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?*

Solution:

$$P(\mu, \omega|y_i) \propto P(y_i|\mu, \omega)P(\mu, \omega) = P(y_i|\mu, \omega)P(\mu|\omega)P(\omega)$$

$$P(\mu, \omega) = \frac{\sqrt{w\kappa_0}}{\sqrt{2\pi}}e^{-\frac{\omega\kappa_0}{2}(\mu-\mu_0)^2}\frac{1}{\Gamma(\alpha_0)}\beta_0^{\alpha_0}\omega^{\alpha_0-1}e^{-\beta_0\omega} = \frac{\sqrt{\kappa_0}}{\Gamma(\alpha_0)\sqrt{2\pi}}\beta_0^{\alpha_0}e^{-\omega(\beta_0+\frac{\kappa_0}{2}(\mu-\mu_0)^2)}\omega^{\alpha_0-\frac{1}{2}}$$

which is a Normal-Gamma distribution

$$\begin{aligned}P(\mu, \omega|y_i) &\propto P(y_i|\mu, \omega)P(\mu, \omega)\\ &= \left\{\frac{\omega}{2\pi}\right\}^{n/2}e^{\frac{-\omega}{2}\sum_{i=1}^n(y_i-\mu)^2}\frac{\sqrt{\kappa_0}}{\Gamma(\alpha_0)\sqrt{2\pi}}\beta_0^{\alpha_0}e^{-\omega(\beta_0+\frac{\kappa_0}{2}(\mu-\mu_0)^2)}\omega^{\alpha_0-\frac{1}{2}}\\ &= \frac{\sqrt{\kappa_0}\beta_0^{\alpha_0}}{\Gamma(\alpha_0)(2\pi)^{\frac{n+1}{2}}}e^{-\omega(\beta_0+\frac{\kappa_0}{2}(\mu-\mu_0)^2+\frac{1}{2}\sum_{i=1}^n(y_i-\mu)^2)}\omega^{\alpha_0+\frac{n-1}{2}}\end{aligned}$$

which is proportional to a new Normal Gamma distribution where:

$$\alpha_n = \alpha_0 + \frac{n}{2}$$

$$\beta_0 + \frac{\kappa_0}{2}(\mu-\mu_0)^2 + \frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2) = \beta_n + \frac{\kappa_n}{2}(\mu-\mu_n)^2$$

$$\beta_0 + \mu^2(\frac{\kappa_0+n}{2}) - 2\mu(\frac{\kappa_0\mu_0 + \sum_{i=1}^{n}(y_i)}{2}) + \frac{\kappa_0\mu_0^2}{2} + \frac{sum_{i=1}^{n}(y_i^2)}{2} = \beta_n + \frac{\kappa_n}{2}(\mu-\mu_n)^2$$

$$\beta_0 + \frac{\kappa_0+n}{2}(\mu^2 - 2\mu(\frac{\kappa_0\mu_0 + \sum_{i=1}^{n}(y_i)}{\kappa_0+n}) + \left\{\frac{\kappa_0\mu_0 + \sum_{i=1}^{n}(y_i)}{\kappa_0+n}\right\}^2) - \left\{\frac{(\kappa_0\mu_0 + \sum_{i=1}^{n}(y_i))^2}{\kappa_0+n}\right\}) + \frac{\kappa_0\mu_0^2}{2} + \frac{\sum_{i=1}^{n}(y_i^2)}{2}$$
$$= \beta_n + \frac{\kappa_n}{2}(\mu-\mu_n)^2$$

Which give:

$$\mu_n = \frac{\kappa_0\mu_0 + \sum_{i=1}^{n}(y_i)}{\kappa_0+n}$$

$$\kappa_n = \kappa_0 + n$$

$$\beta_n = \beta_0 - \left\{\frac{(\kappa_0\mu_0 + \sum_{i=1}^{n}(y_i))^2}{\kappa_0+n}\right\} + \frac{\kappa_0\mu_0^2}{2} + \frac{\sum_{i=1}^{n}(y_i^2)}{2}$$

which can be expressed as:

$$\beta_n = \beta_0 + \frac{1}{2}\left\{\sum_{i=1}^{n}(y_i^2) - \kappa_n\mu_n^2\right\} + \frac{\kappa_0\mu_0^2}{2}$$

