

시계열자료분석팀

5팀

김민
이수린
김동환
서유진
장다연

INDEX

1. 시계열자료
2. 정상성
3. 정상화
4. 정상성 검정

1

시계열 자료

시계열 자료

시계열 자료 (time series)

시간에 따라 관측된 자료의 집합

⋮

$$X_t, t = 1, 2, \dots$$

 t 에 따라 이산형과 연속형으로 구분함

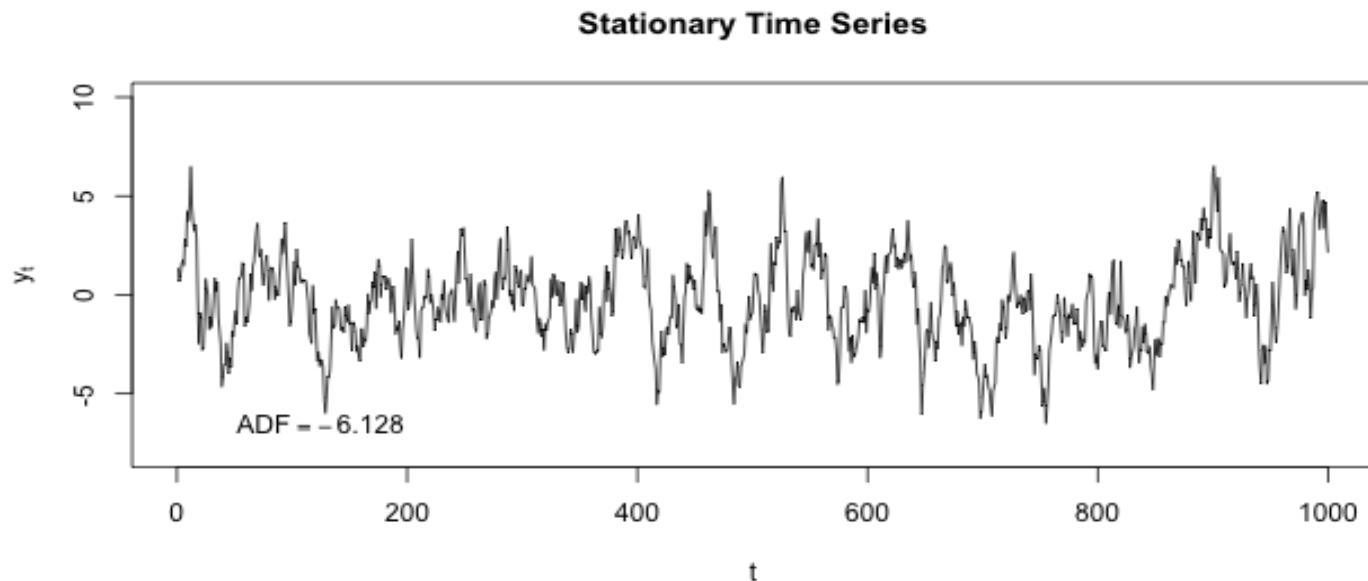
⋮

관측치 사이에 연관성(dependency) 존재



즉, 잔차의 독립성 가정 만족하지 않음

시계열 자료의 구성 요소



추세 변동

순환 변동

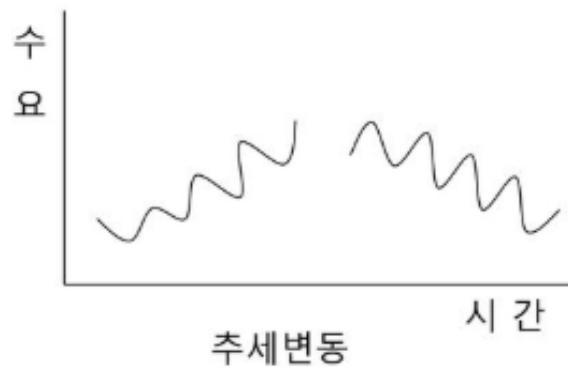
계절 변동

우연 변동
불규칙 성분

시계열 자료의 구성 요소

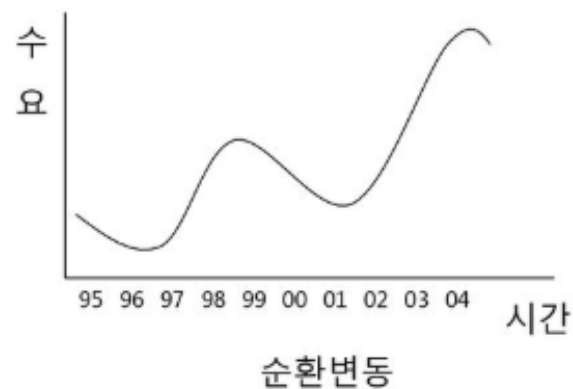
추세 변동

시간이 경과함에 따라
관측치가 증가하거나 감소함



순환 변동

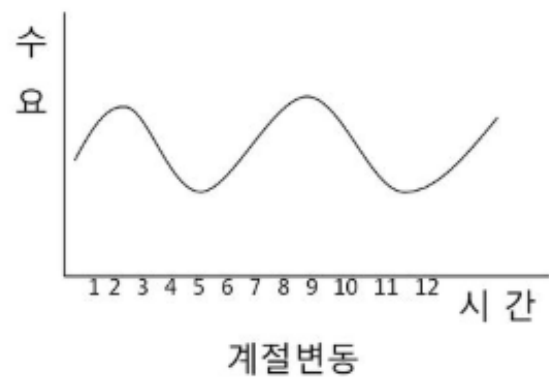
주기적인 변화 ϕ , 주기가 긴 변동
계절에 의한 변화 아님



시계열 자료의 구성 요소

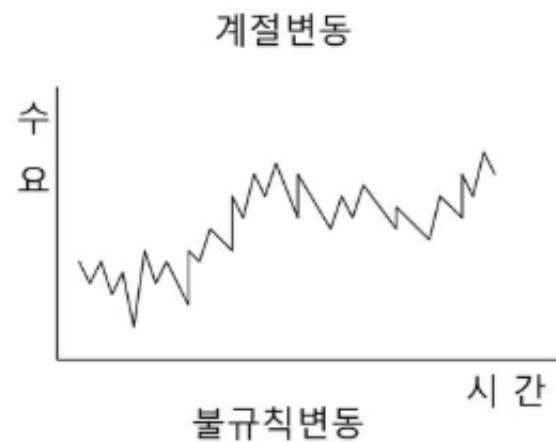
계절 변동

주별·월별·계절별과 같은
주기적인 성분에 의한 변동



우연 변동 / 불규칙 성분

무작위 원인에 의해 나타나는 변동
(시간에 따른 규칙적 움직임과는 무관)



시계열 모형

시계열 모형 (Time Series Model)

시계열 데이터 x_t 가 어떤 확률 구조(=확률 분포)에 따라
생성되는지를 묘사하는 **확률적 모형**

...

