DS-GA 3001.009
Modeling Time Series Data
Lab 2

Artie Shen | Center for Data Science

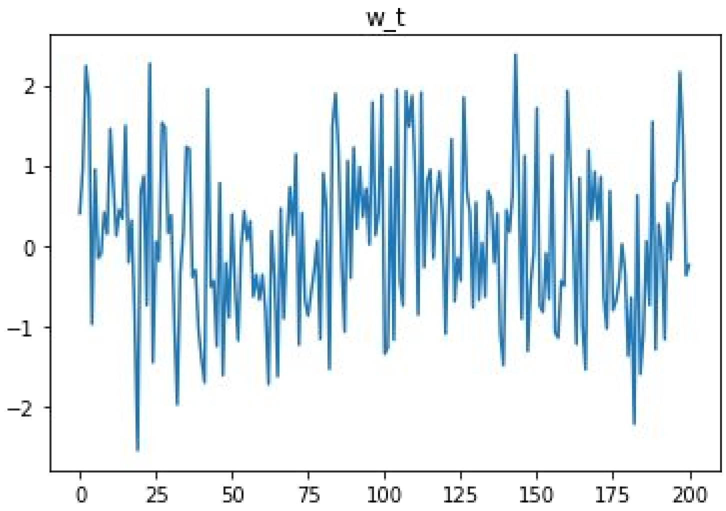
NYU

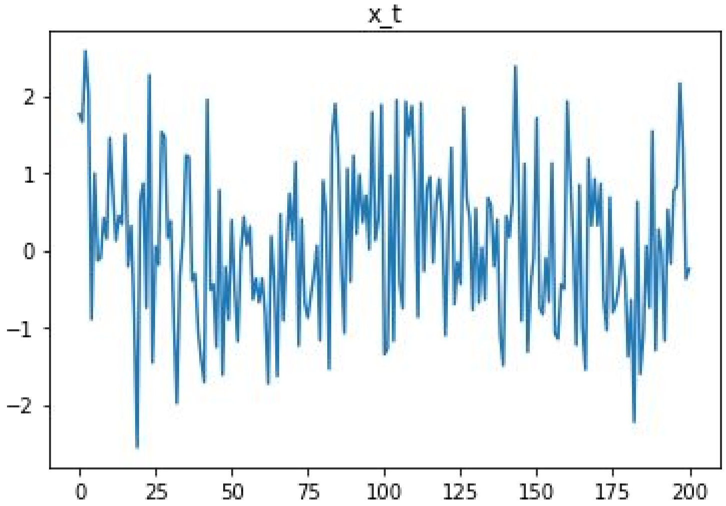
Agenda

- Recap
  - Polynomial Form
  - Forecasting
  - Parameter Estimation

- Programming
  - Yule-Walker Equations
  - Durbin-Levinson Algorithm
  - One-step forward Forecasting

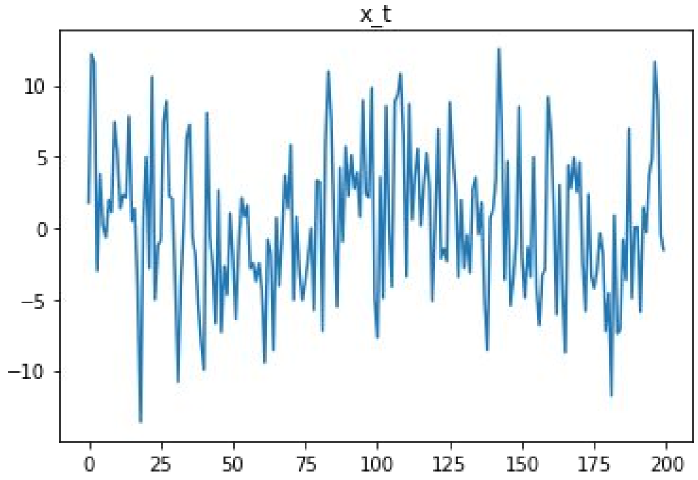
- **Motivation - Parameter Redundancy**
  - $x_t = 0.5x_{t-1} - 0.5w_{t-1} + w_t$
  - Is x_t the same as white noise w_t ?

