# STAT 339
# Probability, Data and Hypotheses

February 21, 2022

Colin Reimer Dawson

# Questions/Administrative Business?

# Outline

# Outline

Probability Spaces

Conditional Probabilities

Bayesian Updating

## Sample Space

A **sample space**, $\Omega$, is

1. **Classical/objectivist/frequentist** defintion: a collection of possible **outcomes** of a **random experiment** (The coin **will** come up heads or tails. Team A **will** win or lose. The train **will** arrive in $x$ minutes (for each value of $x$))
2. **Bayesian/subjectivist** definition: a collection of "possible worlds" that we might be in (The coin is fair and the next flip will be heads. The interval between trains is 30 minutes, the last one was 20 minutes ago, and the next one is operating normally.)

## Event

An **event** is a **subset** of the sample space that does or does not contain a particular outcome/possible world.

- The coin comes up heads.
- The coin is fair.
- The last train arrived 20 minutes ago.
- The next train is operating normally.
- The next train will arrive in 10 minutes.

Think of them as **statements** that are either **true** (if one of their elements is what happens/is true) or **false** (if what happens/is true is not part of the event), but we may not know which

# Three Kinds of Events

1. Some events describe **observations** (or **data**)
   - **Examples:** the coin comes up heads, the next train will arrive in 10 minutes
   - Both Bayesian and frequentist probability use these

# Three Kinds of Events

.1 Some events describe **observations** (or **data**)

- **Examples:** the coin comes up heads, the next train will arrive in 10 minutes
- Both Bayesian and frequentist probability use these

.2 Other events describe **hidden world states** (or **hypotheses**) (whose trueth value may *never* be observable directly, or may be known to some observers but not others)

- **Examples:** The coin is fair, the last train arrived 10 minutes ago, the next train is operating normally
- In frequentist probability, each defines a **separate** sample space. In Bayesian probability, they are subsets of **one**

# Three Kinds of Events

1. Some events describe **observations** (or **data**)
   - **Examples:** the coin comes up heads, the next train will arrive in 10 minutes
   - Both Bayesian and frequentist probability use these
2. Other events describe **hidden world states** (or **hypotheses**) (whose trueth value may *never* be observable directly, or may be known to some observers but not others)
   - **Examples:** The coin is fair, the last train arrived 10 minutes ago, the next train is operating normally
   - In frequentist probability, each defines a **separate** sample space. In Bayesian probability, they are subsets of **one**
3. Still others describe **conjunctions** or **intersections** of data and hypotheses
   - **Examples:** The coin is fair **and** the next flip will be heads, the next train is operating normally **and** it will arrive in 10 minutes

## Probability Space

A **probability space** is a sample space in which every event has been assigned a **probability**, in an internally consistent manner

## Probability Space

A **probability space** is a sample space in which every event has been assigned a **probability**, in an internally consistent manner

## Probability Axioms

1. Probabilities are nonnegative real numbers
2. The entire sample space has probability 1
3. If two events $A$ and $B$ **can't both be true** (they don't share any elements), then their disjunction (union) ($A$ OR $B$, aka $A \cup B$, the event in which at least one is true/happens) is the **sum of their individual probabilities**: $P(A \text{ OR } B) = P(A) + P(B)$

## Probability Axioms

1. Probabilities are nonnegative real numbers
2. The entire sample space has probability 1
3. If two events $A$ and $B$ **can't both be true** (they don't share any elements), then their disjunction (union) ($A$ OR $B$, aka $A \cup B$, the event in which at least one is true/happens) is the **sum of their individual probabilities**: $P(A \text{ OR } B) = P(A) + P(B)$

## Important Consequences

1. The empty event (the set with no elements) has probability 0
2. For every event $A$, $P(\text{ NOT } A) = 1 - P(A)$
3. For every pair of events $A$ and $B$, $P(A \text{ OR } B) = P(A) + P(B) - P(A \text{ AND } B)$.

# Outline

# Restricting the Sample Space

Often we focus our attention on a relevant **subset** of the sample space. For example

- We want to consider the properties of a possible world corresponding to a particular **hypothesis**
- We obtain new **data** which lets us **rule out** parts of the sample space that **contradict** that data

# Restricting the Sample Space

Often we focus our attention on a relevant **subset** of the sample space. For example

- We want to consider the properties of a possible world corresponding to a particular **hypothesis**
- We obtain new **data** which lets us **rule out** parts of the sample space that **contradict** that data

When we restrict our attention to a particular event/statement, $A$ (which in Bayesian probability could be an observation or a hypothesis):

1. The previous sample space is replaced by $A$. That is, $\Omega_{\text{new}} \leftarrow A$
2. The probabilities previously assigned to events $\Omega$ need to be **updated** to this new sample space

# Restricting the Sample Space

Often we focus our attention on a relevant **subset** of the sample space. For example

- We want to consider the properties of a possible world corresponding to a particular **hypothesis**
- We obtain new **data** which lets us **rule out** parts of the sample space that **contradict** that data

