The Binomial Distribution

Advanced

"Bi" means "two" (like a bicycle has two wheels) ...
... so this is about things with **two results**.

Tossing a Coin:

- Did we get Heads (H) or
- Tails (T)

We say the probability of the coin landing **H** is ½
And the probability of the coin landing **T** is ½

Throwing a Die:

- Did we get a four ... ?
- ... or not?

We say the probability of a **four** is 1/6 (one of the six faces is a four).
And the probability of **not four** is 5/6 (five of the six faces are not a four)

Let's Toss a Coin!

Toss a fair coin **three times** ... what is the chance of getting **two Heads**?

Tossing a coin three times (**H** is for heads, **T** for Tails) can get any of these 8 **outcomes**:

HHH 

HHT 

HTH 

HTT

THH 

THT 

TTH 

TTT 

Which outcomes do we want?

"Two Heads" could be in any order: "HHT", "THH" and "HTH" all have two Heads (and one Tail).

So **3 of the outcomes** produce "Two Heads".

What is the probability of each outcome?

Each outcome is equally likely, and there are 8 of them. So each has a probability of 1/8

So the probability of **event** "Two Heads" is:

Number of outcomes we want		Probability of each outcome		
3	×	1/8	=	3/8

We used special words:

- **Outcome**: the result of three coin tosses (8 different possibilities)
- **Event**: "Two Heads" out of three coin tosses (3 possibilities)

Let's Calculate Them All:

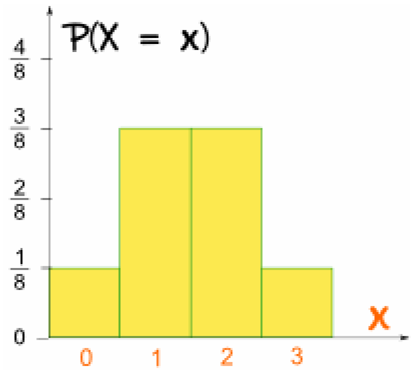
The calculations are (P means "Probability of"):

- P(Three Heads) = P(**HHH**) = **1/8**
- P(Two Heads) = P(**HHT**) + P(**HTH**) + P(**THH**) = 1/8 + 1/8 + 1/8 = **3/8**
- P(One Head) = P(**HTT**) + P(**THT**) + P(**TTH**) = 1/8 + 1/8 + 1/8 = **3/8**
- P(Zero Heads) = P(**TTT**) = **1/8**

We can write this in terms of a Random Variable, X, = "The number of Heads from 3 tosses of a coin":

- P(X = 3) = 1/8
- P(X = 2) = 3/8
- P(X = 1) = 3/8
- P(X = 0) = 1/8

And we can also draw a Bar Graph:



It is symmetrical!

Making a Formula

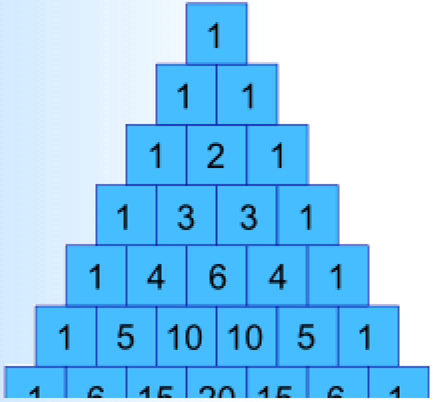
Now ... what are the chances of **5 heads in 9 tosses** ... to list all outcomes (512) will a long time!

So let's make a formula.

In our previous example, how could we get the values 1, 3, 3 and 1 ?

They are actually in the third row of Pascal’s Triangle ... !

Can we make them using a formula?



Sure we can, and here it is:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

It is often called **"n choose k"**

- n = total number
- k = number we want
- the "!" means "factorial", for example 4! = 1×2×3×4 = 24

You can read more about it at [Combinations and Permutations](combinations-permutations).

Let's use it:

Example: 3 tosses getting 2 Heads

We have **n=3** and **k=2**

$$\frac{n!}{k!(n-k)!} = \frac{3!}{2!(3-2)!} = \frac{3\times2\times1}{2\times1 \times 1} = 3$$

So there are 3 outcomes for "2 Heads"

(We knew that already, but now we have a formula for it.)

