Statistics II: Statistical Inference

Prerequisites: Statistics I; Probability I and II

References:
1. Casella, G. and Berger, R. *Statistical Inference*
2. Bickel, D. and Doksum, K. *Mathematical Statistics*
3. Hogg and Craig. *Introduction to Mathematical Statistics*

Grading (Tentative): 20 marks for assignments; 20 marks each for two class tests; 40 marks for the final exam

Lectures: Online lectures will be on Zoom; Time: Monday, 12:10 – 1:10 pm and 2:00 – 3:00 pm; Wednesday, Friday, 2:00 – 3:00 pm

Lecture Notes: Lecture notes will be posted on Moodle. Attempt will be made to post recording of lectures on Moodle too

Assignments: Assignments will also be posted on Moodle. Answers to these must be submitted by uploading to Moodle.

Contact e-mail: *mohan.delampady@gmail.com*

What is statistical inference? Why is a probability model needed? Inference based on modeling data or observations using probability models is of interest. Consider this example.

Example 1. In fisheries and ecology, one uses capture-recapture methods to estimate the size of a population. Let N = total size; this is unknown. First, N_1 of the individuals in this population are caught, tagged and then released. In the next step n out of N are caught again. If n_1 out of n have tags on them, what is an estimate for N? Let X = number of tagged fish out of n. Assuming that the tagged individuals had mixed well with the others in the entire population before sampling and that all individuals in the population have the same chance of being caught, we can write down a probability model for X:

$$P(X = x|N) = \frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}}, x = 0, 1, \ldots n; x \leq N_1; n - x \leq N - N_1.$$

Thus, a statistical model is an approximate but simple theoretical framework to work with.

Statistical inference is mathematical but not mathematics itself, because statistics is inductive reasoning, not deductive as in mathematics. In mathematics, one states certain axioms and then proves (or deduces) certain conclusions, such as theorems. In statistics, one uses observations or data, which are instances of events or occurences. From these one generalizes to other situations (which is induction). The data may be compatible with many models which are just mathematical structures for the generation of data. So, this is a one-to-many problem. Consider an experiment where one tosses a coin 10 times and observes it coming up heads 8 times. What is $\theta = P($ coin comes up heads on any toss $)$? Any $0 < \theta < 1$ can produce the outcome $X = 8$ in $n = 10$ tosses.

It is important to have a theory which can show that statistical inference is valid, consistent and leads to correct conclusions under appropriate assumptions. This is provided by the mathematical theory of probability.

Since statistical inference involves picking a model from a collection of models, it is necessary to study the properties of such classes of models. Optimality results of statistical procedures will be established later for some such classes.

First a few definitions needed in the discussion.

As mentioned previously, a probability model is a structure assumed to model the realization of random observables or data. Simple models involve a few unknown quantities which are used to index or label the distributions in the family of models. These labels are called *parameters*. The set of all parameter values in a family is called the parameter space. Usually, parameters are associated with important features of the distribution, such as mean and variance.

If $\mathcal{P}$ denotes the family of probability models under consideration, then
$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, θ is the unknown parameter and Θ is the parameter space.

Example 2. $X =$ number of times a coin comes up heads when tossed 10 times, $\theta = P($ coin comes up heads on any toss $)$. Then $X \sim \text{Binomial}(10, \theta)$. So,
$\mathcal{P} = \{$ all Binomial(10, θ), $0 < \theta < 1\}$. Note $\Theta = (0, 1) \subset \mathcal{R}^1$.

Example 3. Suppose X denotes the length of time required for a randomly chosen person to recover from common cold, and we model it as $X \sim N(\mu, \sigma^2)$. Then $\mathcal{P} = \{$ all $N(\mu, \sigma^2), -\infty < \mu < \infty, \sigma^2 > 0.\}$. Here $\Theta = \{(\mu, \sigma^2), \mu \in \mathcal{R}^1, \sigma^2 \in \mathcal{R}^+\} = \mathcal{R}^1 \times \mathcal{R}^+ \subset \mathcal{R}^2$. Note that recovery time cannot be negative, but this approximation is reasonable if almost all the probability lies on the positive region.

The idea of a parameter is for it to specify the distribution.

Identifiability. For any θ_1 and θ_2 in Θ, whenever $\theta_1 \neq \theta_2$, we must have $P_{\theta_1} \neq P_{\theta_2}$.

Example 4. Suppose N is the number of tigers in a reserve forest, and we assume $N|\lambda \sim \text{Poisson}(\lambda)$. Let S be the number of tigers sighted by a team of investigators in a study here. Since this involves detection, probability of which is usually less than 1, we can assume, $S|\{(N = n), p\} \sim \text{Binomial}(n, p)$ and therefore, $S|\lambda, p \sim \text{Poisson}(\lambda p)$. (Show this as an exercise.) If $S_1, S_2, \ldots, S_k$ are i.i.d Poisson(λp) can we make inferences about both λ and p? The model for S, $\{\text{Poisson}(\lambda p), \lambda > 0, 0 < p < 1\}$, is not identifiable. Here $\theta = (\lambda, p)$ and $P_\theta = \text{Poisson}(\lambda p)$. Take $\theta_1 = (10, 0.4)$ and $\theta_2 = (20, 0.2)$. Then $\theta_1 \neq \theta_2$, but $P_{\theta_1} = P_{\theta_2}$.