Assignment 11

Exercise 7.1  In this problem, we explore an example in which the MLE is not consistent. Suppose that for \( \theta \in (0, 1) \), \( X \) is a continuous random variable with density

\[
f_\theta(x) = \theta g(x) + \frac{1 - \theta}{\delta(\theta)} h \left( \frac{x - \theta}{\delta(\theta)} \right),
\]

(7.4)

where \( \delta(\theta) > 0 \) for all \( \theta \), \( g(x) = I\{-1 < x < 1\}/2 \), and

\[
h(x) = \frac{3(1 - x^2)}{4} I\{-1 < x < 1\}.
\]

(a) What condition on \( \delta(\theta) \) ensures that \( \{x : f_\theta(x) > 0\} \) does not depend on \( \theta \)?

**Sketch of solution:**  We should have \(-1 < \theta - \delta(\theta) \) and \( \theta + \delta(\theta) < 1 \); however, the former is always satisfied if the latter is satisfied because \( \theta > 0 \). So we need \( \delta(\theta) < 1 - \theta \).

(b) With \( \delta(\theta) = \exp\{-(1 - \theta)^{-1}\} \), let \( \theta = .2 \). Take samples of sizes \( n \in\{50, 250, 500\} \) from \( f_\theta(x) \). In each case, graph the loglikelihood function and find the MLE. Also, try to identify the consistent root of the likelihood equation in each case.

**Hints:**  To generate a sample from \( f_\theta(x) \), note that \( f_\theta(x) \) is a mixture density, which means you can start by generating a standard uniform random variable. If it’s less than \( \theta \), generate a uniform variable on \((-1, 1)\). Otherwise, generate a variable with density \( 3(\delta^2 - x^2)/4\delta^3 \) on \((-\delta, \delta)\) and then add \( \theta \). You should be able to do this by inverting the distribution function or by using appropriately scaled and translated beta(2, 2) variables. If you use the inverse distribution function method, verify that

\[
H^{-1}(u) = 2 \cos \left\{ \frac{4\pi}{3} + \frac{1}{3} \cos^{-1}(1 - 2u) \right\}.
\]

Be very careful when graphing the loglikelihood and finding the MLE. In particular, make sure you evaluate the loglikelihood **analytically** at each of the sample points in \((0, 1)\) and incorporate these analytical calculations into your code; if you fail to do this, you’ll miss the point of the problem and you’ll get the MLE incorrect. This is because the correct loglikelihood graph will have tall, extremely narrow spikes.
Sketch of solution: The loglikelihood function is \( \ell(\theta) = \sum_{i=1}^{n} \log f_{\theta}(x_i) \).

Focusing on the \( i \)th term in this sum, we see that

\[
\log f_{\theta}(x_i) = \begin{cases} 
\log \left\{ \frac{1-\theta}{\delta(\theta)} \right\} \frac{g\left( \frac{x_i - \theta}{\delta(\theta)} \right) + \frac{\theta}{2}}{2} & \text{if } |x_i - \theta| < \delta(\theta) \\
\log \frac{\theta}{2} & \text{if } |x_i - \theta| \geq \delta(\theta)
\end{cases}.
\]

In particular, if \( \theta = x_i \), then \( g[(x_i - \theta)/\delta] = g(0) = 3/4 \), which means that for positive \( x_i \),

\[
\log f_{x_i}(x_i) = \log \left\{ \frac{3(1-\theta)}{4\delta(\theta)} + \frac{\theta}{2} \right\}.
\]

Notice that when \( \theta > .65 \), \( 3(1-\theta)/4\delta > 10^{28} \), which means that for \( x_i > .65 \), we may use the expression above to write

\[
\log f_{x_i}(x_i) \approx \log \frac{3(1-\theta)}{4\delta(\theta)} = \log \frac{3}{4} + \log(1-\theta) + \frac{1}{(1-\theta)^4}
\]

(ignoring the \( \theta/2 \) term). This approximation is extremely close, accurate to at least 28 decimal places. Furthermore, notice that for \( \theta > .65 \), we will never see more than one \( x_i \) in any interval of width \( 2\delta(\theta) \) unless the \( x_i \) are closer together than \( 2\delta(\theta) < 3 \times 10^{-29} \). Using this information, we may now write a function to evaluate \( \ell(\theta) \) approximately correctly for any sample in which no two \( x_i \) are closer than \( 3 \times 10^{-29} \) (which, in computer arithmetic typically accurate to only 14 or 15 decimal places, means that no two \( x_i \) are exactly equal), as shown below.

