Unit 13: ARMA Estimation

Jeffrey Woo

Department of Statistics, University of Virginia

Spring 2020

Readings for Unit 13

Textbook chapter 3.5 (pages 113 to 121).

Last Unit

1. ARMA forecasting.
2. Prediction error.
3. Prediction interval.

This Unit

1. Method of Moments Estimation.
2. Maximum Likelihood Estimation.

Motivation

In this unit, we explore a couple of ways to estimate the parameters for ARMA models: Method of Moments (MOM) estimation and Maximum Likelihood (ML) estimation.

ARMA Estimation

Let's assume that we have an ARMA model (which is of course causal and invertible)

$$\phi(B)X_t = \theta(B)w_t,$$

where the white noise w_t has variance σ_w^2,

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p \quad \text{and} \quad \theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q.$$

Given n observations $x_1, x_2, \ldots, x_n$, we are interested in estimating the parameters $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ and σ_w^2. Initially we assume that the orders p and q are known.

Method of Moments

Let's start with the method of moments (MOM) estimation. The idea behind this is to equate population moments to sample moments and then solving for the parameters in terms of the sample moments.

Method of Moments: Toy Example

Yule-Walker Estimation for AR(p)

The method of moments works well when estimating causal AR(p) models. We consider the causal AR(p) model

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t. \tag{1}$$

Multiply both sides of (1) by x_{t-h}, and taking expectations, we obtain

$$E(x_t x_{t-h}) = \tag{2}$$

Yule-Walker Estimation for AR(p)

By definition of the autocovariance function,

We consider two cases.

Yule-Walker Estimation for AR(p)

Case I: $h \geq 1$. By causality, x_{t-h} only depends on present and past white noise terms, $w_{t-h}, w_{t-h-1}, \cdots,$. Thus, $\mathrm{E}(w_t x_{t-h}) = 0$. Then (2) becomes

$$\gamma(h) = \tag{3}$$

Yule-Walker Estimation for AR(p)

Case II: $h = 0$. Then

$$\begin{aligned} \mathrm{E}(w_t x_t) &= \\ &= \\ &= \end{aligned} \tag{4}$$

due to causality $\mathrm{E}(w_t x_{t-j}) = 0$ for $j \geq 1$.

Yule-Walker Estimation for AR(p)

Case II (continued): Subbing (4) into (2), we obtain

$$\begin{aligned} \gamma(0) &= \\ &= \end{aligned} \tag{5}$$

Yule-Walker Estimation for AR(p)

We combine (3) and (5) to obtain the **Yule-Walker equations**

$$\begin{aligned}\gamma(h) &= \phi_1\gamma(h-1) + \phi_2\gamma(h-2) + \cdots + \phi_p\gamma(h-p), \\ \sigma_w^2 &= \gamma(0) - \phi_1\gamma(1) - \phi_2\gamma(2) - \cdots - \phi_p\gamma(p).\end{aligned}$$

Yule-Walker Estimation for AR(p)

In matrix notation, the Yule-Walker equations are

$$\boldsymbol{\Gamma}_p \boldsymbol{\phi} = \boldsymbol{\gamma}_p$$

and

$$\sigma_w^2 = \gamma(0) - \boldsymbol{\phi}' \boldsymbol{\gamma}_p,$$

where $\boldsymbol{\Gamma}_p = \{\gamma(k-j)\}_{j,k=1}^{p}$ is a $p \times p$ matrix, $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_p)'$ is a $p \times 1$ vector, and $\boldsymbol{\gamma}_p = (\gamma(1), \cdots, \gamma(p))'$ is a $p \times 1$ vector.

