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 $\omega(/\sigma^2)$ – 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?*

Proof: (I used k consistently instead of κ.) For prior we have:

$$\begin{aligned}p(\mu, \omega) &= p(\mu|\omega)p(\omega)\\ &\propto \omega^{\alpha_0-1}e^{-\beta_0\omega}\sqrt{\omega k_0}e^{-\frac{\omega k_0(\mu-\mu_0)^2}{2}}\end{aligned} \tag{2.1}$$

For posterior we have:

$$\begin{aligned}p(\mu, \omega|y_1, ..., y_n) &\propto p(\mu|\omega)p(\omega)p(y_1, ..., y_n|\mu, \omega)\\ &\propto \omega^{\alpha_0-1}e^{-\beta_0\omega}\sqrt{\omega k_0}e^{-\frac{\omega k_0(\mu-\mu_0)^2}{2}}\prod_i \sqrt{\omega}e^{-\frac{(y_i-\mu)^2\omega}{2}}\\ &\propto \omega^{\alpha_0+n/2-1/2}e^{-(2\beta_0+k_0(\mu-\mu_0)^2+\sum(y_i-\mu)^2)\omega/2}\\ &\propto \omega^{\alpha_0+n/2-1/2}\exp\left\{-\left[k_0\mu_0^2+\sum y_i^2+2\beta_0-\frac{(k_0\mu_0+\sum y_i)^2}{k_0+n}\right]\omega/2\right\}\cdots\\ &\exp\{-\left[(k_0+n)(\mu-\frac{k_0\mu_o+\sum y_i}{k_0+n})^2\right]\omega/2\}\end{aligned} \tag{2.2}$$

So, $\alpha_n = \alpha_0 + n/2$, $\beta_n = \frac{k_0\mu_0^2 + \sum y_i^2 + 2\beta_0 - \frac{(k_0\mu_0 + \sum y_i)^2}{k_0+n}}{2}$, $\mu_n = \frac{k_0\mu_0 + \sum y_i}{k_0+n}$, and $k_n = k_0 + n$.

■

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?*

Proof:

$$\begin{aligned}
p(\omega|y_1, ..., y_n) &= \int p(\mu, \omega|y_1, ..., y_n) d\mu \\
&\propto \int \omega^{\alpha_0 + n/2 - 1/2} \exp\left\{-(2\beta_0 + k_0(\mu - \mu_0)^2 + \sum(y_i - \mu)^2)\omega/2\right\} d\mu \\
&\propto \omega^{\alpha_0 + n/2 - 1/2} \exp\left\{-\left[k_0\mu_0^2 + \sum y_i^2 + 2\beta_0 - \frac{(k_0\mu_0 + \sum u_i)^2}{k_0 + n}\right]\omega/2\right\} \cdots \\
&\int \exp\{-\left[(k_0 + n)(\mu - \frac{k_0\mu_o + \sum y_i}{k_0 + n})^2\right]\omega/2\} d\mu \\
&\propto \omega^{\alpha_0 + n/2 - 1} \exp\left\{-\left[k_0\mu_0^2 + \sum y_i^2 + 2\beta_0 - \frac{(k_0\mu_0 + \sum y_i)^2}{k_0 + n}\right]\omega/2\right\}
\end{aligned} \tag{2.3}$$

Therefore, $p(\omega|y_1, ..., y_n) \sim \text{Gamma}(\alpha_n, \beta_n)$, where $\alpha_n = \alpha_0 + n/2$, and $\beta_n = \frac{k_0\mu_0^2 + \sum y_i^2 + 2\beta_0 - \frac{(k_0\mu_0 + \sum y_i)^2}{k_0+n}}{2}$. From equation 2.2, we can see that $p(\mu|\omega, y_1, ..., y_n) \sim \text{Normal}(\mu_n, \frac{1}{\omega k_n})$, where $\mu_n = \frac{k_0\mu_0 + \sum y_i}{k_0+n}$, and $k_n = k_0 + n$.

