Quantitative Methods

Human Sciences, 2020–21

Elias Nosrati

Lecture 3: 29 October 2020

Today

Today

- Recap on conditional probability: prosecutor's fallacy and the Monty Hall problem.

Today

- Recap on conditional probability: prosecutor's fallacy and the Monty Hall problem.
- Random variables and their distributions.

Today

- Recap on conditional probability: prosecutor's fallacy and the Monty Hall problem.
- Random variables and their distributions.
- Problem sheet 2 (tutorial).

Prosecutor's fallacy

Prosecutor's fallacy

- In 1998, Sally Clark was tried for murder after two of her sons died shortly after birth.

Prosecutor's fallacy

- In 1998, Sally Clark was tried for murder after two of her sons died shortly after birth.
- An expert witness for the prosecution testified that the probability of a newborn dying of sudden infant death syndrome (SIDS) was $1/8500$. Hence the probability of two consecutive deaths from SIDS was $(1/8500)^2$: ca. 1 in 73 million.

Prosecutor's fallacy

- In 1998, Sally Clark was tried for murder after two of her sons died shortly after birth.
- An expert witness for the prosecution testified that the probability of a newborn dying of sudden infant death syndrome (SIDS) was $1/8500$. Hence the probability of two consecutive deaths from SIDS was $(1/8500)^2$: ca. 1 in 73 million.
- The concluding argument: the probability of Clark's innocence was 1 in 73 million.

Prosecutor's fallacy

- In 1998, Sally Clark was tried for murder after two of her sons died shortly after birth.
- An expert witness for the prosecution testified that the probability of a newborn dying of sudden infant death syndrome (SIDS) was $1/8500$. Hence the probability of two consecutive deaths from SIDS was $(1/8500)^2$: ca. 1 in 73 million.
- The concluding argument: the probability of Clark's innocence was 1 in 73 million.
- What (if anything) is wrong with this line of reasoning?

Prosecutor's fallacy (cont.)

- Two problems:

Prosecutor's fallacy (cont.)

- Two problems:
  1. Independence assumption: how do we know that the two deaths are independent?

Prosecutor's fallacy (cont.)

- Two problems:
  1. Independence assumption: how do we know that the two deaths are independent?
  2. $\mathbb{P}(\text{innocence} \mid \text{evidence}) \neq \mathbb{P}(\text{evidence} \mid \text{innocence})$.

Prosecutor's fallacy (cont.)

- Two problems:
  1. Independence assumption: how do we know that the two deaths are independent?
  2. $\mathbb{P}(\text{innocence} \mid \text{evidence}) \neq \mathbb{P}(\text{evidence} \mid \text{innocence})$.
- By Bayes' Rule:

$$\mathbb{P}(\text{innocence} \mid \text{evidence}) = \frac{\mathbb{P}(\text{evidence} \mid \text{innocence})\mathbb{P}(\text{innocence})}{\mathbb{P}(\text{evidence})}.$$

Prosecutor's fallacy (cont.)

- Two problems:
  1. Independence assumption: how do we know that the two deaths are independent?
  2. $\mathbb{P}(\text{innocence} \mid \text{evidence}) \neq \mathbb{P}(\text{evidence} \mid \text{innocence})$.
- By Bayes' Rule:

$$\mathbb{P}(\text{innocence} \mid \text{evidence}) = \frac{\mathbb{P}(\text{evidence} \mid \text{innocence})\mathbb{P}(\text{innocence})}{\mathbb{P}(\text{evidence})}.$$

- Note that the prior probability of innocence is extremely high: double deaths due to SIDS might be rare but so are double infanticides!

Prosecutor's fallacy (cont.)

