LECTURE NOTES 21

TRYGGVI JÓNSSON

CONTENT OUTLINE

State space models, 2nd part:

- The Kalman filter when some observations are missing
- ARMA-models on state space form
- ML-estimates of state space models
- Assignment 4

A LINEAR STOCHASTIC STATE SPACE MODEL

System equation: $\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{B}\mathbf{u}_{t-1} + e_{1,t}$

Observation equation: $\mathbf{Y}_t = \mathbf{C}\mathbf{X}_t + e_{2,t}$

System equation	Observation equation
$\mathbf{X}$: State vector	$\mathbf{Y}$: Observation vector
$\mathbf{u}$: Input vector	e_2: Observation noise
e_1: System noise	

A LINEAR STOCHASTIC STATE SPACE MODEL

System equation: $\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{B}\mathbf{u}_{t-1} + e_{1,t}$

Observation equation: $\mathbf{Y}_t = \mathbf{C}\mathbf{X}_t + e_{2,t}$

- $dim(\mathbf{X}_t) = m$ is called the order of the system
- $e_{1,t}$ and $e_{2,t}$ mutually independent white noise
- $\mathbb{V}[e_1] = \mathbf{\Sigma}_1, \mathbb{V}[e_2] = \mathbf{\Sigma}_2$
- $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2$ are known matrices

THE KALMAN FILTER

Initialization:

$$\widehat{\mathbf{X}}_{1|0} = \mathbb{E}[\mathbf{X}_1] = \mu_0$$

$$\mathbf{\Sigma}^{xx}_{1|0} = \mathbb{V}[\mathbf{X}_1] = \mathbb{V}_0$$

$$\Rightarrow$$

$$\mathbf{\Sigma}^{yy}_{1|0} = \mathbf{C}\mathbf{\Sigma}^{xx}_{1|0}\mathbf{C}^T + \mathbf{\Sigma}_2$$

RECONSTRUCTION:

For: $t = 1, 2, 3, \ldots$

$$\mathbf{K}_t = \boldsymbol{\Sigma}^{xx}_{t|t-1} \mathbf{C}^T \left(\Sigma^{yy}_{t|t-1}\right)^{-1}$$

$$\widehat{\mathbf{X}}_{t|t} = \widehat{\mathbf{X}}_{t|t-1} + \mathbf{K}_t \left(\mathbf{Y}_t - \mathbf{C}\widehat{\mathbf{X}}_{t|t-1}\right)$$

$$\boldsymbol{\Sigma}^{xx}_{t|t} = \boldsymbol{\Sigma}^{xx}_{t|t-1} - \mathbf{K}_t \boldsymbol{\Sigma}^{yy}_{t|t-1} \mathbf{K}_t^T$$

PREDICTION:

$$\widehat{\mathbf{X}}_{t+1|t} = \mathbf{A}\widehat{\mathbf{X}}_{t|t} + \mathbf{B}\mathbf{u}_t$$

$$\boldsymbol{\Sigma}^{xx}_{t+1|t} = \mathbf{A}\boldsymbol{\Sigma}^{xx}_{t|t}\mathbf{A}^T + \boldsymbol{\Sigma}_1$$

$$\boldsymbol{\Sigma}^{yy}_{t+1|t} = \mathbf{C}\boldsymbol{\Sigma}^{xx}_{t+1|t}\mathbf{C}^T + \Sigma_2$$

MISSING OBSERVATIONS

- What happens if Y_t is missing for some t?

$$\mathbf{K}_t = \boldsymbol{\Sigma}^{xx}_{t|t-1}\mathbf{C}^T\left(\Sigma^{yy}_{t|t-1}\right)^{-1}$$

$$\widehat{\mathbf{X}}_{t|t} = \widehat{\mathbf{X}}_{t|t-1} + \mathbf{K}_t\left(\mathbf{Y}_t - \mathbf{C}\widehat{\mathbf{X}}_{t|t-1}\right)$$

$$\boldsymbol{\Sigma}^{xx}_{t|t} = \boldsymbol{\Sigma}^{xx}_{t|t-1} - \mathbf{K}_t\boldsymbol{\Sigma}^{yy}_{t|t-1}\mathbf{K}^T_t$$

- Can't use Y_t for reconstruction. Otherwise keep calm and proceed as normal
- Use $\mathbb{E}[\mathbf{Y}_t | \widehat{\mathbf{X}}_{t|t-1}]$ instead of $\mathbf{Y}_t$, i.e. assume $\widehat{\mathbf{X}}_{t|t} = \widehat{\mathbf{X}}_{t|t-1}$

