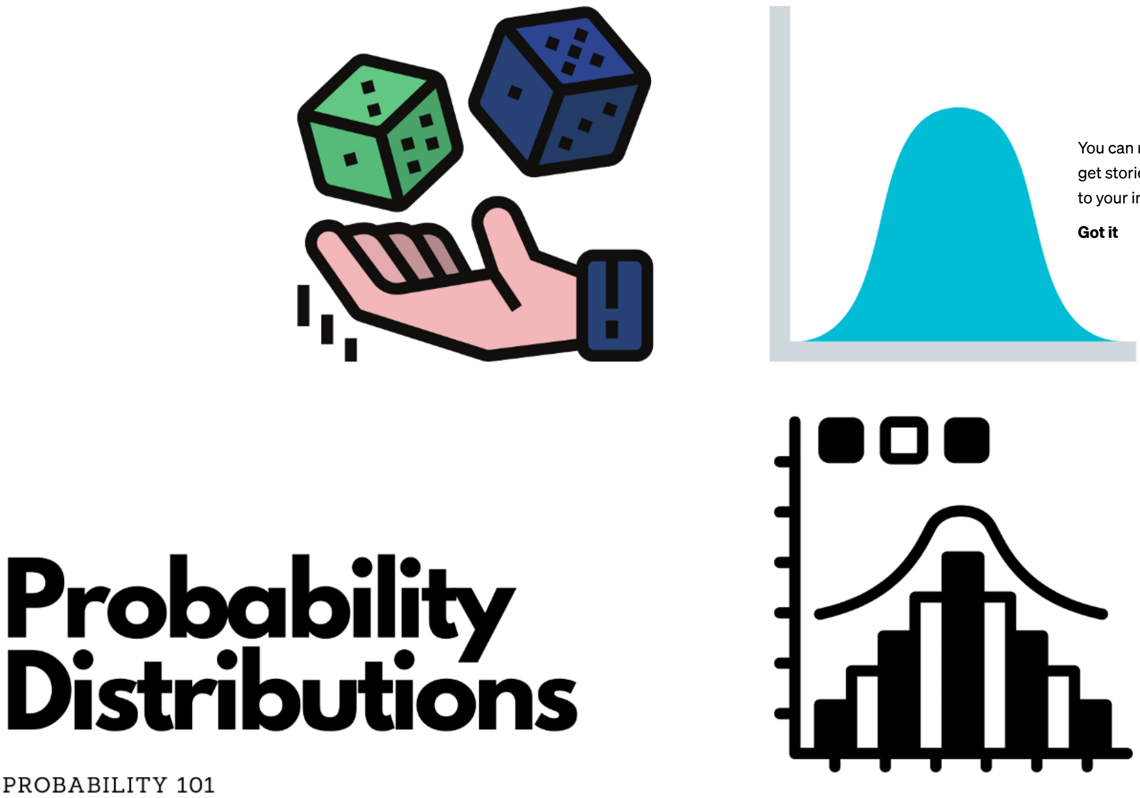
Published in AIGuys · Follow

Chitwan Manchanda · Follow
Feb 7 · 4 min read · Listen

Probability Distributions that every Data Scientist must know



Source

Introduction

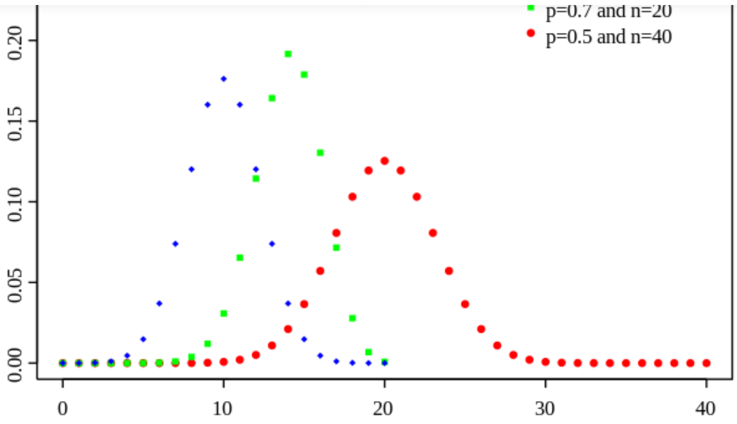
The probability of an event tells us how likely is that, the event will occur. The applications of probability begin with the numbers p0, p1, p2… that give the probability of each possible outcome.

There are dozens of famous and useful possibilities for p. I will discuss four of them in this post. Before going deep into the probability distributions one must be aware of **Random Variables**. A random variable is a variable whose values depend on the outcomes of a random event. There are **two types** of Random Variables,

1. **Discrete Random Variable:** A Random Variable takes a finite or countable number of distinct values.

2. **Continuous Random Variable:** A Random Variable takes an infinite number of values, basically, values that are continuous in nature.

