

Unit 2

Prepared By: Bhavana Hotchandani, DCS, INDUS University

Index

- **Probability:**
 - Dependence and Independence
 - Conditional Probability
 - Bayes's Theorem
 - Random Variables
 - Continuous Distributions
 - The Normal Distribution
 - The Central Limit Theorem

Introduction

- **What is probability?**
- Probability is the chance that something will happen — how likely it is that some event will happen.
- *Probability of an event happening*
- $P(E) = \text{Number of ways it can happen } n(E) / \text{Total number of outcomes } n(T)$
- Probability is the measure of the likelihood that an event will occur. Probability is quantified as a number between 0 and 1, where 0 indicates impossibility and 1 indicates certainty.

Common Terminology

Experiment – are the uncertain situations, which could have multiple outcomes. Whether it rains on a daily basis is an experiment.

Outcome - is the result of a single trial. So, if it rains today, the outcome of today's trial from the experiment is "It rained"

Event - is one or more outcome from an experiment. "It rained" is one of the possible event for this experiment.

Probability - is a measure of how likely an event is. So, if it is 60% chance that it will rain tomorrow, the probability of Outcome "it rained" for tomorrow is 0.6

● Why probability is important?

- Uncertainty and randomness occur in many aspects of our daily life and having a good knowledge of probability helps us make sense of these uncertainties.
- Learning about probability helps us make informed judgments on what is likely to happen, based on a pattern of data collected previously or an estimate

● How Probability is used in Data Science?

- Data science often uses statistical inferences to predict or analyze trends from data, while statistical inferences uses probability distributions of data. Hence knowing probability and its applications are important to work effectively on data science problems.
- Probability requires to know about inferences of certain conclusion. It is the occurrence of particular outcome by measuring it's significance in daily problems. Without probability , you cannot do data science problems easily . We have used Bayesian inference , Monte Carlo methods on the basis of Probability . Without it , no data scientist can work on any problem .

Dependence and Independence

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.

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

- **What is Conditional Probability?**

- Conditional probability is a measure of the probability of an event (some particular situation occurring) given that (by assumption, presumption, assertion or evidence) another event has occurred.

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

Conditional Probability



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

$$P(A|B) = \frac{P(A \cap B)}{P(A \cap B) + P(A' \cap B)}$$

$$P(A \cap B) = P(B)P(A|B)$$

Handwritten calculation:

$$P(H|R) = \frac{P(H \cap R)}{P(R)}$$

13/52 = 1/4

$$\frac{13/52}{1/2} = \frac{13/52 \times 2}{1/2} = \frac{13/26}{1/2} = \frac{13}{13} = 1$$

A and B are independent if $P(A) = P(A|B) = P(A|B')$

Given that:

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

Then if A and B are independent:

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

Handwritten calculation:

$$\frac{1/4}{1/2} = \frac{1}{2}$$

$\hookrightarrow 26/52 = 1/2$



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

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

$$P(A|B) = \frac{P(A \cap B)}{P(A \cap B) + P(A' \cap B)}$$

$$P(A|B) = P(A)$$

A and B are independent if $P(A) = P(A|B) = P(A|B')$

Given that:

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

Then if A and B are independent:

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



Random Experiment:

A random experiment is a physical situation whose outcome cannot be predicted until it is observed.

Sample Space

A sample space is a set of all possible outcomes of a random experiment.

Forexample, we have:

Random Experiment: Tossing of a fair coin

Sample space: {Head, Tail}

As we got a little understanding of Probability, we will now read about Probability Distribution and its types with the help of examples and formulas wherever required.

Distribution

In statistics when we use the term Distribution it usually means Probability distribution.

A Distribution is a function that shows the possible values for a variable and how often they occur.

Or A Probability Distribution is a mathematical function that can be thought of as providing the probabilities of occurrence of different possible outcomes in an experiment.

Good examples are the

Normal Distribution

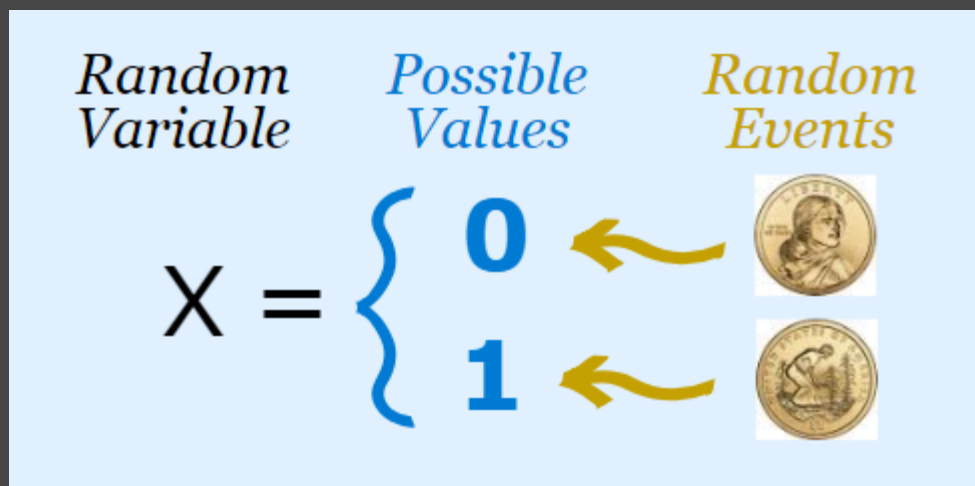
Binomial Distribution

Uniform Distribution

Random Variable

A random variable is a set of possible values from a random experiment.

Random variables can be discrete or continuous. *Discrete random variables* can only take certain values while *continuous random variables* can take any value (within a range).