Solution End

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:
For μ:

$$P(\mu|\omega, y_i) \propto P(y_i|\mu,\omega)P(\mu|\omega)$$

$$= \left\{\frac{\omega}{2\pi}\right\}^{n/2} e^{\frac{-\omega}{2}\sum_{i=1}^{n}(y_i-\mu)^2} \frac{\sqrt{w\kappa_0}}{\sqrt{2\pi}} e^{-\frac{\omega\kappa_0}{2}(\mu-\mu_0)^2} = \frac{\sqrt{\kappa_0}\omega^{\frac{n+1}{2}}}{(2\pi)^{\frac{n+1}{2}}} e^{\frac{-\omega}{2}(\sum_{i=1}^{n}(y_i-\mu)^2 + \kappa_0(\mu-\mu_0)^2)}$$

$$\propto Ce^{\frac{-\omega}{2}(n+\kappa_0)(\mu^2 - 2\mu\frac{\kappa_0\mu_0+\sum_{i=1}^{n}y_i}{n+\kappa_0} + (\frac{\kappa_0\mu_0+\sum_{i=1}^{n}y_i}{n+\kappa_0})^2)} \propto Normal(\frac{\kappa_0\mu_0 + \sum_{i=1}^{n}y_i}{n+\kappa_0}, \frac{1}{\omega(\kappa_0+n)})$$

For ω:

$$P(\omega|y_i) = \int P(\mu,\omega|y_i)d\mu$$

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

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

$$= Gamma(\alpha_n, \beta_n) \int pdf_{normal} = Gamma(\alpha_n, \beta_n)$$

κ_0 can be viewed as the number of samples in prior and μ_0 is their mean. β_0 is a representation of the variance of prior samples, and α_0 is the related to twice the number of prior samples.
Solution End

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?

Solution:

$$P(\mu) = \int P(\mu, \omega) d\omega = \frac{\sqrt{\kappa_0}}{\Gamma(\alpha_0)\sqrt{2\pi}} \int \beta_0^{\alpha_0} e^{-\omega(\beta_0 + \frac{\kappa_0}{2}(\mu-\mu_0)^2)} \omega^{\alpha_0 - \frac{1}{2}} d\omega$$

$$= \frac{\sqrt{\kappa_0} + \beta_0^{\alpha_0}\Gamma(\alpha_0 + \frac{1}{2})}{\Gamma(\alpha_0)\sqrt{2\pi}(\beta_0 + \frac{\kappa_0}{2}(\mu - \mu_0)^2)^{\alpha_0 + \frac{1}{2}}} \int \frac{e^{-\omega(\beta_0 + \frac{\kappa_0}{2}(\mu-\mu_0)^2)}(\beta_0 + \frac{\kappa_0}{2}(\mu - \mu_0)^2)^{\alpha_0+\frac{1}{2}} \omega^{\alpha_0 - \frac{1}{2}}}{\Gamma(\alpha_0 + \frac{1}{2})} d\omega$$

$$= \frac{\sqrt{\kappa_0} + \beta_0^{\alpha_0}\Gamma(\alpha_0 + \frac{1}{2})}{\Gamma(\alpha_0)\sqrt{2\pi}(\beta_0 + \frac{\kappa_0}{2}(\mu - \mu_0)^2)^{\alpha_0 + \frac{1}{2}}} \int pdf_{Gamma} = \frac{\sqrt{\kappa_0} + \beta_0^{\alpha_0}\Gamma(\alpha_0 + \frac{1}{2})}{\Gamma(\alpha_0)\sqrt{2\pi}(\beta_0 + \frac{\kappa_0}{2}(\mu - \mu_0)^2)^{\alpha_0 + \frac{1}{2}}} \propto \left(1 + \frac{1}{\nu}\frac{(\mu - m)^2}{s^2}\right)^{-\frac{\nu+1}{2}}$$

where:

$$\nu = 2\alpha_0$$

$$m = \mu_0$$

$$\frac{1}{\nu s^2} = \frac{\kappa_0}{2\beta_0} \Rightarrow s = \sqrt{\frac{\beta_0}{2\kappa_0\alpha_0^2}}$$

