Week 02 Homework Report

Sneha Patkar

Problem 1

Given the dataset in problem1.csv:

a. calculate the first four moments values by using normalized formula in the "Week1 - Univariate Stats".

b. calculate the first four moments values again by using your chosen statistical package.

c. Is your statistical package functions biased? Prove or disprove your hypothesis respectively. Explain your conclusion.

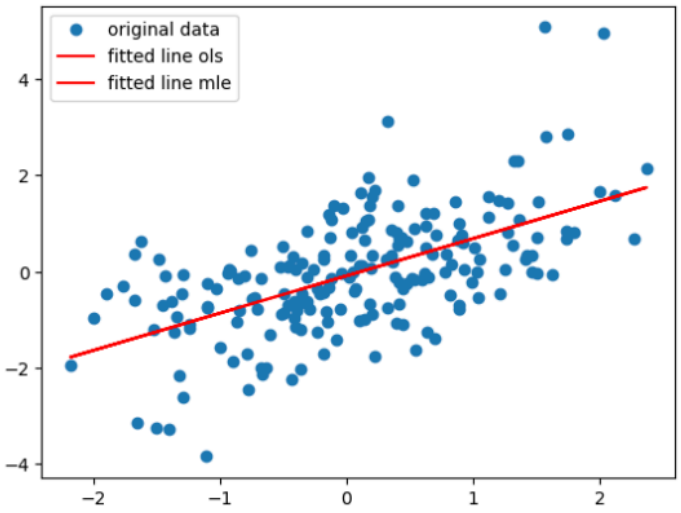
	Mean	Variance	Skew	Kurtosis
My Functions (biased implementation)	1.04897039	5.42179346	0.88060864	23.12220079
Package Functions (Numpy & SciPy)	1.04897039	5.42179346	0.88060864	23.12220079

Yes, my statistical package functions are biased because the numbers from the package match the numbers from my biased implementation.

Problem 2

a. Fit the data in problem2.csv using OLS. Then, fit the data using MLE given the assumption of normality. Compare their beta and standard deviation of the OLS error to the fitted MLE σ. What's your finding? Explain any differences.

	Intercept (B0)	Beta (B1)	Std Deviation
OLS	-0.08738446	0.77527410	1.00627516
MLE	-0.08738068	0.77527495	1.00375450



Beta and intercept for both approaches are nearly identical because of the assumption of normality. Minor variation is likely due to the optimizer function. However, there is a small difference in the std deviation. We see this because the variance is a biased estimate for MLE and is unbiased for Ordinary Least Squares.

b. 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 MLE under normality assumption and T distribution assumption. Which is the best of fit?

	Intercept (B0)	Beta (B1)	Std Deviation	LL	AIC
MLE (w/ Normal)	-0.08738068	0.77527495	1.00375450	-284.53756306	575.19757509
MLE (w/ T)	-0.09726932	0.67501023	0.85510566	-281.29340318	570.79193457

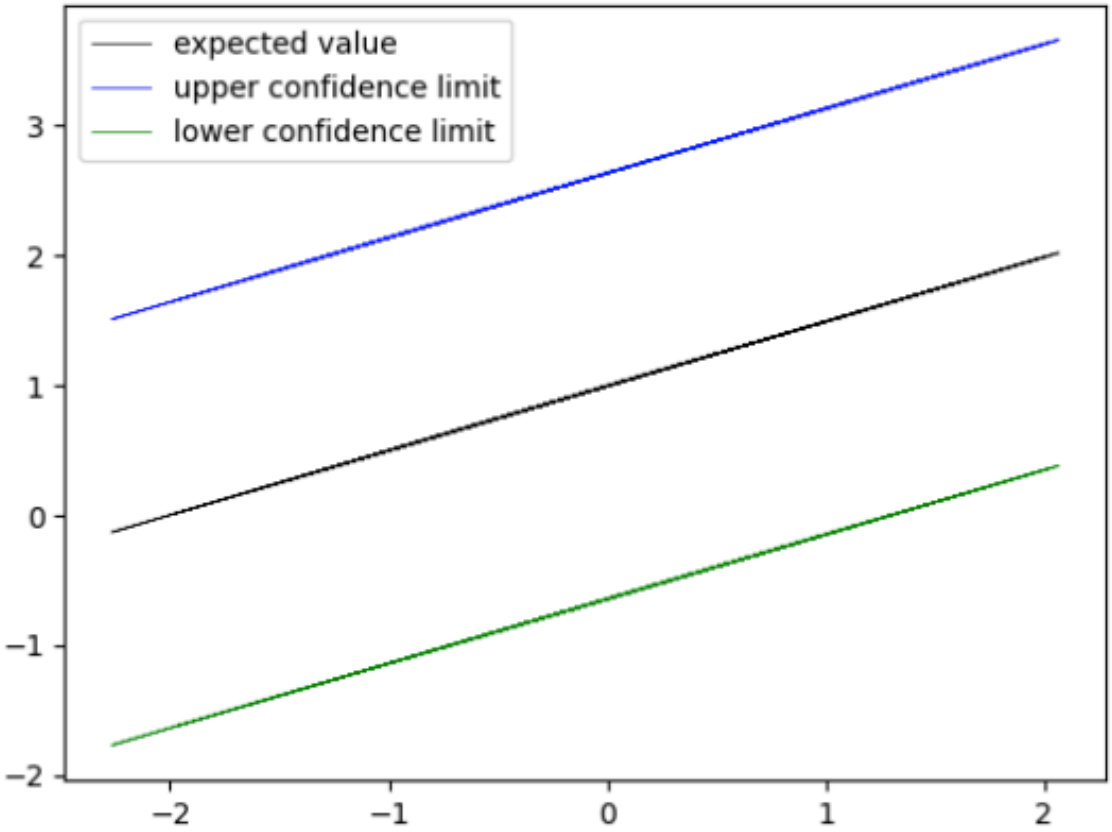
The better fit is with T distribution because it yields a lower AIC value.

