Quantitative Methods

Human Sciences, 2020–2021

Elias Nosrati

Lecture 2: 22 October 2020

Today

Today

- Recap on probability and counting: the birthday problem.

Today

- Recap on probability and counting: the birthday problem.
- Introduction to conditional probability.

Today

- Recap on probability and counting: the birthday problem.
- Introduction to conditional probability.
- Problem sheet 1 (tutorial).

The birthday problem

The birthday problem

- There are k people in a room.

The birthday problem

- There are k people in a room.
- Assume each person's birthday is equally likely to be any of the 365 days of the year and assume people's birthdays are independent.

The birthday problem

- There are k people in a room.
- Assume each person's birthday is equally likely to be any of the 365 days of the year and assume people's birthdays are independent.
- What is the probability that at least one pair of people in the group have the same birthday?

The birthday problem

- There are k people in a room.
- Assume each person's birthday is equally likely to be any of the 365 days of the year and assume people's birthdays are independent.
- What is the probability that at least one pair of people in the group have the same birthday?
- Hint: Recall that $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$.

The birthday problem: solution

The birthday problem: solution

- Total number of ways of assigning birthdays to n people (sampling from 365 days *with* replacement): 365^n.

The birthday problem: solution

- Total number of ways of assigning birthdays to n people (sampling from 365 days *with* replacement): 365^n.
- Complement: the number of ways to assign birthdays to n people such that no two people share the same birthday (sampling from 365 days *without* replacement):

$$365 \times 364 \times \cdots \times (365 - n + 1)$$

for $n \leq 365$.

The birthday problem: solution

- Total number of ways of assigning birthdays to n people (sampling from 365 days *with* replacement): 365^n.
- Complement: the number of ways to assign birthdays to n people such that no two people share the same birthday (sampling from 365 days *without* replacement):

  $$365 \times 364 \times \cdots \times (365 - n + 1)$$

  for $n \leq 365$.
- $\mathbb{P}(\text{no birthday match}) = \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}$.

The birthday problem: solution

- Total number of ways of assigning birthdays to n people (sampling from 365 days *with* replacement): 365^n.
- Complement: the number of ways to assign birthdays to n people such that no two people share the same birthday (sampling from 365 days *without* replacement):

$$365 \times 364 \times \cdots \times (365 - n + 1)$$

  for $n \leq 365$.
- $\mathbb{P}(\text{no birthday match}) = \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}$.
- $\mathbb{P}(\text{at least one birthday match}) = 1 - \mathbb{P}(\text{no birthday match})$.

The birthday problem: solution

- Total number of ways of assigning birthdays to n people (sampling from 365 days *with* replacement): 365^n.
- Complement: the number of ways to assign birthdays to n people such that no two people share the same birthday (sampling from 365 days *without* replacement):

$$365 \times 364 \times \cdots \times (365 - n + 1)$$

  for $n \leq 365$.
- $\mathbb{P}(\text{no birthday match}) = \frac{365 \times 364 \times \cdots \times (365-n+1)}{365^n}$.
- $\mathbb{P}(\text{at least one birthday match}) = 1 - \mathbb{P}(\text{no birthday match})$.
- In this room, $\mathbb{P}(\text{at least one birthday match}) \approx 4\%$.

The birthday problem in R

```r
# Create a function
pmatch <- function(n) {
        1 - prod(365:(365 - n + 1)) / (365 ^ n)
}
```

The birthday problem in R

```r
# Create a function
pmatch <- function(n) {
        1 - prod(365:(365 - n + 1)) / (365 ^ n)
}
```

```r
# Placeholder
probs <- NULL

# For loop
for (i in 1:70) {
  probs[i] <- pmatch(i)
}
```

The birthday problem in **R**

```r
# Create a function
pmatch <- function(n) {
        1 - prod(365:(365 - n + 1)) / (365 ^ n)
}
```

```r
# Placeholder
probs <- NULL

# For loop
for (i in 1:70) {
  probs[i] <- pmatch(i)
}
```

```r
# Alternative method
probs <- sapply(1:70, pmatch)
```

The birthday problem in **R** (cont.)

```
save <- data.frame("n" = 1:70, "prob" = probs)
```

The birthday problem in **R** (cont.)

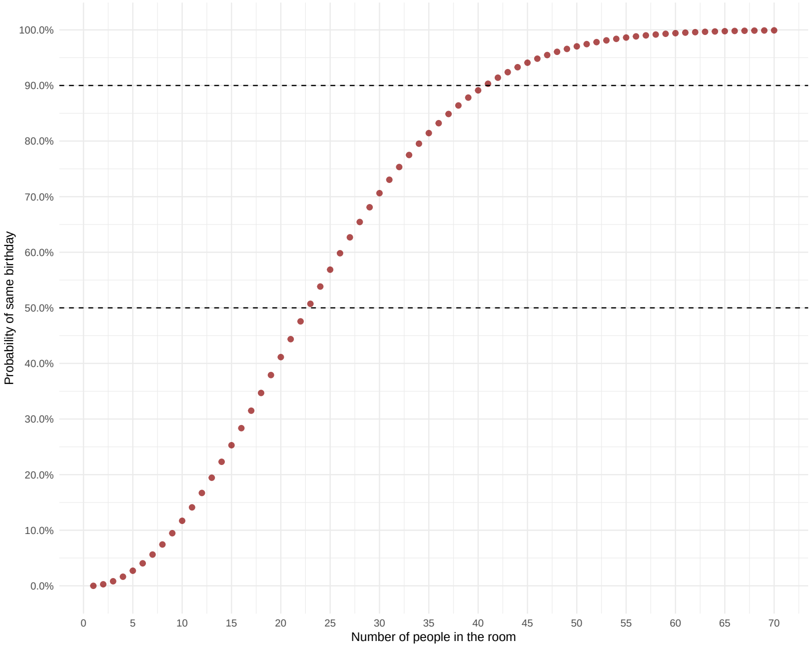
```r
save <- data.frame("n" = 1:70, "prob" = probs)
```

```r
head(save)

##   n        prob
## 1 1 0.000000000
## 2 2 0.002739726
## 3 3 0.008204166
## 4 4 0.016355912
## 5 5 0.027135574
## 6 6 0.040462484
```

