# 부분자기상관계수 함수

## 부분자기상관계수 함수의 정의

**부분자기상관계수 함수(PACF: partial autocorrelation function)**는 AR 모형의 차수(order)를 추정하기 위한 방법의 하나이다.

일반적인 자기상관계수 함수 $\rho_k$ 는 확률 과정의 시각 $t$ 의 값 $Y_t$와 시각 $t-k$의 값 $Y_{t-k}$ 의 상관계수를 뜻한다. 부분자기상관계수 함수는 이와 달리 두 시각 사이의 값 즉 $Y_{t-1}, \cdots, Y_{t-k+1}$ 의 영향을 배제한다.

이 사이 값들의 영향을 없애기 위해 우선 $Y_t$ 값(종속 변수)과 $Y_{t-1}, \cdots, Y_{t-k+1}$ 라는 $k-1$개의 값들(독립 변수) 사이의 선형 회귀 분석을 실시한다.

$$\hat{Y}_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \cdots + \beta_{k-1} Y_{t-k+1}$$

이 식에서 $\beta$는 선형 회귀 분석의 결과로 얻어진 계수이다.

$k=2$인 경우만 따지면 계수 $\beta_1 = \rho_1$ 이 된다. 즉,

$$\hat{Y}_t = \rho_1 Y_{t-1}$$

이와 반대 (시간) 방향으로 $Y_{t-k}$ 값(종속 변수)과 $Y_{t-1}, \cdots, Y_{t-k+1}$ 라는 $k-1$개의 값들(독립 변수) 사이의 선형 회귀 분석을 실시하면 정상 확률 과정의 경우에 결과는 다음과 같다.

$$\hat{Y}_{t-k} = \beta_1 Y_{t-k+1} + \beta_2 Y_{t-k+2} + \beta_3 Y_{t-k+3} + \cdots + \beta_{k-1} Y_{t-1}$$

위 회귀 분석의 계수가 동일하게 나오는 이유는 정상 확률 과정의 특성때문이다.

첫번째 항을 예로 들면 살펴보면,

$$\text{corr}[Y_t, Y_{t-1}] = \text{corr}[Y_t, Y_{t+1}] = \text{corr}[Y_{t-k}, Y_{t-k+1}]$$

로 $Y_t \leftrightarrow Y_{t-1}$ 의 관계가 $Y_{t-k} \leftrightarrow Y_{t-k+1}$ 의 관계와 동일하다.

**부분자기상관계수 함수** $\phi_k$ 는 위와 같은 선형 회귀 분석으로 얻은 추정치 $\hat{Y}_t, \hat{Y}_{t-k}$를 원래의 값 $Y_t, Y_{t-k}$ 에서 뺀 나머지 값 사이의 자기상관계수 함수로 정의한다. 즉,

$$\phi_{kk} = \text{corr}[(Y_t - \hat{Y}_t), (Y_{t-k} - \hat{Y}_{t-k})] \;\; \text{for } k > 1$$

이다. $k=1$인 경우는 1로 정의한다.

$$\phi_{11} = 1$$

## AR 모형의 부분자기상관계수 함수

다음과 같은 AR(p) 모형에 대해

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t$$

$Y_t$ 값(종속 변수)과 $Y_{t-1}, \cdots, Y_{t-k+1}$라는 $k-1$개의 값들(독립 변수) 사이의 선형 회귀 분석을 하면 다음과 같은 결과를 얻는다.

$k=1$인 경우에는,

$$\hat{Y}_t = \phi_1 Y_{t-1}$$

$k=2$인 경우에는,

$$\hat{Y}_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2}$$

$k<p$인 경우에는,

$$\hat{Y}_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_k Y_{t-k}$$

$k \geq p$인 경우에는,

$$\hat{Y}_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + 0 \cdot Y_{t-p-1} + \cdots + 0 \cdot Y_{t-k}$$

이며 이 경우에는 회귀 분석의 잔차항(residual)이 원래의 확률 과정과 아무런 상관관계가 없는 백색 잡음이 된다.

