Statistical Distribution

Distributions

The distribution is a mathematical function that describes the relationship of observations of different heights.

> A distribution is simply a collection of data, or scores, on a variable. Usually, these
> scores are arranged in order from smallest to largest, and then they

> can be presented
> graphically.

Once a distribution function is known, it can be used as a shorthand for describing and calculating
related quantities, such as likelihoods of observations, and plotting the relationship between
observations in the domain.

Density Functions

Density functions are functions that describe how the proportion of data or likelihood of the proportion of observations changes over the range of the distribution.

Two types of density functions are probability density functions and cumulative density functions.

- **Probability Density Function**: calculates the probability of observing a given value.
- **Cumulative Density Function**: calculates the probability of an observation equal or less than a value.

A probability density function, or PDF, can be used to calculate the likelihood of a given observation in a distribution. It can also be used to summarize the likelihood of observations
across the distribution's sample space.

A cumulative density function, or CDF, calculates the cumulative likelihood for the observation and all prior observations in the sample space. It allows you to quickly understand and comment on how much of the distribution lies before and after a given value.

A CDF is often plotted as a curve from 0 to 1 for the distribution.

Both PDFs and CDFs are continuous functions. The equivalent of a PDF for a discrete distribution is called a Probability Mass Function, or PMF.

Gaussian Distribution / Normal Distribution

Symmetric Distribution of Values Around the Mean

A Gaussian distribution can be described using two parameters:

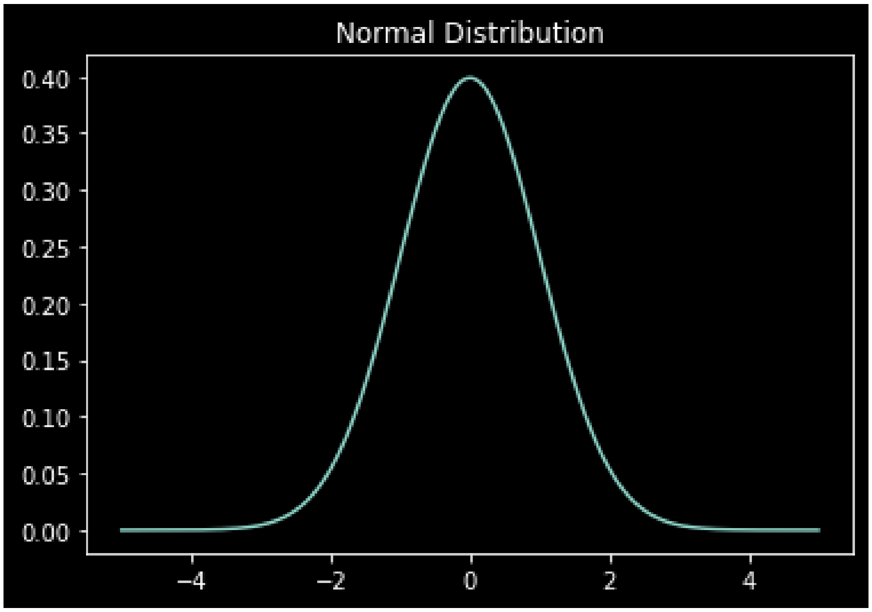
- mean: is the expected value of the distribution.
- variance: describes the spread of observation from the mean

It is common to use a normalized calculation of the variance called the standard deviation

- standard deviation: describes the normalized spread of observations from the mean.

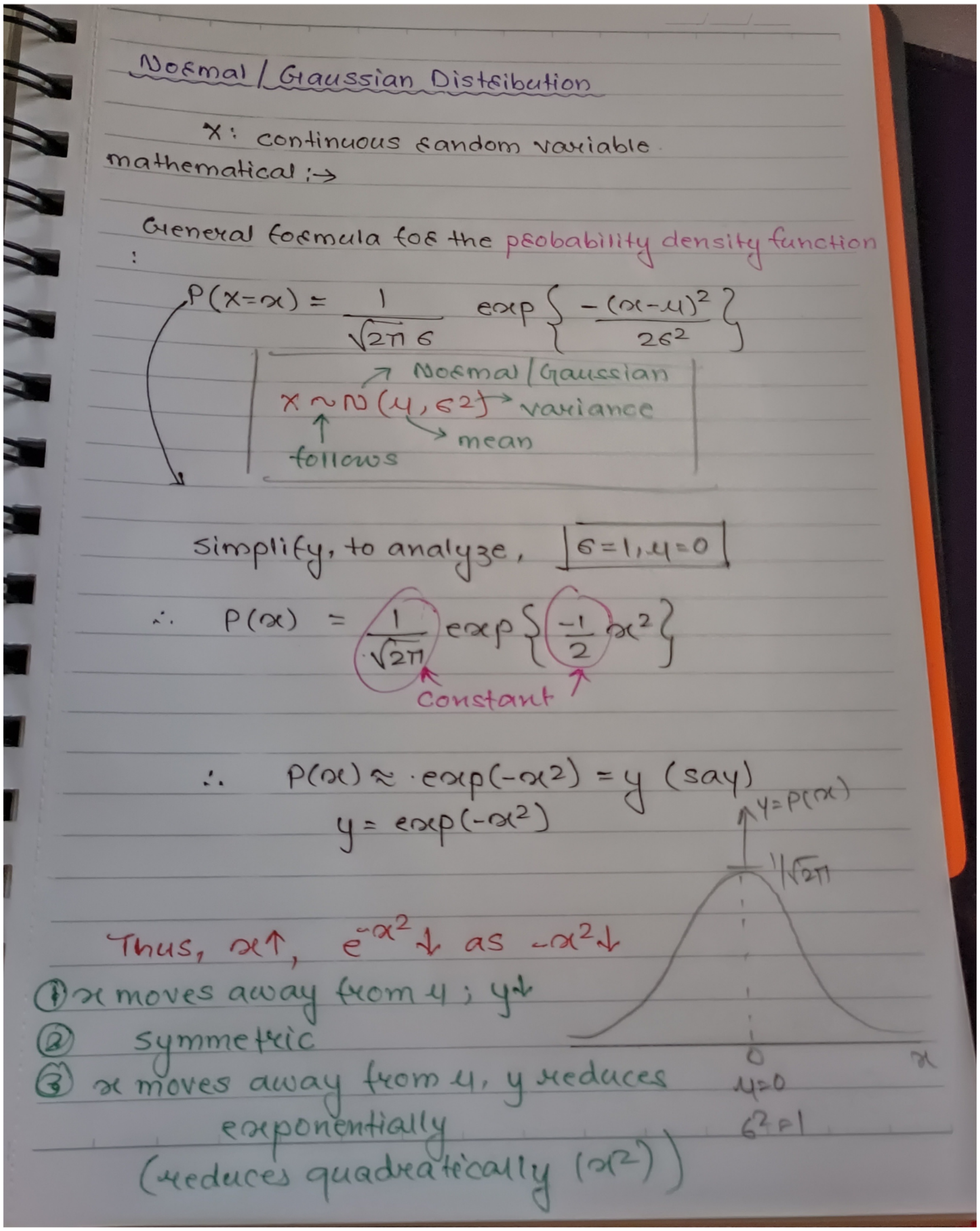
Implementation:

```
# plot the gaussian pdf
from numpy import arange
from matplotlib import pyplot
from scipy.stats import norm

pyplot.style.use('dark_background')
# define the distribution parameters
sample_space = arange(-5, 5, 0.001)
mean = 0.0
stdev = 1.0
# calculate the pdf
pdf = norm.pdf(sample_space, mean, stdev)
# plot
pyplot.plot(sample_space, pdf)
pyplot.show()
```



Density line plot of the Gaussian probability density function.

