Linear Algebra and Probability Theory
Math Camp Day 2

Kiyotaka Yageta [1]

Graduate School of Economics, University of Tokyo

March 30, 2019

[1] Thanks Hikaru Kawarazaki for reffering to your slide. If you find any mistake, please let me know.

Outline

Outline

Introduction

This lecture covers basic linear algebra and probability theory that you need in studying economics, especially

- Econometrics I & II
- Mathematics I

Although I skip many slides because of time constraint, all slides are important enough to spend your time (hopefully).

Today's Goal

- This course is so basic that you do not need to attend the class if you have confidence in your mathematical skill.
- The main focus is not on understanding rigorous proofs of propositions nor on understanding topics in measure theory behind probability theory.
- I omitted many proofs, but proving every proposition is very good practice and will definitely help you understand the materials. So do it by yourself!

Outline

Notation

- $\mathbb{R}$ represents the set of real numbers.
- $\mathbb{C}$ represents the set of complex numbers.
- $\mathbb{K}$ represents $\mathbb{R}$ or $\mathbb{C}$
- $\mathbb{N}$ represents the set of natural numbers.
- Uppercase letters such as A, B denote a matrix.
- Bold lowercase letters such as $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{x}$ denote a column vector.
- A row vector $(a_1, \cdots, a_n)$ is represented by a transpose of a column vector $\boldsymbol{a}$, i.e., $\boldsymbol{a}'$.

Outline

Vector and Matrix

Definition (Matrix)

Let m, n be a natural number. We call A as an $m \times n$ matrix if $a_{ij} \in \mathbb{K}$ are put as follows:

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

$A = (a_{ij} | i = 1, 2, ..., m;\ j = 1, 2, ..., n)$, or $A = (a_{ij})$

If $n = 1$ $(m = 1)$, we call A a column vector[2] (row vector)

[2] It is called an m-dimension column vector or an $m \times 1$ column vector

Transpose

Definition (Transpose)

The transpose of a matrix, denoted as A' or $A^\top$, is the matrix which is obtained by creating the matrix whose kth row is the kth column of the original matrix A:

$$A' = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix}, \text{ where } A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

Transpose

The following holds:

$$\boldsymbol{a_j} \equiv \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} \text{ and } \boldsymbol{b_i'} \equiv \begin{pmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{pmatrix},$$

then

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = \begin{pmatrix} \boldsymbol{a_1} & \boldsymbol{a_2} & \dots & \boldsymbol{a_n} \end{pmatrix} = \begin{pmatrix} \boldsymbol{b_1'} \\ \boldsymbol{b_2'} \\ \vdots \\ \boldsymbol{b_m'} \end{pmatrix}$$

Zero Matrix

Definition (Zero Matrix)

An $m \times n$ zero matrix, usually denoted by $O_{m,n}$, is a matrix in which all components are zero[a]

$$O_{m,n} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

[a]If $m = n$, this is denoted as O_n, usually not $O_{n,n}$

Symmetric Matrix

Definition (Symmetric Matrix)

An $n \times n$ symmetric matrix (a_{ij}) is a square matrix in which $a_{ij} = a_{ji}$ for all i and j.

Definition (Square Matrix)

A square matrix is a matrix with the same number of rows and columns

Example

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 4 \\ 2 & 3 & 6 \\ 4 & 6 & 5 \end{pmatrix}$$

Diagonal Matrix

Definition (Diagonal Matrix)

An $n \times n$ diagonal matrix (a_{ij}) is a square matrix where all the off-diagonal elements are zero, i.e., $\forall j \neq i, a_{ij} = 0$:

$$\begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & d_n \end{pmatrix}, \text{ where } \forall i \; d_i \in \mathbb{K}$$

Unit Vector

Definition(Unit Vector)

An $n \times 1$ jth unit vector, often denoted by $\mathbf{e_j}$, is a vector whose jth element is one and others are zero.

$$\mathbf{e_j} = \begin{pmatrix} 0 & \cdots & 0 & \underbrace{1}_{j\text{th element}} & 0 & \cdots & 0 \end{pmatrix}'$$

Example

Let $\boldsymbol{a'} = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}$, then

$$\boldsymbol{a} = a_1\boldsymbol{e_1} + a_2\boldsymbol{e_2} + \cdots + a_n\boldsymbol{e_n}$$

We will discuss it later. (p.18)

Identity Matrix

Definition (Identity Matrix)

An $n \times n$ identity matrix, often denoted by I or I_n, is a diagonal matrix in which all diagonal components are one.

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{e_1} & \mathbf{e_2} & \cdots & \mathbf{e_n} \end{pmatrix} = \begin{pmatrix} \mathbf{e_1}' \\ \mathbf{e_2}' \\ \vdots \\ \mathbf{e_n}' \end{pmatrix}$$

Submatrix

Definition (Submatrix)

A submatrix of a matrix A is a matrix obtained from A by removing any number of rows or columns from A.

Formally, for a submatrix of $m \times n$ matrix $A = (a_{ij} | i = 1, 2, ..., m;\ j = 1, 2, ..., n)$ is obtained by picking r rows $i_1 < i_2 < ... < i_r$ and s columns $j_1 < j_2 < ... < j_s$ and rearranging them into an $r \times s$ matrix $(a_{i_p j_q} | p = 1, 2, ..., r;\ q = 1, 2, ..., s)$.

Particularly, for a principle submatrix of an $n \times n$ matrix is a submatrix with $r = s$.

In addition to $r = s$, if $i_1 = 1, i_2 = 2, ..., i_r = r$, then it is called a leading principal matrix.

Submatrix

- Example of a submatrix
  - Picking the first and third rows and the second, third, and fourth columns (removing the second row and the first coulumn)

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{pmatrix} \rightarrow \begin{pmatrix} a_{12} & a_{13} & a_{14} \\ a_{32} & a_{33} & a_{34} \end{pmatrix}$$

- Example of a leading principle submatrix
  - Picking the first and second rows and the first and second columns (removing the third row and the third coulumn from a square matrix)

$$B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \rightarrow \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

Triangular Matrix

Definition (Triangular Matrix)

An $n \times n$ matrix is lower triangular matrix if all the components above the main diagonal are zero, i.e., $a_{ij} = 0 \, \forall i, j$ s.t. $i < j$. Similarly, an $n \times n$ matrix is upper triangular matrix if all the entries below the main diagonal are zero, i.e., $a_{ij} = 0 \, \forall i, j$ s.t. $i > j$.

$$\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \qquad \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}$$

The left matrix is a lower triangular matrix and the right matrix is an upper triangular matrix.

Basic Calculation

Definition (Addition of Matrices)

For two $m \times n$ matrices $A = (a_{ij})$ and $B = (b_{ij})$, the sum of A and B, denoted by $A + B$, is a matrix whose (i, j) element is $a_{ij} + b_{ij}$.

Slimilarly, the difference of A and B, denoted by $A - B$, is a matrix whose (i, j) element is $a_{ij} - b_{ij}$.

Definition (Multiplication of a Matrix by a Scalar)

For a $m \times n$ matrix $A = (a_{ij})$ and a scalar c, cA is a matrix whose (i, j) element is $c\, a_{ij}$.

Basic Calculation

For any $m \times n$ matrices A, B, and C, $m \times n$ zero matrix O, and scalars c and d, the followings hold.

1. $(A+B)+C = A+(B+C)$
2. $A+B = B+A$
3. $A+O = A$
4. $A-A = O$
5. $c(A+B) = cA+cB$
6. $(c+d)A = cA+dA$
7. $(cd)A = c(dA)$
8. $1A = A$
9. $0A = O$
10. $cO = O$

Basic Calculation

Definition (Multiplication of Matrices)

For an $l \times m$ matrix $A = (a_{ij})$ and an $m \times n$ matrix $B = (b_{jk})$, the product of A and B, denoted by AB, is an $l \times n$ matrix whose (i,k) element is $\sum_{j=1}^{m} a_{ij} b_{jk}$.

For example, if you multiply 1×2 matrix by 2×3 matrix, then you have a 1×3 matrix as follows:

$$\begin{aligned} & \begin{pmatrix} a_{11} & a_{12} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{pmatrix} \\ = & \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} & a_{11}b_{13} + a_{12}b_{23} \end{pmatrix} \end{aligned}$$

Basic Calculation

- You can define $AB \not\Rightarrow$ You can define BA.
- Suppose that A is $m_A \times n_A$ and that B is $m_B \times n_B$
  - $n_A = m_B \Rightarrow$ You can define AB
  - $m_A = n_B \Rightarrow$ You can define BA
- In the previous example, you can define only AB
- Even if you can define both AB and BA, $AB \neq BA$ in general
  - Suppose that A is 1×2 and that B is 2×1
  - AB is 1×1 and that BA is 2×2
- Even if both are the same shape, they are different as below

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$
$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

Basic Calculation

For an $l \times m$ matrix A, an $m \times n$ matrix B, and an $n \times p$ matrix C, and scalars c and d, the followings hold.

1. $(AB)C = A(BC)$
2. $AI_l = I_m A = A$
3. $AO_{m,q} = O_{l,q},\ O_{q,l}A = O_{q,m}$ for any $q \in \mathbb{N}$
4. $cAB = (cA)B = A(cB)$
5. $A(B+C) = AB + BC$ if C is $m \times n$
6. $(A+B)C = AC + BC$ if B is $l \times m$, C is $m \times n$
7. $(AB)' = B'A'$

Exercise

Compute these products.