A discrete random variable is one which may take on only a countable number of distinct values and thus can be quantified. For example, you can define a random variable X to be the number which comes up when you roll a fair dice. X can take values : $[1,2,3,4,5,6]$ and therefore is a discrete random variable.

A continuous random variable is one which takes an infinite number of possible values. For example, you can define a random variable X to be the height of students in a class. Since the continuous random variable is defined over an interval of values, it is represented by the area under a curve

- **What are probability distributions?**
- 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.
- For a *continuous random variable*, since there is an infinite number of values in any interval, the probability that a continuous random variable will lie within a given interval is considered. So here, the probability distribution is defined by **probability density function**, also denoted by $f(x)$.

Both probability functions must satisfy two requirements::

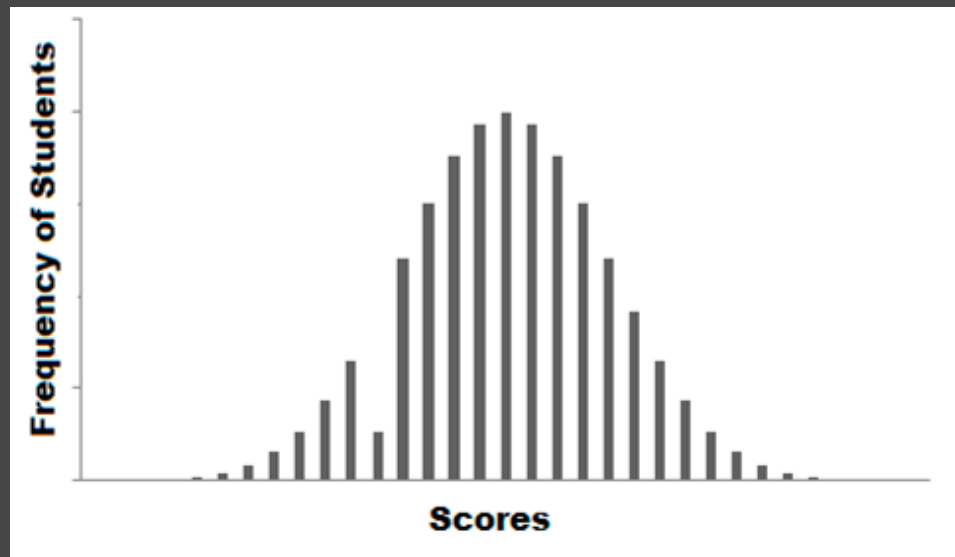
- (1) $f(x)$ must be non-negative for each value of the random variable,
and
- (2) the sum of the probabilities for each value (or integral over all values) of the random variable must equal one.

Example

Suppose you are a teacher at a university. After checking assignments for a week, you graded all the students. You gave these graded papers to a data entry guy in the university and tell him to create a spreadsheet containing the grades of all the students. But the guy only stores the grades and not the corresponding students.

- He made another blunder, he missed a couple of entries in a hurry and we have no idea whose grades are missing. Let's find a way to solve this.
- One way is that you visualize the grades and see if you can find a trend in the data.

S. No.	Scores
1	25
2	27
3	38
4	42
5	16
6	35
7	46
8	48
9	31
10	31



- The graph that you have plot is called the frequency distribution of the data. You see that there is a smooth curve like structure that defines our data, but do you notice an anomaly? We have an abnormally low frequency at a particular score range. So the best guess would be to have missing values that remove the dent in the distribution.
- This is how you would try to solve a real-life problem using data analysis. For any Data Scientist, a student or a practitioner, distribution is a must know concept. It provides the basis for analytics and inferential statistics.
- While the concept of probability gives us the mathematical calculations, distributions help us actually visualize what's happening underneath.

Normal Distribution

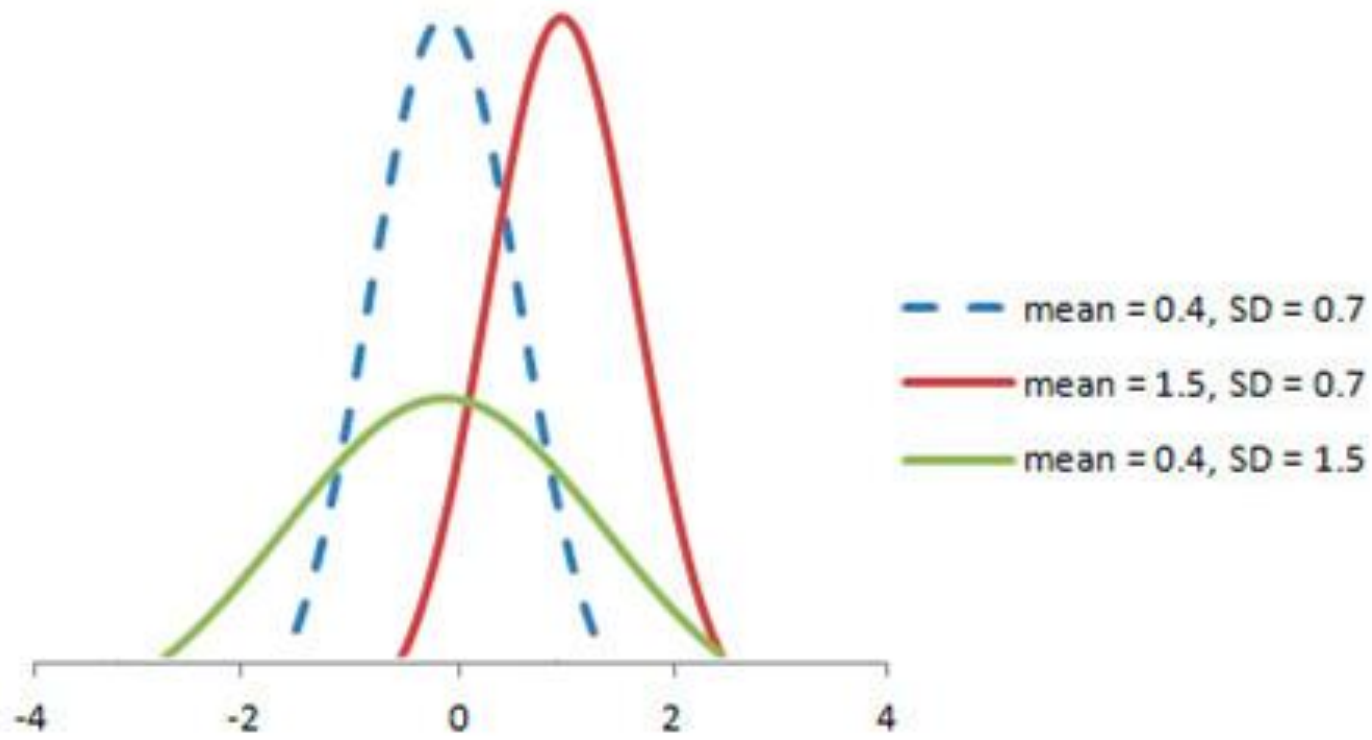
- **Normal distribution** represents the behavior of most of the situations in the universe (That is why it's called a “normal” distribution. I guess!). The large sum of (small) random variables often turns out to be normally distributed, contributing to its widespread application. Any distribution is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

