

STAT 3100 Lecture Note

Week Twelve (Nov 29 & Dec 1, 2022)

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

2022/Nov/29 01:15:00

Asymptotic properties of MLE (con'd)

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

- $\sqrt{n}(\hat{\theta}_{\text{ML}} - \theta_0) \xrightarrow{d} \mathcal{N}(0, 1/I_1(\theta_0))$, provided that $\hat{\theta}_{\text{ML}}$ is the MLE for θ_0 , 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 $\theta \in \Theta$, there exists $c(\theta) > 0$ and $M(x, \theta)$ such that $|\frac{\partial^3}{\partial \theta^3} \ln f_X(x | \theta)| \leq M(x, \theta)$ for all $x \in \text{supp}(X)$ and $\theta \in (\theta - c(\theta), \theta + c(\theta))$.
- In practice,
 - $nI_1(\theta_0) = I_n(\theta_0) \approx I_n(\hat{\theta}_{\text{ML}}) \approx \hat{I}_n(\hat{\theta}_{\text{ML}})$
 - * (Expected) Fisher information (number) $I_n(\theta_0) = -E\{H(\theta_0; \mathbf{X})\}$
 - * 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 $\text{var}(\hat{\theta}_{\text{ML}}) \approx 1/I_n(\theta_0) \approx 1/I_n(\hat{\theta}_{\text{ML}}) \approx 1/\hat{I}_n(\hat{\theta}_{\text{ML}})$

Delta method

- (CB Thm 5.5.24, delta method) If $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$, τ is NOT a function of n , 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 $\text{var}\{\tau(\hat{\theta}_n)\} \approx \{\tau'(\hat{\theta}_n)\}^2 \sigma^2 / n$ if $\tau'(\theta) \neq 0$
- (CB Thm 5.5.26, second-order delta method) If $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$, τ is NOT a function of n , $\tau'(\theta) = 0$, 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 $\text{var}\{\tau(\hat{\theta}_n)\} \approx \{\tau''(\hat{\theta}_n)\}^2 \sigma^4 / (2n^2)$ if $\tau'(\theta) = 0$ but $\tau''(\theta) \neq 0$

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 $\Pr(X_i = 0) = \exp(-\lambda)$.
 - a. 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 .
 - b. Find the UMVUE for $\Pr(X_i = 0)$, say U_n , and then calculate $\text{ARE}(U_n, W_n)$.
 - Hint: $\sqrt{n}(U_n - W_n) \xrightarrow{P} 0$ (derived from S. Portnoy, *The Annals of Statistics*, 1977, 5, pp. 522–529, Theorem 1) and $\sum_{i=1}^n X_i \sim \text{Poisson}(n\lambda)$

Approximation to the variance of $\hat{\theta}_n$

- Why?
 - Reflect the variation or dispersion of $\hat{\theta}_n$
 - Help approximate the distribution of $\hat{\theta}_n$ (and further construct the confidence region for θ) if assuming normality
- How?
 - Utilizing the asymptotic variance of $\hat{\theta}_n$
 - Resampling methods, e.g., bootstrapping

CB Example 10.1.17 & Ex. 10.9 (con'd)

- iid $X_1, \dots, X_n \sim p(x | \lambda) = \lambda^x \exp(-\lambda)/x!$, $x \in \mathbb{Z}^+$, $\lambda > 0$. Define $\theta = \Pr(X_i = 2 | \theta) = \lambda^2 \exp(-\lambda)/2$. Approximate the variance of $\hat{\theta}_{\text{ML}} = \bar{X}_n^2 \exp(-\bar{X}_n)/2$ by delta methods.

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 delta methods.

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

- Nonparametric bootstrap:
 1. For j in $1 : B$, do steps 2–3.
 2. Draw the j th resample \mathbf{x}_j^* of size n from the original sample $\mathbf{x} = \{x_1, \dots, x_n\}$, with replacement, i.e., create a new iid sample \mathbf{x}_j^* from F_n (the empirical cdf of the original sample)
 3. Let $\hat{\theta}_j^* = \hat{\theta}(\mathbf{x}_j^*)$.
 4. $\text{var}(\hat{\theta}) \approx$ the sample variance of $\{\hat{\theta}_1^*, \dots, \hat{\theta}_B^*\}$.
- Parametric bootstrap:
 1. For j in $1 : B$, do steps 2–3.
 2. Draw the j th resample \mathbf{x}_j^* of size n from a fitted model $f(x | \hat{\theta})$.
 3. Let $\hat{\theta}_j^* = \hat{\theta}(\mathbf{x}_j^*)$.
 4. $\text{var}(\hat{\theta}) \approx$ 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.

