Chapter 0: An Overview

ST2334 Probability and Statistics[1]
(Academic Year 2014/15, Semester 1)

Department of Statistics & Applied Probability (DSAP)
National University of Singapore (NUS)[1]

Outline

1. Module Information

2. A Brief Introduction To Probability

3. A Brief Introduction To Statistics

4. Set–Up Of Statistical Problems

5. Relationship Between Probability And Statistics

Module Information I

Instructors
Dr. Ho Man-Wai

Contact
(65) 6516 3953 (Office); stahmw@nus.edu.sg (email)

Office
Block S16, Level 6, Room 101

Consultation Hours
by appointment through email

Module Information II

Aim and Objective

1. Equip students with some fundamental concepts of probability & a solid foundation in statistical theory, &

2. Provide an indication of relevance & importance of statistical theory in solving practical problems in the real world

Syllabus

Basic concepts of probability, conditional probability, independence, random variables, joint & marginal distributions, mean & variance, some common probability distributions, sampling distributions, estimation & hypothesis testing based on a normal population

Module Information III

Lecture Time & Venue

Every Monday & Thursday (from 11 Aug 2014 to 13 Nov 2014 except 22 & 25 Sep, 6 Oct 2014), *12:00nn – 2:00pm*, at *UTown Auditorium 3* [*Room B1-01* between Cinnamon College (USP) and College of Alice & Peter Tan]

Tutorial Time & Venue

1. Monday, 3:00pm – 3:45pm
2. Monday, 4:00pm – 4:45pm
3. Monday, 5:00pm – 5:45pm
4. Wednesday, 1:00pm – 1:45pm
5. Wednesday, 5:00pm – 5:45pm
6. Wednesday, 6:00pm – 6:45pm

Conducted *at S16-06-118* (*i.e.,* Building S16, Level 6, Room 118) starting from 25 or 27 Sep to 10 or 12 Nov except 22, 24 Sep & 6, 22 Oct

Module Information IV

Compulsory Textbook

Introduction to Probability & Statistics for Engineers and Scientists (4th ed.)

- written by Sheldon Ross
- published by Academic Press
- available for download *here via NUS library*
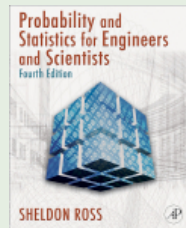
Module webpage

IVLE (https://ivle.nus.edu.sg/) for lecture notes/slides, tutorial problem sheets, suggested solutions, test scores, and so on

Module Information V

Study advice

- *Lecture slides/notes* are to supplement but *NOT to replace the textbook*
- Students are urged to
    - attend most, if not all, lectures & tutorials
    - seriously read through & understand the corresponding materials (especially the examples) in the compulsory textbook after lectures
    - attempt weekly homework problems before attending tutorials, &
    - attempt more problems from the textbook

Assessment

- Mid–semester test (on *2 Oct 2014, Thursday* during lecture): 25%
- End–of–semester examination (on *3 Dec 2014, Wednesday afternoon*): 75%
- Both test & examination are *OPEN BOOK*

Introduction

Learning Outcomes

Questions to Address: ♦ What probability is ♦ What statistics is about ♦ Why we study statistics ♦ Set–up of a statistical problem ♦ Relationship between probability and statistics

Concept & Terminology: ♦ probability/chance/randomness ♦ uncertainty ♦ chance/random phenomenon ♦ probability problem ♦ data ♦ dataset ♦ variable ♦ measurement ♦ observation ♦ statistical problem ♦ experimental unit ♦ population ♦ parameter ♦ sample ♦ statistic

▸ Statistical concepts & methods are not only *useful* but indeed often *indispensable* in understanding the world around us because

Life is full of uncertainty!

