Binomial Random Variable

The first discrete RV that we will discuss is the binomial random variable. The binomial random variable is a result of combining the outcomes from more than one trial. It arises from a binomial experiment. A binomial experiment is defined as having the following properties. Each trial has only two possible outcomes that we refer to as successes and failures. These trials are independent. The random variable is then the number of successes. It is similar to the idea of flipping a coin and recording the number of heads that you see. The binomial RV is like this but there are numerous processes that we can model using a binomial RV.

**Definition**: A binomial RV $X$ is a discrete RV that has the following characteristics:
1. There must be a fixed number of trials, $n$.
2. Each trial can result in one and only one of two possible outcomes, labeled “success” and “failure”.
3. The probability of success in a single trial, $p$, is the same for each trial.
4. The trials are independent, so that the probability of success is unaffected by the result of a previous trial.

$X$ is then the number of success from the $n$ trials.

It should be noted that the terms “success” and “failure” are arbitrary labels and do not connote “good” or “bad” outcomes.

We describe the distribution for a binomial RV with the following formula:

If we know $p$ and $n$,

$$P(X=x) = \binom{n}{x} p^x (1 - p)^{n-x} \text{ for } x = 0, 1, 2, \ldots, n.$$ 

The **mean** of a binomial RV $X$ is $\mu = np$.

The **standard deviation** of a binomial RV $X$ is $\sigma = \sqrt{np(1 - p)}$.

That is the mean is the number of trials multiplied by the probability of success on each trial.

**Example:**
The classic example of a binomial distribution is flipping fair coins. Suppose that we toss a penny 10 times and we assume that the penny is fair, meaning that the probability of heads is the same as the probability of tails, one-half or 50%. The random variable is then the number of heads. Let $X$ be the number of heads from 10 tosses of a fair penny.

Find the probability of getting exactly 4 heads.

The first question we must ask is the ‘Are 4 conditions for this to be a binomial random variable met?’.
1. Is there a fixed number of trials? Yes, $n = 10$.
2. Does each trial result in one of two possible outcomes? Yes since each trial results in either a head or a tail.
3. Is the probability of success the same for each trial? Since the penny is the same one tossed 10 times, it is very likely that this holds. Note that that probability is 0.5, so $p=0.5$.
4. Are the trials independent? This is often the hardest assumption to verify. However it is hard to imagine in this case that getting a head on one trial would effect the outcome of the next. So yes, they would seem to be independent.

We first note that $n$ is 10 and $p = 0.5$, since the total number of tosses is 10 and the probability of success (getting a head) is 0.5. Thus
\[ P(X=4) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{10}{4} 0.5^4 (0.5)^6 = 210 \times 0.0625 \times (0.0156) = 0.205 \]

Thus the probability of getting 4 heads from flipping 10 pennies is 20.5%. Another way to think of this is that if we repeated this experiment over and over, flipping 10 pennies each time, then 20.5% of those experiments we would get exactly 4 heads out of 10 pennies.

**Example:**
A psychologist is doing a study of spatial relationships. Suppose that 15 people will be given a sequence of mazes to complete in 10 minutes. Each person will be given the mazes separately. Past experience suggests that 20% of all people finish all the mazes correctly in the 10 minutes they are allotted. Find the probability that 5 of the 15 people will complete the mazes correctly in the 10 minutes they are given.

Again, we must ask whether or not the four conditions for a binomial RV are met.
1. Is there a fixed number of trials? Yes, since there are 15 people who will attempt to complete the task.
2. Does each trial result in one of two possible outcomes? Yes since each trial results in either correctly finishing the mazes in the time allotted or not. Define success as an individual correctly finishing the mazes in the time allotted.
3. Is the probability of success the same for each trial? This may not be the case since each individual has different abilities.
4. Are the trials independent? Yes, since each individual does their mazes separately.

There is some reason for concern that the third condition is violated. However, since we can go ahead and assume that the probability, \( p \), of success is the same for each individual. With that assumption made, then let \( X \) be the number of people who successfully complete all the mazes on time.

\[ P(X=5) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{15}{5} 0.2^5 (0.8)^{10} = 3003 \times (0.00032) \times (0.10737) = 0.1032. \]

Hence the probability of having 5 people out of twenty complete the mazes in the allotted time is about 10%.

**Calculating cumulative probabilities**
We are often interested in calculating probabilities other than those of the simple events of the two preceding sections. There we focused on calculating the probability of a single value. We might be interested in the probability of getting more than 7 maintenance delays in the next week in the second example of the previous section.

There are two possibilities. First for something like \( P(X<3) \) we could add up the probabilities of the simple events that make up this probability.

\[ P(X<3) = P(X=0) + P(X=1) + P(X=2). \]

Each one of these would have to be calculated using the formula that was appropriate. It gets worse for something like

\[ P(X \leq 15) = P(X=0) + P(X=1) + P(X=2) + \ldots + P(X=15). \]

Our second option is to have a set of tables that will allow us to simplify these calculations. This is what the author of our textbook has done. The drawback to this is that we need to learn how to use the Tables.

