Lecture 16
Review and Further Topics

MATH 8090 Time Series Analysis
November 30 & December 2, 2021

Whitney Huang
Clemson University

Agenda

1. **Review**

2. **Further Topics**

Time Domain Analysis vs Frequency Domain Analysis

- Time Domain:
  - Stationarity, ACVF, and ACF
  - Linear processes, causality, invertibility
  - ARMA models, estimation, forecasting
  - ARIMA, seasonal ARIMA models
- Frequency Domain:
  - Spectral density, Periodogram
  - Nonparametric spectral density estimation
  - Parametric spectral density estimation
  - Lagged regression models

Objectives of Time Series Analysis

- Compact description of data
- Forecasting
- Control
- Hypothesis testing
- Simulation

Time Series Modeling

- First step: plot the time series

  Look for trends, seasonal components, step changes, outliers

- Transform data so that residuals are (approximately) stationary
  - Estimate and remove μ_t and s_t
  - Differencing
  - Nonlinear transformations (e.g., $\log$, $\sqrt{\cdot}$)
- Fit a model to residuals

Stationarity

$\{Y_t\}$ is strictly stationary if, for all k, $t_1, \cdots, t_k$, $y_1, \cdots, y_k$ and h,

$$\mathbb{P}(Y_{t_1} \le y_1, \cdots, Y_{t_k} \le y_k) = \mathbb{P}(Y_{t_1+h} \le y_1, \cdots, Y_{t_k+h} \le y_k).$$

i.e., shifting the time axis does nor affect the joint distribution

We consider second-order properties only:

$\{Y_t\}$ is stationary if its mean function and autocovariance function satisfy

$$\mu_t = \mathbb{E}[Y_t] = \mu,$$
$$\gamma(s,t) = \mathbb{C}\mathrm{ov}(Y_s, Y_t) = \gamma(s-t).$$

Note: it implies constant variance as $\gamma(t,t) = \mathbb{V}\mathrm{ar}(Y_t) = \gamma(0)$

ACF and Sample ACF

The autocorrelation function (ACF) is

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \mathbb{Cor}(Y_{t+h}, Y_t)$$

For observations $y_1, \cdots, y_n$ of a time series, the sample mean is

$$\bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t.$$

The sample autocovariance function (ACVF) is

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}\left(y_{t+|h|} - \bar{y}\right)\left(y_t - \bar{y}\right), \quad \text{for } -n < h < n.$$

The sample autocorrelation function is

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

Linear Processes

Linear process is an important class of stationary time series:

$$Y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j},$$

where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$, and $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

Example: ARMA(p, q)

Causality and Inveribility

A linear porcess $\{Y_t\}$ is causal if there is a

$$\psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \cdots$$

with

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

and

$$Y_t = \psi(B) Z_t.$$

A linear process $\{Y_t\}$ is invertible if there is a

$$\pi(B) = \pi_0 + \pi_1 B + \pi_2 B^2 + \cdots$$

with

$$\sum_{j=0}^{\infty} |\pi_j| < \infty$$

and

$$Z_t = \pi(B) Y_t.$$

Autoregressive Moving Average Models

An ARMA(p,q) process $\{Y_t\}$ is a stationary process that satisfies

$$Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},$$

where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$. Also, $\phi_p, \theta_q \neq 0$ and $\phi(z)$ and $\theta(z)$ have no common factors

Properties:

- A unique stationary solution exists if and only if

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = 0 \Rightarrow |z| \neq 1.$$

- This ARMA(p,q) process is causal if and only if

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = 0 \Rightarrow |z| > 1.$$

- It is invertible if and only if

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q = 0 \Rightarrow |z| > 1.$$

