THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

SDS 321: Introduction to Probability and Statistics

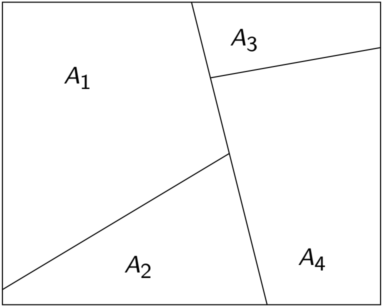
Lecture 4: Total probability theorem and Bayes rule

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

psarkar.github.io/teaching/

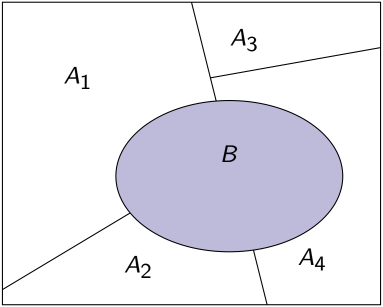
Total Probability Theorem

- Next, we're going to look at ways of obtaining the probability of a subset, using conditional probabilities.
- Let $A_1, \ldots, A_n$ be a partition of Ω, such that $P(A_i) > 0$ for all A_i.



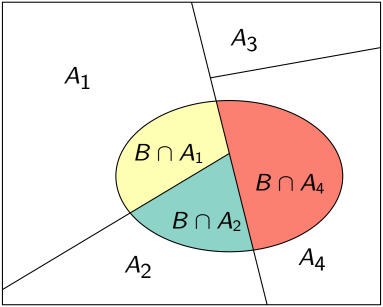
Total Probability Theorem

- Next, we're going to look at ways of obtaining the probability of a subset, using conditional probabilities.
- Let $A_1, \ldots, A_n$ be a partition of Ω, such that $P(A_i) > 0$ for all A_i.
- Let B be an event.



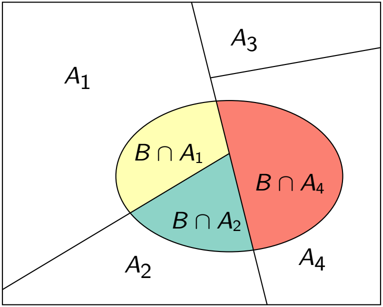
Total Probability Theorem

- Let B be an event.
- Note that $B = \cup_i (A_i \cap B)$.
- Therefore, $P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_3 \cap B)$.



Total Probability Theorem

- By the multiplication rule, $P(A_i \cap B) = P(A_i)P(B|A_i)$.
- So, $P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_n)P(A_n)$.
- This is known as the **Total Probability Theorem**



Revisit Example: Umbrella sales

Anita works for an umbrella company. She gets a bonus (event B) iff she sells more than 10 umbrellas in a day (event W).

- Now $W = B$. We have $P(R) = 0.1$, $P(W|R) = 0.8$ and $P(W|R^c) = 0.25$.
- If you knew that Anita got a bonus, then what is the probability that it rained?

Revisit Example: Umbrella sales

Anita works for an umbrella company. She gets a bonus (event B) iff she sells more than 10 umbrellas in a day (event W).

- Now $W = B$. We have $P(R) = 0.1$, $P(W|R) = 0.8$ and $P(W|R^c) = 0.25$.
- If you knew that Anita got a bonus, then what is the probability that it rained?
- We are interested in $P(R|B)$, which is the same as $P(R|W)$.

Revisit Example: Umbrella sales

Anita works for an umbrella company. She gets a bonus (event B) iff she sells more than 10 umbrellas in a day (event W).

- Now $W = B$. We have $P(R) = 0.1$, $P(W|R) = 0.8$ and $P(W|R^c) = 0.25$.
- If you knew that Anita got a bonus, then what is the probability that it rained?
- We are interested in $P(R|B)$, which is the same as $P(R|W)$.
- First we need $P(R \cap W)$ and then we need $P(W)$.
- $P(R \cap W) = P(W|R)P(R) = 0.8 \times 0.1 = 0.08$.

