STOCHASTIC PROCESSES. SOLUTIONS TO HOME ASSIGNMENTS

6. Non-Linear Filtering of Markov Processes

Problem 6.1

a) Define $Y_n = \Delta\xi_n = \xi_n - \xi_{n-1}$, $n \geq 1$. Clearly $Y_n = \theta_n \varepsilon_n^\alpha + (1-\theta_n)\varepsilon_n^\beta \in \{0,1\}$ and $\pi_n = P(\theta_n = 1|\xi_0^n) = P(\theta_n = 1|Y_1^n)$ P-a.s.

Now let $\pi_n = G(Y_n; Y_1^{n-1})$ and fix pair of numbers h_1 and h_0. Then

$$\mathbb{E}\Big(I(\theta_n = 1)(h_0 I(Y_n = 0) + h_1 I(Y_n = 1))|Y_1^{n-1}\Big) =$$
$$\mathbb{E}\Big(I(\theta_n = 1)(h_0 I(\varepsilon_n^\alpha = 0) + h_1 I(\varepsilon_n^\alpha = 1))|Y_1^{n-1}\Big) =$$
$$\pi_{n|n-1}(1-\alpha)h_0 + \pi_{n|n-1}\alpha h_1$$

Analogously

$$\mathbb{E}\Big(G(Y_n; Y_1^{n-1})(h_0 I(Y_n = 0) + h_1 I(Y_n = 1))|Y_1^{n-1}\Big) =$$
$$h_0 G(0; Y_1^{n-1})\mathbb{E}\Big(I(Y_n = 0)|Y_1^{n-1}\Big) + h_1 G(1; Y_1^{n-1})\mathbb{E}\Big(I(Y_n = 1)|Y_1^{n-1}\Big).$$

Now

$$\mathbb{E}\Big(I(Y_n = 0)|Y_1^{n-1}\Big) = \mathbb{E}\Big(I(\theta_n = 1)I(\varepsilon^\alpha = 0) + I(\theta_n = 0)I(\varepsilon^\beta = 0)|Y_1^{n-1}\Big) =$$
$$\pi_{n|n-1}(1-\alpha) + (1-\pi_{n|n-1})(1-\beta)$$

and similarly

$$\mathbb{E}\Big(I(Y_n = 1)|Y_1^{n-1}\Big) = \pi_{n|n-1}\alpha + (1-\pi_{n|n-1})\beta$$

By arbitrariness of h_0 and h_1, obtain

$$\pi_n = G(Y_n; Y_{n-1}) =$$
$$\frac{(1-\alpha)\pi_{n|n-1}(1-Y_n)}{\pi_{n|n-1}(1-\alpha) + (1-\pi_{n|n-1})(1-\beta)} + \frac{\alpha\pi_{n|n-1}Y_n}{\pi_{n|n-1}\alpha + (1-\pi_{n|n-1})\beta} =$$
$$\frac{(1-\alpha)\pi_{n|n-1}(1-\Delta\xi_n)}{\pi_{n|n-1}(1-\alpha) + (1-\pi_{n|n-1})(1-\beta)} + \frac{\alpha\pi_{n|n-1}\Delta\xi_n}{\pi_{n|n-1}\alpha + (1-\pi_{n|n-1})\beta}$$

The $\pi_{n|n-1}$ is recalculated by familiar transition formula

$$\pi_{n|n-1} = \lambda_2\pi_{n-1} + (1-\lambda_1)(1-\pi_{n-1})$$

b)

Date: Summer, 2004.

