

Hidden Markov Models - Practical Session 4

Exercise 1 - Likelihood evaluation, Poisson HMM

The following function can be used to calculate the log-likelihood of a vector of observations \mathbf{x} under a 2-state Poisson HMM with parameter vector $\boldsymbol{\theta} = (\lambda_1, \lambda_2)$ when $\boldsymbol{\Gamma}$ and $\boldsymbol{\delta}_1$ are known.

```
llk <- function(theta, x){
  Gamma <- matrix(c(0.9, 0.1, 0.2, 0.8), byrow = TRUE, nrow = 2)
  delta <- c(1, 0)
  lambda <- theta[1:2]
  allprobs <- cbind(dpois(x, lambda[1]), dpois(x, lambda[2]))
  foo <- delta %*% diag(allprobs[1,])
  for (t in 2:length(x)){
    foo <- foo %*% Gamma %*% diag(allprobs[t,])
  }
  return(log(sum(foo)))
}
```

- Load “simulatedPoisson.RData” from the *Lernraum* and plot the observation vector \mathbf{x} . We assume the data to be generated by a 2-state Poisson HMM.
- Upon visual inspection, guess the two parameters λ_1 and λ_2 . Calculate the associated log-likelihood value using the function above and try to roughly identify optimal values for λ_1 and λ_2 by trying out different values which might lead to a higher log-likelihood value. Save a plausible set of parameter values in the vector **theta**.
- Now use the following code to calculate the log-likelihood on a grid of possible values of λ_1 and λ_2 , subsequently plotting the log-likelihood as a function of these two parameters:

```
lambda1 <- seq(0, 2, length = 50)
lambda2 <- seq(3, 8, length = 50)
z <- matrix(NA, 50, 50)
for (i in 1:50) {
  for (j in 1:50) {
    z[i, j] <- llk(c(lambda1[i], lambda2[j]), x)
  }
}
contour(lambda1, lambda2, z, nlevels = 200)
```

```
abline(v = theta[1], lty = 2)
abline(h = theta[2], lty = 2)
```

Check if the plot confirms the parameter values you found in b). Use the identified parameter values for λ_1 and λ_2 in d) to f).

- d) What is the probability that the last observation x_{30} is generated by state 1 or state 2, given all the previous observations?

Hint: Remember from the lecture that the forward variables at time t , $\alpha_t(j)$, not only contain information on the likelihood of the observations up to time t , but also on the probabilities of being in the different states at time t (cf. lecture slide 95). In particular, we have

$$f(s_t = j | x_1, \dots, x_t) = \frac{f(x_1, \dots, x_t, s_t = j)}{f(x_1, \dots, x_t)} = \frac{\alpha_t(j)}{\sum_{j=1}^N \alpha_t(j)}$$

- e) Looking at the vector of observations \mathbf{x} , do you think that the 15th observation, $x_{15} = 3$, has been generated by state 1 or state 2? Why? Check which state has a higher probability of having generated x_{15} under the assumed model by calculating the probabilities of being in state 1 or state 2 at x_{15} , given x_1, \dots, x_{15} .
- f) Now calculate for each time point t the probabilities of x_t being in either state 1 or state 2, given all the previous observations, and store them in a matrix. Then plot these probabilities. Finally, plot the observations \mathbf{x} using colour code to indicate the most likely states underlying the observations, given all previous observations.

Bonus exercise - Likelihood evaluation, Gaussian HMM

- a) As an exercise to get used to the likelihood code, change the likelihood function given in lecture slide 101 to calculate the log-likelihood of a 2-state-HMM with normal distributions instead of Poisson distributions.
- b) Use the `simNormMC()` function from the last session to simulate a 2-state HMM with parameters of your choice.
- c) With the simulated data and the extended likelihood function, create a contour plot as in 1c) for the means of the two normal distributions given the true values for $\mathbf{\Gamma}, \delta_1, \sigma_1, \sigma_2$.