Probability and Expected Value

Evan Chang and bissue

September 22, 2021

Contents

1 Introduction to Probability

"A reasonable probability is the only certainty."

- E.W. Howe

1.1 Probability is a big word

Probability is a branch of combinatorics that concerns mathematical analyzation of how likely an event is of occurring, given that certain constraints or assumptions hold true.

Let's break that definition down. I'm feeling hungry, so why not talk about a hypothetical pizza delivery. Suppose you order a pizza, and you want to think about how likely that pizza is going to arrive within 20 minutes. What constraints or assumptions you know are going to hold true? Well, you know that the pizza is pepperoni, the pizza comes from a pizzeria a couple miles away, the weather is sunny today, the delivery man gets to your house using a delivery car, and so on. But there are other constraints that are going to be variable, like how long it takes to package the pizza, and how bad the traffic is. I tend to think of probability as the art of analyzing all of these variables to make one final conclusion about the chance of one particular event.

You see probability in your daily life. A 69 percent chance of snow, rain, sun, etc.? That's probability for you. Casinos use knowledge of probability in their favor to squeeze money about of the pockets of gamblers (which is why gambling is so destructive). Investors use probability all the time to decide what stocks they should invest in, or what stocks they shouldn't invest in.

Of course, if you're reading this you probably aren't interested in all that. I will be walking you through everything you'll need to tackle probability questions on the AMC 10/12, and some higher level competitions.

1.2 Basic definitions and Operations

For any event, the probability of this event occurring is a real number value between 0 and 1. No, it does not have to be a rational number. The higher this value is, the more likely that event is going to occur. 0 means that the event is impossible (for example, me not screaming in pain when I step on a LEGO). 1 means the event is always going to happen (for example, me checking Discord everyday). A probability close to 0 means the event is almost impossible, but not totally out of consideration (for example, getting struck by lightning). A probability close to 1 means the event is almost guaranteed to happen, but it is not a certainty (for example, my webpage loading). What's the probability that if you flip a fair coin, you get heads? Your intuition (and general knowledge) might guess that it's $0.5 = \frac{1}{2}$. As we see later, this is correct.

Before we proceed further, a few basic definitions:

> **Mutually Exclusive**
> A set of events are mutually exclusive if none of them can occur at the same time.

For example, if we roll a die, the events "We roll a six" and "We roll a one" are mutually exclusive. We cannot roll both a six or a one, we must roll one, the other, or neither of them. However, the events "We roll an even number" and "We roll a multiple of three" are not mutually exclusive. (Why?)

Independent Events
We say that a set of events are independent if the probability that one event occurs does not affect the probability of any other event in the set occurring. If events are not independent we say that they are dependent.

If we flip two coins at once, the result of the two coin flips are independent: the result of one does not correlate with the result of the other. Say we choose one card randomly from a deck of 52 cards, and don't put it back in. Then we choose another card randomly from the 51 cards that are left. Surprisingly, the probability that the first card is a Heart, and the probability that the second card is a Heart, are not independent (why)?

Now suppose we have multiple events we have to analyze. We then have the following.

Fact 1.1. Suppose an event P occurs with p probability for a real number $0 \leq p \leq 1$, and an event Q occurs with q probability for a real number $0 \leq q \leq 1$. If P and Q are mutually exclusive, then we have that the probability that P OR Q (one or the other) occurs equals $p + q$.

This is known as the addition principle. A very useful corollary of this is that given an event P, the probability that P occurs or P doesn't occur is 1. Thus, if we have the probability that P doesn't occur, we can subtract that from 1 to get the probability that P does indeed occur. This seems obvious, but it's a crucial step in a great many probability problems. Furthermore, we have the following:

Fact 1.2. Suppose an event P occurs with p probability for a real number $0 \leq p \leq 1$, and an event Q occurs with q probability for a real number $0 \leq q \leq 1$. If P and Q are independent, then we have that the probability that P and Q occur equals $p \cdot q$.

If P and Q are not independent, we can still say the following:

Fact 1.3. Suppose an event P occurs with p probability for a real number $0 \leq p \leq 1$, and suppose an event Q occurs with q probability given that P occurs, for a real number $0 \leq q \leq 1$. We have that the probability that P and Q occur equals $p \cdot q$.

This is known as the multiplication principle.

Remember both of these: they are crucial.

Exercise 1.4. Why is Fact 1.2 a special case of Fact 1.3?

Exercise 1.5. Say we choose one card randomly from a deck of 52 cards, but then we put it back in again. Then we choose another card randomly from the 52 cards in the deck. Are the probability that the first card is a Heart, and the probability that the second card is a Heart, independent?

1.3 Probability with equally occurring outcomes

The most basic type of probability involves probability where all of the possible outcomes are equally occurring, and are also mutually exclusive. In other words, there must always be one and only one of them that occurs. Then, we have the following essential fact:

Fact 1.6. Suppose that we have N mutually exclusive outcomes that occur with equal probability, and we know that one of these outcomes must always occur. the probability that one particular outcome occurs is

$$\frac{1}{N},$$

where N is a positive integer.

Using this, we can see the answer to the following question:

Example 1.7
Suppose we roll a fair six-sided die. What is the probability we get 1?

Solution. We can roll $1, 2, 3, 4, 5, 6$. Since the die is fair each of these outcomes occurs equally often. Since they are mutually exclusive, the probability that we get 1 is just $\boxed{1/6}$. □

Notice that we needed several assumptions before we could find an answer. First, the die had to be fair - each option had to occur with equal chance. Furthermore, there would always be exactly one outcome achieved. If these assumptions did not hold, the solution would be invalidated.

Now, we apply the addition principle to Fact 1.6. We have that

Fact 1.8. Given N mutually exclusive outcomes that occur with equal probability, and we know that one of these outcomes must always occur. The probability that one of M distinct outcomes occur is

$$\frac{M}{N},$$

where N is a positive integer and M is a nonnegative integer less than or equal to N.

Now, we consider the following example.

Example 1.9
Suppose we roll a fair six-sided die. What is the probability we get an even integer?

Solution. We can roll $1, 2, 3, 4, 5, 6$. Since the die is fair each of these 6 outcomes comes up equally often. We want the probability that we get $2, 4$, or 6. Since these outcomes are also mutually exclusive, the probability that we get one of these 3 outcomes is just $\frac{3}{6} = \boxed{1/2}$. □

A lot of times, probability problems just tend to be counting problems that are conjoined with fact 1.8. We have a set of mutually exclusive outcomes that occur with equal probability, and we know that one of these outcomes must always occur. Then we have some condition, and we want to know the probability that it is satisfied. We count the total number of possible outcomes N, then count the number of possible outcomes M that lead to the condition being satisfied. Putting M over N yields our final probability.

Exercise 1.10. Suppose we flip a coin. What is the probability that it is heads?

Exercise 1.11. Suppose we roll a fair six-sided die twice. What is the probability we get an even integer both times? What about exactly one time?

Exercise 1.12. Say we choose one card randomly from a full deck of 52 cards, but then we put it back in again. Then we choose another card randomly from the 52 cards in the deck. What is the probability that both cards are Hearts?

