

## Section 2: Bayesian inference in Gaussian models

## 2.1 Bayesian inference in a simple Gaussian model

Let's start with a simple, one-dimensional Gaussian example, where

$$y_i | \mu, \sigma^2 \sim \text{N}(\mu, \sigma^2).$$

We will assume that  $\mu$  and  $\sigma$  are unknown, and will put conjugate priors on them both, so that

$$\begin{aligned}\sigma^2 &\sim \text{Inv-Gamma}(\alpha_0, \beta_0) \\ \mu | \sigma^2 &\sim \text{Normal}\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right)\end{aligned}$$

or, equivalently,

$$\begin{aligned}y_i | \mu, \omega &\sim \text{N}(\mu, 1/\omega) \\ \omega &\sim \text{Gamma}(\alpha_0, \beta_0) \\ \mu | \omega &\sim \text{Normal}\left(\mu_0, \frac{1}{\omega \kappa_0}\right)\end{aligned}$$

We refer to this as a normal/inverse gamma prior on  $\mu$  and  $\sigma^2$  (or a normal/gamma prior on  $\mu$  and  $\omega$ ). We will now explore the posterior distributions on  $\mu$  and  $\omega$  ( $/\sigma^2$ ) – much of this will involve similar results to those obtained in the first set of exercises.

**Exercise 2.1** *Derive the conditional posterior distributions  $p(\mu, \omega | y_1, \dots, y_n)$  (or  $p(\mu, \sigma^2 | y_1, \dots, y_n)$ ) and show that it is in the same family as  $p(\mu, \omega)$ . What are the updated parameters  $\alpha_n, \beta_n, \mu_n$  and  $\kappa_n$ ?*

Solution: Based on Bayes law,  $p(\mu, \omega | y_1, \dots, y_n) \propto p(y_1, \dots, y_n | \mu, \omega) \cdot p(\mu | \omega) \cdot p(\omega)$  Given conditions

above, we may re-write this equation as following:

$$\begin{aligned}
p(y_1, \dots, y_n | \mu, \omega) \cdot p(\mu | \omega) \cdot p(\omega) &= \prod_i^n N(\mu, 1/\omega) N(\mu, 1/(\omega \kappa_0)) \\
&\propto \prod_i^n \omega^{1/2} \exp(-\frac{\omega}{2}(\mu - y_i)^2) \omega^{1/2} \exp(-\frac{\omega \kappa_0}{2}(\mu - \mu_0)^2) \omega^{\alpha_0 - 1} \exp(-\beta_0 \omega) \\
&= \omega^{n/2} \exp(-\sum_i^n \frac{\omega}{2}(\mu - y_i)^2) \omega^{1/2} \exp(-\frac{\omega \kappa_0}{2}(\mu - \mu_0)^2) \omega^{\alpha_0 - 1} \exp(-\beta_0 \omega) \\
&= \omega^{n/2 + \alpha_0 - 1/2} \exp(-\sum_i^n \frac{\omega}{2}(\mu - y_i)^2) \exp(-\frac{\omega \kappa_0}{2}(\mu - \mu_0)^2) \exp(-\beta_0 \omega) \\
&= \omega^{n/2 + \alpha_0 - 1/2} \exp(-\frac{n\omega}{2}(\mu - y_i)^2 - \frac{\omega \kappa_0}{2}(\mu - \mu_0)^2 - \beta_0 \omega) \\
&= \omega^{n/2 + \alpha_0 - 1/2} \exp(-\frac{n\omega}{2}(\mu^2 - 2y_i\mu + y_i^2) - \frac{\omega \kappa_0}{2}(\mu^2 - 2\mu_0\mu + \mu_0^2) - \beta_0 \omega) \\
&= \omega^{n/2 + \alpha_0 - 1/2} \exp(-\frac{n\omega}{2}\mu^2 + 2\frac{n\omega}{2}y_i\mu - \frac{n\omega}{2}y_i^2 - \frac{\omega \kappa_0}{2}\mu^2 + \frac{\omega \kappa_0}{2}2\mu_0\mu - \frac{\omega \kappa_0}{2}\mu_0^2 - \beta_0 \omega) \\
&= \omega^{n/2 + \alpha_0 - 1/2} \exp(-\frac{n\omega + \omega \kappa_0}{2}\mu^2 + (n\omega y_i + \omega \kappa_0 \mu_0)\mu - \frac{n\omega}{2}y_i^2 - \frac{\omega \kappa_0}{2}\mu_0^2 - \beta_0 \omega) \\
&= \omega^{n/2 + \alpha_0 - 1/2} \exp(-\frac{n\omega + \omega \kappa_0}{2}\mu^2 + (n\omega y_i + \omega \kappa_0 \mu_0)\mu - \frac{n\omega}{2}y_i^2 - \frac{\omega \kappa_0}{2}\mu_0^2 - \beta_0 \omega) \\
&= \omega^{n/2 + \alpha_0 - 1/2} \exp(-\beta_0 \omega) \exp(\mu^2 - \frac{2(n\omega y_i + \omega \kappa_0 \mu_0)}{n\omega + \omega \kappa_0}\mu - \frac{n\omega}{n\omega + \omega \kappa_0}y_i^2 - \frac{\omega \kappa_0}{-n\omega + \omega \kappa_0}\mu_0^2) \\
&= \omega^{n/2 + \alpha_0 - 1/2} \exp(-\beta_0 \omega) \exp((\mu - \frac{(n\omega y_i + \omega \kappa_0 \mu_0)}{n\omega + \omega \kappa_0})^2 - (\frac{(n\omega y_i + \omega \kappa_0 \mu_0)}{n\omega + \omega \kappa_0})^2 \\
&\quad - \frac{n\omega}{n\omega + \omega \kappa_0}y_i^2 - \frac{\omega \kappa_0}{-n\omega + \omega \kappa_0}\mu_0^2) \\
&= \omega^{n/2 + \alpha_0 - 1/2} \exp(-\beta_0 \omega) \exp((\mu - \frac{(ny_i + k_0 \mu_0)}{n + k_0})^2 - (\frac{(ny_i + k_0 \mu_0)}{n + k_0})^2 \\
&\quad - \frac{n}{n + k_0}y_i^2 - \frac{k_0}{-n + k_0}\mu_0^2)
\end{aligned} \tag{2.1}$$

