1. (20 points)
We take a sample of size 2 from a Poisson distribution with parameter $\theta$; the pmf of this distribution is $p(k; \theta) = \frac{e^{-\theta} \theta^k}{k!}$.
We test $H_0 : \theta = 1$ versus $H_1 : \theta = 2$, by accepting $H_1$ if

$$\frac{p(X_1; 2)p(X_2; 2)}{p(X_1; 1)p(X_2; 1)} > 2.$$ 

a. (10 pts) Find the size of this test.

$$\frac{e^{-X_1} X_1 e^{-X_2} X_2}{e^{-1} X_1 e^{-1} X_2} > 2 \iff e^{2} X_1 + X_2 > 2$$ 

$$\iff 2 X_1 + X_2 > 2e^2$$ 

$$\iff X_1 + X_2 > log_2 (2e^2) \approx 3.88,$$

so we accept $H_1$ if $X_1 + X_2 \geq 4$.

If $H_0 = 1$, $X_1 + X_2 \sim \text{Pois}(2)$ so

$$P(X_1 + X_2 \geq 4) = 1 - 0.143 = 0.857.$$ 

b. (10 pts) Find the power of this test when $\theta = 2$, i.e. when $H_1$ is true.

When $\theta = 2$, $X_1 + X_2 \sim \text{Pois}(4)$

So

$$P(X_1 + X_2 \geq 4) = 1 - P(X_1 + X_2 \leq 3) = 1 - 0.433 = 0.567$$.
2. (15 points) \(X_1, \ldots, X_{50}\) are a sample of size \(n = 50\) from \(f(x) = (\theta - 1)x^{\theta-2}\) for \(0 < x < 1\), zero otherwise.

   a. (5 pts) Find \(\mu = E[X_i]\); your answer will be in terms of \(\theta\).
   
   \[
   \begin{align*}
   E[X_i] & = \int_0^1 (\theta - 1) x^{\theta-1} \, dx \\
   & = \frac{(\theta - 1) x^{\theta}}{\theta} \bigg|_0^1 \\
   & = \frac{\theta - 1}{\theta} .
   \end{align*}
   \]

   b. (10 pts) Find an approximate 90% confidence interval for \(\theta\) in terms of the sample mean \(\bar{X}\) and sample variance \(S^2\).

   \text{Hint: First, find an approximate 90% confidence interval for } \mu, \text{ then use your answer to part (a) to find a confidence interval for } \theta.

   Confidence interval for \(\mu\):
   
   \[
   \bar{X} - 1.645 \frac{S^2}{\sqrt{50}} \leq \mu \leq \bar{X} + 1.645 \frac{S^2}{\sqrt{50}}
   \]

   \[
   \mu = 1 - \frac{1}{\theta}
   \]

   \[
   \begin{align*}
   P \left( \bar{X} - 1.645 \frac{S^2}{\sqrt{50}} \leq 1 - \frac{1}{\theta} \leq \bar{X} + 1.645 \frac{S^2}{\sqrt{50}} \right) & \approx 0.9 \\
   & = P \left( \bar{X} - 1.645 \frac{S^2}{\sqrt{50}} \leq \frac{1}{\theta} \leq \bar{X} + 1.645 \frac{S^2}{\sqrt{50}} \right) \\
   & = P \left( \frac{1}{\bar{X} - 1.645 \frac{S^2}{\sqrt{50}}} \leq \theta \leq \frac{1}{\bar{X} + 1.645 \frac{S^2}{\sqrt{50}}} \right)
   \end{align*}
   \]
3. (25 points) Let $X_1, \ldots, X_n$ be a sample from the $\Gamma(1, \beta)$ distribution. Let $Y_1, \ldots, Y_n$ be the order statistics of the sample.

a. (10 pts) Find the pdf of $Y_1$. Identify the distribution of $Y_1$.