■

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:

$$\begin{aligned}
p(\mu) = \int p(\mu|\omega)p(\omega)d\omega &\propto \int \omega^{\alpha_0 - 1/2} \exp\left\{-(\beta_0 + k_0(\mu - \mu_0)^2/2)\omega\right\} d\omega \\
&\propto (\beta_0 + k_0(\mu - \mu_0)^2/2)^{-(\alpha_0 + 1/2)} \propto (1 + k_0(\mu - \mu_0)^2/2\beta_0)^{-(\alpha_0 + 1/2)}
\end{aligned} \tag{2.4}$$

Therefore, $m = \mu_0$, $s = \sqrt{\beta_0/k_0\alpha_0}$, 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.*

Proof: From the last exercise, we know that if $p(\mu|\omega) \sim \text{Normal}(\mu_0, \frac{1}{\omega k_0})$ and $p(\omega) \sim \text{Gamma}(\alpha_0, \beta_0)$, then we have $p(\mu) \sim t(m = \mu_0, s = \sqrt{\beta_0/k_0\alpha_0}, \nu = 2\alpha_0)$.

From exercise 2.1, we know that $p(\mu|\omega, y_1, ..., y_n) \sim \text{Normal}(\mu_n, \frac{1}{\omega k_n})$ and $p(\omega|y_1, ..., y_n) \sim \text{Gamma}(\alpha_n, \beta_n)$, then we have $p(\mu) \sim t(m = \mu_n, s = \sqrt{\beta_n/k_n\alpha_n}, \nu = 2\alpha_n)$.

■

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

Proof:

$$\begin{aligned}
p(y_{n+1}, ..., y_{n+m}|y_1, ..., y_n) &= \int p(y_{n+1}, ..., y_{n+m}|\mu, \omega)p(\mu, \omega|y_1, ..., y_n)d\mu d\omega \\
&\propto \int \omega^{\alpha_n+m/2-1/2} \exp\left\{-(2\beta_n + k_n(\mu-\mu_n)^2 + \sum_{i=n+1}^{n+m}(y_i-\mu)^2)\omega/2\right\} d\mu d\omega \\
&\propto \int \omega^{\alpha_n+m/2-1/2} \exp\left\{-\left[k_n\mu_n^2 + \sum_{i=n+1}^{n+m} y_i^2 + 2\beta_n - \frac{(k_n\mu_n + \sum_{i=n+1}^{n+m} y_i)^2}{k_n+m}\right]\omega/2\right\} \cdots \\
&\int \exp\{-\left[(k_n+m)(\mu - \frac{k_n\mu_n + \sum_{i=n+1}^{n+m} y_i}{k_n+m})^2\right]\omega/2\}d\mu d\omega \\
&\propto \int \omega^{\alpha_n+m/2-1} \exp\left\{-\left[k_n\mu_n^2 + \sum_{i=n+1}^{n+m} y_i^2 + 2\beta_n - \frac{(k_n\mu_n + \sum_{i=n+1}^{n+m} y_i)^2}{k_n+m}\right]\omega/2\right\} d\omega \\
&\propto \left[k_n\mu_n^2 + \sum_{i=n+1}^{n+m} y_i^2 + 2\beta_n - \frac{(k_n\mu_n + \sum_{i=n+1}^{n+m} y_i)^2}{k_n+m}\right]^{-(\alpha_n+m/2)}
\end{aligned} \tag{2.5}$$

■

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

Proof:

$$\begin{aligned}
p(y_1, ..., y_n) &= \int p(y_1, ..., y_n|\mu, \omega)p(\mu, \omega)d\mu d\omega \\
&\propto \int \omega^{\alpha_0+n/2-1/2} \exp\left\{-(2\beta_0 + k_0(\mu-\mu_0)^2 + \sum_{i=1}^{n}(y_i-\mu)^2)\omega/2\right\} d\mu d\omega \\
&\propto \int \omega^{\alpha_0+n/2-1/2} \exp\left\{-\left[k_0\mu_0^2 + \sum_{i=1}^{n} y_i^2 + 2\beta_0 - \frac{(k_0\mu_0 + \sum_{i=1}^{n} y_i)^2}{k_0+n}\right]\omega/2\right\} \cdots \\
&\int \exp\{-\left[(k_0+n)(\mu - \frac{k_0\mu_0 + \sum_{i=1}^{n} y_i}{k_0+n})^2\right]\omega/2\}d\mu d\omega \\
&\propto \int \omega^{\alpha_0+n/2-1} \exp\left\{-\left[k_0\mu_0^2 + \sum_{i=1}^{n} y_i^2 + 2\beta_0 - \frac{(k_0\mu_0 + \sum_{i=1}^{n} y_i)^2}{k_0+n}\right]\omega/2\right\} d\omega \\
&\propto \left[k_0\mu_0^2 + \sum_{i=1}^{n} y_i^2 + 2\beta_0 - \frac{(k_0\mu_0 + \sum_{i=1}^{n} y_i)^2}{k_0+n}\right]^{-(\alpha_0+n/2)}
\end{aligned} \tag{2.6}$$

■

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

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

Proof: In the univariate case, we have Λ_0 and Σ to be scalars, and the pdf becomes:

$$p(\Sigma|\nu_0, \Lambda_0^{-1}) = \frac{\Lambda_0^{\nu_0/2}}{2^{\nu_0/2}\Gamma(\nu_0/2)}\Sigma^{-\frac{\nu_0+2}{2}}e^{-\frac{1}{2}(\Lambda_0\Sigma^{-1})}, \text{when } \Sigma > 0.$$

So $\Sigma \sim \mathrm{InvGamma}(\alpha = \nu_0/2, \beta = \Lambda_0/2)$. ■

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*

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

and let

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

Show that $p(\mu, \Sigma|y_1, \dots, y_n)$ 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.

Proof:

$$
\begin{aligned}
&(\mu, \Sigma | y_1, \ldots, y_n) \propto p(y|\mu,\Sigma)p(\mu,\Sigma) \\
&\propto |\Sigma|^{-\frac{\nu_0+d+2}{2}} e^{-\frac{1}{2}\mathrm{tr}(\Lambda_0\Sigma^{-1}) - \frac{\kappa_0}{2}(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)} \prod_{i=1}^{n} |\Sigma|^{-1/2} e^{-\frac{(y_i-\mu)^T\Sigma^{-1}(y_i-\mu)}{2}} \\
&\propto |\Sigma|^{-(\nu_0+d+n+2)/2} \ldots \\
&\exp\left\{-\frac{1}{2}\left(\mathrm{tr}(\Lambda_0\Sigma^{-1}) + \kappa_0(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0) + \sum_i (y_i-\bar{y})^T\Sigma^{-1}(y_i-\bar{y}) + n(\bar{y}-\mu)^T\Sigma^{-1}(\bar{y}-\mu)\right)\right\}
\end{aligned} \tag{2.7}
$$

Now,

$$
\begin{aligned}
&\exp\left\{-\frac{1}{2}\left(\kappa_0(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0) + n(\bar{y}-\mu)^T\Sigma^{-1}(\bar{y}-\mu)\right)\right\} \\
&= \exp\left\{-\frac{1}{2}\left(\mu^T\Sigma^{-1}(\kappa_0+n)\mu - 2(\kappa_0\mu_0^T\Sigma^{-1} + n\bar{y}\Sigma^{-1})\mu + \kappa_0\mu_0^T\Sigma^{-1}\mu_0 + n\bar{y}^T\Sigma^{-1}\bar{y}\right)\right\} \\
&= \exp\left\{-\frac{1}{2}\left((\kappa_0+n)(\mu-\mu_n)^T\Sigma^{-1}(\mu-\mu_n) - (\kappa_0+n)\mu_n^T\Sigma^{-1}\mu_n + \kappa_0\mu_0^T\Sigma^{-1}\mu_0 + n\bar{y}^T\Sigma^{-1}\bar{y}\right)\right\}, \\
&= \exp\left\{-\frac{1}{2}\left((\kappa_0+n)(\mu-\mu_n)^T\Sigma^{-1}(\mu-\mu_n) + \mathrm{tr}\left(\kappa_0\mu_0\mu_0^T\Sigma^{-1} + n\bar{y}\bar{y}^T\Sigma^{-1} - (\kappa_0+n)\mu_n\mu_n^T\Sigma^{-1}\right)\right)\right\}
\end{aligned} \tag{2.8}
$$

where $\mu_n = (\kappa_0\mu_0 + n\bar{y})/(\kappa_0+n)$.