delta <- function(theta) exp(-1/(1-theta)^4)
ex7.1sample <- function(n, theta=.2) {
  u <- runif(n)
  x <- theta + (rbeta(n, 2, 2)-.5) * 2 * delta(theta)
  x[u<theta] <- runif(sum(u<theta), -1, 1)
  x
}

loglik <- function (theta, x) {
  d <- delta(theta)
  if (theta<.65) {
    ind <- (abs(x-theta) < d)
    ans <- sum(log(ind*3*(1-theta)*(d^2-(x-theta)^2)/4/d^3 + theta/2))
  }
  else if (any(x==theta)) {
    ans <- (length(x)-1)*log(theta/2) + log(3/4) +
  }
}

217
log(1-theta) + 1/(1-theta)^4
}
else {
  ans <- length(x) * log(theta/2)
}
ans

set.seed(1234)
for (n in c(50, 250, 500)) {
  x <- ex7.1sample(n)
  th <- seq(0, 1, len=400)[-c(1,400)]
  th <- sort(c(th, x[x>0]))
  llth <- rep(0, length(th))
  for (i in 1:length(th)) llth[i] <- loglik(th[i], x)
  mle = round(max(x),4)
  localmax = round(th[llth==max(llth[th<.2])], 4)
  ttl = paste("n =",n,"; MLE =", mle)
  subttl = paste("Plausible Local Maximum =", localmax)
  plot(th, llth, type="l", main=ttl, sub=subttl, xlab="", ylab="",
      ylim=c(min(llth), 3*max(llth[th<.3])), cex.axis=1.5, cex.sub=1.5,cex.main=2)
  abline(v=localmax, lty=2)
}

Exercise 7.7  Prove Theorem 7.9.

Hint: Start with \( \sqrt{n}(\delta_n - \theta_0) = \sqrt{n}(\delta_n - \tilde{\theta}_n) + \sqrt{n}(\tilde{\theta}_n - \theta_0) \), then expand \( \ell'(\tilde{\theta}_n) \) in a Taylor series about \( \theta_0 \) and substitute the result into Equation (7.15). After simplifying, use the result of Exercise 2.2 along with arguments similar to those leading up to Theorem 7.8.
Sketch of solution: Begin with two Taylor expansions, each of which depends on the fact that \( \hat{\theta}_n \to \theta_0 \):

\[
\ell'\left(\hat{\theta}_n\right) = \ell'(\theta_0) + (\hat{\theta}_n - \theta_0) \left[ \ell''(\theta_0) + o_P(1) \right] \quad (7.5)
\]

and

\[
\ell''\left(\hat{\theta}_n\right) = \ell''(\theta_0) + o_P(1). \quad (7.6)
\]

(The latter expression is a “zero-order” Taylor expansion and requires only that \( \ell''(\theta_0) \) be continuous, which follows because the derivative at \( \theta_0 \) does not exist unless it is continuous at \( \theta_0 \).) Since the definition of \( \delta_n \) implies \( \delta_n - \hat{\theta}_n = -\ell'\left(\hat{\theta}_n\right) / \ell''\left(\hat{\theta}_n\right) \), we have

\[
\sqrt{n}(\delta_n - \theta_0) = \sqrt{n}(\delta_n - \hat{\theta}_n) + \sqrt{\left(\hat{\theta}_n - \theta_0\right)} = -\sqrt{n} \frac{\ell'(\hat{\theta}_n)}{\ell''(\hat{\theta}_n)} + \sqrt{n}(\hat{\theta}_n - \theta_0).
\]

Substituting (7.5) into the above expression, we obtain

\[
\sqrt{n}(\delta_n - \theta_0) = -\sqrt{n} \frac{\ell'(\theta_0)}{\ell''(\hat{\theta}_n)} + \sqrt{n}(\hat{\theta}_n - \theta_0) \left[ 1 - \frac{\ell''(\theta_0)}{\ell''(\hat{\theta}_n)} + o_P(1) \right].
\]

Now substitute (7.6) into the above to obtain

\[
\sqrt{n}(\delta_n - \theta_0) = -\sqrt{n} \frac{\ell'(\theta_0)}{\ell''(\theta_0) + o_P(1)} + \sqrt{n}(\hat{\theta}_n - \theta_0) \left[ 1 - \frac{\ell''(\theta_0)}{\ell''(\theta_0) + o_P(1)} + o_P(1) \right].
\]

We see that the term in square brackets tends to zero in probability. Therefore, since \( \hat{\theta}_n \) is assumed to be \( \sqrt{n} \) consistent, we know that

\[
\sqrt{n}(\hat{\theta}_n - \theta_0) \left[ 1 - \frac{\ell''(\theta_0)}{\ell''(\theta_0) + o_P(1)} + o_P(1) \right] \overset{P}{\to} 0
\]

(This is the result of Exercise 2.7.) Therefore, by Slutsky’s theorem it remains only to prove that

\[
-\sqrt{n} \frac{\ell'(\theta_0)}{\ell''(\theta_0) + o_P(1)} \overset{d}{\to} N\left(0, \frac{1}{I(\theta_0)}\right).
\]

But this has already been done as part of the proof of Theorem 8.8; see Equation (8.13) and the ensuing argument.

