

Autoregressive (AR) 모형

시계열 자료 중에는 시차가 증가해도 자기 상관계수가 비교적 오랫동안 0이 아닌 값으로 남아 있는 것들이 많다. 이러한 시계열에 MA 모형을 사용하면 차수가 너무 커진다. 이를 극복하기 위한 모형 중의 하나가 Autoregressive (AR) 모형이다. **AR 모형은 백색 잡음의 현재값과 자기 자신의 과거값의 선형 가중합으로 이루어진 정상 확률 모형이다.**

$$Y_t = -\phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \cdots - \phi_p Y_{t-p} + \epsilon_t$$

자기 자신의 과거 값을 사용하기 때문에 Autoregressive 라는 이름이 붙여졌다. AR 모형은 임의의 계수 값을 가질 수 있는 MA 모형과 달리 정상 과정이 되기 위해서는 **계수의 값에 제한 조건**이 붙는다. 구체적인 제한 조건은 차수에 따라 달라진다. AR 모형은 얼핏 보기에 일반 선형 확률 과정(general linear process)의 수식을 따르지 않는 것처럼 보이지만 일반 선형 확률 과정의 형태로 변환이 가능하다.

AR(1) 모형

1차 AR 모형 즉, AR(1) 모형은 백색 잡음의 현재값과 1스텝 과거의 자기 자신의 값만의 가중합으로 이루어진 모형이다. 수식은 다음과 같다.

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

이 수식에서 계수 ϕ 는 다음 조건을 만족해야 한다.

$$-1 < \phi < 1$$

AR(1) 모형은 다음과 같이 일반 선형 확률 과정의 형태로 바꿀 수 있다.

$$\begin{aligned} Y_t &= -\phi Y_{t-1} + \epsilon_t \\ &= -\phi (-\phi Y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \phi^2 Y_{t-2} - \phi \epsilon_{t-1} + \epsilon_t \\ &= \phi^2 (-\phi Y_{t-3} + \epsilon_{t-2}) - \phi \epsilon_{t-1} + \epsilon_t \\ &= -\phi^3 Y_{t-3} + \phi^2 \epsilon_{t-2} - \phi \epsilon_{t-1} + \epsilon_t \\ &\vdots \\ &= \epsilon_t - \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} - \phi^3 \epsilon_{t-3} + \cdots \end{aligned}$$

AR(1) 모형의 기대값은 MA모형과 마찬가지로 0이다.

$$E[Y_t] = 0$$

이는 다음과 같이 증명할 수 있다.

$$\begin{aligned} E[Y_t] &= \mu = E[-\phi Y_{t-1} - \epsilon_{t-1}] \\ &= -\phi E[Y_{t-1}] - E[\epsilon_{t-1}] \\ &= -\phi \mu - 0 \\ (1 + \phi)\mu &= 0 \\ \mu &= 0 \text{ if } \phi \neq -1 \end{aligned}$$

AR(1) 모형의 분산은 다음과 같다.

$$\gamma_0 = \frac{\sigma_\epsilon^2}{1 - \phi^2}$$

이는 다음과 같이 증명할 수 있다.

우선 ϵ_t 는 과거의 값 Y_{t-l} (for $l > 1$)에 영향을 주지 않기 때문에 독립이다. 따라서

$$E[\epsilon_t Y_{t-l}] = E[\epsilon_t]E[Y_{t-l}] = 0$$

이다.

$$\begin{aligned}\text{Var}[Y_t] = \gamma_0 &= E[Y_t^2] = E[(-\phi Y_{t-1} - \epsilon_t)^2] \\ &= E[\phi^2 Y_{t-1}^2 + 2\phi Y_{t-1} \epsilon_t + \epsilon_t^2] \\ &= \phi^2 E[Y_{t-1}^2] + 2\phi E[Y_{t-1} \epsilon_t] + E[\epsilon_t^2] \\ &= \phi^2 \gamma_0 + 0 + \sigma_\epsilon^2\end{aligned}$$

$$\begin{aligned}(1 - \phi^2)\gamma_0 &= \sigma_\epsilon^2 \\ \gamma_0 &= \frac{\sigma_\epsilon^2}{1 - \phi^2} \text{ if } \phi^2 \neq 1\end{aligned}$$

AR(1) 모형의 자기공분산은 다음과 같다.

$$\gamma_l = (-\phi)^l \frac{\sigma_\epsilon^2}{1 - \phi^2}$$

이는 다음과 같이 증명한다.

$$\begin{aligned}\gamma_l &= E[Y_{t-l} Y_t] \\ &= E[-\phi Y_{t-l} Y_{t-1} - Y_{t-l} \epsilon_t] \\ &= -\phi E[Y_{t-l} Y_{t-1}] - E[Y_{t-l} \epsilon_t] \\ &= -\phi \gamma_{l-1}\end{aligned}$$

$\gamma_0 = \frac{\sigma_\epsilon^2}{1 - \phi^2}$ 과 위 성질에 의해 귀납적으로 증명된다.

