Applied Data Science with Python

Probability Distribution

Learning Objectives

By the end of this lesson, you will be able to:

- Explain probability and its different aspects
- Identify the different types of probability distributions
- Examine the different probability distribution functions
- Explain the central limit theorem and estimation theory

Business Scenario

A retailer wants to optimize its product pricing strategy in order to increase sales and revenue. The company collects data from rivals on the prices of similar items and applies statistical analysis to calculate the mean and standard deviation.

Using this information, the analysts of the company can estimate the optimal price range for their own products to stay competitive in the market while maximizing their profit margin. They also use hypothesis testing to determine if there is a significant difference in sales between products priced at the upper end of the optimal range versus those at the lower end.

Based on the results, they adjust their pricing strategy accordingly to drive sales and revenue growth.

Probability, Its Importance, and Probability Distribution

Discussion: Probability

Duration: 10 minutes

- What is probability, and why is it important?
- What is a probability distribution?

What Is Probability?

Probability is a mathematical term for the likelihood of an event happening.

The chance of an event occurring is a number between 0 and 1.

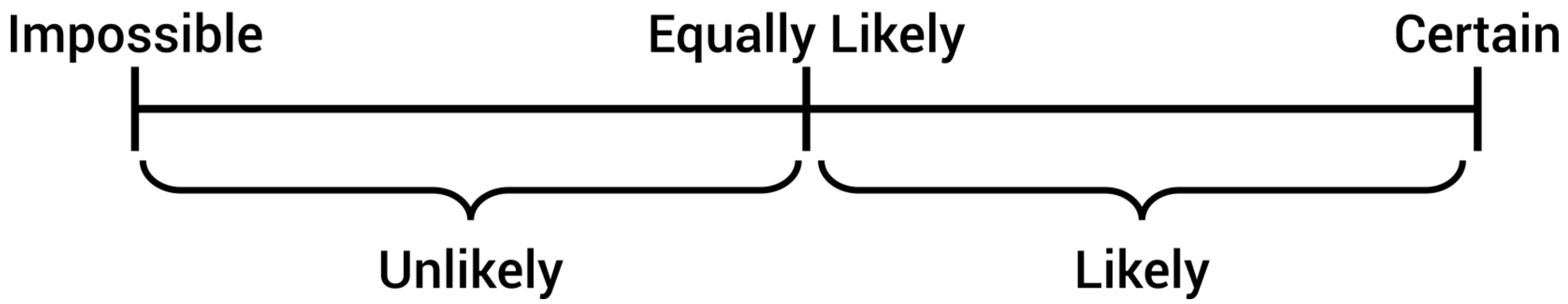
0: Impossibility of occurrence of event

1: Certainty of occurrence of event

What Is Probability?

Mathematically, the probability is defined as the ratio of favorable cases to the total number of cases.

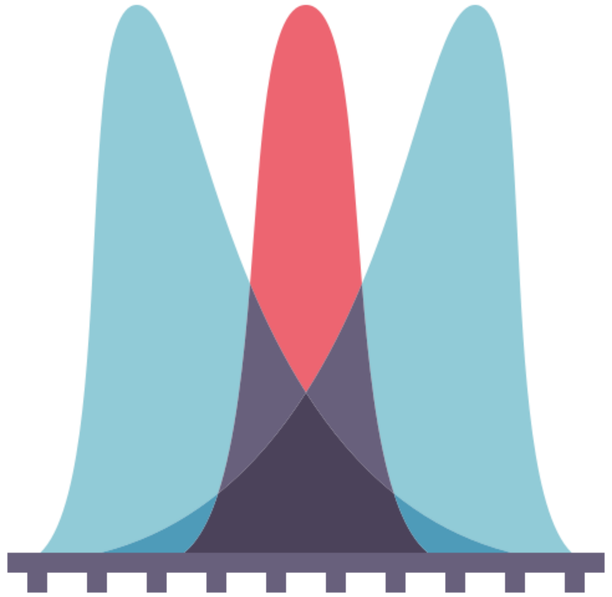
The Probability Scale



$$Probability = \frac{favorable\ cases}{total\ number\ of\ cases}$$

Importance of Probability

Probability helps to:

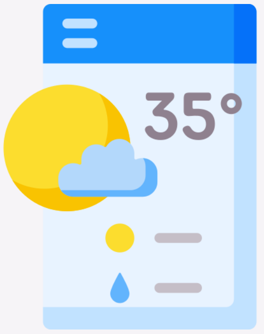


- Anticipate risks and identify ways to manage such risks
- Make forecasts about future events based on their likelihood

Probability is an essential tool in Mathematics and Statistics to address uncertainty in planning and decision-making.

Importance of Probability

It is used in a number of areas such as:



Weather forecasting

Scientific research

Healthcare

Example: Flipping of a Coin

Possession and starting direction are determined at the beginning of a football game using a coin toss

P{Head} = ½ or 50%

P{Tail} = ½ or 50%

The probability of getting the desired outcome while flipping a coin is 0.5 or 50%.

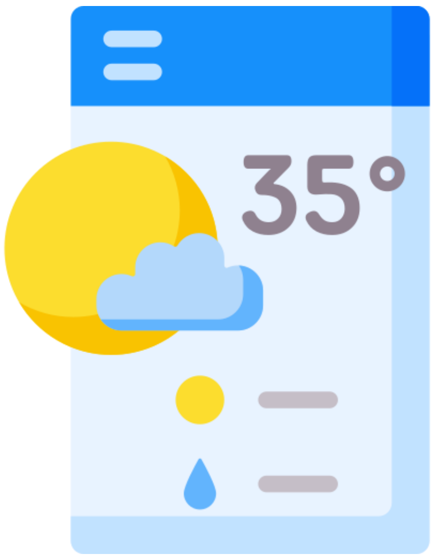
Example: Election Forecasts

The study of probability plays a key role in the prediction of election results.

Political analysts use the results of exit polls to predict whether a certain political party will come into power.

Example: Weather Forecasts

Weather forecasters use specific tools and techniques to predict the weather.

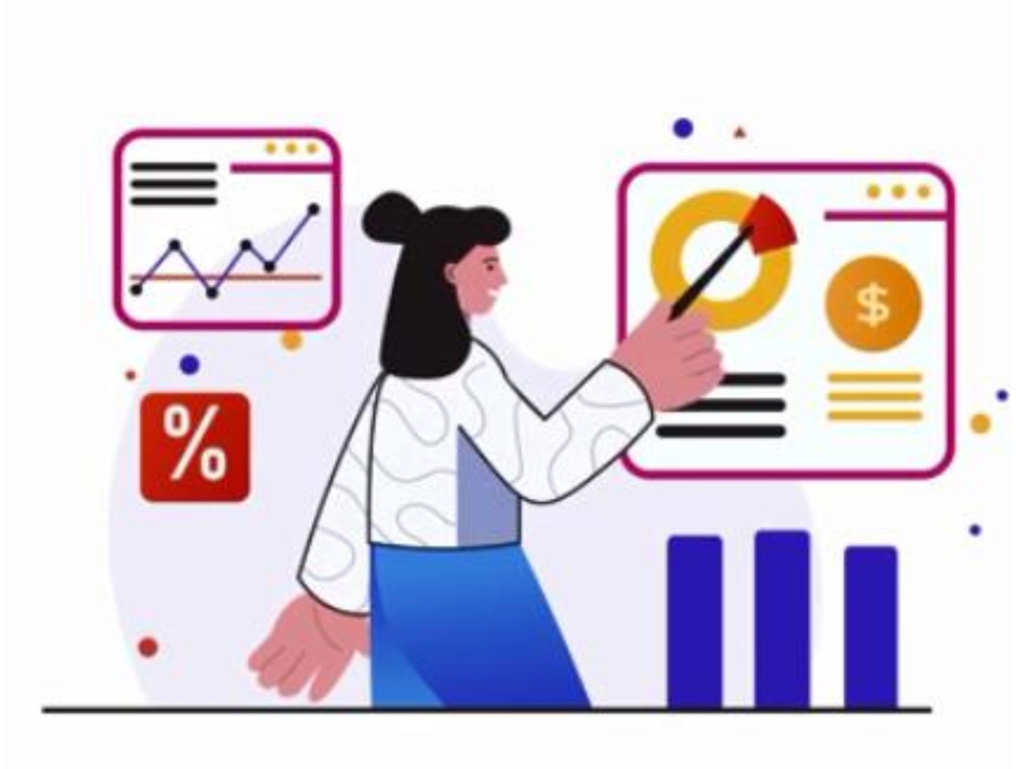


They look at all the historical databases of the days, where following characteristics were similar:

- Temperature
- Pressure
- Humidity

Probability Distribution

It is a discrete variable function whose integral over any interval represents the probability that the random variable described by it will fall inside that interval.



