Statistics 581, Problem Set 9 Solutions

Wellner; 12/01/2017

1. (a) Ferguson, ACILST, problem 17.2, page 117: I would suggest modifying Ferguson's definition of the density to:

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

(b) Do our hypotheses A0-A2 of Chapter 4, Section 1, 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?
In connection with this problem note that the likelihood function $L(\theta)$ given on Ferguson's page 215 is *not correct*: the formula there should be replaced by

$$L_n(\theta|\underline{X}) = \left(\frac{2}{\theta}\right)^k \prod_{j=1}^k X_{(j)} \cdot \left(\frac{2}{1-\theta}\right)^{n-k} \prod_{j=k+1}^n (1-X_{(j)}) \ \text{ if } \ X_{(k)} \le \theta < X_{(k+1)}.$$

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

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

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 $n = 50$.

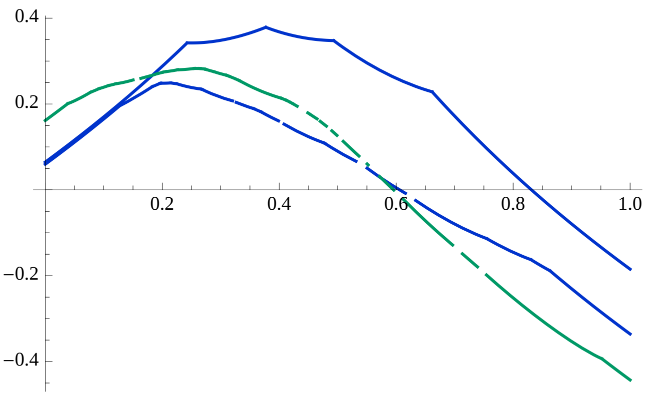

Figure 1: Three log-likelihood functions for problem 3, $n = 4$, $n = 10$, and $n = 50$

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

$$p_\theta(x) = 2\left\{\frac{x}{\theta}1_{[0,\theta]}(x) + \frac{1-x}{1-\theta}1_{(\theta,1]}(x)\right\}, \tag{0.1}$$

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 $\theta \neq \theta^*$, 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

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

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

(d) Since p_θ is given by (0.1),

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

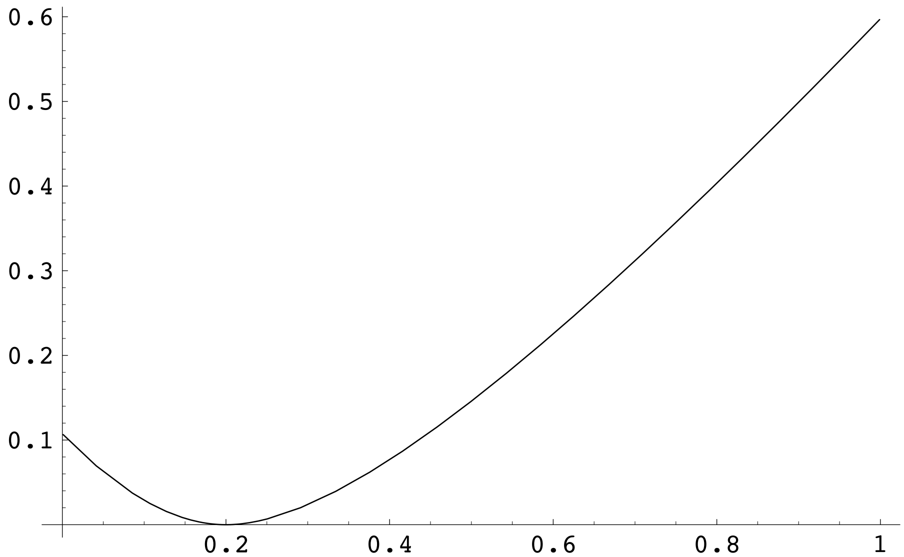

Figure 2: Kullback - Leibler function $K(P_{\theta_0}, P_\theta)$, $\theta_0 = .2$

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

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

we also have

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

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

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

$$
\begin{aligned}
&= \frac{2}{1-\theta}\left\{\frac{1}{3}\theta^2(1-\theta) + \frac{1}{6} - \frac{1}{2}\theta^2 + \frac{1}{3}\theta^3\right\} \\
&= \frac{2}{1-\theta}\left\{\frac{1}{6} - \frac{1}{6}\theta^2\right\} = \frac{1}{3}(1+\theta).
\end{aligned}
$$

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

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

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

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

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

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

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.

