Estimating AR Parameters

Jared Fisher

Lecture 10a

Announcements

- Checkpoint 4 is extended to Friday, April 23 by 11:59pm PDT
- Homework 6 is extended to Friday, April 30 by 11:59pm PDT
- New Grading Policy: Homework drop may be used on a project checkpoint instead.
- My (Fisher's) office hours this week are all moved to Wednesday 8:30am-10:30am (T/Th office hours are cancelled)

Schedule

- Tuesday 4/20: Lecture on parameter estimation
- Thursday 4/22: Lecture on Intro to spectral density
- Friday 4/23: Lab on Parameter estimation and CP4 due
- Tuesday 4/27: Lecture on Spectral density part 2
- Thursday 4/29: Lecture on Extensions, Conclusion
- Friday 4/30: no formal lab but project Q&A, HW6 due
- Monday 5/10: Final Project Report and Forecasts due

Introduction

- To introduce today's topic, let's go back to the Turkey example in R

Estimating Parameters of AR(p)

Estimating AR(p)

Assume our given data $x_1, \ldots, x_n$ was generated by a causal AR(p) model with mean μ, that is,

$$(X_t - \mu) - \phi_1(X_{t-1} - \mu) - \cdots - \phi_p(X_{t-p} - \mu) = W_t.$$

with a white noise process $\{W_t\}$ with variance σ_W^2.

We are interested in finding estimates $\hat{\mu}, \hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\sigma}_W^2$ the parameters $\mu, \phi_1, \ldots, \phi_p, \sigma_W^2$.

Question for the board: How do you think we could estimate these?

Different Methods

We will look at three different methods:

1. Method of moments (Yule-Walker),
2. Least squares (LS), and
3. Maximum Likelihood (MLE).

- Today, instead of estimating a full ARMA(p,q) model, we'll first seek to understand estimation of the more simple AR(p) model.

Yule-Walker Method (Method of Moments)

Method of Moments

The method of moments is using the sample moments to estimate the true/population moments. For example, for sterotypical $X \sim N(\mu, \sigma^2)$:

1. $\hat{\mu} \stackrel{set}{=} \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$
2. $\widehat{\sigma^2} \stackrel{set}{=} s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

Yule-Walker Method

For all t we have that $E(X_t) = \mu$. Therefore, the method of moments simply estimates μ by the sample mean:

$$\hat{\mu} = \overline{x} = \frac{1}{n}\sum_{t=1}^{n} x_t.$$

For estimating the other parameters $\phi_1, \ldots, \phi_p$ and and σ_W^2, recall the Yule-Walker equations from the ARMA-ACVF Lecture

$$\gamma_X(0) - \phi_1\gamma_X(1) - \cdots - \phi_p\gamma_X(p) = \sigma_W^2, \tag{1}$$

$$\gamma_X(k) - \phi_1\gamma_X(k-1) - \cdots - \phi_p\gamma_X(k-p) = 0 \text{ for } k \geq 1. \tag{2}$$

Yule-Walker Method

- Previously, we considered solving these equations to write $\gamma_X(k)$ in terms of σ_W^2 and $\phi_1, \ldots, \phi_p$.
- But these same equations can be used to estimate σ_W^2 and $\phi_1, \ldots, \phi_p$ from the data $x_1, \ldots, x_n$:
- Definition: The Yule-Walker estimates $\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\sigma}_W^2$ for the parameters $\phi_1, \ldots, \phi_p, \sigma_W^2$ in an AR(p) model are obtained by
  1. estimate the autocovariances $\gamma_X(h)$ by the sample autocovariances $\hat{\gamma}_X(h)$.
  2. solve the above equations for the unknown parameters σ_W^2 and $\phi_1, \ldots, \phi_p$.

Yule-Walker Method

- Note that in the definition we have an infinite set of equations in Equation 2 but we only need to estimate $p+1$ parameters.
- So we will only use Equation 1 and the first p of the equations from Equation 2.
- This gives us $p+1$ equations to solve for the $p+1$ unknowns $\phi_1, \ldots, \phi_p$ and σ_W^2.
- Essentially, one is trying to find an AR(p) model whose autocovariance function equals the observed sample autocovariance function at lags $0, 1, \ldots, p$. This is why this method is called the method of moments.

