Probability Theory
Random Variables and Distributions

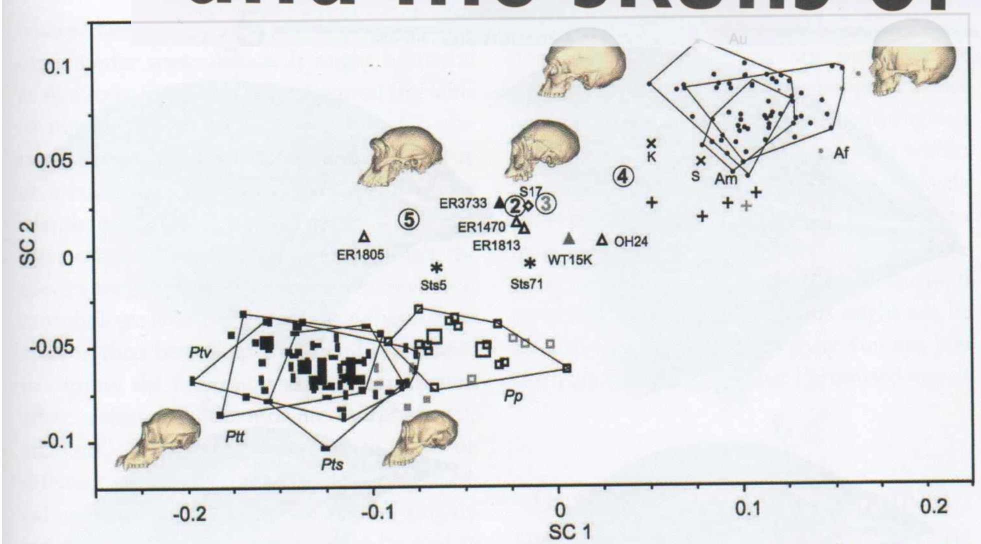
Rob Nicholls

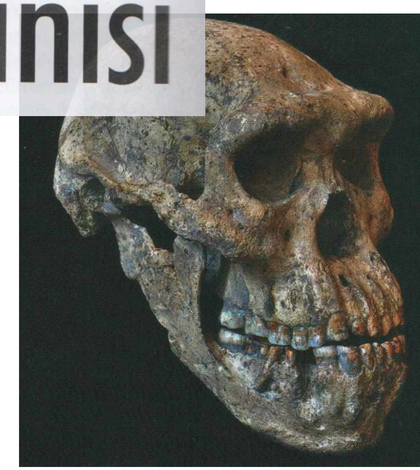
MRC LMB Statistics Course 2014

Introduction

Introduction

Early human evolution and the skulls of Dmanisi



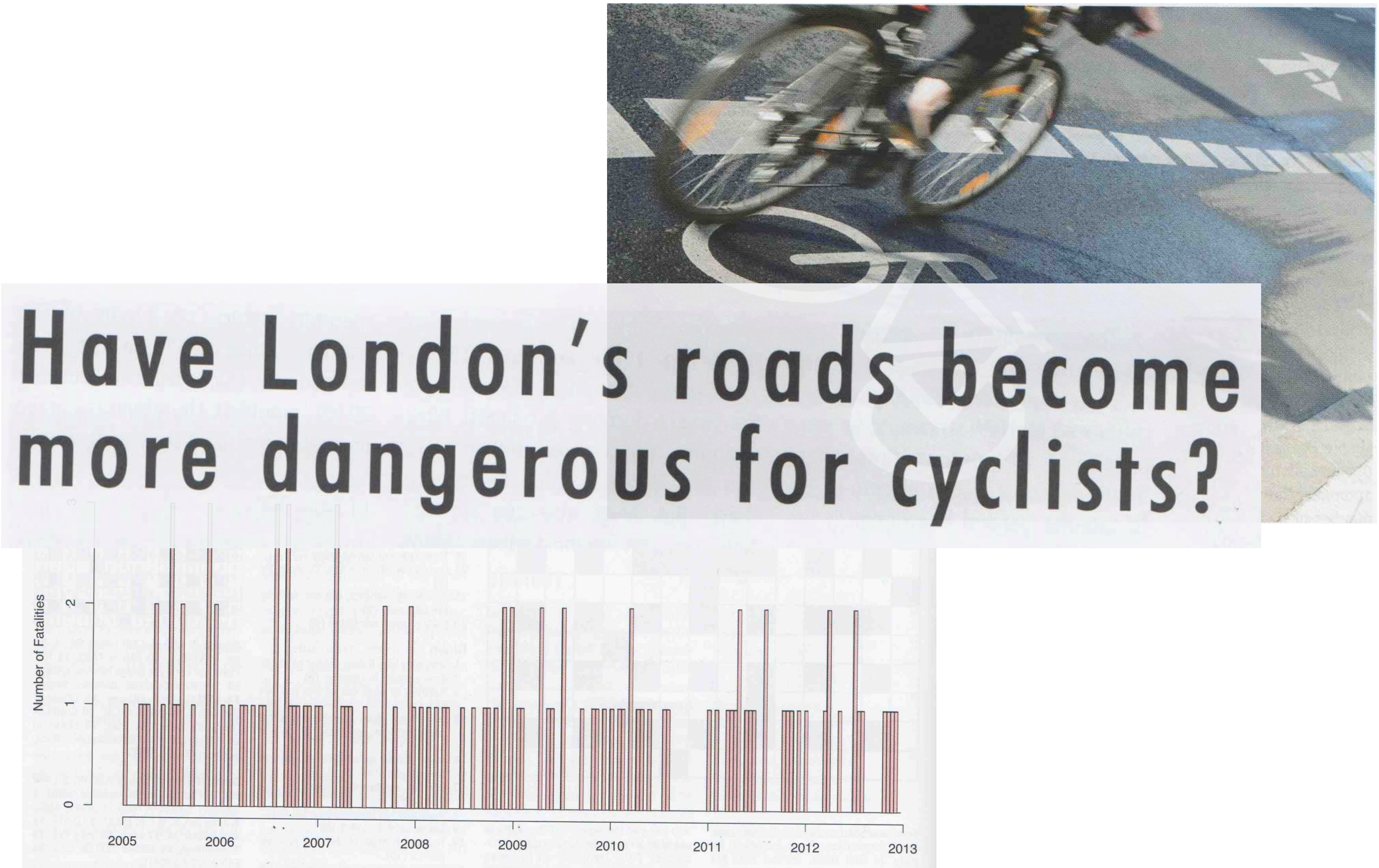


Significance magazine (December 2013) Royal Statistical Society

Introduction

Significance magazine (December 2013) Royal Statistical Society

Introduction



Significance magazine (December 2013) Royal Statistical Society

Contents

Set Notation

A set is a collection of objects, written using curly brackets {}

If *A* is the set of all outcomes, then:

$$A = \{heads, tails\}$$

$$A = \{one, two, three, four, five, six\}$$

A set does not have to comprise the full number of outcomes

E.g. if A is the set of dice outcomes no higher than three, then:

$$A = \{one, two, three\}$$

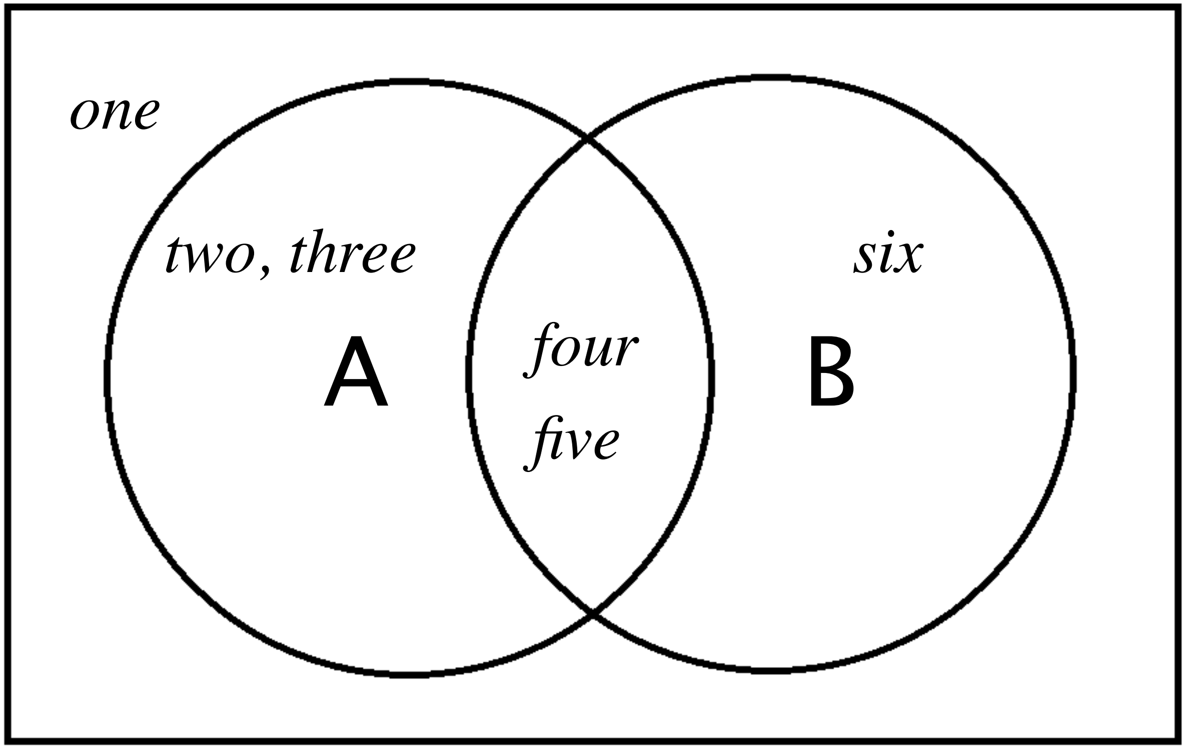
Set Notation

If A and B are sets, then:

A'	Complement – everything but A
$A \cup B$	Union (or)
$A \cap B$	Intersection (and)
$A \setminus B$	Not
$\varnothing$	Empty Set

