Lecture 7
Stationary Processes

Reading: Bowerman, O'Connell, and Koehler (2005): Chapter; Cryer and Chen (2008): Chapter 1.2; Chapter 4.2 and 4.3

MATH 4070: Regression and Time-Series Analysis

Whitney Huang
Clemson University

Notes

Agenda

1. Review
2. Some Examples of Stationary Processes
3. Estimation and Inference for Mean Functions
4. Differencing

Stationary Processes
Review
Some Examples of Stationary Processes
Estimation and Inference for Mean Functions
Differencing

Notes

Time Series Modeling Strategy

Additive Decompstion:

$$Y_t = \mu_t + s_t + \eta_t, \quad t = 1, 2, \cdots, T$$

1. Plot the data y_t to explore the form of μ_t and s_t, and check for non-constant variation in η_t
2. Transform (if necessary) to stabilize variance of η_t
3. Estimate μ_t and s_t to obtain residuals $\hat{\eta}_t$
4. Use residuals to select a time series model for η_t
5. Estimate parameters in μ_t, s_t, and η_t (ideally simultaneously in one step)
6. Check for fit of model (poor fit $\Rightarrow$ return to step 1)
7. Use model for inference: predicting future y_t's, describing changes in y_t over time, hypothesis testing, etc

Stationary Processes
Review
Some Examples of Stationary Processes
Estimation and Inference for Mean Functions
Differencing

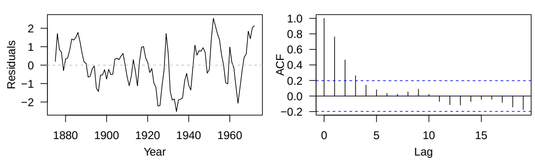
Notes

Recap of the Past Few Lectures

- We discussed the use of regression techniques to model the (deterministic) μ_t and s_t
- Residuals typically suggest temporal dependence in $\{\eta_t\}$
- Time series models concern the modeling of temporal dependence in $\{\eta_t\}$
- Stationarity assumption typically employed to overcome the issue of "one sample"
- Weakly stationary: constant mean and variance over time, with covariance depending only on time lags

Notes

The Implications of Temporal Dependence



- There is a consistent relationship between conservative residuals
- The usual regression assumptions are violated, and t- and F-tests are not valid ☹
- We can get better predictions of future values by modeling autocorrelation ☺

Notes

The White Noise Process

Let's assume $\mathbb{E}(\eta_t) = \mu$ and $\mathrm{Var}(\eta_t) = \sigma^2 < \infty$. $\{\eta_t\}$ is a white noise or $\mathrm{WN}(\mu, \sigma^2)$ process if

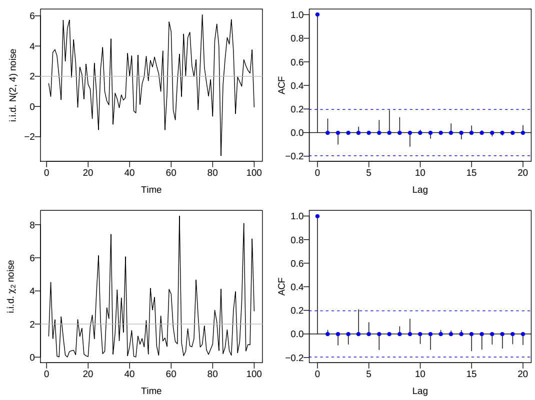
$$\gamma(h) = 0,$$

for $h \neq 0$

- $\{\eta_t\}$ is stationary
- However, distributions of η_t and η_{t+1} can be different!
- All i.i.d. noise with finite variance ($\sigma^2 < 0$) is white noise but the converse need not be true

Notes

Examples Realizations of White Noise Processes



Notes

The Moving Average Process of First Order: MA(1)

Let $\{Z_t\}$ be a $\mathrm{WN}(0, \sigma^2)$ process and θ be some constant $\in \mathbb{R}$. For each integer t, let

$$\eta_t = Z_t + \theta Z_{t-1}.$$

- The sequences of RVs $\{\eta_t\}$ is called the moving average process of order 1 or MA(1) process
- One can show that the MA(1) process $\{\eta_t\}$ is stationary

