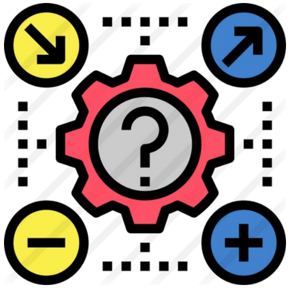
Probability



Tushar B. Kute,
http://tusharkute.com

What is Probability?

- **Probability** is a measure of the likelihood of a random phenomenon or chance behavior.
- Probability describes the long-term proportion with which a certain **outcome** will occur in situations with short-term uncertainty.
- Example:
  - Simulate flipping a coin 100 times. Plot the proportion of heads against the number of flips. Repeat the simulation.

Probability

- Probability deals with experiments that yield random short-term results or outcomes, yet reveal long-term predictability.
- The long-term proportion with which a certain outcome is observed is the probability of that outcome.

Law of large numbers

- As the number of repetitions of a probability experiment increases, the proportion with which a certain outcome is observed gets closer to the probability of the outcome.

Probability and event

- In probability, an **experiment** is any process that can be repeated in which the results are uncertain.
- A **simple event** is any single outcome from a probability experiment. Each simple event is denoted e_i.
- The **sample space, *S*,** of a probability experiment is the collection of all possible simple events. In other words, the sample space is a list of all possible outcomes of a probability experiment.

The event

- An **event** is any collection of outcomes from a probability experiment.
- An event may consist of one or more simple events.
- Events are denoted using capital letters such as *E*.

Example:

- Consider the probability experiment of having two children.

- (a) Identify the simple events of the probability experiment.
- (b) Determine the sample space.
- (c) Define the event E = “have one boy”.

Denoting probability

- The **probability of an event**, denoted *P*(*E*), is the likelihood of that event occurring.

Properties of probabilities

- The probability of any event E, $P(E)$, must be between 0 and 1 inclusive. That is,

  $0 \leq P(E) \leq 1.$

- If an event is **impossible**, the probability of the event is 0.
- If an event is a **certainty,** the probability of the event is 1.
- If $S = \{e_1, e_2, \ldots, e_n\}$, then

  $P(e_1) + P(e_2) + \ldots + P(e_n) = 1.$

Unusual Event

- An **unusual event** is an event that has a low probability of occurring.

Method of probability

- Three methods for determining the probability of an event:

  (1) the classical method

  (2) the empirical method

  (3) the subjective method

Dependence and Independence

- Roughly speaking, we say that two events E and F are dependent if knowing something about whether E happens gives us information about whether F happens (and vice versa). Otherwise they are independent.
- For instance, if we flip a fair coin twice, knowing whether the first flip is Heads gives us no information about whether the second flip is Heads. These events are independent. On the other hand, knowing whether the first flip is Heads certainly gives us information about whether both flips are Tails. (If the first flip is Heads, then definitely it's not the case that both flips are Tails.) These two events are dependent.

Dependence and Independence

- Mathematically, we say that two events E and F are independent if the probability that they both happen is the product of the probabilities that each one happens:

$$P(E, F) = P(E)P(F)$$

- In the example above, the probability of “first flip Heads” is 1/2, and the probability of “both flips Tails” is 1/4, but the probability of “first flip Heads and both flips Tails” is 0.

Conditional Probability

- When two events E and F are independent, then by definition we have:

$$P(E, F) = P(E)P(F)$$

- If they are not necessarily independent (and if the probability of F is not zero), then we define the probability of E “conditional on F” as:

$$P(E \mid F) = P(E, F) / P(F)$$

Conditional Probability

- You should think of this as the probability that E happens, given that we know that F happens.
- We often rewrite this as:

$$P(E, F) = P(E \mid F)P(F)$$

- When E and F are independent, you can check that this gives:

$$P(E \mid F) = P(E)$$

Example:

- One common tricky example involves a family with two (unknown) children.
- If we assume that:
    1. Each child is equally likely to be a boy or a girl
    2. The gender of the second child is independent of the gender of the first child

  then the event “no girls” has probability 1/4, the event “one girl, one boy” has probability 1/2, and the event “two girls” has probability 1/4.

Example:

- Now we can ask what is the probability of the event "both children are girls" (B) conditional on the event "the older child is a girl" (G)? Using the definition of conditional probability:

$$P(B \mid G) = P(B,\ G) / P(G) = P(B) / P(G) = 1/2$$

- since the event B and G ("both children are girls and the older child is a girl") is just the event B. (Once you know that both children are girls, it's necessarily true that the older child is a girl.)

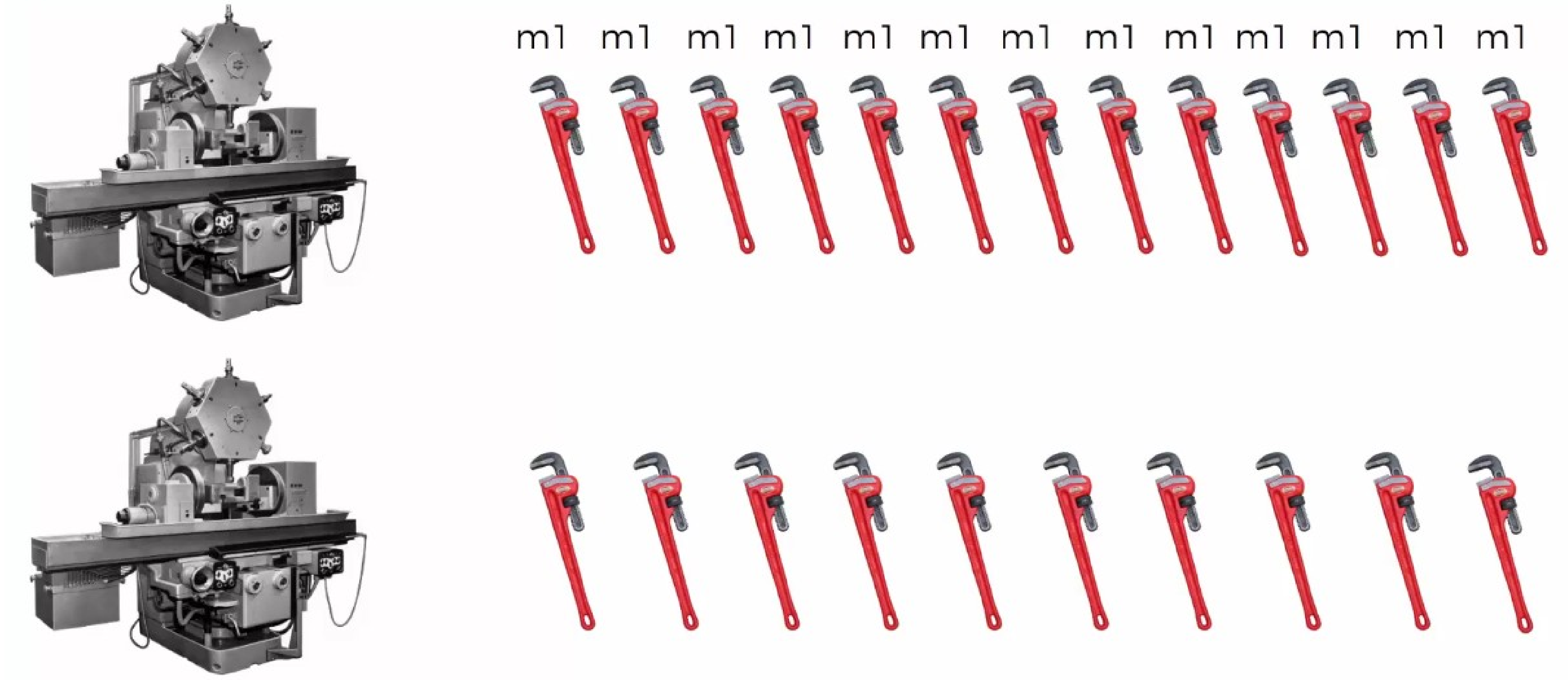
Example:

- We could also ask about the probability of the event "both children are girls" conditional on the event "at least one of the children is a girl" (L). Surprisingly, the answer is different from before!
- As before, the event B and L ("both children are girls and at least one of the children is a girl") is just the event B. This means we have:

$$P(B \mid L) = P(B,\ L) / P(L) = P(B) / P(L) = 1/3$$

- How can this be the case? Well, if all you know is that at least one of the children is a girl, then it is twice as likely that the family has one boy and one girl than that it has both girls.

