BONTA AALAYA

12340530

DSL253 - Statistical Programming

1

INTRODUCTION

In this assignment, we observe the behaviour of battery life of a smartphone model which is said to follow a normal distribution, and use MLE to estimate μ and σ^2 for varying sample sizes

DATA

Generated using in-built matlab function <u>normrnd</u> and considering given parameters

METHODOLOGY

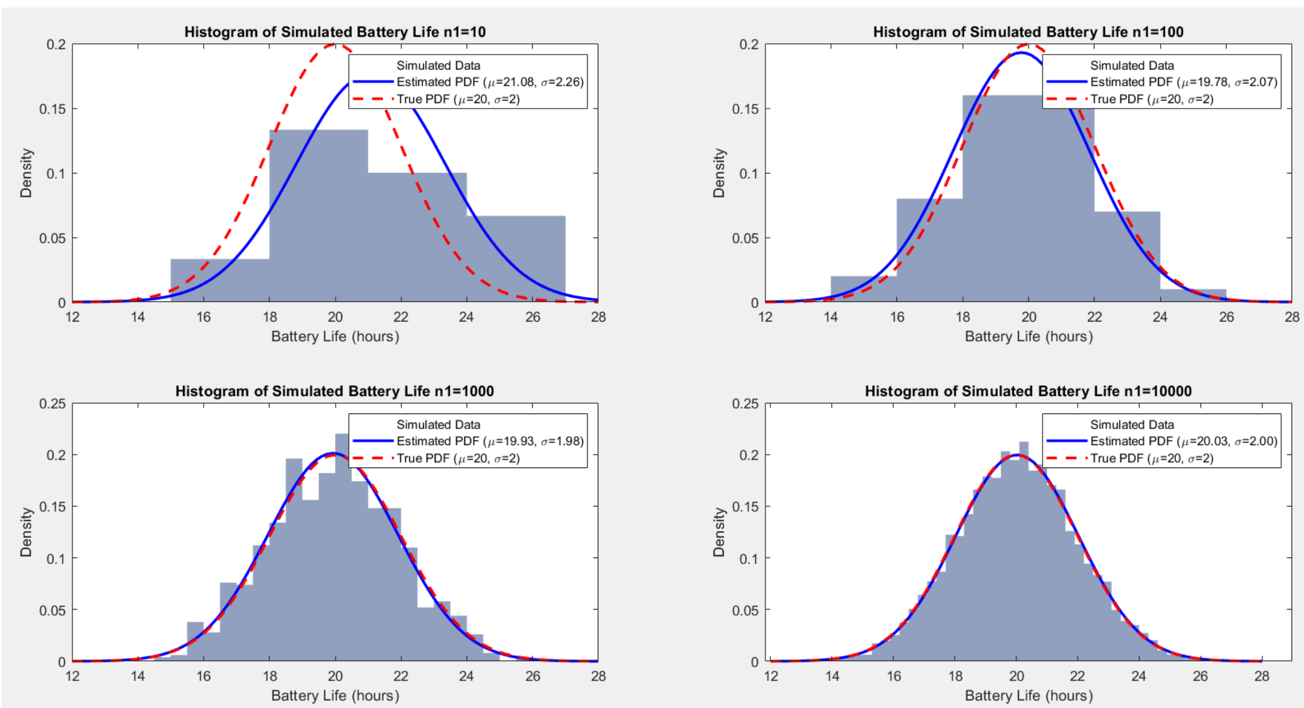
Question 1

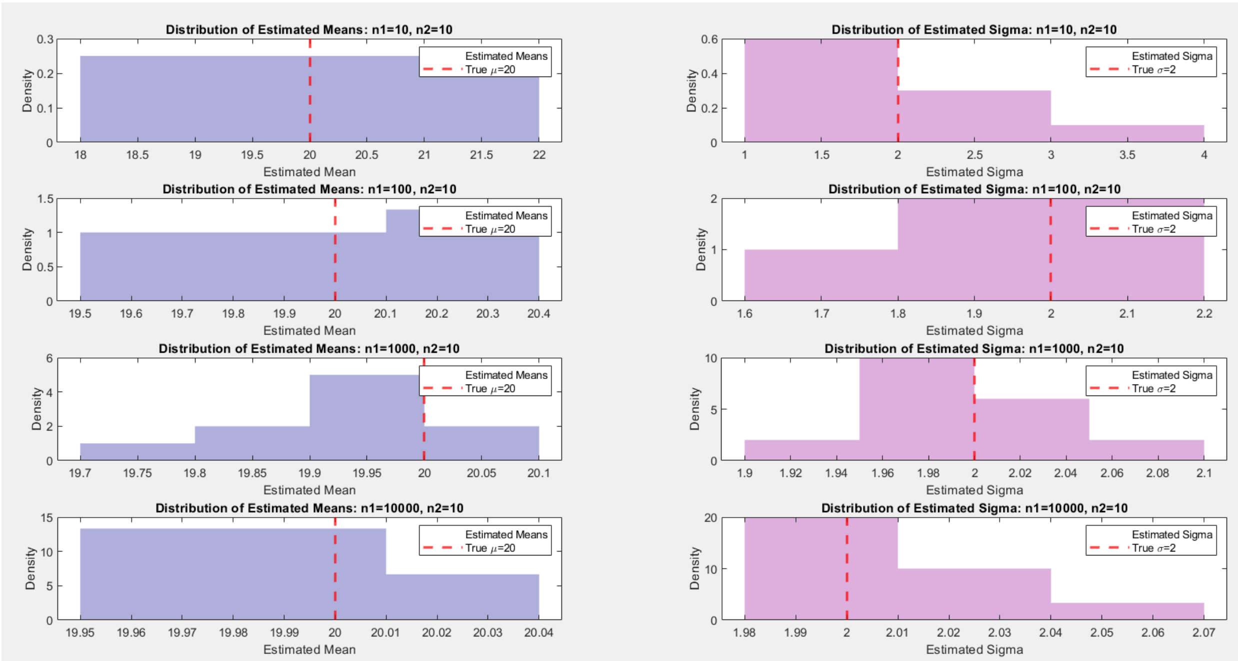
