Data Science 1
Probability
Bayes' Theorem

Edward L. Boone

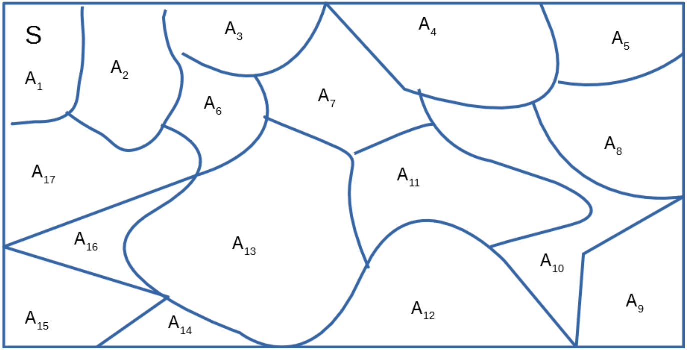
Introduction

Recall the following:

- E is an event in a sample space $\mathcal{S}$.
- $P(\mathcal{S}) = 1$.
- $E \cup F$ is the union of events E and F.
- $E \cap F$ is the intersection of events E and F.
- Probability behaves like area.
- $P(E|F) = \frac{P(E \cap F)}{P(F)}$
- If E and F are independent, then $P(E \cap F) = P(E)P(F)$.

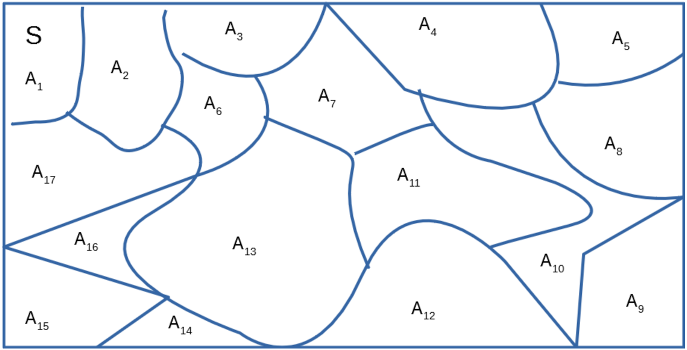
Partition

Suppose we have a **partition** of the sample space $\mathcal{S}$:



Partition

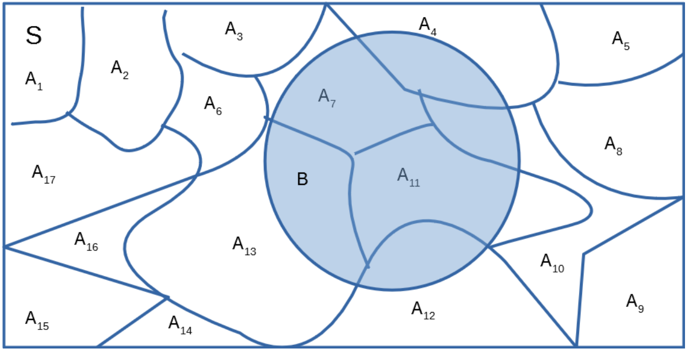
Suppose we have a **partition** of the sample space $\mathcal{S}$:



Here $A_i \cap A_j = \emptyset$ for all $i \neq j$ and $P(\mathcal{S}) = \sum_i P(A_i)$.

Total Probability

Suppose $A_1, A_2, \ldots, A_k$ form a partition of $\mathcal{S}$ and B is an event.



Then $P(B) = \sum_i P(B \cap A_i)$.

Total Probability

Suppose $A_1, A_2, ..., A_k$ form a partition of $\mathcal{S}$ and B is an event. Then:

$$P(B) = \sum_i P(B \cap A_i)$$

Total Probability

Suppose $A_1, A_2, ..., A_k$ form a partition of $\mathcal{S}$ and B is an event. Then:

$$P(B) = \sum_i P(B \cap A_i)$$

which can be rewritten using the rule $P(B \cap A_i) = P(B|A_i)P(A_i)$ as,

Total Probability

Suppose $A_1, A_2, ..., A_k$ form a partition of $\mathcal{S}$ and B is an event. Then:

$$P(B) = \sum_i P(B \cap A_i)$$

which can be rewritten using the rule $P(B \cap A_i) = P(B|A_i)P(A_i)$ as,

$$P(B) = \sum_i P(B|A_i)P(A_i)$$

Total Probability

Suppose $A_1, A_2, ..., A_k$ form a partition of $\mathcal{S}$ and B is an event. Then:

$$P(B) = \sum_i P(B \cap A_i)$$

which can be rewritten using the rule $P(B \cap A_i) = P(B|A_i)P(A_i)$ as,

$$P(B) = \sum_i P(B|A_i)P(A_i)$$

This is called the **Law of Total Probability**.

Example

Example: Suppose our market has three suppliers for mangoes, supplier A, supplier B and supplier C. Based on previous experience we know that $P(A) = 0.2$, $P(B) = 0.5$, $P(C) = 0.3$. Furthermore, from experience we know that rotten mangoes always happen but we noticed that supplier A has about 3% rotten, supplier B has about 4% rotten and supplier C has about 2% rotten. What is the probability that a customer picks up a rotten mango?

Example

Example: Suppose our market has three suppliers for mangoes, supplier A, supplier B and supplier C. Based on previous experience we know that $P(A) = 0.2$, $P(B) = 0.5$, $P(C) = 0.3$. Furthermore, from experience we know that rotten mangoes always happen but we noticed that supplier A has about 3% rotten, supplier B has about 4% rotten and supplier C has about 2% rotten. What is the probability that a customer picks up a rotten mango?

