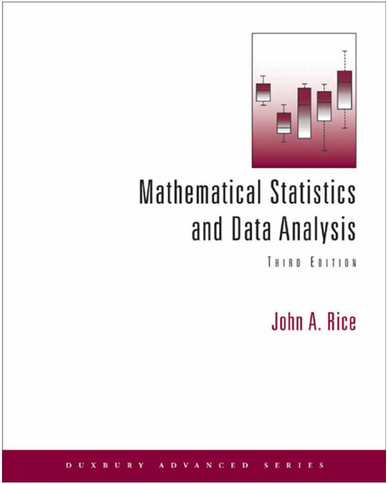
# MH2500 Probability and Introduction to Statistics

## Handout 0 - Course Overview

# *Welcome to MH2500!*

# Course Information

- Textbook: Mathematical Statistics and Data Analysis, *John A. Rice*, 3rd edition, Cengage Learning.



# Course Information

- Lecturer: Chan Song Heng
- Time:
  - Lectures: Mondays 08:30–10:30 in LT23,
    Tuesdays 10:30–11:30 in LT24.
  - Tutorials (Weeks 2–7, 8–13): Thursdays 11:30–12:30, or
    13:30–14:30, or
    16:30–17:30.
- Tutorial participation (weighs 4%)
  - Presenting your solution.

# Course Evaluation

- Test 1 on 30 August*, Lectures up to 29 Aug + Tutorials 1–2.
  Allowed 1 side of an A4-size paper as help sheet.

- Test 2 on 20 September*, Lectures up to 19 Sept + Tutorials 3–5.
  Allowed 1 double-sided A4-size paper as help sheet.

- Test 3 on 18 October*, Lectures up to 17 Oct + Tutorials 6–8.
  Allowed 1 double sided and 1 single-sided A4-size paper as help sheet.

- Final Exam ($\approx$ Chapters 1–7, 9 ).
  Allowed 2 double-sided A4-size papers as help sheet.

- Components of evaluation:
  - 4% Tutorial participation,
  - 17% $\times 3$ Tests 1, 2, 3,
  - 45% Final exam.

Expect the tests to be relatively easy and the Exam to be more difficult.

# About MH2500: Probability and introduction to statistics

Goals

- A first course in Probability with some introduction to statistics.
- Appreciate the subject.
- Preparation for future Statistics courses.

Prerequisites:

- {MTH112, MTH113} or { MH1100, MH1101 }, or { MH1800, MH1801 }, or { MH1101, MH110S }, or { MH1100, MH111S }.

# MH2500 Probability and Introduction to Statistics

## Handout 1 - Probability

# Synopsis

Introduction to Probability, the terminologies, and examples.

- Sample space, events.
- Probability measure.
- Counting methods.
- Conditional probability.
- Independent events.

# Probability and Randomness

Probability has wide applications, for example, in genetics, in designing and analyzing computer operating systems, actuarial science, theory of finance.

Probability theory applies only in situations where the outcomes occur randomly.

Which of the following do probability theory apply?

- Toss a coin and whether head or tail turns up.
- The sex of a baby.
- Will I reach LT27 on time?
- Do I like MH2500?
- Will I pass MH2500?

# Sample space

Probability theory applies only in situations where the outcomes occur randomly.

Such situations are called *experiments*, and the set of all possible outcomes is the **sample space** corresponding to an experiment.

The sample space is denoted by $\Omega$ and an element of $\Omega$ is denoted by $\omega$.

We are often interested in particular subsets of $\Omega$ and these are called **events**.

# Example 1

Experiment: Roll two dice and add the sum of the two numbers on the dice.

The sample space is

$$\Omega = \{ \qquad\qquad\qquad\qquad\qquad\qquad \}.$$

Let $A$ be the event that an odd is obtained. Then

$$A = \{ \qquad\qquad\qquad \}.$$

We can also have other events. E.g.

$$B = \{ \qquad\qquad\qquad\qquad \} \qquad\qquad C = \{ \qquad\qquad\qquad\qquad \}.$$

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Note in this experiment, the sample space is finite.

# Example 2

Experiment: Flip a coin until the first head appears.