Parameters are given, and mean and variance of the generated sample of normal distribution are estimated using MLE and overlay estimated pdf to analyse the distribution of estimated parameters. The distribution of sample standard deviations is not normal but rather follows a chi-square distribution, However, for large n1, it can be approximated as normal with mean μ_{true} and variance $\frac{\sigma_{true}^2}{2n_1}$

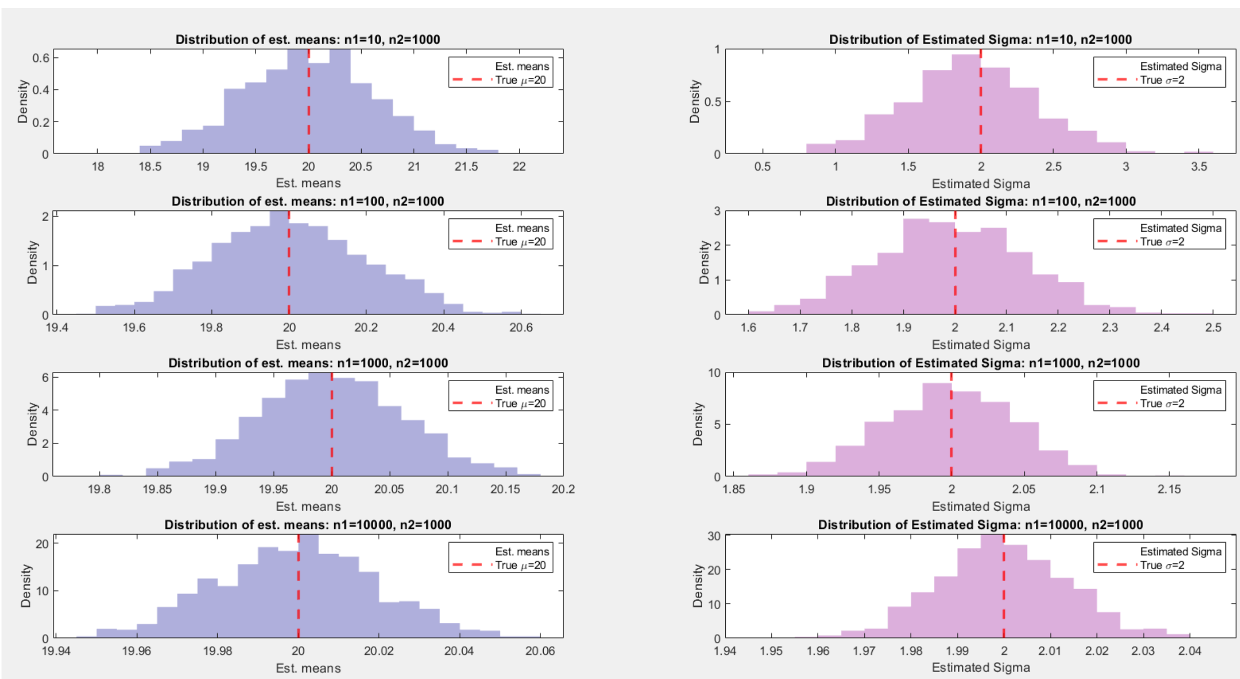
$$\mu = \frac{(x_1 + \ldots + x_n)}{n}$$