The birthday problem in **R** (cont.)

```
ggplot(save, aes(n, prob)) +
  labs( # Axis labels
    x = "Number of people in the room",
    y = "Probability of same birthday") +
  scale_x_continuous( # Modify X-axis
    breaks = seq(0, 70, 5)) +
  scale_y_continuous( # Modify Y-axis
    breaks = seq(0, 1, 0.1),
    label = scales::percent) +
  geom_point( # Modify points
    size = 2,
    colour = "darkred",
    alpha = 0.7) +
  geom_hline( # When is P(match) > 0.5 or 0.9?
    yintercept = c(0.5, 0.9), linetype = "dashed") +
  theme_minimal() # Remove redundant lines
```


Probability of same birthday
100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
Number of people in the room

Conditional probability: motivation

Conditional probability: motivation

- Probability is a means of quantifying uncertainty about events.

Conditional probability: motivation

- Probability is a means of quantifying uncertainty about events.
- Whenever new evidence is observed, we acquire information that may affect out uncertainties.

Conditional probability: motivation

- Probability is a means of quantifying uncertainty about events.
- Whenever new evidence is observed, we acquire information that may affect out uncertainties.
- Conditional probability allows us to update our beliefs in light of new evidence.

Conditional probability: motivation

- Probability is a means of quantifying uncertainty about events.
- Whenever new evidence is observed, we acquire information that may affect out uncertainties.
- Conditional probability allows us to update our beliefs in light of new evidence.
- "Conditioning is the soul of statistics" (Blitzstein and Hwang, 2019: 46).

Defining conditional probability

Definition

If A and B are events, with $\mathbb{P}(B) > 0$, then the *conditional probability* of A given B is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Defining conditional probability

Definition

If A and B are events, with $\mathbb{P}(B) > 0$, then the *conditional probability* of A given B is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

- A is the event whose uncertainty we want to update and B is the (new) evidence we observe.

Defining conditional probability

Definition

If A and B are events, with $\mathbb{P}(B) > 0$, then the *conditional probability* of A given B is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

- A is the event whose uncertainty we want to update and B is the (new) evidence we observe.
- $\mathbb{P}(A)$ is called the *prior* probability of A (*before* updating beliefs based on new evidence).

Defining conditional probability

Definition

If A and B are events, with $\mathbb{P}(B) > 0$, then the *conditional probability* of A given B is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

- A is the event whose uncertainty we want to update and B is the (new) evidence we observe.
- $\mathbb{P}(A)$ is called the *prior* probability of A (*before* updating beliefs based on new evidence).
- $\mathbb{P}(A \mid B)$ is called the *posterior* of A (*after* updating beliefs based on new evidence).

Defining conditional probability

Definition

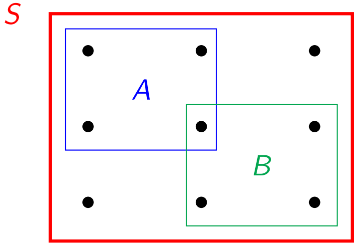
If A and B are events, with $\mathbb{P}(B) > 0$, then the *conditional probability* of A given B is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

- A is the event whose uncertainty we want to update and B is the (new) evidence we observe.
- $\mathbb{P}(A)$ is called the *prior* probability of A (*before* updating beliefs based on new evidence).
- $\mathbb{P}(A \mid B)$ is called the *posterior* of A (*after* updating beliefs based on new evidence).
- Note that $\mathbb{P}(A \mid A) = 1$.

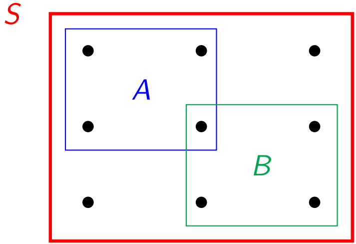
Visualising a conditional probability

Suppose we have a sample space S and two events A and B:



Visualising a conditional probability

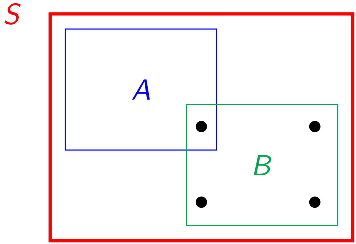
Suppose we have a sample space S and two events A and B:



- Suppose that B occurred: get rid of all elements of B^c.

Visualising a conditional probability

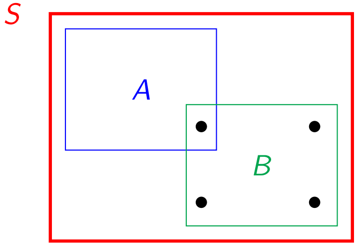
Suppose we have a sample space S and two events A and B:



- Suppose that B occurred: get rid of all elements of B^c.
- The only probability measure remaining that we can assign to A is $\mathbb{P}(A \cap B)$.

