MA and ARMA Spectral Estimation

MA Spectral Estimation

Unlike AR spectral estimation, MA (and more generally, ARMA) spectral estimation requires solving a non-linear system of equations:

$$\gamma_x[k] = \begin{cases} -\sum_{\ell=1}^{p} a[\ell]\,\gamma_x[k-\ell] + \sigma_w^2 \sum_{\ell=0}^{q-k} h^*[\ell]\,b[\ell+k] & k = 0, 1, \ldots, q \\ -\sum_{\ell=1}^{p} a[\ell]\,\gamma_x[k-\ell] & k \geq q+1 \end{cases}$$

where

$a[\ell] \rightarrow$ AR coefficients

$b[\ell] \rightarrow$ MA coefficients

$h[\ell] \rightarrow$ impulse response of $H(z) = \dfrac{B(z)}{A(z)}$,
function of $a[\ell]$, $b[\ell]$

Connections to Blackman-Tukey

Assume $x[n]$ is $MA(q)$. That is,

$$x[n] = \sum_{k=0}^{q} b[k]\, w[n-k]$$

where $w[n]$ is white noise. Then

$$\gamma_x[k] = \begin{cases} \sigma_w^2 \cdot \sum_{k=-q}^{q} b^*[\ell]\, b[\ell+k] & , k = 0, 1, \dots, q \\ \gamma_x^*[-k] & , k = -q, \dots, -1, 0 \\ 0 & , |k| > q \end{cases}$$

Thus, the PSD of $x[n]$ may be written

$$P_{MA}(f) = \sum_{k=-q}^{q} \gamma_x[k] \exp(-2\pi i f k)$$

This is simply a Blackman-Tukey spectral estimator with a lag window = rectangular window of length $2q+1$.

Today we'll consider an MA estimator that estimates the coefficients $b[k]$ explicitly.

Durbin's Method

Recall that the MA(q) process

$$x[n] = \sum_{k=0}^{q} b[k]\, w[n-k]$$

is equivalent to the AR(∞) process

$$x[n] = -\sum_{k=1}^{\infty} a[k]\, x[n-k] + w[n]$$

where

$$A(z) = \frac{1}{B(z)}$$

Darbin's method proceeds as follows:

1. Model $x[n]$ as an AR(L) process where $L \gg q$ (we know how to do this)

2. Convert the AR parameters to MA parameters (we'll see a way to do this)

The hope is that the AR(L) model is sufficiently close to the exact AR(∞) model.

Note we cannot simply take L to be arbitrarily large. The estimate of the AR(L) is still subject "finite sample" effects and we must worry about overfitting.

A reasonable choice might be the geometric mean between q and N:

$$L = \sqrt{qN}$$

Another reasonable choice is

$$L = N/5.$$

Certainly $L = N/2$ is an extreme upper bound.

The Conversion AR(L) → MA(q)

Notation:

$\underline{a}$ → true AR(L) fit (solution to Yule-Walker using true ACV)

$\underline{\hat{a}}$ → estimate of $\underline{a}$ using a method discussed previously

$\underline{b}$ → true MA(q) fit (we're assuming $x[n]$ is truly an MA(q) process)

$\underline{\hat{b}}$ → estimate of $\underline{b}$ (to be determined)

Both $\underline{a}$ and $\underline{\hat{a}}$ define linear predictors of $x[n]$:

$$\hat{x}_{\underline{a}}[n] = -\sum_{k=1}^{L} a[k]\, x[n-k] \qquad \leftarrow \boxed{\text{best predictor}}$$

$$\hat{x}_{\underline{\hat{a}}}[n] = -\sum_{k=1}^{L} \hat{a}[k]\, x[n-k] \qquad \leftarrow \boxed{\text{estimated predictor}}$$

If $\hat{\underline{q}}$ is a good estimate of $\underline{q}$, then

$$Q := E\left\{ \left(\hat{x}_{\underline{q}}[n] - \hat{x}_{\hat{\underline{q}}}[n] \right)^2 \right\}$$

should be small.

Exercise | Simplify Q. The notation

$$\underline{x} = \left[x[n-1] \quad \cdots \quad x[n-L] \right]^T$$ may be

helpful.