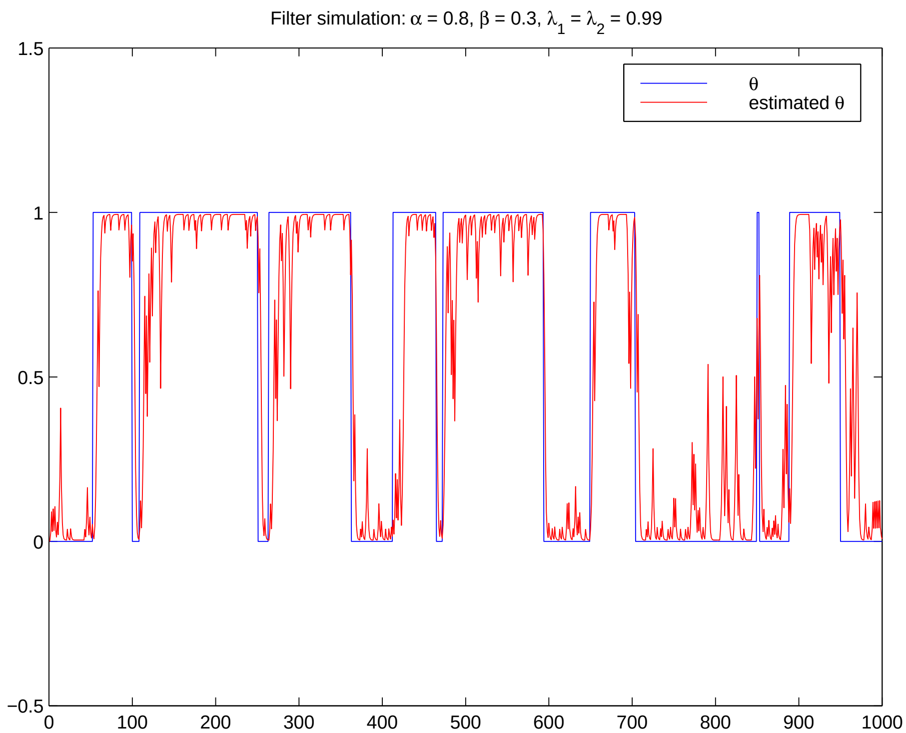

FIGURE 1. Typical original and estimated path

(a) If $\alpha = 1$ and $\beta = 0$, i.e. in one state each time there's an arrival (active state) and in the other no arrivals occur (idle state), the filter gives:

$$\pi_n(\xi) = \Delta\xi_n.$$

That is if the counter is not updated, the idle state is estimated and vise versa.

(b) If $\lambda_1 = 0$ and $\lambda_2 = 1$, i.e. the system is always pushed into one state, we obtain:

$$\pi_n(\xi) \equiv 1$$

regardless of observations.

(c) If $\lambda_1 = 1$ and $\lambda_2 = 1$, that is the system always stays in one of two states, the filter is simplified appropriately (How?)

Problem 6.2

(a) The process $(X_n, Y_n)_{n\geq 0}$ is not necessarily Gaussian. E.g. $a_0 = 0$, $A_0 = 0$, $a_1 = 0$ and $A_1(Y_0^{n-1}) = Y_{n-1}$. Then:

$$Y_1 = Y_0X_0 + B\xi_1$$

Assume that (Y_1, X_0) is Gaussian. Assuming that X_0 and Y_0 independent, $\mathbb{E}(Y_1|X_0) = 0$ and

$$\mathbf{Var}(Y_1|X_0) \quad = \quad X_0^2 \mathbb{E} Y_0^2 + B^2 = \text{funct}(X_0)$$

The latter contradicts the assumption.

(b) Though the process (X_n, Y_n) is not generally Gaussian, it is *conditionally* Gaussian (the dependencies of a_i, A_i on Y_0^{n-1} are omitted for brevity)

$$\begin{aligned}
\varphi_n(\lambda,\mu) &= \mathbb{E}\big(e^{-i\lambda X_n - i\mu Y_n}|Y_0^{n-1}\big) = \\
&= \mathbb{E}\big(\exp\{-i\lambda(a_0 + a_1 X_{n-1} + b\varepsilon_n) - \\
&\quad -i\mu(A_0 + A_1 X_{n-1} + B\xi_n)\}|Y_0^{n-1}\big) = \\
&= \mathbb{E}\Big(\mathbb{E}\big[\exp\{-i\lambda(a_0 + a_1 X_{n-1}) - 1/2 b^2\lambda^2 \\
&\quad -i\mu(A_0 + A_1 X_{n-1}) - 1/2 B^2\mu^2\}|X_{n-1}, Y_0^{n-1}\big]|Y_0^{n-1}\Big)
\end{aligned}$$

