Statistics 581, Problem Set 7

Wellner; 11/7/2018

Reading: Chapter 3, Sections 2-4;
Ferguson, ACILST, Chapters 19-20, pages 126-139;
vdV, Asymp. Statist., pages 108 - 119; Sections 8.1 - 8.7.

Due: Wednesday, November 14, 2018.

1. Suppose that $X \sim \text{Beta}(\alpha, \beta)$; i.e. X has density p_θ given by

$$p_\theta(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} 1_{(0,1)}(x), \quad \theta = (\alpha, \beta) \in (0,\infty) \times (0,\infty) \equiv \Theta .$$

Consider estimation of:
A. $q_A(\theta) \equiv E_\theta X$. B. $q_B(\theta) \equiv F_\theta(x_0)$ for a fixed x_0; here $F_\theta(x) \equiv P_\theta(X \le x)$.
(i) Compute $I(\theta) = I(\alpha, \beta)$; compare Lehmann & Casella page 127, Table 6.1
(ii) Compute $q_A(\theta)$, $q_B(\theta)$, $\dot{q}_A(\theta)$, and $\dot{q}_B(\theta)$.
(iii) Find the efficient influence functions for estimation of q_A and q_B.
(iv) Compare the efficient influence functions you find in (iii) with the influence functions ψ_A and ψ_B of the natural nonparametric estimators $\overline{X}_n$ and $\mathbb{F}_n(x_0)$ respectively. Does $\psi_A \in \dot{\mathcal{P}}$? Does $\psi_B \in \dot{\mathcal{P}}$ hold?

2. Suppose that $X \sim F_\theta = \text{exponential}(\theta)$ with density $f_\theta(x) = \theta e^{-\theta x} 1_{(0,\infty)}(x)$ and $Y \sim G_\eta$ independent of X with densities $\{g_\eta : \eta \in R^+\}$, a regular parametric model on $(0,\infty)$. Consider the following three scenarios for observation of X or functions of X:
(a) Uncensored: we observe X and Y.
(b) Right-censored: we observe

$$T(X,Y) = (X \wedge Y, 1\{X \le Y\} \equiv (\min\{X,Y\}, 1\{X \le Y\}) \equiv (Z, \Delta).$$

(c) Interval-censored (case 1): we observe $S(X,Y) = (Y, 1\{X \le Y\}) \equiv (Y, \Delta)$.
(i) Find the joint density of (X,Y) and joint distributions of $T(X,Y)$ and $S(X,Y)$.
(ii) Find the scores for θ and η in each of the three scenarios (a), (b), and (c). (Let $(\partial/\partial\eta)\log g_\eta(y) \equiv a(y)$ with $a \in L_2^0(G_\eta)$.)
(iii) Compute and compare $I_{X,Y}(\theta)$, $I_{T(X,Y)}(\theta)$, and $I_{S(X,Y)}(\theta)$. Make the comparisons in general and then explicitly by making one or more choices of the family $\{g_\eta\}$.

3. Suppose that we want to model the survival of twins with a common genetic defect, but with one of the two twins receiving some treatment. Let X represent the survival time of the untreated twin and let Y represent the survival time of the treated twin. One (overly simple) preliminary model might be to assume that X and Y are independent with Exponential(η) and Exponential($\theta\eta$) distributions, respectively:

$$f_{\theta,\eta}(x,y) = \eta e^{-\eta x} \eta\theta e^{-\eta\theta y} 1_{(0,\infty)}(x) 1_{(0,\infty)}(y)$$

Compute the Cramér-Rao lower bound for unbiased estimates of θ based on $Z = X/Y$, the maximal invariant for the group of scale changes $g(x,y) = (cx, cy)$ with $c > 0$. Compared this bound to the information bounds for estimation of θ based on observation of (X,Y) when η is known and unknown.

