Chapter 2. Discrete Probability

2.1: Intro to Discrete Probability

Slides (Google Drive) Alex Tsun Video (YouTube)

We're just about to learn about the axioms (rules) of probability, and see how all that counting stuff from chapter 1 was relevant at all. This should align with your current understanding of probability (I only assume you might be able to tell me the probability I roll an even number on a fair six-sided die at this point), and formalize it.

We'll be using a lot of set theory from here on out, so review that in Chapter 0 if you need to!

2.1.1 Definitions

Definition 2.1.1: Sample Space

The **sample space** is the set Ω of all possible outcomes of an experiment.

Example(s)

Find the sample space of...

1. a single coin flip.
2. two coin flips.
3. the roll of a fair 6-sided die.

Solution

1. The sample space of a single coin flip is: $\Omega = \{H, T\}$ (heads or tails).
2. The sample space of two coin flips is: $\Omega = \{HH, HT, TH, TT\}$.
3. The sample space of the roll of a die is: $\Omega = \{1, 2, 3, 4, 5, 6\}$.

$\square$

Definition 2.1.2: Event

An **event** is any subset $E \subseteq \Omega$.

Example(s)

List out the set of outcomes for the following events:

1. Getting at least one head in two coin flips.

2. Rolling an even number on a fair 6-sided die.

Solution

1. Getting at least one head in two coin flips: $E = \{HH, HT, TH\}$

2. Rolling an even number: $E = \{2, 4, 6\}$

□

Definition 2.1.3: Mutual Exclusion

Events E and F are mutually exclusive if $E \cap F = \emptyset$. (i.e. they can't simultaneously happen).

Example(s)

Say E is the event of rolling an even number: $E = \{2, 4, 6\}$, and F is the event of rolling an odd number: $F = \{1, 3, 5\}$. Are E and F mutually exclusive?

Solution E and F are mutually exclusive because $E \cap F = \emptyset$. □

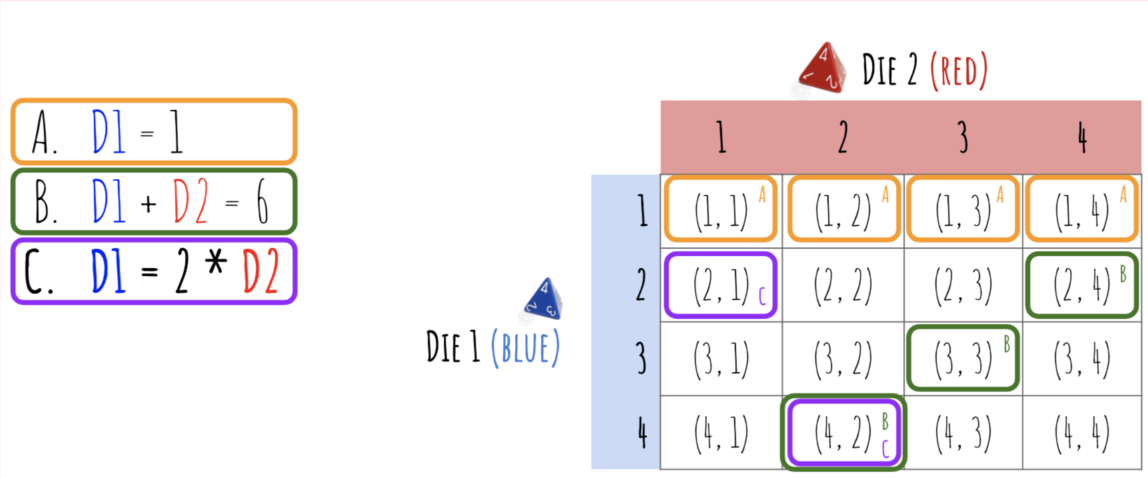
Example(s)

Let's consider another example in which our experiment is the rolling of two fair 4-sided dice, one which is blue $D1$ and one which is red $D2$ (so they are distinguishable, or effectively, order matters). We can represent each element in the sample set as an ordered pair $(D1, D2)$ where $D1, D2 \in \{1, 2, 3, 4\}$ and represent the respective value rolled by the blue and red die.

The sample space Ω is the set of all possible ordered pairs of values that could be rolled by the die ($|\Omega| = 4 \cdot 4 = 16$ by the product rule). Let's consider some events:

1. $A = \{(1, 1), (1, 2), (1, 3), (1, 4)\}$, the event that the blue die, $D1$ is a 1.
2. $B = \{(2, 4), (3, 3), (4, 2)\}$, the event that the sum of the two rolls is 6 ($D1 + D2 = 6$).
3. $C = \{(2, 1), (4, 2)\}$, the event that the value on the blue die is twice the value on the red die ($D1 = 2 * D2$).

All of these events and the sample space are shown below:



Are A and B mutually exclusive? Are B and C mutually exclusive?

Solution Now, let's consider whether A and B are mutually exclusive. Well, they do not overlap, as we can see that $A \cap B = \emptyset$, so yes they are mutually exclusive.

