4.10 Discrete Probability

We now introduce probability, which captures the notion of the likelihood that an event may occur. We begin with the notion of an experiment which is something we conduct or happens, and has one or more outcomes. Each outcome in the context of an experiment is called an elementary event. A sample space is a set of elementary events.

For example, our experiment may be a toss of a two-sided coin, which has one of two possible outcomes, heads or tails. We could associate the symbol H with the elementary event that is the former, and T with the elementary event that is the latter. The set $\{H, T\}$, then, is a sample space. Similarly, we may associate the toss of a 6-sided die with the sample space $\{1, 2, \ldots, 6\}$, where each of its members represents the elementary event that the die lands on that number.

Note: in this course, we deal with space spaces that are finite, and therefore countable, only. Some of the following notions rely on this assumption.

An event is a subset of a sample space. For example, given the sample space $\{HH, HT, TH, TT\}$, that is the set of elementary events associated with tossing two coins, the subset $\{HH, HT, TH\}$ is an event; it is the event that we get at least one heads. A sample space is a subset of itself, and is therefore an event, which we can call the certain event. The emptyset, $\emptyset$, is the null event. Given a sample space S and two events $A, B \subseteq S$, we say that the events A and B are mutually exclusive if $A \cap B = \emptyset$. E.g., in tossing two coins, the event that we get no tails, $\{HH\}$, is mutually exclusive from the event that the first toss is a tails, $\{TH, TT\}$. We can think of each elementary event $s \in S$ as an event $\{s\}$; the elementary events are mutually exclusive from one another.

4.11 Probability

A probability distribution, Pr, is a function from the powerset of a sample space S to the real numbers $\mathbb{R}$ that satisfies the following axioms, which are called the probability axioms.

1. $\Pr\{A\} \geq 0$ for every event $A \subseteq S$.

2. $\Pr\{S\} = 1$. (This is why we call S the certain event.)

3. For pairwise mutually exclusive events $A_1, \ldots, A_n$,

$$\Pr\{A_1 \cup A_2 \cup \ldots \cup A_n\} = \Pr\{A_1\} + \Pr\{A_2\} + \ldots + \Pr\{A_n\}$$

Note: we choose to write $\Pr\{\cdot\}$ rather than $\Pr(\cdot)$, i.e., with the customary round backets that we use for functions, merely to emphasize that while $\Pr\{\cdot\}$ is a function, it is a function that happens to be a probability distribution.

We call $\Pr\{A\}$ the probability of the event A. For example, suppose we associate the sample space $S = \{1, 2, \ldots, 6\}$ with the roll of a 6-sided die. And suppose $\Pr\{1\} = \Pr\{2\} = \ldots = \Pr\{5\} = 1/10$, and $\Pr\{6\} = 1/2$. Then, such a Pr can be a probability distribution. (We need to assert, in addition, that Axiom 3 is satisfied.)

A probability distribution is said to be discrete if it is defined over a sample space that is countable. As our note above says, in this course, we deal with finite, and therefore countable, sample spaces only. Thus, all probability distributions with which we deal are discrete. In a discrete probability distribution over a sample space S, for an event $A \subseteq S$:

$$\Pr\{A\} = \sum_{s \in A} \Pr\{s\}$$

Given the above probability axioms, we can establish a number of claims for a discrete probability distribution, Pr.

Claim 34. *$Pr\{\emptyset\} = 0$.*

Proof. Assume otherwise for the purpose of contradiction, i.e., assume $\Pr\{\emptyset\} > 0$. We observe that if S is the sample space, then $S \cap \emptyset = \emptyset$, that is, S and $\emptyset$ are mutually exclusive. Therefore, $\Pr\{S \cup \emptyset\} = \Pr\{S\} = \Pr\{S\} + \Pr\{\emptyset\} > 1$, a contradiction to the axiom $\Pr\{S\} = 1$. □

Claim 35. *If $A \subseteq S$ is an event, then $Pr\{\overline{A}\} = 1 - Pr\{A\}$.*

Proof. $\overline{A} = S \setminus A$. And $\overline{A} \cap A = \emptyset$. Therefore, $\Pr\{\overline{A} \cup A\} = \Pr\{S\} = 1 = \Pr\{\overline{A}\} + \Pr\{A\} \implies \Pr\{\overline{A}\} = 1 - \Pr\{A\}$. □

The above claim can be useful in intuiting the probability of an event by considering its complement. For example, for the events associated with the toss of two coins, suppose each of the elementary events HH, HT, TH, TT has equal probability of $1/4$. Then, $\Pr\{\text{at least one heads}\} = 1 - \Pr\{\text{no heads}\} = 1 - \Pr\{TT\} = 1 - 1/4 = 3/4$.

