Getting Started with Hidden Markov Models

Vianey Leos(-)Barajas[1] and **Théo** Michelot[2]

[1]**Currently in the traveling state from:**
Depts. of Forestry & Envir Res and Statistics (postdoc)
North Carolina State University (USA)
To
Dept of Statistical Sciences & School of the Environment (assistant prof)
University of Toronto (Canada)

[2]**In the resting state at:**
Centre for Research into Ecological and Environmental Modelling
University of St Andrews (Scotland)

Table of contents

Introduction

- **Claim:**
    - **False:** Hidden Markov models objectively classify behaviors.
    - **True (for the most part):** Hidden Markov models can distinguish between discernible signals and capture temporal dependence.

- **Claim:**
  - **False:** Hidden Markov models objectively classify behaviors.
  - **True (for the most part):** Hidden Markov models can distinguish between discernible signals and capture temporal dependence.
- **Reality**: No model is magic, but **every model tells a story**.

- **Claim:**
  - **False:** Hidden Markov models objectively classify behaviors.
  - **True (for the most part):** Hidden Markov models can distinguish between discernible signals and capture temporal dependence.
- **Reality**: No model is magic, but **every model tells a story**.
- **What's important:**

Where do we start?

- **Claim:**
    - **False:** Hidden Markov models objectively classify behaviors.
    - **True (for the most part):** Hidden Markov models can distinguish between discernible signals and capture temporal dependence.
- **Reality**: No model is magic, but **every model tells a story**.
- **What's important:**
    - Embracing your domain expertise.
    - You know your animals.
    - Constructing a model while thinking about what your animal actually does.

Really. Not even hidden Markov models.

Data Examples
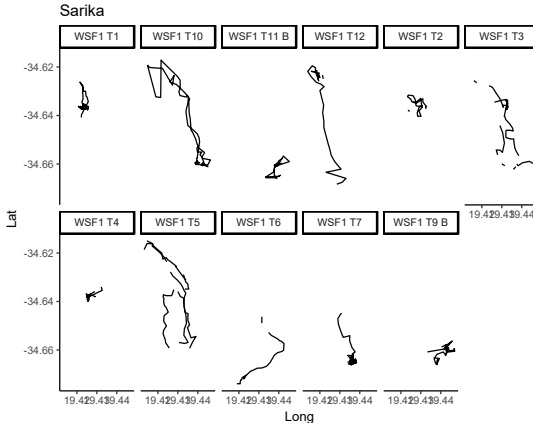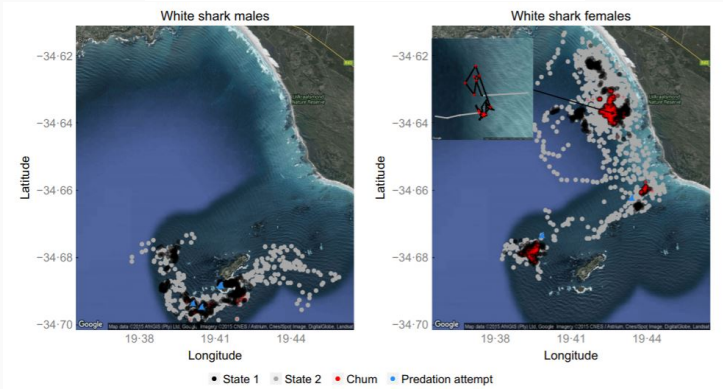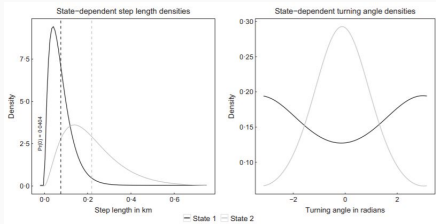
Position every 5 minutes (filled with missing data when needed: 'NA'):

Visually: Some directed, traveling behavior, other times she remained in the same area.

