Uniform Distribution

A uniform distribution is a probability distribution where all outcomes are equally likely with in a given range.

Types

Discrete Uniform → Continous

ex:- A dice rolled

$\{1, 2, 3, 4, 5, 6\}$

$P(X=1) = \frac{1}{6}$, $P(X=2) = \frac{1}{6}$ - - -

Discrete

$X \sim U(a, b)$

Continous

Denotes → $X \sim U(a, b)$ → parameter $(a \text{ and } b)$

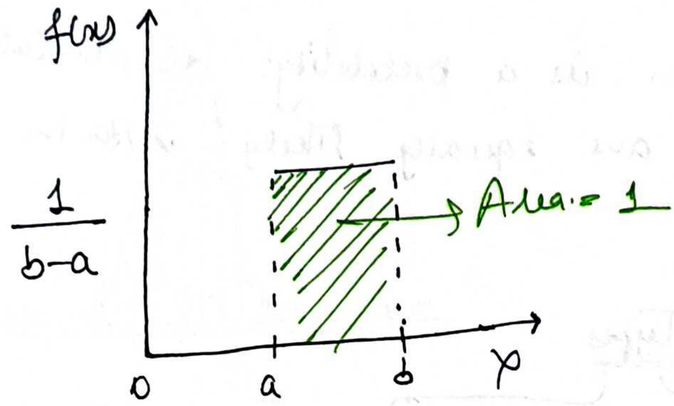
a ↓ lower, b ↓ higher

Example of continous

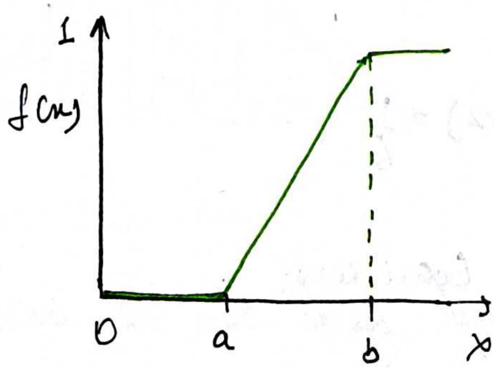
1. The height of a person randomly selected from a group of individuals whose heights range from 5'6 to 6'0 would follow continous uniform distribution.

Range is defining → Continous uniform.

* PDF CDF graphs



PDF



CDF

* Skewness → 0

* Application in Machine learning and Data Science.

a. Random Initialization: In many machine learning algo, such as neural networks and k-means clustering, the initial values of the parameter can have a significant impact on the final result. uniform distribution is often used to randomly initialize the parameter as it ensure that all values in the range have an equal prob of being selected.

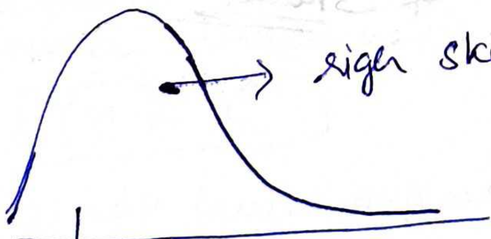
b. **Sampling**: For example, if you have a dataset with an equal number of sample from each class, you can use uniform distbn to randomly select a subset of the data that is representative of all the classes.

c. **Data augmentation**: may want to artificially increase the size of your dataset by generating new examples that are ~~sampling~~ similar to the original data. Uniform distbn can be used to generate new data that are within a specified rang of the original data. (DL → CNN)

d. Hyperparameter **tunning**: where you need to search for the best combination of hyperparameters for a machine learning model.

Log Normal Distribution

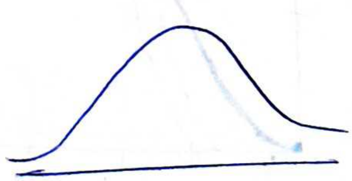
In prob theory and statistics, a lognormal distbn is a heavy tailed continous prob distbn of a random variable whose logarithm is normally distributed.

