Assignment 3

Monday, 3 October 2022 1:59 PM

Q1 $g_\lambda(y) = \begin{cases} \lambda y^{-2} & y \geq \lambda \\ 0 & \text{else} \end{cases}$

a. To show that it's a pdf, we need:

1) $\int_{-\infty}^{\infty} g_\lambda \, dy = 1$
2) $g_\lambda(y) \geq 0 \quad \forall y \geq \lambda, \ \lambda > 0$

For 1), we have that $\lambda > 0$, $y \geq \lambda$, so the lower bound of the integral will be λ

$$\Rightarrow \int_{-\infty}^{\infty} \lambda y^{-2} \, dy = \int_{\lambda}^{\infty} \lambda y^{-2} \, dy$$
$$= \lambda \left[-\frac{1}{y} \right]_{\lambda}^{\infty}$$
$$= -0 - -\frac{\lambda}{\lambda} = 1$$

as required.

For 2), note that $a^2 \geq 0 \ \forall a \in \mathbb{R} \Rightarrow 1/a^2 \geq 0$
$\Rightarrow y^{-2} \geq 0$

and that $\lambda > 0 \Rightarrow \lambda y^{-2} \geq 0 \ \forall y \geq \lambda$

$\therefore g_\lambda$ is a pdf.

b. Given $X = (X_1, \dots, X_n)$, the likelihood function is given by

$$\mathcal{L}(\lambda; x_1, \dots, x_n) = \prod_{i=1}^{n} g_\lambda(\lambda; x_i)$$
$$= \prod_{i=1}^{n} \lambda x_i^{-2}$$
$$= \lambda^n (x_1 \cdot \ldots \cdot x_n)^{-2}$$

c. To find the ML estimate, take the derivative of the loglikelihood function:

$$\frac{\partial \log \mathcal{L}}{\partial \lambda} = \frac{\partial}{\partial \lambda} \left(\log \left(\lambda^n (x_1 \cdot \ldots \cdot x_n)^{-2} \right) \right)$$
$$= \frac{\partial}{\partial \lambda} \left(n \log \lambda - 2 \log (x_1 \cdot \ldots \cdot x_n) \right)$$
$$= \frac{n}{\lambda}$$

We have that $0 < \lambda \leq y$, and so $\frac{\partial \log \mathcal{L}}{\partial \lambda}$ is minimised when $\hat{\lambda} = \min(x_1, \dots, x_n)$ (since λ is then maximised and so λ^{-1} is minimised).

The ML estimator is then $\hat{\lambda} = \min(X_1, \dots, X_n) = T(X)$

d. Since $T(X) = \min(X)$, if $T(X) > t$, then all elements of X are greater than t. Therefore we can take the probability of one sample and multiply it n times to obtain a probability that $T(X) > t$

For one x_i, the pdf is given by g_λ, and so multiply it n times to obtain a probability that $T(X) > t$

For one x_i, the pdf is given by g_λ, and so

$$\mathbb{P}_\lambda(T(X) > t) = 1 - \mathbb{P}_\lambda(T(X) \leq t)$$
$$= 1 - (\mathbb{P}_\lambda(x_i \leq t))^n$$
$$= 1 - \left(\int_{-\infty}^{t} g_\lambda(u)\,du\right)^n$$
$$= 1 - \left(\int_{-\infty}^{t} \lambda u^{-2}\,du\right)^n$$
$$= 1 - \left(\lambda \left[-u^{-1}\right]_{-\infty}^{t}\right)^n$$
$$= 1 - \lambda^n (-t)^{-n}$$

which is valid when $t \geq \lambda$, else $\mathbb{P}_\lambda(T(X) > t) = 1$ since $T(X) \geq 0$

e. $F_\lambda = \mathbb{P}(T(x) \leq t) = 1 - \mathbb{P}(T(x) > t)$

$$= 1 - (1 - \lambda^n (-t)^{-n})$$
$$= \lambda^n (-t)^{-n}$$

f. $f_\lambda = \frac{\partial}{\partial t} F_\lambda$

$$= -1 \cdot \lambda^n \cdot (-n) \cdot (-t)^{-n-1}$$
$$= n \lambda^n (-t)^{-(n+1)}$$

g. The bias of $T(X)$ is given by

$$\text{bias} = \mathbb{E}_\lambda(T(X) - \lambda)$$
$$= \mathbb{E}_\lambda(T(X)) - \lambda$$
$$= \int_{-\infty}^{\infty} t f_\lambda(t)\,dt - \lambda$$
$$= \lambda^n (n+1) \int_{-\infty}^{\infty} t(-t)^{-(n+1)}\,dt - \lambda$$
$$= \lambda^n (n+1) \left[\frac{t(-t)^{-n}}{n-1}\right]_{-\infty}^{\infty} - \lambda$$

