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: $p(\mu,w)$ has normal-gamma distribution as follows:

$$p(\mu,w)=p(\mu|w)p(w)=\frac{\beta_0^{\alpha_0}\sqrt{k_0}}{\Gamma(\alpha_0)\sqrt{2\pi}}w^{\alpha_0-\frac{1}{2}}e^{-\beta_0 w}e^{-\frac{1}{2}k_0 w(\mu-\mu_0)^2}$$

Next, given data $\mathbf{y}=y_1,...,y_n$:

$$\begin{aligned}p(\mu,w|\mathbf{y}) &\propto L(\mathbf{y}|\mu,w)p(\mu,w)\\ L(\mathbf{y}|\mu,w) &\propto w^{n/2}exp(-\frac{w}{2}\sum(y_i-\mu)^2)\\ &\propto w^{n/2}exp\left(-\frac{w}{2}\sum(y_i-\bar{y}+\bar{y}-\mu)^2\right)\\ &\propto w^{n/2}exp\left(-\frac{w}{2}\sum\left[(y_i-\bar{y})^2+(\bar{y}-\mu)^2\right]\right)\\ &\propto w^{n/2}exp\left(-\frac{w}{2}(ns+n(\bar{y}-\mu)^2)\right)\end{aligned}$$

where $\bar{y}$ and s are sample mean and sample variance. Substitute the likelihood expression into the posterior distribution, we get:

$$\begin{aligned} p(\mu, w|\mathbf{y}) &\propto L(\mathbf{y}|\mu, w)p(\mu, w) \\ &\propto w^{\frac{n}{2}+\alpha_0-\frac{1}{2}} exp\left(-w(\frac{1}{2}ns + \beta_0)\right) exp\left(-\frac{w}{2}(k_0(\mu-\mu_0)^2 + n(\bar{y}-\mu)^2)\right) \\ &\propto w^{\alpha_n-\frac{1}{2}} exp(-w\beta_n) exp(-\frac{wk_n}{2}(\mu-\mu_n)^2) \end{aligned}$$

which is in the form of normal-gamma distribution, with:

$$(\mu_n, k_n, \alpha_n, \beta_n) = \left(\frac{k_0\mu_0 + n\bar{y}}{k_0+n}, k_0+n, \alpha_0+\frac{n}{2}, \beta_0+\frac{1}{2}ns+\frac{k_0 n(\bar{y}-\mu_0)^2}{2(k_0+n)}\right)$$

■

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: $p(\mu|w, \mathbf{y}) \propto p(\mu, w, \mathbf{y})$, so $p(\mu|w, \mathbf{y})$ has the same form as the normal-gamma distribution of $p(\mu, w|y)$ in 2.1, but with w and $\mathbf{y}$ as given parameters. So:

$$p(\mu|w, \mathbf{y}) \propto e^{-\frac{wk_n}{2}(\mu-\mu_n)^2} \propto N(\mu_n, wk_n)$$

For $p(w|y)$, integrate out μ from $p(\mu, w|y)$:

$$\begin{aligned} p(w|y) &\propto \int p(\mu, w|y)d\mu \\ &\propto w^{\alpha_n-\frac{1}{2}} e^{-w\beta_n} \int e^{-\frac{wk_n}{2}(\mu-\mu_0)^2} d\mu \\ &\propto w^{\alpha_n-\frac{1}{2}} e^{-w\beta_n} w^{-\frac{1}{2}} \\ &\propto gamma(\alpha_n, \beta_n) \end{aligned}$$

k_0 is like number of pseudo-observations for the prior on μ. μ_0 is the prior expected value of μ. $2\alpha_0$ is like number of pseudo-observations for w. Expected value of w is α_0/β_0 and variance of w is α_0/β_0^2. ■

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: From 2.1, we have:

$$p(\mu, w) = p(\mu|w)p(w) = \frac{\beta_0^{\alpha_0}\sqrt{k_0}}{\Gamma(\alpha_0)\sqrt{2\pi}} w^{\alpha_0+\frac{1}{2}-1} exp(-w(\beta_0 + \frac{1}{2}k_0(\mu-\mu_0)^2))$$

