Statistics 582, Problem Set 1 Solutions

Wellner; 1/11/2006

1. Suppose that $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d. with bivariate normal distribution $N_2(\mu, \Sigma)$ where $\mu \in R^2$ and

$$\Sigma = \begin{pmatrix} \sigma^2 & \sigma\tau\rho \\ \sigma\tau\rho & \tau^2 \end{pmatrix}$$

where $\sigma^2 > 0$, $\tau^2 > 0$, and $\rho \in (-1, 1)$.
A. If we assume that $\mu_1 = \mu_2 \equiv \theta$ and Σ is known, what is the MLE of θ?
B. If we assume that μ is known and $\sigma^2 = \tau^2 \equiv \theta$, what is the MLE of θ?
C. What is the asymptotic distribution of the estimator you found in B?
D. Under the same assumption as in B, what is the MLE of ρ?
E. What is the asymptotic distribution of the estimator you found in D?

Solution: A. When $\mu_1 = \mu_2 = \theta$ and Σ is known, then the log-likelihood for one observation is (relabelling $\mu_1 = \mu$, $\mu_2 = \nu$),

$$\log p(x;\theta) = -\frac{1}{2(1-\rho^2)}\left\{\frac{(x-\theta)^2}{\sigma^2} - 2\rho\frac{(x-\theta)}{\sigma}\frac{(y-\theta)}{\tau} + \frac{(y-\theta)^2}{\tau^2}\right\} + constant.$$

Hence the score for θ for one observation is

$$\begin{aligned}
\dot{l}_\theta(x,y) &= \frac{1}{1-\rho^2}\left\{\frac{(x-\theta)}{\sigma^2} + \frac{(y-\theta)}{\tau^2} - \frac{\rho}{\tau}\frac{(x-\theta)}{\sigma} - \frac{\rho}{\sigma}\frac{(y-\theta)}{\tau}\right\} \\
&= \frac{1}{1-\rho^2}\left\{\frac{(x-\theta)}{\sigma}\left(\frac{1}{\sigma} - \frac{\rho}{\tau}\right) + \frac{(y-\theta)}{\tau}\left(\frac{1}{\tau} - \frac{\rho}{\sigma}\right)\right\}.
\end{aligned}$$

Thus the score equation for θ is

$$0 = \dot{l}_{n\theta}(\theta) = \frac{n}{1-\rho^2}\left\{\frac{(\overline{X}_n-\theta)}{\sigma}\left(\frac{1}{\sigma} - \frac{\rho}{\tau}\right) + \frac{(\overline{Y}_n-\theta)}{\tau}\left(\frac{1}{\tau} - \frac{\rho}{\sigma}\right)\right\},$$

and hence

$$\begin{aligned}
\widehat{\theta}_n &= \frac{\overline{X}_n\left(\frac{1}{\sigma^2} - \frac{\rho}{\sigma\tau}\right) + \overline{Y}_n\left(\frac{1}{\tau^2} - \frac{\rho}{\sigma\tau}\right)}{\frac{1}{\sigma^2} - \frac{2\rho}{\sigma\tau} + \frac{1}{\tau^2}} \\
&= a\overline{X}_n + (1-a)\overline{Y}_n
\end{aligned}$$

where

$$a = \frac{\frac{1}{\sigma^2} - \frac{\rho}{\sigma\tau}}{\frac{1}{\sigma^2} - \frac{2\rho}{\sigma\tau} + \frac{1}{\tau^2}}.$$

Note that this yields

$$Var(\widehat{\theta}) = \frac{1}{n}\left\{a^2\sigma^2 + 2a(1-a)\rho\sigma\tau + (1-a)^2\tau^2\right\}.$$

B. and D. If $\sigma^2 = \tau^2 = \theta$ and μ is known, then the log-likelihood for one observation is (again relabelling $\mu_1 = \mu$, $\mu_2 = \nu$),

$$\begin{aligned}\log p(x;\theta,\rho) &= -\log\theta - \frac{1}{2}\log(1-\rho^2) \\ &\quad - \frac{1}{2(1-\rho^2)\theta}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\} + \text{constant.}\end{aligned}$$