Visualising a conditional probability

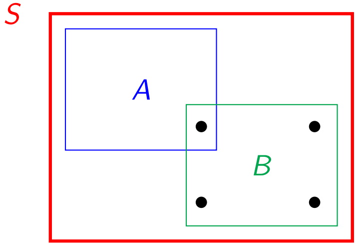
Suppose we have a sample space S and two events A and B:



- Suppose that B occurred: get rid of all elements of B^c.
- The only probability measure remaining that we can assign to A is $\mathbb{P}(A \cap B)$.
- *Renormalise*: create a new probability space that assigns an updated probability measure to each possible event such that all the probabilities add up to 1.

Visualising a conditional probability

Suppose we have a sample space S and two events A and B:

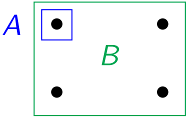


- Suppose that B occurred: get rid of all elements of B^c.
- The only probability measure remaining that we can assign to A is $\mathbb{P}(A \cap B)$.
- *Renormalise*: create a new probability space that assigns an updated probability measure to each possible event such that all the probabilities add up to 1.
- Divide by $\mathbb{P}(B)$, the total mass of the outcomes in B.

Visualising a conditional probability (cont.)

The updated probability measure assigned to the event A is the conditional probability

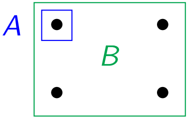
$$\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B).$$



Visualising a conditional probability (cont.)

The updated probability measure assigned to the event A is the conditional probability

$$\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B).$$

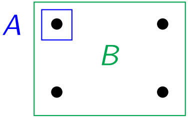


- We have updated our probabilities in accordance with observed evidence.

Visualising a conditional probability (cont.)

The updated probability measure assigned to the event A is the conditional probability

$$\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B).$$

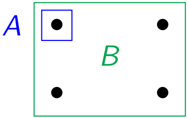


- We have updated our probabilities in accordance with observed evidence.
- Outcomes that contradict the evidence are discarded.

Visualising a conditional probability (cont.)

The updated probability measure assigned to the event A is the conditional probability

$$\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B).$$



- We have updated our probabilities in accordance with observed evidence.
- Outcomes that contradict the evidence are discarded.
- Relative measures of uncertainty are redistributed amongst remaining possible outcomes.

Frequentist interpretation of conditional probability

Frequentist interpretation of conditional probability

- Frequentist interpretation of probability: relative frequency over a large number of repeated trials.

Frequentist interpretation of conditional probability

- Frequentist interpretation of probability: relative frequency over a large number of repeated trials.
- $\mathbb{P}(A \mid B)$ is the fraction of times A occurs, restricting attention to trials where B occurs.

Example

- We flip a coin several times, record the resulting strings of Heads and Tails, and repeat the experiment n times.

Frequentist interpretation of conditional probability

- Frequentist interpretation of probability: relative frequency over a large number of repeated trials.
- $\mathbb{P}(A \mid B)$ is the fraction of times A occurs, restricting attention to trials where B occurs.

Example

- We flip a coin several times, record the resulting strings of Heads and Tails, and repeat the experiment n times.
- B is the event that the first flip is Heads, A is the event that the second flip is Tails.

Frequentist interpretation of conditional probability

- Frequentist interpretation of probability: relative frequency over a large number of repeated trials.
- $\mathbb{P}(A \mid B)$ is the fraction of times A occurs, restricting attention to trials where B occurs.

Example

- We flip a coin several times, record the resulting strings of Heads and Tails, and repeat the experiment n times.
- B is the event that the first flip is Heads, A is the event that the second flip is Tails.
- We condition on B by isolating all the times where B occurs (n_B) and look at the fraction of times A also occurs (n_{AB}).

Frequentist interpretation of conditional probability

- Frequentist interpretation of probability: relative frequency over a large number of repeated trials.
- $\mathbb{P}(A \mid B)$ is the fraction of times A occurs, restricting attention to trials where B occurs.

Example

- We flip a coin several times, record the resulting strings of Heads and Tails, and repeat the experiment n times.
- B is the event that the first flip is Heads, A is the event that the second flip is Tails.
- We condition on B by isolating all the times where B occurs (n_B) and look at the fraction of times A also occurs (n_{AB}).
- Thus

$$\mathbb{P}(A \mid B) = n_{AB}/n_B = (n_{AB}/n)/(n_B/n) = \mathbb{P}(A \cap B)/\mathbb{P}(B).$$

Joint probability and conditional probability

Theorem

For any events A and B with positive probabilities,

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A \mid B) = \mathbb{P}(A)\mathbb{P}(B \mid A).$$

Joint probability and conditional probability

Theorem

For any events A and B with positive probabilities,

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A \mid B) = \mathbb{P}(A)\mathbb{P}(B \mid A).$$

This can be generalised to the intersection of n events:

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_n|A_1 \cap \cdots \cap A_n).$$

Bayes' Rule

Bayes' Rule

Theorem

For any events A and B with positive probabilities,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}.$$

Bayes' Rule

Theorem

For any events A and B with positive probabilities,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}.$$

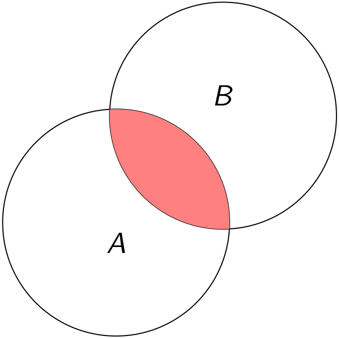
This follows immediately from the previous Theorem (which in turn follows immediately from the definition of conditional probability).

