200320708 宋颖聪 20级自动化7班

1. Casino example

Assume the transition matrix is

$$A = \begin{array}{cc} & \begin{array}{cc} 1 & 0 \end{array} \\ & \begin{bmatrix} & F & L \\ 1\,F & 0.95 & 0.05 \\ 0\,L & 0.1 & 0.9 \end{bmatrix} \end{array}$$

and the emission probability matrix is

$$B = \begin{bmatrix} & 1 & 2 & 3 & 4 & 5 & 6 \\ F & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ L & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{2} \end{bmatrix}$$

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 $Y = \{6, 6, 6\}$, use maximum likelihood and maximum *a posteriori* estimate to estimate the status of the die.

解：

$$\begin{cases} P(X=0) = P(X=0)\cdot 0.9 + P(X=1)\cdot 0.05 \\ P(X=1) = P(X=0)\cdot 0.1 + P(X=1)\cdot 0.95 \\ P(X=0)+P(X=1)=1 \end{cases} \Rightarrow \begin{cases} P(X=0)=\frac{1}{3} \\ P(X=1)=\frac{2}{3} \end{cases}$$

① MLE maximize $P(Y=(6,6,6)\mid X)$

$P(Y=(6,6,6)\mid X=(0,0,0)) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{8}$ $\quad$ $P(Y=(6,6,6)\mid X=(0,0,1)) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{6} = \frac{1}{24}$

$P(Y=(6,6,6)\mid X=(0,1,0)) = \frac{1}{2}\cdot\frac{1}{6}\cdot\frac{1}{2} = \frac{1}{24}$ $\quad$ $P(Y=(6,6,6)\mid X=(0,1,1)) = \frac{1}{2}\cdot\frac{1}{6}\cdot\frac{1}{6} = \frac{1}{72}$

$P(Y=(6,6,6)\mid X=(1,0,0)) = \frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{24}$ $\quad$ $P(Y=(6,6,6)\mid X=(1,0,1)) = \frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{6} = \frac{1}{72}$

$P(Y=(6,6,6)\mid X=(1,1,0)) = \frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{2} = \frac{1}{72}$ $\quad$ $P(Y=(6,6,6)\mid X=(1,1,1)) = \frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6} = \frac{1}{216}$

So $\hat{X}_{ML} = (0,0,0)$

② MAP maximize $P(X\mid Y=(6,6,6)) = \dfrac{P(X, Y=(6,6,6))}{P(Y=(6,6,6))} = \dfrac{P(Y=(6,6,6)\mid X)\cdot P(X)}{P(Y=(6,6,6))}$

$P(X=(0,0,0)) = P(X=0)\cdot 0.9\cdot 0.9 = 0.27$ $\quad$ $\therefore P(X=(0,0,0), Y=(6,6,6)) = \frac{1}{8}\cdot 0.27 = \frac{27}{800} = 0.03375$

$P(X=(0,0,1)) = P(X=0)\cdot 0.9\cdot 0.1 = 0.03$ $\quad$ $\therefore P(X=(0,0,1), Y=(6,6,6)) = \frac{1}{24}\cdot 0.03 = \frac{1}{800}$

$P(X=(0,1,0)) = P(X=0)\cdot 0.1\cdot 0.05 = \frac{1}{600}$ $\quad$ $\therefore P(X=(0,1,0), Y=(6,6,6)) = \frac{1}{24}\cdot\frac{1}{600} = \frac{1}{14400}$

$P(X=(0,1,1)) = P(X=0)\cdot 0.1\cdot 0.95 = \frac{19}{600}$ $\quad$ $\therefore P(X=(0,1,1), Y=(6,6,6)) = \frac{1}{72}\cdot\frac{19}{600} = \frac{19}{43200}$

$P(X=(1,0,0)) = P(X=1)\cdot 0.05\cdot 0.9 = 0.03$ $\quad$ $\therefore P(X=(1,0,0), Y=(6,6,6)) = \frac{1}{24}\cdot 0.03 = \frac{1}{800}$

$P(X=(1,0,1)) = P(X=1)\cdot 0.05\cdot 0.1 = \frac{1}{300}$ $\quad$ $\therefore P(X=(1,0,1), Y=(6,6,6)) = \frac{1}{72}\cdot\frac{1}{300} = \frac{1}{21600}$

$P(X=(1,1,0)) = P(X=1)\cdot 0.95\cdot 0.05 = \frac{19}{600}$ $\quad$ $\therefore P(X=(1,1,0), Y=(6,6,6)) = \frac{1}{72}\cdot\frac{19}{600} = \frac{19}{43200}$

