Types of Distributions - Uniform, Bernoulli, and Binomial

WE'LL COVER THE FOLLOWING

Types of Distributions

A *few* words before we start:

> 📌 ***We will be looking at the mathematical representations for the various distributions. You do NOT need to remember them by heart! We are going to look at them so that we can get a complete picture.***

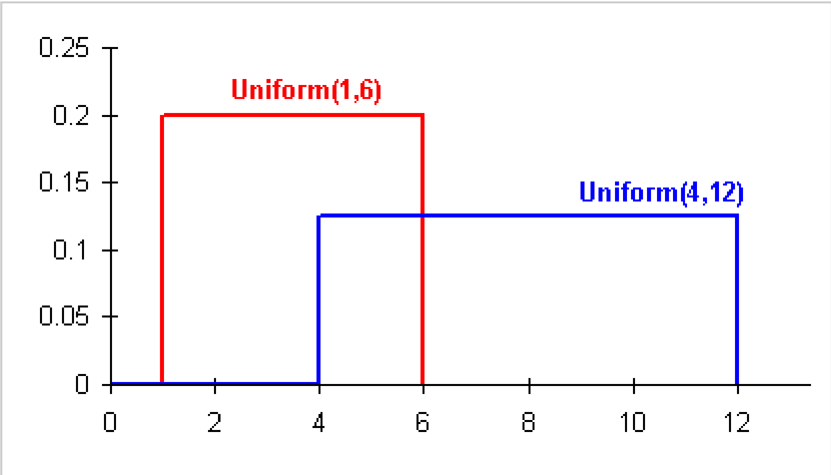
> 📌 ***The important thing is to be able to identify distributions from their graphs and to know their major properties and/or distinguishing features. For example, say you plot the distribution of an interesting variable in your dataset; you should be able to tell what kind of distribution that variable is following — is it Normal or Poisson or something else?***

> 📌 ***Pay special attention to the Normal Distribution and its properties; you should know that one really well as you are likely to encounter it most frequently.***

1. Uniform Distribution

This is a basic probability distribution where all the values have the same probability of occurrence within a specified range; all the values outside that range have probability of 0. For example, when we roll a fair die, the outcomes can only be from 1 to 6 and they all have the same probability, *1/6*. The probability of getting anything outside this range is 0 — you can't get a 7.

The graph of a uniform distribution curve looks like this:



Unifrom Distribution

We can see that the shape of the uniform distribution curve is rectangular. This is the reason why this is often called the rectangular distribution.

The density function, *f(X)*, of a variable *X* that is uniformly distributed can be written as:

$$f(x) = \frac{1}{b-a}$$

where *a* and *b* are the minimum and maximum values of the possible range for *X*.

The mean and variance of the variable *X* can then be calculated like so:

$$Mean = E(X) = \frac{a+b}{2}$$

$$Variance = V(X) = \frac{(b-a)^2}{12}$$

2. Bernoulli Distribution

Although the name sounds complicated, this is an easy one to grasp.

A Bernoulli distribution is a discrete probability distribution. It is used when a random experiment has only two outcomes, "success" or "failure"; "win" or "loss"*, and a **single trial**.

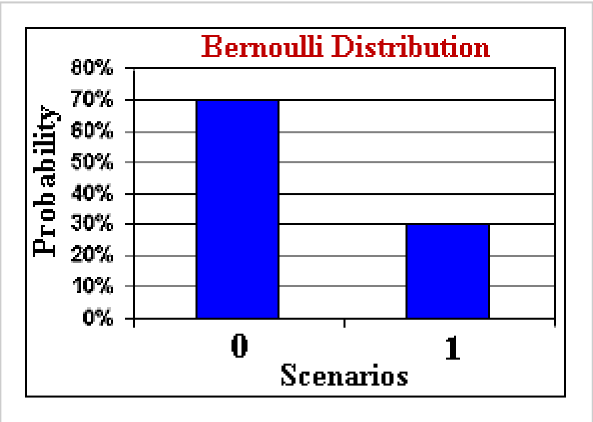
For example, the probability, *P*, of getting *Heads* ("success") while flipping a coin is 0.5. The probability of "failure" is *1 – P*, i.e., 1 minus the probability of success. There's no midway between the two possible outcomes.

A random variable, *X*, with a Bernoulli distribution can take value 1 with the probability of success *p*, and the value 0 with the probability of failure *1-p*.

The probabilities of success and failure do not need to be equally likely, think about the results of a football match. If we are considering a strong team, the chances of winning would be much higher compared to those of a mediocre one. The probability of success, *p*, would be much higher than the probability of failure; the two probabilities wouldn't be the same.

There are many examples of Bernoulli distribution such as whether it's going to rain tomorrow or not (rain in this case would mean success and no rain failure) or passing (success) and not passing (failure) an exam.

