Problem 1: Minimal Sufficient Statistics (25 points)

(a) We know that \( X_1, \ldots, X_n \sim U[\theta - 1/2, \theta + 1/2] \). So the likelihood function can be manipulated as follows:

\[
L(\theta; X_1, \ldots, X_n) = \prod_{i=1}^{n} I\{\theta - 1/2 \leq X_i \leq \theta + 1/2\} \\
= I\{\theta - 1/2 \leq X_{(1)} \leq X_{(n)} \leq \theta + 1/2\} \\
= I\{X_{(n)} - 1/2 \leq \theta \leq X_{(1)} + 1/2\}.
\]

Notice that \( X_{(n)} - X_{(1)} \leq 1 \), hence \( X_{(1)} + \frac{1}{2} \geq X_{(n)} - \frac{1}{2} \).

\( T(x_1, \ldots, x_n) = \{x_{(1)}, x_{(n)}\} \) is minimal sufficient, because given two samples \( x = (x_1, \ldots, x_n) \) and \( y = (y_1, \ldots, y_n) \), the ratio

\[
R(x_1, \ldots, x_n, y_1, \ldots, y_n, \theta) = \frac{p(x_1, \ldots, x_n; \theta)}{p(y_1, \ldots, y_n; \theta)} = \frac{I\{x_{(n)} - 1/2 \leq \theta \leq x_{(1)} + 1/2\}}{I\{y_{(n)} - 1/2 \leq \theta \leq y_{(1)} + 1/2\}}
\]

does not depend on \( \theta \) (or it is constant with respect to \( \theta \); or \( p(x; \theta) \propto p(y; \theta) \)) if and only if \( T(x) = T(y) \). In fact \(^1\):

- Case 1: Without loss of generality we assume \( x_{(n)} - 0.5 < y_{(n)} - 0.5 \), that is

\[
x_{(n)} - 0.5 < y_{(n)} - 0.5 < y_{(1)} + 0.5 < x_{(1)} + 0.5
\]

\[
R(x_1, \ldots, x_n, y_1, \ldots, y_n, \theta) = \begin{cases} 
\frac{0}{0} = 1, & \theta < x_{(n)} - 0.5 \text{ or } \theta > x_{(1)} + 0.5 \\
\frac{0}{\infty} = \infty, & \theta \in [x_{(n)} - 0.5, y_{(n)} - 0.5] \\
\frac{\infty}{\infty} = \infty, & \theta \in (y_{(1)} + 0.5, x_{(1)} + 0.5) \\
\frac{1}{1} = 1, & \theta \in [y_{(n)} - 0.5, y_{(1)} + 0.5] 
\end{cases} \tag{1}
\]

- Case 2: Again without loss of generality we assume \( x_{(n)} - 0.5 < y_{(n)} - 0.5 \), that is

\[
x_{(n)} - 0.5 < y_{(n)} - 0.5 < x_{(1)} + 0.5 < y_{(1)} + 0.5
\]

\[
R(x_1, \ldots, x_n, y_1, \ldots, y_n, \theta) = \begin{cases} 
\frac{0}{0} = 1, & \theta < x_{(n)} - 0.5 \text{ or } \theta > y_{(1)} + 0.5 \\
\frac{0}{\infty} = \infty, & \theta \in [x_{(n)} - 0.5, y_{(n)} - 0.5] \\
\frac{\infty}{1} = \infty, & \theta \in [y_{(n)} - 0.5, x_{(1)} + 0.5] \\
\frac{0}{0} = 0, & \theta \in (x_{(1)} + 0.5, y_{(1)} + 0.5) 
\end{cases} \tag{2}
\]

\(^1\)Notice that since we are actually just checking for proportionality of \( p(x_1, \ldots, x_n; \theta) \propto p(y_1, \ldots, y_n; \theta) \), we will have \( \frac{0}{0} = 1 \) and \( \frac{\infty}{0} = \infty \).
We first show that

\[ x(n) - 0.5 < x(1) + 0.5 < y(n) - 0.5 < y(1) + 0.5 \]

Now given two samples \( x, y \) for the ratio \( R(x, y) = \frac{p(x, y)}{p(x, y)} \) we can use the likelihood function, (4)

\[ R(x_1, ..., x_n, y_1, ..., y_n, \theta) = \begin{cases} \frac{0}{0} = 1, & \theta < x(n) - 0.5 \text{ or } \theta > y(1) + 0.5 \\ \frac{0}{0} = 1, & \theta \in \left( x(n) - 0.5, x(1) + 0.5 \right) \\ \frac{0}{0} = 1, & \theta \in \left( y(n) - 0.5, y(1) + 0.5 \right) \end{cases} \]