- Two problems:
  1. Independence assumption: how do we know that the two deaths are independent?
  2. $\mathbb{P}(\text{innocence} \mid \text{evidence}) \neq \mathbb{P}(\text{evidence} \mid \text{innocence})$.
- By Bayes' Rule:

$$\mathbb{P}(\text{innocence} \mid \text{evidence}) = \frac{\mathbb{P}(\text{evidence} \mid \text{innocence})\mathbb{P}(\text{innocence})}{\mathbb{P}(\text{evidence})}.$$

- Note that the prior probability of innocence is extremely high: double deaths due to SIDS might be rare but so are double infanticides!
- Re-write $\mathbb{P}(\text{evidence})$ using Law of Total Probability:

$$\mathbb{P}(\text{evidence} \mid \text{innocence})\mathbb{P}(\text{innocence}) + \mathbb{P}(\text{evidence} \mid \text{guilt})\mathbb{P}(\text{guilt}).$$

Prosecutor's fallacy (cont.)

- Two problems:
  1. Independence assumption: how do we know that the two deaths are independent?
  2. $\mathbb{P}(\text{innocence} \mid \text{evidence}) \neq \mathbb{P}(\text{evidence} \mid \text{innocence})$.
- By Bayes' Rule:

$$\mathbb{P}(\text{innocence} \mid \text{evidence}) = \frac{\mathbb{P}(\text{evidence} \mid \text{innocence})\mathbb{P}(\text{innocence})}{\mathbb{P}(\text{evidence})}.$$

- Note that the prior probability of innocence is extremely high: double deaths due to SIDS might be rare but so are double infanticides!
- Re-write $\mathbb{P}(\text{evidence})$ using Law of Total Probability:

$$\mathbb{P}(\text{evidence} \mid \text{innocence})\mathbb{P}(\text{innocence}) + \mathbb{P}(\text{evidence} \mid \text{guilt})\mathbb{P}(\text{guilt}).$$

- Note that $\mathbb{P}(\text{guilt}) \approx 0$

Prosecutor's fallacy (cont.)

- Two problems:
  1. Independence assumption: how do we know that the two deaths are independent?
  2. $\mathbb{P}(\text{innocence} \mid \text{evidence}) \neq \mathbb{P}(\text{evidence} \mid \text{innocence})$.
- By Bayes' Rule:

$$\mathbb{P}(\text{innocence} \mid \text{evidence}) = \frac{\mathbb{P}(\text{evidence} \mid \text{innocence})\mathbb{P}(\text{innocence})}{\mathbb{P}(\text{evidence})}.$$

- Note that the prior probability of innocence is extremely high: double deaths due to SIDS might be rare but so are double infanticides!
- Re-write $\mathbb{P}(\text{evidence})$ using Law of Total Probability:

$$\mathbb{P}(\text{evidence} \mid \text{innocence})\mathbb{P}(\text{innocence}) + \mathbb{P}(\text{evidence} \mid \text{guilt})\mathbb{P}(\text{guilt}).$$

- Note that $\mathbb{P}(\text{guilt}) \approx 0$ and hence

$$\mathbb{P}(\text{innocence} \mid \text{evidence}) \approx \frac{\mathbb{P}(\text{evidence} \mid \text{innocence})\mathbb{P}(\text{innocence})}{\mathbb{P}(\text{evidence} \mid \text{innocence})\mathbb{P}(\text{innocence})} = 1.$$

The Monty Hall problem

The Monty Hall problem

- Choose one of three doors: two with goats, one with car.

The Monty Hall problem

- Choose one of three doors: two with goats, one with car.
- After your choice, Monty Hall opens one door with a goat and asks if you want to switch to the remaining door. Should you switch?

The Monty Hall problem

- Choose one of three doors: two with goats, one with car.
- After your choice, Monty Hall opens one door with a goat and asks if you want to switch to the remaining door. Should you switch?
- Solve analytically or by simulation.

