

DSC 40A

Theoretical Foundations of Data Science I

In This Video

- Conditional probability, the probability of one event given that another has occurred

Conditional probabilities

Probability of an event may **change** if have additional information about outcomes.

Suppose E and F are events, and $P(F) > 0$. Then,

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

i.e.,

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

Conditional probabilities

Are these probabilities equal?

The probability that **two siblings are girls** if we know the oldest is a girl.

The probability that **two siblings are boys** if we know that there is a boy.

Assume that each child being a boy or a girl is equally likely.

What do you think?

- A. they are equal
- B. they are not equal

Conditional probabilities

Are these probabilities equal?

The probability that **two siblings are girls** *if we know the oldest is a girl.*

The probability that **two siblings are boys** if we know that there is a boy.

Assume that each child being a boy or a girl is equally likely.

Conditional probabilities

Are these probabilities equal?

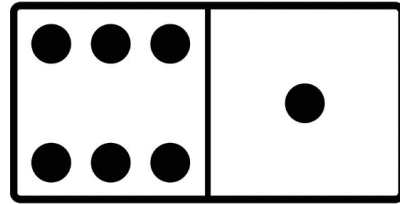
The probability that **two siblings are girls** if we know the oldest is a girl.

The probability that **two siblings are boys** *if we know that there is a boy.*

Assume that each child being a boy or a girl is equally likely.

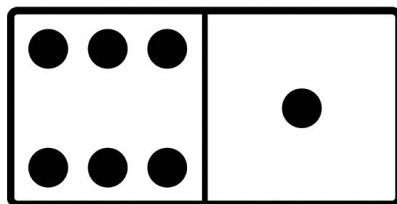
Dominoes

In a set of dominos, each tile has two sides with a number of dots on each side: zero, one, two, three, four, five or six. There are 28 total tiles, with each number of dots appearing alongside each other number (including itself) on a single tile.



Dominoes

In a set of dominos, each tile has two sides with a number of dots on each side: zero, one, two, three, four, five or six. There are 28 total tiles, with each number of dots appearing alongside each other number (including itself) on a single tile.



Question 1: What is the probability of drawing a “double” from a set of dominoes — that is, a tile with the same number on both sides?

Question 2: Now you pick a random tile from the set and uncover only one side, revealing that it has six dots. What’s the probability that this tile is a double, with six on both sides?

Question 3: Now your friend picks a random tile from the set, looks at it, and tells you that they have a six. What is the probability that your friend’s tile is a double, with six on both sides?

Dominoes

Question 1: What is the probability of drawing a “double” from a set of dominoes — that is, a tile with the same number on both sides?

Dominoes

Question 2: Now you pick a random tile from the set and uncover only one side, revealing that it has six dots. What's the probability that this tile is a double, with six on both sides?

Dominoes

Question 2: Now you pick a random tile from the set and uncover only one side, revealing that it has six dots. What's the probability that this tile is a double, with six on both sides?



Dominoes

Question 3: Now your friend picks a random tile from the set, looks at it, and tells you that they have a six. What is the probability that your friend's tile is a double, with six on both sides?

Conditional probabilities: Simpson's Paradox

	Treatment A	Treatment B
Small kidney stones	81 successes / 87 (93%)	234 successes / 270 (87%)
Large kidney stones	192 successes / 263 (73%)	55 successes / 80 (69%)
Combined	273 successes / 350 (78%)	289 successes / 350 (83%)

Which treatment is better?

- A. Treatment A for all cases.
- B. Treatment B for all cases.
- C. A for small and B for large.
- D. A for large and B for small.

Conditional probabilities: Simpson's Paradox

	Treatment A	Treatment B
Small kidney stones	81 successes / 87 (93%)	234 successes / 270 (87%)
Large kidney stones	192 successes / 263 (73%)	55 successes / 80 (69%)
Combined	273 successes / 350 (78%)	289 successes / 350 (83%)

Simpson's Paradox

"When the less effective treatment is applied more frequently to easier cases, it can appear to be a more effective treatment."

Summary

- Today, we studied conditional probability.
- **Next time:** How do we use probability to answer questions about random samples?