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 $p(\mu,\omega|y_1,\dots,y_n)$ (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 α_n,β_n,μ_n and κ_n?*

Answer:

$$p(\mu,\omega|y) = NG(\mu,\omega|\mu_o,\kappa_o,\alpha_o,\beta_o)p(y_i|\mu,\omega) \tag{2.1}$$

$$p(\mu,\omega|y) = \omega^{1/2}\omega^{\alpha_o+n/2-1}e^{(-\omega/2)[\kappa_o(\mu-\mu_o)^2+\sum_i(x_i-\mu)^2]} \tag{2.2}$$

$$\sum_{i=1}^{n}(x_i-\mu)^2 = \dots = n(\mu-\bar{x})^2+\sum_i(x_i-\bar{x})^2 \tag{2.3}$$

$$\kappa_o(\mu-\mu_o)^2+n(\mu-\bar{x}) = (\kappa_o+n)(\mu-\mu_n)^2+\frac{\kappa_o n(\bar{x}-\mu_o)^2}{\kappa_o+n} \tag{2.4}$$

$$\kappa_o(\mu-\mu_o)^2+\sum_i(x_i-\mu)^2 = \kappa_o(\mu-\mu_o)^2+n(\mu-\bar{x})+\sum_i(x_i-\bar{x})^2 \tag{2.5}$$

$$= (\kappa_o+n)(\mu-\mu_n)^2+\frac{\kappa_o n(\bar{x}-\mu_o)^2}{\kappa_o+n}+\sum_i(x_i-x)^2 \tag{2.6}$$

$$p(\mu,\omega|y_i)\ \alpha\ (\omega^{1/2}e^{(\omega/2)(\kappa_o+n)(\mu-\mu_n)^2})(\omega^{\alpha_o+n/2-1}e^{\beta_o\omega-\omega/2\sum_i(x_i-\bar{x})^2}e^{\omega/2})\frac{\kappa_o n(\bar{x}-\mu_o)^2}{\kappa_o+n} \tag{2.7}$$

Of the form:

$$p(\mu,\omega|y_i) = NG(\mu,\omega|\mu_n,\kappa_n,\alpha_n,\beta_n) \tag{2.8}$$

With parameters $\beta_n = \beta_o + (1/2)\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{\kappa_o n(\bar{x}-\mu_o)^2}{2(\kappa_o+n)}$, $\alpha_n = \alpha_o + n/2$, $\kappa_n = \kappa_o + n$, and $\mu_n = \frac{\kappa_o\mu_o + n\bar{x}}{\kappa_o+n}$.

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?*

Answer:

$p(\omega|y_i)$:

$$p(\omega|y_i) \;\alpha\; p(\omega)p(y_i|\omega) \tag{2.9}$$
$$\alpha\; (\omega^{n/2}e^{-(\omega/2)\sum_n(x_n-\mu)^2})(\omega^{\alpha_o-1}e^{\omega\beta_o}) \tag{2.10}$$
$$\alpha\; Gamma(\alpha_o + n/2, \beta_o + \sum_i (y_i-\mu)^2/2) \tag{2.11}$$

$p(\mu|\omega, y_i)$:

$$p(\mu|\omega, y_i) \;\alpha\; p(\mu|\omega)p(y_i|\omega,\mu) \tag{2.12}$$
$$\alpha\; e^{(\omega/2)(n+\kappa_o)(\mu^2 - 2\mu\frac{\kappa_o\mu_o+\sum y_i}{n+\kappa_o} + (\frac{\kappa_o\mu_o+\sum y_i}{n+\kappa_o})^2)} \tag{2.13}$$
$$\alpha\; Normal((\frac{\kappa_o\mu_o+\sum y_i}{n+\kappa_o}), \frac{1}{\omega(\kappa_o+n)}) \tag{2.14}$$

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?

Answer:

$$p(\mu) = \int_0^\infty p(\mu,\sigma^2)d\sigma^2 = \int_0^\infty p(\mu)p(\sigma^2)d\sigma^2 \tag{2.15}$$
$$= \int_0^\infty N(\bar{x},\sigma^2)[inverse\chi^2(\nu,s^2)] \tag{2.16}$$
$$= \int_0^\infty \frac{1}{\sigma^2}exp[\frac{-1}{2\sigma^2}(\mu-\bar{x})^2]\sigma^{-\nu-2}exp[-\frac{\nu s^2}{2\sigma^2}]d\sigma \tag{2.17}$$

Solved for below, where $A = (\mu-\bar{x})^2$ and $Z = exp[-\frac{1}{2\sigma^2}(A)]\sigma^{-\nu-2}$.

$$dz = \frac{A}{2\sigma^4} d\sigma^2 \tag{2.18}$$

$$p(\mu)\ \alpha\ A^{\frac{-\nu-1}{2}} \int_0^\infty z^{\frac{\nu-1}{2}} e^{-z} dz \tag{2.19}$$

$$p(\mu)\ \alpha\ (1 + \frac{(\mu - \bar{x})^2}{\nu s^2})^{\frac{-\nu-1}{2}} \tag{2.20}$$

Where $\nu = 2\alpha_o$, $m = \bar{x}$, $s = \sqrt{\frac{\beta_o}{\alpha_o \kappa_o}}$.