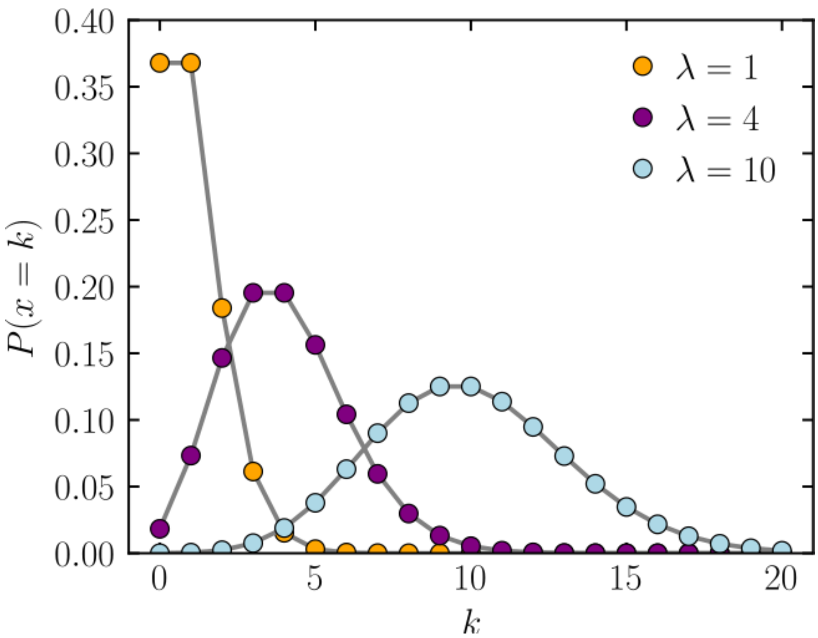
In the next section, we will start discussing the different probability distributions. I have also attached the links for the **PDFs**, **PMFs,** and **the graphs** of these distributions in the **References** section below.



Source

In **Binomial Distribution,** the outcome for each trial is either **0** or **1** (success or failure, heads or tails). The probability of success is given by **p** and the probability of failure is given by **1-p**. If we take an example of a coin toss then, a fair coin has a probability of heads and tails of 1/2 each. There can be n number of trials. **For example**, tossing a fair coin 10 times would probably form a Binomial Distribution.

2. Poisson Distribution



Source

The **Poisson Distribution** is fascinating because there are different ways to approach it. One way is to connect it directly to the binomial distribution (probability p of success in each trial and **pk,n** for **k** successes in n trials. Then Poisson explains this limiting situation of rare events but many trials with λ successes where,

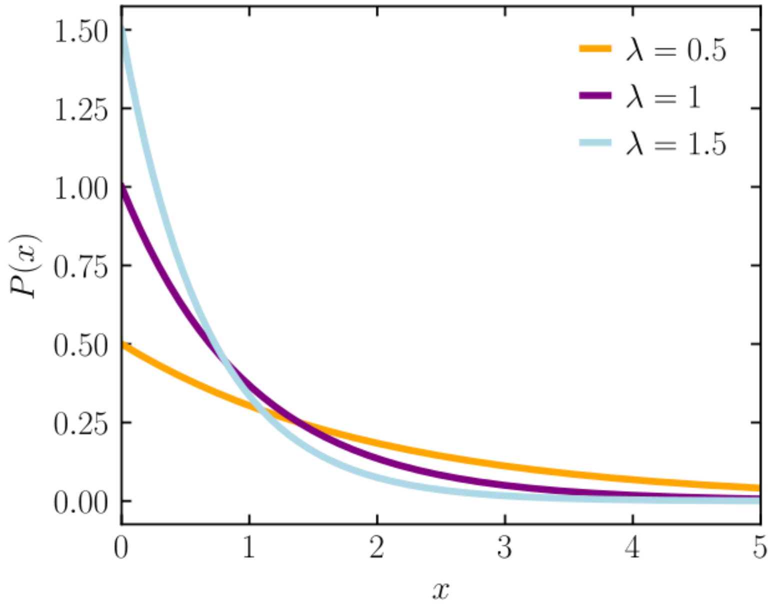
1. **p –> 0** i.e the success probability p is small for each trial (tending to zero).
2. **n –>** ∞ i.e the number of trials n is very large (tending to infinity).
3. **np = λ** i.e the average (expected) number of successes in n trials is λ which is constant.

The **Poisson Distribution** is associated with rare events. Few examples of the events which follow the Poisson Distribution are,

1. The number of big meteors striking the Earth.

One important point to note here is that the Poisson Distribution assumes the independence of events. Some examples above might not fit this assumption. For example, the failure of one bank may be linked to the failure of other banks as well. The assumption of **IID (Independent and Identically distributed)** is not always true.

3. Exponential Distribution



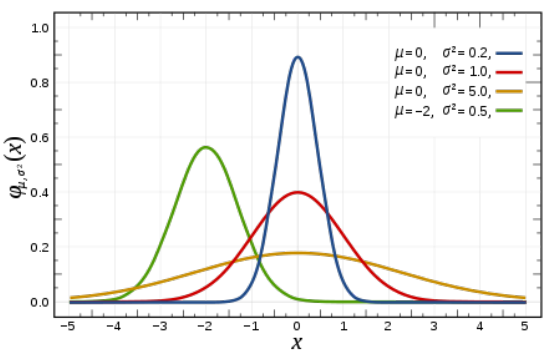
Probability density function

The first continuous distribution in this post is the **Exponential Distribution**. It is the probability distribution of the time between events in a Poisson point process. **For example**, How long until lightning strikes your city? or How long until a big meteor strikes the earth?

> **Note:** The waiting time is independent of the time you have already waited.

The future is independent of the past. If the failure comes from a **slow decay**, this assumption of independence will not hold.

4. Normal Distribution



The distribution which forms the center of probability theory is the **Normal Distribution**. It is also known as the **Gaussian Distribution**. It is the famous bell-shaped curve. The Normal Distribution can be produced by taking random samples from any distribution and creating a distribution from the average of these samples. This is the basis for the **Central Limit Theorem**. An example of an event that follows a Normal Distribution can be the distribution of heights of all people in your city.

References

1. https://www.deeplearningbook.org/contents/prob.html

2. https://en.wikipedia.org/wiki/Probability_distribution

3. https://en.wikipedia.org/wiki/List_of_probability_distributions

4. Binomial Distribution

5. Poisson Distribution

6. Exponential Distribution

7. Normal Distribution