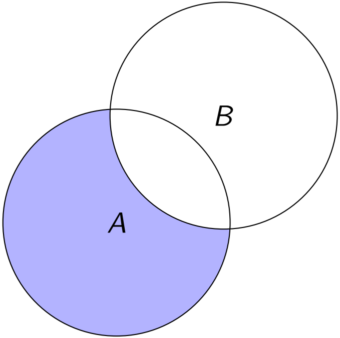
The Law of Total Probability (revisited)

The Law of Total Probability (revisited)

Recall that, for any two events A and B, the Law of Total Probability states that

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c).$$





The Law of Total Probability (revisited)

Viewed differently, the Law of Total Probability relates conditional probability to unconditional probability:

The Law of Total Probability (revisited)

Viewed differently, the Law of Total Probability relates conditional probability to unconditional probability:

Theorem

Let $A_1, \ldots, A_n$ be a partition of the sample space S (quiz: what is a partition again?),

The Law of Total Probability (revisited)

Viewed differently, the Law of Total Probability relates conditional probability to unconditional probability:

Theorem

Let $A_1, \ldots, A_n$ be a partition of the sample space S (quiz: what is a partition again?), with $\mathbb{P}(A_i) > 0$ for all i. Then, for any event B,

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i).$$

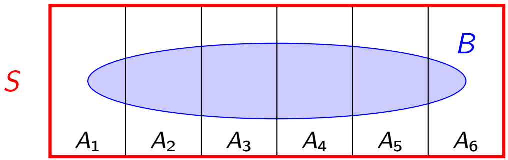
The Law of Total Probability (revisited)

Viewed differently, the Law of Total Probability relates conditional probability to unconditional probability:

Theorem

Let $A_1, \ldots, A_n$ be a partition of the sample space S (quiz: what is a partition again?), with $\mathbb{P}(A_i) > 0$ for all i. Then, for any event B,

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i).$$



Example: testing for a rare disease

Example: testing for a rare disease

- Consider a rare disease D that affects 1% of the population.

Example: testing for a rare disease

- Consider a rare disease D that affects 1% of the population.
- Suppose there is a test T that is 95% accurate such that

$$\mathbb{P}(T \mid D) = 0.95 \quad \text{and} \quad \mathbb{P}(T^c \mid D^c) = 0.95.$$

Example: testing for a rare disease

- Consider a rare disease D that affects 1% of the population.
- Suppose there is a test T that is 95% accurate such that

$$\mathbb{P}(T \mid D) = 0.95 \quad \text{and} \quad \mathbb{P}(T^c \mid D^c) = 0.95.$$

- A randomly selected person takes the test and tests positive. What is the probability that this person has the disease?

Example: testing for a rare disease

- Consider a rare disease D that affects 1% of the population.
- Suppose there is a test T that is 95% accurate such that

$$\mathbb{P}(T \mid D) = 0.95 \quad \text{and} \quad \mathbb{P}(T^c \mid D^c) = 0.95.$$

- A randomly selected person takes the test and tests positive. What is the probability that this person has the disease?

Solution: Apply Bayes' Rule and the Law of Total Probability:

$$\mathbb{P}(D \mid T) = \frac{\mathbb{P}(T \mid D)\mathbb{P}(D)}{\mathbb{P}(T)}$$

Example: testing for a rare disease

- Consider a rare disease D that affects 1% of the population.
- Suppose there is a test T that is 95% accurate such that

$$\mathbb{P}(T \mid D) = 0.95 \quad \text{and} \quad \mathbb{P}(T^c \mid D^c) = 0.95.$$

- A randomly selected person takes the test and tests positive. What is the probability that this person has the disease?

Solution: Apply Bayes' Rule and the Law of Total Probability:

$$\begin{aligned}\mathbb{P}(D \mid T) &= \frac{\mathbb{P}(T \mid D)\mathbb{P}(D)}{\mathbb{P}(T)} \\ &= \frac{\mathbb{P}(T \mid D)\mathbb{P}(D)}{\mathbb{P}(T \mid D)\mathbb{P}(D) + \mathbb{P}(T \mid D^c)\mathbb{P}(D^c)}\end{aligned}$$

Example: testing for a rare disease

- Consider a rare disease D that affects 1% of the population.
- Suppose there is a test T that is 95% accurate such that

$$\mathbb{P}(T \mid D) = 0.95 \quad \text{and} \quad \mathbb{P}(T^c \mid D^c) = 0.95.$$

- A randomly selected person takes the test and tests positive. What is the probability that this person has the disease?

Solution: Apply Bayes' Rule and the Law of Total Probability:

$$\begin{aligned}
\mathbb{P}(D \mid T) &= \frac{\mathbb{P}(T \mid D)\mathbb{P}(D)}{\mathbb{P}(T)} \\
&= \frac{\mathbb{P}(T \mid D)\mathbb{P}(D)}{\mathbb{P}(T \mid D)\mathbb{P}(D) + \mathbb{P}(T \mid D^c)\mathbb{P}(D^c)} \\
&= \frac{0.95 \times 0.01}{(0.95 \times 0.01) + (0.05 \times 0.99)}
\end{aligned}$$

Example: testing for a rare disease

- Consider a rare disease D that affects 1% of the population.
- Suppose there is a test T that is 95% accurate such that

$$\mathbb{P}(T \mid D) = 0.95 \quad \text{and} \quad \mathbb{P}(T^c \mid D^c) = 0.95.$$

- A randomly selected person takes the test and tests positive. What is the probability that this person has the disease?

Solution: Apply Bayes' Rule and the Law of Total Probability:

$$\begin{aligned}
\mathbb{P}(D \mid T) &= \frac{\mathbb{P}(T \mid D)\mathbb{P}(D)}{\mathbb{P}(T)} \\
&= \frac{\mathbb{P}(T \mid D)\mathbb{P}(D)}{\mathbb{P}(T \mid D)\mathbb{P}(D) + \mathbb{P}(T \mid D^c)\mathbb{P}(D^c)} \\
&= \frac{0.95 \times 0.01}{(0.95 \times 0.01) + (0.05 \times 0.99)} \\
&\approx 16\%.
\end{aligned}$$

Conditional probabilities are probabilities

Conditional probabilities are probabilities

$\mathbb{P}(\cdot \mid E) : S \to [0, 1].$

Conditional probabilities are probabilities

$\mathbb{P}(\cdot \mid E) : S \to [0, 1].$

By conditioning on some event E, we have

Conditional probabilities are probabilities

$\mathbb{P}(\cdot \mid E) : S \to [0, 1]$.