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.*

Answer:

This solution is the same as the above, but with these parameters: $\nu = 2\alpha_n$, $m = \mu_n$, and $s = \sqrt{\frac{\beta_n}{2\kappa_n \alpha_n^2}}$

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

Answer: Below, y represents the predicted values, and y_i represents the observations.

$$p(y|y_i) = \int\int p(y|\mu,\omega) p(\mu,\omega|y_i) d\mu d\omega \tag{2.21}$$

$$= \int\int \sqrt{\frac{\omega}{2\pi}} e^{\frac{-\omega(y-\mu)^2}{2}} \frac{\sqrt{\kappa_n}}{\Gamma(\alpha_n)\sqrt{2\pi}} \beta_n^{\alpha_n} e^{-\omega(\beta_n + (\kappa_n/2)(\mu-\mu_n)^2)} \omega^{\alpha_n - (1/2)} d\mu d\omega \tag{2.22}$$

$$\alpha\ (\beta_n + \frac{(\mu - y)^2}{2})^{-\alpha_n - (1/2)} \int\int NG_{pdf}\,\alpha\ (1 + \frac{(\mu - m)^2}{\nu s^2})^{\frac{\nu+1}{-2}} \tag{2.23}$$

In the above, $\nu = 2\alpha_n$, $m = \mu_n$, and $s = \sqrt{\frac{\beta_o}{2\alpha_o^2}}$.

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

Answer: In the below solution, $\tau = \sigma_o^2$, $m = \mu$, $x_i = y_i$, $D = y_{1\to n}$, $s^2 = 1/\sigma^2$, and $T^2 = \tau^{-2}$. To solve for the marginal distribution, I will solve for $p(D|m, \sigma^2, \tau)$.

$$p(D|m,\sigma^2,\tau) = \int [\prod_{i=1}^{n} N(x_i|\mu,\sigma^2)] N(\mu|m,\tau^2) d\mu \tag{2.24}$$

$$\frac{1}{(\sigma\sqrt{2\pi})^n(\tau\sqrt{2\pi})} \int exp[\frac{-1}{2\sigma^2}\sum_i (x_i-\mu)^2 - \frac{1}{2\tau}(\mu-m)^2] d\mu \tag{2.25}$$

$$\frac{1}{(\sqrt{2\pi}/s)^n(2\pi/T)} \int exp[\frac{s^2}{2}(\sum_i x_i^2 + n\mu - 2\mu\sum_i x_i) - \frac{T^2}{2}(\mu^2+m^2-2\mu m)] d\mu \tag{2.26}$$

$$C\int exp[\frac{-1}{2}(s^2\mu^2 n - 2s^2\sum_i x_i + T^2\mu^2 - 2T^2\mu m)] d\mu \tag{2.27}$$

$$C = \frac{exp[\frac{-1}{2}(s^2\sum_i x_i^2 + T^2 m^2)]}{\left(\frac{\sqrt{2\pi}}{s}\right)^n \frac{\sqrt{2\pi}}{T}} \tag{2.28}$$

$$C\int [-\frac{1}{2}(s^2 n + T^2)(\mu^2 - 2\mu\frac{s^2\sum_i x_i + T^2 m}{s^2 n + T^2})] d\mu \tag{2.29}$$

$$Cexp[\frac{(s^2 n\bar{x} + T^2 m)^2}{2(s^2 n + T^2)}]\sqrt{\frac{2\pi}{s^2 n + T^2}} \tag{2.30}$$

$$... \tag{2.31}$$

$$p(D) = \frac{\sigma}{(\sigma\sqrt{2\pi})^n\sqrt{n\tau^2+\sigma^2}} exp[-\frac{\sum_i x_i^2}{2\sigma} - \frac{m^2}{2\tau^2}] exp[\frac{\frac{\tau^2 n^2 \bar{x}^2}{\sigma^2} + \frac{\sigma^2 m^2}{\tau^2} + 2\bar{x}nm}{2(n\tau^2+\sigma^2)} \tag{2.32}$$

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

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

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.*

Answer: In this case, $d = 1$ and $\Sigma = \sigma^2$. Given that Λ_o, ν_o are scalars:

$$p(\Sigma|\nu_o, \Lambda_o^{-1}) = \frac{(\frac{\Lambda_o}{2})^{\nu_o/2}}{\Gamma(\nu_o/2)}(\Sigma)^{\nu_o/2-1}e^{-\Lambda_o/(2\Sigma)} \tag{2.33}$$

$$= \frac{\beta^{\alpha}(\Sigma^{-\alpha-1})e^{\Sigma^{-1}\beta}}{\Gamma(\alpha)} \tag{2.34}$$

Where $\alpha = \nu_o/2$, and $\beta = \Lambda_o/2$, and the distribution is an inverse Gamma.

