SDS 383D: Modeling II **Spring 2017**

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

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

We will assume that μ and σ 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,

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

We refer to this as a normal/inverse gamma prior on μ and σ^2 (or a normal/gamma prior on μ and ω). We will now explore the posterior distributions on μ and $\omega(1/\sigma^2)$ – much of this will involve similar results to those obtained in the first set of exercises.

Exercise 2.1 *Derive the conditional posterior distributions $p(\mu, \omega|y_1, \ldots, y_n)$ (or $p(\mu, \sigma^2|y_1, \ldots, y_n)$) and show that it is in the same family as $p(\mu, \omega)$. What are the updated parameters α_n, β_n, μ_n and κ_n?*

Solution: The joint distribution of μ and ω is

$$p(\mu, \omega) = p(\mu|\omega)p(\omega) \propto \sqrt{\omega\kappa_0}e^{-\frac{\omega\kappa_0(\mu-\mu_0)^2}{2}} \times \omega^{\alpha_0-1}e^{-\beta_0\omega}$$

Then we have the posterior

$$\begin{aligned}p(\mu, \omega|y_1, \ldots, y_n) &\propto p(y_1, \ldots, y_n|\mu, \omega)p(\mu|\omega)p(\omega)\\ &\propto \prod_{i=1}^n \omega^{1/2}e^{-\frac{\omega(y_i-\mu)^2}{2}} \times \sqrt{\omega\kappa_0}e^{-\frac{\omega\kappa_0(\mu-\mu_0)^2}{2}} \times \omega^{\alpha_0-1}e^{-\beta_0\omega}\\ &\propto \omega^{\alpha_0+\frac{n}{2}-\frac{1}{2}} exp\left\{-\omega(2\beta_0 + \kappa_0\mu_0^2 + \sum_{i=1}^n y_i^2 - \frac{(\mu_0\kappa_0 + \sum_{i=1}^n y_i)^2}{\kappa_0+n})/2\right\}\\ &\quad \cdot exp\left\{-\omega(\kappa_0+n)\left(\mu - \frac{\mu_0\kappa_0 + \sum_{i=1}^n y_i}{\kappa_0+n}\right)^2/2\right\}\end{aligned}$$

Therefore, the updated parameters are $\alpha_n = \alpha_0 + \frac{n}{2}$, $\beta_n = \frac{2\beta_0+\kappa_0\mu_0^2+\sum_{i=1}^n y_i^2 - \frac{(\mu_0\kappa_0+\sum_{i=1}^n y_i)^2}{\kappa_0+n}}{2}$, $\mu_n = \frac{\mu_0\kappa_0+\sum_{i=1}^n y_i}{\kappa_0+n}$, and $\kappa_n = \kappa_0 + n$.

Exercise 2.2 *Derive the conditional posterior distribution $p(\mu|\omega, y_1, \ldots, y_n)$ and $p(\omega|y_1, \ldots, y_n)$ (or if you'd prefer, $p(\mu|\sigma^2, y_1, \ldots, y_n)$ and $p(\sigma^2|y_1, \ldots, y_n)$). Based on this and the previous exercise, what are reasonable interpretations for the parameters $\mu_0, \kappa_0, \alpha_0$ and β_0?*

Solution:

$$
\begin{aligned}
p(\omega|y_1, \ldots, y_n) &= \int p(\mu, \omega|y_1, \ldots, y_n) d\mu \\
&\propto \omega^{\alpha_0 + \frac{n}{2} - \frac{1}{2}} exp\left\{ -\omega(2\beta_0 + \kappa_0\mu_0^2 + \sum_{i=1}^n y_i^2 - \frac{(\mu_0\kappa_0 + \sum_{i=1}^n y_i)^2}{\kappa_0 + n})/2 \right\} \\
&\quad \cdot \int exp\left\{ -\omega(\kappa_0 + n)\left(\mu - \frac{\mu_0\kappa_0 + \sum_{i=1}^n y_i}{\kappa_0 + n}\right)^2 /2 \right\} d\mu \\
&\propto \omega^{\alpha_0 + \frac{n}{2} - 1} exp\left\{ -\omega(2\beta_0 + \kappa_0\mu_0^2 + \sum_{i=1}^n y_i^2 - \frac{(\mu_0\kappa_0 + \sum_{i=1}^n y_i)^2}{\kappa_0 + n})/2 \right\}
\end{aligned}
$$

