4F7-STATISTICAL SIGNAL ANALYSIS

S.S. SINGH

Abstract. Optimal filtering material from 3F3 required that the data generating processes satisfy certain stationarity assumptions. Modern filtering theory is introduced through state-space models that do not require any stationarity assumptions. State-space models are thus far more general and more widely applicable in real data settings. An even more general model is the hidden Markov model which will be studied in detail. Inference aims for the hidden Markov model will be defined and exact computation of the probability laws will be addressed. In many applications though exact computation is not possible and the most successful technique to date that addresses this problem is a Monte Carlo method called sequential importance sampling with resampling, also known as particle filtering. The particle filter will be derived and applied to both inference and model calibration for time-series data.

Course Structure

(1) State-space models: definition; examples; inference aims; exact computation via the Kalman filter.

(2) Hidden Markov models: definition; examples; inference aims; exact computation of the filter.

(3) Importance sampling: introduction; weight degeneracy.

(4) Sequential importance sampling and resampling: application to hidden Markov models; filtering; smoothing; exemplar problems in signal processing

(5) Calibrating hidden Markov models: maximum likelihood estimation; numerical implementation.

(6) Examples paper.

In these notes you will find some key learning objectives formulated as Exercises. (These exercises are not to be regarded as optional.) This has been done to emphasise that the statements of these key results can be verified using the knowledge you have acquired up to that point.

1. State Space Models

Definition. We will use the notation $\{\mathbf{W}_n\}_n \sim \mathrm{WN}(0, \{Q_n\}_n)$ for a sequence of random vectors $\mathbf{W}_n$ with zero mean and second moment

$$\mathbb{E}\left(\mathbf{W}_n \mathbf{W}_m^{\mathrm{T}}\right) = \begin{cases} 0 & m \neq n, \\ Q_n & m = n. \end{cases}$$

The sequence is a white noise (WN) sequence.

Bold face here represents a random vector as opposed to a random variable. The next definition is lengthy as it is comprised of a number of components.

Definition. A state space model for a time-series $\{\mathbf{Y}_n\}_n$ consists of two equations. The first is the observation equation

$$\mathbf{Y}_n = G_n \mathbf{X}_n + \mathbf{V}_n, \qquad n = 1, 2, \ldots \tag{1.1}$$

where $\{\mathbf{V}_n\}_n \sim \mathrm{WN}\,(0, \{R_n\}_n)$ and $\{G_n\}$ is a deterministic sequence of matrices. $\{\mathbf{Y}_n\}_n$ is a function of a sequence of unobserved random vectors $\{\mathbf{X}_n\}_n$ which evolve according to their own equation, called the state equation

$$\mathbf{X}_{n+1} = F_n \mathbf{X}_n + \mathbf{W}_n, \qquad n = 1, 2, \ldots \tag{1.2}$$

where $\{\mathbf{W}_n\}_n \sim \mathrm{WN}\,(0, \{Q_n\}_n)$ and $\{F_n\}$ is a deterministic sequence of matrices.

It is further assumed that the initial state $\mathbf{X}_1$ is uncorrelated with all noise terms in $\{\mathbf{W}_n\}_n$ and $\{\mathbf{V}_n\}_n$ and all vectors in $\{\mathbf{W}_n\}_n$ are uncorrelated with all vectors in $\{\mathbf{V}_n\}_n$, that is

$$\mathrm{Cov}(\mathbf{X}_1, \mathbf{W}_n) = 0, \qquad n \geq 1,$$

$$\mathrm{Cov}(\mathbf{X}_1, \mathbf{V}_n) = 0, \qquad n \geq 1,$$

$$\mathrm{Cov}(\mathbf{W}_m, \mathbf{V}_n) = 0, \qquad m \geq 1, n \geq 1. \tag{1.3}$$

Remark. For random vectors $\mathbf{U}$ and $\mathbf{V}$, $\mathrm{Cov}(\mathbf{U},\mathbf{V}) = \mathbb{E}(\mathbf{U}\mathbf{V}^{\mathrm{T}}) - \mathbb{E}(\mathbf{U})\mathbb{E}(\mathbf{V}^{\mathrm{T}})$.