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.

Answer:

$$p(\mu, \Sigma|y_i) \propto p(y_i|\mu, \Sigma)p(\mu, \Sigma) \tag{2.35}$$

$$\propto |\Sigma|^{\frac{\nu_o+d+1+n}{2}} exp[-\frac{1}{2}tr(\Lambda_o\Sigma^{-1}) + \frac{\kappa_o}{2}(\mu-\mu_o)^T\Sigma^{-1}(\mu-\mu_o) - \frac{1}{2}\sum_i(y_i-\mu)^T\Sigma^{-1}(y_i-\mu) \tag{2.36}$$

$$\propto |\Sigma|^{\frac{\nu_o+d+1+n}{2}} \tag{2.37}$$

$$exp[-\frac{1}{2}tr(\Lambda_o\Sigma^{-1} + \frac{1}{2}\sum_{i=1}^{n}(y_iy_i^T) + \frac{\kappa_o+n}{2}\frac{\kappa_o+\sum_{i+1}^{n}y_i}{\kappa_o+n}(\frac{\kappa_o+\sum_{i+1}^{n}y_i}{\kappa_o+n})^T) + \tag{2.38}$$

$$\frac{\kappa_o+n}{2}(\mu - \frac{\kappa_o+\sum_{i+1}^{n}y_i}{\kappa_o+n})^T\Sigma^{-1}(\mu - \frac{\kappa_o+\sum_{i+1}^{n}y_i}{k_o+n})] \tag{2.39}$$

With the parameters:
$\Lambda_n = \Lambda_o + \frac{1}{2}\sum_{i=1}^{N}y_iy_i^T + \frac{\kappa_o+n}{2}(\frac{\kappa_o+\sum_{i+1}^{n}y_i}{\kappa_o+n})(\frac{\kappa_o+\sum_{i+1}^{n}y_i}{\kappa_o+n})^T$
$\nu_n = \nu + n$
$\kappa_n = \kappa + n$
$\mu_n = \frac{\kappa_o+\sum_{i+1}^{n}y_i}{\kappa_o+n}$

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

Answer:

$$p(\beta|\omega, y) \propto p(y|\beta, \omega)p(\beta|\omega) \tag{2.40}$$

Keeping elements dependent on β:

$$p(\beta|\omega, y) \propto exp\Big[\frac{-1}{2}(y - X\beta)^T\omega\Lambda(y - X\beta) + (\beta - \mu)^T\omega K(\beta - \mu)\Big] \tag{2.41}$$

$$p(\beta|\omega, y) \propto exp\Big[\frac{-\omega}{2}(\beta - \mu_p)^T\Sigma_p^{-1}(\beta - \mu_p)\Big] \tag{2.42}$$

$$\tag{2.43}$$

Where:
$\mu_p = \omega(X^T\Lambda X + K)^{-1}(K\mu + X^T\Lambda y)$
$\Sigma_p = \omega(X^T\Lambda X + K)^{-1}$

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

Answer:

$$p(\omega|y_1, \ldots, y_n) = \int p(\beta|\omega, y)d\beta \tag{2.44}$$

$$= det\Big(\frac{\omega\Lambda}{2\pi}\Big)^{1/2} det\Big(\frac{\omega K}{2\pi}\Big)^{1/2} exp\Big[-\frac{\omega}{2}\big(y^T\Lambda y + \mu^T K\mu - \mu_n^T(X^T\Lambda X)\mu_n\big)\Big]\frac{b^a\omega^{a-1}exp[-b\omega]}{\Gamma(a)} \tag{2.45}$$

The density function will integrate to one, except for the constants which are functions of *omega*, removed in exercise 2.9. Below, these constants have been brought back and were combined with the remaining terms from above.

$$\propto \omega^{\frac{n+d}{2}+a-1}exp\Big[-\omega\big(b + \frac{1}{2}[y^T\Lambda y + \mu^T K\mu - \mu_n^T(X^T\Lambda X + K)\mu_n]\big)\Big] \tag{2.46}$$

$$\propto Gamma(a_n, b_n) \tag{2.47}$$

$$a_n = \frac{n+d}{2} + a \tag{2.48}$$

$$b_n = b + \frac{1}{2}[y^T\Lambda y + \mu^T K\mu - \mu_n^T(X^T\Lambda X + K)\mu_n] \tag{2.49}$$