Bayes Theorem



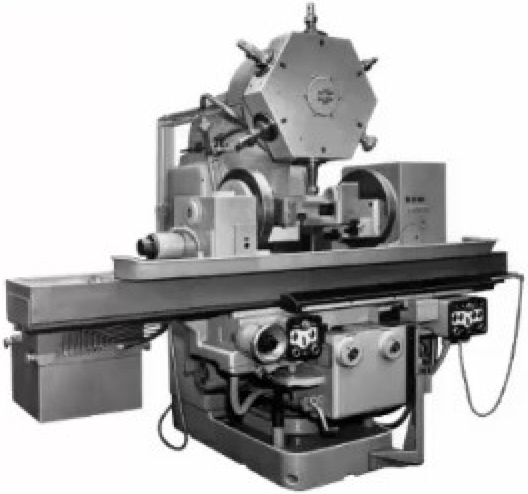
Example Reference: Super Data Science

Bayes Theorem

Defective Spanners

Bayes Theorem

What's the probability?



Bayes Theorem

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Bayes Theorem

Mach1: 30 wrenches / hr -> P(Mach1) = 30/50 = 0.6
Mach2: 20 wrenches / hr -> P(Mach2) = 20/50 = 0.4

Out of all produced parts:
We can SEE that 1% are defective -> P(Defect) = 1%

Out of all defective parts:
We can SEE that 50% came from mach1 -> P(Mach1 | Defect) = 50%
And 50% came from mach2 -> P(Mach2 | Defect) = 50%

Question:
What is the probability that a part produced by mach2 is defective = ? -> P(Defect | Mach2) = ?

Bayes Theorem

$$P(Defect \mid Mach2) = \frac{P(Mach2 \mid Defect) * P(Defect)}{P(Mach2)}$$

$$P(Defect \mid Mach2) = \frac{0.5 * 0.01}{0.4} = 0.0125$$

That’s intuitive

$$P(Defect \mid Mach2) = \frac{P(Mach2 \mid Defect) * P(Defect)}{P(Mach2)} = 1.25\%$$

Let’s look at an example:

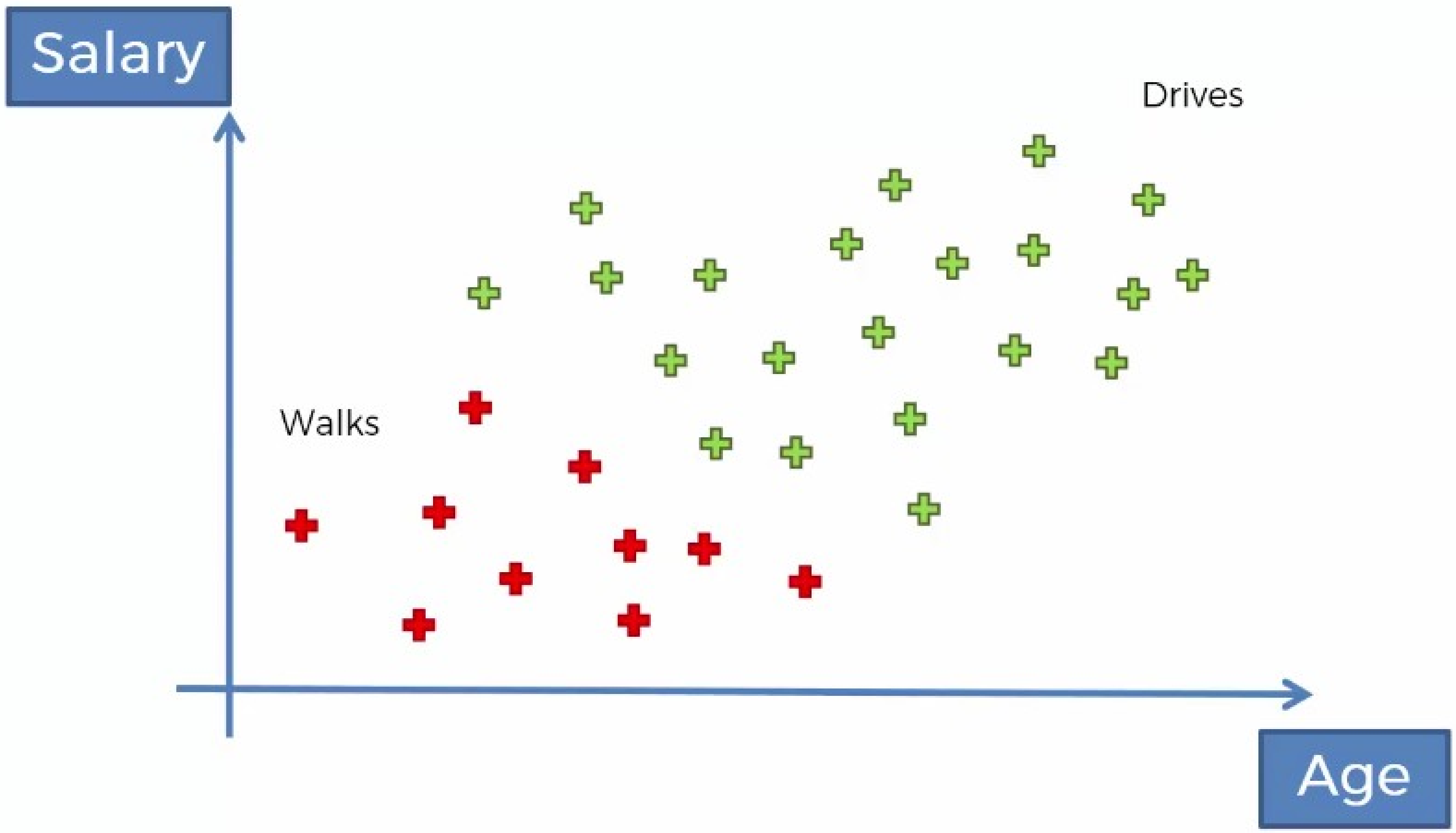
- 1000 wrenches
- 400 came from Mach2
- 1% have a defect = 10
- of them 50% came from Mach2 = 5
- % defective parts from Mach2 = 5/400 = 1.25%

Exercise

Quick exercise:

$$P(Defect \mid Mach1) = ?$$

Example:

