Bayes Theorem

What is Bayes’ Theorem?

Bayes’ Theorem is a way to calculate a conditional probability. It is a simple concept but it has lead to its own branch of statistics (Bayesian) and it is the basis for numerous ML models.

But *what is a conditional probability*? Let’s talk about conditional probabilities using an example.

Thomas Bayes

Example: Estimating Probability

Example Problem: Have you ever wondered how your phone can predict your next location?

Say your phone collects data at regular time intervals and finds the following:

$$P(work) = 40/120 = 33.3\%$$

$$P(home) = 80/120 = 66.7\%$$

Your phone might think, “Well, the probability that they are at home is higher so let’s always predict they are going home”. But can your phone do better than this?

Marginal & Joint Probability

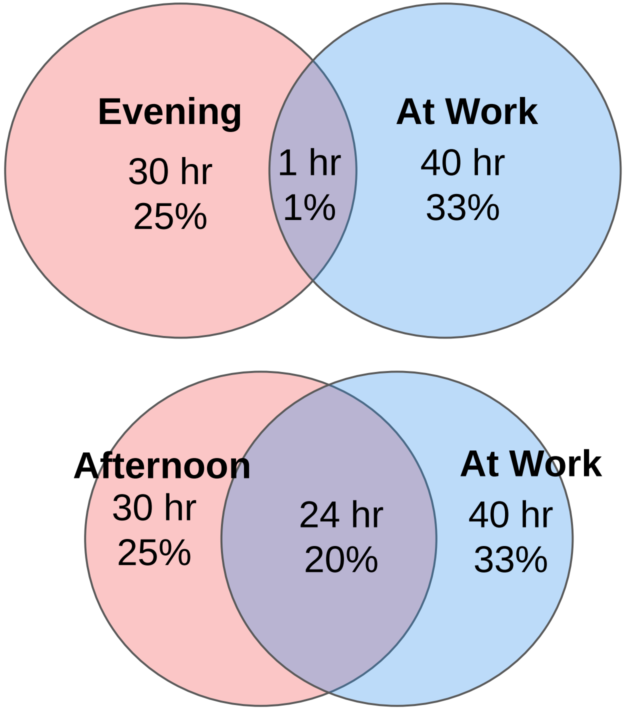
The **marginal probability** of A is denoted P(A) and gives the probability of event A occurring.

The **joint probability** of A and B is denoted P(A and B) or P(A∩B) and gives the probability of both events A and B occurring.

$$P(work\ and\ evening) = 1\%$$

$$P(work\ and\ afternoon) = 20\%$$

What if we consider the time of day?



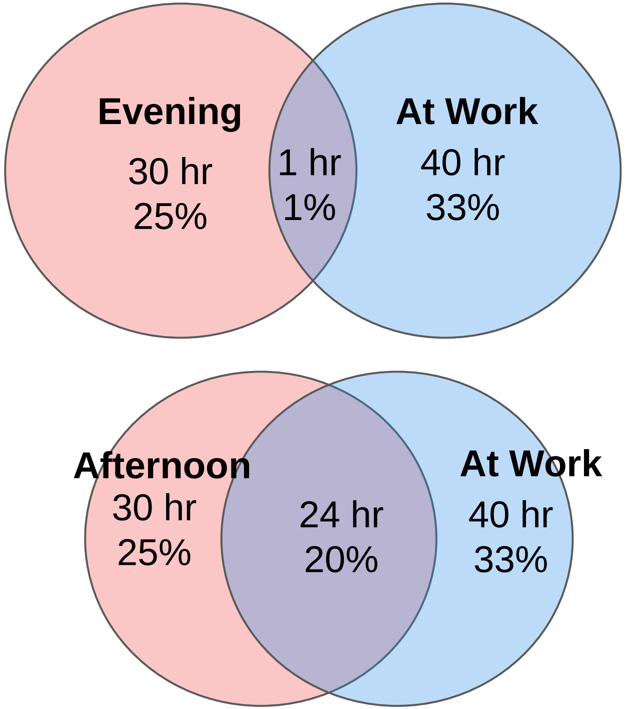
Conditional Probability

The **conditional probability** of A given B is denoted P(A|B) and gives the probability of event A occurring given event B has occurred.

$$P(A|B) = P(A \; and \; B)/P(B)$$

$$P(work|evening) = 1/25 = 4\%$$
$$P(work|afternoon) = 20/25 = 80\%$$

What if we consider the time of day?



Bayes' Theorem

So what role does Bayes' Theorem play in all of this?

$$P(A \; and \; B) = P(A) \cdot P(B)$$

If A and B are independent

$$P(A \; and \; B) = P(B|A) \cdot P(A)$$

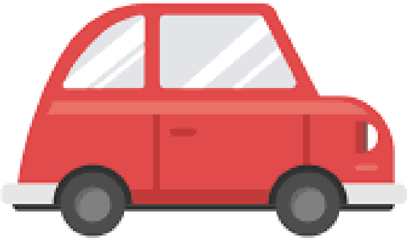
$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

The key to this is we can estimate P(A | B) from P(B | A). Bayes allows us to incorporate additional information to make better predictions.

