TA section 6

May 15th, 2009 by Yang

1. Causal Expression of AR processes

Definition. An ARMA process $\{y_t\}$ such that $\phi(L)y_t = \theta(L)\varepsilon_t$ is causal if it can be expressed as

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

Example. Any MA process $y_t = \theta(L)\varepsilon_t$ is causal.

Definition. An ARMA process $\{y_t\}$ such that $\phi(L)y_t = \theta(L)\varepsilon_t$ is invertible if it can be expressed as

$$\varepsilon_t = \sum_{j=0}^{\infty} \pi_j y_{t-j}$$

Example. Any AR process $\phi(L)y_t = \varepsilon_t$ is invertible.

Theorem. Let $\{y_t\}$ be an ARMA process such that $\phi(L)y_t = \theta(L)\varepsilon_t$ where $\phi(\cdot)$ and $\theta(\cdot)$ have no common zeros. y_t is causal if and only if all zeros of $\phi(\cdot)$ are outside the unit circle, i.e., $\phi(z) \neq 0$ for any $z \in \mathbb{C}$ such that $|z| \leq 1$.

If the condition of the above theorem is satisfied, $\frac{1}{\phi(z)}$ is well defined for any $|z| \leq 1$. So we can write as

$$y_t = \frac{\theta(L)}{\phi(L)}\varepsilon_t$$

which is a causal expression of y_t.

Example. Consider $\phi(z) = 1 - \phi z$ where $\phi \in \mathbb{R}$ and $|\phi| < 1$. Zeros of $\phi(z)$ are z such that $\phi(z) = 0$, so $z = 1/\phi$. Since $|\frac{1}{\phi}| > 1$, we can obtain the causal expression of $\phi(L)y_t = \varepsilon_t$. Note that

$$(1 - \phi z)(1 + \phi z + \phi^2 z^2 + \cdots + \phi^k z^k) = 1 - \phi^{k+1} z^{k+1}$$

As $k \to \infty$, this converges to 1 for $|z| \leq 1$ since $|\phi| < 1$. So

$$\frac{1}{1 - \phi z} = 1 + \phi z + \phi^2 z^2 + \cdots = \sum_{j=0} \phi^j z^j$$

Now

$$y_t = \frac{1}{\phi(L)}\varepsilon_t = \left(\sum_{j=0}^{\infty} \phi^j L^j\right) \varepsilon_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

Example. Consider $\phi(z) = 1 - z + 0.21z^2$. Zeros of $\phi(z)$ are outside the unit circle. Suppose

$$\frac{1}{1 - z + 0.21z^2} = 1 + \psi_1 z + \psi_2 z^2 + \psi_3 z^3 + \cdots = \sum_{j=0}^{\infty} \psi_j z^j$$

This reads as

$$(1 - z + 0.21z^2)(1 + \psi_1 z + \psi_2 z^2 + \psi_3 z^3 + \cdots) = 1$$

Expanding it,

$$1 + (\psi_1 - 1)z + (\psi_2 - \psi_1 + 0.21)z^2 + (\psi_3 - \psi_2 + 0.21\psi_1)z^3 + \cdots = 1$$

Since this holds for any $|z| \leq 1$, by the method of undetermined coefficients, $\psi_1 = 1$, $\psi_2 = 0.79$, $\psi_3 = 0.58$, and so on. So $\phi(L)y_t = \varepsilon_t$ can be written as

$$y_t = \frac{1}{\phi(L)}\varepsilon_t = \varepsilon_t + \varepsilon_{t-1} + 0.79\varepsilon_{t-2} + 0.58\varepsilon_{t-3} + \cdots$$

Verify this satisfies $\phi(L)y_t = \varepsilon_t$.