Let's use it for a harder question:

Example: what are the chances of 5 heads in 9 tosses?

We have **n=9** and **k=5**

$$\frac{n!}{k!(n-k)!} = \frac{9!}{5!(9-5)!} = \frac{9\times8\times7\times6\times5\times4\times3\times2\times1}{5\times4\times3\times2\times1 \times 4\times3\times2\times1} = 126$$

And for 9 tosses there are $2^9 = 512$ total outcomes, so we get the probability:

Number of outcomes we want		Probability of each outcome		
126	$\times$	$\frac{1}{512}$	=	$\frac{126}{512}$

So:

$$P(X=5) = \frac{126}{512} = \frac{63}{256} = 0.24609375$$

About a **25% chance.**

(Easier than listing them all.)

Bias!

So far the chances of success or failure have been **equally likely**.

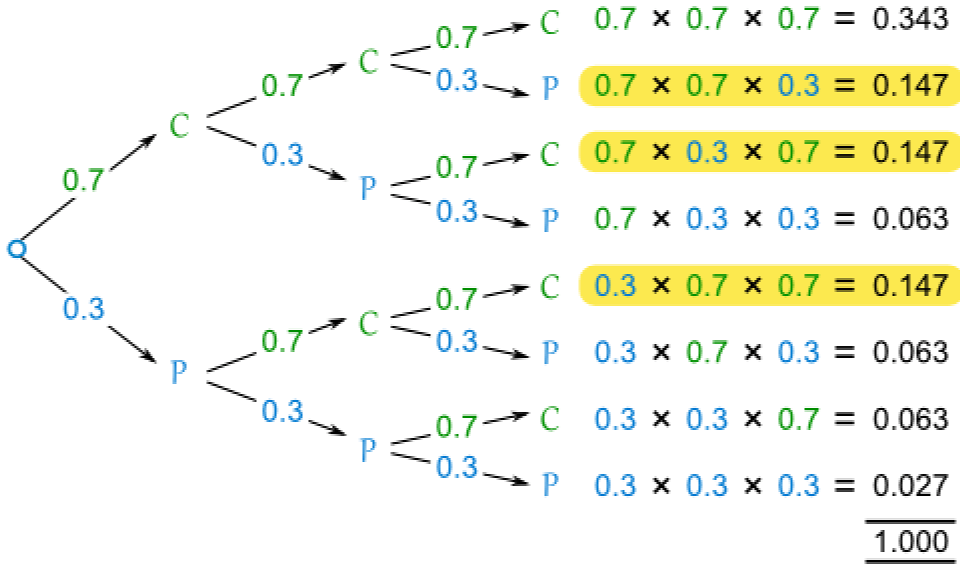
But what if the coins are biased (land more on one side than another) or choices are not 50/50.

Example: You sell sandwiches. 70% of people choose chicken, the rest choose pork.

What is the probability of selling 2 chicken sandwiches to the next 3 customers?

This is just like the heads and tails example, but with 70/30 instead of 50/50.

Let's draw a tree diagram:



The "Two Chicken" cases are highlighted.

Notice that the probabilities for "two chickens" all work out to be **0.147**, because we are multiplying two 0.7s and one 0.3 in each case.

Can we get the **0.147** from a formula? What we want is "two 0.7s and one 0.3"

- 0.7 is the probability of each choice we want, call it **p**
- 2 is the number of choices we want, call it **k**

Probability of "choices we want" (two chickens) is: p^k

And

- The probability of the opposite choice is: **1-p**
- The total number of choices is: **n**
- The number of opposite choices is: **n-k**

Probability of "opposite choices" (one pork) is: $(1-p)^{(n-k)}$

So all choices together is:

$$p^k(1-p)^{(n-k)}$$

Example: (continued)

- p = 0.7 (chance of chicken)
- n = 3
- k = 2

So we get:

$$p^k(1-p)^{(n-k)} = 0.7^2(1-0.7)^{(3-2)} = 0.7^2(0.3)^{(1)} = \mathbf{0.7 \times 0.7 \times 0.3 = 0.147}$$

which is the probability of each outcome.

