STAT 135 MIDTERM EXAMINATION C
SOLUTIONS

July 20, 2021

Problem 1. Let $X_1, \ldots, X_n$ be i.i.d from a population with pdf

$$f(x|\theta) = \begin{cases} \frac{\pi}{2\theta} \sin\left(\frac{\pi|x|}{\theta}\right), & \text{if } -\frac{\theta}{2} \le x \le \frac{\theta}{2} \\ 0, & \text{otherwise.} \end{cases}$$

in which $\theta > 0$ is an unknown parameter.

(a) (3 points) Find a MM estimator of θ; (*Hint*: $\int_0^{\pi/2} t^2 \sin t dt = \pi - 2$.)

Solution. First we calculate the population moments:

$$E(X) = \int_{-\theta/2}^{\theta/2} \frac{\pi x}{2\theta} \sin\left(\frac{\pi|x|}{\theta}\right) dx \overset{\text{odd function}}{=} 0,$$

$$E(X^2) = \int_{-\theta/2}^{\theta/2} \frac{\pi x^2}{2\theta} \sin\left(\frac{\pi|x|}{\theta}\right) dx \overset{\text{even function}}{=} 2\int_0^{\theta/2} \frac{\pi x^2}{2\theta} \sin\left(\frac{\pi x}{\theta}\right) dx$$

$$= \theta^2\left(\frac{1}{\pi} - \frac{2}{\pi^2}\right).$$

(2 pts)

Since the first moment does not contain any information about θ, the MM estimator for θ is obtained using the second moment:

$$\hat{\theta}_{\text{MM}} = \sqrt{\frac{\pi^2}{(\pi-2)n}\sum_{i=1}^n X_i^2}. \qquad \textit{(1 pt)}$$

(b) (4 points) Find the MLE of θ. Is it a sufficient statistic for θ? Explain why;

Solution. The joint likelihood function can be written as

$$L(\theta|\mathbf{X}_n) = \frac{\pi^n}{(2\theta)^n}\prod_{i=1}^n \sin\left(\frac{\pi|X_i|}{\theta}\right)\mathbb{1}\{-\theta/2 \le X_{(1)} \le X_{(n)} \le \theta/2\}$$

$$= \frac{3^n \prod_{i=1}^n X_i^2}{2^n\theta^{3n}}\mathbb{1}\{\theta \ge 2\max\left(-X_{(1)}, X_{(n)}\right)\}.$$

(2 pts)

Since $L(\theta|\mathbf{X}_n)$ is a strictly decreasing function of θ, the maximum likelihood is achieved at $\hat{\theta}_{\text{MLE}} = 2\max\left(-X_{(1)}, X_{(n)}\right)$. *(1 pt)*

$\hat{\theta}_{\text{MLE}}$ is NOT sufficient because the joint likelihood cannot be factorized properly. We need all $|X_i|$'s to contain all information about θ. *(1 pt)*

(c) (5 points) *(Bonus question) Denote the MLE by $\hat{\theta}_{\text{MLE}}$. Prove that $\hat{\theta}_{\text{MLE}}$ can only take values in $[0, \theta]$ and its cdf satisfies

$$P(\hat{\theta}_{\text{MLE}} \leq t) = \left[1 - \cos\left(\frac{\pi t}{2\theta}\right)\right]^n, \quad t \in [0, \theta].$$

Solution. Since both $-X_{(1)}$ and $X_{(n)}$ are in $[-\theta/2, \theta/2]$, we have $\hat{\theta}_{\text{MLE}} \leq \theta$. Also, because $X_{(1)} \leq X_{(n)}$, there must be a positive value between $-X_{(1)}$ and $X_{(n)}$, which means $\hat{\theta}_{\text{MLE}} \geq 0$.

