A simple Gaussian location model

A) We find the marginal distribution $p(\theta)$ by integrating $p(\theta, \omega)$ with respect to ω:

$$\begin{aligned}
p(\theta) &\propto \int p(\theta, \omega)\, d\omega \\
&\propto \int \omega^{(d+1)/2-1} \exp\left(-\omega \cdot \frac{\kappa(\theta-\mu)^2}{2}\right) \exp\left(-\omega \cdot \frac{\eta}{2}\right) d\omega \\
&= \int \omega^{(d+1)/2-1} \exp\left(-\omega \cdot \frac{\kappa(\theta-\mu)^2 + \eta}{2}\right) d\omega \\
&\propto \left(\frac{\kappa(\theta-\mu)^2+\eta}{2}\right)^{-(d+1)/2} \\
&= \left(\frac{\eta}{2} + \frac{\kappa(\theta-\mu)^2}{2}\right)^{-(d+1)/2} \\
&\propto \left(1 + \frac{\kappa(\theta-\mu)^2}{\eta}\right)^{-(d+1)/2} \\
&= \left(1 + \frac{1}{d} \cdot \frac{\kappa(\theta-\mu)^2}{\eta/d}\right)^{-(d+1)/2}
\end{aligned}$$

This has the form

$$\left(1 + \frac{1}{\nu} \cdot \frac{(\theta-m)^2}{s^2}\right)^{-(\nu+1)/2}$$

where

$$\nu = d$$

$$m = \mu$$

$$s^2 = \frac{\eta}{d\kappa}$$

B) To find the joint posterior density $p(\theta, \omega|\mathbf{y})$, we simplify the likelihood times the prior:

$$
\begin{aligned}
p(\theta, \omega|\mathbf{y}) &= p(\theta, \omega) \cdot p(\mathbf{y}|\theta, \omega) \\
&\propto \omega^{(d+1)/2-1} \exp\left(-\omega\left(\frac{\kappa(\theta-\mu)^2}{2}\right)\right) \exp\left(-\omega\left(\frac{\eta}{2}\right)\right) \cdot \omega^{n/2} \exp\left(-\omega\left(\frac{\sum_{i=1}^{n}(y_i-\theta)^2}{2}\right)\right) \\
&= \omega^{(n+d+1)/2-1} \exp\left(-\omega\left(\frac{\kappa(\theta-\mu)^2}{2}\right)\right) \exp\left(-\omega\left(\frac{\eta}{2}\right)\right) \cdot \exp\left(-\omega\left(\frac{S_y + n(\bar{y}-\theta)^2}{2}\right)\right) \\
&= \omega^{(n+d+1)/2-1} \exp\left(-\frac{\omega}{2}\left(\kappa(\theta-\mu)^2 + \eta + S_y + n(\bar{y}-\theta)^2\right)\right)
\end{aligned}
$$

Our goal is now to split the exponent of e, ignoring the $-\omega/2$ term, into part involving θ and part not involving it.

$$
\begin{aligned}
\kappa(\theta-\mu)^2 + \eta + S_y + n(\bar{y}-\theta)^2 &= \kappa\left(\theta^2 - 2\mu\theta + \mu^2\right) + n\left(\bar{y}^2 - 2\bar{y}\theta + \theta^2\right) + \eta + S_y \\
&= (\kappa+n)\theta^2 - 2\left(\kappa\mu + n\bar{y}\right)\theta + \kappa\mu^2 + n\bar{y}^2 + \eta + S_y \\
&= (\kappa+n)\left(\theta^2 - \frac{2\left(\kappa\mu+n\bar{y}\right)\theta}{\kappa+n} + \left(\frac{\kappa\mu+n\bar{y}}{\kappa+n}\right)^2\right) - \frac{\left(\kappa\mu+n\bar{y}\right)^2}{\kappa+n} \\
&\qquad + \kappa\mu^2 + n\bar{y}^2 + \eta + S_y \\
&= (\kappa+n)\left(\theta - \frac{\kappa\mu+n\bar{y}}{\kappa+n}\right)^2 - \frac{\left(\kappa\mu+n\bar{y}\right)^2}{\kappa+n} + \kappa\mu^2 + n\bar{y}^2 + \eta + S_y \\
&= (\kappa+n)\left(\theta - \frac{\kappa\mu+n\bar{y}}{\kappa+n}\right)^2 + \eta + S_y - \frac{\kappa^2\mu^2 + 2\kappa\mu n\bar{y} + n^2\bar{y}^2}{\kappa+n} \\
&\qquad + \frac{\kappa^2\mu^2 + \kappa n\bar{y}^2 + \kappa n\mu^2 + n^2\bar{y}^2}{\kappa+n} \\
&= (\kappa+n)\left(\theta - \frac{\kappa\mu+n\bar{y}}{\kappa+n}\right)^2 + \eta + S_y + \frac{\kappa n\bar{y}^2 - 2\kappa\mu n\bar{y} + \kappa n\mu^2}{\kappa+n} \\
&= (\kappa+n)\left(\theta - \frac{\kappa\mu+n\bar{y}}{\kappa+n}\right)^2 + \eta + S_y + \frac{\kappa n\left(\bar{y}^2 - 2\mu\bar{y} + \mu^2\right)}{\kappa+n} \\
&= (\kappa+n)\left(\theta - \frac{\kappa\mu+n\bar{y}}{\kappa+n}\right)^2 + \eta + S_y + \frac{\kappa n\left(\mu-\bar{y}\right)^2}{\kappa+n}
\end{aligned}
$$

