Statistics for Engineers Lecture 1
Introduction to Probability Theory

Chong Ma

Department of Statistics
University of South Carolina
chongm@email.sc.edu

January 9, 2017

Outline

1. Introduction
   - Deterministic Model
   - Statistical Model
2. Probability
   - Sample Spaces and Events
   - Unions and Intersections
   - Axioms of Probability
   - Conditional Probability
   - Probability Rules
   - Random Variables

What is statistics?

- The art and science of learning from data,i.e., the study of a collection, analysis, interpretation and organization of data. The ultimate goal is to translate data into knowledge and understanding the world around us.
- Partly empirical and partly mathematical involving probability theory, measure theory and other related mathematics. Nowadays statistical is more computational.
- Popular statistical softwares: R, SAS, Python, Julia, Perl . . .
- Broad application: machine learning(Google DeepMind), Biomedical, genetics, econometrics, statistical physics, chemistry, . . .

Examples

1. In a reliability(time to event) study, engineers are interested in describing the time until failure for a jet engine fan blade.
2. In a genetics study involving patients with Prostate cancer, researchers wish to identify genes that are differently expressed compared to non-Prostate cancer patients.
3. In an agricultural experiment, researchers want to know which of four fertilizers varying in their nitrogen contents produces the highest corn yield.
4. In a clinical trial, physicians want to determine which of two drugs is more effective for treating HIV in the early stages of the disease.
5. In a public health study involving "at-risk" teenagers, epidemiologists want to know whether smoking is more common in a particular demographic class.

6. A food scientist is interested in determining how different feeding schedules(for pigs) could affect the spread of salmonella during the slaughtering process.
7. A pharmacist is concerned that administering caffeine to premature babies will increase the incidence of necrotizing enterocolitis.
8. A research dietitian wants to determine if academic achievement is related to body mass index(BMI) among African American students in the fourth grade.

Deterministic Model

Deterministic Model is one that makes no attempt to explain variability. For example,

- In chemistry, the ideal gas law states that

$$PV = nRT$$

  Where p=pressure of a gas, V=volume, n=the amount of substance of gas(number of moles), R=Boltzmann's constant, and T=temperature.

- In circuit analysis, Ohm's law states that

$$V = IR$$

  Where V=voltage, I=current and R=resistant.

Deterministic Model

Remarks:

- In both of these models, the relationship among the variables is completely determined without ambiguity.
- In real life, this is rarely true for the obvious reason: there is natural variation that arises in the measurement process.
- For example, a common electrical engineering experiment involves setting up a simple circuit with a known resistance R. For a given current I, different students will then measure the voltage V.
  - With a sample of 20 students, conducting the experiment in succession, they might very well get 20 different measured voltages.
  - A deterministic model is too simplistic; it does not acknowledge the inherent variability that arises in the measurement process.

Statistical Model

Statistical Model is not deterministic, which incorporates variability and is used to predict future outcomes. Suppose that I am interested in predicting

$$Y = \text{STAT 509 final course percentage}$$

by using x_1 = SAT MATH score and x_2 = MATH 141 grade. The statistical model can be formated as

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

Where ϵ is a term that accounts for not only measurement errors(e.g., incorrect information, data entry errors, grading errors, etc.) but also

- all of the other variables not accounted for(e.g., majoy, study habits, natural ability, etc.) and
- the error induced by assuming a **linear relationship** between Y and x_1 and x_2 when the relationship might actually not be.