1.4 Casework and Complementary Probability

Sometimes we will see that the ideas of Casework and Complementary Counting can be extended to Probability. First, we will often be finding the probability that an event doesn't occur, then subtracting it from 1 to find the probability that it does occur. The key idea is to make sure that the complement is actually easier to compute than the original problem. If the complement is just as complicated, then there's no use to waste your time trying to find it.

Example 1.13
We roll seven coins. What is the probability that not all of the coins flip heads?

Solution. We're trying to find 1 minus the probability that all coins flip heads. Thankfully, the latter is easy to compute, it's just $\left(\frac{1}{2}\right)^7 = \frac{1}{128}$. So our probability is $1 - \frac{1}{128} = \boxed{\frac{127}{128}}$. □

Here's another example.

Example 1.14
Two different positive integers less than 21 are randomly chosen and multiplied together. What is the probability that the resulting product is a multiple of 3?

Solution. We can choose two numbers in $\binom{10}{2} = 190$ ways. Two numbers will have a product which is a multiple of 3 if at least one of them is a multiple of 3. We can more easily count the number of ways for the product to not be a multiple of 3. This happens when neither of the numbers is a multiple of 3.

There are 14 numbers which aren't multiples of 3. The number of ways to choose two of these numbers is $\binom{14}{2} = 91$, so the number of way to choose two numbers where at least one is a multiple of 3 is 99. The final probability is $\boxed{\frac{99}{190}}$. □

An interesting aside: suppose we replace 21 with some other positive integer n. As n increases, our probability actually ends up approaching a certain constant. What is this constant? Think intuitively.

Second, we will often be "splitting" the probability that an event occurs into several disjoint subcases. We find the probability that each subcase occurs, then sum across all subcases to get a final probability.

Example 1.15 (2020 AMC 10B P18)
An urn contains one red ball and one blue ball. A box of extra red and blue balls lie nearby. George performs the following operation four times: he draws a ball from the urn at random and then takes a ball of the same color from the box and returns those two matching balls to the urn. After the four iterations the urn contains six balls. What is the probability that the urn contains three balls of each color?

Solution. There are more clever ways, but the most direct solution is just casework on how George performs his operations to create the desired ball distribution.

In order to get 3 balls of each color, he must select two red balls and two blue balls exactly. We do casework on the order in which he selects balls from the urn.

Let R denote when George selects red, and B denote when George selects blue. There are 6 cases for the order in which balls are selected: $RRBB$, $RBRB$, $RBBR$, $BBRR$, $BRBR$, and $BRRB$. However we can put $RRBB$ and $BBRR$, $RBRB$ and $BRBR$, and $RBBR$ and $BRRB$ together since they occur with equivalent probability. (see "Symmetry" section).

Let's go onto case 1 : either $RRBB$ or $BBRR$. Let's find the probability that he picks the balls in the order of $RRBB$. The probability that the first ball he picks is red is $\frac{1}{2}$.Now there are 2 reds and 1 blue in the urn. The probability that he picks red again is now $\frac{2}{3}$. There are 3 reds and 1 blue now. The probability that he picks a blue is $\frac{1}{4}$. Finally, there are 3 reds and 2 blues. The probability that he picks a blue is $\frac{2}{5}$. So the probability that the $RRBB$ case happens is $\frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{4} \cdot \frac{2}{5} = \frac{1}{30}$. Case 1 has a probability of $\frac{1}{30} \cdot 2 = \frac{1}{15}$ chance of happening.

Case 2: either $RBRB$ or $BRBR$. Let's find the probability that he picks the balls in the order of $RBRB$. The probability that the first ball he picks is red is $\frac{1}{2}$. Now there are 2 reds and 1 blue in the urn. The probability that he picks blue is $\frac{1}{3}$. There are 2 reds and 2 blues now. The probability that he picks a red is $\frac{1}{2}$. Finally, there are 3 reds and 2 blues. The probability that he picks a blue is $\frac{2}{5}$. So the probability that the $RBRB$ case happens is $\frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{2}{5} = \frac{1}{30}$. Case 2 has a probability of $\frac{1}{30} \cdot 2 = \frac{1}{15}$ chance of happening.

Case 3: either $RBBR$ or $BRRB$. Let's find the probability that he picks the balls in the order of $RBBR$. The probability that the first ball he picks is red is $\frac{1}{2}$. Now there are 2 reds and 1 blue in the urn. The probability that he picks blue is $\frac{1}{3}$. There are 2 reds and 2 blues now. The probability that he picks a blue is $\frac{1}{2}$. Finally, there are 2 reds and 3 blues. The probability that he picks a red is $\frac{2}{5}$. So the probability that the $RBBR$ case happens is $\frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{2}{5} = \frac{1}{30}$. Case 3 has a probability of $\frac{1}{30} \cdot 2 = \frac{1}{15}$ chance of happening.

Adding up the cases, we have $\frac{1}{15} + \frac{1}{15} + \frac{1}{15} = \boxed{\frac{1}{5}}$. □

Hmmm, all our cases ended up having the same probability of occurring. No, this is not a coincidence. See if you can justify why this is the case, and generalize the problem!

Exercise 1.16 (MBMT 2018). There are 4 clubs at Blair: the Saxophone, Tambourine, Engineering, and Math clubs. Each club selects another club to be their archnemesis. What is the probability that no two clubs select each other as archnemeses?

Exercise 1.17 (MBMT 2020). Matthew Casertano and Fox Chyatte make a series of bets. In each bet, Matthew sets the stake (the amount he wins or loses) at half his current amount of money. He has an equal chance of winning and losing each bet. If he starts with 256, find the probability that after 8 bets, he will have at least 50.

1.5 Outcomes not occurring with equal probability

Sometimes outcomes will not occur with equal probability. This can be seen with unfair die or coins. Most of these problems can be solved with the help of Fact 1.1 and Fact 1.3.

Example 1.18 (HMMT 1999)
An unfair coin has the property that when flipped four times, it has the same nonzero probability of turning up 2 heads and 2 tails (in any order) as 3 heads and 1 tail (in any order). What is the probability of getting a head in any one flip?

Solution. Let p be the probability of flipping a head with this coin. We can write

$$\binom{4}{2} \cdot p^2 \cdot (1-p)^2 = 4p^3(1-p),$$

so $p = \boxed{\frac{3}{5}}$. □

Sometimes an intuitive way to think about this is taking the "weighted" probabilities and thinking about each them as a particular number of a simpler "unit" of probability.

Say we have a coin that flips heads with probability $2/5$ and flips tails with probability $3/5$. We can treat it like some event that gives out 5 possible different outcomes with equal probability, and 2 of them count as "heads," while the other three count as "tails." We've essentially reduced it to a probability with equally occurring outcomes question again.

Example 1.19
We have a coin that flips heads with probability $2/5$ and flips tails with probability $3/5$. What is the probability that after 4 flips, the number of times we flip heads is even?

Solution. We treat the coin like an event giving out 5 possible different outcomes with equal probability, and 2 of them count as "heads," while the other three count as "tails." We have 625 total outcomes. The number of outcomes that work can be computed to be $2^4 + \binom{4}{2} \cdot 2^2 \cdot 3^2 + 3^4 = 313$. Our answer is $\boxed{\frac{313}{625}}$. □

Exercise 1.20 (2010 AMC 12A P15). A coin is altered so that the probability that it lands on heads is less than $\frac{1}{2}$ and when the coin is flipped four times, the probability of an equal number of heads and tails is $\frac{1}{6}$. What is the probability that the coin lands on heads?