Notes

MA(1): Mean Function

Need to show the mean function is NOT a function of time t

$$\begin{aligned}\mathbb{E}[\eta_t] &= \mathbb{E}[Z_t + \theta Z_{t-1}] \\ &= \mathbb{E}[Z_t] + \theta \mathbb{E}[Z_{t-1}] \\ &= 0 + \theta \times 0 \\ &= 0, \quad \forall t\end{aligned}$$

☺

Notes

MA(1): Covariance Function

Need to show the autovariance function $\gamma(\cdot,\cdot)$ is a function of time lag only

$$\begin{aligned}\gamma(t,t+h) &= \mathrm{Cov}(\eta_t, \eta_{t+h}) \\ &= \mathrm{Cov}(Z_t + \theta Z_{t-1}, Z_{t+h} + \theta Z_{t+h-1}) \\ &= \mathrm{Cov}(Z_t, Z_{t+h}) + \mathrm{Cov}(Z_t, \theta Z_{t+h-1}) \\ &+ \mathrm{Cov}(\theta Z_{t-1}, Z_{t+h}) + \mathrm{Cov}(\theta Z_{t-1}, \theta Z_{t+h-1})\end{aligned}$$

if $h = 0$, we have $\gamma(t,t+h) = \sigma^2 + \theta^2\sigma^2 = \sigma^2(1+\theta^2)$
if $h = \pm 1$, we have $\gamma(t,t+h) = \theta\sigma^2$
if $|h| \geq 2$, we have $\gamma(t,t+h) = 0$

$\Rightarrow \gamma(t,t+h)$ only depends on h but not on t ☺

Notes

MA(1): ACVF & ACF

ACVF:

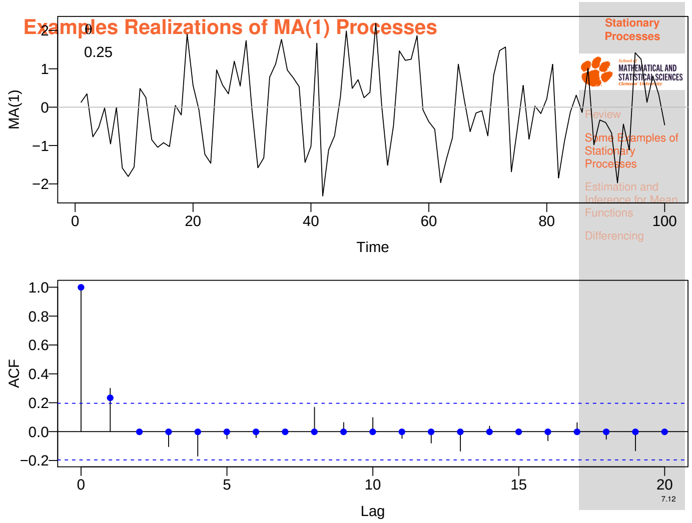
$$\gamma(h) = \begin{cases} \sigma^2(1+\theta^2) & h = 0; \\ \theta\sigma^2 & |h| = 1; \\ 0 & |h| \geq 2 \end{cases}$$

We can get **ACF** by dividing everything by $\gamma(0) = \sigma^2(1+\theta^2)$

$$\rho(h) = \begin{cases} 1 & h = 0; \\ \frac{\theta}{1+\theta^2} & |h| = 1; \\ 0 & |h| \geq 2. \end{cases}$$

Notes

Examples Realizations of MA(1) Processes



Notes

First-order autoregressive process: AR(1)

Let $\{Z_t\}$ be a $\mathrm{WN}(0, \sigma^2)$ process, and $-1 < \phi < 1$ be a constant. Let's assume $\{\eta_t\}$ is a stationary process with

$$\eta_t = \phi\eta_{t-1} + Z_t,$$

for each integer t, where η_s and Z_t are uncorrelated for each $s < t \Rightarrow$ future noise is uncorrelated with the current time point

We will see later there is only one unique solution to this equation. Such a sequence $\{\eta_t\}$ of RVs is called an AR(1) process

Notes

Properties of the AR(1) process

Want to find the mean value μ under the weakly stationarity assumption

$$\begin{aligned} \mathbb{E}[\eta_t] &= \mathbb{E}[\phi\eta_{t-1} + Z_t] \\ \mu &= \phi\mathbb{E}[\eta_{t-1}] + \mathbb{E}[Z_t] \\ \mu &= \phi\mu + 0 \\ \Rightarrow \mu &= 0, \quad \forall t \end{aligned}$$

