1 Probability

1.1 Introduction

The next topic we want to take up is probability. Probability is a mathematical area that studies randomness. The question about what exactly "random" means is something for philosophers to ponder. Basically we call a process a random experiment if its exact outcome is unpredictable, haphazard, and without pattern. We usually can just recognize randomness when we see it. Rolling a pair of dice is a random experiment, as is tossing a coin and drawing a card from a well shuffled deck.

In mathematics something is said to be random when it can reasonably be assumed that individual results are unpredictable. Probability Theory, however, provides a mathematical way to make predictions about the results anyway. Probability can make rather strong predictive statements about repeated random events. The results of an individual event remain a surprise, but after a large enough number of repetitions, the overall results can form very strong patterns. Probability theory is the mathematical framework for describing those patterns. Probability cannot predict exact results, but can make very strong statements about general results. Probability theory is the mathematical background of the normal intuition we all have about random things.

Consider the following example. Suppose you go for coffee each morning with a co-worker. Rather than argue over it, you strike a deal to determine who should pay. The co-worker flips a quarter; if it comes up heads, he pays; if it comes up tails, you pay. All this seems fair until you realize that, after 10 days of this, you have paid every time. On day eleven, you ask to examine the coin. Your intuition about how a fair coin toss works causes you to question the fairness of your co-worker's coin toss. It is not that it is impossible for a coin to come up tails 10 consecutive times, but it seems highly unlikely that it would happen the first ten times you use it in this game.

The result of one coin flip should be totally unpredictable. This means you can assume that the coin flip will produce random results. But randomness is predictable over the long run. In making the deal, you and the co-worker expect that, over the long term, each should pay a roughly equal amount. In 10 tries, it is too hopeful to expect exactly 5 heads and 5 tails. A split of 4-6 or 6-4 should not be that suspicious. Even 3-7 or 7-3 might be chalked up to an odd string of flips. But 0-10 is enough to get even the most trusting person wondering if more than luck is involved. A coin flip is a random event; 10 coin flips are just as random. However, taken collectively 10 coin flips are definitely more predictable.

1.2 Terminology

Before we get started we set out our vocabulary. Probability theory can get complicated, but using the right words can help avoid a lot of confusion.

- Experiment: an experiment is any process that produces unpredictable outcomes.
- Sample Space: a list of every possible individual outcome of an experiment.
- Sample Points: these are individual outcomes in the sample space.
- Event: an event is any outcome or collection of outcomes of an experiment.

Now the same physical experiment can have several different sample spaces, and in fact, one important technique in probability is changing the sample space. A sample space may be selected for several reasons. If we are interested in collecting data on the occurrence of specific results, we may choose a sample space that makes collecting that data easy. If we intend to do additional theoretical probability or statistics work with our experiment, we may choose a sample space just for its theoretical properties. If we have an experiment with a natural sample space that happens to be large or cumbersome, we may choose to combine parts to form a more convenient sample space.

For an example consider the experiment produced by rolling two dice on a Monopoly board. What is a good sample space? One possibility is to list all the possible totals on the two dice. This would give

$$\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$

But we all know that occasionally one die gets hung up on the Chance Deck. Should that be listed as a possible outcome. If we are actually interested in this case, then "yes" it should be in the sample space. However, in the normal play of the game, a "cocked die" leads to a "do-over." Basically that possible outcome is ignored, and it does not count. If we follow this approach, the sample space above is sufficient.

Suppose we find ourselves in Monopoly jail. To get out, we need to roll doubles; that is we need to roll two 1's, two 2's,two 3's,two 4's,two 5's or two 6's. Anything else keeps us in jail. A reasonable sample space for this version of the experiment is

$$\{\textit{Doubles}, \textit{ Not Doubles}\}.$$

Once we realize that it matters in the game if we roll doubles or not, we may need a sample space that keeps track of totals and doubles. Now the collection $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, \textit{Doubles, Not Doubles}\}$ is NOT a sample space. The outcomes in the list are not distinct possible results. If both dice come up 6's, then this outcome appears in the list as both a total of 12 and as doubles. Each possible outcome of the experiment should have a unique identifier in the sample space. To keep track of both the total and doubles, we need to expand our sample space and to introduce events into the mix.