Denote $U = X_{(1)}$ and $V = X_{(n)}$. The joint density between U and V is

$$f(u, v|\theta) = n(n-1)f(u|\theta)f(v|\theta)[F(v|\theta) - F(u|\theta)]^{n-2}, \quad v \geq u,$$

in which the cdf can be calculated as

$$F(u|\theta) = \begin{cases} 1 - \frac{1}{2}\cos\frac{\pi u}{\theta}, & \text{if } u \in (0, \theta/2), \\ \frac{1}{2}\cos\frac{\pi u}{\theta}, & \text{if } u \in (-\theta/2, 0]. \end{cases}$$

Notice that

$$P(\hat{\theta}_{\text{MLE}} \leq t) = P\left[2\max\left(-X_{(1)}, X_{(n)}\right) \leq t\right] = P\left[-t/2 \leq -X_{(1)} \leq X_{(n)} \leq t/2\right],$$

from which we know the domain of integration w.r.t (u, v); see Figure 1. We further divide the domain of integration into three parts, and calculate the integration over these parts separately:

$$\begin{aligned}
P(\hat{\theta}_{\text{MLE}} \leq t) = & \int_{-t/2}^{t/2}\int_{u}^{t/2} f(u,v)dvdu \\
= & n(n-1)\int_{-t/2}^{0}\int_{u}^{0}\left(\frac{\pi}{2\theta}\right)^2 \sin\left(-\frac{\pi u}{\theta}\right)\sin\left(-\frac{\pi v}{\theta}\right)\left[\frac{1}{2}\cos\left(\frac{\pi v}{\theta}\right) - \frac{1}{2}\left(\frac{\pi u}{\theta}\right)\right]^{n-2} dvdu \\
& + n(n-1)\int_{-t/2}^{0}\int_{0}^{t/2}\left(\frac{\pi}{2\theta}\right)^2 \sin\left(-\frac{\pi u}{\theta}\right)\sin\left(\frac{\pi v}{\theta}\right)\left[1 - \frac{1}{2}\cos\left(\frac{\pi v}{\theta}\right) - \frac{1}{2}\left(\frac{\pi u}{\theta}\right)\right]^{n-2} dvdu \\
& + n(n-1)\int_{0}^{t/2}\int_{u}^{t/2}\left(\frac{\pi}{2\theta}\right)^2 \sin\left(\frac{\pi u}{\theta}\right)\sin\left(\frac{\pi v}{\theta}\right)\left[\frac{1}{2}\cos\left(\frac{\pi u}{\theta}\right) - \frac{1}{2}\left(\frac{\pi v}{\theta}\right)\right]^{n-2} dvdu \\
= & \left(\frac{1}{2} - \frac{1}{2}\cos\frac{\pi t}{2\theta}\right)^n + \left(1 - \cos\frac{\pi t}{2\theta}\right)^n - 2\left(\frac{1}{2} - \frac{1}{2}\cos\frac{\pi t}{2\theta}\right)^n + \left(\frac{1}{2} - \frac{1}{2}\cos\frac{\pi t}{2\theta}\right)^n \\
= & \left(1 - \cos\frac{\pi t}{2\theta}\right)^n
\end{aligned}$$

(d) (6 points) Show that $\hat{\theta}_{\text{MLE}}$ is a biased estimator of θ and that $E(\hat{\theta}_{\text{MLE}}) < \theta$ for any sample size n.

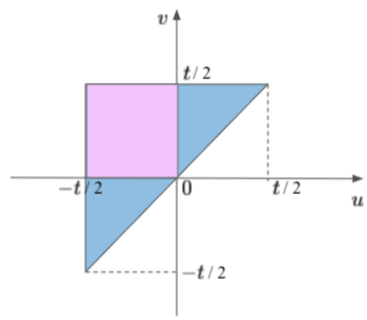

FIGURE 1. Domain of integration to calculate the cdf of $\hat{\theta}_{\text{MLE}}$.

Solution. From (c), we know the pdf of $\hat{\theta}_{\text{MLE}}$ is

$$f_{\text{MLE}}(t) = \frac{n\pi}{2\theta}\left(1 - \cos\frac{\pi t}{2\theta}\right)^{n-1} \sin\frac{\pi t}{2\theta}, \quad t \in [0, \theta].$$

Therefore,

$$\begin{aligned} E(\hat{\theta}_{\text{MLE}}) &= \int_0^\theta \frac{n\pi t}{2\theta}\left(1 - \cos\frac{\pi t}{2\theta}\right)^{n-1} \sin\frac{\pi t}{2\theta} dt && (\textit{1 pt}) \\ &= \frac{2\theta}{\pi}\int_0^{\pi/2} ns\,(1-\cos s)^{n-1}\sin s ds, && (\textit{2 pts}) \\ &= \theta - \frac{2\theta}{\pi}\int_0^{\pi/2} (1-\cos s)^n\, ds && (\textit{2 pts}) \\ &< \theta. && (\textit{1 pt}) \end{aligned}$$

Hence $\hat{\theta}_{\text{MLE}}$ is a biased estimator of θ and $E(\hat{\theta}_{\text{MLE}}) < \theta$ for any sample size n.