The values of the variable will vary based on the underlying probability distribution.

The notation used by statisticians to describe probabilities is p(x), which is the likelihood that a random variable takes a specific value of x.

Probability Distribution

There are two main types of probability distributions that are based on the type of variables:

Discrete variable

It represents counts.
Example: Throwing of a dice

Continuous variable

It represents measurable amounts.
Example: Weight

Single Random Variable

It is a variable that takes on random values with a specified probability distribution.

For a single random variable, distribution is divided into the following two types:

Discrete probability distributions for discrete variables

Probability density functions for continuous variables

The probability distribution indicates how the total probability, 1 or 100%, is distributed across all values.

Example: Rolling a Fair Dice

Its probability distribution is as shown:

Value	Probability
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6
Total	1

Types of Probability Distribution

There are five common types of probability distribution.

1. Bernoulli distribution
2. Binomial distribution
3. Poisson distribution
4. Normal distribution
5. Uniform distribution

Discussion: Probability

Duration: 10 minutes

- What is probability, and why is it important?

Answer: Probability is the mathematical measure of how likely an event will occur. It's vital in mathematics and statistics, as it helps manage uncertainty in planning and decision-making.

- What is a probability distribution?

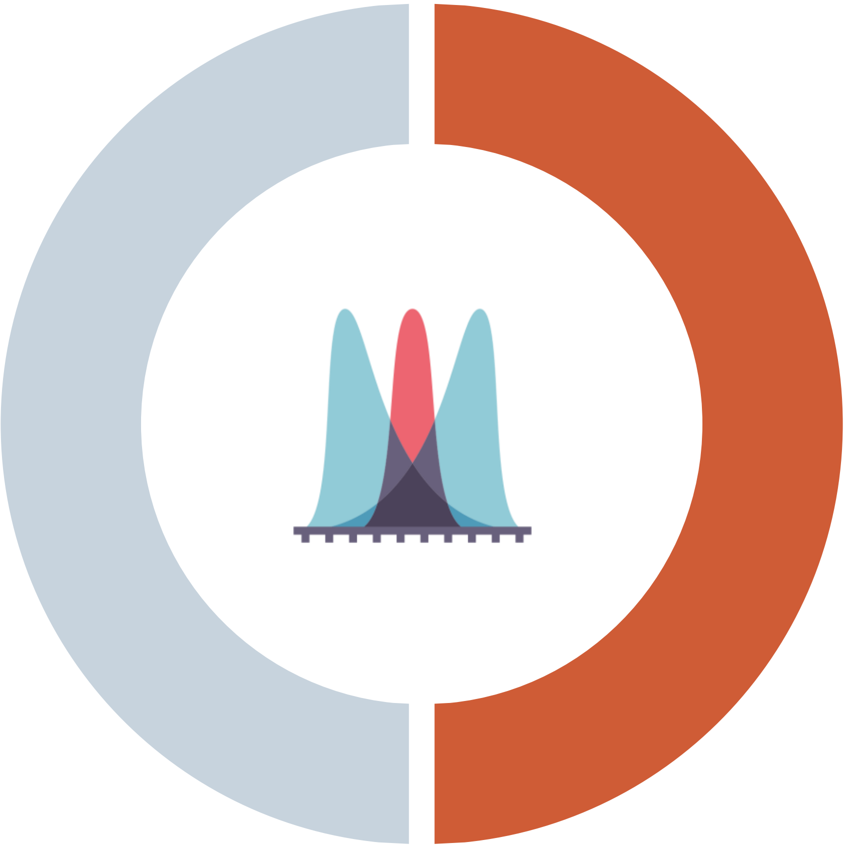
Answer: A probability distribution is a function that describes the likelihood of different outcomes for a random variable. The integral of this function over any given interval represents the chance that the random variable will fall within that range.

Probability Distribution: Bernoulli Distribution

Bernoulli Distribution

Bernoulli distribution is a discrete probability distribution that takes up only two distinct values:

1 denotes success



0 denotes failure

Bernoulli Distribution

Example: A coin flip where a head represents success, and a tail represents failure.

If the probability of success and failure is taken as **p** and **q**, respectively:

$p = 0<p<1$

$q = 1-p, 0<q<1$

Bernoulli Distribution

If **X** is a Bernoulli random variable:

$$Pr(X=1) = p$$

$$Pr(X=0) = 1 - p = q$$

If **f** is the probability mass function and **k** is the possible outcomes for the distribution:

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

Bernoulli Distribution

Cumulative distribution function, **F(X):**

$$F(X) = 0, \text{ if } X < 0$$
$$= 1\text{-}p, \text{ if } 0 \; X \leq 1$$
$$= 1, \text{ if } X \geq 1$$

Mean of **X**:

$$E(X) = Pr(X=1)*1 + Pr(X=0)*0$$
$$= p*1 + (1\text{-}p)*0$$
$$= p$$

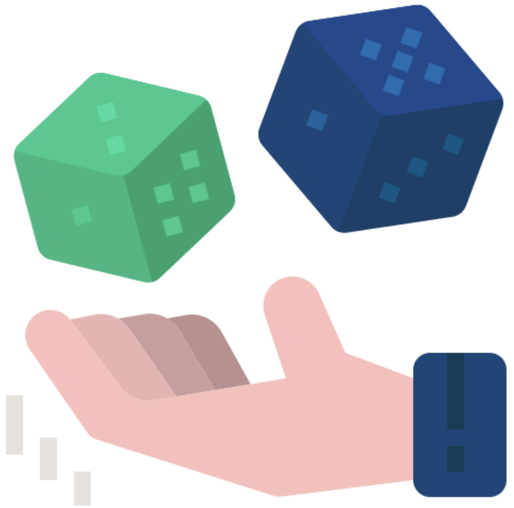
Variance of **X**:

$$E(X^2) = Pr(X=1)*1^2 + Pr(X=0)*0^2$$
$$= p$$

$$Var(X) = E(X^2) - E(X)^2$$
$$= p - p^2$$
$$= p\,(1\text{-}p)$$
$$= pq$$

Variance of a Bernoulli Random Variable

Example: What is the probability of receiving a 1 when a dice is rolled?