Note the pdf is also in the form of $gamma(\alpha_1, \beta_1)$ over w, where $\alpha_1 = \alpha_0 + \frac{1}{2}, \beta_1 = \beta_0 + \frac{1}{2}k_0(\mu - \mu_0)^2$. Integrate out w:

$$\begin{aligned}\int p(\mu, w)dw &= \frac{\beta_0^{\alpha_0}\sqrt{k_0}}{\Gamma(\alpha_0)\sqrt{2\pi}} \int w^{\alpha_1} e^{-w\beta_1} dw \\ &= \frac{\beta_0^{\alpha_0}\sqrt{k_0}\ \Gamma(\alpha_1)}{\Gamma(\alpha_0)\sqrt{2\pi}\ \beta_1^{\alpha_1}} \\ &= \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{\pi v s^2}} \left(1 + \frac{1}{v}(\frac{\mu - \mu_0}{s})^2\right)^{-\frac{v+1}{2}}\end{aligned}$$

$v = 2\alpha_0, m = \mu_0, s^2 = \frac{\beta_0}{k_0\alpha_0}$ ■

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: $p(\mu, w|y)$ has the same form as $p(\mu, w)$. So

$$v = 2\alpha_n, m = \mu_n, s^2 = \frac{\beta_n}{k_n\alpha_n}$$

■

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

Proof: Use the formula from exercise 2.6 and then apply Bayes' Rule:

$$\begin{aligned}p(y_{n+1}, \ldots, y_{n+m}|y_1, \ldots, y_n) &= \frac{p(y_1, \ldots, y_n, y_{n+1}, \ldots, y_{n+m})}{p(y_1, \ldots, y_n)} \\ &= (2\pi)^{-\frac{m}{2}} \frac{\beta_n^{\alpha_n}\Gamma(\alpha_{n+m})k_n^{\frac{1}{2}}}{\beta_{n+m}^{\alpha_{n+m}}\Gamma(\alpha_n)k_{n+m}^{\frac{1}{2}}}\end{aligned}$$

■

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

Proof:

$$p(y) = \frac{p(\mu, w, y)}{p(\mu, w|y)}$$

Both $p(\mu, w, y)$ and $p(\mu, w|y)$ have identical terms involving w and μ, so $p(y)$ is just the ratio of the "constant" terms of two normal-gamma like distributions. $p(\mu, w, y)$ has constant coefficient $(2\pi)^{-n/2}\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}(2\pi)^{-1/2}k_0^{1/2}$; $p(\mu, w|y)$ has constant coefficient $\frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}(2\pi)^{-1/2}k_n^{1/2}$. This gives:

$$p(y) = (2\pi)^{-\frac{n}{2}} \frac{\beta_0^{\alpha_0}\Gamma(\alpha_n)k_0^{\frac{1}{2}}}{\beta_n^{\alpha_n}\Gamma(\alpha_0)k_n^{\frac{1}{2}}}$$

■

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

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

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

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

Proof: In 1-dimension, the formula reduces to:

$$p(x|\lambda, v_0) = \frac{(\lambda/2)^{(v_0/2)}}{\Gamma(v_0/2)}x^{-(v_0/2)-1}e^{-(\lambda/2)x^{-1}}$$

■

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 $p(\mu, \Sigma|y_1, \ldots, y_n)$ 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.

