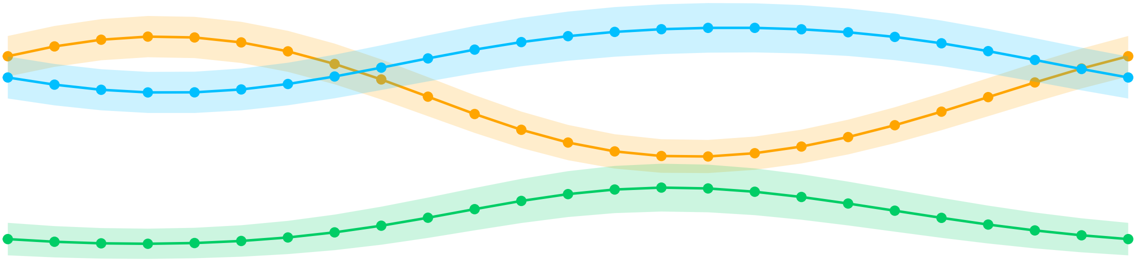
ZeSt Kolloquium

Mitigating consequences of the Markov property

Jan-Ole Koslik

October 24, 2023

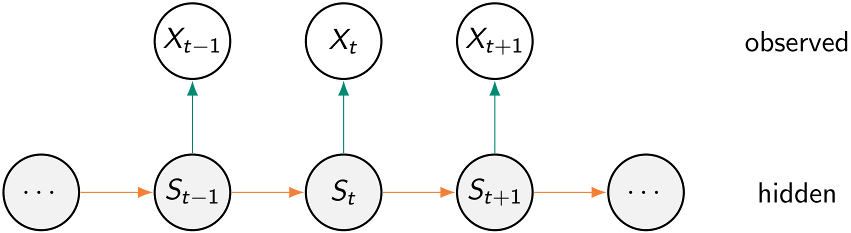


Outline

Outline

Quick Recap on HMMs

Doubly stochastic process:



- every observation is generated by one of N possible distributions $f_1, \ldots, f_N$,
- the state process selects which distribution is active at any given time point
- the state process is a **Markov chain**

Reminder: Markov chains

A Markov chain is a sequence of random variables $S_1, S_2, \dots$ that takes values in the *state space* $\{1, \dots, N\}$ and fulfills the *Markov property*

$$\Pr(S_{t+1} = s_{t+1} \mid S_1 = s_1, \dots, S_t = s_t) = \Pr(S_{t+1} = s_{t+1} \mid S_t = s_t).$$

Reminder: Markov chains

A Markov chain is a sequence of random variables $S_1, S_2, \dots$ that takes values in the *state space* $\{1, \dots, N\}$ and fulfills the *Markov property*

$$\Pr(S_{t+1} = s_{t+1} \mid S_1 = s_1, \dots, S_t = s_t) = \Pr(S_{t+1} = s_{t+1} \mid S_t = s_t).$$

Due to this, we can characterise the state process by specifying the **initial distribution**

$$\boldsymbol{\delta}^{(1)} = (\Pr(S_1 = 1), \dots, \Pr(S_1 = N))$$

and the **transition probabilities**

$$\gamma_{ij}^{(t)} = \Pr(S_{t+1} = j \mid S_t = i).$$

Reminder: Markov chains

A Markov chain is a sequence of random variables $S_1, S_2, \dots$ that takes values in the *state space* $\{1, \dots, N\}$ and fulfills the *Markov property*

$$\Pr(S_{t+1} = s_{t+1} \mid S_1 = s_1, \dots, S_t = s_t) = \Pr(S_{t+1} = s_{t+1} \mid S_t = s_t).$$

Due to this, we can characterise the state process by specifying the **initial distribution**

$$\boldsymbol{\delta}^{(1)} = (\Pr(S_1 = 1), \dots, \Pr(S_1 = N))$$

and the **transition probabilities**

$$\gamma_{ij}^{(t)} = \Pr(S_{t+1} = j \mid S_t = i).$$

We summarise the transition probabilities in the **transition probability matrix** (t.p.m.)

$$\boldsymbol{\Gamma}^{(t)} = (\gamma_{ij}^{(t)})_{i,j=1,\dots,N}.$$

Why deal with the Markov property?

The Markov property and state dwell-times

In a homogeneous Markov chain, leaving a state can be interpreted as a repeated Bernoulli trial

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & 1 - \gamma_{11} \\ 1 - \gamma_{22} & \gamma_{22} \end{pmatrix}$$

The Markov property and state dwell-times

In a homogeneous Markov chain, leaving a state can be interpreted as a repeated Bernoulli trial

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & 1 - \gamma_{11} \\ 1 - \gamma_{22} & \gamma_{22} \end{pmatrix}$$

$$\Pr(R = 1) = 1 - \gamma_{11}$$

The Markov property and state dwell-times

In a homogeneous Markov chain, leaving a state can be interpreted as a repeated Bernoulli trial

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & 1-\gamma_{11} \\ 1-\gamma_{22} & \gamma_{22} \end{pmatrix}$$

$$\Pr(R=1) = 1-\gamma_{11}$$
$$\Pr(R=2) = \gamma_{11}(1-\gamma_{11})$$

The Markov property and state dwell-times

In a homogeneous Markov chain, leaving a state can be interpreted as a repeated Bernoulli trial

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & 1-\gamma_{11} \\ 1-\gamma_{22} & \gamma_{22} \end{pmatrix}$$

$$\begin{aligned}
\Pr(R=1) &= 1-\gamma_{11} \\
\Pr(R=2) &= \gamma_{11}\,(1-\gamma_{11}) \\
&\ldots \\
\Pr(R=r) &= \gamma_{11}^{r-1}\,(1-\gamma_{11})
\end{aligned}$$

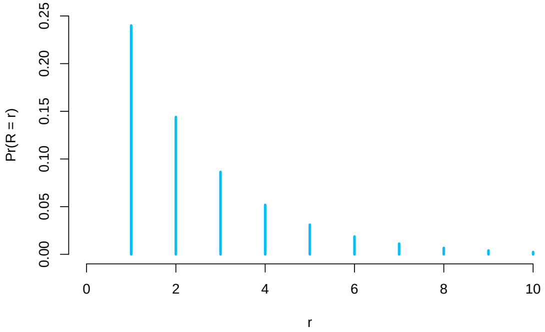
The Markov property and state dwell-times

