PROBABILITY AND RANDOM VARIABLES

1. Probability Basics

We all have an intuitive understanding of *Probability* and use it in our daily interactions and conversations with people. From common things such as weather forecasts we have a finely honed understanding of events that may or may not occur. We know, for instance, that when the forecaster says that there is a 90% chance of snow tomorrow, we can expect to see snow on the ground, though it is known every once in a while that the forecaster has not been right. Likewise, when they predict sunny weather, it is not certain that it will be sunny but more often than not, the forecaster is accurate. Likewise, we make similar snap assessments about many hundreds of events that occur daily. For instance, if our building has a friendly doorman, we know that he will greet us in the morning and maybe we might return his greeting too. It is not certain that he will do it everyday but 9 times out of 10, he would smile. In other words, if 10 people walked into the building, 9 of them would have gotten a friendly greeting from him and the 10th would not have gotten one. We are used to dealing with these kinds of "uncertain" events and logically reasoning about them.

Let's now define these intuitive understandings formally so that we may do algebra on them and make our understanding precise. Probability is usually used to describe outcomes of events that take multiple values and cannot be stated with certainty that a particular outcome will occur at a particular experiment. For instance, the temperature on any given time on any given day at a given place can take several values. Likewise, if you are rolling a die, you are certain that one of the faces will be up but you cannot precisely state which face will be up at any given throw.

Let's introduce some terms into our vocabulary. A *trial* refers to a single event such as a coin toss or a roll of a die. The *elementary* events of a trail are the possible outcomes of a that trial. For instance, when we pick a card at random from a deck of cards, any one of the 52 cards can get picked. From these elementary events, we can construct more complex events. For instance, we can speak of the event of getting two heads in three coin tosses. Likewise, we can construct the event of getting an even number when we roll a die – these are complex events that are constructed from combinations of elementary events. We can construct arbitrarily complex events from these elementary events. For instance, an event that we can easily construct is that of getting the Ace of Hearts in a draw of card after getting 4 heads in a row in the toss of a coin and then follow this with getting a number greater than 11 in two tosses of a die. We can also visualize events that are infinite in the number of trials. Going further, we can even visualize events that take on values in a continuous spectrum (e.g. the temperature at a given time). All of these are example of probabilistic events and get used when we build models to represent possible outcomes that are of interest to us.

When we speak of rolling a fair die, we assume that that the die is symmetric in the space of possible outcomes. That is the ideal die has 6 possible faces and each of them are equally likely. They are symmetric with respect to the die. It has no preference towards one face or the other and all of the faces are equivalent. Only one of the faces can occur in one roll and when that event occurs, all other events do not occur. Since all the events are equally likely, the sum of these events add up to 1 because one of them has to definitely occur. We can conclude logically that for a fair die, the probability of any one elementary event is 1/6. Similarly, the probability of drawing any card from a deck is 1/52 and that of a getting any particular outcome in a single coin toss is 1/2.

The next concept to is that of *independence*. Knowing one coin toss doesn't tell us anything about subsequent coin tosses. We can say that the first coin toss and the second one are independent. Likewise, the outcome of a coin toss has no bearing on tomorrow's weather – they are independent events. When we have independent events, we can reason that the probability of seeing a particular configuration (say a Head followed by a 6) is the product of their individual occurrences: $1/2 \times 1/6$. We can do this by computing all possible configurations of coins and die faces (there are a total of 12 of them) and they are all symmetric or we can use the notion of independence and compute the probability of a particular outcome from there. Here are some examples of events that are not independent of each other: The temperature at a given time is not independent from the temperature 1 hour ago. Knowing that one hour ago, it was 70 degrees, we know that there is a much greater chance that the temperature is close to 70 degrees than 10 degrees. Likewise, if the first two characters in a word is *th*, we know that *e* has a higher probability of occurring than say *z*. The occurrence of *e* is not independent of the previous characters in a piece of English text, unlike the case of a coin toss. Knowing the previous outcomes alters our expectation of a particular future outcome in the case of events that are not independent. Another type of dependent events occur when we sample seemingly independent events but our sampling changes the sample space underneath. For instance, you can construct a scenario where you pick a card at random from a deck of cards. After you pick the first card, you discard it. This changes the probability of the subsequent set of events since the sample space has been reduced by one. If you picked a black card in your first draw and discarded it, the probability of getting a black card in your second draw goes down – as there are only 51 possible elementary events and not 52.

