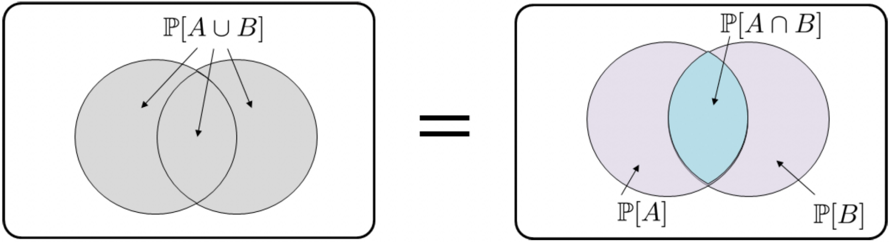
④ For any A and B

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$$



This statement is different from Axiom 3 because A and B are not necessarily disjoint.

Example (Fair coin tossing)

Let $\Omega = \{\epsdice{1}, \epsdice{2}, \epsdice{3}, \epsdice{4}, \epsdice{5}, \epsdice{6}\} = \{1, 2, 3, 4, 5, 6\}$ be the sample space of a fair die.

Let $A = \{\epsdice{1}, \epsdice{2}, \epsdice{3}\}$ and $B = \{\epsdice{3}, \epsdice{4}, \epsdice{5}\}$.

Then

$$P(A \cup B) = P(\{\epsdice{1}, \epsdice{2}, \epsdice{3}, \epsdice{4}, \epsdice{5}\}) = \frac{5}{6}$$

but also

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= P(\{\epsdice{1}, \epsdice{2}, \epsdice{3}\}) + P(\{\epsdice{3}, \epsdice{4}, \epsdice{5}\}) - P(\{\epsdice{3}\}) \\ &= \frac{3}{6} + \frac{3}{6} - \frac{1}{6} = \frac{5}{6} \end{aligned}$$

Properties of Probability

⑤ For any A and B

$$\mathbb{P}[A \cup B] \leq \mathbb{P}[A] + \mathbb{P}[B]$$

Properties of Probability

⑤ For any A and B

$$\mathbb{P}[A \cup B] \leq \mathbb{P}[A] + \mathbb{P}[B]$$

⑥ If $A \subseteq B$, then

$$\mathbb{P}[A] \leq \mathbb{P}[B]$$

Example

$A = \{t \leq 5\}$, and $B = \{t \leq 10\}$, then $\mathbb{P}[A] \leq \mathbb{P}[B]$

Probability & Statistics for DS & AI

Conditional Probability

Michele Guindani

Jio Institute

Summer 2022

Conditional Probability

- In many practical data science problems, we are interested in the relationship between two or more events.
- For example, an event A may make B more likely or less likely to happen, and B may make C more or less likely to happen.
- A legitimate question in probability is then: If A has happened, what is the probability that B also happens?
- Of course, if A and B are correlated events, then knowing one event can tell us something about the other event.
- If the two events have no relationship (they are "independent"), knowing one event will not tell us anything about the other.

Conditional Probability

Consider two events A and B. Assume $\mathbb{P}[B] \neq 0$. The conditional probability of A given B is

$$\mathbb{P}[A \mid B] \stackrel{\text{def}}{=} \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

- Intuitively, the conditional probability of A given B is the probability that A happens when we know that B has already happened. Since B has already happened, the event that A has also happened is represented by $A \cap B$. However, since we are only interested in the relative probability of A with respect to B
- The difference between $\mathbb{P}[A \mid B]$ and $\mathbb{P}[A \cap B]$ is the denominator they carry:

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} \text{ and } \mathbb{P}[A \cap B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[\Omega]}$$

