**1.** (*20 points*)
We take a sample of size 2 from a Poisson distribution with parameter $\theta$; the pmf of this distribution is $p(k;\theta) = \mathbb{P}_\theta(X = k) = \frac{e^{-\theta}\theta^k}{k!}$.
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.

Test is:
$$\frac{e^{-2}\cancel{2}^{X_1} e^{-2}\cancel{2}^{X_2}}{e^{-1}1^{X_1} e^{-1}1^{X_1}} > 2 \iff e^{-2} 2^{X_1+X_2} > 2$$
$$\iff 2^{X_1+X_2} > 2e^2$$
$$\iff X_1+X_2 > \log_2(2e^2) \approx 3.88,$$

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 $\mathbb{P}(X_1+X_2 \geq 4)$
$$= 1 - .857 = .143.$$
$$= \mathbb{P}(X_1+X_2 \leq 3)$$

**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 $\mathbb{P}(X_1+X_2 \geq 4) = 1 - \mathbb{P}(X_1+X_2 \leq 3)$
$$= 1 - .433 = .567$$

**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 $\mu = \mathbb{E}[X_i]$; your answer will be in terms of $\theta$.

$$\mathbb{E}[X_i] = \int_0^1 (\theta-1)\, x^{(\theta-1)}\, dx = \frac{(\theta-1)x^{(\theta)}}{\theta}\Bigg|_0^1 = \frac{\theta-1}{\theta} = \boxed{1 - \frac{1}{\theta}}$$

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

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

Confidence interval for $\mu$:

$$\mathbb{P}\left(\bar{X} - \frac{1.645\sqrt{S^2}}{\sqrt{50}} \le \mu \le \bar{X} + 1.645\sqrt{\frac{S^2}{50}}\right) \approx .9$$

$$\mu = 1 - \frac{1}{\theta}$$

$$\therefore \mathbb{P}\left(\bar{X} - 1.645\sqrt{\frac{S^2}{50}} \le 1 - \frac{1}{\theta} \le \bar{X} + 1.645\sqrt{\frac{S^2}{50}}\right) \approx .9$$

$$= \mathbb{P}\left(\bar{X} - 1.645\sqrt{\frac{S^2}{50}} - 1 \le -\frac{1}{\theta} \le \bar{X} + 1.645\sqrt{\frac{S^2}{50}} - 1\right)$$

$$= \mathbb{P}\left(1 - 1.645\sqrt{\frac{S^2}{50}} - \bar{X} \le \frac{1}{\theta} \le 1 - \bar{X} + 1.645\sqrt{\frac{S^2}{50}}\right)$$

$$= \mathbb{P}\left(\frac{1}{1 - \bar{X} + 1.645\sqrt{\frac{S^2}{50}}} \le \theta \le \frac{1}{1 - \bar{X} - 1.645\sqrt{\frac{S^2}{50}}}\right)$$

$$\sim \left(\frac{1}{1 - \bar{X} + 1.645\sqrt{\frac{S^2}{50}}},\ \frac{1}{1 - \bar{X} - 1.645\sqrt{\frac{S^2}{50}}}\right)$$

**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.

$f_X(x) = \frac{1}{\beta} e^{-x/\beta} \quad 0<x<\infty \qquad F_X(x) = 1 - e^{-x/\beta} \quad 0<x<\infty$

$f_{Y_1}(y) = n \cdot (e^{-x/\beta})^{n-1} \cdot \frac{1}{\beta} e^{-x/\beta} = \frac{n}{\beta} e^{-nx/\beta} \qquad 0<x<\infty$

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

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

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

$\therefore E[nY_1] = \beta$, take $c = n$.

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

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

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

$\beta \approx nY_1 = \underbrace{4 \cdot (1.67)}_{\text{Point estimate for } \beta.}$

**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.

$Q_1$: $\int_0^{Q_1} e^{-x}\,dx = \frac{1}{4} \Rightarrow 1 - e^{-Q_1} = \frac{1}{4}$

$e^{-Q_1} = \frac{3}{4} \qquad Q_1 = -\ln\left(\frac{3}{4}\right) = \ln\left(\frac{4}{3}\right)$

$Q_3$: $\int_0^{Q_3} e^{-x}\,dx = \frac{3}{4} \Rightarrow e^{-Q_3} = \frac{1}{4} \therefore Q_3 = \ln(4)$

$h = \frac{3}{2}(Q_3 - Q_1) = \frac{3}{2}\ln(3)$.

