

STAT 4100 Lecture Note

Week Eleven (Nov 21, 23, & 25, 2022)

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

2022/Nov/20 10:11:36

Asymptotic properties of MLE (con'd)

Consistency of MLE (univariate case, CB Thm 10.1.6)

- $\hat{\theta}_{\text{ML}} \xrightarrow{P} \theta$ under regularity conditions (CB Sec 10.6.2):
 - X_1, \dots, X_n are iid;
 - If $\theta_1 \neq \theta_2$, then $f_X(x | \theta_1) \neq f_X(x | \theta_2)$;
 - $f_X(x | \theta)$ has a common support for each $\theta \in \Theta$ and is differentiable with respect to θ ;
 - * Violated by, e.g., $\text{Unif}(0, \theta)$;
 - true θ is an interior point of parameter space Θ .

Example of inconsistent MLE

There are independent $X_{i1}, X_{i2} \sim \mathcal{N}(\mu_i, \sigma^2)$, $i = 1, \dots, n$. Then $\widehat{\sigma^2}_{\text{ML}}$ is NOT consistent for σ^2 .

Examples of consistent MLE even when the regularity conditions are violated

- iid $X_1, \dots, X_n \sim \text{Ber}(1)$
- iid $X_1, \dots, X_n \sim \text{Unif}(0, \theta)$

Efficiency

- (HMC Def 6.2.2, efficiency) For an estimator, say T_n , unbiased for $\tau(\theta)$, the efficiency of T_n is the ratio of CRLB to the actual variance of T_n , i.e., $[\{\tau'(\theta)\}^2 / I_n(\theta)] / \text{var}(T_n | \theta)$.
 - The higher efficiency the better.
 - (HMC Def 6.2.1, efficient) If the efficiency of T_n is 1, then T_n is called an efficient estimator.
- (CB Def 10.1.11, asymptotically efficient) An estimator T_n for $\tau(\theta)$ is asymptotically efficient $\iff \sqrt{n}\{T_n - \tau(\theta)\} \xrightarrow{d} \mathcal{N}(0, \{\tau'(\theta)\}^2 / I(\theta))$, where
$$I(\theta) = -\text{E} \left\{ \frac{\partial^2}{\partial \theta^2} \ln f(X_i | \theta) \mid \theta \right\} = -\text{E}\{H(\theta; X_i) \mid \theta\}$$
is the Fisher information of one single observation.
- (CB Def 10.1.16, asymptotic relative efficiency, ARE) Denote by T_n and W_n two estimators for $\tau(\theta)$. Suppose that $\sqrt{n}(T_n - \tau(\theta)) \xrightarrow{d} N(0, \sigma_T^2)$ and $\sqrt{n}(W_n - \tau(\theta)) \xrightarrow{d} N(0, \sigma_W^2)$. The asymptotic relative efficiency (ARE) of T_n with respect to W_n is defined as

$$\text{ARE}(T_n, W_n) = \sigma_W^2 / \sigma_T^2.$$

- $\text{ARE}(T_n, W_n) > 1 \iff T_n$ is asymptotically more efficient than W_n

Asymptotic efficiency of MLE (CB Thm 10.1.12 & Ex. 10.7)

- $\sqrt{n}\{\tau(\hat{\theta}_{\text{ML}}) - \tau(\theta)\} \xrightarrow{d} \mathcal{N}(0, \{\tau'(\theta)\}^2 / I(\theta))$, provided that $\hat{\theta}_{\text{ML}}$ is the MLE for θ , τ is differentiable and we have the previous four regularity conditions (for the consistency of MLE) plus the following two more (CB Sec 10.6.2):
 - For each $x \in \text{supp}(X)$, $f(x | \theta)$ is three times continuously differentiable with respect to θ ; and $\int f(x | \theta) dx$ can be differentiated three times under the integral sign;
 - for each $\theta_0 \in \Theta$, there exists $c(\theta_0) > 0$ and $M(x, \theta_0)$ such that $|\frac{\partial^3}{\partial \theta^3} \ln f_X(x | \theta)| \leq M(x, \theta_0)$ for all $x \in \text{supp}(X)$ and $\theta \in (\theta_0 - c(\theta_0), \theta_0 + c(\theta_0))$.
- In practice,
 - $nI(\theta) = I_n(\theta) \approx I_n(\hat{\theta}_{\text{ML}}) \approx \hat{I}_n(\hat{\theta}_{\text{ML}})$
 - * (Expected) Fisher information (number) $I_n(\theta) = -E\{H(\theta; \mathbf{X}) | \theta\}$
 - * Observed Fisher information (number) $\hat{I}_n(\hat{\theta}_{\text{ML}}) = -\frac{\partial^2}{\partial \theta^2} \ln L(\theta; \mathbf{X})|_{\theta=\hat{\theta}_{\text{ML}}} = -H(\hat{\theta}_{\text{ML}}; \mathbf{X})$
 - Hence $\text{var}(\tau(\hat{\theta}_{\text{ML}})) \approx \{\tau'(\theta)\}^2 / I_n(\theta) \approx \{\tau'(\hat{\theta}_{\text{ML}})\}^2 / I_n(\hat{\theta}_{\text{ML}}) \approx \{\tau'(\hat{\theta}_{\text{ML}})\}^2 / \hat{I}_n(\hat{\theta}_{\text{ML}})$

