CHAPTER 1: BACKGROUND

Subject: Introduction to data science

Instructor: Đinh Xuân Trường
Posts and Telecommunications Institute of Technology
Faculty of Information Technology 1

Hanoi, 2024
http://www.ptit.edu.vn

Content

Content

Vectors

Definition

An ordered sequence $(a_1, a_2, \ldots, a_n)$ of real numbers is called an ordered ***n*-tuple**. In other words,

$$(a_1, a_2, \ldots, a_n) = (b_1, b_2, \ldots, b_n) \quad \text{ if and only if}$$

$$a_1 = b_1, a_2 = b_2, \ldots, \text{ and } a_n = b_n.$$

Definition

Let $\mathbb{R}$ denote the set of all real numbers. The set of all ordered n-tuples from $\mathbb{R}$ has a special notation $\mathbb{R}^n$ denoted the set of all ordered n-tuples of real numbers.

Vectors

- ⋆ There are two commonly used ways to denote the n-tuples in $\mathbb{R}^n$: As rows $(r_1, r_2, \ldots, r_n)$ or columns $\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}$.
- ⋆ If x and y are two n-vectors in $\mathbb{R}^n$, it is clear that their matrix sum $\mathrm{x} + \mathrm{y}$ is also in $\mathbb{R}^n$ as is the scalar multiple $k\mathrm{x}$ for any real number k.
- ⋆ $\mathbb{R}^n$ is closed under addition and scalar multiplication.

Content

Definition

A rectangular array of numbers is called a **matrix** (the plural is **matrices**), and the numbers are called the **entries** of the matrix, denoted by

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

- ⋆ In general, a matrix with m rows and n columns is referred to as an $\boldsymbol{m} \times \boldsymbol{n}$ **matrix** or as having **size** $\boldsymbol{m} \times \boldsymbol{n}$.
- ⋆ Matrices of size $n \times n$ for some n are called **square matrices**.
- ⋆ A matrix of size $1 \times n$ is called a **row matrix**, whereas one of size $m \times 1$ is called a **column matrix**.

Matrix-Vector Multiplication

Definition

Let $A = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{bmatrix}$ be an $m \times n$ matrix, written in terms of its columns $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$. If $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is any n-vector, the product $A\boldsymbol{x}$ is defined to be the m-vector given by:

$$A\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n$$

In other words, if A is $m \times n$ and x is an n-vector, the product $A\boldsymbol{x}$ is the linear combination of the columns of A where the coefficients are the entries of $\boldsymbol{x}$ (in order).

Matrix-Vector Multiplication

Theorem

1. *Every system of linear equations has the form* $A\mathbf{x} = \mathbf{b}$ *where* A *is the coefficient matrix,* $\mathbf{b}$ *is the constant matrix, and* $\mathbf{x}$ *is the matrix of variables.*
2. *The system* $A\mathbf{x} = \mathbf{b}$ *is consistent if and only if* $\mathbf{b}$ *is a linear combination of the columns of* A*.*
3. *If* $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ *are the columns of* A *and if* $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$*, then* $\mathbf{x}$ *is a solution to the linear system* $A\mathbf{x} = \mathbf{b}$ *if and only if* $x_1, x_2, \ldots, x_n$ *are a solution of the vector equation*

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$$

Matrix Addition

If A and B are matrices of the same size, their sum $A+B$ is the matrix formed by adding corresponding entries.

If $A=[a_{ij}]$ and $B=[b_{ij}]$, this takes the form

$$A+B=[a_{ij}+b_{ij}]$$

Note that addition is not defined for matrices of different sizes.

Example

If $A=\begin{bmatrix} 2 & 1 & 3 \\ -1 & 2 & 0 \end{bmatrix}$ and $B=\begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & 6 \end{bmatrix}$, compute $A+B$.

Solution.

$$A+B=\begin{bmatrix} 2+1 & 1+1 & 3-1 \\ -1+2 & 2+0 & 0+6 \end{bmatrix}=\begin{bmatrix} 3 & 2 & 2 \\ 1 & 2 & 6 \end{bmatrix}$$

Scalar Multiplication

if A is any matrix and k is any number, the scalar multiple kA is the matrix obtained from A by multiplying each entry of A by k. If $A = [a_{ij}]$, this is

$$kA = [ka_{ij}]$$

Example

If $A = \begin{bmatrix} 3 & -1 & 4 \\ 2 & 0 & 6 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 3 & 2 \end{bmatrix}$ compute $5A$, $\frac{1}{2}B$, and $3A - 2B$. Solution.

$$5A = \begin{bmatrix} 15 & -5 & 20 \\ 10 & 0 & 30 \end{bmatrix}, \quad \frac{1}{2}B = \begin{bmatrix} \frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & \frac{3}{2} & 1 \end{bmatrix}$$

$$3A - 2B = \begin{bmatrix} 9 & -3 & 12 \\ 6 & 0 & 18 \end{bmatrix} - \begin{bmatrix} 2 & 4 & -2 \\ 0 & 6 & 4 \end{bmatrix} = \begin{bmatrix} 7 & -7 & 14 \\ 6 & -6 & 14 \end{bmatrix}$$

Scalar Multiplication

Theorem

Let A, B, and C denote arbitrary $m \times n$ matrices where m and n are fixed. Let k and p denote arbitrary real numbers. Then

1. $A + B = B + A$.
2. $A + (B + C) = (A + B) + C$.
3. *There is an $m \times n$ matrix 0 , such that* $0 + A = A$ *for each* A.
4. *For each A there is an $m \times n$ matrix, $-A$, such that* $A + (-A) = 0$.
5. $k(A + B) = kA + kB$.
6. $(k + p)A = kA + pA$.
7. $(kp)A = k(pA)$.
8. $1A = A$.

