STAT 576 Bayesian Analysis

Lecture 10: State-space Models and Sequential Monte Carlo I

Chencheng Cai

Washington State University

State-space Models

The **state-space model** is a general framework for modeling time series data. It consists of two components:

- The **state equation**: describes the evolution of the latent state variables over time.
- The **observation equation**: describes the relationship between the latent state variables and the observed data.

State-space Models

The **state-space model** is a general framework for modeling time series data. It consists of two components:

- The **state equation**: describes the evolution of the latent state variables over time.
- The **observation equation**: describes the relationship between the latent state variables and the observed data.
- The state-space model is also known as the **hidden Markov model (HMM)** when the state space is finite and the process is Markovian.

State-space Models

$$
\begin{tikzcd}
 & Y_1 & Y_2 & & Y_t & & Y_{T-1} & Y_T \\
X_0 \arrow[r, "f_1"] & X_1 \arrow[u, "g_1"] \arrow[r, "f_2"] & X_2 \arrow[u, "g_2"] \arrow[r, "f_3"] & \cdots \arrow[r, "f_t"] & X_t \arrow[u, "g_t"] \arrow[r, "f_{t+1}"] & \cdots \arrow[r, "f_{T-1}"] & X_{T-1} \arrow[u, "g_{T-1}"] \arrow[r, "f_T"] & X_T \arrow[u, "g_T"]
\end{tikzcd}
$$

State-space Models

$$
\begin{array}{ccccccccccccccc}
 & & Y_1 & & Y_2 & & & & Y_t & & & & Y_{T-1} & & Y_T \\
 & & \uparrow{\scriptstyle g_1} & & \uparrow{\scriptstyle g_2} & & & & \uparrow{\scriptstyle g_t} & & & & \uparrow{\scriptstyle g_{T-1}} & & \uparrow{\scriptstyle g_T} \\
X_0 & \xrightarrow{f_1} & X_1 & \xrightarrow{f_2} & X_2 & \xrightarrow{f_3} & \cdots & \xrightarrow{f_t} & X_t & \xrightarrow{f_{t+1}} & \cdots & \xrightarrow{f_{T-1}} & X_{T-1} & \xrightarrow{f_T} & X_T
\end{array}
$$

- Observed data: $\boldsymbol{Y} = (Y_1, \ldots, Y_T)$
- Latent states: $\boldsymbol{X} = (X_0, X_1, \ldots, X_T)$

State-space Models

$$
\begin{tikzcd}
 & Y_1 & Y_2 & & Y_t & & Y_{T-1} & Y_T \\
X_0 \arrow[r, "f_1"] & X_1 \arrow[u, "g_1"] \arrow[r, "f_2"] & X_2 \arrow[u, "g_2"] \arrow[r, "f_3"] & \cdots \arrow[r, "f_t"] & X_t \arrow[u, "g_t"] \arrow[r, "f_{t+1}"] & \cdots \arrow[r, "f_{T-1}"] & X_{T-1} \arrow[u, "g_{T-1}"] \arrow[r, "f_T"] & X_T \arrow[u, "g_T"]
\end{tikzcd}
$$

- Observed data: $\boldsymbol{Y} = (Y_1, \ldots, Y_T)$
- Latent states: $\boldsymbol{X} = (X_0, X_1, \ldots, X_T)$
- The state equation:

$$p(X_0) = f_0(X_0), \quad p(X_t \mid \boldsymbol{X}_{t-1}) = f_t(X_t \mid \boldsymbol{X}_{t-1})$$

- The observation equation:

$$p(Y_t \mid \boldsymbol{X}_t) = g_t(Y_t \mid X_t)$$

State-space Model

- If the state equation satisfies

$$p(X_t \mid \boldsymbol{X}_{t-1}) = p(X_t \mid X_{t-1})$$

then the state-space model is **Markovian**.

