Time Series Analysis

Andreas Jakobsson

Content:

- Prediction error method
- Maximum likelihood

The prediction error method

We will later on in the course look at how to best predict a time series using its signal model. Here, we will assume we have formed such an estimate, in this case of the one-step prediction of the process.

With it, one may form the prediction error

$$\epsilon_{t+1|t}(\mathbf{\Theta}) = y_{t+1} - \hat{y}_{t+1|t}(\mathbf{\Theta})$$

where $\epsilon_{t+1|t}(\mathbf{\Theta})$ and $\hat{y}_{t+1|t}(\mathbf{\Theta})$ are the prediction error and the one-step prediction, respectively, at time $t+1$, given a collection of measurements up to time t, with $\mathbf{\Theta}$ denoting the parameter vector detailing the model of the process y_t and the collection of all the available measurements up to time t, respectively, i.e.,

$$\mathbf{\Theta} = \left[\begin{array}{cc} \boldsymbol{\theta} & \mathbf{Y}_t \end{array} \right]$$

where $\boldsymbol{\theta}$ denotes the unknown n_θ-dimensional parameter vector, and

$$\mathbf{Y}_t = \left[\begin{array}{ccc} y_1 & \ldots & y_t \end{array} \right]^T$$

The prediction error method

Notice that the prediction error

$$\epsilon_{t+1|t}(\mathbf{\Theta}) = y_{t+1} - \hat{y}_{t+1|t}(\mathbf{\Theta})$$

thus depends on the unknown parameters $\boldsymbol{\theta}$.

The idea behind the prediction error method (PEM) is to choose the parameters $\boldsymbol{\theta}_{PEM}$ such that the variance of this error is minimised.

The prediction error method

Notice that the prediction error

$$\epsilon_{t+1|t}(\boldsymbol{\Theta}) = y_{t+1} - \hat{y}_{t+1|t}(\boldsymbol{\Theta})$$

thus depends on the unknown parameters $\boldsymbol{\theta}$.

The idea behind the prediction error method (PEM) is to choose the parameters $\boldsymbol{\theta}_{PEM}$ such that the variance of this error is minimised.

In order to do so, we estimate the variance of the prediction error as

$$\sum_{t} \left| \epsilon_{t+1|t}(\boldsymbol{\Theta}) \right|^2$$

where the sum is over the available prediction errors.

We thus seek the estimate

$$\hat{\boldsymbol{\theta}}_{PEM} = \arg\min_{\boldsymbol{\theta}} \sum_{t} \left| \epsilon_{t+1|t}(\boldsymbol{\Theta}) \right|^2$$

The prediction error method

Example:

Consider an ARMA(p,q) process

$$
\begin{aligned}
y_t &= \sum_{\ell=0}^{q} c_\ell e_{t-\ell} - \sum_{\ell=1}^{p} a_\ell y_{t-\ell} \\
&= \begin{bmatrix} -y_{t-1} & \ldots & -y_{t-p} & e_{t-1} & \ldots & e_{t-q} \end{bmatrix} \boldsymbol{\theta} + e_t \\
&= \mathbf{x}_t^T \boldsymbol{\theta} + e_t
\end{aligned}
$$

with $c_0 = 1$, and

$$
\boldsymbol{\theta} = \begin{bmatrix} a_1 & \ldots & a_p & c_1 & \ldots & c_q \end{bmatrix}^T
$$

Note that the process depends on the (unknown) noise values. These must therefore be predicted somehow; typically, this is done using the one-step prediction error, $\epsilon_{t+1-\ell|t-\ell}(\boldsymbol{\Theta})$. For the correct model, this makes sense - it will then be an estimate of the noise value; otherwise, it will be something else.

