Advanced Algorithms – COMS31900

Probability recap.

Raphaël Clifford

Slides by Markus Jalsenius

Randomness and probability

Probability

The **sample space** S is the set of *outcomes* of an experiment.

Probability

The **sample space** S is the set of *outcomes* of an experiment.

EXAMPLES

Roll a die: $S = \{1, 2, 3, 4, 5, 6\}$.

Probability

The **sample space** S is the set of *outcomes* of an experiment.

EXAMPLES

Roll a die: $S = \{1, 2, 3, 4, 5, 6\}$.

Flip a coin: $S = \{\mathsf{H}, \mathsf{T}\}$.

Probability

The **sample space** S is the set of *outcomes* of an experiment.

Examples

Roll a die: $S = \{1, 2, 3, 4, 5, 6\}$.

Flip a coin: $S = \{\text{H}, \text{T}\}$.

Amount of money you can win when playing some lottery:

$$S = \{\pounds 0, \pounds 10, \pounds 100, \pounds 1000, \pounds 10,000, \pounds 100,000\}.$$

Probability

The **sample space** S is the set of *outcomes* of an experiment.

Examples

Roll a die: $S = \{1, 2, 3, 4, 5, 6\}$.

Flip a coin: $S = \{\mathsf{H}, \mathsf{T}\}$.

Amount of money you can win when playing some lottery:

$$S = \{£0, £10, £100, £1000, £10,000, £100,000\}.$$

For $x \in S$, the **probability** of x, written $\Pr(x)$,

is a real number between 0 and 1,

such that $\sum_{x \in S} \Pr(x) = 1.$

Probability

The **sample space** S is the set of *outcomes* of an experiment.

EXAMPLES

Roll a die: $S = \{1, 2, 3, 4, 5, 6\}$.

Flip a coin: $S = \{\mathsf{H}, \mathsf{T}\}$.

Amount of money you can win when playing some lottery:

$$S = \{\pounds 0, \pounds 10, \pounds 100, \pounds 1000, \pounds 10,000, \pounds 100,000\}.$$

For $x \in S$, the **probability** of x, written $\Pr(x)$,
is a real number between 0 and 1,
such that $\sum_{x \in S} \Pr(x) = 1$.

$\Pr$ is *'just'* a function which maps each $x \in S$ to $\Pr(x) \in [0, 1]$

Probability

The **sample space** S is the set of *outcomes* of an experiment.

For $x \in S$, the **probability** of x, written $\Pr(x)$,
is a real number between 0 and 1,
such that $\sum_{x \in S} \Pr(x) = 1.$

$\Pr$ is *'just'* a function which maps each $x \in S$ to $\Pr(x) \in [0, 1]$

Probability

The **sample space** S is the set of *outcomes* of an experiment.

EXAMPLE

Roll a die: $S = \{1, 2, 3, 4, 5, 6\}$.

$$\Pr(1) = \Pr(2) = \Pr(3) = \Pr(4) = \Pr(5) = \Pr(6) = \frac{1}{6}.$$

For $x \in S$, the **probability** of x, written $\Pr(x)$,

is a real number between 0 and 1,

such that $\sum_{x \in S} \Pr(x) = 1.$

$\Pr$ is *'just'* a function which maps each $x \in S$ to $\Pr(x) \in [0, 1]$

Probability

The **sample space** S is the set of *outcomes* of an experiment.

For $x \in S$, the **probability** of x, written $\Pr(x)$,
is a real number between 0 and 1,
such that $\sum_{x \in S} \Pr(x) = 1$.

$\Pr$ is *'just'* a function which maps each $x \in S$ to $\Pr(x) \in [0, 1]$

Probability

The **sample space** S is the set of *outcomes* of an experiment.

EXAMPLE

Flip a coin: $S = \{\mathsf{H}, \mathsf{T}\}$.

$$\Pr(\mathsf{H}) = \Pr(\mathsf{T}) = \tfrac{1}{2}.$$

For $x \in S$, the **probability** of x, written $\Pr(x)$,
is a real number between 0 and 1,
such that $\sum_{x \in S} \Pr(x) = 1.$

$\Pr$ is *'just'* a function which maps each $x \in S$ to $\Pr(x) \in [0, 1]$

Probability

The **sample space** S is the set of *outcomes* of an experiment.

For $x \in S$, the **probability** of x, written $\Pr(x)$,

is a real number between 0 and 1,

such that $\sum_{x \in S} \Pr(x) = 1$.

$\Pr$ is *'just'* a function which maps each $x \in S$ to $\Pr(x) \in [0, 1]$

Probability

The **sample space** S is the set of *outcomes* of an experiment.

EXAMPLE

Amount of money you can win when playing some lottery:

$$S = \{\pounds 0, \pounds 10, \pounds 100, \pounds 1000, \pounds 10,000, \pounds 100,000\}.$$

$$\Pr(\pounds 0) = 0.9,\ \Pr(\pounds 10) = 0.08,\ \ldots,\ \Pr(\pounds 100,000) = 0.0001.$$

For $x \in S$, the **probability** of x, written $\Pr(x)$,

is a real number between 0 and 1,

such that $\sum_{x \in S} \Pr(x) = 1$.

$\Pr$ is *'just'* a function which maps each $x \in S$ to $\Pr(x) \in [0, 1]$

Probability

The sample space is not necessarily *finite*.

Probability

The sample space is not necessarily *finite*.

Example

Flip a coin until first tail shows up

Probability

The sample space is not necessarily *finite*.

Example

Flip a coin until first tail shows up:

$$S = \{\mathsf{T}, \mathsf{HT}, \mathsf{HHT}, \mathsf{HHHT}, \mathsf{HHHHT}, \mathsf{HHHHHT}, \dots\}.$$

Probability

The sample space is not necessarily *finite*.

EXAMPLE

Flip a coin until first tail shows up:

$$S = \{\text{T}, \text{HT}, \text{HHT}, \text{HHHT}, \text{HHHHT}, \text{HHHHHT}, \ldots\}.$$

$\Pr(\text{“It takes } n \text{ coin flips”}) = \left(\frac{1}{2}\right)^n$, and

Probability

The sample space is not necessarily *finite*.

Example

Flip a coin until first tail shows up:

$$S = \{\text{T}, \text{HT}, \text{HHT}, \text{HHHT}, \text{HHHHT}, \text{HHHHHT}, \ldots\}.$$

$\Pr(\text{“It takes } n \text{ coin flips”}) = \left(\frac{1}{2}\right)^n$, and

$$\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n$$

Probability

The sample space is not necessarily *finite*.

Example

Flip a coin until first tail shows up:

$$S = \{\text{T}, \text{HT}, \text{HHT}, \text{HHHT}, \text{HHHHT}, \text{HHHHHT}, \ldots\}.$$

$\Pr(\text{“It takes } n \text{ coin flips”}) = \left(\frac{1}{2}\right)^n$, and

$$\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} \cdots$$

Probability

The sample space is not necessarily *finite*.