$P(X=(1,1,1)) = P(X=1)\cdot 0.95\cdot 0.95 = \frac{361}{600}$ $\quad$ $\therefore P(X=(1,1,1), Y=(6,6,6)) = \frac{1}{216}\cdot\frac{361}{600} = \frac{361}{129600} \approx 2.79\times 10^{-3}$

So $\hat{X}_{MAP} = (0,0,0)$

2. Different estimates

(1) Suppose that $\boldsymbol{z} = \boldsymbol{s} + \boldsymbol{v}$, where $\boldsymbol{s}$ and $\boldsymbol{v}$ are independent, jointly distributed RVs with $\boldsymbol{s} \sim \mathcal{N}(\eta, \sigma^2)$ and $\boldsymbol{v} \sim \mathcal{N}(0, V^2)$.
   (a) Derive an expression for $E[\boldsymbol{s}|\boldsymbol{z} = z]$.
   (b) Derive an expression for $E[\boldsymbol{s}^2|\boldsymbol{z} = z]$.

(2) Suppose that $\boldsymbol{z} = \boldsymbol{s} + \boldsymbol{v}$, where $\boldsymbol{s}$ and $\boldsymbol{v}$ are independent, jointly distributed RVs with $\boldsymbol{s} \sim \mathcal{N}(\eta_s, \sigma_s^2)$ and $\boldsymbol{v} \sim \mathcal{N}(0, \sigma_v^2)$. Assume we have measurements $\boldsymbol{z}(1), \ldots, \boldsymbol{z}(n)$,
   (a) Derive the maximum likelihood estimate for $\boldsymbol{s}$;
   (b) Derive the maximum *a posteriori* estimate for $\boldsymbol{s}$;
   (c) Derive the minimum mean square estimate for $\boldsymbol{s}$;
   (d) Derive the linear minimum mean square estimate for $\boldsymbol{s}$;
   (e) Derive the least squares estimate for $\boldsymbol{s}$ provided measurements $z(1), \ldots, z(n)$;
   (f) Suppose at each time k ($k \in \{1, \ldots, n\}$), there is a new measurement $z(k)$, derive the recursive least squares estimate for $\boldsymbol{s}$. (Assume $\hat{s}_0 = E(\boldsymbol{s})$, the initial error covariance is P_0);
   (g) Compare all these 6 kinds of estimates.

(1) (a) $E(z) = E(s+v) = \eta$

$Var(z) = Var(s+v) = Var(s) + Var(v) = \sigma^2 + V^2$

所以

$$f_z(z) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2+V^2}} e^{-\frac{1}{2}\frac{(z-\eta)^2}{\sigma^2+V^2}}$$

$$f(s|z) = \frac{f(s,z)}{f(z)}$$

$$Cov(s,z) = E(sz) - E(s)E(z)$$

$$E(sz) = E(s^2+sv) = E(s^2) + E(sv) = Var(s) + E^2(s) + Cov(sv) + E(s)E(v)$$
$$= \sigma^2 + \eta^2$$

所以 $Cov(s,z) = \sigma^2$

所以 $\Sigma = \begin{bmatrix} \sigma^2+v^2 & \sigma^2 \\ \sigma^2 & \sigma^2 \end{bmatrix}$

$$\therefore \quad f(s,z) = \frac{1}{2\pi\sqrt{|\Sigma|}} \cdot \exp\left\{-\frac{1}{2}\cdot [z \ s]\ \Sigma^{-1} \begin{bmatrix} z \\ s \end{bmatrix}\right\}$$
$$= \frac{1}{2\pi\sigma v} \exp\left\{-\left[\frac{(z-s)^2}{2v^2} + \frac{(s-\eta)^2}{2\sigma^2}\right]\right\}$$

$$\therefore \quad f(s|z) = \frac{f(s,z)}{f(z)} = \frac{1}{\sqrt{2\pi}\frac{\sigma v}{\sqrt{\sigma^2+v^2}}} \cdot \exp\left\{-\frac{1}{2} \frac{\left(s - \left[\eta + \frac{\sigma^2}{\sigma^2+v^2}(z-\eta)\right]\right)^2}{\frac{\sigma^2 v^2}{\sigma^2+v^2}}\right\}$$

$$\therefore \quad E(s|z) = \eta + \frac{\sigma^2}{\sigma^2+v^2}(z-\eta)$$

$$Var(s|z) = \frac{\sigma^2 v^2}{\sigma^2+v^2}$$

(b) $Var(s|z) = E(s^2|z) - E(s|z)^2$

$$\therefore E(s^2|z) = Var(s|z) + E(s|z)^2$$
$$= \frac{\sigma^2 v^2}{\sigma^2+v^2} + \frac{\sigma^4}{(\sigma^2+v^2)^2} z^2 + 2\frac{\sigma^2 v^2}{(\sigma^2+v^2)^2} z\eta + \frac{v^4}{(\sigma^2+v^2)^2}\eta^2$$