The Monty Hall problem: solving by simulation

```
sims <- 10 ^ 5 # Simulations
doors <- 1:3 # Label doors
win_no_switch <- 0 # Win count without switch
win_switch <- 0 # Win count with switch
```

The Monty Hall problem: solving by simulation

```r
sims <- 10 ^ 5 # Simulations
doors <- 1:3 # Label doors
win_no_switch <- 0 # Win count without switch
win_switch <- 0 # Win count with switch
```

```r
for (i in 1:sims) {
  # Pick a door for the car
  car_door <- sample(doors, 1)
  # Pick a door for the contestant
  chosen_door <- sample(doors, 1)
  # If they match, add to "no switch" count
  if (car_door == chosen_door)
    win_no_switch <- win_no_switch + 1
  # Otherwise, add to switch count
  else
    win_switch <- win_switch + 1
}
```

The Monty Hall problem: solving by simulation (cont.)

```
cat("P(car | no switch) =", win_no_switch / sims)

## P(car | no switch) = 0.33102
```

The Monty Hall problem: solving by simulation (cont.)

```
cat("P(car | no switch) =", win_no_switch / sims)

## P(car | no switch) = 0.33102
```

```
cat("P(car | switch) =", win_switch / sims)

## P(car | switch) = 0.66898
```

How is this possible?

The Monty Hall problem: analytic solution

The Monty Hall problem: analytic solution

- Conditioning as wishful thinking: condition on what you wish you knew.

The Monty Hall problem: analytic solution

- Conditioning as wishful thinking: condition on what you wish you knew.
- Let C_i be the event that the car is behind door i, for $1 \leq i \leq 3$, and let W be the event that you win the car.

The Monty Hall problem: analytic solution

- Conditioning as wishful thinking: condition on what you wish you knew.
- Let C_i be the event that the car is behind door i, for $1 \leq i \leq 3$, and let W be the event that you win the car.
- Since the C_i partition the sample space, we can use the Law of Total Probability:

$$\mathbb{P}(W) = \mathbb{P}(W \mid C_1) \times \frac{1}{3} + \mathbb{P}(W \mid C_2) \times \frac{1}{3} + \mathbb{P}(W \mid C_3) \times \frac{1}{3}.$$

The Monty Hall problem: analytic solution

- Conditioning as wishful thinking: condition on what you wish you knew.
- Let C_i be the event that the car is behind door i, for $1 \leq i \leq 3$, and let W be the event that you win the car.
- Since the C_i partition the sample space, we can use the Law of Total Probability:

$$\mathbb{P}(W) = \mathbb{P}(W \mid C_1) \times \frac{1}{3} + \mathbb{P}(W \mid C_2) \times \frac{1}{3} + \mathbb{P}(W \mid C_3) \times \frac{1}{3}.$$

- Suppose, without loss of generality, that you have chosen door 1 but decide to switch.

The Monty Hall problem: analytic solution

- Conditioning as wishful thinking: condition on what you wish you knew.
- Let C_i be the event that the car is behind door i, for $1 \leq i \leq 3$, and let W be the event that you win the car.
- Since the C_i partition the sample space, we can use the Law of Total Probability:

$$\mathbb{P}(W) = \mathbb{P}(W \mid C_1) \times \frac{1}{3} + \mathbb{P}(W \mid C_2) \times \frac{1}{3} + \mathbb{P}(W \mid C_3) \times \frac{1}{3}.$$

- Suppose, without loss of generality, that you have chosen door 1 but decide to switch.
- If C_1, the switching strategy will fail:

$$\mathbb{P}(W \mid C_1) = 0.$$

The Monty Hall problem: analytic solution

- Conditioning as wishful thinking: condition on what you wish you knew.
- Let C_i be the event that the car is behind door i, for $1 \leq i \leq 3$, and let W be the event that you win the car.
- Since the C_i partition the sample space, we can use the Law of Total Probability:

$$\mathbb{P}(W) = \mathbb{P}(W \mid C_1) \times \frac{1}{3} + \mathbb{P}(W \mid C_2) \times \frac{1}{3} + \mathbb{P}(W \mid C_3) \times \frac{1}{3}.$$