Revisit Example: Umbrella sales

Anita works for an umbrella company. She gets a bonus (event B) iff she sells more than 10 umbrellas in a day (event W).

- Now $W = B$. We have $P(R) = 0.1$, $P(W|R) = 0.8$ and $P(W|R^c) = 0.25$.
- If you knew that Anita got a bonus, then what is the probability that it rained?
- We are interested in $P(R|B)$, which is the same as $P(R|W)$.
- First we need $P(R \cap W)$ and then we need $P(W)$.
- $P(R \cap W) = P(W|R)P(R) = 0.8 \times 0.1 = 0.08$.
- Now what about $P(W)$?

Revisit Example: Umbrella sales

Anita works for an umbrella company. She gets a bonus iff she sells more than 10 umbrellas in a day.

- Now what about $P(W)$?

Revisit Example: Umbrella sales

Anita works for an umbrella company. She gets a bonus iff she sells more than 10 umbrellas in a day.

- Now what about $P(W)$?
  - First write W as an union of two disjoint events. Guesses?

Revisit Example: Umbrella sales

Anita works for an umbrella company. She gets a bonus iff she sells more than 10 umbrellas in a day.

- Now what about $P(W)$?
  - First write W as an union of two disjoint events. Guesses?
  - $W = W \cap \Omega = W \cap (R \cup R^c) = (W \cap R) \cup (W \cap R^c)$.

Revisit Example: Umbrella sales

Anita works for an umbrella company. She gets a bonus iff she sells more than 10 umbrellas in a day.

- Now what about $P(W)$?
  - First write W as an union of two disjoint events. Guesses?
  - $W = W \cap \Omega = W \cap (R \cup R^c) = (W \cap R) \cup (W \cap R^c)$.
  - Now additivity gives

$$\begin{aligned} P(W) &= P(W \cap R) + P(W \cap R^c) \\ &= P(W|R)P(R) + P(W|R^c)P(R^c) = 0.08 + 0.25 \times 0.9 \approx 0.3 \end{aligned}$$

Revisit Example: Umbrella sales

Anita works for an umbrella company. She gets a bonus iff she sells more than 10 umbrellas in a day.

- Now what about $P(W)$?
  - First write W as an union of two disjoint events. Guesses?
  - $W = W \cap \Omega = W \cap (R \cup R^c) = (W \cap R) \cup (W \cap R^c)$.
  - Now additivity gives

$$\begin{aligned} P(W) &= P(W \cap R) + P(W \cap R^c) \\ &= P(W|R)P(R) + P(W|R^c)P(R^c) = 0.08 + 0.25 \times 0.9 \approx 0.3 \end{aligned}$$

$$\begin{aligned} P(R|B) = P(R|W) &= \frac{P(R \cap W)}{P(W)} \\ &= \frac{P(W|R)P(R)}{P(W|R)P(R) + P(W|R^c)P(R^c)} = 0.08/0.3 \approx 0.27 \end{aligned}$$

Revisit Example: Umbrella sales

Anita works for an umbrella company. She gets a bonus iff she sells more than 10 umbrellas in a day.

- Now what about $P(W)$?
  - First write W as an union of two disjoint events. Guesses?
  - $W = W \cap \Omega = W \cap (R \cup R^c) = (W \cap R) \cup (W \cap R^c)$.
  - Now additivity gives

$$
\begin{aligned}
P(W) &= P(W \cap R) + P(W \cap R^c) \qquad \text{Theorem of total probability} \\
&= P(W|R)P(R) + P(W|R^c)P(R^c) = 0.08 + 0.25 \times 0.9 \approx 0.3
\end{aligned}
$$

$$
\begin{aligned}
P(R|B) = P(R|W) &= \frac{P(R \cap W)}{P(W)} \\
&= \frac{P(W|R)P(R)}{P(W|R)P(R) + P(W|R^c)P(R^c)} = 0.08/0.3 \approx 0.27
\end{aligned}
$$

Revisit Example: Umbrella sales

Anita works for an umbrella company. She gets a bonus iff she sells more than 10 umbrellas in a day.

