# Solutions to Assignment

Rongfei Jin

April 23, 2025

## Question 1

We use the PACF to determine the order of the AR model. The PACF is defined as the correlation between $Y_t$ and $Y_{t-k}$ after removing the effect of all the intermediate variables $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-(k-1)}$.

Consider the $AR(p+1)$ model:

$$Y_t = a_0 + \sum_{k=1}^{p+1} a_k Y_{t-k} + \epsilon_t, \quad \epsilon_t$$

Since the true model is $AR(p)$, we have $a_{p+1} = 0$ and the PACF should be zero for all lags greater than $p$.

Now to determine the PACF, we can use the Yule-Walker equations, which relate the autocorrelations of the time series to the coefficients of the AR model and we can compute the significant levels to check if PACF is significantly different from zero.

See q1q2.py for the implementation.

# Question 2

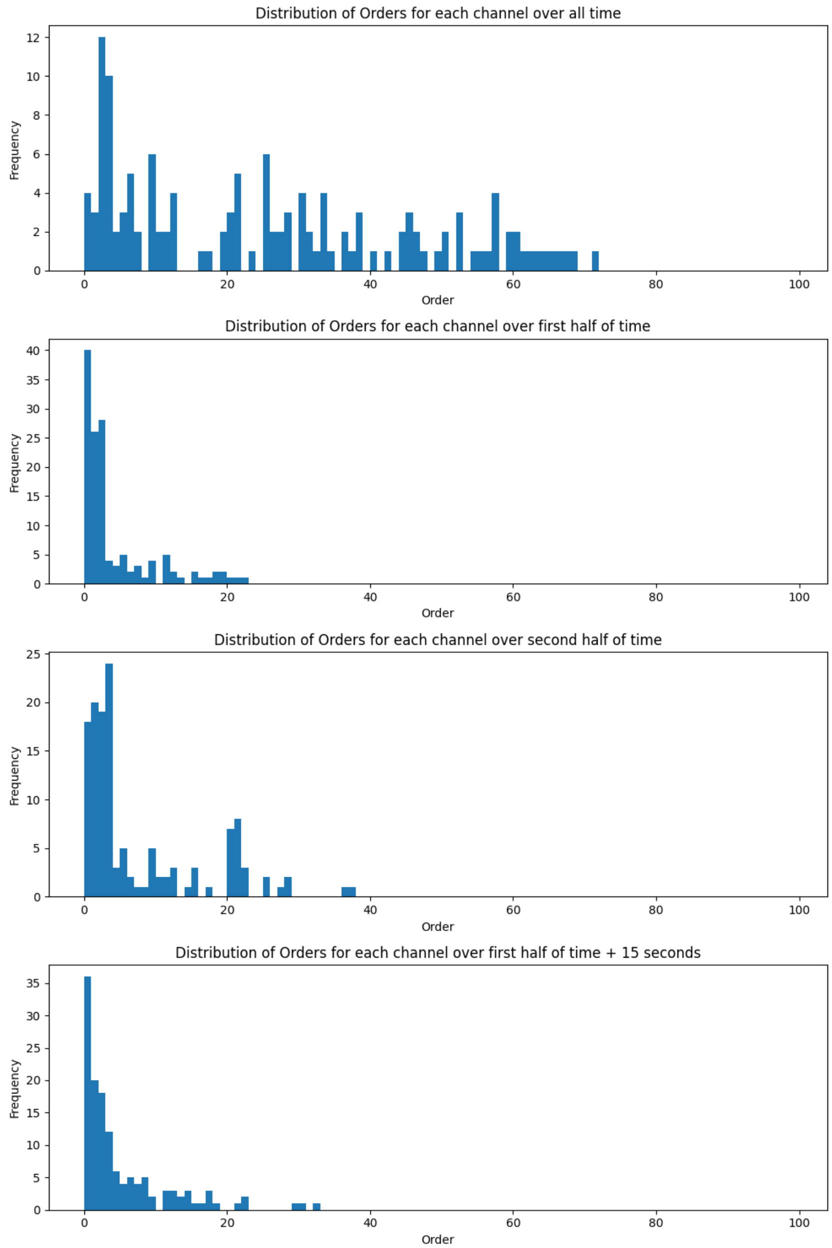


Figure 1: Distribution of Orders for each channel over all time, first half of time, and second half of time

We plot 4 histograms, the first one is the distribution of orders for each channel over all time, the second one is the distribution of orders for each channel over the first half of time, the third one is the distribution of orders for each channel over the second half of time, and the fourth one is the distribution of orders for each channel over the first half of time + 15 seconds. We choose the onset time to be the first half of the data as we expect to see a different pattern between the two halves. The postictal data has more channels matching the AR order of 20, while the preictal data does not, this suggests that the epileptic activity changes the brain signal. We can also see some channels with greather than 20 orders when comparing the first half of time and the first half of time + 15 seconds, which may suggest the effect of the epileptic activity is not immediate.

