클린업 1주차

5팀 시계열자료분석팀

염예빈
한유진
이재현
박세령
이정우

시계열 자료 시계열 분석

시계열 자료란?

[시계열 자료의 종류]

Discrete time series

이산 시계열 : 특정 시점에 측정된 관측값

→ 일반적으로 관측값 사이의 간격이 일정

Continuous time series

연속 시계열 : 시간의 모든 점에서 측정된 관측값

→ 이산 시계열에 비해 자료가 적은 편

회귀분석 vs 시계열분석

Regression Time Series analysis

[회귀분석] [시계열분석]

자기 상관성

- 설명변수를 통한 예측
- 과거 자료를 토대로 예측

어떤 변수에 대해 이전의 값이 이후의 값에 영향을 주는 것

즉, 이전 시점의 값이 이후 시점의 값에 영향을 주는 것

- 독립성을 전제

→ 순서 시경 X

- 자기상관성을

→ 순서가 중요

Earlier Data → Later Data

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$y_t = \delta + \phi_1 y_{1-1} + \varepsilon_t$$

정상성 강정상성 약정상성

강정상성

$$F\left(X_t, \ldots, X_{t+p}\right) = F\left(X_{t+k}, \ldots, X_{t+k+p}\right)$$

동일한 기간의 시계열에 대한 결합확률분포가 모든 시계열 구간에서 동일

→ **지나치게 엄격한 가정** :
이를 만족하는 시계열 자료는 거의 없다

따라서, 강정상성을 만족하기는 어려움!!

약정상성

1. $E(y_t) = \mu < \infty$

: **평균**이 **시점 t에 무관하게** **일정**

2. $Var(y_t) = E(y_t - \mu)^2 = \gamma_0 < \infty$

: **분산**이 **시점 t에 무관하게** **일정**

3. $\mathrm{Cov}(y_t, y_{t+k}) = E(y_t - \mu)(y_{t+k} - \mu) = \gamma_k < \infty \quad \forall k$

: 자기공분산(공분산이 시점이 아닌 **시차에만 의존**)

→ **t에 무관**하고 **k에 의존**

비정상시계열 정상화 방법

추세/계절성을 가질 경우

시계열의 유형과 방법에 따라 과정이 달라진다.

방법	시계열 유형
회귀	추세만 있는 시계열
평활	계절성만 있는 시계열
차분	추세, 계절성 있는 시계열

비정상시계열 정상화 방법

회귀 – 추세만 있는 시계열

STEP 1. 추세성분과 불규칙성분만 있는 시계열 모형을 가정

$$Y_t = \boxed{T_t} + \boxed{I_t}, E(I_t) = 0$$

T_t: 추세성분, I_t : 불규칙성분

STEP 2. 추세성분 T_t 를 시점에 대한 회귀식으로 표현

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

- 추세의 비선형 관계를 고려하기 위해 p차 다항식으로 표현
- P의 값은 플랏을 보고 분석자가 결정

평활 – 추세만 있는 시계열

1. 이동평균 평활법

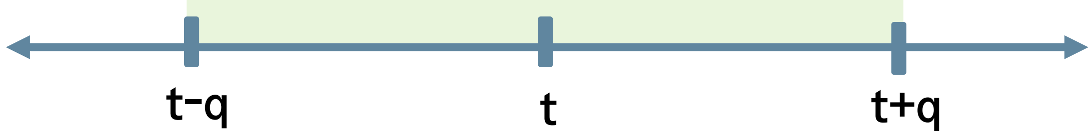
추세성분 T_{t+j}

$$W_t = \frac{1}{2q+1}\sum_{j=-q}^{q}\left(\boxed{T_{t+j}} + I_{t+j}\right)$$

$$\frac{1}{2q+1}\sum_{j=-q}^{q} T_{t+j} = c_0 + c_1 t = T_t\ ,$$

$$t \in [q+1, n-q]$$

시점 t-q부터 t+q까지의 평균 = 시점 t에서의 추세!



