# Tutorial 2: Bayesian Estimation of Linear Regression Models

## Exercise 2: Solution

**Let $y = (y_1, \dots, y_N)'$ be a random sample from a normal distribution with unknown mean $\mu$ and precision $h$,**

$$p(y_i|\mu, h) = (2\pi)^{-\frac{1}{2}} h^{\frac{1}{2}} \exp\left[-\frac{1}{2}h(y_i - \mu)^2\right].$$

**Suppose prior beliefs concerning $\mu$ are, conditional on $h$, represented by a normal distribution with mean $\underline{\mu}$ and precision $h\underline{\kappa}$, $\mu \sim \mathcal{N}(\underline{\mu}, (h\underline{\kappa})^{-1})$, with density**

$$p(\mu|\underline{\mu}, h\underline{\kappa}) = (2\pi)^{-\frac{1}{2}} (h\underline{\kappa})^{\frac{1}{2}} \exp\left[-\frac{1}{2}h\underline{\kappa}(\mu - \underline{\mu})^2\right],$$

**and prior beliefs concerning $h$ are represented by a gamma distribution with parameters $\underline{s}^{-2}$ and $\underline{\nu}$, $h \sim$ Gamma$(\underline{s}^{-2}, \underline{\nu})$, with density**

$$p(h|\underline{s}^{-2}, \underline{\nu}) = \left(\frac{2}{\underline{s}^2\underline{\nu}}\right)^{\frac{\underline{\nu}}{2}} \Gamma\left(\frac{\underline{\nu}}{2}\right)^{-1} h^{\frac{\underline{\nu}-2}{2}} \exp\left[-\frac{1}{2}h\underline{\nu}\underline{s}^2\right].$$

**(a) Find the posterior distribution.**

Step 1: State the prior distribution.
To this end, define $\theta = (\mu, h)'$. Then,

$$\begin{aligned}
p(\theta) &= p(\mu|\underline{\mu}, h\underline{\kappa})p(h|\underline{s}^{-2}, \underline{\nu}) \\
&= (2\pi)^{-\frac{1}{2}} (h\underline{\kappa})^{\frac{1}{2}} \exp\left[-\frac{1}{2}h\underline{\kappa}(\mu - \underline{\mu})^2\right] \left(\frac{2}{\underline{s}^2\underline{\nu}}\right)^{\frac{\underline{\nu}}{2}} \Gamma\left(\frac{\underline{\nu}}{2}\right)^{-1} h^{\frac{\underline{\nu}-2}{2}} \exp\left[-\frac{1}{2}h\underline{\nu}\underline{s}^2\right]
\end{aligned}$$

Recall that multiplicative constants that do not depend on $\theta$ cancel anyway so leave them out for convenience:

$$p(\theta) \propto h^{\frac{\underline{\nu}-1}{2}} \exp\left[-\frac{h}{2}\left\{\underline{\kappa}(\mu-\underline{\mu})^2+\underline{\nu s}^2\right\}\right]$$

Step 2: State the likelihood.

The data are again from a normal distribution:

$$p(y|\theta) \propto h^{\frac{N}{2}} \exp\left[-\frac{1}{2}h\sum_{i=1}^{N}(y_i-\mu)^2\right]$$

Step 3: Find the posterior.

Multiply prior distribution with likelihood and rearrange:

$$\begin{aligned} p(\theta|y) &\propto h^{\frac{\underline{\nu}-1}{2}} \exp\left[-\frac{h}{2}\left\{\underline{\kappa}(\mu-\underline{\mu})^2+\underline{\nu s}^2\right\}\right] \times h^{\frac{N}{2}} \exp\left[-\frac{1}{2}h\sum_{i=1}^{N}(y_i-\mu)^2\right] \\ &\propto h^{\frac{\underline{\nu}+N-1}{2}} \exp\left[-\frac{h}{2}\left\{\underline{\kappa}(\mu-\underline{\mu})^2+\underline{\nu s}^2+\sum_{i=1}^{N}(y_i-\mu)^2\right\}\right] \qquad (1) \end{aligned}$$

Note that this looks very much like the kernel of a Normal-Gamma distribution. Therefore, let us conjecture that the posterior is in fact a Normal-Gamma distribution, i.e., $\theta \sim NG(\bar{\mu}, \bar{\kappa}^{-1}, \bar{s}^{-2}, \bar{\nu})$, which implies that $\mu$ follows, conditional on $h$, a normal distribution with mean $\bar{\mu}$ and precision $h\bar{\kappa}$, and $h$ follows a gamma distribution with parameters $\bar{s}^{-2}$ and $\bar{\nu}$. Hence, we conjecture that the joint posterior distribution is

