

# PSTAT 115 Section 1

10/04/2020

```
library(ggplot2)
```

Key ideas:

1. How to use server to do homework
2. Notations
3. Likelihood function: how they are defined, what they look like,  $L(\theta)$  notation, proportionality and plotting in R
4. Maximum Likelihood Estimator (MLE) and its functional invariance
5. Sufficiency
6. Bayesian and Frequentist differences

**1. Demonstration of how to use the server, pulling the assignment with the link and making a copy of the .rmd file to do homeworks**

**2. Notations:**

**a) Random Variable and Realization of a Random Variable**

- A Random Variable is a variable whose values depend on outcomes of a random event.
  - The height of 100 randomly selected human beings.  $Y_1, Y_2, \dots, Y_{100}$
- Realization of a random variable is the outcome of one random event.
  - We find 100 human beings and measure their heights.  $y_1, y_2, \dots, y_{100}$

**b) Population and Sample**

- Population is a set of similar items or events which is of interest for some questions or experiments.
  - The height of all human beings.
- Sample is a set of data collected from a population by a defined method.
  - If we find 100 students in UCSB and measure their heights, then the heights of those 100 students forms a sample of heights of human beings with sample size 100. Of course, this sample is not representative and we need a better sampling method.
- If we want to know what is the mean height of all human beings (Property of the population, we call it parameter), we have to measure the height of each human beings, which is not feasible. Instead, we take a sample, which has much smaller size compared to the number of human beings and use the sample mean to inference about the population mean.
  - Can you suggest some better ways to get a representative sample than measuring the heights of 100 UCSB students?

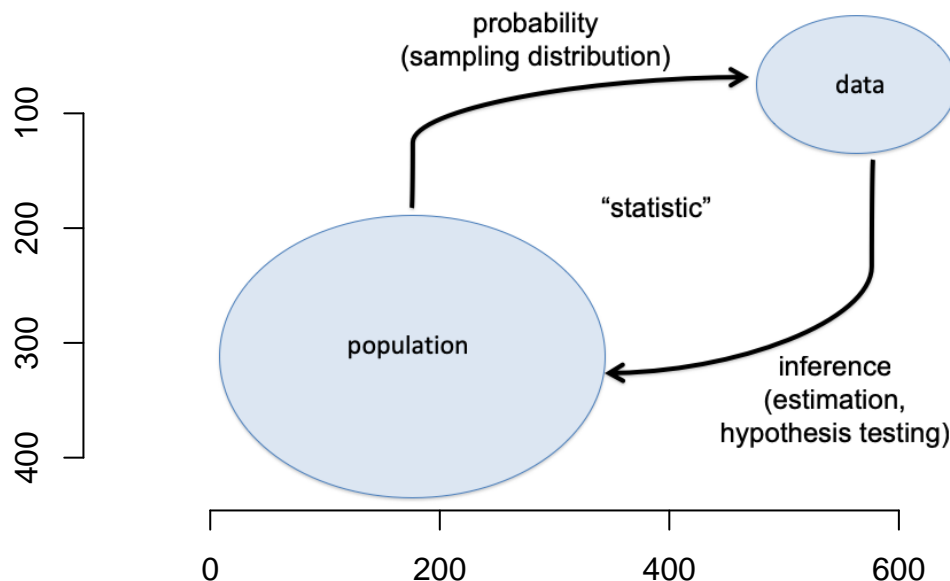
**c) Estimator and Estimate**

- Estimator is a random variable. It gives a rule to estimate the parameter from a sample.
  - We randomly select 10,000 human beings and the mean height is the estimator of mean height of all human beings.

$$\bar{Y} = \sum_{i=1}^{10,000} Y_i/n$$

- Estimates is the corresponding realization of the estimator.
  - Say we have already selected 10,000 human beings and measured their heights  $(y_1, y_2, \dots, y_{10,000})$ . Then the estimate is

$$\bar{y} = \sum_{i=1}^{10,000} y_i / n$$



### 3. Likelihood and Log-likelihood

a) Let  $X_1, X_2, \dots, X_n \sim N(\mu, 1)$ . The dataset obtained is  $x_1, x_2, \dots, x_n$ . Find the likelihood, log-likelihood and proportional simplification. What is the MLE

