Statistics 581, Problem Set 9

Wellner; 11/19/2008

Reading: Chapter 4, Sections 1-2;
Ferguson, ACLST, Chapters 18- 20, pages 119-125, 133-139; Chapter 22, pages 144-150;
Lehmann and Casella, Chapter 6, especially section 6.5, pages 461-468.
Due: Wednesday, November 26, 2008.

1. (a) Exercise 2.1.6, page 10, chapter 2 notes.
   (b) Exercise 2.1.7, page 10, chapter 2 notes.

2. (a) Lehmann and Casella, problem 6.3.1, page 501.
   (b) Lehmann and Casella, problem 6.3.2, page 501.
   (c) Lehmann and Casella, problem 6.3.4, page 501.

3. Lehmann and Casella, problem 6.3.18, page 502. [**Note:** It seems to me that 3.15(b) should be 3.15(c) since $C(0,a)$ is a *scale family*).]

4. Suppose that $(Y|Z) \sim \text{Poisson}(\lambda e^{\gamma Z})$, and $Z \sim Bernoulli(\eta)$, and $\theta = (\lambda, \gamma, \eta)$. Let $X = (Y, Z)$, and suppose that we observe $X_1, \ldots, X_n$ i.i.d. as X.
   (a) Find the score equations for estimation of θ.
   (b) Give conditions on the data $X_1, \ldots, X_n = (Y_1, Z_1), \ldots, (Y_n, Z_n)$ guaranteeing that the score equations have a unique solution which maximizes the likelihood. Call the resulting estimators $\widehat{\theta}_n = (\widehat{\lambda}_n, \widehat{\gamma}_n, \widehat{\eta}_n)$.
   (c) What does theorem 4.1.2 (Chapter 4, page 5), say about the asymptotic distribution of $\sqrt{n}(\widehat{\theta} - \theta_0)$ when the distribution of the data is given by P_{θ_0}?
   (d) Suppose that $\theta_1 \neq \theta_0$ is the "true" value of the parameter θ, and we consider the likelihood ratio $L_n(\theta_1)/L_n(\theta_0)$ where $L_n(\theta) \equiv \prod_{i=1}^n p_\theta(X_i)$. Show that $n^{-1}\log(L_n(\theta_1)/L_n(\theta_0) \to_p$ some constant, and identify the constant explicitly in terms of θ_1, θ_0.

5. For the same set-up as in problem 4, consider taking a "profile likelihood" approach to the estimation of γ as follows:
   (a) Let $l_n(\theta) = l_n(\gamma, \lambda, \eta)$: consider first maximizing this as a function of λ and η for each fixed value of γ to find
   $$(\widehat{\lambda}(\gamma), \widehat{\eta}(\gamma)) \equiv \text{argmax}_{(\lambda,\eta)} l_n(\lambda, \gamma, \eta)\,.$$
   Compute the maximizer $(\widehat{\lambda}(\gamma), \widehat{\eta}(\gamma))$ as explicitly as possible, and then form the "profile log-likelihood" $l_n^{profile}(\gamma)$ defined by
   $$l_n^{profile}(\gamma) \equiv l_n(\widehat{\lambda}(\gamma), \gamma, \widehat{\eta}(\gamma))\,.$$
   (b) Now maximize $l_n^{profile}(\gamma)$ with respect to γ. Find the resulting "profile likelihood" score equation for γ.
   (c) Does the equation you derived in (b) follow from the original score equations?
   (d) Does the "profile score function" which appears in (b) correspond to or relate to the efficient score for γ in any way?

6. (a) Ferguson, ACILST, problem 17.2, page 117.
   (b) Do our hypotheses A0-A2 hold in this example?
   (c) Compute $K(P_{\theta_0}, P_\theta)$ where P_θ has density as given in this problem.
   (d) Do our hypotheses A3 and A4 hold in this example? Why or why not?
   (e) Does there exist an estimator $\overline{\theta}_n$ of θ which is $n^{1/2}-$consistent?

7. **Optional bonus problem 1:** Ferguson, ACILST, problem 17.3, page 118.

8. **Optional bonus problem 2:** Lehmann and Casella, TPE, problem 6.3.22, page 503, reworded as follows. (In other words, prove (vi) of theorem 1.2, pages 5-6, chapter 4 notes). Suppose that $X_1, \ldots, X_n$ are i.i.d. with density p_θ, $\theta \in \Theta \subset R^k$, satisfying the hypotheses of theorem 4.1, page 463 (the Cramér conditions given in (A) - (D) on pages 462-463). Show that the following Local Asymptotic Normality (LAN) result holds for the (local) log- likelihood ratios: with

$$L_n(\theta) \equiv \log(\prod_{i=1}^n p_\theta(X_i)) = \sum_{i=1}^n \log p_\theta(X_i),$$

for a fixed $\theta_0 \in \Theta$,

$$\begin{aligned} L_n(\theta_0 + n^{-1/2}\underline{t}) - L_n(\theta_0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \underline{t}^T \dot{\underline{l}}_\theta(X_i) - \frac{1}{2}\underline{t}^T I(\theta_0)\underline{t} + o_p(1) \\ &\rightarrow_d N(0, t^T I(\theta_0)t) - \frac{1}{2}t^T I(\theta_0)t \\ &\stackrel{d}{=} N(-\sigma^2/2, \sigma^2) \end{aligned}$$

under P_{θ_0} where $\sigma^2 \equiv \underline{t}^T I(\theta_0)\underline{t}$. (The convergence in the last display actually holds under the considerably weaker hypothesis of Hellinger differentiability of p_θ at θ_0, as stated in Corollary 3 of section 3.3, page 28, of the Chapter 3 notes.)