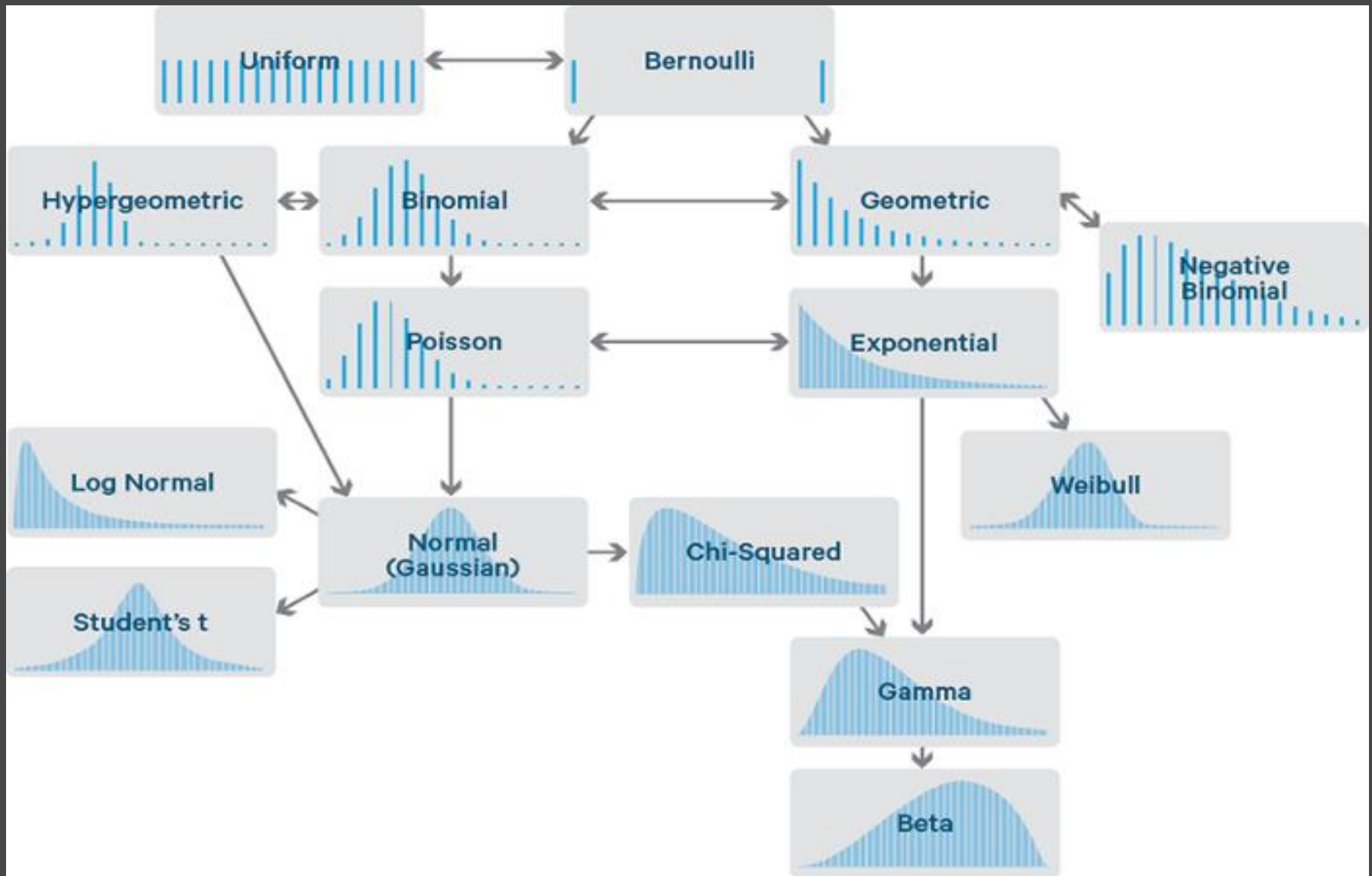
A continuous random variable is one which takes an infinite number of possible values. For example, you can define a random variable X to be the height of students in a class. Since the continuous random variable is defined over an interval of values, it is represented by the area under a curve. The mean and variance of a random variable X which is said to be normally distributed is given by:

$$\text{Mean} \rightarrow E(X) = \mu$$

$$\text{Variance} \rightarrow \text{Var}(X) = \sigma^2$$

Here, μ (mean) and σ (standard deviation) are the parameters. The graph of a random variable $X \sim N(\mu, \sigma)$ is shown below.