In fact, we can use zeros of $\phi(L)$ to obtain a causal expression of $\phi(L)y_t = \varepsilon_t$. Take the previous example. We can factorize $\phi(z)$ into

$$\phi(z) = (1 - 0.3z)(1 - 0.7z)$$

We know that

$$\begin{aligned} \frac{1}{1 - 0.3z} &= 1 + 0.3z + 0.3^2 z^2 + 0.3^3 z^3 + \cdots \\ \frac{1}{1 - 0.7z} &= 1 + 0.7z + 0.7^2 z^3 + 0.7^3 z^3 + \cdots \end{aligned}$$

So

$$\begin{aligned} \frac{1}{\phi(z)} &= \frac{1}{1 - 0.3z} \cdot \frac{1}{1 - 0.7z} \\ &= (1 + 0.3z + 0.3^2 z^2 + 0.3^3 z^3 + \cdots)(1 + 0.7z + 0.7^2 z^3 + 0.7^3 z^3 + \cdots) \\ &= 1 + z + 0.79z^2 + 0.58z^3 + \cdots \end{aligned}$$

Hence

$$\begin{aligned} y_t &= \frac{1}{\phi(L)}\varepsilon_t \\ &= \frac{1}{1 - 0.3L} \cdot \frac{1}{1 - 0.7L}\varepsilon_t \\ &= (1 + 0.3L + 0.3^2 L^2 + 0.3^3 L^3 + \cdots)(1 + 0.7L + 0.7^2 L^3 + 0.7^3 L^3 + \cdots)\varepsilon_t \\ &= (1 + L + 0.79L^2 + 0.58L^3 + \cdots)\varepsilon_t \end{aligned}$$

This holds in general, even if zeros of $\phi(L)$ are not real numbers.

Remark. What does it mean that zeors of $\phi(L)$ are outside the unit circle? We can factorize $\phi(L)$ into $\phi(L) = (1-\alpha_1 z)\cdots(1-\alpha_p z)$ with $|\alpha_k| < 1$ for any $k = 1, \cdots, p$ if zeros of $\phi(L)$ are outside the unit circle. This ensures that $\frac{1}{1-\alpha_k z}$ is well defined for any $|z| \leq 1$.

2. Conditional Maximum Likelihood Estimation of AR processes

Consider an AR(p) model.

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

where $\varepsilon_t \sim N(0, \sigma^2)$. We have observations $(y_{-p+1}, \cdots, y_0, y_1, \cdots, y_T)$. We would like to maximize

$$L(\theta) = \prod_{t=1}^{T} f(y_t | y_{t-1}, \cdots, y_{-p+1}; \theta)$$

where $\theta = (\phi_1, \cdots, \phi_p, \sigma^2)$. Note first that by the AR(p) model,

$$f(y_t|y_{t-1}, \cdots, y_{-p+1}; \theta) = f(y_t|y_{t-1}, \cdots, y_{t-p}; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p})^2}{2\sigma^2}\right)$$

So the log-likelihood function is

$$\begin{aligned}
\mathcal{L}(\theta) &= \sum_{t=1}^{T} \log f(y_t|y_{t-1}, \cdots, y_{t-p}; \theta) \\
&= -\frac{T}{2}\log 2\pi - \frac{T}{2}\log\sigma^2 - \sum_{t=1}^{T} \frac{(y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p})^2}{2\sigma^2}
\end{aligned}$$

Differentiate it with respect to $\phi_1, \cdots, \phi_p, \sigma^2$ to obtain the set of the first order conditions. Define

$$x_t = \begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \end{pmatrix} \text{ and } \Pi = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \end{pmatrix}$$

then,

$$\mathcal{L}(\theta) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log\sigma^2 - \sum_{t=1}^{T} \frac{(y_t - \Pi' x_t)^2}{2\sigma^2}$$

So the first order conditions are

$$\begin{aligned}
\sum_{t=1}^{T} \frac{(y_t - \Pi' x_t) x_t'}{2\sigma^2} &= 0 \\
-\frac{T}{2\sigma^2} + \sum_{t=1}^{T} \frac{(y_t - \Pi' x_t)^2}{2\sigma^4} &= 0
\end{aligned}$$

and the maximum likelihood estimators are

$$\widehat{\Pi} = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1} x_t y_t \quad \text{and} \quad \widehat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \widehat{\Pi}' x_t)^2$$

3. Unconditional Maximum Likelihood Estimation of AR processes

Consider an AR(1) model for example.

