An overview of Distributions

A distribution shows the possible values a random variable can take and how frequently they occur.

Important Notation for Distributions:

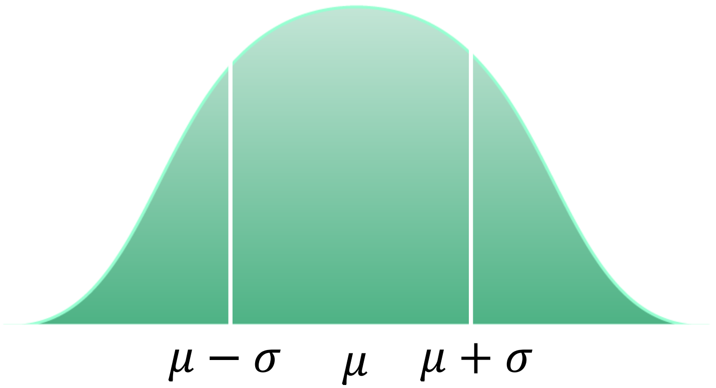
Y actual outcome

y one of the possible outcomes

$P(Y = y)$ is equivalent to $p(y)$.

We call a function that assigns a probability to each distinct outcome in the sample space, a **probability function.**

	Population	Sample
Mean	μ	$\bar{x}$
Variance	σ^2	s^2
Standard Deviation	σ	s



Types of Distributions

Certain distributions share characteristics, so we separate them into **types**. The well-defined types of distributions we often deal with have elegant statistics. We distinguish between two big types of distributions based on the type of the possible values for the variable – discrete and continuous.

Discrete

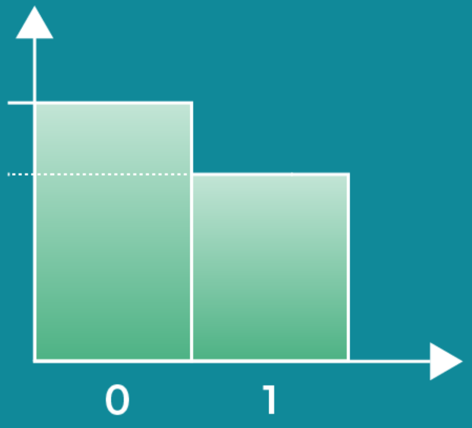
- Have a finite number of outcomes.
- Use formulas we already talked about.
- Can add up individual values to determine probability of an interval.
- Can be expressed with a table, graph or a piece-wise function.
- Expected Values might be unattainable.
- Graph consists of bars lined up one after the other.

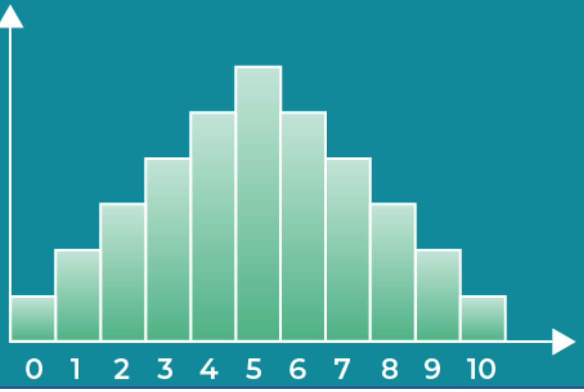
Continuous

- Have infinitely many consecutive possible values.
- Use new formulas for attaining the probability of specific values and intervals.
- Cannot add up the individual values that make up an interval because there are **infinitely many** of them.
- Can be expressed with a graph or a continuous function.
- Graph consists of a smooth curve.

Discrete Distributions

Discrete Distributions have finitely many different possible outcomes. They possess several key characteristics which separate them from continuous ones.





Key characteristics of discrete distribution

- Have a finite number of outcomes.
- Use formulas we already talked about.
- Can add up individual values to determine probability of an interval.
- Can be expressed with a table, graph or a piece-wise function.
- Expected Values might be unattainable.
- Graph consists of bars lined up one after the other.
- $\boldsymbol{P(Y \leq y) = P(Y < y + 1)}$

Examples of Discrete Distributions:

- Discrete Uniform Distribution
- Bernoulli Distribution
- Binomial Distribution
- Poisson Distribution

Uniform Distribution

A distribution where all the outcomes are equally likely is called a **Uniform Distribution**.