- If we go to the distribution table, we can find out that pdf of normal distribution with is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Then the likelihood with  $\sigma^2 = 1$  is

$$\begin{aligned} L(\mu) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2}} \\ &= (2\pi)^{-n/2} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2}} \end{aligned}$$

- The log-likelihood is

$$l(\mu) = \log(L(\mu)) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2$$

- Simplification
  - Likelihood:

$$\begin{aligned}
L(\mu) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2}} \\
&= (2\pi)^{-n/2} \exp\left(-\sum_{i=1}^n (x_i - \mu)^2\right) \\
&\propto \exp\left(-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2\right) \\
&= \exp\left(-\frac{1}{2} \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)^2\right) \\
&= \exp\left(-\frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2 - \frac{1}{2} \sum_{i=1}^n (\bar{x} - \mu)^2 + \sum_{i=1}^n (x_i - \bar{x})(\bar{x} - \mu)\right) \\
&\propto \exp\left(-\frac{n}{2} (\bar{x} - \mu)^2\right)
\end{aligned}$$

- log-likelihood:

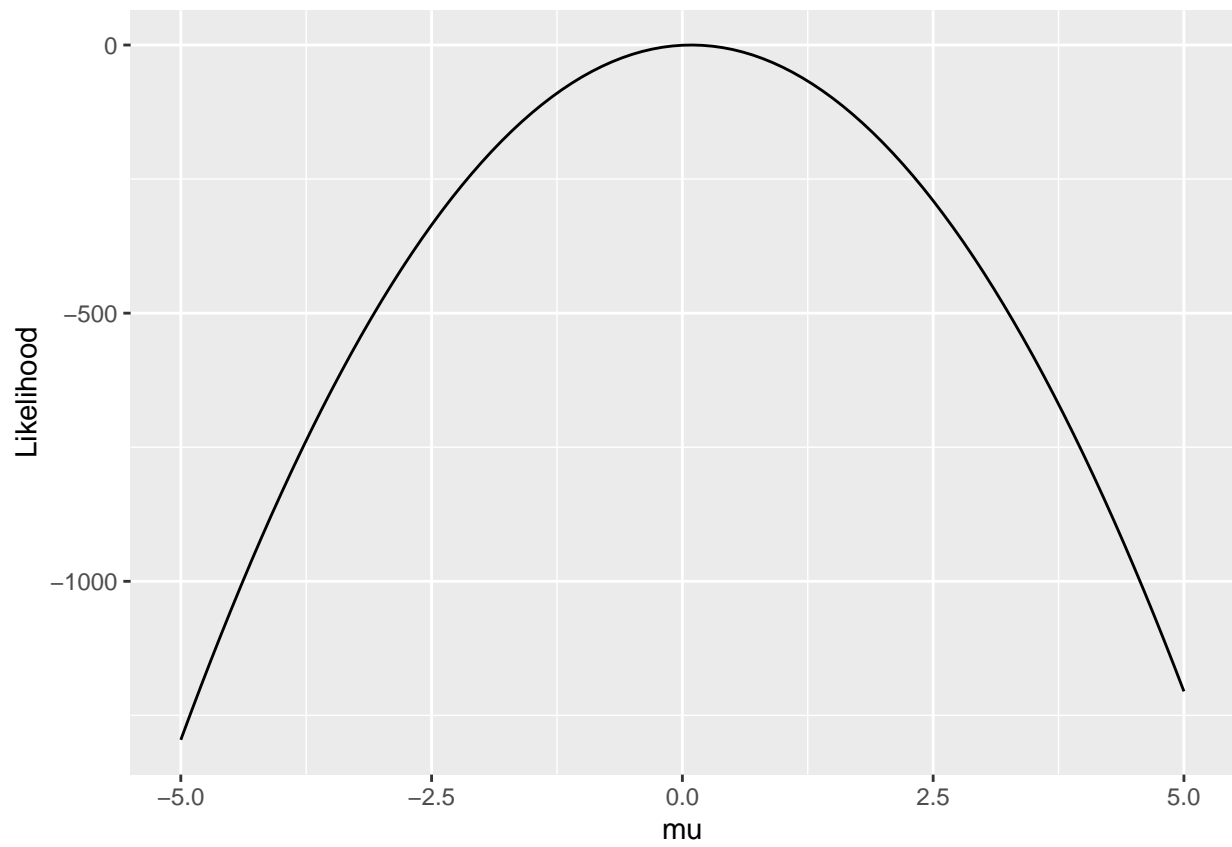
$$l(\mu) = \log(L(\mu)) \propto -\frac{n}{2} (\bar{x} - \mu)^2$$