$$\begin{aligned} p(\theta|y) &= p(\mu|\bar{\mu}, h\bar{\kappa})p(h|\bar{s}^{-2}, \bar{\nu}) \\ &= (2\pi)^{-\frac{1}{2}}\left(\frac{2}{\bar{s}^2\bar{\nu}}\right)^{\frac{\bar{\nu}}{2}}\Gamma\left(\frac{\bar{\nu}}{2}\right)^{-1}\bar{\kappa}^{\frac{1}{2}}h^{\frac{\bar{\nu}-1}{2}}\exp\left[-\frac{h}{2}\left\{\bar{\kappa}(\mu-\bar{\mu})^2+\bar{\nu}\bar{s}^2\right\}\right] \\ &\propto h^{\frac{\bar{\nu}-1}{2}}\exp\left[-\frac{h}{2}\left\{\bar{\kappa}(\mu-\bar{\mu})^2+\bar{\nu}\bar{s}^2\right\}\right], \qquad (2) \end{aligned}$$

where in the last step we again dropped irrelevant constants. By comparing (1) and (2) we can find the parameters $\bar{\mu}$, $\bar{\kappa}$, $\bar{s}$, and $\bar{\nu}$ of the posterior Normal-Gamma distribution as functions of the data and the parameters of the prior distribution.

**Posterior parameters:** To find $\bar{\nu}$, compare $h^{\frac{\underline{\nu}+N-1}{2}}$ in (1) with $h^{\frac{\bar{\nu}-1}{2}}$ in (2). These two expressions must be the same and thus

$$\underline{\nu} + N - 1 = \bar{\nu} - 1 \;\Rightarrow\; \bar{\nu} = \underline{\nu} + N.$$

To find the other parameters of the posterior, note that the exponential functions in (1) and (2) must also be equal. Hence,

$$\underline{\kappa}(\mu - \underline{\mu})^2 + \underline{\nu s}^2 + \sum_{i=1}^{N}(y_i - \mu)^2 = \bar{\kappa}(\mu - \bar{\mu})^2 + \bar{\nu}\bar{s}^2.$$

Note that some terms in the LHS and RHS are quadratic functions in $\mu$. To be equal for all $\mu$, the parameters of these quadratic functions in $\mu$ must be equal. Hence, decompose the quadratic functions from both sides of the equation:

$$\underline{\kappa}\mu^2 - 2\underline{\kappa}\underline{\mu}\mu + \underline{\kappa}\underline{\mu}^2 + \underline{\nu s}^2 + \sum_{i=1}^{N} y_i^2 - 2\mu\sum_{i=1}^{N} y_i + N\mu^2 = \bar{\kappa}\mu^2 - 2\bar{\kappa}\bar{\mu}\mu + \bar{\kappa}\bar{\mu}^2 + \bar{\nu}\bar{s}^2,$$

and finally

$$\underbrace{\Big(\underline{\kappa} + N\Big)\mu^2 - 2\Big(\underline{\kappa}\underline{\mu} + N\bar{y}\Big)\mu + \Big(\underline{\nu s}^2 + \underline{\kappa}\underline{\mu}^2 + N\overline{y^2}\Big)}_{\text{LHS}} = \underbrace{\bar{\kappa}\mu^2 - 2\bar{\kappa}\bar{\mu}\mu + \bar{\kappa}\bar{\mu}^2 + \bar{\nu}\bar{s}^2}_{\text{RHS}},$$

where $\bar{y} \equiv N^{-1}\sum_{i=1}^{N} y_i$ and $\overline{y^2} \equiv N^{-1}\sum_{i=1}^{N} y_i^2$. By comparing the coefficients on the LHS and the RHS, we obtain the three equations

$$\bar{\kappa} = \underline{\kappa} + N \tag{3}$$
$$\bar{\kappa}\bar{\mu} = \underline{\kappa}\underline{\mu} + N\bar{y} \tag{4}$$
$$\bar{\kappa}\bar{\mu}^2 + \bar{\nu}\bar{s}^2 = \underline{\nu s}^2 + \underline{\kappa}\underline{\mu}^2 + N\overline{y^2} \tag{5}$$

From (3), we directly conclude that

$$\bar{\kappa} = \underline{\kappa} + N.$$

Substituting this into (4) yields

$$\bar{\mu} = \frac{1}{\bar{\kappa}}\left(\underline{\kappa}\underline{\mu} + N\bar{y}\right) = \frac{\underline{\kappa}\underline{\mu} + N\bar{y}}{\underline{\kappa} + N}.$$

Finally, we solve (5) for $\bar{s}^2$:

$$\bar{s}^2 = \frac{\underline{\nu}\underline{s}^2 + \underline{\kappa}\underline{\mu}^2 + N\overline{y^2} - \bar{\kappa}\bar{\mu}^2}{\bar{\nu}} = \frac{\underline{\nu}\underline{s}^2 + \underline{\kappa}\underline{\mu}^2 + N\overline{y^2} - (\underline{\kappa} + N)\left(\frac{\underline{\kappa}\underline{\mu}+N\bar{y}}{\underline{\kappa}+N}\right)^2}{\underline{\nu} + N}.$$

