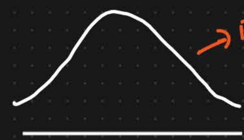
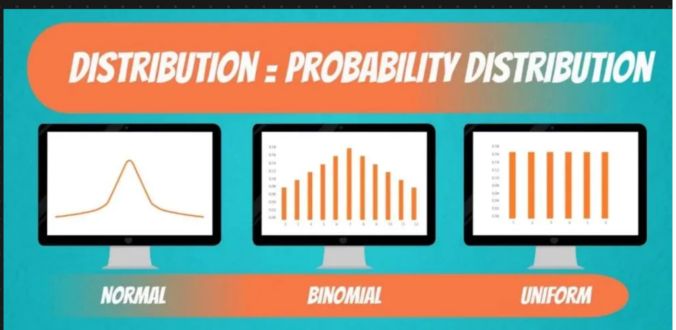
Distribution

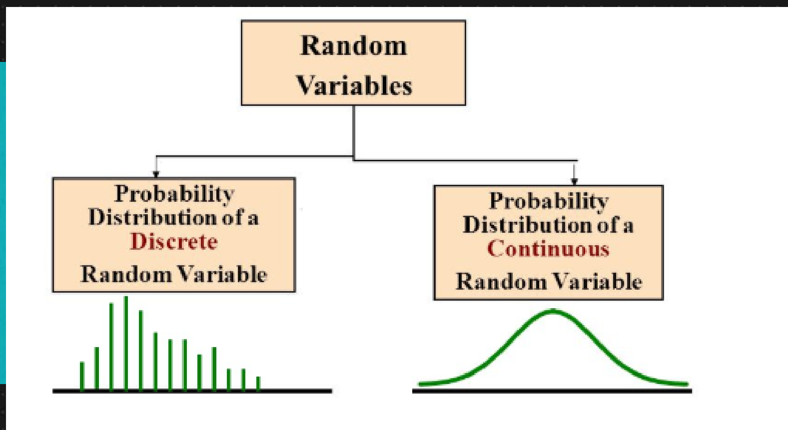
Types Of The Probability Distribution



→ Bell Curve

1. Normal / Gaussian Distribution (pdf)
2. Bernoulli Distribution (pmf) → Outcomes are 2 {Success Or Failure}
3. Uniform Distribution
4. Log Normal Distribution (pdf)
5. Poisson Distribution (pmf)
6. Power Law Distribution (pdf) → 80-20% Rule
7. Binomial Distribution (pmf)