(2)

$z = s+v \quad E(z) = E(s)+E(v) = \eta_s \quad Var(z) = E(s)+E(v) = \sigma_s^2+\sigma_v^2 \quad E(z^2) = Var(z) + E(z)^2 = \sigma_s^2+\sigma_v^2+\eta_s^2$

$E(sz) = E(s^2+sv) = \sigma_s^2+\eta_s^2 \quad Cov(sz) = E(sz) - E(s)E(z) = \sigma_s^2$

令 $X = \begin{bmatrix} s \\ z_1 \\ \vdots \\ z_n \end{bmatrix}_{n\times 1} \quad \Sigma = \begin{bmatrix} \sigma_s^2 & \sigma_s^2 & \cdots & \sigma_s^2 \\ \sigma_s^2 & \sigma_s^2+\sigma_v^2 & & \\ \vdots & & \ddots & \\ \sigma_s^2 & & & \sigma_s^2+\sigma_v^2 \end{bmatrix}_{n+1} \quad \det\Sigma = (\sigma_s^2+\sigma_v^2)^n \cdot \left(\sigma_s^2 - \frac{n\sigma_s^4}{\sigma_s^2+\sigma_v^2}\right)$

$u = \begin{bmatrix} \eta_s \\ \vdots \\ \eta_s \end{bmatrix}_{n\times 1}$

$$\therefore \quad f(z_1,z_2,\cdots,z_n,s) = \frac{1}{(\sqrt{2\pi})^{n+1}(\det\Sigma)^{\frac{1}{2}}} \cdot \exp\left\{-\frac{1}{2}\cdot(X-u)^T\Sigma^{-1}(X-u)\right\}$$

$$f(s) = \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left\{-\frac{1}{2}\frac{(s-\eta_s)^2}{\sigma_s^2}\right\}$$

$$f(z_1,z_2,\cdots,z_n) = \frac{1}{(\sqrt{2\pi})^n(\sigma_s^2+\sigma_v^2)^{\frac{n}{2}}} \exp\left\{-\sum_{i=1}^{n}\frac{(z_i-\eta_s)^2}{2\sigma_s^2+2\sigma_v^2}\right\}$$

(a) Maximizes $f(z_1,z_2,\cdots,z_n|s)$

$$f(z_1,z_2,\cdots,z_n|s) = \frac{f(z_1,z_2,\cdots,z_n,s)}{f(s)} = \frac{1}{(\sqrt{2\pi})^n(\sigma_v^2+\sigma_s^2)^{\frac{n}{2}}\sqrt{1-\frac{n\sigma_s^2}{\sigma_s^2+\sigma_v^2}}} \exp\left\{-\frac{1}{2}(X-u)^T\Sigma^{-1}(X-u) + \frac{1}{2}\frac{(s-\eta_s)^2}{\sigma_s^2}\right\}$$

$$\frac{\partial \ln f(z_1,z_2,\cdots,z_n|s)}{\partial s} = 0 \Rightarrow \frac{\partial\left[(X-u)^T\Sigma^{-1}(X-u) - \frac{(s-\eta_s)^2}{\sigma_s^2}\right]}{\partial s} = 0 \Rightarrow 2(X-u)^T\Sigma^{-1}\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix} - \frac{2(s-\eta_s)}{\sigma_s^2} = 0$$

$$\therefore \hat{s}_{ML} = \frac{1}{n}(z_1+z_2+\cdots+z_n)$$

(b) Maximizes $f(s|z_1,z_2,\cdots,z_n)$

$$f(s|z_1,z_2,\cdots,z_n) = \frac{f(z_1,z_2,\cdots,z_n,s)}{f(z_1,z_2,\cdots z_n)} \qquad \frac{\partial f(s|z_1,z_2,\cdots,z_n)}{\partial s} = 0 \Leftrightarrow \frac{\partial \ln f(z_1,z_2,\cdots,z_n,s)}{\partial s} = 0$$

$$\Leftrightarrow \frac{\partial\left[(X-u)^T\Sigma^{-1}(X-u)\right]}{\partial s} = 0$$