Bayes Theorem

Bayes' Theorem is a way to calculate conditional probability.

Conditional probability is the probability of an event happening, given that it has some relationship to one or more other events.

For example, your probability of getting a parking space is connected to the time of day you park, where you park, and what conventions are going on at any time.

The Theorem was named after English mathematician Thomas Bayes . The formal definition for the rule is:

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

The formula is very simple to calculate, but it can be challenging to fit the right pieces into the puzzle.

The first challenge comes from defining your event (A) and test (B); The second challenge is rephrasing your question so that you can work backwards: turning $P(A|B)$ into $P(B|A)$.

The following image shows a basic example involving website traffic.

Q: What is the probability a person who visits my homepage will end up buying something?

1. Define your "event" and "test":

A (an event): Person buys something

B (a test): Person visits the homepage

2. Define $P(A|B)$:

The probability that the person buys something, given that test B (visiting the homepage) was positive.

$P(B|A)$:
The probability that the person visited the homepage, given that they bought something. Past data tells you among those persons making a purchase, 7% have visited your homepage

$P(A)$:
Past data tells you that 10% of people who visit your site buy a product.

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

$P(B)$:
5% of visitors go to your homepage.

$$P(A|B) = \frac{(0.07)(0.1)}{0.05}$$

$$P(A|B) = 0.14$$

A: If the person visits your homepage, their chances of making a purchase is 0.14 (14%). This is a slight increase from the 10% suggested by past data.

Some more examples:

You might be interested in finding out a patient's probability of having liver disease if they are an alcoholic. "Being an alcoholic" is the **test** (kind of like a litmus test) for liver disease.

A could mean the event "Patient has liver disease." Past data tells you that 10% of patients entering your clinic have liver disease. $P(A) = 0.10$.

B could mean the litmus test that "Patient is an alcoholic." Five percent of the clinic's patients are alcoholics. $P(B) = 0.05$.

You might also know that among those patients diagnosed with liver disease, 7% are alcoholics. This is your **B|A**: the probability that a patient is alcoholic, given that they have liver disease, is 7%.

Bayes' theorem tells you:

$$P(A|B) = (0.07 * 0.1) / 0.05 = 0.14$$

In other words, if the patient is an alcoholic, their chances of having liver disease is 0.14 (14%). This is a large increase from the 10% suggested by past data. But it's still unlikely that any particular patient has liver disease.

In a particular pain clinic, 10% of patients are prescribed narcotic pain killers. Overall, five percent of the clinic's patients are addicted to narcotics (including pain killers and illegal substances). Out of all the people prescribed pain pills, 8% are addicts. *If a patient is an addict, what is the probability that they will be prescribed pain pills?*

Step 1: Figure out what your event “A” is from the question. That information is in the italicized part of this particular question. The event that happens first (A) is being prescribed pain pills. That's given as 10%.

Step 2: Figure out what your event “B” is from the question. That information is also in the italicized part of this particular question. Event B is being an addict. That's given as 5%.

Step 3: Figure out what the probability of event B (Step 2) given event A (Step 1). In other words, find what $P(B|A)$ is. We want to know “Given that people are prescribed pain pills, what's the probability they are an addict?” That is given in the question as 8%, or .8.

Step 4: Insert your answers from Steps 1, 2 and 3 into the formula and solve.

$$P(A|B) = P(B|A) * P(A) / P(B) = (0.08 * 0.1) / 0.05 = 0.16$$

The probability of an addict being prescribed pain pills is 0.16 (16%).

central limit theorem

There are some technical areas that scientists should really be experts in, and of those hypothesis testing ranks at the top. And hypothesis testing relies on the central limit theorem.

Why does it matter? It's the core idea in statistics that lets you use data to evaluate your ideas, even with incomplete information.

The central limit theorem (CLT) is simple. It just says that “with a large sample size, sample means are normally distributed.”

Normal distributed means that a group of numbers follows a bell-shaped curve. Most of the numbers cluster in the middle around the average, and there are fewer numbers at the extremes to the right and left. Looks something like this:



A sample mean is the average of a random subset of a larger group. So if you randomly picked 10 people out of 100 and recorded their heights, the average of those 10 heights would be the sample mean. You could do this many times and, since it is a random selection, the sample mean would be different each time.

What constitutes a large sample is of course subjective, but the frequently cited number here is a sample size should be greater than 30. In practice, your sample might need to be a fair amount bigger, depending on a few factors.

Putting it all together the CLT just says that when you have roughly 30 or more observations in your sample, the average of those numbers is part of a bell-shaped curve. So if you took a bunch of 30+ sized samples and plotted them, they'd look like the normal distribution pictured above — most of the averages would fall toward the center but you'd get a few observations towards the extremes.

The CLT make no assumptions about the distribution of your underlying data. The distribution of people's heights does not need to be normally distributed in order to know that the sample means of the heights are normally distributed.

KEY TAKEAWAYS

- The central limit theorem (CLT) states that the distribution of sample means approximates a normal distribution as the sample size gets larger.
- Sample sizes equal to or greater than 30 are considered sufficient for the CLT to hold.
- A key aspect of CLT is that the average of the sample means and standard deviations will equal the population mean and standard deviation.
- A sufficiently large sample size can predict the characteristics of a population accurately.

Hypothesis and Inference

A hypothesis is the prediction about the outcome of an experiment. An inference is conclusion drawn based on observations and prior knowledge.

Hypothesis: Made before an experiment. "If I do this (has to do with the independent variable), then this will happen (has to do with the dependent variable)."

Inference: "Based on my observations (concrete, provable things found via the 5 senses), I can conclude that this happened."

Statistical Hypothesis Testing

What is hypothesis testing ?

