Introduction to Probability

Probability - Why is it important?

- Quantifying uncertainty
- Dealing with random processes (such as data gathering or sampling)
- Understanding how good our inferences are - estimates have a margin of error.

NASHVILLE SOFTWARE SCHOOL

Probability

Most processes are **stochastic** rather than **deterministic**.

Deterministic: the same inputs always produce the same outputs

Stochastic: the same inputs do **not** always produce the same outputs

Probability

Deterministic Process: Finding the area (A) of a circle based on its radius (r):

$$A = \pi r^2$$

Stochastic Process: Finding (estimating) a man's shoe size based on his height.

NASHVILLE SOFTWARE SCHOOL

Probability - Informal Example

We can reason about random processes by using probability.

Let's think about the possible outcome of a die roll using probability.

Probability - Informal Example

What is the chance of rolling a 4?

Probability - Informal Example

What is the chance of rolling a 4?

Answer: There are 6 equally-likely faces, and only one of them is 4, so the chances of rolling a 4 are 1/6.

NASHVILLE SOFTWARE SCHOOL

Probability - Informal Example

What is the chance of rolling an even number?

Probability - Informal Example

What is the chance of rolling an even number?

Answer: There are 6 equally-likely faces, and three of them are even (2, 4, and 6), so the chances are 3/6.

NASHVILLE SOFTWARE SCHOOL

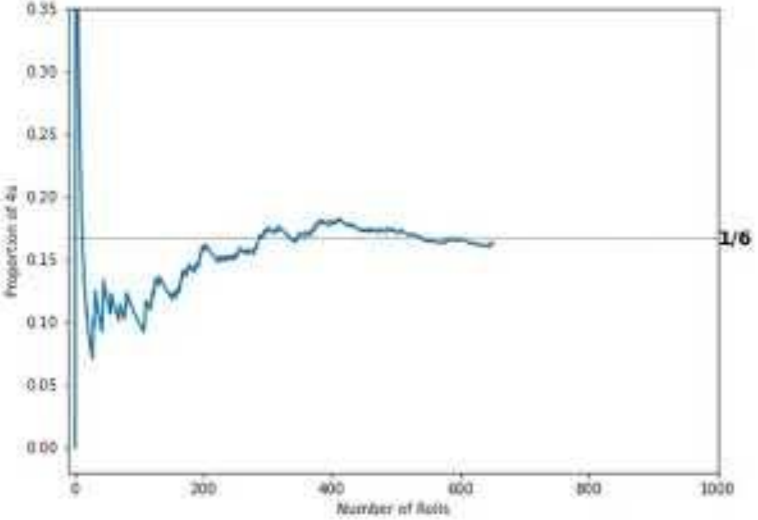
Probability as **Relative Frequency**

What do we mean when we say that the chance of rolling a 4 is 1/6?

Does it mean that if I roll it 6 times, it will land on 4 exactly once?

No! - But, if one were to roll the die a large number of times, then it should land on each face approximately 1/6 of the time.

Python Simulation of 1000 Die Rolls (Click to Open)



face	count	ratio
1	96	0.148
2	118	0.182
3	103	0.158
4	106	0.163
5	118	0.182
6	109	0.168

Probability as **Relative Frequency**

The probability of event *A* is the proportion of times that *A* occurs in a very long run of separate tries.

This notion of probability is called the **frequentist approach**.

Probability Terminology

Notation: For event *A*, the probability of *A* is written as *P(A)*

Probabilities range from 0 to 1, with 0 representing *impossibility* and 1 representing *certainty*.

An event with probability 0.7 will occur in about 70% of tries.