- MLE:  $\bar{x}$

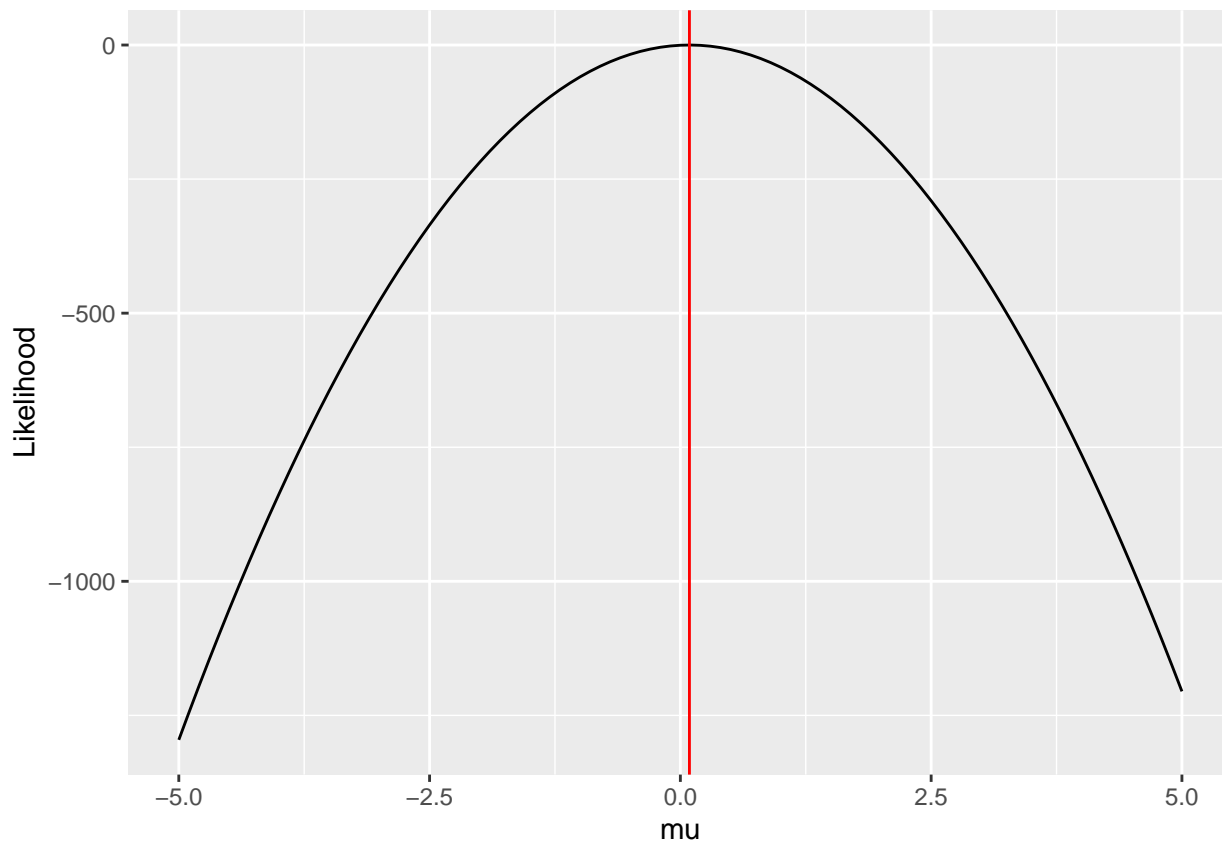
b) Let's assume  $n = 100$  and we get our dataset from the following codes. Plot the log-likelihood function

```
set.seed(123)
n <- 100
x <- rnorm(100, mean = 0, sd=1)
```

```
L <- function(mu){#Log-Likelihood Function
  -n/2*(mean(x)-mu)^2
}
p <- ggplot(data = data.frame(mu = 0), mapping = aes(x=mu))+stat_function(fun = L)+
  xlim(-5,5)+scale_y_continuous(name = "Likelihood")
p #plot log-likelihood function against mu
```



```
p+geom_vline(xintercept = mean(x), color = 'red') #identify the MLE
```