미래 정보가 불확실하여
확률 또는 확률분포로 표현되는 모형

특정 시점에 대한 확률 변수 X_t 는
전체 시점에서의 관측치 집합 $\{x_1, x_2, \dots\}$ 을 모두 고려한 결합 분포

2

정상성

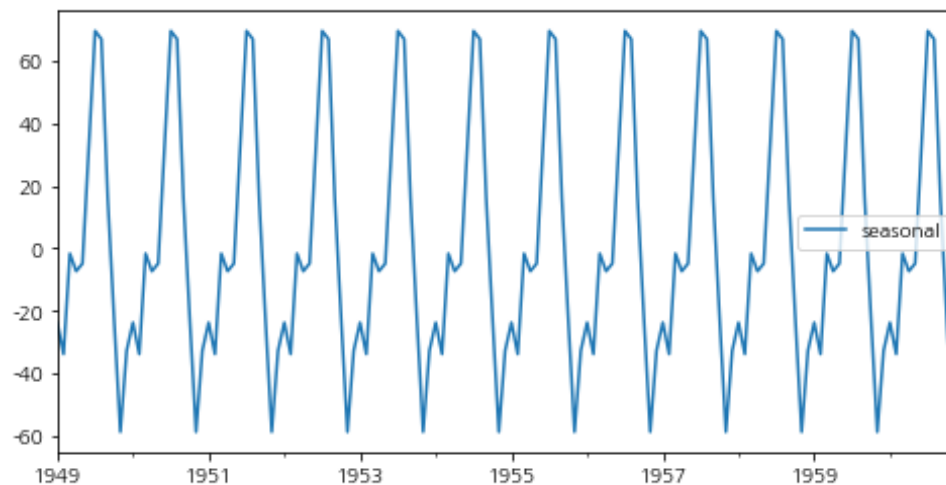
정상성

정상성 (Stationarity)

시계열 자료의 확률적 성질이
시점(t)에 의존하지 않고 **시차(lag)**에만 의존하는 특성



평균과 분산이 시간에 따라 변화하지 않음





정상성

정상성의 필요성

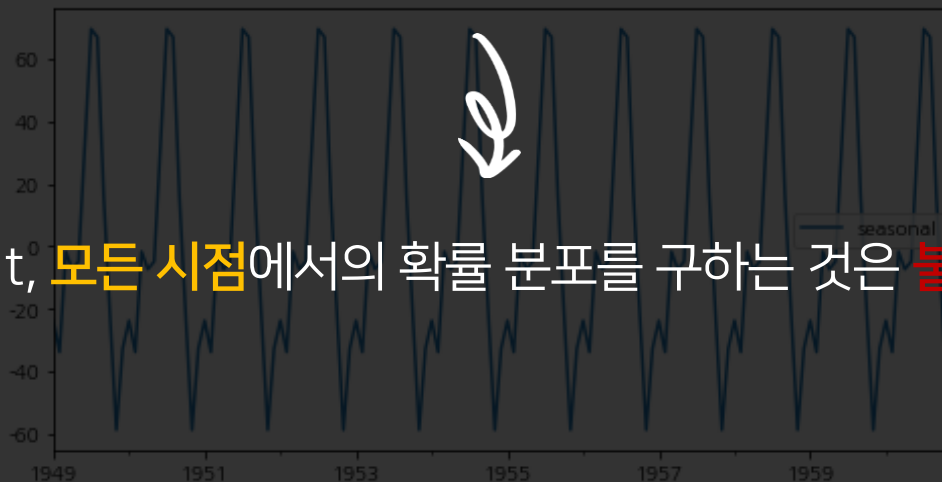
정상성 (Stationarity)

시계열 자료의 확률적 성질이
시점(t)에 의존하지 않고 **시차(lag)**에만 의존하는 특성
시계열 모형은 전체 시점 관측치에 대한 **결합 분포** 확률변수로 구성됨

미래(x_{t+1})를 예측하려면

관측하지 못한 x_{t+1} 를 포함한 결합 분포 필요

평균과 분산이 시간에 따라 변화하지 않음



But, **모든 시점**에서의 확률 분포를 구하는 것은 **불가능**



정상성

정상성의 필요성

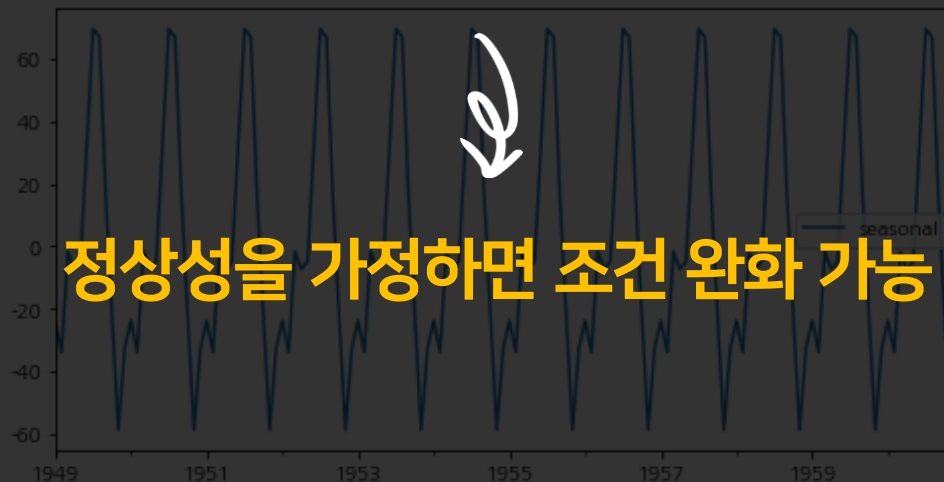
정상성 (Stationarity)

시계열 자료의 확률적 성질이
시점(t)에 의존하지 않고 **시차(lag)**에만 의존하는 특성
시계열 모형은 전체 시점 관측치에 대한 **결합 분포** 확률변수로 구성됨

미래(X_{t+1})를 예측하려면

관측하지 못한 x_{t+1} 를 포함한 결합 분포 필요

평균과 분산이 시간에 따라 변화하지 않음



정상성

정상성 (Stationarity)