From this, we may know that  $\mu_n = \frac{(ny_i + k_0 \mu_0)}{n + k_0}$ ,  $\alpha_n = n/2 + \alpha_0$ ,  
 $\beta_n = \beta_0 + (\frac{(ny_i + k_0 \mu_0)}{n + k_0})^2 + \frac{n}{n + k_0}y_i^2 + \frac{k_0}{-n + k_0}\mu_0^2$ ,  $k_n = k_0 + n$

**Exercise 2.2** Derive the conditional posterior distribution  $p(\mu | \omega, y_1, \dots, y_n)$  and  $p(\omega | y_1, \dots, y_n)$  (or if you'd prefer,  $p(\mu | \sigma^2, y_1, \dots, y_n)$  and  $p(\sigma^2 | y_1, \dots, y_n)$ ). Based on this and the previous exercise, what are reasonable interpretations for the parameters  $\mu_0, \kappa_0, \alpha_0$  and  $\beta_0$ ?

Solution: Based on Bayes law,  $p(\mu | \omega, y_1, \dots, y_n) \propto p(y_1, \dots, y_n | \mu, \omega) \cdot p(\mu | \omega)$  Given conditions above, we

may re-write this equation as following:

$$\begin{aligned}
p(y_1, \dots, y_n | \mu, \omega) \cdot p(\mu | \omega) &\propto \prod_i^n N(\mu, 1/\omega) N(\mu, 1/(\omega \kappa_0)) \\
&\propto \prod_i^n \omega^{1/2} \exp(-\frac{\omega}{2}(\mu - y_i)^2) \omega^{1/2} \exp(-\frac{\omega \kappa_0}{2}(\mu - \mu_0)^2) \\
&= \omega^{n/2} \exp(-\sum_i^n \frac{\omega}{2}(\mu - y_i)^2) \omega^{1/2} \exp(-\frac{\omega \kappa_0}{2}(\mu - \mu_0)^2) \\
&\propto \exp(\frac{-\omega n}{2}(\mu^2 - 2\mu y_i + y_i^2) - \frac{\omega \kappa_0}{2}(\mu^2 - 2\mu \mu_0 + \mu_0^2)) \\
&\propto \exp(\frac{-\omega n}{2}\mu^2 + \frac{\omega n}{2}2\mu y_i + \frac{-\omega n}{2}y_i^2 - \frac{\omega \kappa_0}{2}\mu^2 + \frac{\omega \kappa_0}{2}2\mu \mu_0 - \frac{\omega \kappa_0}{2}\mu_0^2) \\
&\propto \exp(\frac{-\omega n}{2}\mu^2 + \frac{\omega n}{2}2\mu y_i - \frac{\omega n}{2}y_i^2 - \frac{\omega \kappa_0}{2}\mu^2 + \frac{\omega \kappa_0}{2}2\mu \mu_0 - \frac{\omega \kappa_0}{2}\mu_0^2) \\
&\propto \exp(\frac{-\omega(n + \kappa_0)}{2}(\mu - \frac{\sum y_i + \kappa_0 \mu_0}{n + \kappa_0}))
\end{aligned} \tag{2.2}$$

Based on this, we know that it has mean  $\frac{\sum y_i + \kappa_0 \mu_0}{n + \kappa_0}$

On the other hand, we may obtain  $p(\omega | y_1, \dots, y_n)$  by:

$$\begin{aligned}
p(\omega | y_1, \dots, y_n) &\propto p(y_1, \dots, y_n | \mu, \omega) \cdot p(\omega) \\