We now put this back into the posterior formula:

$$
p(\theta, \omega|\mathbf{y}) \propto \omega^{(n+d+1)/2-1} \exp\left(-\frac{\omega}{2}(\kappa+n)\left(\theta - \frac{\kappa\mu+n\bar{y}}{\kappa+n}\right)^2\right) \exp\left(-\frac{\omega}{2}\left(\eta + S_y + \frac{\kappa n\left(\mu-\bar{y}\right)^2}{\kappa+n}\right)\right)
$$

This gives us the following updates from the prior:

$$\mu^* = \frac{\kappa\mu + n\bar{y}}{\kappa + n}$$

$$\kappa^* = \kappa + n$$

$$d^* = d + n$$

$$\eta^* = \eta + S_y + \frac{\kappa n\left(\mu - \bar{y}\right)^2}{\kappa + n}$$

C) To find $p(\theta|\mathbf{y},\omega)$, it suffices to look at the posterior $p(\theta,\omega|\mathbf{y})$, ignore the factors not containing a θ, and see that it is the kernel of a normal distribution:

$$\exp\left(-\frac{\omega}{2}(\kappa + n)\left(\theta - \frac{\kappa\mu + n\bar{y}}{\kappa + n}\right)^2\right)$$

It has mean m and variance v given by

$$m = \frac{\kappa\mu + n\bar{y}}{\kappa + n} = \mu^*$$

$$v = \frac{1}{\omega(\kappa + n)} = \frac{1}{\omega\kappa^*}$$

D) To find $p(\omega|\mathbf{y})$, we start with $p(\theta,\omega|\mathbf{y})$ and integrate out θ. Taking our kernel from part (C), we integrate out θ to get the inverse of the normalizing constant, namely $1/\sqrt{\omega(\kappa + n)}$. Therefore, we get

$$p(\omega|\mathbf{y},\theta) = \int p(\theta,\omega|\mathbf{y})\,d\theta$$

$$\propto \frac{\omega^{(d^*+1)/2-1}\exp\left(-\frac{\omega}{2}\eta^*\right)}{\sqrt{\omega(\kappa + n)}}$$

$$\propto \omega^{d^*/2-1}\exp\left(-\frac{\omega}{2}\eta^*\right)$$

Thus, $p(\omega|\mathbf{y})$ is $\text{Gamma}\left(d^*/2, \eta^*/2\right)$.

E) This is completely analogous to part (A), except now we are starting with

$$\omega^{(d^*+1)/2-1}\exp\left(-\omega\cdot\frac{\kappa^*(\theta - \mu^*)^2}{2}\right)\exp\left(-\omega\cdot\frac{\eta^*}{2}\right)$$

instead of

$$\omega^{(d+1)/2-1}\exp\left(-\omega\cdot\frac{\kappa(\theta - \mu)^2}{2}\right)\exp\left(-\omega\cdot\frac{\eta}{2}\right)$$

and thus, as in (A), we get

$$p(\theta|\mathbf{y}) \propto \left(1 + \frac{1}{\nu^*} \cdot \frac{(\theta - m^*)^2}{(s^2)^*}\right)^{-(\nu^*+1)/2}$$

where

$$\nu^* = d^*$$

$$m^* = \mu^*$$

$$(s^2)^* = \frac{\eta^*}{d^*\kappa^*}$$

F) The prior $p(\omega)$ is $\text{Gamma}\big(d/2, \eta/2\big)$; that is,

$$\frac{\big(\eta/2\big)^{d/2}}{\Gamma\big(d/2\big)} x^{d/2-1} e^{-\eta/2}$$

As d and η approach 0, we have the following four facts:

$$\big(\eta/2\big)^{d/2} \text{ is bounded between 0 and 1}$$

$$\Gamma\big(d/2\big) \to \infty$$

$$x^{d/2-1} \to 1/x \text{ for all } x \neq 0$$

$$e^{-\eta/2} \to 1$$

This means that the prior is shrinking to the zero function everywhere except at $x = 0$, and therefore leads to an improper prior in the limit.