- Now what about $P(W)$?
  - First write W as an union of two disjoint events. Guesses?
  - $W = W \cap \Omega = W \cap (R \cup R^c) = (W \cap R) \cup (W \cap R^c)$.
  - Now additivity gives

$$\begin{aligned} P(W) &= P(W \cap R) + P(W \cap R^c) \qquad \text{Theorem of total probability} \\ &= P(W|R)P(R) + P(W|R^c)P(R^c) = 0.08 + 0.25 \times 0.9 \approx 0.3 \end{aligned}$$

$$\begin{aligned} P(R|B) = P(R|W) &= \frac{P(R \cap W)}{P(W)} \\ &= \frac{P(W|R)P(R)}{P(W|R)P(R) + P(W|R^c)P(R^c)} = 0.08/0.3 \approx 0.27 \end{aligned}$$

The last step is also known as **Bayes Rule**, which we will study next.

Bayes rule

- Simple rule to get conditional probability of A given B, from the conditional formula of B given A.

$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \\ &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} \end{aligned}$$

- This is very useful for **inferring** hidden causes underlying our observations.

When do you use Bayes' rule?

- You have a homework question asking for $P(A|B)$.
- First check what is given to you.
- If you have $P(A|B)$ or $P(A^c|B)$ then you are all set.
- If not, then you have to use Bayes rule.
- For these you will need to know $P(B|A)$ and $P(B|A^c)$ and $P(A)$. If you don't have these, then either the question is incomplete, or you should read it again.

Typical Bayes rule example

- Consider testing for some latent (hidden/unobservable) disease, that won't become symptomatic until a future time point.
- We can directly observe the outcome of the test.
- Assuming the test isn't 100% accurate, we *can't* directly observe whether we have the disease.
- We have two possible **hidden causes** for a positive test result:
  - We have the disease, and the test is correct.
  - We don't have the disease, and the test is a *false positive*.
- We want to **infer** *which* hidden cause underlies our observation.

Inference: Disease testing

Let's add some numbers to this example. Let's assume:

- The disease affects 2% of the population.
- The false positive rate is 1%.
- The false negative rate is 5%.
- If you take the test and the result is positive, you are really interested in the question: Given that you tested positive, what is the chance you have the disease?

Inference: Disease testing

Let's add some numbers to this example. Let's assume:

- The disease affects 2% of the population.
- The false positive rate is 1%.
- The false negative rate is 5%.
- If you take the test and the result is positive, you are really interested in the question: Given that you tested positive, what is the chance you have the disease?
- Let T be the event "tests positive" and "D" be the event "has disease".

Inference: Disease testing

Let's add some numbers to this example. Let's assume:

- The disease affects 2% of the population.
- The false positive rate is 1%.
- The false negative rate is 5%.
- If you take the test and the result is positive, you are really interested in the question: Given that you tested positive, what is the chance you have the disease?
- Let T be the event "tests positive" and "D" be the event "has disease".
- We know that:

$$P(D) = 0.02 \qquad P(T^c|D) = 0.05 \qquad P(T|D^c) = 0.01$$

Another application of Bayes' rule

If you take the test and the result is positive, you are really interested in the question: Given that you tested positive, what is the chance you have the disease?

i.e. what is $P(D|T)$? Bayes' rule gives us:

$$P(D|T) = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^c)P(D^c)}$$

It lets us get from the conditional probability of an observation given a hidden cause (which we usually know), to the conditional probability of a hidden cause given an observation (which we usually care about!)

Inference: Disease testing

So, let's plug in the numbers. Recall

$$P(D) = 0.02 \qquad P(T^c|D) = 0.05 \qquad P(T|D^c) = 0.01$$

So, $P(T|D) = 1 - 0.05 = 0.95$.

$$\begin{aligned}
P(D|T) =& \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^c)P(D^c)} \\
=& \frac{0.95 \times 0.02}{0.95 \times 0.02 + 0.01 \times 0.98} \\
=& \frac{0.019}{0.0288} = .66
\end{aligned}$$

More examples to apply Bayes' Rule (Coding)

Alice is sending a coded message to Bob using “dots” and “dashes”, which are known to occur in the proportion of $3:4$ for Morse codes. Because of interference on the transmission line, a dot can be mistakenly received as a dash with probability $1/8$ and vice-versa. If Bob receives a “dot”, what is the probability that Alice sent a “dot”.