AR(p) 모형은 정상 과정이므로 시간을 반대 방향으로 해도 같은 결과를 얻을 수 있다. 즉,

$Y_t$ 값(종속 변수)과 $Y_{t+1}, \cdots, Y_{t+k-1}$라는 $k-1$개의 값들(독립 변수) 사이의 선형 회귀 분석을 하면 다음과 같은 결과를 얻는다.

$k=1$인 경우에는,

$$\hat{Y}_t = \phi_1 Y_{t+1}$$

$k=2$인 경우에는,

$$\hat{Y}_t = \phi_1 Y_{t+1} + \phi_2 Y_{t+2}$$

$k \leq p$인 경우에는,

$$\hat{Y}_t = \phi_1 Y_{t+1} + \phi_2 Y_{t+2} + \cdots + \phi_k Y_{t+k}$$

$k>p$인 경우에는,

$$\hat{Y}_t = \phi_1 Y_{t+1} + \phi_2 Y_{t+2} + \cdots + \phi_p Y_{t+p} + 0 \cdot Y_{t+p+1} + \cdots + 0 \cdot Y_{t+k}$$

이다.

회귀 분석 결과를 알았으니 이제 부분자기상관계수를 구해보자.

우선 AR(1) 모형을 고려해 보자.

$$Y_t = \phi_1 Y_{t-1} + \epsilon_t$$

$k=1$인 경우, 정의에 의해

$$\phi_{11} = 1$$

$k=2$인 경우,

$$
\begin{aligned}
\phi_{22} &= \text{corr}[(Y_t - \hat{Y}_t), (Y_{t-2} - \hat{Y}_{t-2})] \\
&= \text{corr}[(Y_t - \phi_1 Y_{t-1}), (Y_{t-2} - \phi_1 Y_{t-1})] \\
&\propto \text{Cov}[(Y_t - \phi_1 Y_{t-1}), (Y_{t-2} - \phi_1 Y_{t-1})] \\
&= \text{E}[(Y_t - \phi_1 Y_{t-1})(Y_{t-2} - \phi_1 Y_{t-1})] \\
&= \gamma_0(\rho_2 - \phi_1 \rho_1 - \phi_1 \rho_1 + \phi_1^2 \rho_0)
\end{aligned}
$$

AR(1) 모형의 경우 이론적인 자기상관계수 함수 $\rho_k$ 는 다음과 같이 계산된다.

$$
\rho_k = \phi_1^k
$$

따라서

$$
\rho_2 - \phi_1 \rho_1 - \phi_1 \rho_1 + \phi_1^2 \rho_0 = \phi_1^2 - \phi_1 \cdot \phi_1 - \phi_1 \cdot \phi_1 + \phi_1^2 \cdot 1 = 0
$$

즉, $\phi_{22} = 0$이다.

$k = 3$인 경우도 마찬가지로 $\phi_{33} = 0$이 된다.

$$
\begin{aligned}
\phi_{33} &= \text{corr}[(Y_t - \hat{Y}_t), (Y_{t-3} - \hat{Y}_{t-3})] \\
&= \text{corr}[(Y_t - \phi_1 Y_{t-1}), (Y_{t-3} - \phi_1 Y_{t-2})] \\
&\propto \text{Cov}[(Y_t - \phi_1 Y_{t-1}), (Y_{t-3} - \phi_1 Y_{t-2})] \\
&= \text{E}[(Y_t - \phi_1 Y_{t-1})(Y_{t-3} - \phi_1 Y_{t-2})] \\
&= \gamma_0(\rho_3 - \phi_1 \rho_2 - \phi_1 \rho_2 + \phi_1^2 \rho_1) \\
&= \gamma_0(\phi_1^3 - \phi_1 \cdot \phi_2 - \phi_1 \cdot \phi_2 + \phi_1^2 \cdot \phi_1) \\
&= 0
\end{aligned}
$$

같은 방법으로 $k > 1$인 모든 경우에 대해 $\phi_{kk} = 0$이 된다.

같은 방식으로 AR(p) 모형의 경우에는 $k > p$ 인 모든 경우에 대해 부분자기상관계수 함수 $\phi_k = 0$ 이 됨을 증명할 수 있다.

