Question 1: Chapter 9 # 2

Let $X_1, \dots, X_n \sim \text{Uniform}(a, b)$ where a and b are unknown parameters and $a < b$.

(a) Find the method of moments estimators for a and b.

(b) Find the MLE $\hat{a}$ and $\hat{b}$.

(c) Let $\tau = \int x dF(x)$. Find the MLE of τ.

(d) Let $\hat{\tau}$ be the MLE of τ. Let $\tilde{\tau}$ be the non-parametric plug-in estimator of $\tau = \int x dF(x)$. Suppose that $a = 1$, $b = 3$, and $n = 10$. Find the mean squared error (MSE) of $\hat{\tau}$ by simulation. Find the MSE of $\tilde{\tau}$ analytically. Compare the results.

(a) MoM

Because $X_1, \dots, X_n \sim \text{Uniform}(a, b)$:

- The first moment of X_i is
$$\int_a^b x f_X(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}$$
- The second moment of X_i is
$$\int_a^b x^2 f_X(x) dx = \int_a^b \frac{x^2}{b-a} dx = \frac{x^3}{3(b-a)}\bigg|_a^b = \frac{b^3 - a^3}{3(b-a)} \tag{1}$$
$$= \frac{b^2 + ab + a^2}{3} \tag{2}$$

Let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ be the sample mean of the data. The sample second moment is $\frac{1}{n}\sum_{i=1}^n X_i^2$

We can set up the following systems of equations:

$$\frac{1}{n}\sum_{i=1}^n X_i = \frac{a+b}{2} \tag{3}$$
$$\frac{1}{n}\sum_{i=1}^n X_i^2 = \frac{b^2 + ab + a^2}{3} \tag{4}$$

Solving the first equation in terms of b will give us something we can plug into the second equation.

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{a+b}{2} \tag{5}$$

$$b = 2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - a \tag{6}$$

Using this in the second equation we get:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \frac{\left[2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - a\right]^2 + a\left[2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - a\right] + a^2}{3} \tag{7}$$

This eventually simplifies to

$$\hat{a} = \frac{1}{n}\sum_{i=1}^{n} X_i - \sqrt{3}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 - \frac{1}{n}\sum_{i=1}^{n} X_i^2\right]$$

$$\hat{b} = \frac{1}{n}\sum_{i=1}^{n} X_i + \sqrt{3}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 - \frac{1}{n}\sum_{i=1}^{n} X_i^2\right]$$

$\square$

(b) MLE

Note that the density of a uniform random variable is

$$f_X(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & x \notin [a,b] \end{cases}$$

We can proceed using the likelihood function and try to maximize it.

$$\begin{align} L(\theta \mid X_1, X_2, \ldots, X_n) &= \prod_{i=1}^{n} f(X_i; (a,b)) \tag{8} \\ &= \prod_{i=1}^{n} \begin{cases} 1 & \text{if} \quad X_i \in [a,b] \\ 0 & \text{if} \quad X_i \notin [a,b] \end{cases} \cdot \frac{1}{b-a} \tag{9} \\ &= \begin{cases} 1 & \text{if} \quad X_i \in [a,b] \\ 0 & \text{if} \quad X_i \notin [a,b] \end{cases} \cdot \frac{1}{(b-a)^n} \tag{10} \\ & \tag{11} \end{align}$$

We can see that if $X_i < a$ then the piece-wise term will be 0. As the likelihood is a product of this term and the $\frac{1}{(b-a)^n}$ term, the whole thing will be 0 if any of the X_i's are less than a. By that logic, it is clear that a lower bound for a to maximize this function is that $a \leq \min(X_1, \ldots, X_n)$.