자기공분산으로부터 자기상관계수는 다음과 같다

$$\rho_l = \frac{\gamma_l}{\gamma_0} = (-\phi)^l$$

1과 -1사이의 여러가지 ϕ 에 대해 AR(1) 모형의 자기상관계수 함수 즉, ACF를 그리면 다음과 같다.

- $\phi < 0$ 이면 ACF는 지수함수적으로 감소한다.
- $\phi > 0$ 이면 ACF는 부호를 바꿔가면서 (진동하면서) 지수함수적으로 감소한다.

In [1]:

```
lag = np.arange(12)

plt.subplot(221)
acf = 0.9 ** lag
plt.stem(acf)
plt.xlim(-0.2, 11.2)
plt.ylim(-0.1, 1.1)

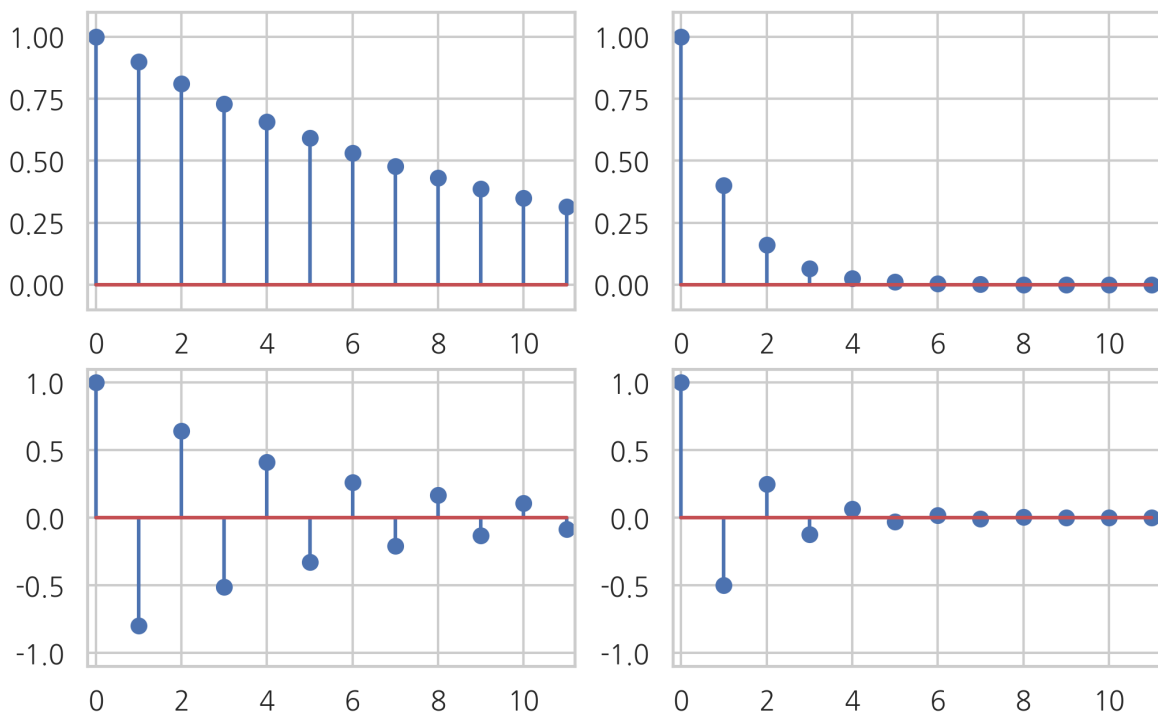
plt.subplot(222)
acf = 0.4 ** lag
plt.stem(acf)
plt.xlim(-0.2, 11.2)
plt.ylim(-0.1, 1.1)

plt.subplot(223)
acf = (-0.8) ** lag
plt.stem(acf)
plt.xlim(-0.2, 11.2)
plt.ylim(-1.1, 1.1)

plt.subplot(224)
acf = (-0.5) ** lag
plt.stem(acf)
plt.xlim(-0.2, 11.2)
plt.ylim(-1.1, 1.1)

plt.suptitle("AR(1) 모형의 자기상관계수 함수의 예")
plt.show()
```