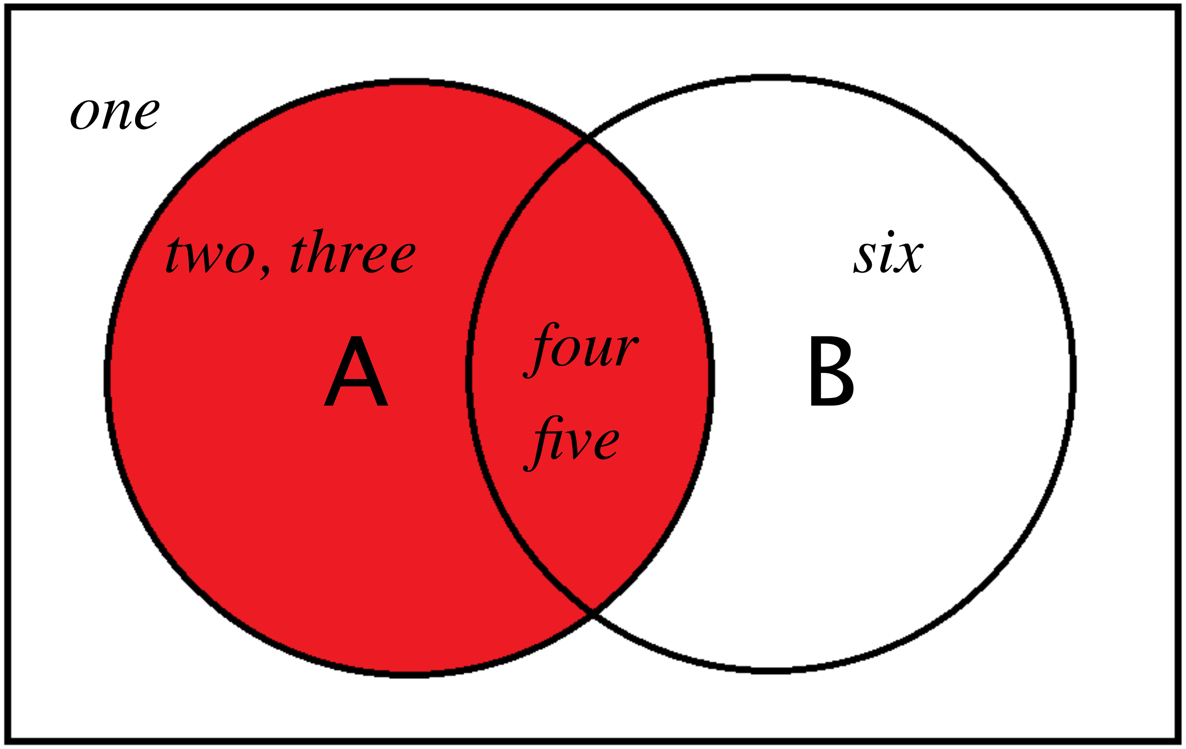
Set Notation

Venn Diagram:



Set Notation

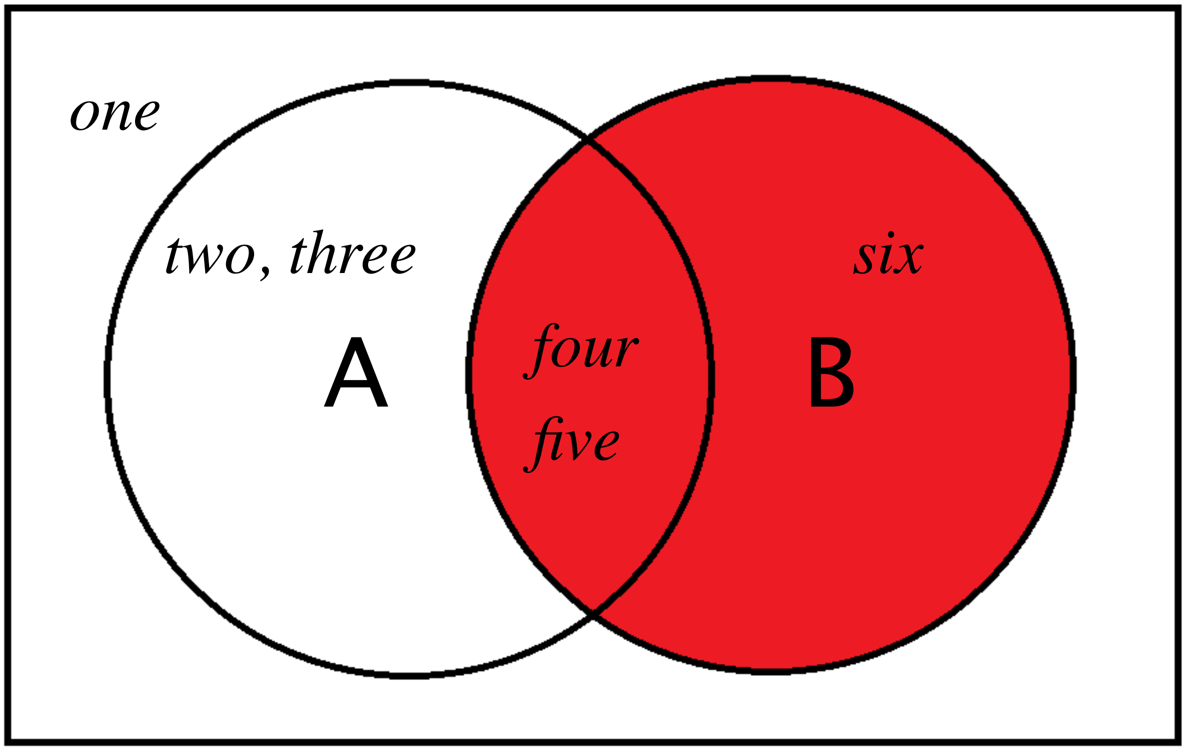
Venn Diagram:



$$A = \{two, three, four, five\}$$

Set Notation

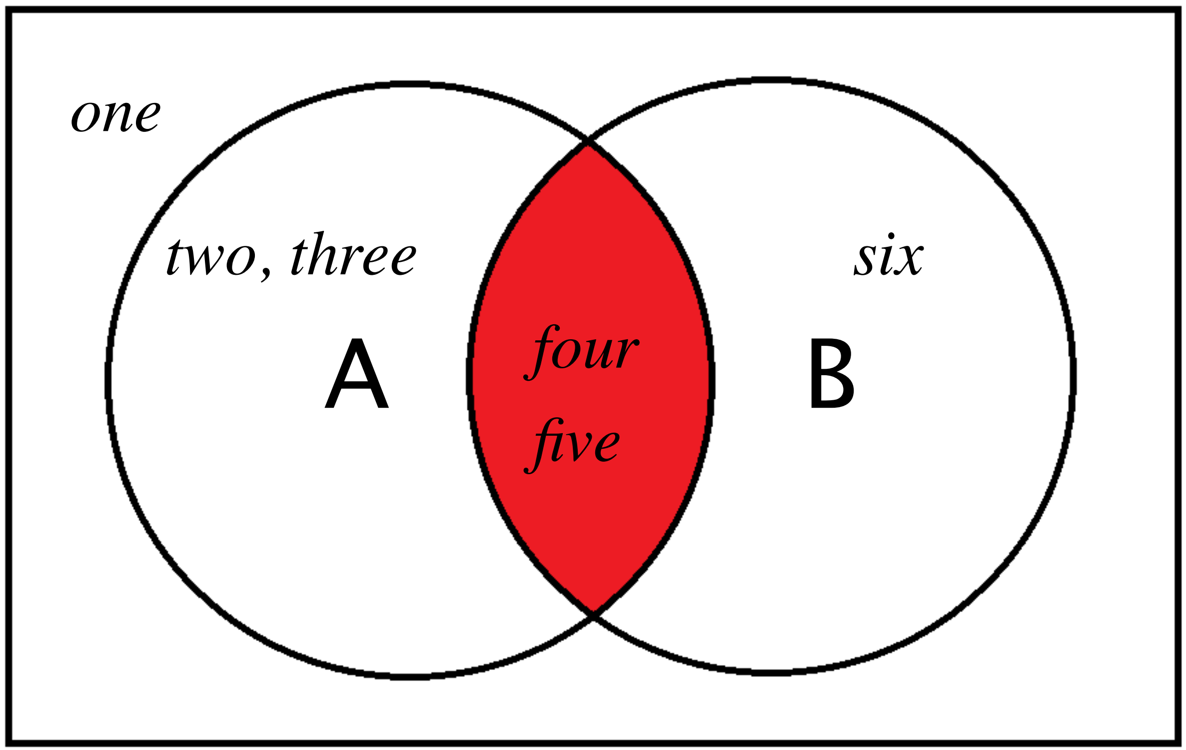
Venn Diagram:



$$B = \{four, five, six\}$$

Set Notation

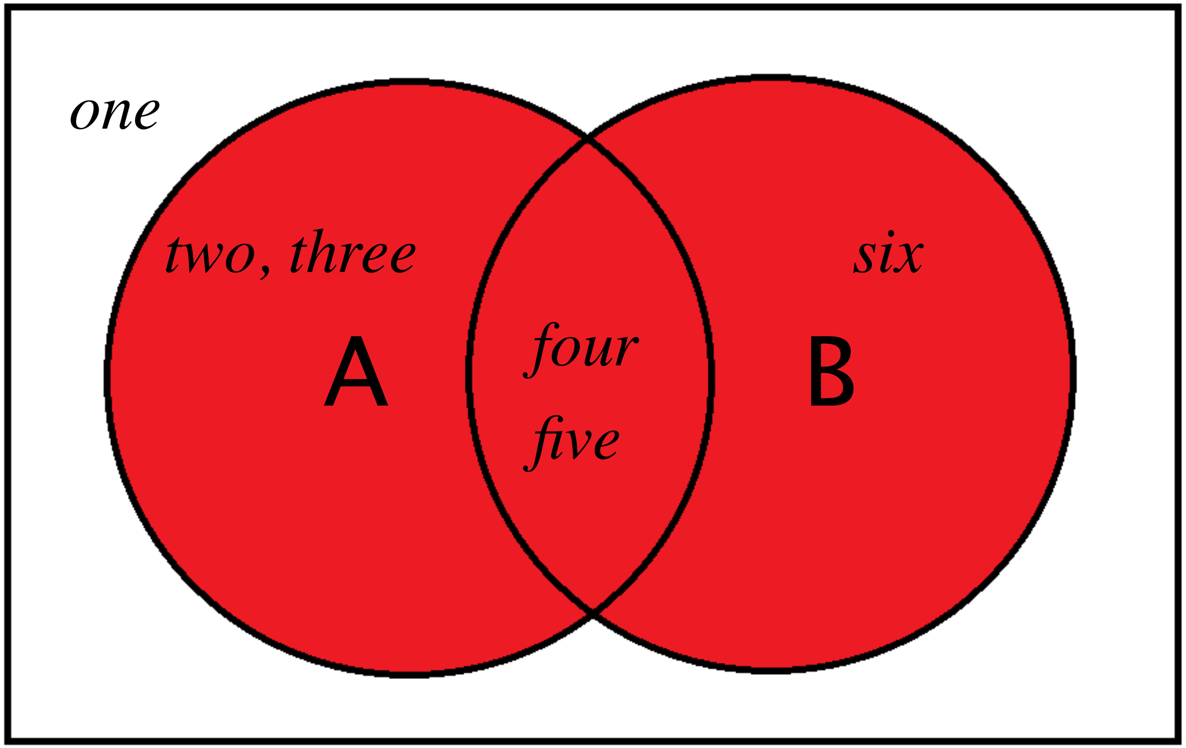
Venn Diagram:



$$A \cap B = \{four, five\}$$

Set Notation

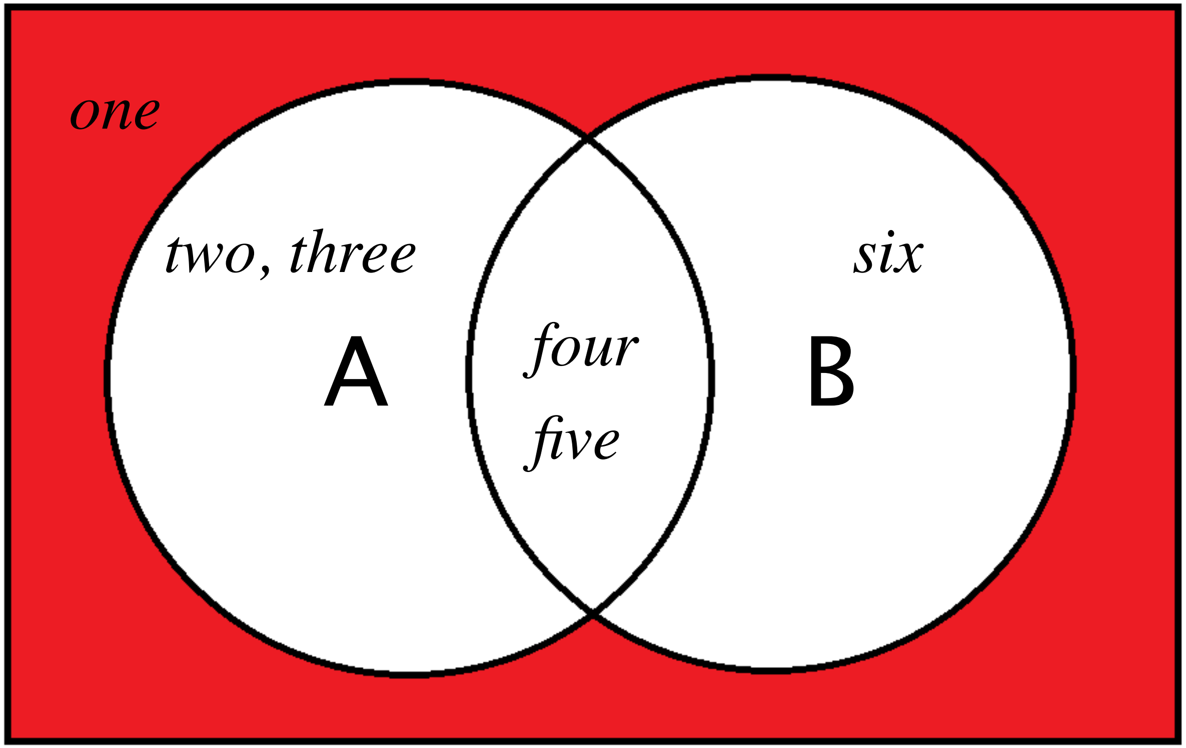
Venn Diagram:



$$A \cup B = \{two, three, four, five, six\}$$

Set Notation

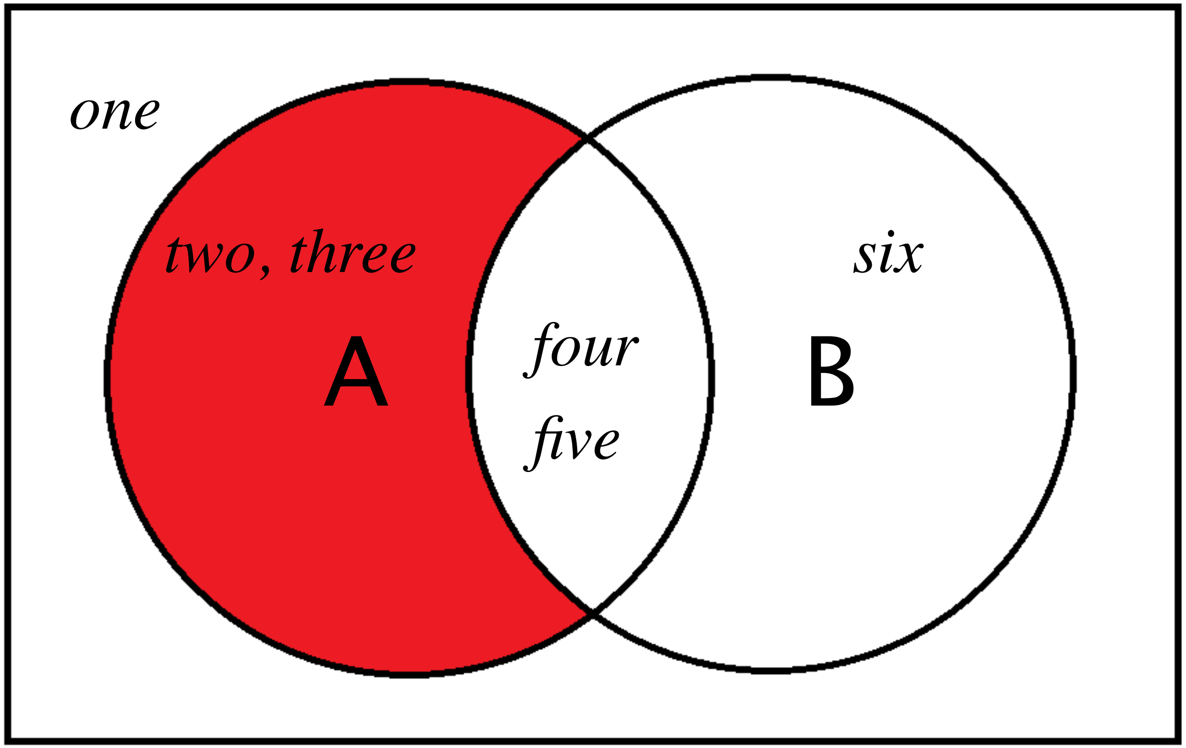
Venn Diagram:



$$(A \cup B)' = \{one\}$$

Set Notation

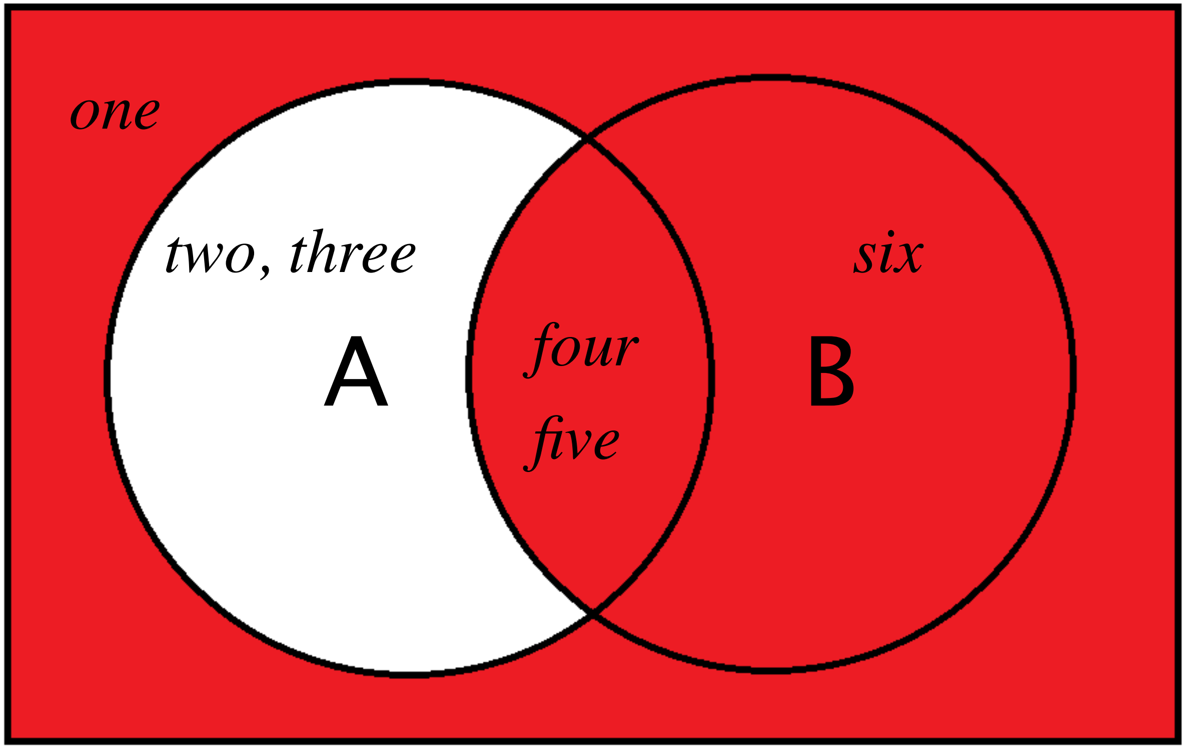
Venn Diagram:



$$A \setminus B = \{two, three\}$$

Set Notation

Venn Diagram:



$$(A \setminus B)' = \{one, four, five, six\}$$

Probability Theory

To consider Probabilities, we need:

1. Sample space: Ω

2. Event space: $\mathcal{F}$

3. Probability measure: P

Probability Theory

To consider Probabilities, we need:

1. Sample space: Ω – the set of all possible outcomes

$$\Omega = \{heads, tails\}$$

$$\Omega = \{one, two, three, four, five, six\}$$

Probability Theory

To consider Probabilities, we need:

2. Event space: $\mathcal{F}$ – the set of all possible events

$$\Omega = \{heads, tails\}$$

$$\mathcal{F} = \{\{heads, tails\}, \{heads\}, \{tails\}, \varnothing\}$$

Probability Theory

To consider Probabilities, we need:

3. Probability measure: *P* $P : \mathcal{F} \rightarrow [0,1]$

P must satisfy two axioms:

$$P(\Omega) = 1$$ Probability of any outcome is 1 (100% chance)

$$P(\bigcup_i A_i) = \sum_i P(A_i)$$ If and only if $A_1, A_2, \ldots$ are disjoint

Probability Theory

To consider Probabilities, we need:

3. Probability measure: *P* $P : \mathcal{F} \rightarrow [0,1]$

P must satisfy two axioms:

$P(\Omega) = 1$ Probability of any outcome is 1 (100% chance)

$P(\bigcup_i A_i) = \sum_i P(A_i)$ If and only if $A_1, A_2, \ldots$ are disjoint

$$P(\{one, two\}) = P(\{one\}) + P(\{two\})$$

$$\frac{1}{3} = \frac{1}{6} + \frac{1}{6}$$

Probability Theory

To consider Probabilities, we need:

1. Sample space: Ω
2. Event space: $\mathcal{F}$
3. Probability measure: P

As such, a *Probability Space* is the triple: $(\Omega, \mathcal{F}, P)$

Probability Theory

To consider Probabilities, we need:

The triple: $(\Omega, \mathcal{F}, P)$

i.e. we need to know:

1. The set of potential outcomes;
2. The set of potential events that may occur; and
3. The probabilities associated with occurrence of those events.