Transpose of a Matrix

Definition

If A is an $m \times n$ matrix, the transpose of A, written A^T, is the $n \times m$ matrix whose rows are just the columns of A in the same order.

Example

Write down the transpose of each of the following matrices.

$$A = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \quad B = \begin{bmatrix} 5 & 2 & 6 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

$$D = \begin{bmatrix} 3 & 1 & -1 \\ 1 & 3 & 2 \\ -1 & 2 & 1 \end{bmatrix}$$

Transpose of a Matrix

This is useful in verifying the following properties of transposition.

Theorem

Let A and B denote matrices of the same size, and let k denote a scalar.

1. *If A is an $m \times n$ matrix, then A^T is an $n \times m$ matrix.*
2. $\left(A^T\right)^T = A$.
3. $(kA)^T = kA^T$.
4. $(A + B)^T = A^T + B^T$.

Matrix Multiplication

Definition

Let A be an $m \times n$ matrix, let B be an $n \times k$ matrix, and write $B = \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_k \end{bmatrix}$ where $\boldsymbol{b}_j$ is column j of B for each j. The product matrix AB is the $m \times k$ matrix defined as follows:

$$AB = A \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_k \end{bmatrix} = \begin{bmatrix} A\boldsymbol{b}_1 & A\boldsymbol{b}_2 & \cdots & A\boldsymbol{b}_k \end{bmatrix}$$

Example

$$\text{Compute } AB \text{ if } A = \begin{bmatrix} 2 & 3 & 5 \\ 1 & 4 & 7 \\ 0 & 1 & 8 \end{bmatrix} \text{ and } B = \begin{bmatrix} 8 & 9 \\ 7 & 2 \\ 6 & 1 \end{bmatrix}$$

Matrix Inverses

Definition

If A is a square matrix, a matrix B is called an inverse of A if and only if

$$AB = I \quad \text{and} \quad BA = I$$

A matrix A that has an inverse is called an invertible matrix.

Example

Show that $B = \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix}$ is an inverse of $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$.

Matrix Inverses

Matrix Inversion Algorithm

If A is an invertible (square) matrix, there exists a sequence of elementary row operations that carry A to the identity matrix I of the same size, written $A \to I$. This same series of row operations carries I to A^{-1}; that is, $I \to A^{-1}$. The algorithm can be summarized as follows:

$$\begin{bmatrix} A & I \end{bmatrix} \to \begin{bmatrix} I & A^{-1} \end{bmatrix}$$

where the row operations on A and I are carried out simultaneously.

Example

Use the inversion algorithm to find the inverse of the matrix

$$A = \begin{bmatrix} 2 & 7 & 1 \\ 1 & 4 & -1 \\ 1 & 3 & 0 \end{bmatrix}$$

Matrix Inverses

Theorem

All the following matrices are square matrices of the same size.

1. I *is invertible and* $I^{-1} = I$.
2. *If* A *is invertible, so is* A^{-1}, *and* $\left(A^{-1}\right)^{-1} = A$.
3. *If* A *and* B *are invertible, so is* AB, *and* $(AB)^{-1} = B^{-1}A^{-1}$.
4. *If* $A_1, A_2, \ldots, A_k$ *are all invertible, so is their product* $A_1A_2\cdots A_k$, *and* $(A_1A_2\cdots A_k)^{-1} = A_k^{-1}\cdots A_2^{-1}A_1^{-1}$.
5. *If* A *is invertible, so is* A^k *for any* $k \geq 1$, *and* $\left(A^k\right)^{-1} = \left(A^{-1}\right)^k$.
6. *If* A *is invertible and* $a \neq 0$ *is a number, then* aA *is invertible and* $(aA)^{-1} = \frac{1}{a}A^{-1}$.
7. *If* A *is invertible, so is its transpose* A^T, *and* $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$.

Content

Cofactors of a Matrix

Definition

Assume that determinants of $(n-1) \times (n-1)$ matrices have been defined. Given the $n \times n$ matrix A, let A_{ij} denote the $(n-1) \times (n-1)$ matrix obtained from A by deleting row i and column j. Then the (i,j)-cofactor $c_{ij}(A)$ is the scalar defined by

$$c_{ij}(A) = (-1)^{i+j} \det (A_{ij})$$

Here $(-1)^{i+j}$ is called the sign of the (i,j)-position.

Example

Find the cofactors of positions $(1,2), (3,1)$, and $(2,3)$ in the following matrix.

$$A = \begin{bmatrix} 3 & -1 & 6 \\ 5 & 2 & 7 \\ 8 & 9 & 4 \end{bmatrix}$$

Cofactor expansion of a Matrix

Definition

Assume that determinants of $(n-1)\times(n-1)$ matrices have been defined. If $A=[a_{ij}]$ is $n\times n$ define

$$\det A = a_{11}c_{11}(A) + a_{12}c_{12}(A) + \cdots + a_{1n}c_{1n}(A)$$

This is called the **cofactor expansion of** $\det A$ **along row 1** .

Theorem

The determinant of an $n\times n$ matrix A can be computed by using the cofactor expansion along any row or column of A. That is $\det A$ *can be computed by multiplying each entry of the row or column by the corresponding cofactor and adding the results.*

Example

Compute the determinant of $A=\begin{bmatrix} 3 & 4 & 5 \\ 1 & 7 & 2 \\ 9 & 8 & -6 \end{bmatrix}$.

