**Moving average and Smoothing** – 시계열을 표현하는 세가지 기본적인 모델

1. **Naïve** : 최근의 관찰
   1. Simple : future value = past value(Y^t+1 = Yt)

1-2 with Trend(추세를 가지는 경우) : future value = past value + ΔYt

future value = past value\*(Yt/Yt-1)

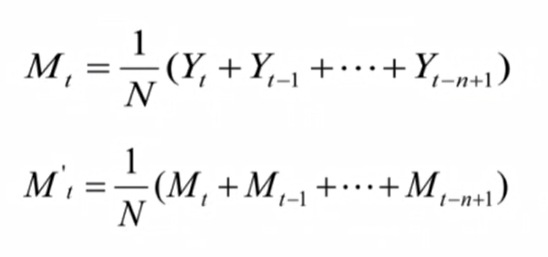
1-3 seasonality(계절성을 띄는 경우) : Yt+1=Yt-3

1. **Moving average** : 이동 평균

2-1 Simple average methods : 전체의 평균

2-2 **Moving average methods** : 최근 n개의 평균

* 1. **Double moving average methods**: 이동평균의 이동평균



1. **Smoothing**

3-1 **Exponential Smoothing** : Y^t+1 = aYt + (1-a)Y^t : estimate past data와 actual data를 이용하는 방법 / a는 사용자가 직접 설정해주는 값

3-2 **Double exponential smoothing** : exponential smoothing의 exponential smoothing

Moving average는 estimate value를 쓰지않고 exact data를 씀

이동평균 큰 숫자를 쓸수록 장기적인 추세를 나타냄

주식 차트에서 5일선이 20일선 아래에서 위로 지나가면 단기적인 상승 추세를 나타냄

60일선이 120일선 아래에서 위로 지나가면 장기적인 상승 추세를 나타냄

**MSE(mean squared error),MAPE(mean average percentage error) : model의 forecast power를 알기 위한 방법(예측값과 실제값이 얼마나 비슷한지 알 수 있음)**

**Mse : 오차의 절대값의 평균**

**Mape : 실제값에 대한 오차의 비율**

Mean squared error에서 제곱을 하거나 절대값을 하는 이유 : error들이 서로 상쇄되지 않게 하려고

**Autoregressive : 자기회귀 : 현재 값은 과거 p개 값의 영향을 받아 실현됨 ->AR(p)**

**MA(이동평균법) : 과거 q개의 값에 의해 형성된 현재값을 error-term으로 구성 ->MA(q)**

**ARMA(자기회귀 이동평균 모형) : AR + MA ->ARMA(p,q)**

**Lags : 시차(p,q)**

**시차를 결정하는 법**

1번 : autocorrelation - MA결정

2번 : partial autocorrelation - AR결정

3번 : correlogram(상관도) - x축에 lags를, y축에 autocorrelation coeffieicent를 그림

If Xt 가 MA(m)를 따르고 , Yt 가 MA(n)를 따르면,

Then Zt = Xt+Yt 는 MA(r)를 따른다. r<= max(m,n)

If Xt 가 AR(p)를 따르고 , Yt 가 AR(q)를 따르면,

Then Zt = Xt+Yt 는 AR(r)를 따른다. r<= p+q

**Backward Operator(B) = Lag Operator(L) = 과거값을 의미함**

Yt-1=BYt = LYt

**I 를 통해 stationary인지 아닌지 판별 가능**

I(0):stationary / I(1) : non-stationary

**Feature of stationary time series : desirable**

1. mean reverting : 시간별 데이터가 평균값을 기준으로 평균 위, 평균 아래가 번갈아가며 나타남
2. data with short memory – 예를 들어 20년 전의 데이터가 현재값에 영향을 미치는 것은 원하지 않기 때문
3. robust long-term forecasting (강건)

비정상시계열에서의 해결 방법: 차분(differencing)

**정보기준(information criteria)**

AIC : 정보기준을 사용하면 best model를 찾을 수 있음

정보기준 : based on RSS / RSS가 작으면 정보기준이 작고 그게 더 좋은모델

ARMA(0,0) ~ ARMA(12,12)의 169가지 경우에서 AIC를 구해 가장 낮은 값이 가장 좋은 모델

iid:identically independently distributed : ideal residuals

**small sample AIC** : n/k가 40 미만일때 사용

aic의 장점 : 다른 오류 분포에도 모델을 비교할 수 있음

aic의 주의사항 : 같은 자료로 모델을 다르게 한 것은 AIC로 비교 가능하지만 다른 자료일때는 안됨

**Unit root: : 단위근 / cointegration : 공적분**

**Individual two times series are cointegrated -> X and Y have 장기관계(균형)**

Stationary + stationary combination : stationary

Non-stationary + stationary : non-stationary

Non-stationary + Nonstationary : non-stationary일수도 있고 stationary일 수도 있음 : cointegrated

Ordinary를 한 팀은 diagnosis를 해야하고 stationary를 봐야함(unit root test)

**Spurious regression : 가성회귀 – all regressors are non-stationary and not cointegrated(관계가 있는것 같지만 아님)**

**Unit root test**

**ADF, pp test** : 귀무가설이 정상시계열인지 아닌지 검사하는 방법(귀무가설이 nonstationary일 때 쓰는 방법) – 단위근 검정

**Kpss** assume null hypothesis as a time series is stationary – 시계열 안정성 검사

단위근 검정과 시계열 안정성 검사가 상호보완적인 이유 : 귀무가설을 기각하지 못한다는 말이 대립가설을 받아들인다는 말이 아님 -> 비정상시계열이 틀렸다고 말할 수 있다. 그러나 정상시계열이다. 라고는 말할 수 없음