$$\sigma^2 = \frac{(x_1 - \mu)^2 + \ldots + (x_n - \mu)^2}{n}$$

RESULTS







DISCUSSION

The histogram shows significant variability in the simulated data for small values of n1. As n1 increases, the histograms become highly symmetric, closely resembling a normal distribution.The estimated parameters are nearly identical to the true values, and the estimated PDF matches the true PDF almost perfectly. For small sample sizes (n1 = 10), the distribution of estimated means is wide, indicating high variability across trials. However, it is centered around the true mean ($\mu = 20$), showing that the sample mean is an unbiased estimator. As n1 increases, the variability decreases, indicating convergence to the true mean. The sample mean is an unbiased estimator of the population mean because its average across simulations aligns with $\mu = 20$. The sample standard deviation is slightly biased for small sample sizes because it underestimates the population standard deviation due to its dependence on degrees of freedom. This bias diminishes as n1 increases

CONCLUSION

The sample mean is an unbiased estimator of the population mean. The sample standard deviation is slightly biased but becomes unbiased as the sample size increases since it follows normal distribution for larger n. For small sample sizes, there is significant variability in the estimates of μ and σ. Larger sample sizes lead to more accurate estimates and better alignment between the simulated data and the claimed distribution. Increasing n1 improves parameter estimation and reduces variability. This experiment demonstrates that MLE provides unbiased estimates of the mean and slightly biased standard deviation, and increasing sample size improves estimation accuracy

2

INTRODUCTION

Temperature calibrations of chemical reactions is an important step to determine the rate of the reaction because catalytic activity is heavily dependent on external temperature. However, sensor noise can distort readings, making it challenging to estimate the true temperature accurately. This study simulates temperature measurements affected by uniform noise and analyzes how this noise impacts Maximum Likelihood Estimation (MLE) of the true temperature parameters. The true temperature follows a normal distribution while the sensor noise is uniformly distributed. The goal is to evaluate how sensor noise affects estimation accuracy and whether the estimators remain unbiased