Solution End

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: Since the posterior distribution has the same form of the marginal distribution, the parameters for the posterior can be given by analogy with Exercise 2.3 as:

$$\nu = 2\alpha_n$$

$$m = \mu_n$$

$$\frac{1}{\nu s^2} = \frac{\kappa_n}{2\beta_n} \Rightarrow s = \sqrt{\frac{\beta_n}{2\kappa_n\alpha_n^2}}$$

Solution End

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

Solution:

$$P(y|y_i) = \int\int P(y|\mu,\omega)P(\mu,\omega|y_i)d\mu d\omega$$
$$= \int\int \left\{\frac{\omega}{2\pi}\right\}^{n/2} e^{\frac{-\omega}{2}\sum_{i=1}^{n}(y_i-\mu)^2} \frac{\sqrt{\kappa_n}}{\Gamma(\alpha_n)\sqrt{2\pi}}\beta_n^{\alpha_n} e^{-\omega(\beta_n+\frac{\kappa_n}{2}(\mu-\mu_n)^2)}\omega^{\alpha_n-\frac{1}{2}}d\mu d\omega$$
$$= \int\int \frac{\sqrt{\kappa_n}}{\Gamma(\alpha_n)(\sqrt{2\pi})^{n+1}}\beta_n^{\alpha_n} e^{-\omega(\beta_n+\frac{\kappa_n}{2}(\mu-\mu_n)^2+\frac{-1}{2}\sum_{i=1}^{n}(y_i-\mu)^2)}\omega^{(\alpha_n+\frac{n}{2})-\frac{n}{2}}d\mu d\omega$$
$$= \frac{\beta_n^{\alpha_n}\Gamma(\alpha_n+\frac{n}{2})}{\Gamma(\alpha_n)\sqrt{2\pi}(\beta_n+\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2)^{\alpha_n+\frac{1}{2}}}\int\int \frac{e^{-\omega(\beta_n+\frac{\kappa_n}{2}(\mu-\mu_n)^2+\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2)}(\beta_n+\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2)^{\alpha_n+\frac{n}{2}}\omega^{(\alpha_n+\frac{n}{2})-\frac{1}{2}}}{\Gamma(\alpha_n+\frac{n}{2})}d\mu d\omega$$
$$= \frac{\beta_n^{\alpha_n}\Gamma(\alpha_n+\frac{n}{2})}{\Gamma(\alpha_n)\sqrt{2\pi}(\beta_n+\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2)^{\alpha_n+\frac{1}{2}}}$$

for the case of n=1

$$\propto C(\beta_n+\frac{1}{2}(\mu-y)^2)^{-\alpha_n-\frac{1}{2}}\int\int pdf_{Normal-Gamma} \propto \left(1+\frac{1}{\nu}\frac{(\mu-m)^2}{s^2}\right)^{-\frac{\nu+1}{2}}$$

where:

$$\nu = 2\alpha_n$$
$$m = \mu_n$$
$$\frac{1}{\nu s^2} = \frac{1}{2\beta_n} \Rightarrow s = \sqrt{\frac{\beta_n}{2\alpha_n^2}}$$

can be written as

$$P(new/old) = p(new\&old)/P(old)$$

Solution End

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

Solution:
This is special case of 2.5, same results but with the prior for normal gamma. Results is the same with 0 instead of n indices:

$$P(\mu,\omega|y) = \frac{P(y,\mu,\omega)}{P(y)}$$
$$P(\mu,\omega,y) = Nomrmal-Gamma(\mu,\omega|\mu_0,\kappa_0,\alpha_0,\beta_0)\prod_{i=1}^{n}N(y|\mu,\omega)$$
$$P(y) = \frac{P(y,\mu,\omega)}{P(y)} = \frac{\Gamma(\alpha_n)}{\Gamma(\alpha_0)}\frac{\beta_n^{\alpha_n}}{\beta_0^{\alpha_0}}\frac{\kappa_n}{\kappa_0}(2\pi)^{-n/2}$$