2. (a) Lehmann and Casella, problem 6.3.1, page 501: Let X have the binomial distribution $Binomial(n,p)$, $0 \le p \le 1$. Determine the MLE of p by:
(i) The usual calculus method of determining the maximum of a function.
(ii) Showing that $p^x(1-p)^{n-x} \le (x/n)^x((n-x)/n)^{n-x}$.
(b) Lehmann and Casella, problem 6.3.2, page 501: In the preceding problem, show that the MLE does not exist when p is restricted to $0 < p < 1$ and when $X = 0$ or $X = n$.
(c) Lehmann and Casella, problem 6.3.4, page 501: Suppose that $X_1, \ldots, X_n$ are i.i.d. $N(\theta, 1)$ with $\theta > 0$. Show that the MLE is $\overline{X}_n$ when $\overline{X}_n > 0$ and does not exist when $\overline{X}_n \le 0$.

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) When $X_1, \ldots, X_n$ are i.i.d. $N(\xi, 1)$ with $\xi > 0$, the log-likelihood is

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

When $\overline{X}_n > 0$, $l_n(\xi)$ achieves its supremum over $(0, \infty)$ at $\xi = \overline{X}_n$; hence the MLE is $\hat{\xi}_n = \overline{X}_n$ in this case. On the other hand when $\overline{X}_n \le 0$, $l_n(\xi)$ fails to achieve its supremum over $(0,\infty)$, and hence the MLE does not exist. (A "reasonable" estimator is $\tilde{\eta}_n \equiv \overline{X}_n 1_{[0,\infty)}(\overline{X}_n)$, but it is not the MLE.)

3. (a) vdV, Chapter 7, problem 1, page 106: Show that the family of Poisson distributions

$$\mathcal{P} = \{P_\theta : \ p_\theta(k) = \exp(-\theta)\theta^k/k!, \ \theta > 0\}$$

satisfies the conditions of Lemma 7.6 (vdV, page 95).
(b) If $X_1, \ldots, X_n$ are i.i.d. $P_{\theta_0} \in \mathcal{P}$, what can you conclude about

$$\log \prod_{i=1}^n \frac{p_{\theta_0 + hn^{-1/2}}}{p_{\theta_0}}(X_i)?$$

(c) Can you prove the same results as in (b) by a direct argument?

Solution: (a) For the given Poisson(θ) family, $\Theta = (0,\infty)$ is open in $\mathbb{R}$, and

$$s_\theta(k) = \sqrt{p_\theta(k)} = \exp(-\theta/2)(\sqrt{\theta})^k/\sqrt{k!}$$

for $k \in \{0,1,\ldots\}$ is indeed continuously differentiable for each integer $k \geq 0$ with

$$\begin{aligned} \dot{s}_\theta(k) &= (-1/2)\exp(-\theta/2)(\sqrt{\theta})^k/\sqrt{k!} + \exp(-\theta/2)(k/2)\theta^{k/2-1}/\sqrt{k!} \\ &= \frac{1}{2}\left(\frac{k}{\theta} - 1\right) s_\theta(k) \\ &= \frac{1}{2}\dot{l}_\theta(k) s_\theta(k) \end{aligned}$$

where $l(\theta,k) \equiv \log p_\theta(k) = k\log\theta - \theta - \log(k!)$ and

$$\dot{l}_\theta(k) = (\dot{p}_\theta/p_\theta)(k) = k/\theta - 1.$$

Furthermore, $I_\theta = E_\theta(\dot{l}_\theta^2(X) = 1/\theta$ which is continuous in θ. It follows from Lemma 7.6 that the Poisson family is differentiable in quadratic mean.
(b) Hence by Theorem 7.2, vdV page 94, under P_{θ_0},

$$\begin{aligned} \log \prod_{i=1}^n \frac{p_{\theta_0 + hn^{-1/2}}}{p_{\theta_0}}(X_i) &= hn^{-1/2} \sum_{i=1}^n \dot{l}_\theta(X_i) - \frac{1}{2} I(\theta_0) h^2 + o_p(1) \\ &= hn^{-1/2} \sum_{i=1}^n \left(\frac{X_i}{\theta_0} - 1\right) - \frac{1}{2}\theta_0^{-1} h^2 + o_p(1) \\ &\to_d N(0, \theta_0^{-1}) - 2^{-1} h^2 \theta_0^{-1} = N(-\sigma^2/2, \sigma^2) \end{aligned}$$