Similarly, if $X_i > b$ then the piece-wise term will be 0, resulting in the likelihood function being 0. Thus, quite similarly, we have an upper bound on what b can be, namely that b must be such that $b \geq \max(X_1, \ldots, X_n)$

Now that we have a and b bounded, we can be assured that this product will now be 0. So we can now analyze $\frac{1}{(b-a)^n}$ to maximize a and b. We can simplify this by taking the log likelihood as follows:

$$\log L(a,b) = \frac{1}{(b-a)^n} = -n \log(b-a)$$

To maximize this function, we want to minimize $b - a$ as the log likelihood is strictly decreasing(as long as n is positive, which it is as it is the number of samples). Therefore, we want the maximum possible a and the minimum possible b. The minimum value b can take (without making the product 0) is $\max(X_1, \ldots, X_n)$ and similarly, the maximum value a can take is $\min(X_1, \ldots, X_n)$ so,

$$\hat{a} = \min(X_1, \ldots, X_n), \quad \hat{b} = \max(X_1, \ldots, X_n)$$

$\square$

(c) Tau

$$\tau = \int x dF(x) = \mathbb{E}(X)$$

τ is the $\mathbb{E}(X)$ and because each X_i is uniform with parameters a and b, $\mathbb{E}(X) = \frac{a+b}{2}$

$\hat{\tau}$ is the estimator that we get from the MLE. τ is in terms of a and b, so we can express $\hat{\tau}$ similarly in terms of $\hat{a}$ and $\hat{b}$:

$$\hat{\tau} = \frac{\hat{b} - \hat{a}}{2} = \frac{\max(X_1, \ldots, X_n) - \min(X_1, \ldots, X_n)}{2}$$

(d) Simulation and SE of Tau

Let $\hat{\tau}$ be the MLE of τ. Let $\tilde{\tau}$ be the non-parametric plug-in estimator of $\tau = \int x dF(x)$. Suppose that $a = 1$, $b = 3$, and $n = 10$. Find the mean squared error (MSE) of $\hat{\tau}$ by simulation. Find the MSE of $\tilde{\tau}$ analytically. Compare the results.

Part 1

```
import numpy as np

a = 1
b = 3
n = 10

X = np.random.uniform(low=a, high=b, size=n)
tau_hat = (X.min() + X.max()) / 2

B = 10000
```

```python
e = np.empty(B)
for i in range(B):
    x = np.random.choice(X, n, replace=True)
    e[i] = (x.min() + x.max()) / 2

se = e.std()
print("MSE for tau_hat:", se)
```

Estimated MSE of $\hat{\tau}$: 0.1536114415970741

Part 2

$\tilde{\tau}$ is the non-parametric plug-in estimator for τ:

$$\tilde{\tau} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Because $\tilde{\tau}$ is unbiased (non-parametric plug-in estimator is unbiased), the MSE of $\tilde{\tau}$ is:

$$MSE(\tilde{\tau}) = Var(\tilde{\tau}) + 0 = Var(\tilde{\tau}) \tag{12}$$

$$= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \tag{13}$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \tag{14}$$

$$= \frac{(b-a)^2}{12n} \tag{15}$$

In this case, we have $b = 3$ and $a = 1$ so:

$$MSE(\tilde{\tau}) = \frac{(2)^2}{12n} \tag{16}$$

$$= \frac{4}{12n} \tag{17}$$

$$= \frac{1}{3n} \tag{18}$$

Thus, as $n = 10$ in this case:

$$MSE(\tilde{\tau}) = \frac{1}{30}$$

Thus

$$SE(\tilde{\tau}) = \sqrt{\frac{1}{30}} \approx 1.825$$

Comparison

From simulation, we got a standard error of 0.153 and with the non-parametric plug in estimator, we calculated 0.1825. The simulation is slightly off the observed, which is expected as there is varience in the simulation

Question 2: Chapter 9 # 4

Let $X_1, \ldots, X_n \sim \text{Uniform}(0, \theta)$. Show that the MLE is consistent.

