1 Bayesian Parameter Estimation: One-Dimensional Case

Consider the case of one uncertain parameters X of a model. Bayes theorem is used to make inference about the values of this parameter based on a set of data D and the background information I. Specifically the posterior distribution of the model parameter is given by

$$p(x|D,I) = \frac{p(D|x,I)\ p(x|I)}{p(D|I)} \tag{1}$$

which completely quantifies the uncertainties in the values x of the model parameter X. Often we are interested in summarizing the uncertainty with a couple of indices (uncertainty descriptors). These indices are the best estimate and the spread of uncertainty about the best estimate or the measure of reliability.

The best estimate $\hat{x}$ of the uncertain parameter X is the most probable value that maximizes the posterior probability distribution function (PDF) $p(x|D,I)$. It is often called the Maximum A Posteriori (MAP) value and corresponds to the most probable value (MPV) of the posterior PDF. The best estimate $\hat{x}$ or the MAP is obtained by postulating the condition

$$\left.\frac{\partial p(x|D,I)}{\partial x}\right|_{x=\hat{x}} = 0$$

The condition for $\hat{x}$ to be a maximum of $p(x|D,I)$ is that the second derivative of $p(x|D,I)$ is negative. To guarantee that the PDF achieves its maximum at $\hat{x}$ one needs to check that the second derivative is negative, i.e.

$$\left.\frac{\partial^2 p(x|D,I)}{\partial x^2}\right|_{x=\hat{x}} < 0$$

Alternatively, introducing the function

$$L(x) = -\log[p(x|D,I)] \tag{2}$$

which is defined as the minus the log of the posterior distribution, the best estimate $\hat{x}$ of the value of the model parameter can equivalently be obtained by minimizing the function $L(x)$. Thus, the best estimates of the model parameter is obtained by solving the following equation

$$\left.\frac{\partial L}{\partial x}\right|_{x=\hat{x}} = 0 \tag{3}$$

and ensuring that the solution $\hat{x}$ corresponds to a minimum of $L(x)$. This is done by verifying that the second derivative is positive, i.e.

$$\left.\frac{\partial^2 L(x)}{\partial x^2}\right|_{x=\hat{x}} > 0$$

The uncertainty in the values of the parameter is obtained by considering the spread of the posterior PDF about the best estimate $\hat{x}$. To obtain as useful estimate of the spread of the uncertainty about the best estimate we expand the posterior PDF $p(x|D,I)$ in Taylor series around the best estimate and keep the lower order terms in the expansion. This is the standard approach for approximating a function by a lower-order polynomial. However, noting that the posterior PDF is peaked around the best estimate, it is better to carry out this expansion for the function $L(x)$ which is varies much more slowly.

The local behavior of the posterior PDF about $\hat{x}$ is thus obtained by the Taylor series expansion of the function $L(x)$ about $\hat{x}$, given by

$$L(x) = L(\hat{x}) + \left.\frac{\partial L}{\partial x}\right|_{x=\hat{x}} (x-\hat{x}) + \frac{1}{2}\left.\frac{\partial^2 L}{\partial x^2}\right|_{x=\hat{x}} (x-\hat{x})^2 + \cdots$$

Using the fact that we expand around the minimum of $L(x)$, the linear term in the Taylor series expansion are zero because of (3). Introducing the function $H(x)$ as

$$H(x) = \frac{\partial^2 L}{\partial x^2}$$

the Taylor series expansion of $L(x)$ takes the form

$$L(x) = L(\hat{x}) + \frac{1}{2} H(\hat{x})(x-\hat{x})^2 + \cdots \qquad (4)$$

Note that at the neighbor of the best estimate, the terms of the order of three or higher in the Taylor series expansion of $L(x)$ can be neglected and the behavior of the function $L(x)$ locally is specified by the behavior of the quadratic term. Specifically the spread of uncertainty around the best estimate $\hat{x}$ is determined by the value of $H(x)$ at $\hat{x}$. Making use of (2) and using only up to the quadratic terms in the expansion (4), the posterior PDF is given by

$$\begin{aligned} p(x|D,I) &= \exp[-L(x)] \\ &\propto \exp\left[-L(\hat{x}) - \frac{1}{2} H(\hat{x})(x-\hat{x})^2\right] \\ &\propto \exp\left[-\frac{1}{2} H(\hat{x})(x-\hat{x})^2\right] \end{aligned} \qquad (5)$$

It is thus obvious that the updated PDF is approximated by a Gaussian distribution

$$p(x|D,I) = \frac{1}{\sqrt{2\pi S}} \exp\left[-\frac{1}{2S}(x-\hat{x})^2\right] \qquad (6)$$

with mean the best estimate $\hat{x}$ or MAP or the most probable value of the posterior PDF and variance $S = H^{-1}(\hat{x})$ which is the inverse of the second derivative of the function $L(x)$ evaluated at the best estimate $\hat{x}$.

Bayesian Central Limit Theorem

It can be shown that asymptotically, for large number of data, the posterior PDF tends to the Gaussian distribution (6), centered at its most probable value and with variance matrix equal to the inverse of the second derivative (Hessian) of the minus the logarithm of the posterior PDF, evaluated at the most probable value. The error of the asymptotic approximation is of order of N^{-1} where N denotes the number of data.

Spread of Uncertainty about the Best Estimate

The spread of the uncertainty in the parameters around the best estimate $\hat{x}$ is completely defined by the standard deviation σ of the Gaussian distribution. That is, it is estimated by the value $\hat{H} = H(\hat{x})$, the second-order derivative of the function $L(x)$, which denotes the variance $\sigma^2 = H^{-1}(\hat{x})$ of the Gaussian distribution.