

# STAT 3100 Lecture Note

Week Four (Sep 27 & 29, 2022)

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca, zhiyanggeezhou.github.io)

2022/Sep/29 01:19:50

---

## Estimating equations

### Parametric models

- A parametric model is a set of distributions indexed by unknown  $\theta \in \Theta \subset \mathbb{R}^p$  with small or moderate  $p$ 
  - Say  $\{f(\cdot | \theta) : \theta \in \Theta \subset \mathbb{R}^p\}$ , where  $f$  is either a pdf or a pmf and  $\Theta$  is the set of all the possible values of  $\theta$
- Believed that the true parameter (vector)  $\theta_0 (\in \Theta \subset \mathbb{R}^p)$  is fixed
  - Rather than making  $\theta_0$  random in the Bayesian philosophy

### Method of moments (MOM, CB Sec 7.2.1)

- Procedure
  1. Equate raw moments to their empirical counterparts.
  2. Solve the resulting simultaneous equations for  $\theta = (\theta_1, \dots, \theta_p)$ .
- Features
  - Easy implementation
  - Start point for more complex methods
  - No constraint
  - Not uniquely defined
  - No guarantee on optimality

### Exercise Lec7.1

- Let  $X_1, \dots, X_n$  iid follow the following distributions. Find MOM estimators for  $(\theta_1, \theta_2)$ .
  - a.  $N(\theta_1, \theta_2)$ ,  $(\theta_1, \theta_2) \in \mathbb{R} \times \mathbb{R}^+$ .
  - b.  $\text{Binom}(\theta_1, \theta_2)$  with pmf

$$p_X(x | \theta_1, \theta_2) = \binom{\theta_1}{x} \theta_2^x (1 - \theta_2)^{\theta_1 - x} \mathbf{1}_{\{0, \dots, \theta_1\}}(x), \quad (\theta_1, \theta_2) \in \mathbb{Z}^+ \times (0, 1).$$

### Exercise Lec7.2

- Let  $X_1, \dots, X_n$  iid follow pdf  $f(x | \theta) = \theta x^{\theta-1} \mathbf{1}_{[0,1]}(x)$ ,  $\theta > 0$ .
  - a. Find an MOM estimator of  $\theta$ .
  - b. Can we employ the second (raw) moment instead of the first one?

## Maximum Likelihood Estimator (MLE, CB Sec 7.2.2)

- Likelihood function:  $L : \Theta \rightarrow \mathbb{R}$  such that, given  $\mathbf{x}$  (a realization of  $\mathbf{X}$ ),

$$L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}; \mathbf{x}) = f_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\theta}),$$

where  $f_{\mathbf{X}}$  is the joint pdf or pmf.

- For each  $\mathbf{x}$ , let  $\hat{\boldsymbol{\theta}}(\mathbf{x})$  be the maximizer of  $L(\boldsymbol{\theta}; \mathbf{x})$  (or log-likelihood  $\ell(\boldsymbol{\theta}; \mathbf{x}) = \ln L(\boldsymbol{\theta}; \mathbf{x})$ ) with respect to  $\boldsymbol{\theta}$  constrained in  $\Theta$ , i.e.,

$$\hat{\boldsymbol{\theta}}(\mathbf{x}) = \arg \max_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}; \mathbf{x}) = \arg \max_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}; \mathbf{x}).$$

Then the statistic  $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{X})$  is the MLE for  $\boldsymbol{\theta} \in \Theta$ .

- Invariance property of MLE (CB Thm 7.2.10): As long as  $\hat{\boldsymbol{\theta}}$  is the MLE of  $\boldsymbol{\theta}$ , for ANY function  $g$ , the  $g(\hat{\boldsymbol{\theta}})$  is the MLE of  $g(\boldsymbol{\theta})$ .

- 
- If  $\ell$  is differentiable, the score function  $\mathbf{S}$  is defined as its gradient

$$\mathbf{S}(\boldsymbol{\theta}) = \mathbf{S}(\boldsymbol{\theta}; \mathbf{x}) = \left[ \frac{\partial}{\partial \theta_1} \ell(\boldsymbol{\theta}; \mathbf{x}), \dots, \frac{\partial}{\partial \theta_p} \ell(\boldsymbol{\theta}; \mathbf{x}) \right]^\top.$$

- If  $\ell$  is twice differentiable, we have hessian of  $\ell(\boldsymbol{\theta}; \mathbf{x})$

$$\mathbf{H}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta}; \mathbf{x}) = \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell(\boldsymbol{\theta}; \mathbf{x}) \right]_{p \times p}.$$

- 
- Maximizing differentiable  $\ell(\boldsymbol{\theta})$  with  $\boldsymbol{\theta} \in \Theta$ 
    1. Find out stationary points, i.e., solutions to simultaneous equations  $\mathbf{S}(\boldsymbol{\theta}) = \mathbf{0}$
    2. Determine the global maximizer within  $\Theta$ : by comparing values of likelihood (or log-likelihood) evaluated at stationary points and boundary points of  $\Theta$

## Exercise Lec7.3

- Suppose  $X_1, \dots, X_n$  are iid as the following distributions. Find MLEs for corresponding parameters.
  - a.  $N(\mu, \sigma^2)$ ,  $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+$ .
  - b. Bernoulli( $\theta$ ):  $p(x \mid \theta) = \theta^x (1 - \theta)^{1-x} \mathbf{1}_{\{0,1\}}(x)$ ,  $\theta \in [0, 1/2]$ .
  - c. Two-parameter exponential:  $f(x \mid \alpha, \beta) = \beta^{-1} \exp\{-(x - \alpha)/\beta\} \mathbf{1}_{(\alpha, \infty)}(x)$ ,  $(\alpha, \beta) \in \mathbb{R} \times \mathbb{R}^+$ .

