State Space Model (Dynamic model)

A/Prof Richard Yi Da Xu
Yida.Xu@uts.edu.au
Wechat: aubedata
https://github.com/roboticcam/machine-learning-notes

University of Technology Sydney (UTS)

February 18, 2018

Content

- **Continous Dynamic Model:** Kalman Filter and Extended KF
- **Discrete Dynamic Model:** Hidden Markov Model

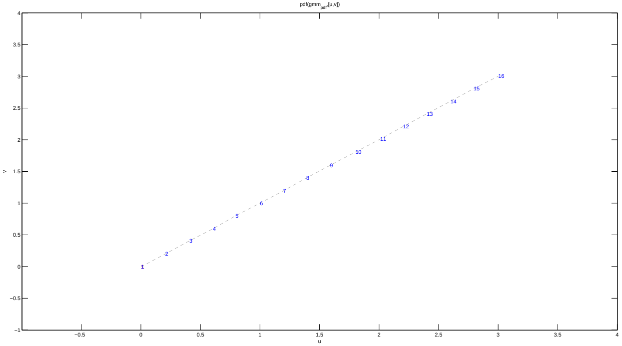
What is time series?

- A collection of observations of well-defined data items obtained through repeated measurements over time.
- Examples of time series?

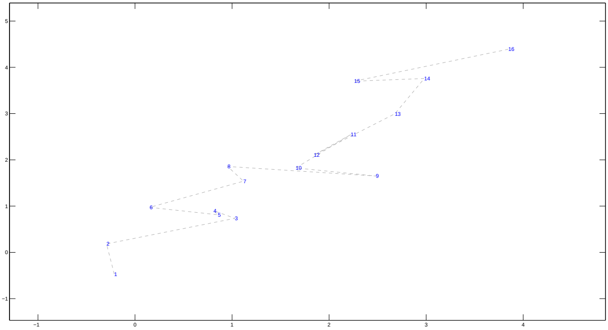
a primary school approach: We have a dynamic model: a robot that is travelling 0.2 meters every minute in both x and y directions:

- At previous time $t-1$, its position (state) is: x_{t-1}
- At current time t, its position (state) is: $x_t = x_{t-1} + \begin{bmatrix} 0.2 \\ 0.2 \end{bmatrix}$

Let's simulate a path:



However, nothing is perfect! The dynamic model always contains a random noise:



State Transitions

Very commonly, we have Gaussian noise in the **transition**:

$$x_t = F(x_{t-1}) + w_t \qquad w_t \sim \mathcal{N}(0, Q_t)$$

In the case of previous example,

$$x_t = x_{t-1} + w_t \qquad w_t \sim \mathcal{N}(0, Q_t) \qquad \text{where } F(x_{t-1}) = x_{t-1} + w_t$$

Making the matter slightly complicated:

We can not measure the states directly, we have to estimate the states via some external measurements:

$$y_t = H(x_t) + v_t \qquad v_t \sim \mathcal{N}(0, R_t)$$

Note that x_t is now a **latent** variable.

A dynamic System

In a general Dynamic System (DS), we have the following equation:

$$\begin{aligned} x_t &= F(x_{t-1}) \\ y_t &= H(x_t) \end{aligned}$$

For Kalman Filter, we are interested in Linear Dynamic System (LDS), and Gaussian noises. We have the following equations:

$$\begin{aligned} x_t &= Ax_{t-1} + B + w_t \qquad & w_t &\sim \mathcal{N}(0, Q_t) \\ y_t &= Hx_t + C + v_t \qquad & v_t &\sim \mathcal{N}(0, R_t) \end{aligned}$$

- A truck on perfectly frictionless, infinitely long straight rails.
- Initially the truck is stationary at position 0, but it is buffeted this way and that by **random acceleration**, i.e., we assume $a_t \sim \mathcal{N}(0, \sigma^2)$. Of course, this does NOT imply $w_t \sim \mathcal{N}(0, \sigma^2)$
- We measure position of the truck every $\triangle$t seconds, but these measurements are imprecise.
- We want to maintain a model of where the truck is and what its velocity.

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + w_t \qquad w_t \sim \mathcal{N}(0, Q_t)$$

$$\begin{bmatrix} x_t \\ \dot{x}_t \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & \triangle t \\ 0 & 1 \end{bmatrix}}_{A_t} \underbrace{\begin{bmatrix} x_{t-1} \\ \dot{x}_{t-1} \end{bmatrix}}_{x_{t-1}} + \underbrace{\begin{bmatrix} \frac{1}{2} a_t (\triangle t)^2 \\ a_t \triangle t \end{bmatrix}}_{w_t}$$

This is using simple high school physics:

$$x_t = x_{t-1} + \dot{x}_{t-1} \triangle t + \frac{1}{2} a_t (\triangle t)^2$$
$$\dot{x}_t = \dot{x}_{t-1} + a_t \triangle t$$

How to compute Q_t

$$\mathbf{x}_t = \begin{bmatrix} x_t \\ \dot{x}_t \end{bmatrix} = \begin{bmatrix} 1 & \triangle t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ \dot{x}_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} \frac{1}{2} a_t (\triangle t)^2 \\ a_t \triangle t \end{bmatrix}}_{w_t}$$

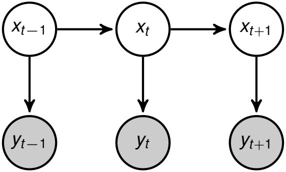
- Assume $a_t \sim \mathcal{N}(0, \sigma^2) \quad \forall t$ and $w_t \sim \mathcal{N}(0, Q_t)$
- What's Q_t?