Mathematical Analysis



Normal / Gaussian Distribution

X : continuous random variable

mathematical :→

General formula for the probability density function :

$$P(X=x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}$$

Normal / Gaussian

$X \sim N(\mu, \sigma^2)$ variance

mean

follows

Simplify, to analyze, $\boxed{\sigma=1, \mu=0}$

$$\therefore \quad P(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{\left(\frac{-1}{2}\right)x^2\right\}$$

constant

$$\therefore \quad P(x) \approx \exp(-x^2) = y \text{ (say)}$$

$$y = \exp(-x^2)$$

Thus, $x\uparrow$, $e^{-x^2}\downarrow$ as $-x^2\downarrow$

① x moves away from μ ; $y\downarrow$

② symmetric

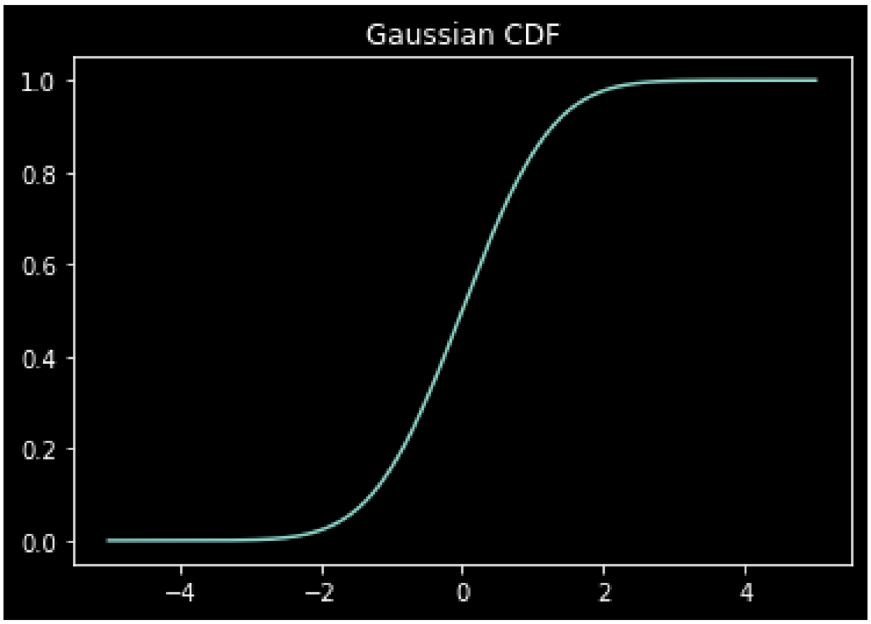
③ x moves away from μ, y reduces exponentially

(reduces quadratically (x^2))

Mathematical Analysis

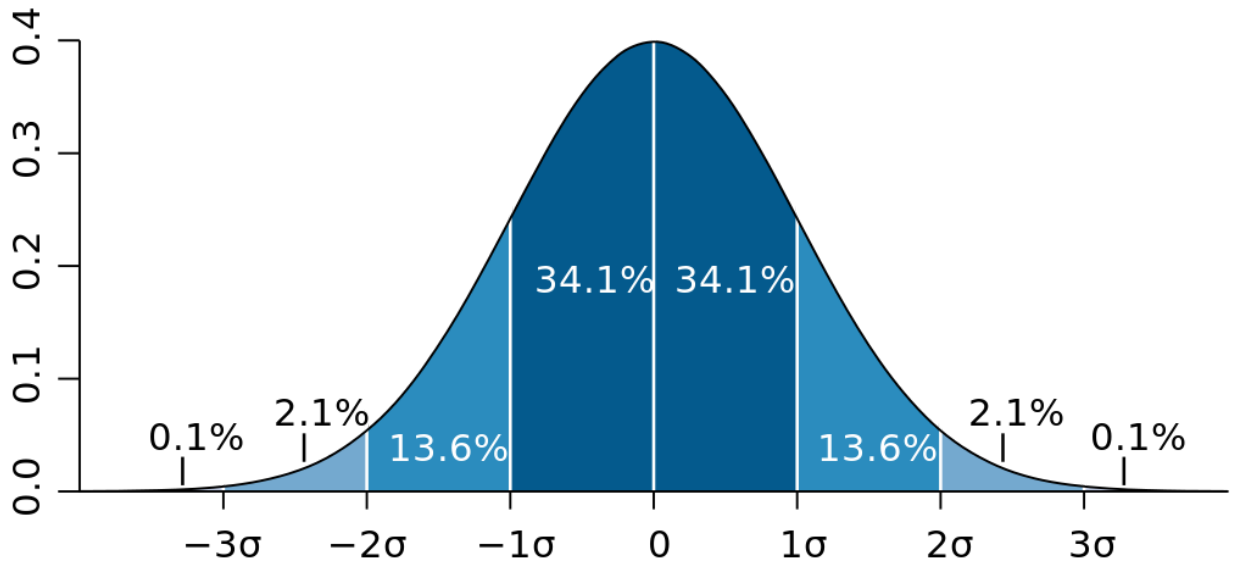
The example below creates a Gaussian CDF for the same sample space.

```
# plot the gaussian cdf
from numpy import arange
from matplotlib import pyplot
from scipy.stats import norm

pyplot.style.use('dark_background')
# define the distribution parameters
sample_space = arange(-5, 5, 0.001)
# calculate the cdf
cdf = norm.cdf(sample_space)
# plot
pyplot.plot(sample_space, cdf)
pyplot.show()
```



Density line plot of the Gaussian cumulative density function.

68-95-99.7 Rule



Probability Density and 68-95-99.7 Rule

While plotting a graph for a normal distribution, 68% of all values lie within one standard deviation from the mean. In the example above, if the mean is 70 and the standard deviation is 10, 68% of the values will lie between 60 and 80. Similarly, 95% of the values lie within two standard deviations from the mean, and 99.7% lie within three standard deviations from the mean. This last interval captures almost all matters. If a data point is not included, it is most likely an outlier.

