

Optimal Estimation homework 2

Li Yude 24S053054

December 4, 2024

1. Casino example

Assume the transition matrix is

$$A = \begin{bmatrix} & F & L \\ F & 0.95 & 0.05 \\ L & 0.1 & 0.9 \end{bmatrix}$$

and the emission probability matrix is

$$B = \left[\begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline F & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ L & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{2} \end{array} \right]$$

in which “ F ” denotes fair die and “ L ” represents loaded die. Denote Y as the number of dies, and X the status of the die, i.e., $X = 0$ means loaded and $X = 1$ indicates fair. If we have the observation $Y = \{6, 2, 6\}$, use maximum likelihood and maximum a posteriori estimate to estimate the status of the die.

Solution

(1) maximum likelihood:

$$\begin{aligned} P(Y = \{6, 2, 6\} | X = \{1, 1, 1\}) &= \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{216} \\ P(Y = \{6, 2, 6\} | X = \{0, 1, 1\}) &= \frac{1}{2} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{72} \\ P(Y = \{6, 2, 6\} | X = \{1, 0, 1\}) &= \frac{1}{6} \times \frac{1}{10} \times \frac{1}{6} = \frac{1}{360} \\ P(Y = \{6, 2, 6\} | X = \{1, 1, 0\}) &= \frac{1}{6} \times \frac{1}{6} \times \frac{1}{2} = \frac{1}{72} \\ P(Y = \{6, 2, 6\} | X = \{0, 0, 1\}) &= \frac{1}{2} \times \frac{1}{10} \times \frac{1}{6} = \frac{1}{120} \\ P(Y = \{6, 2, 6\} | X = \{1, 0, 0\}) &= \frac{1}{6} \times \frac{1}{10} \times \frac{1}{2} = \frac{1}{120} \\ P(Y = \{6, 2, 6\} | X = \{0, 1, 0\}) &= \frac{1}{2} \times \frac{1}{6} \times \frac{1}{2} = \frac{1}{24} \\ P(Y = \{6, 2, 6\} | X = \{0, 0, 0\}) &= \frac{1}{2} \times \frac{1}{10} \times \frac{1}{2} = \frac{1}{40} \end{aligned}$$

So, $X_{ML} = \{0, 1, 0\}$

(2) maximum a posteriori estimate:

For transition matrix

$$A = \begin{bmatrix} F(1) & L(0) \\ F(1) & 0.95 & 0.05 \\ L(0) & 0.1 & 0.9 \end{bmatrix}$$

we can calculate the probability of each state by:

$$\lim_{n \rightarrow \infty} A^n = \lim_{n \rightarrow \infty} \begin{bmatrix} 0.95 & 0.05 \\ 0.1 & 0.9 \end{bmatrix}^n = \lim_{n \rightarrow \infty} \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0.85 \end{bmatrix}^n \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

So we have

$$P(X = 0) = \frac{1}{3}, P(X = 1) = \frac{2}{3}$$

Consider the posteriori estimate:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \propto P(Y|X)P(X)$$

So, we have

$$\begin{aligned} P(X = \{1, 1, 1\} | Y = \{6, 2, 6\}) &\propto P(Y = \{6, 2, 6\} | X = \{1, 1, 1\})P(X = \{1, 1, 1\}) \\ &= \frac{1}{216} \times \frac{2}{3} \times 0.95 \times 0.95 = \frac{19}{129600} \end{aligned}$$

$$\begin{aligned} P(X = \{0, 1, 1\} | Y = \{6, 2, 6\}) &\propto P(Y = \{6, 2, 6\} | X = \{0, 1, 1\})P(X = \{0, 1, 1\}) \\ &= \frac{1}{72} \times \frac{1}{3} \times 0.1 \times 0.95 = \frac{10}{43200} \end{aligned}$$

$$\begin{aligned} P(X = \{1, 0, 1\} | Y = \{6, 2, 6\}) &\propto P(Y = \{6, 2, 6\} | X = \{1, 0, 1\})P(X = \{1, 0, 1\}) \\ &= \frac{1}{360} \times \frac{2}{3} \times 0.05 \times 0.1 = \frac{1}{108000} \end{aligned}$$

$$\begin{aligned} P(X = \{1, 1, 0\} | Y = \{6, 2, 6\}) &\propto P(Y = \{6, 2, 6\} | X = \{1, 1, 0\})P(X = \{1, 1, 0\}) \\ &= \frac{1}{72} \times \frac{2}{3} \times 0.95 \times 0.05 = \frac{19}{43200} \end{aligned}$$