(e) (8 points) Use the cdf in (c) to explain why $\hat{\theta}_{\text{MLE}}$ can not be approximated by a Normal distribution as $n \to \infty$. Is it contradictory to Theorem D of Lecture 4?

Solution. Firstly, the density of $\hat{\theta}_{\text{MLE}}$ is not bell-shaped: it is strictly increasing from 0 to θ, and as $n \to \infty$, the density becomes highly concentrated around θ; see Figure 2. This behavior directly indicates the consistency of $\hat{\theta}_{\text{MLE}}$, whose rate of convergence is much faster than $1/\sqrt{n}$ of the common asymptotic normality (proven in (f)). *(3 pts)*

Secondly, the support of the pdf of $n^a\hat{\theta}_{\text{MLE}}$ with $a > 0$ is restricted in $[0, n^a\theta]$, whereas the density of a Normal variable is the real line. *(1 pt)*

It is not contradictory to Theorem D of Lecture 4 because some vital assumptions needed to ensure asymptotic normality are violated in this case:

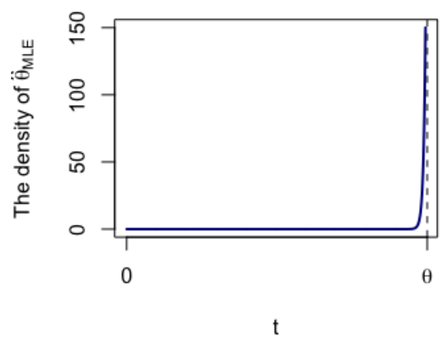

FIGURE 2. Density of of $\hat{\theta}_{\mathrm{MLE}}$.

- The support of $f(x|\theta)$ is not independent of the model parameter θ; *(2 pts)*
- There is no exchangeability between the integration and differentiation. *(2 pts)*

(Assign full marks as long as the student's answer is sufficiently adequate.)

(f) (8 points) Denote $u_k = \int_0^{\pi/2} t^k \sin t dt$. One can estimate

$$\mathrm{Var}(\hat{\theta}_{\mathrm{MM}}) \approx \frac{u_4 - u_2^2}{4nu_2^2}\theta^2, \quad \mathrm{Var}(\hat{\theta}_{\mathrm{MLE}}) \approx \frac{4\theta^2}{\pi}\int_0^{\pi/2}(1-\cos t)^n dt$$

for sufficiently large n. With this result, we can calculate the relative efficiency

$$\mathrm{eff}(\hat{\theta}_{\mathrm{MLE}}, \hat{\theta}_{\mathrm{MM}}) = \frac{\mathrm{Var}(\hat{\theta}_{\mathrm{MLE}})}{\mathrm{Var}(\hat{\theta}_{\mathrm{MM}})}.$$

Show that $\mathrm{eff}(\hat{\theta}_{\mathrm{MLE}}, \hat{\theta}_{\mathrm{MM}}) \to 0$ for large n, which indicates the estimator $\hat{\theta}_{\mathrm{MLE}}$ is a much more efficient estimator than $\hat{\theta}_{\mathrm{MM}}$.
(*Hint*: $\int_0^{\pi/2} n(1-\cos t)^n dt \to 0$ as $n \to \infty$.)

Solution. By the results in the question,

$$\begin{aligned}
\mathrm{eff}(Y, \hat{\theta}_{\mathrm{MM}}) &= \frac{4\theta^2 \int_0^{\pi/2}(1-\cos t)^n dt/\pi}{(u_4 - u_2^2)\theta^2/(4nu_2^2)} && \textit{(2 pts)} \\
&= \frac{16u_2^2}{\pi(u_4-u_2^2)}\int_0^{\pi/2} n(1-\cos t)^n dt && \textit{(3 pts)} \\
&\overset{\text{Hint}}{\to} 0, && \textit{(3 pts)}
\end{aligned}$$

which indeed shows that the unbiased estimator Y is a much more efficient estimator than $\hat{\theta}_{\mathrm{MM}}$.