Hint: Let $Y = \max\{X_1, \ldots, X_n\}$. For any c,

$$P(Y < c) = P(X_1 < c, X_2 < c, \ldots, X_n < c) = P(X_1 < c)P(X_2 < c) \ldots P(X_n < c).$$

****Proof****

The density of a uniform random variable is

$$f_X(x) = \begin{cases} \frac{1}{\theta} & x \in [0, \theta] \\ 0 & x \notin [0, \theta] \end{cases}$$

The MLE is

$$L(\theta \mid X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} f(X_i; (\theta)) \tag{19}$$

$$= \prod_{i=1}^{n} \begin{cases} 1 & \text{if } \; X_i \in [0, \theta] \\ 0 & \text{if } \; X_i \notin [0, \theta] \end{cases} \cdot \frac{1}{\theta} \tag{20}$$

$$= \begin{cases} 1 & \text{if } \; X_i \in [0, \theta] \\ 0 & \text{if } \; X_i \notin [0, \theta] \end{cases} \cdot \frac{1}{\theta^n} \tag{21}$$

Analysing this similarly to question 1, in order for this product to not be 0, $\hat{\theta} \geq \max(X_i, \ldots, X_n)$. We want to make this term, $\frac{1}{\theta^n}$ as small as possible, so we must minimize θ, so we take $\hat{\theta} = max(X_i, \ldots, X_n)$ similar to how we did in question 1 for a and b.

From here we can examine the asymptotic behavior of $\hat{\theta}$. From the hint, we know that, because $\hat{\theta} = max(X_i, \ldots, X_n)$

$$P(\hat{\theta} < c) = P(X_1 < c)P(X_2 < c) \ldots P(X_n < c) \tag{22}$$

$$= \prod_{i=1}^{n} P(X_i < c) \tag{23}$$

If we take $c = \theta - \epsilon$ for some arbitrary $\epsilon > 0$.

$$P(\hat{\theta} < \theta - \epsilon)$$

This signifies, what's the probability that our estimator is less than the real thing for some arbitrarily small ϵ

$$P(\hat{\theta} < \theta - \epsilon) = \prod_{i=1}^{n} P(X_i < \theta - \epsilon) = \left(1 - \frac{\epsilon}{\theta}\right)^n \tag{24}$$

This goes to 0 as $n \to \infty$ as $1 - \frac{\epsilon}{\theta} < 1$, so:

$$\lim_{n\to\infty} P(\hat{\theta} < \theta - \epsilon) = 0$$

Next,

$$P(\hat{\theta} \geq \theta - \epsilon) = 1 - P(\hat{\theta} < \theta - \epsilon) = 1 - \left(1 - \frac{\epsilon}{\theta}\right)^n$$

goes to 1, as

$$\begin{aligned}
&\lim_{n\to\infty} 1 - \left(1 - \frac{\epsilon}{\theta}\right)^n && (25)\\
&= \lim_{n\to\infty} 1 - \lim_{n\to\infty} \left(1 - \frac{\epsilon}{\theta}\right)^n && (26)\\
&= 1 - \lim_{n\to\infty} \left(1 - \frac{\epsilon}{\theta}\right)^n && (27)\\
&= 1 && (28)
\end{aligned}$$

$P(\hat{\theta} < \theta - \epsilon)$ goes to 0 and $P(\hat{\theta} \geq \theta - \epsilon)$ goes to 1 as $n \to \infty$, so $P(|\hat{\theta} - \theta| < \epsilon)$ goes to 0 as $n \to \infty$. Therefore, $\hat{\theta}$ is a consistent estimator for θ. $\square$

Question 3: Chapter 9 # 6

Let $X_1, \ldots, X_n \sim N(\theta, 1)$. Define

$$Y_i = \begin{cases} 1 & \text{if } X_i > 0 \\ 0 & \text{if } X_i \leq 0 \end{cases}.$$

Let $\psi = P(Y_1 = 1)$.

(a) Find the maximum likelihood estimator $\hat{\psi}$ of ψ.

(b) Find an approximate 95 percent confidence interval for ψ.

(c) Define $\bar{\psi} = \frac{1}{n} \sum Y_i$. Show that $\bar{\psi}$ is a consistent estimator of ψ.