&\propto \prod_i^n \omega^{1/2} \exp(-\frac{\omega}{2}(\mu - y_i)^2) \omega^{\alpha_0 - 1} \exp(-\beta_0 \omega) \\
&= \omega^{n/2} \exp(-\frac{\omega}{2} \sum_i^n (\mu - y_i)^2) \omega^{\alpha_0 - 1} \exp(-\beta_0 \omega) \\
&= \omega^{n/2} \exp(-\frac{\omega}{2} \sum_i^n (\mu - y_i)^2) \omega^{\alpha_0 - 1} \exp(-\beta_0 \omega) \\
&= \omega^{\alpha_0 - 1 + n/2} \exp(-\frac{\omega}{2} \sum_i^n (\mu - y_i)^2 - \beta_0 \omega)
\end{aligned} \tag{2.3}$$

Thus, it will follow gamma distribution with  $(\alpha_0 + n/2 - 1, \beta_0 + \sum_i^n (\mu - y_i)^2/2)$

**Exercise 2.3** Show that the marginal distribution over  $\mu$  is a centered, scaled  $t$ -distribution (note we showed something very similar in the last set of exercises!), i.e.

$$p(\mu) \propto \left(1 + \frac{1}{\nu} \frac{(\mu - m)^2}{s^2}\right)^{-\frac{\nu+1}{2}}$$

What are the location parameter  $m$ , scale parameter  $s$ , and degree of freedom  $\nu$ ?

Solution: Since we know  $p(\mu) = \int p(\mu, \omega) d\omega \propto \int p(\mu | \omega) p(\omega) d\omega$ , we can write the equation as following:

$$\begin{aligned}
p(\mu) &= \int p(\mu, \omega) d\omega \propto \int p(\mu | \omega) p(\omega) d\omega \\
&\propto \int \omega^{1/2} \exp(-\frac{\omega \kappa_0}{2}(\mu - \mu_0)^2) \omega^{\alpha_0 - 1} \exp(-\beta_0 \omega) d\omega \\
&\propto \int \omega^{\alpha_0 + 1/2 - 1} \exp(-\frac{\omega \kappa_0}{2}(\mu - \mu_0)^2 - \beta_0 \omega) d\omega
\end{aligned} \tag{2.4}$$

We have the kernel of gamma  $\text{Gamma}(\alpha_0 + 1/2, \frac{k_0}{2}(\mu - \mu_0)^2 + \beta_0)$ . Thus, we can use this to integration.

$$\begin{aligned}
 p(\mu) &\propto \left[ \frac{\Gamma(\alpha_0 + \frac{1}{2})}{\frac{k_0}{2}(\mu - \mu_0)^2 + \beta_0} \right]^{\alpha_0 + 1/2} \\
 &\propto \left[ \beta_0 + \frac{\kappa_0(\mu - \mu_0)^2}{2} \right]^{-(\alpha_0 + 1/2)} \\
 &\propto \left[ 1 + \frac{1}{2\alpha_0} \frac{\alpha_0 \kappa_0 (\mu - \mu_0)^2}{\beta_0} \right]^{-\frac{(2\alpha_0 + 1)}{2}}
 \end{aligned} \tag{2.5}$$

From this, we can know that  $m = \mu_0, \nu = 2\alpha_0, s = \sqrt{\frac{\beta_0}{\alpha_0 \kappa_0}}$ .