So $\quad 2(x-u)^T \cdot \Sigma^{-1} \cdot \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0$

$$\therefore \quad \hat{s}_{MAP} = \eta_s + \frac{\sigma_s^2(z_1+\cdots+z_n) - n\sigma_s^2\eta_s}{\sigma_s^2 + \sigma_v^2}$$

(c) $E(s \mid z_1, z_2, \cdots, z_n)$

$$f(s \mid z_1, z_2, \cdots, z_n) = \frac{f(z_1, z_2, \cdots, z_n, s)}{f(z_1, z_2, \cdots, z_n)}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{\left|\sigma_s^2 - \frac{n\sigma_s^4}{\sigma_s^2+\sigma_v^2}\right|}} \cdot \exp\left\{ -\frac{1}{2} \frac{\left[s - \eta_s - \frac{\sigma_s^2}{\sigma_s^2+\sigma_v^2}(z_1+\cdots+z_n - n\eta_s)\right]}{\sigma_s^2 - \frac{n\sigma_s^4}{\sigma_s^2+\sigma_v^2}} \right\}$$

$$\therefore \quad \hat{s}_{MMSE} = E(s \mid z_1, \cdots, z_n) = \eta_s + \frac{\sigma_s^2}{\sigma_s^2+\sigma_v^2}(z_1+\cdots+z_n - n\cdot\eta_s)$$

(d) $\hat{s} = \lambda z$

$z = \frac{1}{n}(z_1 + z_2 + \cdots + z_n)$

$E(sz) = \sigma_s^2 + \eta_s^2 \qquad E(z^2) = \sigma_s^2 + \sigma_v^2 + \eta_s^2$

$\therefore$

$$\hat{s}_{LMSE} = \frac{E(sz)}{E(z^2)} \cdot z$$

$$= \frac{\sigma_s^2 + \eta_s^2}{\sigma_s^2 + \sigma_v^2 + \eta_s^2} \cdot \frac{1}{n}(z_1 + z_2 + \cdots + z_n)$$

(e)

$$\hat{s}_{LS} = (H^TH)^{-1}H^T[z_1, \cdots, z_n]^T = \frac{1}{n}(z_1 + z_2 + \cdots + z_n)$$

(f)
$$\begin{cases} K_k = P_{k-1}H_k^T(H_kP_{k-1}H_k^T + R_k)^{-1} \\ \hat{x}_k = \hat{x}_{k-1} + K_k(y_k - H_k\hat{x}_{k-1}) \\ P_k = (I - K_kH_k)P_{k-1}(I - K_kH_k)^T \end{cases}$$

Assume $\quad R_k = R$

$\hat{s}_0 = E(s) = \eta_s \qquad H_k = 1$

$K=1$
$$\begin{cases} K_1 = P_0(P_0+R)^{-1} \\ \hat{s}_1 = \hat{s}_0 + \frac{P_0}{P_0+R}(z_1 - \hat{s}_0) \\ P_1 = \frac{P_0R}{P_0+R} \end{cases}$$

$\xRightarrow{K=2}$
$$\begin{cases} K_2 = \frac{P_1}{P_1+R} = \frac{P_0}{2P_0+R} \\ P_2 = \frac{P_1R}{P_1+R} = \frac{P_0R}{2P_0+R} \\ \hat{s}_2 = \hat{s}_1 + \frac{P_1}{P_1+R}(z_2 - \hat{s}_1) \end{cases}$$

$$= \frac{P_0+R}{2P_0+R}\hat{s}_1 + \frac{P_0}{2P_0+R}z_2$$

$\xRightarrow{\cdots}$

$\xRightarrow{K=n}$
$$\begin{cases} P_{n-1} = \frac{P_0R}{(n-1)P_0+R} \\ K_n = \frac{P_0}{nP_0+R} \\ \hat{s}_n = \hat{s}_{n-1} + K_n(z_n - \hat{s}_{n-1}) \end{cases}$$

$$= \frac{(n-1)P_0+R}{nP_0+R}\hat{s}_{n-1} + \frac{P_0}{nP_0+R}z_n$$

(g) MLE：无需先验信息，在小样本情况下可能不稳定

MAP：充分考虑先验信息，对于估计问题的不确定性有更好处理

MMSE：考虑估计误差的平方，适用于线性和非线性

LMMSE：原理基于MMSE，但仅适用于线性模型

LSE：寻找一个参数估计，使观测数据与模型的预测值的平方误差之和最小，适用于回归分析和曲线拟合等问题

RLSE：一种递归估计方法，基于LSE的原理，逐步更新参数估计以适应动态系统。