Chapter 2

Probability

Random experiments

- A random experiment is an action or process whose outcome is uncertain.

Examples: Roll dice, draw cards from shuffled decks, picking a person at random for a survey, choosing a census tract at random

- An (elementary) outcome is one of the possible outcomes of a random experiment.

Roll 2 dice: (2,6) is one possible outcome. 2 on first die, 6 on second

Roulette: 13 is a possible outcome.

Flip a coin five times: H,T,T,H,T is a possible outcome

Census: draw census tract CT34021 (Mercer County, NJ)

Modeling

Some experiments may not be really random. For example, the height of the flood in Holland depends on the moon, currents, temperature,... and many other parameters. It may still be helpful consider it as a random experiment. This is called modeling (we use a simple model for the truth). Other examples are

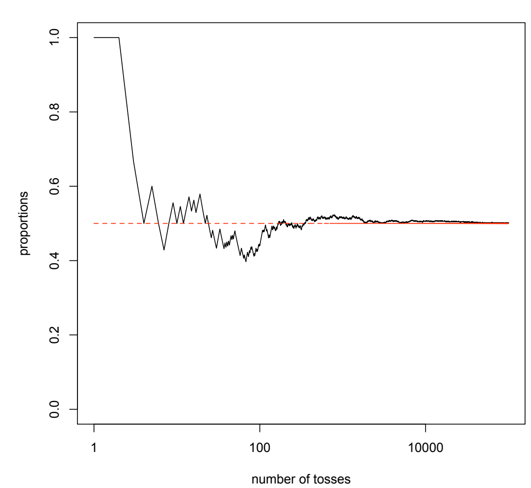
- measurement errors or rounding
- unobservable characteristics of individuals (drug testing)
- stock market

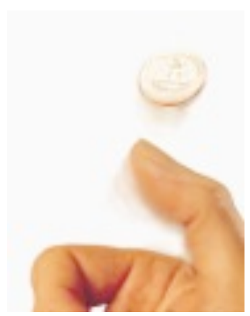
Probability

- The probability of an outcome is the proportion of times the event would occur is we observed the random experiment for an infinite number of repetitions.
- Justified by the law of large numbers

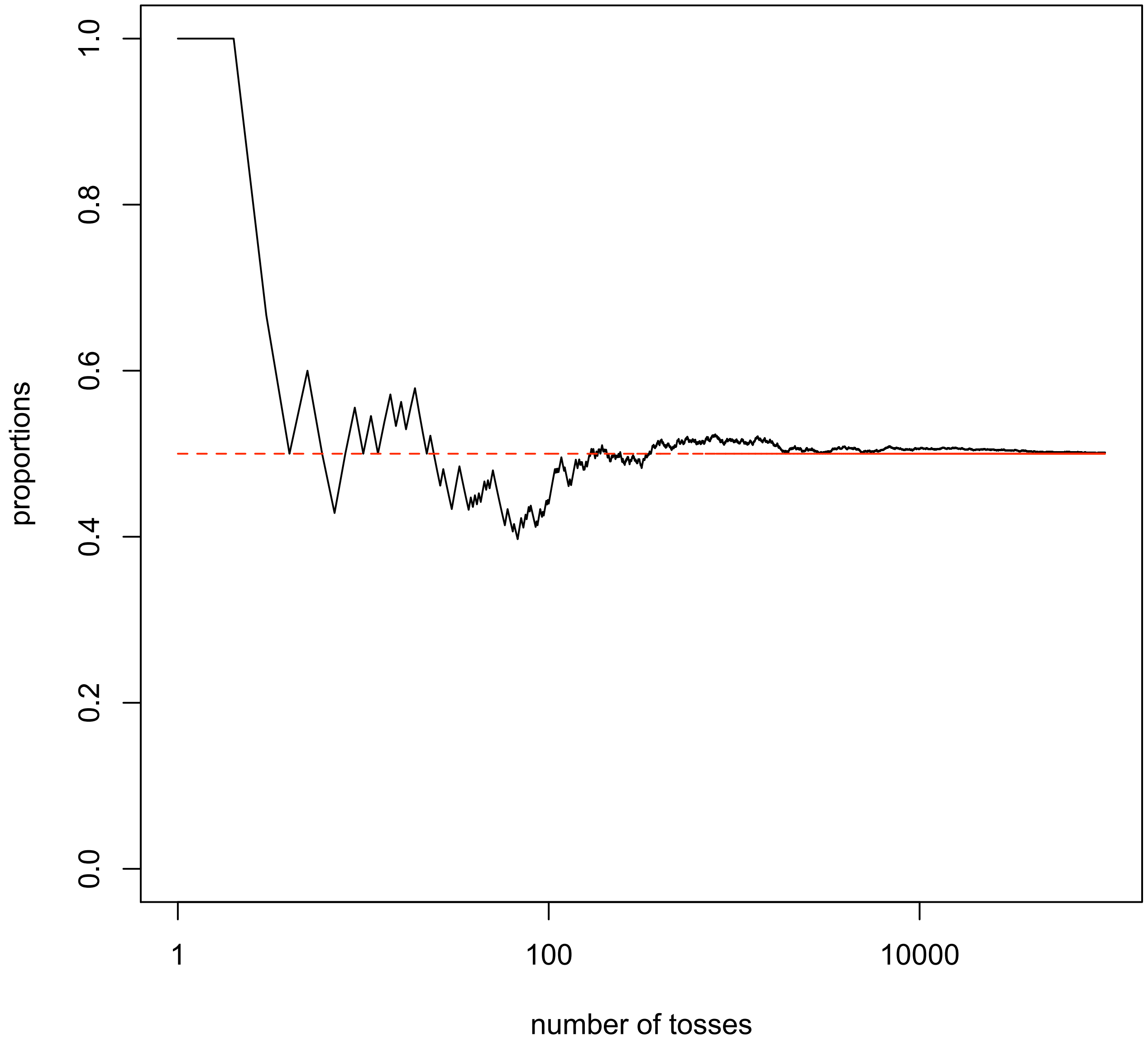
As the number of observations goes to infinity, the proportion of occurrences of a given outcome converges to the probability of this outcome.

Example: flipping a coin.





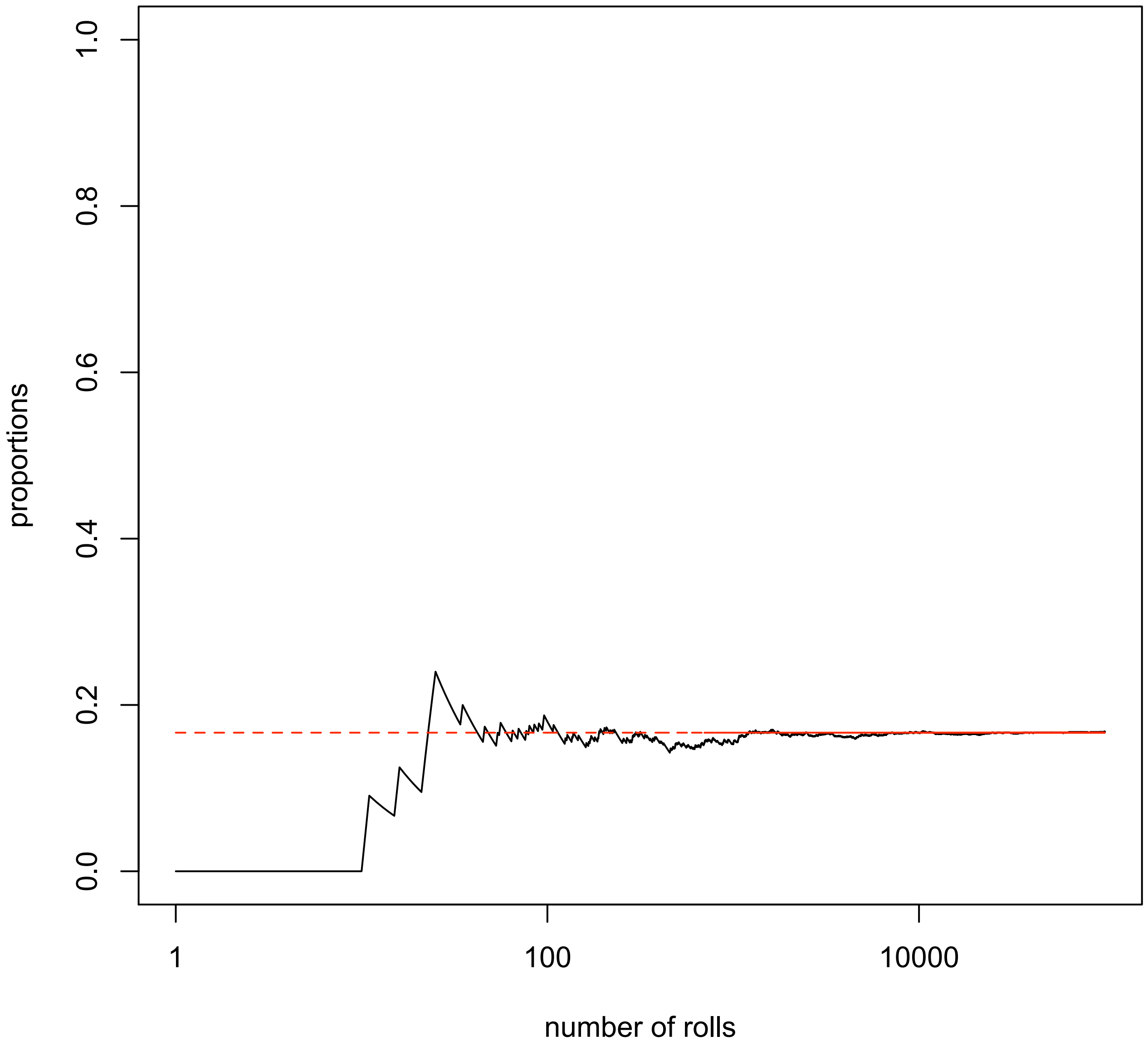
Flipping a coin



The proportion stabilizes around 1/2

```
coins=(runif(100000)<0.5)
proportions=cumsum(coins)/(1:100000)
plot(proportions, log="x", type="l", ylim=c(0,1), xlab="number of tosses")
lines(1:100000, (1:100000)*0+0.5, col=2, lty=2)
```

Rolling a die



The proportion stabilizes around 1/6

```
face_one=(runif(100000)<(1/6))
proportions=cumsum(face_one)/(1:100000)
plot(proportions, log="x", type="l", ylim=c(0,1), xlab="number of rolls")
lines(1:100000, (1:100000)*0+1/6, col=2, lty=2)
```

Probability

- We write P(outcome) the probability of an outcome.

