Introduction to Time Series Analysis. Lecture 14.

Last lecture: Yule-Walker estimation

1. Maximum likelihood estimation
2. Large-sample distribution of MLE

Parameter estimation: Maximum likelihood estimator

One approach:

Assume that $\{X_t\}$ is Gaussian, that is, $\phi(B)X_t = \theta(B)W_t$, where W_t is i.i.d. Gaussian.

Choose ϕ_i, θ_j to maximize the *likelihood*:

$$L(\phi, \theta, \sigma^2) = f(X_1, \ldots, X_n),$$

where f is the joint (Gaussian) density for the given ARMA model.

(c.f. choosing the parameters that maximize the probability of the data.)

Maximum likelihood estimation

Suppose that $X_1, X_2, \ldots, X_n$ is drawn from a zero mean Gaussian ARMA(p,q) process. The likelihood of parameters $\phi \in \mathbb{R}^p$, $\theta \in \mathbb{R}^q$, $\sigma_w^2 \in \mathbb{R}_+$ is defined as the density of $X = (X_1, X_2, \ldots, X_n)'$ under the Gaussian model with those parameters:

$$L(\phi, \theta, \sigma_w^2) = \frac{1}{(2\pi)^{n/2} \left|\Gamma_n\right|^{1/2}} \exp\left(-\frac{1}{2} X' \Gamma_n^{-1} X\right),$$

where $|A|$ denotes the determinant of a matrix A, and Γ_n is the variance/covariance matrix of X with the given parameter values.

The maximum likelihood estimator (MLE) of ϕ, θ, σ_w^2 maximizes this quantity.

Parameter estimation: Maximum likelihood estimator

Advantages of MLE:

Efficient (low variance estimates).

Often the Gaussian assumption is reasonable.

Even if $\{X_t\}$ is not Gaussian, the asymptotic distribution of the estimates $(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2)$ is the same as the Gaussian case.

Disadvantages of MLE:

Difficult optimization problem.

Need to choose a good starting point (often use other estimators for this).

Preliminary parameter estimates

Yule-Walker for AR(p): Regress X_t onto $X_{t-1}, \ldots, X_{t-p}$.

Durbin-Levinson algorithm with γ replaced by $\hat{\gamma}$.

Yule-Walker for ARMA(p,q): Method of moments. Not efficient.

Innovations algorithm for MA(q): with γ replaced by $\hat{\gamma}$.

Hannan-Rissanen algorithm for ARMA(p,q):

1. Estimate high-order AR.
2. Use to estimate (unobserved) noise W_t.
3. Regress X_t onto $X_{t-1}, \ldots, X_{t-p}, \hat{W}_{t-1}, \ldots, \hat{W}_{t-q}$.
4. Regress again with improved estimates of W_t.

Recall: Maximum likelihood estimation

Suppose that $X_1, X_2, \ldots, X_n$ is drawn from a zero mean Gaussian ARMA(p,q) process. The likelihood of parameters $\phi \in \mathbb{R}^p$, $\theta \in \mathbb{R}^q$, $\sigma_w^2 \in \mathbb{R}_+$ is defined as the density of $X = (X_1, X_2, \ldots, X_n)'$ under the Gaussian model with those parameters:

$$L(\phi, \theta, \sigma_w^2) = \frac{1}{(2\pi)^{n/2} \left|\Gamma_n\right|^{1/2}} \exp\left(-\frac{1}{2} X' \Gamma_n^{-1} X \right),$$

where $|A|$ denotes the determinant of a matrix A, and Γ_n is the variance/covariance matrix of X with the given parameter values.

The maximum likelihood estimator (MLE) of ϕ, θ, σ_w^2 maximizes this quantity.

Maximum likelihood estimation

We can simplify the likelihood by expressing it in terms of the *innovations*.

Since the innovations are linear in previous and current values, we can write

$$\underbrace{\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}}_{X} = C \underbrace{\begin{pmatrix} X_1 - X_1^0 \\ \vdots \\ X_n - X_n^{n-1} \end{pmatrix}}_{U}$$

where C is a lower triangular matrix with ones on the diagonal.
Take the variance/covariance of both sides to see that

$$\Gamma_n = CDC' \qquad \text{where } D = \mathrm{diag}(P_1^0, \ldots, P_n^{n-1}).$$

Maximum likelihood estimation

Thus, $|\Gamma_n| = |C|^2 P_1^0 \cdots P_n^{n-1} = P_1^0 \cdots P_n^{n-1}$ and

$$X'\Gamma_n^{-1}X = U'C'\Gamma_n^{-1}CU = U'C'C^{-T}D^{-1}C^{-1}CU = U'D^{-1}U.$$

So we can rewrite the likelihood as

$$\begin{aligned} L(\phi, \theta, \sigma_w^2) &= \frac{1}{\left((2\pi)^n P_1^0 \cdots P_n^{n-1}\right)^{1/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (X_i - X_i^{i-1})^2 / P_i^{i-1} \right) \\ &= \frac{1}{\left((2\pi\sigma_w^2)^n r_1^0 \cdots r_n^{n-1}\right)^{1/2}} \exp\left(-\frac{S(\phi, \theta)}{2\sigma_w^2} \right), \end{aligned}$$

where $r_i^{i-1} = P_i^{i-1}/\sigma_w^2$ and

$$S(\phi, \theta) = \sum_{i=1}^{n} \frac{\left(X_i - X_i^{i-1}\right)^2}{r_i^{i-1}}.$$