- Suppose, without loss of generality, that you have chosen door 1 but decide to switch.
- If C_1, the switching strategy will fail:

$$\mathbb{P}(W \mid C_1) = 0.$$

- If the car is behind doors 2 or 3, switching will succeed:

$$\mathbb{P}(W \mid C_2) = \mathbb{P}(W \mid C_3) = 1.$$

The Monty Hall problem: analytic solution

- Conditioning as wishful thinking: condition on what you wish you knew.
- Let C_i be the event that the car is behind door i, for $1 \leq i \leq 3$, and let W be the event that you win the car.
- Since the C_i partition the sample space, we can use the Law of Total Probability:

$$\mathbb{P}(W) = \mathbb{P}(W \mid C_1) \times \frac{1}{3} + \mathbb{P}(W \mid C_2) \times \frac{1}{3} + \mathbb{P}(W \mid C_3) \times \frac{1}{3}.$$

- Suppose, without loss of generality, that you have chosen door 1 but decide to switch.
- If C_1, the switching strategy will fail:

$$\mathbb{P}(W \mid C_1) = 0.$$

- If the car is behind doors 2 or 3, switching will succeed:

$$\mathbb{P}(W \mid C_2) = \mathbb{P}(W \mid C_3) = 1.$$

- Hence, $\mathbb{P}(W) = 0 \times \frac{1}{3} + 1 \times \frac{1}{3} + 1 \times \frac{1}{3} = \frac{2}{3}$.

Random variables

Random variables

Definition

Given an experiment with sample space S, a *random variable* is a function from the sample space S to the set of real numbers $\mathbb{R}$. Thus a random variable X assigns a numerical value $X(s)$ to each possible outcome s of the experiment:

$$X : S \to \mathbb{R}.$$

Random variables

Definition

Given an experiment with sample space S, a *random variable* is a function from the sample space S to the set of real numbers $\mathbb{R}$. Thus a random variable X assigns a numerical value $X(s)$ to each possible outcome s of the experiment:

$$X : S \to \mathbb{R}.$$

- A concept that simplifies notation.

Random variables

Definition

Given an experiment with sample space S, a *random variable* is a function from the sample space S to the set of real numbers $\mathbb{R}$. Thus a random variable X assigns a numerical value $X(s)$ to each possible outcome s of the experiment:

$$X : S \to \mathbb{R}.$$

- A concept that simplifies notation.
- Expands our ability to quantify uncertainty and summarise the results of experiments.

Random variables

Definition

Given an experiment with sample space S, a *random variable* is a function from the sample space S to the set of real numbers $\mathbb{R}$. Thus a random variable X assigns a numerical value $X(s)$ to each possible outcome s of the experiment:

$$X : S \to \mathbb{R}.$$

- A concept that simplifies notation.
- Expands our ability to quantify uncertainty and summarise the results of experiments.
- The function itself is deterministic.

Random variables

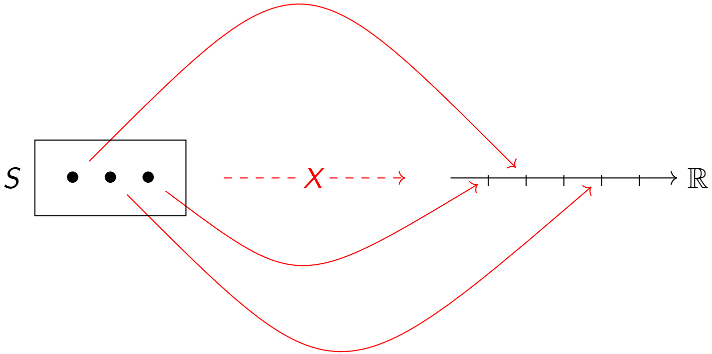
Definition

Given an experiment with sample space S, a *random variable* is a function from the sample space S to the set of real numbers $\mathbb{R}$. Thus a random variable X assigns a numerical value $X(s)$ to each possible outcome s of the experiment:

$$X : S \to \mathbb{R}.$$

- A concept that simplifies notation.
- Expands our ability to quantify uncertainty and summarise the results of experiments.
- The function itself is deterministic.
- The randomness derives from the experiment whose outcomes have probabilities described by the probability function $\mathbb{P}(\cdot)$.

