DS-GA 3001.009
Modeling Time Series Data
Lab 1

Artie Shen | Center Data Science

NYU

Logistics

- **Time & Location**
  - 6:45 PM - 7:35 PM, Thursday, 60 FA C12
- **TA Office Hour**
  - 11:00 AM - 12:00 PM, Thursday, 60 FA 660
- **Lab Agenda**
  - Recitation / Lecture
  - Interactive coding work
- **Submission**
  - Submit your lab work before **6:45 PM at the following Thursday**

Agenda

- Recap
  - Basic Time Series Models
  - Measures of Dependence
  - Stationary
  - Sample Statistics
- Exercise
- Programming
  - ACF Analysis
  - CCF Analysis

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**

$$x_t = x_{t-1} - 0.9x_{t-2} + w_t$$

- **Random Walk**

$$x_t = \delta + x_{t-1} + w_t$$

NYU

Measure of Dependency

- **Autocovariance**
  - Measures the **linear dependence** between two points on the same series

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

- **Autocorrelation (ACF)**

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

- **Cross-covariance**
  - Measures the linear dependence between two series

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

- **Cross-correlation (CCF)**
  - $$\rho_{xy}(s,t) = \frac{\gamma_{xy}(s,t)}{\sqrt{\gamma_x(s,s)\gamma_y(t,t)}}$$

Stationary

- **Strictly Stationary**
  - For any two collections of values: $\{x_{t_1}, x_{t_2}, \ldots, x_{t_k}\}\{x_{t_1+h}, x_{t_2+h}, \ldots, x_{t_k+h}\}$
  - Their joint CDF match

$$Pr\{x_{t_1} \leq c_1, \ldots, x_{t_k} \leq c_k\} = Pr\{x_{t_1+h} \leq c_1, \ldots, x_{t_k+h} \leq c_k\}$$

- **Weakly Stationary**
  - Mean is constant
  - ACF only depends on lag and variance is finite.
  - Let h = s - t,

$$\gamma(s,t) = cov(x_s, x_t) = cov(x_{t+h}, x_t) = cov(x_h, x_0) = \gamma(h)$$

$$\rho(s,t) = \frac{\gamma(t+h,t)}{\sqrt{\gamma(t+h,t+h)\gamma(t,t)}} = \frac{\gamma(h)}{\gamma(0)} = \rho(h)$$

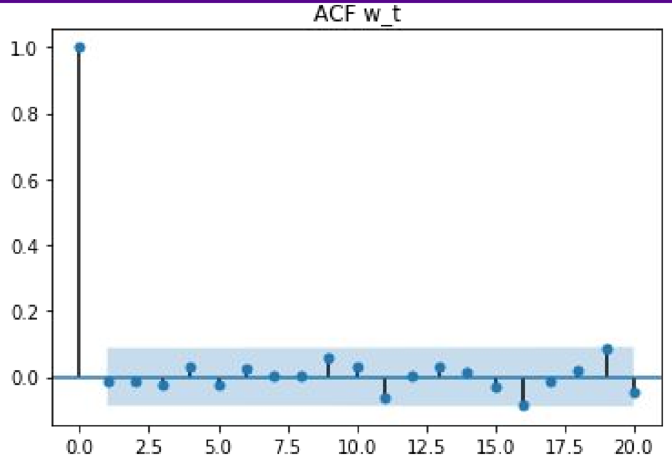
  - ACF is symmetric around the origin:

$$\rho(h) = \rho((t+h)-t) = cov(x_{t+h}, x_t) = cov(x_t, x_{t+h}) = \rho(t-(t+h)) = \rho(-h)$$

