Probability Theory for Data Analytics

Dr. Cahit Karakuş

2024 – İsatnbul

The purpose of the course:

Learning to look through the window of probability and statistics has become important. Because, students who have gained skills in professional applications, who can solve problems and make comments with the help of data analytics and data science should be trained. The traces of change, deviation and transformation from the behavior of the systems shown by the data mass; the direction and intensity of its trajectory should be determined. In order to predict the traces of change in its behavior or deviations in its trajectory, the skills that measure, digitize, store and compare and classify the data should be acquired. Laplace explained the theory of probability as follows: Equations established to calculate the probability of a situation do not ensure certainty of the result, they only serve to find the result with the least margin of error; in other words, they try to minimize the margin of error, not eliminate it, because it is not right to make a perfect system; systems that progress towards perfection need to be developed.

Course Content:

- Variables and functions drawing, derivative, integral, limit operations and interpretations
- Correlation and Regression
- Statistical Data Analysis
- Probability
- Random Variables
- Inferences on Probability Distribution
- Markov Chain Analysis
- Algorithmic Analysis
- Applications on Data Analytics

The course is not based on mathematics but on arithmetic. Analytical calculation and interpretation are emphasized. What you need to know:

1) Decimal arithmetic: Addition, Subtraction, Multiplication, Division, Exponentiation, Integer Division, Division with Carry; Square root, Rounding
2) Binary Arithmetic and logical operations in binary number system (bit: 0/1)
3) Comparison: <,>,=, ≠, ≤, ≥
4) Ratio, Percentage
5) Exponential operations
6) Factorial
7) Linear equations, Matrix, Vector, eigenvalues, eigenvectors
8) Complex numbers
9) Trigonometry
10) Logarithmic operations
11) Basic Functions: Sinusoidal, Exponential, Linear, Polynomial
12) Simplifying and finding results by substituting values in equations

İçindekiler

Introduction

If you evaluate the mass of information without analyzing it, you will get into such trouble that you will sink as you try to get out of it. Nowadays, the decisive power of information, which has strategic importance, is gradually increasing. Collecting information has strategic importance in the struggle for sovereignty (egemenlik mücadelesi). Those who do not want to understand the power of knowledge try to find their way and direction by groping in the dark.

On the other hand, the size of the collected information is growing so rapidly that it is becoming difficult to determine the necessary information from the mass of data. After the information is determined, processing, classifying, accessing the information in a timely manner, and to achieve valuable results by analyzing it have become very important. The vital point is to be able to find and extract the implicit or covert information that no one notices and no one can think of from the mass of data that is in front of everyone's eyes.

It is to be able to determine the direction and intensity of the traces of change, deviation and transformation from the behavior shown by the data mass. In order to learn to make predictions from the traces or trajectory of the behavior, it is necessary to gain the skills to measure, digitize, store and compare and classify the data. Laplace explained the theory of probability as follows: The equations established to calculate the probability of a situation did not provide certainty about the result, they only served to find the result with the least margin of error; in other words, they tried to minimize the margin of error, not eliminate it,

because it is not right to make a perfect system; systems that progress towards perfection need to be developed.

Making decisions is also making choices by purchasing risks. If making predictions is only and solely based on winning, the devastating effect of losses will be much greater than expected. In order to make sound predictions, the functions of the processes should be constantly measured, information should be collected, necessary calculations should be made, the results should be compared and analyzed and decisions should be made. Correct evaluation of the collected information is possible by developing skills that can make interpretations based on statistical calculations. Observations are almost always subject to random errors. Therefore, statistical methods should be used to collect and analyze data.

On the other hand, it should not be forgotten that numbers do not lie, but liars use statistical numbers very well. Data analysis and artificial intelligence applications should be developed from the data mass.

1. Measurement and Evaluation

Measurement is the expression of whether an system has a certain feature or not, and if it has a certain feature, its features are expressed with numbers. Measurement is the values to which the results are compared. Evaluation is the process of comparing the measurement results with a criterion and making decisions about their qualities.

1.1. Data

The aim is to transform data into wisdom. Symbols and signals are converted into binary numbering system also these are symbols. Information is obtained by processing data, and the journey from knowledge to wisdom, from wisdom to skill development and experience; from skill development to awareness begins.

Symbols (Signals, Pictures, Shapes, ...): Symbols that carry messages are represented by signals. Human beings communicate with each other or with computer systems using symbols. Symbols are numbers, words, text, images, shapes, documents, video and audio. Symbols and signals transfer to the computer's memory as binary numbering system (bit: 1/0). They contain messages. When symbols are converted to the binary number system (bit:0/1). Calculation, storage and communication are possible with the binary number system in the computer system.

Data are facts or pieces of information that have not gained meaning, have not been associated, have not been assimilated and have not been processed. They are in forms devoid of any content. Sometimes they are a physical event, uninterpreted observations. They do not carry any comment but are ready to be processed. They are not effective in decision making.

Big Data: The concept of "Big Data" has emerged with the great increase in data in terms of speed, variety and capacity (volume) today and with the support of technology to this increase and the production of new solutions.

Information: What, who, when and where are the questions that need to be answered. Information is processed, organized and meaningful data. When necessary, it is classified, represented and clustered with the coefficients of mathematical equations. Information is organized, meaningful and useful data. During the output phase, the information created is put into a presentation form with printed reports, graphics, documents and visuals. The information is stored in the computer for future use. Generally, information emerges through the interpretation of data.

Knowledge: The answer to the question of how. It is to increase performance in decision making, prediction and search for the truth. It is to ensure continuity of learning.

Understand (Understanding - Consciousness): Evaluation of Why, Why questions. It is becoming conscious by understanding, grasping, feeling. It is sharing of knowledge.

Wisdom: Deep, comprehensive, holistic knowledge that not everyone can reach. It is an evaluated understanding. It is making decisions and interpreting by questioning, making predictions.

Data preparation is the process of making raw data suitable for later processes and analyses. The basic steps include collecting raw data, cleaning it, and labeling it in a way suitable for machine learning (ML) algorithms, and then discovering and visualizing the data.

Steps in the data preparation process:
1- Data collection: Relevant data is collected from operational systems, data warehouses, data lakes, and other data sources. During this step, data scientists must verify that the planned analytical applications are suitable for their goals.

2- Data discovery and profiling: Data profiling is used to investigate the collected data to determine what it contains and prepare it for the intended use, to identify patterns,

relationships, and other attributes in the data, as well as inconsistencies, anomalies, missing values, noise, incorrect values, and other issues, and to address these.

3- Data cleaning: Complete and accurate data sets are created by correcting identified data errors and issues. For example, data cleaning removes or corrects incorrect data, fills in missing values, and reconciles inconsistent entries.

4- Data structuring: At this point, the data needs to be modeled and organized to meet analytical requirements. For example, data stored in comma-separated values (CSV) files or other file formats needs to be converted into tables to be accessible to analysis tools.

5- Data transformation and enrichment: In addition to being structured, data typically needs to be converted into a unified and usable format. For example, data transformation can involve creating new fields or columns that collect values from existing fields. Datasets are developed and optimized as needed through measures such as data enrichment, augmentation, and addition of data.

6- Data validation and publishing: Automated routines are run against the data to verify the consistency, completeness, and accuracy of the data. The prepared data is then stored in a data warehouse, data lake, or other repository and either used directly by the person who prepared it or made available to other users.

7- Splitting into training and test datasets

8- Modeling

1.2. Numbering Systems

Data types, classes, aggregation, and equationalization are the building blocks of programming. They tell the computer how to interpret and store data. The most common data types are integers, strings, booleans, dates, arrays, and objects. Integers are whole numbers, strings are text, booleans can be true or false, date and time data types can represent a specific point in time, and objects can store multiple pieces of data.

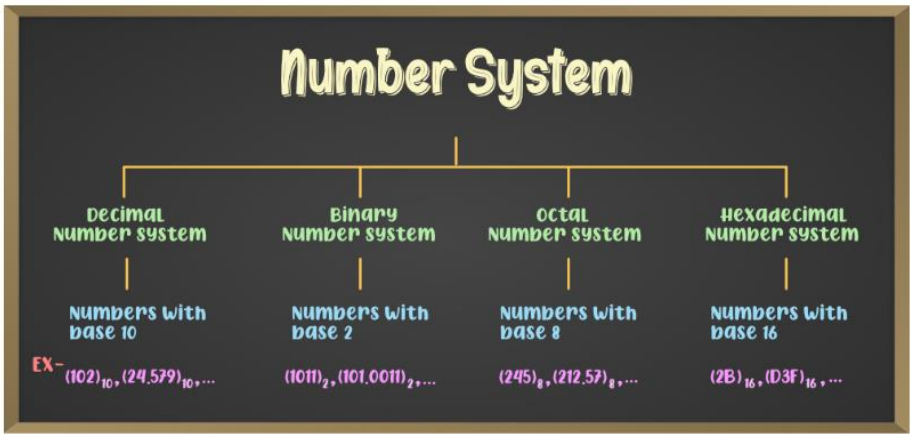
Number System is a method of representing numbers on the number line with the help of a set of Symbols and rules. These symbols range from 0-9 and are termed as digits.

Number System in Maths

Number system in Maths is a writing system for expressing numbers. It is a mathematical notation for representing numbers of a given set, consistently using digits or other symbols.

It allows us to perform arithmetic operations like addition, subtraction, multiplication, and division.
Numbers, letters, arithmetic symbols conversion among different systems: ASCII Coding

Natural Numbers: 1, 2, 3 …
Integers: -3, -2, -1, 0, 1, 2, 3
Rational Numbers: 17, -½, -0.65, 0.333
Real Numbers
Complex Numbers
string — text (“Lorem ipsum”, ...)
boolean — true, false
date — a specific point in time (February 7, 2023)
object — { name: "John", age: "19" }



The four common types of Number System are:

1. Decimal Number System
2. Binary Number System
3. Octal Number System
4. Hexadecimal Number System

Decimal Number System

Number system with base value 10 is termed as Decimal number system. It uses 10 digits i.e. 0-9 for the creation of numbers.
Here, each digit in the number is at a specific place with place value a product of different powers of 10. The place value is termed from right to left as first place value called units, second to the left as Tens, so on Hundreds, Thousands, etc. Here, units has the place value as 100, tens has the place value as 101, hundreds as 102, thousands as 103, and so on.
For example: 10285 has place values as
(1 × 104) + (0 × 103) + (2 × 102) + (8 × 101) + (5 × 100)

1 × 10000 + 0 × 1000 + 2 × 100 + 8 × 10 + 5 × 1
10000 + 0 + 200 + 80 + 5
10285
Decimal: 0 … 9 (Decimal): Decimal number, Fractional number, Positive or negative number, rational or irrational number.

Binary Numbering:
Number System with base value 2 is termed as Binary number system. It uses 2 digits i.e. 0 and 1 for the creation of numbers. The numbers formed using these two digits are termed as Binary Numbers.
Binary number system is very useful in electronic devices and computer systems because it can be easily performed using just two states ON and OFF i.e. 0 and 1.
Decimal Numbers 0-9 are represented in binary as: 0, 1, 10, 11, 100, 101, 110, 111, 1000, and 1001
Examples:
14 can be written as 1110
19 can be written as 10011
50 can be written as 110010

Binary: Bit: 0 or 1
Byte: 8 bits

Octal Number System
Octal Number System is one in which the base value is 8. It uses 8 digits i.e. 0-7 for creation of Octal Numbers. Octal Numbers can be converted to Decimal value by multiplying each digit with the place value and then adding the result. Here the place values are 80, 81, and 82. Octal Numbers are useful for the representation of UTF8 Numbers. Octal: 3 bits (8), 0 … 7

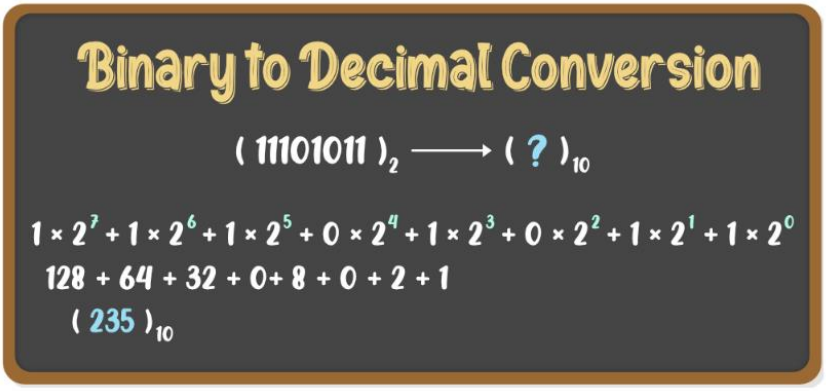
Example:
(135)10 can be written as (207)8
(215)10 can be written as (327)8

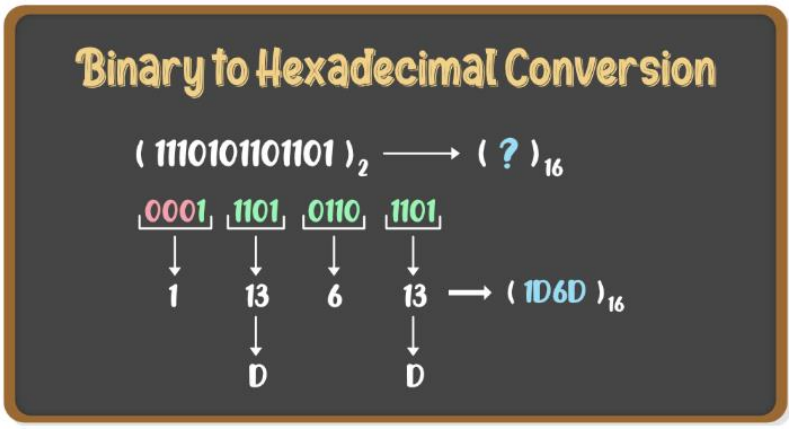
Hexadecimal Number System
Number System with base value 16 is termed as Hexadecimal Number System. It uses 16 digits for the creation of its numbers. Digits from 0-9 are taken like the digits in the decimal number system but the digits from 10-15 are represented as A-F i.e. 10 is represented as A, 11 as B, 12 as C, 13 as D, 14 as E, and 15 as F. Hexadecimal Numbers are useful for handling memory address locations. Hexadecimal: 4 bits, 0 …. 15, $\{0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F\}$ (Hexadecimal number system)

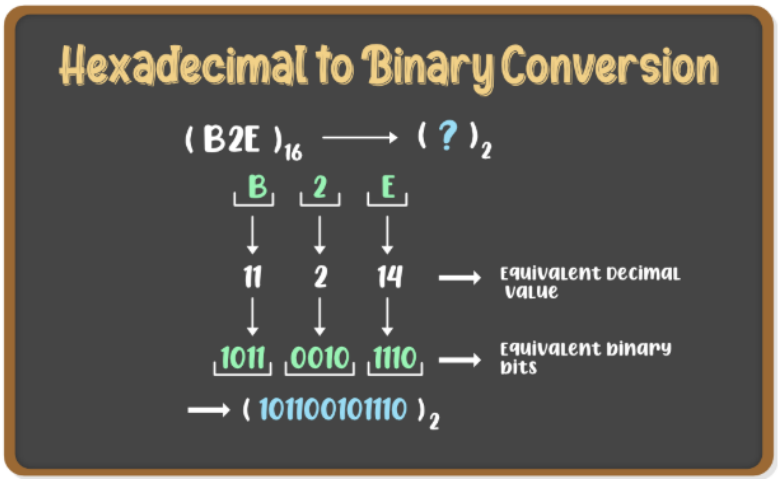
Examples:
(255)10 can be written as (FF)16
(1096)10 can be written as (448)16
(4090)10 can be written as (FFA)16

Conversion from Decimal to Other Number Systems

Decimal Numbers are represented with digits 0-9 and with base 10. Conversion of a number system means conversion from one base to another. Following are the conversion of the Decimal Number System to other Number Systems:







Natural numbers are the numbers that start from 1 and end at infinity. In other words, natural numbers are counting numbers and they do not include 0 or any negative or fractional numbers. For example, 3, 6, 57, 973, 4000, and so on. *Natural numbers are the counting numbers: 1, 2, 3, 4, 5, and so on. They begin at 1, unlike whole numbers, which start at 0.*

Data Types in the software:

Data Types are used to define the type of a variable. It defines what type of data we will store in a variable. Data stored in memory can be of many types. For example, a person's age is stored as a numeric value, and their address is stored as alphanumeric characters.

- Numeric - int, float, complex, bool
- String - str
- Sequence - list, tuple, range
- Binary - bytes, bytearray, memoryview
- Mapping - dict
- Boolean - bool
- Set - set, frozenset
- None - NoneType
- int (signed integers)
- bool (subtype of integers.)
- float (floating point real values)
- complex (complex numbers)

String (string) is a collection of characters, digits, letters and symbols, it is one of the object types and the name of a class. A string object has a sequence of characters enclosed in double quotes, such as "Computer Engineering".

Immutable data types: Number, String, Tuple.

Variable data types: List, Dictionary, Set.

Integer (int), Short integer, Long integer: Integers are whole numbers, and integer variables are used when you know there will never be anything after the decimal point, e.g. if you are writing a lottery ball generator, all the balls have whole numbers on them. The difference between short integers, integers, and long integers is the number of bytes used to store them. This varies depending on the operating system and hardware you are using, but these days you can assume that an integer will be at least 16 bits, and a long integer will probably be at least 32 bits. It is more efficient to use long integers (i.e. a complete word), and many compilers will automatically use long integers unless you specify a short integer.

Float, Single, Double, Real: Floating-point numbers are numbers that have fractional parts - that is, they are not whole numbers. Single and double quantifiers are similar to the short and long quantifiers used with integers - that is, they indicate how many bits are used to store the variable. Floating-point arithmetic can cause problems with rounding and precision, so if you're dealing with a limited number of decimal places, it's probably more efficient to use integers and multiply all your values by a power of 10. If you're dealing with money, it's probably better to work in pence and use integers than to work in pounds and use floating-point variables.

Char: A char variable is a common sight in programs (that can't handle strings) and is used to store a single character of text. The value it actually stores is an integer representing the code (e.g. ASCII) for the character being represented. The ASCII code is 8 bits; it represents each key on the keyboard.

Boolean: A boolean variable can store one of two values, TRUE or FALSE. Like char, it's usually an integer - for example in VisualBASIC, FALSE is 0 and TRUE is 1, and the values TRUE and FALSE are themselves constants.

Arrays (Vectors, matrices): An array is a collection of variables of the same type with the same name, differentiated by a numbered index. Note that the array normally starts at zero, not one. The useful thing about arrays is that you can use the same code to process all the elements in the array using a loop.

Arrays can also be used to implement a simple lookup table. For example, if you wanted to randomly generate a day of the week, you could use an array of strings containing the days of the week. You could then create an integer between 0 and 6 and use that as an index to return the day of the week.

Irrational numbers:
Irrational numbers are numbers that cannot be expressed with integer fractions in any way. It is not possible to find all irrational numbers. However, there is a way to separate irrational

numbers from rational numbers. Numbers given as integers such as 1 – 10 – 256 – 38975 and repeating decimal numbers are rational numbers. Irrational numbers do not have repeating parts. The numbers in their decimal parts do not repeat themselves in a certain pattern.

Irrational numbers,
• The number 'Pi' is an irrational number.
• Square root numbers that are not perfect squares are irrational numbers.
• A decimal number that becomes irrational as a result of the square root with a comma after 0.
• A single digit left to the right of the comma is an irrational number.

1.3. Constants

The constants are the value that never changes. Because of their inflexibility, constants are used less than variables in programming.

A constant can be:
–a number, like 25 or 3.6
–a character, like a,b,c or $
–a character string, like "this is a string"

Programming languages may contain their own predefined constants. In programming languages, constants are values stored in memory like variables.

İnt: a=123
Float:a=9.99
Complex:a=5+6j

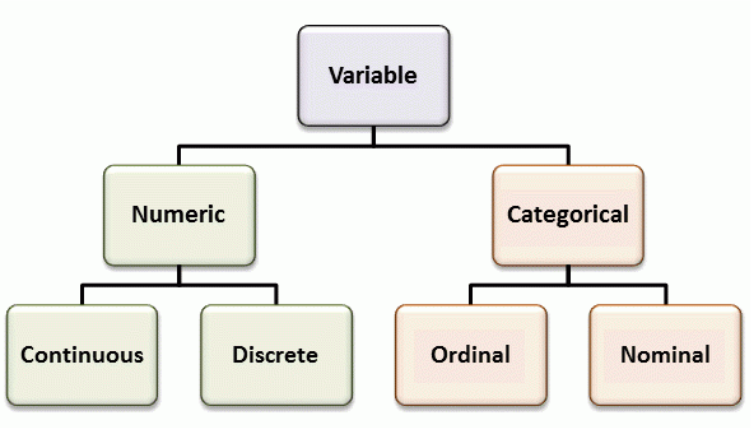
Example: Calculate area of circle
area = 3.142 * r * r
Variable: area, r
Constant: 3.142

1.4. Variables

A variable is a symbol or name that represents a value, a value that changes. A variable is any characteristic, number, quantity, or attribute that can be measured or counted. Examples of variables include age, gender, business income and expenses, country of birth, capital expenditure, class grades, eye color, and vehicle type.

Variables provide temporary storage for data in a computer program or application. Variables are used to store data in a program or to describe stored data. They have a name, and the value stored in a variable can be changed during the execution of a program. A property is a concept that describes an aspect or quality of a variable.

To create a variable, you use a keyword followed by the name of the variable and an assignment operator (=) followed by the value to be stored in the variable. A variable is a named place in memory where a programmer can store data and retrieve the data later using the variable “a, b, c, y, z”. Programmers give names to variables. Variables can represent numerical values, characters, strings of characters, or memory addresses.



Numerical variables:
Numerical variables have values that describe a measurable quantity as a number, such as "how many" or "how much." Therefore, numerical variables are quantitative variables.

Categorical variables:
Categorical variables have values that describe the 'quality' or 'characteristics' of a data unit, such as 'what kind' or 'which category'. Categorical variables are divided into categories that are mutually exclusive (in one category or another) and exhaustive (include all possible options). Therefore, categorical variables are qualitative variables and tend to be represented by a non-numerical value.
Categorical variables can also be defined as either ordinal or nominal:
• An ordinal variable is a categorical variable. Observations can take on a value that can be logically ordered or ranked. Categories associated with ordinal variables can be ranked higher or lower than one another, but do not necessarily create a numerical difference between each category. Examples of ordinal categorical variables include academic grades (i.e. A, B, C), clothing size (i.e. small, medium, large, extra large), and attitudes (i.e. strongly agree, agree, disagree, strongly disagree).

• A nominal variable is a categorical variable. Observations can take on a value that is not arranged in a logical order. Examples of nominal categorical variables are gender, job type, eye color, religion, and brand. The data collected for a categorical variable is qualitative data.

Frequency Distribution: The distribution shown by the data is called the frequency distribution. It is the number of times the same periodic signal repeats itself throughout the whole.

Continuous Variable: Variables that can take all values (infinitely) in the range they are defined. If you measure continuous data and express it with a function, you will obtain a

continuous function with continuous variables. For example, in the function y(t)=Asin(wt+ϕ), w=2πf; f: frequency, φ: phase angle; π, f, φ: constant coefficients, t: independent time variable, y(t): dependent variable. A continuous variable is a numerical variable. Observations can take any value between a certain set of real numbers. The value given to an observation for a continuous variable can contain values as small as the measuring instrument allows. Examples of continuous variables include height, time, age, and temperature.

For example, in the function expression f(t)=at^2+bt+c, if t is a continuous variable, the function f(t) expression is also continuous; if t is discrete, the function f(t) expression is also expressed as discrete. The continuous function is expressed in the form f(t). If the discrete function is f[n]; It is indicated by the index n=0,1,2….

Discrete variable:
A discrete variable is a numerical variable. If you make measurements at certain sampling intervals, you get a series with discrete data. Observations can take a value based on a count from a series of different exact values. An example of a discrete variable is the value of a company on the stock exchange given at certain time intervals during the day. The data collected for a numerical variable is quantitative data. A discrete variable is a discontinuous variable. They are variables that take discrete values at the intervals they are defined. They are series. [1,3,5,2,9,5,4, ...]

Independent Dependent Variables:
A variable that changes by increasing or decreasing without being dependent on another variable is called an independent variable. Some variables take on values depending on another variable and are called dependent variables.

An independent variable (also known as a predictor variable) is a variable that potentially affects, or predicts, another variable. For example, if you are investigating whether age affects income, age is the independent variable.

A dependent variable (also known as an outcome variable) is a variable that is potentially affected or predicted. For example, if you are investigating whether income can be predicted by age, income is the dependent variable.

It is important to be able to distinguish between these two types, as this will determine where you put each variable in tables and graphs. Remember that which variable is which depends on the context; some variables (e.g. age) are always independent, while others (e.g. smoking status) can be either independent or dependent, depending on what you are trying to test.

y=at^2+bt+c

Here, a,b,c: coefficients (ML, range); t: independent variable; y: dependent variable

Example: Z=aX+bY+c is an example of a programming expression.
X, Y and Z are variables. Independent variables can be given a specific value or range of values. Here, a and b are coefficients, and c is a constant value. If a feature takes the same value in each observation, that is, does not change from observation to observation, this situation is called constant. For example, since the value of pi will be an unchanging value, it is defined as constant.

In the expression y=at+b, a and b are constant coefficients, and t and y are variables. y is a variable that depends on t. Here, t is the independent variable and y is the dependent variable.

Measured Variables:
Measurement is the process of comparing an unknown quantity with a unit quantity of the same type and determining how many times it is. In physics, a set of numerical values are needed to define variables quantitatively. These numerical values are called physical quantities. There are many physical quantities such as mass, length, time, speed, acceleration, force, temperature, energy, electric field strength, magnetic flux, etc. Physical quantities can be examined in two groups as "basic physical quantities" and "derived physical quantities".

Basic Quantities: These are quantities that are directly determined, that is, not determined with the help of other physical quantities. For example, mass, time, length, etc.

Derived Quantities: These are quantities that cannot be directly determined, that is, can be derived with the help of basic quantities. For example, area, volume, speed, acceleration, etc.

The necessary precision and accuracy are defined in measurement, and a measuring instrument is used. Measurement is assigning numerical values to objects, events, or abstract concepts according to a set of known rules.

The data collected during the study are known as variables:
• Any physiological, psychological or performance characteristic of an organism
• Any characteristic of the organism's environment.

Levels of Measurement:
• A nominal scale is a measure of identity or category. It is useful for measuring qualitative data. It does not provide information about order or magnitude.
• An ordinal scale is a measure of order or rank. It is used to organize data into series. It does not provide information about magnitude.
• An interval scale is a measure of order and quantity. The difference between values can be calculated. You can also measure the difference between two interval scale values, but there is no natural zero. For example, temperature scales are interval data where 25C is hotter than 20C, and a difference of 5C has some physical meaning. Note that 0C is arbitrary, so it does not make sense to say that 20C is twice as hot as 10C.
• A ratio scale is an interval scale with a true zero. The ratio of any two scale points is independent of the units of measurement.

Accuracy is the degree to which a measured or calculated quantity corresponds to its true (actual) value. Accuracy is closely related to precision, also called repeatability, the degree to which subsequent measurements or calculations show the same or similar results.

Interpreting the measurement result: equality, identity, compatibility, similarity – difference, validity, unity.

The size or magnitude of an object is a property that determines whether the object is larger or smaller than other objects of the same type.

The fabric is measured. The ball is counted. When a drop comes together with another drop, the volume increases, it is not counted. When an apple comes together with an apple, it becomes two apples, it is counted. The numbers are one after another. The magnitude is side by side. Quantity is divided.

When asked what is the basis of physics, the answer is observation. Physics is a branch of science based on observations. It is very important to be able to convey the observed events correctly. For this, correct measurements should be taken and the measured magnitude should be understood and expressed correctly. The unit system was found to eliminate the differences in measurements. Thus, the unit system entered physics as a universal language and became an indispensable part of physics.

In the mathematical world, similarity requires that two objects have the same shape, but not necessarily the same size. For example, two different circles are both circles and similar, but their dimensions make them different. They can be compared as similar shapes, but they cannot be equated with each other. Two objects that are similar will have the same shape, but one may be a scaled-down or scaled-down version of the other. The orientation of the shape may be different, but they remain similar. Mathematically, objects are similar when they have the same shape but are not the same size.

Identity: No matter what unknown number is given, both sides of the equation are the same. In equations, equality is achieved with some real number. The equations mentioned here are of course equations that are not identical. If the result is the same for each value, the equation is said to be identical.

Theories are explanations for some phenomena based on observation, experimentation, and reasoning. They are based on many experiments. Experiments that explain a theory must be repeatable. You must be able to predict from a theory. Laws are summaries of many experimental results and observations. Laws are not the same as theories because laws only tell what happens, not why it happens. Pure Science: An attempt to learn more about the world or an ongoing search for scientific knowledge. Pure science is done by scientists or people with curious minds. It involves experimentation, observation, questioning, and research. It involves technology.

Technology is the application of science to meet the needs of society. Engineers, inventors, and creative people apply scientific knowledge to build new “things” or tools. New technology can lead to new scientific discoveries. For example: Before the invention of the microscope, we could not know about cells.

Congruent shapes have the same measurements and overlap each other when they overlap. Two congruent objects are the same size and shape, but their orientation or placement in a space may be different. This does not change the fact that they are the same, because they have the same physical properties, the same angles, and the same measurements. When two objects are congruent, they can be matched or mapped exactly. They are the same size and have the same shape. Congruent objects are exact in terms of measurement, shape, and size. At first glance, the two shapes being compared may appear to be different because of how they are placed. However, when mapped or rotated, they are exact copies of each other and therefore congruent.

Errors:

Systematic error: Poor accuracy, Definite causes, Reproducible. Errors have consistent signs. Errors have consistent magnitude. Can be corrected by calibration, correcting procedural flaws, checking with a different procedure.

Random error: Poor precision, due to nonspecific causes. Cannot be repeated. Errors have random signs, Small errors are more likely than large errors. Can be corrected by making more measurements, improving technique, and increasing instrumental precision.

1.5. Sequences and Series

A **sequence** is an ordered list of numbers following a specific rule. Each number in a sequence is called a "term." The order in which terms are arranged is crucial, as each term has a specific position, often denoted as an, where n indicates the position in the sequence. For example:

- 2, 5, 8, 11, 14, . . . [Here, each term is 3 more then he previous term.]
- 3, 6, 12, 24, 48, . . . [Here, each term is 2 times of the preceding term]
- 0, 1, 1, 2, 3, 5, 8, 13, 21, . . . [Here, each term is sum of two preceding terms]

A **series** is the sum of the terms of a sequence. If we have a sequence a1, a2, a3, . . . the series associated with it is:

$S = a1 + a2 + a3 + \ldots$

How can you determine if a series converges or diverges?

A series converges if the sum of its terms approaches a finite value as more terms are added. For a geometric series, it converges if the absolute value of the common ratio $| r | < 1$. For other types of series, various tests like the comparison test, ratio test, or integral test are used to determine convergence or divergence.

1.6. Equations and Expressions

Expressions are written as phrases.
For example – the sum of a variable x and 6

Equations are written as sentences.
For example, the sum of 6 and the variable x equals 10.

What is the difference between a numerical expression and an algebraic or variable expression?
Numerical expression:
-3 + 2 + 4 – 5

Algebraic expression:
-3x + 2y – 4z - 5

An algebraic or numerical expression consists of three parts:
• Variable
• Coefficient
• Constant

The letter "x" in the expression 12+x is a variable. A variable is a symbol or letter that represents a number.

What are the variables in the following expression?
f(x,y,z)=- 4x + 3y - 8z + 9
x,y,z: Independent variable
f(x,y,z): Dependent variable

A coefficient is the number multiplied by a variable in an algebraic expression.
Algebraic expression Coefficient
6m + 5 6
8r + 7m + 4 8, 7
14b - 8 14

A coefficient is the number that multiplies a variable. In other words, it is the number in front of a variable.
What are the coefficients in the following expression?
f(x,y,z)=- 4x + 3y - 8z + 9
f(x,y,z)=ax + by + cz + d

A constant is any term that does not have a variable.
The constant in the following expression is -8.
f(x,y)=- 8 + 5y + 3x

A constant is a number that cannot change its value.
expression: 5x + 7y - 2
constant: - 2

Assigning values to variables
a = b = 4
This code allows us to define variables a and b:

Using this information, for example, you can assign the number of days each month has in a year to month names:

jan = march = may = july = august = october = december = 31
april = june = september = november = 30
february = 28

Thus, in one go, we have defined seven variables whose value is 31, four variables whose value is 30, and one variable whose value is 28. You know how to access the values of these variables.

Polynomial, F(x)= an xn + an-1 Xn-1+ … + a1x + a0

Where,
a0: constant
an, an-1, … , a1, a0 : coefficients
x: independent variable
F(x): dependent variable

1.7. Functions

A signal is a model, pattern, template of a change or transformation that carries information. Signals are mathematically represented as a function of one or more independent variables. An image is the brightness as a function of two spatial variables, x and y.

In this lesson, signals that contain a single independent variable are usually expressed as a time, t, and are considered. A signal is a real-valued or scalar-valued function of an independent variable t, although it does not represent time in a particular application.

The mathematical expression of a signal is called a function. A discrete-time signal is called a sequence. A sequence is a function obtained by taking samples of a function at equal intervals in the time domain. Mathematical expressions of variables that are dependent on each other are also signals. The number of deaths from heart attacks by age, the rate of deaths from heart attacks among married people.

Although a sequence is a special case of a function, the term function is also generally used to refer to a function that is not a sequence. The nth element of a sequence x is denoted by x(n), x[n] or xn. The concepts of vectors are critical in functions and sequences. They define the direction and rate of change.

An expression such as "f(t)" refers to the value of the function f evaluated at a point t. Instead of the value of f evaluated at a point t, engineers often use the expression "f(t)" to refer to the function f, and this careless notation can sometimes lead to problems such as ambiguity. Therefore, care should be taken to make a clear distinction between a function and its value.

Let f and g be real-valued functions of a real variable; let t be an arbitrary real number. Let H be a system operator (that maps a function to a function).
The quantity f + g formed by adding the functions f and g is also a function.
The quantity f(t) + g(t) is a number, that is, the sum of the following: the value of the function f evaluated at t and the value of the function g evaluated at t.
The quantity Hx is a function, that is, the output produced by the system represented by H when the input to the system is a function of x. The quantity Hx(t) is a number, that is, the value of the function Hx evaluated at t.

Number of independent variables (i.e., dimensionality):
• A signal with one independent variable is said to be one-dimensional (e.g., sound).
• A signal with more than one independent variable is said to be multidimensional (e.g., image).

Continuous or discrete independent variables:
• A signal with continuous independent variables is said to be continuous time (CT) (e.g., voltage waveform).
• A signal with discrete independent variables is said to be discrete time (DT) (e.g., stock market index).

Continuous or discrete dependent variable:

• A signal with a continuous dependent variable is said to be continuous valued (e.g., voltage waveform).

• A signal with a discrete dependent variable is said to be discrete valued (e.g., digital image).

• A CT signal with a continuous value is said to be analog (e.g., voltage waveform).

• A discrete-valued DT signal is said to be digital (for example, digital audio).

The polynomial Formula gives the standard form of polynomial expressions. It specifies the arrangement of algebraic expressions according to their increasing or decreasing power of variables.

The General Formula of a Polynomial:

$f(x) = anx^n + an{-}1x^{n-1} + \cdots + a1x + a0$

Where,

- an, an−1, …, a1, a0 are the coefficients,
- x is the variable,
- n is the degree of the polynomial (the highest power of x).

A **polynomial** is an algebraic expression consisting of terms with non-negative integer exponents of the variable. It can be expressed as the sum of **monomials**, **binomials**, or more complex expressions.

The **degree of a polynomial** is determined by the highest power of the variable in the expression. For example, in the polynomial f(x) = 3x2 + 4x + 5, the highest power of x is 2, so the degree of the polynomial is 2.

Like and Unlike Terms:

- **Like terms**: Terms that have the same variable raised to the same power (e.g., $3x^2$ and $5x^2$).
- **Unlike terms**: Terms that have different variables or different powers (e.g., $3x^2$ and 4x).

Type of Polynomial	Description	General Formula	Example
Monomial	Polynomials with one term	ax^n	x, y^2, $3y^3$, etc.

Type of Polynomial	Description	General Formula	Example
Binomial	Polynomials with two terms	$ax^n + by^m$	$2x + y^2$, $x + 3y^3$, etc.
Trinomial	Polynomials with three terms	$ax^n + by^m + cz^k$	$2x + z + y^2$, $z - x + 3y^3$, etc.
Quadratic Polynomial	Second-degree polynomial (typically two or three terms)	$ax^2 + bx + c$	

1.8. Units

Unit: A certain and unchangeable part selected as an example of its own kind to measure a quantity; for example, the unit of length measurement is the meter, the unit of Turkish currency is the lira. Each of the entities that constitute the multitude or each element of a set.

International System of Units (SI):
In the "Weights and Measurements" conference held in Paris in 1960 on the needs in science, technology, trade and engineering, the International System of Units or the International System of Measurements (French: Système International d'unités) was defined and given an official status. The acceptance of the SI System of Units facilitated international technical communication. At the conference, seven quantities were determined as "base quantities".

The quantities of the same kind selected for comparison purposes to measure a quantity are called units. The multitude of physical quantities to be measured and their differences at the same time led to the need to establish unit systems based on a small number of base units. Systems consisting of arbitrarily selected base quantities and quantities whose

definitions are derived from these base quantities are called unit systems. There are three important unit systems in general use.

Mainly used unit systems:
FPS Unit System: This system, also known as the British Unit System; It is a unit system in which length is measured in feet (ft), weight in pounds (lb) and time in seconds (s).

CGS Unit System: It is a unit system in which length is measured in centimeters (cm), mass in grams (g) and time in seconds (s).

MKS Unit System: It is a unit system in which length is measured in meters (m), weight in kilogram-force (kg-f) and time in seconds (s). The name of this unit system has been changed to SI and has become one of the most widely used unit systems in calculations.

Definition of kilogram:
Since 1889, the definition of the kilogram has been made with a platinum-based ingot called "Le Grand K", which is kept in a safe in Paris. However, the days of the main definition of the kilogram are numbered. Its weight has changed due to deterioration over the years.

We know that there are differences between the kilograms copied from the kilogram in Paris and the Big K itself. This situation is unacceptable from a scientific point of view. Therefore, the Big K may serve the purpose now, but it will not in 100 years.

The definition of the kilogram is changing... The gravitational force produced by electromagnets is directly related to the electric current going to their coils. In other words, there is a direct relationship between electricity and weight. There is a number called the Planck constant, discovered by German physicist Max Planck, which establishes the relationship between weight and electric current, and is represented by the symbol h. The device known as the Kibble balance has an electromagnet on one side and a weight, such as a kilogram, on the other. The amount of electric current passing through the electromagnet is increased until both sides of the device are perfectly balanced. Experts can measure the amount of electricity passing through the electromagnet with a precision of 0.000001 percent.

Units		Inches	Feet	Yards	Miles	Centimeters	Meters
1 inch	=	1	0.083 333 33	0.027 777 78	0.000 015 782 83	2.54	0.025 4
1 foot	=	12	1	0.333 333 3	0.000 189 393 9	30.48	0.304 8
1 yard	=	36	3	1	0.000 568 181 8	91.44	0.914 4
1 mile	=	63 360	5 280	1 760	1	160 934.4	1609.344
1 centimeter	=	0.393 700 8	0.032 808 40	0.010 936 13	0.000 006 213 712	1	0.01
1 meter	=	39.370 08	3.280 840	1.093 613	0.000 621 371 2	100	1

Common Powers:

Prefix	Symbol	Power of 10	Power of 2	Prefix	Symbol	Power of 10
Kilo	K	1 thousand = 10^3	$2^{10} = 1024$	Milli	m	1 thousandth = 10^{-3}
Mega	M	1 million = 10^6	2^{20}	Micro	μ	1 millionth = 10^{-6}
Giga	G	1 billion = 10^9	2^{30}	Nano	n	1 billionth = 10^{-9}
Tera	T	1 trillion = 10^{12}	2^{40}	Pico	p	1 trillionth = 10^{-12}
Peta	P	1 quadrillion = 10^{15}	2^{50}	Femto	f	1 quadrillionth = 10^{-15}
Exa	E	1 quintillion = 10^{18}	2^{60}	Atto	a	1 quintillionth = 10^{-18}
Zetta	Z	1 sextillion = 10^{21}	2^{70}	Zepto	z	1 sextillionth = 10^{-21}
Yotta	Y	1 septillion = 10^{24}	2^{80}	Yocto	y	1 septillionth = 10^{-24}

$1\ inch\ (in) = \frac{1}{12} foot = 2{,}54\ cm = 0{,}254\ m$

$1\ foot\ (ft) = 30{,}48\ cm = 0{,}3048\ m$

$1\ yard\ (yd) = 3\ foot\ (ft) = 91{,}44\ cm = 0{,}9144\ m$

$1\ mile\ (mi) = 1760\ yard\ (yd) = 1609\ m$

$1\ ound\ (oz) = \frac{1}{16} Pound\ Ib) = 28{,}4\ g = 0{,}00384\ kg$

$1\ pound\ (Ib) = 0{,}454\ kg = 454\ g$

$1\ stone = 14\ Ib = 6{,}35\ kg$

UNITS OF VOLUME

1 gallon ≅ 3.78 liters
≅ 231 cubic inches
≅ 0.1335 cubic ft
≅ 4 quarts
≅ 8 pints

1 fl ounce ≅ 29.57 cubic centimeter (cc) or milliliters (ml)

1 in^3 ≅ 16.387 cc

UNITS OF AREA

1 sq meter ≅ 10.76 sq ft

1 sq in ≅ 645 sq millimeters (mm)
= 1,000,000 sq mil

1 mil = 0.001 inch

1 acre = 43,560 sq ft

UNITS OF WEIGHT

1 kilogram (kg) ≅ 2.2 pounds (lbs)

1 pound ≅ 0.45 Kg
= 16 ounce (oz)

1 oz = 437.5 grains

1 carat ≅ 200 mg

1 stone (U.K.) ≅ 6.36 kg

NOTE: These are the U.S. customary (avoirdupois) equivalents, the troy or apothecary system of equivalents, which differ markedly, was used long ago by pharmacists.

UNITS OF POWER / ENERGY

1 H.P. ≅ 33,000 ft-lbs/min
= 550 ft-lbs/sec
≅ 746 Watts
≅ 2,545 BTU/hr

(BTU = British Thermal Unit)

1 BTU ≅ 1055 Joules
≅ 778 ft-lbs
≅ 0.293 Watt-hrs

Carob seeds are always the same gram. Historically, carat was used for diamond measurement. 1 Carat = 200mg 16 carob seeds make 1 dirham. 1 dirham = 3gr

UNITS OF LENGTH

1 inch (in)	=	2.54 centimeters (cm)
1 foot (ft)	=	30.48 cm = 0.3048 m
1 yard (yd)	≅	0.9144 meter
1 meter (m)	≅	39.37 inches
1 kilometer (km)	≅	0.54 nautical mile
	≅	0.62 statute mile
	≅	1093.6 yards
	≅	3280.8 feet
1 statute mile	≅	0.87 nautical mile
(sm or stat. mile)	≅	1.61 kilometers
	=	1760 yards
	=	5280 feet
1 nautical mile	≅	1.15 statute miles
(nm or naut. mile)	≅	1.852 kilometers
	≅	2025 yards
	≅	6076 feet
1 furlong	=	1/8 mi (220 yds)

UNITS OF SPEED

1 foot/sec (fps)	≅	0.59 knot (kt)*
	≅	0.68 stat. mph
	≅	1.1 kilometers/hr
1000 fps	=	600 knots
1 kilometer/hr	≅	0.54 knot
(km/hr)	≅	0.62 stat. mph
	≅	0.91 ft/sec
1 mile/hr (stat.)	≅	0.87 knot
(mph)	≅	1.61 kilometers/hr
	≅	1.47 ft/sec
1 knot*	≅	1.15 stat. mph
	≅	1.69 feet/sec
	≅	1.85 kilometer/hr
	≅	0.515 m/sec

*A knot is 1 nautical mile per hour.

2. Signals and Systems

A signal is a functional definition of a change or transformation that carries a message. Information is given meaning by a signal. If you measure the signals, you manage the processes.

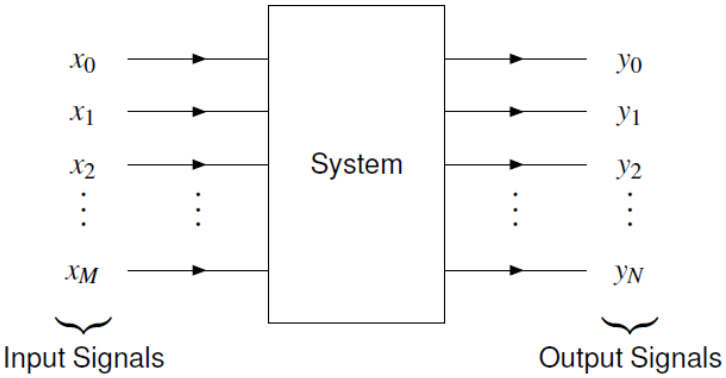
Mathematically, signals are represented as functions of one or more independent variables. A function f is called a sequence (vector) of variables (ti) in which each {ti, i=1,2, ... ,n} is sampled at independent discrete times.

The function f(t) is continuous, f(t), is called a dependent variable of the independent variable t. We usually consider signals that contain a single independent variable called time, t. A signal is a real-valued or discrete-valued function of an independent variable t. Although it does not represent time in a particular application, a signal is a real-valued or discrete-valued function of an independent variable t.

A signal is stored and gains meaning. Information is written on a stone, a book. It is written in a memory or in the brain. What makes information powerful is that it can be carried and stored in integration with signals. The information processed on the clay tablet has been stored for ages and has stopped time.

All components that make up the universe, from subatomic particles to giant planets, produce signals. All components that make up the universe interact with each other through signals. There is a continuous flow of information. We become conscious by discovering the universe and the information hidden in the signals.

In the spatial case depending on the path, there is a weakening of the signal. (frequency and the square of the distance are weakened). Noise produces the errors.



A signal,

- A signal is a pattern of change that carries information.
- A signal is mathematically represented as a function of one or more independent variables.
- There are two types of electrical signal forms: Analog signal, digital signal.
- An analog signal is a real-valued or scalar-valued function of the independent variable t. Frequency, amplitude, phase and time are independent variables.
- A picture is the change of colors as a function of two spatial variables, x and y.
- Digital signals: binary (binary number system), bit:0/1. Signals in the internal structure of computer systems are represented by electrical signals in the binary number system.
- Systems are units that process input signals and convert the input signal into another signal in order to produce output signals in line with the purpose.
- Systems can be physical or hardware, as well as completely software. Software ones are called mathematical models and mathematical programs are also called virtual systems.

Signal types:

Analog signals are continuous in terms of time, frequency, phase and amplitude. It is a signal whose amplitude, phase and frequency change over time. It consists of a mixture of sinusoidal signals. Example: voltage, current, temperature,… The signals that exist in the non-computer world are analog. X(t)=Asin(wt+ø). The components of an analog signal: A: amplitude, f: frequency, ø: phase, t: time. The signal can be temporal (function of time), spatial (function of space), spatiotemporal.

Digital (Numerical) signals are the representation of signals that are discrete in terms of both time and amplitude in the binary number system. (bit:0/1)

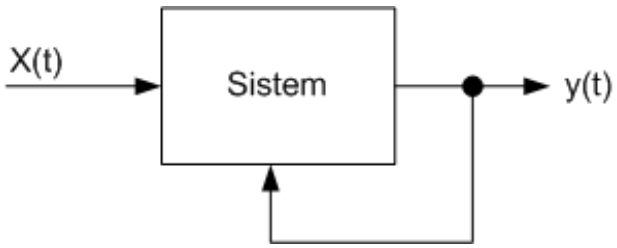
Discrete-time signals are discrete in time and continuous in amplitude. They are amplitude values taken continuously from an analog signal at certain time intervals. These sample

amplitude values taken are represented by a bit group consisting of 1/0s. In theory, it is discrete-time continuous. It is based on sample values taken from amplitude signals. In computer science, its equivalent is vector; array or matrix.

Analog signal is a continuous signal whose amplitude, frequency and phase change over time. All systems and all components that make up the universe produce analog signals.
In general, an analog signal consists of the sum of sinusoidal signals.

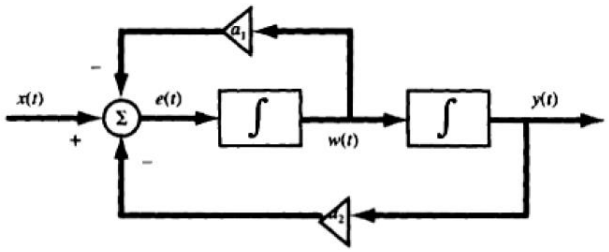
An analog signal generally consists of the sum of sinusoidal signals or their combination with certain arithmetic operations. A sinusoidal signal, X(t)=Asin(wt+ø)
W=2*pi*f, radial frequency
Ø: phase
A: Amplitude
F: frequency, the number of vibrations or periods in one second. Signals are transmitted at different frequencies. The hearing range of the ear: 300Hz to 20 KHz. GSM frequency: 900MHz, 1800Mhz, 2100MHz, 3Ghz and 30GHz, 6G: Terahetrz; Satellite: 9Ghz to 14GHz
Period: The repetition time of a similar analog signal that is repeated continuously.

2.1. Intelligent systems as feedback.



If the outputs of the systems become inputs as feedback, the systems start to become smarter.

Feedback Systems:



$$e(t) = \frac{dw(t)}{dt} = -a_1 w(t) - a_2 y(t) + x(t)$$

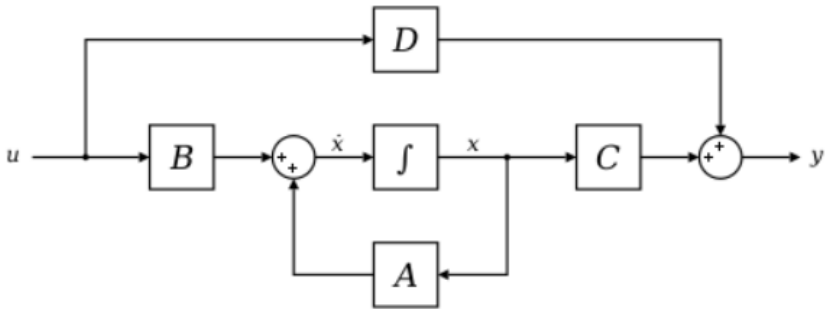
$$w(t)) = \frac{dy(t)}{dt}$$

$$\frac{d^2 y(t)}{dt^2} = -a_1 \frac{dy(t)}{dt} - a_2 y(t) + x(t)$$

$$\frac{d^2 y(t)}{dt^2} + a_1 \frac{dy(t)}{dt} + a_2 y(t) = x(t)$$

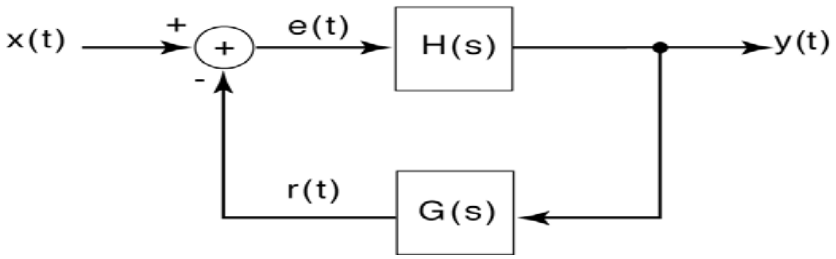
The relationship between integral and derivative plays a critical role in the mathematical modeling of systems. In a system that integrates the input signal, if the output is known, the input is obtained by taking the derivative of the signal. The converse is also true; if the output of a system that takes the derivative of the input signal is known, the input is obtained by integrating the signal. In mathematical modeling, the system is usually described by differential equations.

Mathematical model of linear state space equations:



$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t)$$
$$\mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t)$$

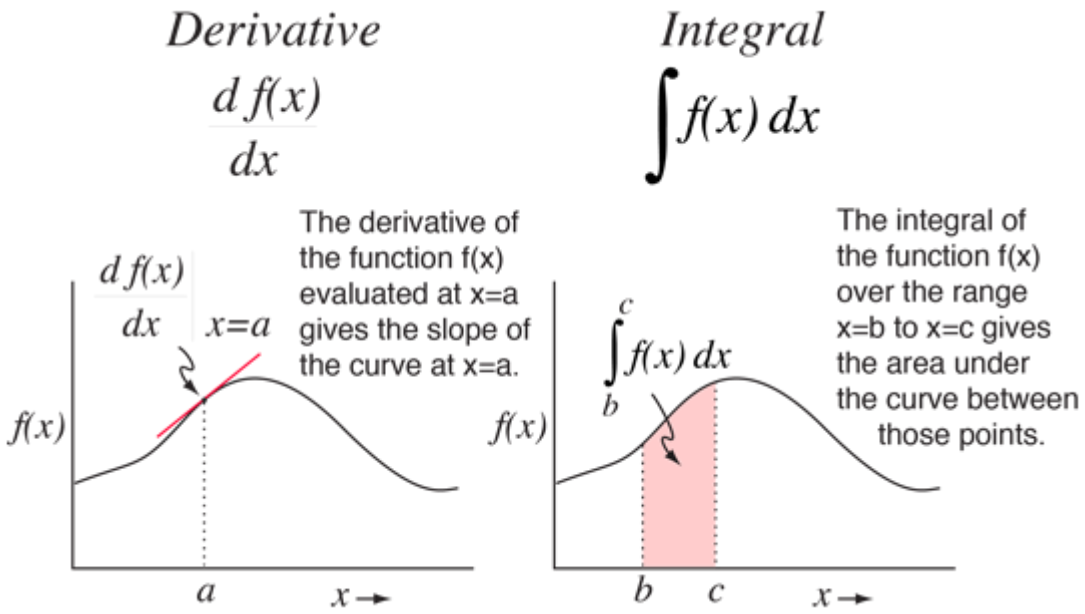
The transfer function of this system:



$$E(s) = X(s) - R(s) = X(s) - G(s)Y(s)$$
$$Y(s) = H(s)E(s) = H(s)[X(s) - G(s)Y(s)]$$
$$Q(s) = \frac{Y(s)}{X(s)} = \frac{H(s)}{1 + G(s)H(s)}$$

2.2. Derivatives and Integrals

Foundational working tools in calculus, the derivative and integral permeate all aspects of system modeling.



The derivative of a function can be geometrically interpreted as the slope of the curve of the mathematical function f(x) plotted as a function of x. But the modeling of system go far deeper than simple geometric application might imply. Derivative importance lies in the fact that many physical entities such as velocity, acceleration, force and so on are defined as instantaneous rates of change of some other quantity. The derivative can give a precise intantaneous value for that rate of change and lead to precise modeling of the desired quantity.

The integral of a function can be geometrically interpreted as the area under the curve of the mathematical function f(x) plotted as a function of x. You can see yourself drawing a large number of blocks to approximate the area under a complex curve, getting a better answer if you use more blocks. The integral gives you a mathematical way of drawing an infinite number of blocks and getting a precise analytical expression for the area. That's very important for geometry - and profoundly important for the system modelling where the definitions of many physical entities can be cast in a mathematical form like the area under a curve. The area of a little block under the curve can be thought of as the width of the strip weighted by (i.e., multiplied by) the height of the strip. Many properties of continuous bodies depend upon weighted sums, which to be exact must be infinite weighted sums - a problem tailor-made for the integral. A vast number of physical problems involve infinite sums in their solutions, making the integral an essential tool for the physical scientist.

3. Statistical Data Analysis

Statistics: It is a branch of science that enables estimating and commenting on risks and opportunities as a result of systematic collection and processing of data. Statistics is the set of methods used in the collection, processing, analysis and interpretation of data. Statistics is a science of uncertainty. Liars use numbers.

Why statistical data analysis? Observing, gathering information.
Responding to questions with numbers.
Commenting. Making predictions, estimating.

Objects or events that contain countable or measurable characteristics (variables) and have many similarities but also differences are called “statistical units”. If uncountable or unmeasurable objects or events are involved, they do not constitute a statistical unit.

The population is the collection of all statistical units defined for the research. For example; in a study conducted on the broken washing machines produced in a factory in a year, each of the broken machines is a statistical unit, while the collection of all these machines is called the population.

Statistical quality control; to produce accurate data at the lowest cost, on time and with the least cost. Thus, statistical analysis provides the opportunity to make decisions based on data in order to initiate corrective and preventive activities. Dr. E. Deming's comment on statistical process control; The basis of poor quality is variability. Quality cannot be achieved all of a sudden, and it is only possible to control system processes with statistical process analysis.

There are many risks in making predictions by analyzing large data sets. Disruptive factors such as noise, anomalies, missing data, incorrect data, intentional data, uncertainties are active in large data sets. Large data sets also produce incorrect results due to overfitting, bias and variance changes. Because the main behavioral characteristics of the data set are lost. The data must be cleaned, prepared, classified, clustered and converted into mathematical equations and expressions with regression. What needs to be done is to take sample data sets by making small-sized samples representing the large data set. Small data sets represent the large data set. A learning model is developed from one of the data sets obtained as a representative. This model is tested with the selected data set for testing purposes and the model to be used is determined by applying hypotheses. Over time, models that make decisions are obtained by improving performance from experiences.

First of all, generalizations should be made about the parameters of the large data set using statistical analysis methods for small data sets obtained from the sample. Whether the selected data sets represent the large data set is determined by statistical analysis methods. This process is called statistical interpretation. Statistical interpretation consists of generalizing two types of problems.

1. Estimation
2. Hypothesis testing

While performing hypothesis testing, it is determined whether the statistics of the sample data set, in other words, the values of which are tried to be known, are suitable for the parameters of the large data set.

There will definitely be a certain level of uncertainty when estimating parameters with the help of a statistic. There will be a difference between the statistics of the sample data set from which limited benefit is aimed and the parameters of the large data set. In this case, there is a risk of making an error while estimating.

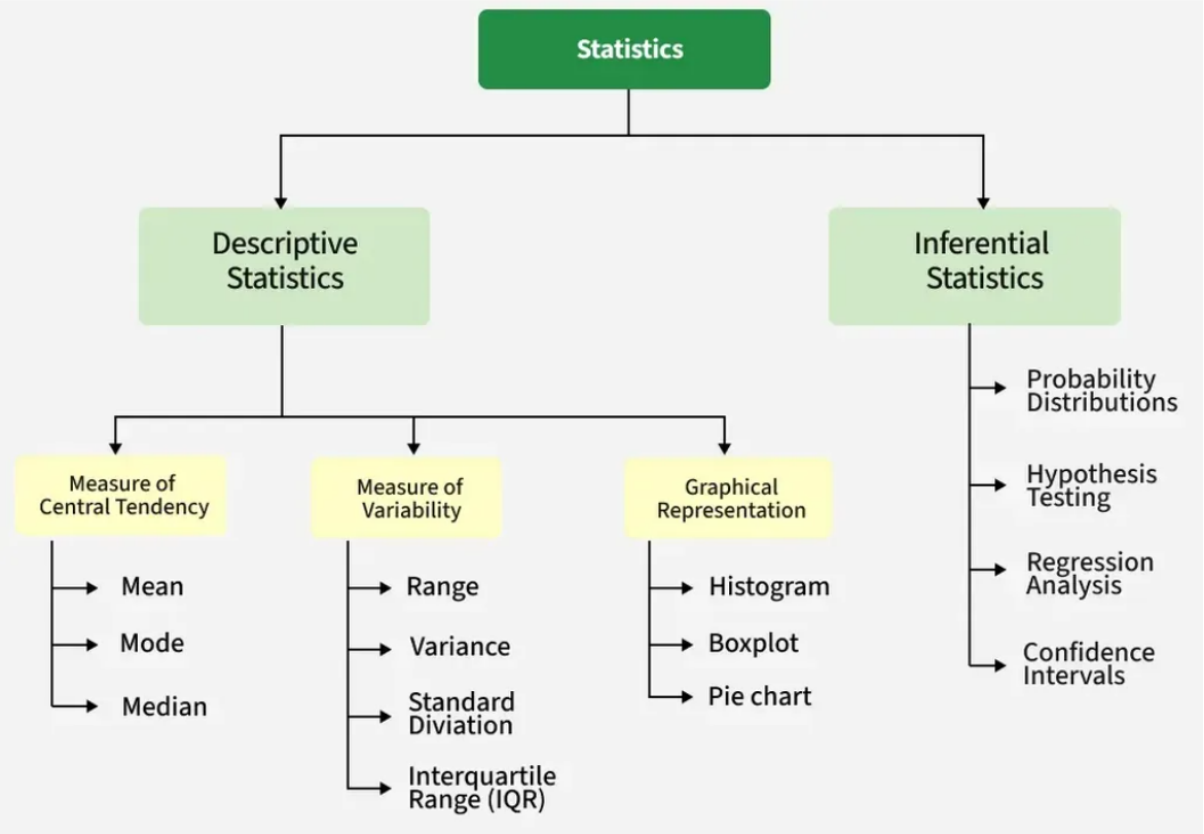
When a hypothesis is established, in order to be able to use an estimation, it is necessary to know how reliable this estimation is. On the other hand, it is also necessary to know what types of errors are encountered.

With statistical process analysis, problems are determined in advance, variability is reduced, it allows monitoring of machine or process adequacy, and it determines the need for corrective and preventive activities. When the decision-making process begins, analysis should be continued to see that the correct decision-making is statistically under control as a representation of the data stack.

Statistics is a branch of science that allows prediction and interpretation of risks with certain sensitivity as a result of observing business processes and systematically collecting and processing information.

Types of Statistics:
Descriptive statistics are procedures used to classify and summarize numerical data. Data is summarized in a meaningful way, such as tables, graphs or numerically. Some data can be organized as a frequency distribution. The average value and some special middle values can be calculated from the data. For example, the median is the value at the middle point that divides a group of numerical data into two (50% - 50%).

Predictive statistics draw conclusions for future situations and make predictions about deviations in the behavior of the population with data obtained by observation (measurement).



3.1. Measurement of Central Tendency

Measures of Central Tendency

- Arithmetic mean
- Width of distribution
- Mode
- Median - Median
- Mean Absolute Deviation (MAD)
- Variance - Standard Deviation
- Coefficient of Variation (CV)
- Coefficient of Skewness
- Coefficient of Kurtosis
- Comparison of measures of central tendency
- Spread

Measures of Dispersion (Variation):

- Variance
- Standard Deviation
- Coefficient of Variation

Visualization methods used in descriptive statistics:

- Frequency Tables
- Figures and Graphs
- Histograms and Frequency Polygons
- Column and Pie Charts

3.2. Arithmetic Mean

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

If some of the values calculated by the arithmetic mean are too high or too low, the arithmetic mean will not give the correct result in making an interpretation. If one or two of the sample values taken are too high, the arithmetic mean will not reflect the tendency of the behavior.

Average absolute deviation, variance, standard deviation and coefficients of variation. In such cases, the tendency of the behavior can be determined by evaluating the median.

Which two central tendency measures are critical for interpreting, predicting and developing hypotheses from the data mass? Weighted mean, standard deviation or variance. These two expressions are critical for interpreting, predicting and drawing conclusions.

Median evaluation is the value that separates the data series into 2 equal frequencies from the middle after the data values of the samples taken are sorted from large to small or from small to large.

Another method of determining behavioral tendency that does not take into account all values in a data set (is not sensitive) is Mode measurement. Mode measurement is the most frequently observed data value in a data set.

Variability measures such as the range of variability of data in the stack, mean deviation and standard deviation are used. The range of change in behaviors is calculated from the

difference between the highest and lowest values in a data series. Mean deviation is the arithmetic average of the absolute deviations from the arithmetic mean of all data values. The criterion that shows the levels at which sample values in a stack vary and change is called variance.

In order to find variability, deviations of behaviors from nominal values must be analyzed well. Since the midpoint in classified data will not always be the weighted midpoint of the behavior, higher deviation values are measured in values grouped according to raw data. Probability calculations and statistical methods are used together to measure the probability of a certain change and to determine deviations very well.

The rate of change in the values of the function against the change of a variable in the function that defines the behavior of the system is shown in tables and graphics. In order to analyze how to approach a given point on the graph, the slope of the tangent at that point should be found. In addition, the integral and derivative graphs of a function whose formula is given should be drawn to support interpretation. Factor analysis, which aims to explain measurement with a small number of factors by gathering variables together, reduces the number of variables and makes the results obtained meaningful. Factor analysis determines how measurement is performed. Regression analysis techniques are used to measure the relationships between variables.

An arithmetic mean is a fancy term for what most people call an "average." When someone says the average of 10 and 20 is 15, they are referring to the arithmetic mean. The simplest formula for a mean is the following: Add up all the numbers you want to average, and then divide by the number of items you just added.

For example, if you want to average 10, 20, and 27, first add them together to get 10+20+27= 57. Then divide by 3 because we have three values, and we get an arithmetic mean (average) of 19.

Want a formal, mathematical expression of the arithmetic mean?

$$\text{arithmetic mean} = \frac{\sum_{n=1}^{k} x_n}{k}$$

That's just a fancy way to say "the sum of k different numbers divided by k."
Check out a few examples of the arithmetic mean to make sure you understand:
Example:
Find the arithmetic mean (average) of the following numbers: 9, 3, 7, 3, 8, 10, and 2.
Solution:

Add up all the numbers. Then divide by 7 because there are 7 different numbers.
9+3+7+3+8+10+2=427=42/7=6

Example:
Find the arithmetic mean of -4, 3, 18, 0, 0, and -10.
Solution:
Add the numbers. Divide by 6 because there are 6 numbers.
The answer is 7/6, or approximately 1.167

Mean ($\bar{x}$ or μ): The mean, or arithmetic average, is calculated by summing all the values in a dataset and dividing by the total number of values. It's sensitive to outliers and is commonly used when the data is symmetrically distributed.

Median (M): The median is the middle value when the dataset is arranged in ascending or descending order. If there's an even number of values, it's the average of the two middle values. The median is robust to outliers and is often used when the data is skewed.

Mode (Z): The mode is the value that occurs most frequently in the dataset. Unlike the mean and median, the mode can be applied to both numerical and categorical data. It's useful for identifying the most common value in a dataset.

Mean of Grouped Data
Mean for the grouped data can be calculated by using various methods. The most common methods used are discussed in the table below:

Direct Method	**Assumed Mean Method**	**Step Deviation Method**
$\bar{x} = \sum f_i x_i / \sum f_i$ Where, **∑fi** is the sum of all frequencies	$\bar{x} = a + \sum f_i x_i / \sum f_i$ Where, **a** is Assumed mean **di** is equal to xi – a **∑fi** the sum of all frequencies	$\bar{x} = a + h\sum f_i x_i / \sum f_i$ Where, a is Assumed mean **ui** = (xi – a)/h **h** is Class size **∑fi** the sum of all frequencies

Differences between Mean, Median and Mode
Mean, median, and mode are measures of central tendency in statistics.

Feature	Mean	Median	Mode
Definition	Mean is the average of all values.	Median is the middle value when data is sorted.	Mode is the most frequently occurring value in the dataset.
Sensitivity	Mean is sensitive to outliers.	Median is not sensitive to outliers .	Mode is not sensitive to outliers.
Calculation	Calculated by adding up all values of a dataset and dividing them by the total number of values in dataset.	Calculated by finding the middle value in a list of data.	Calculated by finding which value occurs more number of times in a dataset.
Representation	Value of mean may or may not be in dataset.	Value of median is always a value from the dataset.	Value of mode is also always a value from the dataset.

How does Mean Median Mode link to Real Life?

In our daily life we came across various instances where we have to use the concept of mean, median and mode. There are various application of mean, median and mode, here's how they link to real life:

- **Mean**: Mean, or average, is used in everyday situations to understand typical values. For example, if you want to know the average income of people in a city, you would calculate the mean income.
- **Median:** Median is in household income data, the median income provides a better representation of the typical income than the mean when there are extreme values. In real estate, the median house price is often used to gauge the affordability of homes in a particular area.
- **Mode:** Mode represents the most frequently occurring value in a dataset and is used in scenarios where identifying the most common value is important. For example, in manufacturing, the mode may be used to identify the most common defect in a production line to prioritize quality control efforts

Question: Find the mean, median, mode, and range for the given data
190, 153, 168, 179, 194, 153, 165, 187, 190, 170, 165, 189, 185, 153, 147, 161, 127, 180

Solution:
For Mean:
190, 153, 168, 179, 194, 153, 165, 187, 190, 170, 165, 189, 185, 153, 147, 161, 127, 180
Number of observations = 18
Mean = (Sum of observations) / (Number of observations)
= (190+153+168+179+194+153+165+187+190+170+165+189+185+153+147
+161+127+180) / 18
= 2871/18
= 159.5
Therefore, the mean is 159.5

For Median:
The ascending order of given observations is,
127, 147, 153, 153, 153, 161, 165, 165, 168, 170, 179, 180, 185, 187, 189, 190, 190, 194
Here, n = 18
Median = 1/2 [(n/2) + (n/2 + 1)]th observation
= 1/2 [9 + 10]th observation
= 1/2 (168 + 170)
= 338/2
= 169
Thus, the median is 169

For Mode:
The number with the highest frequency = 153
Thus, mode = 153

For Range:
Range = Highest value – Lowest value
= 194 – 127
= 67

Question: Find the Median of the data 25, 12, 5, 24, 15, 22, 23, 25
Solution:

25, 12, 5, 24, 15, 22, 23, 25

Step 1: Order the given data in ascending order as:
5, 12, 15, 22, 23, 24, 25, 25

Step 2: Check n (number of terms of data set) is even or odd and find the median of the data with respective 'n' value.

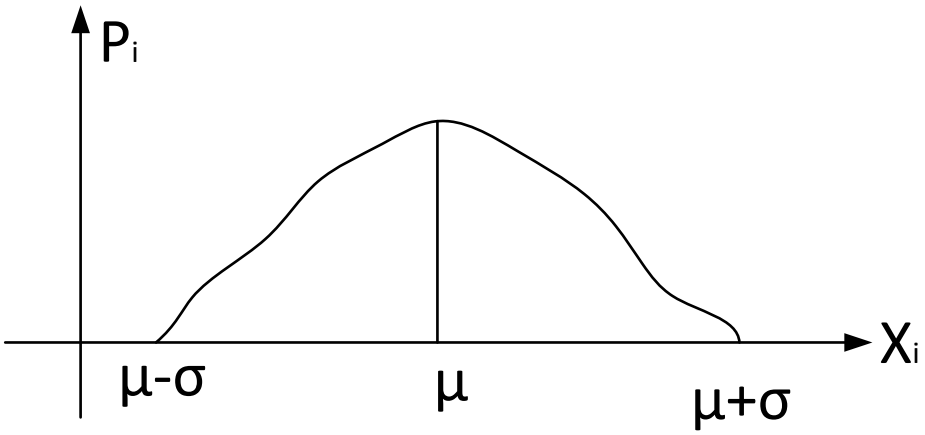
Step 3: Here, n = 8 (even) then,
Median = [(n/2)th term + {(n/2) + 1)th term] / 2
Median = [(8/2)th term + {(8/2) + 1}th term] / 2
= (22+23) / 2
= 22.5

Question: Find the mode of given data 15, 42, 65, 65, 95.
Solution:
Given data set 15, 42, 65, 65, 95
The number with highest frequency = 65
Mode = 65

Mean, Median and Mode are essential statistical measures of central tendency that provide different perspectives on data sets. The mean provides a general average, making it useful for evenly distributed data. The median gives a middle value, providing a better view of central tendency when dealing with skewed distributions or extreme values and, the mode highlights the most frequent value, making it valuable in categorical data analysis.

Mean, in statistical terms, represents the arithmetic average of a dataset. It is calculated by summing up all the values in the dataset and dividing the sum by the total number of values. For instance, if you have the numbers 2, 4, 6, 8, and 10, the mean would be (2 + 4 + 6 + 8 + 10) / 5 = 6.

3.3. Measurements of Dispersion: Variance - Standard Deviation



Standard deviation is the root mean square (RMS) deviation resulting from the arithmetic mean of the values. How much does each value in the data set deviate from the weighted mean? In fact, the square of the deviations from the weighted mean are examined. Why? Variance can be calculated to comment on the distribution of the data. In probability and statistics, the standard deviation of a probability distribution is a measure of the spread of a random variable or a set or values. It is usually indicated with the letter σ (lowercase sigma). Standard deviation is defined as the square root of the variance. Variance is the sum of the squares of the differences between the data and the arithmetic mean. It measures the spread of the measured data to the mean. Standard deviation gives the deviation from the arithmetic mean.

A large data set cannot be analyzed soundly. Why? Noisy, error, missing data, manipulation, ..

A minimum number of sample data sets representing the large data set is taken. Does the sample data set represent the large data set? For this, variance, skewness and kurtosis coefficients are examined.

If the data values are close to the arithmetic mean, the standard deviation is small. Also, if many data points are far from the mean, the standard deviation is large. If all data values are equal, the standard deviation is zero.

An important measure of change in a data distribution is the variance. The standard deviation is obtained by taking the square root of the variance.

The standard deviation shows how close each value in the series is to the arithmetic mean. A small standard deviation shows that the deviations and risk in the mean are small, while a large standard deviation shows that the deviations and risk in the mean are large.
If data is collected from the entire population in a study, it is called a complete count. A sample is a data group consisting of certain elements selected from a population.

Variance
According to layman's words, the variance is a measure of how far a set of data are dispersed out from their mean or average value. It is denoted as 'σ^2'.

Properties of Variance

- It is always non-negative since each term in the variance sum is squared and therefore the result is either positive or zero.
- Variance always has squared units. For example, the variance of a set of weights estimated in kilograms will be given in kg squared. Since the population variance is squared, we cannot compare it directly with the mean or the data themselves.

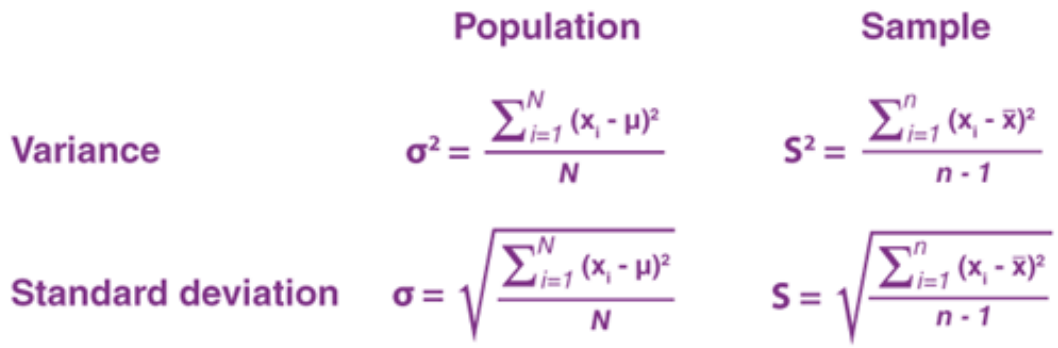
Standard Deviation
The spread of statistical data is measured by the standard deviation. Distribution measures the deviation of data from its mean or average position. The degree of dispersion is computed by the method of estimating the deviation of data points. It is denoted by the symbol, 'σ'.

Properties of Standard Deviation

- It describes the square root of the mean of the squares of all values in a data set and is also called the root-mean-square deviation.
- The smallest value of the standard deviation is 0 since it cannot be negative.
- When the data values of a group are similar, then the standard deviation will be very low or close to zero. But when the data values vary with each other, then the standard variation is high or far from zero.

Variance and Standard Deviation Formula
As discussed, the variance of the data set is the average square distance between the mean value and each data value. And standard deviation defines the spread of data values around the mean.

Variance and Standard Deviation are the important measures used in Mathematics and Statics to find the meaning from a large set of data. The different formulas for Variance and Standard Deviation are highly used in mathematics to determine the trends of various values in mathematics. Variance is the measure of how the data points vary according to the mean while standard deviation is the measure of the central tendency of the distribution of the data. The major difference between variance and standard deviation is in their units of measurement. Standard deviation is measured in a unit similar to the units of the mean of data, whereas the variance is measured in squared units.



The formulas for the variance and the standard deviation for both population and sample data set are given below:

σ^2 = Population variance
N = Number of observations in population
X_i = ith observation in the population
μ = Population mean

s^2 = Sample variance
n = Number of observations in sample
x_i = ith observation in the sample

Standard Deviation Formula
σ = Population standard deviation
s = Sample standard deviation

Variance and Standard deviation Relationship
Variance is equal to the average squared deviations from the mean, while standard deviation is the number's square root. Also, the standard deviation is a square root of

variance. Both measures exhibit variability in distribution, but their units vary: Standard deviation is expressed in the same units as the original values, whereas the variance is expressed in squared units.

Variance

Variance is a measure of how far the observed values in a dataset fall from the arithmetic mean, and is therefore a measure of spread - more specifically, it is a measure of variability. It is denoted by the Greek letter sigma squared, and its formula is given by:

$$\sigma^2 = \frac{\Sigma(x - \bar{x})^2}{n} = \frac{\Sigma x^2}{n} - \bar{x}^2$$

where:

- σ^2 is the variance we are wanting to find
- $\sum$ is the summation function
- x is an observation in the dataset
- $\bar{x}$ is the population mean
- n is the number of observations in the population.

Standard Deviation

Standard deviation is the square root of the variance, and therefore is also a measure of spread - more specifically, it is a measure of dispersion (or, the measure of variability!). Where variance is used to show how much the values in a dataset vary from each other, the standard deviation exists to show how far apart the values in a dataset are from the mean, and therefore can be used to identify outliers.

Standard deviation is denoted by the Greek letter sigma and, being the square root of variance, is written as:

$$\sigma = \sqrt{variance} = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n}} = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$$

where:

- σ^2 is the variance we are wanting to find
- $\sum$ is the summation function
- x is an observation in the dataset
- $\bar{x}$ is the population mean
- n is the number of observations in the population.

Standard Error

Standard error is another measure of spread. The most common standard error is the standard error of the mean, and used to measure sampling error as it measures how accurately the mean of a sample distribution represents the mean of the population. In other words, it shows how much variation there is likely to be between different samples of a population and the population itself.

The main difference between the standard deviation and the standard error is that the standard deviation is a type of descriptive statistics, used to summarise the data, whereas the standard error of the mean describes the random sampling process, and is an estimation rather than a definite value like the standard deviation is. It is useful because you can see how well your sample data represents your population.

The formula is given by:

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \frac{1}{\sqrt{n}}\sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$$

where:

- SE is the standard error
- σ is the standard deviation
- n is the sample size.

Example

Question: If a die is rolled, then find the variance and standard deviation of the possibilities.

Solution: When a die is rolled, the possible outcome will be 6. So the sample space, n = 6 and the data set = { 1;2;3;4;5;6}.

To find the variance, first, we need to calculate the mean of the data set.

Mean, $\overline{x}$ = (1+2+3+4+5+6)/6 = 3.5

We can put the value of data and mean in the formula to get;

$\sigma^2 = \Sigma (x_i - \overline{x})^2/n$

σ^2 = ⅙ (6.25+2.25+0.25+0.25+2.25+6.25)

σ^2 = 2.917

Now, the standard deviation,$\sigma = \sqrt{2.917} = 1.708$

Examples

Example 1

Let's say we have the following dataset:

7, 12, 5, 18, 5, 9, 10, 9, 12, 8, 12, 16

In order to find the variance and standard deviation of this, we need to first find the mean, which is:

$$\frac{7+12+5+18+5+9+10+9+12+8+12+16}{12}=\frac{123}{12}=10.25$$

The variance of this dataset is then given by:

$$\sigma^2=\frac{7^2+12^2+5^2+18^2+5^2+9^2+10^2+9^2+12^2+8^2+12^2+16^2}{12}-10.25^2=14.69$$

to two decimal places.

Then, the standard deviation is:

$$\sigma=\sqrt{\frac{7^2+12^2+5^2+18^2+5^2+9^2+10^2+9^2+12^2+8^2+12^2+16^2}{12}-10.25^2}=3.83$$

to two decimal places, and the standard error is given by:

$$SE=\frac{3.83}{\sqrt{12}}=1.11$$

to two decimal places.

3.4. Anova

Variance Analysis (or ANOVA, abbreviation of the English words ANalysis Of VAriance) is a collection of statistical models used in the field of statistics to analyze group means and processes related to them (such as intra-group and inter-group variation). In its simplest form, variance analysis is an inferential statistical method to test whether the means of several groups are equal or unequal, and this test generalizes the t-test test performed for two groups to multiple groups. If it is desired to perform multiple two-sample t-tests one after the other for multivariate analysis, it is obvious that this results in an increase in the probability of making a type I error. Therefore, it becomes clear that Variance Analysis will be more useful for comparing the statistical significance of three or more (for groups or variables) means with the test.

In short, ANOVA is a parametric inferential method and is used to test whether there is a difference between the population means. For example, the H0 hypothesis 'The average gasoline consumption of Opel and Toyota brand vehicles is the same' is tested. The result is obtained as "averages are the same" or "averages are not the same". It is assumed that there is no slope coefficient for a linear relationship between the two variables in this analysis (as in regression analysis). The most basic condition for conducting ANOVA analysis is that the variances of the populations whose averages will be examined are the same.

ANOVA Test or Analysis of Variance is a statistical method used to test the differences between the means of two or more groups. ANOVA helps determine whether there are statistically significant differences between the means of three or more independent groups. This method was first developed by the English statistician and geneticist Ronald Fisher in the 1920s and 1930s. Since they are generally characterized by their use of the F-distribution in statistical significance tests, this analysis is sometimes called Fisher's Analysis of Variance.

ANOVA analysis is a statistical relevance tool designed to evaluate whether the null hypothesis can be rejected when testing hypotheses. It is used to determine whether the means of three or more groups are equal. The ANOVA test is used to look for heterogeneity within groups and variability between groups. The f test returns the ANOVA test statistic. The ANOVA formula consists of many parts. The best way to handle an ANOVA test problem is to organize the formulas in an ANOVA table. Below are the ANOVA formulas.

Source of Variation	Sum of Squares	Degree of Freedom	Mean Squares	F Value
Between Groups	$SSB = \Sigma n_j(\bar{X}_j - \bar{X})^2$	df1 = k – 1	MSB = SSB / (k – 1)	f = MSB / MSE or, F = MST/MSE
Error	$SSE = \Sigma (\bar{X} - \bar{X}_j)^2$	df2 = N – k	MSE = SSE / (N – k)	
Total	SST = SSB + SSE	df3 = N – 1		

where,

- F = ANOVA Coefficient
- MSB = Mean of the total of squares between groupings
- MSW = Mean total of squares within groupings
- MSE = Mean sum of squares due to error
- SST = total Sum of squares
- p = Total number of populations
- n = The total number of samples in a population
- SSW = Sum of squares within the groups
- SSB = Sum of squares between the groups
- SSE = Sum of squares due to error
- s = Standard deviation of the samples
- N = Total number of observations
- k=Örnek sayısı

Examples of the use of ANOVA Formula

- Assume it is necessary to assess whether consuming a specific type of tea will result in a mean weight decrease. Allow three groups to use three different varieties of tea: green tea, Earl Grey tea, and Jasmine tea. Thus, the ANOVA test (one way) will be utilized to examine if there was any mean weight decrease displayed by a certain group.
- Assume a poll was undertaken to see if there is a relationship between salary and gender and stress levels during job interviews. A two-way ANOVA will be utilized to carry out such a test.

Important Notes on ANOVA Test

- *ANOVA test is used to check whether the means of three or more groups are different or not by using estimation parameters such as the variance.*
- *An ANOVA table is used to summarize the results of an ANOVA test.*
- *There are two types of ANOVA tests - one way ANOVA and two way ANOVA*
- *One way ANOVA has only one independent variable while a two way ANOVA has two independent variables.*

One-Way ANOVA

This test is used to see if there is a variation in the mean values of three or more groups. Such a test is used where the data set has only one independent variable. If the test statistic exceeds the critical value, the null hypothesis is rejected, and the averages of at least two different groups are significant statistically.

The one way ANOVA test is used to determine whether there is any difference between the means of three or more groups. A one way ANOVA will have only one independent variable. The hypothesis for a one way ANOVA test can be set up as follows:
Null Hypothesis, H0: $\mu 1 = \mu 2 = \mu 3 = ... = \mu k$
Alternative Hypothesis, H1: The means are not equal
Decision Rule: If test statistic > critical value then reject the null hypothesis and conclude that the means of at least two groups are statistically significant.
The steps to perform the one way ANOVA test are given below:

- *Step 1: Calculate the mean for each group.*
- *Step 2: Calculate the total mean. This is done by adding all the means and dividing it by the total number of means.*
- *Step 3: Calculate the SSB.*
- *Step 4: Calculate the between groups degrees of freedom.*
- *Step 5: Calculate the SSE.*
- *Step 6: Calculate the degrees of freedom of errors.*
- *Step 7: Determine the MSB and the MSE.*
- *Step 8: Find the f test statistic.*
- *Step 9: Using the f table for the specified level of significance, α, find the critical value. This is given by $F(\alpha, df_1. df_2)$.*
- *Step 10: If f > F then reject the null hypothesis.*

Limitations of One Way ANOVA Test:
The one way ANOVA is an omnibus test statistic. This implies that the test will determine whether the means of the various groups are statistically significant or not. However, it cannot distinguish the specific groups that have a statistically significant mean. Thus, to find the specific group with a different mean, a post hoc test needs to be conducted.

Two-Way ANOVA

Two independent variables are used in the two-way ANOVA. As a result, it can be viewed as an extension of a one-way ANOVA in which only one variable influences the dependent variable. A two-way ANOVA test is used to determine the main effect of each independent variable and whether there is an interaction effect. Each factor is examined independently to determine the main effect, as in a one-way ANOVA. Furthermore, all components are analyzed at the same time to test the interaction impact.

The two way ANOVA has two independent variables. Thus, it can be thought of as an extension of a one way ANOVA where only one variable affects the dependent variable. A two way ANOVA test is used to check the main effect of each independent variable and to see if there is an interaction effect between them. To examine the main effect, each factor is considered separately as done in a one way ANOVA. Furthermore, to check the interaction effect, all factors are considered at the same time. There are certain assumptions made for a two way ANOVA test. These are given as follows:

- *The samples drawn from the population must be independent.*
- *The population should be approximately normally distributed.*
- *The groups should have the same sample size.*
- *The population variances are equal*

Suppose in the two way ANOVA example, as mentioned above, the income groups are low, middle, high. The gender groups are female, male, and transgender. Then there will be 9 treatment groups and the three hypotheses can be set up as follows:

H01: All income groups have equal mean anxiety.

H11: All income groups do not have equal mean anxiety.

H02: All gender groups have equal mean anxiety.

H12: All gender groups do not have equal mean anxiety.

H03: Interaction effect does not exist

H13: Interaction effect exists.

Examples on ANOVA Test

Example: Three types of fertilizers are used on three groups of plants for 6 weeks. We want to check if there is a difference in the mean growth of each group. Using the data given below apply a one way ANOVA test at 0.05 significant level.

Fert1	***Fert2***	***Fert3***
6	***8***	***13***
8	***12***	***9***
4	***9***	***11***
5	***11***	***8***
3	***6***	***7***
4	***8***	***12***

Solution: Hypotheses are determined.

- *H0: $\mu 1 = \mu 2 = \mu 3$*
- *H1: The means are not equal*

The average of each group is calculated separately.
Ort1=5, Ort2=9, Ort3=10

The average of the averages is calculated.
Total means=Tort= (ort1+Ort2+ Ort3)/3= (5+9+10)/3=24/3=8

The number of elements in each set,
N1= N2 = N3= 6,

Total number of event clusters, k = 3
$df_1 = k - 1 = 2$

SSB = Sum of squares between the groups
$SSB = n1(Ort1-Tort)^2 + n2(Ort2 - Tort)^2 + n3(Ort3 - Tort)^2$
$SSB = 6(5-8)^2+6(9-8)^2+6(10-8)^2= 6*9+6*1+6*4=54+6+24=84$

Fert1	***(X - 5)²***	***Fert2***	***(X - 9)²***	***Fert3***	***(X - 10)²***
6	***1***	***8***	***1***	***13***	***9***
8	***9***	***12***	***9***	***9***	***1***
4	***1***	***9***	***0***	***11***	***1***
5	***0***	***11***	***4***	***8***	***4***
3	***4***	***6***	***9***	***7***	***9***
4	***1***	***8***	***1***	***12***	***4***
Ort1 = 5	***E1 = 16***	***Ort2 = 9***	***E2 = 24***	***Ort3 = 10***	***E3 = 28***

SSE = Sum of squares due to error,
SSE = E1+E2+E3=16 + 24 + 28 = 68

Total number of observations, N = observation-1+ observation-2+ observation-3= 6 + 6 +6=18

$df_2 = N - k = 18 - 3 = 15$

MSB = Mean of the total of squares between groupings
$MSB = SSB / df_1 = 84 / 2 = 42$

MSE = Mean sum of squares due to error
$MSE = SSE / df_2 = 68 / 15 = 4.53$

ANOVA test statistic, f = MSB / MSE = 42 / 4.53 = 9.33
Using the f table at α= 0.05 the critical value is given as F(0.05, 2, 15) = 3.68

- *As f > F, thus, the null hypothesis is rejected and it can be concluded that there is a difference in the mean growth of the plants.*
- *Answer: Reject the null hypothesis*

F-table of Critical Values of α = 0.10 for F(df1, df2)																			
	DF1=1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**∞**
DF2=1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.11
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

F-table of Critical Values of α = 0.05 for F(df1, df2)																			
	DF1=1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**∞**
DF2=1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

F-table of Critical Values of α = 0.01 for F(df1, df2)																			
	DF1=1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**∞**
DF2=1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6106.32	6157.29	6208.73	6234.63	6260.65	6286.78	6313.03	6339.39	6365.86
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.51	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.93	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.67	3.59	3.51	3.43	3.34	3.26	3.17
14	8.86	6.52	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.90	3.81	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.85	2.75
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.84	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.02	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.93	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.70	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.77	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.82	2.66	2.59	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.79	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.05	3.73	3.50	3.33	3.20	3.09	3.01	2.87	2.73	2.57	2.50	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.67	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.81
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.04	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.64	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.19	2.04	1.88	1.79	1.70	1.59	1.47	1.33	1.00

Example 2: *A trial was run to check the effects of different diets. Positive numbers indicate weight loss and negative numbers indicate weight gain. Check if there is an average difference in the weight of people following different diets using an ANOVA Table.*

Low Fat	***Low Calorie***	***Low Protein***	***Low Carbohydrate***
8	***2***	***3***	***2***
9	***4***	***5***	***2***
6	***3***	***4***	***-1***
7	***5***	***2***	***0***
3	***1***	***3***	***3***

Solution:

H0: μ1 = μ2= μ3 = μ4

H1: The means are not equal

Low Fat	$(X - 6.6)^2$	***Low Calorie***	$(X - 3)^2$	***Low Protein***	$(X - 3.4)^2$	***Low Carbohydrate***	$(X - 1.2)^2$
8	***2***	***2***	***1***	***3***	***0.2***	***2***	***0.6***
9	***5.8***	***4***	***1***	***5***	***2.6***	***2***	***0.6***
6	***0.4***	***3***	***0***	***4***	***0.4***	***-1***	***4.8***
7	***0.2***	***5***	***4***	***2***	***2***	***0***	***1.4***
3	***13***	***1***	***4***	***3***	***0.2***	***3***	***3.2***
Ort1 = 6.6	***E1 = 21.4***	***Ort2 = 3***	***E2 = 10***	***Ort3 = 3.4***	***E3= 5.4***	***Ort4 = 1.2***	***E4 = 10.6***

Total mean, Tort = 3.6

- *N1= N2 = N3 = N4 = 5, k = 4*

SSB = Sum of squares between the groups
SSB = $n1(Ort1-Tort)^2 + n2(Ort2 - Tort)^2 + n3(Ort3 - Tort)^2$

- *SSB= 75.8*
- *SSE = 21.4 + 10 + 5.4 + 10.6 = 47.4*

- *The ANOVA Table can be constructed as follows:*

Source of Variation	***Sum of Squares***	***Degrees of Freedom***	***Mean Squares***	***F Value***
Between Groups	***SSB= 75.8***	***df_1 = k - 1 = 4 - 1 = 3***	***MSB = SSB / (k - 1) = 25.3***	***f = MSB / MSE = 8.43***
Error	***SSE= 47.4***	***df_2 = N - k = 20 - 4 = 16***	***MSE = SSE / (N - k) = 3***	
Total	***SST = SSB + SSE = 123.2***	***df_3 = N - 1 = 19***		

- *As no significance level is specified, αα = 0.05 is chosen.*
- *F(0.05, 3, 16) = 3.24*
- *As 8.43 > 3.24, thus, the null hypothesis is rejected and it can be concluded that there is a mean weight loss in the diets.*
- *Answer: Reject the null hypothesis*

Example 3: *Determine if there is a difference in the mean daily calcium intake for people with normal bone density, osteopenia, and osteoporosis at a 0.05 alpha level. The data was recorded as follows:*

Normal Density	***Osteopenia***	***Osteoporosis***
1200	***1000***	***890***
1000	***1100***	***650***
980	***700***	***1100***
900	***800***	***900***
750	***500***	***400***
800	***700***	***350***

- ***Using the ANOVA test the hypothesis is set up as follows:***
- ***H0: μ1μ1 = μ2 = μ3***
- ***H1: The means are not equal***

Normal Density	$(X - 938.3)^2$	***Osteopenia***	$(X - 800)^2$	***Osteoporosis***	$(X - 715)^2$
1200	***68,486.9***	***1000***	***40,000***	***890***	***30,625***
1000	***3,806.9***	***1100***	***90,000***	***650***	***4,225***
980	***1,738.9***	***700***	***10,000***	***1100***	***148,225***
900	***1,466.9***	***800***	***0***	***900***	***34,225***
750	***35,456.9***	***500***	***90,000***	***400***	***99,225***

Normal Density	$(X - 938.3)^2$	Osteopenia	$(X - 800)^2$	Osteoporosis	$(X - 715)^2$
800	19,126.9	700	10,000	350	133,225
Ort1 = 938.3	Total = 130,083.3	Ort2 = 800	Total = 240,000	Ort3 = 715	Total = 449,750

- *Total mean, Tort = 817.8*
- *n1 = n2 = n3= 6, k = 3*
- *SSB = 152,477.7*
- *SSE = 130,083.3 + 240,000 + 449,750 = 819,833.3*

- *The ANOVA Table can be constructed as follows:*

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F Value
Between Groups	SSB = 152,477.7	$df_1 = k - 1 = 3 - 1 = 2$	MSB = SSB / (k - 1) = 76,238.6	f = MSB / MSE = 1.395
Error	SSE= 819,833.3	$df_2 = N - k = 18 - 3 = 15$	MSE = SSE / (N - k) = 54,655.5	
Total	SST = SSB + SSE = 972,311	$df_3 = N - 1 = 17$		

Using the F table the critical value is F(0.05, 2, 15) = 3.68

- *As 1.395 < 3.68, the null hypothesis cannot be rejected and it is concluded that there is not enough evidence to prove that the mean daily calcium intake of the three groups is different.*
- *Answer: Do not reject the null hypothesis*

Example: Calculate the ANOVA coefficient for the following data:

Plant	**Number**	**Average span**	**s**
Hibiscus	5	12	2
Marigold	5	16	1
Rose	5	20	4

Solution:

Plant	**n**	**x**	**s**	**s2**
Hibiscus	5	12	2	4
Marigold	5	16	1	1
Rose	5	20	4	16

p = 3

n = 5

N = 15

$\bar{x}$ = 16

$SST = \Sigma n(x-\bar{x})2$

$SST= 5(12 - 16)2 + 5(16 - 16)2 + 11(20 - 16)2 = 160$

MST = SST/p-1 = 160/3-1 = 80

SSE = ∑ (n−1) = 4 (4 + 1) + 4(16) = 84

MSE = 7

F = MST/MSE = 80/7

F = 11.429

Example: The following data show the number of worms quarantined from the GI areas of four groups of muskrats in a carbon tetrachloride anthelmintic study. Conduct a two-way ANOVA test.

I	II	III	IV
338	412	124	389
324	387	353	432
268	400	469	255
147	233	222	133
309	212	111	265

Solution:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
Between the groups	62111.6	8	9078.067
Within the groups	98787.8	16	4567.89
Total	167771.4	24	

Since F = MST / MSE

= 9.4062 / 3.66

F = 2.57

Example: Enlist the results in APA format after performing ANOVA on the following data set:

[nmeansd 3050.2610.45 3045.3212.76 3053.6711.47] **n** 30 30 30**mean**50.2645.3253.67 **sd**10.4512.7611.47

Solution:

Variance of first set = (10.45)2 = 109.2

Variance of second set = (12.76)2 = 162.82

Variance of third set = (11.47)2 = 131.56

MSerror = {109.2 + 162.82 + 131.56} / {3}

= 134.53

MSbetween = (17.62)(30) = 528.75

F = MSbetween / MSerror

= 528.75 / 134.53

F = 4.86

APA writeup: F(2, 87)=3.93, p >=0.01, η2=0.08.

