CHAPTER 6: DISCRETE PROBABILITY DISTRIBUTIONS

[This chapter is based on Chapter 6 from the textbook]

In Chapter 5 we learned to compute the probability of a variety of events, and we defined our events according to the outcomes we wanted to observe. These computations were pretty general, in the sense that we could describe the events with words or the desired outcomes. For example, in the case of rolling a die, we defined our events as $E =$"roll a 5" or, equivalently, $E = \{5\}$. This flexibility allowed us to compute probabilities associated to quantitative and qualitative variables. In this chapter, we dive deeper into quantitative variables; specifically, quantitative-discrete variables.

6.1 Discrete Random Variables

We start with a definition.

Definition 6.1. *A random variable is a numerical measure of the outcome of a probability experiment, so its value is determined by chance. Random variables are typically denoted using capital letters, such as X or Y and their possible values are denoted by lower-case letters, such as x and y.*

Some examples of random variables are:

- $X =$Number of heads in two flips of a coin. In this case, the possible values are $x = 0, 1, 2$; that is, the sample space is $S = \{0, 1, 2\}$
- $X =$Sum of the outcomes of rolling two dice. In this case, the sample space is $S = \{2, 3, \dots, 12\}$
- $T =$Time between arrivals of cars at a drive-through. In this case, the sample space is $S = \mathbb{R}_+$, that is, the possible values of T are all numbers $t > 0$.

Observe the difference between the sample space of the first two examples and the last one. This difference takes us to the next definition.

Definition 6.2.

(i) A discrete random variable has either a finite or a countable number of values.

(ii) A continuous random variable has infinitely many values, and the possible values represent a continuous range (or interval) of values.

Similarly to Chapter 1, a rule of thumb to decide whether a random variable is discrete or continuous is the following. Discrete random variables can easily repeat their value; however, two experiments of a continuous random variable will never give exactly the same outcome because the exact outcome involves many decimals. Let's see some examples.

Example 6.1. *Determine if the following random variables are discrete or continuous:*

(a) Number of cars that travel through a McDonald's drive thru

(b) Speed of the next car that passes a state trooper

(c) Number of people celebrating 4th of July at the First Landing Park

(d) Amount of alcohol (in ounces) a person drinks while celebrating 4th of July

Solution. Using our rule of thumb, we obtain

(a) Discrete

(b) Continuous

(c) Discrete

(d) Continuous

□

In this chapter we are interested in the probability of each of all the possible values of a random variable. We start with a definition.

Definition 6.3. *The probability distribution of a discrete random variable X provides the possible values of the random variable and their corresponding probabilities. A probability distribution can be in the form of a table, graph, or mathematical formula. We usually denote $P(x)$ the probability that the random variable X takes the value x.*

We can interpret the probability distribution of a discrete random variable as a probability model. Hence, it must satisfy the following properties.

Theorem 6.1. *Let $P(x)$ denote the probability that the random variable X equals x, and let S denote the state space. Then,*

(i) $\sum_{x \in S} P(x) = 1$

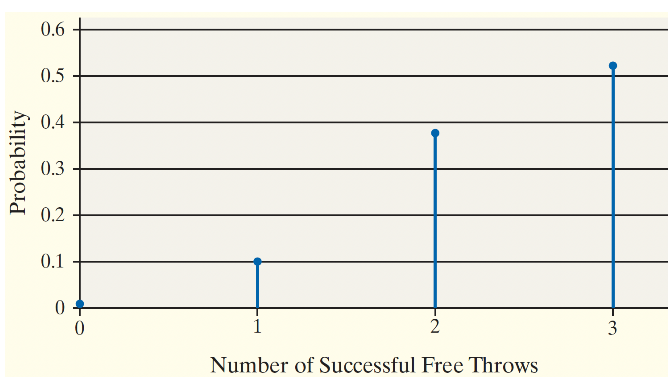
(ii) $0 \leq P(x) \leq 1$ *for all values* x

Let's see an example.

