Compound Probability

Set Theory

A probability measure P is a function that maps subsets of the state space Ω to numbers in the interval $[0,1]$. In order to study these functions, we need to know some basic set theory.

Basic Definitions

Definition 0.0.5. *A **set** is a collection of items, or elements, with no repeats. Usually we write a set A using curly brackets and commas to distinguish elements, shown below*

$$A = \{a_0, a_1, a_2\}$$

*In this case, A is a set with three distinct elements: a_0, a_1, and a_2. The size of the set A is denoted $|A|$ and is called the **cardinality** of A. In this case, $|A| = 3$. The **empty set** is denoted $\varnothing$ and means*

$$\varnothing = \{\quad\}$$

Some essential set operations in probability are the intersection, union, and complement operators, denoted $\cap, \cup$, and ^{c}. They are defined below

Definition 0.0.6. ***Intersection** and **Union** each take two sets in as input, and output a single set. **Complementation** takes a single set in as input and outputs a single set. If A and B are subsets of our sample space Ω, then we write*

(a) $A \cap B = \{x \in \Omega : x \in A \text{ and } x \in B\}$.

(b) $A \cup B = \{x \in \Omega : x \in A \text{ or } x \in B\}$.

(c) $A^c = \{x \in \Omega : x \notin A\}$.

Another concept that we need to be familiar with is that of disjointness. For two sets to be disjoint, they must share no common elements, i.e. their intersection is empty.

Definition 0.0.7. *We say two sets A and B are **disjoint** if*

$$A \cap B = \varnothing$$

It turns out that if two sets A and B are disjoint, then we can write the probability of their union as

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

Set Algebra

There is a neat analogy between set algebra and regular algebra. Roughly speaking, when manipulating expressions of sets and set operations, we can see that " $\cup$ " acts like " $+$ " and " $\cap$ " acts like " $\times$ ". Taking the complement of a set corresponds to taking the negative of a number. This analogy isn't perfect, however. If we considered the union of a set A and its complement A^c, the analogy would imply that $A \cup A^c = \varnothing$, since a number plus its negative is 0. However, it is easily verified that $A \cup A^c = \Omega$ (Every element of the sample space is either in A or not in A.)

Although the analogy isn't perfect, it can still be used as a rule of thumb for manipulating expressions like $A \cap (B \cup C)$. The number expression analogy to this set expression is $a \times (b + c)$. Hence we could write it

$$a \times (b + c) = a \times b + a \times c$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

The second set equality is true. Remember that what we just did was not a proof, but rather a non-rigorous rule of thumb to keep in mind. We still need to actually prove this expression.

Exercise 0.0.3. *Show that* $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Proof. To show set equality, we can show that the sets are contained in each other. This is usually done in two steps.

Step 1: "$\subset$". First we will show that $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$.

Select an arbitrary element in $A \cap (B \cup C)$, denoted ω. Then by definition of intersection, $\omega \in A$ and $\omega \in (B \cup C)$. By definition of union, $\omega \in (B \cup C)$ means that $\omega \in B$ or $\omega \in C$. If $\omega \in B$, then since ω is also in A, we must have $\omega \in A \cap B$. If $\omega \in C$, then since ω is also in A, we must have $\omega \in A \cap C$. Thus we must have either

$$\omega \in A \cap B \text{ or } \omega \in A \cap C$$

Hence, $\omega \in (A \cap B) \cup (A \cap C)$. Since ω was arbitrary, this shows that any element of $A \cap (B \cup C)$ is also an element of $(A \cap B) \cup (A \cap C)$.

Thus we have shown

$$A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$$

Step 2: " $\supset$ ". Next we will show that $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$.

Select an arbitrary element in $(A \cap B) \cup (A \cap C)$, denoted ω. Then $\omega \in (A \cap B)$ or $\omega \in (A \cap C)$. If $\omega \in A \cap B$, then $\omega \in B$. If $\omega \in A \cap C$, then $\omega \in C$. Thus ω is in either B or C, so $\omega \in B \cup C$. In either case, ω is also in A. Hence $\omega \in A \cap (B \cup C)$. Thus we have shown

$$(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$$

Since we have shown that these sets are included in each other, they must be equal. This completes the proof. □

On the website, plug in each of the sets $(A \cap B) \cup (A \cap C)$ and $A \cap (B \cup C)$. Observe that the highlighted region doesn't change, since the sets are the same!