DATA

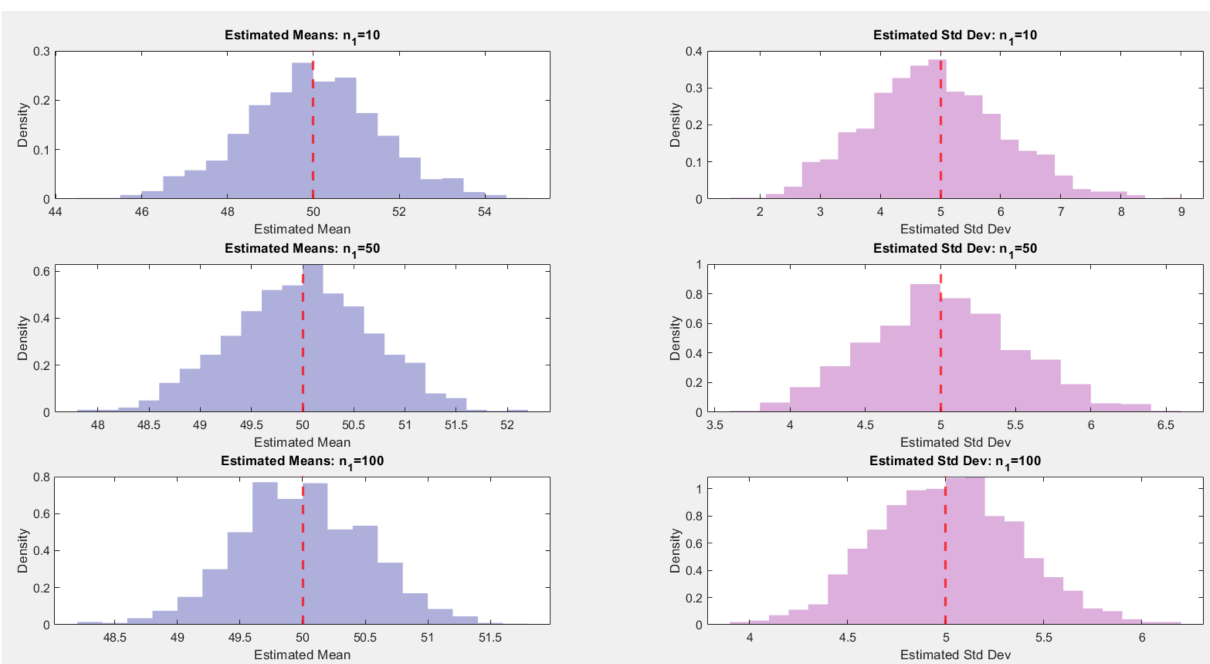
Parameters are given and the pdfs are generated using built-in functions