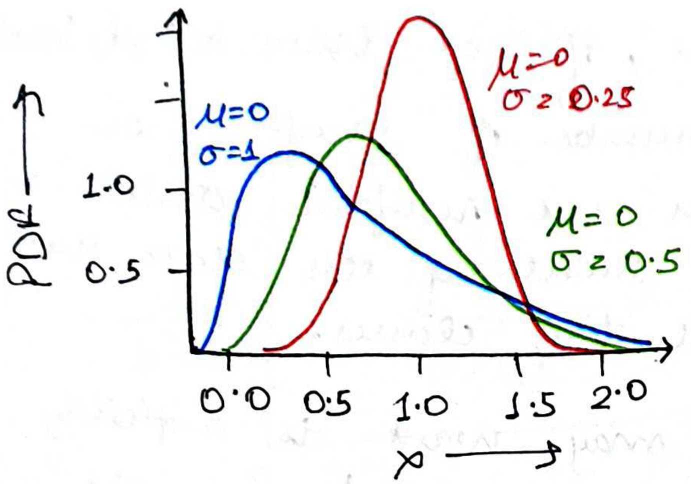


↳ do log of right skewed if and build again graph. If graph is Normal

$\log(x) \sim N(\mu, \sigma)$

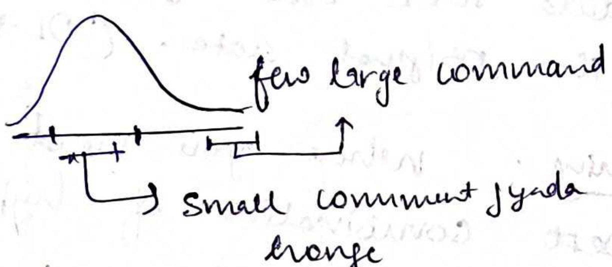
Age
log(–) →
log(–)
log(–).





Example

① Insta, reddit, facebook, youtube → comments

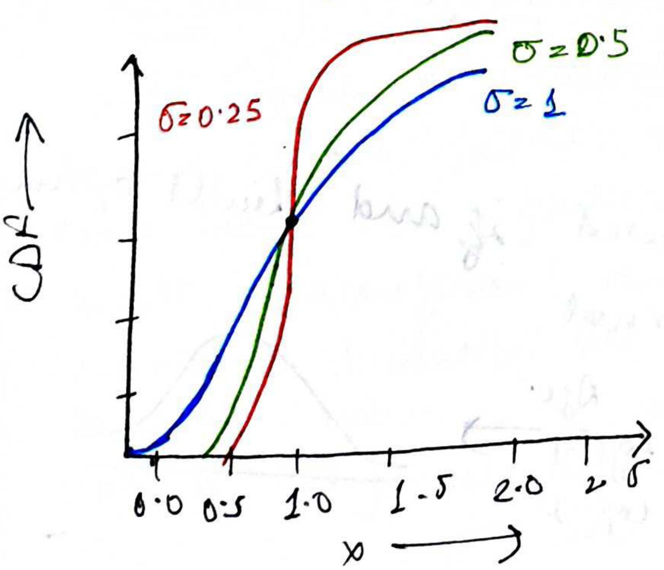


② In economics, there is evidence that the income of 97% - 99% of the population is distributed log normally.

denoty

$x \sim \text{Lognorm}(\mu, \sigma)$

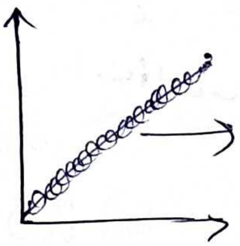
* Skewed distⁿ



* How to check if a random variable is log normally distributed? (Interview qn)

$X \sim$ log normal

$\hookrightarrow \log(x) \rightarrow$ draw graph of these log points if it make Normal distribution then random variable is log normally distributed



Normal distribution

Pareto Distribution

The Pareto distribution is a type of prob distribution that is commonly used to model the distribution wealth, income, and other quantities that exhibit a similar power-law behaviour.