Proof:

$$
\begin{aligned}
&p(\mu,\Sigma|y) \propto p(y|\mu,\Sigma)p(\mu,\Sigma) \\
&\propto \left(\prod exp\Big(-\frac{1}{2}(y_i-\mu)'\Sigma^{-1}(y_i-\mu)\Big)\right) exp\left(-\frac{1}{2}\mathrm{tr}(\Lambda_0\Sigma^{-1}) - \frac{\kappa_0}{2}(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)\right) \\
&\propto exp\left\{-\frac{1}{2}\Big[tr(\Lambda_0\Sigma^{-1}) + k_0(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0) + \sum(y_i-\mu)'\Sigma^{-1}(y_i-\mu)\Big]\right\} \\
&\propto exp\left\{-\frac{1}{2}\Big[tr(\Lambda_0\Sigma^{-1}) + k_0(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0) + n(\bar{y}-\mu)'\Sigma^{-1}(\bar{y}-\mu) + \sum(y_i-\bar{y})'\Sigma^{-1}(y_i-\bar{y})\Big]\right\}
\end{aligned}
$$

Note:

$$
\begin{aligned}
&k_0(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0) + n(\bar{y}-\mu)'\Sigma^{-1}(\bar{y}-\mu) \\
&= (k_0+n)\mu'\Sigma^{-1}\mu - 2(k_0\mu_0' + n\bar{y})\Sigma^{-1}\mu + k_0\mu_0'\Sigma^{-1}\mu_0 + n\bar{y}'\Sigma^{-1}\bar{y}) \\
&= (k_0+n)(\mu-\mu_n)'\Sigma^{-1}(\mu-\mu_n) - (k_0+n)\mu_n'\Sigma^{-1}\mu_n + k_0\mu_0'\Sigma^{-1}\mu_0 + n\bar{y}'\Sigma^{-1}\bar{y}) \\
&= (k_0+n)(\mu-\mu_n)'\Sigma^{-1}(\mu-\mu_n) + tr\big(-(k_0+n)\mu_n\mu_n'\Sigma^{-1} + k_0\mu_0\mu_0'\Sigma^{-1} + n\bar{y}\bar{y}'\Sigma^{-1}\big)
\end{aligned}
$$

with $\mu_n = (k_0\mu_0 + n\bar{y})/(k_0+n)$. Thus:

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

which is normal-inverse Wishart distribution with

$$
\begin{aligned}
\mu_n &= \frac{k_0\mu_0 + n\bar{y}}{k_0+n} \\
\kappa_n &= \kappa_0 + n \\
\nu_n &= \nu_0 + n \\
\Lambda_n &= \Lambda_0 + \sum(y_i-\bar{y})(y_i-\bar{y})' - k_n\mu_n\mu_n' + k_0\mu_0\mu_0' + n\bar{y}\bar{y}' = \Lambda_0 + \sum(y_iy_i') - k_n\mu_n\mu_n' + k_0\mu_0\mu_0'
\end{aligned}
$$

μ_0 is prior mean for μ. k_0 is like prior pseudo-observations, or importance weighting, for μ_0. ■

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

Proof: First, we have $p(\beta|w, y) \propto p(y, \beta, w) = p(y|\beta, w)p(\beta|w)p(w)$

$$\propto exp\left(-\frac{1}{2}(y - X\beta)'(w\Lambda)(y - X\beta)\right) exp\left(-\frac{1}{2}(\beta - \mu)'(wK)(\beta - \mu)\right) w^{a+(d+n)/2-1}exp(-bw)$$

Thus,

$$\begin{aligned}
p(\beta|w, y) &\propto exp\left(-\frac{1}{2}(y - X\beta)'(w\Lambda)(y - X\beta)\right) exp\left(-\frac{1}{2}(\beta - \mu)'(wK)(\beta - \mu)\right) \\
&\propto exp\left\{-\frac{w}{2}\left[\beta'(X'\Lambda X + K)\beta - 2(y'\Lambda X + \mu'K)\beta\right]\right\} \\
&\Lambda_n \equiv w(X'\Lambda X + K), \\
&\mu_n \equiv w\Lambda_n^{-1}(X'\Lambda y + K\mu) = (X'\Lambda X + K)^{-1}(X'\Lambda y + K\mu) \\
&\propto exp\left\{-\frac{1}{2}\left[\beta'\Lambda_n\beta - 2\mu'\Lambda_n\beta\right]\right\} \\
&\propto exp\left\{-\frac{1}{2}\left[\beta'\Lambda_n\beta - 2\mu_n'\Lambda_n\beta + \mu_n'\Lambda_n\mu_n\right]\right\} \\
&\propto exp\left(-\frac{1}{2}(\beta - \mu_n)'\Lambda_n(\beta - \mu_n)\right) \\
&\sim Normal(\mu_n, \Lambda_n)
\end{aligned}$$