Case 3: Again without loss of generality we assume \( x(n) - 0.5 < y(n) - 0.5 \), that is

\[ x(n) - 0.5 < x(1) + 0.5 < y(n) - 0.5 < y(1) + 0.5 \]

R(x_1, ..., x_n, y_1, ..., y_n, \theta) = \begin{cases} 0, & \theta < x(n) - 0.5 \text{ or } \theta > y(1) + 0.5 \\ 1, & \theta \in \left( x(n) - 0.5, x(1) + 0.5 \right) \\ 1, & \theta \in \left( y(n) - 0.5, y(1) + 0.5 \right) \end{cases} \]

Case 4: \( y(n) - 0.5 = x(n) - 0.5 \) and without loss of generality if we assume \( x(1) + 0.5 \)

\[ R(x_1, ..., x_n, y_1, ..., y_n, \theta) = \begin{cases} 0, & \theta < x(n) - 0.5 \text{ or } \theta > y(1) + 0.5 \\ 1, & \theta \in \left( y(n) - 0.5, y(1) + 0.5 \right) \end{cases} \]

Case 5: \( y(n) - 0.5 = x(n) - 0.5 \) and \( x(1) + 0.5 = y(1) + 0.5 \)

\[ R(x_1, ..., x_n, y_1, ..., y_n, \theta) = \begin{cases} 0, & \theta < x(n) - 0.5 \text{ or } \theta > x(1) + 0.5 \\ 1, & \theta \in \left( x(n) - 0.5, x(1) + 0.5 \right) \end{cases} \]

Thus, only in Case 5 have \( R(x_1, ..., x_n, y_1, ..., y_n, \theta) = 1 \) for any \( \theta \). Therefore, statistic \( T(x_1, ..., x_n) = \left\{ x_1, x_n \right\} \) is minimal sufficient for \( \theta \). Note that we can assume \( x(n) - 0.5 < y(n) - 0.5 \) in case 1, 2 and 3 and \( x(1) + 0.5 < y(1) + 0.5 \) in case 4 because the ratio \( R(x_1, ..., x_n, y_1, ..., y_n, \theta) \) is used to check for proportionality of the two likelihood functions \( p(x_1, ..., x_n; \theta) \propto p(y_1, ..., y_n; \theta) \) and we could just switch numerator and denominator of the ratio, i.e. use the ratio \( R(y_1, ..., y_n, x_1, ..., x_n, \theta) = \frac{p(y_1, ..., y_n, \theta)}{p(x_1, ..., x_n, \theta)} \).

(b) We first show that \( T(X_1, ..., X_n) = \sum_{i=1}^n X_i \) is minimal sufficient. First we derive the likelihood function,

\[ L(p; X_1, ..., X_n) = \prod_{i=1}^n p^{X_i}(1-p)^{1-X_i} = p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum_{i=1}^n X_i}. \]

Now given two samples \( x = (x_1, ..., x_n) \) and \( y = (y_1, ..., y_n) \), the ratio

\[ R(x_1, ..., x_n, y_1, ..., y_n, p) = \frac{p(x_1, ..., x_n; p)}{p(y_1, ..., y_n; p)} = \frac{p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}}{p^{\sum_{i=1}^n y_i}(1-p)^{n-\sum_{i=1}^n y_i}} = \left( \frac{p}{1-p} \right)^{\sum_{i=1}^n x_i - \sum_{i=1}^n y_i} \]

does not depend on \( p \) (or it is constant with respect to \( p \); or \( p(x; p) \propto p(y; p) \)) if and only if \( \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \), which implies that \( \sum_{i=1}^n X_i \) is minimal sufficient. Now note that since \( X_i \sim \text{Ber}(p), X_i \in \{0, 1\} \) and so \( X_i = X_i^2 \). Hence \( T(X_1, ..., X_n) = \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i \) is minimal sufficient.

Now there is no function \( g(t) \) such that \( T(X_1, ..., X_n) = g(X_3) \) for any sample \( X_1, ..., X_n \), i.e. the minimal sufficient statistic cannot be derived by \( X_3 \) alone and therefore \( X_3 \) cannot be sufficient.
Alternate proofs

– Alternatively note that
\[ L(p; X_1, ..., X_n) = p^{\sum_{i=1}^n X_i} (1 - p)^{n - \sum_{i=1}^n X_i}, \]
showing that the likelihood function cannot be factorized into \( g(p; X_3) \times h(X) \) since the rest of the data is still required to define the likelihood function.