Notation:

- $Y \sim U(a, b)$
- * alternatively, if the values are categorical, we simply indicate the number of categories, like so: $Y \sim U(a)$

Key characteristics

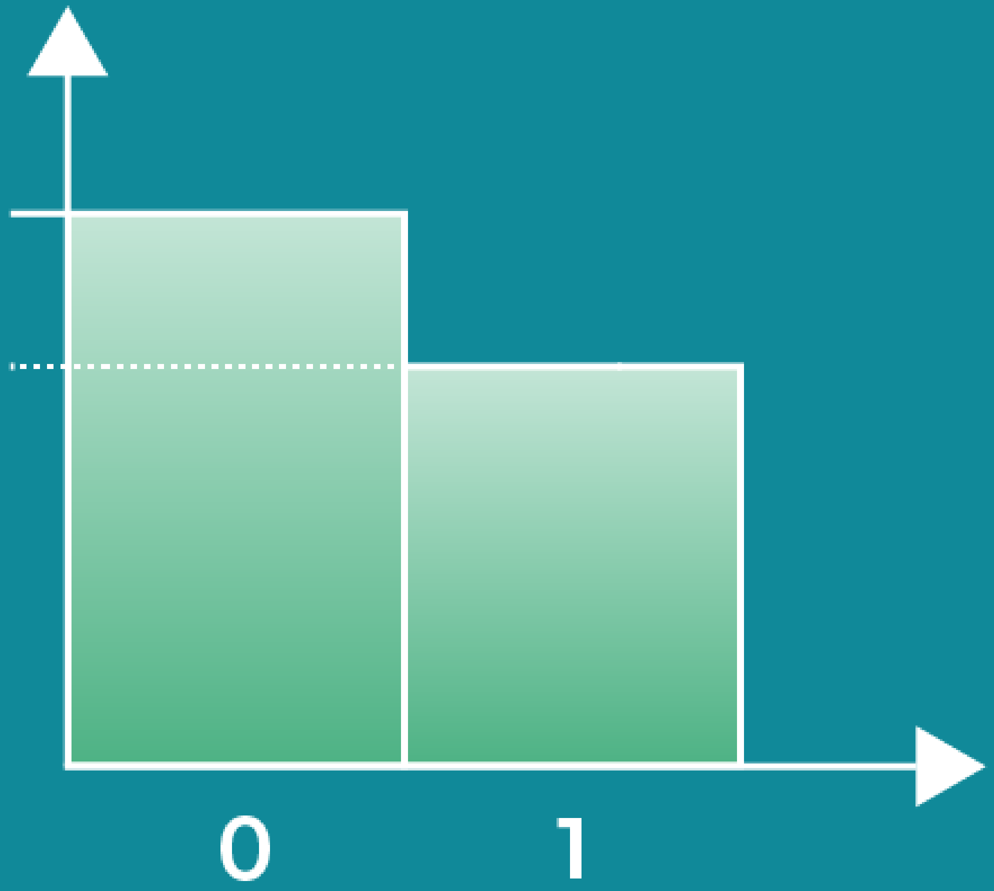
- All outcomes are equally likely.
- All the bars on the graph are equally tall.
- The expected value and variance have no predictive power.

Example and uses:

- Outcomes of rolling a single die.
- Often used in shuffling algorithms due to its fairness.

Bernoulli Distribution

A distribution consisting of a single trial and only two possible outcomes – success or failure is called a **Bernoulli Distribution.**



Notation:

- $Y \sim Bern(p)$

Key characteristics

- One trial.
- Two possible outcomes.
- $E(Y) = p$
- $Var(Y) = p \times (1 - p)$

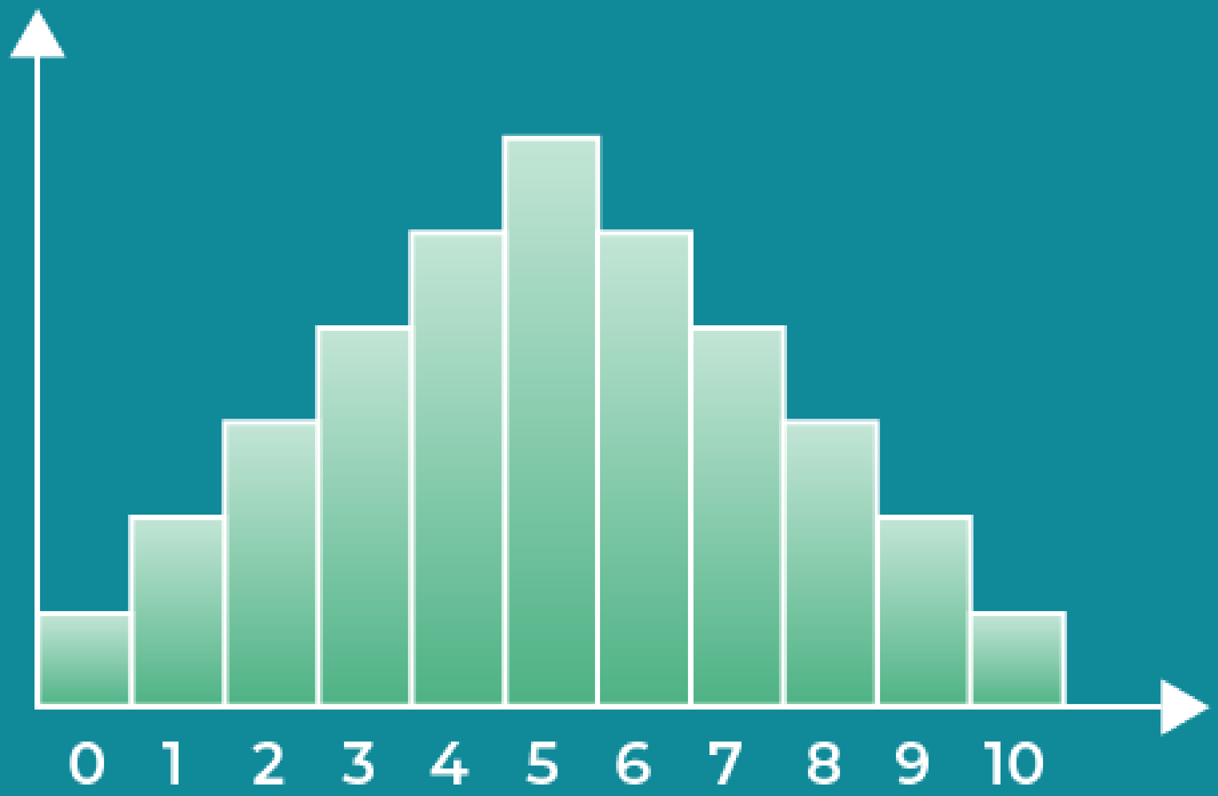
Example and uses:

- Guessing a single True/False question.
- Often used in when trying to determine what we expect to get out a single trial of an experiment.

365√DataScience

Binomial Distribution

A sequence of identical Bernoulli events is called Binomial and follows a **Binomial Distribution**.



Notation:

- $Y \sim B(n, p)$

Key characteristics

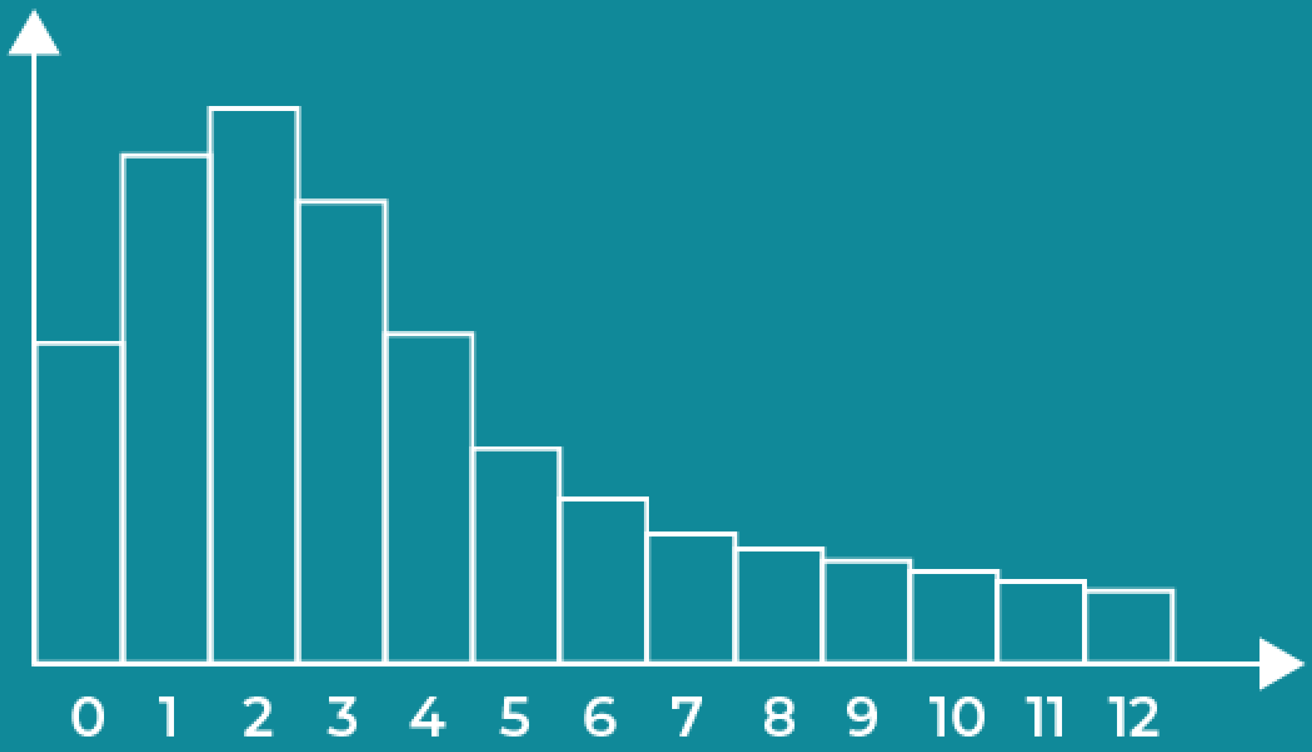
- Measures the frequency of occurrence of one of the possible outcomes over the n trials.
- $P(Y = y) = C(y, n) \times p^y \times (1 - p)^{n-y}$
- $E(Y) = n \times p$
- $Var(Y) = n \times p \times (1 - p)$

Example and uses:

- Determining how many times we expect to get a heads if we flip a coin 10 times.
- Often used when trying to predict how likely an event is to occur over a series of trials.

Poisson Distribution

When we want to know the likelihood of a certain event occurring over a given interval of time or distance we use a **Poisson Distribution**.



Notation:

- $Y \sim Po(\lambda)$

Key characteristics

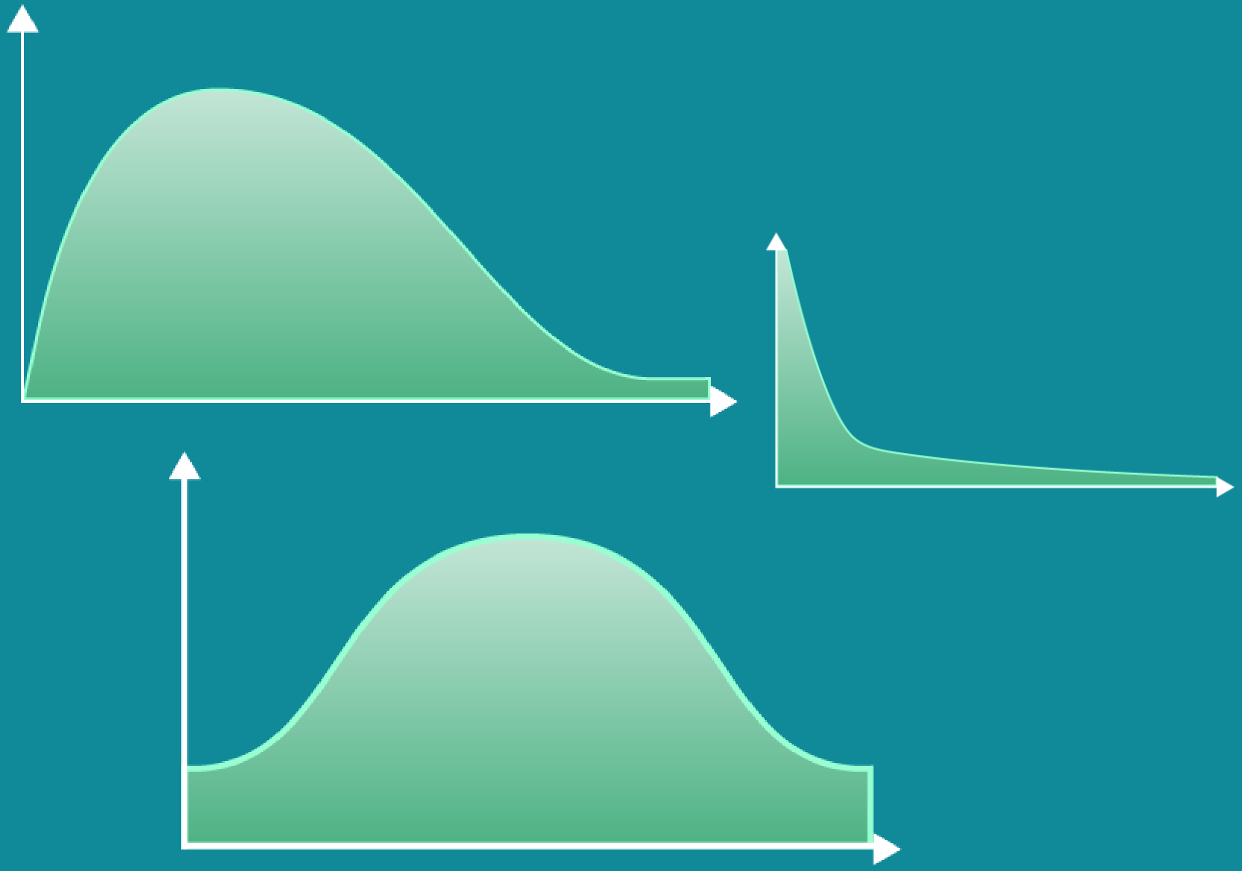
- Measures the frequency over an interval of time or distance. (Only non-negative values.)
- $P(Y = y) = \frac{\lambda^{y} e^{-\lambda}}{y!}$
- $E(Y) = \lambda$
- $Var(Y) = \lambda$