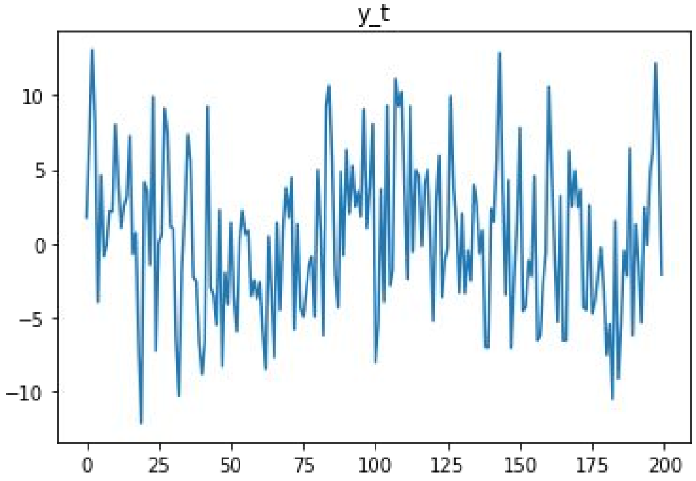




*

- **Motivation - MA Non-uniqueness**
  - $x_t = w_t + \frac{1}{5}w_{t-1}, w_t \sim N(0, 25)$
  - $y_t = v_t + 5v_{t-1}, v_t \sim N(0, 1)$
  - Are those two MA process the same?





Polynomial Form

- **Parameter Redundancy**
  - When a model can be expressed using a simple model.
  - How can we detect parameter redundancy?

- **Other Problems**
  - Stationary series that depends on future
  - MA models that are not unique

Polynomial Form

- **Backshift Operator**

Definition 2.4 *We define the* **backshift operator** *by*

$$Bx_t = x_{t-1}$$

and extend it to powers $B^2 x_t = B(Bx_t) = Bx_{t-1} = x_{t-2}$*, and so on. Thus,*

$$B^k x_t = x_{t-k}.$$

Not a function, nor a polynomial. It’s an operator that behaves like a polynomial.

Polynomial Form

- **Example**

$$x_t = 0.5x_{t-1} - 0.5w_{t-1} + w_t$$

$$x_t - 0.5x_{t-1} = w_t - 0.5w_{t-1}$$

$$x_t - 0.5Bx_t = w_t - 0.5Bw_t$$

$$(1 - 0.5B)x_t = (1 - 0.5B)w_t$$

- **ARMA Model**

Definition 3.5 *A time series* $\{x_t;\ t = 0, \pm 1, \pm 2, \ldots\}$ *is* **ARMA**(p, q) *if it is stationary and*

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}, \tag{3.19}$$

with $\phi_p \neq 0$, $\theta_q \neq 0$, *and* $\sigma_w^2 > 0$. *The parameters p and q are called the autoregressive and the moving average orders, respectively. If* x_t *has a nonzero mean* μ, *we set* $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$ *and write the model as*

- **Polynomial Form**

Definition 3.6 *The* **AR and MA polynomials** *are defined as*

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p, \quad \phi_p \neq 0,$$

and

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q, \quad \theta_q \neq 0,$$

respectively, where z is a complex number.

Polynomial Form

- **Causality**

Property 3.1 Causality of an ARMA(p, q) Process

An ARMA(p, q) model is causal if and only if $\phi(z) \neq 0$ for $|z| \leq 1$. The coefficients of the linear process given in (3.25) can be determined by solving

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1.$$

Another way to phrase Property 3.1 is that *an ARMA process is causal only when the roots of $\phi(z)$ lie outside the unit circle*; that is, $\phi(z) = 0$ only when $|z| > 1$. Finally, to address the problem of uniqueness discussed in Example 3.6, we choose the model that allows an infinite autoregressive representation.