EXAMPLE

Flip a coin until first tail shows up:

$$S = \{\text{T}, \text{HT}, \text{HHT}, \text{HHHT}, \text{HHHHT}, \text{HHHHHT}, \ldots\}.$$

$\Pr(\text{"It takes } n \text{ coin flips"}) = \left(\frac{1}{2}\right)^n$, and

$$\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} \ldots = 1$$

Event

An **event** is a subset V of the sample space S.

Event

An **event** is a subset V of the sample space S.

The probability of event V happening, denoted $\Pr(V)$, is

$$\Pr(V) = \sum_{x \in V} \Pr(x).$$

Event

An **event** is a subset V of the sample space S.

The probability of event V happening, denoted $\Pr(V)$, is

$$\Pr(V) = \sum_{x \in V} \Pr(x).$$

EXAMPLE

Flip a coin 3 times: $S = \{\text{TTT}, \text{TTH}, \text{THT}, \text{HTT}, \text{HHT}, \text{HTH}, \text{THH}, \text{HHH}\}$

For each $x \in S$, $\Pr(x) = \frac{1}{8}$

Event

An **event** is a subset V of the sample space S.

The probability of event V happening, denoted $\Pr(V)$, is

$$\Pr(V) = \sum_{x \in V} \Pr(x).$$

EXAMPLE

Flip a coin 3 times: $S = \{\text{TTT}, \text{TTH}, \text{THT}, \text{HTT}, \text{HHT}, \text{HTH}, \text{THH}, \text{HHH}\}$

For each $x \in S$, $\Pr(x) = \frac{1}{8}$

Define V to be the event "the first and last coin flips are the same"

Event

An **event** is a subset V of the sample space S.

The probability of event V happening, denoted $\Pr(V)$, is

$$\Pr(V) = \sum_{x \in V} \Pr(x).$$

EXAMPLE

Flip a coin 3 times: $S = \{\text{TTT}, \text{TTH}, \text{THT}, \text{HTT}, \text{HHT}, \text{HTH}, \text{THH}, \text{HHH}\}$

For each $x \in S$, $\Pr(x) = \frac{1}{8}$

Define V to be the event "the first and last coin flips are the same"

in other words, $V = \{\text{HHH}, \text{HTH}, \text{THT}, \text{TTT}\}$

Event

An **event** is a subset V of the sample space S.

The probability of event V happening, denoted $\Pr(V)$, is

$$\Pr(V) = \sum_{x \in V} \Pr(x).$$

EXAMPLE

Flip a coin 3 times: $S = \{$TTT, TTH, THT, HTT, HHT, HTH, THH, HHH$\}$

For each $x \in S$, $\Pr(x) = \frac{1}{8}$

Define V to be the event "the first and last coin flips are the same"

in other words, $V = \{$HHH, HTH, THT, TTT$\}$

What is $\Pr(V)$?

Event

An **event** is a subset V of the sample space S.

The probability of event V happening, denoted $\Pr(V)$, is

$$\Pr(V) = \sum_{x \in V} \Pr(x).$$

EXAMPLE

Flip a coin 3 times: $S = \{\text{TTT}, \text{TTH}, \text{THT}, \text{HTT}, \text{HHT}, \text{HTH}, \text{THH}, \text{HHH}\}$

For each $x \in S$, $\Pr(x) = \frac{1}{8}$

Define V to be the event "the first and last coin flips are the same"

in other words, $V = \{\text{HHH}, \text{HTH}, \text{THT}, \text{TTT}\}$

What is $\Pr(V)$?

$$\Pr(V) = \Pr(\text{HHH}) + \Pr(\text{HTH}) + \Pr(\text{THT}) + \Pr(\text{TTT}) = 4 \times \frac{1}{8} = \frac{1}{2}.$$

Random variable

A **random variable** (r.v.) Y over sample space S is a function $S \to \mathbb{R}$

i.e. it maps each outcome $x \in S$ to some real number $Y(x)$.

Random variable

A **random variable** (r.v.) Y over sample space S is a function $S \to \mathbb{R}$

i.e. it maps each outcome $x \in S$ to some real number $Y(x)$.

The probability of Y taking value y is $\displaystyle \Pr(Y = y) = \sum_{\{x \in S \text{ st. } Y(x) = y\}} \Pr(x).$

Random variable

A **random variable** (r.v.) Y over sample space S is a function $S \to \mathbb{R}$

i.e. it maps each outcome $x \in S$ to some real number $Y(x)$.

The probability of Y taking value y is $\Pr(Y = y) = \sum_{\{x \in S \text{ st. } Y(x) = y\}} \Pr(x).$

EXAMPLE

Two coin flips.

S	Y
H H	2
H T	1
T H	5
T T	2

Random variable

A **random variable** (r.v.) Y over sample space S is a function $S \to \mathbb{R}$

i.e. it maps each outcome $x \in S$ to some real number $Y(x)$.

The probability of Y taking value y is $\Pr(Y = y) = \sum_{\{x \in S \text{ st. } Y(x) = y\}} \Pr(x).$

sum over all values of x such that $Y(x) = y$

Example

Two coin flips.

S	Y
H H	2
H T	1
T H	5
T T	2

Random variable

A **random variable** (r.v.) Y over sample space S is a function $S \to \mathbb{R}$

i.e. it maps each outcome $x \in S$ to some real number $Y(x)$.

The probability of Y taking value y is $\Pr(Y = y) = \sum_{\{x \in S \text{ st. } Y(x) = y\}} \Pr(x).$

sum over all values of x such that $Y(x) = y$

EXAMPLE

Two coin flips.

S	Y
H H	2
H T	1
T H	5
T T	2

What is $\Pr(Y = 2)$?

Random variable

A **random variable** (r.v.) Y over sample space S is a function $S \to \mathbb{R}$

i.e. it maps each outcome $x \in S$ to some real number $Y(x)$.

The probability of Y taking value y is $\Pr(Y = y) = \sum_{\{x \in S \text{ st. } Y(x) = y\}} \Pr(x).$

sum over all values of x such that $Y(x) = y$

EXAMPLE

Two coin flips.

S	Y
H H	2
H T	1
T H	5
T T	2

What is $\Pr(Y = 2)$?

$$\Pr(Y = 2) = \sum_{x \in \{\text{HH},\text{TT}\}} \Pr(x)$$

Random variable

A **random variable** (r.v.) Y over sample space S is a function $S \to \mathbb{R}$

i.e. it maps each outcome $x \in S$ to some real number $Y(x)$.

The probability of Y taking value y is $\displaystyle \Pr(Y = y) = \sum_{\{x \in S \text{ st. } Y(x) = y\}} \Pr(x).$

sum over all values of x such that $Y(x) = y$

EXAMPLE

Two coin flips.

S	Y
H H	2
H T	1
T H	5
T T	2

What is $\Pr(Y = 2)$?

$$\Pr(Y = 2) = \sum_{x \in \{\text{HH},\text{TT}\}} \Pr(x) = \frac{1}{4} + \frac{1}{4}$$

Random variable

A **random variable** (r.v.) Y over sample space S is a function $S \to \mathbb{R}$

i.e. it maps each outcome $x \in S$ to some real number $Y(x)$.