– A more direct way to prove that \( X_3 \) is not sufficient is to show that the joint distribution of \( X_1, ..., X_n \mid X_3 \) still depends on \( p \): for \( x_1, ..., x_n, s \in \{0, 1\} \)
\[
P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n \mid X_3 = s) = \frac{P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n, X_3 = s)}{P(X_3 = s)}
= I\{x_3 = s\} \frac{p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i}}{p^{x_3} (1 - p)^{x_3}}
= I\{x_3 = s\} \frac{p^{\sum_{i=3}^n x_i} (1 - p)^{n - \sum_{i=3}^n x_i}}{p^{x_3} (1 - p)^{x_3}},
\]
which still depends on \( p \).

**Problem 2: MLEs for Geometric distribution (35 points)**

(a) Since there is only one parameter to estimate and \( \mathbb{E}[X_1] = \frac{1}{p} \), the method of moments estimator \( \hat{p}_m \) is given by:
\[
\frac{1}{\hat{p}_m} = \frac{1}{n} \sum_{i=1}^n X_i \implies \hat{p}_m = \frac{1}{\overline{X}}.
\]

To find the maximum likelihood estimator \( \hat{p}_{mle} \) calculate the likelihood function:
\[
L(p) = \prod_{i=1}^n (1 - p)^{X_i - 1} p = (1 - p)^{\sum_{i=1}^n X_i - n} p^n.
\]

So the log likelihood function is given by:
\[
l_n(p) = \left( \sum_{i=1}^n X_i - n \right) \log(1 - p) + n \log p.
\]

Its first derivative, the score equals to:
\[
s(p) = \frac{\partial}{\partial p} l_n(p) = -\frac{\sum_{i=1}^n X_i - n}{1 - p} + \frac{n}{p}.
\]

Setting it equal to 0 we get,
\[
\frac{\sum_{i=1}^n X_i - n}{1 - p} = \frac{n}{p} \implies p \sum_{i=1}^n X_i - np = n - np \implies p = \frac{n}{\sum_{i=1}^n X_i},
\]
which implies that the unique solution \( \hat{p}_{mle} = \frac{1}{\overline{X}}. \) This is a global maximum because the second derivative is negative for every \( p \)
\[
\frac{\partial^2}{\partial p^2} l_n(p) = -\frac{\sum_{i=1}^n X_i - n}{(1 - p)^2} - \frac{n}{p^2} < 0.
\]
To get the Fisher information we use the second derivative to get:

\[ I_n(p) = -\mathbb{E} \left[ \frac{\partial^2}{\partial p^2} \log \left( \frac{\mathbb{E}[X_i]}{(1-p)^2} \right) \right] = \frac{\sum_{i=1}^n \mathbb{E}[X_i] - n}{(1-p)^2} + \frac{n}{p^2} \]

\[ = \frac{n-p-n}{(1-p)^2} + \frac{n}{p^2} \]

\[ = \frac{n-np}{p(1-p)^2} + \frac{n}{p^2} \]

\[ = \frac{n}{p(1-p)} + \frac{n}{p^2} = \frac{n}{p^2(1-p)}. \]

(b) The limiting distribution of \( \hat{p}_{mle} \) is given by,

\[ \sqrt{n} (\hat{p}_{mle} - p) \xrightarrow{d} N \left( 0, \frac{1}{I_1(p)} \right). \]

Therefore, in this case

\[ \sqrt{n} (\hat{p}_{mle} - p) \xrightarrow{d} N \left( 0, p^2(1-p) \right). \]

The MLE for \( \theta = p^2 \), by the equivariance property of the MLE is given by:

\[ \hat{\theta}_{mle} = \hat{p}_{mle}^2 = \frac{1}{X^2}. \]

By Delta method the distribution of \( \theta = g(p) = p^2 \) is given by,

\[ \sqrt{n} (g(\hat{p}_{mle}) - g(p)) \xrightarrow{d} N \left( 0, \frac{|g'(p)|^2}{I_1(p)} \right). \]

In this case since \( g'(p) = 2p \),

\[ \sqrt{n} (\hat{\theta}_{mle} - \theta) \xrightarrow{d} N \left( 0, 4p^4(1-p) \right). \]

