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: $<,>,=, \neq, \leq, \geq$
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

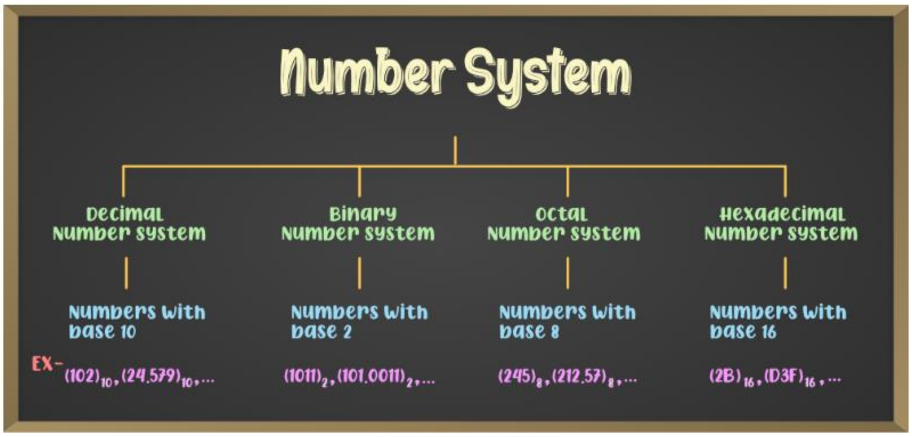
Data types, classes, aggregation, and equationalization are the building blocks of programming. They tell the computer how to interpret and store data. The most common data types are integers, strings, booleans, dates, arrays, and objects. Integers are whole numbers, strings are text, booleans can be 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

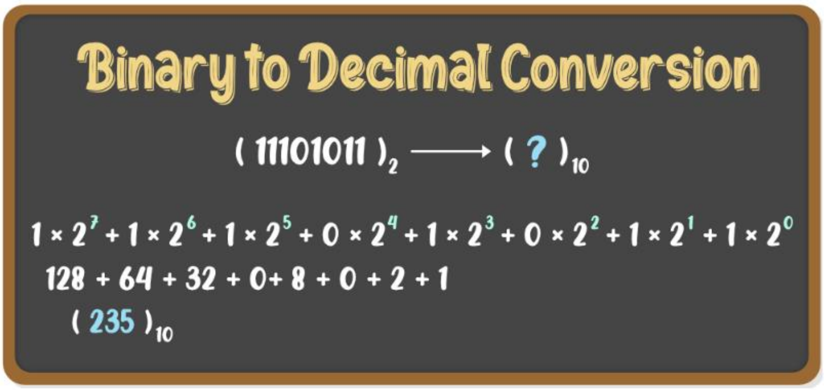
Hexadecimal Number System

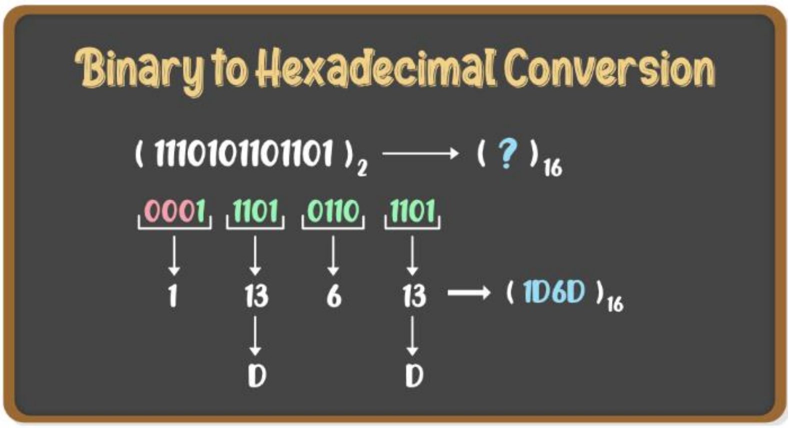
Number System with base value 16 is termed as Hexadecimal Number System. It uses 16 digits for the creation of its numbers. Digits from 0-9 are taken like the digits in the decimal number system but the digits from 10-15 are represented as A-F i.e. 10 is represented as A, 11 as B, 12 as C, 13 as D, 14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

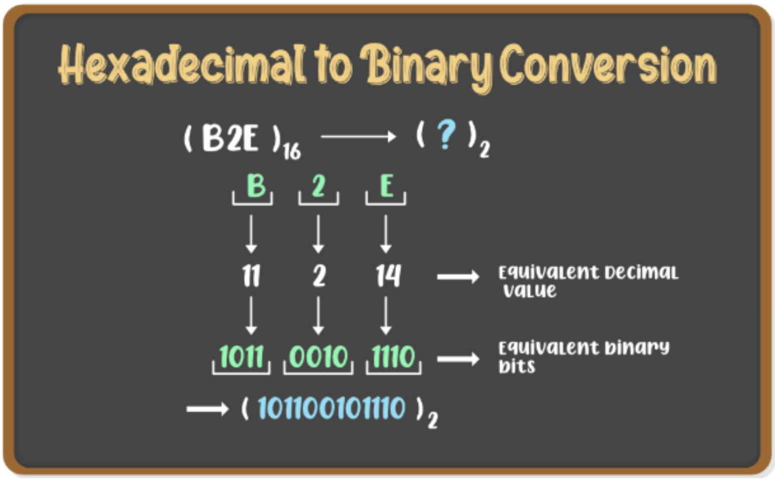
Examples:
(255)10 can be written as (FF)16
(1096)10 can be written as (448)16
(4090)10 can be written as (FFA)16

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

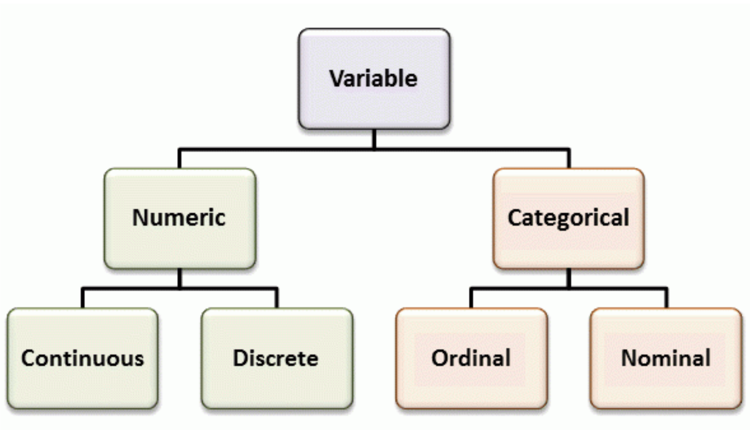
Example: Calculate area of circle
area = 3.142 * r * r
Variable: area, r
Constant: 3.142

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 $y(t)=A\sin(wt+\phi)$, $w=2\pi f$; 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 + . . .

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.
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.

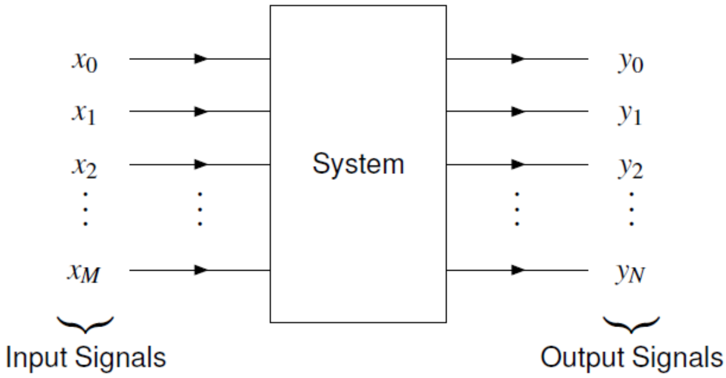
Mathematically, signals are represented as functions of one or more independent variables. A function f is called a sequence (vector) of variables (ti) in which each {ti, i=1,2, ... ,n} is sampled at independent discrete times.

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. $X(t)=A\sin(wt+ø)$. 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.

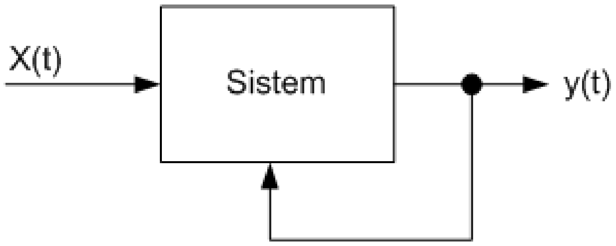
Digital (Numerical) signals are the representation of signals that are discrete in terms of both time and amplitude in the binary number system. (bit:0/1)

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.

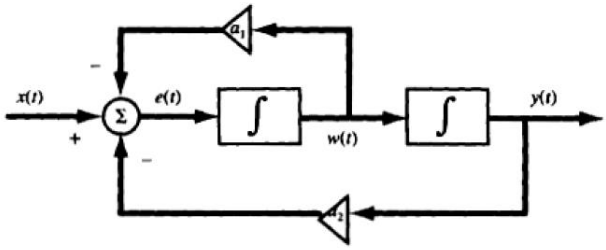
An analog signal generally consists of the sum of sinusoidal signals or their combination with certain arithmetic operations. A sinusoidal signal, X(t)=Asin(wt+ø)
W=2*pi*f, radial frequency
Ø: phase
A: Amplitude
F: frequency, the number of vibrations or periods in one second. Signals are transmitted 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:



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

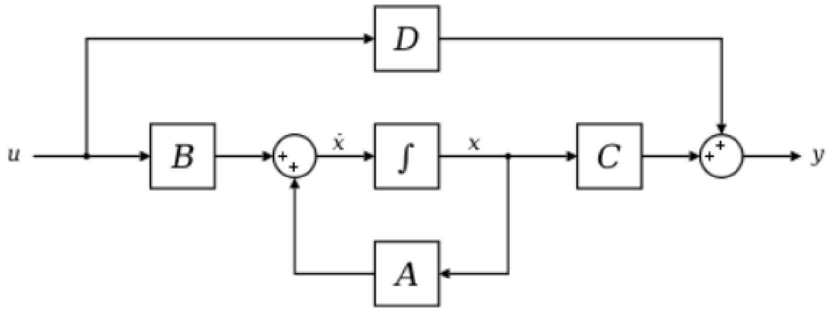
$$w(t)) = \frac{dy(t)}{dt}$$

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

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

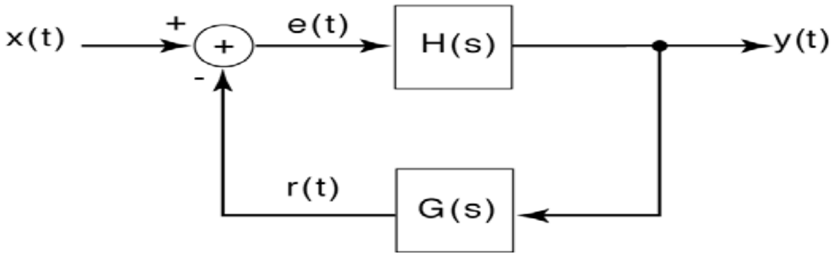
The relationship between integral and derivative plays a critical role in the mathematical modeling of systems. In a system that integrates the input signal, if the output is known, the input is obtained by taking the derivative of the signal. The converse is also true; if the output of a system that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



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

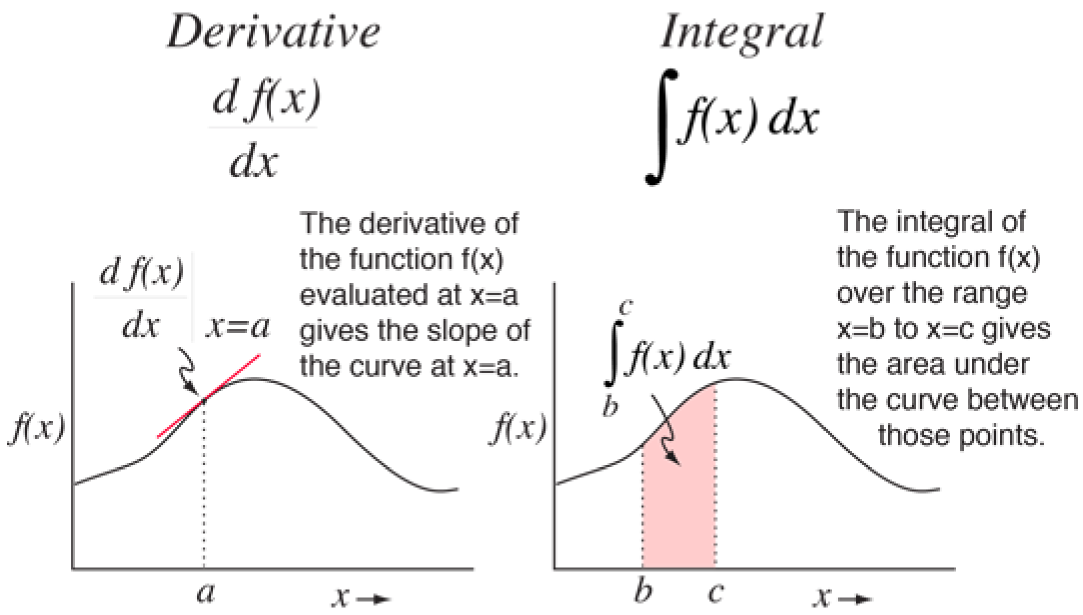
The transfer function of this system:



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

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.

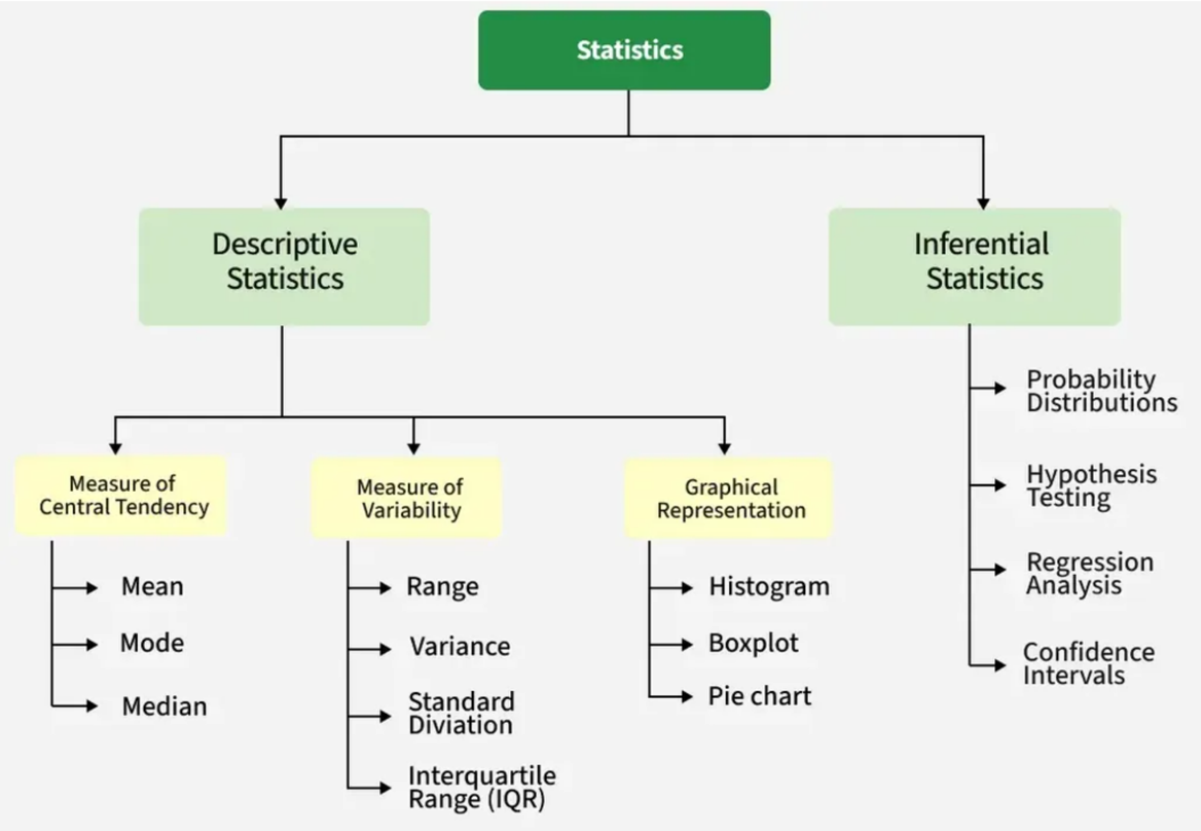
When a hypothesis is established, in order to be able to use an estimation, it is necessary to know how reliable this estimation is. On the other hand, it is also necessary to know what types of errors are encountered.

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.

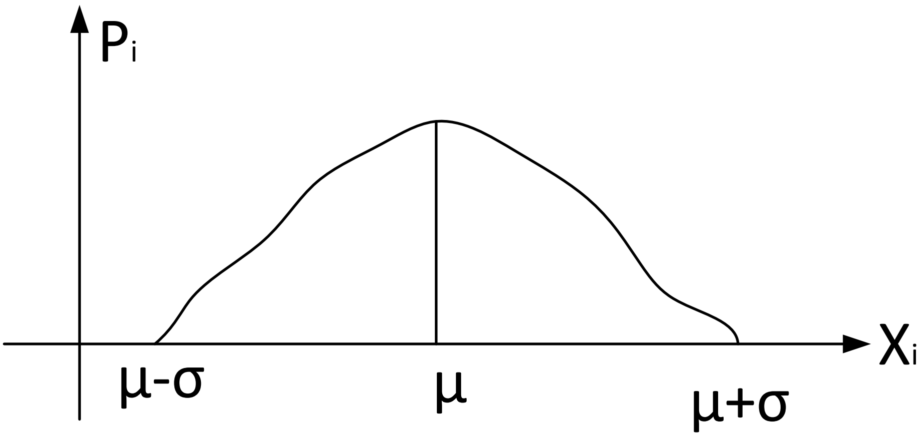
Step 3: Here, n = 8 (even) then,
Median = [(n/2)th term + {(n/2) + 1)th term] / 2
Median = [(8/2)th term + {(8/2) + 1}th term] / 2
= (22+23) / 2
= 22.5

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.

	Population	Sample
Variance	$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$	$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$
Standard deviation	$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$	$S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$

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

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

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

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

σ^2 = 2.917

Now, the standard deviation,σ = √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: μ1 = μ2 = μ3 = ... = μk

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: μ1 = μ2 = μ3*
- *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
x̄ = 16
SST = Σn(x−x̄)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 and B)= p(A)*P(B)
6) The probability of two dependent events occurring together,
   p(A and B)=p(A)*p(B |A)=p(B)*p(A |B);
   p(B |A): Probability of event B occurring after event A,
   p(A |B): Probability of event A occurring after event B.
7) Probability of one of two independent events occurring,
   p(A or B)=$p(A \cup B)=p(A)+p(B)$
8) Probability of one of two dependent events occurring,
   p(A 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?
   $P(W \text{ or } S)=P(B) + P(S) - P(B \cap S)$
   $=14/55 + 20/55 - 7/55 =27/55=0,49$ (%49)

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

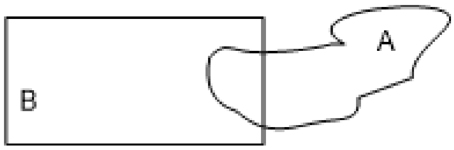
4.4.1. Conditional Probability

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,
$P(B|A)P(A) \quad = P(A|B)P(B)$ 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

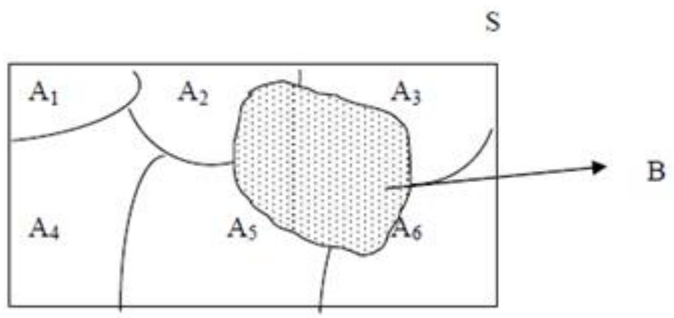
When the probability of a B event cannot be calculated directly, the total probability rule is used. For example, when a random product is taken from the 6 machines in a factory, the probability of this product being defective is investigated. Here, if the B event is the defective product, this product 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 $\cap Aj = 0, i \neq j$



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

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

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 piles:
- Clustered distribution
- Regular (Uniform) distribution
- Random distribution

Clustered distribution: This is the most common distribution in piles. It is the clustering of the elements forming the pile 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 pile 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 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 continuous; y(t) is the dependent variable; a, b, c are the coefficients and 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.

yi(t)=2at+b; this derivative gives the trajectory of the main function.
yiI(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.

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.

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

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}.

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:

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

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

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

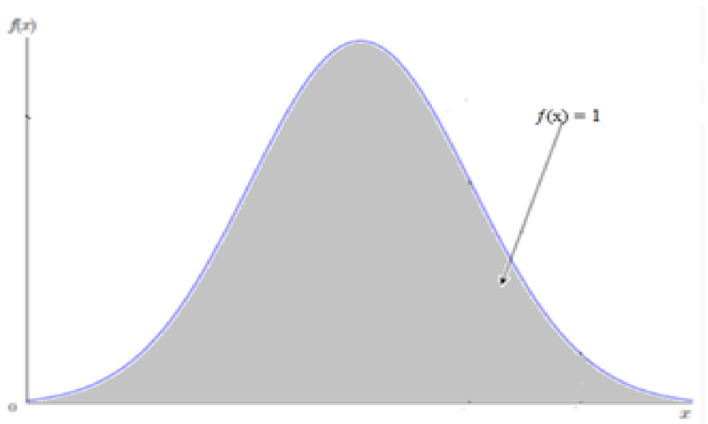
5.1. Probability Density Function in Continuous Random Variables

If X is a continuous random variable; X Є {-∞,∞}, 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".

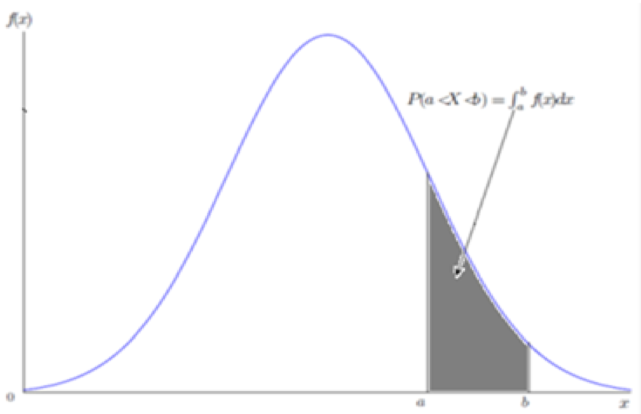
The "Probability Density Function" has the following properties:
The probability density function is denoted by ƒ(x). For all values of x, 1≥ ƒ(x) ≥ 0 (the curve does not intersect the horizontal axis).

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)?

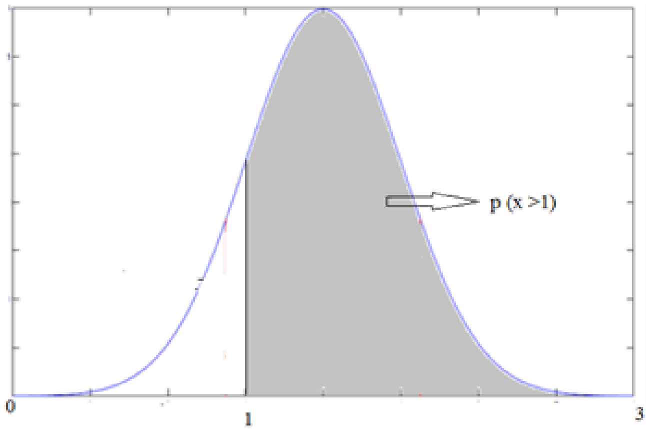
$$F(x) = \int_0^x f(x)dx = \int_0^x \frac{3}{27}x^2 dx = \ = \frac{3}{27}\left(\frac{\mathrm{x}^3}{3}\right)|_0^{\mathrm{x}} = \frac{3}{27}\left(\frac{\mathrm{x}^3}{3}\right) = \frac{x^3}{27}$$

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)?

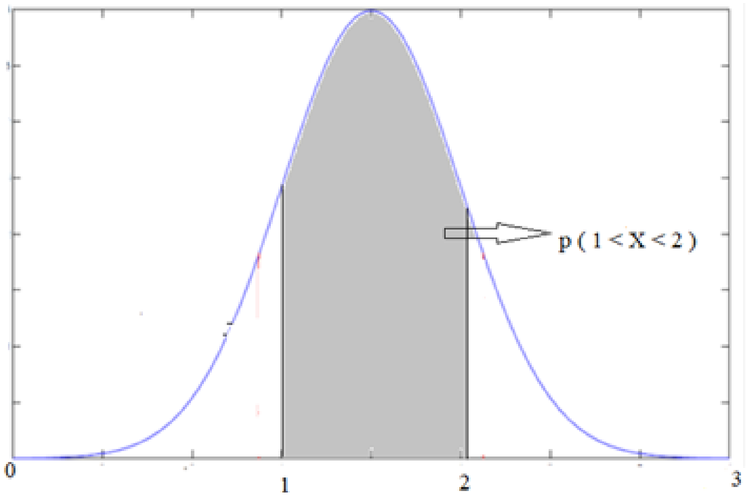


$$p\ (x\ \ge 1) = \int_1^3 \frac{3}{27}x^2 dx\ = \frac{3}{27}\left(\frac{\mathrm{x}^3}{3}\right)|_1^3 = \frac{3}{27}\left(\frac{3^3}{3}\right) - \frac{3}{27}\left(\frac{1^3}{3}\right)$$

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

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

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



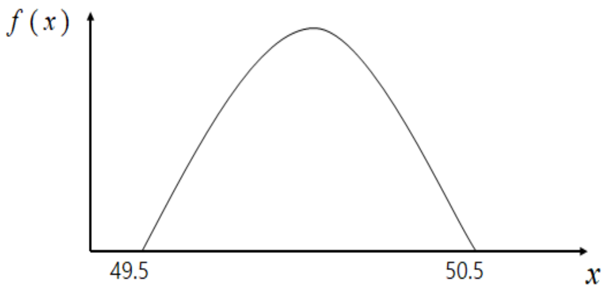
$$p\,(1 \le x \le 2) = \int_1^2 \frac{3}{27} x^2 dx = \frac{3}{27}\left(\frac{x^3}{3}\right)\Big|_1^2$$

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

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.

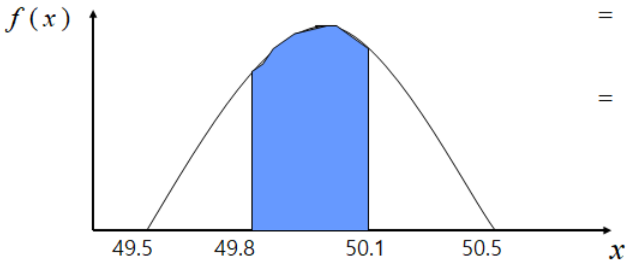
$$f(x) = 1.5 - 6(x - 50.2)^2 \quad \text{for} \quad 49.5 \le x \le 50.5$$
$$f(x) = 0, \quad \text{elsewhere}$$



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

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

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



$$\begin{aligned} &= [1.5\times 50.1-2(50.1-50.0)^3] \\ &\quad -[1.5\times 49.8-2(49.8-50.0)^3] \\ &= 75.148-74.716 = 0.432 \end{aligned}$$

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

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

$$\begin{aligned} P(a < X \le b) &= P(X \le b) - P(X \le a) \\ &= F(b) - F(a) \end{aligned}$$

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

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

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

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

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

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

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

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

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

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

5.2. Probability Density Function 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.

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

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

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

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

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

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

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.

5.3. 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.

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

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

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

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 −∞ 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 $-1 \le x \le 1$ find the CDF F(x).

Solution:

Integrate the PDF to find the CDF:

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

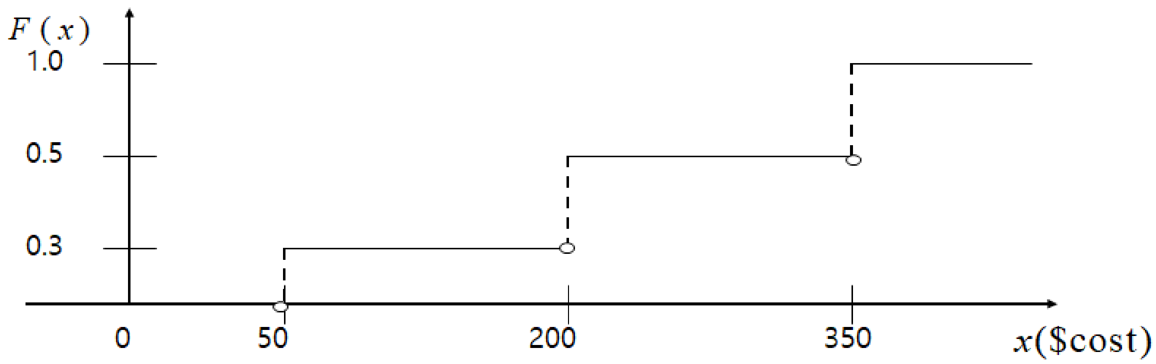
Solving the integral:

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

The CDF (F(x)) is:

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

Example: Machine Failures



$$-\infty < x < 50 \Rightarrow F(x) = P(\text{cost} \le x) = 0$$
$$50 \le x < 200 \Rightarrow F(x) = P(\text{cost} \le x) = 0.3$$
$$200 \le x < 350 \Rightarrow F(x) = P(\text{cost} \le x) = 0.3 + 0.2 = 0.5$$
$$350 \le x < \infty \Rightarrow F(x) = P(\text{cost} \le x) = 0.3 + 0.2 + 0.5 = 1.0$$

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

5.4. 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:

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± b) = aE(X) ± 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 ± 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=$X_1 + X_2 + \ldots + X_N$ is taken E(T)=E($X_1 + X_2 + \ldots + X_N$)=E(X_1)+E(X_2)+ … + E(X_N)

Var(X) = E(X^2)- E^2(X)= E(X^2)- μ^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

5.4.1. 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.

$X=(x_1, x_2, \ldots, x_n)$ 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*(cost=50)=0.3, *P*(cost=200)=0.2, *P* (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:

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

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

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

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

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

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

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

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

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

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

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

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

5.4.2. Expected Value and Variance in Continuous Random Variables

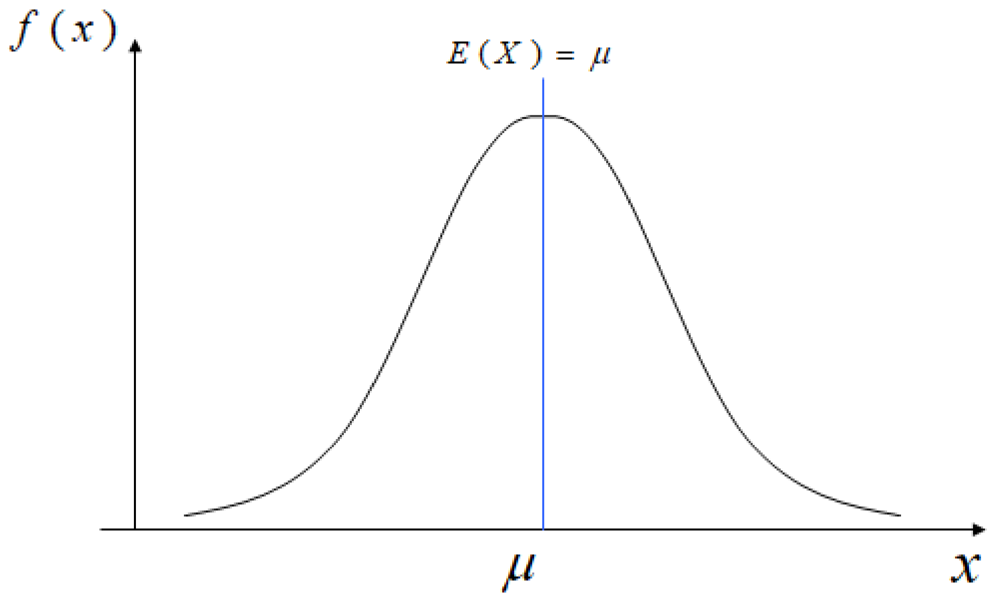
The expected value or weighted average for a continuous random variable (function) is the random variable x, the probability density function of the random variable x, and f(x) is

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

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

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

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



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**.**

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

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

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

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

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

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

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

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

Example:

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

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

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.

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

$$\int_{-\infty}^{\infty} f(x)dx = 2\int_0^1 (1-x)dx = 2\Big(x - \frac{x^2}{2}\Big)\Big|_0^1 = 1$$

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

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

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{\infty} xf(x)dx \\ &= \int_0^1 x\Big[2(1-x)\Big]dx \\ &= 2\int_0^1 \Big(x - x^2\Big)dx \\ &= 2\Big(\frac{x^2}{2} - \frac{x^3}{3}\Big)\Big|_0^1 \\ &= 1/3 \end{aligned}$$

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

$$\operatorname{Var}(X) = \int_{-\infty}^{\infty} \left(x - \mathbb{E}(X)\right)^2 f(x)dx$$
$$\sigma = \sqrt{\operatorname{Var}(X)}$$

$$\begin{aligned}
\operatorname{Var}(X) &= \int_{-\infty}^{\infty} \left(x - \mathbb{E}(X)\right)^2 f(x)dx \\
&= \int_0^1 \left(x - 1/3\right)^2 \cdot 2(1-x)dx \\
&= 2\int_0^1 \left(x^2 - \frac{2}{3}x + \frac{1}{9}\right)(1-x)dx \\
&= 2\int_0^1 \left(-x^3 + \frac{5}{3}x^2 - \frac{7}{9}x + \frac{1}{9}\right)dx \\
&= 2\left(-\frac{1}{4}x^4 + \frac{5}{9}x^3 - \frac{7}{18}x^2 + \frac{1}{9}x\right)\Big|_0^1 \\
&= 2\left(-\frac{1}{4} + \frac{5}{9} - \frac{7}{18} + \frac{1}{9}\right) \\
&= \frac{1}{18}
\end{aligned}$$

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

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

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \\
&= \int_{-\infty}^{\infty} x^2 f(x) dx - \left(\frac{1}{3}\right)^2 \\
&= 2\int_0^1 x^2(1-x)dx - \frac{1}{9} \\
&= 2\int_0^1 (x^2 - x^3)dx - \frac{1}{9} \\
&= 2\left(\frac{1}{3}x^3 - \frac{1}{4}x^4\right)\Big|_0^1 - \frac{1}{9} \\
&= 2\left(\frac{1}{3} - \frac{1}{4}\right) - \frac{1}{9} \\
&= \frac{1}{18}
\end{aligned}$$

Example:

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

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

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

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

$$V(X) = \int_0^1 (x - \frac{2}{3})^2 2x dx = \int_0^1 \left(x^2 - \frac{4}{3}x + \frac{4}{9}\right) 2x dx$$

$$= \int_0^1 \left(2x^3 - \frac{8}{3}x^2 + \frac{8}{9}x\right) dx = \frac{1}{2}x^4 - \frac{8}{9}x^3 + \frac{8}{18}x^2 I_0^1 = \frac{1}{18}$$

$$E(X^2) = \int_0^1 x^2 f(x) dx$$

$$= \int_0^1 x^2 2x dx = \int_0^1 2x^3 dx = \frac{2}{4} x^4 I_0^1 = \frac{1}{2}$$

$$V(X) = E(X^2) - \left(E(X)\right)^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$$

Example:

Find the standard deviation of the random variable X whose probability density function is given below.

$$f(X) = \begin{cases} \frac{3}{10}(3x - x^2) & 0 \leq x \leq 2 \\ 0 & \text{For other X's} \end{cases}$$

$$E(X) = \mu = \int_0^2 x f(x) dx = \int_0^2 x \frac{3}{10}(3x - x^2) dx$$

$$= \left(\frac{3}{10} x^3 - \frac{3}{40} x^4\right)\Bigg|_0^2 = \left(\frac{3}{10} 2^3 - \frac{3}{40} 2^4\right) - 0 = 1.2$$

$$E(X^2) = \int_0^2 x^2 f(x)dx = \int_0^2 x^2 \frac{3}{10}(3x - x^2)dx$$

$$= \left(\frac{9}{40}x^4 - \frac{3}{50}x^5\right)\Bigg|_0^2 = \left(\frac{9}{40}2^4 - \frac{3}{50}2^5\right) - 0 = 1.68$$

$$V(X) = \sigma^2 = E(X^2) - \mu^2 = 1.68 - 1.2^2 = 0.24$$

$$\sigma = \sqrt{0.24} = 0.489$$

5.5. Rassal Değişkenlerde Momentum

Fizikte Momentum:
Bir cismin hareket miktarını belirtmek için momentumdan yararlanılmaktadır. Aynı zamanda momentum **itme kuvveti** olarak da ifade edilebilmektedir. Saniyede kütlenin hareket etme hızını vermektedir. Bir nesnenin sahip olduğu momentumun miktarı, iki fiziksel büyüklüğe bağlıdır: kütlesi ve gözlem çerçevesindeki hızı. Fizikte, momentum için kullanılan sembol genellikle P harfidir ve vektörel büyüklüktür. Momentum şöyle ifade edilebilir;
P=mV
burada P momentum, *m* kütle ve V hızdır. Saniyede kütle ile hızın çarpımı kadar itme kuvvetine sahiptir.