What is Probability? I

- The idea of *probability (prob), chance or randomness* goes back over 300 years in the study of games of chance
- The language of probability is constantly used in an informal manner in both written & spoken contexts:
    - It is *likely* that the Dow–Jones average will increase by the end of the year
    - There is a *50–50 chance* that the incumbent will seek reelection
    - The number (#) of students who show up today at UT-AUD 3 is *probably* more than 150
    - The *odds favor* a quick settlement of the strike
    - It is *expected* that at least 20,000 concert tickets will be sold
    - · · ·

What is Probability? II

▸ Under *uncertainty* (*i.e.,* in any situation in which one of a number of possible *outcomes* may occur), the discipline of probability provides methods for quantifying the *chances or likelihoods*, associated with the various outcomes

▸ *The mathematical theory of probability* has been applied to a wide variety of *chance/random phenomena* in the fields of

  ▸ genetics & bioinformatics (mutations)
  ▸ physics (kinetic theory of gases)
  ▸ operations research (demands on the inventories of goods)
  ▸ finance (volatility of a stock)
  ▸ transportation (utilization rate of the road)
  ▸ queuing theory (arrival rate of customers in a counter)
  ▸ · · ·

Example 1: A Probability Problem

A claim: There are > 120 students present in today's lecture

▶ Of course, before noon today, no one can be 100% certain whether or not the claim is true – number (#) of students who show up is a *chance phenomenon*

▶ It depends on
  - ▶ how many students who enrol in the module in this semester
  - ▶ whether an individual student shows up for all the students who enrol in the module (all of them are chance phenomena)

▶ **Provided that** we know the chances of presence of each individual student, one can *compute the probability (prob)* for the "outcome" that there are > 120 students in theory (though this process is demanding & costly in time)
  - ▶ This is called a *probability problem* – *all the characteristics of the random phenomenon are known* & then *probabilities or likelihoods of some possible outcomes of interest are computed*

Example 2: Real Life Examples of Statistical Nature

▸ Women make 75 cents to every dollar a man makes when they work the same job. Yes or no?

▸ How many lung cancer cases are tobacco–related?

▸ People predict that it is very unlikely there will ever be another baseball player with a batting average over 400. Right?

▸ What is the effectiveness of condom? 94%?

▸ Do people tend to be more persuasive when they look others directly in the eye & speak loudly & quickly?

▸ A surprising new study shows that eating egg whites can increase one's lifespan. Is it true? How many more years?

▸ There is an 95% chance that in a room with 100 people that at least people will share the same birthday. How about here?

▸ ⋯

Why Study Statistics? I

1. To know how to *evaluate published numerical facts* (some of them are valid; others are invalid) that bombard you every day

   ▸ manufacturers' claims for products
   ▸ results of sociological, consumer, political polls
   ▸ published results of scientific research
   ▸ ⋯

2. Your profession or employment may require you to interpret the results of *sampling* (surveys or experimentation) or to employ *statistical methods of analysis to make inferences* in your work

   ▸ to forecast sales & profit in business
   ▸ to monitor product quality in engineering & manufacturing
   ▸ to assist accountants in conducting audits through sampling of accounts
   ▸ ⋯

Why Study Statistics? II

▶ Statistics plays an important role in almost all areas of science, business, & industry

▶ Statistics lies at the heart of the type of *quantitative reasoning* necessary for making important advances in the sciences, such as medicine & genetics, & for making important decisions in business & public policy

What is Statistics About?

- The discipline of statistics is the *science* (&, arguably, also the art!) of *learning from data* to help us to arrive at *intelligent judgements & informed decisions in the presence of uncertainty & variation*

- *Data is everywhere!* Data appear in everyday activities, also in a *huge volume*:

  - quarterly sales figures

  - traffic volumes

  - percent increase in juvenile crime

  - survival rates for patients undergoing medical therapy

  - census figures

  - internet usage

Example 3: U.S. Census Form

Data/A Dataset (about Internet Usage)

Age	Sex	Facebook f a/c	Weibo a/c	"Connected" # of Hours/day
22	M	Y	Y	3.5
24	F	Y	N	2
18	M	N	N	0.5
35	M	Y	Y	4
⋮	⋮	⋮	⋮	⋮

▸ Each row corresponds to information of the same entity (in this case: a person) under consideration

▸ Each column contains information about a *variable*

▸ Each cell contains a *measurement/value* about a variable associated with an entity