Therefore, $p(\omega|y_1, \ldots, y_n) \sim Gamma(\alpha_n, \beta_n)$, where $\alpha_n = \alpha_0 + n/2$ and $\beta_n = \frac{2\beta_0 + \kappa_0\mu_0^2 + \sum_{i=1}^n y_i^2 - \frac{(\mu_0\kappa_0 + \sum_{i=1}^n y_i)^2}{\kappa_0 + n}}{2}$.

$$
\begin{aligned}
p(\mu|\omega, y_1, \ldots, y_n) &= \frac{p(\mu, \omega, y_1, \ldots, y_n)}{p(\omega, y_1, \ldots, y_n)} \\
&= \frac{p(\mu, \omega|y_1, \ldots, y_n)p(y_1, \ldots, y_n)}{p(\omega|y_1, \ldots, y_n)p(y_1, \ldots, y_n)} \\
&= \frac{p(\mu, \omega|y_1, \ldots, y_n)}{p(\omega|y_1, \ldots, y_n)} \\
&\propto exp\left\{ -\omega(\kappa_0 + n)\left(\mu - \frac{\mu_0\kappa_0 + \sum_{i=1}^n y_i}{\kappa_0 + n}\right)^2 /2 \right\}
\end{aligned}
$$

which means that $p(\mu|\omega, y_1, \ldots, y_n) \sim Normal(\mu_n, \frac{1}{\omega\kappa_n})$, where $\mu_n = \frac{\mu_0\kappa_0 + \sum_{i=1}^n y_i}{\kappa_0 + n}$ and $\kappa_n = \kappa_0 + n$.

Exercise 2.3 *Show that the marginal distribution over μ 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 ν?

Proof:

$$
\begin{aligned}
p(\mu) &= \int p(\mu, \omega) d\omega = \int p(\mu|\omega)p(\omega) d\omega \\
&\propto \int \sqrt{\omega\kappa_0} e^{-\frac{\omega\kappa_0(\mu - \mu_0)^2}{2}} \times \omega^{\alpha_0 - 1} e^{-\beta_0\omega} d\omega \\
&\propto \frac{\Gamma(\alpha_0 + \frac{1}{2})}{\left(\beta_0 + \frac{\kappa_0(\mu - \mu_0)^2}{2}\right)^{\alpha_0 + \frac{1}{2}}} \\
&\propto \left(\beta_0 + \frac{\kappa_0(\mu - \mu_0)^2}{2}\right)^{-(\alpha_0 + \frac{1}{2})} \propto \left(1 + \frac{\kappa_0(\mu - \mu_0)^2}{2\beta_0}\right)^{-(\alpha_0 + \frac{1}{2})}
\end{aligned}
$$

where the location parameter, scale parameter, and degrees of freedom are $m = \mu_0$, $s = \sqrt{\frac{\beta_0}{\kappa_0\alpha_0}}$, and $\nu = 2\alpha_0$.

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

Solution: We know that

$$p(\mu,\omega|y_1,\ldots,y_n) = p(\mu,\omega|\mathbf{y}) = \frac{p(\mu,\omega,\mathbf{y})}{p(\mathbf{y})} = \frac{p(\mu|\omega,\mathbf{y})p(\omega|\mathbf{y})p(\mathbf{y})}{p(\mathbf{y})} = p(\mu|\omega,\mathbf{y})p(\omega|\mathbf{y})$$

