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.

ex.) rolling a die

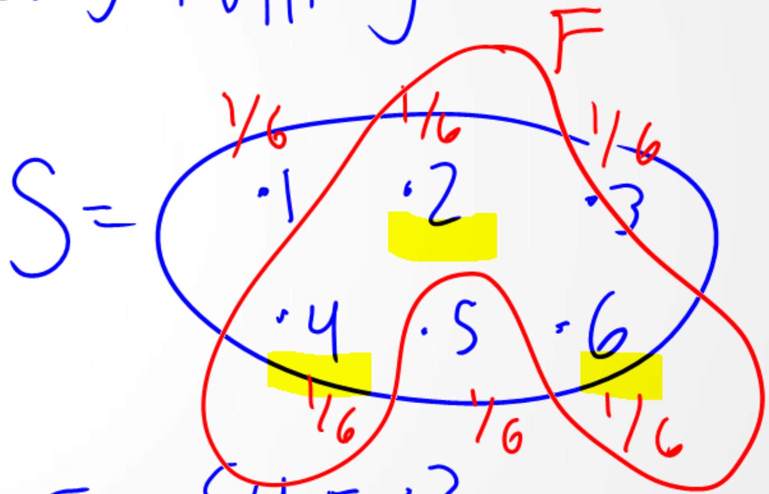
Suppose E and F are events, and P(F) > 0. Then,

$E=\{4,5,6\}$

$F=\{2,4,6\} \rightarrow P(F)=1/2$

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

$E\cap F=\{4,6\} \rightarrow P(E\cap F)=1/3$



i.e., $P(E\cap F)=P(E|F)P(F)$

$E=\{4,5,6\}$

$F=\{2,4,6\}$ $P(>3\,|\,\text{even})=\dfrac{\frac{1}{6}+\frac{1}{6}}{\frac{1}{6}+\frac{1}{6}+\frac{1}{6}}=\boxed{\frac{2}{3}}$ $P(E)=\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{1}{2}$

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

$S = \{ \overset{1/2}{b}, \overset{1/2}{g} \}$ for 2^{nd} sibling

Are these probabilities equal?

The probability that **two siblings are girls** *if we know the oldest is a girl.* $1/2$

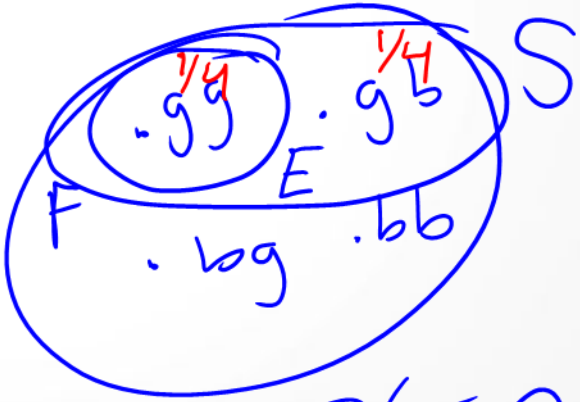
The probability that **two siblings are boys** if we know that there is a boy.

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

$S = \{ \underset{1/4}{gg}, \underset{1/4}{gb}, \underset{1/4}{bg}, \underset{1/4}{bb} \}$

$E = \{gg\} \longrightarrow P(E) = 1/4$

$F = \{gg, gb\} \rightarrow P(F) = 1/2$



$P(E|F) = \dfrac{P(E \cap F)}{P(F)} = \dfrac{1/4}{1/2}$

$P(E \cap F) = P(E) = 1/4$

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.* $1/3$

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

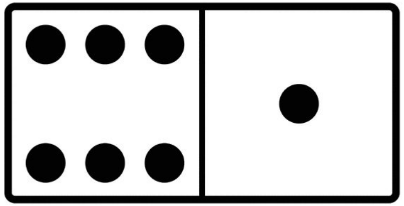
$S = \{gg, bb, gb, bg\}$
$\quad 1/4 \quad 1/4 \quad 1/4 \quad 1/4$