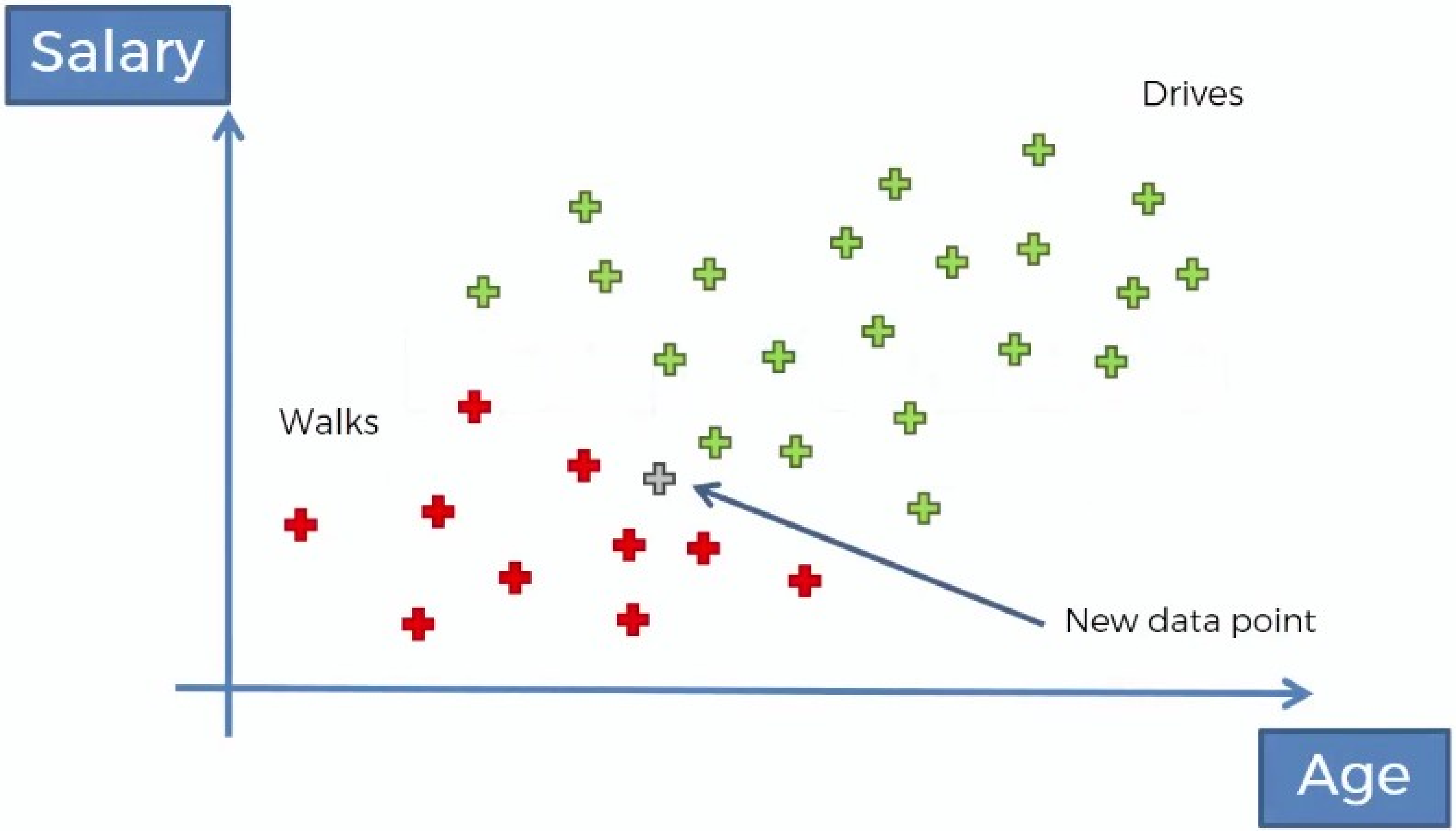


Step-1

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

$$P(Walks|X) = \frac{P(X|Walks) * P(Walks)}{P(X)}$$

Step-1



Step-1

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

$$P(Walks|X) = \frac{P(X|Walks) * P(Walks)}{P(X)}$$

#2 Marginal Likelihood

Step-2

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

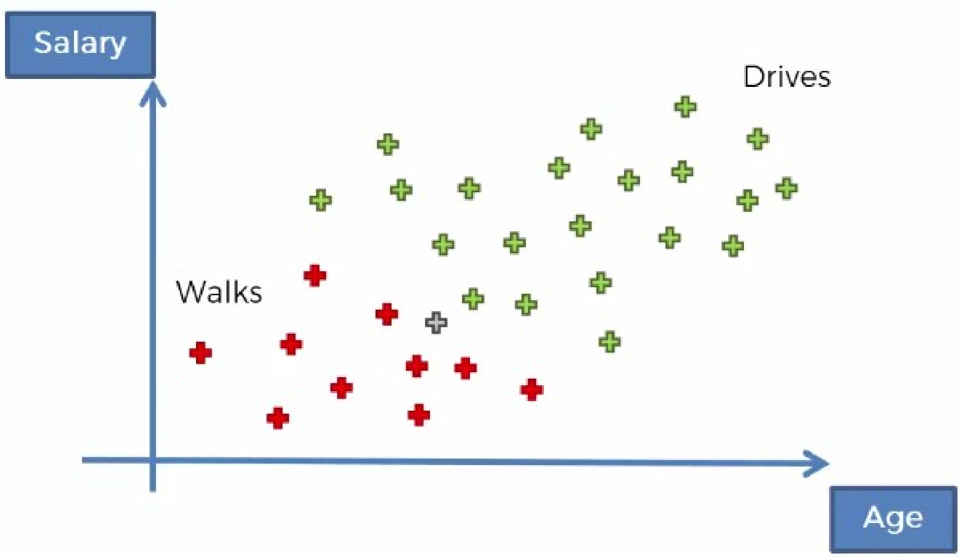
$$P(Drives|X) = \frac{P(X|Drives) * P(Drives)}{P(X)}$$

#2 Marginal Likelihood

Step-3

$$P(Walks|X) \text{ v.s. } P(Drives|X)$$

Naive Bayes – Step-1

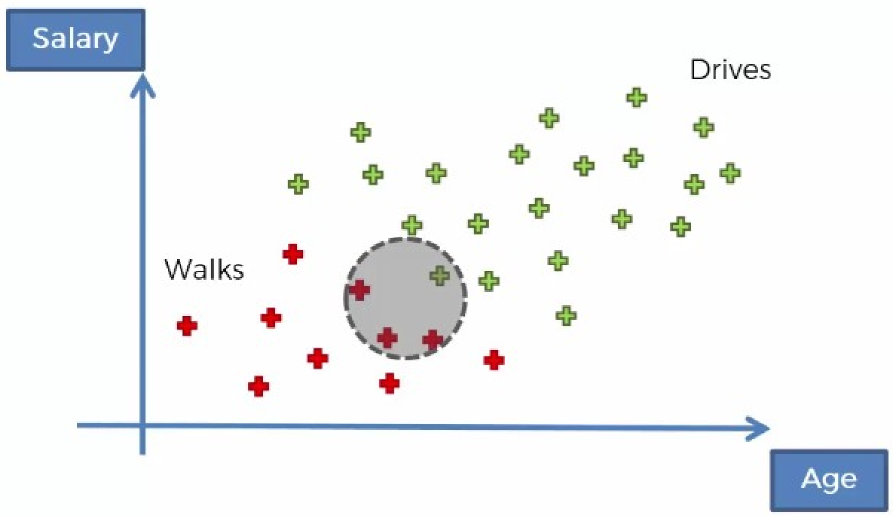


#1. P(Walks)

