Introductory Econometrics

Basic Statistics

Yaohan Chen

School of Big Data and Statistics, Anhui University

Spring, 2025

Linear Algebra

We begin by recalling some elementary notions from linear algebra needed for introducing random vectors.

A **vector** $\mathbf{x}$ of dimension n is an ordered collection of n numbers, which are called **components** or elements:

$$\mathbf{x} = (x_1, \ldots, x_n)$$

Example

$$\mathbf{x} = (2, 3), \quad \mathbf{y} = (-1, 2), \quad \mathbf{z} = (\sqrt{2}, 0, \pi)$$

Linear Algebra

Two vectors $\mathbf{x}$ and $\mathbf{y}$ of common dimension n are *added* component by component:

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, \ldots, x_n + y_n)$$

Example

If $\mathbf{x} = (2, 3)$ and $\mathbf{y} = (-1, 2)$, then

$$\mathbf{x} + \mathbf{y} = (2 + (-1), 3 + 2) = (1, 5),$$

Two vectors $\mathbf{x}$ and $\mathbf{y}$ of common dimension n are *added* component by component:

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, \ldots, x_n + y_n)$$

Linear Algebra

If a is a number, we will sometimes refer to a as a **scalar**.

If a is a scalar and $\mathbf{x}$ is a vector, then the product $a\mathbf{x}$ of a and $\mathbf{x}$ is

$$a\mathbf{x} = (ax_1, \ldots, ax_n)$$

If $a = 0$ we get $a\mathbf{x} = \mathbf{0}$ where $\mathbf{0}$ is the **zero vector**, i.e. the vector of same dimension as $\mathbf{x}$ with all components equal to zero.

A **matrix** can be viewed in different ways: either as an array of numbers ordered into rows and columns, or as a collection of vectors.

Example

$$\mathbf{A} = \begin{pmatrix} 2 & 2 & -1 \\ 3 & 1 & 0 \end{pmatrix}$$

The vectors $(2, 3)$, $(2, 1)$ and $(-1, 0)$ are the columns of $\mathbf{A}$.

Linear Algebra

A matrix has m rows and n columns

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

To indicate the dimension of a matrix we sometimes write $\mathbf{A}_{m\times n}$.

Example

$$\mathbf{B} = \begin{pmatrix} 2 & 2 \\ 3 & 1 \end{pmatrix}$$

$\mathbf{B}$ is a 2×2 matrix with, for example, element $b_{12} = 2$. $\mathbf{B}$ is a so-called **square matrix**, i.e. a matrix with the same number of rows as columns ($m = n$).

Linear Algebra

Matrix addition of two matrices of common dimension is analogous to vector addition, and it can be shown that the previously mentioned properties of vectors also hold for matrices.

Example

$$\begin{pmatrix} 1 & 3 \\ 7 & -1 \end{pmatrix} + \begin{pmatrix} 2 & 5 \\ 8 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 8 \\ 15 & 0 \end{pmatrix}$$

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

We can also define **matrix multiplication**:

Let $\mathbf{A}$ be a $m \times l$ matrix, with row i and column k element a_{ik} , and let $\mathbf{B}$ be a $l \times n$ matrix, with row k and column j element b_{kj} . Then c_{ij} , the row i and column j element of the matrix $\mathbf{C} = \mathbf{AB}$, is given by

$$c_{ij} = \sum_{k=1}^{l} a_{ik} b_{kj}$$

Example

$$\begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 4 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 2 \times 0 + 1 \times 2 & 2 \times 4 + 1 \times 1 \\ 1 \times 0 + 0 \times 2 & 1 \times 4 + 0 \times 1 \end{pmatrix} = \begin{pmatrix} 2 & 9 \\ 0 & 4 \end{pmatrix}$$

Remark

From the definition it follows that the number of columns of $\mathbf{A}$ must equal the number of rows of $\mathbf{B}$ in order for the product $\mathbf{C} = \mathbf{AB}$ to be well-defined:

$$\mathbf{A}_{m \times l}\mathbf{B}_{l \times n} = \mathbf{C}_{m \times n}$$

Moreover, in general, $\mathbf{AB}$ is different from $\mathbf{BA}$.