Exercise 2.11 *Derive the marginal posterior, $p(\beta|y)$*

Answer:

$$p(\beta|y) = \int p(\beta, \omega|\vec{y})d\omega = \int p(\beta|\omega, \vec{y})p(\omega|\vec{y})d\omega \propto \int N(\mu_n \Sigma_n)Gamma(a_n, b_n)d\omega \tag{2.50}$$

$$= \int \frac{1}{(2\pi)^{D/2}}|\Sigma_n|^{-1/2}exp\Big[-\frac{1}{2}(\beta-\mu_n)^T\Sigma_n^{-1}(\beta-\mu_n)\Big]\frac{b_n^{a_n}\omega^{a_n-1}exp[-b_n\omega]}{\Gamma(a_n)}d\omega \tag{2.51}$$

$$\propto \int \omega^{a_n+n/2-1}exp\Big[-\omega(b_n+\frac{1}{2}(\beta-\mu_n)^T(X^T\Lambda X+K)(\beta-\mu_n))\Big]d\omega \tag{2.52}$$

$$\propto b_n + \frac{1}{2}((\beta-\mu_n)^T(X^T\Lambda X+K)(\beta-\mu_n))^{-(a_n+n/2)} \tag{2.53}$$

$$\propto 1 + \frac{1}{2b_n}((\beta-\mu_n)^T(X^T\Lambda X+K)(\beta-\mu_n))^{-(a_n+n/2)} \tag{2.54}$$

$$\tag{2.55}$$

The above is a student's t-distribution if $a_n = b_n = \nu/2$, where $\mu = \mu_n$, and $\Sigma = \omega(X^T\Lambda X + K)^{-1}$.

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?*

Answer: The above model was used to generate a bayesian model for the *dental.csv* dataset, with N=3000 samples. In the results below, "1" refers to the model fitted over both males and females, "2" refers to the model fitted over males only, and "3" refers to the model fitted over females only. Beta is defined by three parameters: $\beta = \{\beta_{int}, \beta_{age}, \beta_{gen}\}$, for use in the following equation:

$$\hat{y} = X\beta + \epsilon \tag{2.56}$$

$$\epsilon = Normal(0, \omega) \tag{2.57}$$

The residual plot and histograms of the parameters are given on the following pages in figures 2.2 and 2.1, respectively. It should be noted that for the male and female only models (2,3), there are only two model coefficients, β, since the gender is held constant.

