8. Non-linear filtering.

In this lecture, we consider the following filtering setting.

The signal $X_n, n=0,1\dots,$ forms random sequence, for every n the random variable X_n takes values in a finite "alphabet": $\{x_1,\dots,x_m\}$, that is X_n coincides with one of x_j's. The observation

$$Y_n = X_n + \eta_n,$$

where $(\eta_n)_{n\ge 1}$ is a random sequence of independent and identically distributed (i.i.d.) random variables sequence of random variables independent of $(X_n)_{n\ge 0}$. Henceforth, for notational convenience Y_0^n is referred as $(Y_0, Y_1, \dots, Y_n)$. For every n, the filtering distribution, in other words conditional distribution for X_n given Y_0^n,

$$\pi_n(1) = \boldsymbol{P}(X_n = x_1|Y_0^n), \dots, \pi_n(m) = \boldsymbol{P}(X_n = x_m|Y_0^n)$$

allows completely describes filtering estimate optimal from many points of view. For instance, the optimal in the mean square sense estimate, being the conditional expectation, is defined as

$$\widehat{X}_n = \boldsymbol{E}(X_n|Y_0^n) = \sum_{j=1}^m x_j \pi_n(j).$$

The maximum a'posterior probability, being popular filtering estimate, is defined by the filtering conditional distribution as well:

$$\widehat{X}_n = x_{j_n};\ j_n = \operatorname*{argmax}_{1\le j\le m} \pi_n(j).$$

A filtering estimate $\widehat{X}_n$, minimizing $\boldsymbol{E}|X_n - \widehat{X}_n|$, called the conditional median, is defined as:

$$\widehat{X}_n = \min_{x\in\{x_1,\dots,x_n\}} \Big\{ P(X_n \le x|Y_0^n) \ge P(X_n > x|Y_0^n) \Big\}.$$

Wonham filter

All these examples convince us of an importance to have algorithm for computing the filtering distribution. We show that such type algorithm is defined by the Wonham filter under assumption that the signal X_n is Markov chain. The Wonham filter is a typical non-linear filter. To derive it, we first fix assumptions on the signal X_n.

1. The random sequence $(X_n)_{n=0,1,\dots,}$ forms the Markov chain, i.e., corresponding to the finite alphabet values of X_n, for every $n\ge 1$, and $j=1,\dots,m$

$$P(X_n = x_j|X_{n-1}\dots,X_1,X_0) = P(X_n = x_j|X_{n-1}),$$

where $P(X_n = x_j|X_{n-1}) := \boldsymbol{E}I(X_n = x_j|X_{n-1})$ and, similarly, $P(X_n = x_j|X_{n-1}\dots,X_1,X_0)$ is defined. The transition probabilities

$$\lambda_{ij} = P(X_n = x_j|X_{n-1} = x_i),\ i,j = 1,\dots,m \tag{8.1}$$

are assumed to be known.

2. The initial distribution

$$p_0(j) = P(X_0 = x_j), j = 1, \dots, m. \tag{8.2}$$

is assumed to be known.

Concerning the noise sequence $(\eta_n)_{n\geq 1}$, being i.i.d. sequence of random variables independent of the signal (X_n), we assume only that the distribution function of η_1: $F(z) = P(\eta_1 \leq z)$ possesses density

$$f(z) = \frac{dF(z)}{dz} \tag{8.3}$$

and the density function $f(z)$ is known.

8.1. **Development of filtering distribution formulae.** We derive these formulae in two steps.

8.1.1. $(n = 0)$. Here, we present the filtering distribution for the time value zero

$$\pi_0(1) = P(X_0 = x_1)|Y_0), \dots, P(X_m = x_m|Y_0).$$

Proposition 8.1.

$$\pi_0(j) = \frac{f(Y_0 - x_j)p_0(j)}{\sum_{j=k}^{m} f(Y_0 - x_k)p_0(k)}, \ j = 1, \dots, m \tag{8.4}$$

where f is the density function given in (8.3) *and $p_0(j)$'s are entries of the initial distribution* (8.2).

Proof. Since $\pi_0(j) = \boldsymbol{E}(I(X_0 = x_j)|Y_0)$, to find $\pi_0(j)$ we apply conditions **(i)**, **(ii)** (Lect. 7) which completely define the conditional expectation. Notice first that by **(i)** $\pi_0(j) = G(Y_0)$ and by **(ii)** the function $G(Y_0)$ satisfies

$$\boldsymbol{E}g(Y_0)\{I(X_0 = x_j) - G(Y_0)\} = 0 \tag{8.5}$$

for any bounded function $g(Y_0)$.