Yere paralel düz bir rotada 0.25 m/s hızına ve 10.000 kg kütleye sahip bir uçağın momentumu yere göre ölçüldüğünde, düz bir rotada 2500kg m/s ‘dir. Oysa kokpitin içindeki bir pilot, kokpit gözlem çerçevesine göre uçağın hızını sıfır ölçeceğinden, momentumunu da sıfır ölçer.

Momentum aynı zamanda enerjinin aktarılma yönünü de göstermektedir. Üstelik bir cisme uygulanan kuvvetle doğru orantılı olarak momentumunda da değişiklik görülebilir. Günlük hayattan bir örnek verilecek olursa iki arabanın çarpışması sonrasında en yüksek hıza sahip olan aracın bulunması ile çarpan aracın tespiti de momentum aracılığı ile hesaplanmaktadır.

Momentum vektörel bir niceliktir ve hız ile aynı yöndedir. Birimi kg·m/s 'dir. Cismin kütlesi ve hızı değişmediği müddetçe cismin momentumu sabittir. Bir cismin momentumunu değiştirmek için cismin hızı değiştirilir. Cismin hızını değiştirebilmek için ise cisme ivme kazandırılması gerekir, bunun içinse cisme net bir kuvvet uygulanmalıdır.

Fizikte ivme, hızın zamana göre türevi olarak tanımlanır, hızın zamanla değişimidir. Büyüklüğü uzaklık/zaman^2 olan bir vektörel niceliktir ve cismin hem hızının hem de yönünün şiddetlerindeki değişimini gösterir.

İvme(a) = Hız (V) / Zaman(t)

Örnek:
Saniyede 150 metre hızla giden 1000 kilogramlık bir araba aşırı hız ve dikkatsizlik sonucu direksiyon hakimiyetini kaybederek karşı yönden gelen kamyon ile çarpışır. Kamyonun ağırlığı 10.000 kilo ve hızı 100 m/s

Kamyonun momentumu: 10.000*100=1.000.000 kg m/s
Arabanın momentumu: 1000*-150=-150.000 kg m/s(Karşı yönden geldiği için hızı – olarak alıyoruz)

Kaza öncesi toplam momentum 1.000.000-150.000=850.000 Kg m/s

Kaza sonrasında momentum yine 850.000 Kg m/s olmak zorundadır.
Buna bakarak kaza sonrasında iki aracın hızını bulabiliriz. İki aracın kütlelerinin birleştiğini varsayıyoruz. Yani toplam kütle 10.000+1.000=11.000 kg oluyor. Ve momentum da 850.000

850.000=11.000*V >>> V=100.000/16.000=77,3m/s

Kısaca iki araç çarpıştıktan sonra 77.3 metre/saniye hızla kamyonun gittiği yöne doğru yola doğru devam eder. Buna bakarak görürüz ki kamyon 100 metre/saniye hızdan 77.3 metre/saniye, otomobilse 150 metre hızdan -77.3 metre/saniye hıza düşüyor. Yani kamyona göre çok daha fazla ivme yaşıyor. f=m*a yani kuvvet=kütle*ivme olduğuna göre ivme ne kadar yüksekse kuvvet ve ona bağlı olarak da hasar çok daha fazla olacaktır.

Açısal momentum, herhangi bir cismin dönüş hareketine devam etme isteğinin bir göstergesidir ve bu nicelik cismin kütlesine, şekline ve hızına bağlıdır. Açısal momentum bir vektör birimidir ve cismin belirli eksenler üzerinde sahip olduğu dönüş eylemsizliği ile dönüş hızını ifade eder.

Beklenen değer işlemine dayalı olarak geliştirilen momentler:
Bir rassal değişkenin olasılık dağılımı, çok kere pratikte anlanması ve uygulanması kolay olan küçük sayıda parametreler ile nitelendirilir. Örneğin, sadece "ortalama değer" olan λ değerini bilmek Poisson dağılımını bilmek için yeterlidir. Ortalama kavramı matematik teoride bir rassal değişkenin beklenen değeri olarak, yani E[X] olarak ifade edilir. Genellikle E[f(X)] ifadesi f(E[X]) ifadesine eşit değildir. "Ortalama değer" bilinince, bu ortalama değerin X tipik değerlerinden ne kadar fazla uzaklıkta olduğu sorusu hemen akla gelir ve bu soruya yanıt bu rassal değişkenin standart sapması ve varyansı ile bulunur.

Matematik kuramı içinde bu (genelleştirilmiş) momentler problemi olarak bilinmektedir: Bilinmekte olan bir sınıf rassal değişkenler olan X için, E[fi(X)] ifadesindeki beklenen değerler ile rassal değişken Xin dağılımını tam olarak nitelendiren bir {fi} fonksiyonlar koleksiyonu bulunması istenmektedir.

Beklenen değer işlemine dayalı olarak geliştirilen momentler, gerek fonksiyonların parametreleri, gerekse ilgilenilen rassal sistemin bütünleşik davranış göstergeleri için türetilen bir genellemedir. Beklenen değer E(X), moment denilen parametreler sınıfının bir özel halidir.

$$E[(X-a)^k]$$

İfadesinde a bir gerçel sayı olmak üzere, X rassal değişkenin a'ya göre k. mertebeden momenti (k. momenti) denir.

Eğer a=0 ise, X'in k. mertebeden momenti $E(X^k)$ beklenen değerine dönüşür ve bu moment m_k ile gösterilir.

X, xi değerlerini Pi olasılıkları ile alan ayrık rasgele değişken ise, k. mertebeden moment şu şekilde formüle edilir;

$$m_k = E(X^k) = \sum_i X_i^k \, P_i$$

X, f(x) olasılık yoğunluk fonksiyonuna sahip sürekli rasgele değişken ise, k. mertebeden moment şu şekilde formüle edilir;

$$m_k = E(X^k) = \int_{-\infty}^{+\infty} \mathrm{X}^k \, f(x) dx$$

Momentin Özellikleri:

1) $E[(aX)^k] = a^k E[(X)^k]$
2) $E[(X-c)^k]$ beklenen değerine c noktasına göre k. mertebeden moment denir. c =0 noktasına göre momentlere adi momentler denir.
3) $\mu_k = E[(X-E(X))^k]$: Beklenen değere (ortalamaya) göre momentlere merkezi moment denir. Buna göre ilk üç moment şöyle bulunur;

$$\mu_1 = E[(X-E(X))^1] = E(X-m_1) = E(X) - m_1 = m_1 - m_1 = 0$$

$$\mu_2 = E[(X-E(X))^2] = E[(X-m_1)^2] = E(X^2) - 2m_1E(X) + m_1^2 = m_2 - m_1^2$$

$$\mu_3 = E[(X-E(X))^3] = E[(X-m_1)^3] = E(X^3) - 3m_1E(X^2) + 3m_1^2E(X) - m_1^3$$
$$= m_3 - 3m_1m_2 + 3m_1^2m_1 - m_1^3 = m_3 - 3m_1m_2 + 2m_1^3$$

4) İkinci mertebeden merkezi momente $\mu_2 = E[(X-E(X))^2]$, X rassal değişkenin varyansı denir.

   $Var(x) = \sigma^2 = m_2 - m_1^2$

5) Standart sapmanın beklenen değere oranına değişim (varyasyon) katsayısı (DK) denir. DK = $\frac{\sigma}{m_1}$ veya DK = $\frac{\sigma}{m_1} * 100$ ile ifade edilir.
   Beklenen değer 1'e eşit olduğunda DK'da standart sapmaya eşit olur.

6) $E[|X-E(X)|^k]$ ifadesine, X rassal değişkenin k. mertebeden mutlak merkezi momenti denir.
7) X, f(x) olasılık fonksiyonuna (veya olasılık yoğunluk fonksiyonuna) sahip ayrık veya sürekli rassal değişken olsun. h>0 olmak üzere | t | < h aralığında her değeri alan t için e^{tx} in beklenen değeri X'in moment çıkaran fonksiyonu olarak tanımlanır. Buna göre X 'in moment çıkaran fonksiyonu;

$$m_x(t) = E(e^{tx}) = \begin{cases} \sum_x e^{tx} \cdot f(x) \;; & \text{X rassal değişkeni kesikli ise.} \\ \int_{-\infty}^{+\infty} e^{tx} \cdot f(x)dx \;; & \text{X rassal değişkeni sürekli ise.} \end{cases}$$

Moment Çıkaran Fonksiyonu

Teorem: X rasgele değişkeninin moment çıkaran fonksiyonu $m_x(t)$ olsun.

$$\left.\frac{d^r m_x(t)}{dt^r}\right]_{t=0} = E(x^r) = m_r$$

- Birinci türevde; $M'(t=0) = \frac{dM(t)}{dt}\Big|_{t=0} = E(X) = m$ (X 'in 1. momentidir.)
- İkinci türevde; $M''(t=0) = \frac{d^2M(t)}{dt^2}\Big|_{t=0} = E(X^2)$ (X 'in 2. momentidir.)

. . .

- n. türevde; $M^{(n)}(t=0) = \frac{d^nM(t)}{dt^n}\Big|_{t=0} = E(X^n)$ (X 'in n . momentidir.)

Moment Çıkaran Fonksiyonu:

Örnek: $P(X=x) = \begin{cases} x/8 & , \quad x=1,3,4 \\ 0 & , \quad \text{diğer durumlarda} \end{cases}$

Yukarıda bir X kesikli rassal değişkenin olasılık fonksiyonu verilmiştir.

a) X rassal değişkenin moment çıkaran fonksiyonunu bulalım.

$$\begin{aligned} m_x(t) = E(e^{tx}) &= \sum_x e^{tx} \cdot f(x) \\ &= \sum_x e^{tx} \cdot \frac{x}{8} \\ &= e^t \cdot \frac{1}{8} + e^{3t} \cdot \frac{3}{8} + e^{4t} \cdot \frac{4}{8} \\ &= \frac{1}{8}(e^t + 3e^{3t} + 4e^{4t}) \end{aligned}$$

b) Moment çıkaran fonksiyonu kullanarak ilk iki momenti bulalım.

➢ Fonksiyonun 1. türevini alırsak;

$$M'(t) = \frac{1}{8}(e^t + 3 \cdot 3e^{3t} + 4 \cdot 4e^{4t})$$

$$M'(0) = \frac{26}{8} = E(X) = \mu$$

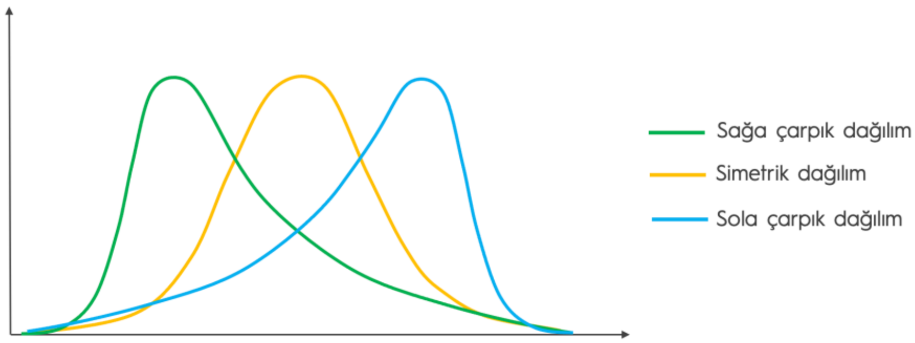
➢ Fonksiyonun 2. türevini alırsak;

$$M''(t) = \frac{1}{8}(e^t + 3 \cdot 3 \cdot 3e^{3t} + 4 \cdot 4 \cdot 4e^{4t})$$

$$M''(0) = \frac{1}{8}(1 + 27 + 64) = \frac{92}{8} = \frac{23}{2}$$

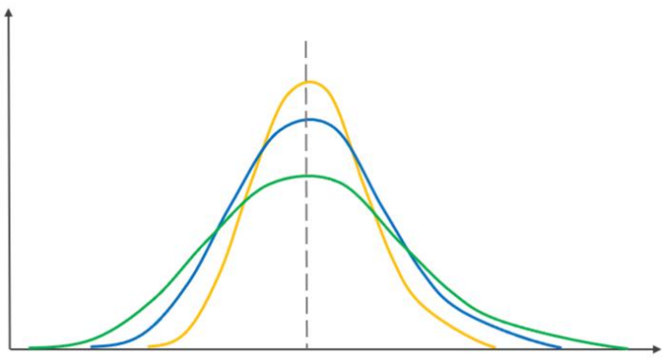
Bir Dağılımda Çarpıklık (Skewness):

Ortalamaya göre üçüncü moment μ_3' e simetrik olmayışın ölçüsü denir. Yani çarpıklık, bir dağılımda simetriden ayrılışın derecesidir.



Bir Dağılımda Sivrilik (Kurtosis):

Ortalamaya göre dördüncü moment μ_4'e düzlük (yassılık) ölçüsü denir. Yani sivrilik, bir dağılımın basıklığının derecesidir.



Markov Eşitsizliği:

X rassal değişkenin olasılık fonksiyonu veya olasılık yoğunluk fonksiyonu bilinmediğinde, bu değişkene ilişkin olasılıkları hesaplamak için geliştirilen bir eşitsizliktir. Markov eşitsizliği bir rassal değişkenin dağılımı hakkında bilgi sağlar ve bu değişkenin değerlerinin ne kadar dağılmış olduğunu ölçer.

Tanım: X, negatif olmayan değerler alabilen bir rassal değişken ise, P(X ≥ C) ≤ E(X)/C eşitsizliğine Markov Eşitsizliği denir.

Örnek:

X, bir çocuk tiyatrosunda rastgele seçilmiş çocukların yaşlarını gösteren bir rassal değişkendir. Bu rassal değişkenin ortalaması E(X) = 5,5 yaş olduğuna göre rastgele seçilen bir çocuğun yaşının en az 17 olma olasılığını hesaplayalım.

P(X ≥ C) ≤ E(X)/C

P X ≥ 17 ≤ 5,5/17 = 0,32

Chebyshev Eşitsizliği:

Chebyshev Eşitsizliği, aritmetik ortalaması ve varyansı (veya standart sapması) verilen bir rassal değişkenin, ortalama etrafında ne kadar dağıldığını ölçen eşitsizliktir. Tanım: X, μ ortalamalı ve σ^2 varyanslı bir rassal değişken olsun. k > 0 değeri için Chebyshev eşitsizliği,

$P((X - \mu)^2 \geq k^2 * \sigma^2) \leq 1/k^2$ şeklinde tanımlanır

Aynı şekilde;

$$P(|X-\mu| \geq k \cdot б) \leq \frac{1}{k^2} \quad \text{veya} \quad P(|X-\mu| \geq k) \leq \frac{б^2}{k^2} \quad \text{veya} \quad P(|X-\mu| < k \cdot б) \geq 1 - \frac{1}{k^2}$$

Örnek: X, bir çocuk tiyatrosunda rastgele seçilmiş çocukların yaşlarını gösteren bir rassal değişkendir. Bu rassal değişkenin ortalaması E(X) = 5,5 yaş ve б = 1 yaş olduğuna göre rastgele seçilen bir çocuğun yaşının en az 17 olma olasılığını hesaplayalım.

$P(X \geq 17) = P(X - \mu \geq 17 - \mu) = P(|X - 5{,}5|) \geq 11{,}5)$ olur. Buradan,

$$= P(|X - 5{,}5|) \geq 11{,}5) \leq \frac{б^2}{(11{,}5)^2} = \frac{1}{(11{,}5)^2} = 0{,}0076$$

Yani, tiyatroya gelen 10000 kişiden en fazla 76 kişi 17 yaşından büyük olabilmektedir.

X rassal değişkeninin ortalaması ve varyansı bilindiğinde, X'in alabileceği değerlerin olasılıklarının hesaplanması için geliştirilen fonksiyonları kullanmadan (veya bu olasılıkların bilinmediği durumlarda), X'in özel aralıklarda değer alma olasılıklarının alt ve üst sınırı Chebyshev eşitsizliği ile bulunabilmektedir.

Örnek: İstatistik dersinin dönem sonu not ortalaması 75 ve standart sapması 8’dir. Öğrencilerin dönem sonu notunun 59 ile 91 aralığında olma olasılığını bulalım.

Çözüm: μ = 75, б = 8, ÜS = 91, AS = 59 olarak verilmiş.

- ÜS = μ + k. б ve AS = μ - k. б olmak üzere;

  μ + k . б = 91 ⟹ 75 + k . 8 = 91
  μ – k . б = 59 ⟹ 75 – k . 8 = 59 } k=2

- Bu durumda P(59 < X < 91) = P(μ - k . б < X < μ + k . б)

  = P(- k . б < X - μ < k . б)

  $= \mathrm{P}(|\mathrm{X}-\mu| < \mathrm{k}\cdot б) \geq 1-\frac{1}{\mathrm{k}^2} = 1-\frac{1}{2^2} = 0{,}75$

Chebyshev eşitsizliği ile X’in ortalamadan her iki yönde belirli bir sapma ile bulunan aralıkta değer alma olasılığının alt sınırı hesaplanabilir. Örnek: Ortalamadan en az 2б uzakta bulunmanın maksimum olasılığı nedir?

k · б = 2 б olmak üzere k=2 ‘dir.

$$\mathrm{P}(|\mathrm{X}-\mu| \geq \mathrm{k}\cdot б) \leq \frac{1}{\mathrm{k}^2}$$

$$\mathrm{P}(|\mathrm{X}-\mu| \geq 2\cdot б) \leq \frac{1}{2^2} = 0{,}25$$

Örnek:

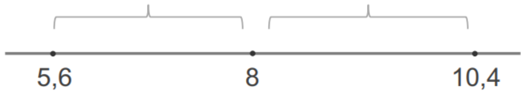
X rassal değişkenin ortalamasının 8 ve standart sapmasının 2 olduğu bilinmektedir. Bu rassal sistemden gelişi güzel çekilen bir birimin; a) 5,6’ya eşit veya küçük ya da 10,4’e eşit veya büyük çıkma olasılığını bulunuz.

a) 5,6’ya eşit veya küçük ya da 10,4’e eşit veya büyük çıkma olasılığını bulunuz.

O_1 = {X|x ≤ 5,6 veya x ≥ 10,4}

P(O_1) = P(x ≤ 5,6 veya x ≥ 10,4)

k.б = 8 – 5,6= 2,4 k.б = 10,4 – 8 = 2,4



⟶ k=1,2

$\mathrm{P}(|\mathrm{X}-\mu| \geq \mathrm{k}\cdot б) \leq \frac{1}{\mathrm{k}^2}$

$\mathrm{P}(|\mathrm{X}-8| \geq 1{,}2\cdot 2) \leq \frac{1}{1{,}2^2}$

$\mathrm{P}(\mathrm{O}_1) \leq 0{,}6944$

b) 4 ile 12 arsında çıkma olasılığını hesaplayınız.

$O_2 = \{X|\ 4 \leq x \leq 12\}$ olayı olmak üzere $P(O_2) = P(4 \leq x \leq 12)$ isteniyor.

$$\left.\begin{array}{l} \text{ÜS} = \mu + k.\,б\ = 8 + k.2 = 12 \\ \text{AS} = \mu - k.\,б\ = 8 - k.2 = 4 \end{array}\right\} k=2$$

Bu durumda $P(4 \leq x \leq 12) = P(\mu - k\,.\,б < X < \mu + k\,.\,б)$

$$P(|X - \mu| < k \cdot б) \geq 1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 0{,}75$$

Bu sonuca göre, rastgele çekilen bir birimin belirtilen aralıkta (4 ile 12 arasında) olma olasılığı %75'den büyüktür.

X, ortalaması μ olan ve negatif değerler almayan rassal değişken olsun. Herhangi bir $k > 0$ için;

$$P(X > k\mu) < \frac{1}{k}$$

şeklinde hesaplanabilir.

5.6. Rassal Değişkenlerde Kovaryans ve Korelasyon

Bir rastlantı değişkenin alabileceği değerlerle, bu değerleri alabilmesi olasılıkları arasındaki ilişkiyi gösteren bir fonksiyondur.

$f(x) \geq 0$ ve $\int_{-\infty}^{+\infty} f(x)dx = 1$

koşulunu sağlayan f(x) fonksiyonuna x'in olasılık yoğunluk fonksiyonu denir.

Sürekli rasgele deşikeni için,

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

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

Not: E(X)>0 ise beklenti uygun, E(X)=0 ise beklenti ortada, E(X)<0 ise beklenti uygun değil.

E(aX)=a E(X)

E(X+a)= E(X)+a

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

Var(X) = $E(X^2)$- $E^2(X)$= $E(X^2)$- μ^2

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

Kovaryans:
iki değişkenin birlikte değişimlerinin ölçüsü kovaryans olarak bilinir ve Cov(X,Y) ya da Kov(X,Y) ile gösterilir.

$Kov(X, Y) = E[(X - E(X)) - (Y - E(Y))] = E(XY) - E(X)E(Y)$

Eğer X ve Y rasgele değişkenleri bağımsız ise

$E(XY) = E(X)E(Y)$ olacağı için $Kov(X, Y) = 0$ olur.

Ancak, $Kov(X, Y) = 0$ olması demek, X ve Y rasgele değişkenlerinin bağımsız olduğu anlamına gelmez. Bu durumda iki değişken arasındaki doğrusal ilişkinin sıfır olduğu yargısına varılır.

Korelasyon katsayısı:
Korelasyon katsayısı iki değişken arasındaki doğrusal ilişkinin derecesini belirleyen ve karşılaştırmaya olanak veren bir katsayıdır.

$\rho_{XY} = Cov(X, Y) /\surd(\sigma_x \sigma_y), \quad -1 \leq \rho_{XY} \leq 1$
$\rho_{XY} = 0$ olması X ve Y rasgele değişkenleri arasında doğrusal bir ilişki olmadığını,
$\rho_{XY} < 0$ olması X ve Y rasgele değişkenleri arasında ters yönlü doğrusal bir ilişki olduğunu,
$\rho_{XY} > 0$ olması X ve Y rasgele değişkenleri arasında aynı yönlü doğrusal bir ilişki olduğunu gösterir.

Örnek: X ve Y değişkenlerinin bileşik olasılık yoğunluk fonksiyonu
$f_{X,Y}(x, y) = x + y$, $0 < x < 1$, $0 < y < 1$ verilmiş olsun.

$Kov(X, Y)$ ve ρ_{XY} değerlerini hesaplayınız.

$E(X) = \int\int x(x + y)dxdy = 7/12$

$E(X^2) = \int\int x^2(x + y)dxdy = 5/12$

$\sigma_x^2 = 5/12 - (7/12)^2 = 11/144$

$E(Y) = \int\int y(x + y)dxdy = 7/12$

$E(Y^2) = \int\int y^2(x + y)dxdy = 5/12$

$\sigma_y^2 = 5/12 - (7/12)^2 = 11/144$

$E(XY) = \int\int xy\, f_{X,Y}(x, y)\, dxdy$

$E(XY) = \int\int xy(x + y)dxdy = 1/3$

$Cov(X, Y) = E(XY) - E(X)E(Y) = 1/3 - (7/12)(7/12) = -1/144$

$\rho_{XY} = Kov(X, Y) / \sqrt{(\sigma_x \sigma_y)} = -1/144 / \sqrt{(11/144)(11/144)} = -0.091$

Ayrık Olasılık Dağılımında Kovaryans Hesaplama:

X ve Y rasgele değişkenler arasındaki ilişkiyi gösteren bir katsayıdır.

Cov(X,Y)= $\sigma_{xy} = E(XY) - E(X)E(Y)$

$E(XY)$: Korelasyondur.

$E(X)$: X rasgele değişkenin beklenen değeri

$E(Y)$: Y rasgele değişkenin beklenen değeri

Örnek: Ayrık ortak olasılık dağılımı veriliyor.

Y \ X	x1	x2
y1	a11	a12
y2	a21	a22

Y \ X	1	2
0	1/12	1/6
1	1/4	1/2

Cov(X,Y)= ?

σ_{xy}=?

Kovaryans=?

1. $E(XY) = x_1y_1a_{11} + x_1y_2a_{21} + x_2y_1a_{12} + x_2y_2a_{22}$
   $E(XY)$ =1*0*1/12 + 1*1*1/4 + 2*0*1/6 + 2*1*1/2 =1/4+1=5/4
2. P(X=x1)= a11 + a21=1/12+1/4=1/3
   P(X=x2)= a12 + a22=1/6 + 1/2=2/3
3. P(Y=y1)= a11 + a12=1/12+1/6=1/4
   P(Y=y2)= a21 + a22=1/4 + 1/2=3/4
4. E(X)= x1* P(X=x1) + x2* P(X=x2) = 1*1/3 + 2*2/3=5/3
   E(Y)= y1* P(Y=y1) + y2* P(Y=y2) = 0*1/4 + 1*3/4=3/4
5. Cov(X,Y)= $E(XY) - E(X)E(Y)$=5/4-5/3*3/4=0

Cov(X,Y)=0 çıktığı için X ve Y rasgele değişkenleri bağımsızdır.

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.

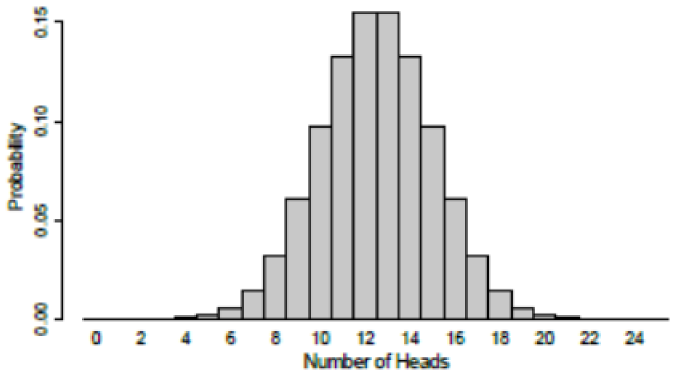
Other than these examples, there are various other examples of random discrete variables. Some of these are as follows:

- 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.), $\boldsymbol{p(x)}$ 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;

$$\mathbf{P(X=x)=p^{x}q^{1-x}}$$

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:

$$ÇK = \frac{q-p}{\sqrt{pq}} \quad \text{ve} \quad BK = \frac{1}{pq} - 6$$

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.

$\mathbf{P(X=x)=0.6^x0.4^{1-x}}$; $x=0,1$

The opposite of not smoking versus drinking = $\dfrac{p}{q}$ = 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 ve 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, The number of heads
- A jar containing 8 black and 4 white balls is replaced and 3 balls are drawn. X, The number of black balls drawn.
- A box containing 3 defective and 7 perfect pieces is replaced and 4 pieces are selected. X, 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.

$P(x, n-x) = p.p.p.........p.(1-p).(1-p).......(1-p) = p^x q^{n-x}$

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:

$$\mathbf{P(X; n,p) = \frac{n!}{x!(n-x)!} p^x q^{n-x}}$$

$$\mathbf{P(X; n,p) = C_n^x\, p^x q^{n-x}}$$

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:

  $\mu_1 = 0$

  $\mu_2 = \sigma_x^2$

  $\mu_3 = npq[q - p]$

  $\mu_4 = npq[1\text{-}6pq+3npq]$

- Asymmetry (Skewness) and Kurtosis Coefficients:

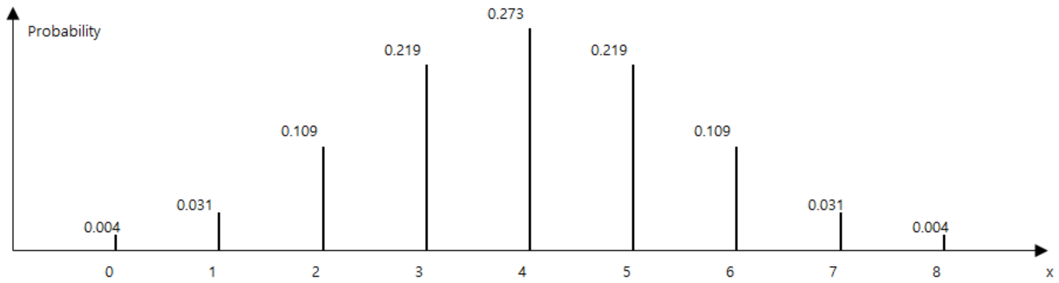
  ÇK = $\frac{q-p}{\sqrt{npq}}$ ve BK = 3 + $\frac{1-6pq}{npq}$

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.

$$P(X = 3) = \binom{8}{3} \times 0.5^3 \times (1-0.5)^5 = \frac{8!}{3!5!} \times 0.5^8 = 0.219$$

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= \binom{8}{0} \times 0.5^0 \times (1-0.5)^8 + \binom{8}{1} \times 0.5^1 \times (1-0.5)^7 \\ &= \frac{8!}{0!8!} \times 0.5^8 + \frac{8!}{1!7!} \times 0.5^8 = 0.004 + 0.031 = 0.035 \end{aligned}$$



x	0	1	2	3	4	5	6	7	8
$P(X \le x)$	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,

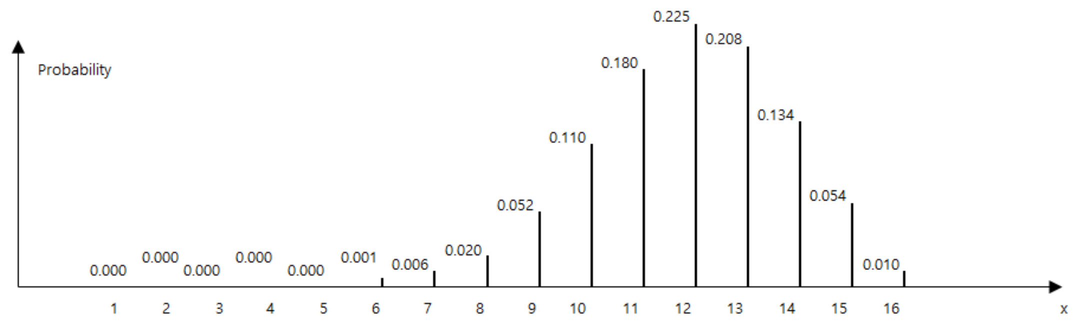
$$P(X=12)=\binom{16}{12}\times 0.75^{12}\times 0.25^{4}=\frac{16!}{12!4!}\times 0.75^{12}\times 0.25^{4}=0.225$$

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

$P(x \leq 4) = 1 - P(x=5)$
$P(x=5) = (5! / (5!*0!))*(0.30)^5 *(0.70)^0 = 0.00243$
$P(x \leq 4) = 1 - 0.00243 = 0.99757$

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.

$$P(x,6,0.05) = \begin{cases} \binom{6}{x}(0.05)^x(0.95)^{6-x} & , \quad x = 0,1,2,\ldots,6 \\ 0, & diğer\ x'leriçin \end{cases}$$

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:

  $ÇK = \frac{\mu_3}{\sigma^3} = \frac{1}{\sqrt{\Lambda}}$ ve $BK = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1}{\Lambda}$

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.

$\lambda = \mu = n.p$

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. ($p \leq 0.01$ and $\lambda = n.p \leq 5$)

Example:

The arithmetic average of the incorrect grades entered in a year in a University with 10,000 students is 0.5.

1. Aritmatrik ortalama, $\Lambda = 0.5$

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.

$$P(x) = \frac{e^{-\lambda}\,\lambda^{x}}{x!} \;, \quad x = 0, 1, 2, \ldots\ldots n$$

$$P(x) = \frac{e^{-0.5}\,(0.5)^{x}}{x!} \;, x = 0, 1, 2, \ldots\ldots n$$

Bu Poisson dağılımının,

- Aritmetik ortalaması: $\mu_x = E(X) = n.p = \lambda$
  μ_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;

Λ = 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 <u>0.0153</u> 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

$P(x) = \frac{e^{-ʎ} ʎ^x}{x!} = \frac{e^{-0.2} 0.2^x}{x!}$, 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.

$$P(X=k)=\frac{e^{-\lambda}\lambda^{k}}{k!}$$

$$\text{a) } P(3)=\frac{e^{-10}10^{3}}{3!}=0.007567$$

$$\begin{aligned}\text{b) } P(3)+P(2)&=0.007567+\frac{e^{-10}10^{2}}{2!}\\&=0.007567+0.00227\\&=0.009837\end{aligned}$$

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.
  $C_n^N = \dfrac{10!}{5!\,(10-5)!} = 252$
- 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.

$$P\,(X = 5) = \frac{C_5^6\, C_0^4}{C_5^{10}} = \frac{\dfrac{6!}{5!\,(6-5)!} * \dfrac{(4)!}{(0)!\,(4)!}}{\dfrac{10!}{5!\,(10-5)!}} = \frac{6 * 1}{252} = 0.0238$$

b) 3 öğrencinin başarılı olma olasılığını bulunuz.

$$P\,(X = 5) = \frac{C_3^6\, C_{5-3}^{10-6}}{C_5^{10}} = \frac{\dfrac{6!}{3!\,(6-3)!} * \dfrac{(4)!}{(2)!\,(4-2)!}}{\dfrac{10!}{5!\,(10-5)!}} = \frac{20 * 6}{252} = 0.476$$

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.

Başarılı öğrenci sayısına göre:

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

Başarısız öğrenci sayısına göre:

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:

$$f(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}} \quad x > 0$$

- Beklenen değer:

$$E(X) = \int_0^\infty x \frac{1}{\mu} e^{-\frac{x}{\mu}} dx = \frac{1}{\mu} \int_0^\infty x e^{-\frac{x}{\mu}} dx \qquad u = x \quad du = dx \qquad dv = e^{-\frac{x}{\mu}} dx \quad v = -\mu e^{-\frac{x}{\mu}}$$

$$\int u\,dv = uv - \int v\,du \quad \text{kismi integrasyon islemi ile}$$

$$E(X) = \frac{1}{\mu}\left[-x\mu e^{-\frac{x}{\mu}} \Big|_0^\infty + \int_0^\infty \mu e^{-\frac{x}{\mu}} dx \right] = x e^{-\frac{x}{\mu}} \Big|_0^\infty - \mu e^{-\frac{x}{\mu}} \Big|_0^\infty$$

$E(X) = \mu$ *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?

$$f(x) = \frac{1}{24} e^{-\frac{x}{24}} \quad x > 0$$

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,

$$E(X) = \int_0^1 x\,dx = \left.\frac{x^2}{2}\right|_0^1 = \boxed{\frac{1}{2}}.$$

Not surprisingly the mean is at the midpoint of the range.