Solution End

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.*

Solution:
With $d = 1$, $\Sigma = \sigma^2$ and $\Lambda_0 \& \nu_0$ are scalers. So,

$$p(\Sigma(or\sigma^2)|\nu_0, \Lambda_0^{-1}) = \frac{\Lambda^{\nu_0/2}}{2^{\nu_0/2}\Gamma(\nu_0/2)}\Sigma^{-\frac{\nu_0+2}{2}}e^{-\frac{1}{2}\Lambda\Sigma^{-1}} = \frac{(\frac{\Lambda}{2})^{\nu_0/2}}{\Gamma(\nu_0/2)}\Sigma^{-\frac{\nu_0+2}{2}}e^{-\frac{\Lambda}{2}\Sigma^{-1}}$$

Which is an inverse Gamma $(\frac{\nu_0}{2}, \frac{\Lambda}{2})$.
Solution End

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:

$$P(\mu, \Sigma|y_i) \propto P(y_i|\mu, \Sigma)P(\mu, \Sigma)$$

$$\propto |\Sigma|^{-\frac{\nu_0+d+1+n}{2}} e^{-\frac{1}{2}\text{tr}(\Lambda_0\Sigma^{-1})+\frac{\kappa_0}{2}(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)-\frac{1}{2}\sum_{i=1}^{N}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu)}$$

$$= |\Sigma|^{-\frac{\nu_0+d+1+n}{2}}$$

$$e^{-\frac{1}{2}\text{tr}(\Lambda_0\Sigma^{-1})+\frac{\kappa_0+n}{2}(\mu-\frac{\mu_0\kappa_0+\sum_{i+1}^{n}y_i}{\kappa_0+n})^T\Sigma^{-1}(\mu-\frac{\mu_0\kappa_0+\sum_{i+1}^{n}y_i}{\kappa_0+n})-\frac{1}{2}\sum_{i=1}^{N}(y_i)^T\Sigma^{-1}(y_i)-\frac{\kappa_0+n}{2}(\frac{\mu_0\kappa_0+\sum_{i+1}^{n}y_i}{\kappa_0+n})^T\Sigma^{-1}(\frac{\mu_0\kappa_0+\sum_{i+1}^{n}y_i}{\kappa_0+n})}$$

$$= \Sigma|^{-\frac{\nu_0+d+1+n}{2}}$$

$$e^{-\frac{1}{2}\text{tr}(\Lambda_0\Sigma^{-1}+\frac{1}{2}\sum_{i=1}^{N}(y_i)(y_i)^T+\frac{\kappa_0+n}{2}(\frac{\mu_0\kappa_0+\sum_{i+1}^{n}y_i}{\kappa_0+n})(\frac{\mu_0\kappa_0+\sum_{i+1}^{n}y_i}{\kappa_0+n})^T)+\frac{\kappa_0+n}{2}(\mu-\frac{\mu_0\kappa_0+\sum_{i+1}^{n}y_i}{\kappa_0+n})^T\Sigma^{-1}(\mu-\frac{\mu_0\kappa_0+\sum_{i=1}^{n}y_i}{\kappa_0+n})}$$