Solution

$$Q = E\left\{\left(\underline{a}^T \underline{x} - \underline{\hat{a}}^T \underline{x}\right)^2\right\}$$

$$= E\left\{(\underline{a} - \underline{\hat{a}})^T \underline{x} \cdot \underline{x}^T (\underline{a} - \underline{\hat{a}})\right\}$$

$$= (\underline{a} - \underline{\hat{a}})^T \Gamma_L (\underline{a} - \underline{\hat{a}})$$

where $\Gamma_L = E\{\underline{x} \cdot \underline{x}^T\}$

$L \times L$

Let us write

$$\Gamma_L = \Gamma_L(\underline{b})$$

since the ACV is defined in terms of $\underline{b}$.

Recall that Q should be small, reflecting our assumption that $\underline{\hat{a}}$ is a good estimate of $\underline{a}$.

Recall Q is defined in terms of the true MA parameters $\underline{b}$

$\underline{\hat{b}}$ close to $\underline{b}$ $\longrightarrow$ Q is small

$\underline{\hat{b}}$ far from $\underline{b}$ $\longrightarrow$ Q is large

This motivates our criterion for converting $\underline{\hat{a}} \longrightarrow \underline{\hat{b}}$:

$$\boxed{\begin{array}{l}\text{Choose } \underline{\hat{b}} \text{ that minimizes} \\ \hat{Q} := (\underline{a} - \underline{\hat{a}})^T \Gamma_L(\underline{\hat{b}})\,(\underline{a} - \underline{\hat{a}})\end{array}}$$

Note: Kay motivates this criterion as a approximate maximum likelihood estimate.

Minimizing $\hat{Q}$

It is convenient to minimize

$$\frac{\hat{Q}}{\sigma^2} = (\underline{a} - \underline{\hat{a}})^T \bar{\Gamma}_L(\underline{\hat{b}})(\underline{a} - \underline{\hat{a}})$$

$$\bar{\Gamma}_L = \frac{\Gamma_L}{\sigma^2}$$

This gets rid of the dependence on σ^2 in Γ_L.

Now, let's write (omitting dependence on $\underline{\hat{b}}$)

$$\frac{\hat{Q}}{\sigma^2} = \underline{\hat{a}}^T \bar{\Gamma}_L \underline{\hat{a}} - 2\underline{\hat{a}}^T \bar{\Gamma}_L \underline{a} + \underline{a}^T \bar{\Gamma}_L \underline{a}$$

By the Yule-Walker equations for an AR(L) process:

$$\bar{\Gamma}_L \underline{a} = -\underline{\bar{\gamma}}_L$$

$$\underline{\bar{\gamma}}_L = \frac{\underline{\gamma}_L}{\sigma^2}$$

Thus

$$\frac{\hat{Q}}{\sigma_\omega^2} = \underline{\hat{a}}^T \,\bar{\Gamma}_L\, \underline{a} + 2\underline{\hat{a}}^T \underline{\bar{\gamma}}_L - \underline{a}^T \underline{\bar{\gamma}}_L$$

Recall for an AR(L) process we have

$$\sigma_\omega^2 = \gamma_x[0] + \sigma_\omega^2\, \underline{a}^T \underline{\gamma}_L$$

$$= \gamma_x[0] + \sigma_\omega^2\, \underline{a}^T \underline{\bar{\gamma}}_L$$

$$\Rightarrow -\underline{a}^T \underline{\bar{\gamma}}_L = \frac{\gamma_x[0]}{\sigma_\omega^2} - 1$$

$$\Rightarrow \frac{\hat{Q}}{\sigma_\omega^2} = \underline{\hat{a}}^T \,\bar{\Gamma}_L\, \underline{a} + 2\underline{\hat{a}}^T \underline{\bar{\gamma}}_L + \underbrace{\frac{\gamma_x[0]}{\sigma_\omega^2}}_{\bar{\gamma}_x[0]} - 1$$