Polynomial Form

- **Invertibility**

Definition 3.8 *An ARMA(p, q) model is said to be* **invertible**, *if the time series* $\{x_t;\ t = 0, \pm 1, \pm 2, \dots\}$ *can be written as*

$$\pi(B)x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j} = w_t, \tag{3.26}$$

where $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$, *and* $\sum_{j=0}^{\infty} |\pi_j| < \infty$; *we set* $\pi_0 = 1$.

Another way to phrase Property 3.2 is that *an ARMA process is invertible only when the roots of* $\theta(z)$ *lie outside the unit circle*; that is, $\theta(z) = 0$ only when $|z| > 1$. The proof of Property 3.1 is given in Section B.2 (the proof of Property 3.2 is similar). The following examples illustrate these concepts.

Polynomial Form

- **Example**

Consider the process

$$x_t = .4x_{t-1} + .45x_{t-2} + w_t + w_{t-1} + .25w_{t-2},$$

or, in operator form,

$$(1 - .4B - .45B^2)x_t = (1 + B + .25B^2)w_t.$$

- **Example - Redundancy**

At first, x_t appears to be an ARMA(2, 2) process. But notice that

$$\phi(B) = 1 - .4B - .45B^2 = (1 + .5B)(1 - .9B)$$

and

$$\theta(B) = (1 + B + .25B^2) = (1 + .5B)^2$$

have a common factor that can be canceled. After cancellation, the operators are $\phi(B) = (1 - .9B)$ and $\theta(B) = (1 + .5B)$, so the model is an ARMA(1, 1) model, $(1 - .9B)x_t = (1 + .5B)w_t$, or

$$x_t = .9x_{t-1} + .5w_{t-1} + w_t. \tag{3.27}$$

- **Example - Invertibility & Causality**

$$\phi(B) = (1 - .9B) \text{ and } \theta(B) = (1 + .5B);$$

The model is causal because $\phi(z) = (1 - .9z) = 0$ when $z = 10/9$, which is outside the unit circle. The model is also invertible because the root of $\theta(z) = (1 + .5z)$ is $z = -2$, which is outside the unit circle.

Proof in textbook section 3.1

Given a time series $X = \{x_t\}$ and an $AR(p)$ model:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi x_{t-p} + w_t$$

- **Model Selection**
  - Select the model family (AR/MA/ARMA/...)
  - Select hyper-parameters. (AR:p, MA:p, ARMA:p,q)
- **Parameter Estimation**
  - Given $X = \{x_t\}$, model family, and hyper-parameters find the set of model parameters $\{\phi_i\}$ that minimizes some loss function.
- **Forecast**
  - Given model and model parameters $\{\phi_i\}$ and a set of data, calculate x_{t+1}.

- **Gaussian Process**
  - $\mathbf{X} \sim N(\mu, \Sigma)$
  - $p_X(x) = (2\pi)^{-n/2}(det\Sigma)^{-\frac{1}{2}} exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))$
  - ARMA, by definition, is a Linear Gaussian Process.
- **Best Linear Predictor for AR(p)**
  - $$\begin{aligned} \hat{x}_{t+1} &= argmax_{x_{t+1}}(p_X(x_{t+1}|x_t,\dots,x_{t-p+1})) \\ &= E(x_{t+1}|x_t,\dots,x_{t-p+1}) \\ &= \mu + \phi_1(x_t-\mu)+\dots+\phi_p(x_{t-p+1}-\mu) \end{aligned}$$

*Reference: http://www.stat.columbia.edu/~rdavis/papers/VAG002.pdf

- **Method 1: MLE**
  - Our predictor $\hat{x}_{t+1} = f_\phi(x_t, x_{t-1}, \ldots, x_{t-p+1})$
  - Likelihood

$$L(\Sigma_n, \mu, \phi) = (2\pi)^{-n/2}(v_0 v_1 \ldots v_{n-1})^{-1/2} exp(-\frac{1}{2}\sum_{j=1}^{n} \frac{(x_j - \hat{x}_j)^2}{v_{j-1}})$$

  - MLE Estimator

$$\phi^* = argmax_\phi L(\Sigma_n, \mu, \phi)$$

Reference: http://www.stat.columbia.edu/~rdavis/papers/VAG002.pdf

- **Method 2: Method of Moments**

Definitions.

