Statistics 581, Problem Set 9 Solution

Wellner; 11/26/2008

1. (a) Exercise 2.1.6, page 10, chapter 2 notes; i.e. show that $d_{TV}(P,Q) = 1 - \int p \wedge q \, d\mu$.
(b) Exercise 2.1.7, page 10, chapter 2 notes; i.e. show that

$$H^2(P,Q) \le d_{TV}(P,Q) \le H(P,Q)\{1 + \rho(P,Q)\}^{1/2} \le \sqrt{2} H(P,Q).$$

Solution: (a) From the proof of proposition 1.13, chapter 2 notes, page 9, we see that

$$\begin{aligned} d_{TV}(P,Q) &= \frac{1}{2}\int |p-q| d\mu = \int_{[p\ge q]} (p-q)\, d\mu = \int_{[p\ge q]} p d\mu - \int_{[p \ge q]} p \wedge q \, d\mu \\ &= \int_{[p\ge q]} p d\mu + \int_{[p<q]} p d\mu - \int_{[p\ge q]} p\wedge q\, d\mu - \int_{[p<q]} p\, d\mu \\ &= \int p d\mu - \int_{[p\ge q]} p \wedge q \, d\mu - \int_{[p<q]} p\wedge q\, d\mu \\ &= 1 - \int p \wedge q \, d\mu \equiv 1 - \eta(P,Q). \end{aligned}$$

Alternatively, use the identity $|a-b| = a + b - 2(a\wedge b)$ for all $a, b \in \mathbb{R}$ to deduce that

$$|p(x) - q(x)| = p(x) + q(x) - 2p(x) \wedge q(x)$$

for each fixed x, and hence

$$\begin{aligned} d_{TV}(P,Q) &= \frac{1}{2}\int |p-q| d\mu = \frac{1}{2}\left(\int p d\mu + \int q d\mu - 2\int p \wedge q d\mu \right) \\ &= 1 - \int p \wedge q d\mu \equiv 1 - \eta(P,Q). \end{aligned}$$

(b) To see the first inequality, note that $H^2(P,Q) = 1 - \rho(P,Q)$ where

$$\rho(P,Q) = \int \sqrt{pq}\, d\mu \ge \int p \wedge q \, d\mu \equiv \eta(P,Q)$$

since $\sqrt{p(x)q(x)} \ge p(x) \wedge q(x)$ for all x. Thus we have

$$H^2(P,Q) \;=\; 1 - \rho(P,Q) \le 1 - \eta(P,Q) = d_T V(P,Q).$$

For the second inequality, write $|p-q| = |(\sqrt{p} - \sqrt{q})(\sqrt{p} + \sqrt{q})|$ and then apply the Cauchy-Schwarz inequality: thus

$$2 d_{TV}(P,Q) \;=\; \int |p-q| d\mu = \int |(\sqrt{p} - \sqrt{q})(\sqrt{p}+\sqrt{q})| d\mu$$

$$
\begin{aligned}
&\le \left(\int |\sqrt{p}-\sqrt{q}|^2 d\mu\right)^{1/2}\left(\int |\sqrt{p}+\sqrt{q}|^2 d\mu\right)^{1/2} \\
&= \sqrt{2}H(P,Q)\left\{\int (p+2\sqrt{pq}+q)d\mu\right\}^{1/2} \\
&= \sqrt{2}H(P,Q)\left\{2+2\rho(P,Q)\right\}^{1/2} \\
&= 2H(P,Q)\left\{1+\rho(P,Q)\right\}^{1/2},
\end{aligned}
$$

and this yields the claimed inequality. The third inequality is easy since $\rho(P,Q) \le 1$ by Cauchy-Schwarz again.

2. (a) Lehmann and Casella, problem 6.3.1, page 501.
   (b) Lehmann and Casella, problem 6.3.2, page 501.
   (c) Lehmann and Casella, problem 6.3.4, page 501.