Yule-Walker Estimation for AR(p)

Using method of moments, we replace $\gamma(h)$ with $\gamma(\hat{h})$ and solve

$$\hat{\phi} = \hat{\mathbf{\Gamma}}_p^{-1} \hat{\gamma}_p$$

and

$$\hat{\sigma}_w^2 = \hat{\gamma}(0) - \hat{\gamma}_p' \hat{\mathbf{\Gamma}}_p^{-1} \hat{\gamma}_p.$$

Yule-Walker Estimation for AR(p)

By factoring $\hat{\gamma}(0)$, the Yule-Walker estimates are

$$\hat{\phi} = \hat{\boldsymbol{R}}_p^{-1} \hat{\boldsymbol{\rho}}_p \tag{6}$$

and

$$\begin{aligned} \hat{\sigma}_w^2 &= \hat{\gamma}(0) \left[1 - \hat{\boldsymbol{\rho}}_p' \hat{\boldsymbol{R}}_p^{-1} \hat{\boldsymbol{\rho}}_p \right] \\ &= \hat{\gamma}(0) \left[1 - \hat{\boldsymbol{\rho}}_p' \hat{\phi} \right], \end{aligned} \tag{7}$$

where $\hat{\boldsymbol{R}}_p = \{\hat{\rho}(k-j)\}_{j,k=1}^{p}$ is a $p \times p$ matrix and $\hat{\boldsymbol{\rho}}_p = (\hat{\rho}(1), \cdots, \hat{\rho}(p))'$ is a $p \times 1$ vector.

Yule-Walker Estimation for AR(p)

The asymptotic behavior of the Yule-Walker estimators for causal AR(p) processes is

$$\sqrt{n}\left(\hat{\phi} - \phi\right) \xrightarrow{d} N\left(\mathbf{0}, \sigma_w^2 \mathbf{\Gamma}_p^{-1}\right) \tag{8}$$

and

$$\hat{\sigma}_w^2 \xrightarrow{p} \sigma_w^2.$$

Yule-Walker Estimation for AR(p)

The variance-covariance matrix for $\hat{\phi}$ is

$$\begin{aligned} Var(\hat{\phi}) &= \frac{\sigma^2}{n}\boldsymbol{\Gamma}_p^{-1} \\ &= \frac{\sigma^2}{n\hat{\gamma}(0)}\boldsymbol{R}_p^{-1} \end{aligned} \tag{9}$$

Simulation Example

Using the `armasim()` function, I simulated $n = 1000$ observations from the following AR(2) process

$$x_t = 1.5x_{t-1} - 0.75x_{t-2} + w_t$$

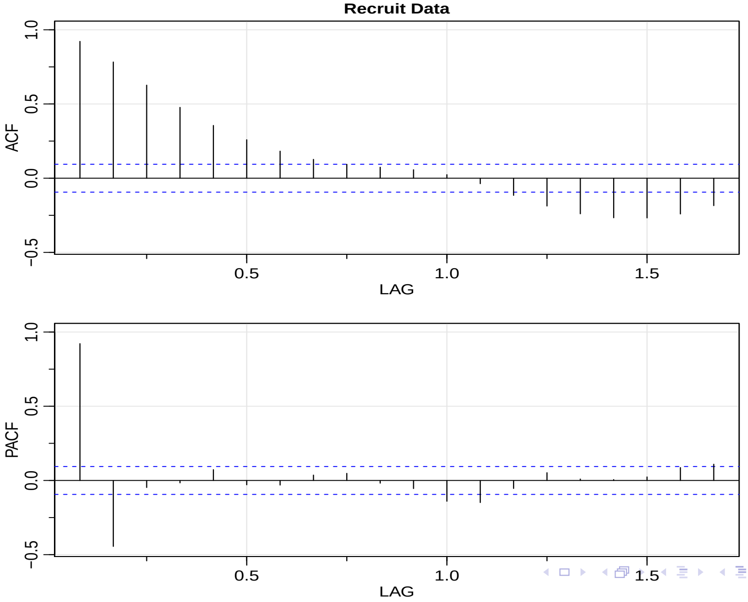
where $\sigma_w^2 = 1$. For the sample, $\hat{\gamma}(0) = 7.69697$, $\hat{\rho}(1) = 0.8456375$, and $\hat{\rho}(2) = 0.5054795$.