Articles Related to ANOVA Formula:

- *Variance and Standard Deviation*
- *How to Calculate Variance?*
- *Frequency Distribution*

ANOVA formula is an important tool for anyone involved in statistical analysis or research. It provides a robust method for comparing the means of multiple groups and determining whether observed differences are statistically significant. By breaking down the total variance into components attributable to various factors and within-group variation, ANOVA helps to identify whether these differences are due to random chance or actual effects. Mastery of the ANOVA formula enables more precise and reliable conclusions.

ANOVA Formula – FAQs

How does one set the hypothesis for ANOVA?

The equality of the means of distinct groups must be tested in an ANOVA test. As a result, the hypothesis is as follows:

- *H0 = 1 = 2 = 3 =… = k = Null Hypothesis*
- *H1 is Alternative Hypothesis: The means are not equal.*

How do you calculate the ANOVA?

ANOVA compares group means' differences. Calculate Grand Mean, Between-Group Variability (SSB), and Within-Group Variability (SSW). Determine significance through variance comparison.

What is meant by ANOVA statistic?

The sample mean of the jth treatment of a grouping or a mass data sample is called the ANOVA statistic. It is denoted by the alphabet f.

What is the p value in ANOVA?

In ANOVA, a shared P-value is initially obtained. A significant P-value in the ANOVA test suggests statistical significance in at least one pair's mean difference. Multiple comparisons are then employed to identify these significant pair(s).

What do you mean by one-way ANOVA?

One-way ANOVA is a form of ANOVA test that is used when just one independent variable is present. It compares the means of the various test groups. A test of this type can only provide information on the statistical significance of the means; it cannot establish which groups have different means.

What is accuracy of the ANOVA test?

Since it is more versatile and requires fewer observations, ANOVA analysis is sometimes thought to be more accurate than t-testing. It is also more suited to employ in more sophisticated studies than those that can be evaluated by testing.

Two-Way ANOVA

ANOVA test is a method that compares means of three or more groups, assessing if differences are significant by analyzing variation within and between groups using F-ratio. Two-way ANOVA is used to estimate how the mean of a quantitative variable changes according to the levels of two categorical variables.

ANOVA formula is made up of numerous parts. The best way to tackle an ANOVA test problem is to organize the formulae inside an ANOVA table. Below are the ANOVA formulae.

Source of Variation	Sum of Squares	Degree of Freedom	Mean Squares	F Value
Between Groups	$SSB = \Sigma nj(\bar{X}j - \bar{X})^2$	df1 = k – 1	MSB = SSB / (k – 1)	• f = MSB / MSE • F = MST/MSE
Error	$SSE = \Sigma nj(\bar{X} - \bar{X}\ j)^2$	df2 = N – k	MSE = SSE / (N – k)	
Total	SST = SSB + SSE	df3 = N – 1		

where,

- F is ANOVA Coefficient
- MSB is Mean of the total of squares between groupings
- MSW is Mean total of squares within groupings
- MSE is Mean sum of squares due to error
- SST is Total Sum of squares
- p is Total number of populations
- n is Total number of samples in a population
- SSW is Sum of squares within the groups
- SSB is Sum of squares between the groups
- SSE is Sum of squares due to error
- s is Standard deviation of the samples
- N is Total number of observations

Limitations of One Way ANOVA Test

Various limitations of one way ANOVA test are:

- Assumes homogeneity of variance.
- Sensitive to outliers.
- Assumes normality of data distribution.
- Works best with equal sample sizes.
- Does not identify specific group differences.
- Multiple post-hoc tests increase Type I errors.
- Cannot assess interactions between variables.

Two Way ANOVA

Two-way ANOVA is a statistical analysis method used to assess how two independent variables (factors) affect a dependent variable simultaneously. It examines both the main effects of each factor and any interaction effects between them.

This method allows researchers to understand not only the individual effects of each factor but also how they may interact to influence the outcome. Two-way ANOVA is commonly applied in experimental research to study complex relationships between variables.

When to Use a Two-Way ANOVA

Two way ANOVA formula is used to in various cases including:

- Use a Two-Way ANOVA when you want to analyze the effects of two categorical independent variables on a continuous dependent variable.
- It helps determine if there are significant interactions between the two independent variables and if each independent variable has a significant main effect.
- Useful when studying how two factors simultaneously influence the dependent variable.
- Allows for comparison of means across multiple groups formed by the combinations of the two independent variables.
- Helps in understanding whether the effects observed are due to one variable, the other, or both.

Two-Way ANOVA Assumptions

Various assumptions to use Two-way ANOVA are:

- **Equal Variance:** The variability within each group is roughly the same.
- **Normality:** The data within each group are normally distributed.
- **Independence:** Observations within each group are independent of each other and across groups.
- **Homogeneity of Regression Slopes:** The relationship between the independent and dependent variables is consistent across all groups.

Two-Way ANOVA: Examples

Various examples where two way ANOVA concepts are used:

- **Medicine Experiment**: Testing the effect of two types of medicine (A and B) on patients from different age groups (young and old).
- **Crop Yield Study**: Analyzing the impact of two fertilizers (X and Y) on crop yield across different soil types (sandy and loamy).
- **Education Intervention**: Evaluating the effectiveness of two teaching methods (traditional and online) on students from different socioeconomic backgrounds (low-income and high-income).
- **Marketing Campaign**: Assessing the influence of two advertising strategies (social media and television) on customer response among different geographical regions (urban and rural).
- **Fitness Program Evaluation**: Investigating the effects of two exercise routines (aerobic and strength training) on fitness levels among various age categories (teens, adults, seniors).

So, two-way ANOVA is a powerful statistical method for conducting analyses on the influence of two independent variables on a dependent variable in the simultaneously. Researchers often utilize two-way ANOVA for two-fold reasons namely to recognize and appreciate the way the different factors impart their influence on the outcomes in experimental studies.

What is two-way ANOVA?

Two-way ANOVA is a statistical analysis method used to evaluate the effects of two independent variables on a dependent variable simultaneously.

How does two-way ANOVA differ from one-way ANOVA?

While one-way ANOVA considers variation in one factor, two-way ANOVA examines the influence of two factors and their interaction on the dependent variable.

When is two-way ANOVA appropriate?

Two-way ANOVA is suitable when studying the combined effects of two factors on an outcome, especially in experimental research with multiple variables.

What insights does two-way ANOVA provide?

Two-way ANOVA helps researchers understand not only the individual effects of each factor but also how they interact to influence the outcome variable.

What are the assumptions of two-way ANOVA?

Assumptions include homogeneity of variance, normality of data distribution, independence of observations, and equal group sizes.

How is two-way ANOVA interpreted?

Interpretation involves assessing main effects of each factor and interaction effects between factors to determine their significance in influencing the dependent variable.

Can two-way ANOVA be used for observational studies?

Yes, two-way ANOVA can be applied to observational studies, but researchers must ensure that assumptions are met and causal inferences are appropriately interpreted.

4. Probability

Probability is the determination of the probability of an event occurring. The birth of the Probability Theory began with the passion of a gambler. A noble Frenchman named Chevalier de Méré was seized by the passion of increasing his wealth by gambling. The winner would be the one who rolled the double dice once in 24 times (total 6+6=12). But he soon saw that this rule earned less. He asked his friend Blaise Pascal (1623-1662) why this was the case. Pascal was one of the best mathematicians of that period. Until then, the world of mathematics did not know that games of chance had anything to do with mathematics. Pascal investigated the answer to the question asked to him with the eyes of a mathematician. Finally, he came up with a simple but definite solution. While Chevalier's chance of winning was 51.8% with the old rule, it was 49.1% with the new rule. This was the reason why Chevalier lost. Pascal did not stop at solving this simple problem. He understood that there was a larger mathematical theory behind the problem. He began exchanging ideas with his contemporary Pierre de Fermat through correspondence. Eventually, they created the Probability Theory, an important branch of mathematics. Today, probability theory has long since outgrown its application to games of chance and has entered every field that affects modern human life, such as science, industry, economics, sports, and management. For example, many fields such as banking, insurance, quality control in industry, genetics, kinetic theory of gases, statistical mechanics, and quantum mechanics cannot survive without probability theory.

Some of the important mathematicians who developed Probability Theory are: Blaise Pascal (1623-1662), Pierre de Fermat (1601-1665), Christiaan Huygens (1629-1695), Jakob Bernoulli (1645-1705), Abraham de Moivre (1667-1754), Daniel Bernoulli (1700-1782), Comte de Buffon (1707-1788), Pierre-Simon Laplace (1749-1827), Augustus De Morgan (1806-1871), Thomas Bayes (1702-1761), Andrei Andreyvich Markov (1856-1922), Richard von Mises (1883-1953).

The birth of probability theory as a branch of mathematics dates back to the mid-17th century. In the 20th century, two important steps were taken towards the formalization of probability. The first was the axioms put forward by Andrey Nikolaevich Kolmogorov (1903–1987) in 1933. These axioms brought probability theory to a measurement space and elevated it to a very general structure that included all discrete calculations related to probability up until then as special cases. The second was the axioms put forward by Richard Threlkeld Cox (1898–1991). Cox took probability as a primitive concept that could not be reduced to anything simpler and revealed the basic properties it provided. In the 1960s, without knowing each other, Ray Solomonoff, Anrey Kolmogorov, and Gregory J. Chaitin

introduced the concept of algorithmic randomness. This new concept not only clarified the concept of randomness on which probability theory is based, but also brought new perspectives to information theory. Algorithmic non-uniformity, which is similar to Gödel's incompleteness theorem, is a very active topic today and seems likely to take probability from where it is and push it even higher.

Basic Laws of Probability:

The mathematical expressions of the laws of probability are given by the following relations, where p is the probability function and A is the event set. A is called a data set.

1) The sum of the probabilities of the events forming a data set of A is equal to 1.
2) The probability of any event in a data set of A occurring, $1 \geq p(A) \geq 0$
3) If the probability of any event occurring in a set is 1, the probability of the others occurring is zero.
4) The sum of the probability of any event occurring and the probability of not occurring in a set is equal to 1, $p(A') + p(A)=1$.
   The probability of not occurring in a set is $p(A')= 1- p(A)$.
5) The probability of two independent events occurring together,
   $p(A \text{ and } B)= p(A)*P(B)$
6) The probability of two dependent events occurring together,
   $p(A \text{ and } B)=p(A)*p(B \mid A)=p(B)*p(A \mid B)$;
   $p(B \mid A)$: Probability of event B occurring after event A,
   $p(A \mid B)$: Probability of event A occurring after event B.
7) Probability of one of two independent events occurring,
   $p(A \text{ or } B)=p(A \cup B)=p(A)+p(B)$
8) Probability of one of two dependent events occurring,
   $p(A \text{ or } B)=p(A \cup B)=p(A)+p(B)-p(A \cap B)$

4.1. Addition Rule in Independent Events

The conjunctions of A and B are prevented. Since A and B are mutually exclusive events, the occurrence of event A or event B is equal to the sum of the simple probabilities of these events. It is denoted by (A+B). The probabilities of mutually exclusive events are,

$$P(A+B) = P(A) + P(B)$$

P(A veya B) = P(A) + P(B)

Example:

A customer entering a store will buy a navy blue suit with a 20% probability, a black suit with a 30% probability, a plaid suit with a 40% probability, and a brown suit with a 10% probability. What is the probability that this customer will buy either a navy blue or a black suit from the store?

$P(A \cup B)=P(A) + P(B)=0.2 +0.3=0.5$

4.2. Addition Rule in Dependent Events

If events are related to each other, the occurrence of event A or event B means that either event A or event B or both events A and B occur together. If A and B do not prevent each other;

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

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

$P(A \cup B \cup C)=P(A) + P(B) + P(C) – P(A \cap B) – P(A \cap C) – P(B \cap C) + P(A \cap B \cap C)$

Probability of A intersection B = $P(A \cap B)$

Probability of A union B = $P(A \cup B)$

Example:

In a department store, there are boxes of T-shirts with the sizes and colors given in the table below. We know that one T-shirt is sold.

	Blue	Red	White	Black
Small	3	4	7	6
Middle	5	4	3	3
Large	6	5	4	5

a) What is the probability that the T-shirt sold is White?
P(W)=14/55=0,2545 (% 25)

b) What is the probability that the T-shirt sold is a small size?
P(S)=20/55=4/11=0,36 (%36)

c) What is the probability that the T-shirt sold is white or small?
P(W or S)=P(B) + P(S) – P(B ∩ S)
=14/55 + 20/55 - 7/55 =27/55=0,49 (%49)

d) What is the probability that the T-shirt sold is white and medium size?
P(B ∩ O)=3/55=0,0545 (%5,45)

4.3. Multiplication Rule for Independent Events

In the multiplication rule, events that do not prevent each other.
The multiplication rule for independent events,

P(A and B)= P(A) * P(B)

The probability of independent events occurring simultaneously is equal to the product of the probabilities of these events occurring separately.

Example:
If Person A has an 80% chance of surviving after 15 years and person B has a 60% chance of surviving after 15 years, what is the probability that both of them will survive after 15 years?

P(A and B)=0.80 * 0.60 = 0.48 (48%)

If P(A)=%1,
P(A and B)=0.01 * 0.60 = 0.006

Example:
There are 8 free tickets and 2 winning tickets in a lottery. What is the probability of a person who buys 2 tickets from this lottery winning the jackpot? (non-refundable)

The probability of the first ticket winning is 2/10. If the first ticket wins the jackpot, there are 9 tickets left, 8 free tickets and 1 winning ticket. The probability of the second ticket winning is 1/9. The probability of both tickets winning the jackpot:

P(B1ve B2)= (2/10)*(1/9) = 1/45

Example:

There are 6 lemons and 4 oranges in a box.

a) What is the probability that the first of two citrus fruits drawn from the bag in order with returns is an orange and the other is a lemon?
   With returns, Independent events:
   P(P and L)= P(P)*P(L), The orange is drawn and put back into the box.
   P(P and L)=4/10 * 6/10=24/100=6/25

b) What is the probability that the first of two citrus fruits drawn from the bag in order without returns is an orange and the other is a lemon?
   Without returns, Dependent events:
   P(P and L)= P(P)*P(L /P), The probability that a lemon is drawn from the remaining oranges after they are drawn and separated from the box.
   The first is an orange, the second is a lemon, P(P and L)=4/10 * 6/9 =24/90=8/30=4/15
   The first is a lemon, the second is an orange, P(L and P)=6/10 * 4 /9 =24/90=4/15

c) What is the probability that two citrus fruits drawn from the bag in sequence without returns are both lemons?
   No returns, Interdependent events: P(L and L)= P(L)*P(L /L), The probability of drawing a lemon from the remaining ones after the lemon is drawn and separated from the box.
   P(L and L)= 6/10 * 5/9=30/90=1/3

4.4. Multiplication Rule in Dependent Events

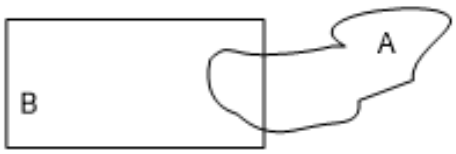
4.4.1. Conditional Probability

Here, the conditional probability of event B occurring under the assumption that event A has occurred, where A and B are two events defined in the same sample space, is

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

Using this definition, we can calculate the probability of events A and B occurring together using conditional probability.

P(A∩B)=P(B|A)P(A) The probability of the common intersection of set A and set B is equal to the product of the probability of set A and the probability of set B after the probability of set A. Thus, the remaining part of set A in B is defined with the expression P(B|A).



in the same way, $P(A \cap B) = P(A|B)P(B)$ is written. Then,
$P(B|A)P(A) = P(A|B)P(B)$ is obtained.

Example:

45% of the candidates failed the statistics exam, 35% failed the computer exam and 25% failed both the computer and statistics exams.
If they failed the computer, the probability of failing the statistics exam,
If they failed the statistics exam, the probability of failing the computer exam,
The one with the smallest probability will be selected. Which one will you choose?

Calculate the probability of failing at least one of these two courses (Statistics or Computer).

İ: Students who failed the statistics course
B: Students who failed the computer course.
P(İ)=0.45, P(B)=0.35, P(A ∩B)=0.25
P(İ|B)=(P(İ∩B))/(P(B))=0.25/0.35=5/7
P(B|İ)=(P(İ∩B))/(P(İ))=0.25/0.45=5/9
If they failed the statistics exam, the one who failed the computer exam will be selected.
P(İ ∪B)=P(i)+P(B)-P(İ∩B)=0.45+0.35-0.25=0.55

4.4.2. Total Probability Rule

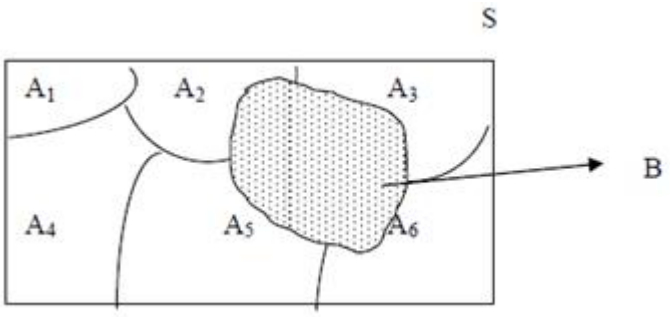
When the probability of a B event cannot be calculated directly, the total probability rule is used. For example, when a random product is taken from the 6 machines in a factory, the probability of this product being defective is investigated. Here, if the B event is the defective product, this product could have been produced in one of the machines A1, A2, ..., A6.

$$P(B) = \sum_{i=1}^{n} P(B|Ai)P(Ai) \quad = \sum_{i=1}^{n} P(Ai \cap B)$$

The above equation is called the total probability rule.

Here, i=1,2, ... ,n

Ai ∩ $Aj = 0, i \neq j$



Example:

In a factory, 50% of the goods produced are produced by machine 1, 30% by machine 2, and 20% by machine 3. It has been observed that 3%, 4%, and 5% of the goods produced by these machines are defective, respectively. What is the probability that one of the goods selected is defective?

Ai: Machines that produce products. (i=1,2,3)
B: A collection set for defective products.

P(A1)=0.50
P(A2)=0.30
P(A3)=0.20

P(B|A1)=0.03; The probability that the defective items in machine A1 will be collected in machine B

P(B|A2)=0.04; The probability that the defective items in machine A2 will be collected in machine B

P(B|A3)=0.05; The probability that the defective items in machine A3 will be collected in machine B

Since the items produced come out of these three machines, event B occurs together with one of the events A1, A2, A3. In this case, it can be written as the sum of three separate events.

$P(B)=P(B \cap A1) + P(B \cap A2) + P(B \cap A3)$

$P(B)= P(B|A1)P(A1) + P(B|A2)P(A2) + P(B|A3)P(A3)$
$P(B)= 0.03*0.50 + 0.04*0.30 + 0.05*0.20$
$P(B)=0.037$

The probability that a randomly selected product is defective is 0.037. In other words, if 1000 products are purchased from this factory, the expected defective value among these 1000 products will be 37.

4.5. Bayesian Theory

It is a theorem developed by Thomas Bayes and used in calculating conditional probabilities. In the event that more than one independent cause is effective in the occurrence of an event, it provides convenience in calculating the probability of which independent cause will occur.

Let P(Ai)>0 and Ai ∩ Aj =0 for each i ≠ j. For any event B defined in the sample space S, provided that P(B)>0, the conditional probability of event A under the assumption that event B has occurred or the probability of event A occurring in the set B,

If defective products are collected in B, the probability that any of these defective products came from the selected machine,

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)} = \frac{P(A_j \cap B)}{\sum_{i=1}^{n} P(A_i \cap B)} = \frac{P(A_j \cap B)}{P(B)}, \qquad 1 \leq j \leq n \; olur.$$

Bayesian theory plays a critical role in determining what is most likely to happen. According to Bayesian theory, in order to find the probability of an undesirable situation (B) occurring due to a known cause (P(Aj |B)), the probability of the intersection of the undesirable situation and the known cause with the undesirable situations (P(Aj∩B)) must be divided by the sum of the probabilities of the undesirable situations (P(B)).

Example:
A large office uses 3 photocopiers. 60% of the shots are made on the first machine, 30% on the second machine, and 10% on the third machine. The waste rates of the machines are 10% for the first machine, 20% for the second machine, and 40% for the third machine. What is the waste rate for all copies?
Let M1, M2, M3 machines, and D represent the faulty copy (waste).

Probability of using the machines:
P(M1) = 0.60
P(M2) = 0.30
P(M3) = 0.10

Waste rates:
P(D\M1) = 0.10
P(D\M2) = 0.20
P(D\M3) = 0.40

Probability of wasting after being pulled on machine M1: P(D\M1)= P(M1 ∩ D) / P(M1)
Probability of wasting after being pulled on machine M2: P(D\M2)= P(M2 ∩ D) / P(M2)
Probability of wasting after being pulled on machine M3: P(D\M3)= P(M3 ∩ D) / P(M3)

If we show the events with A , B , C;
It can be written as A = M1 ∩ D , B = M2 ∩ D , C = M3 ∩ D . P(A) = P(M1 ∩ D) = P(M1)*P(D\M1)= (0.60)*(0.10) = 0.06
P(B) = P(M2 ∩ D) = P(M2)*P(D\M2)= (0.30)*(0.20) = 0.06,
P(C) = P(M3 ∩ D) = P(M3)*P(D\M3)= (0.10)*(0.40) = 0.04

The sum of the calculated waste probabilities for each machine will give us the total waste probability:
P(D) = P(A) + P(B) + P(C) = 0.06 + 0.06 + 0.04 = 0.16

We can say that 16% of all copies are wasted.

Example:
A company produces 100 units from three companies. The distribution of machines is as follows: A1 machine produces 60 units, A2 machine produces 30 units and A3 machine produces 10% of the products produced. A1 machine produces 10% defective products, A2 machine produces 20% and A3 machine produces 10% defective products.
a) What is the probability that one of the products produced by the company is defective?
b) What is the probability that the defective product is Type-B?
c) Find and comment on which machine group the defective products come from with the lowest and highest probability.
B: Probability of a product being defective?
Ai: Let i=1,2,3 be the events that the machine comes from company 1, 2 or 3.
P(A1)=0.60, P(A2)=0.30, P(A3)=0.10

Number of A1 defective products= 60*10/100 = 6
Number of A2 defective products= 30*20/100 = 6
Number of A3 defective products= 10*10/100 = 1
Total number of products =100, 13 out of 100 products are defective.

Probability of A1 products being defective, P(B|A1)=0.10,
Probability of A2 machines being defective, P(B|A2)=0.20
Probability of A3 machines being defective, P(B|A3)=0.10

P(B) =13/100 is the probability that the machine is faulty.

P(B)= P(A1)P(B|A1) + P(A2)P(B|A2) + P(A3)P(B|A3)
P(B)= 0.60*0.10 + 0.30*0.20 + 0.10 * 0.10
P(B)=0.06 + 0.06 + 0.01=0.13
13% of the machines purchased from this company will be defective.

a) If one of the company's products is defective, the probability that this product is A2 can be found by using Bayesian theory.

$$P(A_2|B) = \frac{P(B|A_2)P(A_2)}{\sum_{i=1}^{3} P(B|A_i)P(A_i)} = \frac{0.20 * 0.3}{0.13} = 0.46$$

Although the company produces 30% of the machines with A2 machines, 46% of the faulty machines come from A2 machines.

b) If the machine purchased by the company turns out to be defective, the probability that this machine is A1 can be found by using Bayesian theory.

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{\sum_{i=1}^{3} P(B|A_i)P(A_i)} = \frac{0.10 * 0.60}{0.13} = 0.46$$

Although the company purchases 30% of the machines from Company 2, 46% of the defective machines come from Company A1.

c) If the machine purchased by the company turns out to be defective, the probability that this machine is A3 can be found by using Bayesian theory.

$$P(A_3|B) = \frac{P(B|A_3)P(A_3)}{\sum_{i=1}^{3} P(B|A_i)P(A_i)} = \frac{0.10 * 0.10}{0.13} = 0.08$$

Although the company purchases 10% of the machines from a 3rd company, 8% of the faulty machines come from A3rd company.

5. Random Variables

There are mainly 3 types of distribution in data sets:
• Clustered distribution
• Regular (Uniform) distribution
• Random distribution

Clustered distribution: This is the most common distribution in data sets. It is the clustering of the elements forming the data sets in certain areas. Clustering is usually seen in plants in areas where environmental conditions are suitable for development. Starfish can form groups in tidal pools where they can easily find food and mate. Individuals in a flock form groups, facilitating hunting, feeding, and defense, while also increasing the rate of utilization from the environment where needs are met.

Regular (Uniform) distribution: This is the distribution type seen when there is competition between individuals for insufficient resources in challenging environmental conditions. Individuals are relatively equidistant from each other. This distribution is not a common distribution type. Individuals directly affect each other in distribution. For example, some plants secrete chemicals that prevent the germination, development and growth of individuals they compete with for limited resources. Thus, they create a living space for themselves. Uniform distribution is seen in king penguins and some pine tree species in the Falkland Islands located in the south of the Atlantic Ocean.

Random Distribution: In random distribution, the state of each element in the data sets is independent of that of the other elements. In such distributions, situations where elements interact with each other, such as attracting or repelling each other, also occur. For example, in plants where pollen is pollinated by wind, random distribution is seen in the spring months, which is the pollination period.

Variables that form a space cluster (Data Set) and take certain probability values are called random variables. Random is the equivalent of the word random in English.

Statistics is defined as a branch of science that is used to mathematically model a random event and to make inferences about the unknown characteristic features of the population (Data Set) such as mean and variance with the help of this model. Modeling of a random event is done with the help of variables that are expressed with numerical values and are called random variables. For example, if the grade distribution in a class is between 0 and

100, the number of students who receive their grade probabilities are known; here, the grade ranking is expressed as a discrete random variable.

Even random trial results that cannot be expressed with numbers are transformed into a form that can be expressed with numbers and thus probability functions can be determined. For example, whether a product is defective or not, the winner or loser of a competition, the marital status and educational status of a person, etc. In a production process, the numerical values of probabilty of defective goods = 1 and Non-Faulty goods = 0 can be given.

The probability values that any random event can take when it occurs in a sample space define the probability distribution. Probability values should cover all possible outcomes for the event and the sum of the probabilities of the outcomes related to the event should be equal to one. **The probability value of each event in a sample space is between 0 and 1.**

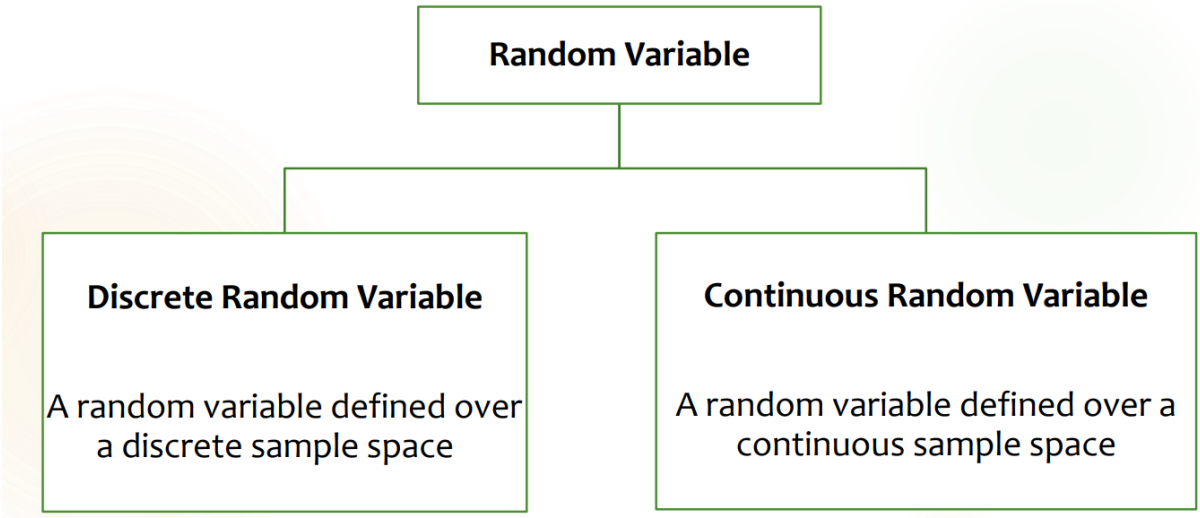
The different values that the units that form a whole in terms of any feature are called variables.

For example, in the function y(t)=at^2+bt+c, t is the independent variable and fonction is continuous; y(t) is the dependent variable; a, b are the coefficients and c is constant values. In the definition range of a random variable, the values that the relevant function will take are variables that are not known in advance but the probabilities of taking these values can be calculated; y(t) is unknown in the example above; but the probability values or probability functions that it will take according to the variables are known.

y'(t)=2at+b; this derivative gives the trajectory of the main function.
y''(t)=2a, It answers the questions of whether it is concave, convex or if it has a region.

In statistical terminology, the set of random variables is shown with capital letters (X, Y, Z,... etc.), and the values of random variables are shown with lower case letters (x, y, z,... etc.).

The function that shows the relationship between the values that a random variable can take and the probabilities of taking these values is called the probability function. For discrete variables, the probability function is the table that shows the values that the variable has taken and the probabilities of taking these values. For continuous variables, the probability function is given names such as probability density function or frequency function.



Examples:

Discrete Random Variable:

1. X= Number of correct answers in a 100-MCQ test= 0, 1, 2, …, 100
2. X= Number of cars passing a toll both in a day= 0, 1, 2, …, ∞
3. X= Number of balls required to take the first wicket = 1, 2, 3, …, ∞

Continuous Random Variable:

1. Mathematical functions
2. X= Monthly Profit. $-\infty<X<\infty$
3. Exchange market, Stock Exchange market

In the calculation of probability values, if discrete probability functions are used, **the sum sign (Σ)** is used, while if the probability density function (continuous) is used, the integral **sign (∫) is used.**

Random Variables:

- Discrete Random Variables
- Continuous Random Variables
- Expectation of a Random Variable
- Variance of a Random Variable
- Jointly Distributed Random Variables
- Combinations and Functions of Random Variables

Example: In a dice-throwing experiment, two dice are thrown at the same time. The sample space (S), which is expressed as the set of all possible outcomes of this experiment, is obtained as shown below.

$$S = \begin{Bmatrix} (1,1),(1,2),(1,3),(1,4),(1,5),(1,6); \\ (2,1),(2,2),(2,3),(2,4),(2,5),(2,6); \\ \dots \\ (6,1),(6,2),(6,3),(6,4),(6,5),(6,6) \end{Bmatrix}$$

The set of values that the random variable X takes is called the "Value Set".

The set of values that the random variable X takes is expressed as {2,3,4,5,6,7,8,9,10,11,12}. If I roll a double dice 100 times and write down the values, I can calculate the probability values.

Types of random variables:
Random variables are called discrete or continuous depending on the values they take. Random variables whose value set is countable are called discrete, and random variables whose value set is uncountable are called continuous.

The probability density function is expressed as the definition range of the random variable X and the probability function for this range. The set of situations X formed by random variables is defined as a function and the values that any random variable in this set can take are shown as xi. If the number of values that a random variable x can take is finite or countable, **X is called a "Discrete Random Variable function".** If a random variable x, which is the result of a random experiment, can take all values in a certain range, that is, if it covers an interval on the number line, X is called a "Continuous Random Variable function". The function of a random variable X consists of a range of possible values.

Examples of values that a discrete random variable (X) can take:

- $x = 0,1$
- $x = 0,1,2,3,\dots$
- $x = k, k+1, k+2,\dots$

Examples of values that a continuous random variable (X) can take:

- $0 < x < 1$
- $0 < x < \infty$
- $-\infty < x < \infty.$

5.1. Probability Distribution

Properties of Probability Distribution:

1.

$$P(x) \geq 0, \quad if \quad X \ is \ discrete.$$

$$f(x) \geq 0, \ if \ X \ is \ continuous$$

2.

$$\sum_x P(X = x) = 1, \quad if \quad X \ is \quad discrete.$$

$$\int_x f(x)dx = 1, \quad if \ is \ continuous.$$

Note:

1. If X is a continuous random variable then

$$P(a < X < b) = \int_a^b f(x)dx$$

2. Probability of a fixed value of a continuous random variable is zero.

$$\Rightarrow P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b)$$

3. If X is discrete random variable then

$$P(a < X < b) = \sum_{x=a+1}^{b-1} P(x)$$

$$P(a \leq X < b) = \sum_{x=a}^{b-1} p(x)$$

$$P(a < X \leq b) = \sum_{x=a+1}^{b} P(x)$$

$$P(a \leq X \leq b) = \sum_{x=a}^{b} P(x)$$

4. Probability means area for continuous random variable.

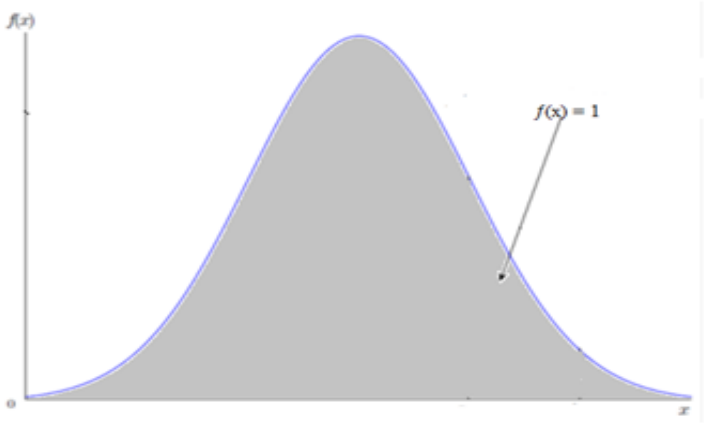
5.1.1. Probability Distribution in Continuous Random Variables

If X is a continuous random variable; X ∈ {-∞,∞}, **the function f(x) is defined as the probability density function of random variable.** The probability of the continuous random variable X being in a certain interval can be calculated with the "Probability Density Function".

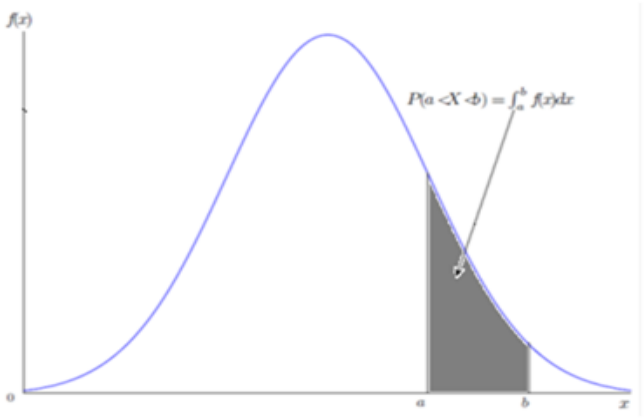
The "Probability Density Function" has the following properties:
The probability density function is denoted by ƒ(x). For all values of x, 1≥ ƒ(x) ≥ 0 (the curve does not intersect the horizontal axis).

The area under the probability density function curve is equal to 1.

$$\int_{-\infty}^{\infty} f(x)dx = 1$$



The probability that the random variable X is a value between the values a and b in the interval (a,b) where a<b is the area between these two values under the Probability Density Function.



$$P(a < X < b) = \int_a^b f(x)dx$$

X rassal değişkeninin R'deki değer kümesi A, sayılmaz küme ise x'e sürekli rassal değişken denir.

If the value set of the random variable x in Real is A and it is an uncountable set, then x is called a continuous random variable.

$$A = \{x | a \leq X \leq b\}$$

These mathematical models, which are a function of random variables, are called probability or probability density functions.

example:

$$f(x) = \begin{cases} cx^2, & 0 \leq x \leq 3 \\ 0, & for\ other\ x\ values \end{cases}$$

What value should c take for the function above to be a "probability density function"? In the probability density function, the integral is between 0 and infinity.

$$f(x) = \int_0^{\infty} f(x)dx = \int_0^3 cx^2 dx = 1$$

$$= \mathrm{c}\left(\frac{\mathrm{x}^3}{3}\right)|_0^3 = \mathrm{c}\left(\frac{3^3}{3}\right) - \mathrm{c}\left(\frac{0^3}{3}\right) = \frac{27}{3}\ c = \frac{27}{3}\ c = 1'den$$

$$c = \frac{3}{27}$$

Örnek:

$$f(x) = \begin{cases} \frac{3}{27}x^2, & 0 \le x \le 3 \\ 0, & for\ other\ x\ values \end{cases}$$

Find the probability distribution function F(x)?

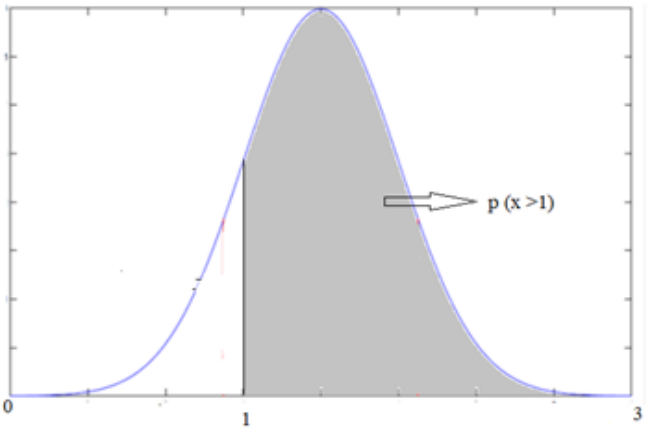
$$F(x) = \int_0^x f(x)dx = \int_0^x \frac{3}{27}x^2 dx = \ = \frac{3}{27}\left(\frac{x^3}{3}\right)|_0^x = \frac{3}{27}\left(\frac{x^3}{3}\right) = \frac{x^3}{27}$$

Calculate F (0≤ x≤2)?

$F(2) = \ p\ (x\ \le 2) = \frac{3}{27}\left(\frac{2^3}{3}\right) \ = \ \frac{8}{27}$ = 0.3

The probability that the random variable X is less than 2 is 30 percent.

Calculate the probability density function of P (x ≥ 1)?

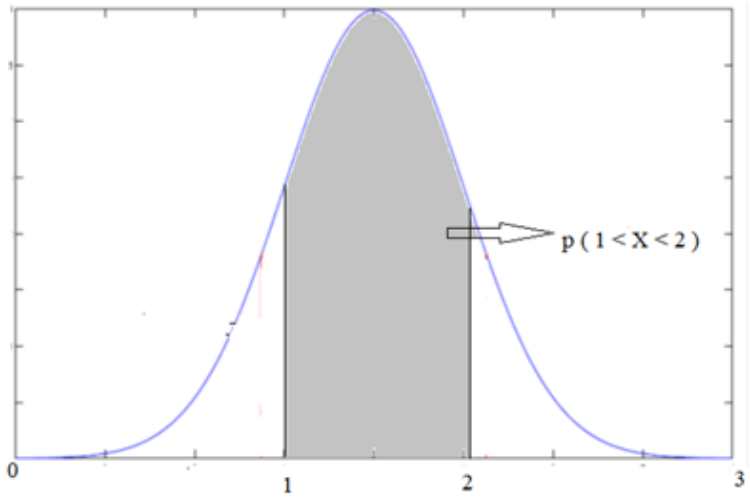


$$p\ (x\ \ge 1) = \int_1^3 \frac{3}{27}x^2 dx \ = \frac{3}{27}\left(\frac{x^3}{3}\right)|_1^3 = \ \frac{3}{27}\left(\frac{3^3}{3}\right) - \frac{3}{27}\left(\frac{1^3}{3}\right)$$

$$= \frac{27}{27} - \frac{1}{27} = \frac{26}{27} = 0.96$$

The probability that the random variable X is greater than 1 is 96 percent.

Calculate the probability density function of P (1 ≤ X ≤ 2)?



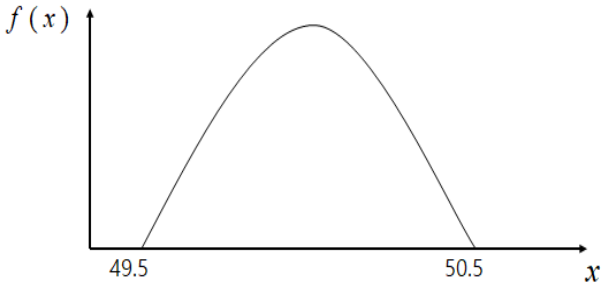
$$p\,(1 \le x \le 2) = \int_1^2 \frac{3}{27}x^2 dx = \frac{3}{27}\left(\frac{x^3}{3}\right)\Big|_1^2$$

$$= \frac{3}{27}\left(\frac{2^3}{3}\right) - \frac{3}{27}\left(\frac{1^3}{3}\right) = \frac{8}{27} - \frac{1}{27} = \frac{7}{27} = 0.26$$

The probability that the random variable X is between 1 and 2 is 26 percent.

Example: Suppose the diameter of a metal cylinder has a probability density function.

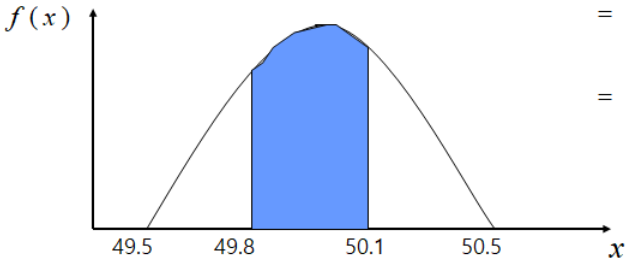
$$f(x) = 1.5 - 6(x - 50.2)^2 \quad \text{for} \quad 49.5 \le x \le 50.5$$
$$f(x) = 0, \quad \text{elsewhere}$$



$$\int_{49.5}^{50.5} (1.5 - 6(x - 50.0)^2)\,dx = [1.5x - 2(x - 50.0)^3]_{49.5}^{50.5}$$
$$= [1.5 \times 50.5 - 2(50.5 - 50.0)^3]$$
$$-[1.5 \times 49.5 - 2(49.5 - 50.0)^3]$$
$$= 75.5 - 74.5 = 1.0$$

The probability that the diameter of a metal cylinder is between 49.8 and 50.1 mm can be calculated as follows:

$$\int_{49.8}^{50.1} (1.5 - 6(x - 50.0)^2)\,dx = [1.5x - 2(x - 50.0)^3]_{49.8}^{50.1}$$
$$= [1.5 \times 50.1 - 2(50.1 - 50.0)^3]$$
$$- [1.5 \times 49.8 - 2(49.8 - 50.0)^3]$$
$$= 75.148 - 74.716 = 0.432$$



$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)\,dy$$

$$f(x) = \frac{dF(x)}{dx}$$

$$P(a < X \le b) = P(X \le b) - P(X \le a)$$
$$= F(b) - F(a)$$

$$P(a \le X \le b) = P(a < X \le b)$$

$$F(x) = P(X \le x) = \int_{49.5}^{x} (1.5 - 6(y - 50.0)^2)\,dy$$
$$= [1.5y - 2(y - 50.0)^3]_{49.5}^{x}$$
$$= [1.5x - 2(x - 50.0)^3] - [1.5 \times 49.5 - 2(49.5 - 50.0)^3]$$
$$= 1.5x - 2(x - 50.0)^3 - 74.5$$

Example: Calculate the probability density functions of the continuous random variable X given below.

$$f(X) = \begin{cases} 3X^2 & 0 \le X \le 1 \\ 0 & \text{other X's} \end{cases}$$

$a)\ P(X \le 0.5)$

$b)\ P(X \ge 0.7)$

$c)\ P(0.2 < X < 0.8)$

$$a)\ P(X \le 0.5) = \int_0^{0.5} 3x^2 dx = x^3 I_0^{0.5} = 0.5^3 = 0.125$$

$$b)\ P(X \ge 0.7) = \int_{0.7}^{1} 3x^2 dx = x^3 I_{0.7}^{1} = 1^3 - 0.7^3 = 0.657$$

$$c)\ P(0.2 < X < 0.8) = \int_{0.2}^{0.8} 3x^2 dx = x^3 I_{0.2}^{0.8} = 0.8^3 - 0.2^3 = 0.504$$

5.1.2. Probability Distribution in Discrete Random Variables

If we distribute all possible outcomes of a random experiment according to their probabilities, we obtain a probability distribution. Probability distributions are the discrete probability distributions of the random variable they belong to, and the binomial, hypergeometric and poisson distributions are the most common discrete probability distributions we encounter in practice. If the random variable is continuous, a continuous probability distribution emerges.

In the probability function of a discrete random variable, x: Discrete random variable, P(x): x is the probability function of the random variable.

The probability function for discrete variables is the table showing the values that the random variable has taken and the probabilities corresponding to these values. Therefore, the probability function for the discrete random variable X can be written as follows.

$$P(X_i) = p_i \quad i = 1,2,3\cdots,n \text{ or } i = 1,2,3,\cdots$$

The probability function for discrete variables can be written in a table form as follows.

$$\text{For every } X_i,\ 0 \leq P(X_i) \leq 1$$

$$\sum_{i=1}^{n} P(X_i) = 1$$

If the values that a random variable X can take in the probability distribution are,
x1 , x2 , x3 ,…. , xn , then the probabilities of these values are defined as, P(x1), P(x2), P(x3),…. , P(xn).

All possible values that the random variable X can take are shown with their probabilities as follows:

$X=x_i$	x1	x2	x3	……	xn
$P(X=x_i)$	P(x1)	P(x2)	P(x3)	……	P(xn)

Example: The probability function for the random variable X is given as follows. What value should the constant k take in order for X to be a probability function? Note: The function (Polynomial) is obtained from the series by the regression method.

$$P(X) = \begin{cases} k(X+1) & X = 1,2,3 \\ 0 & \text{Other X's} \end{cases}$$

$$\sum P(X_i) = 1$$

$$\sum_{i=1}^{3} k(X+1) = k[2+3+4] = 9k = 1 \Rightarrow k = \frac{1}{9}$$

Example: Probability function of the random variable X?

X	P(X)
0	2/7
1	4/7
2	1/7

$$P(X=1) = \frac{4}{7} \quad ; \quad P(X<2) = P(X=0) + P(X=1) = \frac{2}{7} + \frac{4}{7} = \frac{6}{7}$$

Example:

$$p(x) = \left\{ \left(\frac{a}{b}\right)^{x} \left(\frac{c}{d}\right)^{1-x} \right.$$ to be a probability function

- $0 \le a/b \le 1$, $0 \le c/d \le 1$
- $a/b + c/d = 1$ should be.

Example: When a dice is thrown, the result is a random variable (X), and the possible results are xi = 1, 2, 3, 4, 5 and 6, and the probability of each of them occurring is 1/6. In this case, the probability function of this event will be as follows.

$P_x(x) = P(X=x) = 1/6 > 0$

$\sum_{x=1}^{n} P(X = x) = 1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 1/6 = 1$

Example: A pair of dice is thrown. A finite equally probable sample space S containing 36 ordered pairs of numbers between 1 and 6 is obtained. S={(1,1), (1,2), ... , (6,6)}, X(S)={1,2,3,4,5,6}. Let X correspond to the largest of all the numbers (a,b) in S. Find the weighted average of X.

The f distribution of X is,

f(1) = P(X =1) = P{(1,1)} = 1/36
f(2) = P(X =2) = P{(2,1), (2,2), (1,2)} = 3/36
f(3) = P(X =3) = P{(3,1), (3,2), (3,3), (2,3), (1,3)} = 5/36
f(4) = P(X =4) = 7/36
f(5) = P(X =5) = 9/36
f(6) = P(X =6) = 11/36

x_i	x_1	x_2	x_3	x_4	x_5	x_6
$f(x_i)$	1/36	3/36	5/36	7/36	9/36	11/36

$$E(\mathrm{X}) = \mu = \sum_{i=1}^{n} x_i f(x_i) = 1 \cdot \frac{1}{36} + 2 \cdot \frac{3}{36} + 3 \cdot \frac{5}{36} + 4 \cdot \frac{7}{36} + 5 \cdot \frac{9}{36} + 6 \cdot \frac{11}{36} = 4.47$$

Örnek:

$p(x) = \left\{\left(\frac{3}{5}\right)^{x} \left(\frac{2}{5}\right)^{1-x}\right.$, x=0,1 is a discrete random variable. In this case, p(x) is the probability function.

Condition-1:

P(X=0)=2/5, 0≤ P(X=0)≤ 1
P(X=1)=3/5, 0≤ P(X=1)≤ 1

Condition-2:

2/5 + 3/5 =1

Since conditions 1 and 2 are met, p(x) is a probability function.

5.2. Cumulative Probability Distribution Function

In probability theory and statistics, the cumulative distribution function is a function that completely describes the probability distribution of a real-valued random variable X. It is also called the probability distribution function or simply the distribution function.

The **Cumulative Distribution Function (CDF)** of a random variable is a mathematical function that provides the probability that the variable will take a value less than or equal to a particular number.

The CDF starts at 0 for the smallest possible value of X and increases to 1 as x approaches the largest possible value of X. It is a non-decreasing function that provides a complete description of the distribution of the random variable.

For example, if you're looking at the CDF for a test score of 80, and it gives you 0.75, this means there's a 75% chance that a random student's score will be 80 or less.

In short, the CDF helps you understand the likelihood of a random value being within a certain range by summing up probabilities as you go along.

The Cumulative Distribution Function F(x) of the random variable X is defined as:
$F(x)=P(X\leq x)$

Properties of Cumulative Distribution Function

Every cumulative distribution function, F, generally (but not necessarily invariably) exhibits four properties.
F is monotonically increasing.
F is continuous from the right.

Monotonicity: The CDF is a non-decreasing function. This means that for any two values x1 and x2 such that $x1\leq x2$ the corresponding CDF values satisfy $F(x1)\leq F(x2)$.

Limits: As x approaches negative infinity the CDF approaches 0. 2. F is continuous from the right.

$$\lim_{x\to-\infty} F(x) = 0$$

As x approaches positive infinity the CDF approaches 1:

$$\lim_{x\to\infty} F(x) = 1$$

Continuity: The CDF of a continuous random variable is a continuous function while the CDF of the discrete random variable has jumps or discontinuities at specific points.

Non-Negativity: The CDF is always non-negative i.e. F(x)≥0$F(x)$≥0 for the all x.

The probability that X, a discrete random variable, will not exceed any x_0 value can be represented by the Cumulative Probability Function as follows.

$F(x_0) = P(X \leq x_0) = \sum_{x \leq x_0}^{n} P(X = x)$

The cumulative probability function of a discrete random variable has two properties.

- For every x0 value, it is between $0 \leq F(x\ 0) \leq 1$.
- If $x_0 < x_1$ ise , $F(x_0) \leq F(x_1)$

How to Calculate Cumulative Distribution Function?

Steps to find cumulative distribution function are given below-

Step 1. Identify the Distribution: The Determine whether the random variable follows a discrete or continuous distribution.

Step 2. Determine the PDF: For continuous distributions find the Probability Density Function (PDF) f(x).

Step 3. Integrate the PDF: The Integrate the PDF from the $-\infty$ to x to find the CDF:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Step 4. Sum the Probabilities: For discrete distributions sum the probabilities for the all values less than or equal to the x.

Using Probability Density Function (PDF) to Find CDF

For continuous random variables the CDF F(x) is derived from the PDF f(x) by the integrating:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

his process accumulates the probability from the left up to the point x.

Types of Cumulative Distribution Functions

CDF of a Discrete Random Variable

For a discrete random variable X with the possible values x1,x2,…,xn*x*1,*x*2,…,*xn* and corresponding probabilities P(X=xi)=pi the CDF is given by:

$$F(x) = \sum_{x_i \leq x} P_i$$

The CDF of the discrete random variable increases in steps at the points where the variable takes on the specific values.

CDF of a Continuous Random Variable

For a continuous random variable X with the probability density function (PDF) f(x) the CDF is the integral of the PDF.

When to Use CDF vs. PDF?

- CDF: Use the CDF when we need to find the probability that a random variable is less than or equal to the specific value. It provides the cumulative probability up to the certain point.
- PDF: Use the PDF when you need to the find the probability density at a specific value. The PDF represents the rate of the change of the CDF and is useful for the calculating the probabilities over.

Example : Let X be a mixed random variable with the following distribution: P(X=0) = 0.2, P(X=1) = 0.3 and the continuous part is uniformly distributed over [2, 3]. Find the CDF F(x).

Solution:

For x < 0, F(x) = 0.

For $0 \leq x < 1$, *F(x) = 0.2.*

For $1 \leq x < 2$, *F(x) = 0.5.*

For $2 \leq x < 3$ *the continuous part applies:*

$F(x) = 0.5 + 0.5 \times (x - 2) = 0.5 + 0.5\,x - 1$

Thus, $F(x) = 0.5\,x - 0.5 for 2 \leq x \leq 3.$

For $x \geq 3$, *F(x) = 1.*

Example: Given a PDF $f(x) = \frac{3}{4}(1 - x^2)$ for the $-1 \leq x \leq 1$ find the CDF F(x).

Solution:

Integrate the PDF to find the CDF:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-1}^{x} \tfrac{3}{4}(1 - t^2)\,dt$$

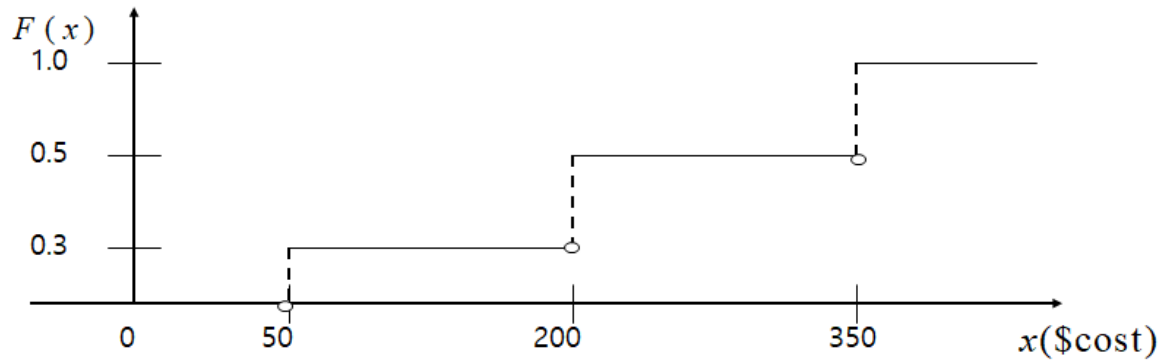
Solving the integral:

$$F(x) = \tfrac{3}{4}\left[t - \tfrac{t^3}{3}\right]_{-1}^{x} = \tfrac{3}{4}\left(x - \tfrac{x^3}{3} + \tfrac{2}{3}\right)$$

The CDF (F(x)) is:

$$F(x) = \begin{cases} 0 & \text{if } x < -1 \\ \frac{3}{4}\left(x - \frac{x^3}{3} + \frac{2}{3}\right) & \text{if } -1 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$

Example: Machine Failures



$-\infty < x < 50 \Rightarrow F(x) = P(\text{cost} \leq x) = 0$

$50 \leq x < 200 \Rightarrow F(x) = P(\text{cost} \leq x) = 0.3$

$200 \leq x < 350 \Rightarrow F(x) = P(\text{cost} \leq x) = 0.3 + 0.2 = 0.5$

$350 \leq x < \infty \Rightarrow F(x) = P(\text{cost} \leq x) = 0.3 + 0.2 + 0.5 = 1.0$

Example:
A child psychologist is interested in the number of times a newborn baby's crying wakes its mother after midnight. For a random sample of 50 mothers, the following information was obtained. Let *X* = the number of times per week a newborn baby's crying wakes its mother after midnight. For this example, *x* = 0, 1, 2, 3, 4, 5.

P(*x*) = probability that *X* takes on a value *x*.

x	P(x)
0	P(x = 0) = 2/50
1	P(x = 1) = 11/50
2	P(x = 2) = 23/50
3	P(x = 3) = 9/50
4	P(x = 4) = 4/50
5	P(x = 5) = 1/50

X takes on the values 0, 1, 2, 3, 4, 5. This is a discrete PDF because we can count the number of values of *x* and also because of the following two reasons:

a. Each *P*(*x*) is between zero and one, therefore inclusive
b. The sum of the probabilities is one, that is,
   2/50+11/50+23/50+9/50+4/50+1/50=1

Example: Write the function of the probability of a thrown dice being less than 3. A thrown dice being less than 3 is the case of a thrown dice being 1 or 2.
$F(x_0) = P(X \leq x_0) = \sum_{x=1}^{2} P(X = 1,\ X = 2) = 1/6 + 1/6 = 2/6$

Example: Write the function of the probability of a thrown dice being greater than 1 and less than 6. A thrown dice being greater than 1 and less than 6 is the case of a 2, 3, 4 or 5.

$F(x_0) = P(X \leq x_0) = \sum_{x=2}^{5} P(X = 2,\ X = 3,\ X = 4,\ X = 5) = 4/6$

Example:

When 2 dice are thrown, if the random variable x represents the sum of the numbers on the upper faces of the dice, the options will be between 2 and 12. The probabilities are as follows:

x(şıklar)	2	3	4	5	6	7	8	9	10	11	12
P(x)	1/36	2/36	3/36	4 /36	5/ 36	6/36	5/36	4/36	3/36	2/36	1/36

5.3. Expected Value and Variance in Random Variables

Variables that can take on various values with certain probabilities are called random variables. When sampling from a large dataset, two parameters are of critical importance: weight average and variance.

The expected value or weight average for a continuous random variable (function), x is the random variable, x is the probability density function of the random variable, and f(x) is

$$\mu = E(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx$$

The expected value or weighted average for a discrete random variable (sequence, vector), where f(xi) are the probability function values. The value of f(xi) is the probability of occurrence of the random variable xi. In the set space of finite random variables X, X=(x1, x2, ... , xn), the probability of occurrence of xi is written as f(xi) and the function defined as P(X= xi) is called the distribution or probability function of X.

$$\mu = E(\mathrm{X}) = \sum_{i=1}^{n} x_i f(x_i)$$

$$E(X^2) = \sum_{i=1}^{n} x_i^2 f(x_i)$$

$$V(X) = E(X^2) - E(X)^2$$

Note: If E(X)>0 then the expectation is appropriate, if E(X)=0 then the expectation is reasonable, if E(X)<0 then the expectation is not appropriate. Here, since f(xi) defines the probabilities of occurrence of xi random variables, according to the basic laws of probability, $1 \geq f(xi) \geq 0$

$$\sum_{i=1}^{n} f(x_i) = 1$$

Let X be a random variable,

a) The expected value of X for a continuous random variable (weight average):

When the value $\int_{-\infty}^{+\infty}|x|f(x)dx < \infty$ is taken instead of the value $\mu = E(X) = \int_{-\infty}^{+\infty} xf(x)dx$, it indicates that X is stable.

b) Expected value of X for discrete random variable (weight Average):

Considering the value of $\sum_{i=1}^{n}|x_i|\ f(x_i) < \infty$ at the value of $\mu = E(X) = \sum_{i=1}^{n} x_i f(x_i)$, it can be said that the random variables xi forming X follow a stable trajectory. In the sample space of X, the xis form trajectories. Trajectory: is the path or time followed.

Moment:

The expression $E(X-\mu)^k$ is called the kth moment of X about point a.

In classical mechanics, momentum is the product of an object 's mass and its velocity ; (p = m v) . Like velocity , momentum is a vector quantity , meaning that it has a direction as well as a magnitude.

Variance:

Variance is a measure of distribution. It is an informative measure about the structure of the distribution of units around the arithmetic mean (homogeneous or heterogeneous). The positive square root of the variance is called the standard deviation. The standard deviation always takes positive values. As the standard deviation value approaches zero, the homogeneity (homogeneity) in the distribution of the relevant variable will increase.

The value of $E(X-\mu)^2$ is called the variance of X and is denoted by Var(X). Var(X)≥ 0,

Variance in the population or data set: $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{N}$

Variance for the selected sample in the dataset: $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{N-1}$

Variance in random variables, $V(X) = \sum_{i=1}^{N}(x_i-\mu)^2 f(x_i)$

While the expected value provides information about the center of the distribution, the variance provides information about the spread around the expected value.

$$V(X) = \sum_{i=1}^{N} (x_i^2 - 2\mu x_i + \mu^2) f(x_i) = \sum_{i=1}^{N} x_i^2 f(x_i) - 2\mu \sum_{i=1}^{N} x_i f(x_i) + \mu^2 \sum_{i=1}^{N} f(x_i)$$

From the fundamental theory of probability $\sum_{i=1}^{N} f(x_i) = 1$ dir.

$$V(X) = \sum_{i=1}^{N} x_i^2 f(x_i) - 2\mu\mu + \mu^2 = E(X^2) - \mu^2 = E(X^2) - E(X)^2$$

$$V(X) = E(X^2) - E(X)^2$$

Expected value and Variance properties:

Let, c is a constant number

X and Y are two independent random variables

1. E(c) = c
2. E(c X) = c E(x)
3. E(X + c) = E(x) + c
4. E(X+Y) = E(X) + E(Y)
5. E(X-Y) = E(X) - E(Y)
6. E(XY) = E(X) . E(Y)

1. Var(c) = o
2. Var(c X) = c^2 Var(x)
3. Var(X + c) = Var(x)
4. Var(X+Y) = Var(X) + Var(Y)
5. Var(X-Y) = Var(X) + Var(Y)

X represents a random variable set space. Here, the following properties can be written for the expected value and variance, where a and b are coefficients. There may be difficulties in interpreting the motion along a line in the orbit, so the weighted average and variance of the motion as a translation or envelope may be desired.

$$E(aX \pm b) = aE(X) \pm b$$

In the X sample set space, if the probabilities of the states in the orbit are known, the probabilities of this new state can be calculated quickly if the orbit is changed with the coefficients a and b. When there is any deviation in the stable states, the weight average and variance of the new state are calculated.

$Var(aX \pm b)=a^2Var(X)$

$E(aX)=aE(X)$

$E(aX+b)= aE(X)+b$

$E(X+Y)= E(X) + E(Y)$

Exemple:

If $E(3X-6)=12$, what is the weighted average?

$3E(X) - 6=12$; $3E(x)=12+6=18$

$E(X)=6$

$E(X^2)=36.5$ ise $var(3X-6)=?$

$V(X) = E(X^2) - E(X)^2$

$V(x)=36.5-36=0.5$

$V(3X-6)=9* (0.5)^2= 2.25$

$T=X_1 + X_2 + \dots + X_N$ is taken $E(T)=E(X_1 + X_2 + \dots + X_N)=E(X_1)+E(X_2)+ \dots + E(X_N)$

$Var(X) = E(X^2)- E^2(X)= E(X^2)- \mu^2$

$\sigma(X) = \sqrt{\mathrm{Var(X)}}$

Properties of variance:

- $V(a)=0$, where a is a constant. In other words, the variance of the constant is zero.
- $V(aX)=a^2V(X)$, where a is a constant
- $V(aX+b)=V(aX)+V(b)=a^2 V(X)$, where a and b are constants

Example: Let the variance of the random variable X be 3. Calculate the variances of the random variables X+4 and -X+8.

V(X)=3 iken V(X+4)=V(X)=3 ve V(-X+8)=V(X)=3

Example: Calculate the expected value and variance of the random variable Y, where Y=3X-5, E(X)=4, Var(X)=2.

E(Y)=E(3X-5)=3E(X)-5=3*4-5=12-5=7

Var(Y)=Var(3X-5)=9Var(X)=9*2=18

Example: What is the expected value of the random variable Y, where Y=X2+3X, E(X)=10, Var(X)=6?

$Var(X)=E(X^2)-E(X)^2$
$6= E(X^2)-100$
$E(X^2)=106$
$E(Y)=E(X^2+3X)=E(X^2)+3E(X)=106+3*10=136$

5.3.1. Expected Value and Variance in Discrete Random Variables

Purpose: To make estimates and predictions about the population based on sample information. The expected value or weighted average and variance are calculated for a discrete random variable.

$p(x)$ has ALL the same properties as a probability and so we have:

1. $0 \leq p(x) \leq 1$ for ALL values of x
2. $\sum all\ x\ p(x) = 1$. This

$$\mu = E(\mathrm{X}) = \sum_{i=1}^{n} x_i f(x_i)$$

$$E(X^2) = \sum_{i=1}^{n} x_i^2 f(x_i)$$

Varyans, $V(X) = E(X^2) - E(X)^2$

Let's consider a sample of n objects that is to be selected from a large data set:

- The selection process that gives each possible sample of n objects an equal chance of being selected is called random sampling.
- These inferences are based on a statistic that is a certain function of the sample information drawn from the population.
- The sampling distribution of this statistic is the probability distribution of the values that the statistic in question can take in all samples of the same size that can be drawn from this population.

$X=(x_1, x_2, \dots, x_n)$ represents the probability space X of a finite random variable. Here, the probability of xi occurring is written as f(xi) and the function defined as P(X= xi) is called the distribution or probability function of X. Here, the probability of f(xi) occurring satisfies the conditions $1 \geq f(x_i) \geq 0$ and $\sum_{i=1}^{n} f(x_i) = 1$

x_1	x_2	...	x_n
$f(x_1)$	$f(x_2)$	...	$f(x_n)$

Example:

A discrete random variable X can assume five possible values: 2, 3, 5, 8, and 10. Its probability function is shown in the table:

x	2	3	5	8	10
$p(x) = P(X = x)$	0.15	0.10	0.25	0.25	0.25

Solve for the expected value and variance of X.

Soln:

$E(X) = (2 \times 0.15) + (3 \times 0.10) + (5 \times 0.25) + (8 \times 0.25) + (10 \times 0.25) = 6.35$

$E(X^2) = (2^2 \times 0.15) + (3^2 \times 0.10) + (5^2 \times 0.25) + (8^2 \times 0.25) + (10^2 \times 0.25) = 48.75$

Therefore $Var(X) = E(X^2) - (E(X))^2 = 48.75 - (6.35)^2 = 8.4275$

Example: A discrete random variable X can assume five possible values: 2, 3, 5, 8, and 10. Its probability function is shown in the table:

x	2	3	5	8	10
$p(x) = P(X = x)$	0.15	0.10	?	0.25	0.25

1. Using the properties of $p(x)$ solve for P(X=5).
2. What is the probability X equals 2 or 10?
3. What is $P(X \leq 8)$?
4. What is $P(X < 8)$?

Soln:

1. $P(X = 5) = 1 - (0.15 + 0.10 + 0.25 + 0.25) = 0.25$
2. $P(X = 2 \cup X = 10) = P(X = 2) + P(X = 10) = 0.15 + 0.25 = 0.4$
3. $P(X \leq 8) = P(X = 2) + P(X = 3) + P(X = 5) + P(X = 8) = 0.75$
4. $P(X < 8) = P(X = 2) + P(X = 3) + P(X = 5) = 0.5$

Example:

Let the data set be: {6,9,12,15,18}. Let's assume f(x)=1/5, i=1,2,3,4,5. The probability function can be written as follows.

The probability of each group occurring is examined.

X	6	9	12	15	18
f(x) = P(X=x)	1/5	1/5	1/5	1/5	1/5

Let's find the mean, variance, and median of the population.

$$E(\mathrm{X}) = \mu = \sum_{i=1}^{n} x_i f(x_i)$$

$$E(X^2) = \sum_{i=1}^{n} x_i^2 f(x_i)$$

E(x)=μ=6*1/5 + 9*1/5 + 12* 1/5 + 15* 1/5 + 18*1/5=60/5=12

$E(x^2)$=36/5 + 81/5 + 144/5 + 225/5 + 324/5 =162

Var(X) = σ^2 = $E(X^2)$- μ^2 =162 – 144 =18

Med(X)=12 (The middle value is considered.)

Example: There are 5 departments in the psychology department of a hospital. The number of patients coming to the psychology department in a day is 100. The probability of examining patients in each department is given in the table below.

Hospital Department Number	1	2	3	4	5
Fee Distribution (TL)	200	500	300	600	400
Probability of being examined in each section	3/20	2/20	6/20	C	5/20

a) According to probability theory, what is the probability of patients in group 4 being examined?
   The sum of their probabilities must be equal to 1.
   3/20+2/20+6/20+C+5/20=1; C=4/20

b) Given the total number of patients, find the number of patients examined in each department.
   1. **Departmant, N1=100*3/20=15**
   2. **Departmant, N2=100*2/20=10**
   3. **Departmant, N3=100*6/20=30**
   4. **Departmant, N4=100*4/20=20**
   5. **Departmant, N5=100*5/20=25**

c) What is the weighted average of the income obtained at the end of the day?

   μ=200*3/20+500*2/20+300*6/20+600*4/20+400*5/20

   μ=10*3+25*2+15*6+30*4+20*5 = 30+50+90 + 120 + 100 =390

Example:

The student grades in a class are {30,40,50,60,70,80,90,100} and the probability of students getting these grades is given in the table below. Calculate the Expected Value and Variance using the Discrete Random Sampling equations. The number of students is 48.

X	30	40	50	60	70	80	90	100
f(x) = P(X=x)	1/8	1/6	1/6	1/12	1/8	1/6	1/12	1/12

$$E(\mathrm{X}) = \mu = \sum_{i=1}^{n} x_i f(x_i)$$

E(x)=μ=30*1/8 + 40*1/6 + 50* 1/6 + 60*1/12 + 70*1/8 + 80*1/6 + 90*1/12 + 100*1/12 = 61,67

$$E(X^2) = \sum_{i=1}^{n} x_i^2 f(x_i)$$

$E(x^2)$= 30*30*1/8 + 40*40*1/6 + 50*50* 1/6 + 60*60*1/12 + 70*70*1/8 + 80*80*1/6 + 90*90*1/12 + 100*100*1/12 = 4283,33

Var(X) = σ^2 = $E(X^2)$- μ^2 =4283,33 – 3802,78 =480

Standard deviation =21.92

Example:

Let the student grades in a class be: {60,70,80,90} and the number of students who got these grades be: {12, 12, 12, 12}. Probability functions can be written as follows.

What is the arithmetic mean?
Find the total number of students. 12+12+12+12=48
Arithmetic mean, μ = (60*12+70*12+80*12+90*12)/48 = 75
Maximum value: 90
Minimum value: 60
Half the number of students: 48/2=24
24th student grade: (70+80)/2=75

Calculate the expected value and variance using discrete random sampling equations.

X	60	70	80	90
Number of students	12	12	12	12
f(x) = P(X=x)	12/48	12/48	12/48	12/48
f(x) = P(X=x)	1/4	1/4	1/4	1/4

$$E(\mathrm{X}) = \mu = \sum_{i=1}^{n} x_i f(x_i)$$

E(x)=μ=+ 60*1/4 + 70*1/4 + 80*1/4 + 90*1/4 = 75

$$E(X^2) = \sum_{i=1}^{n} x_i^2 f(x_i)$$

$E(x^2)$ = 60*60*1/4 + 70*70*1/4 + 80*80*1/4 + 90*90*1/4 = 5750

Var(X) = σ^2 = $E(X^2)$- μ^2 =5750 – 5625 =125
Standard deviation =11

Example: Machine Failures

- $P(\text{cost}=50)=0.3$, $P(\text{cost}=200)=0.2$, $P(\text{cost}=350)=0.5$
- $0.3 + 0.2 + 0.5 = 1$

x_i	50	200	350
p_i	0.3	0.2	0.5

Expected repair cost:

$$E(\text{cost}) = (\$50 \times 0.3) + (\$200 \times 0.2) + (\$350 \times 0.5) = \$230$$

$$\begin{aligned} \text{Var}(X) &= E((X - E(X))^2) = \sum_i p_i (x_i - E(X))^2 \\ &= 0.3(50 - 230)^2 + 0.2(200 - 230)^2 + 0.5(350 - 230)^2 \\ &= 17{,}100 = \sigma^2 \end{aligned}$$

$$\sigma = \sqrt{17{,}100} = 130.77$$

Örnek:

X	P(X)	XP(X)	X^2(P(X)
0	1/4	0	0^2 (1/4)=0
1	2/4	2/4	1^1 (2/4)=2/4
2	1/4	2/4	2^2 (1/4)=4/4
		1.0	6/4=3/2

$$E(X) = \mu = \sum_{i=0}^{2} x_i p(x_i) = 1$$

$$E(X^2) = \mu = \sum_{i=0}^{2} x_i^2 p(x_i) = \frac{3}{2}$$

$$V(X) = E(X^2) - \big(E(X)\big)^2 = \frac{3}{2} - 1^2 = \frac{1}{2}$$

Example: The probability function of the random variable X is given below.

a) $E(X^2)$=? b) $E(X^2+X)$=?

X	P(X)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

$$E(X^2) = \sum_{i=1}^{6} X_i^2 P(X_i)$$

$$E(X^2) = 1^2\frac{1}{6} + 2^2\frac{1}{6} + 3^2\frac{1}{6} + 4^2\frac{1}{6} + 5^2\frac{1}{6} + 6^2\frac{1}{6} = \frac{91}{6}$$

$$E(X^2 + X) = \sum_{i=1}^{6} (X_i^2 + X_i) P(X_i)$$

$$E(X^2+X) = (1^2+1)\frac{1}{6} + (2^2+2)\frac{1}{6} + (3^2+3)\frac{1}{6} + (4^2+4)\frac{1}{6} + (5^2+5)\frac{1}{6} + (6^2+6)\frac{1}{6}$$

$$= \frac{1}{6}(2+6+12+20+30+42) = \frac{112}{6}$$

Örnek:

The probability function for the number of defective parts is given as follows, where the random variable X represents the number of defective parts in a production process. Find the variance of the distribution for the number of defective parts.

xi	0	1	2	3
f(xi)	0.51	0.38	0.1	0.01

$$\sum_{0}^{3} f(xi) = 1$$

$$E(X) = \sum_{x=0}^{3} xP(x) = 0 \times 0.51 + 1 \times 0.38 + 2 \times 0.10 + 3 \times 0.01 = 0.61$$

$$E(X^2) = \sum_{x=0}^{3} x^2P(x) = 0^2 \times 0.51 + 1^2 \times 0.38 + 2^2 \times 0.10 + 3^2 \times 0.01 = 0.87$$

$$V(X) = E(X^2) - \mu^2 = 0.87 - 0.61^2 = 0.498$$

5.3.2. Expected Value and Variance in Continuous Random Variables

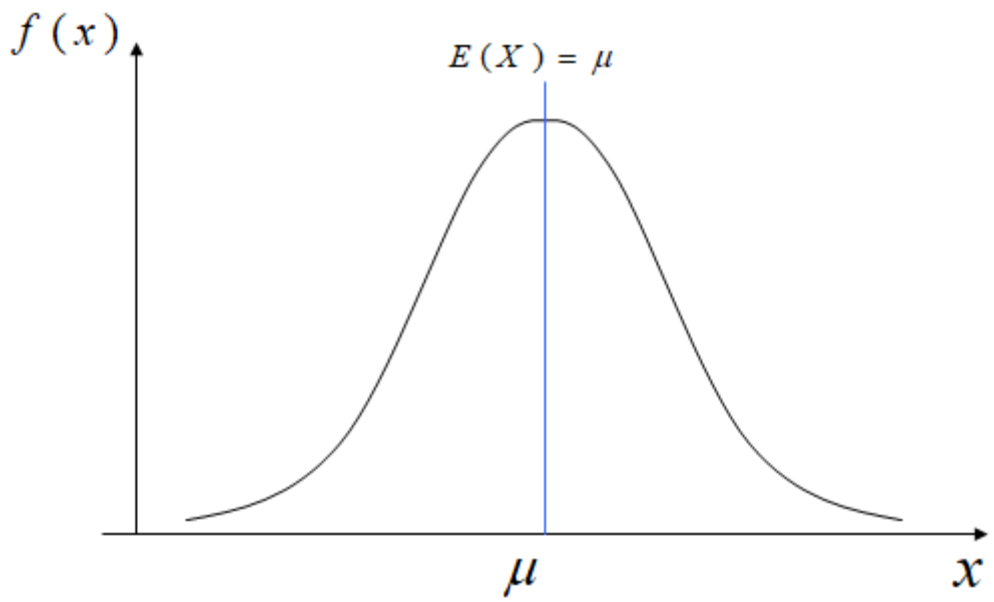
The expected value or weighted average for a continuous random variable (function) is the random variable x, the probability density function of the random variable x, and f(x) is

$$\mu = E(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx$$

$$V(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx = E(X^2) - \mu^2 = E(X^2) - (E(X))^2$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx$$



If x has a probability density function, f(x), with symmetry about the point μ, then the following expressions hold such that the expectation of the random variable is equal to the point of symmetry:

$$f(\mu + x) = f(\mu - x)$$
$$E(X) = \mu$$

Example:

$$f(x) = 1.5 - 6(x - 50.2)^2 \quad \text{for} \quad 49.5 \le x \le 50.5$$
$$f(x) = 0, \quad \text{elsewhere}$$

$$\begin{aligned} F(x) = P(X \le x) &= \int_{49.5}^{x} (1.5 - 6(y - 50.0)^2)\,dy \\ &= [1.5y - 2(y - 50.0)^3]_{49.5}^{x} \\ &= [1.5x - 2(x - 50.0)^3] - [1.5 \times 49.5 - 2(49.5 - 50.0)^3] \\ &= 1.5x - 2(x - 50.0)^3 - 74.5 \end{aligned}$$

$$E(X) = \int_{49.5}^{50.5} x(1.5 - 6(x - 50.0)^2)\,dx$$

Change of variable: x =y+50

$$\begin{aligned} E(x) &= \int_{-0.5}^{0.5} (y + 50)(1.5 - 6y^2)\,dy \\ &= \int_{-0.5}^{0.5} (-6y^3 - 300y^2 + 1.5y + 75)\,dy \\ &= [-3y^4/2 - 100y^3 + 0.75y^2 + 75y]_{-0.5}^{0.5} \\ &= [25.09375] - [-24.90625] = 50.0 \end{aligned}$$

$$F(x) = 1.5x - 2(x - 50.0)^3 - 74.5 = 0.5$$

$$x = 50.0$$

Example: The probability density function of the random variable X is as follows. Calculate the expected value of X**.**

$$f(x) = \begin{cases} \frac{1}{2}x & 0 \le x \le 2 \\ 0 & diğer\ x'leriçin \end{cases}$$

$$E(X) = \int_0^2 xf(x)dx$$

$$E(X) = \int_0^2 x\left(\frac{1}{2}x\right)dx = \frac{1}{2}\int_0^2 x^2dx = \frac{x^3}{6}I_0^2 = \frac{8}{6} = \frac{4}{3}$$

Example: The probability density function of the random variable X is as follows. Calculate the expected value of X.

$$f(x) = \begin{cases} x & 0 < x < 1 \\ 2 - x & 1 \le x \le 2 \\ 0 & diğer\ x'leriçin \end{cases}$$

$$E(X) = \int_0^2 x.\ f(x)\,dx \qquad E(X) = \int_0^1 xxdx + \int_1^2 x(2-x)dx$$

$$= \int_0^1 x^2dx + \int_1^2 (2x - x^2)dx \qquad = \frac{x^3}{3}I_0^1 + \left(x^2 - \frac{x^3}{3}\right)I_1^2$$

$$= \frac{1}{3} + \left[\left(4 - \frac{8}{3}\right) - \left(1 - \frac{1}{3}\right]\right. = \frac{1}{3} + \frac{2}{3} = 1$$

Example:

Calculate the standard deviation of the function f(x).

$$f(x) = \begin{cases} 2(1-x) & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise} \end{cases}$$

Determine whether the function f(x) is a probability function. Note: For the function f(x) to be a probability function, its integral must be equal to 1.

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$\int_{-\infty}^{\infty} f(x)dx = 2\int_{0}^{1} (1-x)dx = 2\left(x - \frac{x^2}{2}\right)\Big|_0^1 = 1$$

Calculate the arithmetic mean of the function f(x). Note: The arithmetic mean of random functions is found with the following expression,

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{\infty} x f(x)dx \\ \mathbb{E}(X) &= \int_{-\infty}^{\infty} x f(x)dx \\ &= \int_{0}^{1} x\Big[2(1-x)\Big]dx \\ &= 2\int_{0}^{1} \Big(x - x^2\Big)dx \\ &= 2\Big(\frac{x^2}{2} - \frac{x^3}{3}\Big)\Big|_0^1 \\ &= 1/3 \end{aligned}$$

Calculate the variance and standard deviation of the function f(x) using the expressions below.

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} \big(x - \mathbb{E}(X)\big)^2 f(x)dx$$

$$\sigma = \sqrt{\mathrm{Var}(X)}$$

$$
\begin{aligned}
\operatorname{Var}(X) &= \int_{-\infty}^{\infty} \big(x - \mathbb{E}(X)\big)^2 f(x)dx \\
&= \int_{0}^{1} \big(x - 1/3\big)^2 \cdot 2(1-x)dx \\
&= 2\int_{0}^{1} \big(x^2 - \frac{2}{3}x + \frac{1}{9}\big)(1-x)dx \\
&= 2\int_{0}^{1} \big(-x^3 + \frac{5}{3}x^2 - \frac{7}{9}x + \frac{1}{9}\big)dx \\
&= 2\big(-\frac{1}{4}x^4 + \frac{5}{9}x^3 - \frac{7}{18}x^2 + \frac{1}{9}x\big)\Big|_0^1 \\
&= 2\big(-\frac{1}{4} + \frac{5}{9} - \frac{7}{18} + \frac{1}{9}\big) \\
&= \frac{1}{18}
\end{aligned}
$$

Therefore, the standard deviation of *X* is

$$
\begin{aligned}
\sigma &= \sqrt{\operatorname{Var}(X)} \\
&= \frac{1}{3\sqrt{2}}
\end{aligned}
$$

f(x) fonksiyonun varyansını ve standart sapmasını aşağıdaki ifadeleri kullanarak hesaplayınız

$$\operatorname{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$$

$$
\begin{aligned}
\operatorname{Var}(X) &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \\
&= \int_{-\infty}^{\infty} x^2 f(x)dx - \left(\frac{1}{3}\right)^2 \\
&= 2\int_{0}^{1} x^2(1-x)dx - \frac{1}{9} \\
&= 2\int_{0}^{1} \big(x^2 - x^3\big)dx - \frac{1}{9} \\
&= 2\big(\frac{1}{3}x^3 - \frac{1}{4}x^4\big)\Big|_0^1 - \frac{1}{9} \\
&= 2\big(\frac{1}{3} - \frac{1}{4}\big) - \frac{1}{9} \\
&= \frac{1}{18}
\end{aligned}
$$

Example:

The probability density function of the random variable X is as follows. Calculate the variance of X.

$$f(X) = \begin{cases} 2x & 0 < X < 1 \\ 0 & \text{For other X's} \end{cases}$$

$$V(X) = \int_0^1 (X-\mu)^2 f(x)dx$$

$$\mu = E(X) = \int_0^1 xf(x)dx = \int_0^1 x\ 2x\ dx = 2\int_0^1 x^2 dx = 2\ \frac{x^3}{3} I_0^1 = \frac{2}{3}$$

$$V(X) = \int_0^1 (x-\frac{2}{3})^2 2xdx = \int_0^1 \left(x^2 - \frac{4}{3}x + \frac{4}{9}\right) 2xdx$$

$$= \int_0^1 \left(2x^3 - \frac{8}{3}x^2 + \frac{8}{9}x\right) dx = \frac{1}{2}x^4 - \frac{8}{9}x^3 + \frac{8}{18}x^2 I_0^1 = \frac{1}{18}$$

$$E(X^2) = \int_0^1 x^2 f(x)dx$$

$$= \int_0^1 x^2 2xdx = \int_0^1 2x^3 dx = \frac{2}{4}x^4 I_0^1 = \frac{1}{2}$$

$$V(X) = E(X^2) - \left(E(X)\right)^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$$

Example:

Find the standard deviation of the random variable X whose probability density function is given below.

$$f(X) = \begin{cases} \frac{3}{10}(3x - x^2) & 0 \le x \le 2 \\ 0 & \text{For other X's} \end{cases}$$

$$E(X) = \mu = \int_0^2 xf(x)dx = \int_0^2 x\frac{3}{10}(3x - x^2)dx$$

$$= \left(\frac{3}{10}x^3 - \frac{3}{40}x^4\right)\Big|_0^2 = \left(\frac{3}{10}2^3 - \frac{3}{40}2^4\right) - 0 = 1.2$$

$$E(X^2) = \int_0^2 x^2 f(x)dx = \int_0^2 x^2\frac{3}{10}(3x - x^2)dx$$

$$= \left(\frac{9}{40}x^4 - \frac{3}{50}x^5\right)\Big|_0^2 = \left(\frac{9}{40}2^4 - \frac{3}{50}2^5\right) - 0 = 1.68$$

$$V(X) = \sigma^2 = E(X^2) - \mu^2 = 1.68 - 1.2^2 = 0.24$$

$$\sigma = \sqrt{0.24} = 0.489$$

6. Probability Distribution of Discrete Random Variables

In probability theory, a discrete random variable is a type of random variable that can take on a finite or countable number of distinct values. These values are often represented by integers or whole numbers, other than this they can also be represented by other discrete values.

A very basic and fundamental example that comes to mind when talking about discrete random variables is the rolling of an unbiased standard die. An unbiased standard die is a die that has six faces and equal chances of any face coming on top. Considering we perform this experiment, it is pretty clear that there are only six outcomes for our experiment.
Thus, our random variable can take any of the following discrete values from 1 to 6. Mathematically the collection of values that a random variable takes is denoted as a set. In this case, let the random variable be X.
Thus, X = {1, 2, 3, 4, 5, 6}

Another popular example of a discrete random variable is the number of heads when tossing of two coins. In this case, the random variable X can take only one of the three choices i.e., 0, 1, and 2.

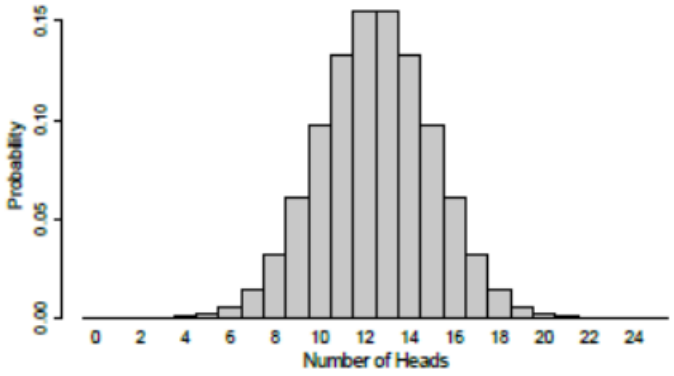
Other than these examples, there are various other examples of random discrete variables. Some of these are as follows:

- The number of cars that pass through a given intersection in an hour.
- The number of defective items in a shipment of goods.
- The number of people in a household.
- The number of accidents that occur at a given intersection in a week.
- The number of red balls drawn in a sample of 10 balls taken from a jar containing both red and blue balls.
- The number of goals scored in a soccer match.

Types of Discrete Random Variables:

- Binomial Random Variable
- Geometric Random Variable
- Bernoulli Random Variable
- Poisson Random Variable

To properly define a discrete random variable and to solve for probabilities we use a Probability mass function (pmf)/ Probability function (pf), probability density function (pdf), $\boldsymbol{p(x)}$ For discrete random variables: $\boldsymbol{p(x)} = \mathbf{P}\ (\boldsymbol{X} = \boldsymbol{x})$ represents the probability that X (the r.v.) takes on the value x. This is called a probability function, and it allocates a probability for every value of x. Visually, it is represented by a histogram



6.1. Bernoulli Distribution

In many experiments, two different results occur. For this reason, the Bernoulli distribution is used to calculate the probability of events with two outcomes. Random variables: 0 and 1. An exam result can be defined in two states as successful and unsuccessful, or a product purchased for quality control can be in two forms as intact or defective. If experiments with two outcomes are tried once, it is called the "Bernoulli Distribution". The Bernoulli process is a process in which one of two mutually exclusive outcomes occurs in each experiment. In the coin toss experiment, only one of the possible outcomes of heads (T) and tails (Y) occurs in each experiment, and the probabilities of P(Y) and P(T) are the same in each toss since the experiments are independent of each other, (1 / 2).

Bernoulli deneylerinde ortaya çıkan iki sonuçtan biri *başarı* diğeri ise *başarısızlık* olarak adlandırılır; p başarı olasılığını, q'da başarısızlık olasılığını göstermektedir, 1- p = q.

X raslantı değişkeni, x alabileceği tüm değerler ise başarı için 1, başarısızlık için 0 değerini alsın. X'in olasılık fonksiyonu;

$P(X=x)=p^x q^{1-x}$

If x=0 or1, this distribution is called the Bernoulli probability distribution function. ($p^0=1$, $q^0=1$)

P(X=1) = p

P(X=0) = q= 1-p

Arithmetic Mean and Variance of Bernoulli Distribution:

Arithmetic mean of Bernoulli random variable: $\mu_x = E(x) = p$
The variance of the Bernoulli random variable is: $\sigma_x^2 = E[(X - \mu_x)^2] = pq$
Asymmetry (Skewness) and Kurtosis Coefficients:

$$ÇK = \frac{q-p}{\sqrt{pq}} \quad ve \quad BK = \frac{1}{pq} - 6$$

For an event with probability p of occurring, the probability of not occurring is 1 − p. The ratio of the probability of an event occurring to the probability of it not occurring is called its reciprocal.

$$karşıtlık = \frac{p}{q} = \frac{p}{1-p}$$

Expected value and Variance:

$$E(X) = \sum_{x=0}^{n} xP(x)$$

$$E(X^2) = \sum_{x=0}^{n} x^2P(x)$$

$$V(X) = E(X^2) - \mu^2$$

$$\begin{aligned} E(X) &= (0 \times P(X = 0)) + (1 \times P(X = 1)) \\ &= (0 \times (1 - p)) + (1 \times p) \\ &= p \end{aligned}$$

$$\begin{aligned} E(X^2) &= (0^2 \times P(X = 0)) + (1^2 \times P(X = 1)) \\ &= p \end{aligned}$$

$$\begin{aligned} Var(X) &= E(X^2) - (E(X))^2 \\ &= p - p^2 = p(1 - p) \end{aligned}$$

Example:
If p(all three successful)=0.075 in three consecutive Bernoulli trials, find the probability that all three will fail. q = 1-p = 1-0.075 =0.925

Example:

What must q be for the following expression to be a Bernoulli probability distribution function?

$P(X = x) = p^x (q)^{1-x}$

$q=1-p$

$p+q=1$

What values does x take in the expression (0 and 1).

Example:

In a population of 200 people, 120 people are non-smokers and 80 people are smokers, meaning 60% are non-smokers. When a random person is selected from this population, the probability that the selected person is a non-smoker is p=0.6, and the random variable X takes the value 1 for non-smoker and 0 for smoker. X is a Bernoulli random variable.

$P(X=x)=0.6^x 0.4^{1-x}$; x=0,1

The opposite of not smoking versus drinking = $\frac{p}{q}$ = 0.6/0.4 = 3/2 For every 3 non-smokers, 2 people smoke.

Example:

A student believes that he has a 70% chance of passing a Physics course. Write the probability distribution function? Find its mean and variance?

If the student passes the course and the random variable X takes the value x = 1 and fails and the random variable X takes the value x = 0, the probability distribution of the random variable X can be written as:

P (x=1) = 0.7 and P (x=0) = 0.3

Probability distribution function:

$P(X = x) = p^x (1 - p)^{1-x} = 0.7^x (1 - 0.7)^{1-x} = 0.7^x * 0.3^{1-x}$ is found as.

The probability distribution of x for the values 1 and 0 is as follows.

$P(x = 1) = p^1 (1 - p)^{1-1} = 0.7^1 (1 - 0.7)^0 = 0.7^1 * 0.3^0 = 0.7$

$P(x = 0) = p^0 (1 - p)^{1-0} = 0.7^0 (1 - 0.7)^1 = 0.7^0 * 0.3^1 = 0.3$

Arithmetic mean: $\mu_x = E(X) = p = 0.7$

Variance: $\sigma_x^2 = E[(X - \mu_x)^2] = p(1 - p) = 0.7*0.3 = 0.21$ is found as..

Example:
If defined as drawing a red ball in 6 balls, the probability of success (P) would be 1/6 or 0.17. The probability of blue (drawing a blue ball) would be 5/6 or 0.83. The probability of failure for any Bernoulli run is always 1 - P.

The probability distribution function is:
$P (X = x) = p^x (1 – p)^{1-x} = 0.17^x (1 – 0.17)^{1-x} = 0.7^x * 0.83^{1-x}$ is found as.

x'in alacağı 0 ve 1 değerlerine göre olasılık dağılımı aşağıdaki gibi bulunur.
$P (x = 0) = p^0 (1 – p)^{1-0} = 0.17^0 (1 – 0.17)^1 = 0.17^0 * 0.83^1 = 0.83$
$P (x = 1) = p^1 (1 – p)^{1-1} = 0.17^1 (1 – 0.17)^0 = 0.7^1 * 0.83^0 = 0.17$

Arithmetic mean: $\mu_x = E (X) = p = 0.17$
Variance: $\sigma_x^2 = E [(X - \mu_x)^2] = p (1 – p) = 0.17*0.83 = 0.14$ is found as.

6.2. Binom Distribution

If a random experiment with two results is repeated n times under the same conditions, a distribution called the "Binomial Distribution" is obtained. It is a special form of the Bernoulli distribution. The binomial distribution is the most widely used of the discrete distributions. Conditions that the binomial distribution must satisfy:

- The experiment is repeated a certain number of times (n).
- Each experiment has two results, successful and unsuccessful.
- The experiments are independent of each other.
- The probability of success is p and the probability of failure is q=1-p.
- The successful results obtained in n experiments are assigned to the x variable.

In the experiments given below, X is the binomial random variable defined:

- A coin is tossed 10 times. X is the number of heads
- A jar containing 8 black and 4 white balls is replaced and 3 balls are drawn. X is the number of black balls drawn.
- A box containing 3 defective and 7 perfect pieces is replaced and 4 pieces are selected. X is the number of defective pieces selected.

If the experiment is repeated n times, the total number of successful cases is a random variable indicated by x. This variable is called a binomial variable. In order to accept the variable x as a binomial variable, the repeated experiments must be the same, the probabilities must not change from experiment to experiment, and the choices must be made with returns.

If the p and q probabilities in the binomial distribution are equal, the shape of the distribution will be symmetrical. Since the p and q probabilities are the same in the coin toss experiment, the distribution will be symmetrical. In cases where p is not equal to q, the shape of the distribution is asymmetrical.

First of all, since there are n trials, there will be n two-probability outcomes. If the event being examined is success or failure, there will be x successful and (n-x) unsuccessful outcomes in n trials, and since the trials are independent of each other, the probability of any sequence of results is equal to the product of the probabilities of the individual results and is as follows.

$P(x, n-x) = p.p.p\ldots\ldots p.(1-p).(1-p)\ldots\ldots.(1-p) = p^x\, q^{n-x}$

Each trial has a probability of success p and a probability of failure (1-p). In n random trials, x successes can occur with (n-x) failures in many different sequences, only one of which was considered above.

If the sequence order is not important, the number of sequences with x successes in n random trials is:

$C_n^x = \frac{n!}{x!(n-x)!}$ (x-grained combinations of n random trials) is found as.

Binomial probability function of x trials (x=0,1,2, …, n) being successful in n random trials:

$P(X; n,p) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$

$P(X; n,p) = C_n^x p^x q^{n-x}$

It is written as and calculated using this formula. x = 0, 1, 2, 3, ……n
n: Number of times the experiment is repeated
X: Number of desired results
p: Probability of desired successful result
q: Probability of failure

Arithmetic Mean, Variance and Moments of Binomial Distribution:

The probability mass function of a B(n,p) random variable is

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

for $x = 0, 1, \ldots, n$, with

$$E(X) = E(X_1) + \cdots + E(X_n) = p + \cdots + p = np$$

$$Var(X) = Var(X_1) + \cdots + Var(X_n)$$

$$= p(1-p) + \cdots + p(1-p) = np(1-p)$$

- Arithmetic mean: μx = E (X) = np
- Variance: σ^2= E [(X - μx)2] =npq= npq
- Moments:
  $\mu_1 = 0$
  $\mu_2 = \sigma_x^2$
  μ_3 = npq[q – p]
  μ_4 = npq[1-6pq+3npq]

- Asymmetry (Skewness) and Kurtosis Coefficients:

  ÇK = $\frac{q-p}{\sqrt{npq}}$ ve BK = 3 + $\frac{1-6pq}{npq}$

Example:

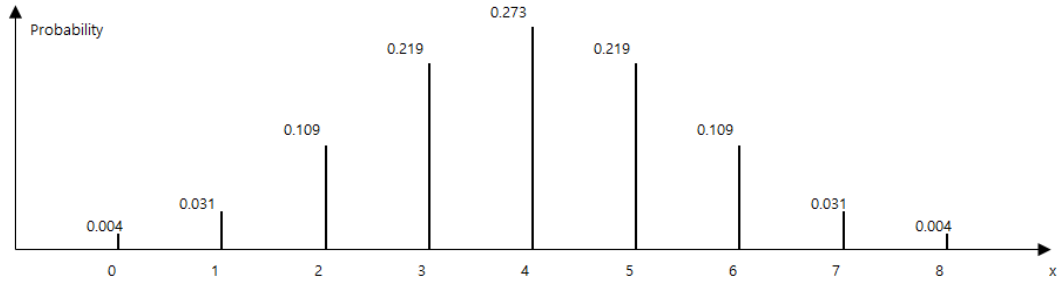
Total 8 trials, what is the probability of 3 trials being successful? The probability of each trial is calculated. The probability of trial intervals can also be calculated.

$$P(X = 3) = \binom{8}{3} \times 0.5^3 \times (1-0.5)^5 = \frac{8!}{3!5!} \times 0.5^8 = 0.219$$

$$P(X \leq 1) = P(X = 0) + P(X = 1)$$

$$= \binom{8}{0} \times 0.5^0 \times (1-0.5)^8 + \binom{8}{1} \times 0.5^1 \times (1-0.5)^7$$

$$= \frac{8!}{0!8!} \times 0.5^8 + \frac{8!}{1!7!} \times 0.5^8 = 0.004 + 0.031 = 0.035$$



x	0	1	2	3	4	5	6	7	8
$P(X \leq x)$	0.004	0.035	0.144	0.363	0.636	0.855	0.965	0.996	0.1000

Example:

Let a coin be tossed 10 times. Calculate the probability of getting heads 4 times

Since we are interested in random events where the binomial distribution is suitable and there are two situations as successful and unsuccessful, it can be defined as:

successful: getting heads (p=0.5)

unsuccessful: not getting heads (q=0.5)

Since n=10; X=4, the desired probability is:

$$P(X; n, p) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$P(4;10,0.5) = \frac{10!}{4!(10-4)!} 0.5^4 (1-0.5)^6 = 0.205$$

Example:

There are 16 missiles. 4 out of 16 missiles do not work. What is the probability of 12 attempts working?