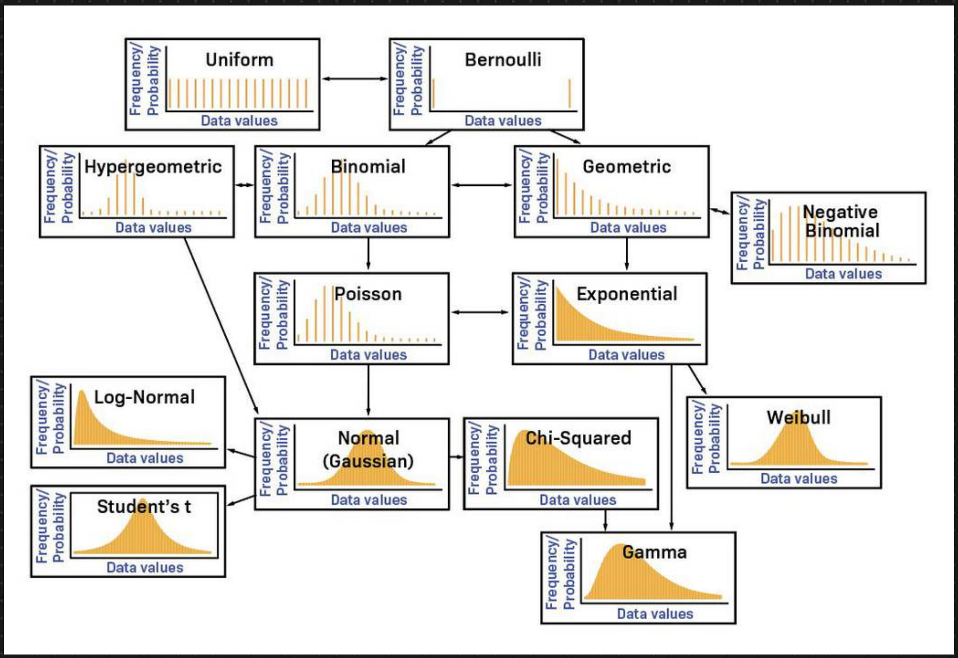
Types of Probability Distribution

Contnous Probability Distribution

- Gaussian Normal Distribution
- Standard Normal Distribution (Z- dist.)
- Student T-distribution
- Uniform Distribution
- Log-Normal Distribution
- Chi-Square Distribution

Discrete Normal Distribution

- Bernoulli Distribution
- Binomial Distribution
- Negative Binomial Distribution
- Geometric Distribution
- Poisson distribution
- Uniform Distribution


Uniform
Bernoulli
Hypergeometric
Binomial
Geometric
Negative Binomial
Poisson
Exponential
Log-Normal
Normal (Gaussian)
Chi-Squared
Weibull
Student's t
Gamma
Frequency/ Probability
Data values

Distribution

Probability Density Function And Probability Mass Function

Cumulative Distribution Function (cdf)

① PMF :