where $\sigma^2 = h^2 I(\theta_0)$.
(c) In this case we have, by expanding $\log(1+x)$ as $x + (1/2)x^2 + O(x^3)$,

$$\begin{aligned} \log \prod_{i=1}^n \frac{p_{\theta_0 + hn^{-1/2}}}{p_{\theta_0}}(X_i) &= \log \prod_{i=1}^n \frac{e^{-\theta_n}\theta_n^{X_i}/X_i!}{e^{-\theta_0}\theta_0^{X_i}/X_i!} \\ &= \sum_{i=1}^n \{X_i \log(\theta_n/\theta_0) - (\theta_n - \theta_0)\} \end{aligned}$$

$$
\begin{aligned}
&= \log(1 + hn^{-1/2}/\theta_0)\sum_{i=1}^n X_i - n^{1/2}h \\
&= \left(\frac{h}{\sqrt{n}} + \frac{1}{2}\frac{h^2}{n} + O(|h|^3/n^{3/2})\right)\sum_{i=1}^n X_i - n^{1/2}h \\
&= \frac{h}{\sqrt{n}}\sum_{i=1}^n\left(\frac{X_i}{\theta_0} - 1\right) - \frac{1}{2}\frac{h^2}{\theta_0^2}\overline{X}_n + O(n^{-1/2}\overline{X}_n) \\
&= hS_n(\theta_0) - \frac{1}{2}\frac{h^2}{\theta_0} + \frac{1}{2}\frac{h^2}{\theta_0}\left(1 - \frac{\overline{X}_n}{\theta_0}\right) + O(n^{-1/2}\overline{X}_n) \\
&= hS_n(\theta_0) - \frac{1}{2}\frac{h^2}{\theta_0} + o_p(1)
\end{aligned}
$$

where $S_n(\theta_0) = n^{-1/2}\sum_{i=1}^n \dot{l}_\theta(X_i)$. This is the same result as in (b).

4. (a) Show that the Laplace location family is differentiable in quadratic mean. What is the consequence of this for the behavior of the local log-likelihood ratios? What is the resulting information for the location parameter θ?
(b) Apply the methods of section 3.5 to show that if $\theta_0 \in \mathbb{R}$ is fixed and $\theta_n = \theta_0 + n^{-1/2}h$, then for any estimator T_n of θ we have

$$\liminf_{n\to\infty}\inf_{T_n}\max\{E_{\theta_n} n|T_n - \theta_n|^2,\ E_{\theta_0} n|T_n - \theta_0|^2\} \geq cI(\theta_0)^{-1}$$

for some choice of h and an absolute constant c.

Solution: (a) For the Laplace location family $s_\theta(x) = 2^{-1/2}\exp(-|x-\theta|/2)$, so s_θ is differentiable at every $\theta \neq x$ with derivative $\dot{s}_\theta(x) = 2^{-1}\text{sign}(x-\theta)s_\theta(x)$. where $\text{sign}(x) \equiv 2\cdot 1_{[0,\infty)}(x) - 1$. (Here the choice $\text{sign}(0) = 1$ is arbitrary.) Thus to show that $\theta \mapsto s_\theta$ is differentiable in quadratic mean we will show that

$$\int_{\mathbb{R}}\{s_{\theta+h} - s_\theta - h\dot{s}_\theta\}^2\,dx = o(|h|^2). \tag{0.2}$$

But the left side of the last display is equal to

$$
\begin{aligned}
&\int_{\mathbb{R}}\left\{2^{-1/2}\exp(-|x-\theta-h|/2) - \exp(-|x-\theta|/2) - 2^{-1}h\text{sign}(x-\theta)\exp(-|x-\theta|/2)\right\}^2 dx \\
&= 2^{-1}\int_{\mathbb{R}}\exp(-|y|)\left\{\exp(-(|y-h|/2 - |y|/2) - 1 - 2^{-1}h\text{sign}(y)\right\}^2 dy
\end{aligned}
$$

by the change of variable $y = x - \theta$ and then factoring out $\exp(-|y|/2)$. Consider first $h \geq 0$. Then to bound the right side in this last display, we break it into three regions: $(-\infty, 0]$, $(0, h]$, and (h, ∞). Then denoting these terms by I, II, and III, we find that

