227-0689-00L System Identification 4 October 2023

Solution 1: Data fit statistics & ML/MAP estimators

Problem 1

Consider a quadrotor on a flight that is trying to estimate the position of the landing site to land. The site is clearly marked with a visual marker and the onboard estimation module of the drone is using two vision based algorithms to estimate the x-coordinate of the site. The two independent measurements can be modelled as $x_1 \sim \mathcal{N}(\theta_x, \sigma_1^2)$ and $x_2 \sim \mathcal{N}(\theta_x, \sigma_2^2)$. Assume the y-coordinate of the site to be perfectly known at this point.

1. Given two measurements x_1 and x_2 from each sensor respectively, provide an expression for the Maximum Likelihood Estimate (MLE) of $\hat{\theta}_x$.

The quadrotor is now given access to a previously conducted accurate Simultaneous Localisation and Mapping (SLAM) environment which provides a more accurate position of the landing site. In particular $\theta_x \sim \mathcal{N}(\mu_0, \sigma_3^2)$.

2. Using the SLAM estimate as a prior, derive expressions for the posterior distribution of θ_x and the maximum *a posteriori* (MAP) estimate $\hat{\theta}_{\text{MAP}}$.

After some developments, a new "grey" box method of estimating directly the complete position of the landing site $\mathbf{p}$ is provided as follows

$$\mathbf{z} = F(\theta)\mathbf{w},$$
$$\mathbf{p} = H(\theta)\mathbf{z} + \mathbf{e},$$

where $\mathbf{w}$ and $\mathbf{e}$ are two independent Gaussian random vectors with zero mean values and unit covariance matrices.

3. Find an expression for the maximum likelihood estimate of the parameter θ given the measurement $\mathbf{p}$. You only need to derive the optimization problem for obtaining the estimate.

Solution

1. Since the two measurements are independent and normally distributed, the joint probability density function is given by

$$f(x_1, x_2; \theta_x) = \prod_{i=1}^{2} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i - \theta)^2}{2\sigma_i^2}\right).$$

The maximum likelihood estimate is then given by

$$\hat{\theta}_x = \arg\max_{\theta_x} \ln f\left(x_1, x_2; \theta_x\right) = \frac{\sigma_1^{-2} x_1 + \sigma_2^{-2} x_2}{\sigma_1^{-2} + \sigma_2^{-2}}.$$

2. The posterior distribution of $\hat{\theta}_x$ is given by

$$\begin{aligned} p(\theta_x | x_1, x_2) &= \frac{p(\theta_x)\, p(x_1|\theta)\, p(x_2|\theta)}{p(x_1, x_2)} \\ &= \frac{\exp\left(-\frac{1}{2}\left(\left(\frac{x_1-\theta_x}{\sigma_1}\right)^2 + \left(\frac{x_2-\theta_x}{\sigma_2}\right)^2 + \left(\frac{\theta_x-\mu_0}{\sigma_1}\right)^2\right)\right)}{p(x_1, x_2)} \\ &= \frac{\exp\left(-\frac{1}{2}\left(\theta_x - \frac{\sigma_1^{-2}x_1+\sigma_2^{-2}x_2+\sigma_3^{-2}\mu_0}{\sigma_1^{-2}+\sigma_2^{-2}+\sigma_3^{-2}}\right)\right)}{\sigma_1^{-2} + \sigma_2^{-2} + \sigma_3^{-2}} \cdot \frac{1}{p(x_1, x_2)}. \end{aligned}$$

The MAP is then given as the value that maximizes the above expression

$$\hat{\theta}_{\text{MAP}} = \frac{\sigma_1^{-2} x_1 + \sigma_2^{-2} x_2 + \sigma_3^{-2} \mu_0}{\sigma_1^{-2} + \sigma_2^{-2} + \sigma_3^{-2}}.$$