1. $$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

2. $$\begin{pmatrix} 4 & 0 & -1 \\ 1 & 2 & -1 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ 3 & 4 \\ -1 & 0 \end{pmatrix}$$

3. $$\begin{pmatrix} -2 & 3 & -1 \\ 1 & 4 & 0 \end{pmatrix} \begin{pmatrix} 4 & 1 \\ 0 & 2 \\ -1 & -1 \end{pmatrix}$$

Answer of Exercise

Compute these products.

1. $$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

2. $$\begin{pmatrix} 4 & 0 & -1 \\ 1 & 2 & -1 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ 3 & 4 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} -7 & 4 \\ 5 & 9 \end{pmatrix}$$

3. $$\begin{pmatrix} -2 & 3 & -1 \\ 1 & 4 & 0 \end{pmatrix} \begin{pmatrix} 4 & 1 \\ 0 & 2 \\ -1 & -1 \end{pmatrix} = \begin{pmatrix} -7 & 5 \\ 4 & 9 \end{pmatrix}$$

Useful Notations for Summation

- Denote an $n \times 1$ column vector $(1, \cdots, 1)'$ by $\boldsymbol{\iota}$.
- For $\boldsymbol{x} = (x_1, \cdots, x_n)'$, $\sum_{i=1}^{n} x_i = \boldsymbol{\iota}'\boldsymbol{x}$
- The sample mean of the elements of $\boldsymbol{x}$, $\bar{x} \equiv \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\boldsymbol{\iota}'\boldsymbol{x}$.
- A squared sum $\sum_{i=1}^{n} x_i^2$ can be written as $\sum_{i=1}^{n} x_i^2 = \boldsymbol{x}'\boldsymbol{x}$.

Idempotent Matrix

Definition (Idempotent Matrix)

An idempotent matrix M is one that is equal to its square, i.e., $MM = M$.

Exercise

Express $\sum_{i=1}^{n}(x_i - \bar{x})^2$ by vectors and matrices.

Idempotent Matrix

1. $$\begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix} = \boldsymbol{x} - \bar{x}\boldsymbol{\iota} = \boldsymbol{x} - \boldsymbol{\iota}\bar{x} = \boldsymbol{x} - \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}'\boldsymbol{x}$$

2. Because $\boldsymbol{x} = \boldsymbol{I}\boldsymbol{x}$, we have

$$\left(\boldsymbol{x} - \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}'\boldsymbol{x}\right) = \left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}'\right)\boldsymbol{x} = M^0\boldsymbol{x}$$

3. We obtain $\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2 = (M^0\boldsymbol{x})'M^0\boldsymbol{x} = \boldsymbol{x}'M^{0\prime}M^0\boldsymbol{x}$

Idempotent Matrix

- M^0 is an $n \times n$ symmetric matrix $(M^0 = M^{0\prime})$
  - diagonal elements are all $1 - \frac{1}{n}$
  - off-diagonal elements are all $-\frac{1}{n}$
- WTS: $M^0 M^0 = M^0$
  - On diagonal:
$$\left(1 - \frac{1}{n}\right)^2 + (n-1)\frac{1}{n^2} = 1 - \frac{1}{n}$$
  - Off diagonal:
$$2\left\{-\frac{1}{n} \times \left(1 - \frac{1}{n}\right)\right\} + (n-2)\frac{1}{n^2} = -\frac{1}{n}$$
- M is an idempotent matrix $\Rightarrow \sum_{i=1}^{n}(x_i - \bar{x})^2 = \boldsymbol{x}' M^0 \boldsymbol{x}$
- $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = (M^0\boldsymbol{x})'(M^0\boldsymbol{y}) = \boldsymbol{x}' M^0 \boldsymbol{y}$

Vector Space

Definition (Vector Space)

A vector space is a set of vectors that is closed under scalar multiplication and addition.

Example

If U is a vector space, then

1. $\forall \boldsymbol{a}, \boldsymbol{b} \in U,\ \boldsymbol{a} + \boldsymbol{b} \in U$
2. $\forall \boldsymbol{a} \in U\ \forall c \in \mathbb{K},\ c\boldsymbol{a} \in U$

$\Rightarrow\ \forall \boldsymbol{a}, \boldsymbol{b} \in U,\ \forall k, h \in \mathbb{K},\ k\boldsymbol{a} + h\boldsymbol{b} \in U$

Throughout this lecture, let U denote a vector space.

Vector Space

Definition (Linear Independence)

For $\boldsymbol{a_1}, \boldsymbol{a_2}, \ldots, \boldsymbol{a_n} \in \mathbb{K}^n$, consider a linear combination

$$c_1\boldsymbol{a_1} + \cdots + c_n\boldsymbol{a_n} = \mathbf{0}, \quad c_1, \ldots, c_n \in \mathbb{K}$$

- A set of vectors $\boldsymbol{a_1}, \cdots, \boldsymbol{a_n}$ is called linearly independent if $c_1 = \cdots = c_n = 0$.
- A set of vectors $\mathbf{a_1}, \ldots, \mathbf{a_n}$ is linearly dependent if at least one of $c_1, \cdots, c_n$ is not zero.

We often write

$$\{c_1\boldsymbol{a_1} + \cdots + c_n\boldsymbol{a_n} \mid c_1, \ldots, c_n \in \mathbb{K}\} = \langle \boldsymbol{a_1}, \cdots, \boldsymbol{a_n} \rangle$$

Subspace

Definition (Spanned Space)

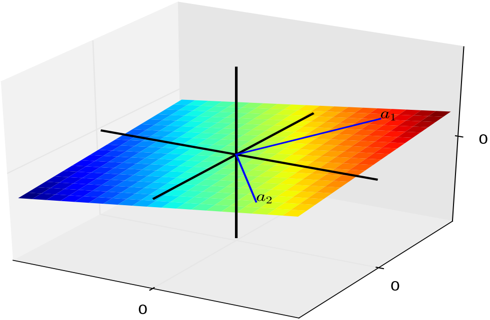
The set of all linear combinations of a set of vectors is the vector space that is spanned by those vectors

For instance, $\mathbb{R}^3$ is spanned by $[1,0,0],[0,1,0],[0,0,1]$.
What is the space spanned by $[1,0,0],[0,1,0]$? That is not $\mathbb{R}^2$ but a plane in $\mathbb{R}^3$, which is called a *subspace* in $\mathbb{R}^3$.

Definition (Subspace)

If $V \subset U$ is closed under scalar multiplication and addition, then V is called a subspace of U.

Subspace



This is a plane spanned by two vectors $\mathbf{a_1}, \mathbf{a_2}$ in $\mathbb{R}^3$.

Basis Vector

Definition (Basis Vectors)

A set of vectors $\boldsymbol{u_1}, \cdots, \boldsymbol{u_n} \in U$ is a basis for U if the following conditions are satisfied:

1. $\boldsymbol{u_1}, \cdots, \boldsymbol{u_n}$ are linearly independent.
2. $\forall \boldsymbol{u} \in U,\ \boldsymbol{u} \in \langle \boldsymbol{u_1}, \cdots, \boldsymbol{u_n} \rangle$

Definition (Dimension)

The number of basis vectors of U is called the dimension of U.

- The choice of basis vectors of a vector space is not unique.
  - Both pairs $[1, 0], [0, 1]$ and $[1, 1], [0, 1]$ are a basis of $\mathbb{R}^2$.
- The dimension of a vector space is unique. (If $U = \{\mathbf{0}\}$, the dimension of U is zero)

Rank

Note that matrix is viewed as a set of column vectors or row vectors. This leads us to the following concepts.

Definition (Column space)

The column space of a matrix is the vector space that is spanned by its column vectors.

Definition (Column rank)

The column rank of a matrix is the dimension of the vector space that is spanned by its column vectors.

- The column rank of a matrix is equal to the largest number of linearly independent column vectors of it.
- The row space and the row rank are defined similarly.

Outline

Permutation

Definition (Permutation)

An n-th permutation, denoted as σ, is an act of rearranging members of a set into another order, i.e., bijective mapping of a set $\{1, 2, ..., n\}$ to a set $\{1, 2, ..., n\}$. Especially, if you change the order of two elements and leave the others unchanged, i.e., for some $i \neq j$, $\sigma(i) = j$, $\sigma(j) = i$, and $\forall k \in \{1, \cdots, n\} \backslash \{i, j\}$ $\sigma(k) = k$, the the act is called a transposition.

I denote the permutation with $\sigma(1) = i_1$, $\sigma(2) = i_2$, ..., $\sigma(n) = i_n$ as

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix},$$

where the numbers in the first row indicates the order (1st number, 2nd number etc.).

Permutation

All permutation can be made by repeating transpositions. How many transpositions we need to make a permutation is not unique, but whether the number is even or odd is determined regardless of its order.

Definition (Even/Odd Permutation)

If the number of transpositions is even, the permutations is called an even permutation, and if it is odd, it is called an odd permutation.

The sign of a permutation σ denoted as

$$\text{sgn}(\sigma) = \begin{cases} +1 & \text{if the permutations is even} \\ -1 & \text{if the permutations is odd} \end{cases}$$

Determinant of Matrix

Definition (Determinant)

The determinant of an $n \times n$ matrix $A = (a_{ij} \mid i, j = 1, 2, ..., n)$, denoted as $\det(A)$ ($|A|,\ D(A)$) is as follows:

$$\det(A) = |A| = \sum_{\sigma} \mathrm{sgn}(\sigma) \cdot \prod_{i=1}^{n} a_{\sigma(i)i},$$

where $\sum_{\sigma}$ is a summation over all n-th permutation σ's.

The above definition is used in almost all textbooks about basic linear algebra but it disgusts us. (I believe that many people agree with my idea)
So, I introduce the derivation of the above definition.