From exercise 2.3, we know that if $\omega \sim Gamma(\alpha_0,\beta_0)$ and $\mu|\omega \sim Normal(\mu_0, 1/\omega\kappa_0)$, then the marginal distribution over μ is a centered, scaled t-distribution with location parameter $m = \mu_0$, scale parameter $s = \sqrt{\frac{\beta_0}{\kappa_0\alpha_0}}$, and degrees of freedom $\nu = 2\alpha_0$. Meanwhile, from exercise 2.2, we have $\omega|\mathbf{y} \sim Gamma(\alpha_n,\beta_n)$ and $\mu|\omega,\mathbf{y} \sim Normal(\mu_n, \frac{1}{\omega\kappa_n})$ with correspondent parameters. Therefore,

$$p(\mu|\mathbf{y}) \propto \left(1 + \frac{1}{\nu_n}\frac{(\mu-m_n)^2}{s_n^2}\right)^{-\frac{\nu_n+1}{2}}$$

where $m_n = \mu_n$, $s_n = \sqrt{\frac{\beta_n}{\kappa_n\alpha_n}}$, and $\nu_n = 2\alpha_n$

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

Solution:Let $\mathbf{y}^* = (y_{n+1},\ldots,y_{n+m})^T$ and $\mathbf{y} = (y_1,\ldots,y_m)^T$, so $\mathbf{y}^*$ is independent of $\mathbf{y}$, we have

$$\begin{aligned}
p(\mathbf{y}^*|\mathbf{y}) &= \iint p(\mathbf{y}^*,\mu,\omega|\mathbf{y})d\mu d\omega \\
&= \iint p(\mathbf{y}^*|\mu,\omega,\mathbf{y})p(\mu,\omega|\mathbf{y})d\mu d\omega \\
&= \iint p(\mathbf{y}^*|\mu,\omega)p(\mu,\omega|\mathbf{y})d\mu d\omega \\
&= \iint (\frac{\omega}{2\pi})^{\frac{m}{2}}\exp\left\{-\frac{\omega}{2}\sum_{i=1}^m(y_i^*-\mu)^2\right\} \times \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\omega^{\alpha_n-1}e^{-\beta_n\omega} \times (\frac{\omega\kappa_n}{2\pi})^{\frac{1}{2}}\exp\left\{-\frac{\omega\kappa_n}{2}(\mu-\mu_n)^2\right\}d\mu d\omega \\
&= (2\pi)^{-\frac{m}{2}}\frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\iint \omega^{\alpha_n+\frac{m}{2}-1}(\frac{\omega\kappa_n}{2\pi})^{\frac{1}{2}}e^{-\beta_n\omega}\exp\left\{-\frac{\omega}{2}\left[(m+\kappa_n)\mu^2 - 2(m\bar{y_i}^*+\mu_n\kappa_n)\mu + \frac{(m\bar{y_i}^*+\mu_n\kappa_n)^2}{m+\kappa_n}\right]\right\} \\
&\quad \times \exp\left\{-\frac{\omega}{2}\left[\sum(y_i^*)^2 + \kappa_n\mu_n^2 - \frac{(m\bar{y_i}^*+\mu_n\kappa_n)^2}{m+\kappa_n}\right]\right\}d\mu d\omega \\
&= (2\pi)^{-\frac{m}{2}}\frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\iint \omega^{\alpha_n+\frac{m}{2}-1}\exp\left\{-\frac{\omega}{2}\left[2\beta_n + \sum(y_i^*)^2 + \kappa_n\mu_n^2 - \frac{(m\bar{y_i}^*+\mu_n\kappa_n)^2}{m+\kappa_n}\right]\right\} \\
&\quad \times (\frac{\omega\kappa_n}{2\pi})^{\frac{1}{2}}\exp\left\{-\frac{\omega(m+\kappa_n)}{2}(\mu - \frac{m\bar{y_i}^*+\mu_n\kappa_n}{m+\kappa_n})\right\}d\mu d\omega \\
&= (2\pi)^{-\frac{m}{2}}\frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}(\frac{\kappa_n}{m+\kappa_n})^{\frac{1}{2}}\iint \omega^{\alpha_n+\frac{m}{2}-1}\exp\left\{-\frac{\omega}{2}\left[2\beta_n + \sum(y_i^*)^2 + \kappa_n\mu_n^2 - \frac{(m\bar{y_i}^*+\mu_n\kappa_n)^2}{m+\kappa_n}\right]\right\}d\omega \\
&= (2\pi)^{-\frac{m}{2}}\left(\frac{\kappa_n}{\kappa_{n+m}}\right)^{\frac{1}{2}}\frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\frac{\Gamma(\alpha_{n+m})}{\beta_{n+m}^{\alpha_{n+m}}}
\end{aligned}$$