The first piece of vital information is that the tables for both binomial random variables and Poisson random variables are cumulative tables. That is, instead of giving \( P(X=4) \) or \( P(Y<3) \), they give \( P(X\leq3) \) and \( P(Y\leq7) \) and \( P(G\leq2) \). As a consequence of this we need to use formulae to determine the other probabilities that we are interested in. Thus the table below. In this table the letter \( r \) represents the value that we need
for the probability we want. For the examples in the last column, what is on the right hand side of the equal sign

<table>
<thead>
<tr>
<th>Probability we want</th>
<th>Calculation we need to perform</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(Y \leq r)$</td>
<td>$P(Y \leq r)$</td>
<td>$P(Y \leq 4)$</td>
</tr>
<tr>
<td>$P(Y &lt; r)$</td>
<td>$P(Y \leq (r-1))$</td>
<td>$P(Y &lt; 4) = P(Y \leq 3)$</td>
</tr>
<tr>
<td>$P(Y \geq r)$</td>
<td>$1 - P(Y \leq (r-1))$</td>
<td>$P(Y \geq 4) = 1 - P(Y \leq 3)$</td>
</tr>
<tr>
<td>$P(Y &gt; r)$</td>
<td>$1 - P(Y \leq r)$</td>
<td>$P(Y &gt; 4) = 1 - P(Y \leq 4)$</td>
</tr>
<tr>
<td>$*P(Y=r)$</td>
<td>$P(Y \leq r) - P(Y \leq (r-1))$</td>
<td>$P(Y=r) = P(Y \leq r) - P(Y \leq (r-1))$</td>
</tr>
</tbody>
</table>

* This probability could also be calculated using the formula for the relevant random variable.

The above rules are the result of two basic ideas. The first is one that we saw in the previous chapter. This is the complement of a probability. Hence

$$ P(X \leq 4) = 1 - P(X > 4). $$

The second is an observation that

$$ P(X < 4) = P(X \leq 3), $$

since both contain the same simple events. The tables give us values that are less than or equal to a value $r$. Consequently, we must use a formula for the probability that we want that include probabilities with less than or equal to values in them.

Having mastered these rules, let’s consider the tables.

The binomial tables (Table 1 in the textbook) begin on page 505 of the textbook. We want to find $P(X \leq r)$ for some binomial random variable $X$. In order to use these tables we need to know what $n$ and $p$ are. Recall that we needed to know the values of those two quantities before we could do any calculations of binomial probabilities. The first step in the process is to find the value of $p$ across the top row of pages 505 to 509. Having found $p$, we then go down the first column on the left until we find the correct $n$. That will place us in a subtable of the entire table. Subtables are separated by a blank row. Having found the right subtable, we then find the appropriate value of $r$ that we need and use that row. We then go across that row until we hit the column for our value of $p$.

**Example:**
Suppose that $n = 10$ and $p = 0.15$. And we want $P(X \leq 2)$

Start on page 505. The second value at the top of the page $p = 0.15$ is the column we need. Go down the first column on the left until you reach $n = 10$. The $r$ that we want is 2, so find the row in the subtable that corresponds to $r = 2$. Then go across that row to the second column (it’s the one underneath $p = 0.15$). The value you find there 0.8202 = $P(X \leq 2)$.

**Example:**
Suppose that $n = 5$ and $p = 0.30$. We want $P(X \leq 4)$

Start on page 505. The fourth column at the top of the page corresponds to $p = 0.30$. Go down the first column on the left until you come to $n = 5$. We will use this subtable to get the probability we need. Find the row corresponding to $r = 4$. Go across that row until you reach the column for $p = 0.30$. The value at that row and column is 0.9976. Consequently $P(X \leq 4) = 0.9976$. 


Example:
Find P(X ≤ 3) with n = 20 and p = 0.08.
P(X≤3) = 0.9294

Example:
Find P(X ≤ 6) with n = 20 and p = 0.25.
P(X≤6) = 0.7858

Example:
Find P(X > 3) with n = 10 and p = 0.4.
Since we want P(X > 3) = 1 - P(X ≤ 2), we need to look up P(X ≤ 2) in the table.
P(X≤2) = 0.1673
So P(X > 3) = 1 - P(X ≤ 2) = 1 - 0.1673 = 0.8627

Example:
Find P(X < 2) with n = 5 and p = 0.20.
Since we want P(X < 2) = P(X ≤ 1), so we need to look up P(X ≤ 1) in the table.
P(X≤1) = 0.7373.
So P(X < 2) = P(X ≤ 1) = 0.7373

Example:
Find P(X ≥ 5) with n = 10 and p = 0.35.
Since we want P(X ≥ 5) = 1 - P(X ≤ 4), we need to look up P(X ≤ 4) in the table.
P(X≤4) = 0.7515
So P(X ≥ 5) = 1 - P(X ≤ 4) = 1 - 0.7515 = 0.2485.

Example:
Find P(X = 4) with n = 10 and p = 0.15.
Since we want P(X = 4) = P(X ≤ 4) - P(X ≤ 3), we need to look up both P(X ≤ 4) and P(X ≤ 3) in the table.
P(X≤4) = 0.9901 and P(X≤3) = 0.9500
So P(X = 4) = P(X ≤ 4) - P(X ≤ 3) = 0.9901 - 0.9500 = 0.0401.