An Introduction to Probability Theory

Ishaan Patkar

Probability theory is the mathematical study of uncertainties. With the use of mathematics, the study of probability becomes a rich theory with many unintuitive results.

This paper concerns itself with the concepts and ideas mathematicians use to understand probability. We formulate an abstract mathematical model of a real-world situation involving randomness, called a *probability space*. This model is the starting point for probability theory, since it gives us a method in which to apply mathematical tools to real-world situations. In order to do that, however, one must be able to translate such situations into probability spaces. We give several examples how one might do that, and hope this gives the reader insight into why the concept is useful.

Most of the material in this paper comes from [GS03] and [Fel68] for probability and [Tao11] for the basic underlying measure theory.

1 The Sample Space

In the real world, situations often arise whose results are uncertain, and cannot be known in advance. Here are some example:

a. A 6-sided die is rolled. What number is rolled?

b. A marble is picked from a jar containing red and blue marbles. What is the color of the marble?

c. Alice thinks of a number between 1 and 100. What is that number?

d. A dart is thrown at a dartboard. Where will it land?

e. A bus is supposed to arrive at a bus stop at 5:00 pm. However, buses are often late. How many minutes late will the bus be?

We refer to these situations like these as *experiments*. Every experiment results in a certain *outcome*, which we do not know in advance. For instance, in experiment (b), the outcome is the color of the marble. In experiment (d), the outcome is the position where the dart lands. In experiment (c), the outcome is the number Alice picks. In each case, we (the observers) cannot know what the outcome is until after the experiment happens.

However, there are some outcomes we can explicitly rule out. For instance, in experiment (b), we can rule out the possibility of picking out a green marble. In (c), we can rule out the possibility that Alice picks the number 101. These may not be strictly impossible in the real-world; however, for the sake of mathematical abstraction, we will always assume experiments behave exactly as defined.

So what exactly are the possible outcomes? The answer to this question is important enough that we give it a special name. We define the set of possible outcomes to be the *sample space*, which we will often denote Ω. Elements of Ω are outcomes of the experiment. For instance, when rolling a 6-sided die, the possible outcomes are the positive integers from 1 through 6, so:

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

The sample space is also helpful in illustrating when real-world problems are ill-defined. For instance, when throwing darts at the dartboard, each point on the dartboard is a possible outcome, and so is an element of the sample space Ω. However, are these the only possible outcomes? There is nothing in the statement of the experiment to tell us one way or the other. In an ill-defined situation like this, we may make some reasonable assumptions about the sample space in order to define it. For instance, we can assume here that the players are good enough to consistently hit the dartboard. So, Ω is exactly the set of points on the dartboard.

However, sometimes we want to consider a smaller collection of outcomes than the sample space. Take experiment (e), for instance. In this experiment, we might desire that the bus arrives within 15 minutes. This describes the set of outcomes $[0, 15]$ (the closed interval of the real line). We call such sets of outcomes *events*. As in this example, events are often given by short descriptions. For instance, in experiment (d), one event is hitting the bullseye. The event this corresponds to is set of points on the dartboard which make up the bullseye.

In particular, two events stand out: the empty set $\emptyset$ which contains no outcomes, and the sample space Ω which contains every outcome. These are, however, the two extremes: events are more often somewhere in between $\emptyset$ and Ω.

With this preliminary terminology out of the way, let us consider a more comprehensive example.

Example 1.1. Consider the experiment of flipping a coin three times. Each outcome is then a sequence of heads and tails: for instance, HHH and HTT are outcomes. The sample space Ω is then the set of all such outcomes:

$$\Omega = \left\{ \begin{array}{cccc} \text{HHH} & \text{HHT} & \text{HTH} & \text{HTT} \\ \text{THH} & \text{THT} & \text{TTH} & \text{TTT} \end{array} \right\}$$

Now, let's consider some examples of events. Consider the event E where the first flip is tails. This is the set

$$E = \{\text{THH}, \text{THT}, \text{TTH}, \text{TTT}\}$$

Another event F we can consider is the event of flipping the sequence TTT. This event has one element: $F = \{\text{TTT}\}$. In this way, every outcome is also represented among the events, and events can be thought of as a way to generalize outcomes.

Thus, talking about events is more general than talking about outcomes alone. So, we introduce the *set of events* to contain all of our events. We typically denote this $\mathcal{F}$.

2 The Probability Measure

Sample spaces, outcomes, and events are helpful in understanding possibilities in an experiment. Now we introduce the notion of probability. We need a way to assign probabilities to outcomes—or more generally, to events. We define a function which does just that. The *probability measure* or *probability function* is a function $P : \mathcal{F} \to \mathbb{R}$ which assigns every event $E \in \mathcal{F}$ to a real number $P(E)$ which is called the *probability*. This number $P(E)$ represents the relative likelihood that the unknown outcome of the experiment will be an element of E.