AR(1) 모형의 자기상관계수 함수의 예



앞의 그림에서 볼 수 있듯이 ϕ 의 절대값이 1에 가까울 수록 자기상관계수의 감소 속도가 작아진다. 즉, 시차가 증가해도 상관계수가 크게 작아지지 않는다. 그럼 만약 ϕ 의 절대값이 1이거나 1보다 크다면 어떻게 될까?

ϕ 값의 크기가 1 이상이 되면 분산이 계속 증가하는 비정상 과정이 된다. $\phi = -1$ 인 경우를 보자.

$$\begin{aligned} Y_t &= Y_{t-1} + \epsilon_t \\ \text{Var}[Y_t] &= \text{Var}[Y_{t-1} + \epsilon_t] \\ &= \text{Var}[Y_{t-1}] + \text{Var}[\epsilon_t] \quad (\text{independence}) \\ \text{Var}[Y_t] &> \text{Var}[Y_{t-1}] \end{aligned}$$

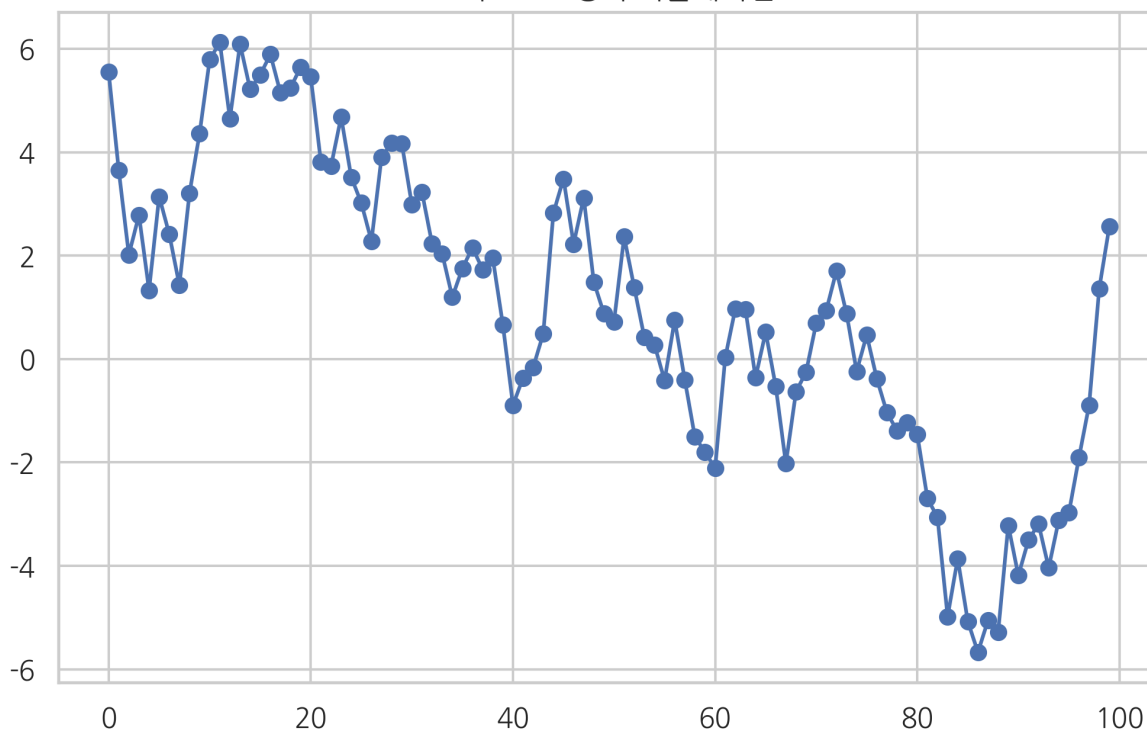
AR(1) 모형의 경우 $|\phi| < 1$ 조건은 **정상상태 조건(stationarity condition)**이 된다.