c. Let  $X_1, X_2, \dots, X_n \sim \text{Bernoulli}(p)$ . The dataset obtained is  $x_1, x_2, \dots, x_n$ . Find the likelihood, log-likelihood, proportional simplification and MLE. Then plot the log-likelihood function with respect to  $p$  assuming we have the following data:

```
set.seed(123)
n <- 100
x <- rbinom(n, 1, 0.5)
```

- pdf of Bernoulli( $p$ ) distribution:  $f(x) = p^x(1-p)^{1-x}$
- Likelihood:

$$L(p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} \\ = p^X (1-p)^{n-X}$$

Where  $X = \sum x_i$

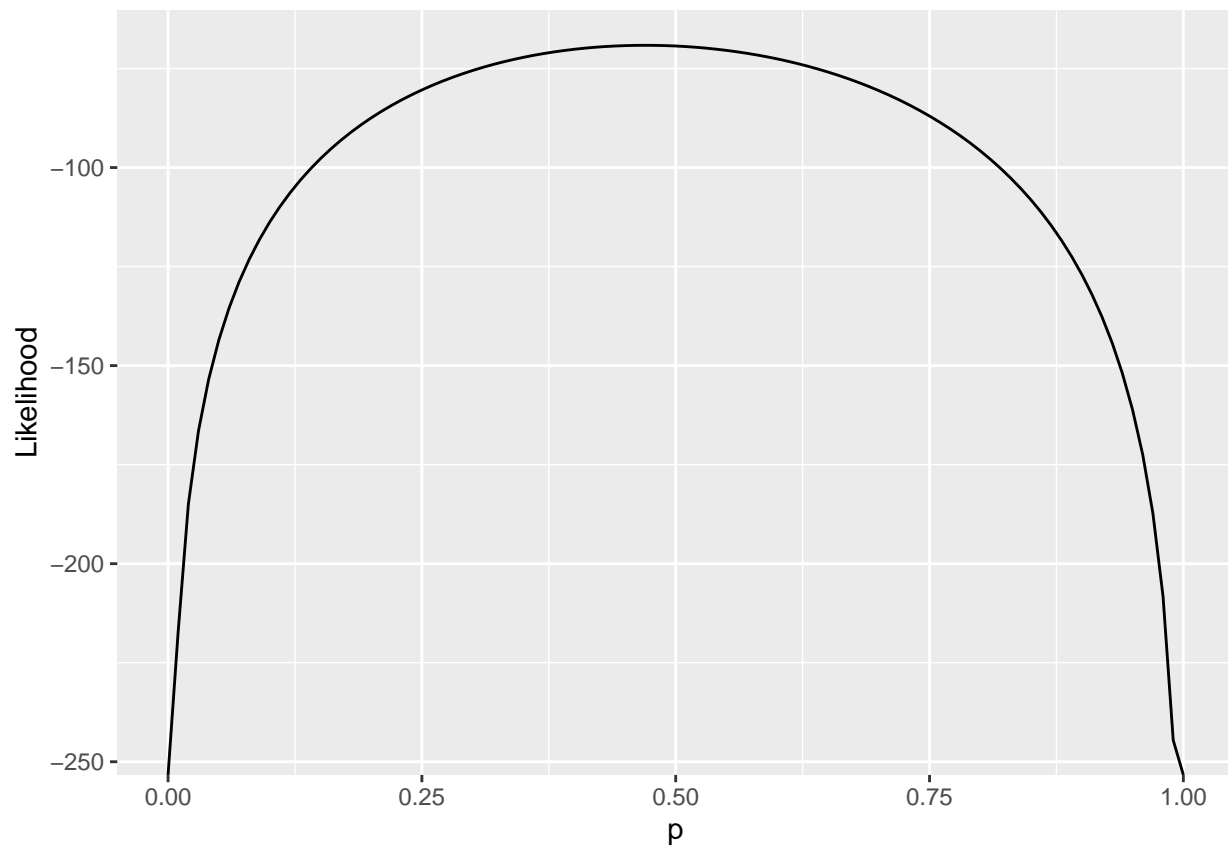