① Discrete Random Variable

PMF

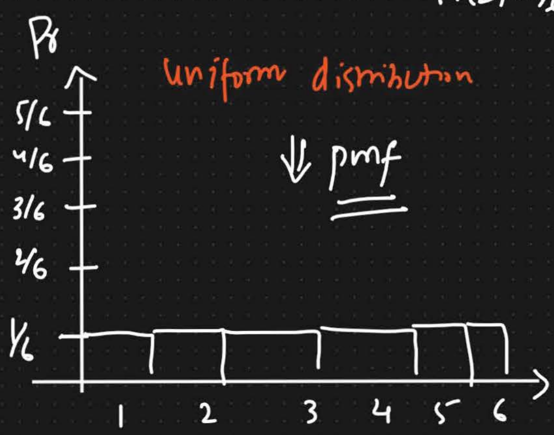
Eg: Rolling a dice

$\{1, 2, 3, 4, 5, 6\}$ $Pr(1) = 1/6$

$Pr(2) = 1/6$

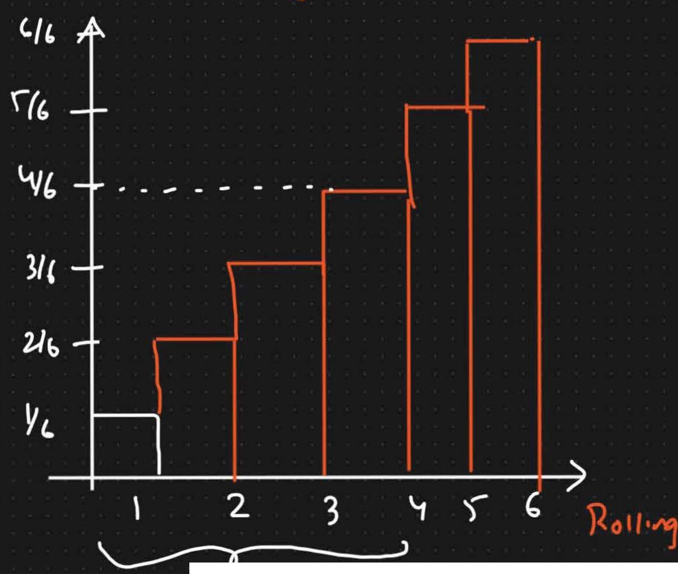
uniform distribution

$\Downarrow$ pmf



CDF

Cumulative Probability



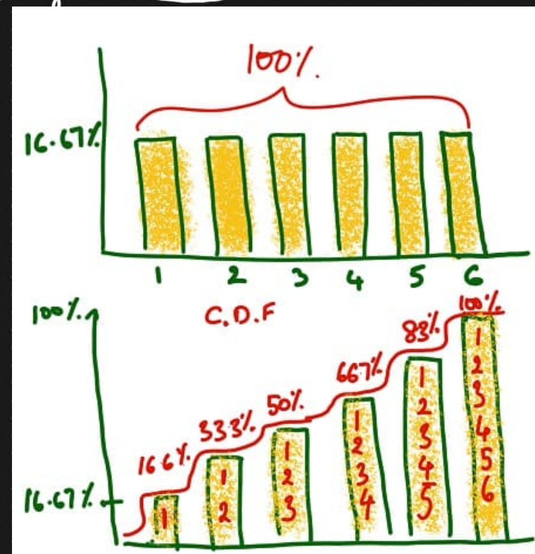
$Pr(X=1) = \frac{1}{6}$

$Pr(x=2) = \frac{1}{6}$

$Pr(X \le 4) = Pr(X=1) + Pr(X=2)$

$+ Pr(X=3) + Pr(X=4)$

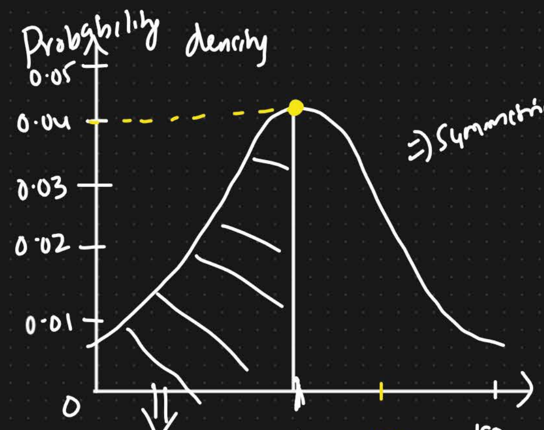
$= 1/6 + 1/6 + 1/6 + 1/6 = 4/6 = 2/3$



$$Pr(X \le 6) = Pr(X=1) + Pr(X=2) + Pr(X=3)$$
$$+ Pr(X=4) + Pr(X=5) + Pr(X=6)$$
$$= 1$$

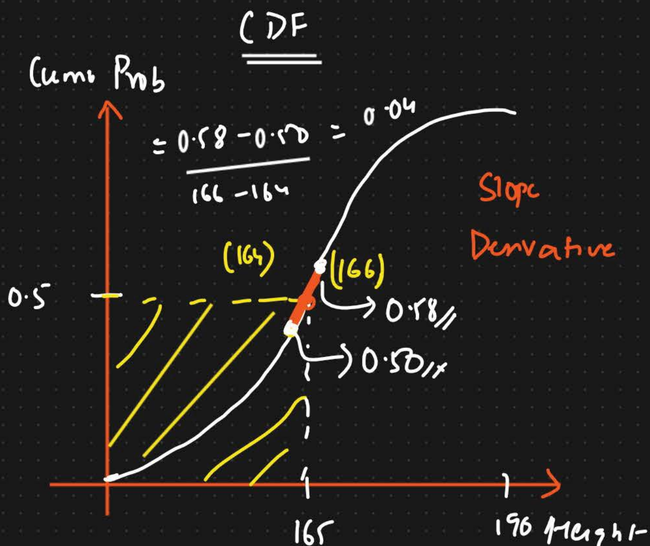
② Probability Density Function (PDF)