Problem 3: Testing using the score statistic (40 points)

(a) We have that \( X_1, \ldots, X_n \sim p(X; \theta) \). Then the score function is given by

\[ s(\theta) = \sum_{i=1}^n \nabla \log \left( p(X_i; \theta) \right). \]

Then the Fisher’s information for \( \theta \in \mathbb{R} \), \( I(\theta) \) is defined as

\[ I(\theta) = \mathbb{E}[s(\theta)^2]. \]

As shown in Lecture 15 page 4, \( \mathbb{E}[s(\theta)] \) and hence the variance of the score statistic is given by,

\[ \text{Var}(s(\theta)) = \mathbb{E} \left[ s(\theta)^2 \right] = I(\theta) = nI_1(\theta). \]
So for the true parameter $\theta$,
\[
\frac{s(\theta)}{\sqrt{n}} \xrightarrow{d} N(0, I_1(\theta)).
\]

Remind yourself that,
\[
\mathbb{E}[s(\theta)] = \sum_{i=1}^{n} \int \nabla_{\theta} \log p(x_i; \theta) p(x_1, \ldots, x_n; \theta) dx_1 \ldots dx_n
\]
\[
= \sum_{i=1}^{n} \int \nabla_{\theta} \log p(x_i; \theta) p(x_i; \theta) dx_i
\]
\[
= n \int \nabla_{\theta} \log p(x_1; \theta) p(x_1; \theta) dx_1,
\]
using the i.i.d. assumption several times. Under some regularity conditions we can switch the derivative and integral (essentially the dominated convergence theorem again but see the Lehmann and Casella book for details) so we obtain,
\[
\int \nabla_{\theta} \log p(x_1; \theta) p(x_1; \theta) dx_1 = \int \frac{\nabla_{\theta} p(x_1; \theta)}{p(x_1; \theta)} p(x_1; \theta) dx_1
\]
\[
= \nabla_{\theta} \int p(x_1; \theta) dx_1 = \nabla_{\theta} 1 = 0.
\]

(b) i. Now we wish to test,
\[
H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.
\]

Then under the null hypothesis since the true $\theta = \theta_0$, we have that
\[
\frac{s(\theta_0)}{\sqrt{n}} \xrightarrow{d} N(0, I_1(\theta_0)).
\]

Therefore,
\[
\frac{s(\theta_0)}{\sqrt{nI_1(\theta_0)}} \xrightarrow{d} N(0, 1).
\]

So to ensure the asymptotic Type I error to be $\alpha$, we can reject $H_0$ if
\[
\frac{|s(\theta_0)|}{\sqrt{nI_1(\theta_0)}} > z_{\alpha/2}.
\]

ii. Suppose for the observed data, $s(\theta_0) = 3.05$, then observed
\[
\frac{s(\theta_0)}{\sqrt{nI_1(\theta_0)}} = \frac{3.05}{\sqrt{nI_1(\theta_0)}},
\]
and let
\[
\frac{s(\theta_0)}{\sqrt{nI_1(\theta_0)}} \xrightarrow{d} Z \sim N(0, 1),
\]

5
then the approximate \( p - value \) is given by

\[
\text{p-value} = P_{\theta_0} \left( \frac{\sqrt{n} s(\theta_0)}{\sqrt{I_1(\theta_0)}} > 3.05 \frac{\sqrt{n}}{\sqrt{I_1(\theta_0)}} \right) 
\approx P \left( |Z| > \frac{3.05}{\sqrt{nI_1(\theta_0)}} \right) 
= P \left( Z < - \frac{3.05}{\sqrt{nI_1(\theta_0)}} \right) + P \left( Z > \frac{3.05}{\sqrt{nI_1(\theta_0)}} \right) 
= \Phi \left( - \frac{3.05}{\sqrt{nI_1(\theta_0)}} \right) + 1 - \Phi \left( \frac{3.05}{\sqrt{nI_1(\theta_0)}} \right) 
= 2 - 2\Phi \left( \frac{3.05}{\sqrt{nI_1(\theta_0)}} \right) = 2\Phi \left( - \frac{3.05}{\sqrt{nI_1(\theta_0)}} \right),
\]

where \( \Phi \) is the CDF of \( N(0,1) \).

