시계열자료분석팀

5팀

김민우
김영호
정승연
조건우
조웅빈

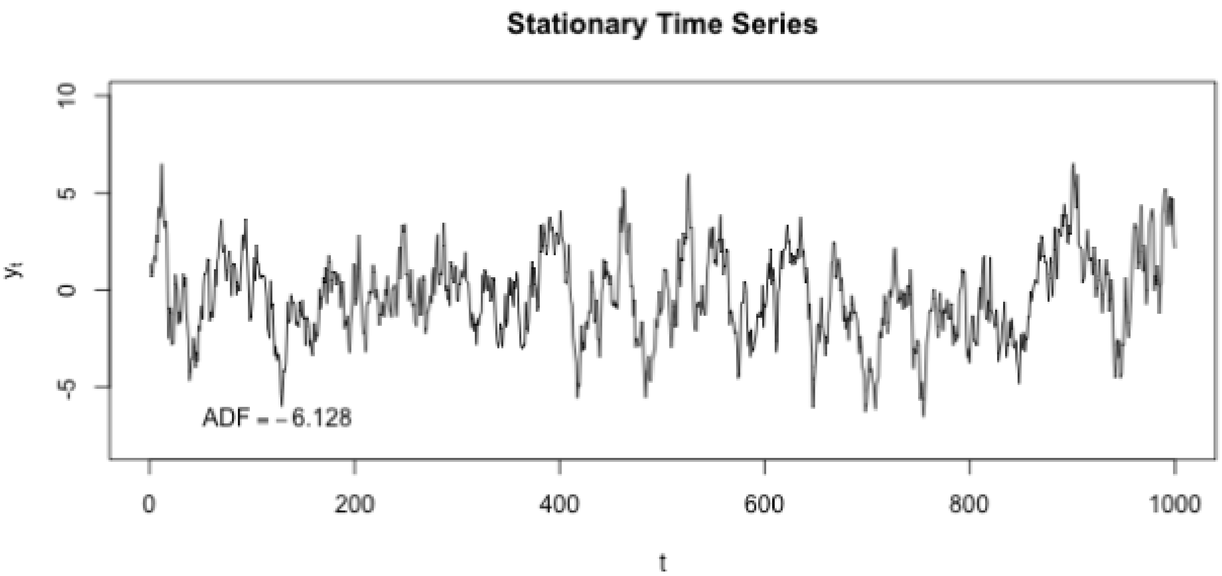
CONTENTS

1

시계열 살펴보기

시계열 자료란?

연도별, 월별, 일별 등 **시간 순서**에 따라 관측된 자료



$$X_t, t = 1, 2, \ldots .$$

시간 t에 따라 이산형과 연속형으로 구분

시계열 자료 분석의 목적

시계열 자료 분석 :

시간 순서대로 정렬된 자료를 분석해 의미있는 요약과 통계적 정보를 도출하는 것

예측을 위한 분석

추세분석, 평활법, 분해법,
ARIMA 모형 등

시스템 이해/제어

스펙트럼 분석, 개입분석,
전이함수 모형 등

시계열 자료 분석의 목적

시계열 자료 분석 :

시간 순서대로 정렬된 자료를 분석해 의미있는 요약과 통계적 정보를 도출하는 것

예측을 위한 분석

추세분석, 평활법, 분해법,

ARIMA 모형 등

시스템 이해/제어

스펙트럼 분석, 개입분석,

전이함수 모형 등

시계열 자료 분석의 목적

시계열 자료 분석 :

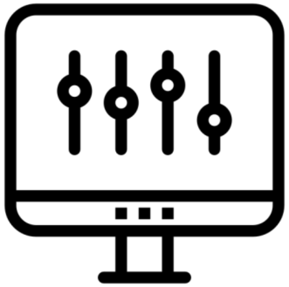
시간 순서대로 정렬된 자료를 분석해 의미있는 요약과 통계적 정보를 도출하는 것

예측을 위한 분석

추세분석, 평활법, 분해법,

ARIMA 모형 등

시스템 이해/제어



스펙트럼 분석, 개입분석,

전이함수 모형 등

시계열 자료 분석의 목적

시계열 자료 분석 :

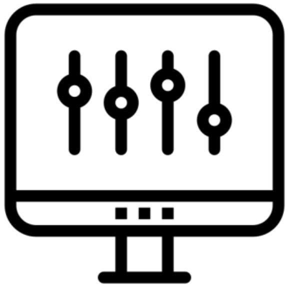
시간 순서대로 정렬된 자료를 분석해 의미있는 요약과 통계적 정보를 도출하는 것

예측을 위한 분석

추세분석, 평활법, 분해법,

ARIMA 모형 등

시스템 **이해/제어**



스펙트럼 분석, 개입분석,

전이함수 모형 등

1 시계열 살펴보기

시계열 자료의 구성요소

시계열 자료

- 시간의 영향 O
  - 체계적 성분
- 시간의 영향 X
  - 불규칙 성분

시계열 자료의 구성요소

시계열 자료

시간의 영향 O
체계적 성분

- 추세변동 (Trend)
- 순환변동 (Cycle)
- 계절변동 (Seasonal Variation)

시간의 영향 X
불규칙 성분

시계열 자료의 구성요소

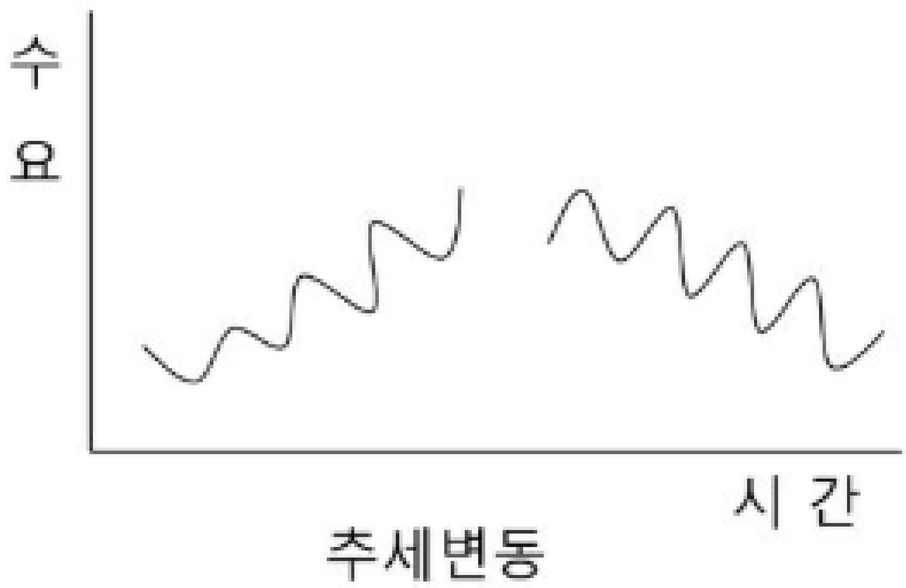
- 시계열 자료
  - 시간의 영향 O
    - 체계적 성분
      - 추세변동 (Trend)
      - 순환변동 (Cycle)
      - 계절변동 (Seasonal Variation)
  - 시간의 영향 X
    - 불규칙 성분
      - 우연변동/불규칙변동 (Random fluctuation)

시계열 자료의 구성요소

- 시계열 자료
  - 시간의 영향 O
    - 체계적 성분
      - ✓ 추세변동 (Trend)
      - ✓ 순환변동 (Cycle)
      - ✓ 계절변동 (Seasonal Variation)
  - 시간의 영향 X
    - 불규칙 성분
      - ✓ 우연변동/불규칙변동 (Random fluctuation)

시계열 자료의 구성요소

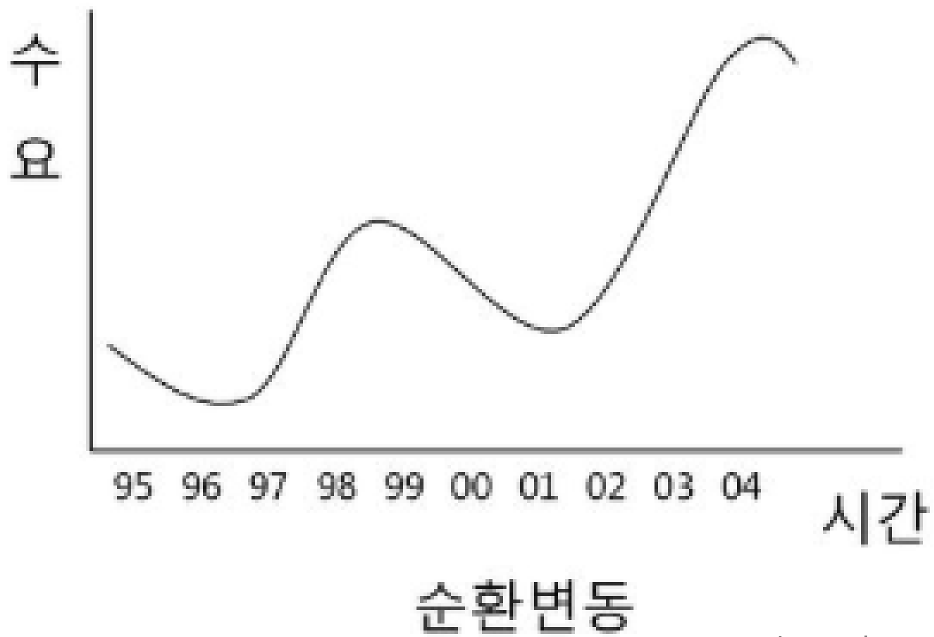
1) 추세변동 (Trend)



시간이 경과함에 따라 **관측값이 증감**하는 변동

시계열 자료의 구성요소

2) 순환변동 (Cycle)

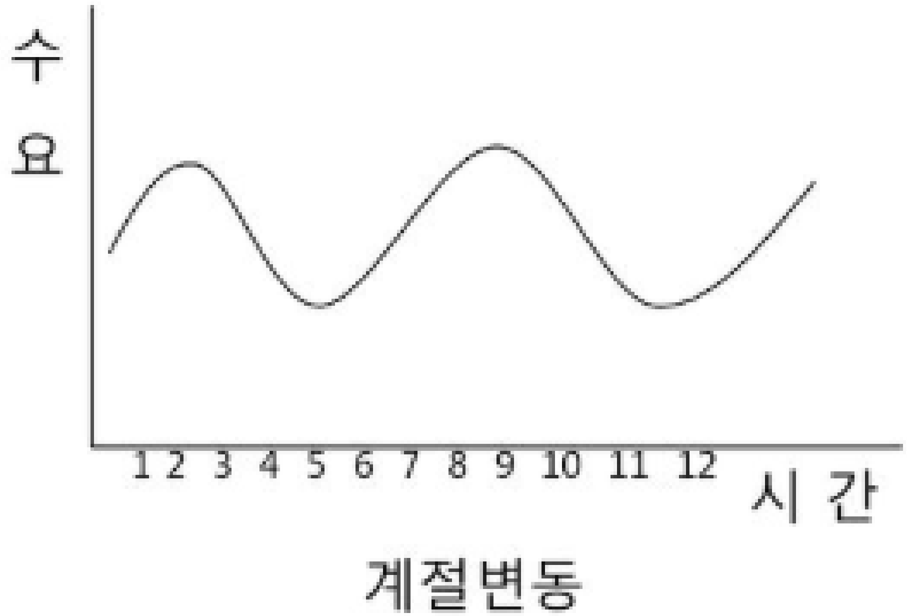


(ex. 남극 빙하의 변화 …)