From the definition of X and Y we have it that $Y_i, \ldots, Y_n \sim \text{Bernoulli}(\Phi(\theta))$ where Φ is the CDF for $N(\theta, 1)$

(a)

We have it that $\psi = P(Y_1 = 1) = \Phi(\theta)$ as the probability that any one Y_i is p where p the the probability associated with the bernoulli random variable

The MLE for ψ is $\hat{\psi}$:

$$L(\psi) = \prod_{i=1}^{n} \psi_i^Y (1-\psi)^{(1-Y_i)} \tag{29}$$

$$l(\psi) = \log \prod_{i=1}^{n} \psi_i^Y (1-\psi)^{(1-Y_i)} \tag{30}$$

$$= \sum_{i=1}^{n} \log \psi_i^Y (1-\psi)^{(1-Y_i)} \tag{31}$$

$$= \sum_{i=1}^{n} Y_i \log \psi + (1 - Y_i) \log(1-\psi) \tag{32}$$

To maximize, we differentiate with respect to ψ and we get,

$$\frac{\sum_{i=1}^{n} Y_i}{\psi} - \frac{\sum_{i=1}^{n} (1 - Y_i)}{1-\psi} \tag{33}$$

We want this to be 0, so

$$\frac{\sum_{i=1}^{n} Y_i}{\psi} - \frac{\sum_{i=1}^{n}(1 - Y_i)}{1 - \psi} = 0$$

This simplifies to:

$$\frac{1}{\psi}\sum_{i=1}^{n} Y_i = \frac{1}{1-\psi}\sum_{i=1}^{n}(1 - Y_i)$$

Further simplifying to:

$$\psi = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

So,

$$\hat{\psi} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

(b)

Y is bernoullli so it converges to a normal distribution. Specifically, because:

$$\bar{Y} \sim N\left(\psi, \frac{\psi(1-\psi)}{n}\right)$$

Thus a confidence interval for ψ can be given as:

$$C_n = \left(\hat{\psi} - z_{0.975}\sqrt{\frac{\hat{\psi}(1-\hat{\psi})}{n}}, \hat{\psi} + z_{0.975}\sqrt{\frac{\hat{\psi}(1-\hat{\psi})}{n}},\right)$$

$$C_n = \left(\hat{\psi} - 1.96\sqrt{\frac{\hat{\psi}(1-\hat{\psi})}{n}}, \hat{\psi} + 1.96\sqrt{\frac{\hat{\psi}(1-\hat{\psi})}{n}},\right)$$

(c)

We can use Chebyshev's inequality as follows

$$P(|\bar{\psi} - E(\bar{\psi})| > \epsilon) \leq \frac{Var(\bar{\psi})}{\epsilon^2} \tag{34}$$

We know that $Var(Y_1) = \psi(1-\psi)$ becuase Y_i is Bernoulli with variable ψ. $Var(\bar{\psi})$:

$$Var(\bar{\psi}) = Var(\frac{1}{n}\sum Y_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(Y_i)) \tag{35}$$

$$= \frac{1}{n^2} n(\psi(1-\psi)) \tag{36}$$

$$= \frac{\psi(1-\psi)}{n} \tag{37}$$

Plugging this in we get:

$$P(|\bar{\psi} - \psi| > \epsilon) \leq \frac{\psi(1-\psi)}{n\epsilon^2} \tag{38}$$

Taking the limit of both sides we get:

$$\lim_{n\to\infty} P(|\bar{\psi} - \psi| > \epsilon) \leq \lim_{n\to\infty} \frac{\psi(1-\psi)}{n\epsilon^2} \tag{39}$$

$$\lim_{n\to\infty} P(|\bar{\psi} - \psi| > \epsilon) \leq 0 \tag{40}$$

$$\lim_{n\to\infty} P(|\bar{\psi} - \psi| > \epsilon) = 0 \tag{41}$$

Therefore, by the law of large numbers we say that $\bar{\psi} \xrightarrow{P} \psi$ and therefore is a consistent estimator for ψ

Question 4: Supplemental

Let $x_1, x_2, \ldots, x_n$ be i.i.d. samples from a random variable X that is Gaussian with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.

(a) Show the MLE estimators for μ, σ^2 are

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2.$$

(b) Show that the MLE estimator $\hat{\sigma}_n^2$ for σ^2 is such that $\mathbb{E}(\hat{\sigma}_n^2) = \frac{n-1}{n} \cdot \sigma^2$, and is thus biased.

