Stochastic Filtering (SS2016) Exercise Sheet 2

Lecture and Exercises: JProf. Dr. Philipp Harms
Due date: May 4, 2016

2.1. Best non-linear estimate

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. We fix a measurable "loss function" $L : \mathbb{X} \times \mathbb{X} \to \mathbb{R}_+$.

a) Let $\mathbb{X}$ be a Borel space. Show that minimizing the functional $\mathbb{E}[L(X, f(Y))]$ over all measurable functions $f : \mathbb{Y} \to \mathbb{X}$ is equivalent to minimizing $\mathbb{E}[L(X, \hat{X})]$ over all $\sigma(Y)$-measurable random variables $\hat{X}$ with values in $\mathbb{X}$.

   Hint. You may use the result [1, Lemma 1.13] on functional representation.

b) Let $\mathbb{X}$ be a separable Hilbert space, $X \in L^2(\Omega; \mathbb{X})$, and $L(x, \hat{x}) = \|x - \hat{x}\|^2$. Use the characterization in a) and Hilbert space projections to calculate the minimizer $\hat{X}$.

2.2. Best linear estimate

Let X and Y be random variables on $(\Omega, \mathscr{F}, \mathbb{P})$ with values in finite-dimensional vector spaces $(\mathbb{X}, \mathscr{X})$ and $(\mathbb{Y}, \mathscr{Y})$, respectively. We fix the "loss function" $L(x, \hat{x}) = \|x - \hat{x}\|^2$.

a) The best linear estimate of $X \in L^2(\Omega; \mathbb{X})$ is the minimizer of the quadratic loss function L over all random variables of the form $\hat{X} = \hat{x} + AY$, where $\hat{x} \in \mathbb{X}$ and $A : \mathbb{Y} \to \mathbb{X}$ is a linear map. Show that the best linear estimate exists, is unique, and can be expressed using orthogonal projections.

b) Give an example where the best linear estimate is strictly worse than the best non-linear estimate.

Remark. The vast majority of filters used in signal processing, pattern recognition, electronics, mechanical systems, and econometrics are linear.

2.3. Non-degenerate observations

Let (X,Y) be a HMM on $(\mathbb{X} \times \mathbb{Y}, \mathscr{X} \otimes \mathscr{Y})$ with parameters (P,K,μ) as in the lecture.

a) Give the definition of the condition that (X,Y) has non-degenerate observations.

b) Show under this condition that for any $n \in \mathbb{N}$ and $B_0,\dots,B_n \in \mathscr{Y}$ satisfying $\phi(B_1) > 0,\dots,\phi(B_n) > 0$ one has $\mathbb{P}[Y_0 \in B_0,\dots,Y_n \in B_n] > 0$.

   Remark. We can interpret this statement as follows: every sequence $y_0,\dots,y_n \in \operatorname{supp}(\phi)$ is a possible observation in the HMM and therefore an admissible input for the filtering algorithm.

2.4. Non-degenerate observations

Let (X,Y) be a HMM as in Exercise 2.3 with non-degenerate observations.

a) Show for each $n \in \mathbb{N}$, $x_{0:n} \in \mathbb{X}^{n+1}$ and $\tilde{x}_{0:n} \in \mathbb{X}^{n+1}$ that the probability measures $P_{Y_{0:n}|X_{0:n}}(x_{0:n},\cdot)$ and $P_{Y_{0:n}|X_{0:n}}(\tilde{x}_{0:n},\cdot)$ are equivalent.

   Remark. We can interpret this as all hidden states being observationally equivalent. In other words, the hidden states are not observable with certainty.

b) Use a) to show that the laws of $Y_{0:n}$ and $\tilde{Y}_{0:n}$ are equivalent for any other hidden Markov model $(\tilde{X},\tilde{Y})$ with parameters $(\tilde{P},K,\tilde{\mu})$.

   Remark. We can interpret this as (X,Y) and $(\tilde{X},\tilde{Y})$ being observationally equivalent. In other words, the parameters P and μ of the HMM are not observable with certainty.

2.5. Non-degenerate observations

Check if the condition of non-degenerate observations is satisfied for the following observation kernels $K(x, dy)$:

a) $\mathbb{X} = \mathbb{Y} = \mathbb{R}$, $K(x, \cdot) \sim N(x, 1)$.

b) $\mathbb{X} = (0, \infty)$, $\mathbb{Y} = \{0, 1, 2, \dots\}$, $K(x, \cdot) \sim \mathrm{Poiss}(x)$.

c) $\mathbb{X} = [0, 1]$, $\mathbb{Y} = \{0, 1\}$, $K(x, \cdot) = \mathrm{Ber}(x)$.

References

[1] Olav Kallenberg. *Foundations of modern probability*. 2nd ed. Springer Verlag, New York, 2002.