Example: AR(1)

For $p = 1$ i.e., the AR(1) case, we just have the two equations:

$$\hat{\gamma}_X(0) - \phi\hat{\gamma}_X(1) = \sigma_W^2 \quad \text{and} \quad \hat{\gamma}_X(1) = \phi\hat{\gamma}_X(0).$$

This of course gives

$$\hat{\phi} = \frac{\hat{\gamma}_X(1)}{\hat{\gamma}_X(0)} = r_1 \quad \text{and} \quad \hat{\sigma_W^2} := \hat{\gamma}_X(0)\left(1 - r_1^2\right).$$

Example: AR(2)

When $p = 2$ i.e., AR(2), we get the three equations:

$$\begin{aligned}
\hat{\gamma}_X(0) - \phi_1 \hat{\gamma}_X(1) - \phi_2 \hat{\gamma}_X(2) &= \sigma_W^2 \\
\hat{\gamma}_X(1) - \phi_1 \hat{\gamma}_X(0) - \phi_2 \hat{\gamma}_X(1) &= 0 \\
\hat{\gamma}_X(2) - \phi_1 \hat{\gamma}_X(1) - \phi_2 \hat{\gamma}_X(0) &= 0
\end{aligned}$$

The last two equations can used to solve for ϕ_1 and ϕ_2 to yield:

$$\hat{\phi}_1 = \frac{r_1(1 - r_2)}{1 - r_1^2} \quad \text{and} \quad \hat{\phi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2}.$$

Plugging these values for ϕ_1 and ϕ_2 into the top equation gives an estimate for σ_W^2.

Least Squares (ols)

Definition: Least Squares

- The (conditional) least squares estimates for the parameters $\mu, \phi_1, \dots, \phi_p$ in an AR(p) model are obtained by minimizing

  $$S_c(\phi, \mu) = \sum_{i=p+1}^{n} \left(x_i - \mu - \phi_1(x_{i-1} - \mu) - \ldots - \phi_p(x_{i-p} - \mu)\right)^2 .$$

  The variance σ_W^2 is then estimated as

  $$\hat{\sigma}_W^2 = \frac{1}{n-p} S_c(\hat{\phi}, \hat{\mu}).$$

- "Conditional" as we condition on the first p values of x_i. Unconditional version shown later.

Example: AR(1)

- To minimize the LS equation, let $\beta_0 = \mu(1-\phi)$ and $\beta_1 = \phi$ and rewrite it as

$$\sum_{i=2}^{n} (x_i - \beta_0 - \beta_1 x_{i-1})^2 .$$

- Minimizing this now is exactly linear regression and the answers are given by

$$\hat{\beta}_1 = \frac{\sum_{i=2}^{n} (x_i - \bar{x}_{(2)})(x_{i-1} - \bar{x}_{(1)})}{\sum_{i=2}^{n} (x_{i-1} - \bar{x}_{(1)})^2}$$

where

$$\bar{x}_{(1)} := \frac{x_1 + \cdots + x_{n-1}}{n-1} \quad \text{and} \quad \bar{x}_{(2)} := \frac{x_2 + \cdots + x_n}{n-1}$$

and $\hat{\beta}_0 := \bar{x}_{(2)} - \hat{\beta}_1 \bar{x}_{(1)}$.

Example: AR(1)

This will give

$$\hat{\phi} = \frac{\sum_{i=2}^{n}(x_i - \bar{x}_{(2)})(x_{i-1} - \bar{x}_{(1)})}{\sum_{i=2}^{n}(x_{i-1} - \bar{x}_{(1)})^2} \quad \text{and} \quad \hat{\mu} := \frac{\bar{x}_{(2)} - \hat{\phi}\bar{x}_{(1)}}{1 - \hat{\phi}}.$$

The parameter σ_W^2 is estimated by

$$\hat{\sigma_W^2} := \frac{\sum_{i=2}^{n}\left(x_i - \hat{\mu} - \hat{\phi}(x_{i-1} - \hat{\mu})\right)^2}{n-1}.$$

It is easily seen that these estimates are very close to those obtained by the Yule-Walker method.