Claim 36. *For events A, B with $A \subseteq B$, it is true that $Pr\{A\} \leq Pr\{B\}$.*

Proof. $B \supseteq A \implies B = A \cup (B \setminus A) \implies \Pr\{B\} = \Pr\{A\} + \Pr\{B \setminus A\} \implies \Pr\{B\} \geq \Pr\{A\}$. □

Claim 37. *$Pr\{A \cup B\} = Pr\{A\} + Pr\{B\} - Pr\{A \cap B\}$.*

Proof.

$$\begin{aligned}
&A = (A \setminus (A \cap B)) \cup (A \cap B) \\
\implies &\Pr\{A\} = \Pr\{A \setminus (A \cap B)\} + \Pr\{A \cap B\} \\
\\
&B = (B \setminus (A \cap B)) \cup (A \cap B) \\
\implies &\Pr\{B\} = \Pr\{B \setminus (A \cap B)\} + \Pr\{A \cap B\} \\
\\
\implies &\Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\} = \\
&\Pr\{A \setminus (A \cap B)\} + \Pr\{B \setminus (A \cap B)\} + \Pr\{A \cap B\} = \\
&\Pr\{A \cup B\}
\end{aligned}$$

□

A corollary to the above claim is: $\Pr\{A \cup B\} \leq \Pr\{A\} + \Pr\{B\}$.

4.12 Uniform probability distribution

Given a sample space $S = \{s_1, \ldots, s_n\}$, if $\Pr\{s_1\} = \ldots = \Pr\{s_n\}$, we call such a Pr a uniform probability distribution. Given such a uniform distribution over a sample space S, and an event $A \subseteq S$, we have a relatively simple formula for $\Pr\{A\}$:

$$\Pr\{A\} = \frac{|A|}{|S|}$$

Example 22. *Suppose we toss three coins, with each outcome equally likely. What is the probability that we have at least two heads?*

If S is the same space, then $|S| = 2^3$. The event A mentioned above occurs when we have either (i) exactly two heads, or, (ii) all three are heads. The number of ways in which (ii) can happen is 1. The number of ways in which (i) can happen is $\binom{3}{2}$. So:

$$Pr\{A\} = \frac{\binom{3}{2} + 1}{2^3} = \frac{1}{2}$$

Example 23. *We have a basketful of apples, oranges, pears and peaches. We reach in and take two pieces of fruit in a manner that every multiset of two pieces of fruit is equally likely. What is the probability that we end up with two different kinds of fruit?*

Let A be the event that we end up with two different kinds of fruit. Then, $Pr\{A\} = 1 - Pr\{\overline{A}\}$, where $\overline{A}$ is the event that we end up with two of the same kind of fruit. The number of ways in which $\overline{A}$ can happen is 4, because we have 4 different kinds of fruit.

All that remains is for us to intuit the size of the same space, call it S, which is all possible multisets of size 2. Our situation corresponds to unordered selection with replacement, and so:

$$\begin{aligned} Pr\{A\} &= 1 - Pr\{\overline{A}\} = 1 - \frac{|\overline{A}|}{|S|} = 1 - \frac{4}{\left(\!\binom{4}{2}\!\right)} \\ &= 1 - \frac{4}{\binom{4+2-1}{2}} = 1 - \frac{4}{10} = \frac{3}{5} \end{aligned}$$

It is somewhat interesting to sanity-check the solution in the above example by changing the number of different kinds of fruit in the basket, call it d. The example considers the case that $d = 4$. The following table gives us the probability of picking two different pieces of fruit for different values of d.

d	$\Pr\{A\} = 1 - \frac{d}{\left(\!\binom{d}{2}\!\right)}$
1	$1 - \frac{1}{\binom{1+2-1}{2}} = 1 - \frac{1}{1} = 0$
2	$1 - \frac{2}{\binom{2+2-1}{2}} = 1 - \frac{2}{3} = 1/3$
3	$1 - \frac{3}{\binom{3+2-1}{2}} = 1 - \frac{3}{6} = 1/2$
4	$1 - \frac{4}{\binom{4+2-1}{2}} = 1 - \frac{4}{10} = 3/5$
5	$1 - \frac{5}{\binom{5+2-1}{2}} = 1 - \frac{5}{15} = 2/3$
6	$1 - \frac{6}{\binom{6+2-1}{2}} = 1 - \frac{6}{21} = 5/7$
7	$1 - \frac{7}{\binom{7+2-1}{2}} = 1 - \frac{7}{28} = 3/4$

The table suggests that as the number of kinds of fruit increases in the basket, the probability of picking two different kinds of fruit increases. This of course appeals to the common sense.