(c) Verify (b) empirically as follows. Let $n = 3, 4, \ldots, 100$. For each n value, generate 100 i.i.d. samples of size n from $N(0,1)$. For each sample, compute the MLE estimate for σ^2; then, average across the 100 trials. This average may be thought of as an estimate for $\mathbb{E}(\hat{\sigma}_n^2)$. Plot your average MLE as a function of n and describe the behavior.

(a) MLE

Likelihood Function

The probability density function (pdf) of the normal distribution is:

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

$$L(\mu,\sigma^2) = \prod_{i=1}^{n} f(x_i;\mu,\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right).$$

Log-Likelihood Function

We can take the log of both sides to get the log-likelihood function which will simplify calculations:

$$\ell(\mu,\sigma^2) = \log L(\mu,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

MLE for μ

To find the MLE for μ, we differentiate the log-likelihood function with respect to μ and set the derivative to zero and we get:

$$\frac{d}{d\mu}\ell(\mu, \sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

This simplifies to:

$$\sum_{i=1}^{n}(x_i - \mu) = 0 \quad \Rightarrow \quad n\mu = \sum_{i=1}^{n} x_i \quad \Rightarrow \quad \hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Thus, the MLE for μ is the sample mean:

$$\hat{\mu}_n = \bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

MLE for σ^2

Next, we differentiate the log-likelihood function with respect to σ^2 and set the derivative equal to zero:

$$\frac{d}{d\sigma^2}\ell(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \mu)^2 = 0.$$

Multiplying by $2\sigma^4$ gives:

$$-n\sigma^2 + \sum_{i=1}^{n}(x_i - \mu)^2 = 0 \quad \Rightarrow \quad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu}_n)^2.$$

Substitute $\hat{\mu}_n$ for μ and expand the expression:

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j\right)^2.$$

This can be rewritten as:

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2.$$

Conclusion

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2.$$

$\square$

(b)

From the last part, we found that:

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$$

Which we can simplify as,

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

where the sample mean is:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

We want to show that $\hat{\sigma}_n^2$ is biased by calculating $\mathbb{E}[\hat{\sigma}_n^2]$.

First, rewrite $\hat{\sigma}_n^2$ as:

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2.$$

Now, compute the expected value:

$$\mathbb{E}[\hat{\sigma}_n^2] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2\right].$$

By linearity of expectation,

$$\mathbb{E}[\hat{\sigma}_n^2] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i^2) - \mathbb{E}(\bar{X}^2).$$

We compute each term separately:

1. Since $X_i \sim \mathcal{N}(\mu, \sigma^2)$, we have:

$$\mathbb{E}[X_i^2] = \mathrm{Var}(X_i) + (\mathbb{E}[X_i])^2 = \sigma^2 + \mu^2.$$

2. For $\mathbb{E}[\bar{X}^2]$, we use:

$$\mathbb{E}[\bar{X}^2] = \mathrm{Var}(\bar{X}) + (\mathbb{E}[\bar{X}])^2 = \frac{\sigma^2}{n} + \mu^2.$$

Now, substitute these results into the expectation of $\hat{\sigma}_n^2$:

$$\mathbb{E}[\hat{\sigma}_n^2] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i^2) - \mathbb{E}(\bar{X}^2) \tag{42}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sigma^2 + \mu^2 - \frac{\sigma^2}{n} + \mu^2 \tag{43}$$

$$= (\sigma^2 + \mu^2) - \left(\frac{\sigma^2}{n} + \mu^2\right) \tag{44}$$

$$= \sigma^2\left(1 - \frac{1}{n}\right) \tag{45}$$

Thus, we have:

$$\mathbb{E}[\hat{\sigma}_n^2] = \frac{n-1}{n} \cdot \sigma^2.$$

This shows that $\hat{\sigma}_n^2$ is a biased estimator of σ^2, and the bias is $\frac{n-1}{n}$.

