Introduction to Time Series Analysis. Lecture 12.

Peter Bartlett

Last lecture:

1. Parameter estimation
2. Maximum likelihood estimator
3. Yule-Walker estimation

Introduction to Time Series Analysis. Lecture 12.

1. Review: Yule-Walker estimators
2. Yule-Walker example
3. Efficiency
4. Maximum likelihood estimation
5. Large-sample distribution of MLE

Review: Yule-Walker estimation

Method of moments: We choose parameters for which the moments are equal to the empirical moments.

In this case, we choose ϕ so that $\gamma(h) = \hat{\gamma}(h)$ for $h = 0, \ldots, p$.

Yule-Walker equations for $\hat{\phi}$:
$$\begin{cases} \hat{\Gamma}_p \hat{\phi} = \hat{\gamma}_p, \\ \hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\phi}' \hat{\gamma}_p. \end{cases}$$

These are the forecasting equations.

We can use the Durbin-Levinson algorithm.

Review: Confidence intervals for Yule-Walker estimation

If $\{X_t\}$ is an AR(p) process,

$$\hat{\phi} \sim AN\left(\phi, \frac{\sigma^2}{n}\Gamma_p^{-1}\right), \qquad \hat{\sigma}^2 \xrightarrow{P} \sigma^2.$$

$$\hat{\phi}_{hh} \sim AN\left(0, \frac{1}{n}\right) \quad \text{for } h > p.$$

Thus, we can use the sample PACF to test for AR order, and we can calculate approximate confidence intervals for the parameters ϕ.

Review: Confidence intervals for Yule-Walker estimation

If $\{X_t\}$ is an AR(p) process, and n is large,

- $\sqrt{n}(\hat{\phi}_p - \phi_p)$ is approximately $N(0, \hat{\sigma}^2 \hat{\Gamma}_p^{-1})$,
- with probability $\approx 1 - \alpha$, ϕ_{pj} is in the interval

$$\hat{\phi}_{pj} \pm \Phi_{1-\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}} \left(\hat{\Gamma}_p^{-1} \right)_{jj}^{1/2},$$

where $\Phi_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal.

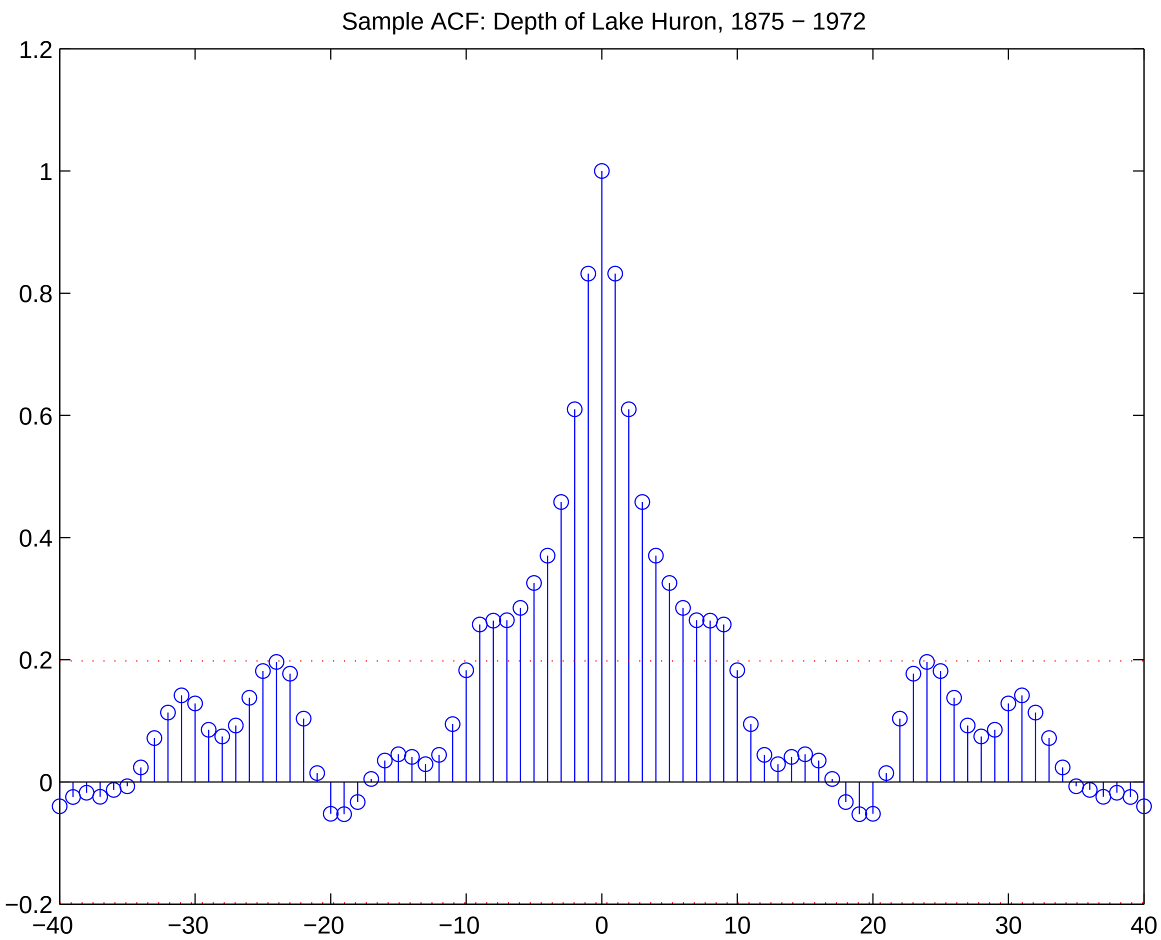
Review: Confidence intervals for Yule-Walker estimation

