Statistics Tutorial: Binomial Distribution

To understand binomial distributions and binomial probability, it helps to understand binomial experiments and some associated notation; so we cover those topics first.

Binomial Experiment

A **binomial experiment** (also known as a **Bernoulli trial**) is a [statistical experiment](http://stattrek.com/Help/Glossary.aspx?Target=Statistical_experiment) that has the following properties:

* The experiment consists of *n* repeated trials.
* Each trial can result in just two possible outcomes. We call one of these outcomes a success and the other, a failure.
* The probability of success, denoted by *P*, is the same on every trial.
* The trials are [independent](http://stattrek.com/Help/Glossary.aspx?Target=Independent); that is, the outcome on one trial does not affect the outcome on other trials.

Consider the following statistical experiment. You flip a coin 2 times and count the number of times the coin lands on heads. This is a binomial experiment because:

* The experiment consists of repeated trials. We flip a coin 2 times.
* Each trial can result in just two possible outcomes - heads or tails.
* The probability of success is constant - 0.5 on every trial.
* The trials are independent; that is, getting heads on one trial does not affect whether we get heads on other trials.

Notation

The following notation is helpful, when we talk about binomial probability.

* *x*: The number of successes that result from the binomial experiment.
* *n*: The number of trials in the binomial experiment.
* *P*: The probability of success on an individual trial.
* *Q*: The probability of failure on an individual trial. (This is equal to 1 - *P*.)
* b(*x*; *n, P*): Binomial probability - the probability that an *n*-trial binomial experiment results in exactly *x* successes, when the probability of success on an individual trial is *P*.
* nCr: The number of [combinations](http://stattrek.com/Help/Glossary.aspx?Target=Combination) of *n* things, taken *r* at a time.

Binomial Distribution

A **binomial random variable** is the number of successes *x* in *n* repeated trials of a binomial experiment. The [probability distribution](http://stattrek.com/Help/Glossary.aspx?Target=Probability_distribution) of a binomial random variable is called a **binomial distribution** (also known as a **Bernoulli distribution**).

Suppose we flip a coin two times and count the number of heads (successes). The binomial random variable is the number of heads, which can take on values of 0, 1, or 2. The binomial distribution is presented below.

|  |  |
| --- | --- |
| **Number of heads**  | **Probability** |
| 0  | 0.25 |
| 1  | 0.50 |
| 2  | 0.25 |

The binomial distribution has the following properties:

* The mean of the distribution (μx) is equal to *n* \* *P* .
* The [variance](http://stattrek.com/Help/Glossary.aspx?Target=Variance) (σ2x) is *n* \* *P* \* ( 1 - *P* ).
* The [standard deviation](http://stattrek.com/Help/Glossary.aspx?Target=Standard%20deviation) (σx) is sqrt[ *n* \* *P* \* ( 1 - *P* ) ].

Binomial Probability

The **binomial probability** refers to the probability that a binomial experiment results in exactly *x* successes. For example, in the above table, we see that the binomial probability of getting exactly one head in two coin flips is 0.50.

Given *x*, *n*, and *P*, we can compute the binomial probability based on the following formula:

**Binomial Formula.** Suppose a binomial experiment consists of *n* trials and results in *x* successes. If the probability of success on an individual trial is *P*, then the binomial probability is:

b(*x*; *n, P*) = nCx \* Px \* (1 - P)n - x

**Example 1**

Suppose a die is tossed 5 times. What is the probability of getting exactly 2 fours?

*Solution:* This is a binomial experiment in which the number of trials is equal to 5, the number of successes is equal to 2, and the probability of success on a single trial is 1/6 or about 0.167. Therefore, the binomial probability is:

b(2; 5, 0.167) = 5C2 \* (0.167)2 \* (0.833)3
b(2; 5, 0.167) = 0.161

Cumulative Binomial Probability

A **cumulative binomial probability** refers to the probability that the binomial random variable falls within a specified range (e.g., is greater than or equal to a stated lower limit and less than or equal to a stated upper limit).

