Stochastic Filtering (SS2016) Exercise Sheet 4

Lecture and Exercises: JProf. Dr. Philipp Harms
Due date: May 25, 2016

Note: Please send me your source code by email or bring a laptop.

4.1. Kalman filter

We consider the same linear Gaussian HMM as in Exercise 3.4: the state and observation processes are assumed to evolve according to

$$X_{k+1} = AX_k + R\xi_k, \qquad Y_k = BX_k + S\eta_k, \qquad k \geq 0$$

with initial state $X_0 \sim \mathcal{N}(m_0, \Sigma_0)$ and parameters $A, R \in \mathbb{R}^{p\times p}$, $B \in \mathbb{R}^{q\times p}$, and $S \in \mathbb{R}^{q\times q}$ with S invertible. The variables $\xi_i \sim \mathcal{N}(0, \mathbb{1}_{p\times p})$, $\eta_j \sim \mathcal{N}(0, \mathbb{1}_{q\times q})$, and X_0 are independent. Recall from Exercise 3.4 that $\pi_k \sim \mathcal{N}(\hat{X}_k, \hat{\Sigma}_k)$.

a) Smoothing step: assume for some $k \leq n$ that $\pi_{k|n} \sim \mathcal{N}(\hat{X}_{k|n}, \hat{\Sigma}_{k|n})$. Show that $\pi_{k-1|n} \sim \mathcal{N}(\hat{X}_{k-1|n}, \hat{\Sigma}_{k-1|n})$, where

$$\begin{aligned}
\hat{X}_{k-1|n} &= \hat{X}_{k-1} + \hat{\Sigma}_{k-1} A M_{k-1}(\hat{X}_{k|n} - A\hat{X}_{k-1}),\\
\hat{\Sigma}_{k-1|n} &= \hat{\Sigma}_{k-1} - \hat{\Sigma}_{k-1} A^\top M_{k-1} A \hat{\Sigma}_{k-1} + \hat{\Sigma}_{k-1} A^\top M_{k-1} \hat{\Sigma}_{k|n} M_{k-1} A \hat{\Sigma}_{k-1},\\
M_{k-1} &= (A\hat{\Sigma}_{k-1}A^\top + RR^\top)^{-1}.
\end{aligned}$$

Hint. You may look this up in the reference of your choice or [1].

b) Explain how Exercises 3.4 and 4.1 can be combined to calculate smoothing distributions $\pi_{k|n}$, $k \leq n$, for fixed n.

Remark. This is known under the names forward-backward algorithm, Baum-Welch algorithm and, in the special case of linear Gaussian models, Rauch-Tung-Striebel smoothing [4].

4.2. Filtering the position of a moving particle

Consider a particle moving in $\mathbb{R}^2$ at constant velocity, subject to small random perturbations in position and velocity. The goal is to estimate the true position of the particle from noisy observations of its position. This estimation can be done in the framework of Exercises 3.4 and 4.1. Complete the Matlab template `kalman.m` on the homepage to perform the following tasks.

a) Generate a sample trajectory of $(X_k, Y_k)_{k=0,\dots,N}$ for a fixed number N.

b) Calculate the filtering distributions π_k and smoothing distributions $\pi_{k|N}$ for $k \le N$.

c) Calculate the prediction distributions $\pi_{k|n}$, $n \le k \le N$, for some fixed $n < N$. Compare this to the distributions $\pi_{k|N}$ to see the effect of dropping all observations between n and N.

4.3. HMMs with finite state space

When the state spaces $\mathbb{X}$ and $\mathbb{Y}$ of the signal and observation processes are finite, the filtering, smoothing, and prediction recursions simplify significantly.

Without loss of generality, we take $\mathbb{X} = \{1,\dots,d\}$ for some $d \in \mathbb{N}$. Any function $f : \mathbb{X} \to \mathbb{R}$ can be identified with a vector $\mathbf{f} = (f(1),\dots,f(d))^\top \in \mathbb{R}^d$. Analogously, any measure μ on $\mathbb{X}$ can be identified with $\boldsymbol{\mu} = (\mu(\{1\}),\dots,\mu(\{d\}))^\top \in \mathbb{R}^d$. Similarly, transition kernels can be represented by matrices. For example, the transition kernel P of the state process $(X_k)_{k\ge 0}$ can be represented by the matrix $\mathbf{P}$ with entries $\mathbf{P}_{ij} = P(i,\{j\})$.

a) Verify the identity $Pf(i) = (\mathbf{Pf})_i$, where P is a transition kernel with matrix representation $\mathbf{P}$, f is a function on $\mathbb{X}$, and $\mathbf{Pf}$ denotes matrix-vector multiplication.

b) Use these concepts to derive a matrix representation for the filtering recursion that we derived in the lecture (see also Exercise 3.1).