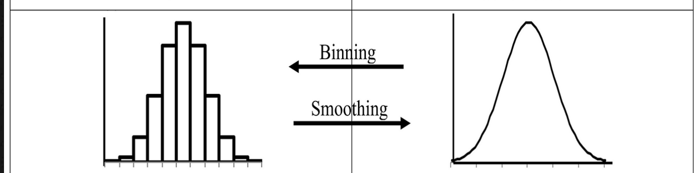
① Continuous Random Variable

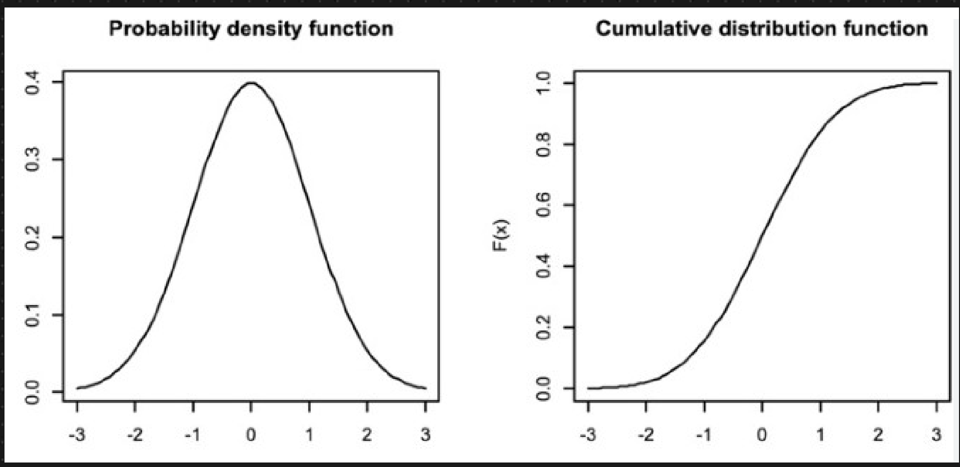


50% of the entire distribution

Probability density
↳ Gradient descent of CDF



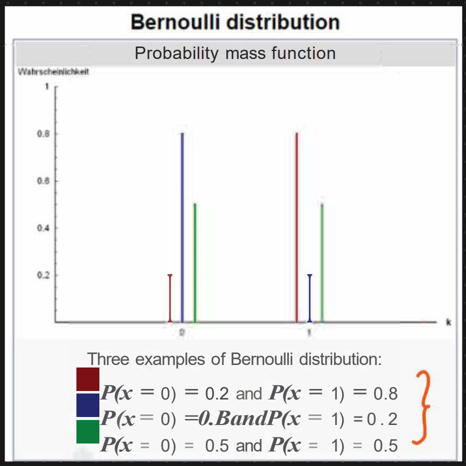
Discrete		Continuous
	Binning ← / Smoothing →	
Probability Mass Function		Probability Density Function
Count,Sum,Proportion		Integration
P(X = x) = f(x)		$P(X=x) = \int f(x).dx$
CMF,PMF = Sum, Difference		CDF,PDF = Integrate, Differentiate



Distribution

① Bernoulli Distribution [Binary outcomes]

In probability theory and statistics, the Bernoulli distribution, named after Swiss mathematician Jacob Bernoulli, is the discrete probability distribution of a random variable which takes the value **1 with probability p** and the **value 0 with probability q=1-p**. Less formally, it can be thought of as a model for the set of possible outcomes of any single experiment that asks a yes–no question. Such questions lead to outcomes that are boolean-valued: a single bit whose value is **success/yes/true/one with probability p and failure/no/false/zero with probability q**.



Outcomes are Binary

Eg: Tossing a Fair Coin $\{0, 1\}$ (H ↓ 0, T ↓ 1)

$P(T) = 0.5 = p$ //

$P(H) = 1 - 0.5 = 1 - p = q$ //

Eg: Whether the Person Pass/Fail

Pmf

pmf vs pdf

pmf ⇓ Discrete random variable

pdf ⇓ Continuous random variable

Pmf $K = 0$ or 1

$$\boxed{P(x = K) = p^{K}(1-p)^{1-K}}$$

⇓

① $P(x = 1) = p^{1}(1-p)^{0}$

$P(x = 1) = p$ //

② $P(x = 0) = p^{0}(1-p)^{1-0}$

$= (1-p) = q$ //.

