Statistics 582, Problem Set 4 Solutions

Wellner; 2/6/98

1. (MLE's and consistency via reparametrization/compactification).
Suppose that $X_1, \ldots, X_n$ are i.i.d. P_{θ_0} $\theta_0 \in \Theta = R$ where $\mathcal{P} = \{P_\theta : p_\theta(x) = (dP_\theta/d\mu)(x) = g(x-\theta), \theta \in R\}$ and where $g(x) = \exp(-x)/(1+\exp(-x))^2$ is the logistic density function.
A. Show that the MLE $\widehat{\theta}_n$ of θ exists and is unique.
B. Show that the MLE $\widehat{\theta}_n$ of θ is consistent by compactifying the parameter space using the reparametrization shown in the following figure, and then applying Wald's theorem. [Hint: see theorem 4.2 and examples 4.3 and 4.4 in section 4.4.]

Solution: A. The function

$$\rho(x,\theta) = \log g(x-\theta) = -(x-\theta) - 2\log(1+e^{-(x-\theta)})$$

is strictly concave, as is easily seen by computing derivatives:

$$\begin{aligned} \frac{\partial}{\partial\theta}\rho(x,\theta) &= 1 - 2\frac{e^{-(x-\theta)}}{1+e^{-(x-\theta)}} = 1 - 2(1-G(x-\theta)) \\ &= 2G(x-\theta) - 1 = 2\frac{1}{1+e^{-(x-\theta)}} - 1 \end{aligned}$$

so that

$$\frac{\partial^2}{\partial\theta^2}\rho(x,\theta) = -2\,\frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2} = -2G(x-\theta)(1-G(x-\theta)) < 0$$

for all $x \in R$, $\theta \in R$. Since the sum of (strictly) concave functions is (strictly) concave, it follows that the log-likelihood

$$l_n(\theta) = \sum_{i=1}^n \rho(X_i,\theta) = -(\overline{X}-\theta) - \frac{2}{n}\sum_{i=1}^n \log(1+e^{-(X_i-\theta)})$$

is stricly concave. Hence if $l_n(\theta)$ has a maximum, it is unique. But

$$\dot{l}_n(\theta) = 2\sum_{i=1}^n G(X_i-\theta) - n$$

converges to $-n$ as $\theta \to +\infty$, and converges to $+n$ as $\theta \to -\infty$; and decreases strictly and continuously for $\theta \in R$ (since $\ddot{l}_n(\theta) < 0$ for all θ). Hence there is a

unique $\hat{\theta}_n$ satisfying $\dot{l}_n(\hat{\theta}_n) = 0$, and this $\hat{\theta}_n$ maximizes $l_n(\theta)$.
B. Let Θ' be the unit circle in R^2 with south pole at $(0,0)$ and north pole at $(0,2)$. The map $c : \Theta' \to \Theta = R$ given by the following picture gives a one-to-one continuous map from Θ' to $\overline{R}$ with identification of the points $-\infty$ and $+\infty$, the "one-point compactification" of R.

The natural sub-probability measure corresponding to $\theta = \pm\infty$ (or $\theta' = (0,2)$) is the measure with density 0 for all x. Thus we can have identified Θ' with $\overline{R}$, and hence $\overline{R}$ is compact. We will continue to denote the parameter by θ.

Note that $\rho(x,\theta) = \log g(x-\theta)$ is maximimized over $\theta \in \overline{R}$ by $\theta_m = x$, and $\rho(x,x) = -2\log 2$. Thus we have

$$\begin{aligned} l(\Theta \big| X) - l(\theta_0 \big| X) &= (X - \theta_0) + 2\log(1 + e^{-(X-\theta_0)}) - 2\log 2 \\ &\le |X - \theta_0| \end{aligned}$$

by considering separately the sets $X \ge \theta_0$ and $X \le \theta_0$. Thus an envelope function F for Wald's theorem is just $F(X) = |X - \theta_0|$, which is easily integrable under $P_0 = P_{\theta_0}$:

$$E_0|X - \theta_0| = \int_{-\infty}^{\infty} |x - \theta_0| g(x - \theta_0)dx < \infty\,.$$

Furthermore $\overline{\Theta} = \overline{R}$ is compact via the identification with the compact set Θ', and the map $p(x,\theta) = g(x-\theta)$ from Θ' (or $\overline{R}$) to R is upper semicontinuous in θ for all x (even continuous since the limits at $\pm\infty$ are equal). Hence by Wald's theorem we conclude that $\hat{\theta}_n \to_{a.s.} \theta_0$.