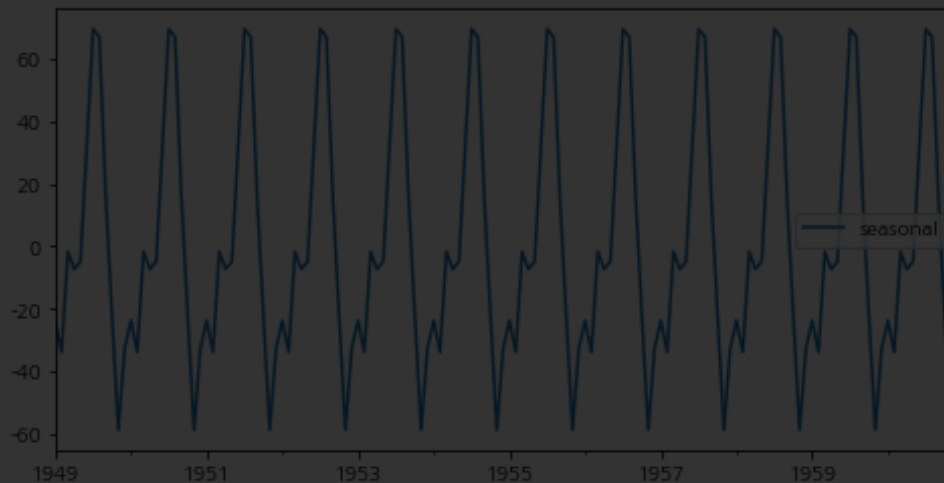
시계열 자료의 확률적 성질이
시점(t)에 의존하지 않고 **시차(lag)**에만 의존하는 특성



확률적 성질이 시점이 아닌 **시차(lag)**에 의존



lag에서의 분포를 구하고 일정 간격마다 반복



강정상성

강정상성 (Strict Stationarity)

$$(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \dots, X_{t_n+h}), \quad h \text{는 시차}$$

시계열 $\{X_t, t \in \mathbb{Z}\}$ 가 위 조건을 만족할 때 강정상성을 만족

분포가 완전히 동일해야 함

⋮

t_1 부터 t_n 까지 n 기간만큼의 시계열에 대한 결합 분포는,

시점을 h 만큼 옮겨도 동일한 기간에 대해서는 같은 결합 분포를 가져야 함

강정상성

강정상성 (Strict Stationarity)

$$(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \dots, X_{t_n+h}), \quad (h \text{는 시차})$$

시계열 $\{X_t, t \in \mathbb{Z}\}$ 가 위 조건을 만족할 때 강정상성을 만족

분포가 완전히 동일해야 함

⋮

t_1 부터 t_n 까지 n 기간만큼의 시계열에 대한 결합 분포는,

시점을 h 만큼 옮겨도 동일한 기간에 대해서는 같은 결합 분포를 가져야 함



강정상성은 지나치게 엄격한 가정이기 때문에

강정상성에 정규성을 가정해 조건 완화

정규성 가정



$$(X_{t_1}, \dots, X_{t_n}) \sim MVN(\mu, \Sigma) \text{ 가정}$$

⋮

평균 벡터 μ 와 공분산 행렬 Σ 만 구하면

분포를 추정할 수 있음

⋮

정규성 가정에 의해

- ① 각 확률변수의 **기댓값은 상수**
- ② 각 확률변수의 **공분산은 시차에만 의존**

∴ 공분산 행렬은 대칭 행렬

약정상성

강정상성 (Strict Stationarity)

$$(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \dots, X_{t_n+h}), \quad (h \text{는 시차})$$

시계열 $\{X_t, t \in \mathbb{Z}\}$ 가 위 조건을 만족할 때 강정상성을 만족

분포가 완전히 동일해야 함

약정상성 (Weakly Stationarity)

정규성 가정으로부터의 조건들을 확장하고 완화한 것



강정상성과 달리 분포의 **동일성 가정 X**

앞으로 다룰 모든 정상 시계열은 약정상성을 만족

약정상성의 3가지 조건



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



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



$$\gamma_X(r, s) = \gamma_X(r + h, s + h), \quad \forall r, s, h \in \mathbb{Z}$$
$$(\gamma_X(r, s) := \text{Cov}(X_r, X_s))$$

약정상성의 3가지 조건



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



2차 적률(분산)이 존재하고,
시점에 관계없이 일정



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



$$\gamma_X(r, s) = \gamma_X(r + h, s + h), \quad \forall r, s, h \in \mathbb{Z}$$

$(\gamma_X(r, s) := \text{Cov}(X_r, X_s))$

약정상성의 3가지 조건



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



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



평균은 상수이며 시점에 관계없이 **일정**



$$\gamma_X(r, s) = \gamma_X(r + h, s + h), \quad \forall r, s, h \in \mathbb{Z}$$

$(\gamma_X(r, s) := \text{Cov}(X_r, X_s))$

약정상성의 3가지 조건



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



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



자기공분산은 **시차 h에만 의존**, 시점 t와 무관

$$\gamma_X(r, s) = \gamma_X(r + h, s + h), \quad \forall r, s, h \in \mathbb{Z}$$

