Statistics 582, Problem Set 2 Solutions

Wellner; 1/20/2010

1. 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 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

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

   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 direct computation. 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 discussed in class). 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.

2. (a) Suppose that $X, X_1, \ldots, X_n$ are i.i.d. with distribution P on $\mathbb{R}$ satisfying $E|X| < \infty$. Consider the example discussed in class on 1/8/2010: if V_n is defined by

$$V_n \equiv \frac{1}{n}\sum_{i=1}^n |X_i - \overline{X}_n| = \mathbb{P}_n|X - \overline{X}_n|,$$

show that

$$V_n \to_{a.s.} v \equiv E|X - \mu|$$

where $\mu = E(X)$.
(b) Now suppose we generalize the problem considered in (a) by considering $X_1, \ldots, X_n$ i.i.d. P on $\mathbb{R}^d$ and powers other than 1: let $\|\cdot\|$ be the usual Euclidean metric in $\mathbb{R}^d$, and consider

$$V_n(r) \equiv \frac{1}{n}\sum_{i=1}^n \|X_i - \overline{X}_n\|^r$$

for $1 \le r \le 2$ where $\overline{X}_n$ is the (multivariate) sample mean of the X_i's. Show that if $E\|X\|^r < \infty$, then $V_n(r) \to_{a.s.} v(r)$ where $v(r) \equiv E\|X_1 - \mu\|^r$ where $\mu = E(X) \in \mathbb{R}^d$?

Solution: (a) Let $\mathcal{F} = \{f_t(x) = |x - t|,\ |t - \mu| \le \delta\}$ where $f_t(x) \equiv |x - t|$. Then $|f_t(x)| = |x - \mu + \mu - t| \le |x - \mu| + \delta \equiv F(x)$ by the triangle inequality, and $EF(X) = E|X - \mu| + \delta < \infty$ since $E|X| < \infty$. Since $t \mapsto f_t(x)$ is continuous for every x and $T \equiv [\mu - \delta, \mu + \delta]$ is compact, it follows from Theorem 4.4.2 that

$$\sup_{t\in T} |\mathbb{P}_n f_t(X) - P f_t(X)| = \sup_{f\in\mathcal{F}} |\mathbb{P}_n(f) - P(f)| \to_{a.s.} 0.$$

Then, since $g(t) \equiv Pf_t = E|X - t|$ is continuous and $\overline{X}_n \to_{a.s.} \mu$, we have for $n \ge N_{\delta,\omega}$

$$\begin{aligned} |V_n - g(\mu)| &= |\mathbb{P}_n f_{\overline{X}_n}(X) - Pf_\mu(X)| \\ &\le |\mathbb{P}_n f_{\overline{X}_n}(X) - Pf_{\overline{X}_n}(X)| + |g(\overline{X}_n) - g(\mu)| \\ &\le \sup_{t\in[\mu-\delta,\mu+\delta]} |\mathbb{P}_n f_t(X) - Pf_t(X)| + |g(\overline{X}_n) - g(\mu)| \\ &\to_{a.s.} 0 + 0 = 0. \end{aligned}$$

(b) Essentially the same argument works for $V_n(r) = n^{-1}\sum_{i=1}^n \|X_i - \overline{X}_n\|^r$ when the X_i's are i.i.d. in $\mathbb{R}^d$ with $E\|X\|^r < \infty$: for $\delta > 0$ let $\mathcal{F} \equiv \{f_t(x) : \ \|t - \mu\| \le \delta\}$. Then with $c_r = 2^{r-1}$,

$$|f_t(x)| = \|x - \mu + \mu - t\|^r \le c_r\{\|x-\mu\|^r + \|\mu - t\|^r\} \le c_r\{\|x-\mu\|^r + \delta^r\} \equiv F(x)$$

by the triangle and c_r-inequalities. Since $E\|X\|^r < \infty$, it follows that

$$EF(X) = c_r\{E\|X - \mu\|^r + \delta^r\} < \infty.$$

Since $t \mapsto f_t(x)$ is continuous for every x and $T \equiv \{t : \ \|t-\mu\| \le \delta\}$ is compact, it follows from Theorem 4.4.2 that

$$\sup_{t\in T} |\mathbb{P}_n f_t(X) - Pf_t(X)| = \sup_{f\in\mathcal{F}} |\mathbb{P}_n(f) - P(f)| \to_{a.s.} 0.$$

