Method

Sun Minni

March 27, 2025

1 model

We use a switching state space model similar to [1]. In our model, leaving and returning states are represented by a discrete latent variable z_t, where $z_t = 0$ indicates returning state, and $z_t = 1$ indicates leaving state. Memory of entry $m_t^{(0)}$, and memory of exit $m_t^{(1)}$ are 2-d continuous latent variables. At time step t, z_t selects memory $m_t^{(z_t)}$ as the current 2-d goal velocity. The joint probability of velocities and latent variables can be factored as:

$$
\begin{aligned}
p(z_{1:n}, m_{1:n}^{(0)}, m_{1:n}^{(1)}, v_{1:n}) = p(z_1) \prod_{t=2}^{n} p(z_t | z_{t-1}, o_t) \\
\cdot \prod_{k=0}^{1} p(m_1^{(k)}) \prod_{t=2}^{n} p(m_t^{(k)} | m_{t-1}^{(k)}, o_{(t-c):t}, v_{(t-c):(t-1)}) \\
\cdot \prod_{t=1}^{n} p(v_t | m_t^{(0)}, m_t^{(1)}, z_t)
\end{aligned}
$$

where hyperparameter c is the interval (in number of time steps) over which we define a crossing (entry or exit) event. The initial state $z_1 \sim \text{Bernoulli}(p_1)$; initial entry memory $m_1^{(0)} \sim \mathcal{N}(\mu_1^{(0)}, \sigma_1^{(0)} I)$; initial exit memory $m_1^{(1)} \sim \mathcal{N}(\mu_1^{(1)}, \sigma_1^{(1)} I)$. We set $\mu_1^{(0)}$ to $\mathbf{0}$ and $\sigma_1^{(0)}$ to 0 so that the initial entry memory is a null vector.

The transition matrix of discrete latent variable:

$$
T = \begin{bmatrix} p(z_t = 0 | z_{t-1} = 0) & p(z_t = 1 | z_{t-1} = 0) \\ \\ p(z_t = 0 | z_{t-1} = 1) & p(z_t = 1 | z_{t-1} = 1) \end{bmatrix}
$$

$$
= \begin{cases} \begin{bmatrix} p_{0\to0} & 1 - p_{0\to0} \\ 0 & 1 \end{bmatrix} \text{ if } o_t = 1 \\ \\ \begin{bmatrix} 1 & 0 \\ 1 - p_{1\to1} & p_{1\to1} \end{bmatrix} \text{ if } o_t = 0 \end{cases}
$$

After training, we refit the duration in returning state after entry $\Delta t_{z=0,o=1}$ with negative binomial distribution $\text{NB}(r_{0\to0}, 1 - p_{0\to0})$, the duration in leaving state after exit $\Delta t_{z=1,o=0}$ with $\text{NB}(r_{1\to1}, 1 - p_{1\to1})$. This allows the mode of the distribution to be larger than 0.

We define an entry event at time step t as:

$$o_i = 0 \text{ for } t-c \leq i < t-\frac{c}{2}, \text{ and } o_i = 1 \text{ for } t-\frac{c}{2} \leq i \leq t$$

an exit event at time step t as:

$$o_i = 1 \text{ for } t-c \leq i < t-\frac{c}{2}, \text{ and } o_i = 0 \text{ for } t-\frac{c}{2} \leq i \leq t$$

For an event at time step t, the crossing travel direction is defined as:

$$\hat{v}_{t-1} = \frac{\overline{v}_{t-1}}{\|\overline{v}_{t-1}\|}, \text{ where } \overline{v}_{t-1} = \frac{1}{c}\sum_{t-c}^{t-1} v_i$$

The continuous latent variable $m_t^{(k)}$ is only updated by $\hat{v}_{t-1}$ at entry (for k=0) or at exit (for k=1).

$$m_t^{(k)} = \begin{cases} a_m^{(k)} m_{t-1}^{(k)} + b_m^{(k)} \hat{v}_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,\ \sigma_m^{(k)} I) & \text{if event} \\ m_{t-1}^{(k)} & \text{otherwise} \end{cases}$$

The scale of $m^{(k)}$ is controlled by the parameter $b_m^{(k)}$, which can be interpreted as $(1-a_m^{(k)}) \cdot s^{(k)}$ where $s^{(k)}$ is the goal speed for state k.