$$y_t = \phi y_{t-1} + \varepsilon_t$$

where $\varepsilon_t \sim N(0, \sigma^2)$. We have observations $(y_0, \cdots, y_T)$. We need to obtain the distribution of y_0 to do the MLE. Let us guess $y_0 \sim N(\mu_y, \sigma_y^2)$. Since $y_1 = \phi y_0 + \varepsilon_t$,

$$y_1 \sim N\left(\phi\mu_y, \phi^2\sigma_y^2 + \sigma^2\right)$$

The AR(1) model is stationary, so every observation has the same mean and variance. This implies that the unconditional mean and variance of y_1 is the same with those of y_0. Therefore,

$$\begin{aligned} \phi\mu_y &= \mu_y \\ \phi^2\sigma_y^2 + \sigma^2 &= \sigma_y^2 \end{aligned}$$

which implies that

$$y_0 \sim N\left(0, \frac{\sigma^2}{1-\phi^2}\right)$$

The likelihood function is now

$$\begin{aligned} L(\theta) &= f(y_0;\theta)\prod_{t=1}^{T} f(y_t|y_{t-1};\theta) \\ &= \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi^2)}}\exp\left(-\frac{y_0^2}{2\sigma^2/(1-\phi^2)}\right)\prod_{t=1}^{T}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_t-\phi y_{t-1})^2}{2\sigma^2}\right) \end{aligned}$$

and thus

$$\begin{aligned} \mathcal{L}(\theta) &= \log f(y_0;\theta) + \sum_{t=1}^{T}\log f(y_t|y_{t-1};\theta) \\ &= \left(-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\frac{\sigma^2}{1-\phi^2} - \frac{y_0^2}{2\sigma^2/(1-\phi^2)}\right) + \left(-\frac{T}{2}\log 2\pi - \frac{T}{2}\log\sigma^2 - \sum_{t=1}^{T}\frac{(y_t-\phi y_{t-1})^2}{2\sigma^2}\right) \\ &= -\frac{T+1}{2}\log 2\pi - \frac{T+1}{2}\log\sigma^2 + \frac{1}{2}\log(1-\phi^2) - \frac{y_0^2(1-\phi^2)}{2\sigma^2} - \sum_{t=1}^{T}\frac{(y_t-\phi y_{t-1})^2}{2\sigma^2} \end{aligned}$$

Obtain the FOC's as follows.

$$\begin{aligned} -\frac{\phi}{1-\phi^2} + \frac{y_0^2\phi}{\sigma^2} + \sum_{t=1}^{T}\frac{(y_t-\phi y_{t-1})y_{t-1}}{\sigma^2} &= 0 \\ -\frac{T+1}{2\sigma^2} + \frac{y_0^2(1-\phi^2)}{2\sigma^4} + \sum_{t=1}^{T}\frac{(y_t-\phi y_{t-1})^2}{2\sigma^4} &= 0 \end{aligned}$$

The above system of equations does not have an analytic solution in general. An exception is the case where $T = 1$. This is why we do the conditional MLE with loss of some observations.

4. Optimal Linear Forecast

A forecast of y_{t+1} based on x_t is nothing but a function of x_t that predicts y_{t+1} well. The goodness of the forecast depends on cases. When we know the distribution of y_{t+1} given x_t, the optimal forecast can be the conditional mean, the conditional median, or something else, according to what risk we care about most. For example, we usually want to minimize the mean squared error of the forecast, and thus the optimal one would be the conditional mean.

The optimal linear forecast is a linear function of x_t that predicts y_{t+1} best in the sense that the mean squared error is minimized. Let $\alpha' x_t$ be the optimal linear forecast of y_{t+1}. Then α satisfies

$$E(y_{t+1} - \alpha' x_t)x_t' = 0$$

This is because we want to minimize

$$\min_a E(y_{t+1} - a' x_t)^2$$

Obtain a FOC with respect to a, then we get a desired property. Another way to prove it is as follows.

$$\begin{aligned} E(y_{t+1} - a'x_t)^2 &= E(y_{t+1} - \alpha' x_t + \alpha' x_t - a' x_t)^2 \\ &= E(y_{t+1} - \alpha' x_t)^2 + 2 \underbrace{E\left[(y_{t+1} - \alpha' x_t)(\alpha' x_t - a' x_t)\right]}_{=E(y_{t+1} - \alpha' x_t)x_t'(\alpha - a) = 0} + E(\alpha' x_t - a' x_t)^2 \end{aligned}$$

So this is minimized when $a = \alpha$.