CS 440/ECE 448 Lecture 5: Probability

Mark Hasegawa-Johnson, 2/2021

All content CC-BY-4.0 unless otherwise noted

CC-BY-3.0 Image by Diacritica, 2010

https://commons.wikimedia.org/wiki/File:6sided_dice.jpg

Outline

- Motivation: Why use probability?
- The axioms of probability
- Random variables
- Conditional probability
- Mutually exclusive vs. Independent vs. Conditionally Independent

Why use probability?

- Stochastic environment: outcome of an action might be truly random.
- Multi-agent environment:
  - If other players are rational and their goals are known, then you don't need probability; you just work out what their rational actions will be.
  - If other players have unknown goals, then model them as random.
- Unknown environment: outcome of an action is not truly random, but you don't know what the outcome will be.
  - In this case, "probability" measures your belief: P(Q|A)=the degree to which you believe that action A will produce outcome Q.
- Computational complexity:
  - Instead of searching 1b paths using A*, you could randomly choose 1k paths to try, and then choose the best of those.

Why NOT use probability?

- Multi-agent environment:
  - Maybe it’s better to find out what the other players really want?
- Unknown environment:
  - Maybe it’s better to learn the rules of the game?
- Computational complexity:
  - Maybe it’s better to do a complete search, instead of just a partial search?

Notice: these are quantitative questions. “Better” requires some metric: how much better, and with what probability?

What is probability?

- Latin *probabilis* = probable, commendable, believable, from *probare* = to test something
- If tested, it will (probably) turn out to be true

Outline

- Motivation: Why use probability?
- The axioms of probability
- Random variables
- Conditional probability
- Mutually exclusive vs. Independent vs. Conditionally Independent

The axioms of probability

A = some future event, e.g., “it will rain tomorrow.”

$P(A)$ = the degree to which we believe that event A, if tested, will turn out to be true.

The axioms of probability

Axiom 1: every event has a non-negative probability.

$$P(A) \geq 0$$

Axiom 2: a certain event has probability 1.

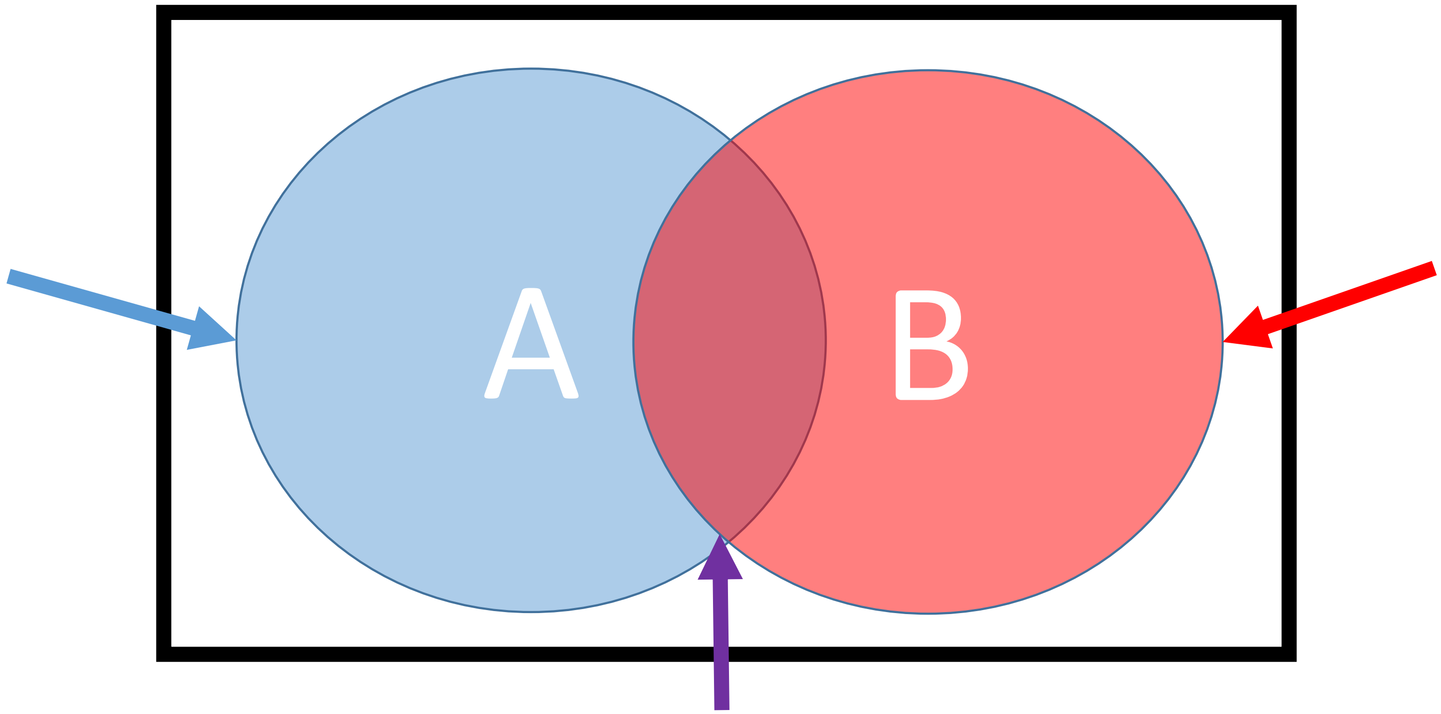
$$P(\mathrm{True}) = 1$$

Axiom 3: probability measures behave like set measures.

