**Statistics 582, Problem Set 1 Solutions**

Wellner; 1/12/2011

1. Lehmann and Casella, problem 3.15, page 502.

(a) A density $f$ is log-concave (or strongly unimodal) if $\varphi(x) = \log f(x)$ is concave. Show that if $f$ is log-concave, then it has a unique mode.

(b) Let $X_1, \ldots, X_n$ be iid with density $f(x-\theta)$. Show that the likelihood function (correction: likelihood **equation**) has a unique root if $f'(x)/f(x)$ is monotone, and the root is a maximum if $f'(x)/f(x)$ is decreasing. Hence densities that are log-concave yield unique MLE's.

(c) Let $X_1, \ldots, X_n$ be positive random variables (or symmetrically distributed about zero) with joint density $a^n \prod_{i=1}^n f(aX_i)$, $a > 0$. Show that the likelihood equation (correction: likelihood **function**) has a unique maximum if $xf'(x)/f(x)$ is strictly decreasing for $x > 0$.

(d) If $X_1, \ldots, X_n$ are i.i.d. with density $f(x-\theta)$ where $f$ is unimodal and if the likelihood equation has a unique root, show that the likelihood equation also has a unique root when the density of each $X_i$ is $af(a(x-\theta))$ with $a$ known.

**Solution:** Unfortunately these problem statements have a variety of difficulties. The following is an effort to give examples and counter-examples to clarify the situation and to provide additional conditions under which the claims do hold.

(a) Unfortunately, this is false as stated. In particular, every uniform density is log-concave: if $f(x) = (b-a)^{-1} 1_{[a,b]}(x)$ with $-\infty < a < b < \infty$, then $\varphi(x) \equiv \log f(x) = -\log(b-a)$ if $x \in [a,b]$, and $\varphi(x) = -\infty$ for $x \in [a,b]^c$; $\varphi$ is concave, so this $f$ is log-concave, but all the points $x \in [a,b]$ are modes of $f$. Here is a different example for which the density does not vanish: suppose

$$f(x) = c1_{[-a,a]}(x) + ce^{-(x-a)}1_{(a,\infty)}(x) + ce^{a+x}1_{(-\infty,-a)}(x)$$

where $c < 1/2$ and $2c(a+1) = 1$. Then

$$\varphi(x) \equiv \log f(x) = c1_{[-a,a]}(x) - (x-a)1_{(a,\infty)}(x) + (a+x)1_{(-\infty,-a)}(x) + \log c$$

is a concave function of $x$, so $f$ is log-concave, but every $x \in [-a,a]$ is a mode of $f$. See Figures 1 and 2. If we change the hypothesis to "$f$ is *strictly* log-concave" (i.e. that $\varphi(x) \equiv \log f(x)$ is a strictly concave function of $x$), then $f$ does have a unique mode. This follows easily by contradiction: if $x_0$ and $x_1$ are two distinct modes of $f$, then $f$ is constant on $[x_0, x_1]$ and hence $\varphi(x) \equiv \log f(x)$ is also constant on $[x_0, x_1]$. But this violates strict concavity of $\varphi$.

(b) This statement is also not quite accurate. If we start with $f$ as in the second example in (a) above, but with $\varphi$ defined (smoothed a bit at $\pm a$!) to be differentiable at $\pm a$, then $\varphi'(x) = f'(x)/f(x)$ is decreasing with $\varphi'(x) = 0$ for $x \in [-a/2, a/2]$. Thus for a sample of size $n = 1$ we have $l(\theta|X) = \log f(X-\theta)$ $\dot{l}(X;\theta) = -\varphi'(X-\theta) = 0$ if $|X-\theta| \le a/2$ or $X - a/2 \le \theta \le X + a/2$. Thus the likelihood equation does not have a unique root. If we strengthen the hypothesis to "$f'(x)/f(x)$ is *strictly*

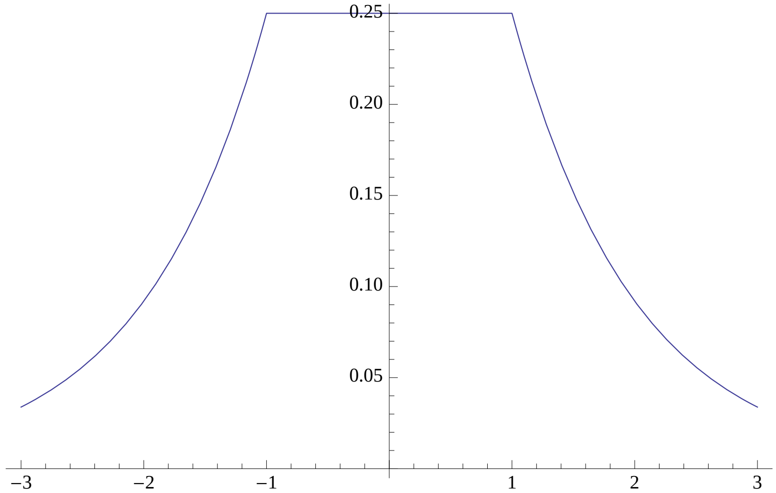


Figure 1: The density $f$ with $a = 1$, $c = 1/4$.