Exercise 1.21 (2015 AMC 12B P17). An unfair coin lands on heads with a probability of $\frac{1}{4}$. When tossed n times, the probability of exactly two heads is the same as the probability of exactly three heads. What is the value of n?

1.6 Symmetry

I'm about to go into some very handwavy stuff, but sometimes a very complicated problem can be greatly reduced through the use of a concept known as symmetry. You may be familiar with reflection symmetry or rotational symmetry in geometry, where a shape looks the same when it undergoes reflections or rotations. The idea of symmetry in probability is a bit of a more conceptual one, but still just as powerful.

A situation in probability, or just combinatorics in general, is a situation that functions the same way when looked at from multiple perspectives. A lot of times, the conditions that have to be satisfied for some outcome to occur, are basically analogous to the conditions that have to be satisfied for some other outcome to occur. This often allows us to conclude that the chances of several different possible outcomes are all the same right off the bat, saving us a lot of time and effort.

Here is a simple example.

Example 1.22
Flip a fair coin 5 times. What is the probability that there are more heads than tails?

Solution. We could treat this as a counting problem of sorts, but notice that the probability of flipping n heads is the same as the probability of flipping n tails, since the situation is symmetric (each time, heads and tails come up with equal probability). So the probability there are more heads than tails is just $\frac{1}{2}$. □

Competitions often give questions similar to this:

Example 1.23
Flip a fair coin 6 times. What is the probability that there are more heads than tails?

Solution. Be careful! The probability that there are more heads than tails is not $\frac{1}{2}$ this time. This is because there can be an equal number of heads and tails: this comes up with probability $\frac{\binom{6}{3}}{2^6} = \frac{5}{16}$. Otherwise, however, assuming that the number of heads and tails is not equal, then we will be able to see that the probability that there are more heads than tails, is the same as the probability that there are more tails than heads. Our answer is $\frac{5}{16} \cdot \frac{1}{2} = \frac{5}{32}$. □

In this type of question, we often find that assuming that some specific outcome does not occur, then the probabilities of the other outcomes must occur with equal probability due to symmetry. So we might end up subtracting the probability that this specific outcome does not occur from 1, then dividing by some integer.

These are all simple problems with easy-to-find solutions that do not require symmetry. The power of symmetry is best seen with a tougher example.

Example 1.24 (SADGIME)
Justin rolls a fair 6 sided die numbered with the integers 1 through 6 repeatedly until he rolls the same number on three consecutive rolls, at which point he stops. Given that his first roll is a 6, find the probability that his last roll is a 6.

Solution. Let our desired probability be P. There is a $\frac{1}{36}$ probability that Justin's second and third rolls are both 6. Otherwise, Justin rolls a number between 1 to 5 before he gets to his third roll. If this occurs, the probability of ending with a 6 thereafter is $\frac{1-P}{5}$ by symmetry. We have that

$$P = \frac{1}{36} + \frac{35}{36}\left(\frac{1-P}{5}\right).$$

Solving this yields $P = \boxed{\frac{8}{43}}$. □

Do you notice what we did here? We set a variable P to be our desired probability. The probability Justin rolls a number between 1 to 5 before he gets to his third roll is $\frac{35}{36}$. Let this number be n.

Now forget about all the rolls that have occurred before - they are irrelevant. This is the exact same situation as what we started with, only our first roll was n instead of 6. So, the probability that from here, our last roll is n is P, by symmetry! Now, here's the zinger: our last roll isn't n, all of the five other outcomes occur with equal probability due to another application of symmetry. Thus, the probability that we roll a 6 on our last roll is always going to be $\frac{1-P}{5}$, assuming that our second and third rolls are not both 6.

Exercise 1.25 (MBMT). Bread draws a circle. He then selects four random distinct points on the circumference of the circle to form a convex quadrilateral. Kwu comes by and randomly chooses another 3 distinct points (none of which are the same as Bread's four points) on the circle to form a triangle. Find the probability that Kwu's triangle does not intersect Bread's quadrilateral, where two polygons intersect if they have at least one pair of sides intersecting.

Exercise 1.26 (DMC). Ryan has an infinite supply of slips and a spinner with letters O, S, and T, where each letter is equally likely to be spun. Each minute, Ryan spins the spinner randomly, writes on a blank slip the letter he spun, and puts it in a pile. Ryan continues until he has written all 3 letters at least once, at which point he stops. What is the probability that after he stops, he can form the words $OTSS$ and $TOST$ using 4 distinct slips from the pile? (Ryan may reuse slips he used for one word in forming the other.)

2 Expected Value

Previously, we've discussed how random events can have certain probabilities of happening. Some of these probabilities might be really high, meaning the event happens quite often, while other probabilities might be really low. A possible "table of probabilities" listing out events and the chances that they happen might look something like this:

Which class will be chosen to participate in the school play?

$$\begin{cases} \text{Mrs. Apple's Class: } \dfrac{1}{6} \text{ probability.} \\ \text{Mr. Bean's Class: } \dfrac{3}{6} \text{ probability.} \\ \text{Mrs. Cherri's Class: } \dfrac{2}{6} \text{ probability.} \end{cases}$$

In the table above, the left side shows a list of **outcomes**, while the right side shows a list of **probabilities.** Using these probabilities, we can determine which class is least or most likely to be chosen, for example. However, there isn't any sense of "value" to any of these outcomes. To get an idea of where we might be going with this, take a look at this table of probabilities:

$$\begin{cases} \text{One Space: } \frac{1}{20} \text{ probability.} \\ \text{Two Spaces: } \frac{1}{10} \text{ probability.} \\ \text{Four Spaces: } \frac{2}{10} \text{ probability.} \\ \text{Seven Spaces: } \frac{6}{10} \text{ probability.} \end{cases}$$

In this case, not only can we tell what outcome will be least or most common, but we also have a sense that, *on average*, we will tend to move pretty far, given how likely the "7" outcome is. On the other hand, with this table of probabilities, we can observe the exact opposite:

$$\begin{cases} \text{One Space: } \frac{4}{20} \text{ probability.} \\ \text{Two Spaces: } \frac{3}{10} \text{ probability.} \\ \text{Four Spaces: } \frac{2}{10} \text{ probability.} \\ \text{Seven Spaces: } \frac{1}{10} \text{ probability.} \end{cases}$$

Here, since the "4" and "7" outcomes don't occur as often, we can expect that this table will *generally* produce lower outcomes than the first table. This isn't to say that the second table will always produce a lower outcome – rather, if we had the ability to choose which table to use, for example, we would most likely choose the first one because its expected value is greater.

See that word we just used? Expected value? This section is all about taking this intuitive concept of some machines being more valuable than others, and making it more rigorous. What if the comparisons in expected values was much vaguer? What if the table of values was much more complex? What if the table of values was infinitely large, even?

2.1 An Aside on Random Variables

On your first read-through, this section might not make a whole lot of sense. Feel free to skip it for now, and come back when we discuss Linearity of Expectation.