**Exercise 2.4** The marginal posterior  $p(\mu|y_1, \dots, y_n)$  is also a centered, scaled  $t$ -distribution. Find the updated location, scale and degrees of freedom.

Solution: Again, what we need to do is integration over  $p(\mu, \omega|y_1, \dots, y_n)$  respect  $\omega$ .

$$\begin{aligned}
 p(\mu|y_1, \dots, y_n) &= \int p(\mu, \omega|y_1, \dots, y_n) d\omega \\
 &= \int \omega^{n/2 + \alpha_0 - 1/2} \exp\left((\mu - \frac{ny_i + k_0\mu_0}{n + k_0})^2\right) \exp(-\beta_0\omega - (\frac{ny_i + k_0\mu_0}{n + k_0})^2) \\
 &\quad - \frac{n}{n + k_0} y_i^2 - \frac{k_0}{-n + k_0} \mu^2) d\omega
 \end{aligned} \tag{2.6}$$

Then, we may approach similarly how we did in previous exercise.

$$\begin{aligned}
 p(\mu|y_1, \dots, y_n) &\propto \int \omega^{n/2 + \alpha_0 - 1/2} \exp\left((\mu - \frac{ny_i + k_0\mu_0}{n + k_0})^2\right) \exp(-\beta_0\omega - (\frac{ny_i + k_0\mu_0}{n + k_0})^2 - \frac{ny_i^2 - k_0\mu^2}{n + k_0}) d\omega \\
 &\propto \left[ 1 + \frac{1}{2(\alpha_0 + \frac{n}{2})} \frac{(\alpha_0 + \frac{n}{2})(k_0 + n)(\mu - \frac{k_0\mu_0 + n\bar{y}}{k_0 + n})^2}{\beta_0 + \frac{1}{2}[\frac{k_0 n(\bar{y} - \mu_0)^2}{k_0 + n} + \sum_{i=1}^n (y_i - \bar{y})^2]} \right]^{-\frac{(2(\alpha_0 + \frac{n}{2}) + 1)}{2}}
 \end{aligned} \tag{2.7}$$

From this, we can know  $m = \mu_n = \frac{k_0\mu_0 + n\bar{y}}{k_0 + n}, \nu = 2(\alpha_0 + \frac{n}{2}) = 2\alpha_n, s = \sqrt{\frac{\beta_n}{\alpha_n \kappa_n}}$

**Exercise 2.5** Derive the posterior predictive distribution  $p(y_{n+1}, \dots, y_{n+m}|y_1, \dots, y_m)$ .

Solution: To obtain  $p(y_{n+1}, \dots, y_{n+m}|y_1, \dots, y_m)$ , we need two distributions:  $p(\mu, \omega|y_1, \dots, y_m)$  and  $p(y_{m+1}, \dots, y_{m+n}|\mu, \omega)$ . More specifically,  $\int \int p(y_{m+1}, \dots, y_{m+n}|\mu, \omega) p(\mu, \omega|y_1, \dots, y_m) d\mu d\omega$  will return us the posterior predictive distribution.

$$\begin{aligned}
 &\int \int p(y_{m+1}, \dots, y_{m+n}|\mu, \omega) p(\mu, \omega|y_1, \dots, y_m) d\mu d\omega \\
 &\propto \omega^{m/2 + \alpha_0 - 1/2} \exp\left(-\sum_{n+1}^{n+m} \frac{\omega}{2} (\mu - y_i)^2\right) \omega^{\alpha_n - \frac{1}{2}} \exp\left((\mu - \frac{ny_i + k_0\mu_0}{n + k_0})^2 + \beta_n\right) \\
 &\propto \omega^{m/2 + \alpha_0 - 1/2} \exp\left(-\sum_{n+1}^{n+m} \frac{\omega}{2} (\mu - y_i)^2\right) \omega^{n/2 + \alpha_0 - 1/2} \exp(-\beta_0\omega) \exp\left((\mu - \frac{ny_i + k_0\mu_0}{n + k_0})^2 - (\frac{ny_i + k_0\mu_0}{n + k_0})^2\right)
 \end{aligned}$$