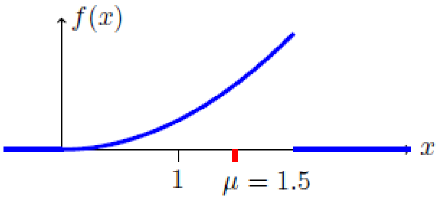
Örnek:

Let X have range $[0,2]$ and density $\frac{3}{8}x^2$. Find $E(X)$.

$$E(X) = \int_0^2 x f(x)\,dx = \int_0^2 \frac{3}{8}x^3\,dx = \left.\frac{3x^4}{32}\right|_0^2 = \boxed{\frac{3}{2}}.$$

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

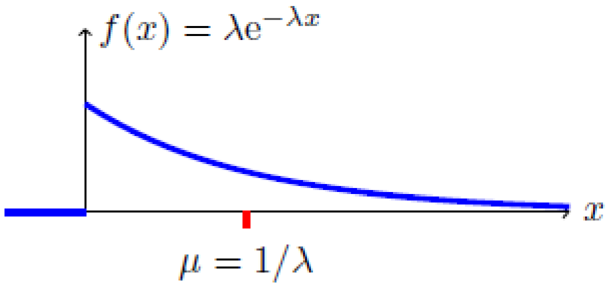
$$E(X) = \int_0^\infty x f(x)\, dx = \int_0^\infty \lambda x e^{-\lambda x}\, dx$$

(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 $E(X^2)$.

$$E(X^2) = \int_0^\infty x^2 f(x)\,dx = \int_0^\infty \lambda x^2 e^{-\lambda x}\,dx$$

(using integration by parts with $u = x^2$, $v' = \lambda e^{-\lambda x} \Rightarrow u' = 2x$, $v = -e^{-\lambda x}$)

$$= -x^2 e^{-\lambda x}\Big|_0^\infty + \int_0^\infty 2x e^{-\lambda x}\,dx$$

(the first term is 0, for the second term use integration by parts: $u = 2x$, $v' = e^{-\lambda x} \Rightarrow u' = 2$, $v = -\frac{e^{-\lambda x}}{\lambda}$)

$$= -2x\frac{e^{-\lambda x}}{\lambda}\Big|_0^\infty + \int_0^\infty \frac{e^{-\lambda x}}{\lambda}\,dx$$

$$= 0 - 2\frac{e^{-\lambda x}}{\lambda^2}\Big|_0^\infty = \frac{2}{\lambda^2}.$$

Örnek:

Let $X \sim \exp(\lambda)$. Find $\mathrm{Var}(X)$ and σ_X.

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x}\,dx = \frac{1}{\lambda} \quad \text{and} \quad E(X^2) = \int_0^\infty x^2 \lambda e^{-\lambda x}\,dx = \frac{2}{\lambda^2}.$$

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \quad \text{and} \quad \sigma_X = \frac{1}{\lambda}.$$

We could have skipped Property 3 and computed this directly from $\mathrm{Var}(X) = \int_0^\infty (x - 1/\lambda)^2 \lambda e^{-\lambda x}\,dx$.

Example:

Suppose that the random variable X has a cumulative distribution function

$$F(x) = \begin{cases} \sin(x) \text{ if } 0 \leq x \leq \frac{\pi}{2} \\ 0 \text{ if } x < 0 \text{ or } x > \frac{\pi}{2} \end{cases}$$

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

$$\begin{aligned} \mathrm{Var}(X) &= \int_0^{\frac{\pi}{2}} x^2 \cos(x) dx - E(X)^2 \\ &= (x^2 \sin(x) - 2\sin(x) + 2x\cos(x))|_{x=0}^{x=\frac{\pi}{2}} - (\frac{\pi}{2} - 1)^2 \\ &= \frac{\pi^2}{4} - 2 - (\frac{\pi^2}{4} - \pi + 1) \\ &= \pi - 3 \end{aligned}$$

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:

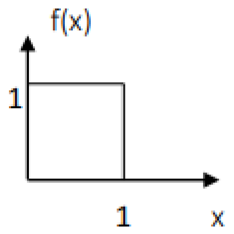
$$V(X) = E\left(\left(X - E(X)\right)^2\right) = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{dx}{b-a} = \frac{(b-a)^2}{12}.$$

Example:

For a random variable x with a uniform distribution between 0 and 1

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & d.d \end{cases}$$

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{1} f(x)dx = \int_{0}^{1} 1\, dx = 1$$

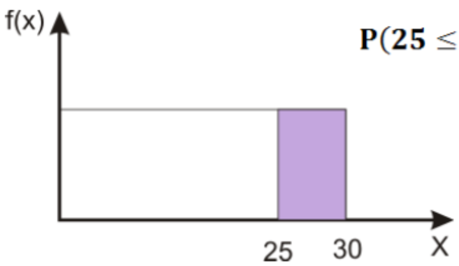


Örnek:

Süper marketteki kasaya 30 dakikalık periyotta bir müşteri gelmiştir. Bu müşterinin son 5 dakikada gelmiş olma ihtimalini hesaplayınız.

Olasılık yoğunluk fonksiyonu:

$$\mathbf{f(x)} = \begin{cases} \dfrac{\mathbf{1}}{\mathbf{b-a}} & \mathbf{a \le x \le b} \\ \mathbf{0} & \text{diğer} \end{cases} \Rightarrow \mathbf{f(x)} = \begin{cases} \dfrac{\mathbf{1}}{\mathbf{30-0}} & \mathbf{0 \le x \le 30} \\ \mathbf{0} & \text{diğer} \end{cases}$$



$$\mathbf{P(25 \le X \le} 30) = \int_{25}^{30} \mathbf{f(x)dx} = \int_{25}^{30} \frac{\mathbf{1}}{\mathbf{30}} \mathbf{dx} = \frac{\mathbf{30-25}}{\mathbf{30}} = \mathbf{1/6}$$

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.

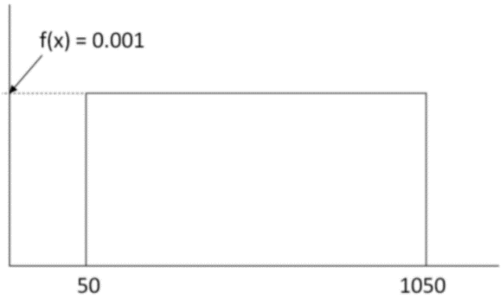
Solution

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:

$$\mu = \frac{a+b}{2} \qquad \sigma^2 = \frac{(b-a)^2}{12} \qquad \sigma = \sqrt{\frac{(b-a)^2}{12}}$$

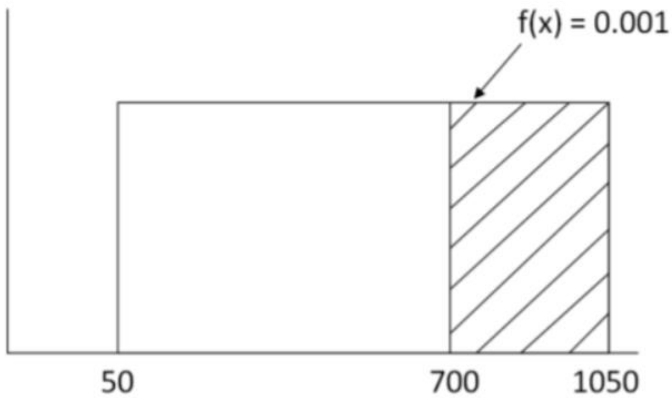
For the loose leaf tea problem:

$$\mu = \frac{50+1050}{2} = 550\text{g}$$

$$\sigma^2 = \frac{(1050-50)^2}{12} = 83,333$$

$$\sigma = \sqrt{83333} = 289\text{g}$$

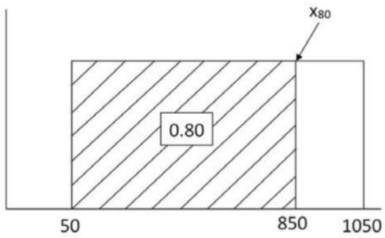
Probability problems can be easily solved by finding the area of rectangles.
Find the probability that there are at least 700 grams of TYin tea in stock.
P(X≥700)= width × height =(1050−700)(0.001)=0.35



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 $\mathrm{Var}(X)$.

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{0} x \times 0 dx + \int_{0}^{1} x \times 1 dx + \int_{1}^{\infty} x \times 0 dx$$

$$= 0 + \frac{1}{2} x^2 |_0^1 + 0 = \frac{1}{2}$$

$$\int_{-\infty}^{\infty} x^2 f(x) dx = \int_{0}^{1} x^2 dx = \frac{1}{3} x^3 |_{x=0}^{x=1} = \frac{1}{3}$$

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - E(X)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

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).

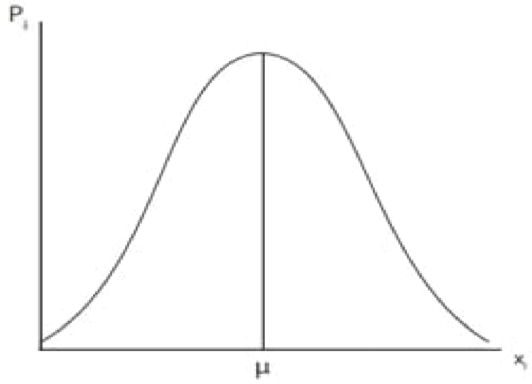


Figure: Normal Distribution

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:

$$\mathbf{f(x)} = \frac{1}{\sqrt{2\pi\sigma^2}}\mathbf{e}^{-\frac{1}{2}\frac{(X-\mu)^2}{\sigma^2}}$$

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.

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.

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,

$$\int_{-\infty}^{+\infty} f(X) = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}((X_i-\mu)/\sigma)^2} = 1$$

eşitliği gerçekleşir.

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,

$$\int_{a}^{b} f(X) = \int_{a}^{b} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}((X_i-\mu)/\sigma)^2}$$

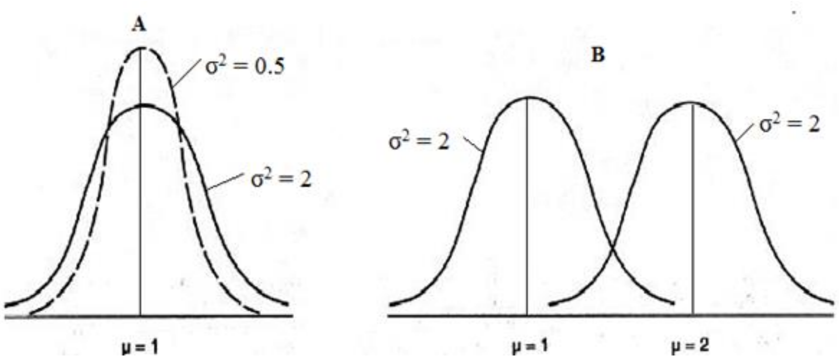
biçimindeki integralin hesaplanmasıyla yapılması gerekecektir.

Normal Dağılımın Özellikleri:

X rassal değişkeninin ortalaması μ ve varyansı σ^2 olan normal dağılıma uyduğu düşünülürse aşağıdaki özelliklere sahip olacaktır:

1. X rassal değişkeninin ortalaması $E(X) = \mu$
2. X rassal değişkeninin varyansı $Var(X) = E[(X-\mu)^2] = \sigma^2$
3. Skewness = 0 (Basıklık katsayısı)
4. Kurtosis = 0 (Çarpıklık Katsayısı)

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.

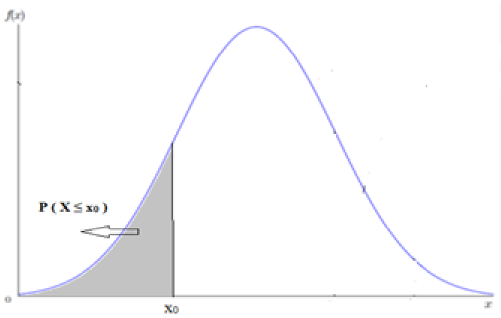


Şekil A'da normal dağılıma sahip ortalamaları aynı varyansları farklı iki tane olasılık yoğunluk fonksiyonu verilmektedir. **Şekilden de görüldüğü üzere varyans azaldıkça olasılık yoğunluk fonksiyonu sivrileşmektedir (Basıklığı azalmaktadır).**

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.

Rassal Değişkenlerin Kümülatif (Birikimli) Olasılık Fonksiyonu:

Ortalaması μ ve varyansı σ^2 olan normal dağılıma sahip bir X rassal değişkeninin Kümülatif olasılık fonksiyonu $F(x) = P(X \leq x_0)$ olarak gösterilir. X rassal değişkeninin belli bir x_0 değerinde küçük olma olasılığını gösterir.

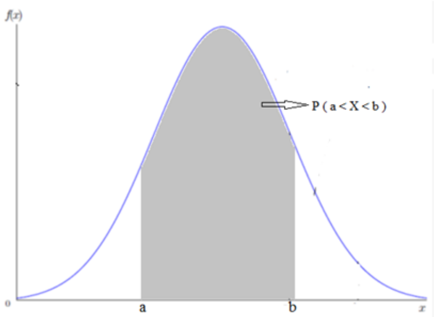


Yukarıdaki şekilde gri taralı alan, normal dağılıma sahip X rassal değişkeninin herhangi bir x_0 değerinden küçük olma olasılığını vermektedir. Bu alan aşağıdaki genel formül yardımıyla bulunabilir.

$$F(x) = P\ (X \leq\ x_0) = \int_{-\infty}^{x_0} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\,\mu)^2}{2\sigma^2}} \quad dx$$

Rassal Değişkenlerin Aralıklı Olasılık Fonksiyonu:

Ortalaması µ ve varyansı σ2 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).



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.

$$f(x) = P\ (a < X < b) = \int_{a}^{b} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\,\mu)^2}{2\sigma^2}} \quad dx$$

Örnek:

Normal dağılıma sahip rasgele değişkenin olasılık yoğunluk fonksiyonunun grafiğini çizdirecek Matlab programını yazınız.

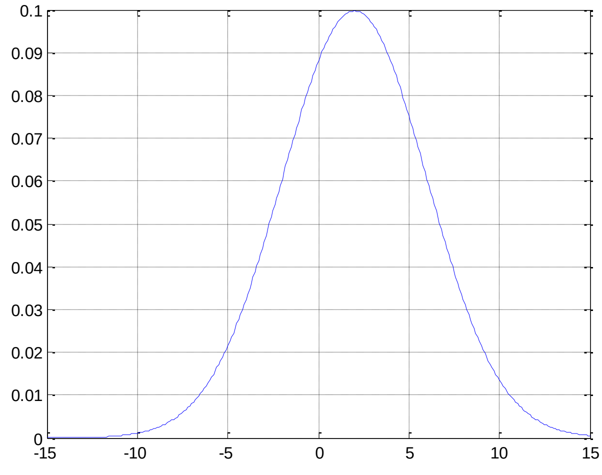
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}x^2}$$

Çözüm -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)
```

Çözüm -2:
```
x=-4:.1:4;
ort=0;
std=1;
figure, plot(x,normpdf(x,ort,std))
```

Örnek:

m=2 ve $\sigma^2 = 16$ için olasılık yoğunluk fonksiyonunun grafiği,
```
clc; clear all; close all
x=-15:.1:15;
plot(x,normpdf(x,2,4))
grid on
```



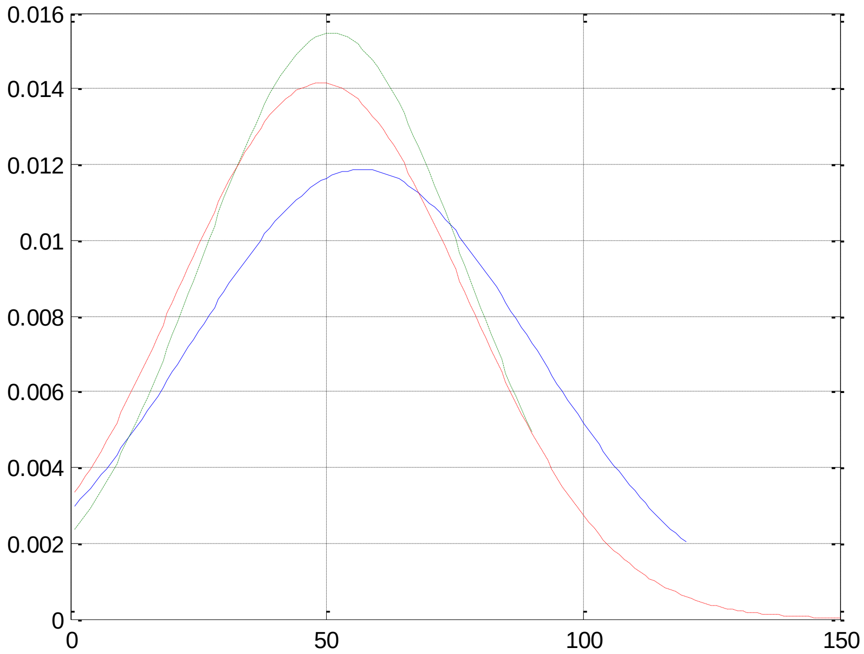
Örnek: Normal dağılımların çizilmesi

N1 = 40

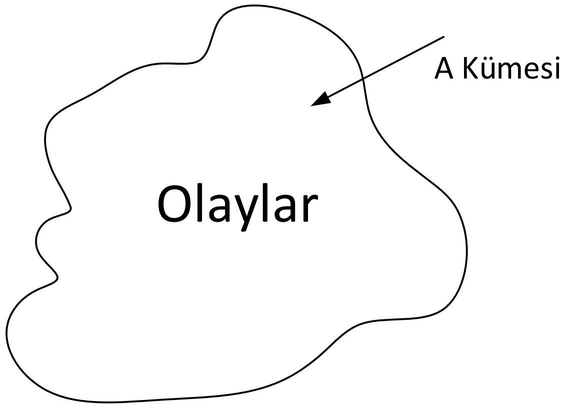
N2 = 30

N3 = 50