Which is a new form of $P(\mu, \Sigma)$ with parameters:

$$\Lambda_n = \Lambda_0 + \frac{1}{2}\sum_{i=1}^{N}(y_i)(y_i)^T - \frac{\kappa_0+n}{2}(\frac{\mu_0\kappa_0+\sum_{i=1}^{n}y_i}{\kappa_0+n})(\frac{\mu_0\kappa_0+\sum_{i=1}^{n}y_i}{\kappa_0+n})^T$$

which can be expressed as:

$$\Lambda_n = \Lambda_0 + \frac{1}{2}\left\{YY^T - \kappa_n\mu_n\mu_n^T\right\}$$

$$\nu_n = \nu + n$$

$$\kappa_n = \kappa + n$$

$$\mu_n = \frac{\mu_0\kappa_0 + \sum_{i=1}^{n}y_i}{\kappa_0 + n}$$

Solution End

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:
Start from:

$$P(\beta,\omega) = \frac{|\omega K|^{\frac{d}{2}}}{\sqrt{2\pi}^d} e^{-\frac{1}{2}(\beta-\mu)^T \omega K(\beta-\mu)} \frac{b^a e^{-\omega b}\omega^{a-1}}{\Gamma(a)} = \frac{|K|^{\frac{d}{2}} b^a \omega^{a-\frac{d}{2}} e}{\Gamma(a)\sqrt{(}2\pi)^d} e^{-\omega(\frac{1}{2}(\beta-\mu)^T K(\beta-\mu)+b)}$$

$$\begin{aligned} P(\omega,\beta|y_1,\ldots,y_n) &\propto P(y_1,\ldots,y_n|\beta,\omega)P(\beta,\omega) \\ &= \frac{\omega^{\frac{n}{2}}|\Lambda|^{\frac{n}{2}}}{\sqrt{2\pi}^n} e^{\frac{-\omega}{2}(\mathbf{y}-X\beta)^T\Lambda(\mathbf{y}-X\beta)} \frac{|K|^{\frac{d}{2}} b^a \omega^{a-\frac{d}{2}} e}{\Gamma(a)\sqrt{(}2\pi)^d} e^{-\omega(\frac{1}{2}(\beta-\mu)^T K(\beta-\mu)+b)} \\ &= \frac{\omega^{\frac{n}{2}}|\Lambda|^{\frac{n}{2}}}{\sqrt{2\pi}^n} \frac{|K|^{\frac{d}{2}} b^a \omega^{a-\frac{d}{2}} e}{\Gamma(a)\sqrt{(}2\pi)^d} e^{\frac{-\omega}{2}((\mathbf{y}-X\beta)^T\Lambda(\mathbf{y}-X\beta)+(\beta-\mu)^T K(\beta-\mu)+2b)} \end{aligned}$$

Looking into the exponential term:

$$\begin{aligned} (\mathbf{y}-X\beta)^T\Lambda(\mathbf{y}-X\beta)+(\beta-\mu)^T K(\beta-\mu)+b &= Y^T\Lambda Y - 2\beta^T X^T AY + \beta^T X^T\Lambda X\beta \\ &\quad +\beta^T K\beta - 2\beta^T K\mu + \mu^T K\mu \\ = (\beta-(X^T\Lambda X+K)^{-1}(K\mu+X^T\Lambda Y))^T&(X^T\Lambda X+K)(\beta-(X^T\Lambda X+K)^{-1}(K\mu+X^T AY)) \\ -(K\mu+X^T\Lambda Y)^T(X^T\Lambda X+K)^{-1}&(K\mu+X^T\Lambda Y)+b+\mu^T K\mu+Y^T\Lambda Y+b \\ = (\beta-\mu_n)^T(K_n)(\beta-\mu_n)&+\mu_n^T K_n\mu_n - \mu^T K\mu + Y^T\Lambda Y + b \end{aligned}$$

