AM 1650: Statistical Inference (I)

www.dam.brown.edu/people/huiwang/classes/am165/am165/html

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: Regular homeworks. Two mid-term and one final (cumulative).

Software: Not required.

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. 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. 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 . . .

4. 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?

Probability and Statistics

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

$$\boxed{\text{Design of data collection mechanism: experiments, survey, observations, etc.}}$$

$$\Downarrow$$

$$\boxed{\text{Data collection: samples}}$$

$$\Downarrow$$

$$\boxed{\text{Inference on various configurations of the population}}$$

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.

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. 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 your family, including yourself.

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.

  Classical probability as a measurement of uncertainty interpreted as "long run relative frequency". Extremely useful (kind of a constructive way of understanding uncertainty).

  Difficulties in interpreting common-sense statements such as "What is the probability that it will rain tomorrow?" "What is the chance that Red Sox will win World Series this year?" "What is Hillary's chance of becoming the next President?"

- "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.

  Anything unknown is given a probability distribution, representing degrees of belief [subjective probability].

  Degrees of belief [subjective probability] can be handled as if it were classical probability, and therefore, *mathematically* there is no difference.

Example: Horse Betting

Horse	Amount bet on horse (in thousands)	"probability"	odds against	payoff for \$1 bet
H1	500			
H2	250			
H3	100			
H4	100			
H5	50			
Total	1000			

$$\text{odd against event } E : O(E) \doteq \frac{1 - P(E)}{P(E)}$$

The odds are set after all the money are collected. So the odds represent a concensus of the bettors' subjective perception on the abilities of horses.