Maximum likelihood estimation

The log likelihood of ϕ, θ, σ_w^2 is

$$\begin{aligned} l(\phi, \theta, \sigma_w^2) &= \log(L(\phi, \theta, \sigma_w^2)) \\ &= -\frac{n}{2}\log(2\pi\sigma_w^2) - \frac{1}{2}\sum_{i=1}^{n}\log r_i^{i-1} - \frac{S(\phi, \theta)}{2\sigma_w^2}. \end{aligned}$$

Differentiating with respect to σ_w^2 shows that the MLE $(\hat{\phi}, \hat{\theta}, \hat{\sigma}_w^2)$ satisfies

$$\frac{n}{2\hat{\sigma}_w^2} = \frac{S(\hat{\phi}, \hat{\theta})}{2\hat{\sigma}_w^4} \qquad \Leftrightarrow \qquad \hat{\sigma}_w^2 = \frac{S(\hat{\phi}, \hat{\theta})}{n},$$

and $\hat{\phi}, \hat{\theta}$ minimize $\qquad \log\left(\frac{S(\hat{\phi}, \hat{\theta})}{n}\right) + \frac{1}{n}\sum_{i=1}^{n}\log r_i^{i-1}.$

Summary: Maximum likelihood estimation

The MLE $(\hat{\phi}, \hat{\theta}, \hat{\sigma}_w^2)$ satisfies

$$\hat{\sigma}_w^2 = \frac{S(\hat{\phi}, \hat{\theta})}{n},$$

and $\hat{\phi}, \hat{\theta}$ minimize $$\log\left(\frac{S(\hat{\phi}, \hat{\theta})}{n}\right) + \frac{1}{n}\sum_{i=1}^{n} \log r_i^{i-1},$$

where $r_i^{i-1} = P_i^{i-1}/\sigma_w^2$ and

$$S(\phi, \theta) = \sum_{i=1}^{n} \frac{\left(X_i - X_i^{i-1}\right)^2}{r_i^{i-1}}.$$

Maximum likelihood estimation

Minimization is done numerically (e.g., Newton-Raphson).

Computational simplifications:

• *Unconditional least squares.* Drop the $\log r_i^{i-1}$ terms.

• *Conditional least squares.* Also approximate the computation of x_i^{i-1} by dropping initial terms in S. e.g., for AR(2), all but the first two terms in S depend linearly on ϕ_1, ϕ_2, so we have a least squares problem.

The differences diminish as sample size increases. For example, $P_t^{t-1} \to \sigma_w^2$ so $r_t^{t-1} \to 1$, and thus $n^{-1} \sum_i \log r_i^{i-1} \to 0$.

Maximum likelihood estimation: Confidence intervals

For an ARMA(p,q) process, the MLE and un/conditional least squares estimators satisfy

$$\begin{pmatrix} \hat{\phi} \\ \hat{\theta} \end{pmatrix} - \begin{pmatrix} \phi \\ \theta \end{pmatrix} \sim AN\left(0, \frac{\sigma_w^2}{n} \begin{pmatrix} \Gamma_{\phi\phi} & \Gamma_{\phi\theta} \\ \Gamma_{\theta\phi} & \Gamma_{\theta\theta}, \end{pmatrix}^{-1}\right),$$

where $\begin{pmatrix} \Gamma_{\phi\phi} & \Gamma_{\phi\theta} \\ \Gamma_{\theta\phi} & \Gamma_{\theta\theta}, \end{pmatrix} = \mathrm{Cov}((X,Y),(X,Y)),$

$$X = (X_1, \ldots, X_p)' \qquad \phi(B)X_t = W_t,$$

$$Y = (Y_1, \ldots, Y_p)' \qquad \theta(B)Y_t = W_t.$$

Introduction to Time Series Analysis. Lecture 14.

1. Maximum likelihood estimation: Gaussian model.

$$\hat{\sigma}_w^2 = \frac{S(\hat{\phi}, \hat{\theta})}{n},$$

and $\hat{\phi}, \hat{\theta}$ minimize

$$\log \left(\frac{S(\hat{\phi}, \hat{\theta})}{n} \right) + \frac{1}{n} \sum_{i=1}^{n} \log r_i^{i-1}.$$

2. Large-sample distribution of MLE

$$\begin{pmatrix} \hat{\phi} \\ \hat{\theta} \end{pmatrix} - \begin{pmatrix} \phi \\ \theta \end{pmatrix} \sim AN \left(0, \frac{\sigma_w^2}{n} \begin{pmatrix} \Gamma_{\phi\phi} & \Gamma_{\phi\theta} \\ \Gamma_{\theta\phi} & \Gamma_{\theta\theta}, \end{pmatrix}^{-1} \right),$$