Determinant of Matrix

Definition (Determinant)

The determinant of an $n \times n$ matrix $A = (a_{ij} \mid i, j = 1, 2, ..., n)$, denoted as $\det(A)$ ($|A|,\ D(A)$) is satisfied the following three properties

(a) $D(\boldsymbol{a_1}, \cdots, k\boldsymbol{a_j^1} + h\boldsymbol{a_j^2}, \cdots, \boldsymbol{a_n}) =$
$kD(\boldsymbol{a_1}, \cdots, \boldsymbol{a_j^1}, \cdots, \boldsymbol{a_n}) + hD(\boldsymbol{a_1}, \cdots, \boldsymbol{a_j^2}, \cdots, \boldsymbol{a_n})$

(b) $\boldsymbol{a_i} = \boldsymbol{a_j} \Rightarrow D(A) = 0$

(c) $D(E) = D(\boldsymbol{e_1}, \cdots, \boldsymbol{e_n}) = 1$

Determinant of Matrix

Property (Determinant)

From the definition of determinant, we can derive the following three properties

(d) $D(\boldsymbol{a_1}, \cdots, \boldsymbol{a_i}, \cdots, \boldsymbol{a_j}, \cdots, \boldsymbol{a_n}) = -D(\boldsymbol{a_1}, \cdots, \boldsymbol{a_j}, \cdots, \boldsymbol{a_i}, \cdots, \boldsymbol{a_n})$

(e) $\boldsymbol{a_1}, \cdots, \boldsymbol{a_n}$ is linearly dependent $\Rightarrow D(A) = 0$

(f) $D(\boldsymbol{a_1}, \cdots, \boldsymbol{a_i}, \cdots, \boldsymbol{a_j}, \cdots, \boldsymbol{a_n}) = D(\boldsymbol{a_1}, \cdots, \boldsymbol{a_i}, \cdots, \boldsymbol{a_j} + c\boldsymbol{a_i}, \cdots, \boldsymbol{a_n})$

You can easily show (d)-(f) by using (a)-(c) (please try by yourself!)

Determinant of Matrix

Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} = \begin{pmatrix} \boldsymbol{a_1} & \boldsymbol{a_2} & \dots & \boldsymbol{a_n} \end{pmatrix}$$

Here,

$$\boldsymbol{a_1} = a_{11}\boldsymbol{e_1} + a_{21}\boldsymbol{e_2} + \cdots + a_{n1}\boldsymbol{e_n}$$

Then,

$$\begin{aligned} D(A) &= D(\boldsymbol{a_1}, \cdots, \boldsymbol{a_n}) \\ &= D(a_{11}\boldsymbol{e_1} + a_{21}\boldsymbol{e_2} + \cdots + a_{n1}\boldsymbol{e_n}, \boldsymbol{a_2}, \cdots, \boldsymbol{a_n}) \\ &= a_{11}D(\boldsymbol{e_1}, \boldsymbol{a_2}, \cdots, \boldsymbol{a_n}) + \cdots + a_{n1}D(\boldsymbol{e_n}, \boldsymbol{a_2}, \cdots, \boldsymbol{a_n}) \quad (\because (a)) \\ &= \sum_{i=1}^{n} a_{i1} D(\boldsymbol{e_i}, \boldsymbol{a_2}, \cdots, \boldsymbol{a_n}) \end{aligned}$$

Determinant of Matrix

Applying the same procedure to $\boldsymbol{a_2}$,

$$D(A) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i1} a_{j2} D(\boldsymbol{e_i}, \boldsymbol{a_j}, \cdots, \boldsymbol{a_n})$$

Iterating the above process to $\boldsymbol{a_3}, \cdots, \boldsymbol{a_n}$,

$$D(A) = \sum_{i_1, \cdots, i_n}^{n} a_{i_1 1} \cdots a_{i_n n} D(\boldsymbol{e_{i_1}}, \cdots, \boldsymbol{e_{i_n}})$$

From (b), (c), and (d), we can derive

$$D(A) = \sum_{\sigma} \mathsf{sgn}(\sigma) \cdot \prod_{i=1}^{n} a_{\sigma(i)i}$$

Determinant of Matrix

- For 1×1 matrix (scalar), the determinant is

$$|a| = a$$

- For 2×2 matrix, the computation of the determinant is

$$\begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - bc.$$

- For 3×3 matrix, the computation of the determinant is

$$\begin{aligned} &\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\ = \;& a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ -\;& a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}. \end{aligned}$$

Determinant of Matrix

Formula (Determinant of Matrix)

Calculating determinant of matrix, we frequently use the following formula.

$$|A| = \sum_{j=1}^{n} a_{ji}(-1)^{i+j} \left| A_{ji} \right|, \quad i = 1, \ldots, n,$$

where i can be arbitrarily chosen and A_{ji} is the submatrix which is obtain by removing the jth row and the ith column.

The above formula is also famous but counterintuitive.
So, I introduce the derivation of the above formula using the definition (a)-(c).

Determinant of Matrix

Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} = \begin{pmatrix} \boldsymbol{a_1} & \boldsymbol{a_2} & \dots & \boldsymbol{a_n} \end{pmatrix}$$

Here,

$$\boldsymbol{a_i} = a_{1i}\boldsymbol{e_1} + a_{2i}\boldsymbol{e_2} + \cdots + a_{ni}\boldsymbol{e_n}$$

Then,

$$\begin{aligned} D(A) &= D(\boldsymbol{a_1}, \cdots, \boldsymbol{a_n}) \\ &= D(\boldsymbol{a_1}, \cdots, a_{1i}\boldsymbol{e_1} + a_{2i}\boldsymbol{e_2} + \cdots + a_{ni}\boldsymbol{e_n}, \cdots, \boldsymbol{a_n}) \\ &= a_{1i}D(\boldsymbol{a_1}, \cdots, \boldsymbol{e_1}, \cdots, \boldsymbol{a_n}) + \cdots + a_{ni}D(\boldsymbol{a_n}, \cdots, \boldsymbol{e_n}, \cdots, \boldsymbol{a_n}) \\ &= \sum_{j=1}^{n} a_{ji}D(\boldsymbol{a_1}, \cdots, \boldsymbol{e_j}, \cdots, \boldsymbol{a_n}) \end{aligned}$$

Determinant of Matrix

Let

$$B = \begin{pmatrix} a_{11} & \dots & a_{1,i-1} & 0 & a_{1,i+1} & \dots & a_{1n} \\ \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ a_{j1} & \dots & a_{j,i-1} & 1 & a_{j,i+1} & \dots & a_{jn} \\ \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ a_{n1} & \dots & a_{n,i-1} & 0 & a_{n,i+1} & \dots & a_{nn} \end{pmatrix}$$

Here, applying to (d),

$$D(B) = (-1)^{i-1} D \begin{pmatrix} 0 & a_{11} & \dots & a_{1,i-1} & a_{1,i+1} & \dots & a_{1n} \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 1 & a_{j1} & \dots & a_{j,i-1} & a_{j,i+1} & \dots & a_{jn} \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 0 & a_{n1} & \dots & a_{n,i-1} & a_{n,i+1} & \dots & a_{nn} \end{pmatrix}$$

Determinant of Matrix

$$
(-1)^{i-1} D \begin{pmatrix}
0 & a_{11} & \dots & a_{1,i-1} & a_{1,i+1} & \dots & a_{1n} \\
\vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\
1 & a_{j1} & \dots & a_{j,i-1} & a_{j,i+1} & \dots & a_{jn} \\
\vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\
0 & a_{n1} & \dots & a_{n,i-1} & a_{n,i+1} & \dots & a_{nn}
\end{pmatrix}
$$

$$
= (-1)^{i+j-2} D \begin{pmatrix}
1 & a_{j1} & \dots & a_{j,i-1} & a_{j,i+1} & \dots & a_{jn} \\
0 & a_{11} & \dots & a_{1,i-1} & a_{1,i+1} & \dots & a_{1n} \\
\vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\
0 & a_{j-1,1} & \dots & a_{j-1,i-1} & a_{j-1,i+1} & \dots & a_{j-1,n} \\
0 & a_{j+1,1} & \dots & a_{j+1,i-1} & a_{j+1,i+1} & \dots & a_{j+1,n} \\
\vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\
0 & a_{n1} & \dots & a_{n,i-1} & a_{n,i+1} & \dots & a_{nn}
\end{pmatrix}
$$

$$
\equiv (-1)^{i+j} D(B^*)
$$

Determinant of Matrix

WTS:

$$D(B^*) = D\begin{pmatrix} 1 & a_{j1} & \dots & a_{j,i-1} & a_{j,i+1} & \dots & a_{jn} \\ 0 & a_{11} & \dots & a_{1,i-1} & a_{1,i+1} & \dots & a_{1n} \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 0 & a_{j-1,1} & \dots & a_{j-1,i-1} & a_{j-1,i+1} & \dots & a_{j-1,n} \\ 0 & a_{j+1,1} & \dots & a_{j+1,i-1} & a_{j+1,i+1} & \dots & a_{j+1,n} \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 0 & a_{n1} & \dots & a_{n,i-1} & a_{n,i+1} & \dots & a_{nn} \end{pmatrix}$$

$$= D\begin{pmatrix} a_{11} & \dots & a_{1,i-1} & a_{1,i+1} & \dots & a_{1n} \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ a_{j-1,1} & \dots & a_{j-1,i-1} & a_{j-1,i+1} & \dots & a_{j-1,n} \\ a_{j+1,1} & \dots & a_{j+1,i-1} & a_{j+1,i+1} & \dots & a_{j+1,n} \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ a_{n1} & \dots & a_{n,i-1} & a_{n,i+1} & \dots & a_{nn} \end{pmatrix} = D(A_{ji})$$