- No dependence on $\mathbf{Y}_t$ in the prediction step:

$$
\begin{aligned}
\widehat{\mathbf{X}}_{t+1|t} &= \mathbf{A}\widehat{\mathbf{X}}_{t|t} + \mathbf{B}\mathbf{u}_t \\
\boldsymbol{\Sigma}^{xx}_{t+1|t} &= \mathbf{A}\boldsymbol{\Sigma}^{xx}_{t|t}\mathbf{A}^T + \boldsymbol{\Sigma}_1 \\
\boldsymbol{\Sigma}^{yy}_{t+1|t} &= \mathbf{C}\boldsymbol{\Sigma}^{xx}_{t+1|t}\mathbf{C}^T + \Sigma_2
\end{aligned}
$$

ESTIMATION IN ARMA(P,Q)-MODELS USING THE KF

- Using the Kalman filter we can get the mean and variance of the one-step predictions of the observations:

$$\widehat{\mathbf{Y}}_{t+1|t} = \mathbf{C}\widehat{\mathbf{X}}_{t+1|t}$$

$$\boldsymbol{\Sigma}^{yy}_{t+1|t} = \mathbf{C}\boldsymbol{\Sigma}^{xx}_{t+1|t}\mathbf{C}^T + \Sigma_2$$

- The Kalman filter can handle missing observations
- An ARMA(p, q)-model can be written as a state space model
- This gives us a way of calculating ML-estimates in the ARMA(p, q)-model even when some observations are missing.

ARMA(P,Q)-MODELS ON STATE SPACE FORM

$$Y_t + \varphi_1 Y_{t-1} + \ldots + \varphi_p Y_{t-p}$$

$$= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon t - q$$

State space form:

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{e}_{1,t}$$

$$\mathbf{Y}_t = \mathbf{C}\mathbf{X}_t$$

$$\mathbf{X}_t = (X_{1,t}, X_{2,t}, \ldots, X_{d,t})^T, d = max(p, q + 1)$$

$$\mathbf{X}_t = \begin{bmatrix} y_t \\ \varphi_2 Y_{t-1} + \ldots + \varphi_m Y_{t-m+1} + \theta_1 \varepsilon_t + \ldots + \theta_{m-1} \varepsilon_{t-m+2} \\ \vdots \\ \varphi_{m-1} Y_{t-1} + \varphi_m Y_{t-2} + \theta_{m-2} \varepsilon_t + \theta_{m-1} \varepsilon_{t-1} \\ \varphi_m Y_{t-1} + \theta_{m-1} \varepsilon_t \end{bmatrix}$$

$$\mathbf{A}= \begin{bmatrix} -\varphi_1 & 1 & 0 & \dots & 0 \\ -\varphi_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\varphi_{d-1} & 0 & 0 & \dots & 1 \\ -\varphi_d & 0 & 0 & \dots & 0 \end{bmatrix}$$

$$\mathbf{e}_{1,t}= \mathbf{G}\varepsilon_t = \begin{bmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{d-1} \end{bmatrix} \varepsilon t \quad \mathbf{C} = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}$$

Other possibilities exist as well, e.g. :

$$\mathbf{X}_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} \boldsymbol{\varphi}^T & \\ \mathbf{I}_{d-1} & \mathbf{0} \end{bmatrix}$$

ML-ESTIMATES IN STATE SPACE MODELS

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{G}\mathbf{e}_{1,t}$$

$$\mathbf{Y}_t = \mathbf{C}\mathbf{X}_t + \mathbf{e}_{2,t}$$

- $\mathbf{e}_{1,t}$ and $\mathbf{e}_{2,t}$ are mutually uncorrelated normally distributed white noise
- $\mathbb{V}(\mathbf{e}_{1,t}) = \Sigma_1$ and $\mathbb{V}(\mathbf{e}_{2,t}) = \Sigma_2$
- For ARMA(p, q)-models we have $\mathbf{A}$, $\mathbf{C}$, and $\mathbf{G}$ as stated on the previous slide. Furthermore, $\mathbf{e}_{1,t} = \varepsilon t$, $\Sigma_1 = \sigma_\varepsilon^2$, and $\Sigma_2 = 0$

MAXIMUM LIKELIHOOD ESTIMATES

- Let Y_{N*} contain the available observations and let $\boldsymbol{\theta}$ contain the parameters of the model
- The likelihood function is the density of the random vector corresponding to the observations and given the set of parameters:

$$L(\boldsymbol{\theta}; Y_{N*}) = f(Y_{N*}|\boldsymbol{\theta})$$

- The ML-estimates is found by selecting $\boldsymbol{\theta}$ so that the density function is as large as possible at the actual observations The random variables $Y_{N*}|Y_{N*-1}$ and Y_{N*-1} are independent:

$$L(\boldsymbol{\theta}; Y_{N*}) = f(Y_{N*}|\boldsymbol{\theta}) = f(Y_{N*}|Y_{N*-1}, \boldsymbol{\theta}) f(Y_{N*-1}|\boldsymbol{\theta})$$
$$= f(Y_{N*}|Y_{N*-1}, \boldsymbol{\theta}) f(Y_{N*-1}|Y_{N*-2}, \boldsymbol{\theta}) \dots f(Y_1|\boldsymbol{\theta})$$

