Problem 1:

a. From given information, the results are:

Mean: 1.0489703904839585
Variance: 5.427220681881727
Skewness: 0.879288059847244
Excess Kurtosis: 23.069982510610533

b. I choose to use SciPy, and my results is like:

Mean: 1.0489703904839585
Variance: 5.427220681881727
Skewness: 0.8806086425277364
Kurtosis: 23.122200789989723

c. I will generate many samples from a standard normal distribution and calculate the sample moments(the mean, variance, skewness, and kurtosis).

And then average those to get an estimate of the expected value of the sample moments. If the statistical package is unbiased, these should converge to the population moments of the normal distribution.

Later, I'll use a one-sample t-test to compare each sample moment's mean estimate against the population moment (mean=0, variance=1, skewness=0, kurtosis=0).

H0: The statistics calculated by Scipy is unbiased.

If the p-value from the t-test is less than 0.05, we can reject the null hypothesis, suggesting that the statistics calculated by Scipy may be biased. If the p-value is greater than or equal to the significance level, then there is not enough evidence to reject the null hypothesis, and it can be considered that the statistics calculated by Scipy are unbiased.

And I got my results as below:

Mean: Scipy unbiased p-value = 0.9652711036026391
Variance: Scipy unbiased p-value = 0.857632242188823
Skewness: Scipy unbiased p-value = 0.12347004083706752
Kurtosis: Scipy unbiased p-value = 0.0038863620988585272

Therefore, the mean, variance, skew gotten from Scipy are unbiased, while, the kurtosis is biased.

Problem 2:

a. For the OLS model, which involves finding β of the linear regression model that minimize the sum of squared residuals btw y and x. The OLS model provides the following estimates:

The slope coefficient (β) for x is 0.7752740987226114.
The intercept is -0.08738446427005075.
The sample standard deviation of the residuals (σ) is 1.0062751598133177.

Next, use MLE for the same model, under the assumption of normally distributed errors. I need to maximize the likelihood function for the normal distribution, and this model has resulted in the following parameter estimates:

The slope coefficient (β) for x is 0.775274090959329.
The intercept is -0.08738443434812372.
The sample standard deviation of the residuals (σ) is 1.0037563160744265.

From the results of these two models, which give me very similar estimates, suggesting that the assumption of normally distributed errors is mostly correct.

And the difference could be explained as: OLS models always try to minimize the sum of squared residuals, which are the differences between observed and predicted values. While, MLE models are trying to maximize the likelihood function, which means finding the parameter values that make the observed data with largest likelihood.

b. If the errors follow T-distribution, with MLE model, the parameter estimates are:

The slope coefficient (β) for x is 0.6750098853428518.
The intercept is -0.0972693511441836.
The sample standard deviation of the residuals (σ) is 0.8551051362349041.
The degrees of freedom (df) of the T-distribution is 7.159864767979876.

And comparing this result with those under noraml distributed errors' assumption, I choose to use AIC value to find which model could fit better.

AIC for normal distribution: 575.0751261088562.
AIC for T-distribution: 570.5868063613983.

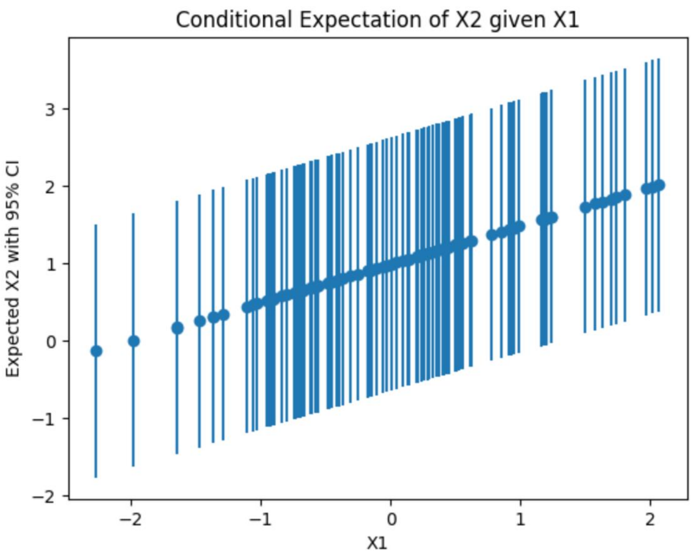
It is clear that the AIC value among MLE under T distribution assumption is lower, indicating a better fit.