A better sample space for the rolled dice might be

$$\begin{aligned}\{&\{1,1\},\{1,2\},\{1,3\},\{1,4\},\{1,5\},\{1,6\},\{2,2\},\\ &\{2,3\},\{2,4\},\{2,5\},\{2,6\},\{3,3\},\{3,4\},\{3,5\},\\ &\{3,6\},\{4,4\},\{4,5\},\{5,5\},\{5,6\},\{6,6\}\}\end{aligned}$$

We use brackets here to signify that the order of the numbers inside does not matter. We are still interested in the dice totals and whether the roll produced doubles or not. Now we are interested in events in this sample more than the sample points. The events we are interested in are

$$
\begin{aligned}
\textit{Total 2} &= \{\{1,1\}\} \\
\textit{Total 3} &= \{\{1,2\}\} \\
\textit{Total 4} &= \{\{1,3\},\{2,2\}\} \\
\textit{Total 5} &= \{\{1,4\},\{2,3\}\} \\
\textit{Total 6} &= \{\{1,5\},\{2,4\},\{3,3\}\} \\
\textit{Total 7} &= \{\{1,6\},\{2,5\},\{3,4\}\} \\
\textit{Total 8} &= \{\{2,6\},\{3,5\},\{4,4\}\} \\
\textit{Total 9} &= \{\{3,6\},\{4,5\}\} \\
\textit{Total 10} &= \{\{4,6\},\{5,5\}\} \\
\textit{Total 11} &= \{\{5,6\}\} \\
\textit{Total 10} &= \{\{6,6\}\} \\
\textit{Doubles} &= \{\{1,1\},\{2,2\},\{3,3\},\{4,4\},\{5,5\},\{6,6\}\} \\
\textit{Not Doubles} &= \{\{1,2\},\{1,3\},\{1,4\},\{1,5\},\{1,6\},\{2,3\} \\
&\quad \{2,4\},\{2,5\},\{2,6\},\{3,4\},\{3,5\},\{3,6\}, \\
&\quad \{4,5\},\{4,6\},\{5,6\}\}
\end{aligned}
$$

There is another good choice for a sample space in this experiment. Its advantages are theoretical, and we will see later what those advantages are. In this sample space possibility, we will keep track of which dice comes up with which number. In practice, we might need to use two different color dice to do this, but theoretically all we need do is assume we can tell the dice apart. We list all possible outcomes keeping track of each die. To signify that the order of the numbers matters, we use parentheses to group the pairs.

$$
\begin{aligned}
\{&(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\
&(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\
&(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\
&(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\
&(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\
&(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}
\end{aligned}
$$

Notice that the sample points in all the previous sample spaces are identified as events in this sample space:

$$\begin{aligned}
\{3,5\} &\equiv \{(3,5),(5,3)\} \\
\{2,2\} &\equiv \{(2,2)\} \\
\textit{Total 4} &= \{(1,3),(2,2),(3,1)\} \\
\textit{Total 5} &= \{(1,4),(2,3),(3,2),(4,1)\} \\
\textit{Doubles} &= \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}
\end{aligned}$$

We have the following terms

- Experiment: an experiment is any process that produces unpredictable outcomes.
- Sample Space: a list of every possible individual outcome of an experiment.
- Sample Points: these are individual outcomes in the sample space.
- Event: an event is any outcome or collection of outcomes of an experiment.
- Mutually Exclusive Events: two events that cannot occur at the same time.
- The Impossible Event or Null Event: the event that nothing happens. We write the null event as $\emptyset$.
- The Intersection Event: The event that two other events occur at the same time. We write the intersection event as $E_1 \cap E_2$.
- The Union Event: The event that either one or both of two other events occur. We write the union event as $E_1 \cup E_2$.
- The Complementary Event: The event that another event does not occur. We write the complementary event as E^c.