$$\begin{aligned} Q_t = \mathbb{COV}(\mathbf{x}_t) &= \mathbb{COV}\left(\begin{bmatrix} 1 & \triangle t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ \dot{x}_{t-1} \end{bmatrix} + \begin{bmatrix} \frac{1}{2} a_t (\triangle t)^2 \\ a_t \triangle t \end{bmatrix} \right) \\ &= \mathbb{COV}\left(\begin{bmatrix} \frac{1}{2} a_t (\triangle t)^2 \\ a_t \triangle t \end{bmatrix} \right) \\ &= \mathbb{E}\left[(a_t)^2 \begin{bmatrix} \frac{1}{2} (\triangle t)^2 \\ \triangle t \end{bmatrix} \begin{bmatrix} \frac{1}{2} (\triangle t)^2 & \triangle t \end{bmatrix} \right] \\ &= \sigma^2 \begin{bmatrix} \frac{1}{4} (\triangle t)^4 & \frac{1}{2} (\triangle t)^3 \\ \frac{1}{2} (\triangle t)^3 & (\triangle t)^2 \end{bmatrix} \end{aligned}$$

Measurement Equation

- Suppose At each time step, a noisy measurement of the true position of the truck is made.
- Let us suppose the noise, v_t is also normally distributed, with mean 0 and standard deviation σ_z

$$\begin{aligned} y_t &= H\mathbf{x}_t + C + v_t \qquad && v_t \sim \mathcal{N}(0, R_t) \\ &= \begin{bmatrix} 1 & 0 \end{bmatrix} + v_t \qquad && v_t \sim \mathcal{N}(0, \sigma_z) \end{aligned}$$



Markov Property:

$$p(x_t|x_1,\dots,x_{t-1},y_1,\dots,y_{t-1}) = p(x_t|x_{t-1})$$
$$p(y_t|x_1,\dots,x_{t-1},x_t,y_1,\dots,y_{t-1}) = p(y_t|x_t)$$

Linear Gaussian Dynamic Model

$$
\begin{aligned}
&\mathbf{x}_t = A\mathbf{x}_{t-1} + B + w_t \qquad w_t \sim \mathcal{N}(0, Q_t) \\
\Longrightarrow\ &\textbf{Transition probability:} \qquad p(\mathbf{x}_t|\mathbf{x}_{t-1}) \sim \mathcal{N}(A\mathbf{x}_{t-1} + B, Q_t)
\end{aligned}
$$

$$
\begin{aligned}
&y_t = H\mathbf{x}_t + v_t \qquad v_t \sim \mathcal{N}(0, R_t) \\
\Longrightarrow\ &\textbf{Measurement probability:} \qquad p(y_t|\mathbf{x}_t) \sim \mathcal{N}(H\mathbf{x}_t, R_t)
\end{aligned}
$$

- Kalman Filter can be used to in this Gaussian, Linear case.
- In general, there are many other Dyanmic models which are non-Gaussian, non-Linear. They can NOT be solved using Kalman Filter.

What do we want to compute?

$$\begin{aligned}
\textbf{Prediction}: &\quad p(x_t|\mathbf{y}_{1:t-1}) = \int_{x_{t-1}} p(x_t|x_{t-1})p(x_{t-1}|\mathbf{y}_{1:t-1}) \\
\textbf{Update}: &\quad p(x_t|\mathbf{y}_{1:t}) = \frac{p(y_t|x_t)p(x_t|\mathbf{y}_{1:t-1})}{\int_{s_t} p(y_t|s_t)p(ds_t|\mathbf{y}_{1:t-1})}
\end{aligned} \tag{1}$$

This is because:

$$\begin{aligned}
p(x_t|\mathbf{y}_{1:t}) &\propto p(x_t, \mathbf{y}_{1:t}) \\
&\propto p(y_t|x_t)p(x_t|\mathbf{y}_{1:t-1}) \\
&= \frac{p(y_t|x_t)p(x_t|\mathbf{y}_{1:t-1})}{\int_{s_t} (y_t|s_t)p(ds_t|\mathbf{y}_{1:t-1})}
\end{aligned} \tag{2}$$

Kalman Filter - Prediction

Following marginal distribution of linear Gaussian (Bishop p.93), given:

- $p(x) \sim \mathcal{N}(x|\mu, \Sigma)$
- $p(y|x) \sim \mathcal{N}(y|Ax + b, L)$

$$\textbf{Marginal} : p(y) = \int_x p(y|x)p(x) \sim \mathcal{N}\left(y|A\mu + b, L + A\Sigma A^T\right)$$

$$\begin{aligned}\textbf{Prediction} : \quad p(x_t|\mathbf{y}_{1:t-1}) \sim \mathcal{N}(\bar{\mu}_t, \bar{\Sigma}_t) &= \int_{x_{t-1}} p(x_t|x_{t-1})p(dx_{t-1}|\mathbf{y}_{1:t-1}) \\ &= \int_{x_{t-1}} \mathcal{N}(x_t|Ax_{t-1} + B, Q_t)\mathcal{N}(x_{t-1}|\hat{\mu}_{t-1}, \hat{\Sigma}_{t-1}) \\ &= \mathcal{N}\left(x_t|A\hat{\mu}_{t-1} + B, A\hat{\Sigma}_{t-1}A^T + Q_t\right)\end{aligned}$$

Question How abour **update**? Let's see an alternative method using Moment representation.