Before moving on, we need to have a rigorous idea of what kind of tables of probabilities we wish to analyze. These machines that produce numerical values by chance are called **random variables,** which we give a definition of below:

Random Variables

A random variable X can be thought of as a number whose state is dependent on chance. (Ever heard of Schrödinger's Cat?) For example, you might define a random variable Y like so:

$$Y = \pi \text{ if it is cloudy, 3 if it is sunny, and 5 if it is neither.}$$

A more natural example of a random variable is a die D; you roll the die, and whatever number comes up is the value of D. The random variable D might be more rigorously defined as such:

$$D = \begin{cases} 1 \text{ with probability } \frac{1}{6} \\ 2 \text{ with probability } \frac{1}{6} \\ 3 \text{ with probability } \frac{1}{6} \\ 4 \text{ with probability } \frac{1}{6} \\ 5 \text{ with probability } \frac{1}{6} \\ 6 \text{ with probability } \frac{1}{6} \end{cases}$$

The key with random variables is the following: **their outcomes must be numbers.** It doesn't matter whether they are rational numbers, irrational numbers, or even imaginary numbers – the outcomes just can't be something like "yes" or "squareman orz".

Exercise 2.1. Which of the following is **not** a random variable?

1. Choose a real number at random. Output 0 if it is even, 1 if it is odd, and 2 otherwise.
2. Choose a point uniformly at random on the coordinate plane, and output its distance from the origin.
3. Choose a question at random, and output 0 if the answer is yes, 1 if the answer is no, and $\emptyset$ otherwise.
4. Choose a bank uniformly at random across the world, and output the number of letters in its name.

Now that we know what we're working with, let's try to compute some expected values!

2.2 How to Expect

2.2.1 The Main Act

Here's our big definition of expected value:

Expected Value

Suppose we have a random variable X with outcomes E_1, E_2, . . ., E_k and respective probabilities $p_1, p_2, \ldots, p_k$ for those outcomes. Then the ***expected value*** of X, denoted $\mathbb{E}[X]$ is equal to:

$$\mathbb{E}[X] = E_1 p_1 + E_2 p_2 + \ldots + E_k p_k.$$

That definition is a little... complex. Let's try to gradually digest it with a few examples:

Example 2.2

What is the expected value of the outcome of a standard six-sided dice roll?

Proof. First, we need to make sure we understand what our random variable is. In this case, it's the outcome of the dice, which we use the letter X to represent.

Now, what are the outcomes of X? If you know what a six-sided die is, you should know that the outcomes are $\{1, 2, 3, 4, 5, 6\}$. Great! Now, what are the probabilities of each of these outcomes? With our knowledge of probability from the previous section, we should be able to determine that each probability equals $\frac{1}{6}$. To be precise, the random variable X is defined like so:

$$\mathrm{X} = \begin{cases} 1 \text{ with probability } \frac{1}{6} \\ 2 \text{ with probability } \frac{1}{6} \\ 3 \text{ with probability } \frac{1}{6} \\ 4 \text{ with probability } \frac{1}{6} \\ 5 \text{ with probability } \frac{1}{6} \\ 6 \text{ with probability } \frac{1}{6} \end{cases}$$

You might be able to tell by intuition that the expected value of X should be the "middle" of this set, or $\frac{7}{2}$. However, let's use the definition of expected value to solve this problem instead. In this case, our **events** are:

$$E_1 = 1,\ E_2, = 2,\ E_3 = 3,\ E_4 = 4,\ E_5 = 5,\ E_6 = 6,$$

and our **probabilities** are:

$$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6}.$$

Therefore, we can compute using the definition that the expected value of X is:

$$\mathbb{E}[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \boxed{\frac{7}{2}.}$$

□

Remark 2.3. If you look at the huge table we made fully defining X, you might recognize that expected value is simply the sum of the products you get by multiplying across each row.

Exercise 2.4. Confirm that the expected value of the outcome of a standard n-sided dice roll, in terms of n, is $\frac{n+1}{2}$. (You can either use your intuition, or use the expected value definition. Try using both!)

You might have noticed that, while $\frac{7}{2}$ is the *expected value* of the six-sided dice roll, it is impossible to roll exactly a $\frac{7}{2}$ on a six-sided dice. This brings us to the following observation:

> Expected value is **not** the most likely outcome! In fact, expected value doesn't even have to be one of the outcomes!

Here are a few more examples, just so we can get an idea of how this works:

Example 2.5
Merlin flips a fair coin 5 times. What is the expected number of heads he flips?

Proof. Here, we define the random variable X to equal the number of heads flipped. (Notice that we intentionally define X to equal a number, not something like "4 heads".)

Now, we simply need to more rigorously define what X is:

$$X = \begin{cases} 0 \text{ with probability } \frac{1}{32} \\ 1 \text{ with probability } \frac{5}{32} \\ 2 \text{ with probability } \frac{10}{32} \\ 3 \text{ with probability } \frac{10}{32} \\ 4 \text{ with probability } \frac{5}{32} \\ 5 \text{ with probability } \frac{1}{32} \end{cases}$$

To finish, we simply plug in these outcomes and probabilities into our expected value "formula":

$$\mathbb{E}[X] = 0 \cdot \frac{1}{32} + 1 \cdot \frac{5}{32} + 2 \cdot \frac{10}{32} + 3 \cdot \frac{10}{32} + 4 \cdot \frac{5}{32} + 5 \cdot \frac{1}{32} = \boxed{\frac{5}{2}}.$$

□

Remark 2.6. You may have had the an idea similar to this when reading this problem:

> *"Well, each coin should give $\frac{1}{2}$ heads on average, so 5 coins should give $\frac{5}{2}$ heads!"*

This logic, based on the answer we achieved in this solution, is actually correct! We'll touch more on this idea when we reach Linearity of Expectation.

Here's a less obvious example:

Example 2.7
A deck of 15 cards contain i copies of a card labeled i, for integers $1 \leq i \leq 5$. Ehuang draws a card from this deck at random. What is the expected value of the number on the card he draws?

Proof. As before, we define X to be a random variable equal to the number on Ehuang's card. The probabilities look like this:

$$X = \begin{cases} 1 \text{ with probability } \frac{1}{15} \\ 2 \text{ with probability } \frac{2}{15} \\ 3 \text{ with probability } \frac{3}{15} \\ 4 \text{ with probability } \frac{4}{15} \\ 5 \text{ with probability } \frac{5}{15} \end{cases}$$

Then, we just compute:

$$\mathbb{E}[X] = 1 \cdot \frac{1}{15} + 2 \cdot \frac{2}{15} + 3 \cdot \frac{3}{15} + 4 \cdot \frac{4}{15} + 5 \cdot \frac{5}{15} = \boxed{\frac{11}{3}}.$$

□

Remark 2.8. We can sort of expect that the expected value would be somewhat "above average", given that higher valued cards are weighted more than lower valued cards. In general, expected value can be thought of as a weighted average of sorts – just how test grades are more weighted than homework grades, outcomes with higher probabilities are more weighted than those with lower probabilities.