- with probability $\approx 1-\alpha$, ϕ_p is in the ellipsoid

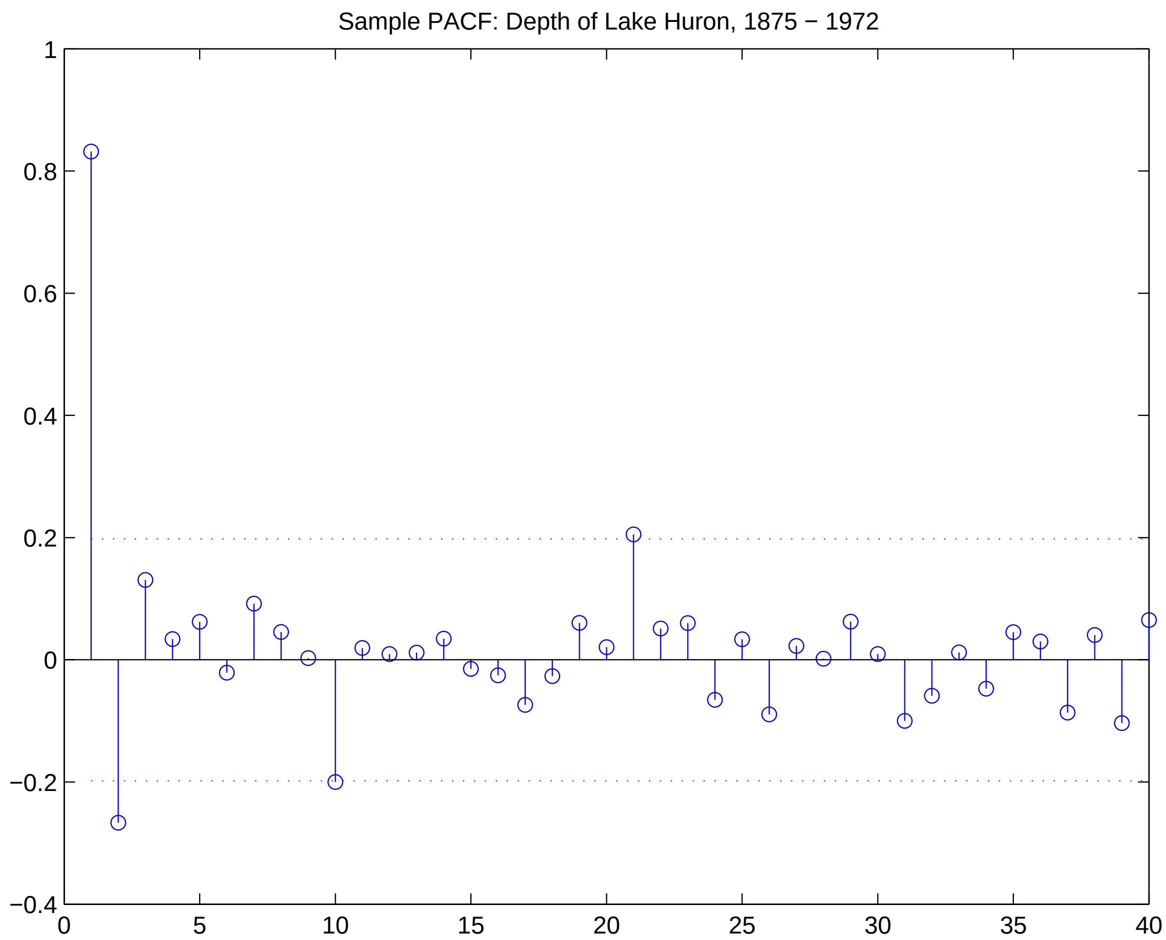
$$\left\{\phi \in \mathbb{R}^p : \left(\hat{\phi}_p - \phi\right)' \hat{\Gamma}_p \left(\hat{\phi}_p - \phi\right) \leq \frac{\hat{\sigma}^2}{n} \chi^2_{1-\alpha}(p)\right\},$$

where $\chi^2_{1-\alpha}(p)$ is the $(1-\alpha)$ quantile of the chi-squared with p degrees of freedom.