Determinant of Matrix

Proof:

Let

$$B^* = \begin{pmatrix} \boldsymbol{e_1} & \boldsymbol{a_1^*} & \dots & \boldsymbol{a_{i-1}^*} & \boldsymbol{a_{i+1}^*} & \dots & \boldsymbol{a_n} \end{pmatrix}$$

Here, from the definition of determinant

$$\begin{aligned} D(B^*) &= D\left(\boldsymbol{e_1}, \boldsymbol{a_1^*}, \dots, \boldsymbol{a_{i-1}^*}, \boldsymbol{a_{i+1}^*}, \dots, \boldsymbol{a_n}\right) \\ &= \sum_{i_1^*, \cdots, i_n^*}^{n} a_{i_1^* 1} \cdots a_{i_n^* n} D(\boldsymbol{e_1}, \boldsymbol{e_{i_1^*}} \cdots, \boldsymbol{e_{i_n^*}}) \\ &= \sum_{\sigma^*} \mathsf{sgn}(\sigma^*) \cdot \prod_{i^*}^{n} a_{\sigma^*(i^*) i^*} = D(A_{ji}) \\ &= D(A_{ji}) \end{aligned}$$

$$\therefore\ D(A) = \sum_{j=1}^{n} a_{ji} (-1)^{i+j} D(A_{ji}) \quad \forall i = 1, \cdots, n$$

Determinant of Matrix

Formula (Determinant of Matrix)

Calculating determinant of matrix, we frequently use the following formula.

$$|A| = \sum_{j=1}^{n} a_{ji}(-1)^{i+j} \left|A_{ji}\right|, \quad i = 1, \ldots, n$$

where i can be arbitrarily chosen and A_{ji} is the submatrix which is obtain by removing the jth row and the ith column.

As an exercise, let's compute the determinant of

$$A = \begin{pmatrix} 5 & 6 & 0 \\ -1 & 0 & 0 \\ 1 & 2 & 2 \end{pmatrix}$$

Determinant of Matrix

I choose $i = 2$.

$$
\begin{aligned}
|A| &= -1 \times (-1)^{2+1}|A_{21}| \\
&\quad +0 \times (-1)^{2+2}|A_{22}| \\
&\quad +0 \times (-1)^{2+3}|A_{23}| \\
&= |A_{21}| \\
&= \begin{vmatrix} 6 & 0 \\ 2 & 2 \end{vmatrix} \\
&= 12.
\end{aligned}
$$

As a sanity check, let's compute the determinant of the same matrix but choose $i = 1$ or $i = 3$ (both is also OK).

Exercise

As an exercise, let's compute the determinants of

1. $$A = \begin{pmatrix} 2 & 3 \\ 5 & -2 \end{pmatrix}$$

2. $$B = \begin{pmatrix} 5 & 6 & 0 \\ -1 & 0 & 0 \\ 1 & 2 & 2 \end{pmatrix}$$

3. $$C = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Answer of Exercise

1. $$A = \begin{pmatrix} 2 & 3 \\ 5 & -2 \end{pmatrix} = -19$$

2. $$B = \begin{pmatrix} 5 & 6 & 0 \\ -1 & 0 & 0 \\ 1 & 2 & 2 \end{pmatrix} = 12$$

3. $$C = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{pmatrix} = 24$$

Determinant of Matrix

The following link (written in Japanese) is very useful for checking properties of determinant.

https://risalc.info/src/determinant-formulas.html

I pick up some important theorems and properties. Try to show them! From now on, let

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = \begin{pmatrix} \boldsymbol{a_1} & \dots & \boldsymbol{a_n} \end{pmatrix} = (a_{ij})$$

and $c \in \mathbb{K}$

Determinant of Matrix

Theorems & Property

1. Let $A^* = (\boldsymbol{a_1}, \dots, \boldsymbol{a_{i-1}}, c\boldsymbol{a_i}, \boldsymbol{a_{i+1}}, \dots, \boldsymbol{a_n})$, then $|A^*| = c|A|$
2. $|cA| = c^n|A|$
3. $|A'| = |A|$
4. If $\boldsymbol{a_1} = a_{11}\boldsymbol{e_1}$, then $|A| = a_{11}|A_{11}|$
5. Let B be a $n \times n$ matrix, then $|AB| = |A||B|$
6. $\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{vmatrix} = a_{11} \cdots a_{nn}$ [a]

[a] Determinant of lower triangle matrix is of course equal

System of Linear Equations

You will face systems of equations many times in Economics.

$$
\begin{aligned}
c_{11}b_1 + \cdots + c_{1k}b_k &= d_1 \\
c_{21}b_1 + \cdots + c_{2k}b_k &= d_2 \\
&\vdots \\
c_{k1}b_1 + \cdots + c_{kk}b_k &= d_k.
\end{aligned}
$$

To calculate it easily, you need to learn "inverse matrix."

Inverse Matrix

Definition (Regular Matrix)

An $n \times n$ matrix A is regular (non-singular, invertible) if there exists an $n \times n$ matrix B which satisfies $AB = BA = I_n$. B is called the inverse matrix of A and we write $B \equiv A^{-1}$.

Examples

We can convert the following system of equations into matrix

$$\begin{cases} a_{11}x_{11} + a_{12}x_{21} = 1 \\ a_{11}x_{12} + a_{12}x_{22} = 0 \\ a_{21}x_{11} + a_{22}x_{21} = 0 \\ a_{21}x_{12} + a_{22}x_{22} = 1 \end{cases} \Rightarrow \underbrace{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}}_{\equiv A} \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Inverse Matrix

Assume $a_{11}a_{22} - a_{12}a_{21} \neq 0$. (If $a_{11}a_{22} - a_{12}a_{21} \neq 0$, there is no solution)
If you solve the simultaneous equations with respect to x_{ij},

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$
$$= \frac{1}{|\mathbf{A}|} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

Formula (Inverse Matrix)

Calculating the inverse matrix of A, we frequently use the following formula

$$A^{-1} = \frac{1}{|A|}\tilde{A},$$

where $\tilde{A}$ is the cofactor matirx of A, i.e., $\tilde{A} = ((-1)^{i+j}|A_{ji}|)$

You can show it by calculating $\tilde{A}A$ and using property 4 of p.53

Inverse Matrix

Theorem (Inverse Matrix)

Let A be an $n \times n$ matrix. The following statements are equivalent.

1. A is non-singular
2. $\text{rank}(A) = n$
3. All column vectors of A are linearly independent
4. All row vectors of A are linearly independent
5. $\det(A) \neq 0$
6. The solution of the system of equations $A\boldsymbol{x} = \boldsymbol{b}$ is unique.
7. The unique solution of the system of equations $A\boldsymbol{x} = \mathbf{0}$ is $\boldsymbol{x} = \mathbf{0}$

Inverse Matrix

Property (Inverse Matrix and Non-singular Matrix)

Suppose A, B, C are non-singular and their dimensions are the same

1. $|A^{-1}| = \frac{1}{|A|}$
2. $(A^{-1})^{-1} = A$
3. $(A^{-1})' = (A')^{-1}$
4. $(AB)^{-1} = B^{-1}A^{-1}$
5. $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$
6. All non-singular matrices have only one inverse matrix
7. $\text{rank}(AB) = \text{rank}(A)$
8. $\text{rank}(CA) = \text{rank}(A)$

Outline

Eigenvalue and Eigenvector

Definition (Eigenvalue and Eigenvector)

For an $n \times n$ matrix A, if nonzero vector $\boldsymbol{x} \in \mathbb{K}^n$ and $\lambda \in \mathbb{K}$ satisfy

$$A\boldsymbol{x} = \lambda \boldsymbol{x},$$

then λ is called an eigenvalue (characteristic root) and $\boldsymbol{x}$ is called an eigenvector (characteristic vector)

- $\boldsymbol{x} = \mathbf{0}$ is not eigenvector
- λ is obtained by solving the characteristic equation $|A - \lambda I_n| = 0$ (You can check it!)
- If $|A - \lambda I_n| \neq 0 \;\Rightarrow (A - \lambda I_n)$ is non-singular $\Rightarrow \boldsymbol{x} = \mathbf{0}$
- If $\boldsymbol{x}$ is an eigenvector, $c\boldsymbol{x}$ $(c \in \mathbb{K} \backslash \{0\})$ is also an eigenvector

Eigenvalue and Eigenvector

Exercise

Let's compute eigenvalues and eigenvectors of the following matrix

$$A = \begin{pmatrix} 8 & 1 \\ 4 & 5 \end{pmatrix}$$

Eigenvalues and Eigenvectors

The characteristic equation is

$$\begin{vmatrix} 8-\lambda & 1 \\ 4 & 5-\lambda \end{vmatrix} = 0$$
$$\Leftrightarrow (8-\lambda)(5-\lambda) - 4 = 0$$
$$\Leftrightarrow (\lambda-4)(\lambda-9) = 0 \quad \therefore \lambda = 4, 9$$

Let

$$\Lambda = \begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix}$$

Eigenvalue and Eigenvector

(i) $\lambda = 4$ case

The eigenvectors $\boldsymbol{x}_{\lambda=4} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ satisfies

$$\begin{pmatrix} 8-4 & 1 \\ 4 & 5-4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 4 & 1 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

This implies $4x_1 + x_2 = 0$. Hence, we obtain

$$\boldsymbol{x}_{\lambda=4} = c_1 \begin{pmatrix} 1 \\ -4 \end{pmatrix}, \text{ where } c_1 \in \mathbb{K} \backslash \{0\}$$

We can arbitrarily choose c_1 without changing the result, so let $c_1 = 1$

Eigenvalue and Eigenvector

(ii) $\lambda = 9$ case

The eigenvectors $\boldsymbol{y}_{\lambda=9} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ satisfies

$$\begin{pmatrix} 8-9 & 1 \\ 4 & 5-9 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} -1 & 1 \\ 4 & -4 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

This implies $-y_1 + y_2 = 0$. Hence, we obtain

$$\boldsymbol{y}_{\lambda=9} = c_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \text{ where } c_2 \in \mathbb{K} \backslash \{0\}$$

We can arbitrarily choose c_2 without changing the result, so let $c_2 = 1$

Eigenvalue and Eigenvector

If we put the eigenvectors into a matrix P, we can obtain the diagonal matrix Λ which satisfies $A = P\Lambda P^{-1}$.

$$\begin{pmatrix} 8 & 1 \\ 4 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -4 & 1 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -4 & 1 \end{pmatrix}^{-1}$$

We call it diagonalization. Using it we can easily calculate A^n.

$$\begin{aligned} A^n &= (P\Lambda P^{-1})^n \\ &= P\Lambda P^{-1} P\Lambda P^{-1} P \cdots P^{-1} \\ &= P\Lambda^n P^{-1} \\ &= \begin{pmatrix} 1 & 1 \\ -4 & 1 \end{pmatrix} \begin{pmatrix} 4^n & 0 \\ 0 & 9^n \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -4 & 1 \end{pmatrix}^{-1} \end{aligned}$$

Exercise

As an exercise, let's diagonalize the following matrices!

