Binomial Distribution

A binomial probability distribution consists of a series of **independent trials**, each of which results in either **success** or **failure**. The probability of success is the same for each trial.

Example

For example, suppose a company manufactures USB cables. They know from historical data that each cable has a 5% probability of being defective. Suppose also that a technician selects 10 cables at random to test. What is the probability that **exactly** 2 of those cables are defective?

Let's answer this in stages. First, let's find the probability that the first 8 cables are good and the last 2 are defective (that is, we'll calculate the probability of getting 2 defective cables in a **specific** order). This probability is

$$\begin{aligned}(0.95)(0.95)(0.95)(0.95)(0.95)(0.95)(0.95)(0.95)(0.05)(0.05) &= (0.95)^8(0.05)^2 \\ &\doteq 0.0017\end{aligned}$$

So there is a 0.17% probability of getting 8 good cables followed by 2 defective cables.

However, the ordering of cables doesn't matter; we need 2 defectives in *any* order. So we can multiply this probability, 0.17%, by the number of orders in which 8 good and 2 defective can occur, which is $\binom{10}{8}$:

$$\begin{aligned}\binom{10}{8}(0.95)^8(0.05)^2 &\doteq 0.075 \\ &= 7.5\%\end{aligned}$$

It's sometimes confusing why we've used a combination to count the number of orders. To see this, think of successes as identical items and failures as identical items. I like to think of them as letters: S for success and F for failure. Our original order can be represented by

$$SSSSSSSSFF$$

How many orders are possible?

$$\frac{10!}{8!\,2!}$$

because there are 8 identical successes and 2 identical failures. This is equivalent to $\binom{10}{8}$.

Another way to think of this is that you are "choosing" 8 positions out of 10 in which to place the 8 successes. The 2 failures will fill in the remaining 2 positions.

In general

For a random variable X representing the number of successes in a series of n independent trials, each with probability of success p, the probability of getting exactly k successes is

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

In this formula, $\binom{n}{k}$ tells us the number of orders in which the successes and failures can occur; p^k tells us the probability of getting k successes in a row; and $(1-p)^{n-k}$ tells us the probability of getting $n-k$ failures in a row.

To describe a probability distribution like this in a table, we only need to know n and p.

Example

Suppose you roll a d6 die 4 times, and you score a point each time you roll a 6. Complete a table that shows the distribution of 6s.

We know the probability of rolling a 6 is $\frac{1}{6}$; this is the probability of success on each trial. There are 4 trials.

k	$P(X = k)$
0	$\binom{4}{0}\left(\frac{1}{6}\right)^{0}\left(1-\frac{1}{6}\right)^{4-0} \doteq 0.48225$
1	$\binom{4}{1}\left(\frac{1}{6}\right)^{1}\left(1-\frac{1}{6}\right)^{4-1} \doteq 0.38580$
2	$\binom{4}{2}\left(\frac{1}{6}\right)^{2}\left(1-\frac{1}{6}\right)^{4-2} \doteq 0.11574$
3	$\binom{4}{3}\left(\frac{1}{6}\right)^{3}\left(1-\frac{1}{6}\right)^{4-3} \doteq 0.01543$
4	$\binom{4}{4}\left(\frac{1}{6}\right)^{4}\left(1-\frac{1}{6}\right)^{4-4} \doteq 0.00077$

In reading this table we can see, for example, that the probability of rolling exactly 2 sixes is about 11.6%.

Expected Value

If you conduct n independent trials, each with a probability of success p, you expect

$$E(X) = np$$

successes. Hopefully this is intuitive. For example, suppose you conduct 20 trials, each with a probability of success 30%. You would expect 30% of those trials to be successes, or $30\% \times 20 = 6$ successes (and therefore $20 - 6 = 14$ failures).

Using our original formula for expected value for a *single* trial, we consider the value of a success to be $x_1 = 1$ and the value of a failure to be $x_2 = 0$.

$$\begin{aligned} x_1 P(x_1) + x_2 P(x_2) &= (1)(p) + (0)(1-p) \\ &= p \end{aligned}$$

Since there are n trials, we have n times this expected value, or

$$E(X) = np$$

One more example

Suppose you hit 70% of free throws you take playing basketball. If you have 6 free throws in a game, what is the probability that you **hit at least 5** of them?

Solution

Hitting at least 5 free throws means hitting exactly 5 or exactly 6. We calculate each case separately and add the results.

For this situation we assume a binomial distribution (independent trials with a fixed probability of success). We have

$$n = 6$$

$$p = 0.7$$

We calculate

$$\begin{aligned} P(5) &= \binom{6}{5}(0.7)^5(1-0.7)^{6-5} \\ &= \binom{6}{5}(0.7)^5(0.3)^1 \\ &\doteq 0.3025 \end{aligned}$$

$$\begin{aligned} P(6) &= \binom{6}{6}(0.7)^6(1-0.7)^{6-6} \\ &= \binom{6}{6}(0.7)^6(0.3)^0 \\ &\doteq 0.1176 \end{aligned}$$

Together we have

$$\begin{aligned} P(5) + P(6) &\doteq 0.3025 + 0.1176 \\ &= 0.4201 \end{aligned}$$

So you have about a 42% chance of hitting at least 5 of your 6 free throws.