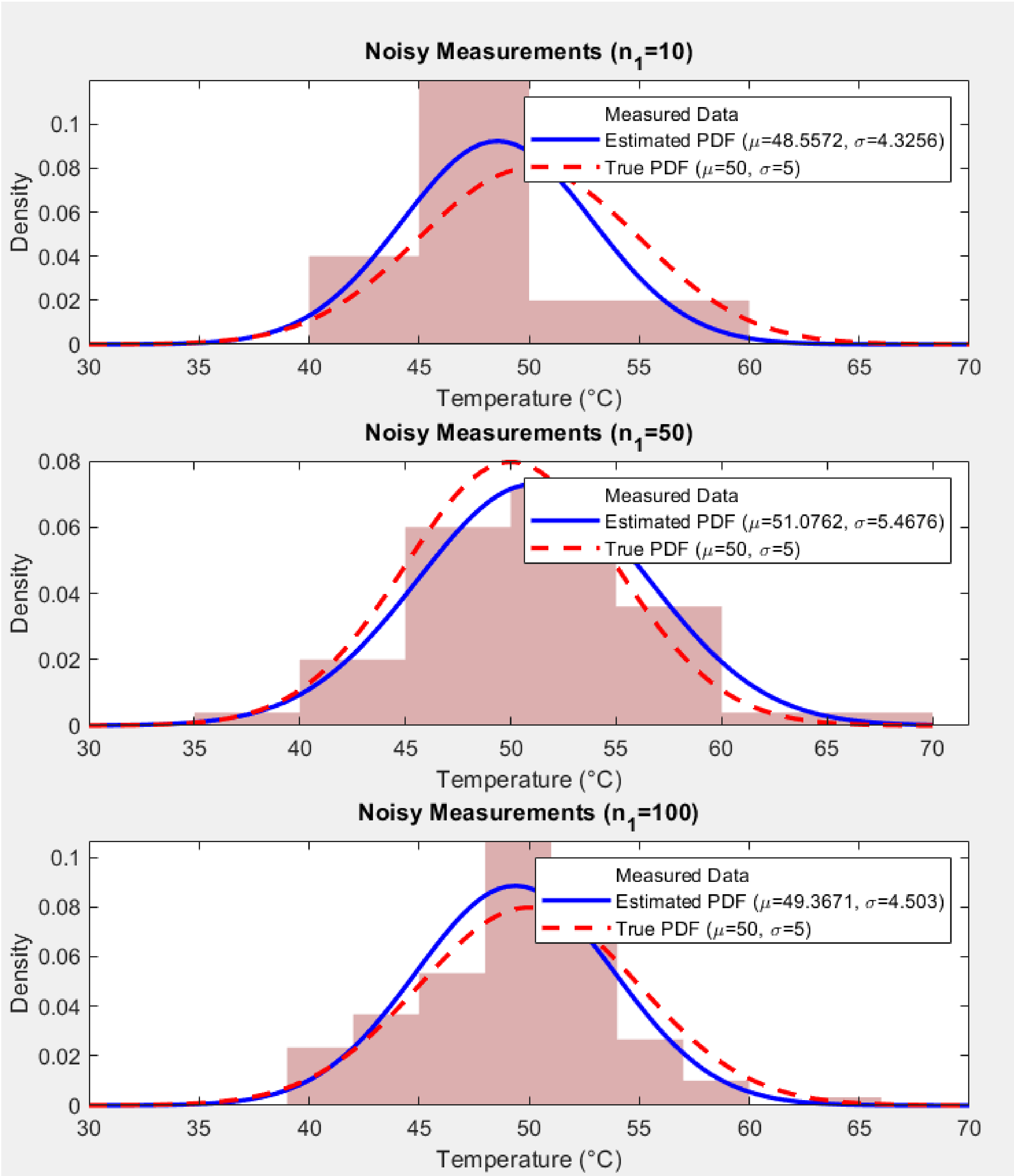
METHODOLOGY

Question 2

True temperature is generated for each value of n1 using a normal distribution with given parameters. Sensor noise (η) is added to the true temperature, where η follows a uniform distribution between -1 and 1. The measured temperature is $Y = X + \eta$. The mean (μ_Y) and standard deviation (σ_Y) of the noisy measurements are estimated using MLE, which corresponds to the sample mean and sample standard deviation. The simulation is repeated $n2 = 1000$ times for each sample size to analyze the variability of the estimated mean and standard deviation

RESULTS




Noisy Measurements (n1=10)
Measured Data
Estimated PDF (μ=48.5572, σ=4.3256)
True PDF (μ=50, σ=5)
Density
Temperature (°C)
Noisy Measurements (n1=50)
Measured Data
Estimated PDF (μ=51.0762, σ=5.4676)
True PDF (μ=50, σ=5)
Density
Temperature (°C)
Noisy Measurements (n1=100)
Measured Data
Estimated PDF (μ=49.3671, σ=4.503)
True PDF (μ=50, σ=5)
Density
Temperature (°C)

DISCUSSION

The sensor noise introduces variability in the temperature measurements, leading to slight biases in MLE estimates. While the mean remains relatively stable, the standard deviation is slightly inflated due to the added noise. Since uniform noise has zero mean, the expected value of Y remains close to $\mu X=50^\circ C$, but its variance increases. The estimated mean μY was close to $50^\circ C$, showing that mean estimation is nearly unbiased. Running the experiment multiple times (with n2 repetitions) allowed us to observe the distribution of estimated parameters. The histogram of σY values showed a wider spread, indicating higher variability in standard deviation estimates.

CONCLUSION

The sensor noise does not significantly affect the mean estimate, meaning the estimator for μ remains unbiased. However, the noise inflates the standard deviation estimate, making σ^Y a biased estimator for σX. The variability in σ^Y across experiments shows that small sample sizes lead to greater uncertainty in standard deviation estimation.

3

INTRODUCTION

Daily stock returns modelling is key to understanding risks and making informed financial decisions. While many stocks follow a normal distribution for returns, high-risk stocks often exhibit heavy tails, which are better modeled using a t-distribution. This study simulates daily returns of a high-risk stock and estimates its mean (μ) and standard deviation (σ) using MLE. Additionally, the effect of market noise on these estimates is analyzed to evaluate potential biases introduced by external factors.

