Statistics 582, Problem Set 1 Solutions

Wellner; 1/12/2010

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 $\tau^2 = c^2\sigma^2 \equiv c^2\theta$ with $c > 0$ fixed and known, what is the MLE of $\theta = \sigma^2$?
(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 $\tau^2 = c^2\sigma^2 = c^2\theta$ and μ is known, then the log-likelihood for one observation is (again re-labelling $\mu_1 = \mu$, $\mu_2 = \nu$),

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

Thus the scores for θ and ρ are given by

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

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

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

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

and

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

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

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

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

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

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

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

This has the solution

$$\widehat{\rho} = \frac{2S_{XY}/c}{S_{XX} + S_{YY}/c^2} = \frac{2cS_{XY}}{c^2 S_{XX} + S_{YY}};$$

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

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

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

and

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

Here

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

and

$$E(X-\mu)\frac{(Y-\nu)}{c} = \rho\theta.$$

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

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

while

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

2. Ferguson, ACILST, page 118, problem 3. [Also see Lehmann and Casella, Example 7.9, page 482.]

Solution: (a) The likelihood for the data is

$$L_n(\mu_1, \ldots, \mu_n, \sigma^2) = \prod_{j=1}^{d}\prod_{i=1}^{n} \frac{1}{\sigma}\phi\left(\frac{X_{ij}-\mu_i}{\sigma}\right),$$

where ϕ is the standard normal density, so the log-likelihood is

$$l_n(\mu_1, \ldots, \mu_n, \sigma^2) = \sum_{j=1}^{d}\sum_{i=1}^{n}\left\{-\frac{1}{2}\log(\sigma^2) - \frac{(X_{ij}-\mu_i)^2}{2\sigma^2}\right\} + \text{constant},$$

and the scores equations are:

$$0 = \frac{\partial}{\partial\mu_i} l_n(\mu_1, \ldots, \mu_n, \sigma^2) = \sum_{j=1}^{d}\frac{(X_{ij}-\mu_i)}{\sigma^2}, \quad i = 1, \ldots, n,$$

$$0 = \frac{\partial}{\partial(\sigma^2)} l_n(\mu_1, \ldots, \mu_n, \sigma^2) = \sum_{j=1}^{d}\sum_{i=1}^{n}\left\{-\frac{1}{2\sigma^2} + \frac{(X_{ij}-\mu_i)^2}{2(\sigma^2)^2}\right\}.$$

Solving the first n score equations for the $\hat{\mu}_i$'s yields

$$\hat{\mu}_i = \frac{1}{d}\sum_{j=1}^{d} X_{ij} \equiv X_{i\cdot},$$

and then, the solution of the score equation for $\hat{\sigma}^2$ yields

$$\hat{\sigma}^2 = \frac{1}{nd}\sum_{j=1}^{d}\sum_{i=1}^{n}(X_{ij}-\hat{\mu}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{d}\sum_{j=1}^{d}(X_{ij}-X_{i\cdot})^2.$$

Since

$$\frac{\partial^2}{\partial\mu_i^2} l_n(\mu_1, \ldots, \mu_n, \sigma^2) = -\frac{d}{\sigma}^2,$$

$$\frac{\partial^2}{\partial\mu_{i'}\mu_i} l_n(\mu_1, \ldots, \mu_n, \sigma^2) = 0, \quad i' \neq i, \quad \text{and}$$

$$\frac{\partial^2}{\partial(\sigma^2)^2} l_n(\hat{\mu}_1, \ldots, \hat{\mu_n}, \hat{\sigma}^2) = -\frac{nd}{\hat{\sigma}^4},$$

it follows that $(\hat{\mu}_1, \ldots, \hat{\mu}_n, \hat{\sigma}^2)$ is the MLE.
(b) When $d > 1$ is fixed, the random vectors

$$(Y_{i,j}, j = 1, \ldots, d) \equiv (X_{i,j} - X_{i,\cdot}, j = 1, \ldots, d)$$

are i.i.d. $N_d(0, \sigma^2(I - \underline{1}\underline{1}^T/d))$. It follows easily that

$$Z_i^2 \equiv \sum_{j=1}^{d} (X_{i,j} - X_{i,\cdot})^2 \stackrel{d}{=} \sigma^2 \chi^2_{d-1}$$

are i.i.d. with expected value $E(Z_1^2) = \sigma^2(d-1)$. Thus it follows by the weak law of large numbers that for fixed d

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{Z_i^2}{d} \rightarrow_p \sigma^2 \frac{d-1}{d}.$$