AR(1) 모형의 시뮬레이션

In [2]:

```
np.random.seed(0)
p1 = sm.tsa.ArmaProcess([1, -0.9], [1])
y1 = p1.generate_sample(100, burnin=100)
plt.plot(y1, 'o-')
plt.title("1차 AR 모형의 시뮬레이션")
plt.show()
```

1차 AR 모형의 시뮬레이션



이 시계열 데이터의 l -시차 상관계수, 즉 Y_t 와 Y_{t-l} 의 상관계수를 조사하기 위해 스캐터플롯을 그려보았다. 지속적으로 강한 상관관계를 보인다.

In [3]:

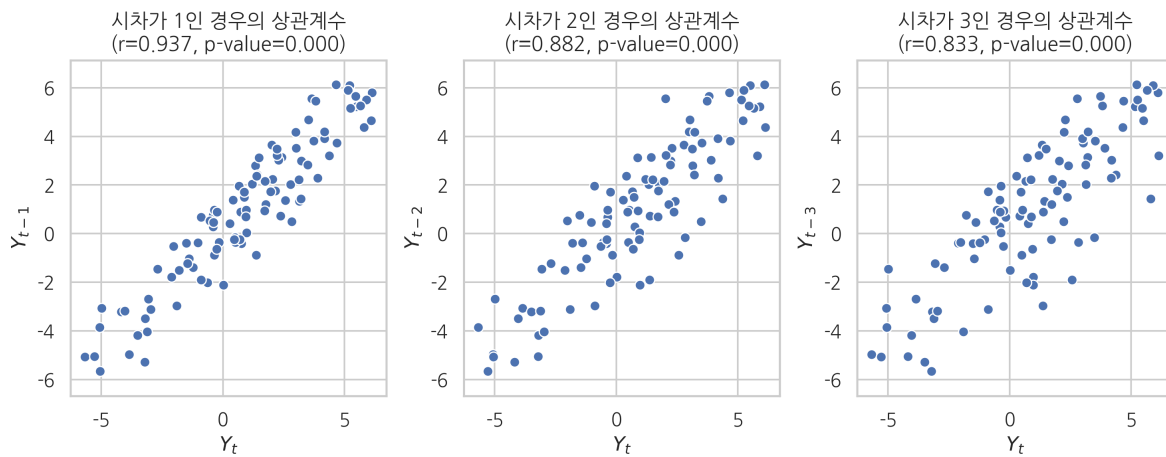
```
plt.figure(figsize=(10, 4))

plt.subplot(131)
r, p = sp.stats.pearsonr(y1[1:], y1[:-1])
sns.scatterplot(y1[1:], y1[:-1])
plt.axis("equal")
plt.xlabel("$Y_t$")
plt.ylabel("$Y_{t-1}$")
plt.title("시차가 1인 경우의 상관계수Wn(r={0:.3f}, p-value={1:.3f})".format(r, p))

plt.subplot(132)
r, p = sp.stats.pearsonr(y1[2:], y1[:-2])
sns.scatterplot(y1[2:], y1[:-2])
plt.axis("equal")
plt.xlabel("$Y_t$")
plt.ylabel("$Y_{t-2}$")
plt.title("시차가 2인 경우의 상관계수Wn(r={0:.3f}, p-value={1:.3f})".format(r, p))

plt.subplot(133)
r, p = sp.stats.pearsonr(y1[3:], y1[:-3])
sns.scatterplot(y1[3:], y1[:-3])
plt.axis("equal")
plt.xlabel("$Y_t$")
plt.ylabel("$Y_{t-3}$")
plt.title("시차가 3인 경우의 상관계수Wn(r={0:.3f}, p-value={1:.3f})".format(r, p))

plt.tight_layout()
plt.show()
```