DATA

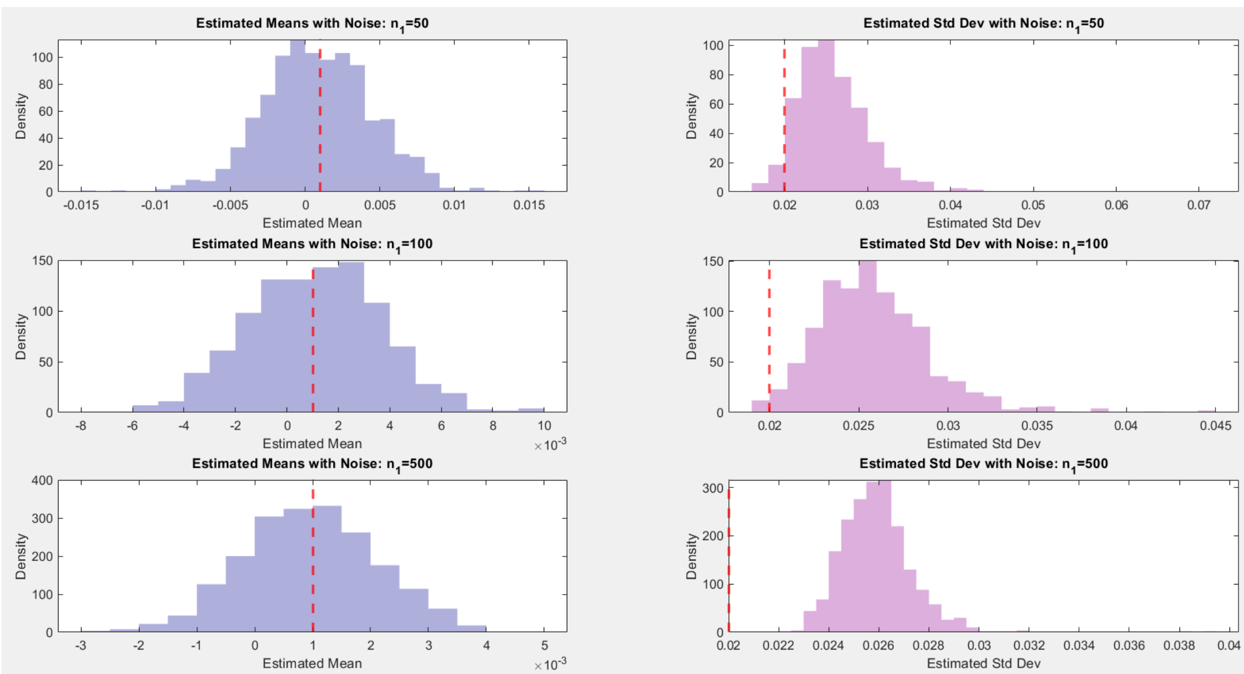
Parameters are given and the pdfs are generated using built-in functions