c. Fit the data in problem2_x.csv using MLE given $XX = [XX1, XX2]$ follows the multivariate normal distribution. Assume X as a random variable, follows the fitted gaussian distribution,
$XX1$ (problem2_x1.csv) are a part of observed value of X, What's the distribution of XX_2 given each observed value? Plot the expected value along with the 95% confidence interval

The fitted parameters are the following:
Mean (mu x1): 0.001023
Mean (mu x2): 0.990244
Covariance (sigma): [[1.06977464 0.53068455]
[0.53068455 0.96147329]]

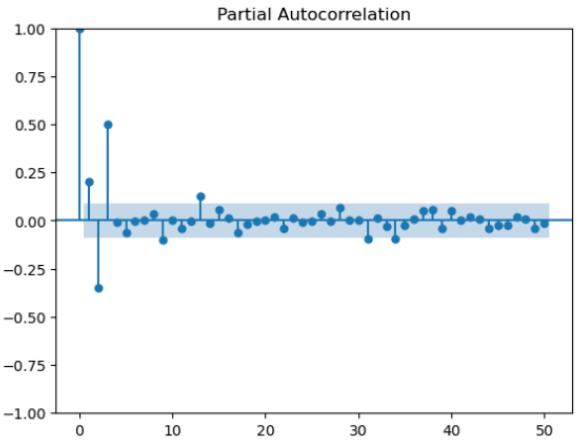
Sigma x1x1 1.06977464
Sigma x1x2 0.53068455

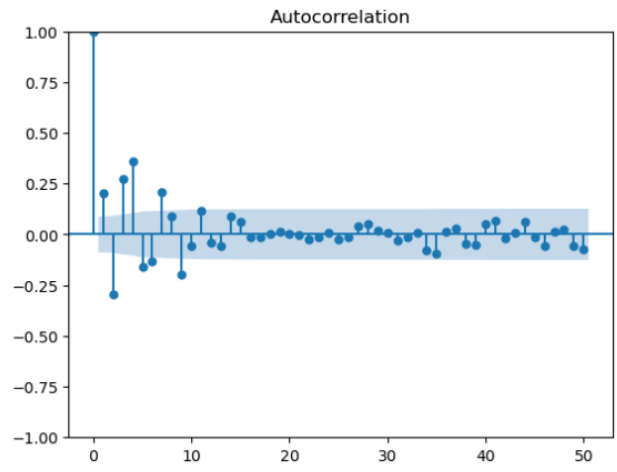
Sigma x2x1 0.53068455
Sigma x2x2 0.96147329



Problem 3

Fit the data in problem3.csv using AR (1) through AR (3) and MA (1) through MA (3), respectively. Which is the best of fit?





AIC Values for various Fits	
AR1	1644.65550476885
AR2	1581.07926590498
AR3	**1436.65980669459**
MA1	1567.40362637079
MA2	1537.94120638074
MA3	1536.86770873503

The best fit is AR3 given that it has the lowest AIC value and we clearly see 3 non-zero autocorrelations in the Partial-Autocorrelation plot above.