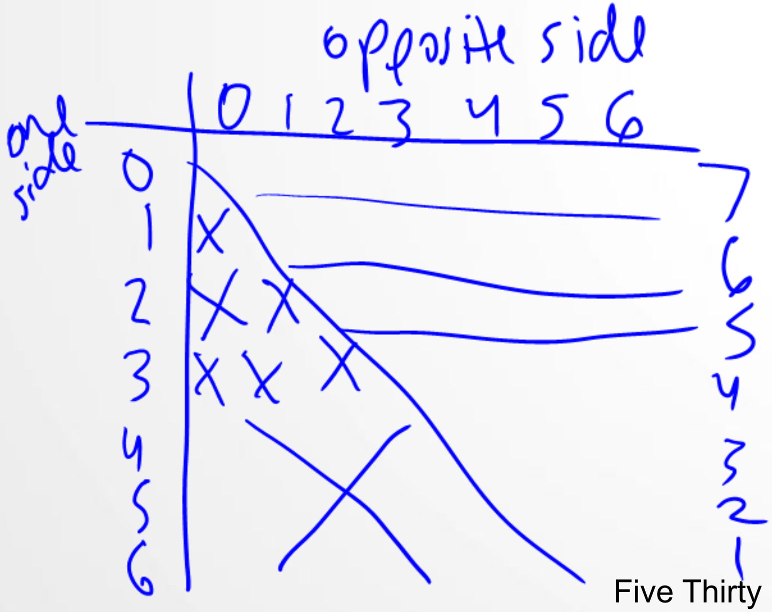
$E = \{bb\} \rightarrow P(E) = 1/4 \qquad P(E \cap F) = P(E) = 1/4$

$F = \{bb, gb, bg\} \rightarrow P(F) = 3/4 \qquad P(E|F) = \frac{P(E\cap F)}{P(F)} = \frac{1/4}{3/4}$

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.

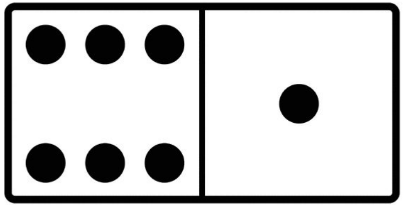




Five Thirty Eight Riddler Express: https://fivethirtyeight.com/features/can-you-eat-an-apple-like-a-toddler/

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?

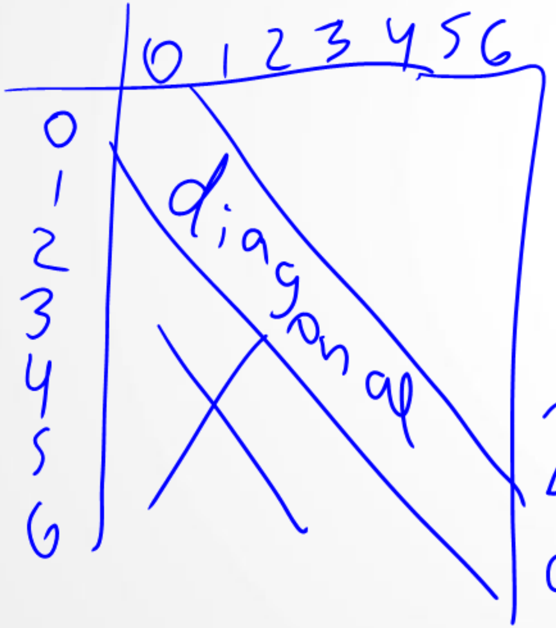
Question 2: Now you pick a random tile from the set and uncover only one side, revealing that it has six dots. What’s the probability that this tile is a double, with six on both sides?

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?

Five Thirty Eight Riddler Express: https://fivethirtyeight.com/features/can-you-eat-an-apple-like-a-toddler/

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?



