Lecture 12-13-14

1 Basic Definition and a simple example

Let $x = (x_1, x_2, \ldots, x_T)$ denote some time series data set. Assume we have a statistical model for X—parameterized by θ—and characterized by a p.d.f. $f(x|\theta)$.

Definition (Likelihood): The likelihood of data x at parameter θ is defined as:

$$\mathcal{L}(\theta; x) \equiv f(x|\theta).$$

The likelihood function (corresponding to x) refers to the map $\mathcal{L}(\cdot; x) : \Theta \to \mathbb{R}_+$.

Intuitively, $\mathcal{L}(\cdot; x)$ tells us how likely is the data that we observe at each different value of θ. The log-likelihood function is simply defined as the natural logarithm of the likelihood: $\ln f(x|\theta)$.

Definition (Maximum Likelihood Estimator): The Maximum Likelihood (ML) Estimator of θ given data x is defined as any value of $\theta \in \Theta$ that maximizes the likelihood; that is:

$$\widehat{\theta} \in \text{argmax}_\theta \, \mathcal{L}(\theta; x).$$

ML estimation is popular because it generates consistent estimators for the parameters of interest (under very general conditions).

Example for i.i.d. data: Suppose that we observe n i.i.d. realizations $(y_1, y_2, \ldots, y_n)$ from the following statistical model $y_i \in \{0, 1\}$ and

$$P_\theta(Y_i = 1) = \theta.$$

In this case we can show that the likelihood function can be written as:

$$\mathcal{L}(\theta, (y_1, \ldots, y_n)) = \theta^{n_1}(1-\theta)^{n_0}$$

where n_1 is the number of 1s in the sample of size n and n_0 is the number of zeros. The maximum likelihood estimator of θ is the value that solves the problem:

$$\max_{\theta} \theta^{n_1}(1-\theta)^{n_0}.$$

This is the same as maximizing:

$$\max_{\theta} \ln\left(\theta^{n_1}(1-\theta)^{n_0}\right)$$

since the logarithm is a monotone function. Expanding the logarithm we get:

$$\max_{\theta} n_1 \ln\theta + n_0 \ln(1-\theta)$$

which has the first-order conditions:

$$\frac{n_1}{\theta} + \frac{n_0}{1-\theta} = 0.$$

Solving for θ in the equation above we get that:

$$\widehat{\theta}_{\mathrm{ML}} = \frac{n_1}{n_1 + n_0}.$$

We can use the LLN for independent data discussed in class to show that

$$\widehat{\theta}_{\mathrm{ML}} \xrightarrow{p} E_{\theta}[Y_i = 1] = \theta.$$

2 ML Estimation for the MA(1)/AR(1) model

We will need two things to compute a Maximum Likelihood estimator: a likelihood function and some computer program to optimize its value. We will try to get a sense of what these requirements entail in the context of the MA(1)/AR(1) models.

2.1 Gaussian MA(1)

We will first discuss the Maximum Likelihood estimation of the MA(1) model:

$$X_t = \mu + \epsilon_t + \theta\epsilon_{t-1} \tag{1}$$

In order to get a likelihood we will assume that the shocks are i.i.d. and:

$$\epsilon_t \sim N(0, \sigma^2).$$

The parameters of this model are (μ, θ, σ^2). It is not difficult to verify that for any t:

$$\begin{pmatrix} X_{t-1} \\ X_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \sigma^2 \begin{pmatrix} 1+\theta^2 & \theta \\ \theta & 1+\theta^2 \end{pmatrix} \right).$$

Note also that X_t is independent of the collections $X_{t-2}, \ldots X_1$ and $X_{t+2}, \ldots X_T$. This means the likelihood of $(X_1, \ldots, X_T)$ at parameters (μ, θ, σ^2) is given by:

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_{T-1} \\ X_T \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \\ \mu \end{pmatrix}, \sigma^2 \begin{pmatrix} 1+\theta^2 & \theta & \ldots & 0 & 0 \\ \theta & 1+\theta^2 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1+\theta^2 & \theta \\ 0 & 0 & \ldots & \theta & 1+\theta^2 \end{pmatrix} \right). \tag{2}$$

The formula above implies that in order to compute the likelihood function of the MA(1) model, all we need is the p.d.f of a multivariate normal distribution where the vector of means and the covariance matrix are functions of μ, θ, and σ^2.[1] We can easily write a Matlab function that evaluates the likelihood of the MA model by using the matlab command `y = mvnpdf(R,MU,SIGMA)`.[2]

[1] See wikipedia for the formula of the p.d.f. of a multivariate normal random vector.