Probability of not working, q=0.25, probability of working, p=0.75 has a binomial distribution.

Number of missiles expected to be launched, weight average or expected value:

$$E(X) = np = 16 \times 0.75 = 12$$

Variance:

$$Var(X) = np(1-p) = 16 \times 0.75 \times 0.25 = 3$$

The probability of exactly 12 missiles flying successfully,

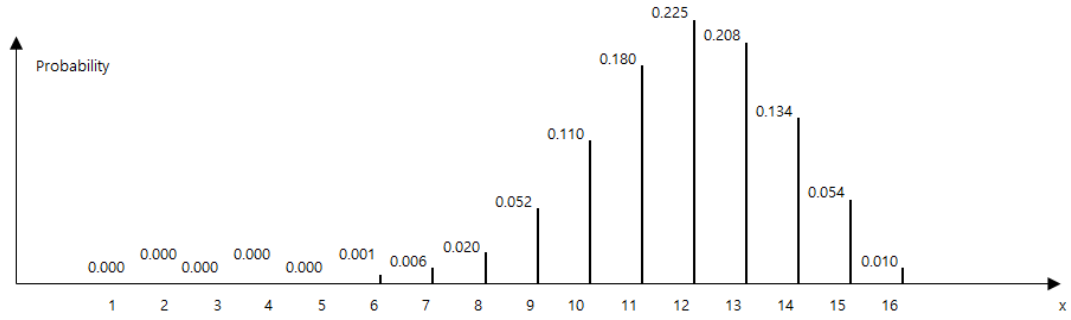
$$P(X = 12) = \binom{16}{12} \times 0.75^{12} \times 0.25^{4} = \frac{16!}{12!4!} \times 0.75^{12} \times 0.25^{4} = 0.225$$

The probability of at least 14 missiles flying successfully,

$$P(X \geq 14) = P(X = 14) + P(X = 15) + P(X = 16)$$

$$P(X \geq 14) = \binom{16}{14} \times 0.75^{14} \times 0.25^{2} + \binom{16}{15} \times 0.75^{15} \times 0.25^{1} + \binom{16}{16} \times 0.75^{16} \times 0.25^{0}$$

$$P(X \geq 14) = 0.134 + 0.054 + 0.010 = 0.198$$



x	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$P(X \leq x)$	0.000	0.000	0.000	0.000	0.000	0.001	0.007	0.027	0.079	0.189	0.369	0.594	0.802	0.936	0.990	0.100

Example:
Calculate the probability of getting a six exactly 12 times if a dice is rolled 20 times.
Successful: Getting a 6 (p=1/6)
Unsuccessful: Not getting a 6 (q=5/6)
Since n=20; X=12, the desired probability is

$$P(X; n, p) = \frac{n!}{x!(n-x)!}\ p^x\ q^{n-x}$$

$$P(12;20,1/6) = \frac{20!}{12!(20-12)!}\ (1/6)^{12}\ (5/6)^{8} = 0.0000135$$

Example:
5 out of 10 tablets in a box are mobile phones. What is the probability that 2 of the 3 tablets are mobile phones when they are taken from the box? Calculate the weighted mean and variance values. Comment.
Here, n=3, p=1/2, P(X=2).

$$P(X=2) = \binom{3}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{3-2} = \frac{3!}{2!\,(3-2)!}\left(\frac{1}{2}\right)^3 = 0.375$$

Example:
A 5000 page encyclopedia has 1500 pages of typographical errors. What is the probability that at most 4 of 5 randomly selected pages from this encyclopedia have typographical errors?
Asymmetric binomial distribution. Typographical error rate:
p= (1500 / 5000) =0.30 q=0.70

P(x ≤ 4) = 1 – P(x=5)
P(x=5) = (5! / (5!*0!))*(0.30)^5 *(0.70)^0 = 0.00243
P(x ≤ 4) = 1 – 0.00243 = 0.99757

Example:
It is known that 3 of 12 bulbs in a box are defective. When 3 bulbs are randomly drawn from this box and returned;
• What is the probability that 2 of them are defective?
• What is the expected average number of defectives and standard deviation at the end of this experiment?
The probability that 2 of them are defective is,
p= 3 / 12 = 0.25
q = 1-0.25=0.75
P(x=2) = (3! /(2!*1!))*(0.25)^2* (0.75)^1 = 0.1406

Example:

It is known that 5% of the parts produced by a machine are defective. 6 products produced by this machine were examined.

$$P(x,6,0.05) = \begin{cases} \binom{6}{x}(0.05)^x(0.95)^{6-x}, & x = 0,1,2,\ldots,6 \\ 0, & diğer\ x'leriçin \end{cases}$$

a) There is no possibility of any product being perfect.

$$P(x=0) = \binom{6}{0}(0.05)^0(0.95)^{6-0}$$

$$\binom{6}{0} = \frac{6!}{0!\,(6-0)!} = 1$$

$$P(x=0) = 1 * 1 * (0.95)^6 = 0.735$$

b) The possibility of a product being defective

$$P(x=1) = \binom{6}{1}(0.05)^1(0.095)^{6-1}$$

$$\binom{6}{1} = \frac{6!}{1!\,(6-1)!} = 6$$

$$P(x=1) = 6(0.05)^1(0.095)^5 = 0.22$$

a) En az iki ürünün kusurlu çıkma olasılığı

$$P(x \geq 2) = 1 - [P(x=0) + P(x=1)] = 1 - 0.735 - 0.22 = 0.045$$

6.3. Poisson Distribution

The Poisson probability distribution function is an expression of the arithmetic mean. The Poisson (pronounced “puason”) distribution is also known as the rare event distribution. In statistics and probability theory, it is a discrete probability distribution that expresses the probability of the number of occurrences in a certain fixed time unit interval. It is assumed that the average number of occurrences of an event in this time interval is known and that the time difference between any event and the event immediately following it occurs independently of the previous time differences.

The Poisson distribution emerges with the Poisson process. The Poisson process takes the form of some events that are intermittent in nature (i.e., occurring 0, 1, 2, 3 .. times) occurring with a fixed probability in a unit of time, area, space or volume. Examples of such events and the application of the Poisson distribution are as follows:

• Number of soldiers killed by horse and mule kicks each year in the Prussian cavalry: This classic example was published in a book by Ladislaus Josephovich Bortkiewicz in 1868 and was given to military and civilian high school students for years.
• Number of connections to a certain Internet site in an hour,
• Number of trucks arriving at a shipping depot for loading and unloading in half an hour,
• Number of cars passing through a certain traffic intersection in 1 minute,
• Number of planes landing at an airport every hour,
• Number of monthly traffic accidents occurring at a certain traffic point,
• Number of defects in a manufactured product,
• Number of abnormal cells in 1 cm3 of blood
• Number of defects in 10 m2 of fabric.

Let X be the random variable for the number of successes in an area or volume in a given time interval. X that satisfies the following conditions is called a Poisson random variable.

• An experiment consists of counting the number of times an event (successes) occurs in a given time, area or volume.

• The number of successes to be achieved in an area or volume in two discrete unit times is independent of each other.
• The probability of success in a unit time interval, area or volume is the same for all units.
• It is almost impossible for two or more successes to occur in a very small time interval, area or volume. That is, the probability of multiple successes in this case approaches zero.
• The average number of occurrences of an outcome in a unit time interval, area or volume is λ.

The random variable that the Poisson distribution generally focuses on is a countable event; this event occurs discretely in a fixed-length (usually time) interval, and the number of events observed in this interval is the random variable for the Poisson distribution. The expected value of the number of events occurring in this fixed interval (the average number of occurrences) is fixed as λ, and this average value is proportional to the interval length. If an average of 5 events occur in every 4-minute time interval, then an average of 10 (=8x5/4) events occur in a fixed 8-minute interval.

The interval between successive Poisson-type events is an exponential distribution, which is mutually related. If x is a random variable and all the values it can take are represented by a non-negative integer x (x = 0, 1, 2, 3...n), the probability of the event occurring is expressed as follows:

$$P(X = x, \lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

Here,

- *e, base of natural logarithm (e = 2.71828...);*
- *n, number of occurrences of the event whose probability is given by the function;*
- *x!, factorial for x*
- *λ is the arithmetic mean, the expected value of the number of occurrences in the given fixed interval; is a positive real number. This function of x is the probability mass function for the Poisson distribution.*

Arithmetic Mean, Variance and Moments of Poisson Distribution:

Arithmetic mean: $\mu_x = E(X) = n.p = \Lambda$

Variance: $\sigma_x^2 = E[(X - \mu_x)^2] = n.p.\ q = \Lambda$

Here, p: probability of occurrence, q: probability of non-occurrence

Moments:

$\mu_1 = 0$

$\mu_2 = \sigma_x^2 = \Lambda$

$\mu_3 = \Lambda$

$\mu_4 = \Lambda + 3\Lambda^2$

- Asymmetry (Skewness) and Kurtosis Coefficients:

$$ÇK = \frac{\mu_3}{\sigma^3} = \frac{1}{\sqrt{\Lambda}} \quad ve \quad BK = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1}{\Lambda}$$

When the conditions n≥ 100 and n.p ≤10 are met for the Poisson distribution, the Binomial distribution approaches and the Poisson distribution can be used instead of the Binomial. The Binomial distribution is used to find the average probability in the Poisson distribution.
λ=μ=n.p
The Poisson distribution gives quite accurate prediction results in problems where the number of experiments is very high and the probability of occurrence is very low. (p ≤ 0.01 and λ=n.p ≤ 5)

Example:

The arithmetic average of the incorrect grades entered in a year in a University with 10,000 students is 0.5.

1. Aritmatrik mean, $\lambda = 0.5$
2. n = 10000
3. p = $\frac{\lambda}{n} = \frac{0.5}{10000} = 0.00005$
4. The expected result of the random trial is two-sided, error-free, and can be repeated n times.

Under these conditions, the Poisson function of this event can be expressed as follows.

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots\ldots n$$

$$P(x) = \frac{e^{-0.5}(0.5)^x}{x!}, \quad x = 0, 1, 2, \ldots\ldots n$$

Bu Poisson dağılımının,

- Aritmetik ortalaması: $\mu_x = E(X) = n.p = \lambda$
  $\mu_x = 0.5$

- Varyansı: $\sigma_x^2 = E[(X - \mu_x)^2] = n.p.(1 - p) = \lambda$
  $\sigma_x^2 = 0.5$

- Momentleri:
  $\mu_1 = 0$
  $\mu_2 = \sigma_x^2 = \lambda = 0.5$
  $\mu_3 = \lambda = 0.5$
  $\mu_4 = \lambda + 3\lambda^2 = 1 + 3(0.5)^2 = 1.75$

- Asimetri (Çarpıklık) ve Basıklık Katsayıları:

$$ÇK = \frac{\mu_3}{\sigma^3} = \frac{1}{\sqrt{\lambda}} = \frac{1}{\sqrt{0.5}} = 1.414$$

$$BK = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1}{\lambda} = 3 + \frac{1}{0.5} = 5$$

Bir yılda gerçekleşebilecek hata sayısının (x) olasılıkları hesaplayabiliriz.

$$P(x=0) = \frac{e^{-0.5}(0.5)^x}{x!}$$

- Bir yıl boyunca hiç hata olmama olasılığı:

  P(x=0) = $\frac{e^{-0.5}\,(0.5)^0}{0!}$ = 0.6065

- Bir yıl boyunca bir adet hata olmama olasılığı:

  P(x=1) = $\frac{e^{-0.5}\,(0.5)^1}{1!}$ = 0.3033

- Bir yıl boyunca iki adet hata olmama olasılığı:

  P(x=2) = $\frac{0.6\,(0.5)^2}{2!}$ = 0.0758

- Bir yıl boyunca üç adet hata olmama olasılığı:

  P(x=3) = $\frac{0.6\,(0.5)^3}{3!}$ = 0.0126

Example:

A bank branch receives an average of 15 customers every half hour. Assuming that the arrival of customers follows a Poisson process, calculate the probability that at least 1 customer will arrive at the bank in 10 minutes.

If 15 customers arrive in 30 minutes, an average of 15/3 = 5 customers will be expected in 10 minutes.

λ = 5 is taken,

P(least 1) = 1 – P(x=0)

$P(x=0) = \frac{e^{-5}\,5^0}{0!}$ =0,00674

P(en az 1) = 1 – 0.00674 = 0.99326

Matlab:

```
clear all
close all
lambda = 5;
x = 0:15;
P = poisspdf(x,lambda)
```

P = 0.0067 0.0337 0.0842 0.1404 0.1755 0.1755 0.1462 0.1044 0.0653 0.0363 0.0181 0.0082 0.0034 0.0013 0.0005 0.0002

Probability of no customers coming, x = 0

Probability of one customer coming, x = 1

Probability of 15 customers coming, x = 15

Example:
An average of 4 failures occur in a power plant per month. What is the probability that no failures will occur in this plant within a month?

P(x) = $\frac{e^{-\lambda}\lambda^{x}}{x!}$, x = 0, 1, 2,……n
λ=4
P(x=0)= 0.01832

Matlab:

```
clear all
close all
lambda = 4;
x = 0
P = poisspdf(x,lambda)
```

P = 0.0183

Example:
On average, 5% of the parts on a production line are defective. What is the probability that 2 of the 22 randomly selected parts are defective?

If we calculate with binomial distribution:

P(x; n,p) = $\frac{n!}{x!(n-x)!}$ $p^x (1-p)^{n-x}$ = C_n^x $p^x (1-p)^{n-x}$
p=0.05,
n=22,
q=1-p=0.95,
x=2,
P(k=2)= 0.2070

If we calculate with Poisson distribution;
ʎ = np, the average number of occurrences of the expected result (Arithmetic mean),
np= λ = 22 * 0.05=1.1

P(x) = $\frac{e^{-\lambda}\lambda^{x}}{x!}$, x = 0, 1, 2,……n

P(x=2)=0.201

Here, e = 2.71828 is the base of the natural logarithm.

Example:
On average, 2% of those who contract the coronavirus die. What is the probability that 2 out of 10 randomly selected people who test positive for the coronavirus will die?

If we calculate with a binomial distribution:

$P(x; n,p) = \frac{n!}{x!(n-x)!}\ p^x\ (1-p)^{n-x} = C_n^x\ p^x\ (1-p)^{n-x}$

p=0.02,
n=10,
q=1-p=0.98,
x=2,

$P(X=2) = C_n^x\ p^x\ (1-p)^{n-x} = \frac{10!}{2!(10-2)!}\ 0.02^2\ (0.98)^{10-2} = 0.015$

Matlab:

```
clear all
close all
defects = 0:10;
P = binopdf(defects,10,.02)
```

P = 0.8171 0.1667 0.0153 0.0008 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

If we calculate with Poisson distribution;
ʎ = np, the average number of occurrences of the expected result (Arithmetic mean),
λ = np= 10 * 0.02=0.2

$P(x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-0.2}\ 0.2^x}{x!}$, x = 0, 1, 2,……n

$P(x=2) = \frac{e^{-0.2}\ 0.2^2}{2!} = 0.016$

Here, e = 2.71828 is the base of the natural logarithm.

Matlab:

```
clear all
close all
lambda = 0.2;
x = 0:10
P = poisspdf(x,lambda)
```

P = 0.8187 0.1637 0.0164 0.0011 0.0001 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

Example:
If 1% of the batteries a factory produces are defective, what is the probability of 3 defective batteries being produced in a production of 200 units?

n=200, p=0.01, np= λ =200*0.01= 2

$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$, x = 0, 1, 2,……n

P(x=3)=0.18045

Example:
An insurance company has insured 1000 people against traffic accidents. If the fatality rate in these accidents is 1%, what is the probability that the insurance company will pay money to 5 people?

p=0.01, λ =1000*0.01=10
P(x=5)=0.0375
$P(x=5) = \frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-10}10^{5}}{5!}$ =0.0375

Örnek:
Bir sınıftaki öğrencilerin %2'sinin boyları 190 cm'nin üzerindedir. Rasgele seçilen 100 öğrenciden; 6'sının, en az 2'sinin boylarının 190 cm'den fazla olması olasılıklarını bulunuz.
$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$, x = 0, 1, 2,……n

λ =2, (100 öğrenciden 2 si)
P(x=6)=0.1203
P(x=0)=0.13534
P(x=1)=0.27068
P(x ≥2) = 1- P(x =0) – P(x =1) = 1 – 0.13584 - 0.27068 = 0.59398

Örnek:

Bir şehirdeki 30 yaşın üzerindeki nüfusun %5'nin üniversite mezunu olduğu bilinmektedir. 30 yaş üzerindekilerden rasgele seçilecek 100 kişi arasında;

a) 5 üniversite mezunu bulunması,

b) Hiç üniversite mezunu bulunmaması olasılığını hesaplayınız.

λ = n*p=100*0.05=5 ,

a) P(X=5) = $\frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-5}\,5^{5}}{5!}$ =0.175

b) P(X=0) = $\frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-5}\,5^{0}}{0!}$ =0.0067

Örnek:

Bir makinenin kusurlu ürün üretim oranı % 0.01 dir. Her saat başında üretim hattından alınan 100 parçanın incelenmesi sonucu 2'den fazla bozuk ürün üretldiğinde üretim durdurulacaktır. Üretimin durdurulma olasılığı nedir?

$$P(x,\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

P(x>2) = 1-P(x<=2) = 1 –[P(x=0) + P(x=1) + P(x=2)]

$$P(x > 2) = 1 - [\frac{e^{-1}1^{0}}{0!} + \frac{e^{-1}1^{1}}{1!} + \frac{e^{-1}1^{2}}{2!}$$

P(x>2) =0.0803

Örnek:

Bir fabrikada çok nadir olan arızadan dolayı, bir hafta içinde arızalanan makine sayısı 4'dür. Belirli bir hafta için bu arızadan

a) Hiç bir makinenin arızalanmama olasılığı nedir?

$P(X = 0)$

b) En az iki makinenin arızalanma olasılığı nedir?

$$P(X \geq 2) = 1 - P(X < 2) = 1 - \big(P(X = 0) + P(X = 1)\big)$$

c) 3 makinenin arızalanma olasılığı nedir?

$P(X = 3)$

X, bir haftada arızalanan makine sayısı

$$f(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0,1,2, \ldots, \lambda = 4$$

a) $P(X = 0) = \frac{e^{-4}4^0}{0!} = 0.0183$

b) $P(X \geq 2) = 1 - P(X < 2) = 1 - \big(P(X = 0) + P(X = 1)\big) = 1 - \left(\frac{e^{-4}4^0}{0!} + \frac{e^{-4}4^1}{1!}\right) =$

$= 1 - (0.0183 + 0.0733) = 1 - 0.0916 = 0.9084$

c) $P(X = 3) = \frac{e^{-4}4^3}{3!} = 0.195$

Örnek:

Tehlikeli bir kavşağın güvenliği araştırılıyor. Geçmiş polis kayıtları bu kavşakta ayda ortalama beş kaza olduğunu göstermektedir. Kazaların sayısı bir Poisson dağılımına göre dağıtılır ve 0, 1, 2, 3 veya 4 kazadan herhangi bir ayda olma olasılığını hesaplayın.

Çözüm: Poisson formülünü kullanarak, kaza olmama olasılığını hesaplayabiliriz:

P(0) = $\frac{\lambda^n e^{-\lambda}}{x!} = \frac{5^0 e^{-5}}{0!} = \frac{(1)(0.0067)}{1}$ = 0.00674

Bir kaza olma olasılığı:

P(1) = $\frac{\lambda^n e^{-\lambda}}{x!} = \frac{5^1 e^{-5}}{1!} = \frac{(5)(0.0067)}{1}$ = 0.03370

İki kaza olma olasılığı:

P(2) = $\frac{\lambda^n e^{-\lambda}}{x!} = \frac{5^2 e^{-5}}{2!} = \frac{(25)(0.0067)}{2\times1}$ = 0.08425

Üç kaza olma olasılığı:

P(3) = $\frac{\lambda^n e^{-\lambda}}{x!} = \frac{5^3 e^{-5}}{3!} = \frac{(125)(0.0067)}{3\times2\times1}$ = 0.14042

Dört kaza olma olasılığı:

P(4) = $\frac{\lambda^n e^{-\lambda}}{x!} = \frac{5^4 e^{-5}}{4!} = \frac{(625)(0.0067)}{4\times3\times2\times1}$ = 0.17552

Herhangi bir ayda 0, 1 veya 2 kaza olasılığını bilmek istiyorsak, bu olasılıkları şu şekilde ekleyebiliriz:

P(0,1, or 2) = P(0) + P(1) + P(2) = 0.00674 + 0.03370 + 0.08425 = 0.12469

P(3 veya daha az kaza) = P(0, 1, 2, or 3) = P(0) + P(1) + P(2) + P(3)

= 0.00674 + 0.03370 + 0.08425 + 0.14042 = 0.26511

Üçten fazla olasılığını hesaplamak istiyorsak, 1- 0.26511=0.73489).

Önemli nokta: Poisson dağılımı, n, 20'den büyük ya da eşit olduğunda ve p, 0.05'e küçük ya da eşit olduğunda binom dağılımının iyi bir yaklaşımdır.

Örnek:

Bir firma, üretmekte olduğu ampullerin son aşamada kontrolünde her gün ortalama 10 ampulün bozuk olduğunu tahmin ediyor.

a. Bir kontrol gününde 3 ampulün bozuk çıkması ,

b.Bir kontrol gününde 2 veya 3 ampulün bozuk çıkması olasılıklarını hesaplayınız?

Bu soru Poisson dağılımı kullanılarak çözülebilir.

$$P(X=k)=\frac{e^{-\lambda}\lambda^{k}}{k!}$$

$$\text{a) } P(3)=\frac{e^{-10}10^{3}}{3!}=0.007567$$

$$\text{b) } P(3)+P(2)=0.007567+\frac{e^{-10}10^{2}}{2!}$$

$$=0.007567+0.00227$$

$$=0.009837$$

Örnek:

Günde 200.000 cıvatanın üretildiği bir işletmede uygunsuzların oranı 0.00003 olarak tespit edilmiştir.

a. Buna göre hiçbir cıvatanın arızalı çıkmaması olasılığını,

b. En az iki cıvatanın bozuk çıkma olasılığını belirleyiniz?

$$\lambda=(200000*0.00003)=6$$

$$\text{a) } P(0)=\frac{e^{-6}6^{0}}{0!}=0.002487$$

$$\text{b) } P(X\geq 2)=1-P(X\leq 1)=1-\left(\frac{e^{-6}6^{0}}{0!}+\frac{e^{-6}6^{1}}{1!}\right)$$

$$=0.9826$$

Örnek:

Bir cıvata üreticisi üretmekte olduğu cıvatalar için kusurlu oranını %5 olarak belirlemiştir. üretimden 60 birimlik bir örnek alındığında bu örnekte 2 tane kusurlu çıkma olasılığını belirleyiniz?

$$n = 60$$

$$p = 0.05$$

$$\lambda = n * p = 60 * 0.05 = 3$$

$$P(x = 2) = \frac{e^{-3} * 3^2}{2!} = 0.224$$

6.4. Hypergeometric Distribution

In probability theory and statistics, the hypergeometric distribution describes the distribution of the number of successes in a finite population of n consecutive objects without replacement, as a discrete probability distribution. Assumptions of the Hypergeometric Distribution: n trials can be repeated under similar conditions. Each trial has two possible outcomes. Non-refundable sampling is done from a finite population. Since sampling is non-refundable, the probability of success (p) varies from experiment to experiment.

Hipergeometrik Dağılımın Olasılık Fonksiyonu:

$P(X = x) = \frac{C_x^B C_{n-x}^{N-B}}{C_n^N}$= *(Başarılı olma)*(Başırısız olma)/ Tüm*

n : örnek hacmi
N : anakütle eleman sayısı
B : yığındaki başarı sayısı
x : örnekteki durum sayısı, x= 0,1, 2, 3, …..,n

- Geri koyulmadan alınabilmesi mümkün örnek sayısı C_n^N'dir.
- Başarılı durum sayısının *x* olması için C_x^B sayıda ihtimal bulunur;
- Geride kalanların başarısız olması için de C_{n-x}^{N-B} ihtimal mevcuttur.

Kombinasyon:

$$C(N,n) = C_n^N = \frac{N!}{(N-n)!\,n!}$$

Hipergeometrik Dağılımının Aritmetik Ortalaması, Varyansı

- Aritmetik ortalaması: μ_x = E (X) = np
- Varyansı: σ_x^2 = E [(X - μ_x)^2] = ($\frac{N-n}{N-1}$) npq

  Burada p = $\frac{B}{N}$ olur.

Örnek:
10 kişilik bir sınıfta 6 öğrenci dersten başarılı olurken 4 öğrenci başarısız olmuştur. İadesiz olarak 5 öğrenci seçilmiş. Seçilen 5 öğrenciden, başarılı ve başarısız öğrencileri sınıflandıran bir olasılık değişkenleri şu şekilde gösterilebilir:

Anakütle eleman sayısı, N=10
Yığındaki başarı sayısı, B=6
Örnek hacmi, n=5
Örnekteki başarı durum sayısı, x = 0,1, 2, 3, 4,5 }

$$P(X = x) = \frac{C_x^B C_{n-x}^{N-B}}{C_n^N}$$

- Geri koyulmadan alınabilmesi mümkün örnek sayısı C_n^N'dir.
  $C_n^N = \frac{10!}{5!\,(10-5)!} = 252$
- Başarılı nesne sayısının *x* olması için C_x^B sayıda ihtimal bulunur;
- Geride kalanların başarısız olması için de C_{n-x}^{N-B} ihtimal mevcuttur.

a) Seçilen 5 öğrencinin de başarılı olma olasılığını bulunuz.

$$P\,(X = 5) = \frac{C_5^6\, C_0^4}{C_5^{10}} = \frac{\frac{6!}{5!\,(6-5)!} * \frac{(4)!}{(0)!\,(4)!}}{\frac{10!}{5!\,(10-5)!}} = \frac{6 * 1}{252} = 0.0238$$

b) 3 öğrencinin başarılı olma olasılığını bulunuz.

$$P\,(X = 5) = \frac{C_3^6\, C_{5-3}^{10-6}}{C_5^{10}} = \frac{\frac{6!}{3!\,(6-3)!} * \frac{(4)!}{(2)!\,(4-2)!}}{\frac{10!}{5!\,(10-5)!}} = \frac{20 * 6}{252} = 0.476$$

c) En fazla 2 öğrencinin geçmiş olma olasılığını bulunuz.

$$P\,(x \le 2) = P\,(X = 0) + P\,(X = 1) + P\,(X = 2) = \frac{C_0^6\, C_{5-0}^{10-6}}{C_5^{10}} + \frac{C_1^6\, C_{5-1}^{10-6}}{C_5^{10}} + \frac{C_2^6\, C_{5-2}^{10-6}}{C_5^{10}}$$

$$P\,(x \le 2) = 0.262$$

d) En az 3 öğrencinin geçmiş olma olasılığını bulunuz.

$$P\,(x \ge 3) = 1 - [P\,(X = 0) + P\,(X = 1) + P\,(X = 2)] = 1 - 0.262 = 0.738$$

e) Sırasıyla seçilen 1, 2, 3, 4 ve 5'inci öğrencinin başarısız olma olasılıklarını bulunuz.
Bu örnekte bulunması istenen seçilen öğrencinin başarısız öğrenci olma olasılığıdır. Başarısız öğrenci sayısı 4 olduğundan B=4 olacaktır.

$$P\ (x = 0) = \frac{C_0^4\ C_{5-0}^{10-4}}{C_5^{10}} = \frac{1 * 6}{252} = 0.0238$$

$$P\ (x = 1) = \frac{C_1^4\ C_{5-1}^{10-4}}{C_5^{10}} = \frac{4 * 15}{252} = 0.238$$

$$P\ (x = 2) = \frac{C_2^4\ C_{5-2}^{10-4}}{C_5^{10}} = \frac{6 * 20}{252} = 0.476$$

$$P\ (x = 3) = \frac{C_3^4\ C_{5-3}^{10-4}}{C_5^{10}} = \frac{1 * 6}{252} = 0.238$$

$$P\ (x = 4) = \frac{4\ C_{5-4}^{10-4}}{C_5^{10}} = \frac{1 * 6}{252} = 0.0238$$

Burada dikkat edilirse öğrencilerin başarısızlığıyla ilgili bütün olasılıklar hesaplanmıştır. Bu olasılıkların toplamı 1'e eşittir.

$P\ (x = 0) + P\ (x = 1) + P\ (x = 2) + P\ (x = 3) + P\ (x = 4) = 1$

f) Bu dağılımın başarılı ve başarısız öğrenci sayısına göre Aritmetik ortalaması ve standart sapmasını bulunuz.

<u>Başarılı öğrenci sayısına göre:</u>

N = 10, B = 6 ve n = 5. Burada p = $\frac{B}{N} = \frac{6}{10}$ = 0.6 olacaktır.

- Aritmetik ortalaması: μ_x = E (X) = n.p = 5*0.6 = 3
- Varyansı: σ_x^2 = E [(X - μ_x)^2] = ($\frac{N-n}{N-1}$) n.p. (1 – p) = ($\frac{10-5}{10-1}$) 3 (1 – 0.6) = 0.666

<u>Başarısız öğrenci sayısına göre:</u>

N = 10, B = 4 ve n = 5. Burada p = $\frac{B}{N} = \frac{4}{10}$ = 0.4 olacaktır.

- Aritmetik ortalaması: μ_x = E (X) = n.p = 5*0.4 = 2
- Varyansı: σ_x^2 = E [(X - μ_x)^2] = ($\frac{N-n}{N-1}$) n.p. (1 – p) = ($\frac{10-5}{10-1}$) 2 (1 – 0.4) = 0.666

Örnek:

Hastaneye gelen 100 kişiden 20 tanesinin koronavirüsü testi pozitif çıkmaktadır. İadesiz rasgele seçilen 12 kişiden 5 kişinin koronavirüs testinin pozitif çıkma olasılığı nedir?

N=100, B=20, n=12, x=0,1,2,….12

$$P\ (x = 5) = \frac{C_x^B C_{n-x}^{N-B}}{C_n^N} = \frac{C_5^{20}\ C_7^{80}}{C_{12}^{100}} = 0.047$$

Hipergeometrik Dağılımının Aritmetik Ortalaması ve Varyansını hesaplayınız.

Burada p = $\frac{B}{N}$ olur, p=B/N=20/100=0.2

- Aritmetik ortalaması: μ_x = E (X) = n.p=12*0.2=2.4
- Varyansı: σ_x^2 = E [(X - μ_x)^2] = ($\frac{N-n}{N-1}$) npq =(88/99)*12*0.2*0.8=1.71

Örnek:

Bir makineden üretilen 100 ürün içinde 60 tanesi testten geçmiştir. İadesiz seçilen 8 üründen 5 tanesinin testten geçme olasılığı nedir?

N=100, B=60, n=8, x=5

$$P\ (X = x) = \frac{C_x^B C_{n-x}^{N-B}}{C_n^N} = \frac{C_x^{60}\ C_{8-x}^{100-60}}{C_8^{100}}$$ x=0,1,2,3, ..., 8

$$P\ (X = 5) = \frac{C_5^{60}\ C_3^{40}}{C_8^{100}} = 0.29$$

6.5. Exponential Distribution Functions

- Üstel dağılım fonksiyonu:

$$f(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}} \qquad x > 0$$

- Beklenen değer:

$$E(X) = \int_0^\infty x \frac{1}{\mu} e^{-\frac{x}{\mu}} dx = \frac{1}{\mu} \int_0^\infty x e^{-\frac{x}{\mu}} dx \qquad u = x \qquad du = dx \qquad dv = e^{-\frac{x}{\mu}} dx \qquad v = -\mu e^{-\frac{x}{\mu}}$$

$$\int udv = uv - \int vdu \quad \text{kismi integrasyo n islemi ile}$$

$$E(X) = \frac{1}{\mu}\left[-x\mu e^{-\frac{x}{\mu}} \Big|_0^\infty + \int_0^\infty \mu e^{-\frac{x}{\mu}} dx \right] = x e^{-\frac{x}{\mu}} \Big|_0^\infty - \mu e^{-\frac{x}{\mu}} \Big|_0^\infty$$

$E(X) = \mu$ elde edilir .

$Var(X) = \mu^2$ olur.

Örnek:

Bir işletmenin üretilmiş olduğu elektronik cihazların arızasız çalışma sürelerinin (saat cinsinden) üstel dağılıma uyduğu görülmüştür ve ortalama arızasız çalışma süresinin 24 saat olduğu hesaplanmıştır. Buna göre

- a)Rastgele seçilen bir cihazın en az 12 saat arızasız çalışma olasılığını hesaplayınız
- b) En fazla 36 saat arızasız çalışması olasılığını bulunuz ?
- c) Seçilen cihazın 30 saatten fazla çalışma olasılığı %80 olabilmesi için bu cihazların ortalama arızasız çalışma süresi ne olmalıdır?

$$f(x) = \frac{1}{24} e^{-\frac{x}{24}} \qquad x > 0$$

a) $$P(X \geq 12) = \int_{12}^\infty \frac{1}{24} e^{-\frac{x}{24}} dx = -e^{-\frac{x}{24}} \Big|_{12}^\infty = e^{-\frac{12}{24}} = e^{-0,5} = 0,6065$$

b) $$P(0 < x < 36) = \int_0^{36} \frac{1}{24} e^{-\frac{x}{24}} dx = -e^{-\frac{x}{24}} \Big|_0^{36} = 1 - e^{-1,5} = 1 - 0,2231 = 0,7769$$

c) $$\int_{30}^\infty \frac{1}{\lambda} e^{-(\frac{x}{\lambda})} dx = \frac{1}{\lambda}(-.\lambda.) e^{-(\frac{x}{\lambda})} \Big|_{30}^\infty = 0,8 \Rightarrow -e^{-(\frac{x}{\lambda})} \Big|_{30}^\infty = 0,8$$

$$-e^{-\frac{\infty}{\lambda}} + e^{-\frac{30}{\lambda}} = 0,8 \qquad e^{-\frac{30}{\lambda}} = 0,8 \Rightarrow -\frac{30}{\lambda} = \ln 0,8 \Rightarrow \lambda = 134 \; saat$$

Örnek:

Let $X \sim \text{uniform}(0,1)$. Find $E(X)$.

X has range $[0,1]$ and density $f(x) = 1$. Therefore,

$$E(X) = \int_0^1 x\,dx = \frac{x^2}{2}\Big|_0^1 = \boxed{\frac{1}{2}}.$$

Not surprisingly the mean is at the midpoint of the range.

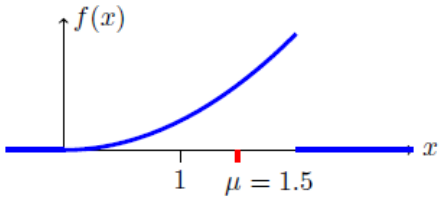
Örnek:

Let X have range $[0,2]$ and density $\frac{3}{8}x^2$. Find $E(X)$.

$$E(X) = \int_0^2 xf(x)\,dx = \int_0^2 \frac{3}{8}x^3\,dx = \frac{3x^4}{32}\Big|_0^2 = \boxed{\frac{3}{2}}.$$

Does it make sense that this X has mean is in the right half of its range?

Yes. Since the probability density increases as x increases over the range, the average value of x should be in the right half of the range.



Example:

Let X be the random variable with probability density function

$$f(x) = \begin{cases} e^x \text{ if } x \le 0 \\ 0 \text{ if } x > 0 \end{cases}.$$

Compute $E(X)$ and $\text{Var}(X)$.

Integrating by parts with $u = x$ and $dv = e^x dx$, we see that $\int xe^x dx = xe^x - e^x + C$. Thus,

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{0} xe^x dx \\ &= \lim_{r\to-\infty}^{0} \int_r^0 xe^x dx = \lim_{r\to-\infty} [-1 - re^r + e^r] \\ &= 1 \end{aligned}$$

[We used L'Hôpital's rule to see that $\lim_{r\to-\infty} re^r = \lim_{r\to-\infty} \frac{r}{e^{-r}} = \lim_{r\to-\infty} \frac{1}{-e^{-r}} = 0$.]

We compute

$$\begin{aligned} \int x^2 e^x dx &= x^2 e^x - 2\int xe^x dx \\ &= x^2 e^x - 2xe^x + 2e^x + C \end{aligned}$$

So,

$$\begin{aligned} \int_{-\infty}^{\infty} x^2 f(x)dx &= \int_{-\infty}^{0} x^2 e^x dx \\ &= \lim_{r\to-\infty} (2 - r^2 e^r + 2re^r - 2e^r) \\ &= 2 \end{aligned}$$

This gives $\text{Var}(X) = 2 - 1^2 = 1$.

Örnek:

Probability Density Function, pdf: f(x)

Let $X \sim \exp(\lambda)$. Find $E(X)$.

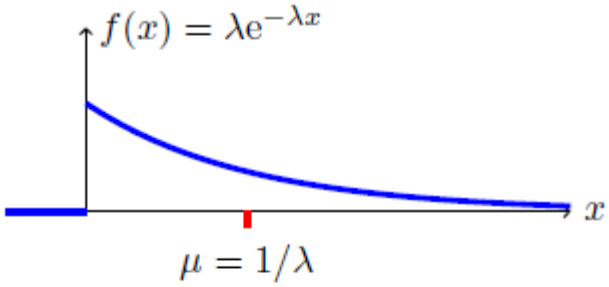
The range of X is $[0, \infty)$ and its pdf is $f(x) = \lambda e^{-\lambda x}$. Therefore

$$E(X) = \int_0^\infty x f(x)\, dx = \int_0^\infty \lambda x e^{-\lambda x}\, dx$$

(using integration by parts with $u = x$, $v' = \lambda e^{-\lambda x} \Rightarrow u' = 1$, $v = -e^{-\lambda x}$)

$$= -x e^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x}\, dx$$
$$= 0 - \frac{e^{-\lambda x}}{\lambda}\Big|_0^\infty = \frac{1}{\lambda}.$$

We used the fact that $x e^{-\lambda x}$ and $e^{-\lambda x}$ go to 0 as $x \to \infty$.



Mean of an exponential random variable

Örnek:

Let $X \sim \exp(\lambda)$. Find $E(X^2)$.

$$E(X^2) = \int_0^\infty x^2 f(x)\,dx = \int_0^\infty \lambda x^2 \mathrm{e}^{-\lambda x}\,dx$$

(using integration by parts with $u = x^2$, $v' = \lambda \mathrm{e}^{-\lambda x} \Rightarrow u' = 2x$, $v = -\mathrm{e}^{-\lambda x}$)

$$= -x^2 \mathrm{e}^{-\lambda x}\Big|_0^\infty + \int_0^\infty 2x\mathrm{e}^{-\lambda x}\,dx$$

(the first term is 0, for the second term use integration by parts: $u = 2x$, $v' = \mathrm{e}^{-\lambda x} \Rightarrow u' = 2$, $v = -\frac{\mathrm{e}^{-\lambda x}}{\lambda}$)

$$= -2x\frac{\mathrm{e}^{-\lambda x}}{\lambda}\Big|_0^\infty + \int_0^\infty \frac{\mathrm{e}^{-\lambda x}}{\lambda}\,dx$$
$$= 0 - 2\frac{\mathrm{e}^{-\lambda x}}{\lambda^2}\Big|_0^\infty = \frac{2}{\lambda^2}.$$

Örnek:

Let $X \sim \exp(\lambda)$. Find $\mathrm{Var}(X)$ and σ_X.

$$E(X) = \int_0^\infty x\lambda \mathrm{e}^{-\lambda x}\,dx = \frac{1}{\lambda} \quad \text{and} \quad E(X^2) = \int_0^\infty x^2 \lambda \mathrm{e}^{-\lambda x}\,dx = \frac{2}{\lambda^2}.$$

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \quad \text{and} \quad \sigma_X = \frac{1}{\lambda}.$$

We could have skipped Property 3 and computed this directly from $\mathrm{Var}(X) = \int_0^\infty (x - 1/\lambda)^2 \lambda \mathrm{e}^{-\lambda x}\,dx$.

Example:

Suppose that the random variable X has a cumulative distribution function

$$F(x) = \begin{cases} \sin(x) \text{ if } 0 \le x \le \frac{\pi}{2} \\ 0 \text{ if } x < 0 \text{ or } x > \frac{\pi}{2} \end{cases}$$

Compute $E(X)$ and $\mathrm{Var}(X)$.

First, we must find the probability density function of X. Differentiating we find that the function

$$f(x) = \begin{cases} \cos(x) \text{ if } 0 \leq x \leq \frac{\pi}{2} \\ 0 \text{ otherwise} \end{cases}$$

is the derivative of F at all but two points. Thus, $f(x)$ is a probability density function for X.

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_{0}^{\frac{\pi}{2}} x \cos(x) dx \\ &= (x \sin(x) + \cos(x))|_{x=0}^{x=\frac{\pi}{2}} \\ &= \frac{\pi}{2} - 1 \end{aligned}$$

Integrating by parts, we compute

$$\begin{aligned} \text{Var}(X) &= \int_{0}^{\frac{\pi}{2}} x^2 \cos(x) dx - E(X)^2 \\ &= (x^2 \sin(x) - 2\sin(x) + 2x\cos(x))|_{x=0}^{x=\frac{\pi}{2}} - (\frac{\pi}{2} - 1)^2 \\ &= \frac{\pi^2}{4} - 2 - (\frac{\pi^2}{4} - \pi + 1) \\ &= \pi - 3 \end{aligned}$$

6.6. Expected Value and Variance in Uniform Distribution Functions

Suppose a random number generator is programmed to produce a real number between 0 and 1, with each number in this range being equally likely. This is an example of a continuous uniform distribution.

If the probability of each value occurring in a certain interval of a continuous random variable X is equal, then the distribution of this random variable is called a uniform distribution.

Types of Uniform Distribution

Types of uniform distribution are:

1. **Continuous Uniform Distribution:** A continuous uniform probability distribution is a distribution that has an infinite number of values defined in a specified range. It has a rectangular-shaped graph so-called rectangular distribution. It works on the values which are continuous in nature. Example: Random number generator

2. **Discrete Uniform Distribution:** A discrete uniform probability distribution is a distribution that has a finite number of values defined in a specified range. Its graph contains various vertical lines for each finite value. It works on values that are discrete in nature. Example: A dice is rolled.

Continuous Uniform Distribution:

Uniform probability density Function (pdf): $F(x) = \frac{1}{b-a}$, $a \leq x \leq b$

Moments

The mean (first raw moment) of the continuous uniform distribution is:

$$E(X) = \int_a^b x \frac{dx}{b-a} = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}.$$

The second raw moment of this distribution is:

$$E(X^2) = \int_a^b x^2 \frac{dx}{b-a} = \frac{b^3 - a^3}{3(b-a)}.$$

In general, the n-th raw moment of this distribution is:

$$E(X^n) = \int_a^b x^n \frac{dx}{b-a} = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}.$$

The variance (second central moment) of this distribution is:

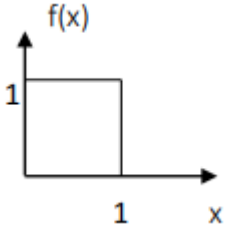
$$V(X) = E\left(\left(X - E(X)\right)^2\right) = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{dx}{b-a} = \frac{(b-a)^2}{12}.$$

Example:

For a random variable x with a uniform distribution between 0 and 1

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & d.d \end{cases}$$

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{1} f(x)dx = \int_{0}^{1} 1\,dx = 1$$

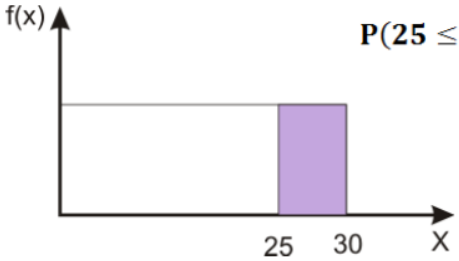


Örnek:

Süper marketteki kasaya 30 dakikalık periyotta bir müşteri gelmiştir. Bu müşterinin son 5 dakikada gelmiş olma ihtimalini hesaplayınız.

Olasılık yoğunluk fonksiyonu:

$$\mathbf{f(x)} = \begin{cases} \dfrac{\mathbf{1}}{\mathbf{b-a}} & \mathbf{a \le x \le b} \\ \mathbf{0} & \text{diğer} \end{cases} \quad \Rightarrow \quad \mathbf{f(x)} = \begin{cases} \dfrac{\mathbf{1}}{\mathbf{30-0}} & \mathbf{0 \le x \le 30} \\ \mathbf{0} & \text{diğer} \end{cases}$$



$$\mathbf{P(25 \le X \le 30)} = \int_{25}^{30} \mathbf{f(x)dx} = \int_{25}^{30} \frac{\mathbf{1}}{\mathbf{30}}\mathbf{dx} = \frac{\mathbf{30-25}}{\mathbf{30}} = \mathbf{1/6}$$

Bu örnek MATLAB komutu yardımıyla kolaylıkla hesaplanabilir: prob=unifcdf(5,0,30)

Continuous uniform cumulative distribution function: unicdf

Example:

A tea lover enjoys Tie Guan Yin loose leaf tea and drinks it frequently. To save money, when the supply gets to 50 grams he will purchase this popular Chinese tea in a 1000 gram package.

The amount of tea currently in stock follows a uniform random variable.

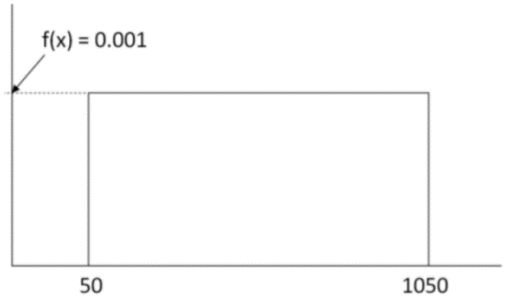
Solution

XX = the amount of tea currently in stock

aa = minimum = 50 grams

bb = maximum = 1050 grams

f(x)=1/(1050 - 50)=0.001



The expected value, population variance and standard deviation are calculated using the formulas:

$$\mu = \frac{a+b}{2} \qquad \sigma^2 = \frac{(b-a)^2}{12} \qquad \sigma = \sqrt{\frac{(b-a)^2}{12}}$$

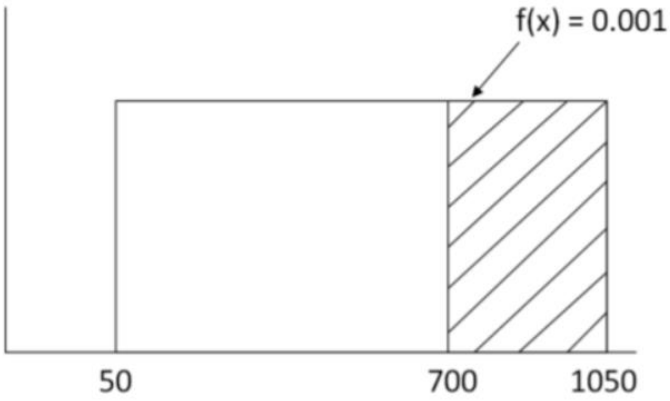
For the loose leaf tea problem:

$$\mu = \frac{50+1050}{2} = 550\text{g}$$

$$\sigma^2 = \frac{(1050-50)^2}{12} = 83,333$$

$$\sigma = \sqrt{83333} = 289\text{g}$$

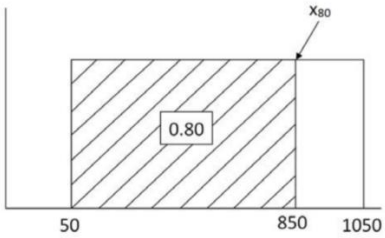
Probability problems can be easily solved by finding the area of rectangles.
Find the probability that there are at least 700 grams of TYin tea in stock.
P(X≥700)= width × height =(1050−700)(0.001)=0.35



The p^{th} percentile of the Uniform Distribution is calculated by using linear interpolation: $x_p = a + p(b-a)$

Find the 80^{th} percentile of Tie Guan Yin in stock:

$x_{80} = 50 + 0.80(1050-50) = 850$ grams



Example:

The *uniform distribution* on the interval $[0, 1]$ has the probability density function

$$f(x) = \begin{cases} 0 \text{ if } x < 0 \text{ or } x > 1 \\ 1 \text{ if } 0 \leq x \leq 1 \end{cases}$$

Letting X be the associated random variable, compute $E(X)$ and $\mathrm{Var}(X)$.

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{0} x \times 0dx + \int_{0}^{1} x \times 1dx + \int_{1}^{\infty} x \times 0dx$$

$$= 0 + \frac{1}{2}x^2|_0^1 + 0 = \frac{1}{2}$$

$$\int_{-\infty}^{\infty} x^2 f(x)dx = \int_{0}^{1} x^2 dx = \frac{1}{3}x^3|_{x=0}^{x=1} = \frac{1}{3}$$

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - E(X)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

Discrete Uniform Distribution: ü

Expected Value

The expected value of a uniform distribution is:

$$\mathrm{E(X)} = \sum_{i=1}^{n} x_i f(x_i) = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_n}{2}$$

In our example, the expected value is $\frac{1+2+3+4+5+6}{6} = \frac{1+6}{2} = 3.5$.

Variance

The variance of a uniform distribution is:

$$\mathrm{Var(X)} = \frac{(b-a+1)^2 - 1}{12}$$

7. Probability Distribution in Continuous Random Variables

7.1. Normal Distribution

The array consists of one-dimensional numerical values. It is called a vector; its behaviors such as direction and speed are discussed. With IoT and the internet of things, everything including humans is turning into smart objects that are sources of information. When a person is sick, we look at it as a model, it is seen as a source of information and communication is established. Data is collected by communication.

In the distribution of observation values, it is desired to gather around the average value and the number of observations to decrease as it moves away from the average or to be in this way. It is desired not to be flat, skewed to the left or right. Variance measures how much each value deviates from the average. Whether the average value can represent the array is decided with variance. The probability of the values forming the array will be calculated. The array elements are ranked from smallest to largest. What is the probability of being more or less than any array element?

For example, when an intelligence test is applied to a large population, it is observed that the observation values are gathered around the average intelligence test score and that the number gradually decreases as the average value is moved away from the average value, and the number of those with the lowest and highest intelligence scores decreases very much. This situation is evaluated as a normal, expected or normal situation.

The normal distribution is a continuous probability distribution. As it is known, the word normal is a word used in the meanings of ordinary, common, and commonly seen, etc. The origin of the normal distribution being called “normal” is that it is a commonly seen distribution.

In all probability distributions, two important measures are used to characterize the distribution. These measures are the expected value of the distribution (arithmetic mean), in other words, the mean of the distribution, and the standard deviation or variance of the distribution.

The normal distribution is also a continuous distribution whose mean is expressed by E(x)=μ and its standard deviation is expressed by or its variance by 2. The shape of the normal

distribution is symmetrical and is also known as the bell curve because its shape resembles a bell (Gaussian curve).

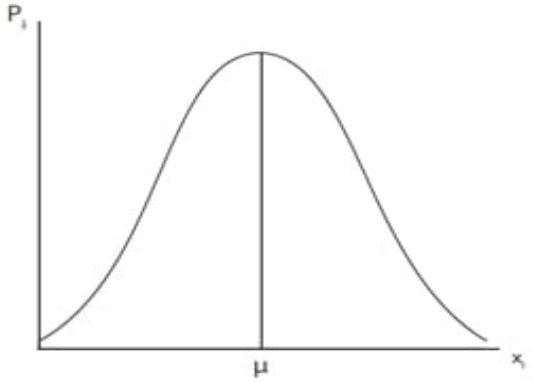


Figure: Normal Distribution

The sum of the probabilities of all variables in the series is equal to 1. The area covered by the curve is equal to 1.

Since there is a continuous random variable in the normal distribution and the state of continuity is defined as the ability of the variable to take infinite values between two values, the values that the random variable can take are very close to each other. This situation transforms the shape of the distribution into a curve. Because the values that the random variable can take are so close to each other that the points on the X axis showing the values that the random variable will take are very close to each other and therefore the shape of the normal distribution is in the form of a bell curve.

As can be seen from the figure above, the normal distribution is a distribution with maximum probability on the mean value (μ) of the distribution, is symmetrical and the right and left tails of the distribution are asymptotes to the X axis, that is, it is assumed that the tails of the distribution approach the X axis at infinity.

In the normal distribution, the probability calculation is done by calculating the area between two values and the total area under the curve is equal to 1, which is the total probability and expresses the full system value. The mean, which is the middle point of the distribution, or μ, divides the area under the curve into two equal parts, and the areas to the right and left of the mean represent a probability of 0.5. In other words, the probability of a randomly selected unit having a value below the mean is 0.5, and similarly, the probability of having a value above the mean is also 0.5.

The probability density function of the normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{1}{2}\frac{(X-\mu)^2}{\sigma^2}}$$

is expressed by the formula. Also, in the normal distribution above, X is the weighted mean and σ is the standard deviation of the random variable. It is seen that the normal probability density function above is a function defined between (-∞,+∞).

formülüyle ifade edilmektedir. Ayrıca, yukarıdaki normal dağılımda, X tesadüfi değişkenin μ ağırlık ortalaması ve σ standart sapmasıdır. Yukarıdaki normal olasılık yoğunluk fonksiyonunun (-∞, +∞) arasında tanımlı bir fonksiyon olduğu görülür.

The expressions in the formula of the normal distribution:
- X, a continuous random variable,
- μ, the mean of the random variables,
- σ, the standard deviation of the random variables,
- π, the constant pi in mathematics, 3.14159265…
- e, the base of the natural logarithm, 2.71828…

Normal dağılımın formülünde yer alan ifadeler:
- X, sürekli bir tesadüfi değişkeni,
- μ, tesadüfi değişkenlerin ortalamasını,
- σ, tesadüfi değişkenlerin standart sapmasını,
- π, matematikte sabit pi sayısını, 3.14159265…
- e, doğal logaritma tabanını, 2,71828… ifade etmektedir.

As can be seen from the formula below, there are two measures that characterize the normal distribution function; mean and standard deviation. Therefore, since all other values in the formula are constant, the determinant or qualifier of the function is the mean and standard deviation expressions. For this reason, there are infinitely many normal distributions with different mean and standard deviation. When the distribution function is integrated in the range of definition in question,

Aşağıdaki formülden de görüleceği gibi, normal dağılım fonksiyonunu niteleyen iki ölçü bulunmaktadır; ortalama ve standart sapma. Dolayısıyla, formülde yer alan diğer tüm değerlerin sabit olması sebebiyle, fonksiyonun belirleyicisi ya da niteleyicisi ortalama ve standart sapma ifadeleri olmaktadır. Bu sebeple de ortalaması ve standart sapması farklı sonsuz sayıda normal dağılımdan söz edilebilmektedir. Dağılım fonksiyonunun söz konusu tanım aralığında integrali alındığında,

$$\int_{-\infty}^{+\infty} f(X) = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}((X_i-\mu)/\sigma)^2} = 1$$

equality is realized.
eşitliği gerçekleşir.

On the other hand, since the probability calculation is done with the area calculation, the probability between two values such as a and b, provided that a < b, that is, the area calculation between a and b, is also,
Öte yandan, olasılık hesabının, alan hesabı ile yapılması sebebiyle a < b olmak koşuluyla a ve b gibi iki değer arasındaki olasılık, yani a ile b arasındaki alan hesabının da,

$$\int_{a}^{b} f(X) = \int_{a}^{b} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}((X_i-\mu)/\sigma)^2}$$

It will need to be done by calculating the integral of the form.
biçimindeki integralin hesaplanmasıyla yapılması gerekecektir.

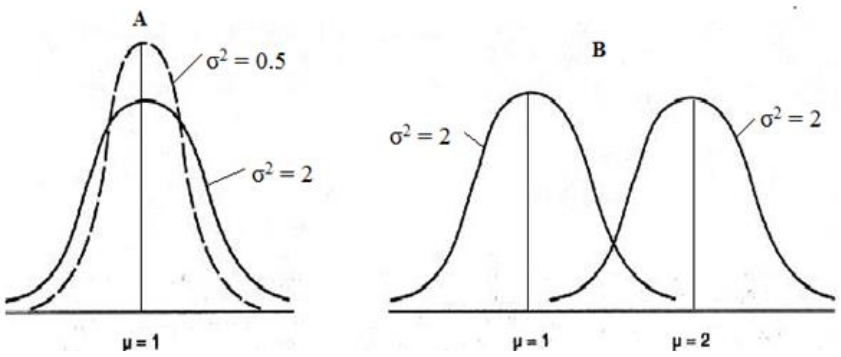
Properties of Normal Distribution:
Normal Dağılımın Özellikleri:

If the random variable X is assumed to follow a normal distribution with mean μ and variance σ^2, it will have the following properties:
X rassal değişkeninin ortalaması μ ve varyansı σ^2 olan normal dağılıma uyduğu düşünülürse aşağıdaki özelliklere sahip olacaktır:

1. Mean of random variable, X, E (X) = μ
2. Variance of random variable X, Var (X) = E [(X- μ)^2] = σ^2
3. Skewness = 0 (Coefficient of Skewness: Çarpıklık Katsayısı)
4. Kurtosis = 0 (Kurtosis coefficient:Basıklık katsayısı)

The mean and variance of the distribution of the random variable X determine the shape of the probability function of this variable. As the two parameters of the distribution, the mean and variance, change, the shape of the normal distribution graph will change, as seen in the examples below.
X rassal değişkeninin dağılımın ortalaması ve varyansı bu değişkenin olasılık fonksiyonunun

şeklini belirler. Dağılımın iki parametresi olan ortalama ve varyansı değiştikçe normal dağılım grafiğinin şeklide aşağıdaki örneklerde görüldüğü üzere değişecektir.



In Figure A, two probability density functions with the same mean and different variances are given. As can be seen from the figure, as the variance decreases, the probability density function becomes sharper (the kurtosis decreases). In Figure B, two probability density functions with the same variance and different means are given. As can be seen from the figure, as the mean increases, the density function shifts to the side, but there is no change in its shape.

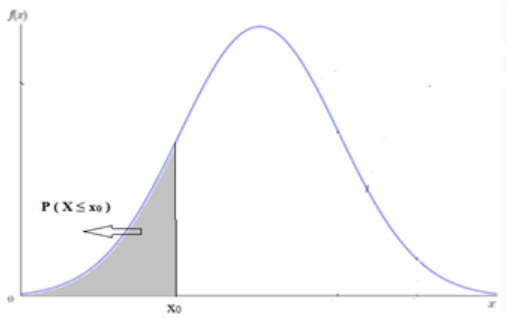
Şekil A'da normal dağılıma sahip ortalamaları aynı varyansları farklı iki tane olasılık yoğunluk fonksiyonu verilmektedir. **Şekilden de görüldüğü üzere varyans azaldıkça olasılık yoğunluk fonksiyonu sivrileşmektedir (Basıklığı azalmaktadır).** Şekil B'de normal dağılıma sahip varyansları aynı ortalamaları farklı iki tane olasılık yoğunluk fonksiyonu verilmektedir. Şekilden de görüldüğü üzere ortalama artıkça yoğunluk fonksiyonu yana kayarken biçiminde herhangi bir değişme söz konusu değildir.

Cumulative Probability Function of Random Variables:

Rassal Değişkenlerin Kümülatif (Birikimli) Olasılık Fonksiyonu:

The cumulative probability function of a random variable X with a normal distribution of mean μ and variance σ2 is given by F (x) = P (X ≤ x0). It shows the probability that the random variable X is small at a certain value of x0.

Ortalaması μ ve varyansı σ^2 olan normal dağılıma sahip bir X rassal değişkeninin Kümülatif olasılık fonksiyonu F **(x) = P (X ≤ x_0)** olarak gösterilir. X rassal değişkeninin belli bir x_0 değerinde küçük olma olasılığını gösterir.



In the figure above, the gray shaded area gives the probability that the normally distributed random variable X is smaller than any x0 value. This area can be found using the general formula below.

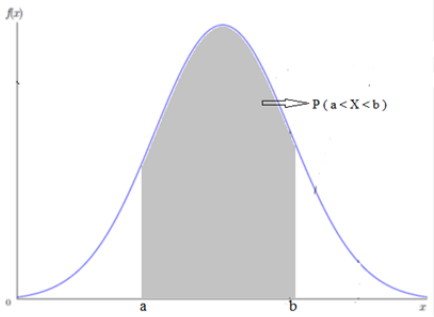
Yukarıdaki şekilde gri taralı alan, normal dağılıma sahip X rassal değişkeninin herhangi bir x_0 değerinden küçük olma olasılığını vermektedir. Bu alan aşağıdaki genel formül yardımıyla bulunabilir.

$$F(x) = P\ (X \leq\ x_0) = \int_{-\infty}^{x_0} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ \ dx$$

Interval Probability Function of Continuous Random Variables:
Sürekli Rassal Değişkenlerin Aralıklı Olasılık Fonksiyonu:

The interval probability function of a random variable X with a normal distribution with mean μ and variance σ^2 is denoted as F(a) – F(b) = P(a < X < b). It shows the probability that the random variable X is between two values a and b (provided that a < b).

Ortalaması μ ve varyansı σ^2 olan normal dağılıma sahip bir X rassal değişkeninin aralıklı olasılık fonksiyonu F (a) – F (b) = P (a < X < b) olarak gösterilir. X rassal değişkeninin a ve b gibi iki değer arasında olma olasılığını gösterir (a < b koşuluyla).



In the figure above, the gray shaded area gives the probability that the normally distributed random variable X is between two values, a and b. This area can be found using the general formula below.

Yukarıdaki şekilde gri taralı alan, normal dağılıma sahip X rassal değişkeninin a ve b gibi iki değer arasında olma olasılığını vermektedir. Bu alan aşağıdaki genel formül yardımıyla bulunabilir.

$$f(x) = P\ (a < X < b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad dx$$

Example:

Write a Matlab program that will plot the probability density function of a random variable with a normal distribution.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}x^2}$$

Solution -1:

```
clc; clear all; close all
x=-4:0.01:4;
s=1/sqrt(2*pi);
y=s*exp(-0.5*x.^2);
figure,plot(x,y)
```

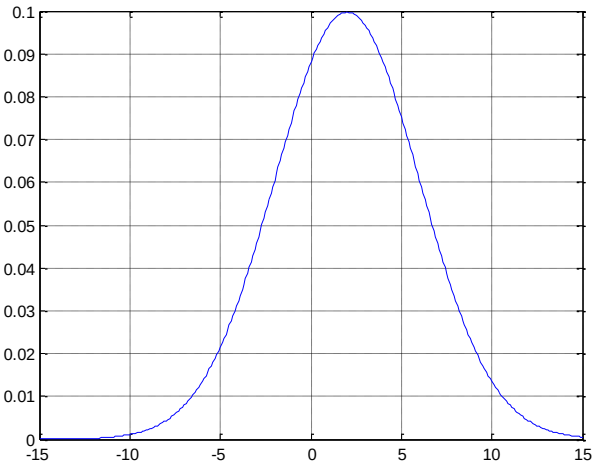
Solution -2:

```
x=-4:.1:4;
ort=0;
std=1;
figure, plot(x,normpdf(x,ort,std))
```

Example:

The graph of the probability density function for am=2 and $\sigma^2 = 16$.

```
clc; clear all; close all
x=-15:.1:15;
plot(x,normpdf(x,2,4))
grid on
```



7.2. Standard normal distribution

Set A consists of probabilities of random variables. It is also defined as a system. Dataset is given as array, vector (Intensity, direction), matrix.

A Kümesi rassal değişkenlerin olasılıklarından oluşmaktadır. Sistem olarak da tanımlanır. Veri yığını dizi, vektör (Şiddet, yön), matris olarak verilmektedir.

Laws of Probability:

1. The sum of the probabilities of each event (random variable) forming a set is equal to 1.
2. The probability of any event occurring in a set is 1≥p(A)≥0
3. If the probability of any event occurring in a set is 1, the probability of the others occurring is zero.
4. The sum of the probability of any event occurring and the probability of not occurring in a set is equal to 1, p(A') + p(A)=1.
5. The probability of any event not occurring in a set is p(A')= 1- p(A).

1. Bir kümeyi oluşturan olayların olma olasılıkları toplama 1'e eşittir.
2. Bir kümedeki herhangi bir olayın olma olasılığı, $1 \geq p(A) \geq 0$
3. Eğer kümedeki herhangi bir olayın olma olasılığı 1 ise diğerlerinin olma olasılığı sıfırdır.
4. Bir kümedeki herhangi bir olayın olma olasılığı ile olmama olasılığı toplamı 1'e eşittir, p(A') + p(A)=1.
5. Bir kümedeki herhangi bir olayın olmama olasılığı p(A')= 1- p(A).

Arithmetic Mean (Expected Value):

$$\mu = \frac{\sum_{i=1}^{n} O_i}{n} = \frac{O_1 + O_2 + \cdots + O_n}{n}$$

$$\mu = \sum_{i=1}^{n} \left(x_1 \frac{O1}{n} + x_2 \frac{O2}{n} + \cdots + x_n \frac{On}{n} \right)$$

$$\mu = E(\mathrm{X}) = \sum_{i=1}^{n} x_i p(x_i)$$

$$E(X^2) = \sum_{i=1}^{n} x_i^2 f(x_i)$$

Varyans, $V(X) = E(X^2) - E(X)^2$

xi: Random variables

Median(Ortanca):

When calculating the median in unclassified series, the data are first sorted from smallest to largest. If n is even, the average of n/2 values and n/2+1 values is the median, if n is odd, the median is (n+1)/2.

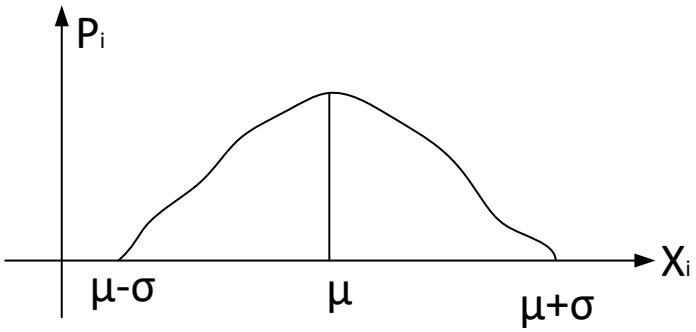
Sınıflandırılmamış serilerde ortanca hesaplanırken, veriler öncelikle küçükten büyüğe doğru sıralanırlar. n çift ise n/2 değer ile n/2+1 değerlerin ortalaması, n tek ise (n+1)/2 değeri ortancadır.

Mode (Peak) value:

Mode is the value with the highest probability in the distribution set. The elements in the array are rearranged according to the number of occurrences. If the highest probability value in the distribution is represented by more than one point, the result is not univocal since the mode corresponds to all of these values. In such cases, the distribution is said to be bimodal, trimodal or multimodal. Frequency is the number of vibrations per second, the number of frequencies.

Mod dağılım kümesinde olasılığı en yüksek değerdir. Dizinin tekrarlanma sayısına göre yeniden düzenlenmesidir. Dağılımda en yüksek olasılık değeri birden fazla nokta ile temsil ediliyorsa, mod bu değerlerin hepsine karşılık geldiğinden sonuç tek anlamlı olmaktan çıkar. Böylesi durumlarda dağılımın bimodal, trimodal ya da multimodal olduğundan söz edilir. Frekans, bir saniyedeki titreşim sayısı, sıklık sayısıdır.

Variance - Standard Deviation (Varyans - Standart Sapma):



Standard deviation is the root mean square (RMS) deviation resulting from the arithmetic mean of the variables. How much does each value in the data set deviate from the weighted mean? In fact, the square of the deviations from the weighted mean are examined. Why? Variance can be calculated to comment on the distribution of the data. In probability and statistics, the standard deviation of a probability distribution is a measure of the spread of a random variable or a set or values. It is usually indicated with the letter σ (lowercase sigma). Standard deviation is defined as the square root of the variance. Variance is the sum of the squares of the differences between the data and the arithmetic mean. It measures the spread of the measured data to the mean. Standard deviation gives the deviation from the arithmetic mean.

Standart sapma, değişkenlerin aritmetik ortalamasından kaynaklanan kök ortalama karesi **(RMS)** sapmasıdır. **Veri yığınındaki her bir değer ağırlık ortalamadan ne kadar sapmaktadır?** Aslında ağırlık ortalamadan olan sapmaları karesine bakılmaktadır. Neden? Varyans hesaplanarak verinin dağılımı hakkında yorum yapılabilir. Olasılık ve istatistikte, bir olasılık dağılımının standart sapması, rasgele değişken veya yığın veya değerlerin yayılmasının bir ölçüsüdür. Genellikle σ harfi ile belirtilir (küçük harf sigma). Standat sapma, varyansın karekökü olarak tanımlanır. Varyans, veriler ile aritmetik ortalama farklarının karelerinin toplamıdır. Ölçülen verilerin ortalamaya yayılmasını ölçer. Standart sapma, aritmetik ortalamadan olan sapmayı verir.

A large data set cannot be analyzed soundly. Why? Noisy, error, missing data, manipulation, … A minimum number of sample data sets representing the large data set is taken. Does the sample data set represent the large data set? For this, variance, skewness and kurtosis coefficients are examined.

Büyük veri yığını sağlıklı analiz edilemez. Neden? Gürültülü, hata, eksik data, manipüle, … Büyük veri yığınını temsil eden minumum sayıda örnek veri yığını alınır. Örnek veri yığını büyük veri yığınını temsil ediyor mu**? Bunun için varyans, çarpıklık ve basıklık katsayılarına bakılır.**

If the data values are close to the arithmetic mean, the standard deviation is small. Also, if many data points are far from the mean, the standard deviation is large. If all data values are equal, the standard deviation is zero. An important measure of change in a data distribution is the variance. The standard deviation is obtained by taking the square root of the variance. The standard deviation shows how close each value in the series is to the arithmetic mean. A small standard deviation shows that the deviations and risk in the mean are small, while a large standard deviation shows that the deviations and risk in the mean are large. If data is collected from the entire population in a study, it is called a complete count. A sample is a data group consisting of certain elements selected from a population.

Veri değerleri aritmetik ortalamaya yakınsa, standart sapma küçüktür. Ayrıca, birçok veri noktası ortalamanın uzağındaysa, standart sapma büyüktür. Tüm veri değerleri eşitse, standart sapma sıfırdır. Bir veri dağılımındaki değişimin önemli bir ölçüsü varyanstır. Varyansın karekökü alınarak standart sapma elde edilir. Standart sapma dizideki her bir değerin aritmetik ortalamaya yakınlığını gösterir. Standart sapmanın küçük olması ortalamalarda sapmaların ve riskin az olduğunu, standart sapmanın büyük olması ortalamalarda sapmaların ve riskin çok olduğunu gösterir. **Eğer bir çalışmada ana kütlenin tümünden veri toplanırsa buna tamsayım denir. Örneklem ise bir ana kütleden seçilen belirli elemanların oluşturduğu veri grubudur.**

Standard deviation of the population (large dataset):

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{N}$$

Here, the relationship between the standard deviation or arithmetic mean and random variables needs to be interpreted. If the standard deviation value is between 0 and 1, there is a close relationship between the data values in the arithmetic mean series, and the arithmetic mean represents the series. If it is greater than 1, it does not represent the sum of squares because it will be much larger.

Burada standart sapma ya da arimetik ortalama ve rassal değişkenler arasındaki ilişkinin yorumlanması gerekmektedir. Eğer standart sapma değeri 0 ile 1 arasında ise aritmetik ortalama dizi içerisindeki veri değerleri arasında yakın bir ilişki vardır, aritmetik ortalama diziyi temsil eder. Eğer 1'den büyük ise kareler toplamı çok daha büyük olacağından temsil etmez.

Instead of calculating probabilities by performing integral calculations, it is possible to convert any normal distribution to a standard normal distribution with the z transformation

and easily calculate the probability of random variables using the probability value table prepared for the standard normal distribution.
İntegral hesaplarını yaparak olasılıkları hesaplamak yerine, Z dönüşümü ile herhangi bir normal dağılımı standart normal dağılıma dönüştürmek ve standart normal dağılım için hazırlanmış olasılık değerleri tablosunu kullanarak rassal değişkenlerin olasılığı kolaylıkla hesaplamak mümkündür.

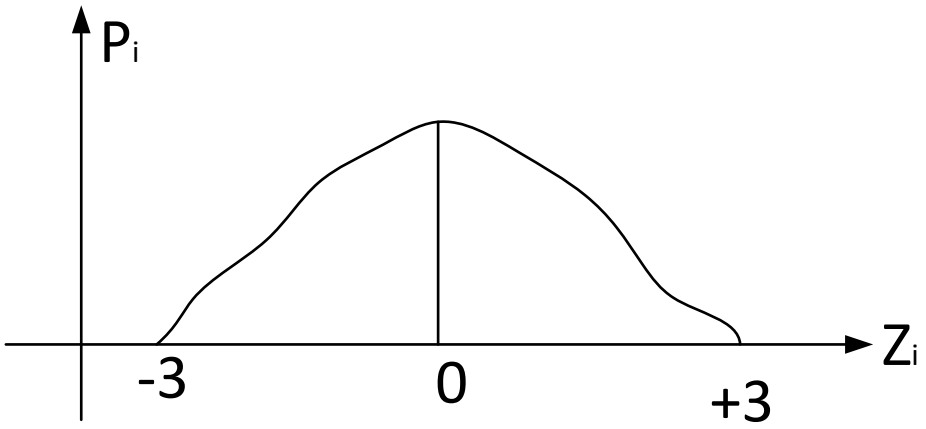
A Zi value corresponding to each Xi value of the random variable X of the normal distribution with arithmetic mean, μ, and standard deviation, σ is calculated, in other words, the random variable X is transformed into the random variable Z. In this case, the Z-transformation is,

Aritmetik ortalaması, μ; standart sapması, σ olan normal dağılımın X tesadüfi değişkenine ilişkin her X_i değerine karşılık gelecek bir Z_i değeri hesaplanmakta, başka bir deyişle X tesadüfi değişkeni Z tesadüfi değişkenine dönüştürülmektedir. Bu durumda Z-dönüşümü,

$$Z_i = \frac{X_i - \mu}{\sigma}$$

It is done with the formula above and is a measure of how many standard deviations a value falls away from its mean. The values that make up the series are normalized. The μ value in the formula represents the weighted average of the distribution or random variables. Xi values are independent variables such as path, time, frequency, score. P(xi)

Yukarıdaki formül ile yapılmaktadır ve bir değerin kendi ortalamasından kaç standart sapma uzağa düştüğünü gösteren bir ölçüdür. Diziyi oluşturan değerleri normalizasyon yapılır. Formülde yer alan μ değeri dağılımın ya da tesadüfi değişkenlerin ağırlık ortalamasını ifade etmektedir. Xi değerleri yol, zaman, sıklık, puan gibi bağımsız değişkenler. P(xi),



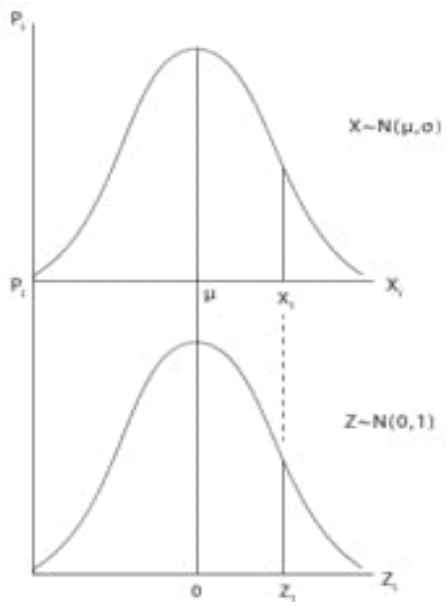


Figure: Normal Distribution and Standard Normal Distribution

Determining Probabilities Using the Standard Normal Distribution Table:

Standart Normal Dağılım Tablosu Kullanılarak Olasılıkların Belirlenmesi:

The basic rule of probability: The total area under the normal curve is always "1". In the case of a normal distribution, the arithmetic mean divides the area under the curve into two equal parts. In other words, there are two areas to the right of the mean with probability 0.5 and to the left of the mean with probability 0.5. The standard normal distribution table gives the area between the mean and the relevant Z value, that is, the probability. Therefore, when it comes to calculating probability other than this, it should not be forgotten that the total area under the curve mentioned above is "1" and that the mean divides the area under the curve into two equal parts with probability 0.5.

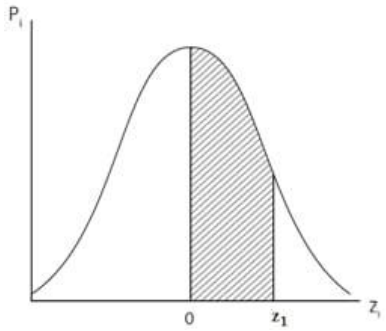
Olasılığın temel kuralı: Normal eğrinin altında kalan toplam alan, daima "1"dir. Normal dağılım söz konusu olduğunda aritmetik ortalama, eğri altındaki alanı iki eşit kısma ayırır. Yani, ortalamanın sağında 0,5 ve solunda 0,5 olasılıkla gerçekleşecek iki alan oluşur. Standart normal dağılım tablosu, ortalama ile ilgili Z değeri arasındaki alanı yani olasılığı vermektedir. Dolayısıyla, bunun dışında olasılık hesabı yapmak gerektiğinde yukarıda bahsedilen eğri altındaki toplam alanın "1" olduğu ve ortalamanın eğri altındaki alanı 0,5 olasılıklı iki eşit kısma ayırdığı unutulmamalıdır.

Since it is based on a continuous variable, the values that a random variable can take in a normal distribution are so close to each other that the probability for a single value is zero. Therefore, the probability must be calculated for a specific range, not for a single value.

Temelinde sürekli bir değişkenin bulunması sebebiyle, normal dağılımda rassal değişkenin alabileceği değerler birbirine o derece yaklaşmıştır ki tek bir değer için olasılık sıfırdır. Dolayısıyla, olasılık tek bir değer için değil belirli bir aralık için hesaplanması gerekmektedir.

The standard normal curve area table is prepared only for positive z values. Since the normal distribution is symmetrical and the arithmetic mean divides the curve area into two equal parts, the fact that the z value is positive or negative only indicates that it is above or below the mean, and since the distance to the mean will be the same even if it is in the + or - region, the probability values will also be the same.

Standart normal eğri alanları tablosu sadece pozitif z değerleri için hazırlanmıştır. Normal dağılımın simetrik olması ve aritmetik ortalamanın eğri alanını iki eşit kısma ayırması sebebiyle, z değerinin pozitif ya da negatif olması sadece ortalamanın altında ya da üstünde bir değer olduğunu ifade eder, + ya da – bölgede olsa da ortalamaya uzaklık aynı olacağı için olasılık değerleri de aynı olacaktır.



Şekil: Standart Normal Dağılım
Figure: Standard Normal Distribution

Properties of the standard normal distribution:

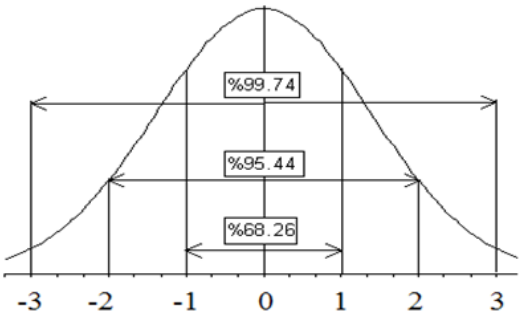
1) The distribution is symmetrical with respect to the mean. 50% is on the right and 50% is on the left.
2) The area under the normal distribution curve is equal to 1.
3) The arithmetic mean, median, peak value (mode) are equal to each other and are located where the maximum height is.
4) The sum of the probabilities of the values between Zi=0 and Zi=Zmax is equal to 0.5. Probability values are given in a table. Zmax=3 is taken.
5) The values of the variables showing the normal distribution:

- 68.26% of the observations fall within the mean and Zmax=± 1 standard deviation range,
- 95.44% of the observations fall within the mean and Zmax=± 2 standard deviation range, and

- 99.74% of the observations fall within the mean and Zmax=± 3 standard deviation range.

Standart normal dağılım Özellikleri:

1) Dağılım ortalamaya göre simetriktir. %50'si sağda, %50'si soldadır.
2) Normal dağılım eğrisinin altında kalan alan 1'e eşittir.
3) Aritmetik ortalama, ortanca, tepe değer (mod) birbirine eşittir ve maksimum yüksekliğin bulunduğu yerdedir.
4) Zi=0 ile Zi=Zmax arasındaki değerlerin olasılıklarının toplamı 0.5'e eşittir. Olasılık değerleri tablo halinde verilir. Zmax=3 alınır.
5) Normal dağılımı gösteren değişkenlerin aldıkları değerler:
   - Gözlemlerin %68.26' i ortalama ile Zmaks=± 1 standart sapma aralığına,
   - Gözlemlerin %95.44' i ortalama ile Zmaks=± 2 standart sapma aralığına, ve
   - Gözlemlerin %99.74 ' i ortalama ile Zmaks= ± 3 standart sapma aralığına düşer.



Normal distributions are easily converted to standard normal distributions. If the arithmetic mean, μ, and standard deviation, σ, of a function with a normal distribution (gauss) are known, the probability function is converted to the standard normal distribution. The selected value, X, is interpreted as being above or below the mean by Z=(X-μ)/σ, a standard deviation. After this conversion, the probabilities are found with the help of the standard normal distribution (Z) table.

Normal dağılımlar, standart normal dağılımlara kolaylıkla çevrilir. Normal dağılıma (gauss) sahip olan bir fonksiyonu aritmetik ortalaması, μ ve standart sapması, σ biliniyorsa olasılık fonksiyonu standart normal dağılıma çevrilir. Seçilen değer, X, ortalamadan Z=(X-μ)/σ, kadar standart sapmanın üstündedir ya da altındadır diye yorumlanır. Bu çevirme işleminden sonra olasılıklar, standart normal dağılım (Z) tablosu yardımıyla bulunur.

Z-table:

- It varies between ∓3.09. In order to convert to a normal standard distribution, the calculated Z value must be between ∓3.09.
- This corresponds to 99.98% of the theoretical universe.

- The Z table shows the standard deviation in 1/10 intervals

Z tablosu:

- ∓3.09 arasında değişmektedir. Normal standart dağılımına çevrilebilmesi için hesaplanan Z değerinin ∓3.09 arasında olması gerekmektedir.
- Bu, teorik evrenin %99.98'ine karşılık geliyor.
- Z tablosu 1/10'luk aralarla standart sapmayı gösteriyor.

The following table calculates the areas of normal curves for various Z values. Probabilities can be calculated directly with the help of these tables. The values given in the table are for the upper or lower half curves of the probability density function (between 0 and 0.5). Aşağıdaki tabloda çeşitli Z değerleri için normal eğrilerin alanlarını hesaplanmıştır. Bu tablolar yardımıyla doğrudan olasılıklar hesaplanabilir. Tabloda verilen değerler olasılık yoğunluk fonksiyonun üst ya da alt yarım eğriler içindir (0 ile 0.5 arasındadır).

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Zi values are found by adding the values on the left and the values on the top of the table.
Important warning: In probability calculation, the probability of any variable in the series is not calculated alone. The probability value for greater than or equal to or less than or equal to values is calculated.

Tablonun en solundaki değerler ile en üstteki değereler toplanarak Zi değerleri bulunur.
Önemli uyarı: Olasılık hesaplamada tek başına dizideki herhangi bir değişkenin olma olasılığı hesaplanmaz. Büyük eşit ya da küçük eşit değerlerine yönelik olasılık değeri hesaplanır.

Example:
Find the probability value P at Z<=0.78. In order to find the probability value from the table, Z=0.7 + 0.08 is distinguished. The value 0.7 is taken as the leftmost value; the value 0.08 is taken as the topmost value.

From the table
P1(0<Z<0.78)=0.2823

P(Z<=0.78)= P1(0<Z<0.78) + 0.5 = 0.2823 + 0.5 = 0.7823, There is a 78.23% probability that the Z value is less than or equal to 0.78.

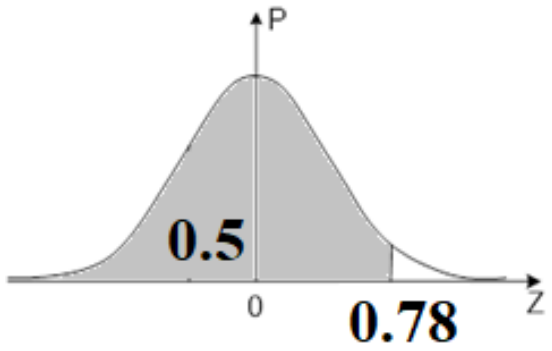
Z>0.78 grater than, Pk=1-P=1-0.7823=0.2177

Z<=0.78'deki P olasılık değerini bulunuz. Tablodan olasılık değerini bulabilmek için Z=0.7 + 0.08 ayrımı yapılır. 0.7 değeri en soldaki değer; 0.08 değeri ise en üstteki değer olarak alınır.

Tabloadan
P1(0<Z<0.78)=0.2823

P(Z<=0.78)= P1(0<Z<0.78) + 0.5 = 0.2823 + 0.5 = 0.7823, %78,23 olasılıkla Z değeri 0.78 den küçük ya da eşittir.

Z>0.78 durumunda Pk=1-P=1-0.7823=0.2177



While the P probability value at Z>0.78 is found, the P probability value at Z<=0.78 is found. The found value is subtracted from 1.
Z>0.78'deki P olasılık değeri bulunurken Z<=0.78'deki P olasılık değeri bulunur. Bulunan değer 1 den çıkarılır.

Example:
Find the probability value P at Z<=0.68. Z=0.6 + 0.08 is distinguished. The value 0.6 is taken as the leftmost value; the value 0.08 is taken as the topmost value.

From the table,
P1(0<Z<0.68)=0.2517

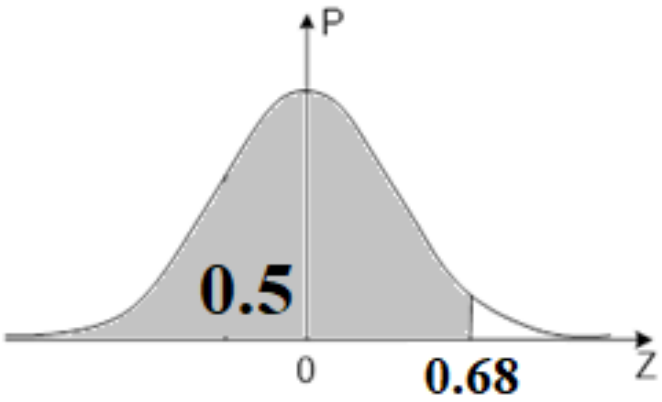
P(Z<=0.68)= P1(0<Z<0.68) + 0.5 = 0.2517 + 0.5 = 0.7517, There is a 75.17% probability that the Z value is less than or equal to 0.68.

In case of Z>0.68, Pk=1-P=1-0.7517=0.2483

Z<=0.68'deki P olasılık değerini bulunuz. Z=0.6 + 0.08 ayrımı yapılır. 0.6 değeri en soldaki değer; 0.08 değeri ise en üstteki değer olarak alınır.
Tablodan,
P1(0<Z<0.68)=0.2517

P(Z<=0.68)= P1(0<Z<0.68) + 0.5 = 0.2517 + 0.5 = 0.7517, %75,17 olasılıkla Z değeri 0.68 den küçük ya da eşittir.

Z>0.68 durumunda Pk=1-P=1-0.7517=0.2483

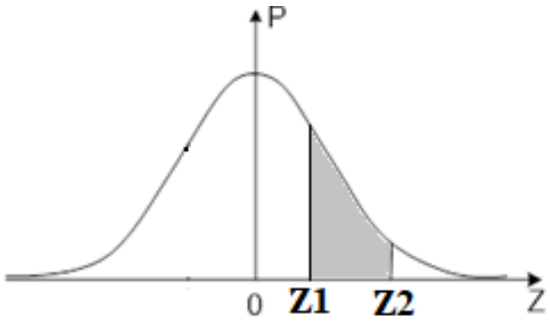


Example:
Find the probability value of P at 0.52<= Z<=0.78.

It is seen from the table that two probability values can be found.
1- P1(0 – Z1)
2- P2(0 – Z2)

P(Z2 – Z1)= P2- P1
From Z-Table, Z2=0.78 için P2= 0.2823
From Z-Tabel Z1=0.78 için P1= 0.1985

P(0.52<= Z<=0.78)= P2(0<Z<0.78) - P1(0<Z1<0.52) = 0.2823 - 0.1985 =0.0838
%8,38

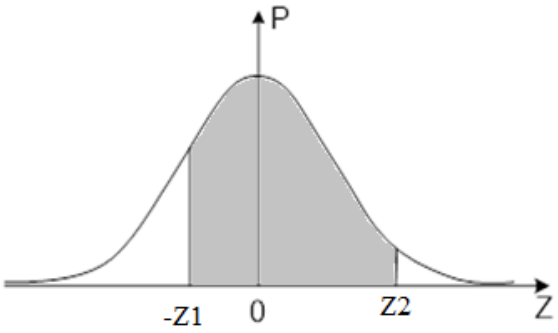


Example:

Find the probability value of P at -0.52<= Z<=0.78.

P(-0.52<= Z<=0.78)= P2(0<Z<0.78) + P1(0<Z1<0.52) = 0.2823 + 0.1985 =0.4808

%48,08



Example:

Instead of calculating probabilities by performing integral calculations, it is possible to convert any normal distribution to a standard normal distribution with the z transformation and easily calculate the probability using the probability value table prepared for the standard normal distribution.

İntegral hesaplarını yaparak olasılıkları hesaplamak yerine, z dönüşümü ile herhangi bir normal dağılımı standart normal dağılıma dönüştürmek ve standart normal dağılım için hazırlanmış olasılık değerleri tablosunu kullanarak olasılığı kolaylıkla hesaplamak mümkündür.

$$Z = \frac{X_i - \mu}{\sigma}$$

The μ value in the formula represents the weighted average of the data set or random variables, Xi represents the random variables in the data set, and σ represents the standard deviation. If the weighted average of a data set is μ=50, its standard deviation is σ=25, and X=60, then Z =?

Formülde yer alan μ değeri veri yığının ya da tesadüfi değişkenlerin ağırlık ortalamasını, X_i veri yığınındaki tesadüfi değişkenleri, σ ifadesi standart sapmayı ifade etmektedir. Bir veri yığının ağırlık ortalaması, μ=50, standart sapması, σ=25, X=60 ise Z =?

Z = (60-50)/25 =10/25 = 0.4

Comment on the situation where Z=0 is obtained even though the standard deviation and weighted mean do not change. If the weighted mean is equal to the random variable in the data set in the standard normal distribution, Z=0. What is the probability for Z <=0.4 (X<=60) from the standard normal distribution table? Express it as a percentage. The values in the table are divided into two as a probability of 0.5. Therefore, 0.5 is added to the value found in the table.

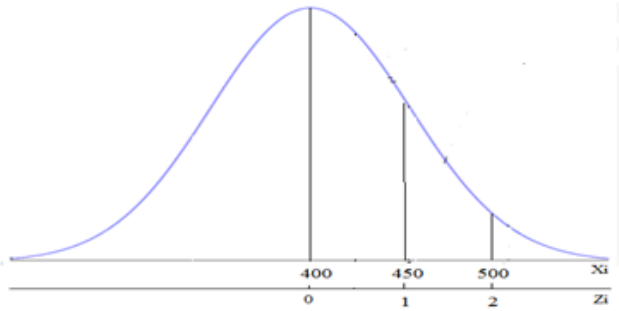
Standart sapma ve ağırlık ortalama değişmediği halde Z=0 çıkması durumunu yorumlayınız. **Standart normal dağılımda ağırlık ortalama veri yığınındaki tesadüfi değişkene eşit olursa Z= 0 olur.** Standart normal dağılım tablosundan Z <=0.4 (X<=60) için olasılık ne olur? Yüzde olarak ifade ediniz. Tablodaki değerler 0.5 olasılık olarak ortadan ikiye ayrılmaktadır. Bu nedenle tablodan bulunan değere 0.5 eklenir.

P=0.1554 + 0.5 = 0.6554 bulunur. %65,54

Example:
Let's say that the income of university students has a normal distribution. Since it is known that the mean, μ=400TL and the standard deviation, σ= 50, let X=500TL be the income of a randomly selected student. Can it be converted to a normal standard distribution?
a) What is the probability that the students' incomes are between 400TL and 500TL?
b) What is the probability that the students' incomes are less than 500TL?
c) What is the probability that the students' incomes are more than 500TL?

Üniversite öğrencilerinin gelirleri normal dağılıma sahip olsun. Ortalaması, μ=400TL ve standart sapması, σ= 50 olduğu bilindiğine göre tesadüfü seçilen bir öğrencinin geliri X= 500 TL ise olsun. Normal standart dağılımına çevrilebilir mi?

a) Öğrencilerin gelirlerinin 400TL ile 500TL arasında olma olasılığı nedir?
b) Öğrencilerin gelirlerinin 500TL'den az olma olasılığı nedir?
c) Öğrencilerin gelirlerinin 500TL'den fazla olma olasılığı nedir?



500TL In order for the normal distribution to be convertible to the standard distribution, Z must be between +/- 3 maximum values.

500TL Normal dağılımın standart dağılımına çevrilebilir olması için Z'in maksimum +/- 3 değerleri arasında olması gerekir.

a)

$$Z = \frac{X-\mu}{\sigma} = \frac{500-400}{50} = 2,$$

In this case, the income of the selected student is 2 standard deviations higher than the mean. The probability value between 400TL and 500TL corresponding to 2 found in the table is,

Bu durumda seçilen öğrencinin geliri, ortalamadan 2 standart sapma daha yüksektir. Tablodan bulanan 2'ye karşılık değer 400TL ile 500TL aralığındaki olasılık değeri,

P(0<Z<2)=P(400<X<500)=0.4772 dir. (%47.72)

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857

What is the probability that the students' income is greater than 400 TL and less than 500 TL? Since the weighted average = 400 TL, Z = 0. The value of Z between 0 and 2 is P(0<Z<2) = 0.4772; 47.72%

Öğrencilerin gelirlerinin 400TL den büyük 500TL den küçük olma olasılığı nedir?
Ağırlık ortalama=400TL, Z=0 olduğu için. Z'in 0 ile 2 arasındaki değeri P(0<Z<2)=0.4772; %47.72

b) What is the probability that the students' income is less than 500TL? The probability value between 0 and 400TL or between -3 and 0 for Z is 0.5.
From the Z-Table, P(Z<2.00)=P(Z<0)+ P(0<Z<2)=0.5+0.4772=0.9772 is obtained.

Öğrencilerin gelirlerinin 500TL den küçük olma olasılığı nedir? 0 ile 400TL ya da Z'in -3 değeri ile 0 arasındaki olasılık değeri 0.5 dir.
Z-Tablosundan, P(Z<2.00)=P(Z<0)+ P(0<Z<2)=0.5+0.4772=0.9772 elde edilir.

$$Z = \frac{X-\mu}{\sigma} = \frac{550-400}{50} = 3.$$

c) What is the probability that the student's income is more than 550 TL?
From the Z-Table, P(Z>3.00)=0.5- P(0<Z<3)=0.5-0.4987=0.0013 is obtained. (% 0.13)
Öğrencinin gelirinin 550 TL'den fazla olma olasılığı nedir?
Z-Tablosundan, P(Z>3.00)=0.5- P(0<Z<3)=0.5-0.4987=0.0013 elde edilir. (% 0,13)

d) What is the probability that student income is less than 450 TL?
Z=(450-400)/50=1.00
Öğrenci gelirlerinin 450TL den küçük olma olasılığı nedir?
Z=(450-400)/50=1.00

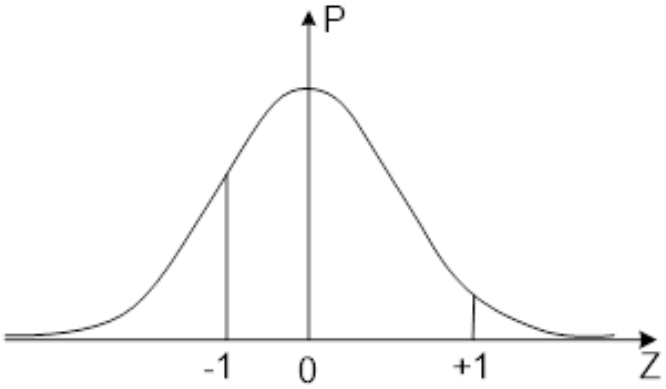
Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830

From the Z-Table, P(Z<1.00)= P(Z<0)+ P(0<Z<1)=0.5+0.3413=0.8413 is obtained.
Z-Tablosundan, P(Z<1.00)= P(Z<0)+ P(0<Z<1)=0.5+0.3413=0.8413 elde edilir.

e) What is the probability that student income is between 350 TL and 500 TL?
Öğrenci gelirlerinin 350TL ile 500 TL arasında olma olasılığı nedir?

