FINTECH 545 HW1

YUQI WANG

yw596

PROBLEM 1

A.
According to the formula in the "Week1 – Univariate Stats":

$$\mu_n = \int_{-\infty}^{\infty} (x - c)^n f(x)dx = E[(x - c)^n]$$

For the first moment – mean, we set $c = 0$. The mean is usually donated μ.

For the higher moments, we set $c = \mu$.

So the results:

Mean	1.0489703904839582
Variance	5.4272206818817255
Skewness	0.8806086425277376
Kurtosis	23.12220078998972

B.
I use pandas package to calculate the four moments directly, and it turns out:

Mean	1.0489703904839585
Variance	5.427220681881727
Skewness	0.8819320922598395
Kurtosis	23.2442534696162

Except the mean, other moments values are a bit different from the results in (a) which I use the biased functions to figure out.

C.

My statistical package is unbiased by default. To prove this point, I make simulations with random samples given the assumption of normality (0,1). The sample size is 100, and I did 1000 simulations.

For this standard normality, the true values of those moments are:

$$\mu_{10} = 0, \qquad \mu_{20} = 1, \qquad \mu_{30} = 0, \qquad \mu_{40} = 0$$

$$H_0: \mu_n - \mu_{n0} = 0, \qquad H_1: \mu_n - \mu_{n0} \neq 0,$$

After simulations, we compare the expected observed value with the true value and note down the p-values. For pandas package, all p-values of those four moments are very close to 0. So, we cannot reject the null hypothesis, which can support that the package functions are unbiased estimators.

PROBLEM 2

A.

Regression Results:

	β_0	β_1	ϵ
OLS	-0.08738	0.77527	1.0037563
MLE (normality)	-0.08738	0.77527	1.0037568

The betas are nearly the same based on OLS and MLE. This is because the assumption of normality in MLE method is like the settings in OLS method.

The standard deviation of the error is also very close.

B.

Regression Results:

	β_0	β_1	ϵ	R^2	Adj-R^2
MLE (T)	-0.09619	0.72658	1.0	0.3456	0.3423
MLE (norm)	-0.08738	0.77527	1.0037568	0.3442	0.3408

I use adjusted R^2 to compare which assumption fits better. The adjusted R-squared is higher for the normality assumption compared to the T-distribution assumption, it generally indicates that the model assuming normal distribution of errors provides a better fit to the data. The adjusted R-squared takes into account the complexity of the model by considering the number of explanatory variables and the sample size, and a higher value suggests a better fitting model.

	β_0	β_1	ϵ	AIC	BIC
MLE (T)	-0.09619	0.72658	1.0	573.075	579.671
MLE (norm)	-0.08738	0.77527	1.0037568	572.166	582.061

Additionally, I use AIC and BIC to compare which assumption fits better.

AIC (Normal) is greater than AIC (T-distribution), it suggests that the T-distribution model provides a relatively better fit to the data (lower AIC values indicate better fit). Additionally, BIC (Normal) is smaller than BIC (T-distribution), it implies that the Normal distribution model has an advantage when considering both the goodness of fit and the complexity of the model (lower BIC values indicate a better trade-off between fit and complexity).

This scenario may be interpreted in different ways:

Model Fit: AIC emphasizes better fit to the data, while BIC balances fit and model complexity. The Normal distribution model may fit the data better, but the T-distribution model is penalized less in BIC, leading to a smaller BIC.

Domain-specific Considerations: Different problem domains may have varying impacts on model selection. Models that are more robust to outliers might perform better in AIC, but BIC could favor simpler models.

Data Characteristics: The nature of the data can influence the performance of AIC and BIC differently. For instance, if there are many outliers, the T-distribution model might be more suitable, but the Normal distribution model could be favored by BIC as outliers might be considered noise.

C.