Basic Probability Properties

For the US,

P(Person has High Blood Pressure) = 0.45

Basic Probability Properties

For the US,

P(Person has High Blood Pressure) = 0.45

P(Person has Diabetes) = 0.11

Basic Probability Properties

For the US,

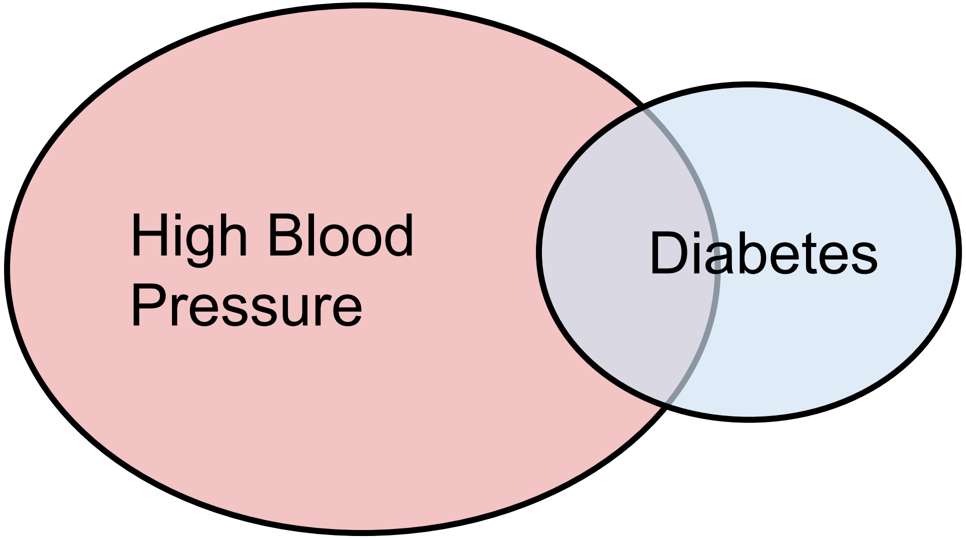
P(Person has High Blood Pressure) = 0.45