When we restrict our attention to a particular event/statement, $A$ (which in Bayesian probability could be an observation or a hypothesis):

1. The previous sample space is replaced by $A$. That is, $\Omega_{\text{new}} \leftarrow A$
2. The probabilities previously assigned to events $\Omega$ need to be **updated** to this new sample space

The probabilities assigned to the new restricted space are called **conditional probabilities**

# Conditional Probability

## Example

$\Omega$ = Instances where a coin is flipped

- $H$ = the coin is fair
- $D$ = the coin comes up heads

- To condition on $H$ restricts our attention to instances of flipping a fair coin
- To condition on $D$ restricts our attention to instances where any kind of coin comes up heads

# Conditional Probability

## Example

$\Omega$ = outcomes of a robot observing color of an object

- $H$ = the true color is "blue"
- $D$ = the sensor reports "blue"
- To condition on $H$ restricts our attention to instances of observing a genuinely blue object
- To condition on $D$ restricts our attention to instances in which the sensor says "blue"

# Conditional Probability

- We write $P(B|A)$ to mean "the probability that $B$ occurs/is true, in the **context of the restricted set of worlds** where $A$ occurs/is true"
- Also known as the **conditional probability** of $B$ **given** $A$
- Same notation and meaning whether $A$ and $B$ are data events or hypotheses

# Example: Marbles

A bag of marbles contains 4 marbles; some black and some white. But we don't know how many of each. We get to draw three marbles, one at a time, replacing the marble between each draw, writing down the color of each draw[1].

[1] Example and images from Richard McElreath: *Statistical Rethinking*

# Example: Marbles

A bag of marbles contains 4 marbles; some black and some white. But we don't know how many of each. We get to draw three marbles, one at a time, replacing the marble between each draw, writing down the color of each draw[1].

Relevant **hypotheses**

(1) [○○○○], (2) [●○○○], (3) [●●○○], (4) [●●●○], (5) [●●●●].

[1] Example and images from Richard McElreath: *Statistical Rethinking*

# Example: Marbles

A bag of marbles contains 4 marbles; some black and some white. But we don't know how many of each. We get to draw three marbles, one at a time, replacing the marble between each draw, writing down the color of each draw[1].

Relevant **hypotheses**

(1) [○○○○], (2) [●○○○], (3) [●●○○], (4) [●●●○], (5) [●●●●].

**If** the bag contains one black marble (Hypothesis 2), what are the possible observations involving three draws (with replacement)?

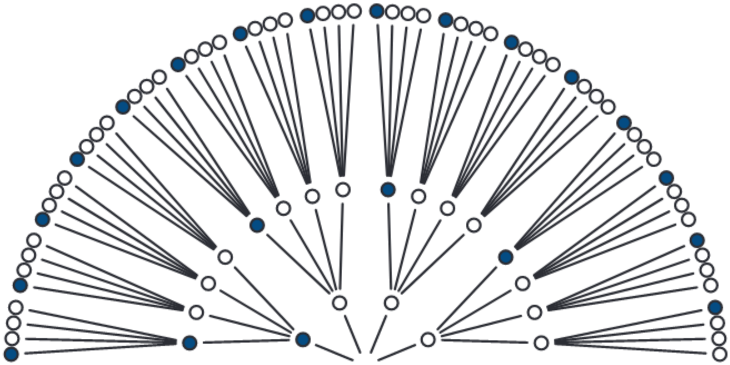
[1] Example and images from Richard McElreath: *Statistical Rethinking*

# Marbles, Continued

We can represent the space of possible draws as a **tree**, where each potential observation is a **path**. After two draws, the tree looks like:

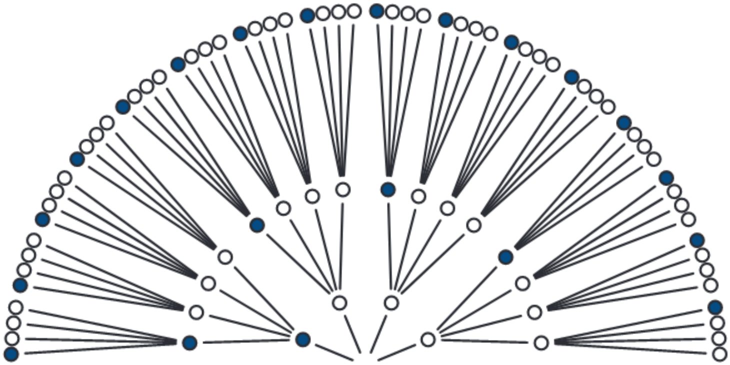
# Marbles, Continued

After the third draw, the tree expands:



# Marbles, Continued

After the third draw, the tree expands:



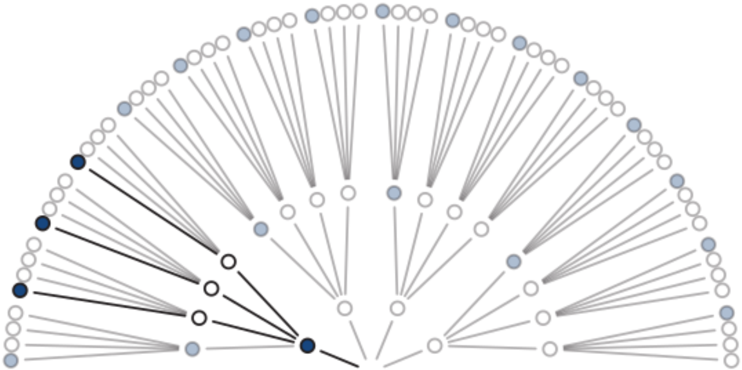
So, **within this context** of the **hypothesis** ($H_2$, say) that the bag has one black and three white marbles, there are **64 distinguishable observations possible**.

# Marbles, Continued

If we observe the data $D$ : (●○●)...

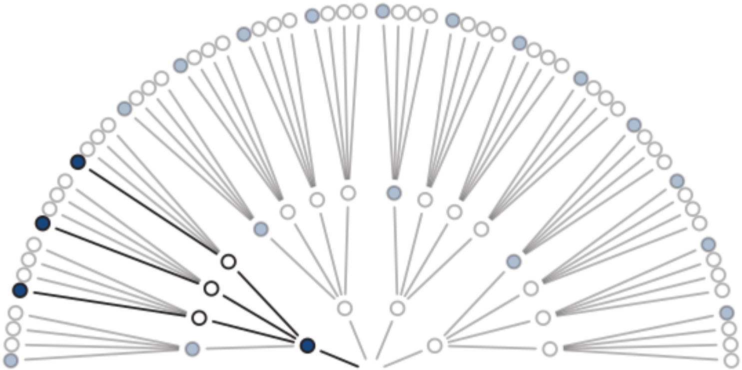
# Marbles, Continued

If we observe the data $D$ : (●○●)... then we've eliminated all but three of these 64



# Marbles, Continued

If we observe the data $D$ : (●○●)... then we've eliminated all but three of these 64



In other words, the event $H_2$ AND $D$ contains 3/64 of the "atomic" possibilities in $H_2$

# Conditional Proportion

- If we think each of these 64 had equal weight going in, then ($H_2$ AND $D$) commands 3/64 of the probability in $H_2$

# Conditional Proportion

- If we think each of these 64 had equal weight going in, then ($H_2$ AND $D$) commands 3/64 of the probability in $H_2$
- **Conditioned on** (in the context of) $H_2$, $H_2$ is the whole sample space, and therefore contains total probability 1

# Conditional Proportion

- If we think each of these 64 had equal weight going in, then ($H_2$ AND $D$) commands 3/64 of the probability in $H_2$
- **Conditioned on** (in the context of) $H_2$, $H_2$ is the whole sample space, and therefore contains total probability 1
- So, the **conditional probability** of $D$ **given** $H_2$ is

$$P(D \mid H_2) = \frac{\#\text{ paths in } (H_2 \text{ AND } D)}{\#\text{ total paths in } H_2} = \frac{3}{64}$$

# Other Hypotheses

Of course, $H_2$ is only one possible description of the bag. Remember that there were five hypotheses to begin with:

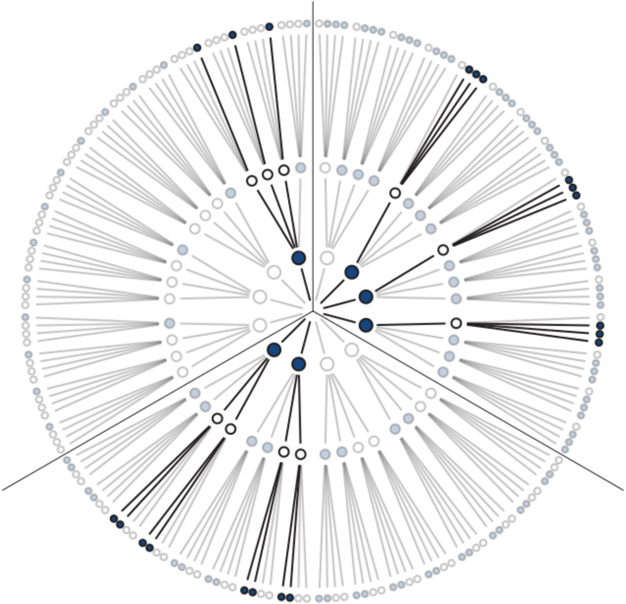
(1) [○○○○], (2) [●○○○], (3) [●●○○], (4) [●●●○], (5) [●●●●].