Then, since $g(t) \equiv Pf_t = E\|X - t\|^r$ is continuous and $\overline{X}_n \to_{a.s.} \mu$, we have for $n \ge N_{\delta,\omega}$

$$\begin{aligned} |V_n(r) - g(\mu)| &= |\mathbb{P}_n f_{\overline{X}_n}(X) - Pf_\mu(X)| \\ &\le |\mathbb{P}_n f_{\overline{X}_n}(X) - Pf_{\overline{X}_n}(X)| + |g(\overline{X}_n) - g(\mu)| \\ &\le \sup_{\{t:\|t-\mu\|\le\delta\}} |\mathbb{P}_n f_t(X) - Pf_t(X)| + |g(\overline{X}_n) - g(\mu)| \\ &\to_{a.s.} 0 + 0 = 0. \end{aligned}$$

3. (a) In connection with Problem 1.3 (Ferguson, ACILST, page 117, problem 2, with parameter space $\Theta = [0,1]$), does Wald's Theorem 4.3, page 28, Chapter 4 notes apply?
Take a continuous version of the density in this problem; i.e.

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

(c) Do our hypotheses A0-A2 (of Section 4.1) hold in this example?
(d) Do our hypotheses A3 and A4 (of Section 4.1) hold in this example?

(e) Does there exist an estimator $\bar{\theta}_n$ of θ which is $n^{1/2}-$consistent?

Solution: (a) **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 θ for $0<x<1$, and two indicator functions of θ 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 θ is (almost surely) consistent: $\widehat{\theta}_n \rightarrow_{a.s.} \theta$.
(b) Deleted.
(c) A0 - A2 all hold in this example: If $\theta \neq \theta^*$, then $p_\theta \neq p_{\theta^*}$ and hence $P_\theta \neq P_{\theta^*}$. The set $A = \{x : \ p_\theta(x) > 0\} = (0,1)$ for all θ, and hence does not depend on θ; thus A1 holds. A2 holds with μ given by Lebesgue measure on $[0,1]$.
(c) Suppose that $\theta_0 < \theta$. Then the Kullback-Leibler information $K(P_{\theta_0}, P_\theta)$ is given by

$$
\begin{aligned}
K(P_{\theta_0}, P_\theta) &= \int_0^{\theta_0} p_{\theta_0}(x) \log(\theta/\theta_0) dx + \int_{\theta_0}^{\theta} p_{\theta_0}(x) \log\left(\frac{1-x}{1-\theta_0} \frac{\theta}{x} \right) dx \\
&\quad + \int_\theta^1 p_{\theta_0}(x) \log \frac{1-\theta}{1-\theta_0} dx \\
&= \theta_0 \log(\theta/\theta_0) + \frac{(1-\theta)^2}{1-\theta_0} \log \frac{1-\theta}{1-\theta_0} \\
&\quad + \frac{1}{1-\theta_0} \{(1-\theta_0)^2 - (1-\theta)^2\} \log\left(\frac{\theta}{1-\theta_0} \right) \\
&\quad + \frac{2}{1-\theta_0} \int_{\theta_0}^{\theta} (1-x) \log\left(\frac{1-x}{x} \right) dx.
\end{aligned}
$$

Similarly, if $\theta_0 > \theta$, then

$$
\begin{aligned}
K(P_{\theta_0}, P_\theta) &= \int_0^{\theta} p_{\theta_0}(x) \log(\theta/\theta_0) dx + \int_{\theta}^{\theta_0} p_{\theta_0}(x) \log\left(\frac{x}{\theta_0} \frac{1-\theta}{1-x} \right) dx \\
&\quad + \int_{\theta_0}^1 p_{\theta_0}(x) \log \frac{1-\theta}{1-\theta_0} dx \\
&= \frac{\theta^2}{\theta_0} \log(\theta/\theta_0) + (1-\theta_0) \log \frac{1-\theta}{1-\theta_0} \\
&\quad + \frac{1}{\theta_0} \{\theta_0^2 - \theta^2\} \log\left(\frac{1-\theta}{\theta_0} \right) + \frac{2}{\theta_0} \int_\theta^{\theta_0} x \log\left(\frac{x}{1-x} \right) dx.
\end{aligned}
$$

See the last page for a plot of $\theta \mapsto K(P_{\theta_0}, P_\theta)$ for $\theta_0 = .2$.