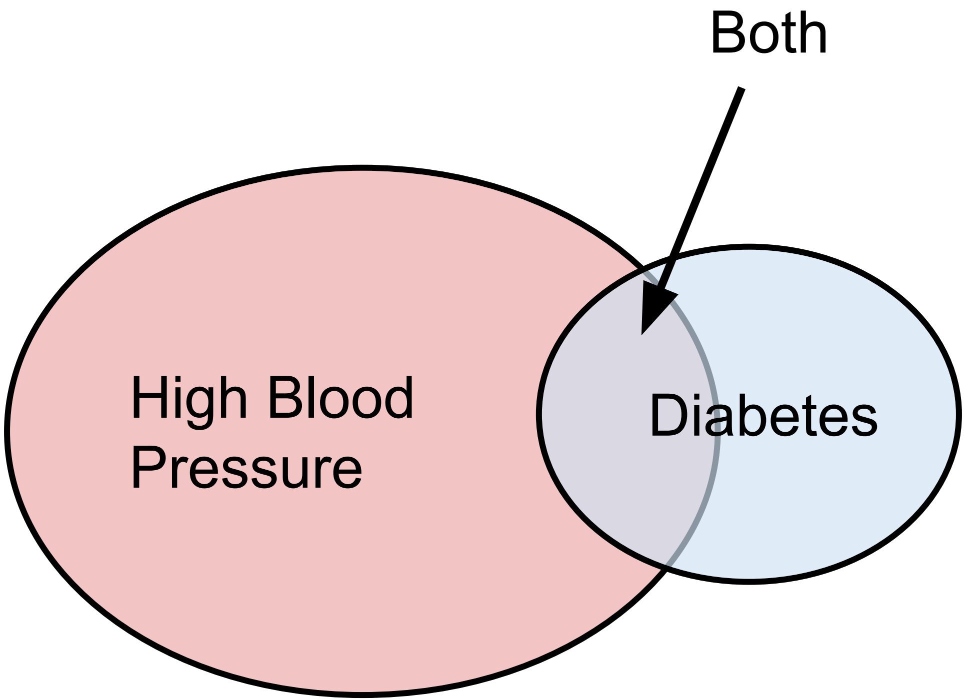
P(Person has Diabetes) = 0.11

Question: Is it true that

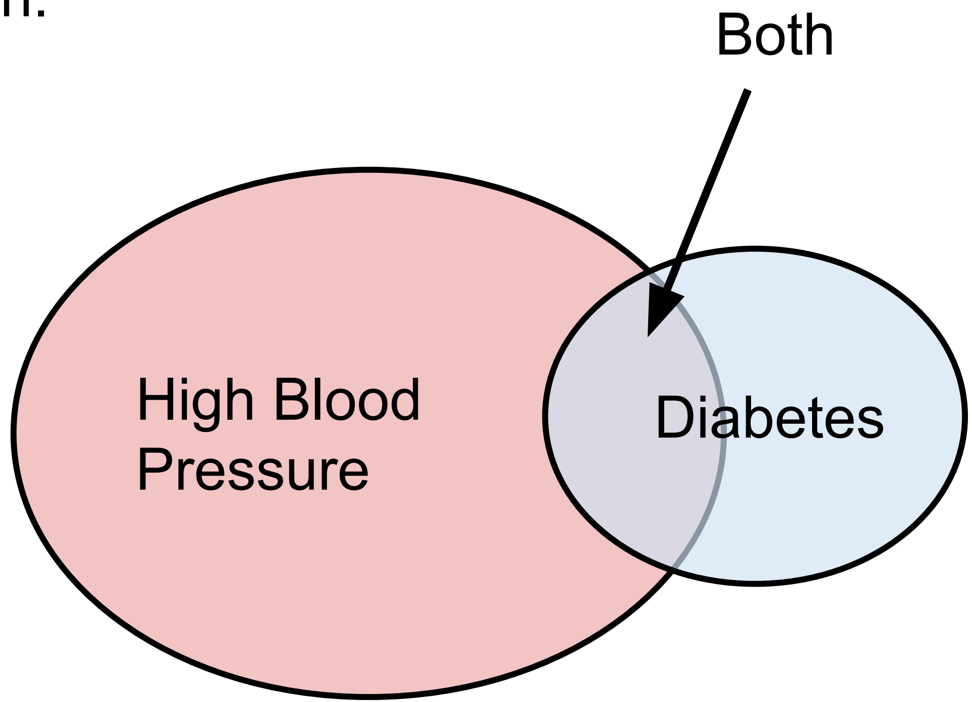
P(Person has High Blood Pressure **or** Diabetes) = 0.56 ?

High Blood
Pressure


High Blood
Pressure
Diabetes


Both
High Blood Pressure
Diabetes

If we just add probabilities, we will be double-counting those with both.



NASHVILLE SOFTWARE SCHOOL

Basic Probability Properties

For the US,

P(Person has High Blood Pressure) = 0.45

P(Person has Diabetes) = 0.11

Basic Probability Properties

For the US,

P(Person has High Blood Pressure) = 0.45

P(Person has Diabetes) = 0.11

P(Person has High Blood Pressure **and** Diabetes) = 0.08

NASHVILLE SOFTWARE SCHOOL

Basic Probability Properties

For the US,

P(Person has High Blood Pressure) = 0.45

P(Person has Diabetes) = 0.11

P(Person has High Blood Pressure **and** Diabetes) = 0.08

P(Person has High Blood Pressure **or** Diabetes)

= 0.45 + 0.11 - 0.08 = 0.48

Basic Probability Properties

If we roll a six-sided die,

$$P(\text{Roll a } 4) = 1/6$$

Basic Probability Properties

If we roll a six-sided die,

P(Roll a 4) = 1/6

P(Roll a 6) = 1/6

Basic Probability Properties

If we roll a six-sided die,

$P(\text{Roll a } 4) = 1/6$

$P(\text{Roll a } 6) = 1/6$

$P(\text{Roll a } 4 \text{ or a } 6) = \frac{1}{6} + \frac{1}{6} = 2/6$

NASHVILLE SOFTWARE SCHOOL

Basic Probability Properties

If we roll a six-sided die,

P(Roll a 4) = 1/6

P(Roll a 6) = 1/6

P(Roll a 4 or a 6) = ⅙ + ⅙ = 2/6

How is this case different?

Basic Probability Properties

If we roll a six-sided die,

$P(\text{Roll a } 4) = 1/6$

$P(\text{Roll a } 6) = 1/6$

$P(\text{Roll a } 4 \text{ or a } 6) = \frac{1}{6} + \frac{1}{6} = 2/6$

How is this case different? The outcomes are **mutually exclusive** - they can't happen simultaneously.

Basic Probability Properties

If we roll a six-sided die,

$P(\text{Roll a 4}) = 1/6$

$P(\text{Roll a 6}) = 1/6$

$P(\text{Roll a 4 or a 6}) = \frac{1}{6} + \frac{1}{6} = 2/6$

$P(\text{Roll a 4 } \textbf{and} \text{ a 6}) = 0$

How is this case different? The outcomes are **mutually exclusive** - they can't happen simultaneously.