Linear Prediction

Given $Y_1, Y_2, \cdots, Y_n$, the best linear predictor $Y_{n+h}^n = \alpha_0 + \sum_{i=1}^n \alpha_i Y_i$ of Y_{n+h} satisfies the prediction equations:

$$\mathbb{E}[Y_{n+h} - Y_{n+h}^n] = 0$$
$$\mathbb{E}\left[(Y_{n+h} - Y_{n+h}^n)Y_i\right] = 0 \quad \text{for } i = 1, \cdots, n.$$

One-step-ahead linear prediction

$$Y_{n+1}^n = \phi_{n1}Y_n + \phi_{n2}Y_{n-1} + \cdots + \phi_{nn}Y_1$$
$$\Gamma_n \phi_n = \gamma_n, \quad P_{n+1}^n = \mathbb{E}(Y_{n+1} - Y_{n+1}^n)^2 = \gamma(0) - \gamma_n^T \Gamma_n^{-1} \gamma_n,$$

with

$$\Gamma_n = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \cdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{bmatrix},$$

where

$$\phi_n = (\phi_{n1}, \phi_{n2}, \cdots, \phi_{nn})^T,$$

and

$$\gamma_n = (\gamma(1), \gamma(n), \cdots, \gamma(n))^T.$$

Yule-Walker Estimation and Maximum Likelihood Estimation

Method of moments: choose parameters for which the moments are equal to the empirical moments. One choose ϕ such that $\gamma = \hat{\gamma}$.

Yule-Walker equations for $\hat{\phi}$: $\begin{cases} \hat{\Gamma}_p \hat{\phi} = \hat{\gamma}_p, \\ \hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\phi}^T \hat{\gamma}_p. \end{cases}$

Maximum Likelihood Estimation: Suppose that $Y_1, \cdots, Y_n$ is drawn from a zero mean Gaussian ARMA(p, q) process. The likelihood of parameters $\phi \in \mathbb{R}^p$ and $\theta \in \mathbb{R}^q$, $\sigma^2 \in \mathbb{R}_+$ is defined as the joint density of $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_n)$:

$$L(\phi, \theta, \sigma^2) = \frac{1}{(2\pi)^{n/2} |\Gamma_n|^{1/2}} \exp\left(-\frac{1}{2} \boldsymbol{Y}^T \Gamma_n^{-1} \boldsymbol{Y} \right).$$

The maximum likelihood estimator (MLE) of ϕ, θ, σ^2 maximizes this quantity.

ARIMA(p, d, q) and Seasonal ARIMA Models

For $p, d, q \geq 0$, we say that a time series Y_t is an ARIMA(p, d, q) process if

$$X_t = \nabla^d Y_t = (1 - B)^d Y_t$$

is ARMA(p, q). We can write

$$\phi(B)(1 - B)^d Y_t = \theta(B) Z_t.$$

For $p, q, P, Q \geq 0$, $s, d, D > 0$, we say a time series $\{Y_t\}$ is a (multiplicative) seasonal ARIMA model (ARIMA(p, d, q) $\times$ $(P, D, Q)_s$) if

$$\Phi(B^s)\phi(B) \nabla_s^D \nabla^d Y_t = \Theta(B^s)\theta(B) Z_t,$$

where the seasonal difference operator of order D is defined by

$$\nabla_s^D Y_t = (1 - B^s)^D Y_t.$$

Spectral Density and Spectral Distribution Function

If $\{Y_t\}$ has $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, then we define its spectral density as

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \omega h}$$

for $-\infty < \omega < \infty$. We have

$$\gamma(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i \omega h} f(\omega)\, d\omega = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i \omega h}\, dF(\omega),$$

where $dF(\omega) = f(\omega)\, d\omega$.

