Instructor and TA’s (weekly consultation location and hours, by appointments only)

Denis HY LEUNG SOE 5047

email: denisleung@smu.edu.sg
phone: 68280396
Tues 4–7pm

TA’s

Mon 333-630pm SOE/SOSS Study Rm 3-13 (Rm 3037) ANG Hui Qi
huiqi.ang.2016@economics.smu.edu.sg

Tue 330-630pm SOE/SOSS Study Rm 3-13 (Rm 3037) Samuel Joel CHIANG
Yuxiang
samueljc.2016@economics.smu.edu.sg

Tue 1215-315pm SOE/SOSS Grp Study Rm 2-5 (Rm 2010) YANG Zhi Yao
zhiyao.yang.2015@accountancy.smu.edu.sg

Wed 12-3pm SOE/SOSS Grp Study Rm 2-5 (Rm 2010) KOH Xue Qi
xueqi.koh.2016@business.smu.edu.sg
Essentials

- Name card (first few weeks)

- Course webpage: http://economics.smu.edu.sg/faculty/profile/9699/Denis%20LEUNG (NOT eLearn!)

- Understanding of basic Calculus and Algebra – Appendix in course notes

- Readings *before* each class

- Projects vs Homework

- Do not disturb others in class

- If you missed a class, it is YOUR responsibility to find out what you have missed from your classmates or course webpage
Assessments

- Class Participation (10%)

- Projects (50%)
  - 2 projects with presentation 25% each
  - Each project’s grade includes 13% individual assessment (quizzes)

- Exam (40%)
  - Closed book but one 2-sided A-4 “cheat sheet” is allowed
Structure of each class

- New materials (< 2.5 hrs)
- Break (15 mins)
- Any other business (15 mins)
Data

- Survival time (in years) in 70 cancer patients

0.67 0.01 0.48 1.06 0.85 0.45 0.19 0.78 1.23 0.18
0.37 2.17 0.15 1.19 0.58 1.22 0.72 0.67 1.42 0.02
0.12 0.50 0.15 0.81 0.64 0.22 0.05 0.55 0.46 0.83
0.46 0.56 0.82 0.07 2.28 0.34 0.64 0.09 0.77 0.26
0.45 0.41 0.23 0.16 0.39 0.29 0.62 1.09 0.14 0.49
0.66 0.89 0.99 0.98 0.95 0.14 0.03 0.01 2.62 0.99
0.08 1.39 1.31 0.50 0.74 1.19 0.15 0.14 1.18 1.53

Orange - Subtype I   Black - Subtype II

- Outcome (H vs. T) in tossing a coin 10 times

H H T H T H T H H

- Occurrence of financial crises

82  84  87  91  97  98  00  07  12
Mexican  S&L  Black Mon. Comm.  RE  AsianLTCM  Dotcom  Subprime  Euro

?
Sample vs. Population

(a) Data are a **sample** from a **population** that we want to study

*e.g.*, Survival time of 70 patients (sample) out of all cancer patients (population)

(b) We are interested in some **characteristics** of the **population**

*e.g.*, average survival time of patients or percentage of patients who live beyond 2 years

(c) Due to **randomness**, we **cannot** make definitive statements about a **particular unit** in the population

*e.g.*, We cannot tell how long the *next* patient would live

(d) We use the sample of data to help us answer the questions in (b)
Observation from survival time data - Why do some patients live longer than others?

- **Subtype I** - 1.06 2.17 1.22 1.42 2.28 2.62 1.39 1.18 1.53
  average $\approx 1.65$

- **Subtype II** - 0.67 0.01 0.48 0.85 0.45 0.19 0.78 1.23 0.18 0.37 0.15 1.19 0.58 0.72 0.67 0.02 0.12 0.50 0.15 0.81 0.64 0.22 0.05 0.55 0.46 0.83 0.46 0.56 0.82 0.07 0.34 0.64 0.09 0.77 0.26 0.45 0.41 0.23 0.16 0.39 0.29 0.62 1.09 0.14 0.49 0.66 0.89 0.99 0.98 0.95 0.14 0.03 0.01 0.99 0.08 1.31 0.50 0.74 1.19 0.15 0.14
  average $\approx 0.50$

- I has a better **chance** of survival because it has a higher average than II
- There are still variations within each subtype - inevitable in the real world
- We attribute (accommodate) these (unexplained) variations as **random**
- Chance and randomness can be described using **probability** theory
A fair coin has a $\frac{1}{2}$ “probability” of observing heads, what does it mean?

Toss 1 2 3 4 5 6 7 8 9 ...  
Outcome H T H T T H T H T ... 

The long run proportion (frequency) of heads is the **probability** of heads.
Probability is the long run frequency of an outcome.