주기적인 변화를 가지나, 그 변화가 계절에 의한 것이 아니며,
주기가 긴 경우의 변동

시계열 자료의 구성요소

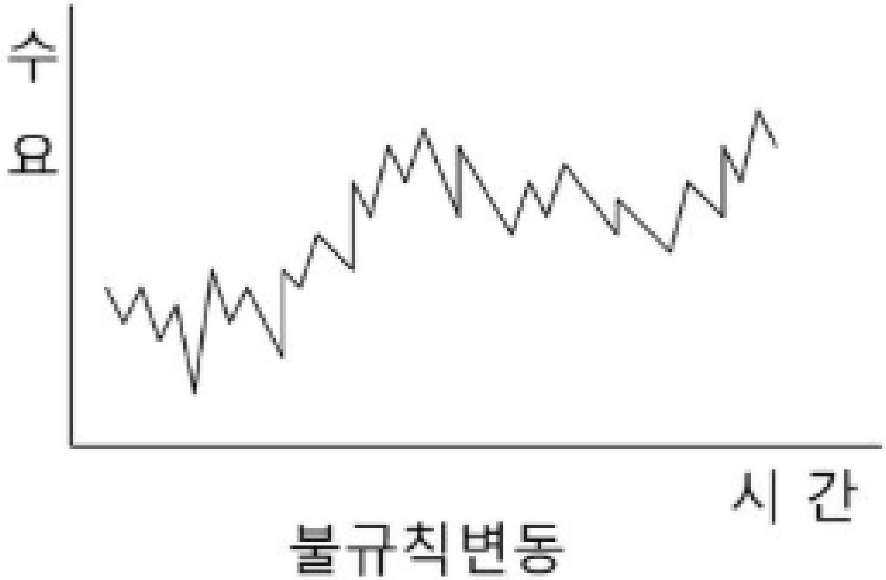
3) 계절변동 (Seasonal Variation)



주별·월별·계절별과 같은 **주기적인 성분**에 의한 변동

시계열 자료의 구성요소

4) 우연변동/불규칙변동 (Random fluctuation)



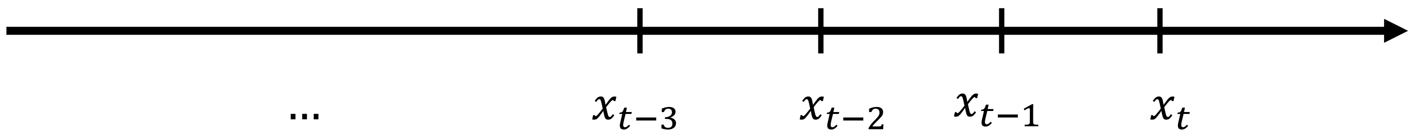
시간과 무관하게, **무작위 원인**에 의해 나타나는 변동

2

정상성

정상성이 필요한 이유

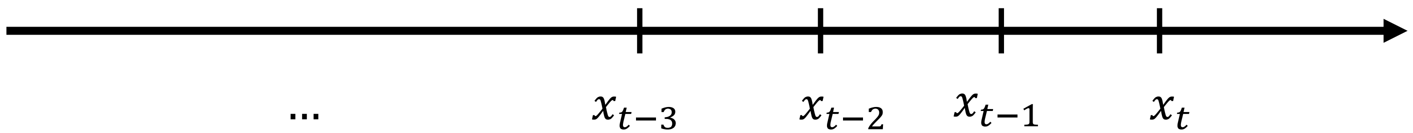
시계열 모형은 각 시점 t의 확률변수 $\{X_t\}$의 관찰값 $\{x_t\}$의 결합분포로 이루어져있음



2 정상성

정상성이 필요한 이유

시계열 모형은 각 시점 t의 확률변수 $\{X_t\}$의 관찰값 $\{x_t\}$의 결합분포로 이루어져있음



미래 예측을 위해서는

무한한 시점의 결합 분포 $\{x_1, x_2, \dots, x_n\}$ 고려 필요!

But, 현실적으로는 불가능한 일 …

2 정상성

정상성이 필요한 이유

시계열 모형은 각 시점 t의 확률변수 $\{X_t\}$의 관찰값 $\{x_t\}$의 결합분포로 이루어져있음

정상성을 통해 몇 가지 가정으로 이를 대체!

(대부분의 시계열 이론들이 정상성을 가정)

미래 예측을 위해서는

무한한 시점의 결합 분포 $\{x_1, x_2, \dots, x_n\}$ 고려 필요!

But, 현실적으로는 불가능한 일 …

정상성이 필요한 이유

“정상성(Stationarity)”

시계열 모형은 각 시점 t의 확률변수 $\{X_t\}$의 관찰값 $\{x_t\}$의 결합분포로 이루어져있음

시계열의 확률적 성질(평균, 분산 등)이

시간의 흐름에 영향을 받지 않는 것

… x_{t-3} x_{t-2} x_{t-1} x_t

강정상성

(Strict Stationarity)

약정상성

(Weak Stationarity)

미래 예측을 위해서는

무한한 시점의 결합 분포 $\{x_1, x_2, \ldots, x_n\}$ 고려 필요!

But, 현실적으로는 불가능한 일 …

강정상성

$$(X_{t_1}, \dots, X_{t_n}) \overset{d}{=} (X_{t_{1+h}}, \dots, X_{t_{n+h}})$$

모든 h와 $n > 0$에 대하여

결합확률밀도가 시간(t)을 바꾸어도 동일

(분포의 불변성)

But, 현실적으로 강정상성을 만족하는 것은 매우 **어려움**

강정상성

$$(X_{t_1}, \dots, X_{t_n}) \overset{d}{=} (X_{t_{1+h}}, \dots, X_{t_{n+h}})$$

모든 h와 $n > 0$에 대하여

결합확률밀도가 시간(t)을 바꾸어도 동일

(분포의 불변성)

But, 현실적으로 강정상성을 만족하는 것은 매우 **어려움**

강정상성보다 완화된 조건인

약정상성 사용!

약정상성

약정상성의 3가지 조건

1. $E[|X_t|]^2 < \infty, \forall t \in \mathbb{Z}$
2. $E[X_t] = m, \forall t \in \mathbb{Z}$
3. $\gamma_X(h) := Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$

2 정상성

약정상성

약정상성의 3가지 조건

1 $E[|X_t|]^2 < \infty, \forall t \in \mathbb{Z}$

분산과 관련된 **2차적률**이 존재하고, **시점 t에 관계없이 일정!**

3 $\gamma_X(h) := Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$

2 정상성

약정상성

약정상성의 3가지 조건

1 $E[|X_t|]^2 < \infty, \forall t \in \mathbb{Z}$

2 $E[X_t] = m, \forall t \in \mathbb{Z}$

3 $\gamma_X(h) := Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$

평균이 상수로 시점 t에 관계없이 일정!

2 정상성

약정상성

약정상성의 3가지 조건

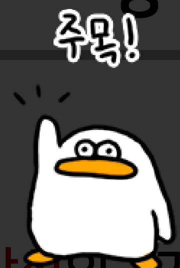
1. $E[|X_t|]^2 < \infty, \forall t \in \mathbb{Z}$

"**자기공분산**"은 **시차 h에만 의존**, 시점 t와는 무관!

2. $E[X_t] = m, \forall t \in \mathbb{Z}$

3. $\gamma_X(h) := Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$

약정상성



약정상성의 3가지 조건

1. $E[|X_t|]^2 < \infty, \forall t \in \mathbb{Z}$
2. $E[X_t] = m, \forall t \in \mathbb{Z}$
3. $\gamma_X(h) = Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$

여기서 잠깐! 자기공분산이란?

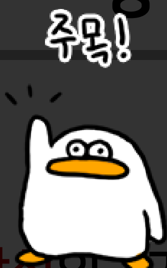
자기자신(X_t)과 몇 시점 이후 자기자신(X_{t+h})과의 공분산!

$$\gamma_X(h) = Cov(X_t, X_{t+h})$$

그렇다면 자기상관도 있겠네?

자기자신(X_t)과 몇 시점 이후 자기자신(X_{t+h})과의 자기상관계수!

$$\rho_X(h) = Corr(X_t, X_{t+h}) = \frac{\gamma_X(h)}{\gamma_X(0)}$$



약정상성

약정상성의 3가지 조건

1 $E[|X_t|]^2 < \infty, \forall t \in \mathbb{Z}$

2 $E[X_t] = m, \forall t \in \mathbb{Z}$

3 $\gamma_X(h) = Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$

여기서 잠깐! 자기공분산이란?

자기자신(X_t)과 몇 시점 이후 자기자신(X_{t+h})과의 공분산!

$$\gamma_X(h) = Cov(X_t, X_{t+h})$$

그렇다면 자기상관도 있겠네?

자기자신(X_t)과 몇 시점 이후 자기자신(X_{t+h})과의 자기상관계수!

$$\rho_x(h) = Corr(X_t, X_{t+h}) = \frac{\gamma_X(h)}{\gamma_X(0)}$$

약정상성

약정상성의 3가지 조건

약정상성을 만족하는지 확인하기 위해서는

1차적률인 $E[X_t]$와 **자기공분산** $\gamma_X(h)$만 고려하면 된다!