[2] This command evaluates the multivariate normal p.d.f. for a row vector R using the vector of means MU and the covariance matrix SIGMA. This command only works if X is

2.2 Tricks to evaluate the likelihood of an MA(1) model

Oftentimes, there are some tricks to evaluate the likelihood in time series models. Note that the definition of conditional density implies that:

$$f(X_1, X_2, \ldots, X_T) = f(X_T | X_{T-1}, X_{T-2}, \ldots, X_1) f(X_1, X_2, \ldots, X_{T-1})$$

Applying the definition one conditional density one more time we have that

$$f(X_1, X_2, \ldots, X_T)$$

equals

$$f(X_T | X_{T-1}, X_{T-2}, \ldots, X_1) f(X_{T-1} | X_{T-2}, X_{T-3}, \ldots, X_1) f(X_1, X_2, \ldots, X_{T-2}).$$

Doing this recursively we get that:

$$f(X_1, X_2, \ldots, X_T) = \Pi_{t=2}^{T} f(X_t | X_{t-1}, X_{t-2}, \ldots, X_1).$$

Why is this useful? Well, MA(1) and AR(1) models make simplifications on the conditional distributions $f(X_t | X_{t-1}, \ldots, X_1)$. Such simplifications often lead to simpler expressions for the likelihood in (2). For instance, note that for an MA(1) model

$$X_t | X_{t-1}, X_{t-2}, \ldots, X_1 \sim X_t | X_{t-1} \sim \mathcal{N}(\mu + \rho(X_{t-1} - \mu) \, , \, (1 - \rho^2)\gamma_0),$$

with $\rho \equiv \gamma_1 / \gamma_0, \gamma_1 \equiv \theta\sigma^2, \gamma_0 \equiv \sigma^2(1 + \theta^2)$. This means that we can easily write the log-likelihood of $(X_2, X_3, \ldots, X_T)$

$$\ln f(X_2, \ldots, X_T | (\mu, \theta, \sigma^2))$$

as

$$c - \frac{T-1}{2} \ln(1 - \rho^2) - \frac{T-1}{2} \ln(\gamma_0) - \frac{1}{2(1-\rho^2)\gamma_0} \sum_{t=2}^{T} (X_t - \mu - \rho(X_{t-1} - \mu))^2. \quad (3)$$

You might not realize it now, but evaluating (3) is computationally less demand-

not very large.

ing than evaluating (2). I will now argue, that the trick is also useful to derive a closed-form expression for the ML estimators (which are typically not available).

Note that there is a one-to-one mapping between (μ, θ, σ^2) and (μ, ρ, γ_0) (by virtue of the results presented in Lecture 10).

Assume, for the sake of exposition, that we have $\mu = 0$ and let us derive the ML estimators of (ρ, γ_0). Note that the first-order conditions for γ_0 are given by:

$$-\frac{T-1}{2\gamma_0} + \frac{1}{2(1-\rho^2)\gamma_0^2} \sum_{t=2}^{T} (X_t - \rho X_{t-1})^2 = 0$$

This implies that the ML estimators of $(\widehat{\rho}, \widehat{\gamma}_0)$—whatever they are—need to satisfy:

$$\widehat{\gamma}_0 = \frac{1}{(1-\widehat{\rho}^2)} \frac{1}{T-1} \sum_{t=2}^{T} (X_t - \widehat{\rho} X_{t-1})^2. \tag{4}$$