(d) Since p_θ is given by (1),

$$\log p_\theta(x) = \begin{cases} \log 2 + \log x - \log\theta, & \text{if } x \le \theta, \\ \log 2 + \log(1-x) - \log(1-\theta), & \text{if } x > \theta, \end{cases}$$

so

$$\dot{\mathbf{l}}_\theta(x) = -\frac{1}{\theta} 1_{[x<\theta]} + \frac{1}{1-\theta} 1_{[x>\theta]},$$

but the derivative does not exist at $\theta = x$ (since the left and right derivatives are different). Similarly, naive differentiation yields

$$\ddot{\mathbf{l}}_{\theta\theta}(x) = \frac{1}{\theta^2} 1_{[x<\theta]} + \frac{1}{(1-\theta)^2} 1_{[x>\theta]},$$

but the second derivative does not exist at $\theta = x$. Note that $\dot{\mathbf{l}}_\theta$ is a discontinuous function of θ for every $0 < x < 1$. Although

$$E_\theta \dot{\mathbf{l}}_\theta(X) = -\frac{1}{\theta}\theta + \frac{1}{1-\theta}(1-\theta) = 0,$$

and

$$E_\theta \dot{\mathbf{l}}_\theta^2(X) = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)},$$

we also have

$$-E_\theta \ddot{\mathbf{l}}_{\theta\theta}(X) = -\frac{1}{\theta} - \frac{1}{1-\theta} = -\frac{1}{\theta(1-\theta)} \neq E_\theta \dot{\mathbf{l}}_\theta^2(X).$$

Thus A3 and A4(iii) fail, while A4(i) and A4(ii) hold. Remarkably, the desired information identity does hold if we take second derivatives in a generalized (or measure) sense: the contribution to $\ddot{\mathbf{l}}_{\theta\theta}$ at $\theta = x$ is the Dirac measure at $\theta = x$ with mass the size of the jump in $\theta \mapsto \dot{\mathbf{l}}_\theta$ at $\theta = x$, namely $(1-x)^{-1} - (-1/x) = 1/(x(1-x))$. Thus in this extended sense

$$\ddot{\mathbf{l}}_{\theta\theta}(x) = \frac{1}{\theta} 1_{[x<\theta]} + \frac{1}{(1-\theta)^2} 1_{[x>\theta]} + \frac{1}{x(1-x)} \delta_x(\theta).$$

Since $P_\theta(X < \theta) = \theta = 1 - P_\theta(X > \theta)$ and the point mass $\delta_x(\theta)$ carries a negative sign when viewed as a function of x, this yields

$$\begin{aligned} -E_\theta \ddot{\mathbf{l}}_{\theta\theta}(X) &= -\left(\frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2}\right) + \frac{1}{\theta(1-\theta)} p_\theta(\theta) = -\frac{1}{\theta(1-\theta)} + \frac{2}{\theta(1-\theta)} \\ &= \frac{1}{\theta(1-\theta)} = E_\theta \dot{\mathbf{l}}_\theta^2(X). \end{aligned}$$

(e) First a $\sqrt{n}-$consistent estimator of θ via moments: note that

$$\begin{aligned}
E_\theta X &= 2\int_0^\theta \frac{x^2}{\theta}dx + 2\int_\theta^1 \frac{x(1-x)}{1-\theta}dx \\
&= \frac{2}{3}\theta^2 + \frac{2}{1-\theta}(\frac{1}{2}x^2 - \frac{1}{3}x^3|_\theta^1 \\
&= \frac{2}{3}\theta^2 + \frac{2}{1-\theta}\left\{\frac{1}{6} - \frac{1}{2}\theta^2 + \frac{1}{3}\theta^3\right\} \\
&= \frac{2}{1-\theta}\left\{\frac{1}{3}\theta^2(1-\theta) + \frac{1}{6} - \frac{1}{2}\theta^2 + \frac{1}{3}\theta^3\right\} \\
&= \frac{2}{1-\theta}\left\{\frac{1}{6} - \frac{1}{6}\theta^2\right\} = \frac{1}{3}(1+\theta).
\end{aligned}$$

Since $\overline{X}_n \rightarrow_p E_\theta X = (1+\theta)/3$, it follows by continuous mapping that $3\overline{X}_n - 1 \rightarrow_p \theta$. Thus with $g(x) \equiv 3x - 1$ we have

$$\sqrt{n}(g(\overline{X}_n) - \theta) \rightarrow_d g'(\theta)\sigma(\theta)Z$$

where $g'(x) = 3$, $\sigma^2(\theta) = Var_\theta(X) = (1 - \theta + \theta^2)/18$, and $Z \sim N(0,1)$. Thus it follows that

$$\sqrt{n}(3\overline{X}_n - 1 - \theta) \rightarrow_d N(0, (1-\theta+\theta^2)/2).$$

Thus the estimator $\overline{\theta}_n \equiv 3\overline{X}_n - 1$ is a $\sqrt{n}-$consistent estimator of θ.

