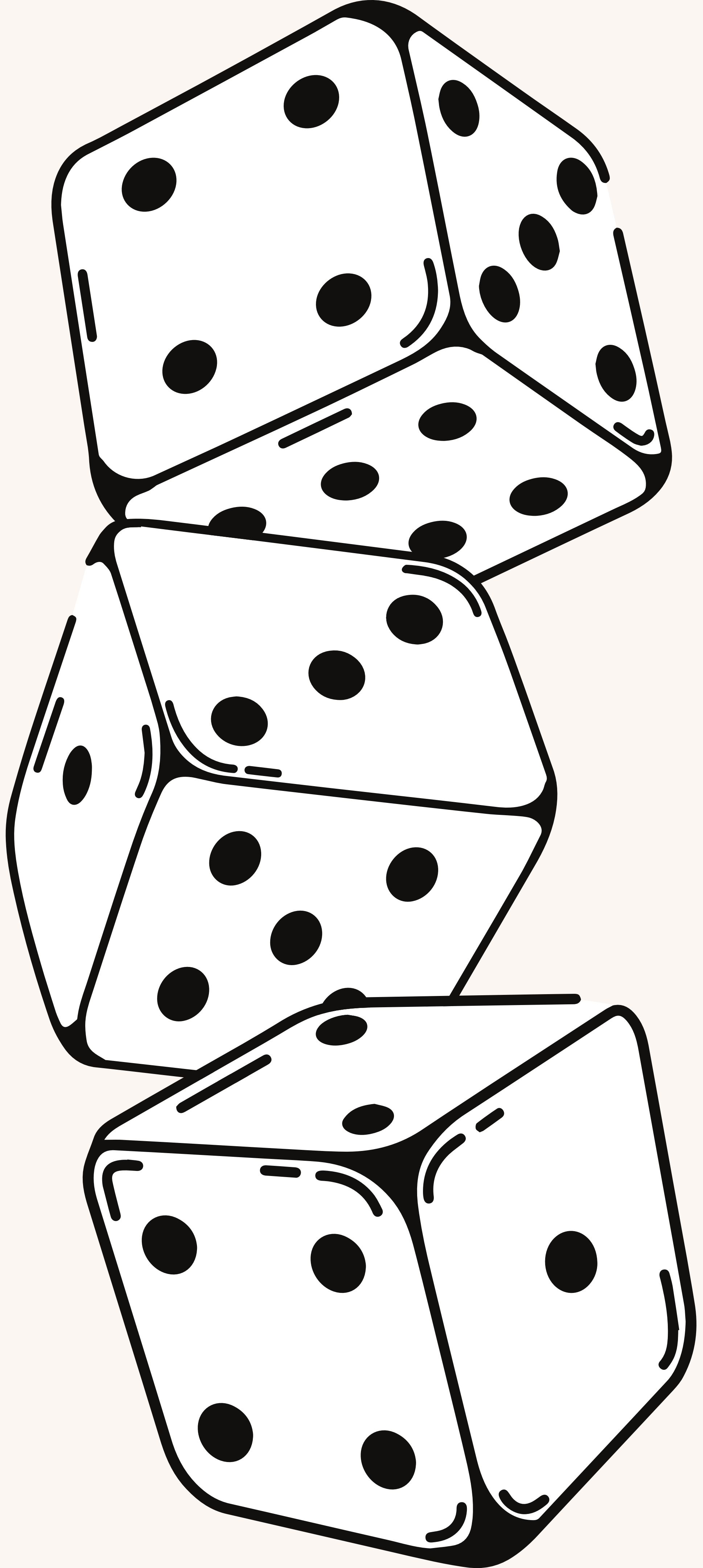
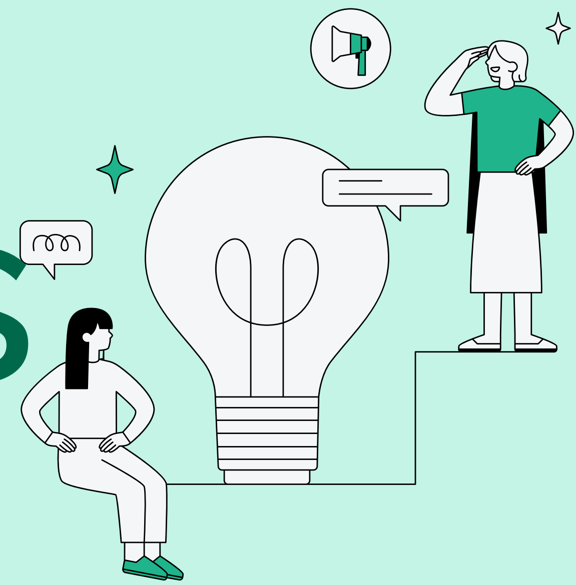
Probability course summarization

