An Introduction to model based time series analysis

S.N. Lahiri

Department of Statistics

snlahiri@ncsu.edu

Aug 24, 2016
Characteristics of Time Series

- A time series is a collection of observations made at different times on a given system.
- For example:
  - Global temperature anomalies from 1856 to 1997 (annual);
  - Investment returns on the New York Stock Exchange (daily).
Global Temperature

![Graph showing global temperature changes from 1860 to 2000. The y-axis represents global temperature (globtemp) with values ranging from -0.4 to 0.4, and the x-axis represents years from 1860 to 2000. The graph depicts a steady increase in global temperature over the years.]
New York Stock Exchange Data
The primary objective of time series analysis is to develop mathematical models that provide plausible descriptions of sample data.

We model a time series as (realizations of) a collection of random variables \( \{X_1, X_2, \ldots \} \), or more generally, as \( \{X_t : t \in T\} \) where \( T \subset (-\infty, \infty) \).

Often the phenomenon being observed evolves in continuous time, but in this talk, we shall consider only the case of discrete, equally spaced time indices. Thus, \( T \subset \mathbb{Z} \equiv \{0, \pm 1, \pm 2, \ldots \} \).

Analyses of irregularly spaced time series data or time series with missing values are more complicated. [Second lecture.]
Characteristics of Time Series

- Time series data are almost always correlated with each other—**autocorrelated**.
- Presence of correlation:
  - facilitates the task of prediction or forecasting (e.g., leads to smaller MSPE)
  - makes it harder to estimate population characteristics (i.e., the effective sample size may be MUCH smaller than the number of data values at hand, particularly when we have strong dependence).
- Thus, we may want to exploit that correlation, or merely to cope with it, depending on the objective.
Covariances

- The **autocovariance function** of a time series \( \{X_t : t \in T\} \) is defined as

\[
\gamma_X(s, t) = E[(X_s - \mu_{X,s})(X_t - \mu_{X,t})], \quad s, t \in T.
\]

- **Symmetry:**

\[
\gamma_X(s, t) = \gamma_X(t, s) \quad \text{for all} \quad s, t \in T.
\]

- **Smoothness:**
  - if a series is smooth, nearby values will be very similar, hence the autocovariance will be *large*;
  - conversely, for a *choppy* series, even nearby values may be nearly uncorrelated.
Example: White Noise

Let $W_t \sim WN(0, \sigma^2_W)$. Then,

$$
\gamma_W(s, t) = \begin{cases} 
\sigma^2_W & \text{if } s = t \\
0 & \text{if } s \neq t
\end{cases}
$$
A realization of an IID WN process
Example: Moving Average

- Consider the moving average process \( \{V_t : t \in \mathbb{Z}\} \):

\[
V_t = \frac{W_t + W_{t-1} + W_{t-2}}{3}
\]

- Then \( EV_t = 0 \) for all \( t \) and

\[
\gamma_V(s, t) = EV_s V_t
= E \left[ \left( W_s + W_{s-1} + W_{s-2} \right) \left( W_s + W_{s-1} + W_{s-2} \right) \right] / 9
= \begin{cases} 
3\sigma_W^2 / 9 & \text{if } s = t \\
2\sigma_W^2 / 9 & \text{if } s = t \pm 1 \\
\sigma_W^2 / 9 & \text{if } s = t \pm 2 \\
0 & \text{otherwise}.
\end{cases}
\]
Moving Average

- Note that from the definition above, the auto-covariance is a function of $s - t$ only; It does not depend on $s$ and $t$ separately.
- A realization of the $V_t$-process:
Stationary Time Series

- Basic idea: the statistical properties of the observations tend to remain the same over time.
- There are two forms: **strong (or strict) stationarity** and **weak stationarity**.
- A time series $X_t$ is called **strongly stationary** if the joint distribution of every finite collection of variables remain the same under time shifts, i.e., the joint distribution of $\{X_{t_1}, \ldots, X_{t_k}\}$ is the same as that of $\{X_{t_1+h}, \ldots, X_{t_k+h}\}$ for all $t_1, \ldots, t_k \in T$, for all $h$ (positive or negative) and for all $k \geq 1$.
- Strong stationarity is hard to verify.
Weak stationarity

- A time series \( \{X_t\} \) is **weakly stationary** if:
  - the mean function \( \mu_X(t) \) is constant; that is, every \( X_t \) has the same mean;
  - the autocovariance function \( \gamma_X(s; t) \) depends on \( s \) and \( t \) only through their difference \( |s - t| \).

- Weak stationarity depends only on the first and second moment functions, so it is also called **second-order stationarity**.

- Strongly stationary (plus finite variance) \( \rightarrow \) weakly stationary.

- In general, weak stationarity does not imply strong stationarity.
A class of stationary time series models, popularized by Box & Jenkins are the *Autoregressive Moving Average* (ARMA) models!!

