9. DISCRETE PROBABILITY DISTRIBUTIONS

• Random Variable: A quantity that takes on different values depending on chance.

Eg: The number of Covid-19 cases next year, worldwide.

The proportion of visitors to the WSJ website who will click on your ad tomorrow.

Demand for iPhones in the next quarter.

The number of times someone will hail an Uber at the Empire State Building in the next hour.
• A random variable is the result of a random experiment in the abstract sense, before the experiment is performed.

The value the random variable actually assumes is called an observation.

Eg: “Next Quarter's Sales of Coca Cola” is a random variable, and the actual value of $3,521,395,576 is an observation of the RV.

• You can think of your data set as observations of a random variable resulting from several repetitions of a random experiment.

We associate the random variable with a population and view observations of the random variable as data.
Eg: Suppose we toss a coin five times. The observed data set is a sequence of zeros and ones, such as 1 1 0 1 0. Each of the five digits in this sequence represents the outcome of the random experiment of tossing a die once, where 1 denotes Heads and 0 denotes Tails. We have five repetitions of the experiment.

- A random variable is discrete if it can assume only a finite or countably infinite number of values.

What does “countably infinite” mean? We won’t try to define this precisely, but an important example (the only one we will consider) of a countably infinite set is the nonnegative integers, \{0, 1, 2, 3, \cdots \}. 
Eg: “World Population 10 Years From Today” is a discrete RV. Its possible values are 0, 1, 2, …

Eg: “Number of Heads in Two Coin Tosses” is a discrete RV, taking values 0, 1, 2 with probabilities 1/4, 1/2, 1/4.

• A random variable is **continuous** if it can assume any value in an interval of real numbers.

Eg: “Weight of a Randomly Selected Quarter Pounder” is a continuous RV. Its possible values are (in principle) all nonnegative real numbers.

Eg: Students: Give examples of random variables. Is the variable discrete or continuous?
A list of the possible values of a discrete RV, together with their associated probabilities.

The probability distribution tells us everything we can know about a random variable, before it becomes an observation.

**Eg: Distribution of # Heads in Two Tosses.**

<table>
<thead>
<tr>
<th>$x$</th>
<th>Prob{Number of Heads = $x$}</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/4</td>
</tr>
<tr>
<td>1</td>
<td>1/2</td>
</tr>
<tr>
<td>2</td>
<td>1/4</td>
</tr>
</tbody>
</table>
Some Notation

$p(x)$ denotes the probability that a discrete RV takes the value $x$. We will use uppercase letters to denote random variables.

**Eg:** $X = \#$ Heads in Two Coin Tosses. Note that $X$ is not a definite number. We don't know what value it will take until we do the experiment. If we do the experiment again, then $X$ might take a different value.

\[
\begin{align*}
\text{Prob}\{X = 0\} &= \text{Prob}\{\# \text{ Heads} = 0\} = p(0). \\
\text{Prob}\{X = x\} &= \text{Prob}\{\# \text{ Heads} = x\} = p(x).
\end{align*}
\]

Note that $X$ is just shorthand for “Number of Heads”, while $x$ represents a possible value for (an observation of) the number of heads.
Eg: \( \text{Prob}\{\text{At Least 1 Head}\} = \text{Prob}\{X \geq 1\} = p(1) + p(2) = \frac{3}{4}. \)

Eg: How many games will the World Series last? For any “Best 4 out of 7” series between two equally matched teams, the duration of the series is a discrete random variable with the following distribution:

<table>
<thead>
<tr>
<th>Duration of series</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.125</td>
</tr>
<tr>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>6</td>
<td>0.3125</td>
</tr>
<tr>
<td>7</td>
<td>0.3125</td>
</tr>
</tbody>
</table>

(We will prove this later).
• Requirements of Discrete Probability Distributions

\[ 0 \leq p(x) \leq 1 \text{ for all values of } x. \]

\[ \sum_{all \, x} p(x) = 1 \]

Explanation:
Since the probability \( p(x) \) is a proportion, it must be between zero (impossibility) and one (certainty).

We are guaranteed to get an outcome when we do the experiment.

Note: If a function \( p \) does not satisfy both requirements, it cannot be a probability distribution.