The latter suggests that, given Y_0^{n-1} and X_{n-1}, the pair (X_n, Y_n) is Gaussian. We proceed by induction: assume that the conditional density of X_{n-1}, given Y_0^{n-1} is Gaussian with

$$\mathbb{E}(X_{n-1}|Y_0^{n-1}) \stackrel{\triangle}{=} m_{n-1}(Y_0^{n-1})$$

and

$$\mathbb{E}\left(\big[X_{n-1} - m_{n-1}\big]^2 \big| Y_0^{n-1}\right) \stackrel{\triangle}{=} P_{n-1}(Y_0^{n-1})$$

Then

$$\begin{aligned}
\varphi_n(\lambda,\mu) &= \mathbb{E}\Big(\exp\{-i\lambda(a_0 + a_1 X_{n-1}) - 1/2 b^2\lambda^2 \\
&\quad -i\mu(A_0 + A_1 X_{n-1}) - 1/2 B^2\mu^2\}|Y_0^{n-1}\Big) = \\
&= \exp\big\{-i\lambda(a_0 + a_1 m_{n-1}) - i\mu(A_0 + A_1 m_{n-1}) \\
&\quad -1/2(a_1^2 P_{n-1} + b^2)\lambda^2 - \lambda\mu A_1 a_1 P_{n-1} \\
&\quad -1/2(A_1^2 P_{n-1} + B^2)\mu^2\big\}
\end{aligned}$$

which implies that the density of X_n given Y_0^n is Gaussian. Hence the optimal filter is given by:

$$m_n = a_0 + a_1 m_{n-1} + \frac{A_1 a_1 P_{n-1}}{A_1^2 P_{n-1} + B^2}(Y_n - A_0 - A_1 m_{n-1}) \tag{6.1}$$

$$P_n = a_1^2 P_{n-1} + b^2 - \frac{A_1^2 a_1^2 P_{n-1}^2}{A_1^2 P_{n-1} + B^2} \tag{6.2}$$

Note 1: the essential difference between Kalman filter and so called *conditionally Gaussian filter* given by (6.1) is that the Riccati equation (6.2) in the latter depends on the observation process $(Y_n)_{n\ge 1}$ and hence cannot be computed *off line*. In a certain sense, it is

an *adaptive* filter, since its parameters vary with the recorded data. This filter is extremely useful in control theory.
Note 2: the equations (6.1) and (6.2) can be derived directly from the conditional density recursion.

(c) If all the functionals are constant the filter gets the form of conventional Kalman filter - the Riccati equation becomes decoupled from the observation process.

Problem 6.3

a) Put $\Delta_n \equiv \theta_n - \widetilde{\theta}_n$ and assume it is small enough to justify:

$$\begin{aligned} h(\theta_{n-1}) &\approx h(\widetilde{\theta}_{n-1}) + h'(\widetilde{\theta}_{n-1})(\theta_{n-1} - \widetilde{\theta}_{n-1}) \\ g(\theta_{n-1}) &\approx g(\widetilde{\theta}_{n-1}) + g'(\widetilde{\theta}_{n-1})(\theta_{n-1} - \widetilde{\theta}_{n-1}) \end{aligned} \tag{6.3}$$

Then the system equations are transformed into a pair of linear (in θ !) recursions:

$$\begin{aligned} \theta_n &= \Big[h(\widetilde{\theta}_{n-1}) - h'(\widetilde{\theta}_{n-1})\widetilde{\theta}_{n-1}\Big] + h'(\widetilde{\theta}_{n-1})\theta_{n-1} + u_n \\ \xi_n &= \Big[g(\widetilde{\theta}_{n-1}) - g'(\widetilde{\theta}_{n-1})\widetilde{\theta}_{n-1}\Big] + g'(\widetilde{\theta}_{n-1})\theta_{n-1} + v_n \end{aligned} \tag{6.4}$$