Velocity v_t is modeled as a linear combination of velocity at the previous time step v_{t-1} and the current goal velocity $m_t^{(z_t)}$.

$$p(v_t | m_t, z_t = k) = \mathcal{N}(v_t \mid a_v^{(k)} v_{t-1} + (1-a_v^{(k)}) m_t^{(k)},\ \sigma_v^{(k)} I)$$

2 variational inference

To learn the parameters of the model and infer the latent variables, we optimize the evidence lower bound $\mathcal{L}_q(\theta)$ on the log probability of observed data $\log p(v|\theta)$:

$$\mathcal{L}_q(\theta) = \mathbb{E}_{q(z,m)}[\log p(z,m,v|\theta) - \log q(z,m)]$$

$\mathcal{L}_q(\theta) \leq \log p(v|\theta)$ from Gibbs' inequality, with equality if and only if $q(z,m|\phi) = p(z,m|v,\theta)$. We use the mean field approximation:

$$q(z,m) = q(z)q(m^{(0)})q(m^{(1)})$$

then we have:

$$\mathcal{L}_q(\theta) = \mathbb{E}_{q(z)q(m^{(0)})q(m^{(1)})}[\log p(z,m,v|\theta) - \log q(z) - \log q(m^{(0)}) - \log q(m^{(1)})]$$

We alternate between updating 1) $q(m^{(0)}), q(m^{(1)})$; 2) $q(z)$; 3) θ.

2.1 $q(m^{(k)})$

Optimal coordinate ascent for $q(m^{(k)})$: $q^*(m^{(k)}) \propto \exp(\mathbb{E}_{q(z)q(m^{(1-k)})}[\log p(z,m,v|\theta)])$.

$$
\begin{aligned}
\mathbb{E}_{q(z)q(m^{(1-k)})}[\log p(z,m,v|\theta)] &= \mathbb{E}_{q(z)q(m^{(1-k)})}[\log p(m_1^{(k)}|\theta) + \sum_{t=2}^{n} \log p(m_t^{(k)}|m_{t-1}^{(k)},\theta) \\
&\quad + \sum_{t=1}^{n} \log p(v_t|m_t^{(k)}, z_t = k, \theta)^{\mathbb{1}\{z_t=k\}}] + \text{const} \\
&= \phi(m_1^{(k)}) + \sum_{t=2}^{n} \phi(m_{t-1}^{(k)}, m_t^{(k)}) + \sum_{t=1}^{n} \phi(m_t^{(k)}, v_t) + \text{const}
\end{aligned}
$$

where we introduce the potentials

$$
\begin{aligned}
\phi(m_1^{(k)}) &= \log p(m_1^{(k)}|\theta) \\
\phi(m_{t-1}^{(k)}, m_t^{(k)}) &= \log p(m_t^{(k)}|m_{t-1}^{(k)},\theta) \\
\phi(m_t^{(k)}, v_t) &= \mathbb{E}_{q(z)}[\log p(v_t|m_t^{(k)}, z_t = k,\theta)^{\mathbb{1}\{z_t=k\}}] \\
&= \mathbb{E}_{q(z)}[\mathbb{1}\{z_t = k\} \log p(v_t|m_t^{(k)}, z_t = k,\theta)] \\
&= h_t^{(k)} \log p(v_t|m_t^{(k)}, z_t = k,\theta), \text{ where } h_t^{(k)} := \mathbb{E}_{q(z)}(\mathbb{1}\{z_t = k\})
\end{aligned}
$$

and the normalizing constant $Z_{m^{(k)}}(\theta)$ such that:

$$
q(m^{(k)}) = \frac{1}{Z_{m^{(k)}}(\theta)} \exp(\ \phi(m_1^{(k)}) + \sum_{t=2}^{n} \phi(m_{t-1}^{(k)}, m_t^{(k)}) + \sum_{t=1}^{n} \phi(m_t^{(k)}, v_t)\ )
$$

we can calculate $q(m_t^{(k)}), q(m_{t-1}^{(k)}, m_t^{(k)})$ using Kalman filter [2] and RTS smoother [4] algorithms with observation noise covariance matrix divided by $h_t^{(k)}$ [1].