Random variables (cont.)



Random variables: coin flip example

Random variables: coin flip example

- We flip a fair coin twice.

Random variables: coin flip example

- We flip a fair coin twice.
- $S = \{HH, HT, TH, TT\}$.

Random variables: coin flip example

- We flip a fair coin twice.
- $S = \{HH, HT, TH, TT\}$.
- Let X be the number of Heads. This is a random variable with possible values 0, 1, and 2:

$$X(HH) = 2, \quad X(HT) = X(TH) = 1, \quad X(TT) = 0.$$

Random variables: coin flip example

- We flip a fair coin twice.
- $S = \{HH, HT, TH, TT\}$.
- Let X be the number of Heads. This is a random variable with possible values 0, 1, and 2:

$$X(HH) = 2, \quad X(HT) = X(TH) = 1, \quad X(TT) = 0.$$

- Let Y be the number of tails. Then $Y = 2 - X$.

Random variables: coin flip example

- We flip a fair coin twice.
- $S = \{HH, HT, TH, TT\}$.
- Let X be the number of Heads. This is a random variable with possible values 0, 1, and 2:

$$X(HH) = 2, \quad X(HT) = X(TH) = 1, \quad X(TT) = 0.$$

- Let Y be the number of tails. Then $Y = 2 - X$.
- Let I be the *indicator random variable* that is 1 if the first flip is Heads and 0 otherwise.

Random variables: coin flip example

- We flip a fair coin twice.
- $S = \{HH, HT, TH, TT\}$.
- Let X be the number of Heads. This is a random variable with possible values 0, 1, and 2:

$$X(HH) = 2, \quad X(HT) = X(TH) = 1, \quad X(TT) = 0.$$

- Let Y be the number of tails. Then $Y = 2 - X$.
- Let I be the *indicator random variable* that is 1 if the first flip is Heads and 0 otherwise.
- Note that two random variables can be defined on the same sample space: the outcomes are the same, but the numerical values assigned to the outcomes are different.

Discrete random variables

Discrete random variables

Definition

A random variable X is said to be *discrete* if it takes on a finite (or at most countably infinite) number of values. In other words, its outcomes are discrete.

Discrete random variables

Definition

A random variable X is said to be *discrete* if it takes on a finite (or at most countably infinite) number of values. In other words, its outcomes are discrete.

Example

The number of Heads resulting from a single coin flip is a *binary* random variable: it takes on only two values (0 or 1). The sum of the number of eyes on two dice or the number of votes won by a presidential candidate are also discrete random variables.

Continuous random variables

Continuous random variables

Definition

A random variable X is said to be *continuous* if it takes on an (uncountably) infinite number of values. In other words, it may take on a continuum of possible values.