Simplified way of PMF

$$\text{PMF} = \begin{cases} q = 1-p & \text{if } K = 0 \\ p & \text{if } K = 1 \end{cases}$$

(2) Mean, Variance And Standard Deviation

Mean

$K = 1$ or 0

$$E(K) = \sum_{i=1}^{K} K \cdot P(K)$$

$P(K=1) = 0.6 = P$

$P(K=0) = 1 - 0.6 = q = 0.4$

$$= 1 * 0.6 + 0 * 0.4$$

$$= 0.6 = P$$

Median of Bernoulli Distribution

$$\text{Median} \begin{cases} 0 & \text{if } p < 1/2 \\ [0,1] & \text{if } p = 1/2 \\ 1 & \text{if } p > 1/2 \end{cases}$$

$p = 0.5 \quad q = 0.5$

Variance & Std

$$\text{Variance} = p(1-p) = pq$$

$$\text{Std} = \sqrt{pq}.$$

Practical Applications

1. **Website Click-through Rate (CTR)**: For an online advertisement, each view can result in a click (success) or no click (failure). The probability of a click can be estimated using the Bernoulli distribution.
2. Product Success Rate: When launching a new product, each customer's decision to purchase (or not purchase) can be modeled as a Bernoulli trial, aiding in estimating the product's market success probability.
3. Email Campaigns: For email marketing campaigns, whether an email recipient opens the email or not can be modeled as a Bernoulli trial, helping to measure and optimize open rates.

The Bernoulli distribution's simplicity and versatility make it a fundamental tool for modeling binary outcomes and conducting probabilistic analysis in numerous domains.

Distribution

① Binomial Distributions

In probability theory and statistics, the binomial distribution with parameters n and p is the discrete probability distribution of the number of successes in a sequence of n independent experiments, each asking a yes–no question, and each with its own Boolean-valued outcome: success (with probability p) or failure (with probability q=1-p). A single success/failure experiment is also called a Bernoulli trial or Bernoulli experiment, and a sequence of outcomes is called a Bernoulli process; for a single trial, i.e., n = 1, the binomial distribution is a Bernoulli distribution.

Eg: Tossing a Coin {Bernoulli Distribution}

$P(H) = 0.5 = 1-p \rightarrow 0$ $\qquad$ $P(T) = 0.5 = p \rightarrow 1$

$\boxed{n = 10}$

Eg: Tossing a Coin for 10 times

1st time	2nd	3rd	4th
$P(T) = p$	p	p	p
$P(H) = 1-p$	1-p	1-p	1-p

⇓

Binomial Distribution

Parameters

$n \in \{0, 1, 2, 3, \ldots\} \rightarrow$ Number of Trials or Experiment

$p \in [0, 1] \rightarrow$ Success probability for each trial

$q = 1-p$

② PMF $\quad P(k) = {}^{n}C_{k}\, p^{k} (1-p)^{n-k}$

$K \in \{0, 1, 2, \ldots n\} \rightarrow$ number of success.

3 Mean of Binomial distribution

$$\text{Mean} = np$$

4 Variance And Std

$$\text{Variance} = npq$$

$$\text{Std} = \sqrt{npq}$$

Defect Testing: Determine the probability of finding a certain number of defective items in a batch.

Examples of Practical Applications

1. Email Campaigns: Determine the probability that a certain number of recipients out of a list of 1000 will open an email, given an open rate of 20%.
2. Call Centers: Model the number of successful customer support calls out of a daily quota, given the probability of successfully resolving an issue on each call.
3. Insurance: Estimate the number of claims to be filed out of a portfolio of policies, each with a certain probability of a claim being made.

Summary

The binomial distribution is a versatile tool for modeling binary outcomes in a fixed number of trials. It is particularly useful for quality control, clinical trials, customer surveys, risk analysis, sports analytics, and educational assessments. By understanding and applying the binomial distribution, organizations can make informed decisions based on probabilistic outcomes and expected frequencies of events.