Presented by Rawan Hatem

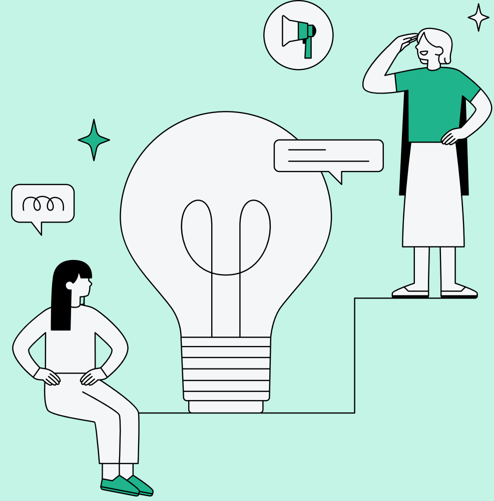


Introduction to results analysis



- Probability distribution provides possible values a variable can take and how frequently they occur.
- Upper-case Y represents actual outcome of an event
- Lowercase y represents one of the possible outcomes
- p(y) or p(y=y) represents likelihood of reaching a particular outcome
- Mean of a distribution is denoted as μ, variance as σ^2
- Finite vs Infinite number of possible outcomes and corresponding methods of determining probabilities:
- Finite: Probability Frequency Distribution, divided by total number of elements in the sample space
- Infinite: Constructing a probability distribution becomes impossible, uses different methods like continuous distributions

Types of Probability distribution

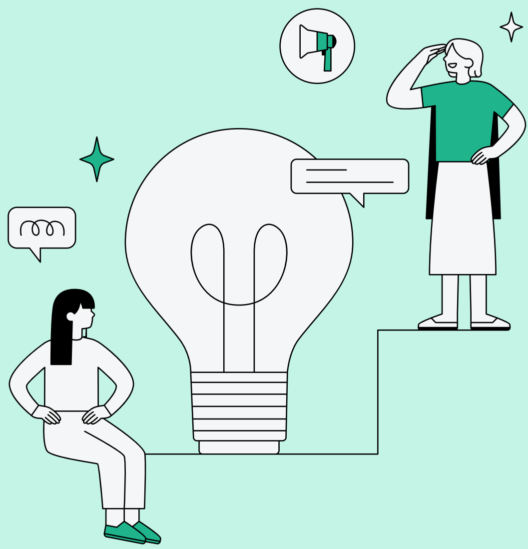


Two types of distributions are:

a. Discrete distributions: used for events with a finite number of outcomes, such as rolling a die or picking a card.
b. Continuous distributions: used for events with infinitely many outcomes, such as recording time and distance.