More examples to apply Bayes' Rule (Coding)

Alice is sending a coded message to Bob using "dots" and "dashes", which are known to occur in the proportion of $3:4$ for Morse codes. Because of interference on the transmission line, a dot can be mistakenly received as a dash with probability $1/8$ and vice-versa. If Bob receives a "dot", what is the probability that Alice sent a "dot".

- Notation: $\text{dotS} = \{\text{dot sent}\}$, $\text{dashS} = \{\text{dash sent}\}$, $\text{dotR} = \{\text{dot received}\}$, $\text{dashR} = \{\text{dash received}\}$.
- $P(\text{dotS}) = \frac{3}{7}$, $P(\text{dashS}) = \frac{4}{7}$ and $P(\text{dashR}|\text{dotS}) = P(\text{dotR}|\text{dashS}) = \frac{1}{8}$.

More examples to apply Bayes' Rule (Coding)

Alice is sending a coded message to Bob using "dots" and "dashes", which are known to occur in the proportion of $3:4$ for Morse codes. Because of interference on the transmission line, a dot can be mistakenly received as a dash with probability $1/8$ and vice-versa. If Bob receives a "dot", what is the probability that Alice sent a "dot".

- Notation: $\text{dotS} = \{\text{dot sent}\}$, $\text{dashS} = \{\text{dash sent}\}$, $\text{dotR} = \{\text{dot received}\}$, $\text{dashR} = \{\text{dash received}\}$.
- $P(\text{dotS}) = \frac{3}{7}$, $P(\text{dashS}) = \frac{4}{7}$ and $P(\text{dashR}|\text{dotS}) = P(\text{dotR}|\text{dashS}) = \frac{1}{8}$.
- We want:
  $P(\text{dotS}|\text{dotR})$

More examples to apply Bayes' Rule (Coding)

Alice is sending a coded message to Bob using "dots" and "dashes", which are known to occur in the proportion of $3:4$ for Morse codes. Because of interference on the transmission line, a dot can be mistakenly received as a dash with probability $1/8$ and vice-versa. If Bob receives a "dot", what is the probability that Alice sent a "dot".

- Notation: $\text{dotS} = \{\text{dot sent}\}$, $\text{dashS} = \{\text{dash sent}\}$, $\text{dotR} = \{\text{dot received}\}$, $\text{dashR} = \{\text{dash received}\}$.
- $P(\text{dotS}) = \frac{3}{7}$, $P(\text{dashS}) = \frac{4}{7}$ and $P(\text{dashR}|\text{dotS}) = P(\text{dotR}|\text{dashS}) = \frac{1}{8}$.
- We want:
  $P(\text{dotS}|\text{dotR}) = \frac{P(\text{dotR}|\text{dotS})P(\text{dotS})}{P(\text{dotR})}$

More examples to apply Bayes' Rule (Coding)

Alice is sending a coded message to Bob using "dots" and "dashes", which are known to occur in the proportion of $3:4$ for Morse codes. Because of interference on the transmission line, a dot can be mistakenly received as a dash with probability $1/8$ and vice-versa. If Bob receives a "dot", what is the probability that Alice sent a "dot".