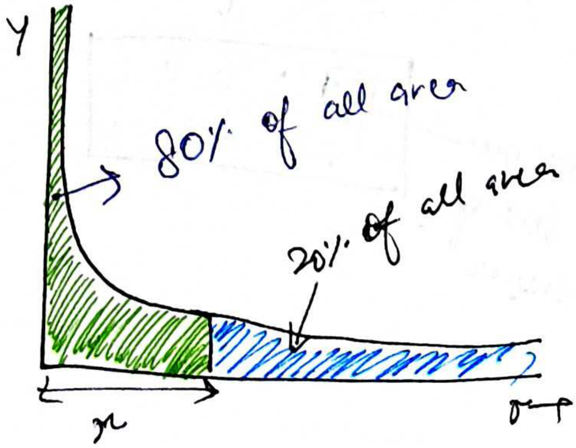
* Pareto is not appliable for all time if $\alpha = 1.16$ then we can say otherwise not

what is Power Law

In mathematics, a power law is a function relationship between two variables, where one variable is proportional to a power of the other. Specifically, if Y and X are two variables related by a power law, then the relationship can be written as:

$$\boxed{Y = k\,x^{\alpha}}$$

Vilfredo Pareto originally used this distribution to describe the allocation of wealth among individuals since it seemed to show rather well the way that a larger portion of the wealth of any society



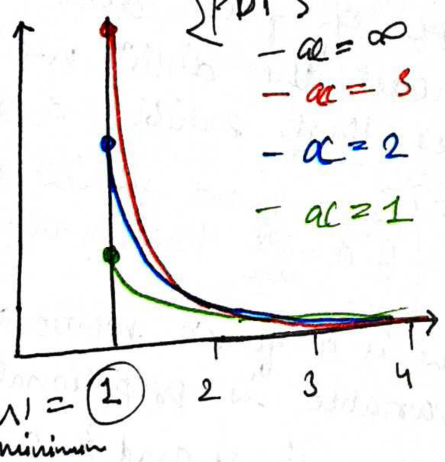
is owned by a smaller percentage of the people in that society. He also used it to describe distrib^n of income. This idea is something expressed more simply as the 'pareto principle' or the "80-20 rule" which says that 20% of the population controls 80% of the wealth.

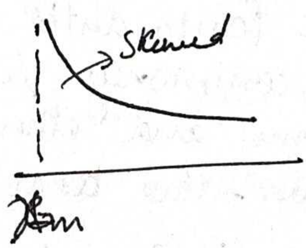
20% of people ⟶ 80% of wealth

80% of people ⟶ 20% of wealth.

Graph & parameter

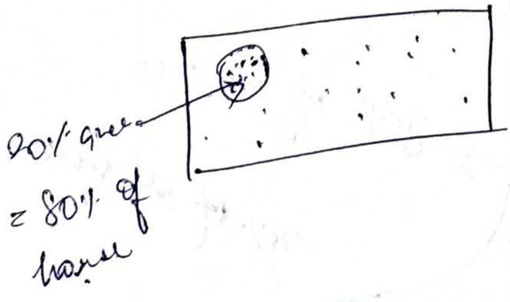
{PDF}



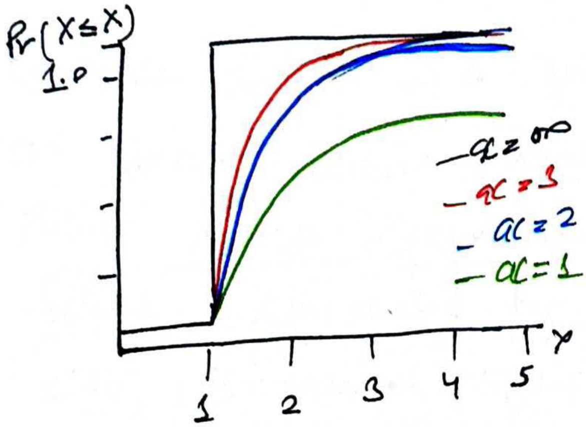


Example

* The size of human settlements (few cities)

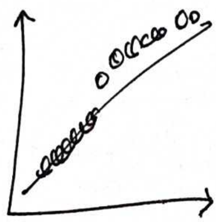


[CDF]



How to detect if a distribution is Pareto Distribution?