**Definition:** A process \( \{X_t : t \in \mathbb{Z}\} \) is called an ARMA(\(p, q\))-process if it is stationary and if for every \(t\),

\[
X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = W_t + \theta_1 W_{t-1} + \ldots + W_{t-q}
\]

where \(W_t \sim WN(0, \sigma^2)\). We shall use the notation \(X_t \sim \text{ARMA}(p, q)\).

- Note that by stationarity, \(EX_t = 0\) for all \(t\).
- Here \(p\) is called the Autoregressive order and \(q\) the Moving Average order!
Moving Average

- A realization of an ARMA(2,0)-process: $\phi_1 = .2$, $\phi_2 = .4$ and $\theta_1 = .5$. 
ARMA-processes

- When $\theta_1 = \ldots = \theta_q = 0$, we get an *Autoregressive process of order* $p$ (or **AR($p$)**-process, in short):

\[ X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + W_t \]

- Thus, $X_t$ is given by a linear regression on the past $p$ values.

- Similarly, if $\phi_1 = \ldots = \phi_p = 0$, we get an **MA($q$)** process:

\[ X_t = W_t + \theta_1 W_{t-1} + \ldots + \theta_q W_{t-q} \]

- **Result:** (Weak) Stationarity of an ARMA ($p,q$) process is guaranteed when the roots of the AR-polynomial

\[ \phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p, z \in \mathbb{C} \]

lie outside the unit circle in the complex plane.
Simplifications

- If \( X_t \) is weakly stationary, then the covariance between \( X_{t+h} \) and \( X_t \) depends on \( h \) but not on \( t \), so that we may write the auto-covariances as

\[
\gamma_X(h) = \text{Cov}(X_{t+h}, X_t) \quad \text{for all} \quad t, h
\]

- Similarly, the correlation between \( X_{t+h} \) and \( X_t \) can be written as

\[
\rho_X(h) = \frac{\gamma_X(t+h, t)}{\sqrt{\gamma_X(t+h, t+h)\gamma_X(t, t)}} = \frac{\gamma_X(h)}{\gamma_X(0)}
\]
Estimation of the Covariance Structure

• If $X_t$ is SOS, we can estimate the autocovariance function $\gamma(h)$ by

$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-h} \left( X_{t+h} - \bar{X}_n \right) \left( X_t - \bar{X}_n \right).$$

• Similarly, we can estimate the autocorrelation function (ACF) $\rho(h)$ by

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

• Caution: These estimators are reasonable only when $h$ is not too large compared to $n$. A rule of thumb is that $|h| < n/2$. 
• The covariance matrix of \((X_1, \ldots, X_k)'\) is

\[
\Gamma_k \equiv \begin{bmatrix}
\gamma(0) & \gamma(1) & \ldots & \gamma(k-1) \\
\gamma(1) & \gamma(0) & \ldots & \gamma(k-2) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma(k-1) & \gamma(k-2) & \ldots & \gamma(0)
\end{bmatrix}.
\]

• It is non-negative definite as

\[
a'\Gamma_k a = \text{Var}\left(a_1 X_1 + \ldots + a_k X_k\right) \geq 0 \text{ for all } a = (a_1, \ldots, a_k)' \in \mathbb{R}^k.
\]

• Note: With the above definition of \(\hat{\gamma}(h), \hat{\Gamma}_k\) is also non-negative definite.
An important Sampling Property for $\hat{\rho}(h)$

- If $X_t$ is white noise and $n$ is large and some mild conditions hold, then:

  for each fixed $h$, $\hat{\rho}(h)$ is approximately normal with zero mean and standard deviation

  $$\sigma_{\hat{\rho}(h)} = \frac{1}{\sqrt{n}}.$$ 

- So we can look for autocorrelations outside $\pm 2/\sqrt{n}$ as evidence of non-zero autocorrelation.
Example 1: ACF

- ACF of a WN(0,1)-series:
Example 2: ACF

- ACF of a MA(3)-series: $\theta_1 = .3 = \theta_2 = \theta_3$. 
Example 3: ACF

- **ACF of a AR(2)-series:** $\phi_1 = .4, \phi_2 = .3$
Example 3: PACF

- Like the ACF of an MA-models, a different function, called the **Partial Autocorrelation Function or PACF** of an AR(p) process cuts off after $p$-lags.
- **PACF of a AR(2)-series:** $\phi_1 = .1, \phi_2 = .6$

![ACF of AR(2)](chart1)

![PACF of AR(2)](chart2)
ACF of ARMA processes

• For an ARMA \((p, q)\) process, the general form of its ACF is

\[
\gamma(h) = \sum_{j=1}^{p} c_j z_j^{-h}
\]

where \(c_1 \ldots, c_p\) are constants and where \(z_1, \ldots, z_p\) are the roots of the AR-polynomial: \(\phi(z) = 0\).

• The correlations decay exponentially fast in \(h\) but never vanishes!

• For \(p \geq 1, q \geq 1\),
  - the estimated ACF does NOT cut off after a finite lag
  - the estimated PACF also does NOT cut off after a finite lag!