$$\begin{aligned} P(X = \{0, 0, 1\} | Y = \{6, 2, 6\}) &\propto P(Y = \{6, 2, 6\} | X = \{0, 0, 1\})P(X = \{0, 0, 1\}) \\ &= \frac{1}{120} \times \frac{1}{3} \times 0.9 \times 0.1 = \frac{1}{4000} \end{aligned}$$

$$\begin{aligned} P(X = \{1, 0, 0\} | Y = \{6, 2, 6\}) &\propto P(Y = \{6, 2, 6\} | X = \{1, 0, 0\})P(X = \{1, 0, 0\}) \\ &= \frac{1}{120} \times \frac{2}{3} \times 0.05 \times 0.9 = \frac{1}{4000} \end{aligned}$$

$$\begin{aligned}
P(X = \{0, 1, 0\} | Y = \{6, 2, 6\}) &\propto P(Y = \{6, 2, 6\} | X = \{0, 1, 0\}) P(X = \{0, 1, 0\}) \\
&= \frac{1}{24} \times \frac{1}{3} \times 0.1 \times 0.05 = \frac{1}{14400}
\end{aligned}$$

$$\begin{aligned}
P(X = \{0, 0, 0\} | Y = \{6, 2, 6\}) &\propto P(Y = \{6, 2, 6\} | X = \{0, 0, 0\}) P(X = \{0, 0, 0\}) \\
&= \frac{1}{40} \times \frac{1}{3} \times 0.9 \times 0.9 = \frac{27}{4000}
\end{aligned}$$

So, $X_{MAP} = \{0, 0, 0\}$

□

2. Different Estimates

(1) Suppose that $\mathbf{z} = s + v$, where s and v are independent, jointly distributed RVs with $s \sim N(\eta, \sigma^2)$ and $v \sim N(0, V^2)$.

(a) Derive an expression for $E[s | \mathbf{z} = z]$.

(b) Derive an expression for $E[s^2 | \mathbf{z} = z]$.

(2) Suppose that $\mathbf{z} = s + v$, where s and v are independent, jointly distributed RVs with $s \sim N(\eta_s, \sigma_s^2)$ and $v \sim N(0, \sigma_v^2)$. Assume we have measurements $z(1), \dots, z(n)$

(a) Derive the maximum likelihood estimate for s ;

(b) Derive the maximum a posteriori estimate for s ;

(c) Derive the minimum mean square estimate for s ;

(d) Derive the linear minimum mean square estimate for s ;

(e) Derive the least squares estimate for s provided measurements $z(1), \dots, z(n)$;

(f) Suppose at each time $k (k \in \{1, \dots, n\})$, there is a new measurement $z(k)$, derive the recursive least squares estimate for s . (Assume $\hat{s}_0 = E(s)$, the initial error covariance is P_0);

(g) Compare all these 6 kinds of estimates.

Solution

(1)(a) We have $\mathbf{z} \sim N(\eta, \sigma^2 + V^2)$

and

$$f_s(s | \mathbf{z} = z) = \frac{f_z(z | s) f_s(s)}{f_z(z)}$$

since $z = s + v$, and s and v are independent, so $f_z(z | s)$ is the pdf of $v = z - s$

since $v \sim N(0, V^2)$, we have

$$f_z(z | s) = \frac{1}{\sqrt{2\pi V^2}} \exp \left[-\frac{(z - s)^2}{2V^2} \right]$$

on the other hand, $s \sim N(\eta, \sigma^2)$

$$f_s(s) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(s - \eta)^2}{2\sigma^2} \right]$$

$$\mathbf{z} \sim N(\eta, \sigma^2 + V^2),$$

$$f_z(z) = \frac{1}{\sqrt{2\pi(\sigma^2 + V^2)}} \exp \left[-\frac{(z - \eta)^2}{2(\sigma^2 + V^2)} \right]$$

so we have,

$$\begin{aligned} f_s(s|\mathbf{z} = z) &= \frac{f_z(z|s)f_s(s)}{f_z(z)} \\ &= \frac{\frac{1}{\sqrt{2\pi V^2}} \exp \left[-\frac{(z-s)^2}{2V^2} \right] \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(s-\eta)^2}{2\sigma^2} \right]}{\frac{1}{\sqrt{2\pi(\sigma^2 + V^2)}} \exp \left[-\frac{(z-\eta)^2}{2(\sigma^2 + V^2)} \right]} \\ &= \frac{1}{\sqrt{2\pi} \sqrt{\frac{\sigma^2 V^2}{\sigma^2 + V^2}}} \exp \left[-\frac{\left(s - \eta - \frac{\sigma^2}{V^2 + \sigma^2} (z - \eta) \right)^2}{2 \frac{\sigma^2 V^2}{\sigma^2 + V^2}} \right] \end{aligned}$$

