

1 Example: Bayesian Estimation of Mean and Variance of Gaussian Process

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

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

Given a set of independent observations/data $D \equiv (\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_N) \equiv \{\hat{Y}_k\}_{1 \rightarrow N}$, we are interesting in updating the uncertainty in the mean μ and variance σ^2 of the model.

1.1 Case 1: Uniform Prior

Bayes Theorem: The problem will be solved assuming that the prior distribution of the mean and the variance is uniform, that is,

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

Posterior: 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

$$p(\mu, \sigma^2 | \{\hat{Y}_k\}_{1 \rightarrow N}, I) \propto p(\{\hat{Y}_k\}_{1 \rightarrow N} | \mu, \sigma^2, I) p(\mu, \sigma^2 | I) \quad (3)$$

Likelihood: The likelihood has already been evaluated in Lecture Notes 2 in the form

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

Estimation of Posterior PDF: Using (4) to replace the first factor in the right hand side (RHS) of (3) and the uniform prior PDF (2), the updates posterior PDF of the uncertain parameters μ and σ^2 takes the form

$$\begin{aligned} p(\mu, \sigma^2 | \{\hat{Y}_k\}_{1 \rightarrow N}, I) &\propto \prod_{k=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(\hat{Y}_k - \mu)^2\right] \\ &\propto \frac{1}{\sigma^N} \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^N (\hat{Y}_k - \mu)^2\right] \end{aligned} \quad (5)$$

Note that this joint distribution of the parameters μ and σ^2 is not Gaussian.

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

$$L(\mu, \sigma^2) = -\log p(\mu, \sigma^2 | \{\hat{Y}_k\}_{1 \rightarrow N}, I) = \frac{N}{2} \log \sigma^2 + \frac{1}{2\sigma^2} \sum_{k=1}^N (\hat{Y}_k - \mu)^2 + \text{constant} \quad (6)$$

The MPVs of $\hat{\mu}$ and $\hat{\sigma}$ maximize the posterior PDF or, equivalently, minimize $L(\mu, \sigma)$. They satisfy the conditions

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

$$\left. \frac{\partial L}{\partial \sigma^2} \right|_{\substack{\mu=\hat{\mu} \\ \sigma=\hat{\sigma}}} = \left[\frac{N}{2\sigma^2} - \frac{1}{2\sigma^4} \sum_{k=1}^N (\hat{Y}_k - \mu)^2 \right]_{\substack{\mu=\hat{\mu} \\ \sigma=\hat{\sigma}}} = \frac{1}{2\hat{\sigma}^2} \left[N - \frac{1}{\hat{\sigma}^2} \sum_{k=1}^N (\hat{Y}_k - \hat{\mu})^2 \right] = 0$$

The solution for the MPVs $\hat{\mu}$ and $\hat{\sigma}$ is readily obtained as

$$\hat{\mu} = \frac{1}{N} \sum_{k=1}^N \hat{Y}_k \quad (7)$$

and

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{k=1}^N (\hat{Y}_k - \hat{\mu})^2 \quad (8)$$

which is the mean arithmetic average and the arithmetic variance of the measurements $(\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_N)$.

Uncertainty in Model Parameters: The uncertainty in the values of the model parameters μ and σ^2 is characterized by the Hessian of the function $L(\mu, \sigma)$ evaluated at the MPVs $\hat{\mu}$ and $\hat{\sigma}$. The components of the Hessian are given by

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

$$\left. \frac{\partial^2 L}{\partial \sigma^2 \partial \mu} \right|_{\substack{\mu=\hat{\mu} \\ \sigma=\hat{\sigma}}} = -\frac{1}{\sigma^4} \sum_{k=1}^N (\mu - \hat{Y}_k) \Big|_{\substack{\mu=\hat{\mu} \\ \sigma=\hat{\sigma}}} = -\frac{1}{\hat{\sigma}^4} \left[N\hat{\mu} - \sum_{k=1}^N (\hat{Y}_k) \right] = 0$$

$$\left. \frac{\partial^2 L}{\partial \sigma^2} \right|_{\substack{\mu=\hat{\mu} \\ \sigma=\hat{\sigma}}} = -\frac{N}{2\hat{\sigma}^4} + \frac{2\hat{\sigma}^2}{2\hat{\sigma}^8} \sum_{k=1}^N (\hat{Y}_k - \hat{\mu})^2 = -\frac{N}{2\hat{\sigma}^4} + \frac{\hat{\sigma}^2}{\hat{\sigma}^8} N\hat{\sigma}^2 = -\frac{N}{2\hat{\sigma}^4} + \frac{N}{\hat{\sigma}^4} = \frac{N}{2\hat{\sigma}^4}$$

so that the Hessian is given by

$$H = \frac{N}{\hat{\sigma}^2} \begin{bmatrix} 1 & 0 \\ 0 & 1/(2\hat{\sigma}^2) \end{bmatrix}$$

and the inverse of the Hessian matrix is given by

$$H^{-1} = \frac{\hat{\sigma}^2}{N} \begin{bmatrix} 1 & 0 \\ 0 & 2\hat{\sigma}^2 \end{bmatrix}$$

Marginal Distribution of the Mean: Following the theoretical developments, the posterior marginal distribution for the mean takes the form

$$\begin{aligned} f(\mu | \{\hat{Y}_k\}_{1 \rightarrow N}, I) &= \int_0^{\infty} f(\mu, \sigma^2 | \{\hat{Y}_k\}_{1 \rightarrow N}, I) d\sigma^2 \\ &\propto \int_0^{\infty} \frac{1}{\sigma^N} \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^N (\hat{Y}_k - \mu)^2\right] d\sigma^2 \\ &= \Gamma\left(\frac{N-1}{2}\right) \left[\frac{1}{2} \sum_{k=1}^N (\hat{Y}_k - \mu)^2\right]^{-\frac{N}{2}+1} \end{aligned}$$

where the last expression is obtained by making use of the integral (17) (see Appendix II). Eventually, one obtains that

$$f(\mu | \{\hat{Y}_k\}_{1 \rightarrow N}, \sigma^2, I) \propto \left[\sum_{k=1}^N (\hat{Y}_k - \mu)^2\right]^{-\frac{N}{2}+1}$$

which is not a Gaussian distribution. It can be demonstrated that the distribution asymptotically, for large number of data N , approaches a Gaussian distribution.