4. Suppose that $\theta = (\theta_1, \theta_2) \in \Theta \subset R^k$ where $\theta_1 \in R$ and $\theta_2 \in R^{k-1}$. Show that:
A. $\mathbf{l}_1^* = \dot{\mathbf{l}}_1 - I_{12} I_{22}^{-1} \dot{\mathbf{l}}_2$ is orthogonal to $[\dot{\mathbf{l}}_2] \equiv \{a' \dot{\mathbf{l}}_2 : a \in R^{k-1}\}$ in $L_2(P_\theta)$.
B. $I_{11\cdot 2} = \inf_{c \in R^{k-1}} E_\theta (\dot{\mathbf{l}}_1 - c' \dot{\mathbf{l}}_2)^2$ and that the infimum is achieved when $c' = I_{12} I_{22}^{-1}$.
Thus

$$I_{11\cdot 2} = E_\theta (\dot{\mathbf{l}}_1 - I_{12} I_{22}^{-1} \dot{\mathbf{l}}_2)^2 = E_\theta [(\mathbf{l}_\theta^*)^2] \,.$$

C. Prove the formulas (15) and (16) on page 21 of the Chapter 3 notes and interpret these formulas geometrically.

5. **Optional bonus problem 1:** Suppose that $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, $\Theta \subset R^k$ is a parametric model satisfying the hypotheses of the multiparameter Cramér - Rao inequality. Partition θ as $\theta = (\nu, \eta)$ where $\nu \in R^m$ and $\eta \in R^{k-m}$ and $1 \le m < k$. Let $\dot{l} = \dot{l}_\theta = (\dot{l}_1, \dot{l}_2)$ be the corresponding partition of the (vector of) scores $\dot{l}$, and, with $\widetilde{l} \equiv I^{-1}(\theta)\dot{l}$, the *efficient influence function* for θ, let $\widetilde{l} = (\widetilde{l}_1, \widetilde{l}_2)$ be the corresponding partition of $\widetilde{l}$. In both cases, $\dot{l}_1$, $\widetilde{l}_1$ are $m-$vectors of functions, and $\dot{l}_2$, $\widetilde{l}_2$ are $k-m$ vectors. Partition $I(\theta)$ and $I^{-1}(\theta)$ correspondingly as

$$I(\theta) = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}$$

where I_{11} is $m \times m$, I_{12} is $m \times (k-m)$, I_{21} is $(k-m) \times m$, I_{22} is $(k-m) \times (k-m)$. Also write

$$I^{-1}(\theta) = [I^{ij}]_{i,j=1,2}.$$

Verify that:
A. $I^{11} = I_{11\cdot 2}^{-1}$ where $I_{11\cdot 2} \equiv I_{11} - I_{12} I_{22}^{-1} I_{21}$,
$I^{22} = I_{22\cdot 1}^{-1}$ where $I_{22\cdot 1} \equiv I_{22} - I_{21} I_{11}^{-1} I_{12}$,
$I^{12} = -I_{11\cdot 2}^{-1} I_{12} I_{22}^{-1}$,
$I^{21} = -I_{22\cdot 1}^{-1} I_{21} I_{11}^{-1}$.
This amounts to formulas (4) and (5) of section 3.2, page 14.
B. Verify that
$\widetilde{l}_1 = I^{11} \dot{l}_1 + I^{12} \dot{l}_2 = I_{11\cdot 2}^{-1} (\dot{l}_1 - I_{12} I_{22}^{-1} \dot{l}_2)$, and
$\widetilde{l}_2 = I^{21} \dot{l}_1 + I^{22} \dot{l}_2 = I_{22\cdot 1}^{-1} (\dot{l}_2 - I_{21} I_{11}^{-1} \dot{l}_1)$. Compare these formulas with those in part C of problem #4 above.

6. **Optional bonus problem 2:** Consider the two parameter location-scale model

$$\mathcal{P} = \left\{ P_\theta : \; \frac{dP_\theta}{d\lambda} = p_\theta : \theta \in \Theta \right\}$$

where $\Theta = \mathbb{R} \times \mathbb{R}^+$,

$$p_\theta(x) = \frac{1}{\theta_2} f\left(\frac{x - \theta_1}{\theta_2} \right),$$

and the (known) density f has a derivative f' almost everywhere with respect to Lebesgue measure λ.
(a) Calculate the information matrix $I(\theta)$ for θ.
(b) For which of the densities in (a)-(e) of problem set #6, problem 1, is $I_{12}(\theta)$ not zero?