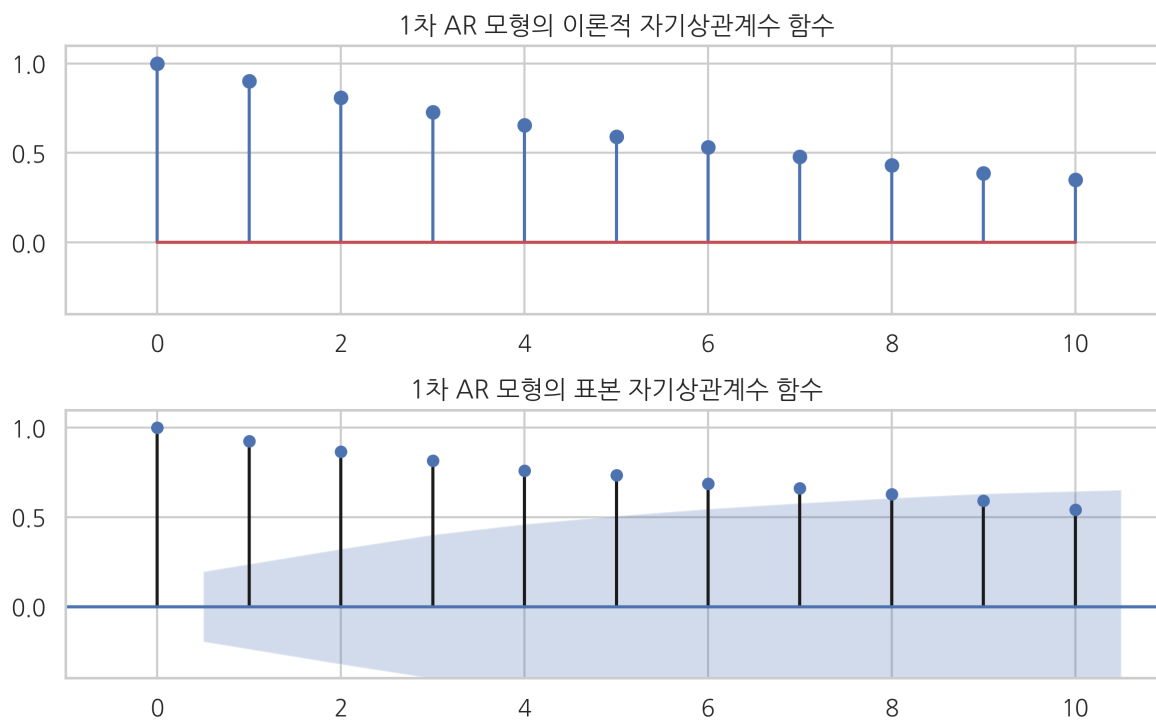
이론적 자기상관계수 함수와 표본 자기상관계수 함수를 그리면 다음과 같다.

In [4]:

```
plt.subplot(211)
plt.stem(p1.acf(11))
plt.xlim(-1, 11)
plt.ylim(-0.4, 1.1)
plt.title("1차 AR 모형의 이론적 자기상관계수 함수")

ax = plt.subplot(212)
sm.graphics.tsa.plot_acf(y1, lags=10, ax=ax)
plt.xlim(-1, 11)
plt.ylim(-0.4, 1.1)
plt.title("1차 AR 모형의 표본 자기상관계수 함수")

plt.tight_layout()
plt.show()
```



AR(2) 모형

2차 AR 모형 즉, AR(2) 모형은 백색 잡음의 현재값과 1스텝 과거의 자기 자신의 값, 그리고 2스텝 과거의 자기 자신의 값의 가중합으로 이루어진 모형이다. 수식은 다음과 같다.

$$Y_t = -\phi_1 Y_{t-1} - \phi_2 Y_{t-2} + \epsilon_t$$

시간 지연 연산자(lag operator) L 을 사용하면 위 식은 다음과 같이 표현할 수 있다.

$$Y_t = -\phi_1 L Y_t - \phi_2 L^2 Y_t + \epsilon_t$$

$$(1 + \phi_1 L + \phi_2 L^2)Y_t = \epsilon_t$$

이 식을 변환한 다음과 같은 식을 **특성 다항식(characteristic polynomial)**이라 하고

$$1 + \phi_1 x + \phi_2 x^2$$

특성 다항식이 0이 되는 방정식을 **특정 방정식(characteristic equation)**이라고 한다.

$$1 + \phi_1 x + \phi_2 x^2 = 0$$

AR(2) 모형이 정상 과정을 나타내려면 **특성 방정식의 해의 크기(modulus)**가 1보다 커야 한다.

사실 이 정상상태 조건은 차수와 관계없이 임의의 AR(p)모형에 대해서도 적용된다.

AR(2) 모형의 경우 특성 방정식의 해는 다음과 같다.

$$\frac{-\phi_1 \pm \sqrt{\phi_1^2 - 4\phi_2}}{2\phi_2}$$

두 해가 모두 절대값이 1보다 커야 한다는 조건을 변환하면 다음 조건들을 동시에 만족하는 것과 같다.

$$\phi_1 + \phi_2 > -1$$

$$\phi_2 - \phi_1 > -1$$

$$|\phi_2| < 1$$

AR(2) 모형의 자기상관계수는 다음과 같이 계산 가능하다.