(I'm so sorry about this question)

Q2

$\theta > 0$, $X_1, X_2, X_3 \overset{\text{iid}}{\sim} \text{Uniform}(\theta, 2\theta)$

a. We have that $\mathbb{E}(X) = \mu = \frac{(\theta + 2\theta)}{2}$ since the distribution is uniform. Hence

$$\frac{3\hat{\theta}}{2} = \mathbb{E}(\underline{X})$$
$$\Rightarrow \hat{\theta} = \frac{2}{3}\mathbb{E}(\underline{X})$$
$$= \frac{2}{3} \sum_{i=1}^{n} X_i / n$$

Which is the method of moments estimator of θ.

b. The pdf of the uniform distribution $U(a, b)$ is given as

$$f(x) = \frac{1}{b-a} \qquad x \in [a, b]$$

as

$$f(x) = \frac{1}{b-a} \qquad x \in [a,b]$$

therefore, for our berries, we have,

$$f(x) = \frac{1}{2\theta - \theta} = \frac{1}{\theta}$$

The likelihood function is then

$$\mathcal{L}(\underline{x};\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

$$= \left(\frac{1}{\theta}\right)^n = \frac{1}{\theta^n}$$

with log likelihood

$$\log \mathcal{L} = -n \log(\theta)$$

and score statistic:

$$\frac{\partial \log \mathcal{L}}{\partial \theta} = -\frac{n}{\theta}$$

This is negative for all values of positive θ, and so the largest θ minimises the score statistic

That is, the MLE is

$$\hat{\theta} = \frac{1}{2} \max(\underline{x})$$

(since the maximum gives 2θ, we have to divide by 2 to find θ).

$\hat{\theta}$ will always be smaller than θ due to the continuous nature of the variables, and so it is biased.

The expectation of $\hat{\theta}$ given θ is

$$E_\theta(\hat{\theta}) = E_\theta\left(\tfrac{1}{2} \max(\underline{x})\right)$$

$$= \frac{1}{2} E_\theta(\max(\underline{x}))$$

$$= \frac{1}{2}\int_\theta^{2\theta} P_\theta(\hat{\theta} \geq \underline{x})\,dx$$

$$= \frac{1}{2}\int_\theta^{2\theta} 1 - P(\hat{\theta} < \underline{x})\,dx$$

$$= \frac{1}{2}\int_\theta^{2\theta} 1 - (P(\hat{\theta} \leq x))^n dx \qquad \text{(due to cont. variables)}$$

$$= \frac{1}{2}\int_\theta^{2\theta} 1 - \left(\int_\theta^x f(u)\,du\right)^n dx \quad (\text{by cdf of } U(\theta, 2\theta))$$

$$= \frac{1}{2}\int_\theta^{2\theta} 1 - \left(\int_\theta^x \frac{1}{\theta}\,du\right)^n dx$$

$$= \frac{1}{2}\int_\theta^{2\theta} 1 - \left(\left[\frac{u}{\theta}\right]_\theta^x\right)^n dx$$

$$= \frac{1}{2}\int_\theta^{2\theta} 1 - \left(\frac{x-\theta}{\theta}\right)^n dx$$

$$= \frac{1}{2}\left[\frac{(\theta - x)(\frac{x}{\theta}-1)^n}{n+1} + x\right]_\theta^{2\theta}$$

$$= \frac{1}{2}\left(\theta - \frac{\theta(2-1)^n}{n+1} - \frac{\theta \cdot (1-1)^n}{n+1}\right)$$

$$= \frac{1}{2}\left(\theta - \frac{\theta}{n+1}\right)$$

$$= \frac{1}{2}\left(\frac{(n+1)\theta - \theta}{n+1}\right) = \frac{n\theta}{2(n+1)}$$