Thus the scores for θ and ρ are given by

$$\dot{l}_\theta(x,y) = -\frac{1}{\theta} + \frac{1}{2(1-\rho^2)\theta^2}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\},$$

$$\begin{aligned}\dot{l}_\rho(x,y) &= \frac{\rho}{(1-\rho^2)} - \frac{\rho}{(1-\rho^2)^2\theta}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\} \\ &\quad + \frac{1}{\theta(1-\rho^2)}(x-\mu)(y-\nu).\end{aligned}$$

Hence the score equations for estimation of θ and ρ are given by

$$0 = \dot{l}_{n\theta}(\theta) = \sum_{i=1}^n \dot{l}_\theta(X_i,Y_i) = -\frac{n}{\theta} + \frac{n}{\theta^2 2(1-\rho^2)}\left\{S_{XX} - 2\rho S_{XY} + S_{YY}\right\},$$

and

$$0 = \dot{l}_{n\rho}(\rho) = \sum_{i=1}^n \dot{l}_\rho(X_i,Y_i) = \frac{n\rho}{1-\rho^2} - \frac{n\rho}{\theta(1-\rho^2)^2}\left\{S_{XX} - 2\rho S_{XY} + S_{YY}\right\} + \frac{n}{\theta(1-\rho^2)}S_{XY}$$

where

$$S_{XX} \equiv n^{-1}\sum_{i=1}^n (X_i-\mu)^2, \quad S_{XY} \equiv n^{-1}\sum_{i=1}^n (X_i-\mu)(Y_i-\nu), \quad S_{YY} \equiv n^{-1}\sum_{i=1}^n (Y_i-\nu)^2.$$

Solving the first of these for $\widehat{\theta}$ yields

$$\widehat{\theta} = \frac{1}{2(1-\widehat{\rho}^2)}\left\{S_{XX} - 2\widehat{\rho} S_{XY} + S_{YY}\right\};$$

Rewriting the score equation for ρ with a common denominator of $\theta(1-\rho^2)^2$ yields

$$\theta\rho(1-\rho^2) - \rho\left\{S_{XX} - 2\rho S_{XY} + S_{YY}\right\} + (1-\rho^2)S_{XY} = 0;$$

and then plugging in the estimator $\widehat{\theta}$ of θ yields the equation

$$(1-\widehat{\rho}^2)S_{XY} = \frac{1}{2}\widehat{\rho}\left\{S_{XX} - 2\widehat{\rho} S_{XY} + S_{YY}\right\}.$$

This has the solution

$$\widehat{\rho} = \frac{2S_{XY}}{S_{XX} + S_{YY}};$$

plugging this (or more precisely the last form of the equation for $\widehat{\rho}$) into the expression for θ yields $\widehat{\theta} = (S_{XX} + S_{YY})/2$.

C. and E. To find the asymptotic distributions of $\widehat{\theta}$ and $\widehat{\rho}$ we could either (i) proceed directly from first principles (central limit theorms and the delta method), or (b) use theorem 4.1.5 concerning the asymptotic behavior of maximum likelikhood estimators. I'll take the second route here. The first step in this direction is to compute the information matrix for (θ, ρ). Now

$$\ddot{l}_{\theta\theta}(x,y) = \frac{1}{\theta^2} - \frac{1}{(1-\rho^2)\theta^3}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\},$$

$$\begin{aligned}\ddot{l}_{\theta\rho}(x,y) &= \frac{2\rho}{2\theta^2(1-\rho^2)^2}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\} \\ &\quad - \frac{1}{2\theta^2(1-\rho^2)^2}2(x-\mu)(y-\nu),\end{aligned}$$

and

$$\begin{aligned}\ddot{l}_{\rho\rho}(x,y) &= \frac{1}{1-\rho^2} + \frac{2\rho^2}{(1-\rho^2)^2} \\ &\quad - \left\{\frac{1}{(1-\rho^2)^2} + \frac{4\rho^2}{(1-\rho^2)^3}\right\}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\} \\ &\quad + \frac{\rho}{\theta(1-\rho^2)^2}2(x-\mu)(y-\nu) + \frac{2\rho}{\theta(1-\rho^2)^2}(x-\mu)(y-\nu) \\ &= \frac{1+\rho^2}{(1-\rho^2)^2} - \frac{1-3\rho^2}{\theta(1-\rho^2)^3}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\} \\ &\quad + \frac{4\rho}{\theta(1-\rho^2)^2}(x-\mu)(y-\nu).\end{aligned}$$