And the total number of those outcomes is:

$$\frac{n!}{k!(n-k)!} = \frac{3!}{2!(3-2)!} = \frac{3\times2\times1}{2\times1 \times 1} = 3$$

And we get:

Number of outcomes we want		Probability of each outcome		
3	×	0.147	=	0.441

So the probability of event "2 people out of 3 choose chicken" = **0.441**

OK. That was a lot of work for something we knew already, but now we can answer harder questions.

Example: You say "70% choose chicken, so 7 of the next 10 customers should choose chicken" ... what are the chances you are right?

- $p = 0.7$
- $n = 10$
- $k = 7$

So we get:

$$p^k(1-p)^{(n-k)} = 0.7^7(1-0.7)^{(10-7)} = 0.7^7(0.3)^{(3)} = 0.0022235661$$

That is the probability of each outcome.

And the total number of those outcomes is:

$$\frac{n!}{k!(n-k)!} = \frac{10!}{7!(10-7)!}$$

$$= \frac{10\times9\times8\times7\times6\times5\times4\times3\times2\times1}{7\times6\times5\times4\times3\times2\times1 \times 3\times2\times1}$$

$$= \frac{10\times9\times8}{3\times2\times1}$$

$$= \mathbf{120}$$

And we get:

Number of outcomes we want		Probability of each outcome		
120	×	0.0022235661	=	0.266827932

In fact the probability of 7 out of 10 choosing chicken is only about **27%**

Moral of the story: even though the long-run average is 70%, don't expect 7 out of the next 10.

Putting it Together

Now we know how to calculate **how many**:

$$\frac{n!}{k!(n-k)!}$$

And the **probability of each**:

$$p^k(1-p)^{(n-k)}$$

We can multiply them together:

Probability of k out of n ways:

$$P(k \text{ out of } n) = \frac{n!}{k!(n-k)!} p^k(1-p)^{(n-k)}$$

The General Binomial Probability Formula

Important Notes:

- The trials are independent,
- There are only two possible outcomes at each trial,

- The probability of "success" at each trial is constant.

Quincunx

Have a play with the Quincunx (then read Quincunx Explained) to see the Binomial Distribution in action.

Throw the Die

A fair die is thrown four times. Calculate the probabilities of getting:

- 0 Twos
- 1 Two
- 2 Twos
- 3 Twos
- 4 Twos

In this case **n=4**, **p = P(Two) = 1/6**

X is the Random Variable 'Number of Twos from four throws'.

Substitute x = 0 to 4 into the formula:

$$\mathbf{P(k\ out\ of\ n)} = \frac{n!}{k!(n-k)!} p^k(1-p)^{(n-k)}$$

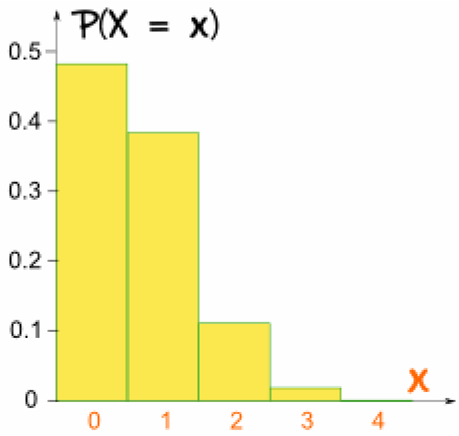
Like this (to 4 decimal places):

- $P(X = 0) = \frac{4!}{0!4!} \times (1/6)^0(5/6)^4 = 1 \times 1 \times (5/6)^4 = 0.4823$
- $P(X = 1) = \frac{4!}{1!3!} \times (1/6)^1(5/6)^3 = 4 \times (1/6) \times (5/6)^3 = 0.3858$
- $P(X = 2) = \frac{4!}{2!2!} \times (1/6)^2(5/6)^2 = 6 \times (1/6)^2 \times (5/6)^2 = 0.1157$

- $P(X = 3) = \frac{4!}{3!1!} \times (1/6)^3(5/6)^1 = 4 \times (1/6)^3 \times (5/6) = 0.0154$
- $P(X = 4) = \frac{4!}{4!0!} \times (1/6)^4(5/6)^0 = 1 \times (1/6)^4 \times 1 = 0.0008$

Summary: "for the 4 throws, there is a 48% chance of no twos, 39% chance of 1 two, 12% chance of 2 twos, 1.5% chance of 3 twos, and a tiny 0.08% chance of all throws being a two (but it still could happen!)"