In a homogeneous Markov chain, leaving a state can be interpreted as a repeated Bernoulli trial

$$\boldsymbol{\Gamma} = \begin{pmatrix} \gamma_{11} & 1-\gamma_{11} \\ 1-\gamma_{22} & \gamma_{22} \end{pmatrix}$$

$$\Pr(R=1) = 1-\gamma_{11}$$
$$\Pr(R=2) = \gamma_{11}(1-\gamma_{11})$$
$$\dots$$
$$\Pr(R=r) = \gamma_{11}^{r-1}(1-\gamma_{11})$$

$R \sim \text{Geom}(1-\gamma_{11})$



The Markov property and state dwell times

In many scenarios, this is unrealistic:

- **Sleeping:** We expect a mode > 1

The Markov property and state dwell times

In many scenarios, this is unrealistic:

- **Sleeping:** We expect a mode > 1
- **Whale foraging:**
  1. Getting to feeding grounds requires a lot of energy: most probable dwell time should not be one (Langrock et al., 2014)
  2. Memorylessness is problematic as whales need to resurface for breathing

The Markov property and state dwell times

In many scenarios, this is unrealistic:

- **Sleeping:** We expect a mode > 1
- **Whale foraging:**
  1. Getting to feeding grounds requires a lot of energy: most probable dwell time should not be one (Langrock et al., 2014)
  2. Memorylessness is problematic as whales need to resurface for breathing
- **Finance:** Time-spent in a particular market regime often shows longer right tail than geometric distribution (J. Bulla and I. Bulla, 2006).

The Markov property and state dwell times

In many scenarios, this is unrealistic:

- **Sleeping:** We expect a mode > 1
- **Whale foraging:**
  1. Getting to feeding grounds requires a lot of energy: most probable dwell time should not be one (Langrock et al., 2014)
  2. Memorylessness is problematic as whales need to resurface for breathing
- **Finance:** Time-spent in a particular market regime often shows longer right tail than geometric distribution (J. Bulla and I. Bulla, 2006).

$\rightarrow$ One of the most important points of criticism for HMMs.

The Markov property and state dwell times

In many scenarios, this is unrealistic:

- **Sleeping:** We expect a mode > 1
- **Whale foraging:**
  1. Getting to feeding grounds requires a lot of energy: most probable dwell time should not be one (Langrock et al., 2014)
  2. Memorylessness is problematic as whales need to resurface for breathing
- **Finance:** Time-spent in a particular market regime often shows longer right tail than geometric distribution (J. Bulla and I. Bulla, 2006).

$\rightarrow$ One of the most important points of criticism for HMMs.

The **Markov property** and **geometric** dwell times are two sides of the same coin: **Memorylessness** of the process

Motivation

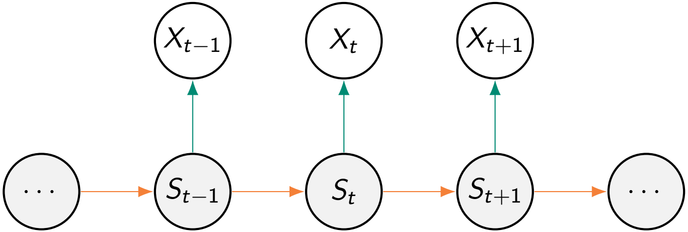
1. How problematic is the Markov property in more complex models? (including periodic variation)

Motivation

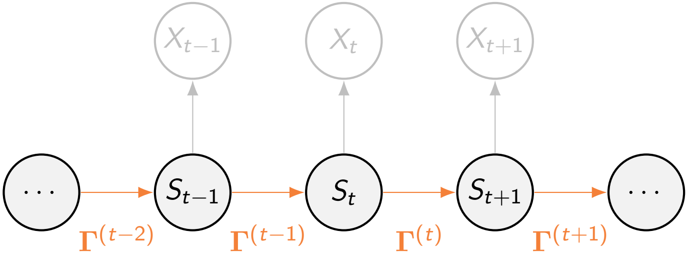
1. How problematic is the Markov property in more complex models? (including periodic variation)
2. What about other options? (Hidden semi-Markov models)

Outline

Inhomogeneity in the state process



Inhomogeneity in the state process



Periodic variation

Periodically varying transition probabilities formally means:

$$\mathbf{\Gamma}^{(t)} = \mathbf{\Gamma}^{(t+L)} \quad \text{for all } t = 1, \ldots, T, \tag{1}$$

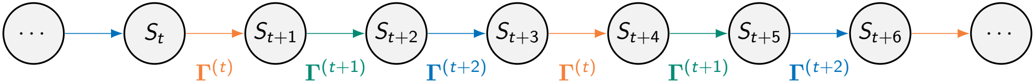
where L is the length of one cycle.

Periodic variation

Periodically varying transition probabilities formally means:

$$\mathbf{\Gamma}^{(t)} = \mathbf{\Gamma}^{(t+L)} \quad \text{for all } t = 1, \ldots, T, \tag{1}$$

where L is the length of one cycle.



Recap: Stationarity

Homogeneous and "well-behaved"[1] Markov chains converge against their stationary distribution.

$$\Pr(S_t = i) \to \delta_i$$

where $\boldsymbol{\delta}$ is the solution to

$$\boldsymbol{\delta}\boldsymbol{\Gamma} = \boldsymbol{\delta}, \quad \text{s.t.} \sum_{i=1}^{N} \delta_i = 1.$$

δ_i is a useful summary statistic: Average time spent in state i.

[1]Finite state space, irreducible and aperiodic.

Periodic stationarity

- We do not expect convergence for inhomogeneous chains ($\mathbf{\Gamma}^{(t)}$ is changing all the time).

Periodic stationarity

- We do not expect convergence for inhomogeneous chains ($\mathbf{\Gamma}^{(t)}$ is changing all the time).
- We are still interested in how $\Pr(S_t = i)$ evolves over the course of one day.