Properties of determinant

Theorem

Let A denote an $n \times n$ matrix.

1. *If A has a row or column of zeros,* $\det A = 0$.
2. *If two distinct rows (or columns) of A are interchanged, the determinant of the resulting matrix is* $-\det A$.
3. *If a row (or column) of A is multiplied by a constant u, the determinant of the resulting matrix is* $u(\det A)$.
4. *If two distinct rows (or columns) of A are identical,* $\det A = 0$.
5. *If a multiple of one row of A is added to a different row (or if a multiple of a column is added to a different column), the determinant of the resulting matrix is* $\det A$.

Determinants and Matrix Inverses

Theorem

1. *If A and B are $n \times n$ matrices, then* $\det(AB) = \det A \det B$.
2. *An $n \times n$ matrix A is invertible if and only if* $\det A \neq 0$. *When this is the case,* $\det\left(A^{-1}\right) = \frac{1}{\det A}$
3. *If A is any square matrix,* $\det A^T = \det A$.

Example

If A and B are $n \times n$ matrices, then $\det A = 2$ and $\det B = 5$, calculate $\det\left(A^3 B^{-1} A^T B^2\right)$.

Solution. We use several of the facts just derived.

$$\begin{aligned}
\det\left(A^3 B^{-1} A^T B^2\right) &= \det\left(A^3\right) \det\left(B^{-1}\right) \det\left(A^T\right) \det\left(B^2\right) \\
&= (\det A)^3 \frac{1}{\det B} \det A (\det B)^2 \\
&= 2^3 \cdot \frac{1}{5} \cdot 2 \cdot 5^2 = 80
\end{aligned}$$

Adjugates

Definition

The **adjugate** of A, denoted adj (A), is the transpose of this cofactor matrix; in symbols, $\text{adj}(A) = [c_{ij}(A)]^T$.

Theorem

If A is any square matrix, then

$$A(\text{adj}\, A) = (\det A)I = (\text{adj}\, A)A$$

In particular, if $\det A \neq 0$, the inverse of A is given by

$$A^{-1} = \frac{1}{\det A}\,\text{adj}\, A$$

Example

Find the (2,3)-entry of A^{-1} if $A = \begin{bmatrix} 2 & 1 & 3 \\ 5 & -7 & 1 \\ 3 & 0 & -6 \end{bmatrix}$.

Solution. First compute

$$\det A = \begin{vmatrix} 2 & 1 & 3 \\ 5 & -7 & 1 \\ 3 & 0 & -6 \end{vmatrix} = \begin{vmatrix} 2 & 1 & 7 \\ 5 & -7 & 11 \\ 3 & 0 & 0 \end{vmatrix} = 3 \begin{vmatrix} 1 & 7 \\ -7 & 11 \end{vmatrix} = 180$$

Since $A^{-1} = \frac{1}{\det A} \operatorname{adj} A = \frac{1}{180} \left[c_{ij}(A)\right]^T$, the $(2,3)$-entry of A^{-1} is the (3,2)-entry of the matrix $\frac{1}{180}\left[c_{ij}(A)\right]$; that is, it equals $\frac{1}{180} c_{32}(A) = \frac{1}{180} \left(- \begin{vmatrix} 2 & 3 \\ 5 & 1 \end{vmatrix} \right) = \frac{13}{180}$.

Exercises

Write python programs that fulfill the following requirements:

(a) Sum two matrices, multiply a matrix by a number, multiply two matrices?

(b) Calculate the determinant of a matrix, calculate the inverse of a matrix?

Note: Each request is made in two ways:

- ⋆ code directly according to definitions, theorems, constructions,...
- ⋆ uses libraries from python.

Content

Content

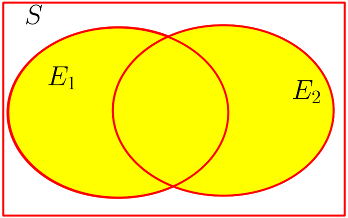
Sample Spaces and Events

Definition

- An experiment that can result in different outcomes, even though it is repeated in the same manner every time, is called a **random experiment**.
- The set of all possible outcomes of a random experiment is called the **sample space** of the experiment. The sample space is denoted as S.
  - A sample space is **discrete** if it consists of a finite or countable infinite set of outcomes.
  - A sample space is **continuous** if it contains an interval (either finite or infinite) of real numbers.
- An **event**, denoted by E, is a subset of the sample space of a random experiment.

Sample Spaces and Events

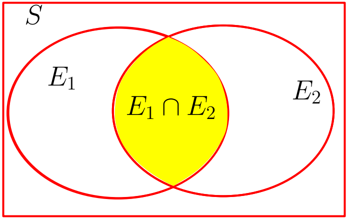
The **union** of two events is the event that consists of all outcomes that are contained in either of the two events. We denote the union as $E_1 \cup E_2$ or $E_1 + E_2$.



$E_1 \cup E_2$

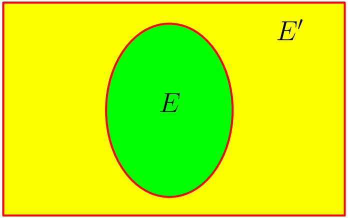
Sample Spaces and Events

The **intersection** of two events is the event that consists of all outcomes that are contained in both of the two events. We denote the intersection as $E_1 \cap E_2$ or $E_1 E_2$.