2. Use Jensen's inequality to extend the proof of theorem 4.6.2 given in class to the case where ties are possible. That is, suppose that $Y_1, \ldots, Y_m$ are the distinct values appearing in the sample $X_1, \ldots, X_n$ and let $m_j \equiv \#\{i \le n : X_i = Y_j\}$, $q_j \equiv Q(\{Y_j\})$ so that $\sum_{j=1}^{k} m_j = n$, and $\sum_{j=1}^{m} q_j \le 1$. Then show that

$$\prod_{j=1}^{k} q_j^{m_j} \le \prod_{j=1}^{k} (\frac{m_j}{n})^{m_j},$$

and that the resulting maximizer yields the empirical measure $\mathbb{P}_n$.

Solution: Let $Y_1, \ldots, Y_k$ be the distinct values of $X_1, \ldots, X_n$, and let $m_j \equiv \#\{i : X_i = Y_j\}$, $j = 1, \ldots, k$. Thus if Q is an arbitrary measure on $(\mathcal{X}, \mathcal{A})$, and

$q_j \equiv Q(\{Y_j\})$, $j = 1, \ldots, k$, the nonparametric likelihood is

$$L(Q|\underline{X}) = \prod_{j=1}^{k} q_j^{m_j} \equiv L(\underline{q}|\underline{X})$$

which we want to maximize as a function of $\underline{q}$. Now $\log x$ is strictly concave, so it follows by Jensen's inequality that

$$\begin{aligned} \sum_{j=1}^{k} (\frac{m_j}{n}) \log(\frac{q_j}{m_j/n}) &\le \log \left\{ \sum_{j=1}^{k} (\frac{m_j}{n})(\frac{q_j}{m_j/n}) \right\} \\ &= \log \left\{ \sum_{j=1}^{k} q_j \right\} \le \log 1 = 0 \end{aligned}$$

with equality iff $\sum_{j=1}^{k} q_j = 1$ and $q_j/(m_j/n) = 1$ for $j = 1, \ldots, k$. Thus the maximum is unique and occurs at Q_{max} with

$$Q_{max}(A) = \sum_{j=1}^{k} (\frac{m_j}{n}) 1_A(Y_j) = \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{m_j} 1_A(Y_j) = \sum_{i=1}^{n} 1_A(X_i) \equiv \mathbb{P}_n(A) .$$

Note that

$$K(\mathbb{P}_n, Q) = \sum_{j=1}^{k} (\frac{m_j}{n}) \log(\frac{m_j/n}{q_j}) \ge 0 ,$$

with equality iff $Q = \mathbb{P}_n$ giving another derivation of (a).

3. (Interval censoring case 1 or "current status data".) Suppose that (X_i, Y_i), $i = 1, \ldots, n$ are i.i.d. pairs of nonnegative random variables with $X_i \sim F$, $Y_i \sim G$, and X_i, Y_i independent for each i. Suppose that we observe $(Y_i, \delta_i) \equiv (Y_i, 1_{[X_i \le Y_i]})$, $i = 1, \ldots, n$. As noted in class on 1/23, $(\delta_i | Y_i) \sim \text{Bernoulli}(F(Y_i))$, and if G has density g with respect to some dominating measure μ, the joint density of (Y_i, δ_i) is

$$p_{F,g}(y, \delta) = F(y)^{\delta} (1 - F(y))^{1-\delta} g(y).$$

Suppose that we order the observed Y_i's as $0 \le Y_{(1)} \le \ldots \le Y_{(n)}$, assume no ties, denote the corresponding δ's by $\delta_{(i)}$, and set $p_i \equiv F(Y_{(i)})$, $q_i \equiv G(\{Y_{(i)}\})$. Then a nonparametric likelihood for estimation of F and G is given by

$$L_n(F, G | \underline{(Y, \delta)}) = \prod_{i=1}^{n} p_i^{\delta_{(i)}} (1 - p_i)^{1-\delta_{(i)}} q_i \equiv L_n(\underline{p}, \underline{q}) ,$$

and hence the log-likelihood is

$$l_n(\underline{p}, \underline{q})) = \sum_{i=1}^{n} \{\delta_{(i)} \log p_i + (1 - \delta_{(i)}) \log(1 - p_i)\} + \sum_{i=1}^{n} q_i \,.$$

