Spring 2015 Statistics 153 (Time Series) : Lecture Fifteen

Aditya Guntuboyina

17 March 2015

1 Recap: Fitting AR models to data

Assuming that the order p is known. Carried out by invoking the function $ar()$ in R.

1. **Yule Walker or Method of Moments**: Finds the AR(p) model whose acvf equals the sample autocorrelation function at lags $0, 1, \dots, p$. Use yw for method in R.

2. **Conditional Least Squares**: Minimizes the conditional sum of squares: $\sum_{i=p+1}^{n}(x_i - \mu - \phi_1(x_{i-1} - \mu) - \cdots - \phi_p(x_{i-p} - \mu))^2$ over μ and $\phi_1, \dots, \phi_p$. And σ^2 is achieved by the average of the squared residuals. Use ols for method in R. In this method, given data $x_1, \dots, x_n$, R fits a model of the form $x_t - \bar{x} = intercept + \phi(x_{t-1} - \bar{x}) + residual$ to the data. The fitted value of intercept can be obtained by calling $\$x.intercept$. One can convert this to a model of the form $x_t = intercept + \phi x_{t-1} + residual$. Check the help page for the R function ar.ols.

3. **Maximum Likelihood**: Here one maximizes the likelihood function. The method is described below. The likelihood is relatively straightforward to write down but which requires an optimization routine to maximize. Use mle for method in R. This method is complicated.

It is usually the case that all these three methods yield similar answers. The default method in R is Yule-Walker.

2 Maximum Likelihood for fitting AR models

MLE is a very general estimation technique in statistics. One just writes down the likelihood function of the observed data in terms of the unknown parameters and estimates the parameters by the maximizers of the likelihood (or its logarithm) over the unknown parameters.

To write a likelihood, we need a distribution assumption on $\{Z_t\}$. Most common assumption is that $\{Z_t\}$ are i.i.d normal with mean 0 and variance σ_Z^2.

Let us write down the likelihood of the observed data $x_1, \dots, x_n$ for the AR(1) model under the gaussian assumption. Clearly $(x_1, \dots, x_n)$ are distributed according to the multivariate normal distribution with mean $(\mu, \dots, \mu)$ and covariance matrix $\Gamma_n := \gamma_X(i - j)$. One can write this covariance in terms of ϕ and σ_Z^2 and write down the formula for the density of the multivariate normal:

$$(2\pi)^{-n/2}|\Gamma|^{-1/2} \exp\left(-\frac{1}{2}(x - \mu)^T \Gamma^{-1}(x - \mu)\right).$$

In the AR(1) case, it is easy to simplify this likelihood. Decompose the joint density as:

$$f_{\mu,\phi,\sigma^2}(x_1,\dots,x_n) := f(x_1)f(x_2|x_1)f(x_3|x_1,x_2)\dots f(x_n|x_1,\dots,x_{n-1}).$$

Because of the gaussian assumption on $\{Z_t\}$, it is easy to see that for $i \geq 2$, the conditional distribution of x_i given $x_1, x_2, \dots, x_{i-1}$ is normal with mean $\mu + \phi(x_{i-1} - \mu)$ and variance $\sigma_X^2(1-\phi^2) = \sigma_Z^2$. Moreover x_1 is distributed as a normal with mean μ and variance $\sigma_Z^2/(1-\phi^2)$. We thus get the following likelihood:

$$L(\mu,\phi,\sigma_Z^2) := (2\pi\sigma_Z^2)^{-n/2}(1-\phi^2)^{1/2}\exp\left(-\frac{S(\mu,\phi)}{2\sigma_Z^2}\right), \tag{1}$$

where

$$S(\mu,\phi) := (1-\phi^2)(x_1-\mu)^2 + \sum_{i=2}^{n}\left(x_i - \mu - \phi(x_{i-1}-\mu)\right)^2. \tag{2}$$

This above sum of squares is called unconditional least squares.

Minimizing the likelihood (1) or its logarithm results in a non-linear optimization problem. R solves it when you choose the method *mle* in the *ar* function.

A compromise between maximum likelihood and the least squares technique (previous section) is to minimize the unconditional least squares $S(\mu,\phi)$. This also results in a non-linear optimization problem.

For the AR model, all these three estimation techniques: Yule-Walker, least squares and mle yield answers that are almost the same.