$$P(Walks) = \frac{Number\ of\ Walkers}{Total\ Observations}$$

$$P(Walks) = \frac{10}{30}$$

Naive Bayes – Step-2

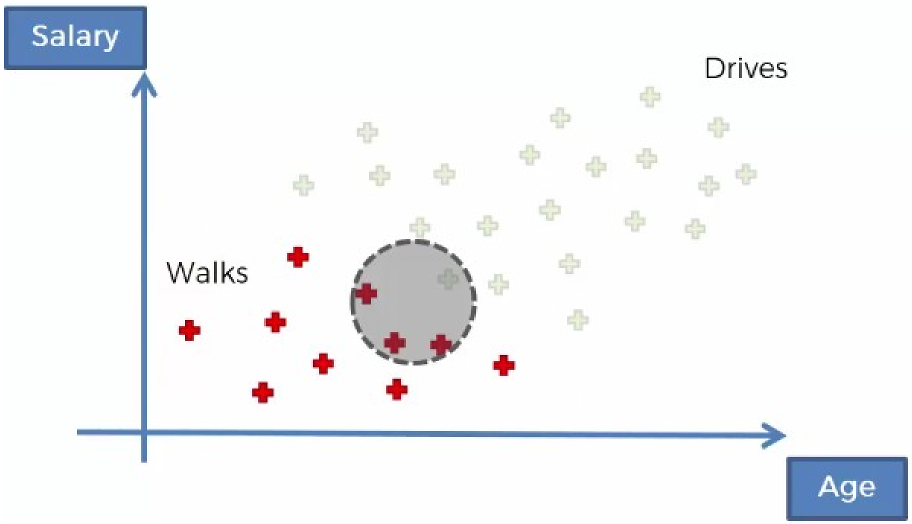


#2. P(X)

$$P(X) = \frac{Number\ of\ Similar\ Observations}{Total\ Observations}$$

$$P(X) = \frac{4}{30}$$

Naive Bayes – Step-3



#3. P(X|Walks)

$$P(X|Walks) = \frac{\begin{matrix}Number\ of\ Similar \\ Observations \\ Among\ those\ who\ Walk\end{matrix}}{Total\ number\ of\ Walkers}$$

$$P(X|Walks) = \frac{3}{10}$$

Combining altogether

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

$$P(Walks|X) = \frac{\frac{3}{10} * \frac{10}{30}}{\frac{4}{30}} = 0.75$$

#2 Marginal Likelihood

Naive Bayes – Step-4

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

$$P(Drives|X) = \frac{P(X|Drives) * P(Drives)}{P(X)}$$

#2 Marginal Likelihood

Naive Bayes – Step-5

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

$$P(Drives|X) = \frac{\frac{1}{20} * \frac{20}{30}}{\frac{4}{30}} = 0.25$$

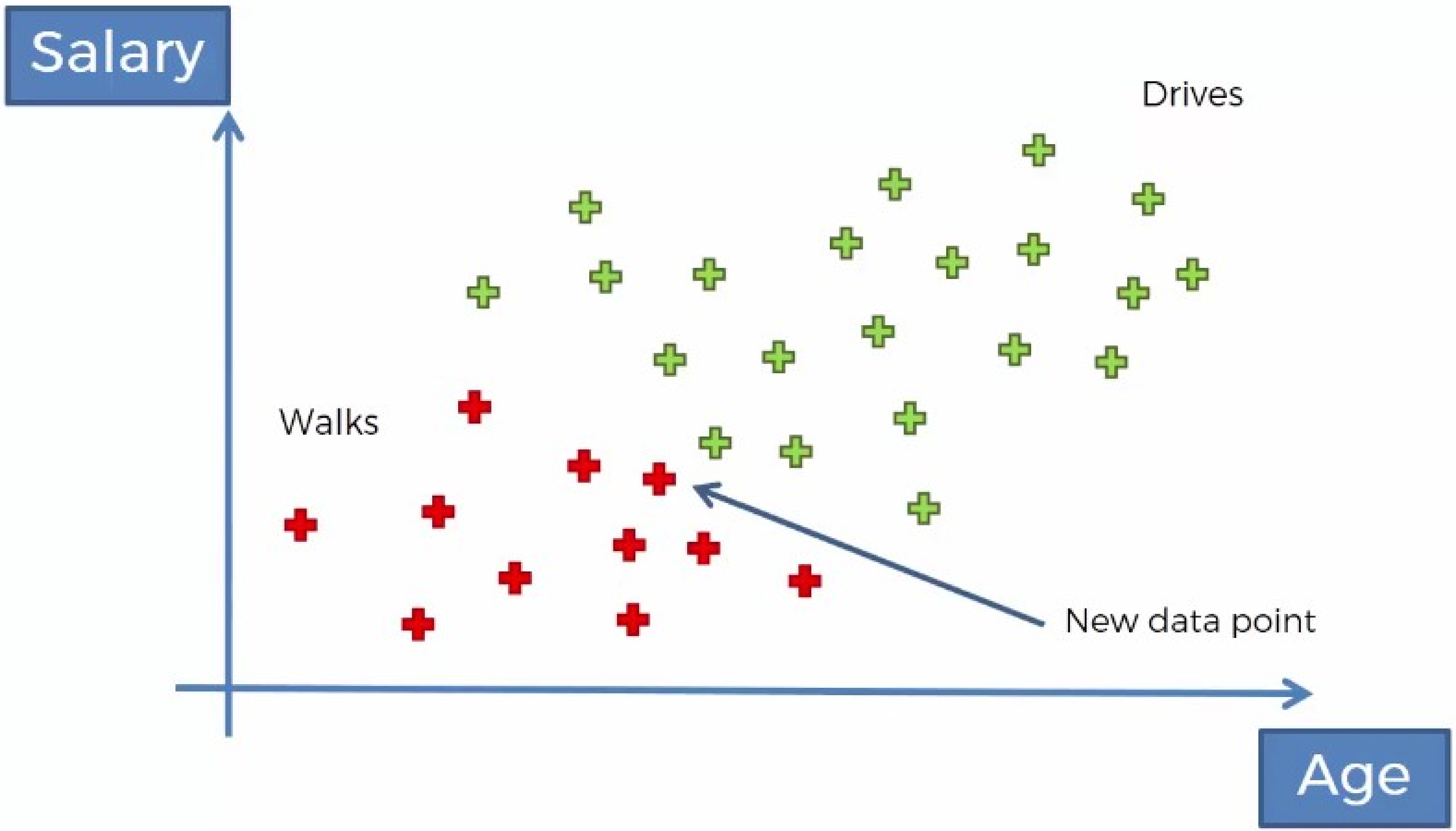
#2 Marginal Likelihood

Types of model

$$P(Walks|X) \; v.s. \; P(Drives|X)$$

$$0.75 \; v.s. \; 0.25$$

Final Classification



Random Variable

- A random variable is a numerical description of the outcome of a statistical experiment.
- A random variable that may assume only a finite number or an infinite sequence of values is said to be discrete; one that may assume any value in some interval on the real number line is said to be continuous.
- For instance, a random variable representing the number of automobiles sold at a particular dealership on one day would be discrete, while a random variable representing the weight of a person in kilograms (or pounds) would be continuous.