Write down what we know:

$$P(A) = 0.2,\ P(B) = 0.5,\ P(C) = 0.3$$

$$P(\text{rotten}|A) = 0.03,\ P(\text{rotten}|B) = 0.04,\ P(\text{rotten}|C) = 0.02.$$

Example

Example: What is the probability that a customer picks up a rotten mango?

Write down what we know:

$$P(A) = 0.2, \; P(B) = 0.5, \; P(C) = 0.3$$

$$P(\text{rotten}|A) = 0.03, \; P(\text{rotten}|B) = 0.04, \; P(\text{rotten}|C) = 0.02.$$

Then,

$$\begin{aligned} P(\text{rotten}) &= P(\text{rotten}|A)P(A) + P(\text{rotten}|B)P(B) + P(\text{rotten}|C)P(C) \\ &= 0.03 \times 0.2 + 0.04 \times 0.5 + 0.02 \times 0.3 \\ &= 0.032 \end{aligned}$$

This is not difficult to work out once you write down what you know.

Bayes' Theorem

Suppose $A_1, A_2, ..., A_k$ form a partition of $\mathcal{S}$ and B is an event. And we know $P(B|A_i)$ and $P(A_i)$ for all i. Then we can use this to "flip" the probability statement.

$$
\begin{aligned}
P(A_j|B) &= \frac{P(A_j \cap B)}{P(B)} \\
&= \frac{P(A_j \cap B)}{\sum_i P(B|A_i)P(A_i)} \\
&= \frac{P(B|A_j)P(A_j)}{\sum_i P(B|A_i)P(A_i)}
\end{aligned}
$$

This is known as Bayes' Theorem.

Additional Rule

Recall the following:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Which gives:

$$P(E|F)P(F) = P(E \cap F)$$

Additional Rules

Recall the following:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Which gives:

$$P(E|F)P(F) = P(E \cap F)$$

And if E and F are **independent** then $P(E|F) = P(E)$ which means:

$$P(E)P(F) = P(E \cap F)$$

Additional Rules

Recall the following:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Which gives:

$$P(E|F)P(F) = P(E \cap F)$$

And if E and F are **independent** then $P(E|F) = P(E)$ which means:

$$P(E)P(F) = P(E \cap F)$$

These rules are very helpful.

Example

Example: Suppose we know that probability of a specific disease in the general population is 0.01. Furthermore, we have a test that reads positive 99% of the time for people who have the disease and the test reads positive for 10% of the time for people who do not have the disease. If I am a physician I want to know what is what is the probability someone has the disease given they read positive on the test.

Example

Example: Suppose we know that probability of a specific disease in the general population is 0.01. Furthermore, we have a test that reads positive 99% of the time for people who have the disease and the test reads positive for 10% of the time for people who do not have the disease. If I am a physician I want to know what is what is the probability someone has the disease given they read positive on the test.

Write down what we know.

Example

Example: Suppose we know that probability of a specific disease in the general population is 0.01. Furthermore, we have a test that reads positive 99% of the time for people who have the disease and the test reads positive for 10% of the time for people who do not have the disease. If I am a physician I want to know what is what is the probability someone has the disease given they read positive on the test.

Write down what we know.

$$\begin{aligned} P(D) &= 0.01 \\ P(+|D) &= 0.99 \\ P(+|D^c) &= 0.10 \end{aligned}$$

Example

Example: Suppose we know that probability of a specific disease in the general population is 0.01. Furthermore, we have a test that reads positive 99% of the time for people who have the disease and the test reads positive for 10% of the time for people who do not have the disease. If I am a physician I want to know what is what is the probability someone has the disease given they read positive on the test.

Write down what we know.

$$
\begin{aligned}
P(D) &= 0.01 \\
P(+|D) &= 0.99 \\
P(+|D^c) &= 0.10 \\
P(D^c) &= 1 - P(D) = 1 - 0.01 = 0.99
\end{aligned}
$$

Example

Example: Suppose we know that probability of a specific disease in the general population is 0.01. Furthermore, we have a test that reads positive 99% of the time for people who have the disease and the test reads positive for 10% of the time for people who do not have the disease. If I am a physician I want to know what is what is the probability someone has the disease given they read positive on the test.

Write down what we know.

$$\begin{aligned} P(D) &= 0.01 \\ P(+|D) &= 0.99 \\ P(+|D^c) &= 0.10 \\ P(D^c) &= 1 - P(D) = 1 - 0.01 = 0.99 \end{aligned}$$

We want $P(D|+)$

Example

Example:

$$P(D) = 0.01,\ \ P(+|D) = 0.99,\ \ P(+|D^c) = 0.10 \text{ and } P(D^c) = 0.99$$

Put it all together...

$$\begin{aligned} P(D|+) &= \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)} \\ &= \frac{0.99 \times 0.01}{0.99 \times 0.01 + 0.1 \times 0.99} \\ &= \frac{0.0099}{0.0099 + 0.099} \\ &= 0.05 \end{aligned}$$

Example

Example:

$$P(D) = 0.01,\ \ P(+|D) = 0.99,\ \ P(+|D^c) = 0.10 \text{ and } P(D^c) = 0.99$$

Put it all together...

$$\begin{aligned} P(D|+) &= \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)} \\ &= \frac{0.99 \times 0.01}{0.99 \times 0.01 + 0.1 \times 0.99} \\ &= \frac{0.0099}{0.0099 + 0.099} \\ &= 0.05 \end{aligned}$$

This is an example of where probabilities are not near what you think they would be.

Summary

The law of Total Probability and Bayes' Theorem are incredibly important ideas in Probability and Statistics.

- Total Probability: $P(B) = \sum_i P(A_i \cap B)$
- Bayes' Theorem:

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_i P(B|A_i)P(A_i)}.$$

These ideas show up again and again in probability, statistics and data science.