lab6

April 23, 2024

1 AR(p) model fitting

[1]:
```
import numpy as np
import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt
```

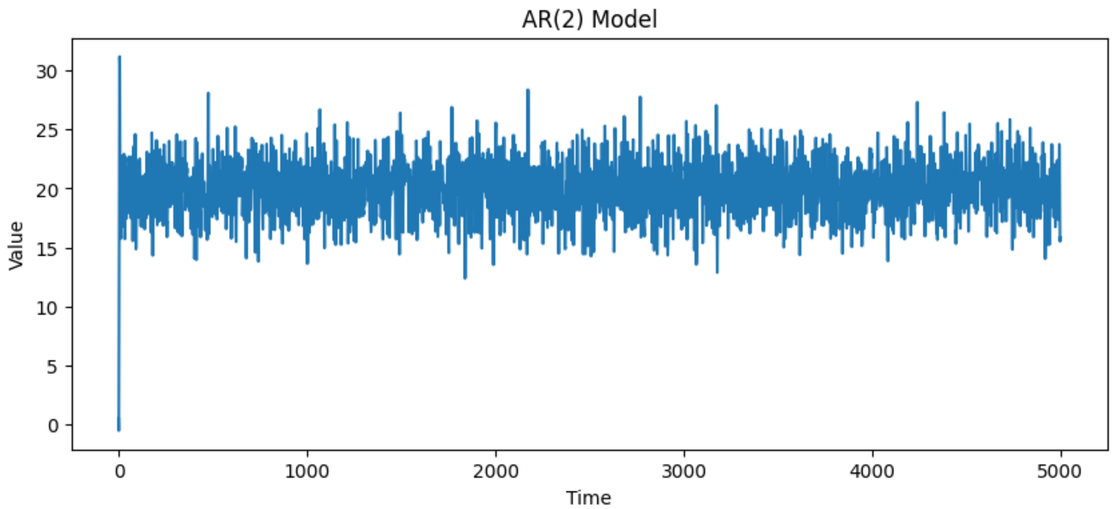
AR function from previous lab

[2]:
```
def ar_function(p, n, phi, c, burnin=0):
    yt = np.zeros(n)
    eps = np.random.normal(0, 1, n + p)
    yt[:p] = eps[:p]
    for i in range(p, n):
         yt[i] = c + np.dot(phi, yt[i-p:i][::-1]) + eps[i]
    return yt[burnin:]
```

[3]:
```
phi1 = 1.3
phi2 = -0.7
c = 8
n = 5000
yt_ar2 = ar_function(2, n, [phi1, phi2], c)
```

Plot the time series

[4]:
```
plt.figure(figsize=(10, 4))
plt.plot(yt_ar2)
plt.title('AR(2) Model')
plt.xlabel('Time')
plt.ylabel('Value')
plt.show()
```



Model selection - Find the model with minimum AIC

[5]:
```
AIC_values = []
for p in range(1, 5):
    model = sm.tsa.ARIMA(yt_ar2, order=(p, 0, 0)).fit()
    AIC_values.append((p, model.aic))

best_model = min(AIC_values, key=lambda x: x[1])
```

[6]:
```
for p, aic in AIC_values:
    print(f"AR({p}) model: AIC = {aic}")

print("\nBest Model:")
print(f"AR({best_model[0]}) model with AIC = {best_model[1]} is the preferred␣
 ↪model.")
```

```
AR(1) model: AIC = 17630.9712286693
AR(2) model: AIC = 14362.927061550417
AR(3) model: AIC = 14364.542717825676
AR(4) model: AIC = 14365.91843308229

Best Model:
AR(2) model with AIC = 14362.927061550417 is the preferred model.
```

Perform log-likelihood ratio test between AR(2) and the best model

[7]:
```
from scipy.stats import chi2

model_ar2 = sm.tsa.ARIMA(yt_ar2, order=(2, 0, 0)).fit()
model_best = sm.tsa.ARIMA(yt_ar2, order=(best_model[0], 0, 0)).fit()
llr = 2 * (model_best.llf - model_ar2.llf)
```

```python
df = best_model[0] - 2  # Difference in degrees of freedom
p_value = 1.0 - chi2.cdf(llr, df)
print(f"\nP-value for LLRT between AR(2) and the best model: {p_value}")
if p_value < 0.05:
    print("Reject the null hypothesis that the AR(2) model fits better.")
else:
    print("Fail to reject the null hypothesis.")
```

```
P-value for LLRT between AR(2) and the best model: nan
Fail to reject the null hypothesis.
```

