STAT 3100 Lecture Note

Week Four (Sep 27 & 29, 2022)

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

2022/Sep/29 12:53:06

Estimating equations

Parametric models

- A parametric model is a set of distributions indexed by unknown $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$ with small or moderate p
  - Say $\{f(\cdot \mid \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p\}$, where f is either a pdf or a pmf and $\boldsymbol{\Theta}$ is the set of all the possbile values of $\boldsymbol{\theta}$
- Believed that the true parameter (vector) $\boldsymbol{\theta}_0$ ($\in \boldsymbol{\Theta} \subset \mathbb{R}^p$) is fixed
  - Rather than making $\boldsymbol{\theta}_0$ 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 $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$.
- Features
  - Easy implementation
  - Start point for more complex methods
  - No constraint
  - Not uniquely defined
  - No guarantee on optimality

Exercise Lec7.1

- Let $X_1, \ldots, X_n$ iid follow the following distributions. Find MOM estimators for (θ_1, θ_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

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

Exercise Lec7.2

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

Maximum Likelihood Estimator (MLE, CB Sec 7.2.2)

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

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

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

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

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

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

- 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 ℓ is differetiable, the score funtion $\mathbf{S}$ is defined as its gradient

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

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

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

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

Exercise Lec7.3

- Suppose $X_1, \ldots, X_n$ 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(θ): $p(x \mid \theta) = \theta^x (1-\theta)^{1-x} \mathbf{1}_{\{0,1\}}(x)$, $\theta \in [0, 1/2]$.
  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