

# PH 712 Probability and Statistical Inference

## Part IV: Point Estimation I

Zhiyang Zhou (zhou67@uwm.edu, zhiyanggeezhou.github.io)

2025/Oct/21 21:13:35

---

### Framework of statistical inference/learning

- Goal: infer/learn the distribution of RV  $X$ , say  $f_X$ , from a random sample  $X_1, \dots, X_n$
- Assumption:  $f_X \approx \hat{f}_X$  (statistical model)
  - E.g.,  $\hat{f}_X = \mathcal{N}(\mu, \sigma^2)$ , reducing the task to estimating  $(\mu, \sigma)$
- Point estimation: make the “best” guess about unknown parameter(s)
  - E.g., estimate  $(\mu, \sigma)$  by  $(\hat{\mu}, \hat{\sigma})$
- Hypothesis testing
  - E.g., confirm whether  $\mu = 0$  by testing  $H_0 : \mu = 0$  vs.  $H_1 : \mu \neq 0$
- Interval estimation: construct an interval likely to cover the unknown parameter
  - E.g., construct an interval, say  $(c_1, c_2)$ , such that  $c_1 < \mu < c_2$  with a high probability

### Point estimation

- $\theta$ : the unknown parameter
  - A unknown scalar (i.e., we only consider cases with one unknown parameter)
- The generation of a guess on the value of  $\theta$  based on the random sample  $X_1, \dots, X_n$
- Estimator: the generated guess, say  $\hat{\theta}$ 
  - A statistic (why?) and hence an RV
  - E.g., sometimes,  $\bar{X} = n^{-1} \sum_{i=1}^n X_i$  (sample mean) is an estimator of certain parameter  $\theta$
- Estimate: plugging the realization of the random sample, say  $x_1, \dots, x_n$ , into the estimator
  - A number (why?) and NOT randomized
  - E.g.,  $n^{-1} \sum_{i=1}^n x_i$  is an estimate of certain parameter  $\theta$

### Maximum Likelihood (ML) Estimator (MLE)

- $\Theta$ : the set of allowed values of  $\theta$
- Likelihood function: an alias of joint pdf/pmf

$$L(\theta) = L(\theta; X_1, \dots, X_n) = f_{X_1, \dots, X_n}(X_1, \dots, X_n | \theta), \quad \theta \in \Theta$$

- $f_{X_1, \dots, X_n}$ : the joint pdf/pmf of  $X_1, \dots, X_n$

- Log-likelihood function: the natural logarithm of likelihood function

$$\ell(\theta) = \ln L(\theta), \quad \theta \in \Theta$$

- $\hat{\theta}_{\text{ML}}$  is the MLE for  $\theta$  if  $\hat{\theta}_{\text{ML}}$  is the maximizer of  $L(\theta)$  (equiv. the maximizer of  $\ell(\theta)$ ) with respect to  $\theta$  constrained in  $\Theta$ 
  - In the math notation,
$$\hat{\theta}_{\text{ML}} = \arg \max_{\theta \in \Theta} L(\theta) = \arg \max_{\theta \in \Theta} \ell(\theta)$$
  - That is to say,  $L(\hat{\theta}_{\text{ML}}) \geq L(\theta)$  and  $\ell(\hat{\theta}_{\text{ML}}) \geq \ell(\theta)$ , for all  $\theta \in \Theta$ .
- Invariance property of MLE: if  $\hat{\theta}_{\text{ML}}$  is the MLE of  $\theta$ , then  $g(\hat{\theta}_{\text{ML}})$  is the MLE of  $g(\theta)$  for any given function  $g(\cdot)$ .