$\therefore$ setting $k = \frac{n+1}{n} = \frac{4}{3}$ (the inverse of the result above, divided by 2 to remove the persisting factor) will give the result

divided by 2 to remove the persisting factor)
will give the result

$$\mathbb{E}_\theta(k\hat{\theta}) = \theta$$

as required.

c. Method of moments:

$$\hat{\theta} = \frac{2}{3} \sum_{i=1}^{n} \frac{x_i}{n}$$

$$= \frac{2}{3} \cdot \left(\frac{1.2 + 1.9 + 1.1}{3} \right)$$

$$= 0.93\bar{3}$$

Maximum likelihood:

$$\hat{\theta} = \frac{1}{2} \max(\underline{x})$$

$$= \frac{1}{2} \max\{1.2, 1.9, 1.1\}$$

$$= \frac{1}{2} \cdot 1.9 = 0.95$$

d. Check audio file in submission.
tl;dr: both become more accurate with higher sample sizes. MoM particularly prone to error in low n limit.
MLE is biased, as shown before, but will converge to a more accurate value faster.
MLE will underestimate, MoM may overestimate

Q3

$$U_1, U_2, \dots, U_n \overset{\text{i.i.d.}}{\sim} U(0,1)$$
$$U_{(n)} = \max\{U_1, \dots, U_n\}$$

a. When $0 \le u \le 1$,

$$P(U_i \le u) = \int_{-\infty}^{u} f_U(x)\, dx$$

$$= \int_{-\infty}^{u} \frac{1}{1}\, dx \qquad \text{(by pdf of } U(0,1))$$

$$= \int_{0}^{u} 1\, dx \qquad \text{(by bounds of } U(0,1))$$

$$= u$$

However, since each $U_i \in \underline{U}$ has the same probability of being the maximum, $P(U_{(n)} \le u) = u^n$ as there are n chances for each r.v. to be the max.
Clearly, if $u < 0$, the probability of an $U \sim U(0,1)$ being smaller is just 0, and the opposite is true too. Hence,

$$P(U_{(n)} \le u) = \begin{cases} 0 & u < 0 \\ u^n & 0 \le u \le 1 \\ 1 & u > 1 \end{cases}$$

b. We want $P(a \le U_{(n)} \le b) = 0.9$

$$P(a \le U_{(n)} \le b) = P(U_{(n)} \le b) - P(U_{(n)} \le a)$$

Arbitrarily set $P(U_{(n)} \le b) = 0.99$

$$\Rightarrow P(U_{(n)} \le a) = 0.09$$

Arbitrarily set $\mathbb{P}(U_{(n)} \le b) = 0.99$
$\Rightarrow \mathbb{P}(U_{(n)} \le a) = 0.09$

Then, for $0 \le b \le 1$,

$$\mathbb{P}(U_{(n)} \le b) = b^n = 0.99$$
$$\Rightarrow b = (0.99)^{1/n}$$

Similarly obtain that $a = (0.09)^{1/n}$

c. Now, $X_1, X_2, \dots, X_n \overset{iid}{\sim} U(0, \theta)$, $\theta \ge 0$

The pdf of the uniform distribution $U(a,b)$ is given as

$$f(x) = \frac{1}{b-a} \qquad x \in [a,b]$$

therefore, for our data we have

$$f(x) = \frac{1}{\theta - 0} = \frac{1}{\theta}$$

The likelihood function is then

$$\mathcal{L}(\underline{x};\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$
$$= \left(\frac{1}{\theta}\right)^n = \frac{1}{\theta^n}$$

with log likelihood

$$\log \mathcal{L} = -n \log(\theta)$$

and score statistic:

$$\frac{\partial \log \mathcal{L}}{\partial \theta} = -\frac{n}{\theta}$$

This is negative for all values of positive θ, and so the largest θ minimises the score statistic.
That is, the MLE is

$$\hat{\theta} = \max\{\underline{X}\}$$
$$= \max\{X_1, \dots, X_n\}$$

d. We would expect that $X_{(n)}/\theta$ is a pivot variable as $X_{(n)}$ is the MLE of θ, and

$$X_{(n)} \sim U(0,\theta)$$
$$\Rightarrow \frac{X_{(n)}}{\theta} \sim U(0,1)$$

which does not depend on any unknown parameter.