  Rolling a die: P(2)=1/6
  Flipping a coin: P(H)=1/2

- If all outcome are equally likely then

$$\mathrm{P(outcome)} = \frac{1}{\text{number of outcomes}}$$

- Consider a more complicated example: rolling 2 dice. What is P(2,6)? It is the same as P(1,1) or P(3,4). All outcomes are equally likely. We need to count the outcomes or have rules to compute probabilities.

Events

An event is a collection of outcomes. It can be described either with words or using formal notation from set theory. Passing from the first one to the second is a necessary skill.

Flipping two coins. We know that the outcomes are (H,H), (H,T), (T, H), (T,T). Consider the **events**

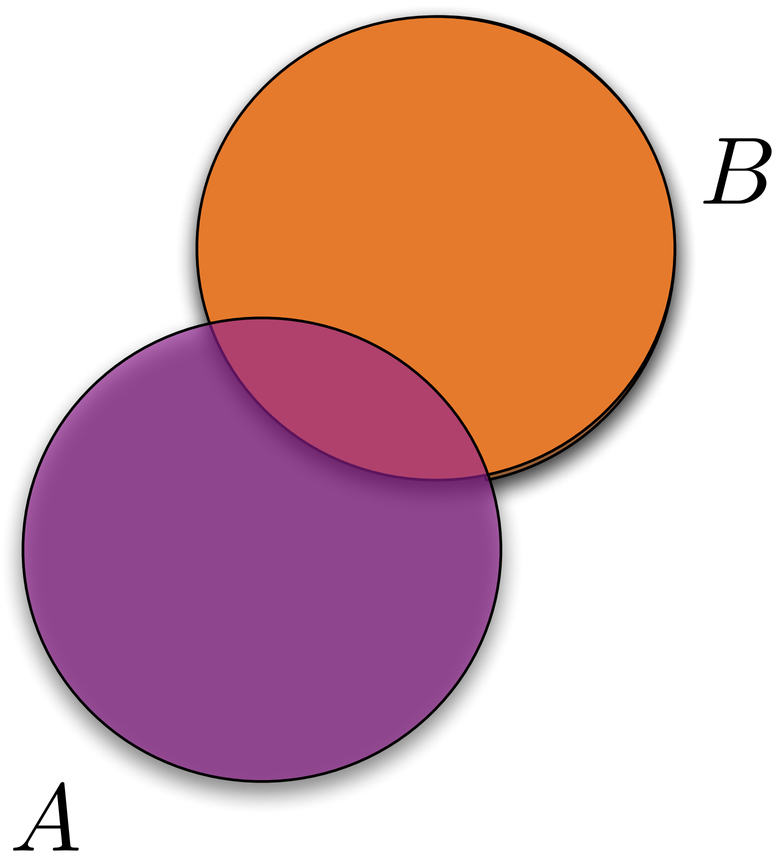
{twice the same}={(H,H), (T,T)}

{heads first}={(H,T), (H,H)}

{no heads}={(T,T)}

We want to find rules to compute the probability of events from the probability of outcomes.

Operations on events



Operations on events

Union of two events A and B:

$A \cup B$ = {outcomes that are either in A or in B or in both}

Intersection of two events A and B:

$A \cap B$= {outcomes that in A and in B}

Complement of the event A:

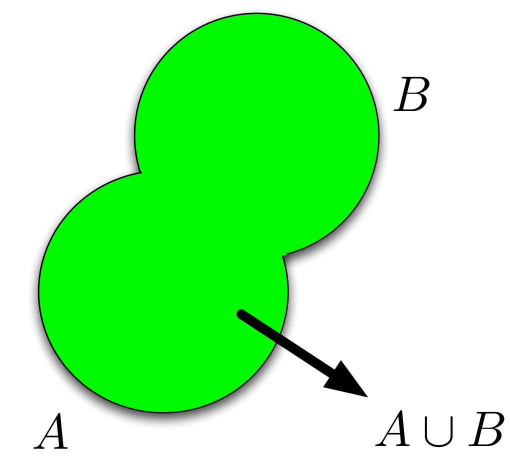
A^c = {outcomes that are not in A}

Operations on events

Union

$A \cup B$

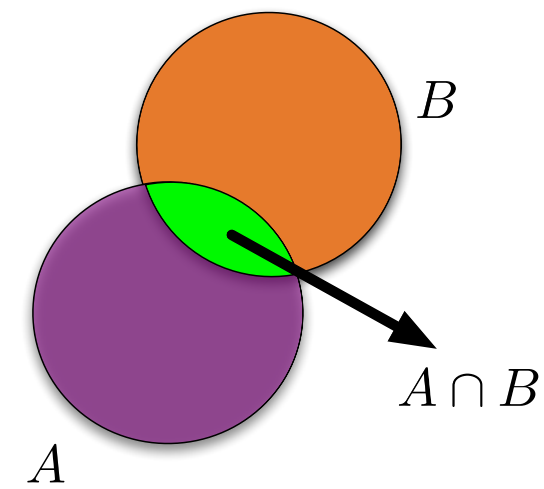
A "or" B



Intersection

$A \cap B$

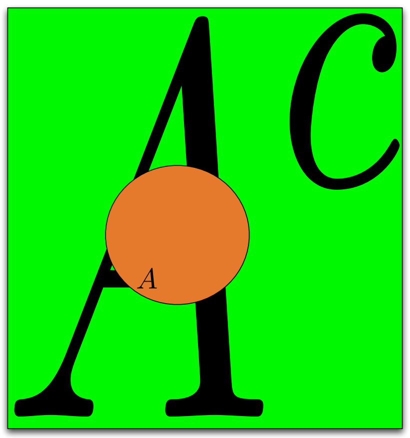
A "and" B



Complement

A^c

"not" A



Operations on events

Flipping two coins.

A={twice the same}={(H,H), (T,T)}
B={heads first}={(H,T), (H,H)}
C={no heads}={(T,T)}

$A \cup B$ ={(H,H), (T,T), (H,T)}
={twice the same or heads first}

$A \cap B$ ={(H,H)}
={twice the same and heads first}
={twice heads}

C^c ={ }
={at least one heads}

Addition rule of disjoint events

Two events are disjoint if they have no outcome in common. For example:
{(H,H), (T,T)} and {(H,T), (H,T)} are disjoint
{(H,H), (T,T)} and {(H,T), (H,H)} are not disjoint

Equivalently, two events are disjoint if their intersection is empty (no outcome)

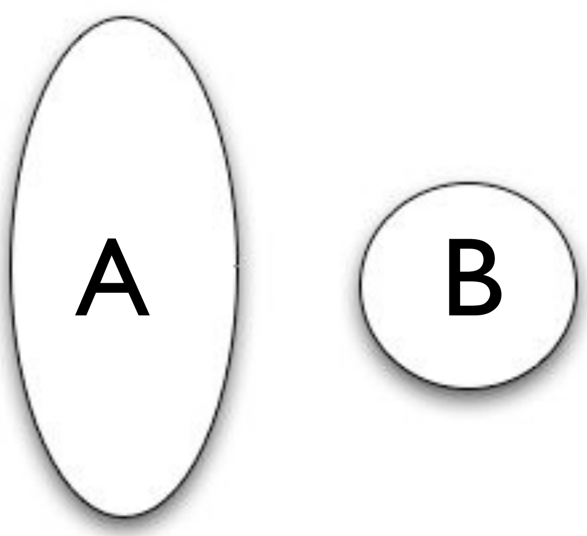
For disjoint events A and B we have the addition rule

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

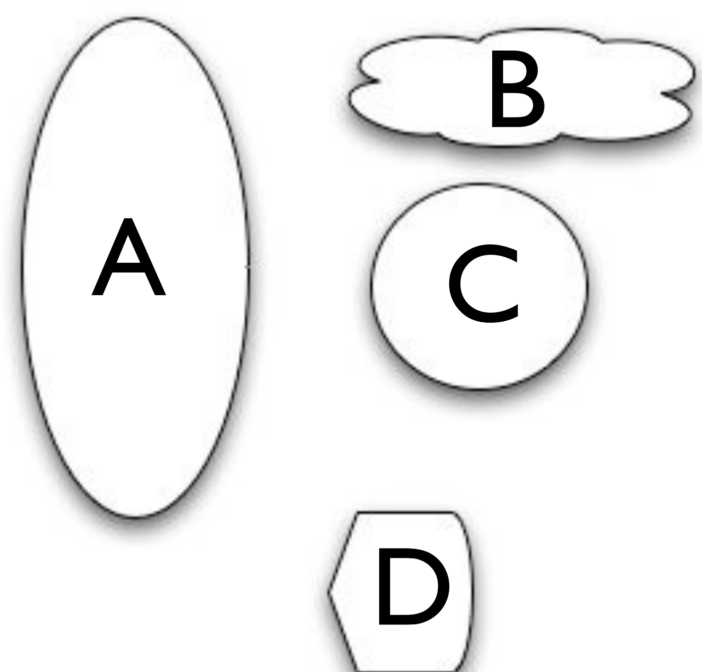
Addition rule of disjoint events

More generally, if k events $A_1, A_2, \ldots, A_k$ are disjoint

$$P(A_1 \cup A_2 \cup \ldots \cup A_k) = P(A_1) + P(A_2) + \ldots + P(A_k)$$



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



$P(A \cup B \cup C \cup D) = P(A) + P(B) + P(C) + P(D)$

Addition rule of disjoint events

Example: Tossing five coins at random. What is the probability of at least four heads?

