Time Series Analysis Class Notes

Dustin Leatherman

January 25, 2020

Contents

1 Characteristics of Time Series (2020/01/09)

- Must be correlation between data points which limits conventional statistical analysis.

- One variable, x_t, will be used in this course

Important Questions to Ask

- What patterns are visible over time?
- How can correlation between observations be used to help with the model?
- Can future state be predicted using this data?

Problem: We don't know how many previous time points should be used to predict the current value.

General Tips

- if non-constant variance, transform the predictors
- Find assumptions, then continue modeling
- Time is generally treated as discrete values instead of continuous

Stochastic Process: collection of random variables, x_t, indexed by t

- **Realization**: Realization of a stochastic process.

Time Series: collection of randome variables indexed and ordered by time

White Noise: $w_t \sim N(0, \sigma_w^2)$

One way to "smooth" a time series is to introduce a moving average.

MA(1): $x_t = \beta w_{t-1} + w_t$

AR(1): $x_t = \beta x_{t-1} + w_t$

$$\begin{aligned} E(x_t) =& E(\beta X_{t-1} + w_t) \\ =& \beta E(x_{t-1}) + E(w_t) \\ =& \ldots \\ =& 0 \end{aligned} \tag{1}$$

- $0 \leq \beta \leq 1$

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

if s == t, $cov(x_s, x_s) = var(x_s)$

$\gamma(s,t) = \begin{cases} \sigma_w^2 & s = t \\ 0 & s \neq t \end{cases}$

- given $w_t \sim ind\ N(0, \sigma_w^2)$

1.1 Moving Average

Let $m_t = \frac{w_t + w_{t-1} + w_{t-2}}{3}$

$$\begin{aligned} E[(m_s - \mu_s)(m_t - \mu_t)] =& E(m_s m_t) \\ =& \frac{1}{9} E[(w_s + w_{s-1} + w_{s-2})(w_t + w_{t-1} + w_{t-2})] \end{aligned} \tag{2}$$

<u>s = t</u>

$$\begin{aligned} E(m_t^2) =& var(m_t) + E(m_t)^2 \\ =& \frac{1}{9} var(w_t + w_{t-1} + w_{t-2}) + 0 \\ =& \frac{1}{9}(var(w_t) + var(w_{t-1} + var(w_{t-2}))) \\ =& \frac{1}{9}(1 + 1 + 1) \\ =& \frac{3}{9} \end{aligned} \tag{3}$$

<u>s = t - 1</u>: $E(m_{t-1}, m_t) = \frac{2}{9}$
<u>s = t - 2</u>: $E(m_{t-2}, m_t) = \frac{1}{9}$
<u>s = t - 3</u>: $E(m_{t-3}, m_t) = 0$

$$\gamma(s,t) = \begin{cases} \frac{3}{9} & s = t \\ \frac{2}{9} & |s-t| = 1 \\ \frac{1}{9} & |s-t| = 2 \\ 0 & |s-t| \geq 3 \end{cases}$$

1.2 Autocorrelation

$\rho_{xy} = \frac{cov(x,y)}{\sqrt{var(x)}\sqrt{var(y)}}$

AR: $\rho(s,t) = \begin{cases} 1 & s = t \\ 0 & s \neq t \end{cases}$

MA: $\rho(s,t) = \begin{cases} 1 & s = t \\ \frac{2}{3} & |s-t| = 1 \\ \frac{1}{3} & |s-t| = 2 \\ 0 & |s-t| \geq 3 \end{cases}$

positve linear dependence = smooth negative linear dependence = choppy

1.3 Stationarity

Strict stationary time series: the probabalistic behavior of $x_t, ..., x_{tk}$ os the exact same as the shifted set $x_{t+h}, ..., x_{tk+h}$ for any collection of time points $[t_1, t_k]$ for any \$k = 1, 2,. . . \$

$P(x_q \leq c_1, x_2 \leq c_2) = P(x_{10} \leq c_q, x_{11} \leq c_2)$

This is almost never used in practice because it is *too* strict.

Weakly Stationary Time Series: The first two moments (mean, covariance) of the time series are invariant to time shifts

$E(x_t) = \mu \forall t$

$\gamma(t, t+h) = \gamma(0, h) \forall t$

- μ and $\gamma(0, h)$ are *not* functions of t
- Assumption of **Equal Variance**
- $\gamma(h) = \gamma(-h)$ if weakly stationary

$$\begin{aligned}\rho(t, t+h) =& \frac{\gamma(t, t+h)}{\sqrt{\gamma(t,t)}\sqrt{\gamma(t+h, t+h)}} \\ =& \frac{\gamma(h)}{\sqrt{\gamma(0)}\sqrt{\gamma(0)}} \\ =& \frac{\gamma(h)}{\gamma(0)}\end{aligned} \tag{4}$$

Is there a correlation between lags? $H_0: \ \rho(h) = 0$ $H_A: \ \rho(h) \neq 0$

Sample Mean: $\bar{x} = \frac{1}{n}\Sigma x_t$

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

2 Time Series Regression, Exploratory Data Analysis, and ARIMA Models (2020/01/16)

2.1 Differences

Taking differences between successive values helps remove trend to help bring a time series to stationarity.