B and C are not mutually exclusive, since there is a case in which they can happen at the same time $B \cap C = \{(4, 2)\} \neq \emptyset$, so they are not mutually exclusive. □

Again, to summarize, we learned that Ω was the sample space (set of all outcomes of an experiment), and $E \subseteq \Omega$ is just a subset of outcomes we are interested in.

2.1.2 Axioms of Probability and their Consequences

Definition 2.1.4: Axioms of Probability and their Consequences

Let Ω denote the sample space and $E, F \subseteq \Omega$ be events.
Axioms:

1. (Nonnegativity) $\mathbb{P}(E) \geq 0$; that is, no event has a negative probability.

2. (Normalization) $\mathbb{P}(\Omega) = 1$; that is, the probability of the entire sample space is always 1 (something is guaranteed to happen)

3. (Countable Additivity) If E and F are *mutually exclusive*, then $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F)$. This actually holds for any countable (finite or countably infinite) collection of pairwise mutually exclusive events $E_1, E_2, E_3, \ldots$.

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)$$

The word "axiom" means: things that we take for granted and assume to be true **without proof**.
Corollaries:

1. (Complementation) $\mathbb{P}\left(E^C\right) = 1 - \mathbb{P}(E)$.

2. (Monotonicity) If $E \subseteq F$, then $\mathbb{P}(E) \leq \mathbb{P}(F)$.
3. (Inclusion-Exclusion) $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$.

The world "corollary" means: results that follow almost immediately from a previous result (in this case, the axioms).

Explanation of Axioms

1. Non-negativity is simply because we cannot consider an event to have a negative probability. It just wouldn't make sense. A probability of 1/6 would mean that on average, something would happen 1 out of every 6 trials. What about a probability of $-1/4$?
2. Normalization is based on the fact that when we run an experiment, there must be *some* outcome, and all possible outcomes are in the sample space. So, we say the probability of observing some outcome from the sample space is 1.
3. Countable additivity is because if two events are mutually exclusive, they don't overlap at all; that is, they don't share any outcomes. This means that the union of them will contain the same outcomes of each together, so the probability of their union is the the sum of their individual probabilities. (This is like the sum role of counting).

Explanation of Corollaries

1. Complementation is based on the fact that the sample space is all the possible outcomes. This means that $E^C = \Omega \setminus E$, so $\mathbb{P}\left(E^C\right) = 1 - \mathbb{P}(E)$. (This is like complementary counting).
2. Monotonocity is because if E is a subset of F, then all outcomes in the event E are in the event F. This means that all the outcomes that contribute to the probability of E contribute to the probability of F, so it's probability is greater than or equal to that of E (since probabilities are non-negative).
3. Inclusion-Exclusion follows because if E and F have some intersection, this would be counted twice by adding their probabilities, so we have to subtract it once to only count it once and not overcount. (This is like inclusion-exclusion for counting).

Proof of Corollaries. The proofs of these corollaries only depend on the 3 axioms which we assume to be true.

1. Since E and $E^C = \Omega \setminus E$ are mutually exclusive,

$$\begin{aligned}
\mathbb{P}(E) + \mathbb{P}\left(E^C\right) &= \mathbb{P}\left(E \cup E^C\right) && \text{[axiom 3]} \\
&= \mathbb{P}(\Omega) && [E \cup E^C = \Omega] \\
&= 1 && \text{[axiom 2]}
\end{aligned}$$

   Now just subtract $\mathbb{P}(E)$ from both sides.

2. Since $E \subseteq F$, consider the sets E and $F \setminus E$. Then,

$$\begin{aligned}
\mathbb{P}(F) &= \mathbb{P}(E \cup (F \setminus E)) && \text{[draw a picture of E inside event F]} \\
&= \mathbb{P}(E) + \mathbb{P}(F \setminus E) && \text{[mutually exclusive, axiom 3]} \\
&\geq \mathbb{P}(E) + 0 && [\text{since } \mathbb{P}(F \setminus E) \geq 0 \text{ by axiom 1}]
\end{aligned}$$

3. Left to the reader. Hint: Draw a picture.

$\square$

2.1.3 Equally Likely Outcomes

Now we'll see why counting was so useful. We can compute probabilities in the special case where each outcome is equally likely (e.g., rolling a *fair* 6-sided die has each outcome in $\Omega = \{1, 2, \ldots, 6\}$ equally likely). If events are equally likely, then in determining probabilities, we only care about the number of outcomes that are in an event. That let's us conclude the following:

Theorem 2.1.1: Probability in Sample Space with Equally Likely Outcomes

If Ω is a sample space such that each of the unique outcome elements in Ω **are equally likely**, then for any event $E \subseteq \Omega$:

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|}$$

Proof of Equally Likely Outcomes Formula. If outcomes are equally likely, then for any outcome in the sample space $\omega \in \Omega$, we have $\mathbb{P}(\omega) = \frac{1}{|\Omega|}$ (since there are $|\Omega|$ total outcomes). Then, if we list the $|E|$ outcomes that make up event E, we can write

$$E = \{\omega_1, \omega_2, \ldots, \omega_{|E|}\}$$

Every set is the union of the (mutually exclusive) singleton sets containing each element (e.g., $\{1, 2, 3\} = \{1\} \cup \{2\} \cup \{3\}$), and so by countable additivity, we get