$$
\begin{aligned}
I &= 2^{-1}\int_{-\infty}^0 e^{-|y|}\{\exp(-h/2) - 1 + (h/2)\}^2 dy = 2^{-1}O((h/2)^4), \\
III &= 2^{-1}\int_h^\infty e^{-|y|}\{\exp(+h/2) - 1 - (h/2)\}^2 dy \leq 2^{-1}O((h/2)^4), \\
II &= 2^{-1}\int_0^h e^{-|y|}\{\exp(h) - 1 - h\}^2 dy \leq 2^{-1}|h|\cdot O(h^2) = O(|h|^3).
\end{aligned}
$$

The argument for $h < 0$ is similar. We conclude that the left side of (0.2) is of order $O(|h|^3) = o(|h|^2)$. Thus (0.2) holds. The resulting information for the location parameter is

$$I(\theta_0) = E_{\theta_0} \dot{l}_\theta^2(X_1;\theta_0) = \int_{\mathbb{R}} [\text{sign}(X_1 - \theta_0)]^2 p_{\theta_0}(x)dx = 1.$$

(b) By Corollary 3, Chapter 3 course notes, page 29 (or by Theorem 7.2, vdV page 94), under P_{θ_0},

$$\log\left(\prod_{i=1}^n \frac{p_{\theta_n}}{p_{\theta_0}}(X_i)\right) = \frac{h}{\sqrt{n}} \dot{l}_\theta(X_i;\theta_0) - \frac{1}{2}h^2 + o_p(1)$$

where $\dot{l}_\theta(x;\theta_0) = \text{sign}(x - \theta_0)$.
Thus by Example 5.1 of the Chapter 3 notes, with $\ell(x) = x^2$ and $c = 1$,

$$\liminf_{n\to\infty} \inf_{T_n} \max\{E_{\theta_n} n|T_n - \theta_n|^2,\ E_{\theta_0} n|T_n - \theta_0|^2\} \ge \frac{e^{-1/2}}{16} \cdot \frac{1}{I(\theta_0)}$$

with $I(\theta_0) = 1$. By Corollary 1 of the convolution Theorem 4.1 of Chapter 3, page 38, the same conclusion holds for (locally) regular estimator sequences $\{T_n\}$ for a larger class of loss functions and without the maximum over risks or the constant factor $e^{-1/2}/16$.

5. Suppose that $X_1, \ldots, X_n$ are i.i.d. log-normal(μ, σ^2) with density

$$p_\theta(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right) 1_{(0,\infty)}(x).$$

Here $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+ \equiv (-\infty, \infty) \times (0, \infty)$.
(a) Find the MLE $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2)$ of $\theta = (\mu, \sigma^2)$.
(b) Show that $\log X \stackrel{d}{=} \mu + \sigma Z \sim N(\mu, \sigma^2)$ where $Z \sim N(0, 1)$.
(c) Suppose that $\nu(P_\theta) = q(\theta) = E_\theta(X)$. Express $q(\theta)$ explicitly as a function of θ.
(d) Suggest a natural nonparametric estimator $\overline{\nu}_n$ of $E_\theta(X)$.
(e) Find the asymptotic variance of $\sqrt{n}(\overline{\nu}_n - \nu(P_\theta))$ for the estimator $\overline{\nu}_n$ you proposed in (d).
(f) What is the MLE $\hat{\nu}_n$ of $\nu = \nu(P_\theta)$ assuming that the log-normal model is true? What do our results in chapter 3 say about the asymptotic distribution of $\sqrt{n}(\hat{\nu}_n - \nu(P_\theta))$ (assuming that the model holds)?
(g) Compare the variances you found in (e) and (f). Which estimator do you prefer if the log-normal model holds?