By conditioning on some event E, we have

- $\mathbb{P}(A \mid E) \geq 0$ for any event A.

Conditional probabilities are probabilities

$\mathbb{P}(\cdot \mid E) : S \to [0, 1]$.

By conditioning on some event E, we have

- $\mathbb{P}(A \mid E) \geq 0$ for any event A.
- $\mathbb{P}(\varnothing \mid E) = 0$, $\mathbb{P}(S \mid E) = 1$.

Conditional probabilities are probabilities

$\mathbb{P}(\cdot \mid E) : S \to [0,1]$.

By conditioning on some event E, we have

- $\mathbb{P}(A \mid E) \geq 0$ for any event A.
- $\mathbb{P}(\varnothing \mid E) = 0$, $\mathbb{P}(S \mid E) = 1$.
- If $A_1, \ldots, A_n$ are disjoint, then $\mathbb{P}(\cup_i A_i \mid E) = \sum_i \mathbb{P}(A_i \mid E)$.

Conditional probabilities are probabilities

$\mathbb{P}(\cdot \mid E) : S \to [0, 1]$.

By conditioning on some event E, we have

- $\mathbb{P}(A \mid E) \geq 0$ for any event A.
- $\mathbb{P}(\varnothing \mid E) = 0$, $\mathbb{P}(S \mid E) = 1$.
- If $A_1, \ldots, A_n$ are disjoint, then $\mathbb{P}(\cup_i A_i \mid E) = \sum_i \mathbb{P}(A_i \mid E)$.

Conclusion:

- Conditional probabilities are probabilities.

Conditional probabilities are probabilities

$\mathbb{P}(\cdot \mid E) : S \to [0, 1]$.

By conditioning on some event E, we have

- $\mathbb{P}(A \mid E) \geq 0$ for any event A.
- $\mathbb{P}(\varnothing \mid E) = 0$, $\mathbb{P}(S \mid E) = 1$.
- If $A_1, \dots, A_n$ are disjoint, then $\mathbb{P}(\cup_i A_i \mid E) = \sum_i \mathbb{P}(A_i \mid E)$.

Conclusion:

- Conditional probabilities are probabilities.
- *All probabilities are conditional probabilities.*

Independence

Definition

Events A and B are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Independence

Definition

Events A and B are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

- If $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$, this is equivalent to

$$\mathbb{P}(A \mid B) = \mathbb{P}(A) \quad \text{and} \quad \mathbb{P}(B \mid A) = \mathbb{P}(B).$$

Independence

Definition

Events A and B are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

- If $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$, this is equivalent to

$$\mathbb{P}(A \mid B) = \mathbb{P}(A) \quad \text{and} \quad \mathbb{P}(B \mid A) = \mathbb{P}(B).$$

- Independence is a symmetric relation: if A is independent of B, B is independent of A.

Independence

Definition

Events A and B are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

- If $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$, this is equivalent to

$$\mathbb{P}(A \mid B) = \mathbb{P}(A) \quad \text{and} \quad \mathbb{P}(B \mid A) = \mathbb{P}(B).$$

- Independence is a symmetric relation: if A is independent of B, B is independent of A.
- Warning: independence $\neq$ disjointness. In fact, disjoint events can only be independent if $\mathbb{P}(A) = 0$ or $\mathbb{P}(B) = 0$. (Why?)

Conditional independence

Definition

Events A and B are said to be *conditionally independent* given a third event E if

$$\mathbb{P}(A \cap B \mid E) = \mathbb{P}(A \mid E)\mathbb{P}(B \mid E).$$

Conditional independence

Definition

Events A and B are said to be *conditionally independent* given a third event E if

$$\mathbb{P}(A \cap B \mid E) = \mathbb{P}(A \mid E)\mathbb{P}(B \mid E).$$

Health warning: independence does *not* imply conditional independence and vice versa.

Conditional independence and complements

Conditional independence and complements

- Suppose there are two types of teachers: those who give grades that reflect student effort (E), and those who randomly assign grades, regardless of student effort (E^c).

Conditional independence and complements

- Suppose there are two types of teachers: those who give grades that reflect student effort (E), and those who randomly assign grades, regardless of student effort (E^c).
- Let W be the event that you work hard and let G be the event that you receive a good grade.

Conditional independence and complements

- Suppose there are two types of teachers: those who give grades that reflect student effort (E), and those who randomly assign grades, regardless of student effort (E^c).
- Let W be the event that you work hard and let G be the event that you receive a good grade.
- Then W and G are conditionally independent given E^c, but they are not conditionally independent given E.

Conditional independence $\not\Rightarrow$ independence

Conditional independence $\not\Rightarrow$ independence

- You have one fair coin and one biased coin which lands Heads with probability 3/4.

Conditional independence $\not\Rightarrow$ independence

- You have one fair coin and one biased coin which lands Heads with probability 3/4.
- You pick one of the coins at random, without knowing which one you've chosen, and flip it several times.

Conditional independence $\not\Rightarrow$ independence

- You have one fair coin and one biased coin which lands Heads with probability 3/4.
- You pick one of the coins at random, without knowing which one you've chosen, and flip it several times.
- Conditional on choosing either the fair or the biased coin, the coin flips are independent.