Some elementary properties of matrices:

Let a be a scalar, and let $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ be matrices of dimensions such that the left-hand side expressions below are well-defined. Then,

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$$
$$\mathbf{A}(a\mathbf{B}) = a(\mathbf{AB})$$
$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$$

The **identity matrix**, $\mathbf{I}$, of dimension n is the $n \times n$ matrix with elements along its main diagonal equal to 1, and all other elements equal to 0:

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

It is easy to verify that

$$\mathbf{I}_m \mathbf{A} = \mathbf{A} = \mathbf{A} \mathbf{I}_n$$

for all $m \times n$ matrices $\mathbf{A}$.

Let $\mathbf{A}$ be a square matrix. If there exists a matrix $\mathbf{A}^{-1}$ such that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix of same dimension as $\mathbf{A}$, then $\mathbf{A}^{-1}$ is called the inverse of $\mathbf{A}$. Not all matrices have an inverse.

Example

If $A_{1\times1} = a_{11}$ and $a_{11} \neq 0$, then $\mathbf{A}^{-1} = \frac{1}{a_{11}}$

If $\mathbf{B} = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$, then $\mathbf{B}^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & -2 \end{pmatrix}$

The **transpose** $\mathbf{A}^\top$ of a matrix $\mathbf{A}_{m\times n}$ is the $n \times m$ matrix whose ith column is the ith row of $\mathbf{A}$:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \mathbf{A}^\top = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}$$

- A square matrix $\mathbf{A}$ with $\mathbf{A}^\top = \mathbf{A}$ is called a **symmetric matrix**.

The **determinant** of a 2×2 matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

is given by the scalar

$$|\mathbf{A}| = a_{11}a_{22} - a_{12}a_{21}$$

The determinant of a general $n \times n$ matrix is more complicated but can be defined recursively.

One case is particularly simple:

The determinant of a diagonal matrix is the product of the elements along its main diagonal:

$$|\mathbf{A}| = a_{11} \times \cdots \times a_{nn}$$

Example

Since $|\mathbf{I}_n| = 1 \times \cdots \times 1$, the determinant of the identity matrix is one.

Sets

We begin by recalling some elementary notions from set theory needed for introducing concepts from probability theory.

- A set is a collection of objects. Sets are usually denoted by upper-case letters such as A, B or C.
- If an object c belongs to a set C we write $c \in C$ (read, "c in C").
- If c does not belong to C we write $c \notin C$ (read, "c not in C").
- If $c_1, \ldots, c_n$ are objects, the set consisting of precisely these n objects is denoted by $\{c_1, \ldots, c_n\}$.

Sets

- The objects belonging to a set A are called its **elements** (or members). The elements can be **anything**, e.g. numbers, outcomes or other sets.
- Let A and B be sets. The **intersection** of A and B is the set whose elements are those objects c such that $c \in A$ **and** $c \in B$. We write $A \cap B$ (read, "A intersect B").
- The **union** of A and B is the set whose elements are those objects c such that c belongs to **at least** one of the two sets A,B (i.e. either $c \in A$ or $c \in B$, or both). We write $A \cup B$ (read, A union B).

Sets

Example

When flipping a coin, 'head' or 'tail' occurs. If $A = \{\text{head}, \text{tail}\}$ and $B = \{\text{head}\}$, then

$$A \cap B = \{\text{head}\} \text{ and } A \cup B = \{\text{head}, \text{tail}\}$$