**Exercise 2.6** Derive the marginal distribution over  $y_1, \dots, y_n$ .

Solution: Marginal distribution over  $y_1, \dots, y_n$  would be  $p(y_1, \dots, y_n) =$

Solution:

## 2.2 Bayesian inference in a multivariate Gaussian model

Let's now assume that each  $y_i$  is a  $d$ -dimensional vector, such that

$$y_i \sim N(\mu, \Sigma)$$

for  $d$ -dimensional mean vector  $\mu$  and  $d \times d$  covariance matrix  $\Sigma$ .

We will put an *inverse Wishart* prior on  $\Sigma$ . The inverse Wishart distribution is a distribution over positive-definite matrices parametrized by  $\nu_0 > d - 1$  degrees of freedom and positive definite matrix  $\Lambda_0^{-1}$ , with pdf

$$p(\Sigma|\nu_0, \Lambda_0^{-1}) = \frac{|\Lambda|^{d/2}}{2^{(\nu_0+d)/2} \Gamma_d(\nu_0/2)} |\Sigma|^{-\frac{\nu_0+d+1}{2}} e^{-\frac{1}{2} \text{tr}(\Lambda \Sigma^{-1})}$$

where  $\Gamma_d(x) = \pi^{d(d-1)/4} \prod_{i=1}^d \Gamma(x - \frac{i-1}{2})$ .

**Exercise 2.7** Show that in the univariate case, the inverse Wishart distribution reduces to the inverse gamma distribution.

Solution:  $p(\Sigma|\nu_0, \Lambda_0^{-1})$  is the univariate,  $d = 1$ ;

$$\begin{aligned} p(\Sigma|\nu_0, \Lambda_0^{-1}) &= \frac{|\Lambda|^{d/2}}{2^{(\nu_0+d)/2} \Gamma_d(\nu_0/2)} |\Sigma|^{-\frac{\nu_0+d+1}{2}} e^{-\frac{1}{2} \text{tr}(\Lambda \Sigma^{-1})} \\ &= \frac{(|\Lambda|/2)^{d/2}}{\sqrt{2} \Gamma_d(\nu_0/2)} |\Sigma|^{-\frac{\nu_0+d+1}{2}} e^{-\frac{1}{2} \text{tr}(\Lambda \Sigma^{-1})} \end{aligned} \quad (2.8)$$

It is inverse-gamma( $\nu/2, \Lambda/2$ )

**Exercise 2.8** Let  $\Sigma \sim \text{Inv-Wishart}(\nu_0, \Lambda_0^{-1})$  and  $\mu|\Sigma \sim N(\mu_0, \Sigma/\kappa_0)$ , so that

$$p(\mu, \Sigma) \propto |\Sigma|^{-\frac{\nu_0+d+2}{2}} e^{-\frac{1}{2} \text{tr}(\Lambda_0 \Sigma^{-1}) - \frac{\kappa_0}{2} (\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0)}$$

and let

$$y_i \sim N(\mu, \Sigma)$$

Show that  $p(\mu, \Sigma|y_1, \dots, y_n)$  is also normal-inverse Wishart distributed, and give the form of the updated parameters  $\mu_n, \kappa_n, \nu_n$  and  $\Lambda_n$ . It will be helpful to note that

$$\begin{aligned}
\sum_{i=1}^n (y_i - \mu)^T \Sigma^{-1} (y_i - \mu) &= \sum_{i=1}^n \sum_{j=1}^d \sum_{k=1}^d (x_{ij} - \mu_j) (\Sigma^{-1})_{jk} (x_{ik} - \mu_k) \\
&= \sum_{j=1}^d \sum_{k=1}^d (\Sigma^{-1})_{ab} \sum_{i=1}^n (x_{ij} - \mu_j) (x_{ik} - \mu_k) \\
&= \text{tr} \left( \Sigma^{-1} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T \right)
\end{aligned}$$

Based on this, give interpretations for the prior parameters.