4.13 Conditional probability and independence

Conditional probability addresses situations that we already have some prior knowledge about some outcomes. Consider the following game, which is from a TV show called "Let's Make a Deal."

There are three curtains, numbered 1, 2 and 3. Behind one of them is a desirable prize. Behind the other two, there is nothing. The game goes as follows. You are first asked to pick one of the curtains. The host then draws back one of the other curtains that does not contain the prize; we know that there is at least one. The host them gives you the opportunity to change your choice to the other curtain that remains closed.

Should we change our choice? Is it rational to do so?

We can pose this as a problem of intuiting the probability of winning if we switch our choice, given our a priori knowledge that the curtain that the host drew back does not contain the prize. If this probability is higher than 1/3, we should switch; otherwise, there is no rational reason to switch. The value 1/3 comes from our assumption that initially, we have a uniform distribution, i.e., the probability that the prize is behind any one of the curtains is 1/3.

The above problem is called "The Monty Hall problem," after the host of the game show. We revisit it after our discussions on conditional probability. A simpler example is: suppose we toss two coins, with every elementary event equally likely, and you know that one of them lands heads. So that is our a priori knowledge. What is the probability that both land heads?

The fact that one of the coins lands heads eliminates the event TT, that both land tails. So, the only possible events are HH, HT, TH. And therefore, the conditional probability in question is 1/3.

The conditional probability of an event A given that an event B occurs, i.e., $\Pr\{B\} \neq 0$, read as "the probability of A given B" is:

$$\Pr\{A \mid B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}}$$

One way to understand the above formula is that we normalize the probability that both A and B occur by the probability that B occurs. For example, for our coin toss example above, A is the event that both coins land heads, and B is the event that one of them lands heads. And we have:

$$\Pr\{A \mid B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}} = \frac{\Pr\{A\}}{\Pr\{B\}} = \frac{1/4}{3/4} = \frac{1}{3}$$

We exploited the fact that $A \cap B = A$, because $A \subseteq B$.

We say that events A and B are said to be independent if $\Pr\{A \cap B\} = \Pr\{A\}\Pr\{B\}$. This is equivalent, if $\Pr\{B\} \neq 0$, to: $\Pr\{A \mid B\} = \Pr\{A\}$.

Example 24. *Suppose we toss a coin once and then again, in a manner that every elementary event, HH, TT, HT, TH, is equally likely. Let A be the event that the first toss lands heads. And B be the event that the two tosses land differently. Are* *the events A and B independent?*

We compare $Pr\{A\}$ with $Pr\{A \mid B\}$.

$$\begin{aligned}
Pr\{A\} &= 1/2 \\
Pr\{A \mid B\} &= \frac{Pr\{A \cap B\}}{Pr\{B\}} = \frac{Pr\{HT\}}{Pr\{HT, TH\}} = \frac{1/4}{2/4} = \frac{1}{2}
\end{aligned}$$

Thus, the events A and B are indeed independent.

Example 25. *You have a coin that you fear may be biased. That is, it lands heads with some probability $p \in (0, 1)$, and tails with probability $1 - p$. You do not know what p is, except that it is neither 0 nor 1. Devise a way to get a fair coin toss.*

Consider the following approach. We repeatedly toss the coin twice till the two outcomes are different. Then we adopt the first of the two tosses as our result.

Why does this work? With every two tosses, we have the sample space $S = \{HH, HT, TH, TT\}$. And we observe that $Pr\{HT\} = Pr\{TH\} = p(1 - p)$. That is, we have the same probability for the two events that correspond to different outcomes for the two tosses.