이는 마치 MA(q) 모형의 경우 $k > q$ 인 모든 경우에 대해 자기상관계수 함수 $\rho_k = 0$ 이었던 것과 유사하다.

## MA 모형의 부분자기상관계수 함수

MA(q)모형의 경우에는 역으로 $k > q$인 경우에도 부분자기상관계수 함수 $\phi_{kk}$가 0이 되지 않는다.

MA(1)모형을 예로 들어 보자.

$$
Y_t = \epsilon_t - \theta_1 \epsilon_{t-1}
$$

인 MA(1) 모형의 경우,

$$
\begin{aligned}
\phi_{22} &= \text{corr}[(Y_t - \hat{Y}_t), (Y_{t-2} - \hat{Y}_{t-2})] \\
&= \text{corr}[(Y_t - \rho_1 Y_{t-1}), (Y_{t-2} - \rho_1 Y_{t-1})] \\
&= \frac{\text{Cov}[(Y_t - \rho_1 Y_{t-1}), (Y_{t-2} - \rho_1 Y_{t-1})]}{\sqrt{\text{Var}[Y_t - \rho_1 Y_{t-1}]\text{Var}[Y_{t-2} - \rho_1 Y_{t-1}]}} \\
&= \frac{\text{Cov}[(Y_t - \rho_1 Y_{t-1}), (Y_{t-2} - \rho_1 Y_{t-1})]}{\text{Var}[Y_t - \rho_1 Y_{t-1}]} \\
&= \frac{\text{E}[(Y_t - \rho_1 Y_{t-1})(Y_{t-2} - \rho_1 Y_{t-1})]}{\text{E}[(Y_t - \rho_1 Y_{t-1})^2]}
\end{aligned}
$$

MA(1)의 성질

$$E[Y_t^2] = \sigma_e^2(1+\theta^2)$$

$$E[Y_t Y_{t-1}] = -\theta\sigma_\epsilon^2$$

$$E[Y_t Y_{t-2}] = 0$$

$$\rho_1 = -\frac{\theta}{1+\theta^2}$$

을 이용하면 이 값은 다음과 같다.

$$\phi_{22} = -\frac{\theta^2}{1+\theta^2+\theta^4} \neq 0$$

$k > q$인 경우에 대해 모두 구하면

$$\phi_{kk} = -\frac{\theta^k(1-\theta^2)}{1-\theta^{2(k+1)}} \neq 0$$

즉 일반적으로 0이 되지 않는다.

이 결과를 정리하면 다음과 같다.

| | AR(p) | MA(q) |
|---|---|---|
| ACF | 점차적으로 감소 | 시차 $q$ 이후에 0 |
| PACF | 시차 $p$ 이후에 0 | 점차적으로 감소 |

표 15.3.1 : MA 모형과 AR 모형의 자기상관계수와 부분자기상관계수

# 부분자기상관계수 함수의 계산

부분자기상관계수 함수는 회귀분석을 사용하여 정의되었지만 실제 계산시에는 다음과 같은 특성을 이용하여 더 효율적으로 계산한다.

부분자기상관계수 함수의 값 $\phi_{kk}$는 $Y_t$를 $Y_{t-1}, \cdots, Y_{t-k}$라는 $k$개의 값들에 대해 회귀분석한 경우 $Y_{t-k}$의 계수와 일치한다.

$$\hat{Y}_t = \phi_{1k}Y_{t-1} + \phi_{2k}Y_{t-2} + \phi_{3k}Y_{t-3} + \cdots + \phi_{kk}Y_{t-k}$$

여기에 $Y_{t-1}, \cdots, Y_{t-k}$를 각각 곱해서 기댓값을 구하면 다음과 같이 Yule-Walker 방정식이 성립한다.

$$\begin{aligned}
\rho_1 &= \phi_{k1}\rho_0 &+ \phi_{k2}\rho_{-1} &\cdots \phi_{kk}\rho_{-k+1} \\
\rho_2 &= \phi_{k1}\rho_1 &+ \phi_{k2}\rho_0 &\cdots \phi_{kk}\rho_{-k+2} \\
\vdots &= \vdots &+ \vdots &\ddots \vdots \\
\rho_k &= \phi_{k1}\rho_{k-1} &+ \phi_{k2}\rho_{k-2} &\cdots \phi_{kk}\rho_0
\end{aligned}$$