Substitute into equation 2.7, we have

$$
\begin{aligned}
&(\mu, \Sigma | y_1, \ldots, y_n) \propto |\Sigma|^{-(\nu_0+d+n+2)/2} \ldots \\
&\exp\left\{-\frac{1}{2}\left(\mathrm{tr}(\left[\Lambda_0 + \sum_i (y_i-\bar{y})(y_i-\bar{y})^T + \kappa_0\mu_0\mu_0^T + n\bar{y}\bar{y}^T - (\kappa_0+n)\mu_n\mu_n^T\right]\Sigma^{-1}) + (\kappa_0+n)(\mu-\mu_n)^T\Sigma^{-1}(\mu-\mu_n)\right)\right\}
\end{aligned} \tag{2.9}
$$

Therefore, the posterior is normal-inverse Wishart distributed, with $\mu_n = (\kappa_0\mu_0 + n\bar{y})/(\kappa_0+n), \kappa_n = \kappa_0 + n, \nu_n = \nu_0 + n, \Lambda_n = \Lambda_0 + \sum_i (y_i-\bar{y})(y_i-\bar{y})^T + \kappa_0\mu_0\mu_0^T + n\bar{y}\bar{y}^T - (\kappa_0+n)\mu_n\mu_n^T$.

∎

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.

$$\begin{aligned} p(\omega, \beta|y_1, ..., y_n) &\propto p(\omega)p(\beta|\omega)p(y_1, ..., y_n|\beta, \omega) \\ &\propto \omega^{a+(d+n)/2-1}e^{-b\omega}e^{-\frac{(\beta-\mu)^T\omega K(\beta-\mu)}{2}}e^{-\frac{(y-X\beta)^T\omega\Lambda(y-X\beta)}{2}} \end{aligned} \tag{2.10}$$

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

Proof:

$$\begin{aligned} p(\beta|\omega, y_1, ..., y_n) &\propto e^{-\frac{(\beta-\mu)^T\omega K(\beta-\mu)}{2}}e^{-\frac{(y-X\beta)^T\omega\Lambda(y-X\beta)}{2}} \\ &\propto \exp\left\{-\frac{1}{2}\left(\beta^T(\omega K + X^T\omega\Lambda X)\beta - 2(\mu^T\omega K + y^T\omega\Lambda X)\beta\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}(\beta-\mu_n)^T\Lambda_n(\beta-\mu_n)\right\} \end{aligned} \tag{2.11}$$

Therefore, $\beta|\omega, y_1, ..., y_n \sim \mathcal{N}(\mu_n, \Lambda_n^{-1})$, where $\mu_n = \Lambda_n^{-1}(\omega K\mu + \omega X^T\Lambda y)$, and $\Lambda_n = \omega K + \omega X^T\Lambda X$ ■

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

Proof:

$$\begin{aligned} p(\omega|y_1, ..., y_n) \propto& \int p(\omega)p(\beta|\omega)p(y_1, ..., y_n|\beta, \omega)d\beta \\ \propto& \omega^{a+(d+n)/2-1}e^{-b\omega}\int e^{-\frac{(\beta-\mu)^T\omega K(\beta-\mu)}{2}}e^{-\frac{(y-X\beta)^T\omega\Lambda(y-X\beta)}{2}}d\beta \\ \propto& \omega^{a+(d+n)/2-1}e^{-b\omega}\int \exp\left\{-\frac{1}{2}\left((\beta-\mu_n)^T\Lambda_n(\beta-\mu_n) - \mu_n^T\Lambda_n\mu_n + \omega\mu^TK\mu + \omega y^T\Lambda y\right)\right\}d\beta \\ \propto& \omega^{a+(d+n)/2-1}e^{-b\omega}\int \exp\left\{-\frac{1}{2}\left((\beta-\mu_n)^T\Lambda_n(\beta-\mu_n) - \mu_n^T\Lambda_n\mu_n + \omega\mu^TK\mu + \omega y^T\Lambda y\right)\right\}d\beta \\ \propto& \omega^{a+n/2-1}e^{-b\omega}\exp\left\{-\frac{\omega}{2}\left(\mu^TK\mu + y^T\Lambda y - (K\mu + X^T\Lambda y)^T(K + X^T\Lambda X)^{-1}(K\mu + X^T\Lambda y)\right)\right\} \end{aligned} \tag{2.12}$$

Therefore, $\omega|y_1, ..., y_n \sim \text{Gamma}(a+n/2, b+(\mu^TK\mu+y^T\Lambda y-(K\mu+X^T\Lambda y)^T(K+X^T\Lambda X)^{-1}(K\mu+X^T\Lambda y))/2)$
■

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

Proof:

$$\begin{aligned} p(\beta|y_1,...,y_n) &\propto \int p(\omega)p(\beta|\omega)p(y_1,...,y_n|\beta,\omega)d\omega \\ &\propto \int \omega^{a+(d+n)/2-1}e^{-\omega(b+\frac{(\beta-\mu)^TK(\beta-\mu)+(y-X\beta)^T\Lambda(y-X\beta)}{2})}d\omega \\ &\propto \frac{\Gamma(a+(d+n)/2)}{(b+\frac{(\beta-\mu)^TK(\beta-\mu)+(y-X\beta)^T\Lambda(y-X\beta)}{2})^{a+(n+d)/2}} \end{aligned} \tag{2.13}$$

■

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?*

Proof: After sampling from the posterior distribution, we find (shown in Figure 2.1)the samples for β concentrate around the ridge regression estimator(with the $\lambda = 1$) and they could be far away from least square estimator.

■

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, $p(\lambda_i|\mathbf{y},\beta,\omega)$?*

Proof:

$$\begin{aligned} p(\lambda_i|y_1,...,y_n,\beta,\omega) &\propto \lambda_i^{\tau-1}e^{-\tau\lambda_i}\lambda_i^{1/2}e^{-\omega(y-X\beta)_i^2\lambda_i/2} \\ &\propto \lambda_i^{\tau-1/2}e^{-(\tau+\omega(y-X\beta)_i^2/2)\lambda_i} \end{aligned} \tag{2.14}$$

Therefore, $\lambda_i|y_1,...,y_n,\beta,\omega \sim \mathrm{Gamma}(\tau+1/2,\tau+\omega(y-X\beta)_i^2/2)$ ■