1 $E[|X_t|]^2 < \infty, \forall t \in \mathbb{Z}$

2 $E[X_t] = m, \forall t \in \mathbb{Z}$

자기공분산 $\gamma_X(h)$에서 $h = 0$인 경우

$$Cov(X_t, X_{t+h}) = Cov(X_t, X_t)$$
$$= Var(X_t) \leftrightarrow E[|X_t|]^2 < \infty$$

3 $\gamma_X(h) := Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$

약정상성

약정상성의 3가지 조건

1 $E[|X_t|]^2 < \infty, \forall t \in \mathbb{Z}$

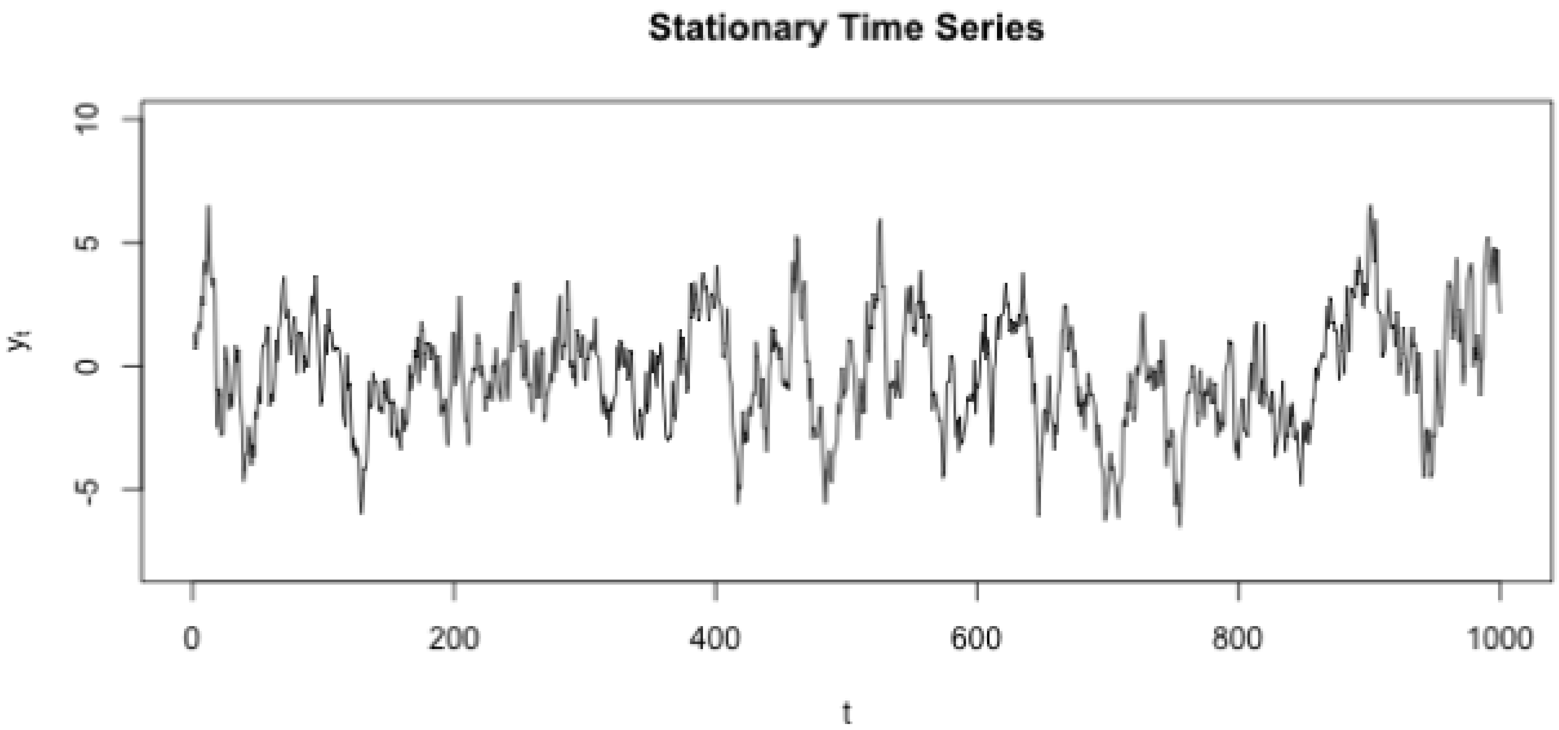
따라서, 3가지 조건 중

2, 3번 조건만 확인하면 된다!

2 $E[X_t] = m, \forall t \in \mathbb{Z}$

3 $\gamma_X(h) := Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$

정상시계열 자료



평균 및 분산이 일정하여 정상성을 만족함

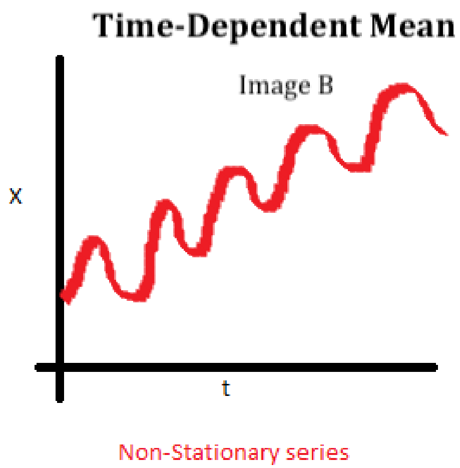
3

정상화

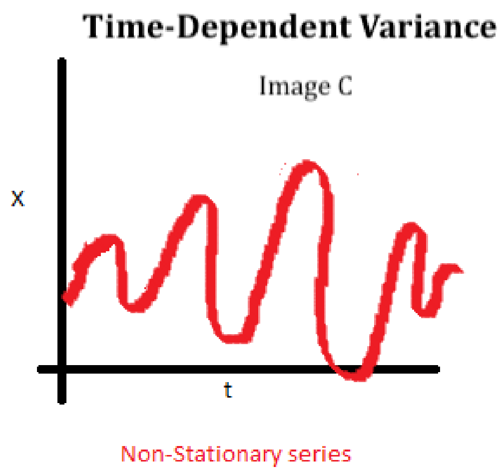
비정상 시계열

: 정상 시계열로 변환하여 예측

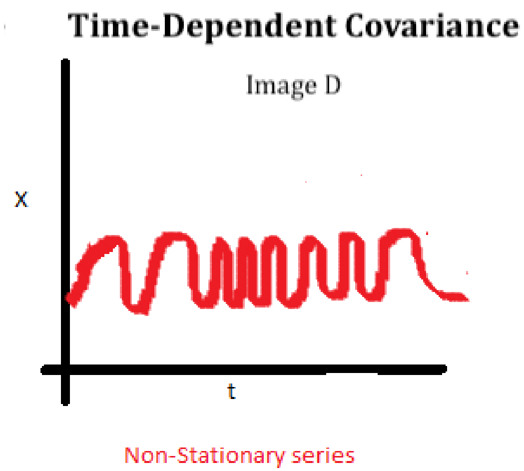
평균이 일정하지 않은 경우



분산이 일정하지 않은 경우



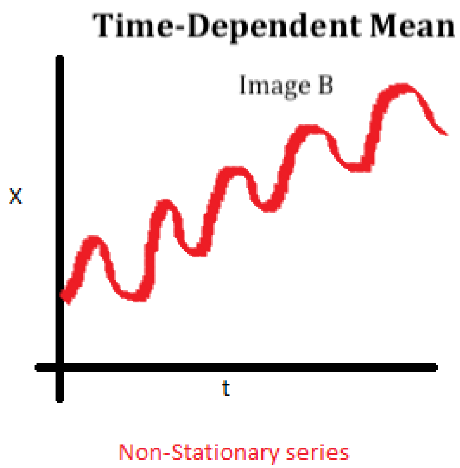
공분산이 시점에 의존하는 경우



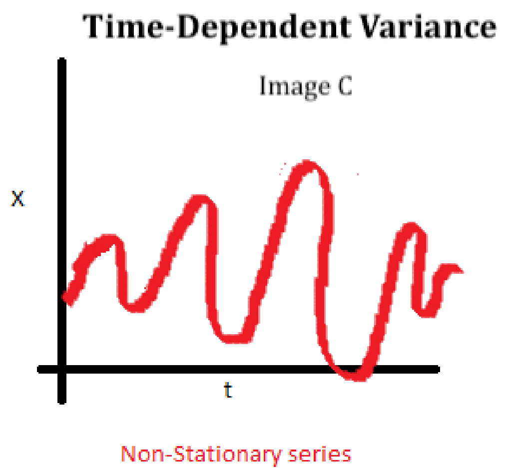
비정상 시계열

: 정상 시계열로 변환하여 예측

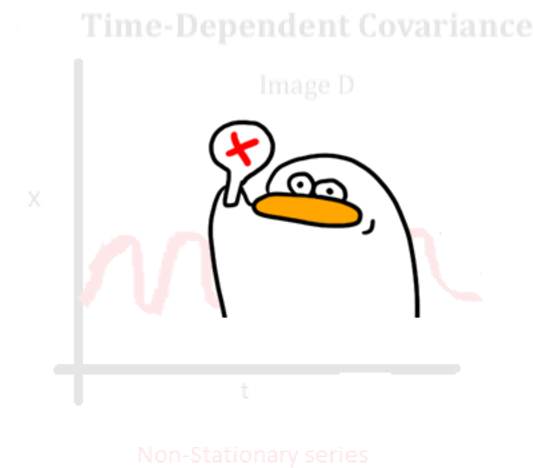
평균이 일정하지 않은 경우



분산이 일정하지 않은 경우



공분산이 시점에 의존하는 경우



정상화 과정

분산이 일정하지 않은 경우

시간의 흐름에 따라 변동폭이 커지거나, 작아지는 자료들이 존재 (**이분산성**)

→ 변환을 통해 시간의 흐름에 따라 분산이 일정하도록 함

정상화 과정

분산이 일정하지 않은 경우

시간의 흐름에 따라 변동폭이 커지거나, 작아지는 자료들이 존재 (**이분산성**)

→ 변환을 통해 시간의 흐름에 따라 분산이 일정하도록 함

분산 안정화 변환

(Variance Stabilizing Transformation; VST)

정상화 과정

분산이 일정하지 않은 경우

분산 안정화 변환 (VST)

- Log-transformation

$$f(X_t) = \log(X_t)$$

- Square root transformation

$$f(X_t) = \sqrt{X_t}$$

- Box-Cox transformation

$$f_\lambda(X_t) = \begin{cases} \dfrac{X_t^\lambda - 1}{\lambda}, X_t \geq 0, \lambda > 0 \\ logX_t, \lambda = 0 \end{cases}$$

