Estimation of Likelihood

<u>Example 1: Scalar Linear Model</u>

Consider the mathematical model

$$Y = \mu + E$$

of a physical process/system, where E is a Gaussian distribution, i.e. $E \sim N(0,\sigma^2)$. Given the values of μ and σ^2 the output quantity of interest Y follows the Gaussian distribution $Y \sim N(\mu,\sigma^2)$ or, equivalently, the uncertainty in y is given by the PDF

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

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 variables μ and σ^2. This involves the estimation of the likelihood.

<u>Bayes Theorem</u>: Using Bayes' theorem, the inference about the values of μ and σ^2 given the data and the information I (I includes the selection of the Gaussian model) is expressed by the posterior PDF

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

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

$$p(b,a \mid I) = p(b \mid a,I)\ \ p(a \mid I) \tag{3}$$

to finally derive that

$$p(\{\hat{Y}_k\}_{1\to N} \mid \mu,\sigma^2,I) = \prod_{k=1}^{N} p(\hat{Y}_k \mid \mu,\sigma^2,I) = \prod_{k=1}^{N}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}(\hat{Y}_k-\mu)^2\right] \tag{4}$$

<u>Proof of (4)</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

$$p(\hat{Y}_k \mid \hat{Y}_{k-1},\hat{Y}_{k-2},\ldots,\hat{Y}_1,\mu,\sigma^2,I) = p(\hat{Y}_k \mid \mu,\sigma^2,I) \qquad \text{for any } k \tag{5}$$

Using now the product rule (3) 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

$$\begin{aligned} p(\{\hat{Y}_k\}_{1\to k} \mid \mu,\sigma^2,I) &= p(\hat{Y}_k,\hat{Y}_{k-1},\hat{Y}_{k-2},\ldots,\hat{Y}_1 \mid \mu,\sigma^2,I) \\ &= p(\hat{Y}_k \mid \hat{Y}_{k-1},\hat{Y}_{k-2},\ldots,\hat{Y}_1,\mu,\sigma^2,I)\ \ p(\hat{Y}_{k-1},\hat{Y}_{k-2},\ldots,\hat{Y}_1 \mid \mu,\sigma^2,I) \\ &= p(\hat{Y}_k \mid \mu,\sigma^2,I)\ \ p(\hat{Y}_{k-1},\hat{Y}_{k-2},\ldots,\hat{Y}_1 \mid \mu,\sigma^2,I) \end{aligned} \tag{6}$$

where the last equality holds due to (5) resulting from the independence of the data. Applying equation (6) with k replaced by $k-1$ one has that the second factor of the left hand side (LHS) of the last equality in (6) is given by

$$p(\{\hat{Y}_k\}_{1\to k-1} \mid \mu,\sigma^2,I) = p(\hat{Y}_{k-1} \mid \mu,\sigma^2,I)\ \ p(\hat{Y}_{k-2},\hat{Y}_{k-3},\ldots,\hat{Y}_1 \mid \mu,\sigma^2,I) \tag{7}$$

Substituting (7) into (6) and continuing this process successively for the resulting factors, one readily derives that

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

The proof of the first equality in (4) follows from (8) by setting $k = N$ and replacing the index ρ by k. The second equality in (4) follows by substituting the value of $p(\hat{Y}_k \mid \mu,\sigma^2,I)$ using the PDF in (1). □

Example 2: Scalar Non-Linear Model

Consider the mathematical model

$$Y = g(\mu) + E \tag{9}$$

of a physical process/system, where E is a Gaussian distribution, i.e. $E \sim N(0,\sigma^2)$. Given the values of μ and σ^2 the output quantity of interest Y follows the Gaussian distribution $Y \sim N(g(\mu),\sigma^2)$ or, equivalently, the uncertainty in y is given by the PDF

$$p(y \mid \mu,\sigma^2,I) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}[y-g(\mu)]^2\right] \tag{10}$$

Given a set of independent 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 variables μ and σ^2. This involves the estimation of the likelihood.