Hypothesis testing is a statistical method that is used in making statistical decisions using experimental data.

Hypothesis Testing is basically an assumption that we make about the population parameter.

Ex : you say avg student in class is 40 or a boy is taller than girls.

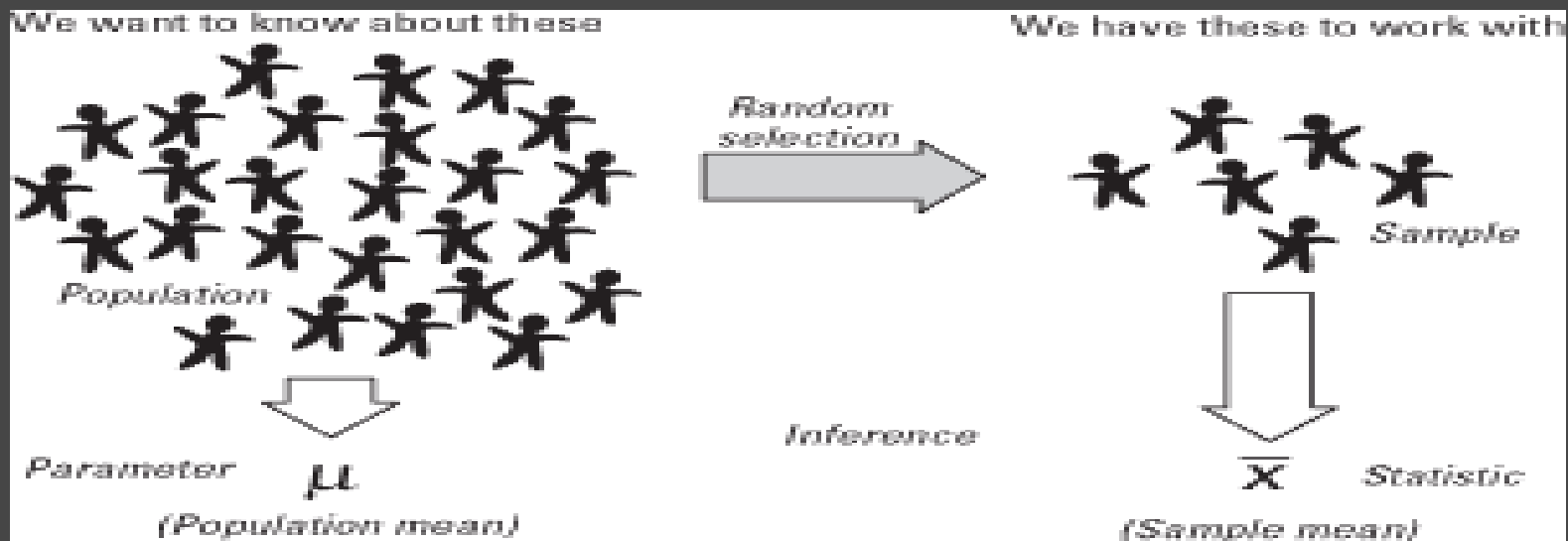
all those example we assume need some statistic way to prove those. we need some mathematical conclusion what ever we are assuming is true.

Statistics is all about data. Data alone is not interesting. It is the interpretation of the data that we are interested in. Using Hypothesis Testing, we try to interpret or draw conclusions about the population using sample data. A Hypothesis Test evaluates two mutually exclusive statements about a population to determine which statement is best supported by the sample data.

Whenever we want to make claims about the distribution of data or whether one set of results are different from another set of results in applied machine learning, we must rely on statistical hypothesis tests.

There are two possible outcomes: if the result confirms the hypothesis, then you've made a measurement. If the result is contrary to the hypothesis, then you've made a discovery

Terminology: Population, sample, statistics



A Statistic is a summary description of a characteristic or measure of the sample. The Sample Statistic is used as an estimate of the population parameter.

Mean (μ), Variance (σ^2), Standard Deviation (σ), Proportion (π)

Sample Mean (\bar{x}), Sample Variance (S^2), Sample Standard Deviation (S), Sample Proportion (p)

Which are important parameter of hypothesis testing ?

Null hypothesis :- In inferential statistics, the null hypothesis is a general statement or default position that there is no relationship between two measured phenomena, or no association among groups

In other words it is a basic assumption or made based on domain or problem knowledge.

Example : a company production is = 50 unit/per day etc.

Alternative hypothesis :-

The alternative hypothesis is the hypothesis used in hypothesis testing that is contrary to the null hypothesis. It is usually taken to be that the observations are the result of a real effect (with some amount of chance variation superposed)

Example : a company production is $\neq 50$ unit/per day etc.

Null vs. Alternative Hypothesis

Null Hypothesis

$$H_0$$

A statement about a population parameter.

We test the likelihood of this statement being true in order to decide whether to accept or reject our alternative hypothesis.

Can include =, ≤, or ≥ sign.

Alternative Hypothesis

$$H_a$$

A statement that directly contradicts the null hypothesis.

We determine whether or not to accept or reject this statement based on the likelihood of the null (opposite) hypothesis being true.

Can include a ≠, >, or < sign.



Level of significance: Refers to the degree of significance in which we accept or reject the null-hypothesis. 100% accuracy is not possible for accepting or rejecting a hypothesis, so we therefore select a level of significance that is usually 5%.

This is normally denoted with alpha (maths symbol α) and generally it is 0.05 or 5%, which means your output should be 95% confident to give similar kind of result in each sample.

Type I error: When we reject the null hypothesis, although that hypothesis was true. Type I error is denoted by alpha. In hypothesis testing, the normal curve that shows the critical region is called the alpha region

Type II errors: When we accept the null hypothesis but it is false. Type II errors are denoted by beta. In Hypothesis testing, the normal curve that shows the acceptance region is called the beta region.

