1   Simulation of Discrete Probabilities

Probability is the study of randomness, of uncertainty.

1.1   Formal Introductory Definitions

More formally, when we think of probability, we think of running a **random experiment** (for example, flipping a coin to see if it's head or tails, or roll a die to see if you get numbers 1 through 6, or even a lottery ticket to see if you win or lose.)

The set of all possible outcomes is called the **sample space**, which we call Ω. Namely, we say that

$$\Omega = \{\omega_1, \ldots, \omega_n\}$$

Here, each ω_i represents a possible outcome. For example, if you are flipping a coin, we would have

$$\Omega_{\text{Coin}} = \{\text{Heads}, \text{Tails}\}$$

In this section, we will assume, for simplicity, that there is only a *finite* number of possible outcomes.

1.2   What is Probability?

The outcomes are assigned **masses** $m(\omega_i) \geq 0$ such that

$$m(\omega_1) + \cdots + m(\omega_n) = 1$$

We think of $m(\omega_i)$ as the **probability** that outcome ω_i occurs. For example, when rolling a "fair" die with numbers 1 through 6, there are 6 possible outcomes, each with an equal chance of occurring, so the mass of each outcome is $\frac{1}{6}$.

1.3   Studying Probability

When studying random experiments, we often use something called a **random variable** (*RV* for short). Formally, a RV X is a function from the sample space Ω to the set of real numbers $\mathbb{R}$; that is,

$$X : \Omega \mapsto \mathbb{R}.$$

(Example.) Suppose we want to flip a fair coin. It's obvious that the sample space Ω is just Heads or Tails. We can define the random variable X like

$$X = \begin{cases} 1 & \text{If Heads} \\ 0 & \text{If Tails} \end{cases}.$$

We could also define the random variable like

$$X = \begin{cases} 100 & \text{If Heads} \\ 703 & \text{If Tails} \end{cases}.$$

The point is that your random variable maps your outcomes (from your sample space) to numbers. In other words, you are quantifying your outcomes.

If the outcome $\omega \in \Omega$ has a **random** occurrence, then the value $X(\omega)$ is also random; this is due to this random experiment. For example, suppose we have a random variable which takes some value depending on the roll of a die. For instance, if our random variable indicates whether we roll an even or odd number, then X will take the value 1 if we roll an even number or 0 if we roll an odd number. So, while X will either only

take the values 0 or 1, which of the two values it takes is random because it depends on the random roll of the die.

The collection of probabilities $\mathbb{P}(X = x)$ is called the **probability distribution** of X. We say that $\mathbb{P}(X = x)$ is the *sum* of the masses $m(\omega)$ over all ω that would make $X(\omega) = x$. So, in the example that we just talked about, the probability distribution would just be $\frac{1}{2}$ and $\frac{1}{2}$.

(Example.) Suppose we're rolling a fair die. Then, we might let $X(i) = i$ if we roll an i. In other words, X is the value of the die roll. It follows that

$$\mathbb{P}(X = i) = m(1) = \frac{1}{6} \text{ for all } 1 \le i \le 6.$$

(Example.) Suppose we're rolling a fair die, and suppose we have the random variable Y that takes the value 1 if $\omega \in \{1\}$ and the value 0 if $\omega \in \{2, 3, 4, 5, 6\}$. Then, $X = 1$ if we roll a 1 and $X = 0$ otherwise. Then,

$$\mathbb{P}(X = 1) = m(1) = \frac{1}{6} \text{ and } \mathbb{P}(X = 0) = m(2) + \cdots + m(6) = \frac{5}{6}.$$

We can use RVs to make more complicated RVs. Similarly, we can often decompose complicated RVs into several simpler ones.

(Example.) Suppose we roll a die 3 times in a row. Let X be the number of 1's that we roll. Then,

$$X = X_1 + X_2 + X_3$$

where each X_i has the same distribution, i.e. $\mathbb{P}(X_i = 1) = \frac{1}{6}$ and $\mathbb{P}(X_i = 0) = \frac{5}{6}$ for each $i \in [1, 3]$.