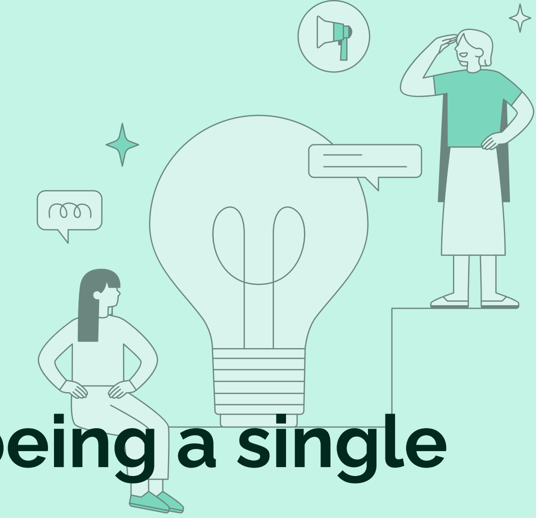
- Discrete distributions include:
  - Discrete Uniform Distribution: used for events with equally likely outcomes, like drawing cards from a deck or flipping a coin.
  - Bernoulli Distribution: used for events with two possible outcomes, like selecting a captain from a group of students.
  - Binomial Distribution: used when repeating a Bernoulli trial several times.
  - Poisson Distribution: used to test event frequency for a given interval.
- Continuous distributions:
  - Normal Distribution: represents the outcomes of many natural events, with the majority of data points falling near the mean.
  - Student's-t Distribution: a small sample approximation of a normal distribution, used when data is limited.
  - Chi-squared Distribution: an asymmetric distribution with non-negative values, used for hypothesis testing.
  - Exponential Distribution: used for events that rapidly change over time, such as online news article hits.
  - Logistic Distribution: useful in forecast analysis, determining a cut-off point for successful outcomes.

Discrete Uniform Distribution



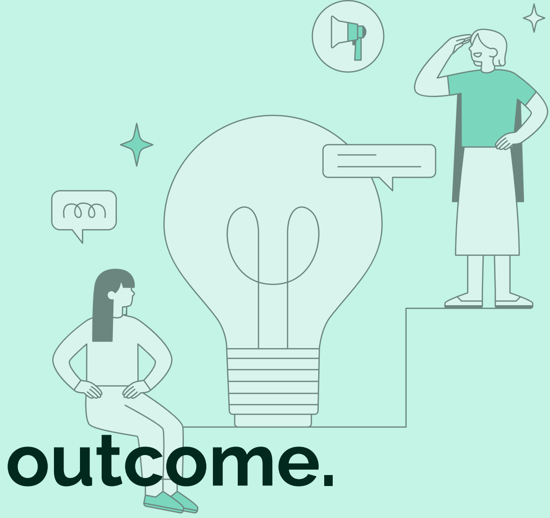
- Discrete distributions involve events with finitely many distinct outcomes.
- Probability distributions for discrete events can be expressed using a table, graph, or formula.
- probability of an interval is often more significant in discrete distributions.
- To calculate the probability of an interval in a discrete distribution, probabilities for all values within that range are added up.
- A peculiarity of discrete events is that the probability of $y \leq y$ equals the probability of $y < y + 1$.
- The uniform distribution is defined by the letter u and the range of values in the dataset.
- Events following the uniform distribution have all outcomes with equal probability.
- An example of a uniform distribution is rolling a single standard six-sided die, where there is an equal chance of getting any value from 1 to 6.
- The uniform distribution has a predictable graph with equally tall bars, but its expected value is not useful in predicting outcomes.
- The mean and variance of the uniform distribution are uninterpretable and possess no predictive power.

Binomial Distribution



- Binomial distributions and Bernoulli distributions are related, with a Bernoulli distribution being a single trial of a binomial distribution.
- Notation for a binomial distribution: "b" followed by the number of trials and the probability of success in each trial.
- The expected value of a binomial distribution represents the number of times a specific outcome is expected.
- The graph of a binomial distribution shows the likelihood of attaining a desired outcome a specific number of times.
- The probability function of the binomial distribution is used to find the associated likelihood of getting a given outcome a precise number of times over the course of n trials.
- Probability function for a binomial distribution is the product of the number of combinations of picking y-many elements out of n, times "p" to the power of y and "1 - p" to the power of "n minus y".
- After calculating the expected value, the variance is calculated using the formula "n times p times 1 minus p".
- Knowing the expected value and standard deviation allows for more accurate future forecasts.