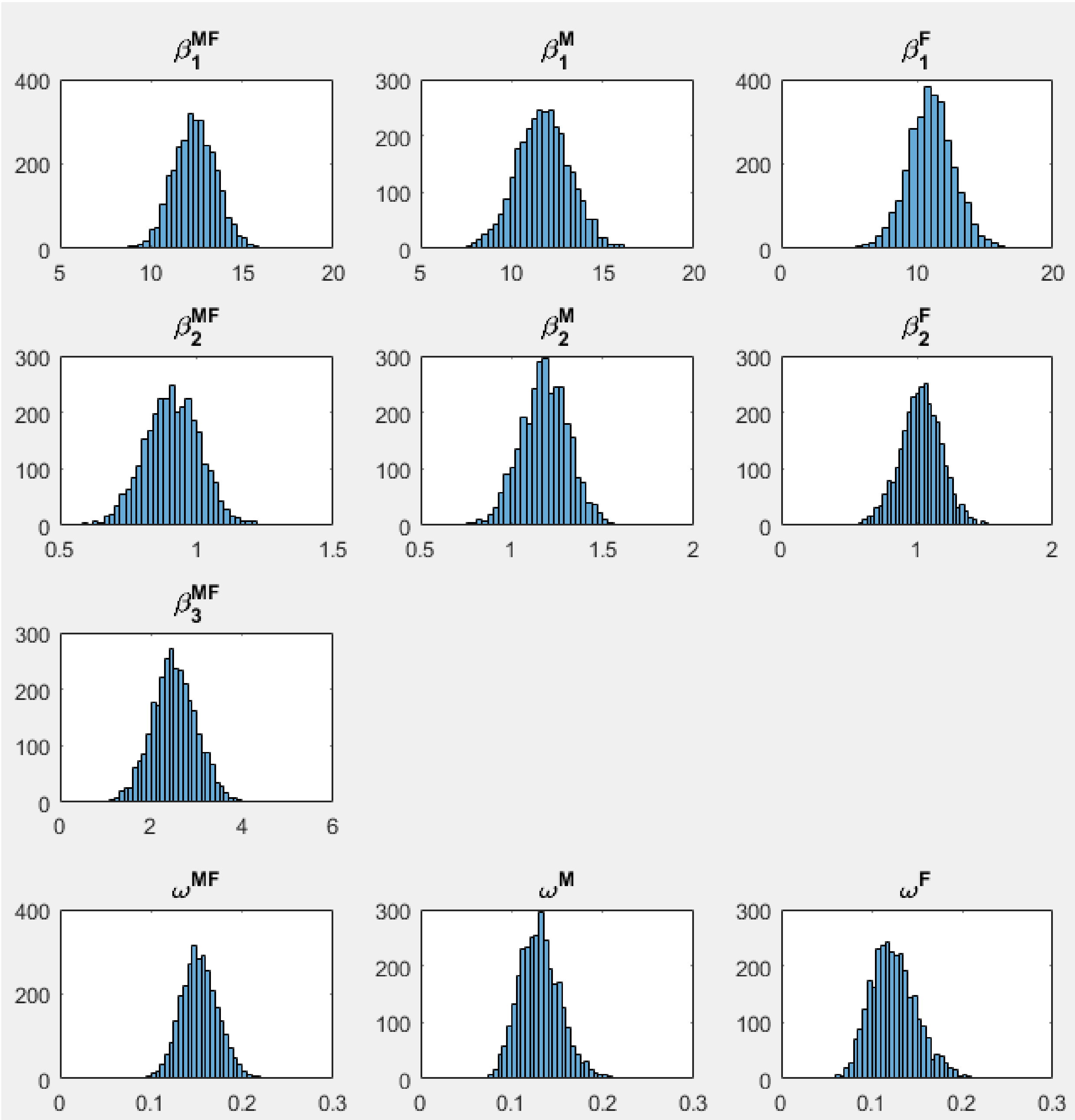

Figure 2.1: Plots showing histograms of the model parameters, where parameters are plotted for all iterations.

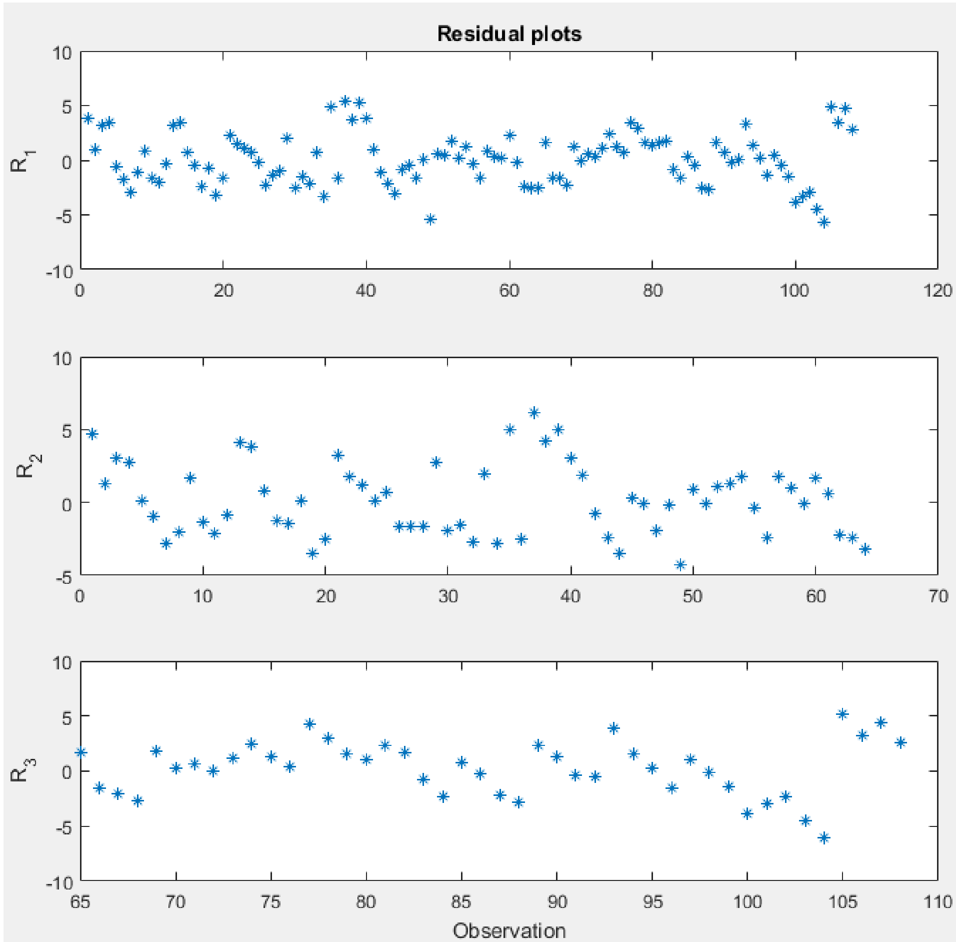

Figure 2.2: Plots showing residuals for each model, where models 2 and 3 were only used to predict their corresponding genders.