- Notation: $\text{dotS} = \{\text{dot sent}\}$, $\text{dashS} = \{\text{dash sent}\}$, $\text{dotR} = \{\text{dot received}\}$, $\text{dashR} = \{\text{dash received}\}$.
- $P(\text{dotS}) = \frac{3}{7}$, $P(\text{dashS}) = \frac{4}{7}$ and $P(\text{dashR}|\text{dotS}) = P(\text{dotR}|\text{dashS}) = \frac{1}{8}$.
- We want:
  $$P(\text{dotS}|\text{dotR}) = \frac{P(\text{dotR}|\text{dotS})P(\text{dotS})}{P(\text{dotR})} = \frac{(1-1/8)\times 3/7}{P(\text{dotR})} = \frac{3/8}{P(\text{dotR})}$$

More examples to apply Bayes' Rule (Coding)

Alice is sending a coded message to Bob using "dots" and "dashes", which are known to occur in the proportion of $3:4$ for Morse codes. Because of interference on the transmission line, a dot can be mistakenly received as a dash with probability $1/8$ and vice-versa. If Bob receives a "dot", what is the probability that Alice sent a "dot".

- Notation: $\text{dotS} = \{\text{dot sent}\}$, $\text{dashS} = \{\text{dash sent}\}$, $\text{dotR} = \{\text{dot received}\}$, $\text{dashR} = \{\text{dash received}\}$.
- $P(\text{dotS}) = \frac{3}{7}$, $P(\text{dashS}) = \frac{4}{7}$ and $P(\text{dashR}|\text{dotS}) = P(\text{dotR}|\text{dashS}) = \frac{1}{8}$.
- We want:
$$P(\text{dotS}|\text{dotR}) = \frac{P(\text{dotR}|\text{dotS})P(\text{dotS})}{P(\text{dotR})} = \frac{(1-1/8)\times 3/7}{P(\text{dotR})} = \frac{3/8}{P(\text{dotR})}$$
- $P(\text{dotR}) = P(\text{dotR}|\text{dotS})P(\text{dotS}) + P(\text{dotR}|\text{dashS})P(\text{dashS})$
$$= \frac{3}{8} + \frac{1}{8}\times\frac{4}{7} = \frac{25}{56}$$

More examples to apply Bayes' Rule (Coding)

Alice is sending a coded message to Bob using "dots" and "dashes", which are known to occur in the proportion of $3:4$ for Morse codes. Because of interference on the transmission line, a dot can be mistakenly received as a dash with probability $1/8$ and vice-versa. If Bob receives a "dot", what is the probability that Alice sent a "dot".

- Notation: $\text{dotS} = \{\text{dot sent}\}$, $\text{dashS} = \{\text{dash sent}\}$, $\text{dotR} = \{\text{dot received}\}$, $\text{dashR} = \{\text{dash received}\}$.
- $P(\text{dotS}) = \frac{3}{7}$, $P(\text{dashS}) = \frac{4}{7}$ and $P(\text{dashR}|\text{dotS}) = P(\text{dotR}|\text{dashS}) = \frac{1}{8}$.
- We want:
  $$P(\text{dotS}|\text{dotR}) = \frac{P(\text{dotR}|\text{dotS})P(\text{dotS})}{P(\text{dotR})} = \frac{(1-1/8)\times 3/7}{P(\text{dotR})} = \frac{3/8}{P(\text{dotR})}$$
- $$P(\text{dotR}) = P(\text{dotR}|\text{dotS})P(\text{dotS}) + P(\text{dotR}|\text{dashS})P(\text{dashS})$$
  $$= \frac{3}{8} + \frac{1}{8}\times\frac{4}{7} = \frac{25}{56}$$
- So we finally have: $P(\text{dotS}|\text{dotR}) = \frac{3/8}{25/56} = \frac{21}{25}$.

Bayes' Rule

Figure: Thomas Bayes, 1701-1761. English statistician, philosopher and Presbyterian minister

Bayes' rule: Let A_1, $A_2, \ldots, A_n$, be a partition of the sample space, and let B be any set. Then, for each $i = 1, 2, \ldots, n$,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}$$

Example: The Monty Hall problem

You are a contestant on a game show, where you have to pick one of three boxes (say A, B, and C) to open. One of the three boxes contains money, the rest are empty. Assume the host knows which box contains the money.

You pick a box, say A. To build suspense, the host opens one of the other two boxes (say B) revealing it is empty. He asks, do you want to stick with your existing box or switch?

What do you do? Does it make any difference?