   Solution: (a)(i) Since $\log P_p(X=x) = x\log p + (n-x)\log(1-p)$, we have $l(p|X) = X\log p + (n-X)\log(1-p)$; differentiating this with respect to p yields

$$\dot{l}(p|X) = \frac{X}{p} - \frac{n-X}{1-p} = \frac{X(1-p)-(n-X)p}{p(1-p)}$$

and this equals 0 if $p = \hat{p} \equiv X/n$. Since the second derivative is

$$\ddot{l}(p|X) = -\frac{X}{p^2} - \frac{n-X}{(1-p)^2} < 0$$

it follows that $\hat{p} = X/n$ is the MLE of $p \in [0,1]$.
(a)(ii) Since $(\prod_{i=1}^n y_i)^{1/n} \le n^{-1}(y_1+\cdots+y_n)$ for any numbers $y_i \ge 0$, it follows, with $y_i \equiv np/X$ for $i=1,\ldots,X$, and $y_i \equiv nq/(n-X)$, $i = X+1,\ldots,n$, that

$$\left\{\left(\frac{np}{X}\right)^X\left(\frac{nq}{n-X}\right)^{n-X}\right\}^{1/n} \le n^{-1}\left\{X\frac{np}{X}+(n-X)\frac{nq}{n-X}\right\} = 1\,,$$

or, equivalently,

$$p^X(1-p)^{n-X} \le \left(\frac{X}{n}\right)^X\left(\frac{n-X}{n}\right)^{n-X}\,,$$

with equality if and only if $p = X/n \equiv \hat{p}$. Thus $\hat{p} = X/n$ is the MLE of $p \in [0,1]$.
(b) When the closed interval $[0,1]$ is replaced by the open interval $(0,1)$, then the MLE exists if $0 < X < n$ and is $\hat{p} = X/n \in (0,1)$ in this case. If $X = 0$, then the log-likelihood equals $n\log(1-p)$, so $\sup_{p\in(0,1)} l(p) = 0$, but this supremum is not achieved (in the set $(0,1)$). Thus the MLE does not exist in this case. Similarly, if $X = n$, the the log-likelihood equals $n\log p$, so $\sup_{p\in(0,1)} l(p) = 0$, but this supremum is not achieved (in the set $(0,1)$).
(c) The log-likelihood is

$$
\begin{aligned}
l_n(\theta) &= -\frac{1}{2}\sum_{i=1}^n (X_i-\xi)^2 - (n/2)\log(2\pi) \\
&= n\xi\overline{X}_n - n\xi^2/2 - (n/2)\log(2\pi) \\
&= -(n/2)(\xi-\overline{X}_n)^2 + (n/2)\overline{X}_n^2 - (n/2)\log(2\pi).
\end{aligned}
$$

When $\overline{X}_n > 0$ this is maximized by $\xi = \hat{\xi} = \overline{X}_n$.
When $\overline{X}_n \le 0$ this is maximized over $\xi \ge 0$ by $\xi = \hat{\xi} = 0$. Since the maximization in the problem as stated is over the open set $\xi > 0$, the supremum is not attained.

3. Lehmann and Casella, problem 6.3.18, page 502. [**Note:** It seems to me that 3.15(b) should be 3.15(c) since $C(0, a)$ is a *scale family*).]

(d) First consider problem 3.15(c): Here the log-likelihood for the scale family for a density f is

$$l_n(a) = n \log a + \sum_{i=1}^n \log f(aX_i), \qquad a > 0 .$$

Hence the score equation is

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

where $g(x) \equiv -xf'(x)/f(x)$. If $xf'(x)/f(x)$ is strictly decreasing in x, then $g(x)$ is strictly increasing in x, and hence $h_n(a)$ is strictly increasing in a. Hence there is at most one value of a satisfying $h_n(a) = n$. If a solution exists, it is unique.
For the particular case of a Cauchy density f, it is easy to compute

$$g(x) = \frac{2x^2}{1 + x^2}$$

which is strictly increasing (from 0 at $x = 0$ to 2 at $x = \infty$). Moreover, in this case the likelihood equation becomes

$$h_n(a) = \sum_{i=1}^n \frac{2a^2 X_i^2}{1 + a^2 X_i^2} = n .$$

But the left side converges to 0 as $a \downarrow 0$, and converges to $2n$ as $a \uparrow \infty$. Since it is monotone increasing and continuous, there is a unique solution $\hat{a}_n$. All the hypotheses of Theorem 1.5 hold in this case, and

$$I_{scale}(f) = \int_{-\infty}^{\infty} \left\{ 1 + x\frac{f'(x)}{f(x)} \right\}^2 f(x)dx = \frac{1}{2} .$$

Thus it follows that

$$\sqrt{n}(\hat{a}_n - a) \to_d N(0, 2a^2) .$$

4. Suppose that $(Y|Z) \sim \text{Poisson}(\lambda e^{\gamma Z})$, and $Z \sim Bernoulli(\eta)$, and $\theta = (\lambda, \gamma, \eta)$. Let $X = (Y, Z)$, and suppose that we observe $X_1, \ldots, X_n$ i.i.d. as X.
(a) Find the score equations for estimation of θ.
(b) Give conditions on the data $X_1, \ldots, X_n = (Y_1, Z_1), \ldots, (Y_n, Z_n)$ guaranteeing that the score equations have a unique solution which maximizes the likelihood. Call the resulting estimators $\widehat{\theta}_n = (\widehat{\lambda}_n, \widehat{\gamma}_n, \widehat{\eta}_n)$.