The mean values of β and ω are given in table 2.3, along with the root mean squared error (RMSE).

Table 2.1: Model parameters.

Model	β_{int}	β_{age}	β_{gen}	ω	RMSE
All	12.40	0.9116	2.507	0.1536	2.326
Male	11.76	1.184	–	0.1308	2.432
Female	11.06	1.031	–	0.1245	2.469

It is not totally surprising that the RMSE for the gender-separated models is marginally higher than that for the model with both genders included. This is likely due to the fact that the male and female models have a smaller sample size than when they are combined. The data including both genders was used to generate a least squares and a ridge regression, with parameters given in the following table.

Table 2.2: Model parameters for least squares (LS) and a ridge regression (R).

Model	β_{int}	β_{age}	β_{gen}
All LS	15.39	0.6602	2.321
Male LS	16.34	0.7844	–
Female LS	17.37	0.4795	–
All R	1.483	1.146	–
Male R	24.97	1.768	–
Female R	22.65	1.085	–

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

Answer:

$$p(\lambda_i|\vec{y},\beta,\omega) = p(\lambda_i|y_i,\beta,\omega) \propto p(y_i|\lambda_i,\beta,\omega)p(\lambda_i|\beta,\omega) \tag{2.58}$$

$$\sqrt{\frac{\omega\lambda_i}{2\pi}} exp\left[\frac{-\omega\lambda_i}{2}(y_i - X_i^T\beta)^2\right]\frac{\tau^\tau\lambda_i^{\tau-1}e^{-\tau\lambda_i}}{\Gamma(\tau)} \tag{2.59}$$

$$\propto \lambda_i^{\tau+1/2-1} exp\left[-\lambda_i\left(\tau + \frac{\omega}{2}(y_i - X_i^T\beta)^2\right)\right] \tag{2.60}$$

$$\propto Gamma(\tau + 1/2, \tau + \omega/2(y_i - X_i^T\beta)^2) \tag{2.61}$$

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.*

Answer: The above model was used to generate a bayesian model for the *dental.csv* dataset, where N=3000 samples. In the results below, "1" refers to the model fitted over both males and females, "2" refers to the model fitted over males only, and "3" refers to the model fitted over females only. Beta is defined by three parameters: $\beta = \{\beta_{int}, \beta_{age}, \beta_{gen}\}$, for use in the following equation:

$$\hat{y} = X\beta + \epsilon \tag{2.62}$$

$$\epsilon = Normal(0, \omega\lambda) \tag{2.63}$$

where X, y, and ϵ are of size [108x3]. The residual plot and histograms of the parameters are given on the following pages in figures 2.3 and 2.4, respectively.