So

$$P(\omega,\beta|y_1,\ldots,y_n) \propto \omega^{a+\frac{n-d}{2}} e^{\frac{-\omega}{2}((\beta-\mu_n)^T(K_n)(\beta-\mu_n)+\mu_n^T K_n\mu_n+\mu^T K\mu+Y^T\Lambda Y+2b)}$$

which has the same form of $P(\omega, \beta)$ with the following updated params:

$$\mu_n = (X^T\Lambda X+K)^{-1}(K\mu+X^T\Lambda Y)$$

$$K_n = X^T\Lambda X + K$$

$$a_n = a + \frac{n}{2}$$

$$b_n = \frac{1}{2}(-(K\mu+X^T\Lambda Y)^T(X^T\Lambda X+K)^{-1}(K\mu+X^T\Lambda Y)+\mu^T K\mu+Y^T\Lambda Y)+b = \frac{1}{2}(-\mu_n^T K_n\mu_n+\mu^T K\mu+Y^T\Lambda Y)+b$$

Solution End

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

Solution:

$$
\begin{aligned}
P(\omega|\mathbf{y}) &= \int P(\omega, \beta|y_1, \ldots, y_n) d\beta \\
&= \frac{b_n^{a_n} e^{-\omega b_n} \omega^{a_n - 1}}{\Gamma(a_n)} \int \frac{|\omega K_n|^{\frac{d}{2}}}{\sqrt{2\pi}^d} e^{-\frac{1}{2}(\beta-\mu)^T \omega K_n (\beta-\mu)} d\beta \\
&= Gamma(a_n, b_n)
\end{aligned}
$$

Solution End

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

Solution:
$P(\beta|\mathbf{y})$:

$$
\begin{aligned}
P(\beta|\mathbf{y}) &= \int P(\omega, \beta|y_1, \ldots, y_n) d\omega \\
&= \int \frac{|K_n|^{\frac{d}{2}} b_n^{a_n} \omega^{a_n - \frac{d}{2}}}{\Gamma(a_n)\sqrt{(2\pi)^d}} e^{-\omega(\frac{1}{2}(\beta-\mu_n)^T K_n(\beta-\mu_n)+b_n)} d\omega \\
&= \frac{|K_n|^{\frac{d}{2}} b_n^{a_n}}{\Gamma(a_n)\sqrt{(2\pi)^d}} \int \omega^{(a_n+\frac{d}{2})-1} e^{-\omega(\frac{1}{2}(\beta-\mu_n)^T K_n(\beta-\mu_n)+b_n)} d\omega \\
&= \frac{|K_n|^{\frac{d}{2}} b_n^{a_n} \Gamma(a_n + \frac{d}{2})}{\Gamma(a_n)\sqrt{(2\pi)^d}(\frac{1}{2}(\beta-\mu_n)^T K_n(\beta-\mu_n)+b_n)^{(a_n+\frac{d}{2})}} \int beta(\alpha_n + \frac{d}{2}, (\beta-\mu_n)^T K_n(\beta-\mu_n) + b_n d\omega = \\
&\frac{|K_n|^{\frac{d}{2}} b_n^{a_n} \Gamma(a_n + \frac{d}{2})}{\Gamma(a_n)\sqrt{(2\pi)^d}(\frac{1}{2}(\beta-\mu_n)^T K_n(\beta-\mu_n)+b_n)^{(a_n+\frac{d}{2})}} \propto \left\{\frac{1}{2b_n}(\beta-\mu_n)^T K_n(\beta-\mu_n)+1\right\}^{-(a_n+\frac{d}{2})}
\end{aligned}
$$

which is the form of multivariate t distribution.
Solution End

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?*

Solution:
The code can be found on Github under python directory with name 'Ch2_12.py'

