Statistics 581, Problem Set 7

Wellner; 11/13/2002

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 20, 2001.

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. 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. $I^{11} = I^{-1}_{11\cdot 2}$ where $I_{11\cdot 2} \equiv I_{11} - I_{12}I^{-1}_{22}I_{21}$,
   $I^{22} = I^{-1}_{22\cdot 1}$ where $I_{22\cdot 1} \equiv I_{22} - I_{21}I^{-1}_{11}I_{12}$,
   $I^{12} = -I^{-1}_{11\cdot 2}I_{12}I^{-1}_{22}$,
   $I^{21} = -I^{-1}_{22\cdot 1}I_{21}I^{-1}_{11}$.
   This amounts to formulas (5) and (6) of section 3.2, page 15.
   B. Verify that
   $\widetilde{l}_1 = I^{11}\dot{l}_1 + I^{12}\dot{l}_2 = I^{-1}_{11\cdot 2}(\dot{l}_1 - I_{12}I^{-1}_{22}\dot{l}_2)$, and
   $\widetilde{l}_2 = I^{21}\dot{l}_1 + I^{22}\dot{l}_2 = I^{-1}_{22\cdot 1}(\dot{l}_2 - I_{21}I^{-1}_{11}\dot{l}_1)$.
   The first of these is (7) on page 15, section 3.2.

4. 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:

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

A. One crude approach to estimation in this problem is to reduce the data to $W = X/Y$, the maximal invariant for the group of scale changes $g(x,y) = (cx, cy)$ with $c > 0$. Find the distribution of W, and compute the Cramér-Rao lower bound for unbiased estimates of θ based on W.
B. Find the information bound for estimation of θ based on observation of (X,Y) pairs when η is known and unknown.
C. Compare the bounds you computed in A and B and discuss the pros and cons of reducing to estimation based on the W.

5. This is a continuation of the preceding problem. A more realistic model involves assuming that the common parameter η for the two twins varies across sets of twins. There are several different ways of modeling this: one approach involves supposing that each pair of twins observed (X_i, Y_i) has its own fixed parameter η_i, $i = 1,\ldots,n$. In this model we observe (X_i, Y_i) with density f_{θ,η_i} for $i = 1,\ldots,n$; i.e.

$$f_{\theta,\eta_i}(x_i,y_i) = \eta_i e^{-\eta_i x_i}\eta_i\theta e^{-\eta_i\theta y_i}1_{(0,\infty)}(x_i)1_{(0,\infty)}(y_i)\,. \tag{0.1}$$

This is sometimes called a "functional model" (or model with incidental nuisance parameters).

Another approach is to assume that $\eta \equiv Z$ has a distribution, and that our observations are from the mixture distribution. Assuming (for simplicity) that $Z = \eta \sim$ Gamma(a,b) with density $g_{a,b}(\eta)$, it follows that the (marginal) distribution of (X,Y) is

$$\begin{aligned} p_{\theta,a,b}(x,y) &= \int_0^\infty f_{\theta,z}(x,y)g_{a,b}(z)dz \\ &= \frac{\theta}{b^2}\left(\frac{b}{b+x+\theta y}\right)^{a+2}\frac{\Gamma(a+2)}{\Gamma(a)}\,. \end{aligned} \tag{0.2}$$

This is sometimes called a "structural model" (or mixture model).
(a) Find the information for θ in the functional model.
(b) Find the information for θ in the structural model.
(c) Compare the information bounds you computed in (a) and (b). When is the information for θ in the functional model larger than the information for θ in the structural model?

6. **Optional bonus problem:** [This is example 7.2.5 and 7.2.7 in Lehmann and Casella, TPE, section 6.2; also see problems 6.2.12 - 6.2.14, Lehmann and Casella, TPE, page 501.] Suppose that $X_1,\ldots,X_n$ are i.i.d. $N(\theta,1)$ so $I(\theta) = 1$. Let $0 < a < 1$ and define $T_n \equiv \overline{X}_n 1_{[|\overline{X}_n| \ge n^{-1/4}]} + a\overline{X}_n 1_{[|\overline{X}_n| < n^{-1/4}]}$. This is Hodges superefficient estimator of θ.
(a) Show that $\sqrt{n}(T_n - \theta) \to_d N(0, V(\theta))$ where

$$V(\theta) = \begin{cases} 1, & \theta \ne 0 \\ a^2, & \theta = 0 \end{cases}$$

(b) Show that T_n is *not a regular estimator* of θ at $\theta = 0$, but that it is regular at every $\theta \neq 0$. If $\theta_n = t/\sqrt{n}$, find the limiting distribution of $\sqrt{n}(T_n - \theta_n)$ under P_{θ_n}.
C. For $\theta_n = t/\sqrt{n}$ show that

$$R_n(\theta_n) = nE_{\theta_n}(T_n - \theta_n)^2 \to E(aZ + t(a-1))^2 = a^2 + t^2(1-a)^2$$

where $Z \sim N(0,1)$. This is *larger* than 1 if $t^2 > (1+a)/(1-a)$, and hence superefficiency also entails *worse* risks in a local neighborhood of the point(s) where the asymptotic variance is smaller.

7. **Optional bonus problem:**
Lehmann and Casella, TPE, Problem 6.6, page 142.