Introduction to Time Series Analysis. Lecture 13.

Peter Bartlett

Last lecture:

1. Yule-Walker estimation
2. Maximum likelihood estimator

Introduction to Time Series Analysis. Lecture 13.

1. Review: Maximum likelihood estimation
2. Computational simplifications: un/conditional least squares
3. Diagnostics
4. Model selection
5. Integrated ARMA models

Review: Maximum likelihood estimator

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:

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

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: Simplifications

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 $n^{-1} \sum_i \log r_i^{i-1} \to 0$.

Review: Maximum likelihood estimation

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 13.

1. Review: Maximum likelihood estimation
2. Computational simplifications: un/conditional least squares
3. Diagnostics
4. Model selection
5. Integrated ARMA models

Building ARMA models

1. Plot the time series.

   Look for trends, seasonal components, step changes, outliers.

2. Nonlinearly transform data, if necessary

3. Identify preliminary values of p, and q.

4. Estimate parameters.

5. Use diagnostics to confirm residuals are white/iid/normal.

6. Model selection: Choose p and q.

Diagnostics

How do we check that a model fits well?

The residuals (innovations, $x_t - x_t^{t-1}$) should be white.
Consider the *standardized innovations*,

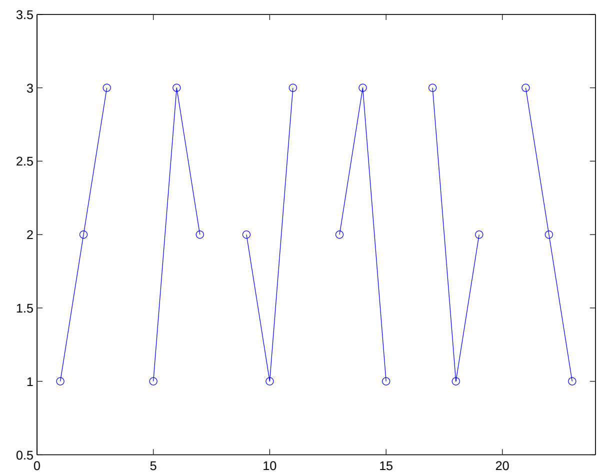
$$e_t = \frac{x_t - \hat{x}_t^{t-1}}{\sqrt{\hat{P}_t^{t-1}}}.$$

This should behave like a mean-zero, unit variance, iid sequence.

- Check a time plot
- Turning point test
- Difference sign test
- Rank test
- Q-Q plot, histogram, to assess normality

Testing i.i.d.: Turning point test

$\{X_t\}$ i.i.d. implies that X_t, X_{t+1} and X_{t+2} are equally likely to occur in any of six possible orders:



(provided X_t, X_{t+1}, X_{t+2} are distinct).

Four of the six are **turning points**.

Testing i.i.d.: Turning point test

Define $T = |\{t : X_t, X_{t+1}, X_{t+2} \text{ is a turning point}\}|$.

$\mathrm{E}T = (n-2)2/3$.

Can show $T \sim AN(2n/3, 8n/45)$.

Reject (at 5% level) the hypothesis that the series is i.i.d. if

$$\left| T - \frac{2n}{3} \right| > 1.96 \sqrt{\frac{8n}{45}}.$$

Tests for positive/negative correlations at lag 1.

Testing i.i.d.: Difference-sign test

$$S = |\{i : X_i > X_{i-1}\}| = |\{i : (\nabla X)_i > 0\}|.$$

$$\mathrm{E}S = \frac{n-1}{2}.$$

Can show $S \sim AN(n/2, n/12)$.

Reject (at 5% level) the hypothesis that the series is i.i.d. if

$$\left| S - \frac{n}{2} \right| > 1.96\sqrt{\frac{n}{12}}.$$

Tests for trend.

(But a periodic sequence can pass this test...)

Testing i.i.d.: Rank test

$$N = |\{(i, j) : X_i > X_j \text{ and } i > j\}|.$$

$$\mathrm{E}N = \frac{n(n-1)}{4}.$$

Can show $N \sim AN(n^2/4, n^3/36)$.

Reject (at 5% level) the hypothesis that the series is i.i.d. if

$$\left| N - \frac{n^2}{4} \right| > 1.96 \sqrt{\frac{n^3}{36}}.$$

Tests for linear trend.

Testing if an i.i.d. sequence is Gaussian: qq plot

Plot the pairs $(m_1, X_{(1)}), \ldots, (m_n, X_{(n)})$,
where $m_j = \mathrm{E}Z_{(j)}$,
$Z_{(1)} < \cdots < Z_{(n)}$ are order statistics from $N(0, 1)$ sample of size n, and
$X_{(1)} < \cdots < X_{(n)}$ are order statistics of the series $X_1, \ldots, X_n$.

Idea: If $X_i \sim N(\mu, \sigma^2)$, then

$$\mathrm{E}X_{(j)} = \mu + \sigma m_j,$$

so $(m_j, X_{(j)})$ should be *linear*.

There are tests based on how far correlation of $(m_j, X_{(j)})$ is from 1.

Introduction to Time Series Analysis. Lecture 13.

1. Review: Maximum likelihood estimation
2. Computational simplifications: un/conditional least squares
3. Diagnostics
4. Model selection
5. Integrated ARMA models