Sample Spaces and Events

The **complement** of an event in a sample space is the set of outcomes in the sample space that are not in the event. We denote the complement of the event E as E'.

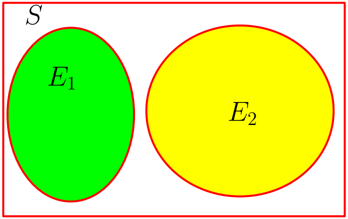


Sample Spaces and Events

Two events, denoted as E_1 and E_2, such that

$$E_1 \cap E_2 = \emptyset.$$

are said to be **mutually exclusive**.



$E_1 \cup E_2 = \emptyset$

Content

Addition Rules

* Any two events E_1, E_2

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

* If A and B are mutually exclusive events,

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

* A collection of events, $E_1, E_2, \ldots, E_k$, is said to be mutually exclusive if for all pairs,

$$E_i \cap E_j = \varnothing$$

For a collection of mutually exclusive events,

$$P(E_1 \cup E_2 \cup \ldots \cup E_k) = P(E_1) + P(E_2) + \ldots P(E_k)$$

Addition Rules

* Any three events E_1, E_2, E_3

$$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) \\ - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3)$$

* Any events $E_1, E_2, \ldots, E_n$

$$P(E_1 + E_2 + \ldots + E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i<j} P(E_i E_j) \quad (1) \\ + \sum_{i<j<k} P(E_i E_j E_k) - \ldots + (-1)^{n-1} P(E_1 E_2 \ldots E_n).$$

Addition Rules

Example

If A, B, and C are mutually exclusive events with $P(A) = 0.2, P(B) = 0.3$, and $P(C) = 0.4$, determine the following probabilities:
(a) $P(A \cup B \cup C)$
(b) $P(A \cap B \cap C)$
(c) $P(A \cap B)$
(d) $P[(A \cup B) \cap C]$
(e) $P(A' \cap B' \cap C')$

Addition Rules

Example

Strands of copper wire from a manufacturer are analyzed for strength and conductivity. The results from 100 strands are as follows:

	high strength	low strength
high conductivity	74	8
low conductivity	15	3

(a) If a strand is randomly selected, what is the probability that its conductivity is high and its strength is high?
(b) If a strand is randomly selected, what is the probability that its conductivity is low or the strength is low?
(c) Consider the event that a strand has low conductivity and the event that the strand has a low strength. Are these two events mutually exclusive?

Conditional Probability

The probability of an event B under the knowledge that the outcome will be in event A is denoted as $P(B|A)$ and this is called the **conditional probability** of B given A, and is calculated by

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \; P(A) > 0.$$

Similar,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \; P(B) > 0.$$

Conditional Probability

Example

Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

	shock resistance	
	high	low
high scratch	70	9
low scratch	16	5

Let A denote the event that a disk has high shock resistance, and let B denote the event that a disk has high scratch resistance. Determine the following probabilities:

(a) $P(A)$
(b) $P(B)$
(c) $P(A \mid B)$
(d) $P(B \mid A)$

Multiplication and Total Probability Rules

* Multiplication rule for any two events A and B:

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

* Total probability rule for any two events A and B:

$$P(B) = P(B \cap A) + P\left(B \cap A'\right) = P(B \mid A)P(A) + P\left(B \mid A'\right)P\left(A'\right)$$

* Assume $E_1, E_2, \ldots, E_k$ are k mutually exclusive and exhaustive sets. Then

$$\begin{aligned} P(B) &= P\left(B \cap E_1\right) + P\left(B \cap E_2\right) + \cdots + P\left(B \cap E_k\right) \\ &= P\left(B \mid E_1\right) P\left(E_1\right) + P\left(B \mid E_2\right) P\left(E_2\right) + \cdots \\ &\quad + P\left(B \mid E_k\right) P\left(E_k\right) \end{aligned}$$

Multiplication and Total Probability Rules

Example

- Suppose that $P(A \mid B) = 0.4$ and $P(B) = 0.5$. Determine the following:
  (a) $P(A \cap B)$ (b) $P(A' \cap B)$
- In the 2004 presidential election, exit polls from the critical state of Ohio provided the following results:

total	Bush, 2004	Kerry, 2004
no college degree (62%)	50%	50%
college graduate (38%)	53%	46%

  What is the probability a randomly selected respondent voted for Bush?

Content

Independence

* Two events are independent if any one of the following equivalent statements is true:
  * $P(A \mid B) = P(A)$
  * $P(B \mid A) = P(B)$
  * $P(A \cap B) = P(A)P(B)$
* The events $E_1, E_2, \ldots, E_n$ are independent if and only if for any subset of these events $E_{i_1}, E_{i_2}, \ldots, E_{i_k}$,

$$P\left(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}\right) = P\left(E_{i_1}\right) \times P\left(E_{i_2}\right) \times \cdots \times P\left(E_{i_k}\right)$$

Independence

Example

* If $P(A \mid B) = 0.3, P(B) = 0.8$, and $P(A) = 0.3$, are the events B and the complement of A independent?
* If $P(A) = 0.2, P(B) = 0.2$, and A and B are mutually exclusive, are they independent?
* A batch of 500 containers for frozen orange juice contains five that are defective. Two are selected, at random, without replacement, from the batch. Let A and B denote the events that the first and second containers selected are defective, respectively.
  (a) Are A and B independent events?
  (b) If the sampling were done with replacement, would A and B be independent?

