SAYAH FARES

LinkedIn: linkedin.com/fares-sayah

INTRODUCTION TO PROBABILITY

WHAT IS PROBABILITY?

Probability is the likelihood of an event occurring. This event can be pretty much anything – getting heads, rolling a 4 or even bench pressing 225lbs.

We measure probability with numeric values between 0 and 1, because we like to compare the relative likelihood of events. Observe the general probability formula:

$$P(X) = \frac{\text{Preferred outcomes}}{\text{Sample space}}$$

Probability Formula:

- The Probability of event X occurring equals the number of preferred outcomes over the number of outcomes in the sample space.
- Preferred outcomes are the outcomes we want to occur or the outcomes we are interested in. We also call refer to such outcomes as “Favorable”.
- Sample space refers to all possible outcomes that can occur. Its size indicates the number of elements in it.

If two events are independent:

- The probability of them occurring simultaneously equals the product of them occurring on their own.

$$P(A♠) = P(A).P(♠)$$

EXPECTED VALUES

Trial	Observing an event occur and recording the outcome.
Experiment	A collection of one or multiple trials.
Experimental Probability	The probability we assign an event, based on an experiment we conduct.
Expected value	the specific outcome we expect to occur when we run an experiment.

Example:

- *Trial*: Flipping a coin and recording the outcome
- *Experiment*: Flipping a coin 20 times and recording the 20 individual outcomes.
- In this instance, the *experimental probability* for getting heads would equal the number of heads we record over the course of the 20 outcomes, over 20 (the total number of trials).
- The *expected value* can be numerical, Boolean, categorical or other, depending on the type of the event we are interested in. For instance, the *expected value* of the trial would be the more likely of the two outcomes, whereas the expected value of the experiment will be the number of times we expect to get either heads or tails after the 20 trials.

Expected value for categorical variables: $E(x) = n \times p$

Expected value for numeric variables: $E(x) = \sum_{i=1}^{n} x_i \cdot P_i$

PROBABILITY FREQUENCY DISTRIBUTION

- *What is a probability frequency distribution?*

A collection of the probabilities for each possible outcome of an event.

- *Why do we need frequency distributions?*

We need the probability frequency distribution to try and predict future events when the expected value is unattainable.

- *What is a frequency?*

Frequency is the number of times a given value or outcome appears in the sample space.

- *What is a frequency distribution table?*

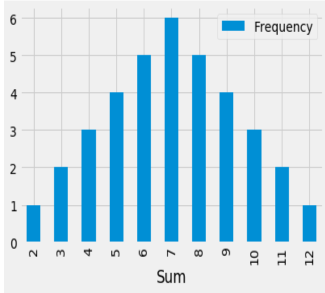
The frequency distribution table is a table matching each distinct outcome in the sample space to its associated frequency.

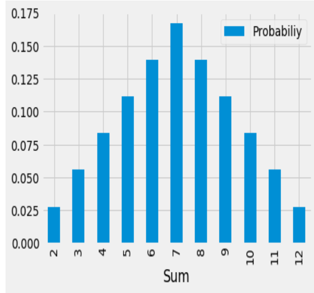
- *How do we obtain the probability frequency distribution from the frequency distribution table?*

By dividing every frequency by the size of the sample space. (Think about the “favored overall” formula.)

Example: Rolling to Dices

Sum	Frequency	Probability
2	1	0.027778
3	2	0.055556
4	3	0.083333
5	4	0.111111
6	5	0.138889
7	6	0.166667
8	5	0.138889
9	4	0.111111
10	3	0.083333
11	2	0.055556
12	1	0.027778





COMPLEMENTS

The complement of an event is everything an event is not. We denote the complement of an event with an apostrophe.

$$A' = \text{Not } A$$

Characteristics of complements:

- Can never occur simultaneously.
- Add up to the sample space. $(A + A' = \text{Sample space})$
- Their probabilities add up to 1. $(P(A) + P(A') = 1)$
- The complement of a complement is the original event. $((A')' = A)$

Example:

- Assume event A represents drawing a spade ♠, so $P(A) = 0.25$.
- Then, A’ represents not drawing a spade, so drawing a club ♣, a diamond ♦ or a heart ♥, $P(A') = 1 - P(A)$, so $P(A') = 0.75$.