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.*

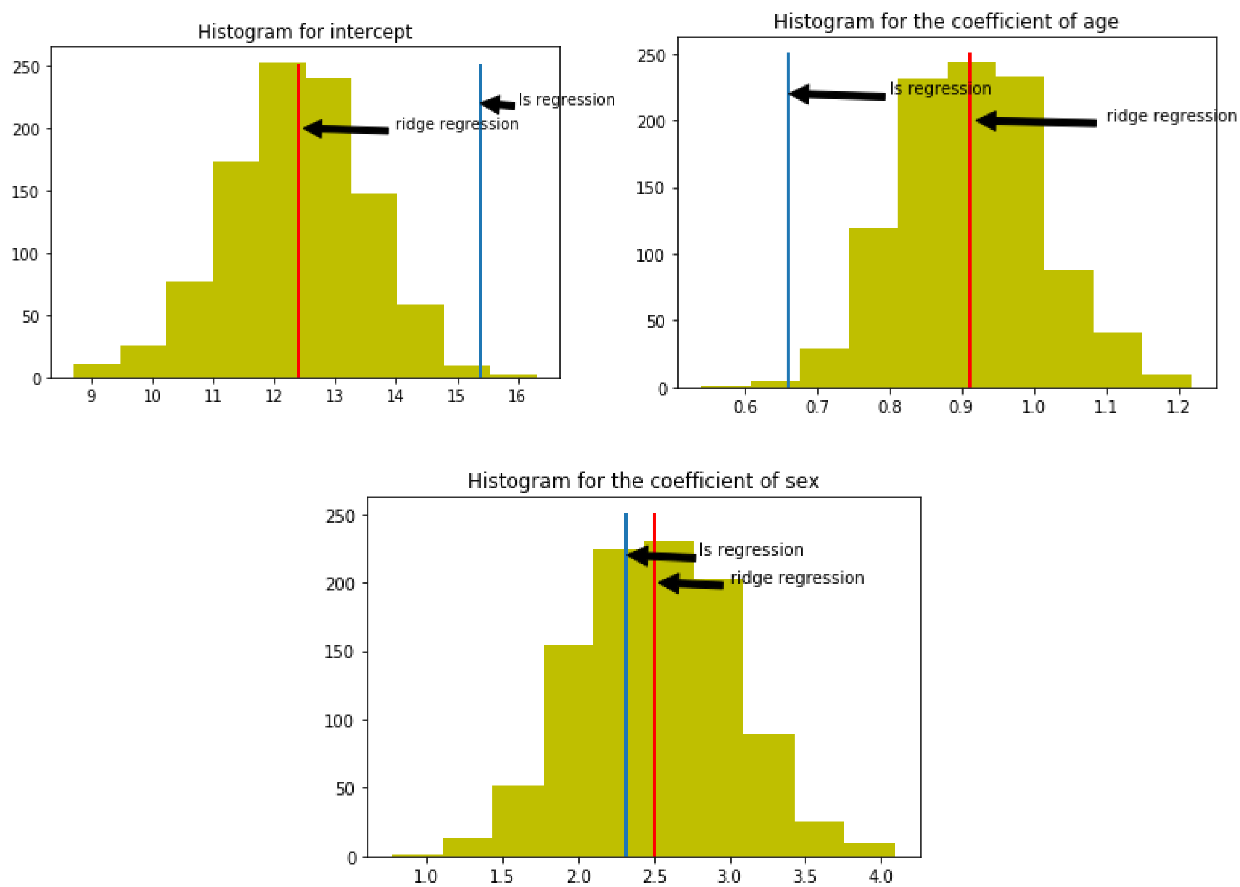

Figure 2.1: Histograms of posterior samples with fixed Λ.

Coefficp4ient	age	sex	intercept
All	0.02333572	0.51532875	3.091482356
Male	0.03844271	7.75366998	7.44173769
Female	0.05427498	13.71422947	6.51420431

Table 2.1: Variance of posterior samples for each variable

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.*

Proof: After sampling from the posterior distribution, we find (shown in Figure 2.2) the samples for β concentrate somewhere between the ridge regression estimator and least square estimator. Also, compared with the histogram shown in Figure 2.1, the posterior of β in this model has higher variance, so that both ridge estimator and least squared estimator are in the concentration area.

Assumption: the prior of $\beta|\omega \sim \mathcal{N}(\mu, (\omega K)^{-1})$, assumes that the variance of β_i is the same across all i. However, we can take a look at the variance in the Table 2.1. It shows that the variances are not the same fro different variables. Therefore, it would natural to put a prior on K. In the simplest case, we can do $K = \text{diag}(k_1, ..., k_p)$, $k_i \overset{i.i.d.}{\sim} \text{Gamma}(\tau, \tau)$.

$$\begin{aligned} p(k_i|y_1, ..., y_n, \beta, \omega, \Lambda) &\propto k_i^{\tau-1} e^{-\tau k_i} k_i^{1/2} e^{-\omega(\beta_i - \mu_i)^2 k_i/2} \\ &\propto k_i^{\tau-1/2} e^{-(\tau+\omega(\beta_i-\mu_i)^2/2)k_i} \end{aligned} \tag{2.15}$$

Therefore, $k_i|y_1, ..., y_n, \beta, \omega, \Lambda \sim \text{Gamma}(\tau + 1/2, \tau + \omega(\beta_i - \mu_i)^2/2)$

The results are shown in Figure 2.3. The variance of each variable has become larger.