Sets

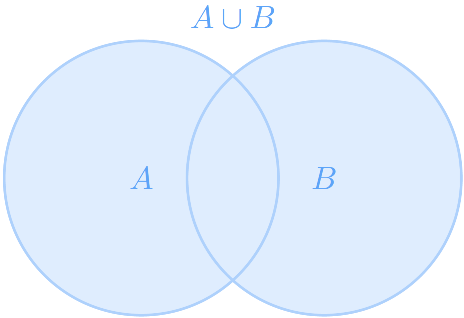

Figure : The union of A and B illustrated using a Venn diagram

Sets

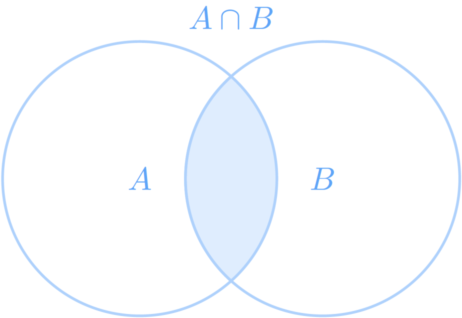

Figure : The intersection of A and B

Sets

- If all elements of a set A are also elements of a set B, we say that A is a **subset** of B, and write $A \subset B$.
- **Empty** set is the set that has no members. The empty set is denoted by $\varnothing$.
- The set $\varnothing$ is a subset of **any** set.

Sets

The **difference** of two sets A and B, written in $A - B$, is the set of all elements that are in A but not in B.

Example

If $A = \{\text{STI}, \text{HSI}, \text{SSE}\}$ and $B = \{\text{HSI}\}$, then $A - B = \{\text{STI}, \text{SSE}\}$.

If A and B have no elements in common we say that the two sets are **disjoint**, $A \cap B = \varnothing$.

Let $A \subset C$. The **complement** of A in C is the set of elements that belong to C but not to A. We write A^c .

Example The possible outcomes of a coin tossing experiment are 'head' and 'tail'. Here $C = \{head, tail\}$. Hence, if $A = \{\text{head}\}$, then $A^c = \{\text{tail}\}$.

Function

A **function** is a rule that associates each member of one set with a member of another set.

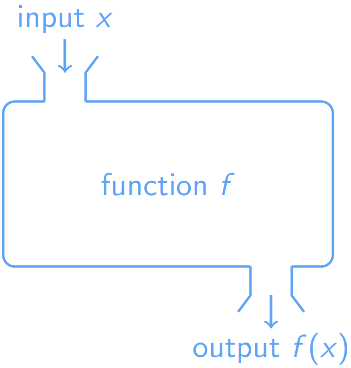

Figure : A function f takes an input x and returns an output $f(x)$.

Random Variables

- Assigning a value to each random outcome.
- When tossing a coin we can write '1' for 'head' and '0' for tail. In this way, we get a **random variable** $X(\omega) \mapsto \{0,1\}$, where ω belongs to the **sample space** $\mathcal{F} = \{\text{head}, \text{tail}\}$.
- $\mathcal{F}$ is a abstract space collecting **all possible outcomes** of the underlying experiment.
- The random variable $X(\omega)$ is nothing but a **real-valued function** defined on $\mathcal{F}$ (i.e. a numerical summary of a random outcome).

Random Variables

- How to make $\mathcal{F}$ complete ? $\mathcal{F}$ should be like:
  - If $A \in \mathcal{F}$, so is its complement A^c;
  - If $A, B \in \mathcal{F}$, so are $A \cap B$, $A \cup B$, $A \cup B^c$, $B \cup A^c$, $A \cap B^c$, $B \cap A^c$, etc.
- In some advanced textbooks $\mathcal{F}$ is called **σ-filed**.

Probability

- Probability is a measure such that for each $A \in \mathcal{F}$, it assigns a number $\mathrm{P}(A) \in [0, 1]$.
- Probability should satisfy:
    - For $A, B \in \mathcal{F}$

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

    - If A and B are disjoint,

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$$

    - Moreover,

$$\mathrm{P}(A^c) = 1 - \mathrm{P}(A), \quad \mathrm{P}(\mathcal{F}) = 1 \quad \text{and} \quad \mathrm{P}(\varnothing) = 0$$

Distribution

- The collection of the probabilities

$$F_X(x) = \mathrm{P}(X \leq x) = \mathrm{P}(\{\omega : X(\omega) \leq x\}), \quad x \in \mathbf{R}$$

  is the **Cumulative Distribution Function (CDF)** $F_X(x)$ of X. $F_X(x)$ of X gives the probability that X belongs to the interval $(a, b]$ as

$$\mathrm{P}(\{\omega : a < X(\omega) \leq b\}) = F_X(b) - F_X(a), \quad a < b$$

- Continuous distributions have **Probability Density Function (PDF)** $f_X(x)$:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \quad x \in \mathbf{R}, \quad \text{and} \int_{-\infty}^{\infty} f_X(t)dt = 1$$

Random Vectors

In the following lectures we will frequently make use of certain finite-dimensional random structures.