Example and uses:

- Used to determine how likely a specific outcome is, knowing how often the event **usually** occurs.
- Often incorporated in marketing analysis to determine whether above average visits are out of the ordinary or not.

Continuous Distributions

If the possible values a random variable can take are a sequence of infinitely many consecutive values, we are dealing with a **continuous distribution**.

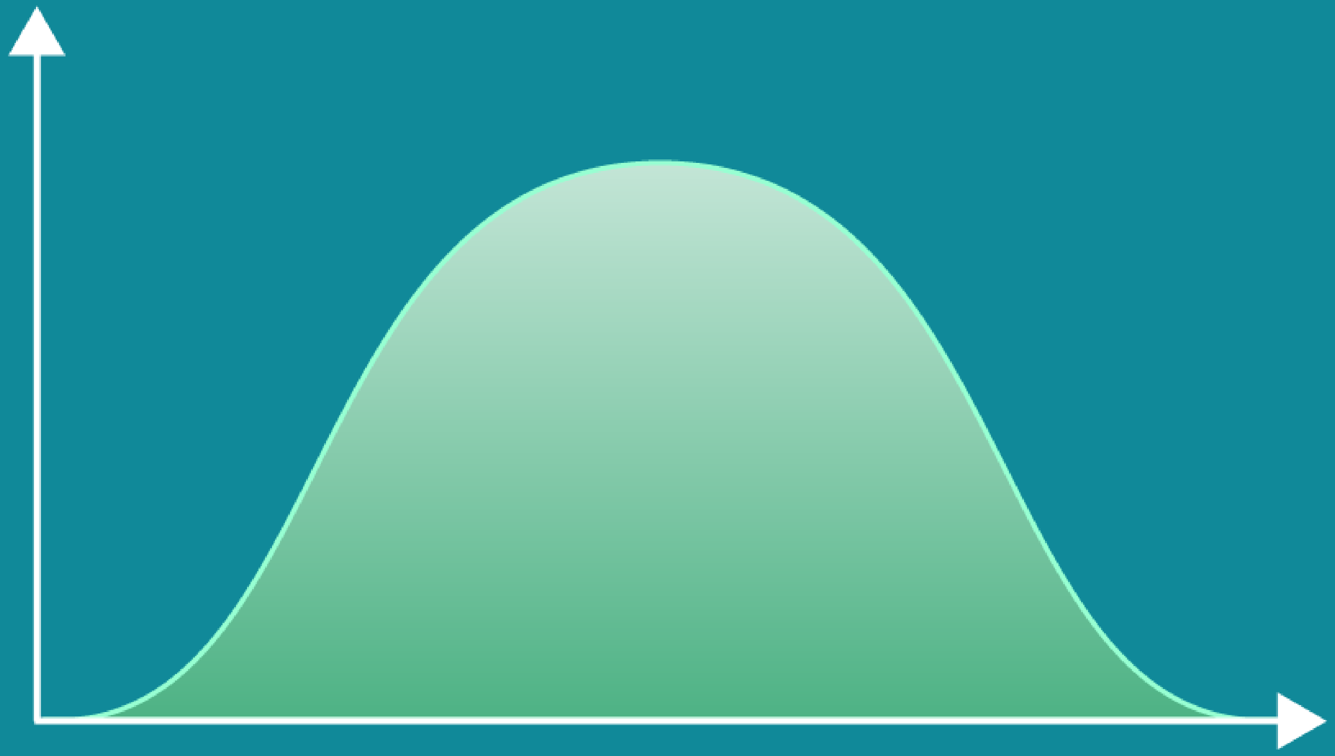


Key characteristics

- Have infinitely many consecutive possible values.
- Cannot add up the individual values that make up an interval because there are **infinitely many** of them.
- Can be expressed with a graph or a continuous function. Cannot use a table, be
- Graph consists of a smooth curve.
- To calculate the likelihood of an interval, we need integrals.
- They have important CDFs.
- $\boldsymbol{P(Y = y)} = 0$ for any individual value y.
