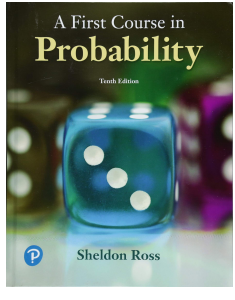
About this course

Textbooks



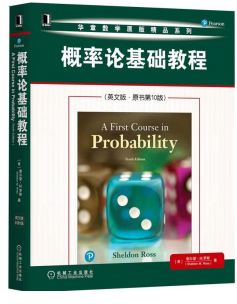
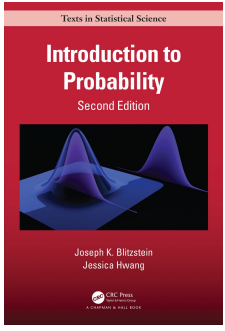
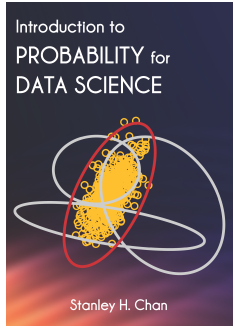
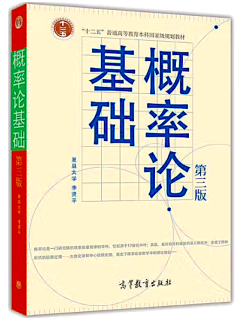


Figure: Textbook: Sheldon Ross, A first course in probability.

Further reading







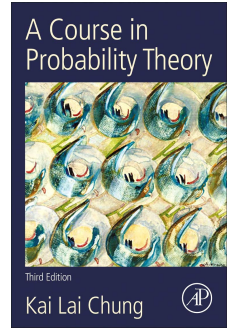


Figure: Further materials

Softwares

In this course, we might use several types of free software:


The LaTeX Project


Overleaf

Lecture 1: Data and Probability

Foundation of Probability Theory/STA 203

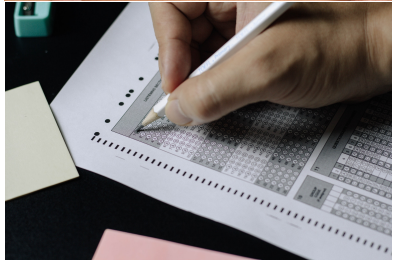
Zhuosong ZHANG

Department of Statistics and Data Science, SUSTech

Fall, 2023

Randomness

Data in our real world



Data in our real world

- We may think Data as any collection of numbers, characters, images, or other items that provide information about something.
- Statisticians are working with data.
- An example of data set.

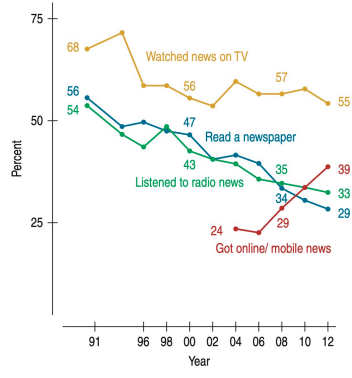
	A	B	C	D	E	F	G	H
1	Purchase ID	Last name	First name	Birthday	Country	Date of purchase	Amount of purchase	
2	1	Davidson	Michael	04/03/1986	United States	10/12/2016	37	
3	2	Vito	Jim	09/01/1994	United Kingdom	02/02/2016	85	
4	3	Johnson	Tom	23/08/1972	France	02/11/2016	83	
5	4	Lewis	Peter	18/10/1979	Germany	22/11/2016	27	
6	5	Koenig	Edward	13/05/1983	Argentina	26/03/2015	43	
7	6	Preston	Jack	16/06/1991	United States	06/11/2016	77	
8	7	Smith	David	11/03/1965	Canada	15/11/2016	23	
9	8	Brown	Luis	03/09/1997	Australia	03/07/2015	74	
10	9	Miller	Thomas	07/01/1980	Germany	07/11/2016	13	
11	10	Williams	Bill	26/07/1960	United States	20/11/2015	80	
12	11	Gemini	Alexia	12/09/1995	Canada	11/03/2017	35	
13	12	Bond	James	25/02/1975	United Kingdom	12/08/2017	40	
14	13	Burgle	Patricia	01/12/1990	United States	18/01/2015	55	
15	14	Reding	Michelle	07/04/1985	Canada	23/02/2017	28	
16	15	Harvey	Billy	14/07/1971	United Kingdom	12/01/2016	41	
17								
18								

Sample and Population

- Target of a statistician: know about an entire group of individuals.

However, the entire group is usually impossible to get access.

A timeplot showing the responses of 18—29 year olds to Pew polls asking them where they get most of their news about national and international news.