Applying the equations of Conditionally Gaussian Filter and setting $\widetilde{\theta}_n$ to be equal to the obtained estimate, we arrive at:

$$\begin{aligned} \widetilde{\theta}_n &= h(\widetilde{\theta}_{n-1}) + \frac{g'(\widetilde{\theta}_{n-1})h'(\widetilde{\theta}_{n-1})P_{n-1}}{\Big(g'(\widetilde{\theta}_{n-1})\Big)^2 P_{n-1} + B^2}\Big[\xi_n - g(\widetilde{\theta}_{n-1})\Big] \\ P_n &= \Big[h'(\widetilde{\theta}_{n-1})\Big]^2 P_{n-1} + b^2 - \frac{\Big[g'(\widetilde{\theta}_{n-1})h'(\widetilde{\theta}_{n-1})P_{n-1}\Big]^2}{\Big(g'(\widetilde{\theta}_{n-1})\Big)^2 P_{n-1} + B^2} \end{aligned} \tag{6.5}$$

b) Since the EKF is a purely heuristic device, in certain cases it will fail to produce reasonable estimate. E.g. if $h(x) = \tanh(x^3)$, then $h(0) = 0$ and $h'(0) = 0$. Once the state estimate $\widetilde{\theta}_n$ is rounded to zero during the calculations the filter will be stuck, i.e. $\widetilde{\theta}_k = 0$, for all $k \geq n$.

Problem 6.4

a. The suitable model is:

$$\begin{aligned} \theta_n &= \theta_{n-1}, \quad \theta_0 = \theta \\ \xi_n &= 1/2\theta_{n-1} + \varepsilon_n \end{aligned}$$

where $\varepsilon_n := \theta_{n-1}(U_n - 1/2)$. Clearly $\mathbb{E}\varepsilon_n = 0$, $\mathbb{E}\varepsilon_n^2 = 1/3 \cdot 1/12 = 1/36$, $\mathbb{E}\varepsilon_n\varepsilon_m = 0$, $n \neq m$ and ε and θ are orthogonal. The corresponding Kalman filter is

$$\begin{aligned}\widehat{\theta}_n &= \widehat{\theta}_{n-1} + \frac{1/2P_{n-1}}{1/4P_{n-1}+1/36}(\xi_n - 1/2\widehat{\theta}_{n-1})\\ P_n &= P_{n-1} - \frac{1/4P_{n-1}^2}{1/4P_{n-1}+1/36} = \frac{1/36P_{n-1}}{1/4P_{n-1}+1/36}\end{aligned}$$

subject to $\widehat{\theta}_0 = 1/2$, $P_0 = 1/12$.

b. Let $Q_n = 1/P_n$, then

$$Q_n = 36/4 + Q_{n-1} = 12 + 9n \implies P_n = \frac{1}{12+9n}$$

Clearly $P_n \to 0$ as $n \to \infty$ with linear rate.

c. Note that

$$\widetilde{\theta}_n = \max(\xi_1, \xi_2, ..., \xi_n)$$

The conditional density is given by (Why?)

$$f_{\widetilde{\theta}_n|\theta}(x;\theta) = \frac{d}{dx}\mathbb{P}(\widetilde{\theta}_n \le x|\theta) = \frac{d}{dx}\left(\frac{x}{\theta}\right)^n I(x\in[0,\theta]) = \frac{n}{\theta^n}x^{n-1}I(x\in[0,\theta])$$

Calculate the *conditional* variance:

$$\mathbb{E}\big((\theta-\widetilde{\theta}_n)^2\big|\theta\big) = \theta^2 - 2\theta\mathbb{E}(\widetilde{\theta}_n|\theta) + \mathbb{E}\big(\widetilde{\theta}_n^2|\theta\big)$$