where $\kappa_{n+m} = \kappa_n + m$, $\alpha_{n+m} = \alpha_n + \frac{m}{2}$, and $\beta_{n+m} = \beta_n + \left[\sum(y_i^*)^2 + \kappa_n\mu_n^2 - \frac{(m\bar{y_i}^*+\mu_n\kappa_n)^2}{m+\kappa_n}\right]/2.$

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

Solution: The marginal is a special case when $m = n$ and $n = 0$, so we can use exercise 2.5 to obtain the distribution:

$$p(\mathbf{y}) = (2\pi)^{\frac{n}{2}}(\kappa_n)^{\frac{1}{2}}\frac{\Gamma(\alpha_n)}{\beta_n^{\alpha_n}}$$

where $\kappa_n = \kappa_0 + n$, $\alpha_n = \alpha_0 + n$, and $\beta_n = \beta_0 + \left[\sum(y_i^*)^2 + \kappa_0\mu_0^2 - \frac{(n\bar{y_i}^* + \mu_0\kappa_0)^2}{n+\kappa_0}\right]/2$

2.2 Bayesian inference in a multivariate Gaussian model

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

$$y_i \sim \mathrm{N}(\mu, \Sigma)$$

for d-dimensional mean vector μ and $d \times d$ covariance matrix Σ.

We will put an *inverse Wishart* prior on Σ. The inverse Wishart distribution is a distribution over positive-definite matrices parametrized by $\nu_0 > d - 1$ degrees of freedom and positive definite matrix Λ_0^{-1}, with pdf

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

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

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

Solution: In the univariate case, $\Sigma = \sigma^2$, $\Lambda_0 = \lambda_0$, and $d = 1$, then $|\Lambda_0|^{-1} = {\lambda_0}^{-1}$ and $\mathrm{tr}(\Lambda_0\Sigma^{-1}) = \lambda_0/\sigma^2$. Also, $\Gamma_{d=1}(x) = \Gamma(x)$, so we have

$$p(\sigma^2|\nu_0, \lambda_0^{-1}) = \frac{\lambda_0^{\nu_0/2}}{2^{\nu_0/2}\Gamma(\nu_0/2)}(\sigma^2)^{-\frac{\nu_0}{2}-1}e^{-\frac{\lambda_0}{2\sigma^2}} = \frac{(\frac{\lambda_0}{2})^{\nu_0/2}}{\Gamma(\nu_0/2)}(\sigma^2)^{-\frac{\nu_0}{2}-1}e^{-\frac{\lambda_0}{2\sigma^2}}$$

which is a inverse gamma distribution with $\alpha = \frac{\nu_0}{2}$ and $\beta = \frac{\lambda_0}{2}$.

Exercise 2.8 *Let $\Sigma \sim$ Inv-Wishart(ν_0, Λ_0^{-1}) and $\mu|\Sigma \sim N(\mu_0, \Sigma/\kappa_0)$, so that*

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

and let

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

Show that $p(\mu, \Sigma|y_1, \ldots, y_n)$ is also normal-inverse Wishart distributed, and give the form of the updated parameters μ_n, κ_n, ν_n and Λ_n. It will be helpful to note that

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

Based on this, give interpretations for the prior parameters.