For example, we might be interested in the cumulative binomial probability of obtaining 45 or fewer heads in 100 tosses of a coin (see Example 1 below). This would be the sum of all these individual binomial probabilities.

b(x < 45; 100, 0.5) =
b(x = 0; 100, 0.5) + b(x = 1; 100, 0.5) + ... + b(x = 44; 100, 0.5) + b(x = 45; 100, 0.5)

Binomial Calculator

As you may have noticed, the binomial formula requires many time-consuming computations. The Binomial Calculator can do this work for you - quickly, easily, and error-free. Use the Binomial Calculator to compute binomial probabilities and cumulative binomial probabilities. The calculator is free. It can be found under the Stat Tables tab, which appears in the header of every Stat Trek web page.

|  |
| --- |
| [Binomial Calculator](http://stattrek.com/Tables/Binomial.aspx)  |

**Example 1**

What is the probability of obtaining 45 or fewer heads in 100 tosses of a coin?

*Solution:* To solve this problem, we compute 46 individual probabilities, using the binomial formula. The sum of all these probabilities is the answer we seek. Thus,

b(x < 45; 100, 0.5) = b(x = 0; 100, 0.5) + b(x = 1; 100, 0.5) + . . . + b(x = 45; 100, 0.5)
b(x < 45; 100, 0.5) = 0.184

**Example 2**

The probability that a student is accepted to a prestigeous college is 0.3. If 5 students from the same school apply, what is the probability that at most 2 are accepted?

*Solution:* To solve this problem, we compute 3 individual probabilities, using the binomial formula. The sum of all these probabilities is the answer we seek. Thus,

b(x < 2; 5, 0.3) = b(x = 0; 5, 0.3) + b(x = 1; 5, 0.3) + b(x = 2; 5, 0.3)
b(x < 2; 5, 0.3) = 0.1681 + 0.3601 + 0.3087
b(x < 2; 5, 0.3) = 0.8369

**Example 3**

What is the probability that the world series will last 4 games? 5 games? 6 games? 7 games? Assume that the teams are evenly matched.

*Solution:* This is a very tricky application of the binomial distribution. If you can follow the logic of this solution, you have a good understanding of the material covered in the tutorial, to this point.

In the world series, there are two baseball teams. The series ends when the winning team wins 4 games. Therefore, we define a success as a win by the team that ultimately becomes the world series champion.

For the purpose of this analysis, we assume that the teams are evenly matched. Therefore, the probability that a particular team wins a particular game is 0.5.

Let's look first at the simplest case. What is the probability that the series lasts only 4 games. This can occur if one team wins the first 4 games. The probability of the National League team winning 4 games in a row is:

b(4; 4, 0.5) = 4C4 \* (0.5)4 \* (0.5)0 = 0.0625

Similarly, when we compute the probability of the American League team winning 4 games in a row, we find that it is also 0.0625. Therefore, probability that the series ends in four games would be 0.0625 + 0.0625 = 0.125; since the series would end if either the American or National League team won 4 games in a row.

Now let's tackle the question of finding probability that the world series ends in 5 games. The trick in finding this solution is to recognize that the series can only end in 5 games, if one team has won 3 out of the first 4 games. So let's first find the probability that the American League team wins exactly 3 of the first 4 games.

b(3; 4, 0.5) = 4C3 \* (0.5)3 \* (0.5)1 = 0.25

Okay, here comes some more tricky stuff, so listen up. Given that the American League team has won 3 of the first 4 games, the American League team has a 50/50 chance of winning the fifth game to end the series. Therefore, the probability of the American League team winning the series in 5 games is 0.25 \* 0.50 = 0.125. Since the National League team could also win the series in 5 games, the probability that the series ends in 5 games would be 0.125 + 0.125 = 0.25.

The rest of the problem would be solved in the same way. You should find that the probability of the series ending in 6 games is 0.3125; and the probability of the series ending in 7 games is also 0.3125.

While this is statistically correct in theory, over the years the actual world series has turned out differently, with more series than expected lasting 7 games. For an interesting discussion of why world series reality differs from theory, see Ben Stein's explanation of [why 7-game world series are more common than expected](http://www.aip.org/isns/reports/2003/080.html).