Let $h$ denotes head and $t$ denotes tail. The sample space is

$$\Omega = \{h, th, tth, ttth, tttth, \ldots\}.$$

Let $A$ be the event that there were at least 3 flips. Then

$$A = \{ \qquad\qquad\qquad\qquad \}.$$

Let $B$ be the event that there were less than 5 flips. Then

$$B = \{ \qquad\qquad\qquad \}.$$

Then

$$A^c = \{ \qquad \} \qquad A \cap B = \{ \qquad\qquad \} \qquad A \cup B =$$

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Note in this experiment, the sample space is infinite.

# Example 3

The sample space can also be continuous.

Experiment: Randomly choose a number between 2 to 5, both numbers inclusive.

Then the sample space,

$$\Omega = \{x \in \mathbb{R} | 2 \leq x \leq 5\} = [2, 5].$$

# Set identities

Recall $\varnothing$ denotes the empty set, and when $A \cap B = \varnothing$ we say that $A$ and $B$ are disjoint.

Recall some of the set identities:

- Commutative laws:

$$A \cup B = B \cup A \qquad\qquad A \cap B = B \cap A.$$

- Associative laws:

$$(A \cup B) \cup C = A \cup (B \cup C) \qquad\qquad (A \cap B) \cap C = A \cap (B \cap C).$$

- Distributive laws:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$
$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C).$$

- etc ... (revise MH1300 if necessary).

# Probability Measures

**Definition**

A **probability measure** on $\Omega$ is a function $P$ from subsets of $\Omega$ to the real numbers that satisfies the following axioms:

1. $P(\Omega) = 1$.
2. If $A \subseteq \Omega$, then $P(A) \geq 0$.
3. If $A_1$ and $A_2$ are disjoint, then

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

   More generally, if $A_1, A_2, \dots A_n, \dots$ are mutually disjoint, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

# Probability measure

Note:

- Point 1 means the "sum of all possible outcome" has probability 1
- Point 2 means probabilities are nonnegative.
- Point 3 means that if two events $A$ and $B$ are disjoint, then the $P(A \cup B)$ is just the sum of $P(A)$ and $P(B)$. This is true even when extended to the limit, i.e., infinitely many mutually disjoint events and the limit of the partial sum.
- Informally, probability measure gives an indication of how likely a certain event occurs, using a score between 0 to 1.

# Properties

**Property A**

For any $A \subseteq \Omega$, $$P(A^c) = 1 - P(A).$$

**Proof:** Clearly $A$ and $A^c$ are disjoint and $A \cup A^c = \Omega$.
Hence

.

Hence $P(A^c) = 1 - P(A)$.

Property B

$$P(\varnothing) = 0.$$

**Proof:** Since $\varnothing = \Omega^c$, by Property A,

$$P(\varnothing) = P(\Omega^c) = 1 - P(\Omega) = 1 - 1 = 0.$$

..........................................................................

Property C

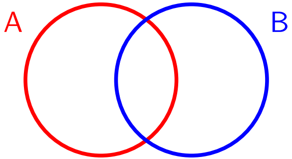
If $A \subseteq B$, then $P(A) \leq P(B)$.

**Proof:** Since $A \subseteq B$, we may express $B$ as a disjoint union of $A$ and $B \cap A^c$. Hence

$$P(B) = P(A \cup (B \cap A^c)) = P(A) + P(B \cap A^c) \geq P(A).$$

## Property D

Addition law:
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$




**Proof:** First, note that we may express $B$ as a disjoint union of $B \cap A^c$ and $A \cap B$. Hence
$$P(B) = P(B \cap A^c) + P(A \cap B),$$
or equivalently, $P(B \cap A^c) = P(B) - P(A \cap B)$.

Next, note that $A \cup B$ is a disjoint union of $A$ and $B \cap A^c$. Hence
$$P(A \cup B) = P(A) + P(B \cap A^c) = P(A) + P(B) - P(A \cap B).$$

# Example

Suppose a fair coin is flipped three times. Let $A$ denote the event that at least two heads were obtained and let $B$ denote the event that the second flip is a tail. Find the probability of $A \cup B$.