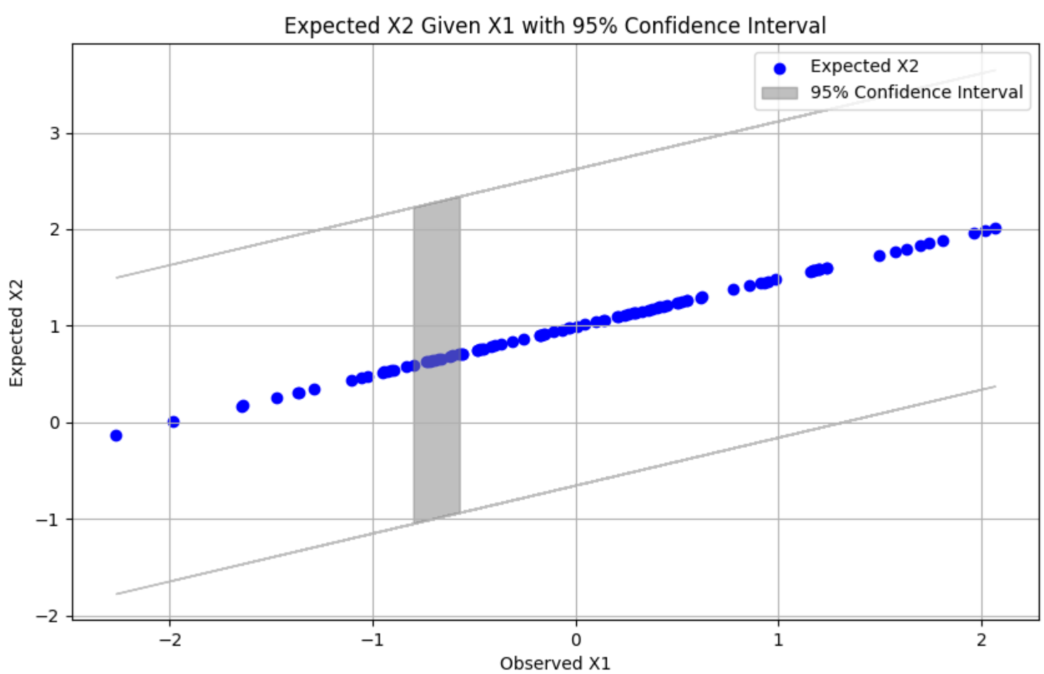
For each observed value of 'x1' (from "problem2_x1.csv"), calculate the conditional distribution of 'x2' given 'x1.' This is done using the properties of multivariate normal distributions.

$$f_{X_2|X_1}(x_2|x_{1,\text{observed}}) = \frac{f_{X_1,X_2}(x_{1,\text{observed}},x_2)}{f_{X_1}(x_{1,\text{observed}})}$$

For a bivariate normal distribution, the joint PDF and marginal PDF are given by:

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sqrt{\det(\Sigma)}}\exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right)$$
$$f_{X_1}(x_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right)$$

Substitute these expressions into the conditional distribution formula to get the conditional PDF of X_2 given $X_1 = x_{\text{observed}}$. The condition mean and variance of X_2 can be calculated from this conditional distribution.



D.

Proof:

$$Y = X\beta + \varepsilon$$

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

The log-likelihood function:

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

Take the derivative of l with respect to β or σ^2, setting it to 0:

$$\frac{\partial l}{\partial \beta} = X^T(Y - X\beta) = 0$$

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

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(Y - X\beta)^T(Y - X\beta) = 0$$

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

PROBLEM 3

Using statsmodels package to Fit the data in problem3.csv using AR (1) through AR (3) and MA (1) through MA (3), respectively.

AR(1): AIC = 1644.6555, BIC = 1657.2993

AR(2): AIC = 1581.0793, BIC = 1597.9377

AR(3): AIC = 1436.6598, BIC = 1457.7328

MA(1): AIC = 1567.4036, BIC = 1580.0475

MA(2): AIC = 1537.9412, BIC = 1554.7996

MA(3): AIC = 1536.8677, BIC = 1557.9407

Since the AIC and BIC of AR(3) are the lowest compared to other models, so AR(3) is the best of fit.