Bernoulli Distribution

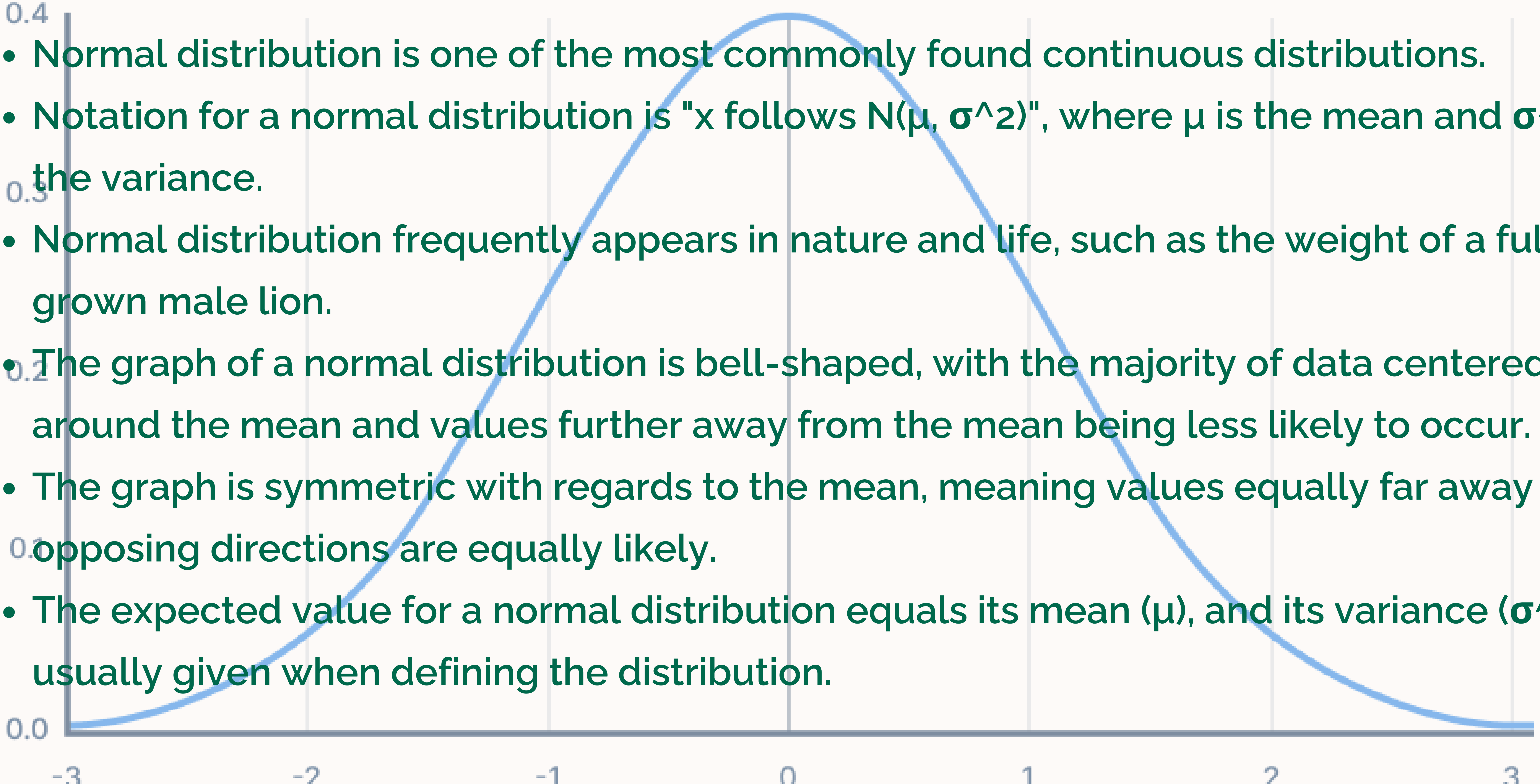


- A Bernoulli distribution is defined as "bern(p)," where "p" is the probability of the preferred outcome.
- Events that follow a Bernoulli distribution have only one trial and two possible outcomes.
- Examples of such events include a coin flip, a single true/false quiz question, or deciding whether to vote for a particular political party.
- When dealing with a Bernoulli distribution, you either have the probabilities of either outcome or past data indicating experimental probability.
- The graph of a Bernoulli distribution is simple, with two bars representing the two possible outcomes.
- The assignment of "0" and "1" to the outcomes is arbitrary, and the expected value will depend on this assignment.
- The expected value is typically denoted as "p" for the higher probability and "1 - p" for the lower probabilit
- The convention is to assign a value of 1 to the event with probability "p" to express the likelihood of the favored event.
- The variance of Bernoulli events is equal to "p * (1 - p)" regardless of the expected value.
- The concepts of variance and standard deviation are demonstrated with an example of an unfair coin, which has a 0.6 probability of getting tails.

Poisson Distribuiton

- The Poisson distribution is denoted as "Po" and has a single value parameter called lambda (λ).
- It focuses on the frequency of event occurrence within a specific interval, rather than the probability of an event.
- The graph of a Poisson distribution plots the number of instances the event occurs in a standard interval of time and the probability for each one, starting from 0 with no upper limit.
- An example is provided of an online course creator who usually gets 4 questions per day but received 7 one day.
- The Poisson distribution's probability function is introduced: $p(y) = (\lambda^y * e^{(-\lambda)}) / y!$
- e is Euler's number or Napier's constant (approximately 2.72)
- A number to the power of negative n is the same as dividing 1 by that number to the power of n
- e is Euler's number or Napier's constant (approximately 2.72)
- A number to the power of negative n is the same as dividing 1 by that number to the power of n
- The expected value of the Poisson distribution is calculated using the formula: the sum of the products of each distinct value in the sample space and its probability. The expected value is equal to lambda (λ).