Example 6.2. *Suppose we ask a basketball player to shoot three free throws. Let the random variable X represent the number of shots made, so $x = 0, 1, 2$ or 3. The following table shows a probability distribution of the random variable X:*

x	$P(x)$
0	0.01
1	0.10
2	0.38
3	0.51

Equivalently, we can present the probability distribution with the following graph:



Then, we can read that the probability of the player shooting 1 throw is

$$P(1) = 0.10$$

Observe the similarity of probability distributions and relative frequency distributions. Using the definition of probability from the law of large numbers, the probability distribution represents the proportion of times we observe each value when we repeat the experiment many many many times. That is, the probability distribution can be interpreted as the relative frequency distribution when we repeat an experiment many many times.

Additionally, since probability distributions describe the possible values of a random variable (instead of values observed from a sample), they represent the population of experiments. With this analogy, we can comput the mean and standard deviation of a random variable as follows.

Theorem 6.2. *Consider a discrete random variable X with state space S and probability distribution $P(x)$. . Then,*

(i) The mean of the random variable X is computed as

$$\mu_X = \sum_{x \in S} x P(x)$$

The mean μ_X is also known as the expected value of X, which is denoted by $E(X)$.

(ii) The standard deviation of X is computed as

$$\sigma_X = \sqrt{\sum_{x \in S} (x - \mu_X)^2 P(x)}$$

or equivalently,

$$\sigma_X = \sqrt{\left(\sum_{x \in S} x^2 P(x)\right) - (\mu_X)^2}$$

Let's do some examples.

Example 6.3. *Let's consider again the example of the basketball player. Recall its probability distribution:*

x	$P(x)$
0	0.01
1	0.10
2	0.38
3	0.51

Compute the mean and the standard deviation of the number of shots.

Solution. First observe that the state space in this example is $S = \{0, 1, 2, 3\}$.
Now we compute the mean. We sum the value of the random variable times its probability. We obtain

$$\mu_X = 0 \cdot 0.01 + 1 \cdot 0.10 + 2 \cdot 0.38 + 3 \cdot 0.51 = 2.39$$

To compute the standard deviation, we use the value of the mean computed above: $\mu_X = 2.39$. We may use any of the two formulas given in the theorem. If we use the first formula, we obtain:

$$\begin{aligned}\sigma_X &= \sqrt{(0 - 2.39)^2 \cdot 0.01 + (1 - 2.39)^2 \cdot 0.10 + (2 - 2.39)^2 \cdot 0.38 + (3 - 2.39)^2 \cdot 0.51} \\ &= \sqrt{0.4979} \approx 0.7\end{aligned}$$

If we use the second formula, we obtain

$$\begin{aligned}\sigma_X &= \sqrt{(0^2 \cdot 0.01 + 1^2 \cdot 0.10 + 2^2 \cdot 0.38 + 3^2 \cdot 0.51) - (2.39)^2} \\ &= \sqrt{6.21 - 2.39^2} \\ &= \sqrt{0.4979} \approx 0.7\end{aligned}$$

□

6.2 The Binomial Probability Distribution

As we learned in the first part of this chapter, different discrete random variables have different probability distributions. In many cases, the probability distributions are presented in the form of a table. However, the tables are specific to each case. In this section, we will learn about a probability model that can be used in several real-life examples, and the probability distribution is presented as a formula. We start with a key definition.

Definition 6.4. *An experiment is said to be a binomial experiment if:*

(i) The experiment is performed a fixed number of times. Each repetition is called a trial

(ii) The trials are independent; that is, the outcome of one trial does not affect the outcome of the other trials.

(iii) For each trial, there are two disjoint outcomes: success or failure

(iv) The probability of success is the same for each trial of the experiment

In binomial experiments, we use the following notation:

n	:	Represents the (fixed) number of trials
p	:	Represents the probability of success in each trial. Then, $1 - p$ represents the probability of failure.
X	:	Represents the random variable that counts the number of successes in the n trials. Then, the possible values of X are $x = 0, 1, 2, \ldots, n$.

Let's see some examples.

Example 6.4. *For each of the following experiments, determine if they are binomial experiments. If yes, identify the number of trials, the probability of success and failure, and the possible values of the random variable X.*

(a) Flip a coin 15 times and count the number of heads.

(b) An experiment in which a basketball player who historically makes 80% of his free throws is asked to shoot three free throws, and the number of free throws made is recorded.