Probability cannot predict individual outcomes.

However, it can be used to predict long run trends.

Probability always lies between 0 and 1, with a value closer to 1 meaning a higher frequency of occurrence.

Probability is numeric in value so we can use it to:

- compare the relative chance between different outcomes (events)
- carry out calculations.
Probability Axioms - Urn model (1)- drawing marbles from an urn (with replacement)
Five possible **Outcomes**: 1  2  3  4  5

Interested in **Event** $A$: ⬜

$A = \{1, 4, 5\}$; hence an event is a collection of outcomes

$$P(A) = \frac{3}{5} = 0.6(60\%) = \frac{\text{Number of marbles in } A}{\text{Total number of marbles}}$$
Complementary events

Marbles in urn: 1 2 3 4 5

Interested in $\bar{A}$: 3 (Not $A$)

$\bar{A} = \{2, 3\}$

$\bar{A}$, sometimes written as $A^C$, is called the **complementary** event of $A$

Chance of 3 = 1 – chance of 2

$\Rightarrow P(\bar{A}) = 1 - P(A) = 1 - \frac{3}{5} = \frac{2}{5}$
Joint probability - independent events - Drawing a blue in draw 1 and a green in draw 2 (with replacement)

*A and B are independent* means the occurrence of one event does not change the chance of the other.

\[
P(A) = \frac{3}{5}; \quad P(B) = \frac{2}{5}
\]

\[
P(A \text{ and } B) = \frac{3}{5} \times \frac{2}{5} = \frac{6}{25} = P(A)P(B)
\]

**Draws**

\[
\begin{array}{c|c}
1 & 2 \\
\hline
\text{●} & \text{●}
\end{array}
\]
Two events $A$ and $B$ are **disjoint** or sometimes called **mutually exclusive** if they cannot occur simultaneously: $P(A \text{ and } B) = 0$

**Example**

$A = \{\text{blue in 1st draw}\}$

$B = \{\text{green in 1st draw}\}$

$P(A \text{ and } B) = 0$
The partition rule is given by:

\[ P(\text{●}) = P(\text{● \ and \ odd}) + P(\text{● \ and \ even}) \]

\[ = P(\{1, 5\}) + P(\{4\}) \]

\[ = \frac{2}{5} + \frac{1}{5} \]

\[ = \frac{3}{5} \]
Conditional probability

**Conditional probability** is a useful quantification of how the assessment of chance changed due to new information: “If \( A \) happened, what is the chance of \( B \)?”

The conditional probability of “\( B \) given \( A \)” is written as \( P(B|A) \)

**Example** Drawing marbles WITHOUT replacement

\[
A = \{ \text{in 1st draw} \} \\
B = \{ \text{in 2nd draw} \}
\]

\[
P(B|A) = \frac{2}{4} = \frac{1}{2}
\]

\[
P(A \text{ and } B) = \frac{2}{4} \times \frac{3}{5} = P(B|A)P(A) = \frac{6}{20} \neq P(B)P(A)
\]
The multiplication rule

- \( P(AB) = P(B|A)P(A) = P(A|B)P(B) \) “Multiplication Rule”

Example

Marbles in urn: 1 2 3 4 5

\[
P(\text{and Odd}) = P(\text{Odd}|\text{Odd})P(\text{Odd}) = \left( \frac{1}{3} \right) \left( \frac{3}{5} \right) = \frac{1}{5}
\]

\[
= P(\text{Odd}|\text{Odd})P(\text{Odd}) = \left( \frac{1}{2} \right) \left( \frac{2}{5} \right) = \frac{1}{5}
\]

- Rearranging the multiplication rule:

\[
P(A|B) = \frac{P(AB)}{P(B)} \quad \text{and} \quad P(B|A) = \frac{P(AB)}{P(A)}
\]
Conditional probability and independence

If $A$ and $B$ are independent, conditional probability becomes unconditional, i.e.,

\[
\begin{align*}
P(A|B) &= \frac{P(AB)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A) \\
P(B|A) &= P(B)
\end{align*}
\]

Independence is NOT the same as mutually exclusive (disjoint), which is $P(A \text{ and } B) = 0$. In fact when $A$ and $B$ are disjoint, they are very dependent.
Union of events (1)

**Union** of events can sometimes be best visualized using a **Venn diagram** (John Venn, 1834-1923)

**Example** What is the probability of drawing a **green** or an odd number?
Union of events (2)

\[
\begin{align*}
P(\text{odd or odd}) &= P(\text{odd}) + P(\text{Odd}) - P(\text{odd and odd}) \\
&= \frac{2}{5} + \frac{3}{5} - \frac{1}{5} \\
&= \frac{4}{5} \\
&= \frac{4}{5}
\end{align*}
\]

In general, if \( A \) and \( B \) are:

- disjoint, then \( P(A \text{ or } B) = P(A) + P(B) \)
- not disjoint, then \( P(A \text{ or } B) = P(A) + P(B) - P(AB) \)
**Probability tree** is useful for studying combinations of events. Branches of a tree are *conditional* probabilities.

