STAT 4100 Lecture Note

Week Six (Oct 12 & 14, 2022)

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

2022/Nov/19 02:26:51

Evaluating estimators (con'd)

Completeness (CB Def 6.2.21)

- Only consider one-dimensional cases
- T is a complete statistic if we have the following identity: for any (measurable) function g,

$$\mathrm{E}(g(T) \mid \theta) = 0 \text{ for all } \theta \in \Theta \Rightarrow \Pr(g(T) = 0 \mid \theta) = 1 \text{ for all } \theta \in \Theta.$$

  - Geometrical interpretation: $\mathrm{span}\{f_{T\mid\theta}(t \mid \theta) : \theta \in \Theta\} = \{g(\cdot) : \text{(measurable) } g \text{ is defined on } \mathrm{supp}(T)\}$
- (CB Thm 6.2.28) Minimal sufficient statistics exist $\Rightarrow$ complete sufficient statistics are minimally sufficient
- (HMC Thm 7.5.2) iid $X_1, \ldots, X_n \sim f(x \mid \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left\{\sum_{i=1}^k w_i(\boldsymbol{\theta})t_i(x)\right\}$, i.e., following the exponential family, $\Rightarrow$ $(\sum_{i=1}^n t_1(X_i), \ldots, \sum_{i=1}^n t_k(X_i))$ is complete sufficient

Example Lec9.2

- Find the complete statistic for the following scenarios:
  a. iid $X_1, \ldots, X_n \sim f(x \mid \theta) = (x!)^{-1}\theta^x e^{-\theta}\mathbf{1}_{\mathbb{R}^+ \times \{0,1,\ldots\}}(\theta, x)$;
  b. iid $X_1, \ldots, X_n \sim \mathrm{Unif}\{1, \ldots, \theta\}$, integer $\theta \geq 2$.

Lehmann-Scheffe (CB Thm 7.3.23 & 7.5.1; HMC Thm 7.4.1)

- The unbiased estimator only depending on complete sufficient statistics is the UMVUE for its expectation.
- Application to the construction of UMVUE
  1. Find the minimal sufficient T.
  2. Check the completeness of T.
  3. Find unbiased $g(T)$, e.g.,
     - $\mathrm{E}(W \mid T)$ with certain unbiased W
     - debiased MLE (if it is a function of T).

Example Lec9.3

- Suppose that iid $X_1, \ldots, X_n$ are following $\mathrm{Unif}\{1, \ldots, \theta\}$, integer $\theta \geq 2$. Prove that $[X_{(n)}^{n+1} - (X_{(n)} - 1)^{n+1}]/[X_{(n)}^n - (X_{(n)} - 1)^n]$ is the UMVUE for θ.

Verifying the independence

Ancillary Statistics

- Statistics whose distribution does not depend on unknown θ.

Example Lec10.1

- Verify the following statistics are ancillary for θ.
  a. Range $X_{(n)} - X_{(1)}$ with $X_1, \ldots, X_n \sim \text{Unif}(\theta, \theta+1)$.
  b. X_1/X_2 with $X_1, X_2 \sim \mathcal{N}(0, \theta^2)$.

Basu's theorem (CB Thm 6.2.4)

- T is complete and sufficient, while S is ancillary. Then T and S are independent of each other.
  - The completeness of T can be relaxed to be bounded completeness.

Example Lec10.2

- Let iid $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$. Deduce the independence of $\bar{X}$ and S^2 by applying Basu's theorem for the case with unknown μ and known σ^2.

How to verify the independence of X and Y

- Joint cdf: $F_{X,Y}(x,y) = F_X(x)F_Y(y)$
- Joint pdf or pmf: $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
- conditional pdf or pmf: $f_{X|Y}(x \mid y) = f_X(x)$
- mgf: $\mathrm{E}(e^{t_1X+t_2Y}) = \mathrm{E}(e^{t_1X})\mathrm{E}(e^{t_2Y})$
- cf: $\mathrm{E}(e^{it_1X+it_2Y}) = \mathrm{E}(e^{it_1X})\mathrm{E}(e^{it_2Y})$
- Basu's theorem
  - Sometimes it is even more complex to find complete statisitcs than to obtain the joint pdf
- Zero covariance matrix for normal cases