Moment Representation (1)

In order to compute **moments**, we introduce a zero-mean variable: $\triangle x_{t-1}$, i.e.,:

- $\triangle x_{t-1} \equiv x_{t-1} - \mathbb{E}[x_{t-1}] \sim \mathcal{N}(0, \hat{\Sigma}_{t-1}) \implies x_{t-1} = \triangle x_{t-1} + \mathbb{E}[x_{t-1}]$

We attempt to write both $\triangle x_t$ and $\triangle y_t$ in terms of $\triangle x_{t-1}$:

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + w_t \qquad w_t \sim \mathcal{N}(0, Q_t) \implies \qquad \begin{aligned} x_t &= A(\triangle x_{t-1} + \mathbb{E}[x_{t-1}]) + w_t \\ &= A\mathbb{E}x_{t-1} + \underbrace{A\triangle x_{t-1} + w_t}_{\triangle x_t} \end{aligned}$$

$$y_t = H\mathbf{x}_t + v_t \qquad v_t \sim \mathcal{N}(0, R_t) \implies \qquad \begin{aligned} y_t &= Hx_t + v_t \\ &= H(A\mathbb{E}x_{t-1} + A\triangle x_{t-1} + w_t) + v_t \\ &= HA\mathbb{E}x_{t-1} + \underbrace{HA\triangle x_{t-1} + Hw_t + v_t}_{\triangle y_t} \end{aligned} \tag{3}$$

The Independence assumptions:

- $\mathbb{COV}(x_{t-1}, w_t) = 0 \qquad \mathbb{COV}(x_{t-1}, v_t) = 0 \qquad \mathbb{COV}(w_t, v_t) = 0$

Moment Representation (2)

$$
\begin{aligned}
\mathbb{E}[\triangle x_t(\triangle x_t)^T|y_{1:t-1}] &= \mathbb{E}[(A\triangle x_{t-1} + w_t)(A\triangle x_{t-1} + w_t)^T] \\
&= A\hat{\Sigma}_{t-1}A^T + Q_t = \bar{\Sigma}_t
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[\triangle y_t(\triangle x_t)^T|y_{1:t-1}] &= \mathbb{E}\left[(HA\triangle x_{t-1} + Hw_t + v_t)(A\triangle x_{t-1} + w_t)^T\right] \\
&= H\left(A\hat{\Sigma}_{t-1}A^T + Q_t\right) = H\bar{\Sigma}_t \\
\implies \mathbb{E}[\triangle x_t(\triangle y_t)^T|y_{1:t-1}] &= \bar{\Sigma}_t H^T
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[\triangle y_t(\triangle y_t)^T|y_{1:t-1}] &= \mathbb{E}\left[(HA\triangle x_{t-1} + Hw_t + v_t)(HA\triangle x_{t-1} + Hw_t + v_t)^T\right] \\
&= H\left(A\hat{\Sigma}_{t-1}A^T + Q_t\right)H^T + R_t = H(\bar{\Sigma}_t)H^T + R_t
\end{aligned}
$$

$$
\mathbb{E}[y_t|y_{1:t-1}] = HA\mathbb{E}[x_{t-1}] = HA\hat{\mu}
$$

$$
\mathbb{E}[x_t|y_{1:t-1}] = A\mathbb{E}[x_{t-1}] = A\hat{\mu}
$$

Kalman Filter Prediction (alternative): $p(x_t|y_1,\dots y_{t-1}) = \mathcal{N}(\bar{\mu}_t, \bar{\Sigma}_t)$

mean: $\bar{\mu}_t = \mathbb{E}[x_t|y_{1:t-1}]$:

$$\begin{aligned}
&\mathbb{E}[Ax_{t-1} + w_t|y_{1:t-1}] \\
&= A\mathbb{E}[x_{t-1}|y_{1:t-1}] + \mathbb{E}[w_t] \\
&= A\hat{\mu}_{t-1}
\end{aligned}$$

covariance:

$$\begin{aligned}
\bar{\Sigma}_t &= \mathbb{E}[(\triangle x_t)(\triangle x_t)^T] \\
&= \mathbb{E}[(A\triangle x_t + \triangle w_t)(A\triangle x_t + \triangle w_t)^T] \\
&= A\mathbb{E}[\triangle x \triangle x_t]A^T + A\mathbb{E}[\triangle x_t(\triangle w_t)^T] + \mathbb{E}[(\triangle x_t)^T \triangle w_t]A^T + \mathbb{E}[\triangle w_t(\triangle w_t)^T] \\
&\text{Since the noises } x \text{ and } w \text{ are assumed independent } \mathbb{E}[\triangle x_t(\triangle w_t)^T] = 0 : \\
&= A\hat{\Sigma}_{t-1}A^T + Q_t
\end{aligned}$$

$$\textbf{Update} \qquad \begin{aligned} & p(x_t|y_1,\dots y_t) \sim \mathcal{N}(\hat{\mu}_t, \hat{\Sigma}_t) \\ & \propto p(y_t|x_t)p(x_t|\mathbf{y}_{1:t-1}) \\ & = \mathcal{N}(y_t|Hx_t, R_t)\mathcal{N}(x_t|\bar{\mu}_t, \bar{\Sigma}_t) \end{aligned}$$

- Say $p(u) = \mathcal{N}(\mu_u, \Sigma_{uu}) \qquad p(v) = \mathcal{N}(\mu_v, \Sigma_{vv})$
- $p(u|v) \sim \mathcal{N}\left(\mu_u + \Sigma_{uv}\Sigma_{vv}^{-1}(v-\mu_v), \Sigma_{uu} - \Sigma_{uv}\Sigma_{vv}^{-1}\Sigma_{vu}\right)$
- Think $p(u) \equiv p(x_t|y_1,\dots,y_{t-1}) \sim \mathcal{N}(x_t|\bar{\mu}_t, \bar{\Sigma}_t) \qquad p(v) \equiv p(y_t|y_1,\dots y_{t-1})$
- We are after $p(u|v) \equiv p(x_t|y_t, y_1,\dots,y_{t-1})$

Kalman Filter Update: $p(x_t|y_1,\dots y_t) = \mathcal{N}(\hat{\mu}_t, \hat{\Sigma}_t)$ (2)

mean: $\hat{\mu}_t = \mathbb{E}[x_t|y_{1:t}]$:

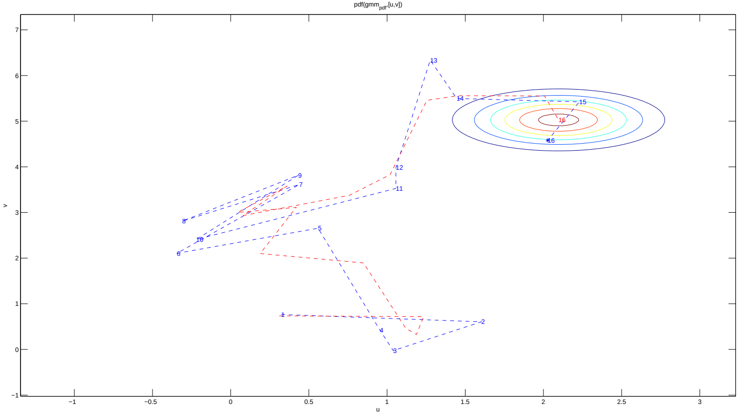
$$
\begin{aligned}
&\mu_u + \Sigma_{uv}\Sigma_{vv}^{-1}(v - \mu_v) \\
&\equiv \mathbb{E}[x_t] + \mathbb{E}[\triangle x_t(\triangle y_t)^T]\mathbb{E}[\triangle y_t(\triangle y_t)^T]^{-1}(y_t - \mathbb{E}[y_t]) \\
&= A\hat{\mu}_{t-1} + \bar{\Sigma}_t^T H(H\bar{\Sigma}_t H^T + R_t)^{-1}(y_t - HA\hat{\mu}_{t-1})
\end{aligned}
$$

covariance: $\hat{\Sigma}_t = \mathbb{COV}[x_t|y_{1:t}]$:

$$
\begin{aligned}
&\Sigma_{uu} - \Sigma_{uv}\Sigma_{vv}^{-1}\Sigma_{vu} \\
&\equiv \mathbb{E}[\triangle x_t(\triangle x_t)^T] - \mathbb{E}[\triangle x_t(\triangle y_t)^T]\mathbb{E}[\triangle y_t(\triangle y_t)^T]^{-1}\mathbb{E}[\triangle y_t(\triangle x_t)^T] \\
&= \bar{\Sigma}_t - \underbrace{\bar{\Sigma}_t H^T(H(\bar{\Sigma}_t)H^T + R_t)^{-1}}_{K} H\bar{\Sigma}_t \\
&= (I - KH)\bar{\Sigma}_t
\end{aligned}
$$

Kalman Filter Demo:

Notice of the **smoothing** effect:



mean: $\hat{\mu}_t = \mathbb{E}[x_t|y_{1:t}]$:

$$\text{k-d:} \quad \hat{\mu}_t = A\hat{\mu}_{t-1} + \bar{\Sigma}_t^T H(H\bar{\Sigma}_t H^T + R_t)^{-1}(y_t - HA\hat{\mu}_{t-1})$$

$$\text{1-d:} \quad \hat{\mu}_t = a\hat{\mu}_{t-1} + \frac{\bar{\sigma}_t^2 h(y_t - ha\hat{\mu}_{t-1})}{h^2\bar{\sigma}_t^2 + R_t} = \frac{a\hat{\mu}_{t-1}(h^2\bar{\sigma}_t^2 + R_t) + \bar{\sigma}_t^2 h(y_t - ha\hat{\mu}_{t-1})}{h^2\bar{\sigma}_t^2 + R_t}$$

$$= \frac{a\hat{\mu}_{t-1}R_t + \bar{\sigma}_t^2 h y_t}{h^2\bar{\sigma}_t^2 + R_t}$$

covariance: $\hat{\Sigma}_t = \mathbb{COV}[x_t|y_{1:t}]$:

$$\text{k-d:} \quad \hat{\Sigma}_t = \bar{\Sigma}_t - \bar{\Sigma}_t H^T(H(\bar{\Sigma}_t)H^T + R_t)^{-1}H\bar{\Sigma}_t$$

$$\text{1-d:} \quad \hat{\sigma}_t = \frac{\bar{\sigma}_t^2(h^2\bar{\sigma}_t^2 + R_t) - (\bar{\sigma}_t^2)^2 h^2}{h^2\bar{\sigma}_t^2 + R_t} = \frac{\bar{\sigma}_t^2(h^2\bar{\sigma}_t^2 + R_t) - (\bar{\sigma}_t^2)^2 h^2}{h^2\bar{\sigma}_t^2 + R_t} = \frac{\bar{\sigma}_t^2 R_t}{h^2\bar{\sigma}_t^2 + R_t}$$

Extended Kalman Filter: Non-Linear Dynamic System and Gaussian model

Kalman Filter: Linear Guassian Model:

$$
\begin{aligned}
\mathbf{x}_t &= A\mathbf{x}_{t-1} + B + w_t \qquad w_t \sim \mathcal{N}(0, Q_t) \\
y_t &= H\mathbf{x}_t + v_t \qquad v_t \sim \mathcal{N}(0, R_t)
\end{aligned}
$$

Extended Kalman Filter: Non-Linear Guassian Model:

$$
\begin{aligned}
\mathbf{x}_t &= F(x_{t-1}) + w_t \qquad w_t \sim \mathcal{N}(0, Q_t) \\
y_t &= H(\mathbf{x}_t) + v_t \qquad v_t \sim \mathcal{N}(0, R_t)
\end{aligned}
$$

$$\mathbf{x}_t = F(\mathbf{x}_{t-1}) + w_t \qquad w_t \sim \mathcal{N}(0, Q_t)$$
$$y_t = H(\mathbf{x}_t) + v_t \qquad v_t \sim \mathcal{N}(0, R_t)$$