In our dice example, the events *Total 5* and *Doubles* are mutually exclusive. They cannot both occur on the same roll of the dice. If you look at these events in the largest possible sample space:

$$\begin{aligned}
\textit{Total 5} &= \{(1,4),(2,3),(3,2),(4,1)\} \\
\textit{Doubles} &= \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}.
\end{aligned}$$

These share no common sample points, and that is exactly what makes them mutually exclusive. This also means that their intersection event is impossible. Thus

$$(\text{Total 5}) \cap (\text{Doubles}) = \emptyset.$$

We see that saying that two events E_1 and E_2 are mutually exclusive is exactly the same as saying that $E_1 \cap E_2 = \emptyset$.

The union event of *Total 5* and *Doubles* is more interesting. Suppose you want to roll either doubles or a total of 5. That means that you are hoping for the event (Total 5)∪(Doubles). Again using the largest of our sample spaces:

$$\begin{aligned} \textit{Total 5} &= \{(1,4),(2,3),(3,2),(4,1)\} \\ \textit{Doubles} &= \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\} \\ \textit{(Doubles)} \cup (\textit{Total 5}) &= \{(1,4),(2,3),(3,2),(4,1), \\ &\quad (1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}. \end{aligned}$$

Not all events are mutually exclusive. It is possible to roll a total of 6 and doubles at the same time. These events are not mutually exclusive. The intersection event is not impossible. In fact we have

$$\textit{(Total 6)} \cap \textit{(Doubles)} = \{(3,3)\}.$$

Finally, the events *Doubles* and *Not Doubles* are examples of complementary events. By their very wording, *Not Doubles* is the event that *Doubles* does not occur. Thus

$$\begin{aligned} \textit{(Doubles)}^c &= \textit{Not Doubles} \\ \textit{(Not Doubles)}^c &= \textit{Doubles.} \end{aligned}$$

Finally we notice a few things about this terminology. First the notions of "Mutually Exclusive", "Intersection", "Union", "Impossible", and "Complementary" do not depend on the sample space we use. When we have two events, the actual list of the sample points that make them up may differ, but the events are the events. The intersection of two events is just the event that both occur at the same time. The list of sample points that make up that event depends on the sample space. For example, the event of rolling an even total is an event in the experiment of rolling two dice. If we use the sample space,

$$\{2,3,4,5,6,7,8,9,10,11,12\},$$

The event is

$$\textit{Even} = \{2,4,6,8,10,12\}.$$

If we use the largest sample space,

$$\begin{aligned} \textit{Even} &= \{(1,1),(1,3),(1,5),(2,2),(2,4),(2,6), \\ &\quad (3,1),(3,3),(3,5),(4,2),(4,4),(4,6), \\ &\quad (5,2),(5,4),(5,6),(6,2),(6,4),(6,6)\}. \end{aligned}$$

Now the event of rolling an even total cannot be expressed in the sample space

$$\{\textit{Doubles}, \textit{Not Doubles}\}.$$

But all this means is that it is not an appropriate sample space to use if we care about this event.

Next, by the requirements of a sample space, all the sample points must be mutually exclusive as events. Sample points cannot occur at the same time. That is the way that sample points and events differ. Sample points in the same sample space must be mutually exclusive, but events need not be. Thus if you have a random process and you are interested in various possible outcomes, you must ask yourself if any of those outcomes can occur at the same time. If the answer is "no," then you may consider the outcomes are the start of a sample space. If, however, it is possible for two of the outcomes to occur at the same time, you must find a larger collection of sample points to form sample space. The collection must be large enough to allow the outcomes you are interested in to be described as events.

1.3 Probability Models

Almost everyone has an intuitive understanding of the mathematical idea of probability. Typically the probability of an event is the proportion of times that event will occur if the experiment is repeated a large number of times. Thus probability is a positive number less than 1. We say that the probability of the toss of a fair coin coming up heads is $\frac{1}{2}$ because we expect that half the times we toss a coin, it will come up heads. Indeed we say that the probability of a heads is $\frac{1}{2}$ because that somehow expresses our experience. Probability is the mathematical theory that tries to put this numerical expression of our experience to work. Among other things, it tries to give us mathematical ways to assign probabilities to random events that will correspond to our experiences. Unfortunately, along the way, the theory requires that we, at least temporarily set that experience aside. We need to allow the possibility that a probability assignment will not be a good expression of our experience, but accept it as a possibility until it is either it is changed or our intuition is changed. This leads to a very general definition.