우선 AR(2)의 자기공분산은 다음 관계를 만족한다. 이 식을 **Yule-Waller 방정식**이라고 한다.

$$\gamma_l = -\phi_1 \gamma_{l-1} - \phi_2 \gamma_{l-2} = 0$$

이는 다음과 같이 증명할 수 있다.

$$\begin{aligned}\gamma_l &= E[Y_t Y_{t-l}] \\ &= E[(-\phi_1 Y_{t-1} - \phi_2 Y_{t-2} + \epsilon_t) Y_{t-l}] \\ &= -E[\phi_1 Y_{t-1} Y_{t-l}] - E[\phi_2 Y_{t-2} Y_{t-l}] + E[\epsilon_t Y_{t-l}] \\ &= -\phi_1 \gamma_{l-1} - \phi_2 \gamma_{l-2}\end{aligned}$$

이 식을 γ_0 로 나누면 자기상관계수에 대한 다음 식을 구할 수 있다.

$$\rho_l = -\phi_1 \rho_{l-1} - \phi_2 \rho_{l-2}$$

$\rho_0 = 1$ 과 $\rho_{-l} = \rho_l$ 를 이용하면

$$\rho_1 = -\phi_1 \rho_0 - \phi_2 \rho_{-1} = -\phi_1 - \phi_2 \rho_1$$

$$\rho_1 = \frac{-\phi_1}{1 + \phi_2}$$

같은 방법으로

$$\rho_2 = \frac{\phi_1^2 - \phi_2(1 + \phi_2)}{1 + \phi_2}$$

도 구할 수 있으며

이를 l 인 경우에 대해 확장하면 다음과 같다.

$$\rho_l = \left(1 + \frac{1 - \phi_2}{1 + \phi_2} \cdot l\right) \left(-\frac{\phi_1}{2}\right)^l$$

여러가지 ϕ_1, ϕ_2 에 대해 AR(2) 모형의 자기상관계수 함수 즉, ACF를 그리면 다음과 같은 특성을 보인다.

- 시차가 증가하면서 자기상관계수의 절대값은 지수함수적으로 감소한다.
- 진동할 수 있다. 진동 주파수는 계수에 따라 달라진다.

In [5]:

```
import statsmodels.api as sm

plt.subplot(221)
p1 = sm.tsa.ArmaProcess([1, -0.5, -0.25], [1])
plt.stem(p1.acf(11))
plt.xlim(-0.1, 10.1)

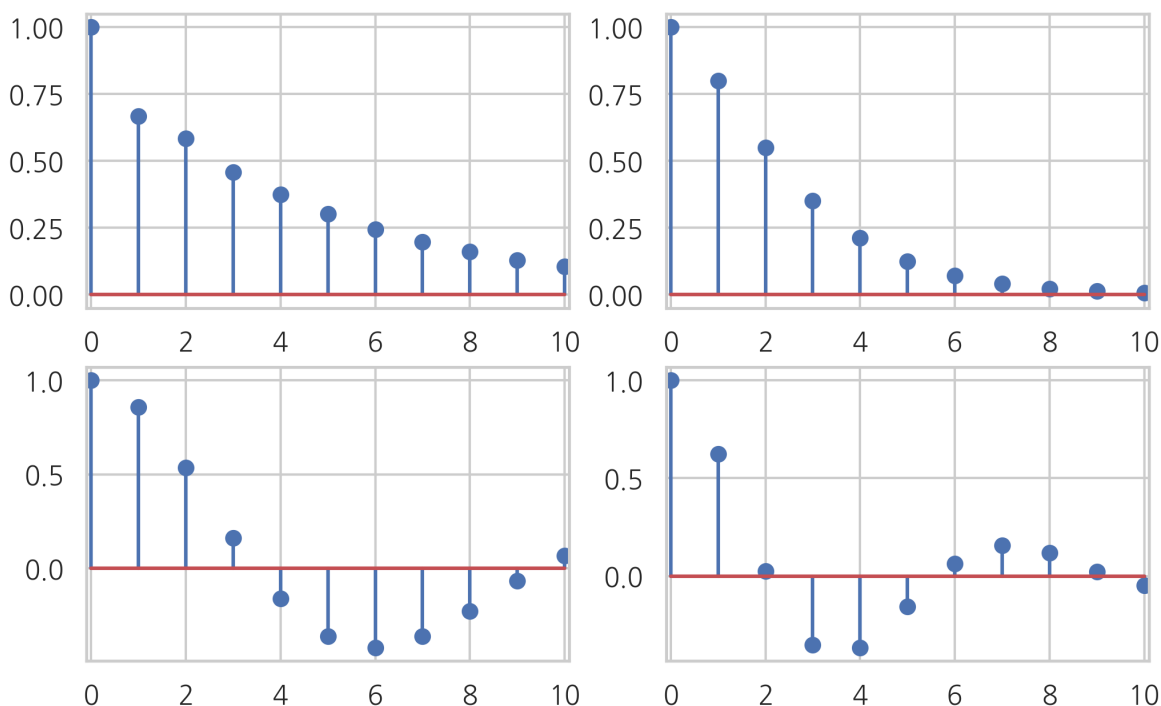
plt.subplot(222)
p1 = sm.tsa.ArmaProcess([1, -1, 0.25], [1])
plt.stem(p1.acf(11))
plt.xlim(-0.1, 10.1)

plt.subplot(223)
p1 = sm.tsa.ArmaProcess([1, -1.5, 0.75], [1])
plt.stem(p1.acf(11))
plt.xlim(-0.1, 10.1)

plt.subplot(224)
p1 = sm.tsa.ArmaProcess([1, -1, 0.6], [1])
plt.stem(p1.acf(11))
plt.xlim(-0.1, 10.1)

plt.suptitle("AR(2) 모형의 자기상관계수 함수의 예")
plt.show()
```