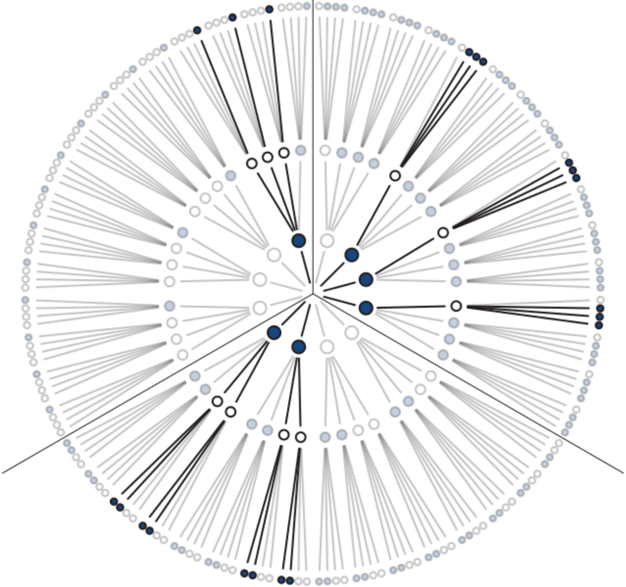
# Other Hypotheses

Of course, $H_2$ is only one possible description of the bag. Remember that there were five hypotheses to begin with:

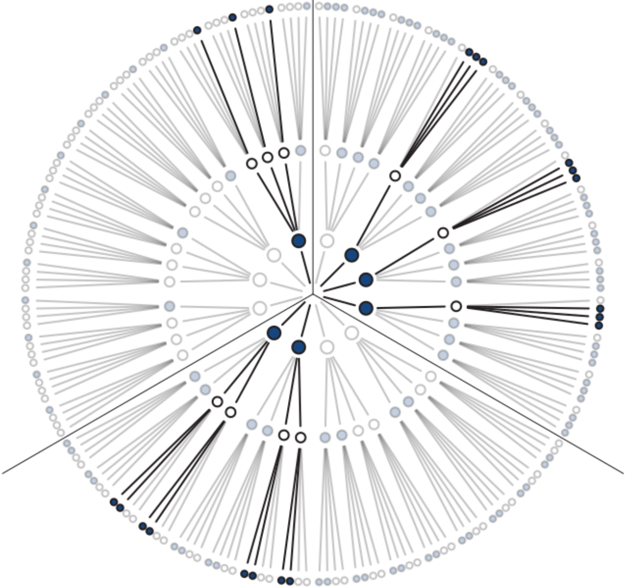
(1) [○○○○], (2) [●○○○], (3) [●●○○], (4) [●●●○], (5) [●●●●].

We can see that two of these (1 and 5) have **no paths** that could produce $D$ : (●○●). But if we sketch the trees for the other two...

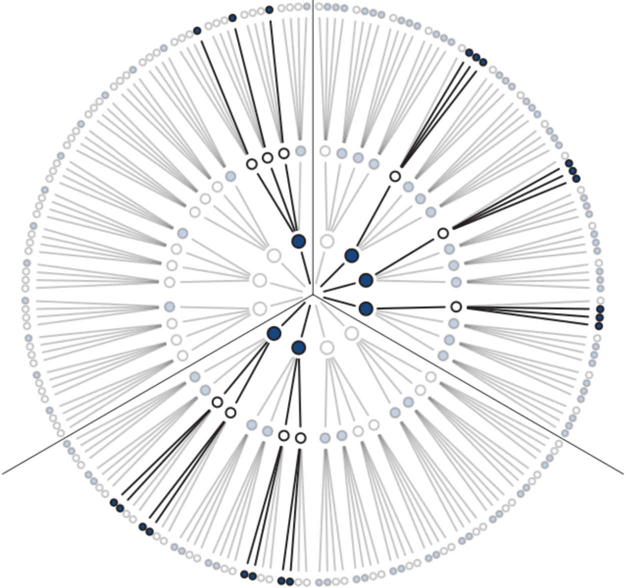




- The highlighted paths constitute **all the ways** we could have gotten $D$. That is, they constitute $D$ **itself** as an **event**



- Notice that ($D$ AND $H_3$) has 3 paths, ($D$ AND $H_3$) has 8 paths, while ($D$ AND $H_4$) has 9



- Intuitively, what should $P(H_2 \mid D)$ be?

# Summarizing the Path Counts

| Conjecture | Ways to produce ●○● |
|---|---|
| [○○○○] | $0 \times 4 \times 0 = 0$ |
| [●○○○] | $1 \times 3 \times 1 = 3$ |
| [●●○○] | $2 \times 2 \times 2 = 8$ |
| [●●●○] | $3 \times 1 \times 3 = 9$ |
| [●●●●] | $4 \times 0 \times 4 = 0$ |

- There is a total of $3 + 8 + 9 = 20$ paths in the "tree world" that give us the observation sequence $D$ : (●○●)