3. Combining both equations, we have

$$\mathbf{p} = H(\theta) F(\theta) \mathbf{w} + \mathbf{e}.$$

Since $\mathbf{w}$ and $\mathbf{e}$ are independent and both subject to $\mathcal{N}\left(\mathbf{0}, I\right)$, we have

$$\mathbf{p}|\theta \sim \mathcal{N}\left(\mathbf{0}, R(\theta)\right),$$

where $R(\theta) = I + H(\theta)F(\theta)F^\top(\theta)H^\top(\theta)$. Substituting the density function of the multivariate normal distribution, the maximum likelihood estimate is given by

$$\begin{aligned} \hat{\theta}_{\text{ML}} &= \operatorname*{argmin}_\theta \; -\log\, p(\mathbf{p}|\theta) \\ &= \operatorname*{argmin}_\theta \; \frac{1}{2}\, \mathbf{p}^\top R^{-1}(\theta) \mathbf{p} + \frac{1}{2} \log\det R(\theta) \end{aligned}$$

Problem 2

1. Consider two random vectors X and Y that are jointly Gaussian. Given the parameters

$$\begin{aligned} &\mathcal{E}\{X\} = m_X,\ \mathcal{E}\{Y\} = m_Y, \\ &\mathcal{E}\left\{(X - m_X)(X - m_X)^\mathsf{T}\right\} = P_X,\ \mathcal{E}\left\{(Y - m_Y)(Y - m_Y)^\mathsf{T}\right\} = P_Y, \\ &\mathcal{E}\left\{(X - m_X)(Y - m_Y)^\mathsf{T}\right\} = P_{XY}, \end{aligned}$$

show that the conditional distribution of X given Y is

$$X \mid Y \sim \mathcal{N}\left(m_X + P_{XY} P_Y^{-1}(Y - m_Y), P_X - P_{XY} P_Y^{-1} P_{XY}^\mathsf{T}\right). \tag{1.1}$$

Hint: If D and $A - BD^{-1}C$ are invertible, the following equalities hold:

$$\det\left(\begin{bmatrix} A & B \\ C & D \end{bmatrix}\right) = \det(D)\det(A - BD^{-1}C), \tag{1.2}$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}. \tag{1.3}$$

2. Consider a static linear model:

$$y = \theta x + v.$$

We have obtained the prior distribution of θ: $\theta \sim \mathcal{N}(\mu, \sigma_\theta^2)$, the noise distribution $v \sim \mathcal{N}(0, \sigma_v^2)$, and a pair of measurements (x_0, y_0). Please use the result in part 1 to calculate the maximum a posteriori (MAP) estimate of θ.

Solution

1. The conditional density of X given $Y = y$ is given by

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x, y)}{f_Y(y)},$$

where

$$Y \sim \mathcal{N}(m_Y, P_Y), \quad \begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m_X \\ m_Y \end{bmatrix}, \begin{bmatrix} P_X & P_{XY} \\ P_{XY}^\mathsf{T} & P_Y \end{bmatrix}\right).$$

Let $P = \begin{bmatrix} P_X & P_{XY} \\ P_{XY}^\mathsf{T} & P_Y \end{bmatrix}$. Substituting the density function of the multivariate normal distribution, we have

$$\begin{aligned} f_{X|Y}(x \mid y) &= \frac{(2\pi)^{-(n_x+n_y)/2} \det(P)^{-1/2} \exp\left(-\frac{1}{2}\begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix}^\mathsf{T} P^{-1} \begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix}\right)}{(2\pi)^{-n_y/2} \det(P_Y)^{-1/2} \exp\left(-\frac{1}{2}(y - m_Y)^\mathsf{T} P_Y^{-1}(y - m_Y)\right)} \\ &= (2\pi)^{-\frac{n_x}{2}} \left(\frac{\det(P)}{\det(P_Y)}\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix}^\mathsf{T} P^{-1} \begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix} - (y - m_Y)^\mathsf{T} P_Y^{-1}(y - m_Y)\right), \end{aligned}$$

where n_x and n_y are the dimension of X and Y respectively. From (1.2), we have

$$\frac{\det(P)}{\det(P_Y)} = \det\left(P_X - P_{XY}P_Y^{-1}P_{XY}^\mathsf{T}\right).$$