Periodic stationarity

- We do not expect convergence for inhomogeneous chains ($\boldsymbol{\Gamma}^{(t)}$ is changing all the time).
- We are still interested in how $\Pr(S_t = i)$ evolves over the course of one day.
- Previously, approximated by $\boldsymbol{\rho}^{(t)}$ which is the solution to

$$\boldsymbol{\rho}^{(t)}\boldsymbol{\Gamma}^{(t)} = \boldsymbol{\rho}^{(t)}, \quad \text{s.t.} \sum_{i=1}^{N} \rho_i^{(t)} = 1, \quad \text{for } t = 1, \ldots, L.$$

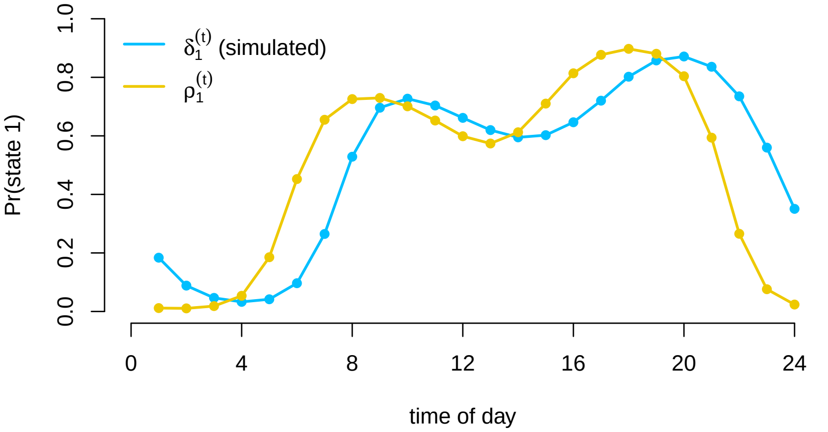
Periodic stationarity

- We do not expect convergence for inhomogeneous chains ($\mathbf{\Gamma}^{(t)}$ is changing all the time).
- We are still interested in how $\Pr(S_t = i)$ evolves over the course of one day.
- Previously, approximated by $\boldsymbol{\rho}^{(t)}$ which is the solution to

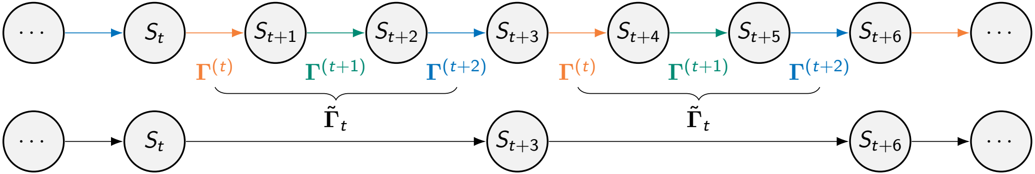
$$\boldsymbol{\rho}^{(t)}\mathbf{\Gamma}^{(t)} = \boldsymbol{\rho}^{(t)}, \quad \text{s.t.} \sum_{i=1}^{N} \rho_i^{(t)} = 1, \quad \text{for } t = 1, \ldots, L.$$

- But this estimate is biased!

Periodic stationarity



Periodic stationarity



Periodic stationarity

- Consider for every $t \in \{1, \dots, L\}$ the thinned Markov chain $S_t, S_{t+L}, S_{t+2L}, \dots$.

Periodic stationarity

- Consider for every $t \in \{1, \dots, L\}$ the thinned Markov chain $S_t, S_{t+L}, S_{t+2L}, \dots$.
- It has constant t.p.m.

$$\tilde{\mathbf{\Gamma}}_t = \mathbf{\Gamma}^{(t)} \cdot \mathbf{\Gamma}^{(t+1)} \cdot \ldots \cdot \mathbf{\Gamma}^{(t+L-1)}.$$

Periodic stationarity

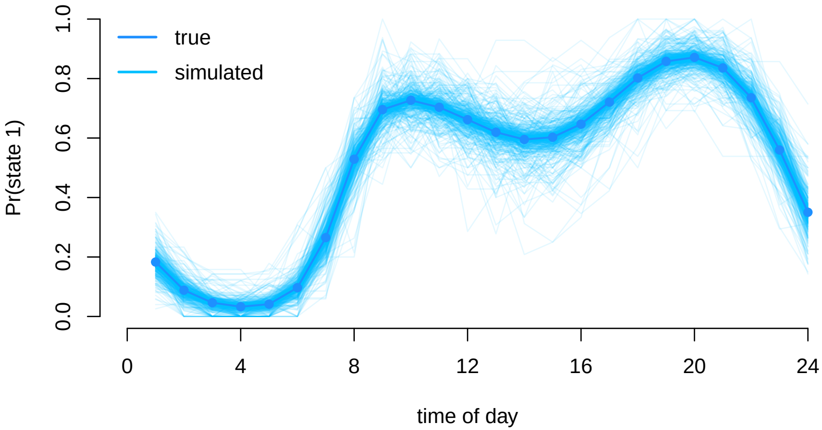
- Consider for every $t \in \{1, \dots, L\}$ the thinned Markov chain $S_t, S_{t+L}, S_{t+2L}, \dots$.
- It has constant t.p.m.

$$\tilde{\boldsymbol{\Gamma}}_t = \boldsymbol{\Gamma}^{(t)} \cdot \boldsymbol{\Gamma}^{(t+1)} \cdot \ldots \cdot \boldsymbol{\Gamma}^{(t+L-1)}.$$

- Thus each thinned chain converges, and we get $\Pr(S_t = i)$ for each t by solving

$$\boldsymbol{\delta}^{(t)} \tilde{\boldsymbol{\Gamma}}_t = \boldsymbol{\delta}^{(t)} \quad \text{s.t.} \sum_{i=1}^{N} \delta_i^{(t)} = 1.$$

Periodic stationarity



Time-varying dwell-time distribution

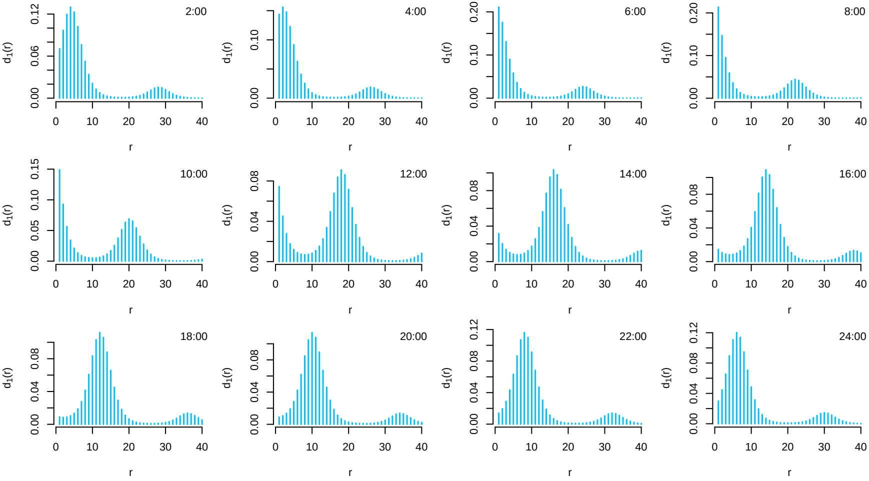
- We are now interested in the implied dwell-time distribution of an inhomogeneous Markov chain.