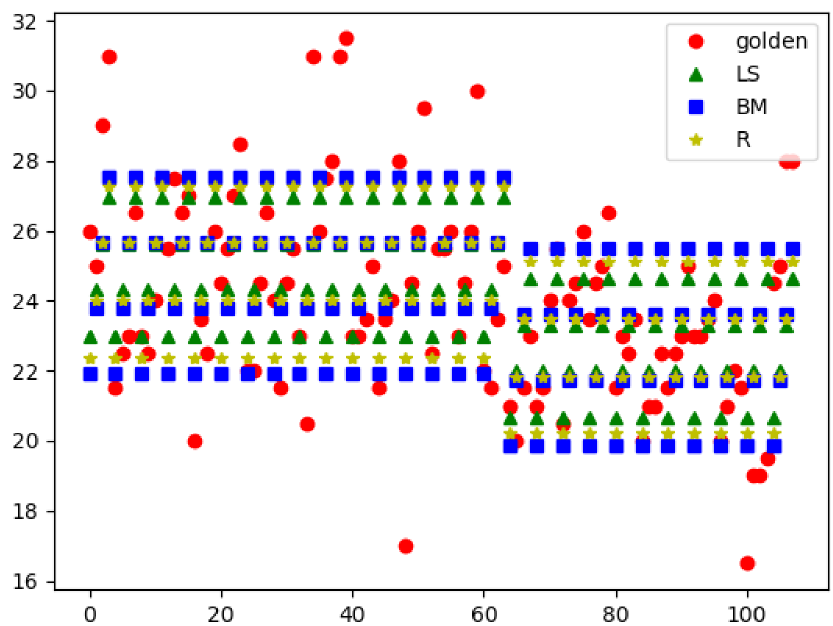

Figure 2.1: Comparison of Models from least square, ridge regression and Bayesian model.

Solution End

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

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

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

Solution:
We know that β and ω are independent of λ_i, then we can write:

$$P(\lambda_i|\mathbf{y},\beta,\omega) \propto P(\mathbf{y},\beta,\omega|\lambda_i)\prod_{i=1}^{n}(\lambda_i) = P(\mathbf{y}|\beta,\omega,\lambda_i)P(\beta,\omega|\lambda_i)\prod_{i=1}^{n}(\lambda_i)$$

Using independence information:

$$P(\lambda_i|\mathbf{y},\beta,\omega) \propto P(\mathbf{y}|\beta,\omega,\lambda_i)P(\beta,\omega)\prod_{i=1}^{n}(\lambda_i)$$

From 2.9 we know that:

$$P(\mathbf{y}|\beta,\omega,\lambda_i)P(\beta,\omega) \propto \frac{\omega^{\frac{n}{2}}(\prod_{i=1}^{n}\lambda_i)^{\frac{n}{2}}}{\sqrt{2\pi}^{n}}\frac{|K|^{\frac{d}{2}}b^a\omega^{a-\frac{d}{2}}e}{\Gamma(a)\sqrt{(2\pi)^d}}e^{\frac{-\omega}{2}(\sum_{i=1}^{n}(\lambda_i(\mathbf{y}-X\beta)^T(\mathbf{y}-X\beta))+(\beta-\mu)^TK(\beta-\mu)+b)}$$

$$P(\lambda_i|\mathbf{y},\beta,\omega) \propto \frac{\omega^{\frac{n}{2}}(\prod_{i=1}^{n}\lambda_i)^{\frac{n}{2}}}{\sqrt{2\pi}^{n}}\frac{|K|^{\frac{d}{2}}b^a\omega^{a-\frac{d}{2}}e}{\Gamma(a)\sqrt{(2\pi)^d}}e^{\frac{-\omega}{2}(\sum_{i=1}^{n}(\lambda_i(\mathbf{y}-X\beta)^T(\mathbf{y}-X\beta))+(\beta-\mu)^TK(\beta-\mu)+b)}$$

$$\prod_{i=1}^{n}\frac{\tau^\tau e^{-\lambda_i\tau}\lambda_i^{\tau-1}}{\Gamma(\tau)} \propto e^{\sum_{i=1}^{n}-\lambda_i\tau}\prod_{i=1}^{n}\lambda_i^{\tau-1}\prod_{i=1}^{n}\lambda_i^{\frac{n}{2}}e^{(\sum_{i=1}^{n}(-\lambda_i\frac{\omega}{2}(\mathbf{y}-X\beta)^T(\mathbf{y}-X\beta)))}$$

$$\propto \prod_{i=1}^{n}\lambda_i^{\tau+\frac{n}{2}-1}e^{\sum_{i=1}^{n}-\lambda_i(\frac{\omega}{2}(\mathbf{y}-X\beta)^T(\mathbf{y}-X\beta)+\tau)}$$

Using the assumption of independence, the conditional posterior of λ_i follows iid $Gamma(\tau+\frac{1}{2},(\frac{\omega}{2}(y_i-X\beta)^T(y_i-X\beta)+\tau))$

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.*

Solution:
The code can be found on Github under python directory with name 'Ch2_14-15.py'. This code is for both exercises 14&15.