*Hint/Note:* You should recognize the distribution if you get the right pdf.

$$ f_Y(y) = n \cdot (e^{-y/\beta})^{n-1} \cdot \frac{1}{\beta} \cdot e^{-y/\beta} = \frac{n}{\beta} \cdot e^{-n \cdot y/\beta} \quad 0 < y < \infty $$

$$ Y_1 \sim \Gamma(1, \frac{\beta}{n}) = \text{exp} \left( \frac{n}{\beta} \right) $$

b. (10 pts) What constant $c$ makes $cY_1$ an unbiased estimator for $\beta$.

$$ E(Y_1) = \frac{\beta}{n} \quad (\text{since it is } \Gamma(1, \frac{\beta}{n})) $$

$$ \therefore E[\ln Y_1] = \beta, \quad \text{take } c = n. $$

c. (5 pts) Suppose $n = 4$ and $X_1, \ldots, X_4$ are

$$ X_1 = 7.59 \quad X_2 = 25.48 \quad X_3 = 2.02 \quad X_4 = 1.67 $$

What is is your point estimate for $\beta$, based on your answer to part (b).

$$ \beta \approx n \cdot \frac{Y_1}{n} = 4 \cdot (1.67) $$

\[\text{Point estimate for } \beta.\]
4. (15 points)
Suppose $X$ has pdf $f(x) = e^{-x}$ for $0 < x < \infty$, zero otherwise. Find the probability that $X$ is a potential outlier of the distribution.

\[
Q_{1/3} = \int_0^{Q_1} e^{-x} \, dx = \frac{1}{4} \approx 1 - e^{-Q_1} = \frac{1}{4}
\]

\[
e^{-Q_1} = \frac{3}{4} \quad Q_1 = -\ln\left(\frac{3}{4}\right) = h\left(\frac{4}{3}\right)
\]

\[
Q_3 = \int_0^{Q_3} e^{-x} \, dx = \frac{3}{4} \approx e^{-Q_3} = \frac{1}{4} \quad Q_3 = h(4)
\]

\[\sim \]

\[h = \frac{3}{2}(Q_3 - Q_1) = \frac{3}{2} h(3)
\]

\[\sim \]

\[LF = Q_1 - h = h(\frac{4}{3}) - h\left(\frac{3^{3/2}}{2}\right) < 0, \quad P(X \leq LF) = 0
\]

\[UF = Q_3 + h = h(4) + \frac{3}{2} h(3)
\]

\[\therefore P(X \leq UF) = 1 - e^{-UF}
\]

\[= 1 - e^{-h(4) + 3 h(3)}
\]

\[= 1 - \frac{1}{4} \cdot 3^{-3/2}
\]

\[\therefore P(X \geq UF \text{ or } X \leq LF) = \frac{1}{4} \cdot 3^{-3/2}
\]
5. (25 points)

$X_1, \ldots, X_{243}$ are a sample of $n = 243$ points in $(0, 1)$. We want to check whether these came from the pdf $f(x) = 4x^3$, $0 < x < 3$, zero otherwise. We observe the number of points in the segments $S_1 = (0, 1/3)$, $S_2 = [1/3, 2/3)$, and $S_3 = [2/3, 1)$, and see the following:

<table>
<thead>
<tr>
<th>Segment</th>
<th>$S_1$</th>
<th>$S_2$</th>
<th>$S_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td># in segment</td>
<td>5</td>
<td>60</td>
<td>168</td>
</tr>
</tbody>
</table>

a. (15 pts) Perform the chi-square goodness of fit test. Does the data support rejecting $H_0$ in favor of $H_1$ at the 0.05 level of significance? How many degrees of freedom are involved in the test?

\[
p_1 = \int_0^{1/3} 4x^3 \, dx = \left[ x^4 \right]_0^{1/3} = \frac{1}{81} \quad p_2 = \int_{1/3}^{2/3} 4x^3 \, dx = \frac{15}{81} \quad p_3 = \int_{2/3}^1 4x^3 \, dx = \frac{65}{81}
\]

$p_1 n = 3, \quad p_2 n = 45, \quad p_3 n = 195$

We have:

\[
\frac{(5-3)^2}{3} + \frac{(60-45)^2}{45} + \frac{(168-195)^2}{195} = 10.072
\]

Under the null hypothesis, this should be $x^2(2)$. Since $10.072 > 5.991$, we reject $H_0$ and accept $H_1$.

b. (10 pts) Suppose the data instead was

<table>
<thead>
<tr>
<th>Segment</th>
<th>$S_1$</th>
<th>$S_2$</th>
<th>$S_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td># in segment</td>
<td>5</td>
<td>$44+b$</td>
<td>$196-b$</td>
</tr>
</tbody>
</table>

Again performing the chi-square test, for what values of $b$ would we reject $H_0$ in favor of $H_1$ at the 0.1 level of significance.