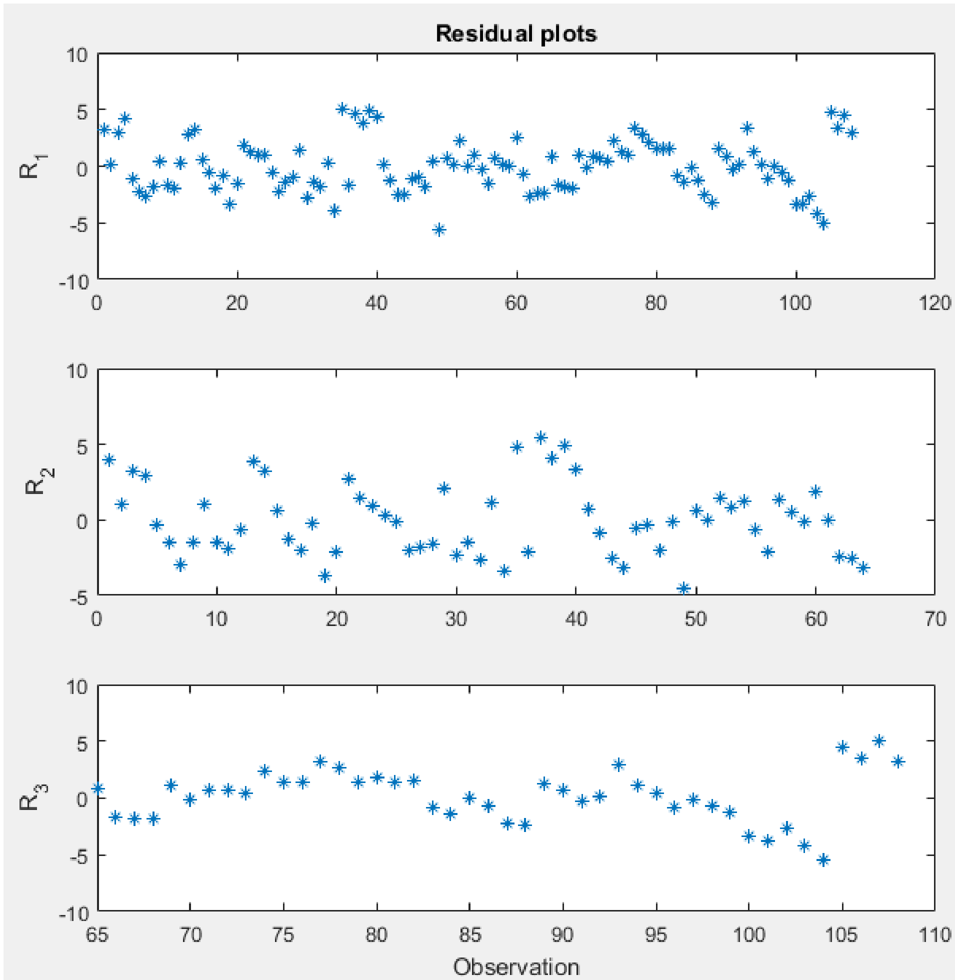

Figure 2.3: Plots showing residuals for each model, where models 2 and 3 were only used to predict their corresponding genders.

The mean values of β and ω are given in table 2.3, along with the root mean squared error (RMSE).

Table 2.3: Gibbs sampler model parameters.

Model	β_{int}	β_{age}	β_{gen}	ω	RMSE
All	13.53	0.7338	0.8124	0.0789	2.283
Male	13.64	1.0326	–	0.0726	2.325
Female	13.69	0.7955	–	0.0750	2.201