$$(\gamma_X(r, s) := \text{Cov}(X_r, X_s))$$

3

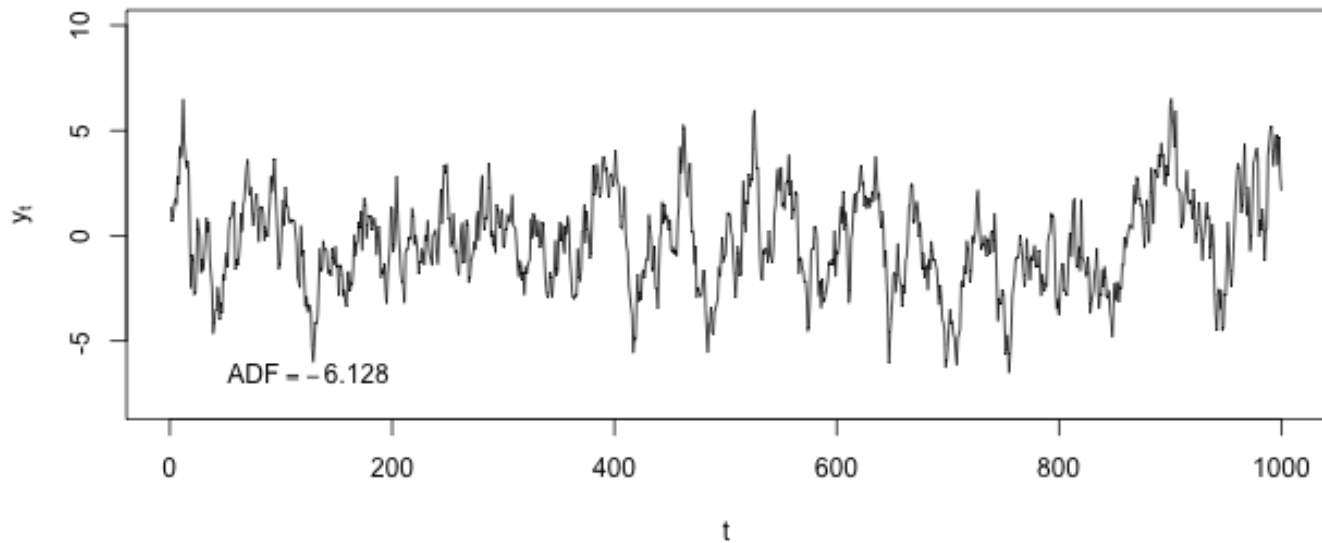
정상화

정상 시계열

정상 시계열

정상성을 만족하는 시계열
추세와 계절성 없이 평균과 분산 일정

Stationary Time Series

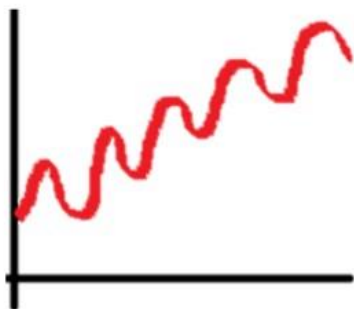


비정상 시계열

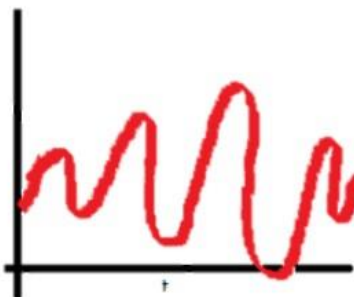
비정상 시계열

약정상성 조건을 만족하지 못하는 시계열

[평균이 일정하지 않은 경우]



[분산이 일정하지 않은 경우]



[공분산이 시점에 의존하는 경우]

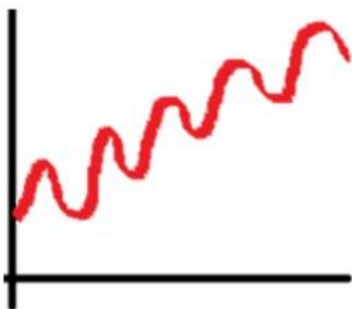


비정상 시계열

비정상 시계열

약정상성 조건을 만족하지 못하는 시계열

[평균이 일정하지 않은 경우]



[분산이 일정하지 않은 경우]



[공분산이 시점에 의존하는 경우]



정상화를 통해 정상 시계열로의 변환 필요

정상화 과정


👉 분산이 일정하지 않은 경우 변환을 통해 분산 안정화

👉 비정상 시계열 X_t 를 추세 m_t , 계절성 s_t ,
stationary error Y_t 로 분해

👉 추세 또는 계절성을 추정하여 제거

정상화 과정

 분산이 일정하지 않은 경우 변환을 통해 분산 안정화

 비정상 시계열
이분산성을 띄는 경우, VST를 통해
분산이 시간의 흐름에 의존하지 않도록 변환

stationary error ϵ_t 로 분해

 추세 또는 계절성을 추정하여 제거

VST : 분산 안정화 변환 (Variance Stabilizing Transformation)

정상화 과정

분산이 일정하지 않은 경우

1) 로그 변환

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

2) 제곱근 변환

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

3) Box-Cox 변환

$$\rightarrow f_{\lambda}(X_t) = \begin{cases} \frac{X_t^{\lambda} - 1}{\lambda}, & X_t \geq 0, \lambda > 0 \\ \log X_t, & \lambda = 0 \end{cases}$$

정상화 과정

평균이 일정하지 않은 경우

비정상 시계열 (X_t)를 비정상 부분과 정상 부분으로 분해

$$X_t = m_t + s_t + Y_t$$

m_t : 추세

s_t : 계절성

Y_t : 정상성을 만족하는 오차

} 비정상 부분

--- 정상 부분

정상화 과정

평균이 일정하지 않은 경우

비정상 시계열 (X_t)를 비정상 부분과 정상 부분으로 분해

$$X_t = m_t + s_t + Y_t$$

m_t : 추세

s_t : 계절성



비정상 부분

Y_t : 정상성을 만족하는 오차

정상 부분



① 추세만 존재

② 계절성만 존재

③ 추세, 계절성 모두 존재

정상화 과정

평균이 일정하지 않은 경우

비정상 시계열 (X_t)를 비정상 부분과 정상 부분으로 분해

$$X_t = m_t + s_t + Y_t$$



비정상 부분 (추세 & 계절성) 추정 및 제거!

1. 회귀 (*Regression*)
2. 평활 (*Smoothing*)
3. 차분 (*Differencing*)

평균이 일정하지 않은 경우 : 회귀

1. 추세만 존재하는 경우 : Polynomial Regression

Step 0) 기본 가정

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

m_t : 추세, Y_t : 정상성을 만족하는 오차

평균이 일정하지 않은 경우 : 회귀

1. 추세만 존재하는 경우 : Polynomial Regression

Step 0) 기본 가정

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

m_t : 추세, Y_t : 정상성을 만족하는 오차



Step 1) 추세 성분 m_t 를 시간 t 에 대한 선형회귀식으로 나타냄

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

평균이 일정하지 않은 경우 : 회귀

1. 추세만 존재하는 경우 : Polynomial Regression

Step 2) 선형회귀식의 계수를 최소제곱법(OLS)을 통해 추정

$$(\hat{c}_0, \dots, \hat{c}_p) = \underset{c}{\operatorname{argmin}} \sum_{t=1}^n (X_t - m_t)^2$$

Step 3) 추정한 추세를 비정상 시계열 (X_t)에서 제거

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

평균이 일정하지 않은 경우 : 회귀

2. 계절성만 존재하는 경우 : Harmonic Regression

Step 0) 기본 가정 : “주기가 d인 계절성만을 가진다”

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

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



Step 1) 계절 성분 s_t 를 시간 t에 대한 회귀식으로 나타냄

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

평균이 일정하지 않은 경우 : 회귀

2. 계절성만 존재하는 경우 : Harmonic Regression

Step 2) 적절한 λ_j 와 k 선택 후, OLS를 통하여 a_j 와 b_j 추정

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

k는 주로 1~4 사이 값 사용



람다(λ_j) 란?

$[-\pi, \pi]$

주기가 2π 인 함수의 주기와 데이터 주기를 맞춰 주기 위한 값

1) 주기 반복 횟수 $f_1 = [n/d]$ (n = 데이터 개수, d = 주기) $\rightarrow f_j = j f_1$

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

평균이 일정하지 않은 경우 : 회귀

2. 계절성만 존재하는 경우 : Harmonic Regression

Step 2) 적절한 λ_j 와 k 선택 후, OLS를 통하여 a_j 와 b_j 추정

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

k는 주로 1~4 사이 값 사용



람다(λ_j) 란?