\[
\frac{(5-3)^2}{3} + \frac{(44 + b - 45)^2}{45} + \frac{(196 - b - 195)^2}{195} \geq 4.6
\]

\[
\Leftrightarrow 2 \left( \frac{(6-1)^2}{45} + \frac{1}{195} \right) > 3.27
\]

\[
\Leftrightarrow (6-1)^2 > 119.56
\]

\[
\Leftrightarrow 6 \geq 12 \quad \text{or} \quad 6 \leq -10
\]
1. (25 points) $X_1, \ldots, X_n$ are drawn from the geometric distribution with pmf $p(x) = (1 - \theta)^{x-1}\theta$, where $x = 1, 2, 3, \ldots$ with $p(x) = 0$ elsewhere.
   a. (15 pts) Find the mle for $\theta$.

   \[
   \ell(\hat{\theta}) = (1 - \theta)^{\sum x_i - n} \theta^n
   \]

   \[
   \ell(\theta) = (\sum x_i - n) \log (1 - \theta) + n \log \theta
   \]

   \[
   \ell'(\theta) = -\frac{\sum x_i - n}{1 - \theta} + \frac{n}{\theta}
   \]

   \[
   \sum x_i - n = \frac{n}{\theta} \Rightarrow \frac{\sum x_i}{n} - 1 = \frac{1 - \theta}{\theta}
   \]

   \[
   \Rightarrow \hat{\theta} = \frac{\sum x_i}{n}, \quad \hat{\theta} = \frac{1}{x}
   \]

   b. (10 pts) Find the mle for $\Pr(X_2 \leq 2)$.

   \[
   \Pr(X_2 \leq 2) = \Pr(X_2 = 1) + \Pr(X_2 = 2)
   \]

   \[
   = \theta + (1 - \theta) \cdot \theta = g(\theta)
   \]

   MLE of $g(\theta)$ is $g(\theta)$

   so mle is

   \[
   \frac{1}{X} + (1 - \frac{1}{X}) \cdot \frac{1}{X}.
   \]
2. (25 points) $X_1, \ldots, X_n$ are drawn from the distribution $f(x; \theta) = \frac{3\theta x^2}{(x+\theta)^2}$ for $0 < x < \infty$ and $0 < \theta < \infty$.

   a. (10 pts) Find the Fisher information $I(\theta)$.
   *Hint/Note: $E[(x+\theta)^k] = \frac{3}{(3-k)\theta^{k-3}}$ if $k < 3$.

   \[
   \log(f(x; \theta)) = \log(3) + 3\log(x) + 4\log(x+\theta)
   \]

   \[
   E[\theta] = \frac{3}{\theta} - \frac{4}{x+\theta} \quad \Rightarrow \quad E[\theta^2] = \frac{3}{\theta^2} - 4E[(x+\theta)^{-2}]
   \]

   \[
   = \frac{3}{\theta^2} - 4 \cdot \frac{3}{5\theta^2} = \frac{3}{5\theta^2}
   \]

   \[\]

   b. (15 pts) $Y = 2\bar{X}$ is an unbiased estimator for $\theta$. Find the efficiency of $Y$.

   *Warning: May be time consuming.*

   Note: $k(\theta) = E[Y] = \theta \quad \Rightarrow \quad k'(\theta) = 1$.

   $E[Y] = 2E[X] = E[X; X]$ so $E[X_i] = \frac{\theta}{2}$.

   \[
   \theta = E[(X_i+\theta)^2] = E[X_i^2 + 2X_i\theta + \theta^2] = E[X_i^2] + \theta^2 + \theta^2
   \]

   \[
   \Rightarrow \quad E[X_i^2] = \theta^2
   \]

   \[
   \Rightarrow \quad \text{Var}(X_i) = \theta^2 - \left(\frac{\theta}{2}\right)^2 = \frac{3}{4}\theta^2 \quad \text{and} \quad \text{Var}(Y) = 4\text{Var}(X) = \frac{3\theta^2}{n}
   \]

   \[
   \text{Rao-Cramér:} \quad \frac{k'(\theta)^2}{nI(\theta)} = \frac{1}{\frac{3}{5}\theta^2} = \frac{3}{5}\theta^2
   \]

   \[
   \Rightarrow \quad \text{Efficiency} = \frac{(\frac{3\theta^2}{5n})}{(\frac{3\theta^2}{n})} = \frac{1}{5}
   \]
3. (20 points)
   a. (10 pts) Explain how to generate a random variable $X$ with the distribution

   
   \[ f_X(x) = \frac{1}{2} \sin(x) \quad \text{for } 0 < x < \pi, \]

   zero elsewhere, from a uniform $(0, 1)$ random variable $U$.