Side notes: we assume the variance of residual is the same for male and female, which is not the case. Therefore, we can change the model to $y_i = \beta_0 + \beta_1 age + \beta_2 I_m + \beta_3 I_m age + \epsilon_i + I_m \rho_i + I_f \rho_i$. ■

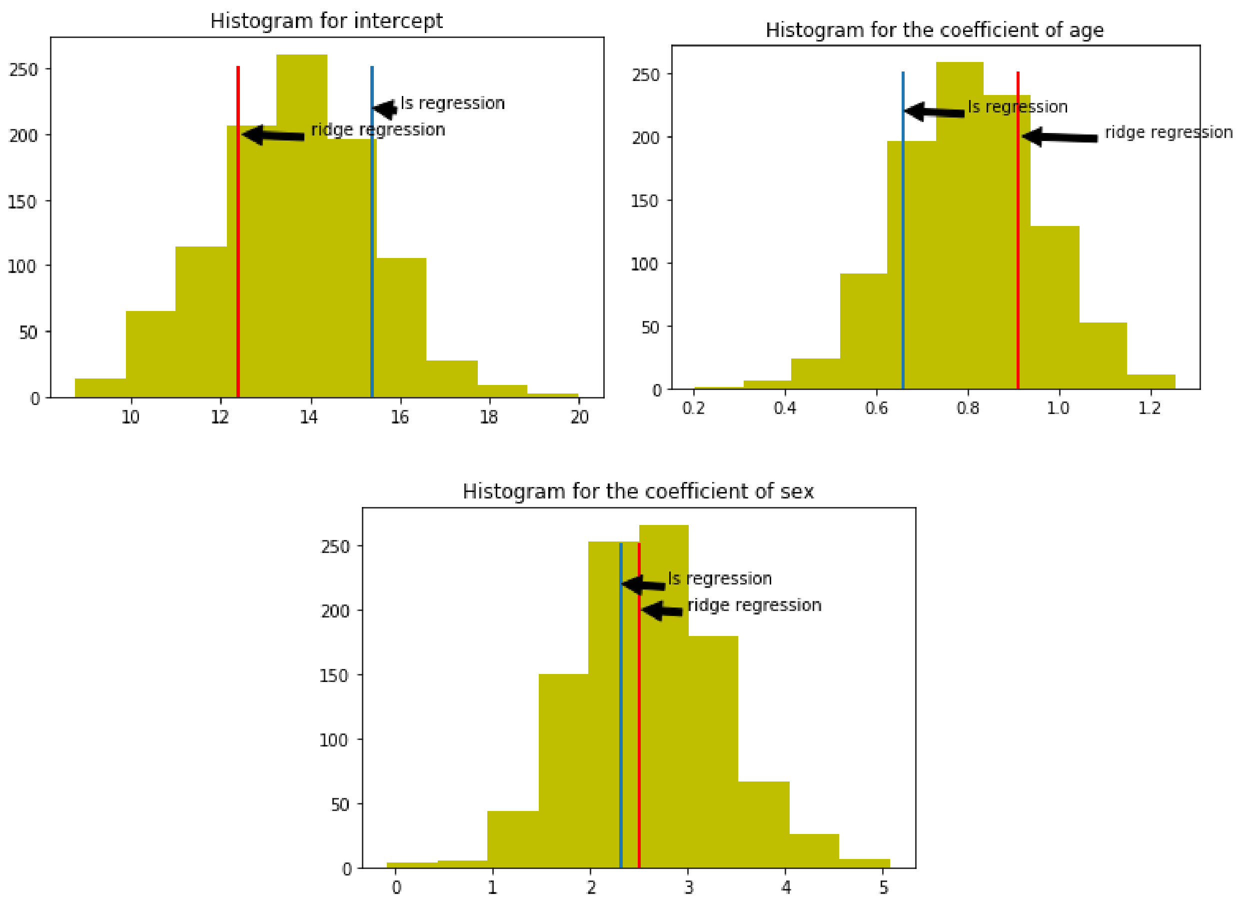

Figure 2.2: Histograms of posterior samples with gamma-priored Λ.