Solution:

$$\begin{aligned}p(\mu,\Sigma|\mathbf{y})&\propto p(\mathbf{y}|\mu,\Sigma)p(\mu,\Sigma)\\ &\propto\prod_{i=1}^{n}|\Sigma|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu)\right\}\cdot|\Sigma|^{-\frac{\nu_0+d}{2}-1}e^{-\frac{1}{2}\mathrm{tr}(\Lambda_0\Sigma^{-1})}\cdot\exp\left\{-\frac{\kappa_0}{2}(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)\right\}\\ &\propto|\Sigma|^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu)\right\}\cdot|\Sigma|^{-\frac{\nu_0+d}{2}-1}e^{-\frac{1}{2}\mathrm{tr}(\Lambda_0\Sigma^{-1})}\cdot\exp\left\{-\frac{\kappa_0}{2}(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)\right\}\end{aligned}$$

Now we focus on μ in the exponential part,

$$\begin{aligned}&\kappa_0(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)+\sum_{i=1}^{n}(\mu-y_i)^T\Sigma^{-1}(\mu-y_i)\\ =&(\kappa_0+n)\mu^T\Sigma^{-1}\mu-2(\kappa_0\mu_0^T\Sigma^{-1}+\sum_{i=1}^{n}y_i^T\Sigma^{-1})\mu+\sum_{i=1}^{n}y_i^T\Sigma^{-1}y_i+\kappa_0\mu_0^T\Sigma^{-1}\mu_0\\ =&(\kappa_0+n)\left(\mu-\frac{\kappa_0\mu_0+\sum_{i=1}^{n}y_i}{\kappa_0+n}\right)^T\Sigma^{-1}\left(\mu-\frac{\kappa_0\mu_0+\sum_{i=1}^{n}y_i}{\kappa_0+n}\right)\\ &+tr\left(\sum_{i=1}^{n}y_iy_i^T+\kappa_0\mu_0\mu_0^T-\frac{(\kappa_0\mu_0+\sum_{i=1}^{n}y_i)(\kappa_0\mu_0+\sum_{i=1}^{n}y_i)^T}{\kappa_0+n}\right)\Sigma^{-1}\end{aligned}$$

Now let

$$\begin{aligned}\mu_n&=\frac{\kappa_0\mu_0+\sum_{i=1}^{n}y_i}{\kappa_0+n}\\ \kappa_n&=\kappa_0+n\\ \nu_n&=\nu_0+n\\ \Lambda_n&=\Lambda_0+\sum_{i=1}^{n}y_iy_i^T+\kappa_0\mu_0\mu_0^T-\frac{(\kappa_0\mu_0+\sum_{i=1}^{n}y_i)(\kappa_0\mu_0+\sum_{i=1}^{n}y_i)^T}{\kappa_0+n}\end{aligned}$$

Then we have

$$p(\mu,\Sigma|\mathbf{y})\propto|\Sigma|^{\frac{\nu_n+d+1}{2}}\exp\left\{-\frac{1}{2}tr(\Lambda_n\Sigma^{-1})+\frac{\kappa_n}{2}(\mu-\mu_n)^T\Sigma^{-1}(\mu-\mu_n)\right\}$$

which is the normal-inverse Wishart distribution with parameters μ_n, κ_n, ν_n and Λ_n.

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 Λ 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* $p(\beta|\omega, y_1, \ldots, y_n)$

Solution:

$$\begin{aligned}
p(\beta|\omega, \mathbf{y}) &= \frac{p(\mathbf{y}|X, \beta, \omega)p(\beta|\omega)p(\omega)}{p(\mathbf{y}, \omega)} \\
&\propto p(\mathbf{y}|X, \beta, \omega)p(\beta|\omega) \\
&\propto \exp\left\{-\frac{1}{2}\omega(\mathbf{y} - X\beta)^T\Lambda(\mathbf{y} - X\beta)\right\} \cdot \exp\left\{-\frac{1}{2}\omega(\beta - \mu)^T K(\beta - \mu)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\omega[\beta^T(X^T\Lambda X + K)\beta - 2\beta^T(X^T\Lambda\mathbf{y} + K\mu)]\right\} \cdot \exp\left\{-\frac{1}{2}\omega(\mathbf{y}^T\Lambda\mathbf{y} + \mu^T K\mu)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\omega[\beta - (X^T\Lambda X + K)^{-1}(X^T\Lambda\mathbf{y} + K\mu)]^T(X^T\Lambda X + K)[\beta - (X^T\Lambda X + K)^{-1}(X^T\Lambda\mathbf{y} + K\mu)]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\omega(\beta - \mu_p)^T K_p(\beta - \mu_p)\right\}
\end{aligned}$$

