Lecture Notes: Posterior distribution of single uncertain parameter: the mean of a Gaussian model, Nov 13, 2013

1 Bayesian Estimation of Mean of a 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 2: Gaussian Prior

1. Compute analytically the posterior PDF of μ assuming a Gaussian prior $p(\mu \mid \sigma^2, I) = N(m, \tau^2)$.

   Show that the posterior PDF of μ is Gaussian with mean

   $$\hat{\mu} = Bm + (1 - B)\mu_0$$

   and variance

   $$S = \left(\frac{N}{\sigma^2} + \frac{1}{\tau^2} \right)^{-1} = B\tau^2 = (1 - B)\frac{\sigma^2}{N}$$

   where

   $$\mu_0 = \frac{1}{N} \sum_{k=1}^{N} \hat{Y}_k$$

   is the sampling mean and

   $$B = \left(\frac{N}{\sigma^2} + \frac{1}{\tau^2} \right)^{-1} \frac{1}{\tau^2} \leq 1 \tag{2}$$

   with σ^2 / N being the sampling variance.

2. Plot the posterior PDF as a function of the number of data. Compare the results for three different priors: a uniform, a Gaussian $N(5, 3^2)$ and a Gaussian $N(7, 1^2)$.

<u>*Gaussian Prior*</u>: 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,

Lecture Notes: Posterior distribution of single uncertain parameter: the mean of a Gaussian model, Nov 13, 2013

$$f(\mu \mid \sigma^2, I) = \frac{1}{\sqrt{2\pi\tau}} \exp\left[-\frac{1}{2\tau^2}(\mu - m)^2\right] \tag{3}$$

Posterior PDF: 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

$$p(\mu \mid \{\hat{Y}_k\}_{1\to N}, \sigma^2, I) \propto p(\{\hat{Y}_k\}_{1\to N} \mid \mu, \sigma^2, I)\ \ p(\mu \mid \sigma^2, I) \tag{4}$$

Likelihood: Note tat the selection of the prior does not affect the form of the likelihood. The likelihood has already been evaluated in the form

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

Estimation of Posterior PDF: The update posterior PDF of the uncertain parameter μ takes the form

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

which, after straightforward manipulations, can be simplified to which again can be shown to by a Gaussian distribution.

$$\begin{aligned} p(\mu \mid \{\hat{Y}_k\}_{1\to N}, \sigma^2, I) &\propto \exp\left[-\frac{1}{2\sigma^2}\sum_{k=1}^{N}(\hat{Y}_k - \mu)^2\right] \exp\left[-\frac{1}{2\tau^2}(\mu - m)^2\right] \\ &= \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}\sum_{k=1}^{N}(\hat{Y}_k - \mu)^2 + \frac{1}{\tau^2}(\mu - m)^2\right)\right] \\ &= \exp\left[-\frac{1}{2} J(\mu)\right] \end{aligned} \tag{6}$$

where

$$J(\mu) = \frac{1}{\sigma^2}\sum_{k=1}^{N}(\hat{Y}_k - \mu)^2 + \frac{1}{\tau^2}(\mu - m)^2$$

It is observed that $J(\mu)$ is a quadratic function of μ. In order to show that the posterior PDF for μ is Gaussian distribution we need to complete the square for the exponent $J(\mu)$. Using the definition of μ_0, this is done by the algebraic manipulations that follow:

$$
\begin{aligned}
J(\mu) &= \frac{1}{\sigma^2}\sum_{k=1}^{N}(\hat{Y}_k-\mu)^2+\frac{1}{\tau^2}(\mu-m)^2 \\
&= \frac{1}{\sigma^2}\left(\sum_{k=1}^{N}\hat{Y}_k^{\,2}-2\mu\sum_{k=1}^{N}\hat{Y}_k+N\mu^2\right)+\frac{1}{\tau^2}\left(\mu^2-2\mu m+m^2\right) \\
&= \left(\frac{N}{\sigma^2}+\frac{1}{\tau^2}\right)\mu^2-2\left(\frac{1}{\sigma^2}\sum_{k=1}^{N}\hat{Y}_k+\frac{1}{\tau^2}m\right)\mu+\left(\frac{1}{\sigma^2}\sum_{k=1}^{N}\hat{Y}_k^{\,2}+\frac{1}{\tau^2}m^2\right) \\
&= \left(\frac{N}{\sigma^2}+\frac{1}{\tau^2}\right)\mu^2-2\left(\frac{N}{\sigma^2}\mu_0+\frac{1}{\tau^2}m\right)\mu+\left(\frac{1}{\sigma^2}\sum_{k=1}^{N}\hat{Y}_k^{\,2}+\frac{1}{\tau^2}m^2\right) \\
&= \left(\frac{N}{\sigma^2}+\frac{1}{\tau^2}\right)\left[\mu^2-2\hat{\mu}\mu+\left(\frac{1}{\sigma^2}\sum_{k=1}^{N}\hat{Y}_k^{\,2}+\frac{1}{\tau^2}m^2\right)\Big/\left(\frac{N}{\sigma^2}+\frac{1}{\tau^2}\right)\right] \\
&= \left(\frac{N}{\sigma^2}+\frac{1}{\tau^2}\right)\left[\mu^2-2\hat{\mu}\mu+\hat{\mu}^2\right]-\left(\frac{N}{\sigma^2}+\frac{1}{\tau^2}\right)\hat{\mu}^2+\left(\frac{1}{\sigma^2}\sum_{k=1}^{N}\hat{Y}_k^{\,2}+\frac{1}{\tau^2}m^2\right)\Big/\left(\frac{N}{\sigma^2}+\frac{1}{\tau^2}\right) \\
&= \left(\frac{N}{\sigma^2}+\frac{1}{\tau^2}\right)\left[\mu-\hat{\mu}\right]^2+\text{Constant (Independent of }\mu)
\end{aligned}
$$