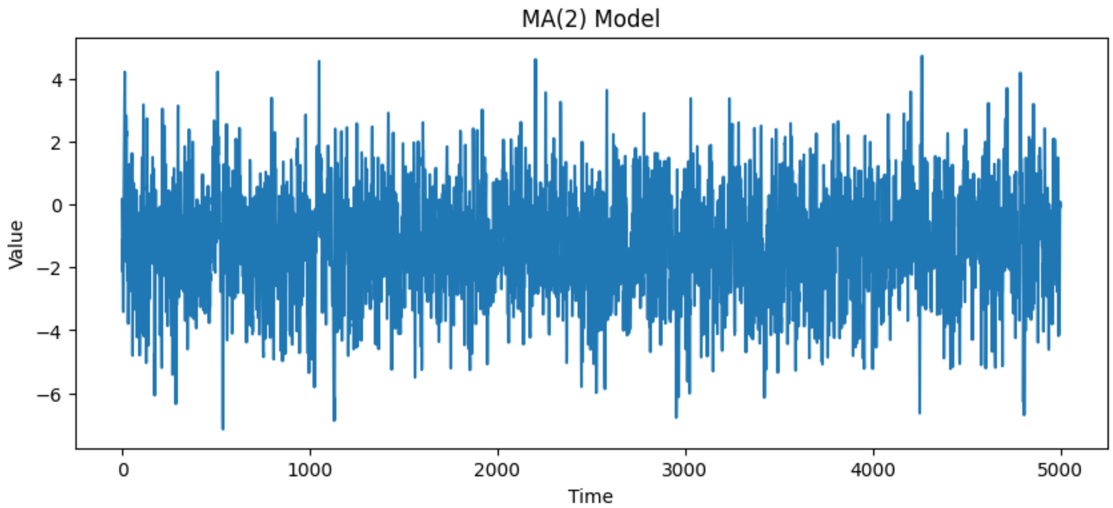
2 MA(q) model fitting

[8]:
```python
def generate_MA(q, n, theta, c, burnin=0):
    yt = np.zeros(n)
    eps = np.random.normal(0, 1, n + q)
    yt[:q] = eps[:q]  # Initialize with random noise
    for i in range(q, n):
        yt[i] = c + np.dot(theta, eps[i-q:i][::-1]) + eps[i]
    return yt[burnin:]
```

[9]:
```python
theta_ma2 = [1, 0.8]
n = 5000
c = np.random.normal(0, 1)
yt_ma2 = generate_MA(2, n, theta_ma2, c)
```

Plot the time series

[10]:
```python
plt.figure(figsize=(10, 4))
plt.plot(yt_ma2)
plt.title('MA(2) Model')
plt.xlabel('Time')
plt.ylabel('Value')
plt.show()
```



Model selection - Find the model with minimum AIC

[11]:
```
AIC_values = []
for q in range(1, 5):
    model = sm.tsa.ARIMA(yt_ma2, order=(0, 0, q)).fit()
    AIC_values.append((q, model.aic))

best_model = min(AIC_values, key=lambda x: x[1])
```

[12]:
```
for q, aic in AIC_values:
    print(f"MA({q}) model: AIC = {aic}")

print("\nBest Model:")
print(f"MA({best_model[0]}) model with AIC = {best_model[1]} is the preferred␣
 ↪model.")
```

```
MA(1) model: AIC = 16793.257719149406
MA(2) model: AIC = 14160.174597449328
MA(3) model: AIC = 14161.38714965257
MA(4) model: AIC = 14160.156285641586

Best Model:
MA(4) model with AIC = 14160.156285641586 is the preferred model.
```

Perform log-likelihood ratio test between MA(2) and the best model

[13]:
```
model_ma2 = sm.tsa.ARIMA(yt_ma2, order=(0, 0, 2)).fit()
model_best = sm.tsa.ARIMA(yt_ma2, order=(0, 0, best_model[0])).fit()
llr = 2 * (model_best.llf - model_ma2.llf)
df = best_model[0] - 2  # Difference in degrees of freedom
p_value = 1.0 - chi2.cdf(llr, df)
```

```python
print(f"\nP-value for LLRT between MA(2) and the best model: {p_value}")
if p_value < 0.05:
    print("Reject the null hypothesis that the MA(2) model fits better.")
else:
    print("Fail to reject the null hypothesis.")
```

```
P-value for LLRT between MA(2) and the best model: 0.13410182173720142
Fail to reject the null hypothesis.
```

3 Conclusions

- MA(4) - the best model for the moving average series: This suggests that the moving average process has a significant dependence on the previous four observations. The MA(4) model captures the noise structure present in the data more effectively compared to other MA models (MA(1), MA(2), MA(3)). The selection of MA(4) indicates that the data exhibits a pattern where the current observation depends on the weighted average of the previous four white noise terms.
- AR(2) - the best model for the autocorrelation series: This indicates that the autocorrelation structure of the data is best represented by an autoregressive model with two lag terms. The AR(2) model suggests that the current observation is linearly dependent on the two previous observations with a certain autoregressive coefficient. This means that the current value is influenced by both the immediate past and the past before that.
- Failure to reject the null hypothesis: In both cases, the log-likelihood ratio test resulted in a p-value higher than the significance level (0.05). This suggests that there is insufficient evidence to reject the null hypothesis that the simpler model (MA(2) for the MA series and AR(2) for the AR series) fits the data better. In other words, the more complex models (MA(4) and AR(2)) do not provide a significantly better fit to the data compared to the simpler models.