Exercise 2.9. What is the expected value of this deck of $\frac{n(n+1)}{2}$ cards, for general choices of n? *(Hint: Make sure you reach the answer of $\frac{n(n+1)}{3}$ in the end.)*

Exercise 2.10. A deck of 15 cards contains i copies of a card labeled $6 - i$, for integers $1 \le i \le 5$. Farley draws a card from this deck at random. What is the expected value of this card? *(Think about this problem in terms of weighted averages to convince yourself that the answer should be at most 3.)*

At this point, we introduce a quick sidenote on gambling. (Don't gamble kids! unless you win in that case gambling is good)

2.2.2 An Aside: "Fair Games"

A more common usage of expected value is determining the cost of a "fair game". Suppose that you played a game, in which playing this game always gives you 4 dollars. If this game were fair, how much should you pay to play? Obviously, the answer should be 4, since otherwise you would either ruin the economy or steal people's money.

Now, what if this game involved a bit of chance? Suppose that you have a $\frac{1}{3}$ chance of receiving 2 dollars, but a $\frac{2}{3}$ chance of receiving 5 dollars. How much should you pay to play this game?

To answer this question, it is pretty intuitive that the amount you pay to play should equal the expected amount you would earn from this game. Thus, the price of this game in order to make it fair would equal:

$$\mathbb{E}[X] = 2 \cdot \frac{1}{3} + 5 \cdot \frac{2}{3} = \frac{12}{3} = \boxed{4.}$$

In other words, to determine the price of a fair game, just set it equal to the expected value of the game! At least, that's what would intuitively be true...

Example 2.11
A certain game costs k dollars to play. There is a $\frac{1}{10^{2021}}$ chance of receiving 10^{2022} dollars, and any other outcome gives 1 dollar. How much should the value of k be to make this game fair?

Proof. As in before, we simply have to compute the expected earnings of this game:

$$10^{2022} \times \frac{1}{10^{2021}} + 1 \times \left(1 - \frac{1}{10^{2021}}\right). = \boxed{11 - \frac{1}{10^{2021}.}} \approx 11.$$

Thus, it would be reasonable to charge 11 dollars for this game, right? Right? □

The issue with using "ideal math" logic in this case is that life often doesn't allow for extremely small chances to occur. If you were to actually play this game in real life, you would most likely run into some serious debt before actually earning any money. An extreme example of this can be found here:

Example 2.12
A game involves flipping a coin and stopping as soon as it lands heads. Suppose that someone playing this game flipped the coin k times before stopping. Then this player receives 2^k dollars in return. How much should this game cost to make it fair?

Proof. In this case, even though there are infinitely many events, we can still compute expected value. Here are a couple pairs showing the relationship between earnings and the probabilities of receiving those earnings:

$$(2, \frac{1}{2}),\ (4, \frac{1}{4}),\ (8, \frac{1}{8}),\ (16, \frac{1}{16}), \ldots$$

Now, to compute expected value, we simply multiply each outcome by the probability it occurs, giving us:

$$\sum_{k=1}^{\infty} 2^k \cdot \frac{1}{2^k} = \sum_{k=1}^{\infty} 1 = \boxed{\infty.}$$

□

Uh, that doesn't make much sense. Who would actually pay infinitely many dollars to play this game that has a 50% chance of giving you just 2 dollars?

The reason why we reach this unreasonable number is because, technically, there *is* a chance that one could receive 2^{999999} dollars by playing this game. Since 2^{999999} is such a huge number, it ends up carrying on much more weight than it should, thus making this game "theoretically" infinitely valuable. However, just as no human would play the game in Example 2.11, no human would play this game, either!

In other words, just because the expected value is high, does not mean that you should play the game over and over.[1]

[1] Of course, math doesn't care about debt – you can just keep playing even when you reach negatives, and expected value tells you that *generally* you should get some earnings in the end.

2.2.3 Weighted in Your Favor

To elaborate just a little more on this weighted average concept, here's one more problem:

Example 2.13
Let r, s, and t be nonnegative real numbers such that $r + s + t = 1$. What is the minimum possible value of $2r + 3s + 5t$?

Proof. There are many simple ways to think about this problem, but we'll go about this by thinking about this problem in terms of probability.

Suppose we had a random variable X defined as follows:

$$X = \begin{cases} 2 \text{ with probability } r \\ 3 \text{ with probability } s \\ 5 \text{ with probability } t \end{cases}$$

Clearly this is valid, since $r + s + t$ guarantees that our probabilities, well, work. Now, the quantity $2r + 3s + 5t$ simply represents:

$$\mathbb{E}[X] = 2r + 3s + 5t = \text{a weighted average of } 2, 3, 5.$$

If we really wanted to make our average as low as possible, what would we do? Obviously, we would want to put all of our weight on the lowest number, which is 2. Thus, we would want $r = 1$ and $s = t = 0$, making the minimum possible value simply $\boxed{2.}$

In general, the weighted average of some set of real numbers $x_1, x_2, \ldots, x_k$ is at least the minimum of all k numbers and at most the maximum. $\square$

Suppose x_1, x_2, x_3, and x_4 are nonnegative real numbers satisfying $x_1 + x_2 + x_3 + x_4 = 10$. What is the maximum possible value of $x_1 + 2x_2 + 3x_3 + 4x_4$?

Remark 2.14. Here's an idea to think about. Suppose, somehow, we managed to figure out that the expected value of some random variable X was equal to 3.14, and that X could only take on positive integer outcomes. Then we should be able to argue that X takes on some outcome that is at most 3, right?

This line of thinking is the motivation for the **Probabilistic Method.** Suppose a problem asks you to show that you can find some configuration in which some quantity equals at least 11, and that quantity is always an integer.

Instead of actually constructing the configuration, you can choose the configuration purely at random, and argue that the expected value of this quantity is $\sqrt{120}$. This would then prove that there existed *at least one* configuration in which the quantity was at least 11, as desired!

At this point, we should be pretty familiar with how to compute expected value. You might come across more complicated examples in which the probabilities of each event are relatively tough to figure out, but the ideas behind each problem should generally go along the lines of something like this:

1. List out all of the outcomes. (Hopefully there's not too many!)

2. Find the probability each outcome occurs. (Every single one of these calculations is pretty much its own probability problem.)

3. Take the weighted average to find the expected value.

When problems flood you with hundreds of cases, most of the time finding the expected value is going to be a huge bash that isn't worth going through. However, there is one very clever trick that can occasionally help out...

2.3 Linearity of Expectation

2.3.1 Proof

We begin with the following scenario:

> Suppose you had two random variables X and Y. Our goal is to compute the expected value of $X+Y$ – that is, suppose you went through the random process of doing X, then doing Y, and summed the two results. What would the expected value of the sum be?

We should first make it clear that the value of $\mathbb{E}[X+Y]$ (the expected value of this sum) is not obviously $\mathbb{E}[X]+\mathbb{E}[Y]$. This does make a lot of sense when the two events X and Y are independent; for example, X is the roll of a six-sided die and Y is a roll of a four-sided die.

But what if X is a random number between 1 and 6, while Y is a random number between 1 and n, where n is the resulting number achieved from X? In this case, it seems harder to believe that the expected value of $X+Y$ would simply be the expected value of X plus the expected value of Y.