Z1=(350-400)/50=-1.00
Z2=(500-400)/50=2.00



From the Z-Table,
P(0<Z1<1) = 0.3413
P(Z1<1) = P(Z1<0) + P(0<Z1<1) =0.5+0.3413=0.8413
P(Z1<-1)= 1-P(Z1<1)=1-0.8413=0.1587

P(0<Z2<2) = 0.4772
P(Z2<2) = P(Z2<0) + P(0<Z2<2) =0.5+0.4772=0.9772

P(-1<Z<2)=P(Z2<2)-P(Z1<-1)= 0.9772-0.1587=0.8185

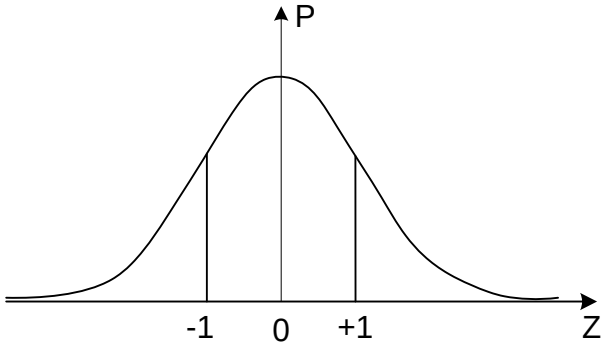
Then, the probability that student income will be between 350 TL and 500 TL is 81.85%.

O halde Öğrenci gelirlerinin 350TL ile 500 TL arasında olma olasılığı %81.85 dır.

f) What is the probability that student income will be between 350 TL and 450 TL?
Öğrenci gelirlerinin 350TL ile 450 TL arasında olma olasılığı nedir?

Z1=(350-400)/50=-1.00

Z2=(450-400)/50=1.00



From Z-Table,

450TL'den az ve 400TL'den büyük olanların yüzdesi ya da olasılığı, P(0<Z1<1) = 0.3413

450TL'den az olan öğrencilerin olasılığı, P(Z1<1) = P(Z1<0) + P(0<Z1<1) =0.5+0.3413=0.8413

350TL'den az olan öğrencilerin olasılığı, P(Z2<-1)= 1-P(Z1<1)=1-0.8413=0.1587

P(0<Z2<1) = 0.3413

P(Z2<1) = P(Z2<0) + P(0<Z2<1) =0.5+0.3413=0.8413

P(-1<Z<1)=P(Z2<2)-P(Z1<-1)= 0.8413-0.1587=0.6826

Then, the probability that student income will be between 350 TL and 500 TL is 68.26%.

O halde Öğrenci gelirlerinin 350TL ile 500 TL arasında olma olasılığı %68.26 dır.

Example:

Let the income of patients applying to the psychology clinic have a normal distribution. Since it is known that the mean is μ = 20,000 TL and the standard deviation is σ = 100, let the income of a randomly selected patient be X = 25,000 TL.

Psikoloji kliniğine başvuran hastaların gelirleri normal dağılıma sahip olsun. Ortalaması, μ=20.000TL ve standart sapması, σ= 100 olduğu bilindiğine göre tesadüfü seçilen bir hastanın geliri X= 25.000 TL ise olsun.

a) Can it be converted to a normal standard distribution? Note: In order to be converted to a normal standard distribution, the calculated Z value must be between ∓3.09. Z=(25000-20000)/100=5000/100=50, It cannot be converted to a normal standard distribution.

Normal standart dağılımına çevrilebilir mi? Not: Normal standart dağılımına çevrilebilmesi için hesaplanan Z değerinin ∓3.09 arasında olması gerekmektedir.

Z=(25000-20000)/100=5000/100=50, Normal standart dağılımına çevrilemez.

b) If it cannot be converted to a normal standard distribution, calculate the Z value again by taking the standard deviation as 5000. In this case, can it be converted to a normal standard distribution? Z=(25000-20000)/5000=5000/5000=1, since the Z value is between ∓3.09, it is converted to a normal standard distribution.

Eğer normal standart dağılıma çevrilemiyorsa standart sapmayı 5000 alarak yeniden Z değerini hesaplayınız, bu durumda normal standart dağılımına çevrilebilir mi?

Z=(25000-20000)/5000=5000/5000=1, Z değeri ∓3.09 arasında olduğu için normal standart dağılımına çevrilir.

c) If the standard deviation is 5,000 TL, what is the probability that patient income will be 30,000 TL less?

Z=(30,000-20,000)/5,000=10,000/5,000=2

From the Z-Table, P(Z<2.00)=P(Z<0)+ P(0<Z<2)=0.5+0.4772=0.9772 is obtained.

Standart sapmanın 5.000TL olması durumunda hasta gelirlerinin 30000TL az olma olasılığı nedir?

Z=(30.000-20.000)/5.000=10.000/5.000=2

Z-Tablosundan, P(Z<2.00)=P(Z<0)+ P(0<Z<2)=0.5+0.4772=0.9772 elde edilir.

d) What is the probability that patient income is more than 10,000 TL?

Z=(10,000-20,000)/5,000=-10,000/5000=-2

Due to the symmetrical feature of the table, Z=2 is taken. (The probability value between 0 and 2 is calculated.)

From the Z-Table, P(0<Z<2.00)= 0.4772

Then the probability of more than 10,000 TL is P(-2<Z)=0.4772+0.5=0.9772

Hasta gelirlerinin 10.000TL den fazla olma olasılığı nedir?

Z=(10.000-20.000)/5.000=-10.000/5000=-2

Tablonun simetrik özelliğinden dolayı Z=2 alınır. (0 ile 2 arasındaki olasılık değeri hesaplanır.)

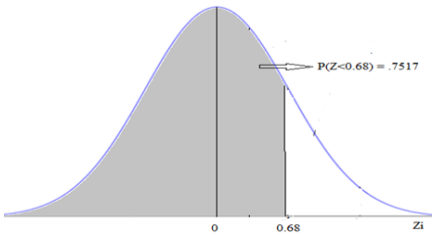
Z-Tablosundan, P(0<Z<2.00)= 0.4772

O halde 10000TL den fazla olma olasılığı, P(-2<Z)=0.4772+0.5=0.9772

Example:

Suppose that a normal distribution P(Z<0.68) is transformed to the standard normal distribution.

Standart normal dağılıma dönüştürülen bir normal dağılımın P(Z<0.68) olduğunu varsayalım.



Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852

For 0.68, the intersection of the .6 row and .08 column (.6 + .08 = .68) is searched.

First, P(0<Z<0.68) = 0.2517 is found.

From here, P(Z<0.68) = P(Z<0) + P(0<Z<0.68) = 0.5 + 0.2517 = .7517 is found.

0.68 için .6 satır ve .08 sütunun kesiştiği nokta (.6 + .08 = .68) aranır. Önce P(0<Z<0.68) = 0.2517 bulunur. Buradan P(Z<0.68) = P(Z<0) + P(0<Z<0.68) = 0.5 + 0.2517 = .7517 bulunur.

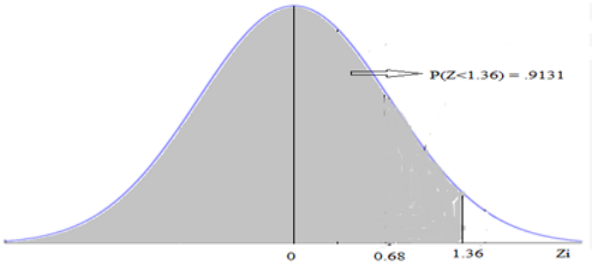
Example:

Let's assume that a normal distribution is P(Z<1.36) which is transformed into a standard normal distribution.

P(Z<1.36), definition: P(Z<1.36)=P(Z<0)+P(0<Z<1.36)

Standart normal dağılıma dönüştürülen bir normal dağılımın P(Z<1.36) olduğunu varsayalım.

P(Z<1.36), tanımı: P(Z<1.36)=P(Z<0)+P(0<Z<1.36)



Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319

It is searched at the intersection of points 1.3 and .06 (=1.3 +0 .06=1,36) in the Z-Table.

P(Z=1.36)=0.4131 is found. From here, P(Z<1.36) = 0.5+0.4131=0.9131 is found.

P(Z<1.36), probability of being 91.31%

P(Z>1.36)=1- P(Z<1.36)=1-0.9131=0.0869

1.3 ve .06 noktalarının (1.3 + .06) kesiştiği yerde aranır. P(Z=1.36)=0.4131 bulunur. Buradan Böylece P(Z<1.36) = 0.5+0.4131=0.9131 bulunur.

P(Z<1.36), olma olasılığı %91.31

P(Z>1.36)=1- P(Z<1.36)=1-0.9131=0.0869

Example:

Suppose that a normal distribution P(Z>2.18) is transformed to the standard normal distribution.

Standart normal dağılıma dönüştürülen bir normal dağılımın P(Z>2.18) olduğunu varsayalım.

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890

- P(0<Z<2.18)=0.4854
- P(Z<2.18)= P(Z<0) + P(0<Z<2.18) = 0.5 + 0.4854 = 0.9854
- P(Z>2.18) = 1 - P(Z<2.18) = 1- 0.9854 = 0.0146 is found.

Example:

Suppose a normal distribution that is transformed to a standard normal distribution is P(-1.26<Z<1.26). What is the probability in these intervals?

Standart normal dağılıma dönüştürülen bir normal dağılımın P(-1.26<Z<1.26) olduğunu varsayalım. Bu aralıklardaki olasılık nedir?

P(0<Z<1.26)=P(-1.26<Z<0); because there is symmetry.

P(0<Z<1.26) = 0.3962

P(-1.26<Z<1.26)=2*0.3962=0.7924

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177

Example:

Suppose a normal distribution P(1.36<Z<1.96) is transformed to the standard normal distribution.

Standart normal dağılıma dönüştürülen bir normal dağılımın P(1.36<Z<1.96) olduğunu varsayalım.

- P(Z<1.96)=0.5+0.475=0.975
- P(Z<1.36)=0.5+0.4131=0.9131

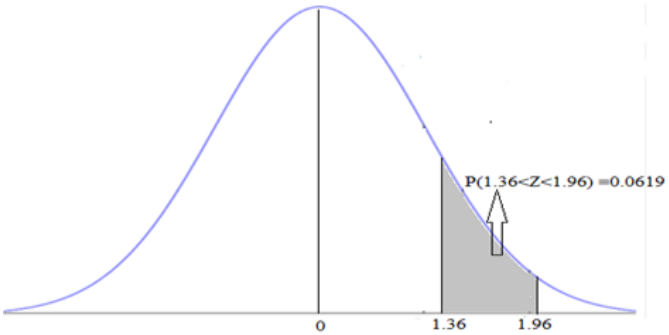
Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319

- Probability in these ranges P(1.36<Z<1.96)= P(Z<1.96) - P(Z<1.36) = 0.9750 - 0.9131 =0.0619 is found.



Example:

It has been determined that the diameters of the shafts produced in a manufacturing plant are normal with a mean of 3.0005 inches and a standard deviation of 0.001 inches. If the shafts produced are outside the range of 3.00 +/- 0.002 inches, these shafts are considered defective production. Accordingly, find the probability of defective products in total production. Desired probability expression: P(Faulty product)=1-P(2.998≤X≤3.002). To calculate this probability value, the continuous normal variable X is converted to standard normal.

Bir imalathanede üretilen millerin çaplarının ortalaması 3.0005 inç ve standart sapmalarının ise 0.001 inç olan normal dağılıma uyduğu tespit edilmiştir. Üretilen miller eğer 3.00 +/- 0.002inç aralığının dışında iseler bu miller hatalı üretim kabul edilmektedir. Buna göre toplam üretimdeki hatalı ürün olasılığını bulunuz. İstenilen olasılık ifadesi: P(Hatalı ürün)=1-P(2.998≤X≤3.002). Bu olasılık değerini hesaplamak için X sürekli normal değişkeni standart normal hale dönüştürülür.

$$Z = \frac{X - \mu}{\sigma}$$

$$Z_1 = \frac{X - \mu}{\sigma} = \frac{2.998 - 3.0005}{0.001} = -2.5$$

$$Z_2 = \frac{X - \mu}{\sigma} = \frac{3.002 - 3.0005}{0.001} = 1.5$$

P(Z ≤ 2.5)= P(Z ≥ -2.5)=0.4938+0.5=0.9938
P(0≥ Z ≥-2.5)=0.4938

P(Z≤1.5)=0.4332+0.5=0.9332
P(1.5 ≥ Z≥ 0)=0.4332

P(1.5 ≥ Z≥ -2.5)= P(0≥ Z ≥-2.5) + P(1.5 ≥ Z≥ 0)= 0.927
1-P(1.5 ≥ Z≥ -2.5)=1-0.927=0.073

Example:

The scores of students in a class are recorded. The mean of the scores received by the students is 60 and the standard deviation is 12. Given that the normal distribution of the scores is known, calculate the desired probabilities in the following options.

1) What is the probability that a randomly selected student will score between 60 and 70?

Tesadüfen seçilecek bir öğrenci 60 ile 70 arasında puan alma olasılığı nedir?

For this, it would be sufficient to first calculate the Z-standard value corresponding to 70 points. Because the 60 point value is the average value, Z60=0 (%50) will be found.

Z70=(70-60)/12=0.83 is found.

As can be seen from the figure, the probability value of any student's standard Z-values in the range of 60-70 is P(0.83 ≥ Z ≥ 0)= P(0.83) – P(Z=0).

Bunun için öncelikle 70 puana karşılık gelen Z-standart değerinin hesaplanması yeterli olacaktı. Çünkü 60 puan değeri ortalama değer olduğundan Z_{60}=0 (%50) çıkacaktır.
Z_{70}=(70-60)/12=0.83 bulunur.

Şekilden de görüldüğü gibi, herhangi bir öğrencinin 60-70 aralığında standart Z-değerlerine olasılık değeri, P(0.83 ≥ Z ≥ 0)= P(0.83) – P(Z=0) dır.

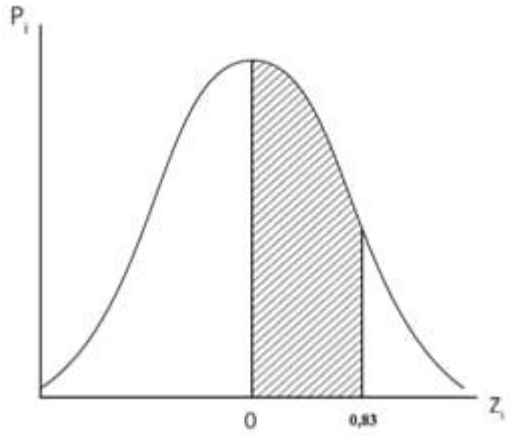


Figure: Probability calculation for the area between the mean and a value greater than the mean

From the standard normal distribution table, P(0.83 ≥ Z ≥ 0)=0.2967 is found. Therefore, the probability of a randomly selected student to score between 60-70 is 29.67%.
Standart normal dağılım tablosundan P(0.83 ≥ Z ≥ 0)=0.2967 bulunur. Dolayısıyla, tesadüfen seçilecek bir öğrencinin 60-70 arasında puan alma olasılığı % 29,67'dir.

1) What is the probability that a student who is selected by chance will get a score between 45-60?

The Z standard value of 45 is found as Z45=(45-60)/12=-1.25.

A negative Z value has no meaning other than showing that the score is below the mean. Since the normal distribution is a symmetric distribution and the mean divides this distribution into two equal parts exactly in the middle, the value of - 1.25 and the value of + 1.25 are equidistant from the mean.

P(1.25 ≥ Z ≥0)=P(0 ≥ Z ≥ -1.25)=0.3944 is found**.**

Tesadüfen seçilecek bir öğrencinin 45-60 arasında puan alma olasılığı nedir?

45'in Z standart değeri, Z_{45}=(45-60)/12=-1.25 bulunur.

Z değerinin negatif çıkmasının, puanın ortalamanın altında olduğunu göstermek dışında bir anlamı yoktur. Zira, normal dağılım simetrik bir dağılım olduğundan ve ortalama bu dağılımı tam ortadan iki eşit kısma ayırdığından, - 1,25 değeri ile + 1,25 değeri ortalamaya eşit uzaklıktadır.

P(1.25 ≥ Z ≥0)=P(0 ≥ Z ≥ -1.25)=0.3944 bulunur.

Therefore, the probability of someone randomly selected from among the selected students getting a score between 45-60 is 39.44%, or in other words, 39.44% of the students got a score between 45-60.

Dolayısıyla, seçilen öğrenciler arasından tesadüfen seçilecek birinin 45-60 arasında puan almaolasılığı % 39,44'dür ya da diğer deyişle, öğrencilerin % 39,44'ü 45-60 arasında puan almıştır.

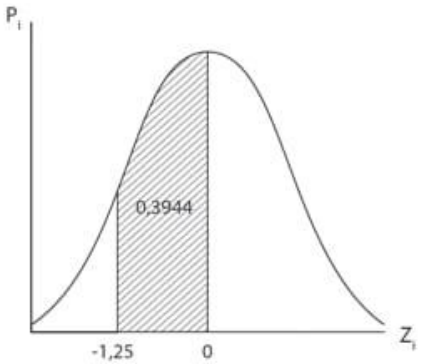


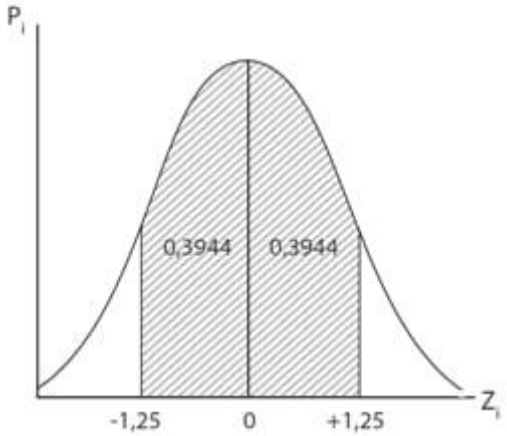
Figure: Determining the probability between the mean and a value less than the mean

2) What is the probability that a randomly selected student will receive a score between 45-75?

In this example, 45 is below the average of 60, and 75 is above it.

Z45=(45-60)/12=-1.25
Z75=(75-60)/12=1.25

Tesadüfen seçilecek bir öğrencinin 45-75 arasında puan alma olasılığı nedir?
Bu örnekte, 45 değeri ortalama değer olan 60'ın altında, 75 değeri ise üzerinde yer almaktadır.
Z_{45}=(45-60)/12=-1.25
Z_{75}=(75-60)/12=1.25



Şekil : P(-1,25 ≤ Z ≤ 1,25)

P(-1,25 ≤ Z ≤ 1,25) = P(0 ≤ Z ≤ 1,25) + P(-1,25 ≤ Z ≤ 0)
P(0 ≤ Z ≤ 1,25) = P(-1,25 ≤ Z ≤ 0)
P(-1,25 ≤ Z ≤ 1,25) = 2*P(0 ≤ Z ≤ 1,25)

From the table, P(0 ≤ Z ≤ 1.25)=0.3944 is found.
P(-1,25 ≤ Z ≤ 1,25) = 2*P(0 ≤ Z ≤ 1,25)=0.7888

Therefore, the probability of a randomly selected student receiving a score between 45 and 75 is 78.88%. In other words, 78.88% of the students received scores between 45 and 75.
Dolayısıyla, tesadüfen seçilecek bir öğrencinin 45 ile 75 arasında puan alma olasılığı % 78,88'dir. Başka bir deyişle öğrencilerden % 78,88'i 45-75 aralığında puan almıştır.

3)

What is the probability that a randomly selected student will receive 80 points more? The Z value corresponding to 80 is calculated as Z=(80-60)/12=1.67. Since the probability of the student receiving more than 80 points is desired, the probability of P(X > 80) needs to be calculated.
P(Z>1.67)=1-P(Z≤1.67)

P(Z≤1.67)=0.5 + 0.4525 =0.9525
P(Z>1.67)=1-0.9525= 0.0475

Therefore, the probability of a randomly selected student receiving more than 80 points is 4.75%. In other words, 4.75% of the students received more than 80 points.

Tesadüfen seçilecek bir öğrencinin 80 fazla puan alma olasılığı nedir?
80 adete karşılık gelen Z değeri, Z=(80-60)/12=1.67 olarak hesaplanır. Öğrencinin 80'den fazla puan alma olasılığı istendiğine göre, P(X > 80) olasılığının hesaplanması gerekiyor.
P(Z>1.67)=1-P(Z≤1.67)

P(Z≤1.67)=0.5 + 0.4525 =0.9525
P(Z>1.67)=1-0.9525= 0.0475

Dolayısıyla, tesadüfen seçilecek bir öğrencinin 80'den fazla puan alma olasılığı, % 4,75 olmaktadır. Başka bir deyişle, öğrencilerden % 4,75'i 80 üzerinde puan almıştır.

Figure: P(z > 1,67)

4)

What is the probability that a randomly selected student will receive less than 45 points?

With the standard value transformation, the Z value corresponding to 45 is calculated as Z=(45-60)/12=-1.25.

P(Z=-1.25)=P(Z=1.25)=0.3944

P(Z ≤ -1.25) = P(Z ≥ 1.25) =0.5-0.3944=0.1056. In this case, the probability that a randomly selected student will receive less than 45 points is 10.56%. In other words, 10.56% of the students received less than 45 points.

Tesadüfen seçilecek bir öğrencinin 45'den az puan alma olasılığı nedir?

Standart değer dönüşümü ile, 45'e karşılık gelen Z değeri, Z=(45-60)/12=-1.25 hesaplanır.

P(Z=-1.25)=P(Z=1.25)=0.3944

P(Z ≤ -1.25) = P(Z ≥ 1.25) =0.5-0.3944=0.1056 olarak hesaplanır. Bu durumda, tesadüfen seçilecek bir öğrencinin 45'den az puan alma olasılığı % 10,56'dır, Başka bir deyişle öğrencilerin % 10,56'sı 45'den az puan almıştır.

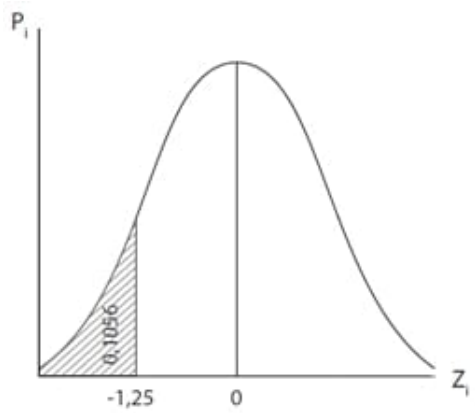


Figure: P(-1,25 < z)

5)

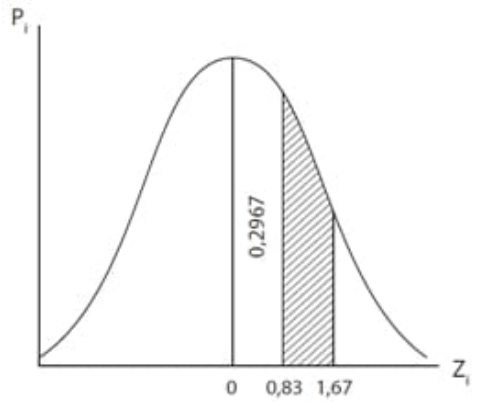
What is the probability that a randomly selected student will score between 70 and 80? First, let's find the standard value equivalents of 70 and 80:

Tesadüfen seçilecek bir öğrncinin 70-80 arasında puan alma olasılığı nedir?

Öncelikle 70 ve 80 puanlarının standart değer karşılıklarını bulalım:

$Z_{70}=(70-60)/12=0.83$

$Z_{80}=(80-60)/12=1.67$



Şekil: P(0,83 ≤ z ≤ 1,67)

P(0,83 ≤ z ≤ 1,67)=P(Z=1.67) – P(Z=0.83) =0.4525-0.2967=0.1558

Therefore, the probability of anyone being selected to score between 70 and 80 is calculated as 15.58%, in other words, 15.58% of the students score between 70-80.

Dolayısıyla, seçilecek herhangi birinin 70 ile 80 arasında puan alma olasılığı % 15,58 olarak hesaplanmakta, başka bir deyişle öğrencilerden % 15,58'i 70-80 arasında puan almaktadır.

6)

What is the probability that a student who is randomly drawn will score between 35-45? First, let's calculate the z values corresponding to the scores of 35 and 45:

Tesadüfen çekilecek bir öğrencinin 35-45 arasında puan alma olasılığı nedir?

Öncelikle 35 ve 45 puanlarına karşılık gelen z değerlerini hesaplayalım:

Z_{35}=(35-60)/12=-2.08

Z_{45}=(45-60)/12=-1.25

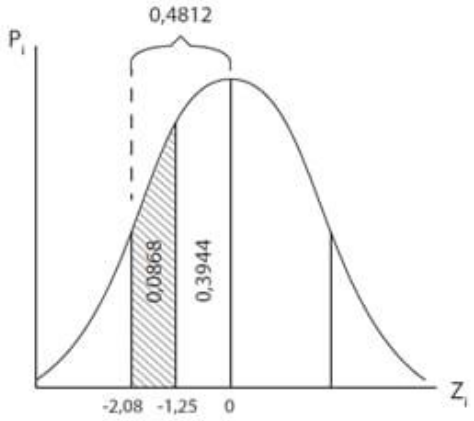


Figure: P(-2,08 ≤ Z ≤ -1,25)

P(-2,08 ≤ Z ≤ -1,25)=P(2,08 ≥ Z ≥ 1,25) = P(Z=2.8)-P(Z=1.25)=0.4812 - 0.3944 =0.0868

This value also indicates that 8.68% of students received scores in the 35-45 range.

Bu değer aynı zamanda, öğrencilerin % 8,68'inin 35-45 aralığında puan aldığının göstergesidir.

7)

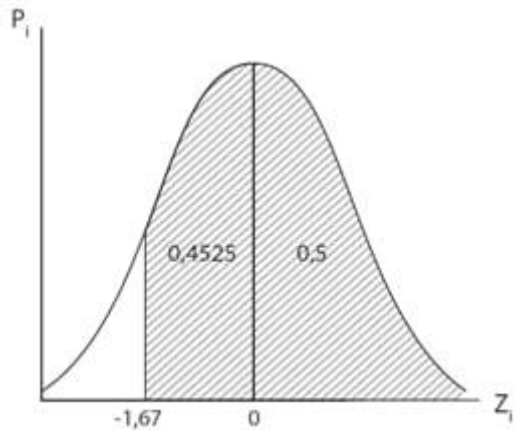
What is the probability that a randomly selected student will score more than 40 points? First, let's calculate the standard value corresponding to 40 points: Z40=(40*60)/12=-1.67

Tesadüfen seçilecek bir öğrencinin 40'dan fazla puan alma olasılığı nedir?

Önce, 40 puana karşılık gelen standart değeri hesaplayalım: Z40=(40*60)/12=-1.67

P(Z ≥ -1.67) = P(0≥ Z ≥ -1.67) + P(Z≥ 0)= P(0≥ Z ≥ -1.67) + 0.5

P(0≥ Z ≥ -1.67) = P(1.67 ≥ Z ≥ 0) =P(Z=1.67)=0.4525 is found as.



Şekil: P(z > -1,67)

P(Z ≥ -1.67) = 0.5 + 0.4525=0.9525

Therefore, the probability that a randomly selected student will have scored above 40 points is 95.25%, or in other words, 95.25% of the examinees scored above 40 points.

Dolayısıyla, tesadüfen seçilecek bir öğrencinin 40 puanın üstünde puan almış olma olasılığı % 95,25'dir veya başka bir deyişle, sınava girenlerin % 95,25'i 40 puandan fazla puan almıştır.

8)

What is the probability that a student who is randomly selected will have a score of less than 45 or more than 80?

Let's draw our shape and scan the desired probability region by finding the Z-values corresponding to 45 and 60 points. Z45=(45-60)/12=-1.25, Z60=(80-60)/12=1.67

Tesadüfen seçilecek bir öğrenci 45 puandan az ya da 80 puandan fazla puan almış olma olasılığı nedir?

45 ve 60 puanlara karşılık gelen Z-değerlerini bularak, şeklimizi çizelim ve istenen olasılık bölgesini tarayalım. Z45=(45-60)/12=-1.25, Z60=(80-60)/12=1.67

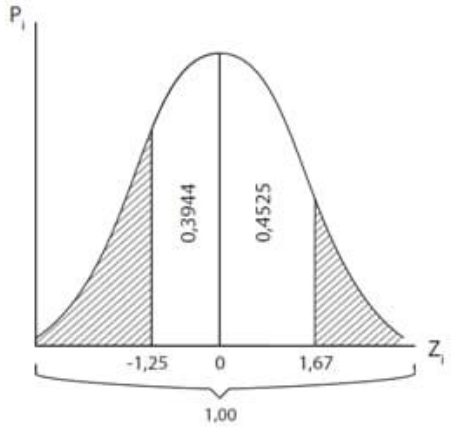


Figure: P(-1,25 < z ; z > 1,67)

P(Z ≥ 1.67) = 1 - P(Z ≤ 1.67)
P(Z ≤ 1.67)= 0.5 + P(Z=1.67) = 0.5 + 4525 =0.9525
P(Z ≥ 1.67) = 1 - 0.9525 =0.0475

P(Z ≤ -1.25) = P(Z ≥ 1.25)
P(Z ≥ 1.25) = 1 - P(Z ≤ 1.25)
P(Z ≤ 1.25)= 0.5 + P(Z=1.25) = 0.5 + 0.3944 =0.8944
P(Z ≥ 1.25) = 1 - 0.8944 =0.1056
P(X < 45 veya X> 80) = 0,1531 is found as.

According to this result, the probability that a randomly selected student will have a score lower than 45 or higher than 80 is 15.31%. In other words, 15.31% of the students who took the exam scored lower than 45 or higher than 80.

Bu sonuca göre, sınava giren öğrencilerarasından tesadüfen seçilecek herhangi birinin 45 puandan düşük ya da 80 puandan yüksek puan almış olma olasılığı % 15,31'dir. Başka bir deyişle, sınava giren öğrencilerin % 15,31'i 45 puandan az ya da 80 puandan yüksek puan almıştır.

Example:

It is known that the weights of margarines produced in a margarine factory are normally distributed with a mean of 250 g (μ= 250 g) and a standard deviation of 8 g (σ =8 g). A margarine package selected randomly from the daily production is

Bir margarin fabrikasında üretilmekte olan margarinlerin ağırlıklarının ortalaması 250 gr (μ= 250 gr) ve standart sapması 8 gr (σ =8 gr) olarak normal dağıldığı bilinmektedir. Günlük üretim içinden tesadüfen seçilecek bir margarin paketinin,

1)

In order to calculate the probability of a margarine to be chosen by chance being between 250-255 gr, we must first calculate the standard Z-value corresponding to the value of 255.

Z255=(255-250)/8=0.625

P(0.625 ≥ Z ≥ 0)=P(Z=0.625) – P(Z=0)=0.2357 – 0 =0.2357 is found.

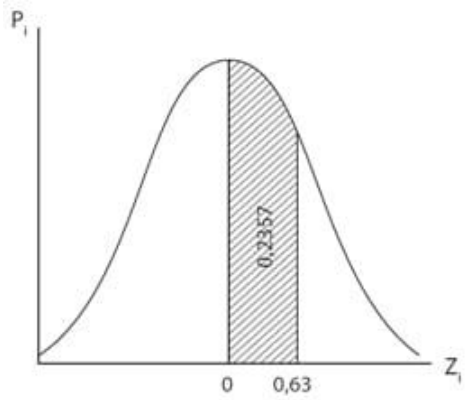
The probability of a margarine to be chosen by chance being between 250-255 gr is calculated as 23.57%.

Tesadüfen seçilecek bir margarinin 250-255gr arasında bulunma olasılığını hesaplayabilmek için öncelikle, 255değerine karşılık gelen standart Z-değerini hesaplamalıyız.

Z_{255}=(255-250)/8=0.625

P(0.625 ≥ Z ≥ 0)=P(Z=0.625) – P(Z=0)=0.2357 – 0 =0.2357 bulunur.

Tesadüfen seçilecek bir margarinin 250-255 gr arasında bulunma olasılığını % 23,57 olarak hesaplamış olunur.



Şekil: P (0 ≤ z ≤ 0,63)

2)

To find the probability that a margarine that will be selected by chance will weigh more than 255 grams,

P(X > 255) = 1-P(X<255)
P(Z>0.625)=1-P(Z<0.625)=1-0.5-0.2357=0.2643 is calculated.
The probability that a margarine that will be selected by chance will weigh more than 255 grams is 26.43%. In other words, 26.43% of the margarines produced in this factory weigh more than 255 grams.

Tesadüfen seçilecek bir margarinin 255 gr'dan fazla olma olasılığını bulmak için,

P(X > 255) = 1-P(X<255)
P(Z>0.625)=1-P(Z<0.625)=1-0.5-0.2357=0.2643 hesaplanır.
Tesadüfen seçilecek bir margarinin 255 gr'dan fazla olması olasılığı % 26,43'tür. Başka bir deyişle, bu fabrikada üretilmekte olan margarinlerin % 26,43'ü 255 gramın üzerinde ağırlığa sahiptir.

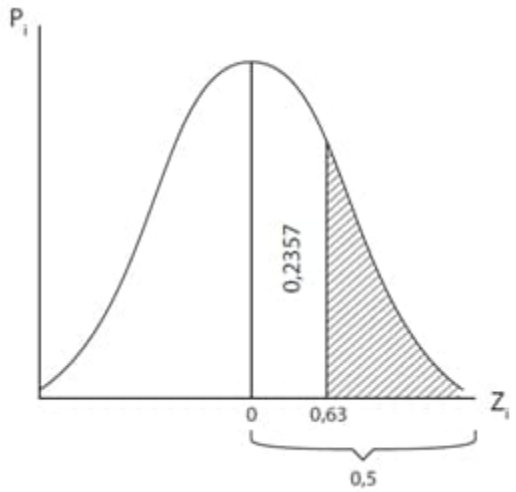


Figure: P (z > 0,63)

3)

To calculate the probability that a margarine selected by chance weighs less than 255 grams,

P(Z<0.625=0.5+P(Z=0.625)=0.5 + 0.2357=0.7357

Therefore, the probability that a margarine selected by chance weighs less than 255 grams is 73.57%. In other words, 73.57% of the margarines produced in this factory weigh less than 255 grams.

Tesadüfen seçilecek bir margarinin 255 gramdan az olması olasılığını hesaplamak için,

P(Z<0.625=0.5+P(Z=0.625)=0.5 + 0.2357=0.7357

Dolayısıyla, tesadüfen seçilecek bir margarinin 255 gramdan az ağırlığa sahip olması olasılığı % 73,57 olmaktadır. Başka bir deyişle bu fabrikada üretilen margarinlerin % 73,57'si 255 gramdan az ağırlığa sahiptir.

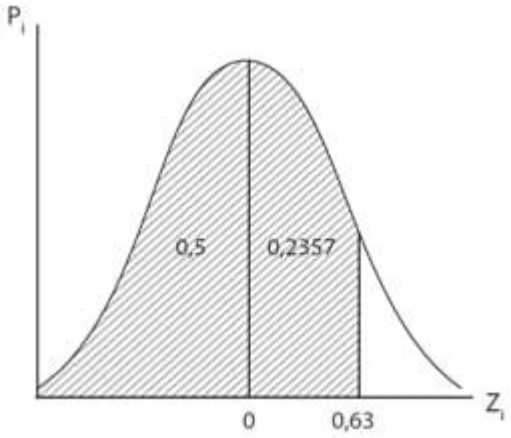


Figure: P (z < 0,63)

.3. Confidence Interval

A sample data set representing a big data set is investigated. Determining whether the main mass arithmetic mean of the sampled data set is within certain intervals with a certain probability is important in data analytics. The weighted mean of either the population, the sample, or a predicted value can be taken as a reference. The standard deviation of the sample data set and the number of samples must also be given.

Büyük veri yığınını temsil eden örneklem veri yığını alınır. Örneklem alınan veri yığının ana kütle aritmetik ortalamasının belirli olasılıkla belirli aralıklarda olup olmadığının belirlenmesi, veri analitiğinde önemlidir. İster anakütlenin, ister örneklemin, isterse öngörüde bulunulan bir değerin ağırlık ortalaması referans olarak alınabilir. Örneklem veri yığının standart sapması ve örnek sayısı da verilmek zorundadır.

It is often impossible to examine the main mass. However, with the help of samples, the limits within which the mean of the sample mass lies can be estimated within the selected probability of error. The "Confidence Interval or Confidence Limits of the Given Arithmetic Mean" can be determined for the limits within which the mean of the sample data mass lies.

Ana kütleyi incelemek çoğu zaman olanaksızdır. Ancak örneklemler yardımıyla örneklem kütlenin ortalamasının içinde bulunduğu sınırlar, seçilen yanılma olasılığında tahmin edilebilir. Örneklem veri yığınına ait ortalamanın içinde bulunduğu sınırlara "**Verilen aritmetik ortalamanın Güven Aralığı ya da Güven Sınırları**" belirlenebilir.

$$\bar{x} = \frac{\sum x}{n}$$

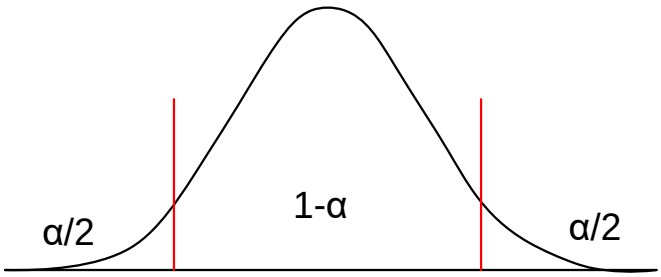
$\bar{x}$ — mean; $\sum x$ — data values; n — sample size

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

σ — standard deviation

It is not expected that the mean of a sample of n numbers taken from a normal distribution of a small population taken from the population will be equal to the mean of the population. In this case, the confidence interval can be estimated. The Z value for the given confidence interval probability is:

Ana kütleden alınan küçük bir yığının normal dağılımından alınan n adet bir örneklemin ortalamasının ana kütlenin ortalamasına eşit olması beklenemez. Bu durumda güven aralığı tahmin edilebilir. Verilen güven aralığı olasılığındaki Z değeri:



When the length of the confidence interval (1-α) is distributed equally on both sides of the mean due to the symmetry property of the normal distribution, in other words, α will be in the form of α/2 and α/2. Therefore, the Confidence Interval will be,

Due to the symmetry property of the normal distribution, the length of the confidence interval (1-α) will be distributed equally on both sides of the mean, in other words, α/2 on the lower side and α/2 on the upper side. Therefore, the Confidence Interval will be,

$$(X - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) \leq \mu \leq (X + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

In this expression, X: The given weight indicates the average.

Sample means must be normally distributed around the population mean. The standard error is nothing more than the standard deviation of the normal distribution of sample means. From the relationship between the mean and standard deviation of a normal distribution, the limits within which any sample mean can be found at certain levels of probability can be estimated.

Örnek ortalamalar, ana kütle ortalaması etrafında normal bir dağılım göstermelidirler. Standart hata ise örnek ortalamaların normal dağılımının standart sapma hatasından başka bir şey değildir. Bir normal dağılımın ortalaması ile standart sapması arasındaki ilişkiden hareketle herhangi bir örneklem ortalamasının belirli olasılık kademelerine göre bulunabileceği sınırlar tahmin edilebilir.

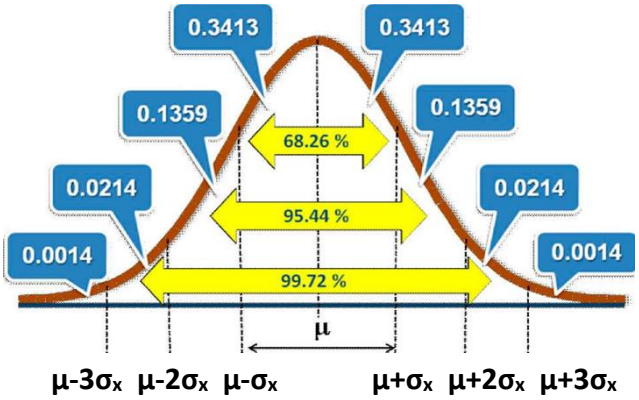


Figure: Various standard error limits of normally distributed sample means.

As can be seen in the figure, 99.72% of the sample means are within the standard error limits of ± 3 σx from the population mean. Similar explanations can be made for other probability levels. Accordingly, if the population mean is shown as μ and the standard error of the sample means is shown as σx, the sample means are calculated as follows:

Şekilde görüleceği gibi, örneklem ortalamalarının % 99.72'si ana kütle ortalamasından ± 3 σ_x standart hata sınırları arasında bulunur. Bunun gibi diğer olasılık kademeleri için de benzer açıklamalar yapılabilir. Buna göre, ana kütle ortalaması μ ve örneklem ortalamaların standart hatasını σ_x şeklinde gösterilecek olursa örnek ortalamaları şu şekilde hesaplanır:

% 68.26' ü	**$\mu - \sigma_x$**	**ile**	**$\mu + \sigma_x$**
% 95.44' i	**$\mu - 2\,\sigma_x$**	**ile**	**$\mu + 2\,\sigma_x$**
% 99.7' si	**$\mu - 3\,\sigma_x$**	**ile**	**$\mu + 3\,\sigma_x$**

The sample means drawn from a population with a normal distribution and a known variance are normally distributed around the population mean. The mean of this distribution (the mean of the sample means) is equal to the population mean and can be expressed as $\overline{X} = \mu$. The standard deviation of the distribution of sample means is known as the standard error and is calculated with the following formula:

Dağılımı normal ve varyansı belli olan yığından çekilen örneklem ortalamaları ana kütle ortalaması etrafında normal dağılım gösterir. Bu dağılımın ortalaması (örnek ortalamalarının ortalaması) ana kütle ortalamasına eşittir ve $\overline{X} = \mu$ şeklinde gösterilebilir. Örnek ortalamalarına ait dağılımının standart sapması, standart hata olarak bilinir ve aşağıdaki formülle hesaplanır:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

The expression $\sqrt{\frac{N-n}{N-1}}$ in the formula is the correction factor and if n/N<0.05, it can be ignored in the standard error calculation since it will not affect the result. In such cases, the standard error is determined as follows:

Formüldeki, $\sqrt{\frac{N-n}{N-1}}$ ifadesi, düzeltme faktörü olup, $\frac{n}{N} < 0.05$ ise, sonucu etkilemeyeceği için, standart hata hesabında ihmal edilebilir. Böyle durumlarda standart hata aşağıdaki gibi belirlenmektedir:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

If σ is unknown, the sample standard deviation error, if Sx is given, is written as $\boldsymbol{\sigma_x} = \frac{\boldsymbol{S_x}}{\sqrt{\boldsymbol{n}}}$. If the main mass mean is unknown, it can be estimated within a certain confidence interval based on the sample.

σ bilinmiyorsa, örneklem standart sapma hatası, S_x verilmiş ise $\boldsymbol{\sigma_x} = \frac{\boldsymbol{S_x}}{\sqrt{\boldsymbol{n}}}$ olarak yazılır. Ana kütle ortalaması bilinmiyorsa örneklemden yola çıkarak belli bir güven aralığında tahmin edilebilir.

Example:

The mean value of test results in a clinic is 100 units. It will be tested how accurately the hypothesis can predict within a certain confidence interval. To test this hypothesis, 1024 patients were randomly selected and the standard deviation was found to be 20 units. In what range does the arithmetic mean vary in the 95% confidence interval?

Bir klinikteki test sonuçlarının ortalama değeri 100 birimdir. Varsayımın belirli bir güven aralığı içerisinde ne kadar isabetli tahmin edebildiği test edilecektir. Bu, hipotezi test etmek için rastgele 1024 hasta seçilmiş ve standart sapmanın 20 birim olduğu bulunmuştur. %95 güven aralığında aritmetik ortalama hangi aralıkta değişir?

Calculate the confidence interval of the population mean using the expression below.
Ana kütle ortalamasının güven aralığını aşağıdaki ifadeyi kullanarak hesaplayınız.

$$(X - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) \leq \mu \leq (X + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

In this expression, X: The weighted mean of the large data set, the standard deviation or standard error of the n sample means:

Bu ifadede, X: Büyük veri yığının ağırlık ortalaması, n örneklem ortalamalarının standart sapması ya da standart hatası:

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_x = \frac{20}{\sqrt{1024}} = \frac{20}{32} = \frac{5}{8} = 0.625gram$$

In order to find the Z value from the table with a 95% confidence interval probability, first P(Zα/2)=0.95/2=0.475 is calculated. The Zα/2 value that meets the P(Zα/2)=0.475 value is determined from the standard normal distribution table. Zα/2=1.96. Then, if the values are substituted,

% 95 güven aralığı olasılıkla Z değeri tablodan bulunması için önce $P(Z_{\alpha/2})=0.95/2=0.475$ hesap edilir. Standart normal dağılım tablosundan $P(Z_{\alpha/2})= 0.475$ değerini karşılayan $Z_{\alpha/2}$ değeri belirlenir. $Z_{\alpha/2}=1.96$ dir. O halde değerler yerine konursa,

100 - (1.96)(0.625) ≤ μ ≤ 100 + (1.96)(0.625) = (100-1.225) ≤ μ≤ (100+1.225) is found as.

Example:

The average weight of the population is 100 grams. The standard deviation of the variability of the weights of 100 products in the sample is known to be σ = 5 grams. Estimate the average weight of the products according to certain probability levels (confidence intervals). Accordingly, first the sample standard error is found,

Anakütlenin ağırlık ortalaması 100gr. Örneklem 100 adet ürünün ağırlıkları ile ilgili değişkenliğin standart sapması, σ = 5 gram olduğu bilinmektedir. Ürünlerin ortalama ağırlığını belirli olasılık kademelerine (güven aralığına) göre tahmin edin. Buna göre önce örneklem standart hata bulunur,

$\boldsymbol{\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{100}} = 0.5\ gr}$ **is obtained.**

Since the sample means are normally distributed around the population mean; by taking advantage of the properties of the normal distribution, the average weight of the population units is,

$$(X - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) \leq \mu \leq (X + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

It can be estimated as follows. The "Z" expression in the formula is the critical values of the standard normal distribution at certain probability levels. The population mean estimate for various probability levels is calculated as shown below.

If the interval estimate of the mean of the population is to be made with a probability of 68%, P(Zα/2)=0.68/2=0.34 is obtained. The Zα/2 value that meets the value of P(Zα/2)=0.34 is determined from the standard normal distribution table. Zα/2=1.

% 68 olasılıkla, ana kütlenin ortalamasının aralık tahmini yapılacak ise, $P(Z_{\alpha/2})=0.68/2=0.34$ elde edilir. Standart normal dağılım tablosundan $P(Z_{\alpha/2})= 0.34$ değerini karşılayan $Z_{\alpha/2}$ değeri belirlenir. $Z_{\alpha/2}=1$ dir.

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830

Interval estimate of the weight average of the population,

Ana kütlenin ağırlık ortalamasının aralık tahmini,

100 - (1)(0.5) ≤ μ ≤ 100 + (1)(0.5) = 99.5gr ≤ μ ≤ 100.5gr is obtained (68% probability).

If the interval estimate of the mean of the population is to be made with 95% probability, P(Zα/2)=0.95/2=0.475 is obtained. The Zα/2 value that meets the value of P(Zα/2)=0.475 is determined from the standard normal distribution table. Zα/2=1.96.

% 95 olasılıkla, ana kütlenin ortalamasının aralık tahmini yapılacak ise,
$P(Z_{\alpha/2})=0.95/2=0.475$ elde edilir. Standart normal dağılım tablosundan $P(Z_{\alpha/2})= 0.475$ değerini karşılayan $Z_{\alpha/2}$ değeri belirlenir. $Z_{\alpha/2}=1.96$ dir.

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817

The interval estimate of the population mean is found as 100 - (1.96)(0.5) ≤ μ ≤ 100 + (1.96)(0.5) = 99.02gr ≤ μ≤ 100.98gr.
Ana kütlenin ortalamasının aralık tahmini,
100 - (1.96)(0.5) ≤ μ ≤ 100 + (1.96)(0.5) = 99.02gr ≤ μ≤ 100.98gr olarak bulunur.

If the interval estimate of the mean of the population is to be made with 99% probability, P(Zα/2)=0.99/2=0.495 is obtained. The Zα/2 value that meets the value of P(Zα/2)=0.495 is determined from the standard normal distribution table. Zα/2=2.58.

% 99 olasılıkla, ana kütlenin ortalamasının aralık tahmini yapılacak ise,
$P(Z_{\alpha/2})=0.99/2=0.495$ elde edilir. Standart normal dağılım tablosundan $P(Z_{\alpha/2})= 0.495$ değerini karşılayan $Z_{\alpha/2}$ değeri belirlenir. $Z_{\alpha/2}=2.58$ dir.

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964

The interval estimate of the population mean is found as 100 - (2.58)(0.5) ≤ μ ≤ 100 + (2.58)(0.5) = 98.71 gr ≤ μ ≤ 101.29 gr.

Example:

The average weight of male students at a university is 80 kilograms and it will be tested how accurately it can predict within a certain confidence interval. Let's assume that 100 male students are randomly selected to test the hypothesis. The standard deviation of the sample data here is assumed to be 13.6 kilograms. Let's assume that the desired confidence interval is 95%. In what interval should the arithmetic mean vary within the 95% interval?

Bir üniversitedeki erkek öğrencilerin ortalama ağırlığı 80 kilogram olduğu varsayımı belirli bir güven aralığı içerisinde ne kadar isabetli tahmin edebildiği test edilecektir. Bu, hipotezi test etmek için rastgele 100 erkek öğrenci seçildiğini varsayalım. Buradaki örneklem verinin standart sapmanın 13,6 kilo olduğunu varsayılır. İstenen güven aralığının %95 seçildiğini varsayalım. %95 aralığında aritmetik ortalama hangi aralıkta değişmelidir?

The sample standard error is,
Örneklem standart hata,

$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{13.6}{\sqrt{100}} = 1.36gr$ is found.

With 95% probability, the interval estimate of the mean of the population will be made. P(Zα/2)=0.95/2=0.475 is obtained. The Zα/2 value that meets the value of P(Zα/2)=0.475 is determined from the standard normal distribution table. Zα/2=1.96.

% 95 olasılıkla, ana kütlenin ortalamasının aralık tahmini yapılacaktır.
$P(Z_{\alpha/2})=0.95/2=0.475$ elde edilir. Standart normal dağılım tablosundan $P(Z_{\alpha/2})= 0.475$ değerini karşılayan $Z_{\alpha/2}$ değeri belirlenir. $Z_{\alpha/2}=1.96$ dir.

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817

Confidence interval estimate of the population mean,
Ana kütlenin ortalamasının güven aralık tahmini,

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

80 - (1.96)(1.36) ≤ μ ≤ 80 + (1.96)(1.36) = 77.33gr ≤ μ ≤ 82.66gr is found as.

Example:

According to a study conducted in Turkey, people watch television for an average of 8 minutes a day and it is known that the standard deviation of the sample data of 64 people is 4 minutes, estimate the mean of the main population with a 95% confidence interval. Accordingly, the sample standard error is,

Yapılan bir araştırmada Türkiye'de insanların ortalama günde 8 dakika televizyon seyrettikleri ve 64 kişi üzerinde yapılan örneklem verinin kütlenin standart sapmasının 4 dakika olduğu bilindiğine göre % 95 güven aralığında ana kütle ortalamasını tahmin ediniz. Buna göre örneklem standart hata,

$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{64}} = 0.5dk$ is calculated.

With 95% probability, the interval estimate of the mean of the population will be made. P(Zα/2)=0.95/2=0.475 is obtained. The Zα/2 value that meets the value of P(Zα/2)=0.475 is determined from the standard normal distribution table. Zα/2=1.96.

% 95 olasılıkla, ana kütlenin ortalamasının aralık tahmini yapılacaktır.

$P(Z_{\alpha/2})=0.95/2=0.475$ elde edilir. Standart normal dağılım tablosundan $P(Z_{\alpha/2})= 0.475$ değerini karşılayan $Z_{\alpha/2}$ değeri belirlenir. $Z_{\alpha/2}=1.96$ dir.

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817

The interval estimate of the population mean is,

Ana kütlenin ortalamasının aralık tahmini,

8 - (1.96)(0.5) ≤ μ ≤ 8 + (1.96)(0.5) = 7.02dk ≤ μ≤ 8.98dk is found as.

Example:

In a study conducted on 49 students at the university, the average IQ was found to be 110 and the standard deviation was found to be 14. In which range will the students' average IQ be at the 99% confidence interval?
The sample standard error is,

Üniversitede 49 öğrenci üzerinde yapılan çalışmada IQ ortalaması 110 ve standart sapması 14 bulunmuştur. %99 güven aralığında öğrencilerin IQ ortalaması hangi aralıkta olacaktır.
Buna göre örneklem standart hata,

$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{14}{\sqrt{49}} = 2$ is calculated.

With 99% probability, the interval estimate of the mean of the population will be made. P(Zα/2)=0.99/2=0.495 is obtained. The Zα/2 value that meets the value of P(Zα/2)=0.495 is determined from the standard normal distribution table. Zα/2=2.58.

% 99 olasılıkla, ana kütlenin ortalamasının aralık tahmini yapılacaktır.
$P(Z_{\alpha/2})=0.99/2=0.495$ elde edilir. Standart normal dağılım tablosundan $P(Z_{\alpha/2})= 0.495$ değerini karşılayan $Z_{\alpha/2}$ değeri belirlenir. $Z_{\alpha/2}=2.58$ dir.

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964

The interval estimate of the population mean,
Anakütlenin ortalamasının aralık tahmini,
110 - (2.58)(2) ≤ μ ≤ 110 + (2.58)(2) = 104.76 ≤ μ ≤ 115.16 is found as.

It can be interpreted as follows: At a 99% confidence level, the IQs of the students are between 104.76 and 115.16.
% 99 güven düzeyinde öğrencilerinin IQ'leri 104.76 ile 115.16 arasındadır, şeklinde yorumlanabilir.

7.4. Hypothesis Testing

Hypothesis is an assumption or prediction put forward about a situation. A hypothesis is put forward; its accuracy is tested, and whether the result is correct or incorrect should be acceptable for decision making. The weight average of the large dataset, the weight average and variance of the sample dataset are known. It is the determination of whether the hypothesis put forward will be accepted one-sided or two-sided under certain probabilities. Probability is given. Can a sample dataset represent a large dataset? A statistical hypothesis is an assumption put forward about the status of the main mass (dataset). The purpose of hypothesis testing is to transform the put forward assumption into decision making. An assumption obtained by performing statistical calculations is formulated in a way that will be accepted or rejected according to the test result.

Hipotez, bir durum hakkında ileri sürülen bir varsayım ya da öngörüdür. Bir hipotez ileri sürülerek; doğruluğu test edilir, sonucun doğru ya da hatalı olması karar vermeye yönelik olarak kabul edilebilir olmalıdır. Büyük veri yığının ağırlık ortalaması, örneklem veri yığının ağırlık ortalaması ve varyansı biliniyor. Belirli olasılıklarda ileri sürülen hipotezin tek taraflı ya da çift taraflı kabul edilip edilmeyeceğinin belirlenmesidir. Olasılık verilmektedir. Örneklem veri yığını büyük veri yığınını temsil edebilir mi? İstatistiksel hipotez, ana kütlenin (Veri yığını) durumu hakkında ileri sürülen bir varsayımdır. Hipotez testinde amaç, ileri sürülen varsayımın karar vermeye dönüştürülmesidir. İstatistik hesaplar yapılarak elde edilen bir varsayım, test sonucuna göre kabul veya reddedilecek şekilde formüle edilir.

In order to test the hypothesis (assumption), a sample set is determined from the large data set. The population is infinitely large (Big Data). It contains infinite problems; missing, faulty, noisy, anomaly, manipulated, uncertain...
The selection is a non-refundable selection and is made with a completely random process.
In order to develop the ability to make the right decision, the number of sample data should be at least 30. The average of the sample population is calculated. Then, the calculated sample average is compared with the known population average and the hypothesis is tried to be verified.

Ortaya konan hipotezi (varsayım) test etmek için, büyük veri yığını içinden örneklem bir kümesi belirlenir. Anakütle sonsuz büyüklüktedir (Büyük Veri). Sonsuz sıkıntılar içerir; eksik, hatalı, gürültülü, anomali, manipule, belirsiz...
Seçim iadesiz seçimdir ve tamamen rassal bir süreçle yapılmıştır.
Doğru karar verme yeteneği geliştirmek için örneklem veri sayısı minimum 30 olmalıdır. Örneklem kütlenin ortalaması hesaplanır. Daha sonra hesaplanan örnek ortalaması bilinen yığın ortalamasıyla karşılaştırılır ve hipotez doğrulanmaya çalışılır.

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Hipotezlerin Belirlenmesi:

Hypotheses are determined: In statistics, H0 is called the null hypothesis, and H1 is called the alternative or research hypothesis. The null hypothesis shows the known or determined value of the population parameter. H1, the alternative hypothesis, is the main hypothesis that directs the research, that is, is desired to be proven. The hypotheses are arranged in such a way that when one of them is rejected, the other is accepted. In the null hypothesis, the main population parameter is determined with the conditions of being equal to, not equal to, small or large to a certain value. In the alternative, situations are put forward that are the opposite of the situation to be proven.

Hipotezler belirlenir: İstatistikte, H0 sıfır hipotezi, H1 ise alternatif veya araştırma hipotezi isimleri ile adlandırılır. Sıfır hipotezi, yığın parametresinin bilinen veya belirlenmiş değerini gösterir. **H1, Alternatif hipotez ise, araştırmayı yönlendiren yani kanıtlanmak istenen asıl hipotezdir.** Hipotezlerden, biri red edildiğinde diğeri kabul edilecek şekilde düzenlenir. Sıfır hipotezinde, ana kütle parametresinin belirli bir değere eşit, eşit değil; küçük ya da büyük koşulları ile belirlenir. Alternatifinde ise kanıtlanacak durumun zıttı olduğu durumlar ileri sürülür.

In a decision-making phase using the given information, the null hypothesis is either accepted or rejected in favor of the opposite hypothesis. In this case, the null hypothesis is compared with the results specified in the table below in the decision-making process.
Verilen bilgiler kullanılarak bir karar verme aşamasında sıfır hipotezi ya kabul edilir ya da karşı hipotezin lehine sıfır hipotez ret edilir. Bu durumda sıfır hipotezi için karar sürecinde aşağıda verilen tabloda belirtilen sonuçlarla karşılaştırılır.

Significance testing:

A mathematical model used to test a hypothesis about a parameter in the dataset to which it belongs, using a sample dataset. Computations are performed on selected samples to gather more precise information about the properties of the population, thus providing a systematic way to test claims or ideas about the entire dataset.

Anlamlılık testi:

Bir örneklem veri seti kullanılarak ait olduğu veri yığın kümesindeki bir parametre hakkındaki bir hipotezi test etmek için kullanılan matematiksel bir modeldir. Yığının özellikleri hakkında daha kesin bilgiler toplamak için seçilen örnekler üzerinde hesaplamalar yapılır, böylece tüm veri kümesiyle ilgili iddiaları veya fikirleri test etmek için sistematik bir yol sağlanmış olunur.

One-sided test, acceptance and rejection regions according to 10% significance level (90% probability): Zk = ±1.28, α=0.10

One-sided test, acceptance and rejection regions according to 5% significance level (95% probability): Zk = ±1.65, α=0.05
One-sided test, acceptance and rejection regions according to 1% significance level (99% probability): Zk = ±2.33, α=0.01

Tek taraflı test, % 10 anlam düzeyine göre (%90 olasılık) kabul ve red bölgeleri: Z_k = ±1.28, α=0.10
Tek taraflı test, % 5 anlam düzeyine göre (%95 olasılık) kabul ve red bölgeleri: Z_k = ±1.65, α=0.05
Tek taraflı test, % 1 anlam düzeyine göre (%99 olasılık) kabul ve red bölgeleri: Z_k = ±2.33, α=0.01

The critical values of Zk are given below according to their importance level.			
Level of Meaning	Sol Kuyruk Testi H_o => 100 H_1 < 100	Left Tail Test H_o <= 100 H_1 > 100	Two-Way Test H_o = 100 H_1 ≠ 100
0.10	-1.28 (%80)	+1.28 (%80)	±1.65 (%90)
0.05	-1.65 (%90)	+1.65 (%90)	±1.96 (0.95)
0.01	-2.33 (%98)	+2.33 (%98)	±2.58 (%99)

Z_k' nin kritik değerleri önem düzeyine göre aşağıda verilmiştir.			
Anlam Düzeyi	Sol Kuyruk Testi H_o => 100 H_1 < 100	Sağ Kuyruk Testi H_o <= 100 H_1 > 100	Çift Yönlü Test H_o = 100 H_1 ≠ 100
0.10	-1.28 (%80)	+1.28 (%80)	±1.65 (%90)
0.05	-1.65 (%90)	+1.65 (%90)	±1.96 (0.95)
0.01	-2.33 (%98)	+2.33 (%98)	±2.58 (%99)

		Decision	
		Protect Hypothesis	Hypothesis Rejected
Accuracy in the Stack	TRUE (H0)	TRUE 1-α	Tip-1 Wrong α
	Wrong (H0)	Tip-2 Wrong β	TRUE 1-β (Güç)

		Karar	
		Hipotezi Koru	Hipotez Ret Edilmiş
Yığındaki Doğruluk	Doğru (H0)	Doğru 1-α	Tip-1 Hata α
	Hatalı (H0)	Tip-2 Hata β	Doğru 1-β (Güç)

As can be seen from the table above, if the null hypothesis is true and this hypothesis is rejected, an incorrect decision has been made. This incorrect decision is called a Type-1 error. The probability of making a Type-1 error will be □. □ is the significance level of the hypothesis. Since the decision will either be accepted or rejected, the probability of accepting the decision will be (1-α).

According to this table, the second error to be made is called a Type-2 error and the probability of accepting the null hypothesis while it is false is indicated by □. The probability of rejecting a false hypothesis is (1-α). (1-α) is called the power of the test.

Yukarıdaki tablodan görüleceği üzere, sıfır hipotezi doğru ve bu hipotez ret edilmiş ise, hatalı bir karar verilmiştir. Bu hatalı karar Tip-1 hata adını alır. Tip-1 hata yapma olasılığı α olacaktır. α, hipotezin **anlamlılık düzeyidir.** Karar ya kabul ya da red edileceğine göre, kararı kabul etme olasılığı da (1-α) olur.

Bu tabloya göre, yapılacak ikinci hataTip-2 hata olarak adlandırılır ve sıfır hipotezi yanlış iken kabul etme olasılığı β ile gösterilir. Yanlış bir hipotezi red etme olasılığı da (1-β) dır. (1-β)' ya **testin gücü** denir.

Statistical "Significance Level" (Risk level, Margin of Error, Margin of Error) α is determined, critical value, Zk is calculated.

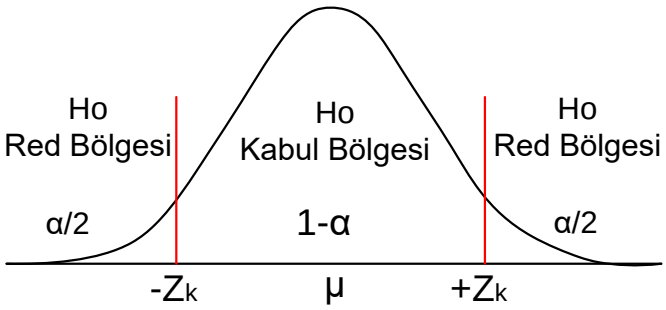
The area that determines the confidence limit in two-sided tests is the middle region of the normal distribution.
A claim such as "The average weight of the products is different from 100 grams" can be established with the null hypothesis as follows:
H0 : μ = 100 gr +/- delta
H1 : μ ≠ 100 gr.

The significance level can also be expressed as the significance level. Since the sample means are normally distributed around the population mean, the areas outside the confidence limits when calculating the sample mean probability are known as the "Significance Level" and these are 10%, 5% and 1%, respectively. The probability values are 90%, 95%, and 99%.

İstatistiksel "Anlam Düzeyi" (Risk düzeyi, Yanılgı Payı, Hata payı) α belirlenir, kritik değer, Z_k hesaplanır.

Çift taraflı testlerde güven sınırını belirleyen alan, normal dağılımın orta bölgesidir.
"Ürünlerin ortalama ağırlığı 100 gramdan farklıdır" şeklindeki bir iddia, sıfır hipotezi ile birlikte şu şekilde kurulabilir:
H_0 : μ = 100 gr +/- delta
H_1 : μ ≠ 100 gr.

Anlam düzeyi, önem seviyesi şeklinde de ifade edilebilir. Örnek ortalamaları yığın ortalaması etrafında normal dağılım gösterdiğinden, örnek ortalama olasılığı hesaplanırken güven sınırları dışında kalan alanlar "Anlam Düzeyi" olarak bilinir ve bunlar sırasıyla % 10, % 5 ve % 1 dir. Olasılık değerleri %90, %95, %99 olur.



Two-sided hypothesis test, 10% significance level (90% probability) acceptance and rejection regions: Zk = ±1.65, α/2=0.05
Two-sided hypothesis test, 5% significance level (95% probability) acceptance and rejection regions: Zk = ±1.96, α/2=0.025
Two-sided hypothesis test, 1% significance level (99% probability) acceptance and rejection regions: Zk = ±2.58, α/2=0.005

Çift taraflı hipotez test, % 10 anlam düzeyine göre (%90 olasılık) kabul ve red bölgeleri: Z_k = ±1.65, α/2=0.05
Çift taraflı hipotez test, % 5 anlam düzeyine göre (%95 olasılık) kabul ve red bölgeleri: Z_k = ±1.96, α/2=0.025

Çift taraflı hipotez test, % 1 anlam düzeyine göre (%99 olasılık) kabul ve red bölgeleri: Z_k = ±2.58, α/2=0.005

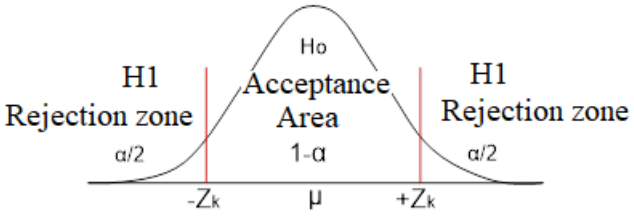
In one-sided tests, the area that determines the confidence limit depends on the direction of the test and is only at one end of the normal curve.
A claim such as "The average weight of the products is equal to 100 grams is light" can be established as follows:
H0 : μ = 100 gr.
H1 : μ ≠ 100 gr.

A claim such as "The average weight of the products is heavier than 100 grams" is shown as follows:
H0 : μ ≤ 100 gr.
H1 : μ > 100 gr.

A claim such as "The average weight of the products is lighter than 100 grams" can be established as follows:
H0 : μ ≥ 100 gr.
H1 : μ < 100 gr.

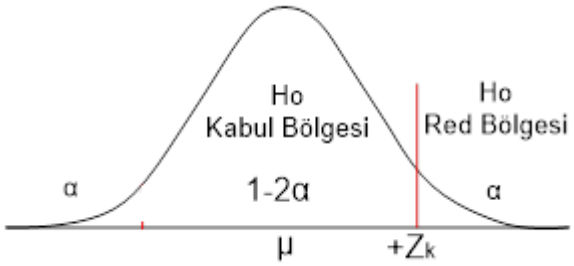
When finding the Z value from the symmetrical property of the table, the value at the other end must be similar to the other end. The value is found from the table by doubling the confidence interval. When the values of being large or small are investigated, a one-sided test is performed.

Tek taraflı testlerde güven sınırını belirleyen alan, testin yönüne bağlı ve normal eğrinin sadece bir ucundadır.
"Ürünlerin ortalama ağırlığı 100 grama eşittir hafiftir" şeklindeki bir iddia şu şekilde kurulabilir:
H_0 : μ = 100 gr.
H_1 : μ ≠ 100 gr.



"Ürünlerin ortalama ağırlığı 100 gramdan hafiftir" şeklindeki bir iddia şu şekilde kurulabilir:
H_0 : μ ≥ 100 gr.

$H_1 : \mu < 100$ gr.



"Ürünlerin ortalama ağırlığı 100 gramdan ağırdır" şeklindeki bir iddia ise şu şekilde gösterilir:
$H_0 : \mu \leq 100$ gr.
$H_1 : \mu > 100$ gr.

Tablonun simetrik özelliğinden Z değeri bulunurken diğer uçtaki değerin benzeri öteki uçtada olmak zorundadır. Güven aralığı iki katı alınarak tablodan değer bulunur. Büyük ve küçük olma değerleri araştırıldığında tek taraflı test yapılır.

Since the confidence limits are the sum of the probabilities at both ends of the normal curve and outside these limits, the area at each end is half of the confidence limit (5% for 10%, 20% for 1%, and 2% for 10%).

Güven sınırları, normal eğrinin her iki ucunda ve bu sınırların dışında kalan olasılıkların toplamı olduğundan, her bir uçtaki alan güven sınırının yarısı (%10 için %5, %1 için sırayla %20, %10 için %2) kadardır.

If the sample mean probability is taken as 90%, α=0.1. In order to find the correct Z value, the value to be considered must be 80%. Because, if the rejection region is 10%, the acceptance region is 80%+10%=90%. P(Zk)=0.80, calculations are made for 80%. The part of 0.80/2=0.40 is found from the table, then the value of 0.5 is added and the Zk value of the 90% part is calculated. In order to find the correct Zk value from the table, the probability expression is also symmetrical.

Örnek ortalama olasılığı, %90 alınırsa, α=0.1 olur. Doğru Z değerini bulabilmek için göz önüne alınacak değer %80 olmak zorundadır. Çünkü, red bölgesi %10 ise kabul bölgesi %80+%10=%90 olur. P(Zk)=0.80, %80'e hesaplamalar yapılır. Tablodan 0.80/2=0.40'lık kısım bulunur, ardından 0.5 değeri de ilave edilrerek %90'lık kısmın Zk değeri hesaplanmış olur. Tablodan Zk değerini doğru bulabilmek için olasılık ifadesi de simetrilk hale getirilir.

Because, as can be seen from the figure, since the Z value will be looked at from the table by taking advantage of the symmetry feature of the curve, the value on the left is within the

probability. The significance level is 10% on the right. The 10% on the left side is in the process. To find the value from the Z-Table, P(Zk)=0.80/2=0.40,

Çünkü, şekilden görüleceği üzere eğrinin simetri özelliğinden faydalanarak tablodan Z değerine bakılacağı için sol tarafdaki değer olasılığın içerisindedir. Anlam düzeyi %10 sağ taraftadır. Sol taraftaki %10 ise işlemin içindedir. Z-Tablosundan değer bulmak için P(Zk)=0.80/2=0.40,

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177

*Z*k= 1.28 değerini alır.

If the sample mean probability is taken as 95%, α=0.05. P(Zk)=0.90
To find the value from the Z-Table, P(Zk)=0.90/2=0.45,
Örnek ortalama olasılığı %95 alınırsa, α=0.05 olur. P(Zk)=0.90
Z-Tablosundan değer bulmak için P(Zk)=0.90/2=0.45,

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633

Zk=1.65 takes the value.

If the sample mean probability is taken as 99%, α=0.01. P(Zk)=0.98
To find the value from the Z-Table, P(Zk)=0.98/2=0.49,
Örnek ortalama olasılığı %99 alınırsa, α=0.01 olur. P(Zk)=0.98
Z-Tablosundan değer bulmak için P(Zk)=0.98/2=0.49,

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936

Zk =2.33 takes the value.

Attention:

- Since the probability calculations found from the Z-Table are between 0 and 0.5, multiplying by 2 will be sufficient in two-sided tests. However, 0.5 should be added to the value found in one-sided tests.
- When determining Z from the probability value, the Z value determined in two-sided tests will be within the confidence limit due to the symmetry feature. In one-sided tests, 2 times the significance level is taken into account to reduce to the symmetry feature

- Z-Tablosundan bulunan olasılık hesaplamaları 0 ile 0.5 arasında olduğundan çift yönlü testlerde 2 ile çarpmak yeterli olacaktır. Oysa Tek yönlü testlerde bulunan değere 0.5 eklemek gerekmektedir.
- Olasılık değerinden Z belirlenirken çift yönlü teslerde simetri özelliğinden dolayı belirlenen Z değeri güven sınırı içerisinde olacaktır. Tek taraflı testler de ise simetri özelliğine indirgemek için anlam düzeyinin 2 katı göz önüne alınır.

Example:

If 90% two-way probability is taken, α/2=0.05. 1- 2*α =P(Zk)=0.90
To find the α value from the symmetric property of the Z-Table, P(Zk)=0.90/2=0.45,

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633

Zk = 1.65 is obtained.

If 95% is taken as two-way, α/2=0.025. P(Zk)=0.95
To find the value from the Z-Table, P(Zk)=0.95/2=0.475,

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817

Zk = 1.96 is obtained.

If 99% is taken as two-way, α/2=0.005. P(Zk)=0.99
To find the value from the Z-Table, P(Zk)=0.99/2=0.495,

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964

Zk = 2.58 elde edilir.

Hypothesis testing

Once a hypothesis is established, it is tested in one of two stages.

Stage -1:

Zh value is calculated based on the sample.

$$Z_h = \frac{X - \mu}{\sigma_x}$$

The calculated Zh value is compared with the Zk value. If Zh > Zk, the H0 hypothesis is rejected and the alternative hypothesis is accepted.

Sample standard error, , $\sigma_x = \frac{\sigma}{\sqrt{n}}$ is calculated. $\overline{X} \pm \sigma_x$ is determined.

Stage -2:

$\bar{X}_h = Z_h * \frac{\sigma}{\sqrt{n}} + \mu$

When the sample mean, μ, is the critical mean value, $\bar{X}_h$, if μ > $\bar{X}_h$, the null hypothesis is rejected and the alternative hypothesis is accepted.

Hipotez test edilmesi

Bir hipotez kurulduktan sonra 2 aşamadan biri ile test edilir.

Örneklem standart hata, $\sigma_x = \frac{\sigma}{\sqrt{n}}$ hesaplanır. $\overline{X} \pm \sigma_x$ belirlenir.

Aşama -1: Örneklemden yola çıkılarak Z_h degeri hesaplanır.

$$Z_h = \frac{X - \mu}{\sigma_x}$$

Hesaplanan Z_h değeri Z_k değeri ile karşılaştırılır. $Z_h > Z_k$ ise H_0 hipotezi reddedilir ve alternatif hipotez kabul edilir.

Aşama-2:

$\bar{X}_h = Z_h * \frac{\sigma}{\sqrt{n}} + \mu$

Örneklem ortalama, μ, kritik ortalam değeri, $\bar{X}_h$ olduğunda $\mu > \bar{X}_h$ ise sıfır hipotezi reddedilir ve ve alternatif hipotez kabul edilir.

Evaluation and interpretation of the test result:

1) When zh<zk, the Ho hypothesis is accepted and;

- The population means from which these two samples were drawn are equal,
- These two populations are random samples drawn from the same population,
- The difference we observe between the two sample means is thought to be a small difference that has emerged as a result of probability and is not statistically significant.

2) When zh>zk, the Ho hypothesis is rejected and;

- The opposite of the idea belonging to the Ho hypothesis is accepted, that is, H1 is accepted.
- The probability (probability) that this large zh value has emerged due to probability is very low. This probability (p value) is also smaller than the α value we have chosen. This z value that emerged with such a small probability is no longer attributed to randomness but to the fact that the population is really different.

Test sonucunun değerlendirilmesi ve yorumlanması:

1) $z_h<z_k$ olduğunda, H_o hipotezi *kabul edilir* ve;

- Bu iki örneklemin çekilmiş olduğu anakütle ortalamalarının birbirlerine *eşit* oldukları,
- Bu iki anakütlenin *aynı anakütleden* çekilmiş birer rassal örneklem olduğu,
- İki örneklem ortalaması arasında gözlediğimiz farkın bir olasılık eseri olarak ortaya çıkmış, istatistik bakımından anlamlı olmayan, önemli olmayan küçük bir fark olduğu düşünülür.

2) $z_h>z_k$ olduğunda, H_o hipotezi *ret edilir* ve;

- Ho hipotezine ait olan düşüncenin tersi kabul edilir, yani H_1'i kabul edilir.
- Bu büyüklükteki z_h değerinin *olasılığa bağlı olarak* ortaya çıkmış olması olasılığı (ihtimali) çok düşüktür. Bu olasılık (p değeri) seçtiğimiz α değerinden de *küçüktür*. Bu

kadar küçük bir olasılıkla ortaya çıkan bu z değerini artık rastgeleliğe değil anakütlenin gerçekten farklı olması sonucuna varılır.

Summary:

1) Null Hypothesis (H0): An idea that is accepted as true until sufficient evidence is found to the contrary.
2) Alternative Hypothesis (H1): A hypothesis that is tested against the null hypothesis and is accepted when the null hypothesis is rejected.
3) One-Sided Opposing Hypothesis: An opposing hypothesis that includes all possible values in the population that are smaller or larger than a value determined by the null hypothesis for the parameter of interest in the population
4) Two-Sided Opposing Hypothesis: An opposing hypothesis that includes all possible values in the population other than the value determined by the null hypothesis for the parameter of interest in the population
5) Hypothesis Test Decision: A decision rule developed to lead the researcher to accept or reject the null hypothesis based on sample evidence
6) Type-1 Error: Rejection of the true hypothesis
7) Type-2 Error: Acceptance of the false hypothesis
8) Significance Level: Probability of rejecting the true null hypothesis, .
9) Power of the Test: Probability of rejecting a false null hypothesis, 1-α.

Özet:

1) Sıfır Hipotezi (H_0): Tersine yeterli kanıt bulununcaya kadar doğru kabul edilen fikirdir.
2) Alternatif Hipotez (H_1): Sıfır hipotezi karşısında test edilen, sıfır hipotezi red edildiğinde kabul edilen hipotez.
3) Tek Yanlı Karşıt Hipotez: Ana kütlede ilgilendiğimiz parametre için sıfır hipotezince, belirlenen bir değerden küçük ya da büyük olanaklı bütün değerleri içeren karşıt hipotez
4) Çift Yanlı Karşıt Hipotez: Ana Kütlede ilgilendiğimiz parametre için sıfır hipotezince belirlenen değer dışında olanaklı bütün değerleri içeren karşıt hipotez
5) Hipotez Testi Kararı: Araştırmacıyı örneklem kanıtına dayanarak, sıfır hipotezini kabul ya da reddetmeye götürecek şekilde geliştirilmiş karar kuralı
6) Tip-1 Hata: Doğru hipotezin red edilmesi
7) Tip-2 Hata: Yanlış hipotezin kabul edilmesi
8) Anlamlılık Düzeyi: Doğru olan sıfır hipotezini reddetme olasılığı, α.
9) Testin Gücü: Yanlış bir sıfır hipotezinin reddedilme olasılığı, 1-α.

Summary:

The weight average of the large data set and the acceptable standard deviation with a certain probability are known. The weight average and standard deviation of the sample data set are known. According to the probability distribution expression, we know the Zk values corresponding to the acceptable probability values. If it is two-sided (=) Zk=1.65 for 90%, Zk=1.96 for 95%, ZK=2.58 for 99%. The weight average of Zh or hypothesis is calculated. Is it not within the acceptable region?

If it is one-sided (=>, <=) 90%, Zk=1.28, Zk=1.65 for 95%, Zk=2.33 for 99%. The weight average of Zh or hypothesis is calculated. Is it within the acceptable region or not?

Büyük veri yığının ağırlık ortalaması ve belirli olasılıkla kabul edilebilir standart sapması biliniyor. Örneklem veri yığının ağırlık ortalaması ve standart sapması biliniyor. Olasılık dağılım ifadesine göre kabul edeilebilecek olasılık değerlerine karşılık gelen Zk değerlerini biliyoruz. Çift taraflı ise (=) %90 için Zk=1.65, %95 için Zk=1.96, %99 için ZK=2.58 alınıyor. Zh ya da hipotezin ağırlık ortalaması hesaplanır. Kabul edilebilir bölgesinin için de değil mi?

Tek taraflı ise (=>, <=) %90, Zk=1.28, %95 için Zk=1.65, %99 için Zk=2.33 alınır.Zh ya da hipotezin ağırlık ortalaması hesaplanır. Kabul edilebilir bölgesinin içinde mi değil mi?

Example:

In order to test a data set representing a system or a model, two parameters are needed;
1- weighted average,
2- variance or standard deviation.

Attention: When generating hypotheses, the number of samples should always be greater than 30.

Bir sistemi ya da bir modeli temsil eden veri yığınını test edebilmek için iki parametreye ihtiyaç var;

1- ağırlık ortalama,
2- varyans ya da standart sapma.

Dikkat: Hipotez üretilirken örnek sayısı her zaman 30'dan büyük olmalıdır.

In a clinic, a device that measures blood sugar levels breaks down while measuring an average of μo=100 mg/dl. Service is called and the device is repaired. Will it still be able to measure an average of μo=100mg/dl?

A trial was conducted and as a result of 64 sample measurements, the weight average was found to be 102.5 mg/dl and the standard deviation was found to be 16mg/dl.

Sample statistics are calculated:
n = 64 bags
Sample average: $\bar{X}$ = 102.5mg/dl
Sample standard deviation: σ = 16 mg/dl
Can this new production methodology be accepted within the range of 100 +/- σx?

Sample standard deviation error,
Bir klinikte kan şekeri ölçü testi yapan cihaz ortalama μ_o=100 mg/dl ölçüm yaparken arızalanır. Servis çağrılır cihaz tamir ettirilir. Acaba yine ortalama μ_o=100mg/dl ölçüm yapabilecek mir?

Deneme yapılıp 64 örneklem ölçüm sonucunda ağırlık ortalama 102.5 mg/dl , standart sapma 16mg/dl olarak bulunmuştur.

Örneklem istatistikleri hesaplanır:
n = 64 torba
Örneklem ortalaması : $\bar{X}$ = 102,5mg/dl
Örneklem standard sapması: σ = 16 mg/dl

Bu yeni üretim metotolojisni 100 +/- σ_x aralığında kabul edebilir mi?

Örneklem standart sapma hatası,

$$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{16}{\sqrt{64}} = \frac{16}{8} = 2mg/dl$$

$\bar{x} \pm \sigma_x$ = 102.5 ± 2mg/dl can the condition be met?

Determining the Statistical "Level of Significance" (Risk level, Margin of Error, Margin of Error) Determining α: Let α=0.05. Let the confidence level of the test = 1 - α = 0.95. The average of the population we are interested in is found as (Zk=1.96).

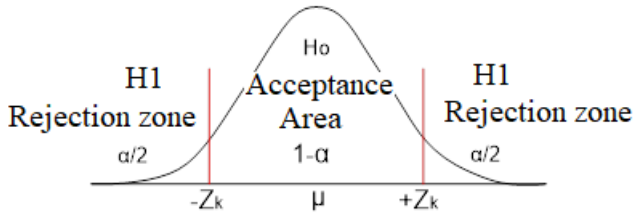
The Zh value of the sample we have is a statistic of the sample and is called the test statistic. Calculate the Zh value. If the Zh value you calculated is smaller than the Zk value, the hypothesis will be accepted. Comment on the Zh value you found.

İstatistiksel "Anlam Düzeyinin" belirlenmesi (Risk düzeyi, Yanılgı Payı, Hata payı) α'nın saptanması: α=0,05 olsun. Testin güven düzeyi = 1 - α = 0,95 olsun. İlgilendiğimiz anakütlenin ortalamasının (Zk=1.96) olarak bulunmuştur.

Elimizdeki örnekleme ait Z_h değeri örneklemin bir istatistiğidir ve test istatistiği adı verilir. Zh değerini hesaplaynız. Hesapladığınız Zh değeri Zk değerinden küçük ise hipotez Kabul edilecektir. Bulduğunuz Zh değerini yorumlayınız.

$$Z_h = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Z_h= (102,5-100)/2 = 1,25. The hypothesis is true.



Example:

A plaster filling machine breaks down while filling plaster with an average weight of µo=20 kg. A repairman is brought in and repaired. Will it be able to fill µo=20 kg again? A trial was conducted and 40 bags were selected according to the simple sampling method and their weights were measured as follows: X1 = 20.5 kg, X2= 21.2 kg, X3= 18.9 kg, ... , X40= 20.8 kg

Bir alçı dolum makinesi μ_o=20 kg ortalama ağırlıklı alçı dolumu yaparken arıza yapar. Tamirci getirip tamir ettirilir. Acaba yine μ_o=20kg'lık dolum yapabilecek midir? Deneme yapıp 40 torba basit örneklem yöntemine göre seçilip ağırlıkları şöyle ölçülmüştür: X_1 = 20,5 kg, X_2= 21,2 kg, X_3= 18,9 kg, ... , X_{40}= 20,8 kg

First, sample statistics are calculated:

n = 40 bags

Sample mean: $\bar{X}$ = 21.4 kg

Sample standard deviation: σ = 3.2 kg. Standardization is the expression representing the deviations in the weight mean.

Can I accept this new production methodology in the range of 20+/- delta (σ_x?

n = 40 torba

Örneklem ortalaması : $\bar{X}$ = 21,4 kg

Örneklem standard sapması: σ = 3,2 kg. Standart yapma, ağırlık ortalamada oluşan sapmaları temsil eden ifadedir.

Bu yeni üretim metotolojisni 20 +/- delta (σ_x) aralığında kabul edebilir miyim?

Sample standard deviation error,

Örneklem standart sapma hatası,

Delta= $\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{3.2}{\sqrt{40}} = 0.506kg$ is calculated.

$\bar{X} \pm \sigma_x$ = 21,4 ± 0,506 kg

Determination of hypotheses:

Ho Hypothesis: The sample we have is a random sample drawn from a population with a population mean of "µo = 20 kg", and the sample mean X- value can be accepted as equal to the population mean. The difference of 21.4-20 = 1.4 kg is a very small difference that can be attributed to chance, is not important, and does not carry any meaning. Therefore, we can write $\bar{X}$= μ_o.

Standart Normal Dağılım Tablosu

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Hipotezlerin belirlenmesi:
Ho Hipotezi: Elimizdeki örneklem anakütle ortalaması "μ_o = 20 kg" olan bir anakütleden çekilmiş bir rassal örneklem olup, örneklem ortalaması X- değeri anakütle ortalamasına eşit olarak kabul edilebilir. Aradaki 21.4-20=1,4 kg'lık fark ise tesadüfe bağlanabilecek, önemli olmayan, anlam taşımayan çok küçük bir farktır. Dolayısıyla $\overline{X}= \mu_o$ yazabiliriz.

H1 Hypothesis: This sample cannot be a random sample drawn from a population with "μo = 20 kg". The difference of 1.4 kg is not due to chance, it occurred because the adjustment was not made. The population from which this sample was drawn cannot be 20 kg. Our sample must have been drawn from another population of its own.
H1 Hipotezi: Bu örneklem " μ_o = 20 kg" olan bir anakütleden çekilmiş bir rassal örneklem olamaz. Aradaki 1,4 kg lık fark tesadüfe bağlı değil, ayarlamanın yapılmamış olması nedeni ile gerçekleşmiştir. Bu örneklemin çekilmiş olduğu anakütle 20 kg olamaz. Örneklemimiz kendine ait başka bir anakütleden çekilmiş olmalıdır.

Determining the Statistical "Level of Significance" (Risk level, Margin of Error, Margin of Error) Determining α: Since we cannot perform an error-free test, we have a certain risk of error in every test. Therefore, statisticians want to perform tests with as little error as possible. Nevertheless, α = 0.05 and α = 0.01 levels are frequently used. Let α = 0.05. Let the confidence level of the test = 1 - α = 0.90.

İstatistiksel "Anlam Düzeyinin" belirlenmesi (Risk düzeyi, Yanılgı Payı, Hata payı) α'nın saptanması: Hatasız bir test yapamayacağımız için her testte bir miktar yanılma riskimiz vardır. O nedenle istatistikçiler olabildiğince az yanılma ile test yapmak isterler. Yine de α =0,05 ve α=0,01 düzeyleri sık kullanılanır. α=0,05 olsun. Testin güven düzeyi = 1 - α = 0,90 olsun.

Test statistic:
We will conclude the hypothesis test with the help of the Zh value of the sample we have. Therefore, we call the Zh value the Test Statistic.

Test istatistiği:
Elimizdeki örnekleme ait Z_h değeri yardımıyla hipotez testini sonuçlandıracağız. O nedenle, Z_h değerine Test İstatistiği adını veriyoruz.

$$Z_h = \frac{\bar{x} - \mu}{\sigma_x}$$

Z_h= (21,4-20)/0,51 = 2,74

Comparison, result and comment:

According to this situation: the mean of the sample we have (Zh=2.74) is in a size that is very far from the mean of the population we are interested in (Zk=1.96). In this case, the Ho hypothesis, which we expressed as $\bar{x}$ = μo and raised to the level of μ=μo, cannot be accepted. So, this machine makes faulty fillings and cannot make fillings with an average of 20 kg.

If the μ+Zh value is smaller than the sample mean, the hypothesis is true. Then, is the hypothesis true in the given question?

Karşılaştırma, sonuç ve yorum:

Bu duruma göre: elimizdeki örneklemin ortalaması (Zh=2,74), ilgilendiğimiz anakütlenin ortalamasından (Zk=1.96) çok uzağa düşen bir büyüklüktedir. Bu durumda $\bar{x}$ = μ_o biçiminde ifade ettiğimiz ve μ=μ_o düzeyine yükselttiğim H_o hipotezini kabul edilemez. Demek ki, bu makine hatalı dolum yapmakta, ortalaması 20 kg olan dolumlar gerçekleştirememektedir.

μ+Zh değeri örneklem ortalamadan küçük ise hipotez doğrudur. O halde verilen soruda hipotez doğru mudur?

Example:

A plaster filling machine breaks down while filling plaster with an average weight of μo=20 kg. A repairman is brought in and repaired. Will it be able to fill μo=20 kg again? A trial was conducted and 40 bags were selected according to the simple sampling method and their weights were measured as follows: X1 = 19.8 kg, X1 = 20.5 kg, X2= 20.2 kg, X3= 19.9 kg, ... , X40= 20.8 kg

Bir alçı dolum makinesi μ_o=20 kg ortalama ağırlıklı alçı dolumu yaparken arıza yapar. Tamirci getirip tamir ettirilir. Acaba yine μ_o=20kg'lık dolum yapabilecek midir? Deneme yapıp 40 torba basit örneklem yöntemine göre seçilip ağırlıkları şöyle ölçülmüştür: X_1 = 19,8 kg, X_1 = 20,5 kg, X_2= 20,2 kg, X_3= 19,9 kg, ... , X_{40}= 20,8 kg

Sample statistics are calculated:

n = 40 bags

Sample mean: $\bar{X}$ = 20.4 kg

Sample standard deviation: σ = 3.2 kg

Can I accept this new production methodology in the range of 20+/- delta (σ_x)?

Sample standard error, Delta= $\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{3.2}{\sqrt{40}} = 0.506kg$ is calculated.

$\bar{x} \pm \sigma_x$ = 20,4 ± 0,506 kg

Örneklem istatistikleri hesaplanır:
n = 40 torba
Örneklem ortalaması : $\bar{X}$ = 20,4 kg
Örneklem standard sapması: σ = 3,2 kg
Bu yeni üretim metotolojisni 20+/- delta (σ_x) aralığında kabul edebilir miyim?

Örneklem standart hata, Delta= $\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{3.2}{\sqrt{40}} = 0.506kg$ hesaplanır.

$\bar{x} \pm \sigma_x$ = 20,4 ± 0,506 kg

Determination of hypotheses:
Ho Hypothesis: The sample we have is a random sample drawn from a population with a population mean of "μo = 20 kg", and the sample mean X- value can be accepted as equal to the population mean. The difference of 20.4-20 = 0.4 kg is a very small difference that can be attributed to chance, is not important, and does not carry any meaning. Therefore, we can write $\bar{X}$= μo.

H1 Hypothesis: This sample cannot be a random sample drawn from a population with "μo = 20 kg". The difference of 0.4 kg is not due to chance, it occurred because the adjustment was not made. The population from which this sample was drawn cannot be 20 kg. Our sample must have been drawn from another population of its own.

Determining the Statistical “Significance Level” (Risk Level, Margin of Error, Margin of Error) Determining α: Let α = 0.05. The confidence level of the test = 1 - α = 0.95.

Hipotezlerin belirlenmesi:
Ho Hipotezi: Elimizdeki örneklem anakütle ortalaması "μ_o = 20 kg" olan bir anakütleden çekilmiş bir rassal örneklem olup, örneklem ortalaması X- değeri anakütle ortalamasına eşit olarak kabul edilebilir. Aradaki 20.4-20=0,4 kg’lık fark ise tesadüfe bağlanabilecek, önemli olmayan, anlam taşımayan çok küçük bir farktır. Dolayısıyla $\bar{X}$= μ_o yazabiliriz.

H1 Hipotezi: Bu örneklem " μ_o = 20 kg" olan bir anakütleden çekilmiş bir rassal örneklem olamaz. Aradaki 0,4 kg lık fark tesadüfe bağlı değil, ayarlamanın yapılmamış olması nedeni ile gerçekleşmiştir. Bu örneklemin çekilmiş olduğu anakütle 20 kg olamaz. Örneklemimiz kendine ait başka bir anakütleden çekilmiş olmalıdır.

İstatistiksel “Anlam Düzeyinin” belirlenmesi (Risk düzeyi, Yanılgı Payı, Hata payı) α’nın saptanması: α=0,05 olsun. Testin güven düzeyi = 1 - α = 0,95 olur.

Test statistic:

The Zh value of the sample we have is a statistic of the sample. We will conclude the hypothesis test with the help of this statistic. Therefore, we call the Zh value the Test Statistic.

Test istatistiği:

Elimizdeki örnekleme ait Z_h değeri örneklemin bir istatistiğidir. Bu istatistik yardımıyla hipotez testini sonuçlandıracağız. O nedenle, Z_h değerine Test İstatistiği adını veriyoruz.

$$Z_h = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Z_h= (20,4-20)/0,51 = 0,8

Comparison, result and comment:

According to this situation: the mean of the sample we have (Zh=0.8) is in a size that is inside the mean of the population we are interested in (Zk=1.96). It has fallen into the inner area of the acceptance region with a 95% probability. In this case, the Ho hypothesis, which we have expressed as $\bar{x}$ = μo and raised to the level of μ=μo, can be accepted. So, this machine does not make faulty fillings, it makes fillings with an average of 20 kg. Therefore; the probability of our decision being correct is 95%, while the probability of it being wrong is at most 5%.

Karşılaştırma, sonuç ve yorum:

Bu duruma göre: elimizdeki örneklemin ortalaması (Zh=0.8), ilgilendiğimiz anakütlenin ortalamasının (Zk=1.96) iç tarafında olan bir büyüklüktedir. %95 olasılıkla kabul bölgesinin iç alanına düşmüştür. Bu durumda $\bar{x}$ = μ_o biçiminde ifade ettiğimiz ve μ=μ_o düzeyine yükselttiğim H_o hipotezini kabul edilebilir. Demek ki, bu makine hatalı dolum yapmamakta, ortalaması 20 kg olan dolumlar gerçekleştirmektedir. Dolayısıyla; verdiğimiz kararın doğru olması olasılığı %95 iken hatalı olması olasılığı en fazla %5 tir.

Example:

There is an investment plan that is thought to bring profit. An investor will invest only if he earns an average monthly income of over $180. He has a sample of 300 monthly returns with an average of $190 and a standard deviation of $75. Should he invest according to this plan? α = 0.05 (i.e. 5% significance level)

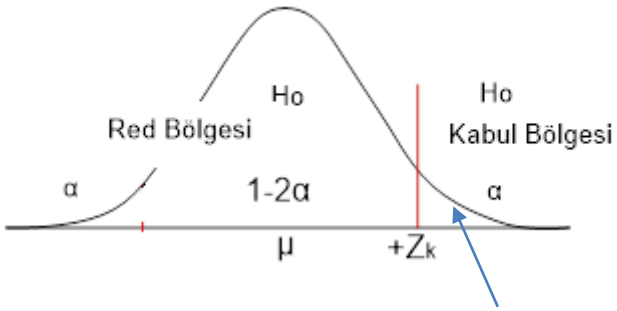
H0: Null Hypothesis, mean ≥ 180
H1: Alternative Hypothesis, mean< 180

In one-sided tests, the area determining the confidence limit depends on the direction of the test and is only at one end of the normal curve.

Kar getireceği düşünülen yatırım planı mevcuttur. Bir yatırımcı, ancak ortalama 180$'ın üstünde aylık gelir elde ederse yatırım yapacaktır. Ortalaması 190$ ve standart sapması 75$ olan 300 aylık getiri örneğine sahiptir. Bu plana göre yatırım yapmalı mı? α = 0,05 (yani% 5 anlamlılık düzeyi)

H0: Boş Hipotez, ortalama ≥ 180
H1: Alternatif Hipotez, ortalama< 180

Tek taraflı testlerde güven sınırını belirleyen alan, testin yönüne bağlı ve normal eğrinin sadece bir ucundadır.