Conditional independence $\not\Rightarrow$ independence

- You have one fair coin and one biased coin which lands Heads with probability 3/4.
- You pick one of the coins at random, without knowing which one you've chosen, and flip it several times.
- Conditional on choosing either the fair or the biased coin, the coin flips are independent.
- However, the coin flips are not unconditionally independent: without knowing which coin we've chosen, each flip gives us new data from which we can predict outcomes of future tosses.

Conditional independence $\not\Rightarrow$ independence

- You have one fair coin and one biased coin which lands Heads with probability 3/4.
- You pick one of the coins at random, without knowing which one you've chosen, and flip it several times.
- Conditional on choosing either the fair or the biased coin, the coin flips are independent.
- However, the coin flips are not unconditionally independent: without knowing which coin we've chosen, each flip gives us new data from which we can predict outcomes of future tosses.
- (Think about the definition of independence.)

Independence $\not\Rightarrow$ conditional independence

Independence $\not\Rightarrow$ conditional independence

- Suppose my friends Alice and Ben are the only two people who call me on my mobile phone.

Independence $\not\Rightarrow$ conditional independence

- Suppose my friends Alice and Ben are the only two people who call me on my mobile phone.
- Each day, they decide independently whether to call me that day.

Independence $\not\Rightarrow$ conditional independence

- Suppose my friends Alice and Ben are the only two people who call me on my mobile phone.
- Each day, they decide independently whether to call me that day.
- Let A be the event that Alice calls me tomorrow and let B be the event that Ben calls me tomorrow.

Independence $\not\Rightarrow$ conditional independence

- Suppose my friends Alice and Ben are the only two people who call me on my mobile phone.
- Each day, they decide independently whether to call me that day.
- Let A be the event that Alice calls me tomorrow and let B be the event that Ben calls me tomorrow.
- Then A and B are unconditionally independent, with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$.

Independence $\not\Rightarrow$ conditional independence

- Suppose my friends Alice and Ben are the only two people who call me on my mobile phone.
- Each day, they decide independently whether to call me that day.
- Let A be the event that Alice calls me tomorrow and let B be the event that Ben calls me tomorrow.
- Then A and B are unconditionally independent, with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$.
- However, given that I receive exactly one call tomorrow (C), A and B are no longer independent:

$$\mathbb{P}(A \mid C) > 0, \quad \text{but} \quad \mathbb{P}(A \mid C \cap B) = 0.$$

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) = 0.33504
```

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

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

## P(car | no switch) = 0.33504
```

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

## P(car | switch) = 0.66496
```

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 have chosen the winning door.

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 have chosen the winning door.
- 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 have chosen the winning door.
- 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 chose door 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 have chosen the winning door.
- 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 chose door 1.
- 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 have chosen the winning door.
- 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 chose door 1.
- 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 have chosen the winning door.
- 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 chose door 1.
- 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, $\mathbb{P}(W) = 0 \times \frac{1}{3} + 1 \times \frac{1}{3} + 1 \times \frac{1}{3} = \frac{2}{3}$.

Problem sheet 1: preamble

```r
# Clear environment, set working directory
rm(list = ls())
setwd("/Users/Elias/Documents/.../QM/Problem sheets")
```

```r
# Load tidyverse packages
library(tidyverse)
```

Question A1

```r
# Construct die and sample space
die <- 1:6
S <- expand.grid(die, die, die)
```

Question A2

```r
# Create new variable
S <- S %>%
  mutate(Value = Var1 + Var2 + Var3)
```

Question A3

```r
# P(A) = # elements in A / # elements in S
sum(S$Value == 12) / nrow(S)

## [1] 0.1157407
```

Question A4

```r
# What if dice are biased?
Prob <- c("1" = 1/8, "2" = 1/8, "3" = 1/8,
          "4" = 1/8, "5" = 1/8, "6" = 3/8)
```

Question A4

```r
# What if dice are biased?
Prob <- c("1" = 1/8, "2" = 1/8, "3" = 1/8,
          "4" = 1/8, "5" = 1/8, "6" = 3/8)
```

```r
# Assign individual and joint probabilities
S$Prob1 <- Prob[S$Var1]
S$Prob2 <- Prob[S$Var2]
S$Prob3 <- Prob[S$Var3]
S$Prob_joint <- S$Prob1 * S$Prob2 * S$Prob3
```

Question A5

```r
# Extract event A and calculate probability
A <- subset(S, Value == 12)
sum(A$Prob_joint)

## [1] 0.1074219
```

Question B1

```r
#  Load data sets
kenya <- read_csv("kenya.csv")
sweden <- read_csv("sweden.csv")
world <- read_csv("world.csv")
```

Question B2

```r
#  Inspect
summary(kenya)
summary(sweden)
summary(world)

glimpse(kenya)
glimpse(sweden)
glimpse(world)

head(kenya)
head(sweden)
head(world)

print(kenya, n = 30)
print(sweden, n = 30)
print(world, n = 30)
```

Question B3

```r
# Calculate age-specific fertility rate
asfr <- function(data) {
  data %>%
    mutate(
      asfr = births / py.women) %>%
    select(period, age, asfr) %>%
    data.frame() # Convert tibble to data frame
}
```

Question B3 (cont.)

```
asfr(kenya)[c(4:10, 19:25), ]

##         period   age       asfr
## 4  1950-1955 15-19 0.16884585
## 5  1950-1955 20-24 0.35596942
## 6  1950-1955 25-29 0.34657814
## 7  1950-1955 30-34 0.28946367
## 8  1950-1955 35-39 0.20644016
## 9  1950-1955 40-44 0.11193267
## 10 1950-1955 45-49 0.03905205
## 19 2005-2010 15-19 0.10057087
## 20 2005-2010 20-24 0.23583536
## 21 2005-2010 25-29 0.23294721
## 22 2005-2010 30-34 0.18087964
## 23 2005-2010 35-39 0.13126805
## 24 2005-2010 40-44 0.05626214
## 25 2005-2010 45-49 0.03815044
```

Question B3 (cont.)

