STAT 3100 Lecture Note

Week Twelve (Nov 29 & Dec 1, 2022)

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

2022/Dec/01 00:29:22

Asymptotic properties of MLE (con'd)

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

- $\sqrt{n}(\hat{\theta}_{\mathrm{ML}} - \theta_0) \xrightarrow{d} \mathcal{N}(0, 1/I_1(\theta_0))$, provided that $\hat{\theta}_{\mathrm{ML}}$ 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 $x \in \mathrm{supp}(X)$, $f(x \mid \theta)$ 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 $|\frac{\partial^3}{\partial\theta^3} \ln f_X(x \mid \theta)| \leq M(x, \theta)$ for all $x \in \mathrm{supp}(X)$ and $\theta \in (\theta - c(\theta), \theta + c(\theta))$.
- In practice,
  - $nI_1(\theta_0) = I_n(\theta_0) \approx I_n(\hat{\theta}_{\mathrm{ML}}) \approx \hat{I}_n(\hat{\theta}_{\mathrm{ML}})$
    * (Expected) Fisher information (number) $I_n(\theta_0) = -\mathrm{E}\{H(\theta_0; \mathbf{X})\}$
    * Observed Fisher information (number) $\hat{I}_n(\hat{\theta}_{\mathrm{ML}}) = -\frac{\partial^2}{\partial\theta^2} \ln L(\theta; \boldsymbol{x})\big|_{\theta=\hat{\theta}_{\mathrm{ML}}} = -H(\hat{\theta}_{\mathrm{ML}}; \boldsymbol{x})$
  - Hence $\mathrm{var}(\hat{\theta}_{\mathrm{ML}}) \approx 1/I_n(\theta_0) \approx 1/I_n(\hat{\theta}_{\mathrm{ML}}) \approx 1/\hat{I}_n(\hat{\theta}_{\mathrm{ML}})$

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 $\mathrm{var}\{\tau(\hat{\theta}_n)\} \approx \{\tau'(\hat{\theta}_n)\}^2\sigma^2/n$ 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 $\mathrm{var}\{\tau(\hat{\theta}_n)\} \approx \{\tau''(\hat{\theta}_n)\}^2\sigma^4/(2n^2)$ if $\tau'(\theta) = 0$ but $\tau''(\theta) \neq 0$

CB Example 10.1.17 & Ex. 10.9

- iid $X_1, \ldots, X_n \sim p(x \mid \lambda) = \lambda^x \exp(-\lambda)/x!$, $x \in \mathbb{Z}^+$, $\lambda > 0$. 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 $\mathrm{ARE}(T_n, W_n)$, 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 $\mathrm{ARE}(U_n, W_n)$.
     - 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 $\sum_{i=1}^n X_i \sim \mathrm{Poisson}(n\lambda)$

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., bootstraping

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

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

CB Example 10.1.15

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

Bootstraping 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 jth resample $\boldsymbol{x}_j^*$ of size n from the original sample $\boldsymbol{x} = \{x_1, \dots, x_n\}$, with replacement, i.e., create a new iid sample $\boldsymbol{x}_j^*$ from F_n (the empirical cdf of the original sample)
  3. Let $\hat{\theta}_j^* = \hat{\theta}(\boldsymbol{x}_j^*)$.
  4. $\mathrm{var}(\hat{\theta}) \approx$ the sample variance of $\{\hat{\theta}_1^*, \dots, \hat{\theta}_B^*\}$.
- (Optional, see, e.g., www.stat.columbia.edu/~bodhi/Talks/Emp-Proc-Lecture-Notes.pdf) Empirical process: theoretical foundation for nonparametric bootstrap
  - (Glivenko-Cantelli) $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \xrightarrow{\text{a.s.}} 0$
  - (Donsker) $\sqrt{n}(F_n - F) \xrightarrow{d} \mathrm{BB} \circ F$, i.e., $\mathrm{E}[g\{\sqrt{n}(F_n - F)\}] \to \mathrm{E}[g(\mathrm{BB} \circ F)]$ for all bounded, continuous and real-valued g
    - BB is a Gaussian process (specifically, standard Brownian bridge process on $[0, 1]$), i.e.,
      - $\mathrm{BB}(0) = \mathrm{BB}(1) = 0$ but $\mathrm{BB}(t) \sim \mathcal{N}(0, t(1-t))$ for $t \in (0, 1)$;
      - fixing $t_1, \dots, t_p \in (0, 1)$, $[\mathrm{BB}(t_1), \dots, \mathrm{BB}(t_p)]^\top$ is of multivariate normal with $\mathrm{cov}(\mathrm{BB}(s), \mathrm{BB}(t)) = \min(s, t) - st$;
      - $\mathrm{BB}(t)$ is continuous in t.
- Parametric bootstrap:
  1. For j in $1:B$, do steps 2–3.
  2. Draw the jth resample $\boldsymbol{x}_j^*$ of size n from a fitted model $f(x \mid \hat{\theta})$.
  3. Let $\hat{\theta}_j^* = \hat{\theta}(\boldsymbol{x}_j^*)$.
  4. $\mathrm{var}(\hat{\theta}) \approx$ 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 Bernoulli(p) is $\tau(p) = p(1-p)$ for which the MLE is $\tau(\hat{p}_{\mathrm{ML}}) = \bar{X}_n(1 - \bar{X}_n)$. Approximate $\mathrm{var}\{\tau(\hat{p}_{\mathrm{ML}})\}$ by the bootstrap.

```r
options(digits = 4)
set.seed(1)
B = 1e4L
```

```r
n = 30
x = rbinom(n, 1, prob = .7)
theta_ml = mean(x)
tau_theta_star_np = numeric(B)
tau_theta_star_p = numeric(B)
# Nonparametric bootstrap
for (j in 1:B){
  x_star = sample(x, size = n, replace = T)
  tau_theta_star_np[j] = mean(x_star)*(1-mean(x_star))
}
var(tau_theta_star_np)
# Parametric bootstrap
for (j in 1:B){
  x_star = rbinom(n, size = 1, prob = theta_ml)
  tau_theta_star_p[j] = mean(x_star)*(1-mean(x_star))
}
var(tau_theta_star_p)
# Estimate via the first-order delta method
theta_ml*(1-theta_ml)*(1-2*theta_ml)^2/n
# Estimate via the second-order delta method
2*theta_ml^2*(1-theta_ml)^2/n^2
```