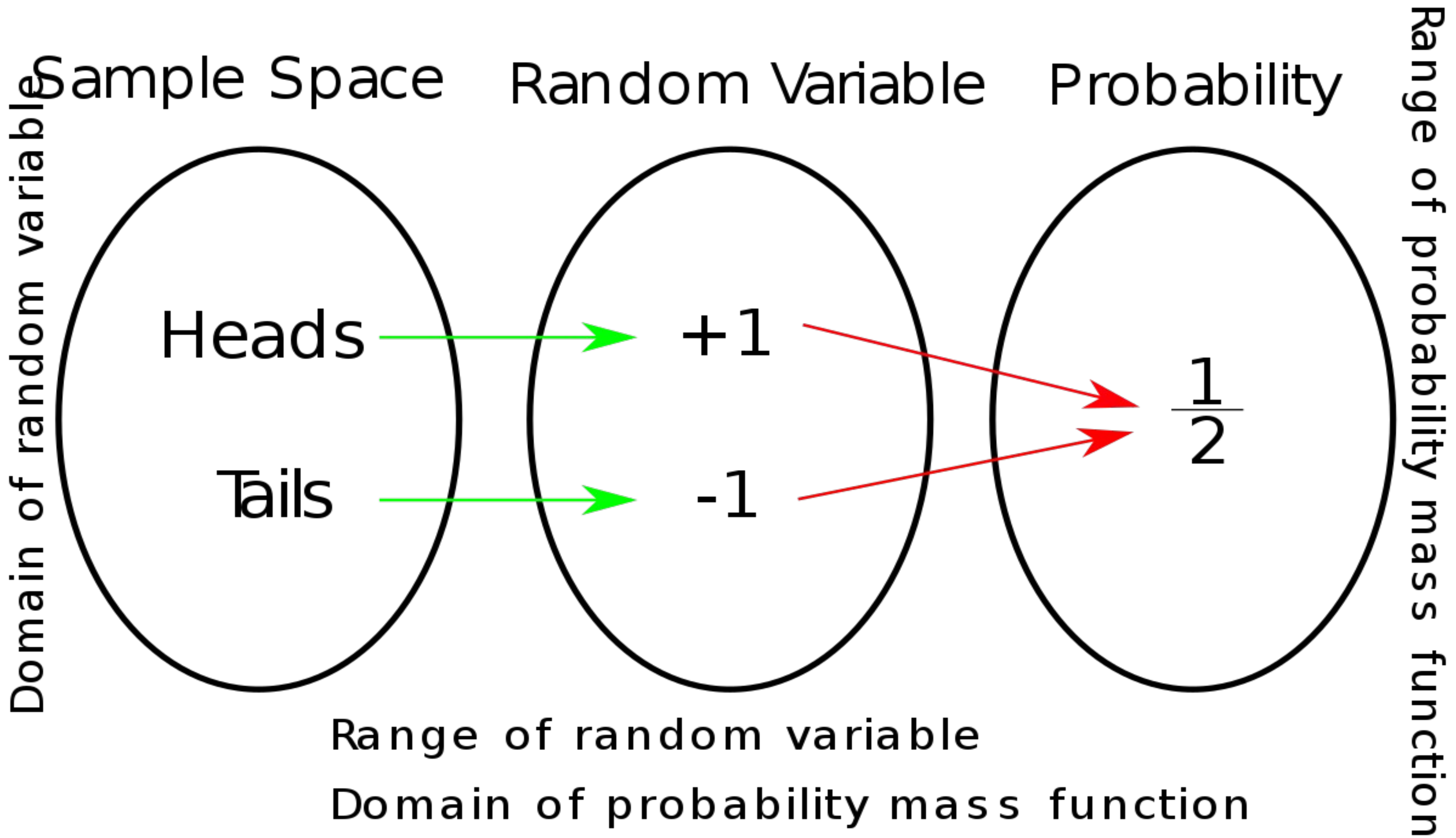
Random Variable

- The probability distribution for a random variable describes how the probabilities are distributed over the values of the random variable.
- For a discrete random variable, x, the probability distribution is defined by a probability mass function, denoted by f(x).
- This function provides the probability for each value of the random variable. In the development of the probability function for a discrete random variable, two conditions must be satisfied:
  - (1) f(x) must be nonnegative for each value of the random variable, and
  - (2) the sum of the probabilities for each value of the random variable must equal one.

Random Variable

- A random variable is a variable whose possible values have an associated probability distribution.
- A very simple random variable equals 1 if a coin flip turns up heads and 0 if the flip turns up tails.
- A more complicated one might measure the number of heads observed when flipping a coin 10 times or a value picked from range(10) where each number is equally likely.

Random Variable



Random Variable

- The associated distribution gives the probabilities that the variable realizes each of its possible values. The coin flip variable equals 0 with probability 0.5 and 1 with probability 0.5.
- The range(10) variable has a distribution that assigns probability 0.1 to each of the numbers from 0 to 9.
- We will sometimes talk about the expected value of a random variable, which is the average of its values weighted by their probabilities.
- The coin flip variable has an expected value of 1/2 (= 0 * 1/2 + 1 * 1/2), and the range(10) variable has an expected value of 4.5.