Therefore,

$$\beta|\omega, \mathbf{y} \sim \mathcal{N}(\mu_p, (\omega K_p)^{-1})$$

where $\mu_p = (X^T\Lambda X + K)^{-1}(X^T\Lambda\mathbf{y} + K\mu)$ and $K_p = X^T\Lambda X + K$.

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

Solution:

$$\begin{aligned}
p(\omega|\mathbf{y}) &\propto p(\mathbf{y}, \omega) \\
&= \int p(\mathbf{y}, \beta, \omega)d\beta \\
&= \int p(\mathbf{y}|X, \beta, \omega)p(\beta|\omega)p(\omega)d\beta \\
&\propto \frac{b^a}{\Gamma(a)}\omega^{a-1}e^{-b\omega}\int \omega^{\frac{n+p}{2}} \exp\left\{-\frac{1}{2}\omega(\beta - \mu_p)^T K_n(\beta - \mu_p)\right\} \exp\left\{-\frac{1}{2}\omega(\mathbf{y}^T\Lambda\mathbf{y} + \mu^T K\mu - \mu_p^T K_p\mu_p)\right\} d\beta \\
&\propto \omega^{a+\frac{n}{2}-1} \exp\left\{-\omega[b + \frac{\mathbf{y}^T\Lambda\mathbf{y} + \mu^T K\mu - \mu_p^T K_p\mu_p}{2}]\right\}
\end{aligned}$$

where $\mu_p = (X^T\Lambda X + K)^{-1}(X^T\Lambda\mathbf{y} + K\mu)$ and $K_p = X^T\Lambda X + K$.Therefore,

$$\omega|\mathbf{y} \sim Gamma(a + \frac{n}{2}, b + \frac{\mathbf{y}^T\Lambda\mathbf{y} + \mu^T K\mu - \mu_p^T K_p\mu_p}{2})$$

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

Solution:

$$\begin{aligned}
p(\beta|\mathbf{y}) &\int p(\mathbf{y}|X,\beta,\omega)p(\beta|\omega)p(\omega)d\omega \\
&\propto \int \omega^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\omega(\mathbf{y}-X\beta)^T\Lambda(\mathbf{y}-X\beta)\right\} \cdot \omega^{\frac{p}{2}} \exp\left\{-\frac{1}{2}\omega(\beta-\mu)^T K(\beta-\mu)\right\} \cdot \omega^{a-1}e^{-b\omega}d\omega \\
&\propto \int \omega^{a+\frac{n+p}{2}-1} \exp\left\{-[\frac{1}{2}(\mathbf{y}-X\beta)^T\Lambda(\mathbf{y}-X\beta)+\frac{1}{2}(\beta-\mu)^T K(\beta-\mu)+b]\omega\right\} d\omega \\
&\propto \left\{1+\frac{(\mathbf{y}-X\beta)^T\Lambda(\mathbf{y}-X\beta)+(\beta-\mu)^T K(\beta-\mu)}{2b}\right\}^{-a+\frac{n+p}{2}}
\end{aligned}$$

which can be seen as non-central t distribution as discussed in exercise 2.4.