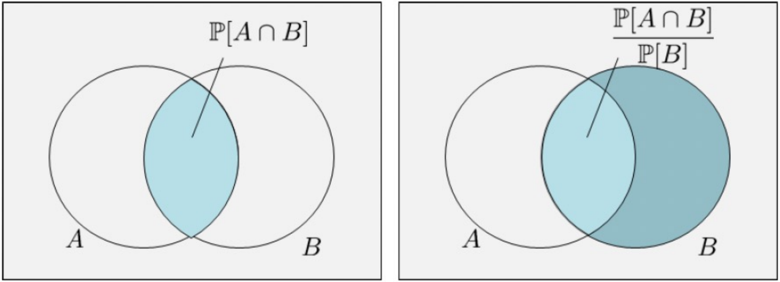


Figure: Illustration of conditional probability and its comparison with $\mathbb{P}[A \cap B]$.

Example (Throwing a die)

Consider throwing a die. Let

$$A = \{ \text{ getting a } 3\} \quad \text{and} \quad B = \{ \text{ getting an odd number } \}$$

The two probabilities are easy to compute:

$$P(A) = \frac{1}{6} \quad P(B) = \frac{3}{6}.$$

Example (Throwing a die)

Consider throwing a die. Let

$$A = \{\text{ getting a } 3\} \quad \text{and} \quad B = \{\text{ getting an odd number }\}$$

The two probabilities are easy to compute:

$$P(A) = \frac{1}{6} \quad P(B) = \frac{3}{6}.$$

The probability of the intersection is also easy

$$P(A \cap B) = \frac{1}{6}$$

Example (Throwing a die)

Consider throwing a die. Let

$$A = \{ \text{ getting a } 3\} \quad \text{and} \quad B = \{ \text{ getting an odd number } \}$$

The two probabilities are easy to compute:

$$P(A) = \frac{1}{6} \quad P(B) = \frac{3}{6}.$$

The probability of the intersection is also easy

$$P(A \cap B) = \frac{1}{6}$$

Then,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

And,

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{6}}{\frac{1}{6}} = 1$$

Independence

Independence

- Conditional probability deals with situations where two events A and B are related. What if the two events are unrelated? In probability, we have a technical term for this situation: statistical independence.

Independence

- Conditional probability deals with situations where two events A and B are related. What if the two events are unrelated? In probability, we have a technical term for this situation: statistical independence.

Example (Throwing a dice)

Throw a dice twice. Let

$$A = \{ \text{ 1st dice is 3} \} \text{ and } B = \{ \text{ 2nd dice is 4} \}$$

Independence

- Conditional probability deals with situations where two events A and B are related. What if the two events are unrelated? In probability, we have a technical term for this situation: statistical independence.

Example (Throwing a dice)

Throw a dice twice. Let

$$A = \{ \text{ 1st dice is } 3\} \text{ and } B = \{ \text{ 2nd dice is } 4\}$$

- What is independence?
- One event does not affect the other event!
- Are A and B independent then?

Independence

Two events A and B are statistically independent if

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]$$

Independence

Two events A and B are statistically independent if

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]$$

- Why define independence in this way?

Independence

Two events A and B are statistically independent if

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]$$

- Why define independence in this way? Recall that $\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$.

  If A and B are independent, then $\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$ and so

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[A]\,\mathbb{P}[B]}{\mathbb{P}[B]} = \mathbb{P}[A]$$

- Intuitively, if the occurrence of B provides no additional information about the occurrence of A, then A and B are independent.

Example (Throwing a dice)

Throw a dice twice. Let

$$A = \{ \text{ 1st dice is } 3\} \text{ and } B = \{ \text{ 2nd dice is } 4\}$$

Example (Throwing a dice)

Throw a dice twice. Let

$$A = \{ \text{ 1st dice is 3}\} \text{ and } B = \{ \text{ 2nd dice is 4}\}$$

Are A and B independent?

