

4.3: Binomial Distribution

A more valuable probability density function with many applications is the binomial distribution. This distribution will compute probabilities for any binomial process. A binomial process, often called a Bernoulli process after the first person to fully develop its properties, is any case where there are only two possible outcomes in any one trial, called successes and failures. It gets its name from the binary number system where all numbers are reduced to either 1's or 0's, which is the basis for computer technology and CD music recordings.

Binomial Formula

$$b(x) = \binom{n}{x} p^x q^{n-x} \quad (4.3.1)$$

where $b(x)$ is the probability of X successes in n trials when the probability of a success in ANY ONE TRIAL is p . And of course $q = (1 - p)$ and is the probability of a failure in any one trial.

We can see now why the combinatorial formula is also called the binomial coefficient because it reappears here again in the binomial probability function. For the binomial formula to work, the probability of a success in any one trial must be the same from trial to trial, or in other words, the outcomes of each trial must be independent. Flipping a coin is a binomial process because the probability of getting a head in one flip does not depend upon what has happened in PREVIOUS flips. (At this time it should be noted that using p for the parameter of the binomial distribution is a violation of the rule that population parameters are designated with Greek letters. In many textbooks θ (pronounced theta) is used instead of p and this is how it should be.)

Just like a set of data, a probability density function has a mean and a standard deviation that describes the data set. For the binomial distribution these are given by the formulas:

$$\begin{aligned} \mu &= np \\ \sigma &= \sqrt{npq} \end{aligned} \quad (4.3.2)$$

Notice that p is the only parameter in these equations. The binomial distribution is thus seen as coming from the one-parameter family of probability distributions. In short, we know all there is to know about the binomial distribution once we know p , the probability of a success in any one trial.

In probability theory, under certain circumstances, one probability distribution can be used to approximate another. We say that one is the limiting distribution of the other. If a small number is to be drawn from a large population, even if there is no replacement, we can still use the binomial even though this is not a binomial process. If there is no replacement it violates the independence rule of the binomial. Nevertheless, we can use the binomial to approximate a probability that is really a hypergeometric distribution if we are drawing fewer than 10 percent of the population, i.e. n is less than 10 percent of N in the formula for the hypergeometric function. The rationale for this argument is that when drawing a small percentage of the population we do not alter the probability of a success from draw to draw in any meaningful way. Imagine drawing from not one deck of 52 cards but from 6 decks of cards. The probability of say drawing an ace does not change the conditional probability of what happens on a second draw in the same way it would if there were only 4 aces rather than the 24 aces now to draw from. This ability to use one probability distribution to estimate others will become very valuable to us later.

There are three characteristics of a binomial experiment.

1. There are a fixed number of trials. Think of trials as repetitions of an experiment. The letter n denotes the number of trials.
2. The random variable, x , number of successes, is discrete.
3. There are only two possible outcomes, called "success" and "failure," for each trial. The letter p denotes the probability of a success on any one trial, and q denotes the probability of a failure on any one trial. $p + q = 1$.
4. The n trials are independent and are repeated using identical conditions. Think of this as drawing WITH replacement. Because the n trials are independent, the outcome of one trial does not help in predicting the outcome of another trial. Another way of saying this is that for each individual trial, the probability, p , of a success and probability, q , of a failure remain the same. For example, randomly guessing at a true-false statistics question has only two outcomes. If a success is guessing correctly, then a failure is guessing incorrectly. Suppose Joe always guesses correctly on any statistics true-false question with a probability $p = 0.6$. Then, $q = 0.4$. This means that for every true-false statistics question Joe answers, his probability of success ($p = 0.6$) and his probability of failure ($q = 0.4$) remain the same.

The outcomes of a binomial experiment fit a binomial probability distribution. The random variable X = the number of successes obtained in the n independent trials.

The mean, μ , and variance, σ^2 , for the binomial probability distribution are $\mu = np$ and $\sigma^2 = npq$. The standard deviation, σ , is then $\sigma = \sqrt{npq}$.

Any experiment that has characteristics three and four and where $n = 1$ is called a Bernoulli Trial (named after Jacob Bernoulli who, in the late 1600s, studied them extensively). A binomial experiment takes place when the number of successes is counted in one or more Bernoulli Trials.

? Exercise 4.3.1

Suppose you play a game that you can only either win or lose. The probability that you win any game is 55%, and the probability that you lose is 45%. Each game you play is independent. If you play the game 20 times, write the function that describes the probability that you win 15 of the 20 times. Here, if you define X as the number of wins, then X takes on the values 0, 1, 2, 3, ..., 20. The probability of a success is $p = 0.55$. The probability of a failure is $q = 0.45$. The number of trials is $n = 20$. The probability question can be stated mathematically as $P(x = 15)$.

$P(x = 15)$ stated more carefully in English would be stated as "the probability of exactly 15 wins in 20 trials." $P(x < 15)$ would need again a more careful statement because if we desire $P(x < 15)$ we need to calculate all the values of x from $P : (x = 0, 1, 2, \dots, 14)$.

