

Hidden Markov Models

Eric Ward

Due Tues May 9 11:59 PM PDT

Contents

Part 1. Hidden Markov Models	1
Part 2. Fitting Multivariate HMMs	2

Part 1. Hidden Markov Models

Identifying regimes using Hidden Markov Models (HMMs)

For the first part of the homework, we'll use data from the Pacific Decadal Oscillation (PDO) to ask questions about identifying regimes. This dataset can be accessed via the `rsoi` package. First, let's grab the data. Run the `install.packages()` code if you need the `rsoi` package.

```
library(dplyr)
#install.packages("rsoi")
pdo <- rsoi::download_pdo()
```

We will look at the winter PDO only. We need to shift the year for Oct-Dec by 1 since Oct-Feb spans 2 calendar years.

```
pdo$Year[which(pdo$Month%in%c("Oct", "Nov", "Dec"))] <- pdo$Year[which(pdo$Month%in%c("Oct", "Nov", "Dec"))]
pdo <- dplyr::group_by(pdo, Year) %>%
  dplyr::summarize(winter_pdo = mean(PDO[which(Month %in% c("Oct", "Nov", "Dec", "Jan", "Feb"))])) %>%
  dplyr::select(winter_pdo, Year)
# The first year will be missing Oct-Dec
pdo <- pdo[-1,]
```

Use `pdo` for your analyses. You will be modeling `winter_pdo`. Use the `hmmTMB` or `depmixS4` packages discussed in the HMM lecture.

1. Fit a 2-state HMM to the annual indices of winter PDO. Assume Gaussian errors (default). See the lecture on HMMs and/or section 3 in the `depmixS4` vignette.
2. Try fitting the model 10-20 times. Does the likelihood seem reasonably stable? (Note `logLik()` gets you the log-likelihood from model fits in R).
3. What is the transition matrix for the best model? What are the persistence probabilities (e.g. probabilities of staying in the same state from step t to $t + 1$)?
4. Plot the predicted values versus year. See slide 50 of the HMM lecture for an example.
5. Plot the posterior probability of being in the various states from your best model (e.g. probability of being in state 1 over time)
6. What is the long-run probability that the PDO is in state 1 versus state 2? You can calculate this from the transition matrix. There is an analytical solution for this (a bit of googling will find it). Or you can

run a `for` loop to find it. Let p_1 be the probability that PDO is in state 1 and p_2 be the probability that PDO is in state 2. Note $p_1 + p_2 = 1$. If P is the transition matrix (in Q3),

$$\begin{bmatrix} p_1 & p_2 \end{bmatrix}_n = \begin{bmatrix} p_1 & p_2 \end{bmatrix}_{n-1} P$$

Note this is a 1x2 matrix times a 2x2 matrix on the right. Start with $p_1 = 1$ and $p_2 = 0$, say. Run a `for` loop until

$$\begin{bmatrix} p_1 & p_2 \end{bmatrix}_n \approx \begin{bmatrix} p_1 & p_2 \end{bmatrix}_{n-1}$$

That $\begin{bmatrix} p_1 & p_2 \end{bmatrix}_n$ is the long-run probability in each state.

Some ideas for optional extra analyses

- Using slide 15 of the HMM lecture as a template and the matrix in Q3, write out your fitted HMM model as equations
- Change the model to a 3-state model. Using AIC as a model selection metric, does the 3-state model perform better (lower AIC) compared to the 2-state model? What about a 1-state model?
- If you include time varying parameters (e.g. year) in the means of each state, or state transition probabilities, does the model do any better?
- Run diagnostics on the best model. Any problems?
- Compare the transition matrices for fits with different random starting conditions. Are the transition matrices stable?

Part 2. Fitting Multivariate HMMs

As part of the California Current Integrated Ecosystem Report, NOAA scientists do annual updates of stoplight charts for ecosystem indicators.

We have included the `stoplight.csv` dataset for this week. One of the columns divides indicators into groups (e.g. Local Physical, Local Biological, etc). Please pick a type of indicators, and develop a 2- or 3-state multivariate HMM. A few tips:

- Assume all responses are Gaussian.
- You're welcome to include covariates (year? Climate variables?) – but fitting a simple model without covariates is also totally fine

Summarize the model you've created. Specifically,

- Does it converge?
- How many states seem to be most supported?
- What are the transition probabilities?
- Anything else interesting that you've discovered?