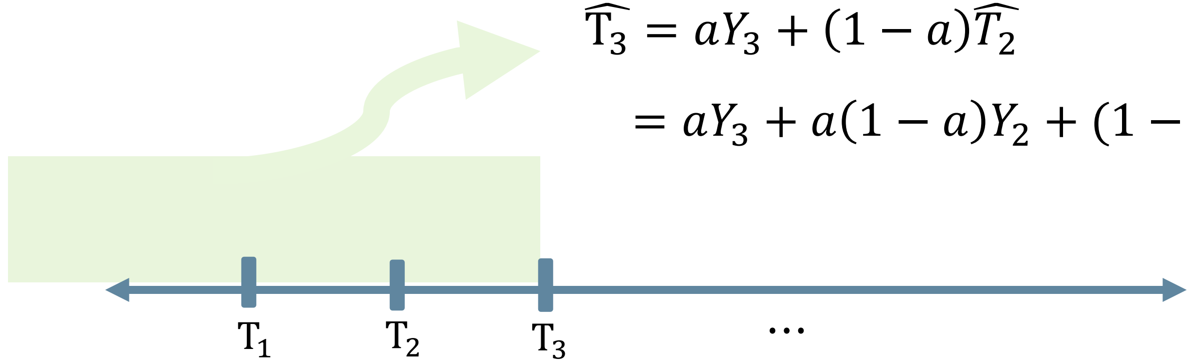
평활 – 추세만 있는 시계열

2. 지수 평활법

✓ 시점별 가중평균으로 추세를 파악하는 방법

$$\widehat{T_3} = aY_3 + (1-a)\widehat{T_2}$$

$$= aY_3 + a(1-a)Y_2 + (1-a)^2 Y_1$$



비정상시계열 정상화 방법

평활 – 추세만 있는 시계열

2. 지수 평활법

시점1 추세 : $\widehat{\mathrm{T}_1} = Y_1$

시점2 추세 : $\widehat{\mathrm{T}_2} = aY_2 + (1-a)\widehat{T_1} = aY_2 + (1-a)Y_1$

시점3 추세 : $\widehat{\mathrm{T}_3} = aY_3 + (1-a)\widehat{T_2} = aY_3 + a(1-a)Y_2 + (1-a)^2Y_1$

…

시점t 추세 : $\widehat{\mathrm{T}_t} = aY_t + (1-a)\widehat{T_{t-1}} = \sum_{j=0}^{t-2}(a(1-a)^jY_{t-j}) + (1-a)^{t-1}Y_1$

- 평활계수 a가 클수록 시계열 변화에 따른 예측값의 변화가 크다.
- 최근자료에 더 큰 가중치를 부여함으로써 이동평균평활법의 단점 보완

비정상시계열 정상화 방법

평활 - 추세와 계절성 둘 다 있는 시계열

<Classical decomposition algorithm>

Step 1. 이동평균평활법을 이용해 추세 추정 (이때, $\sum_{j=1}^{d} S_j = 0$)

Step 2. 관측값에서 추정한 추세를 빼서 계절성과 불규칙성분만 남긴다.

$$Y_t - \widehat{T_t} \approx S_t + I_t$$

Step 3. Seasonal Smoothing을 통해 계절성분 S_t 를 추정

Step 4. 관측값에서 추정한 계절성을 빼서 추세와 불규칙성분만 남긴다.

Step 5. (Step 4)식에서 추세성분 T_t를 회귀를 통해 추정

마무리

(Step 3)에서 추정한 계절성분 S_t와 (Step 5)에서 추정한 추세성분 T_t를 관측치에서 제거한다.

차분의 정의

현재 시계열에서 과거 시계열을 뺀 것!

1차 차분

$$\nabla X_t = X_t - X_{t-1}$$

2차 차분

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

후향연산자(Backshift Operator)

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

앞으로 배울 식들을 간단하게 표현할 수 있다!!

차분 – 추세와 계절성 모두 있는 시계열

STEP 1. 추세와 계절성이 있는 모형을 가정

$$Y_t = T_t + S_t + I_t = (c_0 + c_1 t) + S_t + I_t$$

STEP 2. 1차 차분과 주기가 d인 계절 차분 동시에 적용

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

$$= (1-B)(c_0 + c_1 t + S_t + I_t - c_0 - c_1(t-d) - S_{t-d} - I_{t-d})$$

$$= (1-B)(c_1 d + S_t - S_{t-d} + I_t - I_{t-d})$$

$$= I_t - I_{t-1} - I_{t-d} - I_{t-d-1}$$

자기상관관계 유무 검정

귀무가설: $\mathrm{H}_0:\ \rho_1 = \rho_2 = \cdots = \rho_k = 0$

✓ **ACF그래프를 통한 검정 :**

ACF, PACF그래프가 각 시점에서 신뢰구간안에 존재하고 있을 경우

⇒ 잔차 간의 상관관계가 없다고 판단

✓ **포트맨토 검정(Portmanteau Test, Ljung-Box 검정) :**

$\epsilon_t \sim N(0, \sigma^2)$ 가정하에서 검정통계량 $\mathrm{Q}' = \mathrm{T}(\mathrm{T}+2)\sum_{k=1}^{h}\frac{r_k^2}{T-k} \sim \chi(h-K)$ 를 사용하여 검정

정상성 검정

각 방법의 귀무가설이
모두 다르므로 주의!!

✓ kpss test :

단위근(Unit-root) 검정 방법 중 하나.

귀무가설: '시계열이 정상(stationary)시계열이다.'

✓ ADF(Augmented Dickey-Fuller)Test :

단위근 검정방법 중 하나이며, DF검정을 일반화 한 것

귀무가설: '자료에 단위근이 존재한다.' VS 대립가설: '자료가 정상성을 만족한다.'

✓ PP(Phillips-Perron)Test :

이분산의 경우에도 사용 가능.

귀무가설은 '데이터가 비정상이다.' VS 대립가설: '자료가 정상성을 만족한다.'