(c) According to a recent Harris Poll, 28% of Americans state that chocolate is their favorite flavor of ice cream. Suppose a simple random sample of size 10 is obtained and the number of Americans who choose chocolate as their favorite ice cream flavor is recorded.

(d) A probability experiment in which three cards are drawn from a deck without replacement and the number of aces is recorded.

Solution. In each case, we analyze if the four properties established in Definition 6.4.

(a) The experiment here is observing the outcome of tossing a coin. Let's go through the properties:

  (i) The experiment is performed a fixed number of times. In this case, $n = 15$

  (ii) The trials are independent because the outcome of a flip does not influence the future flips

  (iii) For each trial, there are two possible outcomes: heads and tails. Since we are counting the number of heads, we set heads as success and tails as failure.

  (iv) The probability of success is the same in each flip. Indeed, $p = 1/2$ and $1 - p = 1/2$.

  Hence, the experiment is a binomial experiment with possible values $x = 0, 1, \ldots, 15$.

(b) The experiment here is observing the outcome of three shots. Let's go through all the properties:

  (i) The experiment is performed a fixed number of times; in this case, that number is $n = 3$.

  (ii) The trials are independent because the outcome of the first throw does not affect the other two. Each time the basketball player shoots, they reset and throw again.

  (iii) In each shoot, the player succeeds or fails to make a free throw, and these are the only possible outcomes.

  (iv) The probability of a free throw is always 0.8. Then, the probability of success is $p = 0.8$ and the probability of failure is $1 - p = 0.2$.

Hence, the experiment described is a binomial experiment. Since the basketball player shoots three times, the possible outcomes are $x = 0, 1, 2, 3$.

(c) The experiment here is asking each of the 10 people if their favorite ice cream flavor is chocolate. Let's go through the four properties:

(i) We are going to ask exactly 10 people, so the number of trials is fixed and is $n = 10$.

(ii) The trials are independent because the preference of a specific person does not affect the preference of the other people.

(iii) The possible outcomes are "chocolate is their favorite ice cream flavor" or "chocolate is not their favorite ice cream flavor." Then, we can think of "chocolate is their favorite ice cream flavor" as success, and "chocolate is not their favorite ice cream flavor" as failure.

(iv) The probability of success is the same for each person according to the survey. Further, the probability of success is $p = 0.28$ and the probablity of failure is $1 - p = 0.72$.

Hence, the experiment is a binomial experiment. The random variable X represents the number of people whose favorite ice cream flavor is chocolate, and the possible values are $x = 0, 1, \ldots, 10$.

(d) The experiment here is drawing three consecutive cards from a deck and observing if the outcome is an A, without putting the cards drawn back to the deck before drawing the next. Let's go through the properties:

(i) The experiment is performed a fixed number of times; in this case, that number is $n = 3$

(ii) The trials are **not** independent because:

- If the first outcome is a success, that is, A, then there are three remaining A's in the deck for the second outcome. Hence, the probability that the second outcome is an A is $\frac{3}{51}$
- But if the first outcome is a failure, that is, not A, then there are four remaining A's in the deck for the second outcome. Hence, the probability that the second outcome is an A is $\frac{4}{51}$.

Hence, the outcome of the first trial **does influence** the probability of success of the second trial.

(iii) There are only two possible outcomes from the experiment, where success is drawing an A.

(iv) The probability of success is **not** the same in each trial because the number of possible cards is different: in the first trial we draw one card out of 52, in the second trial out of 51, and in the third trial out of 50. Further, the outcome of the first trial influences the probability of success in the second and third trial (as observed in point (ii)).

Hence, this experiment is **not** a binomial experiment.

Observe that we can modify the experiment to make it a binomial experiment. The problem with the current experiment is that the card we draw does not go back to the deck before drawing another card; then, the outcome of the first card changes the probabilities for the remaining cards. If we draw a card and return it to the deck before drawing another card, then we would have a binomial experiment.