PERMUTATIONS

Permutations represent the number of different possible ways we can arrange a number of elements.

$$P(n) = n! = n \times (n-1) \times (n-2) \times \cdots \times 1$$

Characteristics of Permutations:

- Arranging all elements within the sample space.
- No repetition.
- $P(n) = n \times (n-1) \times (n-2) \times \cdots \times 1 = n!$ (Called "n factorial")

Example:

- If we need to arrange 5 people, we would have $P(5) = 120$ ways of doing so.
- Form a 3 digits number without repetition using 3 numbers (1, 2, 3), We have $3 \times 2 \times 1 = 6$ possibilities.

FACTORIALS

Factorials express the product of all integers from 1 to n and we denote them with the "!" symbol.

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 1$$

Key Values:
- $0! = 1$.
- If $n < 0$, $n!$ does not exist.

VARIATIONS

Variations represent the number of different possible ways we can pick and arrange a number of elements.

1. Variations with repetition:

$$\bar{V}(n,p) = n^p$$

- n: the total number of elements, we have available.
- p: the number of positions we need to fill

Intuition behind the formula:

- We have n-many options for the first element.
- We still have n-many options for the second element because repetition is allowed.
- We have n-many options for each of the p many elements.

Example:

- Form 3 digits number using 5 numbers (1, 2, 3, 4, 5) and repetition is allowed (111, 122…).
- We have $5^3 = 125$ possible ways.

2. Variations without repetition:

$$\bar{V}(n,p) = \frac{n!}{(n-p)!}$$

- We have n-many options for the first element.
- We only have (n-1)-many options for the second element because we cannot repeat the value for, we chose to start with.
- We have less options left for each additional element.

Example:

Form 3 digits number and numbers can't be repeated using 5 numbers (1, 2, 3, 4, 5). We have $\frac{5!}{(5-3)!} = 60$ possible ways.

COMBINATIONS

Combinations represent the number of different possible ways we can pick a number of elements.

$$C(n,p) = \frac{n!}{(n-p)!\,p!}$$

Characteristics of Combinations:

- Takes into account double-counting.
- All the different permutations of a single combination are different variations.
- Combinations are symmetric, so $C_p^n = C_{n-p}^n$, since selecting p elements is the same as omitting n-p elements.

Example:

- In how many ways can a coach choose 3 swimmers from among 5 swimmers?
- The coach can choose the swimmers in $\frac{5!}{(5-3)!3!} = 10$ ways.

COMBINATIONS WITH SEPARATE SAMPLE SPACES

Combinations represent the number of different possible ways we can pick a number of elements.

$$C = n_1 \times n_q \ldots \times n_p$$

Characteristics of Combinations with separate sample spaces:

- The option we choose for any element does not affect the number of options for the other elements.
- The order in which we pick the individual elements is arbitrary.
- We need to know the size of the sample space for each individual element. ($n_1, n_2 \ldots n_p$)

COMBINATIONS WITH REPETITION

In special cases we can have repetition in combinations and for those we use a different formula:

$$\bar{C}(n,p) = \frac{(n+p-1)!}{(n-1)!\,p!}$$

BAYESIAN NOTATION

A set is a collection of elements, which hold certain values. Additionally, any event has a set of outcomes that satisfy it.

The null-set (or empty set), denoted $\emptyset$, is a set which contain no values.