3 Asymptotic Distribution of the Estimates for AR models

The following holds for each of the Yule-Walker, Conditional Least Squares and ML estimates:

For n large, the approximate distribution of $\sqrt{n}\left(\hat{\phi}-\phi\right)$ is normal with mean 0 and variance covariance matrix $\sigma_Z^2\Gamma_p^{-1}$ where Γ_p is the $p \times p$ matrix whose (i,j)th entry is $\gamma_X(i-j)$.

3.1 Proof Sketch

Assume $\mu = 0$ for simplicity. It is easiest to work with the conditional least squares estimates. The AR(p) model is:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t.$$

We may write this model in matrix notation as:

$$X_t = \mathbb{X}_{t-1}^T\phi + Z_t$$

where $\mathbb{X}_{t-1}$ is the $p \times 1$ vector $\mathbb{X}_{t-1} = (X_{t-1}, X_{t-2}, \dots, X_{t-p})^T$ and ϕ is the $p \times 1$ vector $(\phi_1,\dots,\phi_p)^T$. The conditional least squares method minimizes the sum of squares:

$$\sum_{t=p+1}^{n}\left(X_t - \phi^T\mathbb{X}_{t-1}\right)^2$$

with respect to ϕ. The solution is:

$$\hat{\phi} = \left(\sum_{t=p+1}^{n}\mathbb{X}_{t-1}\mathbb{X}_{t-1}^T\right)^{-1}\left(\sum_{t=p+1}^{n}\mathbb{X}_{t-1}X_t\right).$$

Writing $X_t = \mathbb{X}_{t-1}^T \phi + Z_t$, we get

$$\hat{\phi} = \phi + \left(\sum_{t=p+1}^{n} \mathbb{X}_{t-1}\mathbb{X}_{t-1}^T \right)^{-1} \left(\sum_{t=p+1}^{n} \mathbb{X}_{t-1} Z_t \right).$$

As a result,

$$\sqrt{n}\left(\hat{\phi} - \phi\right) = \left(\frac{1}{n} \sum_{t=p+1}^{n} \mathbb{X}_{t-1}\mathbb{X}_{t-1}^T \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \mathbb{X}_{t-1} Z_t \right). \tag{3}$$

The following assertions are intuitive (note that $\mathbb{X}_{t-1}$ and Z_t are uncorrelated and hence independent under the gaussian assumption) and can be proved rigorously:

$$\frac{1}{n} \sum_{t=p+1}^{n} \mathbb{X}_{t-1}\mathbb{X}_{t-1}^T \to \Gamma_p \qquad \text{as } n \to \infty \text{ in probability}$$

and

$$\frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \mathbb{X}_{t-1} Z_t \to N\left(0, \sigma_Z^2 \Gamma_p\right) \qquad \text{as } n \to \infty \text{ in distribution.}$$

These results can be combined with the expression (3) to prove that $\sqrt{n}\left(\hat{\phi} - \phi\right)$ converges in distribution to a normal distribution with mean 0 and variance covariance matrix $\sigma_Z^2 \Gamma_p^{-1}$.

3.2 Special Instances

In the AR(1) case:

$$\Gamma_p = \Gamma_1 = \gamma_X(0) = \sigma_Z^2/(1-\phi^2).$$

Thus $\hat{\phi}$ is approximately normal with mean ϕ and variance $(1-\phi^2)/n$.

For AR(2), using

$$\gamma_X(0) = \frac{1-\phi_2}{1+\phi_2} \frac{\sigma_Z^2}{(1-\phi_2)^2 - \phi_1^2} \qquad \text{and} \qquad \rho_X(1) = \frac{\phi_1}{1-\phi_2},$$

we can show that $(\hat{\phi}_1, \hat{\phi}_2)$ is approximately normal with mean (ϕ_1, ϕ_2) and variance-covariance matrix is $1/n$ times

$$\begin{pmatrix} 1-\phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2 \end{pmatrix}$$

Note that the approximate variances of both $\hat{\phi}_1$ and $\hat{\phi}_2$ are the same. Observe that if we fit AR(2) model to a dataset that comes from AR(1), then the estimate of $\hat{\phi}_1$ might not change much but the standard error will be higher. We lose precision. See Example 3.34 in the book.