Conditional Probability

In English, a conditional probability states "what is the chance of an event E happening given that I have already observed some other event F". It is a critical idea in machine learning and probability because it allows us to update our probabilities in the face of new evidence.

When you condition on an event happening you are entering the universe where that event has taken place. Formally, once you condition on F the only outcomes that are now possible are the ones which are consistent with F. In other words your sample space will now be reduced to F. As an aside, in the universe where F has taken place, all rules of probability still hold!

Definition: Conditional Probability.

The probability of E given that (aka conditioned on) event F already happened:

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

Let's use a visualization to get an intuition for why the conditional probability formula is true. Again consider events E and F which have outcomes that are subsets of a sample space with 50 equally likely outcomes, each one drawn as a hexagon:
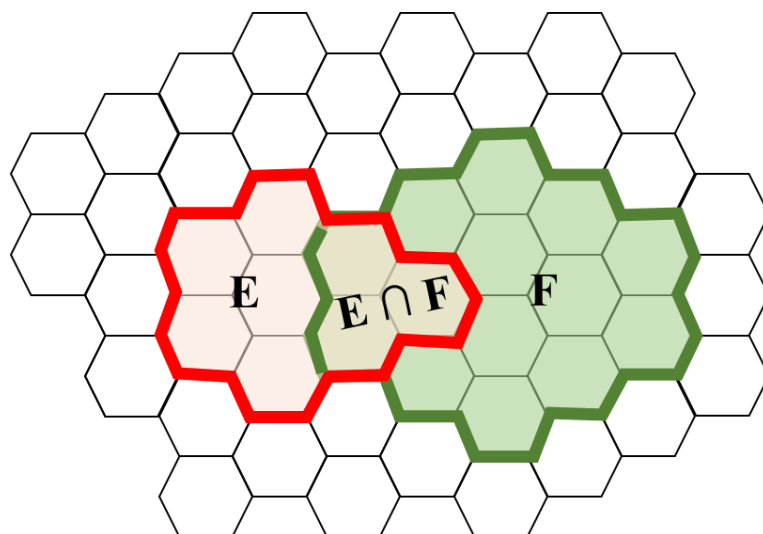
Conditioning on F means that we have entered the world where F has happened (and F, which has 14 equally likely outcomes, has become our new sample space). Given that event F has occurred, the conditional probability that event E occurs is the subset of the outcomes of E that are consistent with F. In this case we can visually see that those are the three outcomes in E and F. Thus we have the:

$$P(E|F) = \frac{P(E \text{ and } F)}{P(F)} = \frac{3/50}{14/50} = \frac{3}{14} \approx 0.21$$

Even though the visual example (with equally likely outcome spaces) is useful for gaining intuition, conditional probability applies regardless of whether the sample space has equally likely outcomes!
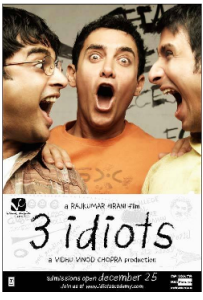
Conditional Probability Example

Let's use a real world example to better understand conditional probability: movie recommendation. Imagine a streaming service like Netflix wants to figure out the probability that a user will watch a movie E (for example, Life is Beautiful), based on knowing that they watched a different movie F (say Amélie). To start lets answer the simpler question, what is the probability that a user watches the movie Life is Beautiful, E? We can solve this problem using the definition of probability and a dataset of movie watching [1]:

$$P(E) = \lim_{n \to \infty} \frac{\text{count}(E)}{n} \approx \frac{\text{\# people who watched movie } E}{\text{\# people on Netflix}}$$
$$= \frac{1,234,231}{50,923,123} \approx 0.02$$

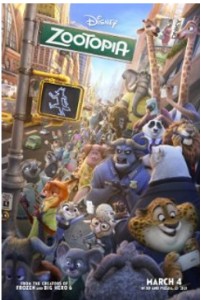In fact we can do this for many movies E:

|				
$P(E) = 0.02$	$P(E) = 0.01$	$P(E) = 0.05$	$P(E) = 0.09$	$P(E) = 0.03$

Now for a more interesting question. What is the probability that a user will watch the movie Life is Beautiful (E), given they watched Amelie (F)? We can use the definition of conditional probability.

$$P(E|F) = \frac{P(E \text{ and } F)}{P(F)} \qquad \text{Def of Cond Prob}$$
$$\approx \frac{(\text{\# who watched } E \text{ and } F)/(\text{\# of people on Netflix})}{(\text{\# who watched movie } F)/(\text{\# people on Netflix})} \qquad \text{Def of Prob}$$
$$\approx \frac{\text{\# of people who watched both } E \text{ and } F}{\text{\# of people who watched movie } F} \qquad \text{Simplifying}$$
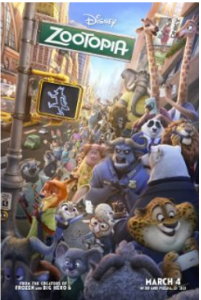
If we let F be the event that someone watches the movie Amélie, we can now calculate $P(E|F)$, the *conditional* probability that someone watches movie E:
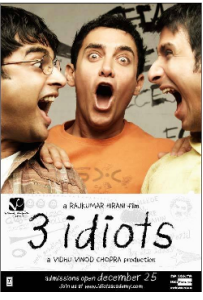
$P(E|F) = 0.09$	$P(E|F) = 0.03$	$P(E|F) = 0.05$	$P(E|F) = 0.02$	$P(E|F) = \mathbf{1.00}$

Why do some probabilities go up, some probabilities go down, and some probabilities are unchanged after we observe that the person has watched Amelie (F)? If you know someone watched Amelie, they are more likely to watch Life is Beautiful, and less likely to watch Star Wars. We have new information on the person!

The Conditional Paradigm

When you condition on an event you enter the universe where that event has taken place. In that new universe all the laws of probability still hold. Thus, as long as you condition consistently on the same event, every one of the tools we have learned still apply. Let's look at a few of our old friends when we condition consistently on an event (in this case G):

Name of Rule	Original Rule	Rule Conditioned on G
Axiom of probability 1	$0 \le P(E) \le 1$	$0 \le P(E|G) \le 1$
Axiom of probability 2	$P(S) = 1$	$P(S|G) = 1$
Axiom of probability 3	$P(E \text{ or } F) = P(E) + P(F)$ for mutually exclusive events	$P(E \text{ or } F|G) = P(E|G) + P(F|G)$ for mutually exclusive events
Identity 1	$P(E^C) = 1 - P(E)$	$P(E^C|G) = 1 - P(E|G)$

Conditioning on Multiple Events

The conditional paradigm also applies to the definition of conditional probability! Again if we consistently condition on some event G occurring, the rule still holds:

$$\mathrm{P}(E|F, G) = \frac{\mathrm{P}(E \text{ and } F|G)}{\mathrm{P}(F|G)}$$

The term $\mathrm{P}(E|F, G)$ is new notation for conditioning on multiple events. You should read that term as "The probability of E occurring, given that both F and G have occurred". This equation states that the definition for conditional probability of $E|F$ still applies in the universe where G has occurred. Do you think that $\mathrm{P}(E|F, G)$ should be equal to $\mathrm{P}(E|F)$? The answer is: sometimes yes and sometimes no.