Statistical Model

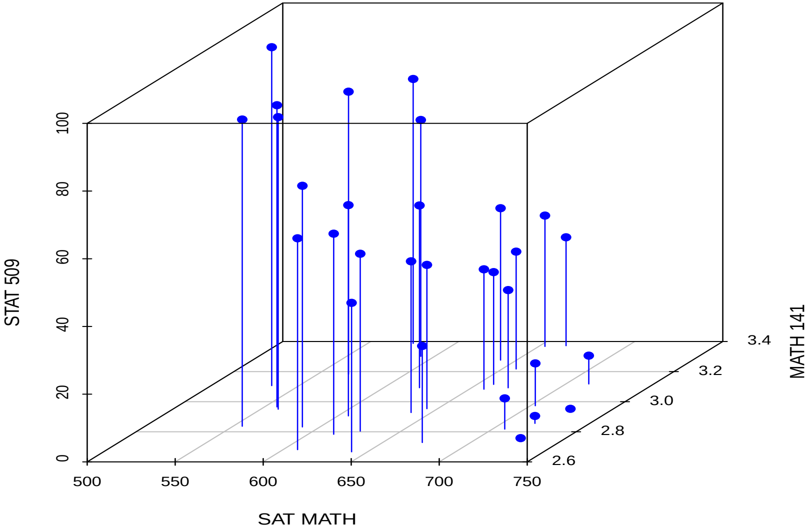


Figure 1: 3-D scatterplot of historical data

Remarks:

- Is this sample of students representative of some larger relevant population? After all, we would like our model to be useful on a larger scale.
- How should we estimate β_0, β_1 and β_2 in the model?
  - If we can do this, then we can produce predictions of Y on a student-by-student basis.
  - This maybe of interest to academic advisers who are trying to model the success of their incoming students.
  - We can also characterize numerical uncertainty with our predictions.
- **Probability** is the "mathematics of uncertainty" and forms the basis for all of statistics.

Outline

1. Introduction
   - Deterministic Model
   - Statistical Model
2. Probability
   - Sample Spaces and Events
   - Unions and Intersections
   - Axioms of Probability
   - Conditional Probability
   - Probability Rules
   - Random Variables

Sample Spaces and Events

Probability is a measure of one's belief in the occurrence of a future event. Here are some events to which we may wish to assign a probability.

- tomorrow's temperature belows 50 degrees
- manufacturing a defective part
- concluding one fertilizer is superior to another when it isn't
- the NASDAQ losing 5 percent of its value
- you being diagnosed with prostate/cervical cancer in next 30 years.

Sample Space is the set of all possible outcomes for a given random experiment, denoted by S. The total number in S is denoted by n_S.
Event is a subset of the sample space, denoted by capital letters such as $A, B, C, \ldots$. The total number in S is denoted by n_S.

Sample Spaces and Events

Suppose a sample space S contains finite outcomes $n_S < \infty$, each of which is equally likely. This is called an **equiprobability model**. If an event A contains n_A outcomes, then

$$P(A) = \frac{n_A}{n_S}$$

Examples

(a) The Michigan state lottery calls for a three-digit integer to be selected:

$$S = \{001, 002, \dots, 999\}$$

Let event $A =$ winning number is a multiple of 5. Thus, we have $n_S = 1000$, $n_A = 200$ and

$$P(A) = \frac{n_A}{n_S} = \frac{200}{1000} = 0.2$$

(b) A USC undergraduate student is tested for chlamydia(0=negative and 1=positive). Thus,

$$S = \{0, 1\}$$

Here $n_S = 2$ possible outcomes. However, is it reasonable to assume that each outcome in S is equally likely?

- The prevalence of chlamydia among college age students is much less than 50%.
- It would be illogical to assign probabilities using equiprobability model.

(c) Four equally qualified applicants(a,b,c,d) are competing for two positions. If the positions are identical(so that selection order doesn't matter). Let A be the event that applicant d is selected for one of the two positions. Thus,