```
a1 = 1:3*N1;
a2 = 1:3*N2;
a3 = 1:3*N3;
ort1 =   56.7750
ort2 =   51.0333
ort3 =   48.9200
Std1 =   33.5960
Std2 =   25.8090
Std3 =   28.1822
figure,
plot(a1,normpdf(a1,ort1,Std1),'-',a2,normpdf(a2,ort2,Std2),'-.',a3,normpdf(a3,ort3,Std3),'--')
grid on
```



7.2. Standart normal dağılım



A Kümesi rassal değişkenlerin olasılıklarından oluşmaktadır. Sistem olarak da tanımlanır. Veri yığı dizi, vektör (Şiddet, yön), matris olarak verilmektedir.

Olasılık Yasaları:

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. Bir kümedeki herhangi bir olayın olmama olasılığı p(A’)= 1- p(A).

Aritmetik Ortalama:

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

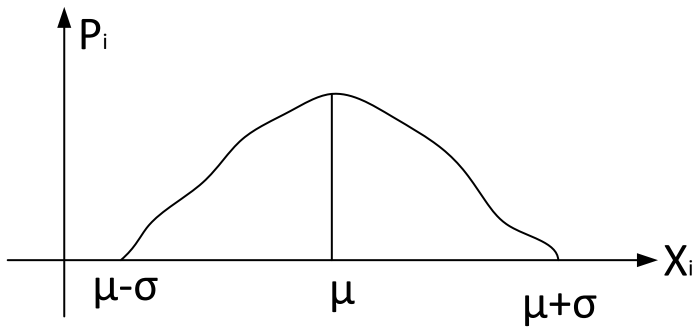
Medyan – Ortanca:

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.

Mod (Tepe) değeri:

Mod dağılım kümesinde olasılığı en yüksek değerdir. Dizide en çok tekrarlanma sayısıdır. 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.

Varyans - Standart Sapma:



Standart sapma, değişkenlerin aritmetik ortalamasından kaynaklanan kök ortalama karesi **(RMS)** sapmasıdır. **Veri yığınındaki herbir 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.

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.**

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 herbir 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.

Ana kütlenin (Büyük veri yığını) standart sapması:

Ana Kütle için Varyans : $$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{N}$$

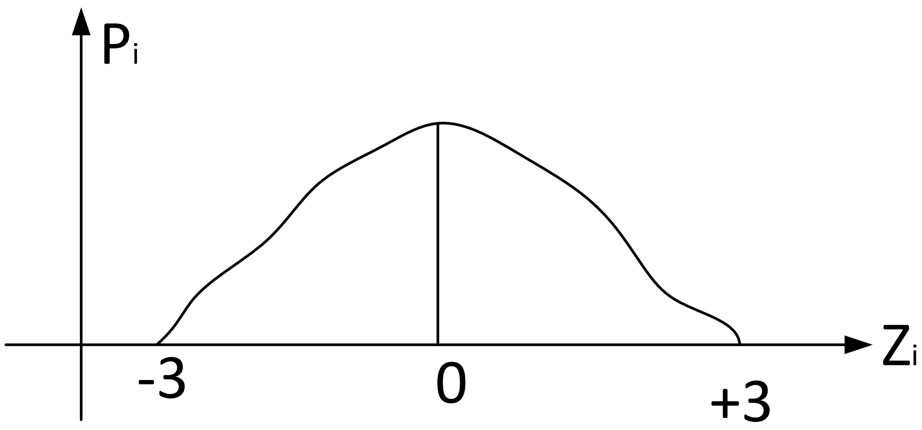
Burada standart sapma ya da arimetik ortalama ve rassal değişkenler arasındaki ilişkinin yorumlanması gerekmektedir. Eğer standart sapma değeri 0 ile 1 arasında ise aritmetik ortalama dizi içerisindeki veri değerleri arasında yakın bir ilişki vardır, aritmetik ortalama diziyi temsil eder. Eğer 1'den büyük ise kareler toplamı çok daha büyük olacağından temsil etmez.

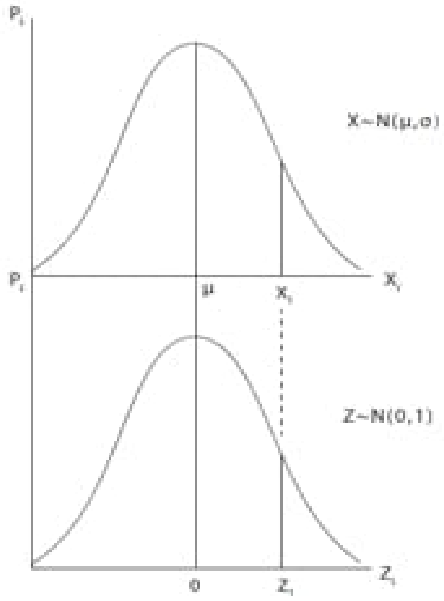
İntegral hesaplarını yaparak olasılıkları hesaplamak yerine, z dönüşümü ile herhangi bir normal dağılımı standart normal dağılıma dönüştürmek ve standart normal dağılım için hazırlanmış olasılık değerleri tablosunu kullanarak rassal değişkenlerin olasılığı kolaylıkla hesaplamak mümkündür.

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,

$$\boldsymbol{Z_i = \frac{X_i - \mu}{\sigma}}$$

formülüyle 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)





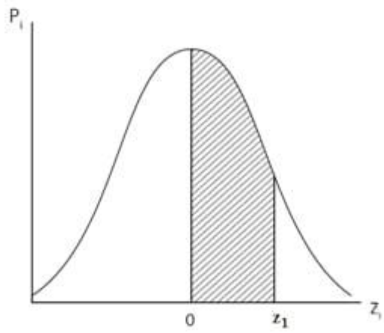
Şekil: Normal Dağılım ve Standart Normal Dağılım

Standart Normal Dağılım Tablosu Kullanılarak Olasılıkların Belirlenmesi:
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.

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.

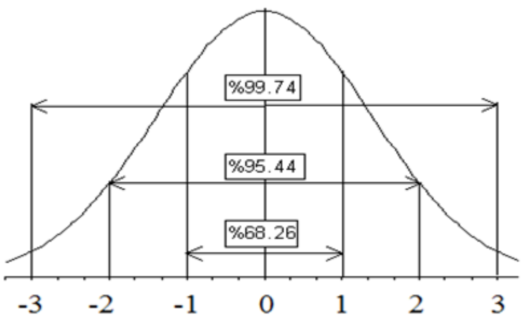
Standart normal eğri alanları tablosu sadece pozitif z değerleri için hazırlanmıştır. Normal dağılımın simetrik olması ve aritmetik ortalamanın eğri alanını iki eşit kısma ayırması sebebiyle, z değerinin pozitif ya da negatif olması sadece ortalamanın altında ya da üstünde bir değer olduğunu ifade eder, + ya da – bölgede olsa da ortalamaya uzaklık aynı olacağı için olasılık değerleri de aynı olacaktır.



Şekil: Standart Normal Dağılım Tablosunda Okunan Olasılıkların

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 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 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.

Örnek: Çan eğrisi ile not hesaplama:

Bu yöntemin kullanılmasındaki temel amaç, belirli bir dersi alan öğrencileri 100 puan üzerinden aldıkları mutlak notlara göre değil, o dersi alan grup içinde gösterdikleri performansa göre notlandırmaktır. Performans, bir öğrencinin sınıf ortalamasına göre hangi konumda olduğudur. Bu sistemde sınıfın ortalaması ve standart sapması çok büyük önem arz ediyor. Notlar hesaplanırken;

Z=(HBN - Sınıfın HBN aritmetik ortalaması) / Standart sapma

formülüyle aldığınız notlar Z Standart Notuna ve Z standart notları da (T=10Z+50) formülüyle T Notuna dönüştürülür. Hesaplanan T Notu (TSkor), yeni not olur.

HBN:Ham Başarı Notu

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

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.

Örnek:

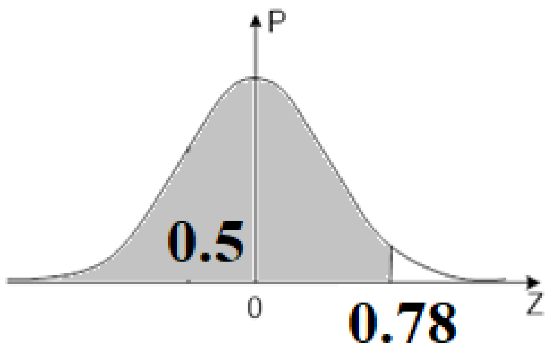
Z<=0.78'deki P olasılık değerini bulunuz. Tablodan olasılık değerini bulabilmek için Z=0.7 + 0.08 ayrımı yapılır. 0.7 değeri en soldaki değer; 0.08 değeri ise en üstteki değer olarak alınır.

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



Z>0.78'deki P olasılık değeri bulunurken Z<=0.78'deki P olasılık değeri bulunur. Bulunan değer 1 den çıkarılır.

Örnek:

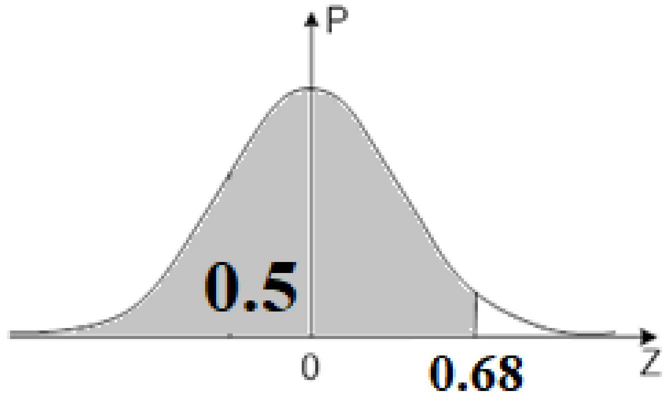
Z<=0.68'deki P olasılık değerini bulunuz. Z=0.6 + 0.08 ayrımı yapılır. 0.6 değeri en soldaki değer; 0.08 değeri ise en üstteki değer olarak alşınır.

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

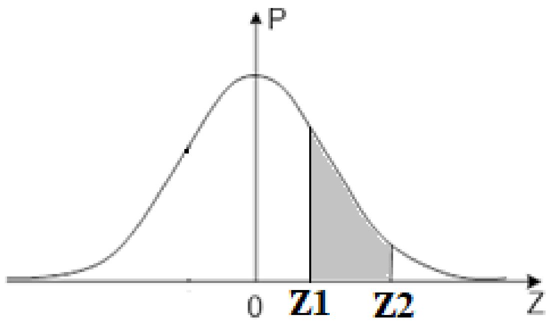


Örnek:
0.52<= Z<=0.78'deki P olma olasılık değerini bulunuz.

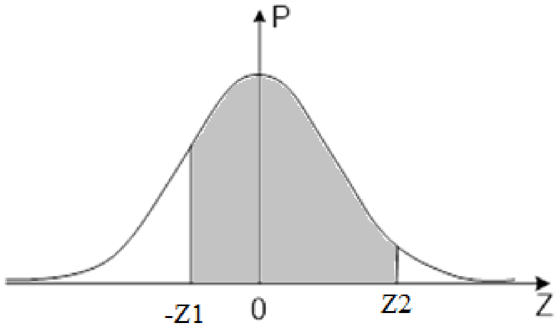
Tabloya bakıldığında iki olasılık değerinin tablodan bulunacağı görülmektedir.
1- P1(0 – Z1)
2- P2(0 – Z2)

P(Z2 – Z1)= P2- P1
Taploadan Z2=0.78 için P2= 0.2823
Taploadan 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



Örnek:
-0.52<= Z<=0.78'deki P olasılık değerini bulunuz.
P(-0.52<= Z<=0.78)= P2(0<Z<0.78) + P1(0<Z1<0.52) = 0.2823 + 0.1985 =0.4808
%48,08



Örnek:
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.

$$Z = \frac{X_i - \mu}{\sigma}$$

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

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.

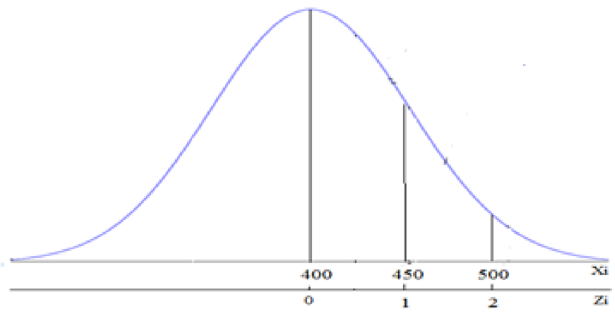
Standart normal dağılım tablosundan Z <=0.4 (X<=60) için olasılık ne olur? Yüzde olarak ifade ediniz. Tablodaki değerler 0.5 olasılık olarak ortadan ikiye ayrılmaktadır. Bu nedenle tablodan bulunan değere 0.5 eklenir.

P=0.1554 + 0.5 = 0.6554 bulunur. %65,54

Örnek:

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 Normal dağılımın standart dağılımına çevrilebilir olması için Z'in maksimum +/- 3 değerleri arasında olması gerekir.

a)

Z = $\frac{X-\mu}{\sigma} = \frac{500-400}{50} = 2$, 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

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) 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) 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) Öğrenci gelirlerinin 450TL den küçük olma olasılığı nedir?

$Z=(450-400)/50=1.00$

Standart Normal Dağılım Tablosu

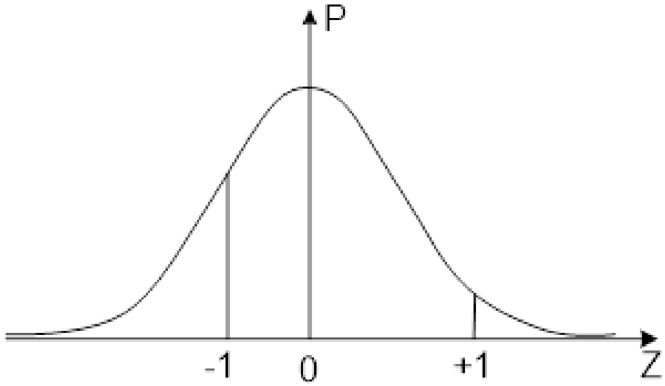
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830

Z-Tablosundan, $P(Z<1.00)= P(Z<0)+ P(0<Z<1)=0.5+0.3413=0.8413$ elde edilir.

e) Öğrenci gelirlerinin 350TL ile 500 TL arasında olma olasılığı nedir?

$Z1=(350-400)/50=-1.00$

$Z2=(500-400)/50=2.00$



Z-Tablosundan,

$P(0<Z1<1) = 0.3413$

$P(Z1<1) = P(Z1<0) + P(0<Z1<1) =0.5+0.3413=0.8413$

$P(Z1<-1)= 1-P(Z1<1)=1-0.8413=0.1587$

$P(0<Z2<2) = 0.4772$

$P(Z2<2) = P(Z2<0) + P(0<Z2<2) =0.5+0.4772=0.9772$

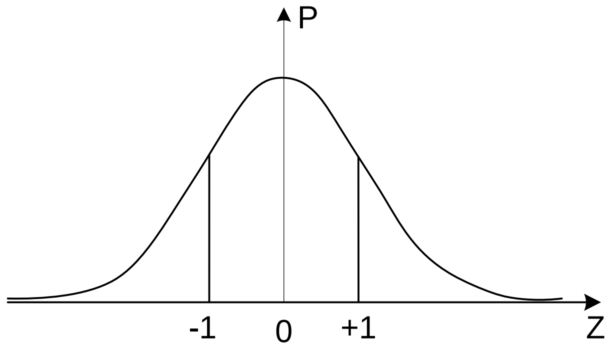
$P(-1<Z<2)=P(Z2<2)-P(Z1<-1)= 0.9772-0.1587=0.8185$

O halde Öğrenci gelirlerinin 350TL ile 500 TL arasında olma olasılığı %81.85 dır.

f) Öğrenci gelirlerinin 350TL ile 450 TL arasında olma olasılığı nedir?

Z1=(350-400)/50=-1.00

Z2=(450-400)/50=1.00



Z-Tablosundan,

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<1) = 0.3413

P(Z2<1) = P(Z2<0) + P(0<Z2<1) =0.5+0.3413=0.8413

P(-1<Z<2)=P(Z2<2)-P(Z1<-1)= 0.8413-0.1587=0.6826

O halde Öğrenci gelirlerinin 350TL ile 500 TL arasında olma olasılığı %68.26 dır.

Örnek:

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) 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) 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) 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) Hasta gelirlerinin 10.000TL den fazla olma olasılığı nedir?

Z=(10.000-20.000)/5.000=-10.000/5000=-2

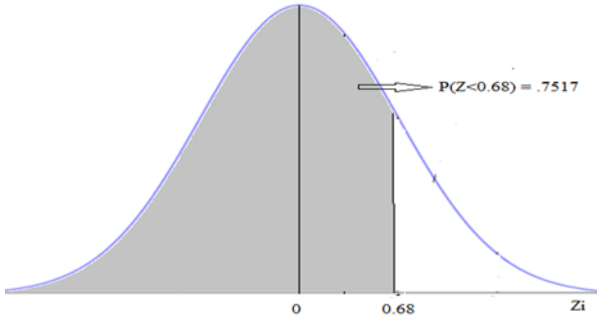
Tablonun simetrik özelliğinden dolayı Z=2 alınır. (0 ile 2 arasındaki olasılık değeri hesaplanır.)

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

Örnek:

Standart normal dağılıma dönüştürülen bir normal dağılımın P(Z<0.68) olduğunu varsayalım.



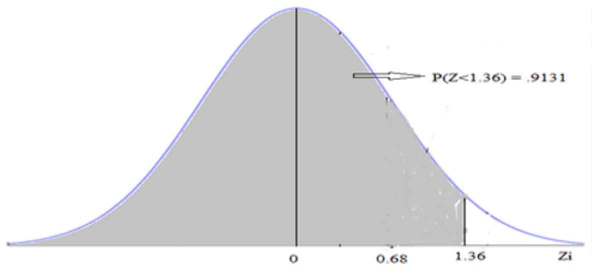
Standart Normal Dağılım Tablosu

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

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.

Örnek:

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

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

Örnek:

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 şeklinde bulunur.

Örnek:

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); çünkü simetri vardır.

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

Örnek:

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

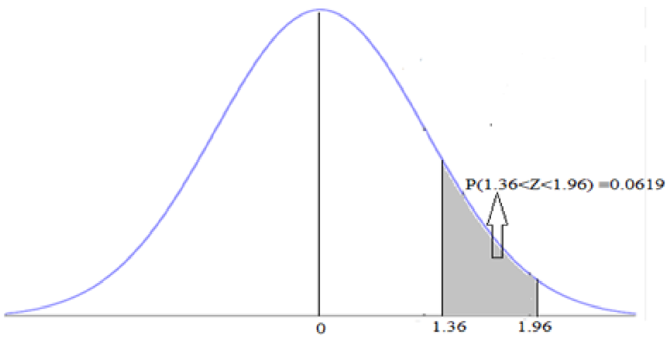
Standart Normal Dağılım Tablosu

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

- Bu aralıklardaki olasılık P(1.36<Z<1.96)= P(Z<1.96) - P(Z<1.36) = 0.9750 - 0.9131 =0.0619 şeklinde bulunur.



Örnek:

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

Örnek:

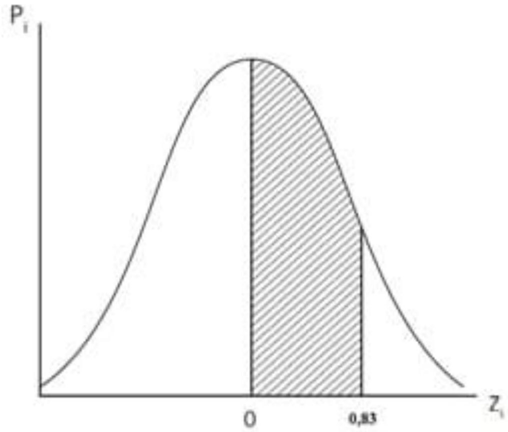
Bir sınıfta öğrencilerin aldığı puanlar kayıt altına alınmıştır. Öğrencilerin aldığı puanların ortalaması 60 ve standart sapması 12 hesaplanmıştır. Puanların normal dağılımı bilindiğine göre, aşağıdaki şıklarda istenen olasılıkları hesaplayınız.

1) Tesadüfen seçilecek bir öğrenci 60 ile 70 arasında puan alma olasılığı nedir?

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.



Şekil: Ortalama ile ortalamadan büyük bir değer arasındaki alan için olasılık hesabı

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.

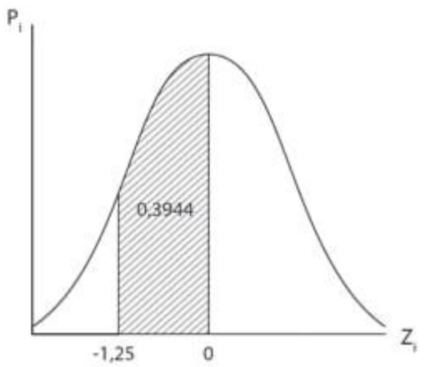
2) Tesadüfen seçilecek bir öğrencinin 45-60 arasında puan alma olasılığı nedir?

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.

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.



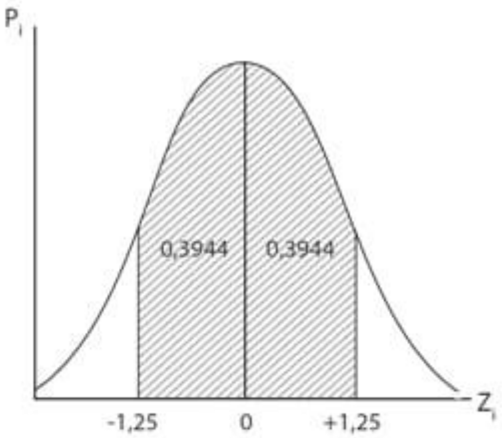
Şekil: Ortalama ile ortalamadan küçük bir değer arasındaki olasılığın belirlenmesi

3) 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)

Tablodan, P(0 ≤ Z ≤ 1,25)=0.3944 bulunur.

P(-1,25 ≤ Z ≤ 1,25) = 2*P(0 ≤ Z ≤ 1,25)=0.7888

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.

4) Tesadüfen seçilecek bir öğrencinin 80 fazla puan alma olasılığı nedir?

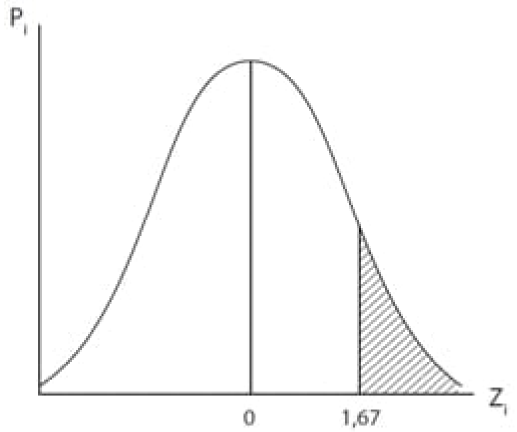
80 adete karşılık gelen Z değeri, Z=(80-60)/12=1.67 olarak hesaplanır. Öğrencinin 80'den fazla puan alma olasılığı istendiğine göre, P(X > 80) olasılığının hesaplanması gerekiyor.

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.



Şekil: P(z > 1,67)

5) 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.

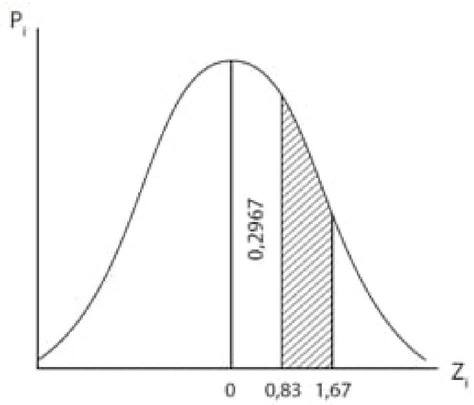
Şekil: P(-1,25 < z)

6) 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

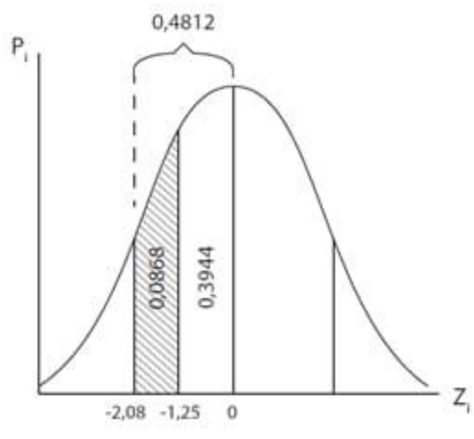
Dolayısıyla, seçilecek herhangi birinin 70 ile 80 arasında puan alma olasılığı % 15,58 olarak hesaplanmakta, başka bir deyişle öğrencilerden % 15,58’i 70-80 arasında puan almaktadır.

7) 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



Şekil: 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

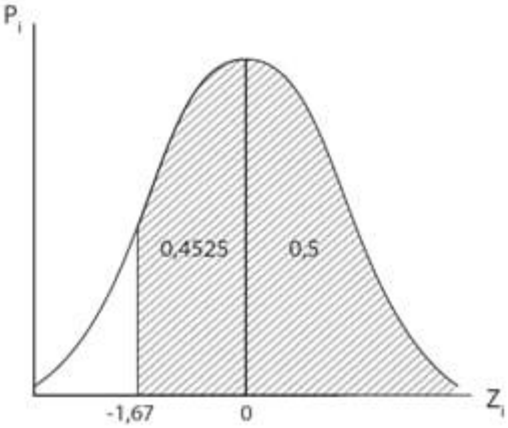
Bu değer aynı zamanda, öğrencilerin % 8,68’inin 35-45 aralığında puan aldğının göstergesidir.

8) 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 olarak bulunur.



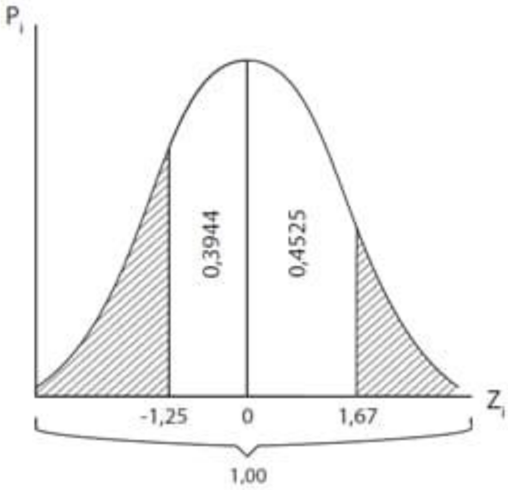
Şekil: P(z > -1,67)

P(Z ≥ -1.67) = 0.5 + 0.4525=0.9525

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.

9) Tesadüfen seçilecek bir öğrencic 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



Şekil: P(-1,25 < z ; z > 1,67)

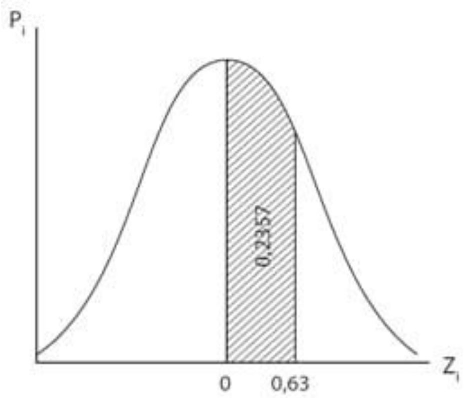
$P(Z \geq 1.67) = 1 - P(Z \leq 1.67)$
$P(Z \leq 1.67) = 0.5 + P(Z=1.67) = 0.5 + 4525 = 0.9525$
$P(Z \geq 1.67) = 1 - 0.9525 = 0.0475$

$P(Z \leq -1.25) = P(Z \geq 1.25)$
$P(Z \geq 1.25) = 1 - P(Z \leq 1.25)$
$P(Z \leq 1.25) = 0.5 + P(Z=1.25) = 0.5 + 0.3944 = 0.8944$
$P(Z \geq 1.25) = 1 - 0.8944 = 0.1056$
$P(X < 45 \text{ veya } X > 80) = 0,1531$ olarak hesaplanır.

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.

Örnek:
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,

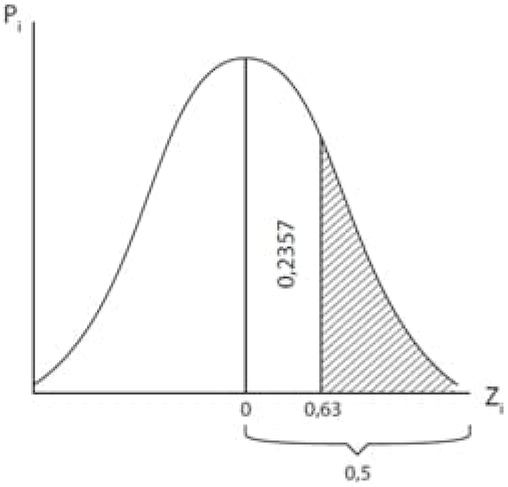
1) Tesadüfen seçilecek bir margarinin 250-255gr arasında bulunma olasılığını hesaplayabilmek için öncelikle, 255değerine karşılık gelen standart Z-değerini hesaplamalıyız.
   $Z_{255}=(255-250)/8=0.625$
   $P(0.625 \geq Z \geq 0)=P(Z=0.625) - P(Z=0)=0.2357 - 0 = 0.2357$ bulunur.
   Tesadüfen seçilecek bir margarinin 250-255 gr arasında bulunma olasılığını % 23,57 olarak hesaplamış olunur.



Şekil: P (0 ≤ z ≤ 0,63)

2) Tesadüfen seçilecek bir margarinin 255 gr’dan fazla olma olasılığını bulmak için,

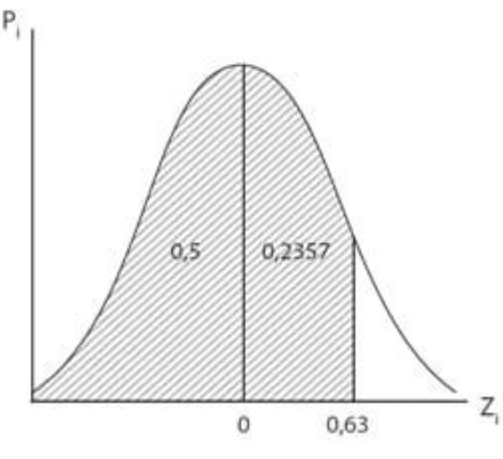
$P(X > 255) = 1\text{-}P(X<255)$
$P(Z>0.625)=1\text{-}P(Z<0.625)=1\text{-}0.5\text{-}0.2357=0.2643$ hesaplanır.
Tesadüfen seçilecek bir margarinin 255 gr’dan fazla olması olasılığı % 26,43’tür. Başka bir deyişle, bu fabrikada üretilmekte olan margarinlerin % 26,43’ü 255 gramın üzerinde ağırlığa sahiptir.



Şekil: P (z > 0,63)

3) 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.



Şekil: P (z < 0,63)

7.3. Normal Dağılım Olasılıkları Kullanılarak Beklenen Değer Hesabı

Normal dağılıma sahip tesadüfi değişkenlerin belirli bir aralıkta değer alma olasılıklarını hesaplamayı öğrendik. Bunun yanında, kitlenin içinde kaç birimin ilgili aralıkta değer alacağını da belirleyebiliriz. Başka bir deyişle, hesapladığımız olasılık değeri üzerinden, ilgili aralıkta yer alacak birimlerin frekansının ne olacağı hesaplanabilir.

Olasılık bölümü içinde öğrendiğimiz beklenen değer kavramını hatırlayalım. Beklenen değer, gözlem sayısı ile olasılığın çarpımıdır. Ağırlık ortalama değerine denktir:

$$E(X) = n.p$$

Dolayısıyla, normal dağılım çözümlemelerinde de hesaplanan olasılık toplam frekans ile çarpılarak ilgili olasılık alanına düşen gözlem sayısı ya da birim sayısı kolaylıkla hesaplanabilir.

Örnek:
Margarin fabrikası örneğinde bir günde toplam 1000 adet margarin üretimi yapıldığını varsayalım.

Tesadüfen seçilecek bir margarinin 250-255 gram arasında olması olasılığını 0,2357 olarak hesaplamıştık. Bu durumda, günlük üretim içinde 250-255 gram aralığında ağırlığa sahip margarin ağırlık ortalaması da,

$E(X) = n.p$

$E(X) = 1000 \,.\, 0{,}2357 = 235{,}7 \cong 236$ gram olarak hesaplanacaktır.

Benzer şekilde, kaç adet margarinin 255 gramdan fazla olacağını da belirleyebiliriz. Bunun için, tesadüfen seçilecek bir margarinin 255 gramdan fazla olması olasılığı olan 0,2643 değerini toplam frekans ile çarpmamız yeterli olacaktır.

$E(X) = n.p$

$E(X) = 1000 \,.\, 0{,}2643 = 264{,}3 \cong 264$

Dolayısıyla bir gün içinde üretilen 1000 adet margarinin 264 tanesinin 255 gramdan daha fazla ağırlığa sahip olmasını bekleyebiliriz.

Örnek:

Bir konserve fabrikasında üretilmekte olan konservelerin ağırlığı normal dağılıyor. Ortalama ağırlık 500 gr ve standart sapma 10 gr olduğuna göre tesadüfen seçilecek bir konserve kutusunun,

1. 525 gramdan fazla olması olasılığı nedir?
2. Günde 5000 adet konserve üretilmesi durumunda kaç konservenin 525 gramdan fazla olması beklenir?

1) Tesadüfen seçilecek bir konserve kutusunun 525 gramdan fazla olması olasılığını hesaplayabilmek için öncelikle 525 değerine karşılık gelen z değerini hesaplıyoruz.

$$z_{525} = \frac{X_i - \mu}{\sigma} = \frac{525 - 500}{10} = +2,5$$

Standart normal eğri alanları tablosunda z=2,5 standart değerine karşılık gelen alan, tablonun ilk sütununda bulunan 2,5 değeri ile ilk satırdaki 0,00 değerlerinin kesişim noktasında yer alan 0,4938 olasılığını vermektedir. Bu değer Z=0 ile Z=2.5 arasındaki değerlerin olma olasılığını verir. 0 ile 3 arasındaki değer 0.5 ise

Bu durumda, tesadüfen seçilecek bir kutunun 525 gramdan fazla olması olasılığı,
0,5 – 0, 4938 = 0,0062 olarak hesaplanmaktadır.

Tesadüfen seçilecek bir konservenin 525 gramdan fazla olması olasılığı % 0,62 olmaktadır.
b-Günde 5000 konserve üretilmesi halinde,
E(X) = n.p = 5000 . 0,0062
E(X) = 31 adet konservenin 525 gramdan fazla olması beklenebilir.

Örnek:

50 öğrencinin bir dersten aldığı puanlar verilmiştir. Ağırlık ortalamasını ve standart sapmasını hesaplayınız. Çözümde Matlab kullanılmıştır.

```
clear all
close all
x = [20 30 50 80 90 60 70 20 60 50 80 70 40 90 30 40 50 40 80 90 65 45 15 35 85 85 90 60 75
75 40 50 60 80 70 80 95 100 65 75 85 35 75 50 75 85 45 80 95 100];
N = length(x); Mx=max(x); Mn=min(x)
ort = mean(x); M = median(x); Std=std(x)
k = kurtosis(x); s = skewness(x)
sort1=sort(x);

