Kexin You (ky125)

Problem 1

a.

First four moments values (normalized):

$\widehat{\mu_1} = E(X) = \frac{1}{n}\sum_i^n x_i$=1.0489703904839585 ≈ 1.0490 (Unbiased)

$\widehat{\mu_2} = E((X - \widehat{\mu_1})^2) = \frac{1}{n-1}\sum_i^n (x_i - \widehat{\mu_1})^2$=5.427220681881726 ≈ 5.4272 (Unbiased)

$\widehat{\mu_3} = E\left((\frac{X-\widehat{\mu_1}}{\sigma})^3\right) = \frac{1}{n}\sum_i^n (\frac{x_i-\widehat{\mu_1}}{\sigma})^3$=0.8806086425277365 ≈ 0.8806 (Biased)

$\widehat{\mu_4} = E\left((\frac{X-\widehat{\mu_1}}{\sigma})^4\right) = \frac{1}{n}\sum_i^n (\frac{x_i-\widehat{\mu_1}}{\sigma})^4$=26.12220078998973 ≈ 26.1222 (Biased)

b.

I use "numpy" package to calculate mean and variance, and "scipy.stats" package (more specifically kurtosis and skew) to calculate skewness and kurtosis. The answers are as follows.

$\widehat{\mu_1}$ =1.0489703904839585 ≈ 1.0490

$\widehat{\mu_2}$ =5.427220681881726 ≈ 5.4272

$\widehat{\mu_3}$ =0.8806086425277365 ≈ 0.8806

$\widehat{\mu_4}$ =26.12220078998973 ≈ 26.1222

c.

Hypothesis: When the "numpy" package are used to calculate the mean and variance with "ddof=1", the estimation of mean and variance are unbiased.
However, when the "scipy.stats" package are used to calculate the skewness and kurtosis with "bias=True", the estimation of skewness and kurtosis are biased.

	Question a	Question b	Same or Not	Biased or Unbiased
Mean	1.0490	1.0490	Same	Unbiased
Variance	5.4272	5.4272	Same	Unbiased
Skewness	0.8806	0.8806	Same	Biased
Kurtosis	26.1222	26.1222	Same	Biased

Table 1: first four moments in problem 1 question a and b

Table 1 compares the first four moments calculated in question a and b, where the first four moments in question a and question b are entirely same. Thus, whether the moments calculated with packages in question b are biased are the same as the moments in question a. Since, I use the unbiased formula to calculate mean and variance as well as biased formula to calculate skewness and kurtosis, the hypothesis can be proved:

The mean is unbiased.
The variance is unbiased.
The skewness is biased.
The kurtosis is biased.
(Other choice could be using t-test for hypothesis with random sampling, and trying to reject H0)

Problem 2

a.

Fit the data with OLS in "statsmodels" package, and the outcome is in table 2.
OLS for beta_x=0.7753; OLS for beta_const=-0.0874

Standard deviation of the OLS error= 1.003756319417732≈1.0038

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.0874	0.071	-1.222	0.223	-0.228	0.054
x	0.7753	0.076	10.226	0.000	0.626	0.925

Table 2: OLS regression in problem 2 question a

To fit the data with MLE given the assumption of normality, I construct

$$ll = -\frac{n}{2}ln\left(\sigma^2 2\pi\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2$$

Then use "minimize" in the "scipy.optimize" package to solve the maximization problem and have the following result:

```
MLE for beta_hat: [-0.08738421  0.77527368]
MLE for sigma_hat: 1.0037567948322634
```

MLE for beta_x≈0.7753; MLE for beta_const≈-0.0874

Standard deviation of the MLE error= $\sigma \approx 1.0038$

Compare the beta and standard deviation between OLS and MLE, the beta are nearly same for OLS and MLE because the estimation of beta are both $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

The standard deviation of the estimation error σ are nearly same because n=1000, which is a large number, so $\frac{1}{n}$ and $\frac{1}{n-1}$ are nearly the same.

b.