Probability of success, p = 1/6

Probability of failure, q = 5/6

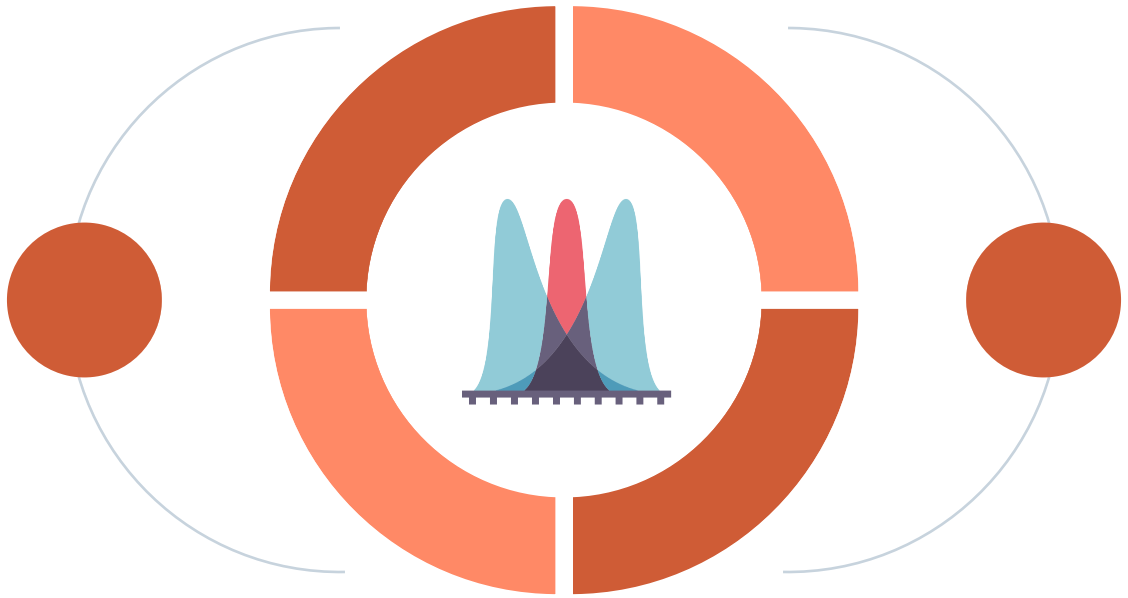
Variance, pq = 1/6 * 5/6 = 5/36

Probability Distribution: Binomial Distribution

Binomial Distribution and Poisson Distribution

Two commonly used discrete probability distributions are:

Binomial Distribution



Poisson Distribution

These are based on the principle of the **Bernoulli trials**.

Binomial and Poisson Distribution

A Bernoulli trial is an experiment that can result in one of two outcomes: success or failure.

Bernoulli test conditions

An independent trial should be conducted.

There is a limited number of trials

The probability of success or failure in each trial should be the same.

Each trial should have one of the two outcomes: success or failure.

Binomial Distribution

The formula for binomial distribution is given below:

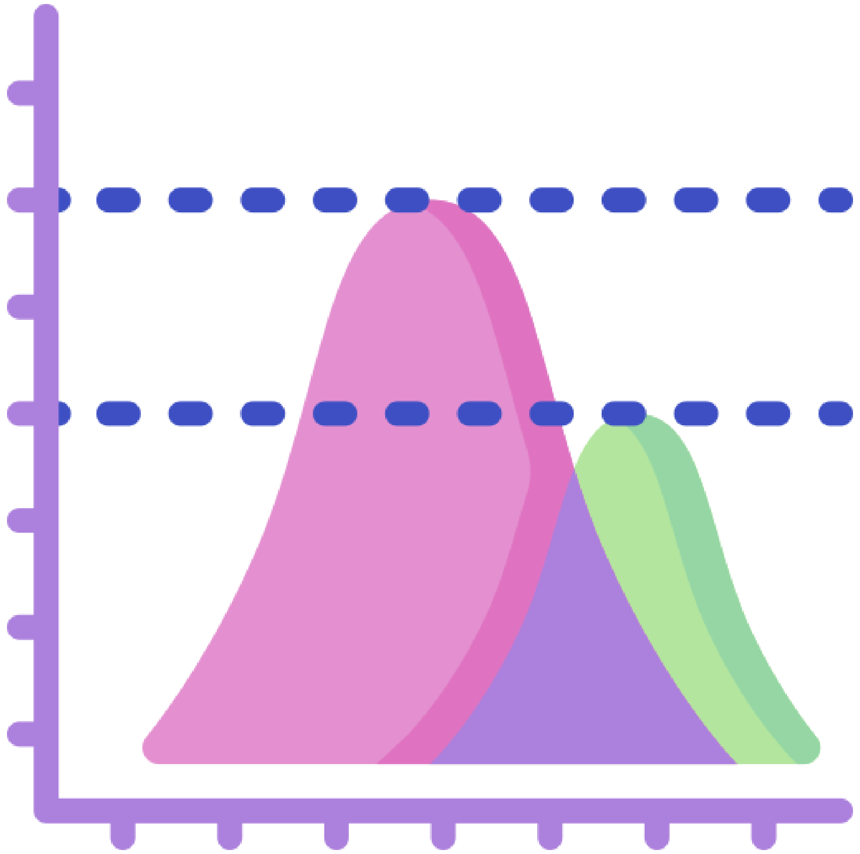
$$p(x) = \left(\frac{n}{x}\right) p^x q^{n-x} = \frac{n}{(n-x)\,!\,x\,!} p^x q^{n-x}$$

Where,

n - the number of trials
x - the number of successes desired
p - probability of getting a success in one trial
q – the probability of getting a failure in one trial

Binomial Distribution

If **p** is the probability of occurrence in a trial, the probability of non-occurrence is **q = 1-p**, which is also a constant.



Binomial Distribution

When **X** follows the binomial with parameters **n** and **p**,
X can take values 0, 1...., r..., n.

The binomial theorem follows the multiplication theorem of probability.

$$P\{X = 0\} = q*q....q*q = q^n$$

The outcomes are independent.

Binomial Distribution

Note: Probabilities of a value can be obtained from statistical tables or by using relevant software.

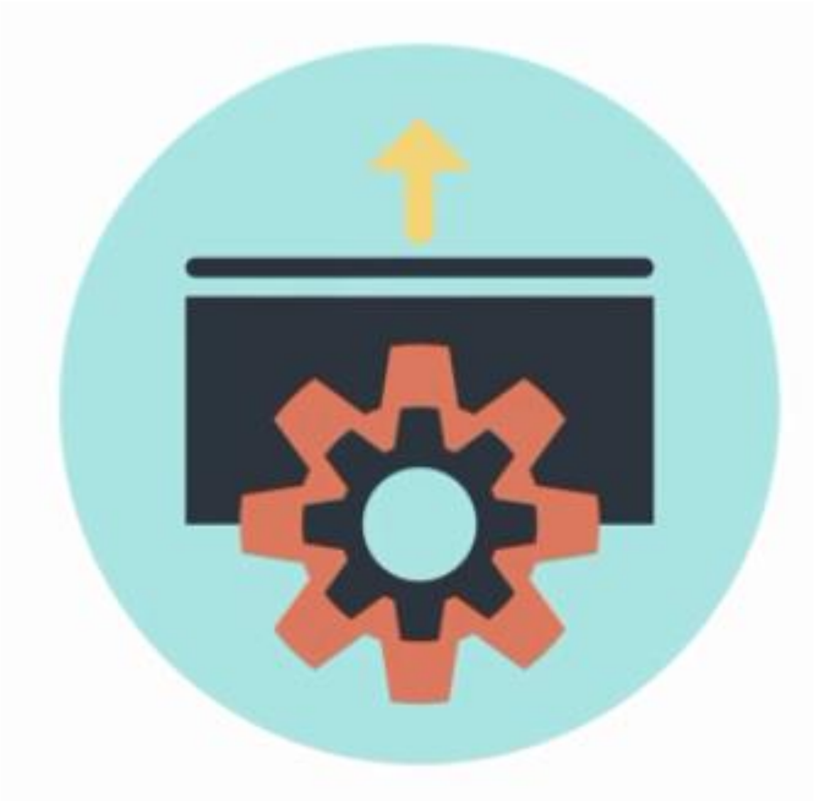
Example:

		p																			
n	*r*	.01	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
2	0	.980	.902	.810	.723	.640	.563	.490	.423	.360	.303	.250	.203	.160	.123	.090	.063	.040	.023	.010	.002
	1	.020	.095	.180	.255	.320	.375	.420	.455	.480	.495	.500	.495	.480	.455	.420	.375	.320	.255	.180	.095
	2	.000	.002	.010	.023	.040	.063	.090	.123	.160	.203	.250	.303	.360	.423	.490	.563	.640	.723	.810	.902

Binomial Distribution: Example

Consider a spare parts manufacturer with a 0.3 probability of producing a defective piece.

