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.