Example 26. *A standard pack of 52 cards includes 12 "face cards" – Queens, Kings and Jacks. Suppose you draw two cards uniformly at random from such a standard pack, and notice that the first is not a face card. What is the probability that the second is a face card?*

Let A be the event that the first is not a face card, and B be the event that the second is. We seek $Pr\{B \mid A\}$.

$$\begin{aligned}
Pr\{B \mid A\} &= \frac{Pr\{B \cap A\}}{Pr\{A\}} = \frac{(40 \times 12)/(52 \times 51)}{40/52} \\
&= \frac{40 \times 12 \times 52}{52 \times 51 \times 40} = \frac{12}{51}
\end{aligned}$$

This makes sense, because once we remove a non-face card, we have a $12/51$ chance *of drawing a face card. Also, the events A and B are not independent. Because:*

$$
\begin{aligned}
Pr\{B\} &= Pr\big\{(B \cap A) \cup \big(B \cap \overline{A}\big)\big\} \\
&= Pr\{B \cap A\} + Pr\big\{B \cap \overline{A}\big\} \\
&= \frac{40 \times 12}{52 \times 51} + \frac{12 \times 11}{52 \times 51} \\
&= \frac{12 \times (40 + 11)}{52 \times 51} \\
&= \frac{12}{52} \neq \frac{12}{51} = Pr\{B \mid A\}
\end{aligned}
$$

In the above example, does it make sense that $\Pr\{B\}$, the probability that the second card that is chosen, is $\frac{12}{52}$? We observe that this is the same probability that, if we choose one card uniformly at random from the pack of 52, it is a face card. What if, for example, we pick 10 cards, one after another uniformly at random, and ask what the probability is that the eighth is a face card? The answer, as per the above mindset, should still be $\frac{12}{52}$.

We argue that this does make sense based on the following reasoning. Suppose we shuffle the cards thoroughly and lay them out left to right on a table. The leftmost card then, can be seen as corresponding to our first pick, the second card from the left as our second pick, and so on. Now, if we ask what the probability is that any one of them is a face card, it is $\frac{12}{52}$. Thus, if we uniformly at random pick n cards out of the 52, and ask what the probability is that the k^{th} of those cards is a face card, for $1 \leq k \leq n$, the answer is the same, $\frac{12}{52}$.

We now articulate Bayes's theorem, which relates $\Pr\{A \mid B\}$ and $\Pr\{B \mid A\}$. It is useful, for example, when one of those probabilities is easier to intuit than the other.

Claim 38 (Bayes's theorem)**.** *Suppose $Pr\{A\} \neq 0, Pr\{B\} \neq 0$. Then:*

$$Pr\{A \mid B\} = \frac{Pr\{A\}Pr\{B \mid A\}}{Pr\{B\}}$$

Proof.

$$\begin{aligned}
\Pr\{A \mid B\} &= \frac{\Pr\{A \cap B\}}{\Pr\{B\}} \\
\Pr\{B \mid A\} &= \frac{\Pr\{A \cap B\}}{\Pr\{A\}} \\
\Longrightarrow \Pr\{B\}\Pr\{A \mid B\} &= \Pr\{A\}\Pr\{B \mid A\} \\
\Longrightarrow \Pr\{A \mid B\} &= \frac{\Pr\{A\}\Pr\{B \mid A\}}{\Pr\{B\}}
\end{aligned}$$

□

Example 27. *We revisit Example 26 and ask, instead, what $Pr\{A \mid B\}$ is, i.e., the probability that the first card we draw is not a face card, given that the second is.*

$$\begin{aligned}
Pr\{A\} &= \frac{40}{52} \\
Pr\{B\} &= \frac{12}{52} \\
Pr\{A \mid B\} &= \frac{Pr\{A\}Pr\{B \mid A\}}{Pr\{B\}} && \because \textit{Bayes} \\
&= \frac{40/52 \times 12/51}{12/52} = \frac{40 \times 12 \times 52}{52 \times 51 \times 12} \\
&= \frac{40}{51}
\end{aligned}$$

Example 28. *We address the Monty Hall problem that we introduced earlier. Recall that the problem is as follows. There are three curtains behind one of which is a prize. We initially pick a curtain, and Monty then opens one of the other curtains that does not contain the prize. He then gives us the option of switching our choice to the third curtain before he reveals behind which curtain the prize is. The question is: should we switch? Or more specifically, does our probability of winning increase by switching?*

Assume that we choose Curtain 1 initially and then Monty opens curtain 2. Consider the following two events:

- *P_1 is the event that the prize is behind Curtain 1.*
- *R_2 is the event that after we have initially chosen Curtain 1, Monty opens Curtain 2.*

Then, we are interested to know $Pr\{P_1 \mid R_2\}$*. Because, if* $Pr\{P_1 \mid R_2\} < 1/2$*, that would be a good rationale to switch to Curtain 3.*