Possible outcomes={HHHHH, THHHH, HTHHH, HHTHH, HHHTH, HHHHT, TTHHH,...,TTTTT}
There are 2^5 possible outcomes (2 for each round)

P(at least 4 heads) =P(4 heads or 5 heads)
=P(4 heads)+P(5 heads)
=1/32 + 5/32 = 6/32=3/16

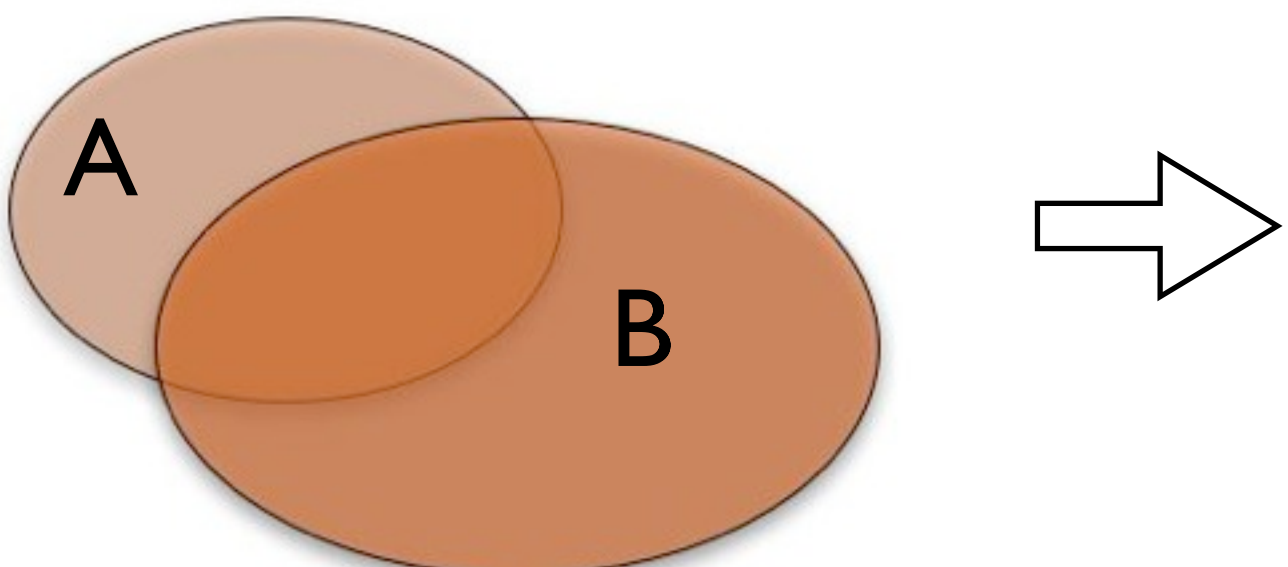
General addition rule

If two events are not disjoint, there is still a rule:

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

If A and B are disjoint, it is impossible to have an outcome in $A \cap B$ and thus $P(A \cap B) = 0$

This rule can be illustrated with the following diagram:



Two layers on the intersection

Equally likely outcomes

We already know that for equally likely outcomes:

$$\text{P(outcome)} = \frac{1}{\text{number of outcomes}}$$

Using the addition rule for disjoint events we find that for an event $E = \{\omega_1, \ldots, \omega_k\} = \{\omega_1\} \cup \cdots \cup \{\omega_k\}$

$$\begin{aligned} P(E) &= P(\omega_1) + \cdots + P(\omega_k) \\ &= kP(\omega_1) \\ &= \frac{\text{number of outcomes in } E}{\text{total number of outcomes}} \end{aligned}$$

Equally likely outcomes

We already know that for equally likely outcomes:

$$\mathrm{P(outcome)} = \frac{1}{\text{number of outcomes}}$$

Using the addition rule for disjoint events we find that for an event $E = \{\omega_1, \ldots, \omega_k\} = \{\omega_1\} \cup \cdots \cup \{\omega_k\}$

$$\begin{aligned} P(E) &= P(\omega_1) + \cdots + P(\omega_k) \\ &= kP(\omega_1) \\ &= \frac{\text{number of outcomes in } E}{\text{total number of outcomes}} \end{aligned}$$

➡ Need to **count** outcomes (in E and total)

Counting rules

Product rule. If the experiment consists of several consecutive parts where the first part has n_1 outcomes, the second part has n_2 outcomes,... then the total number of outcomes is given by the product

$$n_1 \cdot n_2 \cdots$$

Example. Rolling two dice.
What is the probability of getting two numbers less than or equal to 2?

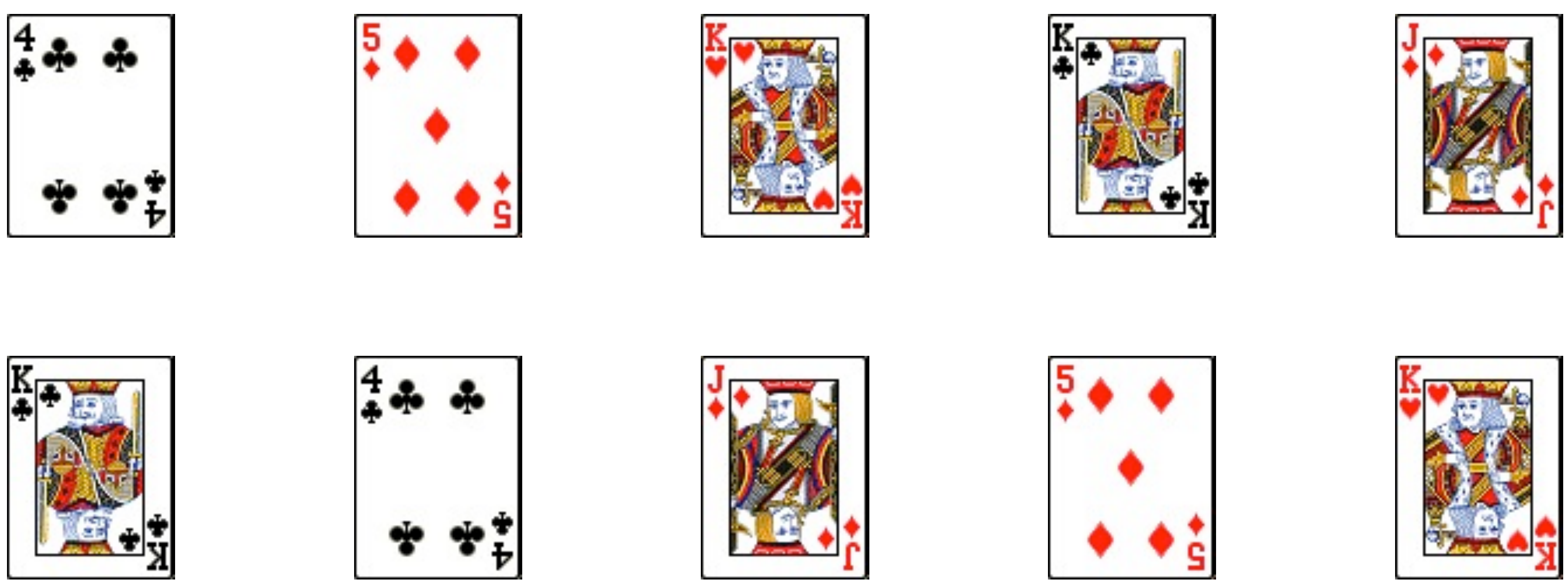
- Total number of outcomes: 6*6=36
- Number of outcomes in event: 2*2=4

Answer is 4/36=1/9

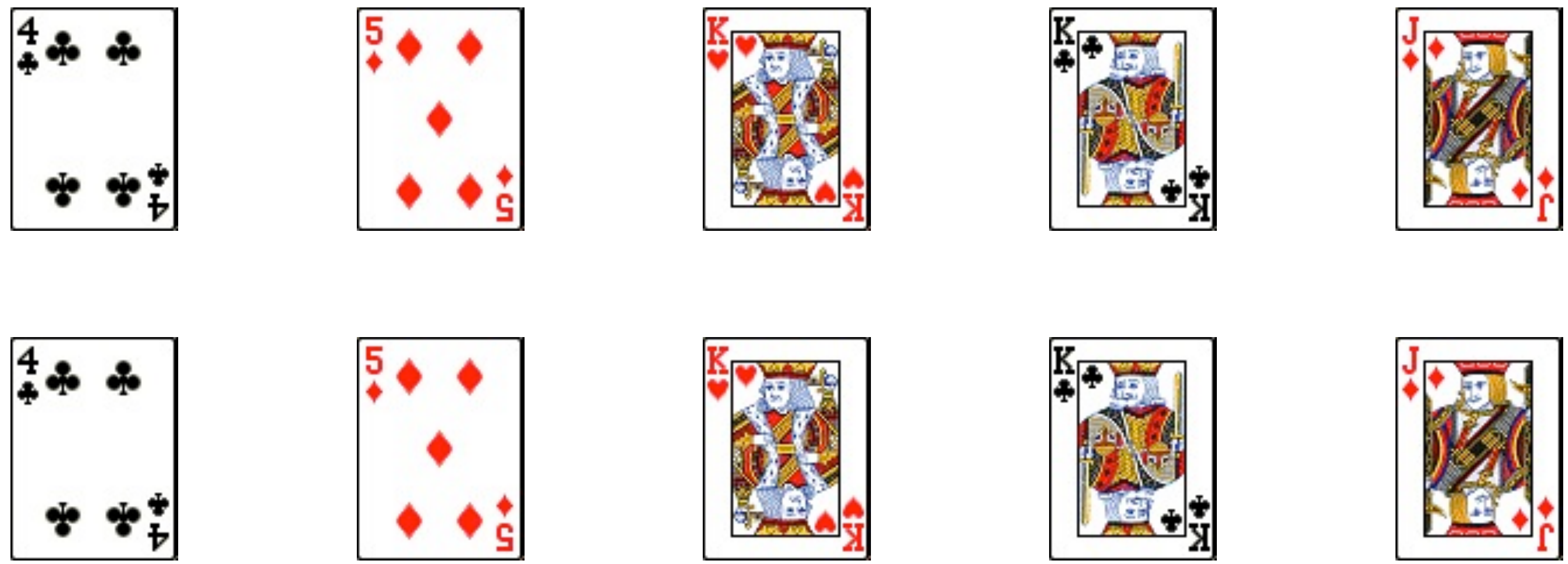
Counting rules

The product rule can take you a long way but here is a very convenient shortcut suitable for many experiments.