정상화 과정

평균이 일정하지 않은 경우

$$X_t = m_t + s_t + Y_t$$

m_t : 추세, s_t : 계절성, Y_t : 정상성을 만족하는 오차

- 추세가 존재
- 계절성이 존재
- 추세와 계절성 모두 존재

추세 및 계절성을 제거

정상화 과정

평균이 일정하지 않은 경우

$$X_t = m_t + s_t + Y_t$$

m_t : 추세, s_t : 계절성, Y_t : 정상성을 만족하는 오차

- 추세가 존재
- 계절성이 존재
- 추세와 계절성 모두 존재

추세 및 계절성을 제거

정상화 과정

평균이 일정하지 않은 경우

$$X_t = m_t + s_t + Y_t$$

m_t : 추세, s_t : 계절성, Y_t : 정상성을 만족하는 오차

1 회귀

2 평활

3 차분

추세 및 계절성을 제거

평균이 일정하지 않은 경우: 회귀

추세만 존재하는 경우

Step1) 추세 성분 m_t를 시간 t에 대한 선형회귀식으로 나타냄

$$X_t = m_t + Y_t , E(Y_t) = 0$$

평균이 일정하지 않은 경우: 회귀

추세만 존재하는 경우

Step1) 추세 성분 m_t를 시간 t에 대한 선형회귀식으로 나타냄

$$X_t = m_t + Y_t \, , E(Y_t) = 0$$

$$m_t = c_0 + c_1 t + c_2 t^2 + \cdots + c_p t^p$$

평균이 일정하지 않은 경우: 회귀

추세만 존재하는 경우

Step2) 선형회귀식의 계수를 최소제곱법(OLS)를 통해 추정

$$\left(\widehat{c_o}, \ldots, \widehat{c_p}\right) = \underset{c}{\operatorname{argmin}} \sum_{t=1}^{n} (X_t - m_t)^2$$

평균이 일정하지 않은 경우: 회귀

추세만 존재하는 경우

Step2) 선형회귀식의 계수를 최소제곱법(OLS)를 통해 추정

$$(\widehat{c_o}, \dots, \widehat{c_p}) = \underset{c}{\operatorname{argmin}} \sum_{t=1}^{n} (X_t - m_t)^2$$

Step3) 추정한 추세를 시계열에서 제거

$$X_t - \widehat{m}_t \approx Y_t$$

평균이 일정하지 않은 경우: 회귀

계절성만 존재하는 경우

Step1) 시계열을 다음과 같이 주기가 d 인 계절성만을 가진다고 가정 후 계절 성분 s_t를 다음과 같이 시간 t 에 대한 회귀식으로 나타냄

$$X_t = s_t + Y_t \,, E(Y_t) = 0$$

$$where\ s_{t+d} = s_t = s_{t-d}$$

$$s_t = a_0 + \sum_{j=1}^{k} \left(a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)\right)$$

평균이 일정하지 않은 경우: 회귀

계절성만 존재하는 경우

Step2) 적절한 λ_j와 k를 선택한 후, OLS를 통하여 a_j와 b_j를 추정

$$s_t = a_0 + \sum_{j=1}^{k} \left(a_j \cos\left(\lambda_j t\right) + b_j \sin\left(\lambda_j t\right)\right)$$

평균이 일정하지 않은 경우: 회귀

계절성만 존재하는 경우

Step2) 적절한 λ_j와 k를 선택한 후, OLS를 통하여 a_j와 b_j를 추정

$$s_t = a_0 + \sum_{j=1}^{k}\left(a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)\right)$$

λ_j는 주기가 2π인 함수의 주기와 데이터의 주기를 맞춰 주기 위한 값으로,

1) $f_1 = [n/d]$ (n = data 수, d = 주기) → $f_j = jf_1$

2) $\lambda_j = f_j(2\pi/n)$

평균이 일정하지 않은 경우: 회귀

계절성만 존재하는 경우

Step3) 추정한 계절성을 시계열에서 제거

$$s_t = a_0 + \sum_{j=1}^{k} (a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t))$$

$$X_t - \hat{s}_t \approx Y_t$$

3

정상화

평균이 일정하지 않은 경우: 회귀

추세와 계절성이 동시에 존재하는 경우

$$X_t = m_t + s_t + Y_t\ , E(Y_t) = 0$$

Step1) 추세만 존재하는 경우의 정상화 과정과
계절성만 존재하는 경우의 정상화 과정을 차례대로 진행

Step2) 이후에도 여전히 추세가 남아 있다면 다시 추세를 제거

평균이 일정하지 않은 경우: 회귀

추세와 계절성이 동시에 존재하는 경우

$$X_t = m_t + s_t + Y_t, E(Y_t) = 0$$

Step1) 추세만 존재하는 경우의 정상화 과정과 계절성만 존재하는 경우의 정상화 과정을 차례대로 진행

Step2) 이후에도 여전히 추세가 남아 있다면 다시 추세를 제거

회귀방법의 단점

최소 제곱법은 오차항의 독립성을 가정하고 전개

But, 시계열의 오차항은 독립성을 가정하지는 않아

추정이 정확하지 않을 수 있음

평활(smoothing)

평활법을 사용하는 이유

앞선 회귀방식의 OLS의 경우 전체 데이터를 한번에 사용하여

추정하는 방식으로, 국소적 변동이 존재하는 경우 부적합

평활(smoothing)

평활법을 사용하는 이유

앞선 회귀방식의 OLS의 경우 전체 데이터를 한번에 사용하여

추정하는 방식으로, 국소적 변동이 존재하는 경우 부적합

이러한 경우 **평활법**을 사용하여 정상성 확보

평활(smoothing)

평활법을 사용하는 이유

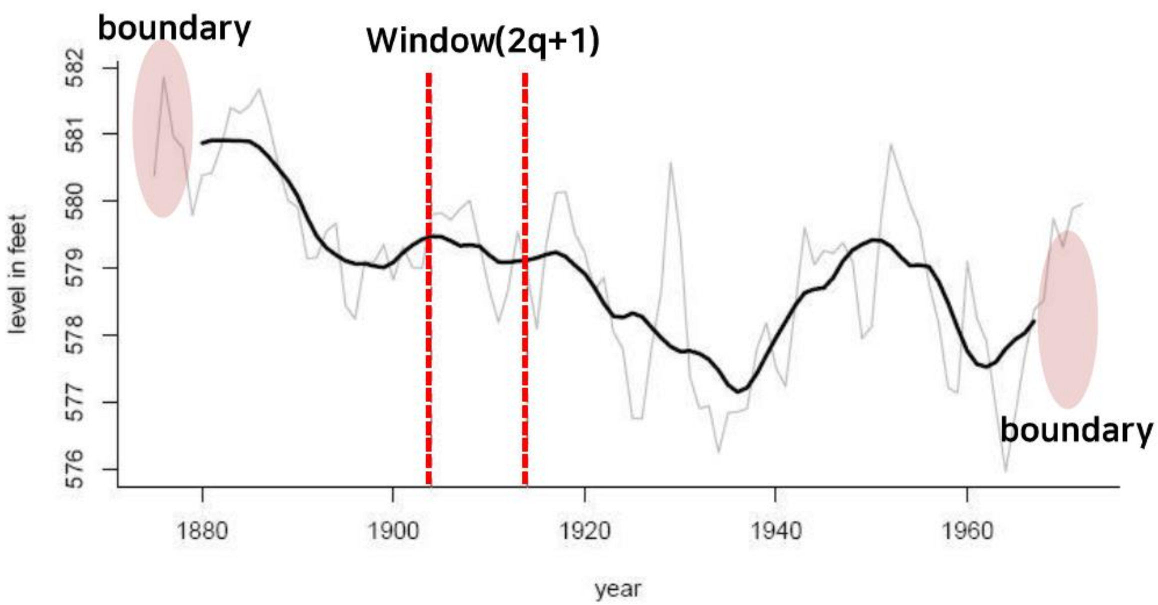
앞선 회귀방식의 OLS의 경우 전체 데이터를 한번에 사용하여 추정하는 방식으로, 국소적 변동이 존재하는 경우 부적합

이러한 경우 **평활법**을 사용하여 정상성 확보

평활법	언제?
이동평균법(MA)	추세만 존재
지수평활법	추세만 존재
계절평활법	계절성만 존재
Classical Decomposition Algorithm	둘 다 존재

이동평균 평활법(Moving Average Smoothing)

추세만 존재하는 경우



추세만 존재하는 경우 활용하며

일정 기간마다 평균을 계산하여 추세를 추정하는 방법

이동평균 평활법(Moving Average Smoothing)

추세만 존재하는 경우

Step1) 길이가 2q+1인 구간의 평균을 구한다.

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^{j=q} (m_{t+j} + Y_{t+j})$$

$$= \frac{1}{2q+1} \sum_{j=-q}^{q} m_{t+j} + \frac{1}{2q+1} \sum_{j=-q}^{q} Y_{t+j}$$

이동평균 평활법(Moving Average Smoothing)

추세만 존재하는 경우

Step2) 앞서 구한 식에 추세 성분 m_t를 대입한다. (추세는 linear하다고 가정)

$$m_t = c_0 + c_1 t$$

$$\frac{1}{2q+1}\sum_{j=-q}^{q} m_{t+j} = c_0 + c_1 t = m_t, \qquad t \in [q+1, n-q]$$

$$\frac{1}{2q+1}\sum_{j=-q}^{q} Y_{t+j} \approx E(Y_t) = 0 \ (by\ WLLN)$$

Step3) 추정한 추세를 제거하여 마무리

이동평균 평활법(Moving Average Smoothing)

추세만 존재하는 경우

Step2) 앞서 구한 식에 추세 성분 m_t를 대입한다. (추세는 linear하다고 가정)

$$m_t = c_0 + c_1 t$$

$$\frac{1}{2q+1}\sum_{j=-q}^{q} m_{t+j} = c_0 + c_1 t = m_t, \qquad t \in [q+1, n-q]$$

$$\frac{1}{2q+1}\sum_{j=-q}^{q} Y_{t+j} \approx E(Y_t) = 0 \ (by\ WLLN)$$

Step3) 추정한 추세를 제거하여 마무리

최근 자료와 과거의 자료의 가중치가 $\frac{1}{2q+1}$로 모두 동일하고

현시점인 t시점을 예측할 때, t시점 이후의 데이터 활용은 불가능

따라서, 과거의 데이터에만 의존하여 추정하는 **지수평활법** 필요!