e. We want a 90% confidence interval on θ from the data,

$$0.9 = \mathbb{P}\left(a \le \frac{X_{(n)}}{\theta} \le b\right)$$

since this is $\sim U(0,1)$, we can use a and b from part b), with $n = 8$ and $X_{(n)} = 2.356$,

$$0.9 = \mathbb{P}\left((0.09)^{1/8} \le \frac{2.3356}{\theta} \le (0.99)^{1/8}\right)$$
$$= \mathbb{P}\left(\frac{2.3356}{(0.99)^{1/8}} \le \theta \le \frac{2.3356}{(0.09)^{1/8}}\right)$$
$$= \mathbb{P}(2.3385 \le \theta \le 3.1559)$$

So our 90% confidence interval for θ is $(2.3385, 3.1559)$

$= \mathbb{P}(2.3385 \leq \theta \leq 3.1559)$

So our 90% confidence interval for θ is $(2.3385, 3.1559)$

f. Check audio file.
tldr: the student is wrong – the true value of θ is a definite number that may or may not be within the interval. We should say that we're 90% confident that the true θ is within the CI.

Q4

$$H_0 : p = 0.4$$

$$H_1 : p \neq 0.4 \quad \text{if} \quad X \in \{[0,15],[26,50]\}$$

a. In this case, the number of failures under the null hypothesis has distribution

$$X \sim Bin(n,p)$$
$$\sim Bin(50, 0.4)$$

(since $n=50$ and $p=0.4$ under the null)
So, the significance level is given by

$$\alpha = \mathbb{P}_{H_0}(X \in C)$$
$$= 1 - \mathbb{P}_{H_0}(15 \leq X \leq 26)$$
$$= 1 - \mathbb{P}_{H_0}(X \leq 26) - \mathbb{P}_{H_0}(X \leq 15)$$
$$= 1 - \left[\left(\sum_{i=0}^{26} \binom{n}{i} p^i (1-p)^{n-i}\right) - \left(\sum_{i=0}^{15} \binom{n}{i} p^i (1-p)^{n-i}\right)\right]$$

(with wolf.alpha.) $= 1 - (0.968594 - 0.0955017)$

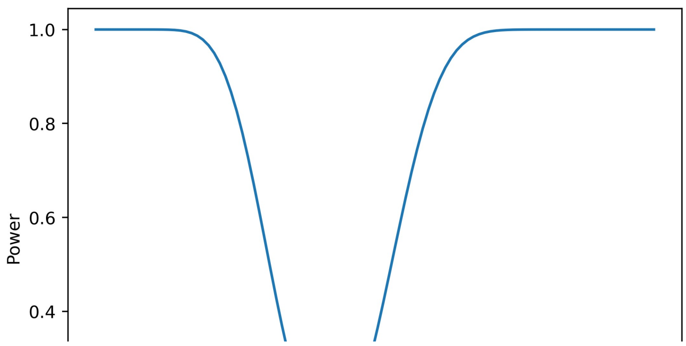
$$\simeq 0.127$$

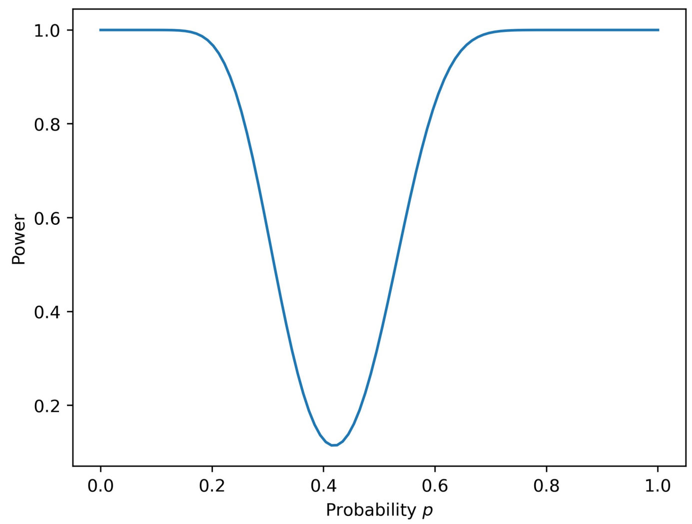
so this is about an 87% confidence interval.

b. The power of the test is given by

$$\text{Power} = 1 - \mathbb{P}_{H_1}(T(\underline{X}) \notin C)$$
$$= 1 - \mathbb{P}_{H_1}(15 \leq X \leq 26)$$
$$= 1 - (\mathbb{P}_{H_1}(X \leq 26) - \mathbb{P}(X \leq 15))$$