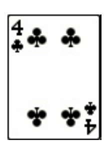
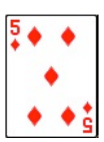
Example. Drawing 5 cards from a 52-card deck. How many hands are there?
We could use the product rule: 52 ways of choosing the first card, 51 ways of choosing the second card,... 52*51*50*49*48



Counting rules



We have counted the same hand several times. How many?

5 positions for , 4 positions for , 3 positions for , ...
that's 5*4*3*2*1.

Finally we obtain that the number of such hands is

$$\frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}$$

Counting rules

For the general case: define "factorial k" by

$$k! = k \cdot (k-1) \cdot (k-2) \cdots 2 \cdot 1$$

We have $$\frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = \frac{\frac{52!}{47!}}{5!} = \frac{52!}{(52-5)!5!}$$

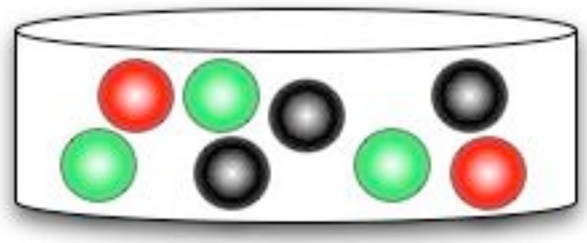
Combination rule. The number of outcomes (combinations) obtained when selecting k different objects from a set of n objects is given by

$$\binom{n}{k} := \frac{n!}{(n-k)!k!}$$

"n choose k"

Counting rules

Exercise. Three balls are selected at random from the jar below.



What is the probability of getting exactly one red ball and two green balls?

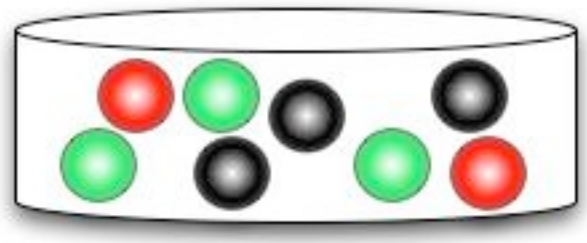
Total number of outcomes: $\binom{8}{3} = \frac{8!}{5!3!} = 56$

Number of outcomes in the event {1 red, 2 greens}:
Choosing the red ball: 2 possibilities
Choosing the green balls: 2 among 3: $\binom{3}{2}$=3

Number of outcomes in the event (product rule): $2 \cdot 3 = 6$

Counting rules

Exercise. Three balls are selected at random from the jar below.

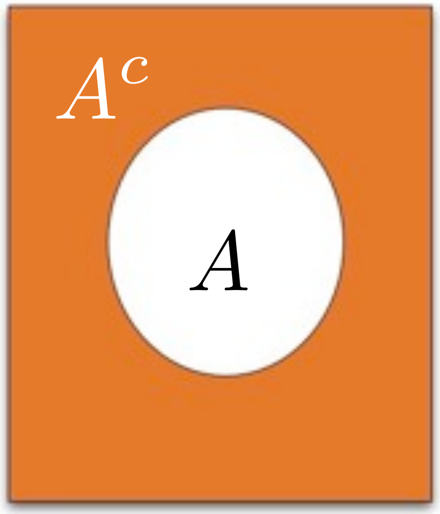


What is the probability of getting exactly one red ball and two green balls?

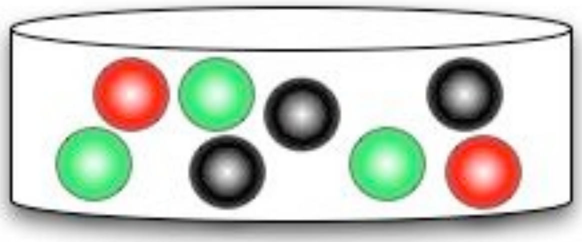
Probability of the event {1 red, 2 greens}: $\frac{6}{56} = \frac{3}{28}$

Rule of complement

$$P(A^c) = 1 - P(A)$$



Exercise. Three balls are selected at random from the jar below.



What is the probability that at least one ball is black or red?
Define the event A={at least one ball is black or red}
Then A^c={The 3 balls are green}

We have $P(A^c) = \dfrac{1}{56}$

So that $P(A) = 1 - P(A^c) = \dfrac{55}{56}$

Independence

Two events A and B are called independent if knowing one does not affect the other. This happens when A and B pertain to two independent experiment

Examples:

- Rolling 2 dice: A={Die one is ⚂} B={Die two is ⚄}
- Two consecutive hands in poker:
  A={Full house in 1st hand} B={two pairs in 2nd hand}
- Sampling randomly two students on campus:
  A={1st student is ORFE} B={2nd student is ORFE}

⚠ Sampling from **small** population $\not\Rightarrow$ independence.

Product rule for independent events

If two events A and B are independent, then

$$P(A \cap B) = P(A) \cdot P(B)$$

Exercises.

1.Rolling two dice. What is the probability of ⚅⚅?

P(⚅⚅)=P(⚅)P(⚅)= $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$

2.Two students are selected at random: mike and jake. P(CS)=60%. What is the probability one is CS and the other one is not?

P(jake is CS and mike is not CS)=P(CS)P(not CS)
=.6(1-.6)=24% (= P(mike is CS and jake is not CS))
P(one is CS and one is not)=2*24%=48% (addition rule).

Conditional probability

What if the events A and B are not independent?
That is: the outcome of A affects the outcome of B and vice-versa.

Example: Consider the contingency table for the Titanic.

Survived

Class

	No	Yes	Total
1st class	122	203	325
2nd class	167	118	285
3rd class	528	178	706
crew	673	212	885
Total	1,490	711	2,201

Survival is not independent of social status!

What is the probability that a randomly selected name on the passenger manifest corresponds to someone in 1st class who survived? ⟹ P(1st class AND survived)?

Conditional probability

We can define the **conditional probability** of A given B by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

We can also define the conditional probability of B given A by

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

Note that $P(B|A) \neq P(A|B)$ if $P(A) \neq P(B)$

In the Titanic example we still need to know either P(1st Class | survived) or P(Survived | first class).

Properties

- If A and B are independent:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

Indeed, knowing B does not affect the probability of A

- For any events A and B (not necessarily independent):

$$P(A \cap B) = P(A|B)P(B)$$

- For any A and B we can find $P(B|A)$ from $P(A|B)$ by:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = P(A|B)\frac{P(B)}{P(A)}$$

Properties

P(• | C) is also a probability. In other words, it satisfies all the rules of a usual probability.

- Addition rule for disjoint events:

$$P(A \cup B|C) = P(A|C) + P(B|C)$$

- General addition rule:

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

- Rule of complement:

$$P(A^c|C) = 1 - P(A|C)$$

- Multiplication rule for independent events:

$$P(A \cap B|C) = P(A|C)P(B|C)$$

Back on the Titanic

Survived

Class

	No	Yes	Total
1st class	122	203	325
2nd class	167	118	285
3rd class	528	178	706
crew	673	212	885
Total	1,490	711	2,201

P(1st class AND survived)= $\frac{203}{2201}$

P(survived | 1st class)= $\frac{203}{325}$

P(1st class)= $\frac{325}{2201}$

P(1st class AND survived)= P(survived | 1st class) P(survived | 1st class)=

$$\frac{203}{325}\frac{325}{2201}=\frac{203}{2201}$$

The game show

Suppose you're on a game show. You're given a chance to choose from 3 different doors.

The game show

Suppose you're on a game show. You're given a chance to choose from 3 different doors.

The game show

Suppose you're on a game show. You're given a chance to choose from 3 different doors.

Behind one of the doors is a
behind the other two:

The game show

Suppose you're on a game show. You're given a chance to choose from 3 different doors.

- Behind one of the doors is a 
- behind the other two:

The game show

Suppose you're on a game show. You're given a chance to choose from 3 different doors.

- Behind one of the doors is a 
- behind the other two:  

The game show

Suppose you're on a game show. You're given a chance to choose from 3 different doors.

Behind one of the doors is a 

behind the other two:  

Which door would you choose?

The game show

You've chosen door number 1 for example.

The game show

You've chosen door number 1 for example.

Next the game show host (who knows what's behind each door) opens a door with a goat: say door #3.

Is it in your interest to switch to door #2?

The game show

To solve the problem, define the events:

C1: {the **car** is behind door 1}
C2: {the **car** is behind door 2}
C3: {the **car** is behind door 3}

H3: {the host opens door 3}.

Simulation exercise: Recover these probabilities from simulation.

Compute: P(C1 | H3) and P(C2 | H3). Which is larger?

All we know is
P(H3 | C1)= , P(H3 | C2)= , P(H3 | C3)=

Random variables

A **random variable** X is a variable whose outcome is random. It is obtained my measuring the outcome of a random experiment.

Examples:

Random variables

A **random variable** X is a variable whose outcome is random. It is obtained my measuring the outcome of a random experiment.

Examples:

1. The number on the face of a die

Random variables

A **random variable** X is a variable whose outcome is random. It is obtained my measuring the outcome of a random experiment.