Review for midterm

Find the distribution of $\mathbf{Y} = \boldsymbol{g}(\mathbf{X})$ given the distribution of $\mathbf{X}$

- First figure out support($\mathbf{Y}$)
- Univariate transformation
  - For discrete Y: find the pmf of Y by definition
  - For continuous Y: find the cdf by definition OR by CB Ex. 2.7(b),

$$f_Y(y) = \sum_{k=1}^{K} f_X\{g_k^{-1}(y)\} \left| J_{g_k^{-1}}(y) \right| \mathbf{1}_{B_k}(y)$$

    * Partition supp(X) into K intervals $A_1, \ldots, A_K$ such that
      · $\bigcup_{k=1}^{K} A_k = \text{supp}(X)$ and $A_k \cap A_{k'} = \emptyset$ if $k \neq k'$
      · g is strictly monotonic and continuously differentiable on A_k
    * $g_k = g_k(x) = g(x)$, $x \in A_k$

* Jacobian of transformation g_k^{-1}

$$J_{g_k^{-1}} = \frac{\mathrm{d}}{\mathrm{d}y} g_k^{-1}(y)$$

* $B_k = \{g(x) : x \in A_k\}$

- Bivariate transformation
  - By definition, e.g., find the cdf of $Y = \min\{X_1, X_2\}$
  - Polar coordinate system, e.g., find the pdf of $Y = X_1^2 + X_2^2$
  - For one-to-one correspondence $\boldsymbol{g}$
    * $\boldsymbol{g}(\cdot) = (g_1(\cdot), g_2(\cdot)) : \mathrm{supp}(\mathbf{X}) \to \mathrm{supp}(\mathbf{Y})$, i.e.,
      · $\boldsymbol{y} = (y_1, y_2) = (g_1(x_1, x_2), g_2(x_1, x_2)) = \boldsymbol{g}(x_1, x_2)$
      · $\boldsymbol{x} = (x_1, x_2) = \boldsymbol{g}^{-1}(y_1, y_2) = (h_1(y_1, y_2), h_2(y_1, y_2))$
    * If $\boldsymbol{g}^{-1}$ is continuously differentiable,

$$f_{\mathbf{Y}}(y_1, y_2) = f_{\mathbf{X}}\{\boldsymbol{g}^{-1}(y_1, y_2)\} |\det\{\mathbf{J}_{\boldsymbol{g}^{-1}}(y_1, y_2)\}| \mathbf{1}_{\mathrm{supp}(\mathbf{Y})}(y_1, y_2)$$

      · $\det\{\mathbf{J}_{\boldsymbol{g}^{-1}}(y_1, y_2)\} = 1/\det[\mathbf{J}_{\boldsymbol{g}}\{\boldsymbol{g}^{-1}(y_1, y_2)\}]$, because

$$\mathbf{J}_{\boldsymbol{g}^{-1}}(y_1, y_2) = \left[\frac{\partial h_i(y_1, y_2)}{\partial y_j}\right]_{2\times 2} = \begin{bmatrix} \frac{\partial h_1(y_1, y_2)}{\partial y_1} & \frac{\partial h_1(y_1, y_2)}{\partial y_2} \\ \frac{\partial h_2(y_1, y_2)}{\partial y_1} & \frac{\partial h_2(y_1, y_2)}{\partial y_2} \end{bmatrix} = \mathbf{J}_{\boldsymbol{g}}^{-1}\{\boldsymbol{g}^{-1}(y_1, y_2)\}$$

Bivariate normal (BVN) distribution

- Random 2-vector $\mathbf{X} = [X_1, X_2]^\top \sim \mathrm{BVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Leftrightarrow \mathbf{X} = \boldsymbol{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu}$ with $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}) \sim \mathrm{BVN}(0, \mathbf{I}_2) \Rightarrow$

$$\mathrm{E}(\mathbf{X}) = [\mathrm{E}(X_1), \mathrm{E}(X_2)]^\top = \boldsymbol{\mu} \quad \text{and} \quad \mathrm{cov}(\mathbf{X}) = [\mathrm{cov}(X_i, X_j)]_{2\times 2} = \boldsymbol{\Sigma}$$

- Random 2-vector $\mathbf{X} \sim \mathrm{BVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Rightarrow \mathbf{B}\mathbf{X} + \boldsymbol{b} \sim \mathrm{BVN}(\mathbf{B}\boldsymbol{\mu} + \boldsymbol{b}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top)$
- If $[X_1, X_2]^\top$ is of BVN, then the marginal distributions of X_1 and X_2 are both normal. The inverse proposition does NOT hold.