☺

Want to find $\gamma(h)$ under the weakly stationarity assumption

$$\begin{aligned} \mathrm{Cov}(\eta_t, \eta_{t-h}) &= \mathrm{Cov}(\phi\eta_{t-1} + Z_t, \eta_{t-h}) \\ \gamma(-h) &= \phi\mathrm{Cov}(\eta_{t-1}, \eta_{t-h}) + \mathrm{Cov}(Z_t, \eta_{t-h}) \\ \gamma(h) &= \phi\gamma(h-1) + 0 \\ \Rightarrow \gamma(h) &= \phi\gamma(h-1) = \cdots = \phi^{|h|}\gamma(0) \end{aligned}$$

Next, need to figure out $\gamma(0)$

Notes

Properties of the AR(1) process Cont'd

$$\begin{aligned} \mathrm{Var}(\eta_t) &= \mathrm{Var}(\phi\eta_{t-1} + Z_t) \\ \gamma(0) &= \phi^2\gamma(0) + \sigma^2 \\ \Rightarrow (1-\phi^2)\gamma(0) &= \sigma^2 \\ \Rightarrow \gamma(0) &= \frac{\sigma^2}{1-\phi^2} \end{aligned}$$

☺

Therefore, we have

$$\gamma(h) = \begin{cases} \frac{\sigma^2}{1-\phi^2} & h = 0; \\ \frac{\phi^{|h|}\sigma^2}{1-\phi^2} & |h| \geq 1, \end{cases}$$

and

$$\rho(h) = \begin{cases} 1 & h = 0; \\ \phi^{|h|} & |h| \geq 1. \end{cases}$$

Notes

Examples Realizations of AR(1) Processes

Notes

The Random Walk Process

Let $\{Z_t\}$ be a $\mathrm{WN}(0, \sigma^2)$ process and for $t \geq 1$ definite

$$\eta_t = Z_1 + Z_2 + \cdots + Z_t = \sum_{s=1}^{t} Z_s.$$

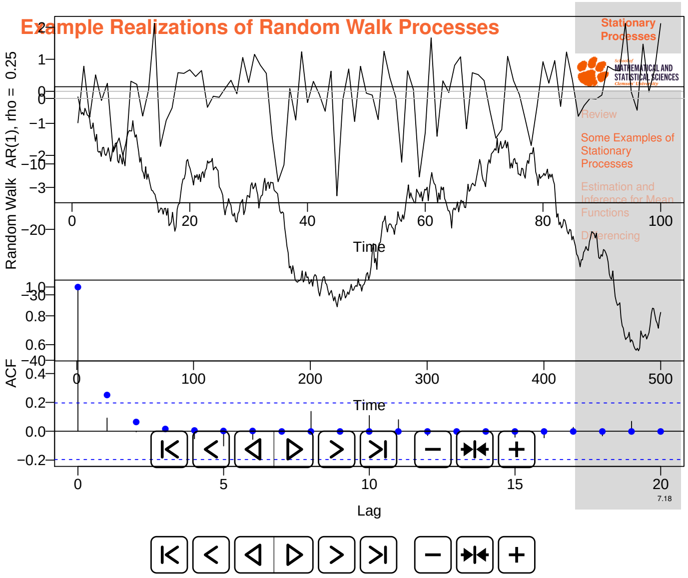
- The sequence of RVs $\{\eta_t\}$ is called a random walk process
- **Special case**: If we have $\{Z_t\}$ such that for each t

$$\mathbb{P}(Z_t = z) = \begin{cases} \frac{1}{2}, & z = 1; \\ \frac{1}{2}, & z = -1, \end{cases}$$

  then $\{\eta_t\}$ is a simple symmetric random walk
- The random walk process is not stationary!

Notes

Example Realizations of Random Walk Processes



Notes

Gaussian Processes

$\{\eta_t\}$ is a Gaussian process (GP) if the joint distribution of any collection of the RVs has a multivariate normal (aka Gaussian) distribution

- The distribution of a GP is fully characterized by $\mu(\cdot)$, the mean function, and $\gamma(\cdot,\cdot)$, the autocovariance function. The joint probability density function of $\boldsymbol{\eta} = (\eta_1, \eta_2, \cdots, \eta_T)^T$ is

$$f(\boldsymbol{\eta}) = \frac{1}{(2\pi)^{\frac{T}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{\eta}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{\eta}-\boldsymbol{\mu})\right),$$

  where $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_T)^T$ and the (i,j) element of the covariance matrix Σ is $\gamma(i,j)$

- If a GP $\{\eta_t\}$ is weakly stationary then the process is also strictly stationary

Notes

Estimating the Mean of Stationary Processes

Let $\{\eta_t\}$ be stationary with mean μ and ACVF $\gamma(s,t) = \gamma(s-t)$

- A natural estimator of μ is the sample mean

$$\bar{\eta} = \frac{1}{T}\sum_{t=1}^{T}\eta_t.$$

  $\bar{\eta}$ is an unbiased estimator of μ, i.e.