Examples:

1. The number on the face of a die
2. The GPS of a randomly selected (r.s.) student

Random variables

A **random variable** X is a variable whose outcome is random. It is obtained my measuring the outcome of a random experiment.

Examples:

1. The number on the face of a die
2. The GPS of a randomly selected (r.s.) student
3. The fat content of a r.s. hoagie

Random variables

A **random variable** X is a variable whose outcome is random. It is obtained my measuring the outcome of a random experiment.

Examples:

1. The number on the face of a die
2. The GPS of a randomly selected (r.s.) student
3. The fat content of a r.s. hoagie
4. {0,1} that indicates if a drug cures a r.s. patient

Random variables

A **random variable** X is a variable whose outcome is random. It is obtained my measuring the outcome of a random experiment.

Examples:

1. The number on the face of a die
2. The GPS of a randomly selected (r.s.) student
3. The fat content of a r.s. hoagie
4. {0,1} that indicates if a drug cures a r.s. patient
5. The median home value of a r.s US census tract

Random variables

A **random variable** X is a variable whose outcome is random. It is obtained my measuring the outcome of a random experiment.

Examples:

1. The number on the face of a die
2. The GPS of a randomly selected (r.s.) student
3. The fat content of a r.s. hoagie
4. {0,1} that indicates if a drug cures a r.s. patient
5. The median home value of a r.s US census tract
6. The gas milage of a r.s. car made in 2010

Random variables

In statistics, we think of our (numerical) data as the realization of random variables (after the random experiment has taken place).

For example, the list of numbers in medv are realizations of the random variable MEDV.

Note that we typically use uppercase for random variables and lowercase for the realizations.

Random variables can also be split into:

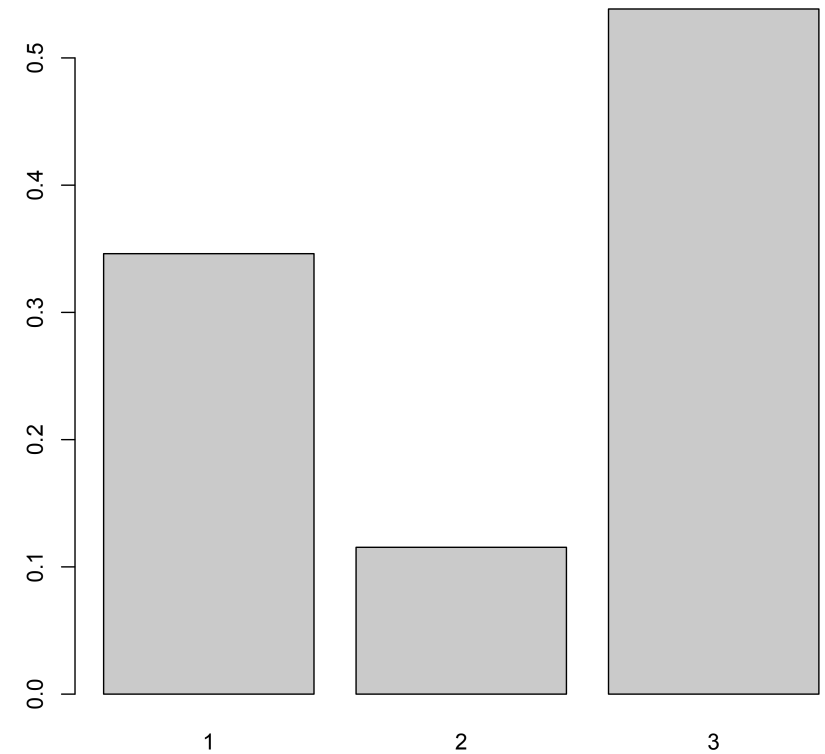
discrete and continuous

Random variables

discrete

Takes values 0, 1, 2, ...
Completely described by
P(X=1), P(X=2), ...

```
x=c(1, 1, 1,1,1,1,1,1,1,2,2,2,3,3,3,3,3,3,3,3,3,3,3,3,3,3)
barplot(prop.tab
```



continuous

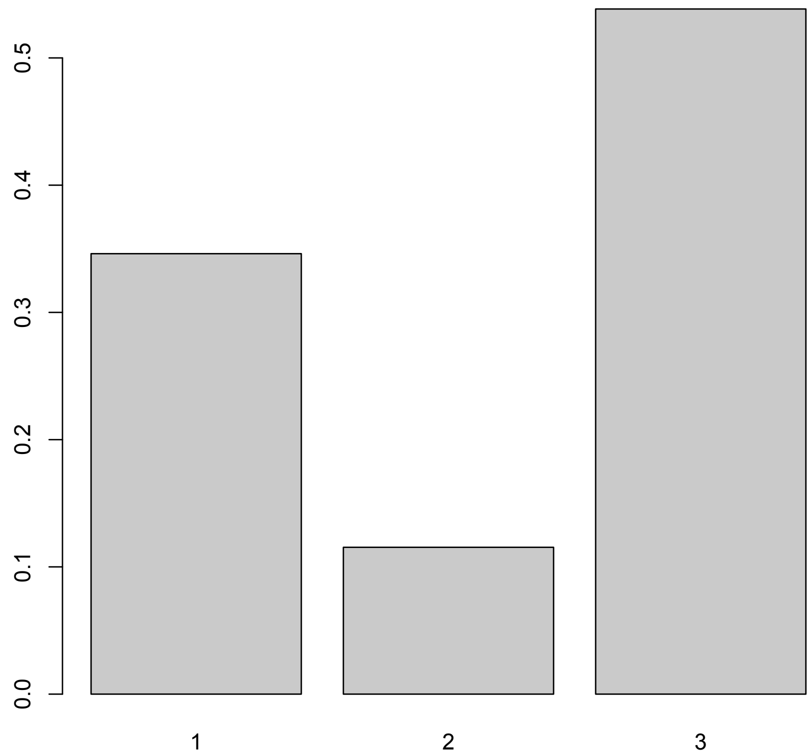
Takes a continuum of values (e.g. interval [a, b])
Can be seen as the limit of a histogram when the bin size goes to 0.

Random variables

discrete

Takes values 0, 1, 2, ...
Completely described by
P(X=1), P(X=2), ...

```
x=c(1, 1, 1,1,1,1,1,1,1,2,2,2,3,3,3,3,3,3,3,3,3,3,3,3,3,3)
barplot(prop.tab
```



continuous

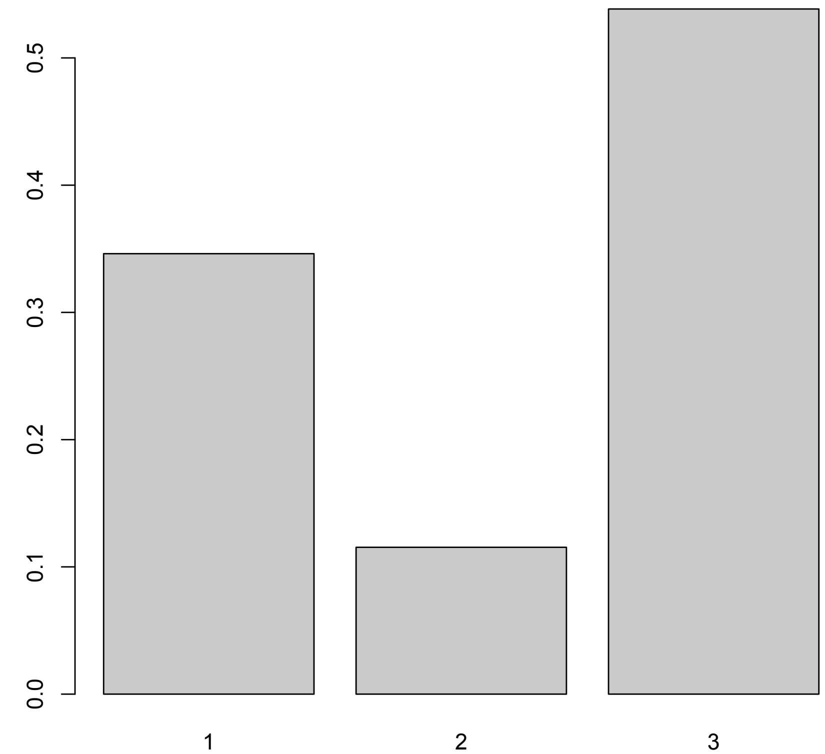
Takes a continuum of values (e.g. interval [a, b])
Can be seen as the limit of a histogram when the bin size goes to 0.

Four histograms of US adults heights with varying bin widths.

Random variables

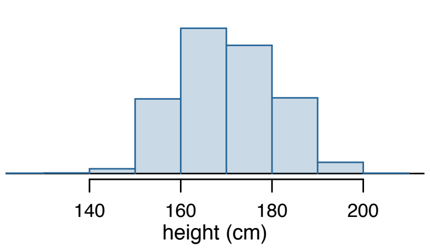
discrete Takes values 0, 1, 2, ...
Completely described by
P(X=1), P(X=2), ...

```
x=c(1, 1, 1,1,1,1,1,1,1,2,2,2,3,3,3,3,3,3,3,3,3,3,3,3,3,3)
barplot(prop.tab
```



continuous Takes a continuum of values (e.g. interval [a, b])
Can be seen as the limit of a histogram when the bin size goes to 0.