Notation	Interpretation	Example
$x \in A$	Element x is a part of set A.	$2 \in$ *All even numbers*
$A \ni x$	Set A contains element x.	*All even numbers* $\ni 2$
$x \notin A$	Element x is NOT a part of set A.	$1 \notin$ *All even numbers*
$\forall x$:	For all/any x such that…	$\forall x$: $x \in$ *All even number*
$A \subseteq B$	A is a subset of B	*Even numbers* $\subseteq$ *Integers*

Every set has at least 2 subsets:
- $A \subseteq A$
- $\emptyset \subseteq A$

MULTIPLE EVENTS

The sets of outcomes that satisfy two events A and B can interact in one of the following 3 ways.

1. Not touch at all.
2. Intersect (Partially Overlap)
3. One completely overlaps the other.

INTERSECTION

The intersection of two or more events expresses the set of outcomes that satisfy all the events simultaneously. Graphically, this is the area where the sets intersect.

We denote the intersection of two sets with the intersect sign: $A \cap B$.

Example:

Event A: ♥, Event B: Q $\Rightarrow$ $A \cap B =$ QUEEN OF HEARTS

UNION

The union of two or more events expresses the set of outcomes that satisfy at least one of the events. Graphically, this is the area that includes both sets.

$$A \cup B = A + B - A \cap B$$

Example:

Event A: ♥, Event B: Q $\Rightarrow A \cup B = 13(\text{all HEARTS}) + 4(\text{QUEEN}) - 1(\text{There intersection})$

MUTUALLY EXCLUSIVE SETS

Sets with no overlapping elements are called mutually exclusive. Graphically, their circles never touch.

If $A \cap B = \emptyset$, then the two sets are mutually exclusive.

All complements are mutually exclusive, but not all mutually exclusive sets are complements.

Example:

Event A: ♥, Event B: ♦ $\Rightarrow A \cap B = \emptyset$

INDEPENDENT AND DEPENDENT EVENTS

If the likelihood of event A occurring ($P(A)$) is affected event B occurring, then we say that A and B are dependent events. Alternatively, if it isn't – the two events are independent.

We express the probability of event A occurring, given event B has occurred the following way $P(A|B)$. We call this the conditional probability.

Example:

Normally	$\Rightarrow P(Q♠) = 1/52$
♠	$\Rightarrow P(Q♠) = 1/13$
Q	$\Rightarrow P(Q♠) = 1/4$

The probability of an event changes depending on the information we have.

Independent:	Dependent:
All the probabilities we have examined so far. The outcome of A does not depend on the outcome of B. $P(A\|B) = P(A)$	New concept. The outcome of A depends on the outcome of B. $P(A\|B) \neq P(A)$

Example:

- Independent: A -> Hearts, B -> Jacks
- Dependent: A -> Hearts, B -> Red

CONDITIONAL PROBABILITY

The likelihood of an event occurring, assuming a different one has already happened.

For any two events A and B, such that the likelihood of B occurring is greater than 0 ($P(B) > 0$), the conditional probability formula states the following.

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

- Only interested in the outcomes where B is satisfied.
- Only the elements in the intersection would satisfy A as well.
- Unlike the union or the intersection, changing the order of A and B in the conditional probability alters its meaning, $P(A|B)$ is not the same as $P(B|A)$, even if $P(A|B) = P(B|A)$ numerically.

Example: Event A: Q♠, Event B: ♠, Event C: Q

- $P(A) = 1/52$
- $P(A|B) = 1/13$
- $P(A|C) = 1/4$

LAW OF TOTAL PROBABILITY

The law of total probability dictates that for any set A, which is a union of many mutually exclusive sets $B_1, B_2, \ldots, B_n$, its probability equals the following sum.

$$P(A) = P(A|B_1) \times P(B_1) + P(A|B_2) \times P(B_2) + \cdots + P(A|B_n) \times P(B_n)$$

Intuition behind the formula:

- Since P(A) is the union of mutually exclusive sets, so it equals the sum of the individual sets.
- The intersection of a union and one of its subsets is the entire subset.
- We can rewrite the conditional probability formula $P(A|B) = \frac{P(A \cap B)}{P(B)}$ to get $P(A \cap B) = P(A|B) \times P(B)$
- Another way to express the law of total probability is: $P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n)$