(c) What does theorem 4.1.2 (Chapter 4, page 5), say about the asymptotic distribution of $\sqrt{n}(\widehat{\theta} - \theta_0)$ when the distribution of the data is given by P_{θ_0}?
(d) Suppose that $\theta_1 \neq \theta_0$ is the "true" value of the parameter θ, and we consider the likelihood ratio $L_n(\theta_1)/L_n(\theta_0)$ where $L_n(\theta) \equiv \prod_{i=1}^n p_\theta(X_i)$. Show that $n^{-1}\log(L_n(\theta_1)/L_n(\theta_0) \rightarrow_p$ some constant, and identify the constant explicitly in terms of θ_1, θ_0.

Solution: (a) The density of $X = (Y, Z)$ is

$$p_\theta(y,z) = f_{\lambda,\gamma}(y|z)g_\eta(z) = e^{-\lambda e^{\gamma z}} \frac{(\lambda e^{\gamma z})^y}{y!} \eta^z (1-\eta)^{1-z},$$

for $y \in \{0, 1, \ldots\}$, $z \in \{0, 1\}$, and hence

$$\log p_\theta(y,z) = y\log(\lambda e^{\gamma z}) - \lambda e^{\gamma z} - \log y! + z\log\eta + (1-z)\log(1-\eta).$$

From this we calculate the scores $\dot{l}_\lambda$, $\dot{l}_\gamma$, and $\dot{l}_\eta$:

$$\begin{aligned}
\dot{l}_\lambda(y,z) &= \frac{y}{\lambda} - e^{\gamma z} = \frac{1}{\lambda}(y - \lambda e^{\gamma z}), \\
\dot{l}_\gamma(y,z) &= yz - \lambda e^{\gamma z} z = z(y - \lambda e^{\gamma z}), \\
\dot{l}_\eta(y,z) &= \frac{z}{\eta} - \frac{1-z}{1-\eta}.
\end{aligned}$$

Thus the score equations for $\theta = (\lambda, \gamma, \eta)$ are

$$\begin{aligned}
0 &= \sum_{i=1}^n \dot{l}_\lambda(Y_i, Z_i) = \frac{1}{\lambda}\sum_{i=1}^n (Y_i - \lambda e^{\gamma Z_i}), \\
0 &= \sum_{i=1}^n \dot{l}_\gamma(Y_i, Z_i) = \sum_{i=1}^n Z_i(Y_i - \lambda e^{\gamma Z_i}), \\
0 &= \sum_{i=1}^n \dot{l}_\eta(Y_i, Z_i) = \sum_{i=1}^n \left\{\frac{Z_i}{\eta} - \frac{1-Z_i}{1-\eta}\right\} = (\eta(1-\eta))^{-1}\sum_{i=1}^n (Z_i - \eta).
\end{aligned}$$

(b) The third equation always has the unique solution $\hat{\eta} = \overline{Z}_n$. It is clear that the score equation for λ has only the solution $\lambda = 0$ if all $Y_i = 0$, and in this case there is clearly no unique solution for γ. So as least one Y_i must be non-zero in order to get a unique solution. If all the Z_i's are equal to 1, then the two equations agree, and it is clear that all we can estimate is the product λe^γ. Similarly, if all the Z_i's are equal to 0, then the score equation for γ is trivially satisfied (for any γ), and the score equation for λ gives just an estimator of λ. For the first and second equations, we compute

$$\begin{aligned}
\ddot{l}_{n,\lambda\lambda} &= -\frac{1}{\lambda^2}\sum_{i=1}^n Y_i, \\
\ddot{l}_{n,\lambda\gamma} &= -\sum_{i=1}^n Z_i e^{\gamma Z_i} = \ddot{l}_{n,\gamma\lambda}, \\
\ddot{l}_{n,\gamma\gamma} &= -\lambda\sum_{i=1}^n Z_i^2 e^{\gamma Z_i}.
\end{aligned}$$