State-space Model

- If the state equation satisfies

$$p(X_t \mid \boldsymbol{X}_{t-1}) = p(X_t \mid X_{t-1})$$

  then the state-space model is **Markovian**.
- The (Markovian) state-space model is **linear** if

$$\mathbb{E}[X_t \mid X_{t-1}] = \boldsymbol{A}_t X_{t-1}$$

  and

$$\mathbb{E}[Y_t \mid X_t] = \boldsymbol{B}_t X_t,$$

  for some matrices $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$.

State-space Model

- If the state equation satisfies

$$p(X_t \mid \boldsymbol{X}_{t-1}) = p(X_t \mid X_{t-1})$$

  then the state-space model is **Markovian**.
- The (Markovian) state-space model is **linear** if

$$\mathbb{E}[X_t \mid X_{t-1}] = \boldsymbol{A}_t X_{t-1}$$

  and

$$\mathbb{E}[Y_t \mid X_t] = \boldsymbol{B}_t X_t,$$

  for some matrices $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$.
- The (Markovian) state-space model is **linear Gaussian** if

$$X_t \mid X_{t-1} \sim \mathcal{N}(\boldsymbol{A}_t X_{t-1}, \boldsymbol{\Sigma}_t) \text{ and } Y_t \mid X_t \sim \mathcal{N}(\boldsymbol{A}_t X_t, \boldsymbol{R}_t)$$

Example: Object Tracking

- Consider the problem that tracks the position of an object moving in a 2D plane.
- The data contains the observed positions (with noise) of the object at different time points. $Y_t = (a_t, b_t)^T$.

Example: Object Tracking

- Consider the problem that tracks the position of an object moving in a 2D plane.
- The data contains the observed positions (with noise) of the object at different time points. $Y_t = (a_t, b_t)^T$.
- We can assume the latent states $X_t = (x_t, y_t)$, the true positions of the object.
- The observation equation is

  $$Y_t = X_t + \epsilon_t$$

  where $\epsilon_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R})$. $\boldsymbol{R}$ is the accuracy of the sensor.
- For the latent states X_t, we can assume a linear Gaussian model (random walk):

  $$X_t = X_{t-1} + \eta_t,$$

  where $\eta_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma}$ is the process noise.

Example: Object Tracking

The previous model has a continuous path, but quite stochastic velocities. We can add a velocity component to the model to stablize the dynamics.

- The latent states $X_t = (x_t, y_t, v_t, u_t)$, where (x_t, y_t) is the position and (v_t, u_t) is the velocity.
- The observation equation is

$$Y_t = (x_t, y_t)^T + \epsilon_t$$

  where $\epsilon_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R})$.

Example: Object Tracking

The previous model has a continuous path, but quite stochastic velocities. We can add a velocity component to the model to stablize the dynamics.

- The latent states $X_t = (x_t, y_t, v_t, u_t)$, where (x_t, y_t) is the position and (v_t, u_t) is the velocity.
- The observation equation is

$$Y_t = (x_t, y_t)^T + \epsilon_t$$

  where $\epsilon_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R})$.
- The state equation is

$$\begin{aligned} x_t &= x_{t-1} + v_{t-1} \\ y_t &= y_{t-1} + u_{t-1} \\ v_t &= v_{t-1} + \eta_t \\ u_t &= u_{t-1} + \xi_t, \end{aligned}$$

  where $\eta_t, \xi_t \sim \mathcal{N}(0, \sigma^2)$.

Example: Object Tracking

The previous model is a linear Gaussian model. We can write it in the matrix form:

$$\boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{\eta}_t$$
$$\boldsymbol{Y}_t = \boldsymbol{B}\boldsymbol{X}_t + \boldsymbol{\epsilon}_t,$$

where

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$
$$\boldsymbol{B} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$
$$\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R})$$
$$\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}).$$

The Probabilities

The state-space model is a full probabilistic model.