We consider finite-dimensional random vectors:

$\mathbf{X} = (X_1, \ldots, X_n)$ is a n-dimensional **random vector** if its components $X_1, \ldots, X_n$ are one-dimensional real-valued random variables.

Random Vectors

Example

Toss a coin. We consider the pairs

$$\text{(head,head)}, \quad \text{(tail,tail)}, \quad \text{(head,tail)}, \quad \text{(tail,head)}$$

as outcomes of the experiment. These four pairs form the sample space $\mathcal{F}$. We can write '1' for 'head' and '0' for 'tail'. In this way, we get two random variables X_1 and X_2, and $\mathbf{X} = (X_1, X_2)$ is a two-dimensional random vector.

Example (Cont'd)

Note that

$$\mathbf{X}(\text{head}, \text{head}) = (1, 1), \quad \mathbf{X}(\text{tail}, \text{tail}) = (0, 0)$$

$$\mathbf{X}(\text{head}, \text{tail}) = (1, 0), \quad \mathbf{X}(\text{tail}, \text{head}) = (0, 1)$$

If the coin is 'fair', we can assign the probability 0.25 to each of the four outcomes, i.e.

$$\mathrm{P}(\{\omega : \mathbf{X}(\omega) = (k, i)\}) = 0.25, \quad k, i \in \{0, 1\}$$

Random Vectors

- The collection of the probabilities

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}) &= \mathrm{P}\left(X_1 \leq x_1, \ldots, X_n \leq x_n\right) \\ &= \mathrm{P}\left(\left\{\omega : X_1(\omega) \leq x_1, \ldots, X_n(\omega) \leq x_n\right\}\right) \end{aligned}$$

  where $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbf{R}^n$, is the **joint CDF** $F_{\mathbf{X}}$ of $\mathbf{X}$.
- $F_{\mathbf{X}}(\mathbf{x})$ is the shorthand for $F_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$.
- Correspondingly, **joint PDF** of $F_{\mathbf{X}}$ is

$$F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{X}}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\mathbf{X}}(w_1, \ldots, w_n)\, dw_1 \cdots dw_n$$

Random Variable: Expectation, Variance, Moment

- Expectation

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

- Variance

$$\sigma_X^2 = \operatorname{Var}(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$$

- I-th moment

$$\mathbb{E}\left(X^I\right) = \int_{-\infty}^{\infty} x^I f_X(x) dx$$

Correlation

- **Correlation** between two random variables X_1 and X_2 is defined as

$$\operatorname{Corr}(X_1, X_2) = \frac{\operatorname{Cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} = \frac{\mathbb{E}\left[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})\right]}{\sigma_{X_1}\sigma_{X_2}}$$

- $-1 \le \operatorname{Corr}(X_1, X_2) \le 1$. Why ?

Independence

Two **events** A_1 and A_2 are **independent** if

$$\mathrm{P}\left(A_1 \cap A_2\right) = \mathrm{P}\left(A_1\right) P\left(A_2\right)$$

Two **random variables** X_1 and X_2 are **independent** if

$$\mathrm{P}\left(X_1 \in B_1, X_2 \in B_2\right) = \mathrm{P}\left(X_1 \in B_1\right) \mathrm{P}\left(X_2 \in B_2\right)$$

for all suitable subsets of B_1 and B_2 of $\mathbf{R}$. This means that the events

$$\{c : X_1(\omega) \in B_1\} \quad \text{and} \quad \{\omega : X_2(\omega) \in B_2\}$$

are independent.