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

Axiom 3: probability measures behave like set measures.

Area of the whole rectangle is $P(\text{True}) = 1$.



Area of this circle is $P(A)$.

Area of this circle is $P(B)$.

Area of their intersection is $P(A \wedge B)$.
Area of their union is $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Example

- A = “it will rain tomorrow.” Suppose $P(A) = 0.4$.
- B = “it will snow tomorrow.” Suppose $P(B) = 0.2$.
- $A \wedge B$ = “it will both rain and snow tomorrow.” Suppose
$$P(A \wedge B) = 0.1$$

Then the probability that it will either rain or snow tomorrow is
$$P(A \vee B) = P(A) + P(B) - P(A \wedge B) = 0.4 + 0.2 - 0.1 = 0.5$$

Outline

- Motivation: Why use probability?
- The axioms of probability
- Random variables
- Conditional probability
- Mutually exclusive vs. Independent vs. Conditionally Independent

Random variables

- A random variable is a function that maps from the outcomes of an experiment to a set of values
- Example: throw four dice, all different colors. X = number of pips showing on the green die.
- Then run the experiment…

- In this particular outcome, X=3.
- In some other outcome, X would have taken a different value.

Notation: $P(X = x)$

- Capital letters are random variables. Small letters are values that the random variable might take.
- "$X = x$" is an event. As such, it has a probability. For example, we can talk about the probability $P(X = x)$:

$$P(X = x) = \frac{1}{6} \quad \forall\, x \in \{1,2,3,4,5,6\}$$

- $\forall$ means "for all." The equation above is shorthand for these six equations:

$$P(X = 1) = \frac{1}{6}, \qquad P(X = 2) = \frac{1}{6}, \qquad P(X = 3) = \frac{1}{6},$$

$$P(X = 4) = \frac{1}{6}, \qquad P(X = 5) = \frac{1}{6}, \qquad P(X = 6) = \frac{1}{6}$$

Notation: $P(X)$

X is not an event, and it's not a value; it's a function. So $P(X)$ is NOT a number. Instead, $P(X)$ is a table, showing all of the values X might take, and the probabilities of each. For example:

$P(X) =$

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Abuse of Notation: Events and Binary Random Variables

- There's one confusing thing. A capital letter might be either an event (A="it will rain tomorrow"), or a random variable (X="number of pips showing").
- P(A) is a number, but P(X) is a table.
- You have to pay attention to whether the capital letter is an event, or a random variable.

Outline

- Motivation: Why use probability?
- The axioms of probability
- Random variables
- Conditional probability
- Mutually exclusive vs. Independent vs. Conditionally Independent

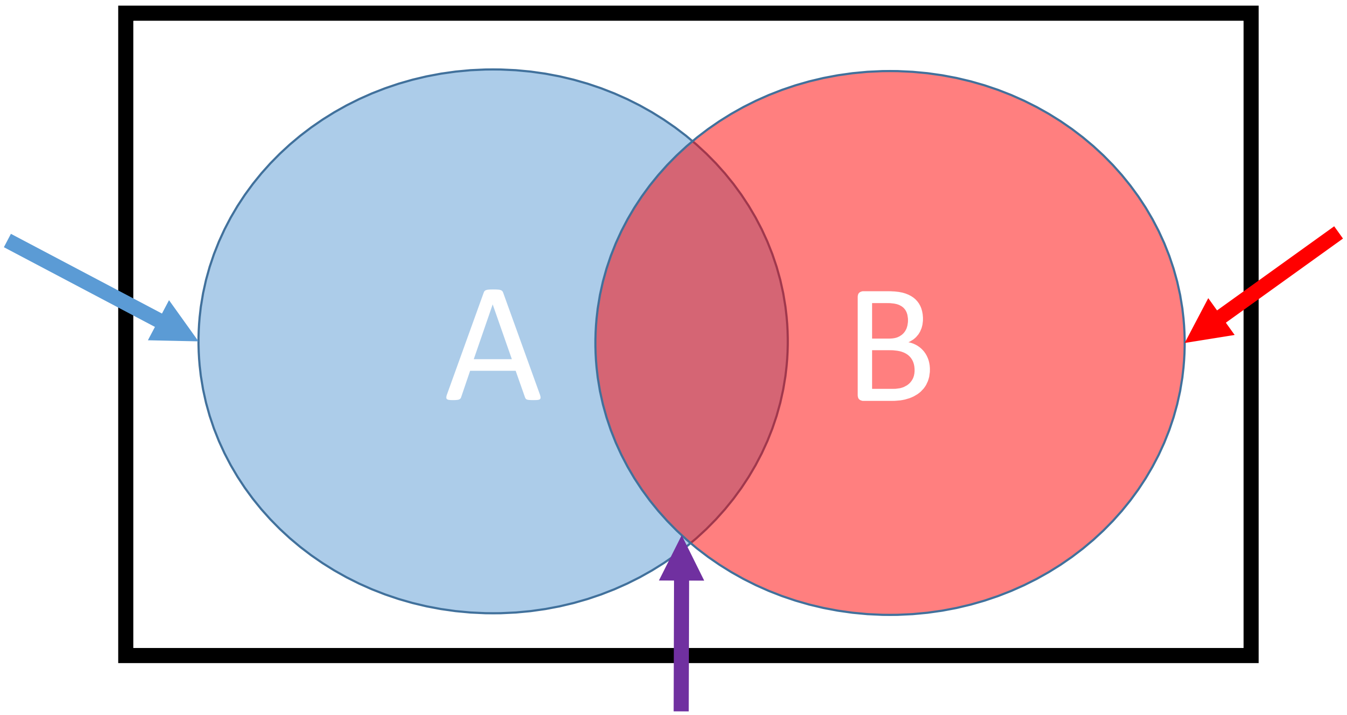
Joint and Conditional probabilities: definitions