In order to resolve this issue, let's try to find a general formula for the expected value of $X+Y$, for two (not necessarily independent!) random variables X and Y. For now, suppose X and Y have outcomes and probabilities that look like this:

$$X : \{(E_1, p_1),\ (E_2, p_2),\ (E_3, p_3)\}.$$

$$Y : \{(F_1, q_1),\ (F_2, q_2),\ (F_3, q_3),\ (F_4, q_4)\}.$$

(Note that even when Y is dependent of X, it is still possible to compute the probability of each outcome of Y.) In this case, X has three outcomes E_1, E_2, E_3, while Y has four outcomes. Notice that, by definition:

$$p_1 + p_2 + p_3 = q_1 + q_2 + q_3 + q_4 = 1.$$

Now, the list of all pairs of events and probabilities for $X+Y$ would look pretty messy. Here's the full list:

$$(E_1+F_1, p_1q_1),\ (E_1+F_2, p_1q_2),\ (E_1+F_3, p_1q_3),\ (E_1+F_4, p_1q_4).$$

$$(E_2+F_1, p_2q_1),\ (E_2+F_2, p_2q_2),\ (E_2+F_3, p_2q_3),\ (E_2+F_4, p_2q_4).$$

$$(E_3+F_1, p_3q_1),\ (E_3+F_2, p_3q_2),\ (E_3+F_3, p_3q_3),\ (E_3+F_4, p_3q_4).$$

Trying to use the weighted average expected value formula and summing all 12 of these terms doesn't exactly look like fun. However, we can make the following observation:

The **expected value** of the event-probability pair:

$$(E_1 + F_1, p_1q_1)$$

is equal, numerically, to the **expected value** of the collection of these two event-probability pairs:

$$(E_1, p_1q_1),\ (F_1, p_1q_1).$$

Basically, we can split up every single one of these 12 pairs into two separate pairs, without changing the expected value, giving us 24 pairs altogether. This, of course, makes it so that the probability of event E_1 occurring is now split up into a bunch of different cases. Let's focus on just the ordered pairs that have outcome E_1:

$$(E_1, p_1q_1),\ (E_1, p_1q_2),\ (E_1, p_1q_3),\ (E_1, p_1q_4).$$

We are looking for the weight that outcome E_1 has.[2] This is simply the sum of all four of these weights, which is:

$$p_1q_1 + p_1q_2 + p_1q_3 + p_1q_4 = p_1(q_1 + q_2 + q_3 + q_4) = p_1.$$

Therefore, we may condense all 4 of these pairs:

$$(E_1, p_1q_1),\ (E_1, p_1q_2),\ (E_1, p_1q_3),\ (E_1, p_1q_4)$$

into just a single pair:

$$(E_1, p_1).$$

See where we're going with this? We can condense other groupings of four E_is similarly, to create these pairs:

$$(E_2, p_2),\ (E_3, p_3),\ (E_4, p_4).$$

And then we can condense the three groupings of F_1s to make these pairs:

$$(F_1, q_1),\ (F_2, q_2),\ (F_3, q_3).$$

Thus, we wish to compute the weighted average of the following "fake random variable":

$$Z = (E_1, p_1), (E_2, p_2), (E_3, p_3), (E_4, p_4), (F_1, q_1), (F_2, q_2), (F_3, q_3).$$

Now we simply compute the "expected value" (or rather, weighted average):

$$\mathbb{E}[X + Y] = \mathbb{E}[Z] = E_1p_1 + E_2p_2 + E_3p_3 + E_4p_4 + F_1q_1 + F_2q_2 + F_3q_3 = \mathbb{E}[X] + \mathbb{E}[Y].$$

Let's take a minute to simply digest the process we have gone through:

1. In an attempt to compute the expected value of two (not necessarily independent) random variables X and Y, we have found the following ordered pairs describing the random variable $Z = X + Y$.

[2] This can be a little confusing. In this case, the probability that outcome E_1 occurs doesn't represent *anything*. When we split up those 12 pairs to make 24 pairs, we essentially created a fake random variable Z (since the sum of its probabilities is not 1), but a fake random variable Z that has the property that $\mathbb{E}[X + Y] = \mathbb{E}[Z]$. We essentially want to compute the weighted average of Z, in a case where the weights don't actually sum to 1.

2. Then, in order to simplify our computations, we split up each of these ordered pairs and then regrouped them in a simpler manner.
3. Finally, when we computed the weighted average again, we noticed that $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

The key is the following: At no point did we use the fact that X and Y were independent. Thus, even if X and Y are heavily dependent on each other, we still have the following crucial claim:

Theorem 2.15 (Linearity of Expectation)
For two (not necessarily independent) random variables X and Y, we have:

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

More generally, a simple induction argument can give the following:

Theorem 2.16 (Generalized Linearity of Expectation)
For some (not necessarily independent) random variables $X_1, X_2, \ldots, X_n$, we have:

$$\mathbb{E}[X_1 + X_2 + \ldots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \ldots + \mathbb{E}[X_n].$$

2.3.2 Applications!

Let's start using this to slaughter some expected value problems!

Example 2.17
Annabel rolls a six-sided die 50 times. What is the expected value of the sum of all numbers that come up?

Proof. It seems fairly obvious that, since the expected value of the number coming up on a six-sided die is simply $\frac{7}{2}$, the expected value of the sum of numbers on 50 of these dice would be $50 \times \frac{7}{2} = 175$. However, for the sake of getting some experience with our new theorem, let's solve this with a bit more rigor.

Let X be the random variable that equals the sum of all numbers that come up on Annabel's dice. Our goal is to compute $\mathbb{E}[X]$. However, computing the exact definition of X (which would require finding the probability that the sum equals, say, 199) would be nearly impossible. Thus, to simplify our work, we use Linearity of Expectation.

Let X_1 be an **indicator variable,** so that X_1 is a random variable that indicates the number appearing on the first die. Then, let X_2 be defined similarly, so that in general, X_i is a random variable indicating the number appearing on the ith die, from $1 \leq i \leq 50$. The key is in the following:

$$X = X_1 + X_2 + \ldots + X_{50}.$$

In other words, the random variable X can be found by simply running random variables $X_1, X_2, \ldots, X_{50}$, and summing the randomly chosen results. By our beloved Linearity of Expectation theorem, we find the following:

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \ldots + \mathbb{E}[X_{50}].$$

Why did we do this? Remember that our goal was to find $\mathbb{E}[X]$, but we ran into a problem when the definition of X was hard to find. If we rewrite $\mathbb{E}[X]$ as a much longer summation, we discover that each individual component of the summation is trivially easy to compute! Just notice that $\mathbb{E}[X_i] = \frac{7}{2}$ for all i. Therefore,

$$\mathbb{E}[X] = 50 \times \frac{7}{2} = \boxed{175.}$$

□

We used the term "indicator variable" in that proof just then. What does that mean?

Indicator Variables

When we want to compute the expected value of a really complicated random variable X, a good trick is to identify X as the sum of a bunch of simpler random variables X_i. These X_is are our indicator variables.

It might be easier to understand why we describe them as "indicating" with the following example:

Example 2.18

Ornikez writes the digits 0 through 9 in a completely random order. What is the expected number of digits i that are ith in line? For example, the string 1053862479 has only 2 digits i that are ith in line, those being 6 and 9.

