Time Series Analysis

Teaching session III : State Space Models, Kalman filtering Kalman Smoothing

Tohid Ardeshiri

Linköping University
Division of Statistics and Machine Learning

September 30, 2019

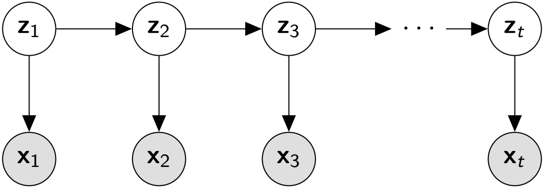
Remaining Course topics

- ARIMA models
- State space models (2 lectures, 1 teaching session with hand-in, 1 computer lab with short report)
  - Linear and Gaussian state space models (Chapter 6.1)
  - Kalman filtering, Kalman smoothing and Forecasting (Chapter 6.2)
  - Maximum likelihood estimate of the state space models (Chapter 6.3)
  - Stochastic volatility (Chapter 6.11)
- Recurrent Neural Networks (RNNs) (1 lecture and 1 Computer lab No examination)
- Summary lecture

State Space models - Linear and Gaussian

Our main focus will be on linear and Gaussian models:

$$\mathbf{z}_t = A\mathbf{z}_{t-1} + e_t, \qquad e_t \sim N(0, Q)$$
$$\mathbf{x}_t = C\mathbf{z}_t + \nu_t, \qquad \nu_t \sim N(0, R)$$



Bayesian Inference

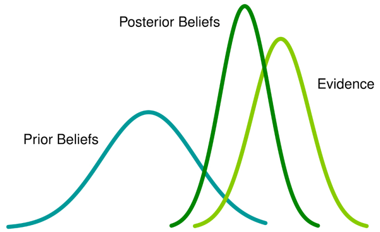
Bayesian inference is a means of combining prior beliefs with the data (evidence) to obtain posterior beliefs.

Example: likelihood update

$$f(\mathbf{z}|\mathbf{x}) \propto f(\mathbf{x}|\mathbf{z})f(\mathbf{z})$$

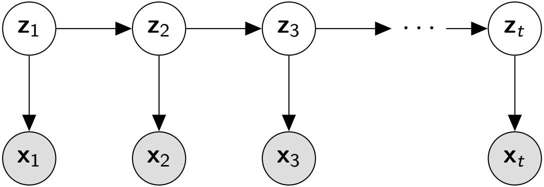
Probability Calculus

$$f(\mathbf{z},\mathbf{x}) = f(\mathbf{z}|\mathbf{x})f(\mathbf{x})$$
$$f(\mathbf{z},\mathbf{x}) = f(\mathbf{x}|\mathbf{z})f(\mathbf{z})$$
$$f(\mathbf{z}) = \int f(\mathbf{z},\mathbf{x})\, \mathrm{d}\mathbf{x}$$



Online recursive algorithms

Consider a stochastic dynamical system represented by the following recursion



$$\begin{aligned}\mathbf{z}_1 &\sim f(\mathbf{z}_1),\\ \mathbf{x}_k &\sim f(\mathbf{x}_k|\mathbf{z}_k),\\ \mathbf{z}_{k+1} &\sim f(\mathbf{z}_{k+1}|\mathbf{z}_k).\end{aligned}$$

The Bayesian filtering recursion corresponds to computing the posterior distributions $f(\mathbf{z}_k|\mathbf{x}_{1:k})$;

$$f(\mathbf{z}_k|\mathbf{x}_{1:k}) = \frac{f(\mathbf{z}_k|\mathbf{x}_{1:k-1})f(\mathbf{x}_k|\mathbf{z}_k)}{\int f(\mathbf{z}_k|\mathbf{x}_{1:k-1})f(\mathbf{x}_k|\mathbf{z}_k)\,\mathrm{d}\mathbf{z}_k}$$

The density $f(\mathbf{z}_k|\mathbf{x}_{1:k-1})$ in the numerator which is called the predicted density of $\mathbf{z}_k$ and is obtained by integration as in

$$f(\mathbf{z}_k|\mathbf{x}_{1:k-1}) = \int f(\mathbf{z}_k|\mathbf{z}_{k-1})f(\mathbf{z}_{k-1}|\mathbf{x}_{1:k-1})\,\mathrm{d}\mathbf{z}_{k-1}.$$