What is the probability that an inspection sample of size **n** will have four or more defectives?



Assuming that the process is Bernoulli:

- **X** indicates the number of defectives in the inspection sample
- **n** (number of trials) **= 10**
- **p** (probability parameter) **= 0.1**

Binomial Distribution: Example

Assume a table that displays the probabilities of **X** taking different values as shown below:

X	0	1	2	3	4	5	6	7	8	9	10	Total
Pr{X=r}	0.0282	0.1211	0.2335	0.2668	0.2001	0.1029	0.0368	0.0090	0.0014	0.0001	0.0000	1

Probability that the inspection sample has four or more defectives is:

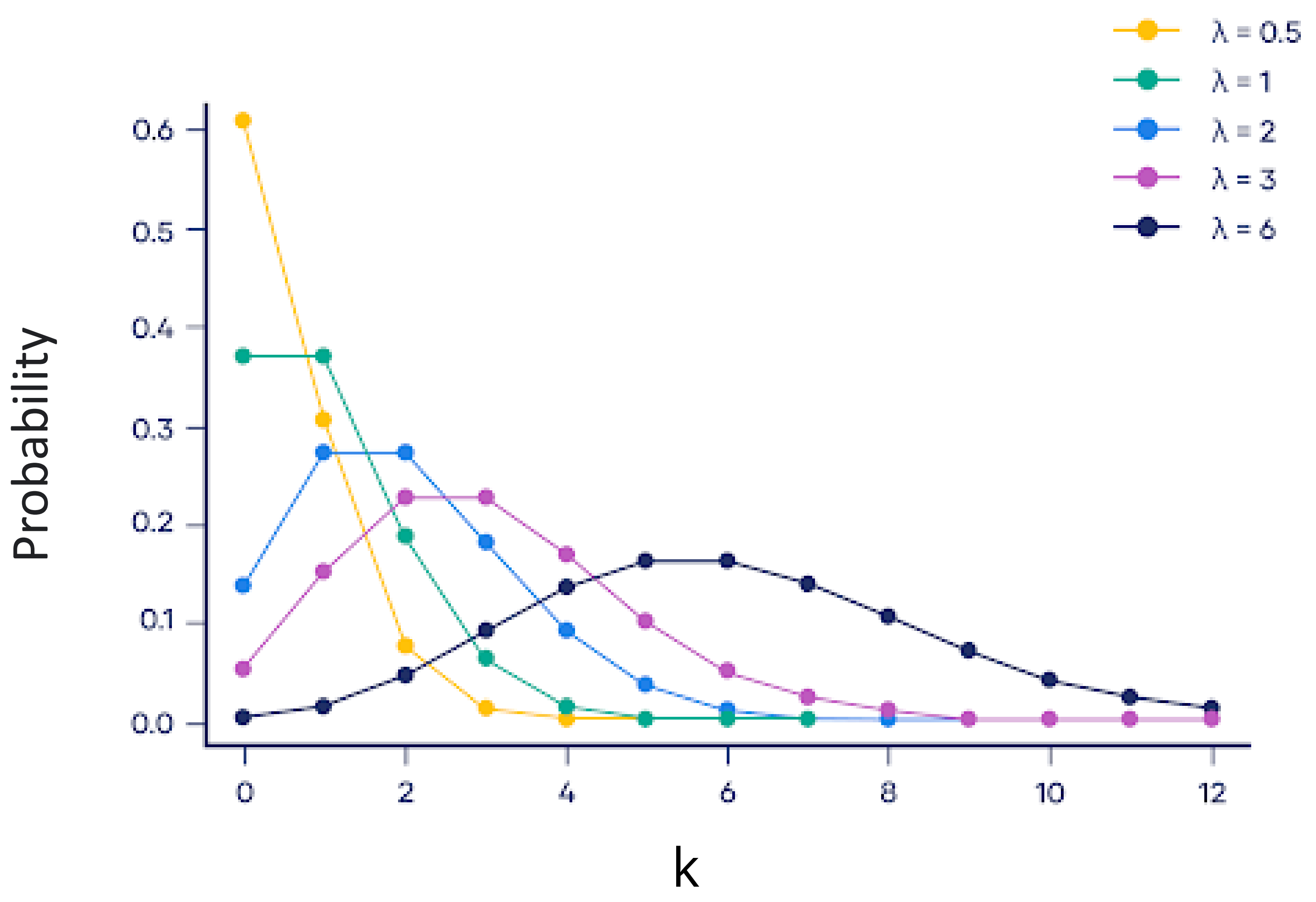
$$\Pr\{X \geq 4\} = 1 - \Pr\{X \leq 3\}$$
$$= 1 - 0.6496$$
$$= 0.3504$$

The result justifies the **assumption** that the process is Bernoulli.

Probability Distribution: Poisson Distribution

Poisson Distribution

Poisson distribution measures the probability of an event occurring over a specified period. It has only one parameter, (lambda), which is the mean number of events.



The graph above depicts examples of Poisson distributions with varying values of λ.

Parameters of Poisson Distribution

Consider that **X** denotes the random variable indicating the number of occurrences which follows a Poisson distribution.

Probabilities of **X** taking a given value (0, 1, 2 .....) depends on its expected value **λ**.

λ is the parameter of the distribution.

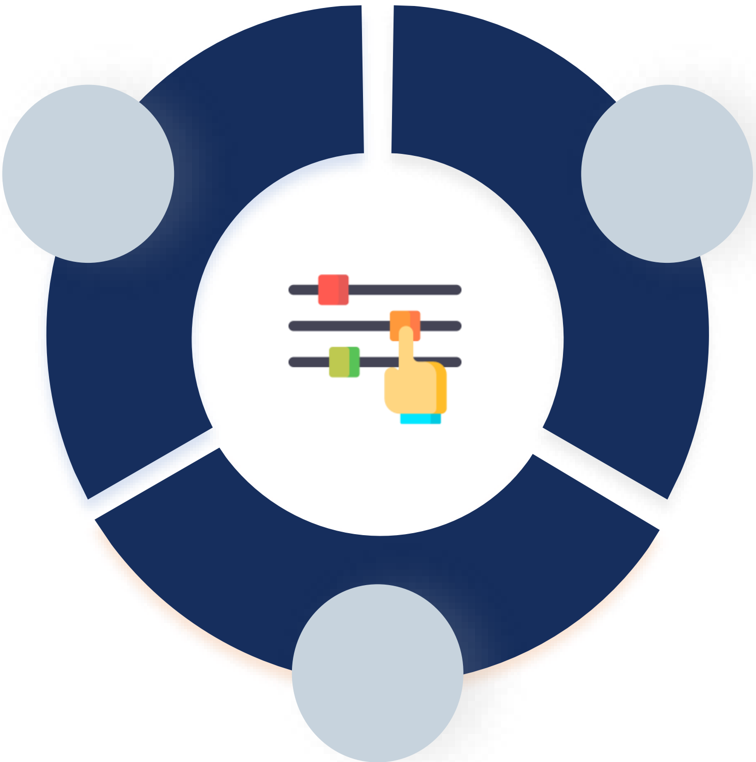
For a given value of **λ**, the probabilities can be obtained from tables or software.