Bayes' Theorem

* If A, B are any two events:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \text{ for } P(B) > 0$$

* If $E_1, E_2, \ldots, E_k$ are k mutually exclusive and exhaustive events and B is any event,

$$P\left(E_1 \mid B\right) = \frac{P\left(B \mid E_1\right) P\left(E_1\right)}{P\left(B \mid E_1\right) P\left(E_1\right) + P\left(B \mid E_2\right) P\left(E_2\right) + \cdots + P\left(B \mid E_k\right) P\left(E_k\right)}$$

for $P(B) > 0$

Bayes' Theorem

Example

* Suppose that $P(A \mid B) = 0.7, \quad P(A) = 0.5$, and $P(B) = 0.2$. Determine $P(B \mid A)$.
* Suppose that $P(A \mid B) = 0.4, \quad P(A \mid B') = 0.2$, and $P(B) = 0.8$. Determine $P(B \mid A)$.
* An inspector working for a manufacturing company has a 99% chance of correctly identifying defective items and a 0.5% chance of incorrectly classifying a good item as defective. The company has evidence that its line produces 0.9% of nonconforming items.
  (a) What is the probability that an item selected for inspection is classified as defective?
  (b) If an item selected at random is classified as nondefective, what is the probability that it is indeed good?

Content

2.8 Random Variables

Definition

- ⋆ A **random variable** is a function that assigns a real number to each outcome in the sample space of a random experiment.
- ⋆ A **discrete random variable** is a random variable with a finite (or countably infinite) range.
- ⋆ A **continuous random variable** is a random variable with an interval (either finite or infinite) of real numbers for its range.

Random Variables

Example

Decide whether a discrete or continuous random variable is the best model for each of the following variables:

(a) The time until a projectile returns to earth.

(b) The number of times a transistor in a computer memory changes state in one operation.

(c) The volume of gasoline that is lost to evaporation during the filling of a gas tank.

(d) The outside diameter of a machined shaft.

(e) The number of cracks exceeding one-half inch in 10 miles of an interstate highway.

(f) The weight of an injection-molded plastic part.

(g) The number of molecules in a sample of gas.

(h) The concentration of output from a reactor.

Probability Distributions and Probability Density Functions

* For a continuous random variable X, a **probability density function** is a function such that
  (1) $f(x) \geq 0$
  (2) $\int_{-\infty}^{\infty} f(x)dx = 1$
  (3) $P(a \leq X \leq b) = \int_a^b f(x)dx =$ area under $f(x)$ from a to b for any a and b
* The **cumulative distribution function** of a continuous random variable X is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du \text{ for } -\infty < x < \infty.$$

Example

Suppose the cumulative distribution function of X is

$$F(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ 0.2x & \text{if} \quad 0 \leq x < 5 \\ 1 & \text{if} \quad 5 \leq x \end{cases}$$

Determine the following:

(a) $P(X < 2.8)$ (b) $P(X > 1.5)$

(c) $P(X < -2)$ (d) $P(X > 6)$

Mean and Variance

X is a continuous random variable with probability density function $f(x)$.

- The **mean** or **expected value** of X, denoted as μ or $E(X)$, is

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

- The **variance** of X, denoted as $V(X)$ or σ^2, is

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

- The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$.

Mean and Variance

Example

- Suppose $f(x) = 1.5x^2$ for $-1 < x < 1$. Determine the mean and variance of X.
- Suppose that $f(x) = x/8$ for $3 < x < 5$. Determine the mean and variance of x.

Content

Continuous Uniform Distribution

* A **continuous random variable** X with probability density function

$$f(x) = 1/(b-a), \quad a \leq x \leq b$$

is a continuous uniform random variable.

* If X is a continuous uniform random variable over $a \leq x \leq b$,

$$\mu = E(X) = \frac{(a+b)}{2} \text{ and } \sigma^2 = V(X) = \frac{(b-a)^2}{12}$$

* The cumulative distribution function of a continuous uniform random variable is

$$F(x) = \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \leq x < b \\ 1 & b \leq x \end{cases}$$

Continuous Uniform Distribution

Example

Suppose the time it takes a data collection operator to fill out an electronic form for a database is uniformly between 1.5 and 2.2 minutes.
(a) What is the mean and variance of the time it takes an operator to fill out the form?
(b) What is the probability that it will take less than two minutes to fill out the form?
(c) Determine the cumulative distribution function of the time it takes to fill out the form.

Content

Normal Distribution

A random variable X with probability density function

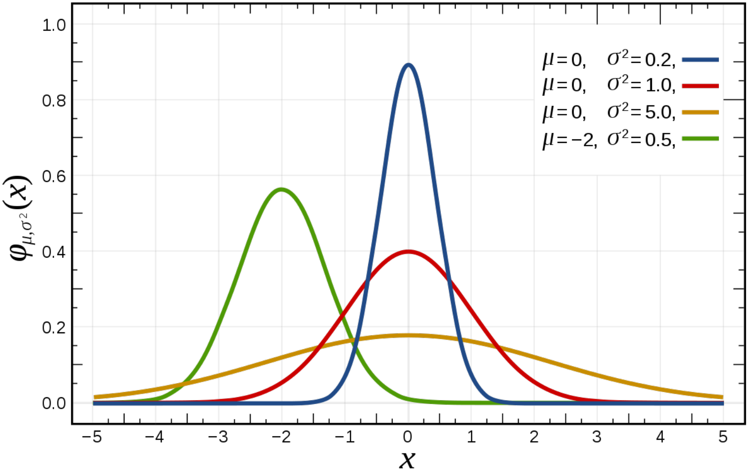
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty$$

is a normal random variable with parameters μ, where $-\infty < \mu < \infty$, and $\sigma > 0$. Also,