One-sided test, acceptance and rejection regions according to 5% significance level: Z_k = +1.65, α=0.05 (%95)

$$1.65 = \frac{\bar{X}_h - 180}{75/\sqrt{300}}$$

$\bar{x}_h = 1.65 * \frac{75}{\sqrt{300}}$ + 180 = 187.12 is found.

Since the sample mean of the business, $\bar{x}_h$ =187.12 is greater than the critical value of 180, the null hypothesis is accepted and the result is that the average monthly return is indeed more than $180, so the investor can consider investing in this business.

İşin örneklem ortalaması, $\bar{x}_h$ =187.12 değeri 180 kritik değerden büyük olduğu için, sıfır hipotezi kabul edilir ve sonuç, ortalama aylık getirinin gerçekten 180$ 'dan fazla olduğu, bu nedenle yatırımcı bu işe yatırım yapmayı düşünebilir.

Method-2:
Standardized test statistics and standardized Z value can be used.

$$\mathbf{Z_h} = \frac{\bar{X} - \mu}{s_{\bar{x}}} = \frac{190 - 180}{75/\sqrt{300}} = 2.309$$

Since Z = 2.309 is greater than Zk = 1.645 with 90% probability, it is within the acceptance region and is accepted with a similar result.
Z = 2.309, %90 olasılıklı Z_k = 1.645'ten büyük olduğundan, kabul bölgesinin içerisindedir, benzer bir sonuçla kabul edilir.

Example:

It is claimed that the average weight of bread produced by a bakery is 500 grams. Municipal authorities who inspected the bakery found the average weight of 100 samples to be 490 grams and the standard error to be 30 grams. Test whether the average weight of the bread is 500 grams at a 1% significance level (99% confidence interval).

Bir fırının ürettiği ekmek ortalama ağırlığı 500 gram olduğu iddia edilmektedir. Fırını denetleyen belediye yetkilileri 100 adet örneğin ortalama ağırlığını 490 gram ve standart hatasını 30 gram bulmuşlardır. %1 anlam düzeyinde (%99 güven aralığında) ekmeğin ortalama ağırlığı 500 gram kabul edilebilir mi, test ediniz.

Let's set up the hypothesis test:

H_0 : μ = 500 gr.

H_1 : μ ≠ 500 gr.

Based on the sample, the Zh value is calculated. If the standard deviation error, Sx is given

$\boldsymbol{\sigma_x = \frac{S_x}{\sqrt{n}}}$ =30/10=3

Örneklemden yola çıkılarak Z_h degeri hesaplanır.

Standart sapma hatası, S_x verilmiş ise $\boldsymbol{\sigma_x = \frac{S_x}{\sqrt{n}}}$ =30/10=3

$$\mathbf{Z_h = \frac{X - \mu}{\sigma_{\bar{x}}} = \frac{490 - 500}{3} = -3.33}$$

At the 99% confidence interval, P(Zα/2) = 0.99/2 = 0.495 is obtained. The Zα/2 value that meets the value P(Zα/2) = 0.495 is determined from the standard normal distribution table. Zk = Zα/2 = 2.58.

%99 güven aralığındaki, $P(Z_{\alpha/2})$ = 0.99/2 = 0.495 elde edilir. Standart normal dağılım tablosundan $P(Z_{\alpha/2})$= 0.495 değerini karşılayan $Z_{\alpha/2}$ değeri belirlenir. $Z_k = Z_{\alpha/2}$ = 2.58 dir.

(- Zk =-2.58, +Zk =+2.58), the symmetry property of Z values is taken into account.

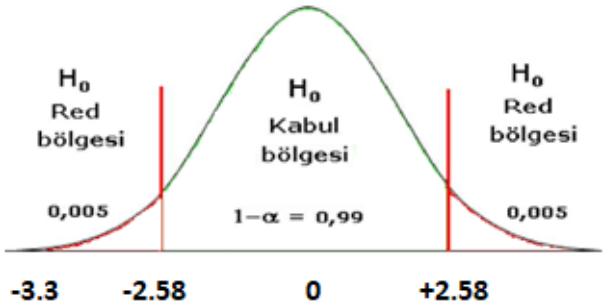
Since Zh =-3.3 > Zk=2.58, H0 hypothesis is rejected and the alternative hypothesis is accepted.

Comment: The average weight of the breads produced by the bakery is different from 500 grams.

(- Z_k =-2,58, +Z_k =+2.58), Z değerlerinin simetri özelliği göz önüne alınır.

Z_h =-3.3 > Z_k=2.58 oldugundan H_0 hipotezi reddedilir ve alternatif hipotez kabul edilir.

Yorum: Fırının ürettiği ekmeklerin ortalama ağırlığı 500 gramdan farklıdır.



Example:

It is claimed that a painkiller will take effect in less than 60 minutes on average. 64 randomly selected patients are given the drug and the mean duration of effect is X=63 minutes and the standard error is 12. Test the accuracy of the claim at the α = 0.05 level of significance (95% confidence interval).

Ağrı kesici bir ilacın ortalama 60 dakikadan daha az bir sürede etkisini göstereceği iddia ediliyor. Rasgele seçilen hastalardan 64'üne ilgili ilaç veriliyor ve ortalama etki süresi 63 dakika ve standart hatası 12 bulunuyor. α=0.05 anlamlılık düzeyinde (%95 güven aralığı) iddianın doğruluğunu test ediniz.

Let's set up the hypothesis test:

H0 : $\mu \leq 60$ minutes

H1 : $\mu > 60$ minutes

Hipotez testini kuralım:

$H_0 : \mu \leq 60$ dakika

$H_1 : \mu > 60$ dakika

Method: With 90% probability, the interval estimate of the population mean will be made. α = 0.05 (i.e. 5% significance level) corresponds to 90%. Due to symmetry, the middle region corresponds to 90% α values, and the extremes correspond to 5% α values.

Yöntem: % 90 olasılıkla, ana kütlenin ortalamasının aralık tahmini yapılacaktır. α = 0,05 (yani% 5 anlamlılık düzeyi), %90'a karşılık gelir. Simetriden dolayı orta bölge %90, uçlar % 5'lik α değerlerine karşılık gelmektedir.

P(Zα)=0.90/2=0.45 is obtained. The Zα value that meets the P(Zα)=0.45 value is determined from the standard normal distribution table. Zα=1.65.

$P(Z_\alpha)=0.90/2=0.45$ elde edilir. Standart normal dağılım tablosundan $P(Z_\alpha)= 0.45$ değerini karşılayan Z_α değeri belirlenir. $Z_\alpha=1.65$ dir.

If the standard deviation error, Sx is given $\boldsymbol{\sigma_x} = \frac{S_x}{\sqrt{n}}$ = 12/8 = 1.5

$$1.65 = \frac{Xh-60}{1.5}$$

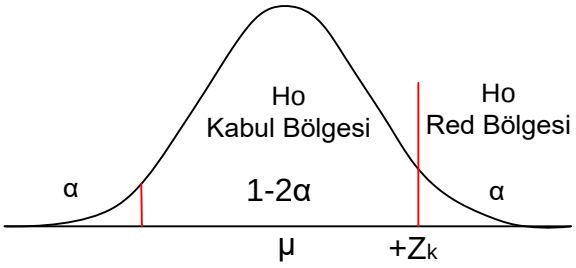
$Xh = 1.65 * 1.5 + 60$ = 62.475 is found.

Since the sample mean (Xk=60) is not greater than the critical value (Xh=62.475) given as hypothesis, H0 hypothesis can be rejected.

Örneklem ortalaması (Xk=60) hipotez olarak verilen kritik değer (Xh=62.475) örneklem ortalamasından küçük olduğu için, H_0 hipotezi reddedilir.

Yorum: %95 güven düzeyinde ilaç ağrıyı en geç 63 dakika içinde geçirmektedir.

Method-2: Zh= (63-60)/1.5=2 Zh değeri Zαdeğerinden büyük olduğu için hipotez ret edilir.



Example:

It is claimed that a known diet can help you lose at least 10 kilos in 3 months. It has been determined that the diet applied to 144 people for 3 months resulted in an average weight loss of 9 kilos and the standard error was 3 kilos. Test the accuracy of the claim with a 99% confidence interval.

Bilinen bir diyet ile 3 ayda en az 10 kilo verildiği iddia edilmektedir. 144 kişi üzerinde 3 ay uygulanan diyetin ortalama 9 kilo zayıflattığı ve standart hatanın 3 kilo olduğu tespit edilmiştir. %99 güven aralığında iddianın doğruluğunu test ediniz.

Let's set up the hypothesis test.

H_0 : μ ≥ 10 kilo

H_0 : μ < 10 kilo

Based on the sample, the Zh value is calculated.

If the standard deviation error, Sx is given $\boldsymbol{\sigma_x} = \frac{S_x}{\sqrt{n}}$ = 4/12 = 1/3

$$Z_h = \frac{X - \mu}{\sigma_{\bar{x}}} = \frac{9 - 10}{1/3} = \frac{-1}{1/3} = -3$$

In the 99% confidence interval, P(Zα/2) = (100 -2α)/2= 0.98/2 = 0.49 is obtained. (Probability=0.49 +0.5=0.99). The Zα/2 value that meets the value P(Zα/2)= 0.49 is determined from the standard normal distribution table. Zk = Zα/2 = 2.33.

The calculated Zh value is compared with the Zk value.

Since Zh = -3 < Zk = -2.33, the H0 hypothesis is rejected.

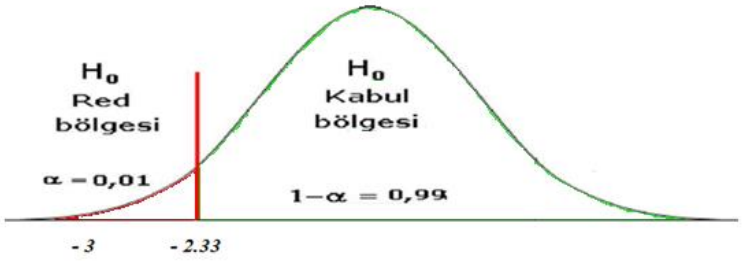
Since Xh=Zh*σx +μ=10+1=11, Xh is greater than the μ value, the hypothesis is rejected.

%99 güven aralığındaki, $P(Z_{\alpha/2})$ = (100 -2α)/2= 0.98/2 = 0.49 elde edilir. (Olasılığı=0.49 +0.5=0.99). Standart normal dağılım tablosundan $P(Z_{\alpha/2})$= 0.49 değerini karşılayan $Z_{\alpha/2}$ değeri belirlenir. $Z_k = Z_{\alpha/2}$ = 2.33 dir.

Hesaplanan Z_h değeri Z_k değeri ile karşılaştırılır.

Z_h = -3 < Z_k = -2.33 olduğundan H_0 hipotezi reddedilir.

Xh=Zh*σx +μ=10+1=11, Xh, μdeğerinden büyük olduğundan hipotez ret edilir.


H_0
Red
bölgesi
$\alpha = 0,01$
H_0
Kabul
bölgesi
$1-\alpha = 0,99$
-3
-2.33

Example:
It is predicted that the monthly water consumption in a town will be at least 20 liters. If this prediction is correct, the town may face a water problem. For this purpose, data from 1000 randomly selected people were collected and it was determined that the monthly average water consumption was 22 liters and the standard deviation was 8 liters. Decide whether there will be a water problem using the significance level of α = 0.01 (99% confidence level).

Bir kasabada aylık su tüketimini en az 20 litre olabileceği öngörülmektedir. Eğer bu öngörü doğruysa kasaba su sorunuyla karşı karşıya kalabilir. Bu amaçla, rassal olarak seçilen, 1000 kişiden veriler derlenmiş ve aylık ortalama su tüketiminin 22 litre ve standart sapmasının da 8 litre olduğu tespit edilmiştir. α = 0.01 anlam düzeyini (% 99 güven düzeyi) kullanarak su sorunu olup olmayacağına karar veriniz.

Let's set up the hypothesis test.
H_0 : μ = 20 litre
H_1 : μ > 20 litre
n = 1000
$\bar{X}$ = 22 litre
μ = 20 litre

Since n>30, it has a normal distribution. The Z value is found as 2.33 for a 99% confidence level from the table. After this information, the solution is reached in 2 stages.
n > 30 olduğu için normal dağılıma sahiptir. Z değeri tablodan % 99 güven düzeyi için 2.33 bulunur. Bu bilgilerden sonra 2 aşama ile çözüme ulaşılır.

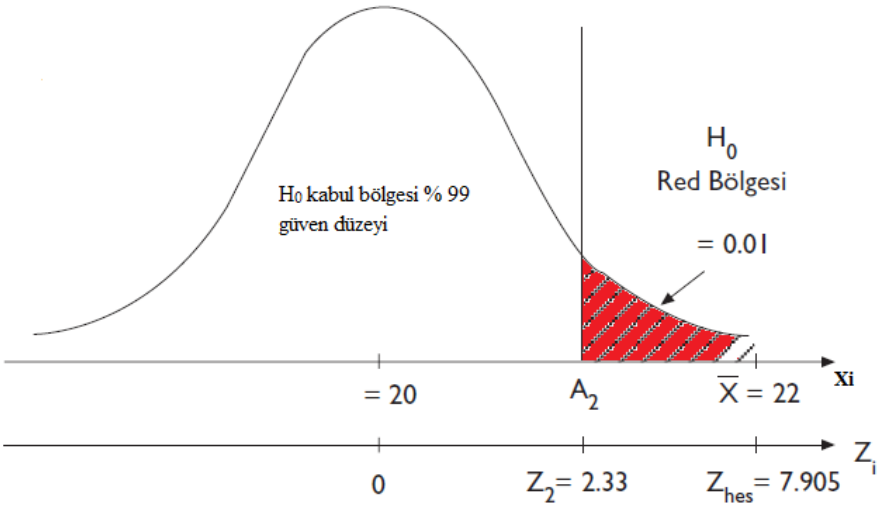
Step 1: The Zh value is calculated.

$$\mathbf{s_{\bar{x}}} = \frac{\mathbf{s}}{\sqrt{n}} = \frac{\mathbf{8}}{\sqrt{\mathbf{1000}}} = \mathbf{0.253}$$

$$\mathbf{Z_h} = \frac{\bar{X} - \mu}{\mathbf{s_{\bar{x}}}} = \frac{\mathbf{22 - 20}}{\mathbf{0.253}} = \mathbf{7.9}$$

Step -2: The calculated Zh value is compared with the Zk value.
Since Zh (7.9) > Zk (2.33), the H0 hypothesis is rejected and the alternative hypothesis is accepted.
Comment: The monthly water consumption of individuals is more than 20 liters.

Adım -2: Hesaplanan Z_h değeri Z_k değeri ile karşılaştırılır.
Z_h (7.9) > Z_k (2.33) oldugundan H_0 hipotezi reddedilir ve alternatif hipotez kabul edilir.
Yorum: Kişilerin aylık su tüketimi 20 litreden fazladır.



Example:

A chocolate company plans to produce 500 gram packages. To check whether the production is carried out as planned, the average weight of 100 randomly selected packages is found to be 495 grams and the standard deviation is 20 grams. Decide whether the production is carried out as planned using the significance level of α = 0.05 (95% confidence level).

Bir çikolata firması 500 gramlık paketler halinde üretim yapmayı planlamaktadır. Üretimin planlandığı gibi gerçekleşip gerçekleşmediğini kontrol etmek için rassal olarak seçilen 100 paketin ortalama ağırlığı 495 gram ve standart sapma da 20 gram olarak bulunmuştur. Üretimin planlandığı gibi gerçekleşip gerçekleşmediğini α = 0.05 anlam düzeyini (% 95 güven düzeyi) kullanarak karar veriniz.

Hypothesis is symmetric

H_0 : μ = 500 gram

H_1 : μ ≠ 500 gram

n = 100

$\bar{X}$ = 495 gram

μ = 500 gram

s = 20 gram

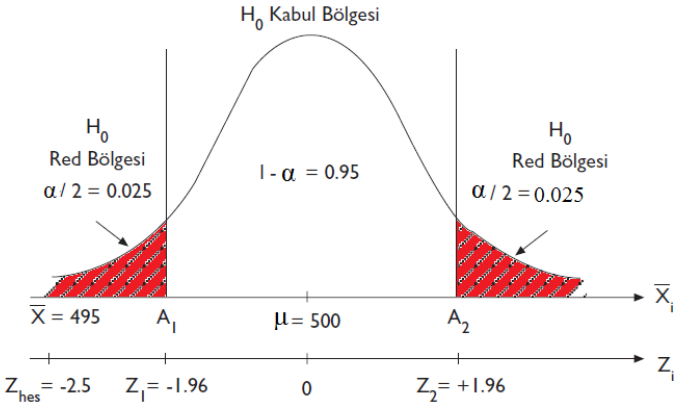
Since n>30, it has a normal distribution. The Z value is found from the table as Zk=1.96 for a 95% confidence level. The Zh value is calculated based on the sample.

n > 30 olduğu için normal dağılıma sahiptir. Z değeri tablodan % 95 güven düzeyi için Zk=1.96 bulunur. Örneklemden yola çıkılarak Z_h degeri hesaplanır.

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{20}{\sqrt{100}} = 2$$

$$Z_h = \frac{\bar{X} - \mu}{s_{\bar{x}}} = \frac{495 - 500}{2} = -2.5$$

The calculated Zh value is compared with the Zk value.
Since Zh (2.5) > Zk (1.96), the H0 hypothesis is rejected and the alternative hypothesis is accepted.
The weight of the packages is different from 500 grams.
Hesaplanan Z_h değeri Z_k değeri ile karşılaştırılır.
Z_h (2.5) > Z_k (1.96) oldugundan H_0 hipotezi reddedilir ve alternatif hipotez kabul edilir.
Paketlerin ağırlığı 500 gramdan farklıdır.



7.4.1. Hypothesis Testing in Data Analytics

Everyone who specializes in Data Analytics hears about hypothesis testing. Hypothesis testing is a statistical method used to make statistical decisions. Hypothesis testing is basically testing the accuracy of an assumption we make about a data set parameter.

Data Analitiği konusunda uzmanlaşan herkes hipotez testinden bahsedildiğini duyar. Hipotez testi, istatistiksel kararlar vermede kullanılan istatistiksel bir yöntemdir. Hipotez Testi temel olarak veri yığını parametresi hakkında yaptığımız bir varsayım doğruluğunun test edilmesidir.

En iyi ortalama performansa sahip algoritmanın, daha kötü ortalama performansa sahip algoritmalardan daha iyi olması beklenir. Peki ya ortalama performanstaki fark istatistiksel bir tesadüften kaynaklanıyorsa? Çözüm, herhangi iki algoritma arasındaki ortalama performans farkının gerçek olup olmadığını değerlendirmek için istatistiksel bir hipotez testi kullanmaktır.

Ortalama model performansına göre model seçimi yapmak yanıltıcı olabilir. Bir hipotez testi, hangi ifadenin örnek veriler tarafından en iyi şekilde desteklendiğini belirlemek için bir popülasyon hakkında birbirini dışlayan iki ifadeyi değerlendirir. Bir bulgunun istatistiksel olarak anlamlı olduğunu söylediğimizde, bu bir hipotez testi sayesindedir.

Modele güvenmek ve tahminlerde bulunmak için hipotez testi kullanılır. Modeli eğitmek için örnek veriler kullanılacağı zaman, popülasyon hakkında varsayımlarda bulunulur. Hipotez testi yapılarak, bu varsayımlar istenen bir önem düzeyi için doğrulanır.

Data Analitiği modelleri, genellikle k-kat çapraz doğrulama kullanılarak hesaplanan ortalama performanslarına göre seçilir. En iyi ortalama performansa sahip algoritmanın, daha kötü ortalama performansa sahip algoritmalardan daha iyi olması beklenir. Peki ya ortalama performanstaki fark istatistiksel bir tesadüften kaynaklanıyorsa? Çözüm, herhangi iki algoritma arasındaki ortalama performans farkının gerçek olup olmadığını değerlendirmek için istatistiksel bir hipotez testi kullanmaktır.

Model seçimi, bir dizi farklı Data Analitiği algoritmasını değerlendirmeyi veya işlem hatlarını modellemeyi ve performanslarına göre karşılaştırmayı içerir. Performans metriğine göre en iyi performansı elde eden model veya modelleme hattı, daha sonra yeni veriler üzerinde tahminler yapmaya başlamak için kullanabilecek son model olarak seçilir. Bu, klasik Data Analitiği algoritmaları ve derin öğrenme ile regresyon ve sınıflandırma tahmine dayalı modelleme görevleri için geçerlidir.

Kusursuz olmasa da, istatistiksel hipotez testi, model seçimi sırasında hem yorumlamaya hem de sonuçların sunumuna olan güveni artırabilir. İstatistiksel hipotez testleri, Data Analitiği modellerini karşılaştırmaya ve nihai bir model seçmeye yardımcı olabilir. İstatistiksel hipotez testlerinin safça uygulanması yanıltıcı sonuçlara yol açabilir. İstatistiksel testlerin doğru kullanımı zordur ve McNemar testini veya değiştirilmiş eşleştirilmiş Student t testi ile 5×2 çapraz doğrulamayı kullanmak için bazı fikir birliği vardır.

Data Analitiği modellerini istatistiksel anlamlılık testleri aracılığıyla karşılaştırmak, kullanılabilecek istatistiksel test türlerini etkileyecek bazı beklentiler getirir; örneğin:

Beceri Tahmini: Model becerisinin belirli bir ölçüsü seçilmelidir. Bu, kullanılabilecek testlerin türünü sınırlayacak sınıflandırma doğruluğu (bir orantı) veya ortalama mutlak hata (özet istatistik) olabilir.
Tekrarlanan Tahminler: İstatistikleri hesaplamak için bir beceri puanı örneği gereklidir. Belirli bir modelin aynı veya farklı veriler üzerinde tekrarlanan eğitimi ve test edilmesi, kullanılabilecek test türünü etkileyecektir.
Tahminlerin Dağılımı: Beceri puanı tahminleri örneğinin dağılımı, belki Gauss ya da olmayabilir. Bu, parametrik veya parametrik olmayan testlerin kullanılıp kullanılamayacağını belirleyecektir.
Merkezi Eğilim: Model beceri genellikle, beceri puanlarının dağılımına bağlı olarak ortalama veya medyan gibi bir özet istatistik kullanılarak açıklanacak ve karşılaştırılacaktır. Test bunu doğrudan hesaba katabilir veya almayabilir.

İstatistiksel bir testin sonuçları genellikle bir test istatistiği ve bir p değeridir ve her ikisi de modeller arasındaki farkın güven veya anlamlılık düzeyini ölçmek için sonuçların sunumunda yorumlanabilir ve kullanılabilir. Bu, istatistiksel hipotez testleri kullanmamaktan ziyade model seçiminin bir parçası olarak daha güçlü iddialarda bulunulmasına izin verir. Model seçiminin bir parçası olarak istatistiksel hipotez testlerinin kullanılmasının istendiği göz önüne alındığında, özel kullanım durumunuza uygun bir testi seçilmelidir.

Regresyon modellerini ele alalım: Doğrusal bir regresyon modeli aracılığıyla düz bir çizgi uydurulduğunda, doğrunun eğimi ve kesişimi elde edilir. Bir lineer regresyon modelinde beta katsayılarının anlamlı olup olmadığını doğrulamak için hipotez testi kullanılır. Doğrusal regresyon modeli her çalıştırıldığında, katsayının anlamlı olup olmadığı kontrol edilerek doğrunun anlamlı olup olmadığı test edilir.

Hipotez testi gerçekleştirmek için temel adımlar aşağıdaki gibidir:

- Bir hipotez formüle edilir.
- Önem düzeyi belirlenir.
- Testin türünü belirlenir.

- Teste göre istatistik değerleri ve p değerleri hesaplanır.
- Karar verilir.

Hipotez testinin türünü seçimi:
Tahmin değişkenine göre test istatistiği türü seçilir: nicel veya kategorik. Aşağıda nicel veriler için yaygın olarak kullanılan test istatistiklerinden birkaçı verilmiştir.

Tahmin değişkeni tipi	Dağılım tipi	İstenen Test	Nitelikler
Nicel	Normal dağılım	Z – Test	• Büyük numune boyutu • Bilinen popülasyon standart sapması
Nicel	T Dağılımı	T-Test	• Örnek boyutu 30'dan az • Popülasyon standart sapması bilinmiyor
Nicel	Pozitif çarpık dağılım	F – Test	• 3 veya daha fazla değişkeni karşılaştırmak istediğinizde
Nicel	Negatif çarpık dağılım	NA	• Bir hipotez testi gerçekleştirmek için özellik dönüşümü gerektirir
Kategorik	NA	Chi-Square test	• Bağımsızlık testi • Formda olmanın güzelliği

Z-istatistiği – Z Testi:
Örnek normal bir dağılım istendiğinde Z-istatistiği kullanılır. Ortalama ve standart sapma gibi veri yığını parametrelerine göre hesaplanır. Bir örneklem ortalamasını bir büyük veri yığını ortalaması ile karşılaştırmak istediğimizde bir örnek Z testi kullanılır. İki örneğin ortalamasını karşılaştırmak istediğimizde iki örnek Z testi kullanılır.

T-istatistiği – T-Testi:
Örnek bir T dağılımını takip ettiğinde ve büyük veri yığını parametreleri bilinmediğinde T istatistiği kullanılır. T dağılımı normal dağılıma benzer, normal dağılımdan daha kısadır ve daha düz bir kuyruğa sahiptir.

F-istatistiği – F testi:
Üç veya daha fazla grup içeren numuneler için F Testini tercih edilir. Birden fazla grup üzerinde t testi yapmak, Tip-1 hata olasılığını artırır. Bu gibi durumlarda ANOVA kullanılır.

Varyans analizi (ANOVA), üç veya daha fazla grubun ortalamalarının farklı olup olmadığını belirleyebilir. ANOVA, araçların eşitliğini istatistiksel olarak test etmek için F-testlerini kullanır.

F-istatistiği, veriler pozitif olarak çarpık olduğunda ve bir F dağılımını takip ettiğinde kullanılır. F dağılımları her zaman pozitiftir ve sağa çarpıktır.

F = Numune ortalamaları arasındaki varyasyon/numuneler içindeki varyasyon
Negatif çarpık veriler için özellik dönüşümü yapmamız gerekir

Ki-kare testi:
Kategorik değişkenler için ki-kare testi yapıyor olacağız.

İki tür ki-kare testi şunlardır:
Ki-kare bağımsızlık testi – İki kategorik değişken arasında anlamlı bir ilişki olup olmadığını belirlemek için Ki-Kare testini kullanırız.
Ki-kare Uyum iyiliği, örnek verilerin popülasyonu doğru bir şekilde temsil edip etmediğini belirlememize yardımcı olur.

Hipotezin temeli normalleştirme ve standart normalleştirmedir, tüm hipotez bu 2 terimin temeli etrafında dönüyor.

Normal dağılım:
Bir değişkenin dağılımı normal bir eğri - özel bir çan şeklindeki eğri - şeklindeyse, bir değişkenin normal olarak dağıldığı veya normal bir dağılıma sahip olduğu söylenir. Sayısal değerleri normalize etmede kullanılmaktadır. Normal dağılımın grafiği, aşağıdaki özelliklerin tümüne sahip olan normal eğri olarak adlandırılır: Ortalama, medyan ve mod.

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Standartlaştırılmış Normal Dağılım:
Standart normal dağılım, ortalaması 0 ve standart sapması olan normal bir dağılımdır.

$$x_{new} = \frac{x - \mu}{\sigma}$$

7.4.2. T-Testi

Normal dağılım gösteren veri yığını parametrelerine dayalı hipotezlerin, Normal dağılım gösteren veri yığınından alınan n birimlik örneklerin ortalamaları ve varyansları kullanılarak test edilmesini sağlayan bir yöntemdir. Bir T-Testi, iki grubun ortalamalarını karşılaştırarak ve sonuçların tesadüfen olma olasılığını bulur.

T test istatistiği test modeline göre hesaplanan serbestlik derecesine (sd) göre farklı t dağılımı gösterir. Üç Farklı türü bulunur:

- Tek örneklem t test modellerinde sd=n-1,
- İki örneklem t test modellerinde sd=n1+n2-2
- Bağımlı iki örneklem t test modellerinde sd=n-1

T-Değerleri ve P-değerleri

Her t-değerinin onunla birlikte hareket eden bir p-değeri vardır. Bir p değeri, örnek verilerinizden elde edilen sonuçların tesadüfen meydana gelme olasılığıdır. Buraki p değerleri 0 ila 1 arasındadır. Tüm p değerleri toplamı 1'e eşittir. Genellikle ondalık olarak yazılırlar. Örneğin, %5'lik bir p değeri 0,05'tir. Düşük p değerleri iyidir; Verilerinizin tesadüfen oluşmadığını belirtirler. Örneğin, .01'lik bir p değeri, bir deneyden elde edilen sonuçların tesadüfen meydana gelme olasılığının yalnızca %1 olduğu anlamına gelir. Çoğu durumda, verilerin geçerli olduğu anlamına gelen 0,05 (%5) p değeri kabul edilir.

Karar verme aşamasında Test istatistiği ve sd hesaplanır. Sd parametreli teorik t dağılımının α yanılma payına göre kritik değerleri belirlenir [t(α,sd)]

İkiyönlü hipotez test sonuçlarına göre;

- |t| < t(0.05,sd) P>0.05 Önemsiz
- t(0.05,sd) <= |t| < t(0.01,sd) P<0.05 Önemli
- t(0.01,sd) <= |t| < t(0.001,sd) P<0.01 Önemli
- t(0.001,sd) <= |t| P<0.001 Önemli

Tek veri yığını Ortalamasına Dayalı Hipotezlerin Testi:

$$H_0 : \mu = \mu_0 \quad H_1 : \mu \neq \mu_0 \quad t = \frac{(\bar{X} - \mu_0) * \sqrt{n}}{S}$$

Örnek:

Çalışan işçilerde iş kazası geçirme yaş ortalaması=35.5 standart sapması=11.3 dür. Bir hastanede iş kazası nedeniyle muayene edilen 40 kişinin yaş ortalaması= 28.0 standart sapması= 5.9 dur. Bu hastaneye yansıyan iş kazaları yaş dağılımı açısından çalışan işçileri yansıtmakta mıdır?

Ortalamalar Arası Farka Ait Hipotez Testi – Student t-test

Birbirinden bağımsız iki veri yığını ortalamasını karşılaştırırken bu hipotez testinden yararlanılır. Örneğin A ve B gibi iki veri yığını çeşidinin verimi karşılaştırılmak istendiğinde bu hipotez testinden yararlanılır. Bu hipotez testlerinde hipotezler aşağıdaki şekilde kurulur ve kullanılacak istatistiklerle ilgili kriterler aşağıdaki gibidir.

a) $H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$ } Çift Yönlü hipotez

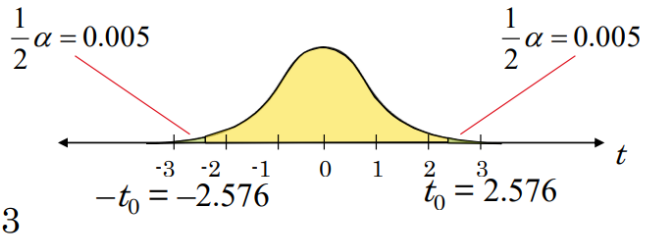
b) $H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 < \mu_2$
c) $H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 > \mu_2$ } Tek Yönlü hipotez

Örnek:

Brownsville'de rasgele seçilen 17 polis memurunun yıllık ortalama geliri 35.800$ ve standart sapması 7.800$'dır. Greensville'de, 18 polis memurundan oluşan rasgele bir örneklemin yıllık ortalama geliri 35.100$ ve standart sapması 7.375$'dır. İki şehirdeki yıllık ortalama gelirlerin aynı olmadığını α = 0.01 anlam düzeyinde test edin. Kitle varyanslarının eşit olduğunu varsayalım.

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$ (İddia)



$$\text{d.f.} = n_1 + n_2 - 2$$
$$= 17 + 18 - 2 = 33$$

Standart hata

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

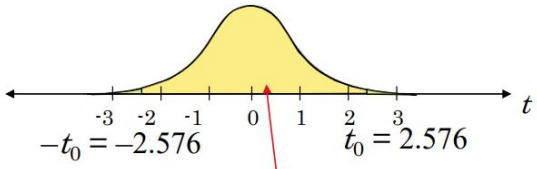
$$= \sqrt{\frac{(17-1)7800^2 + (18-1)7375^2}{17+18-2}} \cdot \sqrt{\frac{1}{17} + \frac{1}{18}}$$

$$\approx 7584.0355(0.3382)$$

$$\approx 2564.92$$

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$ (iddia)



standartlaştırılmış test istatistiği

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{(35800 - 35100) - 0}{2564.92} \approx 0.273$$

H_0 reddedilemez

Ortalama yıllık gelirlerin farklılık gösterdiği iddiasını desteklemek için% 1 düzeyinde yeterli kanıt yoktur.

Örnek:

Add up all of the squared differences

Subject #	Score 1	Score 2	X-Y	(X-Y)^2
1	3	20	-17	289
2	3	13	-10	100
3	3	13	-10	100
4	12	20	-8	64
5	15	29	-14	196
6	16	32	-16	256
7	17	23	-6	36
8	19	20	-1	1
9	23	25	-2	4
10	24	15	9	81
11	32	30	2	4
		SUM:	**-73**	**1131**

$$D = \sum X - Y$$

$$t = \frac{(\sum D)/N}{\sqrt{\frac{\sum D^2 - \left(\frac{(\sum D)^2}{N}\right)}{(N-1)(N)}}}$$

$$t = \frac{-73/11}{\sqrt{\frac{1131 - \left(\frac{(-73)^2}{11}\right)}{(11-1)(11)}}}$$

$$t = \frac{-73/11}{\sqrt{\frac{1131 - \left(\frac{5329}{11}\right)}{110}}}$$

$$t = -\ 2.74$$

Serbestlik derecelerini elde etmek için örneklem boyutundan 1 çıkarın. 11 öğemiz var, yani 11-1 = 10.

Serbestlik derecelerini kullanarak t-tablosunda p değerini bulun. Belirtilen bir alfa seviyeniz yoksa, 0,05 (%5) kullanın. Bu örnek problem için, df = 10 ile t değeri 2.228'dir.

(2.228) t-tablo değerinizi hesapladığınız t-değeriniz (-2.74) ile karşılaştırın. Hesaplanan t değeri, .05'lik bir alfa düzeyinde tablo değerinden daha büyüktür. p değeri, alfa seviyesinden küçüktür: p <.05. Ortalamalar arasında bir fark olmadığı sıfır hipotezini reddedebiliriz.

7.4.3. Chi-Square Bağımsızlık Testi

İki kategorik değişkenin bağımsızlığı testi Chi-kare bağımsızlık testi kullanılarak gerçekleştirilir.

r = satır sayısı, c = sütun sayısı olmak üzere r × c beklenmedik durum tablosu olarak da adlandırılan iki yönlü bir tabloda iki kategorik değişken özetleyebilir. İlgilenilen soru "İki değişken bağımsız mı?"

Sıfır hipotezi: İki kategorik değişken bağımsızdır
Alternatif hipotez: İki kategorik değişken bağımlıdır

Chi-Kare Testi İstatistik

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Burada O, gözlemlenen frekansı temsil eder. E beklenen değer, sıfır hipotezi altında beklenen frekanstır ve şu şekilde hesaplanır:

$$E = \frac{(Satır\ Toplamı) * (Sütun\ Toplamı)}{Örnek\ Boyutu}$$

Serbestlik derecesi = (r - 1) (c - 1) olan X_α^2 kritik değeriyle test istatistiğinin değeri, karşılaştırılır ve eğer $X^2 > X_\alpha^2$ varsa boş hipotez reddedilir.

Örnek:
Cinsiyet eğitim seviyesinden bağımsız mıdır? Rastgele 395 kişiden oluşan bir örneklemle anket yapıldı ve her bir kişiden aldıkları en yüksek eğitim düzeyini bildirmeleri istendi. Anket sonucunda elde edilen veriler aşağıdaki tabloda özetlenmiştir:

	High School	Bachelors	Masters	Ph.d.	Total
Female	60	54	46	41	201
Male	40	44	53	57	194
Total	100	98	99	98	395

Cinsiyet ve eğitim düzeyi %5 önem düzeyinde bağımlı mı? Başka bir deyişle, yukarıda toplanan verilere göre, bireyin cinsiyeti ile almış olduğu eğitim düzeyi arasında bir ilişki var mıdır?

İşte beklenen sayıların tablosu:
60 için beklenen değer için E= (201*100)/395=50.886
54 için beklenen değer için E= (201*98)/395=49.868
46 için beklenen değer için E= (201*99)/395=50.377
41 için beklenen değer için E= (201*98)/395=49.868

40 için beklenen değer için E= (194*100)/395=49.114
44 için beklenen değer için E= (194*98)/395=48.132
53 için beklenen değer için E= (194*99)/395=48.623
57 için beklenen değer için E= (194*98)/395=48.132

	High School	Bachelors	Masters	Ph.d.	Total
Female	50.886	49.868	50.377	49.868	201
Male	49.114	48.132	48.623	48.132	194
Total	100	98	99	98	395

$$X^2 = \frac{(60 - 50.886)^2}{50.886} + \cdots + \frac{(57 - 48.132)^2}{48.132} = 8.006$$

Serbestlik derecesi = (r - 1) (c - 1) olan X_α^2 kritik değeriyle test istatistiğinin değeri, X^2 karşılaştırılır ve eğer $X^2 > X_\alpha^2$ varsa boş hipotez reddedilir.

Serbestlik derecesi = (r - 1) (c - 1) = (4-1)(2-1)=3

3 serbestlik dereceli kritik değeri X_α^2 =7.815'tir. $X^2 > X_\alpha^2$ (8.006 > 7.815) olduğundan, sıfır hipotezini reddedilir ve eğitim düzeyinin %5 anlamlılık düzeyinde cinsiyete bağlı olduğu sonucuna varılır.

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

7.4.4. Güç Analizi

Verilerden hesaplanan p-değerinin 0.12 olduğu bir araştırma deneyi düşünün. Sonuç olarak, bu p değeri α= 0,05'ten büyük olduğu için boş hipotez reddedilemez. Bununla birlikte, sıfır hipotezini reddedemediğimiz iki olası durum vardır:

1) sıfır hipotezi makul bir sonuçtur,
2) örneklem boyutu, sıfır hipotezini kabul edecek veya reddedecek kadar büyük değil, yani ek örnekler ek kanıt sağlayabilir.

Güç analizi, araştırmacıların, testin makul bir sonuca varmak için yeterli gücü içerip içermediğini belirlemek için kullanabilecekleri prosedürdür. Başka bir perspektiften, belirli bir güç seviyesine ulaşmak için gereken numune sayısını hesaplamak için güç analizi de kullanılabilir.

Örnek:
X, rastgele bir üniversite öğrencisinin boyunu göstersin. X'in bilinmeyen ortalama değeri μ ve standart sapması 9 ile normal olarak dağıldığını varsayın. n = 25 öğrenciden rastgele bir örnek alın, böylece α=0.05 durumunda I. Tip hata yapma olasılığını ayarladıktan sonra sıfır hipotezini Ho, μ=170 değerini alternatif hipoteze Ha, μ>175 karşı test edebiliriz.

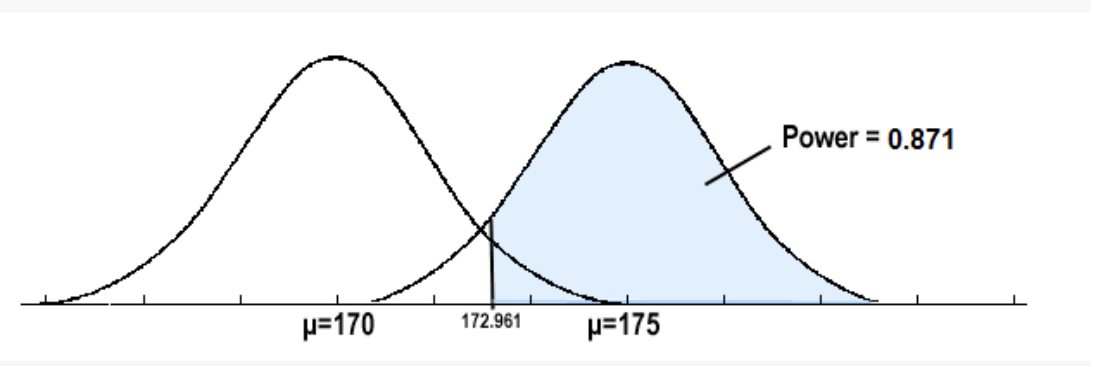
Gerçek popülasyon ortalaması μ=175 ise, hipotez testinin gücü nedir?

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$
$$\bar{x} = \mu + z\left(\frac{\sigma}{\sqrt{n}}\right)$$
$$\bar{x} = 170 + 1.645\left(\frac{9}{\sqrt{25}}\right)$$
$$= 172.961$$

Bu nedenle, gözlemlenen örnek ortalaması 172.961 veya daha büyük olduğunda boş hipotezi reddetmeliyiz:

$$\begin{aligned}\text{Power} &= P(\bar{x} \geq 172.961 \text{ when } \mu = 175) \\ &= P\left(z \geq \frac{172.961 - 175}{9/\sqrt{25}}\right) \\ &= P(z \geq -1.133) \\ &= 0.8713\end{aligned}$$

and illustrated below:



Özetle, gerçek bilinmeyen popülasyon ortalaması gerçekte ise, Ha, μ>175 alternatif hipotez lehine Ho, μ=170 sıfır hipotezini reddetme şansımızın %87,13 olduğu belirlenir.

Numune Boyutu Bölümünün Hesaplanması

Örneklem büyüklüğü sabitse, Tip I α hatasının azaltılması Tip II β hatasını da artıracaktır. Her ikisinin de azalması isteniyorsa, örneklem büyüklüğü artırılmalıdır.

Belirtilen α, β, μ_a değerleri için gereken en küçük örnek boyutunu hesaplamak için, μa, gücü değerlendirmek istediğiniz μ'nin olası değeridir.

Tek Kuyruklu Test için Örnek Boyutu:

$$n = \frac{\sigma^2(Z_\alpha + Z_\beta)^2}{(\mu_0 - \mu_a)^2}$$

İki Kuyruklu Test için Örnek Boyutu:

$$n = \frac{\sigma^2(Z_{\alpha/2} + Z_\beta)^2}{(\mu_0 - \mu_a)^2}$$

Örnek:
X, rastgele bir üniversite öğrencisinin boyunu göstersin. X'in bilinmeyen μ ortalama ve standart sapma 9 ile normal olarak dağıldığını varsayın.

α=0.05 durumunda I. sıfır hipotezi Ho, μ=170 değerini alternatif hipoteze Ha, μ>175 karşı test edebiliriz. μ = 175 alternatifinde 0.90 güce ulaşmak için gerekli olan numune boyutunu n bulun.

$$\begin{aligned} n &= \frac{\sigma^2(Z_\alpha + Z_\beta)^2}{(\mu_0 - \mu_a)^2} \\ &= \frac{9^2(1.645 + 1.28)^2}{(170 - 175)^2} \\ &= 27.72 \\ n &= 28 \end{aligned}$$

Özetle, veriler sıfır hipotezinin reddedilemeyeceğini gösterdiğinde doğru kararın verebilebilmesi için güç analizinin ne kadar önemli olduğu görülmelidir. Ayrıca, araştırmanın ihtiyaçlarını karşılayan bir farkı tespit etmek için gereken minimum örnek boyutunu hesaplamak için güç analizinin nasıl kullanılabileceği de görilmelidir.

8. Markov Chain Analysis

Stochastic (Random) process is a probability model used to describe phenomena that change or evolve over time or space, the existence of which has been experimentally proven. Evolve: To naturally change from one form to another. The weather is sunny now, what will it be like tomorrow? Rainy? Cloudy? Sunny?

Stokastik (Rastgele - Rassal) süreç, varlığı deneysel olarak kanıtlanmış zaman veya mekana göre değişen ya da evrilen olguları tanımlamak için kullanılan bir olasılık modelidir. Evrilmek: Bir biçimden başka bir biçime doğal olarak dönmek. Şu an hava Güneşli yarın nasıl olacak? Yağmurlu mu? Bulutlu mu? Güneşli mi?

If your states and transition probabilities are known, whether in the past, present or future, the probabilities of your next state can be calculated.

İster geçmişte, ister şu an, ister gelecekte durumlarınız ve geçiş olasılıklarınız biliniyorsa bir sonraki adımdaki durumunuzun olasılıkları hesaplanabilir.

A stochastic process is a mathematical model that develops in a probabilistic manner over time. In the study of a special stochastic process called a Markov chain, the next state of the system does not depend on previous states, but only on the current, present state. Traces of tomorrow are in today. You will decide for tomorrow today. The theory of Markov Chains in the analysis of stochastic processes is named after the Russian mathematician A.A. Markov (1856–1922). Each state is a result of the previous state, not of the previous states. How random variables change over time also includes stochastic processes. For example, how the price of a share changes on the stock exchange or how the market share of a firm changes is related to the stochastic process. Markov Chains, an example of a stochastic process, are applied to areas such as education, marketing, health care, accounting and manufacturing. They form the basis of mathematical models developed especially for machine learning.

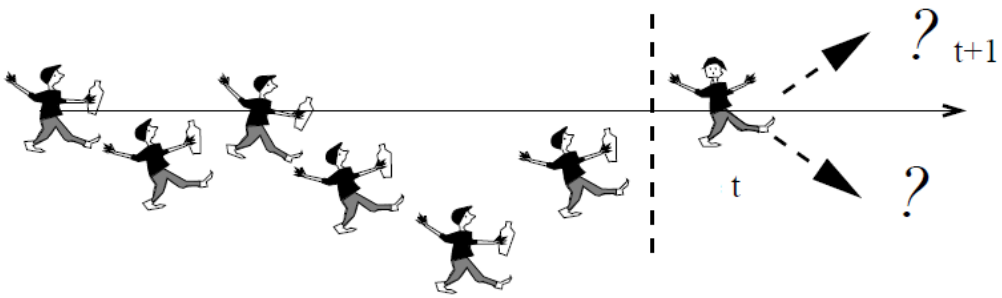
Stokastik bir süreç, zamana bağlı olasılıklı bir şekilde gelişen matematiksel modeldir. Markov zinciri adı verilen özel bir stokastik süreç çalışılmasında, **sistemin bir sonraki durumu önceki durumlara değil, yalnızca mevcut, şu andaki durumuna bağlıdır.** Yarının izleri bugündedir. Yarın için bugün karar vereceksin. Stokastik süreçlerin analizinde Markov Zincirleri teorisi, ismini Rus matematikçi A.A. Markov'dan (1856–1922) almıştır. ***Her bir durum ancak ve ancak bir önceki durumun sonucudur, ondan önceki durumların sonucu değildir***. Rassal değişkenlerin zamanla nasıl değiştiği stokastik süreçleri de içerir. Örneğin borsada bir hissenin fiyatının nasıl değiştiği veya bir firmanın piyasa payının nasıl değiştiği stokastik süreçle ilgilidir. Bir stokastik süreç örneği olan Markov Zincirleri eğitim, pazarlama, sağlık hizmetleri, muhasebe ve üretim alanları gibi alanlara uygulanmaktadır. Özellikle makine öğrenmesine yönelik geliştirilen matematiksel modellerin temelini oluşturmaktadır.

Markov chains are a special type of discrete-time stochastic processes. In simple terms, at any given time a discrete-time stochastic process can be in one of a finite number of states. If a discrete-time stochastic process satisfies the following condition, then the process is a Markov chain.

Markov zincirleri, ayrık zamanlı stokastik süreçlerin özel bir türüdür. Basit bir ifadeyle herhangi bir zamanda ayrık zamanlı stokastik süreç sonlu sayıda durumdan birinde olabilir. Ayrık zamanlı stokastik süreç aşağıdaki koşulu sağlıyorsa süreç Markov zinciridir.

The processes can be written as {X0, X1, X2,. . .}, where t= 0,1,2,… is the state at time t for each state Xt. The processes we look at via the transition diagram of the states have a common property: state Xt+1 depends only on state Xt. It does not depend on states X0, X1, . . . , Xt−1.

Süreçler $\{X_0, X_1, X_2, \ldots\}$, şeklinde yazılabilir, burada t= 0,1,2,… her bir durum için X_t, t zamanındaki durumdur. Durumların geçiş diyagramı üzerinden baktığımız süreçlerin ortak bir özelliği var: X_{t+1} durumu sadece X_t durumuna bağlıdır. $X_0, X_1, \ldots, X_{t-1}$ durumlarına bağlı değildir.



We formulate the Markov Property in mathematical notation as follows:
Markov Özelliğini matematiksel gösterimle şu şekilde formüle ediyoruz:

$$\mathbb{P}(X_{t+1} = s \,|\, X_t = s_t, X_{t-1} = s_{t-1}, \ldots, X_0 = s_0) \;=\; \mathbb{P}(X_{t+1} = s \,|\, X_t = s_t),$$

Properties of the transition matrix:
1) It is a square matrix, because all possible states must be used as both rows and columns.
2) All input data is between 0 and 1; This is because all inputs represent probabilities.
3) The sum of the probabilities of the data in any row must be 1, because the numbers in the row give all the transition probabilities of one state to another.

Geçiş matrisinin özellikleri:

1) Kare matristir, çünkü tüm olası durumların hem satır hem de sütun olarak kullanılmak zorundadır.
2) Tüm giriş verileri 0 ile 1 arasındadır; Bunun nedeni bütün girdilerin olasılıkları temsil edilmesidir.
3) Herhangi bir satırdaki verilerin olasılıkları toplamı 1 olmalıdır, çünkü satırdaki sayılar, bir durumun diğer durum olan tüm geçiş olasılıklarını verir.

$$p_{ij} = \mathbb{P}(X_{t+1} = j \mid X_t = i)$$

For a sequence of trials of an experiment to be a Markov chain,
1) The outcome of each experiment must be one of the specified states,
2) The outcome of an experiment must depend only on the current state and not on any past state. Probability coefficients are obtained from past observations or measurements.

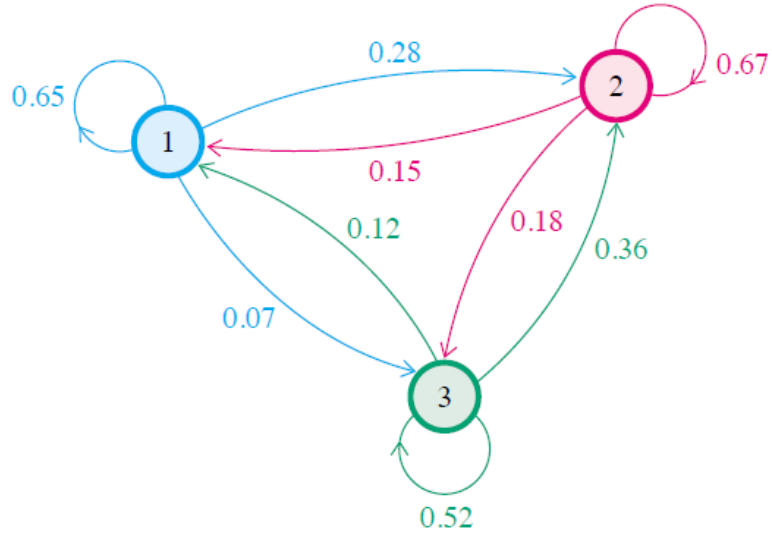
Bir deneyin denemelerinin bir sırası, bir Markov zinciri olabilmesi için

1) Her bir deneyin sonucu, belirli durumlardan bir tanesi olmalıdır,
2) Bir deneyin sonucu, sadece şu andaki duruma bağlı ve geçmiş herhangi bir duruma bağlı olmamalıdır. Olasılık kaysayıları, geçmişteki gözlemlerden ya da ölçümlerden elde edilir.

Transition Table of States:

		Next Generation		
	State	1	2	3
Current	1	0.65	0.28	0.07
Generation	2	0.15	0.67	0.18
	3	0.12	0.36	0.52

Transition or flow diagram of states:



The transition matrix is the matrix that gives the probability of other states occurring after one state occurs.

Geçiş matrisi, bir durum olduktan sonra diğer durumların olma olasılığının veren matrisdir.

The transition matrix:

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.65 & 0.28 & 0.07 \\ 0.15 & 0.67 & 0.18 \\ 0.12 & 0.36 & 0.52 \end{bmatrix} \end{array} = P$$

What does p23 = 0.18 mean? The probability of going from state 2 to state 3 in the next step is 18%.

p_{23} = 0.18 anlamı nedir? 2.durumdan bir sonraki adımda 3. Duruma gitme olasılığı %18 dir.

Scalar multiplication:

Scalar * matrix = scalar multiplication

$$\lambda \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} = \begin{pmatrix} \lambda a & \lambda b & \lambda c \\ \lambda d & \lambda e & \lambda f \end{pmatrix}$$

Matrix multiplication:

Let the row and column vectors be given as follows.

$A = (a \;\; b \;\; c), \qquad B = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$

In this situation, $AB = (a \;\; b \;\; c)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = ax + by + cz$

Similarly, $BA = \begin{pmatrix} x \\ y \\ z \end{pmatrix}(a \;\; b \;\; c) = \begin{pmatrix} xa & xb & xc \\ ya & yb & yc \\ za & zb & zc \end{pmatrix}$

Multiplication of square matrix and column vector:

$A = \begin{pmatrix} a & b & c \\ p & q & r \\ u & v & w \end{pmatrix}$

$B = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$

$AB = \begin{pmatrix} a & b & c \\ p & q & r \\ u & v & w \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} ax + by + cz \\ px + qy + rz \\ ux + vy + wz \end{pmatrix}$

Not: BA is not defined

8.1. Transition matrix in Markov chain analysis

In Markov chain analysis, the transition matrix must meet the following conditions:
1) Each element is a probability and its value is between 0 and 1. Probabilities cannot be negative and greater than one.
2) The sum of the probabilities in each row is equal to 1. Since the elements in a row of the transition matrix give the results resulting from the probabilities of possible events occurring, it is clear that the sum of the probabilities is one.
Markov zincir analizinde geçiş matrisi, aşağıdaki koşulları sağlaması gerekir:

1) Her bir eleman olasılıktır ve değeri 0 ile 1 arasındadır. Olasılıklar negatif ve birden büyük olamaz.
2) Her bir satırdaki olasılıkların toplamı 1'e eşittir. Geçiş matrisinin bir satırdaki elemanları muhtemel olayların gerçeklenme olasılıklarından doğan sonuçları vermesi nedeni ile, olasılıklar toplamının bir olması açıktır.

The transition matrix is given in a general notation as

$$0 \leq P_{ij} \leq 1,$$

$$\sum_{j=1}^{m} P_{iJ} = 1, (i = 1,2,....,m)$$

including,

$$P=\begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1m} \\ \vdots & \vdots & \vdots & \vdots \\ P_{m1} & P_{m2} & \cdots & P_{mm} \end{bmatrix}$$

is written as a square matrix. It is defined as a row vector v i to represent the i row. It should not be forgotten that the row vectors v i are probability vectors.
kare matrisi yazılır. i satırı temsil etmek üzere v^i satır vektörü olarak tanımlanır. v^i satır vektörlerinin birer olasılık vektörü olduğu unutulmamalıdır.

$v^2 = (P_{21}\ P_{22}\ \ldots\ P_{2m})$, The probability vector represents the second row of the transition matrix and the vector elements give the probabilities of transitioning from state-2 to all other states, respectively. Then, the transition matrix P consists of probability vectors v i and satisfies the transition matrix conditions.
$v^2 = (P_{21}\ P_{22}\ \ldots\ P_{2m})$ olasılık vektörü geçiş matrisinin ikinci satırını temsil eder ve vektör elemanları sırasıyla durum-2'den diğer bütün durumlara geçme olasılıklarını verir. O halde P geçiş matrisi v^i olasılık vektörlerinden oluşur ve geçiş matrisi koşullarını sağlar.

Basic properties of the transition matrix:

1- Each probability value of the vector is between 0 and 1.
2- The sum of the coefficients of the vector is equal to 1.

Geçiş matrisin temel özellikleri:

1- Vektörün her bir olasılık değeri 0 ile 1 arasındadır.
2- Vektörün katsayıları toplamı 1'e eşittir.

Example:

Let's classify three weather conditions,
Situation-1: Rainy
Situation-2: Cloudy
Situation-3: Sunny
Assumption: Tomorrow's weather depends only on today's weather! Let's say the probabilities of the weather being Rainy, Cloudy and Sunny are given below.

Current situation (n=0)	Next situation (n=1)		
	Rainy (%)	Cloudy (%)	Sunny (%)
Rainy	40	30	30
Cloudy	20	50	30
Sunny	25	25	50

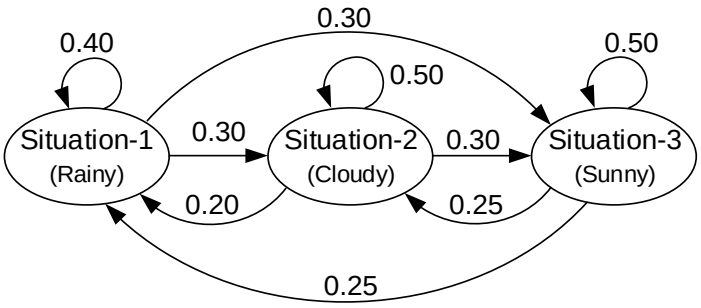
The table gives the current state and the probabilities of the next state. The information in the table provides a transition matrix, which we will denote by P.
Tablo, şu anki durumu ve bir sonraki durumun olma olasılıklarını vermektedir. Tablodaki bilgiler P ile göstereceğimiz bir **geçiş matrisi** bulunur.

$$P = \begin{bmatrix} 0.40 & 0.30 & 0.30 \\ 0.20 & 0.50 & 0.30 \\ 0.25 & 0.25 & 0.50 \end{bmatrix}$$

Each element of the transition matrix gives the probability of transitioning from one state to another, and these elements are called Pij. This is the conditional probability that the process currently in state i will be in state j at the next step.
Geçiş matrisinin her bir elemanı bir durumdan diğer bir duruma geçme olasılığını verir ve bu elemanlara P_{ij} denir. Bu ise, halen i. durumda olan sürecin bir sonraki adımda j. durumda olacağını gösteren şartlı olasılıktır.

The P13=0.30 element gives the probability that if it is raining now, the next day will be sunny. The probability vector v 2 =(0.20 0.50 0.30) represents the second row of the transition matrix and the vector elements give the probabilities of transitioning from state-2 to state-1, state-2 and state-3, respectively. Therefore, the transition matrix P consists of the probability vectors v i and satisfies the transition matrix conditions. The transition diagram of the states is given below. What is the probability of YYGGYBGB?

P_{13}=0.30 elemanının, şu an yağmur yağıyorsa bir sonraki gün güneşli olma olasılığını vermektedir. v^2 =(0.20 0.50 0.30) olasılık vektörü geçiş matrisinin ikinci satırını temsil eder ve vektör elemanları sırasıyla durum-2'den durum-1, durum-2 ve durum-3'e geçme olasılıklarını verir. O halde **P geçiş matrisi v^i olasılık vektörlerinden oluşur ve geçiş matrisi koşullarını sağlar.** Durumların geçiş diyagramı aşağıda verilmiştir. YYGGYBGB olma olasılığı nedir?



8.2. Probability Analysis with Markov Chain

What will be the probability of a weather being Rainy at n=0, Rainy at n=1 and Cloudy at n=2? It is possible to answer the question with probability theory techniques. It is easier to work with a diagram to separate the different options one by one from a weather being Rainy to being Cloudy in the second step.

Bir havanın n=0 da Yağmur olması, n=1 de Yağmur olması ve n=2 de Bulutlu olma olasılığı ne olacaktır? Soruyu olasılık teorisi teknikler ile cevaplamak olanağı vardır. Yağmur olan bir havanın ikinci adımda Bulut olana kadar olan değişik seçeneklerini teker teker ayırmak için bir diyagramla çalışmak kolaylık sağlar.

If the weather is raining now, what is the probability that it will be cloudy in the second step (not tomorrow but the day after)?

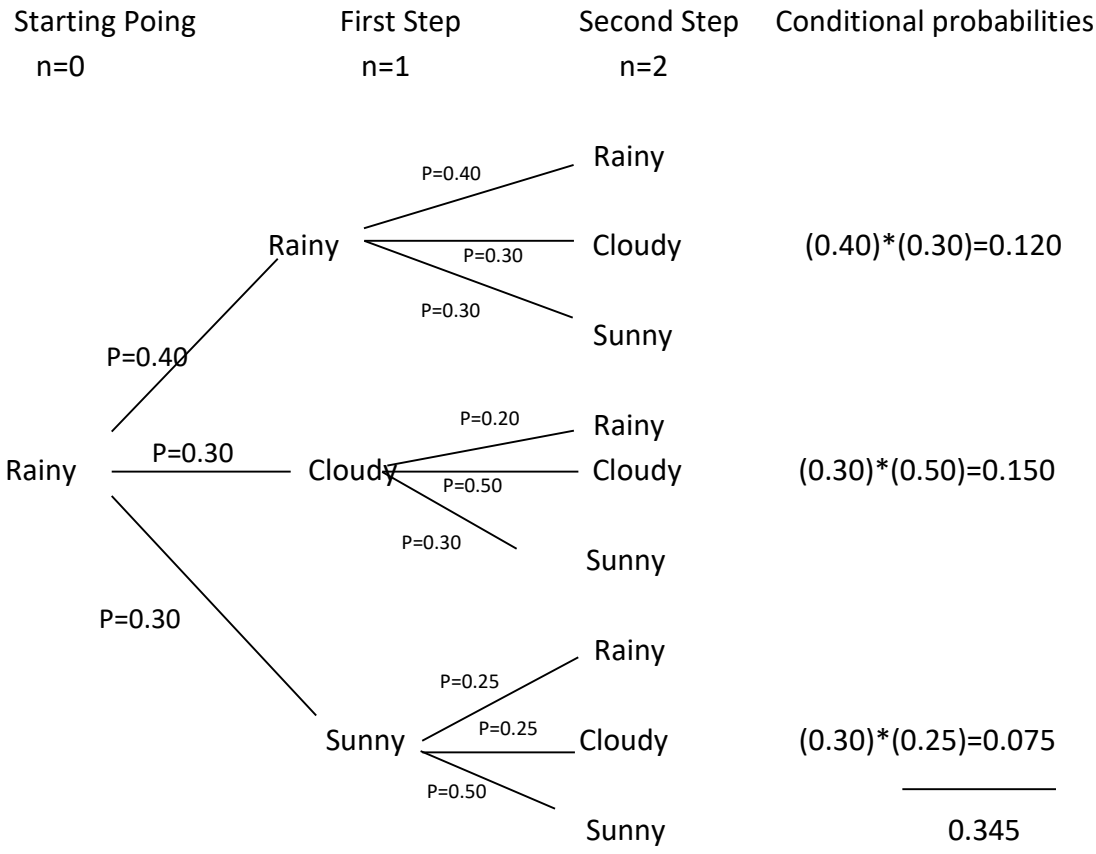


Figure: n=2, possible results in the second step

In the figure, the probability given for cloudy weather in step 2 is added up to find 0.345. This is the conditional probability. The table below summarizes the operations in the graph and the probability of cloudy weather being rainy now in step n=0 is 0.345 in step n=2. The probability of cloudy weather being rainy today but not tomorrow is 34.5%.

Şekilde, 2. adımda Bulutlu olması için verilen olasılıkların toplanması ile 0,345 bulunur. Bu ise şartlı olasılıktır. Aşağıdaki tablo grafikteki işlemleri özetlemektedir ve n=0 adımında Şu an yağmurlu olan havanın n=2 adımda Bulutlu olma olasılığı 0,345 dir. Şu an yağmurlu iken yarın değil öbürgünün bulutlu olma olasılığı:%34.5 dır.

n=0 - Starting	n=1	n=2	Conditional Probability
Rainy	Rainy 0.40 Cloudy 0.30 Sunny 0.30	Cloudy 0.30 Cloudy 0.50 Cloudy 0.25	(0.40)*(0.30)=0.120 (0.30)*(0.50)=0.150 (0.30)*(0.25)=0.075 0.345

The result found can also be obtained with the analysis technique provided by Markov chains. First, it is necessary to explain what the probability vector V1 is used for. The probability vector for the S1 situation is V1= (0.40; 0.30; 0.30). In the n=0 step, the V1 vector will give all the supply probabilities of a weather with rain in the second step. V2 is found by multiplying the V1 vector with the P matrix.

Bulunan sonuç Markov zincirlerinin sağladığı analiz tekniğiyle de elde edilebilir.

Önce V^1 olasılık vektörünün ne anlamda kullanıldığını açıklamaya gerek vardır. S^1 durumu için olasılık vektörü V^1= (0.40; 0.30; 0.30) dur. n=0 adımında Yağmur olan bir havanın ikinci adımdaki tüm tedarik olasılıklarını V^1 vektörü verecektir. V^1 vektörünün P matrisi ile çarpımı ile V^2 bulunur.

$$v^2 = v^1 P = [0.4\,,0.3\,,0.3] \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$$

= [0,295 **0,345** 0,360]

The vector V^2 gives the probabilities of the weather being Rain at step n=0 and being Rain at step n=1 and being Rain, Cloudy and Sunny at step n=2. Markov chains have provided the technique and analysis for formulating special cases of probability problems.

V^2 vektörü n=0 adımda Yağmur olan havanın n=1 adımında Yağmur iken n=2 adımında Yağmur, Bulutlu ve Güneşli olma olasılıklarını vermektedir. Markov zincirleri,olasılık problemlerinin özel halini formüle etmek için teknik ve analiz vermiştir.

V^2=P x P

$$V^2 = \begin{bmatrix} 0.2950 & 0.3450 & 0.3600 \\ 0.2550 & 0.3850 & 0.3600 \\ 0.2750 & 0.3250 & 0.4000 \end{bmatrix}$$

n: When it is rain
n=1st step [Rain, Cloudy, Sunny]=[0.4, 0.3, 0.3]
n=2nd step [Rain, Cloudy, Sunny]=[0.295, 0.345, 0.360]
is found as

n: Yağmur iken
n=1'inci adımda [Yağmur, Bulutlu, Güneşli]=[0.4, 0.3, 0.3]
n=2'inci adımda [Yağmur, Bulutlu, Güneşli]=[0.295, 0.345, 0.360]
olarak bulunur.

Example:
Starting vector of a Markov process, $v^0 = (0.6 \quad 0.4)$ and the transition probability matrix

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix}$$

Since $v^1=v^0 P$ for the first step
Birinci adım için $v^1 = v^0 P$ olduğundan
$v^1 = (0.6 \quad 0.4)\begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix} = (0.86 \quad 0.14) = (v_1^1 \quad v_2^1)$ is found.

The probability of being in state S^1 at the end of a step of that process is 0.86 and the probability of being in state S^2 is 0.14. These results can also be obtained with a tree diagram.
O sürecin bir adım sonunda S^1 durumunda bulunma olasılığı 0.86 ve S^2 durumunda bulunma olasılığı 0.14 dir.Bu sonuşlar ağaç diyagramı ile de elde edilebilir.

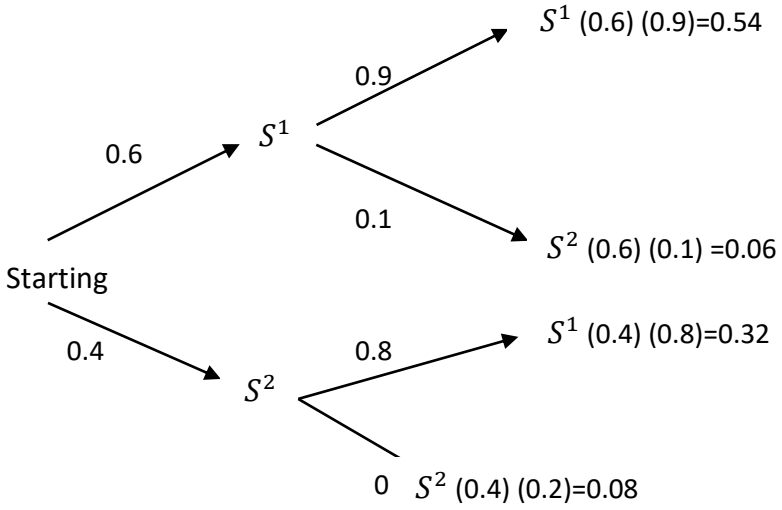


Figure: State space after one step with starting probabilities (0.6 0.4).
Şekil: Başlama olasılıkları (0.6 0.4) olmak üzere bir adım sonra durum uzayı.

$v_1^1 = (0.6 \quad 0.4)\begin{pmatrix}0.9\\0.8\end{pmatrix}$=0.54+0.32=0.86

$v_2^1 = (0.6 \quad 0.4)\begin{pmatrix}0.1\\0.2\end{pmatrix}$=0.06+0.08=0.14

$$v^2 = v^1P = (0.86 \quad 0.14)\begin{bmatrix}0.9 & 0.1\\0.8 & 0.2\end{bmatrix} = (0.886 \quad 0.114) = (v_1^2 \quad v_2^2)$$

or

$$v^2 = v^0P^2 = (0.6 \quad 0.4)\begin{bmatrix}0.9 & 0.1\\0.8 & 0.2\end{bmatrix}\begin{bmatrix}0.9 & 0.1\\0.8 & 0.2\end{bmatrix} = (0.886 \quad 0.114) = (v_1^2 \quad v_2^2)$$

8.3. Long-Term Probability in the Transition Matrix

The long-term probability can be considered as the steady-state probability. Because when a state in the process is stable, we can calculate the steady-state probability. Here in the Markov chain, if the first stage is stable, i.e. once it becomes stable, we can calculate the steady-state probability. Let's say V0 is the initial state probability vector and T is the transition matrix, i.e. the one-shot step estimate can be shown as:
Uzun dönem olasılığı, kararlı durum olasılığı olarak kabul edilebilir. Çünkü süreçteki bir durum kararlı olduğunda kararlı durum olasılığını hesaplayabiliriz. Burada Markov zincirinde, eğer ilk aşama kararlı ise, yani bir kez sabit hale geldiğinde, kararlı durum olasılığını hesaplayabiliriz. Diyelim ki **V0 ilk durum olasılık vektörü** ve **T geçiş matrisi,** yani tek seferlik adım tahmini şu şekilde gösterilebilir:

V1 = V0 x T

Something worth noting here and very simple mathematics is the dot product of a vector and a matrix in a vector, and with this intuition we can say that in the process of predicting a one-shot step we again encounter a vector that is considered as the initial state. Or more formally to say that each one-shot step predicted in the future will only be responsible for the next step.
Burada dikkate değer ve çok basit bir matematik olan bir şey, bir vektördeki vektör ve matrisin nokta çarpımıdır ve bu sezgiyle, bir kerelik adımı tahmin etme sürecinde tekrar bir vektörle karşılaştığımızı söyleyebiliriz. başlangıç durumu olarak kabul edilir. Veya daha resmi olarak gelecekte tahmin edilen her bir tek seferlik adımın yalnızca bir sonraki adımından sorumlu olacağını söylemek.

So if we want to predict the second step, the prediction formula would be:
Dolayısıyla, ikinci adımı tahmin etmek istiyorsak, tahmin formülü şu şekilde olacaktır:

V2 = V1 x T

And here we know the value of V1 from the estimate of one step. Substituting the value of V1 into
Ve burada bir adımın tahmininden V1'in değerini biliyoruz. V1 değerini yerine koyarak

V2 = (V0 x T) x T

V2 = V0 x T^2

Similarly, the estimate for the third step would be:
Benzer şekilde, üçüncü adım için tahmin şu şekilde olacaktır:

V3 = V2 x T = (V0 x T^2)xT

V3 = V0 x T^3

Therefore, when talking about the nth time step prediction, the prediction can be calculated by the following formula.
Bu nedenle, n'inci zaman adımı tahmininden bahsederken, tahmin aşağıdaki formülle hesaplanabilir.

Vn = Vn-1 x T = V0 xT^n

So, the above iterative process helps in estimating the probability of future states of long processes in this way. Here, the long-run probability can be written as:
Dolayısıyla, yukarıda verilen yinelemeli süreç, uzun süreçlerin gelecekteki durum olasılığının tahmininde bu şekilde yardımcı olur. Burada uzun dönem olasılık şu şekilde yazılabilir:

V∞ = V0 x T^∞.

From the above long-run probability formula, we can say that no amount of multiplication by the transition matrix results in changes in the long-run probability vector.
Yukarıdaki uzun dönem olasılık formülünden, geçiş matrisi tarafından yapılan hiçbir çarpma miktarının uzun dönem olasılık vektöründe değişikliklere yol açmadığını söyleyebiliriz.

Advantages of Markov Chain:
- As we have seen above, it is very easy to derive Markov chain from a sequential data.
- We do not need to dive deep into the dynamic change mechanism.

• Markov chain is very insightful. It can tell us the area where we are lacking in any process and also make changes according to the improvement.
• Very low or modest computational requirements, can be easily calculated by any dimension of the system.

Markov Zincirinin Avantajları:

- Yukarıda gördüğümüz gibi**, Markov zincirini ardışık bir veriden türetmek çok kolaydır**.
- Dinamik değişim mekanizmasının derinliklerine dalmamıza gerek yok.
- Markov zinciri çok anlayışlı. Herhangi bir sürecin eksik olduğumuz alanını söyleyebilir ve ayrıca iyileştirmeye göre değişiklikler yapabiliriz.
- Çok düşük veya mütevazı hesaplama gereksinimleri, sistemin herhangi bir boyutu tarafından kolayca hesaplanabilir.

8.4. Ergodic (Regular) Markov Chains

An ergodic chain describes a process in which mathematical transitions from any state to all other states are possible. In this definition, there is no chance of finding an exact step, but a state must be reached regardless of the starting state. The chain must be ergodic to ensure that steady-state conditions are reached.

Ergodik zincir, matematik olarak herhangi bir durumdan diğer bütün durumlara geçişin mümkün olduğu bir süreci tanımlar. Bu tanımda tam bir adımda bulunma şansı yoktur, ama başlama durumuna bakmadan bir duruma erişilmiş olmalıdır. Denge durumu koşullarına erişilmesini sağlamak için zincir mutlaka ergodik olmalıdır.

The bounding case of the ergodic Markov chain is a regular chain. A regular chain requires that the elements in the powers of the P transition probability matrix be nonzero and positive. If the powers of the ergodic chain are taken or if the powers are taken until there are no zero elements left in the matrix, it is seen that the ergodic chain is regular. This process is given below. It is true that all regular chains are ergodic, but the converse need not be true. It should be noted that not all ergodic chains are regular.

Ergodik markov zincirini sınırlayan hal düzenli (=reguler) zincirdir. ***Düzenli zincir, P geçiş olasılıkları matrisinin kuvvetlerinde bulunan elemanların sıfırdan farklı ve pozitif olmasını gerektirir.*** Ergodik zincirin kuvvetleri alınırsa veya matriste sıfır eleman kalmayıncaya kadar kuvvetler alınırsa ergodik zincirin düzenli olduğu görülür. Bu işlem aşağıda verilmiştir.Bütün düzenli zincirlerin ergodik olduğu doğrudur, ama tersininde doğru olmasına ihtiyaç yoktur. Bütün ergodik zincirlerin düzenli olmadığına dikkat etmelidir.

$$P=\begin{bmatrix} x & x & 0 & x & x \\ 0 & x & x & 0 & x \\ 0 & 0 & 0 & x & x \\ x & 0 & x & 0 & x \\ x & x & 0 & 0 & 0 \end{bmatrix}, \quad P^2=\begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & 0 & x \\ x & x & 0 & x & x \\ x & x & x & x & x \end{bmatrix}, \quad P^4=\begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \end{bmatrix}$$

Example:

Örnek:

Explain whether the following transition matrices are a) regular and b) ergodic.

Aşağıdaki geçiş matrislerinin a) düzenli ve b) ergodik olmasını açıklayınız.

$$P=\begin{matrix} 1 \\ 2 \\ 3 \end{matrix}\begin{bmatrix} x & x & 0 \\ x & 0 & x \\ 0 & x & x \end{bmatrix} \qquad P^2=\begin{bmatrix} x & x & x \\ x & x & x \\ x & x & x \end{bmatrix}$$

Since the elements of the P^2 power of the transition matrix P are positive or different from zero, the Markov chain given by the transition matrix P is regular. Regular Markov chains are ergodic. Because there is a direct transition from 1 to 1 or 2, then a transition from 2 to 3 is possible. It is possible to pass from 2 to 1 and from 3 to 2, to 1. Therefore, the chain is ergodic, there is a transition to all states.

P geçiş matrisinin P^2 kuvvetinin elemanları pozitif veya sıfırdan farklı olduğundan P geçiş matrisi ile verilen Markov zinciri düzenlidir. Düzenli Markov zincirleri ise ergodiktir. Çünkü 1 den 1 veya 2 ye doğrudan geçiş vardır,daha sonra da 2 den 3 e geçiş olanaklıdır.2 den 1e geçilebilir ve 3 den 2 ye ,1 e geçiş vardır.Dolayısıyla zincir ergodiktir,bütün durumlara geçiş olanağı vardır.

Example:
Örnek:

$$P=\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} x & 0 & x & 0 \\ 0 & x & 0 & x \\ x & 0 & x & 0 \\ 0 & x & 0 & x \end{bmatrix} \end{array} \quad P^2=\begin{bmatrix} x & 0 & x & 0 \\ 0 & x & 0 & x \\ x & 0 & x & 0 \\ 0 & x & 0 & x \end{bmatrix} \quad P^4\begin{bmatrix} x & 0 & x & 0 \\ 0 & x & 0 & x \\ x & 0 & x & 0 \\ 0 & x & 0 & x \end{bmatrix}$$

The powers of the transition matrix P give the matrix P again. Therefore, the stochastic matrix P is not a regular chain. Because there are zero elements in the first matrix and the zero elements in the powers remain the same. There is also a transition from 1 to 1 or 3, and from 3 to 3 or 1. Therefore, there is no possibility of transition from state 1 to state 2 or state 4, and the chain is not ergodic.

P geçiş matrisinin kuvvetleri P matrisini tekrar vermektedir. Dolayısıyla P stokastik matrisi düzenli bir zincir değildir. Zira ilk matriste sıfır elemanları vardır ve kuvvetlerde sıfır elemanlar aynen kalmıştır. Ayrıca 1 den 1 veya 3 e,ve 3 den 3 veya 1 e geçiş vardır. Dolayısıyle 1. durumdan 2. duruma veya 4. duruma geçiş olanağı yoktur, ve zincir ergodik değildir.

Decide whether the following transition matrices are regular.

(a) $A=\begin{bmatrix} 0.75 & 0.25 & 0 \\ 0 & 0.5 & 0.5 \\ 0.6 & 0.4 & 0 \end{bmatrix}$

Solution Square A.

$$A^2=\begin{bmatrix} 0.5625 & 0.3125 & 0.125 \\ 0.3 & 0.45 & 0.25 \\ 0.45 & 0.35 & 0.2 \end{bmatrix}$$

Since all entries in A^2 are positive, matrix A is regular.

(b) $B=\begin{bmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Solution Find various powers of B.

$$B^2=\begin{bmatrix} 0.25 & 0 & 0.75 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix};\ B^3=\begin{bmatrix} 0.125 & 0 & 0.875 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix};\ B^4=\begin{bmatrix} 0.0625 & 0 & 0.9375 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Further powers of B will still give the same zero entries, so no power of matrix B contains all positive entries. For this reason, B is not regular.

Example: Investigate whether the Markov chain given below is ergodic and regular.

$$P=\begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix} \qquad P^2=\begin{bmatrix} 1 & 0 \\ 3/4 & 1/4 \end{bmatrix}$$

If the powers of the matrix P are taken, the element (1.2) will always be (0). Therefore, the given Markov chain is not regular and ergodic.

Example:

$$P=\begin{bmatrix} 0 & 1 \\ 1/3 & 2/3 \end{bmatrix}$$

Show that it is a regular Markov chain.

$$P^2 = P*P = \begin{bmatrix} 0 & 1 \\ 1/3 & 2/3 \end{bmatrix} * \begin{bmatrix} 0 & 1 \\ 1/3 & 2/3 \end{bmatrix} = \begin{bmatrix} 1/3 & 2/3 \\ 2/9 & 7/9 \end{bmatrix}$$

Since all elements of P^2 are positive, P is a regular, i.e. ergodic Markov chain.
P^2 nin bütün elemanları pozitif olduğundan P düzenli yani ergodik Markov zinciridir.

Example: Find out whether the following matrix is regular or not.
Aşağıdaki matrisin reguler (düzenli) olup olmadığını bulunuz.

$$P = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0 & 0.5 & 0.5 \\ 0.6 & 0.4 & 0 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 0.5625 & 0.3125 & 0.125 \\ 0.3 & 0.45 & 0.25 \\ 0.45 & 0.35 & 0.2 \end{bmatrix}$$

Matrix P is regular when all values in matrix P^2 are positive.

Example:

Find out whether the following matrix is regular or not.

$$P = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 0.25 & 0 & 0.75 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P^3 = \begin{bmatrix} 0.125 & 0 & 0.875 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P^4 = \begin{bmatrix} 0.0625 & 0 & 0.9375 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Since all values in matrix P^2 are non-positive, matrix P is not regular.

P^2 matrisindeki bütün değerler pozitif olmadığından P matrisi düzenli değildir.

8.5. Equilibrium Conditions

In regular ergodic chains, the existence of equilibrium conditions is determined by calculating Pn. As can be seen in the powers of P, as n increases, the Pij values approach a fixed number or limit; the transition matrix approaches the equilibrium state. It corresponds to the eigenvalues and eigenvectors, and provides information about the behavior and direction of the system.

Düzenli ergodik zincirlerde denge durumu koşullarının varlığı P^n hesaplanarak belirlenir. P'nin kuvvetlerinde görüleceği gibi n büyüdükçe Pij değerleri sabit bir sayıya veya limite; geçiş matrisi denge durumuna yaklaşmaktadır. Özdeğer ve özvektörlere denk gelir, sistemin davranışı ve yönü hakkında bilgi verir.

Example:

Calculations are given for values from n=1 to n=8 according to the nth powers of P.

P nin n. kuvvetlerine göre n=1 den n=8 e kadar değerler için ilişkin hesaplar verilmiştir.

$$P=\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.25 & 0.25 & 0.5 \end{bmatrix} \qquad P^2=\begin{bmatrix} 0.295 & 0.345 & 0.360 \\ 0.255 & 0.385 & 0.360 \\ 0.275 & 0.325 & 0.400 \end{bmatrix}$$

$$P^3=\begin{bmatrix} 0.277 & 0.351 & 0.372 \\ 0.269 & 0.359 & 0.372 \\ 0.275 & 0.345 & 0.380 \end{bmatrix} \qquad P^4=\begin{bmatrix} 0.2740 & 0.3516 & 0.3744 \\ 0.2724 & 0.3744 & 0.3744 \\ 0.2740 & 0.3500 & 0.3760 \end{bmatrix}$$

$$P^5=\begin{bmatrix} 0.27352 & 0.35160 & 0.37488 \\ 0.27320 & 0.35192 & 0.37488 \\ 0.27360 & 0.35120 & 0.37520 \end{bmatrix} \quad P^6\begin{bmatrix} 0.273448 & 0.351576 & 0.374976 \\ 0.273384 & 0.356400 & 0.374976 \\ 0.273480 & 0.351400 & 0.375040 \end{bmatrix}$$

$$P^7=\begin{bmatrix} 0.2734384 & 0.3515664 & 0.3749952 \\ 0.2734256 & 0.3515792 & 0.3749952 \\ 0.2734480 & 0.3515440 & 0.3750080 \end{bmatrix}$$

$$P^8=\begin{bmatrix} 0.27343744 & 0.35156352 & 0.37499904 \\ 0.27343488 & 0.35156608 & 0.3749904 \\ 0.27244000 & 0.35155840 & 0.37500160 \end{bmatrix}$$

Calculation of steady-state conditions:

Denge durum koşullarının hesaplanması:

In the equilibrium state, each v_i^n probability vector tends to be equal for all values.
Denge durumunda her bir v_i^n olasılık vektörlerinin bütün değerleri için eşit olmaya meyletmektedir.

Therefore, the following two rules are written:

1) For sufficiently large values of n, the probability vectors v_i^n are the same for all o values and do not change.
2) Since $v_i^{n+1} = V_i P$ and $v_i^{n+1} = V_i^n$, there is an equilibrium vector V such that V*P=V. Vi gives the Eigenvalues.

Dolayısıyle aşağıdaki iki kural yazılır:

1) n nin yeterince daha büyük değerleri için, v_i^n olasılık vektörleri bütün o değerleri için aynıdır ve değişmez.
2) $v_i^{n+1} = V_i P$ ve $v_i^{n+1} = V_i^n$ olduğundan V*P=V eşitliğini sağlayan bir V denge vektörü vardır. $Vi, Özdeğerleri\ verir.$

The vector V contains the probabilities involving the equilibrium conditions.
V vektörü denge durumu koşullarını içeren olasılıkları kapsar.

$\sum_{j=1}^{m} P_i^j = 1$, this relation provides the analytical method to obtain these values. Since V is a probability vector,

$$\sum_{j=1}^{m} v_i^j = 1$$

The relation is valid. From condition (2), [v_1 v_2 v_3 .....v_m]*P=[v_1 v_2 v_3 .....v_m] is written.
Bağıntısı geçerlidir.(2) nolu koşuldan [v_1 v_2 v_3 .....v_m]*P=[v_1 v_2 v_3 .....v_m] yazılır.

By multiplying the row vector on the left side with the transition matrix P, a row vector is found again and each element is written to the vector elements on the right side of the equation to be equal, and (m) equations are obtained. With the condition that the sum of the probabilities is equal to 1, (m) unknowns can be solved from (m+1) equations. However, one of the (m+1) equations is eliminated and does not join the equation set.
Sol tarafta bulunan satır vektörü ile P geçiş matrisi çarpımından yine bir satır vektörü bulunarak eşitliğin sağ tarafındaki vektör elemanlara her bir eleman eşit olması yazılarak (m) adet denklem elde edilir. Olasılıkların toplamıın 1 e eşit olma şartı ile de (m) bilinmeyen, (m+1) denklemden çözülebilir. Fakat (m+1) adet denklemden biri elimine edilerek denklem takımına katılmaz.

Example:

Örnek:

$$P=\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$$

Let's apply the specified rules to our example.

Örneğimize belirlenen kuralları uygulayalım.

$$\sum_{j=1}^{3} v_1 = 1 \qquad v_1 + v_2 + v_3 = 1$$

$$[v_1 \;\; v_2 \;\; v_3] * P = [v_1 \;\; v_2 \;\; v_3]$$

$$0.4v_1 + 0.2v_2 + 0.25v_3 = v_1$$

$$0.3v_1 + 0.5v_2 + 0.25v_3 = v_2$$

$$0.3v_1 + 0.3v_2 + 0.5v_3 = v_3$$

But one of the (3) equations is eliminated.

By solving the equation system with three unknowns; v , v , v is found.

Fakat (3) adet denklemden biri elimine edilir.

Üç bilinmeyenli denklem sisteminin çözümü ile; $v_1 = 0.273$, $v_2 = 0.352$, $v_3 = 0.375$ bulunur.

Example:

Örnek:

Find the equilibrium conditions using the transition matrix below.

Aşağıdaki geçiş matrisinden hareketle denge durumu koşullarını bulunuz.

$$P=\begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

V=[v v v], It is desired to find the balance vector. Therefore,

V=[v_1 v_2 v_3] denge vektörünün bulunması istenmektedir.O halde,

$$v_1+v_2+v_3=1$$

$$0.2v_1+0.1v_2+0.5v_3=v_1$$
$$0.5v_1+0.6v_2+(0)v_3=v_2$$
$$0.3v_1+0.3v_2+0.5v_3=v_3$$

The above equations are written. The second equation is eliminated and

Yukarıdaki denklemler yazılır. İkinci eşitlik elimine edilerek

$$v_1+v_2+v_3=1$$
$$0.5v_1-0.4v_2=0$$
$$0.3v_1+0.3v_2-0.5v_3=0$$

is written and solution

yazılır ve çözüm

$$v_1=0.282$$
$$v_2=0.352$$
$$v_3=0.366$$

is found as . Therefore, the balance vector is: V=[0.282 0.352 0.366].

olarak bulunur. Dolayısiyle denge vektörü: V=[0.282 0.352 0.366] olur.

Example

Örnek:

After manufacturing a standard product on a machine tool, the probability that the following product will be of the same quality is 0.9, and after manufacturing a defective product, the probability that the following product will be standard is 0.8. Establish the transition matrix and examine the process assuming that the results of each stage in the production process will depend on a preliminary result.

Bir imalat tezgahında standart bir mamul imalinden sonra, takip eden mamulünde aynı kalitede olma olasılığı 0.9 ve hatalı bir mamul imalinden sonra takip eden mamulün standart olma olasılığı 0.8 dir. Üretim sürecinde her kademe sonuçları bir ön sonuca bağlı olacağı varsayımı ile geçiş matrisini kurunuz ve süreci irdeleyiniz.

A standard product production is indicated by S_1, a defective product production is indicated by S_2.

Standart bir mamul üretimi S_1 , hatalı bir mamul üretimi S_2 ile gösterilerek

$$P=\begin{bmatrix}0.9 & 0.1\\ 0.8 & 0.2\end{bmatrix}$$

The transition matrix is found. The transition matrix is obtained by taking the powers of the

Geçiş matrisi bulunur.Geçiş matrisi kuvvetleri alınarak

$$P^2=\begin{bmatrix}0.9 & 0.1\\ 0.8 & 0.2\end{bmatrix}*\begin{bmatrix}0.9 & 0.1\\ 0.8 & 0.2\end{bmatrix}=\begin{bmatrix}0.89 & 0.11\\ 0.88 & 0.12\end{bmatrix}$$

$$P^3=P^2P=\begin{bmatrix}0.89 & 0.11\\ 0.88 & 0.12\end{bmatrix}*\begin{bmatrix}0.9 & 0.1\\ 0.8 & 0.2\end{bmatrix}=\begin{bmatrix}0.889 & 0.111\\ 0.888 & 0.112\end{bmatrix}$$

$$P^4=P^3P=\begin{bmatrix}0.889 & 0.111\\ 0.888 & 0.112\end{bmatrix}*\begin{bmatrix}0.9 & 0.1\\ 0.8 & 0.2\end{bmatrix}=\begin{bmatrix}0.8889 & 0.1111\\ 0.8888 & 0.1112\end{bmatrix}$$

It is found and by continuing the calculations it can be shown that the equilibrium vector of the process is V=(8/9,1/9). Then the given process is a regular Markov chain.

Bulunur ve hesaplara devam ederek sürecin denge vektörünün V=(8/9,1/9) olduğu gösterilebilir. O halde verilen süreç düzenli bir Markov zinciridir.

The probability of the process going to state S in a long period is 8/9 and the probability of entering state S is 1/9. In other words, 8/9 of the production in a long period will be standard product and 1/9 will be defective product. For example, in a batch of 999 units, 888 products will be standard and 111 products will be defective. Eigenvalues can be

interpreted in this way. The trajectory of the system is determined with the help of eigenvectors.

Sürecin uzun bir dönemde S_1 durumuna gitme olasılığı 8/9 ve S_2 durumuna girme olasılığı 1/9 dur. Diğer bir deyişle uzun bir dönemde üretimin 8/9 u standart mamul, 1/9 u hatalı mamul olacaktır. Örneğin 999 adetlik bir parti üretiminde 888 mamul standart, 111 mamul ise hatalı olur. Özdeğerler bu biçimde yorumlanabilir. Öz vektörler yardımıyla sistemin izlediği yörünge belirlenir.

Solid Stochastic Matrix:

Katı Stokastik Matris:

$$P=\begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

(m) is the transition matrix P to represent the number of states.

(m) durum sayısını göstermek üzere P geçiş matrisi

$$\sum_{i=1}^{m} P_{ij} = \sum_{j=1}^{m} P_{ij} = 1$$

If it satisfies the relation, it is called a solid stochastic matrix. The sum of the values in the row and column is equal to 1. In solid stochastic matrices, the balance vector components are v for j values in all matrices. In the example, the balance vector is,

bağıntısını sağlarsa katı stokastik matris adını alır. Satırdaki ve sütündaki değerlerin toplamı 1'E eşittir. Katı stokastik matrislerde denge vektörü bileşenleri bütün matrislerde j değerleri için $\mathbf{V}_j = 1/m$ dir.Örnekte denge vektörü,

It becomes V=[v_1 v_2 v_3], $v_1 = v_2 = v_3 = 1$,V=(1/3 1/3 1/3).

Example:
Örnek:

$$P=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$$

The balance vector of the transition matrix P is found as V=(1/2 1/2) from the relation VP=V. If the powers of P are taken
P geçiş matrisinin denge vektörü, VP=V bağıntısından V=(1/2 1/2) olarak bulunur. P nin kuvvetleri alınırsa

n=for odd number, $P^n=P$
n=tek sayı içim, $P^n=P$

n=for even number, $P^n=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$

n=çift sayı için, $P^n=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$

is obtained and since P has zero in its powers, the chain is not regular.
bulunur ve P nin kuvvetlerinde sıfır bulunduğu için zincir düzenli değildir.

8.6. Absorbent Markov Chains

Emici Markov Zincirleri

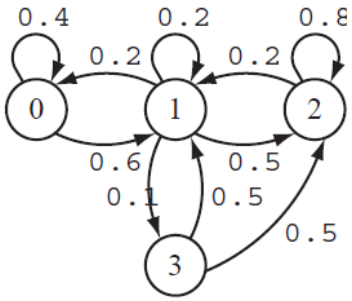
Not all Markov chains are regular. In fact, some of the most important life science applications of Markov chains do not involve transition matrices that are regular. When we use the ideas of Markov chains to model living organisms, the common state is death. Once an organism enters this state, it is impossible to leave. A type of Markov chain commonly used in the life sciences is called an absorbing Markov chain.

Tüm Markov zincirleri düzenli değildir. Aslında, Markov zincirlerinin en önemli yaşam bilimleri uygulamalarından bazıları düzenli olan geçiş matrislerini içermez. Markov zincirlerinin fikirlerini canlı organizmaları modellemek için kullandığımızda, ortak durum ölümdür. Organizma bu duruma girdiğinde, ayrılması mümkün değildir. Yaşam bilimlerinde yaygın olarak kullanılan bir Markov zinciri türü, bir emici Markov zinciri olarak adlandırılır.

Example:

Örnek:

Find the state transition matrix **P** for the Markov chain below.



$$\mathbf{P} = \begin{bmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.2 & 0.2 & 0.5 & 0.1 \\ 0 & 0.2 & 0.8 & 0 \\ 0 & 0.5 & 0.5 & 0 \end{bmatrix}$$

a) a) Is the transition matrix regular (ergodic)?
Geçiş matrisi düzenli (ergodik) mi?

Since the elements of the transition matrix P^4 from P^2 to P are positive or nonzero, the Markov chain given by the transition matrix P is regular. Regular Markov chains are ergodic.

P geçiş matrisinin P^2 den P^4 kuvvetinin elemanları pozitif veya sıfırdan farklı olduğundan P geçiş matrisi ile verilen Markov zinciri düzenlidir. Düzenli Markov zincirleri ise ergodiktir.

An ergodic chain describes a process in which it is mathematically possible to move from any state to all other states. In this definition, there is no chance of finding a complete step, but a state must be reached regardless of the starting state.
Ergodik zincir, matematik olarak herhangi bir durumdan diğer bütün durumlara geçişin mümkün olduğu bir süreci tanımlar. Bu tanımda tam bir adımda bulunma şansı yoktur, ama başlama durumuna bakmadan bir duruma erişilmiş olmalıdır.

The limiting case of the ergodic Markov chain is a regular chain. A regular chain requires that the elements in the powers of the P transition probability matrix be nonzero and positive. If the powers of the ergodic chain are taken or if the powers are taken until there are no zero elements left in the matrix, it is seen that the ergodic chain is regular.
Ergodik markov zincirini sınırlayan hal düzenli (=reguler) zincirdir. Düzenli zincir, P geçiş olasılıkları matrisinin kuvvetlerinde bulunan elemanların sıfırdan farklı ve pozitif olmasını gerektirir. Ergodik zincirin kuvvetleri alınırsa veya matriste sıfır eleman kalmayıncaya kadar kuvvetler alınırsa ergodik zincirin düzenli olduğu görülür.

b) b) Determine whether the following transition matrix is regular or ergodic.
Aşağıdaki geçiş matrisi düzenli mi ve ergodik mi beirleyiniz.

$$B = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Suppose that the Markov chain has the following transition matrix. Draw the transition diagram. If the absorber is a Markov chain, which situation is fatal?

c) Varsayalım ki, Markov zinciri aşağıdaki geçiş matrisine sahip olsun. Geçiş diyagramını çiziniz. Emici bir Markov zinciri ise hangi durum ölümcüldür.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0 & 1 & 0 \\ 0.6 & 0.2 & 0.2 \end{bmatrix} \end{array} = P.$$

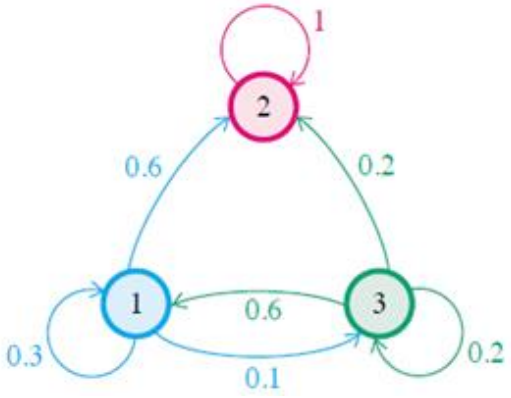
Not all Markov chains are regular. In fact, some of the most important life science applications of Markov chains do not involve transition matrices that are regular. A type of Markov chain commonly used in the life sciences is called an absorbing Markov chain. When we use the ideas of Markov chains to model living organisms, a common situation is death. Once an organism enters this state, it is impossible to separate.