1st diff - $x_t = x_t - x_{t-1}$ (removes linear trend)

2nd diff - $(x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = x_t - x_{t-1} =^2 x_T$ (removes quadratic trend)

Proof $x_t - x_{t-1} = \beta_0 + \beta_1 t - [\beta_0 + \beta_1(t-1)] = \beta_1$
Order of Attempt

1. Transformation
2. Differencing

2.1.1 Backshift

- $Bx_t = x_{t-1}$
- $B^k x_t = x_{t-k}$

$$\begin{aligned}(1-2B+B^2)x_t \\ =& x_t - 2x_{t-1} + x_{t-2} = \\ (x_t - x_{t-1}) - (x_{t-1} - x_{t-2})\end{aligned} \tag{5}$$

A MA model can be expressed using Backshift operators and subsequently, expressed as an AR model.

$$\begin{aligned} m_t =& \frac{w_t + w_{t-1} + w_{t-2}}{3} \\ =& \frac{1}{3}(1 + B + B^2)w_t \end{aligned} \tag{6}$$

1. Properties
   - $BC = C$ for constant C
   - (1 - B) $\mathrm{x_t = x_t}$$
   - $(B \times B) = B^2$
   - $(1-B)^2 x_t =^2 x_t$
   - $(1-B)^0 x_t = x_t$
   - $(1-B)x_t$ - considered a linear filter since it filters out linear trend. i.e. first difference

2.1.2 MA(1)

$x_t = w_t + \theta_1 + w_{t-1} = (1 + \theta_1 B)w_t$ (AR Model Form)

$$\begin{aligned} &(1 - 0.7B)(1 - B)x_t = w_t \\ \rightarrow &(1 - 1.7B + 0.7B^2)x_t = w_t \\ \rightarrow &(x_t = 1.7x_{t-1} - 0.7x_{t-2} + w_t) \end{aligned} \tag{7}$$

Aside: Time series predicts future values. Regression is for estimation within known values.

2.1.3 Functional Differencing

Use $-0.5 \leq B \leq 0.5$ to do differencing

long memory: for $h \rightarrow \infty$, $\rho(h) \rightarrow 0$ *slowly* **short memory**: for $h \rightarrow \infty$, $\rho(h) \rightarrow 0$ *quickly*

2.2 ARIMA

AR-I-MA

AR: Autoregressive I: Integrated (differencing) MA: Moving Average

2.2.1 AR(1)

Uses p past observations to predict future observations. The preset value is predicted by a linear combination of previous time points.

$x_t = \phi_t x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + w_t$

$[\phi_1, \phi_p]$ - unknown parameters

$(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p)x_t = w_t$

$\rightarrow \phi(B)x_t = w_t$

$$\begin{aligned} x_t =& \phi x_{t-2} + w_t \\ =& \phi(\phi x_{t-2} + w_{t-1}) + w_t \\ =& \phi^2(x_{t-2} + \phi w_{t-1} + w_t) \\ ...& \\ =& \sum_{j=0}^{\infty} \phi^j w_{t-j} \end{aligned} \tag{8}$$

$$
\begin{aligned}
E(x_t) =& E(\sum_{j=0}^{\infty} \phi^j w_{t-j}) = 0 \\
\gamma(x_t) =& E(X_t x_{t+h}) - E(x_t)E(x_{t+h}) \\
=& E(x_t x_{t+h}) \text{ when } \mu = 0 \\
\gamma(0) =& \sum_{j=0}^{\infty} \phi^j w_{t-j} \\
=& \sum_{j=0}^{\infty} \phi^{2j} var(w_{t-j}) \\
=& \sigma_w^2 \sum_{j=0}^{\infty} \phi^{2j} = \frac{\sigma_w^2}{1-\phi^2} \text{ where } h = 0
\end{aligned}
\tag{9}
$$

$$
\begin{aligned}
\gamma(h) =& \frac{\phi^h \sigma_w^2}{1-\phi^2} \\
\rho(h) =& \frac{\gamma(h)}{\gamma(0)} = \phi^h
\end{aligned}
\tag{10}
$$

Given $|\phi| < 1$, an AR(1) Model can be expressed as a MA(1) Model (i.e. a sum of w_t's).