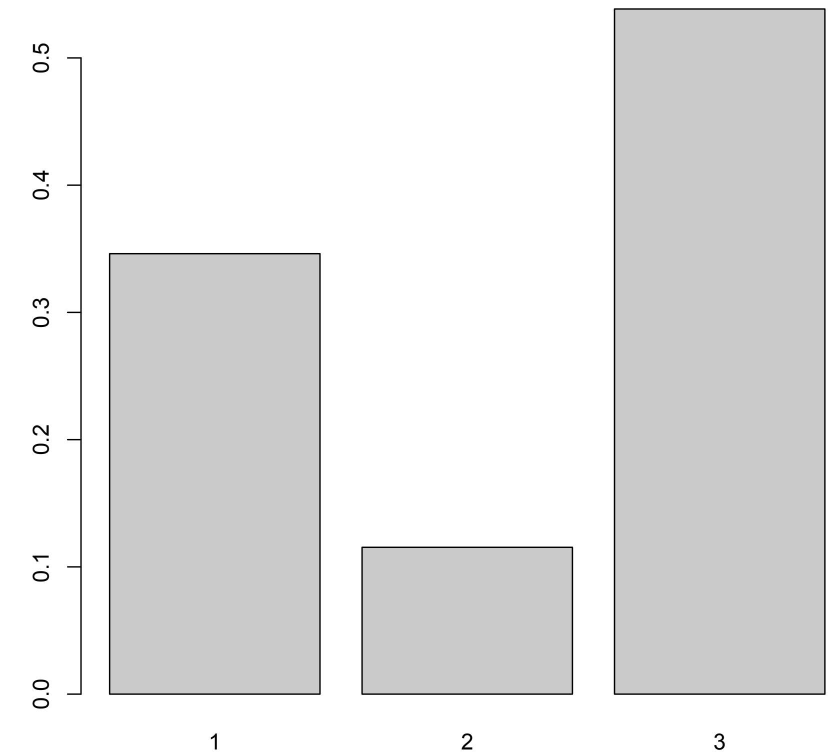
Four histograms of US adults heights with varying bin widths.



Random variables

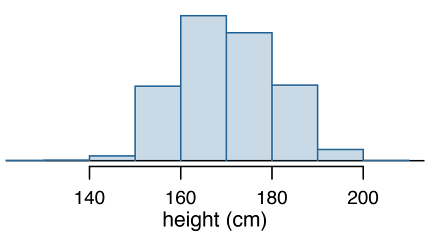
discrete Takes values 0, 1, 2, ...
Completely described by
P(X=1), P(X=2), ...

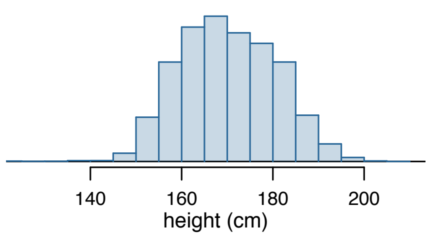
```
x=c(1, 1, 1,1,1,1,1,1,1,2,2,2,3,3,3,3,3,3,3,3,3,3,3,3,3,3)
barplot(prop.tab
```



continuous Takes a continuum of values (e.g. interval [a, b])
Can be seen as the limit of a histogram when the bin size goes to 0.

Four histograms of US adults heights with varying bin widths.

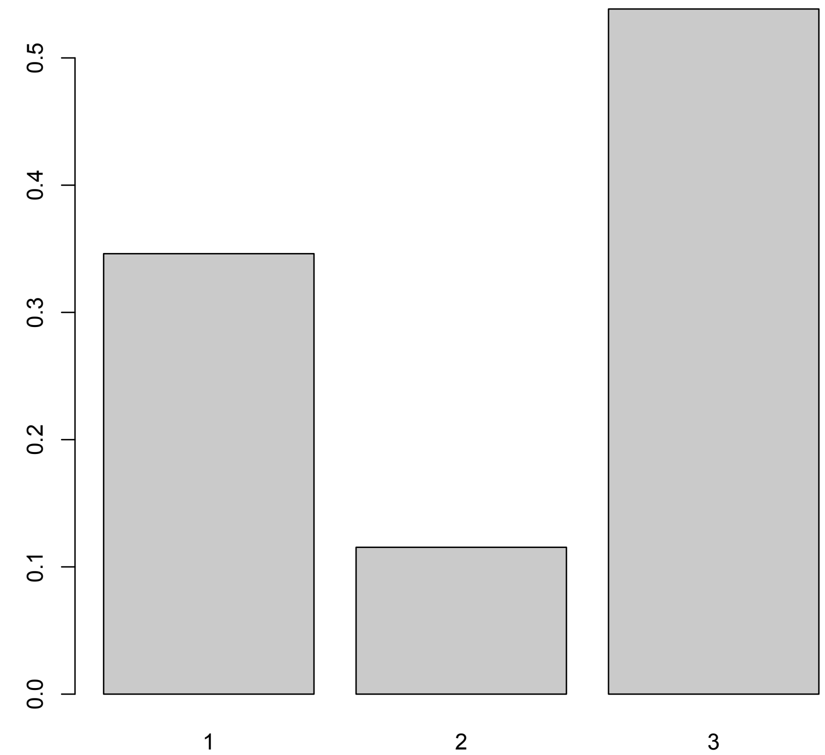




Random variables

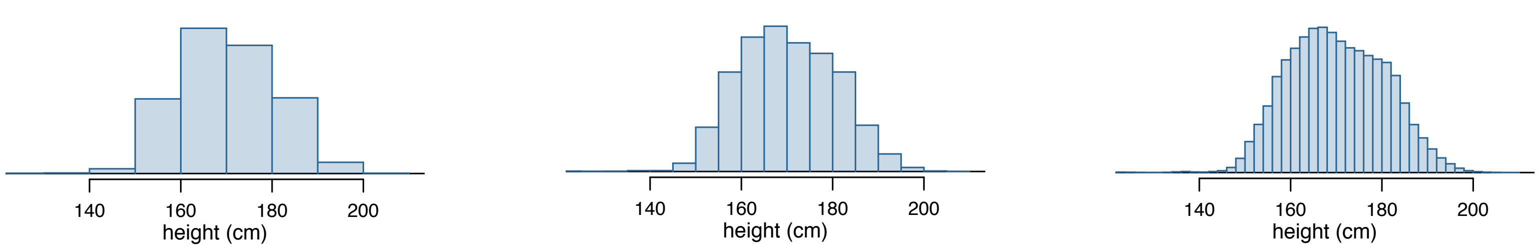
discrete Takes values 0, 1, 2, ...
Completely described by P(X=1), P(X=2), ...

```
x=c(1, 1, 1,1,1,1,1,1,1,2,2,2,3,3,3,3,3,3,3,3,3,3,3,3,3,3)
barplot(prop.tab
```



continuous Takes a continuum of values (e.g. interval [a, b])
Can be seen as the limit of a histogram when the bin size goes to 0.

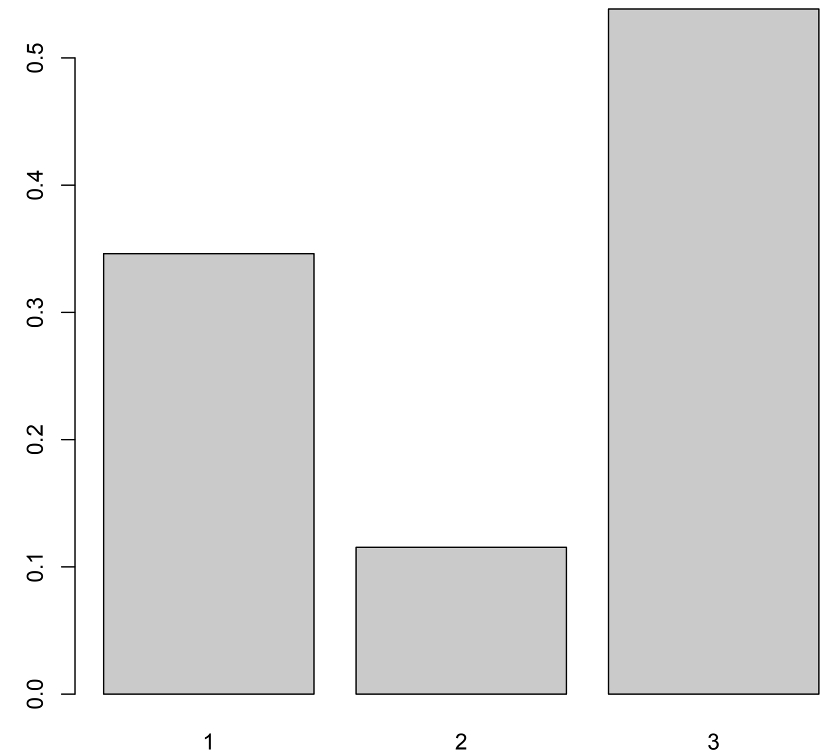
Four histograms of US adults heights with varying bin widths.



Random variables

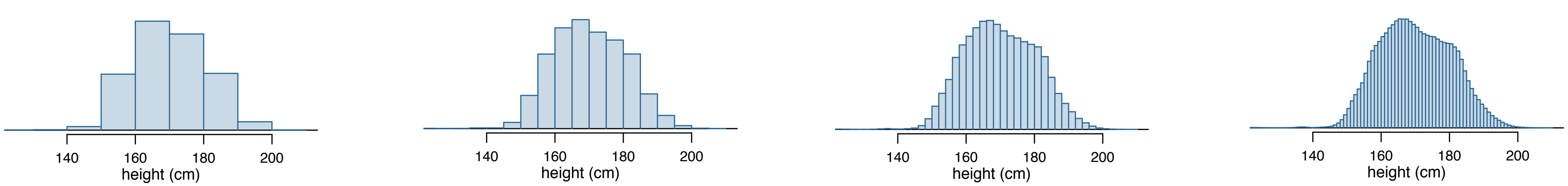
discrete Takes values 0, 1, 2, ...
Completely described by
P(X=1), P(X=2), ...

```
x=c(1, 1, 1,1,1,1,1,1,1,2,2,2,3,3,3,3,3,3,3,3,3,3,3,3,3,3)
barplot(prop.tab
```



continuous Takes a continuum of values (e.g. interval [a, b])
Can be seen as the limit of a histogram when the bin size goes to 0.

Four histograms of US adults heights with varying bin widths.



Discrete random variables

To **fully** describe discrete random variables, we need to give P(X=x), for each value x that the random variable X can take.

