1.

In maximum likelihood estimation, what exactly are you maximizing and why is it reasonable to use this maximum as an estimation of a parameter?

Thankfully, it's what it says on the tin. We're maximizing the likelihood function. This is equivalent to maximizing the joint density of a given sample. The parameters here function as random variables with the data fixed, so by maximizing the joint density what we're really doing is finding the estimators of them that make the data most likely. This is, in a non-technical way, why it's a reasonable approach. Given we have some data, what parameter values are the most likely to have produced it? That's all we're doing here.

2

Suppose $X_1, X_2, ..., X_n$is an iid random sample from the following probability density.

$$f_X(x \mid \lambda) = \lambda e^{-\lambda x}; \quad x \geq 0, \lambda > 0$$

Find the MLE of λ. You may assume $\sum x_i > 0$. Show that the MLE maximizes the likelihood function.

For starters we need the likelihood function, then the log likelihood for easier derivative computation. Our goal here is to take the log likelihood then take the derivative of it with respect to λ, and solve for the critical point. From there we'll use the second derivative to verify that it's the maximum. If it is, that critical point is our MLE for λ

$$\begin{aligned}
L(\lambda \mid \vec{x}) &= f(\vec{x} \mid \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x} && \text{(likelihood function)} \\
&= \lambda^n e^{-\lambda \sum x_i} \\
LL(\lambda \mid \vec{x}) &= \log\left(\lambda^n e^{-\lambda \sum x_i}\right) && \text{(log likelihood function)} \\
&= n \log(\lambda) - \lambda \sum x_i \\
\frac{dLL(\lambda \mid \vec{x})}{d\lambda} &= \frac{n}{\lambda} - \sum x_i && \text{(1st derivative)} \\
0 &= \frac{n}{\lambda} - \sum x_i \\
\lambda &= \frac{n}{\sum x_i} \\
\lambda &= \frac{1}{\bar{x}}
\end{aligned}$$

So our candidate for $\hat{\lambda} = \frac{1}{\bar{x}} = \bar{x}^{-1}$. Now we'll look at the second derivative.

$$\frac{d^2 LL(\lambda \mid \vec{x})}{d\lambda^2} = -\frac{n}{\lambda^2} < 0$$

This is strictly less than 0 as both n and λ are positive values. Therefore, $\hat{\lambda} = \bar{x}^{-1}$ is the MLE of λ.

3

Suppose $X_1, X_2, ..., X_n$is an iid random sample from the Uniform(a,b) distribution. That is, X_i has the following pdf for every i from 1 to n.

$$f_x(x \mid a, b) = \frac{1}{b-a}, \quad a \leq x \leq b, \;\; b > a$$

Find the MLE of a and b.

Same song and dance here, mostly. We'll start with the likelihood function.

$$L(a, b \mid \vec{x}) = \prod_{i=1}^{n} \frac{1}{b-1} I_{(a,b)}(x_i)$$

The indicator function here is the key. How it changes as we pass the product through it is what will provide us with our estimators. Let's think of how this works. We need all of the x_i values to fall between a and b. Any of them falling outside that range results in a likelihood of 0. So we need functions of our sample able to capture the low and high points of our bounds. That will of course be our min and max functions.

So, we have:

$$a \leq X_{(1)} < X_{(n)} \leq b$$

Which gives us,

$$L(a, b \mid \vec{x}) = \left(\frac{1}{b-a}\right)^n I_{(a \leq X_{(1)})} I_{(X_{(n)} \leq b)}$$

Maximizing the likelihood function requires that $(b-a)^{-n}$ is as large as possible which would, ideally, allow b and a to be as close as possible. So we want the smallest interval that captures the entire sample. That happens when, $\hat{a} = X_{(1)}$ and $\hat{b} = X_{(n)}$. Therefore, the MLE for $\vec{\theta}$ is:

$$\hat{a} = X_{(1)}, \;\; \hat{b} = X_{(n)}$$

7.6

Let $X_1, X_2, ..., X_n$be a random sample from the pdf

$$f(x \mid \theta) = \theta x^{-2}, \;\; 0 < \theta \leq x < \infty$$

A

What is a sufficient statistic for θ?