DeMorgan's Laws

In this section, we will show two important set identities useful for manipulating expressions of sets. These rules known as DeMorgan's Laws.

Theorem 0.0.2 (DeMorgan's Laws)**.** *Let A and B be subsets of our sample space Ω. Then*

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

(b) $(A \cap B)^c = A^c \cup B^c$.

Proof.

(a) We will show that $(A \cup B)^c$ and $A^c \cap B^c$ are contained within each other.

   Step 1: "$\subset$". Suppose $\omega \in (A \cup B)^c$. Then ω is not in the set $A \cup B$, i.e. in neither A nor B. Then $\omega \in A^c$ and $\omega \in B^c$, so $\omega \in A^c \cap B^c$. Hence $(A \cup B)^c \subset A^c \cap B^c$.

   Step 2: "$\supset$". Suppose $\omega \in A^c \cap B^c$. Then ω is not in A and ω is not in B. So ω is in neither A nor B. This means ω is not in the set $(A \cup B)$, so $\omega \in (A \cup B)^c$. Hence $A^c \cap B^c \subset (A \cup B)^c$.

   Since $A^c \cap B^c$ and $(A \cup B)^c$ are subsets of each other, they must be equal.

(b) Left as an exercise.

□

If you're looking for more exercises, there is a link on the Set Theory page on the website that links to a page with many set identities. Try to prove some of these by showing that the sets are subsets of each other, or just plug them into the website to visualize them and see that their highlighted regions are the same.

Combinatorics

In many problems, to find the probability of an event, we will have to count the number of outcomes in Ω which satisfy the event, and divide by $|\Omega|$, i.e. the total number of outcomes in Ω. For example, to find the probability that a single die roll is even, we count the total number of even rolls, which is 3, and divide by the total number of rolls, 6. This gives a probability of $\frac{1}{2}$. But what if the event isn't as simple as "roll an even number"? For example if we flipped 10 coins, our event could be "flipped 3 heads total". How could we count the number of outcomes that have 3 heads in them without listing them all out? In this section, we will discover how to count the outcomes of such an event, and generalize the solution to be able to conquer even more complex problems.

Permutations

Suppose there are 3 students waiting in line to buy a spicy chicken sandwich. A question we could ask is, "How many ways can we order the students in this line?" Since there are so few students, let's just list out all possible orderings. We could have any of

$$\text{6 of these} \begin{cases} (1,2,3) \\ (1,3,2) \\ (2,1,3) \\ (2,3,1) \\ (3,1,2) \\ (3,2,1) \end{cases}$$

So there are 6 total possible orderings. If you look closely at the list above, you can see that there was a systematic way of listing them. We first wrote out all orderings starting with 1. Then came the orderings starting with 2, and then the ones that started with 3. In each of these groups of orderings starting with some particular student, there were two orderings. This is because once we fixed the first person in line, there were two ways to order the remaining two students. Denote N_i to be the number of ways to order i students. Now we observe that the number of orderings can be written

$$N_3 = 3 \cdot N_2$$

since there are 3 ways to pick the first student, and N_2 ways to order the remaining two students. By similar reasoning,

$$N_2 = 2 \cdot N_1$$

Since the number of ways to order 1 person is just 1, we have $N_1 = 1$. Hence,

$$N_3 = 3 \cdot N_2 = 3 \cdot (2 \cdot N_1) = 3 \cdot 2 \cdot 1 = 6$$

which is the same as what we got when we just listed out all the orderings and counted them.