Taylor Expansion:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \text{ High order terms}$$

Expand $F(\mathbf{x}_{t-1})$ around a particular point x_{t-1}^p:

$$x_t = F(x_{t-1}^p) + F'(x_{t-1}^p)\left(x_{t-1} - x_{t-1}^p\right) + \text{ High order terms} + w_t$$

Let $J_p \equiv F'(x_{t-1}^p)$:

$$\begin{aligned} x_t &= F(x_{t-1}^p) + J_p\left(x_{t-1} - x_{t-1}^p\right) + \text{ High order terms} + w_t \\ &\approx \underbrace{J_p}_{A} x_{t-1} + \underbrace{\left(F(x_{t-1}^p) - J_p x_{t-1}^p\right)}_{B} + w_t \end{aligned}$$

Extended Kalman Filter: Prediction $p(x_t|y_1,\dots y_{t-1}) = \mathcal{N}(\bar{\mu}_t, \bar{\Sigma}_t)$

Kalman Filter: $x_t = Ax_{t-1} + w_t \qquad w_t \sim \mathcal{N}(B, Q_t)$

- **mean:** $\bar{\mu}_t = \mathbb{E}[x_t|y_{1:t-1}] = A\hat{\mu}_{t-1} + B$
- **covariance**: $\bar{\Sigma}_t = \mathbb{E}[(\triangle x_t)(\triangle x_t)^T] = A\hat{\Sigma}_{t-1}A^T + Q_t$

Extended Kalman Filter: $x_t \approx \underbrace{J_p}_{A} x_{t-1} + w \qquad w_t \sim \mathcal{N}\left(\underbrace{\left(F(x^p_{t-1}) - J_p x^p_{t-1}\right)}_{B}, Q_t\right)$

- **mean:** $\bar{\mu}_t = \mathbb{E}[x_t|y_{1:t-1}] = J_p\hat{\mu}_{t-1} + \left(F(x^p_{t-1}) - J_p x^p_{t-1}\right)$
- **covariance**: $\bar{\mu}_t = \mathbb{E}[(\triangle x_t)(\triangle x_t)^T] = J_p\hat{\Sigma}_{t-1}J_p^T + Q_t$

- **mean:** $\bar{\mu}_t = \mathbb{E}[x_t|y_{1:t-1}] = J_p\hat{\mu}_{t-1} + \left(F(x^p_{t-1}) - J_p x^p_{t-1}\right)$
- **covariance**: $\bar{\Sigma}_t = \mathbb{E}[(\triangle x_t)(\triangle x_t)^T] = J_p\hat{\Sigma}_{t-1}J_p^T + Q_t$

Too complicated, lets simplify it using a trick: x_p is just an arbitary point. So we can choose it to be anything we like. Why not let $x_p = \hat{\mu}_{t-1}$:

- **mean:**
  $\bar{\mu}_t = \mathbb{E}[x_t|y_{1:t-1}] = F'(\hat{\mu}_{t-1})\hat{\mu}_{t-1} + (F(\hat{\mu}_{t-1}) - F'(\hat{\mu}_{t-1})\hat{\mu}_{t-1}) = F(\hat{\mu}_{t-1})$
- **covariance**: $\bar{\Sigma}_t = \mathbb{E}[(\triangle x_t)(\triangle x_t)^T] = F'(\hat{\mu}_{t-1})\hat{\Sigma}_{t-1}F'(\hat{\mu}_{t-1})^T + Q_t$

Extended Kalman Filter: Update $p(x_t|y_1, \dots y_t) = \mathcal{N}(\hat{\mu}_t, \hat{\Sigma}_t)$

Taylor Expansion:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \text{ High order terms}$$

Measurement Equation: $y_t = H(x_t) + v_t \qquad v_t \sim \mathcal{N}(0, R_t)$

$$y_t = H(x_t^p) + H'(x_t^p)\left(x_t - x_t^p\right) + \text{ High order terms} + v_t \qquad v_t \sim \mathcal{N}(0, R_t)$$

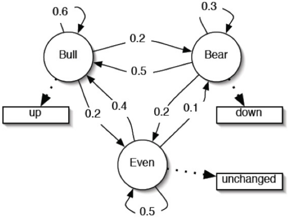
Let $J_p \equiv H'(x_t^p)$:

$$\begin{aligned} y_t &= H(x_t^p) + J_p\left(x_t - x_t^p\right) + \text{ High order terms} + v_t \\ &\approx H(\bar{x}_t) + J_p(x_t - \bar{x}_t) + v_t \qquad \text{let } x_t^p = \bar{x}_t \\ \implies \underbrace{y_t - H(\bar{x}_t) + J_p\bar{x}_t}_{\mathbb{Y}_t} &\approx \underbrace{J_p}_{H} x_t + v_t \end{aligned}$$

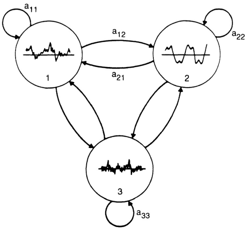
The rest are just following the standard Kalman Filter

Discrete States Dynamic Model: Hidden Markov Model

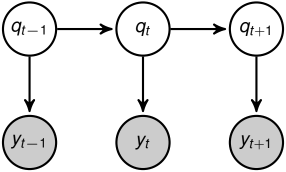
Simple Stock Market:



Speech Recognition:



Hidden Markov Model



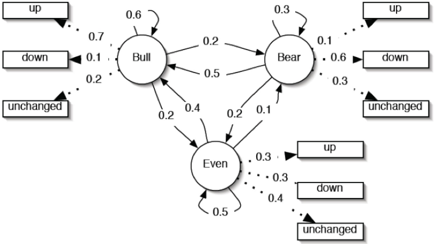
Discrete Transition Probability:

$$p(q_t|q_1, \dots, q_{t-1}, y_1, \dots, y_{t-1}) = p(q_t|q_{t-1})$$

Continous/Discrete Measurement probability:

$$p(y_t|q_1, \dots, q_{t-1}, q_t, y_1, \dots, y_{t-1}) = p(y_t|q_t)$$

HMM's Transition Probability **must be discrete**



$$A = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.5 & 0.3 & 0.2 \\ 0.4 & 0.1 & 0.5 \end{bmatrix}$$

Transition Probability:

- Let Bull = 1, Bear = 2, Even = 3:

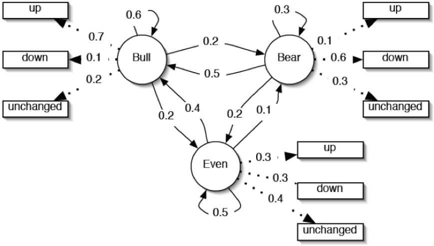
- $p(q_t = 1 | q_{t-1} = 1) = 0.6$
- $p(q_t = 2 | q_{t-1} = 1) = 0.2$
- $p(q_t = 3 | q_{t-1} = 1) = 0.2$

- $p(q_t = 1 | q_{t-1} = 2) = 0.5$
- $p(q_t = 2 | q_{t-1} = 2) = 0.3$
- $p(q_t = 3 | q_{t-1} = 2) = 0.2$

- $p(q_t = 1 | q_{t-1} = 3) = 0.4$
- $p(q_t = 2 | q_{t-1} = 3) = 0.1$
- $p(q_t = 3 | q_{t-1} = 3) = 0.5$

HMM's Measurement Probability

HMM's Measurement Probability can be both discrete or continous



$$B = \begin{bmatrix} 0.7 & 0.1 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}$$

- Let Bull = 1, Bear = 2, Even = 3:
- Let Up = 1, Down = 2, Uneven = 3:

- $p(y_t = 1 | q_t = 1) = 0.7$
- $p(y_t = 2 | q_t = 1) = 0.1$
- $p(y_t = 3 | q_t = 1) = 0.2$

- $p(y_t = 1 | q_t = 2) = 0.1$
- $p(y_t = 2 | q_t = 2) = 0.6$
- $p(y_t = 3 | q_t = 2) = 0.3$

- $p(y_t = 1 | q_t = 3) = 0.3$
- $p(y_t = 2 | q_t = 3) = 0.3$
- $p(y_t = 3 | q_t = 3) = 0.4$

The HMM Parameter λ (discrete measurement case) contains:

$$\lambda = \{A, B, \pi\}$$

π is the probability of the initial state , i.e., $p(q_1)$. We use $\pi_i \equiv p(q_1 = i)$.
Let $Q = q_1, \dots q_T$ and $Y = y_1, \dots y_T$:

Three major operations of HMM:

$$\begin{aligned}
&\text{Evaluate } p(Y|\lambda) \\
&\lambda_{\text{MLE}} = \arg\max_{\lambda} p(Y|\lambda) \\
&\arg\max_{Q} p(Y|Q, \lambda)
\end{aligned}$$

We will discuss Evaluation first.

Evaluate $p(Y|\lambda)(1)$

The usual way to compute this:

$$
\begin{aligned}
p(Y|\lambda) &= \sum_Q [p(Y, Q|\lambda)] = \sum_{q_1=1}^{k} \dots, \sum_{q_T=1}^{k} [p(y_1, \dots, y_T, q_1, \dots q_T|\lambda)] \\
&= \sum_{q_1=1}^{k} \dots, \sum_{q_T=1}^{k} [p(y_1, \dots, y_T, q_0, q_1, \dots q_T|\lambda)] \\
&= \sum_{q_1=1}^{k} \dots, \sum_{q_T=1}^{k} p(q_1)p(y_1|q_1)p(q_2|q_1)\dots p(q_t|q_{t-1})p(y_t|q_t) \\
&= \sum_{q_1=1}^{k} \dots, \sum_{q_T=1}^{k} \pi(q_1) \prod_{t=2}^{T} a_{q_{t-1},q_t} b_{q_t}(y_t)
\end{aligned}
$$

- We let transition probability: $p(q_t = j|q_{t-1} = i) \equiv a_{i,j}$ and
- We let measurement probability $p(y_t|q_t = j) \equiv b_j(y_t)$

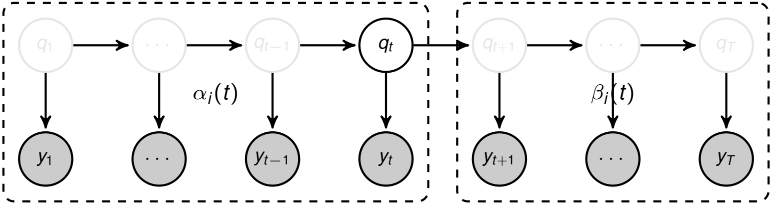
- There are k^T possible values of Q!. We need simpler methods

Forward Algorithm:

$$\alpha_i(t) = p(y_1, y_2, \dots y_t, q_t = i|\lambda)$$

Backward Algorithm:

$$\beta_i(t) = p(y_{t+1}, \dots, y_T|q_t = i, \lambda)$$



Evaluate $p(Y|\lambda)$ (2) Forward and Backward Formula

Therefore, we define **forward** procedure:

$$\alpha_i(t) = p(y_1, y_2, \dots y_t, q_t = i|\lambda) \implies p(Y|\lambda) = \sum_{i=1}^{k} \alpha_i(T)$$