Example: Vegetarian Survey of 100 men and women

	Not Vegetarian	Vegetarian	Total
Women	32	15	47
Men	24	29	53
Total	56	44	100

- A $\Rightarrow$ Vegetarian, B $\Rightarrow$ Women, C $\Rightarrow$ Men
- The likelihood of a women being vegetarian:
  $$P(A|B) = 15/47.$$
- The likelihood of a vegetarian being women:
  $$P(B|A) = 15/44.$$
- $\Rightarrow P(A|B) \neq P(B|A)$
- The law of total probability:

$$P(A) = P(A|B) \times P(B) + P(A|C) \times P(C)$$

$$P(A) = \frac{29}{53} \times \frac{53}{100} + \frac{15}{47} \times \frac{47}{100} = 0.44$$

There is a 44% chance of someone being vegetarian.

ADDITIVE LAW

The probability of the union of two sets is equal to the sum of the individual probabilities of each event, minus the probability of their intersection.

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

The additive law calculates the probability of the union based on the probability of the individual sets it accounts for.

- Recall the formula for finding the size of the Union using the size of the Intersection:
  $$A \cup B = A + B - A \cap B$$
- The probability of each one is simply its size over the size of the sample space.
- This holds true for any events A and B.

Example: Vegetarian Survey of 100 men and women

- ❖ A ⇒ Vegetarian, B ⇒ Women
  $$P(B \cup A) = P(B) + P(A) - P(B \cap A)$$
  $$= 0.47 + 0.44 - 0.15 = 0.76$$
- ❖ There is a 76% chance that a random person from the survey is either female, vegetarian or both.

THE MULTIPLICATION RULE

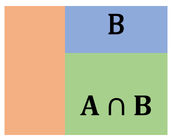
The multiplication rule calculates the probability of the intersection based on the conditional probability.

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

- We can multiply both sides of the conditional probability formula $P(A|B) = \frac{P(A \cap B)}{P(B)}$ to get $P(A \cap B) = P(A|B) \times P(B)$
- If event B occurs in 40% of the time ($P(B) = 0.4$) and event A occurs in 50% of the time B occurs ($P(A|B) = 0.5$), then they would simultaneously occur in 20% of the time ($P(A|B) \times P(B) = 0.5 \times 0.4 = 0.2$).

Example:

- ❖ $P(B) = 0.5$ ⇒ Event B appears 50% of the times.
- ❖ $P(A|B) = 0.8$ ⇒ Event A appears in 80% of those 50% when B occurred
- ❖ $P(A \cap B) = 0.8 \times 0.5 = 0.4$ ⇒ Event A and B appears simultaneously.



BAYES' LAW

Bayes' Law helps us understand the relationship between two events by computing the different conditional probabilities. We also call it Bayes' Rule or Bayes' Theorem.

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

- According to the multiplication rule $P(A \cap B) = P(A|B) \times P(B)$, so $P(B \cap A) = P(B|A) \times P(A)$.
- Since $P(A \cap B) = P(B \cap A)$, we plug in $P(B|A) \times P(A)$ for $P(A \cap B)$ in the conditional probability formula $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- Bayes' Law is often used in medical or business analysis to determine which of two symptoms affects the other one more.
- Helps us make more reasonable arguments about which one causes the other.

Example:

Medical research:

- ❖ There is certain correlation between patient with Back pain (BP) OR vision impairment (VI), $P(VI|BP) = 67\%$, $P(BP|VI) = 41\%$
- ❖ 67% of patient with Back pain wear glasses, 41% of patient who wear glasses have back pain.
- ❖ Even if we can't find a direct causal link, there exist some arguments to support such claims.

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.

