Fintech545 Project1

Jieyang Ran

September 15, 2024

1 Problem 1

1.1 Calculate the first 4 moment values using the normalized formulas in the Week 1 notes.

According to the calculation:
the 1st moment: 1.0489703904839585
the 2nd moment: 5.427220681881727
the 3rd moment: 0.8819320922598392
the 4th moment: 26.17447749985716

1.2 Calculate the first 4 moment values using your chosen statistical package.

According to the calculation:
the 1st moment: 1.0489703904839585
the 2nd moment: 5.4272206818817255
the 3rd moment: 0.8819320922598395
the 4th moment: 23.2442534696162

1.3 Are your statistical package functions biased? Prove or disprove your hypotheses. Explain your conclusion.

Not every equation is biased. For the first moment and second moment, they are unbiased. Because the function in the package pandas, which I used to calculate for the first 4 moment values, adjust the concerning functions for calculating the 2nd moment. For the 3rd and 4th moment, they are biased. We notice that there seems no bias in the calculation of the 3rd moment. That's because the data volume is big. Then, the difference between bias and unbiased one is small.

2 Problem 2

2.1 Fit the data in problem2.csv using OLS. Then fit the data using MLE given the assumption of normality. Compare the beta values and the standard deviation of the OLS errors to the fitted MLE σ. What is your finding? Explain any differences.

Given the assumption of normality:
beta value for OLS: 0.76903146
standard deviation of the OLS errors: 1.0037735049808605
fitted MLE β: 0.7690316631477303
fitted MLE σ: 1.0075356657668637
The estimations for beta are nearly the same, but the estimations for σ have slightly difference. That may due to the reason that OLS focus on minimizing the sum of squared errors, while MLE focus on maximizing the likelihood.

2.2 Fit the data in problem2.csv using MLE given the assumption of a T distribution of errors. Show the fitted parameters. Compare the fitted parameters among the MLE under the normality assumption and T distribution assumption. Which is the best fit?

Given the assumption of T-distribution:
fitted MLE β: 0.6672448846553243
fitted MLE σ: 0.8594379316104537
fitted MLE degree of freedom: 7.156857171146711

We calculate the AIC and BIC to compare the goodness of fit. Below is the calculation results.
AIC for MLE under normality assumption: 574.578
AIC for MLE under T distribution assumption: 570.642
BIC for MLE under normality assumption: 581.175
BIC for MLE under T distribution assumption: 580.537

Both AIC and BIC for MLE under T distribution assumption are smaller that those for MLE under normality assumption. Then, we can get the conclusion that MLE under the T distribution assumption is the best fit.

2.3 Fit a multivariate distribution to the data in problem2_x.csv. Given the values of X_1 what are the conditional distributions for X_2 for each observation. Plot the expected value along with the 95% confidence interval and the observed value.

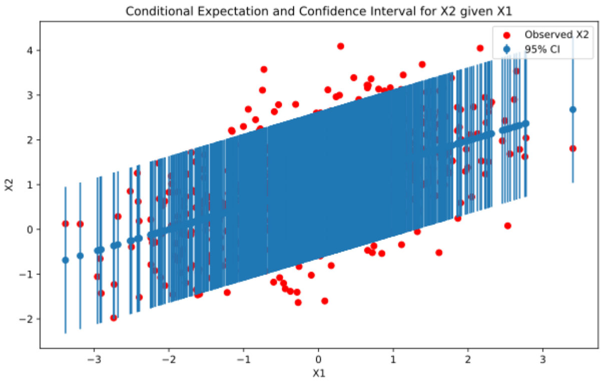

Figure 1: 2.3

2.4 (1 point Extra Credit). Y = Xβ + ϵ and $\epsilon \sim$ N(0, σ^2) . Derive the maximum likelihood estimators for β and σ.

$$\epsilon = Y - X\beta \tag{1}$$

$$\epsilon \sim N(0, \sigma^2) \tag{2}$$

From equation (1) and (2), we can get:

$$Y - X\beta \sim N(0, \sigma^2 I) \tag{3}$$

According to MLE, we need to maximize the following equation:

$$L = \prod_{n} \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} exp(-\frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta)) \tag{4}$$

Then, we take the logarithm and its negative to simply this process. Next, we need to minimize the following equation:

$$-logL = \frac{n}{2}log(2\pi) + nlog(\sigma) + \frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta)) \tag{5}$$

To get the estimator for β and σ, we can take the derivatives of them separately.
For β:

$$\frac{\partial(-logL)}{\partial\beta} = \frac{1}{\sigma^2}X^T(Y - X\beta) = 0 \tag{6}$$

We can get the estimator for β:

$$\hat{\beta}_{MLE} = (XX^T)^{-1}X^TY \tag{7}$$

For σ:

$$\frac{\partial(-logL)}{\partial\sigma^2} = \frac{n}{2\sigma^2} - \frac{1}{2\sigma^4}(Y - X\hat{\beta})^T(Y - X\hat{\beta}) = 0 \tag{8}$$

we can get the estimator for σ^2:

$$\hat{\sigma}_{MLE} = \frac{1}{n}(Y - X\hat{\beta})^T(Y - X\hat{\beta}) \tag{9}$$

3 Problem 3 Examine the data in problem3.csv; which AR(n) or MA(n) model do you expect to fit this data best? Fit the data using AR(1) - AR(3) and MA(1) - MA(3) models. Which is the best fit and does this confirm your hypothesis?

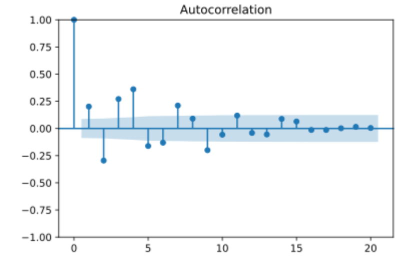

Figure 2: ACF

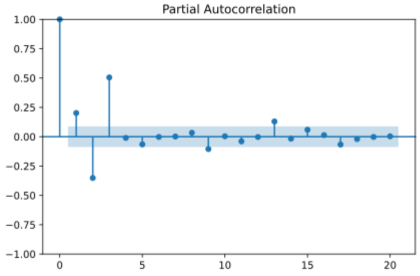

Figure 3: PACF

We still use AIC and BIC to test goodness of fit for each model. AR Model Metrics:
AR(1) - AIC: 1641.09, BIC: 1653.73
AR(2) - AIC: 1574.83, BIC: 1591.68
AR(3) - AIC: 1428.26, BIC: 1449.31

MA Model Metrics:
MA(1) - AIC: 1567.40, BIC: 1580.05
MA(2) - AIC: 1537.94, BIC: 1554.80
MA(3) - AIC: 1536.87, BIC: 1557.94

I expect AR(n) model to fit this data best according to the ACF plot and PACF plot. It shows that the auto-regression feature is more obvious.

From these results, we can find that AR(3), with smallest AIC and BIC, is the best fit, which is align with my hypothesis.