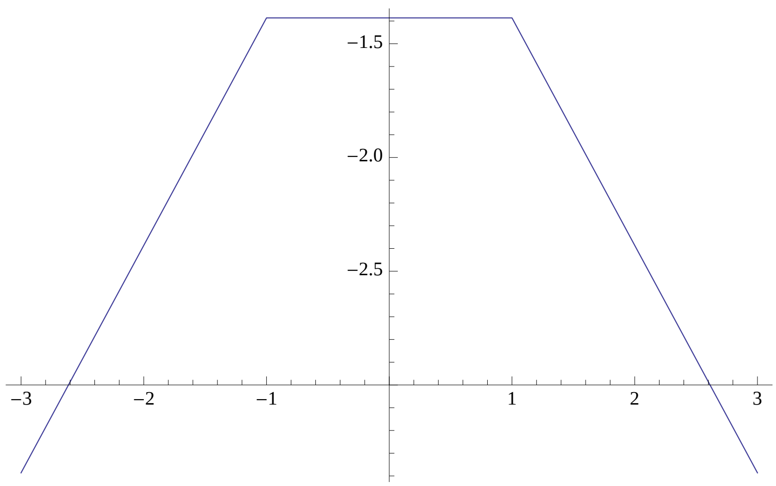


Figure 2: The function $\varphi = \log f$ with $a = 1$, $c = 1/4$.

*monotone*", then the log-likelihood function is

$$l_n(\theta|\underline{X}) = \sum_{i=1}^{n} \log f(X_i - \theta) = \sum_{i=1}^{n} \varphi(X_i - \theta),$$

so with $\varphi'(x) = f'(x)/f(x)$,

$$\dot{l}_n(\theta|\underline{X}) = -\sum_{i=1}^{n} \varphi'(X_i - \theta)$$

is a strictly decreasing function of $\theta$ and hence has at most one root. Since $1 = \int f(x-\theta)dx$, if $\int \varphi'(x)^2 dx < \infty$, then $0 = \int \varphi'(x-\theta) f(x-\theta)dx$ for all $\theta$ and hence $0 = \int \varphi'(y)f(y)dy$. This implies that $\lim_{y\to\infty} \varphi'(y) < 0$ and $\lim_{y\to-\infty} \varphi'(y) > 0$. It follows that a (unique) root of $\dot{l}_n(\theta|\underline{X}) = 0$ exists if $f'(x)/f(x) = \varphi'(x)$ is strictly decreasing.

(c) This part of the problem seems to be correctly stated (with "likelihood equation" replaced by "likelihood function" as noted above): if $X_1, \ldots, X_n$ are i.i.d. with density $af(ax)$, then the log-likelihood is given by

$$l_n(a|\underline{X}) = \sum_{i=1}^n \{\log a + \log f(aX_i)\} = \sum_{i=1}^n \{\log a + \varphi(aX_i)\}$$

so, with $\varphi'(x) = f'(x)/f(x)$,

$$\begin{aligned} \dot{l}_n(a) &= \sum_{i=1}^n \left\{ \frac{1}{a} + X_i \varphi'(aX_i) \right\} \\ &= \frac{1}{a} \sum_{i=1}^n \{1 + (aX_i)\varphi'(aX_i)\} \\ &= \frac{n}{a} \left\{ 1 - \frac{1}{n} \sum_{i=1}^n (-aX_i)\varphi'(aX_i) \right\} \\ &\equiv \frac{n}{a} \{1 - \Psi_n(a)\} = 0 \end{aligned}$$

