Probability and Statistics
Lecture 01

Dao-Hong XIANG

Email: daohongxiang@zjnu.cn
Office: **21-404**

A Brief Outline of this Course

Textbook: R.E. Walpole, R. H. Myers, S. L. Myers and K. Y. Ye, *Essentials of Probability & Statistics for Engineers & Scientists (first Edition)* China Machine Press 2013.

We study Chapters 1--5 of this book

- Probability
- Probability Distribution
- Sampling Distribution and Estimation

- I: attendance rate, written assignments and quizzes (50%)
- II: final exam (50%)
- Written assignments: once per week
- final grade=I$\times$ 50%+II$\times$ 50%

Why Probability and Statistics

- Statistics is the science of *collecting, analyzing, presenting, and interpreting data*.
- Statistics is the *corner stone* of information science.
- It deals with the *extraction of information from data*.
- It *explains and predicts* the model of uncertainty.
- The explanation helps us understand how the nature works.
- The prediction helps us make decisions.
- Statistics has revolutionized virtually all areas of scientific investigation in the 20th century, including: physics, chemistry, biology, sociology, economics, etc, etc...
- It has been becoming even more important in the past two decades, thanks to the wide application of computers and high throughput experimental techniques.

Example 1. Insurance industry

- How does a company decide premium for an insurance policy?
- For medical insurance, how to decide the premium for a person that does or dose not smoke?
- We need statistics!

Example 2. Google, Baidu, Bing, Amazon, and etc

- ``[W]e create as much information in two days now as we did from the dawn of civilization through 2003." ---Eric Schmidt in 2010.
- That's approximately 5 billion GB per two days, and that's in 2010.
- Searching: input a key word, output a ranked list of web pages
- How to adapt? What if most of the people click the second, not the top link?
- Amazon: what might the user need? How to place ads effectively?
- We need statistics!

Example 3. Statistical arbitrage and high frequency trading

- As reported recently, US high frequency traders switched from fiber-optic cable to microwave, saving almost half of the time conveying data.
- Patterns of deviation from equilibrium are detected very quickly so as to create arbitrage opportunity in the next few seconds.
- We need statistics!

The age of Big Data is coming!

- The age of big data is coming, bringing about positive or negative changes to our living place.
- It is the development of new technologies that makes the generating, storing, transporting and processing of huge amount of data possible.
- It is statistics that makes sense of the data.
- ``For Today's Graduate, Just One Word: Statistics." ---New York Times Aug 6, 2009
- A challenging but interesting age is upon us.

- *Experiments*

Experiment	Outcomes
Toss a coin	Head, Tail
Roll a die	1,2,3,4,5,6
Play a football game	win, lose, tie

Experiment

An **experiment** is defined to be any process which randomly generates well defined outcomes.

1. First, each single repetition of the experiment generates one and only one of the possible outcome.
2. Second, the outcome is random, and different repetitions may generate different outcomes. However, we know the entire set of possibilities for each experiment.

Sample space

Definition (1.1)

The set of all possible outcomes of a statistical experiment is called the **sample space** and is represented by the symbol S. Each outcome in a sample space is called an **sample point** or an **element** or a **member** of the sample space.

Definition (1.2)

An **event** is a collection of one or more of the possible outcomes of an experiment, is a *subset* of a sample space.

Usually the event is denoted by a capital letters such as A, B, C, etc.

Example: Rolling a die

- Sample space: $\{1, 2, 3, 4, 5, 6\}$
- In a specific experiment, one has a sample, for example, $A = \{3\}$.
- Examples of events: $\{1\}$, ``faces with odd numbers"=$\{1, 3, 5\}$, ``all the faces"=$\{1, 2, 3, 4, 5, 6\}$.

Example: a point of a square $[0, 1] \times [0, 1]$

(draw a square)

- Sample space: $[0, 1] \times [0, 1]$
- In a specific experiment, one has a sample, for example, ``A=(0.1,0.2)''.
- Examples of events: $\{(0.5, 0.5)\}$, $[0, 0.5] \times [0, 0.5]$, $[0, 1] \times [0, 1]$.
- If the point is ``randomly'' thrown onto the square, we know that the ``probability'' of $\{(0.5, 0.5)\}$ is 0, that of $[0, 0.5] \times [0, 0.5]$ is 1/4, and that of $[0, 1] \times [0, 1]$ is 1.
- The ``probability'' of any specific point is 0 --- less informative.

Example

An experiment consists of flipping a coin three times and each time noting whether it lands heads or tails.