→ QQ Plot



Transformations

Log Transformation

Box-cox Transform

Bernoulli Distribution

Bernoulli distbⁿ is a probability distbⁿ that models a binary outcome, where the outcome can be either success (represented by the value 1) or failure (represented by the value 0). The Bernoulli distbⁿ is named after the Swiss mathematician Jacob Bernoulli who first introduced it in the late 1600s.

① Spam classifier

Spam — $S(1)$

not spam — $F(0)$

② rolling dice getting a 5

5 — $\left(\frac{1}{5}\right)$

~ 5 — $\left(\frac{5}{6}\right)$

PMF (Prob Mass func) $= P(X=x) = \boxed{p^{x}(1-p)^{(1-x)}}$

Coin toss

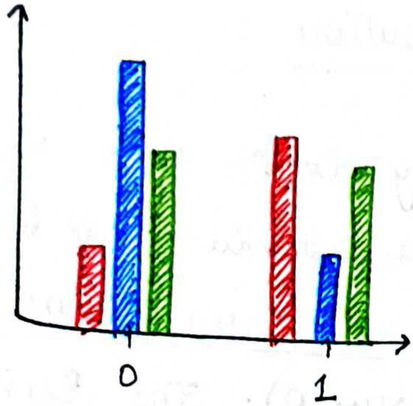
H — $(1)\left(\frac{1}{2}\right)$

T — $(0)\left(\frac{1}{2}\right)$

$P(X=1) = \left(\frac{1}{2}\right)^{1}\left(1-\frac{1}{2}\right)^{1-1}$

$P(X=1) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^{0} = \frac{1}{2}$

$P(X=0) = \left(\frac{1}{2}\right)^{0}\left(1-\frac{1}{2}\right)^{1-0} = \frac{1}{2}$



- $P(x=0)=0.2$, $P(x=1)=0.8$
- $P(x=0)=0.8$, $P(x=1)=0.2$
- $P(x=0)=0.5$, $P(x=1)=0.5$

Bernoulli distbⁿ is commonly used in ML for modeling binary outcomes, such as whether a customer will make a purchase or not

Bernaulli used in Naive Bayse

Binomial Distribution

Binomial distbⁿ is a prob distbⁿ that describe the number of successes in a fixed number of Independent Bernaulli trials with two possible outcomes (often called "success" and "failure"), where the prob of success is constant for each trial. The binomial distribution is characterized by two pramuters: the number of trials n and the prob of success p.

$$\frac{\text{bernaulli trial}}{n \text{ times} \rightarrow 5} \rightarrow \text{Binomial}$$

trial Independent to each other → exampe $\boxed{\text{feedback} \rightarrow 10 \text{ students}}$

1st student give negative feedback and tell to 2nd student to give negative feedback. So this trial is dependent.

$$\boxed{P(X=x) = {}^{n}C_{x}\, p^{x} (1-p)^{n-x}}$$

$n \rightarrow$ # of trials

$p \Rightarrow$ prob of success

$x \rightarrow$ desired result.

