STA 4103/5107 Computational Methods in Statistics II

Department of Statistics
Florida State University

Class 10
February 9, 2017

Chapter 10

Analysis of Dynamic Systems

Basic Ideas

- We have studied the problem of sampling from a probability distribution, using a homogeneous Markov with certain properties.
- Now we consider a similar problem of sampling with a difference that the probability distribution is changing in time.
- In other words, we need to sample from an indexed family of probability distributions, indexed by time.
- Assuming that these probability distribution evolve in time according to some specific transitions, then this task can be simplified greatly.

Basic Ideas

- Let $\{f_t\}$ be a sequence of probability densities that we want to sample from.
- For each t, and let S_t be a collection of samples from f_t.
- A major simplification in the sampling process can be obtained if one can use samples S_t of f_t to obtain samples S_{t+1} of f_{t+1}, rather than sampling from the start.
- This is the topic of study in this chapter.

Time Series

- This topic relates closely to a broader class of problems where one uses observations from a time-series to estimate the underlying process.

- Let $x_1, x_2, \ldots, x_t, x_{t+1}, \ldots,$ form a process of interest, and instead of measuring x_ts directly, one observes variables $y_1, y_2, \ldots, y_t, y_{t+1}, \ldots$

- The goal is to use the observations, and joint probability models of x and y to estimate the unknown x_ts.

- Let $f(x_t \mid y_1, y_2, \ldots, y_\tau)$ be the posterior density function of x_t given a set of observations $(y_1, y_2, \ldots, y_\tau)$.

Three Estimations

- **1. Smoothing:** In this case the goal is to estimate a state x_t using observations up to a time $\tau > t$. In other words,

$$\hat{x}_t = \arg\max_{x_t} f(x_t \mid y_1, \ldots, y_t, \ldots, y_\tau), \text{or } \hat{x}_t = E_f(x_t \mid y_1, \ldots, y_t, \ldots, y_\tau).$$

- **2. Filtering:** In this case the goal is to estimate a state x_t using observations up to t. In other words,

$$\hat{x}_t = \arg\max_{x_t} f(x_t \mid y_1, y_2, \ldots, y_t), \text{or } \hat{x}_t = E_f(x_t \mid y_1, y_2, \ldots, y_t).$$

- **3. Prediction:** In this case the goal is to estimate a state x_t using observations up to a time $\tau < t$. In other words,

$$\hat{x}_t = \arg\max_{x_t} f(x_t \mid y_1, y_2, \ldots, y_\tau), \text{or } \hat{x}_t = E_f(x_t \mid y_1, y_2, \ldots, y_\tau).$$

10.1 Nonlinear Filtering Problem

Nonlinear Filtering Problem

- Let the *state vector* $x_t \in \mathbf{R}^d$ and the observation vector $y_t \in \mathbf{R}^c$.
- Assume that state vectors and observation vectors satisfy the following equations:

$$x_t = \mathrm{F}(x_{t-1}) + w_t \qquad (1)$$

$$y_t = \mathrm{G}(x_t) + q_t \qquad (2)$$

- Eqn. 1 is called the **state equation (prior model)** and it models the evolution of state in time. Here, F: $\mathbf{R}^d \rightarrow \mathbf{R}^d$ is a given function and $w_t \in \mathbf{R}^d$ is a vector of noise variables.
- Assuming that w_t s are statistically independent, the resulting process x_t is a Markov process.

State and Observation Equations

- That is:

$$f(x_t \mid x_1, x_2, \ldots, x_{t-1}) = f(x_t \mid x_{t-1})$$

- Eqn. 2 is called the **observation equation (likelihood model)** and it relates the underlying state x_t with the observation y_t. Here, G: $\mathbf{R}^d \rightarrow \mathbf{R}^c$ is a given function and $q_t \in \mathbf{R}^c$ is the vector of observation noise.