Theorem 17 of Ferguson and our theory from Chapter 4 sections 1 and 2 does not apply because the dimension of the parameter space changes (indeed grows) with the sample size n.
(c) An obvious consistent estimator of σ^2 is given by

$$\tilde{\sigma}^2 \equiv \frac{d}{d-1} \hat{\sigma}^2.$$

3. Ferguson, ACILST, page 117, problem 2, with parameter space $\Theta = [0, 1]$.

Solution: (a) Let $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ denote the order statistics of the sample $X_1, \ldots, X_n$. Now

$$p_\theta(x) = 2 \left(\frac{x}{\theta} 1_{[0,\theta]}(x) + \frac{1-x}{1-\theta} 1_{(\theta,1]}(x) \right),$$

so the likelihood function is

$$\begin{aligned} L_n(\theta|\underline{X}) &= \prod_{i=1}^{n} 2 \left\{ \frac{X_i}{\theta} 1_{[0,\theta]}(X_i) + \frac{1-X_i}{1-\theta} 1_{(\theta,1]}(X_i) \right\} \\ &= \prod_{i=1}^{n} 2 \left\{ \frac{X_{(i)}}{\theta} 1_{[0,\theta]}(X_{(i)}) + \frac{1-X_{(i)}}{1-\theta} 1_{(\theta,1]}(X_{(i)}) \right\} \\ &= \left(\frac{2}{\theta} \right)^k \prod_{j=1}^{k} X_{(j)} \cdot \left(\frac{2}{1-\theta} \right)^{n-k} \prod_{j=k+1}^{n} (1 - X_{(j)}) \ \text{ if } \ X_{(k)} \le \theta < X_{(k+1)}. \end{aligned}$$

Thus

$$l_n(\theta|\underline{X}) = \log L_n(\theta|\underline{X}) = -k \log \theta - (n-k) \log(1-\theta) + \text{const. in } \theta,$$

for $X_{(k)} < \theta < X_{(k+1)}$, and on this interval

$$\dot{l}_{n,\theta}(\theta|\underline{X}) = -\frac{k}{\theta} + \frac{n-k}{1-\theta} = \frac{n\theta - k}{\theta(1-\theta)} \begin{cases} > 0, & \text{if } \theta > k/n, \\ < 0, & \text{if } \theta < k/n, \end{cases}$$

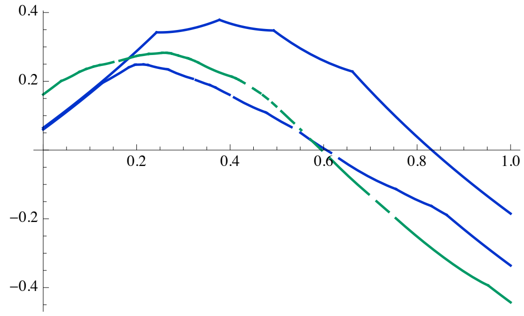

Figure 1: Three log-likelihood functions for problem 3, $n = 4$, $n = 10$, and $n = 50$

so l_n and L_n are decreasing on $X_{(k)} \le \theta < X_{(k+1)} \wedge k/n$ and increasing on $X_{(k)} \vee k/n \le \theta < X_{(k+1)}$.

(b) First note that $L_n(\theta)$ is continuous. Moreover, by the computation in (a), it can only have local *minima* at the solutions of the likelihood equations (which can only occur at the points k/n, $k = 1, \ldots, n$), and hence the local maxima of the likelihood occur only at the order statistics. Furthermore, if $(k-1)/n < X_{(k)} < k/n$, then the log-likelihood $l_n(\theta)$ and the likelihood function $L_n(\theta)$ has a local maximum at $X_{(k)}$: if $k/n > \theta > X_{(k)}$ then $\dot{l}_{n,\theta}(\theta|\underline{X}) < 0$ from (a), while if $(k-1)/n < \theta < X_{(k)}$, then $\dot{l}_{n,\theta}(\theta|\underline{X}) > 0$ also by (a).

Here is a plot of the two examples of the likelihood function for samples of size $n = 4, 10$, and $n = 50$.

4. Ferguson, ACILST, page 124, problem 3. What can you say about the asymptotic distribution of the MLE of $\theta = (\theta_1, \theta_2)$?

Solution: Now $p_\theta(x) = \exp(-\theta_2 \cosh(x - \theta_1) - \zeta(\theta_2))$ where

$$\zeta(\theta_2) \equiv \log\left(\int_{-\infty}^{\infty} \exp\left(-\theta_2 \cosh(y)\right) dy\right).$$

Thus $\log p_\theta(x) = -\theta_2 \cosh(x - \theta_1) - \zeta(\theta_2)$ and the scores for θ_1 and θ_2 (for sample size $n = 1$) are given by

$$\begin{aligned} \dot{l}_1(x) &= \theta_2 \sinh(x - \theta_1), \\ \dot{l}_2(x) &= -\cosh(x - \theta_1) - \zeta'(\theta_2). \end{aligned}$$