N =    50
Mx =   100
Mn =    15
ort =   64.2000
M =    70
Std =   22.8429
k =    2.1097
s =   -0.3833
```

70-80 arasında puan alan öğrenci sayısını standart normal dağılım ile hesaplayınız.

Öncelikle 70 ve 80 puanlarının standart değer karşılıklarını bulalım:

$$Z_{70} = \frac{70 - 64.2}{22.8429} = 0.26$$

$$Z_{80} = \frac{80 - 64.2}{22.8429} = 0.70$$

Her iki z değeri de pozitiftir ve ortalamanın sağında kalan bölgede yer almaktadır.

P(0,26 ≤ z ≤ 0.70)

0 ≤ z ≤ 0.26 alanı standart normal dağılım tablosundan 0.1026 + 0.5=0.6026

0 ≤ z ≤0,70 alanı standart normal dağılım tablosundan 0,2580 + 0.5=0.7580 olarak belirlenir.

P(0,25 ≤ z ≤ 0.69)=0.7580 – 0.6026 = 0,1554 hesaplanır.

Öğrencilerin 70 ile 80 arasında puan alma olasılığı % 15,54 olarak hesaplanmakta, başka bir deyişle öğrencilerin % 15,54’i 70-80 arasında puan almış olmaktadır. O halde toplam öğrenci sayısı 50 olduğuna göre 70 ile 80 puan arasında not alan öğrenci sayısı=50*15.54/100=7.77 yaklaşık 8 öğrenci olarak bulunur.

7.4. Örneklem Büyüklüğünün Belirlenmesi

Büyük bir veri yığınını (Anakütle) temsil edebilecek örnek veri yığını (örnek kütle) oluşturma işlemine örnekleme, oluşturulan temsili veri yığınına da örneklem denir.

Anakütle ya da büyük veri yığını, araştırmaya konu olan tüm verilerin oluşturduğu kütledir. Bir istatistik araştırma ya da incelemede, araştırma konusu ile ilgili bütün verilerin oluşturduğu topluluğa anakütle denir.

Parametre: Anakütleyi karakterize eden ya da başka deyişle niteleyen ve bu anakütleyi diğer anakütlelerden ayırt etmeye yarayan özelliklere ve ölçülere parametre denir.

Örnek kütle: Anakütle içinden anakütleyi temsil etmek üzere seçilmiş daha az sayıda birimden oluşan numune, topluluk.

Anakütleyi oluşturan bütün veriler gözlenerek elde edilen tüm ölçüler parametre değerleridir ve o anakütle için kesin değerlerdir. Dolayısıyla, tam sayım yapılması durumunda parametreye yönelik bir tahmin işlemi söz konusu değildir ve hesaplanan değer doğrudan parametre değeri olmaktadır.

Ancak, bazı hallerde tüm birimlerin gözlenmesi ya da incelenmesi çok maliyetli olmakta veya çok zaman almaktadır. Öte yandan, anakütlenin kaç birimden oluştuğunun tam olarak bilinmediği ve sonsuz birim içeren anakütleler şeklinde nitelenen anakütlelerde, anakütlenin birimlerinin tamamını gözlemek yani tam sayım yapmak mümkün olamamaktadır. Sonuç olarak, yukarıda çeşitli örneklerle açıklamaya çalıştığımız sebeplerle, anakütlenin tamamını gözlemek suretiyle tam sayım yapmak ve anakütleyi karakterize eden değerler olarak nitelediğimiz parametre değerlerini doğrudan belirlemek mümkün olmaz. Böyle durumlarda örnekleme yöntemlerine başvurulur.

Çok sayıda birim içeren bir anakütleden, anakütleyi temsil edecek nitelik ve yeterlilikte örnek birimler seçme ve seçilen birimler üzerinden anakütle parametresini tahmin etme işlemine örnekleme denir. Diğer bir deyişle örnekleme, bir araştırmada anakütleyi oluşturan birimler arasından, anakütlenin yapısını ve özelliklerini yansıtacak şekilde örnek seçme ve seçilen bu örnek içinde yer alan birimleri incelemek suretiyle anakütle parametrelerini tahmin etme sürecidir. Örneğin, bir çuval dolusu pirinç içinden alınacak bir avuç pirincin incelenmesiyle tüm çuval dolusu pirinç hakkında çeşitli özellikleri itibariyle genelleme yapılabilir. Burada, çuvaldaki toplam pirinç miktarı ana kitle (yığın)'dir. İncelemek amacıyla aldığımız bir avuç dolusu pirinç ise örnektir.

Örnekleme kuramı, ana kitleyi oluşturan her davranışın örneğe girme şansının eşit olmasını öngörür. Örnekleme kuramına göre, örneklerden elde edilen bilgilerin matematik ve istatistik tekniklerle test edilip genelleme yapılabilmesi için örneklemenin uygun olarak yapılması gerekir. Ana kitle hakkında tahminde bulunurken yapılan tahminin geçerli olabilmesi için örnekleme tesadüfi (rastgele) olabillir. Tesadüfi örnekleme kavramı, örnekte yer alacak değerlerin belirlenmesinde hiçbir dış etkinin rolünün olmamasını ifade eder.

Elde edilecek bilgi, ana kitleyi oluşturan değişik özellikteki gruplara göre farklılık gösteriyorsa, ana kitleyi oluşturan sınıflara göre örnekleme yöntemi kullanılır. Kümelere göre örneklemede kitle önce kümelere ayrılır, sonra kümelerden örnekler seçilir. Bu tür örnekleme, sağlıklı bir ana kitle çerçevesinin elde bulunmaması ya da çok büyük ana kitleden çekim yapmanın çok zor ve maliyetinin yüksek olması durumunda uygulanır. Bu tür örnekleme tesadüfi örnekleme olmasına karşın, ana kitleyi ne ölçüde temsil ettiği tartışılabilir. Bu nedenle de çok dikkatli yapılması gerekir.

Sistematik örnekleme, seçim işlemlerinin kolay olması nedeniyle özellikle ana kitlenin büyük olduğu durumlarda kullanılan bir örnekleme yöntemidir. Çok sayıda birim içeren kayıt sistemlerinin incelenmesinde. *Örneğin,* hasta dosyaları, hasta ya da işçi kayıtları, kayıt defterleri, fişler , listeler gibi. Bu yöntemde başlangıç sayısı dağılımı büyük oranda etkiler.

Örnekleme Çalışmalarında Ortaya Çıkabilecek Temel Hata Türleri

Örnekleme çalışmalarında ortaya çıkabilecek temel hata türleri tesadüfi ve sistematik hatalar olmak üzere iki ana başlıkta toplanmaktadır.

Tesadüfi hatalar, farklı yönlerde ve genellikle zıt yönlerde ortaya çıkan hatalar olup, genel olarak değerlendirildiğinde çalışmanın bütününe yönelik etkisi düşük düzeyde olan hatalardır. Örneğin, sayım görevlilerinin bir kadını erkek, ya da bir erkeği kadın olarak kodlaması. Bu tür bir hata hep aynı yönde gerçekleşmez. Yani, devamlı olarak erkekler kadın ya da kadınlar erkek olarak kodlanmaz. Dolayısıyla, her iki yönde de yapılabilecek bu tip bir hatanın çalışma üzerinde etkisi, zıt yönde gerçekleşen hatalar birbirini yok edeceği için, çok düşük olacaktır.

Sistematik hatalar ise hep aynı yönde tezahür eden, gerçekleşen hatalardır. Bu tür hatalara örnek olarak, genç görünmek arzusuyla yaşın olduğundan küçük söylenmesi ya da vergi korkusu ile gelirin olduğundan düşük beyan edilmesi verilebilir. Sistematik hataları tesadüfi hatalardan ayıran en temel özellik, hataların hep aynı yönde gerçekleşmesi ve birim sayısını arttırmak suretiyle bu tür hataların azaltılamamasıdır. Oysa, birim sayısı arttıkça tesadüfi hatalar hem oransal olarak azalmakta hem de artı ve eksi yönde gerçekleştikleri için birbirlerinin etkisini nötr hale getirmektedirler.

Örneklemenin amacı, daha yüksek doğrulukta, daha az zaman ve daha az maliyetle anakütle hakkında bilgi sahibi olmaktır diyebiliriz. Örnek seçme işleminin nasıl yapılacağına yönelik, anakütlenin yapısını, birim sayısını ve araştırma yapılan konuyu temel alan çok sayıda örnekleme yöntemi mevcuttur.

Anakütlenin tam sayımının yapılması maliyetli, zaman alıcı ve çoğu zamanda olanaksız olması nedeniyle örnekleme sürecine başvurulmaktadır. Örnekleme, örnek istatistiğinden anakülte parametresinin tahmin edilmesi süreci olarak adlandırılabilmektedir.

Örnek büyüklüğünün belirlenmesi:
Örnek büyüklüğünün hesaplanmasında aşağıdaki formüller kullanılabilir.
Ana kitle varyansı biliyorsa,

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

*A*na kitle varyansı bilinmiyorsa

$$n = \frac{Z^2 s^2}{(\bar{x}-\mu)^2}$$

Burada,
Z: Belirlenen güven düzeyi için Z tablosundan bakılan değer
σ^2: Ana kitle varyansı
S^2: Örnekleme varyansı
$\bar{x}$: Örnekleme ortalaması
μ: Ana kitle ortalaması
$(\bar{x}-\mu)$: Göze alınan örnekleme hatası

Örnek : Bir iş kolunda çalışan işçilerin ortalama saat ücretleri gerçek ana kütle ortalama saat ücretinden en fazla 100 TL sapma gösterecek şekilde ve %90 güven aralığında tahmin edilmek isteniyor. Geçmiş kayıtlara dayanarak bu iş kolu için hesaplanan standart sapma 500 olduğuna göre örnek büyüklüğü ne olmalıdır.

%90 güven aralığında olduğundan olasılık değeri 0.9 dur. 0.9 olasılık değerine karşılık Z değerini tablodan bulabilmek için 0.9/2=0.45 değeri alınır. Çünkü tabloda verilen değerler standart olasılık yoğunluk fonkisyonun yarısına gçre verilmiştir. O halde 0.45 değerine karşılık gelen Z değeri 1.64 dür.

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
1.4	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

$$n = \frac{Z^2\sigma^2}{(\bar{x}-\mu)^2} = \frac{(1.64)^2(500)^2}{(100)^2} = 67.25$$

Bu sonuca göre gerçek ortalamaının ±100TL sınırları şçşnde ve %90 güven aralığında istenen amacı gerçekleştirmek için örnek büyüklüğü 68 olmalıdır.

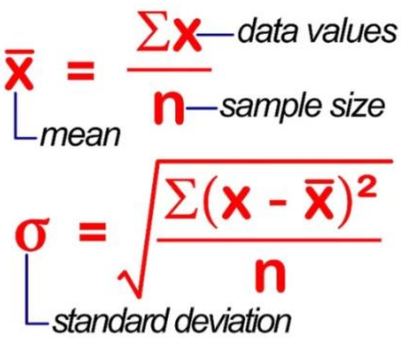
Örnek büyüklüğü de belirlendikten sonra örnekleme planı oluşturulur. Örnekleme yöntemlerine dayalı olarak yapılan tahminlerde örnekleme hataları yapılır. Bunlar örnek sayısının yeterli olmamasından kaynaklanan tesadüfi olarak yapılan hatalar ve sistematik olarak adlandırılan hatalar ise; örnekleme sürecindeki hatalardan kaynaklanır ve sonradan giderilmeleri mümkün değildir. Çünkü,

- Örnekleme yöntemi doğru seçilmemiştir,
- Ana kitle yanlış tanımlanmıştır,
- Örnekleme çerçevesinin yanlış belirlenmiştir,
- Örnekler doğru seçilmemiştir,
- Örnek büyüklüğü doğru hesaplanmamıştır.

7.5. Güvenilirlik Aralığı

Büyük veri yığınından hipotez testi için örnek veri yığını alınsın. Büyük veri yığınının varyansı ve örnek veri yığının ağırlık ortalaması biliniyorsa belirli bir olasılığa göre büyük veri yığının ağırlık ortalamasının güven ağırlığı hesaplanmaktadır.

Ana kütle ortalamasının güven aralığı:
Ana kütleyi incelemek çoğu zaman olanaksızdır. Bu gibi durumda ana kütlenin ortalaması hesaplanmasının bir katma değeri yoktur. Güvenirlilik aralığı sıkıntılıdır. Ancak örneklemler yardımıyla ana kütlenin ortalamasının içinde bulunduğu sınırlar, seçilen yanılma olasılığında tahmin edilebilir. Örneklem kitlesine ait ortalamanın içinde bulunduğu sınırlara "**Ana kütlenin ortalamasının Güven Aralığı ya da Güven Sınırları**" adı verilir.



Not: Karesi almak demek,
Eğer fark 1 den küçükse karesi daha da küçülür.
Eğer fark 1 den büyüukse karesi daha da büyür.

Ana kütleden alınan küçük bir yığının normal dağılımından alınan n adet bir örneklemin ortalamasının ana kütlenin ortalamasına yani μ'ye eşit olması beklenemez. Bu durumda μ için güven aralığı tahmin edilebilir.

Ana kütlenin standart sapması, σ bilindiğinde ana kütlenin ortalamasının aralık tahmini:
Normal dağılımın simetri özelliğinden güven aralığının boyu (1-α), ortalamanın iki yanına eşit dağıtıldığında, başka bir ifadeyle α'nın $\alpha/2$ ve $\alpha/2$ şeklinde iki tarafa bölündüğü durumda minimum olacaktır. Dolayısıyla Güven Aralığı,

$$(X - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) \leq \mu \leq (X + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

Bu ifadede,

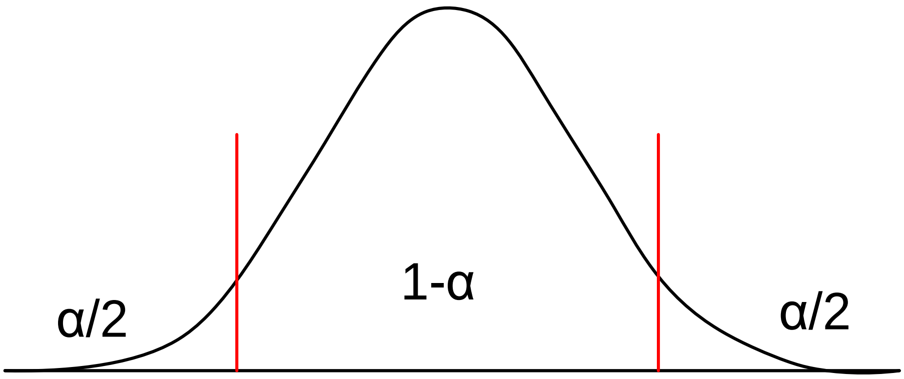
X: Büyük veri yığının ağırlık ortalaması,

Buna göre örneklem standart hata:

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

Verilen güven aralığı olasılığındaki Z değeri:

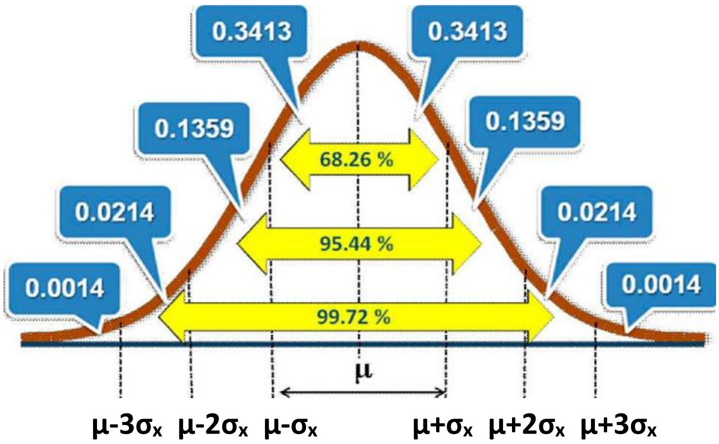
$$Z_{\alpha/2}$$



Örnek Ortalamaların Dağılımı:

Toplam birim sayısı N, varyansı σ^2, aritmetik ortalaması μ ve dağılımı normal olan bir ana kütle N(μ, σ^2) şeklinde gösterilir. Örneklerin basit tesadüfi örnekleme yöntemine göre alındıkları varsayılarak, örnekler için örnekleme dağılımları: Bir yığından ***n*** büyüklüğünde çekilmesi mümkün tüm örneklerin çekilsin ve her biri için aritmetik ortalamanın hesaplansın; C_n^N sayısı kadar olan bu örnek ortalamalarının dağılımı, teorik bir dağılımdır. Söz konusu dağılım, örnek ortalamalarından oluştuğu için buna "örnek ortalamalarının örnekleme dağılımı" adı verilir.

Örnek ortalamalar, ana kütle ortalaması etrafında normal bir dağılım gösterirler. Standart hata ise örnek ortalamaların normal dağılımının standart sapmasından başka bir şey değildir. Bu normal dağılımın ortalaması ile standart sapması arasındaki ilişkiden hareketle herhangi bir örnek ortalamasının belirli olasılık kademelerine göre bulunabileceği sınırlar tahmin edilebilir.



Şekil: Normal dağılan örnek ortalamalarının çeşitli standart hata sınırları

Şekilde görüleceği gibi, örnek ortalamalarının % 99.72'si ana kütle ortalamasından ± 3 $\boldsymbol{\sigma}_x$ standart hata sınırları arasında bulunur. Bunun gibi diğer olasılık kademeleri için de benzer açıklamalar yapılabilir. Buna göre, ana kütle ortalaması μ ve örnek ortalamaların standart hatasını $\boldsymbol{\sigma}_x$ şeklinde gösterilecek olursa örnek ortalamaları şu şekilde hesaplanır:

% 68.26' ü	**$\mu - \sigma_x$**	**ile**	**$\mu + \sigma_x$**	Sınırları arasında
% 95.44' i	**$\mu - 2\,\sigma_x$**	**ile**	**$\mu + 2\,\sigma_x$**	Sınırları arasında
% 99.7' si	**$\mu - 3\,\sigma_x$**	**ile**	**$\mu + 3\,\sigma_x$**	Sınırları arasında

Dağılımı normal ve varyansı belli olan yığından çekilen örnek ortalamaları ana kütle ortalaması etrafında normal dağılım gösterir. Bu dağılımın ortalaması (örnek ortalamalarının ortalaması) ana kütle ortalamasına eşittir ve $\overline{X}$ = μ şeklinde gösterilir.

Örnek ortalamaları dağılımının standart sapması, standart hata olarak bilinir ve aşağıdaki formülle hesaplanır:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

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:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

σ bilinmiyorsa, Standart sapma hatası, S_x verilmiş ise $\boldsymbol{\sigma_x = \frac{S_x}{\sqrt{n}}}$ olarak yazılır. Ana kütle ortalaması bilinmiyorsa örneklemden yola çıkarak belli bir güven aralığında tahmin edilebilir.

Örnek:
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?

Ana kütle ortalamasının güven aralığını aşağıdaki ifadeyi kullanarak hesaplayınız.

$$(X - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) \leq \mu \leq (X + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

Bu ifadesde,
X: Büyük veri yığının ağırlık ortalaması,
Buna göre n örneklem için standart hata:

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_x = \frac{20}{\sqrt{1024}} = \frac{20}{32} = \frac{5}{8} = 0.625 gram$$

% 95 güven aralığı olasılıkla Z değeri:
$P(Z_{\alpha/2})$=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**.**

100 - (1.96)(0.625) ≤ μ ≤ 100 + (1.96)(0.625) = (100-1.225) ≤ μ≤ (100+1.225) olarak bulunur

Örnek.

Ürün ağırlıkları ile ilgili değişkenliğin standart sapması, σ = 5 gram olduğu bilinmektedir. Örnekleme oranı %1 olacak şekilde alınan 100 birimlik (n) örneğin ortalaması 100 gr. ($\overline{X}$) bulunduğuna göre, ana kütledeki ürünlerin ortalama ağırlığını belirli olasılık kademelerine (güven aralığına) göre tahmin edin.

Buna göre önce örneklem standart hata bulunur,

$\boldsymbol{\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{100}} = 0.5\ gr}$ **olur.**

Örnek ortalamaları ana kütle ortalaması etrafında normal dağıldığına göre; normal dağılımın özelliklerinden yararlanarak, ana kütle birimlerin ortalama ağırlığı,

$\overline{X}$: Örneklem ağırlık ortalaması

$$(X - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) \leq \mu \leq (X + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

şeklinde tahmin edilebilir. Formüldeki "Z" ifadesi belirli olasılık kademelerindeki standart normal dağılımın kritik değerleridir. Çeşitli olasılık kademeleri için yığın ortalamasının tahmini ise aşağıda gösterildiği gibi hesaplanır.

% 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

Ana kütlenin ağırlık ortalamasının aralık tahmini,

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

100 - (1)(0.5) ≤ μ ≤ 100 + (1)(0.5) = 99.5gr ≤ μ ≤ 100.5gr olarak bulunur (% 68 olasılıkla).

% 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

Ana kütlenin ortalamasının aralık tahmini,

100 - (1.96)(0.5) ≤ μ ≤ 100 + (1.96)(0.5) = 99.02gr ≤ μ≤ 100.98gr olarak bulunur.

% 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

Ana kütlenin ortalamasının aralık tahmini,

100 - (2.58)(0.5) ≤ μ ≤ 100 + (2.58)(0.5) = 98.71 gr ≤ μ ≤ 101.29 gr olarak bulunur.

Örnek:

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 1.000 erkek öğrenci seçildiğini varsayalım. Buradaki standart sapmanın 13,6 kilo olduğunu varsayılır. İstenen güven aralığının %95 seçildiğini varsayalım. %95 aralığında aritmetik ortalama hangi aralıkta değişiyor.

Buna göre örneklem standart hata,

$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{13.6}{\sqrt{1000}} = 0.43gr$ hesaplanır.

% 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

Ana kütlenin ortalamasının 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)(0.43) ≤ μ ≤ 80 + (1.96)(0.43) = 79.1572gr ≤ μ≤ 80.8428gr olarak bulunur

Örnek:

64 kişi üzerinde yapılan bir araştırmada Türkiye'de insanların ortalama günde 8 dakika televizyon seyrettikleri ve ana kütlenin standart sapmasının 4 dakika olduğu bilindiğine göre % 95 güven aralığında ana kütle ortalamasını tahmin ediniz.

Buna göre örneklem standart hata,

$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{64}} = 0.5dk$ hesaplanır.

% 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

Ana kütlenin ortalamasının aralık tahmini,

8 - (1.96)(0.5) ≤ μ ≤ 8 + (1.96)(0.5) = 7.02dk ≤ μ≤ 8.98dk olarak bulunur

Örnek:

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,

$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{14}{\sqrt{49}} = 2$ hesaplanır.

% 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

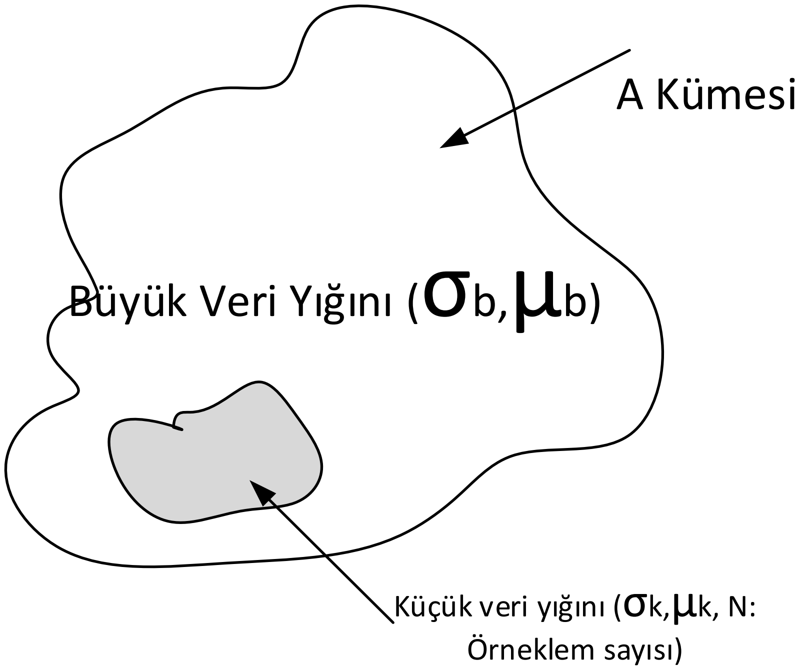
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964

Ana kütlenin ortalamasının aralık tahmini,

110 - (2.58)(2) ≤ μ ≤ 110 + (2.58)(2) = 104.76 ≤ μ ≤ 115.16 olarak bulunur.

% 99 güven düzeyinde öğrencilerinin IQ’leri 104.76 ile 115.16 arasındadır, şeklinde yorumlanabilir.

7.6. Hipotez Testi



Küçük veri yığını büyük veri yığınını hangi hipotez koşulları altında temsil edebilir?

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.

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.

Çözüm adımları:

1) 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...

2) Seçim iadesiz seçimdir ve tamamen rassal bir süreçle yapılmıştır.

3) 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.

4) 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.

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.

		Karar	
		Hipotezi Koru	Hipotez Ret Edilmiş
Yığındaki Doğruluk	Doğru (H0) Olasılığı	Doğru 1-α	Tip-1 Hata α
	Hatalı (H0)	Tip-2 Hata β	Doğru 1-β (Güç)

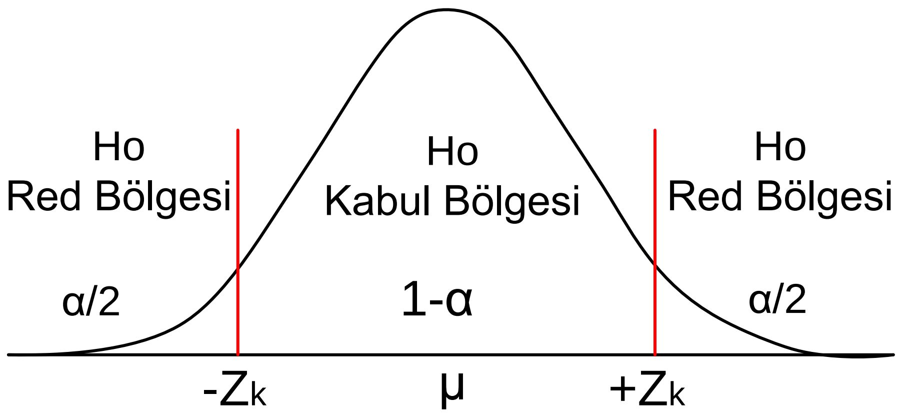
Yukarıdaki tablodan görüleceği üzere, sıfır hipotezi doğru ve bu hipotez ret edilmiş ise, hatalı bir karar verilmiştir. Bu hatalı karar Tip-1 hata adını alır. Tip-1 hata yapma olasılığı α olacaktır. α, hipotezin **anlamlılık düzeyidir.** Karar ya kabul ya da red edileceğine göre, kararı kabul etme olasılığı 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.

5) 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.



Dikkat:

- 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.

Örnek:

%90 çift yönlü olasılık alınırsa, α/2=0.05 olur. 1- 2*α =P(Zk)=0.90

Z-Tablosunun simetrik özelliğinden α değerini 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 elde edilir.

%95 çift yönlü alınırsa, α/2=0.025 olur. P(Zk)=0.95

Z-Tablosundan değer bulmak için 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 elde edilir.

%99 çift yönlü alınırsa, α/2=0.005 olur. P(Zk)=0.99

Z-Tablosundan değer bulmak için 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.

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

Tek taraflı testlerde güven sınırını belirleyen alan, testin yönüne bağlı ve normal eğrinin sadece bir ucundadır.

"Ürünlerin ortalama ağırlığı 100 gramdan hafiftir" şeklindeki bir iddia şu şekilde kurulabilir:

$H_0 : \mu \geq 100$ gr.

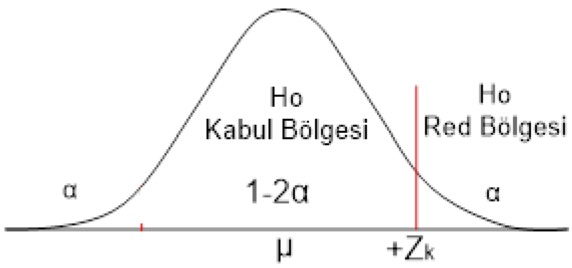
$H_1 : \mu < 100$ gr.

"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.



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.

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.

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.

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 değerini alır.

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 değerini alır.

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

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)

6) Hipotez test edilir:

Bir hipotez kurulduktan sonra 2 aşamada test edilir.

1. Aşama: Örneklemden yola çıkılarak Z_h degeri hesaplanır.

$$Z_h = \frac{X - \mu}{\sigma_x}$$

Örneklem standart hata, $\sigma_x = \frac{\sigma}{\sqrt{n}}$ hesaplanır. $\overline{X} \pm \sigma_x$ belirlenir.

2a. Aşama:

$\bar{X}_h = Z_h * \frac{\sigma}{\sqrt{n}} + \mu$

Örneklem ortalama, μ, kritik ortalam değeri, $\bar{X}_h$ olduğunda $\mu > \bar{X}_h$ ise sıfır hipotezi reddedilir ve ve alternatif hipotez kabul edilir.