Now, there are some conventions when assigning and interpreting probability. We give an event a probability of 0 if it has no chance of happening. On the other hand, we assign a probability of 1 to events which we are completely certain will happen. Note this means $P(\emptyset) = 0$ and $P(\Omega) = 1$. This is since some outcome ω will occur. As $\emptyset$ contains no elements, $\omega \notin \emptyset$, so regardless of the outcome, the event $\emptyset$ will never occur and so we give it probability 0. On the other hand, if ω occurs then it is a possible outcome and so always in Ω. Since the event Ω will always occur, we will assign Ω a probability of 1. Do note that these are not always the only sets which have probabilities 0 and 1.

There is one other important property of probability that we want. Say E and F are events. What is the probability of $E \cup F$? This is the probability of an outcome occurring that is either in E or F. If we assume that E and F are mutually exclusive (so they have no elements in common), we want the individual probabilities of E and F to add together to get the probability of $E \cup F$. That is, whenever E and F are mutually exclusive, we want the following equation to hold:

$$P(E \cup F) = P(E) + P(F) \tag{2.1}$$

This is the property of *finite additivity*, and it is the third requirement we impose on the probability measure.

Example 2.1. We will define the probability space $(\Omega, \mathcal{F}, P)$ of rolling a six-sided die. The only possible outcomes from rolling the die is $\Omega = \{1, 2, 3, 4, 5, 6\}$. The set of all events, $\mathcal{F}$, is then the set of all subsets of Ω.

Now we consider the probability measure. Let us start by finding the probabilities of the most basic type of events: the events with 1 outcome. We will assume that the die is fair, meaning $P(\{\omega\})$ is the same for any $\omega \in \Omega$. Suppose $P(\{\omega\}) = p$. What should p be? Well, assuming our probability measure satisfies finite additivity, we have the following:

$$1 = P(\Omega) = P(\{1\}) + P(\{2\}) + P(\{3\}) + P(\{4\}) + P(\{5\}) + P(\{6\}) = 6p.$$

Therefore, $p = \frac{1}{6}$.

Now, to generalize this to events, consider an event $E \in \mathcal{F}$. The probability of it occurring is the probability of one of its outcomes occurring: this is $\frac{1}{6}$ for each element in E. If we use the convention $|E|$ to refer to the number of elements in E, we then define the probability measure to be:

$$P(E) = \frac{|E|}{6}.$$

Finally, we check that the probability measure satisfies the requirements we set out. It should satisfy these requirements, since we used the requirements to deduce what the probability measure should be. Note that Ω has 6 elements, meaning $P(\Omega) = 1$. On the other hand, $\emptyset$ has 0 elements, and so $P(\emptyset) = 0$. Now, if E and F are two events which are mutually exclusive, then the number of elements in $E \cup F$ will be $|E| + |F|$. This means:

$$P(E \cup F) = \frac{|E \cup F|}{6} = \frac{|E|}{6} + \frac{|F|}{6} = P(E) + P(F).$$

So, this probability measure satisfies the finite additivity property.

Observe that in this example, we had to first assume that the probability space existed. Then, using real-world information—in this case, that the die was fair—we were able to deduce the specific properties of each of the objects Ω, $\mathcal{F}$, and P. This is typical when translating from a real-world experiment to a mathematical space. The deductions here were fairly simple since we had a simple sample space and made the nice assumption that every outcome was equally likely. In more complex situations, the sample space might be far more complicated and the information more limited. In such situations, however, the same technique is useful: first, assume that a probability space modeling the situation exists. Then, use any information we are given to set up equations which we can use to solve for unknown properties of the space.

3 An Infinite Probability Space

Most of the previous examples dealt with probability spaces in which the sample space is finite. Finiteness is a nice property, but not a necessary one. The following example demonstrates this:

Example 3.1. A small ball is dropped from a height onto a 1×1 tile. It is equally likely for the ball to land at any point on the tile. We will find the probability space for this experiment. The sample space Ω is $[0, 1]^2$, the unit square, which represents this 1×1 tile. The events are then subsets of Ω, which are regions in the unit square. Now, the probability that the ball will land in a given region E of the unit square is the ratio of the area of E to the area of Ω. Since the latter is just 1, we have:

$$P(E) = A(E) = \iint_E dA.$$

By using properties of double integrals, we can check that this satisfies the requirements of a probability space.

When the sample space is infinite, probability theory often requires working with calculus, especially integration. By translating real-world situations into probability spaces, we can use the tools of calculus to answer questions about the experiment.

References

[Fel68] William Feller. *An Introduction to Probability Theory and Its Applicatons*. 3rd. Wiley, 1968.

[GS03] Charles M. Grinstead and J. Laurie Snell. *Introduction to Probability*. 2nd. American Mathematical Society, 2003.

[Tao11] Terence Tao. *An Introduction to Measure Theory*. American Mathematical Society, 2011.