We leverage Bayes to determine $Pr\{P_1 \mid R_2\}$*. For that, we need to determine* $Pr\{P_1\}, Pr\{R_2\}$ *and* $Pr\{R_2 \mid P_1\}$*.* $Pr\{P_1\} = 1/3$ *because the prize is equally likely to be behind any of the three curtains.* $Pr\{R_2 \mid P_1\} = 1/2$ *because if the prize is behind Curtain 1, given that we have chosen Curtain 1 initially, Monty can open either Curtain 2 or 3, and we assume he picks one with equal probability.*

As for $Pr\{R_2\}$*, we know that it is* 0 *if the prize is behind Curtain 2. Also, Monty cannot open Curtain 1 as we chose it initially. So the only way the event* R_2 *can occur is if the prize is behind Curtain 3. And this occurs with probability* 1/2 *because the prize may be behind either Curtain 1 or 3 with equal probability. So:*

$$\begin{aligned} Pr\{P_1 \mid R_2\} &= \frac{Pr\{P_1\}Pr\{R_2 \mid P_2\}}{Pr\{R_2\}} \\ &= \frac{1/3 \times 1/2}{1/2} = 1/3 \end{aligned}$$

Therefore, we should switch to Curtain 3, because the probability that the prize is behind Curtain 3 is $1 - 1/3 = 2/3$.

Example 29. *Suppose we have two coins, one of which is fair, and the other always comes up heads. Suppose we pick one of those coins uniformly at random and toss it three times, and it so happens that it comes up heads all of the three times. What is the probability that we happened to pick the coin that always comes up heads?*

Let A *be the event that we pick the coin that always comes up heads. Let* B *be the event that all three tosses of the chosen coin come up heads. We seek* $Pr\{A \mid B\}$*.*

We leverage Bayes, for which we need to know $Pr\{A\}, Pr\{B\}$ *and* $Pr\{B \mid A\}$*.* $Pr\{A\} = 1/2$*, and* $Pr\{B \mid A\} = 1$*. To determine* $Pr\{B\}$*, we observe:*

$$\begin{aligned} Pr\{B\} &= Pr\{B \cap A\} + Pr\{B \cap \overline{A}\} \\ &= Pr\{A\}Pr\{B \mid A\} + Pr\{\overline{A}\}Pr\{B \mid \overline{A}\} \\ &= 1/2 \times 1 + 1/2 \times 1/8 \\ &= 9/16 \end{aligned}$$

So our solution:

$$\frac{Pr\{A\}Pr\{B \mid A\}}{Pr\{B\}} = \frac{1/2 \times 1}{9/16} = \frac{8}{9}$$

We expect that the more tosses we make that are all heads, the higher the probability that we have chosen the biased coin. Of course, if we see even one tails, we immediately know that we have chosen the fair coin.

4.14 Expectation

We conclude our discussions on discrete probability with the notion of expectation, or the expected value of a discrete random variable.

Given a sample space S over which we specify a probability distribution, Pr, a discrete random variable X is a function from the sample space to a real number, $X : S \to \mathbb{R}$.

For example, suppose we toss a coin thrice, and I am to lose \$2 for every tails, and win \$10 for every heads. Then, we can specify a discrete random variable, call it W, which is my total winnings. The sample space is $\{H, T\} \times \{H, T\} \times \{H, T\}$. The range of W is $\{-6, 6, 18, 30\}$.

As we deal with only discrete random variables in this course, we drop the qualifier "discrete," henceforth. Given a random variable X, we define the event $X = x$ to be the set $\{s \in S \mid X(s) = x\}$. In our above example, the event $W = 18$ is $\{THH, HTH, HHT\}$. And then:

$$\Pr\{X = x\} = \sum_{s \in S : X(s) = x} \Pr\{s\}$$

In our above example, if the coin is fair, then $\Pr\{W = 18\} = 3/8$.

The expected value, expectation or mean of a random variable $X : S \to \mathbb{R}$ is denoted $E[X]$, and defined as:

$$\begin{aligned} E[X] &= \sum_{x \in \mathbb{R}} x \cdot \Pr\{X = x\} \\ &= \sum_{s \in \mathbb{S}} X(s) \cdot \Pr\{s\} \end{aligned}$$

As the formula suggests, the expectation of X is a weighted average, where each of the values X can take is weighted by the probability with which X takes that value.