```
asfr(sweden)[c(4:10, 19:25), ]

##        period   age         asfr
## 4  1950-1955 15-19 0.0389089519
## 5  1950-1955 20-24 0.1277108826
## 6  1950-1955 25-29 0.1252436647
## 7  1950-1955 30-34 0.0873641591
## 8  1950-1955 35-39 0.0486037714
## 9  1950-1955 40-44 0.0162101857
## 10 1950-1955 45-49 0.0013418290
## 19 2005-2010 15-19 0.0059709097
## 20 2005-2010 20-24 0.0507320271
## 21 2005-2010 25-29 0.1162085625
## 22 2005-2010 30-34 0.1322744621
## 23 2005-2010 35-39 0.0625923991
## 24 2005-2010 40-44 0.0121600765
## 25 2005-2010 45-49 0.0006143942
```

Question B3 (cont.)

```
asfr(world)[c(4:10, 19:25), ]

##       period   age       asfr
## 4  1950-1955 15-19 0.090295213
## 5  1950-1955 20-24 0.237633702
## 6  1950-1955 25-29 0.252452289
## 7  1950-1955 30-34 0.204164096
## 8  1950-1955 35-39 0.138105344
## 9  1950-1955 40-44 0.063608319
## 10 1950-1955 45-49 0.015190644
## 19 2005-2010 15-19 0.048489719
## 20 2005-2010 20-24 0.151971307
## 21 2005-2010 25-29 0.146980966
## 22 2005-2010 30-34 0.093813813
## 23 2005-2010 35-39 0.046689639
## 24 2005-2010 40-44 0.016268995
## 25 2005-2010 45-49 0.004510245
```

Question B4

```r
# Calculate total fertility rate
tfr <- function(data) {
  out <- asfr(data)
  out %>%
    group_by(period) %>%
    summarise(
      tfr = 5 * sum(asfr))
}
```

Question B4 (cont.)

```
tfr(kenya)

## # A tibble: 2 x 2
##   period      tfr
##   <chr>     <dbl>
## 1 1950-1955  7.59
## 2 2005-2010  4.88

tfr(sweden)

## # A tibble: 2 x 2
##   period      tfr
##   <chr>     <dbl>
## 1 1950-1955  2.23
## 2 2005-2010  1.90
```

Question B4 (cont.)

```
tfr(world)

## # A tibble: 2 x 2
##   period      tfr
##   <chr>     <dbl>
## 1 1950-1955  5.01
## 2 2005-2010  2.54
```

Question B5

```r
# Calculate age-specific death rate
asdr <- function(data) {
  data %>%
    mutate(
      # Convert rates to per 1000 population
      asdr = 1000 * deaths / (py.men + py.women)) %>%
    select(period, age, asdr) %>%
    data.frame() # Convert tibble to data frame
}
```

Question B5 (cont.)

```
sample_n(asdr(kenya), 10)

##       period   age      asdr
## 1  1950-1955 15-19  5.869582
## 2  2005-2010 60-69 25.395531
## 3  1950-1955 10-14  5.972093
## 4  1950-1955 55-59 24.433007
## 5  2005-2010 30-34 10.603913
## 6  1950-1955 60-69 41.996801
## 7  1950-1955 20-24  7.651103
## 8  1950-1955 40-44 12.633744
## 9  1950-1955   0-4 66.826532
## 10 2005-2010 20-24  3.885368
```

Question B5 (cont.)

```
sample_n(asdr(sweden), 10)

##       period   age         asdr
## 1  1950-1955   80+ 167.81702554
## 2  2005-2010   5-9   0.08138094
## 3  2005-2010 40-44   1.03925622
## 4  2005-2010 25-29   0.49414399
## 5  2005-2010 60-69   9.82877193
## 6  1950-1955 10-14   0.48964064
## 7  1950-1955 25-29   1.11409103
## 8  1950-1955 40-44   2.50955411
## 9  1950-1955 70-79  59.98230926
## 10 1950-1955   5-9   0.43205374
```

Question B5 (cont.)

```
sample_n(asdr(world), 10)

##        period   age       asdr
## 1  2005-2010 30-34   2.623982
## 2  1950-1955   80+ 184.364978
## 3  1950-1955 45-49  13.459846
## 4  1950-1955 30-34   7.132501
## 5  2005-2010   80+ 120.679385
## 6  2005-2010 70-79  47.457519
## 7  1950-1955 55-59  24.265320
## 8  2005-2010 20-24   1.832602
## 9  2005-2010 35-39   3.031563
## 10 2005-2010 40-44   3.753402
```

Question B6

```r
# Collect ASFR and ASDR for each country
ken <- left_join(asfr(kenya), asdr(kenya))
swe <- left_join(asfr(sweden), asdr(sweden))
wor <- left_join(asfr(world), asdr(world))
```

Question B6

```r
# Collect ASFR and ASDR for each country
ken <- left_join(asfr(kenya), asdr(kenya))
swe <- left_join(asfr(sweden), asdr(sweden))
wor <- left_join(asfr(world), asdr(world))
```