(1) $E(X^k)$ is the k^{th} **(theoretical) moment** of the distribution (**about the origin**), for $k = 1, 2, ...$

(2) $E\left[(X-\mu)^k\right]$ is the k^{th} **(theoretical) moment** of the distribution (**about the mean**), for $k = 1, 2, ...$

(3) $M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ is the k^{th} **sample moment**, for $k = 1, 2, ...$

(4) $M_k^* = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$ is the k^{th} **sample moment about the mean**, for $k = 1, 2, ...$

Reference: https://onlinecourses.science.psu.edu/stat414/node/193

Parameter Estimation

The basic idea behind this form of the method is to:

(1) Equate the first sample moment about the origin $M_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ to the first theoretical moment $E(X)$.

(2) Equate the second sample moment about the origin $M_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ to the second theoretical moment $E(X^2)$.

(3) Continue equating sample moments about the origin, M_k, with the corresponding theoretical moments $E(X^k)$, $k = 3, 4, ...$ until you have as many equations as you have parameters.

(4) Solve for the parameters.

*Reference: https://onlinecourses.science.psu.edu/stat414/node/193

- **Use Method of Moments on AR(p) model**
  - Express moments as functions of ϕ
  - Replace theoretical moments with empirical moments
  - Solve for ϕ
- **Assumption**
  - AR(p) process with $\mu = 0$

Parameter Estimation

- **Step 1:** Express moments as functions of ϕ

Consider the general AR(p)

$$x_{i+1} = \phi_1 x_i + \phi_2 x_{i-1} + \cdots + \phi_p x_{i-p+1} + \xi_{i+1}.$$

- multiply by x_{i-p-1},

$$x_{i-p+1} x_{i+1} = \sum_{j=1}^{p} \left(\phi_j x_{i-p+1} x_{i-j+1}\right) + x_{i-p+1} \xi_{i+1},$$

- take expectance,

$$\langle x_{i-p+1} x_{i+1} \rangle = \sum_{j=1}^{p} \left(\phi_j \langle x_{i-p+1} x_{i-j+1} \rangle\right) + \langle x_{i-p+1} \xi_{i+1} \rangle$$

*Reference: http://www-stat.wharton.upenn.edu/~steele/Courses/956/ResourceDetails/YWSourceFiles/YW-Eshel.pdf

Parameter Estimation

- **Step 1:** Express moments as functions of ϕ
  - eliminate the zero correlation forcing term

$$\langle x_{i-p+1} x_{i+1} \rangle = \sum_{j=1}^{p} \left(\phi_j \langle x_{i-p+1} x_{i-j+1} \rangle \right)$$

  - divide through by $(N-1)$, and use $c_{-l} = c_l$,

$$c_p = \sum_{j=1}^{p} \phi_j c_{j-p}$$

  - divide through by c_o,

$$r_p = \sum_{j=1}^{p} \phi_j r_{j-p}.$$

Reference: http://www-stat.wharton.upenn.edu/~steele/Courses/956/ResourceDetails/YWSourceFiles/YW-Eshel.pdf

Parameter Estimation

- $$r_p = \sum_{j=1}^{p} \phi_j r_{j-p}.$$

Rewriting all the equations together yields

$$\begin{array}{ccccccccccccc}
r_1 & = & \phi_1 r_o & + & \phi_2 r_1 & + & \phi_3 r_2 & + & \cdots & + & \phi_{p-1} r_{p-2} & + & \phi_p r_{p-1} \\
r_2 & = & \phi_1 r_1 & + & \phi_2 r_o & + & \phi_3 r_1 & + & \cdots & + & \phi_{p-1} r_{p-3} & + & \phi_p r_{p-2} \\
 & & & & & & \vdots & & & & & & \\
r_{p-1} & = & \phi_1 r_{p-2} & + & \phi_2 r_{p-3} & + & \phi_3 r_{p-4} & + & \cdots & + & \phi_{p-1} r_o & + & \phi_p r_1 \\
r_p & = & \phi_1 r_{p-1} & + & \phi_2 r_{p-2} & + & \phi_3 r_{p-3} & + & \cdots & + & \phi_{p-1} r_1 & + & \phi_p r_o
\end{array}$$

