ESE 650, SPRING 2023

HOMEWORK 1

ANH NGUYEN [TUNA28NG@SEAS]

Solution 1 (Time spent: 10 hour)**.** See the code on Gradescope

Solution 2 (Time spent: 8 hour)**.** (a) See the code on Gradescope.

(b) We have

$$\begin{aligned}\xi_k(x,x') &= P(X_k = x, X_{k+1} = x' | Y_1, ..., Y_t, \lambda) \\ &= \frac{P(X_k = x, X_{k+1} = x', Y_1, ..., Y_t, \lambda)}{P(Y_1, ..., Y_t, \lambda)} \\ &= \frac{P(Y_1, ..., Y_t, \lambda | X_k = x, X_{k+1} = x') P(X_k = x, X_{k+1} = x')}{P(Y_1, ..., Y_t, \lambda)}\end{aligned}$$

The numerator:

$P(Y_1, ..., Y_t | X_k = x, X_{k+1} = x') P(X_{k+1} = x' | X_k = x) P(X_k = x)$

$= P(Y_1, ..., Y_k | X_k = x, X_{k+1} = x') P(Y_{k+1}, ..., Y_t | X_k = x, X_{k+1} = x') P(X_{k+1} = x' | X_k = x) P(X_k = x)$

$= P(Y_1, ..., Y_k | X_k = x) P(Y_{k+1}, ..., Y_t | X_{k+1} = x') P(X_{k+1} = x' | X_k = x) P(X_k = x)$ following Markov property

$= \dfrac{P(Y_1, ..., Y_k, X_k = x)}{P(X_k = x)} P(Y_{k+1} | X_{k+1} = x') P(Y_{k+2}, ..., Y_t | X_{k+1} = x') P(X_{k+1} = x' | X_k = x) P(X_k = x)$

$= P(Y_1, ..., Y_k, X_k = x) P(Y_{k+1} | X_{k+1} = x') P(Y_{k+2}, ..., Y_t | X_{k+1} = x') P(X_{k+1} = x' | X_k = x)$

Thus,

$$\xi_k(x,x') = \eta \alpha_k(x) T_{x,x'} M_{x',y_{k+1}} \beta_{k+1}(x')$$

where $\eta^{-1} = P(Y_1, ..., Y_t, \lambda)$

(c) See the code on Gradescope

Solution 3 (Time spent: 4 hour)**.** a) Compute probabilities:

(1)

$$\begin{aligned}P(X_{k+1}=x_j|X_k=x_i,Y_1,...Y_t) &= \frac{P(X_{k+1}=x_j,X_k=x_i,Y_1,...,Y_t)}{P(X_k=x_i,Y_1,...,Y_t)}\\ &= \frac{P(Y_1,...,Y_t|X_k=x_i,X_{k+1}=x_j)P(X_k=x_i,X_{k+1}=x_j)}{P(X_k=x_i|Y_1,...,Y_t)P(Y_1,...,Y_t)}\end{aligned}$$

Here, the denominator can be easily compute from the derivation of the smoothing problem (section 2.4.4 in Chapter 2 note):

$$P(X_k=x_i|Y_1,...,Y_t)P(Y_1,...,Y_t)=\beta_k(x)\alpha_k(x)$$

The numerator:

$$\begin{aligned}&\big(P(Y_1,...,Y_k|X_k=x_i,X_{k+1}=x_j)P(Y_{k+1},...,Y_t|X_k=x_i,X_{k+1}=x_j)\big)\times\\ &\quad\big(P(X_{k+1}=x_j|X_k=x_i)P(X_k=x_i)\big)\end{aligned}$$