# Summarizing the Path Counts

| Conjecture | Ways to produce ●○● |
|---|---|
| [○○○○] | $0 \times 4 \times 0 = 0$ |
| [●○○○] | $1 \times 3 \times 1 = 3$ |
| [●●○○] | $2 \times 2 \times 2 = 8$ |
| [●●●○] | $3 \times 1 \times 3 = 9$ |
| [●●●●] | $4 \times 0 \times 4 = 0$ |

- There is a total of $3 + 8 + 9 = 20$ paths in the "tree world" that give us the observation sequence $D$ : (●○●)
- Of these, $3$ come from $H_2$

# Summarizing the Path Counts

| Conjecture | Ways to produce ●○● |
|---|---|
| [○○○○] | $0 \times 4 \times 0 = 0$ |
| [●○○○] | $1 \times 3 \times 1 = 3$ |
| [●●○○] | $2 \times 2 \times 2 = 8$ |
| [●●●○] | $3 \times 1 \times 3 = 9$ |
| [●●●●] | $4 \times 0 \times 4 = 0$ |

- There is a total of $3 + 8 + 9 = 20$ paths in the "tree world" that give us the observation sequence $D$ : (●○●)
- Of these, $3$ come from $H_2$
- So, **provided we weigh all 20 paths equally**, we get

$$P(H_2 \mid D) = \frac{\#\text{ paths in } (H_2 \text{ AND } D)}{\#\text{ total paths in } D} = \frac{3}{20}$$

# Relationship Between Conditional Probabilities

Notice the similarity

$$P(D \mid H_2) = \frac{\#\text{ paths in } (H_2 \text{ AND } D)}{\#\text{ total paths in } H_2}$$

$$P(H_2 \mid D) = \frac{\#\text{ paths in } (H_2 \text{ AND } D)}{\#\text{ total paths in } D}$$

- The two conditional probabilities have the same **numerator**, but a different **denominator**

# Relationship Between Conditional Probabilities

Notice the similarity

$$P(D \mid H_2) = \frac{\#\text{ paths in }(H_2\text{ AND }D)}{\#\text{ total paths in }H_2}$$

$$P(H_2 \mid D) = \frac{\#\text{ paths in }(H_2\text{ AND }D)}{\#\text{ total paths in }D}$$

- The two conditional probabilities have the same **numerator**, but a different **denominator**
- This suggests a way to translate between one and the other:

$$P(H_2 \mid D) = P(D \mid H_2) \times \frac{\#\text{ total paths in }H_2}{\#\text{ total paths in }D}$$

# From Proportion to Probability

- Note that we can divide top and bottom by total paths in the **original** sample space $\Omega$ to get

$$
\begin{aligned}
P(H_2 \mid D) &= \frac{\#\text{ paths in } (H_2 \text{ AND } D)/\#\text{ paths in } \Omega}{\#\text{ total paths in } D/\#\text{ paths in } \Omega} \\
&= \frac{\text{Share of } \Omega \text{ which is in } (H_2 \text{ AND } D)}{\text{Share of } \Omega \text{ which is in } D}
\end{aligned}
$$

# From Proportion to Probability

- Note that we can divide top and bottom by total paths in the **original** sample space $\Omega$ to get

$$P(H_2 \mid D) = \frac{\#\text{ paths in }(H_2\text{ AND }D)/\#\text{ paths in }\Omega}{\#\text{ total paths in }D/\#\text{ paths in }\Omega}$$
$$= \frac{\text{Share of }\Omega\text{ which is in }(H_2\text{ AND }D)}{\text{Share of }\Omega\text{ which is in }D}$$

- If we didn't weigh the paths all equally to start with, replace the proportions with "share of weight" and the same result follows

# New Observations

- What if we observe a fourth marble, and see that it's black?

# New Observations

- What if we observe a fourth marble, and see that it's black?
- We could start from scratch and sketch tree world with four-marble paths

# New Observations

- What if we observe a fourth marble, and see that it's black?
- We could start from scratch and sketch tree world with four-marble paths
- But the world has "shrunk" to 20 paths from the 192 possible for three observations – no point in sketching paths that have already been eliminated

# New Observations

- What if we observe a fourth marble, and see that it's black?
- We could start from scratch and sketch tree world with four-marble paths
- But the world has "shrunk" to 20 paths from the 192 possible for three observations – no point in sketching paths that have already been eliminated
- Instead, we can simply note that each path in $H_2$ has one way to produce a black marble next. Similarly, each path in $H_3$ has two ways, and each path in $H_4$ has three ways

# New Observations

| Conjecture | Ways to produce ● | Prior counts | New count |
|---|---|---|---|
| [○○○○] | 0 | 0 | $0 \times 0 = 0$ |
| [●○○○] | 1 | 3 | $3 \times 1 = 3$ |
| [●●○○] | 2 | 8 | $8 \times 2 = 16$ |
| [●●●○] | 3 | 9 | $9 \times 3 = 27$ |
| [●●●●] | 4 | 0 | $0 \times 4 = 0$ |