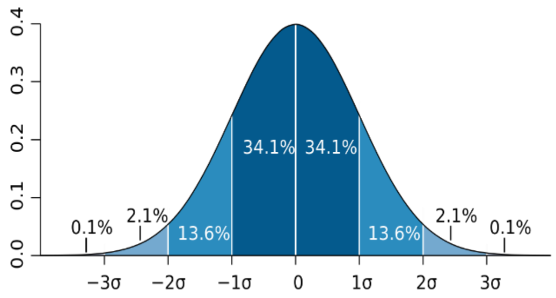
	Population	Sample
Mean	μ	$\bar{x}$
Variance	σ^2	s^2
Standard Deviation	σ	s

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**.

Characteristics:

- Mean: average values (μ)
- Variance: How spread out the data is (σ^2).
- STD (Standard Deviation): square root of the variance ($\sqrt{\sigma^2}$)



- The more congested the middle of the distribution, the more data falls within that interval.
- The less data falls within the interval, the more dispersed the data is.

DISCRETE DISTRIBUTIONS

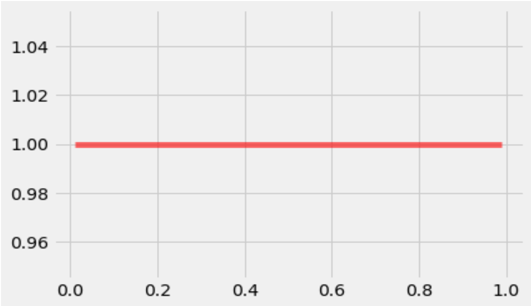
Discrete Distributions have finitely many different possible outcomes. They possess several key characteristics which separate them from continuous ones.

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.
- $P(Y \leq y) = P(Y < y + 1)$

1. Uniform Distribution

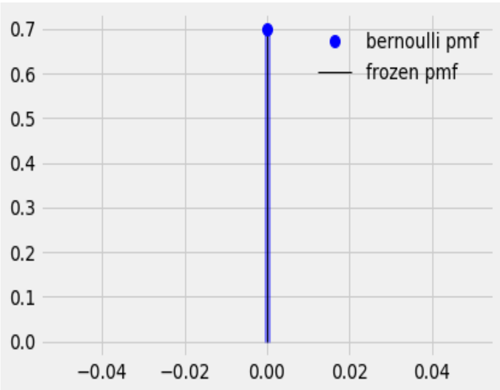
A distribution where all the outcomes are equally likely is called a Uniform Distribution.



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.

2. Bernoulli Distribution

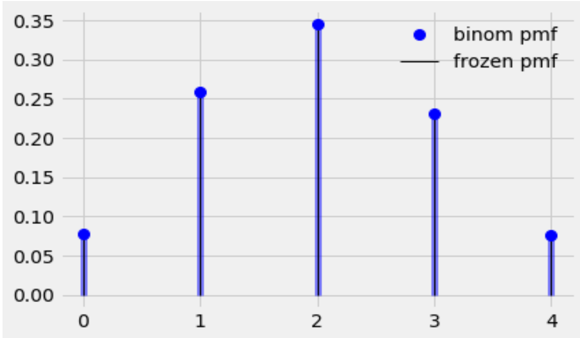
A distribution consisting of a single trial and only two possible outcomes, success or failure is called a Bernoulli Distribution.



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.

3. Binomial Distribution

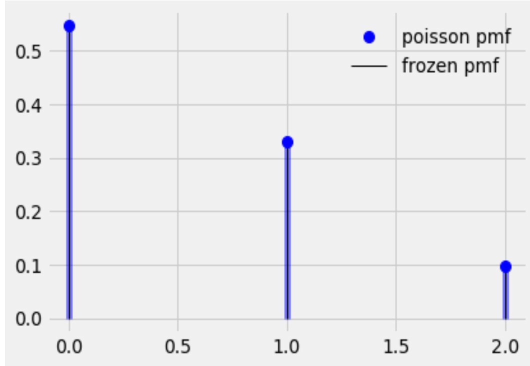
A sequence of identical Bernoulli events is called Binomial and follows a Binomial Distribution.