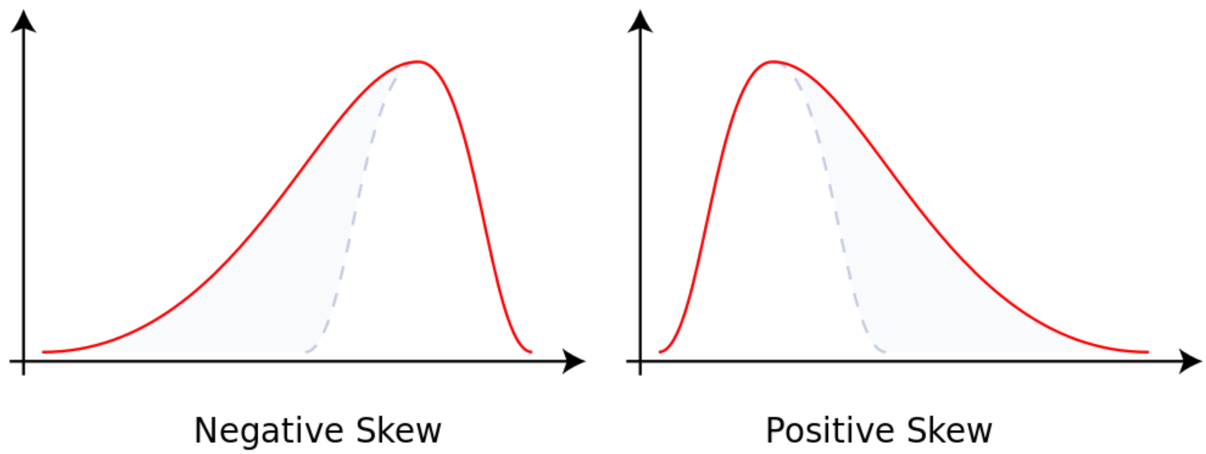
Skewness

A fundamental task in many statistical analyses is to characterize the *location* and *variability* of a data set. Further characterization of the data includes skewness and kurtosis.

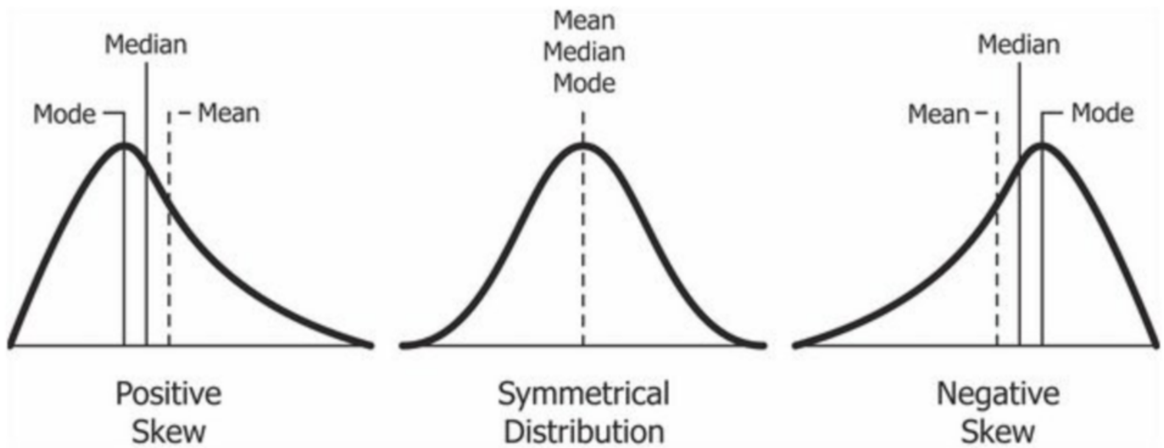
The skewness measures the extent to which the data values are not symmetric around the mean. For normal distribution the portion of the curve below the mean will be a mirror image of the portion of the curve above the mean thus skewness will be zero.

Two kinds of skewness a distribution has:

1. **Negative Skew**: the mass of the distribution is concentrated on the right of the figure. The distribution is said to be *left-skewed*, *left-tailed*, or *skewed to the left.*
2. **Positive Skew:** the mass of the distribution is concentrated on the left of the figure. The distribution is said to be *right-skewed*, *right-tailed*, or *skewed to the right.*



Distribution Curve



A general relationship of mean and median under differently skewed unimodal distribution.

The probability distribution with its tail on the right side is a positively skewed distribution and the one with its tail on the left side is a negatively skewed distribution.

What is Skewness in Statistics? | Statistics for Data Science

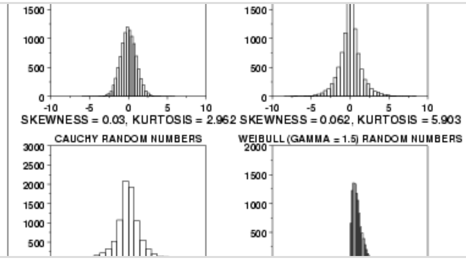
Skewness is a key statistics concept you must know in the data science and analytics fields Learn what is skewness, and why it's important for you as a data science professional The concept of

https://www.analyticsvidhya.com/blog/2020/07/what-is-skewness-statistics/

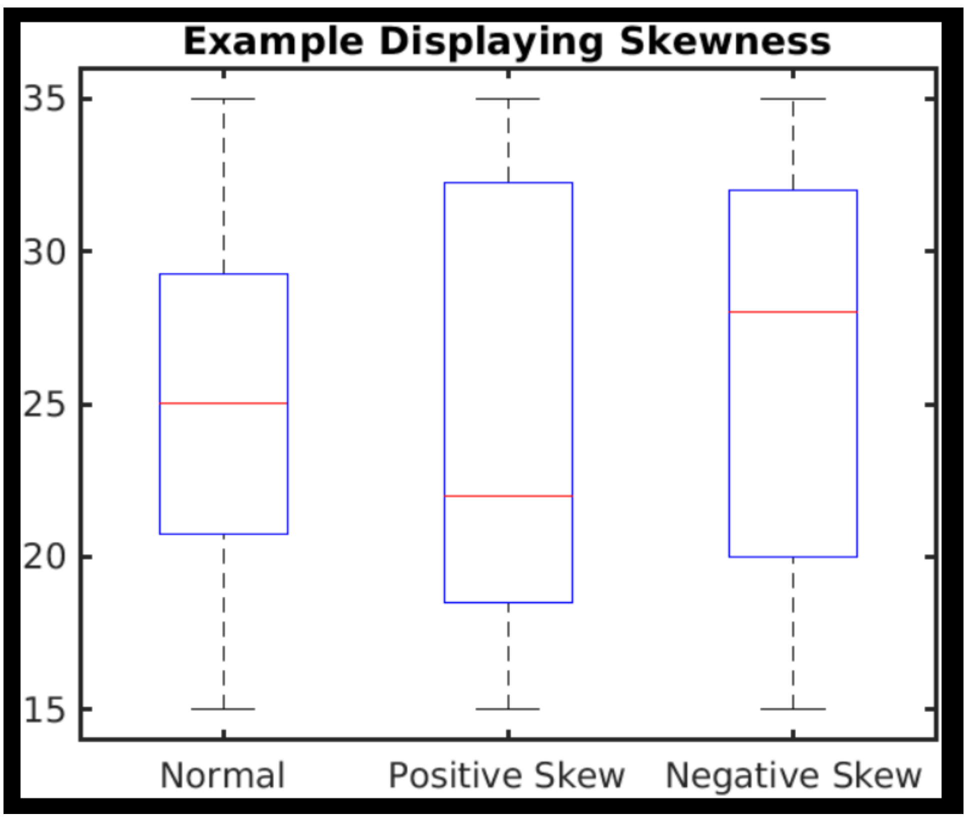
1.3.5.11. Measures of Skewness and Kurtosis

1. Exploratory Data Analysis 1.3. EDA Techniques 1.3.5. Quantitative Techniques Skewness and Kurtosis A fundamental task in many statistical analyses is to characterize the location

https://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm



Box Plot Distribution



- **Symmetric**:- When the median of the distribution is in the middle of the box, and the whiskers are about the same on both sides of the box.