Thus the matrix of second partial derivatives (the Hessian) fails to be negative definite if $\sum_1^n Y_i = 0$ or if

$$\left(\lambda \sum_{i=1}^n Z_i^2 e^{\gamma Z_i}\right)\left(\lambda^{-2}\sum_{i=1}^n Y_i\right) \le \left(\sum_{i=1}^n Z_i e^{\gamma Z_i}\right)^2 .$$

Because the Z_i's are Bernoulli,

$$\sum_{i=1}^n Z_i^2 e^{\gamma Z_i} = \sum_{i=1}^n Z_i e^{\gamma Z_i} = e^{\gamma}\sum_{i=1}^n Z_i .$$

Thus the above inequality holds if and only if

$$\sum_{i=1}^n Y_i \le \lambda e^{\gamma}\sum_{i=1}^n Z_i . \tag{0.1}$$

(Note that upon division by n the right side converges in probability to $\eta_0 \lambda e^{\gamma}$ while the left side (divided by n) converges to $\eta_0\lambda_0 e^{\gamma_0} + \lambda_0(1-\eta_0)$.) At the MLE, the inequality (0.1) holds if and only if

$$\sum_{i=1}^n Y_i \le \sum_{i=1}^n Y_i Z_i .$$

This inequality can occur only if $Z_i = 1$ whenever $Y_i > 0$. Thus it becomes clear that the equations will have a unique solution yielding a maximum of the likelihood if at least one $Z_i = 0$ and $Y_i \ge 1$.
(c) Theorem 4.1.5 says that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d N_3(0, I(\theta_0)^{-1})$$

where

$$I(\theta) = \begin{pmatrix} \lambda^{-1}E(e^{\gamma Z}) & E(Ze^{\gamma Z}) & 0 \\ E(Ze^{\gamma Z}) & \lambda E(Z^2 e^{\gamma Z}) & 0 \\ 0 & 0 & (\eta(1-\eta))^{-1} \end{pmatrix} .$$

(d) When θ_1 is true,

$$\begin{aligned} n^{-1}\log\frac{L_n(\theta_1)}{L_n(\theta_0)} &= n^{-1}\sum_{i=1}^n \log\frac{p_{\theta_1}}{p_{\theta_0}}(X_i) \\ &\to_p E_{\theta_1}\log\frac{p_{\theta_1}}{p_{\theta_0}}(X_1) \\ &= K(P_{\theta_1}, P_{\theta_0}) . \end{aligned}$$

Here

$$\begin{aligned} \log\frac{p_{\theta_1}}{p_{\theta_0}}(x) &= y\log\left(\frac{\lambda_1 e^{\gamma_1 z}}{\lambda_0 e^{\gamma_0 z}}\right) - \lambda_1 e^{\gamma_1 z} + \lambda_0 e^{\gamma_0 z} \\ &\quad + z\log\frac{\eta_1}{\eta_0} + (1-z)\log\frac{1-\eta_1}{1-\eta_0} , \end{aligned}$$

so

$$
\begin{aligned}
K(P_{\theta_1}, P_{\theta_0}) &= E_{\theta_1} \log \frac{p_{\theta_1}}{p_{\theta_0}}(X_1) \\
&= E_{\theta_1} \left(Y \log \frac{\lambda_1 e^{\gamma_1 Z}}{\lambda_0 e^{\gamma_0 Z}} \right) + \eta_1 (\lambda_0 e^{\gamma_0} - \lambda_1 e^{\gamma_1}) + (1-\eta_1)(\lambda_0 - \lambda_1) \\
&\quad + \eta_1 \log \frac{\eta_1}{\eta_0} + (1-\eta_1) \log \frac{1-\eta_1}{1-\eta_0} \\
&= E_{\theta_1} \left(\lambda_1 e^{\gamma_1 Z} \log \left(\frac{\lambda_1 e^{\gamma_1 Z}}{\lambda_0 e^{\gamma_0 Z}} \right) \right) + \eta_1 (\lambda_0 e^{\gamma_0} - \lambda_1 e^{\gamma_1}) + (1-\eta_1)(\lambda_0 - \lambda_1) \\
&\quad + \eta_1 \log \frac{\eta_1}{\eta_0} + (1-\eta_1) \log \frac{1-\eta_1}{1-\eta_0} \\
&= \eta_1 \left(\lambda_1 e^{\gamma_1} \log \left(\frac{\lambda_1 e^{\gamma_1}}{\lambda_0 e^{\gamma_0}} \right) \right) + (1-\eta_1) \left(\lambda_1 \log \left(\frac{\lambda_1}{\lambda_0} \right) \right) \\
&\quad + \eta_1 (\lambda_0 e^{\gamma_0} - \lambda_1 e^{\gamma_1}) + (1-\eta_1)(\lambda_0 - \lambda_1) \\
&\quad + \eta_1 \log \frac{\eta_1}{\eta_0} + (1-\eta_1) \log \frac{1-\eta_1}{1-\eta_0} .
\end{aligned}
$$