Distribution

Poisson Distribution

① Discrete Distribution (pmf)

② Describes the number of events occurring in a fixed time interval

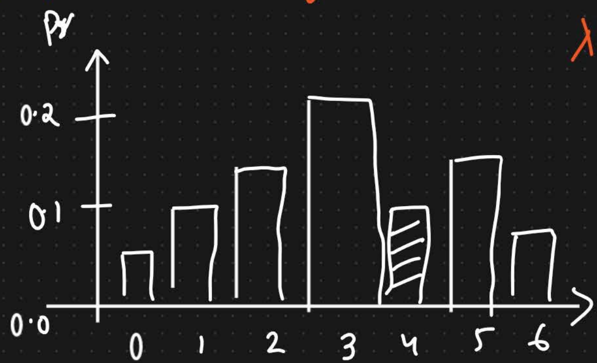
Eg: No. of people visiting hospital every hour

No. of people visiting banks every hour

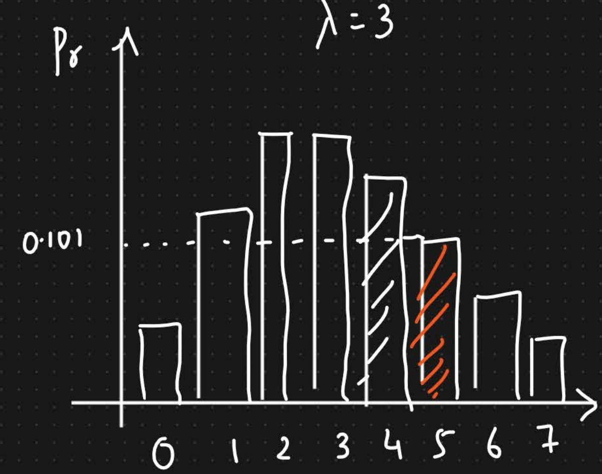
No. of people visiting airport every hour

$\lambda = 3 \Rightarrow$ Expected No. of event to occur at every time interval

pmf



PMF



$$P(X=5) = \frac{e^{-\lambda} \lambda^{x}}{x!}$$

$$= \frac{e^{-3} 3^{5}}{5!} = 0.101$$

$$= 10.1\%$$

$P(x=4) + Pr(x=5) = \_\_\_$

Mean And Variance

λ = Expected No. of events to occur at every time interval

Mean $\Rightarrow E(x) = \mu = \lambda * t$

t = Time Interval

Variance $\Rightarrow E(x) = \mu = \lambda * t$

Examples of Practical Applications

1. Traffic Flow Analysis: Predict the number of cars passing through a toll booth in an hour, given the average rate of arrivals.
2. Sports Analytics: Model the number of goals scored in a soccer match. For instance, the number of goals scored by a team in a match can be described using a Poisson distribution if the average number of goals per match is known.
3. Event Planning: Estimate the number of attendees at an event based on historical data of attendance rates.
4.

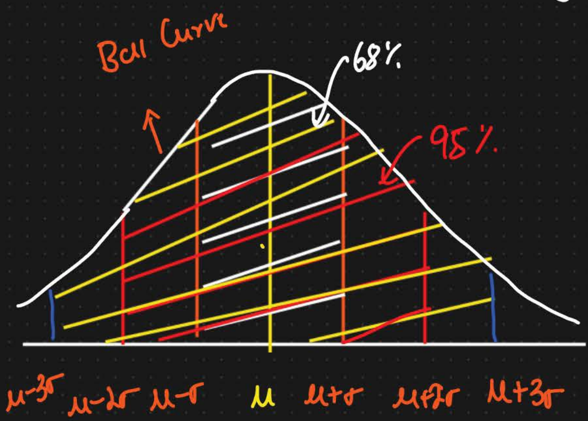
Summary

The Poisson distribution is widely used for modeling the number of events occurring within a fixed interval of time or space. Its applications are diverse, spanning queueing theory, healthcare, telecommunications, manufacturing, finance, retail, and more. By using the Poisson distribution, organizations can effectively predict and manage events and processes that follow a random but average rate of occurrence.