Large-sample hypothesis testing

Recall the LRT

- $H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ v.s. $H_1 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_1$, where $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1$
- LRT statistic

$$\lambda(\mathbf{x}) = \frac{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} L(\boldsymbol{\theta}; \mathbf{x})}{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L(\boldsymbol{\theta}; \mathbf{x})} = \frac{L(\hat{\boldsymbol{\theta}}_{0,\text{ML}}; \mathbf{x})}{L(\hat{\boldsymbol{\theta}}_{\text{ML}}; \mathbf{x})}$$

- $\hat{\boldsymbol{\theta}}_{0,\text{ML}}$: constrained MLE for $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$
- $\hat{\boldsymbol{\theta}}_{\text{ML}}$: unconstrained MLE for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$
- $\{\mathbf{x} : \lambda(\mathbf{x}) \leq c_\alpha\}$: rejection region of level α LRT

- c_α is such defined that $\sup_{\theta \in \Theta_0} \Pr(\lambda(\mathbf{X}) \leq c_\alpha \mid \theta) = \alpha$

Asymptotic distribution of LRT statistic (CB Thm 10.3.1 & 10.3.3)

- As $n \rightarrow \infty$,

$$-2 \ln \lambda(\mathbf{X}) \xrightarrow{d} \chi^2(\nu),$$

where ν = difference of numbers of free parameters in Θ_0 and Θ .

- (CB Thm 10.3.3) $\{\mathbf{x} : -2 \ln \lambda(\mathbf{x}) \geq \chi_{\nu, 1-\alpha}^2\}$: asymptotic rejection region of level α LRT
 - $\chi_{\nu, 1-\alpha}^2$ is the $1 - \alpha$ quantile of $\chi^2(\nu)$.

CB Example 10.3.4

- iid $X_1, \dots, X_n \sim \Pr(X_i = j) = p_j, j = 1, \dots, 5$. Specify the $1 - \alpha$ LRT rejection region for $H_0 : p_1 = p_2 = p_3$ and $p_4 = p_5$ versus H_1 : Otherwise.

Wald test (CB pp. 493)

- $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$
 - Wald statistic: $(\hat{\theta}_n - \theta_0) / \sqrt{\text{var}(\hat{\theta}_n)}$ (if $(\hat{\theta}_n - \theta_0) / \sqrt{\text{var}(\hat{\theta}_n)} \xrightarrow{d} \mathcal{N}(0, 1)$ under H_0 as $n \rightarrow \infty$)
 - * Substitute $\widehat{\text{var}}(\hat{\theta}_n)$ for $\text{var}(\hat{\theta}_n)$ if $\text{var}(\hat{\theta}_n)$ is well approximated by $\widehat{\text{var}}(\hat{\theta}_n)$
 - Level α rejection region: $\{\mathbf{x} : |\hat{\theta}_n - \theta_0| / \sqrt{\text{var}(\hat{\theta}_n)} \geq \Phi_{1-\alpha/2}^{-1}\}$
- Asymptotically equivalent to LRT Under H_0 , but merely involve the (unconstrained) MLE.

Score test (CB pp. 494)

- Merely involve the constrained MLE
- $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$
 - Score statistic: $S(\theta_0; \mathbf{X}) / \sqrt{I_n(\theta_0)}$ ($\xrightarrow{d} \mathcal{N}(0, 1)$ under H_0 as $n \rightarrow \infty$)
 - Level α rejection region: $\{\mathbf{x} : |S(\theta_0; \mathbf{x})| / \sqrt{I_n(\theta_0)} \geq \Phi_{1-\alpha/2}^{-1}\}$.
- If Θ_0 contains more than one points, then substitute $\hat{\theta}_{0, \text{ML}}$ for θ_0 .

CB Examples 10.3.5 & 10.3.6

- iid $X_1, \dots, X_n \sim \text{Bernoulli}(p), p \in (0, 1)$. Derive LRT, Wald and score tests for $H_0 : p = p_0$ versus $H_1 : p \neq p_0$.

Asymptotic confidence regions

- Constructed by reverting rejection regions
- Examples
 - $1 - \alpha$ LRT confidence region for θ : $\{\theta : -2 \ln \{L(\theta; \mathbf{x}) / L(\hat{\theta}_{\text{ML}}; \mathbf{x})\} < \chi_{1, 1-\alpha}^2\}$
 - $1 - \alpha$ Wald confidence region for θ : $\{\theta : |\hat{\theta}_{\text{ML}} - \theta| \sqrt{I_n(\theta)} < \Phi_{1-\alpha/2}^{-1}\}$
 - $1 - \alpha$ score confidence region for θ : $\{\theta : |S(\theta; \mathbf{x})| / \sqrt{I_n(\theta)} < \Phi_{1-\alpha/2}^{-1}\}$

CB Examples 10.4.2, 10.4.3 & 10.4.5

- iid $X_1, \dots, X_n \sim \text{Bernoulli}(p)$, construct $1 - \alpha$ confidence intervals for p .

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

- CB Ex. 10.17(a-c), 10.36, 10.38
- HMC Ex. 6.3.16–6.3.18