A probability model for a random experiment is an assignment of real numbers to events. To simplify things, for now, we will only consider probability models for experiments with finite sample spaces.

Definition 1 *A probability model for a random experiment with sample space* $\{s_1, s_2, s_3, \ldots s_{n-1}, s_n\}$ *is a assignment of real numbers* r_1 *to each sample point. That is*

$$\begin{aligned} p(s_1) &= r_1 \\ p(s_2) &= r_2 \\ p(s_3) &= r_3 \\ \cdots &= \cdots \\ p(s_{n-1}) &= r_{n-1} \\ p(s_n) &= r_n. \end{aligned}$$

These numbers must satisfy

$$0 \leq r_i \leq 1 \quad \textit{for all } i = 1, 2, 3 \ldots (n-1), n$$

and

$$r_1 + r_2 + r_3 + \ldots + r_{n-1} + r_n = 1.$$

We call $p(s_i) = r_i$, *"the probability of* s_i*" but more specifically, "the probability of* s_i *in the model."*

Notice any such assignment is a probability model. In some sense, every probability model is theoretical, and we should keep this in mind. Suppose we consider the toss of a coin. The sample space is

$$\{Heads,\ Tails\}$$

A probability model for this is

$$\begin{aligned} p(Heads) &= 1 \\ p(Tails) &= 0. \end{aligned}$$

This is a probability model because $0 \leq p(Heads) = 1 \leq 1, 0 \leq p(Tails) = 0 \leq 1$ and $p(Heads) + p(Tails) = 1$. It is a probability model all right, but a model for a two-headed coin.

Assigning a probability model to an experiment always involves making an assumption about the experiment. The more common assignment of a probability model to a coin toss is

$$\begin{aligned} p(Heads) &= \frac{1}{2} \\ p(Tails) &= \frac{1}{2}. \end{aligned}$$

This makes assumptions about the coin, the toss, and the recording of the result. For this probability model, we assume the coin, the toss, and the recording are fair. For now, however, we do not want to make any assumptions about any of our random experiments. Any probability model for a coin toss is OK for us to deal with

$$\begin{aligned} p(Heads) &= \frac{1}{3} \\ p(Tails) &= \frac{2}{3}, \end{aligned}$$

or

$$\begin{aligned} p(Heads) &= \frac{4}{5} \\ p(Tails) &= \frac{1}{5} \end{aligned}$$

work as probability models. Either one signals that coins being tossed or the method being used is very strange. But far be it from us to assume that all coins and all tosses are the same.

There is a very practical example of this. Suppose we take a US quarter, and instead of toss it, we spin it on a flat surface and allow it to come to a stop. At first, it seems perfectly reasonable to assign the probability model

$$\begin{aligned} p(Heads) &= \frac{1}{2} \\ p(Tails) &= \frac{1}{2}. \end{aligned}$$

to this experiment and expect it to hold up to our observation. For the most part, it will. But what exactly is the experiment we are conducting? If we are thinking of taking any old coin and spinning it on any old surface one time, this is probably quite reasonable. But if we take one specific coin, and spin it on the same surface over and over, we might want to question that model. First, in tossing a coin, we generally assume that the coin is tossed to caused it to flip in the air. We assume that this flipping is stopped by the random interference of the surface of a hand, the ground or some other object. In the spinning coin experiment, the physics of the spin plays an obvious role in the final outcome. The question is, how much? First not all quarters are the same, there are 50 state quarters, all of which have different backs. Could the different back effect the balance of the coin enough to skew the spinning in favor of one side? Does that matter? Could the state of the coin influence the outcome. Also a well circulated coin may have dents or bevels on the edges that occasionally change the spin of the coin. Could that cause it to fall in one direction more than the other? And what about the surface, is it really flat, or is there a low spot that attracts the coin and influences the spin? Finally there is the spin itself. If you actually try to run this experiment 100 times, you will find that there are inevitably "good" spins and "bad" spins. If the final result is influenced by the quality of the spin, how would you know? The point is that the reasonable assumption that

$$\begin{aligned} p(Heads) &= \frac{1}{2} \\ p(Tails) &= \frac{1}{2}. \end{aligned}$$

is a "good" probability model is not always so completely obvious.