$$\begin{aligned} f(\vec{x} \mid \theta) &= \prod_{i=1}^{n} \theta x_i^{-2} I_{(\theta,\infty)}(x_i) \\ &= \theta^n \left(\prod_{i=1}^{n} x_i^{-2} \right) I_{(\theta,\infty)}(X_{(1)}) \end{aligned}$$

We have,

$$\begin{aligned} T(\vec{x}) &= X_{(1)} \\ g(T(\vec{x}) \mid \theta) &= \theta^n I_{(\theta,\infty)}(X_{(1)}) \\ h(\vec{x}) &= \prod_{i=1}^{n} x_i^{-2} \end{aligned}$$

Thus, by the factorization theorem, $X_{(1)}$ is a sufficient statistic for θ.

B

Find the MLE of θ.

We have the likelihood function

$$L(\theta \mid \vec{x}) = f(\vec{x} \mid \theta) = \theta^n \left(\prod_{i=1}^{n} x_i^{-2} \right) I_{(\theta,\infty)}(X_{(1)})$$

First, we note that $\prod x_i$ does not depend on θ at all. We can think of it as a constant here. θ^n is also increasing increasing in θ.

So, to maximize the likelihood function with respect to theta, we need to maximize θ^n. However, θ is bound by $X_{(1)}$. From this, the maximum of the likelihood function happens when $\hat{\theta} = X_{(1)}$. Therefore, $\hat{\theta} = X_{(1)}$ is the MLE of θ.

C

Find the method of moments estimator of θ.

We have one parameter, so,

$$m_1 = \frac{1}{n} \sum x_i = \bar{X} \equiv E[X]$$

So,

$$\begin{aligned}
E[X] &= \int_{x=\theta}^{x=\infty} x\theta x^{-2} dx \\
&= \theta \int x^{-1} dx \\
&= \theta \left[log(x) \right]_{\theta}^{\infty} \\
&= \theta \left(\log(\infty) - log(\theta) \right) \\
&= \infty
\end{aligned}$$

As the expected value diverges, the method of moment estimator of θ does not exist.

7.10

The independent random variables $X_1, X_2, ..., X_n$ have the common distribution

$$P(X_i \leq x \mid \alpha, \beta) = \begin{cases} 0 & x < 0 \\ (x/\beta)^\alpha & 0 \leq x \leq \beta \\ 1 & x > \beta \end{cases}$$

where the parameters α, β are positive.

A

Find a two dimensional sufficient statistic for (α, β).

First we need the pdf of x. What we have been provided is the cdf.

$$\begin{aligned}
f_X(x \mid \alpha, \beta) &= \frac{d}{dx} F_X(x) \\
&= \frac{d}{dx} \left(\frac{x}{\beta}\right)^\alpha \\
&= \alpha \left(\frac{x}{\beta}\right)^{\alpha-1} \cdot \frac{d}{dx}\frac{x}{\beta} \\
&= \frac{\alpha}{\beta^\alpha} x^{\alpha-1} \cdot I_{(0,\beta)}(x) \\
f_X(\vec{x} \mid \alpha, \beta) &= \prod_{i=1}^{n} \frac{\alpha}{\beta^\alpha} x^{\alpha-1} \cdot I_{(0,\beta)}(x_i) \\
&= \left(\frac{\alpha}{\beta^\alpha}\right)^n \left(\prod_{i=1}^{n} x_i\right)^{\alpha-1} I_{(0 \leq X_{(1)})} I_{(X_{(n)} \leq \beta)}
\end{aligned}$$

So, we have

$$\begin{aligned}
T(\vec{x}) &= \left(\prod_{i=1}^{n} x_i, X_{(n)}\right) \\
g(T(\vec{x})) &= \left(\frac{\alpha}{\beta^\alpha}\right)^n \left(\prod_{i=1}^{n} x_i\right)^{\alpha-1} I_{(X_{(n)} \leq \beta)} \\
h(\vec{x}) &= I_{(0 \leq X_{(1)})}
\end{aligned}$$

By the factorization theorem $T(\vec{x})$ is a sufficient statistic for $\vec{\theta}$.