I will now use the condition above to figure out the expression for $\widehat{\rho}$. This derivative is a little bit more complicated, but it can be shown that the first order condition is given by:

$$\frac{T-1}{2} \frac{2\widehat{\rho}}{1-\widehat{\rho}^2} - \frac{2\widehat{\rho}(T-1)}{2\widehat{\gamma}_0(1-\widehat{\rho}^2)^2} \frac{1}{T-1} \sum_{t=2}^{T} (X_t - \widehat{\rho} X_{t-1})^2 + \frac{2}{2(1-\widehat{\rho}^2)\widehat{\gamma}_0} \sum_{t=2}^{T} (X_t - \widehat{\rho} X_{t-1}) X_{t-1}$$

equals 0. This means that the F.O.C. is satisfied if and only if:

$$\frac{(T-1)\widehat{\rho}}{1-\widehat{\rho}^2} - \frac{\widehat{\rho}(T-1)}{1-\widehat{\rho}^2} + \frac{1}{(1-\widehat{\rho}^2)\widehat{\gamma}_0} \sum_{t=2}^{T} (X_t - \widehat{\rho} X_{t-1}) X_{t-1} = 0.$$

Which implies that:

$$\widehat{\rho} = \sum_{t=2}^{T} X_t X_{t-1} \Big/ \sum_{t=2}^{T} X_{t-1}^2. \tag{5}$$

How do we get the estimators of θ and σ^2? Here is how:

1. Compute $\widehat{\rho}$ from (5) and $\widehat{\gamma}_0$ from (4).

2. Define $\widehat{\gamma}_1$ as $\rho\widehat{\gamma}_0$

3. Use the derivations in Lecture 10 to get $\widehat{\theta}$ and $\widehat{\sigma}^2$ from $\widehat{\gamma}_0$ and $\widehat{\gamma}_1$.

Compare the output with Matlab's `arima` package. Are they the same?

2.3 Gaussian AR(1) model

We will now derive the likelihood of the Gaussian AR(1) model:

$$X_t = \mu + \phi X_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2)$$

It is not difficult to show, that the Gaussian AR(1) model has the following causal linear process representation:

$$X_t = \frac{\mu}{1-\phi} + \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}$$

which suggests that:

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_{T-1} \\ X_T \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \frac{\mu}{1-\phi} \\ \frac{\mu}{1-\phi} \\ \vdots \\ \frac{\mu}{1-\phi} \\ \frac{\mu}{1-\phi} \end{pmatrix}, \frac{\sigma^2}{1-\phi^2} \begin{pmatrix} 1 & \phi & \dots & \phi^{T-2} & \phi^{T-1} \\ \phi & 1 & \dots & \phi^{T-3} & \phi^{T-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi^{T-2} & \phi^{T-1} & \dots & 1 & \phi \\ \phi^{T-1} & \phi^{T-2} & \dots & \phi & 1 \end{pmatrix} \right). \quad (6)$$

2.4 Tricks to evaluate the likelihood of the AR(1) model

We use the same trick that we used for the MA(1) model. Note that the AR(1) model implies that:

$$X_t | X_{t-1} \sim \mathcal{N}(\mu + \phi X_t, \sigma^2).$$

The log-likelihood of $(X_2, X_3, \dots, X_T)$ is then given by:

$$c - \frac{T-1}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=2}^{T} (X_t - \mu - \phi X_{t-1})^2.$$

Once again, and for the sake of exposition, we set $\mu = 0$. The F.O.C for ϕ in this problem imply that:

$$-\frac{1}{2\sigma^2} \sum_{t=2}^{T} 2(X_t - \widehat{\phi} X_{t-1}) X_{t-1} = 0,$$

which implies that

$$\widehat{\phi} = \sum_{t=2}^{T} X_t X_{t-1} / \sum_{t=2}^{T} X_{t-1}^2.$$

The first order conditions for σ^2 give:

$$-\frac{T-1}{2\widehat{\sigma}^2} - \frac{1}{2\widehat{\sigma}^4}(X_t - \widehat{\phi}X_{t-1})^2 = 0,$$

which imply that:

$$\widehat{\sigma}^2 = \frac{1}{T-1}\sum_{t=2}^{T}(X_t - \widehat{\phi}X_{t-1})^2$$