Time-varying dwell-time distribution

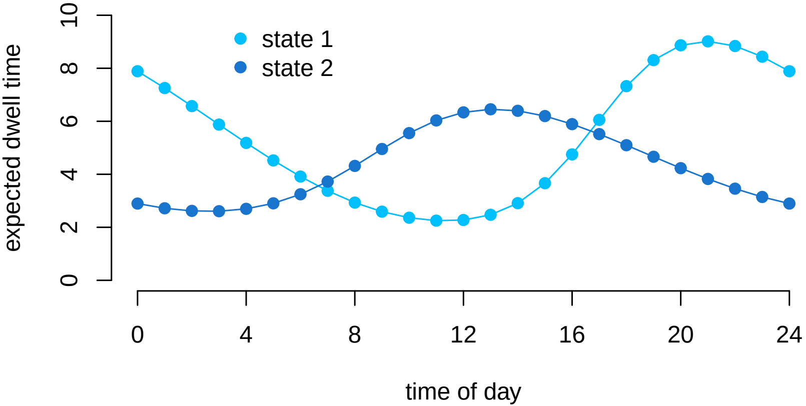
- We are now interested in the implied dwell-time distribution of an inhomogeneous Markov chain.
- To begin with, we can consider the dwell-time distribution at a certain time point:

$$d_i^{(t)}(r) = \underbrace{(1 - \gamma_{ii}^{(t+r-1)})}_{\text{leave}} \cdot \underbrace{\prod_{j=1}^{r-1} \gamma_{ii}^{(t+j-1)}}_{\text{stay r times}}, \qquad r \in \mathbb{N}$$

Time-varying dwell-time distribution



Expected (time-varying) dwell time



Time-varying dwell-time distribution

- The time-varying distribution is already a useful inference tool,

Time-varying dwell-time distribution

- The time-varying distribution is already a useful inference tool,
- but we want something simpler for inference and model checking.

Time-varying dwell-time distribution

- The time-varying distribution is already a useful inference tool,
- but we want something simpler for inference and model checking.
- We want to find the *overall* dwell-time distribution of a given state...

Overall dwell-time distribution

We can obtain the overall dwell-time distribution as a mixture:

$$d_i(r) = \sum_{t=1}^{L} w_i^{(t)} d_i^{(t)}(r), \qquad r \in \mathbb{N}$$

Overall dwell-time distribution

We can obtain the overall dwell-time distribution as a mixture:

$$d_i(r) = \sum_{t=1}^{L} w_i^{(t)} d_i^{(t)}(r), \qquad r \in \mathbb{N}$$

with the mixture weights defined as

$$w_i^{(t)} = \frac{\sum_{l \neq i} \delta_l^{(t-1)} \gamma_{li}^{(t-1)}}{\sum_{t=1}^{L} \sum_{l \neq i} \delta_l^{(t-1)} \gamma_{li}^{(t-1)}}, \qquad t = 1, \ldots, L,$$

where $\boldsymbol{\delta}^{(t)}$ is the periodically stationary distribution.

Overall dwell-time distribution

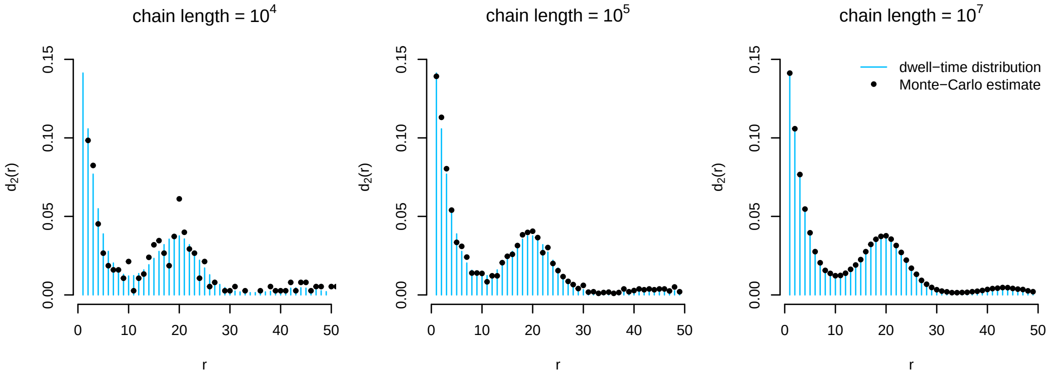
What we have here is

$$\frac{\sum_{t=1}^{L} \Pr(\text{Transition to state } i \text{ at time } t \text{ and stay } r \text{ times})}{\Pr(\text{Transition to state } i \text{ happens at some point of the day})}$$

$$= \frac{\Pr(\text{Transition to state } i \text{ and stay } r \text{ times at some point of the day})}{\Pr(\text{Transition to state } i \text{ happens at some point of the day})}$$

$$= \Pr(\text{Stay } r \text{ times in state } i \mid \text{Transition to } i \text{ happens at some point of the day})$$

Overall dwell-time distribution



Overall dwell-time distribution

Concluding remarks:

- Periodic inhomogeneity can already mitigate undesired consequences of the Markov property.

Overall dwell-time distribution

Concluding remarks:

- Periodic inhomogeneity can already mitigate undesired consequences of the Markov property.
- But this may be limited to scenarios with a large amount of periodic variation.

Overall dwell-time distribution

Concluding remarks:

- Periodic inhomogeneity can already mitigate undesired consequences of the Markov property.
- But this may be limited to scenarios with a large amount of periodic variation.
- Thus, we need even more flexible models...

Outline

Hidden semi-Markov models

- So-called hidden semi-Markov models (HSMMs) are a flexible extension of HMMs explicitly designed to accommodate arbitrary dwell-time distributions.

Hidden semi-Markov models

- So-called hidden semi-Markov models (HSMMs) are a flexible extension of HMMs explicitly designed to accommodate arbitrary dwell-time distributions.
- But estimation and inference become more involved ...