Problem 2. Suppose that we observe two independent random samples: $X_1, \ldots, X_n$ are i.i.d exponential(θ), and $Y_1, \ldots, Y_m$ are i.i.d exponential(μ).

(a) (8 points) Find the LRT of $H_0 : \theta = \mu$ versus $H_1 : \theta \neq \mu$, and show that the rejection region can be based solely on the statistic

$$T(\mathbf{X}_n, \mathbf{Y}_n) = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i + \sum_{i=1}^m Y_i}.$$

Solution. The joint likelihood can be written as

$$L(\theta, \mu|\mathbf{X}_n, \mathbf{Y}_n) = \theta^n \exp\left(-\theta \sum_{i=1}^n X_i\right) \cdot \mu^m \exp\left(-\mu \sum_{i=1}^m Y_i\right). \quad \textit{(1 pt)}$$

The unrestricted parameter space $\Theta = \{\theta > 0, \mu > 0\}$. By the independence between $\mathbf{X}_n$ and $\mathbf{Y}_n$, the unrestricted maximum likelihood is attained when $L(\theta|\mathbf{X}_n)$ and $L(\mu|\mathbf{Y}_n)$ both attain their respective maximum likelihood. Thus,

$$\begin{aligned}\sup_{(\theta,\mu)\in\Theta} L(\theta, \mu|\mathbf{X}_n, \mathbf{Y}_n) &= L(1/\bar{X}_n, 1/\bar{Y}_m|\mathbf{X}_n, \mathbf{Y}_n) \\ &= \left(\frac{n}{\sum_{i=1}^n X_i}\right)^n \left(\frac{m}{\sum_{i=1}^m Y_i}\right)^m \exp(-n-m). \quad \textit{(2 pts)}\end{aligned}$$

The restricted parameter space $\Theta_0 = \{\theta = \mu\}$. Under H_0, $X_1, \ldots, X_n$, $Y_1, \ldots, Y_m \overset{\text{iid}}{\sim}$ exponential(θ). Therefore, the restricted maximum likelihood is attained at $\hat{\theta}_{\text{Res}} = (m+n)/\left(\sum_{i=1}^n X_i + \sum_{i=1}^m Y_i\right)$, and

$$\begin{aligned}\sup_{(\theta,\mu)\in\Theta_0} L(\theta, \mu|\mathbf{X}_n, \mathbf{Y}_n) &= L(\hat{\theta}_{\text{Res}}, \hat{\theta}_{\text{Res}}|\mathbf{X}_n, \mathbf{Y}_n) \\ &= \left(\frac{m+n}{\sum_{i=1}^n X_i + \sum_{i=1}^m Y_i}\right)^{m+n} \exp(-n-m). \quad \textit{(2 pts)}\end{aligned}$$

The likelihood ratio can be calculated

$$\lambda(\mathbf{X}_n) = \frac{\sup_{(\theta,\mu)\in\Theta_0}}{\sup_{(\theta,\mu)\in\Theta}} = \frac{(m+n)^{m+n}}{n^n m^m} T^n(\mathbf{X}_n, \mathbf{Y}_m)\left(1 - T(\mathbf{X}_n, \mathbf{Y}_m)\right)^m, \quad \textit{(1 pt)}$$

which is a uni-modal function of $T = T(\mathbf{X}_n, \mathbf{Y}_m)$ which attains its maximum at $\frac{n}{n+m}$. Therefore, the rejection region is only dependent on T, which becomes the LRT statistic – $\lambda(\mathbf{X}_n, \mathbf{Y}_m) \leq c$ is equivalent to rejecting if $T \leq c_1$ or $T \geq c_2$, where $c_1 < \frac{n}{n+m} < c_2$. *(2 pts)*

(b) (5 points) Find the distribution of T when H_0 is true.