The moral of the story is that, in the beginning, we need to accept any probability model for even the most familiar experiment as possibly valid. We hope that we will one day be able to use probability theory to test the quality of a particular model as applied to one particular experiment. But for now, all probability models are nothing more than theoretical constructs with the right arithmetic properties.

Now we have a few more probability terms to define:

- The Probability of an Event (in a probability model): The probability of an event is the sum of the probabilities of the sample points that make up the event. That is to say,

$$p(\{t_1, t_2, t_3 \dots t_k\}) = p(t_1) + p(t_2) + p(t_3) \dots + p(t_k).$$

- Equiprobable Events: events in the same experiment are equiprobable if they have the same probability.
- Equiprobable Probability Model: a probability model in which all sample points have the same probability.

Equiprobable events are also called "equally likely." Also, both of these definitions only refer to one particular probability model at a time. That means equiprobable events in one model may not be equiprobable in another.

An equiprobable probability model is easy to describe mathematically. From our definition a probability model on a sample space $\{s_1, s_2, s_3, \ldots s_{n-1}, s_n\}$ is a assignment of real numbers r_1 to each sample point that satisfies

$$0 \leq r_i \leq 1 \text{ for all } i = 1, 2, 3 \ldots (n-1), n$$

and

$$r_1 + r_2 + r_3 + \ldots + r_{n-1} + r_n = 1.$$

If the model is equiprobable, then $r_1 = r_2 = r_3 = \ldots = r_{n-1} = r_n = r$ So

$$1 = r_1 + r_2 + r_3 + \ldots + r_{n-1} + r_n = nr.$$

So we have

$$p(s_i) = \frac{1}{n} \text{ for all } i = 1, 2, 3 \ldots n.$$

For example, consider our experiment of rolling two dice. We have proposed four different sample spaces:

$$\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

$$\{Doubles,\ Not\ Doubles\}$$

$$\{\{1,1\},\{1,2\},\{1,3\},\{1,4\},\{1,5\},\{1,6\},\{2,2\},\\ \{2,3\},\{2,4\},\{2,5\},\{2,6\},\{3,3\},\{3,4\},\{3,5\},\\ \{3,6\},\{4,4\},\{4,5\},\{5,5\},\{5,6\},\{6,6\}\}$$

and

$$\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\ (2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\ (3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\ (4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\ (5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\ (6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$$

We can take any one of these and construct a equiprobable probability model. First $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ gives

$$p(2) = p(3) = \ldots p(12) = \frac{1}{11}.$$

The second gives

$$p(\mathit{Doubles}) = p(\mathit{Not\ Doubles}) = \frac{1}{2}.$$

The third gives each unordered pair a probability of $\frac{1}{20}$. From this we can say that, in this model, the probability of the events *Total 4* and *Total 5* are given by

$$\begin{aligned} p(\mathit{Total\ 4}) &= p(\{1,3\},\{2,2\}) = \frac{1}{20} + \frac{1}{20} = \frac{1}{10} \\ p(\mathit{Total\ 5}) &= p(\{1,4\},\{2,3\}) = \frac{1}{20} + \frac{1}{20} = \frac{1}{10} \end{aligned}$$

This makes these two equiprobable events in this model.

Finally the last samples produces a equiprobable model where each order pair has a probability given by

$$p((a,b)) = \frac{1}{36}.$$

From this we can say that, in this model, the probability of the events *Total 4* and *Total 5* are given by

$$\begin{aligned} p(\mathit{Total\ 4}) &= p(\{(1,3),(2,2),(3,1)\} = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{12} \\ p(\mathit{Total\ 5}) &= p(\{(1,4),(2,3),(3,2),(4,1)) = \frac{4}{36} = \frac{1}{9}. \end{aligned}$$

So in the second model, the events *Total 4* and *Total 5* are nor equiprobable.

It takes very little experience tossing pairs of fair dice to realize that the dice come up doubles far less often than they don't. There is no way one would say that these two events are equally likely. It is unlikely that one would accept the equiprobable probability model obtained from the sample space {*Doubles*, *Not Doubles*} as corresponding to experience. All we are saying, however, that there is very real reasons to doubt the assumption that coming up doubles and coming up not doubles are equiprobable events. The model remains mathematically valid, just not applicable to the rolling of two fair dice.