**Example**
Drawing two marbles from urn without replacement:

![Probability Tree Diagram]

\[
P(\bigcap) = \frac{3}{5} \cdot \frac{2}{4} = P(\bullet|\bigcap)
\]

\[
P(\bigcap) = \frac{2}{5} \cdot \frac{3}{4} = P(\bigcap|\bullet)
\]

\[
P(\bigcap) = \frac{3}{5} \cdot \frac{2}{4} = P(\bigcap|\bullet)
\]
Bayes Theorem (Thomas Bayes, 1701-1761)

Trees are useful for visualizing $P(B|A)$ when $B$ follows from $A$ in a natural (time) order. Many problems require $P(A|B)$, **Bayes Theorem** provides an answer.

Example

Testing for an infectious disease.

What is $P(D|T)$ or $P(\bar{D}|\bar{T})$?
Bayes Theorem (2)

\[ P(D|T) = \frac{P(D \cap T)}{P(T)} = \frac{P(T|D)P(D)}{P(T)} \]

\[ = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\bar{D})P(\bar{D})} \]

Particle rule

\[ = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\bar{D})P(\bar{D})} \]

Multiplication rule

In general,

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]
Data revisited - Discrete vs. Continuous Variables

- A **variable** is a characteristic of interest in each unit of the data, *e.g.*, 
  - Survival time of each patient
  - Outcome of each coin toss
  - No. of crises in a decade

- **Discrete** - countable number of possible values, *e.g.*, 
    - Two possible values: H or T
    - Many possible values: 0, 1, 2, 3, 4,...

- **Continuous** - values fall in an interval \((a, b)\), \(a\) could be \(-\infty\) and \(b\) could be \(\infty\), *e.g.*, 
  - Survival time in cancer patients: 0.67, 0.01, 0.48, 1.06, 0.85, 0.45, ...
    - \(0 < \text{survival time} < b \leq \infty\)
What do we do about data?

Coin toss data: H, H, T, H, T, H, T, H, H, H, ...

How can we use the data to find out about “?” in a similar coin?

- Suppose
  - $X$ is used to denote any of the unknown outcomes “?”
  - only two possible values for $X$: H vs. T
  - a random (probability) mechanism generates the data and $X$
  - the random mechanism does not change over time, i.e., $P(X = H)$ is identical for every toss

- Use data to find $P(X = H)$ and $P(X = T)$

- In the data, there are 7 Hs and 3 Ts, we may try

<table>
<thead>
<tr>
<th>$X$</th>
<th>H</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X)$</td>
<td>$\frac{7}{10}$</td>
<td>$\frac{3}{10}$</td>
</tr>
</tbody>
</table>

- The table is called a **probability distribution** of $X$. Since $X$ is discrete, it is an example of a **discrete distribution**

† We will learn in class 6 whether these are good choices for $P(X)$
Discrete distributions (1) - Tossing a coin

(H) or (T)?

<table>
<thead>
<tr>
<th></th>
<th>H</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7/10</td>
<td>3/10</td>
</tr>
</tbody>
</table>

The long run frequencies are 7/10 for H and 3/10 for T – there is higher chance for H than T.
The long run frequencies tell us there is equal chance for 1, 2, 3, 4 and 5 but there is a higher chance for blue than green.
### Discrete distributions (3)

<table>
<thead>
<tr>
<th>$X$</th>
<th>$H$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X)$</td>
<td>$\frac{7}{10}$</td>
<td>$\frac{3}{10}$</td>
</tr>
</tbody>
</table>

A probability distribution summarizes the behavior of a random outcome $X$. 

<table>
<thead>
<tr>
<th>$X$</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$a_3$</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X)$</td>
<td>$P(a_1)$</td>
<td>$P(a_2)$</td>
<td>$P(a_3)$</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$X$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X)$</td>
<td>$\frac{1}{5}$</td>
<td>$\frac{1}{5}$</td>
<td>$\frac{1}{5}$</td>
<td>$\frac{1}{5}$</td>
<td>$\frac{1}{5}$</td>
</tr>
</tbody>
</table>
X is the unknown outcome. X is called a **discrete random variable** if its value can only come from a countable number of possible values: \( a_1, a_2, ..., a_k \).