Example (Throwing a dice)

Throw a dice twice. Let

$$A = \{ \text{ 1st dice is 3}\} \text{ and } B = \{ \text{ 2nd dice is 4}\}$$

Are A and B independent?
Let's compute

$$P(A) = \frac{1}{6} \qquad P(B) = \frac{1}{6}$$

We can also compute

$$P(A \cap B) = P(\text{ 1st dice is 3 and 2nd dice is 4 out of 36 possible combinations}) =$$

Since

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

then the two events are statistically independent.

Example (Throwing a dice twice)

Let

$$A = \{1 \text{ st dice is } 1\} \quad \text{and } B = \{ \text{ sum is } 7\}$$

Are A and B independent?

Example (Throwing a dice twice)

Let

$$A = \{1 \text{ st dice is } 1\} \quad \text{and } B = \{ \text{ sum is } 7\}$$

Are A and B independent?

Let's compute

$$P(A) = \frac{1}{6} \quad P(B) = \frac{6}{36} = \frac{1}{6}$$

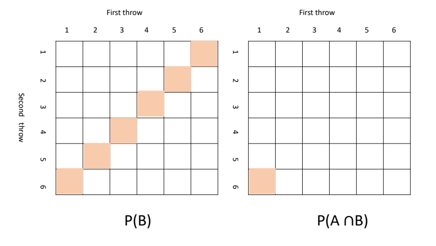
We can also check that

$$P(A \cap B) = \frac{1}{36}$$

So,

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

and the two events are independent.



Example (Throwing a dice twice)

Let

$$A = \{1 \text{ st dice is } 2\} \quad \text{and } B = \{ \text{ sum is } 8\}$$

Are A and B independent?

Example (Throwing a dice twice)

Let

$$A = \{1 \text{ st dice is } 2\} \quad \text{and } B = \{ \text{ sum is } 8\}$$

Are A and B independent?

Let's compute

$$P(A) = \frac{1}{6} \quad P(B) = \frac{5}{36}$$

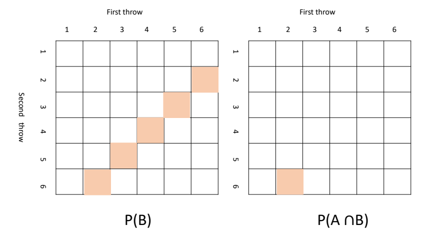
We can also check that

$$P(A \cap B) = \frac{1}{36}$$

So,

$$P(A \cap B) \neq P(A)\, P(B)$$

and the two events are NOT independent.



Example (Throwing a dice twice)

Let

$$A = \{1 \text{ st dice is } 3, 4, \text{ or } 5\} \quad \text{and } B = \{ \text{ sum is } 9\}$$

Are A and B independent?

Example (Throwing a dice twice)

Let

$$A = \{1 \text{ st dice is } 3, 4, \text{ or } 5\} \quad \text{and } B = \{ \text{ sum is } 9\}$$

Are A and B independent?

Let's compute

$$P(A) = \frac{3}{6} = \frac{1}{2} \quad P(B) = \frac{4}{36} = \frac{1}{9}$$

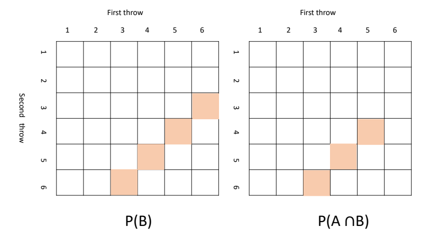
We can also check that

$$P(A \cap B) = \frac{3}{36} = \frac{1}{12}$$

So,

$$P(A \cap B) \neq P(A)\, P(B)$$

and the two events are NOT independent.