Distribution

① Normal / Gaussian Distribution

$X = \{ \quad \}$



⇒ Symmetric Distribution

68 95% 99.7

$X = \{ . \; . \; . \; . \; . \; . \; . \; . \; . \}$

Empirical Rule [3 Sigma Rule]

68 – 95 – 99.7 %

Q-Q plot ⇒ Whether a Distribution is Gaussian / Normal Distribution

Probability

$$Pr(\mu-\sigma \leq X \leq \mu+\sigma) \approx 68\%$$

$$Pr(\mu-2\sigma \leq X \leq \mu+2\sigma) \approx 95\%$$

$$Pr(\mu-3\sigma \leq X \leq \mu+3\sigma) \approx 99.7\%$$

Eg: Weight, Height, IRIS DATASET

Examples of Practical Applications

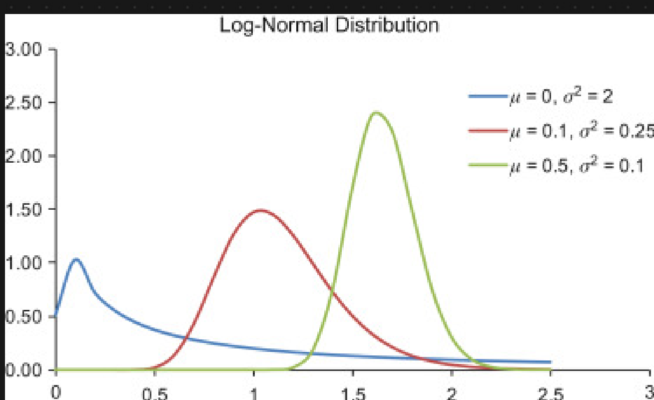
1. Stock Market Returns: Daily returns of stock prices are often modeled as a normal distribution to analyze market behavior and make investment decisions.
2. Climate Data: Temperature and precipitation data often follow a normal distribution, helping meteorologists understand climate patterns and make forecasts.
3. Product Lifetimes: The lifespan of products can be modeled using the normal distribution to predict failure rates and warranty claims.
4. Population Studies: Characteristics such as IQ scores, standardized test results, and other human traits are analyzed using the normal distribution for understanding population health and education trends.

Summary

The normal distribution is essential in many fields due to its mathematical properties and the Central Limit Theorem. Its applications span business, finance, healthcare, education, engineering, and data science, providing a framework for analyzing and interpreting data that follows a bell-shaped curve. Understanding the normal distribution allows for effective decision-making and accurate predictions in various domains.

Log Normal Distributions

A log-normal distribution is a continuous distribution of random variable y whose natural logarithm is normally distributed. For example, if random variable y = exp { y } has log-normal distribution then x = log (y) has normal distribution.



Log-Normal Distribution Statistics[1]

Notation	$\ln N(\mu, \sigma^2)$
	$-\infty < \mu < \infty$
Parameter	$\sigma^2 > 0$
Distribution	$x > 0$
Pdf	$\frac{1}{\sqrt{2\pi}\sigma x}\exp\left\{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right\}$
Cdf	$\frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\ln(x-\mu)}{\sigma}\right)\right]$
Mean	$e^{\left(\mu + \frac{1}{2}\sigma^2\right)}$
Variance	$\left(e^{\sigma^2}-1\right)e^{2\mu+\sigma^2}$
Skewness	$\left(e^{\sigma^2}+2\right)\sqrt{\left(e^{\sigma^2}-1\right)}$
Kurtosis	$e^{4\sigma^2}+2e^{3\sigma^2}+3e^{2\sigma^2}-6$

Examples of Practical Applications

1. Project Management: The completion times of tasks in project management, especially when there are many multiplicative factors, can be modeled using a log-normal distribution. This helps in estimating project timelines more accurately.
2. Insurance: Modeling claim sizes in insurance. Large claims are rare but have significant impacts, which aligns with the properties of the log-normal distribution.
3. E-Commerce: The distribution of purchase amounts by customers in an online store can be log-normally distributed, helping in customer segmentation and targeted marketing.

Summary

The log-normal distribution is a versatile tool for modeling positively skewed data across various fields. Its applicability to phenomena that are the result of multiplicative processes makes it particularly useful in finance, engineering, environmental studies, biology, and social sciences. Understanding and applying the log-normal distribution allows for better analysis and interpretation of real-world data where the normal distribution may not be appropriate due to the presence of skewness and non-negativity constraints.