- $P(A \land B)$ is the probability that both event A and event B happen. This is called their **joint probability**.
- $P(B|A)$ is the probability that event B happens, given that event A happens. This is called the **conditional probability** of B given A.
- Example:
  - A = "it will rain tomorrow"
  - B = "it will snow tomorrow"
  - $P(A \land B)$ = probability that it will both snow and rain
  - $P(B|A)$ = probability that it will snow, given that it rains

Joint probabilities are usually given in the problem statement

Area of the whole rectangle is $P(\text{True}) = 1$.

Suppose $P(A) = 0.4$



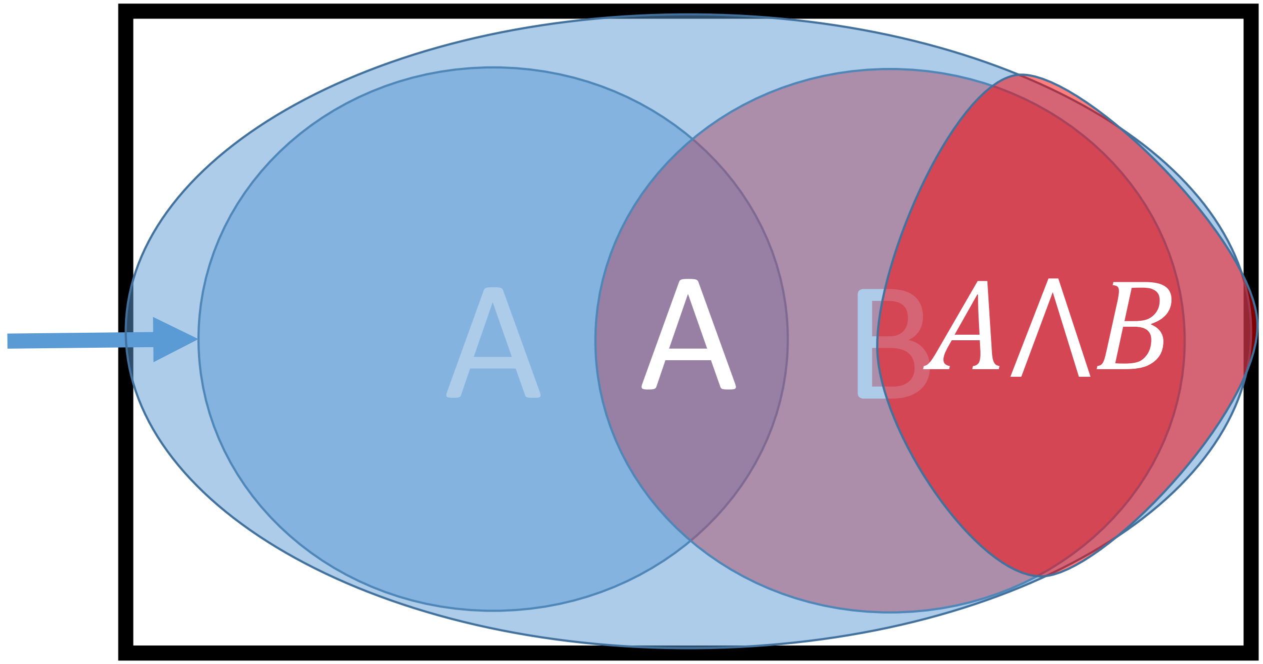
Suppose $P(B) = 0.2$

Suppose $P(A \wedge B) = 0.1$

Conditioning events change our knowledge!
For example, given that A is true…

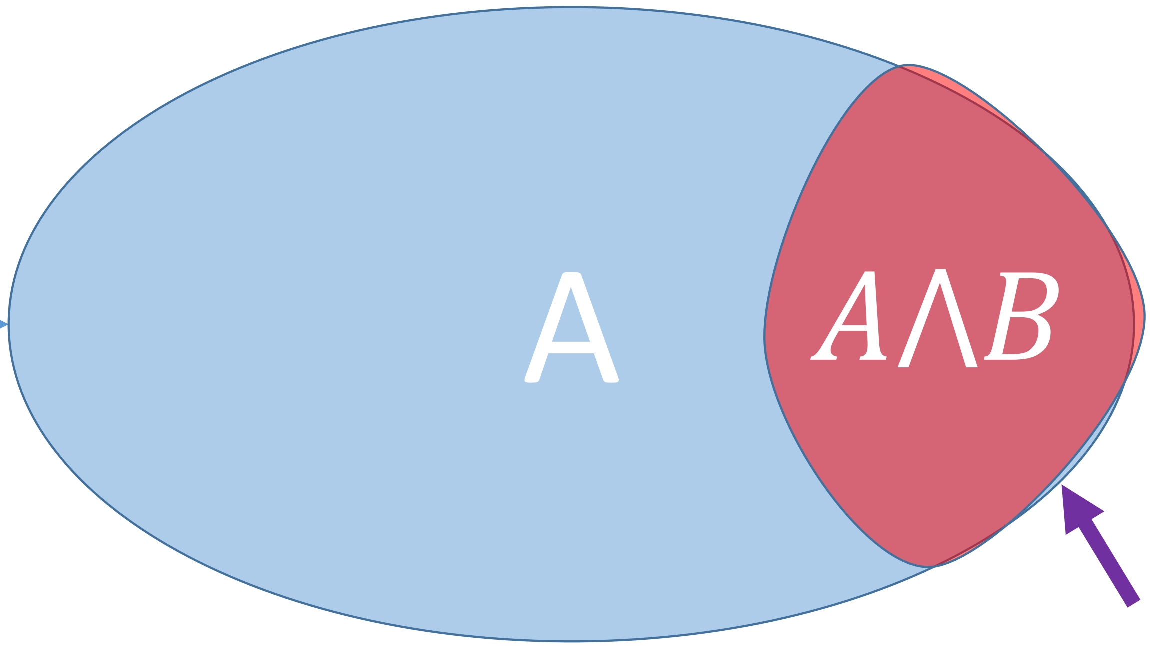
Most of the events in this rectangle are no longer possible!

