Hidden Markov Models - Practical Session 9

Exercise 1: Multivariate HMM for elephant data

We already fitted different univariate HMMs to the step length data of the elephant Habiba. Now, we want to fit an HMM to the bivariate time series containing the step lengths and the turning angles.

a) Read in the elephant dataset:

```
data <- read.table("elephant_full.txt", header=T)
```

Exclude the first and the last observation of the time series from the dataset as these values are missing for the turning angle. Then get an overview of the step length and the turning angle data.

b) Here, we have the log-likelihood function for an N-state univariate HMM as on slide 140:

```
mllk <- function(theta.star, x, N) {
  mu <- exp(theta.star[1:N])
  sigma <- exp(theta.star[N + 1:N])
  Gamma <- diag(N)
  Gamma[!Gamma] <- exp(theta.star[(2 * N + 1):(length(theta.star))])
  Gamma <- Gamma / rowSums(Gamma)
  delta <- solve(t(diag(N) - Gamma + 1), rep(1, N))
  allprobs <- matrix(1, length(x), N)
  ind <- which(!is.na(x))
  for (j in 1:N) {
    allprobs[ind, j] <- dgamma(x[ind], shape = mu[j]^2 / sigma[j]^2,
    scale = sigma[j]^2 / mu[j])
  }
  foo <- delta %*% diag(allprobs[1, ])
  l <- log(sum(foo))
  phi <- foo / sum(foo)
  for (t in 2:length(x)) {
    foo <- phi %*% Gamma %*% diag(allprobs[t, ])
    l <- l + log(sum(foo))
    phi <- foo / sum(foo)
```

```
    }
    return(-l)
  }
  }
```

Modify the R code to fit an N-state bivariate HMM with contemporaneous conditional independence assumption to the time series of step lengths and turning angles. We assume a gamma distribution for the step length (as before) and a von Mises distribution for the turning angle with its mean equal to 0 for all states, i.e.

$$X_{t1}|S_t = j \sim \Gamma(\mu_j^{\text{step}}, \sigma_j) \quad \text{and} \quad X_{t2}|S_t = j \sim \text{von Mises}(\mu_j^{\text{turn}} = 0, \kappa_j).$$

Note that the parameter κ of the von Mises distribution needs to be positive. To calculate the density of the von Mises distribution, you can use the function `dvm` from the package `CircStats`. On lecture slide 228, you find an example of how the likelihood is calculated for several variables with contemporaneous conditional independence assumption.

Hint: You need to include initial parameters for κ into your parameter vector of interest θ^*. In particular, this means that you need to include the following line of code into your likelihood function:

```
kappa <- exp(theta.star[(N-1)*N+2*N+1:N])
```

As the parameter for μ^{turn} is assumed to be zero in this exercise, you don't need to include it in your parameter vector θ. In particular, this means that you need to include the following line of code into your likelihood function:

```
mu.turn <- rep(0, N)
```

As the dataset does not contain any missing values (anymore), you do not have to account for missing data in your likelihood function.

c) Fit a 3-state HMM to the step length and turning angle data of our elephant using your function from b). Then transform the estimates back to their natural parameters.

d) Plot the state-dependent distributions together with a histogram for the time series of step lengths and turning angles.

Exercise 2: Multivariate HMMs for share return data

We consider a bivariate time series of daily returns from Morgan Stanley and the Citigroup in the period from January 1997 to March 2010. The dataset called "share_returns.txt" and an R script called "mllk_shareReturns.R" can be found in the Lernraum. The R script contains two functions to calculate minus the log-likelihood of a multivariate HMM: one with contemporaneous conditional independence assumption and one with longitudinal conditional independence assumption.

a) Read the data into R

```
data <- read.table("share returns.txt", header=T)
```

and plot the two time series and get an overview of the data.

b) Our aim is to model the share returns using multivariate HMMs. Which different dependence assumptions can be made for the bivariate time series? Which likelihood codeWhich of them do you find more suitable for the present data?

c) Use the code from the R script to fit a 2-state HMM with contemporaneous conditional independence assumption to the share returns data. Then transform the estimates back to their natural form.

d) Now use the code to fit a 2-state HMM with longitudinal conditional independence assumption to the data. Again transform the estimates back to their natural form and compare them to the ones from c).

e) Calculate the AIC and BIC values for both models and compare.