HMM vs HSMM: Time-Dependent Transition Kernels

Setup

Base TinyMoA-style sticky matrix:

$$\mathbf{T}_{\text{base}} = \begin{pmatrix} 0.98 & 0.01 & 0.01 \\ 0.01 & 0.98 & 0.01 \\ 0.01 & 0.01 & 0.98 \end{pmatrix}.$$

In a plain HMM, this *same* matrix is used every step. The per-step exit prob. from any state is $1 - 0.98 = 0.02$, so dwell $D \sim \text{Geom}(0.02)$ with

$$\mathbb{E}[D] = \frac{1}{0.02} = 50, \qquad \text{median } \tilde{d} \approx \frac{\ln 2}{0.02} \approx 34.7.$$

HSMM

Objects.

$$D_i \in \{1, 2, \dots\}, \qquad p_i(d) = \Pr(D_i = d), \qquad S_i(d) = \Pr(D_i \geq d) = \sum_{u=d}^{\infty} p_i(u).$$

Meaning: D_i is the segment length (dwell) in state i. $p_i(d)$ is the chance the segment lasts exactly d trials. $S_i(d)$ is the chance it survives *to at least* d.

Hazard (end-now probability).

$$h_i(d) = \Pr(D_i = d \mid D_i \geq d) = \frac{p_i(d)}{S_i(d)}.$$

Meaning: given you have lasted to step d, $h_i(d)$ is the probability the segment *ends at* d. If $h_i(d)$ grows with d, you encode "momentum" (sticky early, more likely to switch later).

Effective one-step transitions at dwell d.

$$\boxed{\tilde{T}_{ii}^{(d)} = 1 - h_i(d), \qquad \tilde{T}_{ij}^{(d)} = h_i(d)\,\frac{T_{ij}}{1 - T_{ii}} \quad (j \neq i)}$$

Meaning: while you are still in i, the total chance of *leaving* is $h_i(d)$; if you leave, the *relative* destinations follow the base HMM row of $\mathbf{T}$ (the off-diagonal proportions are preserved). Thus the kernel becomes time-dependent via $h_i(d)$.

HMM as a special case. If $D_i \sim \mathrm{Geom}(\lambda)$ with $\lambda = 1 - T_{ii}$,

$$h_i(d) \equiv \lambda \quad \Rightarrow \quad \tilde{\mathbf{T}}^{(d)} \equiv \mathbf{T} \quad \text{for all } d.$$

Meaning: geometric durations have a *constant* end-now probability; the transition matrix never changes over time.

Moments.

$$\mathbb{E}[D_i] = \sum_{d\ge 1} S_i(d), \qquad \mathrm{Var}(D_i) = \sum_{d\ge 1}(2d-1)S_i(d) - \mathbb{E}[D_i]^2.$$

Meaning: these give mean/variance of dwell directly from the survival curve; useful for matching empirical dwell statistics.

example: first 50 trials side-by-side

We start both HMM and HSMM at $\mathbf{T}_{\text{base}}$. For HSMM, use a piecewise-constant hazard (same for all states):

$$h(d) = \begin{cases} 0.010, & 1 \le d \le 10, \\ 0.015, & 11 \le d \le 20, \\ 0.020, & 21 \le d \le 30, \\ 0.030, & 31 \le d \le 40, \\ 0.050, & 41 \le d \le 50\ , \end{cases}$$

which still targets a mean near ≈ 50 but greatly increases the chance to end long segments.

HMM: kernels for trials $t = 1, \dots, 50$

$$\tilde{\mathbf{T}}^{(t)}_{\text{HMM}} = \mathbf{T}_{\text{base}} = \begin{pmatrix} 0.98 & 0.01 & 0.01 \\ 0.01 & 0.98 & 0.01 \\ 0.01 & 0.01 & 0.98 \end{pmatrix} \quad \text{for every } t = 1, \dots, 50.$$

Conclusion: identical at all trials.

HSMM: kernels for dwell $d = 1, \dots, 50$

With equal off-diagonals in $\mathbf{T}_{\text{base}}$, each off-diagonal is $\frac{1}{2}h(d)$ and the diagonal is $1 - h(d)$:

$$\tilde{\mathbf{T}}^{(d)}_{\text{HSMM}} = \begin{pmatrix} 1-h(d) & \frac{1}{2}h(d) & \frac{1}{2}h(d) \\ \frac{1}{2}h(d) & 1-h(d) & \frac{1}{2}h(d) \\ \frac{1}{2}h(d) & \frac{1}{2}h(d) & 1-h(d) \end{pmatrix}.$$

Plugging the schedule gives the *first 50* effective matrices as five blocks:

dwell d	$h(d)$	$\tilde{\mathbf{T}}^{(d)}_{\text{HSMM}}$
1:10	0.010	$\begin{pmatrix} 0.990 & 0.005 & 0.005 \\ 0.005 & 0.990 & 0.005 \\ 0.005 & 0.005 & 0.990 \end{pmatrix}$
11:20	0.015	$\begin{pmatrix} 0.985 & 0.0075 & 0.0075 \\ 0.0075 & 0.985 & 0.0075 \\ 0.0075 & 0.0075 & 0.985 \end{pmatrix}$
21:30	0.020	$\begin{pmatrix} 0.980 & 0.010 & 0.010 \\ 0.010 & 0.980 & 0.010 \\ 0.010 & 0.010 & 0.980 \end{pmatrix}$
31:40	0.030	$\begin{pmatrix} 0.970 & 0.015 & 0.015 \\ 0.015 & 0.970 & 0.015 \\ 0.015 & 0.015 & 0.970 \end{pmatrix}$
41:50	0.050	$\begin{pmatrix} 0.950 & 0.025 & 0.025 \\ 0.025 & 0.950 & 0.025 \\ 0.025 & 0.025 & 0.950 \end{pmatrix}$