

Statistics 581, Problem Set 9

Wellner; 11/27/2002

Reading: Chapter 4, Sections 1-2;

Ferguson, ACLST, Chapter 20, pages 133-139; Chapter 22, pages 144-150;

Lehmann and Casella, Chapter 6, especially section 6.5, pages 461-468.

Due: Wednesday, December 4, 2002.

- Lehmann and Casella, problem 6.3.1, page 501.
 - Lehmann and Casella, problem 6.3.2, page 501.
 - Lehmann and Casella, problem 6.3.4, page 501.
 - 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*.]
- Suppose that $(Y|Z) \sim \text{Weibull}(\lambda^{-1}e^{-\gamma Z}, \beta)$, as in problem 2 of problem set # 8, where $Z \sim \text{Bernoulli}(\eta)$. Reparametrize as in problem set # 8 so that the joint density is $p_\theta(y, z)$ with $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$ where $\theta_4 = \eta$. Thus the conditional hazard function is

$$\lambda_\theta(t|z) = \theta_3\theta_2 \exp(\theta_1 z) t^{\theta_3-1},$$

$Z \sim \text{Bernoulli}(\eta)$, and $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$ with $\theta_4 \equiv \eta$. Let $X = (Y, Z)$, and suppose that we observe X_1, \dots, X_n i.i.d. as X .

- Find the score equations for estimation of θ .
 - Give conditions on the data $X_1, \dots, X_n = (Y_1, Z_1), \dots, (Y_n, Z_n)$ guaranteeing that the score equations have a unique solution which maximizes the likelihood. Call the resulting estimators $\hat{\theta}_n$.
 - What does theorem 4.1.2 (Chapter 4, page 5), say about the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ when the distribution of the data is given by P_{θ_0} .
 - 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 $n^{-1} \log(L_n(\theta_1)/L_n(\theta_0)) \rightarrow_p$ some constant, and identify the constant as explicitly as possible in terms of θ_1, θ_0 .
- Consider the Weibull family of example 3.2.5: $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with $\Theta \subset R^{+2}$ given by the (Lebesgue) densities

$$p_\theta(x) = \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left(-\left(\frac{x}{\alpha}\right)^\beta\right) 1_{[0, \infty)}(x)$$

where $\theta \equiv (\alpha, \beta) \in (0, \infty) \times (0, \infty) \subset R^2$. Suppose that X, X_1, \dots, X_n are i.i.d. with density function p_θ .

A. If $X \sim P_\theta \in \mathcal{P}$, show that the distributions of $\log X$ form a location and scale family from a Gumbel (extreme value) density on R .

B. Use the result of A to construct method of moments estimators or quantile based estimators $\bar{\theta}_n$ of $\theta = (\alpha, \beta)$.

C. Show that the method of moments or quantile estimators $\bar{\theta}_n$ of θ are asymptotically normal, and find the asymptotic distribution; i.e. show that

$$\sqrt{n}(\bar{\theta}_n - \theta) \rightarrow_d N_2(0, \Sigma) \quad \text{for some} \quad \Sigma.$$

- D. Does a maximum likelihood estimate of $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ exist? Is it unique?
- E. Compute an approximate (one - step) maximum likelihood estimate $\tilde{\theta}$ of θ using the method of moment estimators $\bar{\theta}_n$ as the preliminary estimators based on the following data (with $n = 19$):

0.19, 0.78, 0.96, 1.31, 2.78, 3.16, 4.15, 4.67, 4.85, 6.50,
7.35, 8.01, 8.27, 12.06, 31.75, 32.52, 33.91, 36.71, 72.89 .

[These are failure times in minutes for an insulating fluid between two electrodes subject to a voltage of 34 kV. – from Nelson, *Applied Life Data Analysis*, page 105.]

- F. Compute the maximum likelihood estimator $\hat{\theta}_n$, and compare it with the one step estimator computed in E.

4. **Optional bonus problem:** Lehmann and Casella, TPE, problem 6.3.22, page 503, reworded as follows. (In other words, prove (vi) of theorem 1.5, page 5, chapter 4 notes). Suppose that X_1, \dots, X_n are i.i.d. with density p_θ , $\theta \in \Theta \subset \mathbb{R}^k$, satisfying the hypotheses of theorem 4.1, page 429 (the Cramér conditions given in (A) - (D) on page 429). Show that the following Local Asymptotic Normality (LAN) result holds for the (local) log- likelihood ratios: with

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

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

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

under P_{θ_0} where $\sigma^2 \equiv \underline{t}^T I(\theta_0) \underline{t}$.