**Solution:**

Let $h$ denote head, and $t$ denote tail. The sample space is

$$\Omega = \{ttt, tth, tht, thh, htt, hth, hht, hhh\}$$

# Computing Probabilities

For finite sample spaces, suppose

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\} \qquad \text{and} P(\omega_i) = p_i,$$

Then for any event $A$,

$$P(A) = \sum_{\omega_i \in A} P(\omega_i).$$

If further that every $\omega_i$ has the same probability, i.e.,

$$P(\omega_i) = \frac{1}{N},$$

and $A \subseteq \Omega$ contains $n$ elements. Then

$$P(A) = \frac{n}{N} = \frac{\text{number of ways } A \text{ can occur}}{\text{total number of outcomes}}.$$

# Example: Simpson's Paradox

Suppose Box 1 contains 5 red and 6 blue balls, and Bag 1 contains 3 red and 4 blue balls. Your aim is to pick a red ball and you may either choose to pick a ball randomly from Box 1, or from Bag 1. Would you choose to pick a ball randomly from Box 1 or from Bag 1?

Suppose Box 2 contains 6 red and 3 blue balls, and Bag 2 contains 9 red and 5 blue balls. Again, your aim is to pick a red ball and you may either choose to pick a ball randomly from Box 2, or from Bag 2. Would you choose to pick a ball randomly from Box 2 or from Bag 2?

Now, the balls in Box 2 are added into Box 1, and likewise, the balls in Bag 2 are added into Bag 1. This time, would you choose to pick a ball randomly from Box 1 or from Bag 1?

# The Multiplication Principle

**Multiplication Principle**

If one experiment has $m$ outcomes and another experiment has $n$ outcomes, then there are $mn$ possible outcomes for the two experiments.

**Proof:**
Suppose the first experiment has outcomes $a_1, a_2, \ldots, a_m$ and the second experiment has outcomes $b_1, b_2, \ldots, b_n$. Then the outcomes for the two experiments are

$$(a_i, b_j), \qquad 1 \le i \le m, \quad 1 \le j \le n.$$

# The Extended Multiplication Principle

**Extended Multiplication Principle**

If there are $p$ experiments with the first experiment having $n_1$ outcomes, the second experiment having $n_2$ outcomes, ..., the $p$-th experiment having $n_p$ outcomes, then there are $n_1 n_2 \cdots n_p$ possible outcomes for the $p$ experiments.

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Example: The number of ways to choose one boy and one girl out of a class of 120 boys and 180 girls is $120x180 =$

Question: How many ways can we choose $r$ objects from $n$ objects? To answer this question, We need to know (1) does ordering matter? and (2) are we allowed to duplicate the objects?

A **permutation** is an ordered arrangement of objects.
A **combination** is a collection of objects where ordering does not matter.

E.g. 123, 213 are two different permutations of $\{1, 2, 3\}$ but the same combination.

If no duplication of objects is allowed, we are **sampling without replacement**.

If duplication of objects is allowed, we are **sampling with replacement**.

E.g. Some 3 digits numbers formed from the set $\{1, 2, 3, 4, 5\}$ without replacements are

$$123, \qquad 135, \qquad 234.$$

Note a number like 112 is not allowed.

# Permutations

**Proposition 1**

For a set of size $n$ and a sample of size $r$, there are $n^r$ different ordered samples with replacement and $n(n-1)(n-2)\cdots(n-r+1)$ different ordered samples without replacement.

**Corollary 2**

The number of orderings of $n$ elements is $n(n-1)(n-2)\cdots 1 = n!$.

**Proof:** Induction. (DIY)

## Example E: Birthday Problem

Suppose that a room contains $n$ people. What is the probability that at least two of them have a common birthday?
What are the probabilities if $n = 23$?  28?  70?

**Solution:**

Let $A$ denote the event that at least two people have a common birthday. Then $A^c$ is the event that no two people have a common birthday, i.e., everyone have a different birthday. Then

$$P(A^c) = \qquad\qquad\qquad\qquad .$$

Hence

$$P(A) = 1 - P(A^c) = 1 - \qquad\qquad\qquad\qquad .$$

| n | 23 | 28 | 70 |
|---|---|---|---|
| P(A) | 0.5073 | 0.6545 | .999 |

# Example F

How many people must you ask to have a 50:50 chance of finding someone who shares your birthday?