- **Positively or Right skewed** - When the median of the distribution is closer to the bottom of the box, and if the whisker is shorter on the lower end of the box.
- **Negatively or skewed** : When the median of the distribution is closer to the top of the box, and if the whisker is shorter on the upper end of the box.

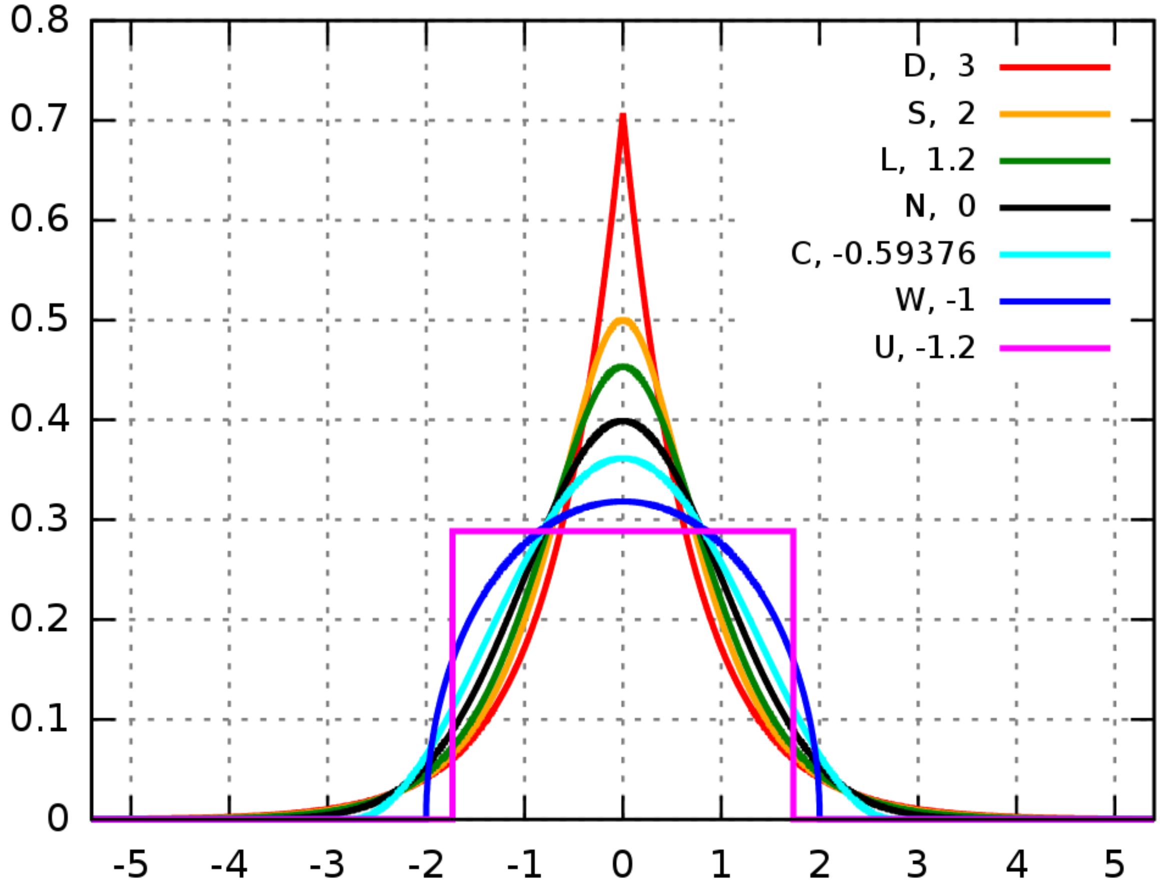
Kurtosis

It measures the peakedness of the curve of the distribution or how sharply the curve raises the approaching center of the distribution. It compares the shape of the peak to the shape of the peak of normal distribution.

Excess kurtosis is defined as kurtosis minus 3. There are 3 distinct regimes as described below.

- **Mesokurtic**:- The kurtosis of any univariate normal distribution is 3 and is called mesokurtic.
- **Leptokurtic**:- The distribution will have a sharper-rising center of the peak, with a higher concentration of values near the mean with a fatter tail that is many values in the tail compared to a normal distribution and kurtosis greater than 3 and a positive excess kurtosis.
- **Platykurtic**:- The distribution will have less center of peak compared to a normal distribution and have a negative excess kurtosis.

https://www.scribbr.com/statistics/kurtosis/#:~:text=Kurtosis is a measure of,(thin tails)



Probability density functions for selected distributions with mean 0, variance 1 and different excess kurtosis

- **D** : Laplace distribution, also known as the double exponential distribution, red curve (two straight lines in the log-scale plot), excess kurtosis = 3
- **S** : hyperbolic secant distribution, orange curve, excess kurtosis = 2
- **L** : logistic distribution, green curve, excess kurtosis = 1.2
- **N** : normal distribution, black curve (inverted parabola in the log-scale plot), excess kurtosis = 0
- **C** : raised cosine distribution, cyan curve, excess kurtosis = −0.593762...
- **W** : Wigner semicircle distribution, blue curve, excess kurtosis = −1
- **U** : uniform distribution, magenta curve (shown for clarity as a rectangle in both images), excess kurtosis = −1.2.