1. $$A = \begin{pmatrix} -2 & 12 \\ -1 & 5 \end{pmatrix}$$

2. $$B = \begin{pmatrix} 5 & -1 & -2 \\ 0 & 2 & 0 \\ 6 & -2 & -2 \end{pmatrix}$$

3. $$C = \begin{pmatrix} 1 & 2 & -1 \\ 2 & -2 & 2 \\ -1 & 2 & 1 \end{pmatrix}$$

Answer of Exercise

1. $$A = \begin{pmatrix} 4 & 3 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 4 & 3 \\ 1 & 1 \end{pmatrix}^{-1}$$

2. $$B = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 3 & 0 \\ 2 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 3 & 0 \\ 2 & 0 & 3 \end{pmatrix}^{-1}$$

3. $$C = \begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{30}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{30}} & -\frac{2}{\sqrt{6}} \\ 0 & \frac{5}{\sqrt{30}} & \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -4 \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{30}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{30}} & -\frac{2}{\sqrt{6}} \\ 0 & \frac{5}{\sqrt{30}} & \frac{1}{\sqrt{6}} \end{pmatrix}$$

Diagonalization

Let A be an $n \times n$ matrix, then $|\lambda I_n - A| = 0$ has n eigenvalues
Since all eigenvalues have at least one eigenvectors, take any eigenvalue and eigenvector, λ_i and $\boldsymbol{p_i}$, we obtain

$$A\boldsymbol{p_i} = \lambda_i \boldsymbol{p_i} \Rightarrow (A\boldsymbol{p_1}, \cdots, A\boldsymbol{p_n}) = (\lambda \boldsymbol{p_1}, \cdots, \lambda \boldsymbol{p_n})$$

$$\Leftrightarrow A(\boldsymbol{p_1}, \cdots, \boldsymbol{p_n}) = (\boldsymbol{p_1}, \cdots, \boldsymbol{p_n}) \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

$$\Leftrightarrow AP = P\Lambda$$

Therefore, P is a regular matrix $\Rightarrow A$ is diagonalizable.

Diagonalization

Let A is $n \times n$ matrix.

Theorem (Diagonalization)

If all eigenvalues of A are different, eigenvectors of A are linearly independent.

From the theorems of p.57 and the fact of p.66, we get

Theorem (Diagonalization)

If all eigenvalues of A are different, A is diagonalizable.

Diagonalization

Even if $|\lambda I_n - A| = 0$ has multiple roots, we can diagonalize A.

Theorem (Diagonalization)

A is diagonalizable if and only if $\dim(V(\lambda_i)) = n_i \ \ \forall i = \{1, \cdots, k\}$, where $|\lambda I_n - A| = (\lambda - \lambda_1)^{n_1} \cdots (\lambda - \lambda_k)^{n_k}$ and $V(\lambda_i)$ is dimension of the space spanned by eigenvectors corresponding to λ_i

If A is diagonalizable, you obtain

$$\begin{aligned} \det(A) &= \det(P\Lambda P^{-1}) \\ &= \det(P)\det(\Lambda)\det(P^{-1}) \quad (\because \text{p.53}) \\ &= \det(\Lambda) \quad (\because \text{p.58}) \end{aligned}$$

Therefore,

$$|A| = \prod_{i=1}^{n} \lambda_i$$

Diagonalization

Here, we consider the case A is not diagonalizable.

Theorem (Schur triangulation)

All $n \times n$ matrix A is triangulable by using a regular matrix Q and an upper (lower) triangle matrix Λ^*

$$A = Q\Lambda^* Q^{-1}$$

Especially, Q is an orthogonal matrix[a].

[a] I omit unitary matrix

Definition (Orthogonal Matrix)

An $n \times n$ orthogonal matrix M is a square matrix where $MM' = I_n$

Diagonalization

Definition (Symmetric Matrix)

An $n \times n$ symmetric matrix[a] M is a square matrix where $M = M'$

[a] I omit Hermitian matrix

You will frequently encounter symmetric matrices in econometrics.

$$\begin{pmatrix} Var(x_1)^2 & Cov(x_1, x_2) & \dots & Cov(x_1, x_n) \\ Cov(x_2, x_1) & Var(x_2)^2 & \dots & Cov(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(x_n, x_1) & Cov(x_1, x_2) & \dots & Var(x_n)^2 \end{pmatrix}$$

Diagonalization

Fortunately, there is an important theorem about a symmetric matrix.

Theorem (Symmetric Matrix)

A symmetric matrix A is diagonalizable by using an orthogonal matrix Q and a diagonal matrix Λ

$$A = Q\Lambda Q^{-1}$$

and their eigenvalues must be real number

Outline

Trace

Definition (Trace)

The trace of an $n \times n$ matrix is the sum of its diagonal elements

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

Theorem (Trace)

The trace of a matrix equals the sum of its eigenvalues.

Trace

The followings hold with $n \times n$ matrices A, B, C, and D and a vector $\boldsymbol{v} \in \mathbb{K}^n$ where $A = \begin{pmatrix} \boldsymbol{a_1} & \cdots & \boldsymbol{a_n} \end{pmatrix}$

1. $\text{tr}(cA) = c\,\text{tr}(A)$
2. $\text{tr}(A') = \text{tr}(A)$
3. $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$
4. $\text{tr}(I_n) = n$
5. $\text{tr}(AB) = \text{tr}(BA)$
6. $\boldsymbol{v}'\boldsymbol{v} = \text{tr}(\boldsymbol{v}'\boldsymbol{v}) = \text{tr}(\boldsymbol{v}\boldsymbol{v}')$
7. $\text{tr}(A'A) = \sum_{k=i}^{n} \boldsymbol{a_i}'\boldsymbol{a_i} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2$
8. $\text{tr}(ABCD) = \text{tr}(BCDA) = \text{tr}(CDAB) = \text{tr}(DABC)$

Kronecker Product

Definition (Kronecker Product)

The Kronecker product of $m \times n$ matrix A and $k \times l$ matrix B is

$$A \otimes B \equiv \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix}$$

Property (Kronecker Product)

1. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
2. $(A \otimes B)(C \otimes D) = (AC \otimes BD)$
3. $(A \otimes B)' = A' \otimes B'$
4. $|A \otimes B| = |A|^n |B|^M \quad$ if A is $M \times M$ and B is $n \times n$

Quadratic Form

Definition (Quadratic Form)

A symmetric $n \times n$ matrix A is

- positive definite if $\boldsymbol{x}' A \boldsymbol{x} > 0$ for any $\boldsymbol{x} \in \mathbb{K}^n \backslash \{0\}$
- positive semi-definite if $\boldsymbol{x}' A \boldsymbol{x} \geq 0$ for any $\boldsymbol{x} \in \mathbb{K}^n$

Theorem (Positive definite)

The following statements are equivalent for symmetric matrix A

1. A is positive definite, i.e., $\boldsymbol{x}' A \boldsymbol{x} > 0$ for any $x \in \mathbb{K}^n \backslash \{0\}$
2. All of its eigenvalues are positive.
3. We can write $A = SS'$, where S is a lower (upper) triangle matrix.[a] (We call it " Cholesky decomposition ")

[a] S is uniquely decided and its diagonal elements are real numbers.

Quadratic Form

The following two theorems are frequently used in Econometrics. (Used when you change $(X'X)\hat{\boldsymbol{\beta}} = X'\boldsymbol{y}$ to $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y}$)

Theorem (Quadratic Form)

If A is $n \times k$ $(n > k)$ with full column rank, then $A'A$ is positive definite and AA' is positive semidefinite.

Theorem (Quadratic Form)

If a symmetric matrix A is positive definite, then A is non-singular and A^{-1} is also positive definite

Calculus and Matrix Algebra

We will consider $f : \mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{g} : \mathbb{R}^m \to \mathbb{R}^n$ which we denote

$$f(\boldsymbol{x}) = f(x_1, \cdots, x_n), \quad \boldsymbol{g}(\boldsymbol{x}) = \begin{pmatrix} g_1(x_1, \cdots, x_m) \\ \vdots \\ g_n(x_1, \cdots, x_m) \end{pmatrix}$$

Definition (Derivative of Multivariate Function)

The ith partial derivative of f is

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}) = \lim_{h \to 0} \frac{f(\boldsymbol{x} + h\boldsymbol{e}_i) - f(\boldsymbol{x})}{h}$$

where $\boldsymbol{e}_i = (0, \cdots, 0, \underbrace{1}_{i}, 0, \cdots, 0)$ *and* $h \in \mathbb{R}$.