**Solution:**

Suppose you ask $n$ people and let $A$ be the event that you find someone whose birthday is the same as yours. Then

$$P(A^c) =$$

# Binomial

Proposition 3

The number of unordered samples of $r$ objects selected from $n$ objects without replacement is

$$\binom{n}{r} = \frac{n!}{(n-r)!r!}$$

They are called the **binomial coefficients** as they are the coefficients in the expansion of

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

In particular,

$$2^k = \sum_{k=0}^{n} \binom{n}{k}.$$

(Note: For combinations with replacements, this is not so straightforward.)

# Example I

Capture/Recapture method to estimate the size of wildlife population.

Suppose 10 animals are captured, tagged, and then released. On a later occasion, 20 animals are captured and it is found that 4 of them are tagged. Estimate the size of the population.

**Solution:**

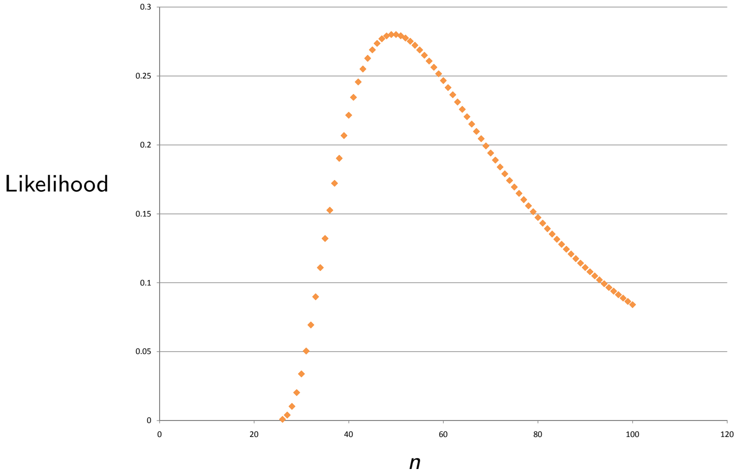
Suppose there are $n$ animals, and amongst them, 10 are tagged while $n - 10$ are not tagged. We assume that any animal has an equal chance of being captured (big assumption).

Then the probability of capturing 20 animals, of which 4 are tagged is

.

This alone doesn't determine $n$. We use the method of **maximum likelihood** to estimate $n$.

# Example I con't




The likelihood is maximized at $n =$

# Multinomial

**Proposition 4**

The number of ways $n$ objects can be grouped into $r$ classes with $n_i$ in the $i$-th class, $i = 1, 2, \ldots, r$, and $\sum_{i=1}^{r} n_i = n$, is

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

The numbers $\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}$ are called **multinomial coefficients** as they occur as the coefficients of

$$(x_1 + x_2 + \cdots x_r)^n = \sum_{\substack{n_1, n_2, \ldots, n_r = 0 \\ n_1 + n_2 + \cdots n_r = n}}^{n} \binom{n}{n_1, n_2, \ldots, n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}.$$

# Multinomial

**Proof of Proposition 4:** This follows from Proposition 3, the multiplication principle, and induction.

There are $\binom{n}{n_1}$ ways to choose $n_1$ objects from $n$ objects and we are left with $n - n_1$ objects. From these $n - n_1$ objects, we choose $n_2$ objects, and there are $\binom{n-n_1}{n_2}$ ways to do that. We continue until the $n_r$ objects are chosen. By the multiplication principle, there are

$$\begin{aligned}\binom{n}{n_1, n_2, \ldots, n_r} &= \binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-\cdots-n_{r-1}}{n_r} \\ &= \frac{n!}{n_1!(n-n_1)!}\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}\times\cdots \\ &\quad\times\frac{(n-n_1-\cdots-n_{r-1})!}{n_r!(n-n_1-n_2-\cdots-n_r)!},\end{aligned}$$

which gives the required result after cancellation.

# Example

Find the number of ways to divide a group of 274 students into 11 tutorial groups of size 23 and 1 group of size 21.