B

Find the MLEs of α and β.

We start with the likelihood and log likelihood as always.

$$L(\theta \mid \vec{x}) = \left(\frac{\alpha}{\beta^\alpha}\right)^n \left(\prod_{i=1}^n x_i\right)^{\alpha-1} I_{(0 \le X_{(1)})} I_{(X_{(n)} \le \beta)}$$

$$LL(\theta \mid \vec{x}) = n \log\left(\frac{\alpha}{\beta^\alpha}\right) + (\alpha - 1)\log\left(\prod_{i=1}^n x_i\right) + \log(I(.))$$

$$= n\log(\alpha) - n\alpha\log(\beta) + (\alpha-1)\log\left(\prod x_i\right) + \log(I(.))$$

Where $I(.) = I_{(0 \le X_{(1)})} I_{(X_{(n)} \le \beta)}$

From here we will need a system of equations due to having two parameters to estimate. We first turn our attention to β as we can logically derive its MLE. We can see from the likelihood function that maximizing it with respect to β involves maximizing α/β^α. This is done by making β as small as possible. However, its lower bound is dictated by $X_{(n)}$. The max of the likelihood then with respect to β occurs when $\hat{\beta} = X_{(n)}$.

Thus, $X_{(n)}$ is the MLE of β.

From here we can now work on finding $\hat{\alpha}$. For this we will use $\hat{\beta} = X_{(n)}$.

$$\begin{aligned}
\frac{d}{d\alpha} LL(\theta \mid \vec{x}) &= \frac{n}{\alpha} - n\log(\beta) + \log\left(\prod x_i\right) = 0 \\
0 &= \frac{n}{\alpha} - n\log(X_{(n)}) + \log\left(\prod x_i\right) \\
-\frac{n}{\alpha} &= -n\log(X_{(n)}) + \log\left(\prod x_i\right) \\
-n &= \alpha(-n\log(X_{(n)}) + \log\left(\prod x_i\right)) \\
\hat{\alpha} &= -\frac{n}{n\log(X_{(n)}) + \log\left(\prod x_i\right)}
\end{aligned}$$

Now for the second derivative.

$$\begin{aligned}
\frac{d^2}{d\alpha^2} LL(\theta \mid \vec{x}) &= \frac{d}{d\alpha}\frac{n}{\alpha} - n\log(\beta) + \log\left(\prod x_i\right) \\
&= -\frac{n}{\alpha^2}
\end{aligned}$$

Since $n, \alpha > 0$, $-n/\alpha^2 < 0$ in α. Therefore, $\hat{\alpha}$ is the MLE for α.

7.12

Let $X_1, X_2, ..., X_n$be a random sample from a population with pmf

$$P_\theta(X = x) = \theta^x (1-\theta)^{1-x}, \quad x \in \{0, 1\},\ 0 \leq \theta \leq 1/2$$

A

Find the method of moments and MLE of θ

For method of moments, we notice that $x_i \sim Bern(\theta)$. Therefore

$$E[X_i] = \theta \equiv \frac{1}{n} \sum x_i$$

This one just kind of falls straight out of the setup really. So $\hat{\theta}_{\text{MOM}} = \frac{1}{n} \sum x_i$ For MLE, things get a bit more interesting because of the bound on θ. I'll be borrowing heavily from example 7.2.7 in the book.

$$\begin{aligned} L(\theta \mid \vec{x}) &= \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} \\ &= \theta^{\sum x_i} (1-\theta)^{n - \sum x_i} \end{aligned}$$