HMM Results:

Example 2: Oceanic Whitetip Shark Acceleration Data

Oceanic Whitetip Shark Data: log ODBA Avg 3 Sec

HMM Results:
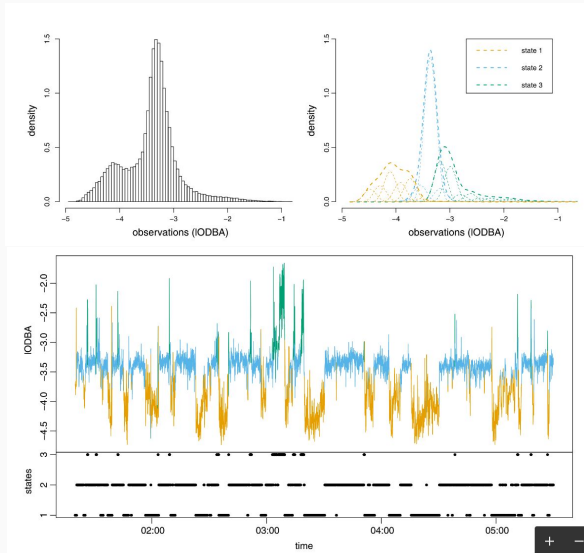
Data Processing is the Most Important Step

Biological Question of Interest $\rightarrow$ Quantification $\rightarrow$ Data Features

Requirements:

- Observations collected at regular temporal scales (can be filled with missing data: in R 'NA')

CAUTION: Do not impute if it's not completely, can not solve the world's problems otherwise – level, needed.

hidden Markov models

Processes: Observation – $\{Y_t\}_{t=1}^{T}$

Figure 1: Observations over time.

Processes: Observation – $\{Y_t\}_{t=1}^{T}$ State – $\{S_t\}_{t=1}^{T}$
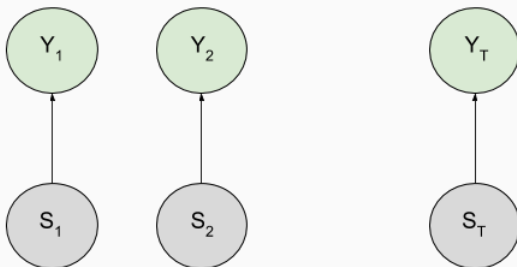
Figure 2: Observation and state (behavior) process.

Finite mixture model, assuming independence. At each point in time, S_t takes on one of N possible values.

hidden Markov models

Processes: Observation – $\{Y_t\}_{t=1}^{T}$ State – $\{S_t\}_{t=1}^{T}$
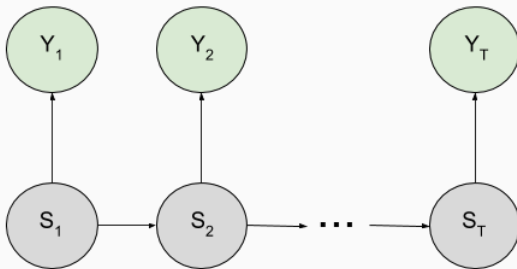
Figure 3: Graphical dependence structure implied by a hidden Markov model.

A basic finite-state, discrete-time HMM is fully specified by four components:

Mathematical Description

- Number of states, N

Biological Interpretation

- Number of movement patterns (proxies for behavior)

A basic finite-state, discrete-time HMM is fully specified by four components:

Mathematical Description

- Number of states, N
- State-dependent distributions: $\{f(\mathbf{y}_t|S_t = n)\}_{n=1}^{N}$

Biological Interpretation

- Number of movement patterns (proxies for behavior)
- Movement pattern distributions

A basic finite-state, discrete-time HMM is fully specified by four components:

Mathematical Description

- Number of states, N
- State-dependent distributions: $\{f(\mathbf{y}_t|S_t = n)\}_{n=1}^{N}$
- Transition probability matrix (t.p.m.), $\boldsymbol{\Gamma}$

Biological Interpretation

- Number of movement patterns (proxies for behavior)
- Movement pattern distributions
- How does the animal switch across behaviors over time?

A basic finite-state, discrete-time HMM is fully specified by four components:

Mathematical Description

- Number of states, N
- State-dependent distributions: $\{f(\mathbf{y}_t|S_t = n)\}_{n=1}^{N}$
- Transition probability matrix (t.p.m.), $\boldsymbol{\Gamma}$
- Initial state distribution, $\boldsymbol{\delta}$

Biological Interpretation

- Number of movement patterns (proxies for behavior)
- Movement pattern distributions
- How does the animal switch across behaviors over time?
- What is the animal doing the first time we see it?

Number of States - number of movement patterns:

- Chosen *a priori* to fit the model. Part of the exploratory search.
- May not be 1-1 to behaviors, but that's okay!

State-dependent (movement patterns) distributions:

Key idea: there are multiple signals that stem from the animal exhibiting a given behavior. We group these signals into behaviors through the use of probability distribution (or mass) functions.