Conditional Probability

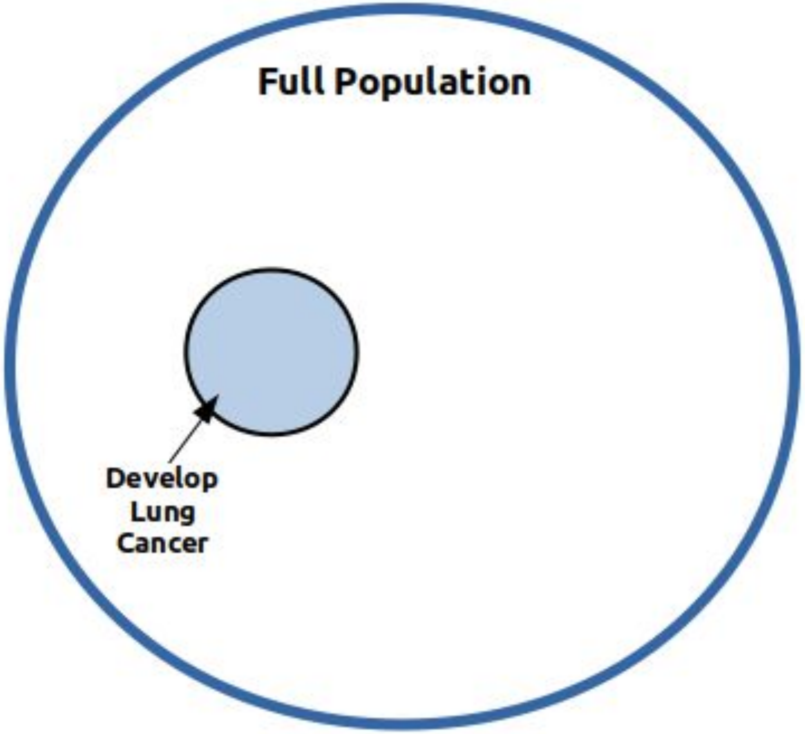
How does the likelihood of a particular outcome change if we are given more information?

Eg. What is the probability that a person develops lung cancer?

What if we know that person smokes? How does the probability change?

NASHVILLE SOFTWARE SCHOOL

Conditional Probability

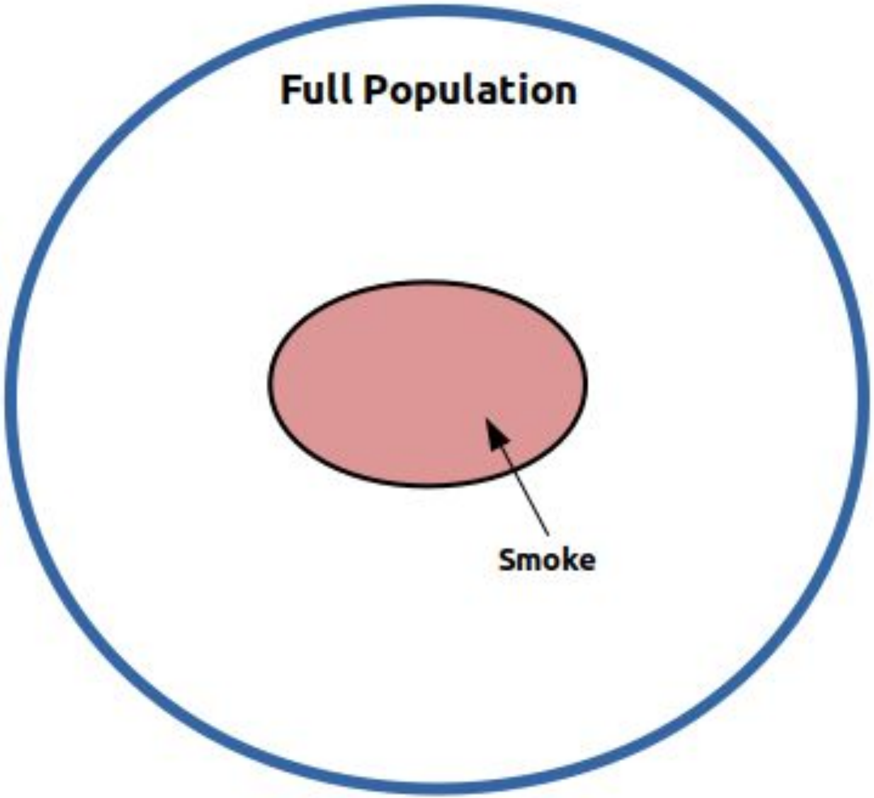


With no further information, our best guess would be the ratio of people from the full population who develop lung cancer.