▸ Measurements in each row constitute an *observation* from an entity

The Discipline of Statistics is Concerned With

Collection of data

Observational studies versus *Experimental designs*

Analysis of data

Selection of *appropriate & well-defined* statistical methodologies

Interpretation of data

Retrieval of *relevant information* to address questions in mind

Effective communication & presentation of results relying on data

Drawing *meaningful conclusions in layman's language*

Set–up of Statistical Problems I

- **The problem** : Of interest is certain characteristic of some entities called a *parameter* (*e.g.,* average weight of students in NUS, & proportion of students in NUS who have a 🅵 a/c)
- A well–defined collection of entities (called *experimental units*) constituting a *population* of interest, which is the source of relevant information/all *measurements* to address the problem
  - When desired information is available for the whole population:
    - we have what is called a *census*
    - any parameter of interest are *deterministic*
- *Data*: ONLY a *subset* (or part) of the designated population, called a *sample*, is available
- *Infer/Guess* the unknown parameter (defined by all entities in the population) based on some *statistics* (computed from the sample – part of, but not all, information in the population)

Set–up of Statistical Problems II
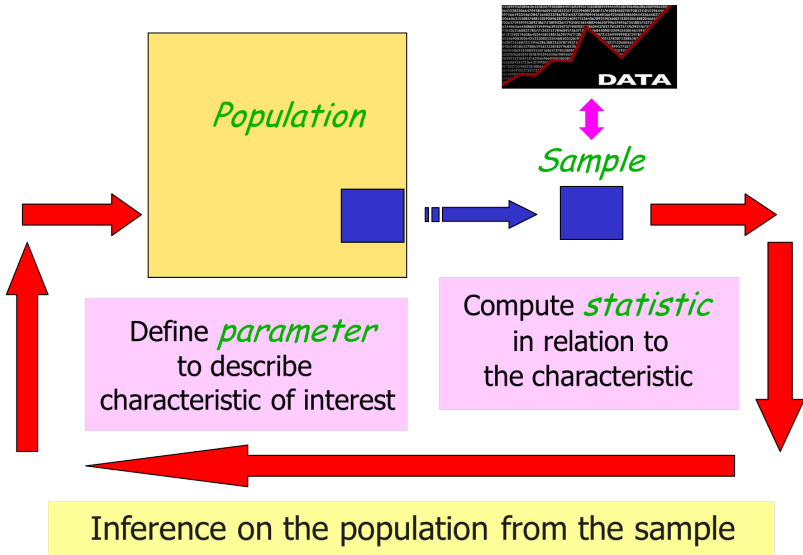
Example 4: *How Popular is Facebook?*

1. *Problem*: Suppose that we are interested in how popular [f] is in Singapore 🇸🇬

2. *Population*: Everyone (👨, 👩) who resides in 🇸🇬 (*i.e., experimental units*)

3. *Parameter*: Proportion of the Singaporean population 🇸🇬 who have a [f] a/c

4. *Sample/data*: 🏛NUS students who are present here at UT AUD 3 in the first lecture of ST2334 today

5. *Statistic*: Proportion of students who have a [f] a/c *among the sample* (all students who are present here at UT AUD 3 in the first lecture of ST2334 today)

Why Learns Probability in Statistics Classes? I

Statistical theory is derived from mathematical theory of probability

▸ Mastery of probability leads to a better understanding of how statistical methodologies are developed & used

▸ A *probability problem* can be *described in terms of population & sample*: View a *chance phenomenon as a population* & *outcome from the chance phenomenon as a sample* ⇔ information of the population under study is known, & questions regarding a sample taken from the population are posed & answered
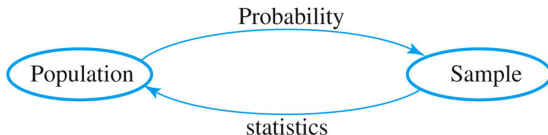
Probability Problem

Chance phenomenon ⇔ Population
⇓
Outcome ⇔ Sample

Why Learns Probability in Statistics Classes? II

▶ The relationship between the two disciplines can be summarized by saying that probability reasons from the population to the sample (*deductive reasoning*), whereas statistics reasons from the sample to the population (*inductive reasoning*)

▶ Before we can retrieve *relevant information about the whole population from any sample*, we should first *understand the uncertainty associated with taking a sample* from a given population

We always study probability before learning statistics!