For varying p, $X \sim Bin(50, p)$. Using Python with scipy to calculate the power for varying p, we get the power curve:





The Python code used to generate this is:

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import binom

def binomCDF(p):
    n = 50
    return 1 - (binom.cdf(26, n, p) - binom.cdf(15, n, p))

x = np.linspace(0, 1, 100)
power = binomCDF(x)

fig, ax = plt.subplots()
ax.plot(x, power)
ax.set_xlabel('Probability $p$'); ax.set_ylabel('Power')

fig.savefig('PowerCurve.png', dpi=400, bbox_inches='tight')
fig.savefig('PowerCurve.pdf', dpi=400, bbox_inches='tight')
```

c. Since we're dealing with binomial data, the likelihood function is given by

$$\mathcal{L}(\underline{x};p) = \prod_{i=1}^{n} \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i}$$

$$= \left(\prod_{i=1}^{n}\binom{n}{x_i}\right) p^{\Sigma x_i}(1-p)^{n^2-\Sigma x_i}$$

Hence, a likelihood ratio test for two different probabilities, p_0 and p_1, would be

$$\frac{\mathcal{L}(\underline{x},p_0)}{\mathcal{L}(\underline{x},p_1)} = \frac{\left(\prod_{i=1}^{n}\binom{n}{x_i}\right) p_0^{\Sigma x_i}(1-p_0)^{n^2-\Sigma x_i}}{\left(\prod_{i=1}^{n}\binom{n}{x_i}\right) p_1^{\Sigma x_i}(1-p_1)^{n^2-\Sigma x_i}}$$

$$= \left(\frac{p_0}{p_1}\right)^{\Sigma x_i}\left(\frac{1-p_0}{1-p_1}\right)^{n^2-\Sigma x_i}$$

With the new model, we have $p_1 = 0.5$ vs $p_0 = 0.4$ from the old model. The LRT becomes $(n=50)$

$$\frac{\mathcal{L}(\underline{x},0.4)}{\mathcal{L}(\underline{x},0.5)} = \left(\frac{4}{5}\right)^{\Sigma x_i}\left(\frac{6}{5}\right)^{50^2-\Sigma x_i}$$

d. Notice that when $\Sigma x_i = x$ is large, the LRT is smaller. The LRT depends only on the value of x, and so if smaller. The LRT depends only on the value of x, and so if

$$\frac{\mathcal{L}(\underline{x}, 0.4)}{\mathcal{L}(\underline{x}, 0.5)} \leq d \qquad d \in \mathbb{R}^+$$

we would find that

$$X \geq c$$

for some $c \in \mathbb{R}^+$

e. We want that

$$\alpha = 0.01 = P(\text{reject } H_0 \mid H_0 \text{ true})$$
$$= P(X \geq c)$$

Therefore, we would want c to be the 99^{th} percentile of the binomial distribution with $n = 50$ and $p = 0.4$. That is,

$$0.99 = \sum_{i=0}^{c} \binom{n}{i} p^i (1-p)^{n-i}$$

Using the below Python code,

```
from scipy.stats import binom

i = 0
c = 0
while i < 0.99:
    i = binom.cdf(c, 50, 0.4)
    if i < 0.99:
        c += 1
print(c)
```

we obtain that $c = 28$ for a 1% significance.

f. The power under the alternative model is

$$\text{Power} = 1 - P_{H_1}(X \ngeq c)$$
$$= 1 - P_{H_1}(X < 28)$$
$$= 1 - P_{H_1}(X \leq 27)$$

Using the Python binomial cdf as above, we obtain that

$$\text{Power} \simeq 0.2399$$

g. Using the below python script, I could find the sample size, n, such that

$$\alpha \leq 0.01 \quad \text{and} \quad \text{power} \geq 0.9$$

```
from scipy.stats import binom
power = 0
n = 1
while power < 0.9:
    c = 0
    i = 0
    while i < 0.99:
        i = binom.cdf(c, n, 0.4)
        if i < 0.99:
            c += 1
    power = 1 - binom.cdf(c, n, 0.5)
    n += 1
print(power, " at n = ", n)
```

I obtain that $n = 323$ satisfies the above criteria.