Constructing a semi-Markov chain

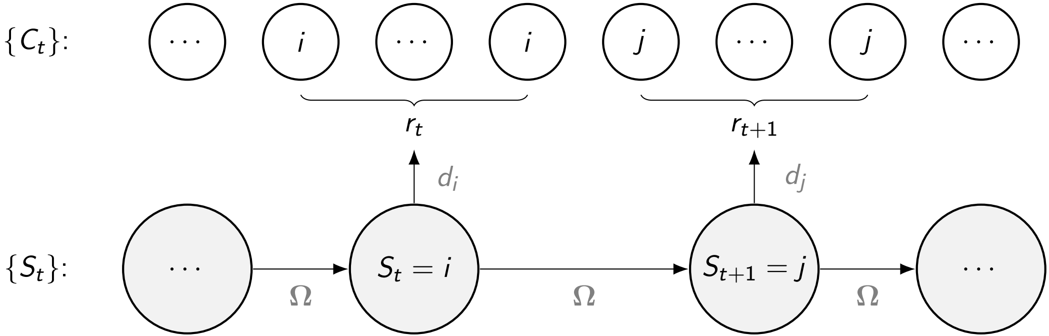
Instead of just a t.p.m., we need two ingredients:

- A set of dwell-time distributions $d_1, d_2, \dots, d_N$ with support $\mathbb{N}$ that determine the time-spent in a state.
- The conditional transition probabilities

$$\omega_{ij} = \Pr(S_{t+1} = j \mid S_t = i, S_{t+1} \neq i), \quad i,j = 1, \dots, N, \quad i \neq j,$$

given that the current state is left.

Constructing a semi-Markov chain



Approximation of a semi-Markov chain (Toy example)

Say we want to approximate a 2-state semi-Markov chain with

- dwell time in state 1 $\sim$ shiftedPois(λ) $\rightarrow$ p.m.f.: $d(r)$
- dwell time in state 2 $\sim$ Geom($1 - \gamma$)

and conditional t.p.m.

$$\boldsymbol{\Omega} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

To approximate the dwell-time distribution in state 1, we replace it with a **state aggregate** of size 3.

Approximation of a semi-Markov chain

We build a new Markov chain with the following block t.p.m.

$$\boldsymbol{\Gamma} = \left(\begin{array}{ccc|c} 0 & 1-c(1) & 0 & c(1) \\ 0 & 0 & 1-c(2) & c(2) \\ 0 & 0 & 1-c(3) & c(3) \\ \hline 1-\gamma & 0 & 0 & \gamma \end{array}\right)$$

where

$$c(r) = \begin{cases} \frac{d(r)}{1-F(r-1)} & \text{for } F(r-1) < 1, \\ 1 & \text{for } F(r-1) = 1 \end{cases} \qquad (= \Pr(R = r \mid R \geq r))$$

Approximation of a semi-Markov chain

We build a new Markov chain with the following block t.p.m.

$$\boldsymbol{\Gamma} = \left(\begin{array}{ccc|c} 0 & 1-c(1) & 0 & c(1) \\ 0 & 0 & 1-c(2) & c(2) \\ 0 & 0 & 1-c(3) & c(3) \\ \hline 1-\gamma & 0 & 0 & \gamma \end{array}\right)$$

where

$$c(r) = \begin{cases} \frac{d(r)}{1-F(r-1)} & \text{for } F(r-1) < 1, \\ 1 & \text{for } F(r-1) = 1 \end{cases} \qquad (= \Pr(R = r \mid R \geq r))$$

Letting d^* denote the p.m.f. of the dwell time in the state aggregate, we see that

$$d^*(1) = c(1) = d(1),$$

Approximation of a semi-Markov chain

We build a new Markov chain with the following block t.p.m.

$$\boldsymbol{\Gamma} = \left(\begin{array}{ccc|c} 0 & 1-c(1) & 0 & c(1) \\ 0 & 0 & 1-c(2) & c(2) \\ 0 & 0 & 1-c(3) & c(3) \\ \hline 1-\gamma & 0 & 0 & \gamma \end{array}\right)$$

where

$$c(r) = \begin{cases} \frac{d(r)}{1-F(r-1)} & \text{for } F(r-1) < 1, \\ 1 & \text{for } F(r-1) = 1 \end{cases} \qquad (= \Pr(R = r \mid R \geq r))$$

Letting d^* denote the p.m.f. of the dwell time in the state aggregate, we see that

$$d^*(1) = c(1) = d(1),$$

and

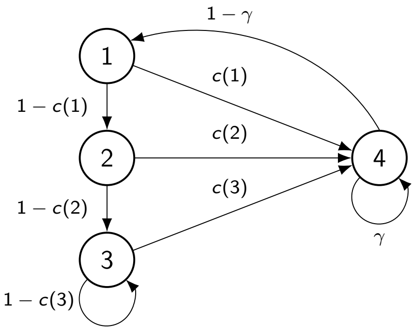
$$d^*(2) = (1 - c(1)) \cdot c(2) = (1 - d(1)) \cdot \frac{d(2)}{1 - d(1)} = d(2).$$

Approximation of a semi-Markov chain

- Paths through the state aggregate yield a **telescoping product** $\rightarrow$ Exact representation up to the chosen aggregate size.
- For dwell times larger than 3, so-called geometric tail:

$$d^*(r) = d(3) \cdot (1 - c(3))^{r-3}$$

Approximation of a semi-Markov chain



$$d^*(1) = c(1) = d(1)$$

$$d^*(2) = (1 - c(1))c(2) = d(2)$$

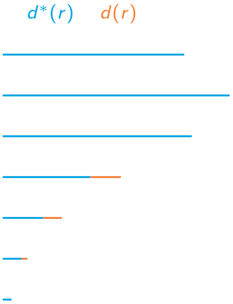
$$d^*(3) = (1 - c(1))(1 - c(2))c(3) = d(3)$$

$$d^*(4) = d(3)(1 - c(3))$$

$$d^*(5) = d(3)(1 - c(3))^2$$

$$d^*(6) = d(3)(1 - c(3))^3$$

$$\ldots$$



Approximation of a semi-Markov chain

- Can be generalised to all states of a given semi-Markov chain, by chosing p.m.f.s $d_1, d_2, \ldots, d_N$ and aggregate sizes $N_1, N_2, \ldots, N_N$.

Approximation of a semi-Markov chain

- Can be generalised to all states of a given semi-Markov chain, by chosing p.m.f.s $d_1, d_2, \ldots, d_N$ and aggregate sizes $N_1, N_2, \ldots, N_N$.
- Each state in a state aggregate is associated with the same state-dependent distribution $f_i, i = 1, \ldots, N$.