Probability Theory

Notable properties of a Probability Space $(\Omega, \mathcal{F}, P)$:

Probability Theory

Notable properties of a Probability Space $(\Omega, \mathcal{F}, P)$:

$$P(A') = 1 - P(A)$$

$$A = \{one, two\}$$

$$A' = \{three, four, five, six\}$$

$$P(A) = 1/3$$

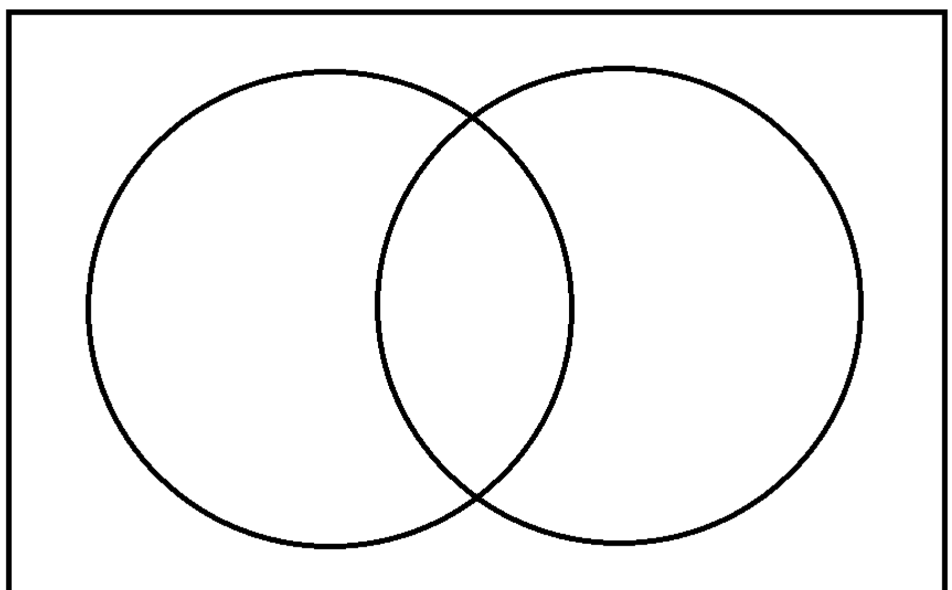
$$P(A') = 2/3$$

Probability Theory

Notable properties of a Probability Space $(\Omega, \mathcal{F}, P)$:

$$P(A') = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

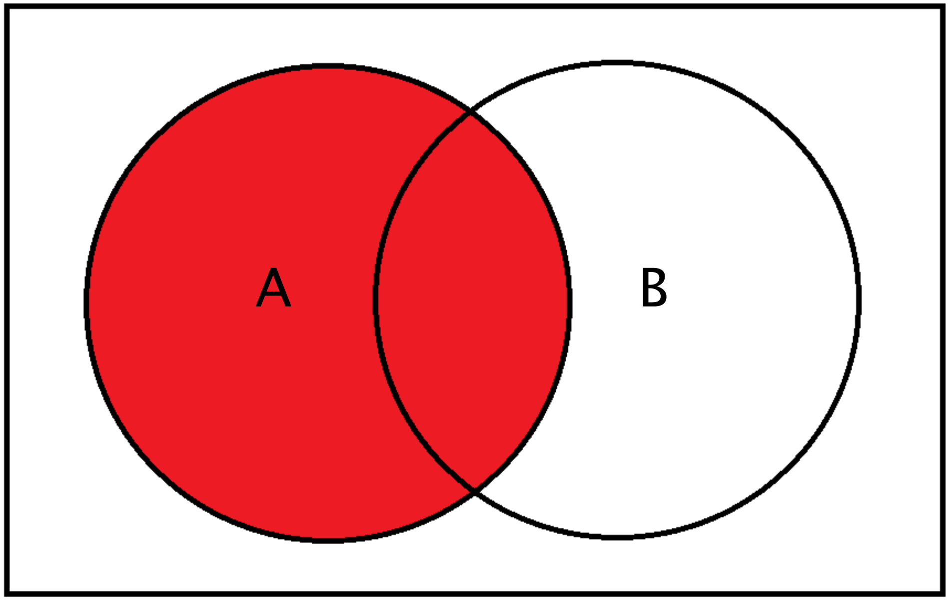


Probability Theory

Notable properties of a Probability Space $(\Omega, \mathcal{F}, P)$:

$$P(A') = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

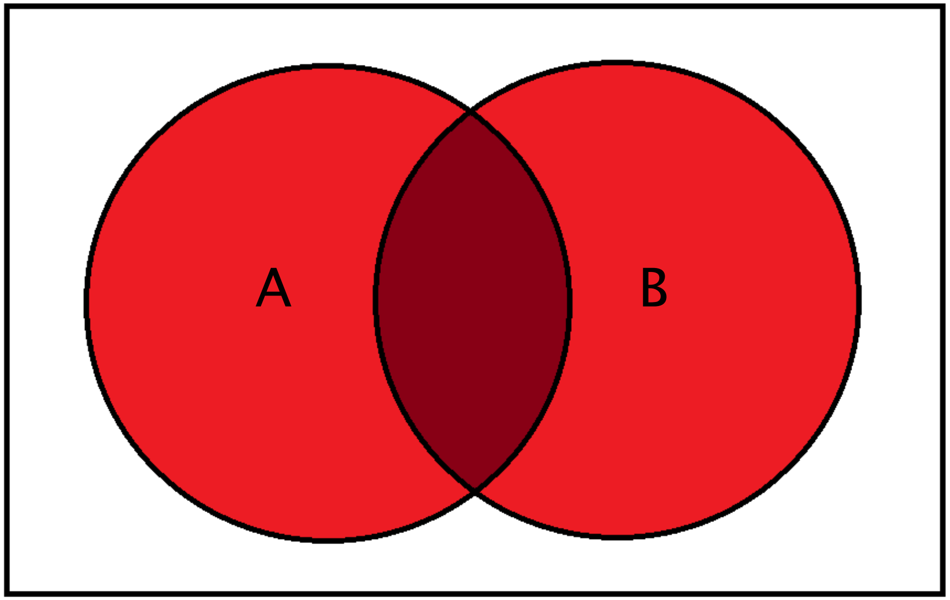


Probability Theory

Notable properties of a Probability Space $(\Omega, \mathcal{F}, P)$:

$$P(A') = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

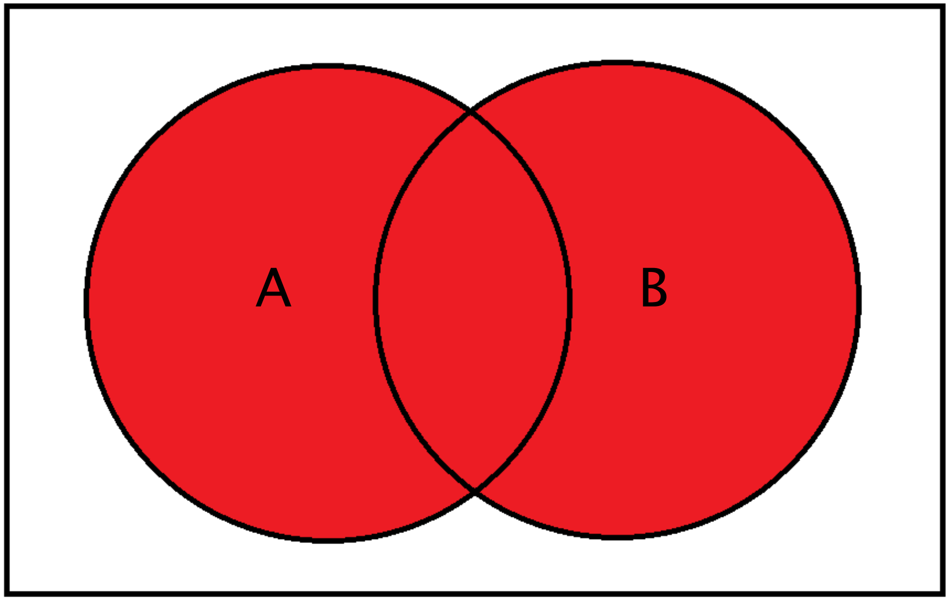


Probability Theory

Notable properties of a Probability Space $(\Omega, \mathcal{F}, P)$:

$$P(A') = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



Probability Theory

Notable properties of a Probability Space $(\Omega, \mathcal{F}, P)$:

$$P(A') = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$A = \{one, two\} \qquad P(A) = 1/3$$

$$B = \{two, three\} \qquad P(B) = 1/3$$

$$A \cup B = \{one, two, three\} \qquad P(A \cup B) = 1/2$$

$$A \cap B = \{two\} \qquad P(A \cap B) = 1/6$$

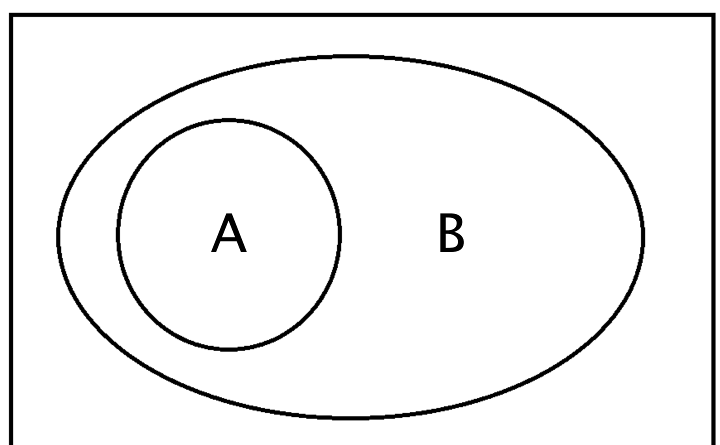
Probability Theory

Notable properties of a Probability Space $(\Omega, \mathcal{F}, P)$:

$$P(A') = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If $A \subseteq B$ then $P(A) \leq P(B)$ and $P(B \setminus A) = P(B) - P(A)$