- It is assumed that the observation noise q_t is statistically independent across times, i.e. q_t is independent of q_s for $t \neq s$, and it is also independent of x_t and w_t. Therefore,

$$f(y_t \mid y_1, y_2, \ldots, y_{t-1}, \mathrm{x}_t) = f(y_t \mid \mathrm{x}_t).$$

Prediction and Update Equations

- Therefore, one can write the full posterior of x_t, as

$$f(x_t \mid y_1, y_2, \ldots, y_t) = \frac{f(x_t, y_1, y_2, \ldots, y_t)}{f(y_1, y_2, \ldots, y_t)}$$

$$= \frac{f(y_t \mid x_t) f(x_t \mid y_1, y_2, \ldots, y_{t-1})}{f(y_t \mid y_1, y_2, \ldots, y_{t-1})}$$

- We also have

$$f(x_t \mid y_1, y_2, \ldots, y_{t-1}) = \int_{x_{t-1}} f(x_t, x_{t-1} \mid y_1, y_2, \ldots, y_{t-1}) dx_{t-1}$$

$$= \int_{x_{t-1}} f(x_t \mid x_{t-1}) f(x_{t-1} \mid y_1, y_2, \ldots, y_{t-1}) dx_{t-1}$$

- The second equation is called **prediction equation**, and the first one is called the **update equation**.

10.2 Kalman Filter

Notation

- $\mathrm{x}_t \in \mathrm{R}^d$: internal state at *t*th frame (hidden random variable, e.g. position of the object in the image).

  $\mathrm{X}_t = [\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_t]^T$: history up to time step *t*

- $\mathrm{y}_t \in \mathrm{R}^c$: measurement at *t*th frame (observable random variable, e.g. the given image).

  $\mathrm{Y}_t = [\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_t]^T$: history up to time step *t*

- Goal:

Estimating the posterior probability $p(\mathrm{x}_t \mid \mathrm{Y}_t)$

Recursive Formula

$$
\begin{aligned}
& p(\mathrm{x}_t | \mathrm{Y}_t) \\
&= p(\mathrm{x}_t | \mathrm{Y}_{t-1}, \mathrm{y}_t) \\
&\propto p(\mathrm{y}_t | \mathrm{x}_t, \mathrm{Y}_{t-1}) p(\mathrm{x}_t | \mathrm{Y}_{t-1}) \\
&\propto p(\mathrm{y}_t | \mathrm{x}_t) p(\mathrm{x}_t | \mathrm{Y}_{t-1}) \\
&\propto p(\mathrm{y}_t | \mathrm{x}_t) \int p(\mathrm{x}_t | \mathrm{x}_{t-1}, \mathrm{Y}_{t-1}) p(\mathrm{x}_{t-1} | \mathrm{Y}_{t-1}) d\mathrm{x}_{t-1} \\
&\propto p(\mathrm{y}_t | \mathrm{x}_t) \int p(\mathrm{x}_t | \mathrm{x}_{t-1}) p(\mathrm{x}_{t-1} | \mathrm{Y}_{t-1}) d\mathrm{x}_{t-1}
\end{aligned}
$$

Bayes rule:
$p(a | b) = p(b | a) p(a) / p(b)$

Assumption:
$p(\mathrm{y}_t | \mathrm{X}_t, \mathrm{Y}_{t-1}) = p(\mathrm{y}_t | \mathrm{x}_t)$

Integration:
$p(a) = \int p(a | b) p(b) db$

Assumption:
$p(\mathrm{x}_t | \mathrm{x}_{t-1}, \mathrm{Y}_{t-1}) = p(\mathrm{x}_t | \mathrm{x}_{t-1})$

Bayesian Formulation

$$\boxed{p(\mathrm{x}_t \mid \mathrm{Y}_t) = \kappa\, p(\mathrm{y}_t \mid \mathrm{x}_t) \int p(\mathrm{x}_t \mid \mathrm{x}_{t-1}) p(\mathrm{x}_{t-1} \mid \mathrm{Y}_{t-1}) d\mathrm{x}_{t-1}}$$

