Written exercise 2

Assignment 2 – 02433 Hidden Markov Models – Anders Hørsted (s082382)

In this exercise two datasets containing noisy observations of log-population sizes are analyzed. The datasets are fitted to a theta logistic model for population growth and various types of decoding are performed. The theta logistic model can be written as a state-space model

$$P_t = P_{t-1} + r_0 \left(1 - \left[\frac{\exp\left(P_{t-1}\right)}{K}\right]^{\theta}\right) + e_t \tag{1}$$

$$X_t = P_t + u_t \tag{2}$$

where P_t is the log-population size and $e_t \sim N(0, Q)$, $u_t \sim N(0, R)$ are iid. and mutually independent. K, r_0, θ are scalar parameters that combined with the two variances Q and R gives a total of five parameters $\boldsymbol{\lambda} = (\theta, r_0, K, Q, R)$.

a) Plotting the data

To start the analysis the two datasets are plotted and shown in figure 1 **Comment on plots**

b) Approximating the SSM by a HMM

By assuming that the state-space is restricted it can be discretizised and by this discretization the state-space model can be expressed as an Hidden markov model instead. For the state-space model given in equation 1 and 2 it is now assumed that $P_t \in [2.1, 8.4]$ and that the state-space is partitioned into $m = 250$ intervals $\Omega_i = (b_{i-1}, b_i)$ where $i = 1, \ldots, m$. The width of each interval is then given by

$$w = \frac{8.4 - 2.1}{250} = 0.0252$$

and the boundaries can then be expressed as $b_i = 2.1 + wi$. If $P_t \in \Omega_i$ it is said to be in state i which is also written $C_t = i$. Each discrete state i is represented by the midpoint p_i of the interval (b_{i-1}, b_i) given by $p_i = 2.1 + w(i - 0.5)$. This links the discrete, integer valued state-space C_t to the original continous state-space P_t.

To express the state-space model as a Hidden markov model the state dependent distibutions should be found. The distribution of $X_t \,|\, C_t = i$ is found by replacing P_t

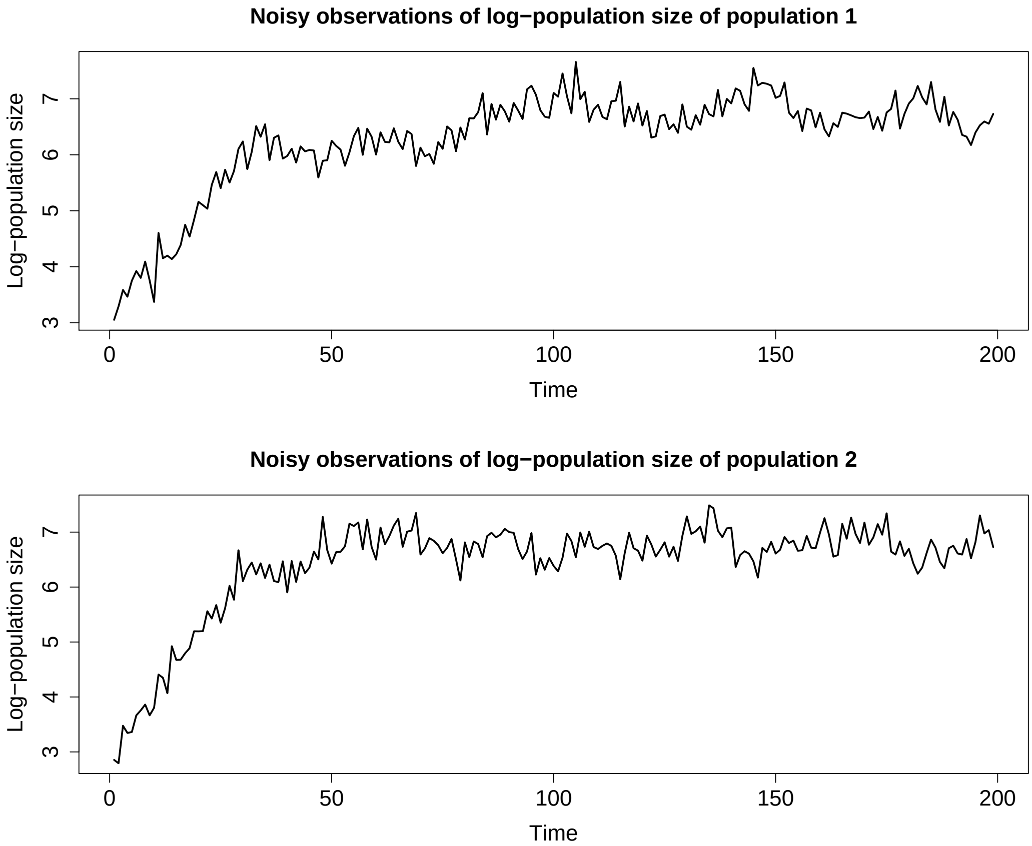

Figure 1: Plot of the two datasets

with the representation point p_i in equation (2). Combined with $u_t \sim N(0, R)$ gives that

$$X_t \,|\, C_t = i \;\sim\; N(p_i, R)$$

or equivalently

$$p(x_t \,|\, C_t = i) = \frac{1}{\sqrt{2\pi R}} \exp\left(-\frac{(x_t - p_i)^2}{2R}\right)$$

Apart from the state dependent distributions the transition probabilities $\Pr(P_t \in \Omega_j \,|\, P_{t-1} \in \Omega_i)$ should also be found. Conditioning on P_{t-1} being in state i ($C_{t-1} = i$) can be represented by replacing P_{t-1} with p_i in equation 1. Using the fact that $e_t \sim N(0, Q)$ then gives

$$P_t \,|\, C_{t-1} = i \;\sim\; N\left(\mu_i, Q\right) \quad \text{with} \quad \mu_i = p_i + r_0 \left(1 - \left[\frac{\exp(p_i)}{K}\right]^\theta\right)$$

Letting $n(\bullet, \mu, \sigma^2)$ denote the Gaussian pdf with mean μ and variance σ^2 the transition probabilities can then be calculated by

$$\Pr(P_t \in \Omega_j \,|\, C_{t-1} = i) = \Pr(C_t = j \,|\, C_{t-1} = i) = \int_{\Omega_j} n(p_t, \mu_i, Q)\, dp_t$$

This integral could be calculated by using the cumulative distribution function for the normal distribution, but as mentioned on page 13 in [1] this is more expensive than approximating the integral using the trapezoidal rule for integration. Using the trapezoidal rule gives

$$\Pr(C_t = j \,|\, C_{t-1} = i) \approx \frac{w}{2}\left(n(b_{j-1}, \mu_i, Q) + n(b_j, \mu_i, Q)\right)$$

c) Computing the likelihood

A Appendices

All R code created for this assignment is included here. All source code incl. latex code for this report can be found at `https://github.com/alphabits/dtu-spring-2012/tree/master/02433/assignment-2`

A.1 Fitting 2-state Poisson-HMM by direct maximization of MLE

References

[1] Martin Wæver Pedersen. HMM analysis of general state-space models. `http://www2.imm.dtu.dk/~mwp/02433/week7_notes.pdf`, June 2011.

[2] Walter Zucchini and Iain L. MacDonald. *Hidden Markov Models for Time Series*. Chapman & Hall/CRC, 1st edition, 2009.