Q. The prob of anyone watching this lecture in the future and then liking it is 0.5. What is the prob that:

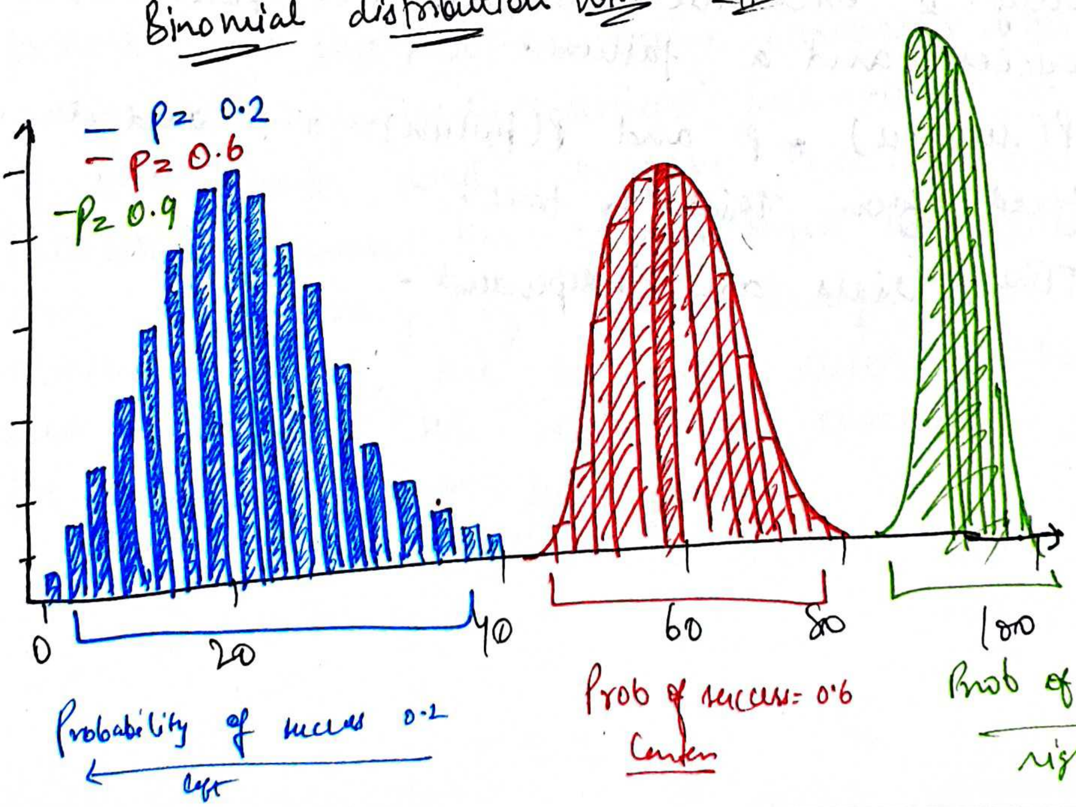
→ 2 out of 3 people will like it

$$= {}^{3}C_{2} \left(\frac{1}{2}\right)^{2} \left(\frac{1}{2}\right)^{3-2}$$

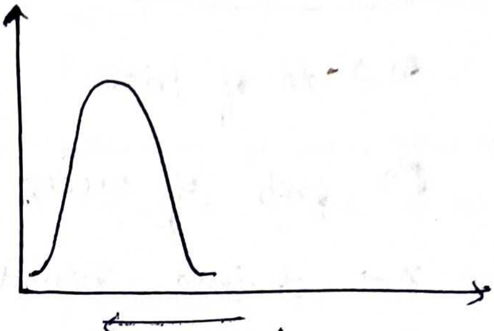
$$= \frac{3!}{2!\,1!} \times \frac{1}{8} \rightarrow 3 \times \frac{1}{8} = \frac{3}{8}$$

Graph of PDF :

Binomial distribution with different probabilities of success

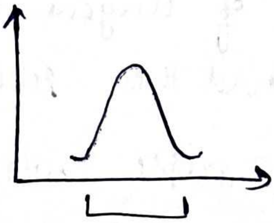


* If Prob of success is less like (0.1, 0.2, ..)
Graph plot at left.



left sid

* If Prob of success is centered like (0.4, 0.5, 0.6)
So graph plot at centre



centre.

* If Prob of success is higher (0.9, 0.8, ..)
So, graph plot at right

right side

Criteria

1. The process consists of n trials.
2. Only 2 exclusive outcomes are possible, a success and a failure.
3. P(success) = p and P(failure) = 1-p and it is fixed from trial to trail.
4. The trials are independent.

Application

1. <u>Binary classification problems</u>: In binary classification problems, we often model the prob of an event happening as a binomial distbⁿ. For example, in a spam detection system, we may model the prob of an email being spam or not spam using a binomial distbⁿ.

2. <u>Hypothesis testing</u>: In statistical hypothesis testing, we use the binomial distbⁿ to calculate the prob of observing a certain number of success in a given no. of trials, assuming a null hypothesis is true. This can be used to make decisions about whether a certain hypothesis is supported by the data or not.