We want to maximize this over $\underline{p} = (p_1, \ldots, p_n)$ and $\underline{q} = (q_1, \ldots, q_n)$ satisfying $0 \le p_1 \le \ldots \le p_n \le 1$ and $q_i \ge 0$, $\sum_{i=1}^{n} q_i \le 1$. The maximum over the q's is easy (since we have done it already) and yields the empirical distribution of the Y_i's as an estimator of G. Suppose that $n = 5$, $Y_{(1)} = 1.2$, $Y_{(2)} = 2.3$, $Y_{(3)} = 2.7$, $Y_{(4)} = 3.1$, $Y_{(5)} = 3.9$, $\delta_{(1)} = \delta_{(3)} = \delta_{(4)} = 1$, $\delta_{(2)} = \delta_{(5)} = 0$. Show that the vector $\underline{\widehat{p}}$ maximizing the loglikelihood is given by $\widehat{p}_1 = \widehat{p}_2 = 1/2$, $\widehat{p}_3 = \widehat{p}_4 = \widehat{p}_5 = 2/3$, and that this corresponds to the left-derivative of the greatest convex minorant of the points $\{(i, \sum_{j \le i} \delta_{(j)}), i = 0, \ldots, n\}$ where $(0, 0)$ corresponds to $i = 0$.

Solution: The cumulative sum diagram for this data, given by the points $\{(i, \sum_{j \le i} \delta_{(j)}), i = 0, \ldots, 5\}$ where $(0, 0)$ corresponds to $i = 0$, is shown in the figure below:

The dotted line indicates the greatest convex minorant H^* of the cumulative sum diagram. The greatest convex minorant H^* has left-derivative $1/2$ at $i = 1$ and $i = 2$, and has left-derivative $2/3$ at $i = 3,\ 4,\ 5$. Thus the MLE of F at $Y_{(j)}$ for $j = 1, \ldots, 5$ is given by $\widehat{p}_1 = \widehat{p}_2 = 1/2$, $\widehat{p}_3 = \widehat{p}_4 = \widehat{p}_5 = 2/3$. Note that this $\underline{\widehat{p}}$ satisfies the conditions of the characterizing proposition proved in class:

$$\begin{aligned}\sum_{j=1}^{5} \widehat{p}_j \left\{ \frac{\delta_{(j)}}{\widehat{p}_j} - \frac{1 - \delta_{(j)}}{1 - \widehat{p}_j} \right\} &= \frac{1}{2}\frac{1}{1/2} + \frac{1}{2}\frac{-1}{1 - 1/2} + \frac{2}{3}\frac{1}{2/3} + \frac{2}{3}\frac{1}{2/3} - \frac{2}{3}\frac{-1}{1 - 2/3} \\ &= 1 - 1 + 1 + 1 - 2 = 0\end{aligned}$$

and

$$S_i \equiv \sum_{j=i}^{5} \left\{ \frac{\delta_{(j)}}{\widehat{p}_j} - \frac{1-\delta_{(j)}}{1-\widehat{p}_j} \right\} \leq 0 \qquad \text{for} \qquad i = 1, \ldots, 5\,;$$

we compute $S_5 = -3$, $S_4 = -3/2$, $S_3 = 0$, $S_2 = -2$, and $S_1 = 0$. Note that the values of i for which $S_i = 0$ correspond (with the exception of the right endpoint) to points where the greatest convex minorant "touches" the cumulative sum diagram. For complete proofs of Claims 1 and 2 from the lecture on 2/2, see Groeneboom and Wellner (1992), pages 40 - 43.