지수평활법(Exponential Smoothing)

추세만 존재하는 경우

$$\hat{m}_1 = X_1, \quad a \in [0,1]$$

$$\hat{m}_2 = aX_2 + (1-a)\hat{m}_1 = aX_2 + (1-a)X_1$$

$$\hat{m}_3 = aX_3 + (1-a)\hat{m}_2 = aX_3 + a(1-a)X_2 + (1-a)^2\, X_1$$

$$\vdots$$

$$\hat{m}_t = aX_t + (1-a)\hat{m}_{t-1} = \sum_{j=0}^{t-2} a\,(1-a)^j X_{t-j} + (1-a)^{t-1} X_1$$

위와 같이 과거 시점의 데이터를 이용한 추세 추정

과거의 값일수록 가중치 값이 지수적으로 줄어드는 방법 사용

지수평활법(Exponential Smoothing)

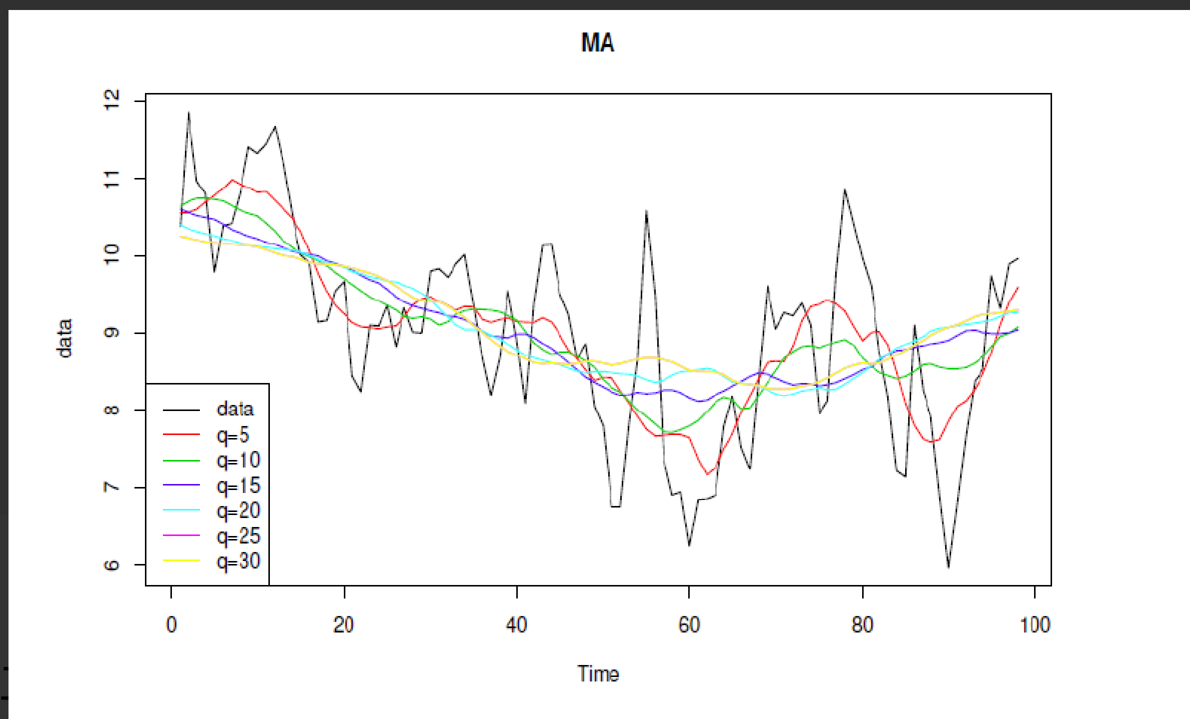
추세만 존재하는 경우

$$\hat{m}_1 = X_1$$

평활법의 Tuning Parameter

이동평균법의 q, 지수평활의 a는 직접 선택하는 문제 발생

가장 일반적으로는 cross-validation을 통해 최적의 parameter 선택



평활법의 Tuning Parameter

Ex) MA 모형의 경우

$$W_t = \frac{1}{2q+1}\sum_{j=-q}^{j=q}(m_{t+j} + Y_{t+j})$$

$$= \underbrace{\frac{1}{2q+1}\sum_{j=-q}^{q} m_{t+j}}_{\approx\, m_t \ if\ q\ is\ small} + \underbrace{\frac{1}{2q+1}\sum_{j=-q}^{q} Y_{t+j}}_{\approx\, 0\ if\ q\ is\ large}$$

지수평활법(Exponential Smoothing)

추세만 존재하는 경우

$$\hat{m}_1 = X_1$$

$$\hat{m}_2 = aX_2 + (1-a)\hat{m}_1 = aX_2 + (1-a)X_1$$

$$\hat{m}_3 = aX_3 + (1-a)\hat{m}_2 = aX_3 + a(1-a)X_2 + (1-a)^2 X_1$$

$$\vdots$$

$$\hat{m}_t = aX_t + (1-a)\hat{m}_{t-1} = \sum_{j=0}^{t-2} a(1-a)^j X_{t-j} + (1-a)^{t-1} X_1$$

위와 같이 과거 시점의 데이터를 이용한 추세 추정

과거의 값일수록 가중치 값이 지수적으로 줄어드는 방법 사용

지수평활법(Exponential Smoothing)

추세만 존재하는 경우

$$\hat{m}_1 = X_1$$

$$\hat{m}_2 = aX_2 + (1-a)\hat{m}_1 = aX_2 + (1-a)X_1$$

$$\hat{m}_3 = aX_3 + (1-a)\hat{m}_2 = aX_3 + a(1-a)X_2 + (1-a)^2 X_1$$

$$\vdots$$

$$\hat{m}_t = aX_t + (1-a)\hat{m}_{t-1} = \sum_{j=0}^{t-2} a(1-a)^j X_{t-j} + (1-a)^{t-1} X_1$$

평활법의 Tuning Parameter

$$W_t = \frac{1}{2q+1}\sum_{j=-q}^{j=q}(m_{t+j} + Y_{t+j})$$

$$= \underbrace{\frac{1}{2q+1}\sum_{j=-q}^{q} m_{t+j}}_{\approx\, m_t \text{ if } q \text{ is small}} + \underbrace{\frac{1}{2q+1}\sum_{j=-q}^{q} Y_{t+j}}_{\approx\, 0 \text{ if } q \text{ is large}}$$

Window Size q가 커지면,

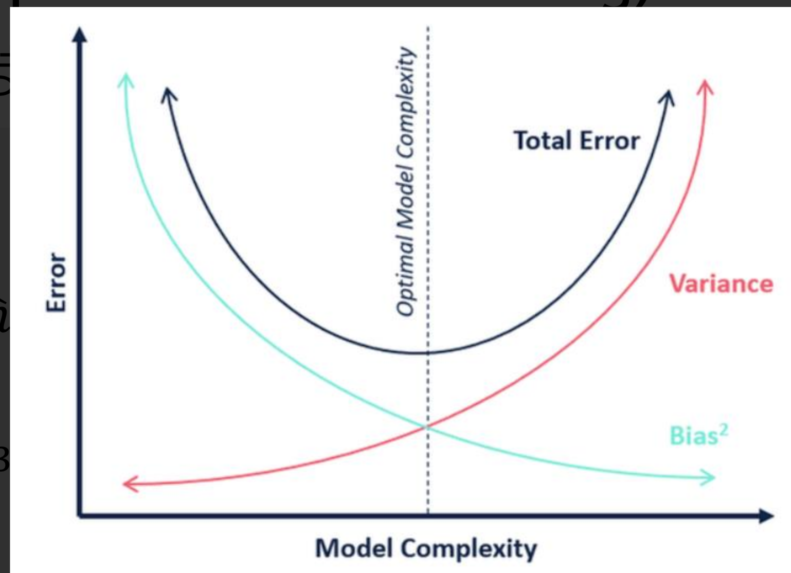
오차항은 0으로 수렴하지만, 정확한 추세 추정 X

Window Size q가 작아지면,

정확한 추세 추정은 가능하지만, 오차항이 0으로 수렴 X

지수평활법(Exponential Smoothing)

평활법의 Tuning Parameter



Bias-Variance Trade off

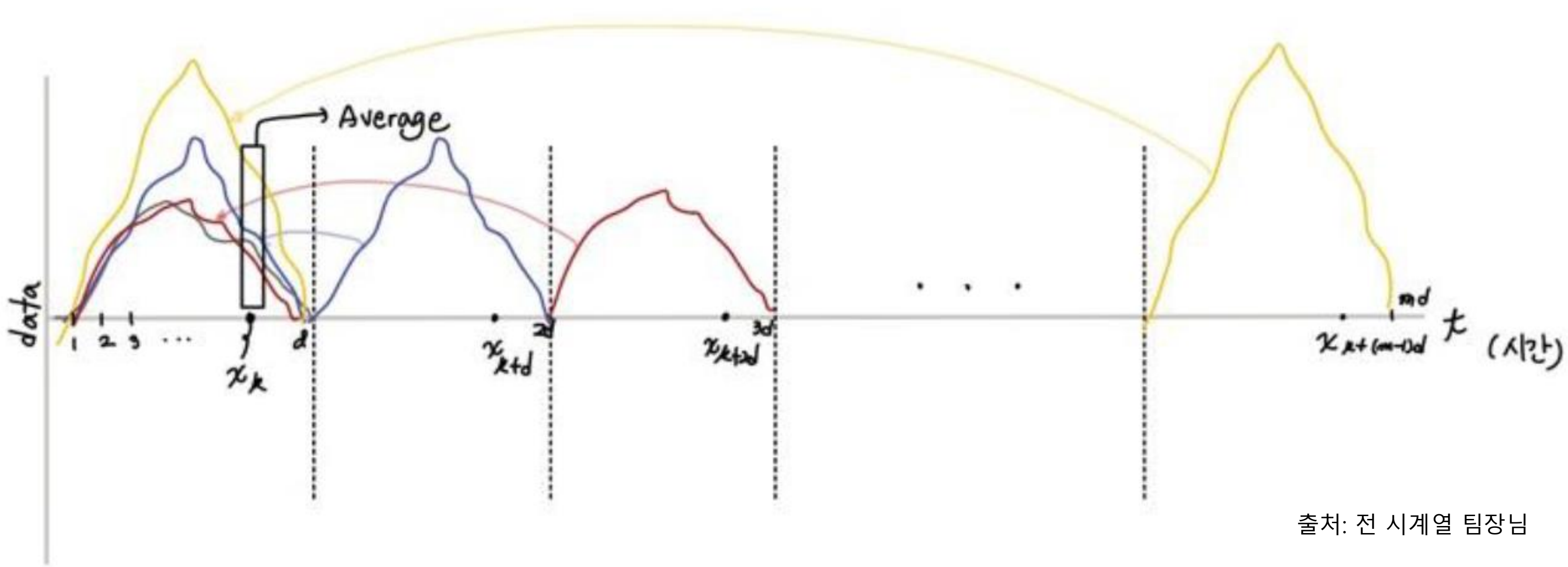
Window Size q가 커지면, Bias는 커지고 Variance는 작아짐

Window Size q가 작아지면, Bias는 작아지고 Variance는 커짐

더 자세한 내용이 궁금하다면 데마 1주차 클린업 참고!