Let $P_{X|Y} = P_X - P_{XY}P_Y^{-1}P_{XY}^\mathsf{T}$. From (1.3), we have

$$\begin{aligned} &\begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix}^\mathsf{T} P^{-1} \begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix} - (y - m_Y)^\mathsf{T} P_Y^{-1}(y - m_Y) \\ &= \begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix}^\mathsf{T} \begin{bmatrix} P_{X|Y}^{-1} & -P_{X|Y}^{-1}P_{XY}P_Y^{-1} \\ -P_Y^{-1}P_{XY}^\mathsf{T}P_{X|Y}^{-1} & P_Y^{-1}P_{XY}^\mathsf{T}P_{X|Y}^{-1}P_{XY}P_Y^{-1} \end{bmatrix} \begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix} \\ &= \left((x - m_X) - P_{XY}P_Y^{-1}(y - m_Y)\right)^\mathsf{T} P_{X|Y}^{-1} \left((x - m_X) - P_{XY}P_Y^{-1}(y - m_Y)\right) \end{aligned}$$

Let $m_{X|Y} = m_X + P_{XY} P_Y^{-1}(y - m_Y)$. We have

$$f_{X|Y}(x \mid y) = (2\pi)^{-\frac{n_x}{2}} \det\left(P_{X|Y}\right)^{-1/2} \exp\left(-\frac{1}{2}(x - m_{X|Y})^\mathsf{T} P_{X|Y}^{-1}(x - m_{X|Y})\right).$$

This is exactly the density function of the multivariate normal distribution (1.1).

2. Let $X = \theta$, $Y = y_0$. We have

$$m_X = \mu, \quad m_Y = x_0\,\mathcal{E}\{\theta\} + \mathcal{E}\{v\} = x_0\mu,$$
$$P_X = \sigma_\theta^2, \quad P_Y = x_0^2\,\mathrm{var}(\theta) + \mathrm{var}(v) = x_0^2\,\sigma_\theta^2 + \sigma_v^2,$$
$$P_{XY} = x_0\,\mathrm{var}(\theta) + \mathrm{cov}(\theta, v) = x_0\,\sigma_\theta^2,$$

where $\mathrm{var}(\cdot)$ and $\mathrm{cov}(\cdot,\cdot)$ denote variance and covariance respectively, and we make use of the property $\mathrm{cov}(aX, bY) = ab\,\mathrm{cov}(X, Y)$.

According to part 1,

$$\theta \mid y_0 \sim \mathcal{N}\left(m_X + P_{XY}P_Y^{-1}(Y - m_Y), P_X - P_{XY}P_Y^{-1}P_{XY}^\mathsf{T}\right).$$

The MAP estimate is obtained by

$$\hat{\theta}_{\mathrm{MAP}} = \mathrm{argmax}_\theta\ f(\theta \mid y_0).$$

This is clearly obtained at the mean of the conditional Gaussian distribution, i.e.,

$$\hat{\theta}_{\mathrm{MAP}} = m_X + P_{XY}P_Y^{-1}(Y - m_Y) = \mu + \frac{x_0\,\sigma_\theta^2}{x_0^2\,\sigma_\theta^2 + \sigma_v^2}(y_0 - x_0\mu).$$

Matlab **Exercise:**

Consider the problem of estimating $\theta = [\theta_1\ \theta_2]^\top$ in the following function

$$y = \theta_1 x + \theta_2(2x^2 - 1) + v\,,$$

where noise v comes from the normal distribution $v \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = 0.5$ and $\sigma^2 = 0.3$.

In order to estimate θ, a set of x and y measurements are collected. The test data are provided as variables `x` and `y` in `SysID_Exercise_1.mat`.

1. Use the maximum likelihood (ML) method to estimate the value of θ.
2. The prior knowledge of θ is characterized by the following distribution,

   $$\theta \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2 \cdot I),\ \text{with}\ \mu_\theta = [1.3\ 0.9]^\top \text{ and } \sigma_\theta^2 = 0.02.$$

   Calculate the maximum a posteriori (MAP) estimate of θ.
3. To assess the accuracy of the ML and MAP estimates, additional measurements are collected for validation, which are provided as variables `x_v` and `y_v` in `SysID_Exercise_1.mat`. Which one is more accurate judging from the validation data?