Solution:

$$\begin{aligned}
p(\mu, \Sigma | y_1, \dots, y_n) &\propto p(y_1, \dots, y_n | \mu, \Sigma) p(\mu, \Sigma) \\
p(y_1, \dots, y_n | \mu, \Sigma) p(\mu, \Sigma) &\propto \prod |\Sigma|^{\frac{1}{2}} \exp\left(-\frac{(y_i - \mu)^T \Sigma^{-1} (y_i - \mu)}{2}\right) |\Sigma|^{-\frac{\nu_0 + d + 2}{2}} e^{-\frac{1}{2} \text{tr}(\Lambda_0 \Sigma^{-1}) - \frac{\kappa_0}{2} (\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0)}
\end{aligned} \tag{2.9}$$

Let's focus on exponential part, first.

$$\begin{aligned}
&\exp\left(\sum_i^n -\frac{(y_i - \mu)^T \Sigma^{-1} (y_i - \mu)}{2} - \frac{1}{2} \text{tr}(\Lambda_0 \Sigma^{-1}) - \frac{\kappa_0}{2} (\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0)\right) \\
&\exp\left(\text{tr}\left(1/2 \sum_{i=1}^n (y_i - \mu) \Sigma^{-1} (y_i - \mu)^T\right) - \frac{1}{2} \text{tr}(\Lambda_0 \Sigma^{-1}) - \kappa_0 (\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0)\right) \\
&\exp\left(\text{tr}/2 \left[\left(\sum_{i=1}^n (y_i - \mu) \Sigma^{-1} (y_i - \mu)^T\right) - (\Lambda_0 \Sigma^{-1}) - \kappa_0 (\mu - \mu_0) \Sigma^{-1} (\mu - \mu_0)^T\right]\right) \\
&\exp\left(\text{tr}/2 \left[\left(\sum_{i=1}^n (y_i - \mu)(y_i - \mu)^T\right) - (\Lambda_0) - \kappa_0 (\mu - \mu_0)(\mu - \mu_0)^T\right] \Sigma^{-1}\right) \\
&\exp\left(\text{tr}/2 \left[n(y_i y_i^T - y_i \mu^T - y_i^T \mu + \mu \mu^T) - (\Lambda_0) - \kappa_0 (\mu \mu^T - \mu_0^T \mu - \mu_0 \mu^T + \mu_0 \mu_0^T)\right] \Sigma^{-1}\right) \\
&\exp\left(\text{tr}/2 \left[n y_i y_i^T - n y_i \mu^T - n y_i^T \mu + n \mu \mu^T - \Lambda_0 - \kappa_0 (\mu \mu^T - \mu_0^T \mu - \mu_0 \mu^T + \mu_0 \mu_0^T)\right]\right) \\
&\exp\left(\text{tr}/2 \left[(n + \kappa_0) \mu^2 - 2 \mu (n y_i^T + \kappa_0 \mu_0^T) - \Lambda_0 - \kappa_0 \mu_0^2 + n y_i^2\right]\right) \\
&\exp\left(\text{tr}/2 \left[\left(\mu - \frac{(n y_i^T + \kappa_0 \mu_0^T)}{(n + \kappa_0)}\right)^2 - \left(\frac{(n y_i^T + \kappa_0 \mu_0^T)}{(n + \kappa_0)}\right)^2 - \Lambda_0 - \kappa_0 \mu_0^2 + n y_i^2\right] \Sigma^{-1}\right)
\end{aligned} \tag{2.10}$$

From this, we can know  $\mu_n = n y_i^T + \kappa_0 \mu_0^T$ ,  $\nu_n = \nu_0 + n$ ,  $\kappa_n = \kappa_0 + n$  and

$$\Lambda_n = -\left(\frac{(n y_i^T + \kappa_0 \mu_0^T)}{(n + \kappa_0)}\right)^2 - \Lambda_0 - \kappa_0 \mu_0^2 + n y_i^2 \Sigma^{-1}$$

## 2.3 A Gaussian linear model

Lets now add in covariates, so that

$$\mathbf{y}|\beta, X \sim \text{Normal}(X\beta, (\omega\Lambda)^{-1})$$

where  $\mathbf{y}$  is a vector of  $n$  responses;  $X$  is a  $n \times d$  matrix of covariates; and  $\Lambda$  is a known positive definite matrix. Let's assume  $\beta \sim \text{Normal}(\mu, (\omega K)^{-1})$  and  $\omega \sim \text{Gamma}(a, b)$ , where  $K$  is assumed fixed.