```r
# Create one data frame with all results
df <- rbind(ken, swe, wor)
df$country <- c(rep("Kenya", 30),
                rep("Sweden", 30), rep("World", 30))
```

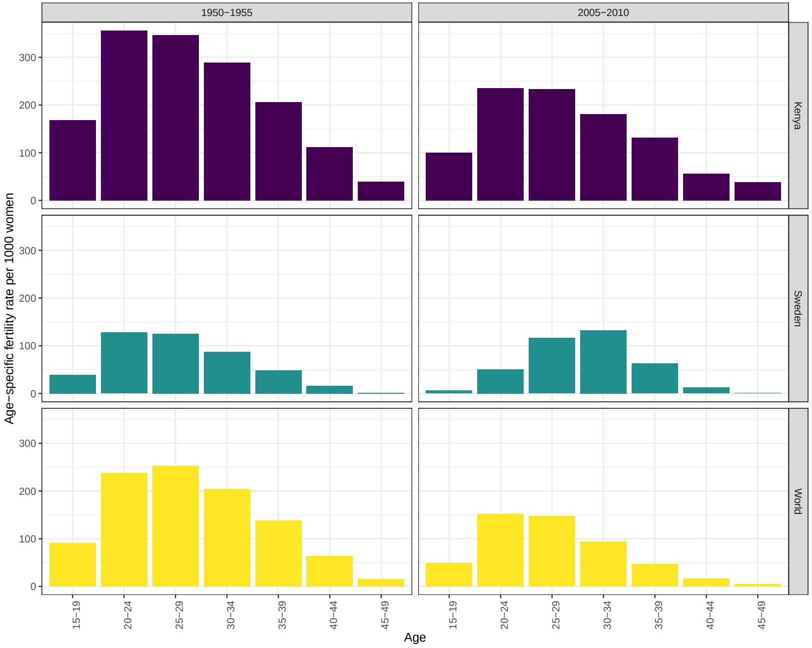
Question B6 (cont.)

```r
# Transform age groups to ordered factor
df$age <- factor(df$age,
                 levels = c("0-4", "5-9", "10-14",
                            "15-19", "20-24", "25-29",
                            "30-34", "35-39", "40-44",
                            "45-49", "50-54", "55-59",
                            "60-69", "70-79", "80+"))

# Age groups for reproductive age range
age_groups <- c("15-19", "20-24", "25-29",
                "30-34", "35-39", "40-44", "45-49")
```

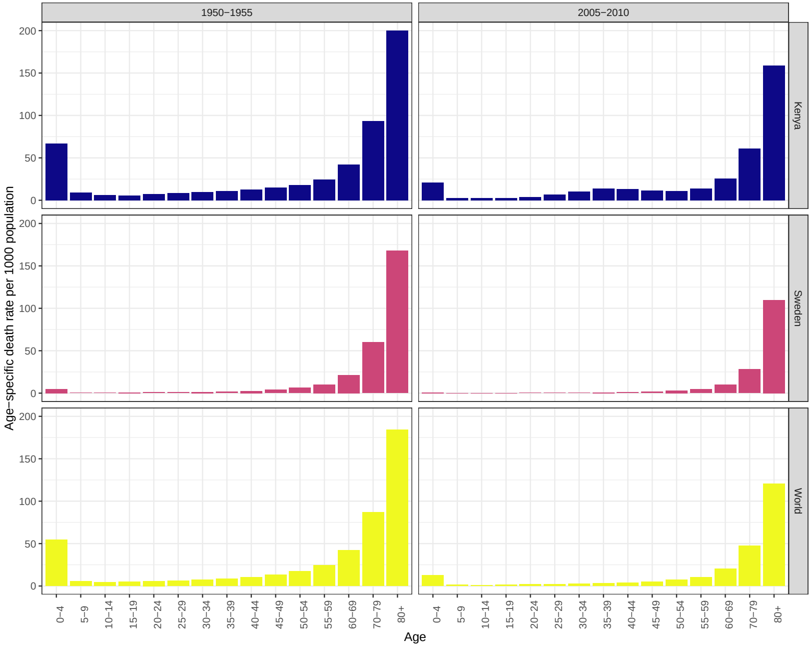
Question B6 (cont.)

```r
# Visualise ASFR
g1 <- ggplot(subset(df, age %in% age_groups), # Age range
             aes(age, 1000 * asfr, # Modify rate
                 fill = country)) + # Colour code
  geom_col() + # Show as columns
  labs( # Axis labels
    x = "Age",
    y = "Age-specific fertility rate per 1000 women") +
  scale_fill_viridis_d() + # Choose a nice colour palette
  facet_grid(country ~ period) + # Stratify
  theme_bw() + # Remove redundant lines
  theme( # Avoid cluttering
    legend.position = "none",
    axis.text.x = element_text(angle = 90))
```


1950−1955
2005−2010
Kenya
Sweden
World
Age−specific fertility rate per 1000 women
0
100
200
300
15−19
20−24
25−29
30−34
35−39
40−44
45−49
Age

Question B6 (cont.)

```
# Visualise ASDR
g2 <- ggplot(df, aes(age, asdr,
                     fill = country)) + # Colour code
  geom_col() + # Show as columns
  scale_y_continuous(breaks = seq(0, 200, 50)) + # Y-axis
  labs( # Axis labels
    x = "Age",
    y = "Age-specific death rate per 1000 population") +
  scale_fill_viridis_d(option = "plasma") + # Colour
  facet_grid(country ~ period) + # Stratify
  theme_bw() + # Remove redundant lines
  theme( # Avoid cluttering
    legend.position = "none",
    axis.text.x = element_text(angle = 90))
```


1950−1955
2005−2010
Kenya
Sweden
World
Age−specific death rate per 1000 population
0
50
100
150
200
0−4
5−9
10−14
15−19
20−24
25−29
30−34
35−39
40−44
45−49
50−54
55−59
60−69
70−79
80+
Age