Clearly

$$\mathbb{E}(\widetilde{\theta}_n|\theta) = \int_0^\infty x f_{\widetilde{\theta}_n|\theta}(x;\theta)dx = \frac{n}{\theta^n}\int_0^\theta x^n dx = \frac{n}{(n+1)}\theta$$

and

$$\mathbb{E}(\widetilde{\theta}_n^2|\theta) = \int_0^\infty x^2 f_{\widetilde{\theta}_n|\theta}(x;\theta)dx = \frac{n}{\theta^n}\int_0^\theta x^{n+1} dx = \frac{n}{(n+2)}\theta^2$$

So

$$\mathbb{E}\big((\theta-\widetilde{\theta}_n)^2\big|\theta\big) = \theta^2\left(1 - 2\frac{n}{(n+1)} + \frac{n}{(n+2)}\right) = \frac{2\theta^2}{(n+1)(n+2)}$$

so that

$$Q_n = \frac{2/3}{(n+1)(n+2)}$$

d. Clearly $Q_n \to 0$ and the convergence rate is n^2. So $\widetilde{\theta}_n$ is more accurate than $\widehat{\theta}_n$ asymptotically. It is quite obvious that for small n, $\widehat{\theta}_n$ is better than $\widetilde{\theta}_n$: note e.g. $\widetilde{\theta}_1 = \xi_1$, i.e. it is linear in ξ_1 and thus is clearly suboptimal. This can be verified also directly via the formulae. So the filtering estimate can be improved if the linear filter is used up to some n^* (determined by the eq. $Q_{n^*} = P_{n^*}$) and afterwards the "maximum" filter is applied.

e. $\widetilde{\theta}$ is clearly suboptimal (since it is even suboptimal with respect to the best linear estimate for small n). In fact in this problem the exact conditional expectation can be found as follows. By the recursive Bayes formula we have:

$$\mathbb{E}(\theta|\xi_1^n) = \frac{\int_0^1 s f_{\xi_1^n|\theta}(\xi_1, ..., \xi_n; s) f_\theta(s) ds}{\int_0^1 f_{\xi_1^n|\theta}(\xi_1, ..., \xi_n; x) f_\theta(x) dx}$$

with obvious notations for conditional densities. Let $\xi_n^* = \max_{i \le n} \xi_i$, then for $n \ge 3$

$$\mathbb{E}(\theta|\xi_1^n) = \frac{\int_0^1 s^{-n+1} \prod_{i=1}^n I(\xi_i \in [0, s]) ds}{\int_0^1 s^{-n} \prod_{i=1}^n I(\xi_i \in [0, s]) ds} = \int_{\xi_n^*}^1 s^{-n+1} ds \Big/ \int_{\xi_n^*}^1 s^{-n} ds =$$

$$= \frac{s^{-n+2}}{-n+2}\Big]_{s=\xi_n^*}^{s=1} \Big/ \frac{s^{-n+1}}{-n+1}\Big]_{s=\xi_n^*}^{s=1} = \frac{1-n}{2-n} \cdot \frac{1-(\xi_n^*)^{-n+2}}{1-(\xi_n^*)^{-n+1}} =$$

$$= \frac{(n-1)\big(\xi_n^* - (\xi_n^*)^{n-1}\big)}{(n-2)\big(1-(\xi_n^*)^{n-1}\big)}$$

(a) Note that the optimal estimate approaches $\widetilde{\theta}_n$ as $n \to \infty$ exponentially fast (Why?), so that it is expected that the minimal mean square error decays to zero as $1/n^2$.

(b) As it was mentioned earlier, generally the recursive optimal filters are infinite dimensional. Remarkably, in this case a one dimensional recursive (since ξ_n^* can be calculated recursively!) filter is available. Moreover observe that $\xi_n^* = \max_{i \le n} \xi_i$ is *sufficient* statistic, i.e. it incorporates all the "information", contained in ξ_1^n, needed for calculation of the optimal estimate.