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{|E|} \{\omega_i\}\right) &= \sum_{i=1}^{|E|} \mathbb{P}(\{\omega_i\}) && \text{[countable additivity axiom]} \\
&= \sum_{i=1}^{|E|} \frac{1}{|\Omega|} && \text{[equally likely outcomes]} \\
&= \frac{|E|}{|\Omega|} && \text{[sum constant } |E| \text{ times]}
\end{aligned}$$

The notation in the first line is like summation or product notation: just union all the sets $\{\omega_1\} \cup \{\omega_2\} \cup \cdots \cup \{\omega_{|E|}\}$. $\square$

Example(s)

If we flip two fair coins independently, what is the probability we get at least one head?

Solution Since the sample space $\Omega = \{HH, HT, TH, TT\}$ is such that events are equally likely and the event of getting at least one head is $E = \{HH, HT, TH\}$, we can say that

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|} = \frac{3}{4}$$

$\square$

Example(s)

Consider the example of rolling the red and blue fair 4-sided dice again (above), a blue die $D1$ and a red die $D2$. What is the probability that the two die's rolls sum up to 6?

Solution We called that event $B = \{(2,4),(3,3),(4,2)\}$. What is the probability of the event B happening?

Well, the 16 possible outcomes that make up all the elements of Ω are each equally likely because each die has an equal chance of landing on any of the 4 numbers. So, $\mathbb{P}(E) = \frac{|B|}{|\Omega|} = \frac{3}{16}$, so the probability is $\frac{3}{16}$. □

2.1.4 Exercises

1. If there are 5 people named A, B, C, D, and E, and they are randomly arranged in a row (with each ordering equally likely), what is the probability that A and B are placed next to each other?

   Solution: The size of the sample space is the number of ways to organize 5 people randomly, which is $|\Omega| = 5! = 120$. The event space is the number of ways to have A and B sit next to each other. We did a similar problem in 1.1, and so the answer was $2! \cdot 4! = 48$ (why?). Hence, *since the outcomes are equally likely*, $\mathbb{P}(E) = \dfrac{|E|}{|\Omega|} = \dfrac{48}{120}$.

2. Suppose I draw 4 cards from a standard 52-card deck. What is the probability they are all aces (there are exactly 4 aces in a deck)?

   Solution: There are two ways to define our sample space, one where order matters, and one where it doesn't. These two approaches are equivalent.

   (a) If *order matters*, then $|\Omega| = P(52,4) = 52 \cdot 51 \cdot 50 \cdot 49$, as the number of ways to pick 4 cards out of 52. The event space E is the number of ways to pick all 4 aces (with order mattering), which is $P(4,4) = 4 \cdot 3 \cdot 2 \cdot 1$. Hence, *since the outcomes are equally likely*, $\mathbb{P}(E) = \dfrac{|E|}{|\Omega|} = \dfrac{P(52,4)}{P(4,4)} = \dfrac{4 \cdot 3 \cdot 2 \cdot 1}{52 \cdot 51 \cdot 50 \cdot 49}$

   (b) If *order does not matter*, then $|\Omega| = \binom{52}{4}$, since we just care which 4 out of 52 cards we get. Then, there is only $\binom{4}{4} = 1$ way to get all 4 aces, and, *since the outcomes are equally likely*, $\mathbb{P}(E) = \dfrac{|E|}{|\Omega|} = \dfrac{\binom{52}{4}}{\binom{4}{4}} = \dfrac{P(52,4)/4!}{P(4,4)/4!} = \dfrac{4 \cdot 3 \cdot 2 \cdot 1}{52 \cdot 51 \cdot 50 \cdot 49}$.

   Notice how it did not matter whether order mattered or not, but we had to be consistent! The 4! accounting for the ordering of the 4 cards gets cancelled out :).

3. Given 3 different spades (S) and 3 different hearts (H), shuffle them. Compute $\mathbb{P}(E)$, where E is the event that the suits of the shuffled cards are in alternating order (e.g., SHSHSH or HSHSHS)

   Solution: The sample space $|\Omega|$ is the number of ways to order the 6 (distinct) cards: 6!. The number of ways to organize the three spades is 3! and same for the three hearts. Once we do that, we either lead with spades or hearts, so we get $2 \cdot 3!^2$ for the size of our event space E. Hence, *since the outcomes are equally likely*, $\mathbb{P}(E) = \dfrac{|E|}{|\Omega|} = \dfrac{2 \cdot 3!^2}{6!}$.

Note that all of these exercises are just counting two things! We count the size of the sample space, then the event space and divide them. It is very important to acknowledge that we can only do this when the outcomes are *equally likely*.

You can see how we can get even more fun and complicated problems - the three exercises above displayed counting problems on the "easier side". The reason we didn't give "harder" problems is because computing probability in the case of equally likely outcomes reduces to doing two counting problems (counting $|E|$ and $|\Omega|$, where computing $|\Omega|$ is generally easier than computing $|E|$). Just use the techniques from Chapter 1 to do this!