Approximation of a semi-Markov chain

- Can be generalised to all states of a given semi-Markov chain, by chosing p.m.f.s $d_1, d_2, \ldots, d_N$ and aggregate sizes $N_1, N_2, \ldots, N_N$.
- Each state in a state aggregate is associated with the same state-dependent distribution $f_i, i = 1, \ldots, N$.
- In doing so we represent an HSMM by a regular HMM with an enlarged state space and structured transition probabilities.

Approximation of a semi-Markov chain

- Can be generalised to all states of a given semi-Markov chain, by chosing p.m.f.s $d_1, d_2, \dots, d_N$ and aggregate sizes $N_1, N_2, \dots, N_N$.
- Each state in a state aggregate is associated with the same state-dependent distribution $f_i, i = 1, \dots, N$.
- In doing so we represent an HSMM by a regular HMM with an enlarged state space and structured transition probabilities.
- Parameter estimation via numerical maximum likelihood and standard inference.

From homogeneous HSMMs to inhomogeneous HSMMs

- Sometimes the assumption of homogeneous dwell-time distributions may be unrealistic.

From homogeneous HSMMs to inhomogeneous HSMMs

- Sometimes the assumption of homogeneous dwell-time distributions may be unrealistic.
- Therefore we now consider HSMMs with inhomogeneous dwell-time distributions:

$$d_i^{(t)}, \quad i = 1, \dots, N, \quad t = 1, \dots, T.$$

From homogeneous HSMMs to inhomogeneous HSMMs

- Sometimes the assumption of homogeneous dwell-time distributions may be unrealistic.
- Therefore we now consider HSMMs with inhomogeneous dwell-time distributions:

$$d_i^{(t)}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T.$$

- Typically by linking the mean to external covariates, e.g. $d_i^{(t)} = \text{shiftedPois}(\lambda_i^{(t)})$ where $\lambda_i^{(t)} = \exp(\boldsymbol{\beta}_i' \boldsymbol{z}_t)$ for some set of external covariates $\boldsymbol{z}_t$, $t = 1, \ldots, T$.

From homogeneous HSMMs to inhomogeneous HSMMs

- Sometimes the assumption of homogeneous dwell-time distributions may be unrealistic.
- Therefore we now consider HSMMs with inhomogeneous dwell-time distributions:

$$d_i^{(t)}, \quad i = 1, \dots, N, \quad t = 1, \dots, T.$$

- Typically by linking the mean to external covariates, e.g. $d_i^{(t)} = \text{shiftedPois}(\lambda_i^{(t)})$ where $\lambda_i^{(t)} = \exp(\boldsymbol{\beta}_i' \boldsymbol{z}_t)$ for some set of external covariates $\boldsymbol{z}_t$, $t = 1, \dots, T$.
- The conditional transition probabilities $\omega_{ij}^{(t)}$ may also be inhomogeneous, but we will ignore that for now (well explored).

Here comes the tricky part

- We will do the same as before (but now with general notation) **and** add the inhomogeneity.
- Now for $i = 1, \ldots, N$ consider the inhomogeneous hazard rates

$$c_i^{(t)}(r) = \begin{cases} \frac{d_i^{(t)}(r)}{1 - F_i^{(t)}(r-1)} & \text{for } F_i^{(t)}(r-1) < 1, \\ 1 & \text{for } F_i^{(t)}(r-1) = 1 \end{cases}$$

Here comes the tricky part

- We will do the same as before (but now with general notation) **and** add the inhomogeneity.
- Now for $i = 1, \dots, N$ consider the inhomogeneous hazard rates

$$c_i^{(t)}(r) = \begin{cases} \frac{d_i^{(t)}(r)}{1-F_i^{(t)}(r-1)} & \text{for } F_i^{(t)}(r-1) < 1, \\ 1 & \text{for } F_i^{(t)}(r-1) = 1 \end{cases}$$

- For $t = 1, \dots, T$ consider the structured block t.p.m.

$$\boldsymbol{\Gamma}^{(t)} = \begin{pmatrix} \boldsymbol{\Gamma}_{11}^{(t)} & \cdots & \boldsymbol{\Gamma}_{1N}^{(t)} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_{N1}^{(t)} & \cdots & \boldsymbol{\Gamma}_{NN}^{(t)} \end{pmatrix}$$

Inhomogeneous HSMMs

The diagonal block matrices ($N_i \times N_i$) are now inhomogeneous and are defined as

$$\mathbf{\Gamma}_{ii}^{(t)} = \begin{pmatrix} 0 & 1 - c_i^{(t)}(1) & 0 & \cdots & & 0 \\ 0 & 0 & 1 - c_i^{(t-1)}(2) & \cdots & & 0 \\ 0 & 0 & 0 & \ddots & & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & \cdots & & 0 & 1 - c_i^{(t-N_i+2)}(N_i - 1) \\ 0 & 0 & \cdots & & 0 & 1 - c_i^{(t-N_i+1)}(N_i) \end{pmatrix}.$$

Inhomogeneous HSMMs

The off-diagonal block matrices ($N_i \times N_j$) are defined as

$$\mathbf{\Gamma}_{ij}^{(t)} = \begin{pmatrix} \omega_{ij} c_i^{(t)}(1) & 0 & \cdots & 0 \\ \omega_{ij} c_i^{(t-1)}(2) & 0 & \cdots & 0 \\ \vdots & & & \\ \omega_{ij} c_i^{(t-N_i+1)}(N_i) & 0 & \cdots & 0 \end{pmatrix}.$$

Example

$$\Gamma_{ii}^{(t)} = \begin{pmatrix} 0 & \mathbf{1 - c_i^{(t)}(1)} & 0 & \cdots & 0 \\ 0 & 0 & 1 - c_i^{(t-1)}(2) & \cdots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ \vdots & \vdots & & & \vdots \end{pmatrix}$$

$$\Gamma_{ii}^{(t+1)} = \begin{pmatrix} 0 & 1 - c_i^{(t+1)}(1) & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{1 - c_i^{(t)}(2)} & \cdots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ \vdots & \vdots & & & \vdots \end{pmatrix}$$

Example

$$\mathbf{\Gamma}_{ij}^{(t+2)} = \begin{pmatrix} \omega_{ij} c_i^{(t+2)}(1) & 0 & \cdots & 0 \\ \omega_{ij} c_i^{(t+1)}(2) & 0 & \cdots & 0 \\ \omega_{ij} \mathbf{c_i^{(t)}(3)} & 0 & \cdots & 0 \\ \vdots & & & \end{pmatrix}$$

The ω_{ij} are summed out by total probability.

