

# Statistic Notes

This is a Notes on Statistical Inference following the book by Mukhopadhyay. The Notes begins with Chapter 7 Point Estimation

## Motivation

Consider a set of iid *observable* random variables  $X_1 \dots X_n$  having common pmf or pdf  $f(x)$ . We can index the whole  $f(x)$  by some parameter  $\vec{\theta}$  therefore we write the pmf or pdf as  $f(x; \vec{\theta})$ . The idea behind writing the whole pmf or pdf with a parameter is characterized by the notion of sufficient statistics.

Suppose the parameter  $\vec{\theta}$  is fixed but unknown and possible values of  $\vec{\theta}$  belong to *parameter space*  $\vec{\Theta} \subset \mathbb{R}^n$

For example, consider the normal distribution  $N(\mu, \sigma^2)$ . If both parameters are unknown, then  $\vec{\theta} = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}^+$

Basically, What we are trying to do is know more about the population by 'doing stuff' on currently known parameters or data. And the meaning of 'doing stuff' is the subject matter of this Chapter

## 1 Definition

**Definition 1.1.** An estimator of a point estimator of the unknown parameter is a function  $\vec{T} = \vec{T}(X_1, \dots, X_n)$  which is allowed to depend only on the observable random variables  $X_1, \dots, X_n$ . Once a particular data  $\vec{X} = \vec{x}$  has been observed, the numerical value of the estimator must be computable.

## 2 Method of Moment

Suppose  $\vec{\theta} = (\theta_1, \dots, \theta_k)$ . Derive the first  $k$  theoretical moments of the distribution  $f(x; \vec{\theta})$  and pretend that they are equal to the corresponding sample moments, thereby obtaining  $k$  equation in  $k$  unknown parameter  $\theta_1, \dots, \theta_k$ . The solution are hence the estimator.

**Example 1.** Consider  $n$  iid sample  $x_1, \dots, x_n$  from  $X \sim U(0, \theta)$ . We can use method of moment to get the estimator for parameter  $\theta$ .

Consider

$$\mathbb{E}(X) = \frac{\theta}{2} = \frac{1}{n} \sum_{i=1}^k x_i$$

Solve  $\theta$

$$\hat{\theta} = \frac{2}{n} \sum_{i=1}^k x_i$$

The solution makes sense when we think about it: The sample mean is just half of  $\theta$

For two parameter distribution we use two moment

**Example 2.** Consider  $x_1, \dots, x_n$  from  $X \sim N(\mu, \sigma^2)$ . Then

$$\begin{aligned}\mathbb{E}(X) &= \mu \\ \mathbb{E}(X^2) &= \mu^2 + \sigma^2\end{aligned}$$

Then

$$\begin{aligned}\hat{\mu} &= \bar{X} \\ \hat{\sigma}^2 &= n^{-1} \sum_{i=1}^n X_i^2 - \bar{X}^2\end{aligned}$$

**Example 3.** Consider  $x_1, \dots, x_n$  from  $X \sim Poi(\lambda)$ . Then

$$\mathbb{E}(X) = \lambda \mathbb{E}(X^2) = \lambda + \lambda^2$$

If we start with considering the second moment instead of the first one. Then

$$\lambda + \lambda^2 = n^{-1} \sum_{i=1}^n X_i^2$$

And solve it with quadratic formula will give something unexpected, as you can imagine. However if we only use first moment everything will be fine.

Notice in above example, MoM didn't give sufficient statistics in two cases (Example 1 and 3).  
However if

1. moment(s) didn't exist (e.g. Cauchy)
2. too difficult to calculate
3. or even being too weird (like Example 3)

Then it is not a good idea to use MoM

### 3 Maximum Likelihood

Consider the *Likelihood function*

$$L(\vec{\theta}) = \prod_{i=1}^n f(x_i; \vec{\theta})$$

**Definition 3.1.** The maximum likelihood estimate of  $\vec{\theta}$  is the value for which the likelihood function attain its maximum. To make life easier, we can also use the log of likelihood function

**Example 4.** Consider  $X \sim \mathbf{N}(\mu, \sigma^2)$  with only  $\mu$  unknown. The likelihood function is

$$L(\mu) = \prod_{i=1}^n \sigma \sqrt{2\pi}^{-1} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) \quad (1)$$

But we can also consider the log likelihood function to make the equation looks better

$$\log L(\mu) = c - (2\sigma^2)^{-1} \sum_{i=1}^n (x_i - \mu)^2$$

for some constant c

Then

$$\frac{d}{d\mu} \log L(\mu) = \sum_{i=1}^n (x_i - \mu) / \sigma^2$$

if  $\frac{d}{d\mu} \log L(\mu) = 0$ , then  $\mu = \bar{x}$ . We can check it is indeed maximum by second derivative test.

**Example 5.** With the same setup as above example but  $\mu$  is also unknown.

Then we have to write the likelihood and log likelihood function in a bit different way

$$L(\mu, \sigma^2) = \prod_{i=1}^n (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$$

$$\log L(\mu, \sigma^2) = c - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Take partial derivatives

$$\frac{\partial}{\partial \mu} \log L(\mu, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)$$

$$\frac{\partial}{\partial \sigma^2} \log L(\mu, \sigma^2) = \frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2$$

We get

$$\mu = \bar{x}$$
$$\sigma^2 = n^{-1} \sum_{i=1}^n (x_i - \mu)^2$$