$LF = Q_1 - h = \ln\left(\frac{4}{3}\right) - \ln\left(3^{3/2}\right) < 0, \therefore P(X \le LF) = 0$.

$UF = Q_3 + h = \ln(4) + \frac{3}{2}\ln(3)$

$\therefore P(X \ge UF) = 1 - e^{-UF}$

$= 1 - e^{-\ln(4) - \frac{3}{2}\ln(3)}$

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

$\therefore P(X \ge UF \text{ or } X \le LF) = \boxed{\frac{1}{4}\cdot 3^{-3/2}}$

**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:

| Segment | $S_1$ | $S_2$ | $S_3$ |
|---|---|---|---|
| # in segment | 5 | 60 | 168 |

**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?

$$P_1 = \int_0^{1/3} 4x^3 = x^4\Big|_0^{1/3} = \frac{1}{81} \qquad P_2 = x^4\Big|_{1/3}^{2/3} = \frac{15}{81} \qquad P_3 = x^4\Big|_{2/3}^{1} = \frac{65}{81}$$

$$P_1 n = 3 \qquad P_2 n = 45 \qquad P_3 n = 195$$

We have:

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

Under the null hypothisis, this should be $\chi^2(2)$.

Since $10.072 > 5.991$ we reject $H_0$ and accept $H_1$

**b.** (*10 pts*) Suppose the data instead was

| Segment | $S_1$ | $S_2$ | $S_3$ |
|---|---|---|---|
| # in segment | 5 | $44 + b$ | $196 - b$ |

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.

We wish to test: When is

$$\frac{(5-3)^2}{3} + \frac{(44+b-45)^2}{45} + \frac{(196-b-195)^2}{195} > 4.6$$

i.e. $\iff (b-1)^2\left(\frac{1}{45} + \frac{1}{195}\right) > 3.27$

$\iff (b-1)^2 > 119.56$

$\iff b-1 > 10.934$ $\quad b > 11.934$

$b-1 < -10.934$ $\quad b < -9.934$

$$\boxed{\begin{array}{c} b \geq 12 \\ b \leq -10 \end{array}}$$

**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 $\theta$.

$$L(\theta) = (1-\theta)^{\sum x_i - n}\theta^n$$

$$\ell(\theta) = \left(\sum x_i - n\right)\log(1-\theta) + n\log(\theta)$$

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

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

$$\Rightarrow \frac{1}{\theta} = \frac{\sum x_i}{n} \quad \boxed{\hat{\theta} = \frac{1}{\bar{X}}}$$

**b.** (*10 pts*) Find the mle for $\mathbb{P}(X_2 \leq 2)$.

$$\mathbb{P}(X_2 \leq 2) = \mathbb{P}(X_2 = 1) + \mathbb{P}(X_2 = 2)$$

$$= \theta + (1-\theta)\cdot\theta = g(\theta)$$

MLE of $g(\theta)$ is $g(\hat{\theta})$

so mle is

$$\boxed{\frac{1}{\bar{X}} + \left(1 - \frac{1}{\bar{X}}\right)\frac{1}{\bar{X}}.}$$

**2.** (*25 points*) $X_1, \ldots, X_n$ are drawn from the distribution $f(x;\theta) = \frac{3\theta^3}{(x+\theta)^4}$. for $0 < x < \infty$ and $0 < \theta < \infty$.