Calculus and Matrix Algebra

Let $f : \mathbb{R}^n \to \mathbb{R}$ is sufficiently smooth and $y = f(\boldsymbol{x})$

Definition (Gradient)

The vector below is called gradient of f

$$\nabla f(\mathbf{x}) = \nabla_x f(\mathbf{x}) = Df(\mathbf{x}) \equiv \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial y}{\partial x_1} & \dots & \frac{\partial y}{\partial x_n} \end{pmatrix}'$$

Definition (Hessian)

The matrix below is called Hessian of f

$$H = \nabla^2 f(\boldsymbol{x}) = D^2 f(\boldsymbol{x}) \equiv \frac{\partial f(\mathbf{x})}{\partial \boldsymbol{x}' \partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial y}{\partial x_1 \partial x_1} & \cdots & \frac{\partial y}{\partial x_n \partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_1 \partial x_n} & \cdots & \frac{\partial y}{\partial x_n \partial x_n} \end{pmatrix}$$

Differentiation

You can show

$$\frac{\partial \boldsymbol{a}'\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{a},$$
$$\frac{\partial A\boldsymbol{x}}{\partial \boldsymbol{x}} = A',$$
$$\frac{\partial \boldsymbol{x}'A\boldsymbol{x}}{\partial \boldsymbol{x}} = (A + A')\boldsymbol{x},$$
$$\frac{\partial \boldsymbol{x}'A\boldsymbol{x}}{\partial \boldsymbol{x}\partial \boldsymbol{x}'} = A + A'.$$

In particular, if A is symmetric, then

$$\frac{\partial \boldsymbol{x}'A\boldsymbol{x}}{\partial \boldsymbol{x}} = 2A\boldsymbol{x},$$
$$\frac{\partial \boldsymbol{x}'A\boldsymbol{x}}{\partial \boldsymbol{x}\partial x'} = 2A.$$

OLS estimator

Let's solve

$$\min_{b}(\boldsymbol{y}-X\boldsymbol{b})'(\boldsymbol{y}-X\boldsymbol{b})=\boldsymbol{y}'\boldsymbol{y}-2\boldsymbol{y}'X\boldsymbol{b}+\boldsymbol{b}'X'X\boldsymbol{b}$$

The first order condition is

$$\mathbf{0}=-2X'\boldsymbol{y}+2X'X\boldsymbol{b}$$

Since $X'X$ is non-singular, the OLS estimator $\hat{\boldsymbol{\beta}}$ is

$$\hat{\boldsymbol{\beta}}=(X'X)^{-1}X'\boldsymbol{y}$$

Outline

Outline

Random Variable

Definition (Random Variable)

X is a random variable if its possible value are outcomes of a random phenomenon

- The concept of random variable is useful when we think about uncertainty or data generating process.
- X is a random variable because it is uncertain what value X will take until the data are actually observed as x.
- We usually use capital letters for the "name" of a random variable and lowercase letters for the "values" it takes. Thus, the probability that X takes a particular value x might be denoted $Pr(X = x)$.

PDF and CDF

Definition (Probability Density Function)

For discrete X, the probability density function (PDF) $f(x)$ is

$$f(x) = Pr(X = x)$$

For continuous X, the probability density function (PDF) $f(x)$ is

$$Pr(a \leq X \leq b) = \int_a^b f(x)dx \text{ with } f(x) \geq 0 \ \ \forall x \in [a, b] \text{ [a]}$$

[a] Since X is continuous and $f(x) = 0$, we define PDF as a range of X

PDF and CDF

Definition (Cumulative Distribution Function)

For discrete X, the cumulative distribution function (CDF) $F(x)$ is

$$F(x) = Pr(X \leq x) = \sum_{t \leq x} f(t) \text{ [a]}$$

For continuous X, the cumulative distribution function (CDF) $F(x)$ is

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

[a] For a discrete X, $f(x_i) = F(x_i) - F(x_{i-1})$

[b] For a continuous X, $f(x) = \frac{dF(x)}{dx}$

Joint Density Function

Definition (Joint Density Function)

The joint density function for X and Y, $f(x,y)$ is

$$Pr(a \leq X \leq b, c \leq Y \leq d) = \begin{cases} \displaystyle\sum_{a \leq x \leq b} \sum_{c \leq y \leq d} f(x,y) & \text{if } X \text{ and } X \text{ are \underline{discrete}} \\ \displaystyle\int_a^b \int_c^d f(x,y) dy dx & \text{if } X \text{ and } Y \text{ are \underline{continuous}} \end{cases}$$

Multivariate Distributions

Definition (Marginal density Function)

The marginal density function for X and Y, $f_Y(y)$ $(f_X(x))$ is

$$f_Y(y) = \begin{cases} \displaystyle\sum_{x \in \mathcal{X}} f(x, y) & \text{if } X \text{ and } Y \text{ are discrete;} \\ \displaystyle\int_{x \in \mathcal{X}} f(x, s) ds & \text{if } X \text{ and } Y \text{ are continuous.} \end{cases}$$

where $\mathcal{X}$ is a set of possible outcomes

Independence

Definition (Independence)

Two random variables are statistically independent if and only if their joint density is the product of the marginal densities, i.e.,

$$f(x,y) = f_X(x)f_Y(y) \;\Leftrightarrow\; X \text{ and } Y \text{ are independent}$$

Conditional Distribution

Definition (Conditional Distribution)

The conditional distribution over y for a value x, $f(y|x)$ is

$$f(y|x) = \frac{f(x,y)}{f_X(x)}$$

Expectation

Definition (Expectation)

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

$$E(X) = \begin{cases} \displaystyle\sum_{x \in \mathcal{X}} x f(x) & \text{if } X \text{ is \underline{discrete}} \\ \displaystyle\int_{x \in \mathcal{X}} x f(x) dx & \text{if } X \text{ is \underline{continuous}} \end{cases}$$

where $\mathcal{X}$ is a set of possible outcomes

Expectation

Definition (Expectation)

For any function $g(X, Y)$, its expected value is given as

$$E[g(X,Y)] = \begin{cases} \displaystyle\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} g(x,y)f(x,y) & \text{if } X \text{ and } Y \text{ are } \underline{\text{discrete}} \\ \displaystyle\int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} g(x,y)f(x,y)dxdy & \text{if } X \text{ and } Y \text{ are } \underline{\text{continuous}} \end{cases}$$

Variance

Definition (Variance)

The variance of X whose mean is μ, denoted as $Var(X)$ or σ_X^2, is

$$
\begin{aligned}
Var(X) &= E[(X-\mu)^2] \\
&= \begin{cases} \displaystyle\sum_{x \in \mathcal{X}} (x-\mu)^2 f(x) & \text{if } X \text{ is discrete;} \\ \displaystyle\int_{x \in \mathcal{X}} (x-\mu)^2 f(x) dx & \text{if } X \text{ is continuous,} \end{cases}
\end{aligned}
$$

where $\mathcal{X}$ is a set of possible outcomes.

Covariance

Definition (Covariance)

The covariance of X and Y, denoted as $Cov(X,Y)$ or σ_{XY} , is

$$Cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)] \ (= E[XY] - \mu_x\mu_y)$$

Let's try showing it!

- X and Y are independent $\Rightarrow Cov(X,Y) = 0$
- $Cov(X,Y) = 0 \not\Rightarrow X$ and Y are independent

Correlation Coefficient

Definition (Correlation Coefficient)

The correlation coefficient of X and Y, denoted as $Corr(X,Y)$, is

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} \ \left(= \frac{\sigma_{XY}}{\sigma_X \sigma_Y}\right)$$

where $\sigma_X = \sqrt{Var(X)}$ and the same for Y

Outline

k-th Dimensional Normal Distribution

Definition (k-th Dimensional Normal Distribution)

If $\boldsymbol{X}$ follows a k-th dimensional normal distribution whose mean and variance are $\boldsymbol{\mu}$ and Σ respectively,

$$\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$$

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left\{-\frac{(1}{2}(\boldsymbol{X}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right\}$$

where Let $\boldsymbol{X} = (X_1, \cdots, X_k)'$

$$\Sigma = \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_k) \\ Cov(X_2, X_1) & Var(X_2) & \cdots & Cov(X_2, X_k) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_k, X_1) & Cov(X_k, X_2) & \cdots & Var(X_k) \end{pmatrix}$$

Bernoulli Distribution

Definition (Bernoulli Distribution)

If X follows a Bernoulli distribution,[a]

$$X \sim Ber(p)$$
$$\mathcal{X} = \{0, 1\}$$
$$\Pr(X = 0) = 1 - p$$
$$E(X) = \Pr(X = 1) = p$$
$$Var(X) = p(1 - p)$$

[a] Bernoulli distribution is a discrete probability distribution which takes the value 1 with probability p and the value 0 with probability $1 - p$

Binomial Distribution

Definition (Binomial Distribution)

If X follows a binomial distribution, [a]

$$X \sim Bin(n, p)$$
$$\mathcal{X} = \{0, \cdots, n\}$$
$$\Pr(X = x) = {}_nC_x p^x (1-p)^{n-x}$$
$$E[X] = np$$
$$Var(X) = np(1-p)$$

[a] Binomial distribution is a discrete probability distribution of the number of successes in a sequence of n times Bernoulli trials

Various Distributions

There are a lot of distributions as follows;

- Poisson distribution
- Exponential distribution
- Beta distribution
- Gamma distribution
- t-distribution
- F-distribution
- Chi-squared distribution

I omit these distributions in this lecture.
Please refer to some textbooks about statistics.