Bayes Theorem

- Sometimes, we may be interested in finding $P(A|B)$ but we only know $P(B|A)$ and the marginal probability $P(B)$

Bayes Theorem

- Sometimes, we may be interested in finding $P(A|B)$ but we only know $P(B|A)$ and the marginal probability $P(B)$

Example (Medical Testing)

- Let D indicate someone who has a disease and H a healthy subject
- Let $+$ indicate that someone tested *positive* for the disease in a screening test, and $-$ that they tested *negative*
- We are interested to know $P(D|+)$. But we may know:

Bayes Theorem

- Sometimes, we may be interested in finding $P(A|B)$ but we only know $P(B|A)$ and the marginal probability $P(B)$

Example (Medical Testing)

- Let D indicate someone who has a disease and H a healthy subject
- Let $+$ indicate that someone tested *positive* for the disease in a screening test, and $-$ that they tested *negative*
- We are interested to know $P(D|+)$. But we may know:

  Prevalence of disease in the population: $Pr(D) = 0.01$

Bayes Theorem

- Sometimes, we may be interested in finding $P(A|B)$ but we only know $P(B|A)$ and the marginal probability $P(B)$

Example (Medical Testing)

- Let D indicate someone who has a disease and H a healthy subject
- Let $+$ indicate that someone tested *positive* for the disease in a screening test, and $-$ that they tested *negative*
- We are interested to know $P(D|+)$. But we may know:

  Prevalence of disease in the population: $Pr(D) = 0.01$

  Sensitivity of the test: how good is the test at identifying people who have the disease, $Pr(+|D) = 0.98$

Bayes Theorem

- Sometimes, we may be interested in finding $P(A|B)$ but we only know $P(B|A)$ and the marginal probability $P(B)$

Example (Medical Testing)

- Let D indicate someone who has a disease and H a healthy subject
- Let $+$ indicate that someone tested *positive* for the disease in a screening test, and $-$ that they tested *negative*
- We are interested to know $P(D|+)$. But we may know:

  Prevalence of disease in the population: $Pr(D) = 0.01$

  Sensitivity of the test: how good is the test at identifying people who have the disease, $Pr(+|D) = 0.98$

  Specificity of the test: how good is the test at correctly identifying healthy individuals: $Pr(-|H) = 0.95$

Bayes Theorem

For any two events A and B such that $\mathbb{P}[A] > 0$ and $\mathbb{P}[B] > 0$,

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[B \mid A]\,\mathbb{P}[A]}{\mathbb{P}[B]}$$

Bayes Theorem

For any two events A and B such that $\mathbb{P}[A] > 0$ and $\mathbb{P}[B] > 0$,

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[B \mid A]\,\mathbb{P}[A]}{\mathbb{P}[B]}$$

- Proof By the definition of conditional probability:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad (*)$$

Bayes Theorem

For any two events A and B such that $\mathbb{P}[A] > 0$ and $\mathbb{P}[B] > 0$,

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[B \mid A]\,\mathbb{P}[A]}{\mathbb{P}[B]}$$

- Proof By the definition of conditional probability:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad (*)$$

But also

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

Bayes Theorem

For any two events A and B such that $\mathbb{P}[A] > 0$ and $\mathbb{P}[B] > 0$,

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[B \mid A]\,\mathbb{P}[A]}{\mathbb{P}[B]}$$

- Proof By the definition of conditional probability:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad (*)$$