Here

$$E\left\{(X-\mu)^2 - 2\rho(X-\mu)(Y-\nu) + (Y-\nu)^2\right\} = 2\theta(1-\rho^2)$$

and

$$E(X-\mu)(Y-\nu) = \rho\theta.$$

Thus we find that

$$I_{\theta\theta} = E(-\ddot{l}_{\theta\theta}(X,Y)) = \theta^{-2},$$

$$I_{\theta\rho} = E(-\ddot{l}_{\theta\rho}(X,Y)) = \frac{\rho}{\theta(1-\rho^2)},$$

and

$$I_{\rho\rho} = E(-\ddot{l}_{\rho\rho}(X,Y)) = \frac{1+\rho^2}{(1-\rho^2)^2}.$$

This yields

$$I_{\theta\theta\cdot\rho} = I_{\theta\theta} - I_{\theta\rho}I_{\rho\rho}^{-1}I_{\rho\theta} = \frac{1}{\theta^2}\frac{1}{1+\rho^2}$$

and

$$I_{\rho\rho\cdot\theta} = I_{\rho\rho} - I_{\rho\theta}I_{\theta\theta}^{-1}I_{\theta\rho} = (1-\rho^2)^{-2}.$$

Hence it follows from theorem 4.1.5 that

$$\sqrt{n}(\widehat{\theta}_n - \theta) \to_d N(0, I^{-1}_{\theta\theta\cdot\rho}) = N(0, \theta^2(1+\rho^2))$$

while

$$\sqrt{n}(\widehat{\rho}_n - \rho) \to_d N(0, I^{-1}_{\rho\rho\cdot\theta}) = N(0, (1-\rho^2)^2)\,.$$

2. Problem 1, page 117, Ferguson, ACILST: Suppose that $X_1, \ldots, X_n$ are i.i.d. Uniform$(0,\theta)$ with $p(x,\theta) = \theta^{-1}1_{[0,\theta]}(x)$ and $\Theta = [1,2]$. Check the conditions of Wald's consistency theorem (theorem 17, Ferguson page 114; theorem 4.3, chapter 4, page 28) in this case. What happens if $\Theta = [1,\infty)$ or $(0,\infty)$?

   Solution: (i) Ferguson's problem: here is a verification of the five conditions (a)-(e) when $\Theta = [1,2]$:
   (a) $\Theta = [1,2]$ is closed and bounded, hence compact.
   (b) For fixed $x \le 1$, $p(x,\theta) = 1/\theta$ is continuous; for $1 < x \le 2$, $p(x,\theta) = \theta^{-1}1\{\theta \ge x$, which is upper semi-continuous. For $x > 2$, $p(x,\theta) = 0$ which is a continuous function of θ. Since log is a continuous function on $(0,\infty)$, these (semi-)continuities carry over to $f(x,\theta) = \log p(x,\theta) - \log p(x,\theta_0)$.
   (c) Fix $\theta_0 \in \$Theta$. Then

$$\frac{p(x,\theta)}{p(x,\theta_0)} = \frac{\theta_0}{\theta}\frac{1_{[0,\theta}}{(}x)1_{[0,\theta_0]} = \begin{cases} (\theta_0/\theta), & x \le 1 \\ (\theta_0/\theta)1_{[x,\infty)}(\theta), & 1 < x \le \theta_0 \\ \infty, & x > \theta_0, \end{cases}$$

   so

$$\sup_{\theta\in\Theta}\frac{p(x,\theta)}{p(x,\theta_0)} \equiv K(x) = \theta_0 1\{x \le 1\} + \frac{\theta_0}{x}1_{(1,\theta_0]}(x) + \infty 1_{(\theta_0,\infty)}(x),$$