Remark. If random variables U and V are independent then $\mathbb{E}(UV) = \mathbb{E}(U)\mathbb{E}(V)$ and so $\mathrm{Cov}(U,V) = 0$. However, $\mathrm{Cov}(U,V) = 0$ does not imply U and V are independent. Here is an example. Let $U \sim \mathcal{N}(0,1)$, and $V = |U|$. Clearly U and $V$are not independent but $\mathrm{Cov}(U,V) = 0$.

To better understand the state-space model, we express $\mathbf{X}_n$ as a function of its initial state and all its driving noise vectors.

$$
\begin{aligned}
\mathbf{X}_n &= F_{n-1}\mathbf{X}_{n-1} + \mathbf{W}_{n-1} \\
&= F_{n-1}\left(F_{n-2}\mathbf{X}_{n-2} + \mathbf{W}_{n-2}\right) + \mathbf{W}_{n-1} \\
&\vdots \\
&= \left(F_{n-1}\cdots F_1\right)\mathbf{X}_1 + \left(F_{n-1}\cdots F_2\right)\mathbf{W}_1 + \ldots + F_{n-1}\mathbf{W}_{n-2} + \mathbf{W}_{n-1}.
\end{aligned}
$$

Thus we see that the state at time n is a linear function of the initial state $\mathbf{X}_1$ and the driving noise sequence. As a memory aid for this composition, we can write $\mathbf{X}_n = f_n(\mathbf{X}_1, \mathbf{W}_1, \ldots, \mathbf{W}_{n-1})$.

Exercise. Show $\mathbb{E}\left(\mathbf{X}_m \mathbf{W}_n^{\mathrm{T}}\right) = 0$ and $\mathbb{E}\left(\mathbf{Y}_m \mathbf{W}_n^{\mathrm{T}}\right) = 0$ for $m \leq n$. Show $\mathbb{E}\left(\mathbf{X}_m \mathbf{V}_n^{\mathrm{T}}\right) = 0$ for all m, n and $\mathbb{E}\left(\mathbf{Y}_m \mathbf{V}_n^{\mathrm{T}}\right) = 0$ for $m < n$.

Not all the results of the exercise are shown but a representative one only, which is $\mathbb{E}\left(\mathbf{X}_m \mathbf{W}_n^{\mathrm{T}}\right) = 0$ for $m \leq n$. Recall that $\mathbf{X}_m$ is a linear function of $\mathbf{X}_1, \mathbf{W}_1, \ldots, \mathbf{W}_{m-1}$. Since $m \leq n$, there is no $\mathbf{W}_n$ term in $\mathbf{X}_m$. The assumptions in the definition of the state-space model state that $\mathbb{E}\left(\mathbf{X}_1 \mathbf{W}_n^{\mathrm{T}}\right) = 0$ and $\mathbb{E}\left(\mathbf{W}_i \mathbf{W}_n^{\mathrm{T}}\right) = 0$ when $i \neq n$, which implies $\mathbb{E}\left(\mathbf{X}_m \mathbf{W}_n^{\mathrm{T}}\right) = 0$.

Occasionally we are only given a time-series $\{\mathbf{Y}_n\}$ (without any mention of a hidden state process.) The next definition states that if the random vectors $\mathbf{Y}_n$ of this time-series can be expressed through a hidden state process then the time-series $\{\mathbf{Y}_n\}$ has a state-space representation.

Definition. A time-series $\{\mathbf{Y}_n\}$ has a state-space representation if $\{\mathbf{Y}_n\}$ can be expressed through a state-space model of the form (1.1)-(1.2).

We will now focus on scalar time-series models and not vector valued ones as we have in this introduction.