5. For the same set-up as in problem 4, consider taking a "profile likelihood" approach to the estimation of γ as follows:
(a) Let $l_n(\theta) = l_n(\gamma, \lambda, \eta)$: consider first maximizing this as a function of λ and η for each fixed value of γ to find

$$(\widehat{\lambda}(\gamma), \widehat{\eta}(\gamma)) \equiv \text{argmax}_{(\lambda,\eta)} l_n(\lambda, \gamma, \eta) .$$

Compute the maximizer $(\widehat{\lambda}(\gamma), \widehat{\eta}(\gamma))$ as explicitly as possible, and then form the "profile log-likelihood" $l_n^{profile}(\gamma)$ defined by

$$l_n^{profile}(\gamma) \equiv l_n(\widehat{\lambda}(\gamma), \gamma, \widehat{\eta}(\gamma)) .$$

(b) Now maximize $l_n^{profile}(\gamma)$ with respect to γ. Find the resulting "profile likelihood" score equation for γ.
(c) Does the equation you derived in (b) follow from the original score equations?
(d) Does the "profile score function" which appears in (b) correspond to or relate to the efficient score for γ in any way?

Solution: (a) The value of γ doesn't influence the maximization with respect to η and we find $\hat{\eta}(\gamma) = \hat{\eta} = \overline{Z}_n$ for all γ. Solving the score equation for λ for a fixed γ yields

$$\sum_1^n Y_i = \lambda \sum_1^n e^{\gamma Z_i} , \qquad \text{or} \qquad \hat{\lambda}(\gamma) = \frac{\sum_1^n Y_i}{\sum_1^n e^{\gamma Z_i}} .$$

Substitution of these into the log-likelihood yield the profile log-likelihood

$$
\begin{aligned}
l_n^{profile}(\gamma) &= l_n(\hat{\lambda}(\gamma), \gamma, \hat{\eta}) \\
&= \sum_{i=1}^n \left\{ Y_i \log(\hat{\lambda}(\gamma) e^{\gamma Z_i}) - \hat{\lambda}(\gamma) e^{\gamma Z_i} - \log(Y_i!) \right\}
\end{aligned}
$$

$$
\begin{aligned}
& + n\overline{Z}\log\overline{Z} + (n - n\overline{Z})\log(1-\overline{Z}) \\
= \quad & n\overline{Y}\log\hat{\lambda}(\gamma) + \gamma\sum_{i=1}^{n} Y_i Z_i - \hat{\lambda}(\gamma)\sum_{i=1}^{n} e^{\gamma Z_i} + \text{constant in } \gamma\,.
\end{aligned}
$$

(b) Differentiating the profile log-likelihood with respect to γ yields

$$
\dot{l}_{n,\gamma}^{profile}(\gamma) = \frac{d}{d\gamma}\log\hat{\lambda}(\gamma) n\overline{Y}_n + \sum_{i=1}^{n} Y_i Z_i = \sum_{i=1}^{n} Y_i Z_i - n\overline{Y}\frac{\sum_1^n Z_i e^{\gamma Z_i}}{\sum_1^n e^{\gamma Z_i}}
$$

since

$$
\frac{d}{d\gamma}\log\hat{\lambda}(\gamma) = -\frac{\sum_1^n Z_i e^{\gamma Z_i}}{\sum_1^n e^{\gamma Z_i}}\,.
$$

Thus the profile score equation for γ becomes: $\hat{\gamma}^{profile} = \hat{\gamma}$ satisfies

$$
\frac{\sum_1^n Y_i Z_i}{\sum_1^n Y_i} = \frac{\sum_1^n Z_i e^{\hat{\gamma} Z_i}}{\sum_1^n e^{\hat{\gamma} Z_i}}\,. \tag{0.2}
$$