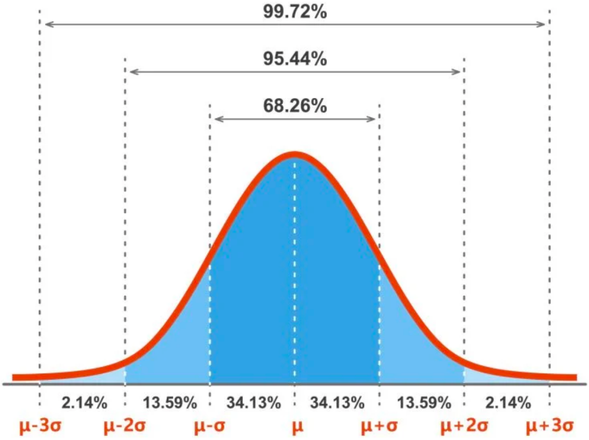
$$E(X) = \mu \text{ and } V(X) = \sigma^2$$

and the notation $N\left(\mu, \sigma^2\right)$ is used to denote the distribution.

Normal Distribution



Normal Distribution



Standard Normal Random Variable

* A normal random variable with

$$\mu = 0 \quad \text{and} \quad \sigma^2 = 1$$

is called a standard normal random variable and is denoted as Z. The cumulative distribution function of Z is denoted as

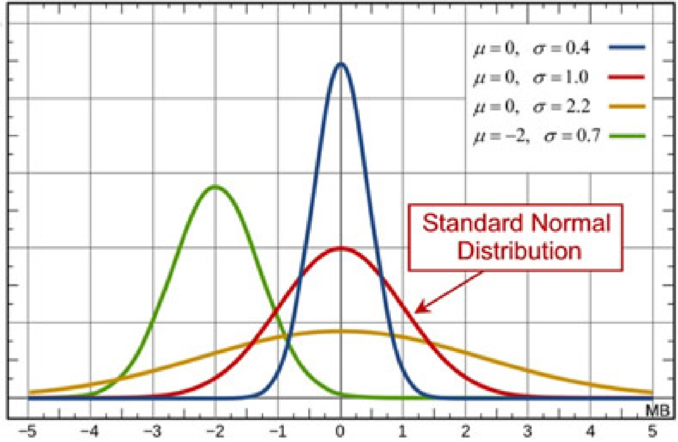
$$\Phi(z) = P(Z \le z)$$

* If X is a normal random variable with $E(X) = \mu$ and $V(X) = \sigma^2$, the random variable

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

is a standard normal random variable.

Standard Normal Random Variable



Standardizing to Calculate a Probability

Suppose X is a normal random variable with mean μ and variance σ^2. Then,

$$P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P(Z \leq z)$$

where Z is a standard normal random variable, and $z = \frac{(x-\mu)}{\sigma}$ is the z **-value** obtained by standardizing X.

Normal Distribution

Example

The compressive strength of samples of cement can be modeled by a normal distribution with a mean of 6000 kilograms per square centimeter and a standard deviation of 100 kilograms per square centimeter.

(a) What is the probability that a sample's strength is less than 6250 $\mathrm{Kg/cm^2}$?

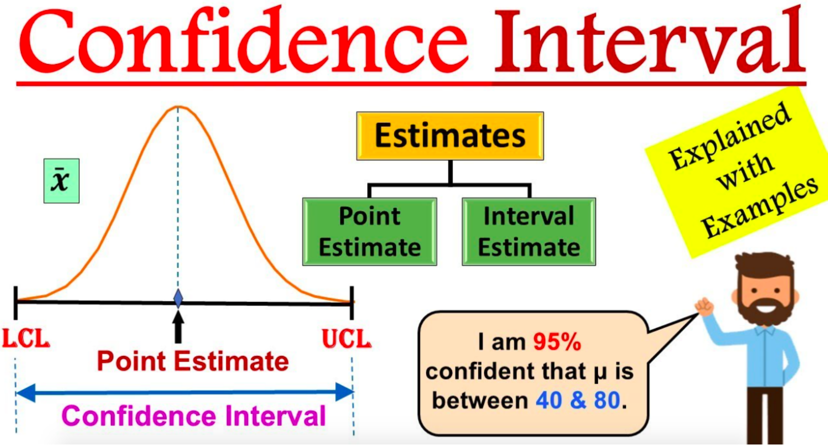
(b) What is the probability that a sample's strength is between 5800 and 5900 $\mathrm{Kg/cm^2}$?

(c) What strength is exceeded by 95% of the samples?

Content

Content

Confidence Interval



Confidence Interval

- A **confidence interval** (CI) estimate for μ is an interval of the form $l \leq \mu \leq u$, where the endpoints l and u are computed from the sample data.
- Suppose that we can determine values of L and U such that

$$P\{L \leq \mu \leq U\} = 1 - \alpha,$$

where $0 \leq \alpha \leq 1$. There is a probability of $1 - \alpha$ of selecting a sample for which the CI will contain the true value of μ.

- Once we have selected the sample, so that $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, and computed l and u, the resulting confidence interval for μ is

$$l \leq \mu \leq u.$$

CI on the Mean, Variance Known

Theorem

If $\bar{x}$ is the sample mean of a random sample of size n from a normal population with known variance σ^2, a $100(1-\alpha)\%$ CI on μ is given by

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \qquad (2)$$

where $z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution.

Note that $\Phi(z_\alpha) = 1 - \alpha$.