Similarly, let us consider the prior for θ:

$$p(\theta|\omega) = \sqrt{\frac{\omega\kappa}{2\pi}} \exp\left(-\frac{\omega\kappa}{2}(\theta - \mu)^2\right)$$

As κ and μ approach 0, this also approaches the zero function, and is therefore also an improper prior. The answer to both true/false questions, therefore, is “false”.

G) These are both true. To see why, let us first note the behavior of the updated hyperparameters as the original hyperparameters approach 0:

$$\mu^* \to \bar{y}$$
$$\kappa^* \to n$$
$$d^* \to n$$
$$\eta^* \to S_y$$

Since $p(\omega|\mathbf{y}, \theta)$ is $\text{Gamma}\big(d^*/2, \eta^*/2\big)$, we see that in the limit, $p(\omega|\mathbf{y}, \theta)$ becomes $\text{Gamma}\big(n/2, S_y/2\big)$. Further, in the notation of (E), we also have

$$\nu^* \to n$$
$$m^* \to \bar{y}$$
$$s^2 \to \frac{S_y}{n^2}$$

We therefore see that the posterior $p(\theta|\mathbf{y})$ approaches a distribution proportional to

$$\left(1 + \frac{1}{n} \cdot \frac{(\theta - \bar{y})^2}{S_y/n^2}\right)^{-(n+1)/2}$$

which is a t distribution with n degrees of freedom and location and scale parameters $\bar{y}$ and S_y/n^2, respectively.

H) True. As noted above, as the hyperparameters all approach 0, m approaches $\bar{y}$ and s approaches $\sqrt{S_y}/n$, which are the statistics used to create a frequentist confidence interval.

The conjugate Gaussian linear model

A) We can find the conditional distribution using some results from the first set of exercises.

$$\begin{aligned}
p(\beta|\mathbf{y},\omega) &\propto p(\beta|\omega)p(\mathbf{y}|\beta,\omega) \\
&\propto \exp\left(-\frac{\omega}{2}(\beta-m)^T K(\beta-m)\right)\exp\left(-\frac{\omega}{2}(\mathbf{y}-X\beta)^T\Lambda(\mathbf{y}-X\beta)\right) \\
&\propto \exp\left(-\frac{\omega}{2}\left(\beta^T K\beta - 2\beta^T Km - 2\beta^T X^T\Lambda\mathbf{y} - (X\beta)^T\Lambda X\beta\right)\right) \\
&= \exp\left(-\frac{\omega}{2}\left(\beta^T(K+X^T\Lambda X)\beta - 2\beta^T(Km+X^T\Lambda\mathbf{y})\right)\right)
\end{aligned}$$

The second line comes from part (F) of the multivariate normal distribution section of the first set of exercises. Using that same result, we recognize this as the kernel of a normal distribution, so

$$\beta|\mathbf{y},\omega \sim N\left((K+X^T\Lambda X)^{-1}(Km+X^T\Lambda\mathbf{y}),\ \ \omega^{-1}(K+X^T\Lambda X)^{-1}\right).$$

B) Let us start by finding the joint posterior distribution $p(\beta,\omega|\mathbf{y})$:

$$\begin{aligned}
p(\beta,\omega|\mathbf{y}) &\propto p(\beta,\omega)p(\mathbf{y}|\beta,\omega) \\
&= p(\omega)p(\beta|\omega)p(\mathbf{y}|\beta,\omega) \\
&\propto \omega^{d/2-1}\exp(-\omega\eta/2)\cdot\omega^{p/2}\exp\left(-\frac{\omega}{2}(\beta-m)^T K(\beta-m)\right)\cdot\omega^{n/2}\exp\left(-\frac{\omega}{2}(\mathbf{y}-X\beta)^T\Lambda(\mathbf{y}-X\beta)\right) \\
&\propto \omega^{(n+p+d)/2-1}\exp\left(-\frac{\omega}{2}\left(\beta^T(K+X^T\Lambda X)\beta - 2\beta^T(Km+X^T\Lambda\mathbf{y})\right)\right)
\end{aligned}$$