# Question 3

We first consdier the Autocovariance function

$$\gamma(h) = \mathrm{Cov}(y_t, y_{t-h}) = \mathrm{E}[y_t y_{t-h}] - \mathrm{E}[y_t]\,\mathrm{E}[y_{t-h}]$$

and the correlation function

$$\mathrm{Corr}(y_t, y_{t-h}) = \frac{\mathrm{Cov}(y_t, y_{t-h})}{\sqrt{\mathrm{Var}(y_t)\,\mathrm{Var}(y_{t-h})}}$$

The ACF is defined as the correlation between $y_t$ and $y_{t-h}$ after removing the effect of all the intermediate variables $y_{t-1}, y_{t-2}, \ldots, y_{t-(h-1)}$. We express the ACF as

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

The PACF is defined as the correlation between $y_t$ and $y_{t-h}$ after removing the effect of all the intermediate variables $y_{t-1}, y_{t-2}, \ldots, y_{t-(h-1)}$. We express the PACF for a staionary time series as

$$\phi_{11} = \gamma(1)$$

and

$$\phi_{hh} = \mathrm{Corr}(y_{t+h} - \hat{y}_{t+h}, y_{t-h} - \hat{y}_{t-h})$$

where $\hat{y}_{t+h}$ and $\hat{y}_{t-h}$ are the predicted values of $y_{t+h}$ and $y_{t-h}$ using the regression model.

$$\hat{y}_{t+h} = \sum_{i=1}^{h-1} \beta_i y_{t+h-i}$$

$$\hat{y}_t = \sum_{i=1}^{h-1} \beta_i y_{t+i}$$

And the we can use the ACF and PACF to determine the order of the AR, MA, or ARMA model.

| Function | AR(p) | MA(q) | ARMA(p,q) |
|---|---|---|---|
| ACF | Tails off | Cuts off after lag q | Tails off |
| PACF | Cuts off after lag p | Tails off | Tails off |

# Question 4

We first derive the ACF for MA(1) model.

$$\begin{aligned}\gamma(1) &= \mathrm{Cov}(y_t, y_{t-1}) = \mathrm{Cov}(\epsilon_t - \theta_1\epsilon_{t-1}, \epsilon_{t-1} - \theta_1\epsilon_{t-2}) \\ &= 0 + 0 + \theta_1\sigma^2 + 0 = \theta_1\sigma^2 \quad \epsilon_k \text{ are independent}\end{aligned}$$

$$\gamma(0) = \mathrm{Var}(y_t) = \mathrm{Var}(\epsilon_t - \theta_1\epsilon_{t-1}) = \sigma^2 + \theta_1^2\sigma^2 = (1+\theta_1^2)\sigma^2$$

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{\theta_1\sigma^2}{(1+\theta_1^2)\sigma^2} = \frac{\theta_1}{1+\theta_1^2}$$

We estimate the ACF using the sample ACF, and we solve the equation for $\hat{\theta}_1$

$$\hat{\theta}_1 = \frac{1 \pm \sqrt{1 - 4\hat{\rho}(1)^2}}{2\hat{\rho}(1)}$$

We then choose the solution $|\hat{\theta}_1| < 1$ for MA(1) model to be invertible, which is important for the equivalent AR model to be causal for further prediction.

See q4q5.py for the implementation.

# Question 5

The simulated data is actually MA(5) model, so both (i) and (ii) models are unable to capture the MA(5) model and produce a incorrect estimate of the parameters. By plotting the ACF and PACF, we can see that the ACF tails off and the PACF cuts off after lag 5, which is the order of the MA(5) model so with the proper order, we can estimate the parameters correctly.

See q4q5.py for the implementation.

# Question 6

We use auto-arima model to find the best model for each channel and plot the ACF and PACF of the residuals. And AR(3) models is the best model for most channels. We also notice that the the largest order of the AR model is 7 where the largest order of the MA model is 4. Based on our knowledge of the epileptic activity and brain signal, it is reasonable to infer that the brain signals have different frequency components. Since the data is collected from subject with epilepsy, we might expect the same pattern of the brain signal among other subjects.