- The variance of the Poisson distribution also equals lambda (λ).
- To compute the probability of an interval of a Poisson distribution, the joint probability of all individual elements within it is found, similar to other discrete distributions.

Normal Distribution



- Normal distribution is one of the most commonly found continuous distributions.
- Notation for a normal distribution is "x follows N(μ, **σ**^2)", where μ is the mean and **σ**^2 is the variance.
- Normal distribution frequently appears in nature and life, such as the weight of a full-grown male lion.
- The graph of a normal distribution is bell-shaped, with the majority of data centered around the mean and values further away from the mean being less likely to occur.
- The graph is symmetric with regards to the mean, meaning values equally far away in opposing directions are equally likely.
- The expected value for a normal distribution equals its mean (μ), and its variance (**σ**^2) is usually given when defining the distribution.

Standard Normal Distribution

- A transformation is a way to alter every element of a distribution to get a new distribution with similar characteristics.
- For normal distributions, addition, subtraction, multiplication, and division of every element can be done without changing the type of the distribution.
- Adding a constant to every element of a normal distribution will move the graph that many places to the right.
- Subtracting a number from every element will move the graph to the left.
- Multiplying the function by a constant will widen the graph, and dividing every element by a number will shrink the graph.
- Standardizing is a special kind of transformation that makes the expected value equal to 0 and the variance equal to 1.
- The distribution obtained after standardizing any normal distribution is called a "standard normal distribution."
- To standardize, we need to subtract the mean from every element and then divide every element of the newly obtained distribution by the standard deviation.
- Every element of the non-standardized distribution is represented in the new distribution by the number of standard deviations it is away from the mean.
- Standardizing is useful when we have a normal distribution, but we cannot always anticipate that the data is spread out that way.
- A small sample size approximation of a normal distribution is the student's t-distribution.