where $\hat{\mu}$ was introduced as

$$\hat{\mu}=\left(\frac{N}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1}\left(\frac{N}{\sigma^2}\mu_0+\frac{1}{\tau^2}m\right)=Bm+(1-B)\mu_0$$

and B is given by (2), with $B\le 1$ and $1-B\le 1$.
Substituting $J(\mu)$ in (6) one readily derives that

$$
\begin{aligned}
p(\mu\,|\,\{\hat{Y}_k\}_{1\to N},\sigma^2,I) &\propto \exp\left[-\frac{1}{2}\left(\frac{N}{\sigma^2}+\frac{1}{\tau^2}\right)(\mu-\hat{\mu})^2\right] \\
&= \exp\left[-\frac{1}{2s^2}(\mu-\hat{\mu})^2\right]
\end{aligned}
$$

which is a Gaussian distribution with mean $\hat{\mu}$ and variance s^2 given by

$$s^2=\left(\frac{N}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1}$$

Note that the mean $\hat{\mu}$ of the posterior PDF of μ is a weighted average of the prior mean and the mean arithmetic average $\frac{1}{N}\sum_{k=1}^{N}\hat{Y}_k$ of the measurements $(\hat{Y}_1,\hat{Y}_2,\ldots,\hat{Y}_N)$, with weights that are inversely proportional to the variances τ^2 (prior variance) and σ^2/N (data variance estimated for non-informative priors). For large prior variance τ^2 relative to the data variance σ^2/N ($\tau^2 >> \sigma^2/N$), corresponding to large uncertainty in the prior, the MPV of the posterior mean is close to the mean arithmetic average of the measurements. In contrast, for small prior variance

τ^2 relative to the data variance σ^2 / N ($\tau^2 << \sigma^2 / N$), corresponding to a highly informative prior, the MPV of the posterior mean is close to the prior mean.

Note also that the variance s^2 of the posterior Gaussian distribution satisfies

$$s^2 = \left(\frac{N}{\sigma^2} + \frac{1}{\tau^2} \right)^{-1} = B\tau^2 = (1-B)\frac{\sigma^2}{N} \tag{7}$$

Note that since $B \le 1$ and $1-B \le 1$, the variance of the posterior PDF of μ is smaller than the variance of either the prior posterior or the data variance σ^2 / N based on the uniform prior.

Figure 1 compares the evolution of the posterior PDF $f(\mu | \{\hat{Y}_k\}_{1\to N}, \sigma^2, I)$ as a function of the number of data for the three prior PDFs. Note that as the number of data N increases the two PDFs converge to each other. This means that for large number of data the form of the prior PDF does not affect the posterior which depends only on the likelihood function.

Exercise 1 (Due March 13, 2013): 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

$$f(y | x, \mu, I) = \frac{1}{\sqrt{2\pi X}} \exp\left[-\frac{1}{2X}(y-\mu)^2 \right] \tag{8}$$

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 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 X and derive the expressions for the

1. Posterior PDF $f(\sigma^2 | \{\hat{Y}_k\}_{1\to N}, \mu, I)$. Note that the posterior PDF follows a inverse gamma distribution $IG(\alpha, \beta)$. What are the values of α and β?
2. The function $L(\sigma^2)$
3. The MPV of $X = \sigma^2$
4. The uncertainty of $X = \sigma^2$
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 $f(\sigma^2 | \{\hat{Y}_k\}_{1\to N}, \mu, I)$
6. 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.

Inverse Gamma Distribution

The Inverse Gamma distribution of a variable X is given by

$$f(x, \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right)$$

and it depends on the two variables α and β. The most probable value $\hat{X}$, the mean $\bar{X}$, the variance $Var(\bar{X})$ *and the uncertainty* $\sqrt{S}$ are given by

$$\hat{X} = \frac{\beta}{\alpha+1}$$

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

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