   and

$$\sup_{\theta\in\Theta}\log\frac{p(x,\theta)}{p(x,\theta_0)} \equiv F(x) = (\log\theta_0)1\{x \le 1\} + \log(\theta_0/x)1_{(1,\theta_0]}(x) + \infty 1_{(\theta_0,\infty)}(x),$$

   satisfies $E_{\theta_0}F(X) < \infty$.
   (d) The function $\varphi(x,\theta,\rho)$ is given by

$$\varphi(x,\theta,\rho) = \sup_{\theta':|\theta'-\theta|<\rho}\frac{1}{\theta'}1_{[0,\theta']}(x) = \begin{cases} 1/(\theta-\rho), & x < \theta-\rho \\ 1/x, & |x-\theta| \le \rho \\ 0, & x > \theta+\rho, \end{cases}$$

   which is clearly measurable.
   (e) If $p(x,\theta') = p(x,\theta)$ a.e. Lebesgue, then

$$0 = \frac{1}{2}\int |p(x,\theta) - p(x,\theta')|dx = d_{TV}(P_\theta, P_{\theta'}) = 1 - \eta(P_\theta, P_{\theta'}) = 1 - \frac{\theta' \wedge \theta}{\theta' \vee \theta}$$

   where the last equality follows by a computation as in the final exam for stat 581, Fall '05. This yields $\theta' \vee \theta = \theta' \wedge \theta$, which implies $\theta = \theta'$.
   (ii) When $\Theta = [1,\infty)$ or $\Theta = (0,\infty)$, then Θ is no longer compact, and Wald's theorem does not apply directly. One way to remedy the problem is to compactify

the set Θ by some appropriate identification of points in Θ with points on the unit half-circle (as shown in class on 1/9/06). Another way to proceed is to show that the MLE is in a compact set eventually with probability 1; see e.g. van der Vaart's re-working of Wald's theorem. In this case the MLE is $\widehat{\theta}_n = \max_{1 \le i \le n} X_i = X_{(n)}$, and it follows that for any $\theta_0 > \delta > 0$

$$P_{\theta_0}(\theta_0 - \delta > \widehat{\theta}_n) = \left(\frac{\theta_0 - \delta}{\theta_0}\right)^n,$$

which has a finite sum on n, and hence by the Borel-Cantelli lemma

$$P_{\theta_0}(\widehat{\theta}_n < \theta_0 - \delta \ \text{ infinitely often}) = 0,$$

or, equivalently,

$$P_{\theta_0}(\theta_0 \ge \widehat{\theta}_n \ge \theta_0 - \delta \ \text{ almost always}) = 1.$$

Note that this argument already yields almost sure consistency of the MLE in this case since δ can be chosen to be arbitrarily small.

3. Suppose that $X, X_1, \ldots, X_n$ are i.i.d. Weibull(α_0, β_0) (if X has the Weibull(θ) distribution where $\theta = (\alpha, \beta)$, then $1 - F_\theta(x) = P_\theta(X > x) = \exp(-(x/\alpha)^\beta)$ for $x \ge 0$). Recall that the MLE $\widehat{\alpha}$ of α is given by

$$\widehat{\alpha} = \left\{\frac{1}{n}\sum_{i=1}^n X_i^{\widehat{\beta}}\right\}^{1/\widehat{\beta}}$$

where $\widehat{\beta}$ is the MLE of β. As a simpler alternative to maximum likelihood, I propose to use the alternative estimator $\overline{\beta}$ of β obtained from the slope of an ordinary least squares fit of a Weibull Q-Q plot, and then estimate α by

$$\overline{\alpha}_n = \left\{\frac{1}{n}\sum_{i=1}^n X_i^{\overline{\beta}}\right\}^{1/\overline{\beta}}.$$

A. Suppose that $\overline{\beta}_n \to_p \beta_0$ is known. Show that $\overline{\alpha}_n \to_p \alpha_0$. [Hint: use a uniform strong law of large numbers.]
B. Show that $\overline{\alpha}_n$ is a "pseudo-MLE" in the sense that $\overline{\alpha}_n$ maximizes $l_n(\alpha, \overline{\beta}_n)$.