Solution End

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.*

Solution:
The code can be found on Github under python directory with name 'Ch2_14-15.py'. This code is for both exercises 14&15.

It is clear that data corresponding to males and females are different. This is missed in the default implementation of the samples. A way that can better fit the data is to use two models instead of one: a Male model and a Female model.

In the figure below we show the comparison for all models applied for this data:

- LS: least square
- R: ridge
- BM: Bayesian Model
- HM: Gibbs samples single model

- HM_male: Gibbs samples male-only model
- HM_female: Gibbs samples female-only model

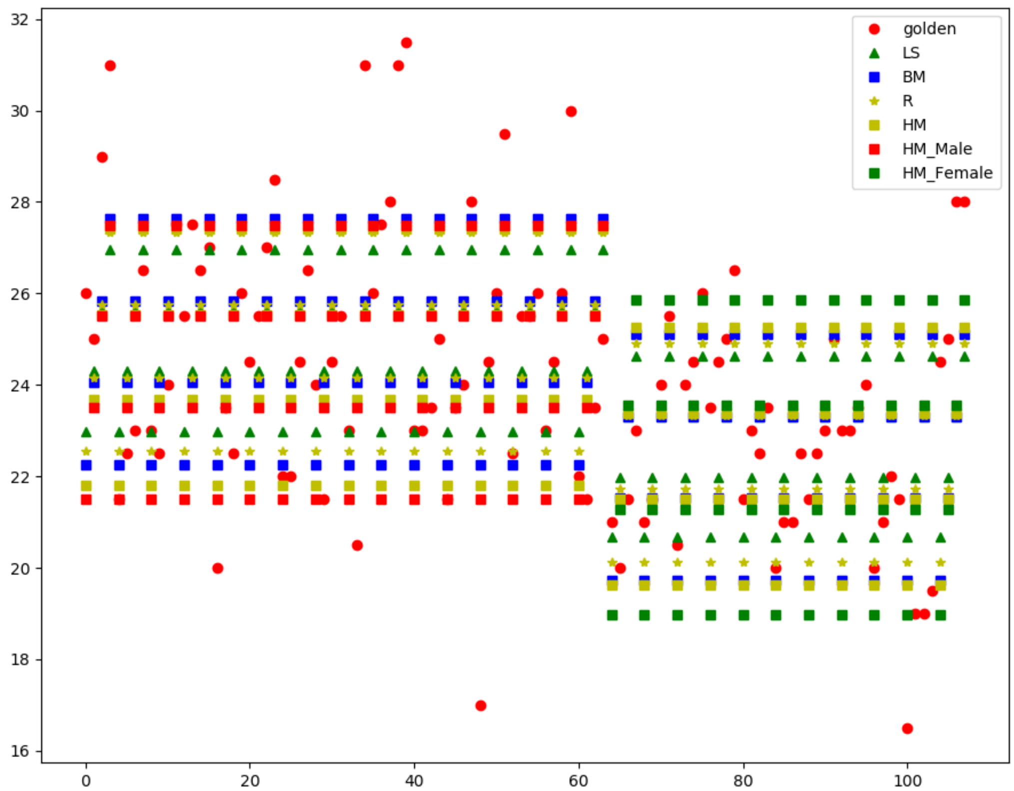

Figure 2.2: Comparison of Models from least square, ridge regression, Bayesian model and three models based on Gibbs Sampling.