Yule Walker estimation: Example



Yule Walker estimation: Example



Yule Walker estimation: Example

$$\hat{\Gamma}_2 = \begin{pmatrix} 1.7379 & 1.4458 \\ 1.4458 & 1.7379 \end{pmatrix} \qquad \hat{\gamma}_2 = \begin{pmatrix} 1.4458 \\ 1.0600 \end{pmatrix}$$

$$\hat{\phi}_2 = \hat{\Gamma}_2^{-1}\hat{\gamma}_2 = \begin{pmatrix} 1.0538 \\ -0.2668 \end{pmatrix}$$

$$\hat{\sigma}_w^2 = \hat{\gamma}(0) - \hat{\phi}_2'\hat{\gamma}_2 = 0.4971$$

Yule Walker estimation: Example

Confidence intervals:

$$\hat{\phi}_1 \pm \Phi_{1-\alpha/2} \left(\hat{\sigma}_w^2 \hat{\Gamma}_2^{-1}/n \right)_{11}^{1/2} = 1.0538 \pm 0.1908$$

$$\hat{\phi}_2 \pm \Phi_{1-\alpha/2} \left(\hat{\sigma}_w^2 \hat{\Gamma}_2^{-1}/n \right)_{22}^{1/2} = -0.2668 \pm 0.1908$$

Yule-Walker estimation

It is also possible to define analogous estimators for ARMA(p,q) models with $q > 0$:

$$\hat{\gamma}(j) - \phi_1\hat{\gamma}(j-1) - \cdots - \phi_p\hat{\gamma}(j-p) = \sigma^2 \sum_{i=j}^{q} \theta_i \psi_{i-j},$$

where $\psi(B) = \theta(B)/\phi(B)$.
Because of the dependence on the ψ_i, these equations are nonlinear in ϕ_i, θ_i.
There might be no solution, or nonunique solutions.
Also, the *asymptotic efficiency* of this estimator is poor: it has unnecessarily high variance.

Efficiency of estimators

Let $\hat{\phi}^{(1)}$ and $\hat{\phi}^{(2)}$ be two estimators. Suppose that

$$\hat{\phi}^{(1)} \sim AN(\phi, \sigma_1^2), \qquad \hat{\phi}^{(2)} \sim AN(\phi, \sigma_2^2).$$

The asymptotic efficiency of $\hat{\phi}^{(1)}$ relative to $\hat{\phi}^{(2)}$ is

$$e\left(\phi, \hat{\phi}^{(1)}, \hat{\phi}^{(2)}\right) = \frac{\sigma_2^2}{\sigma_1^2}.$$

If $e\left(\phi, \hat{\phi}^{(1)}, \hat{\phi}^{(2)}\right) \leq 1$ for all ϕ, we say that $\hat{\phi}^{(2)}$ is a *more efficient* estimator of ϕ than $\hat{\phi}^{(1)}$.

For example, for an AR(p) process, the moment estimator and the maximum likelihood estimator are as efficient as each other.

For an MA(q) process, the moment estimator is less efficient than the innovations estimator, which is less efficient than the MLE.

Yule Walker estimation: Example

AR(1):
$$\gamma(0) = \frac{\sigma^2}{1-\phi_1^2}$$

$$\hat{\phi}_1 \sim AN\left(\phi_1, \frac{\sigma^2}{n}\Gamma_1^{-1}\right) = AN\left(\phi_1, \frac{1-\phi_1^2}{n}\right).$$

AR(2):
$$\begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{pmatrix} \sim AN\left(\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \frac{\sigma^2}{n}\Gamma_2^{-1}\right)$$

and
$$\frac{\sigma^2}{n}\Gamma_2^{-1} = \frac{1}{n}\begin{pmatrix} 1-\phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2 \end{pmatrix}.$$

Yule Walker estimation: Example

Suppose $\{X_t\}$ is an AR(1) process and the sample size n is large.

If we estimate ϕ, we have

$$\mathrm{Var}(\hat{\phi}_1) \approx \frac{1-\phi_1^2}{n}.$$

If we fit a *larger* model, say an AR(2), to this AR(1) process,

$$\mathrm{Var}(\hat{\phi}_1) \approx \frac{1-\phi_2^2}{n} = \frac{1}{n} \quad > \quad \frac{1-\phi_1^2}{n}.$$

We have lost efficiency.

Introduction to Time Series Analysis. Lecture 12.

1. Review: Yule-Walker estimators

2. Yule-Walker example

3. Efficiency

4. Maximum likelihood estimation

5. 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:

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

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

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

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

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

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

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

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

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

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

Introduction to Time Series Analysis. Lecture 12.

1. Review: Yule-Walker estimators
2. Yule-Walker example
3. Efficiency
4. Maximum likelihood estimation
5. Large-sample distribution of MLE