The prediction error method

However, $\epsilon_{t+1-\ell|t-\ell}(\mathbf{\Theta})$ will depend on the earlier noise estimates, as

$$E\left\{y_{t+1}|\mathbf{\Theta}\right\} = \sum_{\ell=0}^{q} c_\ell \epsilon_{t+1-\ell|t-\ell}(\mathbf{\Theta}) - \sum_{\ell=1}^{p} a_\ell y_{t+1-\ell}$$

where the prediction error $\epsilon_{t+1-\ell|t-\ell}(\mathbf{\Theta})$ is computed recursively as

$$\epsilon_{t+1-\ell|t-\ell}(\mathbf{\Theta}) = y_{t+1-\ell} - \hat{y}_{t+1-\ell|t-\ell}(\mathbf{\Theta})$$

Clearly, unless infinitely many past observations are available, it is not possible to form the one-step prediction without making some assumptions of the initial values of the noise process; typically, this is handled by assuming that $\epsilon_{t|t-1}(\mathbf{\Theta}) = 0$, for $t = p-q+1, \ldots, p$.

Note that the PEM estimate results in a multimodal and multi-dimensional minimisation problem.

The maximum likelihood method

As an alternative, one may form an estimate of the unknown parameters such that one maximises the likelihood function of the observed data, i.e.,

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} f(\mathbf{y}; \boldsymbol{\theta})$$

where $f(\cdot)$ is the PDF of $\mathbf{y}$.

We here assume Gaussian distributed measurements, such that for N observed samples, formed as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{e}$$

the PDF is

$$f(\mathbf{y}) = (2\pi)^{-N/2} \det\left(\mathbf{R}_{\mathbf{e}}\right)^{-1/2} \exp\left\{-\frac{1}{2}\left[\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\right]^T \mathbf{R}_{\mathbf{e}}^{-1} \left[\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\right]\right\}$$

The maximum likelihood method

As the $\log(\cdot)$ is a monotone function,

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} f(\mathbf{y};\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta}} \log f(\mathbf{y};\boldsymbol{\theta})$$

where

$$\log f(\mathbf{y};\boldsymbol{\theta}) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log\det\left(\mathbf{R_e}\right) - \frac{1}{2}\left[\mathbf{y}-\mathbf{X}\boldsymbol{\theta}\right]^T \mathbf{R}_{\mathbf{e}}^{-1}\left[\mathbf{y}-\mathbf{X}\boldsymbol{\theta}\right]$$

Thus,

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\min_{\boldsymbol{\theta}} \left\{\log\det\left(\mathbf{R_e}\right) + \left[\mathbf{y}-\mathbf{X}\boldsymbol{\theta}\right]^T \mathbf{R}_{\mathbf{e}}^{-1}\left[\mathbf{y}-\mathbf{X}\boldsymbol{\theta}\right]\right\}$$

To solve this, we here assume that $\mathbf{R_e}$ is known, such that

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\min_{\boldsymbol{\theta}} \left\{\left[\mathbf{y}-\mathbf{X}\boldsymbol{\theta}\right]^T \mathbf{R}_{\mathbf{e}}^{-1}\left[\mathbf{y}-\mathbf{X}\boldsymbol{\theta}\right]\right\}$$

It should be stressed that this is clearly not a realistic assumption.

The maximum likelihood method

In the particular case of ARMA models, one can rewrite this maximisation as

$$\left\{\hat{\boldsymbol{\theta}}_{ML}, \hat{\sigma}_e^2\right\} = \arg\min_{\boldsymbol{\theta}, \sigma_e^2} (N-p)ln\left(\sigma_e^2\right) + \frac{1}{\sigma_e^2}\sum_{t=p+1}^{N} \epsilon_t^2(\boldsymbol{\theta}, \sigma_e^2)$$

Minimising with respect to σ_e^2 yields that

$$\hat{\sigma}_e^2 = \frac{1}{N-p}\sum_{t=p+1}^{N} \epsilon_t^2(\boldsymbol{\theta}, \sigma_e^2)$$

which, if inserted above implies that

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\min_{\boldsymbol{\theta}} \sum_{t=p+1}^{N} \epsilon_t^2(\boldsymbol{\theta}, \sigma_e^2)$$

which thus coincides with the PEM estimate.