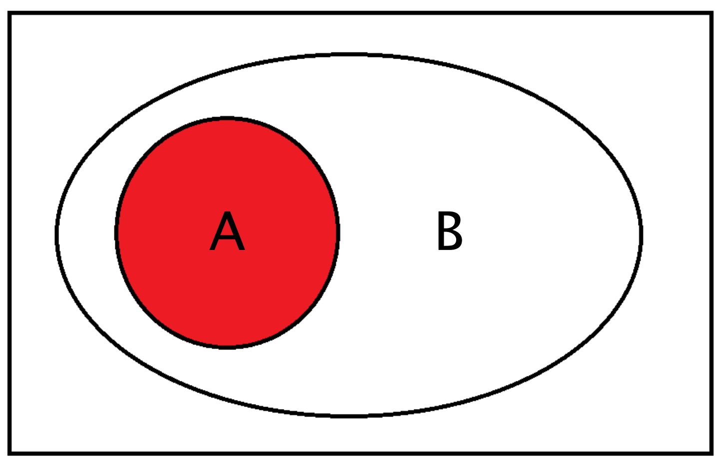
Probability Theory

Notable properties of a Probability Space $(\Omega, \mathcal{F}, P)$:

$$P(A') = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If $A \subseteq B$ then $P(A) \leq P(B)$ and $P(B \setminus A) = P(B) - P(A)$



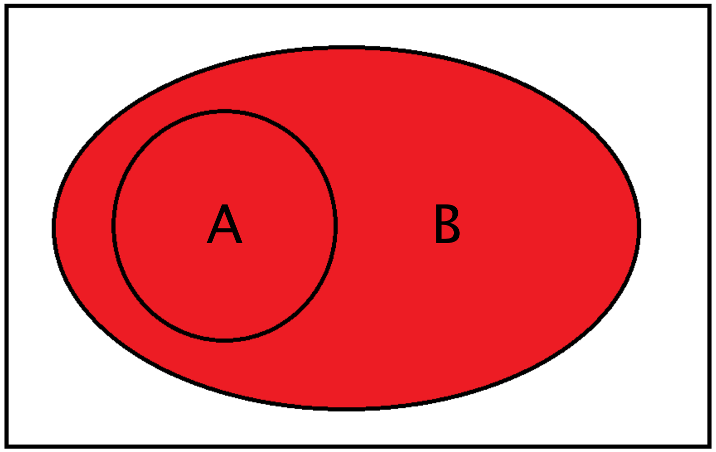
Probability Theory

Notable properties of a Probability Space $(\Omega, \mathcal{F}, P)$:

$$P(A') = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If $A \subseteq B$ then $P(A) \leq P(B)$ and $P(B \setminus A) = P(B) - P(A)$



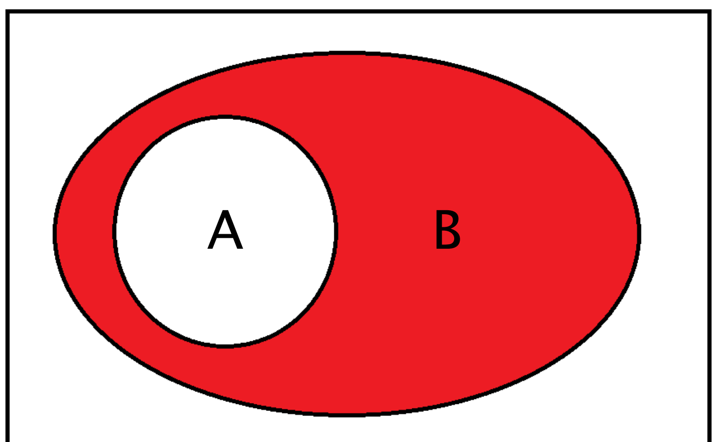
Probability Theory

Notable properties of a Probability Space $(\Omega, \mathcal{F}, P)$:

$$P(A') = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If $A \subseteq B$ then $P(A) \leq P(B)$ and $P(B \setminus A) = P(B) - P(A)$



Probability Theory

Notable properties of a Probability Space $(\Omega, \mathcal{F}, P)$:

$$P(A') = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If $A \subseteq B$ then $P(A) \leq P(B)$ and $P(B \setminus A) = P(B) - P(A)$

$A = \{one, two\}$ $\quad P(A) = 1/3$

$B = \{one, two, three\}$ $\quad P(B) = 1/2$

$B \setminus A = \{three\}$ $\quad P(B \setminus A) = 1/6$

Probability Theory

Notable properties of a Probability Space $(\Omega, \mathcal{F}, P)$:

$$P(A') = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If $A \subseteq B$ then $P(A) \leq P(B)$ and $P(B \setminus A) = P(B) - P(A)$

$$P(\emptyset) = 0$$

Probability Theory

So where's this all going? These examples are trivial!

Probability Theory

So where's this all going? These examples are trivial!

Suppose there are three bags, B_1, B_2 and B_3, each of which contain a number of coloured balls:

- B_1 – 2 red and 4 white
- B_2 – 1 red and 2 white
- B_3 – 5 red and 4 white

A ball is randomly removed from one the bags.

The bags were selected with probability:

- $P(B_1) = 1/3$
- $P(B_2) = 5/12$
- $P(B_3) = 1/4$

What is the probability that the ball came from B_1, given it is red?

Probability Theory

Conditional probability: $$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Partition Theorem: $$P(A) = \sum_i P(A \cap B_i)$$ If the B_i partition A

Bayes' Theorem: $$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Random Variables

A *Random Variable* is an object whose value is determined by chance, i.e. random events

Maps elements of Ω onto real numbers, with corresponding probabilities as specified by P

Random Variables

A *Random Variable* is an object whose value is determined by chance, i.e. random events

Maps elements of Ω onto real numbers, with corresponding probabilities as specified by *P*

Formally, a Random Variable is a function:

$$X : \Omega \rightarrow R$$

Random Variables

A *Random Variable* is an object whose value is determined by chance, i.e. random events

Maps elements of Ω onto real numbers, with corresponding probabilities as specified by *P*

Formally, a Random Variable is a function:

$$X : \Omega \rightarrow \mathrm{R}$$

Probability that the random variable X adopts a particular value x:

$$P(\{w \in \Omega : X(w) = x\})$$

Random Variables

A *Random Variable* is an object whose value is determined by chance, i.e. random events

Maps elements of Ω onto real numbers, with corresponding probabilities as specified by *P*

Formally, a Random Variable is a function:

$$X : \Omega \rightarrow \mathrm{R}$$

Probability that the random variable X adopts a particular value x:

$$P(\{w \in \Omega : X(w) = x\})$$

Shorthand: $P(X = x)$

Random Variables

Example:

If the result is *heads* then WIN – X takes the value 1
If the result is *tails* then LOSE – X takes the value 0

$$\Omega = \{heads, tails\}$$

$$X : \Omega \rightarrow \{0,1\}$$

$$P(X = x) = \begin{cases} P(\{heads\}) & x = 1 \\ P(\{tails\}) & x = 0 \end{cases}$$

$$P(X = x) = 1/2 \qquad x \in \{0,1\}$$

Random Variables

Example:

$$\Omega = \{one, two, three, four, five, six\}$$

Win £20 on a six, nothing on four/five, lose £10 on one/two/three

$$X : \Omega \rightarrow \{-10, 0, 20\}$$

Random Variables

Example:

$$\Omega = \{one, two, three, four, five, six\}$$

Win £20 on a six, nothing on four/five, lose £10 on one/two/three

$$X : \Omega \rightarrow \{-10, 0, 20\}$$

$$P(X = x) = \begin{cases} P(\{six\}) = 1/6 & x = 20 \\ P(\{four, five\}) = 1/3 & x = 0 \\ P(\{one, two, three\}) = 1/2 & x = -10 \end{cases}$$

Note – we are considering the probabilities of events in $\mathcal{F}$

Probability Mass Functions

Given a random variable:

$$X:\Omega \rightarrow A$$

The *Probability Mass Function* is defined as:

$$p_X(x) = P(X = x)$$

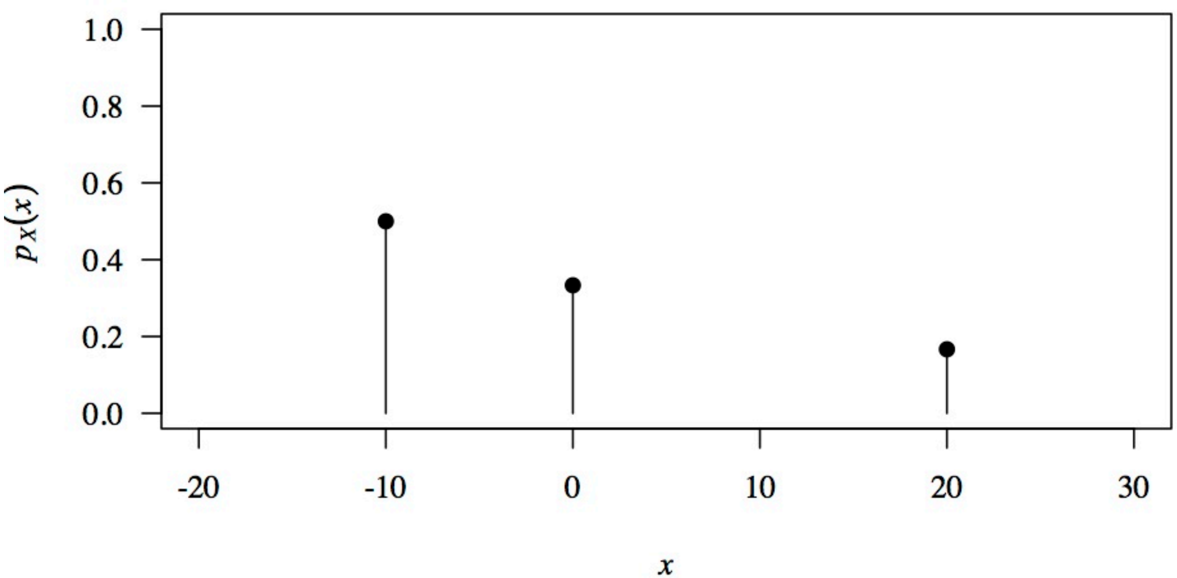
Only for discrete random variables

Probability Mass Functions

Example:

Win £20 on a six, nothing on four/five, lose £10 on one/two/three

$$p_X(x) = \begin{cases} P(\{six\}) = 1/6 & x = 20 \\ P(\{four, five\}) = 1/3 & x = 0 \\ P(\{one, two, three\}) = 1/2 & x = -10 \end{cases}$$

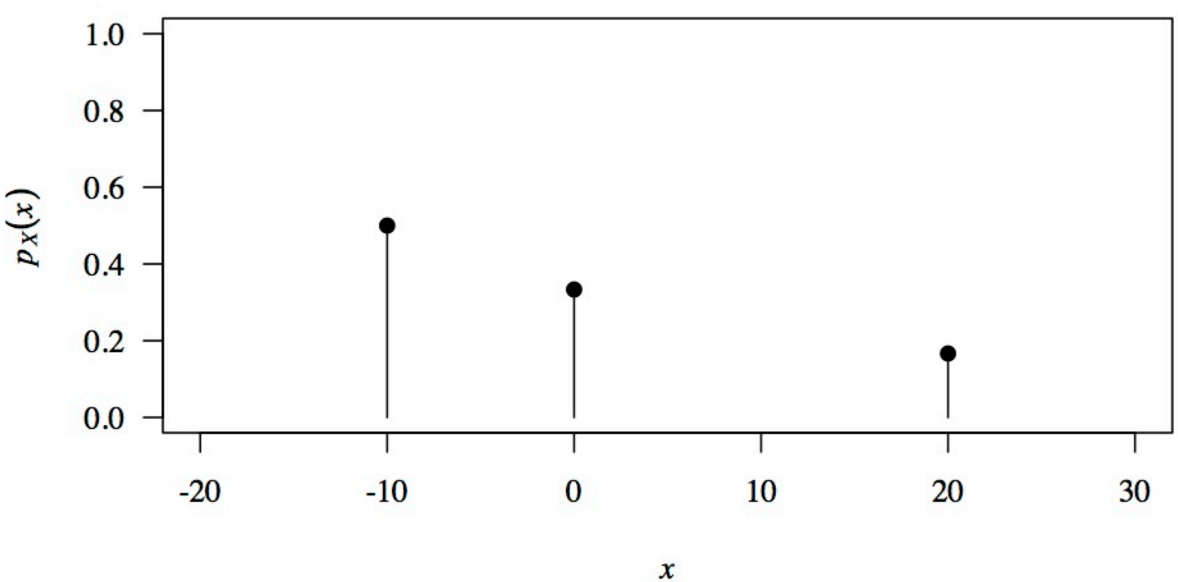


Probability Mass Functions

Notable properties of Probability Mass Functions:

$$p_X(x) \geq 0$$

$$\sum_{x \in A} p_X(x) = 1$$



Probability Mass Functions

Notable properties of Probability Mass Functions:

$$p_X(x) \geq 0$$

$$\sum_{x \in A} p_X(x) = 1$$

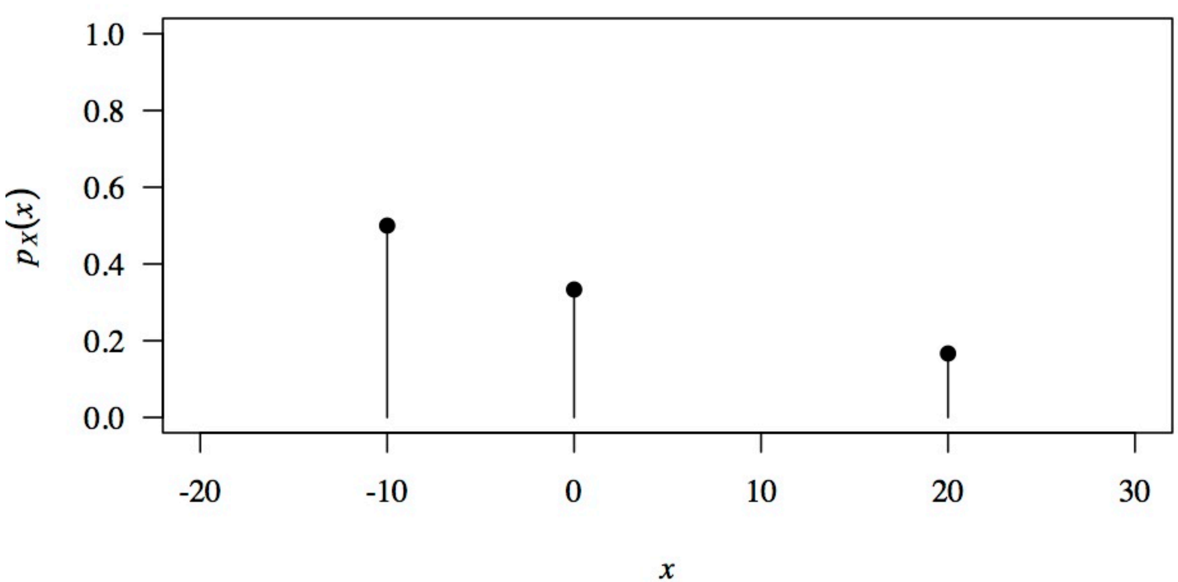
Interesting note:

If *p*() is some function that has the above two properties, then it is the mass function of some random variable…

Probability Mass Functions

For a random variable $X : \Omega \rightarrow A$

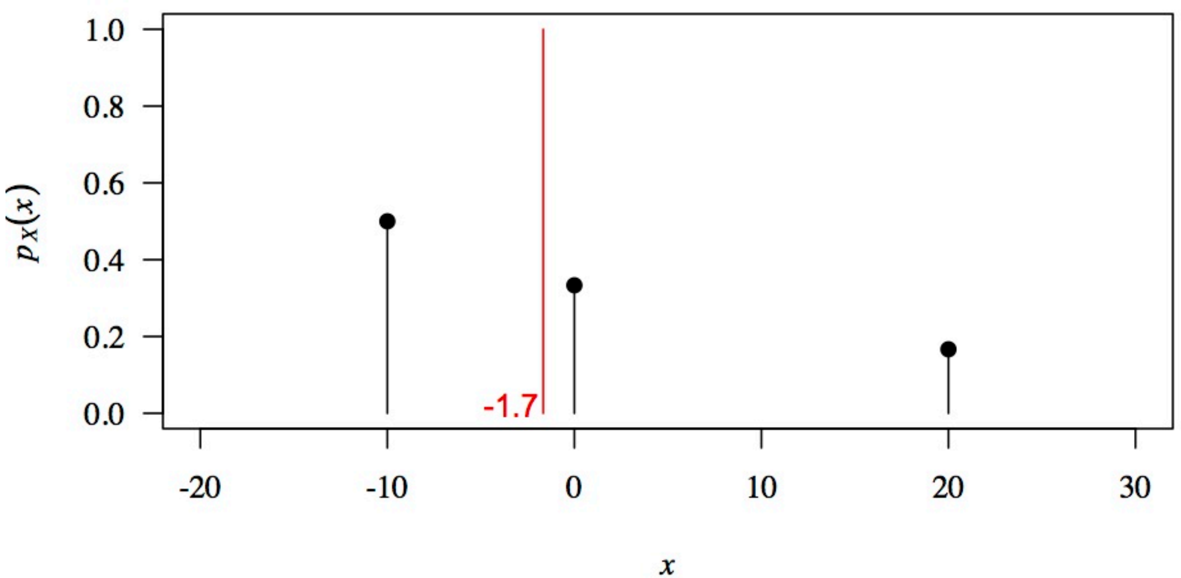
Mean: $$E(X) = \sum_{x \in A} x p_X(x)$$