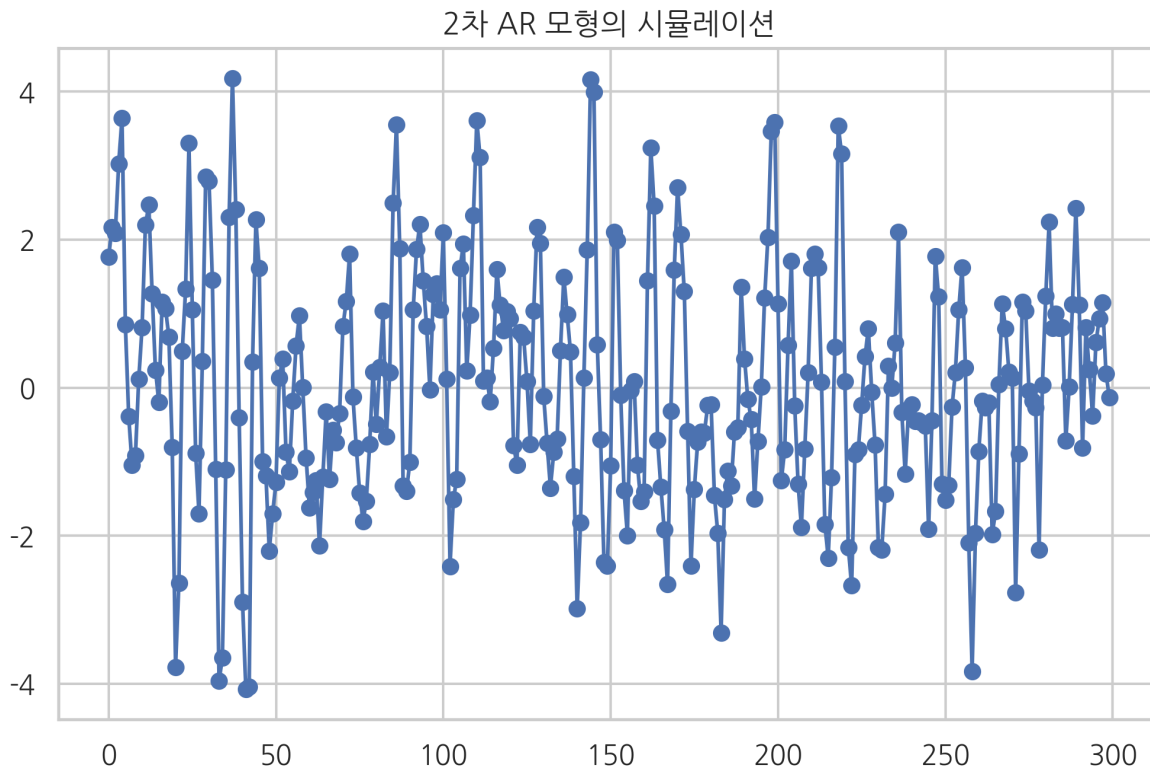
AR(2) 모형의 자기상관계수 함수의 예



AR(2) 모형의 시뮬레이션

In [6]:

```
np.random.seed(0)
p2 = sm.tsa.ArmaProcess([1, -1, 0.6], [1])
y2 = p2.generate_sample(300)
plt.plot(y2, "o-")
plt.title("2차 AR 모형의 시뮬레이션")
plt.show()
```