Reference: http://www-stat.wharton.upenn.edu/~steele/Courses/956/ResourceDetails/YWSourceFiles/YW-Eshel.pdf

Parameter Estimation

which can also be written as

$$\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_{p-1} \\ r_p \end{pmatrix} = \begin{pmatrix} r_o & r_1 & r_2 & \cdots & r_{p-2} & r_{p-1} \\ r_1 & r_o & r_1 & \cdots & r_{p-3} & r_{p-2} \\ & \vdots & & & \vdots & \\ r_{p-2} & r_{p-3} & r_{p-4} & \cdots & r_o & r_1 \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & r_1 & r_o \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{p-1} \\ \phi_p \end{pmatrix}.$$

Recalling that $r_o = 1$, the above equation is also

$$\underbrace{\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_{p-1} \\ r_p \end{pmatrix}}_{\mathbf{r}} = \underbrace{\begin{pmatrix} 1 & r_1 & r_2 & \cdots & r_{p-2} & r_{p-1} \\ r_1 & 1 & r_1 & \cdots & r_{p-3} & r_{p-2} \\ & \vdots & & & \vdots & \\ r_{p-2} & r_{p-3} & r_{p-4} & \cdots & 1 & r_1 \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & r_1 & 1 \end{pmatrix}}_{\mathbf{R}} \underbrace{\begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{p-1} \\ \phi_p \end{pmatrix}}_{\mathbf{\Phi}}$$

or succinctly

$$\mathbf{R}\mathbf{\Phi} = \mathbf{r}.$$

Yule-Walker Equations

Reference: http://www-stat.wharton.upenn.edu/~steele/Courses/956/ResourceDetails/YWSourceFiles/YW-Eshel.pdf

Parameter Estimation

- **Step 2:** Replace theoretical moments with empirical moments
  - Replace auto-correlation r_i with empirical auto-correlations $\hat{r}_i$
  - $R\Phi = r$ now becomes $\hat{R}\Phi = \hat{r}$
  - We can estimate $\hat{R}$ and $\hat{r}$ from data!

Reference: http://www-stat.wharton.upenn.edu/~steele/Courses/956/ResourceDetails/YWSourceFiles/YW-Eshel.pdf

Parameter Estimation

- **Step 3:** Solve for ϕ
  - $\hat{\Phi} = \hat{R}^{-1}\hat{r}$
  - Direct matrix inverse takes $O(n^3)$
  - Use **Durbin-Levinson Algorithm**

Reference: http://www-stat.wharton.upenn.edu/~steele/Courses/956/ResourceDetails/YWSourceFiles/YW-Eshel.pdf

Parameter Estimation

- **Durbin-Levinson Algorithm**
  - Input: observed data $\{x_t\}$
  - Output: p by p **triangular** matrix Φ,where $\Phi_{i,j}$ denotes the estimated parameter for jth lag term in a AR(j) model.

- **Durbin-Levinson Algorithm (continued)**
  - Use data calculate empirical ACF $\hat{\rho}(i)$
  - Initialize $\phi_{00} = 0$,
  - Populate the rest

$$\phi_{nn} = \frac{\rho(n) - \sum_{k=1}^{n-1} \phi_{n-1,k}\, \rho(n-k)}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k}\, \rho(k)},$$

$$\phi_{nk} = \phi_{n-1,k} - \phi_{nn}\phi_{n-1,n-k}, \quad k = 1, 2, \ldots, n-1.$$

Programming

- Github:
  - https://github.com/charlieblue17/timeseries2018
- Due Date 02/08/2018 06:45 pm on NYU Classes
- Please rename your submission to **net_id.ipynb**