1 Example: Bayesian Estimation of Mean of Gaussian Process

Consider a Gaussian distribution with mean X and variance σ^2 to be the mathematical model of a physical process/system. Specifically, an output quantity of interest Y follows the Gaussian distribution $Y \sim N(X, \sigma^2)$ or, equivalently, the measure of the uncertainty in y given that $X = x$ is given by the PDF

$$p(y \mid x, \sigma^2, I) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(y-x)^2 \right] \tag{1}$$

Given a set of <u>independent</u> observations/data $D \equiv (\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_N) \equiv \{\hat{Y}_k\}_{1\to N}$, we are interesting in updating the uncertainty in the mean X of the model. It is assumed that the value of the variance σ^2 is known. For simplicity we use $x = \mu$ to denote the possible values of the uncertain variable X.

1.1 Case 1: Uniform Prior

<u>Bayes Theorem</u>: The problem will be solved assuming that the prior distribution of the mean is uniform, that is,

$$p(\mu \mid \sigma^2, I) = \begin{cases} 1/(\mu_{\max} - \mu_{\min}), & \mu \in [\mu_{\min}, \mu_{\max}] \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

<u>Posterior</u>: Using Bayes' theorem, the inference about the value of μ given the data and the information I (I includes the selection of the Gaussian model) is expressed by the posterior PDF

$$p(\mu \mid \{\hat{Y}_k\}_{1\to N}, \sigma^2, I) \propto p(\{\hat{Y}_k\}_{1\to N} \mid \mu, \sigma^2, I)\ \ p(\mu \mid \sigma^2, I) \tag{3}$$

<u>Likelihood</u>: To evaluate the likelihood $p(\{\hat{Y}_k\}_{1\to N} \mid \mu, \sigma^2, I)$, one uses the fact that the data are independent and applies successively the product rule of the axioms of probability, given by

$$p(b, a \mid I) = p(b \mid a, I)\ \ p(a \mid I) \tag{4}$$

to finally derive that

$$p(\{\hat{Y}_k\}_{1\to N} \mid \mu, \sigma^2, I) = \prod_{k=1}^{N} p(\hat{Y}_k \mid \mu, \sigma^2, I) = \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(\hat{Y}_k - \mu)^2 \right] \tag{5}$$

<u>Proof of (5)</u>: Specifically, the independence of the data allows us to assume that given the values of μ and σ^2 the measurements of one or more data does not influence the inference about the outcome of another datum. Mathematically, this can be written as

$$p(\hat{Y}_k \mid \hat{Y}_{k-1}, \hat{Y}_{k-2}, \ldots, \hat{Y}_1, \mu, \sigma^2, I) = p(\hat{Y}_k \mid \mu, \sigma^2, I) \qquad \text{for any } k \tag{6}$$

Using now the product rule (4) with $b \equiv \hat{Y}_k$ and $a = (\hat{Y}_{k-1}, \hat{Y}_{k-2}, \ldots, \hat{Y}_1)$, conditioned on the fact that μ and σ^2 are known and the background information I, one derives that

$$\begin{aligned} p(\{\hat{Y}_k\}_{1\to k} \mid \mu, \sigma^2, I) &= p(\hat{Y}_k, \hat{Y}_{k-1}, \hat{Y}_{k-2}, \ldots, \hat{Y}_1 \mid \mu, \sigma^2, I) \\ &= p(\hat{Y}_k \mid \hat{Y}_{k-1}, \hat{Y}_{k-2}, \ldots, \hat{Y}_1, \mu, \sigma^2, I)\; p(\hat{Y}_{k-1}, \hat{Y}_{k-2}, \ldots, \hat{Y}_1 \mid \mu, \sigma^2, I) \\ &= p(\hat{Y}_k \mid \mu, \sigma^2, I)\; p(\hat{Y}_{k-1}, \hat{Y}_{k-2}, \ldots, \hat{Y}_1 \mid \mu, \sigma^2, I) \end{aligned} \tag{7}$$

where the last equality holds due to (6) resulting from the independence of the data. Applying equation (7) with k replaced by $k-1$ one has that the second factor of the left hand side (LHS) of the last equality in (7) is given by

$$p(\{\hat{Y}_k\}_{1\to k-1} \mid \mu, \sigma^2, I) = p(\hat{Y}_{k-1} \mid \mu, \sigma^2, I)\; p(\hat{Y}_{k-2}, \hat{Y}_{k-3}, \ldots, \hat{Y}_1 \mid \mu, \sigma^2, I) \tag{8}$$

Substituting (8) into (7) and continuing this process successively for the resulting factors, one readily derives that

$$p(\hat{Y}_k, \hat{Y}_{k-1}, \hat{Y}_{k-2}, \ldots, \hat{Y}_1, \mu, \sigma^2, I) = \prod_{\rho=1}^{k} p(\hat{Y}_\rho \mid \mu, \sigma^2, I) \tag{9}$$