CI on the Mean, Variance Known

Example

In a sample of 36 randomly selected women, it was found that their mean height was 65.3 inches. From previous studies, it is assumed that the standard deviation, σ, is 2.5. Construct the 90% confidence interval for the population mean. Let $z_{0.1} = 1.28$ and $z_{0.05} = 1.64$.

A Large-Sample CI for μ

When n is large (should be greater than 30), the quantity $(\overline{X} - \mu)/(S/\sqrt{n})$ has an approximate standard normal distribution. Consequently,

$$\overline{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}$$

is a **large sample confidence interval** for μ, with confidence level of approximately $100(1-\alpha)\%$.

A Large-Sample CI for μ

Example

A civil engineer is analyzing the compressive strength of concrete. Compressive strength is normally distributed. A random sample of 49 specimens has a mean compressive strength of 3250(psi) and a variance 25(psi^2). Construct approximately a 95% two-sided confidence interval on mean compressive strength? Let $z_{0.05} = 1.64$ and $z_{0.025} = 1.96$

CI on the Mean of a Normal Distribution, Variance Unknown

The t distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with **unknown mean** μ and **unknown variance** σ^2. The random variable

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has a t **distribution with** $n - 1$ **degrees of freedom**.

CI on the Mean of a Normal Distribution, Variance Unknown

Theorem

If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal distribution with unknown variance σ^2, a $100(1-\alpha)\%$ confidence interval on μ is

$$\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

CI on the Mean of a Normal Distribution, Variance Unknown

Example

The brightness of a television picture tube can be evaluated by measuring the amount of current required to achieve a particular brightness level. A sample of 12 tubes results in $\bar{x} = 316.8$ and $s = 15.2$. Find (in microamps) a 99% confidence interval on mean current required. Let
$t_{0.01,12} = 2.681$, $t_{0.005,12} = 3.055$, $t_{0.01,11} = 2.718$, $t_{0.005,11} = 3.106$.

CI for a Population Proportion

Theorem

If $\hat{p}$ is the proportion of observations in a random sample of size n that belongs to a class of interest, an approximate $100(1-\alpha)\%$ confidence interval on the proportion p of the population that belongs to this class is

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Theorem (choice of sample size)

The *required sample size* *that the error estimating* $|\hat{P} - p|$ *not exceed* E *is*

$$n = \left\lceil \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p) \right\rceil \quad (3)$$

At (3): If a previous estimate $\hat{p}$ *is known, change* $p(1-p)$ *by* $\hat{p}(1-\hat{p})$*.*
Else, since $\max p(1-p) = 0.25$*, we use* $n = \lceil (z_{\alpha/2}/E)^2 * 0.25 \rceil$*.*

Example

Of 1000 randomly selected cases of lung cancer, 750 resulted in death within 10 years.

(a) Calculate a 95% two-sided confidence interval on the death rate from lung cancer.

(b) what sample size is needed to be 95% confident that the error in estimating the true value of p is less than 0.04? Let $z_{0.025} = 1.96$.

How to Calculate Confidence Intervals in Python?

- With the given data, determine whether to use the normal distribution function (norm.interval() function) or the t distribution (t.interval() function).
- Determine sample size and reliability.
- Use scipy.stats library in Python.

Example

Estimate with a 95% confidence interval the average height of a population of trees based on an observed sample including the heights of the following 22 trees:

$$11, 12, 12.5, 13, 13, 15.5, 16, 17, 22, 23, 25, 26, 27, 28, 28, 29, 30, 32,$$

$$33, 33, 34, 34.$$

How to Calculate Confidence Intervals in Python?

Example

Estimate with a 95% confidence interval the average height of a population of trees based on an observed sample including the heights of the following 22 trees:

$$11, 12, 12.5, 13, 13, 15.5, 16, 17, 22, 23, 25, 26, 27, 28, 28, 29, 30, 32,$$

$$33, 33, 34, 34.$$

```
import numpy as np
import scipy.stats as st

#define sample data
data = [11, 12, 12.5, 13, 13, 15.5, 16, 17, 22, 23, 25, 26, 27, 28, 28, 29,
        30, 32, 33, 33, 34, 34]

#create 95% confidence interval for population mean weight
a = st.t.interval(alpha=0.95, df=len(data)-1,
                  loc=np.mean(data), scale=st.sem(data))
```

How to Calculate Confidence Intervals in Python?

Example

Estimate with a 95% confidence interval the average height of a population of trees based on an observed sample including the heights of the following 33 trees:

$$11, 11, 11.5, 11.5, 11.5, 12, 12, 12, 12.5, 13, 13, 15.5, 16, 17, 22,$$

$$23, 25, 26, 27, 28, 28, 29, 30, 32, 33, 33, 34, 34, 35, 35, 35.5, 35.5, 36.$$

```
import numpy as np
import scipy.stats as st


#define sample data
data = [11, 11, 11.5, 11.5, 11.5, 12, 12, 12, 12.5, 13, 13, 15.5, 16, 17, 22,
    23, 25, 26, 27, 28, 28, 29, 30, 32, 33, 33, 34, 34, 35, 35, 35.5, 35.5, 36]

#create 95% confidence interval for population mean weight
a = st.norm.interval(alpha=0.95, loc=np.mean(data), scale=st.sem(data
```

Content

Statistical Hypothesis

- A **statistical hypothesis** is a statement about the parameters of one or more populations.
- For a statistical hypothesis: **null hypothesis** H_0 and **alternative hypothesis** H_1.