Below we show the calculation of $Z_m(\theta)$. We omit (k) and write the boundary term $\phi(m_1)$ as $\phi(m_0, m_1)$ for simplicity. We define $f_{t|t}(m_t)$ as:

$$
f_{t|t}(m_t) := \int dm_1 \int dm_2 ... \int dm_{t-1} \exp(\sum_{t=1}^{n} \phi(m_{t-1}, m_t) + \sum_{t=1}^{n} \phi(m_t, v_t))
$$

Terms in the exponent are all quadratic functions in our setting, so $f_{t|t}(m_t)$ is proportional to a Gaussian density function and can be written as

$$
f_{t|t}(m_t) = \mathcal{N}(m_t|\mu_{t|t}, \sigma_{t|t} I) \cdot Z_t, \text{ where } Z_t := \int dm_t f_{t|t}(m_t)
$$

We can see that $Z_m(\theta) = \int dm_n f_{n|n}(m_n) = Z_n$. The calculation can be done recursively. To calculate Z_{t+1} from Z_t at a general time step t, we define $f_{t+1|t}(m_{t+1})$ as:

$$
f_{t+1|t}(m_{t+1}) := \int dm_t f_{t|t}(m_t) \exp(\phi(m_t, m_{t+1})) = Z_t \int dm_t \mathcal{N}(m_t|\mu_{t|t}, \sigma_{t|t} I) p(m_{t+1}|m_t,\theta)
$$

$p(m_{t+1}|m_t,\theta)$ is a Gaussian density function, so the marginal distribution $\int dm_t \mathcal{N}(m_t|\mu_{t|t}, \sigma_{t|t} I) p(m_t|m_{t-1},\theta)$ is also Gaussian, we write it as $\mathcal{N}(m_{t+1}|\mu_{t+1|t}, \sigma_{t+1|t} I)$, then:

$$
f_{t+1|t}(m_{t+1}) = Z_t \cdot \mathcal{N}(m_{t+1}|\mu_{t+1|t}, \sigma_{t+1|t} I)
$$

So for Z_{t+1}:

$$\begin{aligned}
Z_{t+1} &= \int dm_{t+1} f_{t+1|t+1}(m_{t+1}) \\
&= \int dm_{t+1}\ f_{t+1|t}(m_{t+1}) \cdot \exp(\phi(m_{t+1}, v_{t+1})) \\
&= Z_t \int dm_{t+1}\ \mathcal{N}(m_{t+1}|\mu_{t+1|t}, \sigma_{t+1|t} I) \cdot \mathcal{N}(v_{t+1}|a_v v_t + (1-a_v)m_{t+1}, \sigma_v I)^{h_{t+1}}
\end{aligned}$$

We can see that $Z_{t+1} = Z_t$ if $h_{t+1} = 0$.
For a 2-d Gaussian density function to the power of h ($h \neq 0$):

$$\begin{aligned}
\mathcal{N}(v|\mu_v, \sigma_v I)^h &= (\frac{1}{2\pi\sigma_v})^h \exp(-\frac{h}{2\sigma_v}(v-\mu_v)^T(v-\mu_v)) \\
&= (\frac{1}{2\pi\sigma_v})^h\ \frac{2\pi\sigma_v}{h} \frac{h}{2\pi\sigma_v} \exp(-\frac{h}{2\sigma_v}(v-\mu_v)^T(v-\mu_v)) \\
&= (\frac{1}{2\pi\sigma_v})^h\ \frac{2\pi\sigma_v}{h}\ \mathcal{N}(v|\mu_v, \frac{\sigma_v}{h} I)
\end{aligned}$$

So for Z_{t+1} if $h_{t+1} \neq 0$:

$$\begin{aligned}
Z_{t+1} &= Z_t \int dm_{t+1}\ \mathcal{N}(m_{t+1}|\mu_{t+1|t}, \sigma_{t+1|t} I) \cdot (\frac{1}{2\pi\sigma_v})^{h_{t+1}}\ \frac{2\pi\sigma_v}{h_{t+1}}\ \mathcal{N}(v_{t+1}|a_v v_t + (1-a_v)m_{t+1}, \frac{\sigma_v}{h_{t+1}} I) \\
&= Z_t \cdot (\frac{1}{2\pi\sigma_v})^{h_{t+1}}\ \frac{2\pi\sigma_v}{h_{t+1}} \int dm_{t+1}\ \mathcal{N}(m_{t+1}|\mu_{t+1|t}, \sigma_{t+1|t} I)\mathcal{N}(v_{t+1}|a_v v_t + (1-a_v)m_{t+1}, \frac{\sigma_v}{h_{t+1}} I) \\
&= Z_t \cdot (\frac{1}{2\pi\sigma_v})^{h_{t+1}}\ \frac{2\pi\sigma_v}{h_{t+1}}\ \mathcal{N}(v_{t+1}|a_v v_t + (1-a_v)\mu_{t+1|t}, (1-a_v)^2\sigma_{t+1|t} I + \frac{\sigma_v}{h_{t+1}} I) \\
&= Z_t \cdot (\frac{1}{2\pi\sigma_v})^{h_{t+1}}\ \frac{2\pi\sigma_v}{h_{t+1}} \cdot \frac{h_{t+1}}{2\pi(h_{t+1}(1-a_v)^2\sigma_{t+1|t} + \sigma_v)} \\
&\quad \cdot \exp(-\frac{h_{t+1}}{2(h_{t+1}(1-a_v)^2\sigma_{t+1|t} + \sigma_v)} \|v_{t+1} - a_v v_t - (1-a_v)\mu_{t+1|t}\|^2) \\
&= Z_t \cdot (\frac{1}{2\pi\sigma_v})^{h_{t+1}}\ \frac{\sigma_v}{h_{t+1}(1-a_v)^2\sigma_{t+1|t} + \sigma_v} \\
&\quad \cdot \exp(-\frac{h_{t+1}}{2(h_{t+1}(1-a_v)^2\sigma_{t+1|t} + \sigma_v)} \|v_{t+1} - a_v v_t - (1-a_v)\mu_{t+1|t}\|^2)
\end{aligned}$$

The above equation holds true for $h_{t+1} = 0$ as well. Thus for the normalizing constant:

$$Z_n = \prod_{t=1}^{n} (\frac{1}{2\pi\sigma_v})^{h_t}\ \frac{\sigma_v}{h_t(1-a_v)^2\sigma_{t|t-1} + \sigma_v} \cdot \exp(-\frac{h_t}{2(h_t(1-a_v)^2\sigma_{t|t-1} + \sigma_v)} \|v_t - a_v v_{t-1} - (1-a_v)\mu_{t|t-1}\|^2)$$

2.2 $q(z)$

Optimal coordinate ascent for $q(z)$: $q^*(z) \propto \exp(\mathbb{E}_{q(m)}[\log p(z, m, v|\theta)])$.

$$\begin{aligned}\mathbb{E}_{q(m)}[\log p(z, m, v|\theta)] &= \mathbb{E}_{q(m)}[\log p(z_1|\theta) + \sum_{t=2}^{n} \log p(z_t|z_{t-1}, o_t, \theta) + \sum_{t=1}^{n} \log p(v_t|m_t^{(0)}, m_t^{(1)}, z_t, \theta)] + \text{const} \\ &= \phi(z_1) + \sum_{t=2}^{n} \phi(z_{t-1}, z_t) + \sum_{t=1}^{n} \phi(z_t, v_t) + \text{const}\end{aligned}$$

where we introduce the potentials

$$\begin{aligned}\phi(z_1) &= \log p(z_1|\theta) \\ \phi(z_{t-1}, z_t) &= \log p(z_t|z_{t-1}, o_t, \theta) \\ \phi(z_t, v_t) &= \mathbb{E}_{q(m)}[\log p(v_t|m_t, z_t, \theta)] \\ &= \mathbb{E}_{q(m)}[\log \prod_{k=0}^{1} p(v_t|m_t, z_t = k, \theta)^{\mathbb{1}\{z_t = k\}}] \\ &= \mathbb{E}_{q(m)}[\sum_{k=0}^{1} \mathbb{1}\{z_t = k\} \log p(v_t|m_t, z_t = k, \theta)] \\ &= \sum_{k=0}^{1} \mathbb{1}\{z_t = k\} \mathbb{E}_{q(m)}[\log p(v_t|m_t^{(k)}, z_t = k, \theta)] \\ &= \sum_{k=0}^{1} \mathbb{1}\{z_t = k\} \log q_t^{(k)}, \text{ where } \log q_t^{(k)} := \mathbb{E}_{q(m^{(k)})}[\log p(v_t|m_t^{(k)}, z_t = k, \theta)]\end{aligned}$$

and the normalizing constant $Z_z(\theta)$ such that

$$q(z) = \frac{1}{Z_z(\theta)} \exp(\ \phi(z_1) + \sum_{t=2}^{n} \phi(z_{t-1}, z_t) + \sum_{t=1}^{n} \phi(z_t, v_t)\ )$$

We can calculate $q(z_t), q(z_{t-1}, z_t), Z_z(\theta)$ using forward backward algorithm with observation weights $q_t^{(k)}$.