Independence

- The random variables $X_1, \ldots, X_n$ are mutually independent if and only if their joint CDF can be written as

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \times \cdots \times F_{X_n}(x_n), \ (x_1, \ldots, x_n) \in \mathbf{R}^n$$

- If the random vector $\mathbf{X} = (X_1, \ldots, X_n)$ has joint PDF $f_X = f_{X_1,\ldots,X_n}$ with marginal pdfs $f_{X_1}, \ldots, f_{X_n}$, then $X_1, \ldots, X_n$ are mutually independent if and only if

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \times \cdots \times f_{X_n}(x_n), \ (x_1, \ldots, x_n) \in \mathbf{R}^n$$

Correlation and Independence

- Uncorrelated random variables $\Rightarrow$ Independent Random Variables ? Any counterexample ?
- If the random variables $X_1, \ldots, X_n$ are mutually independent and have the *same distribution*, we say that they are **independent and identically distributed (iid)**.

Random Vectors: Expectation,Variance-Covariance Matrix

- Expectation

$$\mu_{\mathbf{X}} = \mathbb{E}(\mathbf{X}) = (\mathbb{E}(X_1), \ldots, \mathbb{E}(X_n))$$

- The Variance-covariance matrix of $\mathbf{X}$ is defined as the matrix $\mathbf{\Sigma}_{\mathbf{X}}$ with row i column j element given by

$$\operatorname{Cov}(X_i, X_j), \quad i, j = 1, \ldots, n$$

where

$$\begin{aligned} \operatorname{Cov}(X_i, X_j) &= \mathbb{E}\left[(X_i - \mu_{X_i})(X_j - \mu_{X_j})\right] \\ &= \mathbb{E}(X_i X_j) - \mu_{X_i}\mu_{X_j} \end{aligned}$$

and $\operatorname{Cov}(X_i, X_i) = \sigma^2_{X_i}$.

Normal Distribution

The most important continuous distribution is the normal or Gaussian distribution:

A random variable X is said to be **normally distributed** or $\mathcal{N}(\mu, \sigma^2)$ with parameters $-\infty < \mu < \infty$ if

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$$

and when $\mu = 0$ and $\sigma = 1$, X is called **standard normal**.

The CDF of standard normal distribution has its own notation $\Phi(x)$.

Multivariate Normal Distribution

The **Multivariate Normal Distribution (n-dimensional normal)** or **Gaussian Distribution** is given by its joint PDF

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})^{\top}\right\}, \quad \mathbf{x} \in \mathbf{R}^n$$

with parameters $\boldsymbol{\mu} \in \mathbf{R}^n$ and $\boldsymbol{\Sigma}$ is symmetric (positive definite) $n \times n$ matrix, $\boldsymbol{\Sigma}^{-1}$ its inverse and $|\boldsymbol{\Sigma}|$ its determinant. Multivariate Normal Distribution is denoted by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Multivariate Normal Distribution

Example

Suppose $\mathbf{X} = (X_1, X_2)$ is 2-dimensional normal with

$$\boldsymbol{\mu} = (0, 0) \quad \text{and} \quad \boldsymbol{\Sigma} = \mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

then

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{1}{2}\left(x_1^2 + x_2^2\right)}$$

for $\mathbf{x} \in \mathbf{R}^2$.

Multivariate Normal Distribution

Example

If $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_n$, then the density $f_{\mathbf{X}}$ is simply the product of n standard normal densities:

$$f_{\mathrm{X}}(x_1, \ldots, x_n) = \varphi(x_1) \times \cdots \times \varphi(x_n), \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Let $\mathbf{X} = (X_1, \ldots, X_n)$ have an $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution and $\mathbf{A}$ be an $m \times n$ matrix. Then $\mathbf{A}\mathbf{X}^\top$ has an $\mathcal{N}(\mathbf{A}\boldsymbol{\mu}^\top, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$ distribution.