The Monty Hall problem

- Our outcome consists of *two* random variables: where the money is, and which box the host opens.
- We will observe which box is opened; we want to infer where the money is.
- Let's write M_A for the event "money in A", M_B for "money in B", M_C for "money in C".
- H_A for "host opens A", H_B for "host opens B", H_C for "host opens C".
- What is $P(M_A)$?

The Monty Hall problem

- Our outcome consists of *two* random variables: where the money is, and which box the host opens.
- We will observe which box is opened; we want to infer where the money is.
- Let's write M_A for the event "money in A", M_B for "money in B", M_C for "money in C".
- H_A for "host opens A", H_B for "host opens B", H_C for "host opens C".
- What is $P(M_A)$?
- $P(M_A) = P(M_B) = P(M_C) = 1/3$.

The Monty Hall problem

- You picked box A (without loss of generality).
- For every possible location of the money, we can calculate the conditional probability of the host opening a given box.
  - We assume that the host opened a box he knew to be empty.
  - We know he's not going to open box A – that's the box we picked.

- *If the money is in A, what is the probability that he opens box B?*

The Monty Hall problem

- You picked box A (without loss of generality).
- For every possible location of the money, we can calculate the conditional probability of the host opening a given box.
  - We assume that the host opened a box he knew to be empty.
  - We know he's not going to open box A – that's the box we picked.

- *If the money is in A, what is the probability that he opens box B?*
- $P(H_B|M_A) = 1/2$ (and similarly, $P(H_C|M_A) = 1/2$).
- *If the money is in B, what is the probability that he opens box B?*

The Monty Hall problem

- You picked box A (without loss of generality).
- For every possible location of the money, we can calculate the conditional probability of the host opening a given box.
  - We assume that the host opened a box he knew to be empty.
  - We know he's not going to open box A – that's the box we picked.

- *If the money is in A, what is the probability that he opens box B?*
- $P(H_B|M_A) = 1/2$ (and similarly, $P(H_C|M_A) = 1/2$).
- *If the money is in B, what is the probability that he opens box B?*
- $P(H_B|M_B) = 0$ (and similarly, $P(H_C|M_B) = 1$).
- *If the money is in C, what is the probability that he opens box B?*

The Monty Hall problem

- You picked box A (without loss of generality).
- For every possible location of the money, we can calculate the conditional probability of the host opening a given box.
  - We assume that the host opened a box he knew to be empty.
  - We know he's not going to open box A – that's the box we picked.

- *If the money is in A, what is the probability that he opens box B?*
- $P(H_B|M_A) = 1/2$ (and similarly, $P(H_C|M_A) = 1/2$).
- *If the money is in B, what is the probability that he opens box B?*
- $P(H_B|M_B) = 0$ (and similarly, $P(H_C|M_B) = 1$).
- *If the money is in C, what is the probability that he opens box B?*
- $P(H_B|M_C) = 1$ (and similarly, $P(H_C|M_B) = 0$).

The Monty Hall problem

- So, if the host opens box B, what's the probability that the money is in box C?
- By Bayes' Rule,

$$P(M_C|H_B) = \frac{P(H_B|M_C)P(M_C)}{P(H_B)}$$

The Monty Hall problem

- So, if the host opens box B, what's the probability that the money is in box C?
- By Bayes' Rule,

$$P(M_C|H_B) = \frac{P(H_B|M_C)P(M_C)}{P(H_B)}$$

- We know that $P(M_C) = 1/3$, and $P(H_B|M_C) = 1$.
- By the law of total probability,

$$\begin{aligned} P(H_B) =& P(H_B|M_A)P(M_A) + P(H_B|M_B)P(M_B) + P(H_B|M_C)P(M_C) \\ =& \frac{1}{2} \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} \\ =& \frac{1}{2} \end{aligned}$$

The Monty Hall problem

- So,

$$P(M_C|H_B) = \frac{1/3 \times 1}{1/2} = 2/3$$

- In other words, given the **partial information** that the host has opened box B, the probability that the money is in box C is 2/3.

- So, we should switch!
- What if I don't specify your choice A and the host's choice B?