3 algorithm

(code implemented based on [3])
Initialization:

- initialize θ:
  $a_m^{(k)}, b_m^{(k)}, \sigma_m^{(k)}, a_v^{(k)}, \sigma_v^{(k)}, \sigma_1^{(1)}$ are initialized to be 0.5; $\mu_1^{(0)} = \mathbf{0}, \sigma_1^{(0)} = 0$;
  $\mu_1^{(1)} \sim \text{Uniform}((-0.5, 0.5), (-0.5, 0.5)), p_{0\to 0} \sim \text{Uniform}(0.1, 0.9), p_{1\to 1} \sim \text{Uniform}(0.1, 0.9)$
- initialize $q(m^{(k)})$: $h_t^{(0)} = 1 - o_t,\ h_t^{(1)} = o_t$

Iteration:

- calculate $q(m_t^{(k)}), q(m_{t-1}^{(k)}, m_t^{(k)})$ for $t = 1, ..., n$; calculate $Z_{m^{(k)}}(\theta)$

- update $q(z)$: $q_t^{(k)} = \exp \mathbb{E}_{q(m^{(k)})}[\log p(v_t|m_t^{(k)}, z_t = k, \theta)]$
- calculate $q(z_t), q(z_{t-1}, z_t)$ for $t = 1, ..., n$; calculate $Z_z(\theta)$
- compute current ELBO to check convergence:

$$\begin{aligned}
\mathcal{L}_q &= \mathbb{E}_{q(z)q(m^{(0)})q(m^{(1)})}[\log p(z, m, v|\theta) - \log q(z) - \log q(m^{(0)}) - \log q(m^{(1)})] \\
&= \mathbb{E}_{q(z)q(m^{(0)})q(m^{(1)})}[\sum_{t=1}^{n}\sum_{k=0}^{1} \mathbb{1}\{z_t = k\} \log p(v_t|m_t^{(k)}, z_t = k, \theta) - \sum_{t=1}^{n}\sum_{k=0}^{1} \mathbb{1}\{z_t = k\} \log q_t^{(k)} \\
&\quad + \log(Z_z(\theta)) - \sum_{t=1}^{n}\sum_{k=0}^{1} h_t^{(k)} \log p(v_t|m_t^{(k)}, z_t = k, \theta) + \sum_{k=0}^{1} \log(Z_{m^{(k)}}(\theta))] \\
&= -\sum_{t=1}^{n}\sum_{k=0}^{1} h_t^{(k)} \log q_t^{(k)} + \log(Z_z(\theta)) + \sum_{k=0}^{1} \log(Z_{m^{(k)}}(\theta))
\end{aligned}$$

- optimize θ with sufficient statistics calculated by $q(m^{(k)}|\theta_{\text{old}})$, $q(z|\theta_{\text{old}})$

$$\theta_{\text{new}} = \arg\max_{\theta} \mathbb{E}_{q(z,m|\theta_{\text{old}})}[\log p(z, m, v|\theta)]$$

- update $q(m)$: $h_t^{(k)} = \mathbb{E}_{q(z)}(\mathbb{1}\{z_t = k\})$

References

[1] Zoubin Ghahramani and Geoffrey E Hinton. "Variational learning for switching state-space models". In: *Neural computation* 12.4 (2000), pp. 831–864.

[2] Rudolph Emil Kalman. "A new approach to linear filtering and prediction problems". In: *Journal of Basic Engineering* 82.1 (1960), pp. 35–45.

[3] Scott Linderman et al. "SSM: Bayesian learning and inference for state space models". In: *github* (2020).

[4] Herbert E Rauch, F Tung, and Charlotte T Striebel. "Maximum likelihood estimates of linear dynamic systems". In: *AIAA journal* 3.8 (1965), pp. 1445–1450.