- Log-likelihood

$$\log L(p) = X \log(p) + (n - X) \log(1 - p)$$

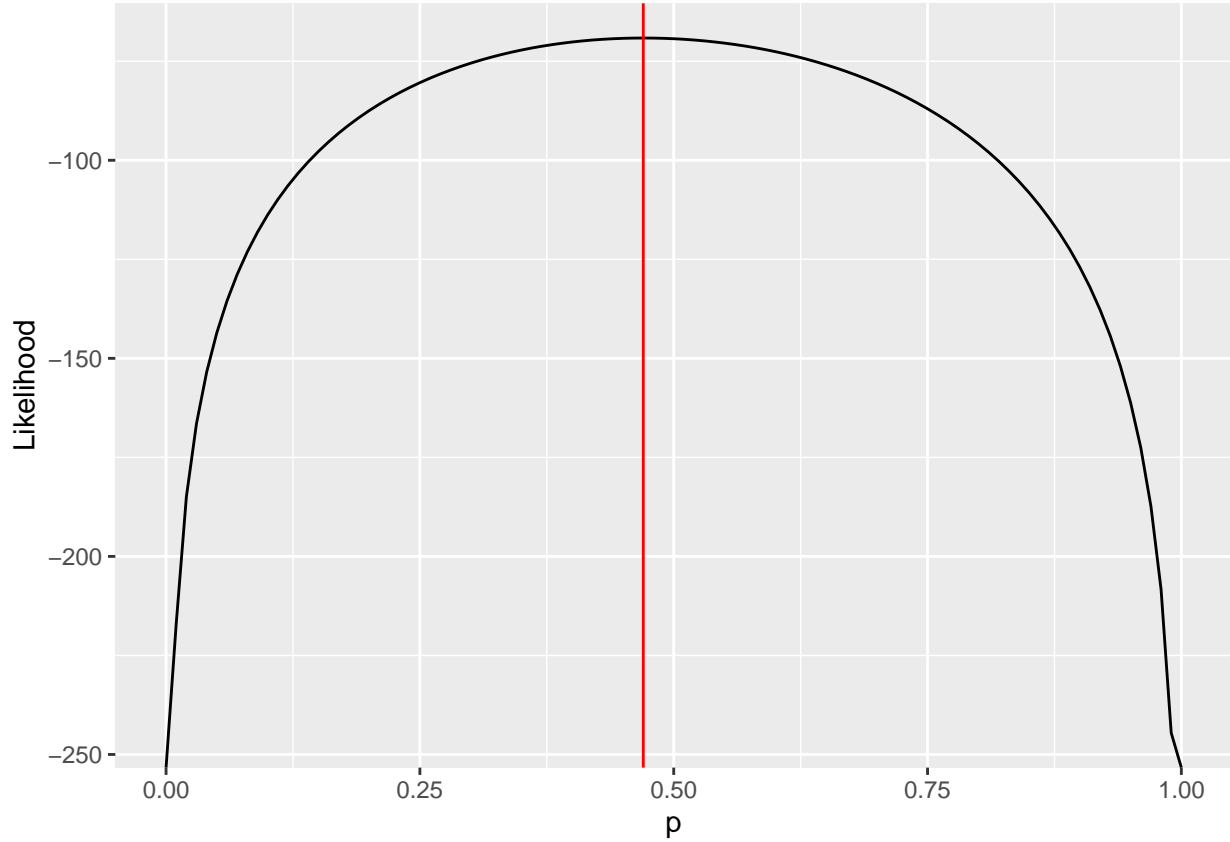
- Simplification: It is the simplest form.
- MLE:  $X/n$
- Plot

```
L <- function(p){#Log-Likelihood Function
  sum(x)*log(p)+(n-sum(x))*log(1-p)
}
```

```
p <- ggplot(data = data.frame(p = 0), mapping = aes(x=p))+stat_function(fun = L)+
  xlim(0,1)+scale_y_continuous(name = "Likelihood")
p #plot log-likelihood function against mu
```



```
p+geom_vline(xintercept = mean(x), color = 'red') #identify the MLE
```