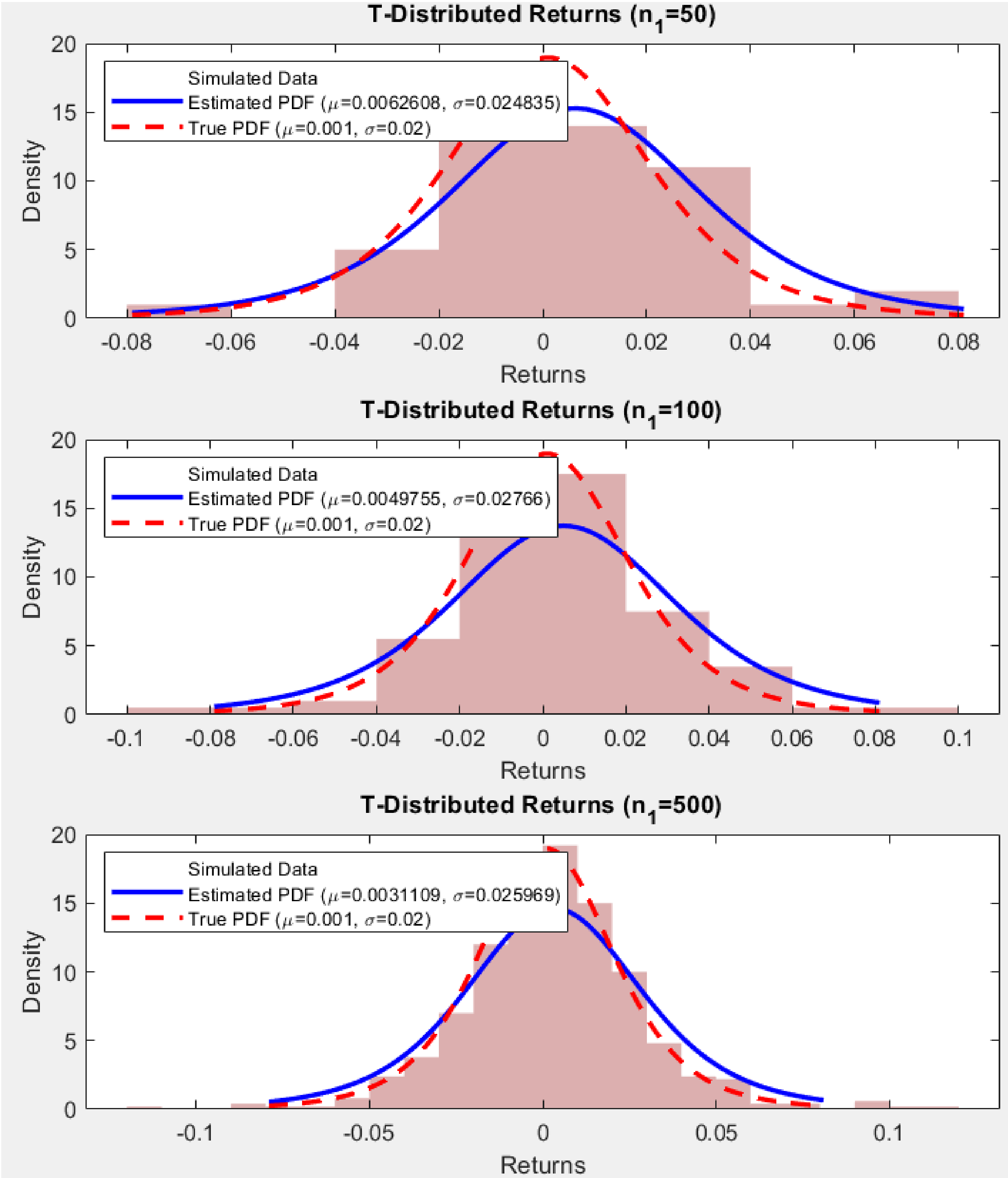
METHODOLOGY

Question 3

n1 daily stock returns are simulated using a t distribution, and uniform noise is added to test its impact on the mean and variance of the pdf (MLE is used to estimate them.) Since degrees of freedom are fixed, we only estimate μ and σ. This is then repeated n2 times to analyse variability across the estimated parameters.

RESULTS




T-Distributed Returns (n1=50)
Simulated Data
Estimated PDF (μ=0.0062608, σ=0.024835)
True PDF (μ=0.001, σ=0.02)
Density
Returns
T-Distributed Returns (n1=100)
Simulated Data
Estimated PDF (μ=0.0049755, σ=0.02766)
True PDF (μ=0.001, σ=0.02)
Density
Returns
T-Distributed Returns (n1=500)
Simulated Data
Estimated PDF (μ=0.0031109, σ=0.025969)
True PDF (μ=0.001, σ=0.02)
Density
Returns

DISCUSSION

The findings indicate how sample size affects the estimation of the mean and standard deviation of t-distributed returns. With an increase in $n1$ from 50 to 500, the estimated mean (μ) converges towards the actual mean (μ=0.001), which demonstrates that the estimator is unbiased and improves with larger sample sizes. But the estimated standard deviation (σ) is always greater than the actual value (σ =0.02) indicating that the presence of noise or the heavy-tailedness of the t-distribution impacts its accuracy. The histograms show that for smaller samples (n1=50) there is more variability in the estimates, whereas for larger samples (n1=500), the estimated pdf is closer to the true distribution. This illustrates the concept that larger sample sizes minimize variability in estimates of parameters and make them more trustworthy

CONCLUSION

The results verify the MLE of the mean to be quite unbiased irrespective of sample size, and the estimator of the standard deviation to be slightly inflated by noise and variability in the sample. With an increase in sample size, the estimates approach the true parameters more accurately, confirming the advantages of larger datasets in enhancing statistical precision. But for deviation in standard error estimates, it appears that when dealing with noisy data or heavy-tailed distributions, extra methods such as bias correction or better estimators would be required to enhance accuracy. Nevertheless, the findings emphasize the significance of sample size for reliable parameter estimation in statistical modeling

____