Bayes Theorem: Using Bayes' theorem, the inference about the values of μ and σ^2 given the data and the information I (I includes the selection of the Gaussian model) is expressed by the posterior PDF

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

Estimation of Likelihood: To estimate the likelihood $p(\{\hat{Y}_k\}_{1\to N} \mid \mu,\sigma^2,I)$, one can use the fact that the data are independent and apply successively the product rule of the axioms of probability. This approach is exactly the same as the approach followed in example 1. Thus

$$p(\{\hat{Y}_k\}_{1\to N} \mid \mu, \sigma^2, I) = \prod_{k=1}^{N} p(\hat{Y}_k \mid \mu, \sigma^2, I)$$

Substituting the value of $p(\hat{Y}_k \mid \mu, \sigma^2, I)$ using the PDF in (10), we derive that

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

Example 3: Scalar Linear Difference Equation of 1st Order

Consider a mathematical model of a physical process/system represented by the difference equation

$$Y_k = \mu Y_{k-1} + E \tag{11}$$

where E is a Gaussian distribution, i.e. $E \sim N(0, \sigma^2)$. Given a particular observation $D \equiv (\hat{Y}_0, \hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_N) \equiv \{\hat{Y}_k\}_{0\to N}$ covering all time instances, we are interesting in updating the uncertainty in the variables μ and σ^2. This involves the estimation of the likelihood.

Bayes Theorem: Using Bayes' theorem, the inference about the values of μ and σ^2 given the data and the information I (I includes the selection of the Gaussian model) is expressed by the posterior PDF

$$p(\mu, \sigma^2 \mid \{\hat{Y}_k\}_{0\to N}, I) \propto p(\{\hat{Y}_k\}_{0\to N} \mid \mu, \sigma^2, I) \;\; p(\mu, \sigma^2 \mid I)$$

Estimation of Likelihood: To estimate the likelihood $p(\{\hat{Y}_k\}_{1\to N} \mid \mu, \sigma^2, I)$, one can use the structure of model (11) to relate the value y_k at the current instant to the value of y_{k-1} at the previous instant and apply successively the product rule of the axioms of probability, to finally derive that

$$p(\{\hat{Y}_k\}_{0\to N} \mid \mu, \sigma^2, I) = \prod_{k=1}^{N} p(\hat{Y}_k \mid \hat{Y}_{k-1}, \mu, \sigma^2, I) = \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2} (\hat{Y}_k - \mu\hat{Y}_{k-1})^2 \right] \tag{12}$$

Proof of (12): Using the product rule (3) with $b \equiv \hat{Y}_k$ and $a = (\hat{Y}_{k-1}, \hat{Y}_{k-2}, \ldots, \hat{Y}_1, \hat{Y}_0)$, conditioned on the fact that μ and σ^2 are known and the background information I, one derives that

$$\begin{aligned} p(\{\hat{Y}_k\}_{0\to k} \mid \mu, \sigma^2, I) &= p(\hat{Y}_k, \hat{Y}_{k-1}, \hat{Y}_{k-2}, \ldots, \hat{Y}_0 \mid \mu, \sigma^2, I) \\ &= p(\hat{Y}_k \mid \hat{Y}_{k-1}, \hat{Y}_{k-2}, \ldots, \hat{Y}_0, \mu, \sigma^2, I) \;\; p(\hat{Y}_{k-1}, \hat{Y}_{k-2}, \ldots, \hat{Y}_0 \mid \mu, \sigma^2, I) \end{aligned} \tag{13}$$