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} = \begin{bmatrix} \rho_0 & \rho_{-1} & \cdots & \rho_{-k+1} \\ \rho_1 & \rho_0 & \cdots & \rho_{-k+2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_k & \rho_{k-1} & \cdots & \rho_0 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix}$$

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} = \begin{bmatrix} \rho_0 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_k & \rho_{k-1} & \cdots & \rho_0 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix}$$

Levinson-Durbin 방법을 사용하면 위의 Yule-Walker 방정식을 다음과 같이 재귀적으로 구할 수 있다.

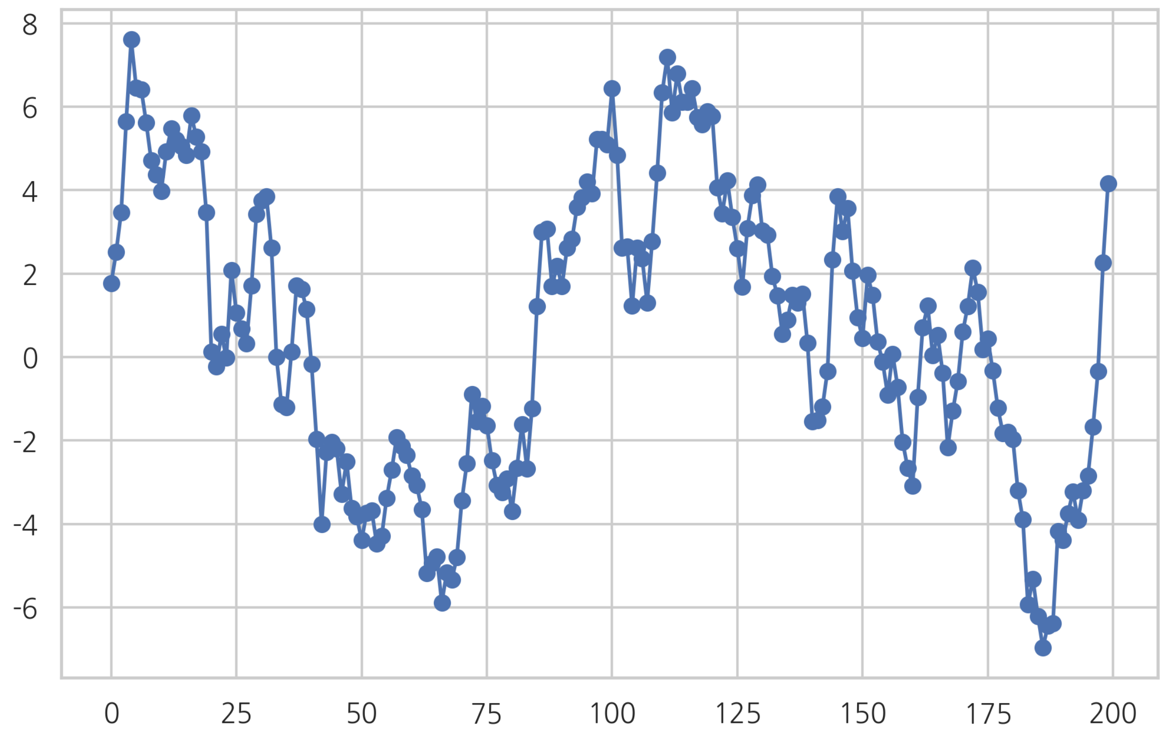
$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_j}$$

$$\phi_{k,j} = \phi_{k-1,j} - \phi_{k,k}\phi_{k-1,k-j}$$

`statsmodels.tsa.pacf` 함수를 사용하면 시계열 자료의 부분자기상관계수 함수를 계산할 수 있다. 플롯을 하려면 `statsmodels.graphics.tsa.plot_pacf` 를 이용한다.