Solution: First part (b): Note that with $Y \equiv \log X$ it follows that the density f_Y of Y is given by

$$\begin{aligned} f_Y(y) &= f_X(e^y)e^y = p_\theta(e^y)e^y \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right); \end{aligned}$$

i.e. $Y = \log X \sim N(\mu, \sigma^2)$.
(a) Since $Y_i \equiv \log X_i$, $i = 1, \ldots, n$ are i.i.d. $N(\mu, \sigma^2)$,

$$\hat{\mu} = n^{-1} \sum_{i=1}^{n} Y_i = n^{-1} \sum_{1}^{n} \log X_i$$

and

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (Y_i - \overline{Y}_n)^2 = n^{-1} \sum_{i=1}^{n} (\log X_i - \overline{\log(X)})^2$$

are the MLE's of μ and σ^2.
(c) Since $Y = \log X \stackrel{d}{=} \mu + \sigma Z$ where $Z \sim N(0,1)$ and the moment generating function of Z is $Ee^{tZ} = \exp(t^2/2)$, it follows that

$$\begin{aligned} \nu(P_\theta) &= q(\theta) \equiv E_\theta(X) = E_\theta \exp(Y) \\ &= E_\theta \exp(\mu + \sigma Z) = e^{\mu} \cdot \exp(\sigma^2/2). \end{aligned}$$

(d) A natural nonparametric estimator of $q(\theta) = E_\theta(X)$ is $\overline{v}_n = \nu(\mathbb{P}_n) = \int x d\mathbb{P}_n(x) = \overline{X}_n$.
(e) Note that the variance of X is:

$$\begin{aligned} Var_\theta(X) &= E_\theta X^2 - (E_\theta X)^2 = E \exp(2\mu + 2\sigma Z) - e^{2\mu + \sigma^2} \\ &= e^{2\mu} \left\{ e^{2\sigma^2} - e^{\sigma^2} \right\} = e^{2\mu} e^{\sigma^2} \left\{ e^{\sigma^2} - 1 \right\}. \end{aligned}$$

Thus it follows from the central limit theorem that

$$\begin{aligned} \sqrt{n}(\overline{v}_n - \nu(P_\theta)) &= \sqrt{n}(\overline{X}_n - E_\theta(X)) \\ &\to N(0, Var_\theta(X)) = N(0, e^{2\mu} e^{\sigma^2} \{e^{\sigma^2} - 1\}). \end{aligned}$$

(f) From our theory in chapter 4 it follows that

$$\sqrt{n}((\hat{\mu}, \hat{\sigma}^2) - (\mu, \sigma^2)) \to_d D \equiv N_2(0, I^{-1}(\theta))$$

where $I^{-1}(\theta) = \text{diag}(\sigma^2, 2\sigma^4)$. Therefore by Corollary 4.1.1 with $\dot{q}(\theta) = (1, 2^{-1}) \exp(\mu + \sigma^2/2)$,

$$\begin{aligned} \sqrt{n}(\hat{\nu} - \nu(P_\theta)) &\to_d \dot{q}(\theta)^T I(\theta)^{-1} Z \\ &\sim N(0, \dot{q}(\theta)^T I(\theta)^{-1} q(\theta)) = N(0, e^{2\mu + \sigma^2}(\sigma^2 + \sigma^4/2)). \end{aligned}$$

(g) The ratio of the asymptotic variances in (e) and (f) is

$$\begin{aligned} \frac{e^{2\mu + \sigma^2}(\sigma^2 + \sigma^4/2)}{e^{\mu + \sigma^2}(e^{\sigma^2} - 1)} &= \frac{\sigma^2(1 + \sigma^2/2)}{e^{\sigma^2} - 1} \\ &< 1 \quad \text{for all} \quad \mu, \sigma^2 \end{aligned}$$

since $e^x > 1 + x + x^2/2$ for all $x > 0$. Note that this ratio become very small as σ^2 grows; see Figure 3.

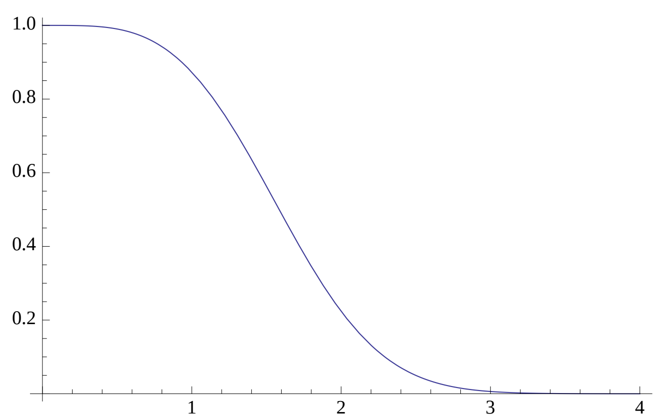

Figure 3: Ratio of asymptotic variance of MLE to variance of $\sqrt{n}(\overline{X}_n - E_\theta(X))$, lognormal mean estimation, as a function of σ.