This is the propbality of partial sequnce $y_1, \dots, y_t$ and ending up in state i at time t. Looking at the following recursion:

$$\alpha_i(1) = p(y_1, q_1 = i|\lambda) = p(q_1)p(y_1|q_1) = \pi_i b_i(y_1)$$

$$\alpha_j(2) = p(y_1, y_2, q_2 = j|\lambda) = \sum_{i=1}^{k} \underbrace{p(q_1 = i)p(y_1|q_1 = i)}_{\alpha_i(1)} \underbrace{p(q_2 = i|q_1 = i)}_{a_{i,j}} \underbrace{p(y_2|q_2 = j)}_{b_j(y_2)} = \left[\sum_{i=1}^{k} \alpha_i(1) a_{i,j}\right] b_j(y_2)$$

$$\dots$$

$$\alpha_j(t+1) = \left[\sum_{i=1}^{k} \alpha_i(t) a_{i,j}\right] b_j(y_{t+1})$$

$$\dots$$

$$\alpha_j(T) = \left[\sum_{i=1}^{k} \alpha_i(T-1) a_{i,j}\right] b_j(y_T)$$

We have now, $k \times T$ summations!

Evaluate $p(Y|\lambda)$ (3) Forward and Backward Formula

We also define a **backward** procedure:

$$\beta_i(t) = p(y_{t+1}, \dots, y_T | q_t = i, \lambda) \implies \sum_{i=1}^{k} \beta_i(1)\pi_i b_i(y_1) = p(Y|\lambda)$$

Propbality of partial sequnce $y_{1+1}, y_{t+2}, \dots y_T$ **given** started at state i at time t:

$$\beta_i(T) = 1$$

$$\beta_i(T-1) = p(y_T|q_{T-1} = i) = \sum_{j=1}^{k} p(q_T = j|q_{T-1} = i)p(y_T|q_T = j) = \sum_{j=1}^{k} a_{i,j} b_j(T)$$

$$\beta_i(T-2) = p(y_T, y_{T-1}|q_{T-2} = i)$$

$$= \sum_{j=1}^{k} \underbrace{\sum_{l=1}^{k} p(q_T = l|q_{T-1} = j)p(y_T|q_T = l)}_{\beta_j(T-1)} \underbrace{p(q_{T-1} = j|q_{T-2} = i)}_{a_{i,j}} \underbrace{p(y_{T-1}|q_{T-1} = j)}_{b_j(T-1)} = \sum_{j=1}^{k} a_{i,j} b_j(y_{T-1})\beta_j(T-1)$$

$\dots$

$$\beta_i(t) = \sum_{j=1}^{k} a_{i,j} b_j(y_{t+1})\beta_j(t+1)$$

$\dots$

$$\beta_i(1) = \sum_{j=1}^{k} a_{i,j} b_j(y_2)\beta_j(2)$$

The probability of being at a particular state

The probability of being in state i at time t for a sequence Y:

$$p(q_t = i|Y, \lambda) = \frac{p(Y, q_t = i|\lambda)}{p(Y|\lambda)} = \frac{p(Y, q_t = i|\lambda)}{\sum_{j=1}^{k} p(Y, q_t = j|\lambda)} = \frac{\alpha_i(t)\beta_i(t)}{\sum_{j=1}^{k} \alpha_j(t)\beta_j(t)}$$

$$\begin{aligned}
p(Y, q_t = i|\lambda) &= p(Y|q_t = i)p(q_t = i) \\
&= p(y_1, \dots y_t|q_t = i)p(y_{t+1}, \dots y_T|q_t = i)p(q_t = i) \quad \text{by its graphical model} \\
&= p(y_1, \dots y_t, q_t = i)p(y_{t+1}, \dots y_T|q_t = i) \quad \text{re-arrange} \\
&= \alpha_i(t)\beta_i(t)
\end{aligned}$$

Parameter Learning

Looking at the E-M algorithm:

$$\Theta^{(g+1)} = \arg\max_{\Theta} \left[Q(\Theta, \Theta^{(g)}) \right] = \arg\max_{\Theta} \left(\int_z \log\left(p(X, Z|\Theta)\right) p(Z|X, \Theta^{(g)}) \mathrm{d}z \right)$$

In HMM, we write it as:

$$\lambda^{(g+1)} = \arg\max_{\lambda} \left(\underbrace{\int_{q \in Q} \ln\left(p(Y, q|\lambda)\right) p(q, Y|\lambda^{(g)})}_{\mathcal{Q}(\lambda, \lambda^{(g)})} \right)$$

$$\begin{aligned}
\mathcal{Q}(\lambda, \lambda^{(g)}) &= \int_{q \in Q} \ln\left(p(Y, q|\lambda)\right) p(q, Y|\lambda^{(g)}) \\
&= \sum_{q_0=1}^{k} \cdots \sum_{q_T=1}^{k} \left(\ln \pi_0 + \sum_{t=1}^{T} \ln a_{q_{t-1}, q_t} + \sum_{t=1}^{T} \ln b_{q_t}(y_t) \right) p(q, Y|\lambda^{(g)})
\end{aligned}$$

