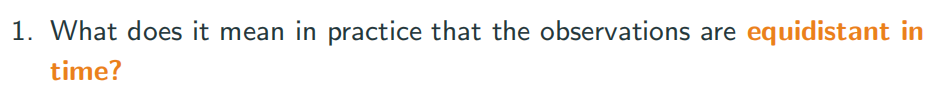
Discussion question 1a



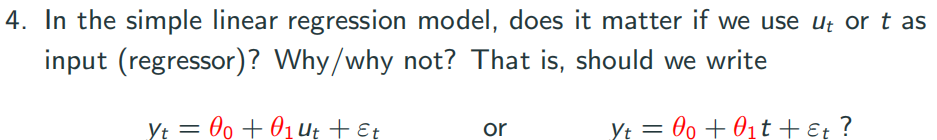
It means we can have the same time gap for each data point. Since we want to have the stable time series data



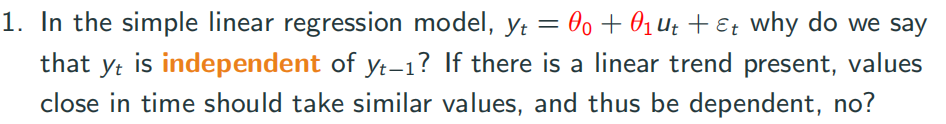
It is a constant.



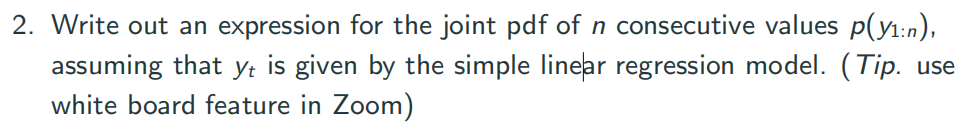
Ut = U0 + const \* (t-1)



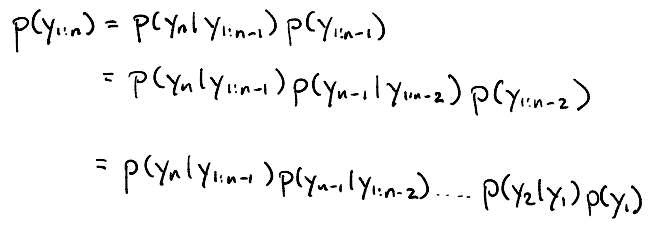
If time is equidistant, then it doesn’t matter.



,and we know that is a fixed number and  **is iid.** So they are independent.

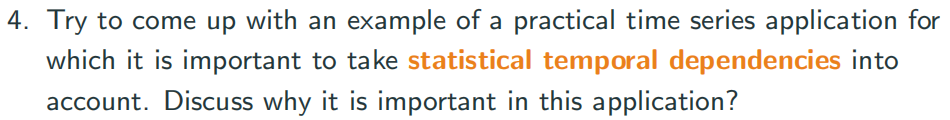


Frist, we have

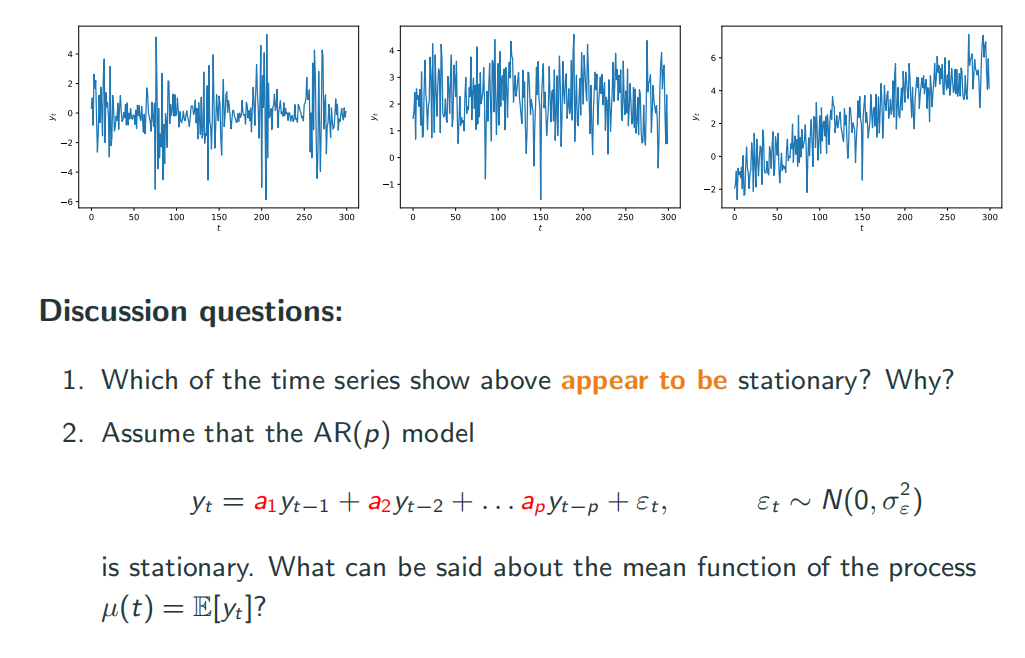


And since we proved that all **Y** are independent, so we can write

and we just multiply them together



Since in the real world data, data are correlated with previous data such as sale conditions and stock price. We need to module these information.