Now suppose we want to count the number of orderings for 10 students. 10 is big enough that we can no longer just list out all possible orderings and count them. Instead, we will make use of our method above. The number of ways to order 10 students is

$$N_{10} = 10 \cdot N_9 = 10 \cdot (9 \cdot N_8) = \cdots = 10 \cdot 9 \cdot 8 \cdot 7 \cdot ... \cdot 2 \cdot 1 = 3,628,800$$

It would have nearly impossible for us to list out over 3 million orderings of 10 students, but we were still able to count these orderings using our neat trick. We have a special name for this operation.

Definition 0.0.8. *The number of **permutations**, or orderings, of n distinct objects is given by the **factorial** expression,*

$$n! = n \cdot (n-1) \cdot ... \cdot 2 \cdot 1$$

The factorial symbol is an exclamation point, which is used to indicate the excitement of counting.

Combinations

Now that we've established a quick method of counting the number of ways to order n distinct objects, let's figure out how to do our original problem. At the start of this section we asked how to count the number of ways we could flip 10 coins and have 3 of them be heads. The valid outcomes include

$$(\mathtt{H}, \mathtt{H}, \mathtt{H}, \mathtt{T}, \mathtt{T}, \mathtt{T}, \mathtt{T}, \mathtt{T}, \mathtt{T}, \mathtt{T})$$
$$(\mathtt{H}, \mathtt{H}, \mathtt{T}, \mathtt{H}, \mathtt{T}, \mathtt{T}, \mathtt{T}, \mathtt{T}, \mathtt{T}, \mathtt{T})$$
$$(\mathtt{H}, \mathtt{H}, \mathtt{T}, \mathtt{T}, \mathtt{H}, \mathtt{T}, \mathtt{T}, \mathtt{T}, \mathtt{T}, \mathtt{T})$$
$$\vdots$$

But its not immediately clear how to count all of these, and it definitely isn't worth listing them all out. Instead let's apply the permutations trick we learned in Section 3.2.2.

Suppose we have 10 coins, 3 of which are heads up, the remaining 7 of which are tails up. Label the 3 heads as coins 1, 2, and 3. Label the 7 tails as coins 4,5,6,7,8,9, and 10. There are 10! ways to order, or permute, these 10 (now distinct) coins. However, many of these permutations correspond to the same string of H's and T's. For example, coins 7 and 8 are both tails, so we would be counting the two permutations

$$(1, 2, 3, 4, 5, 6, \mathbf{7}, \mathbf{8}, 9, 10)$$
$$(1, 2, 3, 4, 5, 6, \mathbf{8}, \mathbf{7}, 9, 10)$$

as different, when they both correspond to the outcome

$$(\texttt{H}, \texttt{H}, \texttt{H}, \texttt{T}, \texttt{T}, \texttt{T}, \texttt{T}, \texttt{T}, \texttt{T}, \texttt{T})$$

hence we are *over counting* by just taking the factorial of 10. In fact, for the string above, we could permute the last 7 coins in the string (all tails) in 7! ways, and we would still get the same string, since they are all tails. To any particular permutation of these last 7 coins, we could permute the first 3 coins in the string (all heads) in 3! ways and still end up with the string

$$(\texttt{H}, \texttt{H}, \texttt{H}, \texttt{T}, \texttt{T}, \texttt{T}, \texttt{T}, \texttt{T}, \texttt{T}, \texttt{T})$$

This means that to each string of H's and T's, we can rearrange the coins in $3! \cdot 7!$ ways without changing the actual grouping of H's and T's in the string. So if there are 10! total ways of ordering the labeled coins, we are counting each unique grouping of heads and tails $3! \cdot 7!$ times, when we should only be counting it once. Dividing the total number of permutations by the factor by which we over count each unique grouping of heads and tails, we find that the number of unique groupings of H's and T's is

$$\text{\# of outcomes with 3 heads and 7 tails} = \frac{10!}{3!7!}$$

This leads us to the definition of the binomial coefficient.

