$P(\limsup E_i) = 1.$$

Proof: $(\limsup E_i)^c := \bigcup_{n=1}^{\infty}(\bigcap_{k=n}^{\infty} E_k^c)$. For each $n$,

$$P(\bigcap_{k=n}^{\infty} E_k^c) = \prod_{k=n}^{\infty} P(E_k^c) = \prod_{k=n}^{\infty}(1-P(E_k)) \leq \prod_{k=n}^{\infty} e^{-P(E_k)} = e^{-\sum_{k=n}^{\infty} P(E_k)} = 0$$

QED

# A simple example of convergence in probability but not a.s.

Step 1: (will use $1-x \leq e^{-x}$ for all $x$)

## Theorem

*(Second Borel Cantelli Lemma)*
*Let $E_1, E_2, \ldots$ be a sequence of independent events in the probability space $(\Omega, \mathcal{M}, P)$. If $\sum_i P(E_i) = \infty$, then*

$$P(\limsup E_i) = 1.$$

Proof: $(\limsup E_i)^c := \bigcup_{n=1}^{\infty}(\bigcap_{k=n}^{\infty} E_k^c)$. For each $n$,

$$P(\bigcap_{k=n}^{\infty} E_k^c) = \prod_{k=n}^{\infty} P(E_k^c) = \prod_{k=n}^{\infty}(1-P(E_k)) \leq \prod_{k=n}^{\infty} e^{-P(E_k)} = e^{-\sum_{k=n}^{\infty} P(E_k)} = 0$$

QED For the unit interval with Lebesgue measure, let $E_n = [0, 1/n]$, what is happening?

# Two examples

# Two examples

Example 1: Let $E_1, E_2, \ldots$ be a sequence of independent events in some probability space $(\Omega, \mathcal{M}, P)$ with $P(E_n) = 1/n^2$.

# Two examples

Example 1: Let $E_1, E_2, \ldots$ be a sequence of independent events in some probability space $(\Omega, \mathcal{M}, P)$ with $P(E_n) = 1/n^2$. Then the sequence of functions $I_{E_n}$ converges to 0 a.s.

# Two examples

Example 1: Let $E_1, E_2, \ldots$ be a sequence of independent events in some probability space $(\Omega, \mathcal{M}, P)$ with $P(E_n) = 1/n^2$. Then the sequence of functions $I_{E_n}$ converges to 0 a.s. This follows immediately from the Borel-Cantelli Lemma.

# Two examples

Example 1: Let $E_1, E_2, \ldots$ be a sequence of independent events in some probability space $(\Omega, \mathcal{M}, P)$ with $P(E_n) = 1/n^2$. Then the sequence of functions $I_{E_n}$ converges to 0 a.s. This follows immediately from the Borel-Cantelli Lemma.

Example 2: Let $E_1, E_2, \ldots$ be a sequence of independent events in some probability space $(\Omega, \mathcal{M}, P)$ with $P(E_n) = 1/n$.

# Two examples

Example 1: Let $E_1, E_2, \dots$ be a sequence of independent events in some probability space $(\Omega, \mathcal{M}, P)$ with $P(E_n) = 1/n^2$. Then the sequence of functions $I_{E_n}$ converges to 0 a.s. This follows immediately from the Borel-Cantelli Lemma.

Example 2: Let $E_1, E_2, \dots$ be a sequence of independent events in some probability space $(\Omega, \mathcal{M}, P)$ with $P(E_n) = 1/n$. Then the sequence of functions $I_{E_n}$ converges to 0 in measure but not a.s.

# Two examples

Example 1: Let $E_1, E_2, \ldots$ be a sequence of independent events in some probability space $(\Omega, \mathcal{M}, P)$ with $P(E_n) = 1/n^2$. Then the sequence of functions $I_{E_n}$ converges to 0 a.s. This follows immediately from the Borel-Cantelli Lemma.

Example 2: Let $E_1, E_2, \ldots$ be a sequence of independent events in some probability space $(\Omega, \mathcal{M}, P)$ with $P(E_n) = 1/n$. Then the sequence of functions $I_{E_n}$ converges to 0 in measure but not a.s.

Why? $P(|I_{E_n} - 0| \geq \epsilon) = P(E_n)$ which goes to 0. Hence convergence in measure. However, the second Borel-Cantelli Lemma says that $P(\limsup E_i) = 1$ and this means that it is not the case that $I_{E_n}$ converges to 0 a.s.. In fact it says that $I_{E_n}$ converges to 0 only on a set of probability 0.

# Two examples

Example 1: Let $E_1, E_2, \dots$ be a sequence of independent events in some probability space $(\Omega, \mathcal{M}, P)$ with $P(E_n) = 1/n^2$. Then the sequence of functions $I_{E_n}$ converges to 0 a.s. This follows immediately from the Borel-Cantelli Lemma.

Example 2: Let $E_1, E_2, \dots$ be a sequence of independent events in some probability space $(\Omega, \mathcal{M}, P)$ with $P(E_n) = 1/n$. Then the sequence of functions $I_{E_n}$ converges to 0 in measure but not a.s.

Why? $P(|I_{E_n} - 0| \geq \epsilon) = P(E_n)$ which goes to 0. Hence convergence in measure. However, the second Borel-Cantelli Lemma says that $P(\limsup E_i) = 1$ and this means that it is not the case that $I_{E_n}$ converges to 0 a.s.. In fact it says that $I_{E_n}$ converges to 0 only on a set of probability 0.

One should contemplate what is happening. For very large $n$, it is very unlikely that $E_n$ occurs.

## Two examples

Example 1: Let $E_1, E_2, \ldots$ be a sequence of independent events in some probability space $(\Omega, \mathcal{M}, P)$ with $P(E_n) = 1/n^2$. Then the sequence of functions $I_{E_n}$ converges to 0 a.s. This follows immediately from the Borel-Cantelli Lemma.

Example 2: Let $E_1, E_2, \ldots$ be a sequence of independent events in some probability space $(\Omega, \mathcal{M}, P)$ with $P(E_n) = 1/n$. Then the sequence of functions $I_{E_n}$ converges to 0 in measure but not a.s.

Why? $P(|I_{E_n} - 0| \geq \epsilon) = P(E_n)$ which goes to 0. Hence convergence in measure. However, the second Borel-Cantelli Lemma says that $P(\limsup E_i) = 1$ and this means that it is not the case that $I_{E_n}$ converges to 0 a.s.. In fact it says that $I_{E_n}$ converges to 0 only on a set of probability 0.

One should contemplate what is happening. For very large $n$, it is very unlikely that $E_n$ occurs. But nonetheless, if you watch things in time (as $n$ moves), $I_{E_n}$ will a.s. pop up to be 1 infinitely often but more and more rarely.