For example, in our above coin-toss game, the expectation of the random variable

W, assuming that the coin is fair, is:

$$\begin{aligned} E[W] &= -6 \times \frac{1}{8} + 6 \times \frac{3}{8} + 18 \times \frac{3}{8} + 30 \times \frac{1}{8} \\ &= \frac{1}{8}(-6 + 18 + 54 + 30) = 12 \end{aligned}$$

The idea behind the expected value is exactly what we associate with the term "expectation." That is, if we play the coin-toss game, we expect to win \$12. And interesting observation is that the expectation is not necessarily one of the values that the random variable can take. That is, in our above example, there is no situation in which we actually win \$12, as our winning from playing the game once is one of $-6, 6, 18$ or 30.

Example 30. *We toss a fair* 6*-sided die whose faces are numbered* $1, \ldots, 6$*. What is the expectation of the toss?*

If T is a random variable that is the value the die lands, we have:

$$E[T] = \frac{1}{6}(1 + 2 + \ldots + 6) = \frac{21}{6} = 3.5$$

The expectation can be used to make decisions that we can argue are rational. Consider the following example.

Example 31. *You need to put in \$15 upfront to play the following game. We toss a fair coin twice. You earn \$4 for every tails and \$10 for every heads. Would you play this game?*

One way to rationally answer this question is to define an appropriate random variable and compute its expected value. Let X be a random variable that is our earnings after the two tosses. If $E[X] \geq 15$, we agree to play the game. If not, we do not play the game.

We observe:

$$E[X] = 8 \times \frac{1}{4} + 14 \times \frac{1}{2} + 20 \times \frac{1}{4} = 14$$

So, if we play the game in the above example, we expect to lose money. This is exactly the kind of set up we see in Casinos. It is not quite true that "the house always wins." Rather, if we play a game in the Casino, we expect to lose money. Of course, we may win as well, and the house may lose. But the expectation captures the long-term trend. That is, provided the Casino is able to stay in business long enough and has sufficiently many visitors, it is highly likely to make a profit. Of course, if the odds are too skewed in favour of the house, no one would visit.

In the above example, we can ask what the probability is that we win more than \$15 so we do not lose money. And the answer is of course that the only way is if we land both heads, which happens with probability 1/4 only.

Example 32. *We revisit Example 25, in which we are given a biased coin, which lands heads with probability $p \in (0, 1)$, and tails with probability $1-p$. Our algorithm to ensure a fair coin toss is: repeatedly toss the coin twice till we see two different results. Choose the first of the two as the result of our fair coin toss.*

As we discuss there, this works because $Pr\{HT\} = Pr\{TH\}$. However, a concern may be the number of times we may have to repeatedly toss the coin before we finally have a result for our fair coin toss. How many could it be?

Of course, in the worst-case, we may never stop – we may get so unlucky that both consecutive coin tosses always have the same result. But what if we ask how many consecutive pair of tosses we expect to have to make before we are able to stop?

Let T be the corresponding random variable. We can intuit $E[T]$ from our definition of expectation. We observe that T takes on values in $\mathbb{N}$, i.e., $1, 2, \ldots$ That is, we may stop after the first pair of coin tosses, or the second, and so on. The only reason we engage in a second pair of coin tosses is that we got the same results for both tosses in the first pair. The probability with which we get the same result in a pair of tosses is $p^2 + (1-p)^2$, i.e., both tails or heads, with each event being mutually exclusive.

So we have the following for $E[T]$, with explanations for the lines with equation *numbers following.*

$$
\begin{aligned}
E[T] = {} & 1 \times 2p(1-p) \; + \\
& 2 \times \left[p^2 + (1-p)^2\right] \times 2p(1-p) \; + \\
& 3 \times \left[p^2 + (1-p)^2\right]^2 \times 2p(1-p) \; + \\
& 4 \times \left[p^2 + (1-p)^2\right]^3 \times 2p(1-p) \; + \\
& \ldots \\
= {} & \sum_{i=1}^{\infty} i \times \left[p^2 + (1-p)^2\right]^{i-1} \times 2p(1-p) \\
= {} & 2p(1-p) \sum_{i=1}^{\infty} i \times \left[p^2 + (1-p)^2\right]^{i-1} \\
= {} & 2p(1-p) \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \left[p^2 + (1-p)^2\right]^{j-1} && (4.1) \\
= {} & 2p(1-p) \sum_{i=1}^{\infty} \frac{\left[p^2 + (1-p)^2\right]^{i-1}}{1 - [p^2 + (1-p)^2]} && (4.2) \\
= {} & \frac{2p(1-p)}{1 - [p^2 + (1-p)^2]} \sum_{i=1}^{\infty} \left[p^2 + (1-p)^2\right]^{i-1} \\
= {} & \frac{2p(1-p)}{\left[1 - (p^2 + (1-p)^2)\right]^2} && (4.3) \\
= {} & \frac{2p(1-p)}{\left[2p(1-p)\right]^2} && (4.4) \\
= {} & \frac{1}{2p(1-p)}
\end{aligned}
$$

Explanations: for clarify, adopt $a = p^2 + (1-p)^2$.