Let $y = \sum x_i$.

$$\begin{aligned} LL(\theta \mid \vec{x}) &= y \log(\theta) + (n-y) \log(1-\theta) \\ \frac{d}{d\theta} LL(\theta \mid \vec{x}) &= \frac{y}{\theta} - \frac{n-y}{1-\theta} = 0 \\ \frac{y}{\theta} &= \frac{n-y}{1-\theta} \\ \frac{y}{\theta}\theta(1-\theta) &= \frac{n-y}{1-\theta}\theta(1-\theta) \\ y(1-\theta) &= (n-y)\theta \\ y - y\theta &= n\theta - y\theta \\ y &= n\theta \\ \frac{y}{n} &= \theta \end{aligned}$$

So, our candidate for the MLE is $\hat{\theta} = y/n = \frac{1}{n} \sum x_i = \bar{x}$. Or, at least you would think. Important to note here is that we aren't done. We haven't accounted for the bounds on θ. The textbook example only looks at the case when $0 \leq \theta \leq 1$. Here, $0 \leq \theta \leq 1/2$.

If we are to use $\bar{x}$ as an estimator for θ, it too must be bounded. If $\bar{x}$ goes above 1/2 for instance, it ceases to be an estimator we can use. So we have two cases for our $\hat{\theta}$, one where $\bar{x}$ falls within the proper range and it is the MLE, and one where it exceeds it. In that case, the MLE of θ will simply be the max of the possible range, or 1/2.

Therefore, what we have as our MLE is the minimum of the two estimators, or $\text{Min}(\bar{x}, \frac{1}{2})$.

B

Find the MSE of each of the estimators.

We'll start with $\hat{\theta}_{\text{MOM}}$. This one is straightforward as this estimator is unbiased. The expected value is simply the parameter, so we can simply the MSE calculation down.

$$\begin{aligned}
E_\theta[\hat{\theta} - \theta]^2 &= Var_\theta(\hat{\theta}) \\
&= Var_\theta(\bar{x}) \\
&= \frac{Var(x_1)}{n} && \text{(Thm 5.2.6)} \\
&= \frac{\theta(1-\theta)}{n}
\end{aligned}$$

This next one is kind of hairy both because it's biased and because there are two possible estimators here. We either use $\bar{x}$ or 1/2. So to solve this we'll actually tackle the expected value directly and see what happens.

$$E_\theta[\hat{\theta} - \theta]^2 = \sum_{y=0}^{y=n} (\hat{\theta} - \theta)^2 \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

So this sum doesn't quite tell the whole story. As mentioned we need to address our two estimators and when they would be used. Our two cases are when:

- $\bar{x} \leq 1/2$ which occurs when $y \leq n/2$
- $\bar{x} > 1/2$ which occurs when $y > n/2$ or, equivalently, $y \geq (n/2) + 1$

So we need to bisect our sum to handle both cases and substitute in our values of $\hat{\theta}$.

$$E_\theta[\hat{\theta} - \theta]^2 = \sum_{y=0}^{y=n/2} (\bar{x} - \theta)^2 \binom{n}{y} \theta^y (1-\theta)^{n-y} + \sum_{y=(n/2)+1}^{y=n} \left(\frac{1}{2} - \theta\right)^2 \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

I uh, don't know how to simplify this down further so I'll be leaving this answer as is.

C

Which estimator is preferred, justify your choice.

We have to wrestle between a few things. Our MOM estimator is clean, simple and unbiased. It's easy to calculate and easy to work with. However, it doesn't respect the bounds of the parameter. We can get sample means that go beyond 1/2 which are entirely invalid.

The MLE estimator however is more complex, but it fully respects the bounds of the paramter. I haven't directly compared the MSEs of the two estimators but I believe the MLE properly accounting for the bounds is too important to pass up. As such I would choose to go with the MLE over the MOM.