시계열자료분석팀

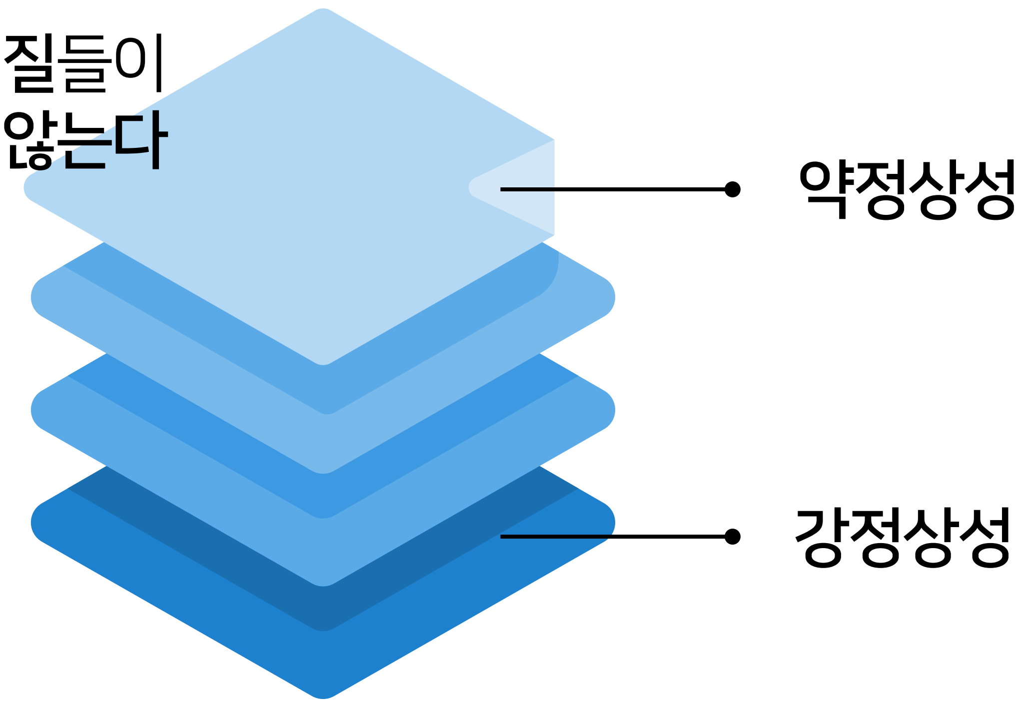
5팀

김지원
권남택
최수진
최순일
홍서영

INDEX

"정상성"

: 시계열의 **확률적인 성질**들이
시간에 따라 **변하지 않는다**



강정상성

$$f(Z_{t1}, Z_{t2}, \cdots, Z_{tn}) = f(Z_{t1+k}, Z_{t2+k}, \cdots, Z_{tn+k})$$

= 시간 축을 k만큼 이동해도 모든n에 대해 결합확률밀도함수가 동일!

= 모든n에 대하여 결합확률밀도함수가 시간대를 바꾸어도 동일하다

약정상성

ⅰ. $V(X_t) = \sigma^2 , \forall_t \in Z$

→ 분산이 유한한 상수로 시점t에 관계없이 일정

ⅱ. $E(X_t) = m, \forall_t \in Z$

→ 평균이 유한한 상수로 시점t에 관계없이 일정

ⅲ. $\gamma(\mathrm{h}) = Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$

→ 공분산이 시차h에만 의존 (시점t에 의존 X)

비정상 시계열 관측

↓

분산 안정화

: 추세와 계절성을 제거하기 전에,
시계열 데이터의 변동을 완화

최적의 λ는
Likelihood Ratio Test를 통해 구함

Box-Cox

$$f(x:\lambda) = \frac{(x^{\lambda}-1)}{\lambda},\ (\lambda \neq 0)$$

$$f(x:0) = log(x),\ (\lambda = 0)$$

로그 변환

$$f(x) = \log(x)$$

제곱근 변환

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

Step 1. 추세와 오차만 있는 시계열 모형 가정

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

Step 2. 추세 성분 m_t를 t에 대한 다항식 근사 고려

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

Step 3. 다항식을 최소제곱법을 통해 계수 추정

$$(\hat{c_0}, \hat{c_1}, \cdots, \hat{c_p}) = argmin \sum (X_t - m_t)^2$$

Step 4. 추정한 추세 $\widehat{m_t}$ 를 X_t 에서 제거!!

- 추세 제거 - 이동평균평활법

1 산정 기준 중심 有 :

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

2 산정 기준 중심 無 :

$$W_t = [X_t + X_{t-1} + \cdots + X_{t-N+1}] / N$$

- 추세 제거 - 이동평균평활법

(2q+1)개

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

: 시점 t에 대해 t-q 시점의 관측치부터 t+q시점의 관측치들의 평균 구하여 값을 대체

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

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

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

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

∴ 추세 성분 m_t만 남고, 이를 제거할 수 있다!