For example, we could have

$f(y_t|S_t = n) \sim Normal(\mu_n, \sigma_n)$ or
$f(y_t|S_t = n) \sim Gamma(\mu_n, \sigma_n)$ or
(any other valid distribution)

A **Markov chain**[1] is a stochastic process $\{S_t, t = 1, 2, \ldots\}$, i.e. a sequence of random variables $S_1, S_2, \ldots$, such that:

- $S_t \in \{1, \ldots, N\}$ for all t (i.e. there are N so-called "states")
- the **Markov property** holds:

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

The **transition probability matrix** (t.p.m.) is given by

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1N} \\ \vdots & \ddots & \vdots \\ \gamma_{N1} & \cdots & \gamma_{NN} \end{pmatrix}$$

There are two important restrictions on the parameters:

- $\gamma_{ij} \in [0, 1]$ for all i, j
- $\sum_{j=1}^{N} \gamma_{ij} = 1$ for all i

[1]more precisely: a discrete-time, finite-state Markov chain

How long do we spend in a state?

Let D_n be a random variable[2] connected with the amount of time that we spend in a state *before switching*. Then, the implied distribution for D_n when assuming the Markov property is given as,

$$Pr(D_n = d_n) \sim Geom(1 - \gamma_{nn})$$

such that $E(D_n) = 1/(1 - \gamma_{nn})$.

---

[2]What's the story here? Do the calculations make sense? Work out together.

Suppose we are dealing with a homogeneous and "well-behaved"[3] Markov chain.

What if we observe this Markov chain at a random point in time, after it has already been running for some time? What state will we find it in?

$\rightarrow$ the probabilities of encountering the process in any of the N states is then given by the **stationary distribution** — this is the vector $\boldsymbol{\delta} \in \mathbb{R}^N$ such that

$$\boldsymbol{\delta}\boldsymbol{\Gamma} = \boldsymbol{\delta} \text{ subject to } \sum_{i=1}^{N} \delta_i = 1$$

---

[3]it needs to be irreducible and aperiodic — properties that are pretty much always met in practice!

HMMs

- an *N*-state HMM is a (doubly) stochastic process in discrete time, with
  - an unobservable **state process** $S_1, S_2, \ldots, S_T$ taking values in $\{1, \ldots, N\}$,
  - an observed **state-dependent process** $Y_1, Y_2, \ldots, Y_T$,
- such that
  - $f(y_t \mid s_1, \ldots, s_t, y_1, \ldots, y_{t-1}) = f(y_t \mid s_t)$
    (conditional independence assumption)
  - $f(s_t \mid s_1, \ldots, s_{t-1}) = f(s_t \mid s_{t-1})$
    (Markov property)

The HMM Story So far…

- We begin at time $t = 1$, the first observation of the animal. The initial distribution assigns probabilities to the animal exhibiting one of the N states (movement patterns/behaviors) during this period.
- Given whatever it's doing at S_1, it produces an observation y_1 from the distribution $f(y_1|S_1 = s_1)$.
- What the animal does during S_2 depends on what it was doing at S_1.
- Given whatever it's doing at S_2, it produces an observation y_2 from the distribution $f(y_2|S_2 = s_2)$.
- What the animal does during S_3 depends on what it was doing at S_2.
- (and so on)

Another slightly different version...

- We begin at time $t = 1$, the first observation of the animal. The initial distribution assigns probabilities to the animal exhibiting one of the N states (movement patterns/behaviors) during this period.

- Given whatever it's doing at S_1, it produces an observation y_1 from the distribution $f(y_1|S_1 = s_1)$.

- We expect that the animal will remain in the same state for a given amount of time, as given by the state duration distribution. Once it 'finishes' its time, it switches to another state with the appropriate probabilities

- If we remain in state n for d_n time points, we draw d_n observations from $f(y_{t-d_n+1:t}|S_{t-d_n+1:t} = n)$.
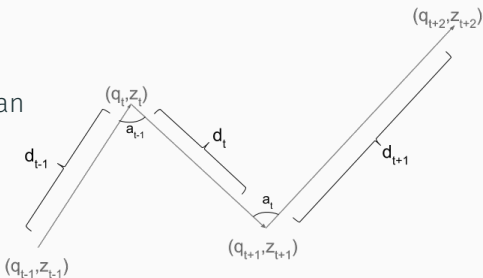
Coordinates: $\{q_t, z_t\}_{t=1}^{T+1}$

Transform positions into:

- **step lengths** $\{d_t\}_{t=1}^{T}$ (euclidean distance between two points)
- **turning angles** $\{a_t\}_{t=1}^{T}$ (using three consecutive positions)

Let $y_t = \{d_t, a_t\}_{t=1}^{T}$

Key Idea:
Observed movements are a result of the underlying behavior that the animal is exhibiting.

