**DS201**

**Statistical Programming**

**Assignment 6**

**Ashutosh Kumar Jha**

**12340390**

**DSAI**

**2nd Year**

**Semester 4**

**---------------------------------------------------------------------------**

**Question 1:** Battery Life Analysis for Smartphone Manufacturer

**Introduction:** The smartphone manufacturer claims that the battery life of its latest model is normally distributed with a mean (μ = 20 hours) and a standard deviation (σ = 2 hours). Due to potential manufacturing variability, it is essential to validate these claims. This study simulates battery life measurements and employs Maximum Likelihood Estimation (MLE) to estimate the parameters. The goal is to assess the effect of sample size and the number of simulation trials on the accuracy and variability of the estimated parameters.

**Data:** The data used in this study is simulated from a normal distribution with:

* True mean (μ) = 20 hours
* True standard deviation (σ) = 2 hours

We are taking different combinations of n1 values i.e., 10,100,1000 and n2 values i.e.,100,1000 and generating the samples and graphs on the basis of these values of n1,n2.

**Methodology:**

1. **Simulation:**
   * Battery life measurements are generated using a normal distribution function..
2. **MLE Estimation:**
   * For each simulation, the sample mean and the population standard deviation (using the MLE formula) are calculated as estimates of μ and σ, respectively.
3. **Visualization:**
   * A histogram of the simulated data from a single trial is plotted along with the estimated normal probability density function (PDF).
   * Histograms of the estimated means and standard deviations from multiple simulations are generated, with the true parameter values indicated by reference lines.

4. **Analysis:**

* + The variability and accuracy of the estimators are evaluated by summarizing the mean and standard deviation of the estimates across the simulation trials.

**Results:**

1. **Single Simulation:** The histogram of the simulated battery life data closely follows the expected normal distribution. The overlaid PDF, based on the estimated parameters, matches the shape of the histogram.
2. **Multiple Simulations:**

* The histogram of estimated means is centered around the true mean of 20 hours, indicating that the sample mean is an unbiased estimator.
* The histogram of estimated standard deviations is centered near the true value of 2 hours.  
   The printed summary shows that the average estimated mean and standard deviation are very close to the true values, confirming that both estimators perform well, especially with a larger sample size.

**Discussion:**

The analysis indicates that:

* **Mean Estimator:** The sample mean is an unbiased estimator of the true mean. The distribution of the estimated means from multiple simulations is narrowly centered around 20 hours.
* **Standard Deviation Estimator:** While the MLE of the standard deviation may exhibit slight bias in very small samples, using a sample size of 100 measurements produces estimates that are nearly unbiased.
* **Impact of Simulation Trials:** Increasing the number of trials (n2) results in smoother histograms, which helps in better visualizing the variability and robustness of the estimates.  
   These findings emphasize the importance of a sufficiently large sample size and multiple simulation runs to obtain reliable parameter estimates.

**Conclusion:** The simulation study confirms that Maximum Likelihood Estimation (MLE) can reliably recover the battery life parameters when the sample size is adequate. Both the mean and standard deviation estimators closely approximate the true values (μ = 20 hours and σ = 2 hours). For practical validation of the manufacturer's claims, it is recommended to collect at least 100 battery life measurements. This approach ensures that the estimates are both accurate and robust, thereby providing a strong empirical basis for the manufacturer's assertions.

**Question 2:** Temperature Estimation in Chemical Reaction using a Noisy Sensor

**Introduction:** In this project, we aim to estimate the true temperature of a chemical reaction, which follows a normal distribution with a mean (μ) of 50°C and a standard deviation (σ) of 5°C. However, the sensor used for measurement introduces random noise due to calibration issues. The noise (η) is uniformly distributed between -1°C and 1°C, and the measured temperature is given by Y = X + η. We use Maximum Likelihood Estimation (MLE) to estimate the true temperature parameters from the noisy measurements, and analyze how sensor noise affects the estimation.

**Data:**

The true temperature, X, is simulated from a normal distribution with:

* Mean (μ) = 50°C
* Standard Deviation (σ) = 5°C

Sensor noise, η, is simulated from a uniform distribution between -1°C and 1°C.  
 The measured temperature is given by:

* Y = X + η

We are taking different combinations of n1 values i.e., 10,100,1000 and n2 values i.e.,100,1000 and generating the samples and graphs on the basis of these values of n1,n2.

**Methodology:**

1. **Simulation:**

* Generate n1 true temperature values X using the normal distribution (μ = 50, σ = 5).
* Generate n1 sensor noise values from a uniform distribution between -1 and 1.
* Compute the measured temperature Y as the sum of X and noise.

1. **Parameter Estimation (MLE):**

* The sample mean of Y is used as the MLE for the true mean μ, since the sensor noise has zero mean.
* The sample variance of Y is computed and the known variance of the sensor noise (4/12 = 0.3333) is subtracted to estimate the variance of X. The square root of this corrected variance gives the estimate for σ.

1. **Visualization:**

* A histogram of the noisy measurements Y is plotted, with the estimated normal probability density function (PDF) for X overlaid.
* Histograms for the estimated means and standard deviations from multiple simulation trials (n2 = 1000) are also plotted, with the true values (μ = 50°C and σ = 5°C) marked on the plots.