With more experience with rolling fair dice, one realizes that rolling a total of 5 and rolling a total of 4 are not equiprobable either. It appears that a 5 comes up more often than a 4 in the long run. Any probability model that makes them equiprobable will not correspond to the experience of rolling fair dice. This calls into question any assumptions made in producing that probability model. The equiprobable probability model based on the sample space $\{2,3,4,5,6,7,8,9,10,11,12\}$ makes all the totals equiprobable. So this model should not be applied to fair dice. The equiprobable probability model based on unordered pairs also makes rolling a total of 5 and rolling a total of 4 equiprobable also. Once we decide that we see that a 5 comes up more often than a 4, we must question the applicability of this model to the rolling of fair dice.

Finally we have the equiprobable probability model based on the ordered pairs. Under this model rolling a total of 5 is assigned a higher probability than the probability assigned to rolling a 4. There is no obvious reason, so far, to say that it is applicable to the rolling of fair dice. We have no reason to question the assumption that the sample points in this sample space are equally likely. We know exactly how many sample points there are, and we therefore know the complete probability model these assumptions produce. Does this make it the "correct" probability model for rolling a pair of fair dice? Well, maybe. In fact, in gambling circles, this probability model is basically the definition of the phrase "fair dice." It is the "correct" model for rolling fair dice because if it does not correspond to the tossing of a pair of dice, those dice are automatically "unfair."

1.4 Final Observations

There are a few conclusions we can draw from the discussion of probability models above. Some are obvious. For example,

Conclusion 2 *If E is an event in a sample space with a probability model, then E and E^c are mutually exclusive.*

Conclusion 3 *If E is an event in a sample space with a probability model, then*

$$p(E^c) = 1 - p(E).$$

Recall that E^c is the event that E does not occur. Since an event cannot both occur and not occur at the same time, any event is mutually exclusive with its complement. Further every sample point that is in the sample space is either part of the event E or part of the event E^c. Thus if $E = \{s_1, s_2, \ldots s_k\}$ is an event in the sample space $\{s_1, s_2, \ldots s_k, s_{k+1} \ldots s_{n-1}, s_n\}$, then $E^c = \{s_{k+1} \ldots s_{n-1}, s_n\}$ This means

$$\begin{aligned} p(E) + p(E^c) &= (p(s_1) + p(s_2) + \ldots p(s_k)) + (p(s_{k+1}) \ldots p(s_{n-1}) + p(s_n)) \\ &= p(s_1) + p(s_2) + \ldots p(s_k) + p(s_{k+1}) \ldots p(s_{n-1}) + p(s_n) \\ &= 1. \end{aligned}$$

So it must be that $p(E^c) = 1 - p(E)$.

Notice that the impossible event has its own probability:

$$p(\emptyset) = 0.$$

The null event has probability zero, and the null event is also called the impossible event. An event may have probability 0 in a model, but not be the null event, and therefore, it may be possible. Consider the following examples.

First let us look at a fair coin toss. One sample space we might use for this is

$$\{\textit{Heads, Tails, Edge}\}$$

All of these are possibilities when we toss a fair coin if the coin is allow to land on surface and come to a rest. It would be quite a remarkable occurrence if a tossed coin were to land on a concrete sidewalk and end up standing on its edge. It is at least possible for this to happen though. Besides, it could land in a crack between concrete slabs, and thus be on its edge. That should count as an edge. It could be a dime flipped onto a grass lawn. A dime is certainly small enough to be held on edge by healthy grass stalks. Once we include the possibility of a coin coming up "edge" in our sample space, it deserves its own probability in every probability model. If this model has any chance of corresponding to our experience in flipping coin, especially onto concrete, that probability should be very small. We might try a probability model like

$$\begin{aligned} p(\mathit{Heads}) &= 0.499 \\ p(\mathit{Tails}) &= 0.499 \\ p(\mathit{Edge}) &= 0.002 \end{aligned}$$

