Stat 982: Module 3, Homework 1

Due on November 22, 2022

Problem 1

Suppose that $X^{'}$ is exponential with mean λ^{-1} (i.e., it has density $f_\lambda(x) = \lambda \exp(-\lambda x) I[x \geq 0]$ with respect to the Lebesgue measure on $\mathbb{R}^1$), but that one only observes $X = X^{'} I[X^{'} \geq 1]$. (There is interval censoring below $X^{'} = 1$.)

Consider the maximum likelihood estimation of λ based on $X_1, \ldots, X_n$, which are iid with the distribution P_λ^X. Let $\delta_i = I[X_i \neq 0]$, for $i = 1, \ldots, n$. Let $M_n = n - \sum_{i=1}^n \delta_i$; that is, M_n is the number of X_i's equal to 0.

(a) Show that there is no maximum likelihood estimate (MLE) of λ when $M_n = n$, but there is an MLE of λ when $M_n < n$.

(b) Show that for any $\lambda \in (0, \infty)$, with λ-probability (that is, $P_\lambda^{X^n}$-probability) tending to 1, the MLE of λ, $\widehat{\lambda}_n$, exists.

(c) Give a simple estimator of λ based on M_n alone. Prove that this estimator is consistent for λ. Then write down an explicit one-step Newton improvement of your estimator based on the likelihood function from part (a).

(d) Discuss what numerical methods you could use to find the MLE from part (a) when it exists.

(e) Give two forms of large-sample (Wald-type) confidence intervals for λ based on the MLE $\widehat{\lambda}_n$ and two different approximations to $I_1(\lambda)$.

Problem 2

Suppose that $X_1, \ldots, X_n$ are iid with the distribution P_θ for $\theta \in \mathbb{R}^1$, where P_θ has the R-N derivative with respect to the counting measure ν on $\mathcal{X} = \{0, 1, 2\}$ given by

$$f_\theta(x) = \frac{\exp(x\theta)}{1 + \exp(\theta) + \exp(2\theta)}.$$

(a) Find an estimator T_n of θ based on $n_0 = \sum_{i=1}^n I[X_i = 0]$ such that T_n is $\sqrt{n}$-consistent (that is, $\sqrt{n}(T_n - \theta)$ is bounded in probability).

(b) Find an explicit one-step Newton improvement of your estimator from part (a).

(c) (Optional) Prove directly that your estimator from part (b), denoted by $\widetilde{\theta}_n$, is asymptotically normal with variance $1/I_1(\theta)$. (Hint: Note that

$$\widetilde{\theta}_n = T_n - \frac{L_n^{'}(T_n)}{L_n^{''}(T_n)},$$

and write $L_n^{'}(T_n) = L_n^{'}(\theta) + (T_n - \theta)L_n^{''}(\theta) + \frac{1}{2}(T_n - \theta)^2 L_n^{'''}(\theta_n^*)$ for some θ_n^* between T_n and θ.)

(d) (Optional) Let $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Show that, if $\overline{X} \in (0, 2)$, the log-likelihood function has a unique maximizer

$$\widehat{\theta}_n = \log\left(\frac{\overline{X} - 1 + \sqrt{-3\overline{X}^2 + 6\overline{X} + 1}}{2(2 - \overline{X})}\right).$$

(e) Prove that an estimator defined to be $\widehat{\theta}_n$ in part (d) when $\overline{X} \in (0, 2)$ is asymptotically normal with variance $1/I_1(\theta)$.

(f) Show that $-\frac{1}{n}L^{''}(\theta) = I_1(\theta)$, for $n = 1, 2, \ldots$ and $\theta \in \mathbb{R}^1$. (Thus the "observed Fisher information" and "expected Fisher information" approximations lead to the same large-sample confidence intervals for λ.)

Note: A version of nearly everything in this problem works in any one-parameter exponential family.

Relevant notes from Stat 643 at ISU

An honest version of the Stat 543 (MS theory course) "MLE's are asymptotically Normal" is next.

Theorem 173. *Suppose that $k = 1$ and there exists an open neighborhood of θ_0, say $\mathcal{O}$, such that*

1. $f_\theta(x) > 0$ $\forall x$ *and* $\forall \theta \in \mathcal{O}$,
2. $\forall x$, $f_\theta(x)$ *is three times differentiable at every point of* $\mathcal{O}$,
3. *there exist* $M(x) \geq 0$ *with* $E_{\theta_0} M(X) < \infty$ *and*
$$\left| \frac{d^3}{d\theta^3} \ln f_\theta(x) \right| \leq M(x) \ \ \forall x \text{ and } \forall \theta \in \mathcal{O},$$
4. $1 = \int f_\theta(x) \, d\mu(x)$ *can be differentiated twice with respect to θ under the integral at θ_0, and*
5. $I_1(\theta) \in (0, \infty)$ $\forall \theta \in \mathcal{O}$.

If with θ_0 probability approaching 1, $\hat{\theta}_n$ is a root of the likelihood equation and $\hat{\theta}_n \to \theta_0$ in θ_0 probability, then under θ_0

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} N\left(0, \frac{1}{I_1(\theta_0)}\right)$$

Corollary 174. *Under the hypotheses of Theorem 173, if $I_1(\theta)$ is continuous at θ_0, then under θ_0*

$$\sqrt{nI_1\left(\hat{\theta}_n\right)}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} N(0,1)$$

Corollary 175. *Under the hypotheses of Theorem 173, under θ_0*

$$\sqrt{-L_n''\left(\hat{\theta}_n\right)}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} N(0,1)$$

What is often a more practically useful result (parallel to Theorem 173) concerns "one-step Newton improvements" on "$\sqrt{n}$-consistent" estimators. (The following is a special case of Schervish's Theorem 7.75.)

Theorem 176. *Under the hypotheses 1-5 of Theorem 173, suppose that under θ_0 estimators T_n are $\sqrt{n}$-consistent, that is $\sqrt{n}(T_n - \theta_0)$ converges in distribution (or more generally, is $O(1)$ in θ_0 probability). Then with*

$$\tilde{\theta}_n = T_n - \frac{L_n'(T_n)}{L_n''(T_n)}$$

under θ_0

$$\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) \xrightarrow{d} N\left(0, \frac{1}{I_1(\theta_0)}\right)$$

It is then obvious that versions of Corollaries 174 and 175 hold where $\hat{\theta}_n$ is replaced by $\tilde{\theta}_n$.