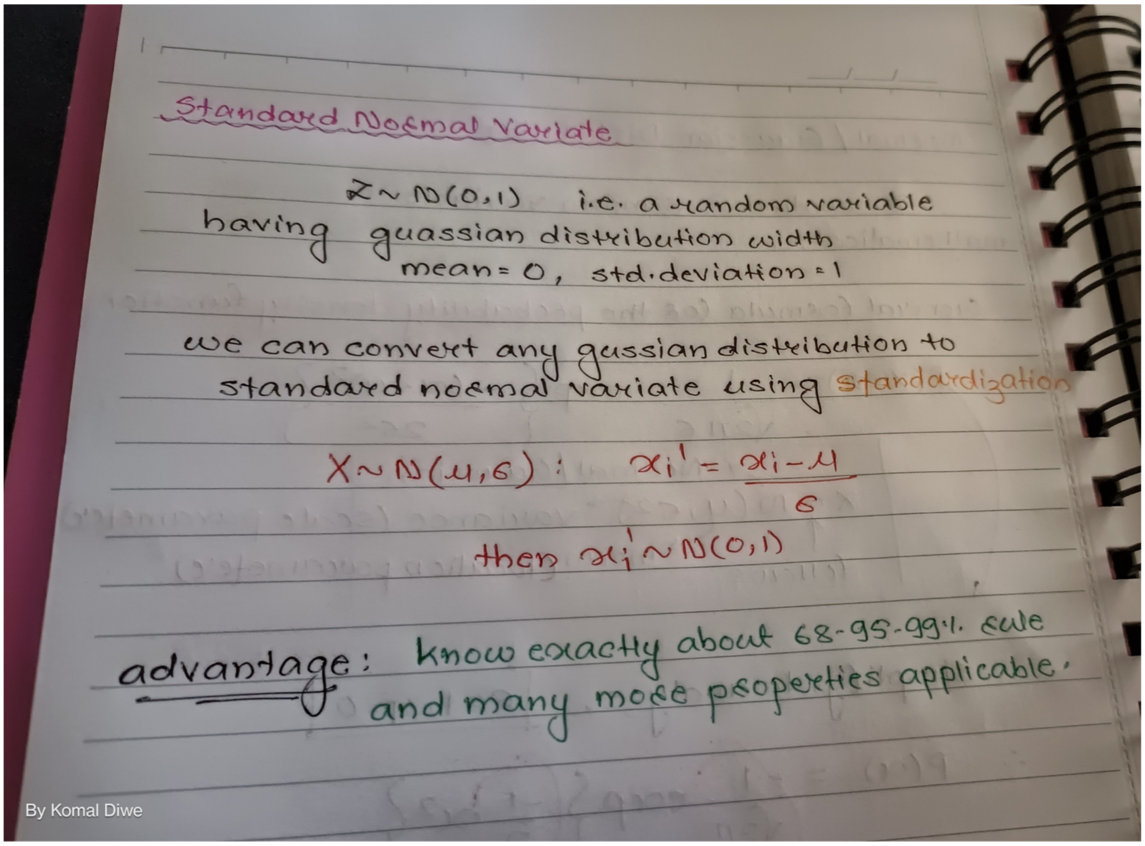
How to handle Skewed Distribution

check this notebook..

Google Colaboratory

https://colab.research.google.com/drive/1HXZR2Vz-Vh29tGj3RIVbNj1r4Of4Jp4P?usp=sharing

Standard Normal Variate



Standard Normal Variate

$Z \sim N(0,1)$ i.e. a random variable having guassian distribution width mean = 0, std.deviation = 1

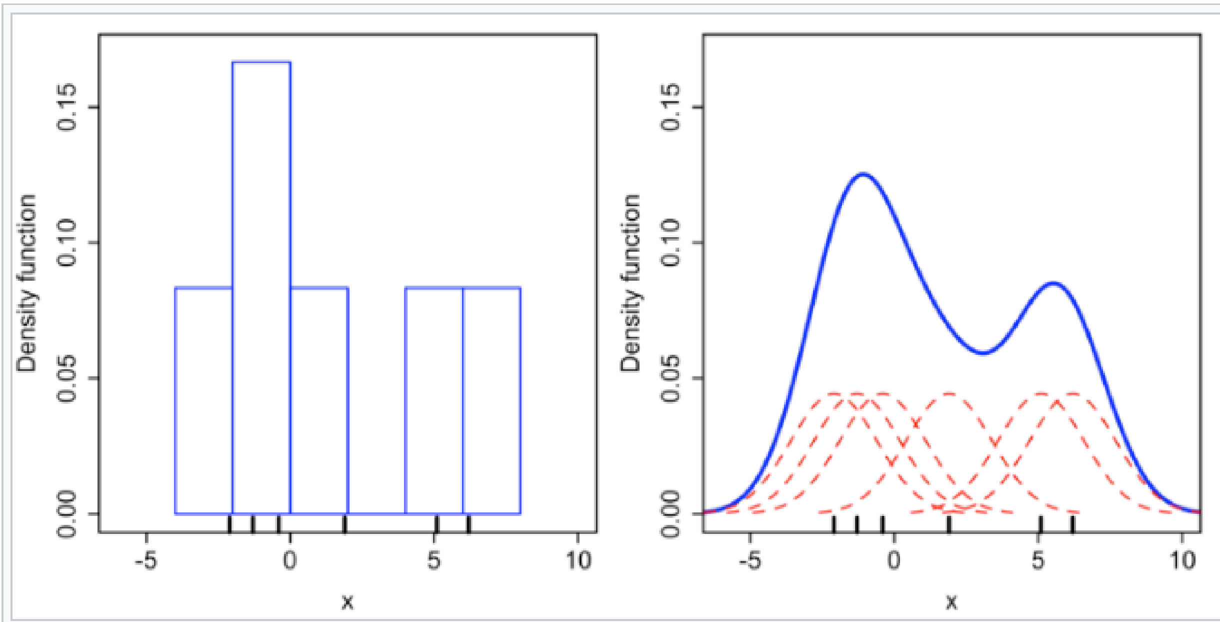
we can convert any gussian distribution to standard normal variate using standardization

$X \sim N(\mu,\sigma)$: $x_i' = \frac{x_i - \mu}{\sigma}$

then $x_i' \sim N(0,1)$

advantage: know exactly about 68-95-99% rule and many more properties applicable.

Kernel Density Estimation



Comparison of the histogram (left) and kernel density estimate (right) constructed using the same data. The six individual kernels are the red dashed curves, and the kernel density estimates the blue curves.

In-Depth: Kernel Density Estimation

In the previous section we covered Gaussian mixture models (GMM), which are a kind of hybrid between a clustering estimator and a density estimator. Recall that a density estimator is an algorithm which takes a D-dimensional dataset and produces an estimate of the D-dimensional probability distribution which that data is

https://jakevdp.github.io/PythonDataScienceHandbook/05.13-kernel-density-estimation.html

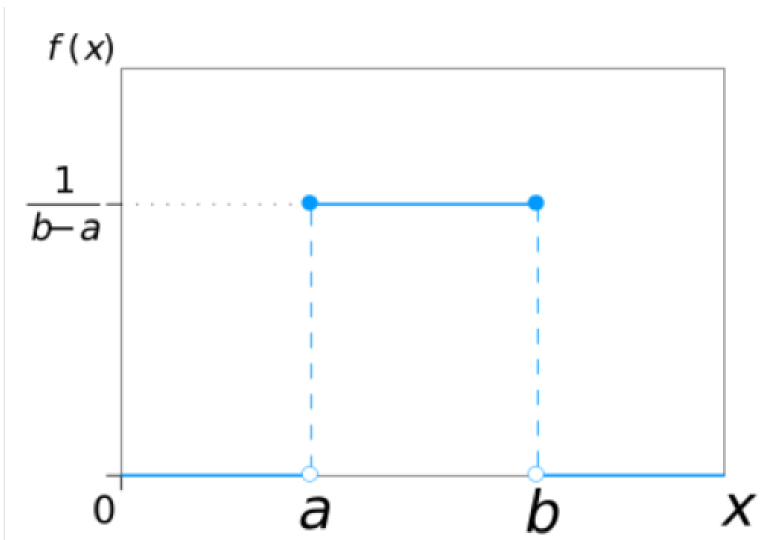
Uniform Distribution

All outcomes are equally likely.