□

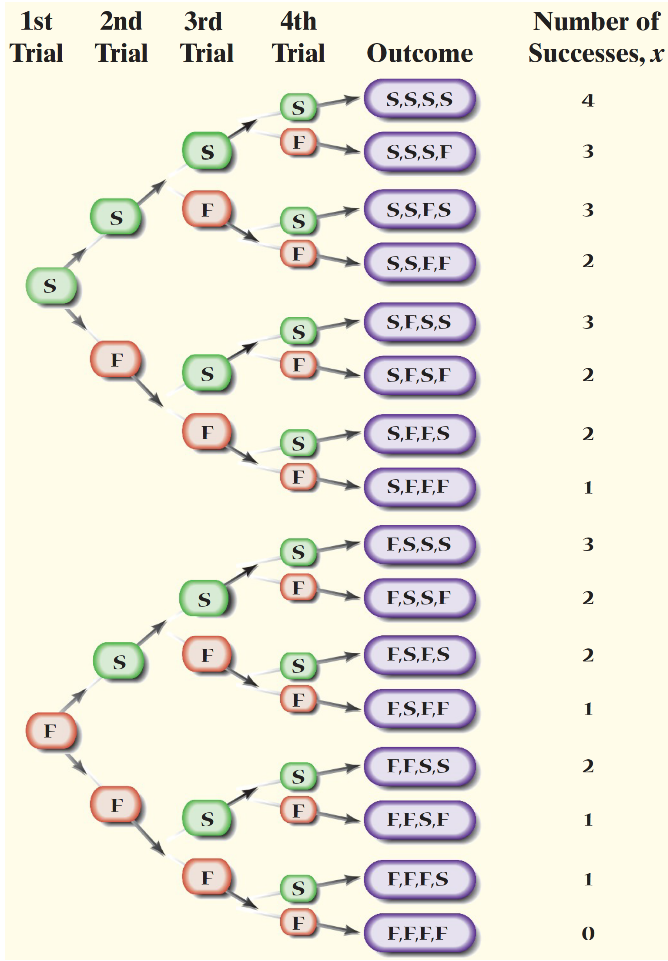
Now that we know what a binomial experiment is, we can define the binomial probability distribution.

Definition 6.5. *The binomial probability distribution is a discrete probability distribution that describes the probabilities for binomial experiments.*

Our next goal is computing the binomial probability distribution. We start with an example.

Example 6.5. *According to the American Red Cross, 7% of people in the United States have blood type O-negative. A simple random sample of size 4 is obtained, and the number of people X with blood type O-negative is recorded. Construct a probability distribution for the random variable X.*

Solution. We have the following possible outcomes of the experiment, where S represents success and F represents failure:



We now compute the probability of each possible value x. We start with $x = 0$, which means that we had 4 failures.

$$\begin{aligned} P(0) &= P(FFFF) \\ &= P(F)P(F)P(F)P(F) \qquad \text{(each trial is independent of the rest)} \\ &= 0.93 \cdot 0.93 \cdot 0.93 \cdot 0.93 \\ &= 0.93^4 \\ &= 0.74805 \end{aligned}$$

When $x = 1$, we have one success and three failures; and the success can be in the first, second, third or fourth trial. Then, we obtain

$$\begin{aligned} P(1) &= P(SFFF \text{ or } FSFF \text{ or } FFSF \text{ or } FFFS) \\ &= P(SFFF) + P(FSFF) + P(FFSF) + P(FFFS) \qquad \text{(disjoint outcomes of the 4 trials)} \\ &= P(S)P(F)P(F)P(F) + P(F)P(S)P(F)P(F) + P(F)P(F)P(S)P(F) + P(F)P(F)P(F)P(S) \qquad \text{(independence)} \\ &= 0.07 \cdot 0.93^3 + 0.07 \cdot 0.93^3 + 0.07 \cdot 0.93^3 + 0.07 \cdot 0.93^3 \\ &= 4 \cdot 0.07 \cdot 0.93^3 \\ &= 0.22522 \end{aligned}$$