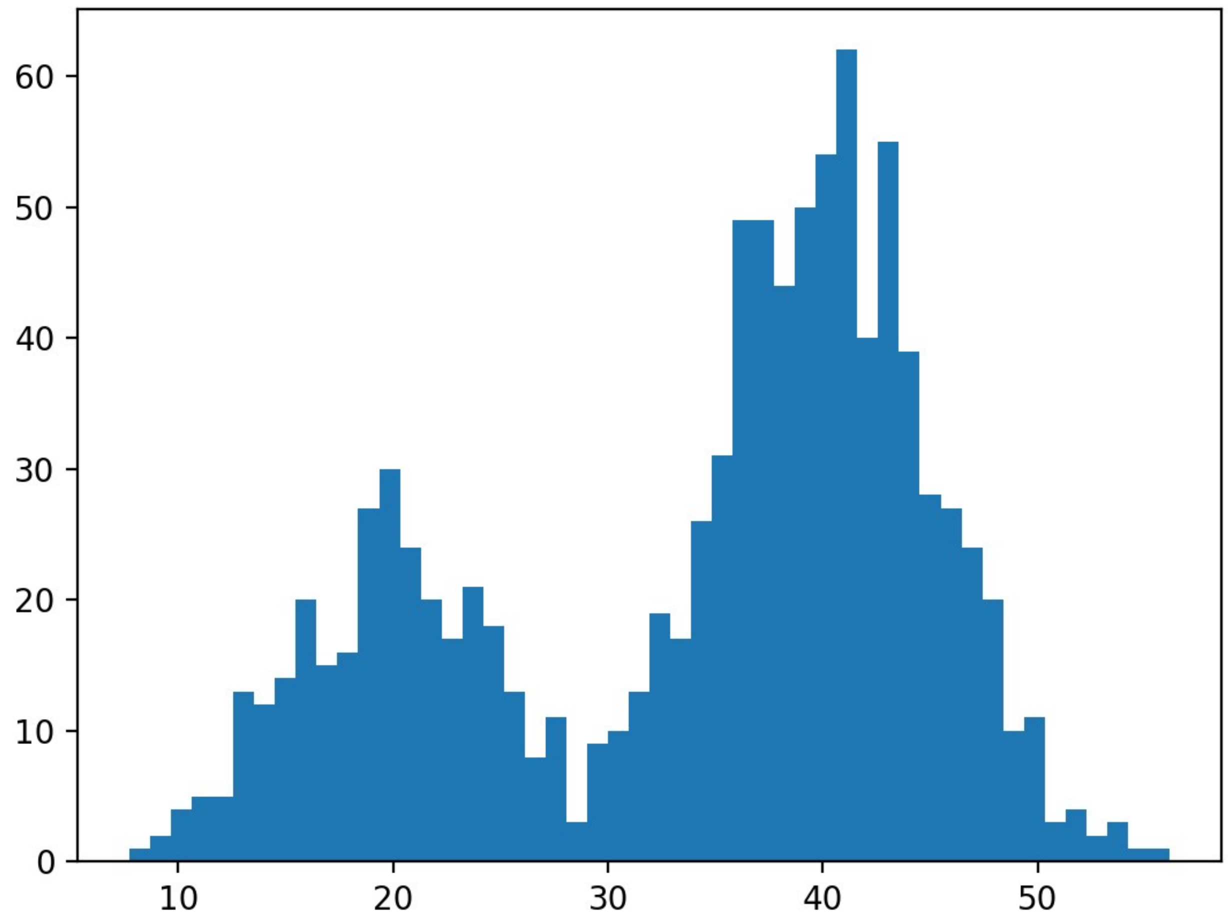
Random Variable

- Random variables can be conditioned on events just as other events can. Going back to the two-child example from "Conditional Probability", if X is the random variable representing the number of girls, X equals 0 with probability 1/4, 1 with probability 1/2, and 2 with probability 1/4.
- We can define a new random variable Y that gives the number of girls conditional on at least one of the children being a girl. Then Y equals 1 with probability 2/3 and 2 with probability 1/3. And a variable Z that's the number of girls conditional on the older child being a girl equals 1 with probability 1/2 and 2 with probability 1/2.

Probability Distribution

- A probability distribution is a function that describes the likelihood of obtaining the possible values that a random variable can assume. In other words, the values of the variable vary based on the underlying probability distribution.
- Suppose you draw a random sample and measure the heights of the subjects. As you measure heights, you can create a distribution of heights. This type of distribution is useful when you need to know which outcomes are most likely, the spread of potential values, and the likelihood of different results.

Probability Distribution



General Properties

- Probability distributions indicate the likelihood of an event or outcome. Statisticians use the following notation to describe probabilities:

  p(x) = the likelihood that random variable takes a specific value of x.

- The sum of all probabilities for all possible values must equal 1. Furthermore, the probability for a particular value or range of values must be between 0 and 1.

Types

- Probability distributions describe the dispersion of the values of a random variable. Consequently, the kind of variable determines the type of probability distribution. For a single random variable, statisticians divide distributions into the following two types:
  - Discrete probability distributions for discrete variables
  - Probability density functions for continuous variables
- You can use equations and tables of variable values and probabilities to represent a probability distribution.

Discrete Probability Distribution

- Discrete probability functions are also known as probability mass functions and can assume a discrete number of values.
- For example, coin tosses and counts of events are discrete functions. These are discrete distributions because there are no in-between values.
- For example, you can have only heads or tails in a coin toss. Similarly, if you're counting the number of books that a library checks out per hour, you can count 21 or 22 books, but nothing in between.

Discrete Probability Distribution

- For discrete probability distribution functions, each possible value has a non-zero likelihood. Furthermore, the probabilities for all possible values must sum to one. Because the total probability is 1, one of the values must occur for each opportunity.
- For example, the likelihood of rolling a specific number on a die is 1/6. The total probability for all six values equals one. When you roll a die, you inevitably obtain one of the possible values.

Discrete Probability Distribution

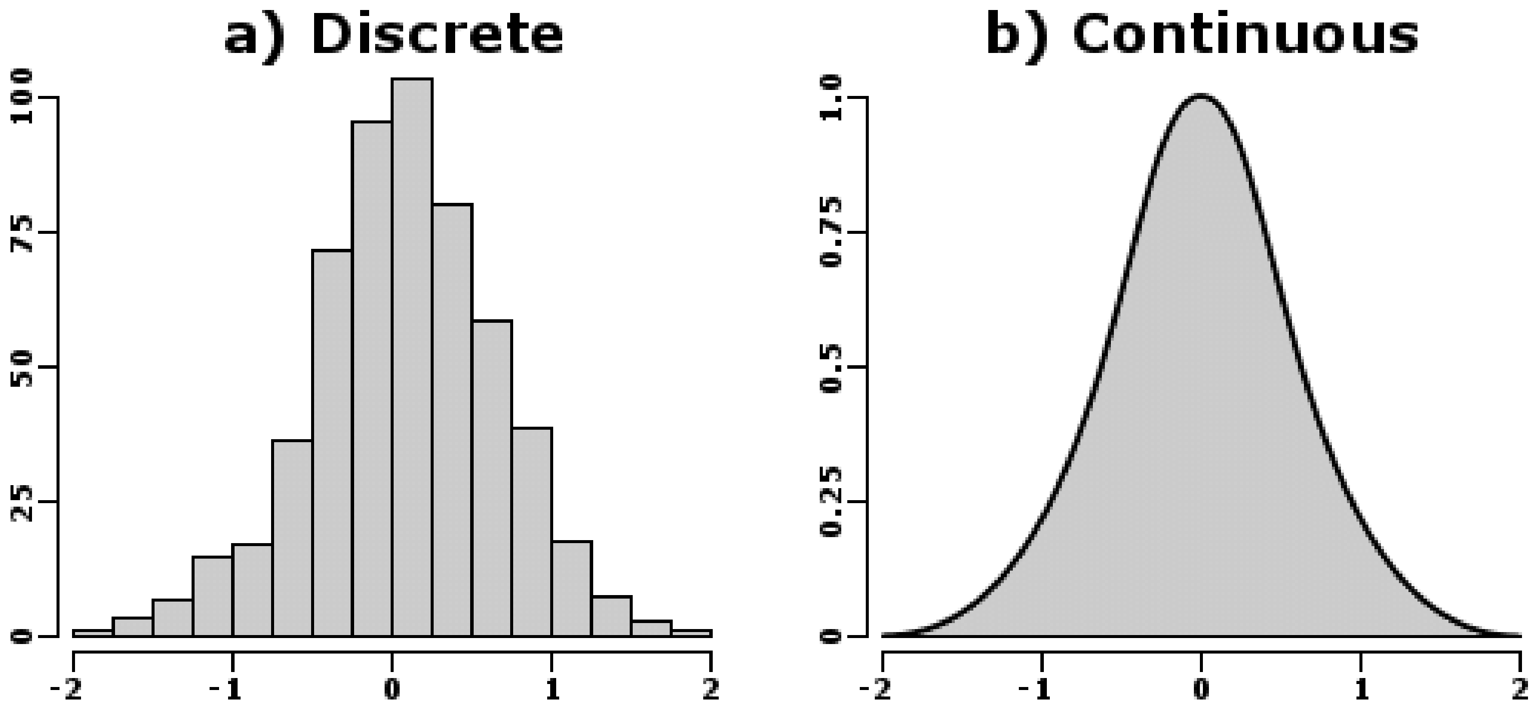
- If the discrete distribution has a finite number of values, you can display all the values with their corresponding probabilities in a table. For example, according to a study, the likelihood for the number of cars in Pune household is the following:

Number of Cars	Probability
0	0.03
1	0.13
2	0.70
3	0.10
4+	0.04

Continous Probability Distribution

- Continuous probability functions are also known as probability density functions. You know that you have a continuous distribution if the variable can assume an infinite number of values between any two values. Continuous variables are often measurements on a scale, such as height, weight, and temperature.
- Unlike discrete probability distributions where each particular value has a non-zero likelihood, specific values in continuous distributions have a zero probability. For example, the likelihood of measuring a temperature that is exactly 32 degrees is zero.