Periodogram and Its Asymptotic Properties

The periodogram is defined as

$$
\begin{aligned}
I(\omega_j) &= |d(\omega_j)|^2 \\
&= \frac{1}{n}|\sum_{t=1}^{n} e^{-2\pi i\omega_j t} y_t|^2 \\
&= \frac{1}{n}\left[(\sum_{t=1}^{n}\cos(2\pi t\omega_j)y_t)^2 + (\sum_{t=1}^{n}\sin(2\pi t\omega_j)y_t)^2\right]
\end{aligned}
$$

Under general conditions, we have

$$\frac{2I(\omega_j)}{f(\omega_j)} \xrightarrow{d} \chi_2^2.$$

Thus,

$$
\begin{aligned}
\mathbb{E}(I(\omega_j)) &\to f(\omega) \\
\mathbb{Var}(I(\omega_j)) &\to f(\omega)^2
\end{aligned}
$$

Smoothed Periodogram

If $f(\omega)$ is approximately constant in the band $[\omega_{j-m}, \omega_{j+m}]$, then the average of the periodogram over the band

$$\bar{f}(\omega_j) = \frac{1}{2m+1} \sum_{k=-m}^{m} I(\omega_{j+k})$$

will be unbiased. This is the averaged periodogram

Smoothed periodogram:

$$\hat{f}(\omega_j) = \sum_{k=-m}^{m} W_m(k) I(\omega_{j+k}).$$

$W_m(k)$ is the spectral window function satisfies $W_m(k) \geq 0, W_m(k) = W_m(-k)$ and $\sum_{k=-m}^{m} W_m(k) = 1$. The averaged periodogram is a special case of smoothed periodogram with

$$W_m(k) = \frac{1}{2m+1} \text{ if } -m \leq k \leq m.$$

Parametric Spectral Density Estimation

Given data $y_1, y_2, \cdots, y_n$,

- Estimate the AR parameters $\phi_1, \cdots, \phi_p, \sigma^2$
- Use the estimates $\hat{\phi}_1, \cdots, \hat{\phi}_p, \hat{\sigma}_2$ to compute the estimated spectral density:

$$\hat{f}(\omega) = \frac{\hat{\sigma}^2}{\left|\hat{\phi}(e^{-2\pi i\omega})\right|^2}$$

- For large n,

$$\mathbb{Var}(\hat{f}(\omega)) \approx \frac{2p}{n} f^2(\omega)$$

Coherence

Cross-spectrum:

$$f_{XY}(\omega) = \sum_{h=-\infty}^{\infty} \gamma_{XY}(h) e^{-2\pi i \omega h}.$$

Cross-covariance:

$$\gamma_{XY} = \int_{-\frac{1}{2}}^{\frac{1}{2}} f_{XY}(\omega) e^{2\pi i \omega h} \, d\omega.$$

Squared coherence:

$$\rho^2_{Y,X}(\omega) = \frac{|f_{YX}(\omega)|^2}{f_X(\omega) f_Y(\omega)}.$$

Lagged Regression Models in the Frequency Domain

Lagged Regression Model:

$$Y_t = \sum_{j=-\infty}^{\infty} \beta_j X_{t-j} + V_t.$$

One can compute the Fourier transform of the series $\{\beta_j\}$ in terms of the cross-spectral density and the spectral density:

$$B(\omega) f_X(\omega) = f_{YX}(\omega).$$

The resulting mean squared error:

$$\mathrm{MSE} = \int_{-\frac{1}{2}}^{\frac{1}{2}} f_Y(\omega) \left(1 - \rho_{Y,X}^2(\omega)\right) d\omega.$$

Thus, $\rho_{Y,x}^2(\omega)$ indicates how the component of the variance of $\{Y_t\}$ at a frequency ω is accounted for by $\{X_t\}$

Multivariate Time Series Analysis

- All the methods presented for univariate time series also apply to multivariate processes

$$\{\boldsymbol{Y}_t \in \mathbb{R}^p\}$$

- The theory is a little more involved as we generalize to the cross-covariance:

$$\mathbb{Cov}(\boldsymbol{Y}_s, \boldsymbol{Y}_t) = \boldsymbol{C}(s,t),$$

where $\boldsymbol{C}(\cdot,\cdot)$ is the $p \times p$ matrix-valued cross-covariance function (CCVF)