Solution: Fix $\delta > 0$ (small). The family of functions $\mathcal{F} = \{f(x, \beta) = x^\beta : \beta \in [\beta_0 - \delta, \beta_0 + \delta]\}$ are indexed by the compact set $[\beta_0 - \delta, \beta_0 + \delta]$, are continuous in β for every $x \ge 0$, and are bounded by

$$\sup_{\beta \in [\beta_0 - \delta, \beta_0 + \delta]} |f(x, \beta)| = x^{\beta_0 + \delta} \vee x^{\beta_0 - \delta} \le x^{\beta_0 + \delta} + x^{\beta_0 - \delta} \equiv F(x)$$

which satisfies $E_0 F(X) < \infty$ if $\delta < 2\beta_0$. Thus by theorem 4.4.1 (of the section 4 revision) the uniform strong law of large numbers holds for $\mathcal{F}$:

$$\sup_{\beta : |\beta - \beta_0| \le \delta} |\mathbb{P}_n f(\cdot, \beta) - P_0 f(\cdot, \beta)| \to_{a.s.} 0\,.$$

If $\overline{\beta}_n \rightarrow_{a.s.} \beta_0$, $\overline{\beta}_n \in [\beta_0 - \delta, \beta_0 + \delta]$, with probability 1 for n sufficiently large, and it follows from the uniform strong law law of large numbers (Theorem 1, section 4.4 revision) together with continuity of $\mu(\beta) \equiv E_0 f(X, \beta)$ that

$$\begin{aligned} \overline{\alpha}_n^{\overline{\beta}_n} &= \frac{1}{n}\sum_{i=1}^n X_i^{\overline{\beta}_n} \\ &\rightarrow_{a.s.} E_0 f(X, \beta_0) = \alpha_0^{\beta_0} . \end{aligned}$$

(If instead $\overline{\beta}_n \rightarrow_p \beta_0$, then for and given $\epsilon > 0$ and $n \geq N_{\epsilon,\delta}$ large, $P_{\theta_0}(\overline{\beta}_n \in [\beta_0 - \delta, \beta_0 + \delta]) > 1 - \epsilon$ and we can simply argue on this set.) But now

$$\overline{\alpha}_n = \{\overline{\alpha}_n^{\overline{\beta}_n}\}^{1/\overline{\beta_n}} = g(\overline{\alpha}^{\overline{\beta}_n}, \overline{\beta_n})$$

where $g(u, v) \equiv u^{1/v}$ is continuous and $(\overline{\alpha}^{\overline{\beta}_n}, \overline{\beta_n}) \rightarrow_{a.s.} (\alpha_0^{\beta_0}, \beta_0)$. Hence by the continuous mapping theorem

$$\overline{\alpha}_n = g(\overline{\alpha}_n^{\overline{\beta}_n}, \overline{\beta_n}) \rightarrow_{a.s.} g(\alpha_0^{\beta_0}, \beta_0) = \alpha_0 .$$

B. The log-likelihood is

$$l_n(\alpha, \beta) = n \log(\beta/\alpha) + (\beta - 1)\sum_{i=1}^n \log(X_i/\alpha) - \sum_{i=1}^n \left(\frac{X_i}{\alpha}\right)^\beta ,$$

and hence

$$\begin{aligned} l_n(\alpha, \overline{\beta}_n) &= n\log(\overline{\beta}_n/\alpha) + (\overline{\beta}_n - 1)\sum_{i=1}^n \log(X_i/\overline{\alpha}) - \sum_{i=1}^n \left(\frac{X_i}{\alpha}\right)^{\overline{\beta}_n} \\ &= -n\overline{\beta}_n \log\alpha - \frac{\sum X_i^{\overline{\beta}_n}}{\alpha^{\overline{\beta}_n}} + \text{constant in } \alpha \\ &= -n\log\eta - \frac{\sum X_i^{\overline{\beta}_n}}{\eta} + \text{constant in } \alpha \text{ and } \eta \end{aligned}$$

where $\eta \equiv \alpha^{\overline{\beta}_n}$. This is easily seen to be maximized by

$$\overline{\eta} \equiv \frac{1}{n}\sum_{i=1}^n X_i^{\overline{\beta}_n}$$

and hence

$$\overline{\alpha}_n = \left\{\frac{1}{n}\sum_{i=1}^n X_i^{\overline{\beta}_n}\right\}^{1/\overline{\beta}_n}$$

as claimed. Thus $\overline{\alpha}_n$ is a pseudo-MLE of α.