Continous Probability Distribution



How to find ?

- Probabilities for continuous distributions are measured over ranges of values rather than single points. A probability indicates the likelihood that a value will fall within an interval. This property is straightforward to demonstrate using a probability distribution plot.
- On a probability plot, the entire area under the distribution curve equals 1. This fact is equivalent to how the sum of all probabilities must equal one for discrete distributions. The proportion of the area under the curve that falls within a range of values along the X-axis represents the likelihood that a value will fall within that range.

Characteristics

- Just as there are different types of discrete distributions for different kinds of discrete data, there are different distributions for continuous data.
- Each probability distribution has parameters that define its shape. Most distributions have between 1-3 parameters.
- Specifying these parameters establishes the shape of the distribution and all of its probabilities entirely.
- These parameters represent essential properties of the distribution, such as the central tendency and the variability.

Characteristics

- The most well-known continuous distribution is the normal distribution, which is also known as the Gaussian distribution or the "bell curve."
- This symmetric distribution fits a wide variety of phenomena, such as human height and IQ scores. It has two parameters—the mean and the standard deviation.
- The Weibull distribution and the lognormal distribution are other common continuous distributions. Both of these distributions can fit skewed data.

Characteristics

- Distribution parameters are values that apply to entire populations.
- Unfortunately, population parameters are generally unknown because it’s usually impossible to measure an entire population.
- However, you can use random samples to calculate estimates of these parameters.

Normal Distribution

- Normal distribution represents the behavior of most of the situations in the universe (That is why it's called a "normal" distribution. I guess!).
- The large sum of (small) random variables often turns out to be normally distributed, contributing to its widespread application.

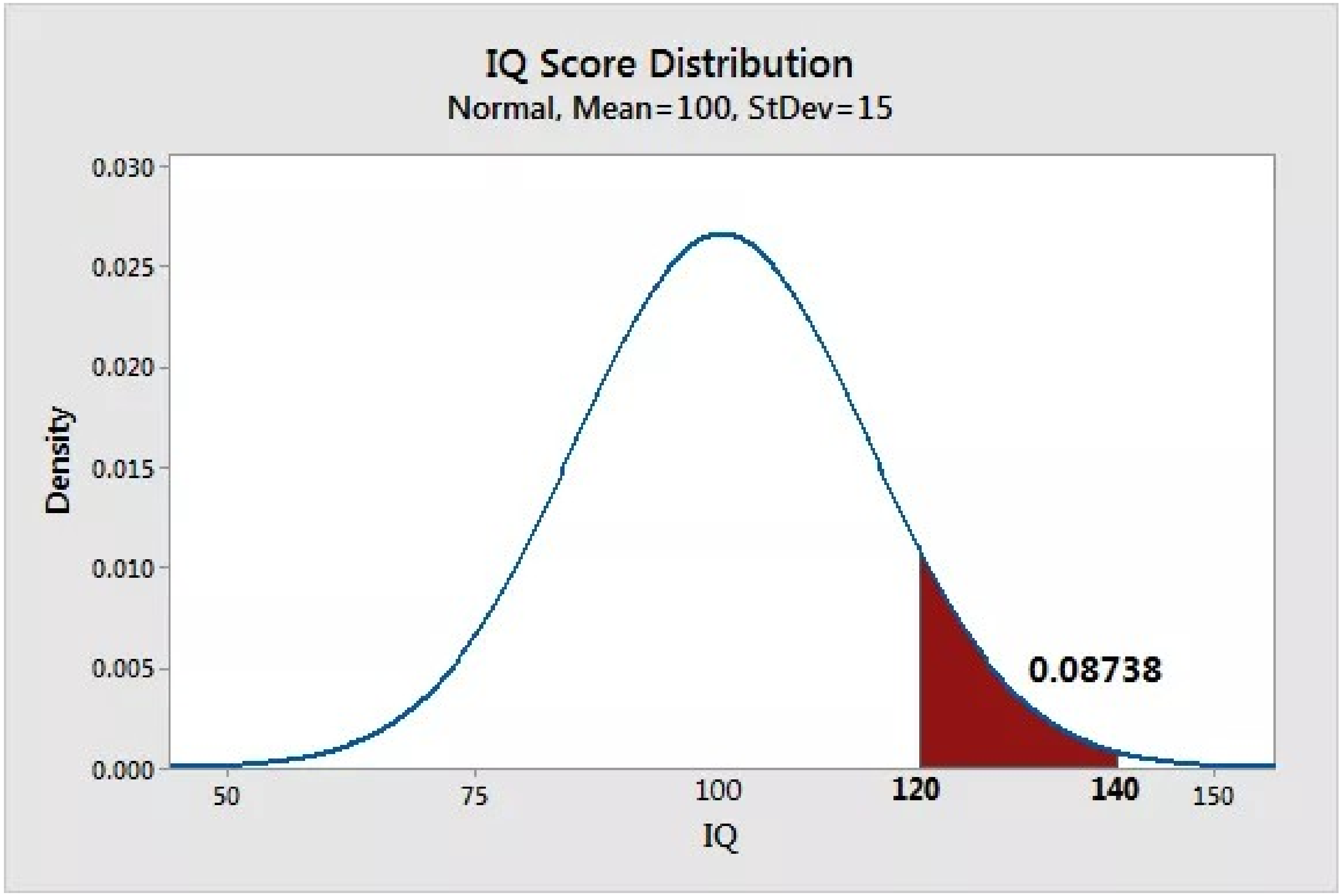
Normal Distribution

- Any distribution is known as Normal distribution if it has the following characteristics:
  - The mean, median and mode of the distribution coincide.
  - The curve of the distribution is bell-shaped and symmetrical about the line $x=\mu$.
  - The total area under the curve is 1.
  - Exactly half of the values are to the left of the center and the other half to the right.

Normal Distribution : Example

- Let's start off with the normal distribution to show how to use continuous probability distributions.
- The distribution of IQ scores is defined as a normal distribution with a mean of 100 and a standard deviation of 15. We'll create the probability plot of this distribution.
- Additionally, let's determine the likelihood that an IQ score will be between 120-140.