Maximum Likelihood

Maximum Likelihood

- To write a likelihood, we need a distribution assumption on $\{W_t\}$. Most common assumption is that $\{W_t\}$ are i.i.d normal with mean 0 and variance σ_W^2.
- Then $(x_1, \ldots, x_n)$ are distributed according to the multivariate normal distribution with mean $(\mu, \ldots, \mu)$ and covariance matrix $\Gamma_n := \gamma_X(i-j)$, which has the likelihood function

$$f_{\mu,\Gamma_n}(x_1, \ldots, x_n) = (2\pi)^{-n/2}|\Gamma|^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)^T \Gamma^{-1}(x-\mu)\right).$$

Definition

Under Gaussian noise assumption, the maximum likelihood estimator for the parameters $\mu, \phi_1, \ldots, \phi_p$ in an AR(p) model are obtained by

- Writing down covariance matrix $\Gamma_n := \gamma_X(i-j)$ as a function of $\phi_1, \ldots, \phi_p, \sigma_W^2$,

$$\Gamma_n = \Gamma_n(\phi_1, \ldots, \phi_p, \sigma_W^2)$$

- Estimate $\mu, \phi_1, \ldots, \phi_p$ by maximizing $f_{\mu, \Gamma_n(\phi_1, \ldots, \phi_p, \sigma_W^2)}(x_1, \ldots, x_n)$

Example: AR(1)

- 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 $\{W_t\}$, 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 σ_W^2.

Example: AR(1)

- Moreover x_1 is distributed as a normal with mean μ and variance $\gamma(0) = \sigma_W^2/(1-\phi^2)$. We thus get the following likelihood:

$$f_{\mu,\phi,\sigma_W^2}(x_1,\dots,x_n) := (2\pi\sigma_W^2)^{-n/2}(1-\phi^2)^{1/2}\exp\left(-\frac{S(\mu,\phi)}{2\sigma_W^2}\right),$$

where

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

- This above sum of squares is called unconditional least squares.

Example: AR(1)

- Maximizing the likelihood 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.

Summary

We have studied three different methods to estimate the parameters in an AR(p) model. Assuming that the order p is known, all three methods can be carried out in R by invoking the function *ar*().

1. **Yule Walker or Method of Moments**: Finds the AR(p) model whose acvf equals the sample autocorrelation function at lags $0, 1, \ldots, p$. Use *yw* for method in R.
2. **Least Squares**: Minimizes the 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 $\phi_1, \ldots, \phi_p$. Use *ols* for method in R. Note the default is $x_t - \bar{x} = \mathit{intercept} + \phi(x_{t-1} - \bar{x}) + \epsilon$.
3. **Maximum Likelihood**: Here one maximizes the likelihood function. Use *mle* for method in R.

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

Asymptotic Distribution of Estimates

Asymptotic Distribution of Estimates

- Recall that an estimator $\hat{\phi}$ of a parameter ϕ is a function of the data $X_1, \ldots, X_n$, that is $\hat{\phi} = \hat{\phi}(X_1, \ldots, X_n)$.
- Thus, the estimator $\hat{\phi}$ is a random variable which depends on the sample size n. The following theorem gives the approximate distribution of the estimators discussed above when n is large.

Thoerem

- Assume a causal AR(p) process $\{X_t\}$ with acvf $\gamma_X(h)$ and define the $p \times p$ matrix Γ with entries $\Gamma_{ij} = \gamma_X(i-j)$.
- Let $\hat{\phi}$ be from any of the three estimators we've discussed (Yule-Walker, least squares, or MLE).
- Then, under some general conditions on the white noise process $\{W_t\}$, with $var(W_t) = \sigma_W^2$, for n large enough, $\hat{\phi}$ is approximately multivariate normal distributed with mean $\phi = (\phi_1, \ldots, \phi_p)^\top$ and covariance matrix $n^{-1}\sigma_W^2\Gamma^{-1}$, that is

$$\sqrt{n}(\hat{\phi} - \phi) \to N(0, \sigma_W^2\Gamma^{-1}) \quad \text{as } n \to \infty.$$

- Proof is Theorem B.4 in Appendix B of TSA4e

Example: AR(1)

In the AR(1) case:

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

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

Example: AR(2)

For AR(2), using

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

we can show that $(\hat{\phi}_1, \hat{\phi}_2)$ is approximately normal with mean (ϕ_1, ϕ_2) and 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.