Parameter Estimation:

Maximum Likelihood

Week XI: Video 31

STAT 485/685, Fall 2020, SFU

Sonja Isberg

Video 31 Learning Objectives

By the end of this video, we should be able to:

- Describe the general Maximum Likelihood Estimation (MLE) approach for parameter estimation
- Apply the MLE approach to estimate ϕ, μ and σ_e^2 in an AR(1) model
- Understand the complications that arise when applying the MLE approach to more complicated ARMA models

An Introduction to Maximum Likelihood Estimation

Suppose we have a dataset $\boldsymbol{y} = \{y_1, \ldots, y_n\}$, which is assumed to be a random sample from some unknown population/model.

The model contains a set of parameters $\boldsymbol{\theta}$, which we will estimate using our dataset.

In the **maximum likelihood estimation (MLE)** approach to parameter estimation, we find the value of $\boldsymbol{\theta}$ that maximizes the **likelihood function**,

$$L(\boldsymbol{\theta}; \boldsymbol{y}) = f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$$

An Introduction to Maximum Likelihood Estimation

Suppose we have a dataset $\boldsymbol{y} = \{y_1, \ldots, y_n\}$, which is assumed to be a random sample from some unknown population/model.

The model contains a set of parameters $\boldsymbol{\theta}$, which we will estimate using our dataset.

In the **maximum likelihood estimation (MLE)** approach to parameter estimation, we find the value of $\boldsymbol{\theta}$ that maximizes the **likelihood function**,

$$L(\boldsymbol{\theta}; \boldsymbol{y}) = f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$$

Therefore, the likelihood is the joint distribution of $\boldsymbol{Y}$, evaluated on the observed dataset. It is a function of $\boldsymbol{\theta}$.

Maximizing the likelihood gives us the value of $\boldsymbol{\theta}$ that is "most likely to have generated the sample".

Oftentimes, the **log-likelihood** $\ell(\boldsymbol{\theta}; \boldsymbol{y}) = \log L(\boldsymbol{\theta}; \boldsymbol{y})$ is maximized instead.

Maximum Likelihood Estimation for AR(1)

Recall the AR(1) model with a non-zero constant mean:

$$(Y_t - \mu) = \phi(Y_{t-1} - \mu) + e_t$$

The likelihood function will be a function of the unknown parameters ϕ, μ and σ_e^2, given the observations $\{Y_1, \ldots, Y_n\}$.

In order to write out the joint pdf of $\{Y_1, \ldots, Y_n\}$, we need to choose a distribution for the random part of our model.

An often safe assumption is that the white noise terms e_t are iid $\mathcal{N}(0, \sigma_e^2)$.

Maximum Likelihood Estimation for AR(1) (cont'd)

Under this normality assumption, it can be shown (see pg. 159) that

$$\begin{aligned} L(\phi, \mu, \sigma_e^2) &= f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n; \phi, \mu, \sigma_e^2) \\ &= (2\pi\sigma_e^2)^{-n/2}(1-\phi^2)^{1/2} \exp\left[-\frac{1}{2\sigma_e^2} S(\phi, \mu)\right] \end{aligned}$$

where

$$S(\phi, \mu) = \sum_{t=2}^{n} [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2 + (1-\phi^2)(Y_1 - \mu)^2$$

Here, $S(\phi, \mu)$ is called the **unconditional sum-of-squares function**.

Maximum Likelihood Estimation for AR(1) (cont'd)

Under this normality assumption, it can be shown (see pg. 159) that

$$\begin{aligned} L(\phi, \mu, \sigma_e^2) &= f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n; \phi, \mu, \sigma_e^2) \\ &= (2\pi\sigma_e^2)^{-n/2}(1-\phi^2)^{1/2} \exp\left[-\frac{1}{2\sigma_e^2} S(\phi, \mu)\right] \end{aligned}$$

where

$$S(\phi, \mu) = \sum_{t=2}^{n} [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2 + (1-\phi^2)(Y_1 - \mu)^2$$

Here, $S(\phi, \mu)$ is called the **unconditional sum-of-squares function**.

We usually focus on maximizing the log-likelihood instead:

$$\begin{aligned} \ell(\phi, \mu, \sigma_e^2) &= \log L(\phi, \mu, \sigma_e^2) \\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma_e^2) + \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma_e^2} S(\phi, \mu) \end{aligned}$$

Maximum Likelihood Estimation for AR(1) (cont'd)

$$\begin{aligned}\ell(\phi, \mu, \sigma_e^2) &= \log L(\phi, \mu, \sigma_e^2) \\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma_e^2) + \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma_e^2}S(\phi, \mu)\end{aligned}$$

Then, temporarily suppose ϕ and μ are known, and optimize with respect to σ_e^2:

$$\frac{\partial \ell(\phi, \mu, \sigma_e^2)}{\partial \sigma_e^2} = -\frac{n}{2\sigma_e^2} + \frac{1}{2(\sigma_e^2)^2}S(\phi, \mu) = 0$$

$$-n + \frac{S(\phi, \mu)}{\hat{\sigma}_e^2} = 0$$

$$\hat{\sigma}_e^2 = \frac{S(\phi, \mu)}{n} \approx \frac{S(\phi, \mu)}{n-2} \quad \text{(less bias)}$$

Let's first estimate ϕ and μ, and then we will be able to obtain $\hat{\sigma}_e^2$!