The proof of the first equality in (5)follows from (9) by setting $k = N$ and replacing the index ρ by k. The second equality in (5) follows by substituting the value of $p(\hat{Y}_k \mid \mu, \sigma^2, I)$ using the PDF in (1). □

Estimation of Posterior PDF: Using (5) to replace the first factor in the reight hand side (RHS) of (3) and the uniform prior PDF (2), the updates posterior PDF of the uncertain parameter μ takes the form

$$p(\mu \mid \{\hat{Y}_k\}_{1\to N}, \sigma^2, I) \propto \prod_{k=1}^{N} \exp\left[-\frac{1}{2\sigma^2} (\hat{Y}_k - \mu)^2 \right] \tag{10}$$

It can be readily shown that this is a Gaussian distribution.

Most Probable Value (MPV) or Best Estimate: The function $L(\mu)$, defined in theory as the minus the logarithm of the posterior PDF, is given by

$$L(\mu) = -\log p(\mu \mid \{\hat{Y}_k\}_{1\to N}, \sigma^2, I) = \sum_{k=1}^{N} \frac{1}{2\sigma^2} (\hat{Y}_k - \mu)^2 + \text{constant} \tag{11}$$

The MPV $\hat{\mu}$ maximizes the posterior PDF or, equivalently, minimizes $L(\mu)$. It satisfies the condition

$$\left.\frac{dL}{d\mu}\right|_{\mu=\hat{\mu}} = \sum_{k=1}^{N} \frac{1}{\sigma^2} (\hat{Y}_k - \hat{\mu}) = \frac{1}{\sigma^2} \left[\sum_{k=1}^{N} (\hat{Y}_k) - N\hat{\mu} \right] = 0$$

The solution for the MPV $\hat{\mu}$ is readily obtained as

$$\hat{\mu} = \frac{1}{N} \sum_{k=1}^{N} \hat{Y}_k \tag{12}$$

which is the mean arithmetic average of the measurements $(\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_N)$. It is worth noting that in this case the MPV does not depends on the magnitude of the variance (error) σ^2 assumed for the Gaussian mathematical model.

Uncertainty in Model Parameter: The uncertainty in the value of the model parameter μ is characterized by the second derivative of the function $L(\mu)$ which is given by

$$\left.\frac{d^2L}{d\mu^2}\right|_{\mu=\hat{\mu}} = \sum_{k=1}^{N}\frac{1}{\sigma^2} = \frac{N}{\sigma^2}$$

The measure of the uncertainty, provided by the square root of the inverse of the second derivative of $L(\mu)$ evaluated at the most probable value, is given by

$$\sqrt{S} = \left(\left.\frac{d^2L}{d\mu^2}\right|_{\mu=\hat{\mu}}\right)^{-1/2} = \frac{\sigma}{\sqrt{N}} \qquad (13)$$

It can be seen that the uncertainty in the estimate depends on the variance σ^2 of the Gaussian mathematical model.

Given the MPV $\hat{\mu}$ and the uncertainty index $\sqrt{S}$ we can write a measure of the uncertainty interval of μ in the form

$$\hat{\mu} \pm \sqrt{S} = \hat{\mu} \pm \frac{\sigma}{\sqrt{N}} \qquad (14)$$

which is a familiar result for the reliability of the mean estimate given N measurements.

Asymptotic Posterior PDF: Following the theoretical result and using the MPV $\hat{\mu}$ and the uncertainty index S, the posterior PDF follows the Gaussian distribution

$$f(\mu \mid \{\hat{Y}_k\}_{1\to N}, \sigma^2, I) = \frac{1}{\sqrt{2\pi H}}\exp\left[-\frac{1}{2H}(\mu-\hat{\mu})^2\right] \qquad (15)$$

Note that this distribution is exact since the function $L(\mu)$ is quadratic in μ.

Figure 1 shows the evolution of the posterior PDF $f(\mu \mid \{\hat{Y}_k\}_{1\to N}, \sigma^2, I)$ (the posterior uncertainty in μ) as a function of the number of data. Note that data affects the values of $\hat{\mu}$ and S, while the posterior PDF in this case is Gaussian for all any N.

Remark 1

The previous analysis can be applied to update the estimates of the uncertain parameter X involved in the mathematical model

$$Y = X + E$$

of a system, where E is the prediction error assumed to follow a zero-mean Gaussian distribution $E \sim N(m, \tau^2)$, Y is the output quantity in the system, and $D \equiv (\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_N) \equiv \{\hat{Y}_k\}_{1\to N}$ are the independent measurements of the output quantity Y.