Multivariate Normal Distribution

Example

If $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_n$, then the density $f_{\mathbf{X}}$ is simply the product of n standard normal densities:

$$f_{\mathrm{X}}(x_1, \ldots, x_n) = \varphi(x_1) \times \cdots \times \varphi(x_n), \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Let $\mathbf{X} = (X_1, \ldots, X_n)$ have an $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution and $\mathbf{A}$ be an $m \times n$ matrix. Then $\mathbf{A}\mathbf{X}^\top$ has an $\mathcal{N}(\mathbf{A}\boldsymbol{\mu}^\top, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$ distribution.

In the Gaussian case, uncorrelatedness and independence are equivalent notions. This statement is wrong for non-Gaussian vectors.

- If $Z_1, \ldots, Z_k$ are **iid standard normal** random variables, then
$$\sum_{i=1}^{k} Z_i^2 = Z_1^2 + Z_2^2 + \cdots + Z_k^2 \sim \chi^2_{(k)}$$
- A random variable T follows the student t distribution with k degrees of freedom, written as $T \sim t(k)$ if $T = \frac{U}{\sqrt{V/k}}$, where $U \sim \mathcal{N}(0,1)$, $V \sim \chi^2_{(k)}$, and **U and V are independent**.
- A random variable F follows the F-distribution with (m, n) degrees of freedom, written as $F \sim F(m,n)$, if $F = \frac{U/m}{V/n}$, where $U \sim \chi^2_{(m)}$ and $V \sim \chi^2_{(n)}$, and **U and V are independent**.

Sampling

- In a typical real-world statistical problem, we have a random variable X of interest, but the PDF $f(x)$ is not known.
- Our lack of knowledge can be classified in on e of two ways:

  - $f(x)$ is completely unknown.
  - The functional form of $f(x)$ is assumed to be known up to a parameter vector θ.

Example

X has a normal distribution $\mathcal{N}(\mu, \sigma^2)$, where $\theta = (\mu, \sigma^2)$.

Sampling

We often write $f(x;\theta)$, where $\theta \in \Omega$ for a specified set Ω, to emphasize that the PDF is known up to θ.

Example

If X has a normal distribution $\mathcal{N}(\mu, \sigma^2)$, then

$$\Omega = \{\theta = (\mu, \sigma^2) : \mu \in \mathbf{R}, \sigma^2 > 0\}$$

- We call θ a **parameter** of the distribution.
- As θ is unknown, we want to **estimate** it.

Sampling

In practice, our information about the unknown distribution of X, or the unknown parameters of the distribution of X, comes from a **sample** of X.

The sample observations have the same distribution as X, and we denote them as the random variables $X_1, \ldots, X_n$.

- n denotes the sample size.
- When the sample is actually drawn, we use lower case letters $x_1, \ldots, x_n$ to denote the values or **realizations** of the sample

Statistics

We often use functions of a sample to summarize the information in it:

Let $X_1, \ldots, X_n$ denote a sample of a random variable X, and let $T = T(X_1, \ldots, X_n)$ be a function of the sample, then T is called a statistic.

- T is a random variable, and **has nothing to do with θ**.
- Once is the sample is drawn, t is called a **realization** of T, where $t = T(x_1, \ldots, x_n)$ and $x_1, \ldots, x_n$ is the realization of the sample.

Statistics

Let $X_1, \ldots, X_n$ denote a sample of random variable X with a pdf $f(x;\theta)$, where $\theta \in \Omega$ for a specified set Ω.

- Statistic T is called a **point estimator** of θ, usually denoted by $\hat{\theta}$.
- The realization t of T is called an **estimate** of θ.

Point estimator might have various properties: unbiasedness, consistency, and efficiency.

Statistics

We temporarily focus on **unbiasedness**, but leave the discussion about consistency and efficiency later.

Let $X_1, \dots, X_n$ denote a sample of a random variable X with pdf $f(x;\theta)$, $\theta \in \Omega$, and let $T = T(X_1, \dots, X_n)$ be a statistic. Then T is an **unbiased** estimator of θ if $\mathbb{E}(T) = \theta$.