if and only $\Psi_n(a) = 1$. Since $x\varphi'(x)$ is strictly decreasing, $a \mapsto \Psi_n(a)$ is strictly increasing, and hence there is at most one solution $\hat{a}$, and this corresponds to a unique maximum (if it exists). The argument for existence is similar to that in the location case: Since $1 = \int af(ax)dx$, it follows that $0 = \int\{f(ax) + ax\varphi'(ax)f(ax)\}dx$, or or equivalently $0 = \int\{1 + y\varphi'(y)\}f(y)dy$. Since $y\varphi'(y)$ is strictly decreasing, this implies that $\lim_{y\to\infty} y\varphi'(y) < 1$ and $\lim_{y\to-\infty} y\varphi'(y) > 1$. It follows that a (unique) solution of $\Psi_n(a) = 1$ exists.
(d) If the likelihood equation for the data with $X_i \sim f(x - \theta)$ has a solution, then we know there exists $\hat{\theta}_n = \hat{\theta}_n(\underline{X})$ such that

$$\dot{l}_n(\hat{\theta}_n) = \sum_{i=1}^n \{-\varphi'(X_i - \hat{\theta}_n) = 0 \tag{0.1}$$

where $\varphi'(x) \equiv f'(x)/f(x)$. Since the likelihood equation for the data with $X_i \sim af(a(x - \theta))$ with $a$ known is

$$\dot{l}_n(\beta; a) = \sum_{i=1}^n \{-\varphi'(a(X_i - \beta))a\} = \sum_{i=1}^n \{-\varphi'(Y_i - \gamma)a\} = 0.$$

where $Y_i \equiv aX_i$ and $\gamma = a\beta$. But this holds with and with $\hat{\gamma} = \hat{\theta}(\underline{Y})$; i.e.

$$\hat{\beta}_n = \frac{\hat{\gamma}}{a} = \frac{\hat{\theta}(a\underline{X})}{a}.$$

2. Problem 1, page 117, Ferguson, ACILST. 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 $\theta$. 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

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

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 $\Theta$ is no longer compact, and Wald's theorem does not apply directly. One way to remedy the problem is to compactify the set $\Theta$ by some appropriate identification of points in $\Theta$ 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 $\delta$ can be chosen to be arbitrarily small.

3. Consider the model introduced in Ferguson, ACILST, problem 17.2, page 117:

$$p(x|\theta) = 2\left(\frac{x}{\theta}1_{[0,\theta]}(x) + \frac{1-x}{1-\theta}1_{(\theta,1]}(x)\right), \quad \theta \in [0,1].$$

Show that Theorem 4.3, page 28, of the Chapter 4 notes (or Theorem 17, Ferguson, ACILST, page 114) applies to the MLE of $\theta$ in this model.

**Solution:** We proceed to verify the conditions (a) - (e) of Theorem 4.3 of the course notes. First (a) holds since $[0,1]$ is compact. (b) holds since for $0 < x < 1$ the function $\theta \mapsto p(x,\theta)$ is continuous (and hence upper-semicontinuous), while for $x = 0$ the function $\theta \mapsto p(0,\theta) = 1_{\{0\}}(\theta)$ is upper-semicontinuous, and similarly for $x = 1$ the function $\theta \mapsto p(1,\theta) = 1_{\{1\}}(\theta)$ is also upper-semicontinuous. To see that (c) holds, consider the two cases $\theta > \theta_0$ and $\theta \le \theta_0$. When $\theta > \theta_0$,

$$\begin{aligned}
f(x,\theta) &= \log p(x,\theta) - \log p(x,\theta_0) \\
&= \begin{cases} \log(\theta_0/\theta) & 0 \le x \le \theta_0 < \theta \\ \log(x/(1-x)) - \log(\theta/(1-\theta_0)) & \theta_0 < x \le \theta \\ \log((1-\theta_0)/(1-\theta)) & \theta < x \le 1 \end{cases} \\
&\le \begin{cases} \log(\theta_0/x) & 0 \le x \le \theta_0 < \theta \\ \log((1-\theta_0)/(1-x)) & \theta_0 < x \le \theta \\ \log((1-\theta_0)/(1-x)) & \theta < x \le 1 \end{cases} \\
&= 1_{[0,\theta_0]}(x)\log(\theta_0/x) + 1_{(\theta_0,1]}(x)\log((1-\theta_0)/(1-x)) \\
&\equiv F(x).
\end{aligned}$$

Here for the middle term we used

$$\begin{aligned}
\log x - \log(1-x) + \log(1-\theta_0) + \log(1/\theta) &\le \log x - \log(1-x) + \log(1-\theta_0) + \log(1/x) \\
&= \log((1-\theta_0)/(1-x))
\end{aligned}$$

since $1/\theta \le 1/x$ on the set $\theta_0 < x \le \theta$. Note that

$$E_{\theta_0}F(X) = 2\int_0^{\theta_0} \frac{x}{\theta_0}\log\left(\frac{\theta_0}{x}\right)dx + 2\int_{\theta_0}^1 \frac{1-x}{1-\theta_0}\log\left(\frac{1-\theta_0}{1-x}\right)dx < \infty.$$