- The conditional densities can be found using the Kalman filter

MLE / KF

- Assuming that at time t we have:
  $\widehat{\mathbf{X}}_{t|t} = \mathbb{E}[\mathbf{X}_t|\mathcal{Y}_t]$ and $\Sigma^{xx}t|t = \mathbb{V}[\mathbf{X}_t|\mathcal{Y}_t]$
- Using the model we obtain predictions for time $t+1$:

$$\widehat{\mathbf{X}}_{t+1|t} = \mathbf{A}\widehat{\mathbf{X}}_{t|t} \quad \mathbf{\Sigma}^{xx}_{t+1|t} = \mathbf{A}\mathbf{\Sigma}^{xx}_{t|t}\mathbf{A}^T + \mathbf{G}\mathbf{\Sigma}_1\mathbf{G}^T$$

$$\widehat{\mathbf{Y}}_{t+1|t} = \mathbf{C}\widehat{\mathbf{X}}_{t+1|t} \quad \mathbf{\Sigma}^{yy}_{t+1|t} = \mathbf{C}\mathbf{\Sigma}^{xx}_{t+1|t}\mathbf{C}^T + \mathbf{\Sigma}_2$$

Due to the normality of the white noise process, $f(\mathbf{Y}_{t+1}|\mathcal{Y}_t, \boldsymbol{\theta})$ is a (multivariate) normal density with mean $\widehat{\mathbf{Y}}_{t+1|t}$ and variance-covariance $\mathbf{\Sigma}^{yy}t+1|t (= \mathbf{R}_{t+1})$

RECONSTRUCTION

At time t + 1 there is two possibilities:

The observation $\mathbf{Y_{t+1}}$ is available: We update the state estimate using the reconstruction step of the Kalman Filter:

$$K_{t+1} = \boldsymbol{\Sigma}^{xx}_{t+1|t} \mathbf{C}^T \left(\boldsymbol{\Sigma}^{yy}_{t+1|t} \right)^{-1}$$

$$\widehat{\mathbf{X}}_{t+1|t+1} = \widehat{\mathbf{X}}_{t+1|t} + \mathbf{K}_{t+1} \left(\mathbf{Y}_{t+1} - \widehat{\mathbf{Y}}_{t+1|t} \right)$$

$$\boldsymbol{\Sigma}^{xx}_{t+1|t+1} = \boldsymbol{\Sigma}^{xx}_{t+1|t} - \mathbf{K}_{t+1} \boldsymbol{\Sigma}^{yy}_{t+1|t} \mathbf{K}^T_{t+1}$$

The observation $\mathbf{Y_{t+1}}$ is missing: We got no new information and we use:

$$\widehat{\mathbf{X}}_{t+1|t+1} = \widehat{\mathbf{X}}_{t+1|t}$$

$$\mathbf{\Sigma}^{xx}_{t+1|t+1} = \mathbf{\Sigma}^{xx}_{t+1|t}$$

And then we predict for time $t+2$

MLE

Using the prediction errors and variances

$$\widetilde{\mathbf{Y}}_t = \mathbf{Y}_t - \widehat{\mathbf{Y}}_{t|t-1}$$

$$\mathbf{R}_t = \mathbf{\Sigma}^{yy}_{t|t-1}$$

The likelihood function can be expressed as

$$L(\theta; \mathcal{Y}_{N*}) = \prod_{t=1}^{N*} \left[(2\pi)^m \det \mathbf{R}_t\right]^{-\frac{1}{2}} exp\left[-\frac{1}{2}\widetilde{\mathbf{Y}}_t^T \mathbf{R}_t^{-1} \widetilde{\mathbf{Y}}_t\right]$$

- In practice optimization is based on $\log L(\theta; Y_{N*})$ and the variance of the estimates can be approximated by the 2'nd order derivatives of log-likelihood.

INITIALIZATION

- The only outstanding issue is “prediction” of $\mathbf{Y}_1$, i.e. calculation of $\widehat{\mathbf{Y}}_{1|0}$
- This can be done by setting $\widehat{\mathbf{X}}_{0|0} = \mathbf{0}$ and $\mathbf{\Sigma}^{xx}_{0|0} = \alpha\mathbf{I}$, where $\mathbf{I}$ is the identity matrix and α is a ’large’ constant (we don’t know what it is)
- Alternatively, we can estimate the initial state $\widehat{\mathbf{X}}_{0|0}$ and set $\mathbf{\Sigma}^{xx}_{0|0} = \mathbf{0}$, whereby $\mathbf{\Sigma}^{xx}_{1|0} = \mathbf{G}\mathbf{\Sigma}_1\mathbf{G}^T$