Only the events inside this circle are now possible.



Conditioning events change our knowledge!
For example, given that A is true…

Given certain knowledge that A has occurred, we now have $P(A) = 1$.



Given certain knowledge that A has occurred, the probability that B also occurs is now

$$P(B|A) = \frac{P(A \land B)}{P(A)}$$

Joint and Conditional distributions of random variables

- $P(X, Y)$ is the **joint probability distribution** over all possible outcomes $P(X = x, Y = y)$.
- $P(X|Y)$ is the **conditional probability distribution** of outcomes $P(X = x|Y = y)$.

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

Joint and Conditional distributions of random variables

Bone die found at Cantonment Clinch. CC-BY-3.0, Colby Kirk, 2007.

Example:

X = number of pips on the bone die.

Y = X modulo 2.

The joint probability $\boldsymbol{P(X = 5, Y = 1) = \frac{1}{6}}$.

Their joint distribution is:

$P(X,Y) =$

$P(X = x, Y = y)$		x					
		1	**2**	**3**	**4**	**5**	**6**
y	**0**	0	$\frac{1}{6}$	0	$\frac{1}{6}$	0	$\frac{1}{6}$
	1	$\frac{1}{6}$	0	$\frac{1}{6}$	0	$\frac{1}{6}$	0

Joint and Conditional distributions of random variables

Bone die found at Cantonment Clinch. CC-BY-3.0, Colby Kirk, 2007.

Example:

X = number of pips on the bone die.

Y = X modulo 2.

The conditional probability $\boldsymbol{P(X = 5|Y = 1) = \frac{1}{3}}$.

Their joint distribution is:

$P(X|Y) =$

$P(X = x \mid Y = y)$		x					
		1	**2**	**3**	**4**	**5**	**6**
y	**0**	0	$\frac{1}{3}$	0	$\frac{1}{3}$	0	$\frac{1}{3}$
	1	$\frac{1}{3}$	0	$\frac{1}{3}$	0	$\frac{1}{3}$	0

Normalization trick

If you're given the joint probability distribution and want to find the conditional distribution, just renormalize so that each row sums to 1.

$P(X = x, Y = y)$		x					
		1	**2**	**3**	**4**	**5**	**6**
y	**0**	0	$\frac{1}{6}$	0	$\frac{1}{6}$	0	$\frac{1}{6}$
	1	$\frac{1}{6}$	0	$\frac{1}{6}$	0	$\frac{1}{6}$	0

$P(X = x \mid Y = y)$		x					
		1	**2**	**3**	**4**	**5**	**6**
y	**0**	0	$\frac{1}{3}$	0	$\frac{1}{3}$	0	$\frac{1}{3}$
	1	$\frac{1}{3}$	0	$\frac{1}{3}$	0	$\frac{1}{3}$	0