Kalman filter

Kalman filter is an algorithm that uses time series data, containing statistical noise and unknown innovations, and produces estimates of latent (hidden) process that tend to be more accurate than those based on a single observations using a probabilistic framework.

$$
\begin{aligned}
\mathbf{z}_t &= A\mathbf{z}_{t-1} + e_t, \\
\mathbf{x}_t &= C\mathbf{z}_t + \nu_t,
\end{aligned}
$$

The Kalman Filter's Foundation

Let $\mathbf{z}$ have a normal prior distribution with mean μ and covariance Σ , i.e., $\mathbf{z} \sim N(\mathbf{z}; \mu, \Sigma)$.

An observation $\mathbf{x}$ with the likelihood function $f(\mathbf{x}|\mathbf{z}) = N(\mathbf{x}; C\mathbf{z}, R)$ is in hand where C is a matrix with proper dimensions and R is a covariance matrix. The posterior distribution of $\mathbf{z}$ can be obtained using the Bayes' rule

$$
\begin{aligned}
f(\mathbf{z}|\mathbf{x}) =& \frac{f(\mathbf{z})f(\mathbf{x}|\mathbf{z})}{\int f(\mathbf{z})f(\mathbf{x}|\mathbf{z})\,\mathrm{d}\mathbf{z}} \\
=& \frac{N(\mathbf{z}; \mu, \Sigma)N(\mathbf{x}; C\mathbf{z}, R)}{\int N(\mathbf{z}; \mu, \Sigma)N(\mathbf{x}; C\mathbf{z}, R)\,\mathrm{d}\mathbf{z}}.
\end{aligned}
$$

The posterior distribution $f(\mathbf{z}|\mathbf{x})$ has an analytical solution and turns out to be the normal distribution $N(\mathbf{z}; \mu', \Sigma')$ where

$$
\begin{aligned}
\mu' =& \mu + K(\mathbf{x} - C\mu), \\
\Sigma' =& \Sigma - KC\Sigma,
\end{aligned}
$$

where

$$
K = \Sigma C^{\mathrm{T}}(C\Sigma C^{\mathrm{T}} + R)^{-1}.
$$

Properties of the normal density function

Property 1: $f(\mathbf{y}_1)f(\mathbf{y}_2|\mathbf{y}_1) = f(\mathbf{y}_1, \mathbf{y}_2)$

$$N(\mathbf{y}_1; \mu, \Sigma)N(\mathbf{y}_2; C\mathbf{y}_1, R) = N\left(\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix}; \begin{bmatrix}\mu\\ C\mu\end{bmatrix}, \begin{bmatrix}\Sigma & \Sigma C^{\mathrm{T}}\\ C\Sigma & C\Sigma C^{\mathrm{T}} + R\end{bmatrix}\right)$$

Property 2: marginalization and conditioning

If $\mathbf{y}_1$, $\mathbf{y}_2$ were jointly normal:

$$f(\mathbf{y}_1, \mathbf{y}_2) = N\left(\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix}; \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\right)$$

then

$$\begin{aligned}
f(\mathbf{y}_1) &= N(\mathbf{y}_1; \mu_1, \Sigma_{11})\\
f(\mathbf{y}_2) &= N(\mathbf{y}_2; \mu_2, \Sigma_{22})\\
f(\mathbf{y}_1|\mathbf{y}_2) &= N(\mathbf{y}_1; \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{y}_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\\
f(\mathbf{y}_2|\mathbf{y}_1) &= N(\mathbf{y}_2; \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{y}_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})
\end{aligned}$$

Kalman filter's derivation

Consider State space model

$$\mathbf{z}_t = A\mathbf{z}_{t-1} + e_t,$$
$$\mathbf{x}_t = C\mathbf{z}_t + \nu_t.$$

And initial prior on the state $\mathbf{z}_1$

$$f(\mathbf{z}_1) = N(\mathbf{z}_1; m_0, P_0)$$

We want to derive a recursive algorithm to compute the posterior filtering density

$$f(\mathbf{z}_t | \mathbf{x}_{1:t}).$$

That is, computing the the posterior density of $\mathbf{z}_t$ using the observations up to time t.