We show now that (8.5) provides (8.4). Substituting $Y_0 = X_0 + \eta_0$ in (8.5), we find

$$\boldsymbol{E}g(X_0 + \eta_0)I(X_0 = x_j) = \boldsymbol{E}g(X_0 + \eta_0)G(X_0 + \eta_0).$$

Owing to an obvious equality

$$g(X_0 + \eta_0)I(X_0 = x_j) = g(x_j + \eta_0)I(X_0 = x_j)$$

and the fact that X_0 and η_0 are independent random variables, we have

$$\begin{aligned}
&\boldsymbol{E}g(x_j+\eta_0)I(X_0=x_j) \\
&\qquad = p_0(j)\int_{\mathbb{R}} g(x_j+y)f(y)dy \\
&\qquad\qquad = P(X_0=x_j)\int_{\mathbb{R}} g(y)f(y-x_j)dy. \qquad (8.6)
\end{aligned}$$

On the other hand, similarly we find

$$\begin{aligned}
&\boldsymbol{E}g(X_0+\eta_0)G(X_0+\eta_0) \\
&\qquad = \int_{\mathbb{R}} \boldsymbol{E}g(X_0+y)G(X_0+y)f(y)dy \\
&\qquad = \int_{\mathbb{R}}\sum_{k=1}^{m} \boldsymbol{E}g(x_k+y)G(x_k+y)I(X_0=x_k)f(y)dy \\
&\qquad = \int_{\mathbb{R}} g(y)G(y)\sum_{k=1}^{m} p_0(k)f(y-x_k)dy. \qquad (8.7)
\end{aligned}$$

By **(ii)**, (8.6)≡(8.7). Owing to an arbitrariness of g, the latter identity is valid not only in the integral form

$$p_0(j)\int_{\mathbb{R}} g(y)f(y-x_j)dy = \int_{\mathbb{R}} g(y)G(y)\sum_{k=1}^{m} p_0(k)f(y-x_k)dy$$

but also integrands coincide

$$p_0(j)f(y-x_j) = G(y)\sum_{k=1}^{m} p_0(k)f(y-x_k).$$

Hence, $G(y) = \frac{p_0(j)f(y-x_j)}{\sum_{k=1}^{m} p_0(k)f(y-x_k)}$ and, since $\pi_0(j) = G(Y_0)$, (8.4) holds true. □

8.1.2. $(n>0)$. Here, we present the filtering distribution for any time value $n>0$.

Theorem 8.2. *With* $\pi_0(j)$, $j=1,\dots,m$, *given in* (8.4), *filtering distributions*

$$\pi_n(1),\dots,\pi_n(m),\ n=1,2,\dots$$

are defined recursively

$$\pi_n(j) = \frac{f(Y_n-x_j)\sum_{i=1}^{m}\lambda_{ij}\pi_{n-1}(i)}{\sum_{i,k=1}^{m} f(Y_n-x_i)\lambda_{ik}\pi_{n-1}(k)}. \qquad (8.8)$$

Proof. Assuming that the filtering distribution $\pi_{n-1}(j),\ j=1,\dots,m$ is known we define first "one step predictable distribution":

$$\pi_{n|n-1}(j)=\boldsymbol{P}(X_n=x_j|Y_0^{n-1}),\ j=1,\dots,m. \tag{8.9}$$

Since $\pi_{n|n-1}(j)=\boldsymbol{E}\big(I(X_n=x_j)|Y_0^{n-1}\big)$, by property *8.* of the conditional expectation (Lect.7) and the fact that $\{Y_0^{n-1}\}\subset\{X_0^{n-1},\eta^{n-1}\}$, we have

$$\pi_{n|n-1}(j)=\boldsymbol{E}\Big(\boldsymbol{E}\big(I(X_n=x_j)|X_0^{n-1},\eta_0^{n-1}\big)\big|Y_0^n\Big). \tag{8.10}$$

Further, by property *15.* of the conditional expectation (Lect.7) and the fact that the signal (X_n) and noise (η_n) are independent we also have

$$P(X_n=x_j|X_0^{n-1},\eta_0^{n-1})=\boldsymbol{P}(X_n=x_j|X_0^{n-1}). \tag{8.11}$$

Finally, by Markov property:

$$\begin{aligned}P(X_n=x_j|X_0^{n-1}) &= P(X_n=x_j|X_{n-1})\\ &= \sum_{i=1}^m P(X_n=x_j|X_{n-1}=i)I(X_{n-1}=i)\\ &= \sum_{i=1}^m \lambda_{ij}I(X_{n-1}=i). \end{aligned}\tag{8.12}$$

Now, applying (8.12)⇒(8.11)⇒(8.10), we find

$$\pi_{n|n-1}(j)=\sum_{i=1}^m\lambda_{ij}\pi_{n-1}(i). \tag{8.13}$$

The next step consists in the definition of $\pi_n(j)=G(Y_0,\dots,Y_{n-1},Y_n)$. By **(ii)** (Lect.7) with a taste function $g(Y_0,\dots,Y_n)=g'(Y_0^{n-1})g''(Y_n)$, we have

$$\boldsymbol{E}\Big(g'(Y_0^{n-1})g''(Y_n)\big\{G(Y_0,\dots,Y_{n-1},Y_n)-I(X_n=x_j)\big\}\Big)=0. \tag{8.14}$$