Simulation Example

Fish Population Example

In Unit 11, we looked at the ACF and PACF of the time series from "recruit.dat", which contains data on fish population in the central Pacific Ocean. The numbers represent the number of new fish for each month in the years 1950-1987. We decided that an AR(2) model is most appropriate.

Fish Population Example

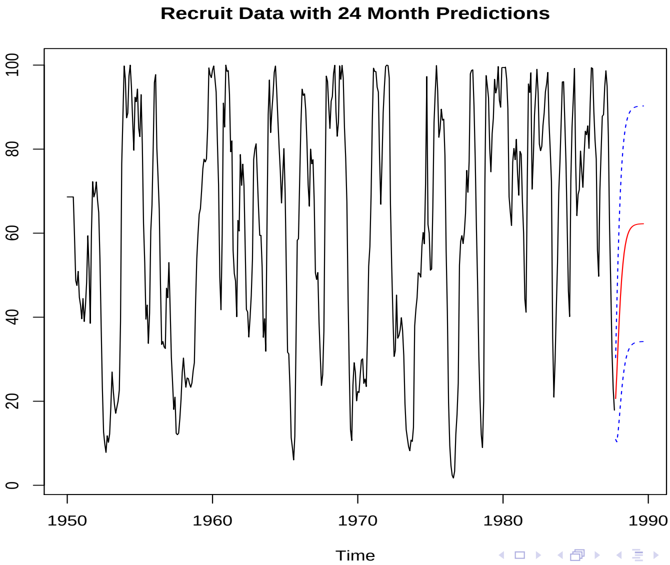


Fish Population Example

Let's check the results of fitting an AR(2) model using Yule-Walker estimation in R.

```
> rec.yw<-ar.yw(rec, order=2)
> rec.yw$x.mean
[1] 62.26278
> rec.yw$ar
[1]  1.3315874 -0.4445447
> sqrt(diag(rec.yw$asy.var.coef))
[1] 0.04222637 0.04222637
```

Fish Population Example



Fish Population Example

```
rec.pred <- predict(rec.yw, n.ahead=24)
ts.plot(rec, rec.pred$pred, col=1:2, main="Recruit Data wit
lines(rec.pred$pred - rec.pred$s, col=4, lty=2)
lines(rec.pred$pred + rec.pred$s, col=4, lty=2)
```

Method of Moments Estimation for MA(q)

Consider an MA(1) process $x_t = w_t + \theta w_{t-1}$. We know that

$$\rho(1) = -\frac{\theta}{1+\theta^2}.$$

Using method of moments, we equate $\hat{\rho}(1)$ to $\rho(1)$ and solve a quadratic equation in θ.

Method of Moments Estimation for MA(q)

The invertible solution is

$$\hat{\theta} = \frac{-1 + \sqrt{1 - 4\hat{\rho}(1)^2}}{2\hat{\rho}(1)}.$$

If $\hat{\rho}(1) = \pm 0.5$, real solutions exist, ∓ 1, neither is invertible.
If $\hat{\rho}(1) > 0.5$, no real solutions exist, the method of moments fails to yield an estimator of θ.

Method of Moments Estimation for MA(q)

For higher order MA(q) models, the method of moments quickly gets complicated. The equations are non-linear in $\theta_1, \cdots, \theta_q$, so numerical methods need to be used.

It turns out that for MA(q) models, method of moments produces poor estimates, in general.

Maximum Likelihood Estimation

To illustrate the main concept with maximum likelihood estimation, we consider the AR(1) model with nonzero mean

$$x_t = \mu + \phi(x_{t-1} - \mu) + w_t, \tag{10}$$

where $|\phi| < 1$ and $w_t \sim$ *iid* N$(0,\sigma_w^2)$.

Maximum Likelihood Estimation

We seek the likelihood

$$L(\mu, \phi, \sigma_w^2) = f_{\mu,\phi,\sigma_w^2}(x_1, x_2, \ldots, x_n). \tag{11}$$

Intuitively speaking, the likelihood function (11) $L(\mu, \phi, \sigma_w^2)$ is formed from the joint probability distribution of the observed data $x_1, x_2, \ldots, x_n$ as a function of the parameters μ, ϕ, σ_w^2.