Now for an estimator of θ based on the median. The distribution function F_θ corresponding to p_θ is

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

and the corresponding quantile function is

$$F_\theta^{-1}(u) = \sqrt{\theta u}1_{[u<\theta]} + (1 - \sqrt{(1-\theta)(1-u)})1_{[u\ge\theta]}.$$

Thus the median is

$$F_\theta^{-1}(1/2) = \sqrt{\theta/2}1_{[1/2<\theta]} + (1 - \sqrt{(1-\theta)/2})1_{[1/2>\theta]} \equiv g(\theta),$$

which has inverse function

$$g^{-1}(x) = 2x^2 1_{[x\ge 1/2]} + (1 - 2(1-x)^2)1_{[x<1/2]} \equiv h(x)$$

Note that $g^{-1}(1/2+) = g^{-1}(1/2-) = 1/2$, so g^{-1} is continuous at $1/2$, and

$$\frac{d}{dx}g^{-1}(x) = \frac{d}{dx}h(x) = 4x1_{[x\ge1/2]} + 4(1-x)1_{[x<1/2]},$$

so the derivative of g^{-1} is also continuous at $x = 1/2$. It follows that $g^{-1}(\mathbb{F}_n^{-1}(1/2)) = h(\mathbb{F}_n^{-1})$ is a consistent and asymptotically normal estimator of θ:

$$g^{-1}(\mathbb{F}_n^{-1}(1/2)) \rightarrow_{a.s.} g^{-1}(F_\theta^{-1}(1/2)) = g^{-1}(g(\theta)) = \theta,$$

and

$$\begin{aligned}\sqrt{n}(g^{-1}(\mathbb{F}_n^{-1}(1/2)) - g^{-1}(F_\theta^{-1}(1/2))) &= \sqrt{n}(h(\mathbb{F}_n^{-1}(1/2)) - h(F_\theta^{-1}(1/2))) \\ &\rightarrow_d h'(F_\theta^{-1})\{-Q'(1/2)\mathbb{U}(1/2)\} \sim N(0, \sigma^2(\theta))\end{aligned}$$

where

$$\sigma^2(\theta) = \{h'(F_\theta^{-1}(1/2))\}^2 \cdot Q'(1/2)^2 \cdot (1/4).$$

There are many other $\sqrt{n}-$consistent estimators of θ in this example and, in fact, the MLE is consistent, $\sqrt{n}-$consistent, and asymptotically efficient. We will return to this example in Stat 582.

4. 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 $\overline{\beta}_n$ of β obtained from the slope of an ordinary least squares fit of a Weibull Q-Q plot, and then estimate α by

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

(a) Suppose that $\overline{\beta}_n \rightarrow_p \beta_0$ is known. Show that $\overline{\alpha}_n \rightarrow_p \alpha_0$. [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 $l_n(\alpha, \overline{\beta}_n)$.
(c) What hypotheses are needed if we assume $X_1, \ldots, X_n$ are i.i.d. P_0 with P_0 not necessarily Weibull? You may continue to assume that $\overline{\beta}_n \rightarrow_p \beta \equiv \beta(P_0)$.

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 $\mathcal{F}$:

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

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

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

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.

5. 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 $\hat{\theta}_n \to_{a.s.} 1$ no matter what θ_0 is true if $\delta(\theta) \to 0$ "fast enough".
(a) Show that $\hat{\theta}_n \to_{a.s.} 1$ continues to hold if

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

with $c > 2$. Note that with this definition of $\delta(\theta)$, δ is continuous at $\theta = 0$. Ferguson shows that $c = 4$ works.
(b) Show that when $c = 2$, Ferguson's argument yields

$$\sup_{0\le\theta\le1} n^{-1}\log L_n(\theta) \ge \frac{n-1}{n}\log(M_n/2) + \frac{1}{n}\log\frac{1-M_n}{\delta(M_n)} \to_d D$$

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

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

if $n(1-M_n)^c \to_{a.s.} 0$. Also note that for $0 < \theta \le 1$ and $x > 1-\theta$ we have

$$P_\theta(X > x) \ge P_0(X > x) \quad \text{or, equivalently} \quad P_\theta(X \le x) \le P_0(X \le x).$$