Exercise 7.8 Suppose that the following is a random sample from a logistic density with distribution function \( F_\theta(x) = (1 + \exp\{\theta - x\})^{-1} \) (I’ll cheat and tell you that I used \( \theta = 2 \)).
(a) Evaluate the unique root of the likelihood equation numerically. Then, taking the sample median as our known $\sqrt{n}$-consistent estimator $\tilde{\theta}_n$ of $\theta$, evaluate the estimator $\delta_n$ in Equation (7.15) numerically.

**Sketch of solution:** The density (found by differentiating the distribution function with respect to $x$) equals

$$f_\theta(x) = \frac{e^{\theta-x}}{(1 + e^{\theta-x})^2},$$

which leads to

$$\frac{d}{d\theta} \log f_\theta(X) = \frac{2}{1 + e^{\theta-x}} - 1 \quad (7.7)$$

$$\frac{d^2}{d\theta^2} \log f_\theta(X) = -\frac{2e^{\theta-x}}{(1 + e^{\theta-x})^2} = -2[f_\theta(X)]^2 \quad (7.8)$$

$$I(\theta) = -E \frac{d^2}{d\theta^2} \log f_\theta(X) = 2 \int_{-\infty}^{\infty} f_\theta^2(x) dx = \frac{1}{3} \quad (7.9)$$

From Equations (7.7) and (7.8), we know that Newton’s method in this problem looks like this:

$$\theta_{new} = \theta_{old} - \frac{\ell'(\theta_{old})}{\ell''(\theta_{old})} = \theta_{old} + \frac{2\sum_i(1 + e^{\theta-x})^{-1} - n}{2\sum_i e^{\theta-x}(1 + e^{\theta-x})^{-2}}.$$  

Therefore, we may write code to implement it as shown below.

```
newton <- function(th, x, iter=1) {
  for (i in 1:iter) {
    etx = exp(th-x)
    th = th + (2*sum(1/(1+etx))-length(x))/(2*sum(etx/(1+etx)^2))
  }
  th
}
```

```
newton <- function(th, x, iter=1) {
  for (i in 1:iter) {
    etx = exp(th-x)
    th = th + (2*sum(1/(1+etx))-length(x))/(2*sum(etx/(1+etx)^2))
  }
  th
}
```

We see that the MLE and $\delta_n$ are 2.392 and 2.385, respectively.
x <- c(1.0944, 1.2853, 1.0422, 6.4723, 1.0439, .169, 3.118, 1.7472, 3.6111, 3.8318, 4.9483, .997, 4.1262, 1.7001, 2.9438)
m <- median(x)
mle <- newton(m, x, iter=100)
delta <- newton(m, x)
mle
[1] 2.39173
delta
[1] 2.385235

(b) Find the asymptotic distributions of \( \sqrt{n}(\tilde{\theta}_n - 2) \) and \( \sqrt{n}(\delta_n - 2) \). Then, simulate 200 samples of size \( n = 15 \) from the logistic distribution with \( \theta = 2 \). Find the sample variances of the resulting sample medians and \( \delta_n \)-estimators. How well does the asymptotic theory match reality?

**Sketch of solution:** From Theorem 6.7, since \( f_0(\theta) = 1/4 \), we know that

\[
\sqrt{n} (\text{sample median} - \theta) \xrightarrow{d} N(0, 4)
\]

because the limiting variance is \((1/2)(1/2)/f^2(\theta) = 4\). Furthermore, from Equation (7.9), we know that

\[
\sqrt{n}(\delta_n - \theta) \xrightarrow{d} N(0, 3)
\]

because \( \delta_n \) is efficient and \( 1/I(\theta) = 3 \). Below is an empirical test of these results, which shows that the theoretical values of 4 and 3 are fairly close. (I notice that if I repeat this experiment, the values jump around quite a bit. We could stabilize this by using more than 200 samples. For instance, when I used 1,000,000 samples, I obtained 3.993 and 3.043, which suggests that the asymptotic approximation is quite good even for \( n = 15 \).)

med200 = rep(0, 200)
delta200 = rep(0, 200)
for (i in 1:200) {
  x = 2+rlogis(15)
  med200[i]=median(x)
  delta200[i] = newton(med200[i], x)
}
15*c(var(med200), var(delta200))
[1] 4.105157 2.841283