# Question 7

Note that $x_k, \mathbf{x_{-k}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Theta}^{-1})$, we can partition the mean and covariance matrix as

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_k \\ \boldsymbol{\mu}_{-k} \end{pmatrix}$$

$$\boldsymbol{\Theta} = \begin{pmatrix} \boldsymbol{\Theta}_{kk} & \boldsymbol{\Theta}_{k,-k} \\ \boldsymbol{\Theta}_{-k,k} & \boldsymbol{\Theta}_{-k,-k} \end{pmatrix}$$

Notice that the kernel of the joint distribution of $x_k$ and $\mathbf{x_{-k}}$ is

$$\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Theta(\mathbf{x}-\boldsymbol{\mu})\right)$$

We can expand the quadratic form as

$$(\mathbf{x}-\boldsymbol{\mu})^T\Theta(\mathbf{x}-\boldsymbol{\mu}) = \Theta_{kk}(x_k-\mu_k)^2 + 2\Theta_{k,-k}(x_k-\mu_k)(\mathbf{x_{-k}}-\boldsymbol{\mu}_{\mathbf{-k}}) + (\mathbf{x_{-k}}-\boldsymbol{\mu}_{\mathbf{-k}})^T\Theta_{-k,-k}(\mathbf{x_{-k}}-\boldsymbol{\mu}_{\mathbf{-k}})$$

Now the conditional expectation of $x_k$ given $\mathbf{x_{-k}}$ is related to the $x_k$ terms in the kernel of the joint distribution of $x_k$ and $\mathbf{x_{-k}}$. We can rewrite the terms in the kernel as

$$-\frac{1}{2}\left((x_k-\mu_k)^2\Theta_{kk} + 2(x_k-\mu_k)(\mathbf{x_{-k}}-\boldsymbol{\mu}_{\mathbf{-k}})\Theta_{k,-k}\right)$$

We can complete the square as

$$-\frac{1}{2}\Theta_{kk}\left((x_k-\mu_k+\frac{\Theta_{k,-k}}{\Theta_{kk}}(\mathbf{x_{-k}}-\boldsymbol{\mu}_{\mathbf{-k}}))^2\right)$$

Based on the kernel, we can derive the conditional mean as

$$\mathrm{E}[x_k|\mathbf{x_{-k}}] = \mu_k + \frac{\Theta_{k,-k}}{\Theta_{kk}}(\mathbf{x_{-k}}-\boldsymbol{\mu}_{\mathbf{-k}})$$

## Question 8

To estimate the $\mathbf{\Theta}$, first notice that the previous expectation shows a linear relationship between $x_k$ and $\mathbf{x_{-k}}$. We can rewrite the expectation as

$$\hat{x}_k = \mu_k + \boldsymbol{\beta}^T(\mathbf{x_{-k}} - \boldsymbol{\mu}_{\mathbf{-k}}) + \epsilon_k$$

where $\boldsymbol{\beta} = \frac{\Theta_{k,-k}}{\Theta_{kk}}$ and $\epsilon_k \sim \mathcal{N}(0, \frac{1}{\Theta_{kk}})$.

Now we can use the lasso regression to estimate the $\mathbf{\Theta}$ based on the linear relationship.

We let $\hat{\mu}_k = x'_k$ and $\hat{\boldsymbol{\mu}}_{-k} = \mathbf{x}'_{-k}$ where $x'_k$ and $\mathbf{x}'_{-k}$ are the sample data, we estimate the $\frac{1}{\Theta_{kk}}$ as the inverse of the variance of the residuals.

$\Theta_{kk} = \frac{\beta}{var(\epsilon_k)}$

See q8.py for the implementation.

## Question 9

We first write the pdf of multivariate normal distribution as

$$f(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} \det(\mathbf{\Sigma})^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$$

The log likelihood function is

$$\log L(\boldsymbol{\mu}, \mathbf{\Sigma}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(\det(\mathbf{\Sigma})) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$$

By using the determiant property, we can rewrite the log likelihood function as

$$\log L(\boldsymbol{\mu}, \mathbf{\Sigma}) = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\log(\det(\mathbf{\Theta}|)) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$$

Notice that the last term is a scalar, so we can use associative properties of trace to rewrite the log likelihood function as

$$\log L(\boldsymbol{\mu}, \mathbf{\Theta}) = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\log(\det(\mathbf{\Theta})) - \frac{1}{2}\operatorname{tr}(\mathbf{S\Theta})$$

We now omit the constant term since we only consider the maximization with respect to $\mathbf{\Theta}$.

$$\log\det(\mathbf{\Theta}) - \mathbf{S\Theta}$$

Finally we can add the penalty term to the log likelihood function to get the penalized log likelihood function.

$$\log\det(\mathbf{\Theta}) - \mathbf{S\Theta} + \rho\|\mathbf{\Theta}\|_1$$

## Question 10

The difference between the two methods is the first method assumes the paritition of the sample precision matrix is the best estimator of the precision matrix but it is not the maximum likelihood estimator, while the second method is based on the maximum likelihood estimator of the precision matrix.

One advantage of the first method is that can be easily run in parallel, while the second method needs to compute the whole matrix. One limitation of the first method is that it may not produce a positive definite matrix. So a good senario for the first method is when the sample size is large and the sample precision matrix is close to the true precision matrix.