(c) If we solve the original score equation for λ for fixed γ, then we obtain

$$
\hat{\lambda}(\gamma) = \frac{\sum_1^n Y_i}{\sum_1^n e^{\gamma Z_i}}
$$

as in (a). Substitution of this into the score equation for γ yields

$$
\begin{aligned}
0 \quad &= \quad \sum_{i=1}^{n} Z_i Y_i - \hat{\lambda}(\gamma)\sum_{i=1}^{n} Z_i e^{\gamma Z_i} \\
&= \quad \sum_{i=1}^{n} Z_i Y_i - \frac{\sum_1^n Y_i}{\sum_1^n e^{\gamma Z_i}}\sum_{i=1}^{n} Z_i e^{\gamma Z_i}\,,
\end{aligned}
$$

and this implies that (0.2) holds.

(d) To see the connection between the profile score function for γ and the efficient score function for γ, note that

$$
\dot{l}_n^{profile}(\gamma) \quad = \quad \dot{l}_{n,\lambda}(\hat{\lambda}(\gamma),\gamma,\hat{\eta})\frac{d}{d\gamma}\hat{\lambda}(\gamma) + \dot{l}_{n,\gamma}(\hat{\lambda}(\gamma),\gamma,\hat{\eta})\,, \tag{0.3}
$$

and, since $0 = \dot{l}_{n,\lambda}(\hat{\lambda}(\gamma),\gamma,\hat{\eta})$, by differentiating with respect to γ we have

$$
0 = \ddot{l}_{n,\lambda\lambda}(\hat{\lambda}(\gamma),\gamma,\hat{\eta})\frac{d}{d\gamma}\hat{\lambda}(\gamma) + \ddot{l}_{n,\gamma\lambda}(\hat{\lambda}(\gamma),\gamma,\hat{\eta})\,,
$$

and hence

$$
\frac{d}{d\gamma}\hat{\lambda}(\gamma) = -\left(\ddot{l}_{n,\lambda\lambda}(\hat{\lambda}(\gamma),\gamma,\hat{\eta})\right)^{-1}\ddot{l}_{n,\gamma\lambda}(\hat{\lambda}(\gamma),\gamma,\hat{\eta})\,. \tag{0.4}
$$

Substitution of (0.4) into (0.3) yields

$$
\begin{aligned}
\dot{l}_n^{profile}(\gamma) \quad &= \quad \dot{l}_{n,\lambda}(\hat{\lambda}(\gamma),\gamma,\hat{\eta})\frac{d}{d\gamma}\hat{\lambda}(\gamma) + \dot{l}_{n,\gamma}(\hat{\lambda}(\gamma),\gamma,\hat{\eta}) \\
&= \quad \dot{l}_{n,\gamma}(\hat{\lambda}(\gamma),\gamma,\hat{\eta}) - \ddot{l}_{n,\gamma\lambda}(\hat{\lambda}(\gamma),\gamma,\hat{\eta})\left(\ddot{l}_{n,\lambda\lambda}(\hat{\lambda}(\gamma),\gamma,\hat{\eta})\right)^{-1}\dot{l}_{n,\lambda}(\hat{\lambda}(\gamma),\gamma,\hat{\eta}) \\
&= \quad \sum_{i=1}^{n}\left\{\dot{l}_\gamma(X_i) - \hat{I}_{\gamma\lambda}\hat{I}_{\gamma\gamma}^{-1}\dot{l}_\lambda(X_i)\right\}\Big|_{\theta=(\hat{\lambda}(\gamma),\gamma,\hat{\eta})}\,.
\end{aligned}
$$

6. (a) Ferguson, ACILST, problem 17.2, page 117.
   (b) Do our hypotheses A0-A2 hold in this example?
   (c) Compute $K(P_{\theta_0}, P_\theta)$ where P_θ has density as given in this problem.
   (d) Do our hypotheses A3 and A4 hold in this example? Why or why not?
   (e) Does there exist an estimator $\overline{\theta}_n$ of θ which is $n^{1/2}-$consistent?