In matrix form:

$$\frac{\hat{Q}}{\sigma_\omega^2} = \begin{bmatrix} 1 \\ \underline{\hat{a}} \end{bmatrix}^T \underbrace{\begin{bmatrix} \bar{\gamma}_x[0] & \underline{\bar{\gamma}}_L^T \\ \underline{\bar{\gamma}}_L & \bar{\Gamma}_L \end{bmatrix}}_{\text{depends on } \hat{\underline{b}}} \begin{bmatrix} 1 \\ \underline{\hat{a}} \end{bmatrix} - 1$$

In quadratic form:

assume $\hat{a}[0] = 1$

$$\frac{\hat{Q}}{\sigma_w^2} = \left(\sum_{i=0}^{L} \sum_{j=0}^{L} \hat{a}[i]\, \hat{a}[j]\, \bar{\gamma}_x[i-j] \right) - 1$$

Now let's plug in

$$\bar{\gamma}_x[i-j] = \sum_{n=-\infty}^{\infty} b[n]\, b[n+i-j]$$

$$= \sum_{n=-\infty}^{\infty} b[n-i]\, b[n-j]$$

where $b[0] = 1$, $b[n] = 0$ for $n < 0, n > q$.

Plugging in, we get

$$\frac{\hat{Q}}{\sigma_w^2} = \sum_{i=0}^{L} \sum_{j=0}^{L} \hat{a}[i]\, \hat{a}[j] \sum_{n=-\infty}^{\infty} \hat{b}[n-i]\, \hat{b}[n-j] \;-1$$

$$= \sum_{n=-\infty}^{\infty} \left(\sum_{i=0}^{L} \hat{a}[i]\, \hat{b}[n-i] \right) \left(\sum_{j=0}^{L} \hat{a}[j]\, \hat{b}[n-j] \right) - 1$$

$$= \sum_{n=-\infty}^{\infty} \left(\sum_{k=0}^{L} \hat{a}[k]\, \hat{b}[n-k] \right)^2 - 1$$

$$= \sum_{n=-\infty}^{\infty} \left(\sum_{k=-\infty}^{\infty} \hat{a}[k]\, \hat{b}[n-k] \right)^2 - 1$$

$\hookrightarrow = 0$ for $k < 0$, $k > L$

$$= \sum_{n=-\infty}^{\infty} \left(\sum_{k=-\infty}^{\infty} \hat{b}[k]\, \hat{a}[n-k] \right)^2 - 1$$

$$= \sum_{n=-\infty}^{\infty} \left(\hat{a}[n] + \sum_{k=1}^{q} \hat{b}[k]\, \hat{a}[n-k] \right)^2 - 1$$

So how should we choose $\hat{b}[k]$ to minimize this expression?

Note that $\dfrac{\left[\dfrac{\hat{Q}}{\sigma_w^2}\right] + 1}{L+1}$ is the prediction error power estimate for the "data" $\{1, \hat{a}[1], \ldots, \hat{a}[L]\}$ used by the autocorrelation method.

Summary of Durbin's Method

1. Choose L, $q \ll L \ll N$, and estimate $AR(L)$ model $\hat{\underline{a}}$.

2. Use autocorrelation method applied to $\{1, \hat{a}[1], \ldots, \hat{a}[L]\}$ to compute $\hat{\underline{b}}$

Optionally: Apply some other AR method in 2, such as modified covariance

Other issues

1. How do we estimate σ_w^2?

$$\hat{\sigma}_w^2 = \hat{\gamma}_x[0] + \sum_{k=1}^{L} \hat{a}[k]\,\hat{\gamma}_x[k]$$

from Yule-Walker (recall $B(z)$ and $\frac{1}{A(z)}$ assume the same driving noise)

2. Model order selection: Choosing q.

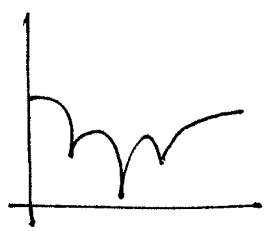
→ AIC, MDL, and various other methods suggested by Kay.

Once again: ART, NOT SCIENCE!

Comments

1. The MA(q) estimate is only as good as the AR(L) estimate.

2. So, why not just use the AR(L) estimate of the PSD? Use the MA(q) estimate if you think your PSD looks like

 and not 

3. Another estimation method called the "method of moments" is possible, but it coincides with the BT estimate discussed earlier.