The probability of a coin landing on its edge is set to a very small $0.002 = \frac{1}{500}$. But on concrete, even that seems a bit high. Maybe

$$\begin{aligned} p(\mathit{Heads}) &= 0.49999 \\ p(\mathit{Tails}) &= 0.49999 \\ p(\mathit{Edge}) &= 0.00002 \end{aligned}$$

is better? But what is the point of all this accuracy, even if it were pretty close to what we would see with experience, who would ever want to gain that much experience flipping coins. They would need to flip a coin 100,000 times just to get a hint of how often it happens. No, probability is just estimation anyway, so why not round off to

$$\begin{aligned} p(\mathit{Heads}) &= 0.5 \\ p(\mathit{Tails}) &= 0.5 \\ p(\mathit{Edge}) &= 0 \end{aligned}$$

Thus we have a zero probability to an event that we will, under duress, admit is actually possible. An event with a zero probability does not mean it is impossible, but it does mean it is highly unlikely.

For another example recall the familiar game Rock-Paper-Scissors. Two people play each other by putting out one hand in a fist (rock), or extended (paper) or with two fingers extended (scissors). The winner is determined by the rule:

- Rock breaks scissors;
- Scissors cut paper
- Paper covers rock.

Assuming the two players are trying to vary their plays, the game can be considered a random process. A sample space for this experiment is

$$\begin{aligned}&\{(\text{Rock, Rock}), (\text{Rock, Paper}), (\text{Rock, Scissors}),\\&(\text{Paper, Rock}), (\text{Paper, Paper}), (\text{Paper, Scissors}),\\&(\text{Scissors, Rock}), (\text{Scissors, Paper}), (\text{Scissors, Scissors})\}.\end{aligned}$$

We use the parentheses to denote the ordered pairs, since we need to know which player threw what to determine a winner. If we build an equiprobable probability model on this sample space, each sample point will be assigned a probability of $\frac{1}{9}$. There are three possible events to worry about:

$$\begin{aligned}\textit{First player wins} &= \{(\text{Rock, Scissors}), (\text{Scissors, Paper}), (\text{Paper, Rock})\}\\\textit{Second player wins} &= \{(\text{Rock, Paper}), (\text{Scissors, Rock}), (\text{Paper, Scissors})\}\\\textit{Tie} &= \{(\text{Rock, Rock}), (\text{Scissors, Scissors}), (\text{Paper, Paper})\}\end{aligned}$$

Then the probabilities under the model are

$$\begin{aligned}p(\textit{First player wins}) &= \frac{3}{9} = \frac{1}{3}\\p(\textit{Second player wins}) &= \frac{3}{9} = \frac{1}{3}\\p(\textit{Tie}) &= \frac{3}{9} = \frac{1}{3}\end{aligned}$$

However, there is another approach. Under the normal rules of this game, a tie does not count and leads to a "do over." If a tie does not count, then why include it in the sample space? We might try the sample space

$$\begin{aligned}&\{(\text{Rock, Paper}), (\text{Rock, Scissors}), (\text{Paper, Rock}),\\&(\text{Paper, Scissors}), (\text{Scissors, Rock}), (\text{Scissors, Paper})\}.\end{aligned}$$

If we build an equiprobable probability model on this sample space, each sample point will be assigned a probability of $\frac{1}{6}$. The events we care about in this sample space are

$$\begin{aligned}\textit{First player wins} &= \{(\text{Rock, Scissors}), (\text{Scissors, Paper}), (\text{Paper, Rock})\}\\\textit{Second player wins} &= \{(\text{Rock, Paper}), (\text{Scissors, Rock}), (\text{Paper, Scissors})\}\\\textit{Tie} &= \emptyset.\end{aligned}$$

So

$$\begin{aligned}p(\textit{First player wins}) &= \frac{3}{6} = \frac{1}{2}\\p(\textit{Second player wins}) &= \frac{3}{6} = \frac{1}{2}\\p(\textit{Tie}) &= 0\end{aligned}$$

We have assigned *Tie* a probability of zero, even though it is a rather common possibility. It is assigned a probability of zero because in the sample space it is a result that is ignored. It can be assigned the probability of zero, because it is an "impossible" outcome under the rules of the game.

There are better ways to deal with this particular example, but it does serve its purpose as a illustration of the difference between "impossible" and "probability zero."

Prepared by: Daniel Madden and Alyssa Keri: May 2009