(Example.) Again, suppose we roll a die 3 times in a row. Then,

$$X = X_1 + X_2 + X_3.$$

Suppose we wanted to find the probability that at most *2* of those rolls are 1's. There are three possibilities:

(a) All 3 rolls are not 1's. Here, the probability that we don't get a 1 is $\frac{5}{6}$. So, the probability that all three rolls are not 1's is given by $(5/6)^3$. There is only $\binom{3}{0} = 1$ way we can possibly obtain zero 1's. To visualize this, let x be some non-one integer. Then,

```
x x x.
```

(b) Exactly 1 of the 3 rolls is a 1. The probability that we get a 1 is $\frac{1}{6}$ and the probability that we don't get a 1 is $\frac{5}{6}$. We note that there are $\binom{3}{1} = 3$ possible ways we can obtain one 1. To visualize this, let x be some non-one integer. Then,

```
1 x x
x 1 x
x x 1.
```

(c) Exactly 2 of the 3 rolls is a 1. We note that there are $\binom{3}{2}$ ways we can possibly get two 1's. To visualize this, let x be some non-one integer. Then,

```
1 1 x
x 1 1
1 x 1.
```

Putting this together, we have:

$$\mathbb{P}(X \leq 2) = \binom{3}{0}\left(\frac{5}{6}\right)^3 + \binom{3}{1}\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2 + \binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)$$

1.4   Simulating Discrete Probability Distributions

Why would we need to simulate anything in probability? In practice, we do not often know a **priori** what the masses $m(\omega_i)$ are. Hence, we often need to *estimate* the probability that an outcome ω_i will occur.

This is related to the *frequentist definition of probability*. Informally, if we run a random experiment many times, then the probability of times that ω_i occurs converges to the probability $m(\omega_i)$ that ω_i occurs in any given experiment.

> (Example.) Suppose that a magician gives you a coin. You do not know if it is fair. Perhaps, you even suspect that it is not real. How would you go about estimating the probability p of "Heads?"
>
> For instance, if you flip the coin 10 times, and "Heads" only occurs once, then you might naturally begin to suspect that the coin is indeed not fair, and that moreover $p < \frac{1}{2}$. If you continue to flip it a thousand times, and only 50 "Heads" come up, then you become more convinced that it is not fair.

The above example is an introduction to the **Law of Large Numbers**. Let $X_i = 1$ if the ith flip is "Heads." Then, by the Law of Large Numbers,

$$\frac{1}{n}\sum_{i=1}^{n} X_i,$$

or the proportion of "Heads" after n flips, will converge to p as $n \mapsto \infty$. This leads to a method of estimating p. Namely, we can flip the coin a large number n of times, and then calculate

$$\frac{1}{n}\sum_{i=1}^{n} X_i,$$

then we will be close to p.

> (Example: Chevalier de Mere's Dice Rolling.) de Mere made some money by betting people that at least one 6 would come up when we rolled 4 dice. The probability of this occurring is
>
> $$1 - (5/6)^4 \approx 51.8\% > 50\%.$$
>
> Here, we're saying that the probability of getting at least one 6 is the same thing as 1 minus the probability of getting no 6's. So, in the long run, he knew that he would win more often than not.
>
> Instead of calculating this probability, de Mere probably rolled 4 dice as many times as possible on his own, and noticed that at least one 6 would come up *slightly* more often than not.

> (Example: Coin Flip "Random Walk.") Suppose Peter and Paul gamble with a fair coin. If it lands "Heads," then Peter must give \$1 to Paul; if it lands "Tails," then Paul must give \$1 to Peter. Suppose they flip the coin 40 times. In each game, Peter's winnings will either increase by \$1 or decrease by \$1, with probability $\frac{1}{2}$. This is known as a unbiased random walk.
>
> What is the probability distribution for Peter's winning after 40 games? Note that he will have won somewhere between \$40 and +\$40 after 40 games. Since the game is fair, intuitively, we expect 0 to be the most likely outcome.
>
> We note that the probability distribution actually looks like a bell-shaped curve.