• The orders \(p\) and \(q\) are determined using information criteria, such as AIC, BIC, etc.
The (weak) stationarity assumption can fail due to:
- a non-constant mean structure, and/or
- a auto-covariance function that is NOT a function of the lag differences!

First, we will consider an example of the first type, involving the Southern Oscillation Index (SOI) series!

The SOI measures changes in the air pressure related to sea surface temperature in the central Pacific Ocean.
Here is a plot of the SOI data.
Does the plot show any patterns, e.g., local dependence, seasonality, ...?
• Consider the ACF of the SOI-data. The ACF suggests that $X_t$ has a correlation of around 0.4 with $X_{t+12}$, $X_{t+24}$, and so on. Further, the magnitude does not seem to die out!
• This “correlation” is caused by the fact that these values all fall in the **same month** of the year, and different months have different means.

• **That is, this series has a non-constant mean function** $\mu_t$.

• Since it is nonstationary, it has no ACF, and the graph cannot be interpreted as an ACF.

• We can estimate $\mu_t$ (for each of the 12 months) and subtract these to give a series with zero mean, which hopefully has a stationary structure!
• The ACF graph now shows correlation dropping progressively from around 0.5 at a one month lag to zero at one year.
• Consider the following (simulated) data set:

Q: Is it (weakly) stationary?
ACF of the series

- The ACF graph shows very slow decay of correlations:
ACF of the series

- This shows that the ACF decays very very slowly. Significant correlation persists even when \( h \approx 30 \).
- A useful device to deal with such situations is to use differencing.
- Define \( \nabla X_t = X_t - X_{t-1} \). Here is a plot of \( \nabla X_t \):
• The ACF graph of the $Z_t = \nabla X_t$ show quick decay of correlations:

• Thus, $Z_t$ can be modeled using an MA(2) model!!
ARIMA models

- **Def:** \( \{X_t\} \) is called an **ARIMA** \((p, d, q)\) **process** (with \(d, p, q \geq 0\) integers) if

\[
\nabla^d X_t \sim \text{ARMA}(p, q)
\]

, where, for \(d \geq 2\), \(\nabla^d X_t = \nabla(\nabla^{d-1} X_t)\).

- Thus, in the above example, \(X_t \sim \text{ARIMA}(0, 1, 2)\).
- Similarly, \(X_t \sim \text{ARIMA}(0, 1, 0)\) gives the **random walk**.
- The random walk \(X_t = X_{t-1} + W_t, \ t \geq 1\) (with \(X_0 = 0\)) is nonstationary, with **ACF**

\[
\gamma(s, t) = \sigma_W^2 \left[ \min\{s, t\} - st \right], \ s, t \geq 1
\]

- Not a function of \((s - t)\).
ARIMA models

- The Box-Jenkins approach to time series analysis is based on successive differencing of a nonstationary time series to reduce it to approximate stationarity, and then apply a suitable low order ARMA \((p, q)\) model.
- Quite successful in many applications.
- However, it is less than satisfactory in some situations...
  - where differencing of any number of orders does not reduce it to the ARMA \((p, q)\)-regime!
The Box-Jenkins approach to time series analysis is based on successive differencing of a nonstationary time series to reduce it to approximate stationarity, and then apply a suitable low order ARMA \((p, q)\) model!

Quite successful in many applications.

However, it is less than satisfactory in some situations...

where differencing of any number of orders does not reduce it to the ARMA \((p, q)\)-regime!
Consider the following series:

Looks stationary?
Fractional Differencing; ARFIMA models

- ACF of the series:

- Decays more slowly than the ACF of an ARMA model, but faster than that of an ARIMA model!
Fractional Differencing; ARFIMA models

- \( \{X_t\} \) is called an ARFIMA(p,d,q) process, with \( p \geq 0, q \geq 0 \) and \( |d| < 0.5 \), if

\[
(1 - B)^d X_t \sim \text{ARMA}(p,q)
\]

where \( B \) is the back-shift operator!

- Equivalently, \( \{X_t\} \sim \text{ARFIMA}(p,d,q) \) if

\[
X_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j}
\]

for some \( \pi_j = \pi_j(d) \), where \( Y_t \sim \text{ARMA}(p,q) \).

- The coefficients \( \pi_j \) are determined by the Binomial series expansion with a fractional/negative exponent!
For $0 < d < 1/2$,

- $\pi_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}$, $j \geq 0$,
- $\gamma_Y(h) \sim h^{2d-1}$
- $\sum_h |\gamma_Y(h)| = \infty$.

The last property says that the time series has **Long memory**!!

- $X_t$ is a $1/f^d$ noise??
An ARFIMA example (Eric Fiegeelson)

KID 010122419
Cleaned and stitched lightcurve
IQR = 51.6

KID 010122419
Differenced lightcurve
IQR = 18

KID 010122419
ARFIMA(1,1,1) model

KID 010122419
ARFIMA residuals
IQR = 11.3
An ARFIMA example (Eric Fiegelson)

Thank you!!