Outline

Convergence of Deterministic Sequences

Definition (Convergence)

A sequence $\{a_n\}$ converges to a if for any $\varepsilon > 0$, there exists N such that $n \geq N$ implies $|a_n - a| < \varepsilon$. We write $a_n \to a$ as $n \to \infty$.

Definition (Convergence in Probability)

A sequence of random variables $\{X_n\}$ converges in probability to the random variable X if for any $\varepsilon > 0$,

$$\Pr(|X_n - X| > \varepsilon) \to 0 \text{ as } n \to \infty.$$

We write $X_n \xrightarrow{p} X$ or $X_n \to_p X$

Convergence in Distribution

Definition (Convergence in Distribution)

A sequence of random variables $\{X_n\}$ converges in distribution to the random variable X if

$$F_n(t) \to F(t) \quad \text{as} \quad n \to \infty$$

for all t at which F is continuous. F_i is the cumulative distribution function of X_i. We write $X_n \xrightarrow{d} X$ or $X_n \to_d X$.

Convergence in Mean Square

Definition (Convergence in Mean Square)

A sequence of random variables $\{X_n\}$ converges in mean square to the random variable X if

$$\lim_{n\to\infty} E[|X_n - X|^2] = 0$$

We write $X_n \xrightarrow{L^2} X$.

Almost Sure Convergence

Definition (Almost Sure Convergence)

A sequence of random variables $\{X_n\}$ converges almost surely to the constant number X if

$$\Pr\left(\lim_{n\to\infty} X_n = X\right) = 1$$

or letting the random variables be a function from $\Omega \to \mathbb{R}$, then

$$\Pr\left(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1$$

We write $X_n \xrightarrow{a.s.} X$.

Property of Convergence

Property (Convergence)

- $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$
- $X_n \xrightarrow{d} X \Rightarrow X_n \xrightarrow{p} X$ if X is constant
- $X_n \xrightarrow{L^2} X \Rightarrow X_n \xrightarrow{p} X$
- $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{p} X$

Outline

Markov's Inequality

Markov's Inequality

If X is a random variable and ε is a positive constant, then

$$\Pr(|X| \geq \varepsilon) \leq \frac{E(|X|)}{\varepsilon}$$

proof (X is discrete case)

$$\begin{aligned} E(|X|) &= \sum_{x \in \mathcal{X}} |x| \Pr(X = x) \\ &\geq \sum_{|x| \geq \varepsilon} |x| \Pr(X = x) \\ &\geq \sum_{|x| \geq \varepsilon} \varepsilon \Pr(X = x) = \varepsilon \Pr(|X| \geq \varepsilon) \end{aligned}$$

Chebychev's Inequality

Chebychev's Inequality

If X is a random variable and μ and ε are constants, then

$$\forall \varepsilon > 0 \quad \Pr(|X-\mu| \geq \varepsilon) \leq \frac{Var(X)^2}{\varepsilon^2}.$$

proof (X is continuous case)

Let $Y = X - \mu$, then $var(V) = E(Y^2)$

$$\begin{aligned} var(X) &= E(Y^2) \\ &= \int_{-\infty}^{\infty} w^2 f(w) dw \\ &\geq \int_{-\infty}^{-\varepsilon} w^2 f(w) dw + \int_{\varepsilon}^{\infty} w^2 f(w) dw \end{aligned}$$

Chebychev's Inequality

proof (continued)

$$
\begin{aligned}
var(X) &\geq \int_{-\infty}^{-\varepsilon} w^2 f(w)dw + \int_{\varepsilon}^{\infty} w^2 f(w)dw \\
&\geq \varepsilon^2 (\int_{-\infty}^{-\varepsilon} f(w)dw + \int_{\varepsilon}^{\infty} f(w)dw) \\
&= \varepsilon^2 Pr(|W| \geq \varepsilon)
\end{aligned}
$$

$$\therefore Pr(|X - \mu| \geq \varepsilon) \leq \frac{var(V)}{\varepsilon^2}$$

Cauchy-Schwarz Inequality

Cauchy-Schwarz Inequality

For two random variables X and Y,

$$E(|XY|) \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}.$$

proof

Let $W = Y + bX$, where b is a constant. Then,

$$E(W^2) = E(Y^2) + 2bE(XY) + b^2E(X^2)$$

Here, let $b = -\frac{E(XY)}{E(X^2)}$ ($E(X^2) \neq 0$) so that

$$E(Y^2) - \frac{\{E(XY)\}^2}{E(X^2)} \geq 0 \quad (\because E(W^2) \geq 0)$$

$$\therefore |E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$$

Law of Large Numbers (LLN)

Theorem (Weak Law of Large Numbers (WLLN))

If $X_1, \ldots, X_n$ is a sequence of independent and identically distributed random variables with $E(X_i) = \mu < \infty$ and $E(X_i^2) < \infty$, then

$$\bar{X}_n \xrightarrow{p} \mu$$

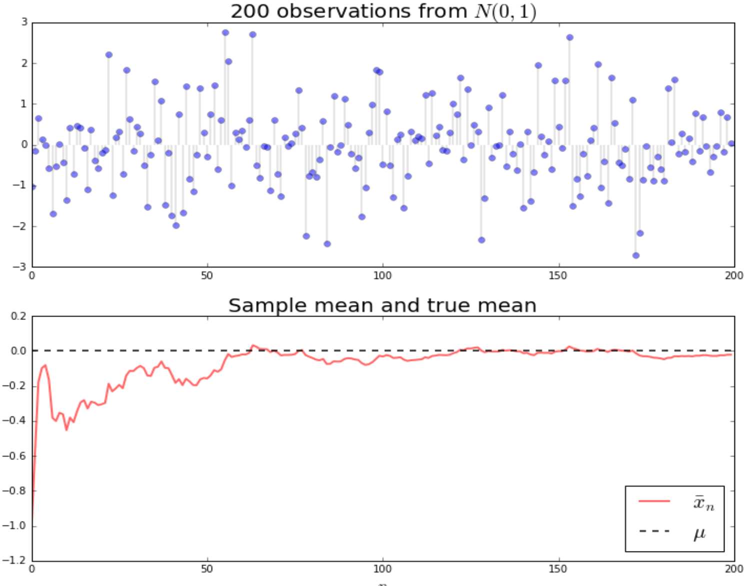
Theorem (Strong Law of Large Numbers (SLLN))

If $X_1, \ldots, X_n$ is a sequence of independent and identically distributed random variables with $E(X_i) = \mu < \infty$ and $E(X_i^2) < \infty$, then

$$\bar{X}_n \xrightarrow{a.s.} \mu$$

SLLN yields stronger conclusions but is hard to prove.

LLN working


200 observations from $N(0,1)$
3
2
1
0
−1
−2
−3
0
50
100
150
200
Sample mean and true mean
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
−1.2
0
50
100
150
200
$\bar{x}_n$
μ

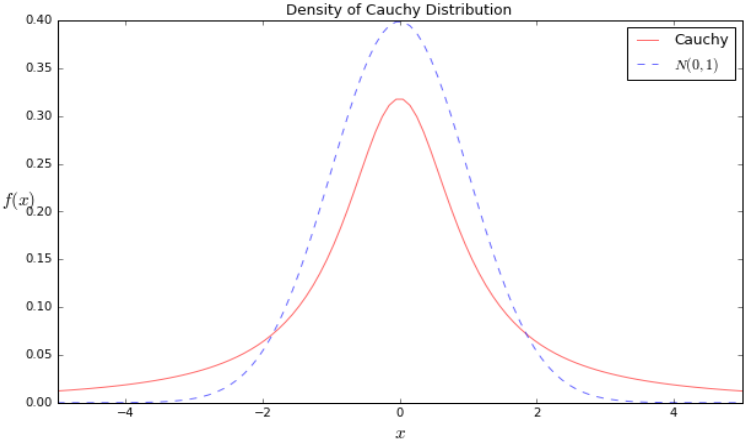
Cauchy Distribution

LLN works only under some conditions. The conditions of LLN might no be satisfied if the underlying distribution is heavy tailed.
The best known example is the Cauchy distribution, which has

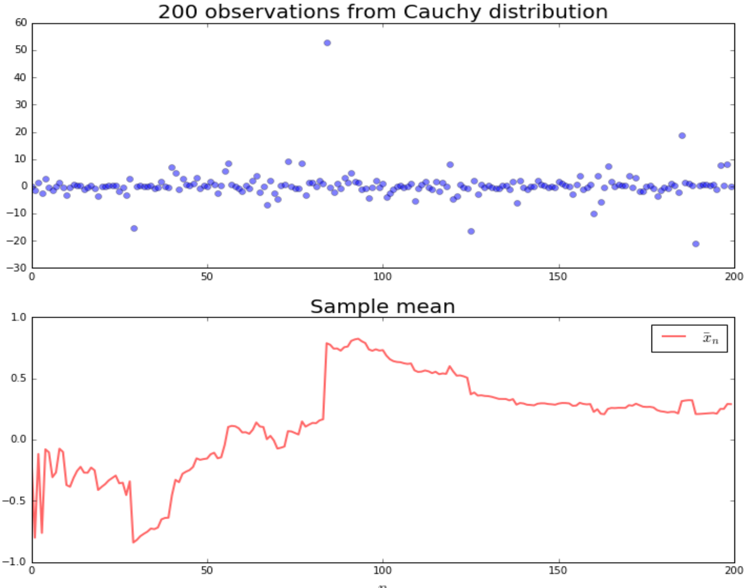
$$f(x) = \frac{1}{\pi(1+x^2)}$$

It is known that $E(X)$ is not defined for Cauchy distribution.