Similarly, when $\theta \le \theta_0$,

$$\begin{aligned}
f(x,\theta) &= \log p(x,\theta) - \log p(x,\theta_0) \\
&= \begin{cases} \log(\theta_0/\theta) & 0 \le x \le \theta \le \theta_0 \\ \log((1-x)/x) - \log((1-\theta)/\theta_0) & \theta < x \le \theta_0 \\ \log((1-\theta_0)/(1-\theta)) & \theta_0 < x \le 1 \end{cases} \\
&\le \begin{cases} \log(\theta_0/x) & 0 \le x \le \theta_0 < \theta \\ \log((1-\theta_0)/(1-x)) & \theta_0 < x \le \theta \\ \log((1-\theta_0)/(1-x)) & \theta \le \theta_0 < x \le 1 \end{cases} \\
&= 1_{[0,\theta_0]}(x)\log(\theta_0/x) + 1_{(\theta_0,1]}(x)\log((1-\theta_0)/(1-x)) \\
&= F(x),
\end{aligned}$$

so the same envelope function works in this case. To verify (d) note that $\theta \mapsto p(x,\theta)$ is a continuous function of $\theta$ for $0 < x < 1$, and two indicator functions of $\theta$ when

$x \in \{0,1\}$, and hence the supremum involved in the condition is measurable. Finally, the identifiability condition (e) holds easily: note that

$$\begin{aligned}\rho(P_{\theta_0}, P_\theta) &= \int_0^1 \sqrt{p_{\theta_0}(x) p_\theta(x)} dx \\ &= \frac{\theta^2}{\sqrt{\theta\theta_0}} + \frac{2}{\sqrt{\theta_0(1-\theta)}} \int_\theta^{\theta_0} \sqrt{x(1-x)} dx + \frac{(1-\theta_0)^2}{\sqrt{(1-\theta_0)(1-\theta)}} \\ &= 1 \qquad \text{if and only if} \quad \theta = \theta_0 .\end{aligned}$$

Thus $p_{\theta_0}(x) = p_\theta(x)$ a.e. Lebesgue implies that $\theta = \theta_0$. Thus the conditions (a) - (e) all hold and we conclude that the MLE $\widehat{\theta}_n$ of $\theta$ is (almost surely) consistent: $\widehat{\theta}_n \rightarrow_{a.s.} \theta$.

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.] Consider the Weibull family of example 3.2.5 (581 Course Notes) $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ 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 $p_\theta$, we know that, for each fixed value of $\beta$ the value of $\alpha$ which maximizes the likelihood as a function of $\alpha$ 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 9.2 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 $\beta$ 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 $\beta$ the value of $\alpha$ 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 $\beta$ 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,

$$\dot{\mathbf{l}}^{\#}_\beta(\underline{X}) = \dot{\mathbf{l}}_{n\alpha}|_{\hat\alpha(\beta)}\dot\alpha(\beta) + \dot{\mathbf{l}}_{n\beta}|_{\hat\alpha(\beta)} = \dot{\mathbf{l}}_{n\beta}|_{\hat\alpha(\beta)} \tag{0.2}$$

since

$$\dot{\mathbf{l}}_{n\alpha}|_{\hat\alpha(\beta)} = 0. \tag{0.3}$$

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

5. Suppose that $X, X_1, \ldots, X_n$ are i.i.d. Weibull$(\alpha_0,\beta_0)$ (if $X$ has the Weibull$(\theta)$ 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 $\alpha$ 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 $\beta$. As a simpler alternative to maximum likelihood, I propose to use the alternative estimator $\overline{\beta}_n$ of $\beta$ obtained from the slope of an ordinary least squares fit of a Weibull Q-Q plot, and then estimate $\alpha$ by

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

(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 $\beta$ for every $x \geq 0$, and are bounded by

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

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 $\mathcal{F}$:

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

If $\overline{\beta}_n \to_{a.s.} \beta_0$, $\overline{\beta}_n \in [\beta_0 - \delta, \beta_0 + \delta]$, 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

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

(If instead $\overline{\beta}_n \to_p \beta_0$, 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 $(\overline{\alpha}^{\overline{\beta}_n}, \overline{\beta_n}) \to_{a.s.} (\alpha_0^{\beta_0}, \beta_0)$. Hence by the continuous mapping theorem

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

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

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

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 $\alpha$.