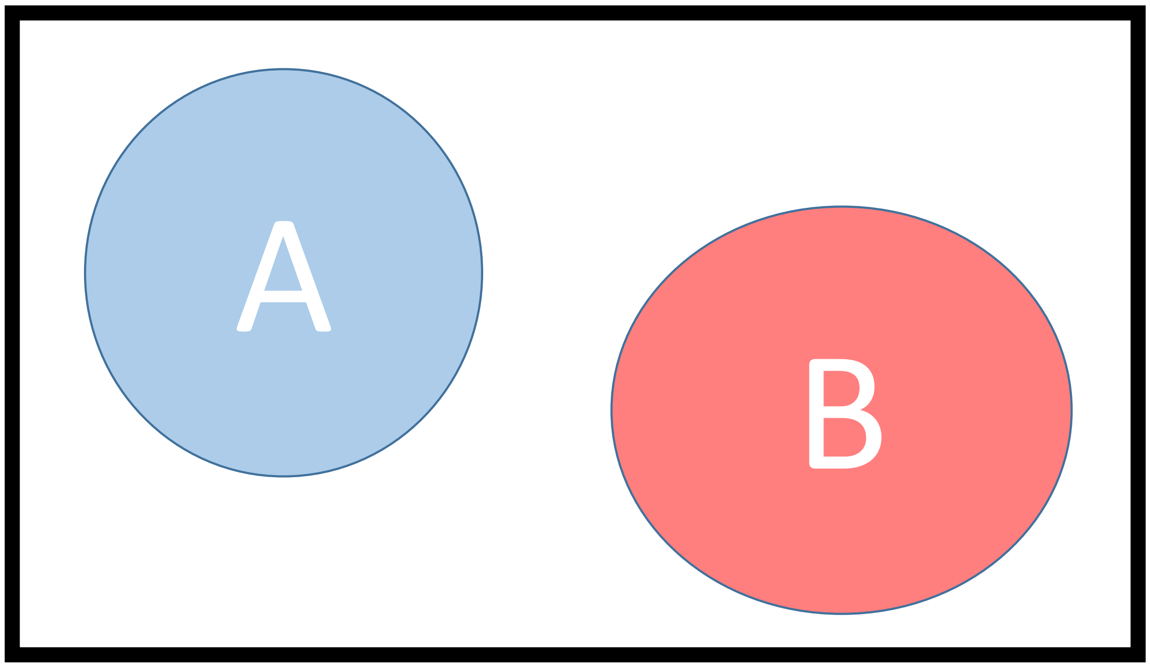
Outline

- Motivation: Why use probability?
- The axioms of probability
- Random variables
- Conditional probability
- Mutually exclusive vs. Independent vs. Conditionally Independent

Mutually exclusive events

Mutually exclusive events never occur simultaneously:

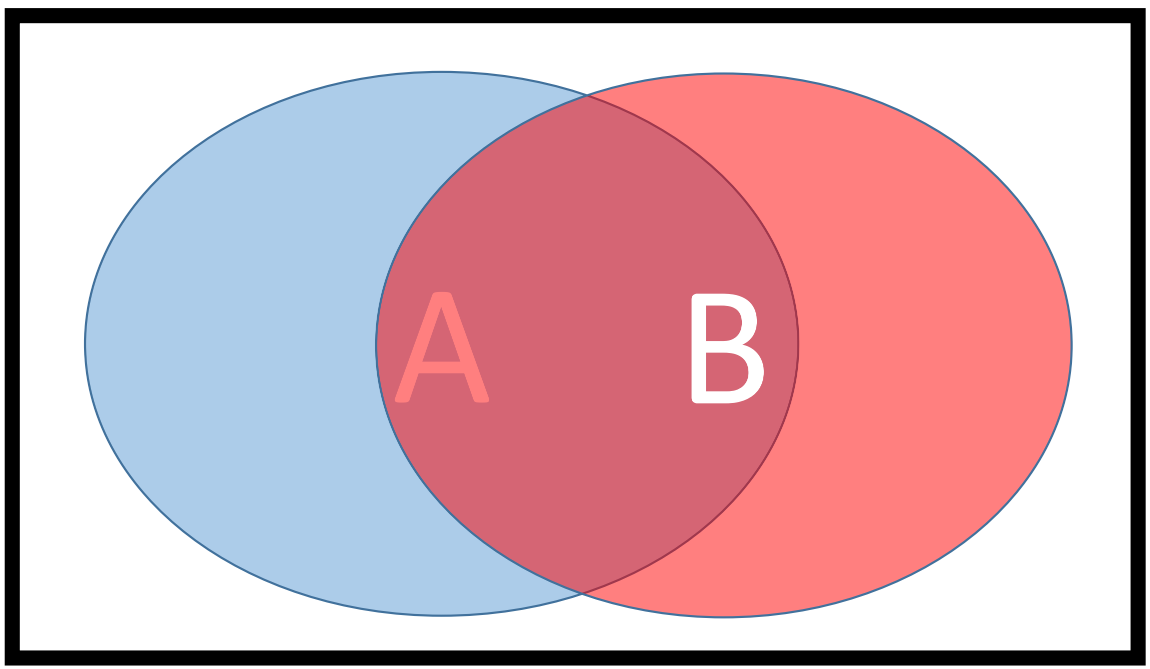
$$P(A \vee B) = P(A) + P(B) - P(A \wedge B) = P(A) + P(B)$$



Independent events

Independent events occur with equal probability, regardless of whether or not the other event has occurred:

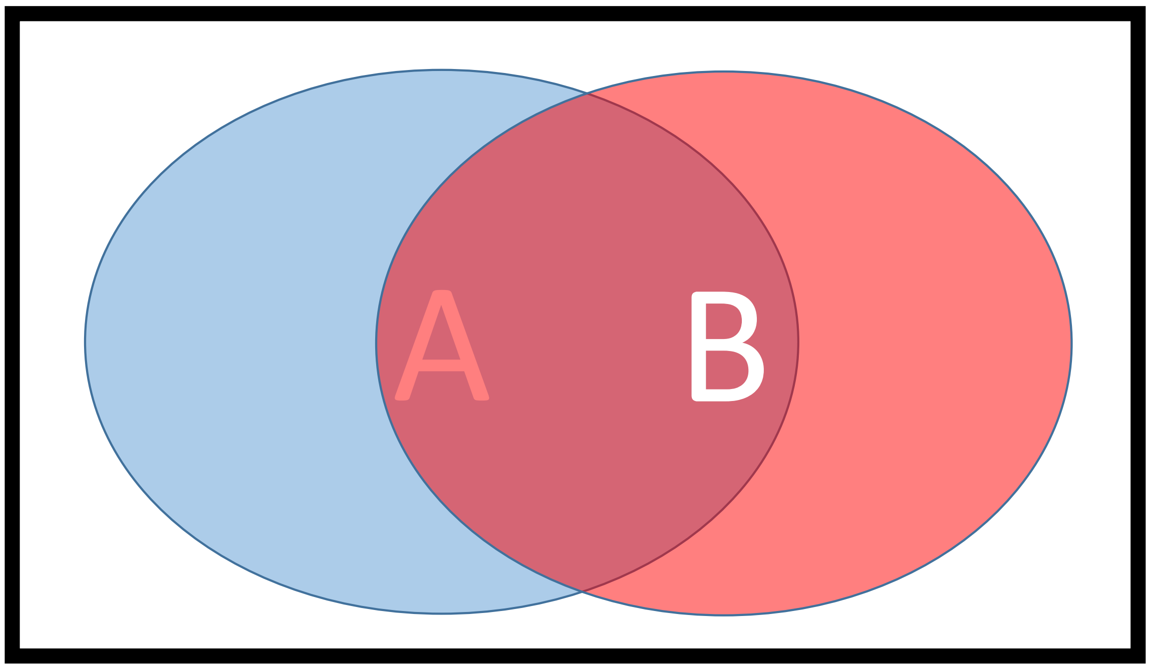
$$P(A|B) = \frac{P(A \wedge B)}{P(B)} = P(A), \qquad P(B|A) = \frac{P(A \wedge B)}{P(A)} = P(B)$$



Independent events: A more useful definition

Re-arranging terms in the previous slide gives us this more useful definition of independent events:

$$P(A \wedge B) = P(A)P(B)$$



Independent vs. Mutually Exclusive

- Independent events:

$$P(A \land B) = P(A)P(B)$$

- Mutually exclusive events:

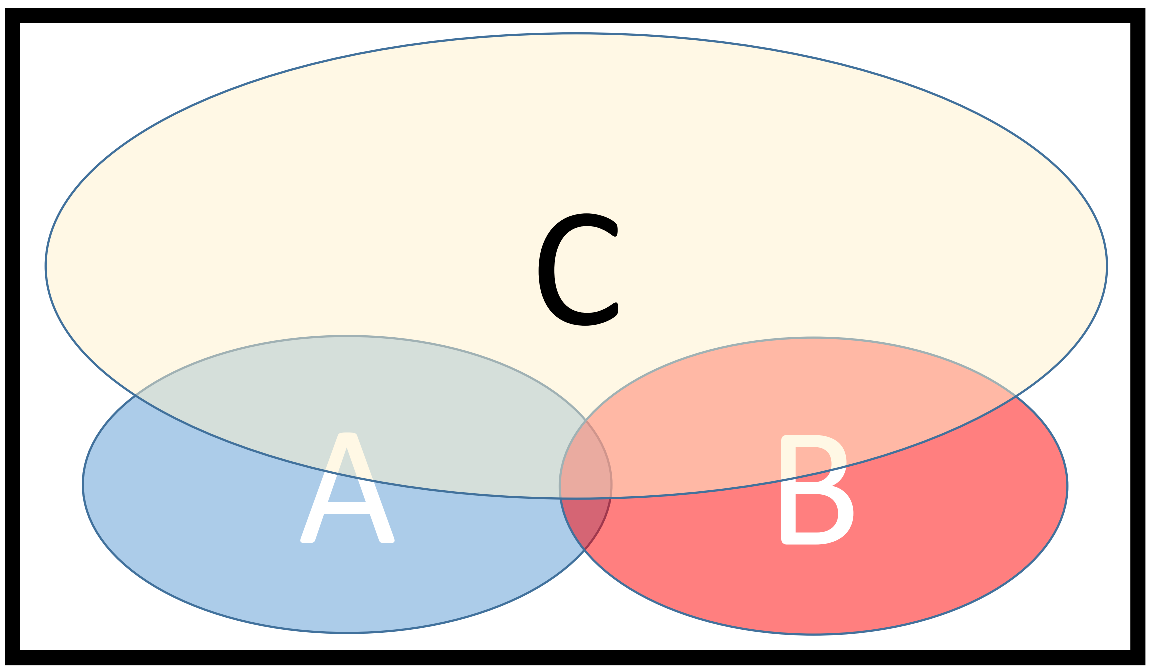
$$P(A \lor B) = P(A) + P(B)$$

Don't confuse them! Mutually exclusive events are not independent. Quite the contrary.