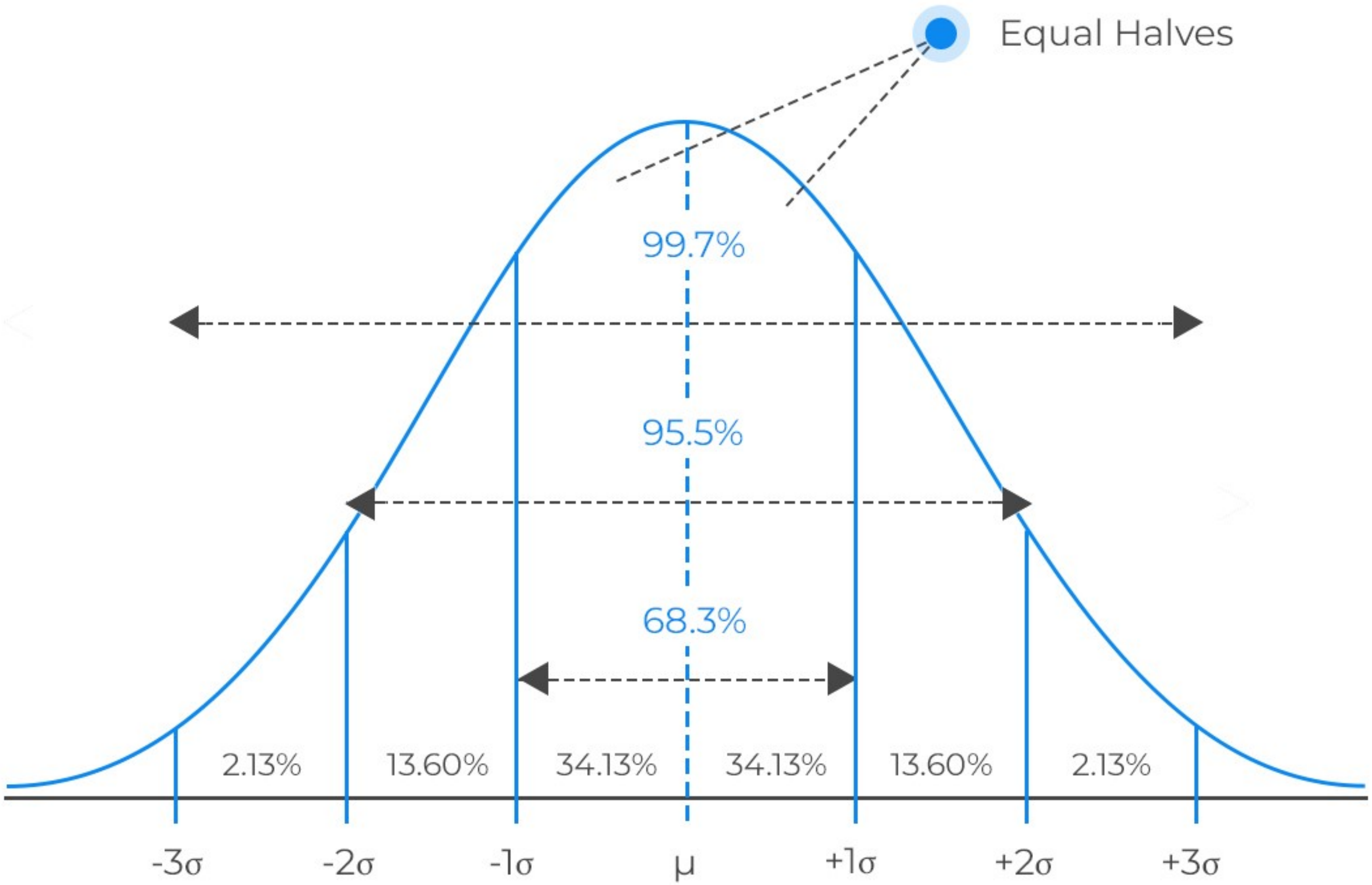
Normal Distribution : Example



Normal Distribution : Example

- We can see that it is a symmetric distribution where values occur most frequently around 100, which is the mean. The probabilities drops-off as you move away from the mean in both directions.
- The shaded area for the range of IQ scores between 120-140 contains 8.738% of the total area under the curve.
- Therefore, the likelihood that an IQ score falls within this range is 0.08738.

Normal Distribution : Shape



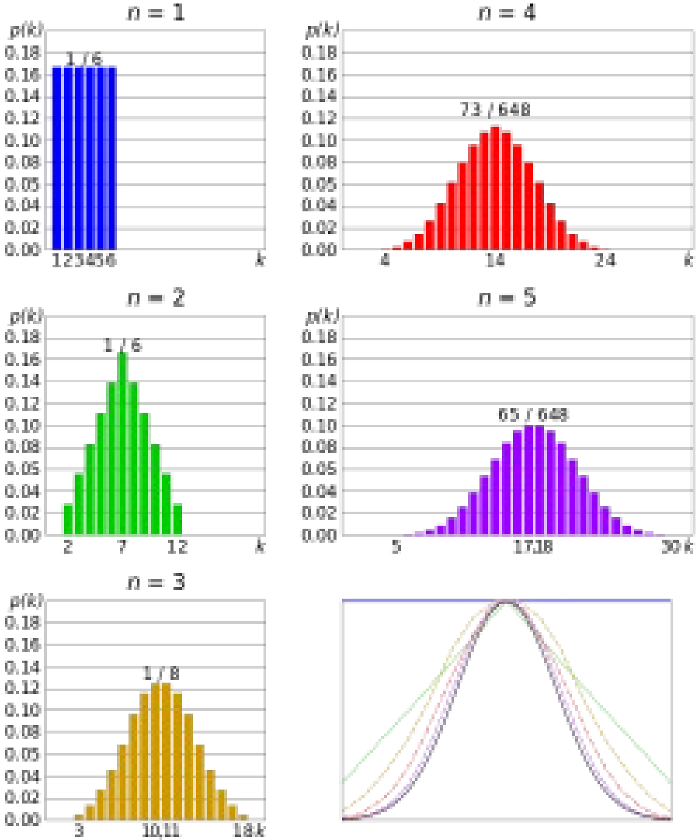
No. of standard deviations from the mean

Central Limit Theorem

- The Central Limit Theorem states that the sampling distribution of the sample means approaches a normal distribution as the sample size gets larger — no matter what the shape of the population distribution. This fact holds especially true for sample sizes over 30.
- All this is saying is that as you take more samples, especially large ones, your graph of the sample means will look more like a normal distribution.

Central Limit Theorem

- Here's what the Central Limit Theorem is saying, graphically. The picture below shows one of the simplest types of test: rolling a fair die.
- The more times you roll the die, the more likely the shape of the distribution of the means tends to look like a normal distribution graph.



Central Limit Theorem

- An essential component of the Central Limit Theorem is that the average of your sample means will be the population mean.
- In other words, add up the means from all of your samples, find the average and that average will be your actual population mean.
- Similarly, if you find the average of all of the standard deviations in your sample, you'll find the actual standard deviation for your population.
- It's a pretty useful phenomenon that can help accurately predict characteristics of a population.

Examples:

- A Central Limit Theorem word problem will most likely contain the phrase “assume the variable is normally distributed”, or one like it. With these central limit theorem examples, you will be given:
  - A population (i.e. 29-year-old males, seniors between 72 and 76, all registered vehicles, all cat owners)
  - An average (i.e. 125 pounds, 24 hours, 15 years, $15.74)
  - A standard deviation (i.e. 14.4lbs, 3 hours, 120 months, $196.42)
  - A sample size (i.e. 15 males, 10 seniors, 79 cars, 100 households)

Thank you

@mitu_skillologies

/mITuSkillologies

@mitu_group

/company/mitu-skillologies

MITUSkillologies

Web Resources
https://mitu.co.in
http://tusharkute.com

contact@mitu.co.in

tushar@tusharkute.com