3. Logistic Regression

4. <u>A/B testing</u>: A/B testing is a common technique used to compare two different version of a product, web page, or marketing campaign. In A/B testing, we randomly assign individuals to one of two groups and compare the outcomes of interest between the groups. Since the outcome are often binary (eg., click-through rate or conversion rate), the binomial distbⁿ can be used to model the distbⁿ of outcomes and test for differences betⁿ the groups.

Sample Distribution

India → pop data → salary

100 times → Sample data (50)

Sample 1 → 50 $\boxed{x_1 \ x_2 \ x_3 \ x_4 \ - - - \ x_{50}}$ → $\bar{x}_1$ (mean)

⋮

Sample99 → 50 $\boxed{x_1 \ x_2 \ x_3 \ - - - \ x_{50}}$ → $\bar{x}_{99}$ (mean)

Sample100 → 50 $\boxed{x_1 \ x_2 \ x_3 \ - - - \ x_{50}}$ → $\bar{x}_{100}$ (mean)

$\boxed{\bar{x}_1 \ \bar{x}_2 \ \bar{x}_3 \ - - - - \ \bar{x}_{100}}$ → Sampling Distbⁿ of mean

or

Sample1 → (50) $\boxed{x_1 \ x_2 \ - - - - \ x_{50}}$ → x_1 (variance)

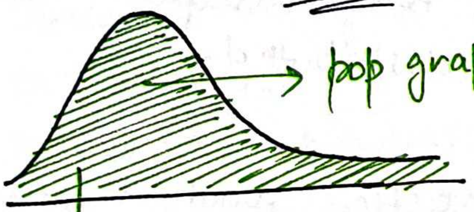
Sample2 → (50) $\boxed{x_1 \ x_2 \ - - - \ x_{50}}$ → x_2 (variance)

|

Sample100 → (50) $\boxed{x_1 \ x_2 \ - - - \ x_{50}}$ → x_{100} (var)

$\boxed{x_1 \ x_2 \ - - - - \ x_{100}}$ → Sampling distbⁿ of Variance def:- Sampling distbⁿ is a prob distbⁿ that describe the Statistical properties of a sample statistic such as the sample mean or sample proportion} computed from multiple independent sample of the sample size from a population.

Central Limit Theorem



→ pop graph (skewed)

If pop $\mu = \mu$ and var $= \sigma^2$. then

Sample size = 100

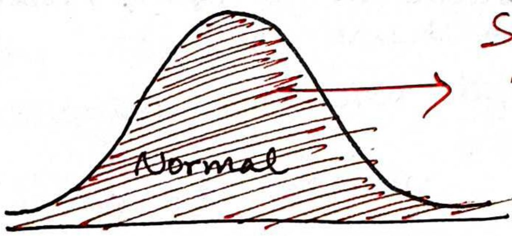
$x_1 \quad x_2 \; \cdots \; x_{100} \rightarrow \bar{x}_1$

$x_1 \quad x_2 \; \cdots \; x_{100} \rightarrow \bar{x}_2$

$\vdots$

$x_1 \quad x_2 \; \cdots \; x_{100} \rightarrow \bar{x}_{1000}$

$\bar{x}_1, \bar{x}_2, \bar{x}_3 \cdots \bar{x}_{1000}$



Sample mean $= \mu$ and var $= \frac{\sigma^2}{n}$

(*) Any graph of pop → always make Normal distbⁿ.

The central limit theorem (CLT) states that the distbⁿ of the sample means of a large no. of independent and identically distributed random variables will approach a normal distbⁿ, regardless of the underlying distbn of the variables.

The conditions required for the CLT to hold are:

1. The sample size is large enough, typically greater than or equal to 30.
2. The sample is drawn from a finite popln or an infinite popln with a finite variance.
3. The random variables in the sample are independent and identically distributed.

Then CLT is important in statistics and machine learning because it allows us to make probabilistic inferences about a population based on a sample of data. For example, we can use the CLT to construct confidence intervals, perform hypothesis tests, and make prediction about the popn mean based on the sample data. The CLT also provide theoritical justification for many commonly used statistical techniques, such as t-test, Anova and linear Regression.