Given the assumption of T distribution of errors, in order to find parameter values to maximize the joint probability of the observed data, I construct $ll = \sum_{i=1}^{n} ln(f(t|\nu))$, where

f(t|ν) is the Probability Density Function (PDF) of t-distribution.
Using "minimize" in the "scipy.optimize" package to solve the maximization problem, I get the estimation of beta and sigma. Then I calculate Akaike Information Criterion (AIC) with the formula $AIC = 2k - 2log(\hat{L})$ to find the best fitted model.

```
MLE for beta_hat_t:  [-0.09726205  0.67500404]
MLE for sigma_hat_t:  0.8550990876723791
MLE for df_hat_t:  7.159239843989632
Log likelihood_normal: 284.53756305449343 Log likelihood_t: 281.2934031988891
AIC_normal: 575.0751261089869 AIC_t: 570.5868063977782
```

Table 3: MLE fitted parameters in problem 2 question b

From table 3, fitted parameters under T distribution assumption are:

MLE for beta_x≈0.6700; MLE for beta_const≈ - 0.0973

Standard deviation of the MLE error =σ≈0.8551

The AIC under T distribution is 570.5868, which is lower than AIC under normal distribution at 575.0751. Thus, the fitted parameters under T distribution assumption is the best of fit.

c.

Distribution of X2 given each observed value is normal distribution.

$$\mu_{X_2|X_1} = \mu_{X_2} + \frac{\sigma_{X_1X_2}}{\sigma^2_{X_1}}(X_1 - \mu_{X_1})$$

$$\sigma^2_{X_2|X_1} = \sigma^2_{X_2} - \frac{\sigma^2_{X_1X_2}}{\sigma^2_{X_1}}$$

The expected value along with 95% confidence interval are in Figure 1.

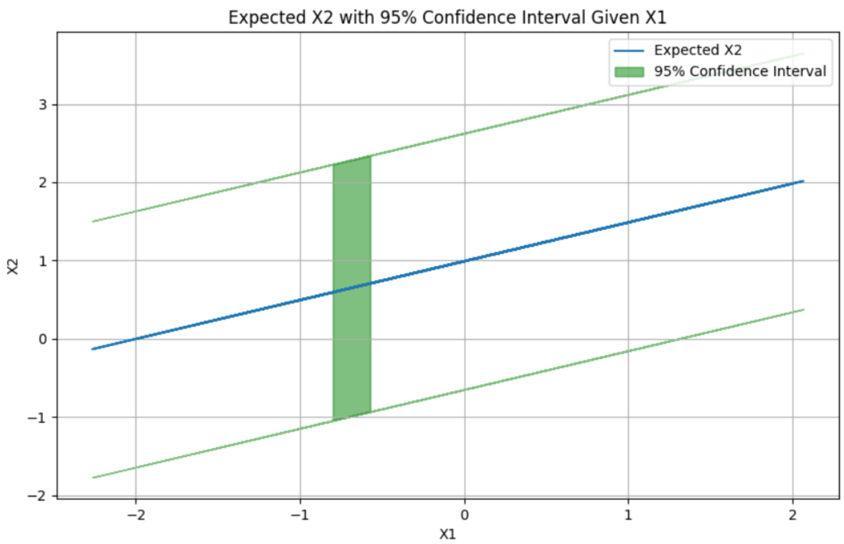

Figure 1: Expected X2 with 95% confidence interval given X1 in problem 2

d.

Likelihood function for a multivariate normal distribution is:

$$L(\beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right)$$

Then log likelihood function is

$$l(\beta, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)$$

In order to maximize this log-likelihood function, calculate the partial derivative of the log-likelihood with respect to β and σ^2 respectively and set them to zero.

From solving $\frac{\partial \ell(\beta,\sigma^2)}{\partial \beta} = 0$, get

$$\hat{\beta} = (X'X)^{-1}X'Y$$

From solving $\frac{\partial \ell(\beta,\sigma^2)}{\partial \sigma^2} = 0$, get

$$\widehat{\sigma^2} = \frac{1}{n}(Y - X\hat{\beta})'(Y - X\hat{\beta})$$

Problem 3

Using the ARIMA model in “statsmodels.tsa.arima.model” package to fit the data with AR(1) through AR(3) and MA(1) through MA(3). Then plot the AIC and BIC for each fitted model. The output is illustrated in Table 4.

	AIC	BIC
AR(1)	1644.6555	1657.2993
AR(2)	1581.0793	1597.9377
AR(3)	1436.6598	1457.7328
MA(1)	1567.4036	1580.0475
MA(2)	1537.9412	1554.7996
MA(3)	1536.8677	1557.9407

Table 4: MLE fitted parameters in problem 3

As AIC at 1436.6598 for AR(3) is the lowest among AIC for all models, and BIC at 1457.7328 for AR(3) is also the lowest among all models’ BIC, it can be concluded that AR(3) is the best of fit.