Continuous random variables

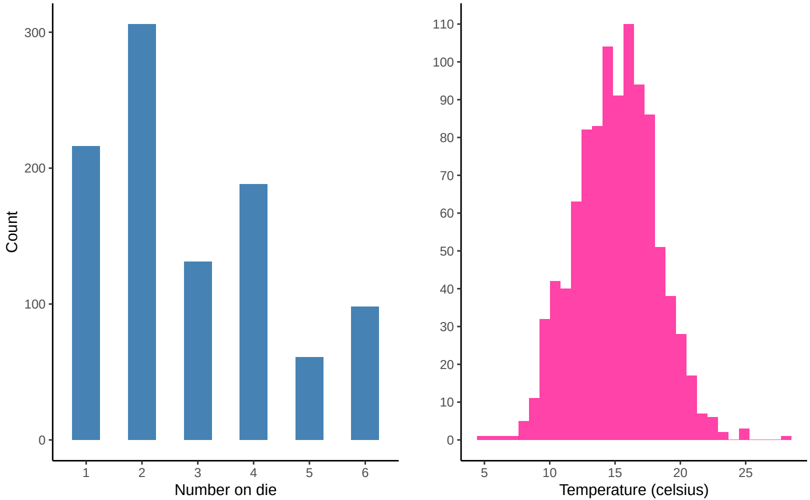
Definition

A random variable X is said to be *continuous* if it takes on an (uncountably) infinite number of values. In other words, it may take on a continuum of possible values.

Example

The total amount of rainfall over the next year is a continuous random variable. So is the temperature on a randomly chosen day, or the height of a randomly selected person.

The distribution of random variables



Probability distributions

Probability distributions

Definition

The *probability distribution* of a random variable X is the assignment of a probability measure to every conceivable realisation of X.

Probability distributions

Definition

The *probability distribution* of a random variable X is the assignment of a probability measure to every conceivable realisation of X.

- The distribution of a random variable X specifies the probabilities of all events associated with X.

Probability distributions

Definition

The *probability distribution* of a random variable X is the assignment of a probability measure to every conceivable realisation of X.

- The distribution of a random variable X specifies the probabilities of all events associated with X.
- Describes which values of X are more likely than other values.

Probability distributions

Definition

The *probability distribution* of a random variable X is the assignment of a probability measure to every conceivable realisation of X.

- The distribution of a random variable X specifies the probabilities of all events associated with X.
- Describes which values of X are more likely than other values.
- Commonly encoded by a *probability density function*.

Probability distributions

Definition

The *probability distribution* of a random variable X is the assignment of a probability measure to every conceivable realisation of X.

- The distribution of a random variable X specifies the probabilities of all events associated with X.
- Describes which values of X are more likely than other values.
- Commonly encoded by a *probability density function*.
- Notation: $f(x) = f(X = x)$.

Probability distributions

Definition

The *probability distribution* of a random variable X is the assignment of a probability measure to every conceivable realisation of X.

- The distribution of a random variable X specifies the probabilities of all events associated with X.
- Describes which values of X are more likely than other values.
- Commonly encoded by a *probability density function*.
- Notation: $f(x) = f(X = x)$.
- Two requirements:

Probability distributions

Definition

The *probability distribution* of a random variable X is the assignment of a probability measure to every conceivable realisation of X.

- The distribution of a random variable X specifies the probabilities of all events associated with X.
- Describes which values of X are more likely than other values.
- Commonly encoded by a *probability density function*.
- Notation: $f(x) = f(X = x)$.
- Two requirements:
    1. $f(x) \geq 0$ for any x.

Probability distributions

Definition

The *probability distribution* of a random variable X is the assignment of a probability measure to every conceivable realisation of X.

- The distribution of a random variable X specifies the probabilities of all events associated with X.
- Describes which values of X are more likely than other values.
- Commonly encoded by a *probability density function*.
- Notation: $f(x) = f(X = x)$.
- Two requirements:
  1. $f(x) \geq 0$ for any x.
  2. $\sum f(x) = 1$ (discrete) or $\int f(x) = 1$ (continuous).

What exactly is a density?

What exactly is a density?

- Let X be a random variable (e.g., age of random person).

What exactly is a density?

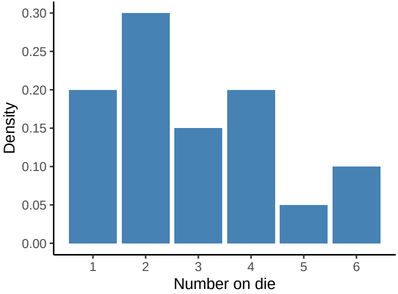
- Let X be a random variable (e.g., age of random person).
- To visualise the distribution of X, we 'discretise' the variable by creating bins or intervals along the variable of interest and count the number of observations that fall within each bin.