**Examples**

- Coin toss: \( X = H \) or \( T \) (2 possible values)
- Marbles: \( X = 1, 2, 3, 4, \) or \( 5 \) (5 possible values)
- Financial crisis: \( X = 0, 1, 2, 3, ... \) (Infinite but countable number of possible values)

\( P(X = a_i) \) gives the probability \( X = a_i \) and is called a **probability distribution function**.

A valid probability distribution function must satisfy the following rules:

- \( P(X = a_i) \) must be between 0 and 1
- We are certain that one of the values will appear, therefore:

\[
P(X = a_1 \text{ or } X = a_2 \text{ or } ... \text{ or } X = a_k) = P(X = a_1) + P(X = a_2) + ... + P(X = a_k) = 1
\]

\( X=a_1,X=a_2,...\) are disjoint events
Calculating probabilities - Example: drawing a marble

<table>
<thead>
<tr>
<th>X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X)</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

- $P(X = 1) = 0.2$
- $P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5) = 0.6$
- $P(X = 1.5) = 0$
- $P(X > 5) = 0$
- $1 = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$
Survival data: 0.67, 0.01, 0.48, ..., 1.18, 1.53, ...

70 patients

How can we use the data to find out about “?” in similar patients?

Suppose
- $X$ is used to denote “?” survival times of similar patients
- many possible values for $X$ – we only know $X > 0$, i.e., $X \in (0, \infty)$
- a random (probability) mechanism generates the data and $X$
- the random mechanism does not change for different patients, i.e., everyone has the same chance of surviving up to time $x$

Use data to find the probability distribution of $X$

In the data, there are only 70 values, but $X$ for similar patients can have infinitely many possible values – we need a different kind of probability distribution
The survival time \( (X) \) of a cancer patient following diagnosis may be any value in a range (e.g., \( 0 < X < \text{some positive number} \)).

\( X \) is called a **continuous random variable** if its value falls in a range \((a, b) \subseteq (-\infty, \infty)\). For a continuous random variable \( X \), \( \Pr(X = x) \) for any value \( x \) is 0, we can only talk about the probability of \( X \) falling in a range.

**Examples**

- \( \Pr(X > 2 \text{ years}) = \text{Probability of surviving beyond 2 years} \)
- \( \Pr(0 < X \leq 1 \text{ year}) = \text{Probability of dying within 1 year} \)
- \( \Pr(X \leq 2 \text{ years} | X > 1 \text{ year}) = \text{Probability of dying within 2 years given surviving for 1 year} \)
Probability distribution for a continuous random variable

A continuous random variable, $X$, is a random variable with outcome that falls within a range or interval, $(a, b)$

**Examples**

- $X =$ survival time of a patient: $(a, b) = [0, 150]$ years
- $X =$ return from an investment of 10000; $(a, b) = [-10000, \infty)$ dollars
- $X =$ per capita income; $(a, b) = [0, 1000000000000]$ dollars

The probability distribution of $X$ is defined by a (probability) density function (PDF), $f(x)$. There are subtle differences between PDF and probability:

- $f(x) \geq 0$ for any value of $x$ in $(a, b)$
- $f(x) \neq P(X = x) = 0$ for any value of $x$
- $P(r \leq X \leq s) = \underbrace{P(r < X < s)} \quad$ probability of $X$ between $r$ and $s$
  
  since $P(X=r)=P(X=s)=0$
- $P(a \leq X \leq b) = 1$ since $X$ must fall within $(a, b)$

The **cumulative distribution function (CDF)**, written as $F(x)$, is defined as $P(X \leq x)$
Calculating probabilities - Example: survival data

- $f(x) = \lambda e^{-\lambda x}$, $\lambda > 0$ is called an **exponential** density.

- Different values of $\lambda$ can be used for different types of data.

- All exponential densities have the same shape but different gradients so it is easy to describe to others.

- Allows us to make probability statements about survival times we have and *have not* observed.

- The “area” under $f(x)$ is 1 which means the probability of dying between time 0 and $\infty$ is 1.

---

‡‡ We will learn in class 6 whether the choice of $\lambda = 1.5$ is good.
$P(X > 2) = 0.049$ is the red area under the curve and can be obtained by analytical or numerical (computer) analysis.
Some calculations

$f(x) = 1.5e^{-1.5x}$

$P(X \leq 1) = P(0 < X \leq 1) = 0.776$ is the red area under the curve and can be obtained by analytical or numerical analysis.
Some calculations

\[ P(X \leq 2|X > 1) = \frac{P(1 < X \text{ and } X \leq 2)}{P(X > 1)} \]

conditional probability

\[ = \frac{P(1 < X \leq 2)}{P(X > 1)} \]

\[ = \frac{1 - P(X \leq 1) - P(X > 2)}{1 - P(X \leq 1)} \]

\[ = \frac{1 - 0.776 - 0.049}{1 - 0.776} \]

\[ = 0.781 \]