$[-\pi, \pi]$

주기가 2π 인 함수의 주기와 데이터 주기를 맞춰 주기 위한 값

① 주기 반복 횟수 $f_1 = [n/d]$ (n = 데이터 개수, d = 주기) $\rightarrow f_j = j f_1$

② $\lambda_j = f_j (2\pi/n)$

평균이 일정하지 않은 경우 : 회귀

2. 계절성만 존재하는 경우 : Harmonic Regression

Step 3) 추정한 계절성을 시계열에서 제거

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

⋮

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

평균이 일정하지 않은 경우 : 회귀

3. 추세와 계절성이 모두 존재하는 경우

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

Step 1) 추세만 존재하는 경우의 정상화 과정과

계절성만 존재하는 경우의 정상화 과정을 차례대로 진행

Step 2) 이후에도 추세가 남아 있다면 다시 추세 제거

평균이 일정하지 않은 경우 : 회귀



3. 추세와 계절성이 모두 존재하는 경우

회귀 방법의 단점

추정에 사용하는 **최소제곱법(OLS)**은 오차항의 독립성 가정

VS

시계열 자료는 오차항의 독립성 가정 X

Step 1) 추세만 존재하는 경우의 정상화 과정과

계절성만 존재하는 경우의 정상화 과정을 차례대로 진행



Step 2) 이후에도 추세가 남아 있다면 다시 추세 제거

추정이 정확하지 않을 수 있음

평균이 일정하지 않은 경우 : 평활

회귀 방식

OLS의 경우 전체 데이터를
한 번에 사용하여 추정



국소적 변동 표현에 부적절

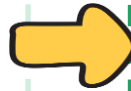
평균이 일정하지 않은 경우 : 평활

회귀 방식

OLS의 경우 전체 데이터를
한 번에 사용하여 추정



국소적 변동 표현에 부적절



평활법

평활법 사용하여 정상성 확보

이동평균평활법

지수평활법

계절성평활법

Classical Decomposition Algorithm

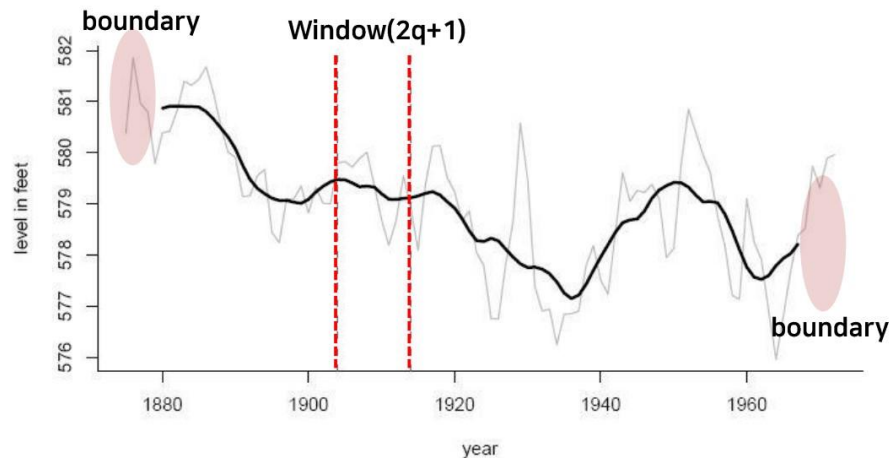
평균이 일정하지 않은 경우 : 평활

1. 추세만 존재하는 경우

이동평균평활법 (Moving Average Smoothing)

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

일정 기간($2q+1$)마다 **평균**을 계산하여 추세를 추정하는 방법



평균이 일정하지 않은 경우 : 평활

1. 추세만 존재하는 경우 : 이동평균평활법 (Moving Average Smoothing)

Step 1) 길이가 $2q+1$ 인 구간의 평균을 구한다.

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

q 는 cross validation을 통해서 찾는다...

평균이 일정하지 않은 경우 : 평활

1. 추세만 존재하는 경우 : 이동평균평활법 (Moving Average Smoothing)

Step 2) 기본 가정 반영



기본 가정

1. 약대수의 법칙

$$E(Y_t) = 0$$

2. 추세는 선형임을 가정

$$m_t = c_0 + c_1 t$$



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

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

약대수의 법칙에 의해 = 0

$W_t = (Y_t, \text{ 즉 Stationary error를 제거한 순수한 선형의 추세만을 보존한 추정량})$

Step 3) 추정한 추세를 비정상 시계열 (X_t)에서 제거

평균이 일정하지 않은 경우 : 평활

1. 추세만 존재하는 경우 : 이동평균평활법 (Moving Average Smoothing)

Step 2) 기본 가정 반영



기본 가정

1. 약대수의 법칙

$$E(Y_t) = 0$$

2. 추세는 선형임을 가정

$$m_t = c_0 + c_1 t$$

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

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

약대수의 법칙에 의해 = 0

$W_t = (Y_t, \text{ 즉 Stationary error를 제거한 순수한 선형의 추세만을 보존한 추정량})$

Step 3) 추정한 추세를 비정상 시계열 (X_t)에서 제거

평균이 일정하지 않은 경우 : 평활

1. 추세만 존재하는 경우 : 이동평균평활법 (Moving Average Smoothing)

Step 2) 기본 가정 반영



기본 가정

1. 약대수의 법칙

$$E(Y_t) = 0$$

2. 추세는 선형임을 가정

$$m_t = c_0 + c_1 t$$

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

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

약대수의 법칙에 의해 = 0

$W_t = (Y_t, \text{ 즉 Stationary error를 제거한 순수한 선형의 추세만을 보존한 추정량})$

Step 3) 추정한 추세를 비정상 시계열 (X_t)에서 제거



평균이 일정하지 않은 경우 : 평활

한계

1. 추세만 존재하는 경우 : 이동평균평활법 (Moving Average Smoothing)

Step 2) 기본 가정 반영

1. Boundary Problem



기본 가정 : 처음 q 개, 마지막 q 개 시점에서는 추정량 구할 수 없음

약대수의 법칙에 의해 = 0

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

1. 약대수의 법칙

$$E(Y_t) = 0$$

2. 추세는 선형임을 가정

$$m_t = c_0 + c_1 t$$

2. T 시점에 대한 예측을 위해 t 시점 이후의 데이터 활용

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



$W_t = (Y_t, \text{ 즉 Stationary error를 제거한 순수한 선형의 추세를 보존한 추정량})$

과거 데이터에만 의존하여 추정하는 **지수평활법 필요!**

Step 3) 추정한 추세를 비정상 시계열 (X_t)에서 제거

평균이 일정하지 않은 경우 : 평활

1. 추세만 존재하는 경우

지수평활법 (Exponential Smoothing)

추세 \hat{m}_t 를 t시점까지의 관찰값만을 이용하여 추정하는 방법
 확률변수(X_t)와 이전 시점의 추세 추정량(\hat{m}_{t-1})의 **가중평균**

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

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

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

⋮

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

추정한 추세를 비정상 시계열 (X_t)에서 제거

평균이 일정하지 않은 경우 : 평활

1. 추세만 존재하는 경우

지수평활법 (Exponential Smoothing)