Now, by property *8.* of the conditional expectation (Lect. 7), (8.14) is transformed into

$$\boldsymbol{E}g'(Y_0^{n-1})\boldsymbol{E}\Big(g''(Y_n)\big\{G(Y_0,\dots,Y_{n-1},Y_n)-I(X_n=x_j)\big\}\big|Y_0^{n-1}\Big)=0.$$

Hence, by an an arbitrariness of $g'(Y_1^{n-1})$,

$$\boldsymbol{E}\Big(g''(Y_n)\big\{G(Y_0,\dots,Y_{n-1},Y_n)-I(X_n=x_j)\big\}\Big|Y_0^{n-1}\Big)=0. \tag{8.15}$$

The equality, given in (8.15), is the main tool in the derivation of $\pi_n(j)$. Let us rewrite (8.15) as:

$$\boldsymbol{E}\Big(g''(Y_n)I(X_n=x_j)\Big|Y_0^{n-1}\Big)=\boldsymbol{E}\Big(g''(Y_n)G(Y_0,\dots,Y_{n-1},Y_n)\Big|Y_0^{n-1}\Big) \tag{8.16}$$

and compute the conditional expectations for both sides of (8.16). Write

$$\boldsymbol{E}\Big(g''(Y_n)I(X_n=x_j)\Big|Y_0^{n-1}\Big) = \boldsymbol{E}\Big(g''(x_j+\eta_n)I(X_n=x_j)\Big|Y_0^{n-1}\Big)$$
$$= \boldsymbol{E}\Big(\int_{\mathbb{R}} g''(x_j+y)f(y)dyI(X_n=x_j)\Big|Y_0^{n-1}\Big)$$
$$= \boldsymbol{E}\Big(I(X_n=x_j)\int_{\mathbb{R}} g(y)f(y-x_j)dy\Big|Y_0^{n-1}\Big)$$
$$= \pi_{n|n-1}(j)\int_{\mathbb{R}} g(y)f(y-x_j)dy$$

and

$$\begin{aligned}
&\boldsymbol{E}\Big(g''(Y_n)G(Y_0,\dots,Y_{n-1},Y_n)\Big|Y_0^{n-1}\Big)\\
=\;& \boldsymbol{E}\Big(g''(X_n+\eta_n)G(Y_0,\dots,Y_{n-1},X_n+\eta_n)\Big|Y_0^{n-1}\Big).\\
=\;& \boldsymbol{E}\Big(\int_{\mathbb{R}} g''(X_n+y)G(Y_0,\dots,Y_{n-1},X_n+y)f(y)dy\Big|Y_0^{n-1}\Big)\\
=\;& \boldsymbol{E}\Big(\int_{\mathbb{R}}\sum_{k=1}^m I(X_n=x_k)g''(x_k+y)G(Y_0,\dots,Y_{n-1},x_k+y)\\
&\quad\times f(y)dy\Big|Y_0^{n-1}\Big)\\
=\;& \sum_{k=1}^m\int_{\mathbb{R}}\pi_{n|n-1}(k)g''(x_k+y)G(Y_0,\dots,Y_{n-1},x_k+y)f(y)dy\\
=\;& \int_{\mathbb{R}}\sum_{k=1}^m\pi_{n|n-1}(k)g''(y)G(Y_0,\dots,Y_{n-1},y)f(y-x_k)dy.
\end{aligned}$$

Thus, both sides of (8.16) are presented now of the form which, by an arbitrariness of g'', provides the equality for integrands

$$\begin{aligned}
&\pi_{n|n-1}(j)f(y-x_j)\\
=\;& \sum_{k=1}^m\pi_{n|n-1}(k)g(y)G(Y_0,\dots,Y_{n-1},y)f(y-x_k),
\end{aligned}$$

that is

$$G(Y_0,\dots,Y_{n-1},y)=\frac{\pi_{n|n-1}(j)f(y-x_j)}{\sum\limits_{k=1}^m\pi_{n|n-1}(i)f(y-x_k)}.$$

This formula, jointly with (8.13), provides the desired result. □

Finally, we give description of (8.8) in the matrix-vector form. Set π_n the vector with entries $\pi_n(1),\dots,\pi_n(1)$, the matrix Λ with entries λ_{ji}, and also the vector $\mathfrak{f}(y)$ with entries $f(y-x_1),\dots,f(y-x_m)$. Denote by $\mathsf{diag}(\cdot)$ diagonal matrix with the diagonal $\cdot$. Then (8.8) possesses

$$\pi_n = \frac{\mathsf{diag}\big(\mathfrak{f}(Y_n)\big)\Lambda^T \pi_{n-1}}{\mathfrak{f}^T(Y_n)\Lambda^T \pi_{n-1}}. \tag{8.17}$$

Remark 8.3. The Wonham filter allows to create the optimal (!) in the mean square sense filter. Obviously, the optimal in the mean square sense linear filter also exits. In spite of, it is worse than the Wonham filter (its mean square error is larger), nevertheless sometimes it makes sense to create a linear filter too. In the case considered the Kalman filter is applicable. It is given at the end of Lect. 9.