(**4.1**) We seek $1 \times a^0 + 2 \times a^1 + 3 \times a^2 + 4 \times a^3 + \ldots$ We rewrite this as $(a^0 + a^1 + a^2 + a^3 + \ldots) + (a^1 + a^2 + a^3 + \ldots) + (a^2 + a^3 + \ldots) + \ldots$ This is exactly what the double summation expresses.

(**4.2**) The inner summation is what is called a geometric series. It is of the form

$a^{i-1} + a^i + a^{i+1} + a^{i+2} + \ldots$ We can intuit what that summation is as follows:

$$\begin{aligned} S &= a^{i-1} + a^i + a^{i+1} + a^{i+2} + \ldots \\ aS &= a^i + a^{i+1} + a^{i+2} + \ldots \\ \hline S - aS &= a^{i-1} \\ \implies S &= \frac{a^{i-1}}{1-a} \end{aligned}$$

(**4.3**) We again have a geometric series, except that the first term in the summation is $a^0 = 1$.

(**4.4**) $1 - [p^2 + (1-p)^2] = 2p(1-p)$. We can intuit this by looking at the binomial expansion of $[p + (1-p)]^2$, or simply by observing that $2p(1-p) = \Pr\{HT, TH\} = 1 - \Pr\{HH, TT\} = 1 - \left[p^2 + (1-p)^2\right]$.

As an example, suppose $p = 1/2$, that is, the coin is fair. Then, the number of pairs of tosses we expect to have to make before we have a result for our fair coin toss is: $\frac{1}{2p(1-p)} = 2$.

If the coin is more skewed, e.g., $p = 1/8$, then the expected number of pairs of tosses is $\frac{1}{2\times 1/8 \times 7/8} = \frac{32}{7}$, which is between 4 and 5. It makes sense that our expected number of pairs of tosses increases as the coin gets more skewed. We get the minimum when the coin is fair, i.e., 2 pairs of tosses only.

In Example 32 above, we could have saved ourselves a whole lot of work on the math if we had been a bit more creative with the random variable we defined, paired with some additional observations about the expected value of random variables.

The first observation is the so-called linearity of expectation: if X, Y are random variables, then $E[X+Y] = E[X] + E[Y]$. The second is about so-called indicator random variables. An indicator random variable is a random variable which takes one of two values only: 0 or 1. Then, if X is an indicator random variable, $E[X] = \Pr\{X = 1\}$.

To prove the linearity of expectation, we recall that a random variable is a function, and rely on how we define addition for functions. We restrict ourselves to functions whose codomain is the real numbers. Given functions $f : A \to \mathbb{R}, g : A \to \mathbb{R}$, we define the function $(f+g) \colon A \to \mathbb{R}$ as $(f+g)(a) = f(a) + g(a)$.

Claim 39. *If $X : S \to \mathbb{R}, Y : S \to \mathbb{R}$ are random variables where S is a sample space, then $E[X+Y] = E[X] + E[Y]$.*

Proof.

$$\begin{aligned}
E[X] &= \sum_{s \in \mathbb{S}} X(s) \cdot \Pr\{s\} \\
E[Y] &= \sum_{s \in \mathbb{S}} Y(s) \cdot \Pr\{s\} \\
E[X+Y] &= \sum_{s \in \mathbb{S}} (X+Y)(s) \cdot \Pr\{s\} \\
&= \sum_{s \in \mathbb{S}} (X(s) + Y(s)) \cdot \Pr\{s\} \\
&= \sum_{s \in \mathbb{S}} ((X(s) \cdot \Pr\{s\}) + (Y(s) \cdot \Pr\{s\})) \\
&= \sum_{s \in \mathbb{S}} X(s) \cdot \Pr\{s\} + \sum_{s \in \mathbb{S}} Y(s) \cdot \Pr\{s\} \\
&= E[X] + E[Y]
\end{aligned}$$

□

Note that the above can be generalized to several random variables. That is:

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Claim 40. *If X is an indicator random variable, i.e., takes on the value 0 or 1 only, then $E[X] = Pr\{X = 1\}$.*

Proof. As X takes on the value 0 or 1 only:

$$\begin{aligned}
E[X] &= 0 \cdot \Pr\{X = 0\} + 1 \cdot \Pr\{X = 1\} \\
&= \Pr\{X = 1\}
\end{aligned}$$

□

We now return to Example 32. Consider the following alternative way of intuiting the expected number of pairs of tosses till we are able to return the result of a fair coin toss.

