STAT 417. lecture Note 5.

① why need statistics? what is indiana annual salary? x

on the road, randomly pick up ten persons, their annual salaries are

121.2 K, 92.1 K, 56.9 K, 72.4 K, 31.7 K,

64.5 K, 21.8 K, 46.9 K, 82.3 K, 51.1 K

Question: How much ~~do~~ ~~[illegible]~~does a typical person in Indiana earn?

of course, we don't know the true indiana annual salary.

but we want to use the above ten samples to infer this ~~[illegible]~~.

~~Denote the~~ The simpliest way is to use the sample mean of the ten samples to estimate x. In the above example, $\bar{x} = 64.1K$

② How do we handle statistics? probability model is used to describe variation in the data.

<u>E.g.1</u> Assuming ~~[illegible]~~ Indiana annual salary x follows some probability model, say, $x \sim$ Exponential (10), then the above ten samples will are considered as random numbers drawn ~~[illegible]~~ from this exponential distribution.

③ main components of statistics:

- Estimation
- testing hypothesis
- confidence interval.

Statistical Models:

E.g.3. X is the lifetime of ~~mercedez~~ BMW. $X \sim$ exponential(θ). $\theta > 0$ unknown.

E.g.2 suppose 8 patients each has ~~poor~~ chance θ to die. $0 \le \theta \le 1$

let X = # of patients dying. Then $X \sim$ Binomial$(8, \theta)$.

In general, we suppose $\{P_\theta : \theta \in \Theta\}$. is a collection of probability models.

In eg2. $\Theta = [0, 1]$. $P_\theta \sim$ Binomial$(8, \theta)$. $S = \{0, 1, \cdots, 8\}$.

eg3. $\Theta = (0, \infty)$. $P_\theta \sim$ Exponential(θ). $S = (0, \infty)$.

E.g.4. $\Theta = \{1, 2\}$. $P_1 \sim$ unif$\{1, 2, \cdots, 1000\}$. $P_2 \sim$ unif$\{1, 2, \cdots, 10{,}000\}$.

Sample space $S = \{1, 2, \cdots\}$. is for a single draw.

~~Eg.~~ sample space $S = \{(1,1), (1,2), \cdots, (2,1), (2,2), \cdots\}$. is for two draws.

~~Step 6~~ ~~Likelihood Inference~~.

E.g.5. (Normal model): $\Theta = \{\theta = (\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$. ~~$P_{(\theta, \sigma^2)} \sim N(\theta, \sigma^2)$~~

$P_\theta \sim N(\mu, \sigma^2)$, where $\theta = (\mu, \sigma^2)$. For a single draw $X \sim P_\theta$.

the samples space $S = \mathbb{R}$.

E.g.6. (poisson model): $P_\theta \sim$ poisson(θ). $\theta \in \Theta = (0, \infty)$.

the sample space of a single draw $X \sim P_\theta$ is $S = \{0, 1, 2, \cdots\}$.

E.g.7. (Bernoulli model). $P_\theta \sim$ Bernoulli(θ). $\theta \in \Theta = [0, 1]$.

the sample space of a single draw $X \sim P_\theta$ is $S = \{0, 1\}$.

Ch6 Likelihood Inference.

A warm-up example; 8 patients. each has probability θ to die, $0 \le \theta \le 1$.

~~the~~ let X = # of patients dying. $X \sim \underbrace{\text{Binomial}(8, \theta)}_{= P_\theta}$. $\theta \in \Theta = [0,1]$.

suppose we know that their are 4 people dying. can you infer θ?

arbitrarily, $\theta = \frac{4}{8} = \frac{1}{2}$.

§6.1. Likelihood function.

suppose. $\{P_\theta : \theta \in \Theta\}$. we ~~kno~~ have data $s \sim P_\theta$

~~In discrete data, let $f_\theta(s)$ be the~~

In discrete model, f_θ is the ~~prodict~~ probability distribution, under model P_θ.

In continuous model, f_θ is the pdf. under model P_θ.

the likelihood function, denoted $L(\theta|s)$; is defined to be

$$L(\theta|s) = f_\theta(s).$$

Eg1. suppose sample space $S = \{1, 2, \ldots\}$. the statistical model is $\{P_\theta : \theta \in \{1,2\}\}$, where $P_1 \sim \text{unif}\{1, 2, \ldots, 1000\}$.

$P_2 \sim \text{unif}\{1, 2, \ldots, 10{,}000\}$. Furthermore, $s = 10$,

then under P_1, $f_1(s) = \frac{1}{1000}$.

under P_2, $f_2(s) = \frac{1}{10{,}000}$.

so the likelihood function. $L(\theta|s) = \begin{cases} \frac{1}{1000}, & \theta = 1, \\ \frac{1}{10000}, & \theta = 2. \end{cases}$