Maximum Likelihood Estimation

For given data, you can think of the likelihood as a function of the parameters. Since we've actually observed the data $x_1, x_2, \ldots, x_n$, we can find parameters (μ, ϕ, σ_w^2) to maximize the likelihood $L(\mu, \phi, \sigma_w^2)$. This is the basic idea behind maximum likelihood estimation.

Likelihood Function

Recall that, for two random variables X and Y, the __________ $f_{X,Y}(x, y)$ of (X, Y) can be written as

$$f_{X,Y}(x, y) = f_X(x) f_{Y|X}(y|x),$$

where $f_X(x)$ is the density function of X, and $f_{Y|X}(y|x)$ is the conditional density function of Y given $X = x$. Similar idea also carries over to multivariate random variables.

Likelihood Function

Since w_t are iid normal random variables with variance σ_w^2, in the case of (10), we have

$$x_2|x_1 = \mu + \phi(x_1 - \mu) + w_2|x_1 \sim N(\mu + \phi(x_1 - \mu), \sigma_w^2)$$

and

$$f_{x_2|x_1}(x_2|x_1) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left\{ - \frac{[x_2 - \mu - \phi(x_1 - \mu)]^2}{2\sigma_w^2} \right\}.$$

Likelihood Function

Similarly,

$$f_{x_t|x_{t-1}}(x_t|x_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left\{-\frac{[x_t - \mu - \phi(x_{t-1} - \mu)]^2}{2\sigma_w^2}\right\}.$$

Likelihood Function

We have

$$
\begin{aligned}
L(\mu, \phi, \sigma_w^2) &= f_{x_1}(x_1) \times f_{x_2|x_1}(x_2|x_1) \times \cdots \times f_{x_n|x_{n-1}}(x_n|x_{n-1}) \\
&= f_{x_1}(x_1)(2\pi\sigma_w^2)^{-(n-1)/2} \exp\left\{ -\frac{S(\mu,\phi)}{2\sigma_w^2} \right\}, \quad (12)
\end{aligned}
$$

where

$$
S(\mu, \phi) = \sum_{t=2}^{n} [x_t - \mu - \phi(x_{t-1} - \mu)]^2.
$$

Likelihood Function

Note that for the AR(1) model (10), we have the causal representation $x_1 = \mu + \sum_{j=0}^{\infty} \phi^j w_{1-j}$. Since w_t are iid normal, x_1 is a normal with mean μ and variance $\sigma_w^2/(1-\phi^2)$, and

$$f_{x_1}(x_1) = \frac{1}{\sqrt{2\pi\sigma_w^2/(1-\phi^2)}} \exp\left\{ -\frac{(x_1-\mu)^2}{2\sigma_w^2/(1-\phi^2)} \right\}$$

Likelihood Function

Thus, for given data $x_1, x_2, \ldots, x_n$, we can find (μ, ϕ, σ_w^2) to maximize the likelihood $L(\mu, \phi, \sigma_w^2)$ in (12). It is worth pointing out that it is more common to consider the log-likelihood

$$\ell(\mu, \phi, \sigma^2) = \log L(\mu, \phi, \sigma^2).$$

Maximizing log-likelihoods may not always have a closed-form solution. ________________ (e.g.: Newton-Raphson, Fisher scoring) needed.

Properties of Maximum Likelihood Estimators

Maximum likelihood estimators are approximately unbiased and normally distributed, for large n.

Properties of Method of Moment Estimators

Advantages:

- Efficient (____________________) for AR(p) models.
- Nice closed-form solutions for AR(p) models.

Disadvantages:

- Poor efficiency for MA(q) models.

Properties of Maximum Likelihood Estimators

Advantages:

- Efficient.
- Even if x_t is not Gaussian, the asymptotic distribution of the MLE is the same when x_t is Gaussian.

Disadvantages:

- Difficult optimization problem.
- Need to choose a good starting point.