Reality

True

False

True

Correct



Type 1 error

False Positive

Measured or Perceived

False

Type 2 error

False Negative

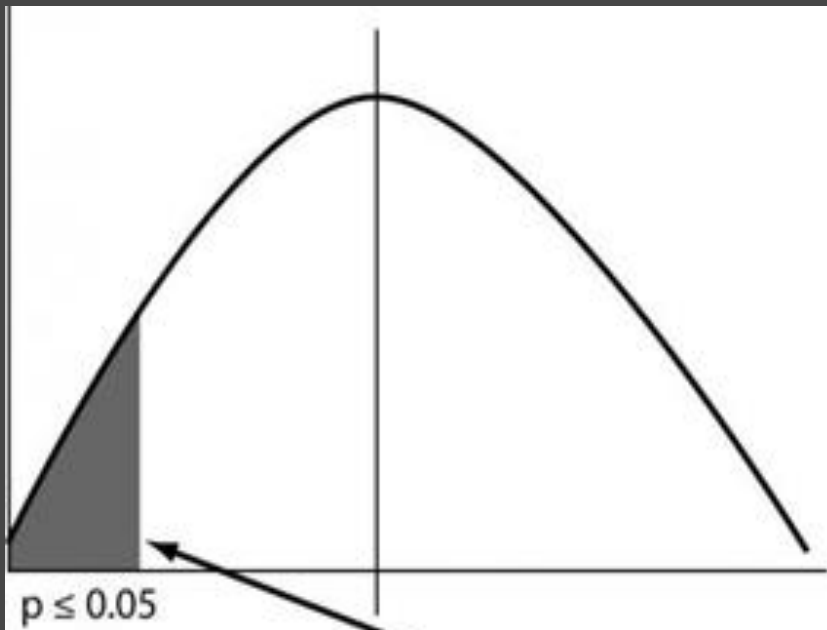
Correct



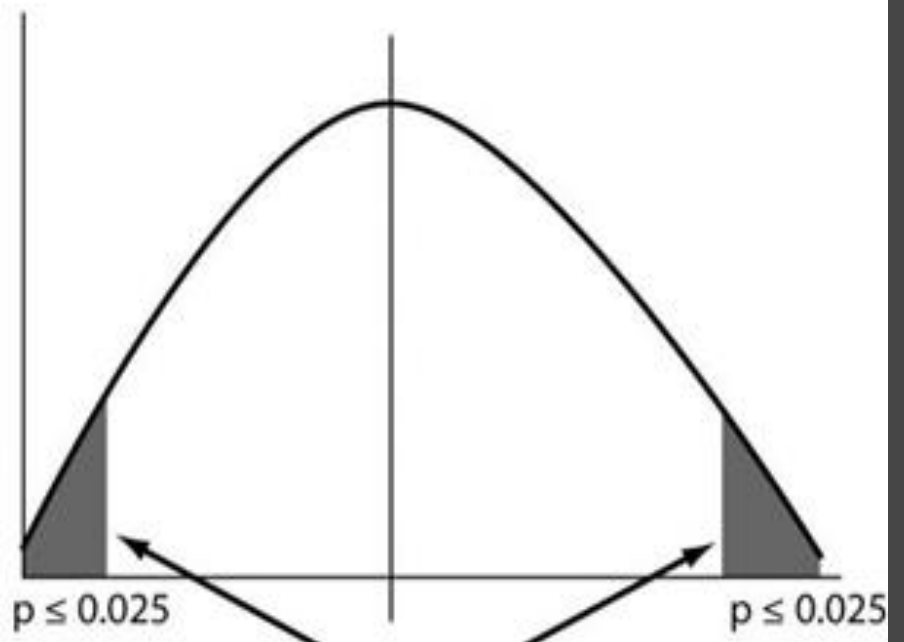
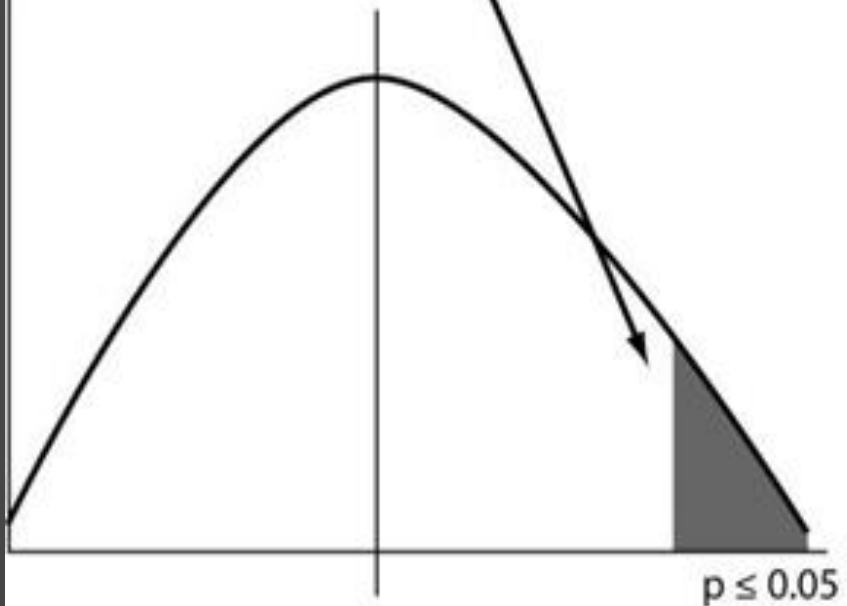
One tailed test :- A test of a statistical hypothesis , where the region of rejection is on only one side of the sampling distribution , is called a one-tailed test.

Example :- a college has ≥ 4000 student or data science $\leq 80\%$ org adopted.