Solution. Under H_0, $X_1, \ldots, X_n$, $Y_1, \ldots, Y_m \overset{\text{iid}}{\sim}$ Gamma$(1, \theta)$. Thus,

$$\sum_{i=1}^n X_i \sim \text{Gamma}(n, \theta), \quad \sum_{i=1}^m Y_i \sim \text{Gamma}(m, \theta), \quad \sum_{i=1}^n X_i \perp\!\!\!\perp \sum_{i=1}^m Y_i,$$

(3 pts)

from which we know

$$T(\mathbf{X}_n, \mathbf{Y}_m) = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i + \sum_{i=1}^m Y_i} \sim \text{Beta}(n, m). \qquad \textit{(2 pts)}$$

(c) (6 points) Suppose $n = 15$ and $m = 17$. Determine the threshold in the LRT rejection region so that the significance level $\alpha = 0.05$. You might need `R` to help compute the upper α quantile of a $\text{Beta}(a, b)$ distribution: `qbeta(alpha, a, b, lower.tail=FALSE)`

Solution. To make sure the LRT has a significance level $\alpha = 0.05$, the rejection region should be

$$R = \{T \leq \text{Beta}(1 - \alpha/2; n, m) \text{ or } T \geq \text{Beta}(\alpha/2; n, m)\}, \qquad \textit{(3 pts)}$$

in which $\text{Beta}(\alpha/2; n, m)$ is the upper α quantile of the $\text{Beta}(a, b)$ distribution.

Since $n = 15$ and $m = 17$, we can use `R` to obtain

$$\text{Beta}(1 - \alpha/2; n, m) = 0.3015, \quad \text{Beta}(\alpha/2; n, m) = 0.6397. \qquad \textit{(3 pts)}$$

(d) (6 points) A researcher was able to collect two samples from the two populations:

$$\begin{aligned}\mathbf{X}_{15} = \{&4.346,\ 0.353,\ 0.361,\ 0.483,\ 1.693,\ 7.044,\ 0.147,\ 1.810,\\ &2.471,\ 0.102,\ 0.869,\ 0.707,\ 1.719,\ 2.139,\ 0.506\},\\ \mathbf{Y}_{17} = \{&1.307,\ 2.442,\ 0.361,\ 0.421,\ 1.358,\ 0.020,\ 0.558,\ 0.605,\\ &0.750,\ 0.993,\ 0.170,\ 0.088,\ 0.291,\ 0.019,\ 1.052,\ 0.155,\ 0.068\}.\end{aligned}$$

Will you reject the null hypothesis? Report the p-value. You might need `R` to help compute the area under the curve of a $\text{Beta}(a, b)$ density: `pbeta(alpha, a, b, lower.tail=TRUE)`

Solution. The LRT statistic is

$$T(\mathbf{X}_{15}, \mathbf{Y}_{17}) = \frac{24.75}{35.408} = 0.6990. \qquad \textit{(2 pts)}$$

Recall that p-value is the probability of observing a statistic as extreme as 0.6990 under H_0. Since $0.6990 > n/(m + n) = 0.4688$, the observed statistic is larger than the mode of the uni-modal function in (1). Therefore, being as extreme as 0.6990 means (also see Figure 3)

$$\begin{aligned}p\text{-value} =& 2 * P\{T(\mathbf{X}_n, \mathbf{Y}_m) \geq 0.6990 \mid H_0\}\\ =& \texttt{2*pbeta(0.6990, 15, 17, lower.tail=FALSE)}\\ =& 0.00706 < 0.05. \qquad \textit{(2 pts)}\end{aligned}$$

Therefore, we reject H_0 at significance level 0.05, and conclude that there is enough evidence to support the alternative hypothesis $\theta \neq \mu$. (*You can also check whether $T(\mathbf{X}_{15}, \mathbf{Y}_{17})$ is in the rejection region in* (c).) *(2 pts)*

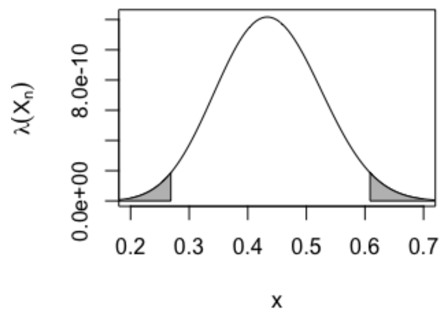

FIGURE 3. What it means to be as extreme as 0.6086678 under H_0.

Problem 3. A research study measured the pulse rates of 62 college women and found a mean pulse rate of 74.4211 beats per minute with a standard deviation S of 10.3628 beats per minute. Let's assume that pulse rates of college women are normally distributed with μ and σ^2.