- The joint distribution of the latent states and the observed data is

$$p(\boldsymbol{X}, \boldsymbol{Y}) = p(X_0) \prod_{t=1}^{T} p(X_t \mid \boldsymbol{X}_{t-1}) p(Y_t \mid X_t) = f_0(X_0) \prod_{t=1}^{T} f_t(X_t \mid \boldsymbol{X}_{t-1}) g_t(Y_t \mid X_t)$$

The Probabilities

The state-space model is a full probabilistic model.

- The joint distribution of the latent states and the observed data is

$$p(\boldsymbol{X}, \boldsymbol{Y}) = p(X_0) \prod_{t=1}^{T} p(X_t \mid \boldsymbol{X}_{t-1}) p(Y_t \mid X_t) = f_0(X_0) \prod_{t=1}^{T} f_t(X_t \mid \boldsymbol{X}_{t-1}) g_t(Y_t \mid X_t)$$

- The joint distribution of the latent states is

$$p(\boldsymbol{X}) = p(X_0) \prod_{t=1}^{T} p(X_t \mid \boldsymbol{X}_{t-1}) = f_0(X_0) \prod_{t=1}^{T} f_t(X_t \mid \boldsymbol{X}_{t-1})$$

The Probabilities

The state-space model is a full probabilistic model.

- The joint distribution of the latent states and the observed data is

$$p(\boldsymbol{X}, \boldsymbol{Y}) = p(X_0) \prod_{t=1}^{T} p(X_t \mid \boldsymbol{X}_{t-1}) p(Y_t \mid X_t) = f_0(X_0) \prod_{t=1}^{T} f_t(X_t \mid \boldsymbol{X}_{t-1}) g_t(Y_t \mid X_t)$$

- The joint distribution of the latent states is

$$p(\boldsymbol{X}) = p(X_0) \prod_{t=1}^{T} p(X_t \mid \boldsymbol{X}_{t-1}) = f_0(X_0) \prod_{t=1}^{T} f_t(X_t \mid \boldsymbol{X}_{t-1})$$

- The joint distribution of the observed data is

$$p(\boldsymbol{Y}) = \int p(\boldsymbol{X}, \boldsymbol{Y})\, d\boldsymbol{X} = \int f_0(X_0) \prod_{t=1}^{T} f_t(X_t \mid \boldsymbol{X}_{t-1}) g_t(Y_t \mid X_t) d\boldsymbol{X}$$

Bayesian Framework

- The prior:

$$p(\boldsymbol{X}) = f_0(X_0) \prod_{t=1}^{T} f_t(X_t \mid \boldsymbol{X}_{t-1})$$

Bayesian Framework

- The prior:

$$p(\boldsymbol{X}) = f_0(X_0) \prod_{t=1}^{T} f_t(X_t \mid \boldsymbol{X}_{t-1})$$

- The likelihood:

$$p(\boldsymbol{Y} \mid \boldsymbol{X}) = \prod_{t=1}^{T} g_t(Y_t \mid X_t)$$

Bayesian Framework

- The prior:

$$p(\boldsymbol{X}) = f_0(X_0) \prod_{t=1}^{T} f_t(X_t \mid \boldsymbol{X}_{t-1})$$

- The likelihood:

$$p(\boldsymbol{Y} \mid \boldsymbol{X}) = \prod_{t=1}^{T} g_t(Y_t \mid X_t)$$

- The posterior:

$$p(\boldsymbol{X} \mid \boldsymbol{Y}) = \frac{p(\boldsymbol{X}, \boldsymbol{Y})}{p(\boldsymbol{Y})} \propto p(\boldsymbol{X}, \boldsymbol{Y}) = f_0(X_0) \prod_{t=1}^{T} f_t(X_t \mid \boldsymbol{X}_{t-1}) g_t(Y_t \mid X_t)$$

Bayesian Framework

- The prior:

$$p(\boldsymbol{X}) = f_0(X_0) \prod_{t=1}^{T} f_t(X_t \mid \boldsymbol{X}_{t-1})$$