**Solution:**
The number of ways is

# Conditional Probability

**Definition**

Let $A$ and $B$ be two events with $P(B) \neq 0$. The conditional probability of $A$ given $B$ is defined to be

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

**Multiplication law**

Let $A$ and $B$ be events and assume $P(B) \neq 0$. Then

$$P(A \cap B) = P(A|B)P(B).$$

# Example

A fair coin is flipped three times. Find the probability of obtaining exactly two head given that the second flip is a head.

**Solution:**

## Law of Total Probability

Let $B_1, B_2, \ldots B_n$ be such that $\bigcup_{i=1}^{n} B_i = \Omega$ and $B_i \cap B_j = \varnothing$ for $i \neq j$ with $P(B_i) > 0$ for all $i$. Then for any event $A$,

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

**Proof:** $P(A) = P(A \cap \Omega)$

$$= P\left(A \cap \bigcup_{i=1}^{n} B_i\right) \qquad \text{(by} \qquad\qquad\qquad \text{)}$$

$$= \qquad\qquad\qquad \text{(by} \qquad\qquad \text{)}$$

$$= \qquad\qquad\qquad \text{(since} \qquad\qquad\qquad\qquad \text{)}$$

$$= \qquad\qquad\qquad \text{(by} \qquad\qquad \text{)}$$

# Example

An urn contains three red balls, two blue balls, and two white balls. Two balls are selected without replacement. What is the probability that the second ball selected is red?

**Solution:**
Let $R_1$ denote the first ball selected is red, and $R_2$ denote the second ball selected is red. Then

$$P(R_2) = P(R_2|R_1)P(R_1) + P(R_2|R_1^c)P(R_1^c)$$

# Bayes' Rule

**Bayes' Rule**

Let $A$ and $B_1, B_2, \ldots, B_n$ be events where the $B_i$'s are disjoint, $\bigcup_{i=1}^n B_i = \Omega$, and $P(B_i) > 0$ for all $i$. Then

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^n P(A|B_i)P(B_i)}.$$

**Proof:**

By definition of conditional probability, the numerator is

$$P(A|B_j)P(B_j) = \qquad\qquad .$$

By the law of total probability, the denominator is

$$\sum_{i=1}^n P(A|B_i)P(B_i) = \qquad .$$

Hence the right side is .

# Example

Suppose an urn contains some red balls and some white balls. Each ball is also labelled with the number 1, 2, or 3. One ball is drawn from the urn. Let $R, W$ denote the event that the ball is red and white respectively, and let $N_i$ denote the event that the ball drawn is labelled with the number $i$. Given the following probabilities, find $P(N_1|R)$.

| $i$ | $P(N_i)$ | $P(R\|N_i)$ | $P(W\|N_i)$ |
|---|---|---|---|
| 1 | 0.2 | 0.2 | 0.8 |
| 2 | 0.5 | 0.2 | 0.8 |
| 3 | 0.3 | 0.6 | 0.4 |

**Solution:**

# Independence

**Definition**

Two events, $A$ and $B$, are said to be **independent** if

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

**Immediate consequences**

If $A$ and $B$ are independent and $P(B) \neq 0$ then

$$P(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Question. In the example on the previous slide, are $N_1$ and $R$ independent?

# Example C

A fair coin is tossed twice. Let $A$ denote the event of heads on the first toss, $B$ the event of heads on the second toss, and $C$ the event that exactly one head is obtained. Clearly $A$ and $B$ are independent events.
(i) Show that $A$ and $C$ are independent.
(ii) Must $P(A \cap B \cap C) = P(A)P(B)P(C)$?

# Defintion

**Definition**

Events $A_1, A_2, \ldots A_n$ are **mutually independent** if for any subcollection $A_{j_1}, A_{j_2}, \ldots, A_{j_n}$,

$$P(A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_n}) = P(A_{j_1})P(A_{j_2}) \cdots P(A_{j_n}).$$

# Ex 1.8.73

A system has $n$ independent units, each of which fails with probability $p$. The system fails only if $k$ or more of the units fail. What is the probability that the system fails?

**Solution:**

The systems fails if $k$ or more fails, so it is given by

# Remarks

What does $P(A) = 0$ mean? How about $P(A) = 1$?