## Other examples of estimating equations

- Least-squares estimator
- Generalized estimating equations (GEE)
- M-estimator

## Evaluating estimators

### Mean squared error (MSE)

- Univariate:  $E(\hat{\theta} - \theta_0)^2 = \{E(\hat{\theta}) - \theta_0\}^2 + \text{var}(\hat{\theta})$
- Multivariate:  $E\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\} = \{E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}_0\}^\top \{E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}_0\} + \text{cov}(\hat{\boldsymbol{\theta}})$

- Best unbiased estimator (a.k.a. (uniform) minimum variance unbiased estimator, abbr. UMVUE/MVUE): if  $\hat{\theta} = \hat{\theta}(\mathbf{X})$  satisfies that
  - $\hat{\theta}$  is unbiased for  $\theta$ , i.e.,  $E(\hat{\theta}) = \theta$ ;
  - $\text{var}(\hat{\theta}) - \text{var}(\hat{\theta}^*) \leq 0$  for all  $\theta \in \Theta$  and all  $\hat{\theta}^*$  such that  $E(\hat{\theta}^*) = \theta$ .
- UMVUE is unique (CB Thm 7.3.19)

### Cramer-Rao lower bound (CB Thm 7.3.9 & Lemma 7.3.11)

- Only consider the univariate case, i.e., one-dimensional unknown parameter  $\theta$
- Fisher information:  $I(\theta) = \text{var}(S(\theta; \mathbf{X})) = E[\{S(\theta; \mathbf{X})\}^2] = -E\{H(\theta; \mathbf{X})\}$ 
  - score function  $S(\theta; \mathbf{X})$  and Hessian  $H(\theta; \mathbf{X})$  both scalar
- Cramer-Rao lower bound:  $\text{var}(\hat{\theta}) \geq \{(d/d\theta)E(\hat{\theta})\}^2 / I(\theta)$  for  $\hat{\theta}$  satisfying regularity conditions
- (CB Coro 7.3.15)  $\hat{\theta}$  attains the lower bound  $\Leftrightarrow \exists a(\theta)$  s.t.  $S(\theta; \mathbf{X}) = a(\theta)\{\hat{\theta} - E(\hat{\theta})\}$
- The unbiased  $\hat{\theta}$  attaining the lower bound is UMVUE.

### Example Lec8.1

- Find the lower bound for unbiased estimators for  $\sigma^2$  in the following cases.
  - $X_1, \dots, X_n \sim N(\mu, \sigma^2)$  with known  $\mu$  and unknown  $\sigma^2$ .
  - $X_1, \dots, X_n \sim N(\mu, \sigma^2)$  with unknown  $(\mu, \sigma^2)$ .

### Sufficiency (CB Sec 6.2.1)

- A statistic  $\mathbf{T} = \mathbf{T}(\mathbf{X})$  is sufficient for  $\theta = (\theta_1, \dots, \theta_p) \Leftrightarrow$  the distribution of  $\mathbf{X}$  conditioning on  $\mathbf{T}$  and  $\theta$ , say  $f_{\mathbf{X}|\mathbf{T},\theta}(\mathbf{x} | \mathbf{t}, \theta)$ , is free of  $\theta$ .
- Fisher-Neyman factorization theorem (CB Thm 6.2.6; HMC Thm 7.2.1):  $\mathbf{T}$  is sufficient for  $\theta \Leftrightarrow$  the likelihood function can be factored into two parts, one of them not depending on  $\theta$ , i.e.,

$$L(\theta; \mathbf{x}) = h(\mathbf{x})g(\mathbf{T}(\mathbf{x}), \theta),$$

for all the possible values of  $\mathbf{x}$  and  $\theta$ .

- (HMC Thm 7.3.2) If  $\mathbf{T}$  is sufficient for  $\theta$  and  $\hat{\theta}$  is the unique MLE of  $\theta$ , then  $\hat{\theta}$  must be a function of  $\mathbf{T}$ .

- 
- Nonuniqueness
    - Trivial examples
      - \*  $\mathbf{X}$  is always sufficient.
      - \*  $(X_{(1)}, \dots, X_{(n)})$  is always sufficient if  $X_i$ 's are iid, with  $X_{(1)} \leq \dots \leq X_{(n)}$ .
    - $\mathbf{T}$  is sufficient and  $g(\cdot)$  is a one-to-one mapping  $\Rightarrow g(\mathbf{T})$  is also sufficient.
  - Minimal sufficiency: a sufficient statistic that is a function of all the other sufficient statistics.
    - How to find a minimal sufficient statistic (CB Thm 6.2.13):
      1. Find the sufficient and necessary condition for  $L(\theta; \mathbf{x})/L(\theta; \mathbf{y})$  to be free of  $\theta$ ;
      2. If the condition is of the form  $\mathbf{T}(\mathbf{x}) = \mathbf{T}(\mathbf{y})$ , then  $\mathbf{T}(\mathbf{X})$  is a minimal sufficient statistic for  $\theta$ .

### Example Lec8.2

- Find the minimal sufficient statistics in the following scenarios.
  - $X_1, \dots, X_n \sim \text{Unif}(1, \dots, \theta)$  with unknown positive integer  $\theta$ .
  - $X_1, \dots, X_n \sim N(\mu, \sigma^2)$  with unknown  $\mu$  and  $\sigma^2$ .