Notice that it's a normal distribution. So,

$$E[s|\mathbf{z} = z] = \eta + \frac{\sigma^2}{V^2 + \sigma^2} (z - \eta)$$

(b)

$$\begin{aligned} E[s^2|\mathbf{z} = z] &= D[s|\mathbf{z} = z] + (E[s|\mathbf{z} = z])^2 \\ &= \frac{\sigma^2 V^2}{\sigma^2 + V^2} + \left(\eta + \frac{\sigma^2}{V^2 + \sigma^2} (z - \eta) \right)^2 \end{aligned}$$

(2)(a) maximum likelihood estimate:

$$f_z(z = z(i)|s) = \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp \left(-\frac{(z(i) - s)^2}{2\sigma_v^2} \right)$$

So, we have likelihood function:

$$f_z(z = z(1), z(2), \dots, z(n)|s) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp \left(-\frac{(z(i) - s)^2}{2\sigma_v^2} \right)$$

Find s that maximizes the likelihood function:

$$\begin{aligned} \ln f_z &= -n \ln(\sqrt{2\pi\sigma_v^2}) + \sum_{i=1}^n -\frac{(z(i) - s)^2}{2\sigma_v^2} \\ \frac{\partial \ln f_z}{\partial s} &= \sum_{i=1}^n \frac{(z(i) - s)}{\sigma_v^2} = 0 \end{aligned}$$

So we have

$$\hat{s}_{ML} = \frac{1}{n} \sum_{i=1}^n z(i)$$

(b) maximum a posteriori estimate:

$$f_s(s|z = z(1), \dots, z(n)) \propto f_z(z = z(1), \dots, z(n)|s) \times f(s) \\ = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left(-\frac{(z(i) - s)^2}{2\sigma_v^2}\right) \times \frac{1}{\sqrt{2\pi\sigma_s^2}} \exp\left(-\frac{(s - \eta_s)^2}{2\sigma_s^2}\right)$$

Find s that maximizes the posteriori estimate:

$$\ln f = -n \ln(\sqrt{2\pi\sigma_v^2}) + \sum_{i=1}^n -\frac{(z(i) - s)^2}{2\sigma_v^2} - \ln(\sqrt{2\pi\sigma_s^2}) - \frac{(s - \eta_s)^2}{2\sigma_s^2} \\ \frac{\partial \ln f}{\partial s} = \sum_{i=1}^n \frac{(z(i) - s)}{\sigma_v^2} + \frac{s - \eta_s}{\sigma_s^2} = 0$$

So, we have

$$\hat{s}_{MAP} = \frac{\sigma_v^2}{n\sigma_v^2 + \sigma_s^2} \sum_{i=1}^n z(i) + \frac{\sigma_s^2}{n\sigma_v^2 + \sigma_s^2} \eta_s$$

(c)

$$\hat{s}_{MMSE} = E[s|z_1 = z(1), \dots, z_n = z(n)]$$

$$f(s = s, z_1 = z(1), \dots, z_n = z(n)) = f(s = s, v_1 = z(1) - s, \dots, v_n = z(n) - s)$$

since s and v are independent, we define $\mathbf{v} = [v_1, \dots, v_n]$, we have

$$\begin{bmatrix} s \\ \mathbf{v} \end{bmatrix} \sim N\left(\begin{bmatrix} \eta_s \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_s^2 & 0 \\ 0 & \sigma_v^2 I \end{bmatrix}\right)$$

$$f(s = s, v_1 = z(1) - s, \dots, v_n = z(n) - s) \\ = \frac{1}{(2\pi)^{\frac{n+1}{2}} \sigma_s \sigma_v^n} \exp\left(\frac{-1}{2} \begin{bmatrix} (s - \eta_s)^T & (\mathbf{z} - s)^T \end{bmatrix} \begin{bmatrix} \sigma_s^2 & 0 \\ 0 & \sigma_v^2 I \end{bmatrix}^{-1} \begin{bmatrix} s - \eta_s \\ \mathbf{z} - s \end{bmatrix}\right)$$