Tüm Markov zincirleri düzenli değildir. Aslında, Markov zincirlerinin en önemli yaşam bilimleri uygulamalarından bazıları düzenli olan geçiş matrislerini içermez. Yaşam bilimlerinde yaygın olarak kullanılan bir Markov zinciri türü, bir emici Markov zinciri olarak adlandırılır. Markov zincirlerinin fikirlerini canlı organizmaları modellemek için kullandığımızda, ortak bir durum ölümdür. Organizma bu duruma girdiğinde, ayrılması mümkün değildir.

For example, suppose a Markov chain has transition matrix

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0 & 1 & 0 \\ 0.6 & 0.2 & 0.2 \end{bmatrix} \end{array} = P.$$

The matrix shows that p_{12}, the probability of going from state 1 to state 2, is 0.6, and that p_{22}, the probability of staying in state 2, is 1. Thus, once state 2 is entered, it is impossible to leave. For this reason, state 2 is called an *absorbing state*. Figure 4 shows a transition diagram for this matrix. The diagram shows that it is not possible to leave state 2.



8.7. Orbit probability

Yörünge olasılığı

Recall that a trajectory is a sequence of values for X_0, X_1, . . . , X_n Because of the Markov property, we can find the probability of any trajectory by multiplying the initial probability and all subsequent one-stage probabilities. The sum of the initial probability values of all states is equal to one.

Bir yörüngenin X_0, X_1, . . . , X_n için bir değerler dizisi olduğunu hatırlayın. Markov özelliğinden dolayı, herhangi bir yörüngenin olasılığını başlangıç olasılığını ve sonraki tüm tek aşamalı olasılıkları çarparak bulabiliriz. Tüm durumların başlangış anındaki olasılık değerleri toplamı bire eşittir.

For each case in the orbit, the initial probability value and orbits are given. Orbits must be continuous. There can be no broken orbits. If the orbit probability is given, the initial probability value is calculated.

Yöründe her bir durum için başlangıç olasılık değeri ve yörüngeler verilir. Yörüngeler sürekli olmak zorundadır. Kopuk yörünge olamaz. Eğer yörünge olasılığı verilmiş ise başlangıç olasılık değeri hesaplanır.

Markov Model: Sequence Prob.

- Conditional probability

$$P(A,B) = P(A \mid B)P(B)$$

- Sequence probability of Markov model

Chain rule

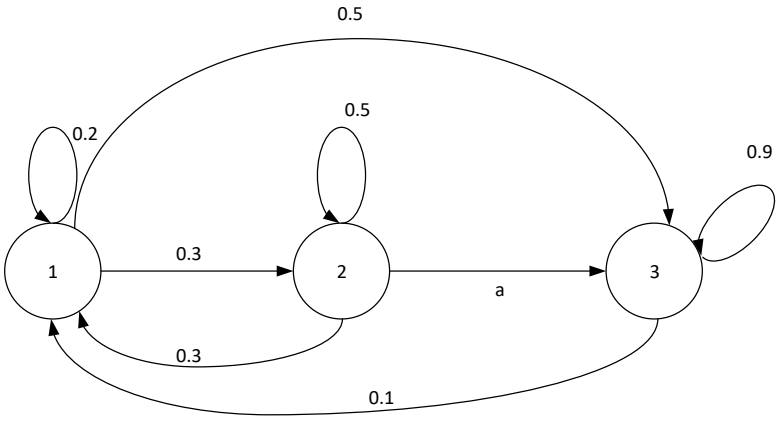
$$P(q_1, q_2, \cdots, q_T)$$
$$= P(q_1)P(q_2 \mid q_1)\cdots P(q_{T-1} \mid q_1, \cdots, q_{T-2})P(q_T \mid q_1, \cdots, q_{T-1})$$
$$= P(q_1)P(q_2 \mid q_1)\cdots P(q_{T-1} \mid q_{T-2})P(q_T \mid q_{T-1})$$

1st order Markov assumption

The value of Pi is the probability of falling into the selected state, given at the beginning.

Pi değeri seçilen duruma düşme olasılığıdır, başlangıçta verilir.

Example:
Örnek:



a) How many states are there in the flow chart above? What is the sum of the transition probabilities from each state to the other states?
Yukarıdaki akış diyagramında kaç durum vardır. Her bir durumdan diğer durumları geçiş olasılıkları toplamı neye eşittir?
There are 3 cases. The sum of the transition probabilities is equal to 1.
3 durum vardır. Geçiş olasılıkları toplamı 1'e eşittir.

b) The sum of the transition probabilities from each state to the other states is equal to 1. Taking this explanation into account, write the sum of the transition probabilities from the 2nd state to the other state, calculate the value of a
Her bir durumdan diğer durumları geçiş olasılıkları toplamı 1'e eşittir. Bu açıklamayı dikkate alarak 2.durumdan diğer duruma geçiş olasılıkları toplamını yazınız, a değerini hesaplayınız
P21+P22+P31=1
0.3+0.5+a=1, a=0.2

c) Create the transition matrix A.
A, geçiş matrisini oluşturun.

$$A=\begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0 & 0.9 \end{pmatrix}$$

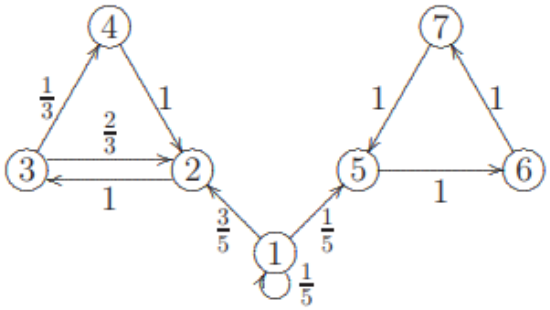
d) Jafar's starting stage Pstart falls into state 1 with probability 0.4, into state 2 with probability 0.5, into state 3 with probability 0.1. After Jafar falls into state 2, what is the probability that he follows the following route? Pcafer=Pstart*P23*P32*P21*P13

Cafer başlama aşaması olan Pstart 0.4 olasılıkla 1.duruma, 0.5 olasılıkla 2.duruma, 0.1 olasılıkla 3.duruma düşüyor. Cafer 2.duruma düştükten sonra, aşağıdaki rotayı takip etme olasılığı nedir? Pcafer=Pstart*P23*P32*P21*P13

Pcafer=0.5*0.2*0*0.3*0.5=0

Example:

Örnek:



For each case, the initial probabilities X0 ~ (3 /4 , 1/2, 1/ 4 , 2/3, 1/5, 1/4, 1/8 are given. What is the probability that the orbit 1, 2, 3, 2, 3, 4 will be followed?

Her bir durum için başlangıç olasılıkları X0 ~ (3 /4 , 1/2, 1/ 4 , 2/3, 1/5, 1/4, 1/8 verilmiştir. 1, 2, 3, 2, 3, 4 yörüngesinin izlenme olasılığı olasılığı nedir?

P(1,2,3,2,3, 4)=P(X_0 = 1) x p12 x p23 x p32 x p23 x p34

=3/4 x 3/5 x 1 x 2/3 x 1 x 1/3=1/10

Example:

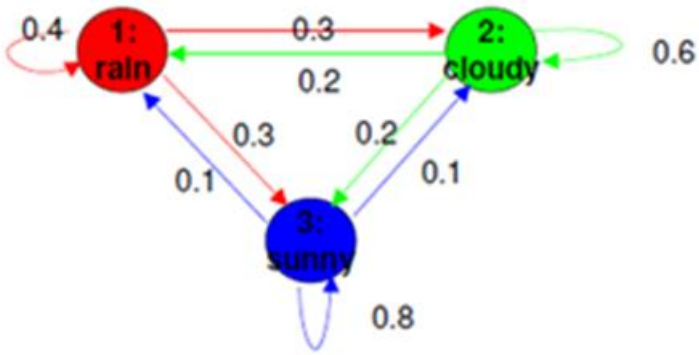
Örnek:

What is the probability that the weather for the next 7 days will be "sun-sun-rain-rain-sun-cloudy-sun" when today is sunny?

$S_1 : rain, \quad S_2 : cloudy, \quad S_3 : sunny$

$$
\begin{aligned}
P(O \mid \text{model}) &= P(S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3 \mid \text{model}) \\
&= P(S_3) \cdot P(S_3 \mid S_3) \cdot P(S_3 \mid S_3) \cdot P(S_1 \mid S_3) \\
&\quad \cdot P(S_1 \mid S_1) P(S_3 \mid S_1) P(S_2 \mid S_3) P(S_3 \mid S_2) \\
&= \pi_3 \cdot a_{33} \cdot a_{33} \cdot a_{31} \cdot a_{11} \cdot a_{13} \cdot a_{32} \cdot a_{23} \\
&= 1 \cdot (0.8)(0.8)(0.1)(0.4)(0.3)(0.1)(0.2) \\
&= 1.536 \times 10^{-4}
\end{aligned}
$$

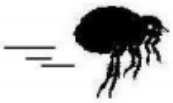
π=1 indicates the initial state as Sunny. Today is sunny.
π=1 başlangıç durumunu Güneşli olarak gösterir. Bugün güneşli.

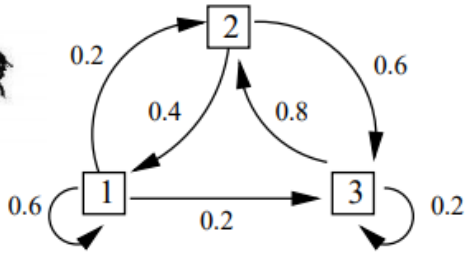


Example:
Örnek:

Worked Example: distribution of X_t and trajectory probabilities

Purpose-flea zooms around the vertices of the transition diagram opposite. Let X_t be Purpose-flea's state at time t $(t = 0, 1, \ldots)$.





(a) Find the transition matrix, P.

Answer: $P = \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.4 & 0 & 0.6 \\ 0 & 0.8 & 0.2 \end{pmatrix}$

(b) Find $\mathbb{P}(X_2 = 3 \,|\, X_0 = 1)$.

$$\mathbb{P}(X_2 = 3 \,|\, X_0 = 1) = \left(P^2\right)_{13} = \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} \cdot & \cdot & 0.2 \\ \cdot & \cdot & 0.6 \\ \cdot & \cdot & 0.2 \end{pmatrix}$$

$$= 0.6 \times 0.2 + 0.2 \times 0.6 + 0.2 \times 0.2$$
$$= 0.28.$$

Note: we only need one element of the matrix P^2, so don't lose exam time by finding the whole matrix.

From the current state X_0=1, the next state X_2=3, (P2)13 is calculated. How to go from state 1 to state 3 in two steps? There are 3 ways to go.
Şu an X_0=1 durumdan, iki sonraki durum olan X_2=3 ise $(P^2)_{13}$ hesaplanır. İki adımda 1 durumundan 3 durumuna nasıl gidilir? 3 yoldan gidilir.

S11 * S13 = 0.6*0.2
S13 * S33 =0.2*0.2
S12 * S23 =0.2*06

Total=0.28
Toplam=0.28

Example:
Örnek:

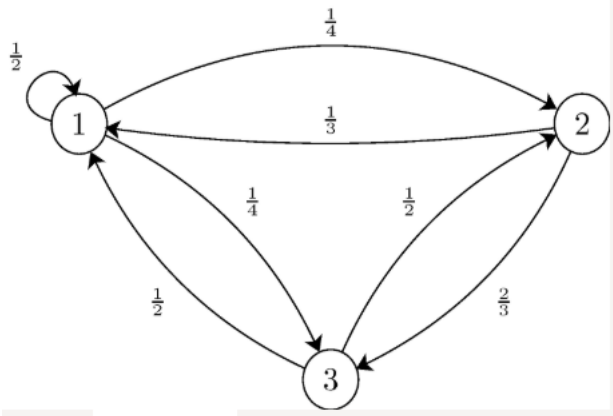
Consider the Markov chain with three states, $S = \{1, 2, 3\}$, that has the following transition matrix

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}.$$

a. Draw the state transition diagram for this chain.
b. If we know $P(X_1 = 1) = P(X_1 = 2) = \frac{1}{4}$, find $P(X_1 = 3, X_2 = 2, X_3 = 1)$.

While it is in state S3 in step $\boldsymbol{X}1$, it will go to state S2 in step $\boldsymbol{X}2$, and finally it will go to state S1 in step $\boldsymbol{X}3$.
$\boldsymbol{X}_1$ adımında S3 durumunda iken, $\boldsymbol{X}_2$ adımından S2 durumuna gidecek, sonunda $\boldsymbol{X}_3$ adımında S1 durumuna gidilecek.

a)



A state transition diagram.

b) Since the orbit is S3, S2, S1, the initial condition for states S1 and S2 is known, the initial condition for S3 is calculated first. The sum of the probability values at the initial moment of all states is equal to one.
Yörüngesi S3, S2, S1 olduğuna göre S1 ve S2 durumları için başlangıç koşulu bilindiğinden önce S3 için başlangıç koşulu hesaplanır. Tüm durumların başlangıç anındaki olasılık değerleri toplamı bire eşittir.

$$
\begin{aligned}
P(X_1 = 3) &= 1 - P(X_1 = 1) - P(X_1 = 2) \\
&= 1 - \frac{1}{4} - \frac{1}{4} \\
&= \frac{1}{2}.
\end{aligned}
$$

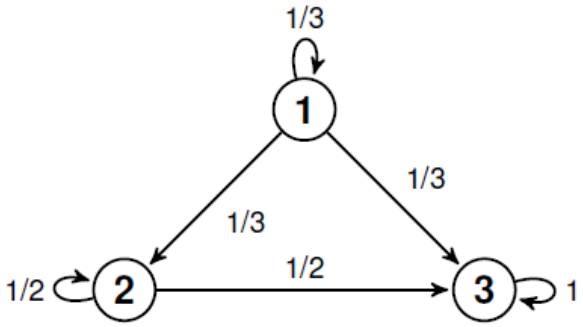
Calculating the probability of case-3, followed by case-2, followed by case-1,
Durum-3, ardından durum-2 ardından durum-1 gelme olasılığının hesaplanması,

$$
\begin{aligned}
P(X_1 = 3, X_2 = 2, X_3 = 1) &= P(X_1 = 3) \cdot p_{32} \cdot p_{21} \\
&= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} \\
&= \frac{1}{12}.
\end{aligned}
$$

Example:
Örnek:
Consider the three-state discrete-time Markov chain corresponding to the transition diagram in Fig. Suppose that the initial distribution of X(0) is given by f(1) = f(2) = 1/2.
Şekildeki geçiş diyagramına karşılık gelen üç durumlu ayrık zamanlı Markov zincirini düşünün. X (0) 'ın başlangıç dağılımının f (1) = f (2) = 1/2 ile verildiğini varsayın.

Compute the following

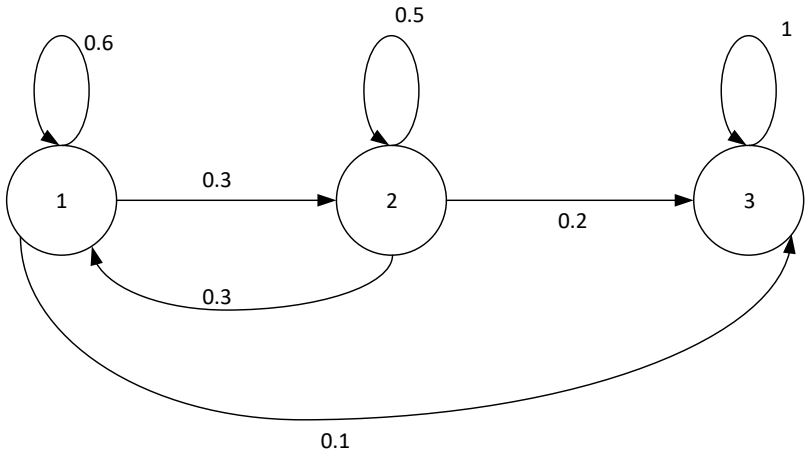
1. P(X(0) = 1, X(1) = 2, X(2) = 3).
2. P(X(2) = i), for i = 1, 2, 3.
3. P(T3 = 2) where



P(X(0) = 1, X(1) = 2, X(2) = 3) = f(1) * P12 * P23 = $\frac{1}{2} * \frac{1}{3} * \frac{1}{2} = \frac{1}{12}$

Example:
Örnek:

In the Markov chain analysis whose transition diagram is given below,
Aşağıda geçiş diyagramı verilen Markov zincir analizinde,



a) Write the state transition matrix.
Durum geçiş matrisini yazınız.

$$P = \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0 & 0 & 1 \end{pmatrix}$$

b) What properties must this matrix have in order to be a transition matrix? Does this matrix satisfy these properties?
c) Bu matris geçiş matrisi olabilmesi için hangi özelliklere sahip olması gerekir? Bu matris bu özellikleri sağlıyor mu?

For a matrix to be a transition matrix, each element value must be in the range $0 \leq P_{ij} \leq 1$
The sum of the values of the elements in a row must be equal to 1.
The number of rows of the matrix must be equal to the number of columns.
Yes, the matrix is a transition matrix.
Bir matrisin geçiş matrisi olabilmesi için her bir eleman değeri $0 \leq P_{ij} \leq 1$ **aralığında olmalıdır.**
Bir satırdaki elemanların değerlerinin toplamı 1'e eşit olmalıdır.
Matrisin satır sayısı sütun sayısına eşit olmalıdır.
Evet matris geçiş matrisidir.

d) What is the probability of going from state 1 to state 3, P13, and the probability of going from state 3 to state 2, P32?
1 durumundan 3 durumuna geçme olasılığı, P13 ve 3 durumundan 2 durumuna geçme olasılığı, P32 nedir?
P13=0.1, P32=0

e) In the Markov chain analysis whose transition diagram is given below,
Geçiş diyagramı aşağıda verilen Markov zincir analizinde,

Calculate the probability of following the orbit P(Orbit) = Sto2*D22*D21*D13*D32. The probability of transition from the initial state to state 2 will be taken as Sto2=0.4.
P(Orbit) = Sto2*D22*D21*D13*D32 =0.4 * 0.5 * 0.3 * 0.1 * 0.0 =0.0
The probability of following this orbit is 0%.
P(Yörünge) = Sto2*D22*D21*D13*D32 yörüngesini takip etme olasılığını hesaplayınız.
Başlangıç durumundan 2 durumuna geçiş olasılığı olan, Sto2=0.4 alınacaktır.
P(Yörünge) = Sto2*D22*D21*D13*D32 =0.4 * 0.5 * 0.3 * 0.1 * 0.0 =0.0
Bu yörüngeyi takip etme olasılığı % 0 dır.

f) What condition must a matrix satisfy for it to be a regular chain or ergodic? The P^8 expression of a transition matrix P is given below. Is this matrix ergodic?
Bir matrisin düzenli zincir ya da ergodik olması için hangi koşulu sağlaması gerekir? Bir P geçiş matrisinin P^8 ifadesi aşağıda verilmiştir. Bu matris ergodik midir?

$$P^8 = \begin{pmatrix} x & x & x \\ x & x & x \\ x & 0 & x \end{pmatrix}$$

The regular chain requires that the elements in the powers of the transition probability matrix P be nonzero and positive. This matrix is not ergodic.
Düzenli zincir, P geçiş olasılıkları matrisinin kuvvetlerinde bulunan elemanların sıfırdan farklı ve pozitif olmasını gerektirir. Bu matris ergodik değildir.

g) In the Markov chain analysis, the state transition matrix is given below, the probability of transition to each state in the initial state is given as the column vector Vs. The product of the initial state transition matrix Vs1=P*Vs1 is given below. What is the probability of transition from the initial state to state 2?
Durum geçiş matrisi aşağıda verilmiş olan Markov zincir analizinde başlangıç durumunda her bir duruma geçiş olasılığı sütun vektörü Vs olarak verilmiştir. Başlangıç durumları geçiş matrisinin çarpımı Vs1=P*Vs1 aşağıda verilmiştir. Başlangıç durumundan 2 durumuna geçiş olasılığı nedir?

$$P = \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$

$$V_s = \begin{pmatrix} 0.2 \\ 0.5 \\ 0.3 \end{pmatrix}$$

$$V_{s1} = \begin{pmatrix} 0.23 \\ 0.35 \\ 0.23 \end{pmatrix}$$

Ps-2=0.35

8.8. Markov zincir modeli ile Bayes Metotunun Bütünleştirilmesi

Bayes Method: In order to calculate which unit an outlier comes from in the set where the outliers from each unit are collected, the probability of the outlier intersection set of each unit in the set where the outliers are collected is first determined. The following relation is used to calculate the probability of which unit a selected outlier originates from. The probability of the intersection with the unit where the outlier comes from is divided by the total probability of all intersections in the outlier set.

Bayes Metotu: Her bir birimden aykırı olanların toplandığı kümede herhangi bir aykırı olanın hangi birimden geldiği hesaplayabilmek:için Önce aykırı olanların topladığı kümede her bir birimin aykırılık kesişim kümesinin olasılığı belirlenir. Seçilen bir aykırılığın hangi birimden kaynaklandığının olasılığını hesaplayabilmek için aşağıdaki bağıntı kullanılır. Aykırı olanın geldiği birim ile kesişim olasılığı, aykırılık kümesindeki tüm kesişimlerim olasılık toplamına bölünür.

$$P(A\ /H) = \frac{P(A \cap H)}{P(A \cap H) + P(B \cap H)}$$

The probability of the outlier intersecting with the unit it came from is equal to the product of the probability of the unit and the probability of the outliers in the unit.

Aykırı olanın geldiği birim ile kesişim olasılığı, birimin olasılığı ile birimdeki aykırı olanların olasılığının çarpımına eşittir.

$$P(A \cap H) = P(A) * P(H\ /A)$$
$$P(B \cap H) = P(B) * P(H\ /B)$$

For example, there are 2 machines in a factory. The daily production rates of each machine are given as A=50%, B=50%. The percentage of defective parts produced is given as A=10%, B=5%. Defective parts are produced independently of each other in machines A and B.

Örneğin, bir fabrikada 2 adet makine bulunmaktadır. Her bir makinenin günlük üretim oranları A=%50, B=%50 olarak verillmiştir. Üretilenlerden kusurlu olanların yüzdesi ise A=%10, B=%5 olarak verilmiştir. Kusurlu parçalar birbirinden bağımsız olarak A, B makinelerinde üretilmektedir.

- Calculate the production probabilities in each production line. P(A)=0.5, P(B)=0.5.
  Her bir üretim bandında üretim olasılıkları hesaplayınız. P(A)=0.5, P(B)=0.5.

- Calculate the probability of error originating from A and B for each production line.
  Her bir üretim bandı için A dan ve B den kaynaklanan hata olasılıklarını hesaplayınız. P(H/A)=0.10

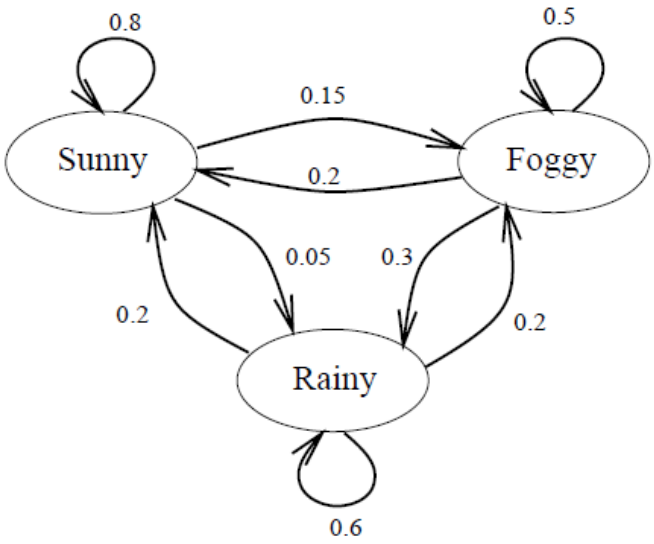
P(H/B)=0.05.

- Calculate the probability of this defective motor coming out of machine A, B. If the probability of being defective is P(T)
  Bu kusurlu motorun A, B makinesinden çıkma olasılıklarını hesaplayınız. Kusurlu olma olasılığı P(T) ise
  P(A∩H)=P(A)*P(H/A)=0.05
  P(B∩H)=P(B)*P(H/B)=0.025
  P(T)= P(A∩H)+ P(B∩H)=0.075

- The probability of the faulty items in A among the total faulty items,
- Toplam hatalı olanlardan A daki hatalı olanların olasılığı,
  P(A/H)= P(A∩H)/P(T)=0.05/0.075=50/75=2/3.

- The probability of the number of errors in B out of the total errors,
  Toplam hatalı olanlardan B daki hatalı olanların olasılığı,
  P(B/H)= P(B∩H)/P(T)=0.025/0.075=25/75=1/3

Example:
Örnek:



It is the probability of the trajectory sequence of the Markov model.
Markov modelin yörünge dizisi olasılığıdır.

$$P(q_1, q_2, \ldots, q_T) = \prod_{i=1}^{T} P(q_i \mid q_{i-1})$$

Conditional probability: P(A∩B)=P(B)P(A|B)=P(A)P(B|A)
Koşullu olasılık: $P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$

$$P(q_1, q_2, \dots, q_T) = P(q_1)P(q_2 \mid q_1) \dots P(q_{T-1} \mid q_{T-2})P(q_T \mid q_{T-1})$$

a) Given that today is sunny, what is the probability that tomorrow will be sunny and the day after will be rainy?
Bugün güneşli olduğu için yarının güneşli ve sonraki günün yağmurlu olma olasılığı nedir?
P(d2= Sunny, d3=Rainy | d1=Sunny)= P(d2=suny | d1=suny)*(d3=rainy | d2=sunny)=0.8*0.05=0.04

b) If it is foggy today, what is the probability that it will rain two days from now? There are three ways for foggy conditions today to rain two days from now:
Bugünün sisli olduğu düşünülürse, bundan iki gün sonra yağmurlu olma olasılığı nedir?
Bugün sisli bir durumdan iki gün sonra yağmurlu hale gelmenin üç yolu var:
P(Foggy – Foggy – Rainy)=0.5*0.3
P(Foggy - Rainy – Rainy)=0.3*0.6
P(Foggy - Sunny – Rainy)=0.2*0.05

P(d3=rainy| d1=foggy)= P(d2=foggy | d1=foggy) * P(d3=rainy | d2=foggy) + P(d2=rainy | d1=foggy) * (d3=rainy | d2=rainy) + P(d2=sunny | d1=foggy) * (d3=rainy | d2=sunny) =0.34

8.9. Hidden Markov Models

Saklı Markov Modeli

The Markov Chain model creates an output for each state in the chain given as input. The Hidden Markov Model is also a chain model, a Probability Chain Model. Because it calculates a probability for each possible output and chooses the most probable one.
Markov Zincir modeli, giriş olarak verilen zincirdeki her duruma bir çıktı yaratır. **Hidden Markov Model** de bir zincir modelidir, **Olasılık Zincir Modelidir**. Çünkü her olası çıktı için bir olasılık hesaplar ve en olası olanı seçer.

The states are $Q = q_1, q_2, \ldots, q_N$, where q1 and qN are the Start and End states. In fact, there are N-2 states in total. The reason for it being +2 is that in the trajectory analysis, the Start and End are considered as one state each. Technically, it has to go from one state to another (a11, a12, a13) and since these are also probabilities, their sum is 1.
Durumlar, $Q = q_1, q_2, \ldots, q_N$, burada q_1 ve q_N: Başlangıç ve Bitiş durumlarıdır. Aslında toplamda N-2 adet durum vardır. +2 olmasının sebebi ise yörünge analizinde Başlangıç ve Bitiş'inde birer durum kabul edilmesidir. Teknik olarak bir durumdan herhangi bir duruma gitmek zorundadır (a11, a12, a13) ve bunlar da birer olasılık olduğu için, toplamları 1 eder.

Transition probability matrix,
Geçiş olasılık matrisi,

$$\begin{pmatrix} a11\ a12 & \cdots & a1N \\ \vdots & \ddots & \vdots \\ aN1\ aN2 & \cdots & aNN \end{pmatrix}$$

Each element in the transition matrix is the probability of transition from state i to state j.
Geçiş matrisininde her bir eleman i-durumundan j-durumuna geçiş olasılığıdır.

$$\sum_{j=1}^{N} a_{ij} = 1 \quad \forall i$$

State transitions are represented by a matrix.
Durum geçişleri bir matris ile gösterilir.

In the Hidden Markov Model, there are also states and transitions between states for another purpose. Can the probability of the desired states be calculated from the data collected for another purpose?
Saklı Markov Modelinde, başka bir amaç için durumlar ve durumlar arası geçişler de söz konusudur. Başka bir amaç için toparlanmış verilerden istenilen durumların olasılık hesabı yapılabilir mi?

History

Tarihçesi

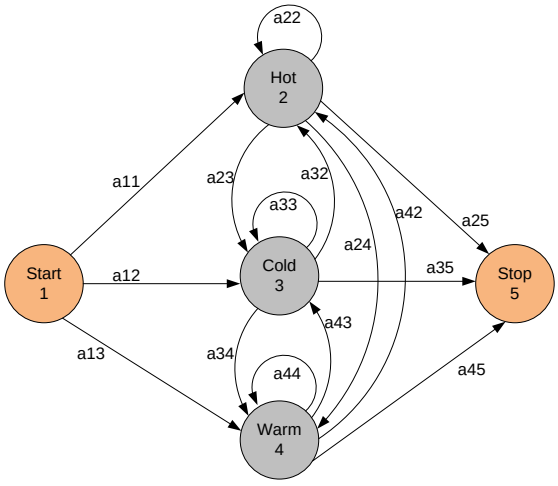
The theory of the Hidden Markov Model was developed in the 1970s by Baum and Eagon (1967), Petrie (1969) and Baum (1972). In fact, the Hidden Markov Model was studied in the 1940s, but its application could not be done because the theory was not fully developed.

Saklı Markov Modelinin teorisi 1970'li yıllarda Baum ve Eagon (1967), Petrie (1969) ve Baum (1972) tarafından geliştirilmiştir. Aslında Saklı Markov Modeli 1940'lı yıllarda çalışılmış fakat teori tam olarak geliştirilemediğinden dolayı uygulaması yapılamamıştır.

Two friends living in cities A and B talk on the phone every day about what they did that day. The one living in city B is only interested in three activities; swimming, walking and cleaning his house. As can be seen, the tasks performed are determined by the weather. On the other hand, the one living in city A does not have precise information about the weather in city B, but he knows the general directions of his friend in city B.

A ve B şehirlerinde yaşayan iki arkadaş her gün telefonla o gün ne yaptıkları hakkında konuşurlar. B şehrinde yaşayan sadece 3 faaliyetle ilgilenir; yüzmek, yürümek ve evini temizlemek. Görüleceği üzere yapılan işler hava durumuna göre belirlenmektedir. Öte yandan A şehrinde yaşayan B şehrindeki hava durumu hakkında kesin bir bilgiye sahip değildir, Fakat B şehrindeki arkadaşının genel yönelmelerini bilmektedir.

- Can a friend who cannot reach his friend guess where his friend is by looking at the weather in the city he lives in?
- Or can a friend who lives in city A guess the weather in city B based on what he does every day?
- Arkadaşına ulaşamayan arkadaşının yaşadığı şehirdeki hava durumuna bakarak arkadaşının nerede olduğunu öngörebilir mi?
- Ya da A şehrinde yaşayan B şehrindeki arkadaşının her gün ne yaptığına dayanarak, oradaki hava durumunu tahmin edebilir mi?



In the Markov Chain example in the picture, each node actually represents a state. There are 5 states in total: Start, Hot, Cold, Warm and Stop. There is a number between 1 and 5 under each state. These are considered as the index of each state. Transition from one state to another can be made. For example, the Hot → Cold transition is made with a23. Since Hot is shown with the index number 2 and Cold with the index number 3, it represents the transition from state 2 to state 3. Each transition that provides a transition between states actually has a value. In the Markov Chain, these values are probabilities: the probability of transitioning from one state to another. Therefore, the sum of all transitions from any state is 1.

Resimde Markov Zinciri örneğinde her node(boğum), aslında bir durum (**state)** belirtir. Toplamda 5 adet durum var: Start, Hot, Cold, Warm ve Stop. Her durumun altında 1 ile 5 arasında bir sayı var. Bunları her durumun indisi olarak kabul edilir. Bir durumdan diğerine geçiş yapılabilir. Örneğin Hot → Cold geçişi **a23** ile yapılır. Hot 2, Cold ise 3 indis numarası ile gösterildiği için 2. durumundan 3. durumuna geçişi temsil eder. Durumlar arası geçiş sağlayan her **geçişin** aslında bir değeri vardır. Markov Zincirinde ise bu değerler birer olasılıktır: bir durumdan diğer duruma geçme olasılığıdır. Bu nedenle *herhangi bir durumdan çıkan tüm geçişlerin toplamı 1 eder.*

Hidden Markov Model Parameters:

Saklı Markov Modeli Parametreleri:

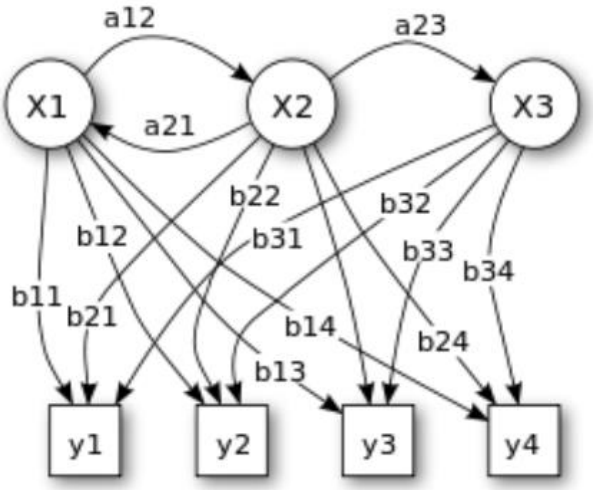
The probability parameters of the Hidden Markov Model shown in the figure below are as follows:

Aşağıdaki şekilde görülen Saklı Markov Modeli'nin olasılık parametreleri şu şekildedir:

x — states
y— possible observations
a— transition probabilities state
b— exit probabilities
x — durumlar
y— olası gözlemler
a— geçiş olasılıkları durumu
b— çıkış olasılıkları

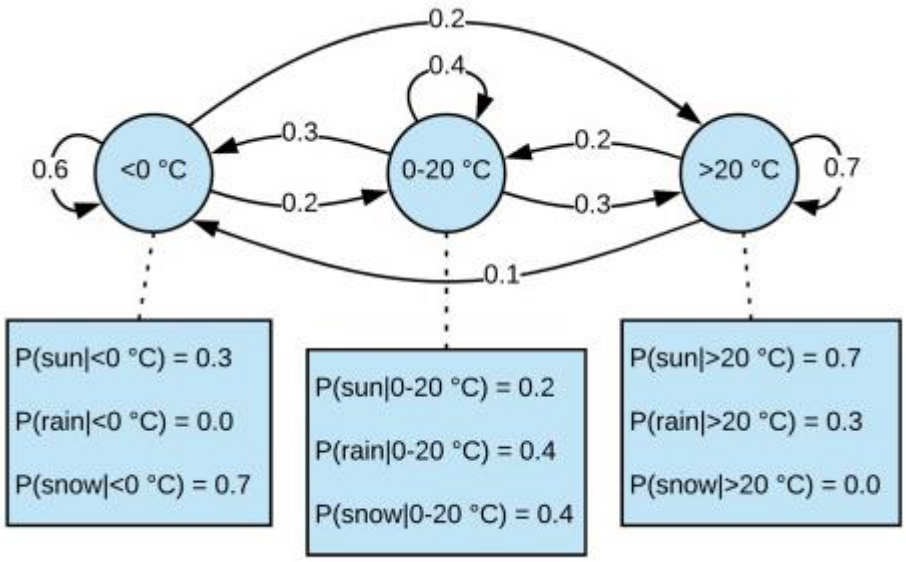
In the normal Markov Model, states are visible to the observer and so the only parameters are the state transition probabilities. In the Hidden Markov Model, state-dependent outputs are visible.

Normal Markov Modelinde, durumlar, gözlemci için görünebilirdir ve bu yüzden tek parametre, durum geçiş olasılıklarıdır. Saklı Markov Modelinde, duruma bağlı çıkışlar görünür.



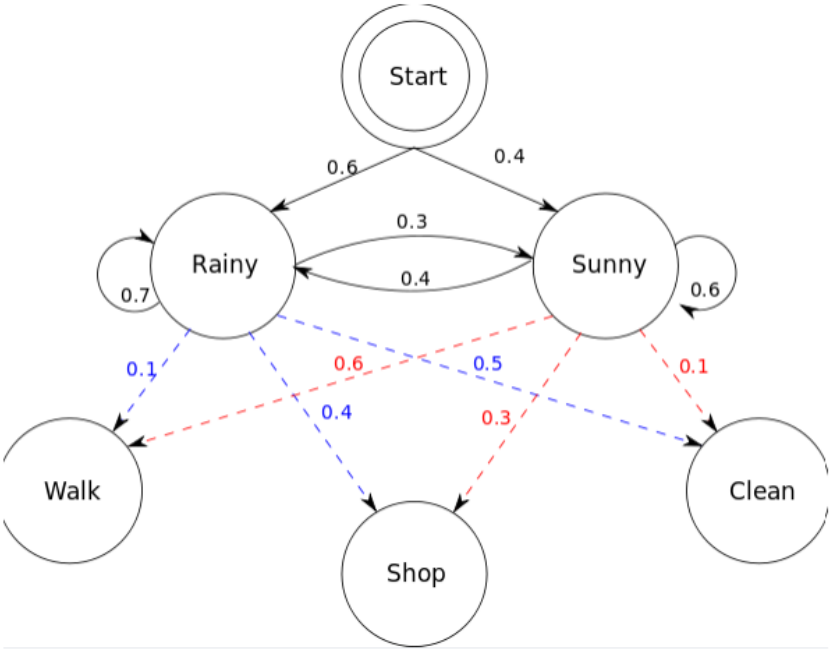
Probability Parameters of the Hidden Markov Model

Hidden Markov Modelinin Olasılık Parametreleri



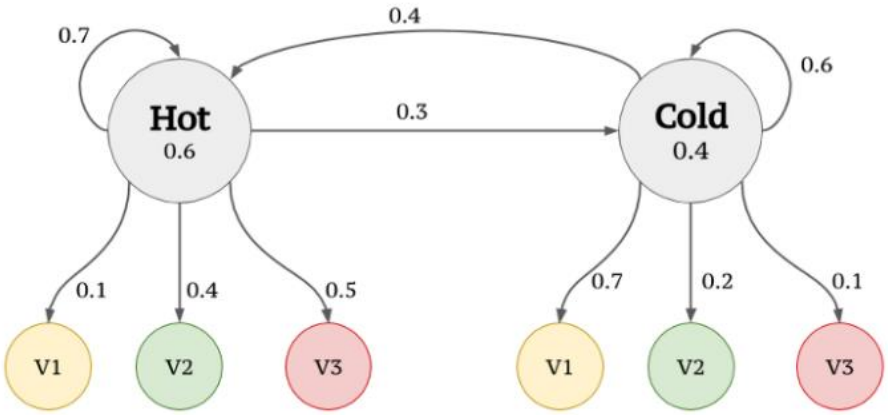
Hidden Markov models (HMMs) are a type of statistical modeling that has been used for several years. It has been applied in different fields such as medicine, computer science, and data science. The hidden Markov model (HMM) is the basis of many modern data science algorithms. It is used in data science to efficiently utilize observations for successful predictions or decision-making processes.

Gizli Markov modelleri (HMM'ler), birkaç yıldır kullanılan bir tür istatistiksel modellemedir. Tıp, bilgisayar bilimi ve veri bilimi gibi farklı alanlarda uygulanmıştır. Gizli Markov modeli (HMM), birçok modern veri bilimi algoritmasının temelidir. Başarılı tahminler veya karar verme süreçleri için gözlemlerden verimli bir şekilde yararlanmak için veri biliminde kullanılmaktadır.



Example: Calculating the termination probabilities in states V1, V2, V3

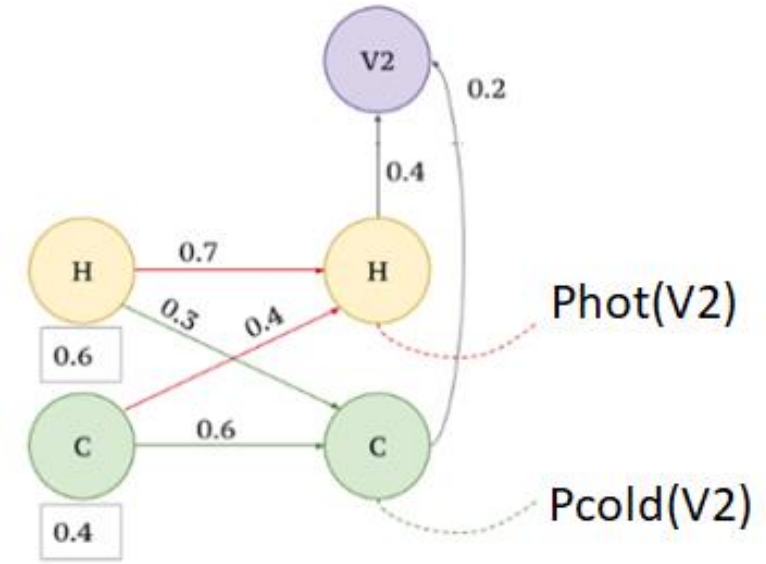
Örnek: V1, V2, V3 durumlarında sonlanma olasılıklarının hesaplanması



$$A \text{ (State Transition Matrix)} = \begin{array}{c} \\ H \\ C \end{array} \begin{array}{c} \begin{array}{cc} H & C \end{array} \\ \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix} \end{array}$$

$$B \text{ (Emission Matrix)} = \begin{array}{c} \\ H \\ C \end{array} \begin{array}{c} \begin{array}{ccc} V1 & V2 & V3 \end{array} \\ \begin{pmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{pmatrix} \end{array}$$

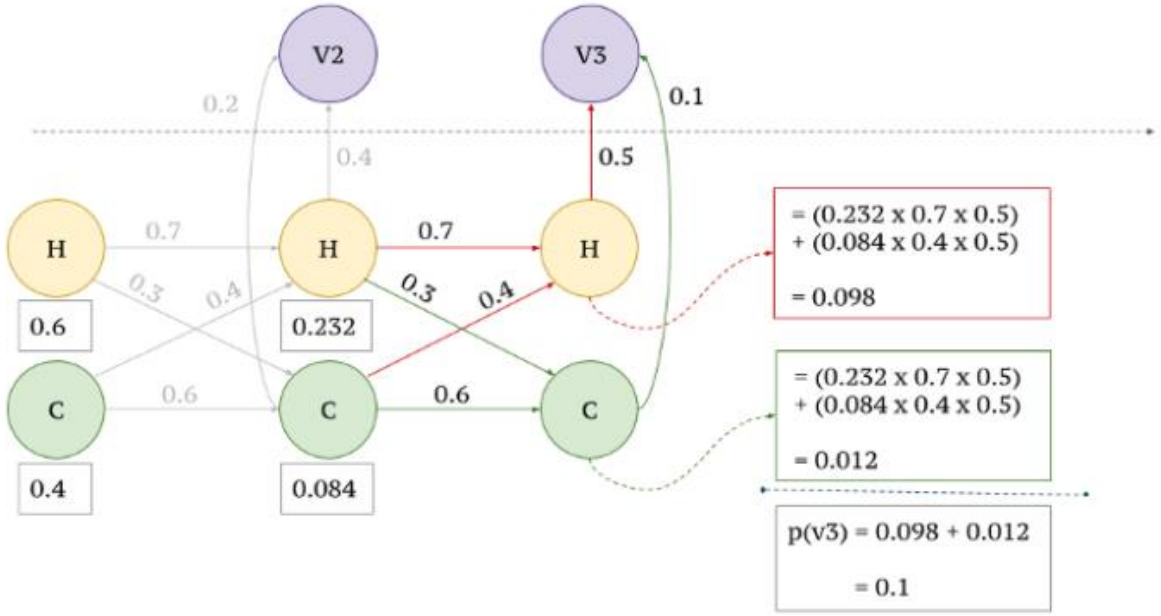
$$\underset{\text{(Initial state S0)}}{\Pi} = \begin{matrix} H & C \\ \{0.6 & 0.4\} \end{matrix}$$

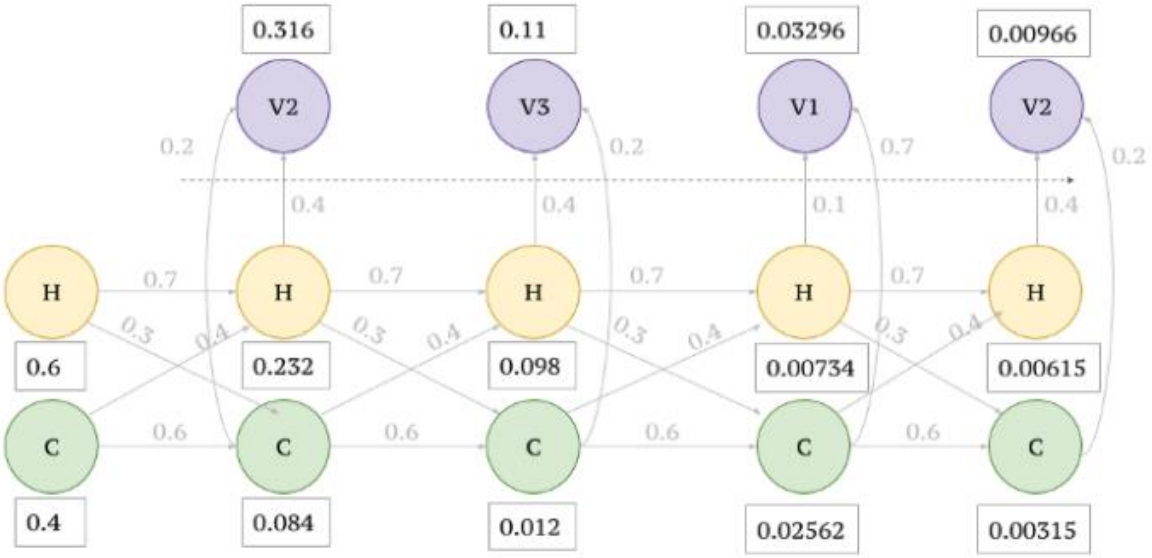


The probability of going from the initial state to the hot state is given as 0.6; the probability of going from the cold state is given as 0.4.
Başlangıç durumundan sıcak durumuna geçme olasılığı, 0.6; soğuk durumuna geçme olasılığı 0.4 olarak verilmiştir.

Let us calculate the decision probability for V2 from the hot state.
Sıcak durumdan V2 için karar verme olasılığını hesaplayalım.
Phot(V2)=0.6*0.7*0.4 + 0.4*0.4*0.4=0.232

Let us calculate the decision probability for V2 from the cold state.
Soğuk durumdan V2 için karar verme olasılığını hesaplayalım.
Pcold(V2)=0.4*0.6*0.2+0.6*0.3*0.2=0.084
P(V2)=Phot(v2) + Pcold(V2)=0.232+0.084=0.316


V2
V3
0.2
0.1
0.4
0.5
H
0.7
H
0.7
H
0.6
0.3
0.4
0.3
0.4
0.232
C
0.6
C
0.6
C
0.4
0.084
= (0.232 x 0.7 x 0.5)
+ (0.084 x 0.4 x 0.5)
= 0.098
= (0.232 x 0.7 x 0.5)
+ (0.084 x 0.4 x 0.5)
= 0.012
p(v3) = 0.098 + 0.012
= 0.1


0.316
0.11
0.03296
0.00966
V2
V3
V1
V2
0.2
0.2
0.7
0.2
0.4
0.4
0.1
0.4
H
0.7
H
0.7
H
0.7
H
0.7
H
0.3
0.4
0.3
0.4
0.3
0.4
0.3
0.4
0.6
0.232
0.098
0.00734
0.00615
C
0.6
C
0.6
C
0.6
C
0.6
C
0.4
0.084
0.012
0.02562
0.00315

Hidden Markov Model and Estimation Algorithms

Saklı Markov Modeli ve Tahmin Algoritmaları

Today, decision making under uncertainty and making future predictions are among the most fundamental problems we face. The uncertainty of a system arises from the fact that it cannot be controlled. There are basically 3 different problems: Likelihood, Decoding, Learning. In the Hidden Markov Model, which is used to interpret under uncertainty, three important algorithms are used to find the state sequence:

- Recognition Problem - Forward Algorithm
- Problem of Finding the State Sequence - Viterbi Algorithm
- Learning Model Parameters - Baum Welch Algorithm

Günümüzde, belirsizlik altında karar alma ve ileriye yönelik tahminde bulunma, karşı karşıya olduğumuz en temel sorunlardandır. Bir sistemin belirsizliği, kontrol altına alınamamasından kaynaklanır. Temelde 3 farklı problem yatar: Olasılık (**Likelihood)**, (Kod Çözme) **Decoding**, (Öğenme) **Learning.** Belirsizlikler altında yorum yapmada kullanılan Saklı Markov Modeli'nde durum dizisini bulabilmek için üç önemli algoritma kullanılır:

- Tanıma Problemi - İleri Yön Algoritması
- Durum Dizisinin Bulunması Problemi - Viterbi Algoritması
- Model Parametrelerinin Öğrenilmesi - Baum Welch Algoritması

Forward Algorithm: Used to find the order of the states in the model. The probabilities of all possible state orders are added.

Viterbi Algorithm: Instead of adding all probabilities as in the Forward Algorithm, the one that best matches the probability vectors from each state order is selected in the Viterbi algorithm. Thus, a more reliable result is obtained.

Baum-Welch Algorithm: Calculates the observation probabilities by going through the observation sequence from beginning to end and again from end to beginning. Thus, it finds more accurate results.

İleri Algoritması: Modeldeki durumların sırasını bulmada kullanılır. Ortaya çıkabilecek tüm durum sıralarının olasılıkları toplanır.

Viterbi Algoritması: İleri Algoritması'ndaki gibi tüm olasılıkları toplamak yerine, Viterbi algoritmasında her durum sıralarından olasılık vektörleriyle en iyi örtüşeni seçilir. Böylece daha sağlıklı bir sonuç elde edilir.

Baum-Welch Algoritması: Gözlem dizisini baştan sona ve tekrar sondan başa geçerek gözlem olasılıklarını hesaplar. Böylece daha kesin sonuçlar bulur.

Maximum likelihood assignment

For a given observed sequence of outputs $x \epsilon V\_T$, we intend to find the most likely series of states $z \epsilon S\_T$. We can understand this with an example found below.

Climate	S	S	S	S	R	R	R	S	S	S	S	R	R	S	S	S
Emotion	G	H	H	H	G	H	H	G	H	H	H	G	H	H	H	H

S - Sunny
R - Rainy
H - Happy
G - Grumpy

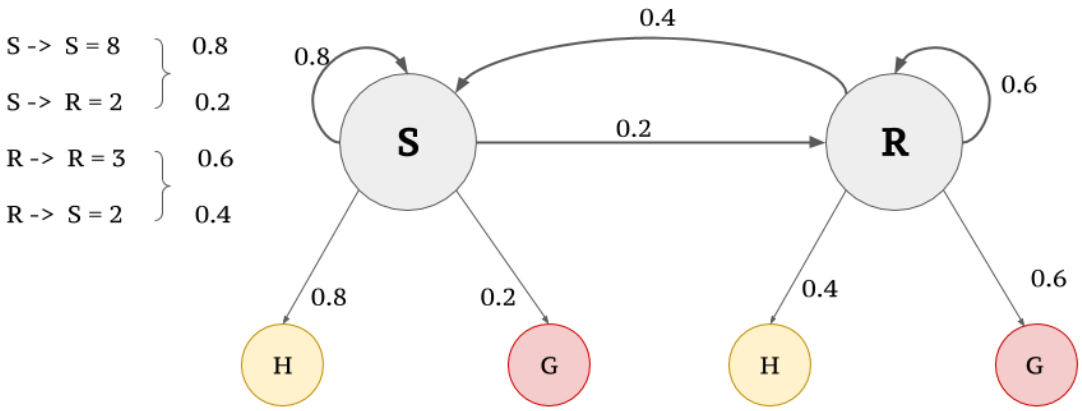


Fig. Data for example

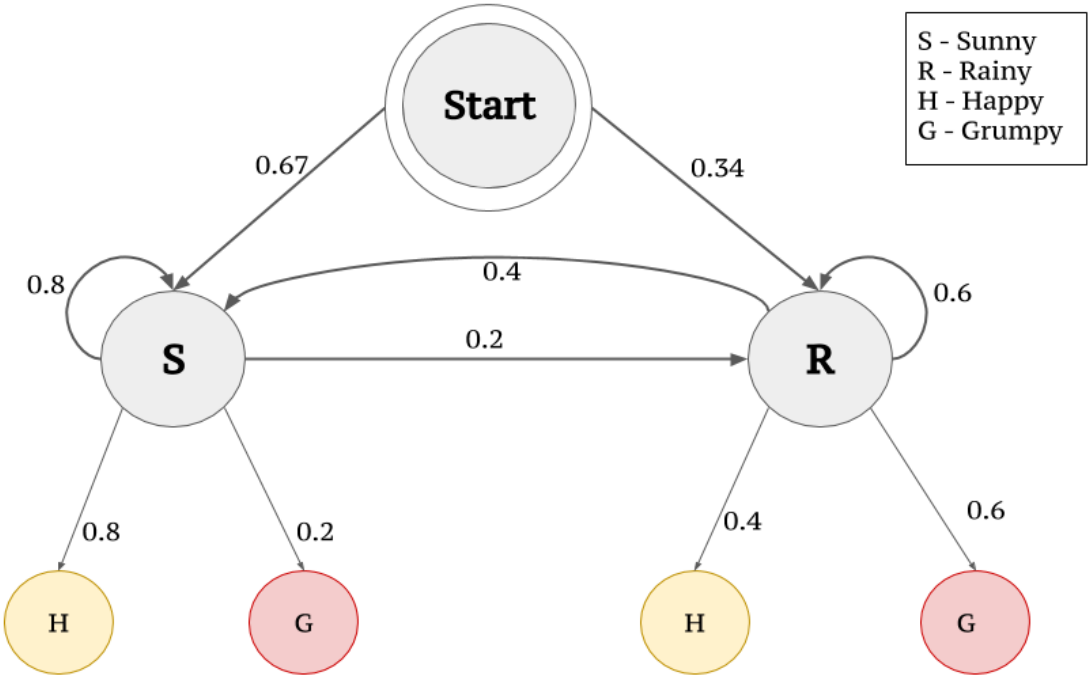


Fig. Markov Model as a Finite State Machine

The Viterbi algorithm is a dynamic programming algorithm similar to the forward procedure often used to find maximum likelihood. Instead of tracking the total probability of observations, it tracks the maximum probability and the corresponding sequence of states.
Viterbi algoritması, genellikle maksimum olasılığı bulmak için kullanılan ileri prosedüre benzer dinamik bir programlama algoritmasıdır. Gözlemlerin toplam olasılığını izlemek yerine, maksimum olasılığı ve karşılık gelen durum dizisini izler.

Consider the sequence of emotions: H,H,G,G,G,H for 6 consecutive days. Using the Viterbi algorithm we will find further probabilities of the series.
Duyguların sırasını düşünün: birbirini takip eden 6 gün boyunca H,H,G,G,G,H. Viterbi algoritmasını kullanarak serinin daha fazla olasılığını bulacağız.

Days	M	T	W	Th	F	S
Emotion	H	H	G	G	G	H

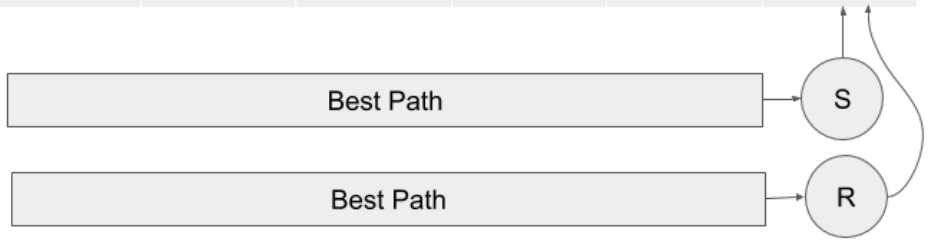


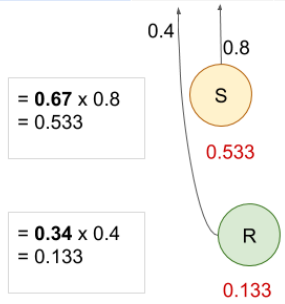
Fig. The Viterbi algorithm requires to choose the best path

There are conditions for Saturday that will be sunny or rainy. Here we aim to determine the best path until Sunny or Rainy Saturday and multiply it by the transition emission probability of Happy (H-Happy) (because Saturday makes one feel Happy).
Cumartesi için güneşli ya da yağmurlu olacak durumlar var. Burada Güneşli veya Yağmurlu Cumartesi'ye kadar en iyi yolu belirlemeyi ve Mutlu'nun (H-Happy) geçiş emisyon olasılığıyla çarpmayı amaçlıyoruz (çünkü Cumartesi kişiyi Mutlu hissettirir).

Let's consider a sunny Saturday. The previous day (Friday) could be sunny or rainy. Then we need to know the best path until Friday and then multiply it by the probability of emissions that lead to feeling Grumpy. Iteratively, we need to find the best path on each day such that the probability of the sequence of days is higher.
Güneşli bir Cumartesi düşünelim. Önceki gün (Cuma) güneşli veya yağmurlu olabilir. Ardından Cuma gününe kadar en iyi yolu bilmemiz ve ardından huysuz (Grumpy feeling) hissetmeye yol açan emisyon olasılıklarıyla çarpmamız gerekiyor. Yinelemeli olarak, her gündeki en iyi yolu, gün dizisi olasılığının daha yüksek olacağı şekilde bulmamız gerekir.

S - Sunny R - Rainy
H - Happy G - Grumpy

Days	Mon	Tues	Wed	Thurs	Fri	Satur
Emotion	H	H	G	G	G	H



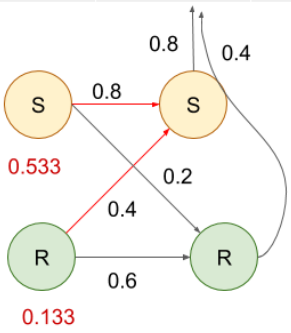
= **0.67** x 0.8
= 0.533

= **0.34** x 0.4
= 0.133

Fig. Step-1

S - Sunny R - Rainy
H - Happy G - Grumpy

Days	Mon	Tues	Wed	Thurs	Fri	Satur
Emotion	H	H	G	G	G	H



If Monday was Sunny:
= 0.533 x 0.8.x 0.8
= **0.341** => Choose this path (since greater prob)

If Monday was Rainy
= 0.133 x 0.4 x 0.8
= **0.00425**

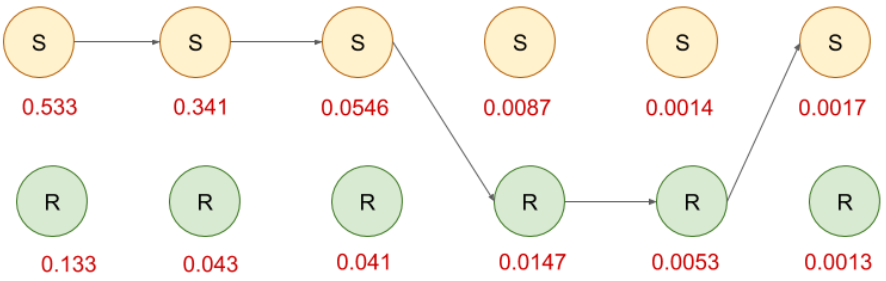
If Monday was Rainy:
= 0.533 x 0.2.x 0.4
= **0.043** => Chosen as more likely for Rainy

If Monday was Rainy
= 0.133 x 0.6 x 0.4
= **0.032**

Fig. Step-2

S - Sunny R - Rainy
H - Happy G - Grumpy

Days	Mon	Tues	Wed	Thurs	Fri	Satur
Emotion	H	H	G	G	G	H



More likelihood sequence = **S S S R R S**
For the given output sequence H H G G G H

Fig. The algorithm is iterated to choose the best path.

Şekil. En iyi yolu seçmek için algoritma yinelenir.

The algorithm leaves you with maximum likelihood values and we can now generate the sequence with maximum likelihood for a given output sequence.

Algoritma size maksimum olabilirlik değerleri bırakır ve artık belirli bir çıktı dizisi için maksimum olabilirliğe sahip diziyi üretebiliriz.

Learning values for HMM parameters A and B

Learning in HMMs involves estimating the state transition probabilities A and output emission probabilities B that make an observed sequence most likely. Expectation-maximization algorithms are used for this purpose. An algorithm known as the Baum-Welch algorithm, which falls into this category and uses the forward algorithm, is widely used.

HMM parametreleri A ve B için değerlerin öğrenilmesi

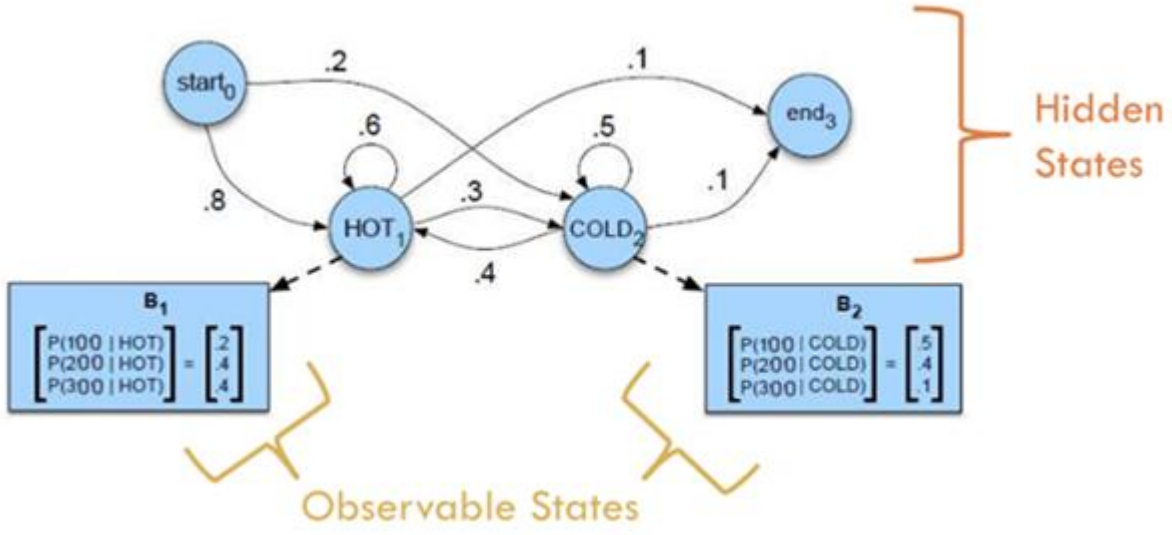
HMM'lerde öğrenme, gözlemlenen bir diziyi büyük olasılıkla yapan durum geçiş olasılıkları A ve çıktı emisyon olasılıkları B'nin tahmin edilmesini içerir. Bu amaçla beklenti-maksimizasyon algoritmaları kullanılmaktadır. Baum-Welch algoritması olarak bilinen ve bu kategoriye giren ve ileri algoritmayı kullanan bir algoritma yaygın olarak kullanılmaktadır.

Example:

Örnek:

A scientist researching the history of global warming in 2029 cannot find weather data for the Porga region of the village of Dedemin in 1956, but he finds a diary. This diary records the number of ice blocks brought from the mountains and sold in the city in hot and cold weather in 1956. The ice blocks sold are divided into 3 groups (300 pieces, 200 pieces, 100 pieces). While the probability of selling 300 ice blocks in cold weather is 10%, the probability of selling the same amount in hot weather is 40%. Can a conclusion be drawn regarding the weather based on this?

2029 yılında küresel ısınmanın geçmişini araştıran bir bilim insanı Dedemin köyü Porga yöresine ait 1956 yılının hava durumu verilerini bulamaz, fakat bir günlük bulur. Bu günlükte 1956 yılında sıcak ya da soğuk havalarda dağdan getirilip şehirde satılan buz blok sayıları yazmaktadır. Satılan buz bloklar 3 gruba (300 adet, 200 adet, 100 adet) ayrılmıştır. Soğuk havalarda 300 adet buz bloğun satılma olasılığı %10 iken sıcak havalarda aynı miktarın satılma olasılığı %40 dir. Bundan yola çıkarak hava durumu ile ilgili bir sonuç bulunabilir mi?



Can the state series be predicted given a series of observations? Each observation in the series represents, as a number, the probability of selling groups of ice blocks each day under hot or cold weather conditions.

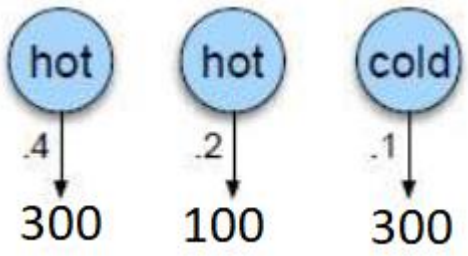
Bir gözlem serisi verildiğinde durum serisi tahmin edilebilir mi? Gözlem serisindeki her gözlem bir sayı olarak, sıcak ya da soğuk hava durumlarına göre her gün buz blok gruplarının satış olasılığını temsil eder.

Likelihood:

When the transition probabilities between the states and the probability of each observation occurring in each state are known, what is the probability of the ice block sales series "300pcs 100pcs 300pcs"? It is difficult to answer this question because it is not known in which state each observation in the observation series was made. However, if this observation series was obtained in the Hot, Hot, Cold states respectively, the probability can be easily calculated. For example,

Olasılık (Likelihood):

Durumlar arası geçiş olasılıklar ve her durumda her gözlemin gerçekleşme olasılığı bilinirken, "300adet 100adet 300adet" buz blok satış serisinin gelme olasılığı nedir? Bu soruyu cevaplamak zor, çünkü gözlem serisindeki her gözlemin hangi durumda yapıldığı bilinmiyor. Ama eğer sırasıyla **Hot, Hot, Cold** durumlarında iken bu gözlem serisini elde edildiyse, olasılık kolayca hesaplanabilir. Söz gelimi,



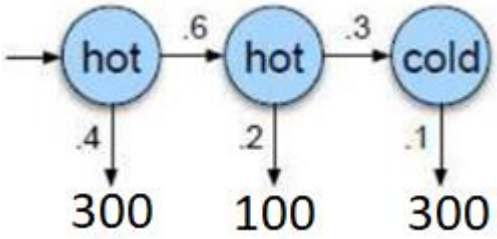
In the Hot state, we can find the probability of being 300 from the Observation State table, 0.4. Similarly, the probability of being 100 in the Hot state is 0.2 and the probability of being 300 in the Cold state is 0.1.

Hot durumunda iken, 300 olma olasılığını **Gözlem Durum** tablosundan bulabiliriz, 0.4. Aynı şekilde Hot iken 100, 0.2 ve Cold iken 300 olma olasılıklarını, 0.1 olarak bulunur.

$$P(O|Q) = \prod_{i=1}^{T} P(o_i \mid q_i)$$

We find the probability of the observation series by multiplying these probabilities with each other. Here, T: is the number of observations. Here, the transition probabilities between states are ignored. If the transition probabilities between states are also taken into account,

Bu olasılıkları birbiri ile çarparak gözlem serisinin olasılığını buluruz. Burada, T: Gözlem sayısıdır. Burada durumlar arası geçiş olasılıkları göz ardı edilmiştir. Durumlar arası geçiş olasılıkları da göz önüne alınırsa,

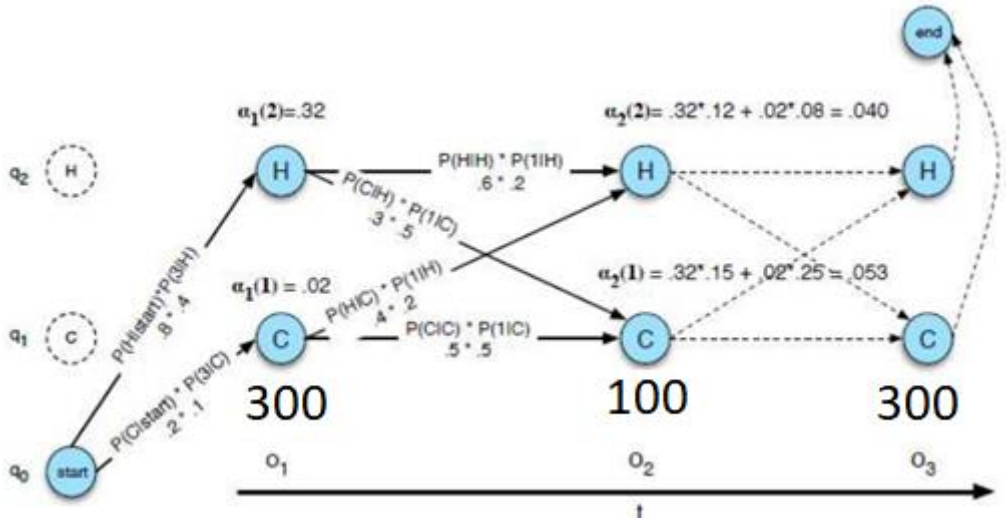


$$P(O,Q) = P(O|Q)P(Q) = \prod_{i=1}^{T} P(o_i \mid q_i) \prod_{i=1}^{T} P(q_i \mid q_{i-1})$$

Start → Hot probability 0.8
Hot → Hot probability 0.6
Hot → Cold probability 0.3
Start → Hot olasılığı 0.8
Hot → Hot olasılığı 0.6
Hot → Cold olasılığı 0.3

When the transition probabilities between states are included in the multiplication process, the real probability is calculated. However, it is not known whether they are in the Hot, Hot, Cold states. In that case, there can be 2 different states for each observation and since there are 3 observations in total, there are 2^3 = 8 different state series. If the above calculation is made for each state series and added together, the probability of the 300 100 300 series is calculated.
Durumlar arası geçiş olasılıkları da çarpma işlemine dahil edildiği zaman gerçek olasılık hesaplanmış olur. Yine de Hot, Hot, Cold durumlarında olup olmadıkları bilinmiyor. O halde her gözlem için 2 farkı durum olabilir ve toplamda 3 gözlem olduğu için 2^3 = 8 farklı durum serisi vardır. Her bir durum serisi için yukarıdaki hesabı yapılıp toplanırsa, 300 100 300 serisinin olasılığı hesaplanmış olur.

To put it more generally, there are N^T possible series of states for N states and T observations. If N and T are large numbers, this approach will take too much time. A more efficient algorithm is the Forward Algorithm. The Forward Algorithm is actually a Dynamic Programming algorithm. It is used to store previously calculated results and solve a larger problem.
Biraz daha genele dökersek, N adet durum ve T adet gözlem için N^T adet olası durum serisi vardır. Eğer N ve T büyük sayılar ise, bu yaklaşım çok fazla süre alacaktır. Daha verimli bir algoritma olan **Forward Algoritması** kullanılır. Forward Algoritması aslında Dinamik Programlama algoritmasıdır. Daha önce hesaplanan sonuçları saklayıp daha büyük bir problemi çözerken kullanılır.



First, we start from the Start state. We go to the Hot or Cold states:

- Since the probability of Start → Hot is 0.8 and the probability of selling 300 ice blocks while Hot is 0.4, there is a total probability of going to Hot and selling 300 ice blocks with probability 0.32.
  P(H | Start)* P(300|H)=0.8*0.4=0.32
- Since the probability of Start → Cold is 0.2 and the probability of selling 300 ice blocks while Cold is 0.1, there is a total probability of going to Cold and selling 300 ice blocks with probability 0.02.

P(C | Start)* P(300|C)=0.2*0.1=0.02

İlk olarak Start durumundan başlanılır. Hot veya Cold durumlarına gidilir:

- Start → Hot olasılığı 0.8 ve Hot iken 300 buz blok satma olasığı 0.4 olduğu için toplamda 0.32 olasılıkla Hot'a gidip 300 buz blok satılma olasılığı bulunur.
  P(H | Start)* P(300|H)=0.8*0.4=0.32
- Start → Cold olasılığı 0.2 ve Cold iken 300 buz blok satma olasılığı 0.1 olduğu için toplamda 0.02 olasılık ile Cold'a gidip 300 buz blok satılma olasılığı bulunur.
  P(C | Start)* P(300|C)=0.2*0.1=0.02

Let's assume that after starting from the Start state, we come to either Hot or Cold state. It doesn't matter which. Now, from Hot state, we go to Hot or Cold state. Similarly, from Cold state, we go to Hot or Cold state.

- If we come to Hot, since Hot → Hot probability is 0.6 and the probability of selling 100 ice blocks while Hot is 0.2, there is a total probability of 0.12 to go to Hot and sell 100 ice blocks.
- If we come to Cold, since Cold → Hot probability is 0.4 and the probability of selling 100 ice blocks while Hot is 0.2, there is a total probability of 0.8 to go to Hot and sell 1000 ice blocks.

Thus, we come to Start → Hot or Cold → Hot states. We calculate the probability of being in Hot right now as follows: (Start → Hot → Hot)+ (Start → Cold → Hot).

Start durumundan yola çıktıktan sonra ya Hot ya da Cold durumlarından birine gelindiğini farz edelim. Hangisi olduğu önemli değil. Şimdi Hot durumundan Hot ya da Cold durumuna gidilir. Aynı şekilde Cold durumundan da Hot ya da Cold durumuna gidilir.

- Eğer Hot'a geldiysek, Hot → Hot olasılığı 0.6 ve Hot iken 100 buz blok satma olasılığı 0.2 olduğu için toplamda 0.12 olasılık ile Hot'a gidip 100 buz blok satılma olma olasılığı bulunur.
- Eğer Cold'a geldiysek, Cold → Hot olasılığı 0.4 ve Hot iken 100 buz blok satma olasılığı 0.2 olduğu için toplamda 0.8 olasılık ile Hot'a gidip 1000 buz blok satılma olasılığı bulunur.

Böylece Start → Hot veya Cold → Hot durumlarına geldik. Şu an Hot'ta olma olasılığını da şu şekilde hesaplarız: (Start → Hot → Hot)+ (Start → Cold → Hot).

We calculate the left part of the sum using the probabilities Start → Hot and Hot → Hot that we calculated earlier: 0.32 * 0.12. We calculate the right part of the sum using the probabilities Start → Cold and Cold → Hot that we calculated earlier: 0.02 * 0.08. Adding these two products gives us the probability of being in the current state.
Toplamın sol kısmını daha önce hesapladığımız Start → Hot ve Hot → Hot olasılıklarını kullanarak hesaplarız: 0.32 * 0.12. Toplamın sağ kısmını daha önce hesapladığımız Start → Cold ve Cold → Hot olasılıklarını kullanarak hesaplarız: 0.02 * 0.08. Bu iki çarpımı toplayınca da mevcut state'de olma ihtimalini buluruz.

If we need to formulate it:
Formüle etmek gerekir ise:

$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij} b_j(o_t)$$

- at-1(i) → Durum i'ye kadar olan toplam olasılık.
- aij → Durum i'den Durum j'ye geçme olasılığı.
- bj(ot) → Mevcut Durum'da o gözlemin olma olasılığı.

- at-1(i) → Total probability up to State i.
- aij → Probability of going from State i to State j.
- bj(ot) → Probability of that observation occurring in the Current State

Decoding:

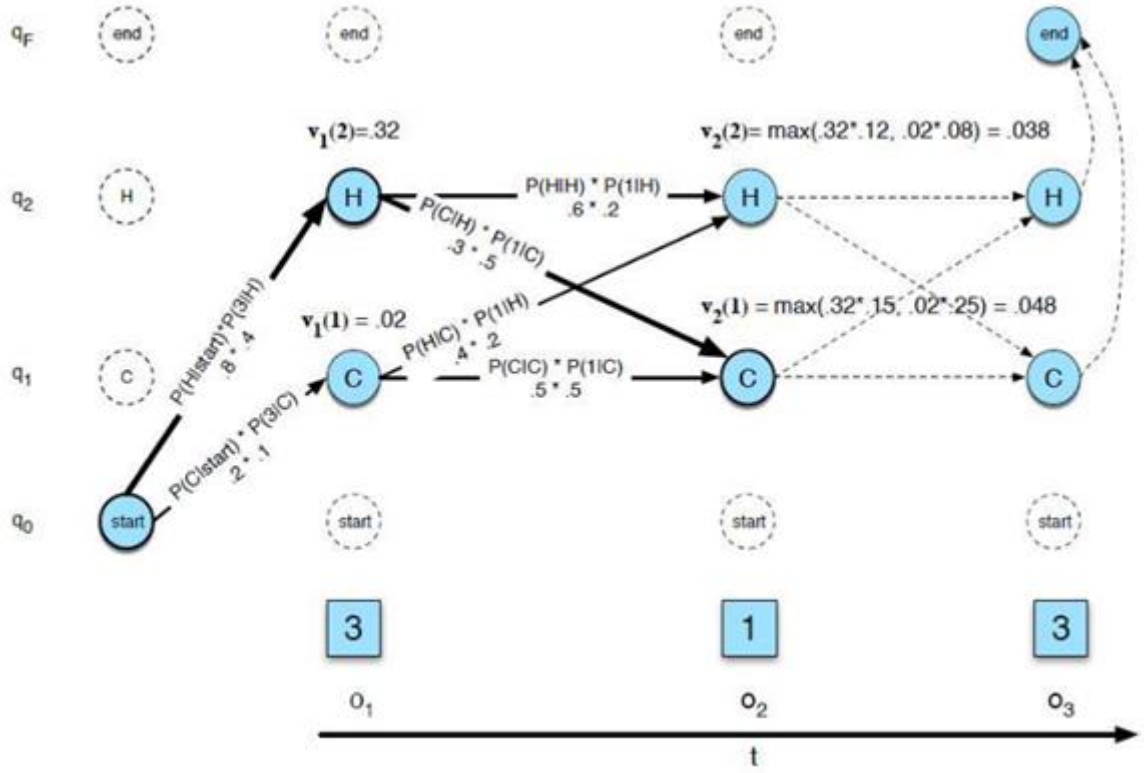
Kod çözme:

300 – 100 – 300 gözlem serisi, en fazla ihtimalle hangi durum serisi tarafından geliştirildi?

En basit çözüm olarak, her durum serisi için İleri Algoritmasını kullanıp, en yüksek olasılığa sahip olan durum serisi bulunabilir. Ama bu pek verimli değil. Viterbi Algoritması ile daha verimli bir şekilde ulaşılabiliriz.

300 – 100 - 300 gözlem serisi, en fazla ihtimalle hangi durum serisi tarafından üretilmiştir?

En basit çözüm olarak, her durum serisi için Forward Algoritmasını kullanıp, en yüksek olasılığa sahip olan durum serisi bulunabilir. Ama bu pek de verimli değildir. **Viterbi Algoritması** ile daha verimli bir şekilde sonuca ulaşabiliriz.



The Viterbi algorithm is similar to the Forward Algorithm, but there is a fundamental difference. When calculating the probability of the Start → Hot or Cold → Hot series, we added the Start → Hot → Hot and Start → Cold → Hot series probabilities. In the Viterbi Algorithm, the probabilities of the Start → Hot → Hot and Start → Hot → Cold series are compared and the highest is taken.

Viterbi algoritması, Forward Algoritmasına benzer, fakat arada temel bir fark vardır. Start → Hot veya Cold → Hot serisinin olasılığını hesaplarken Start → Hot → Hot ve Start → Cold → Hot seri olasılıklarını toplamıştık. Viterbi Algoritmasında ise Start → Hot → Hot ile Start → Hot → Cold serilerinin olasılıklarını karşılaştırıp en fazla olan alınır.

Formulated as:
Formüle edilirse:

$$v_t(j) = max_{i=1}^{N} v_{t-1}(i) a_{ij} b_j(o_t)$$

- at-1(i) → Viterbi probability up to State i.
- aij → Probability of transition from State i to State j.
- bj(ot) → Probability of that observation occurring in the Current State.
- at-1(i) → Durum i'ye kadar olan Viterbi olasılığı.
- aij → Durum i'den Durum j'ye geçme olasılığı.
- bj(ot) → Mevcut Durum'da o gözlemin olma olasılığı.

Learning:
Öğrenme:

In the Learning Problem, the state series and the observation series are not hidden. The aim is to calculate the transition probabilities between states and the probability of which observation will occur in which state.
Öğrenme Probleminde ise durum serisi ve gözlem serisi gizli değildir. Amaç durumlar arası geçiş olasılıklarını ve hangi durumda hangi gözlemin gerçekleşeceği olasılığını hesaplamaktır.

$$a_{ij} = \frac{Count(i \to j)}{\sum_{q\epsilon} Count(i \to q)}$$

To calculate the transition probability from State i to State j, divide the number of transitions i → j by the total number of transitions from State i to any State.
Durum i'den Durum j'ye geçiş olasılığı hesaplamak için i → j geçiş sayısını, Durum i'den herhangi bir Duruma'a yapılan toplam geçiş sayısına bölünür.

Expectation Maximization:

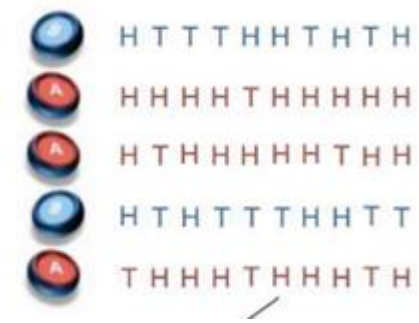
Beklenti maksimizasyon:

Let's say we have two coins: A and B. Both coins are actually fair coins. When we flip a coin with A, it comes up heads with probability Theta(A), while when we flip a coin with B, it comes up heads with probability Theta(B). Our goal is to find the values of Theta(A) and Theta(B). Heads: Heads, Tails: Tails

Elimizde iki adet bozuk para olsun: A ve B. Her iki para da aslında hileli para. A parası ile yazı tura yaptığımız zaman Theta(A) olasılık ile yazı gelirken, B parası ile yazı tura yaptığımız zaman Theta(B) olasılık ile tura gelir. Bizim amacımız ise Theta(A) ile Theta(B) değerlerini bulabilmek. Yazı: Heads, Tura: Tails

To do this, we choose one of these two coins, toss a coin 10 times, and then put the coin back in its place. We repeat this process 5 times. While doing these operations, we note how many times each coin came up heads or tails:

Bunun için bu iki paradan biri seçeriz, 10 kere yazı tura atarız, ardından parayı yerine koyarız. Bu işlemi 5 kere tekrar ederiz. Bu işlemleri yaparken de hangi paradan kaç kere yazı veya tura geldiğimi not ederiz:



To find the value of Theta(A), we divide the number of heads when coin A is tossed by the total number of heads with coin A. That is, we tossed coin A 30 times and it came up heads 24 times, so Theta(A) = 0.8. Similarly, Theta(B) = 0.45 …

Theta(A) değerini bulmak için, A parası ile yazı tura atıldığı zaman gelen yazı sayısını, A parası ile toplam atılan yazı tura miktarına böleriz. Yani, A parası ile 30 kere yazı tura attık ve 24 kere yazı geldiği için Theta(A) = 0.8 olur. Benzer şekilde, Theta(B) = 0.45 …

9. Bilgi Kuramı

Information is used to create order in disorder. Time continued to develop and strengthen the existence of information. The first clue to the true nature of information was noticed thanks to a strange problem. One of the brightest minds of the nineteenth century, a Scottish physicist, had a dream while thinking about something completely different. When the universe tends towards disorder, everything is doomed to fall apart. Despite this, can everything be collected and made orderly? In those years, this idea started a debate. Can order be created without spending any energy?

Düzensizlikte düzen yaratmak için bilgi kullanılır. Zaman bilginin varlığını geliştirmeye ve güçlendirmeye devam ediyordu. Bilginin gerçek doğasına dair ilk ipucu tuhaf bir problem sayesinde fark edilmiştir. On dokuzuncu yüzyılın en parlak zekâlarından İskoç fizikçisi tamamıyla farklı bir şey düşünürken bir hayal kurdu. Evren düzensizlik haline yöneldiğinde her şey parçalanmaya mahkûmdur. Buna rağmen her şey toparlanıp düzgün hale getirilebilir mi? O yıllarda bu fikir bir tartışma başlattı. ***Hiç enerji harcamadan bir düzen yaratılabilir mi?***

The Second Law of Thermodynamics: “Not all energy can be converted into useful work, some of it is used to maintain the internal integrity of the system.” According to the second law, the entropy change in a system and its surroundings in any process is either “zero” or “positive.” The thermodynamic term representing thermal energy that cannot be converted into mechanical work is called entropy. It is the expression of disorder in a system. If there is a thermal process, entropy is either equal to zero or positive. Entropy is used not only in thermodynamics in daily life, but also in many areas from statistics to theology.

Termodinamiğin İkinci Yasası: ***“Enerjinin tamamı faydalı işe çevrilemez, bir kısmı sistemin içsel bütünlüğünü korumak için kullanılır.”*** İkinci yasaya göre, herhangi bir süreçte bir sistem ve çevresindeki entropi değişimi ya “sıfır” ya da “pozitiftir”. Mekanik işe çevrilemeyecek termal enerjiyi temsil eden **termodinamik** terimine **entropi** denilir. Bir sistemdeki düzensizliğin ifadesidir. Termal bir işlem varsa **entropi** ya sıfıra eşit olur ya da pozitif değer alır. Entropi, gündelik hayatta sadece **termodinamikte** değil, istatistikten teolojiye birçok alanda kullanılır.

Scottish physicist James Clark Maxwell (1831 - 1879) was greatly influenced by the science of thermodynamics, apart from many other areas of interest. The study of heat and motion led to the birth of steam engines. Maxwell was one of the first to understand that heat consists of the movement of molecules. While hot molecules gather on one side, cold molecules gather on the other. He explained this with the experiment of two boxes managed by a genie to change the disorder created by molecules moving fast and slow in two separate environments, which constitutes the basis of thermodynamics, to order. Maxwell assumed that he knew what was happening in a box half filled with cold and half

hot air. A genie sits on the box and can easily see the molecules inside the box. A door between the two boxes, which passes the fast ones to one box and the slow ones to the other box, allows the molecules to pass. The genie who manages this follows all the molecules in the boxes with their movements. When the fast molecules and the slow molecules move towards the door, the door opens immediately. Thus, the slow molecules can gather in one box and the fast ones in the other box. This experiment gave rise to a new idea. In the beginning, information was used to provide order in a completely disordered environment. Moreover, this was done without any effort. Here, information should not violate the laws of physics. Maxwell's demon uses nothing but information to create useful energy. This does not mean that something is created from nothing. According to him, you could create an orderly situation from disorder. The demon has to keep the speeds, positions and directions of all molecules, fast and slow. The demon has to keep track of which molecule moves from where to where, and if there is a limit to keeping records, the demon has to delete the records at some point. Deleting information is an irreversible process that increases entropy. Thus, disorder increases. What is discovered here is that the least amount of energy (Dernau limit) must be spent to delete one bit of information. This is a very small value. Thus, the relationship between information and energy is discovered with incredible accuracy. This experiment that Maxwell dreamed of in the steam age is one of the scientific researches today. Maxwell's demon connects the two most important contents of science, energy and information. The science of thermodynamics has clearly shown that the universe tends towards disorder over time, in other words, entropy constantly increases. Everything is doomed to fall apart. The genie claims that you can put everything together without wasting energy.

İskoç fizikçi James Clark Maxwell (1831 – 1879) birçok ilgi alanın dışında termodinamik biliminden çok etkilenmişti. Isı ve hareket çalışmaları buhar makinelerinin doğumuna sebep olmuştu. Maxwell ısının moleküllerin hareketinden ibaret olduğunu anlayan ilk kişilerdendi. Sıcak moleküller bir tarafta toplanırken soğuk moleküller diğer tarafta toplanmaktadır. Bunu termodinamiğin esasını oluşturan iki ayrı ortamda hızlı ve yavaş hareket eden moleküllerin oluşturduğu düzensizlikten düzene geçmek için bir cinin yönettiği iki kutu deneyi ile anlattı. Maxwell teoride yarısı soğuk yarısı sıcak hava ile dolu bir kutuda ne olup bittiğini bildiğini varsaydı. Bir cin kutunun üzerinde oturuyor ve kutunun içerisindeki molekülleri kolayca görebiliyor. İki kutunun arasında hızlı olanları bir kutuya, yavaş olanları diğer kutuya geçiren bir kapı moleküllerin geçişine izin veriyor. Bunu yöneten cin kutulardaki tüm molekülleri hareketleri ile birlikte takip ediyordu. Hızlı olan moleküller ve yavaş olan moleküller kapıya doğru yöneldiğinde kapı hemen açılıyor. Böylece yavaş moleküller bir kutuda, hızlı olanlar ise diğer kutuda toplanabiliyordu. Bu deney yeni bir fikrin doğmasına sebep vermiştir. Başlangıçta tamamen düzensiz ortamda düzeni sağlamak için bilgi kullanılıyordu. Üstelik bu hiçbir çaba harcamadan yapılıyordu. Burada bilgi fizik kanunlarını çiğnememesi gerekir. Maxwell'in cini faydalı enerji yaratmak için bilgiden başka bir şey kullanmıyor. Bu bir şeyin yoktan var edildiği anlamına gelmiyor. Ona göre düzensizlikten düzenli bir durum yaratabilirdiniz. Cin, hızlı ve yavaş tüm molekülerin hızlarını konumlarını, yönlerini

hafızasında tutmak zorunda. Cin hangi molekülün nereden nereye hareket ettiğinin kaydını tutmak zorunda, kayıt tutmanın bir sınırı var ise bir noktada cinin kayıtları silmesi gerekiyor. Bilginin silinmesi entropiyi artıran ve geri dönüşü olmayan bir işlem oluyor. Böylece düzensizlik artar. Burada keşfedilen şey bir bitlik bilgiyi silmek için en az sevideki enerjinin (Dernau sınırı) harcanması gerekir. Bu çok küçük bir değer. Böylece bilgi ile enerji arasındaki ilişki inanılmaz doğrulukta keşfedilmiş oluyor. Maxwell'in buhar çağında hayalinde kurduğu bu deney günümüzde bilimsel araştırmalardan birisidir. Maxwell'in cini bilimin en önemli iki içeriği olan enerji ve bilgiyi birbirine bağlıyor. Termodinamik bilimi açıkça göstermiştir ki, zaman içerisinde evren düzensizlik haline yönelir, yani entropi sürekli artar. Her şey parçalanmaya mahkûmdur. Cin enerji harcamadan her şeyi derleyip toparlayabileceğinizi iddia etmektedir.

Information follows the laws of physics. Information cannot be separated from the physical world. What makes information so powerful is that we can store it in any physical system we want. Information has been carried in rocks and clays for centuries. Information has been transferred rapidly using electricity and light. The devices that carry information have given it extraordinary properties.
Bilgi fizik kanunlarına göre davranır. Bilgi fiziksel dünyadan ayrılamaz. Bilgiyi çok güçlü kılan şey onu istediğimiz herhangi bir fiziksel sistemde saklayabilecek olmamızdır. Taşlarda ve killerde bilgi çağlar boyu taşındı. Elektrik ve ışık kullanarak bilgiyi hızla transfer edildi. Bilgiyi taşıyan aygıtlar ona sıra dışı özellikler kazandırdı.

At the beginning of the twentieth century, a portable device that could easily handle complex operations while processing information began to be considered. This device would be known as a computer. Alan Turing (1912 - 1954) was the first person to create a mathematical model of a computer. A machine that processes and changes information! Turing was actually thinking about solving a mathematical problem. What would happen if problems in mathematics were solved by following a simple set of rules? This made him think about computers. Something unexpected happened and the computer emerged. This machine changed the lives of almost everyone. Turing was interested in solving certain operations in mathematics by following a simple set of rules in the binary number system. Turing's magnificent idea was first published in a 36-page book called "Application of Decision Problems in Computable Numbers", which he wrote in 1936 when he was 24 years old. The idea of a modern computer was put forward thanks to Turing's magnificent logic. In those years, the word computer meant the profession of a person who did arithmetic calculations. In those years, banking and trade were developing rapidly. A large number of people were hired to calculate interest rates and to do navigation calculations. Turing asked a question: What was going on in the mind of a person who was doing calculations and thinking? What was the vital thing for the person doing the calculations? What was the key function of the human brain in the calculation process? He noticed that certain rules were repeated in the calculation process. Turing saw that all calculations were in binary. Turing

had to find a method that would allow machines to understand the instructions that performed arithmetic operations in the same way that humans do. The instructions had to be translated into a language that machines could understand. He concentrated on the instructions that told the machine what to do with the data. Turing wanted to translate arithmetic operations into a language that machines could understand. Turing succeeded in doing this; he showed that when instructions consisting of 1s and 0s were given to the computer as commands on a tape, the machine would perform functions like the human brain. The tapes became the medium in which information and instructions were stored and processed. A computer with a sufficiently large memory could do an almost unlimited number of tasks.

Yirmi yüzyılın başlarında bilgi işlenirken karmaşık işlemleri basitçe halledebilen taşınabilir bir cihaz düşünülmeye başlandı. Bu cihaz bilgisayar olarak bilinecekti. Alan Turing (1912 – 1954) bilgisayarın matematiksel modelini yaratan ilk insandı. ***Bilgiyi işleyen ve değiştiren bir makine!*** Turing aslında matematiksel bir problemin çözümünü düşünüyordu. Matematikteki problemler basit kurallar dizisi takip edilerek çözülürse ne olur? Bu da bilgisayaralar hakkında düşünmesini sağladı. Beklenmedik bir şey oldu ve bilgisayar ortaya çıktı. Bu makine neredeyse tüm insanların hayatını değiştirdi. Turing matematikteki belirli işlemlerin ikili say sisteminde basit kurallar dizisi takip edilerek çözülmesiyle ilgileniyordu. Turing'in muhteşem fikri ilk kez 24 yaşındayken 1936 yılında yazdığı güzümüzde efsane olan "Hesaplanabilir sayılarda karar veren problemlerin uygulanması" isimli 36 sayfalık kitapta yayınlandı. Modern bilgisayar fikri Turing'in muhteşem mantığı sayesinde ortaya atılmıştır. O yıllarda bilgisayarın kelime anlamı aritmetik hesap yapan kişinin mesleğinin adıydı. O yıllarda bankacılık ve ticaret hızlıca gelişiyordu. Çok sayıda insan faiz hesaplamaları, seyrüsefer hesaplamaları için işe alınıyordu. Turing bir soru sordu: ***Hesaplama yapan, düşünen bir insanın zihninde neler oluyor? Hesaplama yapan kişi için hayati öneme sahip ola şey neydi?*** Hesaplama işleminde insan beyninde anahtar işlev neydi? Hesaplama işleminde belirli kuralların tekrar edildiğini fark etti. Turing tüm hesaplamaların ikili boyutta olduğunu gördü. Turing, makinelerin aritmetik işlemleri yapan talimatları insanların ki gibi anlamalarını sağlayan bir metot bulmalıydı. Talimatları makinelerin anlayabileceği bir dile çevrilmesi gerekiyordu. Veri ile ne yapacağını söyleyen talimatlara yoğunlaştı. Turing aritmetik işlemleri makinelerin anlayabileceği bir dile çevirmek istiyordu. Turing bunu başardı; bir şeritte 1 ve 0'lardan oluşan talimatlar bilgisayara komut olarak verildiğinde makinenin insan beyni gibi işlevleri yerine getireceğini gösterdi. Şeritler bilginin ve komutların saklandığı ve işlendiği ortamlara dönüşmüştü. Yeterince büyük hafızası olan bir bilgisayar nerdeyse sınırsız sayıda iş yapabiliyordu.

Today, pictures, music, writings, sounds, images can all be processed by a single machine. Programs, software or applications are nothing more than very long strips of data consisting of 1 and 0 that tell the computer what to do. The machine that converts commands into symbols can create not just a simple picture or sound but even a changing system. Turing,

who thought about how the human brain works and found the methodology to apply it to the machine with commands and instructions, produced one of the most important ideas of the twentieth century. The computer showed that knowledge is power.
Günümüzde resim, müzik, yazılar, ses, görüntü hepsi tek bir makine tarafından işlenebiliyor. Programlar, yazılım ya da uygulamalar dediğimiz bilgisayara ne yapacağını söyleyen 1 ve 0'dan oluşan çok uzun şeritlerdeki verilerden başka bir şey değildir. Komutları sembollere dönüştüren makine sadece basit bir resmi ya da sesi değil değişen bir sistemi bile yaratabiliyor. İnsan beyninin nasıl işlediğini düşünerek onu komut ve talimatlar ile makineye uygulama metodolojisini bulan Turing yirminci yüzyılın en önemli fikirlerinden birini üretti. Bilgisayar bilginin güç olduğunu gösteriyordu.