- Since $\{\eta_t\}$ is stationary, we have

$$\begin{aligned}\mathrm{Var}(\bar{\eta}) &= \frac{1}{T^2}\mathrm{Var}\left(\sum_{i=1}^{T}\eta_t\right)\\ &= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathrm{Cov}(\eta_s,\eta_t)\\ &= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma(s-t)\end{aligned}$$

- **Exercise**: Show

$$\mathrm{Var}(\bar{\eta}) = \frac{1}{T}\sum_{h=-(T-1)}^{T-1}\left(1-\frac{|h|}{T}\right)\gamma(h)$$

Notes

AR(1) Example

Suppose $\{\eta_1, \eta_2, \eta_3\}$ is an AR(1) process with $|\phi| < 1$ and innovation variance σ^2. Show that the variance of $\bar{\eta}$ is $\frac{\sigma^2}{9(1-\phi^2)}(3+4\phi+2\phi^2)$

Solution:

Notes

The Sampling Distribution of $\bar{\eta}$

Let

$$v_T = \sum_{h=-(T-1)}^{(T-1)} \left(1 - \frac{|h|}{T}\right) \gamma(h)$$

- If $\{\eta_t\}$ is Gaussian we have

$$\sqrt{T}(\bar{\eta} - \mu) \sim \mathrm{N}(0, v_T)$$

- The result above is approximate for many non-Gaussian time series
- In practice we also need to estimate $\gamma(h)$ from the data

Notes

Confidence Intervals for μ

- If $\gamma(h) \to 0$ as $h \to \infty$ then

$$v = \lim_{T\to\infty} v_T = \sum_{h=-\infty}^{\infty} \gamma(h) \text{ exists.}$$

- Further, if $\{\eta_t\}$ is Gaussian and

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty,$$

then an approximate large-sample 95% CI for μ is given by

$$\left[\bar{\eta} - 1.96\sqrt{\frac{v}{T}}, \bar{\eta} + 1.96\sqrt{\frac{v}{T}}\right]$$

Notes

Strategies for Estimating v

- Parametric:
  - Assume a parametric model $\gamma_{\boldsymbol{\theta}}(\cdot)$, and calculate

  $$\hat{v} = \sum_{h=-\infty}^{\infty} \gamma_{\hat{\boldsymbol{\theta}}}(h)$$

  based on the ACVF for that model
  - The standard error, v, will depend on the parameters $\boldsymbol{\theta}$ of the parametric model
- Nonparametric:
  - Estimate v by

  $$\hat{v} = \sum_{h=-\infty}^{\infty} \hat{\gamma}(h),$$

  where $\hat{\gamma}(\cdot)$ is an nonparametric estimate of ACVF

Notes

Examples of Parametric Forms for v

- i.i.d. Gaussian Noise: $v = \gamma(0) = \sigma^2 \Rightarrow$ CI reduces to the classical case:

$$\left[\bar{\eta} - 1.96\sqrt{\frac{\sigma^2}{T}}, \bar{\eta} + 1.96\sqrt{\frac{\sigma^2}{T}}\right]$$

- MA(1) process: We have

$$v = \sum_{h=-\infty}^{\infty} \gamma(h) = \gamma(-1) + \gamma(0) + \gamma(1)$$
$$= \gamma(0) + 2\gamma(1)$$
$$= \sigma^2(1 + \theta^2 + 2\theta) = \sigma^2(1+\theta)^2$$

- **Exercise**: Show for an AR(1) process we have

$$v = \frac{\sigma^2}{(1-\phi)^2}$$

Notes

Differencing

Instead of modeling trends, one can consider removing trends by differencing

- Define the first order difference operator ∇ as

$$\nabla Y_t = Y_t - Y_{t-1} = (1-B)Y_t,$$

  where B is the **backshift operator** and is defined as $BY_t = Y_{t-1}$.