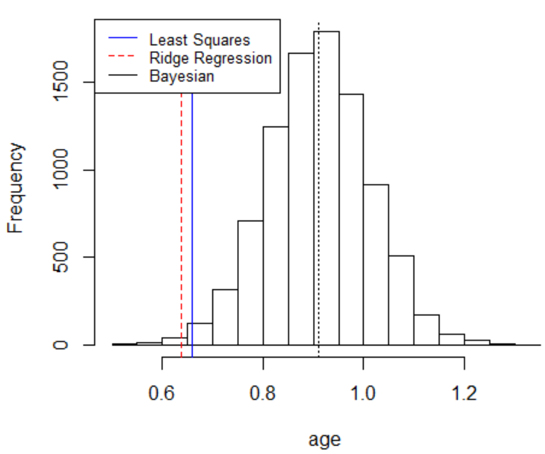
Exercise 2.12 *Download the dataset dental.csv from Github. This dataset measures a dental distance (specifically, the distance between the center of the pituitary to the pterygomaxillary fissure) in 27 children. Add a column of ones to correspond to the intercept. Fit the above Bayesian model to the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

Solution: We know that

$$\beta|\omega,\mathbf{y} \sim \mathcal{N}(\mu_p,(\omega K_p)^{-1})$$

$$\omega|\mathbf{y} \sim Gamma(a+\frac{n}{2}, b+\frac{\mathbf{y}^T\Lambda\mathbf{y}+\mu^T K\mu-\mu_p^T K_p\mu_p}{2})$$

with $\mu_p = (X^T\Lambda X+K)^{-1}(X^T\Lambda\mathbf{y}+K\mu)$ and $K_p = X^T\Lambda X+K$. I set $\Lambda = I$, $K = I$, $a = 2$, and $b = 1$ for the Bayesian model. For ridge regression, I choose tuning parameter $\lambda = 0.1$ (Without loss of generality, I did not perform cross-validation to choose the best λ that minimizing the mean squared error). The estimated values of parameters and comparison of resulting plots are shown in Table 1 and Figure 1.



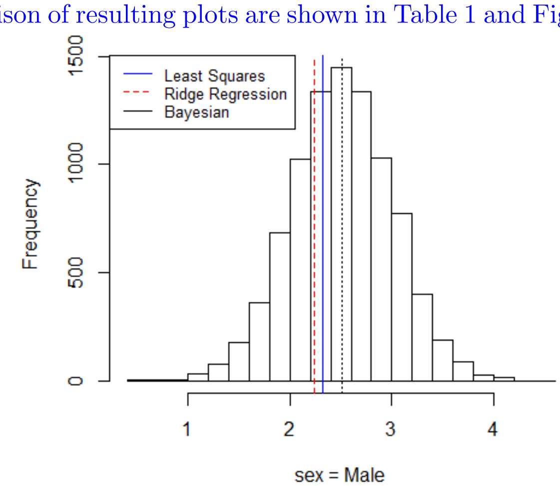


Figure 1: Comparison of the estimated parameters of **age** and **sex**

Model	intercept (β_1)	age (β_2)	sex = Male (β_3)
Least Squares	15.3857	0.6602	2.3210
Ridge Regression ($\lambda = 0.1$)	15.6722	0.6383	2.2440
Bayesian	12.4059	0.9111	2.5106

Table 1: Estimated $\beta = (\beta_1, \beta_2, \beta_3)$ for different regression models

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

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

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

Solution: Since λ_i is only dependent on y_i, we have

$$\begin{aligned}
p(\lambda_i|\mathbf{y}, \beta, \omega) = p(\lambda_i|y_i, \beta, \omega) &\propto p(y_i|\lambda_i, \beta, \omega)p(\lambda_i) \\
&\propto (\omega\lambda_i)^{1/2} e^{-\frac{\omega\lambda_i}{2}(y_i - x_i^T\beta)^2} \cdot \lambda_i^{\tau-1} e^{-\tau\lambda_i} \\
&\propto \lambda_i^{\tau+\frac{1}{2}-1} \exp\left\{-[\tau + \frac{\omega(y_i - x_i^T\beta)^2}{2}]\lambda_i\right\}
\end{aligned}$$

Therefore,

$$\lambda_i \sim Gamma\left(\tau + \frac{1}{2}, \tau + \frac{\omega(y_i - x_i^T\beta)^2}{2}\right)$$

Exercise 2.14 *Write a Gibbs sampler that alternates between sampling from the conditional posteriors of λ_i, β and ω, 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.*