# Law of Large Numbers: Special case

# Law of Large Numbers: Special case

## Theorem

*(Weak and Strong Law of Large Numbers in a SPECIAL case)*

# Law of Large Numbers: Special case

**Theorem**

*(Weak and Strong Law of Large Numbers in a SPECIAL case)*
*Let* $X_1, X_2, \ldots$ *be independent random variables*

# Law of Large Numbers: Special case

## Theorem

*(Weak and Strong Law of Large Numbers in a SPECIAL case)*
*Let $X_1, X_2, \ldots$ be independent random variables each with distribution $(\delta_1 + \delta_{-1})/2$;*

# Law of Large Numbers: Special case

## Theorem

*(Weak and Strong Law of Large Numbers in a SPECIAL case)*
*Let* $X_1, X_2, \ldots$ *be independent random variables each with distribution* $(\delta_1 + \delta_{-1})/2$*; i.e.* $P(\{\omega : X_i(\omega) = 1\}) = P(\{\omega : X_i(\omega) = -1\}) = 1/2$.

# Law of Large Numbers: Special case

## Theorem

*(Weak and Strong Law of Large Numbers in a SPECIAL case)*
*Let* $X_1, X_2, \ldots$ *be independent random variables each with distribution* $(\delta_1 + \delta_{-1})/2$*; i.e.* $P(\{\omega : X_i(\omega) = 1\}) = P(\{\omega : X_i(\omega) = -1\}) = 1/2$.
*Then (WLLN)*

# Law of Large Numbers: Special case

## Theorem

*(Weak and Strong Law of Large Numbers in a SPECIAL case)*
*Let* $X_1, X_2, \ldots$ *be independent random variables each with distribution* $(\delta_1 + \delta_{-1})/2$; *i.e.* $P(\{\omega : X_i(\omega) = 1\}) = P(\{\omega : X_i(\omega) = -1\}) = 1/2$.
*Then (WLLN)*
*(i).*

$$\frac{S_n}{n} := \frac{\sum_{i=1}^{n} X_i}{n} \text{ converges in measure (in probability) } \to 0.$$

# Law of Large Numbers: Special case

## Theorem

*(Weak and Strong Law of Large Numbers in a SPECIAL case)*
*Let $X_1, X_2, \dots$ be independent random variables each with distribution $(\delta_1 + \delta_{-1})/2$; i.e. $P(\{\omega : X_i(\omega) = 1\}) = P(\{\omega : X_i(\omega) = -1\}) = 1/2$.*
*Then (WLLN)*
*(i).*

$$\frac{S_n}{n} := \frac{\sum_{i=1}^{n} X_i}{n} \text{ converges in measure (in probability) } \to 0.$$

*and (SLLN)*

# Law of Large Numbers: Special case

## Theorem

*(Weak and Strong Law of Large Numbers in a SPECIAL case)*
*Let* $X_1, X_2, \ldots$ *be independent random variables each with distribution* $(\delta_1 + \delta_{-1})/2$*; i.e.* $P(\{\omega : X_i(\omega) = 1\}) = P(\{\omega : X_i(\omega) = -1\}) = 1/2$.
*Then (WLLN)*
*(i).*

$$\frac{S_n}{n} := \frac{\sum_{i=1}^n X_i}{n} \text{ converges in measure (in probability) } \to 0.$$

*and (SLLN)*
*(ii).*

$$\frac{S_n}{n} := \frac{\sum_{i=1}^n X_i}{n} \text{ converges almost everywhere (almost surely) } \to 0.$$

# Some remarks

Remarks:

# Some remarks

Remarks:

1. SLLN implies WLLN. Why?

# Some remarks

Remarks:

1. SLLN implies WLLN. Why?
2. WLLN easier to prove but holds a little more generally.

# Some remarks

Remarks:

1. SLLN implies WLLN. Why?
2. WLLN easier to prove but holds a little more generally.
3. WLLN could be formulated in the 19th century while the conceptual framework did not exist in the 19th century to state the SLLN.

# Some remarks

Remarks:

1. SLLN implies WLLN. Why?
2. WLLN easier to prove but holds a little more generally.
3. WLLN could be formulated in the 19th century while the conceptual framework did not exist in the 19th century to state the SLLN.
4. Which is more natural?

# Proof of the WLLN

Proof:
(i) We first compute $E(S_n^2)$.

# Proof of the WLLN

Proof:

(i) We first compute $E(S_n^2)$. We have

$$E(S_n^2) = E((\sum_{i=1}^{n} X_i)(\sum_{j=1}^{n} X_j)) = \sum_{i,j=1}^{n} E(X_i X_j) = n + \sum_{i,j=1, i\neq j}^{n} E(X_i X_j).$$

## Proof of the WLLN

Proof:

(i) We first compute $E(S_n^2)$. We have

$$E(S_n^2) = E((\sum_{i=1}^{n} X_i)(\sum_{j=1}^{n} X_j)) = \sum_{i,j=1}^{n} E(X_i X_j) = n + \sum_{i,j=1, i \neq j}^{n} E(X_i X_j).$$

Each $X_i$ has expectation 0

# Proof of the WLLN

Proof:
(i) We first compute $E(S_n^2)$. We have

$$E(S_n^2) = E((\sum_{i=1}^{n} X_i)(\sum_{j=1}^{n} X_j)) = \sum_{i,j=1}^{n} E(X_i X_j) = n + \sum_{i,j=1, i \neq j}^{n} E(X_i X_j).$$

Each $X_i$ has expectation 0 and due to independence $E(X_i X_j) = 0$ when $i \neq j$.

# Proof of the WLLN

Proof:
(i) We first compute $E(S_n^2)$. We have

$$E(S_n^2) = E((\sum_{i=1}^{n} X_i)(\sum_{j=1}^{n} X_j)) = \sum_{i,j=1}^{n} E(X_iX_j) = n + \sum_{i,j=1, i\neq j}^{n} E(X_iX_j).$$

Each $X_i$ has expectation 0 and due to independence $E(X_iX_j) = 0$ when $i \neq j$. So

$$E(S_n^2) = n$$

# Proof of the WLLN

Proof:

(i) We first compute $E(S_n^2)$. We have

$$E(S_n^2) = E((\sum_{i=1}^{n} X_i)(\sum_{j=1}^{n} X_j)) = \sum_{i,j=1}^{n} E(X_i X_j) = n + \sum_{i,j=1, i\neq j}^{n} E(X_i X_j).$$

Each $X_i$ has expectation 0 and due to independence $E(X_i X_j) = 0$ when $i \neq j$. So

$$E(S_n^2) = n$$

and hence, by linearity

$$E((S_n/n)^2) = 1/n.$$

## Proof of the WLLN

Proof:

(i) We first compute $E(S_n^2)$. We have

$$E(S_n^2) = E((\sum_{i=1}^{n} X_i)(\sum_{j=1}^{n} X_j)) = \sum_{i,j=1}^{n} E(X_i X_j) = n + \sum_{i,j=1, i\neq j}^{n} E(X_i X_j).$$

Each $X_i$ has expectation 0 and due to independence $E(X_i X_j) = 0$ when $i \neq j$. So

$$E(S_n^2) = n$$

and hence, by linearity

$$E((S_n/n)^2) = 1/n.$$

Finally, fixing $\epsilon > 0$, we have, using Markov's inequality

# Proof of the WLLN

Proof:

(i) We first compute $E(S_n^2)$. We have

$$E(S_n^2) = E((\sum_{i=1}^{n} X_i)(\sum_{j=1}^{n} X_j)) = \sum_{i,j=1}^{n} E(X_iX_j) = n + \sum_{i,j=1,i\neq j}^{n} E(X_iX_j).$$

Each $X_i$ has expectation 0 and due to independence $E(X_iX_j) = 0$ when $i \neq j$. So

$$E(S_n^2) = n$$

and hence, by linearity

$$E((S_n/n)^2) = 1/n.$$

Finally, fixing $\epsilon > 0$, we have, using Markov's inequality

$$P(|S_n/n| \geq \epsilon) = P((S_n/n)^2 \geq \epsilon^2) \leq \frac{E((S_n/n)^2)}{\epsilon^2} = \frac{1}{n\epsilon^2}.$$

# Proof of the SLLN

(ii)

$$\sum_n P(|S_n/n| \geq \epsilon)$$

# Proof of the SLLN

(ii)

$$\sum_n P(|S_n/n| \geq \epsilon) < \text{ or } = \infty?$$

# Proof of the SLLN

(ii)

$$\sum_n P(|S_n/n| \geq \epsilon) < \text{ or } = \infty?$$

The previous line does *not* tell us. But let's anyway prove (ii) **assuming** this converges.

# Proof of the SLLN

(ii)

$$\sum_n P(|S_n/n| \geq \epsilon) < \text{ or } = \infty?$$

The previous line does *not* tell us. But let's anyway prove (ii) **assuming** this converges.

The Borel-Cantelli Lemma tells us that for every $\epsilon > 0$,

# Proof of the SLLN

(ii)

$$\sum_n P(|S_n/n| \geq \epsilon) < \text{ or } = \infty?$$

The previous line does *not* tell us. But let's anyway prove (ii) **assuming** this converges.

The Borel-Cantelli Lemma tells us that for every $\epsilon > 0$,

$$P(|S_n/n| \geq \epsilon \text{ i.o.}) = 0.$$

# Proof of the SLLN

(ii)

$$\sum_n P(|S_n/n| \geq \epsilon) < \text{ or } = \infty?$$

The previous line does *not* tell us. But let's anyway prove (ii) **assuming** this converges.

The Borel-Cantelli Lemma tells us that for every $\epsilon > 0$,

$$P(|S_n/n| \geq \epsilon \text{ i.o.}) = 0.$$

Letting $A_k := (|S_n/n| \geq \frac{1}{k} \text{ i.o.})$,

# Proof of the SLLN

(ii)

$$\sum_n P(|S_n/n| \geq \epsilon) < \text{ or } = \infty?$$

The previous line does *not* tell us. But let's anyway prove (ii) **assuming** this converges.

The Borel-Cantelli Lemma tells us that for every $\epsilon > 0$,

$$P(|S_n/n| \geq \epsilon \text{ i.o.}) = 0.$$

Letting $A_k := (|S_n/n| \geq \frac{1}{k} \text{ i.o.})$, we have $P(A_k) = 0$

# Proof of the SLLN

(ii)

$$\sum_n P(|S_n/n| \geq \epsilon) < \text{ or } = \infty?$$

The previous line does *not* tell us. But let's anyway prove (ii) **assuming** this converges.

The Borel-Cantelli Lemma tells us that for every $\epsilon > 0$,

$$P(|S_n/n| \geq \epsilon \text{ i.o.}) = 0.$$

Letting $A_k := (|S_n/n| \geq \frac{1}{k} \text{ i.o.})$, we have $P(A_k) = 0$ and hence $P(\bigcup_k A_k) = 0$.

# Proof of the SLLN

(ii)

$$\sum_n P(|S_n/n| \geq \epsilon) < \text{ or } = \infty?$$

The previous line does *not* tell us. But let's anyway prove (ii) **assuming** this converges.

The Borel-Cantelli Lemma tells us that for every $\epsilon > 0$,

$$P(|S_n/n| \geq \epsilon \text{ i.o.}) = 0.$$

Letting $A_k := (|S_n/n| \geq \frac{1}{k} \text{ i.o.})$, we have $P(A_k) = 0$ and hence $P(\bigcup_k A_k) = 0$. This is the same as $P(\bigcap_k A_k^c) = 1$.

# Proof of the SLLN

(ii)

$$\sum_n P(|S_n/n| \geq \epsilon) < \text{ or } = \infty?$$

The previous line does *not* tell us. But let's anyway prove (ii) **assuming** this converges.

The Borel-Cantelli Lemma tells us that for every $\epsilon > 0$,

$$P(|S_n/n| \geq \epsilon \text{ i.o.}) = 0.$$

Letting $A_k := (|S_n/n| \geq \frac{1}{k} \text{ i.o.})$, we have $P(A_k) = 0$ and hence $P(\bigcup_k A_k) = 0$. This is the same as $P(\bigcap_k A_k^c) = 1$.
One observes that

$$\bigcap_k A_k^c = \{\frac{S_n}{n} \to 0\}.$$

## Proof of the SLLN

(ii)

$$\sum_n P(|S_n/n| \geq \epsilon) < \text{ or } = \infty?$$

The previous line does *not* tell us. But let's anyway prove (ii) **assuming** this converges.

The Borel-Cantelli Lemma tells us that for every $\epsilon > 0$,

$$P(|S_n/n| \geq \epsilon \text{ i.o.}) = 0.$$

Letting $A_k := (|S_n/n| \geq \frac{1}{k} \text{ i.o.})$, we have $P(A_k) = 0$ and hence $P(\bigcup_k A_k) = 0$. This is the same as $P(\bigcap_k A_k^c) = 1$.
One observes that

$$\bigcap_k A_k^c = \{\frac{S_n}{n} \to 0\}.$$

Hence $P(S_n/n \to 0) = 1$.