Normal sampling theory

- $\sum_{i=1}^n X_i^2 \sim \chi^2(n)$ if iid $X_1, \ldots, X_n \sim \mathcal{N}(0, 1)$
- $X/\sqrt{Y/n} \sim t(n)$ if $X \sim \mathcal{N}(0, 1)$ and $Y \sim \chi^2(n)$ are independent
- $(X/m)/(Y/n) \sim F(m, n)$ if $X \sim \chi^2(m)$ and $Y \sim \chi^2(n)$ are independent
- $n^{1/2}(\bar{X} - \mu)/\sigma \sim \mathcal{N}(0, 1)$ if iid $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$
- $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$ if iid $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$
- $\bar{X}$ and S^2 are independent of each other if iid $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$
- $n^{1/2}(\bar{X} - \mu)/S \sim t(n-1)$ if iid $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$

Generating functions

- Univariate mgf $M_X(t) = \mathrm{E}\{\exp(tX)\}$ if $\mathrm{E}\{\exp(tX)\} < \infty$ for all t inside a neighbourhood of 0
  - Characterizing distributions: identical mgfs implying identical distributions
  - $M_Y(t) = \exp(bt)\prod_i M_{X_i}(a_i t)$ if $Y = b + \sum_i a_i X_i$, where b and a_i are constants, $X_1, \ldots, X_p$ are independent, and each $M_{X_i}(\cdot)$ exists

Parametric model

- iid $X_1, \ldots, X_n \sim f(x \mid \boldsymbol{\theta}_0) \in \{f(x \mid \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$
  - fixed unknown $\boldsymbol{\theta}_0$ to be estimated
- Exponential family
  - If the pdf or pmf of X is of the following form

$$f(x \mid \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left\{\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right\}$$

  - (CB Example 3.4.4) $\mathcal{N}(\mu, \sigma^2)$ with μ and σ^2 both unknown
    * $h(x) = \mathbf{1}_{\mathbb{R}}(x)$
    * $c(\mu, \sigma) = (2\pi\sigma^2)^{-1/2} \exp\{-\mu^2/(2\sigma^2)\}\mathbf{1}_{\mathbb{R}\times\mathbb{R}^+}(\mu, \sigma)$
    * $w_1(\mu, \sigma) = \sigma^{-2}\mathbf{1}_{\mathbb{R}^+}(\sigma)$
    * $w_2(\mu, \sigma) = \mu\sigma^{-2}\mathbf{1}_{\mathbb{R}^+}(\sigma)$
    * $t_1(x) = -x^2/2$
    * $t_2(x) = x$
  - (CB Example 3.4.1) Binom(n, p) with known n and unknown p
    * $h(x) = \binom{n}{x}\mathbf{1}_{\{0,\ldots,n\}}(x)$ (What happens if n is also an unknown parameter?)
    * $c(p) = (1-p)^n\mathbf{1}_{(0,1)}(p)$
    * $w_1(p) = \ln\{p/(1-p)\}\mathbf{1}_{(0,1)}(p)$
    * $t_1(x) = x$
  - Other special cases of exponential family: gamma, beta, Poisson, negative binomial
  - $(\sum_{i=1}^{n} t_1(X_i), \ldots, \sum_{i=1}^{n} t_k(X_i))$ is sufficient complete

Point estimation

- Method of moments (MOM)
  - Equate raw moments to their empirical counterparts (Why is it reasonable?)
  - Pros and cons
- Maximum likelihood (ML)
  - $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ is a statistic such that

$$\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L(\boldsymbol{\theta}; \boldsymbol{x}) = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ell(\boldsymbol{\theta}; \boldsymbol{x})$$