- $\boldsymbol{P(Y < y) = P(Y \leq y)}$

Normal Distribution

A **Normal Distribution** represents a distribution that most natural events follow.



Notation:

- $Y \sim N(\mu, \sigma^2)$

Key characteristics

- Its graph is bell-shaped curve, symmetric and has thin tails.
- $E(Y) = \mu$
- $Var(Y) = \sigma^2$
- 68% of all its values should fall in the interval:
    - $(\mu - \sigma,\ \mu + \sigma)$

Example and uses:

- Often observed in the size of animals in the wilderness.
- Could be standardized to use the Z-table.

Standardizing a Normal Distribution

To standardize any normal distribution we need to transform it so that the mean is 0 and the variance and standard deviation are 1.

Ensures mean is 0.

Using a transformation to create a new random variable z.

$$z = \frac{y - \mu}{\sigma}$$

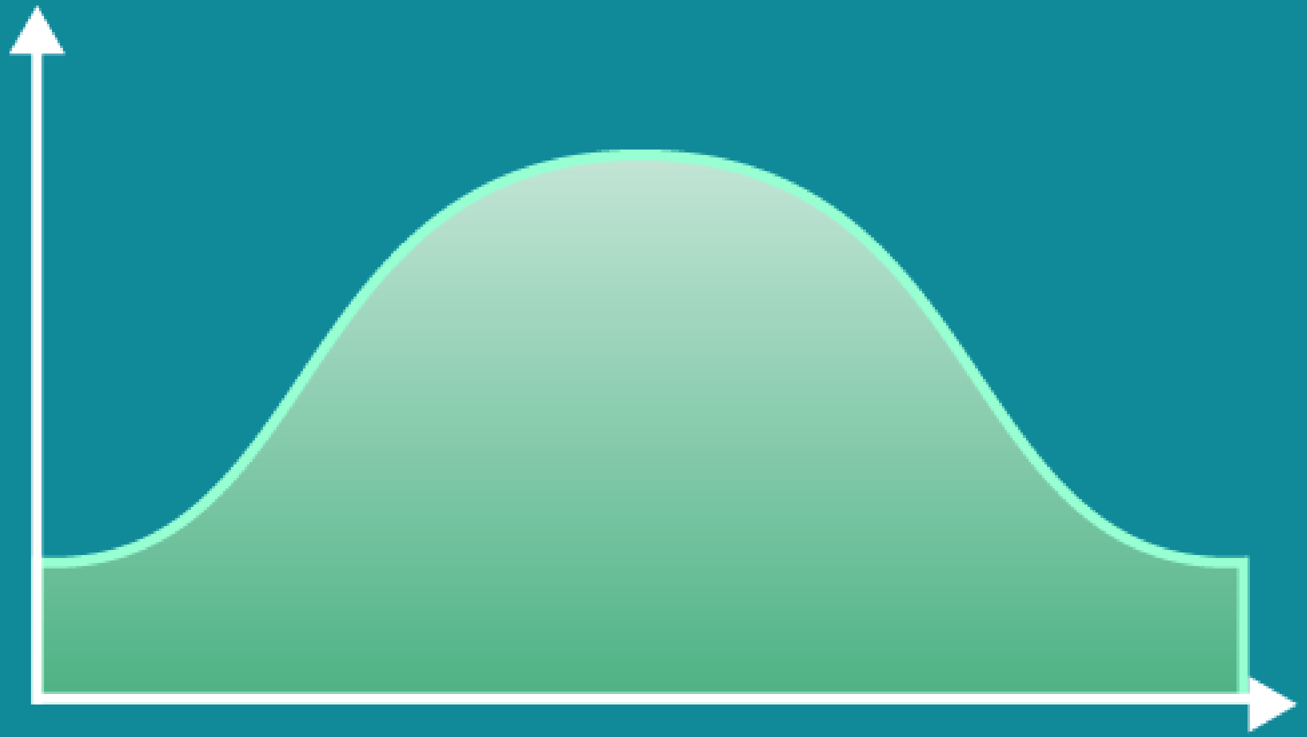
Ensures standard deviation is 1.