Based on the structure of the model (11), given the values of μ and σ^2 as well as the value $Y_{k-1} = \hat{Y}_{k-1}$ at the previous step or time instant $k-1$, the output value Y_k at time instant k is completely described and independent of the values of $\hat{Y}_{k-2}, \ldots, \hat{Y}_1$. This is expressed in mathematical form as

$$p(\hat{Y}_k \mid \hat{Y}_{k-1}, \hat{Y}_{k-2}, \ldots, \hat{Y}_0, \mu, \sigma^2, I) = p(\hat{Y}_k \mid \hat{Y}_{k-1}, \mu, \sigma^2, I)$$

Substituting the last expression in (13), one readily derives that

$$\begin{aligned} p(\{\hat{Y}_k\}_{0\to k} \mid \mu,\sigma^2,I) &= p(\hat{Y}_k \mid \hat{Y}_{k-1},\mu,\sigma^2,I)\ \ p(\hat{Y}_{k-1},\hat{Y}_{k-2},\ldots,\hat{Y}_0 \mid \mu,\sigma^2,I) \\ &= p(\hat{Y}_k \mid \hat{Y}_{k-1},\mu,\sigma^2,I)\ \ p(\{\hat{Y}_k\}_{0\to k-1} \mid \mu,\sigma^2,I) \end{aligned} \tag{14}$$

Applying equation (14) with k replaced by $k-1$ one has that the second factor of the left hand side (LHS) of the last equality in (14) is given by

$$p(\{\hat{Y}_k\}_{0\to k-1} \mid \mu,\sigma^2,I) = p(\hat{Y}_{k-1} \mid \hat{Y}_{k-2},\mu,\sigma^2,I)\ \ p(\hat{Y}_{k-2},\hat{Y}_{k-3},\ldots,\hat{Y}_0 \mid \mu,\sigma^2,I) \tag{15}$$

Substituting (15) into (14) and continuing this process successively for the resulting factors, one readily derives that

$$p(\hat{Y}_k,\hat{Y}_{k-1},\hat{Y}_{k-2},\ldots,\hat{Y}_0 \mid \mu,\sigma^2,I) = \prod_{\rho=1}^{k} p(\hat{Y}_\rho \mid \hat{Y}_{\rho-1},\mu,\sigma^2,I) \tag{16}$$

The proof of the first equality in (12) follows from (16) by setting $k=N$ and replacing the index ρ by k. The second equality in (4) follows by deriving the expression for $p(\hat{Y}_k \mid \hat{Y}_{k-1},\mu,\sigma^2,I)$ using the particular structure of the model (11). Specifically, given the values of μ and σ^2 as well as the value y_{k-1} at the previous step or time instant $k-1$, the output value y_k at time instant k follows the Gaussian distribution $y_k \mid y_{k-1} \sim N(\mu y_{k-1},\sigma^2)$ or, equivalently, the uncertainty in y_k given the value y_{k-1} at the previous instant follows the PDF

$$p(y_k \mid y_{k-1},\mu,\sigma^2,I) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}(y_k-\mu y_{k-1})^2\right] \tag{17}$$

Replacing $y_k=\hat{Y}_k$ and $y_{k-1}=\hat{Y}_{k-1}$ in the last expression and substituting in (16), one readily derives that

$$p(\{\hat{Y}_k\}_{0\to N} \mid \mu,\sigma,I) = \prod_{k=1}^{N}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}(\hat{Y}_k-\mu\hat{Y}_{k-1})^2\right]$$

which completes the proof. □

Example 4: Scalar Non-Linear Difference Equation of 1st Order

Consider a mathematical model of a physical process/system represented by the difference equation

$$Y_k = g(Y_{k-1},\mu) + E$$

where E is a Gaussian distribution, i.e. $E \sim N(0,\sigma^2)$. Given a particular observation $D \equiv (\hat{Y}_0,\hat{Y}_1,\hat{Y}_2,\ldots,\hat{Y}_N) \equiv \{\hat{Y}_k\}_{0\to N}$ covering all time instances, we are interesting in updating the uncertainty in the variables μ and σ^2. This involves the estimation of the likelihood.