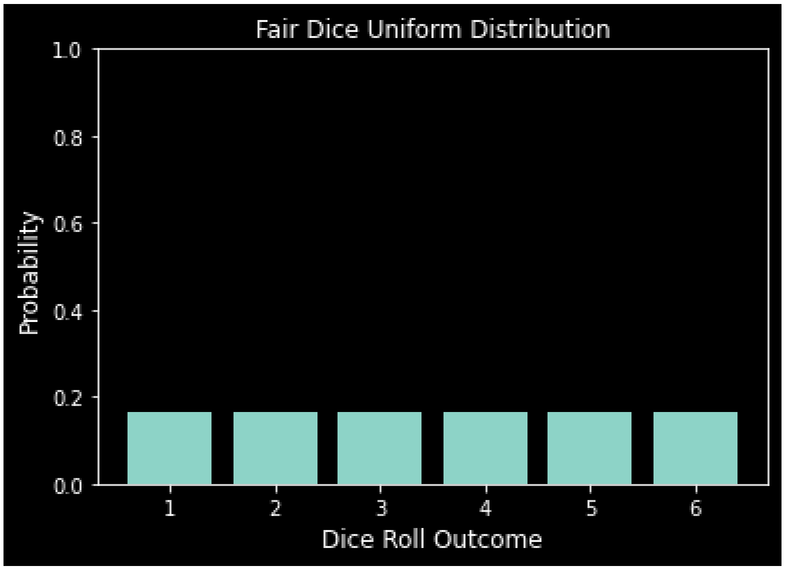
This distribution is a subcategory of a continuous distribution, this distribution represents a similar probability for all the events to occur. The probability density function for uniform distribution is:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b, \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$

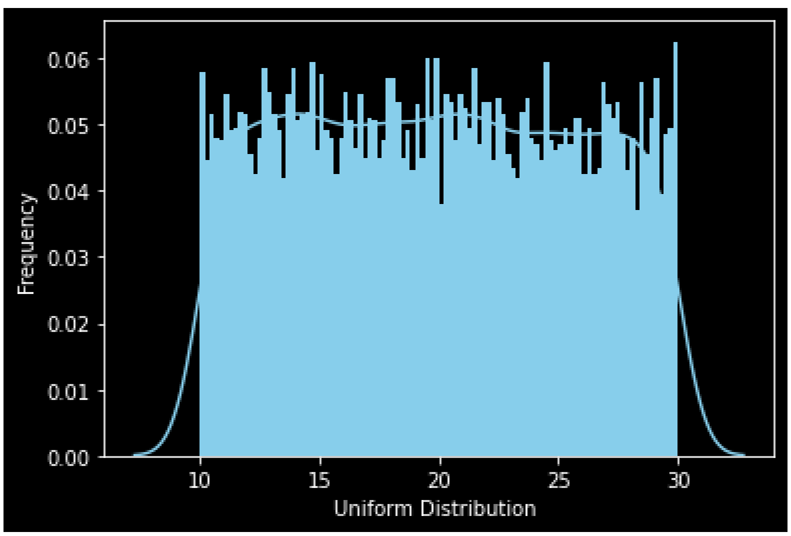
Since any interval of numbers of equal width has an equal probability of being observed, the curve describing the distribution is a rectangle, with constant height across the interval and 0 height elsewhere. Since the area under the curve must be equal to 1, the length of the interval determines the height of the curve. The following figure shows a uniform distribution in intervals (a,b). Notice that the area needs to be 1. The height is set to $\frac{1}{b-a}$.




```
probs = np.full((6), 1/6)
face = [1,2,3,4,5,6]
plt.bar(face, probs)
plt.ylabel('Probability', fontsize=12)
plt.xlabel('Dice Roll Outcome', fontsize=12)
plt.title('Fair Dice Uniform Distribution', fontsize=12)
axes = plt.gca()
axes.set_ylim([0,1])
```




```
# import uniform distribution
from scipy.stats import uniform
# random numbers from uniform distribution
n = 10000
start = 10
width = 20
data_uniform = uniform.rvs(size=n, loc = start, scale=width)
ax = sns.distplot(data_uniform,
                  bins=100,
                  kde=True,
                  color='skyblue',
                  hist_kws={"linewidth": 15,'alpha':1})
ax.set(xlabel='Uniform Distribution ', ylabel='Frequency')
```

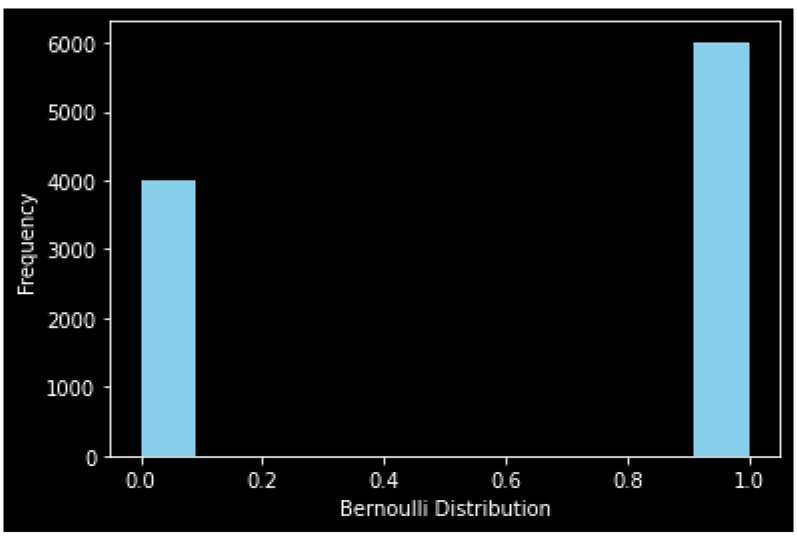


Bernoulli Distribution

A Bernoulli distribution has only two possible outcomes, namely 1 (success) and 0 (failure), and a single trial, for example, a coin toss. So the random variable X which has a Bernoulli distribution can take the value 1 with the probability of success, p, and the value 0 with the probability of failure, q or (1-p). The probabilities of success and failure need not be equally likely. The Bernoulli distribution is a special case of the binomial distribution where a single trial is conducted (n=1). Its probability mass function is given by:

$$f(k;p) = p^k(1-p)^{1-k} \; fork \in \{0,1\}$$

```
from scipy.stats import bernoulli
data_bern = bernoulli.rvs(size=10000,p=0.6)
ax= sns.distplot(data_bern,
                 kde=False,
                 color="skyblue",
                 hist_kws={"linewidth": 15,'alpha':1})
ax.set(xlabel='Bernoulli Distribution', ylabel='Frequency')
```



Binomial Distribution

The Binomial Distribution can be thought of as the sum of outcomes of an event following a Bernoulli distribution.

A distribution where only two outcomes are possible, such as success or failure, gain or loss, win or lose and where the probability of success and failure is the same for all the trials is called a Binomial Distribution. However, The outcomes need not be equally likely, and each trial is independent of the others. The parameters of a binomial distribution are n and p where n is the total number of trials, and p is the probability of success in each trial. Its probability distribution function is given by :

$$f(k,n,p) = Pr(k;n,p) = Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$
$$where \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Binomial vs Bernoulli distribution.