■

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

Proof:

$$\begin{aligned}
p(w|y) \propto p(y, w) &\propto \int p(y, w, \beta)d\beta \\
&\propto w^{a+(d+n)/2-1}e^{-bw}\int exp\left\{-\frac{1}{2}\left[(y - X\beta)'(w\Lambda)(y - X\beta) + (\beta - \mu)'(wK)(\beta - \mu)\right]\right\}d\beta \\
&\propto w^{a+n/2-1}e^{-bw}\int w^{d/2}exp\left\{-\frac{1}{2}\left[(\beta - \mu_n)'\Lambda_n(\beta - \mu_n) - \mu_n'\Lambda_n\mu_n + w\mu'K\mu + wy'\Lambda y\right]\right\}d\beta \\
&\propto w^{a+n/2-1}e^{-bw}exp\left\{-\frac{w}{2}\left[-\frac{\mu_n'\Lambda_n\mu_n}{w} + \mu'K\mu + y'\Lambda y\right]\right\} \\
&\propto w^{(a+n/2)-1}exp\left\{-w\left[b + \frac{1}{2}\left(\mu'K\mu + y'\Lambda y - \mu_n'(X'\Lambda X + K)\mu_n\right)\right]\right\}
\end{aligned}$$

Thus, $p(w|y)$ has gamma distribution. ■

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

Proof: Recall from exercise 2.9, $\Lambda_n \equiv w(X'\Lambda X + K) = wG$:

$$\begin{aligned}
p(\beta|y) &= \int p(\beta, w|y)dw = \int p(\beta|w, y)p(w|y)dw \\
&= \int \frac{b^a}{\Gamma(a)} w^{a-1} e^{-wb} (2\pi)^{-d/2} det(\Lambda_n)^{\frac{1}{2}} e^{-\frac{1}{2}(\beta-\mu_n)'\Lambda_n(\beta-\mu_n)} dw \\
&= \frac{b^a |G|^{0.5}}{\Gamma(a)(2\pi)^{d/2}} \int w^{a+\frac{d}{2}-1} exp\left\{ -w\left[b + \frac{1}{2}(\beta-\mu_n)'G(\beta-\mu_n) \right] \right\} dw \\
&= \frac{b^a \Gamma(a+d/2)|G|^{0.5}}{(b+0.5(\beta-\mu_n)'G(\beta-\mu_n))^{a+d/2}\Gamma(a)(2\pi)^{d/2}} \\
&= \frac{\Gamma(a+d/2)}{\Gamma(a)\pi^{d/2}} \left(\frac{b}{b + \frac{1}{2a}(\beta-\mu_n)'(aG)(\beta-\mu_n)} \right)^{a+d/2} \left| \frac{aG}{b} \right|^{1/2} \frac{1}{(2a)^{d/2}} \\
&= \frac{\Gamma(\frac{v+d}{2})}{\Gamma(\frac{v}{2})\pi^{d/2} v^{d/2}} \left| \Sigma^{-1} \right|^{1/2} \left(1 + \frac{1}{v}(\beta-\mu_n)'\Sigma^{-1}(\beta-\mu_n) \right)^{-\frac{v+d}{2}}
\end{aligned}$$

with $v = 2a, \Sigma = \left(\frac{a}{b}G\right)^{-1} = (X'\Lambda X + K)^{-1}\frac{b}{a}$. This is a multivariate t-distribution. Note a and b are the posterior parameters for $p(w|y)$. ■