Maximum Likelihood Estimation for AR(1) (cont'd)

Note:

$$S(\phi, \mu) = \sum_{t=2}^{n} [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2 + (1 - \phi^2)(Y_1 - \mu)^2$$

Recall the *conditional sum-of-squares function* from Video 30:

$$S_c(\phi, \mu) = \sum_{t=2}^{n} [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2$$

Therefore:

$$S(\phi, \mu) = S_c(\phi, \mu) + (1 - \phi^2)(Y_1 - \mu)^2$$

Maximum Likelihood Estimation for AR(1) (cont'd)

Note:

$$S(\phi, \mu) = \sum_{t=2}^{n} [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2 + (1 - \phi^2)(Y_1 - \mu)^2$$

Recall the *conditional sum-of-squares function* from Video 30:

$$S_c(\phi, \mu) = \sum_{t=2}^{n} [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2$$

Therefore:

$$S(\phi, \mu) = S_c(\phi, \mu) + (1 - \phi^2)(Y_1 - \mu)^2$$

So, usually (except when ϕ is near the stationarity boundary of ± 1):

$$-\ell(\phi, \mu, \sigma_e^2) \approx S(\phi, \mu) \approx S_c(\phi, \mu)$$

Therefore, the MLEs of ϕ and μ will be approximately equal to the LSEs!

Maximum Likelihood Estimation for AR(1) (cont'd)

Note:

$$S(\phi,\mu)=\sum_{t=2}^{n}[(Y_t-\mu)-\phi(Y_{t-1}-\mu)]^2+(1-\phi^2)(Y_1-\mu)^2$$

Recall the *conditional sum-of-squares function* from Video 30:

$$S_c(\phi,\mu)=\sum_{t=2}^{n}[(Y_t-\mu)-\phi(Y_{t-1}-\mu)]^2$$

Therefore:

$$S(\phi,\mu)=S_c(\phi,\mu)+(1-\phi^2)(Y_1-\mu)^2$$

So, usually (except when ϕ is near the stationarity boundary of ± 1):

$$-\ell(\phi,\mu,\sigma_e^2)\approx S(\phi,\mu)\approx S_c(\phi,\mu)$$

Therefore, the MLEs of ϕ and μ will be approximately equal to the LSEs!

Once we have these $\hat{\phi}$ and $\hat{\mu}$, then: $\hat{\sigma}_e^2=S(\hat{\phi},\hat{\mu})/(n-2)$.

Additional Comments

MLE estimation for other models is a lot more complicated:

- The derivation of the likelihood function is a lot more complicated
- The resulting function is a lot more complex, and more difficult to maximize

Additional Comments

MLE estimation for other models is a lot more complicated:

- The derivation of the likelihood function is a lot more complicated
- The resulting function is a lot more complex, and more difficult to maximize

Another approach to parameter estimation, which often yields similar results to MLE and LSE, is **unconditional least squares estimation**:

- This involves minimizing $S(\phi, \mu)$, instead of $S_c(\phi, \mu)$
- However, because of the extra term $(1 - \phi^2)(Y_1 - \mu)^2$, it is much harder to set $\partial S / \partial \phi = 0$ and $\partial S / \partial \mu = 0$ and solve for ϕ and μ. This often has to be done numerically

Further Information & Interesting Reading

For large n, MLEs are approximately unbiased and approximately normally distributed.

They also have some nice asymptotic properties that are easily obtained.

Section 7.4 of the textbook lists some of these results.

Section 7.5 uses these results to compare the performance of the three types of estimators we have learned about.

Section 7.6 describes how *bootstrapping* can be used to obtain these types of properties when there are no analytical results to inform us.

Final Comments

That's all for now!

In this video, we've learned how to estimate the parameters of an AR(1) model using the Maximum Likelihood Estimation method.

We also discussed how this approach works in general, and some of the complications that arise when applying it to more complicated ARMA models.

In Week 12 in STAT 485/685: Model diagnostics, and Forecasting!

Thank you!

References:

[1] Cryer, J. D., & Chan, K. S. (2008). *Time series analysis: with applications in R.* Springer Science and Business Media.

[2] Chan, K. S., & Ripley, B. (2020). TSA: Time Series Analysis. R package version 1.2.1.