Solution hints

1. Let K be the length of data and define $w_k = [x_k \; 2x_k^2 - 1]$, for $1 \leq k \leq K$. Define vectors $\mathbf{y}$, $\mathbf{w}$ and $\mathbf{v}$ as

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_K \end{bmatrix}, \qquad \mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_K \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_K \end{bmatrix}.$$

Therefore, we have $\mathbf{y} = \mathbf{w}\theta + \mathbf{v}$. Note that

$$p(\mathbf{y}|\theta) = \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{1}{2\sigma^2}(y_k - (w_k\theta + \mu))^2}.$$

According to the definition of maximum likelihood estimation, we know that

$$\hat{\theta}_{\mathrm{ML}} = \mathrm{argmax}_\theta \; f(\mathbf{y}|\theta) = \mathrm{argmin}_\theta \; \frac{1}{2\sigma^2} \sum_{k=1}^{K} (y_k - (w_k\theta + \mu))^2 =: \mathrm{argmin}_\theta \; V_1(\theta).$$

To solve this optimization problem, we use the following stationary condition

$$\frac{\mathbf{d}V_1(\theta)}{\mathbf{d}\theta} = \frac{1}{\sigma^2} \sum_{k=1}^{K} -w_k^\top (y_k - w_k\theta - \mu) = -\frac{1}{\sigma^2} \mathbf{w}^\top (\mathbf{y} - \mathbf{w}\theta - \mu) = 0.$$

This gives the closed-form solution of θ.

$$\hat{\theta}_{\mathrm{ML}} = \left(\mathbf{w}^\mathsf{T}\mathbf{w}\right)^{-1} \mathbf{w}^\top (\mathbf{y} - \mu) = [1.4458 \; 0.7924]^\top.$$

2. According to Bayes' rule we have

$$f(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)f(\theta)}{f(\mathbf{y})} = \frac{f(\mathbf{y}|\theta)f(\theta)}{\int_\Theta f(\mathbf{y},\theta)\mathbf{d}\theta} \propto f(\mathbf{y}|\theta)f(\theta).$$

Therefore, we obtain

$$f(\theta|\mathbf{y}) \propto \left(\prod_{k=1}^{K} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_k - (w_k\theta + \mu))^2} \right) \cdot \frac{1}{2\pi\sigma_\theta^2} e^{-\frac{1}{2\sigma_\theta^2}\|\theta - \mu_\theta\|_2^2}.$$

According to the definition of maximum a posteriori estimation, we know that

$$\begin{aligned} \hat{\theta}_{\mathrm{MAP}} &= \mathrm{argmax}_\theta \; f(\theta|\mathbf{y}) \\ &= \mathrm{argmin}_\theta \; \frac{1}{2\sigma^2} \sum_{k=1}^{K} (y_k - (w_k\theta + \mu))^2 \; + \; \frac{1}{2\sigma_\theta^2} \|\theta - \mu_\theta\|_2^2 =: \mathrm{argmin}_\theta \; V_2(\theta). \end{aligned}$$

To solve this optimization problem, we use the following stationary condition

$$\begin{aligned} \frac{\mathbf{d}V_2(\theta)}{\mathbf{d}\theta} &= -\frac{1}{\sigma^2} \sum_{k=1}^{K} -w_k^\top (y_k - w_k\theta - \mu) + \frac{1}{\sigma_\theta^2}(\theta - \mu_\theta) \\ &= -\frac{1}{\sigma^2} \mathbf{w}^\top (\mathbf{y} - \mathbf{w}\theta - \mu) + \frac{1}{\sigma_\theta^2}(\theta - \mu_\theta) = 0. \end{aligned}$$

This gives the closed-form solution of θ.

$$\hat{\theta}_{\text{MAP}} = \left(\frac{1}{\sigma^2}\mathbf{w}^\mathsf{T}\mathbf{w} + \frac{1}{\sigma_\theta^2}I\right)^{-1}\left(\frac{1}{\sigma^2}\mathbf{w}^\top(\mathbf{y}-\mu) + \frac{\mu_\theta}{\sigma_\theta^2}\right) = [1.3991\ 0.8199]^\top.$$

3. The accuracy can be assessed by calculating the mean squared error between the predicted $\hat{y} = \hat{\theta}_1 x + \hat{\theta}_2(2x^2-1) + \mu$ and measured y with the validation data:

$$E = \frac{1}{K_v}\sum_{k=1}^{K_v}(\hat{y}_k - y_k)^2,$$

where K_v is the length of the validation data. The results are

$$E_{\text{ML}} = 0.3699, \quad E_{\text{MAP}} = 0.3708.$$

So the Maximum Likelihood estimate is more accurate in this case.