Inhomogeneous HSMMs

- Dwell time in a state aggregate is again determined by the superdiagonal of $\mathbf{\Gamma}_{ii}^{(t)}$
- Due to the inhomogeneity path runs through $\mathbf{\Gamma}^{(t)}, \mathbf{\Gamma}^{(t+1)}, \mathbf{\Gamma}^{(t+2)}, \dots$
- The implemented time-shift renders the dwell-time distriution $d_i^{(t)}$ when the transition into that state aggregate happends in t.
- We have arrived at **fully inhomogeneous HSMMs**.[2]

[2] Covariates in the dwell-time distributions, conditional transition probabilities and the state-dependent process.

Dwell-time distribution of inhomogeneous HSMMs

- I already mentioned some advantages of approximating HSMMs with HMMs
- Another very nice one: All theory established before applies very similarly when we restrict ourselves to periodic variation as a covariate.

Dwell-time distribution of inhomogeneous HSMMs

- I already mentioned some advantages of approximating HSMMs with HMMs
- Another very nice one: All theory established before applies very similarly when we restrict ourselves to periodic variation as a covariate.
- We can calculate the periodically stationary distributions,

Dwell-time distribution of inhomogeneous HSMMs

- I already mentioned some advantages of approximating HSMMs with HMMs
- Another very nice one: All theory established before applies very similarly when we restrict ourselves to periodic variation as a covariate.
- We can calculate the periodically stationary distributions,
- and the overall state dwell-time distribution...

Dwell-time distribution of inhomogeneous HSMMs

$$d_i(r) = \sum_{t=1}^{L} v_i^{(t)} d_i^{(t)}(r), \quad \text{for } r \leq N_i,$$

with the mixture weights defined as

$$v_i^{(t)} = \frac{\sum_{l \in I_k, k \neq i} \delta_l^{(t-1)} \gamma_{l I_i^-}^{(t-1)}}{\sum_{t=1}^{L} \sum_{l \in I_k, k \neq i} \delta_l^{(t-1)} \gamma_{l I_i^-}^{(t-1)}}, \quad t = 1, \ldots, L,$$

where I_i^- is the lowest state of state aggregate i and $\boldsymbol{\delta}^{(t)}$ is the stationary distribution at t.

Outline

Application: Drosophila melanogaster

Figure 1: Source: https://de.wikipedia.org/wiki/Drosophila_melanogaster.

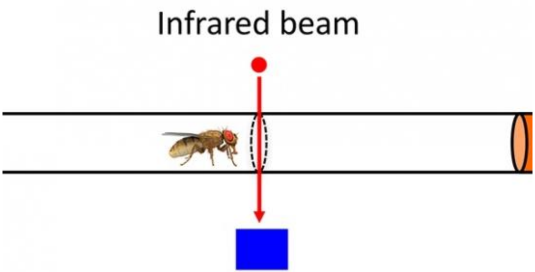


Figure 2: Source: https://www.eurekalert.org/multimedia/670792.

Application: Drosophila melanogaster

- Drosophila flies have a strong circadian rhythm.

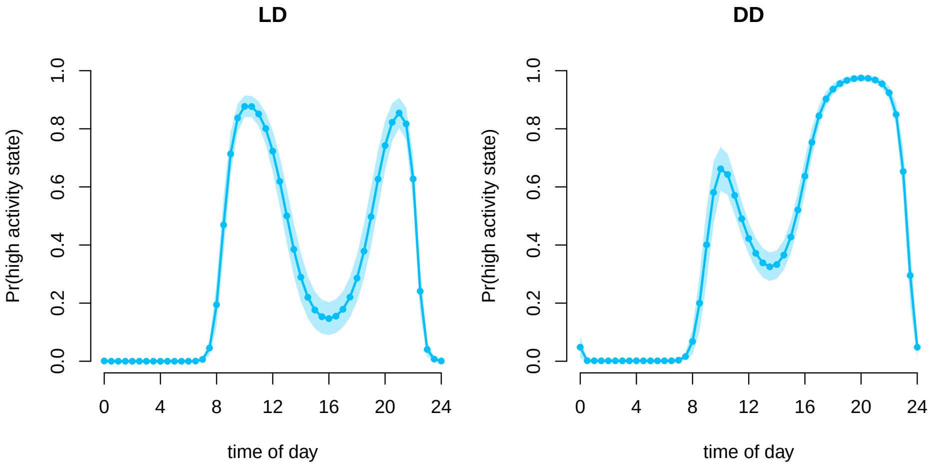
Application: Drosophila melanogaster

- Drosophila flies have a strong circadian rhythm.
- Researchers are interested in its reaction to external variation, therefore
- two conditions: Light-Dark (LD) and only Dark (DD)

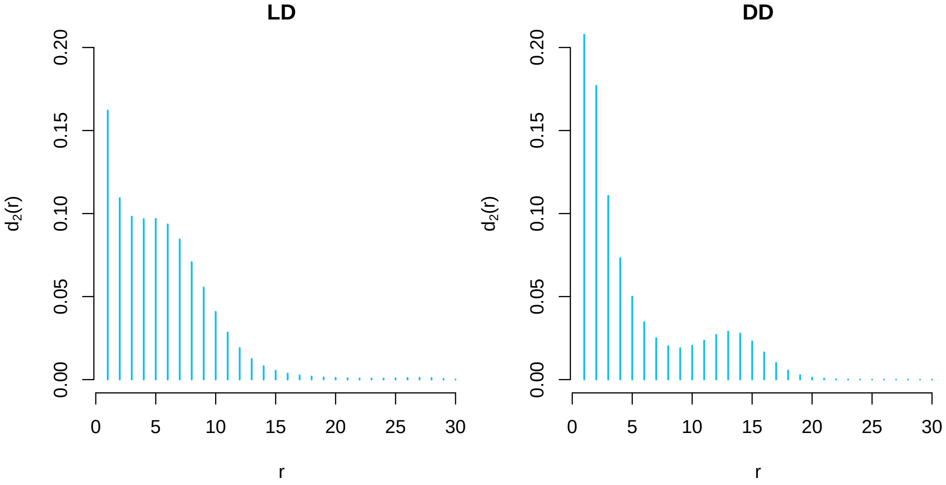
Application: Drosophila melanogaster

- Drosophila flies have a strong circadian rhythm.
- Researchers are interested in its reaction to external variation, therefore
- two conditions: Light-Dark (LD) and only Dark (DD)
- 2-state HMMs: low and high activity state