28 dominoes = S

$$\frac{7}{28} = \frac{1}{4}$$

Five Thirty Eight Riddler Express: https://fivethirtyeight.com/features/can-you-eat-an-apple-like-a-toddler/

Dominoes

$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{2/56}{8/56} = 2/8 = \boxed{1/4}$

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?

~~S = all 28 dominoes~~

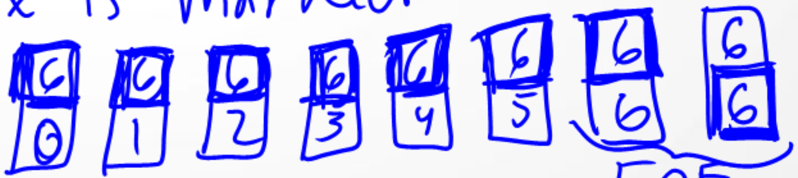
~~E = all doubles~~

~~F = all dominoes with at least one 6~~

S = marked dominoes → 56

E = marked dominoes where both halves are same → 14

F = marked dominoes where a six is marked → 8



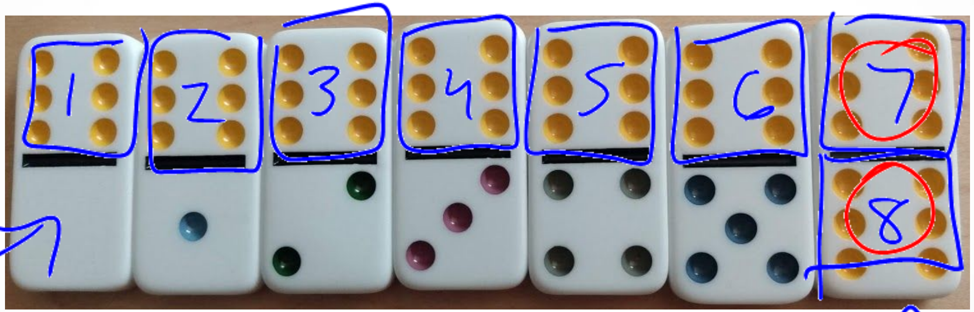
$E \cap F$

Five Thirty Eight Riddler Express: https://fivethirtyeight.com/features/can-you-eat-an-apple-like-a-toddler/

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?

F =



S = halves that you saw → 8

E = halves with a 6 on the other half → 2

$P(E) = \frac{2}{8} = \boxed{\frac{1}{4}}$

S = marked dominoes, where marked domino is domino with one side marked

two halves that could have been looked at

Five Thirty Eight Riddler Express: https://fivethirtyeight.com/features/can-you-eat-an-apple-like-a-toddler/

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?

S = 28 dominoes → 28

E = doubles → (7)

F = dominoes with at least one 6 → 7

$\frac{6}{0}$ $\frac{6}{1}$ $\frac{6}{2}$ $\frac{6}{3}$ $\frac{6}{4}$ $\frac{6}{5}$ $\frac{6}{6}$

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{1/28}{7/28} = \boxed{1/7}$$

Conditional probabilities: Simpson's Paradox

	Treatment A	Treatment B
Small kidney stones easy case	81 successes / 87 (93%)	234 successes / 270 (87%)
Large kidney stones hard case	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.

C. R. Charig, D. R. Webb, S. R. Payne, J. E. Wickham (29 March 1986). "Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy". Br Med J (Clin Res Ed) 292 (6524): 879–882. doi:10.1136/bmj.292.6524.879. PMC 1339981. PMID 3083922. cf. Wikipedia "Simpson's Paradox"

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."

C. R. Charig, D. R. Webb, S. R. Payne, J. E. Wickham (29 March 1986). "Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy". Br Med J (Clin Res Ed) 292 (6524): 879–882. doi:10.1136/bmj.292.6524.879. PMC 1339981. PMID 3083922. cf. Wikipedia "Simpson's Paradox"

Summary

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