Lecture 1. Exercise 11

$X = (X_j)_{j \in \mathbb{Z}_+}$ Markov chain with values in a finite set $S = \{a_1, \dots, a_d\}$

$\Lambda = (\lambda_{ij})$ matrix of transition probabilities and vector of initial distributions γ

Different from notes.

$$P(X_j = a_l \mid X_{j-1} = a_m) = \lambda_{ml}, \quad P(X_0 = a_l) = \gamma_l, \quad 1 \le l, m \le d$$

a) Let p_n be the vector with entries $p_j(i) = P(X_j = a_i)$.

Show that
$$p_j = \Lambda^T p_{j-1} \quad \text{s.t.} \quad p_0 = \gamma \quad j \geq 0.$$

This is just the following formula in matrix form

$$p_j(i) = P(X_j = a_i) = P\left(\{X_j = a_i\} \cap \left\{\biguplus_{k=1}^{d} X_{j-1} = a_k\right\}\right)$$

$$= \sum_{k=1}^{d} P(\{X_j = a_i\} \cap \{X_{j-1} = a_k\})$$

$$= \sum_{k=1}^{d} P(X_j = a_i \mid X_{j-1} = a_k)\, P(X_{j-1} = a_k)$$

$$= \sum_{k=1}^{d} \lambda_{ki}\, p_{j-1}(k) = \left(\Lambda^T p_{j-1}\right)_i$$

where for a matrix $A \in \mathbb{R}^{n \times m}$ $(A)_i$ = i-th row.

b) I_j vectors with entries $I_j(i) = 1\!\!1_{\{X_j = a_i\}}$, $j \geq 0$

Show that $\exists$ a sequence of orthogonal random vectors ε_j such that

$$I_j = \Lambda^T I_{j-1} + \varepsilon_j \qquad j \geq 0.$$

Find its mean and covariance matrix

Define

$$\varepsilon_j := I_j - \Lambda^T I_{j-1}$$

- $E[\varepsilon_j] = E[I_j] - \Lambda^T E[I_{j-1}] = p_j - \Lambda^T p_{j-1} \equiv 0$

- $E[\varepsilon_j \varepsilon_j^T] = E[(I_j - \Lambda^T I_{j-1})(I_j - \Lambda^T I_{j-1})^T]$

$= E[(I_j - \Lambda^T I_{j-1})(I_j^T - I_{j-1}^T \Lambda)]$

$= E[I_j I_j^T - I_j I_{j-1}^T \Lambda - \Lambda^T I_{j-1} I_j^T + \Lambda^T I_{j-1} I_{j-1}^T \Lambda]$

$= ① + ② + ③ + ④$

① $E[I_j I_j^T] = \text{diag}(p_j) := \begin{pmatrix} p_j(1) & & 0 \\ & \ddots & \\ 0 & & p_j(d) \end{pmatrix}$

② $E[I_j I_{j-1}^T \Lambda] = E[E[I_j \mid \sigma(X_{j-1})] I_{j-1}^T \Lambda] = (*)$

$\cdot\ E[I_j(i) \mid \sigma(X_{j-1})] = \sum_{k=1}^{d} \dfrac{E[I_j(i) 1\!\!1_{\{X_{j-1} = a_k\}}]}{P(X_{j-1} = a_k)} 1\!\!1_{\{X_{j-1} = a_k\}}$

$= \sum_{k=1}^{d} \dfrac{P(\{X_j = a_i\} \cap \{X_{j-1} = a_k\})}{P(X_{j-1} = a_k)} 1\!\!1_{\{X_{j-1} = a_k\}}$

$= \sum_{k=1}^{d} P(X_j = a_i \mid X_{j-1} = a_k) 1\!\!1_{\{X_{j-1} = a_k\}}$

$$= \sum_{k=1}^{d} \lambda_{ki} \, 1\!\!1_{\{X_{j-1} = a_k\}} = (\Lambda^T I_{j-1})_i$$

$$(*) = E[\Lambda^T I_{j-1} I_{j-1}^T \Lambda]$$

(3) $E[\Lambda^T I_{j-1} I_j^T] = E[\Lambda^T I_{j-1} E[I_j^T \mid \sigma(X_{j-1})]] = (**)$

By the computations in (2) we can deduce that

$$E[I_j^T \mid \sigma(X_{j-1})] = (\Lambda^T I_{j-1})^T = I_{j-1}^T \Lambda$$

and, hence,

$$(**) = E[\Lambda^T I_{j-1} I_{j-1}^T \Lambda]$$

(4) $E[\Lambda^T I_{j-1} I_{j-1}^T \Lambda] = \Lambda^T E[I_{j-1} I_{j-1}^T] \Lambda$

$$= \Lambda^T \operatorname{diag}(p_{j-1}) \Lambda$$

Hence

$$E[\varepsilon_j \varepsilon_j^T] = \operatorname{diag}(p_j) - \Lambda^T \operatorname{diag}(p_{j-1}) \Lambda$$

Orthogonality

Note that $\varepsilon_j = I_j - \Lambda^T I_{j-1}$ is $\mathcal{F}_j^X = \sigma(X_0, \dots, X_j)$-adapted

and $E[\varepsilon_j \mid \mathcal{F}_{j-1}^X] = E[I_j \mid \mathcal{F}_{j-1}^X] - \Lambda^T I_{j-1}$

$$= \Lambda^T I_{j-1} - \Lambda^T I_{j-1} = 0$$

Hence ε_j is a martingale difference. Let $i < j$

$$E[\varepsilon_i \varepsilon_j^T] = E[\underset{i \le j-1}{\varepsilon_i} E[\varepsilon_j^T \mid \mathcal{F}_{j-1}]] = E[\varepsilon_i 0^T] = 0$$

c) $X_j = h(X_j) + \sigma \xi_j \quad , j \geq 1$

ξ is a white noise and $\sigma > 0$

Let $h \in \mathbb{R}^d$ be the vector with entries $h(a_i)$ $i = 1, \dots, d$.