4. Other MA estimators are discussed briefly in Kay.

ARMA Spectral Estimation

Kay describes several approximate algorithms.
Let's briefly look at one of the simpler ones:

Modified Yule-Walker Equations

Recall:

$$\gamma_x[k] = \begin{cases} -\sum_{\ell=1}^{p} a[\ell]\,\gamma_x[k-\ell] + \sigma_w^2 \sum_{\ell=0}^{q-k} h^*[\ell]\,b[\ell+k] & k = 0, 1, \ldots, q \\ -\sum_{\ell=1}^{p} a[\ell]\,\gamma_x[k-\ell] & k \geq q+1 \end{cases}$$

Idea: View ARMA Model as

$$w[n] \longrightarrow \boxed{\frac{1}{A(z)}} \longrightarrow \boxed{B(z)} \longrightarrow x[n]$$

1. Determine $\hat{A}(z)$ by solving YWE for $k = q+1, \ldots, q+p$

2. Remove the "AR part" by setting
$$\tilde{x}[n] = \{\hat{A}(z)\} \star x[n]$$

3. Estimate MA part by applying Durbin's (or some other) method to $\tilde{x}[n]$.

Step 1 :

$$\underbrace{\begin{bmatrix} \gamma[q] & \gamma[q-1] & \cdots & \gamma[q-p+1] \\ \gamma[q+1] & \gamma[q] & \cdots & \gamma[q-p+2] \\ \vdots & & & \\ \gamma[q+p-1] & & \cdots & \gamma[q] \end{bmatrix}}_{\Gamma_p'} \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[p] \end{bmatrix} = - \begin{bmatrix} \gamma[q+1] \\ \gamma[q+2] \\ \vdots \\ \gamma[q+p] \end{bmatrix}$$

- Toeplitz, Hermitian matrix $\Longrightarrow$ $O(p^2)$

  Levinson-like algorithm exists

- Might be <u>singular</u>!
- Not necessarily min. phase.

Least-Squares Modified YWE

Use $M > p$ equations $\longrightarrow$ $\Gamma_M' \cdot \underline{a}_p = - \underline{\gamma}_M'$

Overdetermined $\longrightarrow$ pseudoinverse

$$\hat{\underline{a}}_p = - \left(\Gamma_M'^{H} \Gamma_M' \right)^{-1} \Gamma_M'^{H} \underline{\gamma}_M'$$

This method tends to have less variance than the MYW estimator.

System Identification Approach

Idea: If we knew the driving noise $w[n]$, then we'd have a linear system of equations for $\{a[\ell]\}$ and $\{b[\ell]\}$!

$$x[n] = -\sum_{k=1}^{p} a[k]\, x[n-k] + \sum_{k=1}^{q} b[k]\, w[n-k] + w[n],$$

$$n = 0, 1, \ldots, N-1$$

(assume $x[n] = u[n] = 0, \quad n < 0, \ n \geq N$).

In matrix form:

$$\underline{x} = H\underline{\theta} + \underline{u}$$

where

$$\underline{x} = [x[0] \ \cdots \ x[N-1]]^T$$

$$\underline{u} = [u[0] \ \cdots \ u[N-1]]^T$$

$$\underline{\theta} = [-a[1] \ \cdots \ -a[p] \ \ b[1] \ \cdots \ b[q]]^T$$

$$H = \begin{bmatrix} x[-1] & \cdots & x[-p] & u[-1] & \cdots & u[-q] \\ \vdots & & \vdots & \vdots & & \\ x[N-2] & \cdots & x[N-p-1] & u[N-2] & \cdots & u[N-q-1] \end{bmatrix}$$

$N \times (p+q)$

This suggests the least squares estimator

$$\hat{\underline{\theta}} = (\hat{H}^H \hat{H})^{-1} \hat{H}^H \underline{x}$$

where $\hat{H}$ is based on an estimate $\hat{\underline{u}}$ of $\underline{u}$.

How to get $\hat{\underline{u}}$? The output of a high order linear prediction (<u>whitening</u>) filter.

Additional variations on this approach discussed in Kay.