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?*

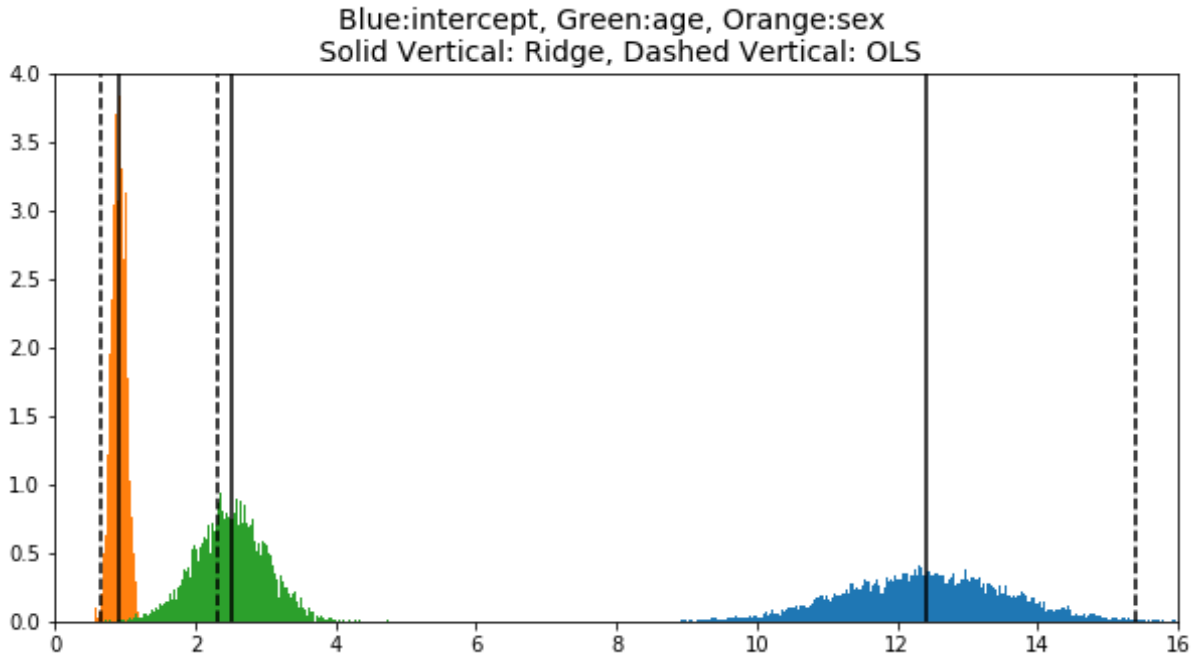
Proof: vague priors: set $a_0 = 1, b_0 = 0.1, \mu_0 = [0, 0, 0]$. Sample β from directly multivariate t-distribution. Code and plot are given below. ■

```
import pandas as pd
import numpy as np
import importlib
import matplotlib.pyplot as plt
import sklearn.linear_model as LM
J = importlib.import_module('my_functions',r'G:\My Drive\Utility Programming')

df0 = pd.read_excel('dental.xlsx')
df1 = df0[['distance','age']].copy()
df1['sex'] = (df0['sex']=='Male')*1
df1['one'] = 1
y = df1['distance'].values
X = df1[['one','age','sex']].values

#%%
```

```
a = 1; b = 0.01
n,d = X.shape
K = np.eye(d)
Lambda = np.eye(n)
Mu0 = np.zeros(d)
G = (X.T@Lambda@X+K)
invG = np.linalg.inv(G)
Mu = invG@(X.T@Lambda@y + K@Mu0)
an = a+n/2
bn = b+0.5*(Mu0.T@K@Mu0+y.T@Lambda@y-Mu.T@G@Mu)
Sigma = invG*bn/an
N = int(1e4)
v = 2*an
x = np.random.chisquare(v, N)/v
Beta = J.multivariate_t_rvs(Mu, Sigma, v, N)
mRidge = LM.Ridge(alpha=1.0,fit_intercept=False).fit(X,y)
mOLS = LM.LinearRegression(fit_intercept=False).fit(X,y)
```