(1) (6 points) Calculate the exact 99% confidence intervals for the population mean μ and variance σ^2;

Solution. The sample mean is $\bar{X}_n = 74.4211$, and the two sample variances can be computed as

$$S^2 = 107.3876, \quad \hat{\sigma}_n^2 = 105.6556. \qquad (\textit{1 pts})$$

When $\alpha = 0.01$,

$$t_{61}(\alpha/2) = 2.66, \quad \chi^2_{61}(1-\alpha/2) = 36.301, \quad \chi^2_{61}(\alpha/2) = 93.186, \qquad (\textit{2 pts})$$

the 99% exact confidence interval for μ is

$$\bar{X}_n \pm \frac{S}{\sqrt{n}} t_{56}(\alpha/2) = [70.907, 77.935], \qquad (\textit{1.5 pts})$$

and the 99% exact confidence interval for σ^2 is

$$\left[\frac{n\hat{\sigma}_n^2}{\chi^2_{61}(\alpha/2)}, \frac{n\hat{\sigma}_n^2}{\chi^2_{61}(1-\alpha/2)}\right] = [70.296, 180.454]. \qquad (\textit{1.5 pts})$$

(2) (6 points) Researchers want to know if the mean pulse rate for all college women is less than the current standard of 76 beats per minute. Construct a null and an alternative hypotheses. Test these hypothesis at significance level $\alpha = 0.05$.

Solution. The null and alternative hypotheses are

$$H_0 : \mu = 76 \text{ versus } H_1 : \mu < 76. \qquad (\textit{2 pts})$$

The LRT rejection region at significance level $\alpha = 0.05$ is

$$R = \left\{ \frac{\sqrt{n}\left(\bar{X}_n - 76\right)}{S} \leq -t_{61}(\alpha) \right\}, \qquad \textit{(2 pts)}$$

in which $t_{61}(\alpha) = 1.6702$.

Since $\sqrt{n}\left(\bar{X}_n - 72\right)/S = -1.1997$ which is not in the rejection region, we fail to reject H_0 and conclude that there is not enough evidence to show the mean pulse rate is less than 76. (*You can reach the same conclusion via calculating the p-value = 0.1174 which is not significant.*) *(2 pts)*

(3) (6 points) Calculate the threshold of the LRT rejection region at significance level $\alpha = 0.05$ for testing

$$H_0 : \sigma^2 = 100 \text{ versus } H_1 : \sigma^2 \neq 100.$$

Is the LRT uniformly most powerful in this case? Does the sample variance fall within the rejection region?

Solution. The LRT of the two-sided test should be

$$R = \left\{ \frac{(n-1)\,S^2}{100} \geq \chi^2_{n-1}(\alpha/2) \;\text{ or }\; \frac{(n-1)\,S^2}{100} \leq \chi^2_{n-1}(1-\alpha/2) \right\},$$

in which

$$\chi^2_{n-1}(1-\alpha/2) = 41.303, \quad \chi^2_{n-1}(\alpha/2) = 84.476 \qquad \textit{(2 pts)}$$

when $\alpha = 0.05$.

Because $(n-1)S^2/100 = 65.506$, the sample is not in the rejection region, and hence we fail to reject H_0. *(2 pts)*

This LRT is not UMP because we are dealing with a two-sided test. In particular, a two-sided test would fail to give us a monotonic power function in the alternative parameter space, which violates the assumption of Karlin-Rubin. *(2 pts)*

Problem 4. Suppose $X_1, \ldots, X_n$ are i.i.d observations from a population with pmf

$$P(X = x|\theta) = \theta^x (1-\theta)^{1-x}, \quad , x = 0 \text{ or } 1, \quad \frac{1}{3} \leq \theta \leq 1$$

.

(a) (8 points) Find the MM estimator and MLE of θ;

Solution. The first moment of this population is

$$\mu = EX = \theta, \qquad \textit{(2 pts)}$$

and thus $\hat{\theta}_{MM} = \bar{X}_n$.