It follows that the likelihood equations are:

$$\begin{aligned} 0 &= l_{n,1}(\underline{X}) = \sum_{i=1}^n \dot{l}_1(X_i) = \theta_2 \sum_{i=1}^n \sinh(X_i - \theta_1), \\ 0 &= l_{n,2}(\underline{X}) = \sum_{i=1}^n \dot{l}_2(X_i) = -\sum_{i=1}^n \left(\cosh(X_i - \theta_1) + \zeta'(\theta_2)\right). \end{aligned}$$

Now $E\dot{l}_1(X_1) = 0$ and $\theta_2 > 0$ imply that $E\sinh(X_1 - \theta_1) = 0$. Similarly, $E\dot{l}_2(X_1) = 0$ implies that $E\cosh(X_1 - \theta_1) = -\zeta'(\theta_2)$. Since we can easily compute

$$\begin{aligned} \ddot{l}_{11}(x) &= -\theta_2 \cosh(x - \theta_1), \\ \ddot{l}_{12}(x) &= -\sinh(x - \theta_1), \\ \ddot{l}_{22}(x) &= -\zeta''(\theta_2), \end{aligned}$$

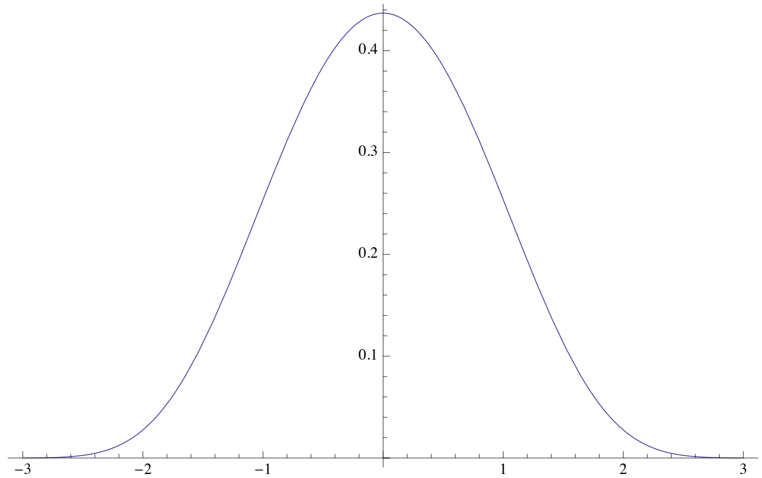

Figure 2: The density $p_\theta(x)$ in problem 4 with $\theta = (0, 1)$

it follows that

$$\begin{aligned} -E\ddot{l}_{11}(X_1) &= \theta_2 E \cosh(X_1 - \theta_1) = -\theta_2 \zeta'(\theta_2), \\ -E\ddot{l}_{12}(X_1) &= E \sinh(X_1 - \theta_1) = 0, \\ -E\ddot{l}_{22}(X_1) &= \zeta''(\theta_2), \end{aligned}$$

and hence the information matrix for a sample of size $n = 1$ is

$$I(\theta) = \begin{pmatrix} -\theta_2 \zeta'(\theta_2) & 0 \\ 0 & \zeta''(\theta_2) \end{pmatrix}.$$

From our theory in Section 4.1 it follows that

$$\sqrt{n}(\hat{\theta}_n - \theta) \to_d N_2(0, I^{-1}(\theta)).$$

The basic density upon which this family is based (with $\theta_1 = 0$ and $\theta_2 = 1$, say), gives a tradeoff between Gaussian behavior near 0 and double-exponential (extreme value) in the tails: note that $\cosh(x) = (e^x + e^{-x})/2 \sim 1 + 2^{-1}x^2$ as $x \to 0$, while $\cosh(x) = (e^x + e^{-x})/2 \sim e^x/2$ as $x \to \infty$. Here is a plot of the basic density $p_{0,1}(x) = \exp(-\cosh(x) - \zeta(1))$:

5. (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 Statistics 581 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 $(\widehat{\alpha}, \widehat{\beta}) = (\widehat{\alpha}(\widehat{\beta}_{\text{profile}}), \widehat{\beta}_{\text{profile}})$.

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

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

(b) It follows that the score function for β corresponding to the profile log-likelihood is

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

and the observed information is

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

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,

$$\dot{\mathbf{l}}_{\beta}^{\#}(\underline{X}) = \dot{\mathbf{l}}_{n\alpha}|_{\hat{\alpha}(\beta)} \dot{\alpha}(\beta) + \dot{\mathbf{l}}_{n\beta}|_{\hat{\alpha}(\beta)} = \dot{\mathbf{l}}_{n\beta}|_{\hat{\alpha}(\beta)} \tag{1}$$

since

$$\dot{\mathbf{l}}_{n\alpha}|_{\hat{\alpha}(\beta)} = 0. \tag{2}$$

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 $\dot{\mathbf{l}}_{n\beta} = 0$.