Two-tailed test :- A two-tailed test is a statistical test in which the critical area of a distribution is two-sided and tests whether a sample is greater than or less than a certain range of values. If the sample being tested falls into either of the critical areas, the alternative hypothesis is accepted instead of the null hypothesis.



one-tail **critical region**



two-tail **critical region(s)**

P-value :- The P value, or calculated probability, is the probability of finding the observed, or more extreme, results when the null hypothesis (H_0) of a study question is true — the definition of ‘extreme’ depends on how the hypothesis is being tested.

If your P value is less than the chosen significance level then you reject the null hypothesis i.e. accept that your sample gives reasonable evidence to support the alternative hypothesis. It does NOT imply a “meaningful” or “important” difference; that is for you to decide when considering the real-world relevance of your result.

Example : you have a coin and you don't know whether that is fair or tricky so let's decide null and alternate hypothesis

H_0 : a coin is a fair coin.

H_1 : a coin is a tricky coin. and $\alpha = 5\%$ or 0.05

Now let's toss the coin and calculate p- value (probability value).

Toss a coin 1st time and result is tail- P-value = 50% (as head and tail have equal probability)

Toss a coin 2nd time and result is tail, now $p\text{-value} = 50/2 = 25\%$ and similarly we Toss 6 consecutive time and got result as P-value = 1.5% but we set our significance level as 95% means 5% error rate we allow and here we see we are beyond that level i.e. our null-hypothesis does not hold good so we need to reject and propose that this coin is a tricky coin which is actually.


```
1. from scipy.stats import ttest_1samp
2. import numpy as np
3. ages = np.genfromtxt("ages.csv")
4. print(ages)
5. ages_mean = np.mean(ages)
6. print(ages_mean)
7. tset, pval = ttest_1samp(ages, 30)
8. print("p-values",pval)
9. if pval < 0.05:      # alpha value is 0.05 or 5%
10.  print(" we are rejecting null hypothesis")
11.else:
12.  print("we are accepting null hypothesis")
```

Perform this

<https://towardsdatascience.com/getting-started-to-data-analysis-with-python-pandas-with-titanic-dataset-a195ab043c77>

Getting Data

In order to be a data scientist you need data. In fact, as a data scientist you will spend an embarrassingly large fraction of your time acquiring, cleaning, and transforming data.

In a pinch, you can always type the data in yourself but usually this is not a good use of your time. In this chapter, we'll look at different ways of getting data into Python and into the right formats.

stdin and stdout: If you run your Python scripts at the command line, you can pipe data through them using `sys.stdin` and `sys.stdout`.

For example, here is a script that reads in lines of text and spits back out the ones that match a regular expression.

```
import sys, re
# sys.argv[0] is the name of the program itself
# sys.argv[1] will be the regex specified at the command line
regex = sys.argv[1]
for line in sys.stdin:
    # if it matches the regex, write it to stdout
    if re.search(regex, line):
        sys.stdout.write(line)
```

And here's one that counts the lines it receives and then writes out the count:

```
# line_count.py
import sys
count = 0
for line in sys.stdin:
    count += 1
    # print goes to sys.stdout
print count
```

In Windows, you'd use:

```
type SomeFile.txt | python egrep.py "[0-9]" | python line_count.py
```

The | is the pipe character, which means “use the output of the left command as the input of the right command.” You can build pretty elaborate data-processing pipelines this way.

Similarly, here's a script that counts the words in its input and writes out the most common ones:

```
import sys
from collections import Counter
# pass in number of words as first argument
try:
    num_words = int(sys.argv[1])
except:
    print "usage: most_common_words.py num_words"
    sys.exit(1) # non-zero exit code indicates error
counter = Counter(word.lower() # lowercase words
                  for line in sys.stdin #
                  for word in line.strip().split() # split on spaces
                  if word) # skip empty 'words'
for word, count in counter.most_common(num_words):
    sys.stdout.write(str(count))
    sys.stdout.write("\t")
    sys.stdout.write(word)
    sys.stdout.write("\n")
```

Reading Files

You can also explicitly read from and write to files directly in your code. Python makes working with files pretty simple.

The Basics of Text Files: The first step to working with a text file is to obtain a file object using `open`:

```
# 'r' means read-only
```

```
file_for_reading = open('reading_file.txt', 'r')
```

```
# 'w' is write—will destroy the file if it already exists!
```

```
file_for_writing = open('writing_file.txt', 'w')
```

```
# 'a' is append—for adding to the end of the file
```

```
file_for_appending = open('appending_file.txt', 'a')
```

```
# don't forget to close your files when you're done
```

```
file_for_writing.close()
```

If you need to read a whole text file, you can just iterate over the lines of the file using for:

```
starts_with_hash = 0
```

```
with open('input.txt','r') as f:
```

```
    for line in file: # look at each line in the file
```

```
        if re.match("^#",line): # a regex to check starts is '#'
```

```
            starts_with_hash += 1 # add 1 to the count
```


Delimited Files

More frequently you'll work with files with lots of data on each line.

These files are very often either comma-separated or tab-separated.

Each line has several fields, with a comma (or a tab) indicating where one field ends and the next field starts.

This starts to get complicated when you have fields with commas and tabs and newlines in them. You should use Python's `csv` module (or the `pandas` library).

If your file has no headers (which means you probably want each row as a list, and which places the burden on you to know what's in each column), you can use `csv.reader` to iterate over the rows, each of which will be an appropriately split list.