Probability Mass Functions

For a random variable $X : \Omega \to A$

Mean: $$E(X) = \sum_{x \in A} x p_X(x)$$



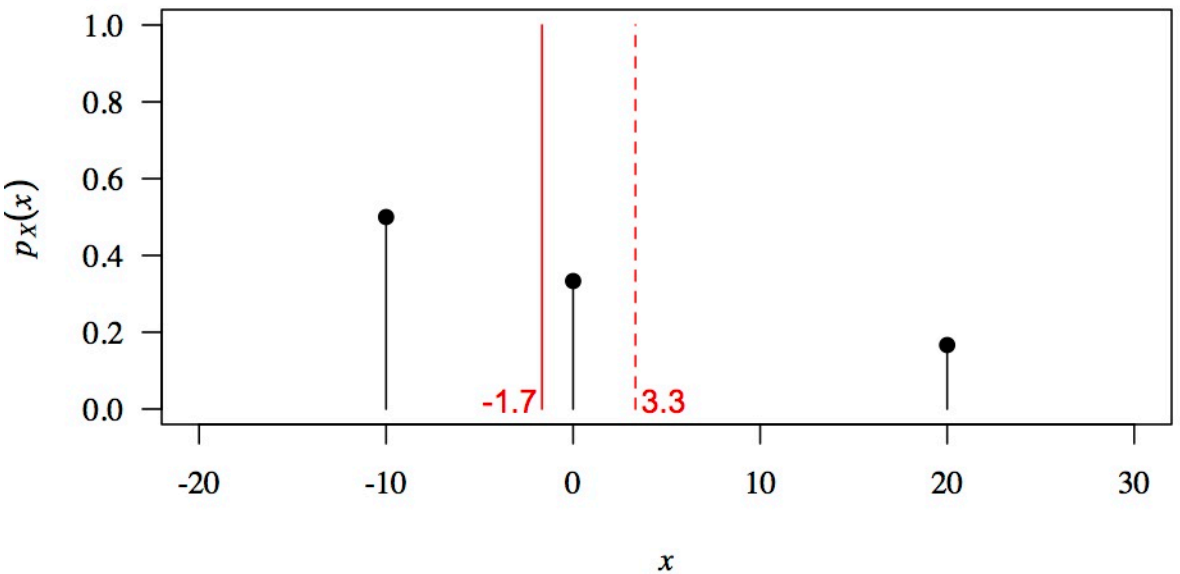
Mean

Probability Mass Functions

For a random variable $X : \Omega \to A$

Mean: $$E(X) = \sum_{x \in A} x p_X(x)$$

Compare with: $$\frac{1}{n}\sum_{i=1}^{n} x_i$$



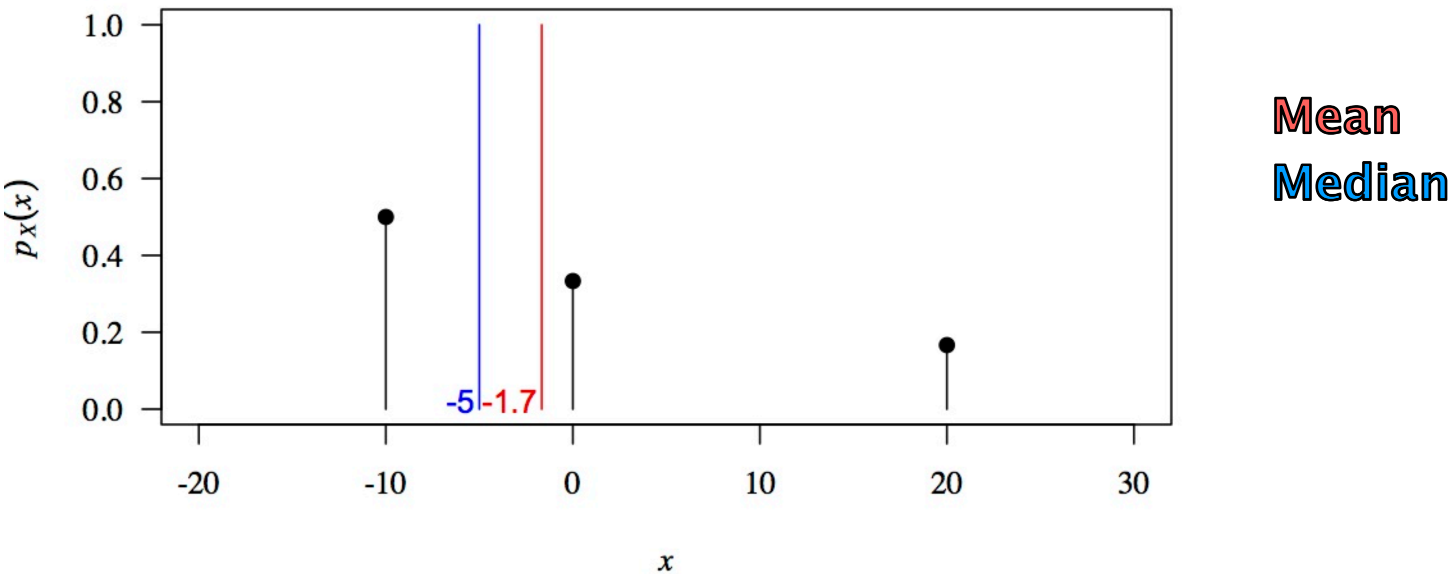
Mean

Probability Mass Functions

For a random variable $X : \Omega \rightarrow A$

Mean: $E(X) = \sum_{x \in A} x p_X(x)$

Median: any m such that: $\sum_{x \leq m} p_X(x) \geq 1/2$ and $\sum_{x \geq m} p_X(x) \geq 1/2$



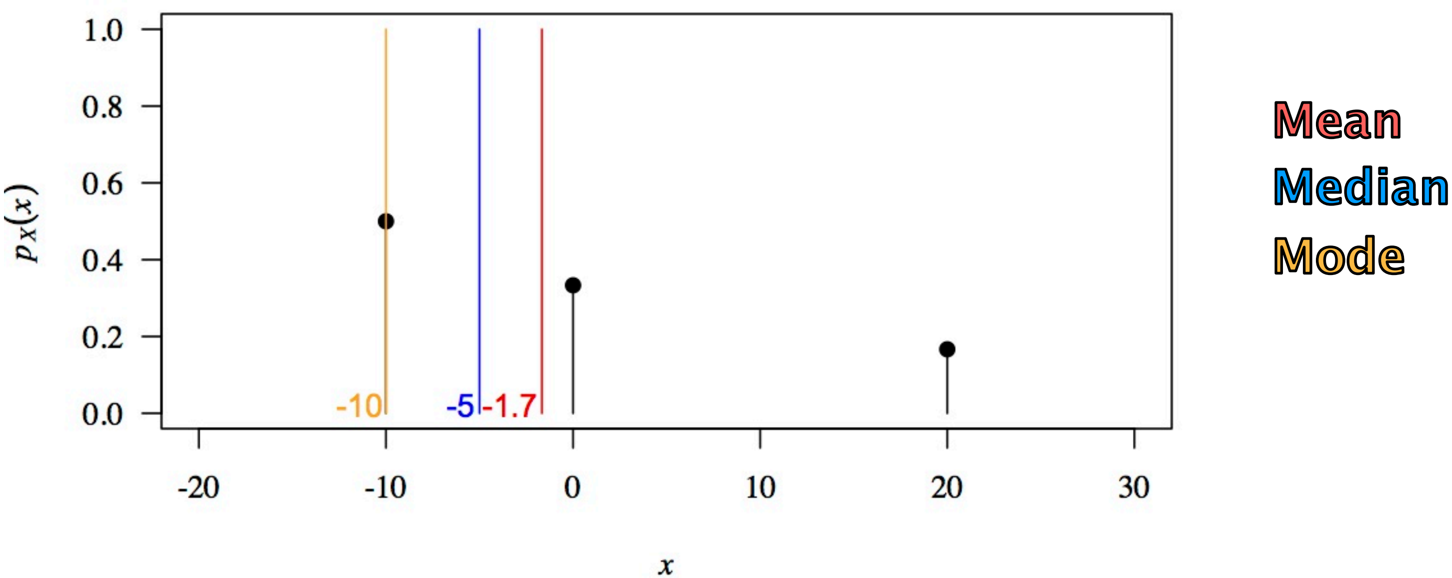
Probability Mass Functions

For a random variable $X : \Omega \rightarrow A$

Mean: $$E(X) = \sum_{x \in A} x p_X(x)$$

Median: any m such that: $\sum_{x \leq m} p_X(x) \geq 1/2$ and $\sum_{x \geq m} p_X(x) \geq 1/2$

Mode: $\underset{x}{\operatorname{argmax}} \left(p_X(x) \right)$: most likely value



Common Discrete Distributions

Common Discrete Distributions

The **Bernoulli** Distribution: $X \sim \mathrm{Bern}(p)$

p : success probability

$$X : \Omega \rightarrow \{0,1\}$$

$$p_X(x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \end{cases}$$

Common Discrete Distributions

The **Bernoulli** Distribution: $X \sim \mathrm{Bern}(p)$

p : success probability

$$X : \Omega \rightarrow \{0,1\}$$

$$p_X(x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \end{cases}$$

Example:

$$X : \{heads, tails\} \rightarrow \{0,1\}$$

$$p_X(x) = 1/2 \qquad x \in \{0,1\}$$

Therefore $X \sim \mathrm{Bern}(1/2)$

Common Discrete Distributions

The **Binomial** Distribution: $X \sim \mathrm{Bin}(n, p)$ $E(X) = np$

n : number of independent trials

p : success probability

$$X : \Omega \rightarrow \{0, 1, \ldots, n\}$$

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Common Discrete Distributions

The **Binomial** Distribution: $X \sim \mathrm{Bin}(n, p)$ $E(X) = np$

n : number of independent trials

p : success probability

$$X : \Omega \rightarrow \{0, 1, \ldots, n\} \qquad p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$p_X(x)$: probability of getting x successes out of n trials

p^x : probability of x successes

$(1-p)^{n-x}$: probability of *(n-x)* failures

$\binom{n}{x} = \frac{n!}{x!(n-x)!}$: number of ways to achieve x successes and *(n-x)* failures (Binomial coefficient)

Common Discrete Distributions

The **Binomial** Distribution: $X \sim \text{Bin}(n, p)$ $\quad E(X) = np$

n : number of independent trials

p : success probability

$$X : \Omega \rightarrow \{0, 1, \ldots, n\} \qquad p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$n = 1 : \qquad p_X(x) = p^x (1-p)^{1-x} = \begin{cases} p & x = 1 \\ 1-p & x = 0 \end{cases}$$

$$X \sim \text{Bin}(1, p) \quad \Leftrightarrow \quad X \sim \text{Bern(p)}$$

Common Discrete Distributions

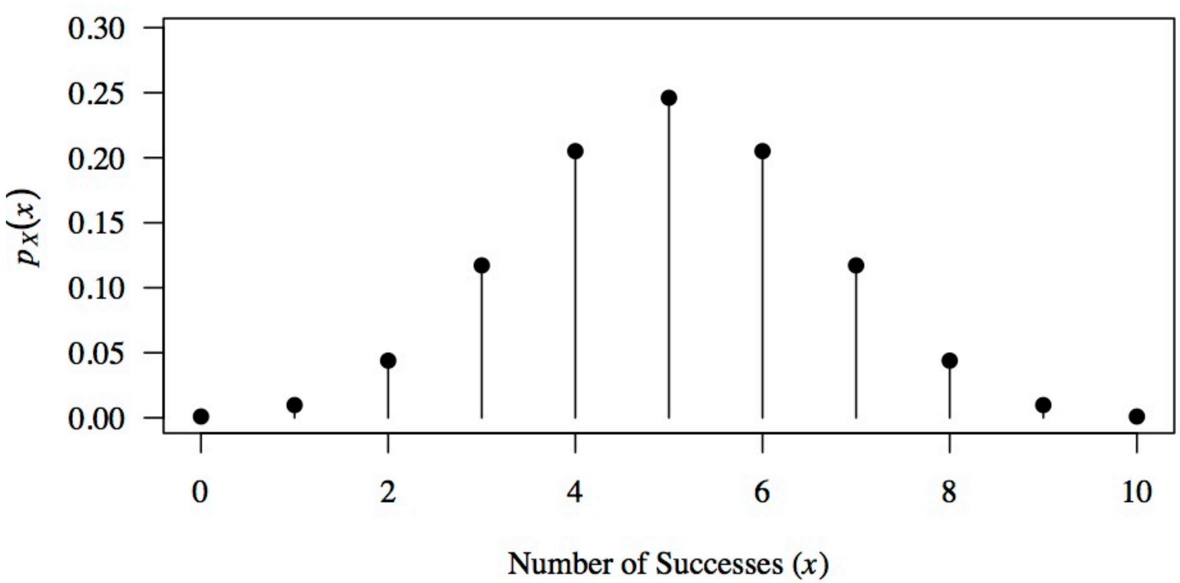
The **Binomial** Distribution: $X \sim \mathrm{Bin}(n, p)$ $E(X) = np$

