Feldmann, Mews June 16, 2023

Hidden Markov Models - Practical Session 7

Exercise 1: Model selection and checking with moveHMM

We consider again the time series of the female elephant Habiba tracked via GPS in Kenya in 2014 (as we did in Practical Session 5). The aim is to find the "right" HMM for our data at hand. This means that first, we need to formulate several candidate models, from which we then choose the "best" one using information criteria, but also taking the model's interpretation into account (model selection). In a second step, we check whether our selected model is also a good model or not in terms of model fit (model checking). This session, we will use the R package `moveHMM` for all these tasks. Install the package:

```
install.packages("moveHMM")
library(moveHMM)
```

a) Read the GPS data

```
rawdata <- read.table("elephant_rawdata.txt", header = T)
```

and use `prepData()` to let `moveHMM` calculate the step lengths for you. Plot the data, e.g. by using the plotting function from `moveHMM`.

Hint: You could have a look at `?prepData` or the e.g. this pdf: `https://cran.r-project.org/web/packages/moveHMM/vignettes/moveHMM-guide.pdf`

b) Fit HMMs with $N = 2, 3, 4, 5$ to the step lengths using `fitHMM()`. Please use the argument `angleDist = "none"` as we do not model the turning angles yet, only the step lengths.

Hint: You still need to choose starting values for the state-dependent distribution, but you don't need to transform them according to some parameter constraints.

c) Fit each candidate model with 10 sets of random starting values to reduce the risk of finding a local mininimum.

d) Using `plot(mod)`, the package will provide you with plots visualizing the results of the HMM. Compare the plotted densities of each candidate model and find reasons for and against some of the modesl. Calculate AIC and BIC values for each candidate model and find out which models are preferred by which information criterion.

e) Decide on one or two models that you continue working with and perform a residual analysis using `plotPR()` or `pseudoRes()`.