DS-GA 3001.001/002
Probabilistic Time Series Analysis
Lab 1: ACF, CCF, AR, MA

Caroline Haimerl

NYU

Logistics

- Online Section: Colin Bredenberg, W 9:00-9:50am
  - Zoom meeting code: 983 3177 9428
- Blended Section In-Person: Ashwin Siripurapu, W 3:30-4:20pm
  - 60 FA, Room 150
- Blended Section Online: Jiyuan (Rupert) Lu, W 3:30-4:20pm
  - Zoom meeting code: 942 1755 0183

Course Materials

- Git repo: https://github.com/savinteachingorg/pTSAFall2020
- Piazza
  - Ask questions and get answers from course staff, other students
- NYU Classes
  - View recorded lectures and lab sections

Please turn in your completed lab1.ipynb notebook on NYU Classes before 09/22/2020 11:59 pm.

Your work will be evaluated based on the code and plots. You don't need to write down your answers to these questions in the text blocks.

python dependency

- Python >= 3.6
- Numpy >= 1.13.3
- Pandas >= 0.20.3
- Statsmodels >= 0.8.0

basic models

White Noise

$$X_t \sim N(0, \sigma^2)$$

Moving Average

$$v_t = \frac{1}{3}(w_{t-1} + w_t + w_{t+1})$$

Autoregressive

$$x_t = x_{t-1} - .9x_{t-2} + w_t$$

basic measures of dependency

autocovariance

$$\gamma_x(s,t) = cov(x_s, x_t) = E[(x_s - \mu_s)(x_t - \mu_t)]$$

cross-covariance

$$\gamma_{xy}(s,t) = cov(x_s, y_t) = E[(x_s - \mu_x)\left(y_t - \mu_y\right)]$$

autocorrelation

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}$$

crosscorrelation

$$\rho_{xy}(s,t) = \frac{\gamma_{xy}(s,t)}{\sqrt{\gamma_x(s,s)\gamma_y(t,t)}}$$

basic measures of dependency - empirical

autocovariance

$$\widehat{\gamma_x}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x})$$

cross-covariance

$$\widehat{\gamma_{xy}}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(y_t - \bar{y})$$

autocorrelation

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

crosscorrelation

$$\hat{\rho}_{xy}(h) = \frac{\hat{\gamma}_{xy}(h)}{\sqrt{\hat{\gamma}_x(0)\hat{\gamma}_y(0)}}$$

ACF

exercise of part I. autocorrelation function

A) implement the ACF
B) ACF of white noise

ACF of moving average models

Given a Moving Average process $x_t = w_{t-1} + 2w_t + w_{t+1}$, where w_t are independent with zero means and variance σ_w^2, determine the autocorrelation function (ACF).

ACF of moving average models

Given a Moving Average process $x_t = w_{t-1} + 2w_t + w_{t+1}$, where w_t are independent with zero means and variance σ_w^2, determine the autocorrelation function (ACF).

- First of all, we need to calculate the autovariance $\gamma(t+h,t) = cov[(w_{t+h-1} + 2w_{t+h} + w_{t+h+1}), (w_{t-1} + 2w_t + w_{t+1})]$.
- Note that because of the independent property of w_t, $cov(w_s, w_t) = \begin{cases} 0 & s \neq t \\ \sigma_w^2 & s = t \end{cases}$
- We can separate the cov of two sums into the sum of several bi-variate covariances

$$cov(aX + bY, cW + dV) = ac * cov(X,W) + ad * cov(X,V) + bc * cov(Y,W) + bd * cov(Y,V)$$

- When $s = t$, we have

$$\begin{aligned}\gamma(t,t) &= cov[(w_{t-1} + 2w_t + w_{t+1}), (w_{t-1} + 2w_t + w_{t+1})] \\ &= cov(w_{t-1}, w_{t-1}) + cov(2w_t, 2w_t) + cov(w_{t+1}, w_{t+1}) \\ &= \sigma_w^2 + 4\sigma_w^2 + \sigma_w^2 \\ &= 6\sigma_w^2\end{aligned}$$

- When $s = t \pm 1$, we have

$$\begin{aligned}\gamma(t+1,t) &= cov[(w_t + 2w_{t+1} + w_{t+2}), (w_{t-1} + 2w_t + w_{t+1})] \\ &= cov(w_t, 2w_t) + cov(2w_{t+1}, w_{t+1}) \\ &= 2\sigma_w^2 + 2\sigma_w^2 \\ &= 4\sigma_w^2\end{aligned}$$