   \[ F_X(x) = \int_0^x \frac{1}{2} \sin(t) \, dt = -\frac{1}{2} \cos(x) \bigg|_0^x = \frac{1}{2} - \frac{1}{2} \cos(x) \]

   \[ y = \frac{1}{2} - \frac{1}{2} \cos(x) \sim x = \arccos(2y - 1) \quad \arccos(1 - 2y) \]

   \[ X = \arccos(1 - 2U) \]

   \[ U \sim \text{unif}(0, 1) \]

   b. (10 pts) $X_1, \ldots, X_n$ are a sample from an unknown pdf $f_X(x)$. Explain, in words, how to use bootstrapping to generate a confidence interval for $\sigma^2 = \text{Var}(X_i)$. Be sure to include an explanation of the bootstrapping procedure and how to find a sample variance in your answer.

   - Generate new samples $X_1^*, \ldots, X_n^*$ by sampling from $X_1, \ldots, X_n$ uniformly at random.
   - Use these to generate sample variances
     \[ S^2_k = \frac{1}{n-1} \sum (X_i^* - \bar{X}^*)^2 \]
   - Take the middle $1-\alpha$ portion to form confidence interval
4. (30 points)
   a. (10 pts) $X_1, \ldots, X_n$ are Bernoulli random variables with parameter $\theta$ (so they are 0/1 valued, and 1 with probability $\theta$). We wish to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. We apply the Wald-type test. Supposing $\bar{X} = \frac{1}{n}$, and $n = 100$, would we accept or reject $H_0$ at the 0.95 approximate confidence level.
   Hint: Recall for this distribution $I(\theta) = \frac{1}{\theta(1-\theta)}$ and $\hat{\theta} = \bar{X}$.

   $\left( \sqrt{n} I(\theta) \right) \left( \hat{\theta} - \theta_0 \right) \sim \left( \frac{1}{\sqrt{\frac{1}{n} \frac{1}{\bar{X}(1-\bar{X})}} \left( \bar{X} - \frac{1}{3} \right)} \right)^2$

   $= \frac{3.704}{3.841}$

   So we would accept $H_0$ at approx. .95 confidence level.

   b. (10 pts) $X_1, \ldots, X_n$ have the $N(0, \theta)$ distribution. We wish to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Show that the likelihood ratio test depends only on $S = \sum \frac{x^2}{\theta_0}$. Give the null distribution of $S$. Explicitly state whether the test should be of the form 'Reject $H_0$ if $S \leq c_1$ or $S \geq c_2$' or 'Reject $H_0$ if $S \geq c_3$'.

   $L(\theta) = \left( \frac{1}{\sqrt{2\pi\theta}} \right)^n e^{-\frac{1}{2\theta} \sum x_i^2}$

   $L(\theta_0) = \left( \frac{1}{\sqrt{2\pi\theta_0}} \right)^n e^{-\frac{1}{2\theta_0} \sum x_i^2}$

   $L(\theta) = \frac{L(\theta_0)^n}{n^{\frac{n}{2}}} e^{-\frac{1}{2\theta} \sum x_i^2}$

   $S^{n/2} e^{-1/2 S}$, const.

   $S^{n/2} - 1/2 S$, depends on $S$

   Null dist of $S$ is $\chi^2(n)$

   Test: Reject $H_0$ if $S \leq c_1$ or $S \geq c_2$.

(Third part on next page)
c. (10 pts) Suppose $X_1 = -1$, $X_2 = 3$ and $X_3 = -2$ is a sample from an $N(0, \theta)$ distribution. Use the test you developed in (b) to test $H_0 : \theta = 1.7$ versus $H_1 : \theta \neq 1.7$. Would you accept or reject $H_0$ at the $\alpha = 0.9$ confidence level.

\[
S = \sum \frac{x_i^2}{\theta} = \frac{1 + 9 + 4}{1.7} = 8.25
\]

Since $S \sim \chi^2(3)$ if $H_0$ true we reject $H_0$ if

\[
S \geq 7.815 \quad \text{or} \quad S \leq 0.352.
\]

\[
S = 8.25 > 7.815 \quad \text{so we reject } H_0
\]