계절 평활(Seasonal Smoothing)

계절성만 존재하는 경우



출처: 전 시계열 팀장님

주기가 d인 관측치를 한 주기 안으로 모두 겹친 후,

겹쳐진 값들을 평균내어 추정

계절 평활(Seasonal Smoothing)

계절성만 존재하는 경우

Step1) 아래와 같이 계절성분 추정

$$\hat{s}_k = \frac{1}{m}\left(x_k + x_{k+d} + \cdots + x_{k+(m-1)d}\right) = \frac{1}{m}\sum_{j=0}^{m-1} x_{k+jd}$$

$$(k = 1, \cdots, d \ and \ \ m \ is \ the \ number \ of \ obs. in \ th \ kth \ seasonal \ component)$$

계절 평활(Seasonal Smoothing)

계절성만 존재하는 경우

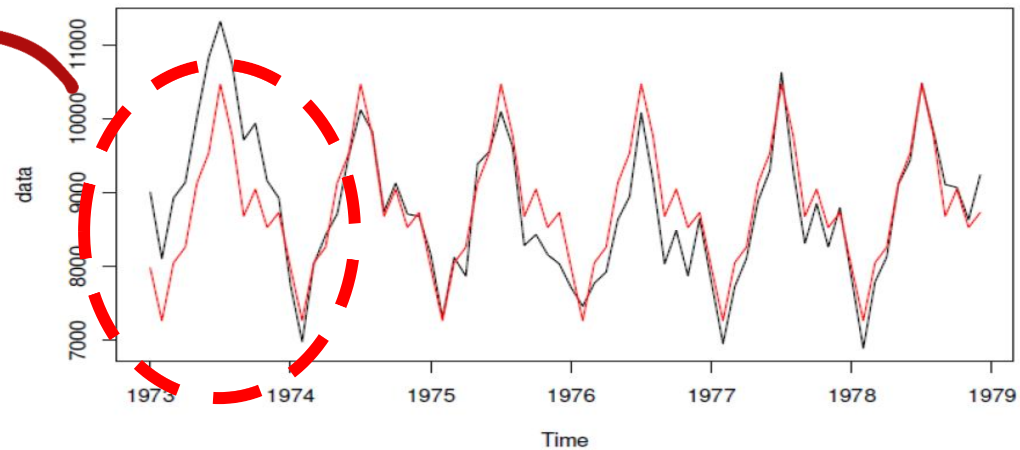
Step1) 아래와 같이 계절성분 추정

$$\hat{s}_k = \frac{1}{m}\left(x_k + x_{k+d} + \cdots + x_{k+(m-1)d}\right) = \frac{1}{m}\sum_{j=0}^{m-1} x_{k+jd}$$

$$(k = 1, \cdots, d \ and \ \ m \ is \ the \ number \ of \ obs. in \ th \ kth \ seasonal \ component)$$

Step2) 추정된 계절성분을 다른 주기에도 적용하여 전체 계절성 추정

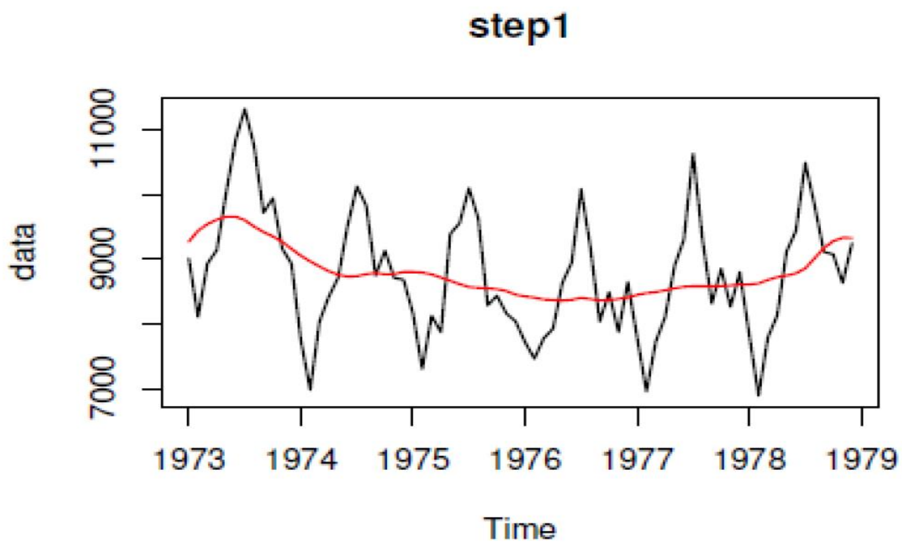
첫 주기의 추세가
반복됨을 확인!



Classical Decomposition Algorithm

추세 및 계절성이 동시에 존재하는 경우

Step1) 이동평균법을 이용하여 추세 예측



단, $\sum_{k=1}^{d} s_j = 0$

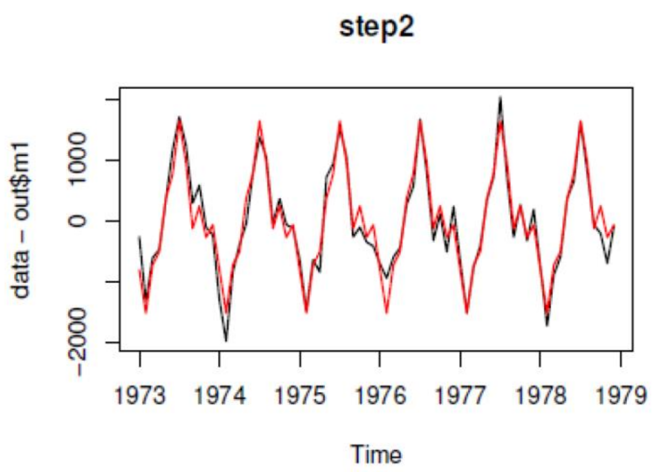
계절성이 존재해도 이동평균법으로 추세를 제거하기 위함

Classical Decomposition Algorithm

추세 및 계절성이 동시에 존재하는 경우

Step2) 추정한 추세 제거 후 계절성 추정

- 추정한 추세 제거 → $z_t = X_t - \widehat{m_t} \approx s_t + Y_t$
- 계절성을 계절평활로 추정 → $\hat{s}_t^* = \frac{1}{m}\Sigma_{k=0}^{m-1} z_{t+kd}, k = 1, \ldots, d$
- 앞의 $\sum_{k=1}^{d} s_j = 0$ 을 만족하기 위한 centering → $\hat{s}_t = \hat{s}_t^* - \widehat{\overline{s^*}}$



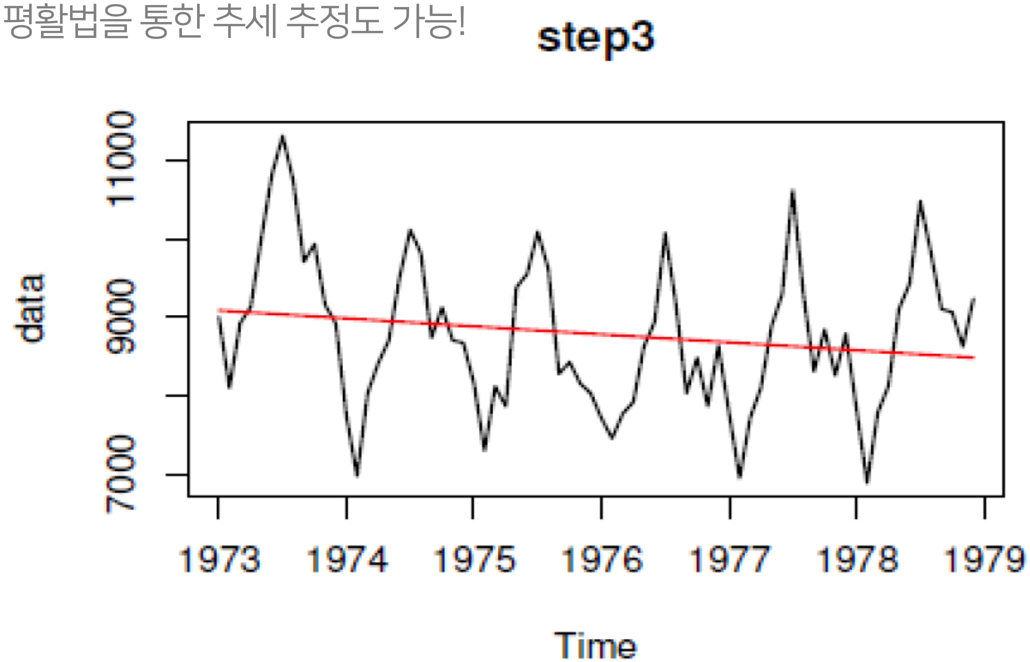
Classical Decomposition Algorithm

추세 및 계절성이 동시에 존재하는 경우

Step3) 계절성이 제거된 시계열에서 OLS로 다시 추세 추정

$$\hat{m}_t^{new} = \underset{c}{\operatorname{argmin}} \Sigma_{t=2}^{n}\left(X_t - \hat{s}_t - c_0 - c_1 t - c_2 t^2 - \cdots - c_p t^p\right)^2$$

OLS 이외에도 평활법을 통한 추세 추정도 가능!