In [1]:

```
np.random.seed(0)
p = sm.tsa.ArmaProcess([1, -1.2, 0.3], [1])
y = p.generate_sample(200)
plt.plot(y, 'o-')
plt.show()
```



In [2]:

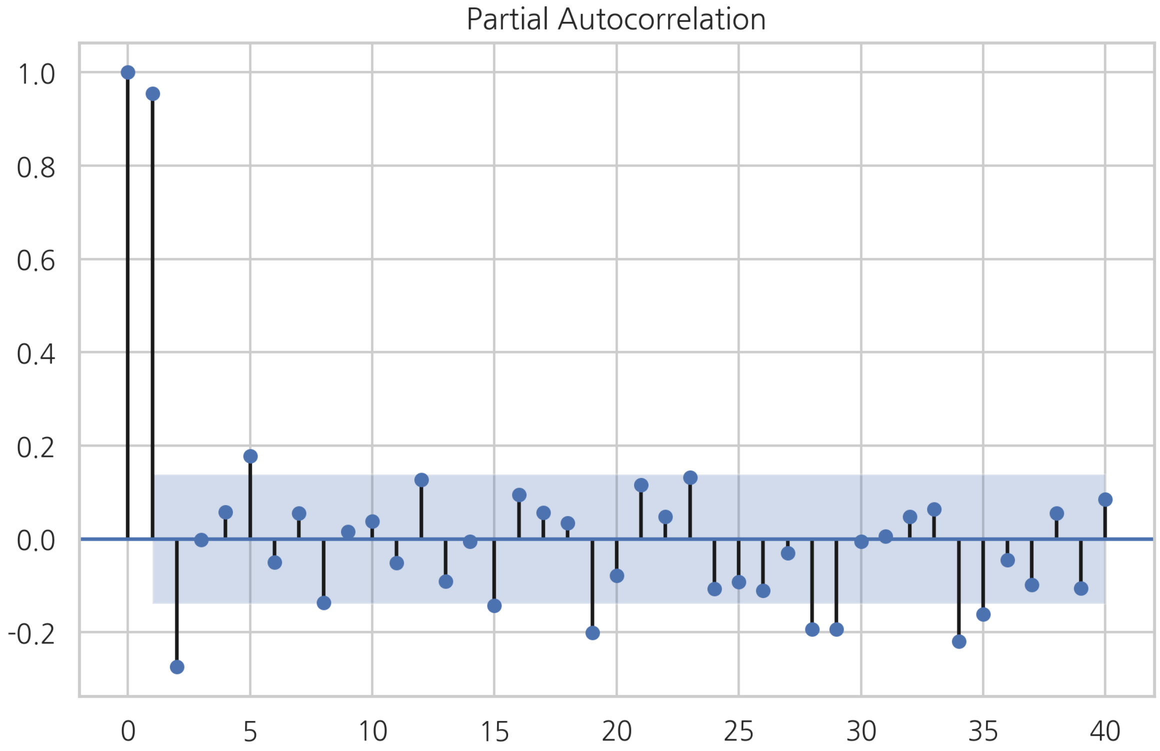
```
pacf = sm.tsa.pacf(y)
pacf
```

Out[2]:

```
array([ 1.        ,  0.954434  , -0.2736326 , -0.00179577,  0.05732259,
        0.17734768, -0.04928944,  0.05504672, -0.13653964,  0.01623869,
        0.03826421, -0.05072837,  0.12656022, -0.09010451, -0.00482013,
       -0.14227634,  0.09523362,  0.05618313,  0.03467563, -0.20119255,
       -0.07870102,  0.11652601,  0.04784023,  0.13198976, -0.10725904,
       -0.09147749, -0.1107981 , -0.02998764, -0.19292734, -0.19328953,
       -0.00497484,  0.00569143,  0.04810996,  0.06388207, -0.21928295,
       -0.16083584, -0.04436043, -0.09845151,  0.05537217, -0.10485931,
        0.08507789])
```

In [3]:

```
sm.graphics.tsa.plot_pacf(y, lags=40)
plt.show()
```




PACF 값을 ACF와 비교하면 차이점을 쉽게 볼 수 있다.

In [4]:

```
sm.graphics.tsa.plot_acf(y, lags=40)
plt.show()
```