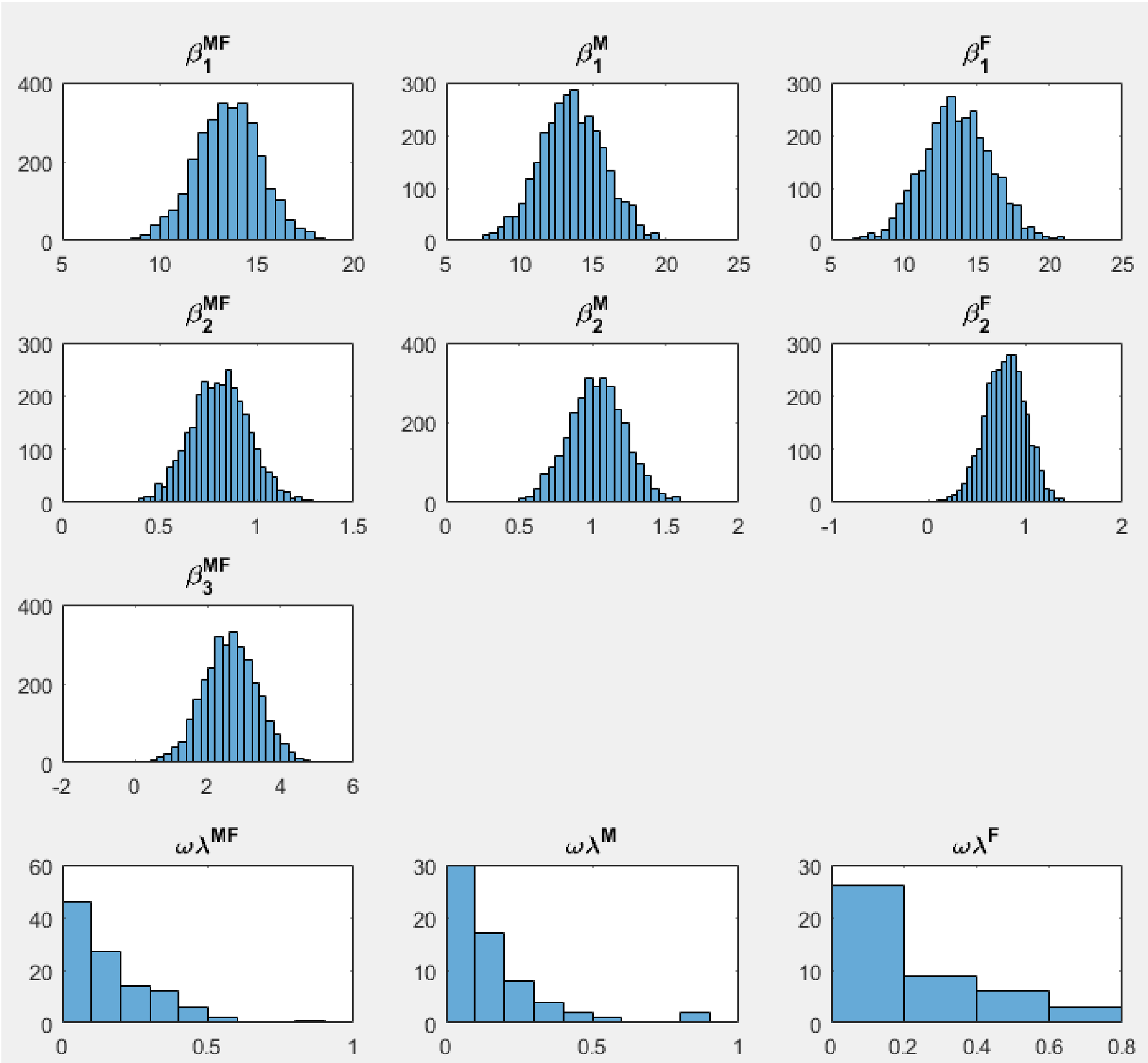

Figure 2.4: Plots showing histograms of the model parameters, where $\omega\lambda$ is for the last iteration and β represents all iterations.

The RMSEs for the three models generated using a Gibbs sampler are similar, where the female model shows to be predicting slightly better than the model with both genders and with just males. It is likely that the small sample size is inhibiting the performance of the Gibbs sampler on modeling this data in addition to the discrete time points (age).

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.*

Answer: The model is okay, but it could be improved. The model does not take in account that there are different subjects in the study, which means that the individual subject background is not accounted for. In order to account for this, I propose to include a mean difference term for each subject. With this method, the y-prediction equation becomes:

$$\hat{y} = X\beta + \epsilon + \Delta\mu \tag{2.64}$$
$$\Delta\mu = \mu_s - \bar{y} \tag{2.65}$$
$$\epsilon = Normal(0, \omega\lambda) \tag{2.66}$$

Where μ_s represents the mean of the distance for each subject, and $\bar{y}$ represents the sample mean distance. The RMSEs for the all gender, male, and female models are: $RMSE = \{1.9228, 2.2682, 1.2619\}$, which have improved on the previous RMSEs, especially for the female model and the both gender model. Lastly, the residuals are plotted below for the improved model.

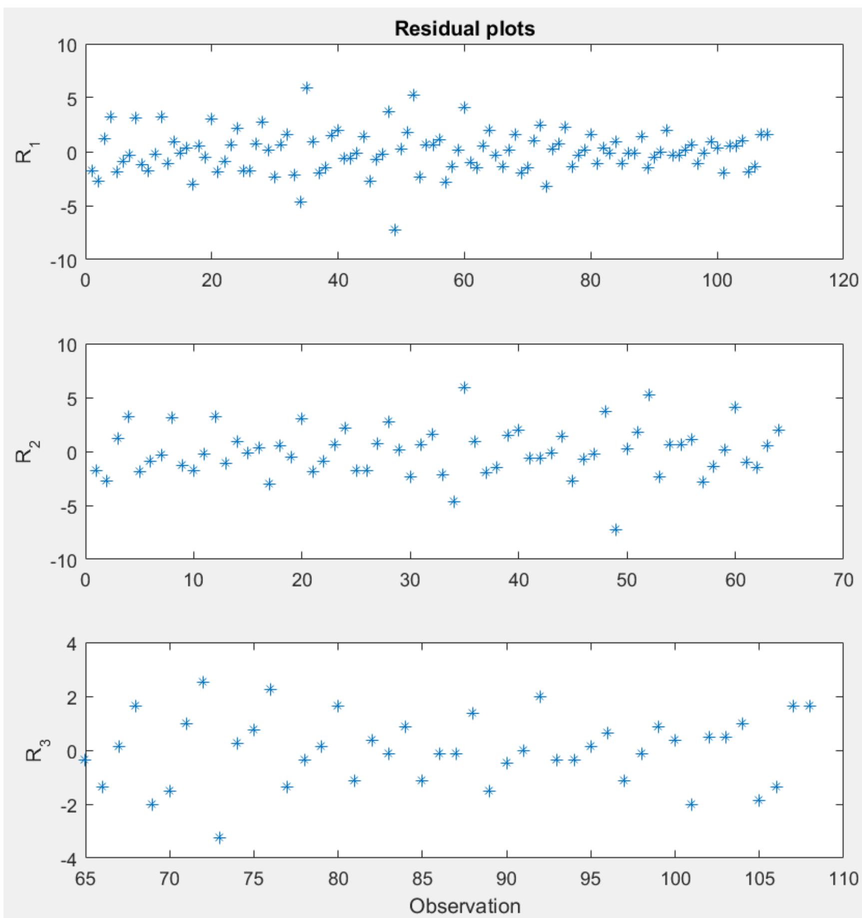

Figure 2.5: Plots showing residuals for each model, where models 2 and 3 were only used to predict their corresponding genders.