On the other hand, $z_i \sim N(\eta_s, \sigma_s^2 + \sigma_v^2)$, $\mathbf{z} \sim N(\eta_s, (\sigma_s^2 + \sigma_v^2)I)$ so,

$$f(\mathbf{z} = z(1), \dots, z(n)) = \frac{1}{(2\pi)^{\frac{n}{2}} (\sigma_s + \sigma_v)^n} \exp\left(\frac{-1}{2} [\mathbf{z} - \eta_s]^T [(\sigma_v^2 + \sigma_s^2)I]^{-1} [\mathbf{z} - \eta_s]\right)$$

So, by bayes equation,

$$f(s|z_1 = z(1), \dots, z_n = z(n)) = \frac{f(s = s, z_1 = z(1), \dots, z_n = z(n))}{f(z_1 = z(1), \dots, z_n = z(n))} \\ = \frac{(\sigma_s + \sigma_v)^n}{(2\pi)^{\frac{1}{2}} \sigma_s \sigma_v^n} \exp(As^2 + Bs + C)$$

where,

$$A = -\frac{n\sigma_s^2 + \sigma_v^2}{2\sigma_s^2 \sigma_v^2}, B = -\frac{\sigma_s^2 \sum_{i=1}^n z(i) + \sigma_v^2 \eta_s}{\sigma_s^2 \sigma_v^2}$$

Notice that it's a Gaussian distribution, we have

$$E(s|z_1 = z(1), \dots, z_n = z(n)) = \frac{\sigma_s^2 \sum_{i=1}^n z(i) + \sigma_v^2 \eta_s}{n\sigma_s^2 + \sigma_v^2}$$

So, we can obtain,

$$\hat{s}_{MMSE} = \eta_s + \frac{\sigma_s^2 \sum_{i=1}^n (z(i) - \eta_s)}{n\sigma_s^2 + \sigma_v^2}$$

(d) here, we donate $\mathbf{z} = [z(1), \dots, z(n)]$, we have

$$\hat{s}_{LMMSE} = E[s\mathbf{z}^T][E(\mathbf{z}\mathbf{z}^T)]^{-1}\mathbf{z}$$

consider

$$E[sz(i)] = E[s(s + v(i))] = E(s^2) + E(s)E(v(i)) = (D(s) + E(s)^2) + 0 = \eta_s^2 + \sigma_s^2$$

on the other hand,

$$\begin{aligned} E[z(i)z(j)] &= E[(s + v(i))(s + v(j))] \\ &= E(s^2) + E(s)E(v(i) + v(j)) + E(v(i)v(j)) \\ &= \eta_s^2 + \sigma_s^2 + 0 + E(v(i)v(j)) \\ E(v(i)v(j)) &= \begin{cases} 0 & \text{if } i \neq j \\ \sigma_v^2 & \text{if } i = j \end{cases} \end{aligned}$$

So we have,

$$E[s\mathbf{z}^T] = [E(sz(1)), \dots, E(sz(n))]^T = [\eta_s^2 + \sigma_s^2, \dots, \eta_s^2 + \sigma_s^2]$$

$$\begin{aligned} E(\mathbf{z}\mathbf{z}^T) &= \begin{bmatrix} E[z(1)z(1)] & E[z(1)z(2)] & \dots & E[z(1)z(n)] \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & E[z(n)z(n)] \end{bmatrix} \\ &= \begin{bmatrix} \eta_s^2 + \sigma_s^2 + \sigma_v^2 & \eta_s^2 + \sigma_s^2 & \dots & \eta_s^2 + \sigma_s^2 \\ \eta_s^2 + \sigma_s^2 & \eta_s^2 + \sigma_s^2 + \sigma_v^2 & \dots & \eta_s^2 + \sigma_s^2 \\ \dots & \dots & \dots & \dots \\ \eta_s^2 + \sigma_s^2 & \eta_s^2 + \sigma_s^2 & \dots & \eta_s^2 + \sigma_s^2 + \sigma_v^2 \end{bmatrix} \end{aligned}$$

$$E(\mathbf{z}\mathbf{z}^T)^{-1} = \frac{1}{\sigma_v^2[\sigma_v^2 + n(\sigma_s^2 + \eta_s^2)]} \begin{bmatrix} \sigma_v^2 + (n-1)(\sigma_s^2 + \eta_s^2) & \dots & -(\sigma_s^2 + \eta_s^2) \\ \vdots & \ddots & \vdots \\ -(\sigma_s^2 + \eta_s^2) & \dots & \sigma_v^2 + (n-1)(\sigma_s^2 + \eta_s^2) \end{bmatrix}$$