- The likelihood:

$$p(\boldsymbol{Y} \mid \boldsymbol{X}) = \prod_{t=1}^{T} g_t(Y_t \mid X_t)$$

- The posterior:

$$p(\boldsymbol{X} \mid \boldsymbol{Y}) = \frac{p(\boldsymbol{X}, \boldsymbol{Y})}{p(\boldsymbol{Y})} \propto p(\boldsymbol{X}, \boldsymbol{Y}) = f_0(X_0) \prod_{t=1}^{T} f_t(X_t \mid \boldsymbol{X}_{t-1}) g_t(Y_t \mid X_t)$$

Direct sampling from this posterior distribution can be difficult. We need to utilize the **sequential** structure of the model.

The Sequential Structure

Suppose we are at time t.

- We have observed data $Y_1, \ldots, Y_t$.
- We have the latent states $X_0, \ldots, X_t$.

The Sequential Structure

Suppose we are at time t.

- We have observed data $Y_1, \ldots, Y_t$.
- We have the latent states $X_0, \ldots, X_t$.
- The **sequential** joint prior for the latent states is

$$p(\boldsymbol{X}_t) = f_t(X_0) \prod_{s=1}^{t} f_s(X_s \mid \boldsymbol{X}_{s-1})$$

The Sequential Structure

Suppose we are at time t.

- We have observed data $Y_1, \ldots, Y_t$.
- We have the latent states $X_0, \ldots, X_t$.
- The **sequential** joint prior for the latent states is

$$p(\boldsymbol{X}_t) = f_t(X_0) \prod_{s=1}^{t} f_s(X_s \mid \boldsymbol{X}_{s-1})$$

- The **sequential** likelihood for the observed data (so far) is

$$p(\boldsymbol{Y}_t \mid \boldsymbol{X}_t) = \prod_{s=1}^{t} g_t(Y_t \mid X_t)$$

The Sequential Structure

Suppose we are at time t.

- We have observed data $Y_1, \ldots, Y_t$.
- We have the latent states $X_0, \ldots, X_t$.
- The **sequential** joint prior for the latent states is

$$p(\boldsymbol{X}_t) = f_t(X_0) \prod_{s=1}^{t} f_s(X_s \mid \boldsymbol{X}_{s-1})$$

- The **sequential** likelihood for the observed data (so far) is

$$p(\boldsymbol{Y}_t \mid \boldsymbol{X}_t) = \prod_{s=1}^{t} g_t(Y_t \mid X_t)$$

- The **sequential** posterior for the latent states up to time t is (also called the **filtering** distribution)

$$p(\boldsymbol{X}_t \mid \boldsymbol{Y}_t) \propto f_t(X_0) \prod_{s=1}^{t} f_s(X_s \mid \boldsymbol{X}_{s-1}) g_t(Y_t \mid X_t)$$

The Sequential Structure

At time t,

- The **predictive** distribution for the latent state at time $t+1$ is

$$p(X_{t+1} \mid \boldsymbol{Y}_t) = \int p(X_{t+1} \mid X_t) p(X_t \mid \boldsymbol{Y}_t) \, dX_t$$

The Sequential Structure

At time t,

- The **predictive** distribution for the latent state at time $t+1$ is

$$p(X_{t+1} \mid \boldsymbol{Y}_t) = \int p(X_{t+1} \mid X_t) p(X_t \mid \boldsymbol{Y}_t)\, dX_t$$

- The joint distribution of the latent states up to time $t+1$ is

$$\begin{aligned} p(\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_t) &= p(X_{t+1} \mid \boldsymbol{Y}_t) p(\boldsymbol{X}_t \mid \boldsymbol{Y}_t) \\ &\propto f_{t+1}(X_{t+1} \mid \boldsymbol{X}_t) f_t(X_0) \prod_{s=1}^{t} f_s(X_s \mid \boldsymbol{X}_{s-1}) g_t(Y_t \mid X_t) \end{aligned}$$