Sampling

- Ideas of sampling:
  a. Examine a part of the whole:

     Select a sample from the population for examining.

  b. Randomizing:

     On average, the sample looks like the rest of the population.

  c. Large sample size:

     Sample size is the number of individuals in the sample.

- Simple random sampling.

  The principle of simple random sampling is that every set of items has the same chance of being chosen.

Randomness

- Statistics and Data Science is based on the theory of “randomness”.
- This world is also filled with randomness.

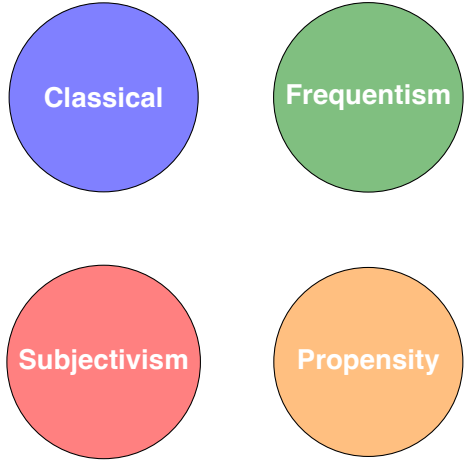
Randomness

- How to understand “randomness”?
- The surprising fact is that in the long run, many truly random phenomena settle down in a way that’s consistent and predictable.
- Random phenomena: we know what outcomes can possibly occur, but we don’t know which particular outcome will happen.
- Trials: each occasion upon which we observe a random phenomenon is called a trial.
- Events: Any Combination of outcomes is called an event.
- Sample Space: The collection of all possible outcomes is called the sample space.

Explanations of probabilities

How likely does an event happen?

We use probability to measure how likely does an event happen.
There are several explanations of the term “probability”:



Classical Probability

- Classical Probability: Pierre-Simon Laplace (拉普拉斯, 1749–1827).

Developed from studies of gambling it states that probability is shared equally between all the possible outcomes, provided these outcomes can be deemed equally likely.

Classical Probability

- If a random experiment can result in N mutually exclusive and equally likely outcomes and if N_A of these outcomes result in the occurrence of the event A, the probability of A is defined by

$$P(A) = \frac{N_A}{N}.$$

- Examples:

Cards

The probability of drawing a face card (JQK) from a deck of playing cards (without Jokers) is

$$P(\text{face card}) = \frac{\#\text{ of face cards}}{\#\text{ of total cards}} = \frac{12}{52} = \frac{3}{13}.$$

Classical Probability

- Limitations:

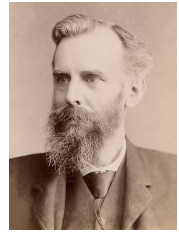
  (a) Only applicable to situations where there is only a "finite" number of possible outcomes.

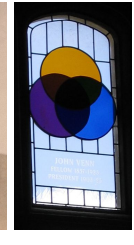
  (b) Require a priori determination that all possible outcomes are equally likely.

Frequency probability

Frequentists posits that the probability of an event is its relative frequency over time (the long-run probability).

John Venn (韦恩, 1834–1923) provided a thorough exposition of frequentist probability in his book, The Logic of Chance (1866):





Probabilities can be found (in principle) by a repeatable objective process (and are thus ideally devoid of opinion).

Frequency probability

Traffic lights

Day	1	2	3	4	5	6	...	100
Light	Green	Red	Green	Green	Red	Red	...	Red
% Green	100	50	66.7	75	60	50	...	33.35

Coin Tossing

In the case of tossing a fair coin, frequentists say that the probability of getting a heads is $1/2$, not because there are two equally likely outcomes but because repeated series of large numbers of trials demonstrate that the empirical frequency converges to the limit $1/2$ as the number of trials goes to infinity.

Frequency probability

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Definition 1 (Frequency probability)

Denote by n_A the number of occurrences of an event A in n trials, then if

$$\lim_{n\to\infty} \frac{n_A}{n} = p$$

we say that $P(A) = p$.

- Limitations:

  (a) It is of course impossible to actually perform an infinity of repetitions of a random experiment to determine the probability of an event.

  (b) If only a finite number of repetitions, different values of probabilities for different trials.

Bayesian probability and Propensity probability

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- Bayesian probability is an interpretation of the concept of probability, in which probability is interpreted as reasonable expectation representing a state of knowledge or as quantification of a personal belief.
- The propensity probability is a probability interpretation in which the probability is thought of as a physical propensity, disposition, or tendency of a given type of situation to yield an outcome of a certain kind, or to yield a long-run relative frequency of such an outcome.
- Will be discussed later.