Kalman filter's derivation

Assume that we have

$$f(\mathbf{z}_t|\mathbf{x}_{1:t}) = N(\mathbf{z}_t; m_{t|t}, P_{t|t}).$$

The state transition density $f(\mathbf{z}_{t+1}|\mathbf{z}_t)$ and the likelihood function $f(\mathbf{x}_{t+1}|\mathbf{z}_{t+1})$ can be written as

$$\begin{aligned} f(\mathbf{z}_{t+1}|\mathbf{z}_t) &= N(\mathbf{z}_{t+1}; A\mathbf{z}_t, Q), \\ f(\mathbf{x}_{t+1}|\mathbf{z}_{t+1}) &= N(\mathbf{x}_{t+1}; C\mathbf{z}_{t+1}, R). \end{aligned}$$

Therefore, the joint posterior $f(\mathbf{z}_t, \mathbf{z}_{t+1}, \mathbf{x}_{t+1}|\mathbf{x}_{1:t})$ can be written as

$$\begin{aligned} f(\mathbf{z}_t, \mathbf{z}_{t+1}, \mathbf{x}_{t+1}|\mathbf{x}_{1:t}) =& N(\mathbf{z}_t; m_{t|t}, P_{t|t}) \\ & \times N(\mathbf{z}_{t+1}; A\mathbf{z}_t, Q) N(\mathbf{x}_{t+1}; C\mathbf{z}_{t+1}, R), \end{aligned}$$

Kalman filter's derivation

The $f(\mathbf{z}_t, \mathbf{z}_{t+1}, \mathbf{x}_{t+1} | \mathbf{x}_{1:t})$ can be rewritten in matrix form as

$$f(\mathbf{z}_t, \mathbf{z}_{t+1}, \mathbf{x}_{t+1} | \mathbf{x}_{1:t}) = N([\mathbf{z}_t^{\mathrm{T}}, \mathbf{z}_{t+1}^{\mathrm{T}}, \mathbf{x}_{t+1}^{\mathrm{T}}]^{\mathrm{T}}; \mu_t, \Sigma_t),$$

where

$$\mu_t = \left[\begin{array}{c} \mu_1 \\ \hline \mu_2 \end{array}\right] = \left[\begin{array}{c} m_{t|t} \\ Am_{t|t} \\ \hline CAm_{t|t} \end{array}\right]$$

and

$$\Sigma_t \triangleq \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right] =$$

$$\left[\begin{array}{cc|c} P_{t|t} & P_{t|t}A^{\mathrm{T}} & (P_{t|t}A^T)C^T \\ AP_{t|t} & AP_{t|t}A^{\mathrm{T}} + Q & (AP_{t|t}A^{\mathrm{T}} + Q)^T C^T \\ \hline C(AP_{t|t}) & C(AP_{t|t}A^{\mathrm{T}} + Q) & C(AP_{t|t}A^{\mathrm{T}} + Q)C^{\mathrm{T}} + R \end{array}\right].$$

Kalman filtering algorithm

Prove the Kalman filtering recursion for the following state space model with initial prior on the state $f(\mathbf{z}_1) = N(\mathbf{z}_1; m_0, P_0)$

$$\begin{aligned} \mathbf{z}_t &= A_{t-1}\mathbf{z}_{t-1} + e_t, & e_t &\sim N(0, Q_t) \\ \mathbf{x}_t &= C_t \mathbf{z}_t + \nu_t, & \nu_t &\sim N(0, R_t) \end{aligned}$$

Inputs: A_t, C_t, Q_t, R_t, m_0, P_0 and $\mathbf{x}_{1:T}$.
initialization
$m_{1|0} \leftarrow m_0$, $P_{1|0} \leftarrow P_0$
for $t = 1$ to T **do**
    observation update step
    $K_t \leftarrow P_{t|t-1} C_t^{\mathrm{T}} (C_t P_{t|t-1} C_t^{\mathrm{T}} + R_t)^{-1}$
    $m_{t|t} \leftarrow m_{t|t-1} + K_t(\mathbf{x}_t - C_t m_{t|t-1})$
    $P_{t|t} \leftarrow (I - K_t C_t) P_{t|t-1}$
    prediction step
    $m_{t+1|t} \leftarrow A_t m_{t|t}$
    $P_{t+1|t} \leftarrow A_t P_{t|t} A_t^{\mathrm{T}} + Q_{t+1}$
end for
Outputs: $m_{t|t}$, $P_{t|t}$ for $t = 1 : T$