The probability of Y taking value y is $\Pr(Y = y) = \sum_{\{x \in S \text{ st. } Y(x) = y\}} \Pr(x)$.

sum over all values of x such that $Y(x) = y$

EXAMPLE

Two coin flips.

S	Y
H H	2
H T	1
T H	5
T T	2

What is $\Pr(Y = 2)$?

$$\Pr(Y = 2) = \sum_{x \in \{\text{HH},\text{TT}\}} \Pr(x) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

Random variable

A **random variable** (r.v.) Y over sample space S is a function $S \to \mathbb{R}$

i.e. it maps each outcome $x \in S$ to some real number $Y(x)$.

The probability of Y taking value y is $\Pr(Y = y) = \sum_{\{x \in S \text{ st. } Y(x)=y\}} \Pr(x).$

sum over all values of x such that $Y(x) = y$

EXAMPLE

Two coin flips.

S	Y
H H	2
H T	1
T H	5
T T	2

$\Pr(Y = 2) = \frac{1}{2}$

What is $\Pr(Y = 2)$?

$$\Pr(Y = 2) = \sum_{x \in \{\text{HH},\text{TT}\}} \Pr(x) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

Random variable

A **random variable** (r.v.) Y over sample space S is a function $S \to \mathbb{R}$

i.e. it maps each outcome $x \in S$ to some real number $Y(x)$.

The probability of Y taking value y is $\Pr(Y = y) = \sum_{\{x \in S \text{ st. } Y(x) = y\}} \Pr(x).$

EXAMPLE

Two coin flips.

S	Y
H H	2
H T	1
T H	5
T T	2

$$\Pr(Y = 2) = \frac{1}{2}$$

Random variable

A **random variable** (r.v.) Y over sample space S is a function $S \to \mathbb{R}$

i.e. it maps each outcome $x \in S$ to some real number $Y(x)$.

The probability of Y taking value y is $$\Pr(Y = y) = \sum_{\{x \in S \text{ st. } Y(x) = y\}} \Pr(x).$$

EXAMPLE

Two coin flips.

S	Y
H H	2
H T	1
T H	5
T T	2

$\Pr(Y = 2) = \frac{1}{2}$

The **expected value** (the mean) of a r.v. Y,

denoted $\mathbb{E}(Y)$, is

$$\mathbb{E}(Y) = \sum_{x \in S} Y(x) \cdot \Pr(x).$$

Random variable

A **random variable** (r.v.) Y over sample space S is a function $S \to \mathbb{R}$

i.e. it maps each outcome $x \in S$ to some real number $Y(x)$.

The probability of Y taking value y is $\Pr(Y = y) = \sum_{\{x \in S \text{ st. } Y(x) = y\}} \Pr(x).$

EXAMPLE

Two coin flips.

S	Y
H H	2
H T	1
T H	5
T T	2

$\Pr(Y = 2) = \frac{1}{2}$

The **expected value** (the mean) of a r.v. Y,

denoted $\mathbb{E}(Y)$, is

$$\mathbb{E}(Y) = \sum_{x \in S} Y(x) \cdot \Pr(x).$$

$$\mathbb{E}(Y) = \left(2 \cdot \tfrac{1}{2}\right) + \left(1 \cdot \tfrac{1}{4}\right) + \left(5 \cdot \tfrac{1}{4}\right) = \tfrac{7}{2}$$

Linearity of expectation

Theorem (Linearity of expectation)

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables. Then

$$\mathbb{E}\Big(\sum_{i=1}^{k} Y_i\Big) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

Linearity of expectation

Theorem (Linearity of expectation)

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables. Then

$$\mathbb{E}\Big(\sum_{i=1}^{k} Y_i\Big) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

Linearity of expectation **always** holds,

(regardless of whether the random variables are independent or not.)

Linearity of expectation

THEOREM (Linearity of expectation)

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables. Then

$$\mathbb{E}\Big(\sum_{i=1}^{k} Y_i\Big) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

Linearity of expectation **always** holds,

(regardless of whether the random variables are independent or not.)

EXAMPLE

Roll two dice. Let the r.v. Y be the sum of the values.

Linearity of expectation

THEOREM (Linearity of expectation)

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables. Then

$$\mathbb{E}\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

Linearity of expectation **always** holds,

(regardless of whether the random variables are independent or not.)

EXAMPLE

Roll two dice. Let the r.v. Y be the sum of the values.

random variable

Linearity of expectation

Theorem (Linearity of expectation)

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables. Then

$$\mathbb{E}\Big(\sum_{i=1}^{k} Y_i\Big) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

Linearity of expectation **always** holds,

(regardless of whether the random variables are independent or not.)

Example

Roll two dice. Let the r.v. Y be the sum of the values.

Linearity of expectation

Theorem (Linearity of expectation)

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables. Then

$$\mathbb{E}\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

Linearity of expectation **always** holds,

(regardless of whether the random variables are independent or not.)

Example

Roll two dice. Let the r.v. Y be the sum of the values.

What is $\mathbb{E}(Y)$?

Linearity of expectation

Theorem (Linearity of expectation)

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables. Then

$$\mathbb{E}\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

Linearity of expectation **always** holds,

(regardless of whether the random variables are independent or not.)

Example

Roll two dice. Let the r.v. Y be the sum of the values.

What is $\mathbb{E}(Y)$?

Approach 1: *(without the theorem)*

Linearity of expectation

Theorem (Linearity of expectation)

Let $Y_1, Y_2, \dots, Y_k$ be k random variables. Then

$$\mathbb{E}\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

Linearity of expectation **always** holds,

(regardless of whether the random variables are independent or not.)

Example

Roll two dice. Let the r.v. Y be the sum of the values. What is $\mathbb{E}(Y)$?

Approach 1: *(without the theorem)*

The sample space $S = \{(1,1),(1,2),(1,3)\dots(6,6)\}$ (36 outcomes)

Linearity of expectation

THEOREM (Linearity of expectation)

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables. Then

$$\mathbb{E}\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

Linearity of expectation **always** holds,

(regardless of whether the random variables are independent or not.)

EXAMPLE

Roll two dice. Let the r.v. Y be the sum of the values. What is $\mathbb{E}(Y)$?

Approach 1: *(without the theorem)*

The sample space $S = \{(1,1),(1,2),(1,3)\ldots(6,6)\}$ (36 outcomes)

$\mathbb{E}(Y) = \sum_{x\in S} Y(x)\cdot \Pr(x)$

Linearity of expectation

Theorem (Linearity of expectation)

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables. Then

$$\mathbb{E}\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

Linearity of expectation **always** holds,

(regardless of whether the random variables are independent or not.)

Example

Roll two dice. Let the r.v. Y be the sum of the values. What is $\mathbb{E}(Y)$?

Approach 1: *(without the theorem)*

The sample space $S = \{(1,1),(1,2),(1,3)\ldots(6,6)\}$ (36 outcomes)

$\mathbb{E}(Y) = \sum_{x\in S} Y(x)\cdot \Pr(x) = \frac{1}{36}\sum_{x\in S} Y(x)$

Linearity of expectation

Theorem (Linearity of expectation)

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables. Then

$$\mathbb{E}\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

Linearity of expectation **always** holds,

(regardless of whether the random variables are independent or not.)

