이번에는 AR model과 MA model이 결합된 형태인 ARMA model과 ARIMA model에 대해 다뤄보려고 합니다. ARMA는 단순히 AR + MA로 유추가 되겠지만, ARIMA의 경우는 I가 있습니다. 이 I는 Integreated를 의미하는데요, 계량경제에서 Integration은 조금 다른 의미를 가지고 있습니다. ARMA, ARIMA model을 하기 앞서 짧게 Integrated의 개념을 짚고 넘어가겠습니다.

이전 글에서 DSP(Difference Stationary Process)에 대해서 알아봤습니다. Non-stationary인 데이터를 차분하여 Stationary하게 된다면 그 data를 일컬어 DSP라고 부릅니다. 단순히 차분만이 아니라 Time trend를 제거해 주어야 statioonary가 되는 TSP(Trend Stationary Process)도 있습니다만, 대부분의 경우 Difference(차분)를 통해 Stationary한 데이터를 만들게 됩니다.

Integrated time series는 바로 이 difference(차분)과 관계된 개념입니다. '몇 번 차분해야 stationary한 time-series가 되는가'가 핵심입니다. 이를 줄여서 'I(n)' 이렇게 표현을 하기도 합니다. 만약 한 번 difference(차분)을 해서 stationary하게 되는 데이터라면 I(1)이라고 표현을 합니다. 한 번 difference(차분)한 데이터를 다시 한 번 difference해야 stationary하게 된다면 I(2)라고 부릅니다. 대부분의 데이터는 1번만 차분해도 stationary하게 되는 것이 일반적이기 때문에 integrated of order one, I(1)인 경우가 가장 많습니다. 교수님께서는 데이터를 분석하실 때 2번까지 차분해야 stationary하게 되는 경우는 딱 한 번 보셨다고 하셨습니다. 만약 difference를 하지 않더라도 이미 stationary한 경우는 Integrated of order zero, I(0)으로 표현됩니다.

다시 ARMA, ARIMA로 돌아와봅시다. ARIMA는 ARMA에 Integration개념이 첨가된 것이기 때문에 ARMA만 이해해도 충분히 이해할 수 있을 것입니다. AR과 MA model은 각각 다음과 같이 쓸 수 있겠습니다.

AR(p) : $Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ldots + \alpha_p Y_{t-p} + u_t$, 단 u는 iid white-noise

MA(q) : $Y_t = u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \ldots + \beta_q u_{t-q}$

↑ PACF에 절단값 존재

↑ ACF에 절단값 존재

이 두 model이 섞여 있는 ARMA model은 다음과 같이 표현되겠습니다.

ARMA(p,q) : $Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ldots + \alpha_p Y_{t-p} + u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \ldots + \beta_q u_{t-q}$

더 나아가 ARIMA model을 생각해봅시다. ARMA model은 AR,MA의 차수를 순서대로 ARMA(p,q)라고 표현을 하는데요.(AR은 보통 p로 MA는 q로 차수를 표현하죠.) ARIMA는 ARIMA(p,d,q)로 표현을 합니다. d번 difference(차분)하면 time-series data가 stationary하게 변하면서 ARMA(p,q)의 형태가 된다는 것이죠.

Or AR(p) Or MA(q)

d번 차분해서 stationary한 time-series data가 ARMA(p,q)를 따르는 경우는 이렇게 생각하셔도 됩니다.

Y_t 가 non-statiaonary하고, d번 difference를 하면 stationary하게 변한다고 합시다.

$Z = (1-L)^d Y_t$ d번 difference를 행한 time-series data Z가 ARMA(p,q)를 따르는 것입니다.