# Proof of the SLLN

# Proof of the SLLN

$$\sum_n P(|S_n/n| \geq \epsilon) < \infty.$$

# Proof of the SLLN

$$\sum_n P(|S_n/n| \geq \epsilon) < \infty.$$

We need to use 4th moments

## Proof of the SLLN

$$\sum_n P(|S_n/n| \geq \epsilon) < \infty.$$

We need to use 4th moments and will prove afterwards

$$E(S_n^4) \leq 3n^2.$$

# Proof of the SLLN

$$\sum_n P(|S_n/n| \geq \epsilon) < \infty.$$

We need to use 4th moments and will prove afterwards

$$E(S_n^4) \leq 3n^2.$$

This gives $E((S_n/n)^4) \leq 3/n^2$.

# Proof of the SLLN

$$\sum_n P(|S_n/n| \geq \epsilon) < \infty.$$

We need to use 4th moments and will prove afterwards

$$E(S_n^4) \leq 3n^2.$$

This gives $E((S_n/n)^4) \leq 3/n^2$. Assuming this, we fix $\epsilon > 0$

# Proof of the SLLN

$$\sum_n P(|S_n/n| \geq \epsilon) < \infty.$$

We need to use 4th moments and will prove afterwards

$$E(S_n^4) \leq 3n^2.$$

This gives $E((S_n/n)^4) \leq 3/n^2$. Assuming this, we fix $\epsilon > 0$ and we obtain

$$P(|S_n/n| \geq \epsilon)$$

# Proof of the SLLN

$$\sum_n P(|S_n/n| \geq \epsilon) < \infty.$$

We need to use 4th moments and will prove afterwards

$$E(S_n^4) \leq 3n^2.$$

This gives $E((S_n/n)^4) \leq 3/n^2$. Assuming this, we fix $\epsilon > 0$ and we obtain

$$P(|S_n/n| \geq \epsilon) = P((S_n/n)^4 \geq \epsilon^4)$$

# Proof of the SLLN

$$\sum_n P(|S_n/n| \geq \epsilon) < \infty.$$

We need to use 4th moments and will prove afterwards

$$E(S_n^4) \leq 3n^2.$$

This gives $E((S_n/n)^4) \leq 3/n^2$. Assuming this, we fix $\epsilon > 0$ and we obtain

$$P(|S_n/n| \geq \epsilon) = P((S_n/n)^4 \geq \epsilon^4) \leq \frac{E((S_n/n)^4)}{\epsilon^4}$$

# Proof of the SLLN

$$\sum_n P(|S_n/n| \geq \epsilon) < \infty.$$

We need to use 4th moments and will prove afterwards

$$E(S_n^4) \leq 3n^2.$$

This gives $E((S_n/n)^4) \leq 3/n^2$. Assuming this, we fix $\epsilon > 0$ and we obtain

$$P(|S_n/n| \geq \epsilon) = P((S_n/n)^4 \geq \epsilon^4) \leq \frac{E((S_n/n)^4)}{\epsilon^4} \leq \frac{3}{n^2\epsilon^4}.$$

# Proof of the SLLN

# Proof of the SLLN

$$E(S_n^4) \leq 3n^2$$

# Proof of the SLLN

$$E(S_n^4) \leq 3n^2$$

$$E(S_n^4) = \sum_{i,j,k,\ell} E(X_i X_j X_k X_\ell).$$

# Proof of the SLLN

$$E(S_n^4) \le 3n^2$$

$$E(S_n^4) = \sum_{i,j,k,\ell} E(X_i X_j X_k X_\ell).$$

We break the index set into three groups,

# Proof of the SLLN

$$E(S_n^4) \leq 3n^2$$

$$E(S_n^4) = \sum_{i,j,k,\ell} E(X_i X_j X_k X_\ell).$$

We break the index set into three groups, (a) $i = j = k = \ell$,

# Proof of the SLLN

$$E(S_n^4) \leq 3n^2$$

$$E(S_n^4) = \sum_{i,j,k,\ell} E(X_i X_j X_k X_\ell).$$

We break the index set into three groups, (a) $i = j = k = \ell$, (b) two of $i, j, k, \ell$ take one value and two take another value

# Proof of the SLLN

$$E(S_n^4) \leq 3n^2$$

$$E(S_n^4) = \sum_{i,j,k,\ell} E(X_i X_j X_k X_\ell).$$

We break the index set into three groups, (a) $i = j = k = \ell$, (b) two of $i, j, k, \ell$ take one value and two take another value and (c) all other possibilities.

Terms in (a) or (b), which are $E(X_i^4)$ and $E(X_i^2 X_j^2)$ are 1.

# Proof of the SLLN

$$E(S_n^4) \leq 3n^2$$

$$E(S_n^4) = \sum_{i,j,k,\ell} E(X_i X_j X_k X_\ell).$$

We break the index set into three groups, (a) $i = j = k = \ell$, (b) two of $i, j, k, \ell$ take one value and two take another value and (c) all other possibilities.

Terms in (a) or (b), which are $E(X_i^4)$ and $E(X_i^2 X_j^2)$ are 1. All terms of type (c) are zero after checking.

# Proof of the SLLN

$$E(S_n^4) \le 3n^2$$

$$E(S_n^4) = \sum_{i,j,k,\ell} E(X_i X_j X_k X_\ell).$$

We break the index set into three groups, (a) $i = j = k = \ell$, (b) two of $i, j, k, \ell$ take one value and two take another value and (c) all other possibilities.

Terms in (a) or (b), which are $E(X_i^4)$ and $E(X_i^2 X_j^2)$ are 1. All terms of type (c) are zero after checking.

The number of terms of type (b) is $n(n-1)3$ (elementary combinatorics).

# Proof of the SLLN

$$E(S_n^4) \leq 3n^2$$

$$E(S_n^4) = \sum_{i,j,k,\ell} E(X_i X_j X_k X_\ell).$$

We break the index set into three groups, (a) $i = j = k = \ell$, (b) two of $i, j, k, \ell$ take one value and two take another value and (c) all other possibilities.

Terms in (a) or (b), which are $E(X_i^4)$ and $E(X_i^2 X_j^2)$ are 1. All terms of type (c) are zero after checking.

The number of terms of type (b) is $n(n-1)3$ (elementary combinatorics). Hence $E(S_n^4)$ is $n + n(n-1)3 \leq 3n^2$. QED

# General SLLN

**Theorem**

*(Strong Law of Large Numbers: General case) Let* $X_1, X_2, \ldots$ *be independent random variables with the same distribution with* $E(|X|) < \infty$*. Then*

$$\frac{S_n}{n} := \frac{\sum_{i=1}^{n} X_i}{n} \text{ converges a.e. to } E(X).$$

# WLLN without an expectation

# WLLN without an expectation

The WLLN holds under slightly weaker assumptions.