Application: Drosophila melanogaster



Application: Drosophila melanogaster



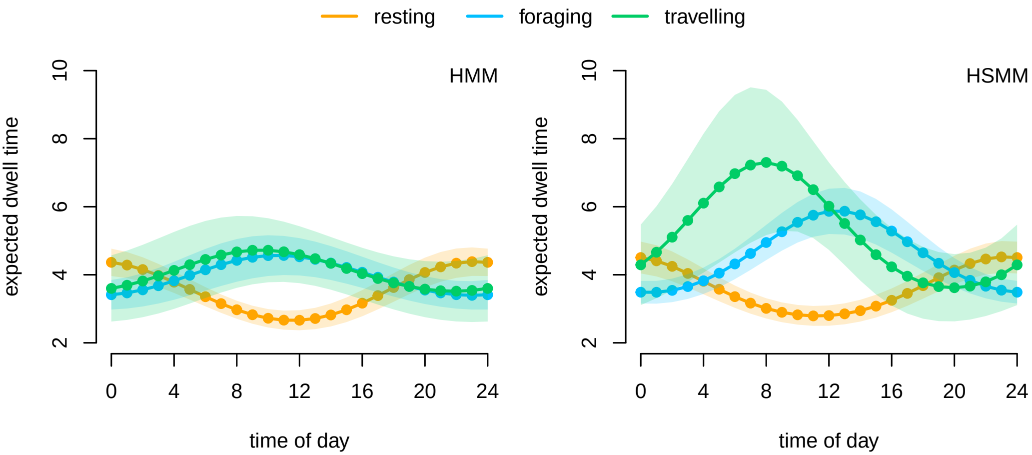
Application: Arctic muskox

Figure 3: Source: https://en.wikipedia.org/wiki/Muskox

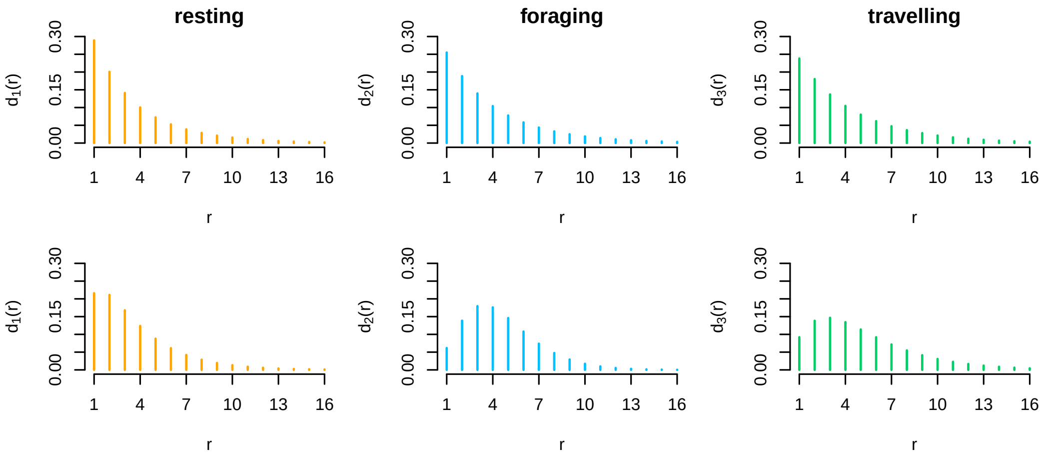
Application: Arctic muskox

- Largest arctic herbivore
- Previous analyses revealed non-geometric dwell times and temporal variation (Pohle, Langrock, et al., 2017; Beumer et al., 2020; Pohle, Adam, et al., 2022)
- 3-state HMMs and HSMMs: resting, foraging and travelling

Application: Arctic muskox



Overall dwell-time distributions



Outline

Conclusion and Outlook

- Having a closer look at widely used HMMs with periodic variation enables us to conduct new inference.

Conclusion and Outlook

- Having a closer look at widely used HMMs with periodic variation enables us to conduct new inference.
- In periodic HMMs, dwell time can be very non-geometric

Conclusion and Outlook

- Having a closer look at widely used HMMs with periodic variation enables us to conduct new inference.
- In periodic HMMs, dwell time can be very non-geometric
- Thus, criticising the Markov assumption by addressing geometric dwell-times falls short of such models' actual potential.

Conclusion and Outlook

- Having a closer look at widely used HMMs with periodic variation enables us to conduct new inference.
- In periodic HMMs, dwell time can be very non-geometric
- Thus, criticising the Markov assumption by addressing geometric dwell-times falls short of such models' actual potential.
- Some scenarios still require flexible modelling of state dwell times that are not well described by periodic variation only.

Conclusion and Outlook

- Having a closer look at widely used HMMs with periodic variation enables us to conduct new inference.
- In periodic HMMs, dwell time can be very non-geometric
- Thus, criticising the Markov assumption by addressing geometric dwell-times falls short of such models' actual potential.
- Some scenarios still require flexible modelling of state dwell times that are not well described by periodic variation only.
- The muskox example showed that in scenarios with non-geometric dwell times and additional inhomogeneity, inhomogeneous HSMMs become worthwile.

Literature

Beumer, Larissa T et al. (2020). "An application of upscaled optimal foraging theory using hidden Markov modelling: year-round behavioural variation in a large arctic herbivore". In: *Movement Ecology* 8, pp. 1–16.

Bulla, Jan and Ingo Bulla (2006). "Stylized facts of financial time series and hidden semi-Markov models". In: *Computational Statistics & Data Analysis* 51.4, pp. 2192–2209.

Langrock, Roland et al. (2014). "Modeling the diving behavior of whales: a latent-variable approach with feedback and semi-Markovian components". In: *Journal of Agricultural, Biological, and Environmental Statistics* 19, pp. 82–100.

Pohle, Jennifer, Timo Adam, et al. (2022). "Flexible estimation of the state dwell-time distribution in hidden semi-Markov models". In: *Computational Statistics & Data Analysis* 172, p. 107479.

Pohle, Jennifer, Roland Langrock, et al. (2017). "Selecting the number of states in hidden Markov models: pragmatic solutions illustrated using animal movement". In: *Journal of Agricultural, Biological and Environmental Statistics* 22, pp. 270–293.

Thank you very much for your attention!