추세 \hat{m}_t 를 t시점까지의 관찰값만을 이용하여 추정하는 방법
 확률변수(X_t)와 이전 시점의 추세 추정량(\hat{m}_{t-1})의 **가중평균**

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

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

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

⋮

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

추정한 추세를 비정상 시계열 (X_t)에서 제거

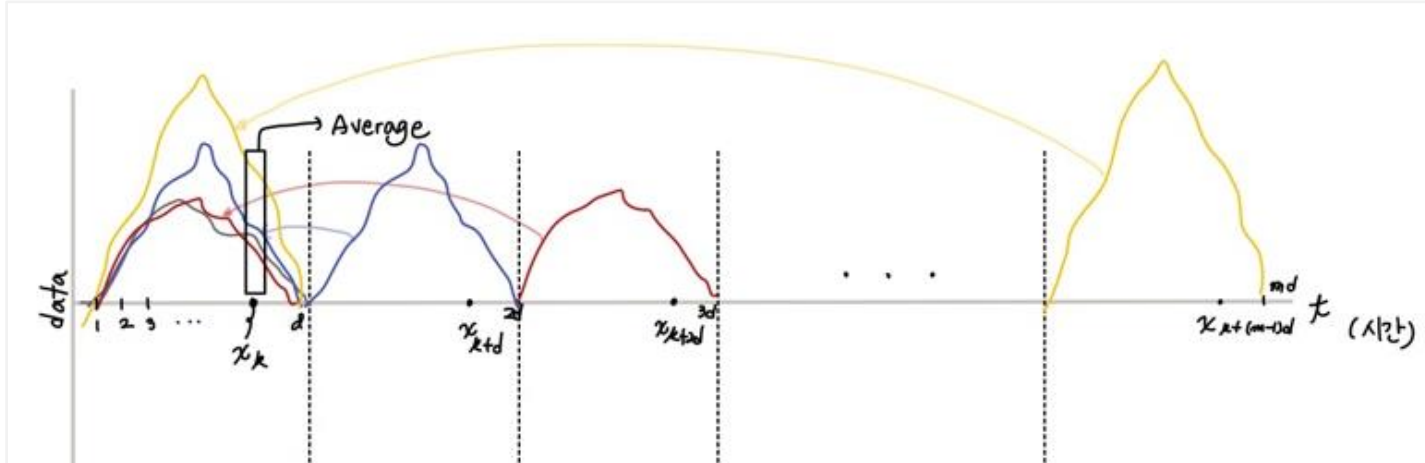
평균이 일정하지 않은 경우 : 평활

2. 계절성만 존재하는 경우

계절성평활법 (Seasonal Smoothing)

동일 기간 d 에 대한 주기의 관측치들을 모두 겹친 후,

겹쳐진 값들을 각 시점마다 평균 내어 계절성 추정



k 시점의 계절성 추정량은 d 간격 떨어진 데이터들의 평균

평균이 일정하지 않은 경우 : 평활

2. 계절성만 존재하는 경우 : 계절성평활법 (Seasonal Smoothing)

Step 1) 계절성 추정량 추정

: d간격 떨어진 데이터들의 평균

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

$$\hat{s}_k = \hat{s}_{k-d}, \quad \text{if } k > d$$

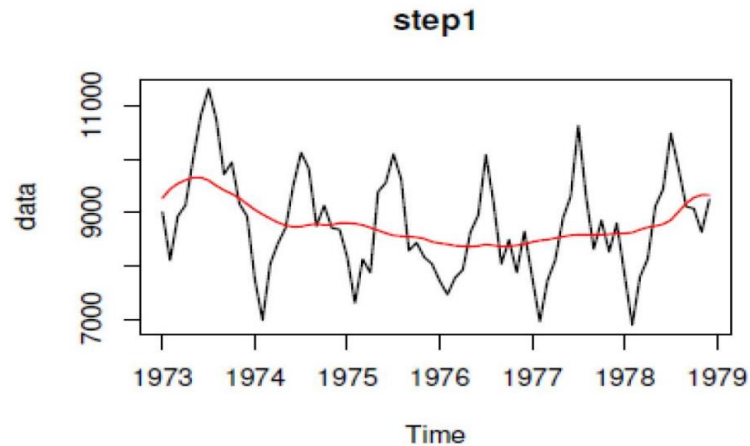
Step 2) 추정량을 다른 주기에도 적용

Step 3) 추정량을 시계열에서 제거

평균이 일정하지 않은 경우 : 평활

3. 추세와 계절성 모두 존재하는 경우 : Classical Decomposition Algorithm

Step 1) 이동평균법을 통해 추세 추정 및 제거



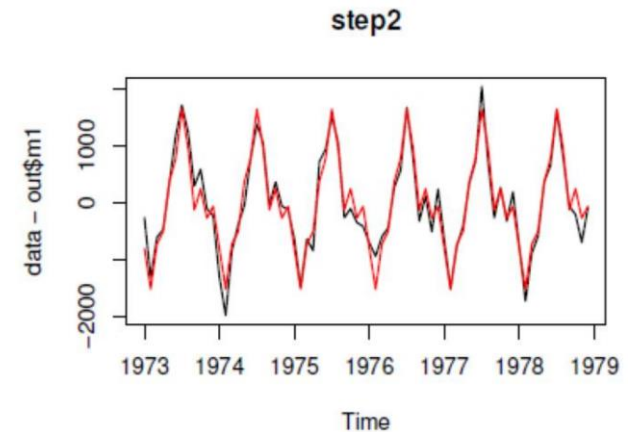
$$\text{if } d = 2q, \quad \hat{m}_t = \frac{0.5X_{t-q} + X_{t-q+1} + \cdots + X_{t+q-1} + 0.5X_{t+q}}{2q}$$

$$\text{if } d = 2q + 1, \quad \hat{m}_t = \frac{X_{t-q} + X_{t-q+1} + \cdots + X_{t+q-1} + X_{t+q}}{2q + 1}$$

평균이 일정하지 않은 경우 : 평활

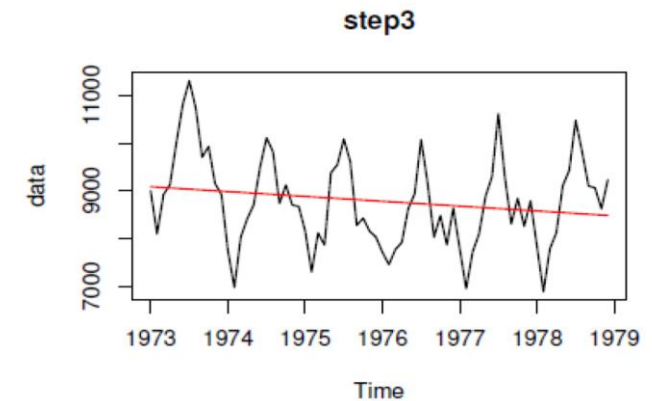
3. 추세와 계절성 모두 존재하는 경우 : Classical Decomposition Algorithm

