7.2 MA모형

q**차 MA(Moving Average) 모형** MA(q)는 확률과정의 현재 값 Y_t이 백색 잡음의 현재 값 ϵ_t부터 q-시간 지연된 ϵ_{t-q}까지 $q+1$개 항의 선형 가중합인 확률과정 모형이다. MA(q)라고 쓴다.

$$Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

1차 MA모형

1차 MA모형 MA(1)의 현재 값 Y_t은 백색 잡음의 현재 값과 1-스텝 지연된 과거 값의 합이다.

$$Y_t = \epsilon_t + \theta \epsilon_{t-1}$$

MA(1) 모형의 기댓값 μ와 분산 γ_0, 그리고 자기공분산 γ_l는 각각 다음과 같다.

$$\mu = \text{E}[Y_t] = 0$$

$$\gamma_0 = \text{Var}[Y_t] = \sigma_\epsilon^2 (1 + \theta^2)$$

$$\gamma_1 = \text{Cov}[Y_t, Y_{t-1}] = \theta \sigma_\epsilon^2$$

$$\gamma_l = 0 \;\; \text{for } l > 1$$

이 식에서 σ_ϵ^2는 백색 잡음의 분산이다.

증명은 기댓값의 성질과 MA(1) 모형식을 이용한다.

$$\begin{aligned} \text{E}[Y_t] &= \text{E}\left[\epsilon_t + \theta \epsilon_{t-1}\right] \\ &= \text{E}[\epsilon_t] + \theta \text{E}[\epsilon_{t-1}] \\ &= 0 + \theta \cdot 0 \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Var}[Y_t] &= \text{E}\left[(\epsilon_t + \theta \epsilon_{t-1})^2\right] \\ &= \text{E}[\epsilon_t^2] + 2\theta \text{E}[\epsilon_t \epsilon_{t-1}] + \theta^2 \text{E}[\epsilon_{t-1}^2] \\ &= \sigma_\epsilon^2 + 2\theta \cdot 0 + \theta^2 \sigma_\epsilon^2 \\ &= \sigma_\epsilon^2 (1 + \theta^2) \end{aligned}$$

시차가 1, 2인 MA(1) 모형의 자기공분산을 계산해보자.

$$\begin{aligned} \text{Cov}[Y_t, Y_{t-1}] &= \text{E}\left[(\epsilon_t + \theta \epsilon_{t-1})(\epsilon_{t-1} + \theta \epsilon_{t-2})\right] \\ &= \text{E}[\epsilon_t \epsilon_{t-1}] + \theta \text{E}[\epsilon_t \epsilon_{t-2}] + \theta \text{E}[\epsilon_{t-1}^2] + \theta^2 \text{E}[\epsilon_{t-1} \epsilon_{t-2}] \\ &= 0 + \theta \cdot 0 + \theta \sigma_\epsilon^2 + \theta^2 \cdot 0 \\ &= \theta \sigma_\epsilon^2 \end{aligned}$$

$$\begin{aligned} \text{Cov}[Y_t, Y_{t-2}] &= \text{E}\left[(\epsilon_t + \theta \epsilon_{t-1})(\epsilon_{t-2} + \theta \epsilon_{t-3})\right] \\ &= \text{E}[\epsilon_t \epsilon_{t-2}] + \theta \text{E}[\epsilon_t \epsilon_{t-3}] + \theta \text{E}[\epsilon_{t-1} \epsilon_{t-2}] + \theta^2 \text{E}[\epsilon_{t-1} \epsilon_{t-3}] \\ &= 0 + \theta \cdot 0 + \theta \cdot 0 + \theta^2 \cdot 0 \\ &= 0 \end{aligned}$$

공분산을 분산으로 나누면 자기상관계수가 다음 성질을 만족한다는 것도 증명할 수 있다.