4. (Profile likelihood) [For nice plots to accompany this exercise, see pages 41 - 43 of Cox, D. R. and Oakes, D. (1984); *Analysis of Survival Data*, Chapman and Hall.]

As in problem 1.3, consider the Weibull family of example 3.2.5: $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^{+2}$ given by the (Lebesgue) densities

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

where $\theta \equiv (\alpha, \beta) \in (0,\infty)\times(0,\infty) \subset \mathbb{R}^2$.
A. For a sample of n observations from p_θ, we know that, for each fixed value of β the value of α which maximizes the likelihood as a function of α is

$$\widehat{\alpha}(\beta) = \{\frac{1}{n}\sum_{i=1}^n X_i^\beta\}^{1/\beta}.$$

Use this to compute the *profile likelihood* $l_{\text{profile}}(\beta) = l_{\text{profile}}(\beta|\underline{X})$ defined by

$$l_{\text{profile}}(\beta) = l(\widehat{\alpha}(\beta), \beta) = l(\widehat{\alpha}(\beta), \beta|\underline{X}).$$

B. Use what we know from problem 10.3 to show that the profile likelihood is strictly concave and hence has a unique maximum. Show that maximizing the profile likelihood as a function of β yields the maximum likelihood estimate: i.e. that $(\widehat{\alpha}, \widehat{\beta}) = (\widehat{\alpha}(\widehat{\beta}_{\text{profile}}), \widehat{\beta}_{\text{profile}})$.

C. What is the relationship of the score function for β from the profile likelihood, $\dot{l}_{\beta,\text{profile}}$ to the (efficient) score for β from the full likelihood? Prove or disprove my claim: the profile score for β (based on n observations) is asymptotically equivalent to the sum of efficient scores for β over the sample in the sense that their difference divided by $\sqrt{n}$ converges to 0 in probability.
D. What is the relationship of the observed information from the profile likelihood $-\ddot{l}_{\beta\beta,\text{profile}}$ to information quantities from the full likelihood?

Solution: A. The log-likelihood is

$$l(\alpha,\beta) = n\log(\beta/\alpha) + (\beta-1)\sum_{i=1}^n \log(\frac{X_i}{\alpha}) - \sum_{i=1}^n(\frac{X_i}{\alpha})^\beta$$

and for fixed β the value of α which maximizes this is

$$\hat{\alpha}(\beta) = (\frac{1}{n}\sum_{i=1}^n X_i^\beta)^{1/\beta}.$$

Thus the profile log-likelihood is

$$l_{profile}(\beta) = l(\hat{\alpha}(\beta),\beta) = n\log\beta - n\log(\sum_{i=1}^n X_i^\beta) + (\beta-1)\sum_{i=1}^n \log X_i + n\log n - n.$$

B. It follows that the score function for β corresponding to the profile log-likelihood is

$$\dot{\text{l}}_{\text{profile},\beta}(\underline{X}) = \frac{n}{\beta} - n\frac{\sum_{i=1}^n X_i^\beta \log X_i}{\sum_{i=1}^n X_i^\beta} + \sum_{i=1}^n \log X_i,$$

and the observed information is

$$-\ddot{\mathbf{l}}_{profile,\beta}(\underline{X}) = \frac{n}{\beta^2} + n\left\{ \frac{\sum_{i=1}^n X_i^\beta (\log X_i)^2}{\sum_{i=1}^n X_i^\beta} - \left(\frac{\sum_{i=1}^n X_i^\beta \log X_i}{\sum_{i=1}^n X_i^\beta} \right)^2 \right\}$$
$$> 0$$

since the term in brackets is a variance, and hence is positive. Thus the profile likelihood is strictly concave and its maximum is unique.