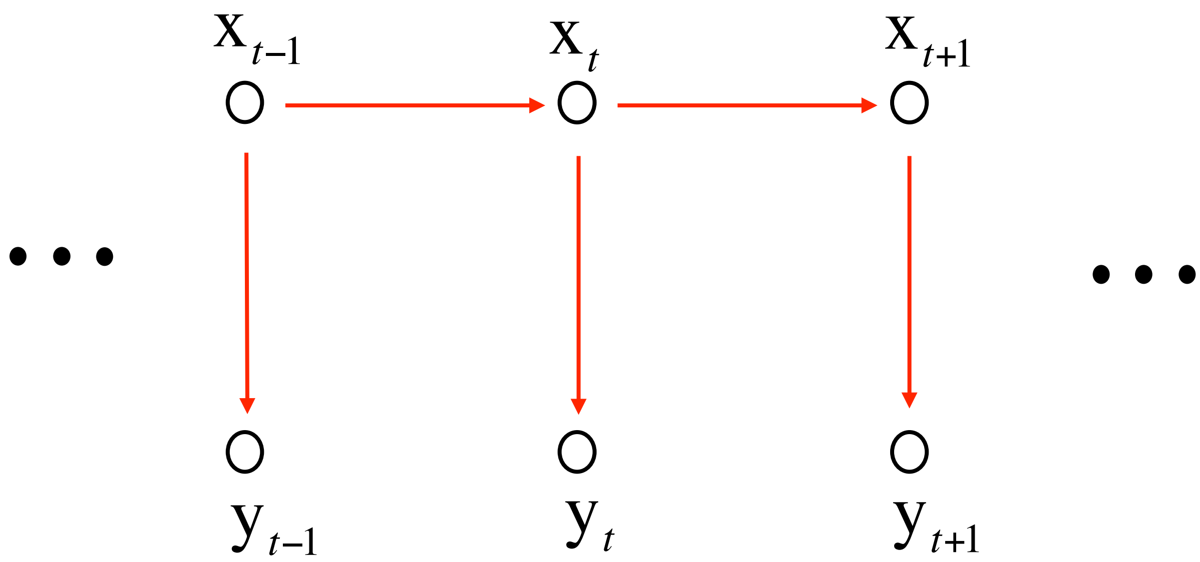
$p(\mathrm{y}_t \mid \mathrm{x}_t)$: *likelihood*

$p(\mathrm{x}_t \mid \mathrm{x}_{t-1})$: *temporal prior*

$p(\mathrm{x}_{t-1} \mid \mathrm{Y}_{t-1})$: posterior probability at previous time step

κ: normalizing term

Bayesian Graphical Model



Markov assumptions:

$$p(y_t | X_t, Y_{t-1}) = p(y_t | x_t)$$

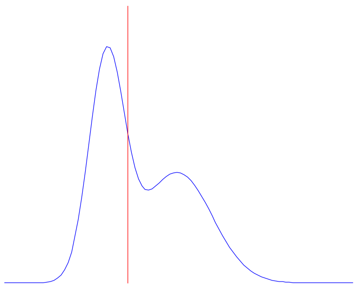
$$p(x_t \mid X_{t-1}, Y_{t-1}) = p(x_t \mid x_{t-1})$$

Estimators

Assume the posterior probability $p(\mathrm{x}_t \mid \mathrm{Y}_t)$ is known:

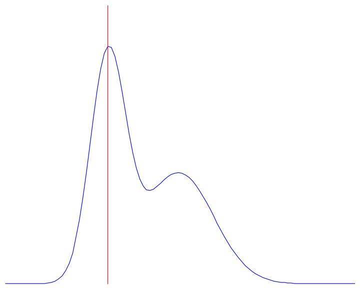
- posterior mean

$$\hat{\mathrm{x}}_t = E(\mathrm{x}_t \mid \mathrm{Y}_t)$$



- maximum *a posterior* (MAP)

$$\hat{\mathrm{x}}_t = \arg\max p(\mathrm{x}_t \mid \mathrm{Y}_t)$$



$p(\mathrm{x}_t \mid \mathrm{Y}_t)$

General Model

- $p(\mathrm{x}_t \mid \mathrm{Y}_t)$ can be an arbitrary, non-Gaussian, multi-modal distribution.

- The recursive equation has no explicit solution, but can be numerically approximated using Monte Carlo techniques.

- If both *likelihood* and *prior* are Gaussian, the solution has closed form and the two estimators (posterior mean & MAP) are the same.

- Such model is known as the Kalman filter. **(Kalman, 1960)**

Broad Applications of Kalman Filter

- Engineering
  - Robotics, spacecraft, aircraft, automobiles

"… *Kalman filtering rapidly became a mainstay of aerospace engineering. It was used, for example, in the Ranger, Mariner, and Apollo missions of the 1960s. In particular, the on-board computer that guided the descent of the Apollo 11 lunar module to the moon had a Kalman filter* …" SIAM news, '93.

- Computer
  - Tracking, real-time graphics, computer vision

- Others
  - Forecasting economic indicators
  - Telephone and electricity loads
  - Encoding/decoding neural signals