Importance of the Standard Normal Distribution.

- The new variable z, represents how many standard deviations away from the mean, each corresponding value is.
- We can transform any Normal Distribution into a Standard Normal Distribution using the transformation shown above.
- Convenient to use because of a table of known values for its CDF, called the Z-score table, or simply the Z-table.

Students' T Distribution

A **Normal Distribution** represents a small sample size approximation of a Normal Distribution.



Notation:

- $Y \sim t\,(k)$

Key characteristics

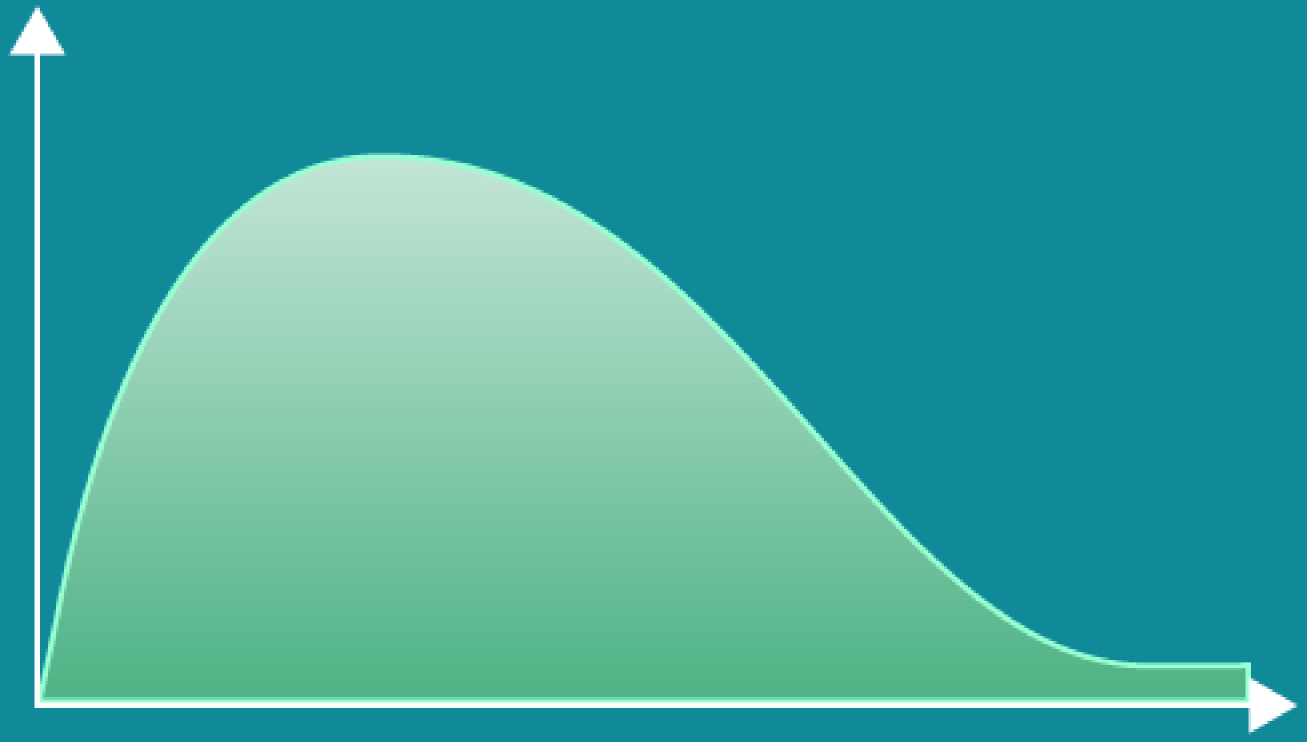
- A small sample size approximation of a Normal Distribution.
- Its graph is bell-shaped curve, symmetric, but has **fat** tails.
- Accounts for extreme values better than the Normal Distribution.
- If k>1: $E(Y) = \mu$ and $Var(Y) = s^2 \times \frac{k}{k-2}$