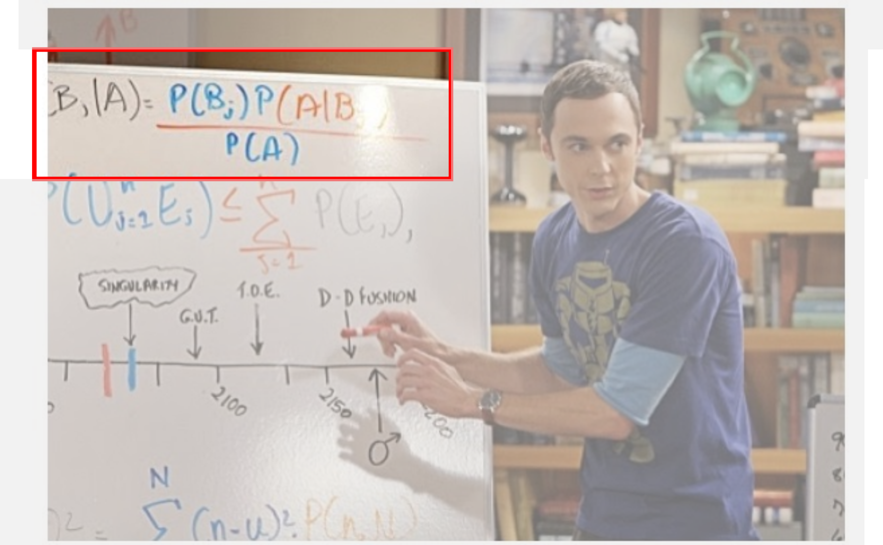
But also

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

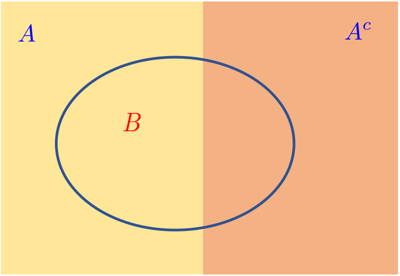
So, substitute in (*)

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

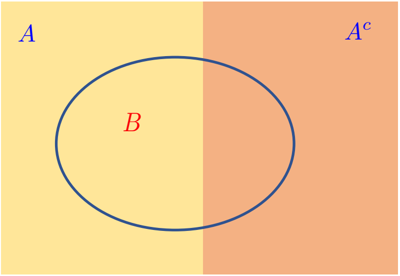
Of course, this is a very important Theorem...



- If we don't know the denominator, i.e. the marginal probability $P(B)$, we can compute it using the law of total probability, in this way:

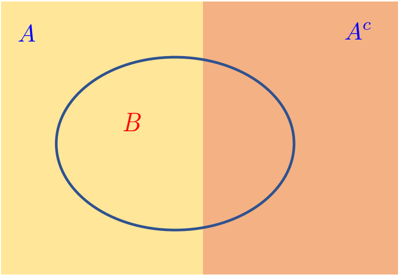


- If we don't know the denominator, i.e. the marginal probability $P(B)$, we can compute it using the law of total probability, in this way:



$$P(B) = P(B \cap A) + P(B \cap A^c)$$

- If we don't know the denominator, i.e. the marginal probability $P(B)$, we can compute it using the law of total probability, in this way:



$$P(B) = P(B \cap A) + P(B \cap A^c) = P(B|A) \times P(A) + P(B|A^c) \times P(A^c)$$

Let's go back to the medical testing example

Example (Medical Testing)

- Recall:
- D indicates someone who has a disease and H a healthy subject
- $+$ indicates that someone tested *positive* for the disease in a screening test, and $-$ that they tested *negative*
- We are interested to know $P(D|+)$. But we may know:

Let's go back to the medical testing example

Example (Medical Testing)

- Recall:
- D indicates someone who has a disease and H a healthy subject
- $+$ indicates that someone tested *positive* for the disease in a screening test, and $-$ that they tested *negative*
- We are interested to know $P(D|+)$. But we may know:

  Prevalence of disease in the population: $Pr(D) = 0.01$

Let's go back to the medical testing example

Example (Medical Testing)

- Recall:
- D indicates someone who has a disease and H a healthy subject
- $+$ indicates that someone tested *positive* for the disease in a screening test, and $-$ that they tested *negative*
- We are interested to know $P(D|+)$. But we may know:

  Prevalence of disease in the population: $Pr(D) = 0.01$

  Sensitivity of the test: how good is the test at identifying people who have the disease, $Pr(+|D) = 0.98$

Let's go back to the medical testing example

Example (Medical Testing)

- Recall:
- D indicates someone who has a disease and H a healthy subject
- $+$ indicates that someone tested *positive* for the disease in a screening test, and $-$ that they tested *negative*
- We are interested to know $P(D|+)$. But we may know:

  Prevalence of disease in the population: $Pr(D) = 0.01$

  Sensitivity of the test: how good is the test at identifying people who have the disease, $Pr(+|D) = 0.98$

  Specificity of the test: how good is the test at correctly identifying healthy individuals: $Pr(-|H) = 0.95$

Example (Medical Testing)

$$
\begin{aligned}
Pr(D|+) &= \frac{Pr(+|D)Pr(D)}{Pr(+)} \\
&= \frac{Pr(+|D)Pr(D)}{Pr(+|D)Pr(D) + Pr(+|H)Pr(H)}
\end{aligned}
$$

Example (Medical Testing)

$$
\begin{aligned}
Pr(D|+) &= \frac{Pr(+|D)Pr(D)}{Pr(+)} \\
&= \frac{Pr(+|D)Pr(D)}{Pr(+|D)Pr(D) + Pr(+|H)Pr(H)}
\end{aligned}
$$

Example (Medical Testing)

$$
\begin{aligned}
Pr(D|+) &= \frac{Pr(+|D)Pr(D)}{Pr(+)} \\
&= \frac{Pr(+|D)Pr(D)}{Pr(+|D)Pr(D) + Pr(+|H)Pr(H)} \\
&= \frac{0.98 \times 0.01}{[0.98 \times 0.01] + [(1-0.95) \times (1-0.01)]}
\end{aligned}
$$

Example (Medical Testing)

$$
\begin{aligned}
Pr(D|+) &= \frac{Pr(+|D)Pr(D)}{Pr(+)} \\
&= \frac{Pr(+|D)Pr(D)}{Pr(+|D)Pr(D) + Pr(+|H)Pr(H)} \\
&= \frac{0.98 \times 0.01}{[0.98 \times 0.01] + [(1-0.95) \times (1-0.01)]} \\
&= \frac{0.98 \times 0.01}{[0.98 \times 0.01] + [0.05 \times 0.99]}
\end{aligned}
$$

Example (Medical Testing)

$$
\begin{aligned}
Pr(D|+) &= \frac{Pr(+|D)Pr(D)}{Pr(+)} \\
&= \frac{Pr(+|D)Pr(D)}{Pr(+|D)Pr(D) + Pr(+|H)Pr(H)} \\
&= \frac{0.98 \times 0.01}{[0.98 \times 0.01] + [(1 - 0.95) \times (1 - 0.01)]} \\
&= \frac{0.98 \times 0.01}{[0.98 \times 0.01] + [0.05 \times 0.99]} \\
&= \frac{0.0098}{0.0098 + 0.0495} \\
&= 0.165
\end{aligned}
$$

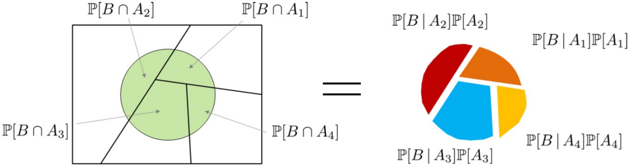
Example (Medical Testing)

$$
\begin{aligned}
Pr(D|+) &= \frac{Pr(+|D)Pr(D)}{Pr(+)} \\
&= \frac{Pr(+|D)Pr(D)}{Pr(+|D)Pr(D) + Pr(+|H)Pr(H)} \\
&= \frac{0.98 \times 0.01}{[0.98 \times 0.01] + [(1-0.95) \times (1-0.01)]} \\
&= \frac{0.98 \times 0.01}{[0.98 \times 0.01] + [0.05 \times 0.99]} \\
&= \frac{0.0098}{0.0098 + 0.0495} \\
&= 0.165
\end{aligned}
$$

- Even after testing positive, there's still a 83.5% chance that the person is healthy!
- Notice the effect of the prevalence: if $Pr(D) = 0.5$, then $Pr(D|+) = 0.95$!!!