**Results:**

1. **Single Simulation:**The histogram of the noisy measurements shows the distribution of Y. The overlaid PDF, based on the MLE estimates for the true temperature parameters, closely approximates the expected normal curve for X.
2. **Multiple Simulations:**

* The histogram of estimated means is centered around 50°C, indicating that the sample mean is an unbiased estimator for μ even in the presence of sensor noise.
* The histogram of estimated standard deviations is centered near 5°C, after correcting for the sensor noise variance.  
   The printed summary confirms that the average estimated mean is close to 50°C and the average estimated standard deviation is close to 5°C.

**Discussion:**Sensor noise introduces additional variability in the measured temperature. However, because the noise has zero mean, the estimator for the mean remains unbiased. The variance of the measured data is inflated by the noise, but by subtracting the known noise variance (4/12), we can accurately recover the true variance of X. Thus, both the mean and the corrected standard deviation estimators remain unbiased when the noise distribution is known. Compared to a scenario without sensor noise, the process now includes an extra correction step for the variance estimation.

**Conclusion:** The simulation study demonstrates that even with sensor noise, it is possible to accurately estimate the true temperature of the chemical reaction using MLE. The mean estimator remains unbiased because the noise is zero-mean, and the standard deviation estimator can be corrected by subtracting the noise variance. These findings highlight the importance of accounting for measurement noise in experimental setups and confirm that the estimators for the true temperature parameters are still unbiased when proper adjustments are made.

**Question 3: Estimation of High-Risk Stock Returns under t-Distribution with Market Noise**

**Introduction:** This project analyzes the daily returns of a high-risk stock whose returns are assumed to follow a t-distribution with heavy tails. Heavy tails imply that extreme gains or losses are more likely than in a normal distribution. The target parameters for the true returns are a mean (μ) of 0.1% and a standard deviation (σ) of 2%. In addition, the effect of market noise is considered. Market noise, uniformly distributed between -0.5% and 0.5%, is added to the returns to simulate real-world conditions. Maximum Likelihood Estimation (MLE) is used to estimate the parameters of the true return distribution, and the impact of noise on the estimates is examined.

**Data:**

The simulated data consists of:

* **True Stock Returns (X):** Generated from a t-distribution with:  
  + Mean (μ) = 0.1%
  + Standard Deviation (σ) = 2%
  + Degrees-of-freedom (df) = 5 (to model heavy tails)
* **Market Noise (η):** Random noise sampled from a uniform distribution between -0.5% and 0.5%.
* **Noisy Returns (Y):** The observed returns are given by Y = X + η.

Two experiments are performed:

1. Simulation of returns without added noise.
2. Simulation of returns with market noise.

**Methodology:**

**1. Simulation of Returns:**

* For the true returns, a t-distribution with df=5 is used. The t-distribution is scaled and shifted to obtain the desired mean and standard deviation.
* For noisy returns, market noise is generated from a uniform distribution and added to the true returns.

**2. Parameter Estimation (MLE):**

* The sample mean is used as the MLE for the true mean μ.
* The sample standard deviation (population standard deviation) is used as the MLE for σ.

**3. Visualization:**

* Histograms of simulated returns (with and without noise) are plotted.
* The t-distribution PDF, using the estimated parameters and an assumed df=5, is overlaid on the histograms.
* Multiple simulation trials (n2 = 1000) are conducted to assess the variability of the estimators.

**4. Comparison:**

* Histograms of the estimated means and standard deviations from both experiments are compared.
* The true values (μ = 0.1% and σ = 2%) are marked on the plots.

**Results:**

1. **Without Market Noise:**The histogram of the simulated returns closely follows a heavy-tailed t-distribution. The overlaid t-distribution PDF, using the MLE estimates, aligns well with the histogram. The sample mean and standard deviation estimates are centered around the true values.
2. **With Market Noise:**Adding uniform market noise increases the overall variability of the observed returns. While the sample mean remains approximately unbiased (since the noise has zero mean), the estimated standard deviation is inflated due to the additional variability from the noise. However, the estimators remain effective after sufficient simulation trials.

The summary of multiple simulations confirms that the mean estimator is unbiased in both cases, while the variance estimator requires correction for the added noise if an accurate measure of the true volatility is desired.

**Discussion:**

The heavy-tailed t-distribution naturally produces a higher likelihood of extreme returns compared to a normal distribution. This behavior is evident in the simulation results. The addition of market noise introduces further variability:

* **Mean Estimation:**Since the noise is symmetric with a zero mean, the estimator for μ remains unbiased.
* **Standard Deviation Estimation:**The noise inflates the overall observed variance. Without adjusting for the known noise variance, the estimate for σ appears higher. Thus, in practice, if the noise characteristics are known, a correction may be applied to recover the true volatility.

Overall, both the t-distribution and noise affect the precision of the estimates. However, with a sufficient sample size and repeated trials, the MLE approach provides robust estimates.

**Conclusion:** This study demonstrates that even for high-risk stocks with heavy-tailed return distributions and market noise, MLE can yield reliable estimates for the true mean and standard deviation. The mean estimator remains unbiased despite the noise, while the standard deviation estimator is influenced by the additional variability. The findings emphasize the importance of accounting for market noise when analyzing stock returns and suggest that proper corrections are necessary for accurate volatility measurement.

Code: [12340390 Ashutosh Asg6.ipynb](https://colab.research.google.com/drive/1QJwp5M_6YBb1uSEIjrmBySBhDyp_zZS7?usp=sharing)