The second one is stationary.

For the first one, no. the mean value is fixed(lets say 0), but the auto-covariance is not only depends on the time lag. We can witness different covariance in different period.

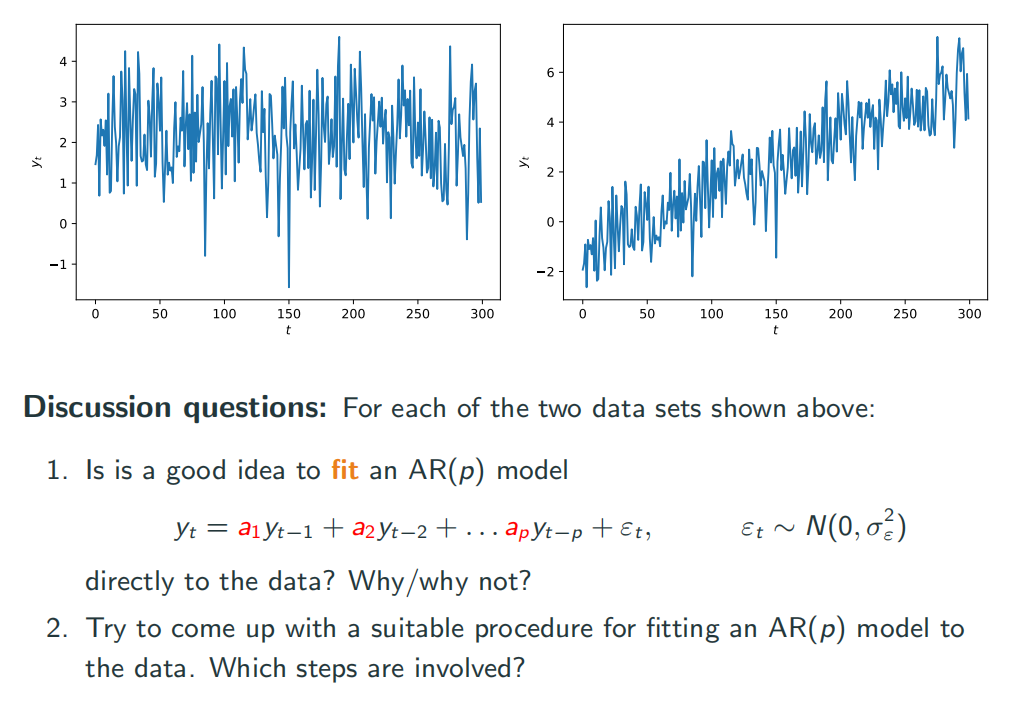
The second one yes.

Third one, no. The mean value is changing

2 If it is stationary, then all the y will have the same mean value.

So , we have = 0

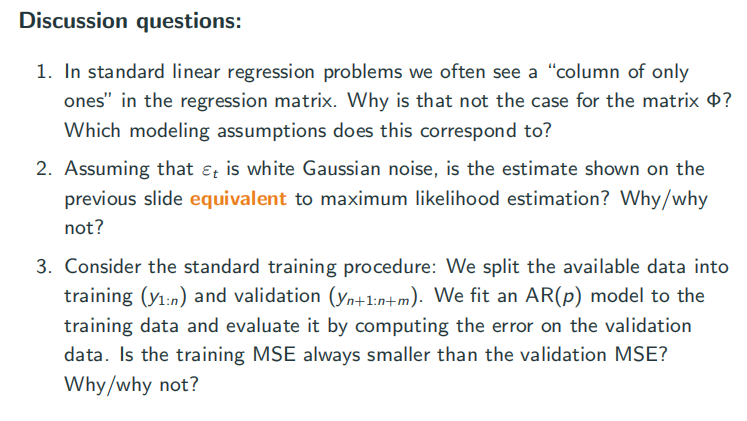
But in most time is hard to control the sum of parameters. So we will try to keep mean value to 0 to make it a stationary time series.



As we said before, it’s not good to directly feed them inside since we can not have stationary model for future data.

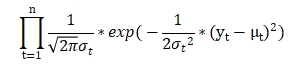
for the first data we need to subtract the mean value

for the second one, we can get a line from linear model and make original data subtract that line.



1 the “column of only ones” works as bias making the line be able to have intercept. However in time series scenario, we have the data mean 0, so we dont need bias. And if there is a bias, the future data will receive accumulated bias from previous calculation. When time goes long, it will face explosion.

2 In the Gaussian noise, we have the likelihood function



If we maximize this function, it’s the same as minimize



This is the same as minimize MSE in the OLS algorithm. So in the Gaussian noise, it’s the same. If we change to another distribution, the likelihood will be different, then they are not same.

3 Normally yes. Since in time series, our training data is normally previous time data and later data works as validation. So the distribution of these two part can be very different. This means the trend or pattern model learned from previous data may not work that well on latter data. But there is no guarantee that it will always work better, maybe we luckily construct a model that fits better on validation data.