Conditionally independent events

Events A and B are conditionally independent, given C, if

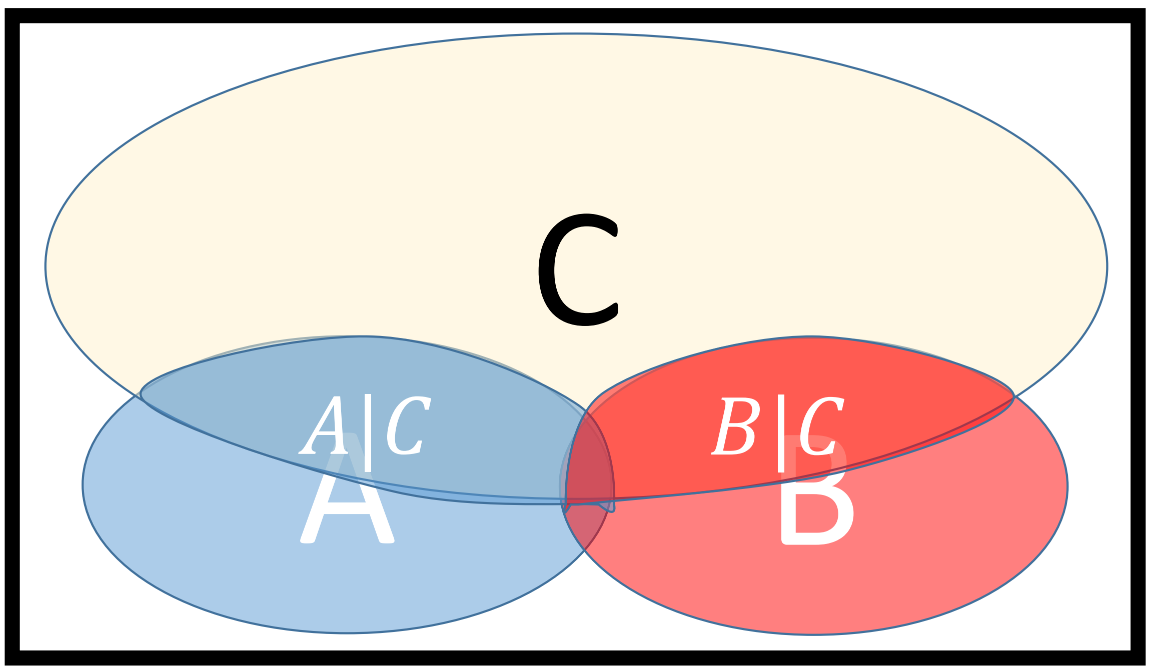
$$P(A|B, C) = P(A|C)$$



Conditionally independent events

Events A and B are conditionally independent, given C, if

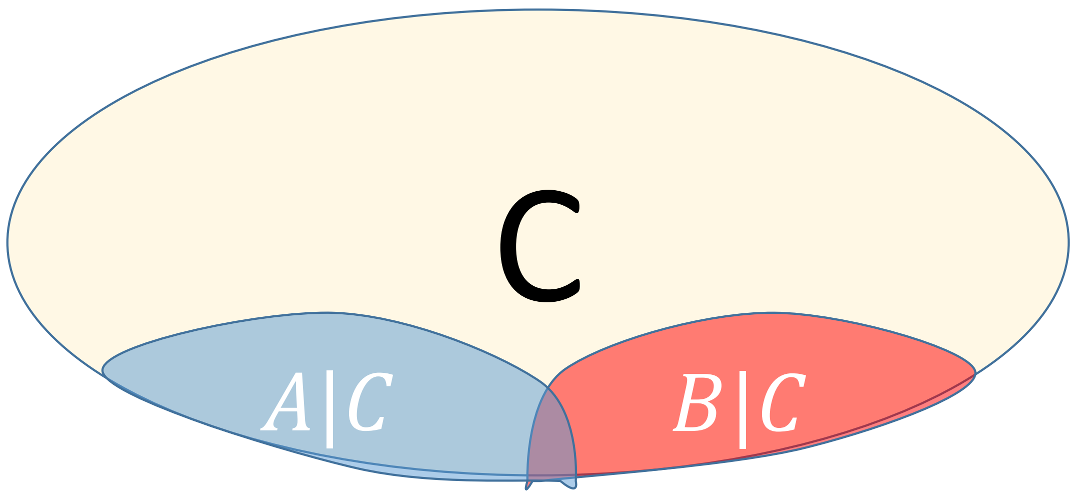
$$P(A|B,C) = \frac{P(A \wedge B|C)}{P(B|C)} = P(A|C)$$



Conditionally independent events

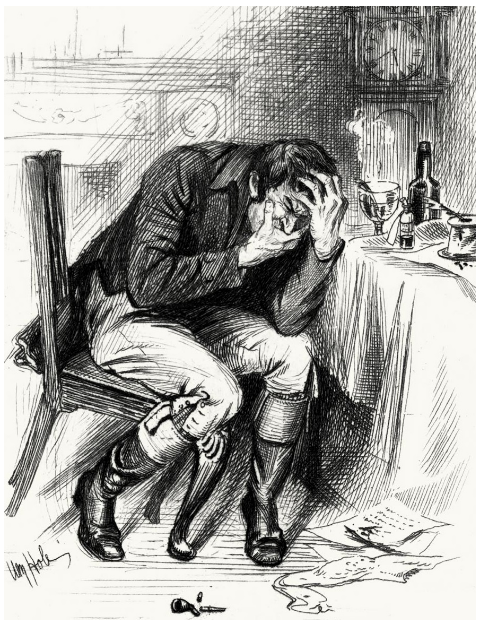
Events A and B are conditionally independent, given C, if

$$P(A, B|C) = P(A|C)P(B|C)$$



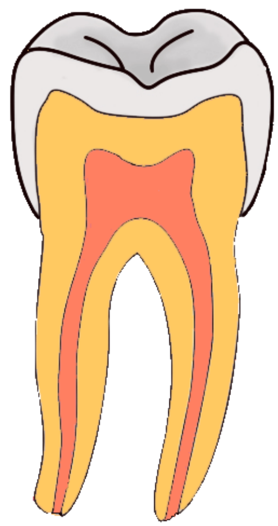
Independence ≠ Conditional Independence

Toothache= patient has a toothache



By William Brassey Hole(Died:1917)

Cavity= the patient has a cavity

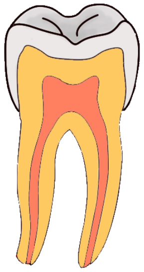


By Aduran, CC-SA 3.0

Catch= dentist's probe catches on something in the mouth

By Dozenist, CC-SA 3.0

These Events are not Independent



- If the patient has a toothache, then it's likely he has a cavity. Having a cavity makes it more likely that the probe will catch on something.