Parameters of Poisson Distribution

When **X** follows a Poisson distribution with parameter **λ**:

Mean of X

$E(X) = \lambda$



Variance of X

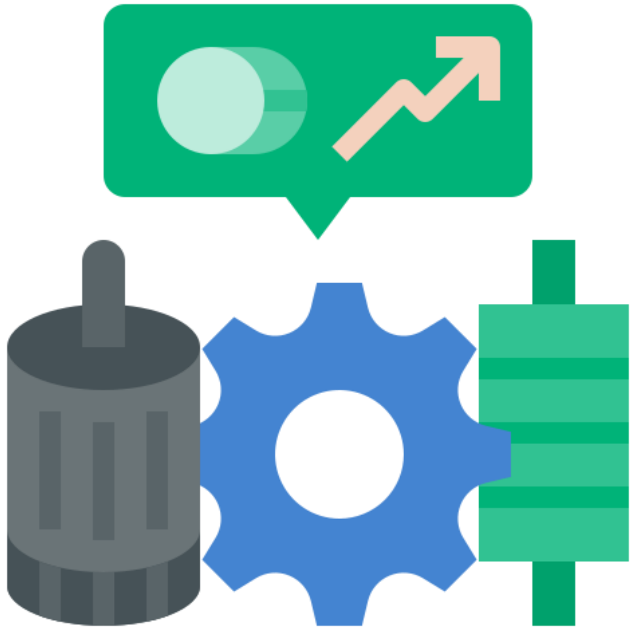
$Var(X) = \lambda$

Standard deviation of X

$S.D(X) = \sqrt{\lambda}$

Poisson Distribution: Example

The number of spares required for any component in a machine follows a Poisson distribution with $\boldsymbol{\lambda} = 2$.



A company that uses such machinery must determine the number of spares required such that the chance of a stockout is less than 0.06.

Poisson Distribution: Example

Solution: Let the random variable **X** denote the number of spares required. **X** follows a Poisson distribution with parameter $\boldsymbol{\lambda} = 2$.

The table shows the stock-out probabilities for different values of stocks.

K	0	1	2	3	4	5	6
P(X=k)	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	0.012
Cumulative	0.1353	0.406	0.6767	0.8571	0.9473	0.9834	0.9954
Stock out probability	0.8647	0.594	0.3233	0.1429	0.0527	0.0166	0.0046

Poisson Distribution: Example

The table provides the following observations:

Stockout probability for a given value = **1 – Cumulative value**

When fewer spare units are stocked, the stock out probabilities are high.

The firm should stock four units for the stock out probability to not exceed 0.06.

Binomial by Poisson Theorem

The binomial can sometimes be approximated by the Poisson theorem.

If

- n is very large
- p is very small

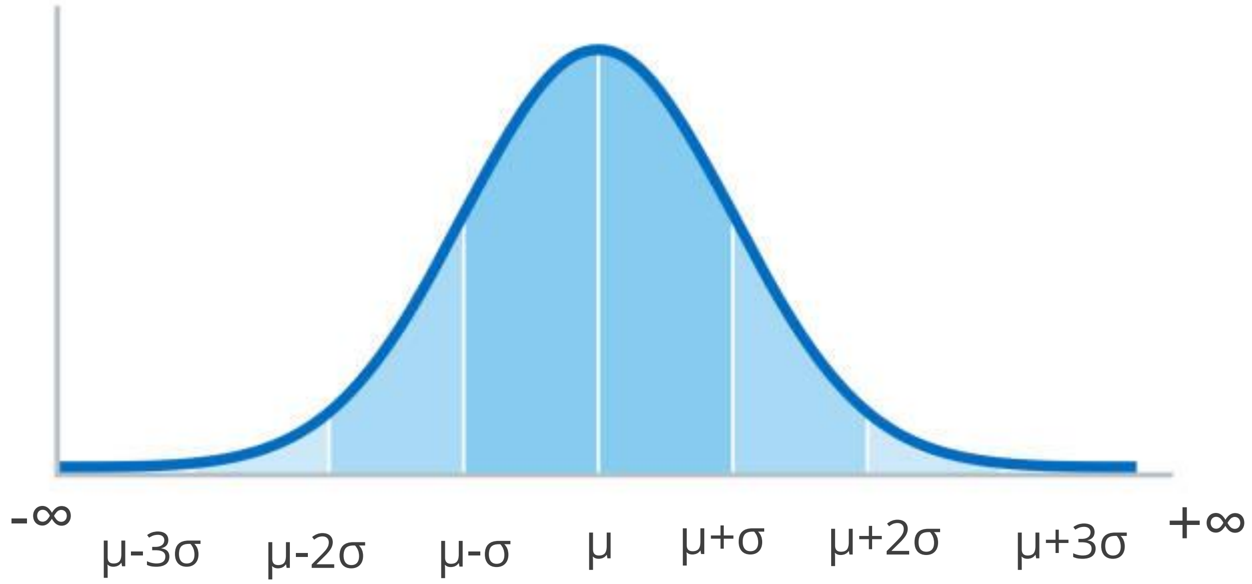
The binomial can be approximated as a Poisson with **λ = n*p**

Probability Distribution: Normal Distribution

Normal Distribution

It is a probability distribution that is symmetric, bell-shaped and describes the random variation of a continuous variable around the mean.

Normal distribution extends from -∞ to +∞.

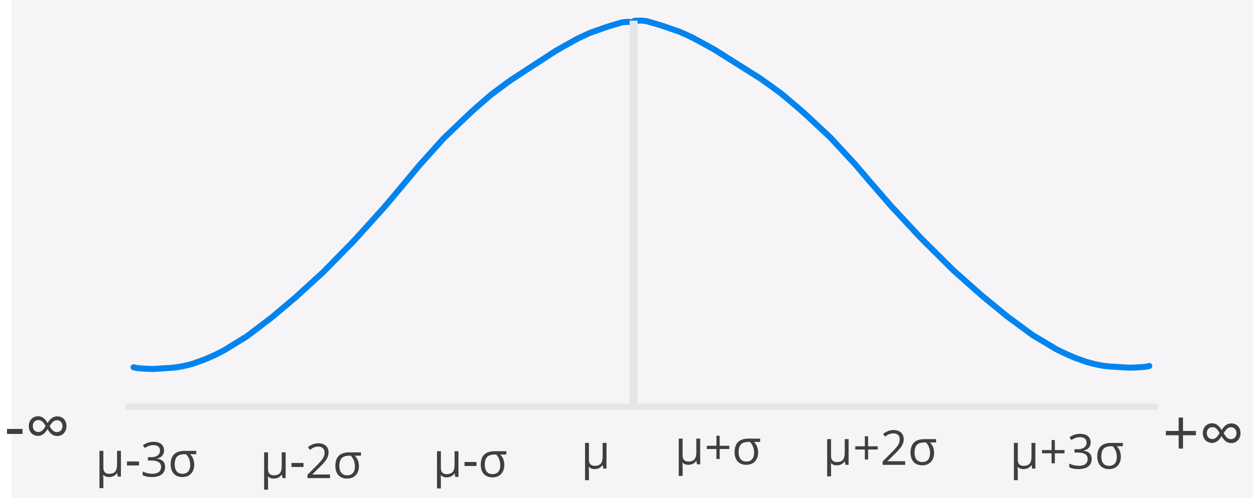


The distribution is completely specified by its mean (μ) and standard deviation (σ).

The value of the population mean (μ) can be positive, negative, or zero.

Normal Distribution

The properties of a normal distribution are:

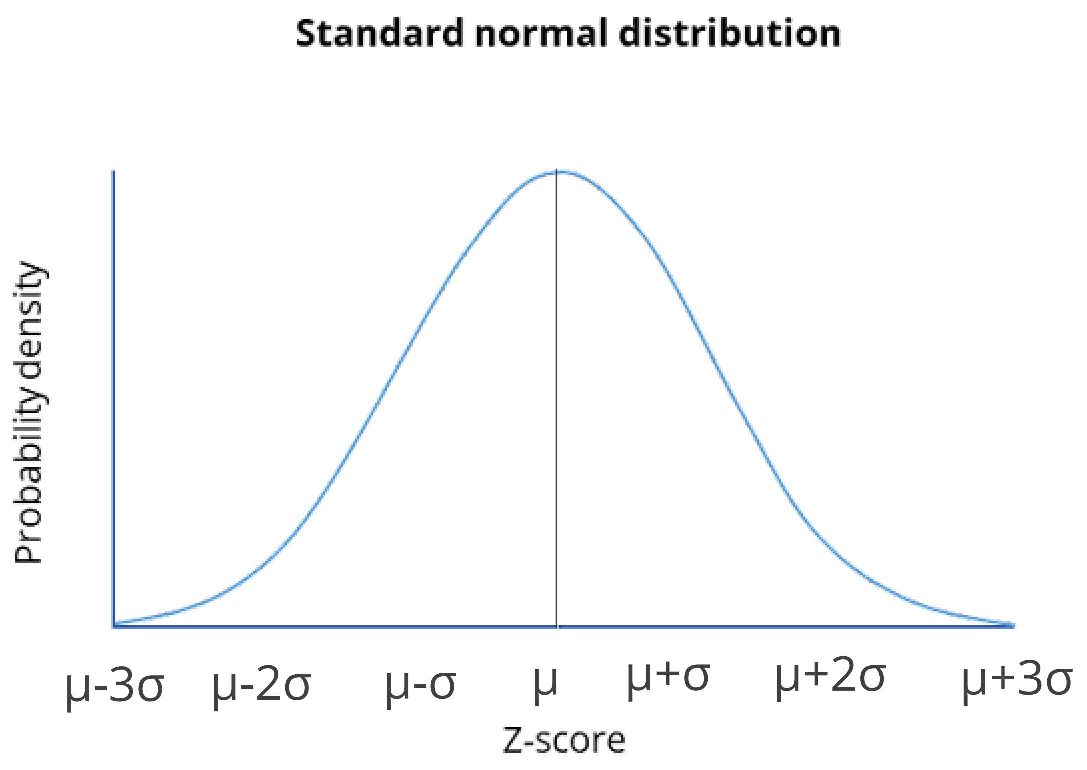


Distribution Curve

- The curve is symmetric about the population mean.
- The mean, median, and mode are equal.
- One half of the population is less than the mean; the other half is greater than the mean.

Standard Normal Distribution

The normal distribution with mean 0 and standard deviation 1 is called standard normal distribution.



When X follows a normal distribution with mean μ and standard deviation σ:

Z = (X – μ)/σ

Standard Normal Distribution

Statistical tables can be used to get probabilities of Z for the normal distribution when:

$\mu=0, \sigma=1$

For other values of mean and standard deviation, probabilities are obtained using:

$Z=(X - \mu)/\sigma$

Normal Distribution: Example

A factory outlet uses a filling machine for low-pressure oxygen shells.

The weights of filled cylinders (in grams) follow a normal distribution with:

Population mean (μ) = 1.433
Standard deviation (σ) = 0.033

Specification limits = 1.460 +- 0.085

Determine the probability that a cylinder will meet the following specification limits:

(1.460-0.085, 1.460 + 0.085) or (1.375, 1.545)

Normal Distribution: Example

Let the random variable X denote the weight of a filled cylinder (grams).

X follows a normal distribution with:

$\mu = 1.433$
$\sigma = 0.033$

Normal Distribution: Example

The probability that a cylinder meets specification limits is :

Here, $\mu = 1.433$ and $\sigma = 0.033$

$$\mathbf{Z = (X - \mu)/\sigma}$$

$$P\{1.375 < X < 1.545\}$$

$$P\{(1.375-1.433)/0.033 < Z < (1.545-1.433)/0.033\}$$

$$P\{-1.7575 < Z < 3.3939\}$$

$$= P\{-1.7575<Z<0\}+P\{0<Z<3.3939\}$$

$$=0.4608 + 0.5$$
$$=0.9608$$

Normal Distribution: Example

Pr {a cylinder meets specification limits}

= 0.9608

Pr {a cylinder does not meet specification limits}

= 1- 0.9608 = 0.0392

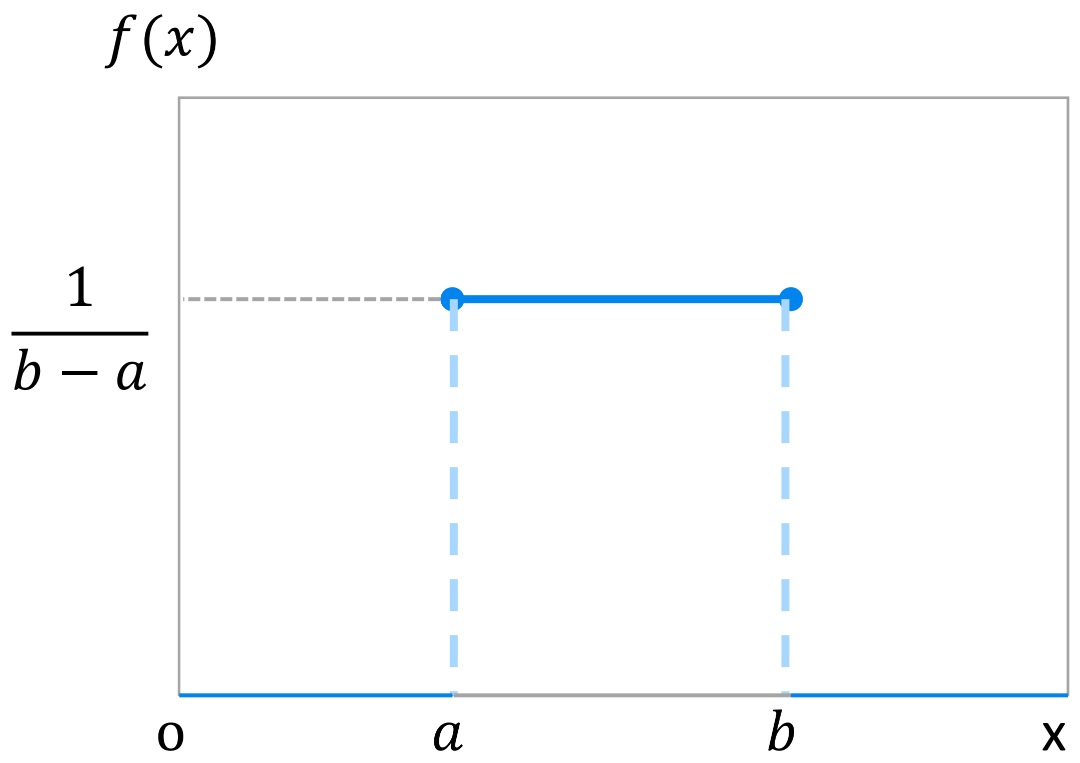
Percentage of cylinders that fall outside the specification limits:

3.92%

Probability Distribution: Uniform Distribution

Uniform Distribution

It is a probability distribution where all values in a given range are equally likely to occur, resulting in a rectangular-shaped probability density function.



Uniform Distribution

A random variable X follows a uniform distribution in the range [a, b], if the probability of lying in the range [c, d] is directly proportional to the length of the interval.

Uniform Distribution

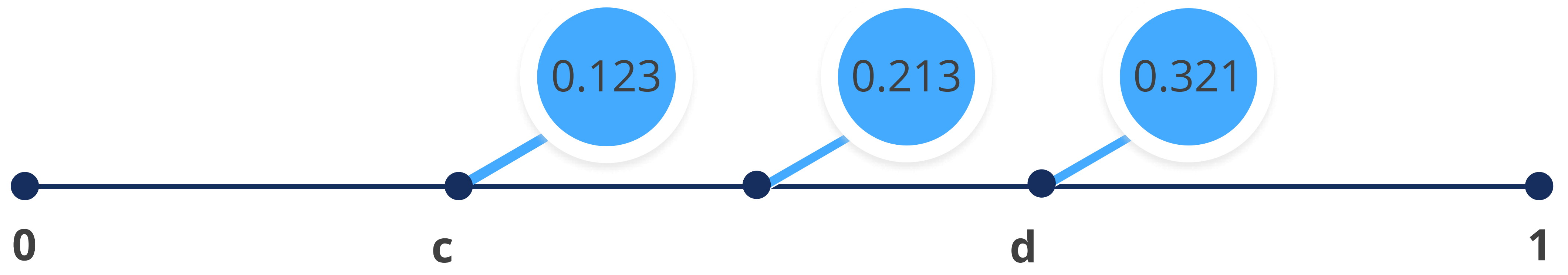
$$\Pr\{c < X < d\} = (d-c)/(b-a)$$

The values a, b, c, and d are all finite, but they can be positive, negative, or zero.