2.2.2 MA(1)

$$
\gamma(h) = \begin{cases} \sigma_w^2(1+\theta_1^2) & h = 0 \\ \theta_1 \sigma_w^2 & h = 1 \\ 0 & h \geq 2 \end{cases}
\tag{11}
$$

$$
\rho(h) = \begin{cases} 1 & h = 0 \\ \frac{\theta_1}{(+\theta_1^2)} & h = 1 \\ 0 & h > 1 \end{cases}
\tag{12}
$$

2.2.3 ARMA(p, q)

$(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p)x_t = (1 + \theta_1 B + \theta_2 B^2 + ... + \theta_q B^q)w_t$
$\rightarrow \phi(B)x_t = \theta(B)w_t$ assuming $\mathrm{x_t}$ is stationary

1. Parameter Redundancy Because AR and MA models can be converted back and forth, parameter redundancy can occur. For example, $ARMA(2,1) == AR(1)$. This mostly happens for theoretical data but R will throw an error if this happens. Can use polyroot() to debug.

3 ARMA Models (2020/01/23)

ARIMA models are reduced to ARMA after differencing.

3.1 AR(p)

$$x_t = (\sum_{j=1}^{p} \phi_j x_{t-j}) + \epsilon \tag{13}$$

3.2 MA(q)

$$\begin{aligned} x_t =& (+\theta_1 B + \theta_2 B^2 + ... + \theta_q B^q) w_t \\ =& (\sum_{j=0}^{q} \theta_j B^j) w_t,\ s.t.\ \theta_0 = 1,\ w_t \sim\ ind.\ N(0\sigma_w^2)\ for\ t = 1, ..., n \end{aligned} \tag{14}$$

$$\begin{aligned} ACF =& \gamma(h) = cov(x_t, x_{t+h}) \\ =& E(x_t x_{t+h}) - E(x_t)E(x_{t+h}) \\ =& E(x_t x_{t+h}) \\ =& ... \\ =& \sigma_w^2 \sum_{i=0}^{q-h} \theta_i \theta_{i+h},\ \text{if j = i + h} \end{aligned} \tag{15}$$

Thus

$$\gamma(h) = \begin{cases} \sigma_w^2 \sum_{i=0}^{q-h} \theta_i \theta_{i+h}, & 0 \leq h \leq q \\ 0, & h > q \end{cases} \tag{16}$$

$$\gamma(0) = \sigma_w^2 \sum_{i=0}^{q} \theta_i^2 \tag{17}$$

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \begin{cases} (\sum_{i=0}^{q} \theta_i^2)^{-1} \sum_{i=0}^{q-h} \theta_i \theta_{i+h}, & 0 \leq h \leq q \\ 0, & h > q \end{cases} \tag{18}$$

3.3 ACF & PACF

These plots are used to find values at lags $h = 0,1,2,...$ for a specific ARMA process. These values can be compared with the *observed* values to determine the appropriate model to use for the data.

ACF plot helps determine q for a MA(q) model. PACF plot helps determine p for an AR(p) model.

PACF plots will "cut off to 0" for an AR(p) model whereas ACF plots will not. ACF plots for AR(p) models are act like PACF plots for MA(q) models.

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off to 0	Cuts off to 0 after lag q	Tails off to 0 after q lags
PACF	Cuts off to 0 after lag p	Tails off to 0	Tails off to 0 after p lags

Examining these plots is the first step to constructing an ARMA model.

3.3.1 PACF

β's are called partial autocorrelations because they measure the dependence of x_t on x_{t+h} removing the effect all other random variables in between. These can be treated like regular correlations.

$$\begin{aligned} \beta_{11} =& Corr(x_t, x_{t+1}) \\ \beta_{22} =& Corr(x_t, x_{t+2}|x_{t+1}) \\ \beta_{33} =& Corr(x_t, x_{t+3}|x_{t+1}, x_{t+2})... \end{aligned} \tag{19}$$

3.4 ARMA(p, q)

Since AR(p) and MA(q) processes are interoperable, it is easier to deal with them as one.

$$\begin{aligned} &\phi(B)x_t = \theta(B)w_t \\ \rightarrow &x_t = [\theta(B)/\phi(B)]w_t \\ \rightarrow &x_t = \psi(B)w_t \\ \rightarrow &\psi(B) = 1 + B\psi_1 + B^2\psi_w + ...+,\ \psi_0 = 1,\ E(x_i) = 0 \end{aligned} \tag{20}$$

$\psi(B)$ is a **constant**.

$$\gamma(h) = \begin{cases} \sum_{i=1}^{p} \theta_j \gamma(h-j) + \sigma_w^2 \sum_{j=h}^{q} \theta_j \psi_{j-h}, & 0 \leq h \leq max(p, q+1) \\ \sum_{i=1}^{p} \phi_j \gamma(h-j), & h \geq max(p, q+1) \end{cases} \tag{21}$$