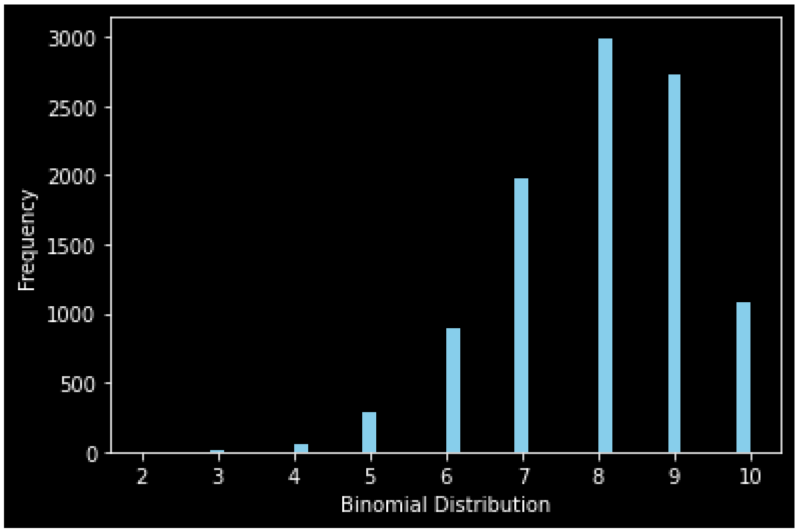
The difference between these distributions can be explained through an example. Consider you’re attempting a quiz that contains 10 True/False questions. Trying a single T/F question would be considered a Bernoulli trial, whereas attempting the entire quiz of 10 T/F questions would be categorized as a Binomial trial. The main characteristics of Binomial Distribution are:

- Given multiple trials, each of them is independent of the other. That is, the outcome of one trial doesn’t affect another one.
- Each trial can lead to just two possible results (e.g., winning or losing), with probabilities p and (1 – p).

A binomial distribution is represented by B (n, p), where n is the number of trials and p is the probability of success in a single trial. A Bernoulli distribution can be shaped as a binomial trial as B (1, p) since it has only one trial. The expected value of a binomial trial “x” is the number of times a success occurs, represented as E(x) = np. Similarly, variance is represented as Var(x) = np(1-p).

```
from scipy.stats import binom
data_binom = binom.rvs(n=10,p=0.8,size=10000)
ax = sns.distplot(data_binom,
```

```
                    kde=False,
                    color='skyblue',
                    hist_kws={"linewidth": 15,'alpha':1})
ax.set(xlabel='Binomial Distribution', ylabel='Frequency')
```



Poisson Distribution

This distribution is named after the mathematician Siméon Denis Poisson. We mainly use this distribution when the variable of interest in data is discrete.

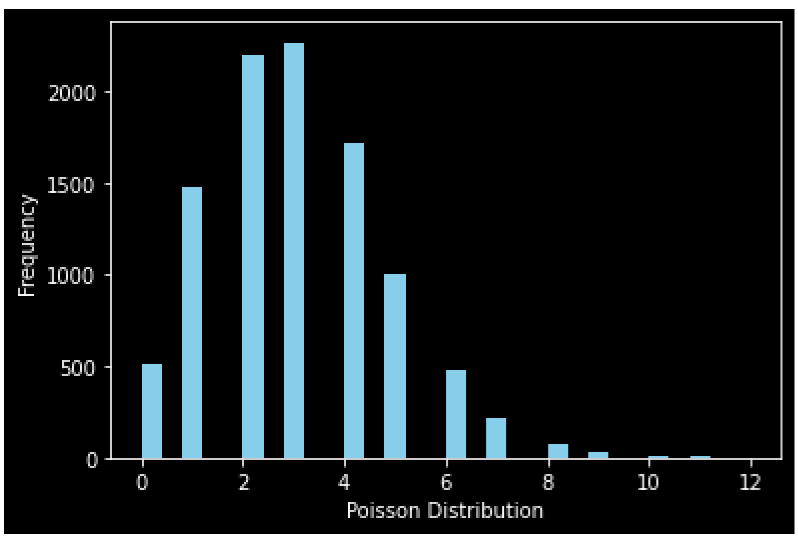
Poisson random variable is typically used to model the number of times an event happened in a time interval. For example, the number of users visited on a website in an interval can be thought of a Poisson process. Poisson distribution is described in terms of the rate (μ) at which the events happen. An event can occur 0, 1, 2, … times in an interval. The average number of events in an interval is designated λ (lambda). Lambda is the event rate, also called the rate parameter. The probability of observing k events in an interval is given by the equation:

$$P(k\ events\ in\ interval) = e^{-\lambda}\frac{\lambda^k}{k!}$$
$$for\ k >= 0$$

note: e to the power -lambda

Note that the normal distribution is a limiting case of Poisson distribution with the parameter $\lambda \to \infty$. Also, if the times between random events follow an exponential distribution with rate λ, then the total number of events in a time period of length t follows the Poisson distribution with parameter λt.

```
from scipy.stats import poisson
data_poisson = poisson.rvs(mu=3, size=10000)
ax = sns.distplot(data_poisson,
                  bins=30,
                  kde=False,
                  color='skyblue',
                  hist_kws={"linewidth": 15,'alpha':1})
ax.set(xlabel='Poisson Distribution', ylabel='Frequency')
```



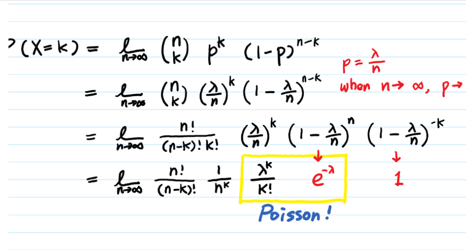
The main characteristics which describe the Poisson Processes are:

- The events are independent of each other.
- An event can occur any number of times (within the defined period).
- Two events can’t take place simultaneously.

Poisson Distribution Intuition (and derivation)

Before setting the parameter λ and plugging it into the formula, let's pause a second and ask a question. Why did Poisson have to invent the Poisson Distribution? To predict the # of events occurring

tds https://towardsdatascience.com/poisson-distribution-intuition-and-derivation-1059aeab90d



Exponential Distribution

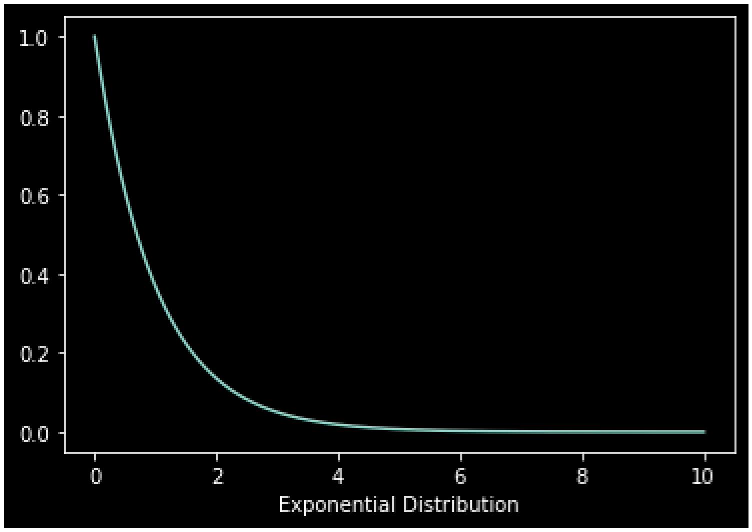
The exponential distribution describes the time between events in a Poisson point process, i.e., a process in which events occur continuously and independently at a constant average rate. It has a parameter λ called rate parameter, and its equation is described as :

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{x} >= 0, \\ 0 & \text{x} < 0. \end{cases}$$

For example, in physics, it is often used to measure radioactive decay; in engineering, to measure the time associated with receiving a defective part on an assembly line; and in finance, to measure the likelihood of the next default for a portfolio of financial assets. Another common application of Exponential distributions in survival analysis (e.g., expected life of a device/machine).