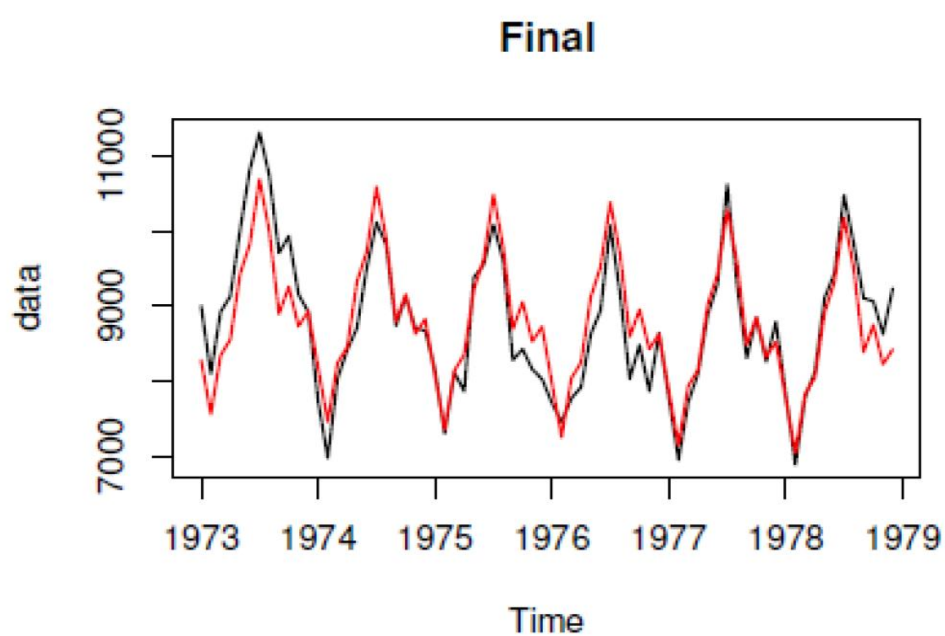


Classical Decomposition Algorithm

추세 및 계절성이 동시에 존재하는 경우

Step4) 새롭게 추정한 추세와 계절성을 제거하여 오차를 추정

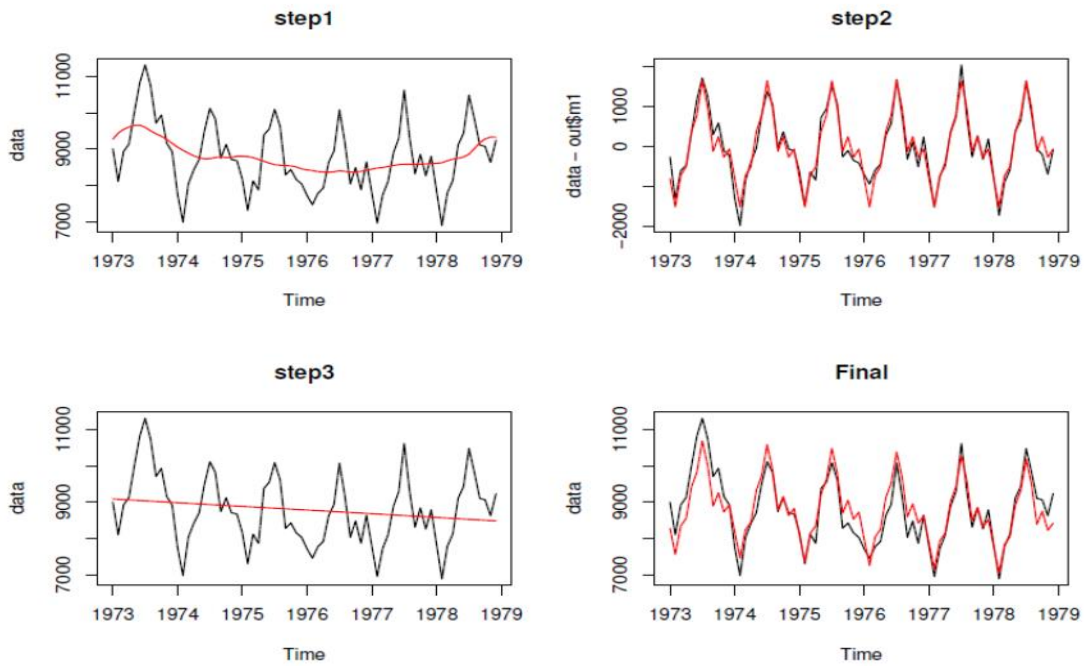
$$\hat{e}_t = X_t - \widehat{m}_t - \hat{s}_t$$



Classical Decomposition Algorithm

추세 및 계절성이 동시에 존재하는 경우

Step5) 추세가 계속 존재한다면 위 과정 반복



추세와 계절성이 모두 제거 될 때까지 추세 추정, 계절성 추정 반복!

후향연산자(Backshift Operator)

후향연산자

관측값을 바로 한 시점 전으로 돌려주는 작용을 하는 연산자

$$BX_t = X_{t-1}$$

후향연산자(Backshift Operator)

후향연산자

관측값을 바로 한 시점 전으로 돌려주는 작용을 하는 연산자

$$BX_t = X_{t-1}$$

후향연산자를 통해 차분이 무엇인지 살펴보자!

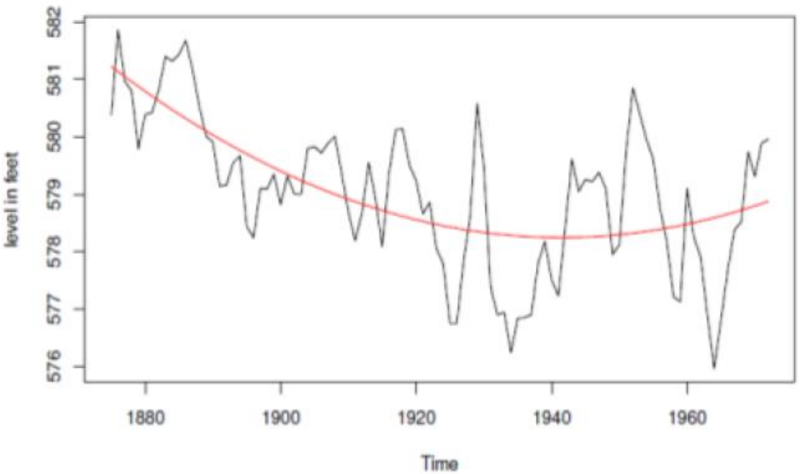
차분(Differencing)

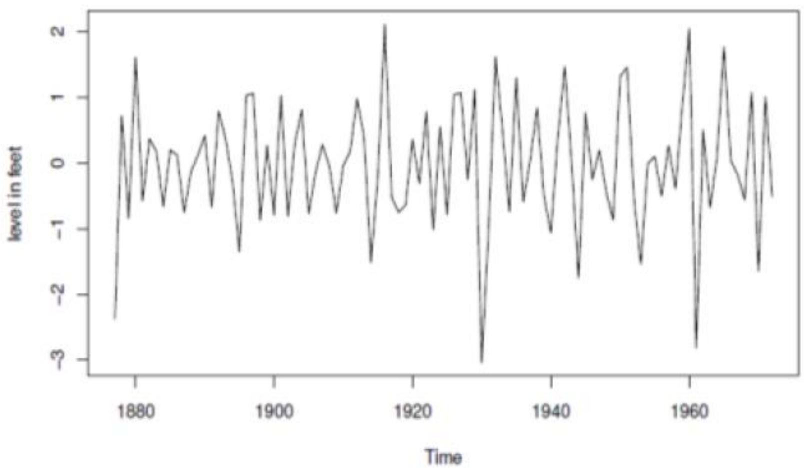
차분

관측값들의 차이를 구하는 것

1차 차분 $\nabla X_t = X_t - X_{t-1} = (1-B)X_t$

2차 차분 $\nabla^2 X_t = \nabla(\nabla X_t) = \nabla(X_t - X_{t-1}) = X_t - 2X_{t-1} + X_{t-2}$





차분(Differencing)

추세만 존재하는 경우

추세를 $m_t = (c_0 + c_1 t)$로 가정 시

$$\nabla m_t = (c_0 + c_1 t) - (c_0 + c_1(\mathrm{t} - 1)) = c_1$$

t 와 상관 없는 상수

차분(Differencing)

추세만 존재하는 경우

추세를 $m_t = (c_0 + c_1 t)$로 가정 시

$$\nabla m_t = (c_0 + c_1 t) - (c_0 + c_1(t-1)) = c_1$$

t 와 상관 없는 상수

일반적으로 k차 차분을 진행하면, k차 추세 제거 가능

1차 차분 $\nabla X_t = X_t - X_{t-1} = (1-B)X_t$

2차 차분 $\nabla^2 X_t = \nabla(\nabla X_t) = \nabla(X_t - X_{t-1}) = X_t - 2X_{t-1} + X_{t-2}$

⋮

차분(Differencing)

계절성만 존재하는 경우

: lag d-differencing을 통해 계절성을 제거

계절성분 $s_t = s_{t+d}$ 라고 가정 시

$$\nabla_d X_t = s_t - s_{t-d} + Y_t - Y_{t-d} = 0 + error$$

오차항만 남음

차분(Differencing)

계절성만 존재하는 경우

: lag-differencing을 통해 계절성

$$\nabla_d X_t \neq \nabla^d X_t$$

시차 d의 계절차분과 d차 차분은 다름에 주의

$$\nabla_d X_t = s_t - s_{t-d} + Y_t - Y_{t-d} = 0 + error$$

오차항만 남음

전자의 경우 주기가 d인 d차 차분을 진행하겠다는 의미

후자의 경우 1차 차분을 d번 진행하겠다는 의미

차분(Differencing)

계절성만 존재하는 경우

: lag-differencing을 통해 계절성

$$\nabla_{\mathrm{d}} \mathrm{X}_{\mathrm{t}} \neq \nabla^{\mathrm{d}} \mathrm{X}_{\mathrm{t}}$$

계절성분 랜덤 가정 시

시차 d의 계절차분과 d차 차분은 다름에 주의

$$\nabla_d X_t = s_t - s_{t-d} + Y_t - Y_{t-d} = 0 + error$$

오차항만 남음

전자의 경우 주기가 d인 d차 차분을 진행하겠다는 의미

후자의 경우 1차 차분을 d번 진행하겠다는 의미

차분(Differencing)

추세 및 계절성이 동시에 존재하는 경우

: 계절차분과 p차 차분을 순차적으로 진행

Step1) 계절차분을 진행

$$\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d}$$

Step2) 남아있는 추세를 제거하기 위해 차분을 진행

$$\nabla_d = (1 - B^d) = (1 - B)(1 + B + B^2 + \cdots + B^{d-1})$$

계절차분은 이미 **1차 차분을 포함**하고 있음에 주의

차분(Differencing)

추세 및 계절성이 동시에 존재하는 경우

: 계절차분과 p차 차분을 순차적으로 진행

Step1) 계절차분을 진행

$$\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d}$$

Step2) 남아있는 추세를 제거하기 위해 차분을 진행

$$\nabla_d = (1 - B^d) = (1 - B)(1 + B + B^2 + \cdots + B^{d-1})$$

계절차분은 이미 **1차 차분을 포함**하고 있음에 주의

4

정상성 검정

자기공분산 함수(ACVF), 자기상관함수(ACF)

정상성을 만족하지 않는다는 것은 시계열의 확률적 성질이 시간 t에 의존한다는 의미이기 때문에 정상성을 충족하는지 확인 필요

이를 확인할 수 있는 지표로 자기공분산함수와 자기상관함수 존재

자기공분산함수 (ACVF) → 표본자기공분산함수 (SACVF)

자기상관함수 (ACF) → 표본자기상관함수 (SACF)

자기공분산함수, 자기상관함수

자기공분산함수 ACVF (Autocovariance Function)

$$\gamma_X(h) = Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$$

표본자기공분산함수 SAVCF (Sample Autocovariance Function)

$$\hat{r}_x(h) = \frac{1}{n}\sum_{j=1}^{n-h}\left(X_j - \bar{X}\right)\left(X_{j+h} - \bar{X}\right)$$