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

$$\rho_l = 0 \;\; \text{for } l > 1$$

1차 MA모형의 시뮬레이션

statsmodels의 tsa 서브패키지는 ARMA 확률과정 모형을 위한 `ArmaProcess` 클래스를 지원한다. 클래스 생성자는 다음과 같다.

```
ArmaProcess(ar_coeff, ma_coeff)
```

생성자의 인수로 사용되는 `ar_coeff` 와 `ma_coeff` 는 모형 수식에 나오는 방정식 $\phi(L)$, $\theta(L)$의 계수로 이루어진 리스트들이다. 가장 앞의 1이라는 숫자는 항상 있어야 한다.

$$Y_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} \cdots + \phi_p Y_{t-p} = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots \theta_q \epsilon_{t-q}$$

$$\phi(L) Y_t = \mu + \theta(L) \epsilon_t$$

$$\begin{aligned} \phi(L) &= 1 + \phi_1 L + \phi_2 L^2 + \cdots \phi_p L^p \\ \theta(L) &= 1 + \theta_1 L + \theta_2 L^2 + \cdots \theta_q L^q \end{aligned}$$

따라서 MA(1) 모형

$$Y_t = \epsilon_t + \theta \epsilon_{t-1}$$

은 다음과 같이 만들 수 있다. 예제 코드에서는 $\theta = 0.9$를 사용하였다.

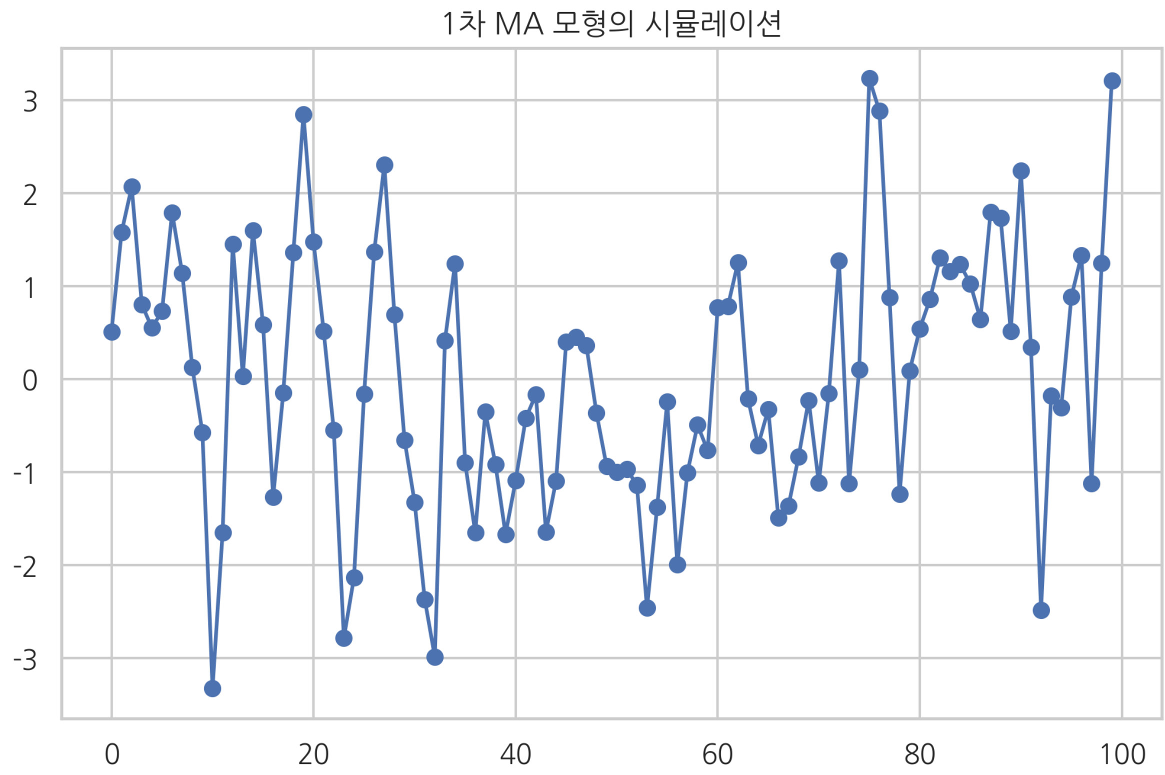
In [1]:

```
theta = 0.9
ar = [1]
ma = [1, theta]
p1 = sm.tsa.ArmaProcess(ar, ma)
```

`ArmaProcess` 클래스의 `generate_sample` 메서드를 이용하여 시계열 표본을 만들 수 있다. `generate_sample` 메서드는 표본 시계열의 길이와 `burnin` 인수를 받는다. MA(q) 모형에서는 `burnin` 인수가 q보다 같거나 커야 한다.