This time the Bar Graph is not symmetrical:



It is not symmetrical!

It is skewed because **p** is not 0.5

Sports Bikes

Your company makes sports bikes. 90% pass final inspection (and 10% fail and need to be fixed).

What is the expected Mean and Variance of the 4 next inspections?

First, let's calculate all probabilities.

- $n = 4$,
- $p = P(\text{Pass}) = 0.9$

X is the Random Variable "Number of passes from four inspections".

Substitute x = 0 to 4 into the formula:

$$\mathbf{P(k \text{ out of } n)} = \frac{n!}{k!(n-k)!} p^k(1-p)^{(n-k)}$$

Like this:

- $P(X = 0) = \frac{4!}{0!4!} \times 0.9^0 0.1^4 = 1 \times 1 \times 0.0001 = 0.0001$
- $P(X = 1) = \frac{4!}{1!3!} \times 0.9^1 0.1^3 = 4 \times 0.9 \times 0.001 = 0.0036$
- $P(X = 2) = \frac{4!}{2!2!} \times 0.9^2 0.1^2 = 6 \times 0.81 \times 0.01 = 0.0486$
- $P(X = 3) = \frac{4!}{3!1!} \times 0.9^3 0.1^1 = 4 \times 0.729 \times 0.1 = 0.2916$
- $P(X = 4) = \frac{4!}{4!0!} \times 0.9^4 0.1^0 = 1 \times 0.6561 \times 1 = 0.6561$

Summary: "for the 4 next bikes, there is a tiny 0.01% chance of no passes, 0.36% chance of 1 pass, 5% chance of 2 passes, 29% chance of 3 passes, and a whopping 66% chance they all pass the inspection."

Mean, Variance and Standard Deviation

Let's calculate the Mean, Variance and Standard Deviation for the Sports Bike inspections.

There are (relatively) simple formulas for them. They are a little hard to prove, but they do work!

The mean, or "expected value", is:

$$\mu = np$$

For the sports bikes:

$$\mu = 4 \times 0.9 = 3.6$$

So we can expect 3.6 bikes (out of 4) to pass the inspection.
Makes sense really ... 0.9 chance for each bike times 4 bikes equals 3.6

The formula for Variance is:

$$\text{Variance: } \sigma^2 = np(1-p)$$

And Standard Deviation is the square root of variance:

$$\sigma = \sqrt{(np(1-p))}$$

For the sports bikes:

$$\text{Variance: } \sigma^2 = 4 \times 0.9 \times 0.1 = 0.36$$

Standard Deviation is:

$$\sigma = \sqrt{(0.36)} = 0.6$$

Note: we could also calculate them manually, by making a table like this:

X	P(X)	X × P(X)	$X^2 \times P(X)$
0	0.0001	0	0
1	0.0036	0.0036	0.0036
2	0.0486	0.0972	0.1944
3	0.2916	0.8748	2.6244
4	0.6561	2.6244	10.4976
	SUM:	**3.6**	**13.32**

The mean is the **Sum of (X × P(X))**:

$$\mu = 3.6$$

The variance is the **Sum of (X^2 × P(X))** minus **Mean^2**:

$$\text{Variance: } \sigma^2 = 13.32 - 3.6^2 = 0.36$$

Standard Deviation is:

$$\sigma = \sqrt{(0.36)} = 0.6$$

And we got the same results as before (yay!)

Summary

- The General Binomial Probability Formula:

$$P(k \text{ out of } n) = \frac{n!}{k!(n-k)!} p^k(1-p)^{(n-k)}$$

- Mean value of X: $\mu = np$
- Variance of X: $\sigma^2 = np(1-p)$
- Standard Deviation of X: $\sigma = \sqrt{np(1-p)}$

Your turn: Question 1 Question 2 Question 3 Question 4 Question 5 Question 6 Question 7 Question 8 Question 9 Question 10 Question 11 Question 12