Say *p=0.3*, we can graphically represent the Bernoulli distribution like so:



The probability density function, *P(X)*, of a Bernoulli distribution is given by:

$$P(x) = p^x * (1-p)^{(1-x)},$$

where x € (0, 1)

This can also be written as:

$$\Pr(X = x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \end{cases}$$

The expected value, *E(X)*, of a random variable, *X*, having a Bernoulli distribution can be found as follows:

$$Mean = E(X) = 1 * p + 0 * (1 - p) = p$$

The variance of a Bernoulli distribution is calculated as:

$$Variance = V(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1 - p)$$

3. Binomial Distribution

Bernoulli distribution allowed us to represent experiments that have two outcomes but only a single trail. ***What if we have multiple trials?*** Say we toss a coin not one but many times. This is where an extension of Bernoulli distribution comes into play, Binomial Distribution.

A Binomial Distribution can be thought of as the probability of having success or failure as outcome in an experiment that is repeated multiple times. In other words, when only two outcomes are possible, success or failure, win or lose, and the probability of success and failure is same for all the trials, the probability of *Heads* when tossing a coin does not change from one toss to another.

Again, just like in case of Bernoulli, the outcomes don't need be equally likely. Also, each trial is independent — the outcome of a previous toss doesn't affect the outcome of the current toss.

Putting it all together, Binomial distributions must meet the following criteria:

- There are **only two possible outcomes** in a trial- either success or failure.
- The **probability of success** (tails, heads, fail, or pass) is exactly the same for all trials.

- The **number of observations** or trials is fixed, a total number of *n* identical trials. In other words, we can only figure out the probability of something happening if we do it a certain number of times. If we toss a coin once, our probability of getting tails is 50%. If we toss a coin 20 times, our probability of getting tails is very close to 100%.
- Each observation or **trial is independent**, none of the trials have an effect on the probability of the next trial.

Since it's easier to understand concepts by looking at graphical representations rather than just heavy formulas, let's look at the graphical representation of a Binomial Distribution first for varying *n* and constant *p* and then vice-versa:

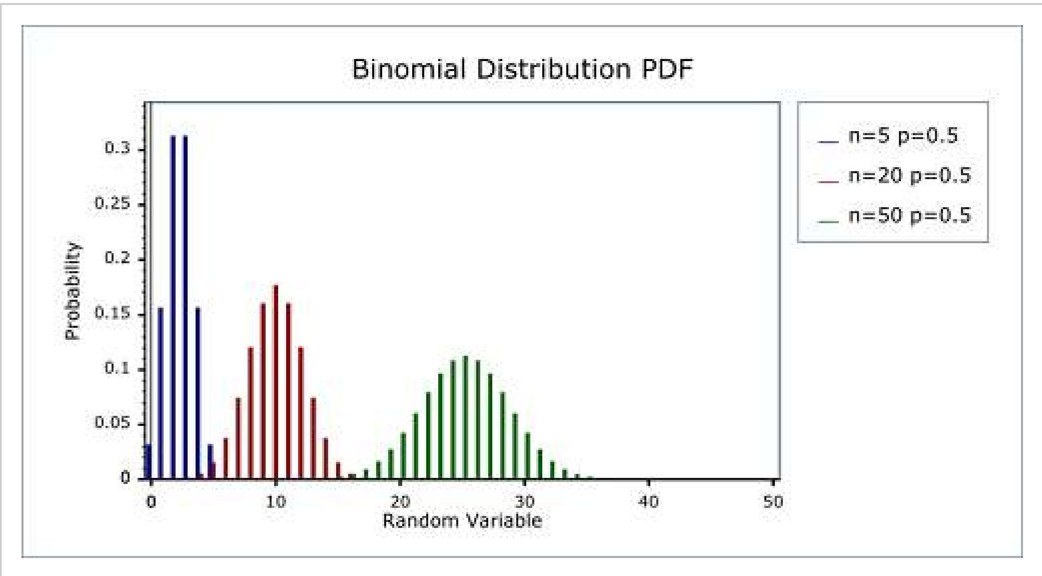


Image Credits: https://www.boost.org

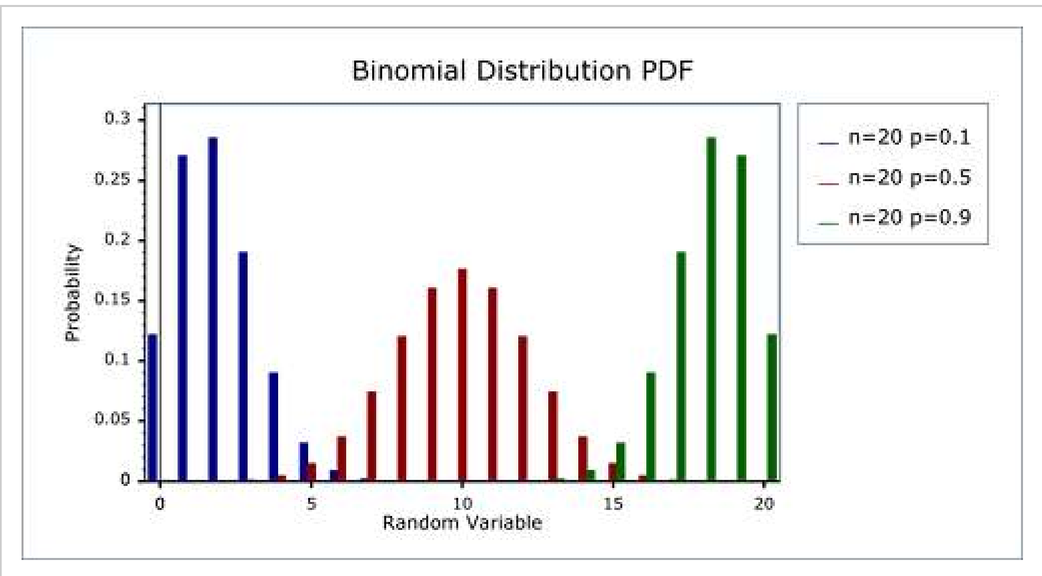
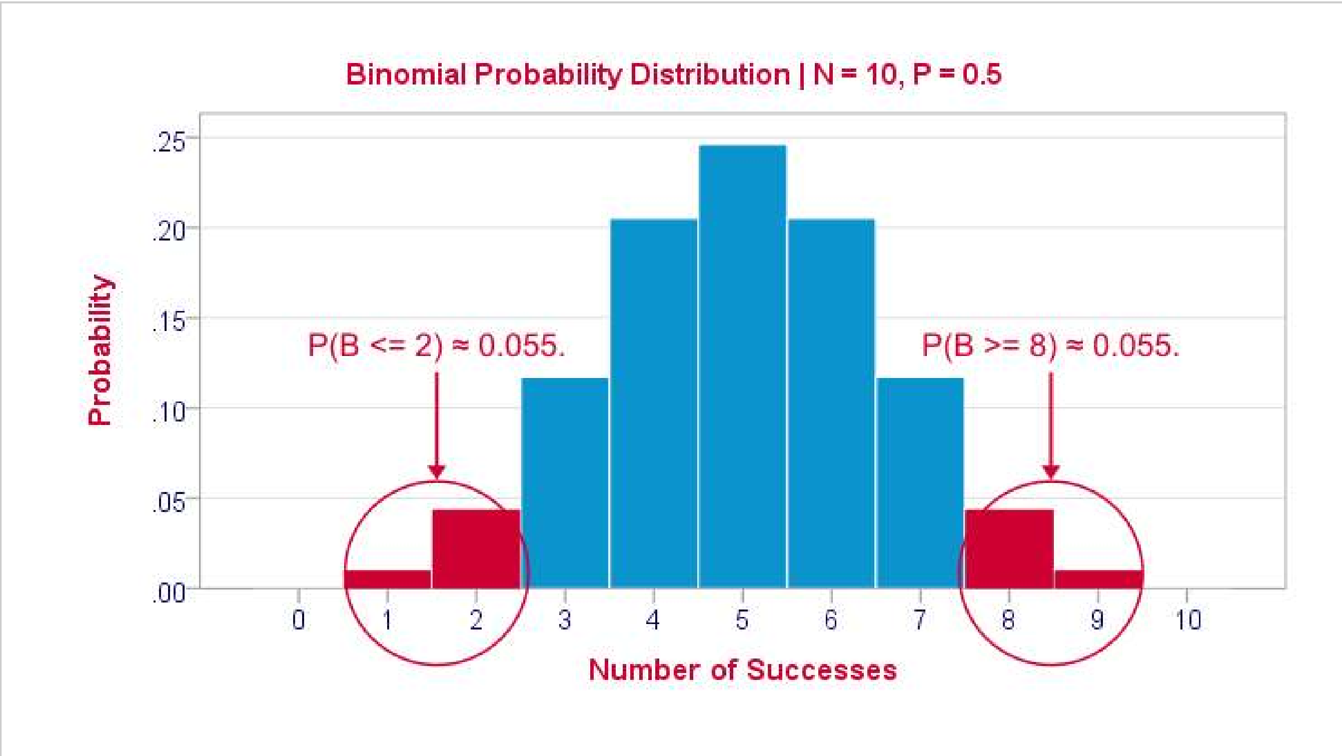


Image Credits: https://www.boost.org

The mean and variance of a binomial distribution are given by:

$$Mean = E(X) = n * p$$
$$Variance = V(X) = n * p * (1 - p)$$

Do you notice something from these formulas? We can observe that the Bernoulli Distribution that we saw earlier was just a special case of Binomial if we set *n=1*.



Example of a binomial distribution chart. Image Credits: https://www.spss-tutorials.com/binomial-test/

We are not done yet! We have 3 more distributions to cover; and we still have to learn about the most important continuous distribution. To the next lesson!