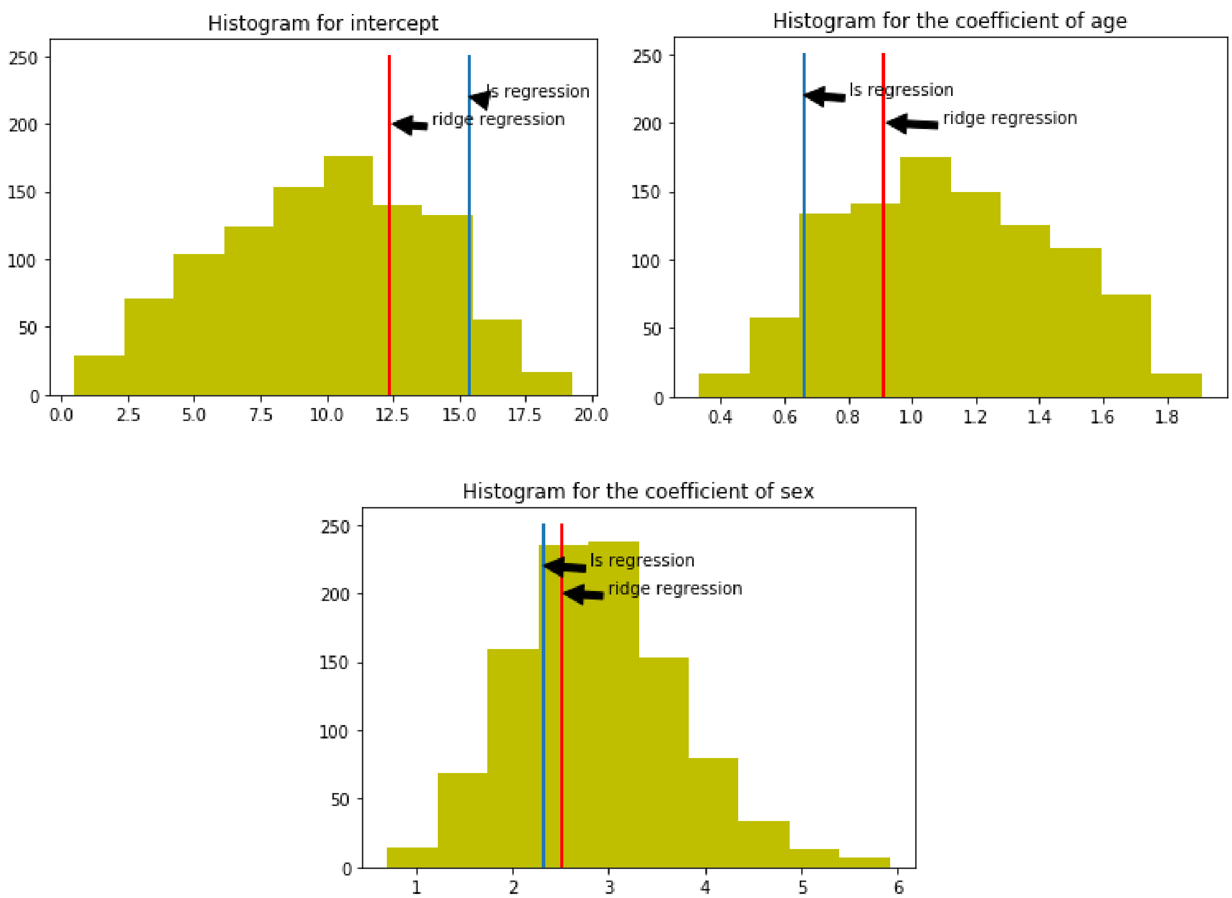

Figure 2.3: Histograms of posterior samples with gamma-priored Λ.