**a.** (*10 pts*) Find the Fisher information $I(\theta)$.

*Hint/Note:* $\mathbb{E}[(x+\theta)^k] = \frac{3}{(3-k)\theta^{-k}}$ if $k < 3$.

$$\log(f(x;\theta)) = \log(3) + 3\log(\theta) - 4\log(x+\theta)$$

$$\partial' = \frac{3}{\theta} - \frac{4}{x+\theta}$$

$$\partial^2 = -\frac{3}{\theta^2} + \frac{4}{(x+\theta)^{-2}}$$

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

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

$$= \frac{3}{5\theta^2}$$

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

*Warning:* May be time consuming.

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

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

$$3\theta^2 = E[(X_i+\theta)^2] = E[X_i^2 + 2X_i\theta + \theta^2] = E[X_i^2] + \theta^2 + \theta^2$$

$$\therefore \quad E[X_i^2] = \theta^2,$$

so $\mathrm{Var}(X_i) = \theta^2 - \left(\frac{\theta}{2}\right)^2 = \frac{3}{4}\theta^2$ and $\mathrm{Var}(Y) = 4\mathrm{Var}(\bar{X})$

$$= \frac{3\theta^2}{n}$$

Rao-Cramer:

$$\frac{k'(\theta)^2}{nI(\theta)} = \frac{1}{n\frac{3}{5\theta^2}} = \frac{3}{n5}\theta^2$$

$$\therefore \text{Efficiency} = \frac{\left(\frac{3\theta^2}{5n}\right)}{\left(\frac{3\theta^2}{n}\right)} = \frac{1}{5}$$

**3.** (*20 points*)

**a.** (*10 pts*) Explain how to generate a random variable $X$ with the distribution

$$f_X(x) = \frac{1}{2}\sin(x) \qquad \text{for } 0 < x < \pi,$$

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

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

$$y = \frac{1}{2} - \frac{1}{2}\cos(x) \rightsquigarrow x = \arccos(1-2y)$$

$$\therefore \quad X = \arccos(1-2U)$$

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

**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 $X_1^* \cdots X_n^*$ by sampling from $X_1, \cdots X_n$ uniformly at random.
- Use these to generate sample variances $S_*^2 = \frac{1}{n-1}\sum (X_i^* - \bar{X}^*)^2$
- 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 $H_0 : \theta = \frac{1}{3}$, versus $H_1 : \theta \neq \theta_0$. We apply the Wald-type test. Supposing $\bar{X} = \frac{1}{4}$, 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 $\hat{\theta} = \bar{X}$.

$$\left\{\sqrt{n I(\hat{\theta})}\,(\hat{\theta} - \theta_0)\right\}^2 = \left\{\sqrt{100 \cdot \frac{1}{\bar{X}(1-\bar{X})}}\left(\bar{X} - \frac{1}{3}\right)\right\}^2$$

$$= 3.704 < 3.841$$

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 $H_1 : \theta \neq \theta$. Show that the likelihood ratio test depends only on $S = \sum \frac{x_i^2}{\theta_0}$. 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 $S \geq c_3$.

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

$$\ell(\theta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\theta) - \frac{1}{2\theta}\sum X_i^2$$

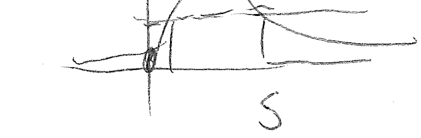
$$\ell'(\theta) = -\frac{n}{2\theta} + \frac{\sum X_i^2}{2\theta^2}$$

$$\Rightarrow \hat{\theta} = \frac{\sum X_i^2}{n}$$

$$\frac{L(\theta_0)}{L(\hat{\theta})} = \frac{\left(\frac{1}{\sqrt{2\pi\theta_0}}\right)^n e^{-\frac{1}{2}S}}{\left(\frac{1}{\sqrt{2\pi \frac{\sum_i X_i^2}{n}}}\right)^n e^{-n}}$$

$$= \underbrace{S^{n/2} e^{-\frac{1}{2}S}}_{\text{depends on } S} \cdot \text{Const.}$$

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




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.

$$S = \sum_i \frac{X_i^2}{\theta_0} = \frac{1+9+4}{1.7} = 8.25.$$

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

$$S \geq 7.815 \text{ or}$$
$$S \leq .352.$$

$S = 8.25 > 7.815$ so we reject $H_0$