When $x = 2$, we have two successes and two failures, and there are 6 ways to organize them according to the tree. We obtain

$$\begin{aligned} P(2) &= P(SSFF \text{ or } SFSF \text{ or } SFFS \text{ or } FSSF \text{ or } FSFS \text{ or } FFSS) \\ &= P(SSFF) + P(SFSF) + P(SFFS) + P(FSSF) + P(FSFS) + P(FFSS) \qquad \text{(disjoint events)} \end{aligned}$$

$$\begin{aligned}
&= P(S)P(S)P(F)P(F) + P(S)P(F)P(S)P(F) + P(S)P(F)P(F)P(S) + P(F)P(S)P(S)P(F) + \\
&\quad + P(F)P(S)P(F)P(S) + P(F)P(F)P(S)P(S) \qquad \text{(independence)} \\
&= 6 \cdot 0.93^2 \cdot 0.07^2 \\
&= 0.02543
\end{aligned}$$

Let's try to find the patterns in terms of $n, p, 1-p$ and x. Observe that:

$$\begin{aligned}
P(0) &= 1 \cdot 0.07^0 \cdot 0.93^{4-0} \\
P(1) &= 4 \cdot 0.07^1 \cdot 0.93^{4-1} \\
P(2) &= 6 \cdot 0.07^2 \cdot 0.93^{4-2}
\end{aligned}$$

where:

- the number in front of the expression represents the number of wais we obtain x successes and $n-x$ failures
- 0.07 represents the probability of success p
- 0.93 represents the probability of failure $1-p$
- x represents the number of successes and appears in the exponent of the probability of success, and in the exponent of the probability of failure to compute the number of failures $(n-x)$

□

Based on the example, we obtain the following theorem.

Theorem 6.3. *In a binomial experiment of n trials with probability of success p, the probability of obtaining x successes is:*

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

where

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} = \frac{1 \cdot 2 \cdots n}{(1 \cdot 2 \cdots x)(1 \cdot 2 \cdots (n-x))}$$

Let's see an example.

Example 6.6. *According to CTIA, 41% of all U.S. households are wireless-only households (no landline). In a random sample of 20 households, what is the probability that:*

(a) exactly 5 are wireless-only?

(b) fewer than 3 are wireless-only?

(c) at least 3 are wireless-only?

Solution. Observing if each of the 20 households selected is wireless-only is a binomial experiment with:

$$n = 20, \qquad p = 0.41, \qquad 1-p = 0.59$$

Then,

(a) The probability that 5 are wireless-only is:

$$P(5) = \binom{20}{5} 0.41^5 \cdot 0.59^{20-5} = 0.0656$$

(b) Fewer than 3 means that 0, 1 or 2 households are wireless-only, that is, we compute the probability that $X < 3$. Then, we compute

$$\begin{aligned} P(X < 3) &= P(0 \text{ or } 1 \text{ or } 2) \\ &= P(0) + P(1) + P(2) \qquad \text{(disjoint events)} \\ &= \binom{20}{0} 0.41^0 \cdot 0.59^{20} + \binom{20}{1} 0.41^1 \cdot 0.59^{20-1} + \binom{20}{2} 0.41^2 \cdot 0.59^{20-2} \\ &= 0.0028 \end{aligned}$$

(c) In this case we are interested in

$$\begin{aligned} P(X \leq 3) &= 1 - P(X > 3) \qquad \text{(complement)} \\ &= 1 - 0.028 = 0.9972 \end{aligned}$$

□

The binomial distribution is well studied, and indeed, we have a formula for its mean and standard deviation.

Theorem 6.4. *A binomial experiment with n independent trials and probability of success p has mean and standard deviation given by the following formulas:*

$$\mu_X = np \qquad \textit{and} \qquad \sigma_X = \sqrt{p(1-p)}$$

The computation of the mean above confirms our intuition. For example, if we toss a coin 20 times, we expect observing 10 heads (on average). Observe that the probability of success (heads) is 0.5 and we have $n = 20$. Then, the average number of heads is $np = 20 \cdot 0.5 = 10$.

Let's do an example.

Example 6.7. *Compute the mean and standard deviation of the number of households that are wireless-only in Example 6.6.*

Solution. We had a random sample of $n = 20$ households, and the probability of success is $p = 0.41$. Then, plugging in the formulas we obtain

$$\begin{aligned} \mu_X &= np = 20 \cdot 0.41 = 8.2 \\ \sigma_X &= \sqrt{p(1-p)} = \sqrt{0.41 \cdot 0.59} = \sqrt{0.2419} \approx 0.4918 \end{aligned}$$

□