- The mean of an unbiased estimators sampling distribution is located at the true (but unknown) value of the parameter of interest.
- An estimator T of θ with $\mathbb{E}(T) \neq \theta$ is **biased**.

Example

Consider the dataset where the variable of interest X is the number of operating hours until the first failure of air-conditioning units for Boeing 720 airplanes.

A sample with size $n = 13$ was obtained, with realized values:

$$359, 413, 25, 130, 90, 50, 50, 487, 102, 194, 55, 74, 97$$

- $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a point **estimator** of θ.
- $\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{13}(359 + 413 + 25 + 130 + 90 + 50 + 50 + 487 + 102 + 194 + 55 + 74 + 97) = 163.5385$ is the corresponding **estimate** of θ.

Confidence Intervals

We use confidence intervals to measure the error of the corresponding estimate:

Let $X_1, \ldots, X_n$ be a sample of a random variable X, where X has pdf $f(x;\theta)$, $\theta \in \Omega$. Let $0 < \alpha < 1$ be specified. Let $L = L(X_1, \ldots X_n)$ and $U = U(X_1, \ldots, X_n)$ be two statistics. The interval (L, U) is a $(1-\alpha)100\%$ **confidence interval** for θ if

$$1 - \alpha = \mathrm{P}_\theta[\theta \in (L, U)]$$

where P_θ refers to the probability when θ is the true parameter. That is, the probability that the random interval includes θ is $1 - \alpha$, which is called the **confidence level** of the interval.

Confidence Intervals

Example

Suppose, for simplicity, that $\hat{\theta}$ denotes the estimator of θ for $\mathcal{N}(\theta, \sigma^2)$ with σ^2 known. Then,

$$P_\theta\left(-z_{\alpha/2} < \frac{\hat{\theta} - \theta}{\sigma} < z_{\alpha/2}\right) = P_\theta\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1 - \alpha$$

where we have used that Z is a standard normal random variable.

A $(1-\alpha)100\%$ confidence interval is now obtained by solving for θ in the above equation. This gives

$$P_\theta\left(\hat{\theta} - z_{\alpha/2}\sigma < \theta < \hat{\theta} + z_{\alpha/2}\sigma\right) = 1 - \alpha$$

Hypothesis Testing

Point estimation and confidence intervals are useful statistical inference procedures. Another type of inference that is frequently used concerns tests of hypotheses.

We label these hypotheses as

$$H_0 : \theta \in \Omega_0 \quad \text{against} \quad H_1 : \theta \in \Omega_1$$

where $\Omega_0 \cap \Omega_1 = \varnothing$ and $\Omega_1 \subset \Omega$, $\Omega_0 \subset \Omega$

H_0 is called the **null hypothesis**, and H_1 the **alternative hypothesis**.

Hypothesis Testing

The decision take H_0 or H_1 is based on a sample $X_1, \ldots, X_n$ from the distribution of X:

Denote the range of the random sample $X = (X_1, \ldots, X_n)$ by $\mathcal{D}$. A **test** of H_0 against H_1 is based on a subset C of $\mathcal{D}$. This set C is called the **critical region** and its corresponding decision rule is

$$\text{Reject } H_0 \text{ (accept } H_1\text{) if } \mathbf{X} \in C$$

$$\text{Accept } H_0 \text{ (reject } H_1\text{) if } \mathbf{X} \in C^c$$

Hypothesis Testing

Since the decision rule to take H_0 or H_1 is based on a random sample, the decision could be wrong.

	True State of Nature	
Decision	H_0 is true	H_1 is true
Reject H_0	Type I error	Correct Decision
Accept H_0	Correct Decision	Type II error

In Econometrics, we often refer to statistical significance by controlling the probability making Type I error, say by setting a probability threshold α.

Summary

- Basic linear algebra operation.
- Basic set theory.
- Random Variables and Random Vectors.
- Some fundamental distributions: Normal Distribution, Multivariate Normal Distribution, χ^2-Distribution, t-Distribution, F-Distribution.
- Sampling, Estimation (Estimator), and Statistics.
- Confidence Intervals and Hypothesis Testing.