2. Discrete Random Variables

Let us now introduce the notion of a *random* variable. A random variable is a variable whose value is subject to some randomness or chance. This variable can take a values from a set of possible values, each of which is associated with a probability of the underlying event occurring. When we speak of discrete random variables, we are restricting ourselves to random variables whose underlying events come from a finite space of possibilities or whose values take countably infinite set of discrete values. For instance, the toss of a coin, the roll of a die, the set of integers, etc – also known as the sample space.

Formally, a random variable is a real-valued function defined over the sample space. It maps the sample space into a real-value between 0 and 1. The events in the sample space cannot have a negative probability. If the event cannot occur, its probability is 0. Likewise, if the event is certain then the probability associated with it is 1, and all other cases lie in between these two extremes.

We introduce a discrete random variable below:

$$Y(\omega) = \begin{cases} 1, \omega = H \\ 0, \omega = T \end{cases} \tag{1}$$

The random variable Y takes a value in $(0, 1)$ based on whether we get a Head or a Tail in the coin toss. Associated with the random variable is the probability of a particular value of the variable. For instance, for Y, we have

$$\begin{aligned} Pr(Y = 1) &= 1/2 \\ Pr(Y = 0) &= 1/2 \end{aligned} \tag{2}$$

Where $Pr(Y)$ is the probability of the discrete random variable taking on a particular value.

Let Y be a random variable denoting the outcomes of the roll of a die. We can now state the probability of Y taking a value larger than 3 or getting a die face of $4, 5, 6$ as $Pr(Y > 3)$. Similarly, we can have a random variable Y that represents the sum of the faces of two dies rolled together. Die 1 produces one of 6 values and Die 2 does the same. So, Y can takes values between 2 and 12. We can compute the probability of $Y == 6$, for instance. $Pr(Y == 6)$ is the probability of getting a sum of 6 by rolling two dies together.

When the events are independent, we can speak of probabilities of both of the events occurring and either one of them occurring: These are very similar notions to the Set theory ideas of *Intersection* and *Union*. If Y is a random variable representing the event of getting a Head in a coin toss and Z is a random variable representing getting a 1 in the roll of a die, then $Pr(Y) = 1/2$ and $Pr(Z) = 1/6$. We can compute $Pr(Y, Z) = 1/12$, or the probability of getting a Head and a 1 when rolling a die and flipping a coin. Similarly, if X represents the event of getting a 2 in a roll of die, the probability of getting either a 1 or a 2 is $Pr(X \cup Z) = 1/6 + 1/6 = 1/3$.

For any random variable Y, $0 \leq Pr(Y) \leq 1$. Similarly, if the Union of several random variables covers the sample space or all the possible elementary events, then it has to have a probability of 1. For example, if Y represents the event of getting a black card from the deck at random, we known $Pr(Y) = 26/52 = 1/2$. If Z represents the event of getting a red card from a card deck, then $Pr(Y \cup Z) = 1$, as they both are independent events and cover the entire sample space. We assume that these events are with replacement. Once you pick a card, you replace it back in the deck. Otherwise, Y and Z are not independent.

Sometimes it is easier to compute the probability of the *complement* of an event. A *complementary* event is the one that covers the sample space given a particular event. For instance, if Y is as above, the event of getting a 1 in a roll of die.
$barY$, or complement of Y is the event of not getting a 1. In any roll, either you get

a 1 or you don't and that covers all the possible outcomes. So, $Pr(Y \cup \bar{Y}) = 1$ and $Pr(\bar{Y}) = 1 - P(Y)$.

3. Continuous Random Variables

When our events take continuous values, such as the temperature at a given time, the speed of the wind on a given day – it doesn't make much sense to talk about the probability of the temperature being precisely 70.2 degrees – in fact that probability will be very close to zero as the sample space is infinite. It makes much more sense to talk about the probability of an event being within a range. We can define the probability of the temperature being between 70 and 71. When we cover calculus, we can formally define this notion using calculus. For now, the somewhat imprecise notion of the random variable taking values within a range will suffice.

There is one fundamental difference that I'd like to highlight for now. With discrete random variables, we have a notion of a probability mass function that maps the value of the random variable to a probability. For instance, in the above examples $Pr(Y = 1) = 1/6$. For continuous random variables, we have the notion of a probability density function and a cumulative distribution function. These functions are continuous (or at least piecewise continuous). We can associate a continuous function with the temperature forecast for instance. This continuous number captures our confidence that the temperature will be in a range around a particular value and *PDF* and *CDF*s capture these notions for us. We'll cover *PDF*s in detail in subsequent weeks when we encounter probability distributions.