This starting probability is called the **prior probability.**

NASHVILLE SOFTWARE SCHOOL

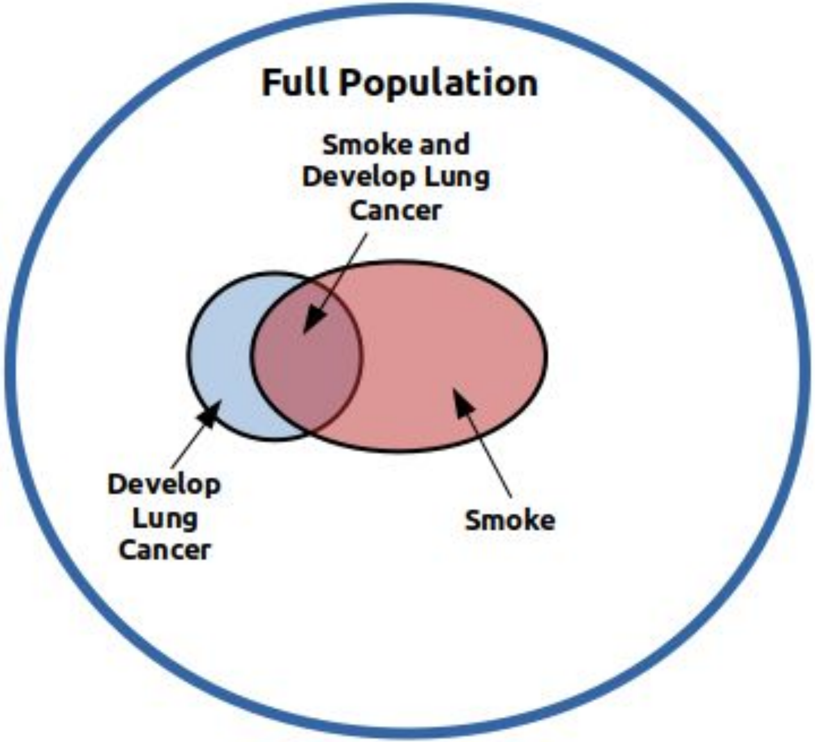
Conditional Probability



But if we had additional information, that they smoke, we could potentially use this information to refine our first guess.

To do this, we need to know the overlap between smokers and those who develop lung cancer.

Conditional Probability

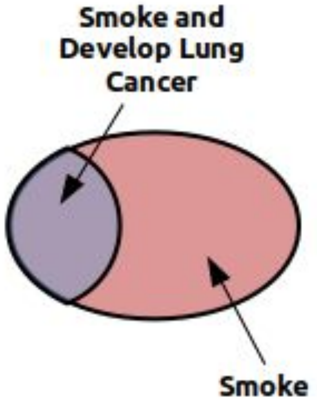


Let's say that these two subpopulations overlap like this.

Now for our guess, we just need to consider the subset of smokers.

NASHVILLE SOFTWARE SCHOOL

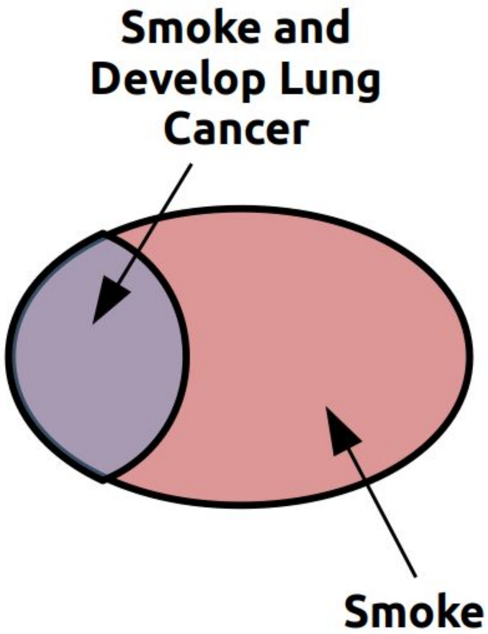
Conditional Probability



Let's say that these two subpopulations overlap like this.

Now for our guess, we just need to consider the subset of smokers.

Conditional Probability



Changing our population of interest to just those people who smoke, our best guess is the ratio of people who smoke *and* develop lung cancer out of the population of smokers.