# New Observations

| Conjecture | Ways to produce ● | Prior counts | New count |
|---|---|---|---|
| [○○○○] | 0 | 0 | $0 \times 0 = 0$ |
| [●○○○] | 1 | 3 | $3 \times 1 = 3$ |
| [●●○○] | 2 | 8 | $8 \times 2 = 16$ |
| [●●●○] | 3 | 9 | $9 \times 3 = 27$ |
| [●●●●] | 4 | 0 | $0 \times 4 = 0$ |

So now,

$$P(H_2 \mid D) = 3/46$$
$$P(H_3 \mid D) = 16/46$$
$$P(H_4 \mid D) = 27/46$$

# From Proportion to Probability

- So, in general, since $\Omega$ has total weight (probability) 1, we can write the "share of $\Omega$" as the original probability

$$P(H_2 \mid D) = \frac{P(H_2 \text{ AND } D)}{P(D)}$$

# From Proportion to Probability

- So, in general, since $\Omega$ has total weight (probability) 1, we can write the "share of $\Omega$" as the original probability

$$P(H_2 \mid D) = \frac{P(H_2 \text{ AND } D)}{P(D)}$$

- The same works for $P(D|H)$ which we found as

$$\begin{aligned} P(D \mid H_2) &= \frac{\#\text{ paths in } (H_2 \text{ AND } D)}{\#\text{ total paths in } H_2} \\ &= \frac{P(H_2 \text{ AND } D)}{P(H_2)} \end{aligned}$$

# Conditional Probability: Formal Definition

## Conditional Probability

For any two events $A$ and $B$ (provided $P(A) > 0$), **conditional probability** of $B$ *given* $A$ is

$$P(B \mid A) = \frac{P(A \text{ AND } B)}{P(A)} \tag{1}$$

# Relationship Between Conditional Probabilities

Revisiting the relationship we had before using probabilities:

$$P(D \mid H_2) = \frac{P(H_2 \text{ AND } D)}{P(H_2)}$$

$$P(H_2 \mid D) = \frac{P(H_2 \text{ AND } D)}{P(D)}$$

- The two conditional probabilities again have the same **numerator**, but a different **denominator**

# Relationship Between Conditional Probabilities

Revisiting the relationship we had before using probabilities:

$$P(D \mid H_2) = \frac{P(H_2 \text{ AND } D)}{P(H_2)}$$

$$P(H_2 \mid D) = \frac{P(H_2 \text{ AND } D)}{P(D)}$$

- The two conditional probabilities again have the same **numerator**, but a different **denominator**
- We can translate from $P(D \mid H_2)$ to $P(H_2 \mid D)$ as:

$$P(H_2 \mid D) = P(D \mid H_2) \times \frac{P(H_2)}{P(D)}$$

# Relationship Between Conditional Probabilities

Revisiting the relationship we had before using probabilities:

$$P(D \mid H_2) = \frac{P(H_2 \text{ AND } D)}{P(H_2)}$$

$$P(H_2 \mid D) = \frac{P(H_2 \text{ AND } D)}{P(D)}$$

- The two conditional probabilities again have the same **numerator**, but a different **denominator**
- We can translate from $P(D \mid H_2)$ to $P(H_2 \mid D)$ as:

$$P(H_2 \mid D) = P(D \mid H_2) \times \frac{P(H_2)}{P(D)}$$

- This relationship is **Bayes' Theorem**

# Bayes Theorem

## Bayes' Theorem

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

- By day, a perfectly ordinary, mild-mannered consequence of the definition of conditional probability

# Bayes Theorem

**Bayes' Theorem**

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

- By day, a perfectly ordinary, mild-mannered consequence of the definition of conditional probability
- But by night, equipped with a **Bayesian Augmented Sample Space**, it becomes...

# Bayes Theorem

## Bayes' Theorem


- The **most powerful epistemic tool ever devised!**

# Outline

# Bayes' Theorem in Bayesian Statistics

**Bayes' Theorem**

$$P(H \mid D) = P(H)\frac{P(D \mid H)}{P(D)}$$

$P(H)$ The **prior probability** of the hypothesis $H$.
How plausible was it **before** observing D?

# Bayes' Theorem in Bayesian Statistics

**Bayes' Theorem**

$$P(H \mid D) = P(H)\frac{P(D \mid H)}{P(D)}$$

| | |
|---|---|
| $P(H)$ | The **prior probability** of the hypothesis $H$. How plausible was it **before** observing D? |
| $P(D \mid H)$ | Called the **likelihood**. How **expected** would $D$ be if $H$ were true? |

# Bayes' Theorem in Bayesian Statistics

## Bayes' Theorem

$$P(H \mid D) = P(H)\frac{P(D \mid H)}{P(D)}$$

| | |
|---|---|
| $P(H)$ | The **prior probability** of the hypothesis $H$. How plausible was it **before** observing D? |
| $P(D \mid H)$ | Called the **likelihood**. How **expected** would $D$ be if $H$ were true? |
| $P(H \mid D)$ | The **posterior probability** of $H$. How plausible is it **after** observing D? |