# WLLN without an expectation

The WLLN holds under slightly weaker assumptions.

For example if the probability density function for $X$ is

$$f(x) := \frac{c}{x^2 \log(|x|)} I_{|x| \geq 2}$$

# WLLN without an expectation

The WLLN holds under slightly weaker assumptions.

For example if the probability density function for $X$ is

$$f(x) := \frac{c}{x^2 \log(|x|)} I_{|x| \geq 2}$$

then $E(|X|) = \infty$

# WLLN without an expectation

The WLLN holds under slightly weaker assumptions.

For example if the probability density function for $X$ is

$$f(x) := \frac{c}{x^2 \log(|x|)} I_{|x| \geq 2}$$

then $E(|X|) = \infty$ and

$$\frac{S_n}{n} \to 0 \text{ in measure but not a.e.}$$

# WLLN without an expectation

The WLLN holds under slightly weaker assumptions.

For example if the probability density function for $X$ is

$$f(x) := \frac{c}{x^2 \log(|x|)} I_{|x| \geq 2}$$

then $E(|X|) = \infty$ and

$$\frac{S_n}{n} \to 0 \text{ in measure but not a.e.}$$

What is happening? How could this be occuring?

# WLLN without an expectation

The WLLN holds under slightly weaker assumptions.

For example if the probability density function for $X$ is

$$f(x) := \frac{c}{x^2 \log(|x|)} I_{|x| \geq 2}$$

then $E(|X|) = \infty$ and

$$\frac{S_n}{n} \to 0 \text{ in measure but not a.e.}$$

What is happening? How could this be occuring?

For very large $n$, $\frac{S_n}{n}$ is very likely to be close to 0,

## WLLN without an expectation

The WLLN holds under slightly weaker assumptions.

For example if the probability density function for $X$ is

$$f(x) := \frac{c}{x^2 \log(|x|)} I_{|x| \geq 2}$$

then $E(|X|) = \infty$ and

$$\frac{S_n}{n} \to 0 \text{ in measure but not a.e.}$$

What is happening? How could this be occuring?

For very large $n$, $\frac{S_n}{n}$ is very likely to be close to 0, but if you watch the trajectory in time, there will be these very rare times at which $\frac{S_n}{n}$ is close to $\infty$ and times close to $-\infty$.

# Infinite number of independent random variables?

# Infinite number of independent random variables?

How do we know that we can have a probability space with an *infinite* number of independent random variables.

# Infinite number of independent random variables?

How do we know that we can have a probability space with an *infinite* number of independent random variables. There are two approaches.

# Infinite number of independent random variables?

How do we know that we can have a probability space with an *infinite* number of independent random variables. There are two approaches.

Approach 1: Constructing an infinite product space (see notes).

# Infinite number of independent random variables?

How do we know that we can have a probability space with an *infinite* number of independent random variables. There are two approaches.

Approach 1: Constructing an infinite product space (see notes).

Approach 2: Use $([0,1], \mathcal{B}_{[0,1]}, m)$ where $m$ is Lebesgue measure as our probability space.

# Infinite number of independent random variables?

How do we know that we can have a probability space with an *infinite* number of independent random variables. There are two approaches.

Approach 1: Constructing an infinite product space (see notes).

Approach 2: Use $([0,1], \mathcal{B}_{[0,1]}, m)$ where $m$ is Lebesgue measure as our probability space.
Given $x \in [0,1]$, $x$ has a binary expansion

$$x = \sum_{n=1}^{\infty} \frac{a_n(x)}{2^n}$$

where each $a_n(x) \in \{0,1\}$.

# Infinite number of independent random variables?

How do we know that we can have a probability space with an *infinite* number of independent random variables. There are two approaches.

Approach 1: Constructing an infinite product space (see notes).

Approach 2: Use $([0,1], \mathcal{B}_{[0,1]}, m)$ where $m$ is Lebesgue measure as our probability space.
Given $x \in [0,1]$, $x$ has a binary expansion

$$x = \sum_{n=1}^{\infty} \frac{a_n(x)}{2^n}$$

where each $a_n(x) \in \{0,1\}$. (Nonuniqueness only occurs at countably many $x$'s and so can ignore.)

## Infinite number of independent random variables?

How do we know that we can have a probability space with an *infinite* number of independent random variables. There are two approaches.

Approach 1: Constructing an infinite product space (see notes).

Approach 2: Use $([0,1], \mathcal{B}_{[0,1]}, m)$ where $m$ is Lebesgue measure as our probability space.
Given $x \in [0,1]$, $x$ has a binary expansion

$$x = \sum_{n=1}^{\infty} \frac{a_n(x)}{2^n}$$

where each $a_n(x) \in \{0,1\}$. (Nonuniqueness only occurs at countably many $x$'s and so can ignore.) Now, for each $n \geq 1$, define the random variable

$$X_n(x)$$

# Infinite number of independent random variables?

How do we know that we can have a probability space with an *infinite* number of independent random variables. There are two approaches.

Approach 1: Constructing an infinite product space (see notes).

Approach 2: Use $([0,1], \mathcal{B}_{[0,1]}, m)$ where $m$ is Lebesgue measure as our probability space.

Given $x \in [0,1]$, $x$ has a binary expansion

$$x = \sum_{n=1}^{\infty} \frac{a_n(x)}{2^n}$$

where each $a_n(x) \in \{0,1\}$. (Nonuniqueness only occurs at countably many $x$'s and so can ignore.) Now, for each $n \geq 1$, define the random variable

$$X_n(x) = 1 \text{ if } a_n(x) = 1 \text{ and } -1 \text{ if } a_n(x) = 0.$$

## Infinite number of independent random variables?

How do we know that we can have a probability space with an *infinite* number of independent random variables. There are two approaches.

Approach 1: Constructing an infinite product space (see notes).

Approach 2: Use $([0,1], \mathcal{B}_{[0,1]}, m)$ where $m$ is Lebesgue measure as our probability space.
Given $x \in [0,1]$, $x$ has a binary expansion

$$x = \sum_{n=1}^{\infty} \frac{a_n(x)}{2^n}$$

where each $a_n(x) \in \{0,1\}$. (Nonuniqueness only occurs at countably many $x$'s and so can ignore.) Now, for each $n \geq 1$, define the random variable

$$X_n(x) = 1 \text{ if } a_n(x) = 1 \text{ and } -1 \text{ if } a_n(x) = 0.$$

One can show that $X_1, X_2, \ldots$ are independent and each has distribution $(\delta_1 + \delta_{-1})/2$.