Example

Roll two dice. Let the r.v. Y be the sum of the values. What is $\mathbb{E}(Y)$?

Approach 1: *(without the theorem)*

The sample space $S = \{(1,1),(1,2),(1,3)\ldots(6,6)\}$ (36 outcomes)

$$\mathbb{E}(Y) = \sum_{x\in S} Y(x)\cdot \Pr(x) = \frac{1}{36}\sum_{x\in S} Y(x) =$$
$$\frac{1}{36}(1\cdot 2 + 2\cdot 3 + 3\cdot 4 + \cdots + 1\cdot 12)$$

Linearity of expectation

Theorem (Linearity of expectation)

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables. Then

$$\mathbb{E}\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

Linearity of expectation **always** holds,

(regardless of whether the random variables are independent or not.)

Example

Roll two dice. Let the r.v. Y be the sum of the values. What is $\mathbb{E}(Y)$?

Approach 1: *(without the theorem)*

The sample space $S = \{(1,1),(1,2),(1,3)\ldots(6,6)\}$ (36 outcomes)

$$\mathbb{E}(Y) = \sum_{x \in S} Y(x) \cdot \Pr(x) = \frac{1}{36}\sum_{x \in S} Y(x) =$$

$$\frac{1}{36}(1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + \cdots + 1 \cdot 12) = 7$$

Linearity of expectation

Theorem (Linearity of expectation)

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables. Then

$$\mathbb{E}\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

Linearity of expectation **always** holds,

(regardless of whether the random variables are independent or not.)

Example

Roll two dice. Let the r.v. Y be the sum of the values.

What is $\mathbb{E}(Y)$?

Linearity of expectation

THEOREM (Linearity of expectation)

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables. Then

$$\mathbb{E}\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

Linearity of expectation **always** holds,

(regardless of whether the random variables are independent or not.)

EXAMPLE

Roll two dice. Let the r.v. Y be the sum of the values.

What is $\mathbb{E}(Y)$?

Approach 2: *(with the theorem)*

Linearity of expectation

Theorem (Linearity of expectation)

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables. Then

$$\mathbb{E}\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

Linearity of expectation **always** holds,

(regardless of whether the random variables are independent or not.)

Example

Roll two dice. Let the r.v. Y be the sum of the values.

What is $\mathbb{E}(Y)$?

Approach 2: *(with the theorem)*

Let the r.v. Y_1 be the value of the first die and Y_2 the value of the second

Linearity of expectation

Theorem (Linearity of expectation)

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables. Then

$$\mathbb{E}\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

Linearity of expectation **always** holds,

(regardless of whether the random variables are independent or not.)

Example

Roll two dice. Let the r.v. Y be the sum of the values. What is $\mathbb{E}(Y)$?

Approach 2: *(with the theorem)*

Let the r.v. Y_1 be the value of the first die and Y_2 the value of the second

$$\mathbb{E}(Y_1) = \mathbb{E}(Y_2) = 3.5$$

Linearity of expectation

THEOREM (Linearity of expectation)

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables. Then

$$\mathbb{E}\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

Linearity of expectation **always** holds,

(regardless of whether the random variables are independent or not.)

EXAMPLE

Roll two dice. Let the r.v. Y be the sum of the values. What is $\mathbb{E}(Y)$?

Approach 2: *(with the theorem)*

Let the r.v. Y_1 be the value of the first die and Y_2 the value of the second

$$\mathbb{E}(Y_1) = \mathbb{E}(Y_2) = 3.5$$

so $\mathbb{E}(Y) = \mathbb{E}(Y_1 + Y_2) = \mathbb{E}(Y_1) + \mathbb{E}(Y_2) = 7$

Indicator random variables

An **indicator random variable** is a r.v. that can only be 0 or 1.

(usually referred to by the letter I)

Indicator random variables

An **indicator random variable** is a r.v. that can only be 0 or 1.

(usually referred to by the letter I)

Fact: $\mathbb{E}(I) = 0 \cdot \Pr(I = 0) + 1 \cdot \Pr(I = 1) = \Pr(I = 1)$.

Indicator random variables

An **indicator random variable** is a r.v. that can only be 0 or 1.

(usually referred to by the letter I)

Fact: $\mathbb{E}(I) = \Pr(I = 1)$.

Indicator random variables

An **indicator random variable** is a r.v. that can only be 0 or 1.

(usually referred to by the letter I)

Fact: $\mathbb{E}(I) = \Pr(I = 1)$.

Often an indicator r.v. I is associated with an event such that

$I = 1$ if the event happens (and $I = 0$ otherwise).

Indicator random variables

An **indicator random variable** is a r.v. that can only be 0 or 1.

(usually referred to by the letter I)

Fact: $\mathbb{E}(I) = \Pr(I = 1)$.

Often an indicator r.v. I is associated with an event such that

$I = 1$ if the event happens (and $I = 0$ otherwise).

Indicator random variables and linearity of expectation work great together!

Indicator random variables

An **indicator random variable** is a r.v. that can only be 0 or 1.

(usually referred to by the letter I)

Fact: $\mathbb{E}(I) = \Pr(I = 1)$.

Often an indicator r.v. I is associated with an event such that

$I = 1$ if the event happens (and $I = 0$ otherwise).

Indicator random variables and linearity of expectation work great together!

EXAMPLE

Roll a die n times.

Indicator random variables

An **indicator random variable** is a r.v. that can only be 0 or 1.

(usually referred to by the letter I)

Fact: $\mathbb{E}(I) = \Pr(I = 1)$.

Often an indicator r.v. I is associated with an event such that

$I = 1$ if the event happens (and $I = 0$ otherwise).

Indicator random variables and linearity of expectation work great together!

EXAMPLE

Roll a die n times.

What is the expected number rolls that show a value that is at least the value of the previous roll?

Indicator random variables

An **indicator random variable** is a r.v. that can only be 0 or 1.

(usually referred to by the letter I)

Fact: $\mathbb{E}(I) = \Pr(I = 1)$.

Often an indicator r.v. I is associated with an event such that

$I = 1$ if the event happens (and $I = 0$ otherwise).

Indicator random variables and linearity of expectation work great together!

EXAMPLE

Roll a die n times.

What is the expected number rolls that show a value that is at least the value of the previous roll?

For $j \in \{2, \ldots, n\}$, let indicator r.v. $I_j = 1$ if the value of the jth roll is at least the value of the previous roll (and $I_j = 0$ otherwise)

Indicator random variables

An **indicator random variable** is a r.v. that can only be 0 or 1.

(usually referred to by the letter I)

Fact: $\mathbb{E}(I) = \Pr(I = 1)$.

Often an indicator r.v. I is associated with an event such that

$I = 1$ if the event happens (and $I = 0$ otherwise).

Indicator random variables and linearity of expectation work great together!

EXAMPLE

Roll a die n times.