2b. Aşama:

Hesaplanan Z_h değeri Z_k değeri ile karşılaştırılır.

$Z_h > Z_k$ ise H_0 hipotezi reddedilir ve alternatif hipotez kabul edilir.

7) Test sonucu değerlendirilir ve yorumlanır:

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.

Hipotez testinde kullanacağımız yukarıda değindiğimiz bilgileri özetlersek:
Sıfır Hipotezi (H_0): Tersine yeterli kanıt bulununcaya kadar doğru kabul edilen fikirdir.
Alternatif Hipotez (H_1): Sıfır hipotezi karşısında test edilen, sıfır hipotezi red edildiğinde kabul edilen hipotez.
Tek Yanlı Karşıt Hipotez: Ana kütlede ilgilendiğimiz parametre için sıfır hipotezince, belirlenen bir değerden küçük ya da büyük olanaklı bütün değerleri içeren karşıt hipotez
Çift Yanlı Karşıt Hipotez: Ana Kütlede ilgilendiğimiz parametre için sıfır hipotezince belirlenen değer dışında olanaklı bütün değerleri içeren karşıt hipotez
Hipotez Testi Kararı: Araştırmacıyı örneklem kanıtına dayanarak, sıfır hipotezini kabul ya da reddetmeye götürecek şekilde geliştirilmiş karar kuralı
Tip-1 Hata: Doğru hipotezin red edilmesi
Tip-2 Hata: Yanlış hipotezin kabul edilmesi
Anlamlılık Düzeyi: Doğru olan sıfır hipotezini reddetme olasılığı, α.
Testin Gücü: Yanlış bir sıfır hipotezinin reddedilme olasılığı, 1-α.

Örnek:
Bir sistemi ya da bir modeli temsil eden veri yığınını test edebilmek için iki parametreye ihtiyaç var;

1- ağırlık ortalama,
2- 2- varyans ya da standart sapma.

Dikkat: Hipotez üretilirken örnek sayısı her zaman 30'dan büyük olmalıdır.

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,

$$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{16}{\sqrt{64}} = \frac{16}{8} = 2mg/dl$$

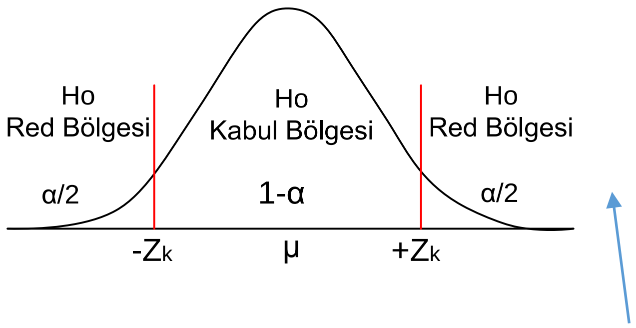
$\bar{x} \pm \sigma_x$ = 102.5 ± 2mg/dl koşulu sağlanabilmektedir mi?

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. Hipotez doğrudur.



Özet:

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?

Örnek:

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= 21,2 kg, X_3= 18,9 kg, ... , X_{40}= 20,8 kg

Örneklem istatistikleri hesaplanır:
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?

Örneklem standart sapma hatası,

Delta= $\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{3.2}{\sqrt{40}} = 0.506kg$ hesaplanır.

$\bar{x} \pm \sigma_x$ = 21,4 ± 0,506 kg

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 $\bar{X}$= μ_o yazabiliriz.

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.

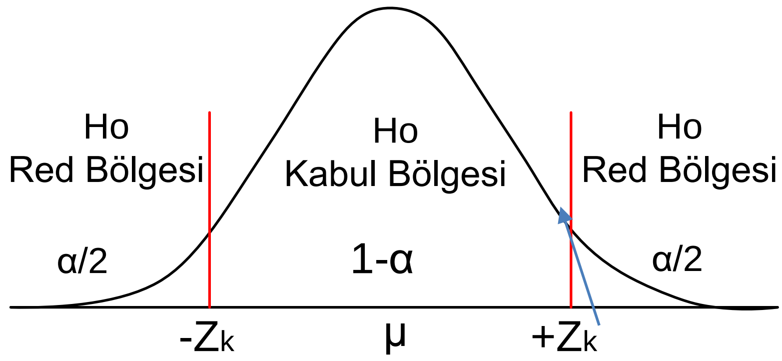
İstatistiksel "Anlam Düzeyinin" belirlenmesi (Risk düzeyi, Yanılgı Payı, Hata payı) α'nın saptanması: Hatasız bir test yapamayacağımız için her testte bir miktar yanılma riskimiz vardır. O nedenle istatistikçiler olabildiğince az yanılma ile test yapmak isterler. Yine de α =0,05 ve α=0,01 düzeyleri sık kullanılanır. α=0,05 olsun. Testin güven düzeyi = 1 - α = 0,90 olsun.

Test 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



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 $\mu=\mu_o$ 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?

Örnek:

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

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

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.

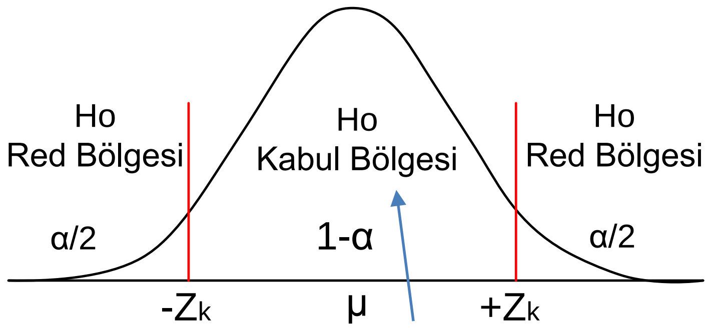
H1 Hipotezi: Bu örneklem " μ_o = 20 kg" olan bir anakütleden çekilmiş bir rassal örneklem olamaz. Aradaki 0,4 kg lık fark tesadüfe bağlı değil, ayarlamanın yapılmamış olması nedeni ile gerçekleşmiştir. Bu örneklemin çekilmiş olduğu anakütle 20 kg olamaz. Örneklemimiz kendine ait başka bir anakütleden 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 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



Karşılaştırma, sonuç ve yorum:

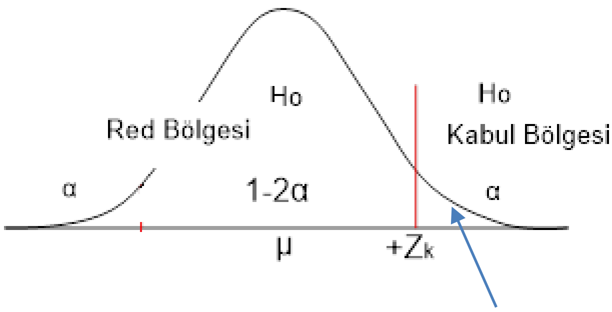
Bu duruma göre: elimizdeki örneklemin ortalaması (Zh=0.8), ilgilendiğimiz anakütlenin ortalamasının (Zk=1.96) iç tarafında olan bir büyüklüktedir. %95 olasılıkla kabul bölgesinin iç alanına düşmüştür. Bu durumda $\bar{x}$ = μ_o biçiminde ifade ettiğimiz ve $\mu=\mu_o$ düzeyine yükselttiğim H_o hipotezini 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.

Örnek:

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 $\geq$ 180
H1: Alternatif Hipotez, ortalama$<$ 180

Tek taraflı testlerde güven sınırını belirleyen alan, testin yönüne bağlı ve normal eğrinin sadece bir ucundadır.



Tek taraflı test, % 5 anlam düzeyine göre kabul ve red bölgeleri: 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 olarak bulunur.

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.

Yöntem-2: Standartlaştırılmış test istatistikleri ve standartlaştırılmış Z değeri kullanılabilir.

$$\mathbf{Z_h} = \frac{\bar{X} - \mu}{s_{\bar{x}}} = \frac{190 - 180}{75/\sqrt{300}} = 2.309$$

Z = 2.309, %90 olasılıklı Z_k = 1.645'ten büyük olduğundan, kabul bölgesinin içerisindedir, benzer bir sonuçla kabul edilir.

Örnek:

Yeni bir borsa firması, X, aracılık ücretlerinin mevcut hisse senedi komisyoncu firması A'nın ücretlerinden daha düşük olduğunu iddia ediyor. Bağımsız bir araştırma firmasından elde edilen veriler, tüm A broker müşterilerinin ortalama ve standart sapmasının sırasıyla 18$ ve standart sapması 6$ olduğunu göstermektedir.

100 A müşterisinden oluşan bir örnek alınır ve aracılık ücretleri yeni X komisyoncusu oranları ile hesaplanır. Örneklemin ortalaması 18.75$ ve standart sapması ise aynı 6$, A ile X komisyoncusu arasındaki ortalama aracılık faturası arasındaki fark hakkında herhangi bir çıkarım yapılabilir mi?

H0: Boş Hipotez: ortalama = 18$

H1: Alternatif Hipotez: ortalama <> 18$ (Kanıtlamak istediğimiz şey budur).Çift taraflı testlerde güven sınırını belirleyen alan, normal eğrinin orta bölgesidir.

Reddetme bölgesi: Z <= - $Z_{2.5}$ ve Z> = $Z_{2.5}$ (% 5 anlamlılık düzeyi varsayılarak, her iki tarafa da 2,5 bölünür).

Zh = (örnek ortalama - ortalama) / (std-dev / sqrt (örnek sayısı))

Zh= (18,75 - 18) / (6 / (sqrt (100)) = 1,25

Hesaplanan bu Z değeri, aşağıdakiler tarafından tanımlanan iki sınır arasındadır:

% 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.

- $Z_{2,5}$ = -1,96 ve $Z_{2,5}$ = 1,96.

Bu, mevcut komisyoncunuz ile yeni komisyoncunuzun oranları arasında herhangi bir fark olduğu sonucuna varmak için yeterli kanıt olmadığı sonucuna varır.

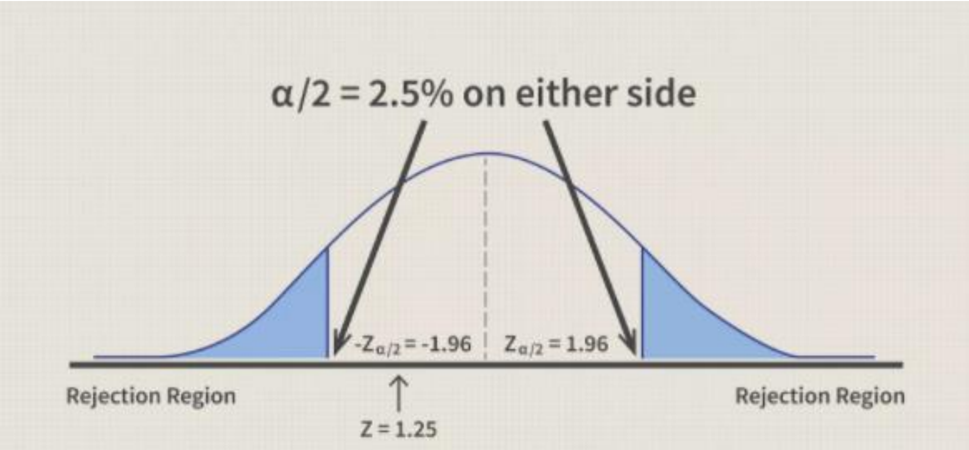
Standart normal dağılımda iki uç nokta değeri, Zh=-1.25 ve Zh=+1.25 ise bu aralıkta ve bu aralık dışındaki olasılıkları nedir?

P (1.25 ≥ Zk ≥ -1.25) = 0.3944*2 =0.7888, %78.88

Alternatif olarak, p-değeri = P (Zh <-1.25) + P (Zh> 1.25)

= 2 * (0.5-0.3944) = 2*0.1056 = 0.2112 =% 21.12, bu da 0.05 veya % 5'ten büyüktür ve aynı sonuca götürür.

Grafik olarak şu şekilde temsil edilir:



Varsayımsal Test Yöntemi İçin Eleştiri Noktaları:

- Varsayımlara dayalı istatistiksel bir yöntem
- Alfa ve beta hataları açısından ayrıntılı olarak hataya açık
- P değerinin yorumlanması belirsiz olabilir ve kafa karıştırıcı sonuçlara yol açabilir

Örnek:

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.

Hipotez testini kuralım.

H_0 : µ = 500 gr.

H_1 : µ ≠ 500 gr.

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

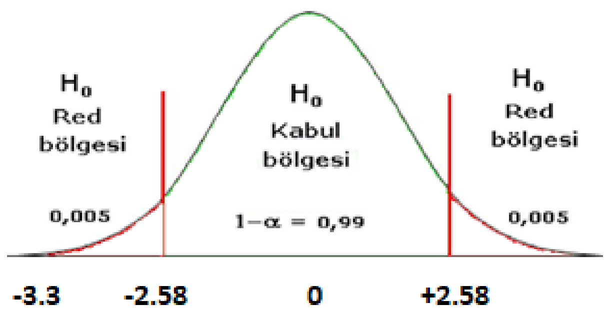
$$\mathbf{Z_h} = \frac{\mathbf{X} - \mu}{\sigma_{\bar{x}}} = \frac{\mathbf{490 - 500}}{\mathbf{3}} = \mathbf{-3.33}$$

%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.

Z_h (-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.



Örnek:

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 ve standart hatası 12 bulunuyor. α=0.05 anlamlılık düzeyinde (%95 güven aralığı) iddianın doğruluğunu test ediniz.

Hipotez testini kuralım:
H_0 : μ ≤ 60 dakika
H_1 : μ > 60 dakika

α = 0,05 (yani% 5 anlamlılık düzeyi), %90'a karşılık gelir. Simetriden dolayı orta bölge %90, uçlar % 5'lik α değerlerine karşılık gelmektedir.

Yöntem:
% 90 olasılıkla, ana kütlenin ortalamasının aralık tahmini yapılacaktır.
$P(Z_\alpha)$=0.90/2=0.45 elde edilir. Standart normal dağılım tablosundan $P(Z_\alpha)$= 0.45 değerini karşılayan Z_α değeri belirlenir. Z_α=1.65 dir.

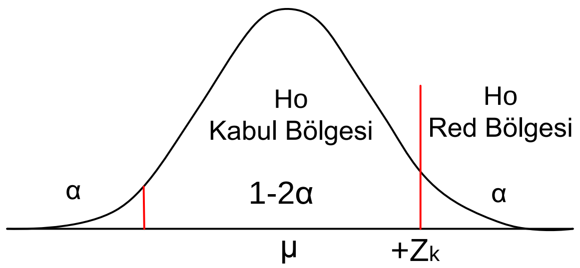
Standart sapma hatası, S_x verilmiş ise $\boldsymbol{\sigma_x} = \frac{\boldsymbol{S_x}}{\sqrt{\boldsymbol{n}}}$ = 12/8 = 1.5

$$1.65 = \frac{\bar{X}_L - 60}{1.5}$$

$\bar{x}_L = 1.65 * 1.5 + 60$ = 62.475 bulunur.

Örneklem ortalaması (63) kritik değerden (62.475) daha büyük olduğu için, H_0 hipotezi reddedilemez.

Yorum: %95 güven düzeyinde ilaç ağrıyı en geç 63 dakika içinde geçirmektedir.



Örnek:

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 4 kilo olduğu tespit edilmiştir. %99 güven aralığında iddianın doğruluğunu test ediniz.

Hipotez testini kuralım.

H_0 : μ ≥ 10 kilo

H_0 : μ < 10 kilo

Örneklemden yola çıkılarak Z_h degeri hesaplanır.

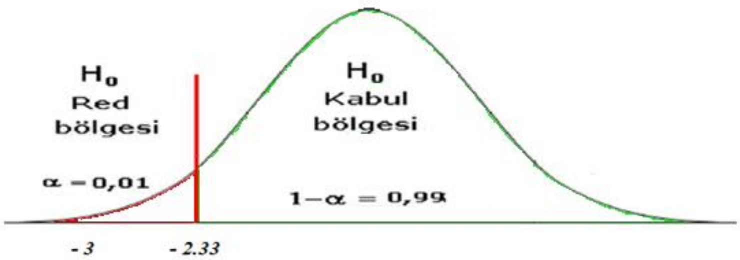
Standart sapma hatası, S_x verilmiş ise $\boldsymbol{\sigma_x} = \frac{\boldsymbol{S_x}}{\sqrt{\boldsymbol{n}}}$ = 4/12 = 1/3

$$Z_h = \frac{\mathrm{X} - \mu}{\sigma_{\bar{x}}} = \frac{9 - 10}{1/3} = \frac{-1}{1/3} = -3$$

%98 güven aralığındaki, $P(Z_{\alpha/2})$ = 0.98/2 = 0.49 elde edilir. 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 (2.58) > Z_k (2.33) olduğundan H_0 hipotezi reddedilir.



Örnek:
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.

Hipotez
$H_0 : \mu = 20$ litre
$H_1 : \mu > 20$ litre
n = 1000
$\bar{X}$ = 22 litre
μ = 20 litre
s = 8 litre
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.

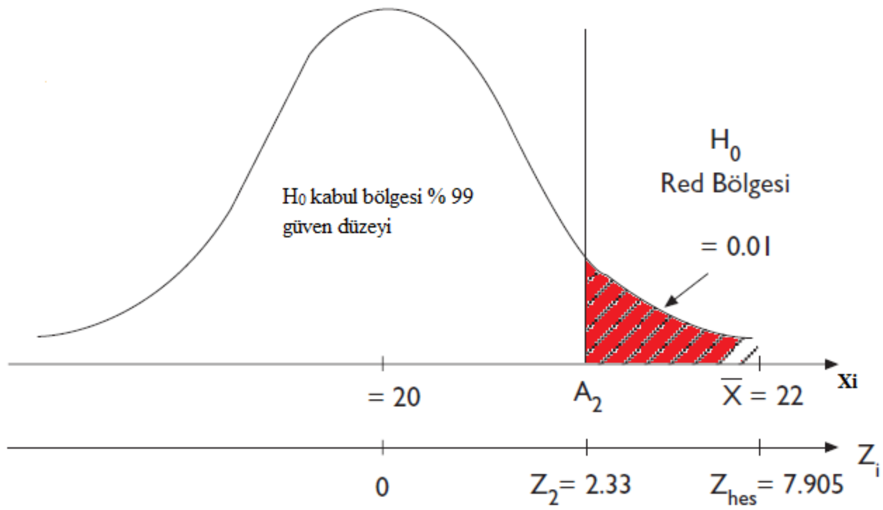
1. **Aşama:** Örneklemden yola çıkılarak Z_h degeri hesaplanır.

$$\mathbf{s_{\bar{x}}} = \frac{s}{\sqrt{n}} = \frac{8}{\sqrt{1000}} = \mathbf{0.253}$$

$$\mathbf{Z_h} = \frac{\bar{X} - \mu}{s_{\bar{x}}} = \frac{22 - 20}{0.253} = \mathbf{7.9}$$

2. **Aşama:** 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.



Örnek:

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.

Hipotez Simetrik

$H_0 : \mu = 500$ gram

$H_1 : \mu \neq 500$ gram

n = 100

$\bar{X}$ = 495 gram

μ = 500 gram

s = 20 gram

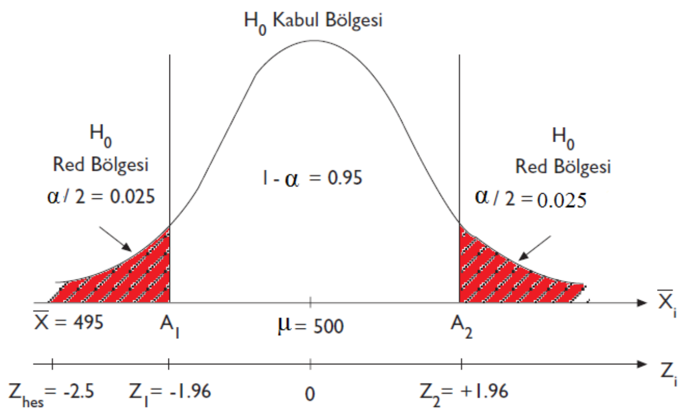
n > 30 olduğu için normal dağılıma sahiptir. Z değeri tablodan % 95 güven düzeyi için Zk=1.96 bulunur. Örneklemden yola çıkılarak Z_h degeri hesaplanır.

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{20}{\sqrt{100}} = 2$$

$$Z_h = \frac{\bar{X} - \mu}{s_{\bar{x}}} = \frac{495 - 500}{2} = -2.5$$

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.



7.6.1. Data Analitiğinde Hipotez Testi

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.

$$x_{new} = \frac{x - \mu}{\sigma}$$

7.6.2. 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) $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 < \mu_2$ — Tek Yönlü hipotez

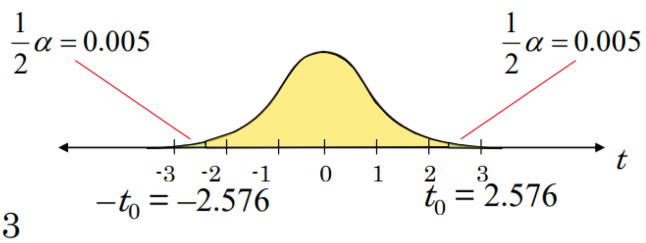
c) $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 > \mu_2$ — Tek Yönlü hipotez

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

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

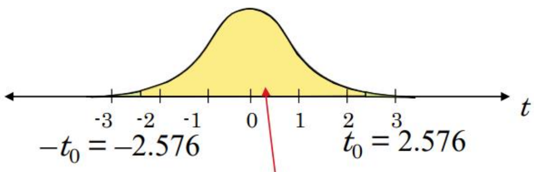
$$= \sqrt{\frac{(17-1)7800^2 + (18-1)7375^2}{17+18-2}} \cdot \sqrt{\frac{1}{17} + \frac{1}{18}}$$

$$\approx 7584.0355(0.3382)$$

$$\approx 2564.92$$

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.

7.6.3. 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

7.6.4. 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?

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$\bar{x} = \mu + z\left(\frac{\sigma}{\sqrt{n}}\right)$$

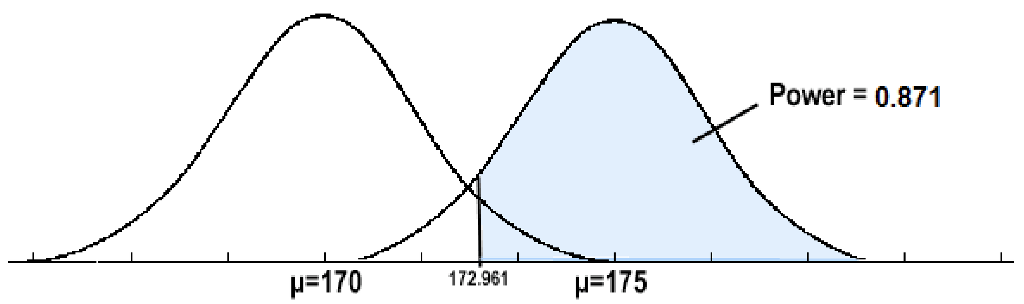
$$\bar{x} = 170 + 1.645\left(\frac{9}{\sqrt{25}}\right)$$

$$= 172.961$$

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 Zincir Analizi

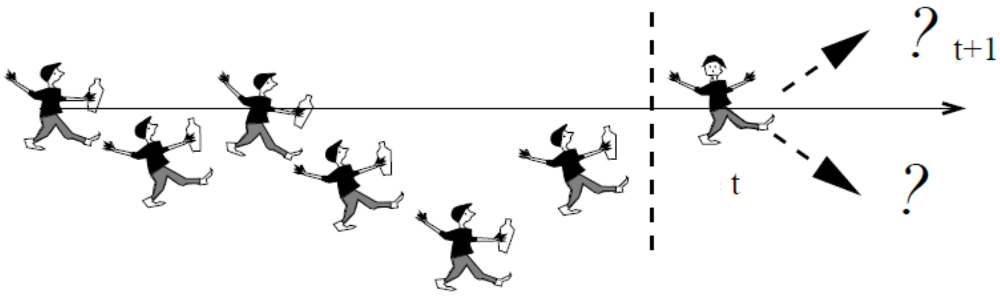
Stokastik (Rastgele - Rassal) süreç, varlığı deneysel olarak kanıtlanmış zaman veya mekana göre değişen ya da evrilen olguları tanımlamak için kullanılan bir olasılık modelidir. Evrilmek: Bir biçimden başka bir biçime doğal olarak dönmek. Şu an hava Güneşli yarın nasıl olacak? Yağmurlu mu? Bulutlu mu? Güneşli mi?

İster geçmişte, ister şu an, ister gelecekte durumlarınız ve geçiş olasılıklarınız biliniyorsa bir sonraki adımdaki durumunuzun olasılıkları hesaplanabilir.

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. ***Herbir 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 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.

Süreçler $\{X_0, X_1, X_2, . . .\}$, 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. $X_0, X_1, . . . , X_{t-1}$ durumlarına bağlı değildir.



Tanımlar:

Tanım-1: t zamanında bir Markov zincirinin durumu X_t değeridir.

Tanım-2: Bir Markov zincirinin durum (S) kümesi, her bir X_t'nin alabileceği değer kümesidir.

Örneğin, S = {1, 2, 3, 4, 5, 6, 7}. S kümesi N duruma sahiptir.

Tanım: Bir Markov zincirinin yörüngesi, X_0, X_1, X_2, için belirli değerler kümesidir. . ..

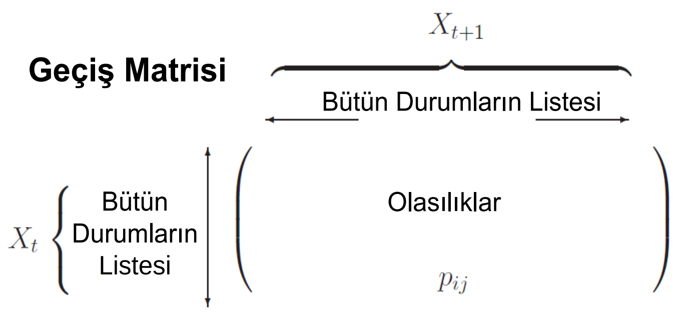
Daha genel olarak, eğer s0, s1, s2, s3, ... yörüngesini takip edersek X0 = s0, X1 = s1, X2 = s2, X3 = s3, değerlerini alır. "Yörünge" sadece "yol" anlamına gelen bir kelimedir.

Markov Özelliğini matematiksel gösterimle şu şekilde formüle ediyoruz:

$$\boxed{\mathbb{P}(X_{t+1} = s \,|\, X_t = s_t, X_{t-1} = s_{t-1}, \ldots, X_0 = s_0) \;=\; \mathbb{P}(X_{t+1} = s \,|\, X_t = s_t),}$$

tüm t = 1, 2, 3, . . . için ve tüm s0, s1,. . . , st, s durumları için.

Eğer {X_0, X_1, X_2, . . .} Markov özelliğini sağlar ise {X_0, X_1, X_2, ... } ayrık rasgele değişkenlerin bir sırası Markov zinciri olur.



$$p_{ij} = \mathbb{P}(X_{t+1} = j \,|\, X_t = i)$$

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 olasıklarını verir.

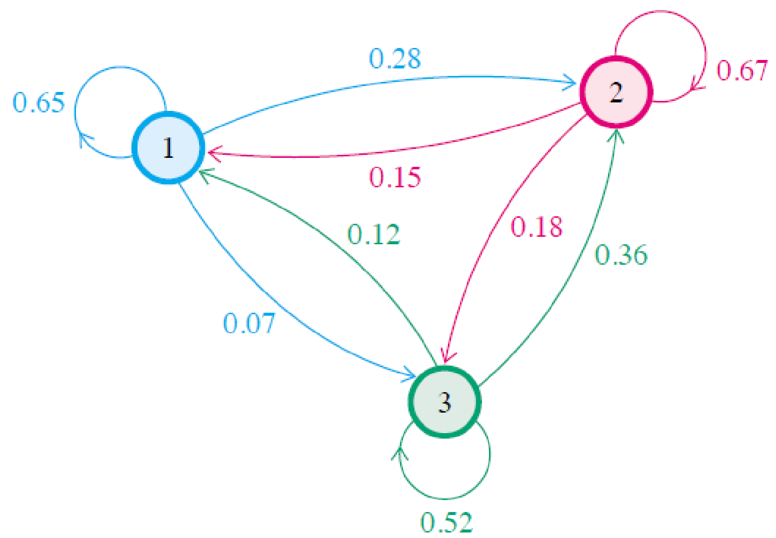
Bir deneyin denemelerinin bir sırası, bir Markov zinciri olabilmesi için

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.

Durumların Geçiş Tablosu:

		Next Generation		
	State	**1**	**2**	**3**
Current Generation	**1**	0.65	0.28	0.07
	2	0.15	0.67	0.18
	3	0.12	0.36	0.52

Durumların Geçiş Diyagramı:



Bir geçiş matrisinde, durumlar yanda ve üstte gösterilir. P tablo için geçiş matrisini temsil ediyorsa, bir durum olduktan sonra diğer durumların olma olasılığının veren matrisdir.

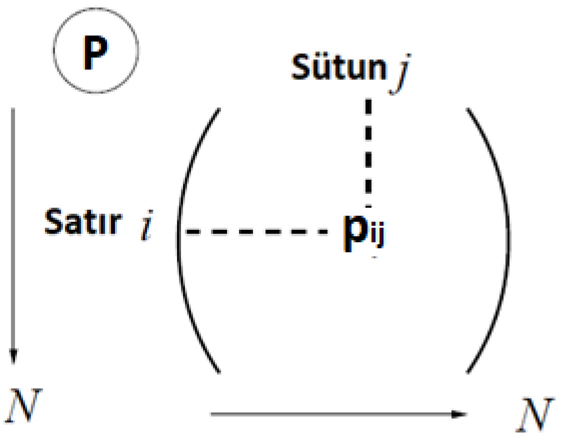
Geçiş Matrisi:

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.65 & 0.28 & 0.07 \\ 0.15 & 0.67 & 0.18 \\ 0.12 & 0.36 & 0.52 \end{bmatrix} \end{array} = P$$

$p_{23} = 0.18$ nedir? 2.durumdan bir sonraki adımda 3. Duruma gitme olasılığı %18 dir.

Matris:

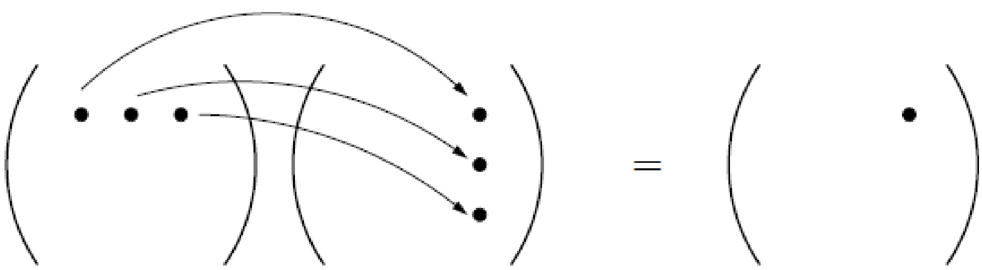
P matrisi N x N matrisi olsun. $P=(p_{ij})$ biçiminde yazılır. P'nin (i,j) elemanı hem p_{ij} hemde $(P)_{ij}$ olarak yazılır.



Matris çarpımı:

$A=(a_{ij})$ ve $B=(b_{ij})$ N x N boyutunda matrisler olsun. A ile B matrislerinin çarpımı A x B =AB dir.

$$(AB)_{ij} = \sum_{k=1}^{N} a_{ik} b_{kj}$$



Bir matrisin karesi:

$$(A^2)_{ij} = \sum_{k=1}^{N} (A)_{ik}(A)_{kj} = \sum_{k=1}^{N} a_{ik} a_{kj}.$$

8.1. Markov Zinciri

Markov Zincirleri, zamana bağlı, uzaya bağlı stokastik süreçleri modellemek için basit ve çok kullanışlı araçlardan biridir. Finans (hisse fiyatı hareketi), satışlar (satış miktarı bilgisi), NLP algoritmaları (sonlu durum dönüştürücüler, POS Etiketleme için Gizli Markov Modeli), hava durumu tahmini vb. gibi birçok alan, tahminlerini kolay ve doğru bir şekilde yapmak için Markov zincirini kullanır.

Markov zinciri, geleceğin geçmişe bağlı olmadığı, şimdiye bağlı olduğu bir stokastik süreçler sınıfını temsil eder. Bir stokastik süreç, eğer süreç geleceği işleyecek olan Markovyen özelliklerden oluşuyorsa, Markov zinciri olarak düşünülebilir. Şimdiki durum ve geçmiş süreçten bağımsız bilgilere ihtiyacımız var.

Xn durumunun n zamanda kaydedildiği bir durumu düşünün. n+1 anındaki gelecek durum, n anındaki duruma bağlıdır. n zamanında vaka sayısının Xn olduğu ve n+1 zamanında vaka sayısının Xn+1 olduğu korona vakalarına bir örnek verelim. Dolayısıyla, Markov zinciri tanımını izliyorsak, n+1 zamanındaki vakaların sayısı, n zamanındaki vakaların sayısına bağlı olacaktır (Xn+1, Xn'ye bağlı olacaktır), geçmişte {Xn−1, Xn olan vakaların sayısına bağlı olacaktır. -2, . . . , X0}. Markov zincirini anlamak için temelde Markov zinciri kavramı tarafından kullanılan bazı terimleri anlamamız gerekebilir. Bu terimler aşağıda açıklanmıştır.

Durum uzayı:
Markov zinciri için durum uzayı S = {1,2,3….., n} olduğunda S ile verilebilirse, sürecin durumu Xn değeri ile verilebilir. Örneğin, Xn = 8 ise, sürecin durumu 8'dir. Dolayısıyla, herhangi bir n anında, sürecin Xn değeri tarafından verildiği durumu söyleyebiliriz.

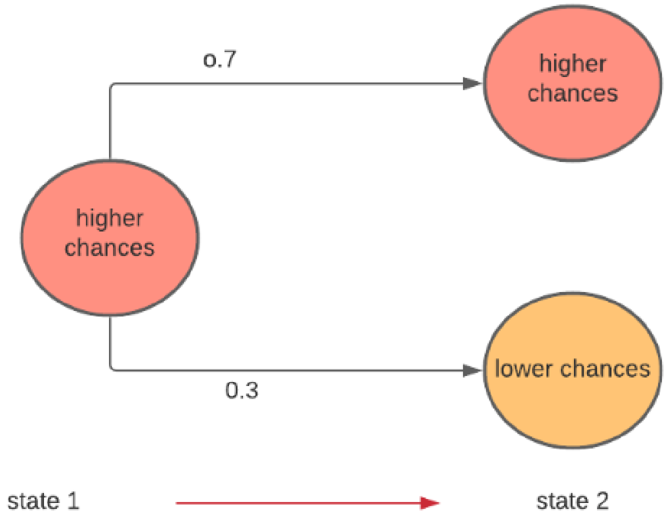
Örneğin, bir öğrenci sınıfında, eski başarısız kaydı olan öğrencilerin başarısız olarak nihai bir sonuç geliştirme olasılığı daha yüksektir ve önceki sınavda daha düşük not alan öğrencilerin sonucu başarısız olarak alma olasılığı daha yüksektir. Dolayısıyla bu durumda, eski başarısız kaydı olan öğrencilerin sınavdan kalma şanslarının daha yüksek, notu düşük olan öğrencilerin ise daha düşük olduğunu söyleyebiliriz. Bu senaryoda iki durumumuz var: daha düşük şans ve daha yüksek şans. Ve S={1,2}.

Yörünge:
Markov zincirinin yörüngesi, stokastik sürecin başından beri var olduğu durumların sırası olarak düşünülebilir.

Yani yörünge değerlerini s0,s1,s2…….sn olarak gösterebilirsek, durum X0=s0, X1 = s1…….Xn=sn olarak değerler alacaktır.

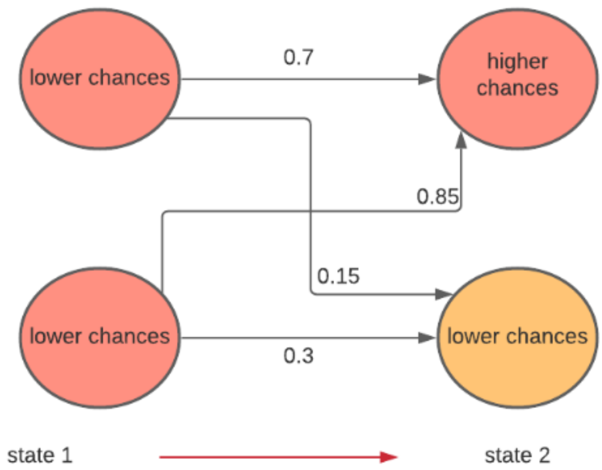
Geçiş olasılığı

Markov zincirleri belirli bir zamanda kayıtsız durumlar olamaz, ancak zamanla durumlarını değiştirebilirler. Durumdaki değişiklik, durum geçişi olarak adlandırılabilir. Aşağıda verilen örnekten, örneğin Markov zinciri ya daha düşük ya da daha yüksek şansa sahip olabilir.



Yukarıdaki resim, durumların geçişinin bir temsilidir. Durum 1'de zincir, devam etmekte olan sınavın başarısızlık şansının daha yüksek olduğu durumda olduğunu söyleyebiliriz. Bir sonraki sınavın başarısızlık şansı daha yüksek olan bir duruma girme olasılığı 0.7'dir ve durumun daha düşük şansa geçme olasılığı 0.3'tür. Öğrencilerin mevcut sınavdan başka bir sınava daha düşük şans durumuna geçme olasılığı 0,3'tür.

Sistemin daha düşük şans durumunda olduğunu ve benzer bir geçiş diyagramının çizildiğini varsayalım. Burada geçiş olasılıkları 0.85 ve 0.15'tir. Her iki diyagramı da kullanarak tam bir süreç çizebiliriz.



Yukarıdaki görüntü, Durum 1'den Durum 2'ye Birleşik-durum geçiş diyagramının bir temsilidir. Bir zaman örneği için, bu süreçler geriye doğru gidemezler, ancak bir sonraki zaman örneğinde geriye gidebilirler.

Durum geçiş matrisi

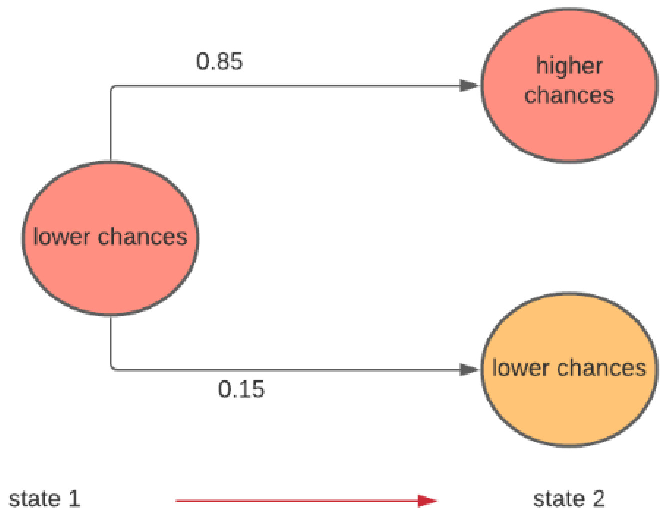
Tüm geçiş olasılıklarının matrisine geçiş matrisi denir. Satırların başlangıç noktası ve sütunların bitiş noktası olduğu yer.

$$\begin{array}{cc} & \begin{array}{cc}\text{lower chances} & \text{higher chances}\end{array} \\ \begin{array}{c}\text{lower chances}\\ \text{higher chances}\end{array} & \begin{Bmatrix} 0.15 & 0.85 \\ 0.3 & 0.7 \end{Bmatrix}\end{array}$$

Yukarıdaki matris, verilen örneğin geçiş matrisinin bir temsilidir. sürecin düşük şans durumundan düşük riske geçişi 0.15 olasılığa sahiptir. Düşük riskten yüksek şansa geçişin olasılığı 0.85'tir.

Markov Zinciri Kullanarak Tahmin

Markov zinciri, gelecekteki değer için tahminler yapmak için çok güçlü bir araçtır. Çeşitli faydalı içgörüler sağladığından, içgörüleri anlamak için geçiş olasılıklarını, geçiş matrisini, durum uzayını ve yörüngeyi bilmek çok gerekli hale gelir.



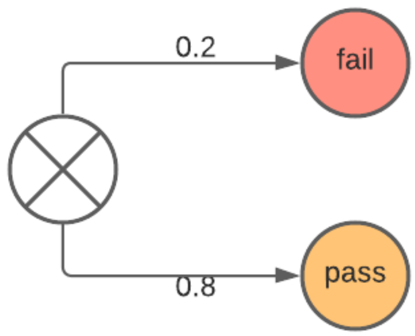
Ayrıca süreç hakkında ön bilgi sahibi olunması gereken temel şeylerden biri de sürecin başlangıç durumudur. Tahmin sürecini açıklamak için, birkaç değişikliğin uygulandığı yukarıdaki öğrenci başarısızlık şansı örneğine bir göz atalım.

İlk Durum ve Tek Adımlı Tahmin

Bu sefer bir mühendislik bölümü sınavlarındaki gözlem şu ki, öğrenciler ilk yıl matematik sınavlarında başarısız olurlarsa, temel derslerinde üç kez başarısız olurlar ve ilk yıl matematiği geçerlerse, çekirdek derslerini geçme olasılıkları daha yüksektir. Örneğin geçiş matrisi

$$\begin{array}{cc} & \begin{array}{cc}\text{fail} & \text{pass}\end{array} \\ \begin{array}{c}\text{fail} \\ \text{pass}\end{array} & \begin{Bmatrix} 0.75 & 0.25 \\ 0.2 & 0.8 \end{Bmatrix} \end{array}$$

Böylece öğrencilerin matematik sınavını geçmeleri halinde sürecin ilk hali şu şekilde olacaktır:



Yukarıdaki başlangıç durumundan, öğrencilerin sadece başlangıç durumu ve geçiş matrisini çarparak çekirdek konuyu geçme olasılığı şeklinde bir gelecek tahmini yapabiliriz.