$$S = \{ab, ac, ad, bc, bd, cd\}, A = \{ad, bd, cd\}$$

And

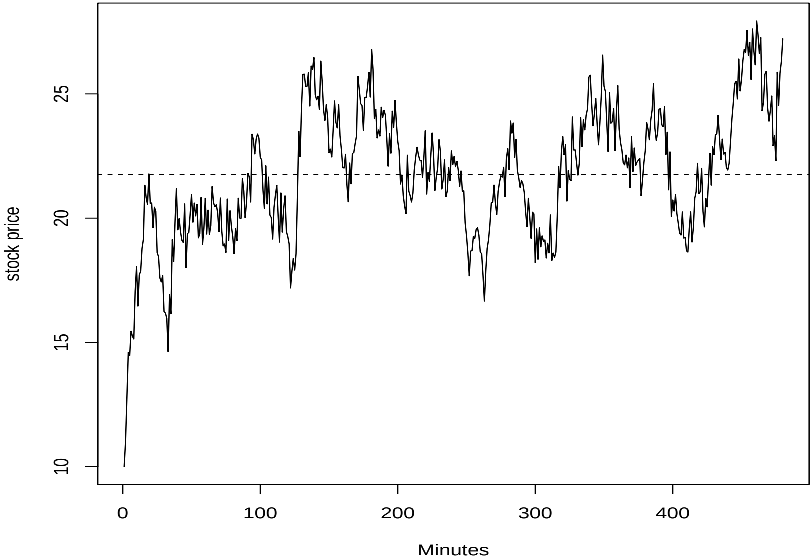
$$P(A) = 3/6 = 0.5$$

Interpretation: What does $P(A)$ measure? There are two main interpretations:

- $P(A)$ measures the likelihood that A will occure on any given experiment.
- If the experiment is performed many times, then $P(A)$ can be interpreted as the percentage of times that A will occur "over the long run". This is called the **relative frequency** interpretation.

Example Suppose a new public company's initial stock price is set at $10. Assume that its average stock price on the first day is more than $21.75 with probability 0.463. The probability 0.463 can also be interpreted as the "**long run**" percentage of stock price more than $21.75. I used R to simulate the stock price on the first transaction day.

Sample Spaces and Events



Unions and Intersections

Null Event denoted by $\emptyset$, is an event that contains no outcomes. Accordingly, $p(\emptyset) = 0$.

Union of two events A and B contains all outcomes ω in either event or in both, denoted by $A \cup B = \{\omega : \omega \in A \, or \, \omega \in B\}$.

Intersection of two events A and B contains all outcomes ω in both events, denoted by $A \cap B = \{\omega : \omega \in A \, and \, \omega \in B\}$.

Disjoint of two events A and B contains no common outcomes, also noted as mutually exclusive. It implies $P(A \cap B) = P(\emptyset) = 0$

Example Hemophilia is a sex-linked hereditary blood defect of males characterized by delayed clotting of the blood. When a woman is a carrier of classical hemophilia, there is a 50 percent chance that a male child will inherit this disease. if a carrier gives birth to two males(not twins), what is the probability that either will have the disease? Both will have the disease?

Unions and Intersections

Consider that the process of having two male children as an experiment with sample space

$$S = \{++, +-, -+, --\}$$

Where " $+$ " means the male offspring has the disease; " $-$ " means that the male does not have the disease. Assume that each outcome in S is equally likely. Note that

$$A = \{\text{first child has disease}\} = \{++, +-\}$$
$$B = \{\text{second child has disease}\} = \{-+, --\}$$

Thus,

$$A \cup B = \{\text{either child has disease}\} = \{++, +-, -+\}$$
$$A \cap B = \{\text{both child have disease}\} = \{++\}$$

Unions and Intersections

The probability that either male child will have the disease is

$$P(A \cup B) = \frac{n_{A \cup B}}{n_S} = \frac{3}{4} = 0.75$$

The probability that both male children will have the disease is

$$P(A \cap B) = \frac{n_{A \cap B}}{n_S} = \frac{1}{4} = 0.25$$

Axioms of Probability

Kolmogorov's Axioms: For any sample space S, a probability P must satisfy

(1) $0 \le P(A) \le 1$, for any event A

(2) $P(S) = 1$

(3) If $A_1, A_2, \ldots, A_n$ are pairwise mutually exclusive events, then

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

Remarks

- The term "pairwise mutually exclusive" means

$$A_i \cap A_j = \emptyset, \text{for all } i \neq j$$

- The event $\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n$ means "at least one A_i occurs".