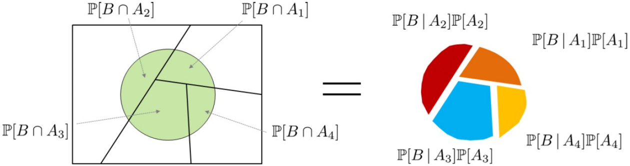
Law of total probability

- A more general version of the law of total probability is the following (on partition of a set B)



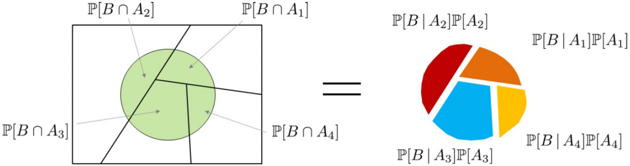
Law of total probability

- A more general version of the law of total probability is the following (on partition of a set B)



Law of total probability

- A more general version of the law of total probability is the following (on partition of a set B)



Theorem

Let $\{A_1, \ldots, A_n\}$ be a partition of Ω, i.e., $A_1, \ldots, A_n$ are disjoint and $\Omega = A_1 \cup \cdots \cup A_n$. Then, for any $B \subseteq \Omega$

$$\mathbb{P}[B] = \sum_{i=1}^{n} \mathbb{P}[B \mid A_i] \mathbb{P}[A_i]$$

Example (Tennis tournament)

Suppose your probability of winning the game is

- 0.3 against $\frac{1}{2}$ of the players (Event A).
- 0.4 against $\frac{1}{4}$ of the players (Event B).
- 0.5 against $\frac{1}{4}$ of the players (Event C).

1. If you play a match in this tournament, what is the probability of your winning the match?

Example (Tennis tournament)

Suppose your probability of winning the game is

- 0.3 against $\frac{1}{2}$ of the players (Event A).
- 0.4 against $\frac{1}{4}$ of the players (Event B).
- 0.5 against $\frac{1}{4}$ of the players (Event C).

1. If you play a match in this tournament, what is the probability of your winning the match?

   Let's see what we know already. We know that:

$$\mathbb{P}[A] = 0.5, \quad \mathbb{P}[B] = 0.25, \quad \mathbb{P}[C] = 0.25$$

Example (Tennis tournament)

Suppose your probability of winning the game is

- 0.3 against $\frac{1}{2}$ of the players (Event A).
- 0.4 against $\frac{1}{4}$ of the players (Event B).
- 0.5 against $\frac{1}{4}$ of the players (Event C).

① If you play a match in this tournament, what is the probability of your winning the match?

Let's see what we know already. We know that:

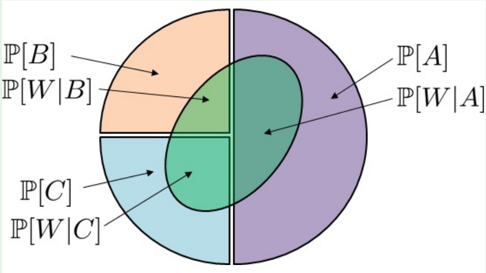
$$\mathbb{P}[A] = 0.5, \quad \mathbb{P}[B] = 0.25, \quad \mathbb{P}[C] = 0.25$$

and we know that

$$\mathbb{P}[W \mid A] = 0.3, \quad \mathbb{P}[W \mid B] = 0.4, \quad \mathbb{P}[W \mid C] = 0.5$$

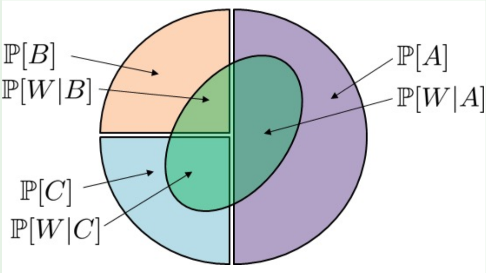
Example (Tennis tournament)