Examples: Possible values for X

1. Rolling one die: {1,2,...,6}
2. Sum of two dice: {1,2,...,12}
3. Number of cars in household: {0,1, 2, ...}
4. Indicator of sickness: {0, 1}
5. Number of courses this semester: {2, ...,7}

Probability distribution

The probability distribution of a **discrete** random variable X is a table (or a formula) that gives P(X=x) for each possible value x of X.

x	2	3	4	5	6	7
P(X=x)	0.04	0.13	0.25	0.39	0.17	0.02

Number of courses

Probability distribution

It is equivalent to add values x for X that are not possible and assign them probability 0

x	0	1	2	3	4	5	6	7
P(X=x)	0.00	0.00	0.04	0.13	0.25	0.39	0.17	0.02

Note that we have

$$\sum_{x=0}^{7} P(X = x) = 1$$

Probability distribution

This is true in general. From the addition rule for disjoint events, we have

$$\sum_{\text{all possible } x} P(X = x) = \text{P(X takes one of its possible values)=1}$$

The probability distribution is a function $p(x)$ defined by

$$p(x) = P(X = x)$$

and is sometimes called "probability mass function".

Probability distribution

This is true in general. From the addition rule for disjoint events, we have

$$\sum_{\text{all possible } x} P(X = x) = \text{P(X takes one of its possible values)=1}$$

The probability distribution is a function $p(x)$ defined by

$$p(x) = P(X = x)$$

and is sometimes called "probability mass function".

This function can also be described by a formula instead of enumerating all possible values x.

Probability distribution

This is true in general. From the addition rule for disjoint events, we have

$$\sum_{\text{all possible } x} P(X = x) = \text{P(X takes one of its possible values)=1}$$

The probability distribution is a function $p(x)$ defined by

$$p(x) = P(X = x)$$

and is sometimes called "probability mass function".

This function can also be described by a formula instead of enumerating all possible values x.

Example: Flip a coin 1000 times. X=number of heads. Possible values {0,...,1000}.

Probability distribution

This is true in general. From the addition rule for disjoint events, we have

$$\sum_{\text{all possible } x} P(X = x) = \text{P(X takes one of its possible values)=1}$$

The probability distribution is a function $p(x)$ defined by

$$p(x) = P(X = x)$$

and is sometimes called "probability mass function".

This function can also be described by a formula instead of enumerating all possible values x.

Example: Flip a coin 1000 times. X=number of heads. Possible values {0,...,1000}.

$$p(x) = \frac{1}{2^{1000}}\binom{1000}{x}$$

Expected Value

For Probability, we had:

As the number of observations goes to infinity, the proportion of occurrences of a given outcome converges to the probability of this outcome.

We now have

As the number of observations goes to infinity, the average of the observed values converges to the **expected value** of the random variable.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \rightarrow E(X)$$

Expected Value

But we can rearrange the sum to get

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{x \text{ possible}} x \cdot (\text{number of } x_i \text{ equal to } x)$$

$$= \sum_{x \text{ possible}} x \cdot (\text{proportion of } x_i \text{ equal to } x)$$

Expected Value

But we can rearrange the sum to get

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{x \text{ possible}} x \cdot (\text{number of } x_i \text{ equal to } x)$$

$$= \sum_{x \text{ possible}} x \cdot (\text{proportion of } x_i \text{ equal to } x)$$

$$\rightarrow$$

Expected Value

But we can rearrange the sum to get

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{x \text{ possible}} x \cdot (\text{number of } x_i \text{ equal to } x)$$

$$= \sum_{x \text{ possible}} x \cdot (\text{proportion of } x_i \text{ equal to } x)$$

$$\to \sum_{x \text{ possible}} xP(X = x)$$

Expected Value

But we can rearrange the sum to get

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{x \text{ possible}} x \cdot (\text{number of } x_i \text{ equal to } x)$$

$$= \sum_{x \text{ possible}} x \cdot (\text{proportion of } x_i \text{ equal to } x)$$

$$\to \sum_{x \text{ possible}} xP(X = x)$$

from the definition of the probability

Expected Value

But $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \rightarrow E(X) = \sum_{x \text{ possible}} xP(X = x)$

Expected Value

The **expected value** of the discrete random variable X is defined by

$$E(X) = \sum_{x \text{ possible}} x P(X = x)$$

Expected Value

The **expected value** of the discrete random variable X is defined by

$$E(X) = \sum_{x \text{ possible}} xP(X = x)$$

In the same way, we can define the expected value of the function h(X) of a discrete random variable X by

Expected Value

The **expected value** of the discrete random variable X is defined by

$$E(X) = \sum_{x \text{ possible}} x P(X = x)$$

In the same way, we can define the expected value of the function h(X) of a discrete random variable X by

$$E(\,h(X)) = \sum_{x \text{ possible}} h(x) \quad P(X = x)$$

Variance

An interesting choice of the function h(X) is

$$h(X) = (X - \mu)^2$$

where $\mu = E(X)$

With this choice, we obtain the variance of X

$$var(X) = \sum_{x \text{ possible}} (x - \mu)^2 P(X = x)$$

It is the limit of s^2 as the number of observations goes to infinity.

The following shortcut formula is useful: $var(X) = E(X^2) - \mu^2$

Standard Deviation $\sqrt{var(X)}$

Example

Going back to the number of courses example:

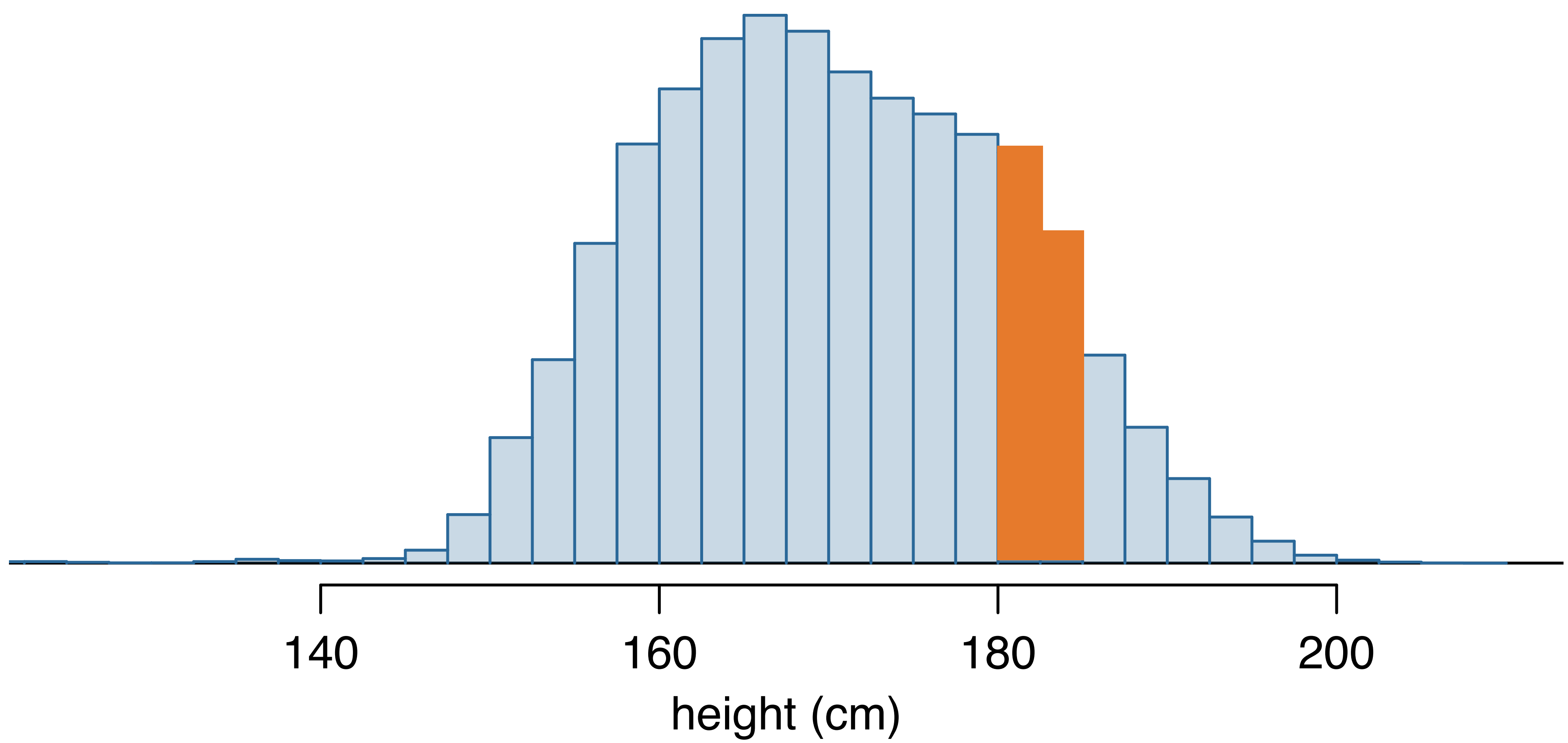
x	2	3	4	5	6	7
P(X=x)	0.04	0.13	0.25	0.39	0.17	0.02

$$E(X) = 2 \cdot 0.04 + 3 \cdot 0.13 + 4 \cdot 0.25 + 5 \cdot 0.39 + 6 \cdot 0.17 + 7 \cdot 0.02 = 4.58$$

$$var(X) = 2^2 \cdot 0.04 + 3^2 \cdot 0.13 + 4^2 \cdot 0.25 + 5^2 \cdot 0.39 + 6^2 \cdot 0.17 + 7^2 \cdot 0.02 - (4.58)^2 = 1.20$$

Continuous random variables

Consider the histogram of heights of 3 million US adults:



What is the probability that a randomly selected US adult has height between 180 and 185 cm?