The modern information age needed another idea. Claude Shannon (1916 – 2001) passion for solving an extraordinary problem led to the emergence of a new power of information. His booklet, "The Mathematical Theory of Communication", written in 1948, is one of the most important scientific booklets of the twentieth century. Shannon found a way to measure and evaluate the amount of information in a message. He realized that the content of the information in a message was not related to its meaning. He needed to give a unit of measurement to information. He showed that a message to be transmitted could be measured when converted to binary number system. The message was a long series of ones and zeros. He realized that converting information to binary number system was a very powerful movement. Bit: 0/1 was defined. Bit is the smallest amount of information in the digital world. All requests have two sides and carry one bit of information: On / Off, Known / Unknown, Heads / Tails, Light / Dark, Stop / Go. Thanks to Shannon, bit became the common language of information. Thus, information became tangible. Information has been transformed into a measurable force, into reality.
Modern bilgi çağının başka bir fikre daha ihtiyacı vardı. Claude Shannon (1916 – 2001) sıra dışı bir problemi çözme tutkusu bilginin yeni bir gücünün ortaya çıkmasına neden oldu. 1948 yılında yazdığı, "İletişimin Matematiksel Teorisi" isimli kitapçığı yirminci yüzyılın en önemli bilimsel kitapçıklarından biridir. Shannon, ***bir mesaj içerisindeki bilgi miktarını ölçmenin ve değerlendirme***nin bir yolunu buldu. Bir mesajdaki bilginin içeriğinin anlamı ile ilgisinin olmadığını fark etti. Bilgiye bir ölçü birimi vermesi gerekiyordu. İletilecek bir mesajın ikili sayı sistemine dönüştürüldüğünde ölçülebileceğini gösterdi. Mesaj bir ve sıfırlardan oluşan uzun bir dizi idi. Bilgi ikili sayı sistemine dönüştürmenin oldukça güçlü bir hareket olduğunu fark etti. Bit: 0/1 tanımlandı. Bit, bilginin sayısal dünyadaki en küçük miktarıdır. Tüm istemlerin iki yüzü vardır ve bir bitlik bilgi taşır: Açık / Kapalı, Bilenen / Bilinmeyen, Yazı /Tura, Aydınlık / Karanlık, Dur / Geç. Shannon sayesinde bit bilginin ortak dili oldu. Böylece bilgi elle tutulabilir hale geldi. Bilgi ölçülebilen bir güce, gerçeğe dönüştürüldü.

Information is not something that only humans create. Any event in the universe can contain incredible information and messages. All physical and chemical events that we

cannot normally see can be shown to us like a normal movie. Information is an inseparable part of the universe. Information is everywhere.
Bilgi sadece insanların var ettiği bir şey değil. Kâinattaki herhangi bir olay inanılmaz bilgiler ve mesajlar içerebilir. Normalde göremeyeceğimiz tüm fiziksel, kimyasal olaylar normal bir film gibi bizlere izletilebiliyor. Bilgi kâinatın ayrılmaz bir parçasıdır. Bilgi her yerdedir.

What is the relationship between information and energy? Information is not abstract. Information is not a formula that you can write on paper. Information is carried and given meaning; information is written on a stone, a book. It is written in a memory or in the brain. After all, information is carried and there is something that carries it. This shows that information behaves according to the laws of physics. Humanity has to learn that information is integrated with the physical world. What makes information powerful is that we can store it in any system. Information has been stored for centuries on a clay tablet and stopped time. We sent information rapidly as electricity and light. The devices that carry information provide it with extraordinary properties.
Bilgi ile enerji arasındaki ilişki nedir? Bilgi soyut değil. Kâğıda yazabileceğiniz bir formül bilgi değildir. Bilgi taşınır ve anlamlandırılır; bilgi bir taşa, bir kitaba yazılır. Bir belleğe ya da beyine yazılır. Sonuçta bilgi taşınır ve onu taşıyan bir şey var. Bu da bilginin fizik kanunlarına göre davrandığını gösterir. İnsanlık bilginin fiziksel dünya ile bütünleşik olduğunu öğrenmek zorunda. ***Bilgiyi güçlü kılan şey onu herhangi bir sistemde saklayabilecek olmamızdır.*** Kil tablette bilgi çağlar boyu saklandı ve zamanı durdurdu. Elektrik ve ışık olarak bilgiyi hızla gönderdik. Bilgiyi taşıyan aygıtlar ona sıra dışı özellikler sağlamaktadır.

Information theory is theoretical part of communication developed by American mathematician Shannon. It deals with mathematical modelling and analysis of a communication system rather than dealing with physical sources (camera, keyboard, microphone etc.) and physical channels (wires, cables, fibre, satellite, radio etc).
Bilgi teorisi, Amerikalı matematikçi Shannon tarafından geliştirilen iletişimin teorik bir parçasıdır. Fiziksel kaynaklar (kamera, klavye, mikrofon vb.) ve fiziksel kanallarla (teller, kablolar, fiber, uydu, radyo vb.) ilgilenmekten ziyade bir iletişim sisteminin (Haberleşme kanalının) matematiksel modellemesi ve analiziyle ilgilenir.

Note: Lot of mathematics is involved in manufacturing of things we use in daily-life… like, manufacturing of pen, paper, flash memory, mobiles, antenna design, garment manufacturing ete etc etc Information theory addresses and answers below fundamental questions of communication theory:

Not: Günlük hayatımızda kullandığımız şeylerin üretiminde çok fazla matematik yer alır. Bilgi teorisi, iletişim teorisinin temel sorularını ele alır ve aşağıdaki cevapları verir:

- Data compression
- Data transmission
- Data storage
- Error detection and correction codes
- Ultimate Data rate over a noisy channel (for reliable communication)

• Veri sıkıştırma
• Veri iletimi
• Veri depolama
• Hata tespiti ve düzeltme kodları
• Gürültülü bir kanal üzerinden nihai veri hızı (güvenilir iletişim için)

Every morning one reads newspaper to receive information. A message is said to convey information, if two key elements are present in it.

- Change in knowledge or meaning
- Uncertainty (unpredictability)

Her sabah bilgi almak için gazete okunur. Bir mesajın, içinde iki temel unsur varsa, bilgi ilettiği söylenir.
• Bilgi veya anlamdaki değişim
• Belirsizlik (öngörülemezlik)

Amount of information contained in a message is inter-related to its probability of occurrence. Information is always about something (occurrence of an event etc.). It may be a true or lie.
Bir mesajda bulunan bilgi miktarı, onun gerçekleşme olasılığıyla ilişkilidir. Bilgi her zaman bir şeyle ilgilidir (bir olayın gerçekleşmesi vb.). Doğru veya yalan olabilir.

Information theory talks about:

- How to measure the amount of information?
- How to measure the correctness of information?
- What to do if information gets corrupted by errors?
- How much memory does it require to store information?

Bilgi teorisi şunlardan bahseder:
• Bilgi miktarı nasıl ölçülür?
• Bilginin doğruluğu nasıl ölçülür?
• Bilgi hatalar nedeniyle bozulursa ne yapılmalı?
• Bilgiyi depolamak için ne kadar bellek gerekir?

The movements and transformations of information, just like those of a fluid, are constrained by mathematical and physical laws. These laws have deep connections with:

- probability theory, statistics, and combinatorics
- thermodynamics (statistical physics)
- spectral analysis, Fourier (and other) transforms
- sampling theory, prediction, estimation theory
- electrical engineering (bandwidth; signal-to-noise ratio)
- complexity theory (minimal description length)
- signal processing, representation, compressibility

Bilginin hareketleri ve dönüşümleri, tıpkı bir akışkanın hareketleri gibi, matematiksel ve fiziksel yasalarla sınırlandırılmıştır. Bu yasaların şunlarla derin bağlantıları vardır:

- olasılık teorisi, istatistik ve kombinatorik
- termodinamik (istatistiksel fizik)
- spektral analiz, Fourier (ve diğer) dönüşümleri
- örnekleme teorisi, tahmin, kestirim teorisi
- elektrik mühendisliği (bant genişliği; sinyal-gürültü oranı)
- karmaşıklık teorisi (minimum açıklama uzunluğu)
- sinyal işleme, gösterim, sıkıştırılabilirlik

9.1. A frequentist version of probability

Olasılığın frekansçı bir versiyonu:

If there are N distinct possible events (x1, x2, . . . , xN), no two of which can occur simultaneously, and the events occur with frequencies (n1, n2, . . . , nN), we say that the probability of event xi is given by

N tane farklı olası olay (x1, x2, . . . , xN) varsa ve bunlardan hiçbiri aynı anda gerçekleşemiyorsa ve olaylar (n1, n2, . . . , nN) sıklıklarında gerçekleşiyorsa, xi olayının olasılığının şu şekilde verildiğini söyleriz:

$$P(x_i) = \frac{n_i}{\sum_{j=1}^{N} n_j}$$

This definition has the nice property that

Bu tanımın güzel bir özelliği var

$$\sum_{i=1}^{N} P(x_i) = 1$$

An observer relative version of probability: In this version, we take a statement of probability to be an assertion about the belief that a specific observer has of the occurrence of a specific event. Note that in this version of probability, it is possible that two different observers may assign different probabilities to the same event. Furthermore, the probability of an event, for me, is likely to change as I learn more about the event, or the context of the event.

Gözlemciye göre olasılığın bir versiyonu: Bu versiyonda, bir olasılık ifadesini belirli bir gözlemcinin belirli bir olayın meydana geleceğine olan inancına dair bir iddia olarak ele alıyoruz. Bu olasılık versiyonunda, iki farklı gözlemcinin aynı olaya farklı olasılıklar atayabileceğini unutmayın. Dahası, benim için bir olayın olasılığı, olay veya olayın bağlamı hakkında daha fazla şey öğrendikçe muhtemelen değişecektir.

In some (possibly many) cases, we may be able to find a reasonable correspondence between these two views of probability. In particular, we may sometimes be able to understand the observer relative version of the probability of an event to be an approximation to the frequentist version, and to view new knowledge as providing us a better estimate of the relative frequencies.

Bazı (muhtemelen birçok) durumda, bu iki olasılık görüşü arasında makul bir ilişki bulabiliriz. Özellikle, bazen bir olayın olasılığının gözlemciye göreli versiyonunun frekansçı versiyona bir yaklaşım olduğunu anlayabilir ve yeni bilgiyi bize göreli frekansların daha iyi bir tahminini sağladığı şeklinde görebiliriz.

Some probability basics, where a and b are events:
Bazı olasılık temelleri, burada a ve b olaydır:
P(not a) = 1 − P(a).
P(a or b) = P(a) + P(b) − P(a and b).

We will often denote P(a and b) by P(a, b). If P(a, b) = 0, we say a and b are mutually exclusive.
P(a ve b)'yi sıklıkla P(a, b) ile gösteririz. Eğer P(a, b) = 0 ise, a ve b'nin karşılıklı olarak dışlayıcı olduğunu söyleriz.

Conditional probability:
P(a|b) is the probability of a, given that we know b. The joint probability of both a and b is given by:
Koşullu olasılık:
P(a|b), b'yi bildiğimizde a'nın olma olasılığıdır. Hem a'nın hem de b'nin birleşik olasılığı şu şekilde verilir:

P(a, b) = P(a|b)P(b).
Since P(a, b) = P(b, a), we have Bayes' Theorem:
P(a|b)P(b) = P(b|a)P(a),
or
P(a|b) = P(b|a)P(a)/P(b).

If two events a and b are such that P(a|b) = P(a), we say that the events a and b are independent. Note that from Bayes' Theorem, we will also have that P(b|a) = P(b), and furthermore, P(a, b) = P(a|b)P(b) = P(a)P(b). This last equation is often taken as the definition of independence.
Eğer iki olay a ve b, P(a|b) = P(a) olacak şekildeyse, a ve b olaylarının
bağımsız olduğunu söyleriz. Bayes Teoremi'nden, P(b|a) = P(b) ve ayrıca, P(a, b) = P(a|b)P(b) = P(a)P(b) olacağını da unutmayın. Bu son denklem genellikle bağımsızlığın tanımı olarak alınır.

We would like to develop a usable measure of the information we get from observing the occurrence of an event having probability p . Our first reduction will be to ignore any particular features of the event, and only observe whether or not it happened. Thus we will think of an event as the observance of a symbol whose probability of occurring is p. We will thus be defining the information in terms of the probability p.
Olasılığı p olan bir olayın meydana gelişini gözlemleyerek elde ettiğimiz bilgilerin kullanılabilir bir ölçüsünü geliştirmek istiyoruz. İlk indirgememiz, olayın herhangi bir özel

özelliğini göz ardı etmek ve yalnızca gerçekleşip gerçekleşmediğini gözlemlemek olacak. Dolayısıyla bir olayı, gerçekleşme olasılığı p olan bir sembolün gözlemlenmesi olarak düşüneceğiz. Dolayısıyla bilgiyi olasılık p açısından tanımlayacağız.
There is a list of four basic axioms to be used. This axiomatic system can be applied in any context that has a set of non-negative real numbers. A special case of interest is probabilities (i.e. real numbers between 0 and 1), which motivated the choice of axioms.
Kullanılacak dört temel aksiyomun bir listesi bulunmaktadır. Bu aksiyomatik sistem, negatif olmayan gerçek sayılar kümesine sahip olan herhangi bir bağlamda uygulayabileceğidir. İlgi çekici özel bir durum olasılıklardır (yani 0 ile 1 arasındaki gerçek sayılar), bu da aksiyomların seçilmesini motive etmiştir.

The information measure I(p) must have several properties:
1. Information is a non-negative quantity: $I(p) \geq 0$.
2. If the probability of an event is 1, we cannot get any information from the event occurring: $I(1) = 0$.
3. If two independent events occur (their joint probability is the product of their individual probabilities), then the information obtained from observing the events is the sum of the two pieces of information: $I(p1 * p2) = I(p1) + I(p2)$. (This is the critical property...)
4. We want the information measure to be a continuous (in fact monotonic) function of probability (small changes in probability should lead to small changes in information).

Bilgi ölçüsü I(p) birkaç özelliğe sahip olmalıdır:
1. Bilgi negatif olmayan bir niceliktir: $I(p) \geq 0$.
2. Bir olayın olasılığı 1 ise, olayın meydana gelmesinden hiçbir bilgi alamayız: $I(1) = 0$.
3. İki bağımsız olay meydana gelirse (ortak olasılıkları, bireysel olasılıklarının çarpımıdır), o zaman olayları gözlemlemekten elde edilen bilgi, iki bilginin toplamıdır: $I(p1 * p2) = I(p1) + I(p2)$. (Bu, kritik özelliktir...)
4. Bilgi ölçüsünün olasılığı sürekli (aslında monotonik) bir fonksiyon olmasını isteriz (olasılıktaki ufak değişiklikler, bilgide ufak değişikliklere yol açmalıdır).

We can therefore derive the following:
Dolayısıyla şu sonuca varabiliriz:
1. $I(p^2) = I(p * p) = I(p) + I(p) = 2 * I(p)$
2. Thus, further, $I(p^n) = n * I(p)$ (by induction . . .)
2. Böylece, daha da ileri giderek, $I(pn) = n * I(p)$ (tümevarım yoluyla . . .)
3. $I(p) = I((p^{1/m})^m) = m * I(p1/m)$,
so $I(p^{1/m}) = (1/m) * I(P)$ and thus in general $I(p^{n/m}) = (n/m) * I(p)$
yani $I(p1/m) = (1/m) * I(P)$ ve dolayısıyla genel olarak $I(pn/m) = (n/m) * I(p)$
4. And thus, by continuity, we get, for $0 < p \leq 1$, and $a > 0$ a real number: $I(p^a) = a * I(p)$
4. Ve böylece, süreklilik sayesinde, $0 < p \leq 1$ ve $a > 0$ için gerçek bir sayı elde ederiz: $I(pa) = a * I(p)$

From this, we can derive the nice property: $I(p) = -\log_b(p) = \log_b(1/p)$ for some base b.
Bundan güzel özelliği türetebiliriz: $I(p) = -\log b(p) = \log b(1/p)$ herhangi bir b tabanı için.

Summarizing: from the four properties,
Özetle: Dört özellikten,
1. $I(p) \geq 0$
2. $I(p1 * p2) = I(p1) + I(p2)$
3. I(p) is monotonic and continuous in p
3. I(p) monotoniktir ve p'de süreklidir
4. $I(1) = 0$

We can derive that
Şunu çıkarabiliriz,
$I(p) = \log b(1/p) = -\log b(p)$,
for some positive constant b. The base b determines the units we are using.
pozitif bir sabit b için. Taban b kullandığımız birimleri belirler.

We can change the units by changing the base, using the formulas, for b1, b2, x > 0,
Tabanı değiştirerek, b1, b2, x > 0 formüllerini kullanarak birimleri değiştirebiliriz.
$x = b1^{\log b1(x)}$

and therefore $\log b_2(x) = \log b_2(b1^{\log b1(x)}) = (\log b2(b1))(\log b1(x))$.
Ve yüzden $\log b_2(x) = \log b_2(b1^{\log b1(x)}) = (\log b2(b1))(\log b1(x))$.

Thus, using different bases for the logarithm results in information measures which are just constant multiples of each other, corresponding with measurements in different units:
1. $\log_2$ units are bits (from 'binary')
2. $\log_3$ units are trits(from 'trinary')
3. $\log_e$ units are nats (from 'natural logarithm') (We'll use ln(x) for loge(x))
4. $\log_{10}$ units are Hartleys, after an early worker in the field.
Bu nedenle, logaritma için farklı tabanlar kullanmak, farklı birimlerdeki ölçümlere karşılık gelen, birbirinin sabit katları olan bilgi ölçümleriyle sonuçlanır:
1. log2 birimleri bitlerdir ('ikili'den)
2. log3 birimleri tritlerdir ('üçlü'den)
3. loge birimleri nats'tır ('doğal logaritma'dan) (loge(x) için ln(x) kullanacağız)
4. log10 birimleri, alanda çalışan ilk kişilerden birinin adını taşıyan Hartley'lerdir.

Unless we want to emphasize the units, we need not bother to specifiy the base for the logarithm, and will write log(p). Typically, we will think in terms of $\log_2(p)$.

Birimleri vurgulamak istemiyorsak, logaritmanın tabanını belirtmeye gerek yok ve log(p) yazacağız. Genellikle, $\log_2(p)$ cinsinden düşüneceğiz.
For example, flipping a fair coin once will give us events h and t each with probability 1/2, and thus a single flip of a coin gives us $-\log_2(1/2) = 1$ bit of information (whether it comes up h or t).
Örneğin, adil bir parayı bir kez havaya atmak bize her biri 1/2 olasılıkla h ve t olaylarını verecektir ve dolayısıyla tek bir para atma bize - $\log_2(1/2) = 1$ bit bilgi (h veya t gelmesi fark etmez) verecektir.

Flipping a fair coin n times (or, equivalently, flipping n fair coins) gives us $-\log_2((1/2)^n) = \log_2(2^n) = n * \log_2(2) = n$ bits of information. We could enumerate a sequence of 25 flips as, for example: hthhtthhhhhtttthhhhhtt or, using 1 for h and 0 for t, the 25 bits 1011001011101000101110100.
Adil bir parayı n kez çevirmek (veya eşdeğer olarak, n adil parayı çevirmek) bize $-\log_2((1/2)^n) = \log_2(2^n) = n * \log_2(2) = n$ bit bilgi verir. 25 çevirmeden oluşan bir diziyi şu şekilde sayabiliriz: hthhtthhhhhtttthhhhhtt veya h için 1 ve t için 0 kullanarak 25 bit 1011001011101000101110100.

We thus get the nice fact that n flips of a fair coin gives us n bits of information, and takes n binary digits to specify. That these two are the same reassures us that we have done a good job in our definition of our information measure.
Böylece, adil bir madeni paranın n kez atılmasının bize n bit bilgi verdiği ve belirtmek için n ikili basamak gerektiği gibi güzel bir gerçeği elde ederiz. Bu ikisinin aynı olması, bilgi ölçütümüzün tanımında iyi bir iş çıkardığımıza dair bize güvence verir.

Most of probability theory was laid down by theologians: Blaise PASCAL (1623-1662) who gave it the axiomatization that we accept today; and Thomas BAYES (1702-1761) who expressed one of its most important and widely-applied propositions relating conditional probabilities.
Probability Theory rests upon two rules:
Olasılık teorisinin çoğu, bugün kabul ettiğimiz aksiyomizasyonu veren teologlar tarafından ortaya konmuştur: Blaise PASCAL (1623-1662); ve koşullu olasılıklarla ilgili en önemli ve yaygın olarak uygulanan önermelerinden birini ifade eden Thomas BAYES (1702-1761).
Olasılık Teorisi iki kurala dayanır:

Product Rule:
p(A, B) = “joint probability of both A and B” = p(A|B)p(B) or equivalently, = p(B|A)p(A)
Çarpım Kuralı:
p(A, B) = “hem A hem de B'nin ortak olasılığı” = p(A|B)p(B) veya eşdeğer olarak, = p(B|A)p(A)

Clearly, in case A and B are independent events, they are not conditionalized on each other and so p(A|B) = p(A) and p(B|A) = p(B), in which case their joint probability is simply p(A, B) = p(A)p(B).
Açıkça, A ve B bağımsız olaylarsa, bunlar birbirlerine bağlı değildir ve dolayısıyla p(A|B) = p(A) ve p(B|A) = p(B) olur; bu durumda bunların ortak olasılığı basitçe p(A, B) = p(A)p(B) olur.

Sum Rule:
If event A is conditionalized on a number of other events B, then the total probability of A is the sum of its joint probabilities with all B: p(A) = X B p(A, B) = XB p(A|B)p(B)
Toplama Kuralı:
Eğer olay A, bir dizi başka olay B'ye bağlıysa, o zaman A'nın toplam olasılığı, tüm B ile olan ortak olasılıklarının toplamıdır: p(A) = X B p(A, B) = XB p(A|B)p(B)

From the Product Rule and the symmetry that p(A, B) = p(B, A), it is clear that p(A|B)p(B) = p(B|A)p(A). Bayes' Theorem then follows: p(B|A) = p(A|B) p(B) / p(A)
Çarpım Kuralı ve p(A, B) = p(B, A) simetrisinden, p(A|B)p(B) = p(B|A)p(A) olduğu açıktır. Bayes Teoremi şu şekildedir: p(B|A) = p(A|B) p(B) / p(A)

The importance of Bayes' Rule is that it allows us to reverse the conditionalizing of events, and to compute p(B|A) from knowledge of p(A|B), p(A), and p(B). Often these are expressed as prior and posterior probabilities, or as the conditionalizing of hypotheses upon data.
Bayes Kuralı'nın önemi, olayların koşullandırılmasını tersine çevirmemize ve p(A|B), p(A) ve p(B) bilgisinden p(B|A)'yı hesaplamamıza olanak sağlamasıdır. Bunlar sıklıkla önsel ve sonradan olasılıklar veya hipotezlerin veriler üzerine koşullandırılması olarak ifade edilir.

9.1.1. Kanal Kapasitesi

Bilgi: Güç, fiziksel büyüklük
Binary numbering system (İkili sayı sistemi), bit: 1/0
Bilgisayar bilgiyi bit olarak alır, işler, saklar (bellek), transfer eder.
Bilgiyi işleme: Aritmetik, mantıksal, karşılaştırma, öteleme

Bigi teorisi Shannon ile başlar.
Bilgiyi transfer etmek bit/saniye
Haberleşme ortamı: Kablolu, kablosuz
Kanal: Band genişliği (Maksimum frekans ile minumum frekans arasındaki fark)
Kanal kapasitesi

Buhar makinesi: sıcak sudan buhar ve buharın döndürdüğü pervaneler, sıcak – soğuk ısı transferi
Termodinamiğin 2. yasası
Sıcağın tümünü sistemde kullanmak mümkün mü? Hayır, bir kısmı kullanılmaktadır (%80). Geri kalan kısmını sistem kullanır. (Su – Kap: su ısıtılırken önce kap ısınır.)
Kullanılmayan ısı entropidir.

Kanal kapasitesi entropi ilişkisi:
Gürültüsüz kanal kapasitesi: Nyquist, C=2Bm; 2^m=sembol sayısı, m: bir semboldeki bit sayısı, B: Band genişliği
Gürültüsüz kanal kapasitesi mümkün mü? Hayır.

Gürültülü kanal kapasitesini hesaplama Shannon, Entropi
Enerji=Güç * Zaman

C=B* Log2 (1+S/N)
A=1+S/N
C=B* Log10 (A) / Log 10 (2) ≈ 3*B*Log10 (A)

Kanal Kapasitesi:

- Data rate is the speed at which data is transmitted in one second.
- Frequency (f) is the number of vibrations or periods in one second. Hertz (Hz) or cycles per second.
- Data rate at which data can be communicated: In bits per second
- Bandwidth: The difference between the maximum frequency and minimum frequency of a signal.
- Constrained by transmitter and medium
- Period = time for one wave repetition (T), T = 1/f (Second)

- Phase: Indicates the relative positions of multiple signals at t=0. Relative position in time, from 0-2*pi

- Veri hızı, bir verinin iletilme hızıdır, bps, bit/saniye. Verilerin iletilebildiği veri hızı: Saniye başına bit cinsinden verilir. Veri iletim hızı, verici ve ortam tarafından kısıtlanmıştır.
- B, Bant genişliği: Bir sinyalin maksimum frekansı ile minimum frekansı aralığındaki farktır. Frekans (f), bir sinyali bir saniyedeki titreşim ya da peryod sayısıdır. Hertz (Hz) veya saniye başına devir.
- Periyot = bir dalganın tekrarı için geçen süre (T), T = 1/f (Saniye)
- Analog sinyalde faz, Faz: Birden fazla sinyallerin birbirlerine göre durumlarını t=0 noktasına göre belirtir. Zamandaki göreceli konum, 0-2*pi arası

9.1.2. Gürültüsüz Kanal Kapasitesi

Nyquist Örnekleme Teoremi:

When an analog signal is sampled, converted to digital and then converted back to an analog signal, the sampling frequency must be equal to or greater than twice the bandwidth or maximum frequency of the analog signal to obtain the same original signal.
Bir analog sinyal örneklenip sayısala dönüştürüldükten sonra yeniden analog işarete dönüştürüldüğünde aynı orijinal sinyalin elde edilmesi için örnekleme frekansı, analog sinyalin band genişliğinin ya da maksimum frekansının iki katına eşit ya da büyük olması gerekmektedir.

$$fs \geq 2*fm$$
$$fs \geq 2*B$$
$$Ts=1/fs$$

Here,
fs is the sampling frequency (Hz)
fm is the maximum frequency in the signal (Hz)
B is the bandwidth (Hz).
Burada,
fs örnekleme frekansıdır (Hz)
fm sinyaldeki maksimum frekanstır (Hz)
B bant genişliğidir (Hz).

Analog signals are propagated by vibration. The frequency of an analog signal is the number of vibrations or periods per second. In telephone communication, the bandwidth is taken as 4KHz. Because the basic characteristics of sound, which are understanding, recognition and feeling, are covered by the 0-4KHz range. Therefore, fs=8KHz. Its period is

T=1/8KHz=125microseconds. In voice transmission, it is sufficient to send 8 bits from a channel every 125 microseconds.
Analog sinyal titreşerek yayılır. Analog sinyalin frekansı bir saniyedeki titreşim ya da peryod sayısıdır. Telefon haberleşmesinde band genişliği 4KHz alınır. Çünkü sesin temel özelliği olan anlama, tanıma, hssetme özelliklerini 0-4KHz aralığı kapsamaktadır. O halde fs=8KHz olur. Peryodu, T=1/8KHz=125mikrosaniyedir. Ses iletiminde 125 mikrosaniyede bir kanaldan 8 bit gönderilmesi yeterlidir.

Channel capacity in noiseless environment:

Gürültüsüz Ortamda kanal kapasitesi:

The Nyquist theorem calculates the channel capacity for a noiseless channel:
Nyquist teoremi gürültüsüz bir kanal için kanal kapasitesini hesaplar:

$C = 2\ B \log_2 2^m = 2\ B\ m$

m: number of bits representing a symbol or a sample (total number of symbols, $n=2^m$)
m: bir sembolün ya da bir örneklemin temsil edildiği bit sayısı (toplam sembol sayısı, $n=2^m$)

n: is the number of amplitude intervals in the sampling process. Quantization is the number of amplitude levels.
n: örneklem işleminde genlik aralığı sayısıdır. Quantalama genlik seviye sayısıdır.

C: is the channel capacity, the number of bits transferred from a channel in one second. Its unit is: bit/second.
C: kanal kapasitesidir, bir kanaldan bir saniyede transfer edilen bit sayısıdır. Birimi: bit/saniye dir.

B: Bandwidth, its unit is Hz.
B: Band genişliğidir, birimi Hz dir.

If m is the number of bits represented in a symbol, the total number of amplitude levels or the number of symbols is $n = 2^m$, where n is the number of sampling intervals.
m: bir sembolde temsil edilen bit sayısı ise toplam genlik seviye sayısı ya da sembol sayısı, $n=2^m$ dir, burada n: örnekleme aralığı sayısıdır.

Example:

B=10MHz, m=8bit ise C=?

C=2*B*m=2*10*10^6*8=160*10^6=160Mbps

9.1.3. Noisy

Gürültü:

Noise is an external interference with the signals used for communication. Noise is unwanted signals that disrupt signals. Noise is random, unwanted electrical energy that enters communication systems and environments through the medium of communication and interferes with the message being transmitted. It is additional signals added between the transmitter and receiver. Noise (also called random electromagnetic radiation) permeates the environment. Even communication systems produce small amounts of electrical noise as a side effect of normal operation.

Gürültü, iletişim için kullanılan sinyallere dışarıdan müdahale edebilmesidir. Gürültü, sinyalleri bozan istenmeyen işaretlerdir. Gürültü, iletişim sistemlerine ve ortamlarına iletişim aracıyla giren ve iletilen mesaja müdahale eden rastgele, istenmeyen elektriksel enerjidir. Verici ve alıcı arasına eklenen ek sinyallerdir. Gürültü (rastgele elektromanyetik radyasyon olarak da adlandırılır) çevreye nüfuz eder. İletişim sistemleri bile normal çalışmanın bir yan etkisi olarak küçük miktarlarda elektriksel gürültü üretir.

There are different types of noise:

Induced: From motors and devices, devices are in the environment as transmitter antenna and receiver antenna.

Thermal: It is formed due to thermal radiation of electrons. It is also called white noise.

Intermodulation: Unwanted signals that are the sum and difference of the original frequencies sharing a medium.

Crosstalk: The signal from one line is received by the other

Impulse: Irregular pulses or sudden spikes. For example. External electromagnetic interference. It is short-term. It is high amplitude.

Farklı gürültü türleri vardır:

Endüklenen (Induced): Motorlardan ve cihazlardan, cihazlar verici anten ve alıcı anten olarak ortamdadır.

Termal: Elektronların termal ışımasından dolayı oluşur. Beyaz gürültü olarak da adlandırılır.

Intermodülasyon: Bir ortamı paylaşan orijinal frekansların toplamı ve farkı olan istenmeyen sinyallerdir.

Karışma (Crosstalk): Bir hattan gelen sinyal, diğeri tarafından alınır

Dürtü: Düzensiz darbeler veya ani yükselmeler. Örneğin. Harici elektromanyetik girişim. Kısa sürelidir. Yüksek genliklidir.

Signal to Noise Ratio (SNR):
Sinyalin Gürültüye Oranı(SNR):

SNR is often used to measure the quality of a system. It shows the signal strength relative to the noise strength in the system. It is the ratio between the two powers.
Bir sistemin kalitesini ölçmek için SNR sıklıkla kullanılır. Sistemdeki gürültü gücüne göre sinyalin gücünü gösterir. İki güç arasındaki orandır.

It is usually given in dB and referred to as SNRdB.
Genellikle dB olarak verilir ve SNRdB olarak adlandırılır.

Example:
Örnek:
If the signal level is 100 milliwatts and the noise level is 0.01 milliwatts, calculate the signal to noise ratio in dB. If the ratio is less than 30dB, the system is not working. Then comment on the value given in the example.
Sinyal seviyesi 100miliwatt, gürültü seviyesi ise 0.01miliwatt ise Sinyalin gürültüye oranını dB olarak hesaplayınız. Oran 30dB'den küçük ise sistem çalışmamaktadır. O zaman örnekte verilen değeri yorumlayınız.

KdB=10Log(Signal/Noise)
KdB=10Log(Sinyal/Gürültü)

KdB=10Log(100/0.01)=10Log(10000)=40dB

Since KdB>=30, the system works.
KdB>=30 olduğundan sistem çalışır.

Example:

Örnek:

If the signal level is 10 milliwatts and the noise level is 0.1 milliwatts, calculate the signal to noise ratio in dB. If the ratio is less than 30dB, the system is not working. Then comment on the value given in the example.

Sinyal seviyesi: 10miliwatt, gürültü seviyesi: 0.1miliwatt ise Sinyalin gürültüye oranını dB olarak hesaplayınız. Oran 30dB'den küçük ise sistem çalışmamaktadır. O zaman örnekte verilen değeri yorumlayınız.

KdB=10Log(Signal/Noise)

KdB=10Log(Sinyal/Gürültü)

KdB=10Log(10/0.1)=10Log(100)=20dB

Since KdB<30, the system does not work.

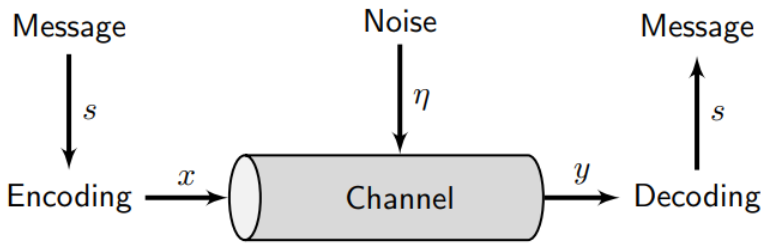
KdB<30 olduğundan sistem çalışmaz.

9.1.4. Shannon Channel Capacity Theorem

Shannon Teoremi:

In 1948, Claude Shannon published a paper called "A Mathematical Theory of Communication". This paper gave birth to the field of Information Theory. Information Theory, by definition, is the study of the quantification, storage and communication of information. But it is much more than that. It has made significant contributions to fields such as Statistical Physics, Computer Science, Economics. The main focus of Shannon's paper was the general communication system, as he was working at Bell Labs when he published the paper. He established several important concepts such as information entropy and redundancy. Today, the foundations of the theorem are applied in the fields of lossless data compression, lossy data compression and channel coding.

1948'de Claude Shannon, "A Mathematical Theory of Communication" adlı bir makale yayınladı. Bu makale Bilgi Teorisi alanını doğurdu. Bilgi Teorisi, tanım olarak, bilginin nicelenmesi, depolanması ve iletişiminin incelenmesidir. Ama bundan çok daha fazlası. İstatistiksel Fizik, Bilgisayar Bilimleri, Ekonomi gibi alanlara önemli katkılar sağlamıştır. Shannon'ın makalesinin ana odak noktası, makaleyi yayınladığında Bell Laboratuvarlarında çalıştığı için genel iletişim sistemiydi. Bilgi entropisi ve fazlalık gibi birkaç önemli kavram oluşturdu. Günümüzde teoremin temelleri kayıpsız veri sıkıştırma, kayıplı veri sıkıştırma ve kanal kodlama alanlarında uygulanmaktadır.



A message (data) is encoded before being used as input to a communication channel, which causes errors, as data interacts with noise as it is transmitted through the channel. The channel output is decoded by a receiver to recover the message.

Bir mesaj (veri), bir iletişim kanalına girdi olarak kullanılmadan önce kodlanır ve bu da hataya neden olur, veri kanaldan iletilirken gürültü ile etkileşime girer. Kanal çıktısı, mesajı kurtarmak için bir alıcı tarafından çözülür.

Information theory defines absolute, insurmountable limits on how much information can be transmitted between any two components of any system. The theorems of information theory are so important that they deserve to be considered information laws. The basic information laws for any communication channel can be summarized as follows.

Bilgi teorisi, herhangi bir sistemin herhangi iki bileşeni arasında ne kadar bilginin iletilebileceğine dair kesin, aşılamaz sınırlar tanımlar. Bilgi teorisinin teoremleri o kadar

önemlidir ki, bilgi yasaları olarak kabul edilmeyi hak ederler. Herhangi bir iletişim kanalı için temel bilgi yasaları aşağıdaki gibi özetlenebilir.

There is a certain upper limit to the amount of information that can be transmitted over a communication channel and the channel capacity, this limit decreases as the amount of noise in the channel increases, this limit can be almost reached by clever packaging or coding of the data.
Bir iletişim kanalından iletilebilecek bilgi miktarının kesin bir üst sınırı ve kanal kapasitesi vardır, bu sınır, kanaldaki gürültü miktarı arttıkça küçülür, bu sınıra, verilerin akıllıca paketlenmesi veya kodlanmasıyla neredeyse ulaşılabilir.

Shannon Channel Capacity Theorem:
The Shannon channel capacity theorem defines the theoretical maximum bit rate (number of bits second) for a noisy channel. Every communication channel has a speed limit, measured in bps. This famous Shannon limit and the formula for capacity of communication channel is given as.
Shannon Kanal Kapasitesi Teoremi:
Shannon kanal kapasite teoremi, gürültülü bir kanal için teorik maksimum bit hızını (saniyedeki bit sayısı) tanımlar. Her iletişim kanalının bps cinsinden ölçülen bir hız sınırı vardır. Bu ünlü Shannon sınırı ve iletişim kanalının kapasitesi formülü şu şekilde verilmiştir.

Capacity $C = B\ x\ Log_2\ (1 + \frac{S}{N})$

Where, C: Channel Capacity (bps)
B: Bandwidth of signal (Hz)
S: Signal power (Watt)
N: Noise power (Watt)
Burada, C: Kanal Kapasitesi (bps)
B: Sinyalin bant genişliği (Hz)
S: Sinyal gücü (Watt)
N: Gürültü gücü (Watt)

SNR=S/N=Signal Power / Noise Power (no unit)
Channel capacity is the ultimate transmission rate of a communication system.
Channel capacity is defined as the ability of a channel to convey information.
Kanal kapasitesi, bir iletişim sisteminin nihai iletim hızıdır.
Kanal kapasitesi, bir kanalın bilgi iletme yeteneği olarak tanımlanır.

Bad news:
It is matematically impossible to get error free communication above the limit.
No matter how sophisticated error correction scheme we use.
No matter how much we compress the data.
We can not maket he channel go faster than the limit without losing some information.
Kötü haber:
Sınırın üzerinde hatasız iletişim elde etmek matematiksel olarak imkansızdır.
Ne kadar karmaşık hata düzeltme şeması kullanırsak kullanalım.
Veriyi ne kadar sıkıştırırsak sıkıştıralım.
Biraz bilgi kaybetmeden kanalın sınırdan daha hızlı gitmesini sağlayamayız.

Good news:
Below the Shannon limit, it is possible to transmit the information with zero error. Shannon mathematicilally proved that use of encoding techniques allow us to reach maximum limit of channel capacity without any errors regardless of amount of noise.
İyi haber:
Shannon sınırının altında, bilgileri sıfır hatayla iletmek mümkündür. Shannon, kodlama tekniklerinin kullanımının, gürültü miktarından bağımsız olarak herhangi bir hata olmadan kanal kapasitesinin maksimum sınırına ulaşmamızı sağladığını matematiksel olarak kanıtladı.

If the (1+ SNR) value is written as $Log_2\ 2^n=n$, taking logarithms in base 2 becomes simpler.
(1+ SNR) değeri $Log_2\ 2^n=n$ olarak yazılırsa 2 tabanında logaritma alma basitleşmiş olur.

$C = B\ log_2(1 + SNR)$
C=n*B olur.

For example, if SNR=20000, B=10MHz, calculate the channel capacity in a noisy environment. 20000 >= 2^14, n=14 is taken. C=n*B=14*10*10^6=140*10^6=140Mbps is found.
Örneğin, SNR=20000, B=10MHz ise gürültülü ortamda kanal kapasitesini hesaplayınız.
20000 >= 2^14, n=14 alınır. C=n*B=14*10*10^6=140*10^6=140Mbps bulunur.

Note: SNR is usually given in decibels in base 10.
If 10Log10 SNR=SNRdB then SNR=10SNRdB/10.
If 10SNRdB/10=2n then SNRdB *Log1010 /10=n*log10 2. Log10 2=0.3. SNRdB/3=n is obtained.
Not: SNR genellikle 10 tabanında desibel olarak verilir.
10Log10 SNR=SNRdB ise SNR=10SNRdB/10 dir.
10SNRdB/10=2n ise SNRdB *Log1010 /10=n*log10 2 dir. Log10 2=0.3 dir. SNRdB/3=n elde edilir.

SNRdB/3=n is found. Capacity is found as bps from C=n*B. Shannon capacity gives us the upper limit; Nyquist formula tells us how many signal levels we need. SNR level cannot be lower than the threshold value. Threshold values in communication systems must be greater than 6dB.

SNRdB/3=n bulunur. C=n*B den kapasite bps olarak bulunur. Shannon kapasitesi bize üst sınırı veriyor; Nyquist formülü bize kaç tane sinyal seviyesine ihtiyacımız olduğunu söyler. SNR seviyesi eşik değerden düşük olamaz. Haberleşme sistemlerinde eşik değerler 6dB'den büyük olmak zorundadır.

Example:

Calculate the bit rate for a noisy channel with SNR=300 and bandwidth of 3KHz.

SNR=300 ve bant genişliği 3KHz, n=8bit olan sinyalin gürültüsüz ve gürültülü bir kanal için bit hızını hesaplayınız. Yorumlayınız.

Gürültüsüz iletişim ortamı için kanal kapasitesi:

C=2*n*B=2*8*3000=48000bps is found.

C=2*n*B=2*8*3000=48000bps = 48 Kbps bulunmuş olur.

The bit rate for noisy channel according to Shannon theorem can be calculated as follows:

Ya da gürültülü kanal için bit hızı Shannon teoremine göre aşağıdaki şekilde hesaplanabilir:

Capacity $C = B\ x\ Log_2\ (1 + \frac{S}{N})$

$$Log_a x = \frac{Log_{10}\ x}{Log_{10} a}$$

$$Log_2 301 = \frac{Log_{10}\ 301}{Log_{10} 2}$$

Log_{10} (301)=2.48

Log_{10} (2)=0.3

Capacity C=3000* 2.48/0.3

C=10000*2,48=24 800 bit/saniye =24,8Kbit/s

Comment: While the channel capacity in a quiet environment was 48Kbps, the channel capacity in a noisy environment was found to be 24.8Kbps. Note: Do not get emotional in your comments.

Yorum: Gürültüsüz ortamda kanal kapasitesi 48Kbps iken gürültülü ortamda kanal kapasitesi 24,8Kbps olarak bulunmuştur.

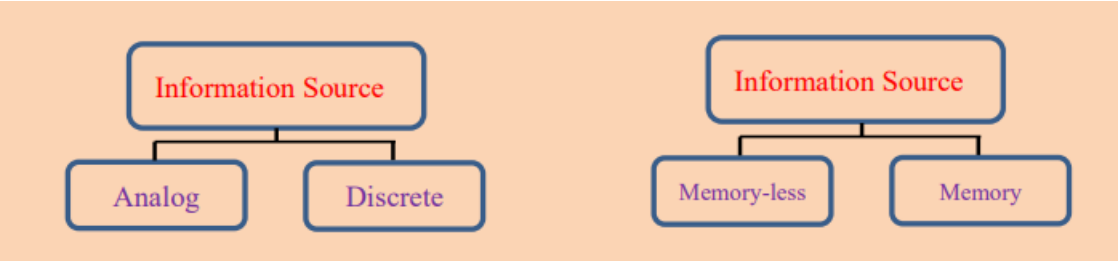
Not: Sakın ha yorumlarınıza duygusallık karıştırmayın.

9.1.5. DMS (Discrete Memory-less Source)

DMS (Ayrık Belleksiz Kaynak)

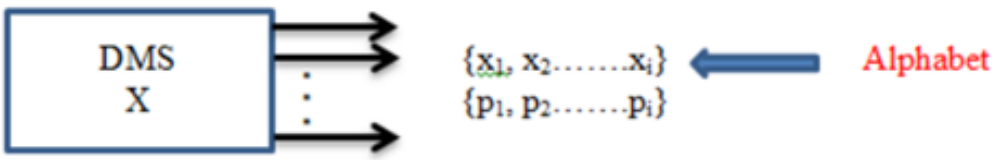
Information sources can be classified as either memory or memory-less. A memory-less source is one for which each symbol produced is independent of previous symbols. A memory source is one for which a current symbol depends on the previous symbols. An information source is an object that produces an event. A practical source in a communication system is a device that produces messages. It can be either analog or digital (discrete). Analog sources can be transformed into discrete sources through the use of sampling and quantization techniques.

Bilgi kaynakları hafızalı veya hafızasız olarak sınıflandırılabilir. Hafızasız kaynak, üretilen her sembolün önceki sembollerden bağımsız olduğu kaynaktır. Hafıza kaynağı, geçerli sembolün önceki sembollere bağlı olduğu kaynaktır. Bilgi kaynağı, bir olay üreten bir nesnedir. Bir iletişim sistemindeki pratik kaynak, mesajlar üreten bir cihazdır. Analog veya dijital (ayrık) olabilir. Analog kaynaklar, örnekleme ve niceleme tekniklerinin kullanımıyla ayrık kaynaklara dönüştürülebilir.



A discrete information source is a source that has only a finite set of symbols as possible outputs.

Ayrık bilgi kaynağı, olası çıktıları yalnızca sonlu sayıda sembolden oluşan bir kaynaktır.



Let X having alphabets {x1, x2,.....xm}. Note that set of source symbols is called source alphabet.

X'in {x1, x2,.....xm} alfabeleri olduğunu varsayalım. Kaynak sembolleri kümesine kaynak alfabesi adı verildiğini unutmayın.

A DMS is described by the list of symbols, probability assignment to these symbols and the specification of rate of generating these symbols by source. A discrete information source

consists of a discrete set of letters or symbols. In general, any message emitted by the source consists of a string or sequence of symbols.
Bir DMS, semboller listesi, bu sembollere olasılık ataması ve bu sembollerin kaynak tarafından üretilme oranının belirtilmesiyle tanımlanır. Ayrık bir bilgi kaynağı, ayrı bir harf veya sembol kümesinden oluşur. Genel olarak, kaynak tarafından yayılan herhangi bir mesaj bir dizi veya sembol dizisinden oluşur.

Measure of information:

Bilgi ölçüsü:

The amount of information in a message depends only upon the uncertainty of the event. Amount of information received from the knowledge of occurrence of an event may be related to the probability of occurrence of that event. Few messages received from different sources will contain more information than others. Information should be proportional to the uncertainty (doubtfulness) of an outcome. According to Hartley, information in a message is a logarithmic function. Let m1 is any message, then information content of this message is given as:
Bir mesajdaki bilgi miktarı yalnızca olayın belirsizliğine bağlıdır. Bir olayın meydana geldiği bilgisinden alınan bilgi miktarı, o olayın meydana gelme olasılığıyla ilişkili olabilir. Farklı kaynaklardan alınan birkaç mesaj diğerlerinden daha fazla bilgi içerecektir. Bilgi, bir sonucun belirsizliğiyle (şüpheliliğiyle) orantılı olmalıdır. Hartley'e göre, bir mesajdaki bilgi logaritmik bir fonksiyondur. m1 herhangi bir mesaj olsun, o zaman bu mesajın bilgi içeriği şu şekilde verilir:

$I(m1) = \log_2 1/p(m1)$
I(m1) : Information in message m1
p(x1) : probability of occurrence of message m1
$I(m1) = \log 2\ 1/p(m1)$
I(m1) : m1 mesajındaki bilgi
p(x1) : m1 mesajının oluşma olasılığı

Let source, X having alphabets {x1, x2,…..xm}
Information of symbol x1 = I(x1)
Information of symbol x2 = I(x2)
Information of symbol x3 = I(x3)
Information of symbol xi = I(xi)
Kaynak, {x1, x2,…..xm} alfabelerine sahip X olsun
Sembol bilgisi x1 = I(x1)
Sembol bilgisi x2 = I(x2)
Sembol bilgisi x3 = I(x3)
Sembol bilgisi xi = I(xi)

$I(x1) = \log_2 1/ P(x1)$
$I(x2) = \log_2 1/P(x2)$
$I(x3) = \log_2 1/P(x3)$
$I(x\mathrm{m}) = \log_2 1/P(xi)$

p(x1) = probability of occurrence of symbol x1
p(x2) = probability of occurrence of symbol x2
p(x3) = probability of occurrence of symbol x3
p(xi) = probability of occurrence of symbol xi
p(x1) = x1 sembolünün oluşma olasılığı
p(x2) = x2 sembolünün oluşma olasılığı
p(x3) = x3 sembolünün oluşma olasılığı
p(xi) = xi sembolünün oluşma olasılığı

Properties of I(xi)
I(xi)'nin özellikleri
i. I(x1) = 0 if p(x1) =1
ii. I(x1) ≥ 0
iii. I(x1, x2) = I(x1) + I(x1) if x1 and x1 are independent
iii. x1 ve x1 bağımsızsa I(x1, x2) = I(x1) + I(x1)
iv. I(x1) > I(x2) if p(x1) < p(x2)

Example:
A source generates one of the 5 possible messages during each message interval. The probabilities of these messages are:
Bir kaynak, her mesaj aralığında 5 olası mesajdan birini üretir. Bu mesajların olasılıkları şunlardır:
$p1$ =1/2, $p2$ =1/16, $p3$ =1/8, $p4$ =1/4, $p5$ =1/16.

Find the information content of each message.
$I(m1) = \log_2 (1/p1) = \log_2 2 = 1\ bit$ Information in message m1
$I(m2) = \log_2 (1/p2) = \log_2 16 = 4\ bits$ Information in message m2
$I(m3) = \log_2 (1/p3) = \log_2 8 = 3\ bit\ s$ Information in message m3
$I(m4) = \log_2 (1/p4) = \log_2 4 = 2\ bit\ s$ Information in message m4
$I(m5) = \log_2 (1/p5) = \log_2 16 = 4\ bit\ s$ Information in message m5
Her mesajın bilgi içeriğini bulun. $I(m1) = \log2\ (1/p1) = \log2\ 2 = 1\ bit$ m1 mesajındaki bilgi
$I(m2) = \log2\ (1/p2) = \log2\ 16 = 4\ bits$ m2 mesajındaki bilgi
$I(m3) = \log2\ (1/p3) = \log2\ 8 = 3\ bit\ s$ m3 mesajındaki bilgi
$I(m4) = \log2\ (1/p4) = \log2\ 4 = 2\ bit\ s$ m4 mesajındaki bilgi
$I(m5) = \log2\ (1/p5) = \log2\ 16 = 4\ bit\ s$ m5 mesajındaki bilgiler

Example:

In a binary PCM if 0 occurs with probability ¼ and 1 occurs with probability ¾, then calculate the amount of information carried by each bit.

$$I(bit\ 0) = \log_2 \frac{1}{P(x_1)} = \log_2 \frac{1}{\frac{1}{4}} = \log_2 4 = \log_2 2^2 = 2\log_2 2 = 2 * 1 = 2\ bits$$

$$I(bit\ 1) = \log_2 \frac{1}{P(x_2)} = \log_2 \frac{1}{\frac{3}{4}} = \log_2 \frac{4}{3} = \log_2 1.33 = \frac{\log_{10} 1.33}{\log_{10} 2} = \frac{0.125}{0.301} = 0.415\ bits$$

- Bit 0 has probability ¼ and it has 2 bits of information
- Bit 1 has probability ¾ and it has 0.415 bits of information

Example:

In a binary PCM if 0 occurs with probability ¼ and 1 occurs with probability ¾, then calculate the amount of information carried by each bit.

$I(bit\ 0)$ = $\log_2 1/P(x1)$= $\log_2 1/(1/4)$= $\log_2 4$ = $\log_2 2^2$ = 2 $\log_2 2$ = 2 $*$ 1 = 2 $bits$

$I(bit\ 1)$ = $\log_2 1/P(x2)$= $\log_2 1/(3/4)$ = $\log_2 4/3$ = $\log_2 1.33$ = $\log_{10} 1.33$ /$\log_{10} 2$ =0.125/0.301 = 0.415 $bits$

Bit 0 has probability ¼ and it has 2 bits of information

Bit 1 has probability ¾ and it has 0.415 bits of information

Example:

If I(x1) is the information carried by symbol x1 and I(x2) is the information carried by message x2 then prove that the amount of information carried compositely due to x1 and x2 is I(x1, x2) = I(x1) + I(x2)

Symbol = message

$I(x1)$ = $\log_2 1/p(x1)$

$I(x2)$ = $\log_2 1/p(x2)$

p(x1) = probability of occurrence of message (or symbol x1)

p(x2) = probability of occurrence of message (or symbol x2)

Note that messages x1 and x2 are independent, therefore

Joint probability p(x1, x2) = p(x1) * p(x2)

Information carried compositely due to x1 and x2 is

$I(x1, x2)$ = $\log_2 1/P(x1,x2)$= $\log_2 1/(p(x1).p(x2))$= $\log_2$ [$1/p(x1)$ * $1/p(x2)$] = $\log_2 1/P(x1)$ + $\log_2 1/P(x2)$= $I(x1)$+ $I(x2)$

[$since$ $\log_2 ab$ = $\log_2 a$ + $\log_2 b$]

Thus proved.

Information contained in independent outcomes should add. So, Information given by two independent messages is the sum of the information contained in individual messages.

Bağımsız sonuçlarda bulunan bilgiler eklenmelidir. Yani, iki bağımsız mesaj tarafından verilen bilgiler, bireysel mesajlarda bulunan bilgilerin toplamıdır.

9.1.6. Entropy theory

Entropi teorisi:

The Second Law of Thermodynamics: "Not all energy can be converted into useful work, some of it is used to maintain the internal integrity of the system." According to the second law, the entropy change in a system and its surroundings in any process is either "zero" or "positive." The thermodynamic term representing thermal energy that cannot be converted into mechanical work is called entropy. It is the expression of disorder in a system. If there is a thermal process, entropy is either equal to zero or positive. Entropy is used not only in thermodynamics in daily life, but also in many areas from statistics to theology. The entropy of the universe tends to constantly increase.
Termodinamiğin İkinci Yasası: ***"Enerjinin tamamı faydalı işe çevrilemez, bir kısmı sistemin içsel bütünlüğünü korumak için kullanılır."*** İkinci yasaya göre, herhangi bir süreçte bir sistem ve çevresindeki entropi değişimi ya "sıfır" ya da "pozitiftir". Mekanik işe çevrilemeyecek termal enerjiyi temsil eden **termodinamik** terimine **entropi** denilir. Bir sistemdeki düzensizliğin ifadesidir. Termal bir işlem varsa **entropi** ya sıfıra eşit olur ya da pozitif değer alır. Entropi, gündelik hayatta sadece **termodinamikte** değil, istatistikten teolojiye birçok alanda kullanılır. Evrenin entropisi sürekli artma eğilimindedir.

Entropy is average information (bits/symbol).
Entropi ortalama bilgidir (bit/sembol).

Suppose now that we have n symbols {a1, a2, . . . , an}, and some source is providing us with a stream of these symbols. Suppose further that the source emits the symbols with probabilities {p1, p2, . . . , pn}, respectively. For now, we also assume that the symbols are emitted independently (successive symbols do not depend in any way on past symbols).
What is the average amount of information we get from each symbol we see in the stream?
Şimdi n tane sembolümüz olduğunu varsayalım {a1, a2, . . . , an} ve bir kaynak bize bu sembollerden oluşan bir akış sağlıyor. Ayrıca kaynağın sembolleri sırasıyla {p1, p2, . . . , pn} olasılıklarıyla yaydığını varsayalım. Şimdilik sembollerin bağımsız olarak rasgele yayıldığını da varsayıyoruz (ardışık semboller hiçbir şekilde geçmiş sembollere bağlı değildir). **Akışta gördüğümüz her sembolden aldığımız ortalama bilgi miktarı nedir?**

What we really want here is a weighted average. If we observe the symbol a_i, we will get be getting log(1/pi) information from that particular observation. In a long run (say N) of observations, we will see (approximately) $N * pi$ occurrences of symbol a_i (in the frequentist

sense, that's what it means to say that the probability of seeing a_i is p_i). Thus, in the N (independent) observations, we will get total information I of

Burada gerçekten istediğimiz şey ağırlıklı bir ortalamadır. Eğer ai sembolünü gözlemlersek, o belirli gözlemden log(1/pi) bilgi elde ederiz. Uzun bir gözlem serisinde (diyelim ki N), ai sembolünün (yaklaşık olarak) N ∗ pi oluşumunu göreceğiz (frekansçı anlamda, bu, ai'yi görme olasılığının pi olduğu anlamına gelir). Bu nedenle, N (bağımsız) gözlemde, toplam bilgi I elde edeceğiz.

$$I = \sum_{i=1}^{n} (Np_i) * Log(\frac{1}{p_i})$$

But then, the average information we get per symbol observed will be

Ancak o zaman, gözlemlenen sembol başına elde ettiğimiz ortalama bilgi şu olacaktır:

$$I/N = (\frac{1}{N}) \sum_{i=1}^{n} (Np_i) * Log(\frac{1}{p_i})$$

$$I/N = \sum_{i=1}^{n} (p_i) * Log(\frac{1}{p_i})$$

Note that limx→0 x ∗ log(1/x) = 0, so we can, for our purposes, define pi ∗ log(1/pi) to be 0 when pi = 0.

limx→0 x ∗ log(1/x) = 0 olduğunu unutmayın, dolayısıyla bizim amacımız doğrultusunda pi = 0 olduğunda pi ∗ log(1/pi)'nin 0 olduğunu tanımlayabiliriz.

This brings us to a fundamental definition. This definition is essentially due to Shannon in 1948, in the seminal papers in the field of information theory. As we have observed, we have defined information strictly in terms of the probabilities of events. Therefore, let us suppose that we have a set of probabilities (a probability distribution) P = {p1, p2, . . . , pn}. We define the entropy of the distribution P by:

Bu bizi temel bir tanıma getiriyor. Bu tanım esasen Shannon'ın 1948'de bilgi teorisi alanındaki öncü makalelerinde ortaya atılmıştır. Gördüğümüz gibi, bilgiyi kesin olarak olayların olasılıkları açısından tanımladık. Bu nedenle, bir olasılık kümesine (bir olasılık dağılımına) sahip olduğumuzu varsayalım (P = {p1, p2, . . . , pn}). Dağılımın entropisini şu şekilde tanımlarız:

$$H(p) = \sum_{i=1}^{n} (p_i) * Log2\left(\frac{1}{p_i}\right) = -\sum_{i=1}^{n} (p_i) * Log2(p_i)$$

If we are talking about a continuous probability distribution rather than a discrete probability distribution, the obvious generalization for P(x) will be stated here:
Ayrık bir olasılık dağılımı değilde sürekli bir olasılık dağılım söz konuş ise, P(x) için bariz genelleme burada belirtilecektir:

$$H(P) = \int P(x) * Log\left(\frac{1}{P(x)}\right) dx.$$

Another worthwhile way to think about this is in terms of expected value. Given a discrete probability distribution P = {p1, p2, . . . , pn}, with pi ≥ 0 and
Bunu düşünmenin bir diğer değerli yolu beklenen değer açısından düşünmektir. Ayrık bir olasılık dağılımı P = {p1, p2, . . . , pn}, pi ≥ 0 ve

$$\sum_{i=1}^{n}(p_i) = 1$$

or a continuous distribution P(x) with P(x) ≥ 0 and $\int P(x) = 1$, we can define the expected value of an associated discrete set F = {f1, f2, . . . , fn} or function F(x) by:
veya P(x) ≥ 0 ve ∫P(x) =1 olan sürekli bir dağılım P(x) için, ilişkili ayrık küme F = {f1, f2, . . . , fn} veya fonksiyon F(x)'in beklenen değerini şu şekilde tanımlayabiliriz:

$$< F >= \sum_{i=1}^{n}(f_i p_i)$$

or

$$< F(x) >= \int F(x)P(x)dx.$$

With these definitions, we have that: H(P) =< I(p) > . In other words, the entropy of a probability distribution is just the expected value of the information of the distribution.
Bu tanımlarla şunu elde ederiz: H(P) =< I(p) > . Başka bir deyişle, bir olasılık dağılımının entropisi, dağılım bilgisinin beklenen değeridir.

What this means is that 0 ≤ H(P) ≤ log(n).
We have H(P) = 0 when exactly one of the pi's is one and all the rest are zero.
We have H(P) = log(n) only when all of the events have the same probability 1/n.
That is, the maximum of the entropy function is the log() of the number of possible events, and occurs when all the events are equally likely.
Bunun anlamı, 0 ≤ H(P) ≤ log(n)'dir.
Pi'lerden tam olarak biri bir ve geri kalanların hepsi sıfır olduğunda H(P) = 0'a sahibiz.
Tüm olayların aynı olasılığa sahip olduğu 1/n olduğunda H(P) = log(n)'e sahibiz.

Yani, entropi fonksiyonunun maksimum değeri olası olayların sayısının log()'udur ve tüm olaylar eşit derecede olası olduğunda meydana gelir.

Entropy H(X) = Average information of a source X
• Here average information rather than information of a single symbol.
• This average information is called entropy.
• Also note that Entropy defines ultimate data compression of a source. Data compression provides a practical means for the efficient storage and transmission of data (text, audio, video etc.)
• Larger entropy represents larger average information.
Entropy is defined as H(X) and is given by:
Entropi H(X) = Bir kaynak X'in ortalama bilgisi
• Burada tek bir sembolün bilgisi yerine ortalama bilgi.
• Bu ortalama bilgiye entropi denir.
• Ayrıca Entropi'nin bir kaynağın nihai veri sıkıştırmasını tanımladığını unutmayın. Veri sıkıştırma, verilerin (metin, ses, video vb.) verimli bir şekilde depolanması ve iletilmesi için pratik bir araç sağlar.
• Daha büyük entropi daha büyük ortalama bilgiyi temsil eder.

Entropi H(X) olarak tanımlanır ve şu şekilde verilir:

$$H(x) = \sum_{i=1}^{M} p(x_i) I(x_i) = \sum_{i=1}^{M} p(x_i) Log_2 \frac{1}{p(x_i)} = -\sum_{i=1}^{M} p(x_i) Log_2\, p(x_i)$$

Where X = source and M = size of message

Shannon Entropy (or just Entropy)
Entropy is the measure of uncertainty of a random variable or the amount of information required to describe a variable. Suppose *x* is a discrete random variable, and it can take any value defined in the set, χ. Let's assume the set is finite in this scenario. The probability distribution for *x* will be $p(x) = \Pr\{\chi = x\}, x \in \chi$. With this in mind, entropy can be defined as
Shannon Entropisi (veya sadece Entropi)
Entropi, bir rastgele değişkenin belirsizliğinin ölçüsü veya bir değişkeni tanımlamak için gereken bilgi miktarıdır. x'in ayrık bir rastgele değişken olduğunu ve kümede tanımlanan herhangi bir değeri, χ, alabileceğini varsayalım. Bu senaryoda kümenin sonlu olduğunu varsayalım. x için olasılık dağılımı p(x) = Pr{χ = x}, x ∈ χ olacaktır. Bunu akılda tutarak, entropi şu şekilde tanımlanabilir

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x).$$

Conditional Entropy:

Intuitively, this is the average of the entropy of *Y* given *X* over all possible values of *X*. Considering the fact that (*X*, *Y*) ~ *p*(*x*, *y*), the conditional entropy can also be expressed in terms of expected value.

Koşullu Entropi:

Sezgisel olarak, bu, X verildiğinde Y'nin tüm olası X değerleri üzerindeki entropisinin ortalamasıdır. (X, Y) ~ p(x, y) gerçeğini göz önünde bulundurarak, koşullu entropi beklenen değer açısından da ifade edilebilir.

$$H(Y|X) = -E \log p(Y|X).$$

Conditional Entropy (Expected Value form):

Let's try an example to understand Conditional Entropy better. Consider a study where subjects were asked:

I) if they smoked, drank or didn't do either.

II) if they had any form of cancer

Koşullu Entropi (Beklenen Değer biçimi):

Koşullu Entropiyi daha iyi anlamak için bir örnek deneyelim. Deneklere şu soruların sorulduğu bir çalışmayı ele alalım:

I) sigara içip içmedikleri, içki içip içmedikleri veya ikisini de yapıp yapmadıkları.

II) herhangi bir kanser türüne sahip olup olmadıkları

Now, I will represent these questions' response as two different discrete variables belonging to a joint distribution.

Şimdi bu soruların yanıtlarını, birleşik bir dağılıma ait iki farklı ayrık değişken olarak göstereceğim.

Activity	Cancer
Smoking	Yes
Smoking	Yes
Smoking	Yes
Alcohol	No
Neither	Yes
Alcohol	Yes
Alcohol	Yes
Smoking	No
Alcohol	Yes
Neither	No

Marginal Probability of X:

X'in Marjinal Olasılığı:

Based on the probability table, we can plug in the value in the conditional entropy formula.

Olasılık tablosuna dayanarak değeri koşullu entropi formülüne koyabiliriz.

$$\begin{aligned} H(Y|X) &= p(\text{"Smoking"})H(Y|X=\text{"Smoking"}) + p(\text{"Alcohol"})H(Y|X=\text{"Alcohol"}) + p(\text{"Neither"})H(Y|X=\text{"Neither"}) \\ &= -\frac{4}{10}\left\{\frac{3}{10}\log\left(\frac{3}{10}\right)+\frac{1}{10}\log\left(\frac{1}{10}\right)\right\} - \frac{4}{10}\left\{\frac{3}{10}\log\left(\frac{3}{10}\right)+\frac{1}{10}\log\left(\frac{1}{10}\right)\right\} - \frac{2}{10}\left\{\frac{1}{10}\log\left(\frac{1}{10}\right)+\frac{1}{10}\log\left(\frac{1}{10}\right)\right\} \\ &= 0.4 * 0.857 + 0.4 * 0.857 + 0.2 * 0.664 \\ &= 0.8184\ bits \end{aligned}$$

Entropy, known as the uncertainty measure of a random variable, is the expected value of the information contained in all samples for a process. Information is a measure of information about the occurrence of a random event. Equally probable situations represent high uncertainty. According to Shannon, when a system changes, the change in entropy defines the information gained. Accordingly, the change in the maximum uncertainty situation will probably provide maximum information. Shannon used logarithm in base two because he represented information with bits.

Rassal bir değişkenin belirsizlik ölçütü olarak bilinen Entropi, bir süreç için tüm örnekler tarafından içerilen enformasyonun beklenen değeridir. ***Enformasyon ise rassal bir olayın gerçeklesmesine ilişkin bir bilgi ölçütüdür.*** Eşit olasılıklı durumlar yüksek belirsizliği temsil eder. Shannon'a göre bir sistemdeki durum değiştiğinde entropideki değişim kazanılan enformasyonu tanımlar. Buna göre maksimum belirsizlik durumundaki değişim muhtemelen maksimum enformasyonu sağlayacaktır. Shannon bilgiyi bitlerle temsil ettiği için logartimayı iki tabanında kullanımıştır.

$$I(x) = \log_2 \frac{1}{P(x)} = -\log_2 P(x)$$

According to Shannon, entropy is the expected value of the information carried by a transmitted message. The term called Shannon Entropy (H) is a value that depends on the probabilities of all xi states P(xi).

Shannon'a göre entropi, iletilen bir mesajın tasıdığı enformasyonun beklenen değeridir.

Shannon Entropisi (H) adıyla anılan terim, tüm x_i durumlarına ait $P(x_i)$ olasılıklarına bağlı bir değerdir.

$$H(X) = E(I(X)) = \sum_{1\le i\le n} P(x_i).I(x_i) = \sum_{i=1}^{n} P(x_i)\log_2 \frac{1}{P(x_i)} = -\sum_{i=1}^{n} P_i \log_2 P_i$$

$$Log_2(P) = \frac{10}{3} Log_{10}(P)$$

$$H(X) = -\frac{10}{3} \sum_{i=1}^{n} P_i \, Log_{10} P_i$$

Logarithmic expressions in base 10:

10 tabanında logaritmik ifadeler:

- Log1=0, Log 2 ≈ 0.3, Log 3 ≈ 0.477, Log 5 ≈ 0.7, Log 7 ≈ 0.845, Log10=1
- log(a*b)=loga + logb; $\log a^n = n * \log a$

Example: Let the event of tossing a coin represent the random process X. Since the probabilities of getting heads (1/2) and tails (1/2) are equal, when the coin toss event (X) occurs, which has an entropy of 1 for the X process, 1 bit of information will be gained.

Örnek: Bir paranın havaya atılması olayı, rassal X sürecini temsil etsin. Yazı (1/2) ve tura (1/2) gelme olasılıkları eşit olduğu için X sürecinin entropisi 1 olan para atma olayı (X) gerçeklestiğinde 1 bitlik bilgi kazanılacaktır.

$$H(X) = -\sum_{i=1}^{2} p_i \log_2 p_i \quad = -(0.5 \log_2 0.5 + 0.5 \log_2 0.5) = 1$$

Example: A decision was made to go on a picnic or not based on values such as weather, humidity, wind, water temperature. As a result of 4 events, the question "Did they go on a picnic?" has two answers. The probability of a yes answer is ¾, the probability of a no answer is 1/4.

Örnek: Hava, nem, rüzgar, su sıcaklığı gibi değerlere göre pikniğe gidip gitmeme kararı verilmiş 4 olay sonucunda pikniğe gidildi mi? sorusunun iki yanıtı var. Evet yanıtının olasılığı: ¾, hayır yanıtının olasılığı:1/4.

Olay No	Hava	Nem	Rüzgar	Su sıcaklığı	Pikniğe gidildi mi?
1	güneşli	normal	güçlü	ılık	Evet
2	güneşli	yüksek	güçlü	ılık	Evet
3	yağmurlu	yüksek	güçlü	ılık	Hayır
4	güneşli	yüksek	güçlü	soğuk	Evet

$$H(X) = -\sum_{i=1}^{2} p_i \log_2 p_i$$

Entropy(S)=H(x)=-(3/4)$\log_2$(3/4)-(1/4)$\log_2$(1/4)=0.811

Example: The machine learning algorithm needs to make a decision based on probability calculation. There are two cases. Calculate the probability of the first case occurring, P1=0.6, and the probability of not occurring, P2=0.4, and calculate its entropy.

Örnek: Makine öğrenmesi algoritmasının olasılık hesaplanması sonucunda karar vermesi gerekmektedir. İki durum söz konusudur. Birinci durumun olma olasılığı, P1=0.6, olmama olasığı, P2=0.4 ise entropisini hesaplayınız.

$$Log_2(0.6) = -0.743$$

$$Log_2(0.4) = -4/3$$

Entropi, H(x)=0.6*0.743+0.4*4/3=0.979

Data Table

On the left, you can see the data table where 10 subjects answered the questions. We have three different possibilities for the variable Activity (I will call it X). The second column represents whether the subject has/had cancer (variable Y). There are two possibilities here, i.e. Yes or No. Since we are not dealing with continuous variables yet, I have kept these variables discrete. Let's create a probability table that will make the scenario clearer.

Veri Tablosu

Solda, 10 denek tarafından soruların cevaplandığı veri tablosunu görebilirsiniz. Aktivite değişkeni için üç farklı olasılığımız var (ben buna X diyeceğim). İkinci sütun, deneklerin kanser olup olmadığını (değişken Y) temsil eder. Burada iki olasılık vardır, yani Evet veya Hayır. Henüz sürekli değişkenlerle uğraşmadığımız için bu değişkenleri ayrı tuttum. Senaryoyu daha net hale getirecek bir olasılık tablosu oluşturalım.

	Yes	No
Smoking	$^3/_{10}$	$^1/_{10}$
Alcohol	$^3/_{10}$	$^1/_{10}$
Neither	$^1/_{10}$	$^1/_{10}$

Probability Table for the above example

Next, I will calculate the value of the marginal probability *p*(*x*) for all the possible value of X.

Yukarıdaki örnek için Olasılık Tablosu

Daha sonra, X'in tüm olası değerleri için marjinal olasılık p(x) değerini hesaplayacağım.

$$p(X = \text{"Smoking"}) = \frac{4}{10}$$
$$p(X = \text{"}Alcohol\text{"}) = \frac{4}{10}$$
$$p(X = \text{"}Neither\text{"}) = \frac{2}{10}$$

Relative Entropy

Relative Entropy is somewhat different as it moves on from random variables to distributions. It is a measure of the distance between two distributions. A more instinctive way to put it would be: Relative entropy or KL-Divergence, denoted by $D(p||q)$, is a measure of the inefficiency of assuming that the distribution is q when the true distribution is p. It can be defined as:

Göreceli Entropi

Göreceli Entropi, rastgele değişkenlerden dağılımlara doğru hareket ettiği için biraz farklıdır. İki dağılım arasındaki mesafenin bir ölçüsüdür. Bunu daha içgüdüsel bir şekilde ifade etmek gerekirse: Göreceli entropi veya KL-Diverjans, D(p||q) ile gösterilir ve gerçek dağılım p olduğunda dağılımın q olduğunu varsaymanın verimsizliğinin bir ölçüsüdür. Şu şekilde tanımlanabilir:

$$D(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$$

Relative entropy is always non-negative and can be 0 only if $p = q$. Although a point to note here is that it is not a true distance since it is not symmetric in nature. But it is often considered as a “distance” between distributions.

Göreceli entropi her zaman negatif değildir ve yalnızca p = q ise 0 olabilir. Burada dikkat edilmesi gereken bir nokta, doğası gereği simetrik olmadığı için gerçek bir mesafe olmamasıdır. Ancak genellikle dağılımlar arasındaki bir "mesafe" olarak kabul edilir.

Lets take an example to solidify this concept! Let $X = \{0,1\}$ and consider two distributions p and q on X. Let $p(0) = 1 - r$, $p(1) = r$ and let $q(0) = 1 - s$, $q(1) = s$. Then,

Bu kavramı sağlamlaştırmak için bir örnek alalım! X = {0,1} olsun ve X üzerinde iki dağılım p ve q düşünelim. p(0) = 1 — r, p(1) = r ve q(0) = 1 — s, q(1) = s olsun. O zaman,

$$D(p||q) = (1-r)\log\left[\frac{1-r}{1-s}\right] + r\log\frac{r}{s}$$

Relative Entropy for p||q:

I would also like to demonstrate the non-symmetric property, so I will also calculate $D(q||p)$.

p||q için Göreceli Entropi:

Ayrıca simetrik olmayan özelliği de göstermek istiyorum, bu yüzden D(q||p)'yi de hesaplayacağım.

$$D(q||p) = (1-s)\log\left[\frac{1-s}{1-r}\right] + s\log\frac{s}{r}$$

If r = s, $D(p||q) = D(q||p) = 0$. But I will take some different values, for instance, $r = 1/2$ and $s = 1/4$.

Eğer r = s ise, D(p||q) = D(q||p) = 0. Ancak bazı farklı değerler alacağım, örneğin, r = 1/2 ve s = 1/4.

$$D(p||q) = \left(1-\frac{1}{2}\right)\log\left[\frac{1-\frac{1}{2}}{1-\frac{1}{4}}\right] + \frac{1}{2}\log\frac{\frac{1}{2}}{\frac{1}{4}} = 0.2075 \text{ bit}$$

$$D(q||p) = \left(1-\frac{1}{4}\right)\log\left[\frac{1-\frac{1}{4}}{1-\frac{1}{2}}\right] + \frac{1}{4}\log\frac{\frac{1}{4}}{\frac{1}{2}} = 0.1887 \text{ bit}$$

As you can see, $D(p||q) \neq D(q||p)$.

Gördüğünüz gibi, D(p||q) ≠ D(q||p).

Now, since we have discussed the different types of entropies, we can move onto Mutual Information.

Şimdi, entropilerin farklı türlerini tartıştığımıza göre, Karşılıklı Bilgiye geçebiliriz.

Example: A sample space of 5 messages with probabilities are given as: P(s) = {0.25, 0.25, 0.25, 0.125, 0.125}. Find Entropy of the source.

Örnek: Olasılıkları olan 5 mesajlık bir örnek uzayı şu şekilde verilmiştir: P(s) = {0.25, 0.25, 0.25, 0.125, 0.125}. Kaynağın Entropisini bulun.

$$H(x) = \sum_{i=1}^{5} p(x_i) Log_2 \frac{1}{p(x_i)} = P(x_1) Log_2 \frac{1}{P(x_1)} + P(x_2) Log_2 \frac{1}{P(x_2)} + P(x_3) Log_2 \frac{1}{P(x_3)} + P(x_4) Log_2 \frac{1}{P(x_4)} + P(x_5) Log_2 \frac{1}{P(x_5)}$$

$$= 0.25\log_2 \frac{1}{0.25} + 0.25\log_2 \frac{1}{0.25} + 0.25\log_2 \frac{1}{0.25} + 0.125\log_2 \frac{1}{0.125} + 0.125\log_2 \frac{1}{0.125}$$

$$= 3 \text{ x } 0.25\log_2 \frac{1}{0.25} + 2 \text{ x } 0.125\log_2 \frac{1}{0.125}$$

$$= 3 \text{ x } 0.25 \text{ x } \log_2 4 + 2 \text{ x } 0.125 \text{ x } \log_2 8$$

$$= 3 \text{ x } 0.25 \text{ x } 2 + 2 \text{ x } 0.125 \text{ x } 4$$

$$= 2.25 \text{ bits/ symbol}$$

Example:

An unfair dice with four faces and p(1) = 1/2, p(2) = 1/4, p(3) = p(4) = 1/8. Find entropy H (answer= 7/4).

Örnek:

Dört yüzü olan ve p(1) = 1/2, p(2) = 1/4, p(3) = p(4) = 1/8 olan bir haksız zar. Entropi H'yi bulun (cevap= 7/4).

Example:

A Discrete Memory-less Source (DMS) X has four symbols x_1, x_2, x_3, x_4 with probabilities $p(x_1) = 0.4$, $p(x_2) = 0.3$, $p(x_3) = 0.2$ and $p(x_4) = 0.1$. Calculate H(X)

$$H(X) = \sum_{1}^{4} P(x_i) \log_2 \frac{1}{P(x_i)} = P(x_1) \log_2 \frac{1}{P(x_1)} + P(x_2) \log_2 \frac{1}{P(x_2)} + P(x_3) \log_2 \frac{1}{P(x_3)} + P(x_4) \log_2 \frac{1}{P(x_4)}$$

$$= 0.4 \log_2 \frac{1}{0.4} + 0.3 \log_2 \frac{1}{0.3} + 0.2 \log_2 \frac{1}{0.2} + 0.1 \log_2 \frac{1}{0.1}$$

$$= 0.4 \log_2 \frac{1}{0.4} + 0.3 \log_2 \frac{1}{0.3} + 0.2 \log_2 \frac{1}{0.2} + 0.1 \log_2 \frac{1}{0.1} = 1.85 \text{ bits/symbol}$$

9.1.7. Information Rate

When an analog signal is converted to a digital signal, samples are taken at certain intervals. Baud rate is the number of samples taken from an analog signal per second. It is the number of symbols sent per second. Because each sample taken is a symbol.

Bir analog sinyal, sayısal sinyale dönüştürülürken belli aralıklarla örnek alınır. Baud rate, bir saniyede bir analog sinyalden alınan örnek sayısıdır. Bir saniyede gönderilen sembol sayısıdır. Çünkü alınan her bir örnek bir semboldür.

R=fs* H(x)
Where
R: Information rate (bps=bits per second)
fs: baud rate = symolic/sec
H(x): entropy or average information (bits/symbol)
R=fs* H(x)
Burada
R: Bilgi hızı (bps=saniye başına bit)
fs: baud hızı = sembolik/sn
H(x): entropi veya ortalama bilgi (bit/sembol)

Example:
If Baud rate=1000, it is said that 1000 samples are taken in one second. If each sample is represented by 4 bits, calculate the Bit rate ratio.
Baud rate = 1000 bauds per second (baud/sec)
Information Rate (Bit rate) = Baud rate (Number of samples taken in one second) x Number of bits representing a sample taken from an analog signal = 1000 x 4= 4000 bps
Örnek:
Baud rate=1000 ise bir saniyede 1000 adet örnek alınmış denmektedir. Her bir örnek 4 bit ile temsil edilirse, Bit rate rate oranını hesaplayınız.
Baud rate = 1000 bauds per second (baud/sec)
Information Rate (Bit rate) = Baud rate (Bir saniyede alınan örnek sayısı) x Analog sinyalden alınan bir örneği temsil eden bit sayısı = 1000 x 4= 4000 bps

Example:
The bit rate of a signal is 4000. If each signal unit carries 4bits, what is the baud rate?
Baud rate = 4000/4 =1000 bauds/sec
Örnek:
Bir sinyalin bit hızı 4000'dir. Her sinyal birimi 4 bit taşıyorsa, baud hızı nedir?
Baud hızı = 4000/4 =1000 baud/sn

Example:
An analog signal is band limited to 1000Hz and sampled Nyquist rate. The samples are quatized into 4 levels. Each level represents 1 symbol. Probabilities of the symbols are: p(x1)=p(x4)=1/8 and p(x2)=p(x3)=3/8. Obtain information rate of the source.
Örnek:

Analog bir sinyal 1000Hz ile bant sınırlı ve örneklenmiş Nyquist oranına sahiptir. Örnekler 4 seviyeye bölünür. Her seviye 1 sembolü temsil eder. Sembollerin olasılıkları: p(x1)=p(x4)=1/8 ve p(x2)=p(x3)=3/8'dir. Kaynağın bilgi oranını elde edin.

$$Entropy, H(X) = \sum_{1}^{4} p(x1)Log_2\left(\frac{1}{p(x1)}\right) + p(x2)Log_2\left(\frac{1}{p(x2)}\right) + p(x3)Log_2\left(\frac{1}{p(x3)}\right) + p(x4)Log_2\left(\frac{1}{p(x4)}\right)$$

$$entropy, H(x) = \frac{1}{8}Log_2 8 + \frac{3}{8}Log_2\frac{8}{3} + \frac{1}{8}Log_2 8 + \frac{3}{8}Log_2\frac{8}{3}$$

Entrop, H(x)=1.5 bits/symbol
Fm=1000Hz (given)
Nyquist rate r=fs=2*fm=2*1000=2000 symbols/sec. This is nothing but number of symbols by source =fs.
R=fs * H(x)
Entropi, H(x)=1,5 bit/sembol
Fm=1000Hz (verilen)
Nyquist oranı r=fs=2*fm=2*1000=2000 sembol/sn. Bu, kaynak =fs'ye göre sembol sayısından başka bir şey değildir.
R=fs * H(x)
Where,
R: Information rate (bps=bits per second)
fs: baud rate = symolic/sec
H(x): entropy or average information (bits/symbol)
We know that R=fs * H(x)
R=2000*1,5=3000bps
Burada,
R: Bilgi hızı (bps=saniye başına bit)
fs: baud hızı = sembolik/sn
H(x): entropi veya ortalama bilgi (bit/sembol)
R=fs * H(x) olduğunu biliyoruz
R=2000*1,5=3000bps

9.1.8. Karşılıklı Bilgi (Mutual information)

Mutual information about two messages measures the amount of information that one conveys about the other. It is defined as:
İki mesaj hakkındaki karşılıklı bilgi, birinin diğeri hakkında ilettiği bilgi miktarını ölçer. Şöyle tanımlanır:

$$I(X;Y) = \sum_{x,y} p(x,y) Log_2 \frac{p(x,y)}{p(x)p(y)}$$

X = information source
Y = another information source
x= message generated by source X
y = message generated by source Y
p(x) = probability of occurrence of x
p(y) = probability of occurrence of y
X = bilgi kaynağı
Y = başka bir bilgi kaynağı
x= X kaynağı tarafından üretilen mesaj
y = Y kaynağı tarafından üretilen mesaj
p(x) = x'in oluşma olasılığı
p(y) = y'nin oluşma olasılığı

Mutual information means message x says about y and vice versa.
If X and Y are independent sources, then joint probability p(x, y) = p(x) . p(y)
Now mutual information between x and y is given as:
Karşılıklı bilgi, x'in y hakkında söylediği mesaj ve bunun tersi anlamına gelir.
Eğer X ve Y bağımsız kaynaklarsa, o zaman ortak olasılık p(x, y) = p(x) . p(y)
Şimdi x ve y arasındaki karşılıklı bilgi şu şekilde verilir:

$$I(X;Y) = \sum_{x,y} p(x,y) Log_2 \frac{p(x,y)}{p(x)p(y)} = \sum_{x,y} p(x,y) Log_2 \frac{p(x)p(y)}{p(x)p(y)}$$
$$= \sum_{x,y} p(x,y) Log_2 1 = 0$$

Digital communication system (Shannon model of communication):
Dijital iletişim sistemi (Shannon iletişim modeli):

- The basic goal of a communication system is to transmit some information from source to the destination.
- Message = Information

- Information consists of letters, digits, symbols, sequence of letters, digits, symbols etc.
- Information theory gives an idea about what can be achieved or what cannot be achieved -in a communication system.
- Bir iletişim sisteminin temel amacı, kaynaktan hedefe bazı bilgileri iletmektir.
- Mesaj = Bilgi
- Bilgi, harflerden, rakamlardan, sembollerden, harf dizilerinden, rakamlardan, sembollerden vb. oluşur.
- Bilgi teorisi, bir iletişim sisteminde neyin başarılabileceği veya neyin başarılamayacağı hakkında bir fikir verir.

Shannon gave ideas on:
• Signal processing operations such as compressing data
• Storing and communicating data on a noisy channel (channel with errors)
Shannon şu konularda fikir verdi:
• Verileri sıkıştırma gibi sinyal işleme işlemleri
• Gürültülü bir kanalda (hatalı kanal) veri depolama ve iletişimi

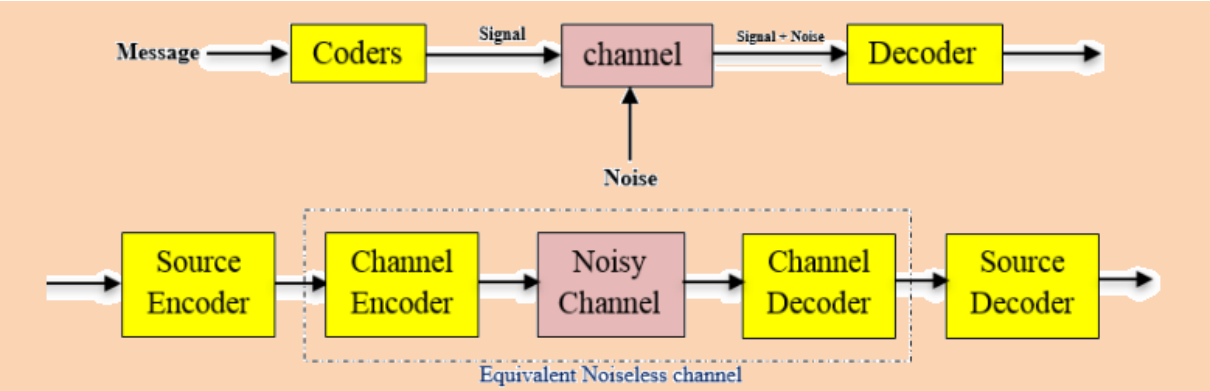


Fig: Digital communication system (Shannon model of communication)

A practical source in a communication system is a device that produces messages. It can be either analog or digital (discrete). So, there is a source and there is a destination. Messages are transferred from one point to another.
Bir iletişim sistemindeki pratik bir kaynak, mesajlar üreten bir cihazdır. Analog veya dijital (ayrık) olabilir. Yani bir kaynak ve bir hedef vardır. Mesajlar bir noktadan diğerine aktarılır.

We deal mainly with discrete sources since analog sources can be transformed to discrete sources through the use of sampling and quantization techniques.
Analog kaynaklar örnekleme ve kantizasyon tekniklerinin kullanımıyla ayrık kaynaklara dönüştürülebildiğinden, esas olarak ayrık kaynaklarla ilgileniyoruz.

Mutual Information is a measure of the amount of information that one random variable contains about another random variable. **Alternatively, it can be defined as the reduction in uncertainty of one variable due to the knowledge of the other.** The technical definition for it would be as follows:

Karşılıklı Bilgi, bir rastgele değişkenin başka bir rastgele değişken hakkında içerdiği bilgi miktarının bir ölçüsüdür. Alternatif olarak, bir değişkenin diğerinin bilgisinden dolayı belirsizliğinin azalması olarak tanımlanabilir. Bunun için teknik tanım şu şekilde olacaktır:

Consider two random variables X and Y with a joint probability mass function $p(x, y)$ and marginal probability mass functions $p(x)$ and $p(y)$. The mutual information $I(X; Y)$ is the relative entropy between the joint distribution and the product distribution $p(x)p(y)$.

Ortak olasılık kütle fonksiyonu p(x, y) ve marjinal olasılık kütle fonksiyonları p(x) ve p(y) olan iki rastgele değişken X ve Y'yi ele alalım. Karşılıklı bilgi I (X; Y), ortak dağılım ile ürün dağılımı p(x)p(y) arasındaki bağıl entropidir.

$$I(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$
$$= D(p(x, y)||p(x)p(y))$$

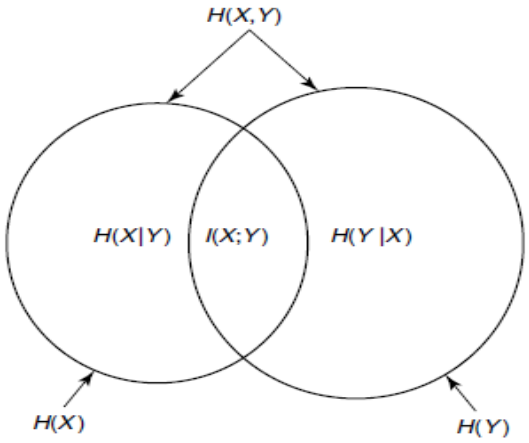
Mutual Information can also be expressed in terms of Entropy.

Karşılıklı Bilgi ayrıca Entropi açısından da ifade edilebilir.

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

From the above equation, we see that mutual information is the reduction in the uncertainty of X due to the knowledge of Y. There is a Venn diagram perfectly describing the relationship.

Yukarıdaki denklemden, karşılıklı bilginin, Y'nin bilinmesi nedeniyle X'in belirsizliğindeki azalma olduğunu görüyoruz. İlişkiyi mükemmel bir şekilde tanımlayan bir Venn diyagramı var.



Relationship between Mutual Information and Entropy:
Let's take an example to understand it better. I can use the example of relating Smoking, Drinking, Neither with having cancer that I used while explaining entropy. We had seen that the $H(Y|X)$ = 0.8184 bits. To calculate Mutual Information, I require one more term $H(Y)$. $H(Y)$, in this case, will be:
Karşılıklı Bilgi ve Entropi Arasındaki İlişki:
Daha iyi anlamak için bir örnek alalım. Entropiyi açıklarken kullandığım Sigara İçmek, İçki İçmek, Hiçbiri'ni kansere sahip olmakla ilişkilendirme örneğini kullanabilirim. H(Y|X) = 0,8184 bit olduğunu görmüştük. Karşılıklı Bilgiyi hesaplamak için bir terime daha ihtiyacım var H(Y). Bu durumda H(Y), şöyle olacaktır:

$$\begin{aligned} H(Y) &= -\sum_{y \in Y} p(y) \log p(y) \\ &= \frac{7}{10} \log \frac{7}{10} + \frac{3}{10} \log \frac{3}{10} \\ &= 0.876 \end{aligned}$$

Therefore, Mutual Information is defined by:
Bu nedenle, Karşılıklı Bilgi şu şekilde tanımlanır:

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= 0.876 - 0.8184 \\ &= 0.0576 \end{aligned}$$

Calculation of Mutual Information:
Alternatively, I can also use $H(X)$ and $H(X|Y)$ to calculate Mutual Information, and it will yield the same result. We can see how knowing X means so little for the uncertainty of variable Y. Let me change the passage of this example and lead you to how it all makes sense in Machine Learning. Suppose X is a predictor variable and Y is the predicted variable. Mutual Information between them can be a great precursor to check how useful the feature will be for predictions. Let us discuss the implications of Information Theory in Machine Learning.
Karşılıklı Bilginin Hesaplanması:
Alternatif olarak, Karşılıklı Bilgiyi hesaplamak için H(X) ve H(X|Y)'yi de kullanabilirim ve aynı sonucu verecektir. X'i bilmenin Y değişkeninin belirsizliği için ne kadar az şey ifade ettiğini

görebiliriz. Bu örneğin geçişini değiştireyim ve sizi Makine Öğrenmesinde her şeyin nasıl mantıklı olduğuna götüreyim. X'in bir tahmin edici değişken ve Y'nin tahmin edilen değişken olduğunu varsayalım. Aralarındaki Karşılıklı Bilgi, özelliğin tahminler için ne kadar yararlı olacağını kontrol etmek için harika bir öncü olabilir. Makine Öğrenmesinde Bilgi Teorisinin etkilerini tartışalım.

Applications of Information Theory in Machine Learning:

There are quite a few applications around, but I will stick with a few popular ones.

Makine Öğrenmesinde Bilgi Teorisinin Uygulamaları:

Etrafta oldukça fazla uygulama var, ancak ben birkaç popüler olana sadık kalacağım.

Coding theory:

Coding is the most important application of information theory. DMS output is converted into binary codes. Device that performs this conversion is called the source encoder.

- Source encoder takes care of data compression
- Channel encoder do reliable data transmission
- Decoding is exactly the reverse process of encoding
- Note that communication channel is at the heart of the communication problem. Channel noise corrupts the transmitted signal, causing unavoidable decoding errors at the receiver.

Kodlama teorisi:

Kodlama, bilgi teorisinin en önemli uygulamasıdır. DMS çıktısı ikili kodlara dönüştürülür. Bu dönüşümü gerçekleştiren cihaza kaynak kodlayıcı denir.

- Kaynak kodlayıcı veri sıkıştırma işlemini gerçekleştirir
- Kanal kodlayıcı güvenilir veri iletimi yapar
- Kod çözme, kodlamanın tam tersi bir işlemdir
- İletişim kanalının iletişim sorununun merkezinde olduğunu unutmayın. Kanal gürültüsü iletilen sinyali bozar ve alıcıda kaçınılmaz kod çözme hatalarına neden olur.

Source coding = Entropy Coding.

- Job: Data compression. Compression algorithms are of great importance when processing and transmitting audio, images and video
- Note that during compression no significant information is lost
- Entropy defines the minimum amount of necessary information
- Source coding is done at transmitter

Various source coding techniques are Huffman coding, Shannon-Fano, Lempel Ziv coding, PCM, DPCM, DM and adaptive DM (ADPCM).

Kaynak kodlama = Entropi Kodlaması.

• İş: Veri sıkıştırma. Sıkıştırma algoritmaları ses, görüntü ve video işlerken ve iletirken büyük önem taşır
• Sıkıştırma sırasında önemli bir bilginin kaybolmadığını unutmayın
• Entropi, gerekli olan minimum bilgi miktarını tanımlar
• Kaynak kodlama vericide yapılır
Çeşitli kaynak kodlama teknikleri Huffman kodlaması, Shannon-Fano, Lempel Ziv kodlaması, PCM, DPCM, DM ve adaptif DM'dir (ADPCM).

Channel coding = Error-Control Coding.
Encoder adds additional bits to actual information in order to detect and correct transmission errors.
Various channel coding techniques are: Hamming codes, cyclic codes, BCH codes, block coding, convolutional coding, turbo coding etc.
Assume that 1-bit of information is transmitted from source to destination. If bit – 0 is transmitted, bit – 0 must be received.
Kanal kodlama = Hata Kontrol Kodlaması.
Kodlayıcı, iletim hatalarını tespit etmek ve düzeltmek için gerçek bilgilere ek bitler ekler.
Çeşitli kanal kodlama teknikleri şunlardır: Hamming kodları, döngüsel kodlar, BCH kodları, blok kodlama, evrişimli kodlama, turbo kodlama vb.
Kaynaktan hedefe 1 bitlik bilgi iletildiğini varsayalım. Bit – 0 iletilirse, bit – 0 alınmalıdır.

Transmitter	Receiver	Remark
0	0	Good
0	1	Error
1	0	Error
1	1	Good

Error correcting code adds just the right kind of redundancy as possible (i.e., error correction) needed to transmit the data efficiently across noisy channel.
Hata düzeltme kodu, gürültülü kanallar üzerinden veriyi verimli bir şekilde iletmek için gereken doğru türde yedekliliği (yani hata düzeltmeyi) ekler.
Source coding:
Also known as entropy coding. Here the information generated by the source is compressed. Note that during compression no significant information is lost. Entropy defines the minimum amount of necessary information. Source coding is done at transmitter.
Source encoder transforms information from source into different information bits, while implementing data compression. Various source coding techniques are Huffman coding, Lempel Ziv coding, PCM, DPCM, DM and adaptive DM (ADPCM).
Kaynak kodlama:

Entropi kodlaması olarak da bilinir. Burada kaynak tarafından üretilen bilgi sıkıştırılır. Sıkıştırma sırasında önemli bir bilginin kaybolmadığını unutmayın. Entropi, gerekli olan minimum bilgi miktarını tanımlar. Kaynak kodlaması vericide yapılır.
Kaynak kodlayıcı, veri sıkıştırmayı uygularken kaynaktan gelen bilgiyi farklı bilgi bitlerine dönüştürür. Çeşitli kaynak kodlama teknikleri Huffman kodlaması, Lempel Ziv kodlaması, PCM, DPCM, DM ve adaptif DM'dir (ADPCM).