   Hint. You may look this up in the reference of your choice or [1, 2].

4.4. Fluorescence resonance energy transfer (FRET)

FRET is used to estimate the distance between light-sensitive molecules by measuring energy transfer between them. As an example,[1] consider two strands of DNA dyed in different fluorescent colors: one red and one green, say. The distance between the strands increases when a binding protein wedges itself between the strands. When the red dye is excited with red laser light, some energy is transferred to the neighboring molecule, which emits green light. Counting the number of emitted green photons allows one to estimate the number of binding proteins.

This estimation can be done in the framework of Exercise 4.3 with X_k representing the number of binding proteins and Y_k the number of measured protons at time k. Assume that $\mathbb{X} = \{0,1,2,3\}$, $\mathbb{Y} = \mathbb{Z}_+$, X is a Markov chain with transition matrix $\mathbf{P}$ and initial distribution $\boldsymbol{\mu}_0$ given by

$$\mathbf{P} = \begin{pmatrix} .94 & .05 & .01 & .00 \\ .03 & .94 & .02 & .01 \\ .05 & .14 & .80 & .01 \\ .05 & .15 & .30 & .50 \end{pmatrix}, \qquad \boldsymbol{\mu}_0 = \begin{pmatrix} .25 \\ .25 \\ .25 \\ .25 \end{pmatrix},$$

and Y_k is Poisson distributed with rate $\lambda_k = 50 - 10X_k$. Complete the Matlab template `fret.m` on the homepage to perform the following tasks.

a) Implement the Baum-Welch algorithm, which calculates the smoothing distributions $\pi_{k|n}$ for $k \leq n$.

b) Calculate the maximum likelihood estimator $\hat{X}_k$ of X_k, $k \leq n$, with respect to the smoothing distribution and compare it to the true signal.

4.5. Bayes' formula and measure changes

Let X and Y be random variables on $(\Omega, \mathscr{F}, \mathbb{P})$ with values in measurable spaces $(\mathbb{X}, \mathscr{X})$ and $(\mathbb{Y}, \mathscr{Y})$, respectively.

[1] The example is borrowed from [2] and based on [3].

a) Let $\mathbb{P} \ll \tilde{\mathbb{P}}$ with density $d\mathbb{P}/d\tilde{\mathbb{P}} = \Lambda(X,Y)$ for some measurable function Λ. Given a regular conditional probability $\tilde{P}_{X|Y}$ of X given Y under $\tilde{\mathbb{P}}$, show that a regular conditional probability $P_{X|Y}$ of X given Y under $\mathbb{P}$ exists and is given by

$$P_{X|Y}(y,A) = \frac{\int_A \Lambda(x,y)\tilde{P}_{X|Y}(y,dx)}{\int \Lambda(x,y)\tilde{P}_{X|Y}(y,dx)}, \qquad y \in \mathbb{Y}, A \in \mathscr{X},$$

where the fraction is set to zero if the denominator vanishes.

b) How can Bayes' formula from Exercise 1.4 be obtained as a special case of this?

References

[1] Olivier Cappé, Eric Moulines, and Tobias Ryden. *Inference in Hidden Markov Models*. Springer Series in Statistics. Springer Verlag, New York, 2005.

[2] Ramon van Handel. *Hidden Markov Models*. Lecture Notes. Princeton University, 2008.

[3] Sean A McKinney, Chirlmin Joo, and Taekjip Ha. “Analysis of single-molecule FRET trajectories using hidden Markov modeling”. In: *Biophysical journal* 91.5 (2006), pp. 1941–1951.

[4] H. Rauch, F. Tung, and C. Striebel. “Maximum likelihood estimates of linear dynamical systems”. In: *AIAA Journal* 3 (1965), pp. 1445–1450.