n : number of independent trials

p : success probability

$$X : \Omega \rightarrow \{0, 1, \ldots, n\}$$

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$



$n = 10$

$p = 0.5$

Common Discrete Distributions

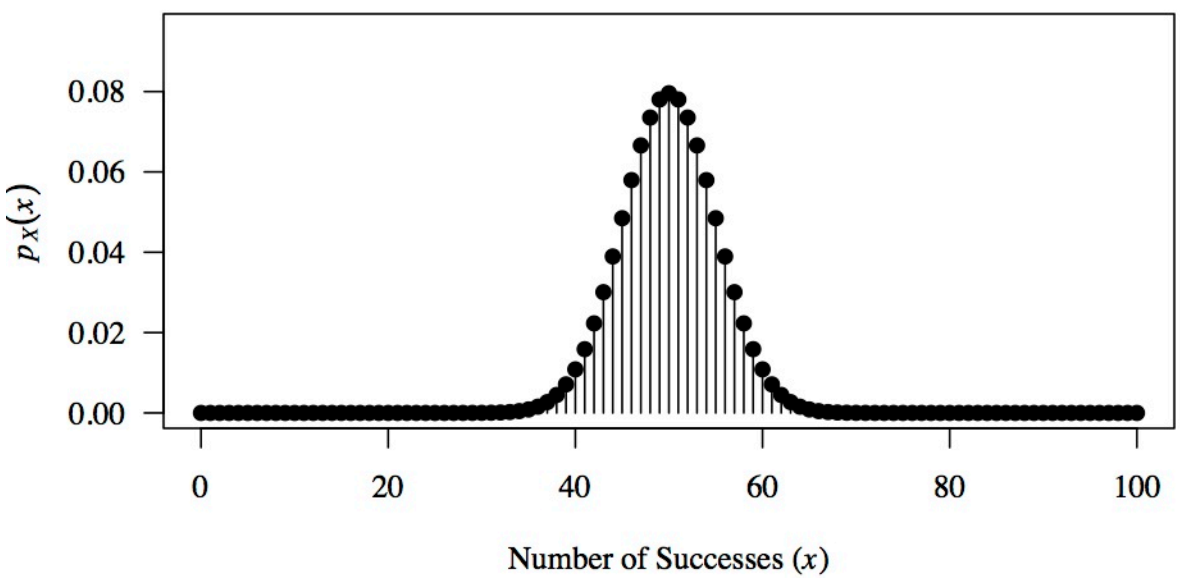
The **Binomial** Distribution: $X \sim \text{Bin}(n, p)$ $E(X) = np$

n : number of independent trials

p : success probability

$$X : \Omega \rightarrow \{0,1,\ldots,n\}$$

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$



$n = 100$

$p = 0.5$

Common Discrete Distributions

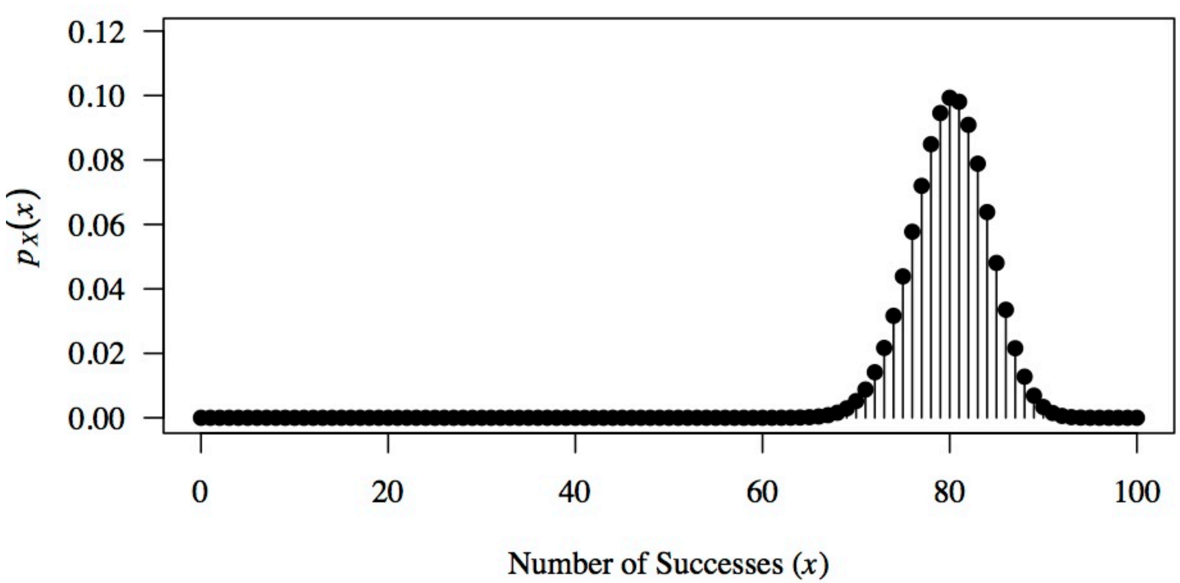
The **Binomial** Distribution: $X \sim \mathrm{Bin}(n, p)$ $E(X) = np$

n : number of independent trials

p : success probability

$$X : \Omega \rightarrow \{0,1,...,n\}$$

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$



$n = 100$

$p = 0.8$

Common Discrete Distributions

Example:

Number of heads in n fair coin toss trials

$$X : \Omega \rightarrow \{0,1,...,n\}$$

$n = 2$ $\quad\quad$ $\Omega = \{heads : heads, heads : tails, tails : heads, tails : tails\}$

In general: $\quad |\Omega| = 2^n$

Common Discrete Distributions

Example:

Number of heads in n fair coin toss trials

$$X : \Omega \rightarrow \{0,1,...,n\}$$

$n = 2$ $\quad\Omega = \{heads : heads, heads : tails, tails : heads, tails : tails\}$

In general: $\quad |\Omega| = 2^n$

Notice: $\quad X \sim \mathrm{Bin}(n, 1/2)$

$$p_X(x) = \binom{n}{x} 0.5^n \qquad E(X) = n/2$$

Common Discrete Distributions

The **Poisson** Distribution: $X \sim Pois(\lambda)$ $E(X) = \lambda$

Used to model the number of occurrences of an event
that occur within a particular interval of time and/or space

λ : average number of counts (controls rarity of events)

$$X : \Omega \rightarrow \{0,1,...\}$$

$$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Common Discrete Distributions

The **Poisson** Distribution:

- Want to know the distribution of the number of occurrences of an event $\Rightarrow$ Binomial?

- However, don't know how many trials are performed – could be infinite!

- But we do know the average rate of occurrence: $E(X) = \lambda$

$$X \sim \mathrm{Bin}(n, p) \quad \Rightarrow \quad E(X) = np$$

$$\Rightarrow \quad \lambda = np$$

$$\Rightarrow \quad p = \frac{\lambda}{n}$$

Common Discrete Distributions

Binomial: $$p_X(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$p = \frac{\lambda}{n} \quad \Rightarrow \quad p_X(x) = \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

Common Discrete Distributions

Binomial: $$p_X(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$p = \frac{\lambda}{n} \quad \Rightarrow \quad p_X(x) = \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$p_X(x) = \frac{\lambda^x}{x!} \frac{n!}{n^x (n-x)!} \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^x}$$

Common Discrete Distributions

Binomial: $$p_X(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$p = \frac{\lambda}{n} \quad \Rightarrow \quad p_X(x) = \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$p_X(x) = \frac{\lambda^x}{x!} \underbrace{\frac{n!}{n^x (n-x)!}}_{\to 1} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\underbrace{\left(1 - \frac{\lambda}{n}\right)^x}_{\to 1}} \qquad \text{as } n \to \infty$$

Common Discrete Distributions

Binomial: $$p_X(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$p = \frac{\lambda}{n} \quad \Rightarrow \quad p_X(x) = \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n} \right)^x \left(1 - \frac{\lambda}{n} \right)^{n-x}$$

$$p_X(x) = \frac{\lambda^x}{x!} \underbrace{\frac{n!}{n^x (n-x)!}}_{\to 1} \frac{\left(1 - \frac{\lambda}{n} \right)^n}{\underbrace{\left(1 - \frac{\lambda}{n} \right)^x}_{\to 1}} \qquad \text{as } n \to \infty$$

$$p_X(x) = \lim_{n \to \infty} \left(\frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n} \right)^n \right)$$

Common Discrete Distributions

Binomial: $$p_X(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$p = \frac{\lambda}{n} \quad \Rightarrow \quad p_X(x) = \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$p_X(x) = \frac{\lambda^x}{x!} \underbrace{\frac{n!}{n^x (n-x)!}}_{\to 1} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\underbrace{\left(1 - \frac{\lambda}{n}\right)^x}_{\to 1}} \qquad \text{as } n \to \infty$$

$$p_X(x) = \lim_{n \to \infty} \left(\frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \right) \qquad = \frac{\lambda^x}{x!} e^{-\lambda}$$

$\square$

Common Discrete Distributions

The Poisson distribution is the Binomial distribution as $n \to \infty$

If $X_n \sim \text{Bin}(n, p)$ then $X_n \xrightarrow{d} \text{Pois}(np)$

If n is large and p is small then the Binomial distribution can be approximated using the Poisson distribution

This is referred to as the:

- "Poisson Limit Theorem"
- "Poisson Approximation to the Binomial"
- "Law of Rare Events"

λ: fixed $\quad n \to \infty \quad \Rightarrow \quad p \to 0$

Poisson is often more computationally convenient than Binomial

References

Countless books + online resources!

Probability theory and distributions:

- Grimmett and Stirzker (2001) Probability and Random Processes. Oxford University Press.

General comprehensive introduction to (almost) everything mathematics (including a bit of probability theory):

- Garrity (2002) All the mathematics you missed: but need to know for graduate school. Cambridge University Press.