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 ahead 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.

4. 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.) • $P(Y = y) = \frac{\lambda^y}{y! e^{-\lambda}}$ • $E(Y) = \lambda$ • $Var(Y) = \lambda$
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 potential values that a random variable can take are an infinite series of consecutive values, we are dealing with a continuous distribution.

Key characteristics of Continuous Distributions:

- 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.
- $P(Y = y) = 0$ for any individual value y.
- $P(Y < y) = P(Y \leq y)$

1. Normal Distribution

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

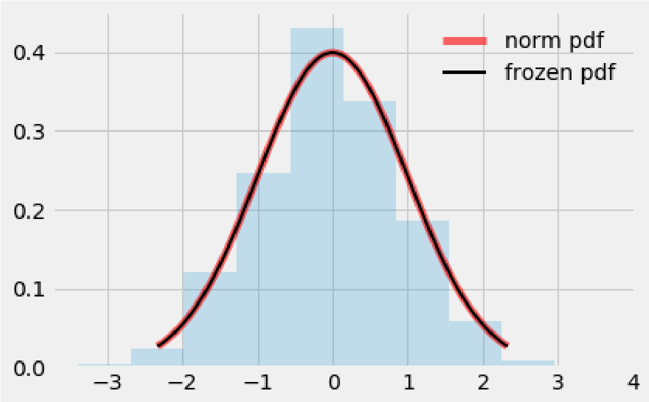
Notation	$Y \sim N(\mu, \sigma^2)$
Key characteristics	• Its graph is bell-shaped curve, symmetric and has thin tails. • $E(Y) = \mu$ • $Var(Y) = \sigma^2$ • 68% of all its values should fall in the interval: $(\mu - \sigma, \mu + \sigma)$
Example and uses	• Often observed in the size of animals in the 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.

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

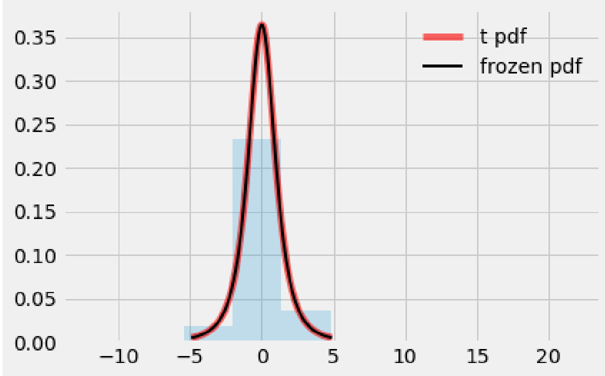
Importance of the Standard Normal Distribution:

- 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.



2. Students' T Distribution

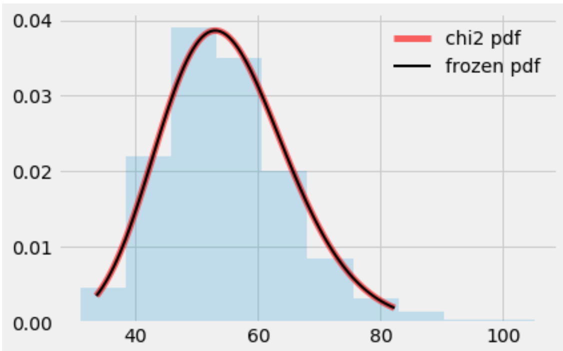
A Normal Distribution represents a small sample size approximation of a Normal Distribution.



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.

3. Chi-Squared Distribution

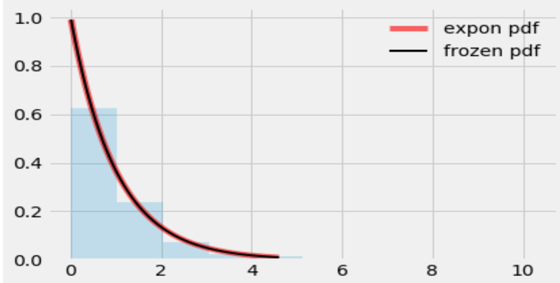
A Chi-Squared distribution is often used.