To obtain the MLE, write out the log-likelihood function

$$l(\theta) = \left(\sum_{i=1}^n X_i\right) \log\theta + \left(n - \sum_{i=1}^n X_i\right) \log(1-\theta). \qquad \textit{(2 pts)}$$

Take the first derivative, and we know

$$\frac{\partial l}{\partial \theta} = \sum_{i=1}^{n} X_i \Big/ \theta - \left(n - \sum_{i=1}^{n} X_i\right) \Big/ (1-\theta) > 0 \qquad \textit{(2 pts)}$$

is equivalent to $\theta < \bar{X}_n$. Note that $1/3 \le \theta \le 1$, which gives

$$\hat{\theta}_{MLE} = \max\{\bar{X}_n, 1/3\}. \qquad \textit{(2 pts)}$$

(b) (10 points) Try to find the MSE of each of the estimators (*See Page 29 of Lecture 1 slides for the definition of MSE*);

Solution. By the definition of MSE,

$$\begin{aligned}\mathrm{MSE}(\hat{\theta}_{MM}) =& E(\hat{\theta}_{MM} - \theta)^2 \\ =& \sum_{k=0}^{n} \binom{n}{k} \left(\frac{k}{n} - \theta\right)^2 \theta^k (1-\theta)^{n-k} = \frac{\theta(1-\theta)}{n}. \qquad (1)\end{aligned}$$

(2 pts)

For the MLE, let's first examine the sampling distribution. We know that all possible values of $\hat{\theta}_{MLE}$ are $1/3, \lceil n/3 \rceil / n, \ldots, (n-1)/n, 1$, in which $\lceil \cdot \rceil$ is the ceiling function. *(1 pts)*

$$P(\hat{\theta}_{MLE} = p) = P(\hat{\theta}_{MLE} = p, \bar{X}_n \le 1/3) + P(\hat{\theta}_{MLE} = p, \bar{X}_n > 1/3)$$

$$= \begin{cases} P(\bar{X}_n \le 1/3), & p = 1/3, \\ P(\bar{X}_n = k), & p = \lceil n/3 \rceil / n, \ldots, (n-1)/n, 1, \end{cases}$$

$$\overset{\sum_{i=1}^{n} X_i \text{ is binomial}}{=} \begin{cases} \sum_{k=0}^{\lceil n/3 \rceil - 1} \binom{n}{k} \theta^k (1-\theta)^{n-k}, & p = 1/3, \\ P(\bar{X}_n = k), & p = \lceil n/3 \rceil / n, \ldots, (n-1)/n, 1. \end{cases}$$

(4pts)

Therefore,

$$\begin{aligned}&\mathrm{MSE}(\hat{\theta}_{MLE}) = E(\hat{\theta}_{MLE} - \theta)^2 \\ &= \sum_{k=0}^{\lceil n/3 \rceil - 1} \binom{n}{k} \left(\frac{1}{3} - \theta\right)^2 \theta^k (1-\theta)^{n-k} + \sum_{k=\lceil n/3 \rceil}^{n} \binom{n}{k} \left(\frac{k}{n} - \theta\right)^2 \theta^k (1-\theta)^{n-k} \qquad (2)\end{aligned}$$

(3pts)

(c) (10 points) Which estimator is preferred? Justify your choice.

Solution. From equations (1) and (2), we have

$$\begin{aligned}
\mathrm{MSE}(\hat{\theta}_{MM}) - \mathrm{MSE}(\hat{\theta}_{MLE}) &= \sum_{k=0}^{\lceil n/3 \rceil - 1} \binom{n}{k} \left[\left(\frac{k}{n} - \theta \right)^2 - \left(\frac{1}{3} - \theta \right)^2 \right] \theta^k (1-\theta)^{n-k} \\
&= \sum_{k=0}^{\lceil n/3 \rceil - 1} \binom{n}{k} \left(\frac{k}{n} - \frac{1}{3} \right) \left(\frac{k}{n} + \frac{1}{3} - 2\theta \right) \theta^k (1-\theta)^{n-k}.
\end{aligned}$$

(4pts)

The facts that $k/n < 1/3$ when $k \leq \lceil n/3 \rfloor - 1$ and $\theta \geq 1/3$ imply that every term in the sum is positive; that is,

$$\mathrm{MSE}(\hat{\theta}_{MM}) - \mathrm{MSE}(\hat{\theta}_{MLE}) > 0 \qquad \textit{(4 pts)}$$

for any θ in [1/3,1]. Therefore, MLE is better than MM estimator because it always has a smaller MSE. *(2 pts)*