What is the expected number rolls that show a value that is at least the value of the previous roll?

For $j \in \{2, \ldots, n\}$, let indicator r.v. $I_j = 1$ if the value of the jth roll is at least the value of the previous roll (and $I_j = 0$ otherwise)

$\Pr(I_j = 1) = \frac{21}{36} = \frac{7}{12}$. (by counting the outcomes)

Indicator random variables

An **indicator random variable** is a r.v. that can only be 0 or 1.

(usually referred to by the letter I)

Fact: $\mathbb{E}(I) = \Pr(I = 1)$.

Often an indicator r.v. I is associated with an event such that

$I = 1$ if the event happens (and $I = 0$ otherwise).

Indicator random variables and linearity of expectation work great together!

EXAMPLE

Roll a die n times.

What is the expected number rolls that show a value that is at least the value of the previous roll?

For $j \in \{2, \ldots, n\}$, let indicator r.v. $I_j = 1$ if the value of the jth roll is at least the value of the previous roll (and $I_j = 0$ otherwise)

$\Pr(I_j = 1) = \frac{21}{36} = \frac{7}{12}$. (by counting the outcomes)

$$E\left(\sum_{j=2}^{n} I_j\right) = \sum_{j=2}^{n} \mathbb{E}(I_j) = \sum_{j=2}^{n} \Pr(I_j = 1) = (n-1) \cdot \frac{7}{12}$$

Indicator random variables

An **indicator random variable** is a r.v. that can only be 0 or 1.

(usually referred to by the letter I)

Fact: $\mathbb{E}(I) = \Pr(I = 1)$.

Often an indicator r.v. I is associated with an event such that

$I = 1$ if the event happens (and $I = 0$ otherwise).

Indicator random variables and linearity of expectation work great together!

Linearity of Expectation

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables. Then

$$\mathbb{E}\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

rolls that show a value

ast the value of the previous roll?

of the jth roll

revious roll (and $I_j = 0$ otherwise)

$\Pr(I_j = 1) = \frac{21}{36} = \frac{7}{12}$. (by counting the outcomes)

$$E\left(\sum_{j=2}^{n} I_j\right) = \sum_{j=2}^{n} \mathbb{E}(I_j) = \sum_{j=2}^{n} \Pr(I_j = 1) = (n-1)\cdot\frac{7}{12}$$

Indicator random variables

An **indicator random variable** is a r.v. that can only be 0 or 1.

(usually referred to by the letter I)

Fact: $\mathbb{E}(I) = \Pr(I = 1)$.

Often an indicator r.v. I is associated with an event such that

$I = 1$ if the event happens (and $I = 0$ otherwise).

Indicator random variables and linearity of expectation work great together!

EXAMPLE

Roll a die n times.

What is the expected number rolls that show a value that is at least the value of the previous roll?

For $j \in \{2, \ldots, n\}$, let indicator r.v. $I_j = 1$ if the value of the jth roll is at least the value of the previous roll (and $I_j = 0$ otherwise)

$\Pr(I_j = 1) = \frac{21}{36} = \frac{7}{12}$. (by counting the outcomes)

$$E\left(\sum_{j=2}^{n} I_j\right) = \sum_{j=2}^{n} \mathbb{E}(I_j) = \sum_{j=2}^{n} \Pr(I_j = 1) = (n-1) \cdot \frac{7}{12}$$

Indicator random variables

An **indicator random variable** is a r.v. that can only be 0 or 1.

(usually referred to by the letter I)

Fact: $\mathbb{E}(I) = \Pr(I = 1).$

Often an indicator r.v. I is associated with an event such that

$I = 1$ if the event happens (and $I = 0$ otherwise).

Indicator random variables and linearity of expectation work great together!

EXAMPLE

Roll a die n times.

What is the expected number rolls that show a value that is at least the value of the previous roll?

For $j \in \{2, \ldots, n\}$, let indicator r.v. $I_j = 1$ if the value of the jth roll is at least the value of the previous roll (and $I_j = 0$ otherwise)

$\Pr(I_j = 1) = \frac{21}{36} = \frac{7}{12}$. (by counting the outcomes)

$$E\left(\sum_{j=2}^{n} I_j\right) = \sum_{j=2}^{n} \mathbb{E}(I_j) = \sum_{j=2}^{n} \Pr(I_j = 1) = (n-1) \cdot \frac{7}{12}$$

Indicator random variables

An **indicator random variable** is a r.v. that can only be 0 or 1.

(usually referred to by the letter I)

Fact: $\mathbb{E}(I) = \Pr(I = 1)$.

Often an indicator r.v. I is associated with an event such that

$I = 1$ if the event happens (and $I = 0$ otherwise).

Indicator random variables and linearity of expectation work great together!

EXAMPLE

Roll a die n times.

What is the expected number rolls that show a value that is at least the value of the previous roll?

For $j \in \{2, \ldots, n\}$, let indicator r.v. $I_j = 1$ if the value of the jth roll is at least the value of the previous roll (and $I_j = 0$ otherwise)

$\Pr(I_j = 1) = \frac{21}{36} = \frac{7}{12}$. (by counting the outcomes)

$$E\left(\sum_{j=2}^{n} I_j\right) = \sum_{j=2}^{n} \mathbb{E}(I_j) = \sum_{j=2}^{n} \Pr(I_j = 1) = (n-1) \cdot \frac{7}{12}$$

Markov's inequality

Example

Suppose that the average (mean) speed on the motorway is 60 mph.

Markov's inequality

Example

Suppose that the average (mean) speed on the motorway is 60 mph.

It then follows that at most

Markov's inequality

EXAMPLE

Suppose that the average (mean) speed on the motorway is 60 mph.

It then follows that at most

$\frac{1}{2}$ of all cars drive at least 120 mph,

Markov's inequality

EXAMPLE

Suppose that the average (mean) speed on the motorway is 60 mph.

It then follows that at most

$\frac{1}{2}$ of all cars drive at least 120 mph,

...otherwise the mean must be higher than 60 mph. (a contradiction)

Markov's inequality

Example

Suppose that the average (mean) speed on the motorway is 60 mph.

It then follows that at most

$\frac{2}{3}$ of all cars drive at least 90 mph,

...otherwise the mean must be higher than 60 mph. (a contradiction)

Markov's inequality

Example

Suppose that the average (mean) speed on the motorway is 60 mph.

It then follows that at most

$\frac{2}{3}$ of all cars drive at least 90 mph,

. . . otherwise the mean must be higher than 60 mph. (a contradiction)

Theorem (Markov's inequality)

If X is a non-negative r.v., then for all $a > 0$,

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

Markov's inequality

EXAMPLE

Suppose that the average (mean) speed on the motorway is 60 mph.

It then follows that at most

$\frac{2}{3}$ of all cars drive at least 90 mph,

. . . otherwise the mean must be higher than 60 mph. (a contradiction)

THEOREM (Markov's inequality)

If X is a non-negative r.v., then for all $a > 0$,

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

EXAMPLE

From the example above:

- $\Pr($speed of a random car ≥ 120 mph$) \leq \frac{60}{120} = \frac{1}{2}$,
- $\Pr($speed of a random car ≥ 90mph$) \leq \frac{60}{90} = \frac{2}{3}$.

