1 Bayesian Estimation of Variance of a Gaussian Process

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

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

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 variance X of the model. It is assumed that the value of the mean μ is known. For simplicity we use $x = \sigma^2$ to denote the possible values of the uncertain variable X. Assume a uniform prior for σ^2 and derive the expressions for the

1. Posterior PDF $p(\sigma^2 \mid D, \mu, I)$. Note that the posterior PDF follows a inverse gamma distribution $IG(\alpha, \beta)$. What are the values of α and β? (Already done in Homework 1)
2. The function $L(\sigma^2) = -\ln p(\sigma^2 \mid D, \mu, I)$
3. The MPV (or best estimate) $\hat{\sigma}^2$ of σ^2
4. The uncertainty of σ^2
5. Retain up to the quadratic terms in the Taylor series expansion of $L(\sigma^2)$ about the most probable value $\hat{\sigma}^2$ and derive the Gaussian asymptotic approximation for the posterior PDF of $p(\sigma^2 \mid \{\hat{Y}_k\}_{1\to N}, \mu, I)$
6. [THIS QUESTION WAS NOT COVERED IN CLASS] Compare the posterior PDF with the asymptotic Gaussian posterior PDF for the following values of $N = 1, 2, 3, 4, 10, 100, 1000$. To facilitate comparisons, plot the two posterior PDFs (exact and asymptotic) so that the maximum value of each equals unity.

<u>Prior PDF</u>: The uniform PDF for σ^2 is given by

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

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

$$p(\sigma^2 \mid D, \mu, I) \propto p(D \mid \sigma^2, \mu, I)\; p(\sigma^2 \mid \mu, I) \tag{3}$$

<u>Likelihood</u>: The likelihood has already been evaluated in Lecture Notes 2 in the form

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

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 posterior PDF of the uncertain parameter σ^2 given the mean value μ takes the form

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

Note that the distribution of the parameter σ^2 is not Gaussian. In fact, it has been shown in Homework 1 that it follows a inverse gamma distribution.

Most Probable Value (MPV) or Best Estimate: The function $L(\sigma^2)$, defined in theory as the minus the logarithm of the posterior PDF of σ^2, is given by

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

The MPV of $\hat{\sigma}^2$ maximize the posterior PDF or, equivalently, minimize $L(\sigma^2)$. It satisfies the condition

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

The solution for the MPV $\hat{\sigma}^2$ is readily obtained as

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{k=1}^{N} (\hat{Y}_k - \mu)^2 \tag{7}$$

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

Uncertainty in Model Parameters: The uncertainty in the value of the model parameters σ^2 given the value of the mean μ is characterized by the Hessian of the function $L(\sigma^2)$ evaluated at the MPV $\hat{\sigma}^2$. The Hessian is given by

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

The measure of the uncertainty, provided by the square root of the inverse of the Hessian of $L(\sigma^2)$ evaluated at the most probable value $\hat{\sigma}^2$, is given by

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

Given the MPV $\hat{\sigma}^2$ and the uncertainty index $\sqrt{S}$ we can write a measure of the uncertainty interval of σ^2 in the form

$$\hat{\sigma}^2 \pm \sqrt{S} = \hat{\sigma}^2 \pm \frac{\sqrt{2}\hat{\sigma}^2}{\sqrt{N}} \tag{8}$$

Asymptotic Posterior PDF: Following the theoretical result for the Bayesian Central Limit Theorem and using the MPV $\hat{\sigma}^2$ and the uncertainty index S, the posterior PDF follows asymptotically for large number of data N the Gaussian distribution

$$p(\sigma^2 \mid D, \mu, I) = \frac{\sqrt{N}}{\sqrt{2\pi}\sqrt{2}\hat{\sigma}^2} \exp\left[-\frac{N}{4\hat{\sigma}^4} (\sigma^2 - \hat{\sigma}^2)^2 \right] \tag{9}$$

[THIS QUESTION WAS NOT COVERED IN CLASS] Figure 1 shows the evolution of the posterior PDF $p(\sigma^2 \mid D, \mu, I)$ (the posterior uncertainty in σ^2) as a function of the number of data. Note that data affects the values of $\hat{\sigma}^2$ and S, while the posterior PDF in this case is asymptotically approaching a Gaussian distribution $N(\hat{\sigma}^2, 2\hat{\sigma}^4 / N)$ for large values of N.