Proof. As always, we define the X to be a random variable equal to the number of digits i that are ith in line, and wish to discover the value of $\mathbb{E}[X]$. In this case, it is much harder to define the exact value of X, so we look back at Linearity of Expectation for help. Is there a way to write X as the sum of a bunch of indicator variables X_i, each of which has a much easier expected value to compute?

It turns out that the answer is yes! We define ten indicator variables $X_0, X_1, \ldots, X_9$ as follows:

$$X_i = \begin{cases} 1 \text{ if the digit } i \text{ is in the } i\text{th position.} \\ 0 \text{ otherwise.} \end{cases}$$

Obviously, the number of digits i that are ith in line is simply the sum of all X_is, meaning that:

$$X = X_0 + X_1 + \ldots + X_9.$$

Remark 2.19. You might be asking how we came up with these indicator variables X_i. Think of it this way: Say that you gave a robot some random example of Ornikez's string. How would that robot compute the number of digits i that are ith in line?

Well, it would simply go through each digit i and ask, "Is this digit in the ith position?" If yes, then output 1. If no, then output 0. Finally, it would sum all ten of these results to give the final answer.

You can see why we call the X_is indicator variables now. Each X_i essentially indicates whether the digit i is in the ith position or not, and we count the number of indicators that say "yes". Note that this means that most indicator variables we define will either have 0 or 1 as an output.

Okay, we now have an identity writing X as the sum of a bunch of indicator variables, which by Linearity of Expectation implies that:

$$\mathbb{E}[X] = \mathbb{E}[X_0] + \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_9].$$

You might interject by noting that the X_is are not independent of each other, so you can't just sum them like this. However, remember our proof of Linearity of Expectation? No flaws show up when the X_is are dependent, so this summing of expectations still holds!

Now all we have to do is compute $\mathbb{E}[X_i]$. This is much simpler, since now X_i only has two events, while X had ten. All we seek is the probability that the digit i is in the ith position.

Exercise 2.20. Use some logic to convince yourself that the probability that digit i is in the ith position equals $\frac{1}{10}$.

Therefore, we can redefine X_i as such:

$$X_i = \begin{cases} 1 \text{ with probability } \frac{1}{10} \\ 0 \text{ with probability } \frac{9}{10}. \end{cases}$$

Obviously $\mathbb{E}[X_i] = \frac{1}{10}$ for all i, so:

$$\mathbb{E} = \mathbb{E}[X_0] + \mathbb{E}[X_1] + \ldots + \mathbb{E}[X_9] = \frac{1}{10} + \frac{1}{10} + \ldots + \frac{1}{10} = \boxed{1.}$$

If you feel like this is cheating, you're completely right! □

Exercise 2.21. Squorz randomly arranges the numbers from 1 to 5 in a line. What is the expected number of numbers that are greater than all numbers before it?
(Hint: If you ever find the number 137*, you should be correct.)*

Let's kill another problem with Linearity of Expectation:

Example 2.22

Five rooks are placed on an 8×8 chessboard, each independently at random. (It is permitted to have multiple rooks on the same square.) What is the expected number of squares on this chessboard that are not being attacked[a] by any rook?

[a] A square is said to be attacked by a rook if it is in the same row or column as the rook.

Proof. Just as before, we define X to be a random variable equal to the number of squares on the chessboard not being attacked by a rook. Just as we did in the previous example, we ask the following question to help determine what our indicator variables should be:

> How would a robot determine the number of squares not being attacked by any rook?

The answer is that it would most likely go through every single square of the board and ask the question, "Is this square attacked by a rook?" Let's try programming this robot with 64 indicator variables $X_1, X_2, \ldots, X_{64}$, defined as follows:

$$X_i = \begin{cases} 1 \text{ if the } i\text{th square is not attacked by any rook.} \\ 0 \text{ otherwise.} \end{cases}$$

Then, we have the following:

$$X = X_1 + X_2 + \ldots + X_{64},$$

so we also have:

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \ldots + \mathbb{E}[X_{64}].$$

All we have to do now is compute $\mathbb{E}[X_i]$, meaning we simply want the probability that any given square is not being attacked by a rook.

> *Remark 2.23.* Make sure you define your indicator variables so that each X_i has an equal expected value, meaning they all function similarly. Otherwise, computing this sum would be much harder than just finding $64\mathbb{E}[X_1]$.

To compute this probability, we note that for every square i, no rook attacks this square if and only if every rook is on one of 49 other squares of the board. Thus, the probability that none of these five rooks attack square i equals $\left(\frac{49}{64}\right)^5$, so:

$$X_i = \begin{cases} 1 \text{ with probability } \frac{49^5}{64^5} \\ 0 \text{ with probability } 1 - \frac{49^5}{64^5}. \end{cases}$$

Thus, $\mathbb{E}[X_i] = \frac{49^5}{64^5}$, so:

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \ldots + \mathbb{E}[X_{64}] = 64 \times \mathbb{E}[X_i] = 64 \times \frac{49^5}{64^5} = \boxed{\frac{49^5}{64^4}.}$$

□

Here is a slightly less obvious application:

Example 2.24

A deck of 52 cards contains exactly 4 aces. If the cards are placed on a table in a completely random order, what is the expected number of cards that are placed before the second ace placed onto the table?

Proof. Let X be a random variable equal the the number of cards before the second ace is placed onto the table. Here's the main idea:

> For every single card that is not an ace, define a random variable X_i as follows:
>
> $$X_i = \begin{cases} 1 \text{ if card } i \text{ is before the second ace.} \\ 0 \text{ otherwise.} \end{cases}$$
>
> Observe that we then have 48 indicator variables.

Obviously, we then have that:

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \ldots + \mathbb{E}[X_{48}] = 48\mathbb{E}[X_1].$$

Now we simply need to determine the probability that any given card is before the second ace. To do this, notice that the four aces essentially split the remaining set of 48 cards into 5 (possibly empty) sets. We want to know the probability that any given card is contained within the first 2 sets.

Exercise 2.25. A card is placed into one of 5 sets. What is the probability that it is placed within one of the first 2? (Don't overthink this: the answer is just $\frac{2}{5}$.)

Therefore, we have:

$$\mathbb{E}[X] = 48 \times \frac{2}{5} = \frac{96}{5}.$$

To finish, notice that the question does not ask for the number of non-ace cards before the second ace. Thus, to find the actual answer, we simply need to add in the first ace, giving $\boxed{\frac{101}{5}}$ as our final answer. □

Before we finish, let's take a look at one of my favorite applications of Linearity of Expectation:

Example 2.26
Consider an infinite number of horizontal lines, all spaced 1 inch apart, thus dividing the plane into infinitely many rectangular regions with width 1.

Suppose a line segment of length 1 is dropped randomly on this plane. What is the probability that this line segment hits one of the lines?

Proof. At first, it may seem that Linearity of Expectation cannot help us in this problem – after all, no mention of expected value is visible, besides a vague reference to probability. Let's try to take that vague reference and use it as much as possible.

Instead, what if we asked for the *expected number of intersections of the line segment with the lines?* Since the line segment either intersects the lines at 0 or 1 location, finding this expected value should give us the answer!