This is known as the **posterior probability.**

Conditional Probability

For two events *A* and *B*, the **conditional probability of *A*, given that *B* has occurred** is $P(A \mid B)$

Probability of developing lung cancer:

$$P(\text{lung cancer})$$

Probability of developing lung cancer, given that a person smokes:

$$P(\text{lung cancer} \mid \text{smokes})$$

NASHVILLE SOFTWARE SCHOOL

Conditional Probability

We can calculate the conditional probability as

$$P(A|B) = P(A \text{ and } B) \,/\, \mathrm{P}(B)$$

We want *both* A and B to occur.

We only care about cases where B occurred.

NASHVILLE SOFTWARE SCHOOL

$P(A \mid B) = P(A \text{ and } B) \,/\, \mathrm{P}(B)$

Conditional Probability

Example: You roll a fair die. What is the probability that the roll is odd, given that it is greater than 1?

$$P(A \mid B) = P(A \text{ and } B) / \mathrm{P}(B)$$

Conditional Probability

Example: You roll a fair die. What is the probability that the roll is odd, given that it is greater than 1?

$$\mathrm{P}(\text{odd} \mid >1) =$$

$P(A|B) = P(A \text{ and } B) / \mathrm{P}(B)$

Conditional Probability

Example: You roll a fair die. What is the probability that the roll is odd, given that it is greater than 1?

$$\mathrm{P}(\text{odd} \mid >1) = \mathrm{P}(\text{odd and } >1) / \mathrm{P}(>1)$$

$$P(A \mid B) = P(A \text{ and } B) / P(B)$$

Conditional Probability

Example: You roll a fair die. What is the probability that the roll is odd, given that it is greater than 1?

$$P(\text{odd} \mid >1) = P(\text{odd and } >1) / P(>1)$$

$$= (2/6) / (5/6)$$

$P(A \mid B) = P(A \text{ and } B) \,/\, \mathrm{P}(B)$

Conditional Probability

Example: You roll a fair die. What is the probability that the roll is odd, given that it is greater than 1?

$$P(\text{odd} \mid >1) = P(\text{odd and } >1) \,/\, P(>1)$$

$$= (2/6) \,/\, (5/6)$$

$$= 2/5 = 0.4$$

$$P(A|B) = P(A \text{ and } B) \,/\, P(B)$$

Conditional Probability

You can also rearrange the formula to get one for the **joint probability**:

$$P(A|B) = P(A \text{ and } B) \,/\, P(B)$$

$$\Updownarrow$$

$$P(A \text{ and } B) = P(A|B) \cdot P(B)$$

$$P(A \text{ and } B) = P(A \mid B) \cdot P(B)$$

Conditional Probability

Example: An online retailer know that historically, 30% of visitors to their site will add at least one item to their shopping cart. They also know that 60% of the time when someone has added an item to their cart, they will complete their purchase. What is the overall probability that a visitor completes a purchase?

Conditional Probability

$P(A \text{ and } B) = P(A \mid B) \cdot P(B)$

Example: An online retailer know that historically, 30% of visitors to their site will add at least one item to their shopping cart. They also know that 60% of the time when someone has added an item to their cart, they will complete their purchase. What is the overall probability that a visitor completes a purchase?

$P(\text{purchase} \mid \text{cart}) = 0.6$

$P(A \text{ and } B) = P(A \mid B) \cdot P(B)$

Conditional Probability

Example: An online retailer know that historically, 30% of visitors to their site will add at least one item to their shopping cart. They also know that 60% of the time when someone has added an item to their cart, they will complete their purchase. What is the overall probability that a visitor completes a purchase?

$P(\text{purchase} \mid \text{cart}) = 0.6$

$P(\text{cart}) = 0.3$

$$P(A \text{ and } B) = P(A \mid B) \cdot P(B)$$

Conditional Probability

Example: An online retailer know that historically, 30% of visitors to their site will add at least one item to their shopping cart. They also know that 60% of the time when someone has added an item to their cart, they will complete their purchase. What is the overall probability that a visitor completes a purchase?

P(purchase | cart) = 0.6

P(cart) = 0.3

P(purchase and cart) =

$$P(A \text{ and } B) = P(A \mid B) \cdot P(B)$$

Conditional Probability

Example: An online retailer know that historically, 30% of visitors to their site will add at least one item to their shopping cart. They also know that 60% of the time when someone has added an item to their cart, they will complete their purchase. What is the overall probability that a visitor completes a purchase?