c. To figure out the distribution of X2, which is related to the observed value of X1. I need to fit the multivariate normal distribution with MLE to find the mean vector and covariance matrix firstly, which is:

mean: [0.0010227 0.99024382]
covariance:[[1.06977464 0.53068455]
[0.53068455 0.96147329]]

As the conditional distribution of a multivariate normal distribution is also normal, with the mean vector and covariance matrix I have already gotten above, I will figure out the expected value of X2 for each X1 and the 95% confidence interval around the expected value.

X1 observed: -0.799934199
Conditional X2 mean: 0.5929120487352633
95% CI: (-1.044850496925984, 2.2306745943965103)



d. The process is shown in the figure below:

d. As $\epsilon \sim N(0, \sigma^2 I_n)$

The likelihood function for the multivariate normal distribution is:

$$L(\beta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \cdot e^{-\frac{1}{2\sigma^2}(y-X\beta)^T(y-X\beta)}$$

$$\ln L(\beta, \sigma^2) = -\frac{n}{2} \times \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \times (y-X\beta)^T(y-X\beta)$$

To find max value of $\ln L(\beta, \sigma^2)$, we need to get partial derivatives of both β and σ^2, and make these equal zero.

$$\frac{\partial \ln L(\beta, \sigma^2)}{\partial \beta} = -\frac{1}{2\sigma^2} \frac{\partial}{\partial \beta}[(y-X\beta)^T(y-X\beta)] = \frac{1}{\sigma^2} X^T(y-X\beta) = 0.$$

As $(y-X\beta)^T(y-X\beta) = y^Ty - y^TX\beta - \beta^TX^Ty + \beta^TX^TX\beta = y^Ty - y^TX\beta - (X\beta)^Ty + \beta^TX^TX\beta$

and $y^T(X\beta)$ & $y(X\beta)^T$ are scalars, thus, $= y^Ty - 2y^TX\beta + \beta^TX^TX\beta$.

Therefore, $X^Ty = X^TX\beta \quad \therefore \hat{\beta} = (X^TX)^{-1} \cdot X^Ty$

$$\frac{\partial \ln L(\beta, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(y-X\beta)^T(y-X\beta) = 0.$$

$$\hat{\sigma}^2 = \frac{1}{n}(y-X\hat{\beta})^T(y-X\beta)$$

$$\begin{cases} \hat{\beta} = (X^TX)^{-1}X^Ty \\ \hat{\sigma}^2 = \frac{1}{n}(y-X\hat{\beta})^T(y-X\beta) \end{cases}$$

Problem 3:

To assess the best fit among these MA and AR models, there are criterias like AIC and BIC, and I will use these two to help me, with the information that lower values of AIC and BIC generally indicate a better model fit. With the values shown in the form below, we can see that the AR(3) model appears to be the best fit for the data. It has the lowest AIC and BIC values among all the fitted models, indicating a good balance between model complexity and fit to the data.

AR(1) model AIC: 1644.6555047688475, BIC: 1657.299329064114
AR(2) model AIC: 1581.079265904978, BIC: 1597.9376982986669
AR(3) model AIC: 1436.6598066945867, BIC: 1457.7328471866977
MA(1) model AIC: 1567.4036263707874, BIC: 1580.047450666054
MA(2) model AIC: 1537.9412063807388, BIC: 1554.7996387744276
MA(3) model AIC: 1536.8677087350309, BIC: 1557.9407492271419