Uniform Distribution

Consider a large set of random numbers drawn from a random number table or a computer.

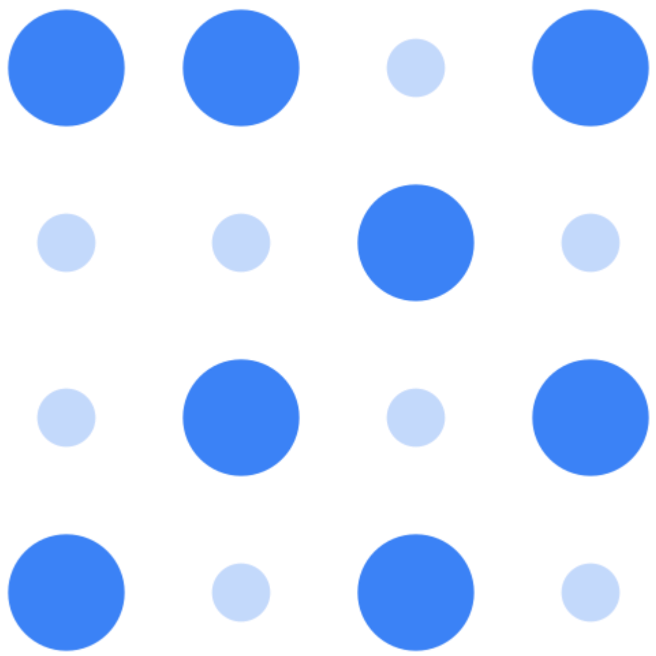
If a decimal is positioned at the start, the random numbers constitute a random sample from a uniform distribution with parameters 0 and 1.



The proportion of values lying between c and d, where $0 \le c \le d \le 1$, would be approximately d-c.

Use of Uniform Distribution

These values can be used to obtain random numbers from other distributions besides the uniform distribution.

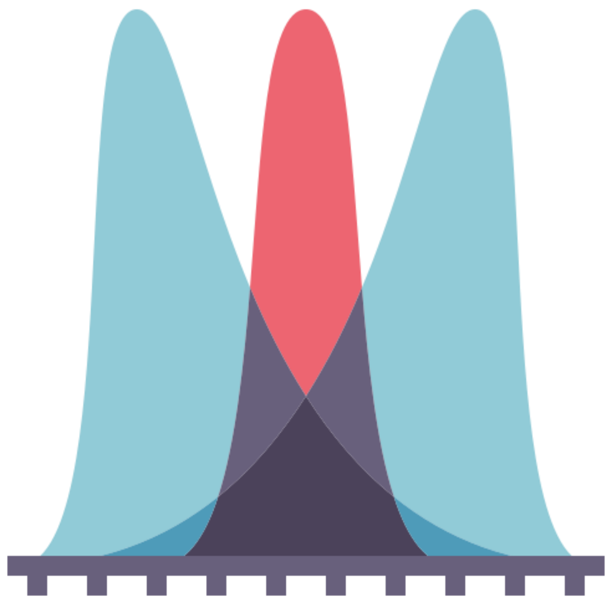


Random numbers are useful in several applications.

Probability Density Function and Mass Function

Probability Density Function (PDF)

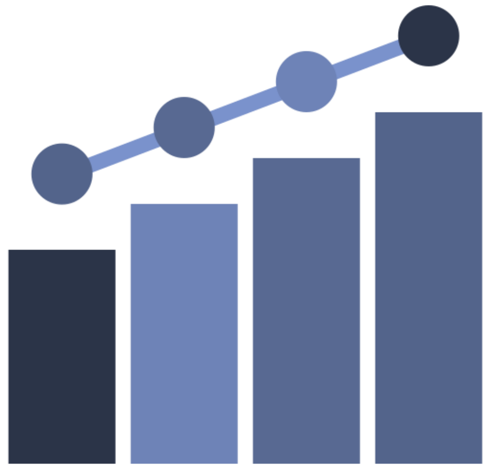
It is used to specify the probability that the random variable falls within a range of values as opposed to taking a particular value.



For a continuous random variable, the absolute likelihood that it takes on a particular value is 0, as there are infinite values it can take.

Probability Density Function (PDF)

When two samples are considered, the integral of the probability density function determines how likely it is to fall in one sample compared to the other.



It is the area under the density function above the horizontal axis between the lowest and highest values of that range.

The probability density function is non-negative everywhere and its integral over the entire space is equal to 1.

Probability Mass Function

When the random variables take only discrete values, the same function is called the probability mass function.

For a continuous random variable **X** with probability density function **f(x)** :

$$P(a \le X \le b) = \int_a^b f(x)\,dx$$

The probability density function is non-negative for all values of **X**.
Hence, $f(x) \geq 0$ for all **X**.

Probability Mass Function

Since the area under the density curve and above the horizontal axis over the entire space is equal to 1:

$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

Example

X is a continuous random variable with the following density function:

Let's assume that $f(x) = x^2 / 9$ if $0 < x < 3$, 0. To calculate the probability that X lies between 1 and 2, that is, $P(1 < X < 2)$:

$$P(1 < X < 2) = x^2 / 9\, dx = (2^3 / 27) - (1^3 / 27) = 7 / 27$$

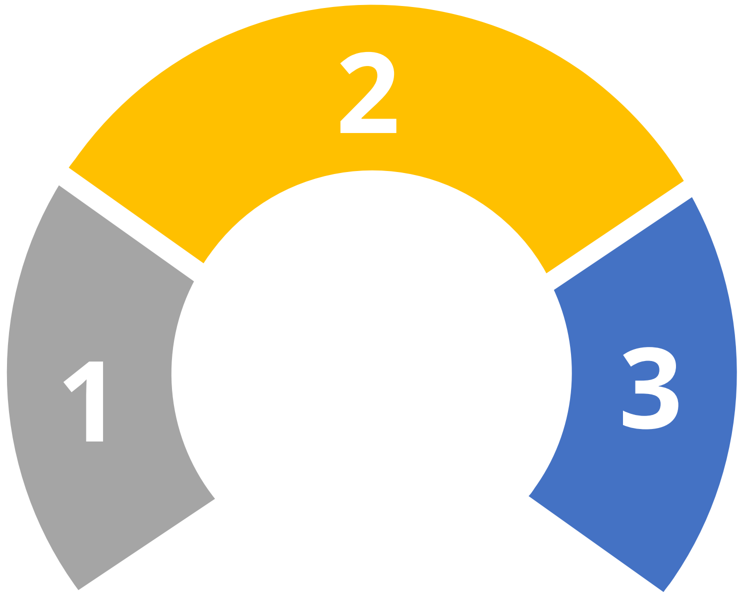
Cumulative Distribution Function

Cumulative Distribution Function

It is used to describe the probability.

It can be used to describe:

Continuous random variables



Discrete random variables

Mixed random variables

The cumulative distribution function of a real-valued random variable **X**, evaluated at **x**, is the probability that **X** will take a value less than or equal to **x**.

Cumulative Distribution Function

The cumulative distribution function (CDF) of a random variable **X** is defined as:

$$F_x(x) = P(X \leq x), \text{ for all } x \in R$$

The CDF is monotonously increasing, that is, if $x_1 \leq x_2$, then $F_X(x_1) \leq F_X(x_2)$.