$P(\text{purchase} \mid \text{cart}) = 0.6$

$P(\text{cart}) = 0.3$

$P(\text{purchase and cart}) = P(\text{purchase} \mid \text{cart}) \cdot P(\text{cart}) =$

$$P(A \text{ and } B) = P(A \mid B) \cdot P(B)$$

Conditional Probability

Example: An online retailer know that historically, 30% of visitors to their site will add at least one item to their shopping cart. They also know that 60% of the time when someone has added an item to their cart, they will complete their purchase. What is the overall probability that a visitor completes a purchase?

P(purchase | cart) = 0.6

P(cart) = 0.3

P(purchase and cart) = P(purchase | cart) · P(cart) = 0.6 · 0.3 = 0.18

NASHVILLE SOFTWARE SCHOOL

Probability as **Degree of Belief**

Thinking in terms of conditional probabilities, we can understand probability in a different way.

Probability quantifies how *certain* we are about a given hypothesis.

By incorporating more information, we can update our probabilities.

This interpretation of probability is known as the **Bayesian approach**.

Bayesian Statistics

Thomas Bayes, an 18th century statistician, minister and philosopher.

Formulated a special case of what is now known as *Bayes' Theorem*.

Bayes' Theorem

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

Bayes' Theorem can be viewed as a recipe for updating our belief about event A by incorporating information about event B.

Bayes' Theorem

Prior

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

Bayes' Theorem can be viewed as a recipe for updating our belief about event A by incorporating information about event B.

Bayes' Theorem

Prior

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

Posterior

Bayes' Theorem can be viewed as a recipe for updating our belief about event A by incorporating information about event B.

Bayes' Theorem

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

Example: You roll a fair die. What is the probability that the roll is odd, given that it is greater than 1?

P(odd | >1) =

Bayes' Theorem

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

Example: You roll a fair die. What is the probability that the roll is odd, given that it is greater than 1?

$$P(\text{odd} \mid >1) = P(>1 \mid \text{odd}) \,/\, P(>1) \cdot P(\text{odd})$$

Bayes' Theorem

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

Example: You roll a fair die. What is the probability that the roll is odd, given that it is greater than 1?

$$P(\text{odd} \mid >1) = P(>1 \mid \text{odd}) \,/\, P(>1) \cdot P(\text{odd})$$

$$= (2/3) \,/\, (5/6) \cdot (3/6)$$

Bayes' Theorem

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

Example: You roll a fair die. What is the probability that the roll is odd, given that it is greater than 1?

$$P(\text{odd} \mid >1) = P(>1 \mid \text{odd}) \,/\, P(>1) \cdot P(\text{odd})$$

$$= (2/3) \,/\, (5/6) \cdot (3/6)$$

$$= (12/15) \cdot (3/6)$$

NASHVILLE SOFTWARE SCHOOL

Bayes' Theorem

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

Example: You roll a fair die. What is the probability that the roll is odd, given that it is greater than 1?

$$P(\text{odd} \mid >1) = P(>1 \mid \text{odd}) \,/\, P(>1) \cdot P(\text{odd})$$

$$= (2/3) \,/\, (5/6) \cdot (3/6)$$

$$= (12/15) \cdot (3/6)$$

It is less likely for the roll to be >1 when we know that it is odd.

NASHVILLE SOFTWARE SCHOOL

Bayes' Theorem

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

Example: You roll a fair die. What is the probability that the roll is odd, given that it is greater than 1?

$$P(\text{odd} \mid >1) = P(>1 \mid \text{odd}) \,/\, P(>1) \cdot P(\text{odd})$$

$$= (2/3) \,/\, (5/6) \cdot (3/6)$$

$$= (12/15) \cdot (3/6)$$

$$= (4/5) \cdot (3/6) = 2/5 = 0.4$$

Conditional Probability

Events A and B are **independent** if the occurrence of B in no way informs us about the probability of A.

Conditional Probability

Events *A* and *B* are **independent** if the occurrence of *B* in no way informs us about the probability of *A*.

That is, $P(A|B) = P(A)$

Conditional Probability

Events *A* and *B* are **independent** if the occurrence of *B* in no way informs us about the probability of *A*.