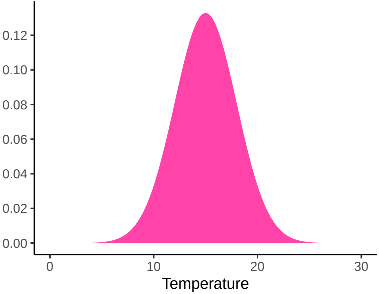
What exactly is a density?

- Let X be a random variable (e.g., age of random person).
- To visualise the distribution of X, we 'discretise' the variable by creating bins or intervals along the variable of interest and count the number of observations that fall within each bin.
- The *density* for each bin is then defined as

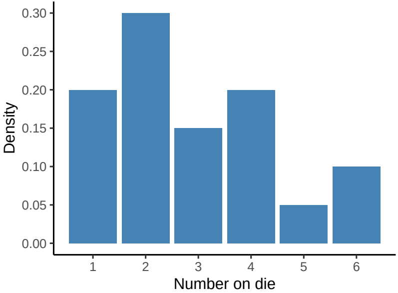
$$\text{density} = \frac{\text{proportion of observations in the bin}}{\text{width of the bin}}.$$

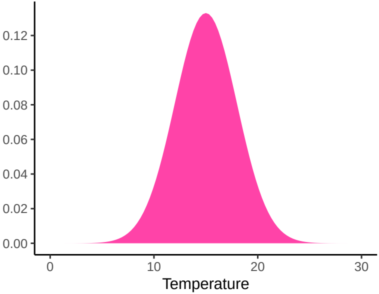
What exactly is a density? (cont.)





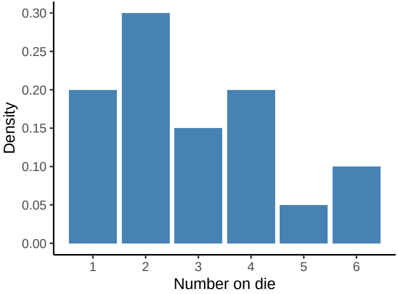
What exactly is a density? (cont.)

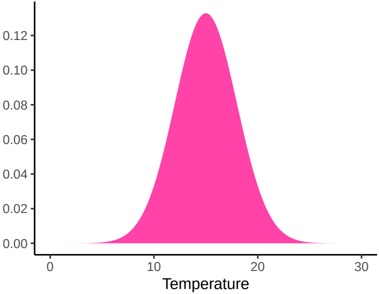




- We interpret the density scale as a probability (or percentage) per horizontal unit.

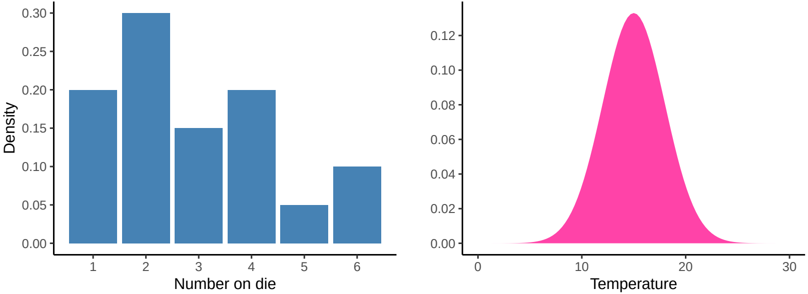
What exactly is a density? (cont.)





- We interpret the density scale as a probability (or percentage) per horizontal unit.
- For a discrete random variable, probability = density.

What exactly is a density? (cont.)



- We interpret the density scale as a probability (or percentage) per horizontal unit.
- For a discrete random variable, probability = density.
- For a continuous random variable, probability = area under curve.