

STAT 3100 Lecture Note

Week Four (Oct 4 & 6, 2022)

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

2022/Oct/06 13:14:19

Evaluating estimators

Mean squared error (MSE)

- Univariate: $E(\hat{\theta} - \theta)^2 = \{E(\hat{\theta}) - \theta\}^2 + \text{var}(\hat{\theta})$
- Multivariate: $E\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\} = \{E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}\}^\top \{E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}\} + \text{cov}(\hat{\boldsymbol{\theta}})$
- Best unbiased estimator (i.e., (uniform) minimum variance unbiased estimator, abbr. UMVUE/MVUE): if $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{X})$ satisfies that
 - $\hat{\boldsymbol{\theta}}$ is unbiased for $\boldsymbol{\theta}$, i.e., $E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$
 - $\text{var}(\hat{\boldsymbol{\theta}}) - \text{var}(\hat{\boldsymbol{\theta}}^*) \leq 0$ for all unbiased $\hat{\boldsymbol{\theta}}^*$
- UMVUE is unique (CB Thm 7.3.19)

Cramer-Rao lower bound (CB Thm 7.3.9 & Lemma 7.3.11)

- Only consider the univariate case, i.e., one-dimensional unknown parameter θ
 - Score function $S(\theta; \mathbf{X})$ and Hessian $H(\theta; \mathbf{X})$ both scalar
- Cramer-Rao lower bound: $\text{var}(\hat{\theta}) \geq \{(d/d\theta)E(\hat{\theta})\}^2 / I(\theta)$ for $\hat{\theta}$ satisfying regularity conditions
 - Fisher information: $I(\theta) = \text{var}(S(\theta; \mathbf{X})) = E[\{S(\theta; \mathbf{X})\}^2] = -E\{H(\theta; \mathbf{X})\}$
 - Proof: Applying the Cauchy-Schwarz inequality (CB Thm 4.7.3)
- (CB Coro 7.3.15) $\hat{\theta}$ attains the lower bound \Leftrightarrow there is $a(\theta)$ such that $S(\theta; \mathbf{X}) = a(\theta)\{\hat{\theta} - E(\hat{\theta})\}$
- The unbiased $\hat{\theta}$ attaining the lower bound is UMVUE

Example Lec8.1

- Find the lower bound for unbiased estimators for σ^2 in the following cases.
 - a. iid $X_1, \dots, X_n \sim \mathcal{N}(\mu_0, \sigma^2)$ with known μ_0 and unknown σ^2 .

Sufficiency (CB Sec 6.2.1)

- A statistic $\mathbf{T} = \mathbf{T}(\mathbf{X})$ is sufficient for $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p) \Leftrightarrow$ the distribution of \mathbf{X} conditioning on \mathbf{T} and $\boldsymbol{\theta}$, say $f_{\mathbf{X}|\mathbf{T},\boldsymbol{\theta}}(\mathbf{x} | \mathbf{t}, \boldsymbol{\theta})$, is free of $\boldsymbol{\theta}$.

- Fisher-Neyman factorization theorem (CB Thm 6.2.6; HMC Thm 7.2.1): \mathbf{T} is sufficient for $\theta \Leftrightarrow$ the likelihood function can be factored into two parts, one of them not depending on θ , i.e.,

$$L(\theta; \mathbf{x}) = f_{\mathbf{X}}(\mathbf{x} | \theta) = h(\mathbf{x})g(\mathbf{T}(\mathbf{x}), \theta), \text{ for all } \mathbf{x} \text{ and } \theta$$

- (HMC Thm 7.3.2) If \mathbf{T} is sufficient for θ and $\hat{\theta}$ is the unique MLE of θ , then $\hat{\theta}$ must be a function of \mathbf{T} .

- Nonuniqueness of sufficient statistics
 - Trivial examples
 - * \mathbf{X} is always sufficient.
 - * $(X_{(1)}, \dots, X_{(n)})$ is always sufficient if X_i 's are iid, with $X_{(1)} \leq \dots \leq X_{(n)}$.
 - \mathbf{T} is sufficient and $g(\cdot)$ is a one-to-one mapping $\Rightarrow g(\mathbf{T})$ is also sufficient.
- Minimal sufficiency: a sufficient statistic that is a function of all the other sufficient statistics.
 - (CB Thm 6.2.13) How to find a minimal sufficient statistic:
 1. Find the sufficient and necessary condition for $L(\theta; \mathbf{x})/L(\theta; \mathbf{y})$ to be free of θ ;
 2. If the above condition is of the form $\mathbf{T}(\mathbf{x}) = \mathbf{T}(\mathbf{y})$, then $\mathbf{T}(\mathbf{X})$ is a minimal sufficient statistic for θ .

Example Lec8.2

- Find the minimal sufficient statistics in the following scenarios.
 - a. iid $X_1, \dots, X_n \sim \text{Unif}\{1, \dots, \theta\}$ with unknown positive integer θ .
 - b. iid $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ with unknown μ and σ^2 .

Rao-Blackwellization (CB Thm 7.3.17)

- Only consider one-dimensional cases
- Improve the variance of statistic W , an estimator of θ : take use of $E(W | T)$ (a function of T only) instead with sufficient T
- T sufficient for $\theta \Rightarrow E(W | T, \theta) = E(W | T) \Rightarrow \text{var}\{E(W | T, \theta) | \theta\} = \text{var}\{E(W | T) | \theta\} \leq \text{var}(W | \theta)$ for all $\theta \in \Theta$
 - No impact on the bias
 - Not working if W is already a function of T

Example Lec9.1

- Improve statistic W in terms of variance.
 - a. $W = X_1$, where iid $X_1, X_2 \sim \mathcal{N}(\theta, 1)$ with unknown θ .
 - b. $W = 2X_1 - X_2$, where iid $X_1, X_2 \sim f(x | \theta) = \theta^{-1} \exp(-x\theta^{-1})\mathbf{1}_{\mathbb{R}^+ \times \mathbb{R}^+}(x, \theta)$ with unknown θ .