Vector Autoregressive (VAR) Models

VAR(p) model:

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + A_1\boldsymbol{Y}_{t-1} + \cdots + A_p\boldsymbol{Y}_{t-p} + \boldsymbol{W}_t, \quad t = 0, \pm 1, \pm 2, \cdots,$$

where

- $\boldsymbol{Y}_t = (Y_{1t}, \cdots, Y_{pt})^T$ is a $(p \times 1)$ random vector
- A_i are $(p \times p)$ coefficient matrices
- $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_p)^T$ is the intercept vector
- $\boldsymbol{W}_t = (W_{1t}, \cdots, W_{pt})^T$ is a p-dimensional white noise, i.e., $\mathbb{E}[\boldsymbol{W}_t] = \mathbf{0}$, $\mathbb{E}[\boldsymbol{W}_t\boldsymbol{W}_t^T] = \Sigma_{\boldsymbol{W}}$ and $\mathbb{E}[\boldsymbol{W}_s\boldsymbol{W}_t^T] = \mathbf{0}$ for $s \neq t$.

An Example of Identifiability Issue of VARMA

The following bivariate AR(1) and MA(1) models are identical:

VAR(1):

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} + \begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix}$$

VMA(1):

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ -2 & 0 \end{bmatrix} \begin{bmatrix} W_{1,t-1} \\ W_{2,t-1} \end{bmatrix}$$

$$\Rightarrow Y_{1t} = W_{1t},\ Y_{2t} = 2Y_{1,t-1} + W_{2,t} = 2W_{1,t-1} + W_{2t}$$

Such a exchangeable forms between AR and MA models cannot occur in the univariate case [Tsay, 2000]

Long-Memory Processes

A stationary process $\{Y_t\}$ is called long-memory with parameter $d \in (0, 0.5)$, if

$$C(h) = \mathbb{C}\mathrm{ov}(Y_t, Y_{t+h}) \sim ch^{2d-1} \quad (h \to \infty)$$

- Long-memory processes are time series models in which ACF decay slowly with increasing lags
- Visual features of the data:
  - Relatively long periods of large and small values
  - Looking at short periods of time, there is evidence of trends and seasonality. These disappear as the period length increases

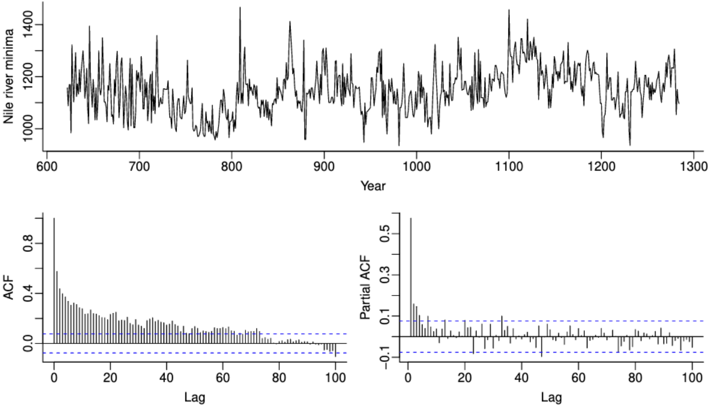
Autoregressive Fractionally Integrated Moving Average (ARFIMA) Model

When we extend d in ARIMA to be real-valued we obtain the autoregressive fractionally integrated moving average (ARFIMA) model:

- This is an example of a long-memory process
- The parameter d is called the long-memory parameter
- The process $\{Y_t\}$ is non-stationary when $d \geq 1/2$

Example: Nile River Ninima

Nile river annual minimal water levels for the years 622 to 1281, measured at the Roda gauge near Cairo [Tousson, 1925, p.366-385]