(a). What is the sample space of this experiment?

(b). What is the event that tails occur more often than heads?

Solution: Let us use ``$H$'' to denote the head, and ``T'' to denote the tail.

(a). Sample space: $\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$. The size of sample space is 8.

(b). We put the samples with more tails together, to give the event.

$$\{HTT, TTH, THT, TTT\}$$

Set notation: representing a set

Set: a collection of elements

Method 1: list all the elements in the set.

For example, $S = \{\sqrt{2}, 34, \pi\}$ denotes the set containing three elements. Note that, the order of the elements is not important. For example, $\{\sqrt{2}, 34, \pi\}$ and $\{\pi, \sqrt{2}, 34\}$ are the SAME set.

Method 2: use the *set-builder notation, or rule method*. For example

$$A = \{x : x \text{ is an integer and } 0 < x < 7\},$$
$$S = \{x : 2x^2 - 5x - 3 = 0\}.$$

In the above two examples,

$$A = \{1, 2, 3, 4, 5, 6\},$$
$$S = \{3, -\frac{1}{2}\}.$$

Interval: a piece on the real line

Intervals may or may not include their end points.

- $[-1.5, 5.3] = \{x : -1.5 \le x \le 5.3\}$

- $(-1.5, 5.3] = \{x : -1.5 < x \le 5.3\}$

- $[-1.5, 5.3) = \{x : -1.5 \le x < 5.3\}$

- $(-1.5, 5.3) = \{x : -1.5 < x < 5.3\}$

Notation ∞: $[a, \infty) = \{x : a \le x\}$. In particular, $\mathbb{R} = (-\infty, \infty)$.

Event Relations = Set Relations

Notation

- $x \in S$: x is an element of the set S.
- $x \notin S$: x is not an element of the set S.

For example, ``$x \in \mathbb{N}$'' means ``x is an element of $\mathbb{N}$'', or equivalently, ``x is a positive integer''.

Notation

A is a subset of B if and only if every element of A is also an element of B. This subset relation is denoted by $A \subset B$, or $B \supset A$.

- $A = B$ if and only if $A \subset B$ and $B \subset A$. Example: $\{2, 1, 3\} = \{1, 3, 2\} = \{\frac{6}{2}, \frac{19}{19}, 5 - 3\}$.
- When $A \subset B$ but $B \not\subset A$, we call A a proper subset of B.
- Recall that $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$, and these inclusions are all proper.

Event Intersection

Event Intersection

$A \cap B$: intersection of A and B.
That is
$A \cap B = \{x : x \in A \text{ and } x \in B\}$.

Example

$\{\sqrt{2}, \pi\} \cap \{\pi, 5\} = \pi$,
$[-1, 3] \cap (0, 5] = (0, 3]$.

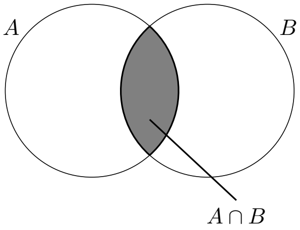

Figure: Venn diagram for the set intersection.

Empty Set

$\emptyset$: empty set (a set with no element).

Disjoint

If $A \cap B = \phi$, we said A and B are **mutually exclusive** or **disjoint**.

For example, $\{1, 2\} \cap \{3, 4\} = \emptyset$.

Event Union

Event Union

$A \cup B$: the union of two events A and B. That is,

$$A \cup B = \{x : x \in A \text{ or } x \in B\}.$$

Example

Let A be the set of odd positive integers and B be the set of even positive integers, then $A \cup B = \mathbb{N}$.

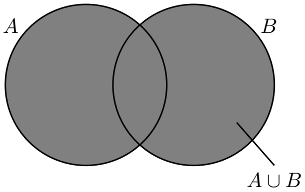

Figure: Venn diagram for the set union.

complement

$A' = \{x \in A : x \notin S\}$: the complement of an event A with respect to S is the subset of all elements of sample space S that are not in A.

- Multiplicative Rule
- Permutations Rule
- Combinations Rule

Multiplicative Rule

If one task needs k steps to finish, and has n_1 ways to finish the first step, n_2 ways to do the second step, etc., then he has in total

$$n_1 \times n_2 \times \cdots \times n_k$$

ways to finish the task.

appetizer	**main course**	**drinks**
egg	pork	juice
apple	steak	beer
	fish	water
		wine

how many possible combinations?