Markov's inequality

EXAMPLE

n people go to a party, leaving their hats at the door.

Each person leaves with a random hat.

Markov's inequality

EXAMPLE

n people go to a party, leaving their hats at the door.

Each person leaves with a random hat.

How many people leave with their own hat?

Markov's inequality

EXAMPLE

n people go to a party, leaving their hats at the door.

Each person leaves with a random hat.

How many people leave with their own hat?

For $j \in \{1, \ldots, n\}$, let indicator r.v. $I_j = 1$ if the jth person gets their own hat, otherwise $I_j = 0$.

Markov's inequality

EXAMPLE

n people go to a party, leaving their hats at the door.

Each person leaves with a random hat.

How many people leave with their own hat?

For $j \in \{1, \ldots, n\}$, let indicator r.v. $I_j = 1$ if the jth person gets their own hat,

otherwise $I_j = 0$.

By linearity of expectation...

$$\mathbb{E}\left(\sum_{j=1}^{n} I_j\right) = \sum_{j=1}^{n} \mathbb{E}(I_j) = \sum_{j=1}^{n} \Pr(I_j = 1) = n \cdot \frac{1}{n} = 1.$$

Markov's inequality

EXAMPLE

n people go to a party, leaving their hats at the door.

Each person leaves with a random hat.

How many people leave with their own hat?

For $j \in \{1, \ldots, n\}$, let indicator r.v. $I_j = 1$ if the jth person gets their own hat,

otherwise $I_j = 0$.

By linearity of expectation...

Fact: $\mathbb{E}(I) = \Pr(I = 1)$.

$$\mathbb{E}\left(\sum_{j=1}^{n} I_j\right) = \sum_{j=1}^{n} \mathbb{E}(I_j) = \sum_{j=1}^{n} \Pr(I_j = 1) = n \cdot \frac{1}{n} = 1.$$

Markov's inequality

Example

n people go to a party, leaving their hats at the door.
Each person leaves with a random hat.

How many people leave with their own hat?

For $j \in \{1, \ldots, n\}$, let indicator r.v. $I_j = 1$ if the jth person gets their own hat,
otherwise $I_j = 0$.

By linearity of expectation...

$$\mathbb{E}\left(\sum_{j=1}^{n} I_j\right) = \sum_{j=1}^{n} \mathbb{E}(I_j) = \sum_{j=1}^{n} \Pr(I_j = 1) = n \cdot \frac{1}{n} = 1.$$

Markov's inequality

Example

n people go to a party, leaving their hats at the door.

Each person leaves with a random hat.

How many people leave with their own hat?

For $j \in \{1, \ldots, n\}$, let indicator r.v. $I_j = 1$ if the jth person gets their own hat,

otherwise $I_j = 0$.

By linearity of expectation...

$$\mathbb{E}\left(\sum_{j=1}^{n} I_j\right) = \sum_{j=1}^{n} \mathbb{E}(I_j) = \sum_{j=1}^{n} \Pr(I_j = 1) = n \cdot \frac{1}{n} = 1.$$

By Markov's inequality (recall: $\Pr(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$),

Markov's inequality

Example

n people go to a party, leaving their hats at the door.

Each person leaves with a random hat.

How many people leave with their own hat?

For $j \in \{1, \ldots, n\}$, let indicator r.v. $I_j = 1$ if the jth person gets their own hat,

otherwise $I_j = 0$.

By linearity of expectation...

$$\mathbb{E}\left(\sum_{j=1}^{n} I_j\right) = \sum_{j=1}^{n} \mathbb{E}(I_j) = \sum_{j=1}^{n} \Pr(I_j = 1) = n \cdot \frac{1}{n} = 1.$$

By Markov's inequality (recall: $\Pr(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$),

$$\Pr(\text{5 or more people leaving with their own hats}) \leq \frac{1}{5},$$

Markov's inequality

EXAMPLE

n people go to a party, leaving their hats at the door.

Each person leaves with a random hat.

How many people leave with their own hat?

For $j \in \{1, \ldots, n\}$, let indicator r.v. $I_j = 1$ if the jth person gets their own hat,

otherwise $I_j = 0$.

By linearity of expectation...

$$\mathbb{E}\left(\sum_{j=1}^{n} I_j\right) = \sum_{j=1}^{n} \mathbb{E}(I_j) = \sum_{j=1}^{n} \Pr(I_j = 1) = n \cdot \frac{1}{n} = 1.$$

By Markov's inequality (recall: $\Pr(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$),

$\Pr($5 or more people leaving with their own hats$) \leq \frac{1}{5}$,

$\Pr($at least 1 person leaving with their own hat$) \leq \frac{1}{1} = 1$.

Markov's inequality

EXAMPLE

n people go to a party, leaving their hats at the door.

Each person leaves with a random hat.

How many people leave with their own hat?

For $j \in \{1, \ldots, n\}$, let indicator r.v. $I_j = 1$ if the jth person gets their own hat,

otherwise $I_j = 0$.

By linearity of expectation...

$$\mathbb{E}\Big(\sum_{j=1}^{n} I_j\Big) = \sum_{j=1}^{n} \mathbb{E}(I_j) = \sum_{j=1}^{n} \Pr(I_j = 1) = n \cdot \frac{1}{n} = 1.$$

By Markov's inequality (recall: $\Pr(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$),

$$\Pr(\text{5 or more people leaving with their own hats}) \leq \tfrac{1}{5},$$

$$\Pr(\text{at least 1 person leaving with their own hat}) \leq \tfrac{1}{1} = 1.$$

(sometimes Markov's inequality is not particularly informative)

Markov's inequality

EXAMPLE

n people go to a party, leaving their hats at the door.

Each person leaves with a random hat.

How many people leave with their own hat?

For $j \in \{1,\ldots,n\}$, let indicator r.v. $I_j = 1$ if the jth person gets their own hat, otherwise $I_j = 0$.

By linearity of expectation...

$$\mathbb{E}\left(\sum_{j=1}^{n} I_j\right) = \sum_{j=1}^{n} \mathbb{E}(I_j) = \sum_{j=1}^{n} \Pr(I_j = 1) = n \cdot \frac{1}{n} = 1.$$

By Markov's inequality (recall: $\Pr(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$),

$$\Pr(\text{5 or more people leaving with their own hats}) \leq \tfrac{1}{5},$$

$$\Pr(\text{at least 1 person leaving with their own hat}) \leq \tfrac{1}{1} = 1.$$

(sometimes Markov's inequality is not particularly informative)

In fact, here it can be shown that as $n \to \infty$, the probability that at least one person leaves with their own hat is $1 - \frac{1}{e} \approx 0.632$.

Markov's inequality

COROLLARY

If X is a non-negative r.v. that only takes integer values, then

$$\Pr(X > 0) = \Pr(X \geq 1) \leq \mathbb{E}(X).$$

For an indicator r.v. I, the bound is tight ($=$), as $\Pr(I > 0) = \mathbb{E}(I)$.