Bayesian Smoothing

The purpose of Bayesian smoothing is to compute the marginal posterior distribution of $\mathbf{z}_t$ at time t after receiving observations up to time T where $T > t$:

$$f(\mathbf{z}_t|\mathbf{x}_{1:T})$$

The Rauch-Tung-Striebel smoother (RTS smoother) which is also called the Kalman smoother is used to compute

$$f(\mathbf{z}_t|\mathbf{x}_{1:T}) = N(\mathbf{z}_t; m_{t|T}, P_{t|T})$$

The RTS smoother uses a Kalman filter in its forward path. In its backwards path it updates the densities using the relation

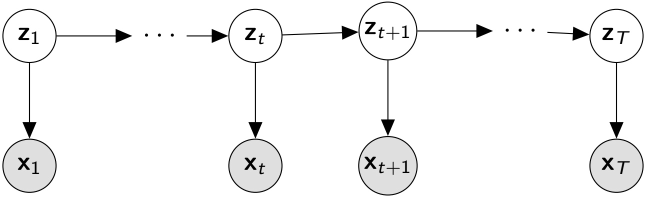
$$\mathbf{z}_t = A_{t-1}\mathbf{z}_{t-1} + e_t$$

RTS Smoother's derivation

Assume $f(\mathbf{z}_{t+1}|\mathbf{x}_{1:T})$ is available as in

$$f(\mathbf{z}_{t+1}|\mathbf{x}_{1:T}) = N(\mathbf{z}_{t+1}; m_{t+1|T}, P_{t+1|T})$$

For example $f(\mathbf{z}_T|\mathbf{x}_{1:T})$ which is the filtering density of $\mathbf{z}_T$ is available after filtering.



The objective is to compute $f(\mathbf{z}_t, \mathbf{z}_{t+1}|\mathbf{x}_{1:T})$.

RTS Smoother's derivation

The joint posterior $f(\mathbf{z}_t, \mathbf{z}_{t+1}|\mathbf{x}_{1:t})$ can be written as

$$
\begin{aligned}
f(\mathbf{z}_t, \mathbf{z}_{t+1}|\mathbf{x}_{1:t}) =& N(\mathbf{z}_t; m_{t|t}, P_{t|t}) N(\mathbf{z}_{t+1}; A\mathbf{z}_t, Q) \\
=& N\left(\begin{bmatrix} \mathbf{z}_t \\ \mathbf{z}_{t+1} \end{bmatrix}, \begin{bmatrix} m_{t|t} \\ Am_{t|t} \end{bmatrix}, \begin{bmatrix} P_{t|t} & P_{t|t}A^{\mathrm{T}} \\ AP_{t|t} & AP_{t|t}A^{\mathrm{T}} + Q \end{bmatrix} \right)
\end{aligned}
$$

Using the conditioning property of the multivariate normal distribution $f(\mathbf{z}_t|\mathbf{z}_{t+1}, \mathbf{x}_{1:t})$ can be computed as a normal density as given in the following:

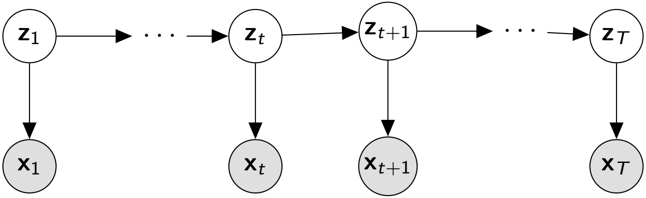
$$
f(\mathbf{z}_t|\mathbf{z}_{t+1}, \mathbf{x}_{1:t}) = N(\mathbf{z}_t; \tilde{m}_t, \tilde{P}_t)
$$

where $\tilde{m}_t$ is a function of $\mathbf{z}_{t+1}$.