# Bayes' Theorem in Bayesian Statistics

## Bayes' Theorem

$$P(H \mid D) = P(H)\frac{P(D \mid H)}{P(D)}$$

| | |
|---|---|
| $P(H)$ | The **prior probability** of the hypothesis $H$. How plausible was it **before** observing D? |
| $P(D \mid H)$ | Called the **likelihood**. How **expected** would $D$ be if $H$ were true? |
| $P(H \mid D)$ | The **posterior probability** of $H$. How plausible is it **after** observing D? |
| $P(D)$ | Called the **marginal likelihood**. How **expected** was $D$ **overall** across **all** possible worlds? |

# Bayes' Theorem in Bayesian Statistics

## Bayes' Theorem

$$P(H \mid D) = P(H)\frac{P(D \mid H)}{P(D)}$$

Notice that Bayes theorem written this way can be seen as telling us how to **update** the plausibility of a possible world, $H$ in light of new information, $D$

$$P(H) \xrightarrow{\text{Multiply by } \frac{P(D \mid H)}{P(D)}} P(H \mid D)$$

# Summarizing the Path Counts

| Conjecture | Ways to produce ●○● |
|---|---|
| [○○○○] | $0 \times 4 \times 0 = 0$ |
| [●○○○] | $1 \times 3 \times 1 = 3$ |
| [●●○○] | $2 \times 2 \times 2 = 8$ |
| [●●●○] | $3 \times 1 \times 3 = 9$ |
| [●●●●] | $4 \times 0 \times 4 = 0$ |

- **Provided we weigh all 20 paths equally**, we get

$$P(H_2 \mid D) = \frac{\#\text{ paths in } (H_2 \text{ AND } D)}{\#\text{ total paths in } D} = \frac{3}{20}$$

# Summarizing the Path Counts

| Conjecture | Ways to produce ●○● |
|---|---|
| [○○○○] | $0 \times 4 \times 0 = 0$ |
| [●○○○] | $1 \times 3 \times 1 = 3$ |
| [●●○○] | $2 \times 2 \times 2 = 8$ |
| [●●●○] | $3 \times 1 \times 3 = 9$ |
| [●●●●] | $4 \times 0 \times 4 = 0$ |

- **Provided we weigh all 20 paths equally**, we get

$$P(H_2 \mid D) = \frac{\#\text{ paths in } (H_2 \text{ AND } D)}{\#\text{ total paths in } D} = \frac{3}{20}$$

- By weighing all paths equally, we are implicitly saying that we viewed all $H$s to have **equal plausibility** going in.

# Summarizing the Path Counts

| Conjecture | Ways to produce ●○● |
|---|---|
| [○○○○] | $0 \times 4 \times 0 = 0$ |
| [●○○○] | $1 \times 3 \times 1 = 3$ |
| [●●○○] | $2 \times 2 \times 2 = 8$ |
| [●●●○] | $3 \times 1 \times 3 = 9$ |
| [●●●●] | $4 \times 0 \times 4 = 0$ |

- **Provided we weigh all 20 paths equally**, we get

$$P(H_2 \mid D) = \frac{\#\text{ paths in } (H_2 \text{ AND } D)}{\#\text{ total paths in } D} = \frac{3}{20}$$

- By weighing all paths equally, we are implicitly saying that we viewed all $H$s to have **equal plausibility** going in.
- That is, the **prior plausibilities** $P(H_1)$, $P(H_2)$, $P(H_3)$, $P(H_4)$ and $P(H_5)$ were all equal

# Summarizing the Path Counts

| Conjecture | Ways to produce ●○● |
|---|---|
| [○○○○] | $0 \times 4 \times 0 = 0$ |
| [●○○○] | $1 \times 3 \times 1 = 3$ |
| [●●○○] | $2 \times 2 \times 2 = 8$ |
| [●●●○] | $3 \times 1 \times 3 = 9$ |
| [●●●●] | $4 \times 0 \times 4 = 0$ |

- **Provided we weigh all 20 paths equally**, we get

$$P(H_2 \mid D) = \frac{\#\text{ paths in } (H_2 \text{ AND } D)}{\#\text{ total paths in } D} = \frac{3}{20}$$

- By weighing all paths equally, we are implicitly saying that we viewed all $H$s to have **equal plausibility** going in.
- That is, the **prior plausibilities** $P(H_1)$, $P(H_2)$, $P(H_3)$, $P(H_4)$ and $P(H_5)$ were all equal
- Since the $H$s are mutually exclusive, this means each $P(H_i)$ is 1/5

# Summarizing the Path Counts

| Conjecture | Ways to produce ●○● |
|---|---|
| [○○○○] | $0 \times 4 \times 0 = 0$ |
| [●○○○] | $1 \times 3 \times 1 = 3$ |
| [●●○○] | $2 \times 2 \times 2 = 8$ |
| [●●●○] | $3 \times 1 \times 3 = 9$ |
| [●●●●] | $4 \times 0 \times 4 = 0$ |