Marginal Distribution of the Variance: Similarly, the posterior marginal distribution of the variance σ^2 is obtained as

$$\begin{aligned} f(\sigma^2 | \{\hat{Y}_k\}_{1 \rightarrow N}, I) &= \int_0^{\infty} f(\mu, \sigma^2 | \{\hat{Y}_k\}_{1 \rightarrow N}, I) d\mu \\ &\propto \int_0^{\infty} \frac{1}{\sigma^N} \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^N (\hat{Y}_k - \mu)^2\right] d\mu \end{aligned}$$

1.2 Case 2: Gaussian Prior for Mean and Inverse Gamma Prior for Variance

Gaussian Prior for mean: The problem will next be solved assuming that the prior distribution of the mean is Gaussian with mean m and variance τ^2 , that is,

$$f(\mu | \sigma^2, I) = \frac{1}{\sqrt{2\pi\tau}} \exp\left[-\frac{1}{2\tau^2} (\mu - m)^2\right] \quad (9)$$

The parameters m and τ^2 are called often *hyperparameters*. In our case we assume them to be known.
Inverse Gamma for Variance: The inverse Gamma distribution for the variance is given by (16) in Appendix II.

Example: Regression

Consider a mathematical model of a physical process/system represented by the linear equations

$$(10)(11) \quad Y = \sum_{k=1}^n a_k X_k + E \quad (12)$$

where E is a Gaussian distribution, i.e. $E \sim N(0, \sigma^2)$. Given a particular observation $D \equiv (\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_N) \equiv \{\hat{Y}_k\}_{1 \rightarrow N}$ covering all time instances, we are interesting in updating the uncertainty in the parameter set $a = [a_1, \dots, a_n]$.

Example: Regression

Consider a mathematical model of a physical process/system represented by the linear system of equations

$$Y = AX + E \quad (13)$$

where $A \in R^{N \times n}$ is a given matrix, $X \in R^n$ the uncertain parameters of the model, $Y \in R^N$ the measured data set, considered independent, and E is a multi-variable zero-mean Gaussian distribution with covariance matrix Σ , i.e. $E \sim N(0, \Sigma)$. Given a particular observation $D \equiv \{\hat{Y}\}$, we are interesting in updating the uncertainty in the parameter set X . Let also $Z = BX + H$ be response quantities of interest, $Z \in R^M$, $B \in R^{M \times n}$ and $H \in R^M$, where $H \sim N(0, \Sigma_0)$. Estimate the uncertainty in Z . (14) *equation reference goes here*

Exercise: N-DOF System Represented by Autoregressive Model

Consider a mathematical model of a physical process/system represented by the linear equations

$$Y_k = \sum_{\rho=1}^m a_\rho Y_{k-\rho} + E_k \quad (15)$$

where E_k are independent identically distributed (i.i.d) zero-mean Gaussian variables, i.e. $E_k \sim N(0, \sigma^2)$. Given a particular observation set $D \equiv (\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_N) \equiv \{\hat{Y}_k\}_{1 \rightarrow N}$ covering all time instances, we are interesting in updating the uncertainty in the parameter set $a = [a_1, \dots, a_m]$.

Exercise: Single DOF Mechanical Oscillator

Consider the mathematical model of an oscillator, with equation of motion given by

$$\ddot{y} + 2\zeta\omega_0\dot{y} + \omega_0^2 y = \frac{1}{m} F(t)$$

Let $F(t) = 0$ and the initial conditions are $y(0) = y_0$ and $\dot{y}(0) = v_0$. Given a set of independent observations/data $D \equiv (\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_N) \equiv \{\hat{Y}_k\}_{1 \rightarrow N}$, we are interesting in estimating the uncertainty (best estimate and co-variance) in the modal frequency ω_0 and damping ratio ζ of the model. Assume that the mass of the oscillator and the initial conditions are given.

APPENDIX II:

A. Inverse Gamma Distribution

The Inverse Gamma distribution of a variable X is given by

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right) \quad (16)$$

with $x > 0$. The Inverse Gamma distribution depends on the two variables α and β . The mean \bar{X} and the variance $Var(X)$ are given by

$$\bar{X} = \frac{\beta}{\alpha - 1} \quad \text{for } \alpha > 1$$

$$Var(X) = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)} \quad \text{for } \alpha > 2$$

B. Useful Integrals

$$\int_0^\infty t^{-a-1} \exp[-\beta/t] dt = \Gamma(a) [\beta]^{-a} \quad (17)$$

$$\int_0^\infty t^{-a} \exp[-\beta/t^2] dt = \frac{1}{2} \Gamma\left[\frac{a-1}{2}\right] [\beta]^{-\frac{a-1}{2}} \quad (18)$$

where $\Gamma(x)$ is the Gamma function given by

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx \quad (19)$$