Definition (1.7 permutation)

A **permutation** of a set of objects, is any arrangement of these objects in a definite order.

Example

If $S = \{a, b, c\}$, then

$$abc, acb, bca, bac, cab, cba$$

are all the permutations of the elements of S.

Factorial

$$n! = n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1.$$

Specially, we define $0! = 1$.

Theorem (1.1)

The number of permutations of n different elements is $n!$.

Theorem (1.2)

The number of permutations of n distinct objects taken r at a time when $1 \le r \le n$, is

$$_nP_r = \frac{n!}{(n-r)!} = n(n-1)(n-2)\cdots(n-r+1).$$

Example

A company has eight applicants to six different jobs. In how many different ways can the jobs be filled with the eight applicants?

Solution.

$$_8P_6 = \frac{8!}{(8-6)!} = \frac{8!}{2!} = \frac{8\times 7\times 6\times 5\times 4\times 3\times 2\times 1}{2\times 1} = 20160$$

Theorem (1.3 Circular permutations)

The number of permutations of n objects arranged in a circle is $(n-1)!$

Example

If $S = \{a, b, c\}, b = c = x$, then

$$axx, axx, xxa, xax, xax, xxa$$

we have $\frac{3!}{2!} = 3$ distinct permutations

Theorem (1.4)

The number of distinct permutations of n things of which n_1 *are of one kind,* n_2 *of a second kind,* $\cdots$ *,* n_k *of a kth kind is* $\frac{n!}{n_1!n_2!\cdots n_k!}$.

Example

Consider the set $S = \{a, e, i, o, u\}$, then the possible partitions into two cells in which the first cell contains 4 elements and the second cell 1 element are

$$\{(a, e, i, o), (u)\}, \{(a, i, o, u), (e)\}, \{(e, i, o, u), (a)\},$$
$$\{(a, e, o, u), (i)\}, \{(a, e, i, u), (o)\}$$

we have $\frac{5!}{4!1!} = 5$ distinct permutations

Theorem (1.5 cells)

The number of ways of partitioning a set of n objects into r cells with n_1 elements in the first cell, n_2 elements in the second, and so forth, is

$$\binom{n}{n_1, n_2, \cdots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}$$

where $n_1 + n_2 + \cdots + n_r = n$.

Definition

A **combination** of a set of objects is a group or subset of the objects disregarding their order.

$$ {}_nC_r, \text{ or } \binom{n}{r} = \frac{n!}{r!(n-r)!}. $$

Example

How many ways can an executive committee of 5 be chosen from a board of directors consisting of 15 members?

Solution.

$$ {}_nC_r = \binom{15}{5} = \frac{15!}{5!10!} = 3003. $$

Example

To select three letters (the order is not important) from 5, $\{A, B, C, D, E\}$, one has ${}_5P_3 = \frac{5!}{2!} = 60$ different ways of permutations; within these permutations, each group of 3 letters appears EXACTLY ${}_3P_3 = 3! = 6$ times.

ABC	ABD	ABE	ACD	ACE	ADE	BCD	BCE	BDE	CDE
ACB	ADB	AEB	ADC	AEC	AED	BDC	BEC	BED	CED
CAB	BDA	BEA	CDA	CEA	DEA	CDB	CEB	DEB	DEC
CBA	BAD	BAE	CAD	CAE	DAE	CBD	CBE	DBE	DCE
BAC	DAB	EAB	DAC	EAC	EAD	DBC	EBC	EBD	ECD
BCA	DBA	EBA	DCA	ECA	EDA	DCB	ECB	EDB	EDC

$$ {}_5C_3 = \frac{{}_5P_3}{{}_3P_3} = \frac{5!}{(5-3)!} \times \frac{1}{3!} = \frac{5!}{(5-3)!3!} $$

In general, we have for $r \leq n$,

$$ {}_nC_r = \binom{n}{r} = \frac{{}_nP_r}{{}_rP_r} = \frac{n!}{(n-r)!r!}. $$

Example

The Student Council at a certain college has one student representative from each of its five departments: Math, Stats, Physics, Chemistry, and Biology. In how many ways can

- Two members be selected to meet the college President
- A Council President and a vice president be selected
- A Council President and two vice presidents be selected

Solutions:

- $\binom{5}{2} = 10$
- $5 \times 4 = 20$
- $\binom{5}{2} \times 3$, or $5 \times \binom{4}{2} = 30$.

Assignment

Page 18-19: 1.1, 1.12, 1.14

Page 24-25: 1. 1.30, 1.32