상관관계를 위한 스캐터플롯, 이론적 자기상관계수 함수와 표본 자기상관계수 함수를 그리면 다음과 같다.

In [7]:

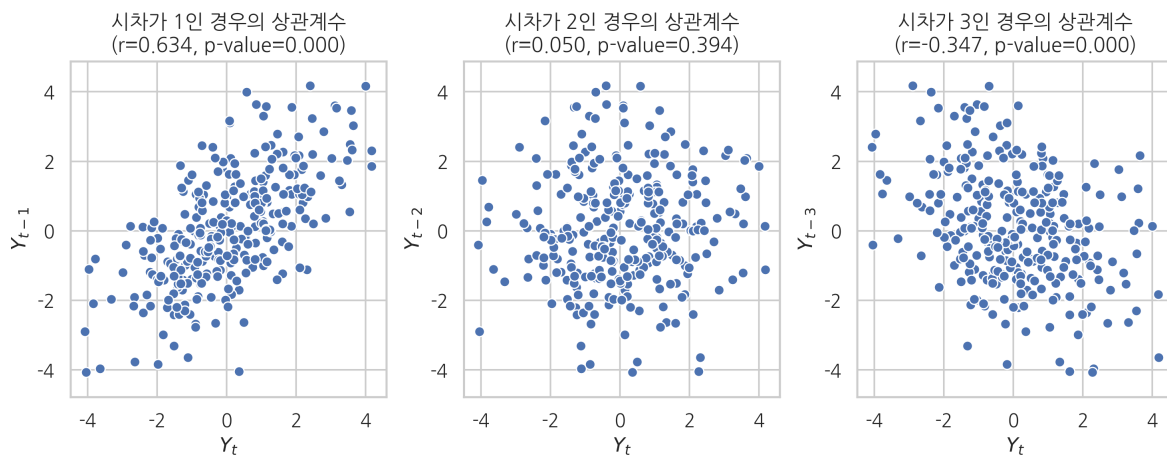
```
plt.figure(figsize=(10, 4))

plt.subplot(131)
r, p = sp.stats.pearsonr(y2[1:], y2[:-1])
sns.scatterplot(y2[1:], y2[:-1])
plt.axis("equal")
plt.xlabel("$Y_t$")
plt.ylabel("$Y_{t-1}$")
plt.title("시차가 1인 경우의 상관계수Wn(r={0:.3f}, p-value={1:.3f})".format(r, p))

plt.subplot(132)
r, p = sp.stats.pearsonr(y2[2:], y2[:-2])
sns.scatterplot(y2[2:], y2[:-2])
plt.axis("equal")
plt.xlabel("$Y_t$")
plt.ylabel("$Y_{t-2}$")
plt.title("시차가 2인 경우의 상관계수Wn(r={0:.3f}, p-value={1:.3f})".format(r, p))

plt.subplot(133)
r, p = sp.stats.pearsonr(y2[3:], y2[:-3])
sns.scatterplot(y2[3:], y2[:-3])
plt.axis("equal")
plt.xlabel("$Y_t$")
plt.ylabel("$Y_{t-3}$")
plt.title("시차가 3인 경우의 상관계수Wn(r={0:.3f}, p-value={1:.3f})".format(r, p))

plt.tight_layout()
plt.show()
```



In [8]:

```
plt.subplot(211)
plt.stem(p2.acf(11))
plt.xlim(-1, 11)
plt.title("2차 AR 모형의 이론적 자기상관계수 함수")

ax = plt.subplot(212)
sm.graphics.tsa.plot_acf(y2, lags=10, ax=ax)
plt.xlim(-1, 11)
plt.title("2차 AR 모형의 표본 자기상관계수 함수")

plt.tight_layout()
plt.show()
```