Example

A cereal company claims that the mean weight of the cereal in its packets is 14.2 oz.

Null hypothesis H_0: $\mu = 14.2$
Alternative hypothesis H_1: $\mu \neq 14.2$.
This is a two-sided alternative hypothesis.

Tests of Statistical Hypotheses

We have 2 forms on the alternative hypothesis:

(1) **two-sided alternative hypothesis**

$$H_0 : \mu = \mu_0 \quad H_1 : \mu \neq \mu_0$$

(2) **one-sided alternative hypothesis**

$$H_0 : \mu = \mu_0 \qquad H_1 : \mu > \mu_0 \text{ or}$$
$$H_0 : \mu = \mu_0 \qquad H_1 : \mu < \mu_0.$$

Example

The owner of a football team claims that the average attendance at games is over 45,000. Express the null hypothesis and the alternative hypothesis in symbolic form.

$$H_0 : \mu = 45000 \qquad H_1 : \mu > 45000.$$

Tests of Statistical Hypotheses

Definition

- The **critical region** for the test: ($\bar{x} < 48.5$ or $\bar{x} > 51.5$) is the region that rejects H_0.
- The **acceptance region** for the test: ($48.5 \leq \bar{x} \leq 51.5$) is the region that fails to reject H_0.
- The **critical values**: (48.5 and 51.5) are the boundaries between the critical regions and the acceptance region.

Tests of Statistical Hypotheses

Test of a Hypothesis

- ⋆ Rejecting the null hypothesis H_0 when it is true is defined as a **type I error**.
- ⋆ Failing to reject the null hypothesis when it is false is defined as a **type II error**.

Probability of Type I Error

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true}).$$

Sometimes the type I error probability is called the **significance level**, or the α-**error,** or the **size** of the test.

Tests of Statistical Hypotheses

Probability of Type II Error

$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false})$$

P-value

The P-value is the smallest level of significance that would lead to rejection of the null hypothesis H_0 with the given data.

Tests on the Mean of a Normal Distribution, Variance known

Null hypothesis : $H_0 : \mu = \mu_0$
Test statistic value: $z_0 = (\bar{x} - \mu_0)/(\sigma/\sqrt{n})$.

Alternative hypothesis	P-value	Critical values	Reject H_0
$H_1 : \mu \neq \mu_0$	$2[1 - \Phi(z_0)]$	$z_{\alpha/2}, -z_{\alpha/2}$	$z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$
$H_1 : \mu > \mu_0$	$1 - \Phi(z_0)$	z_α	$z_0 > z_\alpha$
$H_1 : \mu < \mu_0$	$\Phi(z_0)$	$-z_\alpha$	$z_0 < -z_\alpha$

Example

Suppose you want to test the claim that $\mu \neq 8$. Given a sample size of $n = 82$ and a level of significance of $\alpha = 0.02$. When should you reject H_0?

Tests on the Mean of a Normal Distribution, Variance Unknown

Null hypothesis: $H_0 : \mu = \mu_0$
Test statistic value: $t_0 = (\bar{x} - \mu_0)/(s/\sqrt{n})$.

Alternative hypothesis	P-value	Critical values	Reject H_0
$H_1 : \mu \neq \mu_0$	$2P(T_{n-1} > \lvert t_0 \rvert)$	$t_{\alpha/2,n-1}, -t_{\alpha/2,n-1}$	$t_0 > t_{\alpha/2,n-1}$ or $t_0 < -t_{\alpha/2,n-1}$
$H_1 : \mu > \mu_0$	$P(T_{n-1} > t_0)$	$t_{\alpha,n-1}$	$t_0 > t_{\alpha,n-1}$
$H_1 : \mu < \mu_0$	$P(T_{n-1} < -t_0)$	$-t_{\alpha,n-1}$	$t_0 < -t_{\alpha,n-1}$

Example

Given a sample with $n = 10, \bar{x} = 7.9, s = 1.2$ and alternative hypothesis $H1 : \mu \neq 8.2$ and the level of significance $\alpha = 0.05$. Should we reject H_0?

Tests on a Population Proportion

Let X be the number of observations in a random sample of size n that belongs to the class associated with p. Then, if the null hypothesis $H_0 : p = p_0$ is true, we have $X \approx N[np_0, np_0(1 - p_0)]$.

- ⋆ Null hypothesis : $H_0 : p = p_0$
- ⋆ Test statistic :

$$Z_0 = \frac{X - np_0}{\sqrt{np_0(1 - p_0)}}$$

Alternative hypothesis	P-value	Critical values	Reject H_0
$H_1 : p \neq p_0$	$2[1 - \Phi(z_0)]$	$z_{\alpha/2}, -z_{\alpha/2}$	$z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$
$H_1 : p > p_0$	$1 - \Phi(z_0)$	z_α	$z_0 > z_\alpha$
$H_1 : p < p_0$	$\Phi(z_0)$	$-z_\alpha$	$z_0 < -z_\alpha$

Example

A random sample of 200 circuits generated 9 defectives. Use the data to test $H_0 : p = 0.05$ versus $H_1 : p \neq 0.05$. Use $\alpha = 0.05$. Let $z_{0.05} = 1.65, z_{0.025} = 1.96$.

(a) Find the critical values for this test.

(b) Should we reject H_0?

Exercises

- Find the library to solve statistical hypothesis testing problems in Python?
- Find data and set headings for data and solve using Python program with the following 3 contents:

1. Standardize found data.
2. Estimate confidence interval for the population mean.
3. Statistical hypothesis testing for the population mean.