CB Example 10.1.17 & Ex. 10.9

- iid $X_1, \dots, X_n \sim p(x | \lambda) = \lambda^x \exp(-\lambda) / x!$, $x \in \mathbb{Z}^+$, $\lambda > 0$. To estimate $\Pr(X_i = 0) = \exp(-\lambda)$.
 - Consider $T_n = n^{-1} \sum_i \mathbf{1}_{\{0\}}(X_i)$ and MLE $W_n = \exp(-\bar{X}_n)$. Compute $\text{ARE}(T_n, W_n)$, the ARE of T_n with respect to W_n .
 - Find the UMVUE for $\Pr(X_i = 0)$, say U_n , and then calculate $\text{ARE}(U_n, W_n)$.

Approximation to variances

Delta method

- (CB Thm 5.5.24, delta method) If $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ and $\tau'(\theta) \neq 0$, then

$$\sqrt{n}\{\tau(\hat{\theta}_n) - \tau(\theta)\} \xrightarrow{d} \mathcal{N}(0, \{\tau'(\theta)\}^2 \sigma^2).$$
 - Hence $\text{var}\{\tau(\hat{\theta}_n)\} \approx \{\tau'(\theta)\}^2 \sigma^2 / n$ if $\tau'(\theta) \neq 0$
- (CB Thm 5.5.26, second-order delta method): If $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$, $\tau'(\theta) = 0$, and $\tau''(\theta) \neq 0$, then

$$n\{\tau(\hat{\theta}_n) - \tau(\theta)\} \xrightarrow{d} \frac{\tau''(\theta)\sigma^2}{2} \chi^2(1).$$
 - Hence $\text{var}\{\tau(\hat{\theta}_n)\} \approx \{\tau''(\theta)\}^2 \sigma^4 / (2n^2)$ if $\tau'(\theta) = 0$ but $\tau''(\theta) \neq 0$

CB Example 10.1.15

- Holding iid $X_i \sim \text{Bernoulli}(p)$, the variance of $\text{Bernoulli}(p)$ is $\tau(p) = p(1-p)$ whose MLE is $\tau(\hat{p}_{\text{MLE}}) = \bar{X}_n(1 - \bar{X}_n)$. Approximate $\text{var}\{\tau(\hat{p}_{\text{MLE}})\}$ by the delta method.

Bootstrapping the variance of $\hat{\theta} = \hat{\theta}(\mathbf{X})$ (CB Sec. 10.1.4)

- Nonparametric bootstrap:
 - For j in $1 : B$, do steps a–b.
 - Draw the j th resample \mathbf{x}_j^* of size n from the original sample $\mathbf{x} = \{x_1, \dots, x_n\}$, with replacement.
 - Let $\hat{\theta}_j^* = \hat{\theta}(\mathbf{x}_j^*)$.
 - Update j with $j + 1$.
 - $\text{var}\{\hat{\theta}(\mathbf{X})\} \approx$ the sample variance of $\{\hat{\theta}_1^*, \dots, \hat{\theta}_B^*\}$.
- Parametric bootstrap:
 - For j in $1 : B$, do steps a–b.

- a. Draw the j th resample \mathbf{x}_j^* of size n from a fitted model $f(x \mid \hat{\theta})$.
- b. Let $\hat{\theta}_j^* = \hat{\theta}(\mathbf{x}_j^*)$.
2. Update j with $j + 1$.
3. $\text{var}\{\hat{\theta}(\mathbf{X})\} \approx$ the sample variance of $\{\hat{\theta}_1^*, \dots, \hat{\theta}_B^*\}$.

CB Example 10.1.15

- Holding iid $X_i \sim \text{Bernoulli}(p)$, the variance of $\text{Bernoulli}(p)$ is $\tau(p) = p(1 - p)$ for which the MLE is $\tau(\hat{p}_{\text{mle}}) = \bar{X}_n(1 - \bar{X}_n)$. Approximate $\text{var}\{\tau(\hat{p}_{\text{mle}})\}$ by the bootstrap.

Take-home exercises (NOT to be submitted; to be potentially covered in labs)

- CB Ex. 10.3, 10.17(a-c)