From Markovian property, $Y_1,...,Y_k$ is not dependent on $X_{k+1}=x_j$, and all the future observations $Y_k+1,...,Y_t$ can relate to $X_k=x_i$ through $X_{k+1}=x_j$. So we can write the first term in the numerator as follows:

$$\begin{aligned}&\big(P(Y_1,...,Y_k|X_k=x_i)P(Y_{k+1},...,Y_t|X_{k+1}=x_j)\big)\\ &=\frac{P(Y_1,...,Y_k,X_k=x_i)}{P(X_k=x_i)}P(Y_{k+1}|X_{k+1}=x_j)P(Y_{k+2},...,Y_t|X_{k+1}=x_j)\end{aligned}$$

Bring all the together and cancel out $P(X_k=x_i)$ we have:

$$\begin{aligned}P(X_{k+1}=x_j|X_k=x_i,Y_1,...Y_t) &= \frac{\alpha_k(x)M_{ij}\beta_{k+1}(x)T_{ij}}{\beta_k(x)\alpha_k(x)}\\ &= \frac{M_{ij}\beta_{k+1}(x)T_{ij}}{\beta_k(x)}\end{aligned}$$

(2)

$$\begin{aligned}&P(X_k=x_i|X_{k+1}=x_j,Y_1,...Y_t)\\ &=\frac{P(X_{k+1}=x_j,X_k=x_i,Y_1,...,Y_t)}{P(X_{k+1}=x_j,Y_1,...,Y_t)}\end{aligned}$$

The numerator here is computed similarly to last question. And the denominator:

$$P(X_{k+1}=x_j|Y_1,...,Y_t)P(Y_1,...,Y_t)=\beta_{k+1}(x)\alpha_{k+1}(x)$$

Thus,

$$P(X_{k+1} = x_j | X_k = x_i, Y_1, ...Y_t) = \frac{\alpha_k(x) M_{ij} \beta_{k+1}(x) T_{ij}}{\beta_{k+1}(x)\alpha_{k+1}(x)}$$
$$= \frac{\alpha_k(x) M_{ij} T_{ij}}{\alpha_{k+1}(x)}$$

(3)

$$P(X_{k-1} = x_i, X_k = x_j, X_{k+1} = x_l | Y_1, ...Y_t)$$
$$= \frac{P(X_{k-1} = x_i, X_k = x_j, X_{k+1} = x_l, Y_1, ..., Y_t)}{P(Y_1, ..., Y_t)}$$
$$= \frac{P(Y_1, ..., Y_t | X_{k-1} = x_i, X_k = x_j, X_{k+1} = x_l) P(X_{k-1} = x_i, X_k = x_j, X_{k+1} = x_l)}{P(Y_1, ..., Y_t)}$$
$$= \frac{P(Y_1, ..., Y_t | X_{k-1} = x_i, X_k = x_j, X_{k+1} = x_l) P(X_{k-1} = x_i, X_k = x_j, X_{k+1} = x_l)}{P(Y_1, ..., Y_t)}$$

The numerator:

$$P(Y_1, ..., Y_t | X_{k-1} = x_i, X_k = x_j, X_{k+1} = x_l) P(X_{k-1} = x_i, X_k = x_j, X_{k+1} = x_l)$$
$$= P(Y_1, ..., Y_{k-1} | X_{k-1} = x_i) P(Y_k | X_k = x_j) P(Y_{k+1} | X_{k+1} = x_l) P(Y_{k+2}, ..., Y_t | X_{k+1} = x_l) \times$$
$$P(X_{k+1} = x_l | X_k = x_j) P(X_k = x_j | X_{k-1} = x_i) P(X_{k-1} = x_i)$$
$$= \frac{P(Y_1, ..., Y_{k-1}, X_{k-1} = x_i)}{P(X_{k-1} = x_i)} P(Y_k | X_k = x_j) P(Y_{k+1} | X_{k+1} = x_l) P(Y_{k+2}, ..., Y_t | X_{k+1} = x_l) \times$$
$$P(X_{k+1} = x_l | X_k = x_j) P(X_k = x_j | X_{k-1} = x_i) P(X_{k-1} = x_i)$$