Let $l^{\#}(\beta) = l_{profile}(\beta|\underline{X})$. Then, by the chain rule,

$$\dot{\mathbf{l}}_\beta^{\#}(\underline{X}) = \dot{\mathbf{l}}_{n\alpha}|_{\hat{\alpha}(\beta)}\dot{\alpha}(\beta) + \dot{\mathbf{l}}_{n\beta}|_{\hat{\alpha}(\beta)} = \dot{\mathbf{l}}_{n\beta}|_{\hat{\alpha}(\beta)} \tag{0.1}$$

since

$$\dot{\mathbf{l}}_{n\alpha}|_{\hat{\alpha}(\beta)} = 0. \tag{0.2}$$

Hence solving the profile score equation $\dot{\mathbf{l}}_\beta^{\#}(\underline{X}) = 0$ yields a solution of the likelihood equations $\dot{\mathbf{l}}_{n\alpha} = 0$ and $\dot{\mathbf{l}}_{n\beta} = 0$.

C. As we have seen in B, (0.1) and (0.2) hold by definition of profile (log-)likelihood. Expanding (0.2) about α yields

$$0 = \dot{\mathbf{l}}_{n\alpha}|_{\hat{\alpha}(\beta)} = \dot{\mathbf{l}}_{n\alpha} + \ddot{\mathbf{l}}_{n\alpha\alpha}(\alpha^*)(\hat{\alpha}(\beta) - \alpha),$$

and hence

$$\hat{\alpha}(\beta) - \alpha = -\ddot{\mathbf{l}}_{n\alpha\alpha}^{-1}(\alpha^*)\dot{\mathbf{l}}_{n\alpha}. \tag{0.3}$$

Expanding (0.1) about α yields

$$\dot{\mathbf{l}}_\beta^{\#}(\underline{X}) = \dot{\mathbf{l}}_{n\beta}|_{\hat{\alpha}(\beta)} = \dot{\mathbf{l}}_{n\beta} + \ddot{\mathbf{l}}_{n\beta\alpha}(\alpha^{**})(\hat{\alpha}(\beta) - \alpha). \tag{0.4}$$

Substitution of (0.3) into (0.4) yields

$$\begin{aligned} \dot{\mathbf{l}}_\beta^{\#}(\underline{X}) &= \dot{\mathbf{l}}_{n\beta} + \ddot{\mathbf{l}}_{n\beta\alpha}(\alpha^{**})(\hat{\alpha}(\beta) - \alpha) \\ &= \dot{\mathbf{l}}_{n\beta} - \ddot{\mathbf{l}}_{n\beta\alpha}(\alpha^{**})\ddot{\mathbf{l}}_{n\alpha\alpha}^{-1}(\alpha^*)\dot{\mathbf{l}}_{n\alpha}. \end{aligned}$$

It follows easily from this relationship (under the conditions A0 - A4, the Cramér conditions) that

$$\begin{aligned} &\frac{1}{\sqrt{n}}\{\dot{\mathbf{l}}_\beta^{\#}(\underline{X}) - \mathbf{l}_{n\beta}^*\} \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \{I_{\beta,\alpha}I_{\alpha\alpha}^{-1}\dot{\mathbf{l}}_\alpha(X_i) - (-\ddot{\mathbf{l}}_{n\beta\alpha}(\alpha^{**}))(-\ddot{\mathbf{l}}_{n\alpha\alpha}^{-1}(\alpha^*))\dot{\mathbf{l}}_\alpha(X_i)\} \\ &= o_p(1). \end{aligned}$$

D. Differentiation across (0.2) with respect to β yields

$$\ddot{\mathbf{l}}_{n\alpha\alpha}\dot{\alpha}(\beta) + \ddot{\mathbf{l}}_{n\alpha\beta} = 0,$$

and hence

$$\dot{\alpha}(\beta) = -\ddot{\mathbf{l}}_{n\alpha\beta}\ddot{\mathbf{l}}_{n\alpha\alpha}^{-1}. \tag{0.5}$$