In [16]:

```
np.random.seed(0)
y1 = p1.generate_sample(100, burnin=10)
plt.plot(y1[:100], 'o-')
plt.title("1차 MA모형의 시뮬레이션")
plt.show()
```



이 시계열 데이터의 l-시차 상관계수, 즉 Y_t와 Y_{t-l}의 상관계수를 조사하기 위해 스캐터플롯을 그려보았다. $l = 1$일 때만 상관관계가 있고 $l > 1$이면 상관관계가 없다는 것을 확인할 수 있다.

In [17]:

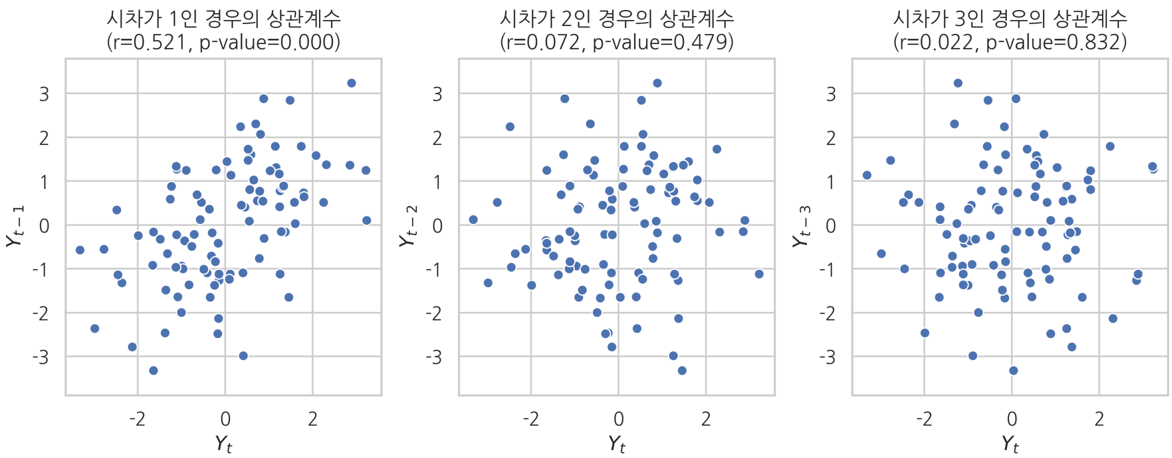
```
plt.figure(figsize=(10, 4))

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 상관계수₩n(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 상관계수₩n(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 상관계수₩n(r={0:.3f}, p-value={1:.3f})".format(r, p))

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



이론적 자기상관계수 함수와 표본 자기상관계수 함수를 그리면 다음과 같다.

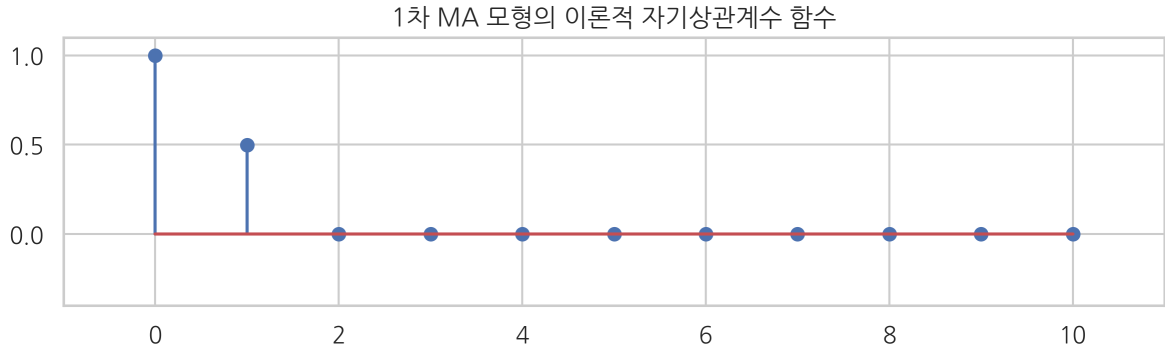
ArmaProcess 클래스의 acf 메서드는 이론적 자기상관계수 함수의 값을 출력한다. statsmodels.graphics.tsa.plot_acf 명령은 표본 시계열 데이터로부터 표본 자기상관계수 함수의 값을 계산한다.

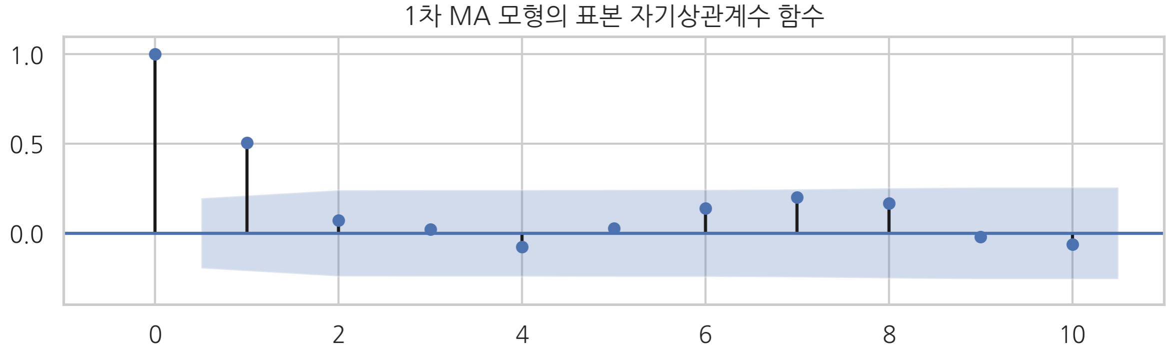
In [18]:

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

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





2차 MA모형

2차 MA모형은 다음과 같이 백색 잡음의 현재 값과 1-시간 지연된 과거 값, 그리고 2-시간 지연된 과거 값의 선형 가중합으로 나타난다.

$$Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$$

MA(2) 모형의 기댓값과 분산은 다음 성질을 만족한다.

$$\mu = \text{E}[Y_t] = 0$$

$$\gamma_0 = \text{Var}[Y_t] = \sigma_\epsilon^2 (1 + \theta_1^2 + \theta_2^2)$$

이 식에서 σ_ϵ^2는 백색 잡음의 분산이다.

이는 다음과 같이 증명할 수 있다.

$$
\begin{aligned}
\text{E}[Y_t] &= \text{E}\left[\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}\right] \\