Bayes Theorem Example

The police have a breathalyzer test that is able to identify a drunk person as being drunk 100% of the time. This test also incorrectly identifies a non-drunk person as being drunk 5% of the time. (Note - this is called the false positive rate). We also know that 1 out of 100 people on the road are driving drunk.

If someone gets pulled over and gets a positive breathalyzer, what is the probability that the person is drunk?



Bayes Theorem Example

Say 100 people are given the breathalyzer. We know that 1 out of the 100 will be drunk and that person will have a positive breathalyzer. We also know that 5% of the remaining 99 will get a positive breathalyzer even though they are not drunk.

1 Drunk Person Positive Breathalyzer

~5 Non Drunk People Positive Breathalyzer

The probability that someone is drunk given a positive breathalyzer is ~1/6 = 17%!

Bayes Theorem Example

Say 100 people are given the breathalyzer. We know that 1 out of the 100 will be drunk and that person will have a positive breathalyzer. We also know that 5% of the remaining 99 will get a positive breathalyzer even though they are not drunk.

1 Drunk Person Positive Breathalyzer

~5 Non Drunk People Positive Breathalyzer

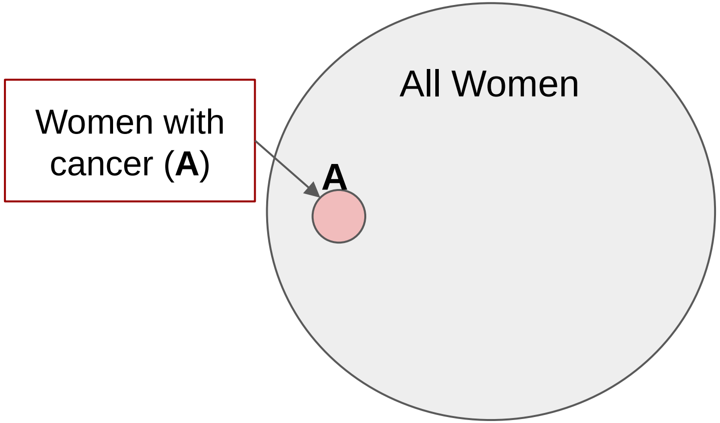
The probability that someone is drunk given a positive breathalyzer is ~1/6 = 17%!

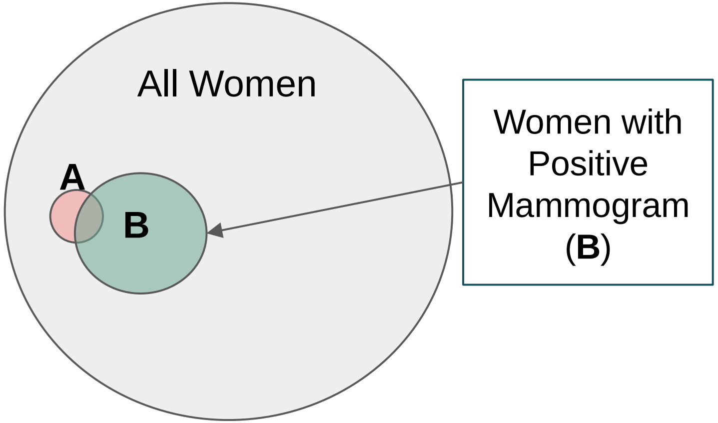
A = Person is Drunk
B = Positive Breathalyzer

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{100\% \cdot \frac{1}{100}}{\frac{1}{100} + \frac{5}{99}} \approx 17\%$$

Bayes' Theorem Example 2

1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammograms. 9.6% of women without breast cancer will also get positive mammograms. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?





Bayes' Theorem Example 2 Cont.

Let A be the event the woman has breast cancer, and B be the event a women gets a positive mammogram.

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

- 1% of women have breast cancer.

$$P(A) = 0.01$$

- 80% of those women get a positive mammogram, and 9.6% of the women without breast cancer get a positive mammogram too.

$$P(B) = 0.8P(A) + 0.096(1 - P(A))$$
$$P(B) = 0.008 + 0.09504$$
$$P(B) = 0.10304$$

Bayes' Theorem Example Cont.

- We can get P(B | A) straight from the problem statement. Remember that 80% of women with breast cancer get a positive mammogram.

$$P(B|A) = 0.80$$

- Now we can plug everything into Bayes' theorem.

$$P(A|B) = \frac{0.8 \cdot 0.01}{0.10304} = 0.0776$$

There is about a 7.8% chance of actually having breast cancer given a positive mammogram.