- Similarly the general order difference operator $\nabla^q Y_t$ is **defined recursively** as $\nabla[\nabla^{q-1} Y_t]$

- The backshift operator of power q is defined as $B^q Y_t = Y_{t-q}$

In next slide we will see an example regarding the relationship between ∇^q and B^q

Notes

Difference and Backshift Notation

The second order difference is given by

$$\nabla^2 Y_t = \nabla[\nabla Y_t]$$

Notes

Difference and Backshift Notation

The second order difference is given by

$$\begin{aligned}\nabla^2 Y_t &= \nabla[\nabla Y_t] \\ &= \nabla[Y_t - Y_{t-1}]\end{aligned}$$

Notes

Difference and Backshift Notation

Stationary Processes

The second order difference is given by

$$\begin{aligned}\nabla^2 Y_t &= \nabla[\nabla Y_t] \\ &= \nabla[Y_t - Y_{t-1}] \\ &= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})\end{aligned}$$

Notes

Difference and Backshift Notation

Stationary Processes

The second order difference is given by

$$\begin{aligned}\nabla^2 Y_t &= \nabla[\nabla Y_t] \\ &= \nabla[Y_t - Y_{t-1}] \\ &= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \\ &= Y_t - 2Y_{t-1} + Y_{t-2}\end{aligned}$$

Notes

Difference and Backshift Notation

The second order difference is given by

$$\begin{aligned}\nabla^2 Y_t &= \nabla[\nabla Y_t]\\ &= \nabla[Y_t - Y_{t-1}]\\ &= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})\\ &= Y_t - 2Y_{t-1} + Y_{t-2}\\ &= (1 - 2B + B^2)Y_t\end{aligned}$$

In the next slide we will see an example of using differening to remove the trend

Notes

Removing Trend via Differening

Stationary Processes
School of MATHEMATICAL AND STATISTICAL SCIENCES Clemson University
Review
Some Examples of Stationary Processes
Estimation and Inference for Mean Functions
Differencing

Consider a time series data with a linear trend (i.e., $\{Y_t = \beta_0 + \beta_1 t + \eta_t\}$) where η_t is a stationary time series. Then first order differencing results in a stationary series with no trend. To see why

$$\begin{aligned}\nabla Y_t &= Y_t - Y_{t-1}\\ &= (\beta_0 + \beta_1 t + \eta_t) - (\beta_0 + \beta_1(t-1) + \eta_{t-1})\\ &= \beta_1 + \eta_t - \eta_{t-1}\end{aligned}$$

This is the sum of a stationary series and a constant, and therefore we have successfully remove the linear trend.

Notes

Notes on Differening

Stationary Processes
School of MATHEMATICAL AND STATISTICAL SCIENCES Clemson University
Review
Some Examples of Stationary Processes
Estimation and Inference for Mean Functions
Differencing

- A polynomial trend of order q can be removed by q-th order differencing
- By q-th order differencing a time series we are shortening its length by q
- Differencing does not allow you to estimate the trend, only to remove it. *Therefore it is not appropriate if the aim of the analysis is to describe the trend*

Notes

Seasonal Differening

- The lag-d difference operator, ∇_d, is defined by

$$\nabla_d Y_t = Y_t - Y_{t-d} = (1 - B^d)Y_t.$$

Note: This is NOT ∇^d!

- **Example**: Consider data that arise from the model $Y_t = \beta_0 + \beta_1 t + s_t + \eta_t$, which has a linear trend and seasonal component that repeats itself every d time points. Then by just seasonal differencing (lag-d differening here) this series becomes stationary.

$$\begin{aligned}\nabla_d Y_t &= Y_t - Y_{t-d}\\ &= [\beta_0 + \beta_1 t + s_t + \eta_t] - [\beta_0 + \beta_1(t-d) + s_{t-d} + \eta_{t-d}]\\ &= d\beta_1 + \eta_t - \eta_{t-d}\end{aligned}$$

Notes

Summary

In this lecture, we discuss

- White Noise Processes, MA(1), AR(1)
- Estimation and Inference of the Mean of Stationary Processes
- Differencing to Remove Trend and Seasonality

The most important `R` function for this lecture is `arima.sim`, which can be used to simulate MA(1), AR(1), and more general ARIMA models

Stationary Processes
Review
Some Examples of Stationary Processes
Estimation and Inference for Mean Functions
Differencing

Notes

Notes