To see this note that $P_0(X > x) = (1-x)^2/2$, while

$$\begin{aligned} P_\theta(X > x) &= \theta(1-x)/2 + \int_x^1 \frac{1-\theta}{\delta(\theta)}\left(1 - \frac{|y-\theta|}{\delta(\theta)}\right)1_{[|y-\theta|\le\delta(\theta)]}dy \\ &\ge \theta(1-x)/2 \ge (1-x)^2/2 = P_0(X > x) \end{aligned}$$

if $x > 1-\theta$ and $0 < \theta \le 1$. But then we have

$$\begin{aligned} P_\theta(n(1-M_n)^c \ge \epsilon) &= P_\theta(X_1 \le 1-(\epsilon/n)^{1/c})^n \\ &\le P_0(X_1 \le 1-(\epsilon/n)^{1/c})^n \qquad \text{for } \ n \ge N(\epsilon, c, \theta) \\ &= (1 - P_0(X_1 > 1-(\epsilon/n)^{1/c})^n \\ &= \left(1 - \frac{1}{2}(\epsilon/n)^{2/c}\right)^n \le \exp(-(1/2)\epsilon^{2/c}n^{1-2/c}) \end{aligned}$$

which has a finite sum on n if $c > 2$. (Note that $\sum_{n \geq n_0+1} \exp(-An^r) < \infty$ for $A > 0$ and $r > 0$ since

$$
\begin{aligned}
\sum_{n \geq n_0+1} \exp(-An^r) &\leq \sum_{n \geq n_0+1} \int_{n-1}^{n} \exp(-Ay^r) dy = \int_{n_0}^{\infty} \exp(-Ay^r) dy \\
&= \int_{n_0^r}^{\infty} r^{-1} x^{(1/r)-1} \exp(-Ax) dx \quad \text{by the change of variables } \ x = y^r \\
&= \int_{An_0^r}^{\infty} \frac{1}{rA^{1/r}} z^{(1/r)-1} \exp(-z) dz \leq \frac{1}{rA^{1/r}} \Gamma(1/r) < \infty.)
\end{aligned}
$$

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

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

or $(1/2)n(1-M_n)^2 \rightarrow_d E \sim \text{Exponential}(1)$ under P_0. Therefore, again under P_θ, by a similar argument,

$$
\begin{aligned}
&P_\theta(\sup_{0 \leq \theta' \leq 1} n^{-1} \log L_n(\theta') \leq t) \\
&\leq \ P_\theta\left(\frac{n-1}{n}\log(M_n/2) + \frac{1}{n(1-M_n)^2} - \frac{1}{n} \leq t\right) \\
&= \ P_\theta\left(n(1-M_n)^2 \geq \frac{1}{t + \frac{1}{n} - \frac{n-1}{n}\log(M_n/2)}\right) \\
&\leq \ P_0\left(n(1-M_n)^2 \geq \frac{1}{t + \frac{1}{n} - \frac{n-1}{n}\log(M_n/2)}\right) \\
&\rightarrow \ P\left(2E \geq \frac{1}{2(t + \log 2)}\right) \\
&= \ \exp\left(-\frac{1}{2(t+\log 2)}\right) = \exp\left(-\frac{1}{2(t - \log(1/2))}\right), \quad t > \log(1/2).
\end{aligned}
$$

It seems that my statement of this is off by a minus sign (or, equivalently, $\log 2$ should be $\log(1/2)$).

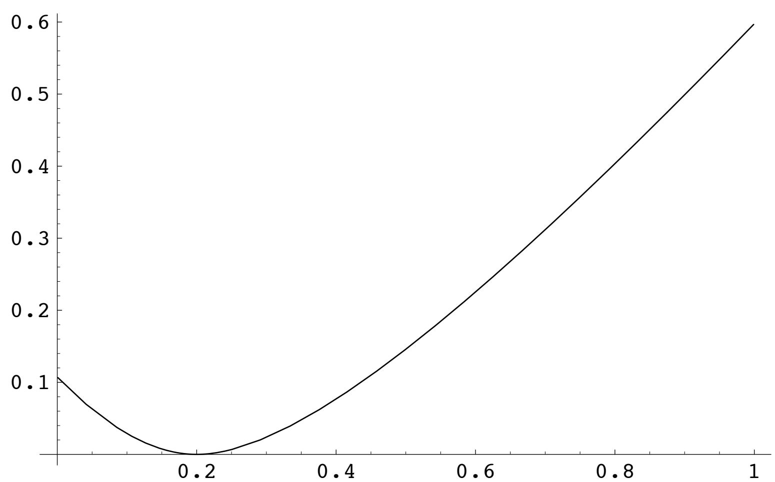

Figure 1: Kullback - Leibler function $K(P_{\theta_0}, P_\theta)$, $\theta_0 = .2$