Step 2) 계절성평활법으로 계절성 추정 및 제거 →



Step 3) OLS를 활용하여 다시 추세 추정 및 제거 →

OLS 이외에도 평활법을 통한 추세 추정도 가능



평균이 일정하지 않은 경우 : 평활

3. 추세와 계절성 모두 존재하는 경우 : Classical Decomposition Algorithm

Step 2) 계절성평활법으로 계절성 추세를 제거

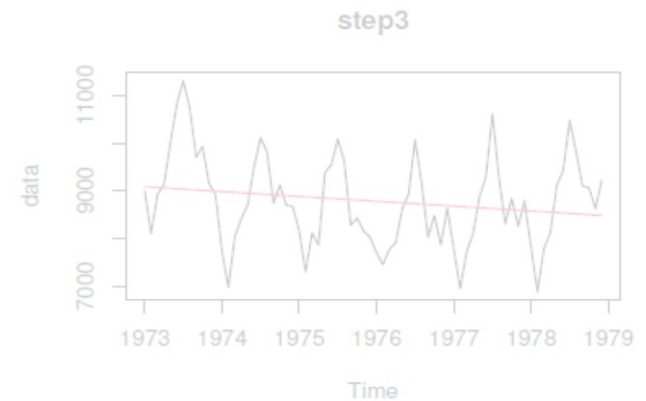
추세가 계속 존재한다면

추세와 계절성이 모두 제거될 때까지 반복!



Step 3) OLS를 활용하여 다시 추세 추정 및 제거 →

OLS 이외에도 평활법을 통한 추세 추정도 가능



평균이 일정하지 않은 경우 : 차분

차분 (Differencing)

관측값들의 차이를 구하는 것

과거 시점을 간단하게 표현 가능!

후향 연산자

관측값을 바로 한 시점 전으로 돌려주는 작용을 하는 연산자


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

평균이 일정하지 않은 경우 : 차분

차분 (Differencing)

관측값들의 차이를 구하는 것

⋮

1차 차분

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

2차 차분

$$\begin{aligned}\nabla^2 X_t &= \nabla(\nabla X_t) = \nabla(X_t - X_{t-1}) \\ &= X_t - 2X_{t-1} + X_{t-2} = (1 - B)^2 X_t\end{aligned}$$

평균이 일정하지 않은 경우 : 차분

1. 추세만 존재하는 경우 : 차분 (Differencing)

추세를 $m_t = (c_0 + c_1 t)$ 와 같이 선형으로 가정 후 **1차 차분** 적용

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

t의 영향을 받지 않는 상수
→ 정상 시계열



k차 차분을 적용하면 k차 추세 제거 가능!

$$\nabla^k X_t = k! c_k + \nabla^k Y_t = \text{const.} + \text{error}$$

정상성을 만족하는 오차

평균이 일정하지 않은 경우 : 차분

2. 계절성만 존재하는 경우 : lag-d 차분 (lag-d Differencing)

Lag-d 차분 (lag-d Differencing)

$s_t = s_{t+d}$ 를 가정하고,

계절성만 있는 비정상 시계열 $X_t = s_t + Y_t$ 에 lag-d 차분 적용



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

오차항만 남는 정상 시계열!



평균이 일정하지 않은 경우 : 차분

2. 계절성만 존재하는 경우 : lag-d 차분 (lag-d Differencing)

Lag-d 차분 (lag-d Differencing)

$s_t = s_{t+d}$ 를 가정하고,

계절성만 있는 비정상 시계열 $X_t = s_t + Y_t$ 에 lag-d 차분 적용



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

오차항만 남는 정상 시계열!





평균이 일정하지 않은 경우 : 차분

2. 계절성만 존재하는 경우 : lag-d 차분 ($\nabla_d \neq \nabla^d$)

Lag-d 차분 (lag-d Differencing)
d차 차분과 lag-d 차분은 다름에 주의

$s_t = s_{t+d}$ 를 가정하고,

계절성만 있는 비정상 시계열 $X_t = s_t + Y_t$ 에 lag-d 차분 적용



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

오차항만 남는 정상 시계열!



평균이 일정하지 않은 경우 : 차분

2. 계절성만 존재하는 경우 : lag-d 차분 ($\nabla_d \neq \nabla^d$)

Lag-d 차분 (Lag-d Differencing)
d차 차분과 lag-d 차분은 다름에 주의

$s_t = s_{t+d}$ 를 가정하고,

계절성만 있는 비정상 시계열 $s_t = s_t + Y_t$ 에 lag-d 차분 적용

d차 차분

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

lag-d차분

$$\nabla_d = (1 - B^d)$$

주기가 d인 d차 차분을 진행 $s_{t-d} + Y_t - Y_{t-d}$ = 1차 차분을 d번 진행

오차항만 남는 정상 시계열!

평균이 일정하지 않은 경우 : 차분

3. 추세와 계절성 모두 존재하는 경우 : 계절차분(lag-d 차분) + p차 차분

Step 1) 계절 차분 진행 → 계절성 제거

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

Step 2) p차 차분 진행 → 추세 제거

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

계절 차분 자체에 1차 차분이 포함되어 있기 때문에
p차의 추세를 제거하고자 할 때 p-1차 차분 적용!



4

정상성 검정

자기공분산함수(ACVF), 자기상관함수(ACF)

비정상 부분이 제대로 제거되었다면 시계열데이터에는
정상성을 만족하는 오차 y_t 만이 남아있어야 함

①평균과 분산이 일정

②시계열의 확률적 성질(공분산 함수)이 시간 t 에 의존하지 않고 시차 h 에만 의존

자기공분산함수(ACVF), 자기상관함수(ACF)

비정상 부분이 제대로 제거되었다면 시계열데이터에는
정상성을 만족하는 오차 y_t 만이 남아있어야 함

①평균과 분산이 일정

②시계열의 확률적 성질(공분산 함수)이 시간 t 에 의존하지 않고 시차 h 에만 의존



자기공분산함수와 자기상관함수로 확인
(ACVF) (ACF)

자기공분산함수(ACVF), 자기상관함수(ACF)

자기공분산함수 ACVF (auto-covariance function)

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

자기상관함수 ACF (auto-correlation function)

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

자기공분산함수(ACVF), 자기상관함수(ACF)

자기공분산함수 ACVF (auto-covariance function)

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



Sample을 이용한

표본자기공분산함수와 표본자기상관함수

정의할 수 있음

자기상관함수 ACF (auto-correlation function)

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

자기공분산함수(ACVF), 자기상관함수(ACF)

자세한 내용은 2주차 클린업에서...