RTS Smoother's derivation

Note the Markov property

$$f(\mathbf{z}_t|\mathbf{z}_{t+1}, \mathbf{x}_{1:T}) = f(\mathbf{z}_t|\mathbf{z}_{t+1}, \mathbf{x}_{1:t})$$



Assume $f(\mathbf{z}_{t+1}|\mathbf{x}_{1:T})$ is available as in

$$f(\mathbf{z}_{t+1}|\mathbf{x}_{1:T}) = N(\mathbf{z}_{t+1}; m_{t+1|T}, P_{t+1|T})$$

Recall that

$$\begin{aligned} f(\mathbf{z}_{t+1}, \mathbf{z}_t|\mathbf{x}_{1:T}) &= f(\mathbf{z}_{t+1}|\mathbf{x}_{1:T}) {\color{red} f(\mathbf{z}_t|\mathbf{z}_{t+1}, \mathbf{x}_{1:T})} \\ &= f(\mathbf{z}_{t+1}|\mathbf{x}_{1:T}) {\color{red} f(\mathbf{z}_t|\mathbf{z}_{t+1}, \mathbf{x}_{1:t})} \\ &= N(\mathbf{z}_{t+1}; m_{t+1|T}, P_{t+1|T}) N(\mathbf{z}_t; \tilde{m}_t, \tilde{P}_t) \end{aligned}$$

RTS Smoother's derivation

where

$$
\begin{aligned}
G_t &= P_{t|t} A_t^{\mathrm{T}} (A P_{t|t} A^{\mathrm{T}} + Q)^{-1} = P_{t|t} A_t^{\mathrm{T}} P_{t+1|t}^{-1} \\
\tilde{m}_t &= m_{t|t} + G_t (x_{t+1} - A m_{t|t}) \\
\tilde{P}_t &= P_{t|t} - G_t (A P_{t|t} A^{\mathrm{T}} + Q) G_t^{\mathrm{T}} = P_{t|t} - G_t P_{t+1|t} G_t^{\mathrm{T}}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
f(\mathbf{z}_{t+1}, \mathbf{z}_t | \mathbf{x}_{1:T}) &= N(\mathbf{z}_{t+1}; m_{t+1|T}, P_{t+1|T}) N(\mathbf{z}_t; \tilde{m}_t, \tilde{P}_t) \\
&= N\left(\begin{bmatrix} \mathbf{z}_t \\ \mathbf{z}_{t+1} \end{bmatrix}, \begin{bmatrix} \cdot \\ m_{t+1|T} \end{bmatrix}, \begin{bmatrix} \cdot & \cdot \\ \cdot & P_{t+1|T} \end{bmatrix} \right)
\end{aligned}
$$

RTS smoother's backwards recursion

Prove the backwards recursion of the RTS smoother for following state space model with initial prior on the state $f(\mathbf{z}_1) = N(\mathbf{z}_1; m_0, P_0)$

$$\mathbf{z}_t = A_{t-1}\mathbf{z}_{t-1} + e_t, \qquad e_t \sim N(0, Q_t)$$
$$\mathbf{x}_t = C_t\mathbf{z}_t + \nu_t, \qquad \nu_t \sim N(0, R_t)$$

```
Inputs: A_t, Q_t, m_{t|t}, P_{t|t}, m_{t+1|t}, P_{t+1|t} for 1 ≤ t ≤ T
initialization
for t = T-1 down to 1 do
    G_t ← P_{t|t} A_t^T P_{t+1|t}^{-1}
    m_{t|T} ← m_{t|t} + G_t (m_{t+1|T} − A_t m_{t|t})
    P_{t|T} ← P_{t|t} + G_t (P_{t+1|T} − P_{t+1|t}) G_t^T
end for
Outputs: m_{t|T}, P_{t|T}
```

Inputs: A_t, Q_t, $m_{t|t}$, $P_{t|t}$, $m_{t+1|t}$, $P_{t+1|t}$ for $1 \leq t \leq T$
initialization
for $t =$ T-1 down to 1 **do**
  $G_t \leftarrow P_{t|t}A_t^{\mathrm{T}}P_{t+1|t}^{-1}$
  $m_{t|T} \leftarrow m_{t|t} + G_t(m_{t+1|T} - A_t m_{t|t})$
  $P_{t|T} \leftarrow P_{t|t} + G_t(P_{t+1|T} - P_{t+1|t})G_t^{\mathrm{T}}$
end for
Outputs: $m_{t|T}$, $P_{t|T}$

Read home

- Shumway and Stoffer, Chapters 6.1 and 6.2