For example, if we had a tab-delimited file of stock prices:

6/20/2014 AAPL 90.91

6/20/2014 MSFT 41.68

6/20/2014 FB 64.5

6/19/2014 AAPL 91.86

6/19/2014 MSFT 41.51

6/19/2014 FB 64.34

we could process them with:

```
import csv
```

```
with open('tab_delimited_stock_prices.txt', 'rb') as f:
```

```
    reader = csv.reader(f, delimiter='\t')
```

```
    for row in reader:
```

```
        date = row[0]
```

```
        symbol = row[1]
```

```
        closing_price = float(row[2])
```

```
        process(date, symbol, closing_price)
```

You can similarly write out delimited data using `csv.writer`:

```
today_prices = { 'AAPL' : 90.91, 'MSFT' : 41.68, 'FB' : 64.5 }  
with open('comma_delimited_stock_prices.txt','wb') as f:  
    writer = csv.writer(f, delimiter=',')  
    for stock, price in today_prices.items():  
        writer.writerow([stock, price])
```

Scraping the Web

Getting Data from API

why use an API instead of a static CSV dataset you can download from the web? APIs are useful in the following cases:

- 1. The data is changing quickly.** An example of this is stock price data. It doesn't really make sense to regenerate a dataset and download it every minute — this will take a lot of bandwidth, and be pretty slow.
- 2. You want a small piece of a much larger set of data.** Reddit comments are one example. What if you want to just pull your own comments on Reddit? It doesn't make much sense to download the entire Reddit database, then filter just your own comments.
- 3. There is repeated computation involved.** Spotify has an API that can tell you the genre of a piece of music. You could theoretically create your own classifier, and use it to compute music categories, but you'll never have as much data as Spotify does.

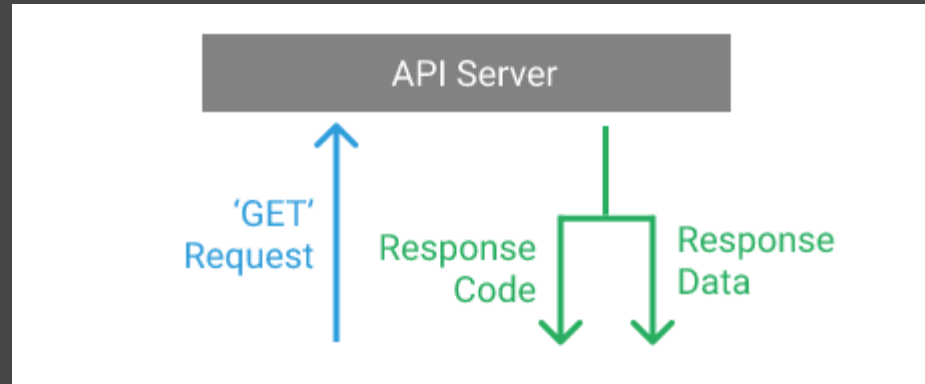
In cases like the ones above, an API is the right solution.

What is an API?

- An API, or Application Programming Interface, is a server that you can use to retrieve and send data to using code. APIs are most commonly used to retrieve data.
- When we want to receive data from an API, we need to make a request. Requests are used all over the web. For instance, when you visit any blog post, your web browser made a request to the web server, which responded with the content of web page.
- API requests work in exactly the same way – you make a request to an API server for data, and it responds to your request.

Making API Requests in Python

- In order to work with APIs in Python, we need tools that will make those requests. In Python, the most common library for making requests and working with APIs is the requests library. The requests library isn't part of the standard Python library, so you'll need to install it to get started.
- There are many different types of requests. The most commonly used one, a GET request, is used to retrieve data. Because we'll just be working with retrieving data, our focus will be on making 'get' requests.
- When we make a request, the response from the API comes with a response code which tells us whether our request was successful. Response codes are important because they immediately tell us if something went wrong.



- To make a 'GET' request, we'll use the `requests.get()` function, which requires one argument — the URL we want to make the request to. We'll start by making a request to an API endpoint that doesn't exist, so we can see what that response code looks like.
- `response = requests.get("http://api.open-notify.org/this-api-doesnt-exist")`
- The `get()` function returns a response object. We can use the `response.status_code` attribute to receive the status code for our request:
- `print(response.status_code)`

API Status Code

Status codes are returned with every request that is made to a web server. Status codes indicate information about what happened with a request.

- 200: Everything went okay, and the result has been returned
- 301: The server is redirecting you to a different endpoint. This can happen when a company switches domain names
- 400: The server thinks you made a bad request.
- 401: The server thinks you're not authenticated. Many APIs require login credentials, so this happens when you don't send the right credentials to access an API.
- 403: The resource you're trying to access is forbidden: you don't have the right permissions to see it.
- 404: The resource you tried to access wasn't found on the server.
- 503: The server is not ready to handle the request.

Good Links to refer

<https://acadgild.com/blog/7-probability-distributions-every-data-science-expert-should-know>

<https://luminousmen.com/post/data-science-probability-distributions>

<https://www.datacamp.com/community/tutorials/probability-distributions-python>

<https://luminousmen.com/post/data-science-bayes-theorem>

<https://www.analyticsvidhya.com/blog/2019/06/introduction-powerful-bayes-theorem-data-science/>

<https://towardsdatascience.com/bayes-theorem-with-example-for-data-science-professionals-55b4d52f8967>

<https://towardsdatascience.com/probability-learning-i-bayes-theorem-708a4c02909a>

<https://data-flair.training/blogs/bayes-theorem-data-science/>

<https://machinelearningmastery.com/bayes-theorem-for-machine-learning/>

<https://www.analyticsjobs.in/machine-learning/bayes-theorem-data-science/>

<https://www.kdnuggets.com/2019/10/bayes-theorem-applied-machine-learning.html>

[vhttps://www.youtube.com/watch?v=VQI5_K7V44g](https://www.youtube.com/watch?v=VQI5_K7V44g)

https://www.youtube.com/watch?v=3OJEae7Qb_o

<https://luminousmen.com/post/data-science-central-limit-theorem>

<https://medium.com/analytics-vidhya/demystifying-statistics-for-data-scientists-central-limit-theorem-ee9d420fc154>