Suppose we carry out n pairs of such tosses. We first ask in how many we expect to have TH or HT. We then ask what n must be so that this expectation is at least

1. We proceed as follows, assuming that we carry out n pairs of tosses. Define n random variables, $X_1, \ldots, X_n$ as follows:

$$X_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ pair is } TH \text{ or } HT \\ 0 & \text{otherwise} \end{cases}$$

Let R be a random variable that is the number of such pairs of tosses for which we are able to return a result. Then:

$$\begin{aligned} R &= \sum_{i=1}^{n} X_i \\ \implies E[R] &= E\left[\sum_{i=1}^{n} X_i\right] \\ &= \sum_{i=1}^{n} E[X_i] \\ &= \sum_{i=1}^{n} \Pr\{X_i = 1\} \\ &= \sum_{i=1}^{n} 2p(1-p) \\ &= 2p(1-p)\sum_{i=1}^{n} 1 \\ &= 2np(1-p) \end{aligned}$$

And so,

$$E[R] \geq 1 \iff 2np(1-p) \geq 1 \iff n \geq \frac{1}{2p(1-p)}$$

The notion of an indicator random variable is related closely to the notion of a Bernoulli trial. A Bernoulli trial is an experiment which has one of two outcomes only: success or failure. And, if p is the probability of success in a Bernoulli trial, the expected number of trials before a success is $1/p$. This can be proved easily by leveraging an appropriately defined indicator random variable. Our experiment in Example 32 can be seen as a Bernoulli trial: success is a pair of coin tosses with different outcomes. As the probability of success is $2p(1-p)$, the expected number of trials before a success is $\frac{1}{2p(1-p)}$.

Example 33. *We revisit the situation in Example 29. We have a fair coin, and a coin that always lands heads. Suppose we pick one of the two uniformly at random, toss it, and repeat both of those steps till we get tails. What is the expected number of tosses?*

We adopt the notion of a Bernoulli trial. That is, success is when we get a tails from our randomly chosen coin. If we are able to intuit the probability of success, call it p, then $1/p$ is the expectation we seek.

Thus, the success event is when: (i) we choose the fair coin, and, (ii) a toss results in tails. And this probability is $1/2 \times 1/2 = 1/4$, and therefore, our expected number of tosses is 4.

The approach in the above example may be used to distinguish the coins. Following is another approach. We toss both coins simultaneously till one of them lands tails. We would then have immediately identified which coin is which. What is the expected number of tosses of each coin in this approach?

The expectation in this case is the same as the expected number of tosses of the fair coin till it lands tails. We can perceive this as a Bernoulli trial: we toss the fair coin, and success is that it lands tails. The probability of success, then, is $1/2$, and therefore the expectation is 2. Thus, we expect to have to toss each coin twice before we identify which is which. Thus, the total number of tosses is 4.

We conclude with an example from algorithms. Suppose you are given a set of n distinct integers, where n is odd. You are asked for an algorithm to find the median of those integers. The median is the middlemost value from amongst the members of the set. E.g., the median of $\{-42, 17, 4, 5, 6\}$ is 5.

Consider the following randomized algorithm. We pick a number from the set, call it i, uniformly at random from amongst the n numbers. We then check whether i is indeed the median. We can do this, for example, by comparing i to every other number, and counting how many are smaller than i. The number of integers in the set that is smaller than i is $(n-1)/2$ if and only if i is the median. If we find out that i is not the median, we repeat the entire process. That is, we pick an integer uniformly at random and test it. (Of course, we may again pick i.)

This algorithm may not seem good, but it is simple, and somewhat surprisingly good in expectation. We can ask, for example, how many trials, i.e., random pick and subsequent check, we expect to make before we find the median. To answer this, we perceive a random pick as a Bernoulli trial. Success is if we picked the median. The probability of success, then, is $1/n$. Therefore, the expected number of trials before success is n.