4. Let  $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$ . The dataset obtained is  $x_1, x_2, \dots, x_n$ .

a) Find the MLE of  $\mu, \sigma^2$ .

$$\begin{aligned}
 L(\mu, \sigma) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \\
 &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) \\
 \log(L(\mu, \sigma)) &= -\frac{n}{2}(\log(2\pi\sigma^2)) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2
 \end{aligned}$$

Since we want to find the  $\sigma$  and  $\mu$  that maximize the log-likelihood, we compute the derivatives with respect to the two parameters

$$\begin{aligned}
 \frac{\partial \log(L(\mu, \sigma))}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \\
 \frac{\partial \log(L(\mu, \sigma))}{\partial \sigma^2} &= \frac{n}{(\sigma^2)^2} \left( \sigma^2 - \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \right)
 \end{aligned}$$

Set both derivatives to zero we get:

$$\mu_{MLE} = \bar{x}$$

$$\sigma_{MLE}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

**b) Find the MLE of  $\mu^3$**

Theorem (Invariance Property of MLE estimators): If  $\hat{\theta}$  is an MLE for  $\theta$  and  $g$  is a function, then  $g(\hat{\theta})$  is an MLE of  $g(\theta)$ . So the answer is  $\bar{x}^3$ .

**5. Sufficiency**

Sufficient statistics summarize all the information in a sample about a target parameter.

Let  $Y_1, Y_2, \dots, Y_n$  denote a random sample from a probability distribution with unknown parameter  $\theta$ . Then the statistic  $U = g(Y_1, Y_2, \dots, Y_n)$  is said to be sufficient for  $\theta$  if the conditional distribution of  $Y_1, Y_2, \dots, Y_n$  given  $U$ , does not depend on  $\theta$ .

Let  $U$  be a statistic based on the random sample  $Y_1, Y_2, \dots, Y_n$ .

Then  $U$  is a sufficient statistic for the estimation of a parameter  $\theta$  if and only if the likelihood

$L(\theta) = L(y_1, y_2, \dots, y_n)$  can be factored into two nonnegative functions,

$L(y_1, y_2, \dots, y_n | \theta) = g(u, \theta) * h(y_1, y_2, \dots, y_n)$

where  $g(u, \theta)$  is a function only of  $u$  and  $\theta$  and  $h(y_1, y_2, \dots, y_n)$  is not a function of  $\theta$ .

**a)**

Let  $Y_1, Y_2, \dots, Y_n$  be a random sample in which  $Y_i$  possesses the probability density function  $f(y_i | \theta) = \frac{1}{\theta} e^{-\frac{y_i}{\theta}}$  where  $0 \leq y_i < \infty$  and  $\theta > 0, i = 1, 2, \dots, n$ . Find a sufficient statistic.

$$L(y_1, y_2, \dots, y_n | \theta) = f(y_1, y_2, \dots, y_n | \theta) = \prod_{i=1}^n \frac{1}{\theta} e^{-\frac{y_i}{\theta}} = \frac{e^{-\sum_{i=1}^n y_i}}{\theta^n}$$

$$= \frac{e^{-n\bar{y}/\theta}}{\theta^n}$$

$$g(\bar{y}, \theta) = \frac{e^{-n\bar{y}/\theta}}{\theta^n} \text{ and } h(y_1, y_2, \dots, y_n) = 1$$

Hence,  $\bar{Y}$  is sufficient.

**6. Comparison between Bayesian and Frequentist**

- In frequentist inference, unknown parameters treated as constants
  - Estimators are random (due to sampling variability)
- In Bayesian inference, unknown parameters are random variables.
  - Need to specify a prior distribution for  $\theta$  - The prior can be uninformative or informative.
  - Gather data
  - Update your prior distribution with the data using Bayes' theorem, resulting in posterior distribution. The posterior distribution is a probability distribution that represents your updated beliefs about the parameter after having seen the data.
  - Analyze the posterior distribution and summarize it (mean, median, sd, quantiles...)



- Confidence level differences
  - Bayesian confidence intervals describe information about the location of the true value of  $\theta$  after observing  $Y$ . (Post experimental coverage)
  - Frequentist confidence intervals describe the probability that the interval will cover the true value before the data is observed. (Pre experimental coverage)