Note that this is already a solution because $\bar{s}^2$ is written as a function of the data and the parameters of the prior distribution. However, it is possible to simplify it considerably. To this end, define the "frequentist" parameter $\nu = N - 1$ (degrees of freedom) and the "frequentist" estimator of the sample variance $s^2 \equiv (N-1)^{-1}\sum_{i=1}^{N}(y_i - \bar{y})^2 = \frac{N}{N-1}\overline{y^2} - \frac{N}{N-1}\bar{y}^2$. Now substitute $N\overline{y^2} = \nu s^2 + N\bar{y}^2$ in the above expression:

$$\begin{aligned}
\bar{s}^2 &= \frac{\underline{\nu}\underline{s}^2 + \underline{\kappa}\underline{\mu}^2 + \nu s^2 + N\bar{y}^2 - \frac{1}{\underline{\kappa}+N}\left(\underline{\kappa}^2\underline{\mu}^2 + 2N\underline{\kappa}\underline{\mu}\bar{y} + N^2\bar{y}^2\right)}{\underline{\nu} + N} \\
&= \frac{\underline{\nu}\underline{s}^2 + \nu s^2 + \frac{\underline{\kappa}^2 + \underline{\kappa}N - \underline{\kappa}^2}{\underline{\kappa}+N}\underline{\mu}^2 - \frac{2N\underline{\kappa}}{\underline{\kappa}+N}\underline{\mu}\bar{y} + \frac{N\underline{\kappa} + N^2 - N^2}{\underline{\kappa}+N}\bar{y}^2}{\underline{\nu} + N}
\end{aligned}$$

Rearrange as a quadratic expression in $\underline{\mu} - \bar{y}$:

$$\bar{s}^2 = \frac{\underline{\nu}\underline{s}^2 + \nu s^2 + \frac{\underline{\kappa}N}{\underline{\kappa}+N}\left(\underline{\mu}^2 - 2\underline{\mu}\bar{y} + \bar{y}^2\right)}{\underline{\nu} + N} = \frac{\underline{\nu}\underline{s}^2 + \nu s^2 + \frac{\underline{\kappa}N}{\underline{\kappa}+N}\left(\underline{\mu} - \bar{y}\right)^2}{\underline{\nu} + N}$$

To summarize, the parameters of the posterior distribution are:

$$\begin{aligned}
\bar{\mu} &= \frac{\underline{\kappa}\underline{\mu} + N\bar{y}}{\underline{\kappa} + N} \\
\bar{\kappa} &= \underline{\kappa} + N \\
\bar{s}^2 &= \frac{\underline{\nu}\underline{s}^2 + \nu s^2 + \frac{\underline{\kappa}N}{\underline{\kappa}+N}\left(\underline{\mu} - \bar{y}\right)^2}{\underline{\nu} + N} \\
\bar{\nu} &= \underline{\nu} + N
\end{aligned}$$

**(b) Find and interpret** $\mathrm{E}(\mu|y)$ **and** $\mathrm{E}(h|y)$.

To find $\mathrm{E}(\mu|y)$, use the LIE

$$\mathrm{E}(\mu|y) = \mathrm{E}[\mathrm{E}(\mu|h, y)|y] = \mathrm{E}[\bar{\mu}|y] = \bar{\mu} = \frac{\underline{\kappa}\underline{\mu} + N\bar{y}}{\underline{\kappa} + N}.$$

Interpretation: the posterior mean is the weighted average of the prior and sample mean with respective weights given by the prior precision and the sample size. Hence, the more imprecise the prior knowledge of the mean is, the closest the posterior mean gets to the sample mean (the frequentist estimator).

To find $\mathrm{E}(h|y)$, just use the posterior marginal Gamma distribution of $h$,

$$\mathrm{E}(h|y) = \bar{s}^{-2} = \left( \frac{\underline{\nu}\underline{s}^2 + \nu s^2 + \frac{\underline{\kappa}N}{\underline{\kappa}+N}\left(\underline{\mu} - \bar{y}\right)^2}{\underline{\nu} + \nu + 1} \right)^{-1}.$$

Hence, the posterior mean of the precision $h$ is the inverse of a kind of posterior variance estimator which is a weighted average of the prior variance $\underline{s}^2$, the sample variance $s^2$, and the MSE between the prior mean and the sample mean, $\left(\underline{\mu} - \bar{y}\right)^2$. The weights depend on the "prior degrees of freedom" $\underline{\nu}$, the sample degrees of freedom $\nu$, and the prior precision, $\underline{\kappa}$, concerning the prior mean $\underline{\mu}$.