Solution: (a) (i) Since the density function p_θ is given by

$$p_\theta(x) = 2\left\{\frac{x}{\theta}1_{[0,\theta]}(x) + \frac{1-x}{1-\theta}1_{(\theta,1]}(x)\right\}, \tag{0.5}$$

it follows that for $X_{(k)} < \theta < X_{(k+1)}$ the likelihood is given by

$$L_n(\theta) = 2^k\theta^{-k}\prod_{i\le k} X_{(i)}(1-\theta)^{-(n-k)}\prod_{i>k}(1-X_{(i)}).$$

(This corrects the expression on page 215 of Ferguson in several respects: it changes Ferguson's $+$ to $\cdot$, and it changes the second product from $\prod X_{(i)}$ to $\prod(1-X_{(i)})$.) Thus for $X_{(k)} < \theta < X_{(k+1)}$ we compute

$$\dot{\mathrm{l}}_n(\theta) = -\frac{k}{\theta} + \frac{n-k}{1-\theta},$$

which is < 0 if $\theta < k/n$ and > 0 if $\theta > k/n$.
(a) (ii) Similarly, for $X_{(k)} < \theta < X_{(k+1)}$,

$$\ddot{\mathrm{l}}_n(\theta) = \frac{k}{\theta^2} + \frac{n-k}{(1-\theta)^2} > 0,$$

so the roots (or zeros) of the likelihood equation correspond to *local minima* of the (log-)likelihood, and any local maxima of the log-likelihood occur at the order statistics $X_{(k)}$. It is easily seen that a local maximum occurring at an observation $X_{(k)}$ must correspond to a cusp in the (log-)likelihood: i.e. a point at which $\dot{\mathrm{l}}_n(\theta)$ is positive to the left of $X_{(k)}$ and negative to the right of $X_{(k)}$. Therefore if $\theta = X_{(k)}$ yields a local maximum we have

$$\lim_{\theta\nearrow X_{(k)}}\left\{-\frac{(k-1)}{\theta} + \frac{n-k+1}{1-\theta}\right\} = -\frac{k-1}{X_{(k)}} + \frac{n-k+1}{1-X_{(k)}} > 0,$$

and

$$\lim_{\theta\searrow X_{(k)}}\left\{-\frac{k}{\theta} + \frac{n-k}{1-\theta}\right\} = -\frac{k}{X_{(k)}} + \frac{n-k}{1-X_{(k)}} < 0.$$

But these two inequalities imply that

$$\frac{k-1}{n} < X_{(k)} < \frac{k}{n} \qquad \text{or} \qquad \frac{k-1}{n} < \mathbb{F}_n^{-1}(k/n) < \frac{k}{n}.$$

(b) A0 - A2 all hold in this example: If $\theta \ne \theta^*$, then $p_\theta \ne p_{\theta^*}$ and hence $P_\theta \ne P_{\theta^*}$. 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

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

Here is a plot of $\theta \mapsto K(P_{\theta_0}, P_\theta)$ for $\theta_0 = .2$.

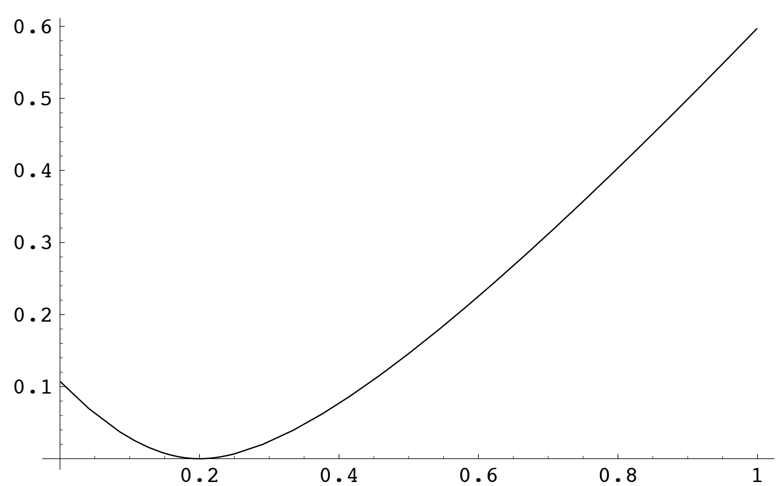

Figure 1: Kullback - Leibler function $K(P_{\theta_0}, P_\theta)$, $\theta_0 = .2$

(d) Since p_θ is given by (0.5),

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

so

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

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

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

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

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

Thus A3 and A4(iii) fail, while A4(i) and A4(ii) hold.

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

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

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

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

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

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

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

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

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

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

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.