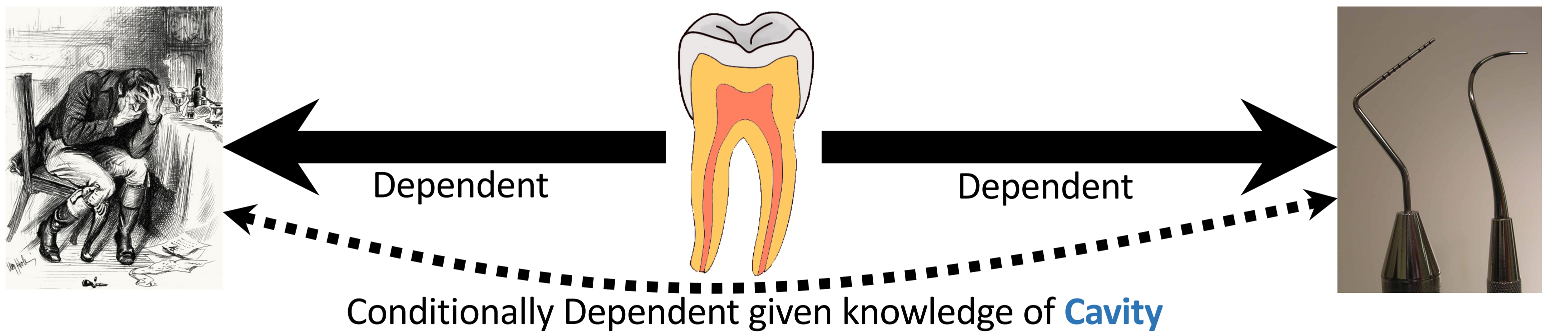
$$P(\text{Catch}|\text{Toothache}) > P(\text{Catch})$$

- If the probe catches on something, then it's likely that the patient has a cavity. If he has a cavity, then he might also have a toothache.

$$P(\text{Toothache}|\text{Catch}) > P(\text{Toothache})$$

- So Catch and Toothache are not independent

…but they are Conditionally Independent

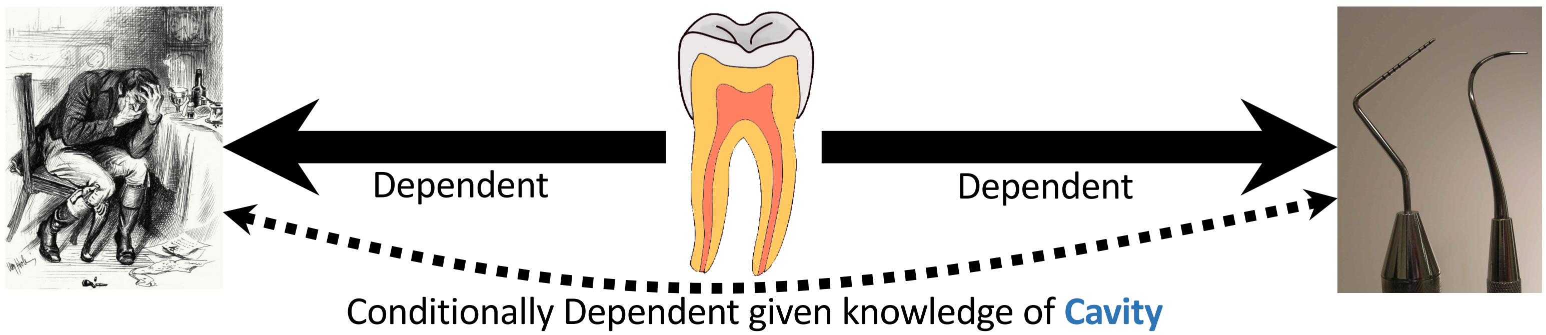


- Here are some reasons the probe might not catch, despite having a cavity:
  - The dentist might be really careless
  - The cavity might be really small
- Those reasons have nothing to do with the toothache!

$$P(\text{Catch}|\text{Cavity}, \text{Toothache}) = P(\text{Catch}|\text{Cavity})$$

- **Catch** and **Toothache** are **<u>conditionally independent</u>** given knowledge of **Cavity**

...but they are Conditionally Independent



These statements are all equivalent:

$$P(\text{Catch}|\text{Cavity}, \text{Toothache}) = P(\text{Catch}|\text{Cavity})$$

$$P(\text{Toothache}|\text{Cavity}, \text{Catch}) = P(\text{Toothache}|\text{Cavity})$$

$$P(\text{Toothache}, \text{Catch}|\text{Cavity}) = P(\text{Toothache}|\text{Cavity})\, P(\text{Catch}|\text{Cavity})$$

...and they all mean that **Catch** and **Toothache** are **conditionally independent** given knowledge of **Cavity**

Summary

Here's today's most important equation:

$$P(B|A) = \frac{P(A \wedge B)}{P(A)}$$

If you haven't seen this stuff since high school, read appendix A.3.