자기공분산함수 ACVF (auto-covariance function)

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

표본자기공분산함수 SACVF (sample auto-covariance function)

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

자기상관함수 ACF (auto-correlation function)

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

표본자기상관함수 SACF (sample auto-correlation function)


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

백색잡음

White Noise Process (백색잡음)

평균이 0이고 분산이 σ^2 인 확률변수의 집합이며,

확률 변수 간에 상관계수가 없어야 함


$$\text{Cov}(X_t, X_s) = 0, \text{ if } t \neq s$$



정상화 과정을 거친 시계열이 백색잡음 조건을 만족한다면,
해당시계열은 정상 시계열임

백색잡음과 IID

또 다른 정상시계열인 IID process

—> 각 확률변수가 독립성을 만족하면서도 동일한 분포를 따름

백색잡음과 IID의 관계는?

—> IID이면 백색잡음이지만, 그 역은 성립하지 않음

—> 백색잡음은 확률변수가 상관관계가 존재하지 않음, 반드시 독립일 필요는 없음



IID와 비교하여 독립성 조건이 완화된 것이 백색잡음

백색잡음 검정

정상화 과정을 거친 뒤, 남은 오차항이 WN 또는
 IID 인 경우 해당 시계열은 정상시계열

정상성을 만족한다면
추가적인 모델링 과정없이,
분산 $\sigma^2 = r(0)$ 만
추정하면 됨

정상성을 만족하지 않는다면,
 $non-stationary\ error$ 에
대한 추가적인 모델링이 필요

백색잡음 검정

오차항이 WN sequence의 조건을 만족하는지 검정



- ① 자기 상관의 유무 확인
- ② 정규성 만족하는지 확인
- ③ 정상성 만족하는지 확인

백색잡음 검정

→ 자기상관 유무 확인

표본자기상관함수(SACF)

오차가 백색잡음을 따른다면, $\hat{\rho} \approx N(0, \frac{1}{n})$

⋮

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

⋮



만약 $\rho(h)$ 가 $1.96\sqrt{n}$ 범위 내 있다면, 귀무가설을 기각하지 못하므로 오차항에 자기상관이 없다고 할 수 있음

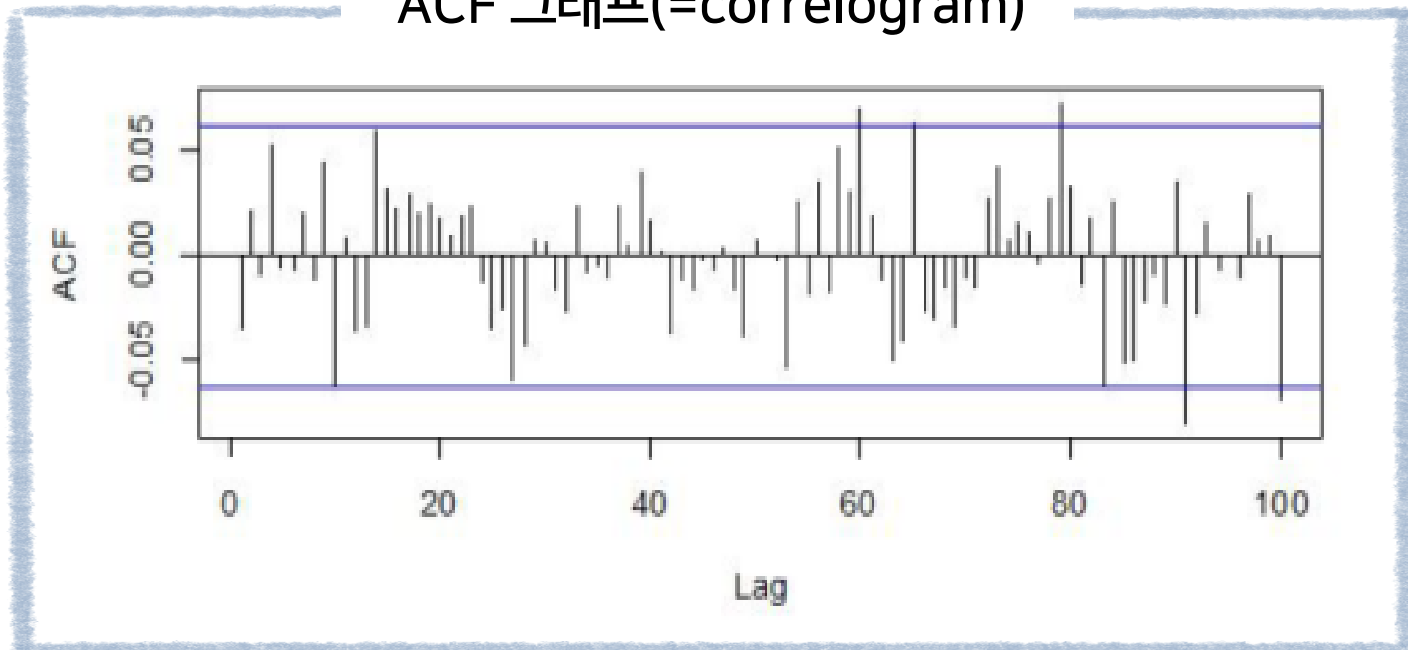
h 는 최소 $n/4$ 보다 크게 설정하는 것을 권장

시간간격이 너무 넓으면 SACF에 대한 신뢰성이 떨어짐

백색잡음 검정

—> 자기상관 유무 확인

ACF 그래프(=correlogram)



대부분의 값이 신뢰구간인 파란선 내에 분포함으로
대체적으로 정상성을 만족함

백색잡음 검정

—> 정규성을 만족하는지 확인

QQ plot

표본과 정규분포의 quantile을 시각적으로 비교

KS plot

표본과 모집단의 cdf가 얼마나 유사한지 비교

Jarque-Bera test

왜도와 첨도의 정규분포로서의 적합도를 검정

백색잡음 검정



→ 정규성을 만족하는지 확인

위의 검정 결과 오차항이 **정규성**을 만족하지 않는다면...?

QQ plot

표본과 정규분포의 quantile을 시각적으로 비교
Box-Cox 변환과 같은 변환 기법을 활용하여

정규분포에 가깝게 변형할 수 있고

KS plot

t분포와 같이 다른 분포를 가정하여 **MLE**를 구하는 방법도 가능함

표본과 모집단의 cdf가 얼마나 유사한지 비교

Jarque-Bera test

왜도와 첨도의 정규분포로서의 적합도를 검정

백색잡음 검정

—> 정상성을 만족하는지 확인

Kpss test

귀무가설로 정상 시계열임을 가정

ADF test

대립가설로 정상 시계열임을 가정

PP test

이분산이 있을 때도 사용가능한 검정방법

대립가설로 정상 시계열임을 가정

다음 주 예고

1. 선형과정
2. AR 모형
3. MA 모형
4. ARMA 모형

감사합니다!