Thus,

$$P(X_{k-1} = x_i, X_k = x_j, X_{k+1} = x_l | Y_1, ...Y_t)$$
$$= \frac{\alpha_{k-1}(x) M_{jl} M_{l(l+1)} \beta_{k+1}(x) T_{jl} T_{ij}}{P(Y_1, ..., Y_t)}$$

where the denominator: $P(Y_1, ..., Y_t) = \sum_x \alpha_t(x)$

Solution 4 (Time spent: 3 hour)**.** (a)

$$\hat{X} = a_1Y_1 + a_2Y_2 = (a_1 + 2a_2)X + a_1e_1 + a_2e_2$$
$$E[\hat{X}] = E[(a_1 + 2a_2)X + a_1e_1 + a_2e_2] = (a_1 + 2a_2)X$$

To make sure the estimate $\hat{X}$ is unbiased, we need $E[\hat{X}] = X$, thus

$$(a_1 + 2a_2)X = X$$
$$a_1 = 1 - 2a_2$$

We have the MSE

$$\begin{aligned}
E[(X - \hat{X})^2] &= E[(X - (a_1 + 2a_2)X - a_1e_1 - a_2e_2)^2] \\
&= E[(-(1 - 2a_2)e_1 - a_2e_2)^2] \\
&= (1 - 2a_2)^2\sigma_1^2 + a_2^2\sigma_2^2 \\
&= (1 - 4a_2 + 4a_2^2)\sigma_1^2 + a_2^2\sigma_2^2
\end{aligned}$$

To minimize this MSE, we derive the above equation and find the values of a_1 and a_2 that make this first-order derivative equal 0

$$\frac{dE[(X - \hat{X})^2]}{da_2} = (-4 + 8a_2)\sigma_1^2 + 2a_2\sigma_2^2 = 0$$
$$-4\sigma_1^2 + 2a_2(4\sigma_1^2 + \sigma_2^2) = 0$$

$$\frac{d^2E[(X - \hat{X})^2]}{da_2^2} = 2(4\sigma_1^2 + \sigma_2^2) > 0$$

$$\begin{aligned}
a_2 &= \frac{2\sigma_1^2}{4\sigma_1^2 + \sigma_2^2} \\
a_1 &= 1 - \frac{4\sigma_1^2}{4\sigma_1^2 + \sigma_2^2} = \frac{\sigma_2^2}{4\sigma_1^2 + \sigma_2^2} \\
\hat{X} &= \frac{\sigma_2^2}{4\sigma_1^2 + \sigma_2^2}Y_1 + \frac{2\sigma_1^2}{4\sigma_1^2 + \sigma_2^2}Y_2
\end{aligned}$$

This estimator has variance

$$\begin{aligned}
\sigma_{\hat{X}}^2 &= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 \\
&= \frac{\sigma_2^4}{(4\sigma_1^2 + \sigma_2^2)^2}\sigma_1^2 + \frac{4\sigma_1^4}{(4\sigma_1^2 + \sigma_2^2)^2}\sigma_2^2 \\
&= \frac{\sigma_1^2\sigma_2^2}{4\sigma_1^2 + \sigma_2^2}
\end{aligned}$$

(b)

(i) If $\sigma_2 >> \sigma_1$ then the measurement Y_2 gets less weight, and the estimator mathematically mainly depends on Y_1. This is correct with the intuition that if this measurement is too noisy, we should rely less upon it.

(ii) If $\sigma_2 = \sigma_1$, the 2 measurements have the same variance of noise distribution, and $\hat{X} = \frac{1}{5}Y_1 + \frac{2}{5}Y_2$. This means Y_2 is weighted twice Y_1. We would also have $\sigma_{\hat{X}^2} = \sigma_1^2/5$.

(iii) If $\sigma_1 >> \sigma_2$, this is similar to (i), and we would rely less on the measurement Y_1 because of the much larger noise variance of this measurement.