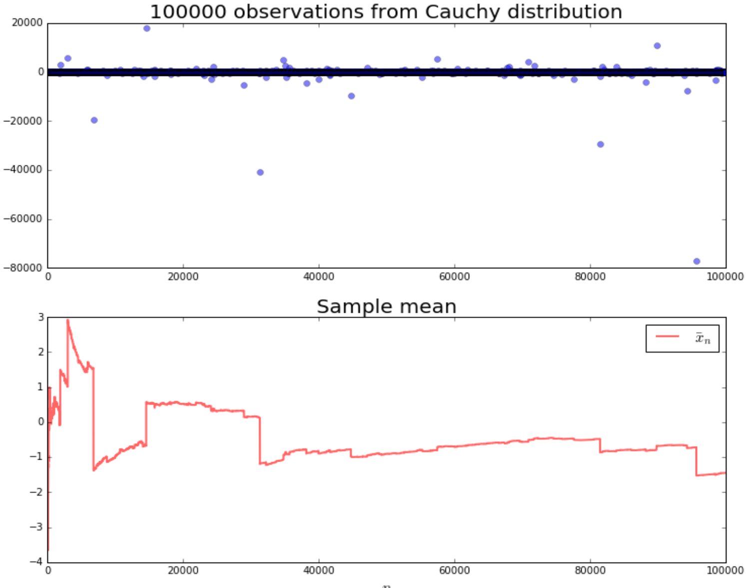
Shape of Cauchy Distribution



LLN under Cauchy Distribution ($N = 200$)



LLN under Cauchy Distribution ($N = 100000$)



Central Limit Theorem

Theorem (Central Limit Theorem)

If $X_1, \ldots, X_n$ is a sequence of independent and identically distributed random variables with $E(X) = \mu < \infty$ and $Var(X) = \sigma^2 < \infty$, then

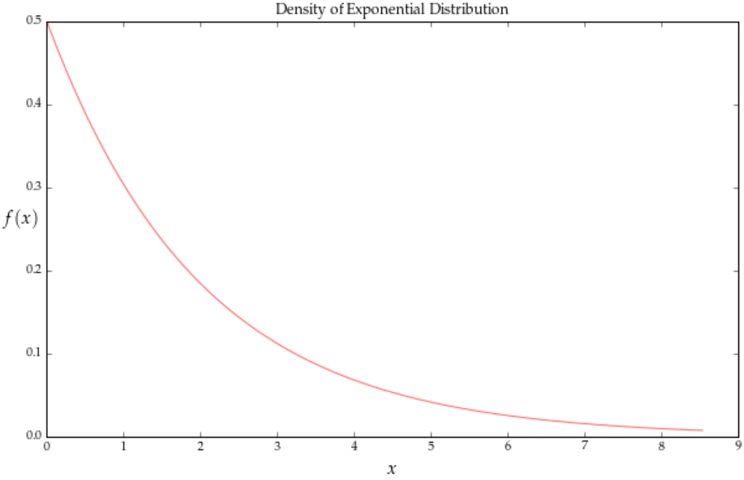
$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2).$$

Since the distribution of $X_1, \cdots, X_n$ is arbitrary, I use the following exponential distribution to simulate CLT

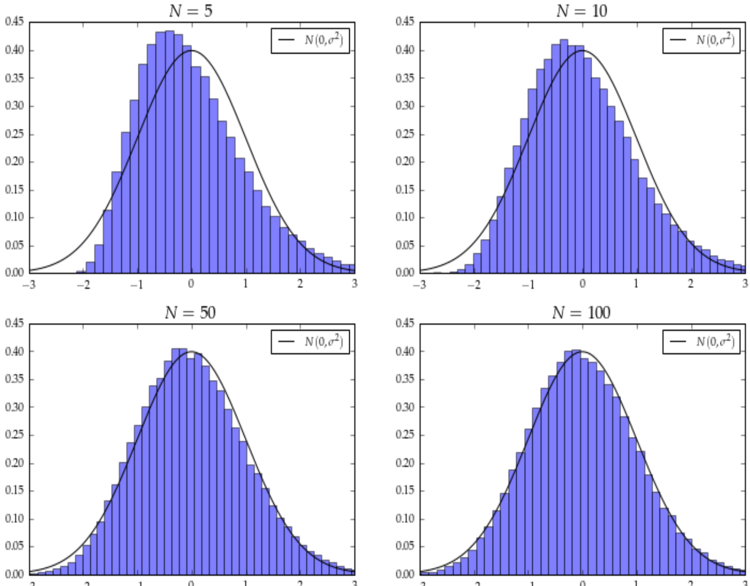
$$f(x) = \lambda \exp(-\lambda x).$$

I set $\lambda = 1/2$ and the number of simulations $T = 100000$.

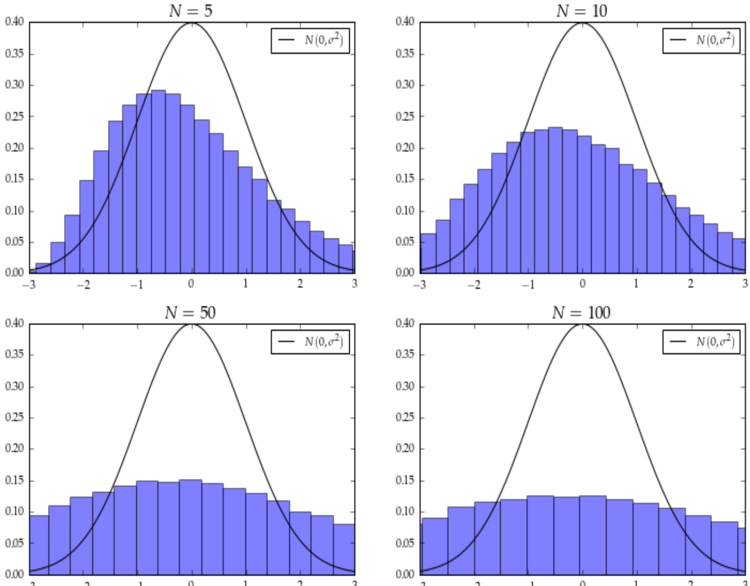
Exponential Distribution



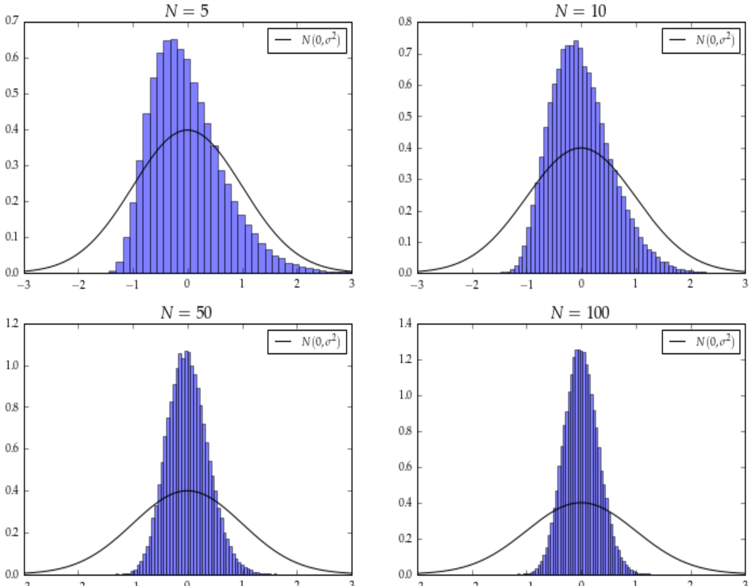
Histogram of $\sqrt{n}(\bar{X}_n - \mu)$



Histogram of $n^{\frac{3}{4}}(\bar{X}_n - \mu)$



Histogram of $n^{\frac{1}{4}}(\bar{X}_n - \mu)$



Slutsky's Theorem

Theorem (Slutsky's theorem)

Suppose that $a_n \xrightarrow{p} a$, where a is constant, and $S_n \xrightarrow{d} S$. Then

- $a_n + S_n \xrightarrow{d} a + S$
- $a_n S_n \xrightarrow{d} aS$
- If $a \neq 0$, $\frac{S_n}{a_n} \xrightarrow{d} \frac{S}{a}$

Continuous Mapping Theorem

Continuous mapping theorem

Suppose that g is a continuous function of a sequence of random variables, S_n. Then

- If $S_n \xrightarrow{p} a$, then $g(S_n) \xrightarrow{p} g(a)$
- If $S_n \xrightarrow{d} S$, then $g(S_n) \xrightarrow{d} g(S)$

e.g.

- If $\bar{X}_n \xrightarrow{p} \mu$, then $\bar{X}_n^2 \xrightarrow{p} \mu^2$
- If $S_n \xrightarrow{d} Z$, where $Z \sim N(0,1)$, then $S_n^2 \xrightarrow{d} \chi_1^2$

References

- Greene, William H.,"*Econometric analysis (7th edition)*", Pearson Education, 2011
- Sargent, Thomas, and John Stachurski, "*Lectures in Quantitative Economics*", `https://lectures.quantecon.org`
- Strang, Gilbert, "*Linear Algebra and Its Applications (4th edition)*", Brooks Cole, 2006
- Wooldridge, Jeffrey M.,"*Econometric analysis of cross section and panel data*", MIT press, 2010
- 石井恵一,"線形代数講義 [増補版]", 日本評論社, 2013 (in Japanese)

I made most of the figures in this lecture based on sample codes of Python in "*Lectures in Quantitative Economics*" by Sargent and Stachurski.