$$\begin{aligned} E[s\mathbf{z}^T][E(\mathbf{z}\mathbf{z}^T)]^{-1} &= \frac{\sigma_s^2 + \eta_s^2}{\sigma_v^2[\sigma_v^2 + n(\sigma_s^2 + \eta_s^2)]} [\sigma_v^2, \sigma_v^2, \dots, \sigma_v^2] \\ &= \frac{\sigma_s^2 + \eta_s^2}{\sigma_v^2 + n(\sigma_s^2 + \eta_s^2)} [1, \dots, 1] \end{aligned}$$

So,

$$\hat{s}_{LMMSE} = \frac{\sigma_s^2 + \eta_s^2}{\sigma_v^2 + n(\sigma_s^2 + \eta_s^2)} \sum_{i=1}^n z(i)$$

(e) we have

$$\begin{bmatrix} z(1) \\ z(2) \\ \dots \\ z(n) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix} s$$

LS estimate is given by

$$\hat{s}_{LS} = (H^T H)^{-1} H^T [z(1), \dots, z(N)]^T = \frac{1}{N} \sum_{i=1}^n z(i)$$

(f)

$$z_k = s + v_k, \hat{s}_0 = E(s) = \eta_s, H_k = 1, R_k = D(v_k) = \sigma_v^2$$

Iteration:

$$\begin{cases} k_k = P_{k-1} H_k^T (H_k P_{k-1} H_k^T + R_k)^{-1} = P_{k-1} (P_{k-1} + R_k)^{-1} \\ \hat{s}_k = \hat{s}_{k-1} + k_k (z_k - H_k \hat{s}_{k-1}) = (1 - k_k) \hat{s}_{k-1} + k_k z_k \\ P_k = (I - k_k H_k) P_{k-1} (I - k_k H_k)^T + k_k P_{k-1} k_k^T = P_{k-1} (1 - k_k)^2 + \sigma_v^2 k_k^2 \end{cases}$$

since $\varepsilon_{s,0} = s - \eta_s \sim N(0, \sigma_s^2)$, we have $P_0 = \sigma_s^2$. By iteration, we can have

$$\begin{cases} k_k = \frac{\sigma_s^2}{n\sigma_s^2 + \sigma_v^2} \\ \hat{s}_n = \frac{\sigma_s^2 \sum_{i=1}^n z(i) + \sigma_v^2 \eta_s}{n\sigma_s^2 + \sigma_v^2} \\ P_n = \frac{\sigma_s^2 \sigma_v^2}{n\sigma_s^2 + \sigma_v^2} \end{cases}$$

So we can concluded that

$$\hat{s}_{RLS} = \frac{\sigma_s^2 \sum_{i=1}^n z(i) + \sigma_v^2 \eta_s}{n\sigma_s^2 + \sigma_v^2} = \eta_s + \frac{\sigma_s^2 \sum_{i=1}^n (z(i) - \eta_s)}{n\sigma_s^2 + \sigma_v^2}$$

(g)

- Maximum Likelihood Estimate (MLE):
It maximizes the likelihood function based on data only, ignoring prior knowledge.
Advantage: Simple concept and easy to calculate in many cases.
Disadvantage: Lacks consideration of prior info.
- Maximum a Posteriori (MAP) Estimate:
Combines data likelihood and parameter's prior distribution.
Advantage: Utilizes prior knowledge.
Disadvantage: Depends on prior choice.
- Minimum Mean Square Estimate (MMSE):
Minimizes mean squared error using posterior distribution.
Advantage: Optimal for mean squared error.
Disadvantage: Computationally complex.

- Linear Minimum Mean Square Estimate (LMMSE):
A linear form of MMSE, simpler in some cases.
Advantage: Computationally easier when linearity assumed.
Disadvantage: May be less accurate for non-linear relations.
- Least Squares Estimate:
Minimizes sum of squared differences between observed and predicted data.
Advantage: Easy to understand and implement.
Disadvantage: Sensitive to outliers.
- Recursive Least Squares Estimate:
Updates estimate as new data comes in.
Advantage: Saves computational resources.
Disadvantage: Depends on initial estimate and algorithm stability.

□