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

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

  - $\hat{\boldsymbol{\theta}}_{0,\mathrm{ML}}$: constrained MLE for $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$
  - $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$: unconstrained MLE for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$
- $\{\boldsymbol{x} : \lambda(\boldsymbol{x}) \le c_\alpha\}$: rejection region of level α LRT
  - c_α is such defined that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} \Pr(\lambda(\mathbf{X}) \le c_\alpha \mid \boldsymbol{\theta}) = \alpha$

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

- Under H_0, as $n \to \infty$,

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

  where $\nu =$ difference of numbers of free parameters in $\boldsymbol{\Theta}_0$ and $\boldsymbol{\Theta}$.
- (CB Thm 10.3.3) $\{\boldsymbol{x} : -2 \ln \lambda(\boldsymbol{x}) \ge \chi^2_{\nu,1-\alpha}\}$: asymptotic rejection region of level α LRT
  - $\chi^2_{\nu,1-\alpha}$ is the $1-\alpha$ quantile of $\chi^2(\nu)$.

CB Example 10.3.4

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

Wald test (CB pp. 493)

- $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$
  - Wald statistic: $(\hat{\theta}_n - \theta_0)/\sqrt{\mathrm{var}(\hat{\theta}_n)}$ (if $(\hat{\theta}_n - \theta_0)/\sqrt{\mathrm{var}(\hat{\theta}_n)} \xrightarrow{d} \mathcal{N}(0, 1)$ under H_0 as $n \to \infty$)
    - Asymptotically equivalent to LRT for this two sided test if $\hat{\theta}_n = \hat{\theta}_{\mathrm{ML}}$
    - Substitute $\widehat{\mathrm{var}}(\hat{\theta}_n)$ for $\mathrm{var}(\hat{\theta}_n)$ if $\mathrm{var}(\hat{\theta}_n)$ is well approximated by $\widehat{\mathrm{var}}(\hat{\theta}_n)$
  - Level α rejection region: $\{\boldsymbol{x} : |\hat{\theta}_n - \theta_0|/\sqrt{\mathrm{var}(\hat{\theta}_n)} \geq \Phi^{-1}_{1-\alpha/2}\}$

Score test (CB pp. 494)

- $H_0 : \theta = \theta_0$ 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 $n \to \infty$)
  - Level α rejection region: $\{\boldsymbol{x} : |S(\theta_0; \boldsymbol{x})|/\sqrt{I_n(\theta_0)} \geq \Phi^{-1}_{1-\alpha/2}\}$.
- If $\boldsymbol{\Theta}_0$ contains more than one points, then substitute $\hat{\theta}_{0,\mathrm{ML}}$ for θ_0. So the score test at most involves the constrained MLE.

CB Examples 10.3.5 & 10.3.6

- iid $X_1, \ldots, X_n \sim \mathrm{Bernoulli}(p)$, $p \in (0, 1)$. 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: $\{\theta : -2 \ln\{L(\theta; \boldsymbol{x})/L(\hat{\theta}_{\mathrm{ML}}; \boldsymbol{x})\} < \chi^2_{1,1-\alpha}\}$
  - $1 - \alpha$ Wald confidence region for θ: $\{\theta : |\hat{\theta}_n - \theta|/\sqrt{\mathrm{var}(\hat{\theta}_n)} < \Phi^{-1}_{1-\alpha/2}\}$
  - $1 - \alpha$ score confidence region for θ: $\{\theta : |S(\theta; \boldsymbol{x})|/\sqrt{I_n(\theta)} < \Phi^{-1}_{1-\alpha/2}\}$

CB Examples 10.4.2, 10.4.3 & 10.4.5

- iid $X_1, \ldots, X_n \sim \mathrm{Bernoulli}(p)$, 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