(c)

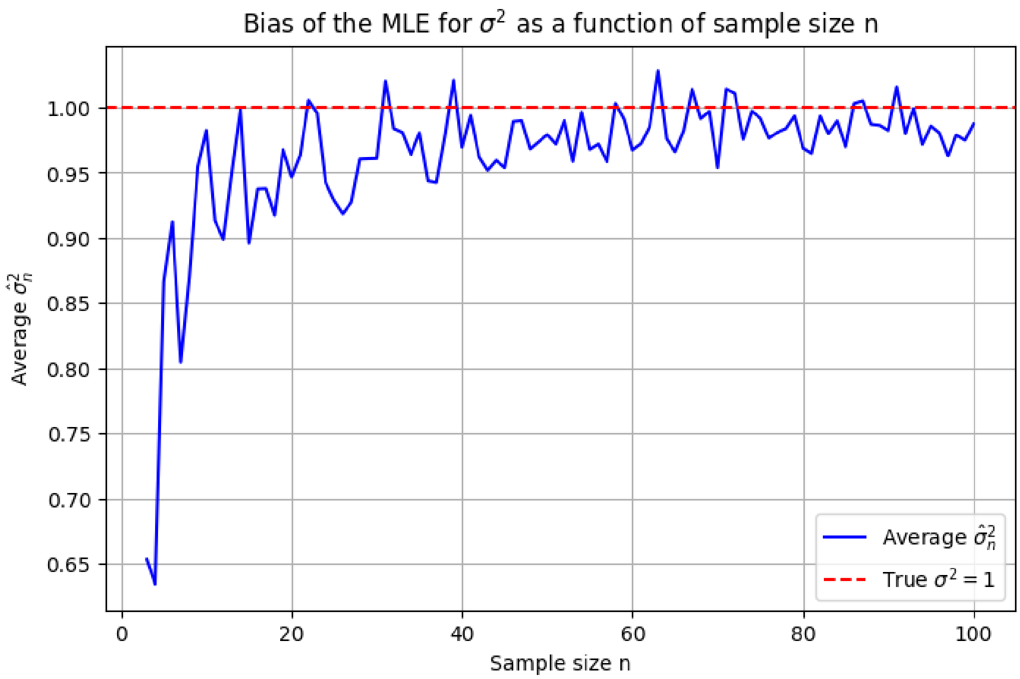

Figure 1: Bias of MLE as function of sample size

Initially for small n, the bias that we calculated in part b is very pronounced, as for small n, the MLE estimation for σ^2 is far off the true $\sigma^2 = 1$ that the samples are being generated from.

As n increases, we can see that the MLE for σ^2 converges to the true value. This is what we can expect because the MLE for σ^2 is consistent. As $n \to \infty$, we expect the deviations between the estimate and the real to eventually go to 0.

```
import numpy as np
import matplotlib.pyplot as plt

# Number of trials for each sample size n
num_trials = 100

# Sample sizes ranging from n = 3 to n = 100
n_values = np.arange(3, 101)

# Store the average MLE of sigma^2 for each n
average_mle = []

# Loop over each sample size n
for n in n_values:
    mle_values = []

    # Perform 100 trials for each n
    for _ in range(num_trials):
        # Generate n samples from N(0, 1)
        samples = np.random.normal(0, 1, n)

        # Compute the MLE of sigma^2
        mle_sigma2 = np.mean(samples**2)
            - np.mean(samples)**2
        mle_values.append(mle_sigma2)

    # Compute the average MLE for the current n
    average_mle.append(np.mean(mle_values))

# Plot the average MLE as a function of n
plt.figure(figsize=(8, 5))
plt.plot(n_values, average_mle,
    label=r'Average $\hat{\sigma}^2_n$', color='blue')
plt.axhline(y=1,
    color='red',
    linestyle='--',
    label=r'True $\sigma^2 = 1$')
```

```
plt.xlabel('Sample size n')
plt.ylabel(r'Average $\hat{\sigma}^2_n$')
plt.title('Bias of the MLE for $\\sigma^2$ as a function of sample size n')
plt.legend()
plt.grid(True)
plt.show()
```