Verify that $\quad Y_j = h^T I_j + \sigma \xi_j$

$$h(X_j) = h(X_j) \, \mathbb{1}_\Omega = h(X_j) \, \mathbb{1}_{\biguplus_{i=1}^d \{X_j = a_i\}}$$

$$= \sum_{i=1}^d h(X_j) \, \mathbb{1}_{\{X_j = a_i\}} = \sum_{i=1}^d h(a_i) \, \mathbb{1}_{\{X_j = a_i\}}$$

disjoint union

$$= h^T I_j$$

d) Derive the Kalman-Bucy filter for

$$\hat{I}_j = \hat{E}[I_j \mid \mathcal{L}^Y_j]$$

The model is

$$I_j = \Lambda^T I_{j-1} + \Sigma_j \varepsilon_j \qquad I_j \in \mathbb{R}^d$$

$$Y_j = h^T I_j + \sigma \xi_j \qquad Y_j \in \mathbb{R}$$

$$= h^T \Lambda^T I_{j-1} + h^T \Sigma_j \varepsilon_j + \sigma \xi_j$$

where Σ_j is such that

$$\Sigma_j \Sigma_j^T = \operatorname{diag}(p_j) - \Lambda^T \operatorname{diag}(p_{j-1}) \Lambda = \operatorname{Cov}(\varepsilon_j)$$

Then,

$a_0(j) \equiv 0 \quad a_1(j) = \Lambda^T \quad a_2(j) = 0 \;, b_1(j) = \Sigma_j \;, b_2(j) = 0$

$A_0(j) = 0 \quad A_1(j) = h^T \Lambda^T \;, A_2(j) = 0 \;, B_1(j) = h^T \Sigma_j \;, B_2(j) = \sigma$

and

$$a_l P_{j-l} A_l^T = \Lambda^T P_{j-l} \Lambda h$$

$$b \circ B = \Sigma_j (h^T \Sigma_j)^T = \Sigma_j \Sigma_j^T h = cov(\varepsilon_j) h$$

$$A_l P_{j-l} A_l^T + B \circ B = h^T \Lambda^T P_{j-l} \Lambda h + h^T \Sigma_j \Sigma_j^T h$$

$$= h^T (\Lambda^T P_{j-l} \Lambda + cov(\varepsilon_j)) h$$

$$(A_l P_{j-l} A_l^T + B \circ B)^{\oplus} = \frac{1}{h^T (\Lambda^T P_{j-l} \Lambda + cov(\varepsilon_j)) h}$$

$$a_l P_{j-l} a_l^T = \Lambda P_{j-l} \Lambda^T$$

$$b \circ b = \Sigma_j \Sigma_j^T = cov(\varepsilon_j)$$

Hence,

$$\hat{I}_j = \Lambda^T \hat{I}_{j-l} + (\Lambda^T P_{j-l} \Lambda + cov(\varepsilon_j)) h \frac{(Y_j - h^T \Lambda^T \hat{I}_j)}{h^T (\Lambda^T P_{j-l} \Lambda + cov(\varepsilon_j)) h}$$

$$P_j = \Lambda P_{j-l} \Lambda^T + cov(\varepsilon_j)$$

$$+ \underbrace{\frac{1}{h^T (\Lambda^T P_{j-l} \Lambda + cov(\varepsilon_j))}}_{} (\Lambda^T P_{j-l} \Lambda + cov(\varepsilon_j)) h \left(h^T (\Lambda^T P_{j-l} \Lambda + cov(\varepsilon_j))^T\right)$$

$$E[I_0] = V \qquad E[Y_0] = 0 \qquad Cov(I_0, Y_0) = 0$$

$$cov(Y_0)^{\oplus} = 0 \quad , \quad cov(X_0) = E[I_0 I_0^T] - VV^T$$

$$= diag(V) - VV^T$$

Then,

$$\hat{I}_0 = V \quad \text{and} \quad \hat{P}_0 = diag(V) - VV^T$$

e) What would be the estimate of $\hat{E}[g(X_j) \mid \mathcal{L}^Y_j]$ for any $g : S \to \mathbb{R}$ in terms of $\hat{I}_j$?
In particular, $\hat{X}_j = \hat{E}[X_j \mid \mathcal{L}^Y_j]$?

Note that $g(X_j)$, even for nonlinear g, can be written as a linear combination of the components of I_j, i.e.,

$$g(X_j) = \sum_{i=1}^{d} g(a_i) \cdot \mathbb{1}_{\{X_j = a_i\}} = g^T I_j ,$$

where g is the vector with components $g(a_i)$, $i = 1, \dots, d$.

Therefore,

$$\widehat{g(X_j)} = \hat{E}[g(X_j) \mid \mathcal{L}^Y_j] = \hat{E}[g^T I_j \mid \mathcal{L}^Y_j]$$

$$\underset{\text{linearity of } \hat{E}[\cdot \mid \mathcal{L}^Y_j].}{=} g^T \hat{E}[I_j \mid \mathcal{L}^Y_j] = g^T \hat{I}_j$$

For $g(x) = Id(x)$

$$\hat{X}_j = \hat{E}[Id(X_j) \mid \mathcal{L}^Y_j] = a^T \hat{I}_j \qquad a = (a_1, \dots, a_d)^T$$