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

4. Exponential Distribution

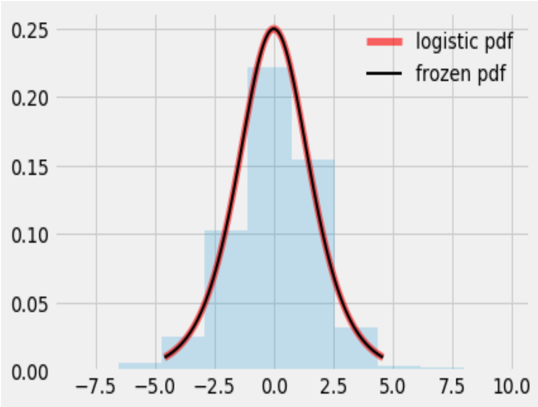
The Exponential Distribution is usually observed in events which significantly change early on.



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.

5. 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	$Y \sim \text{Logistic}(\mu, s)$
Key characteristics	• A small sample size approximation of a Normal Distribution. • $\mathrm{Var}(Y) = \frac{s^2 \times \pi^2}{3}$ • $E(Y) = \mu$ • 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.

◘ PROBABILITY IN OTHER FIELDS

1. Probability in finance

Probability has big uses in finance. In finance usually, we are trying to predict future prices of uncertain events. We take as an example the option pricing, which is an agreement between two parties for the price of stock or item at a future point in time. It allows one of the sides to decide.

Example: Google stock

Let say you want to buy 10 stocks, $ 1,100 each in a week. We have 40% that the stock's price will increase to $1,200, and 60% that the stock's price will decrease to $1,000. You pay an expert $100 for consultations.

40% ⇒ $1,200	You would take advantage of the deal you struck.	Call ⇒ $900 Not Call ⇒ -$100
60% ⇒ $ 1,000	Better off buying the 10 stocks at the market price.	Call ⇒ -$1,100 Not Call ⇒ -$100

Expected Payoffs

- **E(P) < 0** Disadvantageous (Avoid buying this option)
- **E(P) = 0** Fair deal (You expect to make as much as you paid)
- **E(P) > 0** Favorable (Go through with the deal)

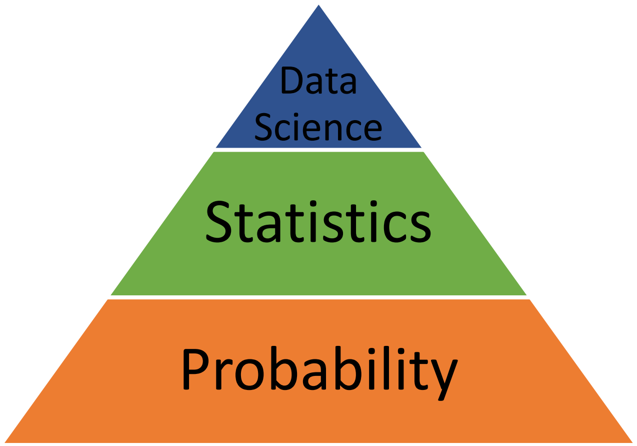
In our example:

E(P) = 0.6 × (-100) + 0.4 × (900) = $300

E(P) = $300 > 0, This deal is **Favorable** for us.

2. Probability in statistics

Statistics focus predominantly on samples and incomplete data which bring some uncertainty. This uncertainty leads us to relays on certain probability concepts like expected events or prediction intervals.



3. Probability in Data Science

The same way probability sets the foundations for statistics, statistics constructs the pillars on which data science is built.

Data science is an expansion of probability, statistics, and programming that implements computational technology to solve more advanced questions. That why it is fundamental to understand probability.

In data science we usually try to analyze past data and use some insight we find to make a reasonable prediction about the future.