

Statistics 581, Problem Set 7

Wellner; 11/13/2009

Reading: Chapter 3, Section 2;

Ferguson, ACILST, Chapter 19, pages 126-132, Chapter 20, pages 133-134;

Lehmann and Casella, pages 113-129, and 439- 443.

Due: Wednesday, November 18, 2009.

1. Compute and plot the *score for location*, $-(f'/f)(x)$ when:
 - A. $f(x) = \phi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$, (normal or Gaussian);
 - B. $f(x) = \exp(-x)/(1 + \exp(-x))^2$, (logistic);
 - C. $f(x) = \frac{1}{2} \exp(-|x|)$, (double exponential);
 - D. $f = t_k$, the t -distribution with k degrees of freedom;
 - E. $f(x) = \exp(-x) \exp(-\exp(-x))$, Gumbel or extreme value.
2. Compute $I_f = \int (f'(x)/f(x))^2 f(x) dx$, the information for location, for each of the densities in problem 1.
3. Consider the two parameter location-scale model

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

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 1 is $I_{12}(\theta)$ not zero?

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

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

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 $F_\theta(x) \equiv P_\theta(X \leq x)$.

(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 \bar{X}_n and $\mathbb{F}_n(x_0)$ respectively; in particular, show that $\psi_A \in \dot{\mathcal{P}}$, while $\psi_B \notin \dot{\mathcal{P}}$.

5. Lehmann and Casella, TPE, Problem 6.6, page 142.

6. 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 $1 \leq m < k$. 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 $\tilde{l} \equiv I^{-1}(\theta)\dot{l}$, the *efficient influence function* for θ , let $\tilde{l} = (\tilde{l}_1, \tilde{l}_2)$ be the corresponding partition of \tilde{l} . In both cases, \dot{l}_1, \tilde{l}_1 are m -vectors of functions, and \dot{l}_2, \tilde{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. $I^{11} = I_{11.2}^{-1}$ where $I_{11.2} \equiv I_{11} - I_{12}I_{22}^{-1}I_{21}$,
 $I^{22} = I_{22.1}^{-1}$ where $I_{22.1} \equiv I_{22} - I_{21}I_{11}^{-1}I_{12}$,
 $I^{12} = -I_{11.2}^{-1}I_{12}I_{22}^{-1}$,
 $I^{21} = -I_{22.1}^{-1}I_{21}I_{11}^{-1}$.

This amounts to formulas (3) and (4) of section 3.2, page 14.

B. Verify that

$$\tilde{l}_1 = I^{11}\dot{l}_1 + I^{12}\dot{l}_2 = I_{11.2}^{-1}(\dot{l}_1 - I_{12}I_{22}^{-1}\dot{l}_2), \text{ and}$$

$$\tilde{l}_2 = I^{21}\dot{l}_1 + I^{22}\dot{l}_2 = I_{22.1}^{-1}(\dot{l}_2 - I_{21}I_{11}^{-1}\dot{l}_1).$$

7. **Optional bonus problem 1:** 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 $\text{Exponential}(\eta)$ and $\text{Exponential}(\theta\eta)$ 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.