&= \text{E}[\epsilon_t] + \theta_1\text{E}[\epsilon_{t-1} + \theta_2\text{E}[\epsilon_{t-2}] \\
&= 0 + \theta_1 \cdot 0 + \theta_2 \cdot 0 \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
\text{Var}[Y_t] &= \text{E}\left[(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2})^2\right] \\
&= \text{E}[\epsilon_t^2] + \theta_1^2\text{E}[\epsilon_{t-1}^2] + \theta_2^2\text{E}[\epsilon_{t-2}^2] + 2\theta_1\text{E}[\epsilon_t\epsilon_{t-1}] + 2\theta_2\text{E}[\epsilon_t\epsilon_{t-2}] + 2\theta_1\theta_2\text{E}[\epsilon_{t-1}\epsilon_{t-2}] \\
&= \sigma_\epsilon^2 + \theta_1^2\sigma_\epsilon^2 + \theta_2^2\sigma_\epsilon^2 + 2\theta_1 \cdot 0 + 2\theta_2 \cdot 0 + 2\theta_1\theta_2 \cdot 0 \\
&= \sigma_\epsilon^2(1 + \theta_1^2 + \theta_2^2)
\end{aligned}
$$

MA(2) 모형의 자기공분산은 다음과 같은 성질을 만족한다.

$$\gamma_1 = \text{Cov}[Y_t, Y_{t-1}] = (\theta_1 + \theta_1\theta_2)\sigma_\epsilon^2$$

$$\gamma_2 = \text{Cov}[Y_t, Y_{t-2}] = \theta_2\sigma_\epsilon^2$$

$$\gamma_l = 0 \;\; \text{for } l > 2$$

위 성질에서 MA(2) 모형의 자기상관계수는 다음 성질을 만족한다.

$$\rho_1 = \frac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

$$\rho_2 = \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

$$\rho_l = 0 \;\; \text{for } l > 2$$

2차 MA모형의 시뮬레이션

다음과 같은 MA(2) 모형의 예를 살펴보자

$$Y_t = \epsilon_t - \epsilon_{t-1} + 0.6\epsilon_{t-2}$$

앞의 성질을 이용하면 이 모형에서 생성된 시계열의 이론적인 자기상관계수는 다음과 같다.

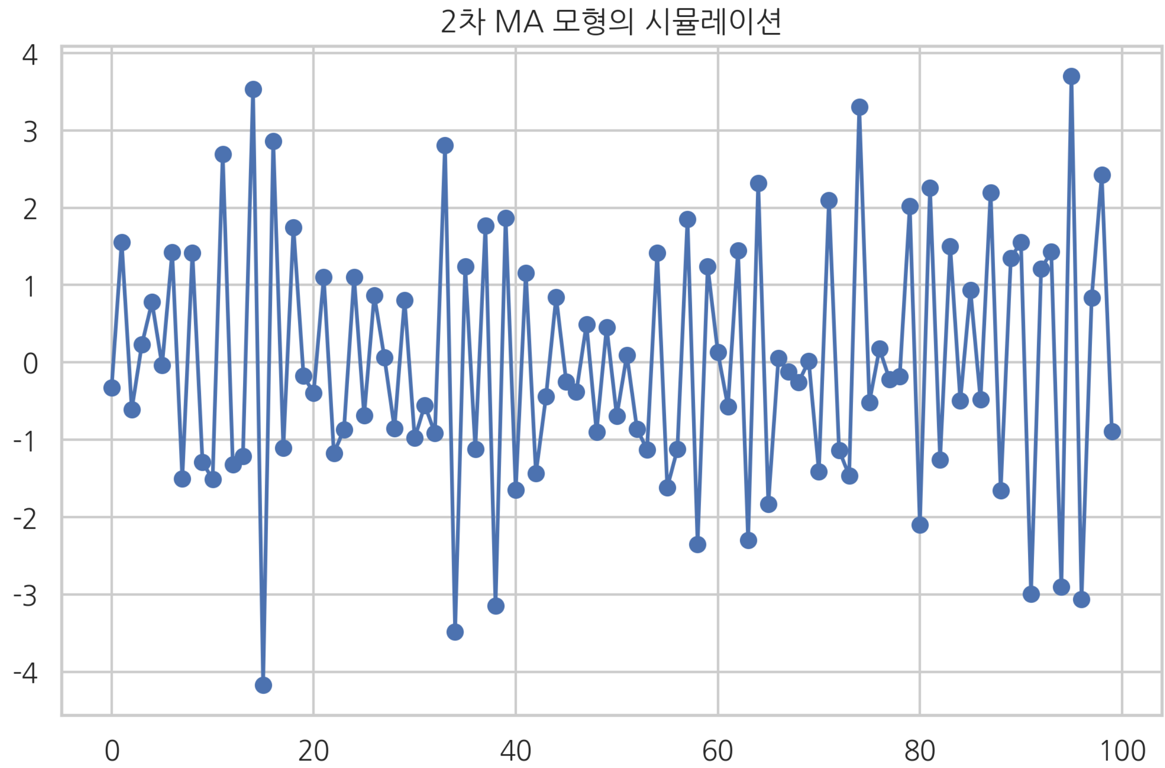
$$\rho_1 = -0.678$$

$$\rho_2 = 0.254$$

$$\rho_l = 0 \;\; \text{for } l > 2$$

In [5]:

```
np.random.seed(0)
p2 = sm.tsa.ArmaProcess([1], [1, -1, 0.6])
y2 = p2.generate_sample(100, burnin=10)
plt.plot(y2, 'o-')
plt.title("2차 MA모형의 시뮬레이션")
plt.show()
```



이 시계열 데이터의 l-시차 상관계수, 즉 Y_t와 Y_{t-l}의 상관계수를 조사하기 위해 스캐터플롯을 그려보았다. $l = 1, 2$일 때만 상관관계가 있고 $l > 2$이면 상관관계가 없다는 것을 확인할 수 있다.

In [6]:

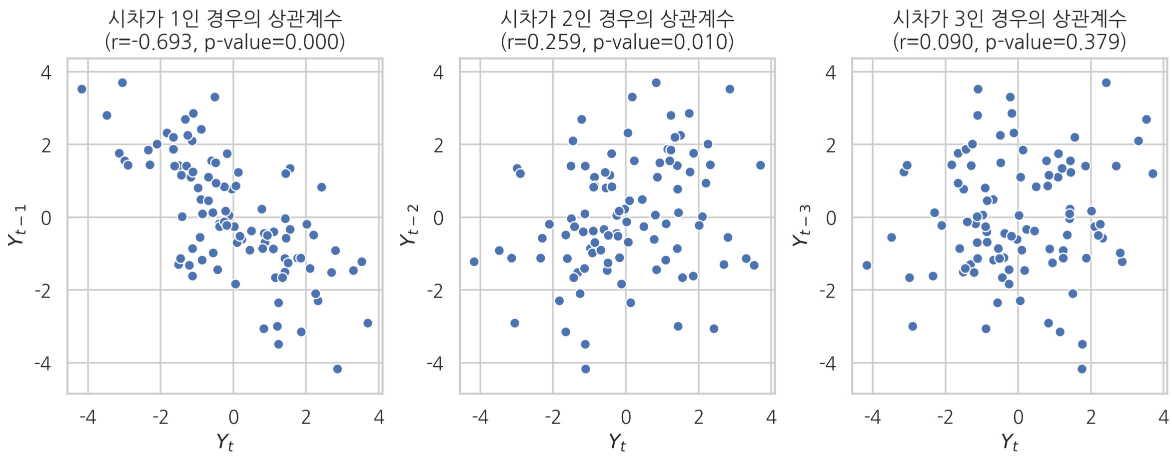
```
plt.figure(figsize=(10, 4))