Example and uses:

- Often used in analysis when examining a small sample of data that usually follows a Normal Distribution.

Chi-Squared Distribution

A **Chi-Squared** distribution is often used.



Notation:

- $Y \sim \chi^2(k)$

Key characteristics

- Its graph is asymmetric and skewed to the right.
- $E(Y) = k$
- $Var(Y) = 2k$
- The Chi-Squared distribution is the square of the t-distribution.

Example and uses:

- Often used to test goodness of fit.
- Contains a table of known values for its CDF called the χ^2-table. The only difference is the table shows what part of the table

Exponential Distribution

The **Exponential Distribution** is usually observed in events which significantly change early on.

Notation:

- $Y \sim Exp\ (\lambda)$

Key characteristics

- Both the PDF and the CDF plateau after a certain point.
- $E(Y) = \frac{1}{\lambda}$
- $Var(Y) = \frac{1}{\lambda^2}$
- We often use the natural logarithm to transform the values of such distributions since we do not have a table of known values like the Normal or Chi-Squared.

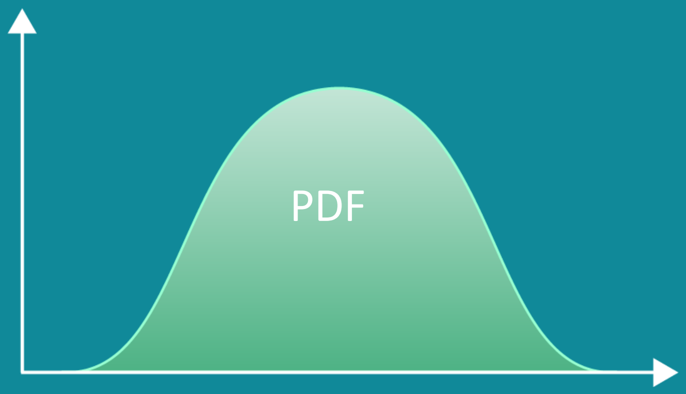
Example and uses:

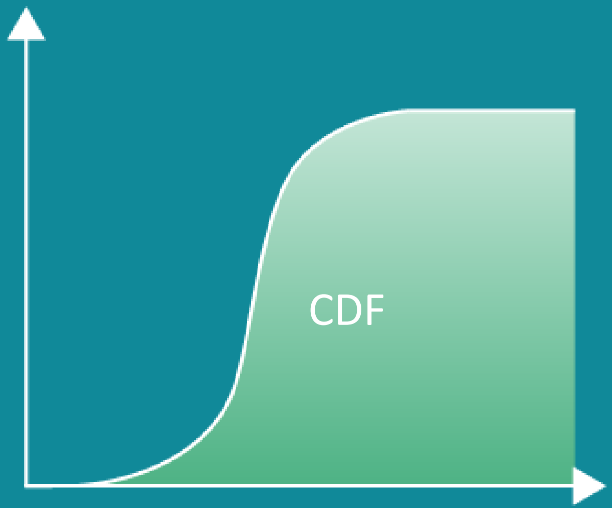
- Often used with dynamically changing variables, like online website traffic or radioactive decay.

365√DataScience

Logistic Distribution

The **Continuous Logistic Distribution** is observed when trying to determine how continuous variable inputs can affect the probability of a binary outcome.





Notation:

- $Y \sim Logistic\ (\mu, s)$

Key characteristics.

- $E(Y) = \mu$
- $Var(Y) = \frac{s^2 \times \pi^2}{3}$
- The CDF picks up when we reach values near the mean.
- The smaller the scale parameter, the quicker it reaches values close to 1.

Example and uses:

- Often used in sports to anticipate how a player's or team's performance can determine the outcome of the match.