Conditional Probability

Let A and B be events in a sample space S with $P(B) > 0$. The **conditional probability** of A, given that B has occurred, is

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

Example In a company, 36% of employees have a degree from a SEC university, 22% of those employees with a degree from the SEC are engineers, and 30% of employees are engineers. An employee is selected at random.

(a) Compute the probability that the employee is an engineer **and** is from the SEC.

(b) Compute the conditional probability that the employee is from the SEC, **given** that he/she is an engineer.

Conditional Probability

Note that

$$A = \{\text{employee has a degree from SEC}\},\ P(A) = 0.36$$
$$B = \{\text{employee is an engineer}\},\ P(B) = 0.3$$

And we also know from the problem

$$P(B|A) = 0.22$$

The probability that the employee is an engineer **and** is from SEC

$$P(A \cap B) = P(B|A)P(A) = 0.22 \times 0.36 = 0.0792$$

The conditional probability that the employee is from the SEC, **given** that he/she is an engineer

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.0792}{0.3} = 0.264$$

Conditional Probability

Remark: In the example, the conditional probability $P(A|B)$ and the unconditional probability $P(A)$ are not equal.

- In some situation, knowledge that "B has occurred" has changed the likelihood that A occurs.
- In other situations, it might be that the occurrence(or non-occurrence) of a companion event has no effect on the probability of the event of interest. This leads to the definition of independence.

When the occurrence or non-occurrence of B has no effect on whether or not A occurs, and vice-versa, we say the events A and B are **independent**. Mathematically, we define A and B are independent if and only if

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

Equivalently, if A and B are independent,

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

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

Conditional Probability

Example In an engineering system, two components are placed in a **series**; that is, the system is functional as long as both components are. Each component is functional with probability 0.95. Define the events

$$A_1 = \{\text{component 1 is functional}\}$$
$$A_2 = \{\text{component 2 is functional}\}$$

So that $P(A_1) = P(A_2) = 0.95$. Because we need both components to be functional, the probability that the system is functional is given by $P(A_1 \cap A_2)$.

- If the components operate independently, then A_1 and A_2 are independent events and the system reliability is

$$P(A_1 \cap A_2) = P(A_1)P(A_2) = 0.95 \times 0.95 = 0.9025$$

- If the components do not operate independently; failure of one component "wears on the other", we can not compute $P(A_1 \cap A_2)$ without additional knowledge.

Extension: The notion of independence extends to any finite collection of events $A_1, A_2, \ldots, A_n$. **Mutually independence** means that the probability of the intersection of any sub-collection of $A_1, A_2, \ldots, A_n$ equals the product of the probabilities of the events in the sub-collection. For example, if A_1, A_2, A_3, and A_4 are mutually independent, then

$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$
$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$
$$P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1)P(A_2)P(A_3)P(A_4)$$

Probability Rules

Let S is a sample space and A is an event. The complement of A, denoted by $\overline{A}$, is the collection of all outcomes in S not in A. That is,

$$\overline{A} = \{\omega \in S : \omega \notin A\}$$

1. **Complement rule:** $P(\overline{A}) = 1 - P(A)$
2. **Additive law:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
3. **Multiplicative law:** $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
4. **Law of Total Probability:**

$$P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B})$$

5. **Bayes' rule:**

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

Probability Rules

Example The probability that train 1 is on time is 0.95. The probability that train 2 is on time is 0.93. The probability that both are on time is 0.90. Define the events

$$A_1 = \{\text{train 1 is on time}\}, A_2 = \{\text{train 2 is on time}\}$$

We are given that $P(A_1) = 0.95, P(A_2) = 0.93, P(A_1 \cap A_2) = 0.90$

(a) What is the probability that train 1 is **not on time**?

$$P(\overline{A_1}) = 1 - P(A_1) = 1 - 0.95 = 0.05$$

(b) What is the probability that **at least one** train is on time?