Union bound

Theorem (union bound)

Let $V_1, \ldots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \leq \sum_{i=1}^{k} \Pr(V_i).$$

Union bound

THEOREM (union bound)

Let $V_1, \ldots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \leq \sum_{i=1}^{k} \Pr(V_i).$$

This is the probability at least one of the events happens

Union bound

Theorem (union bound)

Let $V_1, \ldots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \leq \sum_{i=1}^{k} \Pr(V_i).$$

Union bound

THEOREM (union bound)

Let $V_1, \ldots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \leq \sum_{i=1}^{k} \Pr(V_i).$$

This bound is tight $(=)$ when the events are all disjoint.

(V_i and V_j are disjoint iff $V_i \cap V_j$ is empty)

Union bound

Theorem (union bound)

Let $V_1, \ldots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \leq \sum_{i=1}^{k} \Pr(V_i).$$

This bound is tight $(=)$ when the events are all disjoint.

(V_i and V_j are disjoint iff $V_i \cap V_j$ is empty)

Proof

Union bound

THEOREM (union bound)

Let $V_1, \ldots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \leq \sum_{i=1}^{k} \Pr(V_i).$$

This bound is tight $(=)$ when the events are all disjoint.

(V_i and V_j are disjoint iff $V_i \cap V_j$ is empty)

PROOF

Define indicator r.v. I_j to be 1 if event V_j happens, otherwise $I_j = 0$.

Union bound

Theorem (union bound)

Let $V_1, \ldots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \leq \sum_{i=1}^{k} \Pr(V_i).$$

This bound is tight $(=)$ when the events are all disjoint.

(V_i and V_j are disjoint iff $V_i \cap V_j$ is empty)

Proof

Define indicator r.v. I_j to be 1 if event V_j happens, otherwise $I_j = 0$.

Let the r.v. $X = \sum_{j=1}^{k} I_j$ be the number of events that happen.

Union bound

Theorem (union bound)

Let $V_1, \ldots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \leq \sum_{i=1}^{k} \Pr(V_i).$$

This bound is tight $(=)$ when the events are all disjoint.

(V_i and V_j are disjoint iff $V_i \cap V_j$ is empty)

Proof

Define indicator r.v. I_j to be 1 if event V_j happens, otherwise $I_j = 0$.

Let the r.v. $X = \sum_{j=1}^{k} I_j$ be the number of events that happen.

$$\Pr\left(\bigcup_{j=1}^{k} V_j\right) = \Pr(X > 0) \leq \mathbb{E}(X) = \mathbb{E}\left(\sum_{j=1}^{k} I_j\right) = \sum_{j=1}^{k} \mathbb{E}(I_j)$$

$$= \sum_{j=1}^{k} \Pr(V_j)$$

Union bound

THEOREM (union bound)

Let $V_1, \ldots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \leq \sum_{i=1}^{k} \Pr(V_i).$$

This bound is tight $(=)$ when the events are all disjoint.

(V_i and V_j are disjoint iff $V_i \cap V_j$ is empty)

PROOF

Define indicator r.v. I_j to be 1 if event V_j happens, otherwise $I_j = 0$.

Let the r.v. $X = \sum_{j=1}^{k} I_j$ be the number of events that happen.

$$\Pr\left(\bigcup_{j=1}^{k} V_j\right) = \Pr(X > 0) \leq \mathbb{E}(X) = \mathbb{E}\left(\sum_{j=1}^{k} I_j\right) = \sum_{j=1}^{k} \mathbb{E}(I_j)$$

$$= \sum_{j=1}^{k} \Pr(V_j)$$

by previous Markov corollary

Union bound

THEOREM (union bound)

Let $V_1, \dots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \le \sum_{i=1}^{k} \Pr(V_i).$$

This bound is tight $(=)$ when the events are all disjoint.

(V_i and V_j are disjoint iff $V_i \cap V_j$ is empty)

PROOF

Define indicator r.v. I_j to be 1 if event V_j happens, otherwise $I_j = 0$.

Let the r.v. $X = \sum_{j=1}^{k} I_j$ be the number of events that happen.

$$\Pr\left(\bigcup_{j=1}^{k} V_j\right) = \Pr(X > 0) \le \mathbb{E}(X) = \mathbb{E}\left(\sum_{j=1}^{k} I_j\right) = \sum_{j=1}^{k} \mathbb{E}(I_j)$$

$$= \sum_{j=1}^{k} \Pr(V_j)$$

by previous Markov corollary

Linearity of expectation

Union bound

THEOREM (union bound)

Let $V_1, \ldots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \leq \sum_{i=1}^{k} \Pr(V_i).$$

This bound is tight $(=)$ when the events are all disjoint.

(V_i and V_j are disjoint iff $V_i \cap V_j$ is empty)

PROOF

Define indicator r.v. I_j to be 1 if event V_j happens, otherwise $I_j = 0$.

Let the r.v. $X = \sum_{j=1}^{k} I_j$ be the number of events that happen.

$\mathbb{E}(I) = \Pr(I = 1)$

$$\Pr\left(\bigcup_{j=1}^{k} V_j\right) = \Pr(X > 0) \leq \mathbb{E}(X) = \mathbb{E}\left(\sum_{j=1}^{k} I_j\right) = \sum_{j=1}^{k} \mathbb{E}(I_j)$$

$$= \sum_{j=1}^{k} \Pr(V_j)$$

by previous Markov corollary

Linearity of expectation

Union bound

THEOREM (union bound)

Let $V_1, \ldots, V_k$ be k events. Then

$$\Pr\Big(\bigcup_{i=1}^{k} V_i\Big) \le \sum_{i=1}^{k} \Pr(V_i).$$

This bound is tight $(=)$ when the events are all disjoint.

(V_i and V_j are disjoint iff $V_i \cap V_j$ is empty)

Union bound

Theorem (union bound)

Let $V_1, \ldots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \leq \sum_{i=1}^{k} \Pr(V_i).$$

This bound is tight $(=)$ when the events are all disjoint.

(V_i and V_j are disjoint iff $V_i \cap V_j$ is empty)

Example

$S = \{1, \ldots, 6\}$ is the set of outcomes of a die roll.

Union bound

Theorem (union bound)

Let $V_1, \ldots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \leq \sum_{i=1}^{k} \Pr(V_i).$$

This bound is tight $(=)$ when the events are all disjoint.

(V_i and V_j are disjoint iff $V_i \cap V_j$ is empty)

Example

$S = \{1, \ldots, 6\}$ is the set of outcomes of a die roll.

We define two events:
$$V_1 = \{3, 4\}$$
$$V_2 = \{1, 2, 3\}$$

Union bound

THEOREM (union bound)

Let $V_1, \dots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \leq \sum_{i=1}^{k} \Pr(V_i).$$

This bound is tight ($=$) when the events are all disjoint.

(V_i and V_j are disjoint iff $V_i \cap V_j$ is empty)

EXAMPLE

$S = \{1, \dots, 6\}$ is the set of outcomes of a die roll.

