

# Statistical Inference Note

May 14, 2023

## Notation

1.  $\chi$  - sample space
2.  $X$  - random variable
3.  $F_X(x)$  - cdf of  $X$
4.  $f_X(x)$  - pdf of  $X$
5.  $X = (X_1, \dots, X_n)$  -  $X$  is a random sample of size  $n$

## Chapter 5: Properties of a Random Sample

**Definition 1 (Random Sample)** : The random variables  $X_1, \dots, X_n$  are called a random sample of size  $n$  from the population  $f(x)$  if  $X_1, \dots, X_n$  are mutually independent variables and the marginal pdf or pmf of each  $X_i$  is the same  $f(x)$ .  $\{X_i\}$  are called iid rv with pdf or pmf  $f(x)$ .

**Definition 2 (Statistics)** : Let  $X_1, \dots, X_n$  be a random sample of size  $n$  from a population and let  $T(x_1, \dots, x_n)$  be a real-valued or vector-valued function whose domain includes the sample space of  $(X_1, \dots, X_n)$ . Then the random variable or random vector  $Y = T(X_1, \dots, X_n)$  is called a **statistics**. The probability distribution of a statistic  $Y$  is called the sampling distribution of  $Y$ .

**Definition 3 (Sample mean and variance)**

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (\text{Sample mean})$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (\text{Sample variance})$$

$\bar{x}, s^2$  denote to observed values of  $\bar{X}, S^2$

**Result 1** Let  $X_1, \dots, X_n$  be a random sample from a population with mean  $\mu$  and variance  $\sigma^2 < \infty$ . Then

1.  $E\bar{X} = \mu$
2.  $\text{Var}\bar{X} = \frac{\sigma^2}{n}$
3.  $ES^2 = \sigma^2$

**Remark:** The statistics  $\bar{X}$  is unbiased estimator of  $\mu$ .  $S^2$  is unbiased estimator of  $\sigma^2$  due to the  $n-1$  denominator. If we use  $n$  as denominator,  $ES^2$  would be  $\frac{n-1}{n}\sigma^2$ .

# Chapter 7: Point Estimation

Motivation: we want to find a good estimator for  $\theta$  or  $\tau\theta$  using samples from a pdf  $p(x|\theta)$  since  $\theta$  yields knowledge of the entire population.

**Definition 4 (Point Estimator)** A point estimator is any function  $W(X_1 \dots X_n)$  of a sample; that is any statistic is a point estimator.

Remark: When a sample is taken, estimator is a function of the rv  $X_1, \dots, X_n$  while an estimate is a function of realized values  $X_1, \dots, X_n$ .

## Method of Finding Estimator

### 0.0.1 Method of Moments

**Result 2 (Method of Moments)** Let  $X_1, \dots, X_n$  be a sample from population  $f(x|\theta_1, \dots, \theta_n)$ . The method of moments estimators are found by equating the first  $k$  sample moments ( $m_k = \frac{1}{n} \sum_{i=1}^n X_i^k$ ) to the corresponding  $k$  population moments ( $\mu'_k = EX^k$ )

Example: Suppose  $X_1, \dots, X_n$  are iid from  $f(x|\theta, \sigma^2)$ , we have sample moment  $m_1 = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$ ,  $m_2 = \frac{1}{n} \sum_{i=1}^n X_i^2$  and population moment  $\mu'_1 = EX = \theta$ ,  $\mu'_2 = EX^2 = \theta^2 + \sigma^2$ .

Then we have

$$\begin{aligned}\theta &= \bar{X} & \rightarrow \theta &= \frac{1}{n} \sum_{i=1}^n X_i \\ \theta^2 + \sigma^2 &= \frac{1}{n} \sum_{i=1}^n X_i^2 & \rightarrow \sigma^2 &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \theta^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2\end{aligned}$$

### 0.0.2 Maximum Likelihood Estimator

**Definition 5 (Likelihood function)** If  $X_1, \dots, X_n$  is an iid sample from a population  $f(x|\theta_1, \dots, \theta_n)$ , the likelihood function is defined by

$$L(\theta|x) = L(\theta_1, \dots, \theta_k|x_1, \dots, x_n) = \prod_{i=1}^n f(x_i|\theta_1, \dots, \theta_k)$$

**Definition 6** For each sample point  $x$ , let  $\hat{\theta}(x)$  be a parameter value at which  $L(\theta|x)$  attains its maximum as a function of  $\theta$  with  $x$  held fixed. A maximum likelihood estimator(MLE) of the parameter  $\theta$  based on a sample  $X$  is  $\hat{\theta}(X)$

If the likelihood function is differentiable wrt  $\theta_i$ , the candidate extrema are  $\frac{\partial}{\partial \theta_i} L(\theta|x) = 0$  (Extrema can occur on boundary, we need to check those as well).

Remark: The drawbacks are:

1. The problem of actually finding the global maximum and verifying it.
2. Numerical sensitivity; how sensitive is the estimate to the change in data. This can occur when the likelihood function is very flat in the neighborhood of its maximum or when there is no finite maximum. When using numerical methods, spend some time investigating the stability of the solution.

**Result 3 ( invariance property of MLE )** If  $\hat{\theta}$  is the MLE of  $\theta$ , then for any function  $\tau$ ,  $\tau(\hat{\theta})$  is the MLE of  $\tau(\theta)$ .

### 0.0.3 Bayes Estimators

In the Bayesian approach,  $\theta$  is not thought to be fixed but a quantity whose variation can be described by a probability distribution (**the prior distribution**).

**Definition 7 (Prior distribution)** *is a subjective distribution based on experimenter's belief and is formulated before the data are seen. Denoted as  $\pi(\theta)$*

**Definition 8 (Posterior distribution)** *is the updated prior distribution with the information after a sample taken from the population. Denoted as  $\pi(\theta|x)$ .*

The prior distribution and posterior distribution are related by Bayesian rule:

$$\pi(\theta|x) = \frac{f(x, \theta)}{m(x)} = \frac{f(x|\theta)\pi(\theta)}{m(x)}$$

where  $m(x) = \int f(x|\theta)\pi(\theta)d\theta$

Remark: Posterior distribution is a conditional distribution based on observing the sample. It can be used to make statements about  $\theta$ . E.g. the mean of posterior distribution can be used as a point estimate of  $\theta$ .