Cumulative Distribution Function

The CDF of a continuous random variable can be expressed as an integral of its probability density function f_x.

$$F_x(x) = \int_{-\infty}^{\infty} f_x(t)dt$$

Every CDF is non-decreasing and right-continuous.

$$\lim_{x \to -\infty} F_x(x) = 0 \text{ and } \lim_{x \to \infty} F_x(x) = 1$$

Cumulative Distribution Function

Example 1

If **X** is uniformly distributed on the unit interval [0,1] then its CDF is:

$$
\begin{aligned}
F_x(x) &= 0 \text{ if } x < 0 \\
&= x \text{ if } 0 \le x \le 1 \\
&= 1 \text{ if } x > 1
\end{aligned}
$$

Example 2

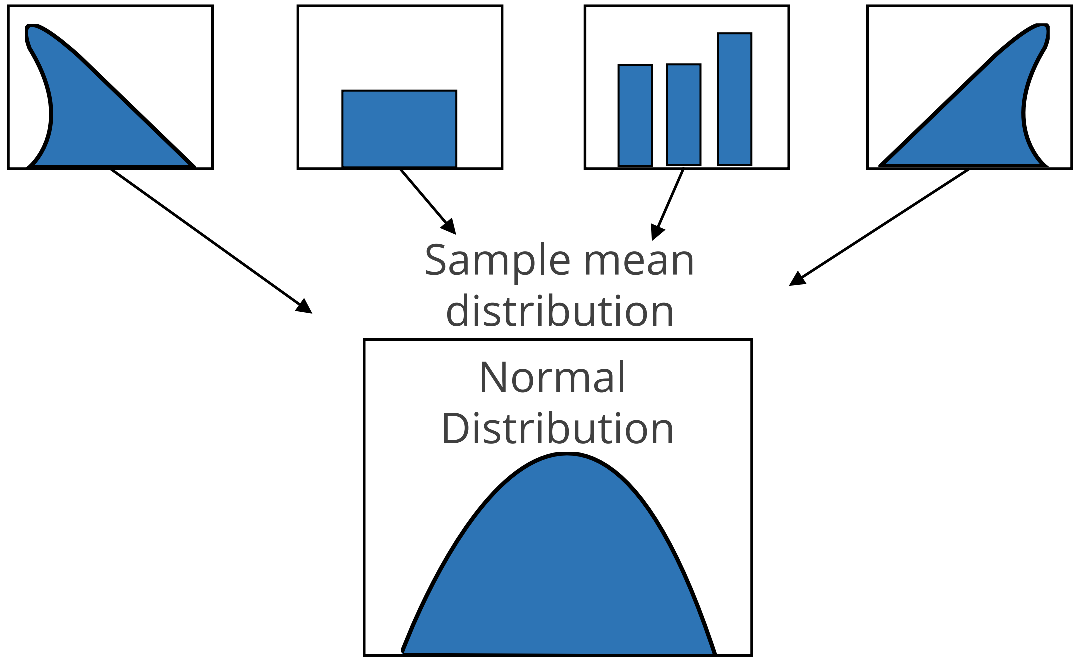
If **X** takes only the discrete values 0 and 1 with equal probability, then its CDF is:

$$
\begin{aligned}
F_x(x) &= 0 \text{ if } x < 0 \\
&= 1/2 \text{ if } 0 \le x < 1 \\
&= 1 \text{ if } x \ge 1
\end{aligned}
$$

Central Limit Theorem

Central Limit Theorem

The central limit theorem states that if sufficiently large samples are drawn from a population, the means of the samples will be normally distributed even if the population is not.



Consider the math scores of 1000 students from a school.

If 100 students from these are tested, the results will not deviate radically from the overall student population.

The average test score of these 100 students will match the average test score of the entire student population.

Central Limit Theorem

Assume that the test scores of 1000 students are not available and 100 students are tested.

The average test score of these 100 students reflects the population mean.

Central Limit Theorem

The central limit theorem helps calculate the probability that a particular sample was drawn from a given population.

If the probability is low, it can be concluded that the sample isn't from that population or isn't representative of that population and needs to be collected again.

Central Limit Theorem

Central limit theorem states that if X_1, X_2, X_3,......, X_n are independent random variables that are identically distributed and have finite mean (**μ**) and variance (**σ** ^{2}):

$$S_n = X_1+X_2+X_3+......+X_n \quad (n = 1,2,.....)$$

$$\lim_{n\to\infty} P(a \leq (S_n - n^{\mu})/\sigma\sqrt{n} \leq b) = (1/\sqrt{2\Pi}) \int_a^b e^{(-u2/2)}\, du$$

The random variable which is the standardized variable corresponding to $\mathbf{S_n}$ is asymptotically normal.

The theorem holds true when X_1, X_2, X_3,......, X_n are independent random variables with the same mean and variance but not necessarily identically distributed.

Estimation Theory

Estimation Theory

Estimation theory is the science of guessing or estimating the properties of a population from which data is collected.

The guesswork happens by determining the approximate value of a population parameter based on a sample statistic.

Estimation Theory

The estimator is a rule or formula that is used to calculate the estimate based on the sample.

Good estimators:

Are unbiased	Have minima variance
The average value of the estimator equals the parameter to be estimated.	The best one among all the unbiased estimators has a sampling distribution with the smallest standard error.

Types of Estimators

There are two types of estimators:

Point estimator

Interval estimator

Point Estimators

Point estimators are functions that are used to estimate the value of a population parameter using random samples of the population.

Unknown population parameter	Symbol	Empirical point estimator	Symbol
Population mean	μ	Sample mean	$\bar{x}$
Population standard deviation	σ	Sample SD	s
Population proportion	p	Sample proportion	$\hat{p}$

Point estimates are never perfect, and they always have an error component (margin of error)

Point Estimators

A point estimator is assessed on three criteria:

Unbiasedness (mean): It is a measure of whether the mean of this estimator is close to the actual parameter.

Efficiency (variance): It denotes whether the standard deviation of this estimator is close to the actual parameter.

Consistency (size): It indicates whether the probability distribution of the estimator is concentrated on the parameter with an increase in the sample size.

Interval Estimators

An interval estimator of a population parameter under random sampling consists of two random variables.

The upper limit

The lower limit

These values decide intervals that expect to contain the parameter estimated.

Interval Estimators

Interval estimates are all the ranges that an interval estimator can assume.

The range within which a population parameter probably lies is captured by the interval estimate.

Example

The mean income of a class of graduate trainees is between $775 and $950 per week.

Interval Estimators

An interval estimator can be assessed through its:

Accuracy or confidence level

Precision or margin of error

The design of an interval estimator consists of evaluating an unbiased point estimator and designating an interval of logical width around it.

Key Takeaways

- Probability is a mathematical term for the likelihood of any event happening.
- The different probability distribution are Binomial, Poisson, Normal, Bernoulli, and Uniform distributions.
- The probability density function is used to specify the probability that the random variable falls within a range of values.
- The central limit theorem helps calculate the probability that a particular sample was drawn from a given population.

Knowledge Check

Knowledge Check 1

What is the formula for finding the probability of X taking a given value in a binomial distribution?

A. Pr {X = n} = pn

B. Pr {X = 0} = qn

C. Pr {X= r} = nCr * pr* qn-r

D. Pr {X ≥ 4} = 1 - Pr {X ≤ 3}

Knowledge Check 1

What is the formula for finding the probability of X taking a given value in a binomial distribution?

A. Pr {X = n} = pn

B. Pr {X = 0} = qn

C. Pr {X= r} = nCr * pr* qn-r

D. Pr {X ≥ 4} = 1 - Pr {X ≤ 3}

The correct answer is **C**

The formula for finding the probability of X taking a given value in a binomial distribution is Pr {X= r} = nCr * pr* qn-r , where r can vary from 0 to n.

Knowledge Check 2

What is the Poisson distribution parameter representing the expected value of occurrences?

A. p

B. λ

C. q

D. σ

Knowledge Check 2

What is the Poisson distribution parameter representing the expected value of occurrences?

A. p

B. λ

C. q

D. σ

The correct answer is **B**

Poisson distribution parameter λ represents the expected value of occurrences of an event.

Knowledge Check

3

What are the two criteria for a good estimator?

A. Unbiasedness and consistency

B. Variance and bias

C. Bias and consistency

D. Variance and efficiency

Knowledge Check

A

What are the two criteria for a good estimator?

A. Unbiasedness and consistency

B. Variance and bias

C. Bias and consistency

D. Variance and efficiency

The correct answer is **A**

A good estimator is one that is unbiased and consistent.

Thank You