Similarly we may wish to know the probability "greater than" some number, which requires multiple calculations of the values of x . ? Tr

Try it 4.3.1

A trainer is teaching a rescued dolphin to catch live fish before returning it to the wild. The probability that the dolphin successfully catches a fish is 35%, and the probability that the dolphin does not successfully perform the trick is 65%. Out of 20 attempts, you want to find the probability that the dolphin succeeds 12 times. Find the $P(X=12)$ using the binomial Pdf.

? Exercise 4.3.2

A coin has been altered to weight the outcome from 0.5 to 0.25 and is flipped 5 times. Each flip is independent. What is the probability of getting more than 3 heads? Let X = the number of heads in 5 flips of the fair coin. X takes on the values 0, 1, 2, 3, 4, 5. Since the coin is altered to result in $p = 0.25$, q is 0.75. The number of trials is $n = 5$. State the probability question mathematically.

Answer

$$P(x > 3)$$

First develop fully the probability density function and graph the probability density function. With the fully developed probability density function we can simply read the solution to the question $P(x > 3)$ heads. $P(x > 3) = P(x = 4) + P(x = 5) = 0.0146 + 0.0007 = 0.0153$. We have added the two individual probabilities because of the addition rule from Probability Topics.

Figure 4.3.1 also allows us to see the link between the probability density function and probability and area. We also see in Figure 4.2 the skew of the binomial distribution when p is not equal to 0.5. In Figure 4.2 the distribution is skewed right as a result of $\mu = np = 1.25$ because $p = 0.25$.

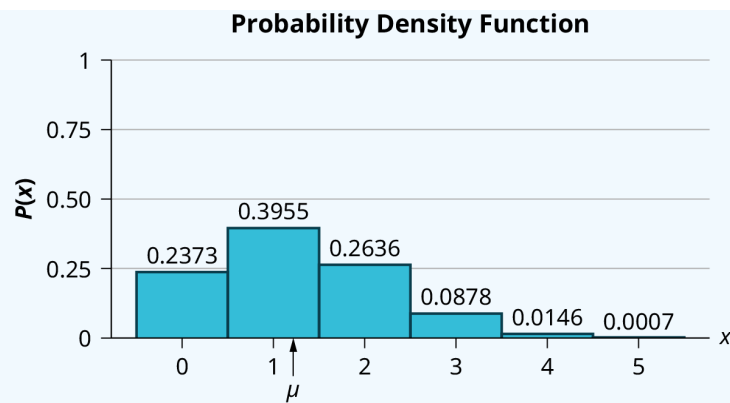


Figure 4.3.1

$$\begin{aligned}
 P(x = x_0) &= \binom{n}{x} p^x (1-p)^{n-x} \\
 &= \binom{5}{x_0} \cdot 25^{x_0} \cdot 75^{5-x_0} \\
 &\text{etc.} \\
 \mu &= np = 1.25
 \end{aligned}$$

Try it 4.3.2

A fair, six-sided die is rolled ten times. Each roll is independent. You want to find the probability of rolling a one more than three times. State the probability question mathematically.

? Exercise 4.3.3

Approximately 70% of statistics students do their homework in time for it to be collected and graded. Each student does homework independently. In a statistics class of 50 students, what is the probability that at least 40 will do their homework on time? Students are selected randomly.

- This is a binomial problem because there is only a success or a _____, there are a fixed number of trials, and the probability of a success is 0.70 for each trial.
- If we are interested in the number of students who do their homework on time, then how do we define X ?
- What values does x take on?
- What is a "failure," in words?
- If $p + q = 1$, then what is q ?
- The words "at least" translate as what kind of inequality for the probability question $P(x \text{ ____ } 40)$.

Answer

- failure
 - X = the number of statistics students who do their homework on time
 - 0, 1, 2, ..., 50
 - Failure is defined as a student who does not complete their homework on time.
- The probability of a success is $p = 0.70$. The number of trials is $n = 50$.
- $q = 0.30$
 - greater than or equal to (\geq)

The probability question is $P(x \geq 40)$.

Try iT 4.3.3

1. Sixty-five percent of people pass the state driver's exam on the first try. A group of 50 individuals who have taken the driver's exam is randomly selected. Give two reasons why this is a binomial problem.
2. During a certain NBA season, a player for the Los Angeles Clippers had the highest field goal completion rate in the league. This player scored with 61.3% of his shots. Suppose you choose a random sample of 80 shots made by this player during the season. Let X = the number of shots that scored points.
 - a. What is the probability distribution for X ?
 - b. Using the formulas, calculate the (i) mean and (ii) standard deviation of X .
 - c. Find the probability that this player scored with 60 of these shots.
 - d. Find the probability that this player scored with more than 50 of these shots.

This page titled [4.3: Binomial Distribution](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.