Verilen örnek için bir sonraki adım için tahmin olacaktır.

$$\begin{Bmatrix} 0.2 & 0.8 \end{Bmatrix} \begin{Bmatrix} 0.75 & 0.25 \\ 0.2 & 0.8 \end{Bmatrix} = \begin{Bmatrix} 0.31 & 0.69 \end{Bmatrix}$$

Yukarıdaki sezgiden yola çıkarak, ilk durumdan sonraki ilk durum için tahminin aşağıdaki formülle verilebileceğini söyleyebiliriz.

İlk Durum X Geçiş Matrisi = Tahmin

Uzun Dönem Olasılık

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

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.

Dolayısıyla, ikinci adımı tahmin etmek istiyorsak, tahmin formülü şu şekilde olacaktır:

V2 = V1 x T

Ve burada bir adımın tahmininden V1'in değerini biliyoruz. V1 değerini koyarak

V2 = (V0 x T) x T

V2 = V0 x T^2

Benzer şekilde, üçüncü adım için tahmin şu şekilde olacaktır:

V3 = V2 x T = (V0 x T^2)xT

V3 = V0 x T^3

Bu nedenle, n'inci zaman adımı tahmininden bahsederken, tahmin aşağıdaki formülle hesaplanabilir.

$V_n = V_{n-1} \times T = V_0 \times T^n$

Dolayısıyla, yukarıda verilen yinelemeli süreç, uzun süreçlerin gelecekteki durum olasılığının tahmininde bu şekilde yardımcı olur. Burada uzun dönem olasılık şu şekilde yazılabilir:

$V_\infty = V_0 \times T^\infty$.

Yukarıdaki uzun dönem olasılık formülünden, geçiş matrisi tarafından yapılan hiçbir çarpma miktarının uzun dönem olasılık vektöründe değişikliklere yol açmadığını söyleyebiliriz.

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.

Markov Zincirinin Uygulanması

Markov zincirleri, hava durumu, sıcaklık, satış vb. her türlü tahminde bulunabilen tahmin için kullanılabilir. Bu, müşteri davranışını tahmin etmek için kullanılabilir. Bildiğimiz gibi, sıralı verilerle iyidir, bu nedenle POS etiketleme gibi birçok NLP problem çözümü ile birleştirilebilir. Marka sadakati ve tüketici davranışı analiz edilebilir. Oyun alanında şans oyununda çeşitli modeller geliştirilebilir.

Görüldüğü gibi, Markov zinciri kavramını anlamak ve hesaplamak zor değil, bu nedenle her boyutta hesaplama için de çok kolay olduğunu söyleyebiliriz. Markov zincirinin kolaylığı ve doğruluğu nedeniyle daha birçok kullanım alanı olabilir, çok popüler bir araştırma konusu haline gelmiştir.

8.2. Markov zincir analizinde geçiş matrisi

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.

Geçiş matrisi genel bir notasyon ile

$$0 \leq P_{ij} \leq 1,$$

$$\sum_{j=1}^{m} P_{iJ} = 1, (i = 1,2,....,m)$$

olmak üzere,

$$P=\begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1m} \\ \vdots & \vdots & \vdots & \vdots \\ P_{m1} & P_{m2} & \cdots & P_{mm} \end{bmatrix}$$

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} ... 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.

1- Vektörün herbir olasılık değeri 0 ile 1 arasındadır.
2- Vektörün katsayıları toplamı 1'e eşittir.

Örnek:

Bir havanın üç durumunu sınıflandıralım,

Durum-1: Yağmurlu

Durum-2: Bulutlu

Durum-3: Güneşli

Varsayım: Yarınki hava durumu sadece bugünün hava durumuna bağlıdır! Hava durumuna ait Yağmurlu, Bulutlu ve Güneşli olma olasıkları aşağıdaki verilmiş olsun.

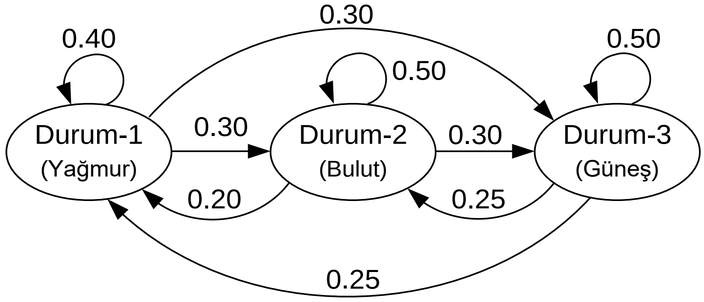
Şimdiki durum (n=0)	Bir sonraki durum (n=1)		
	Yağmur (%)	Bulut (%)	Güneş (%)
Yağmur	40	30	30
Bulut	20	50	30
Güneş	25	25	50

Tablo, şu anki durumu ve bir sonraki durumun olma olasılıklarını vermektedir. Tablodaki bilgiler P ile göstereceğimiz bir **geçiş matrisi** bulunur.

$$P = \begin{bmatrix} 0.40 & 0.30 & 0.30 \\ 0.20 & 0.50 & 0.30 \\ 0.25 & 0.25 & 0.50 \end{bmatrix}$$

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.

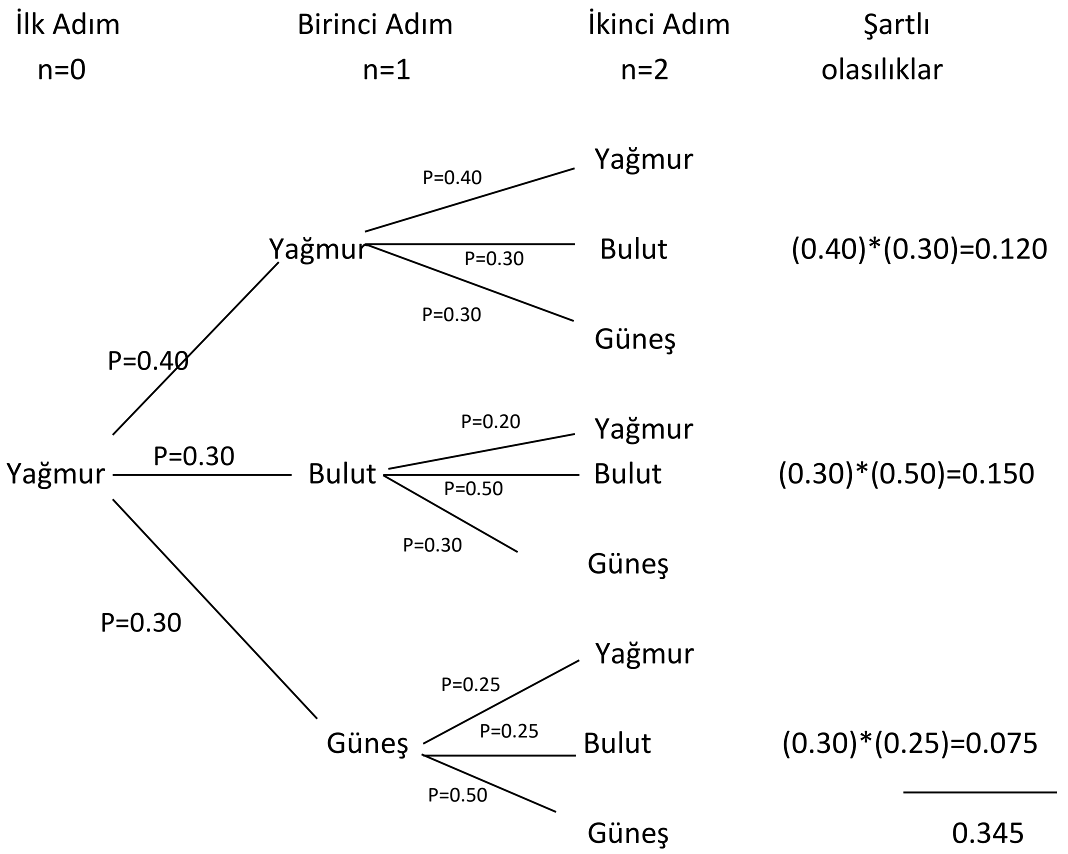
P_{13}=0.30 elemanının, şu an yağmur yağıyorsa bir sonraki gün güneşli olma olasılığını vermektedir. v^2 =(0.20 0.50 0.30) olasılık vektörü geçiş matrisinin ikinci satırını temsil eder ve vektör elemanları sırasıyla durum-2'den durum-1, durum-2 ve durum-3'e geçme olasılıklarını verir. O halde **P geçiş matrisi v^i olasılık vektörlerinden oluşur ve geçiş matrisi koşullarını sağlar.** Durumların geçiş diyagramı aşağıda verilmiştir. YYGGYBGB olma olasılığı nedir?



8.3. Markov Zinciri ile Olasılık Analizi

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.

Şu an hava **Yağmur** ise ikinci adımda (Yarın değil öbürsü gün) **Bulutlu** olma olasılığı nedir?



Şekil: n=2, ikinci adımda mümkün sonuçlar

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	n=1 de	n=2 de	Şartlı Olasılık
Yağmur	Yağmur 0.40 Bulut 0.30 Güneş 0.30	Bulut 0.30 Bulut 0.50 Bulut 0.25	(0.40)*(0.30)=0.120 (0.30)*(0.50)=0.150 (0.30)*(0.25)=0.075 0.345

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]

V^2 vektörü n=0 adımda Yağmur olan havanın n=1 adımında Yağmur iken n=2 adımında Yağmur, Bulutlu ve Güneşli olma olasıklarını vermektedir.

Markov zincirleri,olasılık problemlerinin özel halini formüle etmek için teknik ve analiz vermiştir.

V^2=P x P

$$V^2 = \begin{pmatrix} 0.2950 & 0.3450 & 0.3600 \\ 0.2550 & 0.3850 & 0.3600 \\ 0.2750 & 0.3250 & 0.4000 \end{pmatrix}$$

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.

Örnek:

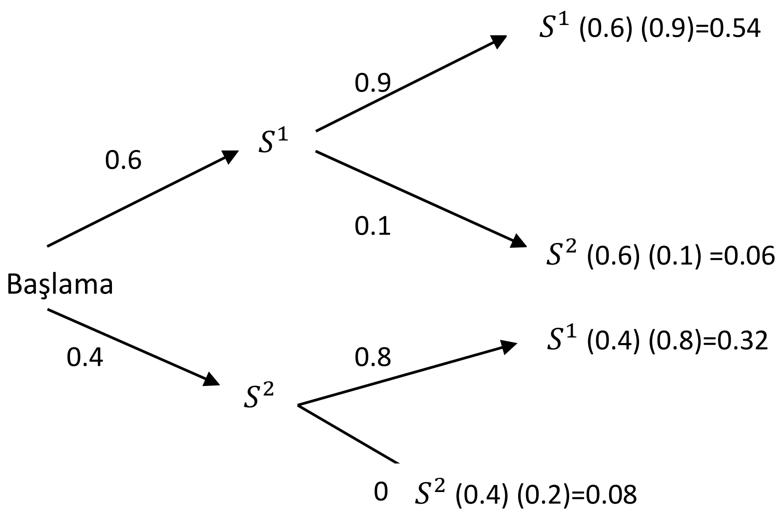
Bir markov sürecinin başlama vektörü $v^0 = (0.6 \quad 0.4)$ ve geçiş olasılık matrisi

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix}$$

Birinci adım için $v^1 = v^0 P$ olduğundan

$v^1 = (0.6 \quad 0.4) \begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix} = (0.86 \quad 0.14) = (v_1^1 \quad v_2^1)$ bulunur.

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.



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

$v^2 = v^1 P = (0.86 \quad 0.14) \begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix} = (0.886 \quad 0.114) = (v_1^2 \quad v_2^2)$

veya

$v^2 = v^0 P^2 = (0.6 \quad 0.4) \begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix} = (0.886 \quad 0.114) = (v_1^2 \quad v_2^2)$

8.3.1. Ergodik (Düzenli) Markov Zincirleri

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.

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.

$$\mathrm{P}=\begin{bmatrix} x & x & 0 & x & x \\ 0 & x & x & 0 & x \\ 0 & 0 & 0 & x & x \\ x & 0 & x & 0 & x \\ x & x & 0 & 0 & 0 \end{bmatrix}, \quad \mathrm{P}^2=\begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & 0 & x \\ x & x & 0 & x & x \\ x & x & x & x & x \end{bmatrix}, \quad \mathrm{P}^4=\begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \end{bmatrix}$$

Örnek: Aşağıdaki geçiş matrislerinin a) düzenli ve b) ergodik olmasını açıklayınız.

$$\mathrm{P}= \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \begin{bmatrix} x & x & 0 \\ x & 0 & x \\ 0 & x & x \end{bmatrix} \qquad \mathrm{P}^2=\begin{bmatrix} x & x & x \\ x & x & x \\ x & x & x \end{bmatrix}$$

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. Ayrıca 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.

Örnek:

$$
\begin{array}{cc}
 & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\
P= \begin{array}{c} 1 \\ \\ 2 \\ 3 \end{array} & \begin{bmatrix} x & 0 & x & 0 \\ 0 & x & 0 & x \\ x & 0 & x & 0 \\ 0 & x & 0 & x \end{bmatrix}
\end{array}
\quad
P^2 = \begin{bmatrix} x & 0 & x & 0 \\ 0 & x & 0 & x \\ x & 0 & x & 0 \\ 0 & x & 0 & x \end{bmatrix}
\quad
P^4 \begin{bmatrix} x & 0 & x & 0 \\ 0 & x & 0 & x \\ x & 0 & x & 0 \\ 0 & x & 0 & x \end{bmatrix}
$$

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.

Örnek:

Aşağıda verilen Markov zincirinin ergodik ve düzenli olmasını araştırınız.

$$P=\begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix} \qquad P^2=\begin{bmatrix} 1 & 0 \\ 3/4 & 1/4 \end{bmatrix}$$

P matrisinin kuvvetleri alınırsa (1.2) elemanı daima (0) olacaktır. Dolayısıyla verilen Markov zincir düzenli ve ergodik değildir.

Örnek:

$$P=\begin{bmatrix} 0 & 1 \\ 1/3 & 2/3 \end{bmatrix}$$

Düzenli Markov zinciri olduğunu gösteriniz.

$$P^2 = P*P = \begin{bmatrix} 0 & 1 \\ 1/3 & 2/3 \end{bmatrix} * \begin{bmatrix} 0 & 1 \\ 1/3 & 2/3 \end{bmatrix} = \begin{bmatrix} 1/3 & 2/3 \\ 2/9 & 7/9 \end{bmatrix}$$

P^2 nin bütün elemanları pozitif olduğundan P düzenli yani ergodik Markov zinciridir.

Örnek:

Aşağıdaki matrisin reguler (düzenli) olup olmadığını bulunuz.

$$P = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0 & 0.5 & 0.5 \\ 0.6 & 0.4 & 0 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 0.5625 & 0.3125 & 0.125 \\ 0.3 & 0.45 & 0.25 \\ 0.45 & 0.35 & 0.2 \end{bmatrix}$$

P^2 matrisindeki bütün değerler pozitif olduğunda P matrisi düzenlidir.

Örnek:

Aşağıdaki matrisin reguler (düzenli) olup olmadığını bulunuz.

$$P = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 0.25 & 0 & 0.75 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P^3 = \begin{bmatrix} 0.125 & 0 & 0.875 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P^4 = \begin{bmatrix} 0.0625 & 0 & 0.9375 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

P^2 matrisindeki bütün değerler pozitif olmadığından P matrisi düzenli değildir.

8.3.2. Denge Durumu Koşulları

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.

Örnek:

P nin n. kuvvetlerine göre n=1 den n=8 e kadar değerler için ilişkin hesaplar verilmiştir.

$$P=\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.25 & 0.25 & 0.5 \end{bmatrix} \qquad P^2=\begin{bmatrix} 0.295 & 0.345 & 0.360 \\ 0.255 & 0.385 & 0.360 \\ 0.275 & 0.325 & 0.400 \end{bmatrix}$$

$$P^3=\begin{bmatrix} 0.277 & 0.351 & 0.372 \\ 0.269 & 0.359 & 0.372 \\ 0.275 & 0.345 & 0.380 \end{bmatrix} \qquad P^4=\begin{bmatrix} 0.2740 & 0.3516 & 0.3744 \\ 0.2724 & 0.3744 & 0.3744 \\ 0.2740 & 0.3500 & 0.3760 \end{bmatrix}$$

$$P^5=\begin{bmatrix} 0.27352 & 0.35160 & 0.37488 \\ 0.27320 & 0.35192 & 0.37488 \\ 0.27360 & 0.35120 & 0.37520 \end{bmatrix} \quad P^6\begin{bmatrix} 0.273448 & 0.351576 & 0.374976 \\ 0.273384 & 0.356400 & 0.374976 \\ 0.273480 & 0.351400 & 0.375040 \end{bmatrix}$$

$$P^7=\begin{bmatrix} 0.2734384 & 0.3515664 & 0.3749952 \\ 0.2734256 & 0.3515792 & 0.3749952 \\ 0.2734480 & 0.3515440 & 0.3750080 \end{bmatrix}$$

$$P^8=\begin{bmatrix} 0.27343744 & 0.35156352 & 0.37499904 \\ 0.27343488 & 0.35156608 & 0.3749904 \\ 0.27244000 & 0.35155840 & 0.37500160 \end{bmatrix}$$

Denge durum koşullarının hesaplanması:

Denge durumunda her bir v_i^n olasılık vektörlerinin bütün değerleri için eşit olmaya meyletmektedir.

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.
3) $Vi,$ Özdeğerleri verir.

V vektörü denge durumu koşullarını içeren olasılıkları kapsar.

$\sum_{j=1}^{m} P_i^j = 1$ bağıntısı bu değerleri elde etmek için analitik yöntemi sağlar. V bir olasılık vektörü olduğundan

$$\sum_{j=1}^{m} v_i^j = 1$$

bağıntısı geçerlidir.(2) nolu koşuldan [v_1 v_2 v_3 ......v_m]*P=[v_1 v_2 v_3 .....v_m] yazılır.

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.

Örnek:

$$P=\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$$

Örneğimize belirlenen kuralları uygulayalım.

$$\sum_{j=1}^{3} v_1 = 1 \qquad v_1 + v_2 + v_3 = 1$$

$$[v_1 \ \ v_2 \ \ v_3] * P = [v_1 \ v_2 \ v_3]$$

$0.4v_1 + 0.2v_2 + 0.25v_3 = v_1$

$0.3v_1 + 0.5v_2 + 0.25v_3 = v_2$

$0.3v_1 + 0.3v_2 + 0.5v_3 = v_3$

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.

Örnek:

Aşağıdaki geçiş matrisinden hareketle denge durumu koşullarını bulunuz.

$$P=\begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$V=[v_1 \ \ v_2 \ \ v_3]$ denge vektörünün bulunması istenmektedir.O halde,

$$v_1+v_2+v_3=1$$

$$0.2v_1+0.1v_2+0.5v_3=v_1$$
$$0.5v_1+0.6v_2+(0)v_3=v_2$$
$$0.3v_1+0.3v_2+0.5v_3=v_3$$

bağıntıları yazılır. İkinci eşitlik elimine edilerek

$$v_1+v_2+v_3=1$$
$$0.5v_1-0.4v_2=0$$
$$0.3v_1+0.3v_2-0.5v_3=0$$

yazılır ve çözüm

$$v_1=0.282$$
$$v_2=0.352$$
$$v_3=0.366$$

olarak bulunur. Dolayısiyle denge vektörü: V=[0.282 0.352 0.366] olur.

Örnek:

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.

Standart bir mamul üretimi S_1 , hatalı bir mamul üretimi S_2 ile gösterilerek

$$P=\begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix}$$

Geçiş matrisi bulunur.Geçiş matrisi kuvvetleri alınarak

$$P^2=\begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix}*\begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix}=\begin{bmatrix} 0.89 & 0.11 \\ 0.88 & 0.12 \end{bmatrix}$$

$$P^3=P^2P=\begin{bmatrix} 0.89 & 0.11 \\ 0.88 & 0.12 \end{bmatrix}*\begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix}=\begin{bmatrix} 0.889 & 0.111 \\ 0.888 & 0.112 \end{bmatrix}$$

$$P^4=P^3P=\begin{bmatrix} 0.889 & 0.111 \\ 0.888 & 0.112 \end{bmatrix}*\begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix}=\begin{bmatrix} 0.8889 & 0.1111 \\ 0.8888 & 0.1112 \end{bmatrix}$$

Bulunur ve hesaplara devam ederek sürecin denge vektörünün V=(8/9,1/9) olduğu gösterilebilir. O halde verilen süreç düzenli bir Markov zinciridir.

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.

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) durum sayısını göstermek üzere P geçiş matrisi

$$\sum_{i=1}^{m} P_{ij} = \sum_{j=1}^{m} P_{ij} = 1$$

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 $V_j = 1/m$ dir.Örnekte denge vektörü,

V=[v_1 $v_2 v_3$], $v_1 = v_2 = v_3 = 1$,V=(1/3 1/3 1/3) olur.

Örnek:

$$P=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$$

P geçiş matrisinin denge vektörü, VP=V bağıntısından V=(1/2 1/2) olarak bulunur. P nin kuvvetleri alınırsa

n=tek sayı için P^n=P

n=çift sayı için $P^n = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$

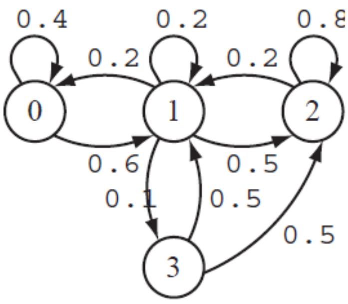
bulunur ve P nin kuvvetlerinde sıfır bulunduğu için zincir düzenli değildir.

8.3.3. Emici Markov Zincirleri

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.

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) Geçiş matrisi düzenli (ergodik) mi?

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.

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.

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) Aşağıdaki geçiş matrisi a) düzenli mi ve b) ergodik mi beirleyiniz.

$$B = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

c) Varsayalık ki, Markov zinciri aşağıdaki geçiş matrisine sahip olsun. Geçiş diyagramını çiziniz. Emici bir Markov zinciri ise hangi durum ölümcüldür.

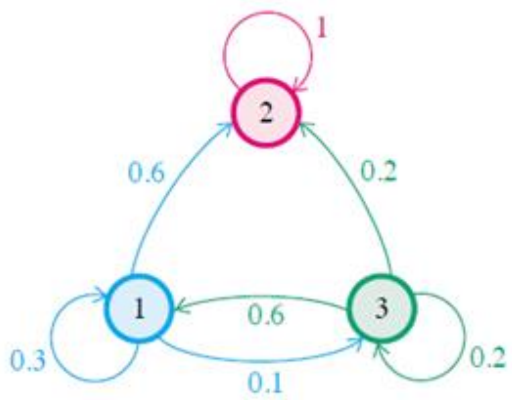
$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0 & 1 & 0 \\ 0.6 & 0.2 & 0.2 \end{bmatrix} \end{array} = P.$$

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.



8.4. Yörünge olasılığı

Bir yörüngenin $X_0, X_1, \ldots, X_n$ 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.

Yöründe herbir 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

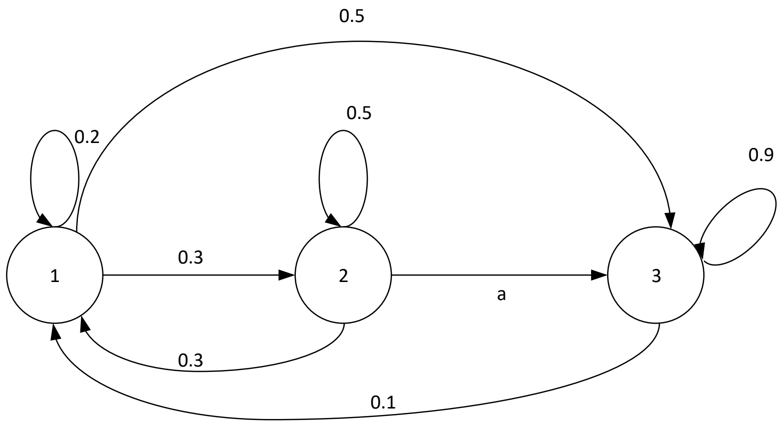
Chain rule

$$\begin{aligned}
&P(q_1, q_2, \cdots, q_T) \\
&= P(q_1)P(q_2 \mid q_1)\cdots P(q_{T-1} \mid q_1, \cdots, q_{T-2})P(q_T \mid q_1, \cdots, q_{T-1}) \\
&= P(q_1)P(q_2 \mid q_1)\cdots P(q_{T-1} \mid q_{T-2})P(q_T \mid q_{T-1})
\end{aligned}$$

1st order Markov assumption

Pi değeri seçilen duruma düşme olasılığıdır, başlangıçta verilir.

Örnek:



a) Yukarıdaki akış diyagramında kaç durum vardır. Her bir durumdan diğer durumları geçiş olasılıkları toplamı neye eşittir?

3 durum vardır. 1'e eşittir.

b) 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

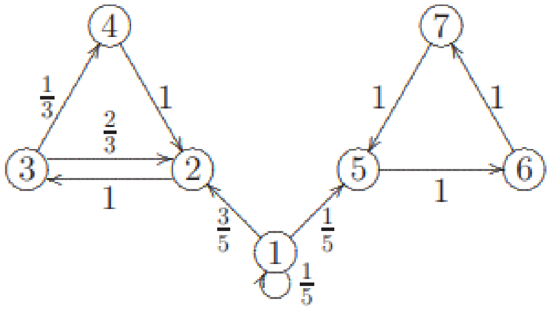
c) A, geçiş matrisini oluşturun.

A=$\begin{pmatrix} \mathbf{0.2} & \mathbf{0.3} & \mathbf{0.5} \\ \mathbf{0.3} & \mathbf{0.5} & \mathbf{0.2} \\ \mathbf{0.1} & \mathbf{0} & \mathbf{0.9} \end{pmatrix}$

d) 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

Örnek:



X0 ~ (3 /4 , 1/2, 1/ 4 , 2/3, 1/5, 1/4, 1/8: Başlangıç olasılıkları). 1, 2, 3, 2, 3, 4 yörüngesinin olasılığı nedir?

P(1,2,3,2,3, 4)=P(X_0 = 1) x p12 x p23 x p32 x p23 x p34

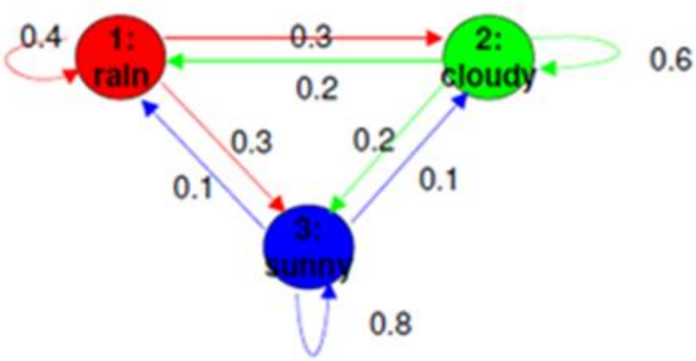
=3/4 x 3/5 x 1 x 2/3 x 1 x 1/3=1/10

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?

$S_1 : rain,\quad S_2 : cloudy,\quad S_3 : sunny$

$$\begin{aligned}P(O \mid \text{model}) &= P(S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3 \mid \text{model}) \\ &= P(S_3)\cdot P(S_3 \mid S_3)\cdot P(S_3 \mid S_3)\cdot P(S_1 \mid S_3) \\ &\quad \cdot P(S_1 \mid S_1) P(S_3 \mid S_1) P(S_2 \mid S_3) P(S_3 \mid S_2) \\ &= \pi_3 \cdot a_{33} \cdot a_{33} \cdot a_{31} \cdot a_{11} \cdot a_{13} \cdot a_{32} \cdot a_{23} \\ &= 1\cdot(0.8)(0.8)(0.1)(0.4)(0.3)(0.1)(0.2) \\ &= 1.536\times 10^{-4}\end{aligned}$$

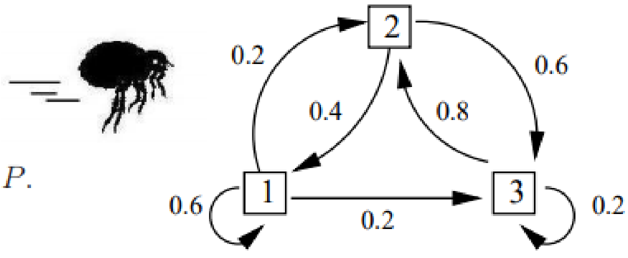


$\pi_3 = 1, \qquad başlangıç\ durumunu\ belirtmektedir. Today\ is\ sunny.$

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.

Şu an $\boldsymbol{X}_0$=1 durumdan, iki sonraki durum olan $\boldsymbol{X}_2$=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

Toplam=0.28

Örnek:

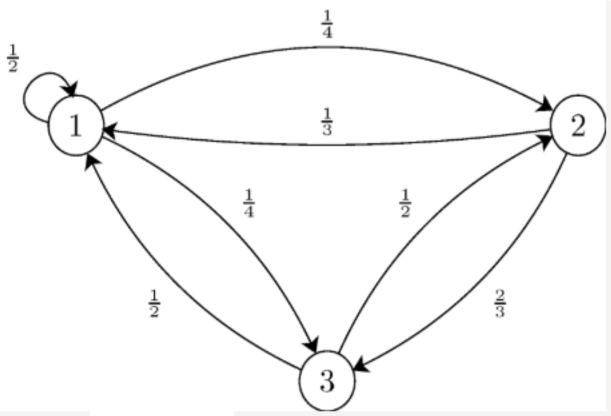
Consider the Markov chain with three states, $S = \{1, 2, 3\}$, that has the following transition matrix

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}.$$

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 $P(X_1 = 3, X_2 = 2, X_3 = 1)$.

$\boldsymbol{X}_1$ adımında S3 durumunda iken, $\boldsymbol{X}_2$ adımında S2 durumuna gidecek, sonunda $\boldsymbol{X}_3$ adımında S1 durumuna gidilecek.

c)



A state transition diagram.

d) 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.

$$\begin{aligned} P(X_1 = 3) &= 1 - P(X_1 = 1) - P(X_1 = 2) \\ &= 1 - \frac{1}{4} - \frac{1}{4} \\ &= \frac{1}{2}. \end{aligned}$$

Durum-3, ardından durum-2 ardından durum-1 gelme olasılığının hesaplanması,

$$\begin{aligned} P(X_1 = 3, X_2 = 2, X_3 = 1) &= P(X_1 = 3) \cdot p_{32} \cdot p_{21} \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} \\ &= \frac{1}{12}. \end{aligned}$$

Örnek:

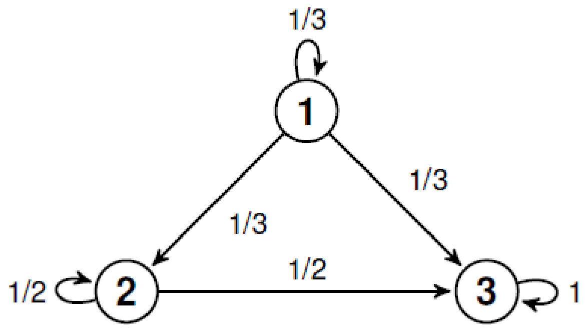
Şekildeki geçiş diyagramına karşılık gelen üç durumlu ayrık zamanlı Markov zincirini düşünün. X (0) 'ın başlangıç dağılımının f (1) = f (2) = 1/2 ile verildiğini varsayın.

Compute the following

1. P(X(0) = 1, X(1) = 2, X(2) = 3).

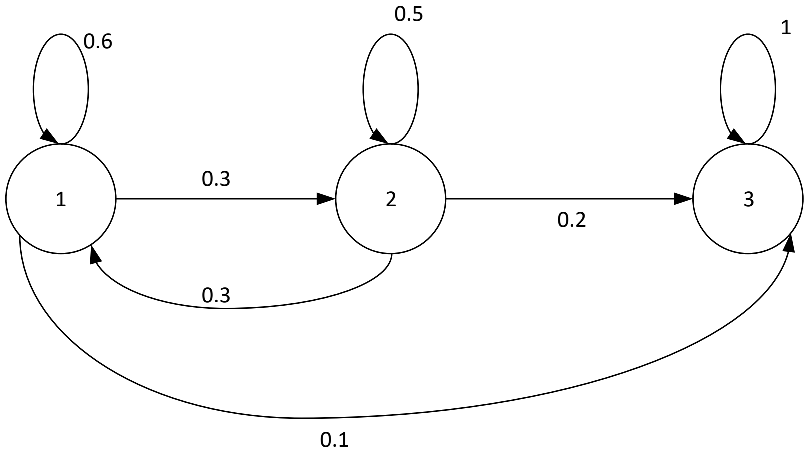
2. P(X(2) = i), for i = 1, 2, 3.

3. P(T3 = 2) where



P(X(0) = 1, X(1) = 2, X(2) = 3) = f(1) * P12 * P23 $= \frac{1}{2} * \frac{1}{3} * \frac{1}{2} = \frac{1}{12}$

Örnek:
Aşağıda geçiş diyagramı verilen Markov zincir analizinde,



a) 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) Bu matris geçiş matrisi olabilmesi için hangi özelliklere sahip olması gerekir? Bu matris bu özellikleri sağlıyor mu?

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.

c) 1 durumundan 3 durumuna geçme olasılığı, P13 ve 3 durumundan 2 durumuna geçme olasılığı, P32 nedir?

P13=0.1, P32=0

d) Geçiş diyagramı aşağıda verilen Markov zincir analizinde,
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.

e) 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?

$$P^8 = \begin{pmatrix} x & x & x \\ x & x & x \\ x & 0 & x \end{pmatrix}$$

Düzenli zincir, P geçiş olasılıkları matrisinin kuvvetlerinde bulunan elemanların sıfırdan farklı ve pozitif olmasını gerektirir. Bu matris ergodik değildir.

f) Durum geçiş matrisi aşağıda verilmiş olan Markov zincir analizinde başlangıç durumunda herbir 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?

$$P = \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$

$$V_s = \begin{pmatrix} 0.2 \\ 0.5 \\ 0.3 \end{pmatrix}$$

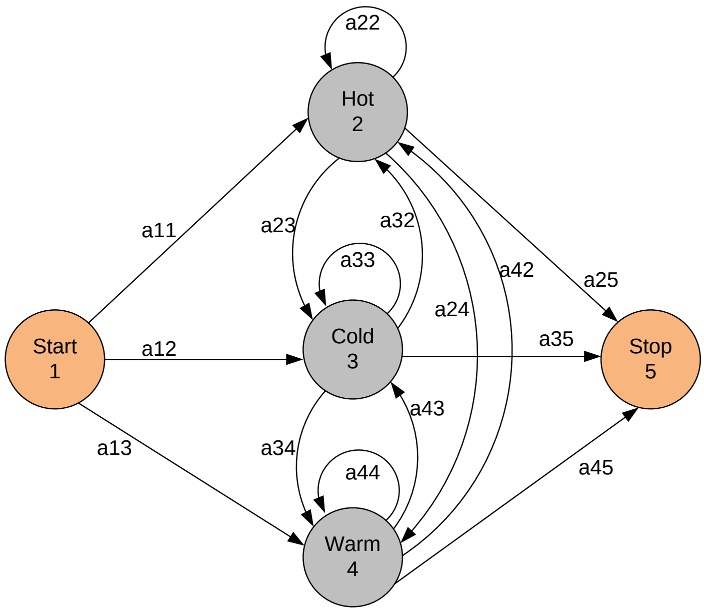
$$V_{s1} = \begin{pmatrix} 0.23 \\ 0.35 \\ 0.23 \end{pmatrix}$$

Ps-2=0.35

8.5. Saklı Markov Modeli (Hidden Markov Models)

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.

Bir durumdan diğer bir duruma gitme olasılığı 0 ile 1 arasında bir değer alır. Bir durumdan diğer tüm durumlara gitme olasılıkları toplamı 1'e eşittir. Herhangi bir durumun meydana gelme olasılığı diğer durumların bu duruma olasılıklarına bağlı olarak kendisi de dahil durumlardan herhangi biri olabilir.



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.*

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.

Geçiş olasılık matrisi,

$$\begin{pmatrix} a11\ a12 & \cdots & a1N \\ \vdots & \ddots & \vdots \\ aN1\ aN2 & \cdots & aNN \end{pmatrix}$$

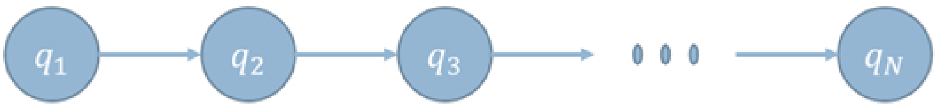
Geçiş matrisininde her bir eleman i-durumundan j-durumuna geçiş olasılığıdır.

$$\sum_{j=1}^{N} a_{ij} = 1 \quad \forall i$$

Durum geçişleri bir matris ile gösterilir.

Markov varsayımına göre, mevcut durumun olasılığı sadece kendinden önceki durumlara bağlıdır.

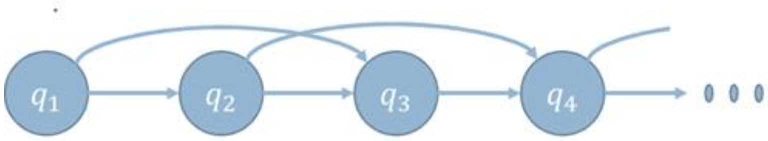
1. Derece Markov Varsayımı:



$$P(q_i|q_1 \ldots q_{i-1}) = P(q_i|q_{i-1})$$

Durum i’nin olasılığını hesaplarken sadece i-1'inci durum kullanılır. Yani kendinden 1 önceki durum kullanılır.

2. Derece Markov Varsayımı:

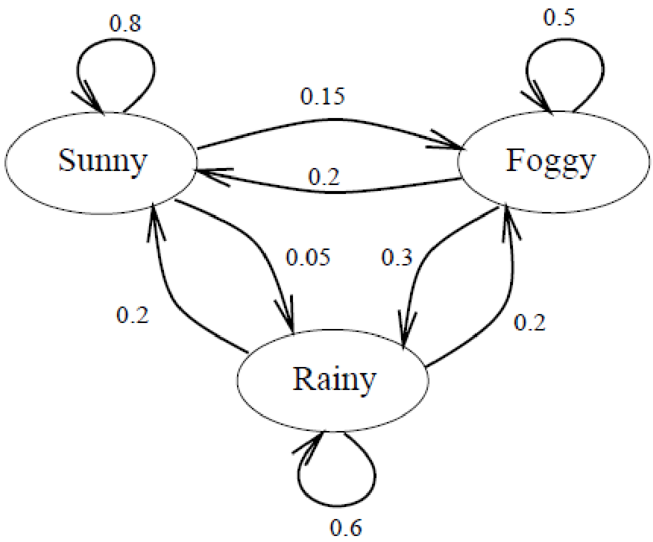