# Random series: a fascinating aside

# Random series: a fascinating aside

Of course $\sum_n \frac{1}{n}$ diverges

# Random series: a fascinating aside

Of course $\sum_n \frac{1}{n}$ diverges while $\sum_n \frac{(-1)^n}{n}$ converges.

# Random series: a fascinating aside

Of course $\sum_n \frac{1}{n}$ diverges while $\sum_n \frac{(-1)^n}{n}$ converges.
What happens if we put a random sign in front of $\frac{1}{n}$?

# Random series: a fascinating aside

Of course $\sum_n \frac{1}{n}$ diverges while $\sum_n \frac{(-1)^n}{n}$ converges.
What happens if we put a random sign in front of $\frac{1}{n}$?

Let $\{X_n\}_{n\geq 1}$ be independent random variables with
$P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$ for each $n$.

# Random series: a fascinating aside

Of course $\sum_n \frac{1}{n}$ diverges while $\sum_n \frac{(-1)^n}{n}$ converges.
What happens if we put a random sign in front of $\frac{1}{n}$?

Let $\{X_n\}_{n\geq 1}$ be independent random variables with $P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$ for each $n$.

$$\text{Does } \sum_{n=1}^{\infty} \frac{X_n}{n} \text{ converge or not?}$$

# Random series: a fascinating aside

Of course $\sum_n \frac{1}{n}$ diverges while $\sum_n \frac{(-1)^n}{n}$ converges.
What happens if we put a random sign in front of $\frac{1}{n}$?

Let $\{X_n\}_{n\geq 1}$ be independent random variables with
$P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$ for each $n$.

$$\text{Does } \sum_{n=1}^{\infty} \frac{X_n}{n} \text{ converge or not?}$$

**Theorem**

$\sum_{n=1}^{\infty} \frac{X_n}{n}$ *converges a.e.*

## Random series: a fascinating aside

### Theorem

*(i) If* $\alpha > 1/2$*,*

# Random series: a fascinating aside

**Theorem**

*(i) If* $\alpha > 1/2$*, then* $\sum_{n=1}^{\infty} \frac{X_n}{n^{\alpha}}$ *converges a.e.*

# Random series: a fascinating aside

**Theorem**

*(i) If* $\alpha > 1/2$, *then* $\sum_{n=1}^{\infty} \frac{X_n}{n^{\alpha}}$ *converges a.e.*

*(ii) If* $\alpha \in (0, 1/2]$, *then* $\sum_{n=1}^{\infty} \frac{X_n}{n^{\alpha}}$ *diverges a.e..*

## Random series: a fascinating aside

### Theorem

*(i) If* $\alpha > 1/2$*, then* $\sum_{n=1}^{\infty} \frac{X_n}{n^\alpha}$ *converges a.e.*

*(ii) If* $\alpha \in (0, 1/2]$*, then* $\sum_{n=1}^{\infty} \frac{X_n}{n^\alpha}$ *diverges a.e..*
*More specifically, one has that a.e.,* $\limsup_{n\to\infty} \sum_{k=1}^{n} \frac{X_k}{k^\alpha} = \infty$ *and* $\liminf_{n\to\infty} \sum_{k=1}^{n} \frac{X_k}{k^\alpha} = -\infty$*.*

# Random series: a fascinating aside

## Theorem

*(i) If* $\alpha > 1/2$*, then* $\sum_{n=1}^{\infty} \frac{X_n}{n^\alpha}$ *converges a.e.*

*(ii) If* $\alpha \in (0, 1/2]$*, then* $\sum_{n=1}^{\infty} \frac{X_n}{n^\alpha}$ *diverges a.e..*
*More specifically, one has that a.e.,* $\limsup_{n\to\infty} \sum_{k=1}^{n} \frac{X_k}{k^\alpha} = \infty$ *and* $\liminf_{n\to\infty} \sum_{k=1}^{n} \frac{X_k}{k^\alpha} = -\infty$.

Explanation: The variance of $\sum_{k=1}^{n} \frac{X_k}{k^\alpha}$

# Random series: a fascinating aside

**Theorem**

*(i) If* $\alpha > 1/2$, *then* $\sum_{n=1}^{\infty} \frac{X_n}{n^\alpha}$ *converges a.e.*

*(ii) If* $\alpha \in (0, 1/2]$, *then* $\sum_{n=1}^{\infty} \frac{X_n}{n^\alpha}$ *diverges a.e..*
*More specifically, one has that a.e.,* $\limsup_{n\to\infty} \sum_{k=1}^{n} \frac{X_k}{k^\alpha} = \infty$ *and* $\liminf_{n\to\infty} \sum_{k=1}^{n} \frac{X_k}{k^\alpha} = -\infty$.

Explanation: The variance of $\sum_{k=1}^{n} \frac{X_k}{k^\alpha} = \sum_{k=1}^{n} \frac{1}{k^{2\alpha}}$ converges to $\infty$ if and only if $\theta \leq 1/2$.

# A few words about the variance

### Definition

If $X$ is a random variable on a probability space $(\Omega, \mathcal{M}, P)$ with finite expectation, then the **variance** of $X$, $Var(X)$,

# A few words about the variance

**Definition**

If $X$ is a random variable on a probability space $(\Omega, \mathcal{M}, P)$ with finite expectation, then the **variance** of $X$, $Var(X)$, is

$$Var(X) = \int (X - E(X))^2 \, dP.$$

# A few words about the variance

## Definition

If $X$ is a random variable on a probability space $(\Omega, \mathcal{M}, P)$ with finite expectation, then the **variance** of $X$, $Var(X)$, is

$$Var(X) = \int (X - E(X))^2 \, dP.$$

- Assuming $X$ has finite expectation, $Var(X) < \infty$ if and only if $X \in L^2(\Omega, \mathcal{M}, P)$

# A few words about the variance

### Definition

If $X$ is a random variable on a probability space $(\Omega, \mathcal{M}, P)$ with finite expectation, then the **variance** of $X$, $Var(X)$, is

$$Var(X) = \int (X - E(X))^2 \, dP.$$

- Assuming $X$ has finite expectation, $Var(X) < \infty$ if and only if $X \in L^2(\Omega, \mathcal{M}, P)$
- $Var(X) = E(X^2) - (E(X))^2$, which is something you might have seen, is actually the pythagorean theorem, viewed properly.

# A few words about the variance

In $R^n$, there is a dot product $x \cdot y := \sum_i x_i y_i$.

# A few words about the variance

In $R^n$, there is a dot product $x \cdot y := \sum_i x_i y_i$. This can be used to compute projections in order to find the closest point to a given point which sits in some plane.

# A few words about the variance

In $R^n$, there is a dot product $x \cdot y := \sum_i x_i y_i$. This can be used to compute projections in order to find the closest point to a given point which sits in some plane.