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 상관계수₩n(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 상관계수₩n(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 상관계수₩n(r={0:.3f}, p-value={1:.3f})".format(r, p))

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

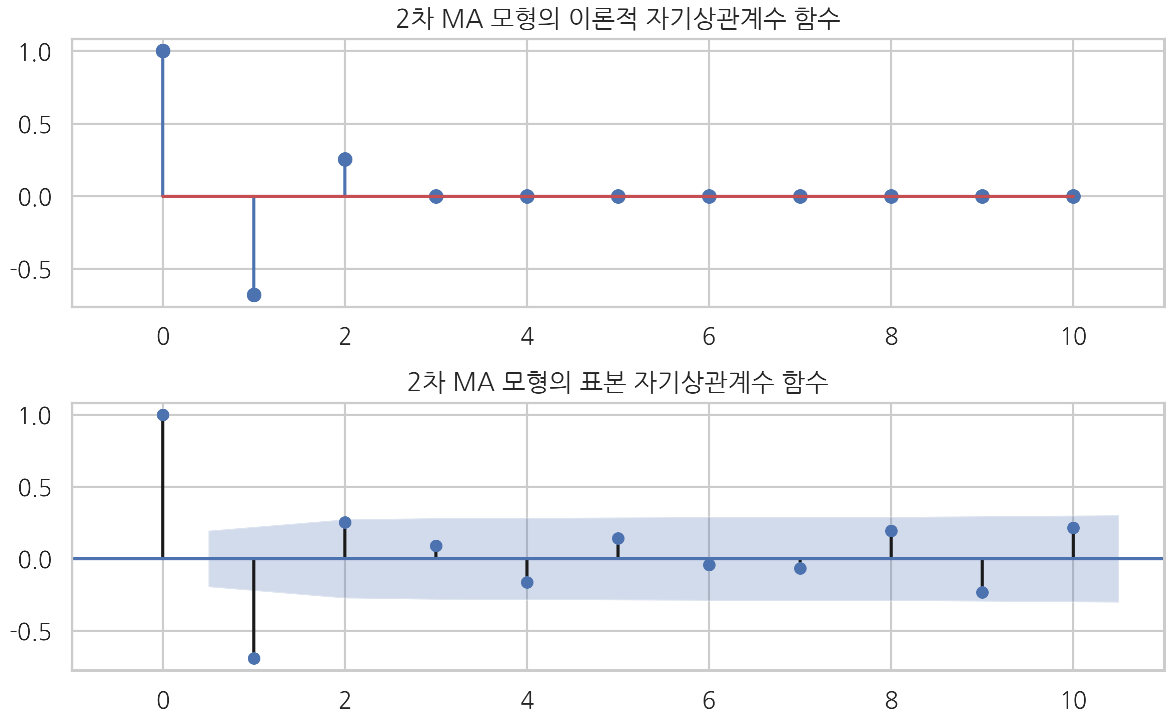


In [7]:

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

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

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



q차 MA모형

일반적인 q차의 MA모형은 MA(q)의 기댓값은 0이며 자기공분산과 자기상관계수는 다음 성질을 만족한다.

