AM 165-166: Statistical Inference

www.dam.brown.edu/people/huiwang/classes/AM165-166

Textbook: "Mathematical Statistical with Applications" (6th ed.) by D.D. Wackerly, W. Mendenhall III, and R.L. Scheaffer.

Office hours: Thursday & Friday 11:00-12:00, or by appointment.

Email address: Hui Wang, huiwang@cfm.brown.edu.

Exams and homeworks: We will have regular homeworks. Tentative schedule for exams: one mid-term and one final (cumulative).

Teaching Assistant: TBA.

Software: Matlab, or Splus.

Syllabus

1st Semester: Probability, point estimation, and hypothesis testing.

2nd Semester: Least square regression, analysis of variance, non-parametric statistics, sample survey, Baysian statistics, and other topics of choice.

Probability and Statistics

- What is probability?

$$\boxed{\text{A system with randomness}} + \boxed{\text{Conditions and assumptions}}$$

$$\Downarrow$$

$$\boxed{\text{Various distributional properties}}$$

Fundamentally mathematics.

The answer can only be right or wrong.

Examples of Probability Questions

1. Coin Toss. Assumption $P(\text{Heads}) = P(\text{Tails}) = 1/2$ and tosses are independent of each other.
   (a) $P(\text{HH})$, $P(\text{HT})$, and so on.
   (b) Average number of Heads in 100 tosses? What about 101 tosses?
   (c) Waiting time to get the first Heads toss. What is the average waiting time?

2. The flippant Juror. A three-man jury has two members each of whom independently has probability p of making correct decision and a third member who flips a fair coin for each decision (majority rules). A one-man jury has probability p of making the correct decision. Which jury has the better probability of making the correct decision? [from "Fifty challenging problems in probability – with solution" by Frederick Mosteller]

3. All in. No-limit Hold'em, headsup. All-in on flop.

   Opponent's holding: ♠ A ♡ A.

   Board: ♣ A ♣ K ♢ J.

   Your holding: ♣ 5 ♣ 6.

   What is your chance of winning? When this all in is profitable to you?

4. Which box? You are playing a game, and are presented three sealed boxes, one contains a prize, two empty. You randomly pick one box. After that, the host will open an empty one in the other two boxes. Now you have a choice . . .

   A related question of Principle of restricted choice in bridge. Dummy $***$, South AJ10. Starts from dummy, east play low, you finesse Jack, caught by West's King. Now South is preparing to finesse 10, what's his chance of success? Better than, equal to, or less than 50%?

5. Accurate lie detectors? A collection of suspects, 90% are honest men, 10% are thieves. An 80% accurate lie detector – an honest man will pass the test with probability 80%, a thief will fail with probability 80%. Randomly pick a suspect and put him through the test. What is the probability that the suspect is a thief if he fails the lie detector test?

Another example of this phenomenon is the high false-positive rate in medical testing.

Probability and Statistics

- What is Statistics? Target of Interest: Population. Need information on the population.

Design of data collection mechanism: experiments, survey, observations, etc.

$\Downarrow$

Data collection: samples

$\Downarrow$

Inference on various configurations of the population

Probability serves as an important tool, but very different ideas.

The answer is usually grey. For the same set of data, different conclusions are possible with different legitimate methods.

Two Camps of Statisticians

- "objectivist" or "frequentists": Probability is objective (or inherent) properties of the system. Consider tossing a coin, the probability of Heads has its own existence, separate from data.
- "subjectivist" or "Bayesians": Probability is subjective, reflecting the "degree of belief". Consider tossing a coin, the probability of Heads is largely your subjective opinion (prior), and each toss "update" your opinion. Whence the probability is inseparable from data.

Reference: *Topics in the Foundation of Statistics* (1997), Edited by Bas C. van Fraassen. Kluwer Academic Publishers, Boston.

A few random thoughts on Statistics

- Do not believe everything on face value.
- The data collection mechanism is extremely important, but often neglected.
- The conclusion from statistics should not be understood in its absolute terms. For a statistician, "no" means "highly unlikely from data" and "yes" means "consistent with the data".

Examples of statistical questions

1. Coin toss: $P(\text{Heads}) = p = 1 - P(\text{Tails})$, with p unknown. Toss coin 100 times, come up with 53 Heads.

   (a) Your estimate of p.

   (b) Are you confident about your estimate?

   (c) If someone claims $p = 0.9$, do you think you can say "no"?

   (d) If someone claims $p = 0.5$, do you think you can say "no"? or "Yes"?

   (e) Is it possible that we get the true value of p through statistical procedure?

2. Is this treatment really working? "Gastric freezing" is a treatment for ulcers in upper intestine. The idea is that cooling the stomach will reduce the acid production will and so relieve ulcers. The justification was the positive experimental results reported in *Journal of the American Medical Association*.

Gastric freezing $\Rightarrow$ Observe pain relief.

The treatment is safe, easy, and widely used for several years. Does this experiment really show that the treatment is effective?

Experiment with a control group: 34% improve with gastric freezing treatment, 38% improve with sham treatment. The use of gastric freezing was abandoned.

A sidenote on the dilemma for medical company: when can a new medicine be accessible to general public?

3. Sample survey: how large is your family? We would like to estimate the average number of children per family around 20 years ago. Sample survey: the number of brothers and sisters in yours family, including yourself.

4. Kidney cancer. See the handout.

5. Anchoring effect:

How to draw a histogram?