For $L^2(\Omega, \mathcal{M}, P)$, there is a similar dot product defined by $X \cdot Y := E(XY)$.

# A few words about the variance

In $R^n$, there is a dot product $x \cdot y := \sum_i x_i y_i$. This can be used to compute projections in order to find the closest point to a given point which sits in some plane.

For $L^2(\Omega, \mathcal{M}, P)$, there is a similar dot product defined by $X \cdot Y := E(XY)$. It satisfies all the usual properties that the dot product in $R^n$ satisfies.

# A few words about the variance

In $R^n$, there is a dot product $x \cdot y := \sum_i x_i y_i$. This can be used to compute projections in order to find the closest point to a given point which sits in some plane.

For $L^2(\Omega, \mathcal{M}, P)$, there is a similar dot product defined by $X \cdot Y := E(XY)$. It satisfies all the usual properties that the dot product in $R^n$ satisfies. The length of a random variable is defined to be, exactly as in $R^n$, $(X \cdot X)^{1/2}$ or $(E(X^2))^{1/2}$.

# A few words about the variance

In $R^n$, there is a dot product $x \cdot y := \sum_i x_i y_i$. This can be used to compute projections in order to find the closest point to a given point which sits in some plane.

For $L^2(\Omega, \mathcal{M}, P)$, there is a similar dot product defined by $X \cdot Y := E(XY)$. It satisfies all the usual properties that the dot product in $R^n$ satisfies. The length of a random variable is defined to be, exactly as in $R^n$, $(X \cdot X)^{1/2}$ or $(E(X^2))^{1/2}$. The distance between $X$ and $Y$ is the length of $X - Y$.

# A few words about the variance

One can consider the 1-dimensional space of constant random variables and ask which random variable in this subspace is closest to a given random variable $X$.

# A few words about the variance

One can consider the 1-dimensional space of constant random variables and ask which random variable in this subspace is closest to a given random variable $X$. It turns out that this is simply the constant random variable $E(X)$,

# A few words about the variance

One can consider the 1-dimensional space of constant random variables and ask which random variable in this subspace is closest to a given random variable $X$. It turns out that this is simply the constant random variable $E(X)$, i.e., $E(X)$ is the projection of $X$ onto the 1-dimensional space of constant random variables.

# A few words about the variance

One can consider the 1-dimensional space of constant random variables and ask which random variable in this subspace is closest to a given random variable $X$. It turns out that this is simply the constant random variable $E(X)$, i.e., $E(X)$ is the projection of $X$ onto the 1-dimensional space of constant random variables.

Hence $E(X)$ and $X - E(X)$ are orthogonal.

# A few words about the variance

One can consider the 1-dimensional space of constant random variables and ask which random variable in this subspace is closest to a given random variable $X$. It turns out that this is simply the constant random variable $E(X)$, i.e., $E(X)$ is the projection of $X$ onto the 1-dimensional space of constant random variables.

Hence $E(X)$ and $X - E(X)$ are orthogonal. The pythagorean theorem tells us that $E(X^2) = E(X - E(X))^2 + (E(X))^2 = Var(X) + (E(X))^2$.

# A few words about the variance

One can consider the 1-dimensional space of constant random variables and ask which random variable in this subspace is closest to a given random variable $X$. It turns out that this is simply the constant random variable $E(X)$, i.e., $E(X)$ is the projection of $X$ onto the 1-dimensional space of constant random variables.

Hence $E(X)$ and $X - E(X)$ are orthogonal. The pythagorean theorem tells us that $E(X^2) = E(X - E(X))^2 + (E(X))^2 = Var(X) + (E(X))^2$. So, the variance is the "squared distance from $X$ to its projection onto the 1-dimensional space of constant functions".

# A fun aside: the arc sign law for coin tossing

# A fun aside: the arc sign law for coin tossing

You play a game. Each minute you win or lose a dollar, each with probably $1/2$, independently each time.

# A fun aside: the arc sign law for coin tossing

You play a game. Each minute you win or lose a dollar, each with probably $1/2$, independently each time. $X_i$ is what you received at time $i$ (either 1 or -1)

## A fun aside: the arc sign law for coin tossing

You play a game. Each minute you win or lose a dollar, each with probably $1/2$, independently each time. $X_i$ is what you received at time $i$ (either 1 or -1) and $S_n := \sum_{i=1}^n X_i$ is your total winnings at time $n$.

# A fun aside: the arc sign law for coin tossing

You play a game. Each minute you win or lose a dollar, each with probably $1/2$, independently each time. $X_i$ is what you received at time $i$ (either 1 or -1) and $S_n := \sum_{i=1}^{n} X_i$ is your total winnings at time $n$.

If you played a very very large number $n$ of times and it turns out that $\frac{S_n}{n} < -.001$, would you have a right to say you have been very unlucky?

## A fun aside: the arc sign law for coin tossing

You play a game. Each minute you win or lose a dollar, each with probably $1/2$, independently each time. $X_i$ is what you received at time $i$ (either 1 or -1) and $S_n := \sum_{i=1}^n X_i$ is your total winnings at time $n$.

If you played a very very large number $n$ of times and it turns out that $\frac{S_n}{n} < -.001$, would you have a right to say you have been very unlucky?

Answer: yes.

# A fun aside: the arc sign law for coin tossing

You play a game. Each minute you win or lose a dollar, each with probably $1/2$, independently each time. $X_i$ is what you received at time $i$ (either 1 or -1) and $S_n := \sum_{i=1}^{n} X_i$ is your total winnings at time $n$.

If you played a very very large number $n$ of times and it turns out that $\frac{S_n}{n} < -.001$, would you have a right to say you have been very unlucky?

Answer: yes.

Since $\frac{S_n}{n}$ approaches 0 in probability (WLLN),

# A fun aside: the arc sign law for coin tossing

You play a game. Each minute you win or lose a dollar, each with probably $1/2$, independently each time. $X_i$ is what you received at time $i$ (either 1 or -1) and $S_n := \sum_{i=1}^{n} X_i$ is your total winnings at time $n$.

If you played a very very large number $n$ of times and it turns out that $\frac{S_n}{n} < -.001$, would you have a right to say you have been very unlucky?

Answer: yes.

Since $\frac{S_n}{n}$ approaches 0 in probability (WLLN),
$P(\frac{S_n}{n} < -.001) \leq P(|\frac{S_n}{n} - 0| \geq .001)$ which goes to 0 as $n \to \infty$.

# A fun aside: the arc sign law for coin tossing

# A fun aside: the arc sign law for coin tossing

Now, lets say you played a very large number of times and you noticed that you were behind most of the time.