The Sequential Structure

At time t,

- The **predictive** distribution for the latent state at time $t+1$ is

$$p(X_{t+1} \mid \boldsymbol{Y}_t) = \int p(X_{t+1} \mid X_t) p(X_t \mid \boldsymbol{Y}_t) \, dX_t$$

- The joint distribution of the latent states up to time $t+1$ is

$$\begin{aligned} p(\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_t) &= p(X_{t+1} \mid \boldsymbol{Y}_t) p(\boldsymbol{X}_t \mid \boldsymbol{Y}_t) \\ &\propto f_{t+1}(X_{t+1} \mid \boldsymbol{X}_t) f_t(X_0) \prod_{s=1}^{t} f_s(X_s \mid \boldsymbol{X}_{s-1}) g_t(Y_t \mid X_t) \end{aligned}$$

- The incremental likelihood for the observed data at time $t+1$ is

$$p(Y_{t+1} \mid \boldsymbol{X}_{t+1}) = g_{t+1}(Y_{t+1} \mid X_{t+1})$$

The Sequential Structure

At time t,

- The **predictive** distribution for the latent state at time $t+1$ is

$$p(X_{t+1} \mid \boldsymbol{Y}_t) = \int p(X_{t+1} \mid X_t) p(X_t \mid \boldsymbol{Y}_t)\, dX_t$$

- The joint distribution of the latent states up to time $t+1$ is

$$\begin{aligned} p(\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_t) &= p(X_{t+1} \mid \boldsymbol{Y}_t) p(\boldsymbol{X}_t \mid \boldsymbol{Y}_t) \\ &\propto f_{t+1}(X_{t+1} \mid \boldsymbol{X}_t) f_t(X_0) \prod_{s=1}^{t} f_s(X_s \mid \boldsymbol{X}_{s-1}) g_t(Y_t \mid X_t) \end{aligned}$$

- The incremental likelihood for the observed data at time $t+1$ is

$$p(Y_{t+1} \mid \boldsymbol{X}_{t+1}) = g_{t+1}(Y_{t+1} \mid X_{t+1})$$

- The filtering distribution for the latent states up to time $t+1$ is

$$\prod^{t+1}$$

The Sequential Structure

The sequential structure of the state-space model allows us to update the latent states one by one.

- $p(\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_t)$ is the prior
- $p(Y_{t+1} \mid X_{t+1})$ is the likelihood
- $p(\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_{t+1})$ is the posterior

The Sequential Structure

The sequential structure of the state-space model allows us to update the latent states one by one.

- $p(\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_t)$ is the prior
- $p(Y_{t+1} \mid X_{t+1})$ is the likelihood
- $p(\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_{t+1})$ is the posterior

A rudiment of sequential Monte Carlo:

- If we have a sample from $\boldsymbol{X}_t \mid \boldsymbol{Y}_t$.
- We can draw a sample from $\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_t$ by drawing X_{t+1} from $p(X_{t+1} \mid \boldsymbol{X}_t)$.
- We can update the sample to $\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_{t+1}$ by adjusting its weight according $p(Y_{t+1} \mid X_{t+1})$.

The Sequential Structure

The sequential structure of the state-space model allows us to update the latent states one by one.

- $p(\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_t)$ is the prior
- $p(Y_{t+1} \mid X_{t+1})$ is the likelihood
- $p(\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_{t+1})$ is the posterior

A rudiment of sequential Monte Carlo:

- If we have a sample from $\boldsymbol{X}_t \mid \boldsymbol{Y}_t$.
- We can draw a sample from $\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_t$ by drawing X_{t+1} from $p(X_{t+1} \mid \boldsymbol{X}_t)$.
- We can update the sample to $\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_{t+1}$ by adjusting its weight according $p(Y_{t+1} \mid X_{t+1})$.

Remark:

- The distribution $p(\boldsymbol{X}_t \mid \boldsymbol{Y}_t)$ is called the **filtering** distribution.
- The distribution $p(\boldsymbol{X}_t \mid \boldsymbol{Y})$ is called the **smoothing** distribution.