3.3 Conditional Probability

Suppose we had a bag that contained two coins. One coin is a fair coin, and the other has a bias of 0.95, that is, if you flip this biased coin, it will come up heads with probability 0.95 and tails with probability 0.05. Holding the bag in one hand, you blindly reach in with your other, and pick out a coin. You flip this coin 3 times and see that all three times, the coin came up heads. You suspect that this coin is "likely" the biased coin, but how "likely" is it?

This problem highlights a typical situation in which new information changes the chance of an event happening. The original event was "we pick the biased coin". Before reaching in to grab a coin and then flipping it, we would reason that the probability of this event occurring (picking the biased coin) is $\frac{1}{2}$. After flipping the coin a couple of times and seeing that it landed heads all three times, we gain new information, and our probability should no longer be $\frac{1}{2}$. In fact, it should be much higher. In this case, we "condition" on the event of flipping 3 heads out of 3 total flips. We would write this new probability as

$$P(\text{picking the biased coin} \mid \text{flipping 3 heads out of 3 total flips})$$

The bar "|" between the two events in the probability expression above represents "conditioned on", and is defined below.

Definition 3.9. *The probability of an event A conditioned on an event B is denoted and defined*

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

The intuition of this definition can be gained by playing with the visualization on the website. Suppose we drop a ball uniformly at random in the visualization. If we ask "What is the probability that a ball hits the orange shelf?", we can compute this probability by simply dividing the length of the orange shelf by the length of the entire space. Now suppose we are given the information that our ball landed on the green shelf. What is the probability of landing on the orange shelf now? Our green shelf has become our "new" sample space, and the proportion of the green shelf that overlaps with the orange shelf is now the only region in which we could have possibly landed on the orange shelf. To compute this new conditional probability, we would divide the length of the overlapping, or "intersecting", regions of the orange and green shelves by the total length of the green shelf.

3.3.1 Bayes Rule

Now that we've understood where the definition of conditional probability comes from, we can use it to prove a useful identity.

Theorem 3.10 (Bayes Rule)**.** *Let A and B be two subsets of our sample space Ω. Then*

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

Proof. By the definition of conditional probability,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Similarly,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

Multiplying both sides by $P(A)$ gives

$$P(B \mid A)P(A) = P(A \cap B)$$

Plugging this into our first equation, we conclude

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

$\square$

3.3.2 Coins in a Bag

Let's return to our first example in this section and try to use our new theorem to find a solution. Define the events

$$\begin{aligned} A &\doteq \{\text{Picking the biased coin}\} \\ B &\doteq \{\text{Flipping 3 heads out of 3 total flips}\} \end{aligned}$$

We were interested in computing the probability $P(A \mid B)$. By Bayes Rule,

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

$P(B \mid A)$, i.e. the probability of flipping 3 heads out of 3 total flips given that we picked the biased coin, is simply $(0.95)^3 \approx 0.857$. The probability $P(A)$, i.e. the probability that we picked the biased coin, is $\frac{1}{2}$, since we blindly picked a coin from the bag. Now all we need to do is compute $P(B)$, the overall probability of flipping 3 heads in this experiment. Remember from the set theory section, we can write

$$B = B \cap \Omega = B \cap (A \cup A^c) = (B \cap A) \cup (B \cap A^c)$$

So

$$P(B) = P((B \cap A) \cup (B \cap A^c)) = P(B \cap A) + P(B \cap A^c)$$

since the two sets $B \cap A$ and $B \cap A^c$ are disjoint. By the definition of conditional probability, we can write the above expression as

$$= P(B \mid A)P(A) + P(B \mid A^c)P(A^c)$$

We just computed $P(B \mid A)$ and $P(A)$. Similarly, the probability that we flip 3 heads given that we *didn't* pick the biased coin, denoted $P(B \mid A^c)$, is the probability that we flip 3 heads given we picked the fair coin, which is simply $(\frac{1}{2})^3 = 0.125$. The event A^c represents the event in which A does not happen, i.e. the event that we pick the fair coin. We have $P(A^c) = 1 - P(A) = 1 - \frac{1}{2} = \frac{1}{2}$. Hence

$$\begin{aligned} P(B) &= P(B \mid A)P(A) + P(B \mid A^c)P(A^c) \\ &= 0.857 \cdot 0.5 + 0.125 \cdot 0.5 \\ &= 0.491 \end{aligned}$$

Plugging this back into the formula given by Bayes Rule,

$$P(A \mid B) = \frac{0.857 \cdot 0.5}{0.491} = 0.873$$

Thus, given that we flipped 3 heads out of a total 3 flips, the probability that we picked the biased coin is roughly 87.3%.

3.3.3 Conditional Poker Probabilities

Within a game of poker, there are many opportunities for us to flex our knowledge of conditional probability. For instance, the probability of drawing a full house is 0.0014, which is less than 2%. But suppose we draw three cards and find that we have already achieved a pair. Now the probability of drawing a full house is higher than 0.0014. How much higher you ask? With our new knowledge of conditional probability, this question is easy to answer. We define the events

$$\begin{aligned} A &\doteq \{\text{Drawing a Full House}\} \\ B &\doteq \{\text{Drawing a Pair within the first three cards}\} \end{aligned}$$

By Bayes Rule,

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

$P(B \mid A)$, i.e. the probability that we draw a pair within the first three cards given that we drew a full house eventually, is 1. This is because every grouping of three cards within a full house must contain a pair. From Section 3.2.3, the probability of drawing a full house is $P(A) = 0.0014$.

It remains to compute $P(B)$, the probability that we draw a pair within the first three cards. The total number of ways to choose 3 cards from 52 is $\binom{52}{3}$. The number of ways to choose 3 cards containing a pair is $\binom{13}{1}\binom{4}{2}\binom{50}{1}$ (since there are $\binom{13}{1}$ ways to choose the value of the pair, $\binom{4}{2}$ ways to pick which two suits of the chosen value make the pair, and $\binom{50}{1}$ ways to pick the last card from the remaining 50 cards). Hence the probability of the event B is

$$P(B) = \frac{\binom{13}{1}\binom{4}{2}\binom{50}{1}}{\binom{52}{3}} \approx 0.176$$

Plugging this into our formula from Bayes Rule,

$$P(A \mid B) = \frac{1 \cdot 0.0014}{0.176} \approx 0.00795$$

It follows that our chance of drawing a full house has more than quadrupled, increasing from less than 2% to almost 8%.