(c) Suppose now under the alternate hypothesis, the true parameter \( \theta = \theta \neq \theta_0 \), then

\[
\frac{s(\theta_1)}{\sqrt{nI_1(\theta_1)}} \xrightarrow{d} N(0,1).
\]

The power of the test when computed at \( \theta_1 \) is given by

\[
\text{Power}(\theta_1) = P_{\theta_1} \left( \frac{|s(\theta_0)|}{\sqrt{nI_1(\theta_0)}} > z_{\alpha/2} \right) 
= P_{\theta_1} \left( \frac{s(\theta_0)}{\sqrt{n}} < -z_{\alpha/2}\sqrt{I_1(\theta_0)} \right) + P_{\theta_1} \left( \frac{s(\theta_0)}{\sqrt{n}} > z_{\alpha/2}\sqrt{I_1(\theta_0)} \right).
\]

Now by Taylor series expansion if \( \theta_1 \) is close to \( \theta_0 \), we can say that

\[
\frac{s(\theta_0)}{\sqrt{n}} \approx \frac{s(\theta_1)}{\sqrt{n}} + \sqrt{n}(\theta_0 - \theta_1) \frac{s'(\theta_1)}{n}.
\]

Now, \( \mathbb{E}[-s'(\theta_1)] = I_1(\theta_1) = nI_1(\theta_1) \) and hence by WLLN,

\[
\frac{s'(\theta_1)}{n} \xrightarrow{p} -I_1(\theta_1).
\]

Therefore,

\[
\frac{s(\theta_0)}{\sqrt{n}} \approx \frac{s(\theta_1)}{\sqrt{n}} - \sqrt{n}(\theta_0 - \theta_1)I_1(\theta_1).
\]
So plugging into the expression for the power we get,

\[
\text{Power}(\theta_1) = P_{\theta_1} \left( \frac{s(\theta_0)}{\sqrt{n}} < -\frac{\alpha}{2}\sqrt{I_1(\theta_0)} \right) + P_{\theta_1} \left( \frac{s(\theta_0)}{\sqrt{n}} > \frac{\alpha}{2}\sqrt{I_1(\theta_0)} \right)
\]

\[
\approx P_{\theta_1} \left( \frac{s(\theta_1)}{\sqrt{n}} - \sqrt{n}(\theta_0 - \theta_1)I_1(\theta_1) < -\frac{\alpha}{2}\sqrt{I_1(\theta_0)} \right) + P_{\theta_1} \left( \frac{s(\theta_1)}{\sqrt{n}} - \sqrt{n}(\theta_0 - \theta_1)I_1(\theta_1) > \frac{\alpha}{2}\sqrt{I_1(\theta_0)} \right)
\]

\[
= P_{\theta_1} \left( \frac{s(\theta_1)}{\sqrt{nI_1(\theta_1)}} < -\frac{\alpha}{2}\sqrt{\frac{I_1(\theta_0)}{I_1(\theta_1)}} + \sqrt{n}(\theta_0 - \theta_1)\sqrt{\frac{I_1(\theta_0)}{I_1(\theta_1)}} \right) + P_{\theta_1} \left( \frac{s(\theta_1)}{\sqrt{nI_1(\theta_1)}} > \frac{\alpha}{2}\sqrt{\frac{I_1(\theta_0)}{I_1(\theta_1)}} + \sqrt{n}(\theta_0 - \theta_1)\sqrt{\frac{I_1(\theta_0)}{I_1(\theta_1)}} \right)
\]

\[
= \Phi \left( \Delta - \frac{\alpha}{2}\sqrt{\frac{I_1(\theta_0)}{I_1(\theta_1)}} \right) + 1 - \Phi \left( \Delta + \frac{\alpha}{2}\sqrt{\frac{I_1(\theta_0)}{I_1(\theta_1)}} \right),
\]

where \( \Delta = \sqrt{nI_1(\theta_1)}(\theta_0 - \theta_1) \). The last approximation is due to \( \frac{s(\theta_1)}{\sqrt{nI_1(\theta_1)}} \xrightarrow{d} N(0, 1) \).

Note that:

1. If the difference between \( \theta_0 \) and \( \theta_1 \) is very small the power will tend to \( \alpha \), i.e. if \( \Delta \approx 0 \) and \( I_1(\theta_1) \approx I_1(\theta_0) \), then the test will have trivial power.
2. As \( n \to \infty \), the two \( \Phi \) terms will approach either 0 or 1, and so the power will approach 1.
3. As a rule of thumb the test will have non-trivial power if \( |\theta_0 - \theta_1| \gg \frac{1}{\sqrt{nI_1(\theta_1)}} \).