Definition 0.0.9. *The* ***binomial coefficient*** *is defined*

$$\binom{n}{k} \doteq \frac{n!}{k!(n-k)!}$$

The binomial coefficient, denoted $\binom{n}{k}$, represents the number of ways to pick k objects from n objects where the ordering within the chosen k objects doesn't matter. In the previous example, $n = 10$ and

$k = 3$. We could rephrase the question as, "How many ways can we pick 3 of our 10 coins to be heads?" The answer is then

$$\binom{n}{k} = \binom{10}{3} = \frac{10!}{3!(10-3)!} = \frac{10!}{3!7!} = 120$$

We read the expression $\binom{n}{k}$ as "n choose k". Let's now apply this counting trick to make some money.

Poker

One application of counting includes computing probabilities of poker hands. A poker hand consists of 5 cards drawn from the deck. The order in which we receive these 5 cards is irrelevant. The number of possible hands is thus

$$\binom{52}{5} = \frac{52!}{5!(52-5)!} = 2,598,960$$

since there are 52 cards to choose 5 cards from.

In poker, there are types of hands that are regarded as valuable in the following order form most to least valuable.

1. Royal Flush: A, K, Q, J, 10 all in the same suit.
2. Straight Flush: Five cards in a sequence, all in the same suit.
3. Four of a Kind: Exactly what it sounds like.
4. Full House: 3 of a kind with a pair.
5. Flush: Any 5 cards of the same suit, but not in sequence.
6. Straight: Any 5 cards in sequence, but not all in the same suit.
7. Three of a Kind: Exactly what it sounds like.
8. Two Pair: Two pairs of cards.
9. One Pair: One pair of cards.
10. High Card: Anything else.

Let's compute the probability of drawing some of these hands.

Exercise 0.0.4. *Compute the probabilities of the above hands.*

Solution.

1. There are only 4 ways to get this hand. Either we get the royal cards in diamonds, clubs, hearts, or spades. We can think of this has choosing 1 suit from 4 possible suits. Hence the probability of this hand is

$$P(\text{Royal Flush}) = \frac{\binom{4}{1}}{\binom{52}{5}} \approx 1.5 \cdot 10^{-6}$$

2. Assuming hands like K, A, 2, 3, 4 don't count as consecutive, there are in total 10 valid consecutive sequences of 5 cards (each starts with any of A,2,...,10). We need to pick 1 of 10 starting values, and for each choice of a starting value, we can pick 1 of 4 suits to have them all in. This gives a total of $\binom{10}{1} \cdot \binom{4}{1} = 40$ straight flushes. However, we need to subtract out the probability of a royal flush, since one of the ten starting values we counted was 10 (10, J, Q, K, A is a royal flush). Hence the probability of this hand is

$$P(\text{Straight Flush}) = \frac{\binom{10}{1}\binom{4}{1} - \binom{4}{1}}{\binom{52}{5}} \approx 1.5 \cdot 10^{-5}$$

3. There are 13 values and only one way to get 4 of a kind for any particular value. However, for each of these ways to get 4 of a kind, the fifth card in the hand can be any of the remaining 48 cards. Formulating this in terms of our choose function, there are $\binom{13}{1}$ ways to choose the value, $\binom{12}{1}$ ways to choose the fifth card's value, and $\binom{4}{1}$ ways to choose the suit of the fifth card. Hence the probability of such a hand is

$$P(\text{Four of a Kind}) = \frac{\binom{13}{1}\binom{12}{1}\binom{4}{1}}{\binom{52}{5}} \approx 0.00024$$

4. For the full house, there are $\binom{13}{1}$ ways to pick the value of the triple, $\binom{4}{3}$ ways to choose which 3 of the 4 suits to include in the triple, $\binom{12}{1}$ ways to pick the value of the double, and $\binom{4}{2}$ ways to choose which 2 of the 4 suits to include in the double. Hence the probability of this hand is

$$P(\text{Full House}) = \frac{\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}}{\binom{52}{5}} \approx 0.0014$$

5. through 10. are left as exercises. The answers can be checked on the Wikipedia page titled "Poker probability".

□

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 likelihood of an event. 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 0.0.10. *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.

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 0.0.3 (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

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

□

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

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

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%.

Conditional Poker Probabilities

Within a game of poker, there are many opportunities 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 $\binom{13}{1}\binom{4}{2}\binom{50}{1}$. There are $\binom{13}{1}$ 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%.