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = 0.95 + 0.93 - 0.90 = 0.98$$

(c) What is the probability that train 1 is on time **given** that train 2 is on time?

$$P(A_1 | A_2) = \frac{P(A_1 \cap A_2)}{P(A_2)} = \frac{0.90}{0.93} \approx 0.968$$

(d) What is the probability that train 2 is on time **train 1** is not on time?

$$P(A_2|\overline{A_1}) = \frac{P(\overline{A_1} \cap A_2)}{P(\overline{A_1})} = \frac{P(A_2) - P(A_1 \cap A_2)}{1 - P(A_1)} = \frac{0.93 - 0.90}{1 - 0.95} = 0.6$$

(d) Are A_1 and A_2 independent?
They are not independent because $P(A_1|A_2) \neq P(A_1)$. In other words, knowledge that A_2 has occurred changes the likelihood that A_1 occurs.

Example An insurance company classifies people as “accident-prone” and “non-accident-prone”. For a fixed year, the probability that an accident-prone person has an accident is 0.4, and the probability that a non-accident-prone person has an accident is 0.2. The population is estimated to be 30% accident-prone. Define the events

$$A = \{\text{policy holder has an accident}\}, B = \{\text{policy holder is accident-prone}\}$$

Probability Rules

Note that

$$P(B) = 0.3, P(A|B = 0.4), P(A|\overline{B}) = 0.2$$

(a) What is the probability that a new policy-holder will have an accident?

$$\begin{aligned} P(A) &= P(A|B)P(B) + P(A|\overline{B})P(\overline{B}) \\ &= 0.4 * 0.3 + 0.2 * 0.7 = 0.26 \end{aligned}$$

(b) Suppose that the policy-holder does have an accident. What is the probability that he/she was "accident-prone"?

$$\begin{aligned} P(B|A) &= \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\overline{B})P(\overline{B})} \\ &= \frac{0.4 * 0.3}{0.4 * 0.3 + 0.2 * 0.7} \approx 0.46 \end{aligned}$$

Random Variables

A **random variables**(denoted as r.v.) Y is a variable whose value is determined by chance. The **distribution** of a r.v.consists of two parts:

- an elicitation of the set of all possible values of Y(called **support**).
- a function that describes how to assign probabilities to events induced by Y.

A r.v. Y can be categorized by two types:

discrete If Y can assume only a finite (or countable) number of values.

continuous If can envision Y as assuming in an interval off numbers.

Remarks: By convention, we denote random variables by upper case letters towards the end of the alphabet, e.g., W, X, Y, Z, etc. A possible value of Y is denoted generally by its according lower case letter. In words,

$$P(Y = y)$$

is read, "the probability that the random variable Y equals the value y."

Random Variables

Example Classify the following random variables as **discrete** or **continuous** and specify the support of each random variable.

V = number of broken eggs in a randomly selected carton(dozen)

W = pH of an aqueous solution

X = length of time between accident at a factory

Y = whether or not you pass this class

Z = number of cans of beer that an adult has

- The r.v. V is **discrete**. It can assume values in

$$\{v : v \in 0, 1, 2, \ldots, 12\}$$

- The r.v. W is **continuous**. It can appropriately assume values in

$$\{\omega : 0 \leq \omega \leq 14\}$$

Of course, its support can assumed as $\{\omega : -\infty < \omega < \infty\}$

- The r.v. X is **continuous**. It can assume values in

$$\{x : 0 < x < \infty\}$$

The key feature here is that a time cannot be negative.

- The r.v. Y is **discrete**. It can assume values in

$$\{y : y = 0, 1\}$$

Where 0 can arbitrarily label for passing and 0 for failing. R.V. that assume exactly two values are called **binary**.

- The r.v. Z is **discrete**. It can assume values in

$$\{z : z = 0, 1, 2, \ldots\}$$

Here a large amount of beer is assumed.