자기공분산함수, 자기상관함수

자기상관함수 ACF(Autocorrelration Function)

$$\rho_x(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = Corr(X_t, X_{t+h}) = \frac{Cov(X_t, X_{t+h})}{\sqrt{var(X_t)}\sqrt{var(X_{t+h})}} = \frac{\gamma(h)}{\gamma(0)}$$

표본자기상관함수 SACF(Sample Autocorrelation Function)

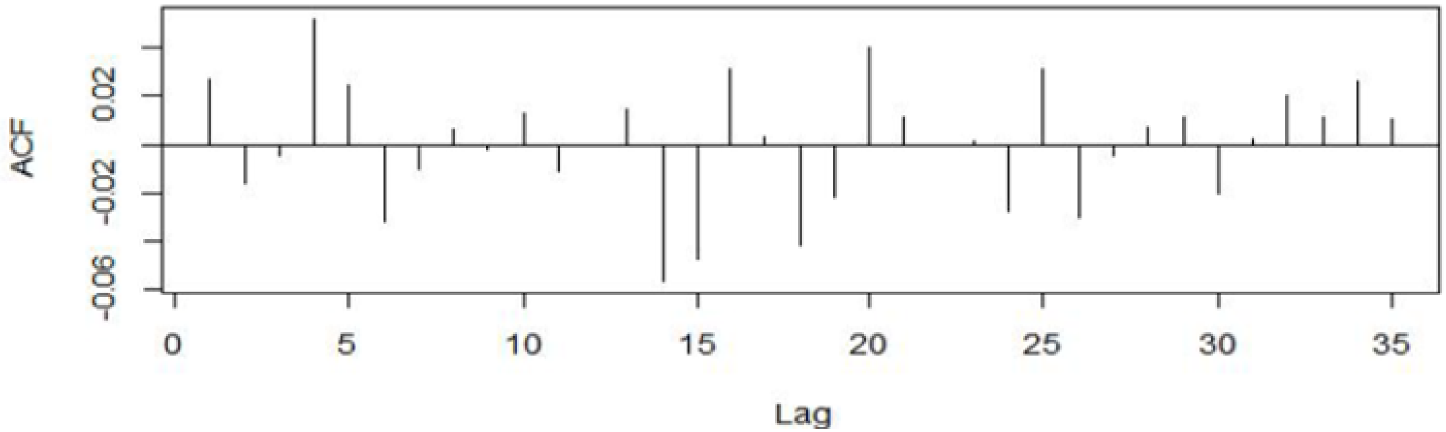
$$\hat{\gamma}_x(h) = \frac{\hat{\gamma}_x(h)}{\hat{\gamma}_x(0)}, \hat{\rho}(0) = 1$$

백색잡음

백색잡음

자기상관이 없는 시계열 → 대표적인 정상 시계열

$$\{X_t\} \sim WN(0, \sigma^2)$$



백색잡음

백색잡음의 조건

1	$E(X_t) = 0$	평균이 0
2	$\mathrm{Var}(X_t) = \sigma^2$	분산이 σ^2
3	$Cov(X_t, X_{t+h}) = 0$	상관관계가 존재하지 않음 (uncorrelated)

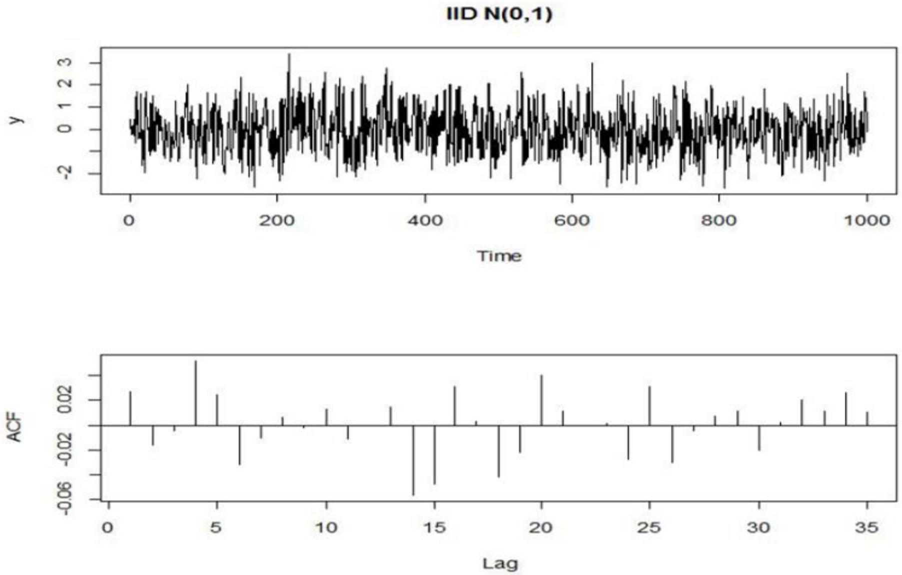
백색잡음

백색잡음과 $IID(0, \sigma^2)$ 의 관계

: $IID(0, \sigma^2)$는 백색잡음이라고 할 수 있지만, 그 역은 참이 아님

$$IID(0, \sigma^2) \rightarrow WN(0, \sigma^2)$$

$$IID(0, \sigma^2) \nleftarrow WN(0, \sigma^2)$$



백색잡음 검정

남은 오차항이 WN이거나 IID인 경우

σ^2는 $\gamma_X(0)$만 추정하면 구할 수 있음

백색잡음 검정

남은 오차항이 WN이거나 IID인 경우

σ^2는 $\gamma_X(0)$만 추정하면 구할 수 있음

⇩

남은 오차항이 WN나 IID가 아닌 경우

여전히 존재하는 시차 간의 의존성을 해결하기 위해 복잡한 모형이 필요

백색잡음 검정

남은 오차항이 WN이거나 IID인 경우

σ^2는 $\gamma_X(0)$만 추정하면 구할 수 있음

남은 오차항이 WN나 IID가 아닌 경우

여전히 존재하는 시차 간의 의존성을 해결하기 위해 복잡한 모형이 필요

1 오차 독립성 검정

2 정규성 검정

3 정상성 검정

백색잡음 검정

오차 독립성 검정

오차가 백색잡음이라면, $\hat{\rho} \approx \mathrm{N}\left(0, \frac{1}{n}\right)$

(n이 증가하면 평균이 0, 분산이 1/n인 정규분포를 따름)

$\boldsymbol{H_0: \ \rho(h) = 0 \ vs \ H_1: \ \rho(h) \neq 0}$ 형태의 가설 검정 실시

만약 $|\hat{\rho}(h)|$가 $1.96\sqrt{n}$ 범위 내 있다면,

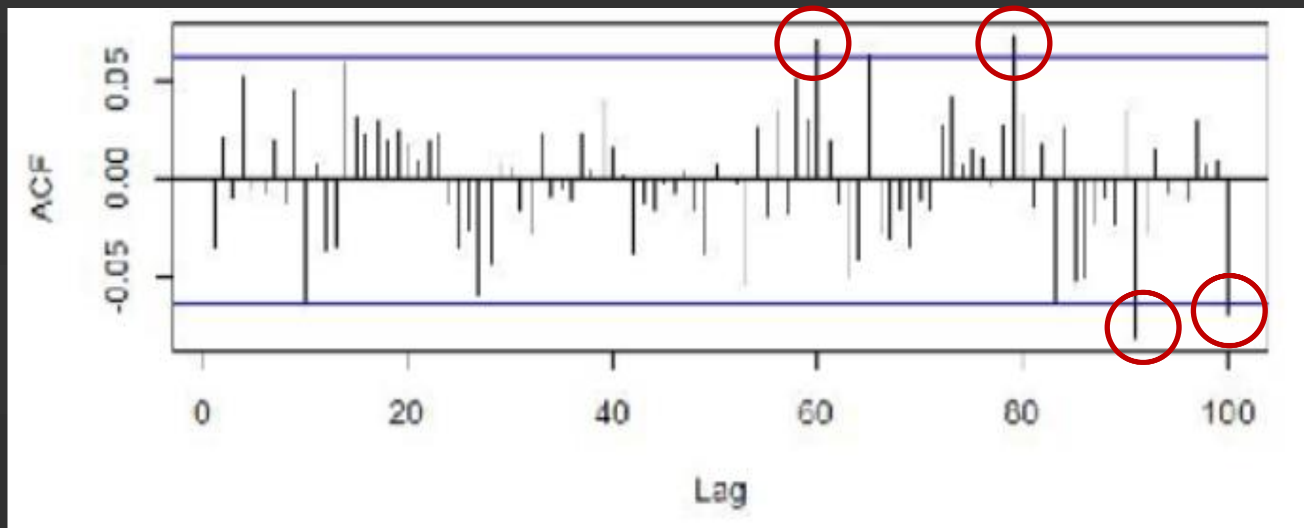
근사적으로 오차들이 uncorrelated 되어있다고 할 수 있음

4 정상성 검정

백색잡음 검정

ACF 그래프를 통한 확인

오차 독립성 검정



(X축 : 시차, Y축 : acf, 파란 실선 : ACF의 신뢰구간)

대부분의 시차에서 신뢰구간 이내의 ACF 값을 가져 자기상관 만족 O

만약 $|\hat{\rho}(h)|$가 $1.96\sqrt{n}$ 범위 내 있다면

그러나 신뢰구간을 벗어나는 시차도 있어 완벽하게 자기상관 만족 X

근사적으로 오차들이 uncorrelated 되어있다고 할 수 있음

백색잡음 검정

정규성 검정

QQ plot

시각적으로 정규성을 만족하는지 확인하는 방법

Kolmogorov-Sminorv test(KS test)

표본의 누적확률분포(CDF)가 모집단(여기서는 정규분포)의 누적확률분포와 얼마나 유사한지를 비교하는 방법

Jarque-Bera test

왜도와 첨도가 정규분포로 보기에 적합한지에 대한 적합도 (Goodness-of-fit) 검정 방법

백색잡음 검정

정상성 검정

Kpss test

단위근 검정방법 → H_0: 정상성을 가진다(정상 시계열이다)

ADF test

단위근 검정방법 → H_1: 정상성을 가진다(정상 시계열이다)

PP test

이분산이 있을 경우에 사용가능한 검정방법

→ H_1: 정상성을 가진다(정상 시계열이다)

다음주 예고 

1. 1주차 복습
2. 모형의 식별
3. 선형과정
4. AR
5. MA
6. ARMA
7. 모형의 적합절차

THANK YOU