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

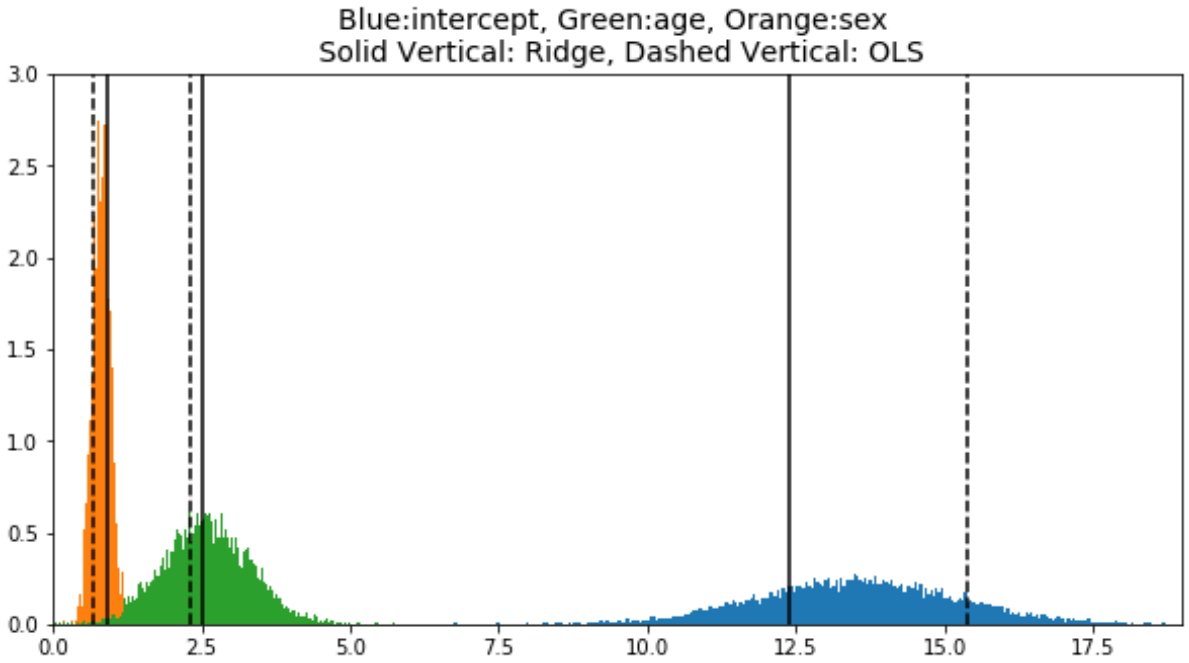
Proof:

$$\begin{aligned}
p(\lambda_i|\mathbf{y},\beta,w) &\propto p(\lambda_i,y_i,\beta,w) \propto p(y_i|\beta,w,\lambda_i)p(\lambda_i) \\
&\propto (w\lambda_i)^{1/2}e^{-\frac{w\lambda_i}{2}(y-X\beta)_i^2}\lambda_i^{\tau-1}e^{-\tau\lambda_i} \\
&\propto \lambda_i^{\tau+1/2-1}exp\left\{-\lambda_i(\tau+\frac{w}{2}(y-X\beta)_i^2)\right\}
\end{aligned}$$

which is Gamma distribution. ■

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.*

Now the distributions of beta have greater variance as Λ becomes another source of uncertainty.



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.*

First, average beta over all Gibb samples beyond the first 1000. The residual density plot on the left show a somewhat asymmetric distribution for males and females. One of the underlying assumptions in the hierarchical model is variance of $(y|\beta, w, \Lambda)$ is different across observations. Another sensible assumption is all males have one variance and all females have another variance; i.e. $\Lambda = diag(\lambda_m, \lambda_f)$. With this pooled variance assumption, the residual density plot appears much more symmetric for both sexes.