That is, $P(A|B) = P(A)$

Equivalently, $P(A \text{ and } B) = P(A) \cdot P(B)$

Conditional Probability

Events *A* and *B* are **independent** if the occurrence of *B* in no way informs us about the probability of *A*.

That is, $P(A|B) = P(A)$

Equivalently, $P(A \text{ and } B) = P(A) \cdot P(B)$

Otherwise, we say that *A* and *B* are **dependent**.

Conditional Probability

From the Bayesian point of view, if events A and B are independent, knowing that event A has occurred does not allow us to update our belief about the likelihood of B occurring.

Conditional Probability

Recall from the previous example:

$$P(\text{odd} \mid >1) = 0.4 < 0.5 = P(\text{odd})$$

Conditional Probability

Recall from the previous example:

$$P(\text{odd} \mid >1) = 0.4 < 0.5 = P(\text{odd})$$

Thus, rolling an odd number and rolling a number that is greater than 1 are **dependent events**.

Conditional Probability

Example: You roll a fair die. What is the probability that the roll is even, given that it is less than 5?

$$P(\text{even} \mid <5) =$$

Conditional Probability

Example: You roll a fair die. What is the probability that the roll is even, given that it is less than 5?

$$P(\text{even} \mid <5) = P(\text{even and } <5) \;/\; P(<5)$$

Conditional Probability

Example: You roll a fair die. What is the probability that the roll is even, given that it is less than 5?

$$P(\text{even} \mid <5) = P(\text{even and } <5) \,/\, P(<5)$$

$$= (2/6) \,/\, (4/6)$$

Conditional Probability

Example: You roll a fair die. What is the probability that the roll is even, given that it is less than 5?

$$P(\text{even} \mid <5) = P(\text{even and } <5) \,/\, P(<5)$$

$$= (2/6) \,/\, (4/6)$$

$$= 2/4 = 0.5$$

Conditional Probability

Example: You roll a fair die. What is the probability that the roll is even, given that it is less than 5?

$$P(\text{even} \mid <5) = P(\text{even and } <5) \,/\, P(<5)$$

$$= (2/6) \,/\, (4/6)$$

$$= 2/4 = 0.5 = P(\text{even})$$

Conditional Probability

Example: You roll a fair die. What is the probability that the roll is even, given that it is less than 5?

$$P(\text{even} \mid <5) = P(\text{even and } <5) \,/\, P(<5)$$

$$= (2/6) \,/\, (4/6)$$

$$= 2/4 = 0.5 = P(\text{even})$$

The conditional probability is the same as the unconditional one, so these events are **independent.**

Conditional Probability

Example: Flipping a coin twice. What is P(Both Coins Landing on Heads)?

Conditional Probability

Example: Flipping a coin twice. What is P(Both Coins Landing on Heads)?

P(Second Coin Landing on Heads) = 0.5

Conditional Probability

Example: Flipping a coin twice. What is P(Both Coins Landing on Heads)?

P(Second Coin Landing on Heads) = 0.5

P(Second Coin Landing on Heads | First Coin Landing on Heads) = 0.5

Conditional Probability

Example: Flipping a coin twice. What is P(Both Coins Landing on Heads)?

P(Second Coin Landing on Heads) = 0.5

P(Second Coin Landing on Heads | First Coin Landing on Heads) = 0.5

That is, knowing the outcome from the first flip gives us no additional information about the next one.

Conditional Probability

Example: Flipping a coin twice

P(Both Coins Landing on Heads)

Conditional Probability

Example: Flipping a coin twice

P(Both Coins Landing on Heads)

= P(First Coin Landing on Heads) · P (Second Coin Landing on Heads)

NASHVILLE SOFTWARE SCHOOL

Conditional Probability

Example: Flipping a coin twice

$$P(\text{Both Coins Landing on Heads})$$

$$= P(\text{First Coin Landing on Heads}) \cdot P\,(\text{Second Coin Landing on Heads})$$

$$= 0.5 \cdot 0.5$$

$$= 0.25$$

NASHVILLE SOFTWARE SCHOOL

Conditional Probability

Why do we care?

When drawing a sample, we usually assume that the individuals are drawn *independently*.

When we're looking at a dataset, we usually assume that our observations are independent of each other.

Conditional Probability

We will usually make the further assumption that all of our observations arose from the same data generation process.

The combination of this assumption plus the assumption of independence is usually shortened to **iid**, standing for *independent and identically distributed*.