STAT 4100 Lecture Note

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

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

2022/Nov/21 18:22:48

Asymptotic properties of MLE (con'd)

Consistency of MLE (univariate case, CB Thm 10.1.6)

- $\hat{\theta}_{\mathrm{ML}} \xrightarrow{p} \theta_0$, provided that $\hat{\theta}_{\mathrm{ML}}$ is the MLE for θ_0, under regularity conditions (CB Sec 10.6.2):
  - iid $X_1, \ldots, X_n \sim f(x \mid \theta_0)$;
  - for $\theta_1, \theta_2 \in \Theta$, if $\theta_1 \neq \theta_2$, then $f_X(x \mid \theta_1) \neq f_X(x \mid \theta_2)$;
  - $f_X(x \mid \theta)$ has a common support for each $\theta \in \Theta$ and is differentiable with respect to θ;
    - Violated by, e.g., $\mathrm{Unif}(0, \theta)$;
  - θ_0 is an interior point of parameter space Θ.

Example of inconsistent MLE

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

Examples of consistent MLE with the regularity conditions violated

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

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., $[\{\tau'(\theta)\}^2 / I_n(\theta)] / \mathrm{var}(T_n \mid \theta)$.
  - 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 $\iff$ $\sqrt{n}\{T_n - \tau(\theta)\} \xrightarrow{d} \mathcal{N}(0, \{\tau'(\theta)\}^2 / I_1(\theta))$, where

  $I_1(\theta) = -\mathrm{E}\left\{\frac{\partial^2}{\partial \theta^2} \ln f(X_i \mid \theta) \mid \theta\right\} = -\mathrm{E}\{H(\theta; X_i) \mid \theta\}$ 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

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

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

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

- $\sqrt{n}\{\tau(\hat\theta_{\rm ML}) - \tau(\theta_0)\} \xrightarrow{d} \mathcal{N}(0, \{\tau'(\theta_0)\}^2/I_1(\theta_0))$, provided that $\hat\theta_{\rm ML}$ is the MLE for θ_0, 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 $x \in \operatorname{supp}(X)$, $f(x \mid \theta)$ is three time continuously differentiable with respect to θ; and $\int f(x \mid \theta)\mathrm{d}x$ can be differentiated three times under the integral sign;
  - for each $\theta \in \Theta$, there exists $c(\theta) > 0$ and $M(x, \theta)$ such that $|\frac{\partial^3}{\partial\theta^3} \ln f_X(x \mid \theta)| \le M(x, \theta)$ for all $x \in \operatorname{supp}(X)$ and $\theta \in (\theta - c(\theta), \theta + c(\theta))$.
- In practice,
  - $nI_1(\theta_0) = I_n(\theta_0) \approx I_n(\hat\theta_{\rm ML}) \approx \hat I_n(\hat\theta_{\rm ML})$
    - (Expected) Fisher information (number) $I_n(\theta_0) = -\mathrm{E}\{H(\theta_0; \mathbf{X})\}$
    - Observed Fisher information (number) $\hat I_n(\hat\theta_{\rm ML}) = -\frac{\partial^2}{\partial\theta^2} \ln L(\theta; \boldsymbol{x})\big|_{\theta=\hat\theta_{\rm ML}} = -H(\hat\theta_{\rm ML}; \boldsymbol{x})$
  - Hence $\operatorname{var}(\tau(\hat\theta_{\rm ML})) \approx \{\tau'(\theta_0)\}^2/I_n(\theta_0) \approx \{\tau'(\hat\theta_{\rm ML})\}^2/I_n(\hat\theta_{\rm ML}) \approx \{\tau'(\hat\theta_{\rm ML})\}^2/\hat I_n(\hat\theta_{\rm ML})$

CB Example 10.1.17 & Ex. 10.9

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

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 $\operatorname{var}\{\tau(\hat\theta_n)\} \approx \{\tau'(\hat\theta_n)\}^2\sigma^2/n$ 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 $\operatorname{var}\{\tau(\hat\theta_n)\} \approx \{\tau''(\hat\theta_n)\}^2\sigma^4/(2n^2)$ 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 $\tau(\hat p_{\rm mle}) = \bar X_n(1 - \bar X_n)$. Approximate $\operatorname{var}\{\tau(\hat p_{\rm mle})\}$ by the delta method.

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

- Nonparametric bootstrap:
  1. For j in $1 : B$, do steps a–b.
     a. Draw the jth resample $\boldsymbol{x}_j^*$ of size n from the original sample $\boldsymbol{x} = \{x_1, \ldots, x_n\}$, with replacement.
     b. Let $\hat\theta_j^* = \hat\theta(\boldsymbol{x}_j^*)$.
  2. Update j with $j + 1$.
  3. $\operatorname{var}\{\hat\theta(\mathbf{X})\} \approx$ the sample variance of $\{\hat\theta_1^*, \ldots, \hat\theta_B^*\}$.
- Parametric bootstrap:
  1. For j in $1 : B$, do steps a–b.

a. Draw the jth resample $\boldsymbol{x}_j^*$ of size n from a fitted model $f(x \mid \hat{\theta})$.
b. Let $\hat{\theta}_j^* = \hat{\theta}(\boldsymbol{x}_j^*)$.

2. Update j with $j+1$.
3. $\text{var}\{\hat{\theta}(\mathbf{X})\} \approx$ the sample variance of $\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\}$.

CB Example 10.1.15

- Holding iid $X_i \sim \text{Bernoulli}(p)$, the variance of 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)