- 추세 제거 - 지수평활법

시점 1일 때의 추세 추정값 $\widehat{m_1} = X_1$

시점 2일 때의 추세 추정값 $\widehat{m_2} = a * X_2 + (1-a) * \widehat{m_1}$

⋮

시점 t일 때의 추세 추정값 $\widehat{m_t} = a * X_t + (1-a) * \widehat{m_{t-1}}$

$$= a * (X_t + (1-a) * X_{t-1} + \cdots + (1-a)^{n-1} * X_1)$$

$(0 < a < 1)$

∴ 과거 시점일수록 더 작은 가중치가 부여된다!

4 정상성 검정

① 자기공분산함수(ACVF)

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

= 시점 t 의 실현값 (X_t) 과 시점 $t+h$ 의 실현값(X_{t+h})사이의 공분산

BUT 1. 각 시점에 대하여 한 개의 실현값...

BUT 2. 평균 μ 역시 모수...

∴ **표본자기공분산함수**(Sample ACVF)를 사용하기로 하자!

SACVF = $\hat{\gamma}(h) = \frac{1}{n}\sum_{j=1}^{n-h}(x_j - \bar{x})(x_{j+\mathrm{h}} - \bar{x})$

= 시차가 h일 때의 ACVF 추정치

4 정상성 검정

② 자기상관함수(ACF)

$$\rho(h) = corr(X_t, X_{t+h}) = \frac{\gamma(h)}{\gamma(0)}$$

= 시점 t 의 실현값 (X_t) 과 시점 $t+h$ 의 실현값(X_{t+h})사이의 상관계수

BUT 1. 각 시점에 대하여 한 개의 실현값…

BUT 2. 평균 μ 역시 모수…

∴ **표본자기상관함수**(Sample ACF)를 사용하기로 하자!

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

= 시차가 h일 때의 ACF 추정치

" 백색잡음 "

: 대표적인 **정상**시계열의 예시로, 서로 독립 & 동일한 분포를 따르는(iid) 확률변수들로 구성된 확률과정

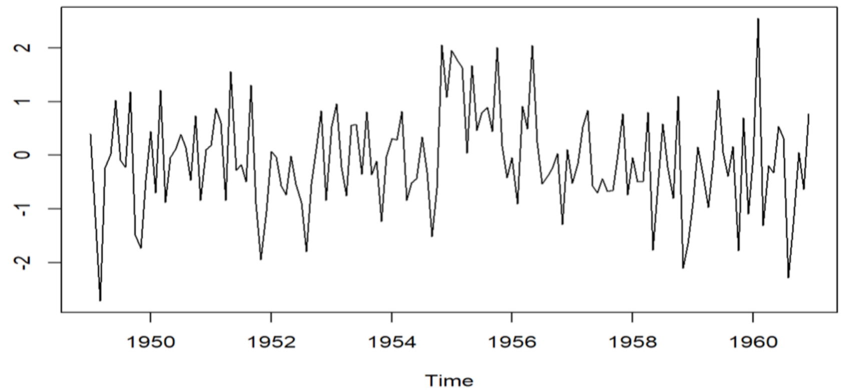
$$X_t = \varepsilon_t,\ \varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$$

<ε_t의 조건 >

① $E(\varepsilon_t) = 0$

② $\gamma_0 = Var(\varepsilon_t) = \sigma_\varepsilon^2$

③ $\gamma_k = Cov(\varepsilon_t, \varepsilon_{t+k}) = \mathbf{0}$



백색잡음 검정 : $X_t \sim WN(0,1)$ 이고, n이 충분히 클 경우 $\hat{\boldsymbol{\rho}}(\mathbf{h}) \approx \boldsymbol{N}(\mathbf{0}, \frac{\mathbf{1}}{\boldsymbol{n}})$

→ 이 특징을 이용해 백색잡음인지 여부 확인!

① 시차 = h인 잔차끼리 상관관계가 X!
② 정규성을 가져야함! ($\hat{\rho}(\mathrm{h})$가 정규분포를 따라야 함!)

$\hat{\rho}(\mathrm{h})$가Uncorrelated?	추세가 존재?	정규성을 가지는지?
• 포트맨토 검정 (Portmanteau test) • Ljung-Box test • McLeod and Li test • Different Sign test	• Turning Point test • Rank test	• QQ plot (시각적) • Jarque – Bera test • Kolmogorov-Smirnov test

"정상성 검정"

: 정상화 과정을 진행한 이후 남은 오차항 Y_t가 정상성을 따르는지 검정하는 과정

$$Xt = \underbrace{m_t + s_t}_{} + Y_t$$

정상화 과정에서 추세(m_t)와 계절성(s_t) 제거!

검정방법! kpss test / adf test / pp test