We can apply the law of total probability:



Example (Tennis tournament)

We can apply the law of total probability:



$$
\begin{aligned}
\mathbb{P}[W] &= \mathbb{P}[W \mid A]\mathbb{P}[A] + \mathbb{P}[W \mid B]\mathbb{P}[B] + \mathbb{P}[W \mid C]\mathbb{P}[C] \\
&= (0.3)(0.5) + (0.4)(0.25) + (0.5)(0.25) = 0.375
\end{aligned}
$$

Example (Tennis tournament - ctd)

2. Supposing that you have won a match, what is the probability that you played against an A player?

Example (Tennis tournament - ctd)

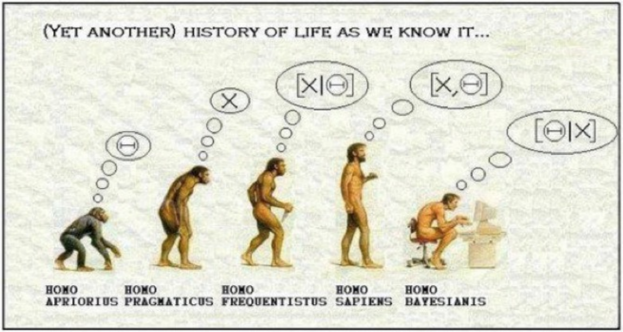
2. Supposing that you have won a match, what is the probability that you played against an A player?
   We can use the Bayes theorem to reply to this question. Given that you have won the match, the probability of A given W is

$$\mathbb{P}[A \mid W] = \frac{\mathbb{P}[W \mid A]\mathbb{P}[A]}{\mathbb{P}[W]} = \frac{(0.3)(0.5)}{0.375} = 0.4$$

Bayesian Analysis!

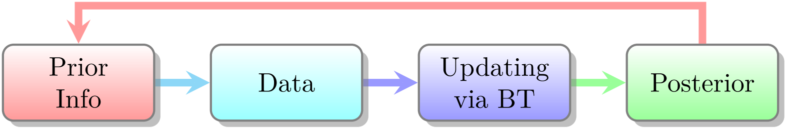
Bayesian Analysis!

- The Bayes Theorem is at the basis of a specific field of Statistics: Bayesian Analysis!

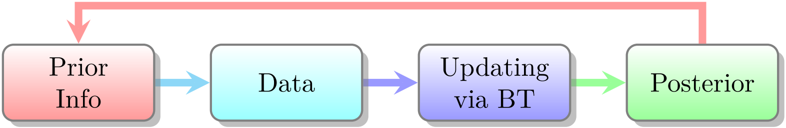


- The next evolution of Statistics! 🤪 😆

The tenets of Bayesian analysis

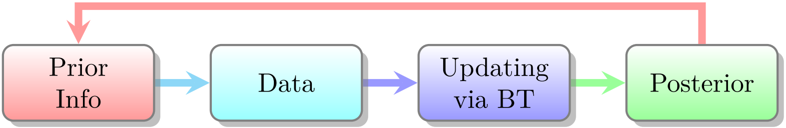


The tenets of Bayesian analysis



- Bayesian statistics starts by using (prior) probabilities to describe your current state of knowledge, say $P(\text{(available) info})$
- It then incorporates information through the collection of data, say $P(\text{data} \mid \text{(available) info})$

The tenets of Bayesian analysis



- Bayesian statistics starts by using (prior) probabilities to describe your current state of knowledge, say $P(\text{(available) info})$
- It then incorporates information through the collection of data, say $P(\text{data} \mid \text{(available) info})$
- By combining the prior probabilities with the data, you can obtain new (posterior) probabilities to describe an updated state of knowledge:

$$P(\text{(updated) info}|\text{data}) = \frac{P(\text{data} \mid \text{(available) info}) \times P(\text{(available) info})}{P(\text{data})}.$$