Similarly, differentiation of (0.1) with respect to β and then plugging in (0.5), yields

$$\ddot{\mathbf{l}}_{\beta\beta}^{\#} = \ddot{\mathbf{l}}_{n\beta\beta} + \ddot{\mathbf{l}}_{n\beta\alpha}\dot{\alpha}(\beta) = \ddot{\mathbf{l}}_{n\beta\beta} - \ddot{\mathbf{l}}_{n\beta\alpha}\ddot{\mathbf{l}}_{n\alpha\alpha}^{-1}\ddot{\mathbf{l}}_{n\alpha\beta}.$$

With some added minus signs, this becomes

$$-\ddot{\mathbf{l}}_{\beta\beta}^{\#} = \hat{I}_{n\beta\beta\cdot\alpha}|_{\hat{\alpha}(\beta)}.$$

5. **Optional bonus problem.** On pages 116-117 of ACILST, Ferguson (see also Ferguson, T. S. (1982). An inconsistent maximum likelihood estimate. *J. Amer. Statist. Assoc.* **77**, 831–834) shows that $\hat{\theta}_n \rightarrow_{a.s.} 1$ no matter what θ_0 is true if $\delta(\theta) \rightarrow 0$ "fast enough".
(a) Show that $\hat{\theta}_n \rightarrow_{a.s.} 1$ continues to hold if

$$\delta(\theta) = (1-\theta)\exp(-(1-\theta)^{-c} + 1)$$

with $c > 2$. (Ferguson shows that $c = 4$ works.)
(b) Show that when $c = 2$, Ferguson's argument yields

$$\sup_{0\le\theta\le 1} n^{-1}\log L_n(\theta) \ge \frac{n-1}{n}\log(M_n/2) + \frac{1}{n}\log\frac{1-M_n}{\delta(M_n)} \rightarrow_d D$$

where

$$P(D \le y) = \exp\left(-\frac{1}{2(y-\log 2)}\right), \quad y \ge \log(2).$$

That is, $D \stackrel{d}{=} \log 2 + 1/(2E)$ where E is an Exponential(1) random variable.

Solution (a) Note that $\log[(1-\theta)/\delta(\theta)] = (1-\theta)^{-c} - 1$, so

$$\frac{1}{n}\log\frac{1-M_n}{\delta(M_n)} = \frac{1}{n(1-M_n)^c} - \frac{1}{n} \rightarrow_{a.s.} \infty$$

if $n(1-M_n)^c \rightarrow_{a.s.} 0$. But

$$\begin{aligned} P_0(n(1-M_n)^c > \epsilon) &= P(X_1 < 1-(\epsilon/n)^{1/c})^n = (1 - P_0(X_1 \ge 1-(\epsilon/n)^{1/c})^n \\ &= \left(1 - \frac{1}{2}(\epsilon/n)^{2/c}\right)^n \le \exp(-(1/2)\epsilon^{2/c}n^{1-2/c}) \end{aligned}$$

which has a finite sum on n if $c > 2$. Thus by Borel-Cantelli, $P(n(1-M_n)^c > \epsilon \ \ i.o.) = 0$ and $n(1-M_n)^c \rightarrow_{a.s.} 0$ if $c > 2$.
(b) When $c = 2$, the above argument shows that

$$P_0(n(1-M_n)^2 > 2t) = \left(1 - \frac{1}{2}\frac{2t}{n}\right)^n \rightarrow \exp(-t), \qquad \text{for } \ t > 0,$$

or $(1/2)n(1-M_n)^2 \rightarrow_d E \sim \text{Exponential}(1)$. Therefore

$$\begin{aligned}\sup_{0\le\theta\le1} n^{-1}\log L_n(\theta) &\ge \frac{n-1}{n}\log(M_n/2) + \frac{1}{n}\log\frac{1-M_n}{\delta(M_n)} \\ &= \frac{n-1}{n}\log(M_n/2) + \frac{1}{2(1/2)n(1-M_n)^2} - \frac{1}{n} \\ &\rightarrow_d \log(1/2) + \frac{1}{2E} = -\log 2 + \frac{1}{2E}.\end{aligned}$$