More precisely, let's define the random variable X to equal the number of intersections of the line segment with the lines. Then we have:

$$\mathbb{E}[X] = \text{the answer}.$$

Okay, so how does this help? Let's call our line segment of length 1 "ℓ", and suppose we divided ℓ into three congruent pieces of length $\frac{1}{3}$, which we label a, b, and c. Also, let's define A, B, and C to be random variables equal to the number of intersections that the tinier pieces a, b, and c have with the lines, respectively. Observe that:

$$X = A + B + C \implies \mathbb{E}[X] = \mathbb{E}[A] + \mathbb{E}[B] + \mathbb{E}[C].$$

Now here comes the big trick. Notice that $\mathbb{E}[A]$ really just asks about what happens when you drop a line segment of length $\frac{1}{3}$, so we can completely ignore the line segment of length 1 we had earlier. Consider the following (seemingly unrelated) scenario:

> Drop an equilateral triangle of length $\frac{1}{3}$ onto the same plane. What is the expected number of intersections of this equilateral triangle with the lines?

To solve this problem, let's define Y to be a random variable equal to the number of intersections with the equilateral triangle. If we say E_1, E_2, and E_3 are random variables equal to the number of intersections of each of the three sides with the lines, then we have:

$$Y = E_1 + E_2 + E_3 \implies \mathbb{E}[Y] = \mathbb{E}[E_1] + \mathbb{E}[E_2] + \mathbb{E}[E_3].$$

But wait! Notice that $\mathbb{E}[E_i]$ is the exact same as $\mathbb{E}[A]$, since both describe what happens when a line segment of length $\frac{1}{3}$ is dropped. Therefore,

$$\mathbb{E}[X] = \mathbb{E}[A] + \mathbb{E}[B] + \mathbb{E}[C] = \mathbb{E}[E_1] + \mathbb{E}[E_2] + \mathbb{E}[E_3] = \mathbb{E}[Y].$$

Great! So now, in order to find the answer to our original problem, we can instead just compute the expected number of intersections this equilateral triangle has with the lines. But, uh, that doesn't exactly seem much easier.

No need to fear, however. We can see that, in general, if we define Z to be a random variable equal to the number of intersections of a randomly dropped regular n-gon of perimeter 1 with the lines, then:

$$\mathbb{E}[X] = \mathbb{E}[Z].$$

Previously, we had worked with $n = 3$, but we can really use whatever n we want! In fact, we can even work with **infinitely large** values of n.

What happens when n grows to infinity? Our regular n-gon of perimeter 1 now becomes a circle of circumference 1. And at this point, finding the expected value should not be so hard:

> **Exercise 2.27.** A circle of circumference 1 has a radius of $\frac{1}{2\pi}$. If we were to drop this circle onto our grid of lines, what is the probability that the circle would meet one of the lines? (Make sure you get $\frac{1}{\pi}$ as your answer!)

Now, to finish, we just need to find the **expected value** of the **number** of intersections with this circle. If $\mathcal{C}$ is a random variable describing the number of intersections with the circle, we have:

$$\mathcal{C} = \begin{cases} 0 \text{ with probability } 1 - \frac{1}{\pi}. \\ 1 \text{ with probability } 0. \\ 2 \text{ with probability } \frac{1}{\pi}. \end{cases}$$

Thus, the expected value of $\mathcal{C}$ (which is the same as that of X) must equal:

$$\mathbb{E}[X] = \mathbb{E}[\mathcal{C}] = \boxed{\frac{2}{\pi}.}$$

If you don't think this is a wonderful proof, I don't know what to tell you. □

3 Problems

Minimum is [30 ♟]. Problems denoted with ♜ are required. (They still count towards the point total.)

[2 ♟] Problem 1 (AIME 2010 I Problem 1). Maya lists all the positive divisors of 2010^2. She then randomly selects two distinct divisors from this list. Find the probability that exactly one of the selected divisors is a perfect square.

[3 ♟] Problem 2 (2016 AIME I P2). Two dice appear to be standard dice with their faces numbered from 1 to 6, but each die is weighted so that the probability of rolling the number k is directly proportional to k. Find the probability of rolling a 7 with this pair of dice.

[3 ♟] Problem 3 (2012 AMC 10A Problem 20). A 3×3 square is partitioned into 9 unit squares. Each unit square is painted either white or black with each color being equally likely, chosen independently and at random. The square is the rotated 90° clockwise about its center, and every white square in a position formerly occupied by a black square is painted black. The colors of all other squares are left unchanged. What is the probability that the grid is now entirely black?

[4 ♟] Problem 4 (2013 AIME 1 Problem 6). Melinda has three empty boxes and 12 textbooks, three of which are mathematics textbooks. One box will hold any three of her textbooks, one will hold any four of her textbooks, and one will hold any five of her textbooks. If Melinda packs her textbooks into these boxes in random order, the probability that all three mathematics textbooks end up in the same box can be written as $\frac{m}{n}$, where m and n Are relatively prime positive integers. Find $m + n$.

[4 ♟] Problem 5 (AMC 12A 2010 Problem 16). Bernardo randomly picks 3 distinct numbers from the set $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and arranges them in descending order to form a 3-digit number. Silvia randomly picks 3 distinct numbers from the set $\{1, 2, 3, 4, 5, 6, 7, 8\}$ and also arranges them in descending order to form a 3-digit number. What is the probability that Bernardo's number is larger than Silvia's number?

[4 ♟] Problem 6 (DMC). Bill and Ben each have 2 fair coins. Each minute, both Bill and Ben flip all their coins at the same time, if they have any. If a coin lands heads, then the other person gets that coin. If a coin lands tails, then that coin stays with the same person. What is the probability that after exactly 3 minutes, they each end up with 2 coins?

[4 ♟] Problem 7 (2019 AMC 10A Problem 20). The numbers $1, 2, \ldots, 9$ are randomly placed into the 9 squares of a 3×3 grid. Each square gets one number, and each of the numbers is used once. What is the probability that the sum of the numbers in each row and each column is odd?

[6 ♟] Problem 8 (2021 AMC 10B Problem 18). A fair 6-sided die is repeatedly rolled until an odd number appears. What is the probability that every even number appears at least once before the first occurrence of an odd number?

[7 ♟] Problem 9 (2021 OMMC Round 2). An infinitely large grid is filled such that each grid square contains exactly one of the digits $\{1, 2, 3, 4\}$, each digit appears at least once, and the digit in each grid square equals the digit located 5 squares above it as well as the digit located 5 squares to the right. A group of 4 horizontally adjacent digits or 4 vertically adjacent digits is chosen randomly, and depending on its orientation is read left to right or top to bottom to form an 4-digit integer. The expected value of this integer is also a 4-digit integer N. Given this, find the last three digits of the sum of all possible values of N.

[9 ♟] Problem 10 (2013 AMC 12A Problem 24). Three distinct segments are chosen at random among the segments whose end-points are the vertices of a regular 12-gon. What is the probability that the lengths of these three segments are the three side lengths of a triangle with positive area?

[9 ♟] Problem 11 (2019 MBMT). There are 4 traffic lights placed uniformly at random on a 4-mile road. The traffic lights, all in sync, follow a 1-minute loop where they are red for 1 minute and then momentarily flash green so that only cars already stopped at the light can go on. A car arrives at one end of the road just as the lights are flashing green. If the car, when not stopped, always travels at 1 mile per minute, what is the expected number of minutes it takes for the car to traverse the road?