Parameter Learning: First term

$$\mathcal{Q}^{\text{term 1}} = \sum_{q_0=1}^{k} \cdots \sum_{q_T=1}^{k} \ln \pi_{q_0} p(q, Y|\lambda^{(g)}) = \sum_{i=1}^{k} \ln \pi_i p(q_0 = i, Y|\lambda^{(g)})$$

arg max$(\mathcal{Q}^{\text{term 1}})$ with $\sum_{i=1}^{k} \pi_i = 1$, using Lagrange Multiplier:

$$\mathbb{LM}^{\text{term 1}} = \sum_{i=1}^{k} \ln \pi_i p(q_0 = i, Y|\lambda^{(g)}) + \tau \left(\sum_{i=1}^{k} \pi_i - 1 \right)$$

$$\frac{\partial \mathbb{LM}^{\text{term 1}}}{\partial \pi_i} = \frac{p(q, Y|\lambda^{(g)})}{\pi_i} + \tau = 0 \qquad \frac{\partial \mathbb{LM}^{\text{term 1}}}{\partial \tau} = \sum_{i=1}^{k} \pi_i - 1 = 0$$

$$p(q_0 = i, Y|\lambda^{(g)}) = -\tau \pi_i$$

sum both sides: $\sum_{i=1}^{k} p(q_0 = i, Y|\lambda^{(g)}) = -\tau \sum_{i=1}^{k} \pi_i = -\tau$

substitute: $\pi_i = \frac{p(q_0 = i, Y|\lambda^{(g)})}{-\tau} \implies \pi_i = \frac{p(q_0 = i, Y|\lambda^{(g)})}{\sum_{i=1}^{k} p(q_0 = i, Y|\lambda^{(g)})}$

Parameter Learning: Second term

$$\mathcal{Q}^{\text{term 2}} = \sum_{q_0=1}^{k} \cdots \sum_{q_T=1}^{k} \sum_{t=1}^{T} \ln a_{q_{t-1},q_t} p(q, Y|\lambda^{(g)}) = \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{t=1}^{T} \ln a_{i,j} p(q_{t-1} = i, q_t = j, Y|\lambda^{(g)})$$

$$\mathbb{LM}^{\text{term 2}} = \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{t=1}^{T} \ln a_{i,j} p(q_{t-1} = i, q_t = j, Y|\lambda^{(g)}) + \sum_{i=1}^{k} \tau_i \left(\sum_{j=1}^{k} a_{i,j} - 1 \right)$$

$$\frac{\partial \mathbb{LM}^{\text{term 2}}}{\partial a_{i,j}} = \frac{\sum_{t=1}^{T} p(q_{t-1} = i, q_t = j, Y|\lambda^{(g)})}{a_{i,j}} + \sum_{i=1}^{k} \tau_i = 0$$

$$\frac{\partial \mathbb{LM}^{\text{term 2}}}{\partial \tau_i} = \sum_{i=1}^{k} a_{i,j} - 1 = 0$$

$$\sum_{t=1}^{T} p(q_{t-1} = i, q_t = j, Y|\lambda^{(g)}) = -a_{i,j} \sum_{i=1}^{k} \tau_i \implies a_{i,j} = \frac{\sum_{t=1}^{T} p(q_{t-1} = i, q_t = j, Y|\lambda^{(g)})}{-\sum_{i=1}^{k} \tau_i}$$

sum both sides: $\sum_{j=1}^{k} \sum_{t=1}^{T} p(q_{t-1} = i, q_t = j, Y|\lambda^{(g)}) = \sum_{j=1}^{k} -a_{i,j} \sum_{i=1}^{k} \tau_i = -\sum_{i=1}^{k} \tau_i \sum_{j=1}^{k} a_{i,j} = -\sum_{i=1}^{k} \tau_i$

substitute: $a_{i,j} = \frac{\sum_{t=1}^{T} p(q_{t-1} = i, q_t = j, Y|\lambda^{(g)})}{\sum_{j=1}^{k} \sum_{t=1}^{T} p(q_{t-1} = i, q_t = j, Y|\lambda^{(g)})} = \frac{\sum_{t=1}^{T} p(q_{t-1} = i, q_t = j, Y|\lambda^{(g)})}{\sum_{t=1}^{T} p(q_{t-1} = i, Y|\lambda^{(g)})}$

Parameter Learning: Third term

$$\mathcal{Q}^{\text{term 3}} = \sum_{q_0=1}^{k} \cdots \sum_{q_T=1}^{k} \sum_{t=1}^{T} \ln b_{q_t}(y_t) p(q, Y|\lambda^{(g)}) = \sum_{j=1}^{k} \sum_{t=1}^{T} \ln b_j(y_t) p(q_t = j, Y|\lambda^{(g)})$$

$$\mathbb{LM}^{\text{term 3}} = \sum_{j=1}^{k} \sum_{t=1}^{T} \ln b_j(y_t) p(q_t = j, Y|\lambda^{(g)}) + \tau \left(\sum_{j=1}^{k} b_j(y_t) - 1 \right)$$

$$\frac{\partial \mathbb{LM}^{\text{term 3}}}{\partial b_j(y_t)} = \frac{\sum_{t=1}^{T} p(q_t = j, Y|\lambda^{(g)})}{b_j(y_t)} + \sum_{i=1}^{k} \tau = 0$$

$$\frac{\partial \mathbb{LM}^{\text{term 3}}}{\partial \tau} = \sum_{i=1}^{k} b_j(y_t) - 1 = 0$$

$$\sum_{t=1}^{T} p(q_t = j, Y|\lambda^{(g)}) = -b_j(y_t)\tau \implies b_j(y_t) = \frac{\sum_{t=1}^{T} p(q_t = j, Y|\lambda^{(g)})}{-\tau}$$

sum both sides: (can't use index j) : $\sum_{l=1}^{k} \sum_{t=1}^{T} p(q_t = j, Y = v_l|\lambda^{(g)}) = -\tau \sum_{l=1}^{k} b_j(y_t = v_l) = -\tau$

substitute: $b_j(y_t = v_l) = \dfrac{\sum_{t=1}^{T} p(q_t = j, Y = v_l|\lambda^{(g)})}{\sum_{j=1}^{k} \sum_{t=1}^{T} p(q_t = j, Y = v_l|\lambda^{(g)})} = \dfrac{\sum_{t=1}^{T} p(q_t = j, Y|\lambda^{(g)})\delta_{y_t, v_l}}{\sum_{t=1}^{T} p(q_t = j, Y|\lambda^{(g)})}$