### How to locate the ML estimator (MLE) constrained in $\Theta$ ?

- If  $L(\theta)$  (or equiv.  $\ell(\theta)$ ) is monotonic with respect to  $\theta \in \Theta$ , then the MLE lies at one boundary point of  $\Theta$
- If  $\ell(\theta)$  is non-monotonic but differentiable with respect to  $\theta \in \Theta$ , then
  1. Collect all the candidates including:
    - Stationary points, i.e., solutions to the equation  $S(\theta) = 0$  subject to  $\theta \in \Theta$ 
      - \* Where  $S(\theta) = \ell'(\theta)$  is called the score/gradient
    - Boundary points of  $\Theta$
  2. Compare the values of log-likelihood or likelihood evaluated at all the above candidates

### How to locate the ML estimate constrained in $\Theta$ ?

- Reachable only when the realization of  $X_1, \dots, X_n$  are available
- Theoretical way: figuring out the MLE before plugging the realization of  $X_1, \dots, X_n$  into the MLE
- Numerical way: R function `optim()`

### Example Lec5.1

- Suppose  $X_1, \dots, X_n$  is an iid sample following  $\mathcal{N}(\mu, \sigma^2)$ , i.e.,  $f_{X_i}(x | \theta) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$ ,  $x \in \mathbb{R}$ , with unknown  $\mu$  and known  $\sigma > 0$ .
  - Find the MLE of  $\mu$ .
  - If the realization of the sample is  $1, \dots, 10$ , find the ML estimate of  $\mu$ .

---
- Suppose  $X_1, \dots, X_n$  is an iid sample following  $\mathcal{N}(\mu, \sigma^2)$ , i.e.,  $f_{X_i}(x | \theta) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$ ,  $x \in \mathbb{R}$ , with known  $\mu$  and unknown  $\sigma > 0$ .
  - Find the MLE of  $\sigma$ .
  - If the realization of the sample is  $1, \dots, 10$ , find the ML estimate of  $\sigma$ .

---
- Suppose  $X_1, \dots, X_n$  is an iid sample following  $p_{X_i}(x | \theta) = \theta^x(1-\theta)^{1-x}\mathbf{1}_{\{0,1\}}(x)$ ,  $\theta \in [0, 1/2]$ .
  - Find the MLE of  $\theta$ .
  - If the realization of the sample is  $0, 1, 1, 0, 0$ , find the ML estimate of  $\theta$ .

---
- Suppose  $X_1, \dots, X_n$  is an iid sample following an exponential distribution, i.e.,  $f_X(x | \beta) = \beta^{-1} \exp(-x/\beta)\mathbf{1}_{(0,\infty)}(x)$ ,  $\beta > 0$ .

- Find the MLE of  $\beta$ .
  - If the realization of the sample is  $1, \dots, 10$ , find the ML estimate of  $\beta$ .
- 

- Suppose  $X_1, \dots, X_n$  is an iid sample following a beta distribution, i.e.,  $f_X(x | \theta) = \theta x^{\theta-1} \mathbf{1}_{[0,1]}(x)$ ,  $\theta > 0$ .
    - Find the MLE of  $\theta$ .
    - If the realization of the sample is  $0.1, \dots, 0.9$ , find the ML estimate of  $\theta$ .
- 

- The simplest linear model (or linear regression) is a collection of independent random variables  $Y_1, \dots, Y_n$  such that

$$Y_i = \beta x_i + \varepsilon_i, \quad i = 1, \dots, n,$$

where  $x_1, \dots, x_n$  are nonrandomized, and  $\varepsilon_1, \dots, \varepsilon_n \stackrel{\text{iid}}{\sim} f_\varepsilon(t) = \sqrt{2\pi} \exp(-t^2/2)$  (i.e.,  $\mathcal{N}(0, 1)$ ).

- Find the MLE of  $\beta$ , say  $\hat{\beta}_{\text{ML}}$ . (Hint: create the likelihood by noting that  $Y_i \sim \mathcal{N}(\beta x_i, 1)$ .)
- Suppose  $x$ -values are  $1, \dots, 10$ . Correspondingly, observed  $Y$ -values are  $2, \dots, 11$ . Find the ML estimate of  $\beta$ .