Probability of student’s t-Distribution

- The student's t-distribution is represented by the lower-case letter "t" followed by a single parameter in parenthesis called "degrees of freedom."
- It is a small sample size approximation of a normal distribution.
- The average lap times for the entire season of a formula 1 race follow a normal distribution, but the lap times for the first lap of the Monaco Grand Prix would follow a student's t-distribution.
- The curve of the student's t-distribution is bell-shaped and symmetric, but it has fatter tails to accommodate the occurrence of values far away from the mean.
- Apart from the mean and variance, the degrees of freedom must also be defined for the student's t-distribution.
- As long as there are at least 2 degrees of freedom, the expected value of a t-distribution is the mean "mu."
- The variance of the distribution equals the variance of the sample times the number of degrees of freedom divided by degrees of freedom minus two.
- The student's t-distribution is frequently used when conducting statistical analysis, especially for hypothesis testing with limited data.
- A table summarizing the most important values of its cumulative distribution function (CDF) is available.

Sample Space & Events

- "sample space" is the set of all possible outcomes of a random experiment. It is denoted by the symbol S.
- Each outcome in the sample space is called a "sample point" or "elementary event".
- An "event" in probability theory is a subset of the sample space, representing a collection of possible outcomes of the experiment. Events are denoted by capital letters
- Events can be simple (consisting of only one outcome) or compound (consisting of more than one outcome).
- Union of Events ($A \cup B$):
  - The union of two events A and B is the event that either A occurs, or B occurs, or both occur.
- Intersection of Events ($A \cap B$):
- The intersection of two events A and B is the event that both A and B occur.
- Symbolically, $A \cap B$ represents the set of outcomes that are in both A and B.
- Probability of Events:
- Probability is a measure of the likelihood of an event occurring.
- The probability of an event A, denoted by P(A), is a number between 0 and 1, where 0 indicates impossibility and 1 indicates certainty.
- Probability is calculated as the ratio of the number of outcomes favorable to the event to the total number of outcomes in the sample space.

Sample Space & Events

- Counting techniques refer to various methods used to determine the number of possible outcomes or arrangements in a given scenario ,These counting techniques are foundational in solving various combinatorial problems, such as counting the number of arrangements, selections, or possibilities .
- Common counting techniques:
  - Multiplication Principle:
    - The multiplication principle states that if there are n1 ways to perform the first task and n2 ways to perform the second task independently of the first, then there are n1×n2 ways to perform both tasks together.
    - This principle is often applied when dealing with sequences of choices or tasks.
  - Permutations:
    - Permutations refer to the arrangements of objects in a specific order.
    - When arranging n distinct objects into r positions (where $r \le n$), the number of permutations is given by
    - P(n, r)=(n−r)!n! , where n! denotes the factorial of n.
    - Permutations are typically used when order matters.
  - Combinations:
    - Combinations refer to the selections of objects without considering the order.
    - When selecting *r* objects from a set of *n* distinct objects (where $r \le n$), the number of combinations is given by !*C*(*n*, *r*)=*r*!×(*n*−*r*)!*n*!.
    - Combinations are typically used when order does not matter.

Rules

Rules

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cap B`) = P(A) - P(A \cap B)$$

$$P(A` \cap B) = P(B) - P(A \cap B)$$

$$P(A` \cap B`) = P(A \cup B)` = 1 - P(A \cup B)$$

Demorgan’s Law

$$(A \cup B)` = A` \cap B`$$

$$(A \cap B)` = A` \cup B`$$

Conditional Probability

- conditional probability, which is the probability of an event (event A) occurring given that another event (event B) has already occurred.
- A simple way to calculate conditional probability is by finding the ratio of the elements of A that are in B to the total elements of B.
- The formula for calculating conditional probability is: $P(A|B) = P(A \cap B) / P(B)$
- The probability of an event occurring given a condition can be calculated using the formula: $P(A|B) = P(A \cap B) / P(B)$ or by finding the ratio of the elements of A that are in B to the total elements of B.

Bayes' Theorem

- Bayes' Theorem is introduced as a formula to calculate the conditional probability of an event if the reverse conditional probability is known.
- The concept of conditional probability is related to Bayes' theorem, with the probability of an event scaled by the knowledge of the event.
- The importance of Bayes' theorem lies in the situations where all data is not available, and it forms the basis for Bayesian statistics.
- The formula for Bayes' Theorem is: $p(a|b) = p(b|a) * p(a) / p(b)$