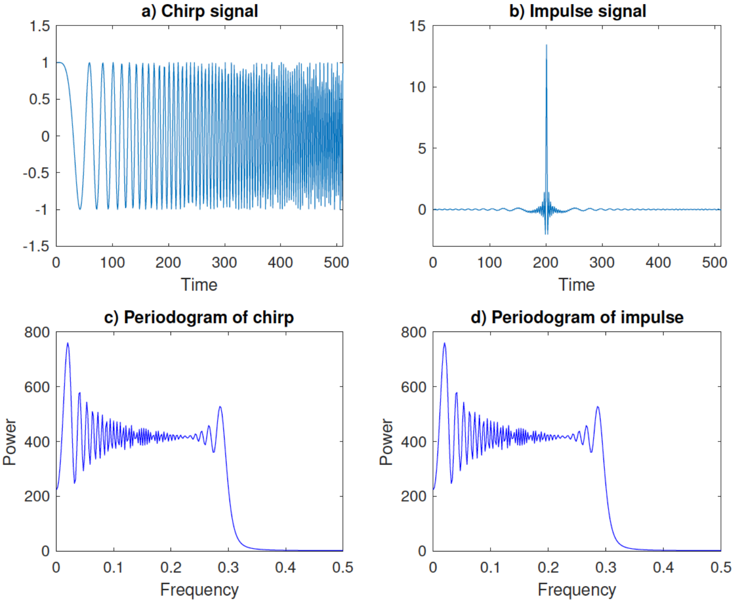


Source: Craigmile's short course in Spatio-temporal methods, Extreme value modeling and water resources summer school, Universite Lyon 1, France, Jun 2016

Bootstraps for Time Series

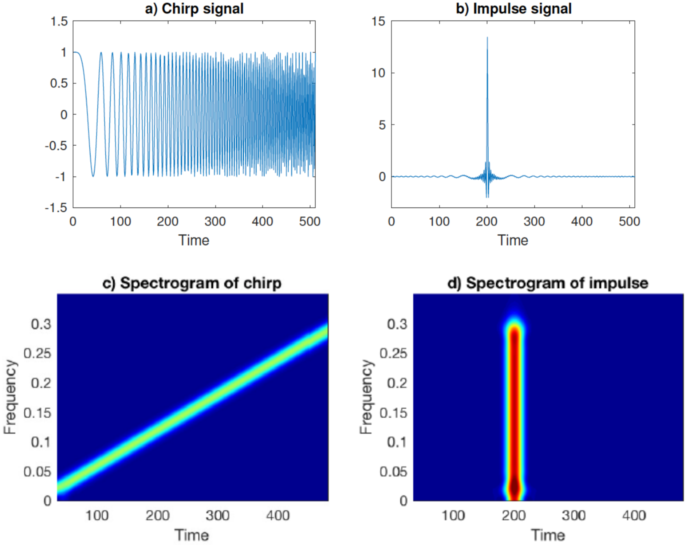
- Bootstrap [Efron, 1979]: simulation-based methods for frequentist inference.
- Moving block bootstrap [Künsch, 1989]: data $\{y_1, y_2, \cdots, y_n\}$ is split into $n - b + 1$ overlapping blocks of length b. Then from these $n - b + 1$ blocks, n/b blocks will be drawn at random with replacement to form the bootstrap observations
- Not stationary by construction. Varying randomly b can avoid this problem and it is known as the stationary bootstrap [Politis and Romano, 1994]

Time-Frequency Analysis: A Motivation Example



Time-Frequency Analysis: Spectrogram

A spectrogram is a visual representation of the spectrum of frequencies of a signal as it varies with time



Non-Gaussian Time Series Methods

Some selected references:

- Regression models for time series analysis, Kedem and Fokianos, 2002
- Handbook of discrete-valued time series, edited by Davis, Holan, Lund, Ravishanker, 2016
- Bayesian Dynamic Generalized Linear Models, Gamerman *et. al*, 2016
- Count Time Series: A Methodological Review, Davis *et. al.*, 2021