$$P(q_i|q_1 \ldots q_{i-1}) = P(q_i|q_{i-1}q_{i-2})$$

Durum i’nin olasılığı hesaplanırken i-1'inci ve i-2'nci durumlar kullanılır. Yani kendinden 1 önceki ve 2 önceki durumlar kullanlıır.

Özetlemek gerekirse, ardı ardına gelen olayların gerçekleşme olasılığı hesaplanırken Markov Zinciri yararlıdır.

Örnek:



Markov modelin yörünge dizisi olasılığıdır.

$$P(q_1, q_2, \dots, q_T) = \prod_{i=1}^{T} P(q_i \mid q_{i-1})$$

Koşullu olasılık: $P(A \cap B) = P(A|B)P(B)$

$$P(q_1, q_2, \dots, q_T) = P(q_1)P(q_2 \mid q_1) \dots P(q_{T-1} \mid q_{T-2})P(q_T \mid q_{T-1})$$

a) 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) 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

Markov zincir modeli ile Bayes Metotunun Bütünleştirilmesi:

Bayes Metotu: Herbir birimden aykırı olanların toplandığı kümede herhangi bir aykırı olanın hangi birimden geldiği hesaplayabilmek:
Önce aykırı olanların topladığı kümede herbir 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.

$$P(A\ /H) = \frac{P(A \cap H)}{P(A \cap H) + P(B \cap H)}$$

Aykırı olanın geldiği birim ile kesişim olasılığı, birimin olasılığı ile birimdeki aykırı olanların olasılığının çarpımına eşittir.

$$P(A \cap H) = P(A) * P(H\ /A)$$
$$P(B \cap H) = P(B) * P(H\ /B)$$

Örneğin, bir fabrikada 2 adet makine bulunmaktadır. Herbir 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.

- Herbir üretim bandında üretim olasıkları hesaplayınız. P(A)=0.5, P(B)=0.5.
- Herbir üretim bandı için A dan ve B den kaynaklanan hata olasılıklarını hesaplayınız.
  P(H/A)=0.10
  P(H/B)=0.05.

- Bu kusurlu motorun A, B makinesinden çıkma olasıklarını 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

- Toplam hatalı olanlardan A daki hatalı olanların olasılığı,
  P(A/H)= P(A∩H)/P(T)=0.05/0.075=50/75=2/3.

- Toplam hatalı olanlardan B daki hatalı olanların olasılığı,
  P(B/H)= P(B∩H)/P(T)=0.025/0.075=25/75=1/3

Saklı Markov Modeli Nedir?
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?

Örnek: 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.

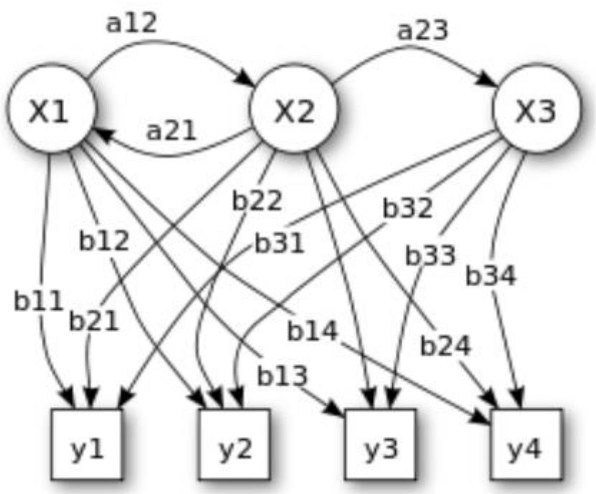
- Arkadaşına ulaşamayan arkadaşının yaşadığı şehirdeki hava durumuna bakarak arkadaşının nerede olduğunu öngörebilir mi?
- Ya da A şehrinde yaşayan B şehrindeki arkadaşının her gün ne yaptığına dayanarak, oradaki hava durumunu tahmin edebilir mi?

Tarihçesi
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.

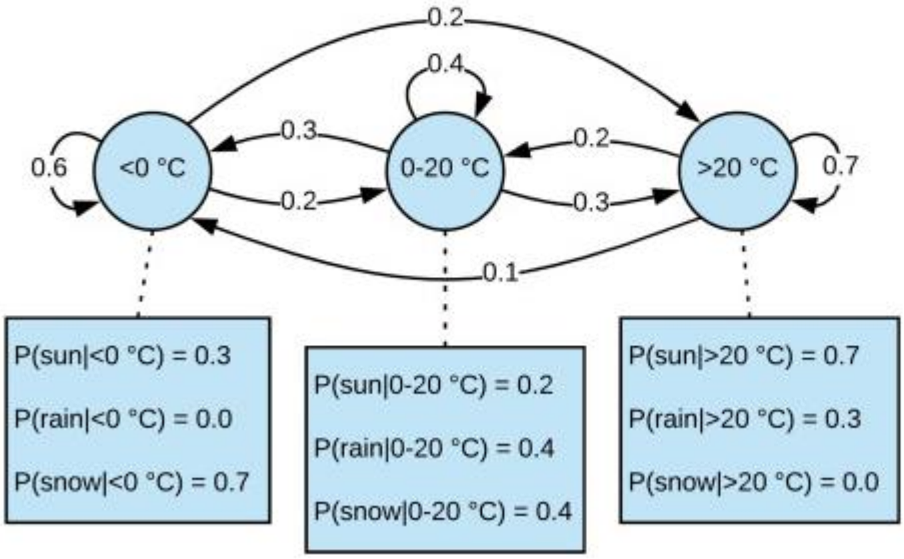
Saklı Markov Modeli Parametreleri:
Aşağıdaki şekilde görülen Saklı Markov Modeli'nin olasılık parametreleri şu şekildedir:

x — durumlar
y— olası gözlemler
a— geçiş olasılıkları durumu
b— çıkış olasılıkları

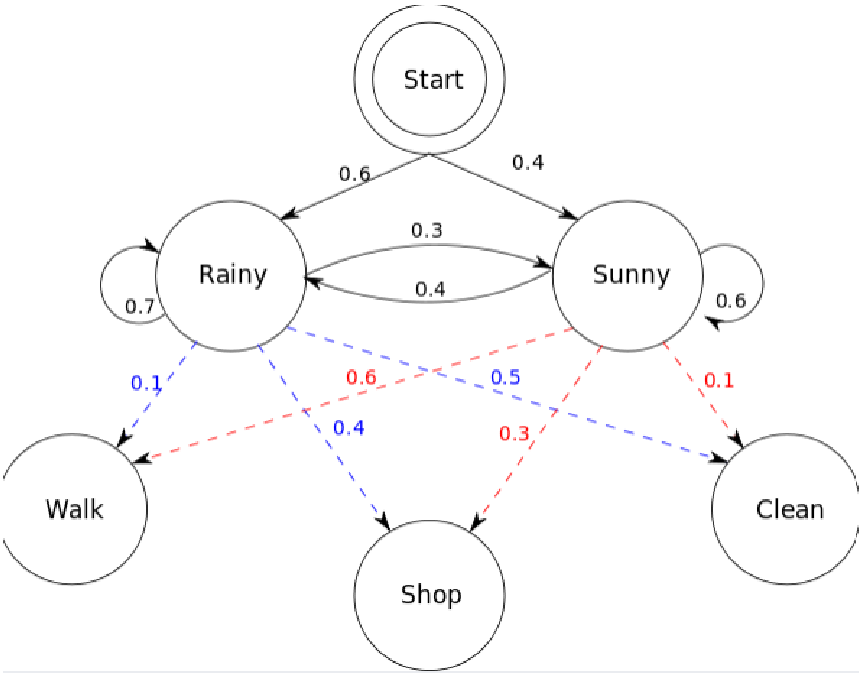
Normal Markov Model'inde, durumlar, gözlemci için görünebilirdir ve bu yüzden tek parametre, durum geçiş olasılıklarıdır. Saklı Markov Modelinde, duruma bağlı çıkışlar görünür.



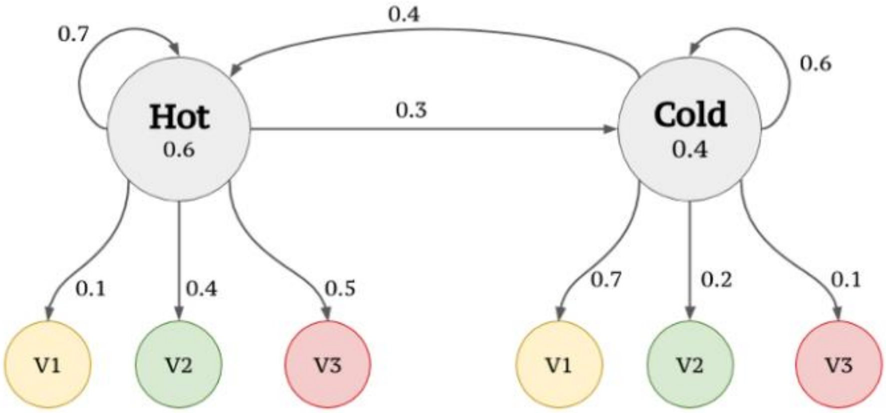
Hidden Markov Modelinin Olasılık Parametreleri



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.


Start
0.6
0.4
Rainy
0.3
0.4
Sunny
0.7
0.6
0.1
0.6
0.5
0.1
0.4
0.3
Walk
Clean
Shop

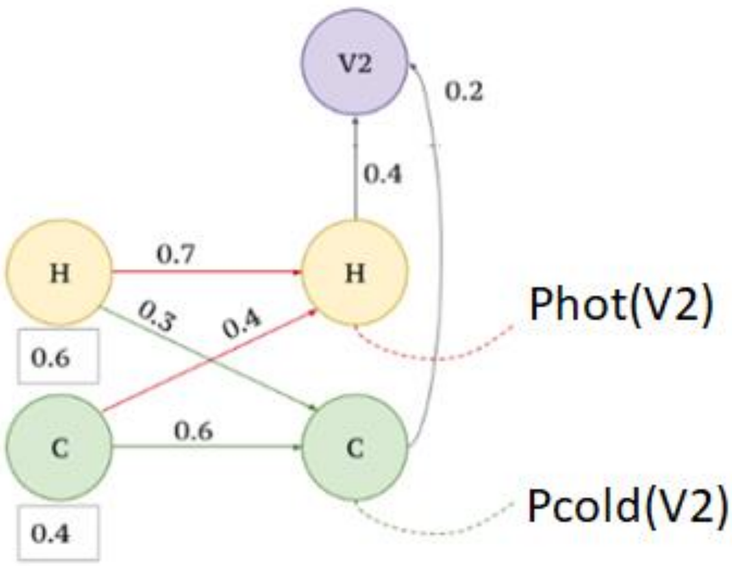
Örnek: V1, V2, V3 durumlarında sonlanma olasılıklarının hesaplanması



$$\begin{array}{c} A \\ \text{(State Transition Matrix)} \end{array} = \begin{array}{cc} & \begin{array}{cc} H & C \end{array} \\ \begin{array}{c} H \\ C \end{array} & \left\{ \begin{array}{cc} 0.7 & 0.3 \\ 0.4 & 0.6 \end{array} \right\} \end{array}$$

$$\begin{array}{c} B \\ \text{(Emission Matrix)} \end{array} = \begin{array}{cc} & \begin{array}{ccc} V1 & V2 & V3 \end{array} \\ \begin{array}{c} H \\ C \end{array} & \left\{ \begin{array}{ccc} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{array} \right\} \end{array}$$

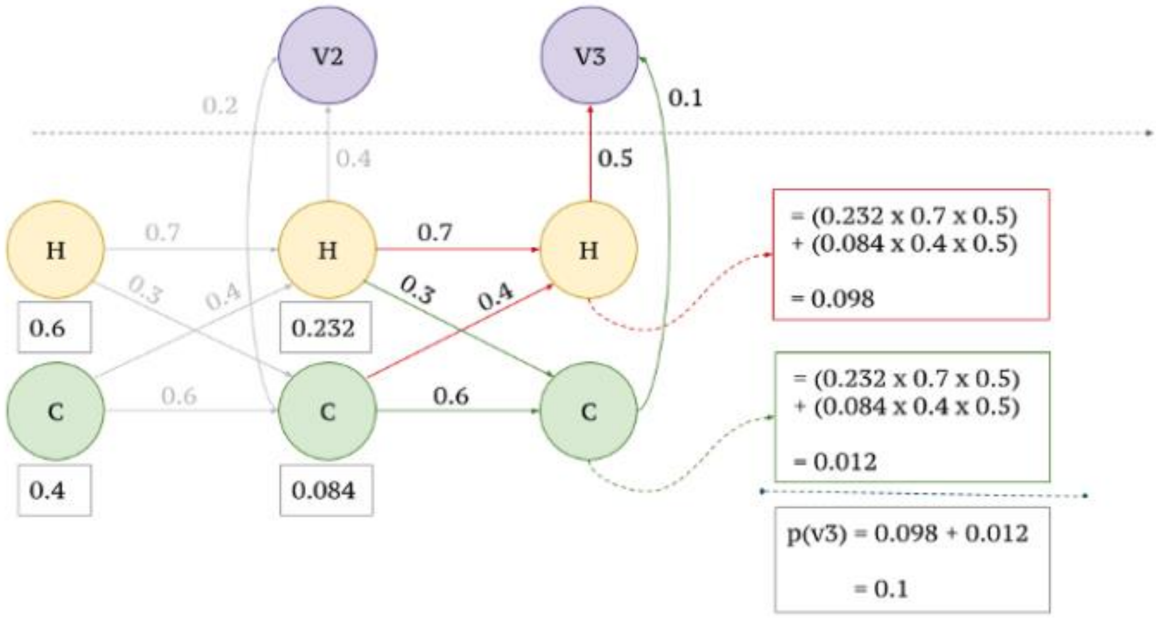
$$\begin{array}{c} \pi \\ \text{(Initial state S0)} \end{array} = \begin{array}{c} \begin{array}{cc} H & C \end{array} \\ \left\{ \begin{array}{cc} 0.6 & 0.4 \end{array} \right\} \end{array}$$



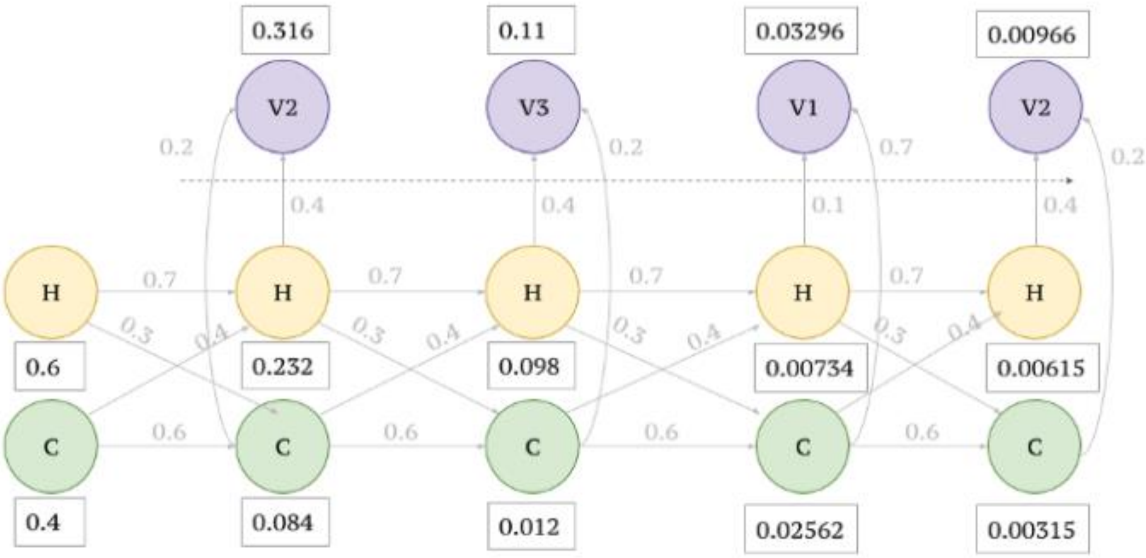
Başlangıç durumundan sıcak durumuna geçme olasılığı, 0.6; soğuk durumuna geçme olasılığı 0.4 olarak verilmiştir.

Sıcak durumdan V2 için karar verme olasılığını hesaplayalım.
Phot(V2)=0.6*0.7*0.4 + 0.4*0.4*0.4=0.232

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
0.3
0.4
0.3
0.4
0.6
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
0.6
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

Saklı Markov Modeli ve Tahmin Algoritmaları (Hidden Markov Model and Estimation Algorithms)

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

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 $x \in V\_T$, we intend to find the most likely series of states $z \in S\_T$. We can understand this with an example found below.

S - Sunny
R - Rainy
H - Happy
G - Grumpy

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 -> S = 8 0.8
S -> R = 2 0.2
R -> R = 3 0.6
R -> S = 2 0.4

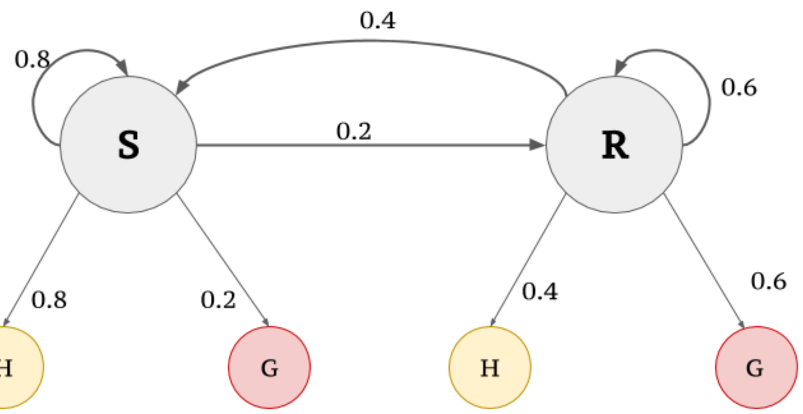


Fig. Data for example

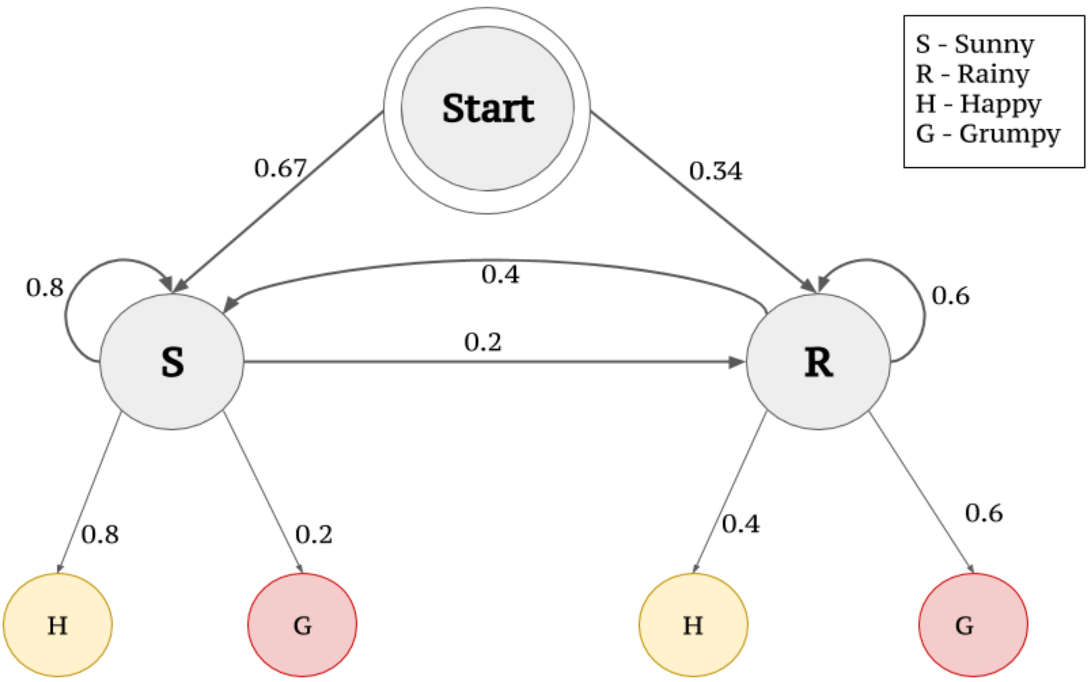


Fig. Markov Model as a Finite State Machine

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.

Duyguların sırasını düşünün: birbirini takip eden 6 gün boyunca H,H,G,G,G,H. Viterbi algoritmasını kullanarak serinin daha fazla olasılığını bulacağız.

Fig. The Viterbi algorithm requires to choose the best path

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).

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.

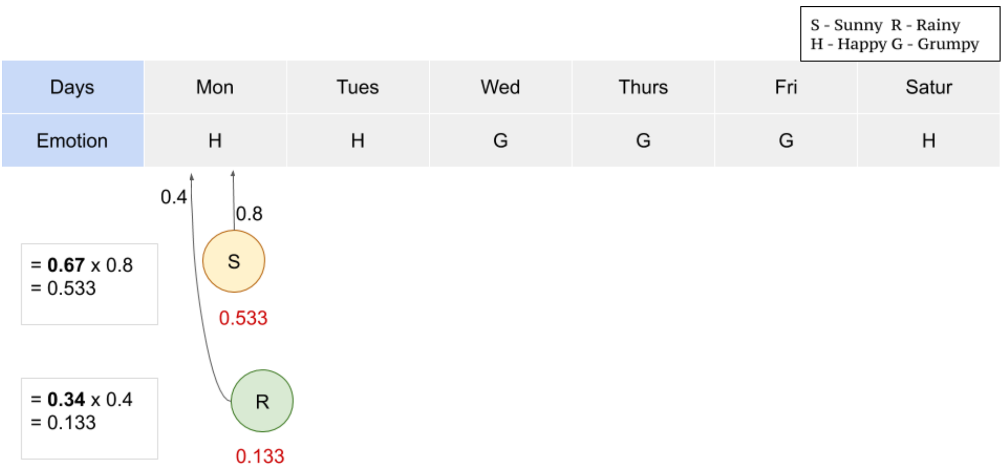


Fig. Step-1

S - Sunny R - Rainy
H - Happy G - Grumpy

Days	Mon	Tues	Wed	Thurs	Fri	Satur
Emotion	H	H	G	G	G	H

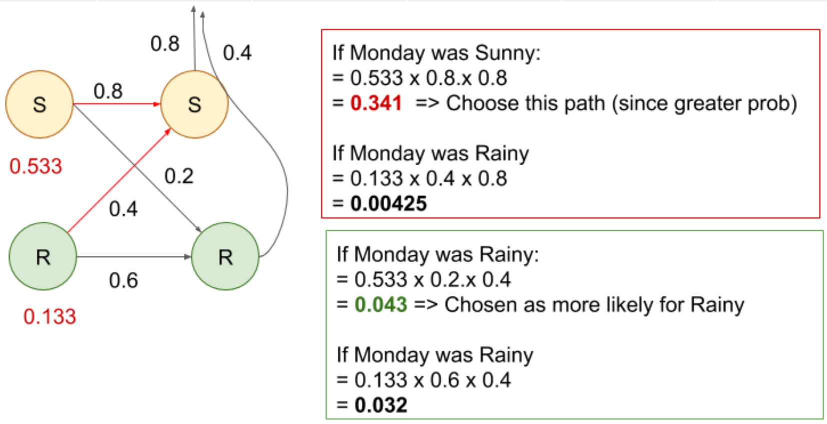
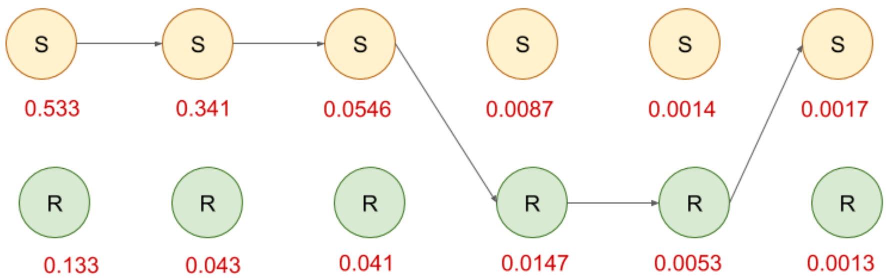


Fig. Step-2

S - Sunny R - Rainy
H - Happy G - Grumpy

Days	Mon	Tues	Wed	Thurs	Fri	Satur
Emotion	H	H	G	G	G	H



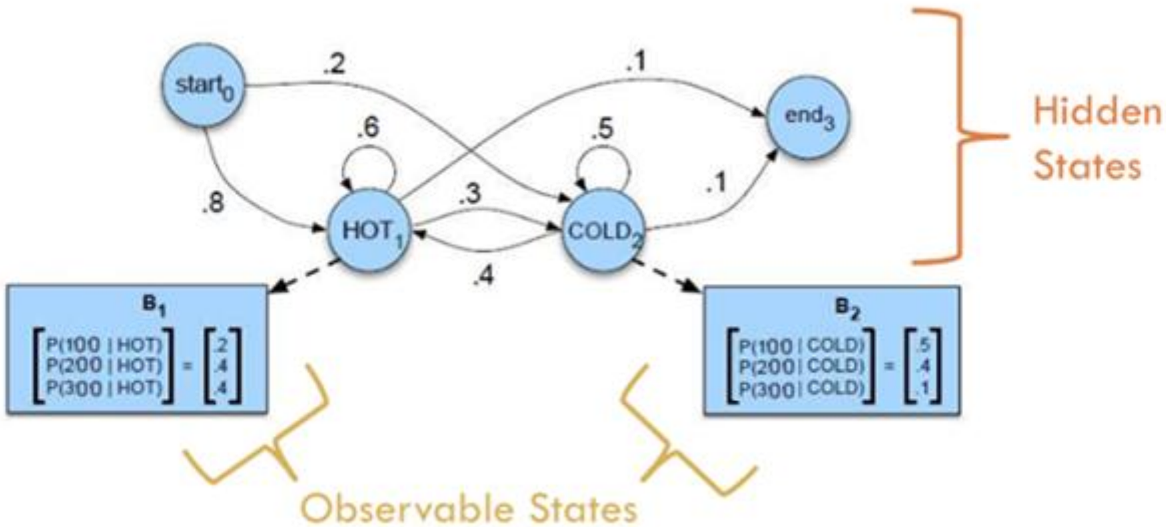
More likelihood sequence = **S S S R R S**
For the given output sequence H H G G G H

Şekil. En iyi yolu seçmek için algoritma yinelenir.

Algoritma size maksimum olabilirlik değerleri bırakır ve artık belirli bir çıktı dizisi için maksimum olabilirliğe sahip diziyi üretebiliriz.

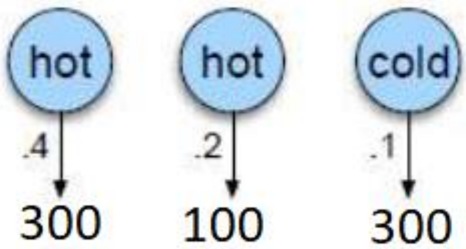
HMM parametreleri A ve B için değerlerin öğrenilmesi
HMM'lerde öğrenme, gözlemlenen bir diziyi büyük olasılıkla yapan durum geçiş olasılıkları A ve çıktı emisyon olasılıkları B'nin tahmin edilmesini içerir. Bu amaçla beklenti-maksimizasyon algoritmaları kullanılmaktadır. Baum-Welch algoritması olarak bilinen ve bu kategoriye giren ve ileri algoritmayı kullanan bir algoritma yaygın olarak kullanılmaktadır.

Örnek:
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?



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.

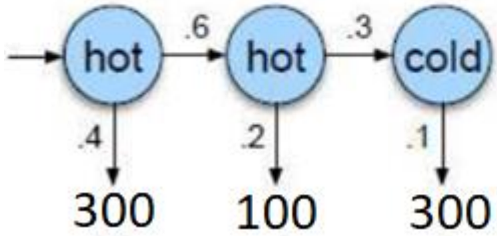
Olasılık (Likelihood):
Durumlar arası geçiş olasılıklar ve her durumda her gözlemin gerçekleşme olasılığı bilinirken, "300adet 100adet 300adet" buz blok satış serisinin gelme olasılığı nedir? Bu soruyu cevaplamak zor, çünkü gözlem serisindeki her gözlemin hangi durumda yapıldığı bilinmiyor. Ama eğer sırasıyla **Hot, Hot, Cold** durumlarında iken bu gözlem serisini elde edildiyse, olasılık kolayca hesaplanabilir. Söz gelimi,



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.

$$P(O|Q) = \prod_{i=1}^{T} P(o_i \mid q_i)$$

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,

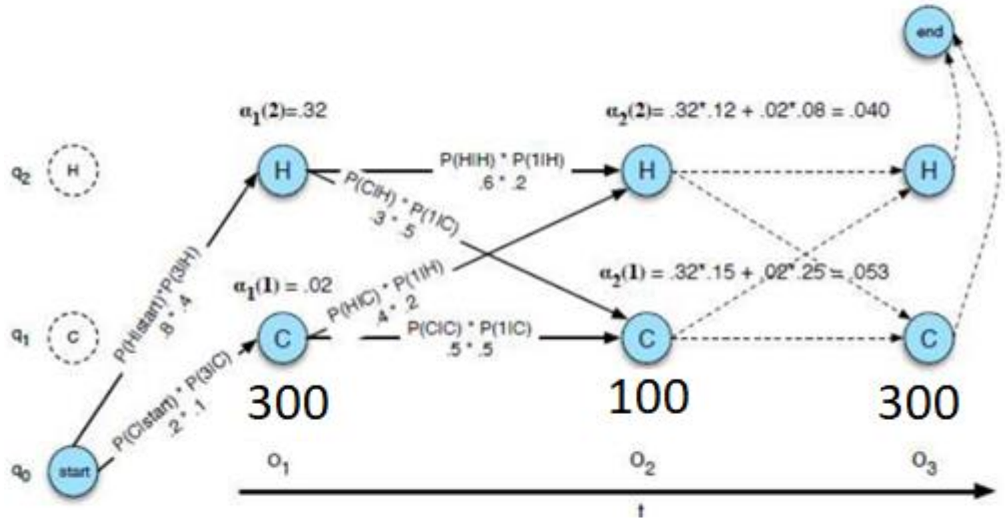


$$P(O,Q) = P(O|Q)P(Q) = \prod_{i=1}^{T} P(o_i \mid q_i) \prod_{i=1}^{T} P(q_i \mid q_{i-1})$$

Start → Hot olasılığı 0.8
Hot → Hot olasılığı 0.6
Hot → Cold olasılığı 0.3

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.

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.



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

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).

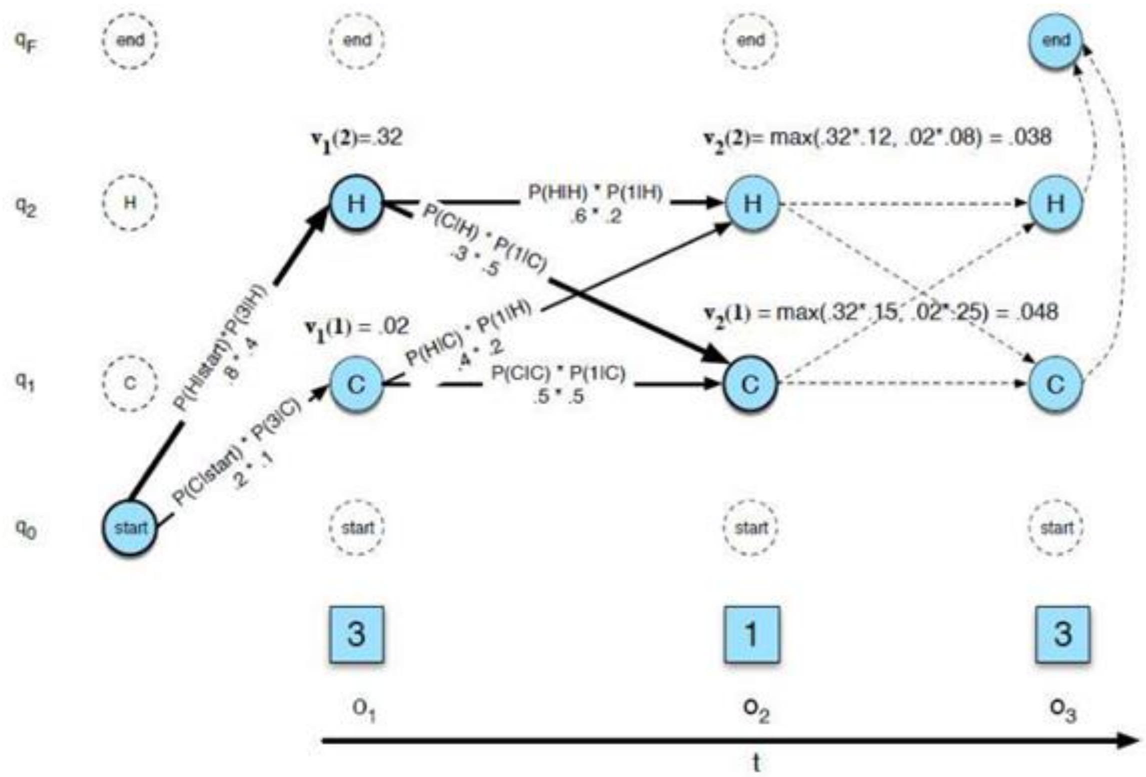
Toplamın sol kısmını daha önce hesapladığımız Start → Hot ve Hot → Hot olasılıklarını kullanarak hesaplarız: 0.32 * 0.12. Toplamın sağ kısmını daha önce hesapladığımız Start → Cold ve Cold → Hot olasılıklarını kullanarak hesaplarız: 0.02 * 0.08. Bu iki çarpımı toplayınca da mevcut state'de olma ihtimalini buluruz.

Formüle dökmek 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.

Decoding:
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.



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.
Formüle dökersek:

$$v_t(j) = max_{i=1}^{N} v_{t-1}(i) a_{ij} b_j(o_t)$$

- 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.

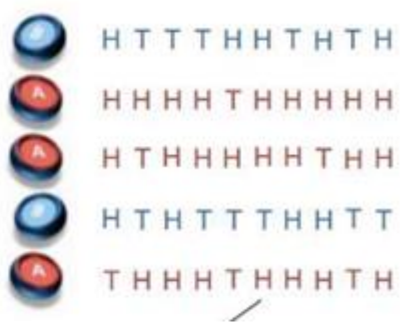
Öğrenme (Learning):
Öğrenme Probleminde ise durum serisi ve gözlem serisi gizli değildir. Amaç durumlar arası geçiş olasılıklarını ve hangi durumda hangi gözlemin gerçekleşeceği olasılığını hesaplamaktır.

$$a_{ij} = \frac{Count(i \rightarrow j)}{\sum_{q\epsilon} \; Count(i \rightarrow q)}$$

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.

Beklenti maksimizasyon (Expectation Maximization):
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

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:



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. 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 niceliksellеştirir. 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.

9.1. Kanal Kapasitesi

Claude Shannon (1916 – 2001) 1948 yılında yazdığı, "İletişimin Matematiksel Teorisi" isimli kitapcığı yirminci yüzyılın en önemli bilimsel kitapcıklarından biridir. Shannon, bir mesaj içerisindeki bilgi miktarını ölçmenin ve değerlendirmenin 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. ***Bilgiyi ikili sayı sistemine dönüştürmenin oldukça güçlü bir hareket olduğunu fark etti.*** Bit: 0/1 tanımlandı. Bit, bilginin sayısal dünyadaki en küçük miktarıdır. Bilgi ölçülebilen bir güce, gerçeğe dönüştürüldü.

Bilgi ölçülebiliyordu. Bilgi sadece insanların var ettiği birşey değil. Kainattaki herhangi bir olay inanılmaz bilgiler ve mesajlar içerebilir. Normalde göremeyeceğimiz tüm fiziksel, kimyasal olyalar normal bir film gibi bizlere izletilebiliyor. Bilgi kainatın ayrılmaz bir parçasıdır. Bilgi heryerdedir. Sinyal ile bilgi sonsuzluğa ilerler.

Bilgi sinyal olarak taşınır ve anlamlandırılır; bilgi bir taşa, bir kitaba yazılır. Bir belleğe ya da beyine yazılır. Bilgi taşınıyor ve onu taşıyan sinyaldir. Sinyal, bilginin fizik kanunlarına göre nasıl 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 bilgi taşıyan sinyalleri hızla bir yerlerden bir yerlere gönderdik. Bilgiyi taşıyan sinyaller ona sıra dışı özellikler sağlamaktadır. Yakın gelecekte bilgi, elektron olarak taşınacak ve işlenecektir.

Bilgi ile enerji arasındaki ilişki nedir? Bilgi soyut değil. Kağıda yazabileceğiniz bir formül bilgi değildir. Bilgi taşınıyor ve anlamlandırılıyor; bilgi bir taşa, bir kitaba yazılır. Bir belleğe ya da beyine yazılır. Souçta bilgi taşınıyor 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.

Nyquist Teoremi:

Nyquist teoremi gürültüsüz bir kanal için kanal kapasitesini hesaplar. m: bir sembolün temsil edildiği bit sayısıdır. Örnekleme frekansı, sinyalin band genişliğinden ya da maksimum frekansından 2 kat fazla olmak zorundadır.

Given bandwidth B, highest signal rate is 2B. Given binary signal, data rate supported by B Hz is 2B bps. Can be increased by using M signal levels, C= 2B log2M

Gürültüsüz kanal kapasitesi C, bps olan veri transferi için n örnekleme aralığı seviye sayısı için band genişliğini hesaplar.

Örnek:

C=1Mbps kaç bps eder?

C=1Mbps=10^6 bps

Örnekleme aralığı seviye sayısı n=1024 ve n=2^m ise m'i bulunur. m=10

C=2Bm formülünden B band genişliğini KHz olarak hesaplayınız.

10^6 =2B10 ise B=50x 10^3 Hz= 50KHz dir.

Shannon's Theorem:

Gürültülü kanalda işaretin gürültüye oranı Shannon Kapasite teoremi ile hesaplanır. 1 + SNR değeri 2' nin üssü biçimine ötelenir. Çünkü 2 tabanında logaritma alma basitleşmiş olur. $\text{Log}_2\ 2^n$=n dir. C = B $\log_2$(1 + SNR). C=n*B olur.

SNR genellikle 10 tabanında desibel olarak verilir.

10Log_{10} SNR=SNRdB ise SNR=$10^{SNRdB/10}$ dir.

$10^{SNRdB/10}=2^n$ ise SNRdB *Log_{10}10 /10=n*$\log_{10}$ 2 dir. Log_{10} 2=0.3 dir. SNRdB/3=n elde edilir.

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.

9.2. 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.

9.3. 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 U B)=P(A) + P(B) – P(A∩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(Ω\E)=1-P(E)

Bir başka deyimle herhangi bir olayın **olmama** olasılığı **1** eksi olayın **olma** (ortaya çıkma) olasılığıdır.

9.4. 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.

9.5. 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.

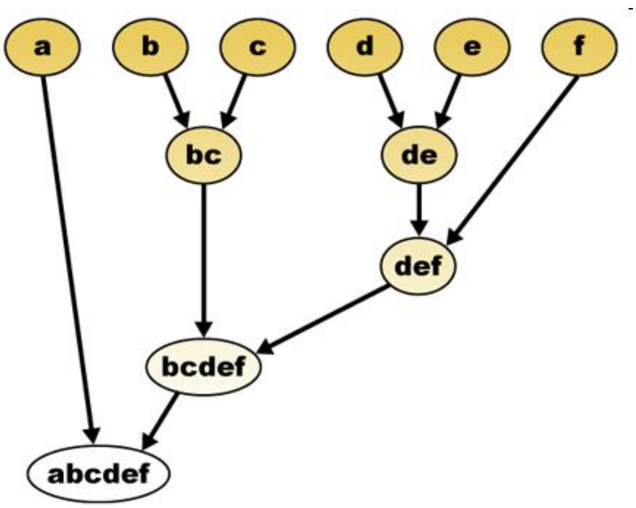
Önce *P* algoritmasının uzunluğunu irdeleyelim. *n* sayısının algoritmik karmaşıklığı yaklaşık olarak log_2n dir. Yaklaşık diyoruz, çünkü algoritmanın gerçek uzunluğu kullanılan makina diline bağlıdır. Yeterince büyük *n* sayıları için log_2n sayısı *n* sayısından çok küçüktür, dolayısıyla algoritmanın 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.



9.6. 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 Ω=∑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.

9.7. Information Theory

Araştırmacılar, 1900'lerin başından beri bilgiyi nicelleştirme üzerine kafa yordular ve 1948'de Claude Shannon, "A Mathematical Theory of Communication" adlı olağanüstü 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 vb. 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 temel temelleri kayıpsız veri sıkıştırma, kayıplı veri sıkıştırma ve kanal kodlama alanlarında uygulanmaktadır.

Bilgi Teorisinde kullanılan teknikler doğası gereği olasılıklıdır ve genellikle 2 spesifik nicelik ile ilgilenir, yani. Entropi ve Karşılıklı Bilgi. Bu iki terime daha derin bir dalış yapalım.

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{*χ* = *x*}, *x* ∈ *χ*. With this in mind, entropy can be defined as

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x).$$

Entropy

The unit of entropy is bit. If you observe the formula, entropy is entirely dependent on the **probability of the random variable** and not on the value of *x* itself. There is a negative sign in front of the formula, making it perpetually positive or 0. If entropy is 0, there is no new information to be gained. I will demonstrate the implementation of this formula through an example.

Consider the scenario of a coin toss. There are two probable outcomes, heads or tails, with equal probabilities. If we quantify that, *x* ∈ {heads,tails}, and *p*(heads) = 0.5 and *p*(tails) = 0.5. If we plug in these values in the formula:

$$\begin{aligned} H(x) = & -\left[\left(p(x_{heads}) * \log\left(p(x_{heads})\right)\right) + \left(p(x_{tails}) * \log\left(p(x_{tails})\right)\right)\right] \\ = & -[(0.5 * log_2 0.5) + (0.5 * log_2 0.5)] = -[(-0.5) + (-0.5)] = 1 \end{aligned}$$

Calculating Entropy in a coin toss event

Therefore, entropy is 1 bit, i.e., the coin toss's outcome can be expressed completely in 1 bit. So, to intuitively express Shannon entropy's concept, it is understood as "how long does a message need to be to convey its value completely". I want to dive a little deeper and discuss the concepts of Joint Entropy, Conditional Entropy and Relative Entropy.

Joint and Conditional Entropy

Previously, I defined Entropy for a single random variable, but now I will extend it to a pair of random variables. It is a simple aggregation as we can define the pair of variables (*X*, *Y*) as a single vector-valued random variable.

The joint entropy *H*(*X*, *Y*) of a pair of discrete random variables (*X*, *Y*) with a joint distribution *p*(*x*, *y*) is defined as

$$H(X, Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y)$$

Joint Entropy

This can also be represented in terms of expected value.

$$H(X, Y) = -E \log p(X, Y)$$

Joint Entropy (Expected Value form)

Similarly, for conditional entropy, *H*(*Y*|X) is defined as:

$$H(Y|X) = \sum_{x \in \mathcal{X}} p(x) H(Y|X = x)$$

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.

$$H(Y|X) = -E \log p(Y|X).$$

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

Now, I will represent these questions' response as two different discrete variables belonging to a joint distribution.

Activity	**Cancer**
Smoking	Yes
Smoking	Yes
Smoking	Yes
Alcohol	No
Neither	Yes
Alcohol	Yes
Alcohol	Yes
Smoking	No
Alcohol	Yes
Neither	No

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.

	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.

$$p(X = \text{"Smoking"}) = \frac{4}{10}$$

$$p(X = \text{"}Alcohol\text{"}) = \frac{4}{10}$$

$$p(X = \text{"}Neither\text{"}) = \frac{2}{10}$$

Marginal Probability of X

Based on the probability table, we can plug in the value in the conditional entropy formula.

$$\begin{aligned}
\mathrm{H(Y|X)} &= \mathrm{p(\text{"Smoking"})H(Y|\,X = \text{"Smoking"})} + \mathrm{p(\text{"Alcohol"})H(Y|\,X = \text{"Alcohol"})} + \mathrm{p(\text{"Neither"})H(Y|\,X = \text{"Neither"})} \\
&= -\frac{4}{10}\left\{\frac{3}{10}\log\left(\frac{3}{10}\right) + \frac{1}{10}\log\left(\frac{1}{10}\right)\right\} - \frac{4}{10}\left\{\frac{3}{10}\log\left(\frac{3}{10}\right) + \frac{1}{10}\log\left(\frac{1}{10}\right)\right\} - \frac{2}{10}\left\{\frac{1}{10}\log\left(\frac{1}{10}\right) + \frac{1}{10}\log\left(\frac{1}{10}\right)\right\} \\
&= 0.4 * 0.857 + 0.4 * 0.857 + 0.2 * 0.664 \\
&= 0.8184\ bits
\end{aligned}$$

Conditional Probability in the above example

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:

$$D(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$$

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.

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,

$$D(p||q) = (1 - r)\log\left[\frac{1 - r}{1 - s}\right] + rlog\frac{r}{s}$$

Relative Entropy for p||q
I would also like to demonstrate the non-symmetric property, so I will also calculate *D*(*q*||*p*).

$$D(q||p) = (1-s)\log\left[\frac{1-s}{1-r}\right] + s log\frac{s}{r}$$

Relative Entropy for q||p
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.

$$D(p||q) = \left(1-\frac{1}{2}\right)\log\left[\frac{1-\frac{1}{2}}{1-\frac{1}{4}}\right] + \frac{1}{2}log\frac{\frac{1}{2}}{\frac{1}{4}} = \ 0.2075 \text{ bit}$$

$$D(q||p) = \left(1-\frac{1}{4}\right)\log\left[\frac{1-\frac{1}{4}}{1-\frac{1}{2}}\right] + \frac{1}{4}log\frac{\frac{1}{4}}{\frac{1}{2}} = \ 0.1887 \text{ bit}$$

As you can see, *D*(*p*||*q*) ≠ *D*(*q*||*p*).

Now, since we have discussed the different types of entropies, we can move onto Mutual Information.

Mutual Information
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:

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*).

$$\begin{aligned} I(X;Y) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\ &= D(p(x,y)||p(x)p(y)) \end{aligned}$$

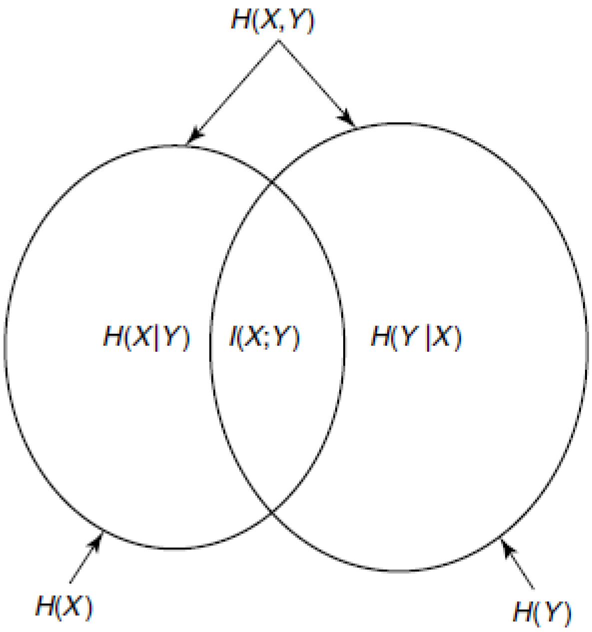
Mutual Information

Mutual Information can also be expressed in terms of Entropy. The derivation is quite fun, but I will refrain myself as it might clutter the article.

$$I\,(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Mutual Information w.r.t. Entropy

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.



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:

$$\begin{aligned} H(Y) &= -\sum_{y \in Y} p(y) \log p(y) \\ &= \frac{7}{10} \log \frac{7}{10} + \frac{3}{10} \log \frac{3}{10} \\ &= 0.876 \end{aligned}$$

Therefore, Mutual Information is defined by:

$$
\begin{aligned}
I\,(X;Y) &= H(Y) - H(Y|X) \\
&= 0.876 - 0.8184 \\
&= 0.0576
\end{aligned}
$$

Entropi (Bilgi Kazancı)

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.

$$I(x) = \log_2 \frac{1}{P(x)} = -\log_2 P(x)$$

Shannon'a göre entropi, iletilen bir mesajın taşıdığı enformasyonun beklenen değeridir.
Shannon Entropisi (H) adıyla anılan terim, tüm x_i durumlarına ait $P(x_i)$ olasılıklarına bağlı bir değerdir.

$$H(X) = E(I(X)) = \sum_{1 \le i \le n} P(x_i).I(x_i) \;\; = \sum_{i=1}^{n} P(x_i) \log_2 \frac{1}{P(x_i)} = -\sum_{i=1}^{n} P_i \log_2 P_i$$

$$Log_2(P) = \frac{10}{3} Log_{10}(P)$$

$$H(X) = -\frac{10}{3} \sum_{i=1}^{n} P_i \, Log_{10} P_i$$

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; $\log a^n = n*\log a$

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.

$$H(X) = -\sum_{i=1}^{2} p_i \log_2 p_i \quad = -(0.5\log_2 0.5 + 0.5\log_2 0.5) = 1$$

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

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

10. 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