We define two events: $V_1 = \{3, 4\}$
$V_2 = \{1, 2, 3\}$

$$\Pr(V_1 \cup V_2) \leq \Pr(V_1) + \Pr(V_2) = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}$$

Union bound

Theorem (union bound)

Let $V_1, \ldots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \leq \sum_{i=1}^{k} \Pr(V_i).$$

This bound is tight ($=$) when the events are all disjoint.

(V_i and V_j are disjoint iff $V_i \cap V_j$ is empty)

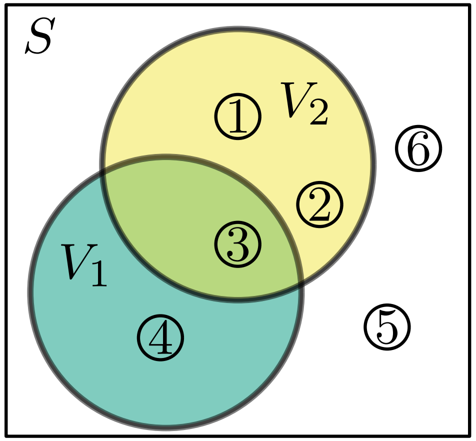
Example

$S = \{1, \ldots, 6\}$ is the set of outcomes of a die roll.

We define two events: $V_1 = \{3, 4\}$

$V_2 = \{1, 2, 3\}$

$$\Pr(V_1 \cup V_2) \leq \Pr(V_1) + \Pr(V_2) = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}$$



Union bound

THEOREM (union bound)

Let $V_1, \ldots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \le \sum_{i=1}^{k} \Pr(V_i).$$

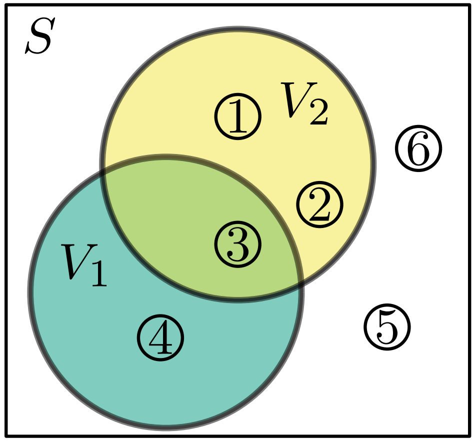
This bound is tight $(=)$ when the events are all disjoint.

(V_i and V_j are disjoint iff $V_i \cap V_j$ is empty)

EXAMPLE

$S = \{1, \ldots, 6\}$ is the set of outcomes of a die roll.

We define two events: $V_1 = \{3, 4\}$
$V_2 = \{1, 2, 3\}$



$$\Pr(V_1 \cup V_2) \le \Pr(V_1) + \Pr(V_2) = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}$$

in fact, $\Pr(V_1 \cup V_2) = \frac{2}{3}$

Union bound

THEOREM (union bound)

Let $V_1, \ldots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \leq \sum_{i=1}^{k} \Pr(V_i).$$

This bound is tight $(=)$ when the events are all disjoint.

(V_i and V_j are disjoint iff $V_i \cap V_j$ is empty)

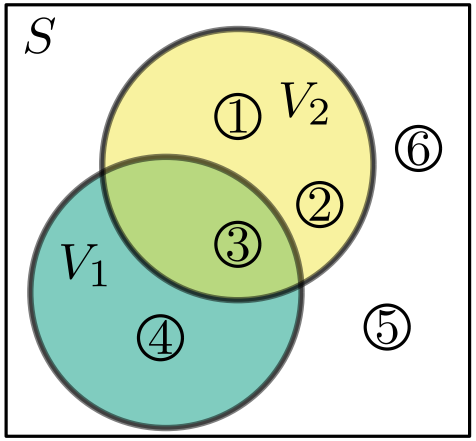
EXAMPLE

$S = \{1, \ldots, 6\}$ is the set of outcomes of a die roll.

We define two events: $V_1 = \{3, 4\}$
$V_2 = \{1, 2, 3\}$

$$\Pr(V_1 \cup V_2) \leq \Pr(V_1) + \Pr(V_2) = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}$$

in fact, $\Pr(V_1 \cup V_2) = \frac{2}{3}$ (3 was 'double counted')



Union bound

THEOREM (union bound)

Let $V_1, \ldots, V_k$ be k events. Then

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \leq \sum_{i=1}^{k} \Pr(V_i).$$

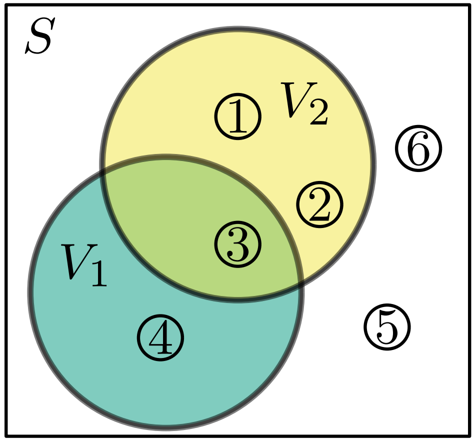
This bound is tight $(=)$ when the events are all disjoint.

(V_i and V_j are disjoint iff $V_i \cap V_j$ is empty)

EXAMPLE

$S = \{1, \ldots, 6\}$ is the set of outcomes of a die roll.

We define two events: $V_1 = \{3, 4\}$
$V_2 = \{1, 2, 3\}$



$$\Pr(V_1 \cup V_2) \leq \Pr(V_1) + \Pr(V_2) = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}$$

in fact, $\Pr(V_1 \cup V_2) = \frac{2}{3}$ (3 was 'double counted')

Typically the union bound is used when each $\Pr(V_i)$ is *much* smaller than k.

Summary

The **sample space** S is the set of *outcomes* of an experiment.

For $x \in S$, the **probability** of x, written $\Pr(x)$,is a real number between 0 and 1,

such that $\sum_{x \in S} \Pr(x) = 1$.

An **event** is a subset V of the sample space S, $\Pr(V) = \sum_{x \in V} \Pr(x)$

A **random variable** (r.v.) Y is a function which maps $x \in S$ to $S(x) \in \mathbb{R}$

The probability of Y taking value y is $\Pr(Y = y) = \sum_{\{x \in S \text{ st. } Y(x) = y\}} \Pr(x)$.

The **expected value** (the mean) of Y is $\mathbb{E}(Y) = \sum_{x \in S} Y(x) \cdot \Pr(x)$.

An **indicator random variable** is a r.v. that can only be 0 or 1.

Fact: $\mathbb{E}(I) = \Pr(I = 1)$.

THEOREM (Linearity of expectation)

Let $Y_1, Y_2, \ldots, Y_k$ be k random variables then,

$$\mathbb{E}\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} \mathbb{E}(Y_i)$$

THEOREM (union bound)

Let $V_1, \ldots, V_k$ be k events then,

$$\Pr\left(\bigcup_{i=1}^{k} V_i\right) \leq \sum_{i=1}^{k} \Pr(V_i).$$

THEOREM (Markov's inequality)

If X is a non-negative r.v., then for all $a > 0$,

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