# A fun aside: the arc sign law for coin tossing

Now, lets say you played a very large number of times and you noticed that you were behind most of the time. More specifically, you noticed that 90% of the time you were losing;

## A fun aside: the arc sign law for coin tossing

Now, lets say you played a very large number of times and you noticed that you were behind most of the time. More specifically, you noticed that 90% of the time you were losing; i.e.

$$|\{i \in \{1, 2, \ldots, n\} : S_i < 0\}| \geq .9n.$$

Could you claim you are unlucky?

## A fun aside: the arc sign law for coin tossing

Now, lets say you played a very large number of times and you noticed that you were behind most of the time. More specifically, you noticed that 90% of the time you were losing; i.e.

$$|\{i \in \{1, 2, \dots, n\} : S_i < 0\}| \geq .9n.$$

Could you claim you are unlucky?

After all, things should even out in the end and you should be leading about half the time.

## A fun aside: the arc sign law for coin tossing

Now, lets say you played a very large number of times and you noticed that you were behind most of the time. More specifically, you noticed that 90% of the time you were losing; i.e.

$$|\{i \in \{1, 2, \ldots, n\} : S_i < 0\}| \geq .9n.$$

Could you claim you are unlucky?

After all, things should even out in the end and you should be leading about half the time.

Being able to claim you are very unlucky should mean that

# A fun aside: the arc sign law for coin tossing

Now, lets say you played a very large number of times and you noticed that you were behind most of the time. More specifically, you noticed that 90% of the time you were losing; i.e.

$$|\{i \in \{1, 2, \dots, n\} : S_i < 0\}| \geq .9n.$$

Could you claim you are unlucky?

After all, things should even out in the end and you should be leading about half the time.

Being able to claim you are very unlucky should mean that

$$\lim_{n\to\infty} P(|\{i \in \{1, 2, \dots, n\} : S_i < 0\}| \geq .9n) = 0.$$

# A fun aside: the arc sign law for coin tossing

Now, lets say you played a very large number of times and you noticed that you were behind most of the time. More specifically, you noticed that 90% of the time you were losing; i.e.

$$|\{i \in \{1, 2, \ldots, n\} : S_i < 0\}| \geq .9n.$$

Could you claim you are unlucky?

After all, things should even out in the end and you should be leading about half the time.

Being able to claim you are very unlucky should mean that

$$\lim_{n \to \infty} P(|\{i \in \{1, 2, \ldots, n\} : S_i < 0\}| \geq .9n) = 0.$$

False:

# A fun aside: the arc sign law for coin tossing

Now, lets say you played a very large number of times and you noticed that you were behind most of the time. More specifically, you noticed that 90% of the time you were losing; i.e.

$$|\{i \in \{1, 2, \ldots, n\} : S_i < 0\}| \geq .9n.$$

Could you claim you are unlucky?

After all, things should even out in the end and you should be leading about half the time.

Being able to claim you are very unlucky should mean that

$$\lim_{n\to\infty} P(|\{i \in \{1, 2, \ldots, n\} : S_i < 0\}| \geq .9n) = 0.$$

False: the above limit is not zero

## A fun aside: the arc sign law for coin tossing

Now, lets say you played a very large number of times and you noticed that you were behind most of the time. More specifically, you noticed that 90% of the time you were losing; i.e.

$$|\{i \in \{1, 2, \ldots, n\} : S_i < 0\}| \geq .9n.$$

Could you claim you are unlucky?

After all, things should even out in the end and you should be leading about half the time.

Being able to claim you are very unlucky should mean that

$$\lim_{n\to\infty} P(|\{i \in \{1, 2, \ldots, n\} : S_i < 0\}| \geq .9n) = 0.$$

False: the above limit is not zero and rather equals

$$\frac{2}{\pi} \arcsin(\sqrt{.1})$$

# A fun aside: the arc sign law for coin tossing

# A fun aside: the arc sign law for coin tossing

Doesn't the Weak Law of Large Numbers say

$$\lim_{n\to\infty} P(|\{i \in \{1, 2, \ldots, n\} : S_i < 0\}| \geq .9n) = 0?$$

# A fun aside: the arc sign law for coin tossing

Doesn't the Weak Law of Large Numbers say

$$\lim_{n\to\infty} P(|\{i \in \{1, 2, \ldots, n\} : S_i < 0\}| \geq .9n) = 0?$$

Let $Y_i = 1$ if you are leading at time $i$ and $Y_i = 0$ if you are losing at time $i$.
The WLLN *should* say that

# A fun aside: the arc sign law for coin tossing

Doesn't the Weak Law of Large Numbers say

$$\lim_{n\to\infty} P(|\{i \in \{1, 2, \ldots, n\} : S_i < 0\}| \geq .9n) = 0?$$

Let $Y_i = 1$ if you are leading at time $i$ and $Y_i = 0$ if you are losing at time $i$.
The WLLN *should* say that

$$\frac{\sum_{i=1}^{n} Y_i}{n}$$

converges in probability to $1/2$

# A fun aside: the arc sign law for coin tossing

Doesn't the Weak Law of Large Numbers say

$$\lim_{n\to\infty} P(|\{i \in \{1,2,\ldots,n\} : S_i < 0\}| \geq .9n) = 0?$$

Let $Y_i = 1$ if you are leading at time $i$ and $Y_i = 0$ if you are losing at time $i$.

The WLLN *should* say that

$$\frac{\sum_{i=1}^{n} Y_i}{n}$$

converges in probability to $1/2$ and then the above limit should be 0.

# A fun aside: the arc sign law for coin tossing

Doesn't the Weak Law of Large Numbers say

$$\lim_{n\to\infty} P(|\{i \in \{1,2,\ldots,n\} : S_i < 0\}| \geq .9n) = 0?$$

Let $Y_i = 1$ if you are leading at time $i$ and $Y_i = 0$ if you are losing at time $i$.

The WLLN *should* say that

$$\frac{\sum_{i=1}^{n} Y_i}{n}$$

converges in probability to $1/2$ and then the above limit should be 0.

What's happening? The WLLN is **not** applicable since the $Y_i$'s are not independent.

# A fun aside: the arc sign law for coin tossing

Doesn't the Weak Law of Large Numbers say

$$\lim_{n\to\infty} P(|\{i \in \{1, 2, \dots, n\} : S_i < 0\}| \geq .9n) = 0?$$

Let $Y_i = 1$ if you are leading at time $i$ and $Y_i = 0$ if you are losing at time $i$.

The WLLN *should* say that

$$\frac{\sum_{i=1}^{n} Y_i}{n}$$

converges in probability to $1/2$ and then the above limit should be 0.

What's happening? The WLLN is **not** applicable since the $Y_i$'s are not independent. In fact, they are very correlated.