- Since all 5 hypotheses have 64 equally likely paths, $P(D) = 20/(5 \times 64)$. So

$$\begin{aligned} P(H_2 \mid D) &= P(H_2)\frac{P(D \mid H_2)}{P(D)} \\ &= \left(\frac{1}{5}\right) \times \left(\frac{3/64}{20/(5 \times 64)}\right) \\ &= \frac{3 \cdot 5 \cdot 64}{5 \cdot 64 \cdot 20} \\ &= \frac{3}{20} \end{aligned}$$

# The Law of Total Probability

Let's examine this "update multiplier", $\frac{P(D \mid H)}{P(D)}$. Note that $D$ can be represented has

$$D = (D \text{ AND } H) \text{ OR } (D \text{ AND } (\text{NOT } H))$$

# The Law of Total Probability

Let's examine this "update multiplier", $\frac{P(D \mid H)}{P(D)}$. Note that $D$ can be represented has

$$D = (D \text{ AND } H) \text{ OR } (D \text{ AND } (\text{NOT } H))$$

Since $(D \text{ AND } H)$ and $(D \text{ AND } (\text{NOT } H))$ are mutually exclusive, it follows that

$$P(D) = P(D \text{ AND } H) + P(D \text{ AND } (\text{NOT } H))$$

# The Law of Total Probability

Let's examine this "update multiplier", $\frac{P(D \mid H)}{P(D)}$. Note that $D$ can be represented has

$$D = (D \text{ AND } H) \text{ OR } (D \text{ AND } (\text{NOT } H))$$

Since $(D \text{ AND } H)$ and $(D \text{ AND } (\text{NOT } H))$ are mutually exclusive, it follows that

$$P(D) = P(D \text{ AND } H) + P(D \text{ AND } (\text{NOT } H))$$

Moreover, the share of probability in $\Omega$ which is in $(D \text{ AND } H)$ can be found by taking the probability in $H$ and splitting off the part of it which is also in $D$; that is

$$P(D \text{ AND } H) = P(H)P(D \mid H)$$

# The Law of Total Probability

Let's examine this "update multiplier", $\frac{P(D \mid H)}{P(D)}$. Note that $D$ can be represented has

$$D = (D \text{ AND } H) \text{ OR } (D \text{ AND } (\text{NOT } H))$$

Since $(D \text{ AND } H)$ and $(D \text{ AND } (\text{NOT } H))$ are mutually exclusive, it follows that

$$P(D) = P(D \text{ AND } H) + P(D \text{ AND } (\text{NOT } H))$$

Moreover, the share of probability in $\Omega$ which is in $(D \text{ AND } H)$ can be found by taking the probability in $H$ and splitting off the part of it which is also in $D$; that is

$$P(D \text{ AND } H) = P(H)P(D \mid H)$$

Similarly

$$P(D \text{ AND } (\text{NOT } H)) = P(\text{NOT } H)P(D \mid \text{NOT } H)$$

# The Law of Total Probability

Since ($D$ AND $H$) and ($D$ AND (NOT $H$)) are mutually exclusive, it follows that

$$P(D) = P(D \text{ AND } H) + P(D \text{ AND } (\text{NOT } H))$$

Moreover, the share of probability in $\Omega$ which is in ($D$ AND $H$) can be found by taking the probability in $H$ and splitting off the part of it which is also in $D$; that is

$$P(D \text{ AND } H) = P(H)P(D \mid H)$$

Similarly

$$P(D \text{ AND } (\text{NOT } H)) = P(\text{NOT } H)P(D \mid \text{NOT } H)$$

So, all together:

$$P(D) = P(H)P(D \mid H) + P(\text{NOT } H)P(D \mid \text{NOT } H)$$

## The Law of Total Probability

$$P(D) = P(H)P(D \mid H) + P(\text{NOT } H)P(D \mid \text{NOT } H)$$

Notice that this means $P(D)$ is a **weighted average** of $P(D \mid H)$ and $P(D \mid \text{NOT } H)$, with weights given by $P(H)$ and $P(\text{NOT } H) = 1 - P(H)$.

## The Law of Total Probability

$$P(D) = P(H)P(D \mid H) + P(\text{NOT } H)P(D \mid \text{NOT } H)$$

Notice that this means $P(D)$ is a **weighted average** of $P(D \mid H)$ and $P(D \mid \text{NOT } H)$, with weights given by $P(H)$ and $P(\text{NOT } H) = 1 - P(H)$.

Implies that the update multpilier $\frac{P(D \mid H)}{P(D)}$ will be **bigger than 1** (leading to the plausibility of $H$ going up) if **hypothesizing $H$ makes $D$ more expected than it was on average**.

Similarly it will be **less than 1** if hypothesizing $H$ makes $D$ **less** expected than it was on average