Continuous random variables

It is given by the area in orange
(sum of 2 rectangles)

Continuous random variables

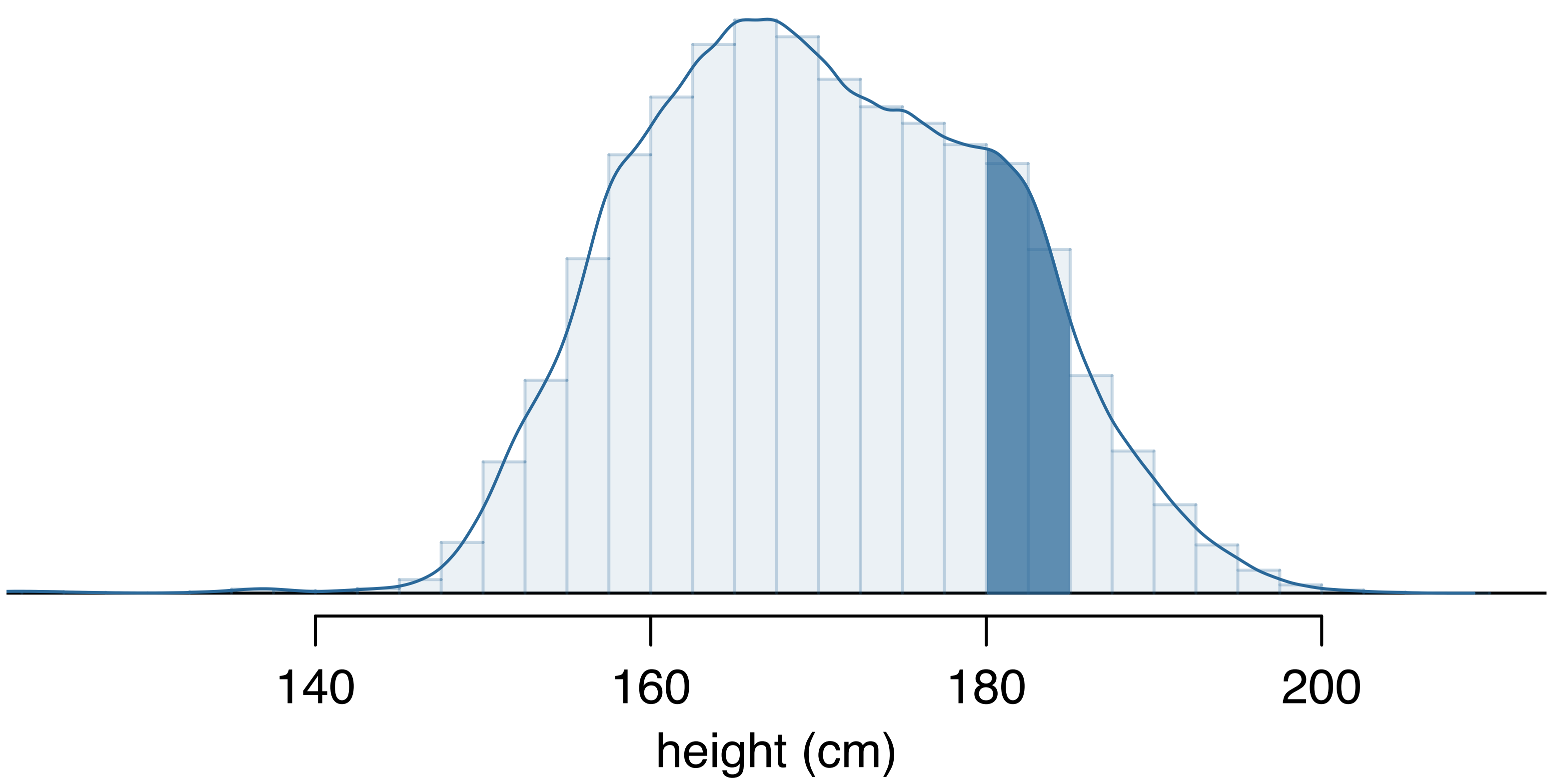
It is given by the area in orange (sum of 2 rectangles)

What if we let the bin width go to 0?

Continuous random variables

It is given by the area in orange (sum of 2 rectangles)

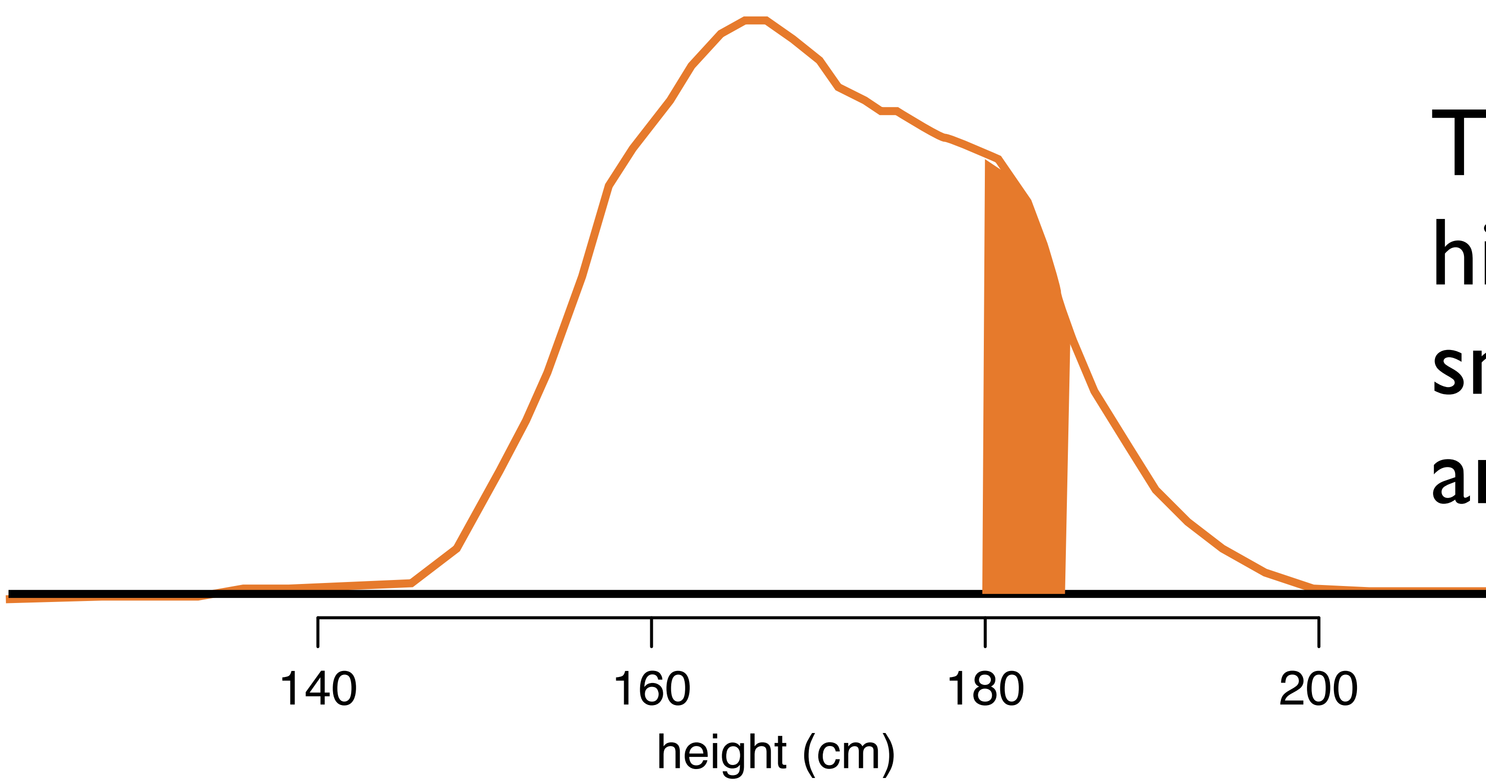
What if we let the bin width go to 0?



Continuous random variables

It is given by the area in orange (sum of 2 rectangles)

What if we let the bin width go to 0?



The contour of the histogram converges to a smooth curve $f(x)$. The area becomes an integral

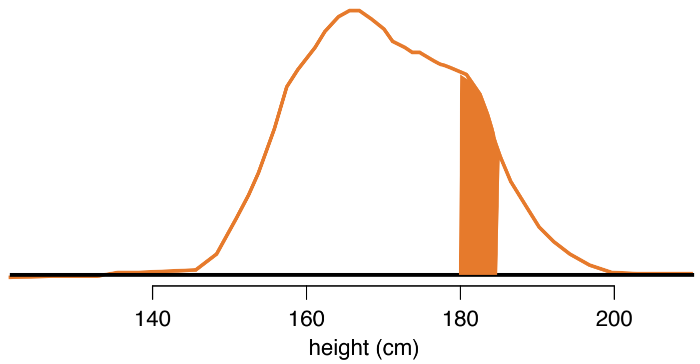
$$\int_{180}^{185} f(x) \ dx$$

Probability density function

The function $f(x)$ is called **probability density function (pdf)** of the continuous random variable X. In our example, X is the (random) height of a randomly selected US adult.

We use it to compute the probability that X is in a given interval [a,b]:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$



In particular, the total area under the curve of $f(x)$ is 1.

$$\text{total area} = \int_{-\infty}^{\infty} f(x)dx = P(-\infty \leq X \leq \infty) = 1$$

Expected Value

The **expected value** of a continuous random variable is also the limit of $\bar{x}$ as the number of observations goes to infinity
In this case, we have

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

For a general function $h(X)$ we have

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) f(x) dx$$

Variance

The **variance** of a continuous random variable is obtained by taking $h(X) = (X - \mu)^2$

$$var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

We still have the shortcut formula

$$var(X) = E(X^2) - \mu^2 = \int x^2 f(x) dx - \left(\int x f(x) dx \right)^2$$

Moreover, the variance is the limit of s^2 as the number of observations goes to infinity.

Rules

If the function h(X) is of the form h(X)=aX+b for some numbers a and b, we have other shortcuts:

$$E(aX + b) = aE(X) + b$$

$$var(aX + b) = a^2 var(X)$$

These rules apply whether the random variable is discrete or continuous.