**Exercise 2.9** Derive the conditional posterior  $p(\beta|\omega, y_1, \dots, y_n)$

Solution:

$$\begin{aligned} p(\beta|\omega, y_1, \dots, y_n) &\propto p(y_1, \dots, y_n|\beta, \omega) \cdot p(\beta|\omega) \\ &\propto \exp(-1/2[(y - X\beta)^T(\omega\Lambda)(y - X\beta) + (\beta - \mu)^T(\omega K)(\beta - \mu)]) \\ &\propto \exp(-1/2[\beta^T X^T \omega \Lambda X \beta - 2\beta X^T \omega \Lambda y + \beta^T K \beta - 2\beta^T K \mu]) \\ &\propto \exp(-1/2[\beta^T (X^T \omega \Lambda X + K) \beta - 2\beta X^T \omega \Lambda y - 2\beta^T K \mu]) \\ &\propto \exp(-1/2[\beta^T (X^T \omega \Lambda X + K) \beta - 2\beta (X^T \omega \Lambda y + K \mu) (X^T \omega \Lambda X + K)^{-1} (X^T \omega \Lambda X + K)]) \\ &\propto \exp(-1/2(X^T \omega \Lambda X + K) [\beta^T \beta - 2\beta (X^T \omega \Lambda y + K \mu) (X^T \omega \Lambda X + K)^{-1}]) \\ &\propto \exp(-1/2[\beta - \frac{(X^T \omega \Lambda y + K \mu)}{(X^T \omega \Lambda X + K)}]^T (X^T \omega \Lambda X + K) [\beta - \frac{(X^T \omega \Lambda y + K \mu)}{(X^T \omega \Lambda X + K)}]) \end{aligned} \quad (2.11)$$

Then we may say that  $\mu_n = (X^T \omega \Lambda y + K \mu) (X^T \omega \Lambda X + K)^{-1}$  and  $\Sigma^{-1} = \omega (X^T \omega \Lambda X + K)$

**Exercise 2.10** Derive the marginal posterior  $p(\omega|y_1, \dots, y_n)$

Solution:

$$\begin{aligned} p(\omega|y_1, \dots, y_n) &\propto \int p(\beta, \omega|y_1, \dots, y_n) d\beta \propto \int p(\omega) p(\beta|\omega) p(y_1, \dots, y_n|\omega, \beta) d\beta \\ &\propto \int \omega^{(a + \frac{d+n}{2} - 1)} \exp(-\omega b) \exp(-\frac{1}{2}((\beta - \mu)^T (\omega \kappa) (\beta - \mu) + (y - X\beta)^T (\omega \Lambda) (y - X\beta))) d\beta \\ &\propto \omega^{(a + \frac{d+n}{2} - 1)} \exp(-\omega b) \int \exp(-\frac{1}{2}((\beta - \mu)^T (\omega \kappa) (\beta - \mu) + (y - X\beta)^T (\omega \Lambda) (y - X\beta))) d\beta \\ &\propto \omega^{\alpha + \frac{d+n}{2} - 1} \exp(-b\omega) \exp(-\frac{1}{2}\omega(\mu^T K \mu + y^T \Lambda y)) \\ &\quad \int \exp(-\frac{1}{2}\omega(\beta^T K \beta - 2\mu^T K \beta - 2y^T \Lambda X \beta + (X\beta)^T \Lambda X \beta)) d\beta \\ &\propto \omega^{\alpha + \frac{d+n}{2} - 1} \exp(-\omega(b + \frac{1}{2}(\mu^T K \mu + y^T \Lambda y))) \\ &\quad \int \exp(-\frac{1}{2}\omega(K + X^T \Lambda X)[(\beta - \frac{\mu K + Y^T \Lambda X}{K + X^T \Lambda X})^2 - (\frac{\mu K + y^T \Lambda X}{K + X^T \Lambda X})^2]) d\beta \\ &\propto \omega^{\alpha + \frac{d+n}{2} - 1} \exp(-\omega(b + \frac{1}{2}(\mu^T K \mu + y^T \Lambda y) - \frac{1}{2}\mu_n^T (K + X^T \Lambda X) \mu_n)) \end{aligned} \quad (2.12)$$