Quick Check: What are the four components of an HMM and how do we construct an HMM for positional data?

MULTINOMIAL LOGISTIC
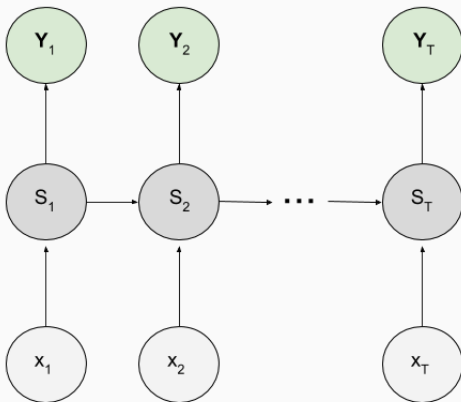REGRESSION:

$$\gamma_{ij}^t(x_t) = \frac{\exp(\rho_{ij}^t)}{\sum_{j=1}^{N} \exp(\rho_{ij}^t)}$$

$$\rho_{ij}^t = \begin{cases} \tau_0^{(ij)} + \tau_1^{(ij)} x_t & \text{if } i \neq j; \\ 0 & \text{o.w.} \end{cases}$$

for $i, j \in \{1, \ldots, N\}$

State decoding in HMMs

Given a fitted model, it is often of interest to **decode the hidden states** underlying the observed time series.

Forward-Backward Algorithm (Local Decoding):

- quantity obtained $\Pr(S_t = i | y_1, \ldots, y_T)$
- probability of the states at each point in time, given the observations

(looks at each time point in isolation)

Forward-Filtering Backward-Sampling (Global Decoding):

- consider $\Pr(S_1 = i_1, \ldots, S_T = i_T | y_1, \ldots, y_T)$
- draws from the joint posterior distribution of the states (sample paths!! woohoo!!)

(looks at the sequence as a whole)

RMarkdown file.

What else can you do?

HMMs are often used to model **multivariate time series**,
$\mathbf{Y}_t = (Y_{t1}, \ldots, Y_{tK})$

$$\begin{pmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1K} \end{pmatrix}, \begin{pmatrix} Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2K} \end{pmatrix}, \ldots, \begin{pmatrix} Y_{T1} \\ Y_{T2} \\ \vdots \\ Y_{TK} \end{pmatrix}.$$

In many scenarios it can be assumed that the *K* variables are **driven by the same underlying state process** S_t

EMBRACE THE RANDOM EFFECTS!!

- Random effects allow us to capture individual-level variation (in particular if we do not have the covariates that explain it).

- Understanding what the animals generally tend to do, and what each animal actually did is part of biological/ecological understanding, and a key part of the story! Make the most of your data (the effort, sweat and tears that go into collection), and ask your models to do the same.

Overview

Why did we want to fit an (unsupervised) HMM?

There are two main features of the data that an (unsupervised) HMM captures:

- the marginal distribution of the observation at time t
- the autocorrelation structure

Using an unsupervised HMM for animal movement, we NEVER validate our model by how well it captures 'actual behavior'. Let's resist the narrative that HMMs are tools for automatic identification of animal behavior.

Oftentimes, with the right data, the right domain expertise, we can capture simple, biologically relevant patterns in the movement data that can serve as good proxies for some very general behaviors of interest (active/not active – area restricted search/directed traveling).

Main options to check if a fitted HMM is adequate:

1. graphical comparison of marginal distribution under fitted HMM and empirical distribution, to check adequacy of state-dep. distributions

2. simulate data from the fitted model, then compare the patterns found in the simulated data with those of the real data (patterns to look for: marginal distribution, autocorrelation, etc.)

3. a residual analysis

Learning through our failures

Model building for HMMs can be hard. Fitting an HMM can be hard, especially when misspecified (this will always happen!). (Advice: Work toward a model with a single mode in the likelihood/posterior.)

BUT these 'failures' are supremely enlightening! When the likelihood/posterior is multimodal – what do the different modes biologically imply? Does it make sense? How can you then extend the model? Are some animals behaving very differently from each other? (there's an always an oddball in the bunch)

Take home message: We'll keep making progress if we do more to embrace HMMs as a generative model! Random effects – to learn about individuals. Hierarchical structures in the state process to capture broad patterns.

Go forth and fit some HMMs!

Some links:

- Statistical Sciences

On the growing interdisciplinary nature of the Statistical Sciences department by Dept. Chair, Radu Craiu: The Hiring Gambit: In Search of the Twofer Data Scientist

- School of the Environment