A decreasing exponential distribution looks like :

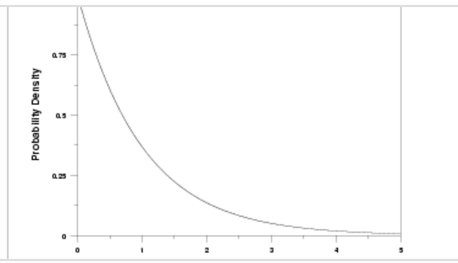
```
from scipy.stats import expon
import matplotlib.pyplot as plt

x = np.arange(0, 10, 0.001)
plt.plot(x, expon.pdf(x))
```



1.3.6.6.7. Exponential Distribution

The general formula for the probability density function of the exponential distribution is\(f(x) = \frac{1} {\beta} e^{-(x - \mu)/\beta} \hspace{.3in} x \ge \mu; \beta > 0 \) where μ is the location

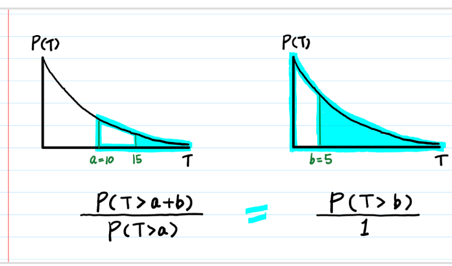
https://www.itl.nist.gov/div898/handbook/eda/section3/eda3667.htm



What is Exponential Distribution

We always start with the "why" instead of going straight to the formulas. If you understand the why, it actually sticks with you and you'll be a lot more likely to apply it in your own line of work. To

tds https://towardsdatascience.com/what-is-exponential-distribution-7bdd08590e2a



Gamma Distribution

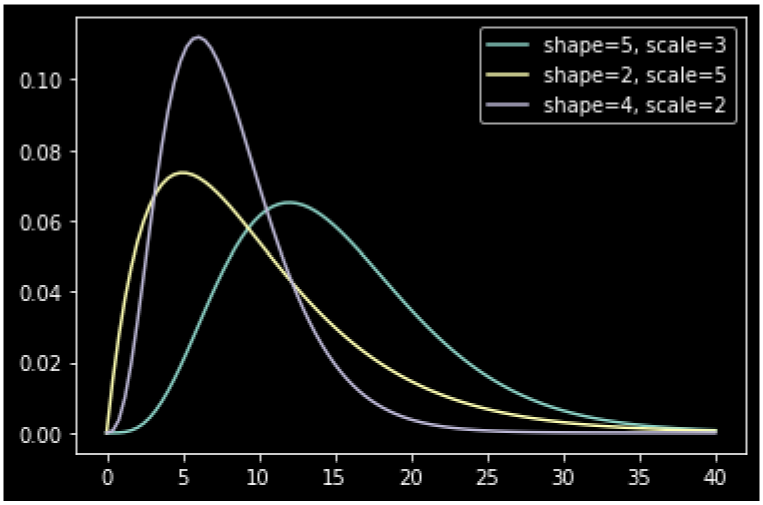
The gamma distribution is a two-parameter family of continuous probability distributions. While it is used rarely in its raw form but other popularly used distributions like exponential, chi-squared, and erlang distributions are special cases of the gamma distribution. The gamma distribution can be parameterized in terms of a shape parameter α = k and an inverse scale parameter β = 1/θ, called a rate parameter., the symbol Γ(n) is the gamma function and is defined as (n-1)! :

The corresponding probability density function in the shape-rate parameterization is

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}e^{-\beta x}\beta^{\alpha}}{\Gamma(\alpha)} \quad \text{for } x > 0 \quad \alpha,\beta > 0,$$

where $\Gamma(\alpha)$ is the gamma function. For all positive integers, $\Gamma(\alpha) = (\alpha - 1)!$.

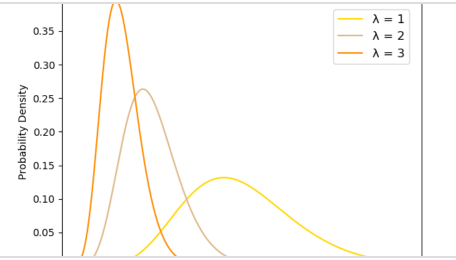
```
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt

#define three Gamma distributions
x = np.linspace(0, 40, 100)
y1 = stats.gamma.pdf(x, a=5, scale=3)
y2 = stats.gamma.pdf(x, a=2, scale=5)
y3 = stats.gamma.pdf(x, a=4, scale=2)

#add lines for each distribution
plt.plot(x, y1, label='shape=5, scale=3')
plt.plot(x, y2, label='shape=2, scale=5')
plt.plot(x, y3, label='shape=4, scale=2')

#add legend
plt.legend()

#display plot
plt.show()
```



Gamma Distribution - Intuition, Derivation, and Examples

and why does it matter? Before setting Gamma's two parameters α, β and plugging them into the formula, let's pause for a moment and ask a few questions... Why did we have to invent the Gamma

https://towardsdatascience.com/gamma-distribution-intuition-derivation-and-examples-55f407423840



Weibull Distribution

As mentioned before, the Poisson and Exponential distributions are only useful as long as the number of occurrences of an instance is relatively consistent over time, which isn't the case when talking about things like mechanical failure. Think about it, why is engine mileage such an important metric when determining the price of used cars? It's because parts of a car a more likely to fail as more stress is put on them. To simulate things like this, we use a Weibull distribution. It's an extension of the Exponential distribution where the even rate can change through a parameter called the shape parameter β. If $\beta > 1$ then the probability of an event happening increases over time while if $\beta < 1$ then the probability decreases over time. There is also the scale parameter η which is affected by the unit of age whether that be hours, miles, cycles etc.

What is the scale parameter?

In Weibull analysis, what exactly is the scale parameter, η (Eta)? And why, at t = η , will 63.21% of the population have failed, regardless of the value of the shape parameter, β (Beta)? η (Eta) is

https://www.livingreliability.com/en/posts/what-is-the-scale-parameter/

Weibull Plot - GeeksforGeeks

Weibull plot is a graphical technique to determining if the dataset comes from a population that is logically be fit by a 2-parameter Weibull distribution. Before, discussing the Weibull plot in detail, we

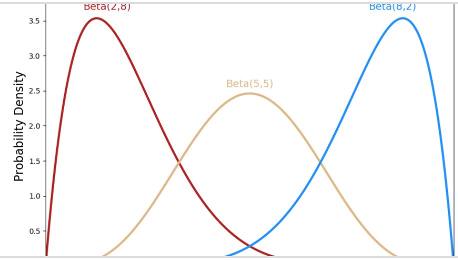
https://www.geeksforgeeks.org/weibull-plot/

Some Blogs:

Beta Distribution - Intuition, Examples, and Derivation

The Beta distribution is a probability distribution on probabilities . For example, we can use it to model the probabilities: the Click-Through Rate of your advertisement, the conversion rate of

https://towardsdatascience.com/beta-distribution-intuition-examples-and-derivation-cf00f4db57af



Notebook

Google Colaboratory

https://colab.research.google.com/drive/1YMGzZ6I4gsvctNIXbCYv1Sp1ictLc1Tf?usp=sharing