Thus, we have following result,

$$\mu_n = \frac{\mu K + y^T \Lambda X}{K + X^T \Lambda X}, a_n = a + \frac{n+d}{2}, b_n = b + \frac{1}{2}[(\mu^T K \mu + Y^T \Lambda Y) - \mu_n^T (K + X^T \Lambda X) \mu_n]$$

**Exercise 2.11** Derive the marginal posterior,  $p(\beta|y_1, \dots, y_n)$

*Solution:*

$$\begin{aligned} p(\beta|y_1, \dots, y_n) &\propto \int p(\beta|\omega, y_1, \dots, y_n) p(\omega|y_1, \dots, y_n) d\omega \\ &\propto \int \exp(-1/2[\beta - \frac{(X^T \Lambda y + K \mu)}{(X^T \Lambda X + K)}]^T (X^T \omega \Lambda X + K) [\beta - \frac{(X^T \Lambda y + K \mu)}{(X^T \Lambda X + K)}]) \\ &\quad \omega^{\alpha + \frac{d+n}{2} - 1} \exp(-\omega(b + \frac{1}{2}(\mu^T K \mu + y^T \Lambda y) - \frac{1}{2}\mu_n^T (K + X^T \Lambda X) \mu_n)) d\omega \end{aligned} \quad (2.13)$$

We need to rewrite the term for convenience of further drive:

$$\begin{aligned} \mu_n &= \frac{(X^T \Lambda y + K \mu)}{(X^T \Lambda X + K)}, \alpha_n = \alpha + \frac{d+n}{2}, b_n = -(b + \frac{1}{2}(\mu^T K \mu + y^T \Lambda y) - \frac{1}{2}\mu_n^T (K + X^T \Lambda X) \mu_n) \\ &\propto \int \exp(-1/2[\beta - \mu_n]^T (X^T \omega \Lambda X + K) [\beta - \mu_n]) \omega^{\alpha_n - 1} \exp(\omega b_n) d\omega \end{aligned} \quad (2.14)$$

**Exercise 2.12** Download the dataset `dental.csv` from Github. This dataset measures a dental distance (specifically, the distance between the center of the pituitary to the pterygomaxillary fissure) in 27 children. Add a column of ones to correspond to the intercept. Fit the above Bayesian model to the dataset, using  $\Lambda = I$  and  $K = I$ , and picking vague priors for the hyperparameters, and plot the resulting fit. How does it compare to the frequentist LS and ridge regression results?

## 2.4 A hierarchical Gaussian linear model

The dental dataset has heavier tailed residuals than we would expect under a Gaussian model. We've seen previously that we can model a scaled  $t$ -distribution using a scale mixture of Gaussians; let's put that into effect here. Concretely, let

$$\begin{aligned} \mathbf{y}|\beta, \omega, \Lambda &\sim N(X\beta, (\omega\Lambda)^{-1}) \\ \Lambda &= \text{diag}(\lambda_1, \dots, \lambda_n) \\ \lambda_i &\stackrel{iid}{\sim} \text{Gamma}(\tau, \tau) \\ \beta|\omega &\sim N(\mu, (\omega K)^{-1}) \\ \omega &\sim \text{Gamma}(a, b) \end{aligned}$$

**Exercise 2.13** What is the conditional posterior,  $p(\lambda_i|\mathbf{y}, \beta, \omega)$ ?

**Exercise 2.14** Write a Gibbs sampler that alternates between sampling from the conditional posteriors of  $\lambda_i$ ,  $\beta$  and  $\omega$ , and run it for a couple of thousand samplers to fit the model to the dental dataset.

**Exercise 2.15** Compare the two fits. Does the new fit capture everything we would like? What assumptions is it making? In particular, look at the fit for just male and just female subjects. Suggest ways in which we could modify the model, and for at least one of the suggestions, write an updated Gibbs sampler and run it on your model.