- When $s = t \pm 2$, we have

$$\begin{aligned}\gamma(t+2,t) &= cov[(w_{t+1} + 2w_{t+2} + w_{t+3}), (w_{t-1} + 2w_t + w_{t+1})] \\ &= cov(w_{t+1}, w_{t+1}) \\ &= \sigma_w^2\end{aligned}$$

- Therefore, our autocovariance is a function of lag $h = s - t$:

$$\gamma(h) = \begin{cases} 6\sigma_w^2 & h = 0 \\ 4\sigma_w^2 & h = \pm 1 \\ \sigma_w^2 & h = \pm 2 \end{cases}$$

- Using the definition of autocorrelation function (ACF), we have

$$\rho(h) = \frac{\gamma_x(h)}{\gamma_x(0)} = \begin{cases} 1 & h = 0 \\ \frac{2}{3} & h = \pm 1 \\ \frac{1}{6} & h = \pm 2 \end{cases}$$

ACF of moving average models

exercise of part I. autocorrelation function

C) ACF of moving average

ACF of signal in noise

exercise of part I. autocorrelation function

D) ACF of signal in noise

CCF

Suppose that series y_t is linearly determined by series x_t with a lag l: $y_t = Ax_{t-l} + w_t$. How can we determine the value of l?

CCF

$\rho_{xy}(h)$ gives us the dependency of the two series on each other with different time lags h

$\rho_{xy}(h)$ will reach its maximum when $h = l$

→ we can determine the lag l from the max/min of the CCF plot

CCF of signal in noise

exercise of part II. crosscorrelation function

A) CCF of signal with noise
B) CCF of data

autoregressive models

current value is a linear function of past values

Definition 3.1 *An* **autoregressive model** *of order p, abbreviated* **AR(p)**, *is of the form*

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t, \tag{3.1}$$

where x_t is stationary, $w_t \sim wn(0, \sigma_w^2)$, and $\phi_1, \phi_2, \ldots, \phi_p$ are constants ($\phi_p \neq 0$).

autoregressive models

ACF of an AR(1) is

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

autoregressive models

ACF of an AR(2)?

autoregressive models

ACF of an AR(2)?

Suppose $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$ is a causal AR(2) process. Multiply each side of the model by x_{t-h} for $h > 0$, and take expectation:

$$\mathrm{E}(x_t x_{t-h}) = \phi_1 \mathrm{E}(x_{t-1} x_{t-h}) + \phi_2 \mathrm{E}(x_{t-2} x_{t-h}) + \mathrm{E}(w_t x_{t-h}).$$

autoregressive models

The result is

$$\gamma(h) = \phi_1 \gamma(h-1) + \phi_2 \gamma(h-2), \quad h = 1, 2, \ldots . \tag{3.38}$$

In (3.38), we used the fact that $E(x_t) = 0$ and for $h > 0$,

$$E(w_t x_{t-h}) = E\Big(w_t \sum_{j=0}^{\infty} \psi_j w_{t-h-j}\Big) = 0.$$

autoregressive models

Divide (3.38) through by $\gamma(0)$ to obtain the difference equation for the ACF of the process:

$$\rho(h) - \phi_1 \rho(h-1) - \phi_2 \rho(h-2) = 0, \quad h = 1, 2, \ldots. \tag{3.39}$$

now compute $\rho(1)$ and $\rho(2)$

autoregressive models

exercise of part III. AR

- determine p given ACF and PCF
- fit $AR(p)$
- use p datapoints to predict $tsteps$ into the future
- plot ACF and PCF of $AR(p)$ model
- relate $AR(p)$ parameters to ACF of data for lag 0 to p

moving average models

filered white noise

average over previous and following value

$$v_t = \frac{1}{3}\left(w_{t-1} + w_t + w_{t+1}\right)$$

moving average models

exercise of part IV. moving average

look at two processes and try to predict whether they are the same or not

then look at ACFs

Please turn in your completed notebook on NYU Classes before 09/22/2020 11:59 pm.

Your work will be evaluated based on the code and plots. You don't need to write down your answers to these questions in the text blocks.

python dependency

- Python >= 3.6
- Numpy >= 1.13.3
- Pandas >= 0.20.3
- Statsmodels >= 0.8.0