  - Maximizing $L(\boldsymbol{\theta}; \boldsymbol{x})$ or $\ell(\boldsymbol{\theta}; \boldsymbol{x})$ with respect to $\boldsymbol{\theta} \in \boldsymbol{\Theta}$
    * For discrete $\boldsymbol{\Theta}$: compare $L(\boldsymbol{\theta}; \boldsymbol{x})$ or $\ell(\boldsymbol{\theta}; \boldsymbol{x})$ over all the possible values of $\boldsymbol{\theta}$
    * For continuous $\boldsymbol{\Theta}$:
      · If $\mathbf{S}(\boldsymbol{\theta})$ has no zero point: utilize the monotonicity of $L(\boldsymbol{\theta}; \boldsymbol{x})$ or $\ell(\boldsymbol{\theta}; \boldsymbol{x})$
      · If $\mathbf{S}(\boldsymbol{\theta})$ has zero point: solve $\mathbf{S}(\boldsymbol{\theta}) = \mathbf{0}$ for $\boldsymbol{\theta}$ (to obtain stationary points) and then compare $L(\boldsymbol{\theta}; \boldsymbol{x})$ or $\ell(\boldsymbol{\theta}; \boldsymbol{x})$ over all the stationary points and boundary points
  - Invariance property: $\widehat{g(\boldsymbol{\theta})}_{\mathrm{ML}} = g(\hat{\boldsymbol{\theta}}_{\mathrm{ML}})$

Evaluating estimators

- Mean squared error (MSE): $\mathrm{E}(\hat{\theta} - \theta)^2 = \{\mathrm{E}(\hat{\theta}) - \theta\}^2 + \mathrm{var}(\hat{\theta})$
  - UMVUE/MVUE/Best unbiased estimator: minimize MSE subject to $\mathrm{E}(\hat{\theta}) = \theta$
- Cramer-Rao lower bound (one-dimesional case): $\mathrm{var}(\hat{\theta}) \geq \{(\mathrm{d}/\mathrm{d}\theta)\mathrm{E}(\hat{\theta})\}^2/I(\theta)$
  - Fisher information: $I(\theta) = \mathrm{var}\{S(\theta; \mathbf{X})\} = \mathrm{E}[\{S(\theta; \mathbf{X})\}^2] = -\mathrm{E}\{H(\theta; \mathbf{X})\}$
    * $I(\theta) = n\mathrm{var}\{S(\theta; X_1)\} = n\mathrm{E}[\{S(\theta; X_1\}^2] = -n\mathrm{E}\{H(\theta; X_1)\}$ for iid sample $\mathbf{X} = [X_1, \ldots, X_n]$
  - For unbiased estimators
    * $\mathrm{var}(\hat{\theta}) \geq 1/I(\theta)$
    * The unbiased estimator attaining the lower bound is UMVUE
- Alternative ways to find UMVUE

  - Rao-Blackwellization with sufficient complete statistics
    * Minimal sufficiency: find the sufficient and necessary condition for the likelihood ratio to be free of unknown parameters
    * Completeness: find sufficient complete statistics for exponential family
  - Debiasing MLE if the MLE is a function only based on sufficient complete statistics

Checking independence

- Joint cdf: $F_{X,Y}(x, y) = F_X(x)F_Y(y)$
- Joint pdf or pmf: $f_{X,Y}(x, y) = f_X(x)f_Y(y)$
- conditional pdf or pmf: $f_{X|Y}(x \mid y) = f_X(x)$
- mgf: $\mathrm{E}(e^{t_1X+t_2Y}) = \mathrm{E}(e^{t_1X})\mathrm{E}(e^{t_2Y})$
- cf: $\mathrm{E}(e^{it_1X+it_2Y}) = \mathrm{E}(e^{it_1X})\mathrm{E}(e^{it_2Y})$
- Basu's theorem
  - Sometimes it is even more complex to find complete statisitcs than to obtain the joint pdf
- Zero $\mathrm{cov}(X, Y)$ for joint normal (X, Y)

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

- CB Ex. 7.46, 7.48, 7.58, 7.66
- HMC Ex. 7.9.4, 7.9.13 (a-d)