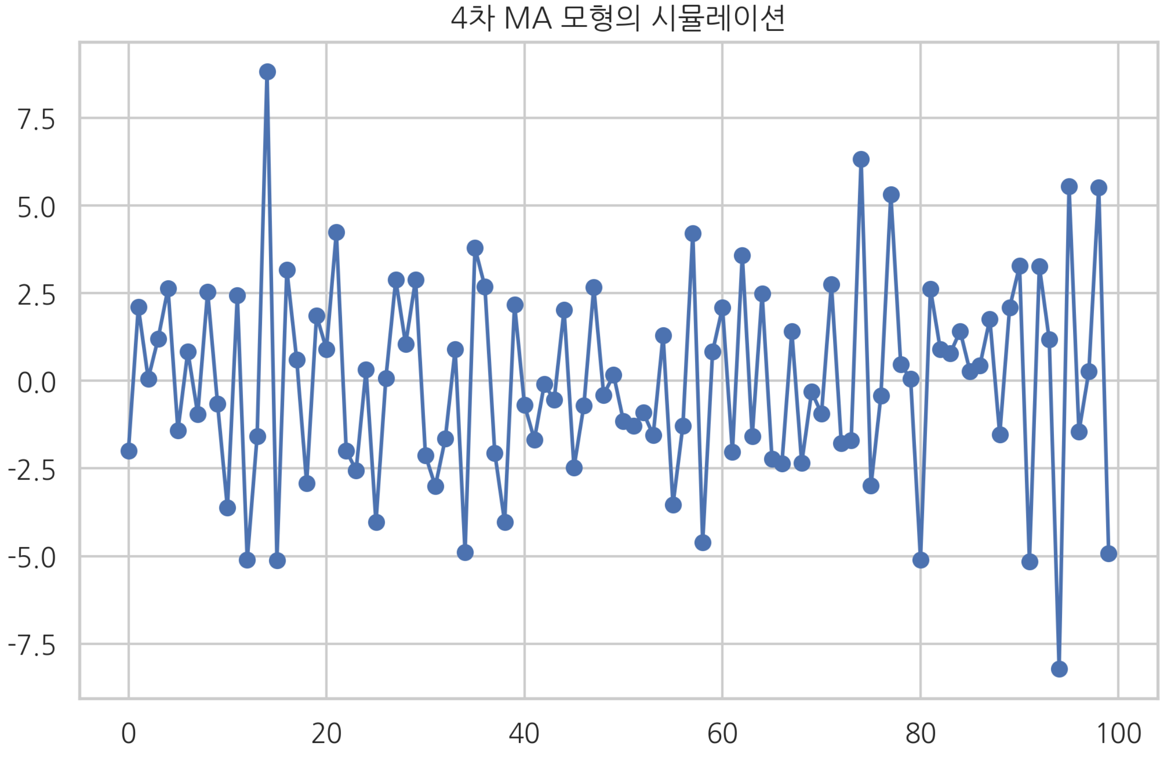
$$\gamma_0 = \sigma_\epsilon^2(1 + \theta_1^2 + \cdots + \theta_q^2)$$

$$\rho_l = \begin{cases} \dfrac{\theta_l + \theta_1\theta_{l-1} + \theta_2\theta_{l-2} + \cdots + \theta_q\theta_{l-q}}{1 + \theta_1^2 + \cdots + \theta_q^2} & \text{for } l = 1, 2, \cdots, q \\ 0 & \text{for } l > q \end{cases}$$

4차 MA모형의 시계열 데이터와 자기상관계수 함수의 예를 아래에 보였다.

In [8]:

```
np.random.seed(0)
p3 = sm.tsa.ArmaProcess([1], [1, -1, 1.6, 0.9, -1.5])
y3 = p3.generate_sample(100, burnin=10)
plt.plot(y3, 'o-')
plt.title("4차 MA모형의 시뮬레이션")
plt.show()
```



In [9]:

```
plt.subplot(211)
plt.stem(p3.acf(11))
plt.xlim(-1, 11)
plt.title("4차 MA모형의 이론적 자기상관계수 함수")

ax = plt.subplot(212)
sm.graphics.tsa.plot_acf(y3, lags=10, ax=ax)
plt.xlim(-1, 11)
plt.title("4차 MA모형의 표본 자기상관계수 함수")

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