Channel coding:

Also known as error-control coding. Encoder adds additional bits to actual information in order to detect and correct transmission errors. Channel coding is done at receiver. Various channel coding techniques are: block coding, convolutional coding, turbo coding etc.

Kanal kodlaması:

Hata kontrol kodlaması olarak da bilinir. Kodlayıcı, iletim hatalarını tespit etmek ve düzeltmek için gerçek bilgilere ek bitler ekler. Kanal kodlaması alıcıda yapılır. Çeşitli kanal kodlama teknikleri şunlardır: blok kodlama, evrişimli kodlama, turbo kodlama vb.

Example:

Örnek:

In a binary PCM if 0 occurs with probability ¼ and 1 occurs with probability ¾, then calculate the amount of information carried by each bit.

$$I(bit\ 0) = \log_2 \frac{1}{P(x_1)} = \log_2 \frac{1}{\frac{1}{4}} = \log_2 4 = \log_2 2^2 \quad = 2\ \log_2 2 = 2 * 1 = 2\ bits$$

$$I(bit\ 1) = \log_2 \frac{1}{P(x_2)} = \log_2 \frac{1}{\frac{3}{4}} = \log_2 \frac{4}{3} = \log_2 1.33 \quad = \frac{\log_{10} 1.33}{\log_{10} 2} = \frac{0.125}{0.301} = 0.415\ bits$$

- Bit 0 has probability ¼ and it has 2 bits of information
- Bit 1 has probability ¾ and it has 0.415 bits of information

Example:

Örnek:

A Discrete Memory-less Source (DMS) X has four symbols x_1, x_2, x_3, x_4 with probabilities $p(x_1) = 0.4$, $p(x_2) = 0.3$, $p(x_3) = 0.2$ and $p(x_4) = 0.1$. Calculate H(X)

$$H(X) = \sum_{1}^{4} P(x_i) \log_2 \frac{1}{P(x_i)} = P(x_1) \log_2 \frac{1}{P(x_1)} + P(x_2) \log_2 \frac{1}{P(x_2)} + P(x_3) \log_2 \frac{1}{P(x_3)} + P(x_4) \log_2 \frac{1}{P(x_4)}$$

$$= 0.4 \log_2 \frac{1}{0.4} + 0.3 \log_2 \frac{1}{0.3} + 0.2 \log_2 \frac{1}{0.2} + 0.1 \log_2 \frac{1}{0.1}$$

$$= 0.4 \log_2 \frac{1}{0.4} + 0.3 \log_2 \frac{1}{0.3} + 0.2 \log_2 \frac{1}{0.2} + 0.1 \log_2 \frac{1}{0.1} = 1.85 \ bits/symbol$$

A source produces one of the 4 possible symbols during each interval having probabilities: $p_1 = \frac{1}{2}, p_2 = \frac{1}{4}, p_3 = p_4 = \frac{1}{8}$. Obtain information content of each of these symbols.

Example:

Örnek:

A source emits independent sequences of symbols from a source alphabet containing five symbols with probabilities 0.4, 0.2, 0.2, 0.1 and 0.1. Compute the entropy of the source.

Solution: Source alphabet = (s1, s2, s3, s4, s5)
Probs. of symbols = p1, p2, p3, p4, p5 = 0.4, 0.2, 0.2, 0.1, 0.1

(i) Entropy of the source = H = $-\sum$pi log pi bits / symbol

Substituting we get,
H = - [$p_1 \log p_1 + p_2 \log p_2 + p_3 \log p_3 + p_4 \log p_4 + p_5 \log p_5$]
= - [0.4 log 0.4 + 0.2 log 0.2 + 0.2 log 0.2 + 0.1 log 0.1 + 0.1 log 0.1]
H = 2.12 bits/symbol

Example:

Örnek:

An analog signal is band limited to 1000 Hz and sampled at Nyquist rate. The samples are quantized into 4 levels. Each level represents 1 symbol. Thus, there are 4 levels (symbols) are:
$p(x_1) = p(x_4) = 1/8$ and $p(x_2) = p(x_3) = 3/8$. Obtain information rate of the source.

$$Entropy, H(X) = \sum_{1}^{4} P(x_i) \log_2 \frac{1}{P(x_i)} = P(x_1) \log_2 \frac{1}{P(x_1)} + P(x_2) \log_2 \frac{1}{P(x_2)} + P(x_3) \log_2 \frac{1}{P(x_3)} + P(x_4) \log_2 \frac{1}{P(x_4)}$$

$$= \frac{1}{8} \log_2 8 + \frac{3}{8} \log_2 \frac{8}{3} + \frac{1}{8} \log_2 8 + \frac{3}{8} \log_2 \frac{8}{3} = 1.5 \ bits/symbol$$

Information rate (R)

R = r H(X)
Where R = information rate (bps) = rate at which symbols are generated.
r = Number of symbols generated by source (symbols/sec)
H(X) = entropy or average information (bits/symbol)

f_m = 1000 Hz (given)
Nyquist rate $f_s = 2f_m$ = 2000 Hz = 2000 symbols/sec. This is nothing but number of symbols generated by source, which is **r**

We know that **R = r H(X)**
$= 2000 \frac{symbols}{sec} \times 1.5 \frac{bits}{symbol} = 3000 \frac{bits}{sec} = 3000 \ bps$

9.1.9. Information Gain

Bilgi kazancı, entropideki bu değişimin bir ölçüsüdür. Entropy: a common way to measure impurity. Entropy comes from information theory. The higher the entropy the more the information content.
Entropi: safsızlığı ölçmenin yaygın bir yolu. Entropi bilgi teorisinden gelir. Entropi ne kadar yüksekse bilgi içeriği de o kadar fazladır.

Information Gain (IG) is a measure used in decision trees to quantify the effectiveness of a feature in splitting the dataset into classes. It calculates the reduction in entropy (uncertainty) of the target variable (class labels) when a particular feature is known. In simpler terms, Information Gain helps us understand how much a particular feature contributes to making accurate predictions in a decision tree. Features with higher Information Gain are considered more informative and are preferred for splitting the dataset, as they lead to nodes with more homogenous classes.
Bilgi Kazancı (IG), veri kümesini sınıflara ayırmada bir özelliğin etkinliğini ölçmek için karar ağaçlarında kullanılan bir ölçüdür. Belirli bir özellik bilindiğinde hedef değişkenin (sınıf etiketleri) entropisindeki (belirsizlik) azalmayı hesaplar. Daha basit bir ifadeyle, Bilgi Kazancı, belirli bir özelliğin bir karar ağacında doğru tahminler yapmaya ne kadar katkıda bulunduğunu anlamamıza yardımcı olur. Daha yüksek Bilgi Kazancına sahip özellikler daha bilgilendirici olarak kabul edilir ve daha homojen sınıflara sahip düğümlere yol açtıkları için veri kümesini bölmek için tercih edilir.

IG(X,A)=H(X)−H(X|A)

Where,

- IG(X, A) is the Information Gain of feature A concerning dataset X.
- H(X) is the entropy of dataset X.
- H(X|A) is the conditional entropy of dataset X given feature A.

Burada,
• IG(X, A), veri kümesi X ile ilgili özellik A'nın Bilgi Kazancıdır.
• H(X), veri kümesi X'in entropisidir.
• H(X|A), özellik A verildiğinde veri kümesi X'in koşullu entropisidir.

Entropy H(X):

$$H(X) = \sum_{i=1}^{n} P(x_i) Log_2(\frac{1}{P(x_i)})$$

- n represents the number of different outcomes in the dataset.
- P(xi) is the probability of outcome xi occurring.

- n, veri kümesindeki farklı sonuçların sayısını temsil eder.
- P(xi), xi sonucunun gerçekleşme olasılığıdır.

Conditional Entropy H(X|A):

$$\mathbf{H(X|A)} = \sum_{j=1}^{m} \boldsymbol{p(a_j)}\mathrm{H(X|aj)}$$

- P(aj) is the probability of feature value aj in feature A,and
- H(X|aj) is the entropy of dataset X given feature A has value aj.
- P(aj), özellik A'da özellik değeri aj'nin olasılığıdır ve
- H(X|aj), özellik A'nın değeri aj olduğunda veri kümesi X'in entropisidir.

Example:

The easiest way to understand this is with an example. Consider a dataset with 1 blue, 2 greens, and 3 reds: . Then

Bunu anlamanın en kolay yolu bir örnekle olur. 1 mavi, 2 yeşil ve 3 kırmızıdan oluşan bir veri kümesini ele alalım: . Sonra

$H(X)= -(p_b*Log_2p_b + p_g*Log_2p_g + p_r*Log_2p_r)$

We know pb=1/6 because 1/6 of the dataset is blue. Similarly, pg=2/6 (greens) and pr=3/6 (reds).

pb=1/6 olduğunu biliyoruz çünkü veri setinin 1/6'sı mavi. Benzer şekilde, pg=2/6 (yeşiller) ve pr=3/6 (kırmızılar).

Thus,

Böylece,

H(X)=−(1/6log2(1/6)+2/6log2(2/6)+3/6log2(3/6))=1.46*E*))=1.46

What about a dataset of all one color? Consider 3 blues as an example: The entropy would be

Peki ya hepsi tek renkten oluşan bir veri kümesi hakkında ne düşünüyorsunuz? Örnek olarak 3 maviyi düşünün: Entropi şu şekilde olurdu:

$H(X)=-(1\log_2 1)=0$

Example: Information Gain

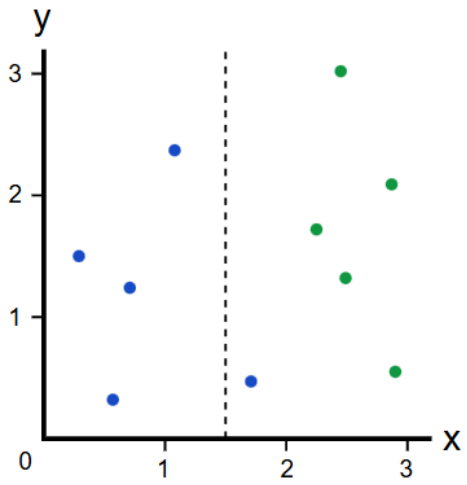
Örnek: Bilgi Kazancı

How can we quantify the quality of a split?

Bir bölünmenin kalitesini nasıl ölçebiliriz?

Let's consider this split:

Bu ayrımı düşünelim:



Before the split, we had 5 blues and 5 greens, so the entropy was

Bölünmeden önce 5 mavi ve 5 yeşilimiz vardı, bu yüzden entropi

$H(X_{before}) = -(0.5\log_2(0.5) + 0.5\log_2(0.5)) = 1$

After the split, we have two branches.

Left Branch has 4 blues, so Eleft=0 because it's a dataset of all one color.

Bölünmeden sonra iki dalımız var.

Sol Dalda 4 mavi var, bu yüzden Eleft=0 çünkü bu tek bir renkten oluşan bir veri kümesi.

Since there are 4 blue ones in the left section, probability = 1. Log2(1) = 0. The entropy of the left side is 0.

Soldaki bölmede 4 adet mavi olduğundan olasılık=1 çıkar. Log2(1)=0. Sol tarafın entropisi 0'dır.

Right Branch has 1 blue and 5 greens, so

Sağ Dal'da 1 mavi ve 5 yeşil var, bu yüzden

$H(X_{right}) = -(1/6\log_2(1/6) + 5/6\log_2(5/6)) = 0.65$

Now that we have the entropies for both branches, we can determine the quality of the split by **weighting the entropy of each branch by how many elements it has**. Since Left Branch has 4 elements and Right Branch has 6, we weight them by 0.4 and 0.6, respectively:

Artık her iki dalın entropilerine sahip olduğumuza göre, her dalın entropisini sahip olduğu eleman sayısına göre ağırlıklandırarak bölünmenin kalitesini belirleyebiliriz. Sol Dal'ın 4 elemanı ve Sağ Dal'ın 6 elemanı olduğundan, bunları sırasıyla 0,4 ve 0,6 ile ağırlıklandırıyoruz:

H(Xsplit)=0.4∗0 + 0.6∗0.65=0.39

We started with Ebefore=1 entropy before the split and now are down to 0.390!

Bölünmeden önce Ebefore=1 entropi ile başlamıştık ve şu an 0,390'a düştük!

Information Gain = how much Entropy we removed, so

Gain=H(Xbefore) - H(Xright) =1−0.39=0.61

Bilgi Kazancı = ne kadar Entropiyi kaldırdığımız, bu nedenle

Kazanç=H(Xönce) - H(Xsağ) =1−0.39=0.61

This makes sense: **higher Information Gain = more Entropy removed**, which is what we want. In the perfect case, each branch would contain only one color after the split, which would be zero entropy!

Bu mantıklıdır: daha yüksek Bilgi Kazanımı = daha fazla Entropi kaldırıldı, ki bu da istediğimiz şeydir. Mükemmel durumda, her dal bölünmeden sonra yalnızca bir renk içerecektir, bu da sıfır entropi olacaktır!

Example:

Örnek:

For set X = {a,a,a,b,b,b,b,b}

Total examples: 8

b: 5 examples

a: 3 examples

X = {a,a,a,b,b,b,b,b} kümesi için

Toplam örnek: 8

b: 5 örnekleri

a: 3'ün örnekleri

$$Entropy H(X) = -\left[\left(\frac{3}{8}\right) log_2 \frac{3}{8} + \left(\frac{5}{8}\right) log_2 \frac{5}{8}\right]$$

```
= -[0.375 * (-1.415) + 0.625 * (-0.678)]
=-(-0.53-0.424)
= 0.954
```

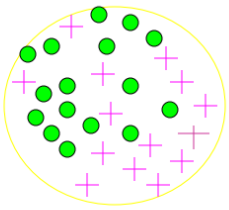
$$Log_2(a) = b \; ise \; 2^b = a \; dır.$$
$$Log_{10}(2^b) = Log_{10}(a)$$

$$b = \frac{1}{0.3} Log_{10}(a)$$

Example:

Information Gain = entropy(parent) – [average entropy(children)]

Entropy = $\sum_i -p_i \log_2 p_i$



p_i is the probability of class i

Compute it as the proportion of class i in the set.

16/30 are green circles; 14/30 are pink crosses

$\log_2(16/30) = -.9$; $\log_2(14/30) = -1.1$

Entropy = -(16/30)(-.9) –(14/30)(-1.1) = .99

In the first stage, the entropy of the main cluster is calculated. Green circle: 16, Red cross: 14 pieces.

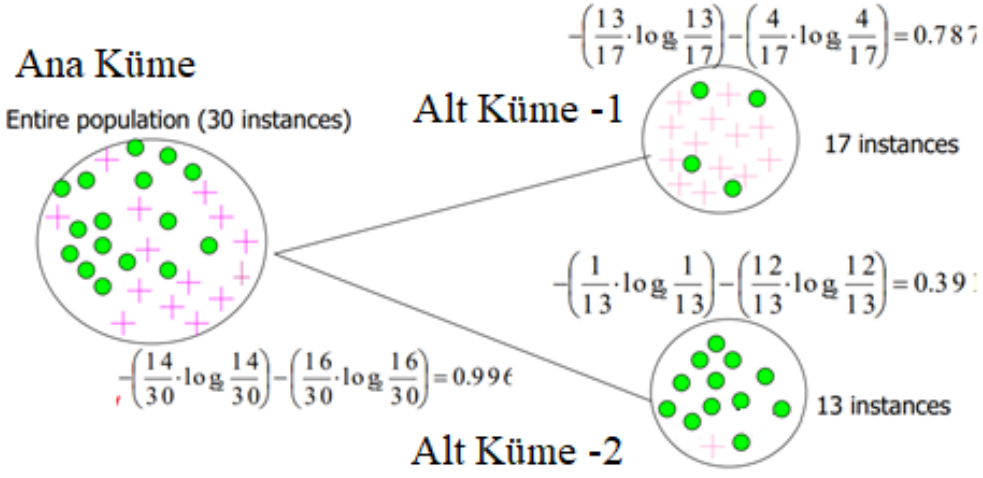
1) İlk aşamada ana kümenin entropisi hesaplanır. Yeşi daire: 16, Kırmızı artı:14 adet.

H(x)=(16/30)*Log2 (30/16) + (14/30)* Log2 (30/14)=0.9

H(x)=(16/30)*3* Log10 (30/16) + (14/30)*3*Log10 (30/14)=0.9

2) Then, since this main cluster is divided into two separate clusters, the entropy of each cluster is calculated (E1, E2).

2) Ardından bu ana küme iki ayrı kümeye ayrıldığından her bir kümenin entropisi hesaplanır (E1, E2).



3) Average entropy is calculated from subsets.
N1= Subset -1 element count
N1= Subset -2 element count
N=Main Set element count
Average entropy= (N1 / N) * E1 + (N2/N)* E2

3) Alt kümelerden ortalama entropi hesaplanır.
N1= Alt küme -1 eleman sayısı
N1= Alt küme -2 eleman sayısı
N=Ana Küme eleman sayısı
Ortalama entropi= (N1 / N) * E1 + (N2/N)* E2

$$\text{Ortalama Entropi} = \left(\frac{17}{30}\cdot 0.787\right)+\left(\frac{13}{30}\cdot 0.391\right)=0.615$$

4) Information Gain = Entropy of the Population – Average Entropy
4) Bilgi Kazancı = Ana Kütlenin Entropisi – Ortalama Entropi

Information Gain= 0.996 - 0.615 = 0.38 for this split

Comment: If the gain goes towards 1, it is interpreted as good. A value of 0.38 indicates that the entropy is not good.
Yorum: Kazanç 1'a doğru giderse iyi olarak yorumlanır. 0,38 değeri entropinin iyi olmadığını gösterir.

Example:
Örnek:

a) There are elements from two types of sets (A, B) in a dataset. If the number of elements in set A is 200 and the number of elements in set B is 50, calculate the entropy.
a) Bir veri yığını içinde iki tür kümeden (A, B) elemanlar bulunmaktadır. A kümesindeki eleman sayısı 200, B kümesindeki eleman sayısı 50 ise entropy'i hesaplayınız.

X1 = 200/250 = 0.8
X2 = 50/250 = 0.2

H(X)=-(0.8*Log2 (0.8) + 0.2* Log2(0.2)) = 0,72193
Log2(a)=log10(a)/Log10(2)

b) These two clusters are classified,

b) Bu iki küme sınıflandırılmış,

In the first classification, there are 195 elements from set A and 3 from set B. In the second classification, there are 5 elements from set A and 47 in set B.

Calculate the entropy of the sets.

İlk sınıflandırmada A kümesinden 195, B kümesinden 3

İkinci sınıflandırmada ise A kümesinden 5, B kümesinde 47 eleman bulunmaktadır.

Kümelerin entropy'lerini hesaplayınız.

H(X1)= -((195/200)*Log2 (195/200) + (5/200)*Log2(5/200)) = 0,16866

H(X2)= - ((3/50)*Log2(3/50) + (47/50)*Log2(47/50)) = 0,3274

Average entropy =(198/200)*H(X1)+ (52/200)*H(X2)=0.252

Ortalama entropy =(198/200)*H(X1)+ (52/200)*H(X2)=0,252

Information Gain = H(X)-Average Entropy = 0.72193 – 0.252=0.47

Bilgi Kazancı= H(X)-Ortalama Entropy = 0, 72193 – 0,252=0,47

10. Algoritmik Olasılık

Bilgi kuramı; Claude E. Shannon tarafından güvenli şekilde veri sıkıştırma, depolama ve iletme gibi sinyal işleme işlemlerinin kısıtlarını bulmak için geliştirilmiştir.

Bilginin önemli bir ölçütü, genellikle depolama ve iletişim için gerekli olan parçaların ortalama sayısı olan entropidir. Entropi, bir rastgele değişkenin değerini tahmin ederken belirsizliği nicelikselleştirir. Örneğin, bir yazı tura oyunun sonuç için sağladığı bilgi, bir zar atma oyunun sonuç için sağladığı bilgiden daha azdır. Yazı tura oyununda eşit olasılıklı iki sonuç vardır, zar atma oyununda ise eşit olasılıklı altı sonuç. Bu nedenle yazı tura oyunu daha düşük entropiye sahiptir.

Bilgi kuramının alanı matematik, istatistik, bilgisayar bilimi, fizik, nörobiyoloji ve elektrik mühendisliği ile kesişir. Voyager derin uzay görevleri, kompakt diskin geliştirilmesi, cep telefonlarının yapılabilirliği, internetin geliştirilmesi, dilbilimi araştırmaları gibi pek çok konuda başarının üzerinde büyük etkisi olmuştur. Bilgi kuramının önemli alt dalları; kaynak kodlaması, kanal kodlaması, algoritmik karmaşıklık kuramı, algoritmik bilgi kuramı gibi alanlardır. Ayrıca Data Analitiği alanında sınıflandırma problemlerinde; özellikle de karar ağaçlarının oluşturulmasında bilgi kuramı ve buna bağlı entropi kavramı kullanılmaktadır.

Ray Solomonoff'a kadar olasılık hesabı klasik yönteme dayanmaktadır. Sonrasında ise modern olasılık hesabı başlar. 1960'lı yılların ortalarında ise birbirinden bağımsız ve habersiz iki bilim insanı Kolmogorov ve Chaitin'de (ABD'li matematikçi ve bilgisayar bilimci, 1947-) algoritmik olasılığın kurallarını ortaya koyar. Bu kavram enformasyon kuramına da (bilgi teorisi) yeni bir bakış açısı kazandırır.

Ray Solomonoff (25 Temmuz 1926 - 7 Aralık 2009), algoritmik olasılığın, Genel Tümevarım Teorisi'nin mucididir ve algoritmik bilgi teorisinin kurucusudur. Data Analitiği, tahmin ve olasılığa dayalı yapay zeka dalının yaratıcısıdır. Anlamsal olmayan Data Analitiği hakkındaki ilk çalışması 1956 yılında yayınlanmıştır.

Solomonoff ilk olarak 1960 yılında Kolmogorov karmaşıklığı ve algoritmik bilgi teorisini başlatan teoremi yayınlayarak algoritmik olasılığı tanımladı. Bu sonuçları ilk olarak 1960'da Caltech'te bir konferansta ve Şubat 1960'ta "A Preliminary Report on a Preliminary Report on a General Theory of Endüktif Çıkarım" adlı kitabında açıklamıştı. yayınlar, "A Formal Theory of Endüktif Çıkarım", Kısım I ve Kısım II.

Algoritmik olasılık, en basit hipotezin (en kısa program) en yüksek olasılığa sahip olduğu ve giderek daha karmaşık hipotezlerin giderek daha küçük olasılıklar aldığı, belirli bir gözlemi açıklayan her bir hipoteze (algoritma/program) bir olasılık değeri atamak için makineden bağımsız bir yöntemdir.

Solomonoff, sağlam felsefi temellere dayanan ve kökeni Kolmogorov karmaşıklığı ve algoritmik bilgi teorisinde bulunan evrensel tümevarımsal çıkarım teorisini kurdu. Teori, Bayesian çerçevesinde algoritmik olasılık kullanır. Evrensel öncelik, tüm hesaplanabilir ölçüler sınıfının üzerine alınır; hiçbir hipotezin sıfır olasılığı olmayacaktır. Bu, Bayes kuralının (nedensellik) bir dizi olaydaki en olası bir sonraki olayı ve bunun ne kadar olası olacağını tahmin etmek için kullanılmasına olanak tanır.[

En iyi algoritmik olasılık ve genel tümevarımsal çıkarım teorisi ile tanınmasına rağmen, hayatı boyunca, çoğu yapay zekadaki hedefine yönelik birçok başka önemli keşif yaptı: olasılıksal yöntemler kullanarak zor problemleri çözebilecek bir makine geliştirmek.

Algoritmik Olasılık

Bir bilgi parçasının karmaşıklığını ölçmek mümkün mü? Matematikçiler buna evet cevabını verirler. Öncelikle bilgiye bilgisayarın bakış açısından baktılar. Bir bilgisayar için bir bilgi parçası yalnızca bir sembol dizisidir ve ikili sayı dizisine (0 ve 1) dayanır. İkili sayı dizgesini kullanarak her bilgiyi karşı tarafa iletme ya da işleme şansı doğar. Bilginin iletilmesi ya da işlenebilmesi için bilgisayara komut girmek gereklidir. Bu komutlara da algoritma denir.

Şimdi bir örneğe bakalım. 20 kez bir bozuk para atalım. Yazı gelince 0, tura gelince 1 yazalım. 2^20 farklı say dizisi elde edilebilir:
0000 0000 0000 0000 0000..... 1111 1111 1111 1111 1111.

Sonuçlar 01010101010101010101 veya 01101100110111100010 biçiminde olabilir. İnsanlar ilk diziye bakınca bunun rasgele bir dizi olmadığını anlar. Ancak ikinci dizi için tesadüfi dizilim olduğunu söylerler. Oysaki bu insan algısı için geçerli bir durumdur. Bilgisayar iki dizilim arasındaki farkı algılamaz.

Bilgisayara göre bir paranın bir kere atılması deneyinde 2^1=2 olasılık vardır; ya 0 ya da 1 dir. Bilgisayara göre bir paranın iki kere atılması deneyinde 2^2=4 olasılık vardır; 00, 01, 10, 11 durumlarından biri olur.

Bilgisayara göre bir paranın üç kere atılması deneyinde 2^3=8 olasılık vardır; 000, 001, 010, 011, 100, 101, 110, 111 durumlarından biri olur.

O halde bilgisayara göre bir paranın yirmi kere atılması deneyinde 2^20 olasılık vardır. 01010101010101010101 veya 01101100110111100010 dizilimde bu olasılıklardan ikisini

temsil eder. Sonuçta bilgisayarlar dizilimler hakkında yorum yapmak yerine sadece verilen komutları algılar.

Bu algı probleminin temel çözümü, algoritmik olasılıkta yatar. Diyelim ki 20 kere değil de 20 katrilyon zar atalım. Sonuçlar yukardaki gibi olsun. Bu durumda ilk mesaj için komut olarak 10 katrilyon kere 01 yaz demek yeterli olur. Ancak ikinci mesaj için kısaltma yapılamaz. Tüm sayı dizisinin yazılması yani 20 katrilyonun hepsinin komut olarak girilmesi gerekir. Kısacası Kolmogorov karmaşıklığı teorisi olarak da bilinen Algoritmik Bilgi Teorisi (AIT) basit bir gözleme dayanmaktadır: Karmaşık nesneler kısa bir programla tanımlanamaz. Özellikle, algoritmik bilgi teorisi rastgele dizgi ve rastgele sonsuz dizilerin resmi ve özenli tanımlarını verir.

10.1. Akıl Yürütme

Mantıkta akıl yürütme, muhakeme ya da uslamlama bilinen olgular ve kurallar kullanılarak yeni bilgiye ulaşılmasıdır. Akıl yürütme üç başlıkta incelenebilir: tümdengelim (dedüksiyon), tümevarım (indüksiyon) ve analoji. Klasik mantığın temelinde tümdengelim vardır.

Tümdengelim öncül bilgilerden yeni sonuçlar çıkarmaktır. Öncüller halihazırda bilinen ya da varsayılan olgulardır. Örneğin, "Arabalar dört tekerlidir" ve "Kara Şimşek bir arabadır" öncüllerinden tümdengelim yoluyla "Kara Şimşek dört tekerlidir" sonucu elde edilebilir.

Tümevarım özelden yola çıkılarak genel hakkında bilgi elde edilmesidir. Bilimsel yöntemin temeli olan tümevarımda, çok sayıda tikel bilgi kullanılarak tümel bir kural doğrulanır. Örneğin, "Serçe, güvercin, karga uçar", "Serçe, güvercin, karga kuştur" tikel bilgilerinden "Kuşlar uçar" tümel bilgisine varılabilir. Ancak, tümevarımsal akıl yürütme, tümdengelimde olduğu gibi kesin bir sonuca götürmez. Tümevarım var olan gözlemlerle tutarlı olmasına karşın, henüz yapılmamış gözlemlerle çelişmesi mümkündür. Örneğin, "Penguen uçamaz" ve "Penguen kuştur" bilgileri edinildiğinde, daha dar bir öncül kümesiyle çıkarsanmış olan "Kuşlar uçar" önermesi geçersiz olacaktır.

Dışaçekim, belli koşullara uyan olgular için bir açıklama üretmek için kullanılan akıl yürütme metodudur. En iyi açıklamanın üretilmesi şeklinde de ifade edilebilir. Dışaçekim, eldeki -çoğu zaman eksik olan- bilgi ile mümkün olanın en iyisinin yapılmaya çalışıldığı günlük karar verme süreçlerinin temelini teşkil eden bir muhakeme metodudur. Bir dizi semptomun varlığı bilinirken bunları en iyi şekilde açıklayan teşhisi bulmaya yönelik tıbbi tanı çabası dışaçekimin en yaygın örneklerinden biri olarak verilebilir. Bir hâkimin, mahkemeye sunulan delilleri, savcının mı yoksa savunmanın mı argümanlarının daha iyi açıkladığını tespiti çabası dışaçekim örneğidir. Peirce'ye göre her bilimsel araştırma şaşırtıcı gelen bir olgunun gözemlenmesi ile başlar. Bilimsel araştırmanın ilk adımı olan olan dışaçekim, gözlemlenen olgunun nasıl olup ta ortaya çıktığını açıklamaya yönelik hipotez veya hipotezlerin

geliştirilmesi sürecidir. Bu durumda Peirce, dışaçekimin salt bir akıl yürütme çeşidi olmayıp aynı zamanda bilimsel buluş metodu oluduğunu da ifade eder.

Analoji iki şey arasındaki benzerliğe dayanarak birisi hakkında verilen bir hükmü diğeri hakkında da vermek şeklindeki akıl yürütme yoludur. Örneğin, "Dünya gezegeninin atmosferi vardır ve üzerinde canlılar yaşar", "Mars gezegeninde atmosfer vardır" öncüllerinden hareketle "O zaman Mars gezegeninde canlılar bulunması gerekir" sonucunun elde edilmesi bir analoji örneğidir. Dünya ile Mars arasındaki bir benzerlikten hareketle Dünya için geçerli olan bir durumun Mars için de geçerli olması gerektiği kabul edilerek Marsta canlılar olması gerektiği hükmüne varılmıştır. Analojinin hem indüktif hem dedüktif metodun kullanıldığı bir akıl yürütme yoludur. "Eğer atmosfer var ise canlı olmalıdır" ve "Dünya ve Mars'ın atmosferleri yapı olarak aynıdır" hükümleri induktif akıl yürütme ile zımni (örtülü) olarak üretilmiş daha sonra bu hükümden "O zaman Mars gezegeninde canlılar bulunması gerekir" hükmü dedüktif metotla üretilmiştir. Analoji varsayımsal (hypothetique) bir dedüksiyon olup dayandığı zımni hükümler ispat edilmiş değil varsayılmıştır. Bu sebeple analoji ile verilen hüküm bir zorunluluk değil ancak bir ihtimal bildirir.

Algoritmik bilgi teorisi

Algoritmik bilgi teorisi, bilgi işlem ve bilgi arasındaki ilişki ile ilgilenen bilgi teorisi ve bilgisayar biliminin bir alt alanıdır. Gregory Chaitin'e göre, "Shannon'un bilgi teorisini ve Turing'in hesaplanabilirlik teorisini bir kokteyl çalkalayıcısına koymanın ve şiddetle sallamanın sonucudur."

Algoritmik bilgi teorisi, temel olarak diziler (veya diğer veri yapıları) üzerindeki karmaşıklık ölçümlerini inceler. Çoğu matematiksel nesne, diziler cinsinden veya dizi dizilerinin sınırı olarak tanımlanabildiğinden, tamsayılar da dahil olmak üzere çok çeşitli matematiksel nesneleri incelemek için kullanılabilir. Teori Ray Solomonoff tarafından kuruldu.

Algoritmik olasılık

Algoritmik bilgi teorisinde, Solomonoff olasılığı olarak da bilinen algoritmik olasılık, belirli bir gözleme önceden bir olasılık atamanın matematiksel bir yöntemidir. 1960'larda Ray Solomonoff tarafından icat edildi.

Endüktif çıkarım teorisinde ve algoritma analizlerinde kullanılır. Genel tümevarımsal çıkarım teorisinde, Solomonoff, bu formülle elde edilen önceliği, Bayes'in tahmin kuralında kullanır.

Kullanılan matematiksel formalizmde, gözlemler sonlu ikili diziler biçimindedir ve evrensel öncelik, sonlu ikili diziler kümesi üzerindeki bir olasılık dağılımıdır. Öncelik evrenseldir Turing-hesaplanabilirlik duygusu, yani hiçbir dizgenin sıfır olasılığı yoktur. Hesaplanamaz, ancak tahmin edilebilir.

Kolmogorov karmaşıklığı

Algoritmik bilgi teorisinde (bilgisayar bilimi ve matematiğin bir alt alanı), bir metin parçası gibi bir nesnenin Kolmogorov karmaşıklığı, nesneyi çıktı olarak üreten en kısa bilgisayar programının (önceden belirlenmiş bir programlama dilinde) uzunluğudur.

Nesneyi belirtmek için gereken hesaplama kaynaklarının bir ölçüsüdür ve ayrıca tanımlayıcı karmaşıklık, Kolmogorov-Chaitin karmaşıklığı, algoritmik karmaşıklık, algoritmik entropi veya program boyutu karmaşıklığı olarak da bilinir. Adını bu konuda ilk kez 1963'te yayınlayan Andrey Kolmogorov'dan almıştır.

Kolmogorov karmaşıklığı kavramı, Cantor'un köşegen argümanına, Gödel'in eksiklik teoremine ve Turing'in durma problemine benzer imkansızlık sonuçlarını ifade etmek ve kanıtlamak için kullanılabilir.

Özellikle, hemen hemen tüm nesneler için, bırakın kesin değerini, Kolmogorov karmaşıklığı (Chaitin 1964) için bir alt sınır bile hesaplamak mümkün değildir.

Bayes çıkarımı

Bayes çıkarımı, daha fazla kanıt veya bilgi elde edildikçe bir hipotezin olasılığını güncellemek için Bayes teoreminin kullanıldığı bir istatistiksel çıkarım yöntemidir. Bayes çıkarımı, istatistikte ve özellikle matematiksel istatistikte önemli bir tekniktir. Bayes güncellemesi, bir veri dizisinin dinamik analizinde özellikle önemlidir. Bayes çıkarımı, bilim, mühendislik, felsefe, tıp, spor ve hukuk dahil olmak üzere çok çeşitli faaliyetlerde uygulama bulmuştur. Karar teorisi felsefesinde, Bayes çıkarımı, genellikle "Bayes olasılığı" olarak adlandırılan öznel olasılık ile yakından ilişkilidir.

Olasılık

Olasılık, bir olayın meydana gelme olasılığının ölçüsüdür. Olasılık ve istatistik sözlüğüne bakın. Olasılık, 0 ile 1 arasında bir sayı olarak ölçülür, burada genel olarak 0 imkansızlığı, 1 ise kesinliği gösterir. Bir olayın olma olasılığı ne kadar yüksekse, olayın olma olasılığı da o kadar yüksektir. Basit bir örnek, adil (tarafsız) bir madeni paranın havaya atılmasıdır. Madeni para adil olduğundan, iki sonucun da ("yazı" ve "tura") eşit derecede olasıdır; "tura" olasılığı, "tura" olasılığına eşittir; ve başka hiçbir sonuç mümkün olmadığından, "tura" veya "tura" olasılığı 1/2'dir (bu, 0,5 veya %50 olarak da yazılabilir).

Bu kavramlara matematik, istatistik, finans, kumar, bilim (özellikle fizik), yapay zeka/Data Analitiği, bilgisayar bilimi, oyun teorisi, ve felsefe, örneğin, olayların beklenen sıklığı hakkında çıkarımlar yapmak. Olasılık teorisi, karmaşık sistemlerin altında yatan mekanikleri ve düzenlilikleri tanımlamak için de kullanılır.

Gregory Chaitin

Gregory John Chaitin (/ˈtʃaɪtɪn/; 15 Kasım 1947 doğumlu) Arjantinli-Amerikalı bir matematikçi ve bilgisayar bilimcisidir. 1960'ların sonundan başlayarak, Chaitin algoritmik

bilgi teorisine ve metamatematiğe, özellikle Gödel'in eksiklik teoremine eşdeğer bir bilgisayar teorik sonucuna katkılarda bulundu. Andrei Kolmogorov ve Ray Solomonoff ile birlikte bugün Kolmogorov (veya Kolmogorov-Chaitin) karmaşıklığı olarak bilinen şeyin kurucularından biri olarak kabul edilir. Bugün, algoritmik bilgi teorisi, herhangi bir bilgisayar bilimi müfredatında ortak bir konudur.

Hipotez

Bir hipotez (çoğulu hipotezler), bir fenomen için önerilen bir açıklamadır. Bir hipotezin bilimsel bir hipotez olması için, bilimsel yöntem, kişinin onu test edebilmesini gerektirir. Bilim adamları genellikle bilimsel hipotezleri, mevcut bilimsel teorilerle tatmin edici bir şekilde açıklanamayan önceki gözlemlere dayandırırlar. "Hipotez" ve "teori" kelimeleri sıklıkla eş anlamlı olarak kullanılsa da, bilimsel bir hipotez, bilimsel bir teori ile aynı şey değildir. Çalışan bir hipotez, daha fazla araştırma için önerilen geçici olarak kabul edilen bir hipotezdir.

Evrensel Turing makinesi

Bilgisayar biliminde, evrensel bir Turing makinesi (UTM), isteğe bağlı bir girişte rastgele bir Turing makinesini simüle edebilen bir Turing makinesidir. Üniversal makine bunu esasen hem simüle edilecek makinenin açıklamasını hem de kendi bandından girişini okuyarak başarır. Alan Turing, 1936-1937'de böyle bir makine fikrini ortaya attı. Bu ilke, 1946'da John von Neumann tarafından "Elektronik Hesaplama Aleti" için kullanılan ve şimdi von Neumann'ın adını taşıyan depolanmış programlı bir bilgisayar fikrinin kaynağı olarak kabul edilir: von Neumann mimarisi. Hesaplama karmaşıklığı açısından, çok bantlı evrensel bir Turing makinesinin, simüle ettiği makinelere kıyasla yalnızca logaritmik faktör açısından daha yavaş olması gerekir.

10.2. Kolmogorov Aksiyomları

Olasılık teorisinde Kolmogorov aksiyomları, temelde üç aksiyomdur:

- Birinci aksiyom: Bir olayın olasılığının negatif-olmayan bir reel sayı olduğunu söyler.
- İkinci aksiyom: Örnekleme uzayının tümünü kapsayan bir basit olayın ortaya çıkması için olasılık her zaman birdir.
- Üçüncü aksiyom: Birbiri ile bağlantısız olayların olasılıkları toplanır.

Ancak olasılık ile ilgili bilinmesi gereken şeyler bu kavramlar ile sınırlı değildir. 1960 yılında Ray Solomonoff (ABD'li matematikçi 1926-2009), algoritmik olasılığı keşfeder. **Algoritmik olasılık, felsefik yapı olarak Bayes'in nedensellik kuralını temel alır:**

- Hangi durum ya da olay, diğer bir olayın ya da durumu ortaya çıkarır?
- Aralarındaki ilişki nasıldır? gibi sorulara yanıt arar.

Olasılık teorisinde Kolmogorov aksiyomları, temel üç aksiyomdur. Belirli bir E olayı için P olasılığı varken matematik notasyonla olarak ifade edilirken Kolmogorov aksiyomlarını tatmin etmesi temeline bağlanmıştır. Bu aksiyomlar, ilk defa 20. yüzyılda Rus istatistikçisi Andrey Kolmogorov tarafından ortaya atılmıştır.

Bu aksiyomları açıklamak için matematiksel şekilde ve notasyonla şu kavramların varsayılması gereklidir: (Ω, F, P) ifadesi bir ölçüm uzayı olsun ve burada P(Ω) = 1 olduğu kabul edilsin. Bu hâlde (Ω, F, P) bir olasılık uzayıdır ve Ω örneklem uzayı, F olay uzayı ve P olasılık ölçüsü olarak tanımlanırlar.

Birinci aksiyom: Bir olayın olasılığı bir negatif-olmayan reel sayıdır ve bu sayı şöyle ifade edilir:

$$P(E) \geq 0 \qquad \forall E \subseteq F \;\; Burada, F\ olay\ uzayıdır.$$

İkinci aksiyom: Bu birim-ölçüsü varsayımıdır: Örnekleme uzayının tümünü kapsayan bir basit olay ortaya çıkması için olasılık 1dir. Daha belirli bir şekilde ifadeyle;

Örneklem uzayını taşan hiçbir basit olay mümkün değildir: P(Ω)=1
Bu aksiyom bazı hatalı olasılık hesaplamalarında çok kere temel bir hatanın ortaya çıkmasına neden olmuştur. Eğer tüm örneklem uzayı kesinlikle tanımlanamıyorsa bunun herhangi bir alt setinin tanımlanması da imkânsızdır.

Üçüncü aksiyom: Bu σ-toplanabilirlik varsayımıdır. Herhangi bir ikişerli bağlantısız ortaya çıkan sayılabilir olaylar dizisi, E1, E2, ... şu eşitliği tatmin eder:
P(E1 U E2 U ...)=$\sum_i P(E_i)$
Aksiyomun daha genel olması için σ-cebiri gereklidir.

Kolmogorov aksiyomları kullanılarak olasılıkların hesaplanması için diğer kullanışlı kurallar ortaya çıkartılabilir. Bunlardan en önemlisi
$P(A \cup B)=P(A) + P(B) – P(A \cap B)$

Bu kurala **toplama kuralı** veya **olasılık için toplama yasası** adı verilir. Buna gore bir *A* olayi **veya** bir *B* olayının olması olasılığı *A* olayı için olasılık artı *B* olayı için olasılık eksi hem *A* hem de *B* olayının birlikte olasılığına eşittir.

Bu yasadan **Kapsama-dışlama prensibi** adı verilen şu sonuç çıkartılır:
$P(\Omega \setminus E)=1-P(E)$

Bir başka deyimle herhangi bir olayın **olmama** olasılığı **1** eksi olayın **olma** (ortaya çıkma) olasılığıdır.

The axioms of Kolmogorov. Let S denote a sample space with a probability measure P defined over it, such that probability of any event $A \subset S$ is given by P(A). Then, the probability measure obeys the following axioms:
(1) $P(A) \geq 0$,
(2) $P(S)=1$,
(3) If $\{A1, A2,...Aj ,...\}$ is a sequence of mutually exclusive events such that
$Ai \cap Aj = \emptyset$ for all i, j, then $P(A1 \cup A2 \cup \cdots \cup Aj \cup \cdots) = P(A1)+P(A2)+$
$\cdots + P(Aj) + \cdots$.
The axioms are supplemented by two definitions:
(4) The conditional probability of A given B is defined by
$P(A|B) = P(A \cap B)/ P(B)$,
(5) The events A, B are said to be statistically independent if
$P(A \cap B) = P(A)P(B)$.
This set of axioms was provided by Kolmogorov in 1936.

Example. The Manager of Ffyfes, who import bananas to the U.K. from many sources, has discovered an unmarked crate, and he wishes to determine its origin. 40 percent of the crates in stock come from Guatemala and 60 percent from Cuba. On average, 1/2 the Guatemalan bananas are bad and 1/6 of the Cuban bananas are bad. The manager opens the crate and pulls out a banana that happens to be bad. In the light of this evidence, what it is the most likely origin of the crate?

Answer. Let H_1 denote the hypothesis that the crate if from Cuba and let H_2 denote the hypothesis that it is from Guatemala. Let E be the event of discovering a rotten banana. There are the following items of information:

$$P(H_1) = \frac{60}{100} = \frac{3}{5} \qquad P(H_2) = \frac{20}{100} = \frac{2}{5},$$

$$P(E|H_1) = \frac{1}{6} \qquad P(E|H_2) = \frac{1}{2},$$

$$P(E|H_1)P(H_1) = \frac{1}{6} \times \frac{3}{5} = \frac{1}{10} \simeq P(H_1|E),$$

$$P(E|H_2)P(H_2) = \frac{1}{2} \times \frac{2}{5} = \frac{1}{5} \simeq P(H_2|E).$$

In the light of the evidence, it seems twice as likely that Guatemala is where the crate is from.

10.3. Kolmogorov Karmaşıklığı

Kolmogorov karmaşıklığı (tanımsal karmaşıklık, Kolmogorov-Chaitin karmaşıklığı, stokastik karmaşıklık, algoritmik entropi veya program boyu karmaşıklığı olarak da bilinir), bilgisayar biliminde, bir metin parçası gibi bir nesneyi tanımlamak için kullanılması gereken bilgi işlemsel kaynakların ölçüsü.

Belirli bir dizeyi üreten ve sonra duran (sonsuz döngüler istemiyoruz) en kısa bilgisayar programının uzunluğu Kolmogorov karmaşıklığı olarak adlandırılır. Diğer bir deyiş ile bir dizinin karmaşıklığı, o diziyi ileten minimal algoritmanın uzunluğuna eşittir. Dolayısıyla bit sayısı Kolmogorov karmaşıklığına yaklaşık olarak eşit olan dizi, rasgeledir. Bir dizinin Kolmogorov karmaşıklığını, bu dizideki bit cinsinden ölçülen "bilgi miktarı" olarak yorumlamak bizi bazı pratik sonuçlara götürür. keyfi sonlu nesnelerin (grafikler vb.) Kolmogorov karmaşıklığını, ikili kodlamalarının karmaşıklığı olarak tanımlamamızı sağlar.

Konuyu daha basit bir biçimde düşünmek gerekirse Kolmogorov karmaşıklığı "sıkıştırılmış boyut" anlamına gelir. Zip, gzip, bzip2, sıkıştırma, rar, arj, vb. gibi programlar, bir dosyayı (metin, resim veya diğer bazı veriler) muhtemelen daha kısa bir dosya biçiminde sıkıştırır. Orijinal dosya daha sonra bir "açma" programı ile geri yüklenir. Düzenli bir yapıya sahip bir dosya önemli ölçüde sıkıştırılabilir. Sıkıştırılmış boyutu, eski uzunluğuna kıyasla çok daha küçüktür. Öte yandan, düzenliliği olmayan bir dosya pek sıkıştırılamaz ve sıkıştırılmış boyutu orijinal boyutuna yakındır. Bu örneklem çok yüzeysel olsa da yine de oldukça karmaşık bir konu hakkında bir ön fikir olmasını sağlayacaktır.

Gözlemlerden ya da deneylerden elde edilen verilerin tablolar halinde yazılması bir teori yaratmıyor. Söz konusu ham verilerin yorumlanarak, herkesin anlayacağı kısa bir dille anlatılması gerekiyor. O da yetmiyor, teorinin gelecekte olacaklar hakkında bilgi içermesi gerekiyor.

Gezegenin yörüngesini biliyorsam, onun ne zaman nerede olacağını hesaplayabiliyorum. Bu iş, kehanetten çok farklı bir şeydir. Bunlar olduğunda, ham veriler bir teoriye dönüşmüş oluyor.

Elbette, toplanan verilerin duyarlığı, kullanılan gözlem/deney aletlerinin gelişmişliğine bağlı olduğu gibi, verilerin yorumlanması da bilim adamının bilgi ve yetenekleriyle sınırlıdır. Şu anda, bir teorinin doğru ya da yanlışlığı amacımız için önem taşımıyor. Yanlış teoriler, nasıl olsa, bir gün bilimsel bilgilerin biriktiği ambardan atılacaktır. Bilimin gücü burada yatar. Bilimin bilgi ambarı çok dinamiktir, yanlış olduğu kanıtlanan teoriler hemen yerlerini yeni teorilere kendiliğinden bırakırlar. Şimdi, bir teori kurma olgusunu algoritmik seçkisizlik kavramıyla ifade edeceğiz. Algoritmik seçkisizlik tanımını vermeden önce, Solomonoff"un

bilimsel teoriyi açıklamak için kullandığı “inductive inference: endüktif çıkarım” yönteminden söz etmeliyiz. Bilim adamı bir sürü deney/gözlem yapar. Bunları bitlerden oluşan bir dizi (mesaj) olarak düşünelim. Bilim adamı bu mesajı iletmek istemektedir. Bu mesajı gönderen en az bir tane algoritma vardır ve o da dizinin kendisidir. Bundan başka algoritmalar da olabilir. Bilim adamı mesajı gönderen bir algoritma kurmuş olsun. Algoritma, ilettiği mesajdan kısa değilse bir teori olamaz. Algoritma ilettiği mesajdan daha kısa ise, diziyi aynen iletmekle kalmayıp gelecek gözlemler için de öngörü yapıyorsa, bu algoritma bir teoridir. Bu koşulu sağlayan birden çok algoritma varsa, daima en kısa (bit sayısı en az) olan algoritma tercih edilir. Bu tercih Occam’s razor diye bilinir: Aynı işi yapan teoriler arasından en basiti tercih edilmelidir.

Algoritmik Seçkisizlik: Yukarıdaki örneklerden hereketle, Chaitin ve Kolmogorov, seçkisiz diziyi şöyle tanımladılar: Kendisinden daha kısa bir algoritma ile yazılamayan dizi seçkisizdir. Bu tanım, sezgisel kavrama dayalı olasılık kuramını sağlam bir temel üzerine taşımaktadır. Elbette, seçkisizliğin bu yeni tanımı, olasılık kuramının aksiyomlarını yok etmiyor, olasılığın hiç bir uygulamasını değiştirmiyor, yalnızca temeli sağlamlaştırıyor. Olasılık açısından peşinde olduğumuz şey, gözlemlerden çıkan seçkisiz bir x1 , x2 , x3 , ... , xn dizisi verilmişken, bir sonraki terimin, yani xn+1 teriminin ne olacağını öngörebilmektir. x1 , x2 , x3 , ... , xn dizisini ileten ve xn+1 teriminin ne olacağını öngören ve diziden kısa olan algoritma bir teoridir.

Verilen bir diziyi ileten sonsuz sayıda algoritma kurulabilir. Örneğin, “233’e 1 ekle”, “235’ten 1 çıkar”, 117’yi 2 ile çarp” gibi algoritmaların hepsi 234 dizisini iletir. Bu tür algoritmalardan sonsuz sayıda yazılabileceği açıktır. Bizim için ilginç olanı en küçük olanıdır. Aynı diziyi ileten algoritmalar arasında en kısa olana minimal algoritma diyeceğiz. Bir dizi için bir tane minimal algoritma olabileceği gibi, bir çok minimal algoritma da olabilir.

İlettiği dizi ister seçkili, ister seçkisiz olsun minimal bir algoritmanın kendisi daima seçkisiz olmak zorundadır. Aksi taktirde, onu ileten daha kısa bir algoritma var olur ve dolayısıyla söz konusu algoritma minimal olamaz.

Karakterlerden (harf ve simgeler) oluşan bir s dizisi düşünelim. s dizisini yazdıran bir P programına s dizisinin iletiyor diyelim. P’nin uzunluğu, P içindeki karakterlerin sayısıdır. s dizisini ileten en kısa P programının uzunluğuna s dizisinin karmaşıklığı denir.

s dizisini, yukarıdakiler gibi bitlerden oluşan bir dizi olarak düşünürsek, bu dizinin karmaşıklığı o diziyi ileten minimal algoritmanın uzunluğuna eşit olur. Bit sayısı Kolmogorov karmaşıklığına eşit olan dizi seçkisizdir. Tabii, buradaki eşitlik yaklaşıklık anlamındadır. Dizilerin bit sayıları çok çok büyüdüğünde, aradaki farkın önemi kalmamaktadır. Kolmogorov karmaşası, seçkisizliği tanımlamakla kalmıyor, seçkisizliğin ölçümünü de veriyor.

10.4. Seçkisiz Sayıların Çokluğu

Klasik anlamda, olasılık ile seçkisizlik (randomness) eşanlamlıdır. Seçkisiz bir süreçte, olayların (çıktıların) olma olasılıkları birbirlerine eşittir. Başka bir deyişle, bir süreçte seçkisizlik "amaç, neden, sıra ve öngörü yokluğu" diye tanımlanabilir. Bu nedenle, seçkisiz süreç, çıktısı öngörülebilir bir biçime (pattern) sahip olmayan ardışık oluşumlar zinciridir. Özel olarak, istatistikte, yanlı (bias) olmayan ya da bağlılaşımlı (correlated) olmayan olayları belirlemek için kullanılır. İstatistiksel seçkisizlik, daha sonraları informasyon kuramında informasyon entropisi kavramı içine alınmıştır.

Seçkisizlik kavramı, başlangıçta şans oyunlarından çıkmıştır. Örneğin zar atma, rulet oyunu, oyun kartlarını karma vb. Daha sonra yapılan elektronik kumar makinalarında da seçkisiz sayı üretimi işin esasıdır. Ama bu işte çok hile yapılabileceği için, bu tür oyun makinaları bir çok ülkede ya yasaklıdır ya da devletin sıkı denetimi altındadır. Bizde olduğu gibi, bazı ülkelerde, seçkisiz sayı üretimine dayalı piyangolar ülke genelinde serbestçe oynanabilir. Bundan farklı olarak, çıktısı önceden öngörülemeyecek spor karşılaşmaları, at yarışları vb. oyunlar da seçkisizliğin (olasılığın) ilgi alanındadır.

Olasılık kavramının geçtiği her yerde, olabilecek olayları sayılarla ifade etmek mümkündür. Dolayısıyla, konu, esasında seçkisiz sayı üretimine dayalıdır. Stephan Wolfram' a göre, seçkisiz sayı üretme işi üç ayrı sınıfa ayrılabilir. Bu üç sınıfta üretilen sayıların nitelikleri birbirlerinden farklıdır.

1. *Çevreden gelen seçkisizlik*. Örneğin, çoğalmayı açıklayan hareket (Brownian motion).
2. *Başlangıç koşullarına hassas bağlı seçkisizlik.* Örneğin, kaos.
3. *Sözde seçkisizlik.* Bu sayılar tasarlanan bir sistem tarafından üretilir. Örneğin, bir bilgisayarla üretilen seçkisiz sayılar... Bu tür sayılar belli bir algoritma ile üretilir. Algoritma çok ağır hesaplamalara dayandırılarak, çıktı hiç kimsenin öngöremeyeceği duruma kolayca getirilebilir. Ama üretilen sayılar gerçek anlamıyla seçkisiz sayılamaz.

1960 yılında *Solomonoff*, bilimsel teorinin basit bir açıklamasını vermeye uğraşırken, *algoritmik olasılık* kavramını ilk ortaya atan kişidir. Bundan 5 yıl sonra, Solomonoff'dan ve birbirlerinden habersiz olarak *Kolmogorov* ve *Chaitin* aynı algoritmik seçkisizlik kavramını ortaya koydular. Kolmogorov o zamanının en ünlü matematikçilerinden birisidir, Chaitin ise henüz üniversitede matematik bölümü son sınıf öğrencisidir. Bu öğrencinin, daha sonra yaptığı çalışmalar olasılığa ve informasyon teorisine büyük katkılar sağlayacaktır.

***"Algoritmik seçkisizlik"* ya da *"algoritmik olasılık"* kavramı:**

Örnek 1. Dünyadaki Uzay Merkezi (UM) çok uzaktaki bir gezegene bir araştırmacı göndermiştir. Dalgın araştırmacımız, yapacağı hesaplar için kendisine mutlaka gerekli olan trigonometri cetvelini yanına almayı unutmuştur ve onun bir iletişim aracıyla kendisine acele gönderilmesini istemektedir. Bu uzak gezegenle telgraf, telefon, faks vb iletişim araçlarıyla iletilen mesajların çok pahalıdır. UM, pahalı iletişim ücretini ödemeyi göze alarak, *sin, cos, tan, cot, sec, cosec* fonksiyonları için hazırlanmış, yirmi haneli geniş bir trigonometri cetvelini bir iletişim aracıyla ile göndermek zorunda kalmıştır. Ama UM'de bir matematikçi varsa, işi çok ucuza getirebilir. Koca bir kitap olan trigonometri cetvelini göndermek yerine, *exp(ix) =cosx + isinx* formülünü göndermesi sorunu çözecektir. Bu kısa mesaj, araştırmacının istediği bütün bilgiyi içermektedir.

Örnek 2. Aradan binlerce yıl geçmiş olsun. Bilginimiz yorulmuştur ve hobilerine biraz zaman ayırmak için geçmiş yıllara ait basketbol maçlarını, skorları ve kimin hangi maçta kaç sayı yaptığını bilmek istemektedir. UM bu isteği çok haklı görmüş ve istenen bilgilerin gönderilmesini emretmiştir. Bu kez, matematikçiler de dahil olmak üzere, hiç kimse istenen maçlarla ilgili bilgileri tamamen içeren daha kısa bir mesaj (formül) yazamamıştır. Çaresiz, yüksek ücretler ödenerek, istenen bilgi gönderilecektir.

Bu iki örnekten çıkardığımız sonuç şudur. Bazı mesajları, anlamını aynen koruyarak, kısaltabiliriz. Bazı mesajları asla kısaltamayız.

Algoritmik Seçkisizlik tanımı, yukarıda verilen tanımda olduğu gibi insan sezgisine dayalı olmasın diye bilgisayar terminolojisine dayandırılacaktır. Günümüz bilgisayarları ikili (binary) sayıtlama dizgesine dayanır. İkili (binary) sayıtlama dizgesinde yalnızca 0 ve 1 sayakları (digit) vardır. Bilgisayar terminolojisinde ikili sayı sistemindeki hanelere *bit* denir. Bir bit'te (hanede) ya 0 ya da 1 sayağı yer alır.

İkili sayı dizgesini kullanarak her bilgiyi (mesajı) karşı tarafa gönderebiliriz. Başka bir deyişle, 0 ile 1 lerden oluşan dizilerle istediğimiz her bilgiyi yazabiliriz. Bunun için, örneğin, bir dildeki harfleri, kelimeleri, cümleleri, kavramları,... vb 0 ile 1 lerin belirli bir sırada sıralanmasıyla oluşan birer diziye karşılık getirmek yetecektir. Diziler sonlu ya da sonsuz olabilir. Mesajın ne kadar uzağa gideceğinin ve mesajın anlamının, şu andaki hedefimiz için bir önemi yoktur. O nedenle, mesajları 0 ile 1 lerden oluşan diziler olarak, uzak gezegeni de bilgisayarın çıktısı olarak düşüneceğiz. Amacımız, mesajın (dizinin) bilgisayar çıktısı olarak elde edilmesidir. O zaman mesajı yerine iletilmiş varsayacağız. Mesajı iletmek için, bilgisayara komutlar vermeliyiz. Verilecek komutlar herhangi bir bilgisayar dilinde yazılmış bir programdır. Biz buna *algoritma* diyeceğiz. Fiziksel kısıtlamaları yok sayıp, mesajın gönderilmesi için gerekli zamanın olduğunu ve algoritma doğru ise mesajın daima yerine ulaşığını varsayalım.

Bilgilerimize ya da sezgilerimize dayalı olarak bir dizi hakkında vereceğimiz *seçkili/seçkisiz* kararlarımızın ne kadar yanıltıcı olabileceğine bir çok örnek gösterebiliriz. Örneğin, 3,1451... dizisini gören iki kişi düşünelim. Bunlardan birisi *pi* sayısını biliyor olsun, ötekisi bilmiyor olsun. Birinci kişi bu diziyi istençle yazılmış (seçkili) bir dizi olarak, yani *pi* sayısı olarak algılarken, ikinci kişi bunu tamamen rasgele dizilmiş (seçkisiz) bir dizi olarak görebilir.

Kolmogorov, seçkisiz diziyi şöyle tanımlar: *Kendisinden daha kısa bir algoritma ile yazılamayan dizi seçkisizdir*. Bu tanım, sezgisel kavrama dayalı olasılık kuramını sağlam bir temel üzerine taşımaktadır. Elbette, seçkisizliğin bu yeni tanımı, olasılık kuramının aksiyomlarını yoketmiyor, olasılığın hiç bir uygulamasını değiştirmiyor, yalnızca temeli sağlamlaştırıyor. Olasılık açısından peşinde olduğumuz şey, gözlemlerden çıkan seçkisiz bir x_1 , x_2 , x_3 , ... , x_n dizisi verilmişken, bir sonraki terimin, yani x_{n+1} teriminin ne olacağını öngörebilmektir. x_1 , x_2 , x_3 , ... , x_n dizisini ileten ve x_{n+1} teriminin ne olacağını öngören ve diziden kısa olan algoritma bir teoridir.

Bazı dizilerde tekrarlanan patternler olabilir. *d* dizisi *s* içinde periyodik olarak tekrarlanan bir altdizi olsun. Örneğin, *s = 01010101010101010101* dizisi *d = 01* altdizisinin *10* kez tekrarlanmasıyla oluşmuştur. *P ="n kez d yaz"* algoritması *s* dizisini ileten algoritmalardan birisidir. Ohalde, *s* dizisinin algoritmik karmaşıklığı, *P* algoritmasının uzunluğundan büyük olamaz. Asıl amacımız, *P* nin uzunluğu ile *s* dizisinin bit uzunluğunu karşılaştırmaktır. Bu karşılaştırma bize, *s* dizisinin seçkisiz olup olmadığı konusunda bir ölçü verecektir.

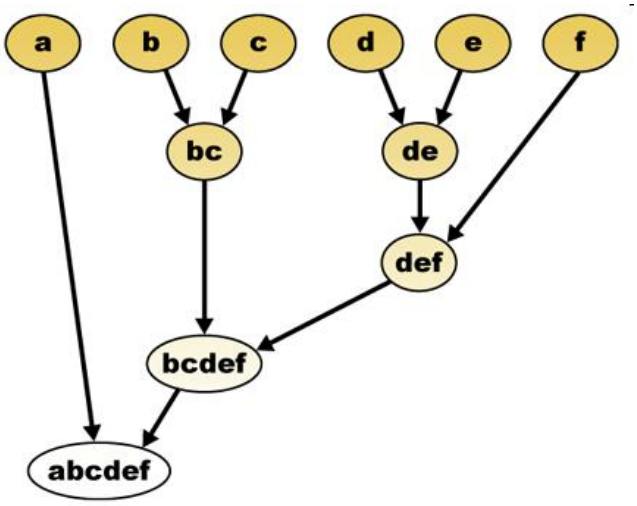
Önce *P* algoritmasının uzunluğunu irdeleyelim. *n* sayısının algoritmik karmaşıklığı yaklaşık olarak $log_2 n$ dir. Yaklaşık diyoruz, çünkü algoritmanın gerçek uzunluğu kullanılan makina diline bağlıdır. Yeterince büyük *n* sayıları için $log_2 n$ sayısı *n* sayısından çok küçüktür, dolayısıyla algoritmanın uzunluğu *s* dizisinin uzunluğu ile karşılaştırılırken göreli olarak ihmal edilebilir. Geriye kalan *"kez"* ve *"yaz"* stringlerinin uzunluğu zaten yok denilecek kadardır, onlar da ihmal edilebilir. Ohalde, *P ="n kez d yaz"* algoritmasının uzunluğunu, *s* dizisinin uzunluğu ile karşılaştırırken belirleyici olan tek etmenin *d* dizisinin uzunluğu (bit sayısı) olduğu sonucuna varırız.

Buradan yola çıkarak *n* bit uzunluğundaki dizilerin algoritmik karmaşıklıklarını *n-1, n-10, n-100, n-1000, ...* gibi sınıflara ayırabiliriz. Artık *P* nin uzunluğu ile *d* nin uzunluğunu yaklaşık eşit sayarak, aşağıdaki inductive yöntemi uygulayabiliriz.

Uzunluğu *1* olmak üzere *n* bitlik dizi ileten kaç tane algoritma vardır? *"n kez 0 yaz"* ve *"n kez 1 yaz"* algoritmaları bu işi yapan iki algoritmadır. Birincisi *000...0* dizisini, ikincisi ise *111...1* dizisini iletir. Bu algoritmaların ilkinde *d* dizisi yalnızca *'0'* dan, ikincisinde ise yalnızca *'1'* den ibarettir. Her ikisinin de uzunluğu *1* bittir. Bir bitlik başka algoritma yoktur. 1 bitlik algoritmaların sayısını 2^1 biçiminde gösterebiliriz. Benzer olarak, 0 ile 1 sayaklarından elde edilecek 2 bitlik dizilerin sayısı 4 dür: 00, 10, 11, 10. Ohalde, İki bitlik algoritmaların sayısı 2^2

dir. Benzer düşünüşle, üç bitlik algoritmaların sayısı 2^3 , ... , $n-11$ bitlik algoritmaların sayısı 2^{2-11} olacaktır. Bunların toplamı $(2^1 + 2^2 + 2^3 + ... 2^{n-11}) = 2^{n-10} - 2$ dir. Demek ki, uzunluğu $n-10$ dan az olan algoritmaların sayısı 2^{n-10} dan daha azdır. Öte yandan n bit uzunluğundaki dizilerin sayısı 2^n dir. Görüldüğü gibi, bunlar arasında ancak 2^{n-10} tanesinin algoritmik karmaşıklığı $n-10$ dan küçüktür. $2^{n-10} / 2^n = 1 / 1024$ olduğuna göre, 1024 diziden ancak 1 tanesinin algoritmik karmaşası $n-10$ dan küçüktür. Bundan anlaşılıyor ki seçkili sayılar çok seyrektir, sayıların çoğunluğu seçkisizdir.

Yukarıda yaptıklarımızdan şu sonuç çıkmaktadır: Bir dizi verildiğinde onun seçkili olduğunu göstermek için, diziyi ileten ve diziden daha kısa olan bir algoritma olduğunu göstermek yetecektir. Bulunacak bu algoritmanın minimal olması gerekmiyor. Ama bir dizinin seçkisiz olduğunu göstermek için onu ileten daha kısa bir algoritmanın var olmadığını göstermek gerekir.



10.5. Algoritmik Olasılık Algoritmaları

1. Tahmin, Olasılık ve Tümevarım

Klasik olasılık kavramlarını revize etmek için güçlü bir motivasyon, insan problem çözme analizinden geldi. Zor bir problem üzerinde çalışırken, kişi olası hareket tarzlarını seçmek zorunda olduğu bir labirentin içindedir. Sorun tanıdık bir sorunsa, seçimler kolay olacaktır. Tanıdık değilse, her seçimde çok fazla belirsizlik olabilir, ancak seçimler bir şekilde yapılmalıdır. Seçimler için bir temel, hızlı bir çözüme yol açan her bir seçimin olasılığı olabilir - bu olasılık, bu problemdeki ve buna benzer problemlerdeki deneyime dayanmaktadır.

Olasılığı hesaplamanın genel yöntemi, geçmişteki olumlu seçeneklerin sayısının toplam seçim sayısına oranını almaktır.

İndüksiyon da var. Aslında tahmin genellikle endüktif modeller bularak yapılır. Bunlar, tahmin için deterministik veya olasılıksal kurallardır. Bize bir dizi veri veriliyor - tipik olarak bir dizi sıfır ve birler ve bizden önce gelen veri noktalarının bir fonksiyonu olarak veri noktalarından herhangi birini tahmin etmemiz isteniyor.

2. Sıkıştırma ve Algoritmik Olasılık

Sembol tahmininin önemli bir uygulaması metin sıkıştırmadır. Bir tümevarım algoritması bir metne S olasılığı atarsa, tüm metni hatasız olarak yeniden oluşturabilen bir kodlama yöntemi - Aritmetik Kodlama - vardır.

3. Hesaplanamazlık

Genel olarak, herhangi bir kesinlik ile gerçekten en iyi modelleri bulmanın imkansız olduğuna dikkat edilmelidir - test edilecek sonsuz sayıda model vardır ve bazılarının değerlendirilmesi kabul edilemez derecede uzun zaman alır. Aramada herhangi bir zamanda, şimdiye kadarki en iyilerini bileceğiz, ancak biraz daha fazla zaman harcamanın çok daha iyi modeller vermeyeceğinden asla emin olamayız! Sınırlı sayıda model kullanarak her zaman algoritmik olasılığa yaklaşımlar yapabileceğimiz açıkken, bu yaklaşımların “Gerçek algoritmik olasılığa” ne kadar yakın olduğunu asla bilemeyiz. algoritmik olasılık gerçekten de biçimsel olarak hesaplanamaz.

4. Öznellik

Olasılığın öznelliği, a priori bilgide bulunur - istatistikçinin tahmin edilecek verileri görmeden önce sahip olduğu bilgiler. Bu, ne tür istatistiksel teknikler kullandığımızdan bağımsızdır.

Tahminlerde bulunurken, a priori bilgi eklemek için yaygın olarak kullanılan birkaç teknik vardır. İlk olarak, dikkate alınacak tümevarım modellerini kısıtlayarak veya genişleterek. Bu kesinlikle en yaygın yoldur. İkinci olarak, ayarlanabilir parametrelere sahip tahmin

fonksiyonları seçilerek ve bu parametrelerle ilgili geçmiş deneyimlere dayalı olarak bu parametreler üzerinde bir yoğunluk dağılımı varsayılarak. Üçüncüsü, bilimlerimizdeki bilgilerin çoğunun tanımlar - dilimize eklemeler - olarak ifade edildiğini not ediyoruz. algoritmik olasılık veya yaklaşıklıkları, kod uzunluklarını ve dolayısıyla modellere a priori olasılıkları atamaya yardımcı olmak için bu tanımları kullanarak bu bilgilerden yararlanır. Bilgisayar dilleri genellikle modelleri tanımlamak için kullanılır ve dilin bir parçası olarak keyfi tanımlar yapmak nispeten kolaydır.

5. Çeşitlilik ve Anlayış

Algoritmik olasılığın, olasılık tahmininin doğruluğunun yanı sıra, AI'nın bir başka önemli değeri daha vardır: Modellerin çokluğu, verilerimizi anlamamız için bize birçok farklı yol sunar. Çok geleneksel bir bilim adamı, bilimini tek bir "mevcut paradigma" kullanarak anlar - şu anda en moda olan anlama yolu. Daha yaratıcı bir bilim insanı bilimini pek çok şekilde anlar ve “mevcut paradigma” artık mevcut verilere uymadığında yeni teoriler, yeni anlayış yolları daha kolay yaratabilir.

Algoritmik olasılığın uygulamaları nelerdir?

Algoritmik olasılığın bir dizi önemli teorik uygulaması vardır; Bunlardan bazıları aşağıda listelenmiştir.

Solomonoff İndüksiyon: Solomonoff (1964), evrensel Turing makinelerine (hesaplamak, nicelleştirmek ve ilgili tüm niceliklere kod atamak için) ve algoritmik karmaşıklığa (basitlik/karmaşıklığın ne anlama geldiğini tanımlamak için) dayanan nicel bir biçimsel tümevarım teorisi geliştirdi. Bu, Solomonoff'un tümevarımsal çıkarım sisteminin, yalnızca mutlak minimum miktarda veri ile herhangi bir hesaplanabilir diziyi doğru bir şekilde tahmin etmeyi öğreneceği anlamına geldiği için dikkat çekicidir. Bu nedenle, bir anlamda, sadece hesaplanabilir olsaydı, mükemmel evrensel tahmin algoritması olurdu.

AIXI ve Zekanın Evrensel Tanımı: M'nin daha genel stokastik ortamlarda mükemmel tahminlere ve kararlara yol açtığı da gösterilebilir. Esasen AIXI, Solomonoff indüksiyonunun pekiştirici öğrenme ortamına, yani ajanın eylemlerinin çevrenin durumunu etkileyebileceği bir genellemedir. AIXI'nin bir anlamda, rastgele bilinmeyen bir ortama gömülü optimal bir pekiştirmeli öğrenme aracı olduğu kanıtlanabilir. Solomonoff indüksiyonu gibi, AIXI ajanı hesaplanabilir değildir ve bu nedenle pratikte sadece yaklaşık olarak tahmin edilebilir.

Optimal bir genel ajan tanımlamak yerine, işleri tersine çevirmek ve evrensel zeka olarak bilinen hesaplanabilir ortamlarda hareket eden ajanlar için evrensel bir performans ölçüsü tanımlamak mümkündür.

Evrensel Dağılım Altında Algoritmaların Beklenen Zaman/Uzay Karmaşıklığı: Her algoritma için, m-ortalama durum karmaşıklığı, zaman ve depolama alanı gibi hesaplama kaynakları için her zaman en kötü durum karmaşıklığı ile aynı büyüklüktedir. Li ve Vitanyi'nin bu sonucu, eğer girdiler alışılageldiği gibi tekdüze dağılım yerine evrensel dağılıma göre dağıtılırsa, beklenen durumun en kötü durum kadar kötü olduğu anlamına gelir. Örneğin, sıralama prosedürü Quicksort'un en kötü durum çalışma süresi n2'dir, ancak n anahtarın düzgün dağılmış permütasyonlarından rastgele çizilen permütasyonlar için n anahtarlık bir listede beklenen çalışma süresi nlogn'dur. Ancak permütasyonlar evrensel dağılıma göre dağıtılırsa, yani Occam'ın, pratikte sıklıkla olduğu gibi basit permütasyonların en yüksek olasılıklara sahip olduğu şeklindeki sözüne göre, beklenen çalışma süresi en kötü n2 durumuna yükselir. Deneyler bu teorik öngörüyü doğruluyor gibi görünüyor.

Öğrenme Aşamasında Evrensel Dağıtımı Kullanan PAC Öğrenimi: PAC-öğrenmede öğrenme aşamasında örnekleme dağılımı olarak m'nin kullanılması, hesaplanabilir dağılımlar altındaki ayrık kavram sınıfları ailesi için öğrenme gücünü ve benzer şekilde M kullanarak hesaplanabilir ölçümler üzerinden sürekli kavramların PAC öğrenmesi için öğrenme gücünü arttırır. Bu Li ve Vitanyi modeli, ilk önce basit örnekleri veren bir öğretmenin öğrenme aşamasında kullanılmasına benzer.

Durdurma Olasılığı: Biçimsel olarak ilişkili bir nicelik, giriş bandında adil yazı turaları sağlandığında U'nun durma olasılığıdır (yani, rastgele bir bilgisayar programının sonunda duracağı). Bu durma olasılığı $\Omega=\sum xm(x)$ aynı zamanda Chaitin sabiti veya "bilgelik sayısı" olarak da bilinir ve çok sayıda dikkate değer matematiksel özelliğe sahiptir. Örneğin, Goedel'in Eksiklik Teoremini ölçmek için kullanılabilir.

10.6. Kolmogorov-Smirnov Goodness-of-Fit Test

We'll learn how to conduct a test to see how well a hypothesized distribution function F(x) fits an empirical distribution function Fn(x). The "goodness-of-fit test" that we'll learn about was developed by two probabilists, Andrey Kolmogorov and Vladimir Smirnov, and hence the name of this lesson. In the process of learning about the test, we'll:

- learn a formal definition of an empirical distribution function
- justify, but not derive, the Kolmogorov-Smirnov test statistic
- try out the test on a few examples
- learn how to calculate a confidence band for a distribution function F(x)

10.6.1. The Test

Before we can work on developing a hypothesis test for testing whether an empirical distribution function Fn(x) fits a hypothesized distribution function F(x) we better have a good idea of just what is an empirical distribution function Fn(x). Therefore, let's start with formally defining it.

Empirical distribution function

Given an observed random sample X1,X2,…,Xn, an **empirical distribution function** Fn(x) is the fraction of sample observations less than or equal to the value *x*. More specifically, if y1<y2<⋯<yn are the order statistics of the observed random sample, with no two observations being equal, then the empirical distribution function is defined as:

$$F_n(x) = \begin{cases} 0, & \text{for } x < y_1 \\ k/n, & \text{for } y_k \le x < y_{k+1}, \quad k = 1, 2, \ldots, n-1 \\ 1, & \text{for } x \ge y_n \end{cases}$$

That is, for the case in which no two observations are equal, the empirical distribution function is a "step" function that jumps 1/n in height at each observation xk. For the cases in which two (or more) observations are equal, that is, when there are nk observations at xk, the empirical distribution function is a "step" function that jumps nk/n in height at each observation xk. In either case, the empirical distribution function Fn(x) is the fraction of sample values that are equal to or less than *x*.

Such a formal definition is all well and good, but it would probably make even more sense if we took at a look at a simple example.

Example-1

A random sample of *n* = 8 people yields the following (ordered) counts of the number of times they swam in the past month:

0 1 2 2 4 6 6 7

Calculate the empirical distribution function Fn(x).

Answer

As reported, the data are ordered, therefore the order statistics

are y1=0,y2=1,y3=2,y4=2,y5=4,y6=6,y7=6 and y8=7. Therefore, using the definition of the empirical distribution function, we have:

Fn(x)=0 for x<0

and:

Fn(x)=18 for 0≤x<1 and Fn(x)=28 for 1≤x<2

Now, noting that there are two 2s, we need to jump 2/8 at *x* = 2:

Fn(x)=28+28=48 for 2≤x<4

Then:

Fn(x)=58 for 4≤x<6

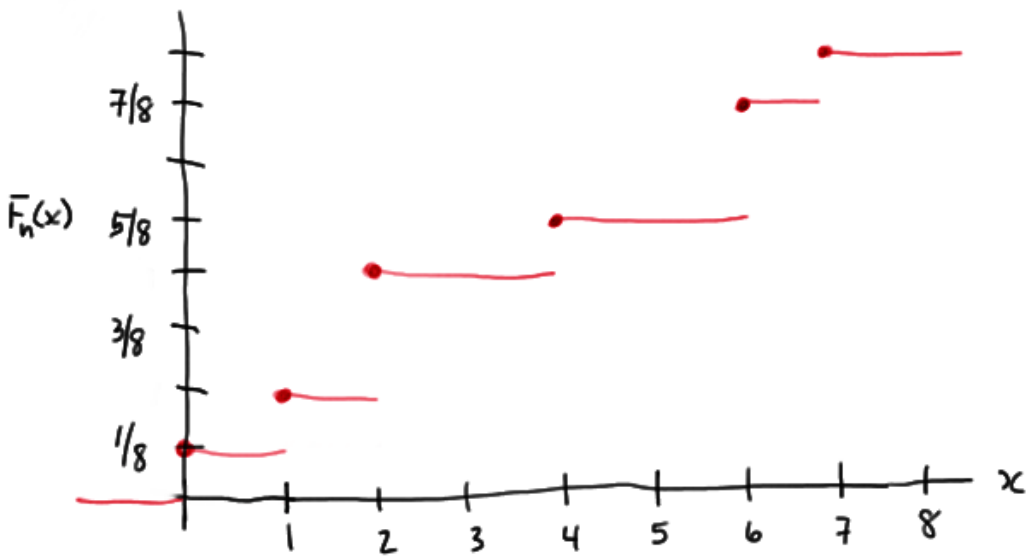
Again, noting that there are two 6s, we need to jump 2/8 at *x* = 6:

Fn(x)=58+28=78 for 6≤x<7

And, finally:

Fn(x)=78+18=88=1 for x≥7

Plotting the function, it should look something like this then:



Now, with that behind us, let's jump right in and state and justify (not prove!) the Kolmogorov-Smirnov statistic for testing whether an empirical distribution fits a hypothesized distribution well.

Kolmogorov-Smirnov test statistic

Dn=supx[|Fn(x)−F0(x)|]

is used for testing the null hypothesis that the cumulative distribution function F(x) equals some hypothesized distribution function F0(x), that is, H0:F(x)=F0(x), against all of the possible alternative hypotheses HA:F(x)≠F0(x). That is, Dn is the least upper bound of all pointwise differences |Fn(x)−F0(x)|.

Justification

The bottom line is that the Kolmogorov-Smirnov statistic makes sense, because as the sample size *n* approaches infinity, the empirical distribution function Fn(x) converges, with probability 1 and uniformly in *x*, to the theoretical distribution function F(x). Therefore, if there is, at any point *x*, a large difference between the empirical distribution Fn(x) and the hypothesized distribution F0(x), it would suggest that the empirical distribution Fn(x) does not equal the hypothesized distribution F0(x). Therefore, we reject the null hypothesis:

H0:F(x)=F0(x)

if Dn is too large.

Now, how do we know that Fn(x) converges, with probability 1 and uniformly in *x*, to the theoretical distribution function F(x)? Well, unfortunately, we don't have the tools in this course to officially prove it, but we can at least do a bit of a hand-waving argument.

Let X1,X2,…,Xn be a random sample of size *n* from a continuous distribution F(x). Then, if we consider a fixed *x*, then W=Fn(x) can be thought of as a random variable that takes on possible values 0,1/n,2/n,…,1. Now:

- *nW* = 1, if and only if exactly 1 observation is less than or equal to *x,* and *n*−1 observations are greater than *x*
- *nW* = 2, if and only if exactly 2 observations are less than or equal to *x,* and *n*−2 observations are greater than *x*
- and in general...
- *nW* = *k*, if and only if exactly *k* observations are less than or equal to *x,* and *n*−*k* observations are greater than *x*

If we treat a success as an observation being less than or equal to *x*, then the probability of success is:

P(Xi≤x)=F(x)

Do you see where this is going? Well, because X1,X2,…,Xn are independent random variables, the random variable *nW* is a binomial random variable with *n* trials and probability of success *p = F*(x). Therefore:

P(W=kn)=P(nW=k)=(nk)[F(x)]k[1−F(x)]n−k

And, the expected value and variance of *nW* are:

E(nW)=np=nF(x) and Var(nW)=np(1−p)=n[F(x)][1−F(x)]

respectively. Therefore, the expected value and variance of *W* are:

E(W)=nF(x)n=F(x) and Var(W)=n[(F(x)][1−F(x)]n2=[(F(x)][1−F(x)]n

We're very close now. We just now need to recognize that as *n* approaches infinity, the variance of *W*, that is, the variance of Fn(x) approaches 0. That means that as *n* approaches infinity the empirical distribution Fn(x) approaches its mean F(x). And, that's why the argument for rejecting the null hypothesis if there is, at any point *x*, a large difference

between the empirical distribution Fn(x) and the hypothesized distribution F0(x). Not a mathematically rigorous argument, but an argument nonetheless!

Notice that the Kolmogorov-Smirnov (KS) test statistic is the supremum over *all* real x---a very large set of numbers! How then can we possibly hope to compute it? Well, fortunately, we don't have to check it at every real number but only at the sample values, since they are the only points at which the supremum can occur. Here's why:

First the easy case. If $x \geq y_n$, then $F_n(x)=1$, and the largest difference between $F_n(x)$ and $F_0(x)$ occurs at y_n. Why? Because $F_0(x)$ can never exceed 1 and will only get closer for larger x by the monotonicity of distribution functions. So, we can record the value $F_n(y_n)-F_0(y_n)=1-F_0(y_n)$ and safely know that no other value $x \geq y_n$ needs to be checked.

The case where $x<y_1$ is a little trickier. Here, $F_n(x)=0$, and the largest difference between $F_n(x)$ and $F_0(x)$ would occur at the largest possible x in this range for a reason similar to that above: $F_0(x)$ can never be negative and only gets farther from 0 for larger x. The trick is that there is no largest x in this range (since x is strictly less than y_1), and we instead have to consider lefthand limits. Since $F_0(x)$ is continuous, its limit at y_1 is simply $F_0(y_1)$. However, the lefthand limit of $F_n(y_1)$ is 0. So, the value we record is $F_0(y_1)-0=F_0(y_1)$, and ignore checking any other value $x<y_1$.

Finally, the general case $y_{k-1} \leq x < y_k$ is a combination of the two above. If $F_0(x)<F_n(x)$, then $F_0(y_{k-1}) \leq F_0(x) < F_n(x) = F_n(y_{k-1})$, so that $F_n(y_{k-1})-F_0(y_{k-1})$ is at least as large as $F_n(x)-F_0(x)$ (so we don't even have to check those x values). If, however, $F_0(x)>F_n(x)$, then the largest difference will occur at the lefthand limits at y_k. Again, the continuity of F_0 allows us to use $F_0(y_k)$ here, while the lefthand limit of $F_n(y_k)$ is actually $F_n(y_{k-1})$. So, the value to record is $F_0(y_k)-F_n(y_{k-1})$, and we may disregard the other x values.

Whew! That covers all real x values and leaves us a much smaller set of values to actually check. In fact, if we introduce a value y_0 such that $F_n(y_0)=0$, then we can summarize all this exposition with the following rule:

Rule for computing the KS test statistic:

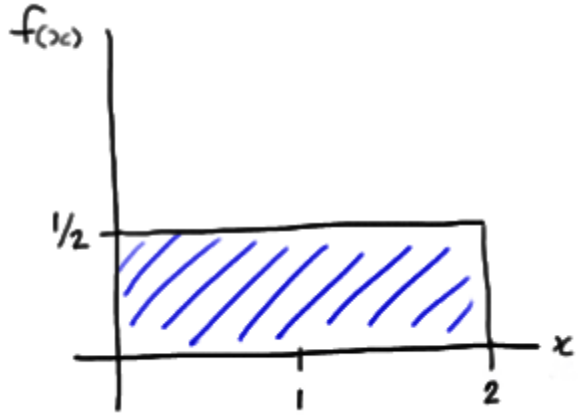
For each ordered observation y_k compute the differences

$|F_n(y_k)-F_0(y_k)|$ and $|F_n(y_{k-1})-F_0(y_k)|$.

The largest of these is the KS test statistic.

The easiest way to manage these calculations is with a table, which we now demonstrate with two examples.

Example -2



We observe the following *n* = 8 data points:

1.41 0.26 1.97 0.33 0.55 0.77 1.46 1.18

Is there any evidence to suggest that the data were not randomly sampled from a Uniform(0, 2) distribution?

Answer

The probability density function of a Uniform(0, 2) random variable *X*, say, is:

f(x)=12

for 0 < *x* < 2. Therefore, the probability that *X* is less than or equal to *x* is:

P(X≤x)=∫0x12dt=12x

for 0 < *x* < 2, and we are interested in testing:

- the null hypothesis H0:F(x)=F0(x) against
- the alternative hypothesis HA:F(x)≠F0(x)

where *F(x)* is the (unknown) cdf from which our data were sampled, and

$$F_0(x) = \begin{cases} 0, & \text{for } x < 0 \\ \frac{1}{2}x, & \text{for } 0 \le x < 2 \\ 1, & \text{for } x \ge 2 \end{cases}$$

Now, in working towards calculating dn, we first need to order the eight data points so that y1≤⋯≤y8. The table below provides all the necessary values for finding the KS test statistic. Note that the empirical cdf satisfies Fn(yk)=k/8, for k=0,1,…,8.

k	y_k	$F_n(y_{k-1})$	$F_0(y_k)$	$F_n(y_k)$	$\lvert F_n(y_{k-1}) - F_0(y_k)\rvert$	$\lvert F_0(y_k) - F_n(y_k)\rvert$
1	0.26	0.000	0.130	0.125	0.130	0.005
2	0.33	0.125	0.165	0.250	0.040	0.085
3	0.55	0.250	0.275	0.375	0.025	0.100
4	0.77	0.375	0.385	0.500	0.010	0.115
5	1.18	0.500	0.590	0.625	0.090	0.035
6	1.41	0.625	0.705	0.750	0.080	0.045
7	1.46	0.750	0.730	0.875	0.020	**0.145**
8	1.97	0.875	0.985	1.000	0.090	0.015

The last two columns represent all the differences we need to check. The largest of these is d8=0.145. From the table below with α=0.05, the critical value is 0.46. So, we can not reject the claim that the data were sampled from Uniform(0,2).

$$D_n = \sup_x[|F_n(x) - F_0(x)|]$$
$$\alpha = 1 - P(D_n \leq d)$$

	α			
n	0.20	0.10	0.05	0.01
1	0.90	0.95	0.98	0.99
2	0.68	0.78	0.84	0.93
3	0.56	0.64	0.71	0.83
4	0.49	0.56	0.62	0.73
5	0.45	0.51	0.56	0.67
6	0.41	0.47	0.52	0.62
7	0.38	0.44	0.49	0.58
8	0.36	0.41	0.46	0.54
9	0.34	0.39	0.43	0.51
10	0.32	0.37	0.41	0.49

You might recall that the appropriateness of the *t*-statistic for testing the value of a population mean μ depends on the data being normally distributed. Therefore, one of the most common applications of the Kolmogorov-Smirnov test is to see if a set of data does follow a normal distribution. Let's take a look at an example.

Example-3

Each person in a random sample of *n* = 10 employees was asked about *X*, the daily time wasted at work doing non-work activities, such as surfing the internet and emailing friends. The resulting data, in minutes, are as follows:

108 112 117 130 111 131 113 113 105 128

Is it okay to assume that these data come from a normal distribution with mean 120 and standard deviation 10?

Answer

We are interested in testing the null hypothesis, H0: X is normally distributed with mean 120 and standard deviation 10, against the alternative hypothesis, HA: X is not normally distributed with mean 120 and standard deviation 10. Now, in working towards calculating dn, we again first need to order the ten data points so that y1=105, y2=108, etc. Then, we need to calculate the value of the hypothesized distribution function F0(yk) at

each of the values of yk. The standard normal table can help us do this. The probability that X is less than or equal to 105, for example, equals the probability that Z is less than or equal to −1.5:

F0(y1)=P(X≤105)=P(Z≤105−12010)=P(Z≤−1.5)=.0668.

The table below summarizes the relevant quantities for finding the KS test statistic. Note that the empirical cdf satisfies Fn(yk)=k/10, except at k=5 because of the tie: y5=y6=113.

k	y_k	$F_n(y_{k-1})$	$F_0(y_k)$	$F_n(y_k)$	$\lvert F_n(y_{k-1}) - F_0(y_k)\rvert$	$\lvert F_0(y_k) - F_n(y_k)\rvert$
1	105	0.0	0.0668	0.1	0.0668	0.0332
2	108	0.1	0.1151	0.2	0.0151	0.0849
3	111	0.2	0.1841	0.3	0.0159	0.1159
4	112	0.3	0.2119	0.4	0.0881	0.1881
5	113	0.4	0.2420	0.6	0.1580	**0.3580**
6	113	0.6	0.2420	0.6	**0.3580**	**0.3580**
7	117	0.6	0.3821	0.7	0.2179	0.3179
8	128	0.7	0.7881	0.8	0.0881	0.0119
9	130	0.8	0.8413	0.9	0.0413	0.0587
10	131	0.9	0.8643	1.0	0.0357	0.1357

The last two columns represent all the differences we need to check. The largest of these is 0.3580. From the table below with α=0.10, the critical value is 0.37. Because d10=0.358, which is just barely less than 0.37, we do not reject the null hypothesis at the 0.10 level. There is not enough evidence at the 0.10 level to reject the assumption that the data were randomly sampled from a normal distribution with mean 120 and standard deviation 10.

$$D_n = \sup_x[|F_n(x) - F_0(x)|]$$
$$\alpha = 1 - P(D_n \le d)$$

	α			
n	0.20	0.10	0.05	0.01
1	0.90	0.95	0.98	0.99
2	0.68	0.78	0.84	0.93
3	0.56	0.64	0.71	0.83
4	0.49	0.56	0.62	0.73
5	0.45	0.51	0.56	0.67
6	0.41	0.47	0.52	0.62
7	0.38	0.44	0.49	0.58
8	0.36	0.41	0.46	0.54
9	0.34	0.39	0.43	0.51
10	0.32	0.37	0.41	0.49

A Confidence Band

Another application of the Kolmogorov-Smirnov statistic is in forming a confidence band for an unknown distribution function F(x). To form a confidence band for F(x), we basically need to find a confidence interval for each value of *x*. The following theorem gives us the recipe for doing just that.

Theorem

A 100(1−α)% **confidence band for the unknown distribution function** F(x) is given by FL(x) and FU(x) where *d* is selected so that P(Dn≥d)=α and:

$$F_L(x) = \begin{cases} 0, & \text{if } F_n(x) - d \leq 0 \\ F_n(x) - d, & \text{if } F_n(x) - d > 0 \end{cases}$$

and:

$$F_u(x) = \begin{cases} F_n(x) + d, & \text{if } F_n(x) + d < 1 \\ 1, & \text{if } F_n(x) + d \geq 1 \end{cases}$$

Proof

We select *d* so that:

$P(D_n \geq d) = \alpha$

Therefore, using the definition of Dn, and the probability rule of complementary events, we get:

$P(\sup x |F_n(x) - F(x)| \leq d) = 1 - \alpha$

Now, if the largest of the absolute values of $F_n(x) - F(x)$ is less than or equal to *d*, then all of the absolute values of $F_n(x) - F(x)$ must be less than or equal to *d*. That is:

$P(|F_n(x) - F(x)| \leq d \text{ for all } x) = 1 - \alpha$

Rewriting the quantity inside the parentheses without the absolute value, we get:

$P(-d \leq F_n(x) - F(x) \leq d \text{ for all } x) = 1 - \alpha$

And, subtracting $F_n(x)$ from each part of the resulting inequality, we get:

$P(-F_n(x) - d \leq -F(x) \leq -F_n(x) + d \text{ for all } x) = 1 - \alpha$

Now, when we divide through by −1, we have to switch the order of the inequality, getting:

$P(F_n(x) - d \leq F(x) \leq F_n(x) + d \text{ for all } x) = 1 - \alpha$

We could stop there and claim that: $F_n(x) - d \leq F(x) \leq F_n(x) + d$ for all x is a $100(1-\alpha)\%$ confidence band for the unknown distribution function F(x). There's only one problem with that. It is possible that the lower limit is less than 0 and it is possible that the upper limit is greater than 1. That's not a good thing given that a distribution function must be sandwiched between 0 and 1, inclusive. We take care of that by rewriting the lower limit to prevent it from being negative:

$$F_L(x) = \begin{cases} 0, & \text{if } F_n(x) - d \leq 0 \\ F_n(x) - d, & \text{if } F_n(x) - d > 0 \end{cases}$$

and by rewriting the upper limit to prevent it from being greater than 1:

$$F_u(x) = \begin{cases} F_n(x) + d, & \text{if } F_n(x) + d < 1 \\ 1, & \text{if } F_n(x) + d \geq 1 \end{cases}$$

As was to be proved!
Let's try it out an example.

Example-4

Each person in a random sample of *n* = 10 employees was asked about *X*, the daily time wasted at work doing non-work activities, such as surfing the internet and emailing friends. The resulting data, in minutes, are as follows:

108 112 117 130 111 131 113 113 105 128

Use the data to find a 95% confidence band for the unknown cumulative distribution function F(x).

Answer

As before, we start by ordering the *x* values. The formulas for the lower and upper confidence limits tell us that we need to know *d* and Fn(x) for each of the 10 data points. Because the α-level is 0.05 and the sample size *n* is 10, the table of Kolmogorv-Smirnov Acceptance Limits in the back of our text book, that is, Table VIII, tells us that d=0.41. We already calculated Fn(x) in the previous example.

Now, in calculating the lower limit, FL(x)=Fn(x)−d=Fn(x)−0.41, we see that the lower limit would be negative for the first four data points:

0.1−0.41 = −0.31 *and* 0.2−0.41 = −0.21 *and* 0.3−0.41 = −0.11 *and* 0.4−0.41 = −0.01

Therefore, we assign the first four data points a lower limit of 0, and then just calculate the remaining lower limit values. Similarly, in calculating the upper
limit, FU(x)=Fn(x)+d=Fn(x)+0.41, we see that the upper limit would be greater than 1 for the last six data points:

0.6+0.41 = 1.01 *and* 0.7+0.41 = 1.11 *and* 0.8+0.41 = 1.21 *and* 0.9+0.41 = 1.31 *and* 1.0+0.41 = 1.41

Therefore, we assign the last six data points an upper limit of 1, and then just calculate the remaining upper limit values. To summarize,

k	y_k	$F_n(y_k)$	$F_L(y_k)$	$F_U(y_k)$
1	105	0.1	0	0.51
2	108	0.2	0	0.61
3	111	0.3	0	0.71
4	112	0.4	0	0.81
5	113	0.6	0.19	1
6	113	0.6	0.19	1
7	117	0.7	0.29	1
8	128	0.8	0.39	1
9	130	0.9	0.49	1
10	131	1.0	0.59	1

The last two columns together give us the 95% confidence band for the unknown cumulative distribution function F(x).

11. Kaynaklar

1. https://tr.wikipedia.org
2. https://www.wiley.com/legacy/Australia/Landing_Pages/c12ContinuousProbabilityDistributions_web.pdf
3. Olasılık ve İstatistik, Aydın Üstün, 2014.
4. İslamoğlu, A.H.(2009). Sosyal Bilimlerde Araştırma Yöntemleri (SPSS Uygulamalı), Beta Basım Yayım Dağıtım A.Ş., İstanbul.
5. Tütek, H., Gümüşoğlu, Ş., 2008, İ letme İstatistiği, Beta Basım Yayım Dağıtım A.Ş., İstanbul.
6. Olasılığın Matematiksel Temelleri , Timur KARAÇAY, Başkent Üniversitesi, Fen-Edebiyat Fakültesi
7. Olasılığın Temelleri, Timur Karaçay, Başkent Üniversitesi. Mantık, Matematik ve Felsefe IV.Ulusal Sempozyumu Foça, 5-8 Eylül 2006.
8. https://www.probabilitycourse.com/chapter11/11_2_7_solved_probs.php
9. http://en.wikipedia.org/wiki/Table_of_mathematical_symbols
10. https://cdn-acikogretim.istanbul.edu.tr/
11. Yılmaz Özkan, Uygulamalı İstatistik 1, Sakarya Kitapevi, 2008.
12. Özer Serper, Uygulamalı İstatistik 1, Filiz Kitapevi, 1996.
13. Meriç Öztürkcan, İstatistik Ders notları, YTÜ.
14. https://tr.wikipedia.org/wiki/Hipotez_testi