Measure of Dependency

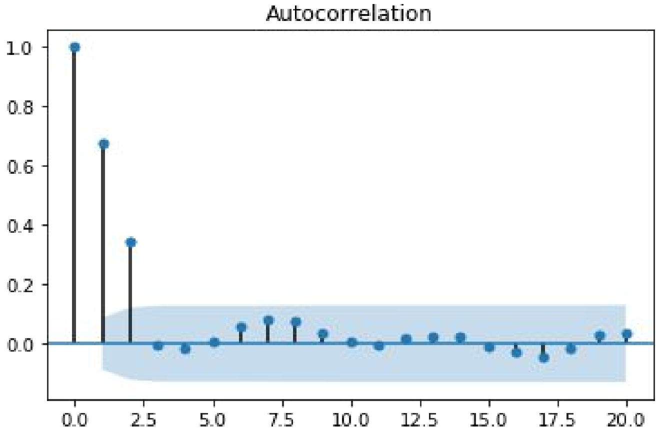
- **ACF of a white noise**

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



- **ACF of a Moving Average**

$$v_t = \frac{1}{3}(w_t, w_{t+1}, w_{t+2})$$



Sample Statistics

- **Sample Mean**

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t \qquad var(\bar{x}) = \frac{1}{n}\sum_{h=-n}^{n}(1 - \frac{|h|}{n})\gamma_x(h)$$

- **Sample Autocovariance / Cross-covariance**

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

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

- **Sample ACF / CCF**

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

Prediction Using 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?

- By definition, cross-covariance is defined as $\gamma_{yx}(h) = cov(y_{t+h}, x_t) = cov(Ax_{t+h-l} + w_{t+h}, x_t) = cov(Ax_{t+h-l}, x_t) = A\gamma_x(h-l)$
- As a matter of fact, $\gamma_x(h-l)$ reaches its maximum when $h = l$.
- Since $\gamma_{yx}(h)$ is linearly depends on $\gamma_x(h-l)$, at $h = l$, $\gamma_{yx}(h)$ will also reaches at its maximum / minimum depends on A.
- We can determine the lag l by locating the maximum / minimum on the CCF plot.

Derivation 1.7

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 break the covariance of two summation into the sum of several bi-variate covariance using property of covariance:

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

Derivation 1.7 (contd)

- 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

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

Derivation 1.7 (contd)

- 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

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

Derivation 1.8

For a Random Walk process $x_t = \delta + x_{t-1} + w_t$, where w_t is white noise with variance σ_w^2,

- (a) find the mean and autovariance function of x_t
- (b) argue that x_t is not stationary
- (c) show $\rho_x(t-1, t) = \sqrt{\frac{t-1}{t}}$
- (d) suggest a transformation to make the series stationary

Derivation 1.8 (cntd.)

- First of all, this series can be re-written as $x_t = \delta t + \sum_{k=1}^{t} w_k$.
- $\mu_t = E(x_t) = E(\delta t) + \sum_{k=1}^{t} E(w_k) = \delta t$
- Remember that because of the independent property of w_t, $cov(w_s, w_t) = \begin{cases} 0 & s \neq t \\ \sigma_w^2 & s = t \end{cases}$
-

$$\begin{aligned} \gamma(s,t) &= cov(\delta s + \sum_{k=1}^{s} w_k, \delta t + \sum_{j=1}^{t} w_j) \\ &= cov(\sum_{k=1}^{s} w_k, \sum_{j=1}^{t} w_j) \\ &= cov(w_1 + w_2 + \ldots + w_s, w_1 + w_2 + \ldots + w_t) \\ &= min(s,t)\sigma_w^2 \end{aligned}$$

Derivation 1.8 (cntd)

- $\rho_x(t-1,t) = \frac{\gamma(t-1,t)}{\sqrt{\gamma(t-1,t-1)\gamma(t,t)}} = \frac{(t-1)\sigma_w^2}{\sqrt{(t-1)t\sigma_w^2}} = \sqrt{\frac{t-1}{t}}$. As t grows, the value of x_t linearly depends more on its previous turn x_{t-1}.

- One transformation technique is differencing. $\hat{x}_t = x_t - x_{t-1} = \delta + w_t$. $E(\hat{x}_t) = \delta$

  $\gamma(x_t, x_s) = cov(\delta + w_t, \delta + w_s) = cov(w_t, w_s) = \begin{cases} 0 & s \neq t \\ \sigma_w^2 & s = t \end{cases}$

Programming

- **Github:**
  - **https://github.com/charlieblue17/timeseries2018**
- **Dependency**
  - Python 3.6
  - Numpy >= 1.13.3
  - Pandas >= 0.20.3
  - Statsmodels >= 0.8.0
- **Due Date 02/01/2018 06:45 pm on NYU Classes**